



UNIVERSITAT DE
BARCELONA

DISSENY DE RECERCA

Dissenys experimentals i quasi- experimentals

Idees bàsiques per al seu ús en Psicologia

Febrer 2024

Dr. Manel Viader Junyent
Secció de Psicologia Quantitativa
Facultat de Psicologia

INDEX

I Aspectes generals (pàg. 5)

- 1 Introducció a la metodologia experimental
- 2 Les variables experimentals
- 3 Variables de confusió, control experimental i validesa
- 4 Principals tipus de dissenys experimentals i quasi-experimentals. Classificació inicial
- 5 Consideracions estadístiques

II Dissenys experimentals clàssics o estadístics (pàg. 47)

Aspectes generals

- 1 Dissenys de grups aleatoritzats (DGA)
- 2 Reducció de l'error experimental i control de variables de confusió. Anàlisi de covariància. Dissenys de grups homogenis (DGH)
- 3 Dissenys intra-subjecte o de mesures repetides (DMR)

III Dissenys quasi-experimentals (pàg. 99)

Aspectes generals

- 1 Dissenys quasi-experimentals transversals
2. Dissenys quasi-experimentals de sèries temporals

IV Dissenys de replicació intra-subjecte o N = 1 (pàg. 121)

- 1 Introducció
- 2 Terminologia general i alguns aspectes metodològics dels dissenys N=1
- 3 Classificació i revisió dels principals dissenys de replicació intra-subjecte
- 4 Anàlisi de dades en dissenys N=1

I Aspectes generals

1.- Introducció a la metodologia experimental

La complexitat dels fenòmens psicològics fa necessari tenir a disposició diferents formes i instruments de recerca. La investigació psicològica s'articula entorn de tres idees o objectius fonamentals: Identificar i mesurar les variables d'interès, planificar les actuacions necessàries per obtenir resultats i analitzar aquests resultats de forma que ens ajudin a arribar a conclusions. Esquemàticament:

Elements bàsics de la recerca en Psicologia

Element	Objectius	Tipus de validesa implicada
Mesura	Identificació, definició i mesura de variables	Validesa de constructe
Disseny	Planificació destinada a obtenir resultats vàlids	Validesa interna Validesa externa
Anàlisi	Estudi de resultats a partir de tècniques estadístiques	Validesa de conclusió estadística

El contingut d'aquest document es refereix sobretot al segon dels elements que s'indiquen, però també necessita fer referència al tercer per tal de mostrar com es poden analitzar els resultats que s'hagin obtingut a partir de l'ús dels diferents dissenys possibles. No obstant això, no es pot perdre de vista tampoc el tema de la mesura de les variables, ja que les característiques de cadascuna de les variables utilitzades té repercussions importants en aspectes com la selecció de valors, el registre de les variables o l'anàlisi dels resultats.

La diversitat dels fenòmens que es poden estudiar en Psicologia i l'existència de diferents perspectives i objectius de la recerca porten necessàriament a l'existència de diverses estratègies de recerca, que es poden agrupar en tres grans nuclis: Un plantejament descriptiu, destinat principalment a descriure les característiques de les variables i dels fets estudiats; un plantejament associatiu, on es poden formular i contrastar hipòtesis sobre la covariació entre diferents variables; i finalment un plantejament manipulatiu, on es formulen i es valoren hipòtesis de caràcter causal. Aquesta darrera perspectiva fa ús de dissenys de caràcter experimental o quasi-experimental, que són els que seran tractats a les pàgines següents. L'ús d'una o altra estratègia té també conseqüències importants pel que fa al nivell de control que tindrem de les diferents situacions.

De forma esquemàtica:

Tres grans estratègies de recerca

Estratègia	Hipòtesi	Grau de control	Dissenys
Descriptiva		Mínim o baix	Observacionals Selectius Comparatius
Associativa	Covariació	Mitjà	Predictius Explicatius Altres (terminologia diversa)
Manipulativa	Causal	Alt	Experimentals clàssics Experimentals N=1 Quasi-experimentals

En aquest context, la metodologia experimental intenta abordar els fenòmens psicològics a partir d'un enfocament que es basa en dues característiques fonamentals:

- 1.- La manipulació de determinades variables per intentar valorar el seu impacte sobre d'altres, i per establir d'aquesta forma una relació causal entre unes i altres
- 2.- L'assoliment d'un nivell de control de la situació suficient com per pensar que la relació entre variables causa i variables efecte és real, i no està contaminada o influïda per altres variables que puguin actuar simultàniament

L'ús d'aquesta metodologia implica, doncs, la identificació d'una o diverses variables causals, que poden rebre denominacions diverses: variables antecedents, independents, predictores, explicatives, etc.; i també, naturalment, la determinació d'una o diverses variables efecte, on esperem veure el possible impacte de la manipulació de les variables causa. Aquestes variables efecte reben també denominacions diverses, com variables conseqüents, variables de resposta, variables criteri o variables dependents, entre d'altres.

Per a major simplicitat, a la resta del text designarem les variables causa com a **variables independents** i les variables efecte com a **variables dependents**.

És important assenyalar des del principi que la manipulació de la variable o variables independents és un element definitori fonamental, ja que permet distingir entre els estudis que són experimentals (i que inclouen aquesta manipulació) i els que no. Cal definir des del principi i amb claredat què significa el terme "manipulació": manipular una variable independent significa que l'investigador és qui determina els valors que ha d'assumir aquesta variable en l'experiment, i qui ha d'establir les actuacions necessàries per posar en joc aquests valors en la situació experimental. Proposem un cas senzill: Si un investigador vol saber com afecta l'interval

de temps al record de llistes de paraules, haurà de decidir en primer lloc quins intervals vol utilitzar. Per exemple, pot decidir que entre la presentació de cada llista i la prova de record passin 10, 20 o 30 segons. Aquests seran doncs els tres valors que assumirà la variable independent en aquest experiment. La segona cosa que haurà de fer és organitzar l'estudi de manera que aquests tres valors es posin en joc: en alguns assajos l'interval serà de 10 segons, en altres de 20 i en altres de 30. Això es pot aconseguir fàcilment amb el suport informàtic necessari i programant la seqüència d'assajos. Naturalment, a partir d'aquí caldrà executar l'estudi, registrar els resultats i, lògicament, comparar el percentatge de record obtingut per a cadascun d'aquests valors de la variable independent.

És evident, doncs, que en una situació experimental els valors d'una variable independent qualsevol venen predeterminats per la selecció que fa l'investigador. El més habitual és que es seleccioni un número relativament petit de valors de la variable, d'entre tots els valors possibles, tot i que, com es veurà més endavant, també pot haver-hi casos on s'utilitzin tots els valors possibles de la variable, sempre que el seu número no sigui molt elevat. Pel que fa a la variable o variables dependents, clarament els seus valors no estan predeterminats ni són coneguts abans de fer l'experiment (tot i que podem tenir una idea prèvia del rang de valors possibles); els valors que presenti la variable dependent per a cada subjecte experimental no són sinó els **resultats** de l'experiment, i és en aquests resultats on podrem estudiar l'impacte dels diversos valors de la variable o variables independents. En el cas proposat abans, l'investigador determina que utilitzarà tres valors de la variable independent (10, 20 i 30 segons d'interval de record), tot i que òbviament hi ha altres valors possibles. Els resultats de l'experiment consisteixen en el percentatge de paraules recordades per cada subjecte experimental, i aquest són, naturalment, els valors de la variable dependent estudiada.

Naturalment, el disseny d'un estudi de caràcter experimental es fa sempre per alguna raó concreta, amb uns objectius determinats, ja siguin de tipus teòric (contrastar una hipòtesi derivada d'un cert marc teòric, provar les conseqüències contrastables d'un determinat model, etc.) o pràctic (per exemple, valorar l'eficàcia de diferents tractaments clínics, o de determinades intervencions psicosocials, etc.). Normalment els estudis pràctics tenen també un component teòric més o menys implícit, ja que les intervencions proposades haurien de tenir alguna base conceptual prèvia.

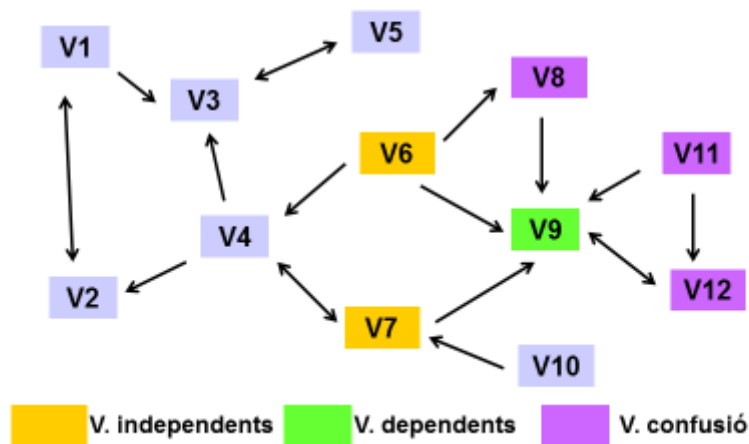
En definitiva, la recerca experimental intenta descobrir lleis o regularitats a partir de l'establiment de relacions causals entre variables. La manipulació de certes variables pot produir eventualment un efecte sobre altres. L'experiment intenta detectar aquest possible efecte, determinar la seva magnitud, veure si hi ha factors que poden modificar-lo, etc. Per exemple, pot ser de molt interès esbrinar si hi ha determinades característiques individuals que modifiquen l'impacte de les variables independents (per exemple: l'eficàcia de l'ús de programes d'estímul cognitiu depèn de l'edat dels subjectes?). Com es veurà més endavant, així com les variables externes a la persona normalment es poden manipular, les característiques individuals dels subjectes experimentals no són manipulables, sinó que l'únic que podrem fer és classificar aquests subjectes en funció de les característiques que ens puguin ser d'interès (edat, extraversió, intel·ligència, etc.).

Com s'ha indicat, el concepte de control experimental es refereix a la necessitat d'evitar que altres variables (diferents de les independents) puguin tenir un impacte sobre els resultats (és a dir, sobre els valors registrats de les variables dependents), de forma que puguin confondre's amb el possible efecte de les variables independents. Per aquesta raó les variables que poden

influir en les VD i que no són les que ens interessin (VI) s'anomenen variables de confusió o variables estranyes, i l'obligació de l'experimentador és intentar controlar-les en la mesura en què sigui possible.

Esquemàticament:

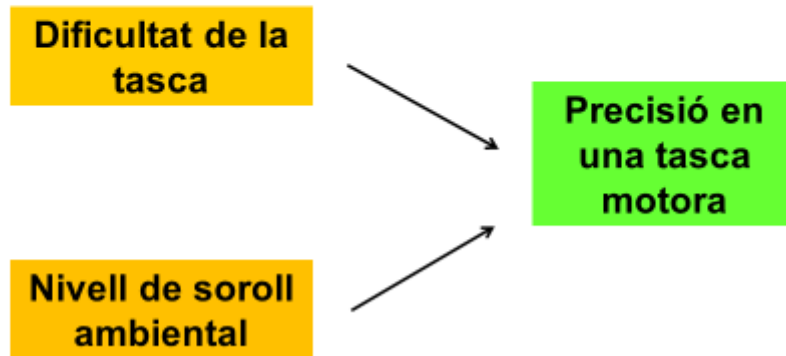
Variables independents, dependents i de possible confusió



A qualsevol situació psicològica hi ha un gran número de variables que poden manifestar-se d'una o altra forma, ja siguin variables ambientals o de la situació, ja siguin variables personals (característiques de les persones, història vital, etc, etc.). El que es fa en una situació experimental és *simplificar* en alguna mesura aquesta situació complexa, i seleccionar un conjunt concret de variables que volem estudiar. En aquest esquema, identifiquem una variable dependent (V9) on esperem detectar el possible efecte de dues variables independents (V6 i V7). Al mateix temps, hi ha altres variables (V8, V11, V12) que, sense ser manipulades per la persona que fa la recerca, poden tenir també una incidència directa sobre els resultats de V9 i, per tant, poden actuar com a variables de confusió. El que haurem de fer és, doncs, *manipular* les variables independents, *registrar* els resultats de la variable dependent i *controlar* d'una o altra forma el possible impacte de les variables de confusió.

Un exemple senzill d'estudi experimental, relacionat amb l'exactitud de la realització d'una tasca motora, amb dues variables independents i una variable dependent, seria el següent:

Exemple d'un estudi experimental



Lògicament en aquest cas, com en qualsevol altra, caldrà valorar quina o quines variables de confusió podrien incidir sobre els resultats, tant pel que fa a variables externes al subjecte (característiques de l'entorn, possibilitat de presència d'estímuls distractors, característiques de la tasca experimental, etc, etc.) com a característiques pròpies de la persona (diferències individuals en habilitat motora, capacitat de concentració, motivació, etc.).

Com qualsevol altre tipus de recerca, un estudi experimental segueix, com a mínim idealment, una certa seqüència lògica, que es pot representar fàcilment de forma esquemàtica:

Fases d'una recerca

Determinació d'objectius: Què volem saber?

Quines expectatives tenim? (hipòtesis). En què es basen?. Estudis exploratoris i confirmatoris

Planificació de la recerca en funció dels objectius: Variables, participants, controls, etc.

Execució de la recerca i obtenció de resultats

Anàlisi de resultats. Valoració a partir dels objectius i hipòtesis plantejades

Com ja s'ha indicat, un estudi experimental es pot planificar i executar a partir d'objectius que poden ser tant teòrics com pràctics (o una barreja dels dos). En tot cas, si el marc teòric o l'evidència empírica ho permeten, moltes vegades els investigadors poden tenir una hipòtesi de partida sobre els resultats esperats, de forma que l'experiment és de caràcter **confirmatori**, és a dir, destinat a confirmar (o no) la hipòtesi establerta. Si el baix nivell de coneixement existent o la manca d'un marc teòric potent no permeten establir una hipòtesi de partida, l'experiment es pot considerar **exploratori**.

En el moment de planificar la recerca s'estableix quin és el disseny que es vol utilitzar. El disseny no és sinó un conjunt de decisions sobre el tractament de les diferents variables importants, els elements de control a introduir, els criteris de selecció dels subjectes participants, etc. Als apartats següents es farà referència als aspectes més importants relatius al disseny d'experiments.

Naturalment, una vegada feta la planificació (disseny) de l'estudi, el pas següent és procedir a la seva execució, recollir els resultats corresponents i procedir a la seva anàlisi (habitualment mitjançant tècniques estadístiques) i valoració.

És fonamental la coherència global entre les diferents fases de la recerca. Per exemple, la valoració de resultats s'ha de fer considerant el fonament de l'estudi i les possibles hipòtesis establertes. La selecció del disseny s'ha de plantejar en funció dels objectius de l'estudi i característiques de la situació estudiada, etc.

2.- Les variables experimentals

D'acord amb l'indicat a l'apartat anterior, en un estudi experimental treballarem amb tres grans tipus de variables, cadascuna d'elles amb el seu sentit i objectius:

Variabls experimentals

Independents	<ul style="list-style-type: none">• Manipulació• Execució
Dependents	<ul style="list-style-type: none">• Registre• Anàlisi de resultats
De possible confusió	<ul style="list-style-type: none">• Identificació• Control

. Variables independents

Les variables independents poden ser de diferents tipus: Poden tenir relació amb característiques ambientals (per exemple: com influeixen la lluminositat o el soroll ambiental en la realització d'una determinada tasca?); amb aspectes estimuladors concrets (p.e., influeix la grandària del tipus de lletra en la rapidesa de lectura d'un text?); en una simulació de conducció,

es reacciona més o menys ràpidament en funció del color dels senyals de trànsit?, etc.); amb la dificultat de la tasca que han de realitzar els participants en l'estudi (p.e., són més difícils els problemes d'adquisició de conceptes basats en la regla conjuntiva o en la disjuntiva?), etc. De vegades la variable independent té un caràcter complex, en el sentit que la definim a partir d'un conjunt d'elements: per exemple, un programa de prevenció contra les drogodependències es pot compondre de diferents elements (accions de difusió i publicitat, xerrades en escoles, etc.), però moltes vegades el considerarem com un tot i valorarem el seu impacte, per exemple, comparant els resultats d'un grup al qual s'ha aplicat el programa amb els d'un altre grup on no s'ha implantat. En aquest cas els valors de la variable són dos: Aplicació del programa (grup experimental) versus No aplicació (grup control), entenem per tant que el programa s'aplica i es valora de forma global.

És possible treballar amb variables pròpies del subjecte com si fossin variables independents?. Per exemple: És cert que els subjectes introvertits realitzen millor tasques que impliquen atenció (per exemple, vigilar un radar) que no pas els subjectes extravertits?. A la literatura experimental hi ha certa discussió sobre aquest tema, des d'autors que admeten aquesta possibilitat (parlant, per exemple, de que la extraversió-introversió és una variable independent de subjecte o "assignada") fins a d'altres que consideren que això descarta totalment el caràcter experimental de l'estudi. Està clar que en aquest cas no hi ha manipulació de la variable (cada persona és més o menys extravertida o introvertida com a característica personal, no manipulable), per tant no es pot parlar estrictament d'un estudi experimental, sinó més aviat d'un estudi associatiu o correlacional (quina relació hi ha entre el nivell d'extraversió d'una persona i el seu rendiment a una tasca de vigilància?. Això no exclou que, si tenim un fonament teòric per fer-ho, puguem pensar en una relació causal entre el factor de personalitat extraversió-introversió i el rendiment a la tasca, però no la podem establir amb el mateix nivell de seguretat o de potència que obtenim si podem manipular directament la variable independent.

Hi ha una certa confusió terminològica i conceptual en relació amb aquesta qüestió. Com a criteri general, entendrem que un estudi com el que s'ha citat (relació extraversió/introversió amb rendiment en una tasca que requereix atenció) no es pot considerar com a experimental, ja que no hi ha una manipulació real de la variable presumptament causal. Això no significa que l'estudi no es pugui fer, ni que no tingui valor, sinó que cal prudència en la seva interpretació ja que, com s'ha indicat, no podem establir una relació causal tan evident i inequívoca com en el cas de que manipulem directament la variable independent.

Ara bé, una variable de subjecte es pot combinar amb variables manipulables en un estudi experimental: per exemple, suposem que definim tres o quatre nivells de dificultat diferents en la tasca de vigilància de radar i volem establir com ho fan els subjectes extravertits i introvertits en cadascun dels casos. En aquest estudi podem parlar d'una variable independent activa o manipulable (el nivell de dificultat de la tasca) i d'una variable "classificatòria", "independent assignada", "independent no manipulable" o "predictora" (l'extraversió-introversió). Qualsevol d'aquests termes es pot trobar a la literatura experimental, i tots ells són parcialment indicatius de la naturalesa d'aquest tipus de variables. Per exemple, el terme "variable predictora" pot ser força útil, ja que deixa clar, segons quins siguin els resultats, que es podrien fer prediccions sobre la variable dependent (rendiment en la tasca, mesurat per exemple pel número d'errors) a partir de la variable extraversió-introversió, sense acceptar però una relació causal demostrada entre una i altra variable. No obstant això, cal tenir en compte que les variables independents manipulades també són variables predictores, i des d'aquest punt de vista no hi ha una diferència decisiva entre variables manipulades i no manipulades. El concepte de "variable

classificatòria” també és indicatiu, ja que per treballar amb la variable extraversió-introversió hem de *classificar* els subjectes (per exemple, en grups d’extraversió baixa, mitjana i alta) en funció d’aquesta característica, i veure si els resultats d’aquests grups a la tasca de vigilància són diferents o no. Finalment, el concepte de “variable independent no manipulable” (o assignada, o denominacions similars) també s’utilitza amb certa freqüència, ja que incideix en el fet que aquestes variables tenen alguns aspectes en comú amb les variables independents actives (per exemple, l’investigador/a selecciona els valors de la variable que li interessa treballar, aquests valors s’estableixen prèviament a la realització de l’experiment, valorem la relació entre els valors d’aquesta variable i els resultats de la variable dependent). Per tant, el problema és en part terminològic i, sobretot, de fer una interpretació acurada dels resultats, sobretot en el sentit d’evitar interpretacions causals abusives en el cas de les variables no manipulades.

En tot cas, ha de quedar clar quin és l’estatus de la variable o variables no manipulades a l’estudi. A l’exemple proposat, queda molt clar que la variable classificatòria forma part dels objectius de l’experiment, és a dir, volem estudiar-la igual que fem amb les variables actives. Explícitament volem saber si el rendiment de les persones en una tasca de vigilància es modifica en funció de la dificultat de la tasca (variable independent activa) i del nivell d’extraversió-introversió (variable independent no manipulada, o classificatòria). Fins i tot podem formular la hipòtesi de que el rendiment de les persones serà superior quan la tasca sigui fàcil, i també que el rendiment mitjà de les persones introvertides serà millor que el de les extravertides (basant-nos, per exemple, en la teoria de la personalitat d’Eysenck). Per tant, la variable classificatòria també ha de formar part de les hipòtesis experimentals que es puguin formular, perquè el nostre objectiu és *estudiar el seu impacte en els resultats*. S’ha de distingir bé aquesta situació d’aquelles altres on una variable de subjecte és simplement una variable que pot confondre els resultats (variable de confusió), de manera que el nostre únic interès seria el de tenir-la controlada precisament per evitar o neutralitzar aquest possible efecte sobre la variable dependent. Per tant, davant de la presència de qualsevol variable de subjecte en un estudi experimental, la pregunta pertinent és: *Quin és el paper d’aquesta variable?*. Es tracta d’una variable que volem *estudiar*, que forma part de la hipòtesi experimental, i de la qual volem conèixer quin és el seu impacte concret sobre la variable o variables dependents?. O bé és únicament una variable que volem *controlar*, és a dir, que sabem prèviament o suposem que podria impactar sobre la variable dependent, i el que volem és precisament eliminar o neutralitzar aquest impacte per evitar qualsevol confusió a l’hora d’interpretar els resultats?.

Pel que fa a les variables independents (manipulades o no), el nostre objectiu és aconseguir que el seu efecte sigui tan clar com sigui possible, és a dir, *maximitzar* l’efecte de la variable independent (si l’efecte existeix, naturalment), per tal de poder detectar-lo amb major facilitat i claredat. Això dependrà bàsicament de tres coses: la selecció dels valors de la pròpia variable independent, el nivell de control assolit (per evitar l’impacte de possibles variables de confusió) i una selecció correcta de la variable o variables dependents, com es veurà més endavant.

Pel que fa a la selecció dels valors de la variable independent, cal distingir diferents coses. En principi qualsevol variable independent té un número de valors possibles (que pot anar des de dos en el cas d’una variable dicotòmica fins a un número teòricament infinit en una variable quantitativa contínua), però òbviament l’experimentador només podrà treballar amb alguns d’aquests valors. Naturalment el primer pas és determinar quants valors s’utilitzaran, i això depèn tant de consideracions conceptuals com pràctiques. Conceptualment, quants més valors agafem de la variable independent major serà la riquesa de la informació obtinguda. Al mateix temps, des d’un punt de vista pràctic, a més valors de la variable independent, major complexitat

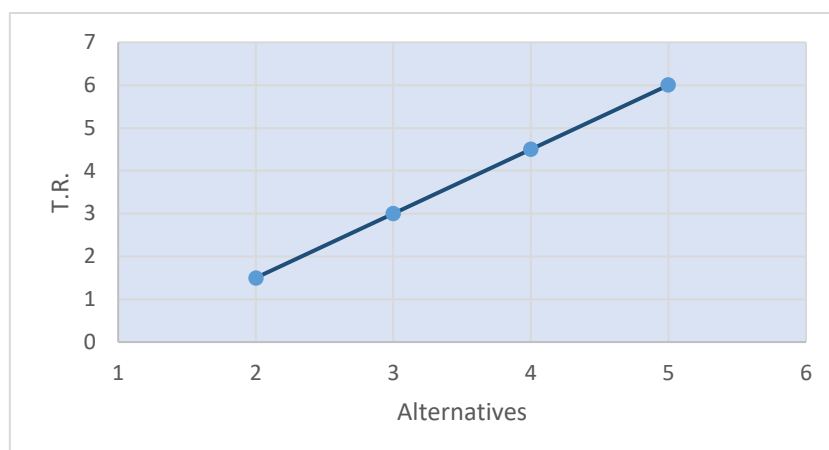
de la recerca en termes de preparació, materials, subjectes participants, temps de realització, etc. Tot això estableix alguns límits.

Una vegada decidit el nombre de valors de la variable amb els quals es vol treballar, la pregunta següent és, com escollir-los d'entre el conjunt de valors possibles de la variable?. El més habitual és que l'experimentador/a seleccioni un conjunt de valors que li interessin de forma concreta. Si, per exemple, es tracta de valorar l'eficàcia relativa de tres tipus d'intervenció (farmacològica, cognitivo-conductual, familiar) sobre un determinat problema psicològic, és evident que seran precisament aquests tres tipus d'intervenció les que es posaran en joc. És a dir, es definirà una variable independent "Tipus d'intervenció" que tindrà tres valors (farmacològica, cognitivo-conductual i familiar). Si es vol fer una rèplica de l'estudi realitzat, amb els mateixos objectius, es tornaran a utilitzar aquests mateixos valors de la variable independent. Quan es treballa amb valors concrets de la variable independent, que són els que interessin específicament a l'experimentador, diem que la selecció de valors de la variable independent s'ha fet amb un *model d'efectes fixes*, o que la variable independent és un *factor d'efectes fixes*. Una altra situació d'efectes fixes es produeix quan seleccionem tots els valors possibles d'una variable independent qualsevol: per exemple, si en la situació indicada els tres tipus d'intervenció esmentats fossin els únics possibles (i no només tres possibilitats concretes dins d'un conjunt més ampli de formes d'intervenció), també estaríem davant d'un model d'efectes fixes. En aquests casos, els resultats obtinguts són únicament vàlids per als valors concrets de la variable independent amb els quals haguem treballat.

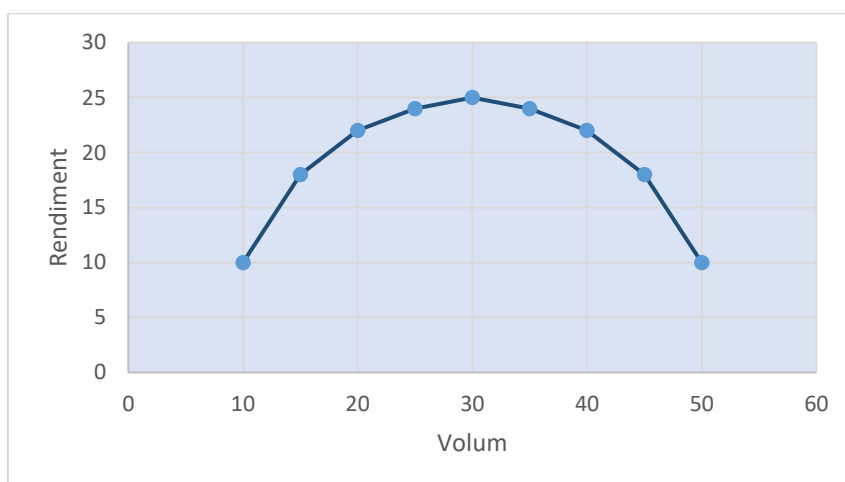
Una opció diferent seria seleccionar a l'atzar, dins d'un conjunt ampli de valors possibles, quins valors concrets utilitzarem realment. Si, per exemple, volem saber si hi ha una relació entre el nivell de soroll ambiental i el rendiment en una tasca cognitiva, podem definir un rang de valors possibles de la variable soroll (per exemple, entre 0 i 80 decibels) i escollir a l'atzar dins d'aquest rang els valors a utilitzar. Si volem emprar cinc valors de la variable independent, aquest podrien ser, per exemple: 7, 21, 41, 49 i 55 decibels. Si féssim rèpliques de l'estudi, a cadascuna d'elles tornariem a seleccionar cinc valors a l'atzar. L'objectiu d'aquest tipus d'enfocament és obtenir informació que pugui ser *generalitzable* a tot el rang de valors de la variable independent (per exemple, veure si obtenim una relació lineal negativa entre soroll i rendiment amb diferents conjunts de valors de la variable independent). Si ho fem d'aquesta manera, estem treballant amb un *model d'efectes aleatoris*, i podem dir que la variable independent és un *factor d'efectes aleatoris*.

Cal dir que a la pràctica el plantejament de variables independents d'efectes aleatoris no és molt habitual. Generalment, l'investigador selecciona els valors de la variable independent d'acord amb un criteri propi, no aleatòriament, sinó pensant en aquells valors que més li puguin interessar des d'un punt de vista teòric o pràctic. Considerant això, si no es diu el contrari, a la resta d'aquest text suposarem que es treballa amb models d'efectes fixes pel que fa a la selecció de valors de les variables independents.

La darrera pregunta seria, quins valors de la variable independent hem de seleccionar, considerant l'objectiu de maximitzar la visibilitat de l'impacte de la variable independent sobre les dependents?. No hi ha una resposta única a aquesta qüestió, ja que això depèn del tipus de relació que pugui existir entre variable independent i dependent (cosa que lògicament en molt casos no coneixem a priori). Considerem els dos casos següents:



Aquest primer exemple mostra la relació que suposadament podria existir, en una prova de reacciometria, entre el número d'alternatives de resposta i el temps de reacció obtingut. Es pot observar fàcilment que a major nombre d'alternatives, més llarg és el temps de reacció, i que la relació entre una variable i l'altra és lineal. Per tant, si en una situació donada haguéssim de triar dos valors d'aquesta variable independent per utilitzar-los en un experiment, sembla evident que seria millor escollir els valors de 2 i 5 alternatives, més que no qualsevol altre. La diferència de resultats esperable entre els valors 2 i 5 és clarament superior, per exemple, a la que obtindríem si treballem amb els valors 3 i 4. En aquest cas un bon criteri és el d'escollir **valors extrems**, dins del rang de valors possibles o d'aquells dels quals tinguem informació (no podem assegurar què passaria, per exemple, amb un número d'alternatives superior a 5).



Aquest segon gràfic mostra la possible relació existent entre el volum de la música ambiental en un determinat recinte i el rendiment de les persones en una determinada tasca (per exemple, de tipus manual). El gràfic ens indica que els volums molt baix o molt alt de la música no ajuden al rendiment de les persones i produeixen resultats baixos, en canvi valors intermedis sembla que produeixen resultats millors. La relació entre les dues variables és, doncs, curvilínia o no lineal (com es veurà més endavant, és concretament una relació quadràtica). En conseqüència, en aquest cas el criteri de valors extrems no seria adient, al contrari, clarament seria erroni, ja

que els valors extrems produeixen resultats similars. Aquí semblaria millor agafar tant valors extrems com centrals, per copsar millor la relació entre les variables.

Naturalment, en molts casos l'investigador pot no conèixer, ni tan sols aproximadament, com és la relació entre la variable independent i la dependent, i per tant no té un criteri clar sobre quins valors de la variable independent escollir. En aquests casos la literatura prèvia pot ajudar, però en tot cas és prudent utilitzar un número no massa petit de valors de la variable, ja que si només n'escollim dos, per exemple, és fàcil perdre informació que sigui rellevant. En general, és recomanable escollir tres o més valors de la variable independent, si això és possible i adient a la situació, és a dir, escollir **valors múltiples**. En principi quants més valors de la variable independent tinguem en joc més informació (i més acurada) obtindrem, però això té limitacions pràctiques importants.

En molts textos els valors que pren una variable independent en un experiment s'anomenen *tractaments* de la variable. Per tant, si donem a una variable independent quatre valors, es pot dir també que la variable té quatre tractaments, i que aquests tractaments definiran quatre *condicions experimentals* diferents. Com veurem més endavant, quan treballem amb més d'una variable independent, en molts casos les condicions experimentals són definides per la combinació de valors o tractaments de les diverses variables independents.

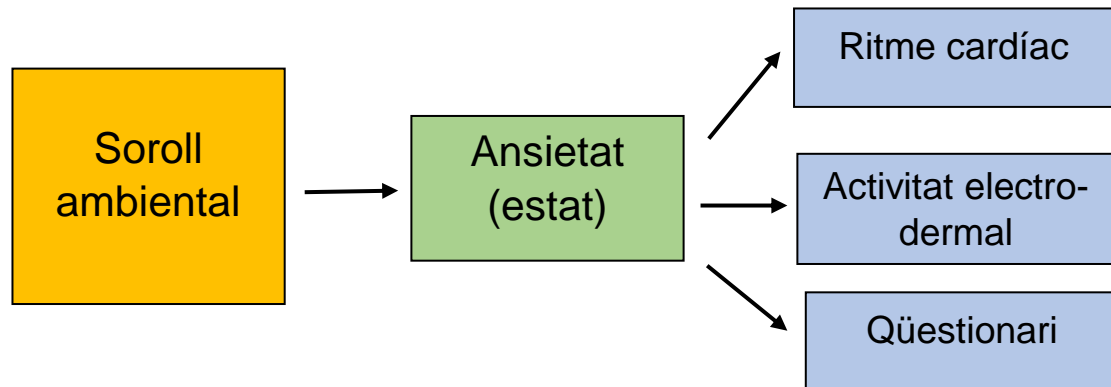
. Variables dependents

La variable o variables dependents són variables efecte que mostren el possible impacte o efecte de les variables independents. Per tant, és evident que els valors de les variables dependents no són escollits per l'investigador, sinó que constitueixen el *resultat* de l'experiment. Per tant, per a cada subjecte participant a l'experiment caldrà recollir un cert número de registres de la variable o variables dependents.

En un experiment pot haver-hi una o més variables dependents, en funció de quins siguin els seus objectius. Per exemple, en una empresa pot ser interessant valorar si determinats canvis tecnològics produeixen canvis en la productivitat. Ara bé, la productivitat té diferents aspectes o dimensions: es pot mesurar en termes de *quantitat* de productes, però també de la seva *qualitat*. Per tant, en una situació com la indicada ens pot interessar registrar la variable "quantitat diària de productes fabricats" i, al mateix temps, introduir una mesura de qualitat com el "percentatge diari de productes que no superen el control de qualitat". D'aquesta forma la informació obtinguda serà molt més acurada. En aquest cas treballarem doncs amb dues variables dependents.

A l'exemple que s'acaba de plantejar les dues variables dependents tenen relació amb un constructe únic que podem anomenar "productivitat", i cadascuna de les variables constitueix la mesura d'algun aspecte o dimensió d'aquest constructe. En aquest cas es diu de vegades que les variables dependents fan la funció d'**indicadors** dels possibles canvis que afecten aquest constructe.

Un altre exemple molt clar de la idea d'indicador es el següent:



En aquest estudi volem saber si els nivells continuats de soroll ambiental poden produir canvis en l'estat d'ansietat de les persones. Treballem amb tres variables dependents, amb la idea de que totes elles ens puguin informar sobre els canvis que es produeixen en l'estat intern d'ansietat del subjecte. Acceptant que les tres variables es relacionen amb canvis en el nivell d'ansietat, llavors si trobem que totes tres canvien de forma correlacionada en funció dels valors de la variable independent podrem argumentar que el soroll ambiental té un impacte sobre els estats d'ansietat dels subjectes experimentals. El ritme cardíac, l'activitat electro-dermal i una escala psicomètrica d'ansietat-estat són tres variables dependents que fan la funció d'**indicadors** de canvis en el que realment ens interessa, que és el nivell d'ansietat de les persones.

Com es pot veure, és possible i fins i tot freqüent (però no obligatori) que s'adjudiqui a les variables dependents la funció d'indicador d'algun procés o constructe no directament observable.

L'expectativa de l'investigador és que les variables dependents mostrin de forma clara el possible impacte de les variables independents. Per tal que això sigui així, cal escollir correctament quina o quines variables dependents s'utilitzaran i assegurar-se de que compleixen determinades condicions:

- Validesa: La variable dependent ha de reaccionar a allò que ens interessa i reflectir els processos implicats. Seguint amb l'exemple anterior, si intento establir el possible efecte del soroll ambiental sobre els nivells d'ansietat-estat dels subjectes, necessito definir una o diverses variables dependents que m'indiquin si es produeixen canvis en aquest nivell d'ansietat. Com s'ha indicat, existeixen algunes variables fisiològiques que poden fer aquesta funció, mentre que d'altres clarament no ens servirien, ja que es relacionen més amb altres processos o no tenen res a veure amb l'ansietat. El mateix es pot dir de l'ús d'un qüestionari psicomètric, del qual hem d'estar totalment segurs de que mesura realment estats o nivells d'ansietat, i no qualsevol altra cosa.
- Fiabilitat: Mentre que la validesa es refereix a que la variable dependent "mesuri allò que ha de mesurar" o "indiqui allò que ha d'indicar", la fiabilitat es refereix a la qualitat i exactitud de la mesura. Fins a quin punt un qüestionari d'ansietat-estat reacciona de forma estable i fiable davant dels canvis en la variable?. Es pot dir el mateix de l'activitat electrodermal i del ritme cardíac?. Per exemple, sempre que hi hagi un increment d'un 20% del nivell d'ansietat, cadascuna de les variables tindrà una reacció consistent davant d'aquest canvi?.

- **Sensibilitat:** Es refereix a la capacitat de la variable dependent de reaccionar davant de l'impacte de la independent. Si el canvi produït a la variable dependent en relació amb els diversos valors de la variable independent és molt petit, segurament no ens serà d'utilitat per comparar l'impacte de la V.I., ja que probablement no serem capaços d'identificar un canvi consistent, considerant a més que les dades sempre tendeixen a mostrar un cert nivell de variabilitat

Es pot resumir fàcilment tot l'indicat respecte de les variables dependents:

Idees principals en relació a les variables dependents	
Aspectes a considerar: <ul style="list-style-type: none"> ➤ Número de variables dependents (una o més d'una?) ➤ Paper de les variables dependents com a possibles indicadors de processos interns ➤ Adequació de la variable dependent: Relació amb el fenomen que es vol estudiar 	Característiques que ha de tenir una variable dependent: <ul style="list-style-type: none"> ○ Validesa ○ Fiabilitat ○ Sensibilitat

. Variables estranyes o de confusió

Com s'ha indicat anteriorment, en qualsevol situació experimental existeixen sempre altres variables, a més de les independents, que poden influir en els resultats obtinguts. Aquestes variables poden ser pròpies dels subjectes o unitats experimentals (per exemple, diferències individuals en factors de personalitat, intel·ligència, història personal, etc, etc.) o externes a ells, pròpies de l'ambient o la situació. En qualsevol situació experimental és indispensable identificar quines variables, al marge de les independents, poden influir en els resultats obtinguts. El control d'aquestes possibles **variables estranyes o de confusió** és fonamental per a la realització correcta d'un estudi experimental i per donar valor als seus resultats. Aquestes qüestions, per la seva especificitat, es tracten a l'apartat següent, dedicat al problema del control experimental i la validesa.

3.- Variables de confusió, control experimental i validesa

El control és una de les característiques que defineixen la metodologia experimental, sempre amb la finalitat d'assegurar tant com es pugui que els canvis registrats a les variables dependents són causats per la manipulació de les independents, i no per cap altra causa. A la pràctica la seguretat completa respecte d'això no és possible a la majoria dels experiments psicològics, i només podrem plantejar una bona aproximació.

És important assenyalar des del principi, com s'ha indicat abans, que hi ha dos grans tipus de possibles variables de confusió en qualsevol experiment psicològic: Les que són externes al subjecte, relacionades amb l'entorn on es realitza l'estudi, les característiques de la tasca experimental, els materials utilitzats, etc.; i les que són internes, és a dir, característiques individuals dels subjectes participants que poden influir en els resultats de l'estudi.

Control experimental

2 grans fonts de confusió:

- ▶ Variables externes o de l'entorn
- ▶ Variables pròpies del subjecte

Tenint present la consideració anterior, en el marc de la Psicologia podem observar dues concepcions diferents del control experimental, en part complementàries:

. Un concepte de control basat en les ciències naturals, que podríem anomenar **control experimental directe**, on es treballa amb molt poques unitats experimentals i en condicions ambientals molt controlades. Aquest enfocament té antecedents importants en Psicologia, com és una bona part del treball de Wundt, Fechner o Pavlov, entre d'altres. Un exemple molt conegut és l'estudi del condicionament operant a l'escola skinneriana. La forma habitual de treball consisteix en introduir un animal en una situació ambiental totalment controlada (caixa de Skinner) i valorar com diferents programes de reforçament poden produir canvis en l'aprenentatge d'aquest animal. El control assolit d'aquesta forma és molt gran, no és necessari treballar amb molts casos sinó estudiar intensivament i individualment un petit nombre d'animals, sota el supòsit de que els resultats obtinguts sota control estricte podran ser generalitzables a altres individus, a partir del concepte de regularitat de la naturalesa. És fàcil comparar aquest raonament amb el propi de la Física o d'altres ciències naturals.

És fàcil veure que aquest plantejament assegura un control molt potent de les possibles variables de confusió externes al subjecte (gran control ambiental i de situació) i també de les de caràcter intern, ja que es treballa continuadament amb un mateix subjecte o unitat experimental, les característiques individuals del qual presumiblement es mantenen constants en el temps.

Una limitació evident d'aquest plantejament a l'àmbit de la Psicologia es posa de manifest quan s'aborden fenòmens d'elevada complexitat (processos cognitius, situacions socials, etc.), on no hi ha la possibilitat d'un control tan exhaustiu com el que s'indica.

A més dels exemples que s'han citat, aquest plantejament s'utilitza també a l'àmbit aplicat, sobretot al camp de la psicologia clínica, on molt sovint és necessari treballar amb persones individuals. El problema principal és que, en aquest cas, és pràcticament impossible garantir el nivell de control que es podia aconseguir en altres situacions, ja que hi haurà moltes variables externes al subjecte que no podran ser controlades exhaustivament.

Cal dir que aquest plantejament ha donat lloc a un tipus concret de disseny experimental, que s'anomena **disseny experimental de replicació intra-subjecte o N=1**, que serà tractat a la secció IV d'aquest document.

. Un segon concepte de control, actualment predominant als experiments psicològics, és el del **control probabilístic o estadístic**. En aquest plantejament no es treballa amb un únic cas, o amb un petit nombre de casos tractats individualment, sinó amb conjunts més o menys amplis de subjectes participants. El principi bàsic és l'assignació aleatòria de molts casos a les diferents condicions experimentals i la comparació dels resultats obtinguts. Per tal de comparar les diferents condicions experimentals, el que farem normalment és constituir tants grups experimentals com condicions haguem establert, amb un cert nombre de participants a cadascun d'ells, amb la premissa fonamental que l'assignació d'una persona a un o altre grup es farà **aleatòriament** (tot i que, com es veurà, en molts casos pot ser útil establir límits a aquesta aleatorització). Si, per exemple, volem comparar l'efectivitat de tres sistemes d'aprenentatge sobre el rendiment en una determinada tasca, i disposem de 24 possibles persones participants, el que farem és construir tres grups experimentals amb 8 subjectes cadascun, repartint els 24 participants aleatòriament entre els diferents grups. Cadascun dels subjectes que formen part de cada grup constitueix una rèplica de l'estudi. D'aquesta forma, el que esperem és que les variables de confusió associades amb les diferències individuals quedin repartides de forma més o menys equitativa entre els diferents grups, és a dir, que aquests siguin **equivalents** pel que fa a aquestes variables i, per tant, la comparació de resultats finals entre els grups sigui possible. Aquesta és la forma més senzilla (veurem que n'hi ha més) amb la qual el control probabilístic intenta controlar les variables pròpies del subjecte. Aquesta forma de procedir defineix el que s'anomena habitualment **dissenys experimentals clàssics o estadístics**. La denominació "estadístics" prové del fet que gairebé sempre l'aplicació de tècniques estadístiques és la forma de valorar els resultats obtinguts a través de l'aplicació d'aquest tipus de disseny, de forma coherent amb el seu plantejament probabilístic.

Naturalment, el fet d'utilitzar aquest plantejament és totalment compatible amb un control experimental directe de les variables externes al subjecte. També en aquest cas és fonamental aconseguir que les condicions ambientals, el material utilitzat i altres variables externes de possible confusió estiguin ben controlades, i la forma més potent de fer-ho és intentar aconseguir una estandardització tan gran com es pugui de la situació i entorn experimental. Tot i això, com ja s'ha dit, en situacions complexes el control que es pot aconseguir no és tan exhaustiu com el que es plantejava a l'alternativa radical plantejada anteriorment.

D'acord amb tot l'anterior, la relació entre els diferents tipus de control i les dues grans categories de dissenys experimentals es pot esquematitzar fàcilment:

Control experimental

- ▶ Control experimental directe (Wundt, Fechner, Pavlov, Skinner.....)
- ▶ Control probabilístic

Dissenys experimentals de replicació intra-subjecte: Control directe

Dissenys experimentals clàssics: Control probabilístic + control directe

A partir de la identificació de les variables que potencialment poden confondre els resultats, cal establir els mecanismes de control necessaris per evitar el seu impacte. Això es pot fer mitjançant diferents tècniques, però el més important es veure que les tècniques de control ens ofereixen dues possibles solucions: L'*eliminació* de l'efecte de la possible variable de confusió (o la seva conversió en una *constant*) i la seva *neutralització*.

Exemples:

Eliminació/Constància: Si utilitzem una cambra insonoritzada evitem que qualsevol soroll exterior pugui distreure els subjectes participants i modificar els seus resultats. La variable "soroll ambiental" queda eliminada. En un sentit més lax, podem no "eliminar" estrictament una variable sinó deixar-la constant, de forma que el seu efecte sobre els resultats sigui també constant i no confongui els resultats: per exemple, mantenir constant la lluminositat de la sala, obtenir els registres sempre a la mateixa hora del dia, utilitzar sempre el mateix tipus de materials, etc. També es poden plantejar casos on no deixem totalment constant la variable que volem controlar, però sí que introduïm certes restriccions: per exemple, per reduir el possible impacte de l'edat dels subjectes en els resultats, podem treballar només amb persones d'entre 25 i 35 anys. El problema d'aquest tipus d'actuació en relació amb característiques del subjecte és que limitem les possibilitats de generalització dels resultats; a l'exemple plantejat, els resultats serien extrapolables només a la població de persones d'entre 25 i 35 anys d'edat. Naturalment, si prenem una decisió radical i treballem només amb persones d'una edat concreta (per exemple, 28 anys), la possibilitat de generalització encara serà menor.

Neutralització: En aquest cas no es tracta d'eliminar o deixar constant una certa variable de confusió, sinó de "distribuir-la" equitativament de forma que no confongui els resultats. Per exemple, si he de formar dos grups experimentals que hagin de rebre dos tractaments diferents, i dispenso de persones de diferents edats, una bona pràctica seria distribuir els participants de forma que les diverses edats quedin repartides de forma semblant en els dos grups. El mateix es pot dir d'altres característiques individuals que puguin ser rellevants. Fent les coses d'aquesta

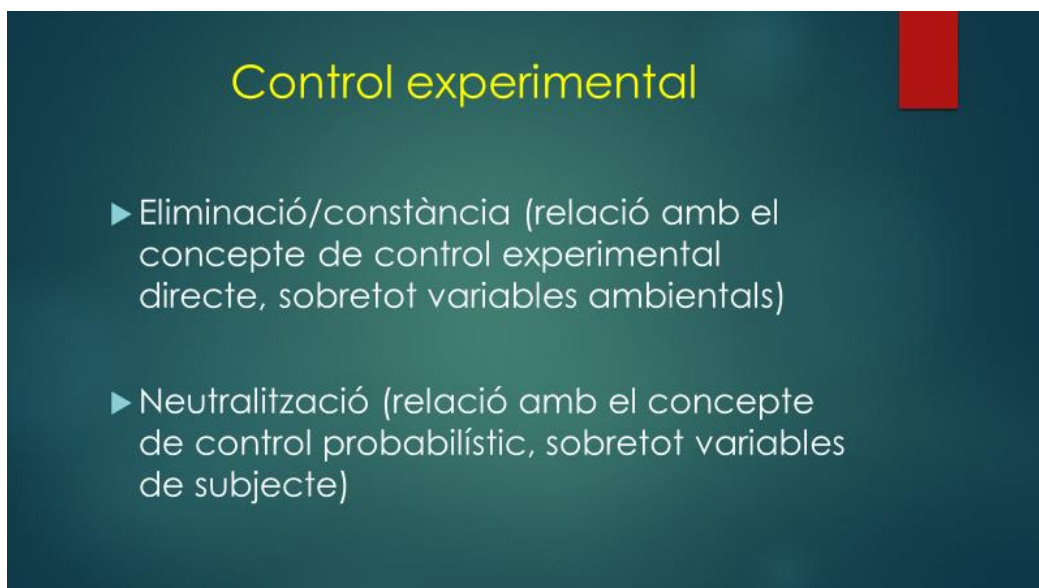
forma, la variable edat continua influint en els resultats, però ho fa de forma semblant en els dos grups, de manera que si comparem la mitjana de resultats dels grups pel que fa a la variable dependent, la diferència entre les mitjanes no s'hauria de veure afectada de forma important per la presència de persones de diferents edats. En altres paraules, i anticipant alguns aspectes que es veuran més endavant, la variable edat produiria diferències *dins dels grups o intra-grup*, però no tindria per què produir diferències *entre els grups*.

Un exemple interessant de l'ús de la neutralització és el de molts estudis neuro-psicològics i clínics, on per valorar el possible impacte d'un tractament en la millora d'un cert trastorn es treballa sovint amb un grup experimental (que rebrà tractament) i un grup control (que no el rebrà), i es controlen molt exhaustivament una sèrie de variables pròpies dels subjectes (edat, gènere, nivell d'estudis, procedència geogràfica, existència o no d'antecedents familiars, etc.), de forma que quedin repartits equitativament entre els dos grups.

En general, en el cas de variables pròpies de la situació el millor abordatge és eliminar o deixar constants totes les variables ambientals que es pugui (soroll, lluminositat, hora del dia, entorn físic on es realitza l'estudi, etc.), i fer el mateix amb tots els aspectes del procediment de realització de l'experiment (els aparells a utilitzar han de ser iguals per a tots els participants, la forma de presentació del material també, així com les característiques de la tasca, etc.), és a dir, **estandaritzar** el procediment experimental. La idea bàsica és que, si es pot, tot ha de ser igual per a tots els participants, excepte allò que realment interessa, és a dir, la variable o variables independents que volem estudiar.

En el cas de variables pròpies del subjecte, la forma més habitual de procedir és via neutralització, mitjançant un repartiment equitatiu de les diferències entre els diferents grups experimentals, tal i com ja s'ha indicat. Com s'ha indicat, també és possible fer constants algunes variables (per exemple, treballar només amb homes, o només amb subjectes amb estudis universitaris, etc.). Això és bastant freqüent, ja sigui plantejat de forma deliberada o bé per limitacions en l'accés a diferents tipologies de subjectes, tot i que òbviament cal assumir les limitacions que això suposa a l'hora de generalitzar els resultats

Esquemàticament:



Control experimental

- ▶ Eliminació/constància (relació amb el concepte de control experimental directe, sobretot variables ambientals)
- ▶ Neutralització (relació amb el concepte de control probabilístic, sobretot variables de subjecte)

Com s'ha vist, un aspecte molt important relacionat amb el control de variables de confusió pròpies dels subjectes és la manera en què es formen els diferents grups experimentals. L'opció que s'ha plantejat inicialment és la de formar aquests grups a l'atzar, però aquesta no és l'única possibilitat que tenim a la nostra disposició. Il·lustrarem aquest punt a partir d'un exemple:

Suposem que a un estudi clínic definim tres tipus de teràpies de les quals volem valorar la seva eficàcia davant d'un determinat problema o trastorn (teràpies A, B i C). En aquest tipus d'estudis les diferències individuals en relació amb diferents aspectes (edat, estudis, característiques de l'entorn familiar, etc.) poden ser molt rellevants. A grans trets, des de la perspectiva dels dissenys experimentals clàssics podem abordar aquest problema de tres formes, que ens definiran tres tipus diferents de disseny:

- Com s'indicava abans, l'opció més bàsica és repartir aleatòriament les persones participants en tres grups, cadascun dels quals rebrà una de les teràpies. Al fer això, esperem que les diferències individuals rellevants quedin repartides de forma més o menys equitativa entre els grups. S'aplica aquí, doncs, una tècnica d'*aleatorització*, i el disseny resultant és un *disseny de grups aleatoritzats*
- Repartir controladament les persones participants en tres grups, assegurant-nos de que les principals característiques individuals queden equilibrades. Per exemple, si tenim tres persones amb una edat similar, mateix nivell d'estudis i un entorn familiar similar, podem assignar-ne a l'atzar una a cadascun dels grups experimentals, i fer el mateix sistemàticament amb la resta de participants. Quan procedim d'aquesta forma apliquem una tècnica d'*aparellament* o de *bloqueig* (més endavant es veurà la diferència entre una i l'altra), i el disseny resultant és un *disseny de grups homogenis*. Una versió més dèbil d'aquest plantejament seria l'ús de la tècnica de balanceig, que consisteix simplement en equilibrar la presència d'una certa característica en els diversos grups experimentals (per exemple, si s'han de formar diversos grups, fer que a cadascun d'ells hi hagi la meitat d'homes i la meitat de dones)
- Finalment, podem optar per una solució més radical, i fer que totes les persones participants passin per les tres teràpies en moments diferents, recollint els diferents resultats obtinguts. D'aquesta manera no tindrem tres grups diferents, sinó un únic grup que passarà per totes les condicions. L'avantatge d'aquesta solució és que els diferents "grups" queden totalment igualats (ja que són el mateix grup), i l'inconvenient principal (que pot ser molt greu) són les interferències que es poden produir entre les diferents teràpies, al ser aplicades a les mateixes persones. En aquest cas estem treballant amb un *disseny de mesures repetides*, i la tècnica de control utilitzada és la del *subjecte com a control propi*

Naturalment, en el cas dels dissenys de replicació intra-subjecte el plantejament és totalment diferent, ja que al treballar amb subjectes individuals que són seguits en el temps el problema de les diferències individuals queda descartat, si més no inicialment. En aquest tipus de dissenys utilitzem també la tècnica del subjecte com a control propi.

Cal indicar que algunes de les tècniques citades són "multi ús". L'aleatorització pot ser útil a l'hora de configurar una sèrie de grups experimentals, com s'ha vist, però també ens pot ajudar a l'hora de decidir en quin ordre es presenta una llista de paraules, o quins ítems s'han d'utilitzar

a cada sessió experimental, etc. La constància es pot aplicar a un gran ventall de variables diferents: podem deixar constants variables ambientals o de procediment (lluminositat de la sala, tipus de material utilitzat, duració de la sessió, hora del dia a la qual es realitza l'experiment, etc., etc.). Altres tècniques són més específiques, com és la possibilitat d'utilitzar un grup control, i algunes només tenen sentit en dissenys concrets (per exemple, l'ús d'una línia base en un disseny de replicació intra-subjecte, o del contrabalanceig en el disseny de mesures repetides), de manera que s'estudiaran a l'abordar els dissenys corresponents.

Com és lògic, les diferents tècniques de control que estem citant s'apliquen a priori, formen part de la planificació de l'estudi i s'introdueixen abans de començar a recollir els resultats. Ara bé, disposem també d'algunes *tècniques "indirectes"* de caràcter estadístic, que ens permeten complementar l'acció de les tècniques de control aplicades a priori. Probablement la tècnica indirecta més coneguda és l'*anàlisi de covariància*, una tècnica estadística que, com es veurà més endavant, ens permetrà ajustar a posteriori uns resultats experimentals per eliminar l'efecte d'alguna variable estranya que no haguem pogut controlar prèviament. Altres tècniques estadístiques poden jugar un paper auxiliar en determinats moments del procés; per exemple, una vegada formats els grups experimentals podríem fer registres pre-test (registres de la variable dependent obtinguts abans d'introduir els tractaments experimentals) i comprovar mitjançant una prova estadística si els grups són equivalents.

El nivell de control que es pugui obtenir en un estudi experimental determina de forma directa el seu nivell de **validesa**. Direm que un experiment és vàlid quan podem assegurar que s'han manipulat correctament les variables independents, que s'han registrat de forma adient les variables dependents i que s'han controlat amb prou seguretat les possibles variables de confusió. En aquestes condicions, podem afirmar, amb un bon nivell de certesa, que els canvis registrats en la variable o variables dependents són produïts per la manipulació de la variable o variables independents, acceptant però l'existència d'un component d'*error experimental* que no és possible eliminar totalment. Tot això configura el que anomenem **validesa interna** de l'estudi.

És important definir en aquest punt el concepte d'**error experimental** i les seves implicacions. Per fer-ho, imaginem inicialment una situació "ideal", en la qual tenim un control absolut de les variables ambientals i, a més, totes les persones participants a l'estudi són absolutament iguals pel que fa a les seves característiques individuals. Suposem també que els instruments de registre que utilitzem per obtenir els resultats de les variables dependents són perfectament fiables i, per tant, que no es produeix cap tipus d'error de mesura. En aquestes circumstàncies, l'únic element que actuaria sobre els resultats seria la variable independent. Per tant, coneixent quin tractament de la variable independent ha rebut una persona, podríem fer una predicció perfecta sobre el seu resultat experimental. Naturalment, aquesta situació és fictícia i és difícil imaginar que es pugui produir a la pràctica. Les diferències individuals existeixen, els errors de mesura o registre també, i la influència de variables ambientals o de situació no controlades també. Per tant, existeix sempre un cert nivell de *variabilitat aleatòria* a les dades. Les diferents persones que reben el mateix tractament experimental no obtindran en general els mateixos resultats, i fins i tot la mateixa persona mesurada en moments diferents tampoc ens oferirà resultats idèntics. Definim aquesta variabilitat com a aleatòria no perquè no tingui les seves causes, sinó perquè aquestes causes en general no són conegudes per l'investigador/a i, per

tant, no podem preveure de forma exacta el seu impacte. En tot cas, el resultat d'això és que la predicció que puguem fer sobre el resultat d'una persona a partir del tractament que hagi rebut de la variable independent no serà perfecta, com ho era en el cas "ideal", i es produirà una desviació respecte d'aquesta predicció perfecta. Aquesta variabilitat aleatòria de les dades, o aquesta desviació dels resultats reals respecte dels podríem preveure en una situació ideal, és el que anomenem error experimental.

És important indicar que l'existència d'error experimental no implica que l'estudi estigui mal plantejat o mal executat, i no significa que no pugui existir un bon nivell de validesa interna. Si treballem amb diferents grups experimentals, i si aconseguim que aquests grups siguin bastant o molt equivalents pel que fa a les característiques individuals de les persones participants, llavors l'impacte de les diferències individuals sobre els resultats es produirà a nivell de variabilitat intra-grups, com s'indicava abans, i no de diferències globals entre els grups. Si això és així, podem tenir un bon nivell de confiança en que les diferències entre les mitjanes de resultats dels diferents grups són degudes a l'efecte de la variable independent. Per tant, podem pensar que tenim un bon nivell de validesa interna, malgrat la presència d'error experimental.

Naturalment, en situacions molt controlades (com la que es plantejava amb la caixa de Skinner) i treballant amb subjectes experimentals individuals (evitant l'ús de grups) podem aconseguir una reducció molt important (però no total) de la magnitud de l'error experimental, tot i que aquest plantejament també té limitacions importants, com ja s'ha indicat.

A més de la validesa interna, hi ha un altre component o un altre tipus de validesa que també és molt important. Si en parlar de validesa interna ens preguntem si l'experiment està ben fet, si ens garanteix una relació clara i no contaminada entre VI i VD, al parlar de **validesa externa** ens fem una pregunta diferent: Fins a quin punt els resultats obtinguts són generalitzables?. Cal tenir present que els resultats d'un experiment s'obtenen treballant amb uns determinats subjectes participants, amb uns certs valors de les variables independents i en una situació concreta, de vegades força artificial (per exemple, una sala aïllada en un centre de recerca). Cal preguntar-se, doncs, si els resultats obtinguts es mantindrien si els participants fossin uns altres, o si l'experiment es portés a terme en una situació més "natural" (a l'escola o a l'empresa, per exemple), o si els valors de la VI no fossin exactament els que s'han escollit (per exemple, si hi ha una relació lineal entre VI i VD, aquesta relació es mantindria si agaféssim un altres valors de la VI?).

Hi ha moltes consideracions a fer en relació amb el tema de la validesa externa. Simplificant molt les coses, podem dir el següent:

- . La qüestió de la generalització a altres subjectes és un problema de representativitat de la mostra, per tant, de la seva grandària i de la forma en què ha estat escollida (si treballem amb mostres, i no amb subjectes individuals, com als dissenys de replicació intra-subjecte)
- . El problema de l'artificialitat de les situacions només es podria resoldre realitzant l'estudi en contextos més naturals. Cal tenir molt present, però, que en els contextos "naturals" és més difícil controlar algunes variables de possible confusió, i el que guanyem en validesa externa ho podem perdre en validesa interna
- . Pel que fa als valors de les variables independents, moltes vegades l'investigador ja escull directament els que l'interessen, i no es planteja més problemes en relació amb això. Si hi ha

una preocupació per la generalització a altres valors de la VI, caldria plantejar un model d'efectes aleatoris, seleccionant de forma aleatòria els valors utilitzats de la variable independent a cadascun dels experiments i rèpliques que es realitzin

Esquemàticament:

Validesa interna	Validesa externa
<ul style="list-style-type: none"> ○ Establiment de relació inequívoca entre VI i VD ○ Paper central del control experimental ○ Relació entre nivell de control i oficialitat ○ Caràcter prioritari 	<ul style="list-style-type: none"> ○ Representativitat dels resultats ○ Interès de la generalització a: <ul style="list-style-type: none"> ● Altres subjectes ● Altres valors de la V.I. ● Altres situacions (validesa ecològica)

Cal fer una darrera reflexió. Sovint pot produir-se una certa contraposició entre la validesa interna i l'externa, en el sentit de que en buscar situacions més naturals es pot perdre nivell de control i, per tant, el que es guanya en validesa externa es perd en validesa interna. Si es produeix algun dilema d'aquest tipus, cal remarcar la importància de preservar un nivell acceptable de validesa interna o, dit d'una altra forma, en termes generals la validesa interna hauria de tenir una certa prioritat, ja que no tindria sentit plantejar la generalització d'uns resultats que no puguem considerar prou segurs o vàlids. Això cal valorar-ho acuradament en cada situació concreta.

La validesa interna i la validesa externa són els dos tipus de validesa més directament vinculats al disseny de recerca que s'utilitzi. Com s'ha indicat al principi, al referir-nos al tres grans elements de la recerca en Psicologia (mesura, disseny i anàlisi) hi ha altres tipus de validesa, com són la validesa de constructe, relacionada amb el tema de la mesura de les variables, o la validesa de conclusió estadística, vinculada a l'anàlisi de dades a partir de tècniques estadístiques. Tot i que es farà alguna referència a aquests dos tipus de validesa, no seran tractades de forma extensa en aquest text.

4.- Principals tipus de dissenys experimentals i quasi-experimentals: Classificació inicial

Recollint algunes idees ja comentades, una primera classificació general dels dissenys experimentals ens obliga a distingir-ne dos grans tipus o categories: els dissenys experimentals clàssics o estadístics, basats en el treball amb grups de subjectes, i els dissenys experimentals de replicació intra-subjecte o $N=1$, que treballen sempre amb subjectes o unitats individuals. Dins de cadascuna d'aquestes grans categories es poden establir també distincions importants. Els dissenys experimentals clàssics seran tractats a l'apartat II d'aquest text, mentre que els de replicació intra-subjecte es tractaran a l'apartat IV.

Les diferències entre el treball amb grups i el treball amb subjectes individuals són evidents, i es poden resumir fàcilment a partir d'un exemple, que es presenta de forma esquemàtica:



Individual-grups

Grups:

- Diferències individuals
- Mostra representativa?
- Generalització estadística
- Comparació amb grup control

Es vol comparar la possible eficàcia de dos tipus d'intervenció pensades per a pacients amb dolor crònic. Les intervencions són: CBT (teràpia cognitivconductual) i exposició (EXP). La CBT s'adreça a aspectes com relaxament, focalització de l'atenció, etc., mentre que EXP està més enfocada a aspectes emocionals, i es basa en l'exposició controlada a situacions potencialment doloroses. Es registren diferents indicadors fisiològics del dolor i es demanava també una valoració subjectiva de dolor mitjançant un qüestionari.

- ▶ Treball amb grups (exemple 4 Dossier de casos): 12 persones, organitzades en dos grups, un per a CBT i un altre per a EXP. Possibilitat de grup control.
- ▶ Treball amb casos individuals: Un pacient, diferents fases: Fase inicial sense tractament– Aplicació EXP– Retirada EXP– Aplicació CBT– Retirada CBT i seguiment

Cas únic:

- No diferències individuals
- No representativitat
- Generalització?
- Comparació amb fase no tractada

Com es pot veure, a l'exemple proposat es plantegen dues alternatives: Constituir dos grups diferents, cadascun dels quals rebrà un dels tractaments (aquest és el cas més senzill, encara que no l'únic possible), o bé treballar amb un subjecte únic, que passarà pels dos tractaments en fases diferents. L'element bàsic d'un disseny experimental clàssic són els *grups*, mentre que en un disseny de replicació intra-subjecte són les *fases*. Això fa que el tipus de plantejaments i de limitacions de cada tipus de disseny siguin diferents:

. En el plantejament de grups, partim de la base que treballem amb una mostra de subjectes o unitats experimentals, que hem de suposar que és representativa d'una certa població global, a la qual volem generalitzar els resultats a través dels procediments propis de la inferència estadística. El fet de treballar amb diferents subjectes introdueix la qüestió de les diferències individuals, la seva influència en els resultats i la necessitat de procedir al seu control. Per millorar el coneixement de l'efecte dels diferents tractaments, un procediment possible en aquest plantejament és la introducció d'un *grup control*, és a dir, d'un grup del qual es fan registres de la variable dependent sense introduir cap tractament experimental, de manera que els resultats d'aquest grup serveixen com a element de comparació amb els resultats dels grups que sí han rebut algun tipus d'intervenció experimental.

. Si treballem amb subjectes individuals, l'administració dels diferents tractaments experimentals s'ha de fer en fases diferents. Cada fase inclou un cert nombre de registres de la variable dependent, espaiats en el temps (registres per hora, per dia, per setmana, etc.). Naturalment, la comparació dels resultats obtinguts a les diferents fases és la base per poder veure si l'efecte dels diversos tractaments és diferent o no. En aquest tipus de situacions és extremadament útil (gairebé indispensable) introduir també fases de "no tractament", és a dir, fases a les quals es fan registres de la variable dependent sense haver introduït cap tractament experimental. Aquestes fases de no tractament, anomenades fases de *línia base*, juguen un paper comparatiu similar al del grup control en els dissenys de grups, ja que permeten obtenir informació més ajustada sobre l'efecte dels tractaments comparant els resultats de les fases de tractament amb les de línia base.

El treball amb grups de subjectes dona lloc als dissenys experimentals clàssics o estadístics, tal i com s'ha indicat abans, però també es produeix en dissenys que no són pròpiament

experimentals, com és el cas dels dissenys quasi-experimentals. La diferència bàsica entre els dissenys experimentals clàssics i els dissenys quasi-experimentals, com s'explica una mica més endavant, és que en el primer cas els grups són formats a partir d'un criteri establert per l'investigador/a, amb la finalitat d'obtenir grups tan equivalent com es pugui, mentre que en el segon cas la formació dels grups no està sota control de les persones que realitzen la recerca.

Esquemàticament:



Formació de grups

Sota control de l'experimentador/a: **Dissenys experimentals clàssics**

. Aleatorització: Grups formats a l'atzar, podem suposar que les característiques individuals s'han distribuït de forma més o menys equitativa. **Dissenys de grups aleatoritzats**

. Homogeneïtzació dels grups a partir de característiques individuals identificades. **Dissenys de grups homogenis**

(també possibilitat de control a posteriori via tècniques estadístiques: Anàlisi de covariància)

. El mateix grup experimental rep totes les condicions experimentals en moments diferents. **Dissenys de mesures repetides**

Sense control de l'experimentador/a (grups "naturals" o no aleatoritzats): **Dissenys quasiexperimentals**

Com s'indica a l'esquema, si per qualsevol motiu l'investigador/a no pot intervenir en la formació dels grups, i per tant es veu en l'obligació de treballar amb grups formats prèviament o sense el seu control, i sense possibilitat d'intervenir en la seva composició, llavors estem en el terreny dels dissenys quasi-experimentals, que seran tractats a l'apartat III d'aquest text. Per exemple, si vol comprovar l'eficàcia de dos sistemes didàctics diferents a una escola, probablement s'haurà d'aplicar un dels sistemes a un grup de classe ja format i l'altre sistema a un altre grup, i comparar posteriorment els resultats. Aquests grups no els ha format l'experimentador, sinó que ja estaven configurats prèviament i no són modificables, de manera que no en podem garantir l'equivalència. Un disseny quasi-experimental comparteix moltes característiques amb els dissenys experimentals clàssics, però amb la diferència fonamental de treballar amb grups no aleatoritzats i, per tant, sense garantia d'equivalència. Cal dir també que, en utilitzar-se moltes vegades en contextos "reals" (escola, empresa, centre de salut, etc.) molt sovint no és possible aplicar totes les tècniques de control habituals en situacions més artificials o de laboratori.

Com es veurà més endavant, en els dissenys quasi-experimentals s'introdueixen alguns elements que intenten compensar d'alguna forma el problema fonamental de la possible no equivalència dels grups experimentals. Això dona lloc a una certa diversitat en els tipus de dissenys disponibles, com es mostrarà a l'apartat corresponent.

Per la seva banda, els dissenys de replicació intra-subjecte, basats en l'estructura de fases, són més flexibles i es poden organitzar de formes diferents en funció del número de fases que es vulguin introduir i del contingut de cadascuna d'elles. Bàsicament caldrà distingir entre fases

sense intervenció o de línia base i fases amb intervenció i, dins d'aquestes últimes, caldrà definir quin tipus d'intervenció es fa en cada fase. Tot i així, com es veurà a l'apartat IV, també es poden introduir alguns criteris classificatoris dins d'aquests dissenys, sobretot a partir de la possibilitat o no de reversió de conducta, i de l'existència d'una o diverses sèries de dades.

Resumint el que s'ha indicat fins ara, els casos que podem trobar són principalment els següents:

Denominació	Aspectes bàsics	Classificació general
Dissenys experimentals clàssics o estadístics	<ul style="list-style-type: none"> . Treball amb grups formats segons criteri de l'investigador. Es busca l'equivalència dels grups . Control experimental directe + control probabilístic 	<ul style="list-style-type: none"> . Grups aleatoritzats . Grups homogenis . Mesures repetides Altres criteris de classificació: <ul style="list-style-type: none"> . Simple-Factorials . Univariables-Multivariables . Transversals - Longitudinals
Dissenys quasi-experimentals	<ul style="list-style-type: none"> . Grups naturals o no aleatoritzats. Risc de no equivalència dels grups . Manca de control experimental complet 	<ul style="list-style-type: none"> . Transversals . Longitudinals (sèries temporals) Identificació de característiques concretes (grup control no equivalent, variables dependents no equivalents, etc.)
Dissenys experimentals N=1 o de replicació intra-subjecte	<ul style="list-style-type: none"> . Treball amb subjectes individuals . Control experimental directe 	2 criteris de classificació: <ul style="list-style-type: none"> . Dissenys amb i sense reversió de conducta . Dissenys intra-sèries i entre-sèries

Als apartats II, III i IV s'estudiaran de forma més àmplia els diferents tipus de dissenys i s'aprofundirà en els criteris d'identificació i classificació.

5.- Consideracions estadístiques

. Lògica general

En el marc dels dissenys experimentals clàssics (i també en bona part dels dissenys quasi-experimentals), les eines de caràcter estadístic són un instrument molt important per valorar els resultats. Un motiu important d'això, entre d'altres, és que el plantejament de treballar amb grups de subjectes implica la presència de diferents elements d'aleatorietat: quan seleccionem una mostra a l'atzar dins de la població que ens interessa estudiar; quan decidim amb un criteri aleatori quines persones formaran part de cadascun dels grups experimentals; quan utilitzem

instruments de registre que necessàriament incorporen errors aleatoris de mesura; quan no podem controlar de forma completa les variables externes als subjectes, ambientals o situacionals, etc. En altres paraules, en qualsevol conjunt de resultats experimentals tindrem necessàriament una variabilitat aleatòria que, com s'ha indicat en parlar de validesa, anomenem *error experimental*.

Quan introduïm a la situació una determinada variable independent, amb els seus valors o tractaments, el nostre objectiu és saber si aquesta variable té o no un impacte sobre els resultats de la variable o variables dependents. El primer que cal tenir en compte és que, a diferència del que passa amb les diferents fonts d'error experimental, l'impacte de la variable independent sobre la dependent té un caràcter **sistemàtic**. Això significa que si, per exemple, una teràpia psicològica és millor que una altra, les persones que rebin la primera teràpia tendiran a mostrar millors resultats que no els que rebin la segona. Això no significa que la reacció de cada persona davant de la teràpia rebuda sigui exactament igual, però sí que a nivell del conjunt del grup, la mitjana de resultats del primer grup serà millor que el del segon.

En conseqüència, podem trobar dues situacions: Si la variable independent no tingúes impacte sobre els resultats, els valors obtinguts per a la variable dependent a cada subjecte serien el resultat de tot un conjunt d'influències aleatòries (error experimental); en canvi, si la variable independent té un efecte sobre els resultats, el resultat de cada persona serà conseqüència de l'efecte del tractament de la variable independent que aquesta persona hagi rebut, tot i que continuaran estan presents tots els efectes aleatoris propis de l'error experimental.

Esquemàticament:

Situació	Efectes presents	Model
No hi ha efecte de la V.I.	Variabilitat aleatòria	$Y_{ij} = \varepsilon_{ij}$
Hi ha efecte de la V.I.	Variabilitat sistemàtica + Variabilitat aleatòria	$Y_{ij} = \alpha_j + \varepsilon_{ij}$

Y_{ij} representa el resultat obtingut pel subjecte i quan rep el tractament j ; α_j representa l'efecte del tractament j sobre aquest resultat; ε_{ij} representa l'impacte sobre el resultat del subjecte i de tots els inputs aleatoris que formen part de l'error experimental (característiques individuals d'aquesta persona, possibles canvis en estats interns del subjecte, possibles canvis en factors ambientals, errors de mesura, etc.).

Pràcticament tots els estudis experimentals basats en el plantejament clàssic contenen una o més proves estadístiques que ens permeten avaluar els resultats i fonamentar la seva interpretació. És molt habitual trobar-hi proves molt conegudes com la t de Student, per a la comparació de dues mitjanes, o l'anàlisi de la variància (AVAR o ANOVA) per a la comparació de múltiples mitjanes, entre d'altres. Aquestes proves ens permeten valorar el possible impacte de les variables independents sobre les dependents i , per tant, ens permeten distingir entre diferents models que poden ajustar-se a les dades. Són d'aplicació en aquest context els conceptes bàsics de la inferència i la decisió estadística, que ja són coneguts: Hipòtesi nul·la i hipòtesi alternativa, risc d'error i tipus d'errors (tipus I tipus II), distribució de probabilitat de l'estadístic, nivell de significació i criteris de decisió.

La majoria de les proves estadístiques utilitzades habitualment en el context experimental estan basades en el que anomenem *model lineal clàssic*. El model lineal clàssic inclou diferents

tècniques (models de regressió, anàlisi de la variància, anàlisi de covariància) i generalitzacions dels anteriors, com seria l'anàlisi multivariable de la variància o altres. El model lineal clàssic ens permet treballar amb variables dependents de caràcter quantitatiu, per a les quals s'assumeixen tres condicions bàsiques: la *normalitat de les dades*, l'*homocedasticitat* o homogeneïtat de variàncies per als resultats dels diferents nivells o factors, i la *independència dels registres* (els resultats obtinguts per part dels diferents subjectes a les diverses condicions experimentals són independents entre si). Naturalment, tractant-se d'un model lineal s'assumeix també que la relació entre variables independents i dependents pot ser expressada correctament amb un model d'aquest tipus.

Algunes tècniques associades al model lineal clàssic són:

. Els models de regressió lineal, on tant les variables explicatives (en el nostre context, bàsicament les variables independents) com les variables conseqüència (variables dependents) són de caràcter quantitatiu. Cal dir, però, que els models de regressió lineal es poden adaptar, mitjançant sistemes de codificació, al seu ús amb variables independents qualitatives

. Els models d'anàlisi de la variància (AVAR o ANOVA), on les variables explicatives són de caràcter qualitatiu, i s'anomenen factors (tot i que també es podria aplicar si la variable independent té un número limitat de valors quantitatius)

. Els models d'anàlisi de covariància (ACOVAR o ANCOVA) els quals, com es veurà més endavant, permeten introduir en el model ANOVA variables concomitants de caràcter quantitatiu

En realitat els diferents models exposats no són excloents i comparteixen molts elements en comú. Com s'ha indicat és possible, per exemple, incloure variables de caràcter qualitatiu en un model de regressió. També és possible, amb precaucions concretes, utilitzar variables explicatives quantitatives al fer una anàlisi de la variància. En realitat, els models ANOVA són una simplificació dels models de regressió i, de fet, poden ajustar-se utilitzant les tècniques pròpies dels models de regressió.

Cal indicar també que el model lineal clàssic no és sinó un exemple particular de diferents tipus de models que es poden plantejar per al modelatge estadístic: El model lineal generalitzat (que permet incorporar procediments específics per a variables resposta de tipus categòric, com és la regressió logística), el model lineal mixt (que permet analitzar dades que no compleixin algunes de les condicions indicades en el model lineal clàssic i modelitzar situacions amb diferents components aleatoris) o el model d'equacions estructurals (que permeten treballar tant amb variables observables com amb variables latents, entre altres aspectes).

De forma molt resumida:

Models estadístics

- ▶ Identificar quin model s'ajusta millor a les dades
- ▶ Model lineal clàssic
 - ▶ Models de regressió lineal
 - ▶ Models ANOVA o AVAR (anàlisi de la variància)
 - ▶ Models ANCOVA (anàlisi de covariància)
- ▶ Altres models: Model lineal generalitzat, models mixtos, models d'equacions estructurals

Fetes les consideracions anteriors, s'il·lustrarà l'ús dels models ANOVA en un context experimental utilitzant alguns exemples numèrics. L'esquema que s'utilitzarà és el següent:

Models ANOVA

- ▶ Lògica de la prova
- ▶ Model subjacent
- ▶ Mecànica i càlculs
- ▶ Proves complementàries: Comparacions parcials, anàlisi de tendències
- ▶ Extensions: Anàlisi multivariable de la variància

Suposem que utilitzem un disseny de grups aleatoritzats amb una variable independent amb 4 valors o tractaments, i una única variable dependent. Anomenarem la variable independent com a "Tractaments" i la variable dependent senzillament com a "Resultats". Tenim en total 20 subjectes, que són distribuïts aleatòriament en quatre grups, cadascun d'ells amb 5 subjectes. Imaginem que els resultats obtinguts, una vegada fets els registres de la variable dependent, són els següents:

	Resultats experimentals			
Tractaments	Tractament 1	Tractament 2	Tractament 3	Tractament 4
Resultats individuals	8	13	18	17
	10	15	20	21
	9	10	22	21
	7	13	18	23
	6	14	17	18
Mitjanes	8	13	19	20
Desviacions	-7	-2	4	5
	Mitjana global = 15			

La mitjana global dels 20 resultats obtinguts és de 15. Com es pot veure, s'indiquen també les desviacions de la mitjana de cada grup respecte de la mitjana global de resultats. Això ens indica que els resultats dels grups 1 i 2 estan per sota de la mitjana global, mentre que els grups 3 i 4 estan per sobre. Igualment es poden calcular fàcilment les desviacions de cada puntuació concreta respecte de la mitjana global i respecte del seu grup. Per exemple: La puntuació del primer subjecte del grup 1 (8) té una desviació de -7 respecte de la mitjana global ($8-15=-7$), mentre que té una desviació 0 respecte de la mitjana del seu grup ($8-8=0$), i igualment es pot fer amb totes les puntuacions.

Des de la perspectiva dels models ANOVA, la pregunta a fer-se és la següent: Considerant que la mitjana global dels resultats obtinguts és de 15, per què les diferents puntuacions no són totes elles iguals, sinó que es distribueixen entorn d'aquesta mitjana global?. Per què es produeixen **desviacions** respecte d'aquesta mitjana global?. Com hem vist, la desviació del resultat del primer subjecte del primer grup respecte de la mitjana global és $(8-15) = -7$, mentre que la desviació del resultat de l'últim subjecte del grup 4 és $(18-15)=3$. Podem observar que també les mitjanes globals de cadascun dels grups es desvien respecte de la mitjana general.

En el disseny plantejat només disposem d'una informació respecte dels subjectes participants: sabem a quin grup experimental han estat assignats, és a dir, quin tractament experimental han rebut. No tenim informació, però, sobre les seves característiques individuals, però podem pensar amb tota seguretat que influeixen en els resultats. Considerant això, podem concretar el model següent:

$$Y_{ij} - \mu = \alpha_j + \varepsilon_{ij}$$

O, d'una altra forma:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

És a dir, la desviació entre la puntuació obtinguda per un subjecte i que pertany al grup j (Y_{ij}) i la mitjana global (μ) es pot explicar per dos elements: L'efecte del tractament experimental que ha rebut (α_j) i la influència de les característiques individuals del propi subjecte i altres elements propis de l'error experimental (ε_{ij}).

Si agafem com a exemple el darrer subjecte del grup 4, podem dir que:

$$18 = 15 + 5 - 2 ,$$

és a dir, que la puntuació del subjecte (18) és igual a la mitjana global de tots els resultats (15), més la desviació de la mitjana del seu grup respecte de la mitjana global (5) i la desviació del propi subjecte respecte de la mitjana del seu grup (-2). En altres paraules, aquesta persona pertany a un grup que té un resultat global superior en 5 punts a la mitjana general, i dins del seu grup està a la “banda baixa”, és a dir, el seu resultat està per sota (-2) de la mitjana del seu propi grup.

Podem constatar doncs diferents coses:

1.- Les puntuacions individuals són diferents i mostren un cert grau de desviació respecte de la mitjana global de resultats. Podem parlar doncs d’una variació o variabilitat **total** dels resultats

2.- Les mitjanes dels diferents grups són diferents entre sí i també mostren desviacions respecte de la mitjana global. Existeix doncs una variació o variabilitat **entre grups**, i l’explicació lògica d’aquesta variabilitat és que els diferents tractaments experimentals tenen efectes diferents sobre els resultats

3.- Finalment, les puntuacions individuals també són diferents dins de cada grup, i existeix una desviació entre cada resultat individual i la mitjana del seu propi grup. Hi ha doncs una variació o variabilitat **intra-grups o dins dels grups**. Això es pot atribuir al fet que les persones són diferents entre sí i, tot i estar sotmeses al mateix tractament experimental, els resultats no tenen per què ser iguals, i també a altres elements propis de l’error experimental (canvis no controlats en la situació, errors de mesura, etc.), és a dir, inclou qualsevol circumstància aleatòria que pugui haver afectat al resultat d’un subjecte o unitat en concret

En definitiva, podem dir que la variació global de les puntuacions és el resultat d’un **component sistemàtic** (l’efecte dels tractaments, que afecta al conjunt de persones que pertany a cada grup) i un **component aleatori** (l’error experimental, i particularment l’impacte de les diferències individuals, que considerem aleatòries pel fet de que els subjectes són assignats als diferents grups a l’atzar, i per tant no sabem qui anirà a parar a cada grup ni les seves característiques individuals).

. L’anàlisi de la variància (AVAR o ANOVA)

S’explicaran aquí els fonaments i la lògica de la tècnica més utilitzada per a l’anàlisi de dades experimentals, que és l’anàlisi de la variància (AVAR o ANOVA). Com a exemple de plantejament basat en el model lineal clàssic, abans de procedir a realitzar l’ANOVA caldria fer una comprovació de les condicions de normalitat de les dades i d’homoscedasticitat. En el primer cas es pot utilitzar qualsevol des les proves habituals (particularment la prova de normalitat de Shapiro-Wilk, especialment útil quan treballem amb mostres petites) per comprovar si la condició de normalitat es compleix per al conjunt de les dades i també dins de cadascun dels grups. Pel que fa a la condició d’homoscedasticitat o homogeneïtat de variàncies, tenim diferents proves a la nostra disposició, com per exemple les de Levene, Bartlett o Brown-Forsythe, aquesta última especialment indicada si les dades no s’ajusten bé a la condició de normalitat.

Pel que fa a la condició d’independència dels registres, cal tenir en compte que en determinats tipus de dissenys experimentals es poden plantejar dubtes sobre el seu compliment. Concretament, a un disseny de mesures repetides, considerant que les mateixes persones passen per les diferents condicions experimentals, podem pensar que els diferents resultats

obtinguts per una mateixa persona no són independents, sinó que mostren determinades regularitats relacionades amb les característiques individuals d'aquest subjecte. Com es veurà més endavant, el dubte sobre el compliment de la condició d'independència dels registres en aquest tipus de disseny fa que el plantejament de l'anàlisi de les dades sigui una mica diferent al d'altres casos.

Com totes les proves estadístiques de significació, el que fa l'ANOVA és indicar-nos si les diferències observades entre les diferents puntuacions són explicables purament per la presència d'una variabilitat aleatòria (hipòtesi nul·la) o bé si existeix un efecte diferencial i sistemàtic dels tractaments que explica una part de les diferències (hipòtesi alternativa). Expressat d'una altra forma:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1 : Alguna o algunes de les igualtats anteriors són falses

μ_1, μ_2, μ_3 i μ_4 representen les mitjanes de resultats obtingudes sota cadascuna de les 4 condicions experimentals. Cal remarcar que representem aquestes mitjanes en notació poblacional (lletres gregues, indicant paràmetres poblacionals), ja que el nostre objectiu és saber si els quatre tractaments experimentals produïrien o no resultats diferents *en el conjunt de la població*, no únicament a la mostra utilitzada. Podem arribar a plantejar aquesta generalització dels resultats a partir dels principis i operacions de l'estadística inferencial.

Tot i que les modernes aplicacions estadístiques (SPSS, R, SAS, etc.) realitzen fàcilment els càlculs corresponents a l'ANOVA, per entendre millor la seva lògica es mostrarà el procés de càlcul "manual" que ens permetrà completar l'anàlisi. En primer lloc, el quadre de dades anterior s'ha de completar de la forma següent, calculant els totals de cada grup i el total general.

	Càlculs inicials ANOVA			
Tractaments	Tractament 1	Tractament 2	Tractament 3	Tractament 4
Resultats individuals	8	13	18	17
	10	15	20	21
	9	10	22	21
	7	13	18	23
	6	14	17	18
Mitjanes	8	13	19	20
Totals	40	65	95	100
	Mitjana global = 15 Total global = 300			

D'acord amb la lògica indicada anteriorment, l'ANOVA ens permet separar, dins de la variabilitat general de les dades, una variabilitat entre-grups (que es podria explicar per l'acció dels tractaments experimentals) i una variabilitat intra-grups, atribuïble sobretot a les característiques individuals dels participants i a altres variables aleatòries no conegudes.

La forma més habitual de mesurar la variabilitat d'un conjunt de dades és a partir del càlcul de la seva variància, que es pot fer a partir d'una fórmula ben coneguda:

$$s^2 = \frac{\sum(Y_i - \mu)^2}{n - 1}$$

El numerador d'aquesta fórmula, anomenat variació o suma de quadrats, s'obté fàcilment a partir de les dades disponibles, amb dues possibilitats alternatives:

Càlcul a partir de puntuacions i mitjanes

$$\text{S.Q. Total} = (8-15)^2 + 10-15)^2 + (9-15)^2 + \dots + (18-15)^2 = 534$$

$$\text{S.Q. Entre} = 5 [(8-15)^2 + (13-15)^2 + (19-15)^2 + (20-15)^2] = 470$$

$$\text{S. Q. Intra} = (8-8)^2 + (10-8)^2 + \dots + (23-20)^2 + (18-20)^2 = 64$$

Càlcul a partir de puntuacions i totals

$$\text{S.Q. Total} = 8^2 + 10^2 + 9^2 + \dots + 23^2 + 18^2 - (300^2/20) = 534$$

$$\text{S.Q. Entre} = 40^2/5 + 65^2/5 + 95^2/5 + 100^2/5 - 300^2/20 = 470$$

$$\text{S.Q. Intra} = 534 - 470 = 64$$

És fàcil veure que la variació o suma de quadrats total ens indica fins a quin punt les puntuacions estan disperses entorn de la mitjana global de resultats. La suma de quadrats entre-grups ens mostra fins a quin punt les *mitjanes* dels diferents grups estan disperses entorn de la mitjana global (i, per tant, fins a quin punt són diferents entre si). Finalment, la suma de quadrats intra-grup ens indica fins a quin punt les diferents puntuacions dins d'un grup estan disperses entorn de la mitjana *del propi grup*. Com es pot comprovar fàcilment, hem procedir a una descomposició de la variació total, de forma que S.Q. Total = S.Q. entre-grups + S.Q. intra-grups.

Pel que fa al denominador de la fórmula de la variància, anomenat graus de llibertat, és igual al número de puntuacions implicades en el càlcul de la suma de quadrats, menys una (n-1). En el nostre cas:

Total: $20 - 1 = 19$ (número total de puntuacions menys una). Fórmula general: $N - 1$

Entre grups: $4 - 1 = 3$ (número de grups experimentals menys un). Fórmula general: $k - 1$

Intra-grups: $20 - 4 = 16$. Fórmula general: $N - k$

L'explicació del cas dels graus de llibertat de la variació intra-grups és la següent: A cada grup hi ha 5 puntuacions, que són comparades amb la mitjana del seu grup. Per tant, a cada grup hi ha $(n-1) = 5-1 = 4$ graus de llibertat. Això és cert per a cadascun dels 4 grups, per tant, $4 \times 4 = 16$. Tot i que la lògica és aquesta, la manera més fàcil de calcular-ho és a partir de l'expressió $N - k$, és a dir, número total de subjectes menys número de tractaments. $N - k = 20 - 4 = 16$.

Tota aquesta informació es trasllada habitualment al quadre resum de l'ANOVA:

Quadre-resum de l'anàlisi de la variància					
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	P
Entre-grups	470	3	156,67	39,17	0,000
Intra-grups	64	16	4,00		
TOTAL	534	19			

Càlcul dels graus de llibertat: Entre-grups: $k - 1 = 4 - 1 = 3$
 Intra-grups: $N - k = 20 - 4 = 16$
 Total: $N - 1 = 20 - 1 = 19$

Càlcul variàncies: Var. Entre = S.Q. entre / g.ll. entre = $470 / 3 = 156,67$
 Var. Intra = S.Q. intra / g.ll. intra = $64 / 16 = 4,00$

Càlcul F de Snedecor: $F = \text{Var entre} / \text{Var intra} = 156,67 / 4,00 = 39,17$

Una vegada tenim calculades les variàncies entre-grups i intra-grups, l'únic pas que ens queda és fer la comparació entre totes dues. La lògica d'això és la següent: Si la variable independent no té cap efecte sobre la dependent, llavors la variància entre-grups serà únicament una forma més de variabilitat aleatòria i, per tant, no té per què ser gaire diferent de la variància intra-grups, que també és de caràcter aleatori. En el cas contrari, si la variable independent té un efecte sistemàtic sobre la variable dependent, les mitjanes de resultats obtingudes als diferents grups seran apreciablement diferents (més del que es podria explicar per atzar) i, en conseqüència, la variància entre-grups serà major que la variància intra-grups. Per tant, les hipòtesis nul·la i alternativa de l'ANOVA es poden expressar també com a:

$$H_0: \sigma^2_{\text{ENTRE}} \approx \sigma^2_{\text{INTRA}} \quad H_1: \sigma^2_{\text{ENTRE}} \neq \sigma^2_{\text{INTRA}}$$

La forma de realitzar la comparació entre dues variàncies és fer el seu quocient, ja que la teoria estadística demostra que el quocient de dues variàncies segueix una distribució coneguda, concretament la F de Fisher-Snedecor (coneguda de forma més simple com a F de Snedecor), cosa que ens permet, com en altres proves estadístiques, calcular la probabilitat p de que el resultat obtingut per a l'estadístic (o un resultat més extrem) s'hagi produït si la hipòtesi nul·la fos certa. En el nostre cas, el quocient obtingut és $156,67 / 4,00 = 39,17$. Observem que el valor de p corresponent a $F > 39,17$ és molt petit (naturalment, si poséssim més decimals en trobaríem algun diferent de 0, el resultat concretament és 0,000000134) i, en tot cas, és molt inferior al nivell de significació que s'agafa normalment com a referència ($p=0,05$). Com que òbviament $0,000000134 < 0,05$, podem concloure que el resultat de la prova és significatiu i, en conseqüència, podem pensar que les diferències de resultats entre els diferents grups són producte d'un efecte sistemàtic dels tractaments experimentals.

Com ja s'ha indicat anteriorment, l'anàlisi de la variància és una concreció i aplicació concreta del model lineal clàssic. Aquest mateix cas es podria plantejar a partir d'un model de regressió lineal. Com és sabut, els models de regressió permeten modelitzar la relació entre variables (en el cas experimental, entre variables independents i dependents) a partir d'un o diversos models matemàtics. Sense entrar en detalls, si la variable independent i la dependent fossin quantitatives es podria plantejar un model de regressió convencional, del tipus:

$$y_i = \beta_0 + \beta_1 X + \varepsilon_i$$

on y_i seria el registre obtingut per a la variable dependent per al subjecte i , X el valor de la variable independent que ha estat assignat a aquest subjecte, β_0 i β_1 els coeficients o paràmetres del model, i ε_i el component d'error, no explicat per la resta de components del model.

Si, com és freqüent, la variable independent és categòrica, llavors cal utilitzar algun tipus de codificació numèrica per plantejar el model de regressió. Si la variable independent utilitzada a l'exemple fos qualitativa, llavors caldria derivar 4 variables codificades, de la forma següent:

$X_1 = 1$ si el subjecte rep el tractament 1, en cas contrari $X_1 = 0$

$X_2 = 1$ si el subjecte rep el tractament 2, en cas contrari $X_2 = 0$

$X_3 = 1$ si el subjecte rep el tractament 3, en cas contrari $X_3 = 0$

$X_4 = 1$ si el subjecte rep el tractament 4, en cas contrari $X_4 = 0$

O, més esquemàticament:

Tractament rebut	X_1	X_2	X_3	X_4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1

El model de regressió corresponent seria, per a qualsevol subjecte i :

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i$$

A la pràctica el model podria ser més senzill, ja que no calen els quatre components X per representar les dades (per exemple, si per a un subjecte qualsevol tenim que X_2 , X_3 i X_4 són iguals a zero, això vol dir automàticament que X_1 és igual a 1).

Una vegada identificat el model de regressió i estimats els valors dels seus paràmetres, la seva valoració (és a dir, la decisió sobre si ha existit o no impacte de la variable independent) es pot fer de diverses formes (per exemple, analitzant si els diferents paràmetres β són o no significativament diferents de zero). Una de les possibilitats de valoració del model seria a través d'una anàlisi de la variància, que ens oferiria resultats idèntics als que ja hem obtingut abans.

Un problema associat a les proves estadístiques de significació, com és el cas de l'anàlisi de la variància, és el de la seva potència, és a dir, la seva capacitat per detectar un efecte realment existent. La potència depèn de diferents aspectes, entre els quals destaquen la magnitud de la variància de l'error i també la grandària de la mostra utilitzada. Una variabilitat molt gran de les dades dificulta la detecció de possibles efectes dels tractaments, i per aquest motiu és important aconseguir mantenir l'error experimental dins de límits raonables, a través de diferents tècniques que s'estudiaran més endavant. Pel que fa al tema de la grandària de la mostra, es poden produir dos tipus de problemes: Si la mostra és molt petita, es pot passar per alt algun efecte real dels tractaments experimentals, ja la prova estadística pot no tenir la potència suficient, que hauria d'estar aproximadament entorn de 0,80. Al contrari, si la mostra és molt gran la prova ens pot mostrar com a significatius efectes relativament petits, que potser no són

rellevants des d'un punt de vista teòric o d'aplicació pràctica. Per aquest motiu, és molt recomanable (i cada vegada més habitual) complementar la prova estadística amb algun tipus d'índex suplementari que permeti valorar millor la magnitud del possible efecte dels tractaments. En aquesta qüestió s'han fet diferents propostes, en forma d'índexs directes (d de Cohen, g de Hedges, etc.), o de diferents tipus d'índex d'associació (el més conegut és omega-quadrat o ω^2).

Una mesura de magnitud de l'efecte molt senzilla i força utilitzada com a complement de l'ANOVA és *eta-quadrat* (η^2), que no és altra cosa que el quocient entre la suma de quadrats deguda als tractaments i la suma de quadrats total. A l'exemple anterior, el càlcul és:

$$\eta^2 = S.Q. \text{ entre-grups} / S.Q. \text{ total} = 470 / 534 = 0,88$$

A partir d'aquest càlcul, eta-quadrat es pot interpretar com una aproximació a la proporció de variabilitat total de les dades associada a la variable independent que s'estigui estudiant. En el nostre cas, podem considerar que una proporció de 0,88 (o un percentatge del 88%) de la variació total està associat a l'efecte dels tractaments. Ara bé, s'ha de dir que aquest índex és un estimador positivament esbiaixat d'aquesta proporció a nivell de la població (tot i que s'han proposat alguns ajustos), i en particular es pot veure afectat per la grandària de la mostra. Per altra banda, no hi ha acord entre els autors a l'hora d'interpretar eta-quadrat com un índex de magnitud de l'efecte; per exemple, alguns autors consideren que a partir d'una eta-quadrat de 0,14 podem considerar que l'efecte és gran, mentre que d'altres demanen valors superiors a 0,25. Tot i que en el cas plantejat el valor de 0,88 és molt alt i no ofereix dubtes, en altres casos sí que es poden produir problemes d'interpretació. En tot cas, segurament la utilitat més rellevant d'eta-quadrat és la de permetre fer comparacions entre els resultats de diferents estudis del mateix tipus (sempre que es compleixin certs criteris de comparabilitat), per veure en quins casos s'obtenen efectes més grans.

Com es veurà més endavant, quan tenim més d'un factor a l'anàlisi de la variància (per exemple, si treballem amb dues variables independents), és útil l'ús de l'índex *eta-quadrat parcial*. Aquest índex es calcula fent el quocient entre la suma de quadrats de l'efecte que ens interessa (numerador) i la suma d'aquesta mateixa suma de quadrats més la suma de quadrats de l'error experimental (denominador), deixant de banda la resta de factors que puguin estar presents a l'anàlisi. Quan només hi ha un factor a l'ANOVA, els valors d'eta-quadrat i eta-quadrat parcial són idèntics.

Hi ha diverses alternatives a l'índex eta-quadrat. Una d'elles és omega-quadrat (ω^2), un estimador menys esbiaixat que eta-quadrat, i que per a un disseny amb una sola variable independent es pot obtenir amb l'expressió:

$$\omega^2 = \frac{(k - 1)(F - 1)}{(k - 1)(F - 1) + N}$$

, on k és el número de grups, N el número total de subjectes i F és el valor de l'estadístic F de Snedecor obtingut al fer l'anàlisi de la variància.

En el nostre cas:

$$\omega^2 = \frac{(4 - 1)(39,17 - 1)}{(4 - 1)(39,17 - 1) + 20} = 0,851$$

Aquest resultat ens indica que una proporció de 0,851 (o un percentatge del 85,1%) de la variabilitat total de les dades està associada amb l'impacte de la variable independent. Com es pot veure, ω^2 ens ofereix un resultat una mica inferior a η^2 , ja que té un menor biaix positiu com a estimador poblacional.

Si l'anàlisi de resultats es fa a partir del corresponent model de regressió lineal, disposem dels índex R^2 (coeficient de determinació) i R^2 ajustat per valorar l'ajust del model i veure fins a quin punt la variància de les dades és explicada pel model que hem plantejat.

. Contrastos parcials de mitjanes i anàlisi de tendències

Cal remarcar que l'anàlisi de la variància ens informa únicament de l'efecte global dels tractaments experimentals, però no ens diu res sobre les diferències parcials que es puguin haver produït. A l'exemple numèric presentat abans hem esbrinat que l'efecte global dels tractaments és altament significatiu, però no podem respondre encara a preguntes més concretes. Per exemple, hi ha una diferència significativa entre l'efecte dels dos primers tractaments?. Hi ha diferència entre el tractament 3 i el 4?. Hi ha diferència entre els resultats dels dos primers tractaments (agafats conjuntament) i el dels dos últims?, etc. Per contestar aquestes preguntes disposem de la possibilitat de fer *contrastos parcials* de les dades.

Una forma senzilla de valorar qualsevol contrast és a partir de les mateixes tècniques de l'ANOVA. Si agafem com a exemple les tres comparacions proposades al paràgraf anterior, es poden definir i representar els tres contrastos corresponents i calcular per a cadascun d'ells la corresponent suma de quadrats, de la forma següent:

	Tractament 1	Tractament 2	Tractament 3	Tractament 4
Contrast 1	1	-1	0	0
Contrast 2	0	0	1	-1
Contrast 3	1	1	-1	-1

Càlcul sumes de quadrats dels contrastos:

Contrast 1: $S.Q_{c1} = 5 [(1 \cdot 8) + (-1 \cdot 13) + (0 \cdot 19) + (0 \cdot 20)]^2 / (1^2 + -1^2 + 0^2 + 0^2) = 62,5$

Contrast 2: $S.Q_{c2} = 5 [(0 \cdot 8) + (0 \cdot 13) + (1 \cdot 19) + (-1 \cdot 20)]^2 / (0^2 + -0^2 + 1^2 + -1^2) = 2,5$

Contrast 3: $S.Q_{c3} = 5 [(1 \cdot 8) + (1 \cdot 13) + (-1 \cdot 19) + (-1 \cdot 20)]^2 / (1^2 + 1^2 + -1^2 + -1^2) = 405$

Com es pot veure, el que s'ha fet és calcular les sumes de quadrats corresponents a cada contrast aplicant els coeficients que defineixen el contrast a les mitjanes de resultats dels diferents grups experimentals. D'aquesta forma, la suma de quadrats entre-grups queda descomposta en les sumes de quadrats parcials per a cadascun dels contrastos.

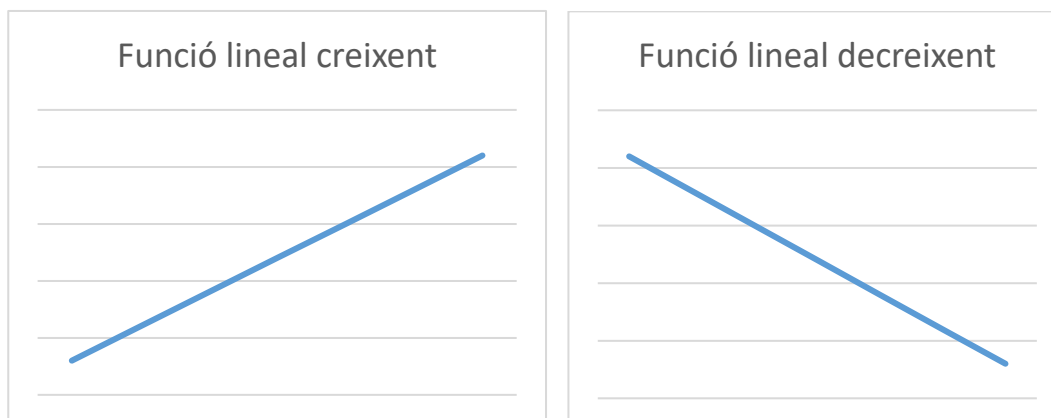
A partir d'aquí, l'anàlisi de la variància realitzat anteriorment es pot completar de la forma següent:

Quadre-resum de l'anàlisi de la variància					
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	P
Entre-grups	470	3	156,67	39,17	0,000
Contrast 1	62,5	1	62,5	15,63	0,001
Contrast 2	2,5	1	2,5	0,62	0,441
Contrast 3	405	1	405	101,25	0,000
Intra-grups	64	16	4,00		
TOTAL	534	19			

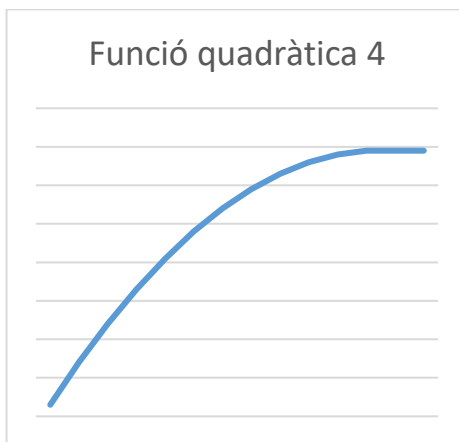
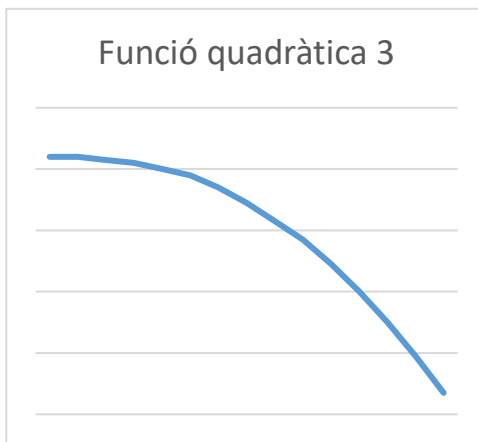
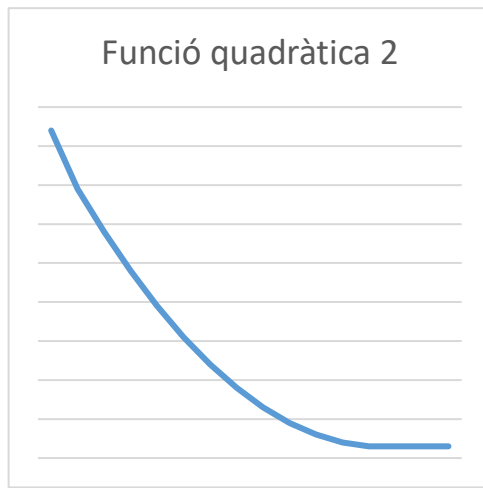
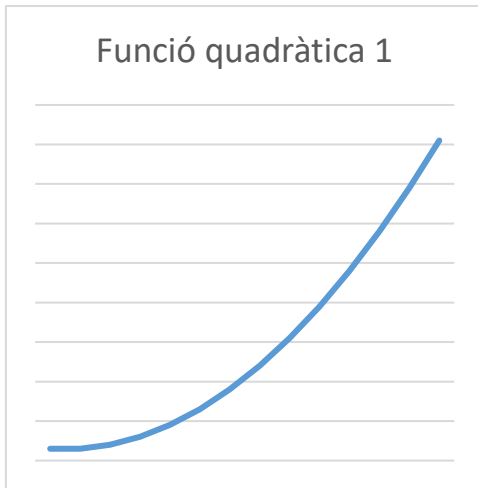
Es pot observar que, de la mateixa forma en què es podia valorar l'efecte global del conjunt dels tractaments experimentals, també és possible valorar la significació dels diferents contrastos parcials. En el nostre cas, es pot constatar que els contrastos 1 i 3 mostren un resultat significatiu, mentre que el contrast 2 no resultat significatiu estadísticament. Això vol dir que la diferència entre els resultats dels grups 3 i 4 no és significativa.

La tècnica de contrastos parcials pot resultar útil també per a finalitats especials. Per exemple, si tant la variable independent com la dependent són de caràcter quantitatiu, ens podria interessar esbrinar quin **tipus** de relació s'estableix entre les dues, és a dir, si la relació és lineal (els resultats de la VD s'incrementen o decreixen linealment a mesura que augmenta el valor de la variable independent), o bé si és una relació més complexa, representable amb una funció quadràtica, cúbica o més complicada. Quan el nostre objectiu és descobrir quin tipus de funció relaciona la variable independent i la dependent, realitzarem una **anàlisi de tendències**. Per tal de realitzar-lo és necessari que la variable independent tingui com a mínim tres valors, ja que amb dos valors no hi hauria forma de discriminar entre les diverses funcions possibles.

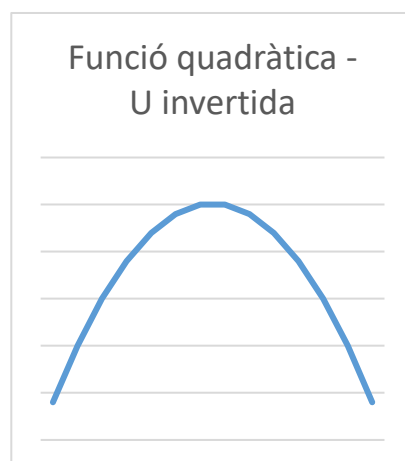
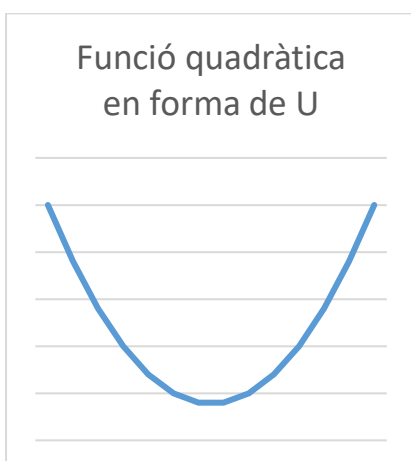
A la pràctica, en els resultats obtinguts a la recerca psicològica gairebé sempre podem trobar tendències lineals o bé quadràtiques, és molt estrany enfrontar-se a casos amb tendències més complicades. Gràficament, les tendències lineal i quadràtica ja són conegudes, tot i que poden oferir algunes variants. En el cas lineal, la tendència pot ser creixent o decreixent:



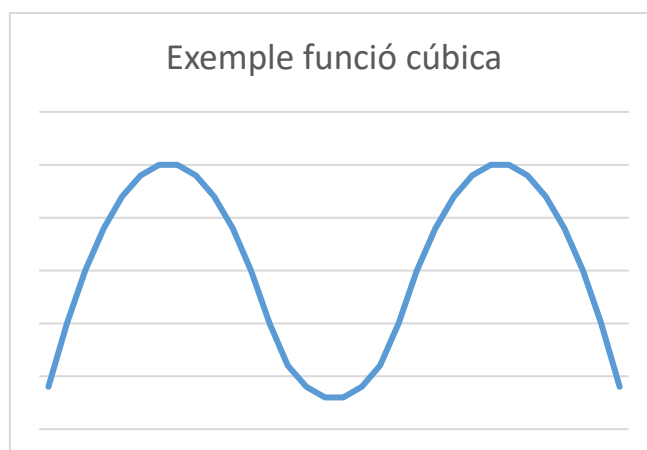
En funció del rang de valors utilitzats per a la variable independent, les funcions quadràtiques poden adoptar diferents formes:



Si tenim un rang suficient de valors, les tendències quadràtiques es poden mostrar en forma de U o de U invertida, com als gràfics següents:



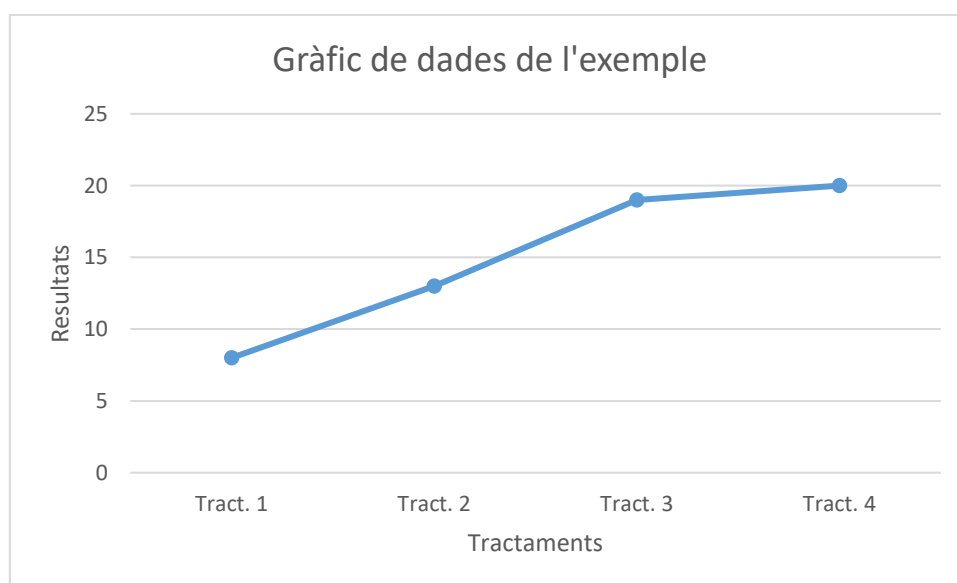
Les tendències cúbiques o superiors són més complexes, sovint amb forma sinusoidal:



La forma més senzilla de portar a terme l'anàlisi de tendències és utilitzant la tècnica de contrastos parcials. En aquest cas, però, no podem plantejar qualsevol contrast, sinó que cal fer ús d'uns contrastos especials, que ja ens venen donats en unes taules (taules de Fisher-Yates). El número de contrastos que es pot fer és igual al número de condicions experimentals menys una (k-1). Per exemple, si tenim una variable independent amb 4 valors, podem valorar si les dades s'ajusten a una tendència lineal, quadràtica o cúbica. Els coeficients que trobem a les taules de Fisher-Yates per a aquest cas són:

	Tractament 1	Tractament 2	Tractament 3	Tractament 4
Lineal	-3	-1	1	3
Quadràtic	1	-1	-1	1
Cúbic	-1	3	-3	1

Recuperant l'exemple numèric que s'ha estat analitzant, la representació gràfica dels resultats per als 4 grups experimentals és la següent:



Suposant que la variable independent sigui quantitativa, es podria realitzar una anàlisi de tendències per valorar quina funció s'aproxima millor als resultats obtinguts.

La simple inspecció del gràfic ens mostra una tendència creixent que per als tres primers valors de la variable independent sembla força lineal, però amb un canvi de dinàmica en el quart valor de la variable independent, ja que el creixement es fa molt més lent.

De la mateixa forma en que ho hem fet abans per altres contrastos parcials, podem calcular les sumes de quadrats que correspondrien als contrastos lineal, quadràtic i cúbic. Una vegada fet això (per simple aplicació de les fórmules que s'han vist abans, aplicant els coeficients corresponents), el resultat és el següent:

Quadre resum de l'ANOVA per a l'anàlisi de tendències:

Quadre-resum de l'ANOVA per a l'anàlisi de tendències					
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	P
Entre-grups	470	3	156,67	39,17	0,000
Lineal	441	1	441	110,25	0,001
Quadràtic	20	1	20	5,00	0,040
Cúbic	9	1	9	2,25	0,153
Intra-grups	64	16	4,00		
TOTAL	534	19			

Com era de preveure a la vista del gràfic de resultats, el contrast lineal mostra un nivell de significació molt gran i recull la major part de la variabilitat observada. Això ens indica que la tendència és predominantment lineal. Ara bé, el fet que la tendència quadràtica estigui també dins dels límits de significació ens fa veure que no hi ha una tendència lineal perfecta, cosa que s'explica fàcilment pel canvi observat en el resultat del darrer grup experimental. Queda la incògnita de saber què passaria si s'afegissin més valors de la variable independent, i com podria afectar això a la funció que relaciona la variable independent amb la dependent.

Tots els contrastos que s'han plantejat fins el moment estaven destinats a provar hipòtesis específiques. Volíem saber, per exemple, si els resultats dels dos primers grups diferien entre sí, o si els dos primers grups mostraven resultats diferents als dos últims. O, en el cas de l'anàlisi de tendències, volíem aproximar-nos al tipus de relació que relacionava la variable independent i la dependent, i per això fèiem servir uns contrastos especials. En tots aquests casos, hem definit uns contrastos concrets per respondre a preguntes també concretes. Quan fem això, es pot dir que hem plantejat una sèrie de **contrastos o comparacions específics, planificats o a priori**, en el sentit de que hem definit quins ens interessaven des d'un punt de vista teòric o pràctic. Ara bé, en molts estudis, sobretot si són de tipus exploratori, podem tenir interès en obtenir tota la informació que ens poden oferir les dades, sense fixar-nos en cap hipòtesi o contrast concret. En el cas que s'està analitzant, per exemple, podríem tenir interès en valorar tots els contrastos possibles entre dos grups (grup 1- grup 2, grup 1 – grup 3, grup 1 – grup 4, grup 2 – grup 3, grup 2-grup 4, grup 3-grup 4) i saber quines diferències són significatives i quines no. Quan volem valorar un nombre gran de contrastos, és més senzill recórrer a determinades proves que estan pensades especialment per això, ja que fer-ho amb la tècnica que s'ha explicat abans seria molt

llarg. Hi ha doncs, proves específiques que serveixen per valorar **contrastos múltiples, no planificats o a posteriori**.

Un dels problemes de realitzar comparacions múltiples entre mitjanes a partir d'uns resultats és que la probabilitat cometre un error tipus I s'incrementa en funció del número de comparacions realitzades. Si fem c contrastos, cadascun d'ells amb el criteri habitual de $\alpha = 0,05$, la probabilitat de cometre com a mínim un error tipus I és igual a $1-(1-\alpha)^c$. Si fem dos contrastos aquesta probabilitat és igual a 0,0975, i si en fem 3 la probabilitat de cometre com a mínim un error puja a 0,143.

S'han proposat proves molt diverses que intenten resoldre o com a mínim reduir el problema de l'acumulació de la probabilitat de cometre un error tipus I. Hi ha diferents característiques que cal tenir en compte per distingir entre les diferents proves: si assumeixen o no la condició d'homoscedasticitat, si serveixen només per a comparacions dos a dos o per comparacions més complexes, si exigeixen o no que el número de subjectes per grup sigui igual, si es poden fer en una sola etapa de càlcul o en més d'una, etc. Donada aquesta complexitat, i a títol purament il·lustratiu, es presenta una taula amb moltes de les proves existents (no totes). Cal indicar que les més utilitzades a la pràctica són probablement les proves de Scheffé, DHS de Tukey i Bonferroni, tot i que les aplicacions estadístiques ens poden oferir el resultat de gairebé totes elles. Per simplificar, es distingeix únicament entre les proves que necessiten el compliment de la condició d'homoscedasticitat i les que no.

Assumeixen homoscedasticitat	No assumeixen homoscedasticitat
DHS de Tukey	T2 de Tamhane
Scheffé	T3 de Dunnett
Bonferroni	C de Dunnett
DMS de Fisher	Games-Howell
Sidak	
GT2 de Hochberg	
Gabriel	
Dunnet	
Duncan	
R-E-G-W F	
R-E-G-W Q	
S-N-K	
Tukey-b	
Waller-Duncan	

Cadascuna d'aquestes proves té el seu procediment de càlcul, i el seu funcionament no es detallarà aquí.

. L'anàlisi multivariable de la variància

De vegades en un disseny experimental pot ser interessant no utilitzar una sola variable dependent, sinó més d'una. Com es veurà més endavant, això ens permetrà obtenir informació sobre l'impacte de la variable o variables independents sobre les diferents variables dependents, i també ens indicarà si el comportament de les diferents variables dependents està relacionat o

no (per exemple, si diferents VD reaccionen de forma similar davant la introducció de determinats tractaments experimentals).

En aquests casos es podrien realitzar anàlisis de la variància separats per a cadascuna de les variables dependents, però tenim una alternativa més acurada, pensada específicament per a aquest cas. Aquesta alternativa és l'anàlisi multivariable de la variància (AMVAR o MANOVA), que no és sinó una generalització de l'anàlisi de la variància als casos en els quals treballem amb més d'una variable dependent.

Un exemple típic d'estudi multivariable és aquell en el qual es vol valorar si les instruccions rebudes per les persones influeixen en el balanç entre rapidesa i exactitud quan es realitza una determina tasca. Imaginem que a una tasca de vigilància (per exemple, amb una pantalla de radar) demanem a les persones que premin un botó cada vegada que aparegui un senyal a la pantalla. Registrem el temps de reacció (temps que tarden en respondre) i l'exactitud en l'execució de la tasca (percentatge d'encerts). Suposem també que volem provar el possible impacte de tres tipus d'instruccions diferents que podem donar als subjectes: a una d'elles es posa l'èmfasi en la rapidesa de resposta, a una altra s'insisteix especialment en l'exactitud, i un tercer tipus d'instrucció seria neutre o equilibrat. El que podem esperar és que les instruccions dirigides a la rapidesa faran disminuir el temps de reacció, però també l'exactitud, ja que una resposta més ràpida probablement comportarà una major probabilitat d'error; al contrari, les instruccions que incentiven l'exactitud tindrien un efecte contrari. En tot cas, el que això hauria de produir és una evolució coordinada de les dues variables dependents (TR i percentatge d'encerts), un temps de reacció ràpid probablement vindrà acompanyat per un percentatge d'encerts inferiors, i al contrari quan la resposta sigui més lenta.

La taula de resultats d'aquest exemple tindria un aspecte com el següent:

Instruccions 1 (demanar rapidesa)		Instruccions 2 (demanar exactitud)		Instruccions 3 (neutres)	
T.R.	% encerts	T.R.	% encerts	T.R.	% encerts
15,2	85,0
13,7	78,7
18,9	81,2
21,5	83,4
.....
.....
\bar{Y}_{tr1}	\bar{Y}_{enc1}	\bar{Y}_{tr2}	\bar{Y}_{enc2}	\bar{Y}_{tr3}	\bar{Y}_{enc3}

Els càlculs implicats en el MANOVA són força més complicats que en el cas de l'anàlisi de la variància convencional, ja que cal tenir en compte no solament la variació o la variància dels resultats tant entre els grups com dins dels grups per a cada variable dependent, sinó també la covariació o la covariància entre les diverses variables dependents. Per tant, un component fonamental per al MANOVA és la *matriu de variàncies i covariàncies* dels resultats. En realitat, el que fa el MANOVA és similar al que es feia en el cas de l'anàlisi de la variància: A l'anàlisi de la variància convencional, hem descompost la variació total en la variació vinculada als tractaments i la vinculada a l'error experimental. Aquesta descomposició es realitza també a l'anàlisi multivariable de la variància, però aquí el que es descompon no és solament la variació,

sinó la variació i covariació dels resultats. D'aquesta manera, la matriu de variacions i covariacions total (o la corresponent de variàncies i covariàncies) es descompon (en el cas més senzill) en dos elements: Una matriu H vinculada als tractaments i una matriu E vinculada a l'error experimental.

No es descriurà aquí tot el procediment de càlcul del MANOVA, es farà únicament una descripció molt resumida dels punts principals i de la terminologia bàsica. De forma molt sintètica, els passos a realitzar són:

. A partir de les matrius H i E, es planteja una *equació determinant*, que pren la forma següent:

$$| H - \lambda E | = 0$$

. Coneixent els valors de H i E, la resolució d'aquesta equació ens permet obtenir els valors de λ , coneguts com a *arrels característiques* de l'equació

. A partir de les arrels característiques, es poden obtenir diferents índexs que ens permetran valorar la significació o no dels resultats: Lambda de Wilks, traça de Hotelling-Lawley, traça de Pillai-Bartlett, criteri d'arrel característica major de Roy, i F generalitzada

. Alguns dels índexs anteriors es poden valorar directament, ja que s'ha calculat la seva distribució de probabilitat i per tant es disposa de taules estadístiques per fer-ho. Una opció més senzilla i molt utilitzada és la de transformar els índexs anteriors en una F de Snedecor convencional o a una chi-quadrat, cosa que en facilita la interpretació, tot i que s'ha d'indicar que no sempre aquestes aproximacions són totalment exactes.

Tot i la certa complexitat dels càlculs implicats, cal indicar que les aplicacions estadístiques més habituals permeten realitzar l'anàlisi multivariable de la variància amb tota facilitat.

Cal remarcar una vegada més que no es tracta únicament de "llegir" els resultats de la prova estadística i decidir si hi ha o no un o més efectes significatius, sinó també d'interpretar correctament les dades. En el cas dels dissenys multivariables, com ja s'ha dit, un aspecte molt important serà valorar si existeix covariació entre les variables dependents, ja que això ens donarà una informació suplementària sobre el possible impacte de la variable o variables independents que no tindríem en cas de treballar amb un plantejament univariable.

Com en el cas de l'anàlisi de la variància, la MANOVA té algunes condicions d'aplicació: Multinormalitat de les dades (ajust a una distribució normal multivariable), homogeneïtat de les matrius de variàncies-covariàncies de cada grup (valorable amb diferents estadístics, per exemple M de Box), independència dels registres dins de cada variable dependent (no entre les diferents variables dependents, ja que com s'ha vist aquestes poden estar correlacionades).

II Dissenys experimentals clàssics o estadístics

Com s'ha indicat anteriorment, els dissenys experimentals clàssics o estadístics es basen en la idea de treballar amb diferents grups de subjectes als quals s'apliquen les diverses condicions experimentals (excepte en el cas del disseny de mesures repetides, on els diversos tractaments són aplicats a un mateix grup de subjectes). La lògica implícita en aquesta forma de treballar és que si distribuïm equitativament les persones participants en els diferents grups, les diferències individuals rellevants (és a dir, les possibles variables de confusió relacionades amb el subjecte) quedaran neutralitzades i els diferents grups seran equivalents pel que fa aquestes variables. Això ens permetrà fer una comparació adient entre els resultats experimentals dels diferents grups, amb la finalitat de determinar el possible impacte de la variable o variables independents.

La forma normal de treballar en el marc dels dissenys clàssics és la següent (per simplificar, de moment suposarem que treballem amb una sola variable independent i una sola variable dependent):

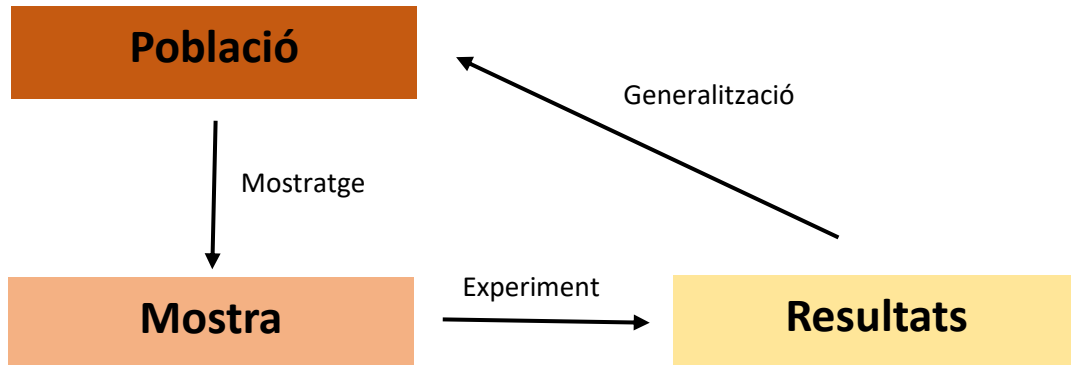
- . Seleccionar la mostra de subjectes participants a l'estudi. Si es volen generalitzar els resultats (com seria desitjable normalment), aquesta mostra ha de ser representativa de la població de referència i, per tant, cal posar molta atenció a la forma de selecció de la mostra
- . Definir els valors de la variable independent i establir com es faran operatius. Establir també la forma de registre de la variable dependent, i la seva validesa, fiabilitat i sensibilitat
- . Definir tants grups com valors tingui la variable independent, de forma que cada grup rebrà posteriorment un dels tractaments experimentals (novament amb l'excepció dels dissenys de mesures repetides). Com ja s'ha indicat, el criteri fonamental és que els grups siguin equivalents pel que fa a qualsevol possible variable de confusió rellevant. La forma d'aconseguir això serà un dels elements més importants que ens permetran distingir entre els diferents tipus de dissenys experimentals clàssics
- . Aplicar els tractaments i obtenir els resultats per a cada subjecte i, a partir d'ells, calcular els resultats globals per a cada grup (mitjana i variància, com a mínim, si la variable dependent és quantitativa)
- . Analitzar les dades mitjançant tècniques estadístiques, de forma que s'estableixi si la diferència entre els resultats obtinguts als diferents grups és prou gran com per considerar-la significativa, és a dir, per pensar que podem rebutjar la hipòtesi nul·la i considerar que hi ha un efecte real dels tractaments experimentals.

. El problema del mostratge

És important fer una referència a la qüestió del mostratge i als problemes que porta associats. Com s'ha indicat, l'objectiu natural d'un estudi experimental és el d'obtenir uns resultats que siguin vàlids no solament per a les persones que participen a l'estudi, sinó que puguin ser generalitzats a la població de referència. Per aconseguir això és fonamental que els subjectes experimentals constitueixin una **mostra representativa** de la població a la qual es volen generalitzar els resultats. Per donar resposta a aquesta necessitat s'han plantejat diferents tipus de mostratge (aleatori simple, aleatori sistemàtic, aleatori per conglomerats, per rutes

aleatòries, per quotes, etc.), sempre amb la idea de que la mostra escollida reproduïx de la millor forma possible les característiques de la població.

L'esquema general és ben conegut:



Hi ha diferents problemes o limitacions que sovint fan difícil la realització d'un mostratge adient:

. Com és sabut, la representativitat de la mostra s'incrementa a mesura que la seva grandària és superior. Ara bé, en moltes ocasions simplement no és possible treballar amb el número de persones que seria ideal des d'un punt de vista de la representativitat mostral. A l'hora de realitzar un estudi experimental és habitual que hi hagi limitacions de temps, econòmiques, de material disponible, d'accessibilitat als subjectes, etc. Per aquest motiu, molt sovint es treballa amb grandàries mostrals que són inferiors a les que serien ideals des d'un punt de vista de la seva representativitat

. Estrictament, fer un mostratge aleatori pur és impossible ja que es necessita, com és evident, el consentiment de les persones que han de participar en l'estudi (o dels seus familiars, en el cas de menors, persones incapacitades, etc.). En altres paraules, els subjectes experimentals han de ser voluntaris i, per tant, això descarta automàticament a totes les persones que no tinguin interès o motivació en participar a l'estudi. Per animar la participació de les persones en els estudis experimentals de vegades es recorre a determinats tipus d'incentius, per exemple de tipus econòmic, acadèmic (crèdits de curs, millores de qualificacions, etc.), temps alliberat de feina a una empresa, o altres. Això pot introduir un biaix a l'hora d'obtenir participants: si, per exemple, hi ha una recompensa de tipus acadèmic, és possible que les persones que es presentin voluntàries siguin les que més necessiten d'aquest incentiu (per exemple, estudiants amb baixes qualificacions), cosa que posa en dubte encara més la representativitat de la mostra.

. Un mostratge aleatori implica una definició adient de la població de referència i la possibilitat d'accedir al conjunt d'aquesta població per escollir els participants a l'estudi. Això en molts casos és senzillament impossible. Si, per exemple, es treballa en el tema dels trastorns de memòria associats a la malaltia d'Alzheimer, el plantejament lògic seria tenir com a població de referència totes les persones afectades en un conjunt poblacional ampli (una ciutat, un país, etc.). En moltes ocasions, per raons purament pràctiques de coneixement i accessibilitat, els investigadors s'han de limitar a treballar amb els casos als quals puguin tenir més fàcil accés (per exemple, les persones afectades que són tractades al centre de salut o a l'hospital on treballa l'investigador/a). Aquí no hi ha, doncs, un mostratge aleatori. En aquests casos es parla de

mostratge incidental o de conveniència, tot i que aquesta definició inclou també alguna altra situació.

Un exemple evident de mostratge incidental és aquell en el qual els investigadors que treballen a la universitat tendeixen moltes vegades a tenir com a subjectes els seus propis estudiants. Aquest fenomen s'ha produït de forma molt àmplia, de forma que s'ha arribat a dir, de forma humorística, que *la psicologia és la ciència de la conducta..... dels estudiants universitaris*, modificant així una de les definicions clàssiques de la psicologia. Afortunadament, hi ha una consciència creixent d'aquest problema, i des de fa anys és més fàcil trobar estudis psicològics amb mostres no universitàries.

. Tipus de dissenys experimentals clàssics

Hi ha una diversitat important de dissenys experimentals clàssics, com ja s'ha indicat abans. Per tenir-ne una visió més completa es poden utilitzar diferents criteris de classificació, com a la taula següent:

Classificació dels dissenys experimentals clàssics	
Criteris	Tipus de dissenys
Nº variables independents	Una: Simples Més d'una: Factorials
Nº variables dependents	Una: Univariables Més d'una: Multivariables
Nº registres per unitat i condició	Un: Estàtics o transversals Més d'un: Longitudinals
Tècnica utilitzada per millorar l'equivalència dels grups	Aleatorització: Grups aleatoritzats Bloqueig, aparellament: Grups homogenis Subjecte control propi: Mesures repetides

Cal indicar que tots aquests criteris es poden combinar per donar lloc a dissenys experimentals concrets. Per exemple, es pot plantejar un disseny simple, multivariable, estàtic i de grups homogenis; o un disseny factorial, univariable, longitudinal i de grups aleatoritzats; o un disseny simple, univariable, estàtic i de mesures repetides, etc, etc.

. Conceptes bàsics relacionats amb cadascun dels criteris de classificació

- Simples - Factorials

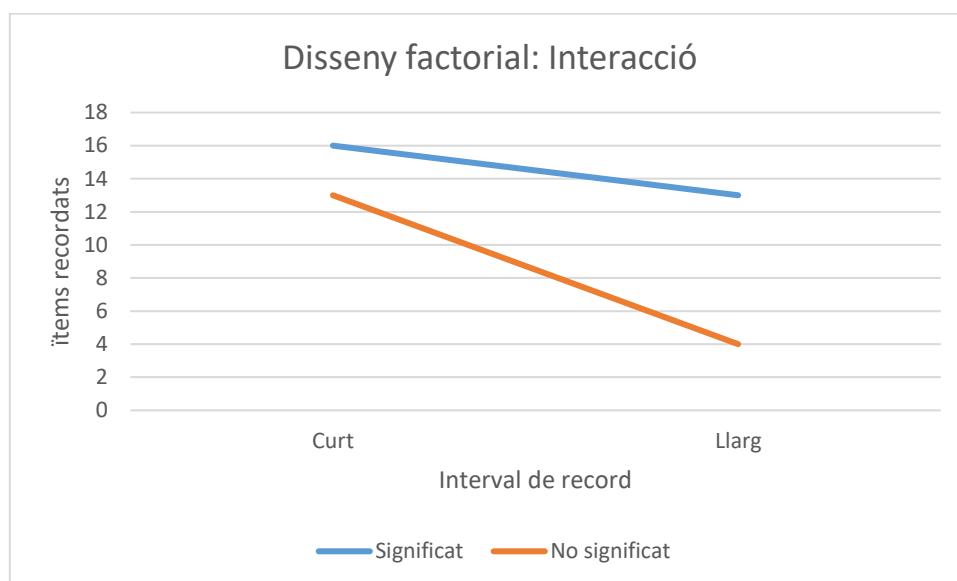
El disseny simple és el cas més elemental, en el qual s'intenta veure quin és l'impacte d'una determinada variable independent sobre una o diverses variables dependents. Per tant, no implica una complicació especial, però al mateix temps ens ofereix una informació limitada. Existeix, però, la possibilitat de realitzar experiments que incorporin més d'una variable independent, cosa que ens ofereix noves i importants possibilitats.

Suposem que en un estudi sobre memòria humana es planteja un disseny factorial amb dues variables independents: Tipus de material (material amb significat –paraules- vs material sense

significat - pseudoparaulas) i interval de record (interval curt i interval llarg). Es pretén esbrinar si aquestes variables tenen algun impacte sobre el record posterior del material, mesurat en termes del número d'ítems recordats d'entre els presentats a cada assaig.

En un disseny factorial és possible analitzar no solament el possible impacte de cadascuna de les variables independents sobre la dependent o dependents, sinó també la possible **interacció** entre les variables independents. Es parla d'interacció entre dues variables independents quan l'impacte d'una d'elles sobre la dependent és modificat per la presència de l'altra variable independent.

A l'exemple plantejat, suposem que els resultats experimentals obtinguts es representen gràficament de la forma següent:

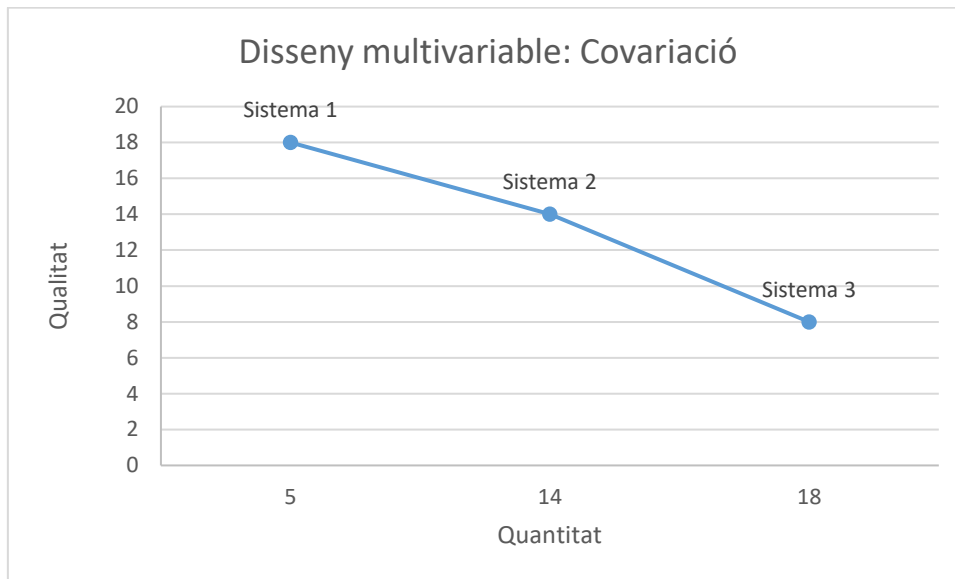


La simple inspecció d'aquest gràfic mostra diferents coses: En primer lloc, es recorda millor el material després d'un interval curt que després d'un interval llarg (això és cert globalment, tant si el material és amb significat com si no); per altra banda, de forma general el material amb significat es recorda millor que el material sense significat (tant si l'interval és curt com llarg). Ara bé, es pot veure com l'efecte de cadascuna de les variables independents es veu modificada per la presència de l'altra. Per exemple, es pot constatar que, tot i que els materials amb significat sempre es recorden millor que els materials sense significat, aquesta diferència és més gran quan l'interval de record és llarg que quan és curt. També podem dir que sempre es recorda millor en intervals curts que en intervals llargs, però aquesta diferència és més gran quan es treballa amb material sense significat que quan s'utilitza material amb significat. Per tant, les variables independents s'influeixen mútuament, en el sentit de que l'impacte d'una es veu modificat per la presència de l'altra.

- Univariables – Multivariables

L'ús d'un disseny multivariable (amb més d'una variable dependent) ofereix algunes possibilitats que no tenim en un disseny univariable. Com s'ha dit abans, la més important és la capacitat de

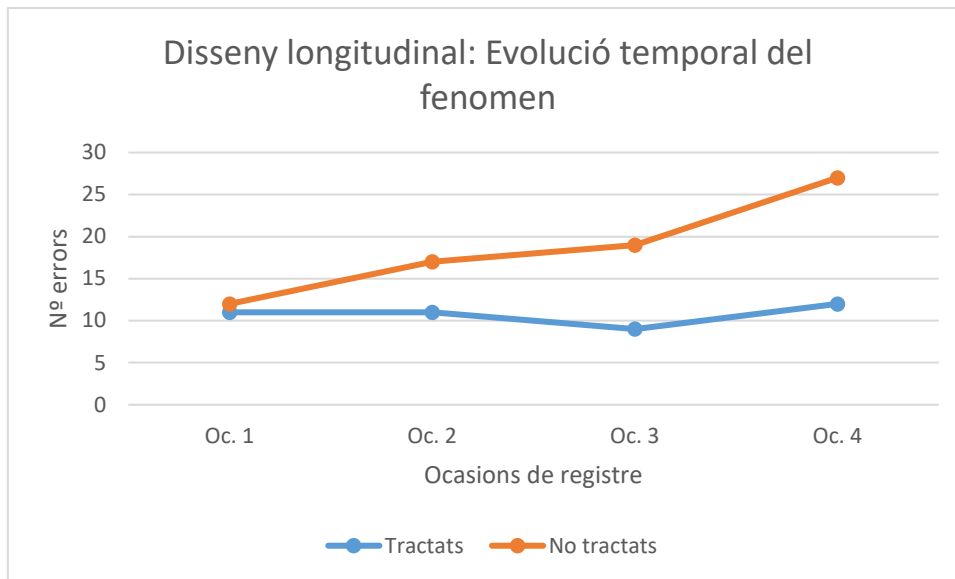
valorar la possible covariació entre els resultats de les diferents variables dependents, com es planteja al gràfic següent:



L'estudi de la covariació (o la correlació) dels resultats de les diverses variables dependents ens dona informació valuosa sobre l'impacte dels tractaments experimentals. A l'exemple del gràfic, veiem l'impacte de tres formes diferents d'organitzar la feina d'una empresa (sistema o tractament 1, sistema 2 i sistema 3) sobre la quantitat i la qualitat de productes fabricats. Els resultats mostren que el primer sistema organitzatiu ofereix un gran nivell de qualitat, però una quantitat de producció baixa; al contrari, el sistema 3 té com a resultat un nivell de producció alt, però amb qualitat baixa, mentre que el sistema 2 està en una posició intermèdia. En funció dels resultats podem escollir quina és la millor opció organitzativa, considerant que el sistema 2 probablement sigui el més equilibrat, tenint en compte que en aquest cas la covariació entre les dues variables dependents és negativa i que tant la qualitat com la quantitat són importants.

- Estàtics – Longitudinals

A diferència dels dissenys estàtics, els dissenys longitudinals ens permeten fer el seguiment de l'evolució en el temps del fenomen que ens interessi. Imaginem, per exemple, que es vol esbrinar si l'aplicació d'un programa de suport cognitiu a persones que presenten un procés progressiu de deteriorament permet evitar o com a mínim alentar aquest procés. Com a variable dependent s'utilitza el nombre d'errors comesos en una prova de raonament lògic. En tractar-se d'un procés que es desenvolupa progressivament, es fan 4 registres de la variable dependent (ocasions de registre), amb intervals de tres mesos. Suposem que els resultats obtinguts es representen al gràfic següent:



Com és evident, el caràcter longitudinal del disseny ens informa de l'evolució temporal d'allò que volíem estudiar. Sembla clar que el programa de suport aconsegueix establir el procés de deteriorament, com a mínim dins dels marges temporals de l'estudi. També és molt clar que si només s'hagués fet un registre després de començar a aplicar el tractament (ocasió 1), com seria el cas d'un disseny estàtic, no es podria apreciar l'impacte de la intervenció, que només es posa de manifest en les ocasions de registre posteriors.

Cal dir que l'estudi plantejat implica problemes ètics rellevants, ja que el grup control no és tractat i, per tant, està subjecte al procés de deteriorament que es pot produir (tot i que abans de fer l'estudi tampoc sabem si el tractament serà eficaç o no). Per tant, fer l'experiment d'aquesta manera té sentit des d'un punt de vista metodològic, ja que permet valorar el possible efecte del tractament, però necessita de reflexió des d'una perspectiva ètica.

Com intentar assegurar l'equivalència dels grups?: Disseny experimental de grups aleatoritzats, de grups homogenis i de mesures repetides

1.- Disseny de grups aleatoritzats (DGA)

Els dissenys de grups aleatoritzats intenten neutralitzar l'efecte de les possibles variables de confusió relacionades amb el subjecte a partir de l'aplicació de la tècnica d'aleatorització en el moment de formar els grups experimentals. Si disposem d'un total de N subjectes experimentals i volem valorar l'impacte de k tractaments o condicions, el que farem és distribuir els N subjectes en k grups utilitzant qualsevol tècnica aleatòria: una taula de números aleatoris, un sorteig, un dau, una moneda o qualsevol altre. Per tant, cada persona té la mateixa probabilitat de ser assignada a qualsevol dels grups experimentals, i d'aquesta forma evitem qualsevol biaix a l'hora de fer aquesta assignació. Cal tenir en compte, però, una restricció: és convenient que el número de subjectes sigui igual o similar per als diferents grups, és a dir, assegurar-se de que les

diferents condicions experimentals són aplicades a un número de subjectes més o menys semblant (o que el número de rèpliques dins de cada grup és similar).

Una vegada formats els grups, cadascun d'ells rep un dels tractaments experimentals i s'obtenen els corresponents resultats. Cal dir que la determinació de quin tractament s'aplica a cadascun dels grups també s'ha de fer aleatòriament. Per tant, l'aplicació de la tècnica d'aleatorització es produeix en dos moments: a l'hora d'assignar els subjectes als grups i a l'hora de determinar quina condició experimental s'aplica a cadascun dels grups formats.

La tècnica d'aleatorització té dues limitacions importants:

. En primer lloc, en tractar-se d'una tècnica probabilística, la seva eficàcia depèn de forma crucial de la grandària de la mostra disponible. Si el número de subjectes N és petit, augmenta la possibilitat de que la distribució en els diferents grups no sigui prou equitativa. Si disposem de 15 subjectes que hem de repartir en tres grups de 5, la probabilitat de que, per exemple, les persones amb major ansietat vagin a parar majoritàriament a un dels grups no és despreciable, i si l'ansietat és una variable que pot influir en els resultats, tenim un risc evident de confusió. En canvi, si en el mateix cas disposéssim de 150 subjectes, a repartir en tres grups de 50, és evident que la probabilitat de que es produeixi aquest mateix problema és molt menor. La dificultat principal, òbviament, és que a la pràctica no és fàcil disposar de mostres grans de subjectes, i això crea un problema rellevant per a l'ús de dissenys de grups aleatoritzats.

. En segon lloc, moltes vegades utilitzem un disseny de grups aleatoritzats perquè no disposem d'informació (o no podem utilitzar-la) sobre les característiques individuals de les persones, de forma que ens limitem a distribuir-les aleatòriament en els grups, confiant en l'actuació de l'atzar per assolir grups més o menys equivalents. Com veurem més endavant, si podem aconseguir alguna informació prèvia sobre els participants (per exemple, sobre característiques de personalitat, intel·ligència, variables sociodemogràfiques, etc.) tenim la possibilitat de no confiar només en l'atzar, sinó d'intentar aconseguir grups equivalents a partir de l'ús d'aquesta informació preliminar que tenim (veure apartat de dissenys de grups homogenis). Com es veurà a l'apartat corresponent, aquesta forma de procedir ens dona més garanties que el simple ús de la tècnica d'aleatorització.

En el marc dels dissenys de grups aleatoritzats de vegades s'obtenen registres pretest, és a dir, registres de la variable dependent abans d'introduir els tractaments experimentals. D'aquesta forma es pot constatar si realment els grups són més o menys equivalents abans de començar l'experiment, com a mínim pel que fa als seus resultats en els registres pre-test. Si, per exemple, es tracta d'intentar modificar les actituds d'estudiants de secundària respecte del problema de les drogues mitjançant l'aplicació de diferents tipus de programes informatius/educatius, pot ser interessant fer una mesura d'actituds (a través d'una escala psicomètrica) abans d'introduir els diversos programes. Això ens permetrà veure si el punt de partida és similar en els diferents grups, i també ens donarà un bon element de comparació amb els resultats experimentals posteriors. Ara bé, cal dir que si disposem d'aquesta informació prèvia i encara tenim la possibilitat de modificar els grups, el més intel·ligent seria utilitzar-la per "reconstruir" els grups i fer-los més equivalents, com es veurà en parlar dels dissenys de grups homogenis.

En un disseny de grups aleatoritzats, com també en altres tipus de dissenys, pot ser útil també la introducció d'un grup control, és a dir, un grup que no rebrà tractament. Això ens ofereix un altre element de comparació interessant per valorar millor l'impacte dels diferents tractaments.

Al darrer exemple plantejat, un grup control seria aquell que no rep cap mena de programa informatiu o educatiu i, per tant, no esperaríem cap canvi en els seus resultats. La comparació amb altres grups, que sí haurien rebut algun tipus d'intervenció, ens seria molt útil per valorar el possible efecte dels diferents programes.

L'exemple numèric que s'ha utilitzat abans en parlar de l'anàlisi de la variància corresponia precisament a un disseny simple de grups aleatoritzats. L'estructura de la taula de dades d'un disseny com aquest pot ser aproximadament com aquesta (s'utilitzen les mateixes dades ja mostrades a l'apartat de bases estadístiques):

Taula de dades per a un DGA simple				
Tractaments	Tractament 1	Tractament 2	Tractament 3	Tractament 4
Resultats individuals	8	13	18	17
	10	15	20	21
	9	10	22	21
	7	13	18	23
	6	14	17	18
Mitjanes	8	13	19	20
TOTAL = Tractaments + Error				

Com ja s'ha indicat també anteriorment, la informació disponible en un disseny de grups aleatoritzats ens permet descompondre la variació total de les dades en dos components: Una variació entre-grups, que ens indica el possible impacte dels tractaments experimentals, i una variació intra-grups, que reflecteix l'efecte de tot un conjunt de variables que poden influir en els resultats i que no hem controlat de forma directa. Com ja s'ha indicat repetidament, el disseny de grups aleatoritzats es basa en el supòsit que la distribució aleatòria dels subjectes en els diferents grups permet equilibrar o neutralitzar l'efecte d'aquestes variables, de manera que no produeixin diferències entre els grups experimentals, però el seu impacte es pot constatar amb l'existència d'una variació intra-grups o error experimental.

La forma de l'anàlisi de la variància en un disseny de grups aleatoritzats també és coneguda, i respon a l'esquema general següent:

DGA Simple	Quadre-resum de l'anàlisi de la variància				
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	p
Tractaments (entregups)		$k - 1$		Var Tract / Var error	
Error (ntragrups)		$N - k$			
TOTAL		$N - 1$			
k = número de valors de la variable independent N = número total de subjectes o casos					

Cal recordar que la valoració de la significació de l'efecte dels tractaments es realitza a partir de la comparació (mitjançant quocient) entre la variància entre-grups i la variància intra-grups, de

forma que intentem veure si la variància entre-grups és producte únicament d'una variabilitat aleatòria de les dades, o bé si reflecteix realment un efecte dels tractaments experimentals (veure apartat de consideracions estadístiques).

És important fer una referència al problema de la sensibilitat del disseny de grups aleatoritzats (no s'ha de confondre això amb el tema de la sensibilitat de la variable dependent, que és una qüestió diferent). La sensibilitat del disseny és la seva capacitat per detectar un efecte real de la variable independent, cosa que té a veure sobretot amb tres aspectes: En primer lloc, amb la magnitud de l'efecte dels tractaments, si existeix; naturalment, un efecte més gran és també més fàcil de detectar. En segon lloc, amb la magnitud de l'error experimental; quan més gran sigui la variabilitat produïda per les diferents variables que poden incidir en la situació, més difícil serà detectar el possible efecte del tractament. Cal recordar que en analitzar les dades fem una comparació entre la variància entre-grups (que pot contenir l'efecte sistemàtic dels tractaments) i la variància intra-grups o error experimental, que ens indica l'impacte dels factors aleatoris que poden incidir en els resultats; lògicament, quan més gran sigui l'error experimental, més difícil serà detectar un possible efecte sistemàtic que es pugui haver produït. Finalment, un tercer element, relacionat amb els anteriors, és la potència de la prova estadística utilitzada. La potència depèn dels factors ja indicats (magnitud de l'efecte i magnitud de l'error experimental), però també d'aspectes com la grandària de la mostra, ja que mostres més grans ofereixen majors nivells de potència. Com a criteri general caldria dir que incrementar el número de casos de la mostra és una manera d'incrementar la potència de la prova en un disseny de grups aleatoritzats (i també a la resta de dissenys). En relació amb això cal prevenir, però, un efecte que tampoc és desitjable i que s'ha esmentat abans: Si es treballa amb mostres molt grans, és possible que la prova estadística detecti com a significatius efectes molt petits dels tractaments experimentals. Aquests efectes podrien ser gairebé negligibles des d'un de vista teòric o pràctic. Per això és important fer dues coses: per un costat, com s'indicava anteriorment, acompanyar l'anàlisi de la variància amb algun índex de magnitud de l'efecte; per un altre, acompanyar també els criteris estadístics amb criteris valoratius propis de cada camp d'estudi (per exemple, criteris clínics, socials, adaptatius, educatius etc.) per tal de fer una avaluació correcta dels resultats obtinguts. D'aquesta manera podrem determinar de la millor manera si un determinat resultat és realment important per a les finalitats teòriques o pràctiques de l'estudi experimental.

En el cas del disseny de grups aleatoritzats, en haver format els grups a l'atzar, sense tenir en consideració les característiques dels diferents subjectes experimentals, correm el risc de tenir un nivell de variabilitat important dels resultats, és a dir, un nivell elevat d'error experimental. Per aquest motiu el DGA és el disseny menys sensible dels que anirem estudiant. Com veurem més endavant, en el cas dels dissenys de grups homogenis i dels dissenys de mesures repetides realitzarem algunes accions destinades a reduir (o a analitzar més acuradament) l'error experimental, afavorint així la possibilitat de detectar el possible impacte dels tractaments. En aquest sentit, es pot afirmar que, a igualtat de grandària mostral, els dissenys de grups homogenis i de mesures repetides són més sensibles que el de grups aleatoritzats a l'hora de detectar un determinat efecte dels tractaments experimentals.

La versió factorial del disseny de grups aleatoritzats, tot i que comparteix les característiques generals de qualsevol disseny d'aquest tipus, té un nivell de complexitat superior. En un disseny factorial es treballa amb més d'una variable independent. Habitualment, les condicions experimentals es defineixen a partir del creuament dels valors de les diverses variables, tot i que existeixen altres estructures factorials menys freqüents (factors niats, dissenys fraccionats).

Per exemple, suposem que es vol esbrinar si la forma de presentació del material i la seva naturalesa (material amb significat o sense significat) influeixen en el seu record en una prova de memòria. Imaginem també que plantegem tres formes de presentació diferents (presentació visual, presentació auditiva, o presentació simultània pels dos canals) i que, tal com s'indica, es treballa amb material amb significat (paraules) i amb material sense significat (pseudo-paraules o similar). Esquemàticament:

Estructura factorial de creuament

Variable A:	A₁ : Presentació visual
	A₂ : Presentació auditiva
	A₃ : Tots dos canals
Variable B:	B₁ : Material sense significat
	B₂ : Material amb significat
Condicions experimentals	
A₁ B₁ :	Presentació visual, material sense significat
A₁ B₂ :	Presentació visual, material amb significat
.....	
A₃ B₂ :	Tots dos canals, material amb significat

Com es pot observar, en un disseny factorial (en el cas més usual, utilitzant una estructura de creuament) les condicions experimentals es defineixen per la combinació dels valors de les variables implicades. En el nostre cas, a la primera condició es fa una presentació visual de material sense significat, mentre que a l'última es presenta simultàniament pels dos canals (visual i auditiu) un material amb significat. De la mateixa forma es defineixen la resta de condicions.

En tenir una variable independent amb 3 valors i una altra amb 2, direm que es tracta d'un disseny factorial 3 x 2, i el número de condicions experimentals es determina fàcilment multiplicant el número de valors o tractaments de cadascuna de les variables: $3 \times 2 = 6$ condicions experimentals.

Per exemple, en un disseny factorial 4 x 4 x 2 tindriem tres variables independents, la primera amb quatre valors, la segona també amb quatre, i la tercera amb dos valors. El número de condicions experimentals seria de 32 ($4 \times 4 \times 2 = 32$). Es tractaria òbviament d'un disseny molt complex, ja que si ens movem dins la lògica dels dissenys de grups aleatoritzats, caldria disposar de 32 grups experimentals diferents, a cadascun dels quals s'aplicaria una de les condicions.

És fàcil veure que a la pràctica resulta molt complicat utilitzar dissenys factorials amb moltes variables independents, ja que el número de condicions experimentals creix molt ràpidament. Quan hi ha moltes condicions experimentals, hi ha la possibilitat de recórrer a algun tipus especial de disseny factorial (disseny factorial fraccionat) on només s'utilitza una part de les condicions (per exemple, la meitat). L'anàlisi i interpretació dels resultats d'aquests dissenys és

complicat i es perd molta informació, de manera que no seran tractats aquí. Tampoc es tractaran altres estructures factorials diferents de les de creuament (per exemple, estructures d’anidament), que poden ser útils només en circumstàncies molt concretes.

En un disseny factorial es poden analitzar bàsicament dos tipus d’efectes: El de cadascuna de les variables independents per sí mateixa, i la interacció que es pot produir entre elles, tot i que veurem immediatament que per arribar a valorar la interacció és útil el pas intermedi dels anomenats efectes simples. Això s’analitzarà a partir de tres exemples de resultats diferents, corresponents a un estudi hipotètic fet amb el disseny factorial més senzill possible (un disseny factorial 2 x 2). A cada casella s’indica la mitjana de resultats obtinguda pel grup experimental corresponent. A les files i columnes trobem les mitjanes corresponents a cadascun dels valors de les diferents variables (per simplicitat suposarem que tots els grups experimentals tenen el mateix nombre de subjectes).

. Exemple 1:

Disseny factorial 2 x 2

	B ₁	B ₂	Mitjanes
A ₁	8	4	6
A ₂	14	6	10
Mitjanes	11	5	8

Efectes principals	Efectes simples
Variable A: 10 – 6 = 4	Var. A / B ₁ : 14 – 8 = 6
Variable B: 5 – 11 = -6	Var. A / B ₂ : 6 - 4 = 2
	Var. B / A ₁ : 4 - 8 = -4
	Var. B / A ₂ : 6 - 14 = -8

Com s’ha indicat, els efectes que es poden estudiar són els següents:

Efectes principals: Efecte de cada variable per sí mateixa. S’obté comparant les mitjanes globals dels resultats obtinguts per als dos tractaments de cada variable (A₂ – A₁) i (B₂ – B₁). Per tant, per a la variable A: 10 – 6 = 4 ; i per a la variable B: 5 – 11 = -6. Convencionalment fem la diferència (A₂ – A₁), però naturalment es podria fer també la contrària (A₁ – A₂), de forma que l’únic que canviaria és el signe del resultat.

Per entendre la idea d’interacció és molt útil calcular els efectes simples de les variables. L’efecte simple de la variable A és l’impacte d’aquesta variable, però no en general, sinó quan ve acompanyada pels diferents valors de l’altra variable. En el nostre cas, per exemple, la variable A té dos efectes simples:

Efecte simple de A sota B₁: (A₂ – A₁) amb B₁ 14 – 8 = 6

Efecte simple de A sota B₂: (A₂ – A₁) amb B₂ 6 – 4 = 2

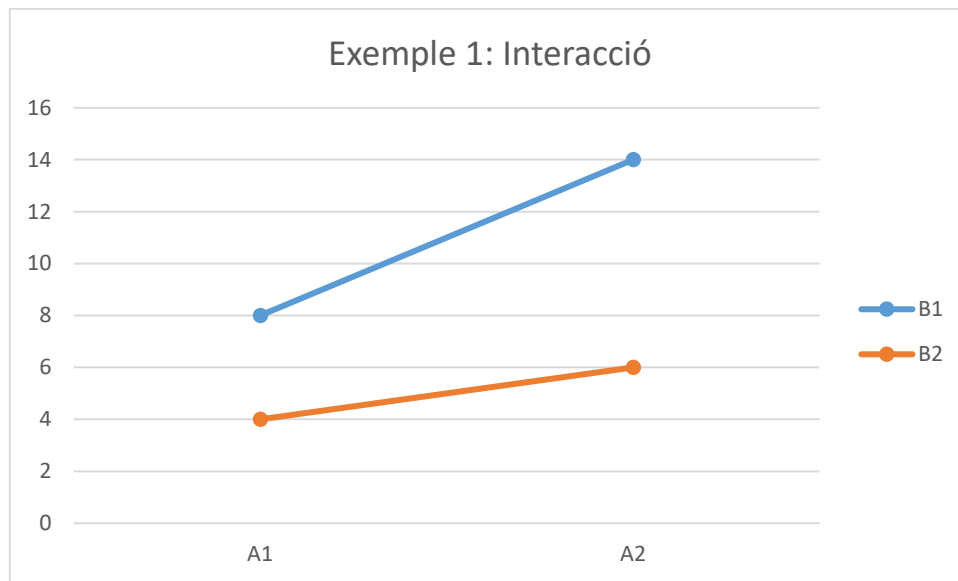
Es veu immediatament que l’efecte de A no és el mateix quan ve acompanyada pels diversos valors de B, cosa que fa pensar en una possible interacció. Naturalment, la mateixa operació es pot fer valorant els efectes simples de la variable B acompanyada pels diferents valors de A, els quals també resulten ser diferents:

Efecte simple de B sota A_1 : $(B_2 - B_1)$ amb A_1 $4 - 8 = -4$

Efecte simple de B sota A_2 : $(B_2 - B_1)$ amb A_2 $6 - 14 = -8$

Les diferències observades entre els efectes simples són un primer indicatiu de que pot existir interacció entre les dues variables, ja que la interacció consisteix precisament en el fet que la presència d'una segona variable independent modifica l'impacte de la primera (o viceversa).

La representació gràfica de les mitjanes dels resultats ens pot ajudar a interpretar-los millor, especialment si la fem de la forma següent:



En el gràfic es poden observar fàcilment els efectes principals, tant per a la variable A com per a la variable B, els efectes simples (per exemple, l'efecte de la variable B és clarament diferent segons quin valor de la variable A l'acompanyi), i la possible interacció de les dues variables. La interacció es posa de manifest pel fet de que les dues línies no són paral·leles, com a conseqüència lògica de que els efectes simples són diferents: en aquest cas, la diferència entre els resultats obtinguts amb els dos valors de la variable B és més gran sota la condició A_2 que sota la condició A_1 .

És important indicar que, en tot cas, tant aquesta possible interacció com els efectes principals de les variables s'han de valorar a la corresponent anàlisi de la variància, la qual ens indicarà si els efectes són estadísticament significatius o no. Els càlculs realitzats fins ara només ens ofereixen una primera aproximació, però és indispensable valorar els resultats a partir de l'anàlisi de la variància corresponent, que s'exposarà una mica més endavant.

Es mostra a continuació un segon exemple amb un patró d'interacció més complicat entre les variables independents.

. Exemple 2:

Disseny factorial 2 x 2

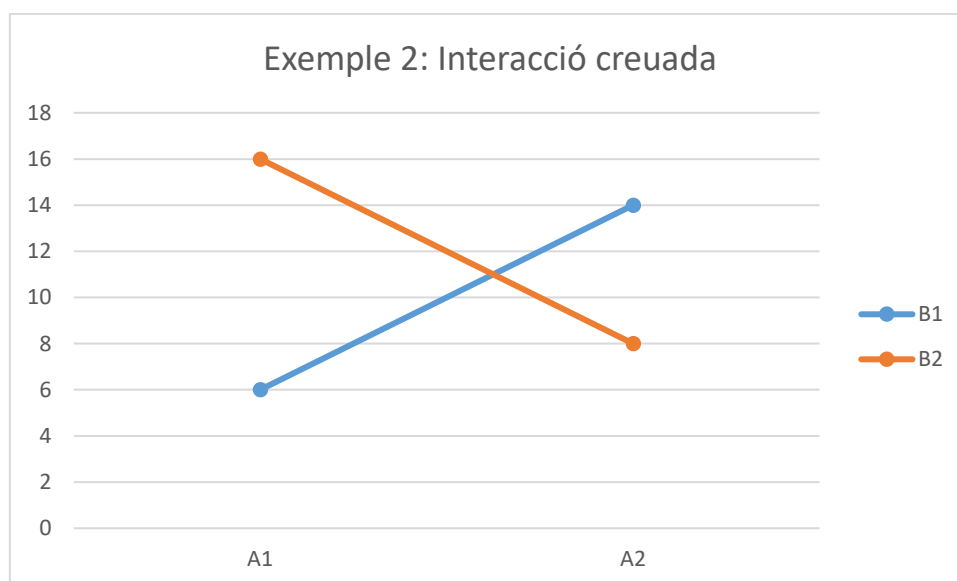
	B ₁	B ₂	Mitjanes
A ₁	6	16	11
A ₂	14	8	11
Mitjanes	10	12	11

Efectes principals	Efectes simples
Variable A: 11 – 11 = 0	Var. A / B ₁ : 14 - 6 = 8
Variable B: 12 – 10 = 2	Var. A / B ₂ : 8 - 16 = -8
	Var. B / A ₁ : 16 - 6 = 10
	Var. B / A ₂ : 8 - 14 = -6

Es planteja aquí un exemple molt particular, en el qual els efectes principals de les dues variables són molt petits. Ara bé, es pot observar que això no es deu a una manca d'impacte de les variables, sinó a la probable existència d'una interacció molt potent. Es pot veure que els efectes simples de la variable A, per exemple, són de magnitud similar però de signe diferent, per tant, són efectes totalment inversos. Una cosa semblant es pot trobar a la variable B. Això significa que l'efecte d'una variable no solament és modificat per l'altra, sinó que és invertit per la segona variable. Hi ha doncs, una interacció molt potent, anomenada "inversa" o "creuada", ja que l'efecte d'una variable és invers segons quin sigui el valor de l'altra.

Aquest és un cas particular d'interacció, no molt freqüent però possible. És important veure que una interacció creuada fa que els efectes principals de les dues variables siguin petits, cosa que podria fer pensar, erròniament, que les variables no tenen impacte sobre els resultats. Per això, al fer l'anàlisi d'un disseny factorial sempre és millor començar estudiant la interacció, i només en un segon moment posar atenció sobre els efectes principals.

El gràfic de resultats corresponent a aquest segon exemple aclareix totalment la situació:



El gràfic mostra amb total claredat la naturalesa de la interacció creuada, amb l'efecte d'inversió de l'impacte de les variables que ja s'ha explicat i que es posa de manifest a la taula de resultats.

L'existència d'interacció és una possibilitat real (i habitual) a un disseny factorial, però no té per què produir-se necessàriament. Es planteja a continuació un exemple on no hi ha interacció entre les variables independents.

. Exemple 3

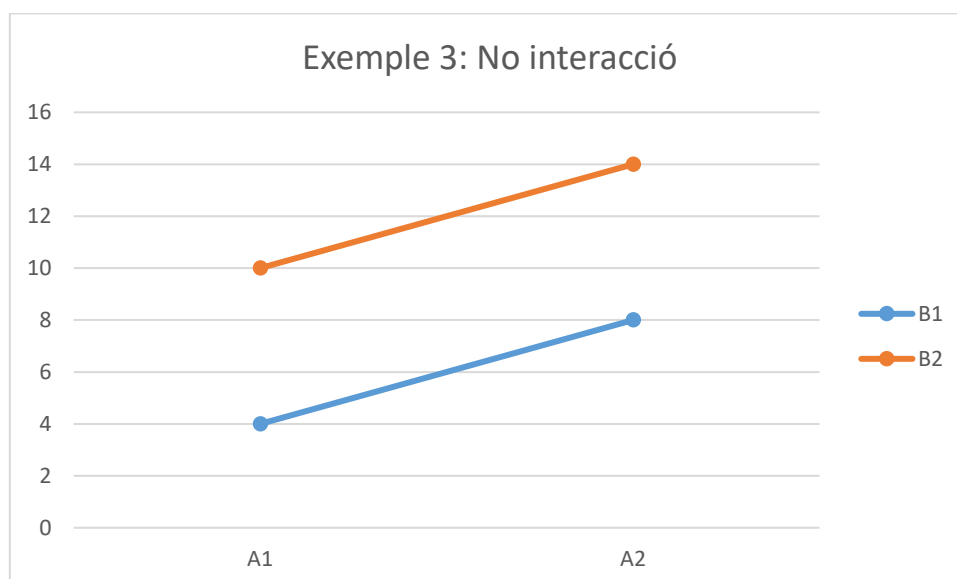
Disseny factorial 2 x 2

	B ₁	B ₂	Mitjanes
A ₁	4	10	7
A ₂	8	14	11
Mitjanes	6	12	9

Efectes principals	Efectes simples
Variable A: $11 - 7 = 4$	Var. A / B ₁ : $8 - 4 = 4$
Variable B: $12 - 6 = 6$	Var. A / B ₂ : $14 - 10 = 4$
	Var. B / A ₁ : $10 - 4 = 6$
	Var. B / A ₂ : $14 - 8 = 6$

En aquest cas es pot observar que els efectes simples de cada variable són idèntics entre sí. Això significa que les variables no s'influeixen mútuament, és a dir, que no existeix cap mena d'interacció entre elles.

El gràfic corresponent és el següent:



La inexistència d'interacció es posa de manifest en el fet que les dues línies del gràfic són paral·leles, de forma que queda clar que l'efecte de cadascuna de les variables independents no es veu modificat per la presència de l'altra. Sempre que s'observin línies paral·leles (o gairebé paral·leles) podem pensar en l'absència d'interacció entre les variables.

Com s'ha indicat, en qualsevol cas la valoració de la significació o no dels diferents efectes s'ha de fer mitjançant la corresponent anàlisi de la variància. En el cas de qualsevol disseny factorial de grups aleatoritzats amb dues variables independents, l'anàlisi de la variància prendrà la forma següent:

DGA Factorial (2 V.I.)	Quadre-resum de l'anàlisi de la variància				
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	p
Variable A		a - 1		Var A / Var intra	
Variable B		b - 1		Var B / Var intra	
Interacció A x B		(a-1)(b-1)		Var Ax B / Var intra	
Intra-grups		N - k			
TOTAL		N - 1			
a = número de valors de la variable A b = número de valors de la variable B k = número total de condicions experimentals (a x b) N = número total de subjectes o casos					

A l'anàlisi de la variància s'inclouen normalment els efectes principals de les variables i la seva possible interacció. En cas necessari també es podrien incloure els efectes simples, tot i que no és gaire habitual, ja que la interacció ja ens informa sobre les possibles diferències en l'impacte d'una variable en funció de la presència de l'altra.

És important veure que es manté la lògica bàsica dels dissenys de grups aleatoritzats: La variació total es descompon en dos elements: variació entre-grups i variació intra-grups (o error experimental). La novetat en el cas factorial és que la variació entre-grups es descompon en els elements que la poden explicar: Efecte de la variable A, efecte de la variable B i interacció entre les dues. L'ANOVA ens indicarà quins efectes són significatius i quins no. Com passava en el disseny simple, la variància intra-grups és el terme de comparació per valorar quins efectes són significatius i quins no.

Alguns autors recomanen mantenir algunes precaucions a l'hora de valorar els diferents efectes factorials: S'hauria de començar en primer lloc veient si la interacció és o no significativa. Si ho és, cal centrar la interpretació dels resultats en la interacció, de forma que els efectes principals s'haurien de valorar amb precaució ja que, com hem vist abans, es poden veure modificats per l'existència d'interacció. Si la interacció no és significativa, es pot procedir a la interpretació dels efectes principals sense major problema, tot i que alguns autors recomanen utilitzar com a terme de contrast (per al càlcul de la F de Snedecor) no només la variància intra-grups, com fem habitualment, sinó l'agregació d'aquesta variància intra-grups amb la variància corresponent a la interacció, ja que aquesta, si no és significativa, es pot considerar també com a producte de l'atzar.

El càlcul dels diferents elements de l'anàlisi de la variància és en part semblant al que ja s'ha vist en el cas simple, considerant però que en el cas factorial tenim més efectes que cal estudiar. Suposem que tenim un disseny factorial 2 x 2 de grups aleatoritzats, amb 5 subjectes per grup. Obtenim els resultats corresponents i calculem el primer lloc els totals de cada grup i de les files i columnes, a més del total general:

	B ₁	B ₂	Totals
A ₁	13, 11, 13, 15, 13 Total: 65	15, 16, 17, 19, 14 Total: 81	146
A ₂	11,13,12, 9, 15 Total: 60	19, 19, 20, 17, 22 Total: 97	157
TOTALS	125	178	303

Calculem a continuació les sumes de quadrats per als diferents efectes:

$$\text{S.Q. Total} = 13^2 + 11^2 + 13^2 + \dots + 17^2 + 22^2 - 303^2/20 = 224,55$$

$$\text{S.Q. A} = 146^2/10 + 157^2/10 - 303^2/20 = 6,05$$

$$\text{S.Q. B} = 125^2/10 + 178^2/10 - 303^2/20 = 140,45$$

$$\text{S.Q. AxB} = 65^2/5 + 81^2/5 + 60^2/5 + 97^2/5 - 303^2/10 - 6,05 - 140,45 = 22,05$$

$$\text{S.Q. Intra-grups} = 224,55 - 6,05 - 140,45 - 22,05 = 56,00$$

I a partir d'aquí construïm el quadre resum de l'anàlisi de la variància:

Quadre-resum de l'anàlisi de la variància					
Fonts variació	S.Q.	g.ll.	Var.	F	p
Variable A	6,05	1	6,05	1,73	0,207
Variable B	140,45	1	140,45	40,13	0,000
Interacció AxB	22,05	1	22,05	6,30	0,023
Error	56,00	16	3,50		
TOTAL	224,55	19			

Naturalment, és possible generalitzar aquest procediment d’anàlisi al cas on tinguem més de dues variables independents. Si treballem, per exemple, amb un disseny amb tres variables independents (A, B i C), que tenen a, b i c valors, respectivament, el quadre resum de l’anàlisi de la variància prendria la forma següent:

DGA Factorial (3 V.I.)	Quadre-resum de l’anàlisi de la variància				
Fonts variació	S. Quadrats	G. llibertat	Variàncies	F	p
Variable A		a – 1		Var A / Var intra	
Variable B		b – 1		Var B / Var intra	
Variable C		c . 1		Var Ax B / Var intra	
Interacció A x B		(a-1)(b-1)			
Interacció A x C		(a-1)(c-1)			
Interacció B x C		(b-1)(c-1)			
Interacció A x B x C		(a-1)(b-1)(c-1)			
Intra-grups		N – k			
TOTAL		N - 1			
a = número de valors de la variable A b = número de valors de la variable B c = número de valors de la variable C k = número total de condicions experimentals (a x b x c) N = número total de subjectes o casos					

Com passava en el cas de 2 V.I., si treballem amb tres variables independents podrem estudiar l’efecte principal de cadascuna d’elles i també les seves interaccions. Naturalment, en aquest cas tenim un major número d’interaccions possibles. Podem estudiar la interacció de parells de variables (A x B, A x C, B x C), les anomenades interaccions de primer ordre, tal i com fèiem en el disseny amb dues variables independents. Però podríem estudiar també una interacció més complexa, com és la que es podria produir entre les tres variables simultàniament (A x B, C), que anomenarem interacció de segon ordre. Normalment les interaccions de segon ordre són difícils d’interpretar i són de poc interès per a l’investigador, de forma que no entrarem en el seu comentari. En tot cas, cal dir que en dissenys més complexos podríem trobar interaccions d’ordre superior (per exemple, A x B x C x D), que haurien de formar part de l’anàlisi de dades però que pràcticament mai ofereixen elements interpretatius d’interès.

Per fer una valoració correcta dels diversos efectes a un disseny factorial cal tenir en compte també si el disseny que es planteja és balancejat o no. Un disseny factorial és balancejat si el número de casos inclosos a cada grup és igual al de la resta de grups o bé hi manté una proporcionalitat similar:

Cas balancejat: Número de casos igual

	B ₁	B ₂	
A ₁	n = 5	n = 5	n _{1.} = 10
A ₂	n = 5	n = 5	n _{2.} = 10
	n _{.1} = 10	n _{.2} = 10	N = 20

Cas balancejat: Número de casos proporcional

	B ₁	B ₂	
A ₁	n = 8	n = 8	n _{1.} = 16
A ₂	n = 5	n = 5	n _{2.} = 10
	n _{.1} = 13	n _{.2} = 13	N = 26

Cas **no** balancejat: Número de casos no proporcional

	B ₁	B ₂	
A ₁	n = 5	n = 9	n _{1.} = 14
A ₂	n = 8	n = 4	n _{2.} = 12
	n _{.1} = 13	n _{.2} = 13	N = 26

Es pot comprovar fàcilment si un disseny és balancejat o no mitjançant la regla següent: Si un disseny és balancejat, el producte del total de fila i del total de columna de qualsevol casella, dividit pel número total de casos, és igual al número de casos del grup:

$$n_{jk} = \frac{n_j n_k}{N}$$

Quan les grandàries dels grups són iguals: $10 \cdot 10 / 20 = 5$, per a tots els grups

Amb grandàries proporcionals: $16 \cdot 13 / 26 = 8$ (primer grup). Es pot comprovar que també es compleix per a la resta de grups.

Disseny no balancejat: $14 \cdot 13 / 26 \neq 8$. Tampoc es compleix per a la resta de grups.

Amb un exemple molt senzill es pot veure per quina raó un disseny no balancejat pot complicar la valoració d'alguns efectes. Per exemple, en el cas no balancejat que s'exposa, suposem que la variable A són dos sistemes d'aprenentatge d'idiomes, i que s'apliquen a dos idiomes diferents (variable B). Suposem també que l'idioma B₁ és més difícil que l'idioma B₂. Per tant, cal suposar que si valorem l'efecte principal de la variable B, qualsevol indicador d'aprenentatge que s'utilitzi (per exemple, nivell de vocabulari adquirit durant el primer curs) mostrarà millors resultats per a idioma 2 que no per a l'idioma 1. El problema es produeix si, com és lògic, volem valorar també l'efecte principal de la variable A, és a dir, la diferència global de resultats per als dos idiomes. Es pot produir un error d'interpretació, ja que el mètode 1 es veure afavorit per sobre del mètode 2. La raó és que la majoria d'estudiants que utilitzen el mètode 1 estudien l'idioma més fàcil, mentre que la majoria dels que utilitzen el mètode 2 estudien l'idioma més difícil. Per tant, la diferència entre A₁ i A₂ no reflecteix únicament el funcionament dels dos mètodes, sinó que es pot veure contaminada per la presència diferent dels dos idiomes. Caldria, per tant, tenir això en compte a l'hora de valorar els resultats d'aquest estudi. Cal dir també que això no té res a veure amb el fet de que les dues variables puguin interactuar o no; aquesta contaminació dels resultats pel que fa a l'efecte principal de A no és el resultat d'una possible interacció amb B (que pot existir o no), sinó de la distribució dels estudiants en els diferents grups.

2.- Reducció de l'error experimental i control de variables de confusió. Anàlisi de covariància. Disseny de grups homogenis (DGH)

Una vegada exposats, a l'apartat anterior, els fonaments dels dissenys de grups aleatoritzats, és important recordar i insistir en els problemes que hi estan associats. Aquests problemes són bàsicament dos:

. Com ja s'ha indicat, l'aleatorització és una tècnica probabilística. La seva eficàcia a l'hora d'aconseguir grups experimentals equivalents depèn sobretot de la grandària dels grups. En grups petits, pot existir un dubte raonable sobre l'eficàcia de la tècnica, existeix un risc important de que els grups formats no siguin equivalents, amb tots els problemes de possible confusió en la interpretació de resultats que això implica

. El segon problema és refereix a la variància de l'error o error experimental. Una de les fonts més importants d'error experimental són les diferències individuals entre els diferents subjectes participants. Al disseny de grups aleatoritzats ens limiten a assignar els subjectes aleatòriament als diferents grups, però en principi no sabem si existeixen diferències importants entre ells ni en quins aspectes poden ser més o menys diferents. Per això en principi no podem fer gaire per reduir la magnitud de l'error experimental, cosa que pot arribar a suposar un problema de manca de sensibilitat del disseny, com s'ha indicat anteriorment

Per donar resposta a aquests problemes hi ha algunes actuacions possibles:

. En primer lloc, si podem obtenir registres d'una variable de confusió important, fins i tot durant o després de la realització de l'experiment, encara tindriem l'opció de millorar l'anàlisi estadística de les dades, introduint aquesta informació a l'anàlisi mitjançant la tècnica anomenada *anàlisi de covariància*, sense modificar el disseny utilitzat.

. En segon lloc, si podem identificar alguna o algunes variables de confusió importants, i si tenim la possibilitat d'obtenir algun tipus de registre sobre aquesta variable abans de formar els grups experimentals, podem utilitzar aquesta informació per assegurar l'equivalència dels grups pel que fa a aquesta variable. Això ens portarà a un tipus de disseny diferent, el *disseny de grups homogenis*.

. Anàlisi de covariància

L'anàlisi de covariància (ACOVAR) es pot utilitzar quan identifiquem alguna possible variable de confusió de caràcter quantitatiu, sempre que puguem mesurar-la en algun moment del procés experimental. En aquestes condicions, és possible completar l'anàlisi de les dades incorporant el possible efecte d'aquesta variable de confusió. Si s'aplica l'anàlisi de covariància, la variable de confusió controlada mitjançant aquesta tècnica s'anomena variable concomitant o, de forma més precisa, *covariant*. Aquest és bon exemple del que podem anomenar tècnica de control indirecta, que no implica una actuació directa sobre la variable de confusió però sí l'ajust de les dades experimentals per eliminar el seu impacte sobre els resultats.

Imaginem, per exemple, que es volen provar tres condicions ambientals diferents per veure si influeixen en el número de tasques correctament executades en un temps limitat. Les tasques a realitzar són de tipus motor i necessiten d'un nivell d'atenció elevat. Suposem que utilitzem un disseny de grups aleatoritzats i que els resultats obtinguts són els següents (10 subjectes per grup):

Tractament 1	Tractament 2	Tractament 3
13	20	19
17	18	22
13	22	23
16	16	17
20	15	22
18	18	23
22	13	26
8	21	15
24	9	11
10	24	30

Si realitzem l'anàlisi de la variància per a aquestes dades, el resultat serà el següent:

	Sumes quadrats	g.ll.	Variàncies	F	p
Tractaments	115,27	2	57,63	2,25	0,125
Intra-grups	692,90	27	25,66		
Total	808,17	29			

Els resultats de l'anàlisi no són estadísticament significatius, de forma que no es pot afirmar que hi hagi diferència entre els resultats dels tres tractaments experimentals.

Ara bé, és conegut el fet que el rendiment en una tasca que implica atenció continuada es veu afectat per determinades característiques personals, i especialment pel nivell d'extraversió-introversió dels subjectes. Aquesta és una variable de potencial confusió. Si no s'ha pogut controlar aquesta variable prèviament, és possible aplicar una tècnica estadística indirecta, com és l'ACOVAR, si es poden obtenir registres de la variable estranya.

Suposem que l'investigador pot administrar als subjectes participants, uns dies després de l'experiment, una prova psicomètrica d'extraversió-introversió, puntuada de 0 a 10. Els resultats de l'estudi es poden completar, doncs, amb aquesta informació (entre parèntesi, els valors de la covariant):

Tractament 1	Tractament 2	Tractament 3
13 (6)	20 (9)	19 (5)
17 (4)	18 (7)	22 (7)
13 (5)	22 (6)	23 (4)
16 (6)	16 (7)	17 (4)
20 (8)	15 (5)	22 (5)
18 (6)	18 (6)	23 (6)
22 (6)	13 (5)	26 (9)
8 (5)	21 (6)	15 (4)
24 (6)	9 (5)	11 (5)
10 (6)	24 (8)	30 (6)

A partir d'aquí és possible aplicar l'ACOVAR, que ofereix els resultats següents:

ACOVAR	Sumes quadrats	g.ll.	Variàncies	F	p
Tractaments	167,53	2	83,76	4,26	0,025
Covariant	181,41	1	181,41	9,22	0,005
Intra-grups	511,49	26	19,67		

Com es pot observar, l'aïllament o la separació de l'efecte de la covariant ha modificat els resultats de l'anàlisi, sobretot pel que fa a una reducció de la variància de l'error. D'aquesta forma, l'efecte dels tractaments, que no semblava significatiu a l'anàlisi original, sí ho és després de fer l'ACOVAR. No obstant això, és també motiu de reflexió que l'impacte de la variable extraversió-introversió sigui fins i tot superior al de la variable independent, això relativitza la importància de les diferències entre les tres condicions ambientals, encara que finalment hagi estat significativa.

L'ACOVAR implica l'ampliació del model que s'ha vingut utilitzant normalment a l'anàlisi de la variància, amb la introducció d'un nou element (la covariant) de caràcter quantitatiu. Tenim, doncs, una barreja entre un factor que normalment és categòric o és tractat com a tal (la variable independent) i una variable quantitativa com és la covariant, de forma que el resultat és una barreja entre el model típic de l'ANOVA i un model de regressió lineal. El model resultant seria el següent (on X_{ij} són els valors de la covariant i β és el coeficient de regressió que s'aplica a aquest nou element):

$$Y_{ij} = \mu + \alpha_j + \beta X_{ij} + \varepsilon_i$$

No s'entrarà en els aspectes de mecànica i càlcul de la realització de l'ACOVAR, tot i que no són particularment complexos. Cal indicar, únicament, que la covariant porta associat un grau de llibertat, que es resta dels graus de llibertat del component d'error experimental. Als efectes d'aquest text és suficient amb comprendre de forma general la lògica de l'ACOVAR i la manera en que pot modificar l'anàlisi dels resultats obtinguts, tal i com s'ha vist a l'exemple exposat. Al final d'aquest apartat es farà una breu comparació entre les característiques de l'ACOVAR i dels dissenys de grups homogenis.

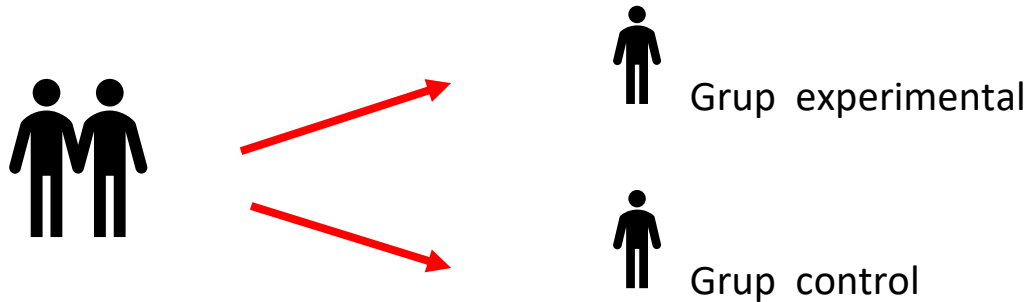
. Dissenys de grups homogenis (DGH)

. Conceptes generals i exemples

A diferència del que passava als dissenys de grups aleatoritzats, en els dissenys de grups homogenis l'assignació dels participants als grups experimentals no és completament aleatòria, sinó que es realitza de forma que s'igualin aquests grups en relació amb alguna variable o variables que potencialment poden influir en els resultats.

Hi ha diferents situacions on això es pot produir:

- a) En determinades ocasions (per exemple, en estudis neuropsicològics), l'investigador introdueix un conjunt de variables descriptives dels diferents subjectes experimentals, amb la idea d'igualar els diferents grups experimentals. Suposem, per exemple, que s'han de construir un grup experimental i un grup control, i que es considera que l'edat, nivell d'estudis i possibles antecedents familiars són variables que interessa controlar. Una pràctica força habitual en aquests casos consisteix en formar "parells" de subjectes amb característiques similars, i després assignar a l'atzar un membre de cada parell a un dels grups i l'altre membre a l'altre grup. Per exemple, si trobem dues persones amb una edat similar, amb un nivell d'estudis equivalent i sense antecedents neuropsicològics familiars, les considerarem "equivalents", i n'enviarem una al grup experimental i l'altra al grup control, amb un criteri aleatori.



Naturalment, si en lloc de treballar amb dos grups necessitem fer-ho amb tres, per exemple, llavors necessitarem trobar com a mínim tres persones amb característiques similars, i després les repartirem aleatòriament en els tres grups. Si trobem 6 persones similars, n'enviarem dues a cada grup, si en tenim 9, n'enviarem tres a cada grup, i així successivament. El més important, per assegurar l'eficàcia del disseny, és mantenir l'equitat en el número de persones similars que s'adjudiquen a cadascun dels grups experimentals.

b) Una altra situació és aquella en la qual l'emparellament dels subjectes fa a partir del rendiment en una prova prèvia. Per exemple, en un estudi en el qual la variable dependent sigui el temps de reacció ens poden preocupar les diferències individuals en la rapidesa de reacció de les diferents persones. Una forma d'evitar aquest problema és fer una prova prèvia (sense aplicació de cap tractament experimental), establir quin és el nivell de rapidesa que mostren les diferents persones, i emparellar-les en funció d'això.

Els dissenys en els quals es formen parells de subjectes amb característiques similars, tal i com s'ha mostrat als exemples anteriors, es coneixen de vegades com a *dissenys de grups aparellats* o *emparellats*, i no són sinó un exemple específic de disseny de grups homogenis. Tot i això, cal dir que res no impedeix utilitzar aquesta tècnica quan hem de formar més de dos grups experimentals, tot i que en aquest cas la denominació de "grups aparellats" pot portar a confusió (perquè ja no parlem de parells de subjectes, sinó d'agrupacions més grans, de 3, 4 o més persones), i per això molts autors prefereixen utilitzar la denominació de *dissenys de blocs aleatoritzats*, que com veurem s'empra també en altres casos.

c) Una altra variant de disseny de grups homogenis es produeix quan el que ens interessa controlar és una característica concreta, estable i identificada de les persones, mesurable mitjançant instruments que no tenen cap semblança amb la situació experimental. Per exemple, en molts estudis en psicologia cognitiva es considera que les diferències individuals en el nivell d'intel·ligència de les persones són una variable rellevant. El que es pot fer és administrar una escala psicomètrica d'intel·ligència, definir una sèrie de nivells (per exemple, CI de 80 a 90, de 90 a 100, etc.) i classificar cada persona al nivell que li correspongui. Una vegada fets els nivells, distribuïrem les persones de cadascun d'ells als diferents grups experimentals de forma equitativa. D'aquesta manera, els grups experimentals quedaran equilibrats pel que fa a aquesta possible variable de confusió.

Com es pot intuir a partir dels exemples anteriors, hi ha una certa confusió terminològica en els dissenys de grups homogenis, ja que autors diferents utilitzen terminologies també diverses. Com ja s'ha dit, un terme que s'utilitza sovint és el de "disseny de blocs aleatoritzats" (DBA), que de vegades es planteja com a sinònim de qualsevol disseny de grups homogenis, i en altres casos es reserva per a situacions concretes. Alguns autors reserven la denominació de DBA només per

al cas c), mentre que altres l'utilitzarien indistintament per a qualsevol dels tres casos plantejats. En molts textos, com ja s'ha indicat, es proposa l'ús del terme grups aparellats per al cas on només hi hagi dos grups, una idea que terminològicament té sentit però limita molt el concepte. En definitiva, és del tot probable trobar una àmplia diversitat terminològica als treballs teòrics dedicats a aquest tema, i també en els estudis empírics que utilitzen aquest tipus de dissenys.

Des d'un punt de vista pràctic, utilitzarem indistintament els termes *disseny de grups homogenis* i *disseny de blocs a l'atzar o de blocs aleatoritzats*. Caldrà tenir present, però, els diferents criteris que es poden utilitzar per formar aquests blocs, i que s'han pogut veure als exemples anteriors: Blocs basats en diverses característiques dels subjectes (exemple a), blocs basats en el rendiment en una prova experimental prèvia (exemple b) o blocs basats en una característica estable i identificada dels subjectes (exemple c). Això no exclou que es pugui utilitzar la denominació de *grups aparellats* quan es formen dos grups homogenis. Cal insistir en que els tres casos considerats són dissenys de grups homogenis, ja que tenen en comú el fet que els grups experimentals no es formen directament a l'atzar, sinó que prèviament es fa una classificació dels subjectes mitjançant algun criteri, i a partir d'aquí es formen posteriorment els grups de forma equitativa en funció del criteri de classificació. El resultat d'un DGH és la definició d'una sèrie de nivells o estrats de la variable concomitant, i per aquest motiu també es parla d'*estratificació* de la variable. I, per altra banda, també s'utilitza el terme *aleatorització restringida* per referir-se a la naturalesa dels DGH, pel fet que els subjectes no són assignats als grups experimentals de forma completament aleatòria, sinó partint del nivell, estrat o bloc al qual pertany. Per això els dissenys de blocs a l'atzar són denominats també *dissenys estratificats*.

Tot i que en els casos plantejats fins ara ens hem concentrat en el control de diferències individuals dels subjectes (precisió, edat, intel·ligència, etc.), també és possible utilitzar el disseny de grups homogenis per controlar variables externes al subjecte, pròpies del context o de la situació (per exemple, si l'experiment es fa al matí, a la tarda o al vespre, el lloc de realització de l'experiment, o altres variables relacionades amb la situació experimental que no s'hagin pogut deixar constants). No obstant això, aquesta no és una pràctica gaire habitual, i normalment els dissenys de grups homogenis es dirigeixen al control de variables pròpies del subjecte.

Cal dir també, com es pot veure als exemples anteriors, que la variable o variables concomitants que es volen controlar via homogeneïtzació poden ser tant qualitatives (per exemple, l'existència o no d'antecedents familiars) com quantitatives (per exemple l'edat). És important tenir en compte això, ja que quan la variable sigui quantitativa ens podríem plantejar com a alternativa l'ús d'una anàlisi de covariància. Aquest tema es discutirà breument al final d'aquest apartat.

A continuació es presenten alguns exemples per il·lustrar aspectes importants dels dissenys de grups homogenis.

. Suposem que volem portar a terme un estudi sobre l'impacte de tres condicions ambientals diferents en la precisió dels subjectes en una tasca motora, un cas similar al que hem vist al parlar de l'ACOVAR. Sospitem que poden existir diferències individuals importants entre les diferents persones pel que fa a la seva habilitat motora, i abans de començar l'estudi fem una prova prèvia de precisió, similar o igual a la tasca experimental posterior (però sense introduir les condicions ambientals que volem estudiar) amb puntuacions de 0 a 30. En aquest cas, doncs, la variable concomitant, és a dir, la variable que volem controlar mitjançant l'homogeneïtzació

dels grups, són les diferències individuals en precisió motora, mesurades mitjançant una prova prèvia amb una tasca semblant o idèntica a la tasca experimental. Imaginem que en funció dels resultats d'aquesta prova acabem definint cinc nivells de precisió, i que trobem tres persones que encaixen en cadascun dels nivells. Ja que hem de formar tres grups experimentals (hi ha tres condicions ambientals que volem provar), la distribució de subjectes podria ser:

Classificació dels subjectes en 5 nivells (N = 15)			
Resultats registre pre-tractament: 12, 12, 14, 15, 17, 17, 18, 19, 19, 21, 21, 23, 24, 26, 26			
	Límits	Subjectes	Mitjanes nivells
Nivell 1	12-14	(12)(12)(14)	(12,7)
Nivell 2	15-17	(15)(17)(17)	(16,3)
Nivell 3	18-20	(18)(19)(19)	(18,7)
Nivell 4	21-23	(21)(21)(23)	(21,7)
Nivell 5	24-26	(24)(26)(26)	(25,3)

Com es pot veure, no hem trobat tres subjectes “exactament iguals” (per exemple, tres subjectes amb un resultat de 12 punts a la prova prèvia) i, per tant, ens veiem obligats a treballar no amb nivells exactes sinó amb intervals (de 12 a 14, de 15 a 17, etc.). Lògicament això ens fa perdre una mica de precisió, ja que considerem com a equivalents a subjectes que no han obtingut exactament la mateixa puntuació.

En tot cas, el fet de disposar de tres subjectes a cada nivell ens permet adjudicar-ne aleatòriament un a cada grup experimental, per tant podem fer-ho de forma equitativa:

Formació de tres grups experimentals (N = 15, n = 1)			
	Tractament 1	Tractament 2	Tractament 3
Nivell 1	(14)	(12)	(12)
Nivell 2	(17)	(17)	(15)
Nivell 3	(18)	(19)	(19)
Nivell 4	(21)	(23)	(21)
Nivell 5	(26)	(26)	(24)
Mitjanes grups	(19,2)	(19,4)	(18,2)

Es pot observar que les mitjanes dels tres grups pel que fa als resultats de la prova prèvia són relativament similars, que és el que es volia aconseguir. Cal tenir clar que aquests encara no són els resultats experimentals, sinó els de la prova prèvia que ens ha permès classificar els subjectes i igualar els grups. A partir d'aquí s'haurien d'aplicar les tres condicions experimentals i obtenir els resultats corresponents per a la variable dependent.

Es pot veure també que, en aquest cas, per a cada “casella” del disseny (és a dir, per a cada combinació nivell-grup) hi ha un sol subjecte. Per exemple, per al tractament 1 hi ha una persona provinent del nivell 1, una persona provinent del nivell 2, i així successivament per a totes les caselles del disseny. Per això hem de dir que aquest és un cas de disseny de grups homogenis amb $n = 1$, on n és el número de subjectes per casella. Com veurem més endavant, això és important a l'hora d'analitzar els resultats experimentals. Cal notar també que en el disseny de

grups a l'atzar la lletra n representava el número de subjectes per grup experimental, mentre que en el disseny de grups homogenis representa el número de subjectes per casella.

Si en lloc de tenir solament tres subjectes per a cada nivell en poguéssim aconseguir més, per exemple 9 a cada nivell, la distribució resultant podria ser la següent:

Formació de tres grups experimentals (N = 45, n > 1)			
	Tractament 1	Tractament 2	Tractament 3
Nivell 1	(14)(13)(13)	(14)(12)(12)	(12)(12)(14)
Nivell 2	(16)(17)(17)	(15)(17)(17)	(16)(15)(17)
Nivell 3	(18)(18)(20)	(19)(19)(18)	(20)(19)(18)
Nivell 4	(22)(21)(21)	(21)(23)(23)	(22)(21)(22)
Nivell 5	(26)(26)(24)	(25)(26)(24)	(26)(24)(25)
Mitjanes grups	(19,07)	(19)	(18,87)

En aquest cas tenim tres subjectes a cada casella, per tant, és un cas $n > 1$ o, més exactament, $n=3$.

Altres criteris de classificació poden ser menys idonis. Si tornem al cas inicial, on teníem en total 15 subjectes, i en lloc de classificar-los en cinc nivells ho fem en tres, el resultat seria el següent:

Classificació dels subjectes en 3 nivells (N = 15)			
Resultats registre pre-tractament: 12, 12, 14, 15, 17, 17, 18, 19, 19, 21, 21, 23, 24, 26, 26			
	Límits	Subjectes	Mitjanes nivells
Nivell 1	12-17	(12)(12)(14)(15)(17)	(14,0)
Nivell 2	17-21	(17)(18)(19)(19)(21)	(18,8)
Nivell 3	21-26	(21)(23)(24)(26)(26)	(24,0)

És fàcil veure que en aquest cas tenim dos problemes importants:

- En primer lloc, hi ha una pèrdua de precisió força rellevant, ja que considerem com a equivalents subjectes que en realitat són bastant diferents (per exemple, al primer nivell considerem equivalent a una persona que té una puntuació de 12 i a una altra que té una puntuació de 17)
- En segon lloc, el número de subjectes per nivell és de 5, cosa que no permet fer una distribució equitativa entre els tres grups experimentals que necessitem formar

. Un segon exemple de disseny de grups homogenis seria el següent: Imaginem que volem comparar els resultats que s'obtenen amb tres tipus de problemes de raonament, suposadament de nivells de dificultat diferents. Al plantejar aquest estudi, podem pensar que el nivell d'intel·ligència dels subjectes és una variable rellevant que pot afectar els resultats, i podem tenir interès en controlar-la. El disseny de grups homogenis ens permet fer-ho, d'acord amb el procediment que ja s'ha vist: En primer lloc, classificarem els subjectes en funció del seu

nivell d'intel·ligència; posteriorment, formarem els grups experimentals de manera que quedin tan igualats com es pugui pel que fa a aquest criteri.

Novament podem trobar diferents situacions. Per exemple si, com a l'exemple anterior, disposem de 15 subjectes i els classifiquem en cinc nivells o blocs en funció del seu CI (coeficient d'intel·ligència), el resultat podria ser el següent:

Registres C.I.: 91, 94, 95, 97, 97, 100, 102, 102, 103, 106, 106, 107, 111, 112, 115		
Nivells C.I.	Subjectes	Mitjanes nivells
90 - 95	(91)(94)(95)	(93,3)
96 - 100	(97)(97)(100)	(98,0)
101 - 105	(102)(102)(103)	(102,3)
106-110	(106)(106)(107)	(106,3)
111 - 115	(111)(112)(115)	(112,7)

Suposem que necessitem formar tres grups experimentals, que volem que siguin aproximadament homogenis pel que fa al C.I. de les persones participants. Si repartim els subjectes de cada nivell a l'atzar entre els diferents grups, el resultat podria ser:

Formació de tres grups experimentals N = 15 n = 1			
	Tractament 1	Tractament 2	Tractament 3
Bloc 1 (90-95)	(94)	(95)	(91)
Bloc 2 (96-100)	(97)	(97)	(100)
Bloc 3 (101-105)	(103)	(102)	(102)
Bloc 4 (106-110)	(106)	(106)	(107)
Bloc 5 (111-115)	(112)	(111)	(115)
Mitjanes grups	(102,4)	(102,2)	(103,0)

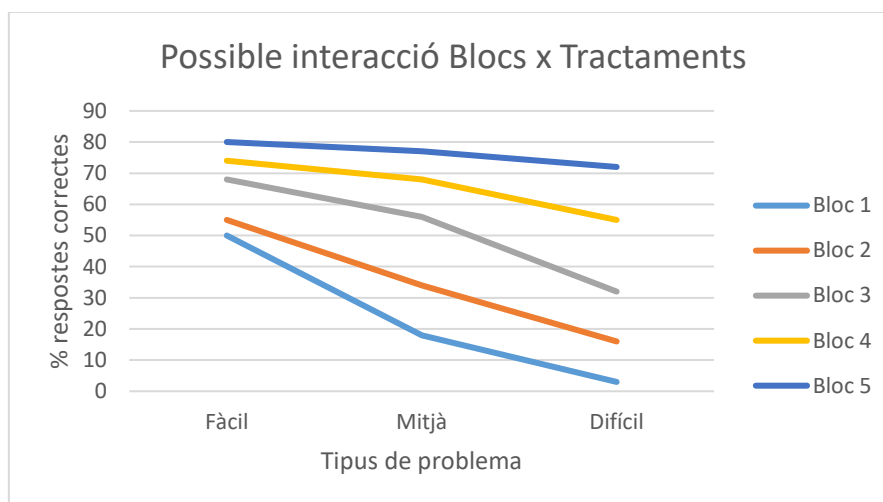
Clarament estem en un cas n = 1, ja que hi ha un únic subjecte a cadascuna de les caselles definides pels blocs i els grups experimentals.

Com es pot veure, utilitzem en aquest cas el terme "Blocs" per referir-nos als diferents nivells de la variable que s'ha utilitzat com a criteri d'homogeneïtzació, tot i que també es podrien anomenar "Nivells", com s'ha fet en casos anteriors. Fem això per il·lustrar la diversitat terminològica habitual en aquest tipus de dissenys.

Novament, si disposem de més subjectes podrem incrementar el número de casos per casella. Per exemple, si podem trobar 12 persones de cada nivell d'intel·ligència, la distribució podria ser:

Formació de tres grups experimentals N = 60 n = 4			
	Tractament 1	Tractament 2	Tractament 3
Bloc 1 (90-95)	(94)(95)(91)(94)	(95)(90)(92)(93)	(91)(94)(94)(90)
Bloc 2 (96-100)	(97)(99)(100)(98)	(97)(96)(99)(99)	(100)(97)(96)(98)
Bloc 3 (101-105)	(103)(103)(102)(104)	(102)(101)(105)(103)	(102)(104)(101)(103)
Bloc 4 (106-110)	(107)(106)(108)(109)	(107)(110)(109)(106)	(106)(106)(108)(109)
Bloc 5 (111-115)	(112)(113)(115)(112)	(111)(114)(114)(113)	(115)(111)(112)(114)

Hi ha una qüestió que cal tenir present en els dissenys de grups homogenis, i és la possibilitat que es produeixi algun tipus d'interacció entre la variable independent i el criteri d'homogeneïtzació. Per exemple, en el cas que s'acaba de citar, podria existir la possibilitat de que l'impacte dels diferents nivells de dificultat dels problemes de raonament no fos exactament el mateix per als diversos nivells d'intel·ligència. Suposem que els resultats experimentals obtinguts són els que es representen al gràfic següent (la variable dependent és el percentatge de problemes resolts correctament):



Això ens mostra que l'impacte de la dificultat del problema (eix d'abscisses) no és idèntic per a cadascun dels nivells d'intel·ligència que hem definit. En tots els casos, a mesura que creix la dificultat del problema el rendiment és inferior, però aquest efecte sembla molt més clar a mesura que ens movem amb nivells d'intel·ligència menors. Aquesta és una situació problemàtica, com a mínim per dues raons:

- En primer lloc, ens indica que la variable Nivell d'intel·ligència és més important del que podíem pensar inicialment. Cal tenir present que l'hem introduïda en el disseny simplement com una variable que volíem controlar, sense donar-li cap rellevància suplementària, igual que en altres exemples hem volgut controlar altres coses. Ara bé, els resultats ens indiquen que no solament el C.I. fa que el rendiment de les persones en aquesta tasca sigui diferent (cosa que ja esperàvem), sinó que també interactua amb la variable independent. Per tant, en estudis futurs probablement calgui replantejar-se l'estatus d'aquesta variable, no sembla que pugui ser una variable que ens limitem a controlar i puguem "oblidar" posteriorment, sinó que juga un paper crucial en aquest tipus d'estudis.

- En segon lloc, si es produeix una interacció entre la variable independent i el criteri d'homogeneïtzació, podem tenir certes dificultats en l'anàlisi estadística de les dades en cas que treballem amb un cas $n = 1$, com veurem al sub-apartat següent.

. Anàlisi de dades en dissenys de grups homogenis

La forma més habitual d'abordar l'anàlisi de dades en dissenys de grups homogenis és a partir d'una anàlisi de la variància que contempli tres elements: El possible efecte dels tractaments

experimentals, les diferències atribuïbles als nivells del criteri d'homogeneïtzació i, finalment, la variabilitat no explicada per cap dels altres dos factors. Ara bé, com ja s'ha indicat, existeix també la possibilitat d'existència d'interacció entre Tractaments i Nivells, fet que també s'hauria d'incorporar en el seu cas en el model.

Per tant, esquemàticament:

Model additiu (sense interacció Tractaments x Nivells):

$$Y_{ijk} = \mu + \alpha_k + \delta_j + \varepsilon_{ijk}$$

En aquest moment s'inclouen la mitjana global, el possible efecte del tractament (α), l'efecte del nivell del criteri d'homogeneïtzació (δ) i l'error experimental (ε).

Model no additiu:

$$Y_{ijk} = \mu + \alpha_k + \delta_j + (\alpha\delta)_{jk} + \varepsilon_{ijk}$$

En aquest segon model s'inclou la possible interacció entre Tractaments i Nivells.

En el cas $n > 1$ és possible estimar tots els efectes inclosos al model, ja que les dades disponibles permeten plantejar una anàlisi de la variància amb els components següents:

DGH simple Cas $n > 1$	Fonts variació	S.Q.	g.ll.	Var	F	p
	Tractaments Nivells Nivells x Tract. Intra-caselles		k-1 r-1 (k-1)(r-1) kr(n-1)		Var Tract / Var Intrac	
	TOTAL		N - 1			

k és el número de tractaments, r el número de nivells i n el número de subjectes per casella.

Per exemple, en el cas de tres grups experimentals, cinc nivells d'homogeneïtzació i tres subjectes per casella, els 44 graus de llibertat totals es reparteixen entre els 2 corresponents als tractaments, 4 per als nivells, 8 per al component interactiu i 30 per a la variància intra-caselles.

Es pot observar que en realitat hi ha dos components fonamentals: La variabilitat entre-caselles, explicada pels tractaments, els nivells i la seva possible interacció, i la variabilitat intra-caselles (diferències entre els resultats dins de cada casella), no explicable per cap dels factors indicats. Aquí cal introduir un raonament similar al que ja hem utilitzat en els DGA: El component intra-caselles respon a la presència de variables aleatòries que poden influir en els resultats i és, per tant, una aproximació al que anomenem *error experimental*. En canvi, la variància entre-caselles recull la presència de possibles efectes sistemàtics: l'efecte dels tractaments de la variable independent, la diferència de resultats entre els diversos nivells de la variable de confusió controlada, i la possible interacció entre les dues. Per aquest motiu, així com en el DGA la valoració dels diferents efectes es feia agafant com a terme de contrast la variància intra-grups, en un DGH l'element de comparació equivalent és la variància intra-caselles; per això la valoració

del possible efecte de la variable independent es fa mitjançant el quocient entre la variància corresponent a la VI i la variància intra-caselles, com s'indica al quadre.

Per la seva banda, la valoració del component Nivells ens indicarà si la variable de confusió que hem controlat afecta en major o menor mesura els resultats, i la significació (o no) del component interactiu ens dirà si el model al qual s'ajusten les dades és additiu o no.

En el cas $n = 1$ es planteja un problema important. En aquest cas el component de variabilitat intra-caselles no pot existir, ja que només hi ha un subjecte (i , per tant, un resultat) a cada casella. El resultat d'això és que en un disseny $n=1$ només es poden calcular tres components: L'efecte dels tractaments (diferències entre els grups), l'efecte dels nivells d'homogeneïtzació (diferència entre nivells) i un component residual, que recull tot allò no explicat pels altres dos.

Esquemàticament, l'anàlisi corresponent al cas $n = 1$ seria:

DGH simple Cas $n = 1$	Fonts variació	S.Q.	g.ll.	Var	F	p
	Tractaments Nivells Residual		$k-1$ $r-1$ $(k-1)(r-1)$		Var Tract / Var Resid	
	TOTAL		$N - 1$			

Si el model al qual s'ajusten les dades és additiu aquest no és un problema insoluble, ja que el component residual permet contrastar els efectes tant de tractaments com de nivells. Però si les dades s'ajusten a un model no additiu la situació és complicada, per una raó de concepte: Com s'ha indicat repetidament, per contrastar un efecte qualsevol necessitem com a terme de comparació una estimació de la variabilitat aleatòria o variància de l'error. En el cas dels dissenys de grups aleatoritzats, aquest element de contrast era la variància intra-grups, i en el disseny de grups homogenis amb $n > 1$ és la variància intra-caselles. En el cas del disseny de grups homogenis amb $n=1$ aquest element de contrast només pot ser la variància residual. Ara bé, si el model al qual s'ajusten les dades és no additiu, el problema és que la variància residual no contindrà només variabilitat aleatòria, sinó també l'impacte de la interacció entre tractaments i nivells. Aquesta interacció no és aleatòria, és un component sistemàtic i, per tant, "contamina" la variància residual i fa que no sigui un bon element de contrast.

Per tant, a un disseny de blocs amb $n = 1$:

Model additiu: Variància residual = Variància de l'error

Model no additiu: Variància residual = Variància de l'error + Interacció Tractaments x Blocs

Existeixen algunes formes d'esbrinar, ni que sigui de forma aproximada, si les dades s'ajusten o no a un model additiu, ja sigui mitjançant el gràfic d'interaccions convencional, o relacionant els valors pronosticats pel model amb els residuals corresponents. Si sospitem que existeix un component no additiu, i si no és possible millorar el disseny passant a un cas $n > 1$, cal assumir que hi ha el risc que la variància residual estigui "inflada" i la valoració de l'efecte dels tractaments i dels nivells estigui subestimada, de manera que es pot deixar de detectar algun efecte significatiu. Tot i que s'han plantejat alguns intents de solució a aquest problema, d'efectivitat bastant limitada (per exemple, la prova de no additivitat de Tukey, que sota

determinades condicions pot detectar l'existència d'un component no additiu), cal tenir clara l'existència d'aquesta dificultat. Si l'efecte dels tractaments resulta ser significatiu a l'ANOVA el problema és menor, ja que podem confiar en què, malgrat que el component residual pugui estar inflat, l'efecte de la variable independent és prou gran com per ser detectat a l'anàlisi. El problema principal es produeix quan el resultat de l'ANOVA per a la variable independent no és significatiu, ja que en aquest cas no tenim una interpretació segura, perquè això es pot deure a que realment la variable no té un impacte sobre els resultats, però també a que un residual "inflat" hagi emmascarat un efecte que realment existeix, i no ens permeti detectar-lo.

Cal indicar que els problemes indicats es produeixen quan el factor Blocs és d'efectes fixes, i que el seu plantejament si escollim els blocs de forma aleatòria podria ser diferent. No entrarem aquí en aquesta complicació, tot i que això serà rellevant al parlar dels dissenys de mesures repetides. Un cas en el qual es podria considerar el factor Blocs o Nivells com d'efectes aleatoris seria aquell en el qual fem un emparellament o agrupació a partir de diferents característiques (edat, procedència geogràfica, etc.). Si els subjectes s'han escollit a l'atzar dins de la població, els "nivells" generats a partir de les característiques dels subjectes es podrien considerar aleatoris.

Els casos que s'han vist fins ara pel que fa a dissenys de grups homogenis són dissenys de tipus simple, és a dir, amb una única variable independent. Naturalment, també és possible plantejar dissenys de grups homogenis factorials tot i que, com es veurà immediatament, això introdueix alguns elements de complicació.

Suposem, per exemple, que tenim dues variables independents, una amb tres valors i l'altra amb dos, i que classifiquem a les persones en quatre nivells pel que fa a una determinada variable de confusió que volem controlar. Esquemàticament, el disseny resultat és el següent:

	A ₁		A ₂		A ₃	
	B ₁	B ₂	B ₁	B ₂	B ₁	B ₂
Nivell 1						
Nivell 2						
Nivell 3						
Nivell 4						

Està clar que hem de configurar 6 grups experimentals. En funció del número de subjectes que tinguem a cada nivell de la variable de confusió podem estructurar un tipus de disseny o altre: si només tenim 6 subjectes a cada nivell, el disseny resultant serà $n = 1$; si tenim números superiors, podem pensar en un disseny $n > 1$ (per exemple, amb 12 subjectes per nivell podríem plantejar un disseny $n = 2$, amb 18 subjectes per nivell un disseny $n = 3$, etc.).

Sense entrar en detalls, és fàcil veure que el cas factorial inclou algunes complicacions afegides. En el cas $n = 1$, la possibilitat de que es produeixi alguna interacció entre la variable bloquejada i alguna de les variables independents (o amb les dues) s'incrementa, amb el consegüent problema d'anàlisi de dades que ja s'ha comentat; en el cas $n > 1$ es podria fer una anàlisi correcta de les dades utilitzant la variància intra-caselles com a terme de comparació, amb el petit inconvenient d'una certa multiplicació d'elements de l'anàlisi (interaccions entre les variables independents i la variable bloquejada). Esquemàticament:

DGH fact. Cas n = 1	Fonts variació	S.Q.	g.ll.	Var	F	p
	Variable A Variable B A x B Nivells Residual		a-1 b-1 (a-1)(b-1) r-1 (ab-1)(r-1)		Var A / Var Resid Var B / Var Resid Var Ax B / Var Resid	
	TOTAL		N - 1			

DGH fact. Cas n > 1	Fonts variació	S.Q.	g.ll.	Var	F	p
	Variable A Variable B A x B Nivells Nivells x A Nivells x B Nivells x A x B Intra-caselles		a-1 b-1 (a-1)(b-1) r-1 (a-1)(r-1) (b-1)(r-1) (a-1)(b-1)(r-1) abr(n-1)		Var A / Var Intra Var B / Var Intra Var Ax B / Var Intra	
	TOTAL		N - 1			

Naturalment, també seria possible plantejar dissenys de grups homogenis multivariats (en lloc d'univariats) o longitudinals (en lloc d'estàtics), amb les complicacions pertinents en l'estructuració del disseny i la posterior anàlisi de les dades.

També es poden plantejar dissenys de grups homogenis més complexos en cas que es vulgui bloquejar més d'una variable de confusió (doble o triple bloqueig); això donaria lloc a dissenys de quadrat llatí i similars, però el seu escàs ús a la pràctica no fa necessari tractar-los aquí.

Com s'ha indicat abans, és necessari fer una última consideració sobre els dissenys de grups homogenis, pel que fa a la seva comparació amb la tècnica de l'anàlisi de covariància, que s'ha estudiat anteriorment. Com s'ha indicat, l'ACOVAR ens permet fer un ajust dels resultats experimentals, de forma que podem aïllar o "eliminar" l'efecte d'una possible variable de confusió, quan aquesta variable és quantitativa. En aquest sentit, els seus objectius coincideixen amb els dels DGH, amb la diferència que en el cas del disseny de grups homogenis podem treballar tant amb variables quantitatives (establint intervals o estrats) com qualitatives.

Quan la variable concomitant és qualitativa o categòrica, l'ACOVAR no és aplicable i, per tant, cal plantejar un disseny de grups homogenis i fer l'anàlisi de dades corresponent via ANOVA, d'acord amb el que s'ha vist al llarg d'aquest apartat. Ara bé, quan la variable concomitant (o covariant) és quantitativa (com passava en els exemples de la precisió motora i de la intel·ligència), l'ACOVAR seria una alternativa raonable. Això ens porta al problema de quina de les dues vies d'acció (disseny de grups homogenis o ACOVAR) és preferible en aquesta situació.

La literatura experimental ens ofereix diferents perspectives d'aquest problema. Per una part, diferents estudis indiquen que l'ACOVAR és més potent que el DGH quan la correlació entre la

variable de confusió i la variable dependent és elevada (aproximadament per sobre de 0,6), mentre que amb correlacions inferiors el DGH sembla funcionar igual o millor. Això depèn també del grau de precisió amb què es classifiquen els subjectes en el DGH, és a dir, de l'amplitud dels intervals que es formen. Com ja s'ha indicat, és millor treballar amb intervals estrets (per exemple, en el cas del CI, de 90 a 95, de 96 a 100, etc.) que amb intervals amplis (de 90 a 100, de 101 a 110, etc.). Si els intervals amb els quals es treballa són amplis, és evident que el DGH perd precisió respecte de l'ACOVAR, ja que l'anàlisi de covariància conserva el valor quantitatiu exacte, mentre que en el DGH formem intervals i, per tant, considerem com a "iguals" persones o casos que no ho són totalment. Per tant, en aquesta situació l'ACOVAR aconseguiria una major reducció de l'error experimental. Ara bé, des d'una altra perspectiva, el DGH té la gran virtut de construir des del principi grups experimentals aproximadament equivalents per que fa a la variable concomitant (aquest és en realitat el seu gran objectiu, més enllà del tipus d'anàlisi de dades que es faci després). Això també té conseqüències sobre l'ACOVAR, ja que no és el mateix fer una anàlisi de covariància sobre la base de grups aleatoritzats (com en el DGA) que fer-lo sobre grups ja homogeneïtzats prèviament. Alguns autors proposen combinar les dues tècniques quan la variable concomitant sigui quantitativa: fer primer una homogeneïtzació establint una sèrie de nivells o blocs, i fer després una ACOVAR sobre els resultats. D'aquesta forma s'aconseguiria la màxima eficàcia de les dues tècniques.

3.- Dissenys intra-subjecte o de mesures repetides (DMR)

. Característiques generals, avantatges i inconvenients

Els dissenys de mesures repetides (anomenats també dissenys intra-subjecte) plantegen una resposta radical al problema d'aconseguir l'equivalència dels grups. Si volem grups tant semblants com sigui possible, la manera més potent i segura d'aconseguir-ho és treballar amb un **grup únic**, al qual aplicarem de forma successiva (amb intervals de temps que s'hauran de determinar) els diferents tractaments experimentals.

Cal indicar que el terme "mesures repetides" pot portar de vegades a alguna confusió, ja que també és aplicable a altres situacions. Per exemple, a un disseny longitudinal de grups aleatoritzats, cada subjecte passa solament per una condició experimental, però fem diferents registres dels seus resultats al llarg del temps, de manera que també podríem utilitzar el terme de "mesures repetides". El mateix podríem dir dels dissenys quasi-experimentals de sèries temporals, i d'altres situacions. En aquest text, el terme "mesures repetides" l'utilitzarem en referència als dissenys experimentals en els quals cada subjecte passa per les diferents condicions experimentals (motiu pel qual s'anomenen també dissenys intra-subjecte).

L'esquema corresponent al disseny intra-subjecte més senzill (disseny simple de mesures repetides, una sola variable independent amb k valors, aplicats a n subjectes) és el següent:

Disseny simple de mesures repetides amb k tractaments i n subjectes				
	Tractament 1	Tractament 2	Tractament k
Subjecte 1				
Subjecte 2				
Subjecte 3				
.....				
Subjecte n				

Òbviament, l'equivalència dels k "grups" és absoluta, ja que la seva composició és idèntica i, naturalment, cada persona "és igual a sí mateixa". Per aquest motiu es diu sovint que en els dissenys de mesures repetides s'aplica la tècnica del *subjecte com a control propi*.

Com veurem a l'apartat d'anàlisi de dades, l'estructura del DMR ens permet obtenir una reducció radical de l'error experimental. Als dissenys de grups homogenis aconseguíem una certa reducció de l'error al fer una anàlisi separada de la variable bloquejada. En el cas del DMR, l'anàlisi del factor *Subjectes*, com es veurà més endavant, ens permet extreure de l'error totes les diferències associades a les característiques individuals estables dels subjectes, i per tant la reducció de l'error experimental serà superior.

Es pot observar que el disseny de mesures repetides té una estructura similar (files i columnes) al d'un disseny de grups homogenis amb $n=1$, amb una diferència molt important: En el cas del disseny de grups homogenis cada fila representava un nivell del criteri d'homogeneïtzació, però dins de la fila hi havia diversos subjectes, en el cas $n=1$ un a cada grup; en canvi, en el disseny de mesures repetides cada fila es un subjecte únic, que passa per les diferents condicions experimentals. En el cas del disseny de mesures repetides, l'equivalència entre els "subjectes" d'una mateixa fila és absoluta, ja que en tractar-se del mateix subjecte totes les seves característiques individuals són idèntiques; en canvi, en el disseny de grups homogenis els subjectes d'una mateixa fila són iguals o semblants en alguna característica concreta (el criteri d'homogeneïtzació), però poden ser diferents en molts altres aspectes (personalitat, edat, capacitat de concentració, etc. etc.).

Una altra forma de veure la diferència entre DGH i DMR és pensar que, en el primer cas, es fa un control adreçat a variables o característiques específiques, que defineixen el criteri d'homogeneïtzació, mentre que en el segon el control és inespecífic, en el sentit que es controlen totes les característiques individuals dels subjectes, i no solament algunes.

Una altra diferència entre els DGH i els DMR és que en condicions normals els n subjectes que participen a l'experiment s'haurien d'escollir aleatòriament a partir de la població de referència, de forma que el "factor Subjectes" seria un cas de factor d'efectes aleatoris. En canvi, als dissenys de grups homogenis el factor Blocs en molts casos (no sempre) és d'efectes fixes, ja que és l'experimentador qui determina quins blocs o nivells vol utilitzar.

Com es veurà més endavant, la semblança estructural (encara que amb diferències metodològiques molt clares) entre els DGH amb $n=1$ i el DMR serà rellevant a l'hora de plantejar l'anàlisi de dades.

Tot i que els avantatges del disseny de mesures repetides són evidents, també són clars els problemes que planteja. El fet que els mateixos subjectes passin per totes les condicions pot produir efectes no desitjables, com són els d'aprenentatge i fatiga, que poden influir en els resultats i produir confusions. Aquests efectes seqüencials poden ser de dos tipus:

. Efectes residuals: Es produeixen en aplicar un tractament experimental quan l'efecte d'un tractament anterior encara no ha desaparegut totalment. Un cas molt típic és el dels estudis farmacològics, on cal esperar un temps prudencial entre l'aplicació d'un tractament i el següent.

. Efectes d'ordre: Tenen a veure amb la posició concreta d'un tractament dins de la seqüència experimental. Els casos més típics són els de l'aprenentatge, habituació a la situació o similars,

que afavoreixen als tractaments situats en els darrers llocs, i el de fatiga, on els últims tractaments són els més desafavorits.

En moltes situacions els efectes seqüencials són inevitables. Per exemple, suposem que es volen provar dos tipus de tècniques de suport en psicologia de l'esport, una d'elles basada en la visualització de situacions i l'altra en tècniques de concentració. Si s'utilitza un disseny de mesures repetides, això voldrà dir que tots els esportistes que participin en l'estudi hauran de passar per les dues condicions, és a dir, aprendre a utilitzar les dues tècniques. Ara bé, és evident que si un esportista aprèn a utilitzar tècniques de visualització, per exemple, no és realista suposar que "desaprengui" aquestes tècniques i deixi d'aplicar-les, si li són útils. Per tant, si després se li fa aprendre la tècnica de concentració, els resultats obtinguts reflectiran l'ús d'aquesta tècnica, però també el possible record i ús de la tècnica de visualització apresada anteriorment, i això és molt difícil d'evitar.

Un problema relacionat amb l'anterior és el del número de condicions experimentals del disseny. És fàcil veure que en dissenys complexos l'ús del disseny de mesures repetides presenta problemes. Suposem que es vol plantejar un disseny factorial $4 \times 2 \times 2$, i que es vol utilitzar la tècnica de mesures repetides. Això significaria que cada subjecte ha de passar per totes les condicions experimentals, és a dir, per un total de 16 condicions. Això és molt difícil a la pràctica, i a més és evident que l'ús d'un número de condicions tan gran multiplica els possibles problemes d'aprenentatge, habituació, fatiga, etc., és a dir, empitjora els problemes relacionats amb els efectes seqüencials. Com es veurà més endavant, existeixen algunes possibilitats d'actuació quan el número de condicions experimentals és elevat.

De forma resumida, les característiques bàsiques del disseny de mesures repetides són:

Idees bàsiques en relació als dissenys de mesures repetides o intra-subjecte

- . Subjecte com a control propi. Tècnica de control inespecífica
- . Equivalència total dels "grups" experimentals (grup únic)
- . Podrem obtenir una reducció radical de l'error experimental
- . Possibilitat d'efectes seqüencials: Efectes residuals i efectes d'ordre
- . Dificultat quan el número de condicions experimentals és gran

. Control dels efectes seqüencials

Com s'ha indicat anteriorment, els efectes seqüencials són probablement el problema metodològic fonamental dels dissenys de mesures repetides. Existeixen algunes coses que es poden fer al respecte.

Com s'ha indicat, en alguns cas pot ser útil separar en el temps l'aplicació dels diferents tractaments. Això pot funcionar en el cas de tractaments que produeixen efectes reversibles, de

manera que quan deixen d'aplicar-se el seu efecte desapareix en un temps determinat. Com s'ha assenyalat abans, els tractaments de tipus farmacològic són un exemple molt clar d'això.

Pel que fa als efectes d'ordre, una tècnica que pot ser útil és el *contrabalanceig*. El que pretén aquesta tècnica no és fer desaparèixer els efectes de fatiga o aprenentatge, sinó neutralitzar-los, de forma que l'impacte que aquests efectes puguin tenir sobre els diferents tractaments sigui equivalent.

El contrabalanceig consisteix simplement en modificar l'ordre d'aplicació dels diferents tractaments per als diferents subjectes. Si, per exemple, hem d'aplicar dos tractaments i treballem amb vuit subjectes, la solució òbvia es aplicar a quatre d'ells la seqüència Tractament 1-Tractament 2, i als altres quatre la seqüència Tractament 2-Tractament 1. Esquemàticament:

Subjectes	Seqüència
1	$T_1 - T_2$
2	$T_1 - T_2$
3	$T_1 - T_2$
4	$T_1 - T_2$
5	$T_2 - T_1$
6	$T_2 - T_1$
7	$T_2 - T_1$
8	$T_2 - T_1$

En aquesta disposició es pot veure que cadascun dels tractaments ocupa el mateix número de vegades cadascuna de les posicions. És a dir, el tractament 1 està la meitat de les vegades a la primera posició i l'altra meitat a la segona, i el tractament 2 exactament igual.

Això es pot aconseguir amb qualsevol número de tractaments, sempre que es disposi del número de subjectes necessari. En el cas de 2 tractaments, n'hi hauria prou amb dos subjectes, de manera que un passi la seqüència $T_1 - T_2$ i l'altre la seqüència $T_2 - T_1$. Lògicament dos subjectes són molt pocs, però qualsevol número parell de subjectes ens permetrà complir amb la condició, com s'ha fet a l'exemple amb 8 subjectes.

Si treballem amb tres tractaments, una organització possible és la següent:

Subjectes	Seqüència
1	$T_1 - T_2 - T_3$
2	$T_2 - T_3 - T_1$
3	$T_3 - T_1 - T_2$

Com es pot veure, cadascun dels tres tractaments està una vegada a cadascuna de les tres posicions possibles, per tant el disseny queda equilibrat. Quan es compleix aquesta condició

parlem d'un *contrabalanceig complet*. El número de subjectes mínim necessari per fer un contrabalanceig complet quan treballem amb k condicions és igual a k o qualsevol múltiple d'aquest número. Per exemple, si treballem amb 3 tractaments o condicions experimentals el número mínim de subjectes per fer un contrabalanceig complet és de 3, com en el cas que s'acaba de veure, però també ens serviria tenir-ne 6 (repetiríem dues vegades el mateix esquema), 9, 12, o qualsevol múltiple de 3.

Cal veure que si disposem d'un número de subjectes que no és igual a k o un dels seus múltiples llavors no és possible fer un contrabalanceig complet. Si, per exemple, $k=3$ i disposem de cinc subjectes, l'esquema resultant no és equilibrat:

Subjectes	Seqüència
1	$T_1 - T_2 - T_3$
2	$T_2 - T_3 - T_1$
3	$T_3 - T_1 - T_2$
4	$T_1 - T_2 - T_3$
5	$T_2 - T_3 - T_1$

Com es pot apreciar, en aquest cas els tractaments 1 i 2 estan dues vegades a la primera posició, mentre que el tractament 3 només hi és una vegada. A la segona posició hi ha dues vegades els tractaments 2 i 3, i només una vegada l'1, etc. La situació clarament no està equilibrada, i ens trobem doncs amb un *contrabalanceig incomplet*. En cas que el contrabalanceig complet no sigui possible, és preferible fer un contrabalanceig incomplet abans que no fer res, ja que sempre s'aconseguirà un cert equilibri dels efectes seqüencials, encara que no sigui perfecte.

Hi ha una variant especial del contrabalanceig complet, que anomenarem *contrabalanceig complet fort*. En aquest cas, no solament cada tractament ocupa les diferents posicions el mateix número de vegades, sinó que és precedit per cadascun dels altres tractaments el mateix número de vegades. Per exemple, si treballem amb tres tractaments, un contrabalanceig complet fort seria el següent:

Subjectes	Seqüència
1	$T_1 - T_2 - T_3$
2	$T_2 - T_3 - T_1$
3	$T_3 - T_1 - T_2$
4	$T_1 - T_3 - T_2$
5	$T_2 - T_1 - T_3$
6	$T_3 - T_2 - T_1$

Com podem veure, en aquest cas cadascun dels tractaments ocupa dues vegades cadascuna de les posicions possibles però, a més, cada tractament és precedit de forma immediata dues

vegades per cadascun dels altres. Per exemple, el tractament 2 es precedit dues vegades pel tractament 1 (subjectes 1 i 3), i dues vegades pel tractament 3 (subjectes 4 i 6), i el mateix passa amb totes les altres combinatòries.

Per aconseguir un contrabalanceig complet fort, si treballem amb k tractaments necessitarem com a mínim $k!$ subjectes, és a dir, el factorial de k , o bé un múltiple d'aquesta xifra. En el cas exposat, $3! = 3 \times 2 \times 1 = 6$, per tant necessitarem com a mínim 6 subjectes, però també podem treballar amb 12 (es repeteix dues vegades la seqüència), amb 18, i en general amb qualsevol múltiple de 6.

Cal indicar que normalment no és necessari utilitzar un contrabalanceig complet fort, ja que per neutralitzar els efectes d'ordre n'hi ha prou amb un contrabalanceig complet ordinari. La variant forta pot ser útil en casos molt particulars, per exemple quan sospitem que pot existir algun efecte residual que no s'ha dissipat amb prou rapidesa i no podem allargar el temps d'espera entre tractaments per assegurar la desaparició d'aquest efecte. En aquest cas el contrabalanceig fort ens pot ajudar, però tampoc garanteix una neutralització eficaç d'aquests possibles efectes residuals, ja que no sabem si l'efecte residual de cada tractament és el mateix o si els diferents tractaments s'influeixen mútuament sempre de la mateixa manera.

Cal indicar que en alguns casos el contrabalanceig, encara que sigui complet, no és ni l'única i ni tan sols la millor tècnica per intentar controlar els efectes seqüencials. Imaginem, per exemple, que treballem en un experiment sobre memòria i volem saber si la longitud de les llistes de paraules afecta el seu record posterior. Suposem també que volem utilitzar llistes de 3, 5, 7 i 9 paraules, i que volem fer 10 assajos amb cadascuna d'aquestes condicions. Es ben conegut que quan es treballa amb paraules es produeixen sovint intrusions, confusions entre llistes, i altres fenòmens que tenen a veure amb el fet de que els subjectes es vegin exposats a moltes paraules, que a més poden tenir relacions semàntiques entre elles. En aquest cas, si apliquem estrictament la tècnica de contrabalanceig, caldria que un subjecte determinat passés primer, per exemple, totes les llistes de longitud 3, després les de longitud 5, i així fins el final, mentre que altres subjectes farien el mateix però en ordre diferent. Però hi ha una possibilitat alternativa interessant, i és senzillament barrejar aleatòriament les diferents longituds, de manera que cada subjecte passi els 40 assajos totals en un ordre escollit a l'atzar (per exemple: 5, 9, 9, 3, 7, 3, 9, 5, 7... etc.). Cada subjecte passaria un ordre diferent. En aquest cas, no parlariem pròpiament d'un contrabalanceig sinó d'una *aleatorització* completa dels assajos, que ens permetria assolir els mateixos objectius de forma més fàcil i més eficaç. Naturalment, això ho podem fer en aquest cas perquè cada persona ha de passar diferents assajos sota cada condició, i això ens permet barrejar-los completament. Una vegada obtinguts els resultats, la pràctica habitual és reduir-los a un únic resultat per a cada condició. És a dir, si un subjecte passa 10 assajos amb llistes de longitud 5, el seu resultat final per a aquesta condició seria la mitjana de resultats dels 10 assajos, i així amb la resta de condicions.

. Tractament estadístic dels dissenys de mesures repetides

L'anàlisi estadística dels dissenys de mesures repetides es pot abordar des de diferents perspectives, considerant la naturalesa particular de les dades generades en aquest tipus de disseny.

Un aspecte molt important a considerar és el possible incompliment, en el cas del DMR, d'una de les condicions bàsiques per a l'ús de les proves estadístiques més habituals, i en particular de

l'anàlisi de la variància. Com ja s'ha indicat, una de les condicions per a l'aplicació correcta de l'ANOVA és la d'independència dels registres. Això significa que el resultat d'un subjecte que rep una determinada condició experimental no s'ha de veure afectat de cap manera pels resultats obtinguts en altres registres. Això és obvi, per exemple, en un disseny de grups aleatoritzats. Si un subjecte obté un resultat de 18,2, això no té per què tenir relació amb el resultat obtingut per qualsevol altre subjecte, que serà totalment independent. Ara bé, en el disseny de mesures repetides hi ha una dificultat en relació amb això, ja que els mateixos subjectes ens aporten diversos resultats, un sota cadascuna de les condicions experimentals.

Això és fàcil de veure si recuperem l'esquema general d'un disseny de mesures repetides i suposem que ja hem obtingut els resultats experimentals (el resultat d'un subjecte i quan rep el tractament j es representa com a Y_{ij}):

Disseny simple de mesures repetides amb k tractaments i n subjectes				
	Tractament 1	Tractament 2	Tractament k
Subjecte 1	Y_{11}	Y_{12}	Y_{1k}
Subjecte 2	Y_{21}	Y_{22}	Y_{2k}
Subjecte 3	Y_{31}	Y_{32}	Y_{3k}
.....
Subjecte n	Y_{n1}	Y_{n2}	Y_{nk}

Ha de quedar clar que els resultats vistos "en vertical", és a dir, els resultats dels diferents subjectes dins d'un mateix tractament, són independents entre sí. El resultat que obtingui una persona no té cap relació amb el resultat obtingut per una altra. Ara bé, això deixa de ser cert si mirem les dades "en horitzontal". Els resultats obtinguts per un mateix subjecte poden variar quan s'apliquen els diferents tractaments, però des del moment en que són generats per una mateixa persona, és evident que entren en joc les seves característiques individuals estables, que són les mateixes al llarg del temps, i poden generar patrons de resultats reconeixibles. Un subjecte concret pot tendir, per exemple, a obtenir puntuacions més altes que la resta en funció de les seves característiques individuals, i probablement això es mantindrà al llarg dels diferents tractaments. No es pot afirmar, doncs, que les diferents puntuacions obtingudes són completament independents. Les dades obtingudes sota les diferents condicions experimentals poden mostrar, per tant, una certa covariació.

En coherència amb el que s'acaba d'indicar, una informació fonamental en l'anàlisi de resultats en un DMR és la matriu de variàncies i covariàncies de les dades. Com és conegut, en aquesta matriu es representen a la diagonal principal les variàncies que trobem dins dels resultats per a cadascun dels tractaments, mentre que fora de la diagonal principal trobarem les covariàncies entre els resultats de les diferents condicions experimentals. Si tenim un DMR amb k tractaments (sigui k igual a 2, 3, 4 o el número que sigui), l'aspecte de la matriu de la variàncies-covariàncies serà el següent (naturalment, tant les variàncies com les covariàncies serien valors numèrics que caldria calcular):

$$\Sigma = \begin{pmatrix} \sigma^2_1 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1k} \\ & \sigma^2_2 & \sigma_{23} & \dots & \sigma_{2k} \\ & & \sigma^2_3 & \dots & \sigma_{3k} \\ & & & \dots & \cdot \\ & & & & \cdot \\ & & & & \sigma^2_k \end{pmatrix}$$

És important veure que si treballéssim amb un disseny de grups aleatoritzats, el que podríem esperar lògicament és que a la diagonal principal hi hagi les variàncies intra-grup per a cadascun dels grups experimentals, i com que les diferents puntuacions són independents, les covariàncies haurien de ser petites o tendir a 0. La condició d'homoscedasticitat, citada anteriorment, consistiria en el fet que les diferents variàncies haurien de ser similars, o no significativament diferents. En canvi, en un DMR la situació és diferent, ja que al tractar-se dels mateixos subjectes que passen per totes les condicions, les seves puntuacions podrien no ser independents i les covariàncies no tindrien per què ser properes a 0.

A partir d'aquesta consideració, hi ha tres formes bàsiques d'abordar l'anàlisi de dades en un disseny de mesures repetides:

. L'enfocament "clàssic", utilitzant l'anàlisi de la variància de manera similar a com es feia en els dissenys de grups homogenis. Per poder fer-ho correctament, cal que les diferents variàncies (diagonal principal) siguin similars (condició d'homoscedasticitat) i també que les diferents covariàncies siguin similars entre sí (homogeneïtat de variàncies i covariàncies). Per ser més precisos, tècnicament el que es necessita és que es compleixi una condició una mica menys restrictiva anomenada *esfericitat* de la matriu de variàncies-covariàncies, com es comentarà una mica més endavant

. Un enfocament que podem anomenar "multivariable", no perquè en un DMR hagi d'haver-hi per força més d'una variable dependent, sinó perquè s'utilitza una variant de l'anàlisi multivariable de la variància (MANOVA). L'ús d'aquest enfocament no necessita del compliment de la condició d'esfericitat

. Un enfocament basat en els anomenats models mixtos. Aquesta aproximació té diferents avantatges de tipus tècnic (que no seran tractats aquí), i es pot aplicar siguin quines siguin les característiques de la matriu de variàncies-covariàncies, però al mateix temps té algunes dificultats, sobretot una certa complexitat dels càlculs i, en general, el fet de ser força diferent dels procediments que són més familiars dins de l'àmbit de l'anàlisi de dades, incloent l'ús d'un sistema d'estimació que no és el més habitual (procediment de màxima versemblança)

Considerant el que s'acaba d'indicar, en aquest text només tractarem breument els dos primers enfocaments, i deixarem l'aproximació basada en el model mixt per a cursos més avançats.

Com s'ha assenyalat, els teòrics de l'estadística que han analitzat aquest problema han conclòs que l'anàlisi de la variància convencional és aplicable al disseny de mesures repetides sempre que es compleixi una condició necessària i suficient, coneguda com a *esfericitat de la matriu de variàncies i covariàncies*. Esquemàticament, aquesta condició s'expressaria de la forma següent:

Condicció d'esfericitat

$$\Sigma = \begin{pmatrix} \sigma^2_1 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1k} \\ & \sigma^2_2 & \sigma_{23} & \dots & \sigma_{2k} \\ & & \sigma^2_3 & \dots & \sigma_{3k} \\ & & & \dots & \cdot \\ & & & & \cdot \\ & & & & \sigma^2_k \end{pmatrix}$$

Es compleix la condició si:

$$C'^T \Sigma C^* = \delta I$$

C' : Matriu de contrastos ortonormal
 δ = Constant I = Matriu identitat

No s'entrarà en el detall de l'explicació d'aquesta condició, ja que tècnicament és complicada. N'hi ha prou amb tenir present que es refereix a certes condicions que ha de complir la matriu de variàncies-covariàncies quan es realitzen determinades operacions matricials (el producte d'aquesta matriu amb una matriu de coeficients ortonormals i la seva transposta ha de donar com a resultat una matriu diagonal). A partir d'aquesta definició teòrica s'han proposat algunes proves estadístiques que ens permeten valorar aquesta condició, la més coneguda de les quals és la prova de Mauchly. El que fa aquesta prova és comparar les variàncies de les diferències de resultats entre les diverses condicions experimentals; el resultat de la prova és un estadístic W, acompanyat del corresponent p-value. Si les variàncies de les diferències són similars podem considerar que la condició d'esfericitat es compleix raonablement. Per tant, si el resultat de la prova de Mauchly no és significatiu ($p > 0,05$), això voldrà dir que les variàncies de les diferències són semblants i, per tant, que es compleix la condició d'esfericitat.

En cas que es compleixi la condició d'esfericitat de la matriu de variàncies-covariàncies es pot utilitzar l'anàlisi de la variància convencional, que en el cas d'un disseny simple de mesures repetides prendria la forma següent:

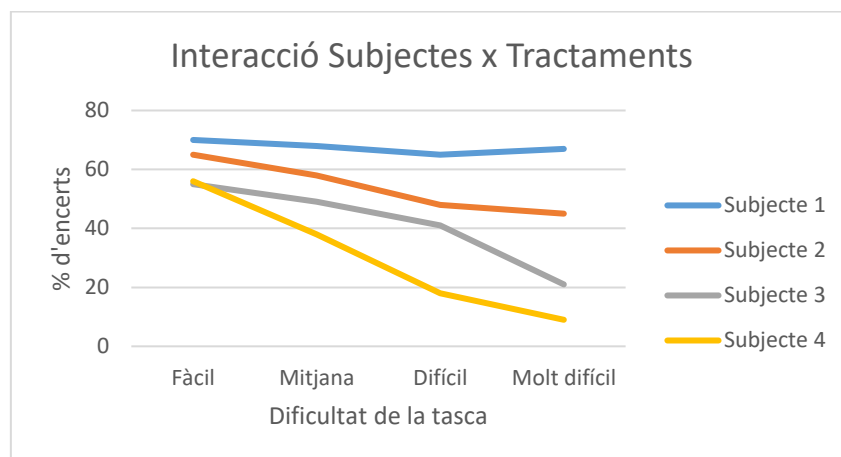
DMR simple	Fonts variació	S.Q.	g.ll.	Var	F	p
	Tractaments Subjectes Residual		k-1 n-1 (k-1)(n-1)		Var Tract / Var Resid	
	TOTAL		kn -1			

En cas d'incompliment de la condició d'esfericitat de la matriu de variàncies-covariàncies, no seria correcte utilitzar directament aquesta anàlisi. Diversos autors proposen fer una correcció

dels graus de llibertat utilitzats a l'ANOVA, i aporten diferents criteris per fer-ho. Probablement els més coneguts són el de Greenhouse-Geisser i el de Huyhn-Feldt. En tot cas, com ja s'ha indicat, en cas d'incompliment de la condició d'esfericitat tenim alternatives d'anàlisi que poden ser molt útils.

Es pot veure que l'anàlisi de la variància que es planteja és idèntica a la que es realitzava amb el disseny de grups homogenis amb $n = 1$. La diferència entre un disseny i l'altre és metodològica, com s'explicava al principi d'aquest apartat, però hi ha una semblança important des del punt de vista estadístic. Per tant, són aplicables aquí algunes de les coses que s'han indicat al parlar de l'anàlisi de dades en els DGH amb $n = 1$.

En particular, en els DGH es parlava del problema que suposava la possible existència d'interacció entre Nivells i Tractaments. Cal indicar que en un disseny de mesures repetides es pot produir també una interacció entre Subjectes i Tractaments. Això voldria dir que les diferències de resultats entre els diferents tractaments poden ser també diferents per als diversos subjectes experimentals. Per exemple, suposem que quatre subjectes diferents passen per quatre nivells de dificultat d'una tasca de concentració o vigilància, i es registra el percentatge d'encerts. Suposem que els resultats obtinguts es representen al gràfic següent:



Es pot veure fàcilment que no solament els diversos subjectes mostren resultats globals diferents, sinó que l'impacte de l'increment de la dificultat no és el mateix per a tots ells. Per exemple, el subjecte 1 mostra molt poca afectació per la dificultat de la tasca, mentre que els subjectes 3 i 4 mostren un impacte molt important. Un primer problema que planteja això és la necessitat d'explicar el perquè d'aquestes diferències tan marcades, que probablement tenen a veure amb característiques individuals que seria interessant identificar.

Hi ha relació entre l'existència o no d'interacció Subjectes x Tractaments i el compliment de les condicions d'aplicació de l'anàlisi de la variància. En efecte, quan no hi ha interacció entre subjectes i tractaments i, per tant, les dades s'ajusten a un model additiu, la condició d'esfericitat de la matriu de variàncies-covariàncies tendeix a complir-se; al contrari, quan hi ha interacció entre subjectes i tractaments, com en el cas del gràfic de dades anterior, es pot produir una desviació més o menys important respecte de la condició d'esfericitat de la matriu.

Al parlar dels dissenys de grups homogenis amb $n=1$ s'indicava un problema que tenia a veure amb la interacció entre Nivells i Tractaments la qual, en cas de produir-se, tenia com a

conseqüència un cert nivell d'“inflació” del terme residual, que inclouria també l'impacte d'aquesta interacció, amb el que això implica (veure apartat de DGH amb $n=1$). En el cas del disseny de mesures repetides aquest problema també existeix, però s'aborda d'una forma diferent, ja que el seu impacte depèn de diferents elements (especialment, de si els subjectes han estat escollits a l'atzar dins de la població i, per tant, es poden considerar com un factor d'efectes aleatoris). Tot i que hi ha una certa discussió sobre aquest punt, per simplificar podem entendre que la valoració de la condició d'esfericitat (que, com s'ha dit, es relaciona amb l'existència o no d'interacció Subjectes x Tractaments) és suficient per abordar correctament l'anàlisi de dades en un DMR. En cas de compliment de la condició d'esfericitat l'anàlisi de la variància convencional es pot considerar correcte, i en cas contrari tenim alternatives a la nostra disposició (correccions dels graus de llibertat, o ús de l'anàlisi multivariable de la variància).

Com s'ha indicat en apartats anteriors, l'anàlisi multivariable de la variància està pensat per ser utilitzat quan treballem amb dissenys experimentals multivariables, és a dir, quan registrem resultats de dues o més variables dependents, que eventualment poden mostrar relació entre sí. Un disseny de mesures repetides no és en sí mateix multivariable, ja que podem treballar perfectament amb una sola variable dependent, però té en comú amb els dissenys multivariables que per a cada subjecte obtenim diferents registres, i que aquests registres poden estar relacionats entre sí. Per aquest raó té sentit l'aplicació del MANOVA en un disseny de mesures repetides, encara que treballéssim amb una única variable dependent. Naturalment, també podria existir el cas més complicat d'un disseny de mesures repetides multivariable, que no serà abordat aquí.

Vista la complexitat dels càlculs implicats a la MANOVA per al disseny de mesures repetides, el seu desenvolupament no es fa en aquest document. Cal dir, però, que les consideracions que s'han fet abans per a la MANOVA al parlar de dissenys multivariables, i especialment els passos fonamentals per a la seva realització, són vàlids per al seu ús en els dissenys de mesures repetides.

L'avantatge fonamental de la MANOVA és que no es veu afectat pel compliment o no de la condició d'independències dels registres, ni tampoc per un possible incompliment de la condició d'esfericitat, de forma que el seu ús és sempre possible en un disseny de mesures repetides. Cal recordar que en el cas de la MANOVA partim de la base de que els registres dels diversos subjectes dins de cada variable dependent són independents, però no representa cap problema l'existència de covariacions entre les diferents VD. Traslladant aquesta lògica al disseny de mesures repetides, els registres obtinguts sota cada tractament per part dels diferents subjectes els podem considerar com a independents, però no hi ha dificultat amb el fet que els diferents resultats d'un mateix subjecte estiguin correlacionats entre sí, com pot passar en aquest tipus de disseny.

El desavantatge principal de la MANOVA és que habitualment el seu nivell de potència és inferior al de l'anàlisi de la variància convencional, de manera que existeix la possibilitat de no detectar algun efecte significatiu dels tractaments. Per aquest motiu una opció habitual és realitzar tant l'anàlisi de la variància convencional com la MANOVA, i valorar la coincidència o no dels resultats. Això no es refereix a que els estadístics i nivells de significació trobats siguin idèntics (no ho seran), sinó a que les dues proves coincideixin indicant que l'efecte dels tractaments és significatiu, o bé que no ho és.

Com s’ha dit abans, quan no es compleix la condició d’esfericitat s’han proposat altres possibles solucions en el marc de l’AVAR convencional (correcció dels graus de llibertat), però l’existència d’una alternativa com és la MANOVA (sense entrar en el tema més complicat del l’anàlisi basada en models mixtos) ens permet no aprofundir en aquestes solucions alternatives, que a més tenen també limitacions rellevants.

Resumint el que s’acaba d’indicar, les principals alternatives per a l’anàlisi de dades en un disseny de mesures repetides serien, esquemàticament, les següents:

Es compleix la condició d’esfericitat?	
SI	NO
<ul style="list-style-type: none"> . ANOVA . MANOVA (menys potent) . Enfocament basat en models lineals mixtos 	<ul style="list-style-type: none"> . ANOVA amb correccions (Greenhouse-Geisser, Huyhn-Feldt) . MANOVA . Enfocament basat en models lineals mixtos
<p>Forma fàcil de procedir: Fer tant ANOVA com MANOVA. Si els resultats coincideixen a l’hora d’indicar si l’efecte dels tractaments és o no significatiu, no hi ha problema. Si no coincideixen, considerar el resultat de l’ANOVA si es compleix la condició d’esfericitat i el de MANOVA si no es compleix</p>	

Cal dir, finalment, que les diferents tècniques que s’han comentat en aquest apartat no són vàlides únicament per al cas del disseny simple de mesures repetides, sinó que són aplicables també (amb un grau de complexitat superior) a altres dissenys de mesures repetides, com els que veurem als apartats següents.

. Dissenys factorials de mesures repetides o intra-subjecte

És perfectament possible plantejar un disseny de mesures repetides on es treballi amb més d’una variable independent, és a dir, un disseny factorial de mesures repetides. Com ja s’ha indicat abans, el problema principal d’aquest tipus de disseny té a veure amb el número de condicions experimentals. És fàcil veure que fins i tot un disseny factorial senzill implica l’increment del número de condicions experimentals. El disseny més senzill possible (factorial 2 x 2) treballa amb quatre condicions experimentals, el disseny 3 x 2 n’implica 6, i així successivament. Això és un problema, ja que suposa que els subjectes experimentals han de passar per moltes condicions. Les dificultats que deriven d’això tenen a veure sobretot amb la pèrdua de subjectes si han de participar en múltiples sessions experimentals (“mortalitat” experimental) i amb l’agreujament dels problemes relacionats amb els efectes seqüencials, especialment fatiga i aprenentatge.

L’esquema d’un disseny factorial de mesures repetides amb dues variables independents i n subjectes seria el següent (cas d’un disseny factorial 3 x 2):

Disseny factorial de mesures repetides amb 2 variables independents						
	A ₁		A ₂		A ₃	
	B ₁	B ₂	B ₁	B ₂	B ₁	B ₂
Subjecte 1						
Subjecte 2						
Subjecte 3						
.....						
Subjecte n						

En aquest cas hi ha una variable A amb tres valors i una variable B amb dos, de forma que es configuren 6 condicions experimentals. Els *n* subjectes participants passen per les 6 condicions, en l'ordre que s'estableixi.

Com passava en el cas simple, l'anàlisi de dades d'un disseny factorial de mesures repetides es pot abordar des de diferents perspectives: O bé mitjançant l'anàlisi de la variància tradicional, sempre que es compleixi la condició d'esfericitat, o bé a través de l'anàlisi multivariable de la variància. Hi ha, però, algunes complicacions suplementàries, ja que la condició d'esfericitat s'ha de valorar tant per a la matriu global de variàncies i covariàncies, que inclou totes les condicions experimentals, com per a cadascuna de les variables independents per separat. Per això en els dissenys factorials de mesures repetides es pot parlar d'*esfericitat global*, quan considerem la matriu de variàncies-covariàncies per al conjunt de les condicions experimentals, i d'*esfericitat local*, quan s'analitzen per separat les matrius de variàncies-covariàncies generades per a cadascuna de les variables independents.

Per aquesta raó, en aquests tipus de dissenys hi hauria dues maneres de realitzar l'anàlisi de la variància convencional, en funció de les característiques de les diferents matrius de variàncies-covariàncies.

En cas que es compleixi la condició d'**esfericitat global**:

DMR factorial	Fonts variació	S.Q.	g.ll.	Var	F	p
	Variable A		a-1		Var A / Var Resid	
	Variable B		b-1		Var B / Var Resid	
	A x B		(a-1)(b-1)		Var Ax B / Var Resid	
	Subjectes		n-1			
	Residual		(ab-1)(n-1)			
	TOTAL		abn -1			

En cas d'incompliment de la condició d'esfericitat global, però quan les matrius parcials per a cada variable independent sí que són esfèriques (**esfericitat local**):

DMR factorial	Fonts variació	S.Q.	g.ll.	Var	F	p
	Variable A		a-1		Var A / Sub x A	
	Variable B		b-1		Var B / Sub x B	
	A x B		(a-1)(b-1)		Var Ax B / Sub x A x B	
	Subjectes		n-1			
	Subjectes x A		(a-1)(n-1)			
	Subjectes x B		(b-1)(n-1)			
	Subjectes x A x B		(a-1)(b-1)(n-1)			
	TOTAL		abn - 1			

Com es pot observar, la diferència fonamental entre les dues formes de realitzar l'AVAR és que en el primer cas tots els efectes que es volen estudiar es contrasten amb una variància residual general, mentre que en el segon cada efecte factorial té el seu propi terme de contrast.

En cas d'incompliment de les condicions d'esfericitat tant global com local, disposem de les alternatives ja indicades al parlar del cas simple: Correcció dels graus de llibertat o bé, preferentment, càlcul de l'anàlisi multivariable de la variància. L'estratègia d'anàlisi pot ser similar a la indicada en el disseny simple, amb preferència per la MANOVA en cas d'incompliment de la condició d'esfericitat. Per la seva complexitat i per evitar un allargament excessiu del text no es presentarà el procediment a seguir per a la realització de la MANOVA en un cas factorial.

. Disseny factorial mixt o de mesures parcialment repetides (DFM)

El disseny factorial mixt és, com indica el seu nom, un cas particular en el qual es combinen dues estratègies: la de grups aleatoritzats i la de mesures repetides.

Suposem, per exemple, que necessitem realitzar un experiment amb dues variables independents: Per una banda (variable A), volem comparar el rendiment de les persones al realitzar dos tipus de problemes diferents (per exemple, problemes aritmètics i problemes espacials); per l'altra, ens interessa saber si el rendiment de les persones al resoldre aquests problemes canvia en funció de si treballen individualment, en parella o en grup (variable B). Tenim, per tant, un disseny 2 x 3, i si volem utilitzar una estratègia de mesures repetides, això significa que les persones que participin en aquest estudi hauran de passar per les 6 condicions experimentals definides pel disseny. Si considerem que això no és possible o convenient, el disseny factorial mixt ens ofereix una alternativa interessant. En el cas que s'exposa, la idea consistiria en que una de les variables es treballés com en un disseny de grups a l'atzar (per tant, formant grups diferents per a cada condició), mentre que l'altre es tractaria com de mesures repetides.

Imaginem que decidim plantejar la variable "Tipus de problema" com una variable de grups aleatoritzats. Si disposem, per exemple, de 12 subjectes experimentals, això significa que construirem dos grups aleatoritzats amb 6 persones cadascun. Per tant, cada persona treballarà únicament amb un dels tipus de problema (aritmètics o espacials). En canvi, l'altra variable la tractarem com de mesures repetides, de manera que tots els subjectes treballaran, en sessions

diferents, de les tres formes que s'han indicat: individualment, en parella o en grup. Tenim, doncs, una variable *entre-subjectes o entre-grups* (la variable A, tipus de problema) i una variable *intra-subjectes o de mesures repetides* (la variable B, forma de treball). Esquemàticament, l'organització que es planteja és:

		Individual	Parella	Grup
Grup Problemes aritmètics	Sub. 1 Sub. 2 Sub. 3 Sub. 4 Sub. 5 Sub. 6			
Grup Problemes espacials	Sub. 7 Sub. 8 Sub. 9 Sub. 10 Sub. 11 Sub. 12			

Aquest esquema ens indica, com ja s'ha explicat, que un subjecte determinat qualsevol pertany a un grup de la variable A (per tant, només treballa amb un tipus de problema) i, en canvi, passa per totes les condicions de la variable B (és a dir, en algunes sessions treballa sol, en altres en parella i en altres en grup). D'aquesta forma cada subjecte passa únicament per tres condicions experimentals, i no per les sis que hauria de passar en un disseny factorial de mesures repetides. Per exemple, el subjecte 11 passa per les condicions Problemes espacials-Treball individual, Problemes espacials-Treball en parella i Problemes espacials-Treball en grup.

Hi ha una altra situació en la qual un disseny mixt té sentit, i és quan alguna de les variables que interessen a l'investigador/a correspon a una característica del subjecte i, per tant, no pot ser ni manipulada ni tractada com de mesures repetides. Això correspondria a la situació que anteriorment s'ha anomenat com a *variable classificatòria*, és a dir, una variable de subjecte que ens interessa estudiar i que, en no poder ser manipulada, ens permet únicament dividir als subjectes en grups en funció dels diferents valors de la variable. Per exemple, suposem que en el cas anterior deixem de banda la variable tipus de situació ambiental i, en canvi, volem saber si les persones treballen millor en solitari, en parella o en grup, però en funció del seu gènere (homes o dones). En aquest cas tenim una variable A (gènere) que és de subjecte i, per tant, hem de definir necessàriament dos grups separats, un d'homes i un de dones. En canvi, la variable B pot ser tractada en forma de mesures repetides, com es feia a l'exemple anterior.

Suposem que es treballa amb 5 homes i 7 dones. Esquemàticament:

		Individual	Parella	Grup
Homes	Sub. 1 Sub. 2 Sub. 3 Sub. 4 Sub. 5			
Dones	Sub. 6 Sub. 7 Sub. 8 Sub. 9 Sub. 10 Sub. 11 Sub. 12			

Es pot veure que estructuralment els dos casos són semblants. Quan es treballa amb una variable classificatòria, de vegades el disseny mixt s’anomena disseny *split-plot*.

Com passava en els casos anteriors, l’anàlisi de dades del disseny mixt es pot abordar des de l’anàlisi de la variància convencional (univariable) o des de la perspectiva multivariable a través de la MANOVA, aplicable a la variable o variables de mesures repetides que hi hagi. Per simplificar s’exposarà únicament, de forma esquemàtica, la solució mitjançant l’anàlisi de la variància convencional.

El quadre-resum de l’anàlisi de la variància en un disseny factorial mixt amb una variable entre-subjectes (entre-grups) i una variable intra-subjectes (de mesures repetides) és el següent:

DFM	Fonts variació	S.Q.	g.ll.	Var	F	p
	Variable A (grups)		a-1		Var A / Intra-grups	
	Intra-grups		a(n-1)			
	Total entre-subjectes		an-1			
	Variable B (intrasubjecte)		b-1		Var B / Var Resid Var Ax B / Var Resid	
	A x B		(a-1)(b-1)			
	Residual		a(b-1)(n-1)			
	Total intra-subjectes		an(b-1)			
	TOTAL		abn - 1			

En tractar-se d’un disseny que combina diferents lògiques, l’anàlisi de la variància d’un disseny factorial mixt també és complicat. Bàsicament es pot dir que es divideix en dues parts: La que correspon a la variable o variables de grups, i la que correspon a la variable o variables de mesures repetides. En el cas exposat al quadre (amb una variable A de grups, i una variable B de mesures repetides) es pot observar que l’anàlisi té dues seccions diferenciades:

. Pel que fa a la variable A, l'anàlisi és similar al que es feia en un disseny de grups aleatoritzats.
Variació total = Variació entre-grups + Variació intra-grups

. En relació a la variable B (de mesures repetides) hi ha alguns elements particulars que cal indicar. Normalment, a l'ANOVA per a un disseny de mesures repetides l'anàlisi seria: Variació total = Variació entre-tractaments + Variació entre-subjectes + Variació residual. Cal dir, però, que cal incorporar a aquesta part de l'anàlisi qualsevol interacció on intervingui una variable de mesures repetides, com és el cas de la interacció A x B. Per altra banda, el factor "Subjectes" sembla haver desaparegut de l'anàlisi, tot i que realment no és així. De fet la variabilitat entre els subjectes està incorporada a un altre lloc, concretament a la variació Intra-grups de la primera part de l'anàlisi de la variància. Com ja s'ha indicat, la variació intra-grups indica les diferències individuals entre els diversos participants, i això és precisament el que es recull al factor Subjectes en un disseny de mesures repetides. Per tant, per no duplicar l'anàlisi d'aquesta font de variació, no s'inclou aquest factor explícitament a l'ANOVA del disseny mixt, tot i que, com ja s'ha dit, està incorporat a la variació intra-grups.

Naturalment, pel que fa a la part de mesures repetides de l'anàlisi, continuen en vigor totes les consideracions ja realitzades sobre la necessitat de compliment de la condició d'esfericitat de la matriu de variàncies-covariàncies, i és perfectament possible abordar l'anàlisi d'aquesta part mitjançant l'anàlisi multivariable de la variància. De fet, les principals aplicacions estadístiques ofereixen aquesta possibilitat, i permeten fer una anàlisi global dels resultats d'aquest tipus d'estudis, mantenint el tractament convencional de la part entre-grups i introduint el plantejament multivariable a la part de mesures repetides, tal i com s'ha vist en exemples anteriors.

. Aplicació de la lògica dels dissenys mixtos a estudis longitudinals

A l'exposar de forma general els diferents tipus de dissenys experimentals s'ha parlat de la possibilitat de treballar amb dissenys de caràcter longitudinal, és a dir, d'estudis en els quals el registre de la variable o variables dependents no es faci només en un moment temporal concret, sinó en diferents ocasions de registre al llarg del temps.

L'exemple que s'ha proposat a l'apartat corresponent es referia a un estudi en el qual es volia esbrinar si l'aplicació d'un programa de suport cognitiu a persones que presenten un procés progressiu de deteriorament permet evitar o com a mínim alentir aquest procés. Com a variable dependent s'utilitzava el nombre d'errors comesos en una prova de raonament lògic i, al tractar-se d'un procés que es desenvolupa progressivament, es feien 4 registres de la variable dependent (ocasions de registre), amb intervals de tres mesos.

La forma més normal d'abordar aquest estudi seria establir un grup experimental (que rebria el programa de suport cognitiu) i un grup control (que no el rebria) i comparar els resultats obtinguts en els dos grups pel que fa a la variable dependent establerta. Si els dos grups (experimental i control) es construeixen a l'atzar, estaríem davant del cas d'un disseny experimental simple (ja que només hi ha una variable independent), univariable (només hi ha una variable dependent), de grups aleatoritzats (els grups s'han format a l'atzar) i longitudinal (hi ha diferents registres de la variable dependent al llarg del temps).

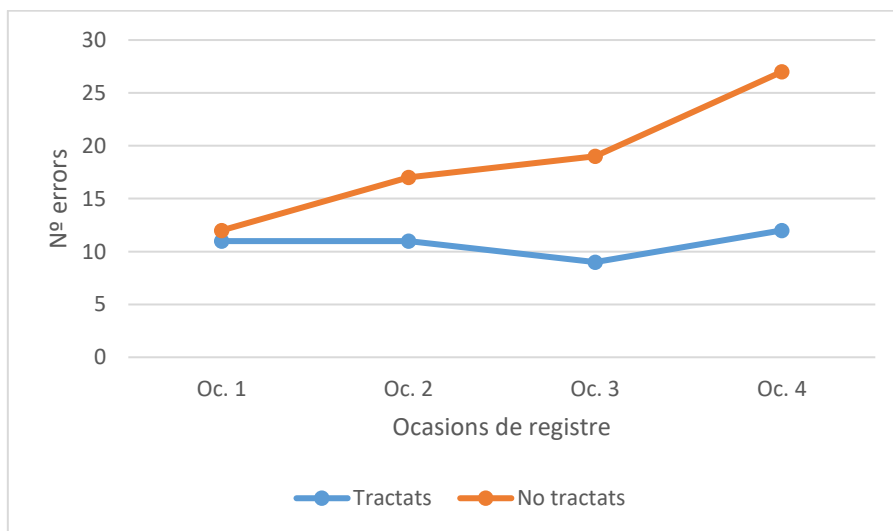
Tot i que es tracta d'un disseny de grups aleatoritzats, cal remarcar que els subjectes són mesurats en diferents ocasions al llarg del temps, de forma que s'introdueix també un

component de mesures repetides, encara que en un sentit una mica diferent del que hem estat veient. Cal insistir, doncs, en la diferència bàsica entre aquest cas longitudinal i un disseny de mesures repetides convencional (intra-subjecte). En un disseny de mesures repetides intra-subjecte, els subjectes passen per diferents condicions experimentals, i obtenim resultats corresponents a cadascuna d’elles; en canvi, en el disseny longitudinal que s’exposa, obtenim també diferents registres de cada subjecte, però *sota una única condició experimental*, aquesta condició experimental no canvia en el temps, i cada condició s’aplica a un grup diferent.

El fet que hi hagi diferents registres al llarg del temps ens mostra que el disseny longitudinal que es planteja té una estructura idèntica a la d’un disseny mixt. Si tenim, per exemple, 8 subjectes al grup experimental i uns altres 8 al grup control, esquemàticament el que fem és:

		Ocasíó 1	Ocasíó 2	Ocasíó 3	Ocasíó 4
Grup control	Sub. 1				
	Sub. 2				
	Sub. 3				
	Sub. 4				
	Sub. 5				
	Sub. 6				
	Sub. 7				
	Sub. 8				
Grup experimental	Sub. 9				
	Sub. 10				
	Sub. 11				
	Sub. 12				
	Sub. 13				
	Sub. 14				
	Sub. 15				
	Sub. 16				

Uns resultats hipotètics d’un estudi com aquest podrien tenir, a una representació gràfica, la forma següent:



Es pot observar que en el grup control el rendiment en la tasca proposada va empitjorant, probablement com a resultat d'un procés de deteriorament, mentre que en el grup experimental aquesta pèrdua progressiva no es produeix, cosa que fa pensar en un possible efecte positiu del programa de suport cognitiu.

Cal recordar, com ja s'havia dit anteriorment, que aquest estudi en concret plantejaria problemes ètics, ja que existeix un grup no tractat que eventualment podria continuar patint un procés de deteriorament. Aquest dilema ètic es produeix sovint en situacions similars.

Vista l'estructura del disseny, l'anàlisi dels resultats obtinguts es podria abordar com en el disseny mixt, ja sigui amb l'ANOVA o amb el MANOVA. Si utilitzem l'anàlisi de la variància convencional, el quadre corresponent seria (suposant que tenim a tractaments i p ocasions de registre):

Long	Fonts variació	S.Q.	g.ll.	Var	F	p
	Tractaments		$a-1$		Var A / Intra-grups	
	Intra-grups		$a(n-1)$			
	Total entre-subjectes		$an-1$			
	Ocasions		$p-1$		Var B / Var Resid Var AxB / Var Resid	
	Tract. x Ocasions		$(a-1)(p-1)$			
	Residual		$a(p-1)(n-1)$			
	Total intra-subjectes		$an(p-1)$			
	TOTAL		$apn - 1$			

Es pot observar que l'anàlisi és idèntica a la realitzada per al disseny factorial mixt, amb la diferència de que la variable convencional de mesures repetides és substituïda per la "variable" Ocasions de registre.

Cal fer alguna consideració suplementària sobre els dissenys experimentals longitudinals, sobretot per diferenciar-los dels dissenys de mesures repetides convencionals. És necessari fer com a mínim dues precisions:

. Es podria pensar que els dissenys de mesures repetides intra-subjecte són sempre longitudinals, ja que per a cada subjecte obtenim un registre sota cadascuna de les condicions experimentals, i per tant tenim tants registres per subjecte com condicions experimentals hi hagi. Cal indicar que en aquest cas no es tracta d'un disseny longitudinal, ja que per a cada condició s'obté només un registre, i per tant no hi ha possibilitat de valorar si els resultats mostren algun tipus de tendència, estabilització, nivell asimptòtic o cap característica semblant. Per fer això, caldria que sota cada condició obtinguéssim diversos registres de la variable dependent en moments temporals diferents, però sense modificar la condició experimental aplicada. Només en aquest cas el disseny de mesures repetides es podria considerar com a longitudinal.

. En moltes situacions es poden obtenir múltiples registres de la variable dependent per a cada subjecte sota una condició experimental determinada, sense que això vulgui dir que el disseny

sigui longitudinal. Com s'indicava a un exemple anterior, en un estudi de memòria podem administrar a cada subjecte 20 llistes de paraules sota una certa condició i registrar el nivell de record de cadascuna d'elles, però el que farem normalment és obtenir un resultat global mitjà per al conjunt dels assajos: si, per exemple, tenim 20 llistes de 10 paraules, podríem trobar que globalment el subjecte ha recordat correctament 150 ítems dels 200 presentats, cosa que significa un percentatge del 75%. Aquest serà el resultat final per a aquest subjecte, de forma que els múltiples registres realitzats es resumeixen en un únic resultat, sense cap mena de plantejament longitudinal.

Cal insistir en que només tenen caràcter longitudinal aquells registres fets deliberadament amb la finalitat de disposar d'un seguiment temporal del fenomen estudiat, i sense cap intenció de reduir-los a una sola dada mitjana ni cap altra operació similar. En aquest sentit, tant els dissenys de grups aleatoritzats, com els de grups homogenis com els intra-subjecte poden ser longitudinals, si l'investigador ho planteja així, però també poden no ser-ho, i de fet aquest és el cas més habitual.

III Dissenys quasi-experimentals

Com ja s'ha indicat al primer apartat d'aquest document, els dissenys quasi-experimentals comparteixen una part important de les característiques dels dissenys experimentals, en particular el fet de manipular les variables independents, però en canvi no poden complir amb tots els requisits necessaris per a ser considerats pròpiament com a experiments.

La característica més important que defineix un disseny quasi-experimental és que la formació dels grups experimentals no és controlada per l'investigador, ja sigui amb tècniques d'aleatorització, emparellament, etc, sinó que es treballa amb grups "naturals" o ja formats prèviament i, més en general, amb grups sobre la formació dels quals l'experimentador no té control. Hi ha diferents exemples de grups naturals: dues o més classes d'una escola, diferents hospitals, diverses seccions d'una empresa, etc. Normalment el fet de treballar amb grups naturals no és una elecció de l'investigador, sinó una necessitat derivada del tipus d'estudi que es vol realitzar. Si, per exemple, es vol comparar l'eficàcia de dos mètodes d'ensenyament diferents per a una determinada assignatura, normalment serà impossible formar dos grups aleatoritzats dins d'una escola determinada. El que es podrà fer és utilitzar un dels mètodes a una classe i l'altre mètode a una altra classe, i comparar els resultats. Més en general, considerarem normalment com a quasi-experimentals les situacions en les quals no es pugui realitzar una aleatorització dels grups i es faci una assignació no aleatòria per qualsevol raó. Per exemple, en algunes situacions terapèutics es permet, a la cerca de la major eficàcia, que els propis subjectes participants escullin entre dos o més tipus d'intervenció (cognitivo-conductual, farmacològica, etc.). En aquest cas tampoc és l'experimentador qui decideix el criteri de formació dels grups, i clarament no es tracta de grups aleatoritzats.

Naturalment, el problema fonamental d'utilitzar grups no aleatoritzats és que no podem garantir la seva equivalència abans d'aplicar els tractaments experimentals. Poden existir diferències entre els grups abans d'aplicar els tractaments (per exemple, una de les dues classes pot tenir un rendiment previ millor que l'altra), i això podria confondre els resultats obtinguts.

Per altra banda, és molt habitual que els dissenys quasi-experimentals s'utilitzin en situacions de camp o externes al laboratori (entorn laboral, educatiu, social, etc.), cosa que també dificulta la introducció de tècniques de control potents per evitar el possible efecte de variables de confusió.

Davant d'aquests problemes, el que fan els dissenys quasi-experimentals és intentar introduir alguns mecanismes complementaris de control, per intentar evitar com a mínim les dificultats més importants. Un exemple important és el dels registres pretest. Com que no estem segurs de l'equivalència inicial dels grups, el que podem fer és un registre de dades previ a l'aplicació dels tractaments, per tenir informació sobre si els grups són o no molt diferents entre sí. Aquestes possibles diferències inicials s'hauran de tenir en compte a l'hora de valorar els resultats experimentals. Per altra banda, l'obtenció de registres pre-test permet fer posteriorment comparacions entre resultats pre-test i postest, cosa que ens ofereix un font d'informació molt rellevant sobre el possible efecte dels tractaments experimentals.

El més habitual i aconsellable és obtenir els registres pre-test amb els mateixos instruments de mesura i els mateixos subjectes que els posteriors registres postest. Ara bé, en determinades situacions això no és possible o desitjable, i podem trobar dues variants importants:

. Mesures pretest substitutives: En aquest cas no s'utilitza el mateix instrument de registre per obtenir els resultats pretest i post-test. Per exemple, si es vol valorar si una determinada

intervenció produeix canvis en l'actitud dels subjectes cap a un determinat fenomen social (per exemple, el consum de drogues), es pot utilitzar una escala psicomètrica d'actituds per fer el registre de dades, abans i després de l'aplicació dels tractaments. Ara bé, és aconsellable no utilitzar exactament els mateixos ítems abans i després del tractament, per evitar possibles efectes re-test. En aquest cas és habitual utilitzar escales diferents (però relacionades), o emprar ítems diferents dins de la mateixa escala. Naturalment, és fonamental assegurar que els dos instruments de mesura utilitzats són tan comparables com sigui possible.

. Mostres separades: Els subjectes als quals es realitzen els registres pretest no són els mateixos que aquells amb els que s'obtenen els registres post-test. Normalment això no és deliberat, sinó el resultat de les característiques de l'estudi. Per exemple: Suposem que volem esbrinar si l'horari de funcionament d'un servei sanitari influeix en el nivell de satisfacció dels usuaris. Obtenim mesures de satisfacció (qüestionari d'opinió) amb l'horari habitual; posteriorment s'introdueix una modificació en l'horari, i s'obtenen registres de satisfacció després del canvi. És evident que no podem garantir que les persones que utilitzen el servei després del canvi són les mateixes que ho feien abans (probablement algunes sí, però altres no). Per tant, no són la mateixa "mostra" de subjectes, és un cas de mostres separades.

En resum, les característiques principals dels dissenys quasi-experimentals serien:

- Manipulació de les variables independents
- Treball amb grups "naturals" o, més en general, amb grups no aleatoritzats
- Ús freqüent en situacions poc controlades (estudis de camp)
- Substitució del control experimental més potent per altres elements de control més dèbils

Els dissenys quasi-experimentals es poden classificar a partir de diferents criteris, tot i que els més habituals són els següents:

1. En funció de la dimensió temporal, dissenys transversals vs. dissenys longitudinals o de sèries temporals. Com ja és conegut, en els dissenys transversals es fa un únic registre de la variable dependent en el temps (o, com a molt, un registre abans i un altre després de l'aplicació del tractament), mentre que en el cas longitudinal es fan diversos registres de la variable dependent en el temps, per valorar la seva evolució
2. En funció dels elements de control complementaris aplicats al disseny. D'acord amb això es pot pensar principalment en tres grups de dissenys quasi-experimentals: Disseny de comparació estàtica (treballaran sempre amb més d'un grup), dissenys amb variables dependents no equivalents (treballaran amb més d'una variable dependent), i dissenys que no reuneixen cap de les dues condicions anteriors, és a dir, dissenys amb un sol grup i una sola variable dependent. Aquest criteri no és perfecte, ja que alguns dissenys no encaixen exactament amb cap dels tipus indicats o poden combinar-ne més d'un

Els dos criteris són combinables, tot i que no de manera totalment perfecta. Esquemàticament podem considerar que els principals tipus de dissenys quasi-experimentals són els següents, d'acord amb la terminologia més habitual:

<p>Dissenys transversals:</p> <ul style="list-style-type: none"> . Disseny amb grup control no equivalent . Disseny de discontinuïtat a la regressió o de regressió discontinua . Disseny de variables dependents no equivalents . Dissenys amb retirada de tractament i de replicació múltiple
<p>Dissenys longitudinals o de sèries temporals:</p> <ul style="list-style-type: none"> . Disseny simple de sèries de temps . Disseny de sèries de temps amb grup control no equivalent . Disseny de sèries de temps amb variables dependents no equivalents . Dissenys de sèries de temps de regressió discontinua . Dissenys de sèries de temps amb retirada de tractament i de replicació múltiple . Dissenys de sèries de temps amb replicacions asincròniques

Es farà a continuació una revisió dels principals casos de disseny quasi-experimental, agafant com a guia la classificació indicada.

1-. Dissenys quasi-experimentals transversals

. Disseny amb grup control no equivalent

El disseny quasi-experimental més habitual (sigui en la seva versió transversal o longitudinal) és el que inclou un o més grups de control no equivalents. Aquest disseny es basa en l'existència d'un o més grups experimentals (depenent dels tractaments que es vulguin aplicar) i de com a mínim un (en algun cas poder ser dos o més) grup control que no rep tractament i serveix com a element de comparació. Tant en el grup o grups que reben tractament com en el grup control es realitzen registres pre-test, que serveixen per valorar si els diferents grups són més o menys equivalents i també com a element de comparació entre els resultats pretest i els postest, cosa que ens ajuda a valorar el possible efecte de la intervenció.

Esquemàticament, el disseny transversal amb grup control no equivalent més senzill seria el següent:

Esquema del disseny transversal amb grup control no equivalent				
Grups	Formació	Pretest	Condicions	Postest
Experimental	No aleatoritzada	R ₁	X ₁	R ₂
Control	No aleatoritzada	R ₃	X ₀	R ₄

Característiques principals:

- Existència d'un grup control que no rep tractament (X_0)
- No es pot garantir l'equivalència dels grups
- Existència d'un registre pretest i un registre posttest per als dos grups

La denominació de “grup control no equivalent” es refereix al fet que no podem garantir l'equivalència inicial entre els grups experimental i control, tot i que els registres pre-test ens orienten sobre el possible grau de les diferències entre un i l'altre.

Si volem aplicar més d'un tractament caldrà disposar de tants grups experimentals com intervencions previstes. També hi ha altres variants: per exemple, en alguns estudis el mateix tractament s'aplica a diferents grups, de forma que podem obtenir informació suplementària i veure si l'impacte d'aquest tractament és similar a cadascun dels grups. També es poden trobar casos en els quals s'utilitza més d'un grup control. Tot plegat ens aportarà més dades, que podem interpretar amb un major nivell de seguretat.

Pel que fa a l'anàlisi de dades, en el cas d'un disseny transversal amb grup control no equivalent es pot fer de manera semblant a la d'un disseny experimental, és a dir, realitzant una anàlisi de la variància de les puntuacions post-test. Ara bé, hi ha algunes alternatives interessants que utilitzen també la informació obtinguda amb els registres pretest. Esquemàticament, les principals alternatives per a l'anàlisi de dades en un disseny com aquest són:

Anàlisi de dades en un disseny amb grup control no equivalent	
Anàlisi de la variància de puntuacions posttest	Anàlisi similar al d'un disseny de grups aleatoritzats
Anàlisi de la variància de puntuacions de diferència	Per a cada subjecte: $D = R_{\text{post}} - R_{\text{pre}}$ Anàlisi de la variància de les puntuacions D
Anàlisi de la variància mixt (similar al d'un disseny factorial mixt)	Condicions com a variable entre-grups i pretest i posttest com a mesures repetides
Anàlisi de covariància	ACOVAR sobre les puntuacions posttest, introduint les puntuacions pre-test com a covariant

En el cas de les puntuacions de diferència, la idea consisteix en fer una anàlisi de la variància, però no directament amb les puntuacions post-test, sinó de les diferències entre la puntuació post-test i la puntuació pretest de cada subjecte. D'aquesta manera, el que es valora és el canvi produït entre pretest i post-test. Podríem trobar, per exemple, que el grup experimental ha canviat els seus resultat entre el pretest i el post-test, mentre que el grup control ha canviat poc o gens. Aquesta seria una evidència clara d'un possible impacte del tractament experimental. A més, d'aquesta forma solucionem en part el problema de les possibles diferències inicials entre els dos grups, ja que no fem una comparació directa entre les puntuacions post-test dels dos grups, sinó entre puntuacions de canvi, de manera que podem saber si un grup ha canviat més que l'altre. Cal tenir en compte, però, que en algunes situacions la magnitud del canvi que es pugui produir depèn del nivell inicial de cada subjecte, per tant si els subjectes dels dos (o més)

grups mostren nivells inicials molt diferents pel que fa a les puntuacions pretest, això pot influir en la seva capacitat posterior de canvi i confondre els resultats post-test.

Una opció interessant és utilitzar una anàlisi de la variància semblant al d'un disseny factorial mixt, partint de la base que les dues condicions (Experimental i Control) constitueixen una variable entre-grups (s'apliquen a grups diferents) i considerant les puntuacions obtingudes per a cada subjecte (pretest i posttest) com una "variable" de mesures repetides.

Una darrera alternativa és aplicar una anàlisi de covariància (ACOVAR) a les puntuacions post-test. Com s'ha vist en apartats anteriors, l'ACOVAR ens permet separar o aïllar l'efecte d'una possible variable de confusió, de manera que l'anàlisi dels resultats experimentals no es vegi afectada pel possible impacte d'aquesta variable. La particularitat en el cas del disseny de grup control no equivalent és que *la covariant (és a dir, el factor a "eliminar") són les puntuacions pretest*. És a dir, l'ACOVAR l'utilitzem per "eliminar" de les puntuacions post-test la possible influència de les puntuacions pretest. D'aquesta manera, el fet de que existeixin diferències en les puntuacions pretest entre els diferents subjectes i, sobretot, entre els dos grups, deixa de ser un problema, i podem comparar els resultats post-test dels dos grups amb major garantia.

. Disseny de discontinuïtat a la regressió o de regressió discontinua

Disseny de discontinuïtat a la regressió
<p>Característiques principals:</p> <ul style="list-style-type: none"> ➤ Els grups són formats a partir del rendiment obtingut a una prova pretest ➤ El grup de rendiment inferior rep un tractament diferencial, l'altre no ➤ Aplicació especial en recerca ATI (interacció Aptituds x Tractaments) ➤ Els grups no són "naturals", però tampoc aleatoritzats

Com s'ha indicat anteriorment, el disseny de discontinuïtat en la regressió és un cas particular, difícil de classificar, tot i que habitualment s'emmarca entre els dissenys quasi-experimentals. En el disseny de discontinuïtat, els grups experimentals no són formats a l'atzar, però tampoc són "naturals", sinó que són formats a partir dels resultats obtinguts en un pretest, i després són tractats diferencialment.

Exemples:

. Si es vol millorar el rendiment dels estudiants que mostren pitjors resultats dins d'un grup de classe, es poden intentar introduir mesures de reforç del seu aprenentatge. Per fer-ho, es formarien dos sub-grups: Un de baix rendiment i un d'alt rendiment. Els estudiants del grup de baix nivell reben l'actuació de reforç i els altres no. El resultat hauria de ser una millora del rendiment dels estudiants del grup tractat.

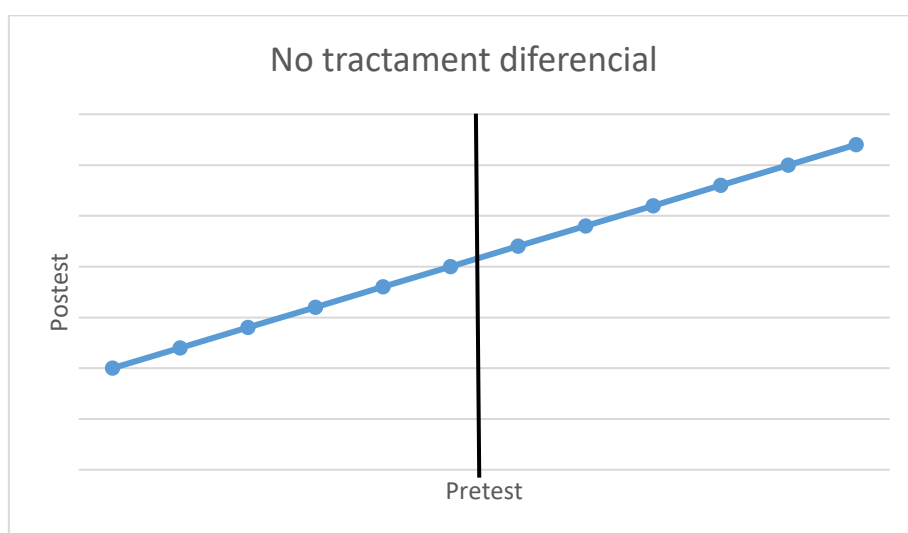
. Es pot fer una cosa semblant en altres contextos. Per exemple, en el cas d'atletes, distingint entre aquells que tenen marques pitjors i els que mostren resultats millors, i introduint posteriorment un sistema d'entrenament diferenciat per al grup de menor nivell.

És important no confondre aquest disseny amb altres vistos anteriorment. Per exemple, en el disseny experimental de grups homogenis també es classificava els subjectes en diferents nivells

a partir dels resultats d'una prova prèvia. Però en aquell cas els subjectes dels diferents nivells eren repartits equitativament entre els diversos grups experimentals, de forma que cada grup quedava format per persones de nivells diferents i, per tant, eren equivalents. En el cas del disseny de discontinuïtat en la regressió, els grups experimentals són deliberadament molt diferents entre sí, i són tractats també diferencialment precisament a partir d'això.

Tenint en compte la seva naturalesa i objectius, els dissenys de discontinuïtat a la regressió reben també altres denominacions, com són les de “disseny compensatori”, “disseny aptituds x tractaments”, etc., tot i que algunes d'aquestes denominacions poden ser equívokes o no correspondre sempre exactament a un disseny de discontinuïtat en la regressió.

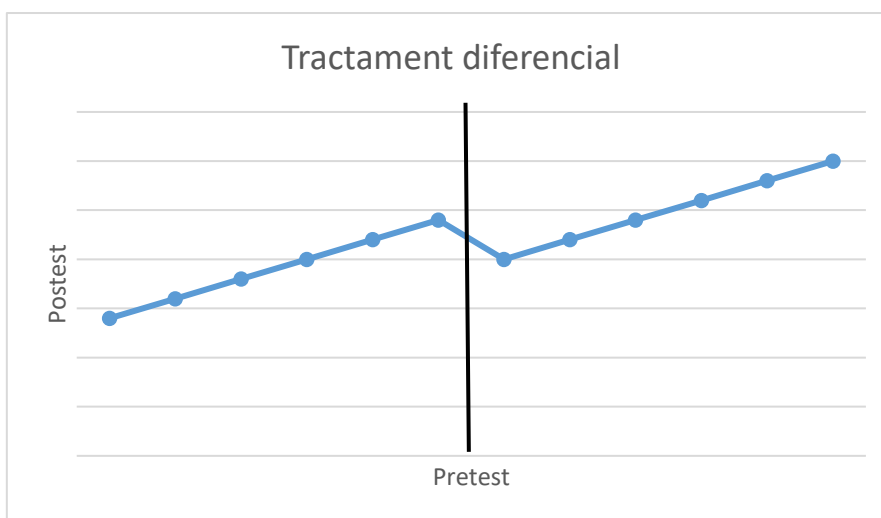
A més de l'aplicació de proves estadístiques, per valorar el possible impacte del tractament en un disseny de regressió discontinua és especialment útil aquest tipus de gràfic:



El grup situat a l'esquerra de la línia (format per 6 persones) és el de nivell pretest més baix, i el situat a la dreta (també amb 6 persones) és el de nivell pretest més alt. Cada punt de la línia representa els resultats d'un subjecte concret. Com es pot veure, per a cada subjecte participant es representen al gràfic la seva puntuació pretest (eix d'abscisses) i post-test (eix d'ordenades).

En aquest cas simplificat hem suposat que tots els subjectes mostren una millora entre pretest i posttest, sense que hi hagi diferències en els canvis produïts entre el grup de nivell més baix i el de nivell més alt. Si no s'aplica cap tractament, els resultats previsibles són els que es mostren al gràfic anterior: Els participants de baix nivell al registre pretest continuen mostrant pitjors resultats al post-test, tot i que òbviament podria haver-hi excepcions per evolució o millora de persones concretes. Igualment, els subjectes amb nivells alts al pretest els continuen mantenint al post-test.

Si el grup de baix rendiment ha rebut algun tipus d'intervenció “compensatòria”, mentre que el grup de major nivell no rep aquest tractament diferencial, un resultat possible seria aquest:



En aquest cas, es pot observar que els membres del grup tractat han millorat els seus resultats en major mesura que els del grup no tractat. Aquesta discontinuïtat o “trencament” de la línia és un indicatiu que ens fa pensar que el tractament diferencial ha estat eficaç, tot i que això s’haurà de valorar també a partir de les proves estadístiques pertinents, que poden ser similars a les indicades per al disseny de comparació estàtica. Probablement l’estratègia més habitual d’anàlisi de dades en aquest disseny és l’ús de l’anàlisi de covariància utilitzant els resultats pretest com a covariant.

. Disseny amb variables dependents no equivalents

En el disseny quasi-experimental de variables dependents no equivalents l’element principal de definició no és la presència d’un grup control, sinó l’existència d’una (o més) “variable dependent no equivalent”. Si treballem amb una variable dependent qualsevol, la variable dependent no equivalent serà, o bé una segona variable dependent diferent o bé la mateixa variable dependent original, mesurada en situacions diferents. La idea bàsica és que el tractament hauria de tenir un impacte sobre els resultats de la variable dependent original, però no sobre la variable no equivalent, i que això ens ofereix un element suplementari de control.

Esquemàticament:

Esquema del disseny transversal amb variables dependents no equivalents				
Grups	Formació	Pretest	Condicions	Postest
Experimental	No aleatoritzada	R ₁₍₁₎ R ₁₍₂₎	X (X)	R ₂₍₁₎ R ₂₍₂₎

. Els subíndexs (1) i (2) es refereixen als resultats obtinguts per a cadascuna de les dues variables dependents registrades
 . El símbol (X) es refereix al fet que, tot i que s’aplica el tractament X, en el cas de la segona variable dependent no esperem trobar un impacte del tractament

Exemple:

. A un institut de secundària s’introdueix un programa destinat a disminuir el consum de begudes refrescants ensucrades. Es fa un registre del consum d’aquest tipus de producte a la cafeteria

del centre abans i després de la introducció del programa. Ara bé, paral·lelament es registra també (abans i després del tractament) el consum d'un conjunt de productes diferents (suc de fruita, etc.). La idea és que, per aplicació del programa, el consum de begudes ensucrades pugui disminuir, mentre que la resta de consums es pot mantenir sense canvis o fins i tot incrementar-se. Naturalment, si tots els consums baixen (tant els "tractats" com els "no tractats") la interpretació hauria de ser diferent, i podria associar-se amb una disminució global del consum de begudes, produïda per qualsevol altra causa. Aquesta "variable dependent no equivalent" (el consum d'altres begudes) ens ofereix la possibilitat de contrastar molt millor el possible efecte del tractament i controlar l'efecte de possibles variables de confusió que puguin afectar els resultats.

Com es pot observar, la idea del disseny de variable dependent no equivalent és ajudar-nos a detectar el possible efecte específic del tractament, i distingir-lo de canvis que es podrien atribuir a altres causes diferents del propi tractament. En aquest sentit, la variable dependent no equivalent juga un paper semblant al del grup control en els dissenys que hem vist abans, és a dir, és un element bàsic de comparació.

En el disseny de variable dependent no equivalent és habitual treballar amb un sol grup, com és el cas de l'exemple exposat. Això no exclou la possibilitat d'introduir suplementàriament un grup control que no rebi tractament, cosa que reforçaria les possibilitats de detectar correctament el possible impacte del tractament experimental, però aquesta no és la situació més habitual. De fet, el disseny de variable dependent no equivalent ens pot ser útil precisament en casos en els quals no sigui possible disposar d'un grup control, i ens ofereix una alternativa interessant.

De forma resumida, les característiques dels dissenys transversals de variable dependent no equivalent serien:

Disseny transversal amb variables dependents no equivalents
<p>Característiques principals:</p> <ul style="list-style-type: none"> ➤ Normalment, un sol grup experimental. No és indispensable la presència d'un grup control, encara que es podria introduir ➤ Registre pretest i postest de dues (o més) variables dependents, o de la mateixa variable dependent mesurada en situacions diferents ➤ S'espera que el tractament produeixi canvis en una de les variables dependents, però no a l'altra o altres ➤ La variable dependent no equivalent juga un paper de comparació semblant al del grup control

. Dissenys de retirada de tractament i de replicació múltiple (o tractament repetit)

Els dissenys quasi-experimentals estudiats fins ara comptaven, o bé amb dos o més grups (disseny amb grup control no equivalent) o bé amb dues o més variables dependents (disseny de variables dependents no equivalents), en els dos casos amb la idea d'obtenir elements de comparació que ens permetin identificar el possible efecte de la intervenció experimental.

Hi ha un petit conjunt de dissenys quasi-experimentals transversals que no inclouen cap dels elements citats (ni grup control ni variable dependent no equivalent). En aquests casos,

l'element de control que s'introdueix és diferent, i consisteix en la comparació dels resultats obtinguts després de procedir a una introducció i retirada repetida del tractament.

Les variants més habituals d'aquest tipus de dissenys són dues: el disseny de retirada de tractament i el de tractament repetit o replicació múltiple, que no és més que una extensió o repetició de l'anterior. Esquemàticament:

Disseny de retirada de tractament	Disseny de tractament repetit
Exp R ₁ X R ₂ \bar{X} R ₃	Exp R ₁ X R ₂ \bar{X} R ₃ X R ₄ \bar{X} R ₅
Comparacions pretest-postest-retirada X: Aplicació del tractament \bar{X} : Retirada del tractament	Comparació dels canvis produïts a cada introducció i retirada del tractament

Com es pot veure, l'única diferència entre el disseny de retirada i el de tractament múltiple és que en el primer cas la seqüència és Pretest – Tractament – Post-test – Retirada, mentre que en el segon aquest procés s'amplia amb una nova introducció del tractament i, eventualment, la seqüència podria continuar amb una nova retirada i altres re-introduccions i retirades posteriors. Cada nova implantació del tractament suposa una replicació de l'estudi, i per això es parla de replicacions múltiples.

Aquests dissenys es basen en la suposició de que els efectes del tractament experimental puguin ser reversibles. A partir d'aquí, la lògica a utilitzar és senzilla: El que esperem és que cada vegada que s'implanti el tractament es produeixi un canvi similar en els resultats, i que cada vegada que es retiri, el resultat sigui una certa reversió cap a la situació d'origen. Si això passa repetidament, podem tenir una evidència de que és el tractament experimental, i no cap altre factor, el responsable dels canvis observats.

Alguns exemples senzills d'aquest tipus de dissenys serien l'estudi del possible impacte d'un canvi en l'horari laboral d'una empresa en la productivitat, o d'un canvi en la regulació semafòrica en una zona de la ciutat en la fluïdesa del trànsit, etc. Es pot veure fàcilment que en aquests casos la intervenció experimental es pot implantar i retirar, i que els seus efectes poden ser reversibles.

De forma resumida, per tant, les característiques d'aquest grup de dissenys serien les que es mostren a l'esquema següent:

Disseny transversal de retirada de tractament i tractament repetit

Característiques principals:

- Normalment, un sol grup experimental. No sol introduir-se grup control, ni tampoc variables dependents no equivalents, de forma que no es disposa d'aquests elements de comparació
- Aquests elements són substituïts per la retirada i re-introducció del tractament, amb els registres de resultats corresponents
- Cada re-introducció del tractament es pot considerar com una rèplica de l'estudi
- La retirada del tractament permet valorar la possible reversió dels canvis, i ofereix un altre element d'evidència

2-. Dissenys quasi-experimentals de sèries temporals

Com s'ha indicat en apartats anteriors, els dissenys longitudinals ens permeten, a diferència dels transversals, fer una valoració de l'evolució temporal del fenomen que estiguem estudiant. Això s'aconsegueix a partir de l'obtenció de registres repetits de les variables dependents dins de cadascuna de les situacions definides en el disseny. Això pot posar de manifest aspectes com l'estabilitat o no dels registres, la possible existència de tendències, la rapidesa d'impacte d'una intervenció o la persistència del seu efecte, etc.

En primer lloc es faran unes breus consideracions generals sobre els dissenys de caràcter longitudinal. Aquestes apreciacions són vàlides, naturalment, pels dissenys quasi-experimentals longitudinals o de sèries temporals, però ho són també per a qualsevol disseny longitudinal, sigui experimental, quasi-experimental o no experimental.

En general cal considerar que qualsevol plantejament longitudinal necessita com a mínim sèries amb tres registres de dades. Difícilment es pot considerar útil una sèrie només amb dos registres, ja que ens ofereix molt poca informació i ni tan sols fa possible detectar de forma fiable cap mena de tendència que puguin mostrar les dades.

En principi els diferents registres de resultats es fan sobre els mateixos subjectes, però això no sempre és possible. Si, per exemple, es registra cada setmana el temps mitjà de recuperació dels pacients que són donats d'alta d'un hospital, evidentment els subjectes no són els mateixos d'una setmana a l'altra, per tant la unitat global (o "grup"), és a dir, l'hospital, és el mateix, però els subjectes van canviant. O si es fan enquestes setmanals a les persones assistents a un determinat espectacle musical, està clar que aquestes persones poden variar d'una setmana a una altra.

Per altra banda, una decisió fonamental es refereix al nombre de registres i a l'interval que volem establir entre ells. El nombre de registres ha de ser suficient per detectar els aspectes d'evolució que volem estudiar (tendències, estabilització de l'efecte d'un tractament, etc.), cosa no sempre fàcil d'establir, ja que depèn de les característiques temporals del fenomen que estiguem estudiant. Per altra banda, els intervals entre registres han de ser els correctes per obtenir un nivell de "sensibilitat" suficient. Si, per exemple, es treballa amb casos de tabaquisme, sembla més raonable fer registres diaris, que ens donen informació força afinada, que no registres setmanals o mensuals. En canvi, en un registre d'actes delictius a una ciutat probablement n'hi haurà prou amb registres setmanals o mensuals, o fins i tot de menor periodicitat. Per la seva

banda, els registres de resultats acadèmics d'estudiants es solen fer habitualment de curs en curs, és a dir, amb periodicitat anual, etc.

És fàcil veure que el nombre de registres i el seu interval determinen directament la riquesa de la informació de la qual disposem. Un element interessant, si es disposa de dades suficients, és la possibilitat d'identificar cicles. Per exemple, en el cas del tabaquisme, el fet de disposar de registres diaris ens permetria veure si hi ha o no diferències consistents de consum entre la setmana laboral i el cap de setmana, i si aquestes diferències es repeteixen sistemàticament. Amb uns intervals de registre superiors, aquesta informació no es podria obtenir. Per altra banda, si un determinat fenomen mostra una tendència marcada, sigui creixent o decreixent, el fet de disposar d'un nombre de registres llarg en el temps ens pot indicar si aquest creixement (o decreixement) és constant, si es va accelerant o alentint, si té algun nivell asimptòtic, etc.

Se sol afirmar que quantes més ocasions de registre hi hagi, millor i més completa és la informació obtinguda. Això és cert en termes generals, però a l'hora de decidir quants registres es faran caldrà tenir molt presents consideracions pràctiques (limitacions de temps i recursos), ètiques (per exemple, no es pot allargar indefinidament una situació negativa o conflictiva per tal de tenir molts registres pretest) i metodològiques (l'obtenció de registres múltiples amb els mateixos subjectes pot produir efectes no desitjats, si s'allarga excessivament una sèrie la informació obtinguda pot ser redundant, etc.).

En el cas de dissenys quasi-experimentals longitudinals, i també en alguns dissenys experimentals, s'utilitza sovint l'expressió "dissenys de sèries temporals interrompudes". La idea de sèrie interrompuda es refereix al fet de que, en moltes ocasions, tindrem una sèrie de registres abans de la intervenció experimental i una altra sèrie després, amb la finalitat de comparar-les. Per això es pot dir que el conjunt de la sèrie temporal està "interrompuda" per la introducció del tractament, de forma que queda dividida en dues parts, la sèrie pretest o pretractament i la sèrie post-test.

Els tipus de dissenys quasi-experimentals longitudinals o de sèries temporals interrompudes utilitzats més habitualment són els següents:

Principals dissenys quasi-experimentals de sèries temporals

- Disseny simple de sèries de temps
- Disseny de sèries de temps amb grup control no equivalent
- Disseny de sèries de temps amb variables dependents no equivalents
- Disseny de sèries temporals amb retirada de tractament
- Disseny de sèries de temps amb replicacions múltiples
- Disseny de sèries de temps amb replicacions asincròniques

Es pot veure fàcilment que molts dels dissenys quasi-experimentals de sèries temporals són similars als seus homòlegs transversals (grup control no equivalent, variables dependents no equivalents, retirada de tractament i replicacions múltiples o tractament repetit), cosa que permetrà no repetir coses ja indicades anteriorment. Altres dissenys són propis només del plantejament longitudinal, com és el cas del disseny simple de sèries de temps i del disseny amb

replicacions asincròniques. Cal dir també que pot plantejar-se una versió longitudinal del disseny de regressió discontinua, tot i que no entrarem en les seves especificitats.

. Disseny simple de sèries de temps

Esquema del disseny simple de sèries de temps											
Exp	R ₁	R ₂	R ₃	R _p	X	R _{p+1}	R _{p+2}	R _{p+3}	R _q
<ul style="list-style-type: none"> ○ Es treballa normalment amb un grup experimental, tot i que se'n podrien introduir més ○ Els registres pretest mostren l'evolució de les dades abans de l'aplicació del tractament (nivell, variabilitat, possibles tendències) ○ L'absència de grup control o altres elements de comparació no permet assegurar que els canvis siguin causats per la intervenció ○ Els registres múltiples poden ajudar a minimitzar l'efecte de variables de confusió d'acció puntual 											

És un cas molt senzill, on per a un grup de subjectes es fan múltiples registres de la variable dependent abans i després de la introducció del tractament. Com a qualsevol altre disseny de sèries temporals, és recomanable que els nombre de registres abans i després de la intervenció no sigui massa diferent, per assegurar la comparabilitat dels resultats. En cas de dubte, el més habitual és que es facin més registres post-test que no pretest (per valorar millor l'evolució de les coses una vegada aplicat el tractament), però no convé desequilibrar molt les dues fases.

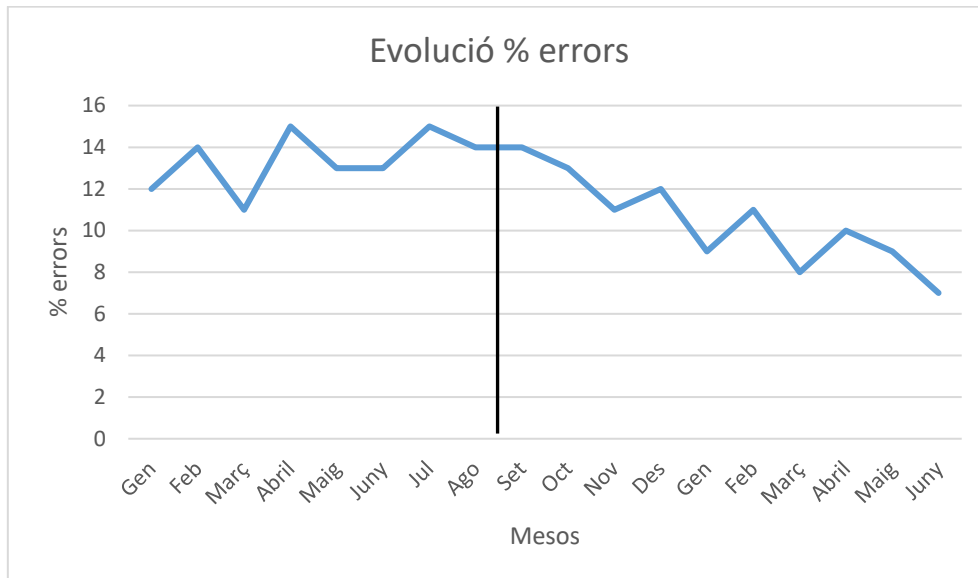
El disseny simple de sèries de temps és el més dèbil d'entre els quasi-experimentals de sèries temporals, ja que no disposem de grup control, ni de variable dependent no equivalent o qualsevol altre element de comparació. Fins i tot si s'observa un canvi clar i persistent dels resultats entre el pretest i el post-ttest, no es pot assegurar que la intervenció experimental sigui la causa real d'aquest canvi, ja que poden existir altres factors que hagin coincidit aproximadament en el temps i que puguin explicar els canvis produïts a les dades, i no tenim manera de contrastar aquesta possibilitat comparant els resultats amb els d'un grup control o una variable dependent no equivalent.

El fet de tenir sèries de registres tant abans com després de la intervenció és important, ja que ens permet evitar errors importants d'interpretació. Podríem trobar, per exemple, que després de la intervenció els resultats de la variable dependent mostren una tendència creixent, i pensar que això és degut al tractament experimental. Però aquesta interpretació canviaria totalment si sabem que aquesta tendència creixent ja existia en els registres pretest. En aquest segon cas l'efecte del tractament experimental s'ha de posar totalment en dubte.

Exemple: En una empresa s'observa que el percentatge de productes amb algun defecte, que necessiten revisió posterior, és excessivament elevat. Es fan registres mensuals del % de productes erronis al llarg de 8 mesos, i posteriorment s'implanta un canvi de sistema que inclou

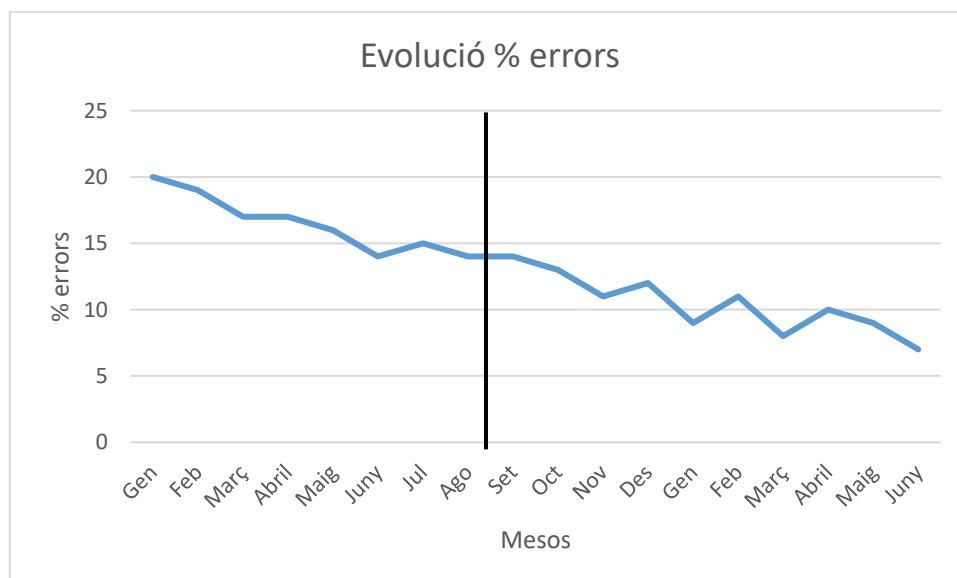
modificacions ambientals i de certes condicions de treball. Durant la implantació d'aquest canvi es continuen realitzant registres mensuals del % de productes defectuosos.

Uns resultats hipotètics d'aquest estudi es poden representar de la forma següent:



Els resultats mostren una reducció progressiva dels errors després de la implantació del nou sistema, cosa que fa pensar en la seva possible eficàcia.

Ara bé, aquesta mateixa reducció no s'interpretaria igual si els resultats fossin els següents:



En aquest segon cas és evident que la reducció del percentatge d'errors ja es venia produint abans de la introducció del nou sistema, i per tant no podem pensar que sigui aquesta intervenció la responsable de la caiguda que es segueix produint després de la seva aplicació.

. Disseny de sèries temporals amb grup control no equivalent

Esquema del disseny de sèries de temps amb grup control no equivalent											
Exp	R ₁	R ₂	R ₃	R _p	X ₁	R _{p+1}	R _{p+2}	R _{p+3}	R _q
Control	R _{q+1}	R _{q+2}	R _{q+3}	R _r	X ₀	R _{r+1}	R _{r+2}	R _{r+3}	R _s

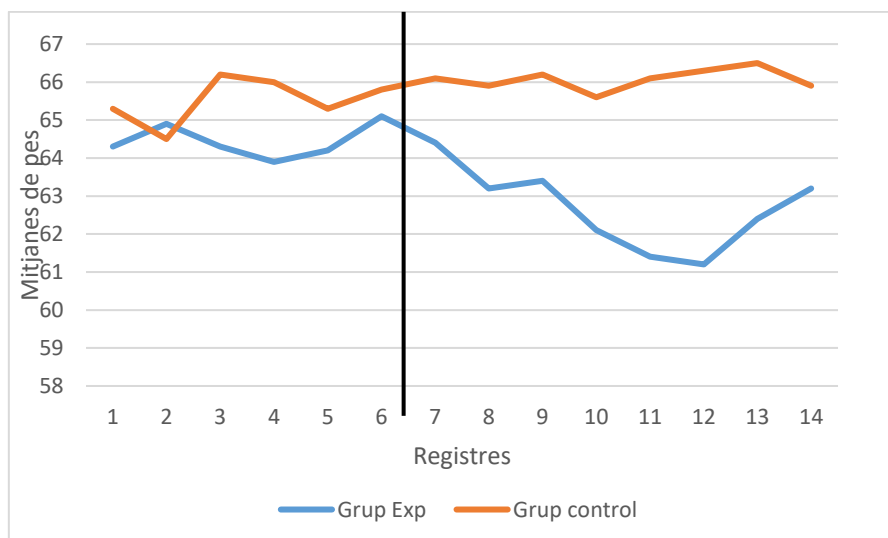
○ La comparació entre el grup experimental i el control permet establir amb més seguretat si els canvis observats en el grup experimental són deguts o no a la introducció del tractament

És equivalent al cas transversal, amb la diferència de que es fan diversos registres pretest i post-test per als grups experimental i control. El grup control és un element bàsic de comparació, ja que esperem que, si el tractament experimental és eficaç, el grup experimental mostri canvis (més o menys ràpids i intensos) a partir del moment de la intervenció, cosa que no hauria de passar en el grup control. Naturalment, el fet de disposar de sèries temporals ens ofereix informació sobre aspectes molt importants, com pot ser la presència o no de tendències abans de la intervenció, i la comparació entre els dos grups pel que fa a aquestes possibles tendències; o l'evolució en el temps de la dinàmica dels dos grups després de l'aplicació del tractament. Els patrons de resultats poden ser diversos.

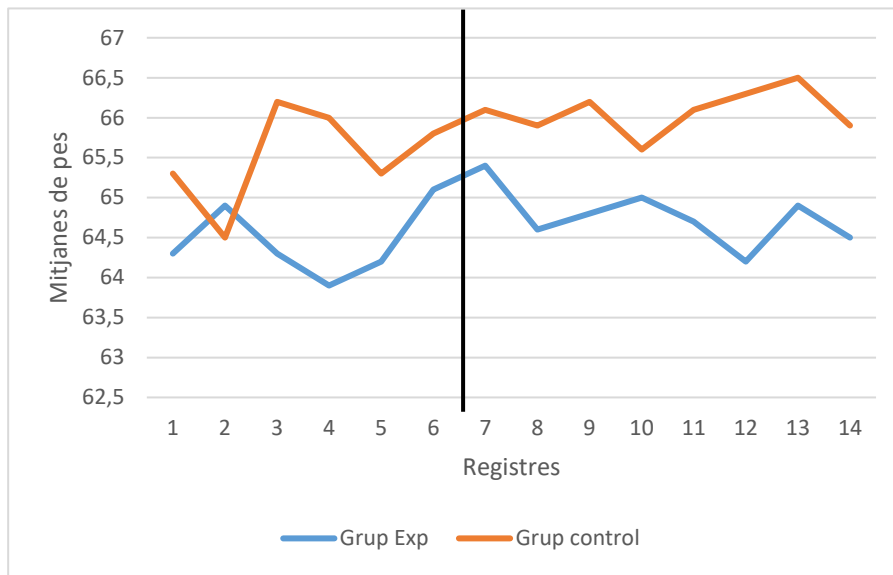
Suposem, per exemple, que es treballa amb dos centres de dietètica diferents, i que es vol valorar la possible eficàcia d'un nou sistema motivacional per a la reducció de pes. Inicialment es fan registres setmanals de la mitjana de pes del conjunt dels participants a cadascun dels dos grups, durant 6 setmanes (setmanes 1-6). A la setmana 7 s'introdueix al grup experimental el nou sistema, mentre que amb l'altre no es fa cap canvi. Es continuen fent registres setmanals de pes fins a la setmana 14.

Als gràfics següents es mostren dos resultats hipotètics d'aquest estudi:

Resultats hipotètics 1:



Resultats hipotètics 2:



El primer gràfic de resultats mostra una possibilitat més clara d'efecte del nou sistema que no pas el segon. En aquest últim cas, és cert que els resultats del grup experimental estan per sota dels del grup control, però això ja passava abans de la introducció del nou sistema (només amb l'excepció del registre número 2) i, per altra banda, no s'observa cap canvi en la dinàmica dels resultats abans i després de la intervenció.

. Disseny de sèries temporals amb variables dependents no equivalents

Esquema del disseny de sèries de temps amb variables dependents no equivalents

Exp	$R_{1(1)}$	$R_{2(1)}$	$R_{3(1)}$	$R_{p(1)}$	X	$R_{p+1(1)}$	$R_{p+2(1)}$	$R_{p+3(1)}$	$R_{q(1)}$
	$R_{1(2)}$	$R_{2(2)}$	$R_{3(2)}$	$R_{p(2)}$	(X)	$R_{p+1(2)}$	$R_{p+2(2)}$	$R_{p+3(2)}$	$R_{q(2)}$

- Mateixa lògica que el disseny transversal corresponent, amb l'afegit d'informació i control que suposen les sèries de temps
- Simbologia similar al disseny transversal: Els subíndexs (1) i (2) es refereixen a les dues variables dependents registrades, i el símbol (X) indica que, tot i aplicar-se el tractament, en el cas de la segona variable dependent no s'esperen canvis en els resultats

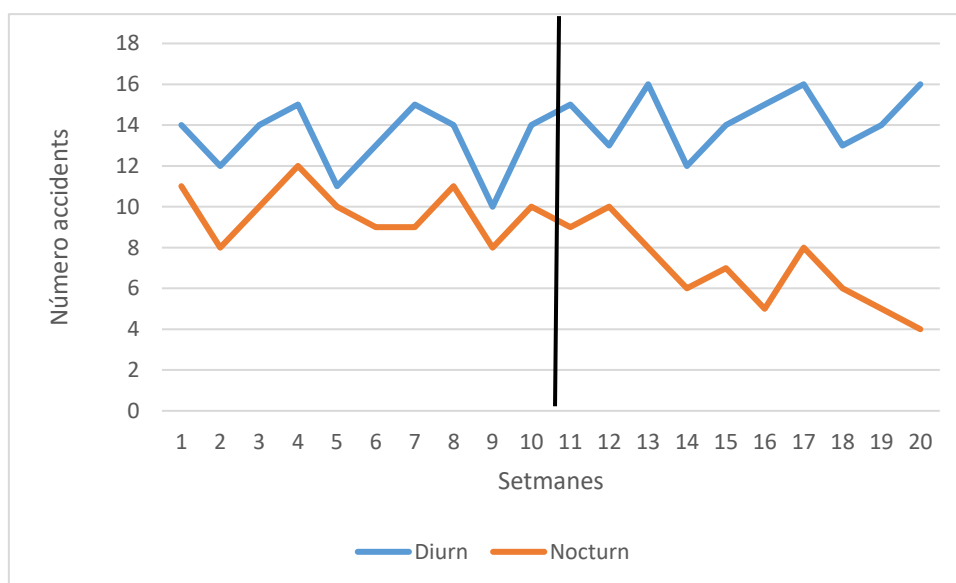
Novament se segueix aquí una lògica semblant a la del cas transversal, amb la diferència que per a les diferents variables dependents implicades es faran diversos registres tant abans com després de la intervenció. És molt important que el registre de les diferents variables dependents es faci de forma simultània, per assegurar la funció de control que exerceix la variable dependent no equivalent.

A l'esquema del disseny les variables dependents s'indiquen entre parèntesi, de forma que $R_{1(1)}$ representa el primer registre de la variable dependent 1, mentre que $R_{1(2)}$ és el primer registre per a la variable dependent 2. Com en altres casos, el signe X indica l'aplicació del tractament, però en el cas de la variable dependent no equivalent es pot substituir per (X), indicant que no esperem que la intervenció tingui efecte sobre aquesta segona variable.

Com s'ha indicat al parlar de la versió transversal del disseny, en un disseny de variables dependents no equivalents podem trobar dues variants. La més habitual és que a la variable dependent que es vol estudiar si li afegeixi una segona on esperem que no hi hagi canvis. A l'exemple que s'ha presentat abans, la variable dependent que es volia estudiar era el consum de begudes ensucrades, ja que es volia introduir una campanya per reduir-lo; la segona variable dependent introduïda (consum d'altres begudes) tenia una funció de control i comparació, ja que no esperàvem trobar-hi canvis a partir de la campanya. Ara bé, hi ha una segona variant del disseny de variables dependents no equivalents, i és aquella en la qual treballem amb una única variable dependent, però mesurada en dues o més situacions diferents, de forma que esperem que es produeixin canvis diferencials en una i altra situació. L'exemple que es presenta aquí correspon a aquesta segona modalitat.

Exemple: Les autoritats de trànsit d'un determinat territori decideixen intentar reduir el nombre d'accidents de trànsit que es produeixen a la nit millorant la visibilitat nocturna dels senyals de trànsit de perill (millora de la reflectància i altres característiques). Disposen de registres del número d'accidents a les 10 setmanes anteriors a la introducció del canvi, i la idea és obtenir una sèrie de registres durant 10 setmanes més després de la intervenció. Ara bé, per assegurar que qualsevol canvi és degut realment a la intervenció, i no a qualsevol altre factor, decideixen també fer un registre dels accidents que es produeixen durant les hores diürnes. La lògica de l'estudi és que si s'actua sobre aspectes de la visibilitat nocturna dels senyals, una possible reducció d'accidents s'hauria de produir només a les hores nocturnes, i no hi ha cap raó per esperar una afectació a les hores diürnes.

Uns resultats hipotètics que donarien suport a aquesta expectativa són les següents:



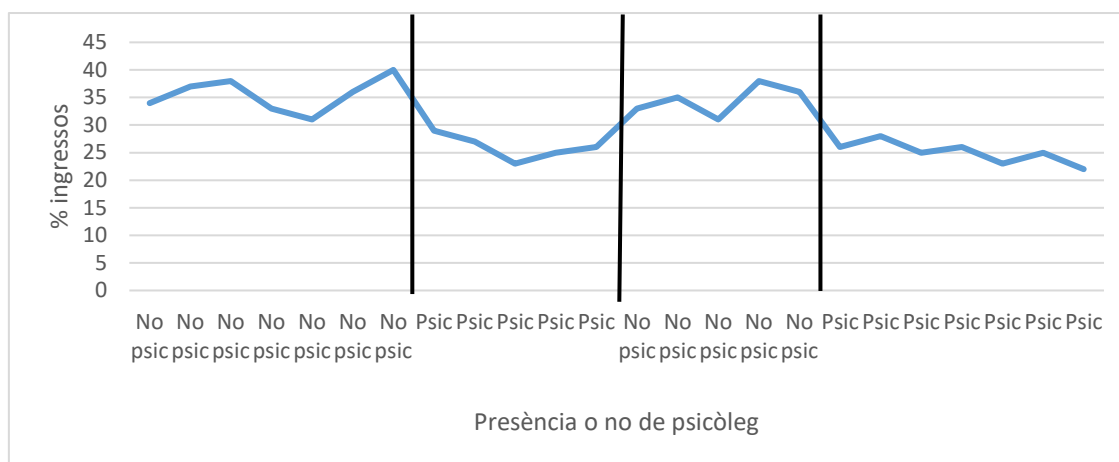
. Disseny de retirada de tractament i disseny de tractament repetit

Esquema dels dissenys de retirada de tractament i tractament repetit	
Retirada de tractament:	
Exp	$R_1 R_2 \dots R_p X R_{p+1} R_{p+2} \dots R_q \bar{X} R_{q+1} R_{q+2} \dots R_r$
Tractament repetit:	
Exp	$R_1 \dots R_p X R_{p+1} \dots R_q \bar{X} R_{q+1} \dots R_r X R_{r+1} \dots R_s$
<ul style="list-style-type: none"> ○ Mateixa lògica que els corresponents dissenys transversals, amb l'element afegit de les sèries temporals ○ El disseny de tractament repetit es pot allargar amb successives introduccions i retirades del tractament experimental 	

Són semblants també als corresponents casos transversals. Es produeix la introducció del tractament experimental i posteriorment la seva retirada, amb la possibilitat de tornar a repetir el procés diferents vegades, si és possible i convenient (tractament repetit o replicacions múltiples). Es realitzen sèries de registres en cada fase de l'estudi (pretest, després d'introduir el tractament, després de retirar-lo, etc.). Cal insistir en la conveniència que el nombre de registres a cada fase no sigui gaire diferent, novament per assegurar la comparabilitat dels resultats; en tot cas cal evitar que el número de registres a qualsevol de les fases sigui molt petit, per assegurar el mínim d'informació necessari.

Exemple: A un determinat districte sanitari es decideix incorporar un/a psicòleg o psicòloga als equips d'intervenció immediata, amb l'expectativa de que això podria reduir el nombre d'ingressos hospitalaris. La idea és que un número significatiu de casos on participen els equips d'intervenció corresponen a quadres d'ansietat, factors psicosocials o altres aspectes que poden ser abordats sense necessitat d'ingrés hospitalari. Per assegurar els resultats es procedeix a una retirada del tractament després d'un cert període, i es torna a introduir més endavant. Suposem que es registra el percentatge d'intervencions que acaben en ingrés hospitalari.

Uns resultats hipotètics serien els següents:



És fàcil observar que en els dos períodes on hi ha presència de psicòlegs als equips es redueix el percentatge d'ingressos hospitalaris, i també que quan deixa d'haver-hi presència (tercera fase), aquest percentatge torna a incrementar-se fins als nivells inicials. Disposem, per tant, de diferents evidències que ens indiquen que probablement la presència de psicòlegs als equips d'intervenció té un efecte positiu.

Cal recordar que habitualment en aquests dissenys no hi ha grup control ni tampoc variable dependent no equivalent, per tant la valoració del possible impacte de la intervenció s'ha de fer a partir del comportament dels resultats quan s'introdueix i quan es retira el tractament.

Com s'ha indicat abans, és fonamental tenir present que aquest tipus de dissenys només tenen sentit quan l'impacte de la intervenció pot ser reversible, com és el cas de l'exemple plantejat. Si l'efecte del tractament no és reversible (o no se sap amb seguretat) el disseny no és útil, ja que la retirada del tractament no té per què venir acompanyada d'un retorn dels resultats al seu nivell inicial, i per tant no disposem de l'element principal de comparació que ens ofereix aquest disseny.

. Disseny de replicacions asincròniques

Com es veurà immediatament, el disseny de replicacions asincròniques és bastant diferent de tots els que s'han vist anteriorment, sobretot pel fet de fer un ús diferent de la dinàmica temporal pròpia dels dissenys de sèries temporals.

Esquemàticament, l'estructura i les característiques principals del disseny de replicacions asincròniques serien les següents:

Esquema del disseny de sèries de temps amb replicacions asincròniques

Exp	R ₁	R ₂	R ₃	R _p	X	R _{p+1}	R _{p+2}	R _{p+3}	R _q
Exp	R _{q+1}	R _{q+2}	R _{q+3}	R _r	X	R _{r+1}	R _{r+2}	R _{r+3}	R _s

- Dos o més grups experimentals reben el mateix tractament en moments diferents
- Inicialment es fan registres pretest dels diversos grups, i posteriorment es va introduint el tractament en moments diferents per a cada grup, sense deixar de fer registres de la variable dependent
- Quan un grup rep tractament i l'altre no, aquest segon grup fa la funció comparativa pròpia del grup control
- Es pot combinar amb altres actuacions: per exemple, amb retirades i reintroduccions del tractament es podria anar alternant el paper de grup experimental i control per als diferents grups

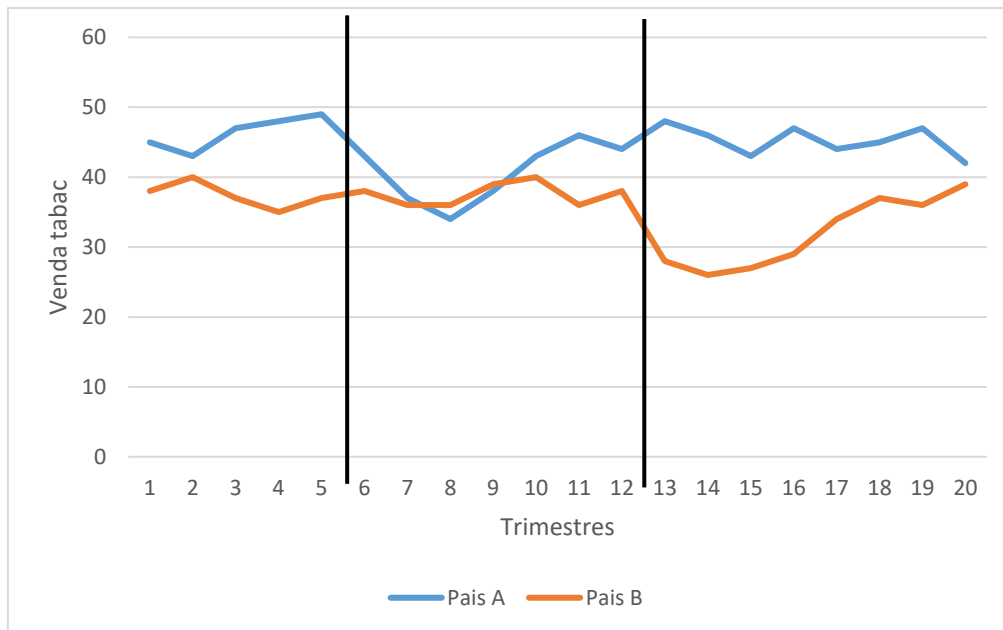
Les diferències més importants d'aquest disseny respecte dels anteriors són les següents: En primer lloc, hi ha com a mínim dos grups, però no es tracta de grups experimental i control, com en el cas del disseny de grup control no equivalent, sinó de diversos grups experimentals. Tots ells reben la mateixa intervenció, i l'element diferencial és el *moment* de l'aplicació del tractament, ja que s'introdueix en moments diferents per a cadascun dels grups. En l'esquema proposat, el primer grup rep el tractament abans que el segon (i, per tant, el segon grup té un major número de registres pretest que el primer).

És habitual que aquest disseny es combini amb la idea de retirada i repetició del tractament, però sempre mantenint un cert nivell d'asincronicitat, de manera que les retirades i possibles reintroduccions del tractament no es produïrien simultàniament en els diversos grups. D'aquesta manera, quan un grup està sota tractament l'altre pot no estar-ho, i més endavant es pot invertir aquesta situació.

Òbviament, l'element principal per valorar el possible efecte de la intervenció és la comparació entre el comportament dels dos grups a les diferents fases.

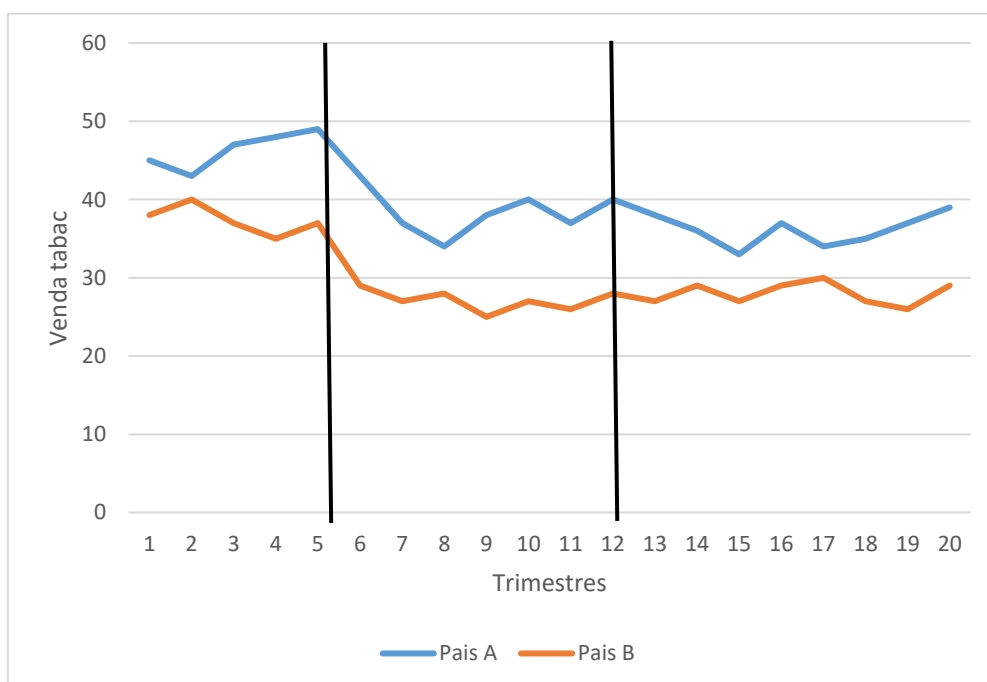
Un exemple senzill de disseny de replicacions asincròniques seria el següent: Suposem que es vol esbrinar si la introducció d'etiquetes d'avertència sobre el perill del tabac als paquets de cigarretes té algun impacte en el consum. Imaginem també que aquestes etiquetes, amb un format similar, són introduïdes en un cert país en un determinat moment, mentre que a un altre país del seu entorn s'introdueixen 7 mesos després. Suposem que disposem de mesures mensuals de les vendes de tabac a cadascun dels dos països.

Els següents resultats hipotètics ens mostren una de les situacions que es poden produir:



Es pot observar fàcilment que el comportament en els dos països ha estat semblant: La introducció del tractament ha produït una reducció de les vendes de tabac, però aquest efecte ha estat transitori, i s'ha retornat lentament als nivells inicials. L'important és veure que en cadascun dels països el canvi s'ha produït en el moment d'introduir el tractament, no abans, de forma que els resultats ens fan pensar que la intervenció ha tingut un impacte, tot i que només hagi estat temporal.

Un altre patró de resultats molt diferent seria el següent:



En aquest segon cas sembla probable que els canvis observats no siguin deguts a la intervenció. Això ens ho indiquen dos factors:

- Simultàniament a la introducció del tractament al país A (primera línia vertical) es produeix una reducció de vendes, però no solament en aquest país, sinó també al país B, en el qual la intervenció encara no ha estat introduïda.
- La introducció del tractament al país B (segona línia vertical) no produeix cap canvi

En aquest cas es podria pensar en algun factor desconegut, diferent a la intervenció introduïda, que pot haver influït en els resultats, i que caldria identificar.

Breus consideracions sobre l'anàlisi de dades en els dissenys quasi-experimentals de sèries temporals

Com es pot observar en molts dels exemples plantejats, sovint els dissenys quasi-experimentals de sèries temporals no mostren un patró de resultats comparable amb el dels dissenys experimentals clàssics. Tant en els dissenys experimentals com en els quasi-experimentals es treballa amb grups de subjectes, però hi ha una diferència important: En els dissenys experimentals coneixem exactament la composició dels grups, i disposem de resultats individuals per a cadascun dels subjectes participants. D'aquesta forma és aplicable la lògica de proves com l'anàlisi de la variància, que es basen en comparar, per exemple, la variància entre els grups amb la variància intra-grups. En alguns dissenys quasi-experimentals això també pot passar (veure, per exemple, el cas del sistema motivacional per a la reducció de pes, on tenim resultats individuals que podem estudiar en forma de mitjanes i variàncies), de forma que seria possible utilitzar proves estadístiques convencionals com l'anàlisi de la variància. Però en molts altres casos els resultats obtinguts no permetran fer-ho. Per exemple, en el darrer cas exposat (impacte de l'etiquetat sobre la venda de tabac), la variable dependent registrada és el número (en milers) de paquets de tabac venuts. Aquest és un resultat global per al conjunt del país, encara que vingui donat per les decisions individuals de moltes persones. Per tant, per a cada ocasió de registre tenim una dada global única, i en conseqüència no són aplicables les proves estadístiques convencionals (tot i que, com es veurà a l'apartat de dissenys de replicació intra-subjecte, tenim la possibilitat d'utilitzar algunes eines de modelització, com són els models de regressió lineal).

En aquests casos el més habitual és adoptar una perspectiva doble per a l'anàlisi de resultats: Per una banda, la valoració directa dels patrons de canvi que es mostren als gràfics de dades (canvis de nivell, de tendència, etc.) i, per una altra, l'aplicació d'algunes tècniques pensades específicament per a l'anàlisi de sèries temporals. Considerant l'element de subjectivitat que es pot produir en la valoració dels gràfics, un enfocament complementari entre les dues possibilitats és el millor.

Al següent apartat (dissenys de replicació intra-subjecte) es plantejarà una problemàtica similar, i allà es farà referència tant als principals patrons de canvi que es poden observar en els gràfics de resultats com a algunes tècniques numèriques o gràfiques que poden ser útils també per als dissenys quasi-experimentals de sèries temporals.

IV Dissenys experimentals de replicació intra-subjecte o N=1

1.- Introducció

L'element fonamental que diferencia els dissenys N = 1, de cas únic o de replicació intrasubjecte de tots els anteriors és el fet de no treballar amb grups de subjectes sinó amb casos individuals. Per tant, tots els elements que es plantejaven en altres casos (mostratge, formació dels grups, inferència estadística, etc.) deixen d'estar presents o canvien substancialment en aquest tipus de dissenys.

La idea bàsica és l'obtenció de múltiples registres al llarg del temps d'una mateixa persona o cas en diferents situacions (per exemple, aplicació i no aplicació d'un determinat tractament). Es tracta, per tant, de dissenys longitudinals o de sèries temporals, amb la diferència que els resultats no corresponen a cap agrupació de subjectes (grup format per l'experimentador, escola, empresa, hospital, territori, etc.) sinó a casos individuals (N=1).

Com passava en els dissenys quasi-experimentals de sèries temporals, en el dissenys N = 1 els resultats obtinguts a les fases de no-tractament (anomenades aquí *línia base*) serveixen com element de comparació amb les dades registrades durant l'aplicació de la intervenció (*fases de tractament*).

Esquemàticament:

Característiques generals dissenys N = 1
<ul style="list-style-type: none"> - Un sol subjecte o unitat experimental - Mesura de les variables dependents en múltiples ocasions de registre <ul style="list-style-type: none"> → Sèries temporals, disseny longitudinal - Presència de fases de no intervenció (línia base) i d'intervenció - Les fases de línia base tenen una funció comparativa i de control (paper semblant al del grup control en altres dissenys) - La comparació entre fases de línia base i d'intervenció permet valorar l'efecte del tractament

Els dissenys de cas únic tenen el seu origen en un enfocament de la recerca psicològica que prefereix posar l'èmfasi en l'estudi repetit de casos particulars, més que en l'anàlisi dels resultats obtinguts en grups de subjectes. Aquest enfocament individual de la recerca té una llarga tradició en Psicologia (Wundt, Fechner, Pavlov, Köhler, Watson, etc.), i va ser portat al seu grau màxim de sistematicitat per part de l'escola del conductisme radical encapçalada per B.F. Skinner. D'acord amb el plantejament de l'escola skinneriana, el comportament animal i humà està regit en gran part per factors de caràcter ambiental que poden ser controlats, de manera que és possible preparar situacions molt estandaritzades (per exemple, la caixa de Skinner) on es faci efectiu aquest control. Si això és així, s'obre la possibilitat de realitzar a l'àmbit de la Psicologia estudis experimentals similars als d'altres ciències, és a dir, l'estudi de casos individuals en situacions altament controlades. Si, per exemple, s'introdueix un programa de reforçament a un animal i s'observen canvis subsegüents en el seu comportament en una situació molt controlada, es poden treure conclusions sobre l'eficàcia d'aquest programa, i a

més aquestes conclusions podrien ser generalitzades a partir del supòsit bàsic de comportament regular de la naturalesa, com es fa en altres disciplines.

En altres paraules, si en el plantejament clàssic o de grups la generalització es planteja a partir de l'estudi del comportament de mostres suposadament representatives, en el cas $N=1$ els principis de generalització són el control estricte de la situació i el principi de regularitat dels fenòmens de la naturalesa.

Per altra banda, i recordant el que s'indicava en parlar dels diferents tipus de control experimental (apartat I.3), està clar que des d'aquest enfocament es fa una aposta pel control experimental directe, i no es pren normalment en consideració el control probabilístic.

Cal indicar, però, que els dissenys $N=1$ no s'utilitzen només des d'aquesta perspectiva, molt lligada a la recerca bàsica, sinó que també poden ser útils en moltes situacions aplicades. Els dissenys de cas únic poden ser utilitzats en situacions on hi hagi una intervenció adreçada a un subjecte en particular (o a diferents participants, però tractats cadascun d'ells individualment), sempre que es puguin obtenir registres fiables i continuats d'una o diferents variables dependents abans i després de la introducció de la intervenció, en condicions que permetin un cert control de variables de confusió. Un camp d'ús molt evident dels dissenys $N=1$ és el de la psicologia clínica, on es treballa gairebé sempre amb pacients individuals o amb unitats identificades (per exemple, una família), però també poden ser utilitzats en altres camps on tingui sentit una intervenció individualitzada (àmbit educatiu, psicosocial, etc.).

Cal remarcar, en tot cas, que a les situacions aplicades és molt difícil o impossible arribar als nivells de control que es poden assolir en situacions de laboratori molt estandaritzades. Per tant, tot i que l'estructura del disseny pugui ser experimental (amb la introducció de diferents fases que permetin obtenir elements de comparació suficients), caldrà posar molta atenció a les possibles variables de confusió que puguin actuar en cada cas.

A les situacions aplicades (p.e. clíniques, educatives, etc.) normalment no es pot plantejar una generalització automàtica dels resultats a partir d'un cas particular, ja que el nivell de control assolit difícilment serà òptim, i el número de factors que poden influir en els resultats és ampli. En tot cas la possibilitat de generalització passa per l'acumulació d'evidència a partir de la replicació de l'estudi amb altres casos, i la comprovació de si els resultats són similars o no.

Esquemàticament:

<p>Recerca bàsica</p> <ul style="list-style-type: none"> - Ús de metodologies pròpies de les ciències experimentals clàssiques: Control estricte de les situacions - Generalització d'acord amb el patró habitual de les ciències experimentals (generalització a partir de casos molt controlats, regularitat de la naturalesa)
<p>Recerca aplicada</p> <ul style="list-style-type: none"> - Caràcter individual de moltes aplicacions psicològiques - Control gairebé sempre menys estricte que en recerca bàsica - Generalització per acumulació d'evidència a partir de rèpliques de l'estudi

A més del que s'ha indicat, els dissenys N=1 mostren algunes limitacions o dificultats que tenen a veure amb el registre de resultats. Cal pensar que aquest plantejament obliga a registrar repetidament una mateixa variable o variables dependents amb la fiabilitat necessària. Resumidament, els aspectes principals serien:

Consideracions respecte del registre de les variables dependents

- Habitualment no és fàcil obtenir sèries llargues de dades, per raons pràctiques o per la naturalesa de la pròpia variable dependent
- El tipus de variables dependents utilitzades estableix limitacions més o menys grans a l'hora d'obtenir sèries de dades: Distinció entre variables conductuals, fisiològiques, cognitives, actitudinals, etc, etc.
- Necessitat crucial d'assegurar la fiabilitat del registre de les variables dependents
- Possibles efectes reactius del registre continuat

Com es pot veure, aquestes dificultats tenen a veure amb el registre repetit de la mateixa variable. Les variables dependents que es poden registrar amb major fiabilitat i en sèries llargues en un disseny d'aquest tipus són les de caràcter conductual, és a dir, registres de comportaments manifestos del subjecte (consum de tabac, número de comportaments inadequats, quilòmetres recorreguts en un programa d'entrenament, etc.) o les de tipus biològic o fisiològic (pes corporal, hores de son, etc.), tot i que en determinats casos també poden existir problemes (p.e. el registre de consum de tabac pot basar-se en auto-informes, amb el perill de manca de precisió que això suposa). Hi ha altres variables que presenten una dificultat especial: per exemple, és molt poc adient utilitzar repetidament un instrument psicomètric per valorar un possible canvi d'actitud, ja que l'aplicació repetida d'un mateixa escala (o fins i tot d'una escala similar) pot produir tota classe de distorsions en la resposta del subjecte.

Algunes decisions importants que cal prendre a l'hora de plantejar un disseny N = 1 són les següents:

Algunes decisions bàsiques en un disseny N = 1

- . Número de fases de cada tipus (línia base i intervenció) que s'han d'introduir
- . Establir criteris per tal de decidir el moment de començar la intervenció
- . Número de registres per fase
- . Intervals entre registres: Sensibilitat al canvi

A l'hora de prendre decisions en el disseny de l'estudi es pot produir una certa contraposició entre el que metodològicament seria més correcte i el que la situació real permet. Per exemple,

quan més llargues siguin les sèries de dades, major informació obtindrem, però és obvi que en una situació real no ens podrem plantejar períodes de registre molt llargs (sobretot si hi ha necessitat d'una intervenció urgent, com passa sovint en l'àmbit clínic). Per altra part, quantes més vegades puguem introduir i retirar el tractament experimental (per tant, quantes més fases de línia base i tractament tinguem), millor informació tindrem, però també està clar que això té limitacions pràctiques.

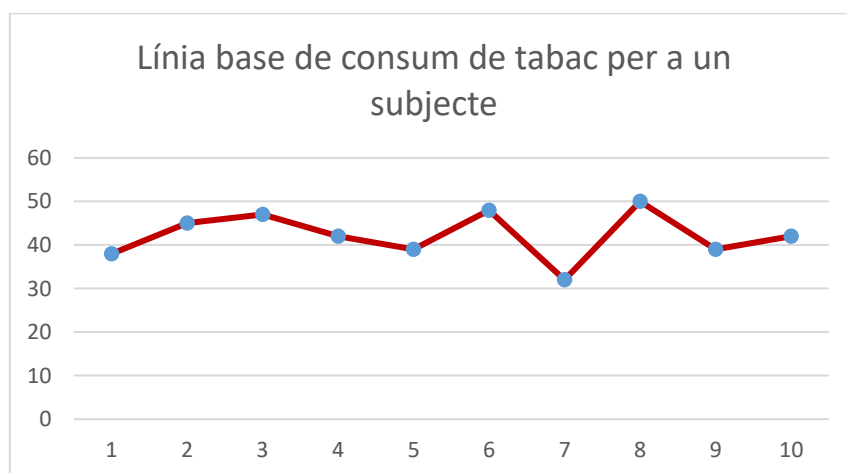
Com s'indicava en el cas d'altres dissenys longitudinals, també és important assegurar que a cada fase hi ha un número suficient de registres per assegurar un coneixement mínim de la dinàmica temporal implicada: veure si hi ha tendències, si els resultats són poc o molt variables, etc.

Per altra banda, l'interval entre registres és també fonamental per afinar la informació. Intervals massa curts poden ser redundants i innecessaris, mentre que intervals massa llargs ens poden fer perdre informació rellevant, com s'ha indicat ja anteriorment a partir d'alguns exemples.

2.- Terminologia general i alguns aspectes metodològics dels dissenys N = 1

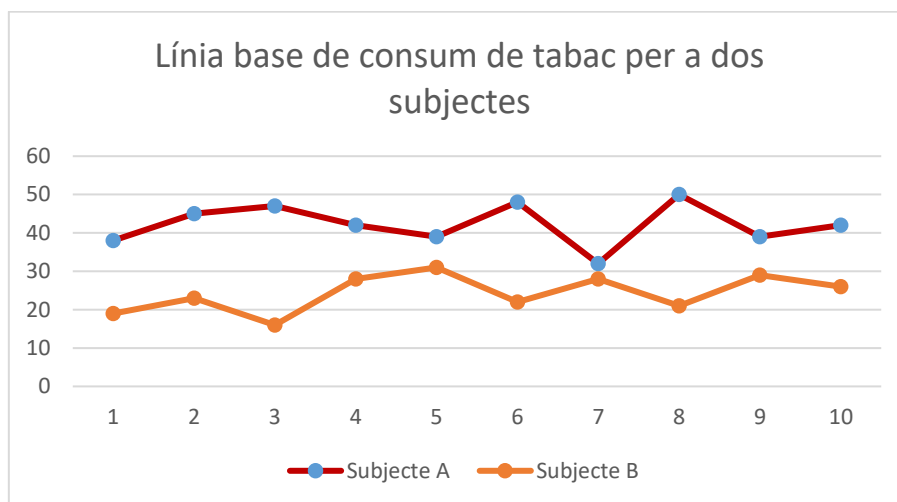
Com s'ha indicat, els dissenys de replicació intra-subjecte s'organitzen en un conjunt de fases, entre les quals cal distingir-ne de dos tipus: fases de no intervenció (línia base) i fases d'intervenció o tractament. Aquestes últimes no tenen per què ser idèntiques entre sí, en fases diferents es poden aplicar tractaments també diferents, o fins i tot combinacions de més d'un tractament.

La línia base no és altra cosa que una sèrie de registres de la variable dependent que s'hagi definit, obtinguts en situació de no intervenció per part de l'investigador. Lògicament, el seu objectiu és esbrinar quin és el comportament inicial del subjecte experimental pel que fa a la variable estudiada. Gràficament els resultats de la línia base es representen com qualsevol altra sèrie temporal. Per exemple, si s'estudia el consum diari de tabac durant 10 dies per a una determinada persona (abans d'introduir qualsevol intervenció), els resultats podrien ser:

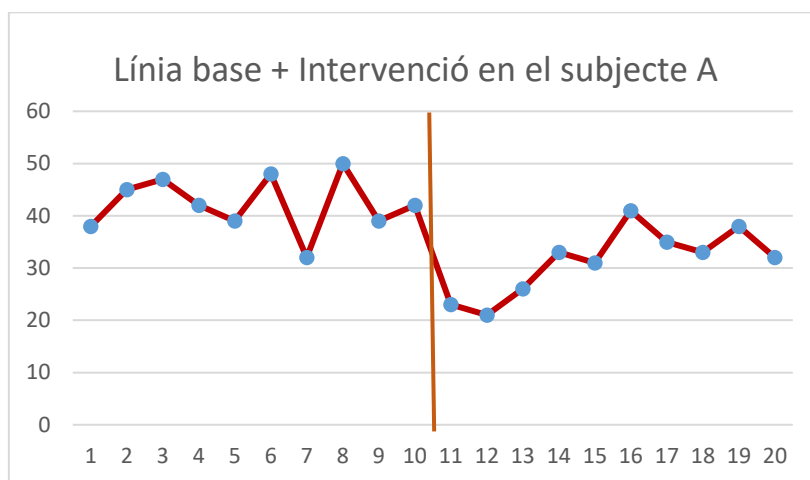


Si treballem amb més d'un subjecte participant, hem de mantenir sempre el caràcter individualitzat de l'anàlisi per a cadascun d'ells, el que no farem mai és obtenir mitjanes ni altres

estadístics similars. Si, per exemple, treballem amb dos pacients de tabaquisme, obtindrem les línies base corresponents a cadascun d'ells:



Naturalment, el fet de disposar d'una línia base prou extensa ens oferirà la possibilitat d'utilitzar-la com a terme de comparació per al possible efecte de qualsevol tractament. Això es pot veure fàcilment, per exemple, en el gràfic següent:



Com en altres dissenys de sèries temporals, la línia vertical separa les dues fases del disseny, en aquest cas línia base i intervenció. En aquest exemple sembla que la introducció del tractament pugui contribuir a una reducció del consum, però aquesta reducció sembla transitòria, ja que uns dies després es produeix un cert retrocés, tot i no arribar (de moment) als nivells inicials.

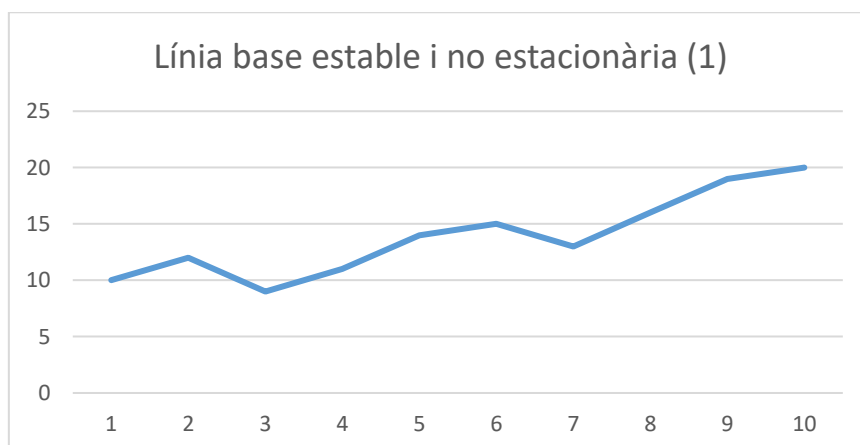
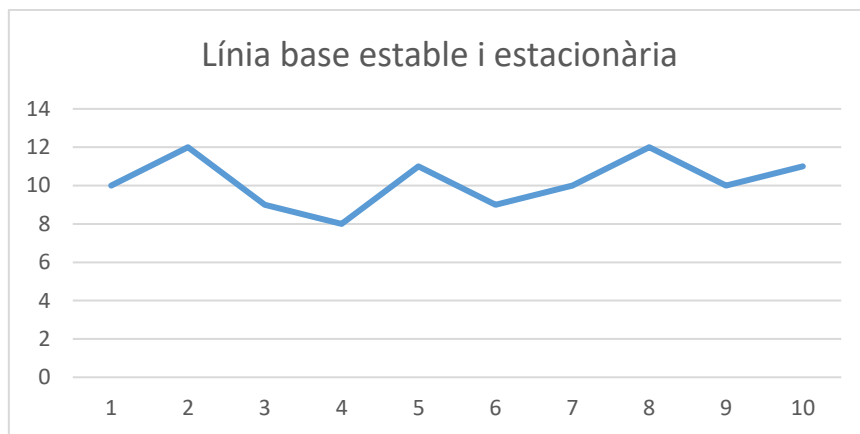
Cal dir, en tot cas, que un disseny tan senzill com aquest (fase de línia base + fase d'intervenció) no es considera suficient per establir amb seguretat l'impacte del tractament, serà necessari introduir fases posteriors de retirada i re-introducció de la intervenció i veure si la resposta del subjecte es repeteix o no.

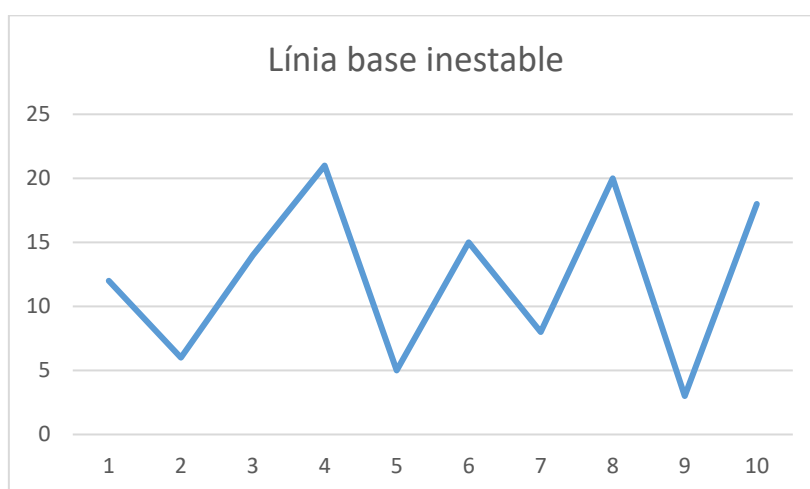
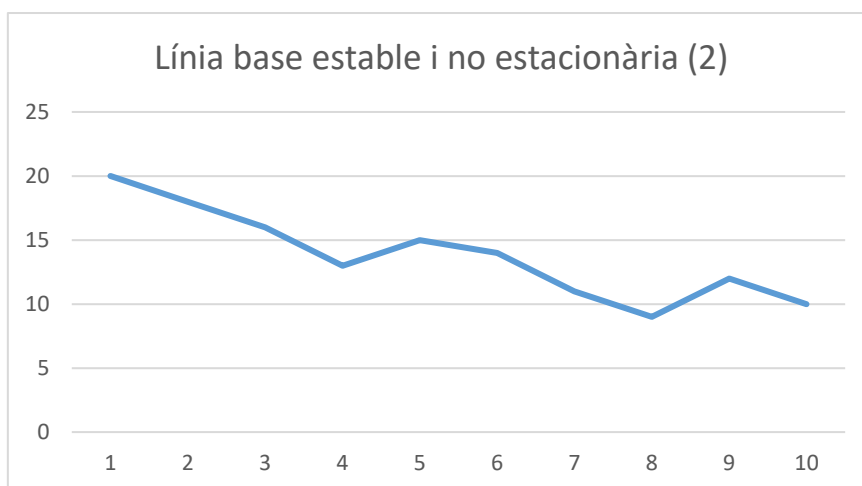
La línia base juga una funció comparable amb la del grup control en altres dissenys: ofereix un conjunt de resultats obtinguts en situació de no intervenció, cosa que li permet oferir una possibilitat de comparació amb els resultats obtinguts quan s'apliqui el tractament.

És fàcil veure que les característiques de la línia base són molt importants a l'hora d'utilitzar-la com a element de comparació per al possible efecte de qualsevol intervenció experimental. Hi ha dues característiques bàsiques que cal tenir en compte:

- . Estabilitat: Es refereix al fet que la línia base tingui una variabilitat limitada (tot i que no hi ha un criteri estricte per establir aquests límits) i un patró reconeixible
- . Estacionarietat: L'estacionarietat d'una línia base significa que no presenta cap mena de tendència, ni creixent ni decreixent

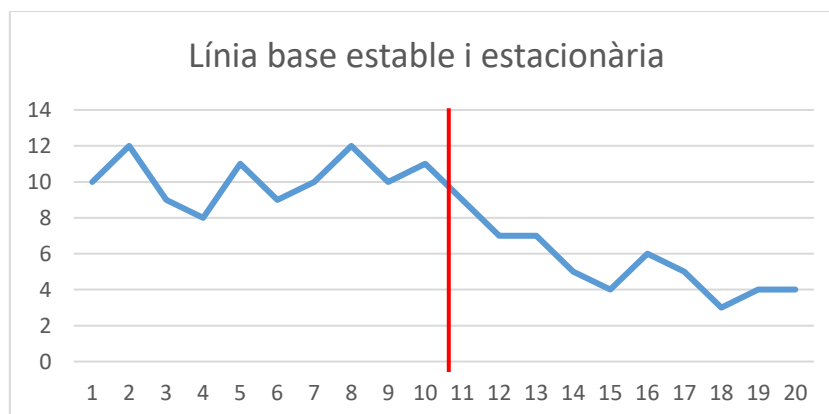
Els gràfics següents mostren diferents possibilitats:

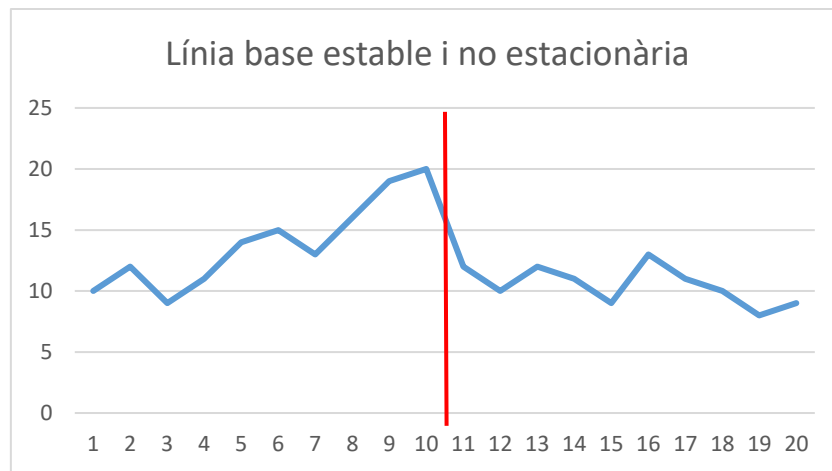




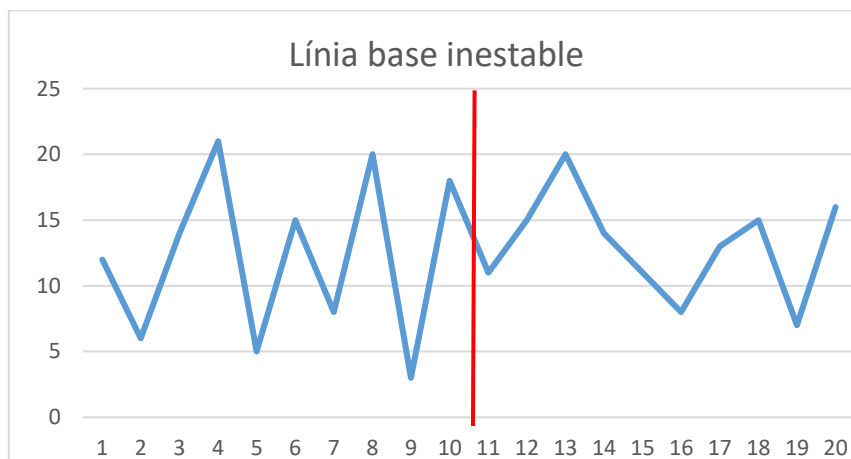
És fàcil veure que la característica més important de la línia base és la seva estabilitat. Una línia base estable (tant si és estacionària com si no) permetrà probablement una comparació més o menys clara amb els resultats de les fases d'intervenció, mentre que una línia base inestable serà molt més difícil d'utilitzar amb aquesta finalitat. Alguns exemples senzills poden il·lustrar aquesta idea:

Gràfics amb interpretació relativament simple (línia base estable i estacionària, o bé estable però no estacionària):





En canvi, amb una línia base inestable la interpretació pot ser menys clara:



. Terminologia per identificar els dissenys N = 1

Per tal d'identificar qualsevol disseny de replicació intra-subjecte el que es fa normalment és representar la seva estructura de fases, d'acord amb una terminologia estàndard que és la següent:

Terminologia bàsica de fases
A : Fase de línia base
A' : Línia base en situacions especials (placebo, incentiu no contingent, etc.)
B : Fase d'intervenció
B, B', B'', etc.: Nivells quantitativament diferents del mateix tractament
B, C, D, etc.: Tractaments qualitativament diferents
B ₁ , B ₂ , B ₃ , etc. : Canvi de criteri

Cal indicar que no a tots els articles de recerca s'utilitza sempre aquesta terminologia, o bé es pot emprar amb alguna variació, per tant és molt important veure detingudament quina és la descripció que es fa de cadascuna de les fases per entendre l'estructura i lògica del disseny utilitzat.

Per exemple, un disseny A – B – A – C inclou una primera fase de línia base, una segona fase en la qual s'aplica un cert tractament, una tercera fase a la qual es retira aquest tractament i es torna a situació de línia base, i una quarta fase on s'introdueix una segona intervenció, diferent de la primera.

Cal posar atenció especial a la diferència entre B, B', B''..., per un costat, i entre B, C, D..., per un altre. En el primer cas, es tracta de nivells quantitativament diferents d'una mateixa intervenció (per exemple, dues dosis diferents d'un fàrmac, durades diferents de sessions terapèutiques, etc.), mentre que en el segon (B, C, D....) s'indiquen intervencions qualitativament diferenciades (per exemple, un tractament farmacològic i una teràpia psicològica, aplicats a un mateix pacient en fases diferents).

El canvi de criteri és un cas especial que es tractarà més endavant.

3.- Classificació i revisió dels principals dissenys de replicació intra-subjecte

Es pot observar fàcilment que poden existir moltes variants dels dissenys de replicació intrasubjecte, ja que per definició són molt flexibles. Per tal de comprendre millor les diferents possibilitats es poden plantejar alguns criteris de classificació dels dissenys N=1. Tot i que no hi ha acord entre diferents autors, un criteri senzill i útil seria considerar que existeixen dos grans tipus de dissenys de cas únic:

- . Dissenys amb comparacions intrasèrie o d'estructura simple
- . Dissenys amb comparacions intrasèrie i entresèries, o d'estructura complexa

La idea de comparacions intrasèrie es refereix a que es comparen resultats del mateix subjecte o unitat obtingudes en diferents fases de línia base i tractament, amb les variants que sigui. En canvi, en els dissenys d'estructura complexa, a més de les típiques comparacions intrasèrie, es plantegen també comparacions entre sèries diferents de dades, referides a comportaments diferents, a situacions diverses o fins i tot a subjectes diferents, tal i com es veurà més endavant.

Esquemàticament, els principals tipus de dissenys de cadascuna de les categories són els següents:

En el cas del dissenys d'estructura simple:

Estructura simple – Comparacions intra-sèrie
A - B (quasi-experimental)
A – B – A , A – B – A – B , A – B – A – B – A , etc.
B – A – B i les seves variants
Diversos nivells: A – B – A – B' (B'', B''',)
Tractaments múltiples: A – B – A – C (D, E,)
Interactiu: A – B – A – C – BC
Canvi de criteri: A – B ₁ – B ₂ – B ₃

Com es pot veure, en els dissenys d'estructura simple es tendeix a introduir, sempre que es pugui, línies base entre les diferents fases de tractament. Això té sentit en el cas que se suposi que l'efecte de la intervenció experimental és o pot ser reversible. En aquest cas, si al retirar un tractament es produeix un retorn (total o parcial) al nivell de la línia base, tenim un element suplementari per pensar que la intervenció ha tingut impacte. A més, això permet comparar cada fase de tractament amb fases de línia base anteriors i posteriors.

Tot i que és habitual considerar que els dissenys de replicació intra-subjecte estructuralment són experimentals, es fa una excepció en el cas del disseny A-B, ja que és un disseny molt dèbil que no permet assegurar que qualsevol canvi sigui degut realment a la intervenció, i no a qualsevol altre factor. Per remarcar aquesta debilitat del disseny, és útil considerar-lo com a quasi-experimental, en oposició a l'estructura experimental de la resta de dissenys N=1.

Pel que fa als dissenys d'estructura complexa, hi ha dues variants principals:

Estructura complexa – Comparacions intra i entre sèries

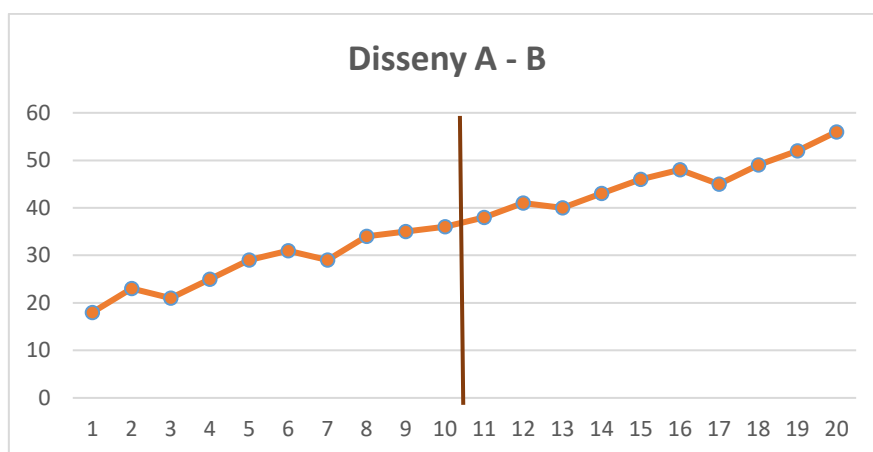
Disseny de línia base múltiple: Entre subjectes, entre conductes, entre situacions

Disseny de tractament altern

Els dissenys d'estructura complexa s'utilitzen especialment quan no és possible o desitjable introduir línies base intercalades entre les fases de tractament, quan se sospita que l'efecte de la intervenció pot ser irreversible, o en altres situacions especials.

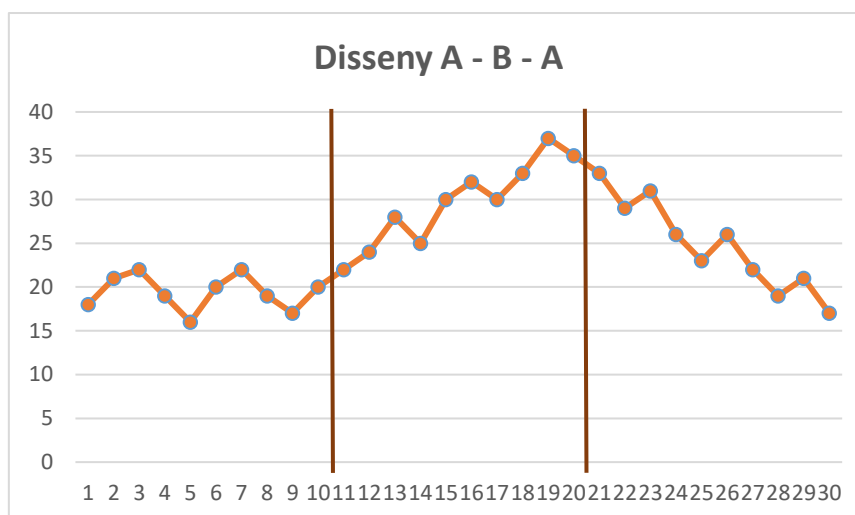
A continuació es farà una breu revisió dels principals tipus de dissenys $N=1$, indicant algunes característiques bàsiques i aportant algunes claus per a la interpretació dels resultats.

. Disseny amb comparacions intrasèrie o d'estructura simple



Com ja s'ha indicat, és el disseny $N=1$ més dèbil, i en general el seu ús no és recomanable. El problema principal és que qualsevol canvi que s'observi en els resultats no es pot atribuir de forma fiable al tractament, ja que poden existir variables de confusió que poden explicar totalment o parcialment aquest canvi. Per aquest motiu alguns autors consideren aquest disseny no com a experimental sinó com a "quasi-experimental", en el sentit de que li falten elements com per ser considerat pròpiament com a un disseny experimental.

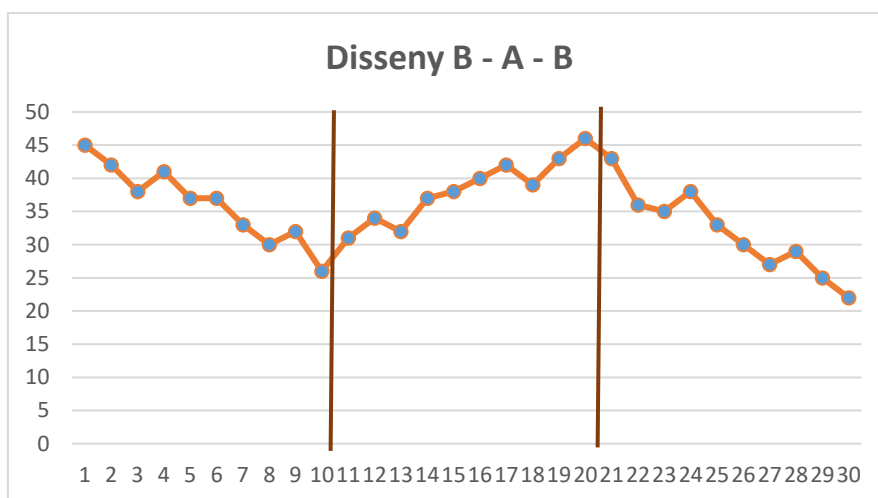
Aquest disseny pot ser útil per detectar un "no-efecte" del tractament. Per exemple, en el gràfic proposat es pot veure que la línia base presenta una tendència creixent, i que aquesta tendència no es modifica en absolut per la introducció del tractament. Per tant, es pot pensar que la intervenció implantada no té cap impacte sobre els resultats, ja que no s'observa cap modificació en el comportament dels resultats entre les dues fases.



El disseny A-B-A introdueix un element de control fonamental, que és la retirada del tractament, de manera que hi ha una segona fase de línia base, sense intervenció. Aquesta segona fase de línia base és especialment útil quan suposem que l'efecte del tractament experimental pot ser reversible, ja que llavors la seva retirada hauria de produir un retorn total o parcial al nivell inicial de resultats, és a dir, una *reversió de conducta*, com és el cas del gràfic.

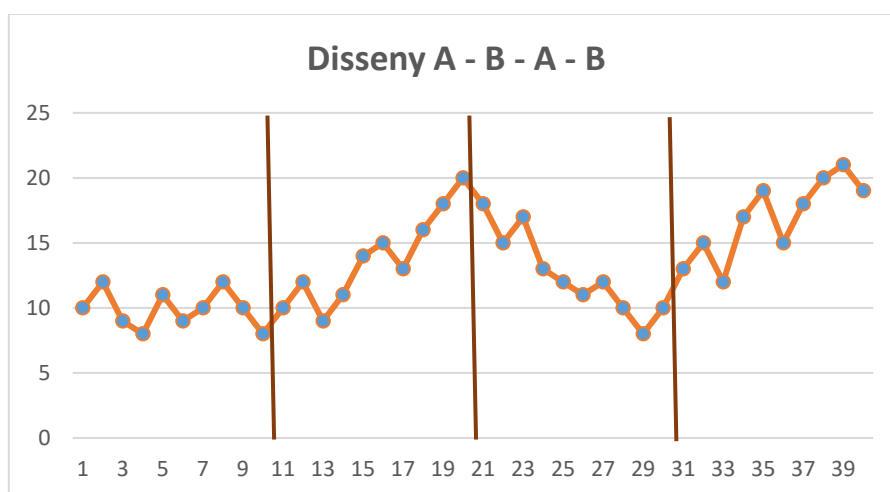
Si, per exemple, en un lloc de treball s'introdueix un canvi ambiental que produeix un increment de productivitat d'un treballador, la retirada d'aquest canvi podria tenir com a conseqüència un retorn al nivell de productivitat inicial. D'aquesta forma l'impacte de la intervenció queda mostrada per una doble via: El canvi produït amb la seva introducció i la reversió d'aquest canvi que es produeix amb la retirada del tractament.

Si al retirar la intervenció no es produeix una reversió de conducta, hi ha bàsicament dues interpretacions possibles: O bé l'efecte del tractament és irreversible, i per tant la seva retirada no produeix un retorn al nivell inicial; o bé el canvi produït entre les fases A i B no és deguda al tractament, sinó a qualsevol altre factor, i la posterior retirada del tractament no influeix en els resultats. De vegades pot no ser fàcil distingir entre les dues interpretacions, la clau està en fins a quin punt és versemblant que l'impacte d'un tractament sigui irreversible, i això depèn de la seva naturalesa. En general, una idea que pot ajudar és la d'allargar la segona línia base introduint més ocasions de registre, ja que a la llarga és probable que l'impacte de la majoria de tractaments tendeixi a disminuir amb el temps.



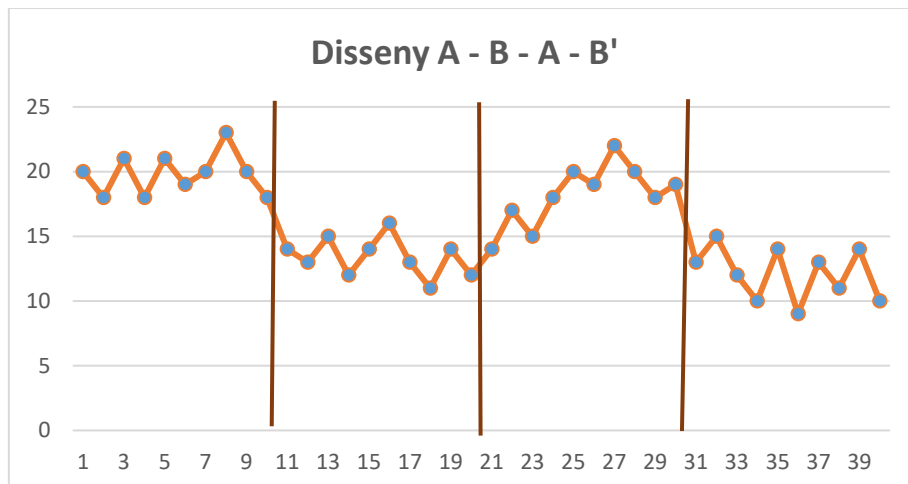
Aquest disseny s'utilitza quan per alguna raó (normalment d'urgència) no és possible o desitjable endarrerir l'aplicació de la intervenció experimental, i per tant no és factible obtenir una línia base inicial. Naturalment, això impedeix conèixer el nivell inicial de la variable dependent estudiada. La interposició d'una retirada del tractament pot ajudar, sempre que l'efecte de la intervenció sigui totalment o parcialment reversible, però a no ser que els resultats siguin molt clars, no és un disseny fàcil d'interpretar. Pot ajudar també la introducció d'una nova fase de línia base al final, plantejant un disseny B – A – B – A, per veure si el patró de canvi entre fases es repeteix.

A l'exemple exposat sí tenim uns resultats força aclaridors, ja que durant les dues fases de tractament s'observa una tendència decreixent en els registres de la variable dependent, mentre que quan es retira el tractament aquesta tendència canvia totalment i passa a ser creixent. Sovint els resultats no són tan clars i es poden produir problemes d'interpretació.

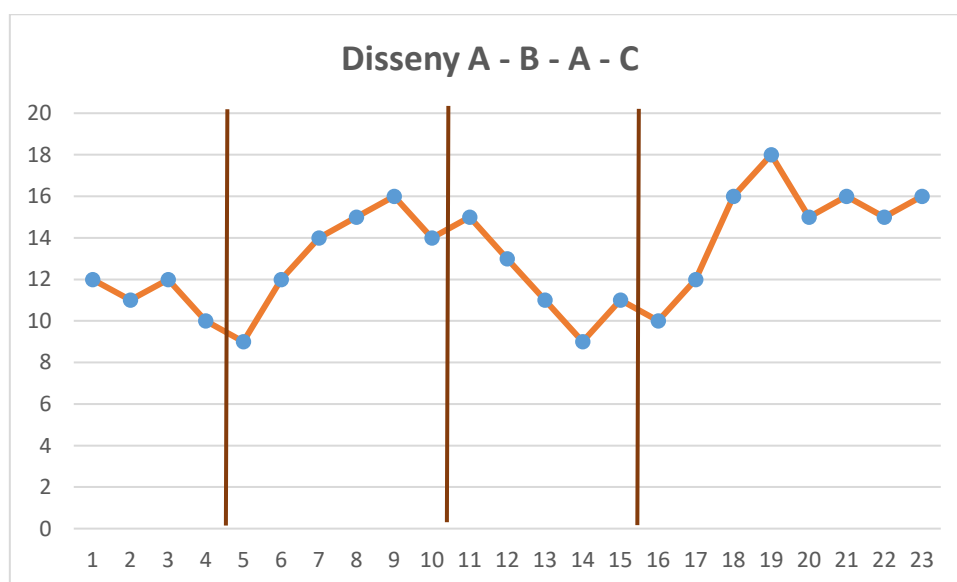


Es poden fer les mateixes consideracions que en el disseny A – B – A, amb l'element de reforç suplementari de comptar amb una segona introducció del tractament a la darrera fase, de forma que es pot valorar la possible repetició del patró de canvi observat. Per tant, és necessari

comparar no únicament les fases de línia base amb les d'intervenció, sinó comparar també les dues fases d'intervenció per veure si els resultats són similars o bé hi ha alguna diferència. Naturalment, si és possible i necessari el disseny es podria allargar amb noves fases de línia base i tractament (per exemple, A – B – A – B – A – B – A).

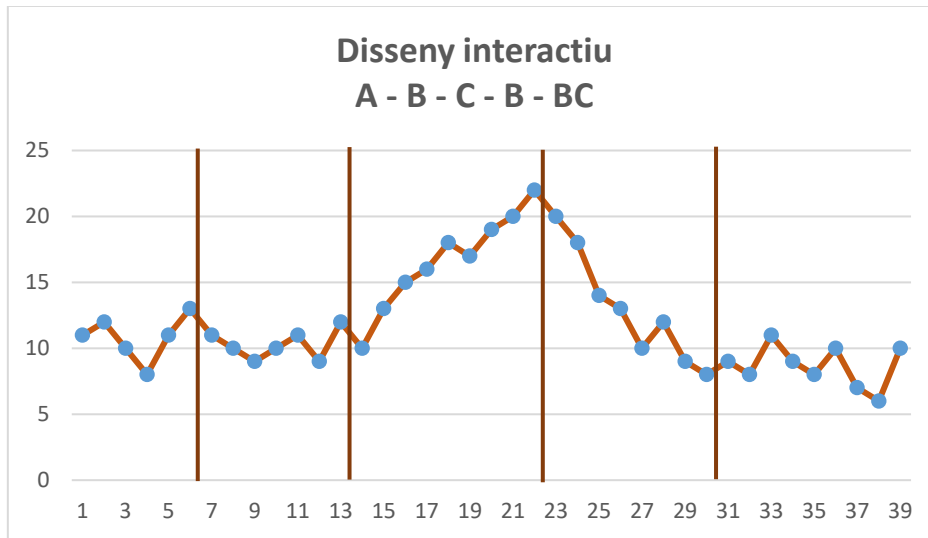


Quan s'introdueixen dos o més nivells quantitativament diferents d'una mateixa intervenció, torna a ser important comparar no solament els canvis que es puguin produir a l'introduir o al retirar cada intervenció, sinó també les diferents fases de tractament entre sí, per poder valorar si hi ha diferències en l'impacte dels diferents nivells d'intervenció (per exemple, si dues dosis diferents d'una determinada substància tenen o no un efecte diferencial sobre els resultats). Per tant, la comparació entre les fases B i B' és important, com també ho seria si hi haguessin altres nivells de tractament (B'', B''', etc.).



Es poden fer les mateixes consideracions que en el cas anterior, amb la diferència que aquí es comparen dues intervencions qualitativament diferents (B i C) i no dos nivells quantitativament

diferents de la mateixa intervenció. Naturalment, si és possible i necessari es pot ampliar el disseny introduint nous tractament (per exemple, en un disseny A – B – A – C – A – D). Com es pot veure, és convenient, sempre que sigui possible, intercalar línies base entre dues fases de tractament, per assegurar l'existència de prou elements de comparació.



Es caracteritza pel fet d'aplicar simultàniament més d'un tractament en alguna de les fases. La idea de "disseny interactiu" té una certa relació amb la idea de les interaccions que s'estudiaven als dissenys experimentals clàssics, (els dos tractaments poden potenciar-se mútuament, anul·lar-se, o modificar-se entre sí d'alguna manera), tot i que la manera de plantejar-ho i analitzar-ho és molt diferent. Aquí ens limitarem a estudiar la fase BC com una fase més, comparant-la amb els resultats de la resta de fases, sense cap particularitat especial.

Els resultats del gràfic mostren que el tractament B no sembla gens efectiu, mentre que el C produeix un canvi de tendència molt clar. La combinació BC tampoc sembla resultar eficaç.

Per poder analitzar correctament els resultats d'aquest disseny és important intentar seguir el criteri de fer un sol canvi per fase, ja que això facilita molt la interpretació. Per exemple, un disseny A – B – A – BC – A comportaria problemes, ja que després de la segona línia base s'introdueixen simultàniament els dos tractament, i després es retiren també tots dos de cop, i a més no disposem de cap fase C, de forma que no serà fàcil saber si qualsevol possible impacte que es produeixi a la fase BC és deguda a la combinatòria dels dos tractaments, o a l'efecte només d'un d'ells.

Seguint el criteri anterior, en el cas del gràfic proposat el disseny es podria millorar amb la següent estructura: A – B – A – C – A – B – BC, intercalant línies base entre totes les fases de tractament, tot i que lògicament això complicaria l'estudi. Si no es fa així, cal pensar que al passar directament, per exemple, de la fase B a la fase C, en realitat es fan dues coses: retirar el tractament B i introduir el C, cosa que segons el patró de resultats que s'obtingui podria complicar la interpretació.

Hi un cas especial dins dels dissenys d'estructura simple, que és el disseny de canvi de criteri. De forma resumida, les seves característiques són:

Disseny de canvi de criteri

- . Correspon a la tècnica d'aproximacions successives per arribar a un objectiu
- . Increment progressiu dels requeriments conductuals necessaris per obtenir un incentiu o conseqüència constant
- . És diferent del disseny amb diferents intensitats del tractament (A – B – B' – B''...), tot i que es podria considerar com un cas especial d'aquest tipus de disseny, però amb característiques pròpies
- . És freqüent no introduir línies base intercalades per afavorir l'eficàcia de la intervenció i progressar en l'assoliment progressiu de l'objectiu
- . Possible representació: A – B₁ – B₂ – B₃ – B₄

El disseny de canvi de criteri es basa en la tècnica d'aproximacions successives. S'estableix un objectiu conductual concret; per exemple, reduir un 80% el nombre de faltes d'ortografia que fa un nen a les seves redaccions, en un cas on els errors s'atribueixen sobretot a distraccions, no a un desconeixement de les regles ortogràfiques. Per arribar a aquest objectiu, s'utilitza algun tipus d'actuació (per exemple, un reforçament positiu o negatiu) per anar assolint objectius intermedis. Suposem que el nen ha de fer una redacció diària a l'escola. A la primera setmana es registra el nivell inicial d'errors ortogràfics a la redacció diària, sense intervenció. A la segona setmana, el nen pot obtenir un punt de bonificació (bescanviable per determinats beneficis escolars) si comet menys de 15 errors. A la tercera setmana el punt s'aconsegueix només si comet menys de 10 errors, i a la quarta setmana si en comet menys de 5. D'aquesta forma, es va "modelant" el comportament fins arribar a l'objectiu establert. Per això la representació del disseny és diferent a la dels anteriors, i s'utilitza la representació A – B₁ – B₂ – B₃, etc. on B₁, B₂, etc. indiquen el criteri utilitzat a cada fase.

Tot i que en principi es podrien introduir fases de línia base entre les fases de tractament, en molts casos de canvi de criteri això no és gaire convenient, ja que podria interferir amb la dinàmica de les aproximacions successives a un objectiu, i posar en risc l'assoliment dels resultats desitjats.

Cal fer notar que la intervenció (en aquest cas un reforç positiu concret) es manté constant al llarg de tot el procés (no es modifica el número de punts obtinguts, ni s'utilitza cap altre incentiu), el que canvia és l'exigència conductual que es fa a canvi. Per tant, no es pot considerar pròpiament com un disseny amb diferents nivells de tractament (B, B', B'', etc.), tot i que també es podria entendre com un cas especial d'aquest tipus de disseny, en el qual l'aplicació del tractament (que és constant) es modifica en funció dels objectius conductuals establerts. Les seves característiques particulars fan que aquest disseny tingui una denominació i una simbologia pròpies.

A diferents camps de treball no és infreqüent l'ús de la idea d'aproximació successiva a un objectiu. A l'àmbit clínic, per exemple, es pot utilitzar com a part del tractament de les fòbies, de forma que el temps que el pacient pot passar davant de l'objecte fòbic es va incrementant

progressivament, tot i que caldria veure en cada cas concret si s'ha utilitzat realment un disseny de canvi de criteri o no.

. Disseny amb comparacions intra i entresèries, o d'estructura complexa

Tot i que existeix alguna altra variant, molt poc utilitzada, aquí només es tractaran els dos casos principals: el disseny de línia base múltiple (sense dubte el més freqüent dins d'aquesta categoria), i el disseny de tractaments alterns o de programa múltiple.

Cal remarcar novament que, a diferència dels dissenys d'estructura simple, en els d'estructura complexa s'assumeix moltes vegades que l'efecte dels tractaments pot ser irreversible i, ja sigui per aquesta raó o per altres de conveniència, es renuncia a la introducció de línies base intercalades entre les fases de tractament (tot i que, com es veurà, pot haver-hi alguna excepció en el cas del disseny de tractaments alterns).

Esquemàticament, les idees principals relacionades amb el **disseny de línia base múltiple** són les següents:

Disseny de línia base múltiple
<ul style="list-style-type: none"> ➤ Entre conductes ➤ Entre situacions ➤ Entre subjectes
<p>Basat en la idea de replicacions asincròniques: Intervenció successiva sobre diversos subjectes, conductes o situacions, de forma que la comparació entre elles estableix un element de control</p>
<p>Cal prendre algunes precaucions: conductes independents i sensibles, subjectes sense interacció i comparables, situacions mínimament equiparables</p>

Com s'indica a l'esquema general, la idea fonamental del disseny de línia base múltiple consisteix inicialment en prendre línies base diferents, que poden correspondre a diverses persones, a la mateixa persona valorant diversos aspectes comportamentals, o també a una mateixa persona en situacions diferents. A continuació s'introdueix la intervenció, però no de forma simultània sinó en moments diferents per a cadascun dels subjectes, comportaments o situacions. La nostra expectativa és que els elements que siguin tractats mostrin algun tipus de canvi, mentre que els altres no es modificarien.

Exemples:

. Disseny de línia base múltiple entre conductes. Es volen millorar els hàbits d'una nena en relació a tres aspectes diferents: temps que tarda a vestir-se, temps que dedica a rentar-se les dents i temps que tarda en menjar a migdia. S'estableix un sistema d'economia de fitxes que s'aplica als tres comportaments citats. Inicialment es prenen registres dels tres comportaments sense cap intervenció. Després, es tracta només el temps per vestir-se (els altres dos elements no són tractats), en segon lloc es comença a tractar el temps per rentar-se les dents (el darrer hàbil encara no es tracta, el primer es continua tractant) i finalment es comença a tractar el temps dedicat a menjar (els tres hàbits són tractats).

Esquemàticament:

	Fase 1	Fase 2	Fase 3	Fase 4
	Tractament	Tractament	Tractament	Tractament
Conducta 1	NO	SI	SI	SI
Conducta 2	NO	NO	SI	SI
Conducta 3	NO	NO	NO	SI

. Disseny de línia base múltiple entre subjectes: Es vol millorar un dels hàbits citats al cas anterior (per exemple, el temps dedicat a menjar) en tres nenes que dinen a la mateixa escola. S'intervé primer amb una d'elles (mitjançant el programa d'economia de fitxes) sense coneixement de les altres dues, després s'introdueix la intervenció a la segona nena participant, i finalment a la tercera.

	Fase 1	Fase 2	Fase 3	Fase 4
	Tractament	Tractament	Tractament	Tractament
Subjecte 1	NO	SI	SI	SI
Subjecte 2	NO	NO	SI	SI
Subjecte 3	NO	NO	NO	SI

. Disseny de línia base múltiple entre situacions: Per a una mateixa nena i un únic comportament (temps dedicat a menjar a migdia), es defineixen dues situacions diferents: dinar a casa o dinar a l'escola. S'introdueix el tractament en una de les situacions però no a l'altra, i posteriorment s'introdueix per a la segona situació.

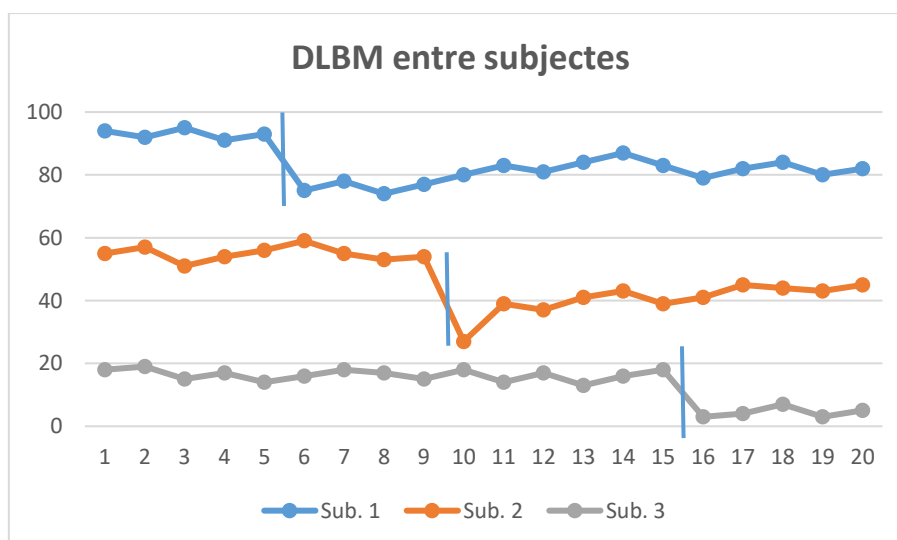
	Fase 1	Fase 2	Fase 3
	Tractament	Tractament	Tractament
Situació 1	NO	SI	SI
Situació 2	NO	NO	SI

Malgrat les diferències, hi ha dos principis comuns a les tres variants:

. El primer és que els components no tractats (siguin altres subjectes, altres conductes o altres situacions) serveixen com a control o element de comparació, de forma que podem trobar que el component tractat mostra un canvi, mentre que els altres es mantenen sense modificacions. Això permet incrementar la seguretat respecte de que els canvis registrats són deguts realment a la intervenció experimental, i no a qualsevol altre factor.

. El segon és que, per tal d'exercir aquesta funció de comparació, cal prendre determinades precaucions, que depenen de cada cas. Per exemple, en el cas entre subjectes és important que les diferents persones participants no estiguin en contacte o com a mínim no coneguin la situació de les altres, per evitar contaminar els resultats. En el cas entre conductes, els diferents comportaments han de ser independents entre sí per evitar interferències (l'actuació sobre un dels comportaments no ha de produir canvis en els altres), cosa que podria ser dubtosa a l'exemple plantejat. En el cas del disseny entre situacions, els diferents escenaris no han de ser radicalment diferents, però també cal assegurar-se que el que passa en una situació no influeix de forma important en el que es produeix a l'altra, cosa que no sempre resulta fàcil.

Els resultats esperables en un disseny de línia base múltiple, en cas que el tractament sigui eficaç, serien els següent (s'utilitza l'exemple del disseny de línia base múltiple entre subjectes amb tres participants que s'ha citat anteriorment):



En aquest cas, el fet que les tres participants hagin mostrat un canvi en la direcció esperada (reducció del temps dedicat a menjar) i que aquest canvi s'hagi produït a partir de la implantació del tractament, i no abans, reforça la idea de que la intervenció ha estat eficaç. Naturalment, en molts casos reals els resultats no són tan evidents com els que es mostren en aquest gràfic, i necessiten d'interpretació acurada.

Pel que fa al **disseny de tractaments alterns o programa múltiple**, les idees principals són:

Disseny de tractaments alterns
<ul style="list-style-type: none"> ➤ Alternància ràpida de dos o més tractaments, habitualment en fases de duració molt curta (1 o 2 registres) ➤ Sovint, aplicació aleatòria dels tractaments a cada sessió de registre <p>Exemple: B – C – B – B – C – B - C – C – C – B – B – C.....</p> <p>Es poden introduir també sessions de línia base: B – A – B – C – C – A – B – A – A – C – B – B ...</p>
<ul style="list-style-type: none"> ➤ Ofereix un bon control d'alguns factors de confusió lligats al temps ➤ Atenció a la interferència de tractaments múltiples

Aquest disseny es pot utilitzar únicament quan es disposa de dos o més tractaments que puguin ser implantats i retirats de forma ràpida i senzilla, en fases molt curtes, de pocs registres. Per aquest motiu no es pot aplicar a tractaments complexos (per exemple, comparar dues teràpies psicològiques), però sí a estímuls ambientals o situacionals que es puguin modificar fàcilment.

Per exemple, podem intentar esbrinar si el resultat d'una sessió de teràpia de relaxament per reduir la tensió muscular d'un determinat pacient dona millors resultats si el terapeuta és un home o una dona, si se sospita que aquest pot ser un aspecte rellevant. En aquest cas, a les diferents sessions s'anirien alternant aleatòriament els dos terapeutes, per valorar si hi ha diferències en els resultats.

En aquest exemple segurament seria un error utilitzar un esquema més tradicional. Es podria pensar, per exemple, en un disseny A – B – A – C (registre sense tractament, fase de registre amb terapeuta home, registre sense tractament, fase de registre amb terapeuta dona). En aquest cas el tractament aplicat és el mateix, però la distinció entre B i C es fa perquè el que es vol veure és si el gènere del terapeuta té influència en els resultats. El problema d'aquest plantejament és que el tractament aplicat podria tenir un efecte progressiu, de manera que els resultats a la fase C podrien ser millors que a la fase B, però no perquè sigui millor treballar amb una terapeuta femenina, sinó perquè hi ha un efecte positiu acumulatiu de l'aplicació del tractament. En canvi, en el disseny de tractaments alterns això no es pot produir, per tant es pot dir que aquest tipus de disseny permet controlar alguns factors de confusió lligats amb el temps.

Naturalment, una condició per a l'ús d'aquest disseny és que els diferents tractaments no s'interfereixin entre sí, és a dir, que els resultats obtinguts a les sessions B no es vegin afectats pels obtinguts a les sessions C, i viceversa. A l'exemple citat això sembla molt clar, però en altres casos, en funció de la naturalesa dels tractaments, això podria arribar a ser un problema.

4.- Anàlisi de dades en dissenys N = 1

Com passava en part en el cas dels dissenys quasi-experimentals de sèries temporals, l'anàlisi dels resultats en un disseny de replicació intrasubjecte es basa en dos elements bàsics:

- a) La inspecció visual dels gràfics de resultats, a partir d'un conjunt de criteris de valoració
- b) Anàlisi estadística de sèries temporals: Aplicació dels models de regressió lineal, anàlisi multinivell
- c) Ús de tècniques pensades específicament per als dissenys N = 1, diferents de les pròpies de l'estadística clàssica. Es distingirà entre dos grans tipus de tècniques: Numèriques i gràfiques

a) Inspecció visual de gràfics

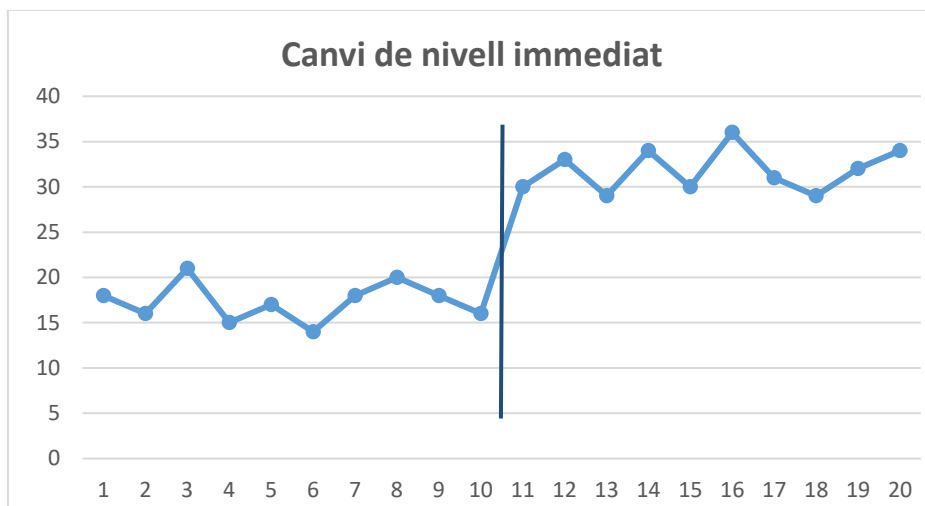
Els aspectes més importants a tenir en compte a l'hora de valorar un gràfic de resultats són els següents:

Inspecció visual de gràfics: Criteris principals
<ul style="list-style-type: none">➤ Estabilitat de la línia base➤ Variabilitat intrafases➤ Variabilitat entre fases➤ Grau de solapament entre fases adjacents➤ Canvis de tendència➤ Canvis de nivell➤ Comparació entre fases de tipus diferent (línia base- fases de tractament, diferents tractaments, etc.)➤ Comparació entre fases similars

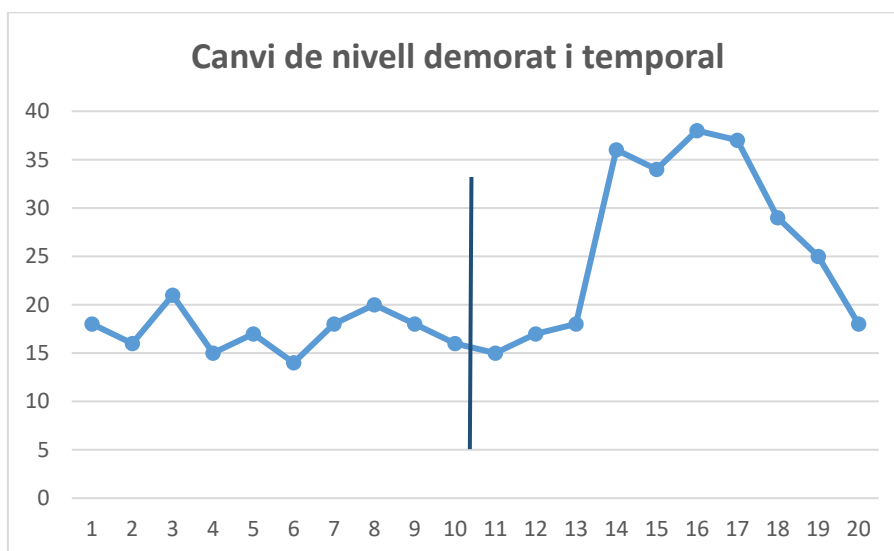
Com es pot veure, la inspecció visual de gràfiques en un disseny N = 1 (i en altres dissenys longitudinals) es basa en dos grans elements:

- a) La valoració del comportament de les dades dins de cada fase, bàsicament la seva estabilitat i la seva estacionarietat, tant a les línies base com a les fases d'intervenció
- b) La comparació dels resultats entre unes fases i altres, que es pot referir a diferents coses:
 - La comparació de fases de línia base amb fases de tractament
 - La comparació de tractaments diferents (per exemple, en un disseny A-B-A-C té tot el sentit comparar les fases B i C per veure si l'impacte dels dos tractaments és o no similar), o bé entre nivells diferents del mateix tractament (B, B', B'', etc.) i altres possibilitats
 - La comparació entre fases del mateix tipus: Per exemple, en un disseny A-B-A les dues línies base mostren resultats similars?, és a dir, hi ha reversió de conducta?; o en un disseny A-B-A-B, les dues vegades que s'implanta el mateix tractament es produeix el mateix canvi?, etc.

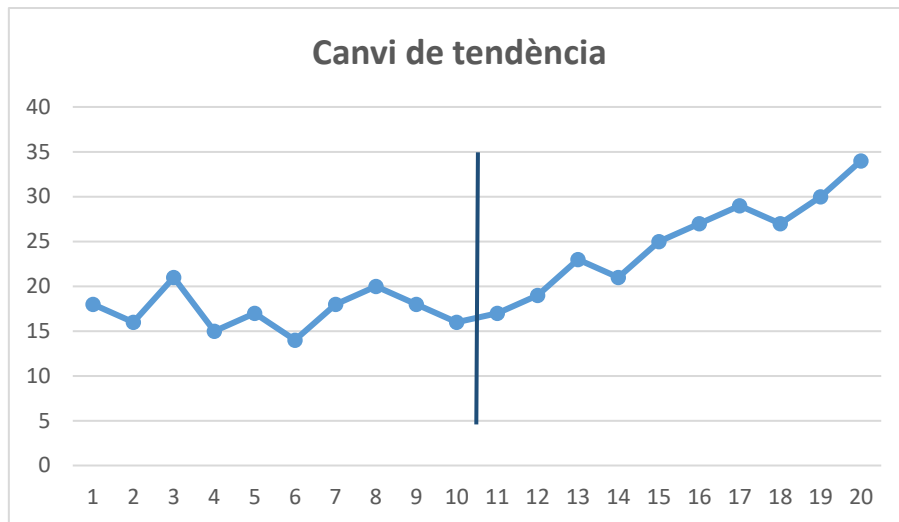
La comparació entre diverses fases (siguin del mateix tipus o diferents) es basa fonamentalment en l'anàlisi dels *patrons de canvi* que s'observin entre una fase i una altra. Els patrons de canvi més habituals són els següents (suposarem en tots els casos que la primera fase és de línia base i la segona d'intervenció):



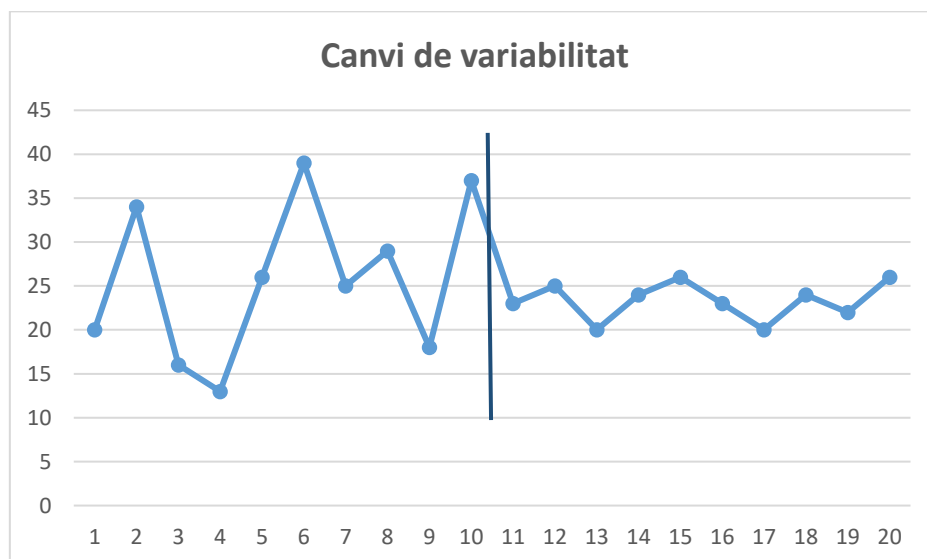
El canvi de nivell implica un “salt”, una modificació que porta de forma ràpida i sobtada els resultats de la segona fase a un nivell més alt o més baix que els de la primera. El canvi de nivell pot ser immediat (com en el cas del gràfic), és a dir, produir-se ràpidament una vegada implantada la intervenció, o bé demorat, amb un cert retard respecte del moment de la intervenció. El canvi pot ser permanent (és a dir, mantenir-se durant tota la fase, com és el cas del gràfic), però també temporal, és a dir, retornar totalment o parcialment al nivell basal, tot i estar aplicant-se la intervenció. Per exemple:



Com es pot veure, en aquest cas el canvi de nivell no es produeix immediatament quan s'introdueix el tractament (és un canvi demorat), i cap al final de la segona fase es produeix una reversió cap als nivells inicials.



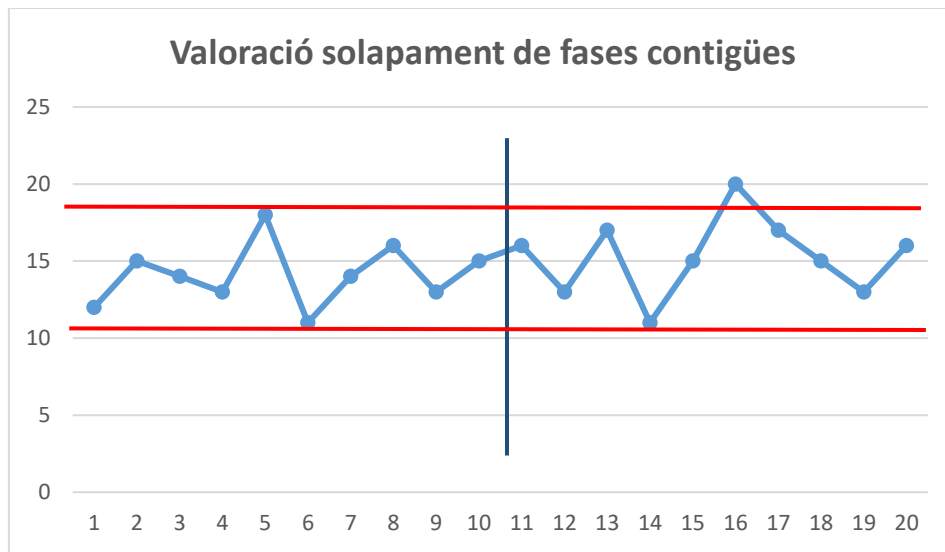
A diferència del canvi de nivell, el canvi de tendència és progressiu, de manera que la modificació es produeix de forma menys ràpida. Com passava en el canvi de nivell, el canvi de tendència pot ser immediat (com és el cas del gràfic) o bé demorat, i igualment pot ser permanent o temporal.



Com és evident, en aquest cas el que es produeix és una modificació important en el grau de variabilitat de les dades al canviar de fase. En el gràfic proposat, la variabilitat dels resultats disminueix de forma molt important a l'introduir el tractament, sense que sigui possible (per la inestabilitat de la línia base) detectar amb claredat altres patrons de canvi.

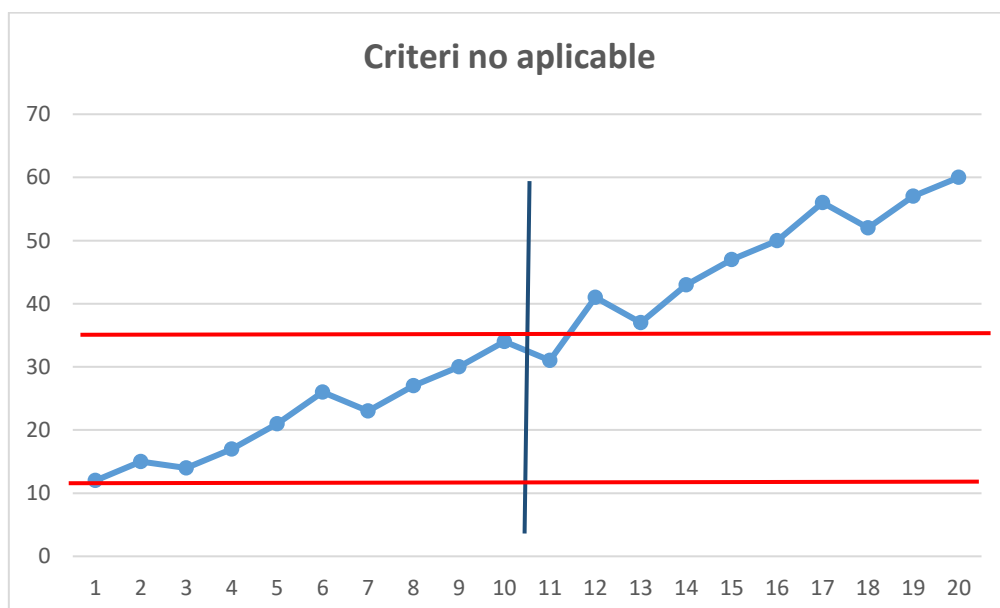
A més dels tres patrons de canvi assenyalats, un element interessant per a la inspecció visual de gràfics és el concepte de *solapament de fases contigües*. Això es refereix al grau en el qual les dades d'una determinada fase es mouen dins dels límits marcats per la fase anterior.

Per exemple:



Com es pot observar, s’han traçat dues línies horitzontals (en vermell), la de dalt passant pels registres més alts de la primera fase, i la de sota passant per les puntuacions més baixes. Les dues línies es perllonguen a la segona fase. Es pot veure que totes les puntuacions de la segona fase excepte una es mantenen dins de les línies marcades per la primera. Per tant, el grau de solapament és important, i es pot pensar que no hi ha un canvi important d’una fase a l’altra.

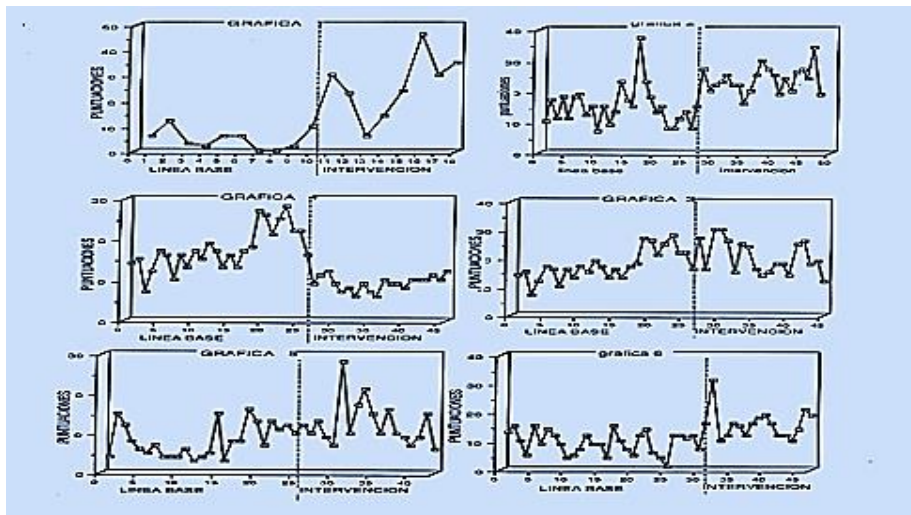
La valoració del nivell de solapament entre fases contigües té algunes limitacions evidents. La més òbvia és que no es pot utilitzar en molts casos en els quals la línia base no es estacionària, ja que pot portar a greus errors d’interpretació. Un exemple molt evident és el cas següent:



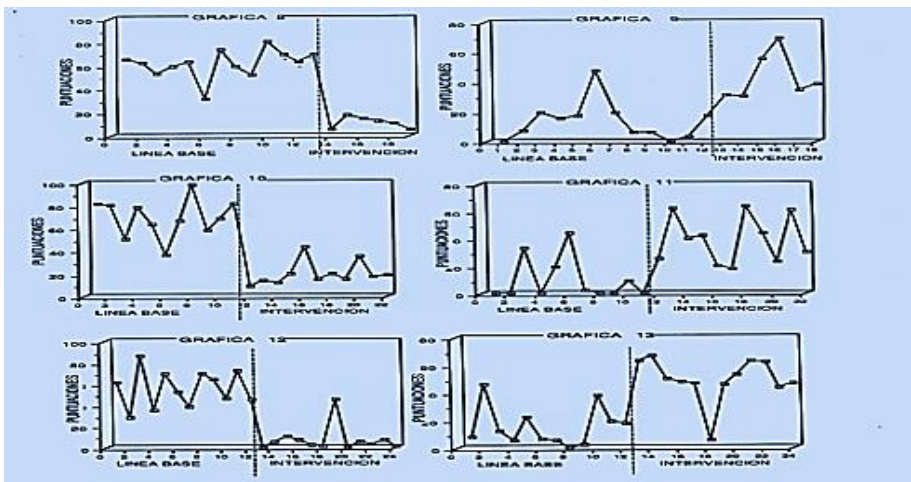
Si, en contra de la lògica i el sentit comú, apliquem a aquest cas la lògica del solapament entre fases contigües, caldria pensar que hi ha un impacte molt important per efecte del tractament, ja que gairebé totes les puntuacions de la segona fase estan fora dels límits marcats per la primera. Però la simple inspecció del gràfic ens indica que això no és així, ja que hi ha una tendència creixent que explica aquest fet i que no es trenca amb la introducció del tractament. Per tant, en un cas com aquest l'aplicació del criteri de solapament o no entre les puntuacions de les dues fases és del tot inadequat.

Els exemples plantejats presenten patrons de resultats força clars. En situacions reals, però, és freqüent que els resultats obtinguts siguin força més difícils d'interpretar, com a mínim en alguns casos. Es presenten a continuació, a títol il·lustratiu, alguns gràfics reals de resultats, on es pot apreciar la diversitat de patrons de resultats que es poden trobar. Per simplificar, es presenten només gràfics amb dues fases, la primera de línia base i la segona de tractament, encara que en molts casos puguin formar part de dissenys més amplis, amb més fases:

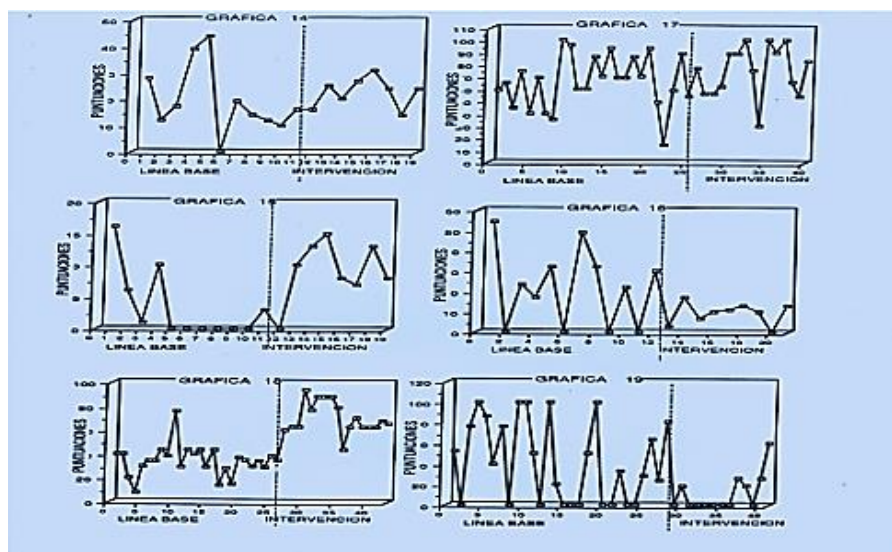
Gràfics corresponents a casos reals (1):



Gràfics corresponents a casos reals (2):



Gràfics corresponents a casos reals (3)



És molt important indicar que la inspecció visual de gràfics no es fa normalment de forma purament mecànica, aplicant simplement els criteris anteriors, sinó que ha de venir acompanyada per una interpretació adient a cada context. Per exemple, en el tractament d'una patologia s'han de posar en joc criteris de caràcter clínic i de vegades social, sovint qualitius (per exemple, millora auto-percebuda pel propi subjecte o percebuda pel seu entorn immediat, possible millora en la dinàmica de l'entorn familiar o social, etc.); o en altres casos es poden valorar aspectes com la millora qualitativa en l'adaptació del subjecte, el benestar subjectiu, o altres.

b) Anàlisi estadística de sèries temporals

Com és evident, la inspecció visual de gràfiques té les seves limitacions, especialment en els casos on el patró de resultats no és clar. Per aquest motiu és important disposar d'altres tècniques que es puguin utilitzar amb caràcter complementari per aconseguir una interpretació correcta dels resultats.

Cal indicar que l'aplicació de tècniques d'anàlisi a les dades dels dissenys $N=1$ és objecte d'investigació intensa en els darrers anys, i que es van introduint progressivament noves propostes. Considerant el caràcter introductor d'aquest text, no es farà en absolut una anàlisi exhaustiva de les diferents possibilitats, però cal indicar que hi ha un desenvolupament important d'idees relacionades amb camps com l'anàlisi multinivell o l'ampliació dels àmbits d'aplicació de les proves d'aleatorització, entre d'altres.

Cal dir en primer lloc que la majoria de les tècniques estadístiques clàssiques no són generalment útils en els dissenys $N=1$ (exceptuant els models de regressió lineal i tècniques relacionades), per raons tècniques i conceptuals. S'han fet algunes propostes per utilitzar tècniques estadístiques convencionals (t d'Student, anàlisi de la variància) per analitzar dades de dissenys de replicació intra-subjecte. És fàcil veure, però, que es produeix un incompliment evident d'algunes de les condicions d'ús d'aquestes proves, especialment pel que fa a la condició d'independència dels registres, al ser generats tots ells per un mateix subjecte experimental. Per

una altra, els habituals índexs de centralització utilitzats a l'estadística convencional (per exemple, la mitjana aritmètica), perden una gran part del seu poder representatiu en situacions de gran variabilitat i poden portar a greus problemes quan es produeixen canvis de tendència o altres dinàmiques de canvi. És obvi que la reducció de les dades a termes estrictes de mitjanes, variàncies o altres estadístics convencionals significa deixar de banda la informació referent al caràcter dinàmic de la sèrie de dades (canvis immediats o demorats, tendències, etc.).

Alguns autors han proposat modificacions de les proves estadístiques convencionals (sobretot de l'anàlisi de la variància) per adaptar-les a les característiques de les dades dels dissenys $N = 1$, però en general els resultats no han estat convincents i l'ús d'aquestes proves és mínim.

Una excepció important a aquesta situació es troba en els models de regressió lineal, que es poden utilitzar, amb algunes limitacions o modificacions, per a la modelització i valoració de les dades en dissenys $N = 1$. Això dona lloc al que sovint s'anomenen *models de sèries temporals*.

Com ja és conegut, els models de regressió permeten representar la relació entre diferents variables a partir de models matemàtics relativament senzills. Suposem, per exemple, que treballem amb un disseny $N=1$ on apliquem només un tractament B, i que tenim algunes fases de línia base (A) i algunes fases de tractament (B). Una forma senzilla de representar els resultats en un model de regressió lineal seria:

$$y_t = \beta_0 + \beta_1 X + \varepsilon_t$$

En aquesta expressió, y_t és la puntuació obtinguda en el registre t , β_0 és una constant, X representa la presència o no del tractament ($X=0$ si estem en línia base, $X=1$ si estem en fase de tractament), β_1 és un coeficient aplicat a X , i ε_t és el terme d'error, és a dir, aquella part de la puntuació que no és explicada per la resta de components del model. A partir de les dades és possible estimar els valors de β_0 i de β_1 i, posteriorment, valorar si l'impacte del tractament és o no efectiu, a partir de proves de significació convencionals. Aquest seria doncs un model de regressió lineal estàndard, amb l'única diferència de que la variable X no és quantitativa sinó qualitativa i dicotòmica, tot i que pot ser introduïda en el model amb la codificació senzilla que s'ha explicat.

Com s'ha indicat abans, un dels principals problemes de les dades en els dissenys $N=1$ és que molt probablement, considerant que tots els resultats provenen del mateix subjecte, els resultats estaran correlacionats, és a dir, no seran independents. Això voldria dir, per exemple, que el resultat en un registre t potser no seria independent de l'obtingut en el registre $(t-1)$ o anteriors.

El que plantegen els models de sèries temporals és utilitzar els models estàndard de regressió, però incorporant al model la possible correlació que presentin les dades. Imaginem, per exemple, que trobem que en una sèrie temporal existeix una correlació entre les puntuacions t i les puntuacions $(t-1)$, és a dir, una auto-correlació de retard 1. Una forma senzilla d'abordar aquest problema seria modificar el model anterior de la forma següent:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 X + \varepsilon_t$$

Com es pot veure, en aquest model es pressuposa que la puntuació obtinguda en el registre t és funció de l'aplicació o no de tractament (valor de X), però també del resultat del registre immediatament anterior $(t-1)$. D'aquesta manera, la no independència dels registres és

incorporada al model i en principi es podria valorar correctament si el tractament ha estat eficaç o no.

Aquest model és anomenat model auto-regressiu d'ordre 1, o bé AR(1), ja que parteix de la base que, al marge del tractament, l'única influència que rep una puntuació és la de la puntuació immediatament anterior. Naturalment, aquest és només el cas més senzill de model de sèries temporals. Moltes vegades és necessari incloure altres elements en el model, representant auto-regressions d'ordre superior, o fins i tot altres tipus de components (mitjanes mòbils), que no seran tractats aquí. En tot cas, aquesta família de models de sèries temporals són anomenats models ARIMA (models auto-regressius integrats i de mitjanes mòbils). El model AR(1) és només el cas més senzill d'aquest conjunt de possibles models.

Tot i que aquesta és una bona aproximació, cal dir que no deixa de tenir problemes. El més important és que l'aplicació correcta del procediment estàndard per estimar un model ARIMA necessita d'un número de registres important. Normalment es cita com a xifra ideal la de 50 registres o més a cada fase, tot i que sovint s'admet l'ús d'aquests models amb xifres inferiors, entre 20 i 50 puntuacions. Tot i així, la literatura aplicada ens indica que en molts casos aquestes xifres no s'assoleixen, de manera que l'ús d'aquest tipus de models seria dubtós. Davant d'aquest problema s'han proposat algunes correccions o alternatives als models ARIMA, tot i que no hi ha un consens complet a la literatura sobre quines són les millors opcions.

Una alternativa rellevant als models de regressió és l'anàlisi multinivell, útil quan es treballa amb més d'un subjecte experimental (encara que cadascun s'hagi tractat individualment), i que permet estimar un efecte mitjà del tractament experimental per al conjunt dels subjectes. Tot i que per la seva complexitat no s'exposarà aquí el seu funcionament, aquesta és una alternativa que pot ser profitosa quan, per exemple, s'aplica un mateix tractament a un cert número de subjectes i es vol fer una valoració del seu possible efecte global i de la magnitud del seu efecte per a cadascun dels subjectes i les possibles diferències entre ells.

Una altra línia de recerca interessant en aquest àmbit es relaciona amb el desenvolupament d'índexs de magnitud de l'efecte (per exemple, l'índex de Hedges), amb una idea semblant a la que es planteja en els dissenys experimentals clàssics. Això és especialment útil quan es volen comparar els resultats de diferents estudis.

c) Tècniques específiques per a dissenys de replicació intrasubjecte

Les tècniques comentades a l'apartat anterior són de caràcter relativament general, i s'han plantejat inicialment fora de l'àmbit dels dissenys de replicació intra-subjecte. Hi ha també un bon número de tècniques més específiques, sovint pensades de forma especial per al seu ús en els dissenys $N = 1$. Vista la diversitat de propostes existent, aquestes tècniques específiques s'han classificat en dues categories: Tècniques numèriques (basades en una anàlisi numèrica de les dades) i tècniques gràfiques (basades en algun tipus de valoració o anàlisi dels gràfics de resultats, més enllà de la simple inspecció). Es presentarà només un llistat general (i no complet) de cadascuna de les dues categories i un exemple concret d'una tècnica de cada tipus, a mode d'il·lustració.

Esquemàticament:

Algunes tècniques específiques N=1

Tècniques numèriques

- . Proves d'aleatorització (Eddington, Revusky, altres)
- . Dependència serial (Dufour)
- . Tendències i solapament de fases (tau de Kendall, prova de rècords, etc.)
- . Altres, de vegades multiús (Estadístic C, percentatges de canvi, proves de signes, etc.)

Tècniques gràfiques

- . Tendències lineals: Divisió en meitats (Split-middle)
- . Tendències no lineals: Mediane mòbils
- . Variabilitat: Línies i gràfics de rangs

. Tècniques numèriques

Un conjunt de tècniques numèriques força interessant és el que es basa en la idea d'aleatorització. Aquest tipus de proves es poden aplicar quan s'ha pogut aleatoritzar algun dels aspectes del disseny. Per exemple, en un disseny de línia base múltiple es pot aleatoritzar el moment en el qual s'introdueix la intervenció en cadascun dels subjectes, conductes o situacions, i seria d'aplicació una prova com la de Revusky o altres variants. En dissenys d'estructura simple amb un número gran de fases (per exemple, A-B-A-B-A) es podria determinar a l'atzar el moment d'inici de cada fase. Probablement un dels casos més evidents de possible aplicació d'aquestes proves és el disseny de tractaments alterns, on el tractament que s'aplica a cada sessió es decideix habitualment de forma aleatòria.

Per entendre el fonament d'aquest tipus de proves, es planteja com a exemple la prova d'aleatorització d'Edgington per a dissenys de tractament altern o programa múltiple. La idea bàsica de la prova d'aleatorització es basa en el raonament estadístic clàssic: Si tenim dos tractaments (B i C), la hipòtesi nul·la "clàssica" seria que no hi ha diferència en els resultats obtinguts sota cadascun d'ells. Si això fos cert, el fet que a una determinada sessió s'apliqui el tractament B o el tractament C seria indiferent, ja que els resultats serien semblants.

Suposem, per exemple, que s'utilitza un disseny de tractaments alterns amb 7 sessions de registre, organitzades de la forma següent: C – B – B – C – B – B – C, i que els resultats obtinguts per a la variable dependent a cada sessió són 13, 10, 8, 15, 12, 10, 14.

Suposem que es fa la mitjana dels resultats de totes les sessions B i de totes les sessions C. El resultat seria: $\bar{y}_b = ((10+8+12+10)/4=10$; $\bar{y}_c=(13+15+14)/3= 14$. Per tant, la diferència C-B seria $14 - 10 = 4$ (naturalment es podria usar igualment la diferència B – C, que seria de -4)). La pregunta a fer-se és: Aquesta diferència de 4 punts en els resultats és producte purament de l'atzar (hipòtesi nul·la) o bé existeix una diferència real en l'impacte dels dos tractaments (hipòtesi alternativa)?.

El que fa la prova d'aleatorització és calcular quina seria la diferència obtinguda per a totes les combinacions possibles de distribució dels tractaments a les sessions. Mantenint la idea de que hi hagi 7 sessions, i que 4 d'elles siguin B i 3 siguin C, el número de combinacions possibles seria: $7! / 4! 3! = 35$. Si calculem la diferència entre C i B que s'obtingria en cada cas, es construeix una taula com la següent (es remarca el resultat de la combinació que s'ha utilitzat realment: : C – B – B – C – B – B – C):

Combinacions possibles	Diferència $\bar{y}_c - \bar{y}_b$
B-B-B-B-C-C-C	0,5
B-B-B-C-B-C-C	2,25
B-B-B-C-C-B-C	3,42
B-B-B-C-C-C-B	1,08
.....	
C-B-B-C-B-B-C (combinació real)	4
.....	
C-C-C-B-B-B-B	-2,42

Després caldrà ordenar les 35 diferències de major a menor, mantenint el signe de la diferència. El resultat d'això seria: 4, 3.42, 2.83, 2.25,, -2.42, -3, -4.16.

Es pot observar que la diferència corresponent a la combinació utilitzada realment és igual a 4, i que ocupa una posició extrema a la llista ordenada de diferències, és la diferència positiva més gran dins de les obtingudes. Quina probabilitat hi ha de que això hagi passat per atzar?. Com que hi ha 35 combinacions possibles, aquesta probabilitat és $p = 1/35 = 0,029$. A partir d'aquí, s'aplicarà el raonament típic de l'estadística: Ja que aquesta probabilitat és molt petita (de fet és més petita que el criteri convencional de $p=0,05$), podem pensar que el resultat obtingut no és producte de l'atzar, sinó que realment els resultats obtinguts al tractament C són superiors als obtinguts amb el tractament B.

Si el resultat de la combinació real hagués estat el segon més gran (3,42), llavors la probabilitat d'obtenir una diferència com aquesta o superior seria: $p = 2/35 = 0,057$. En aquest cas la probabilitat està lleugerament per sobre del criteri de significació, tot i que pot indicar-nos també una possible superioritat dels resultats de C per sobre dels de B. Naturalment, si el resultat obtingut fos a l'altre extrem (-4,16) la interpretació seria la mateixa, ja que la probabilitat d'obtenir-la per atzar seria $1/35=0,029$. En aquest cas, però, això ens diria que els resultats obtinguts al tractament B són superiors als del tractament C.

Esquemàticament:

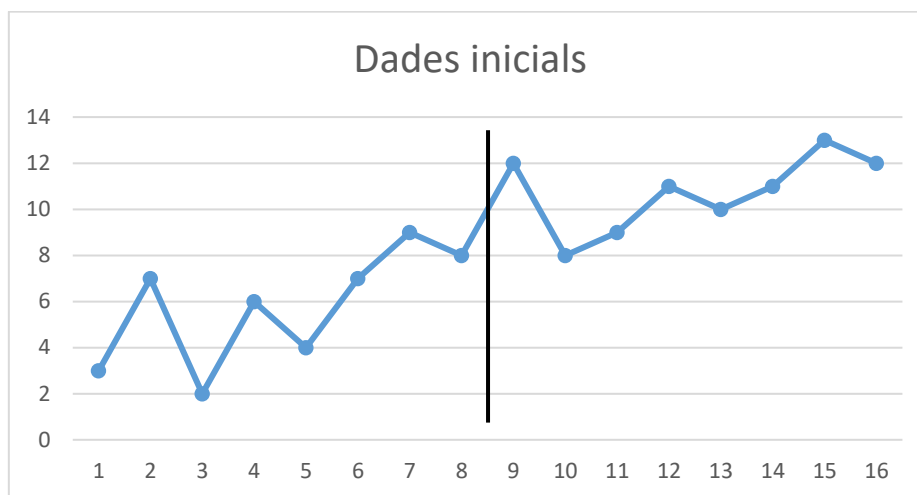
Sessions	C	B	B	C	B	B	C
Resultat	13	10	8	15	12	10	14

- Nº combinacions possibles: $7! / 4! 3! = 35$
- Diferències teòriques ordenades (C – B): 4, 3.42, 2.83, 2.25,, -2.42, -3, -4.16
- Diferència real: $(13 + 15 + 14 / 3) - (10 + 8 + 12 + 10 / 4) = 4$
- Probabilitat d'obtenir aquesta diferència a l'atzar $p = 1 / 35 = 0,029$
- Conclusió: Resultats C > Resultats B (diferència significativa)

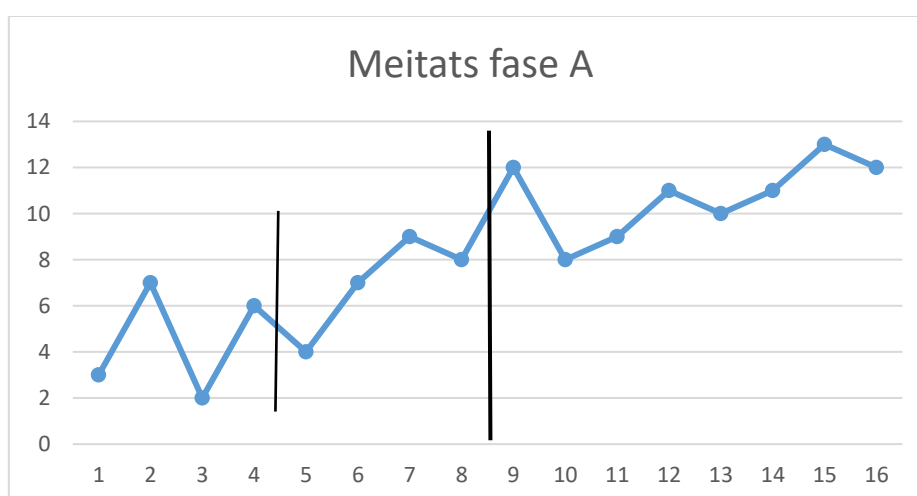
. Tècniques gràfiques

La tècnica que es presentarà és la de divisió en meitats, o *split-middle*, que serveix per detectar possibles canvis de tendència entre una fase i la següent.

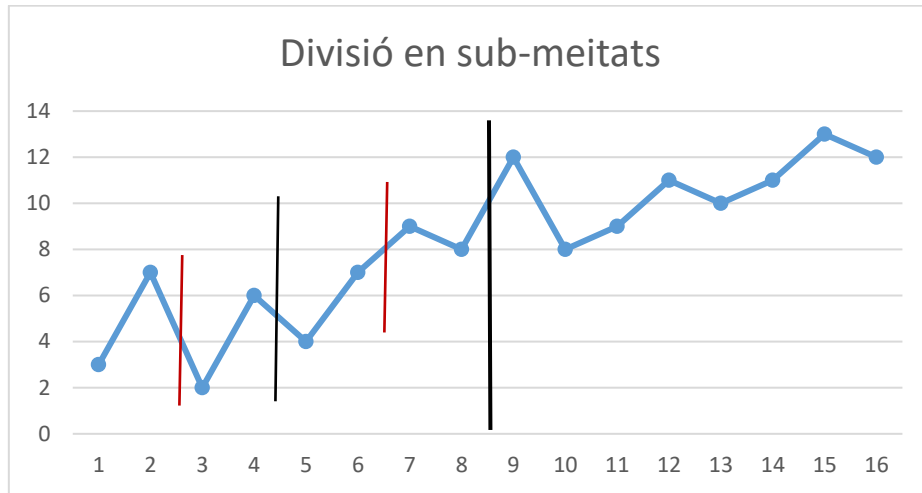
Suposem que a un determinat estudi s'obtenen els resultats següents en dues fases concretes (la primera fase és de línia base i la segona de tractament). Una primera inspecció dels resultats ens pot fer pensar que no sembla existir un efecte potent de la intervenció. La tècnica de divisió en meitats ens permetrà valorar més acuradament el gràfic obtingut.



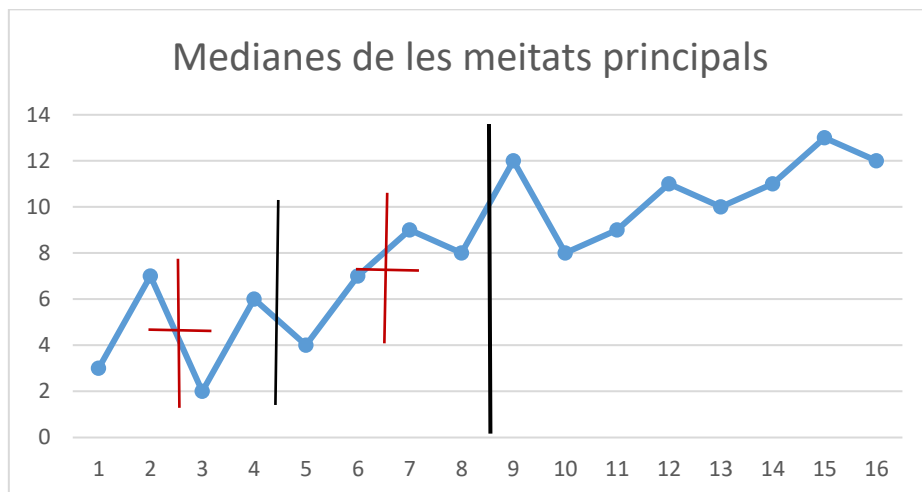
El primer pas és subdividir la primera fase en dues meitats, deixant la meitat de les puntuacions a l'esquerra i la meitat a la dreta de la línia de divisió. Si el número de dades fos imparell, la línia de divisió hauria de passar per sobre de la puntuació situada al mig de la fase. En el nostre cas, com que hi ha 8 puntuacions a la primera fase, en podem deixar 4 a l'esquerra i 4 a la dreta:



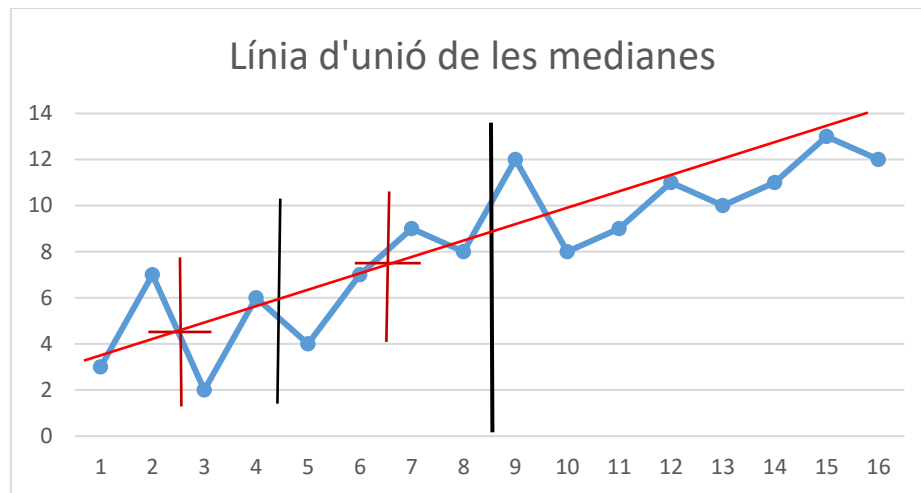
A continuació les dues meitats obtingudes a la primera fase es divideixen novament en dues sub-meitats:



Després es calcula la mediana (no la mitjana) de cadascuna de les meitats principals, i es traça una línia horitzontal a l'alçada de la mediana corresponent a cada meitat. Naturalment, la línia horitzontal de cada meitat es creuarà amb la línia vertical que marca la divisió en sub-meitats:



Finalment, s'uneixen els dos punts de creuament identificats al pas anterior amb una línia recta, i es prolonga aquesta línia també a la segona fase:



El que s'aconsegueix amb tots aquests elements és traçar una línia recta que representa la tendència observada de les dades a la primera fase: molt aproximadament, la meitat de les puntuacions de la línia base queden per sobre de la línia i la meitat per sota. Si el tractament introduït a la segona fase no té cap efecte, cal suposar que aquest patró es mantindrà, és a dir, que la línia traçada continuarà essent representativa i que la meitat de les puntuacions de la segona fase hauria d'estar per sobre i la meitat per sota d'aquesta línia. Com es pot veure, això no és així, només hi ha una puntuació per sobre i 7 per sota. Per tant, en contra del que podria semblar inicialment, els resultats indiquen que el tractament pot haver tingut algun efecte sobre els resultats. En tot cas, aquest efecte sembla bastant discret (algunes puntuacions de la fase d'intervenció s'aproximen a la línia, tot i quedar per sota), per això la seva vàlua s'haurà d'establir a partir del coneixement del cas concret, és a dir, a partir de la valoració clínica, educativa, psico-social, etc. que sigui pertinent.

Alguns autors han proposat l'ús de la distribució binomial de probabilitat per establir quina és la probabilitat d'obtenir els resultats trobats. Per exemple, en el nostre cas caldria establir la probabilitat de que, d'entre 8 puntuacions, per atzar una hagi quedat per sobre de la línia recta i totes les altres per sota. Si aquesta probabilitat és prou baixa (com és el cas), es pot pensar en un efecte significatiu. Cal dir, però, que l'ús de la distribució binomial ha estat també criticada per diferents autors, ja que no és evident que es comporti de forma adient en dades no independents, com ho són probablement les pròpies d'una sèrie temporal.