TREBALL DE FI DE GRAU

Grau en Estadística

Títol: Construcció d'un paquet R per l'estudi de variables binaries compostes

Autor: Raquel Rovira Salvat

Director: Marta Fairén i Marta Bofill

Departament: Departament llenguatges i sistemes informàtics (UPC), Departament d'estadística i investigació operativa (UPC)

Convocatòria: Gener 2020

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Per totes aquelles abraçades pendents
que sempre tindrem al nostre cor
i que mai ningú ens podrà robar.

ABSTRACT

This work focuses on the study of compound events and, in particular, on the design of clinical trials where the main variable is a compound binary variable. The aim is to build an R library with functions dedicated to the calculation of sample size and effect size; in terms of relative risk, odds ratio and risk difference. Once the entire structure by the library has been validated, a numerical example of the parameters required to calculate the sample size based on the postulated design will be given.

RESUM

Aquest treball es centra en l'estudi d'esdeveniments compostos i, en particular, en el disseny d'assajos clínics on la variable principal és una variable binària composta. Es pretén construir una llibreria R amb funcions dedicades al càlcul de mida mostral i del tamany del efecte; en termes del risc relatiu, l'odds rati i la diferència de proporcions. Un cop validada tota l'estructura creada per la llibreria, es donarà un exemple numèric dels paràmetres que es requereixen anticipadament pel càlcul de la mida mostral.

TAULA DE CONTINGUTS

I. INTRODUCCIÓ

Els recents estudis mèdics manifesten una clara tendència a l'alça de la taxa de l'esperança de vida. Avui en dia la societat viu cada vegada més anys. De fet es preveu que aquesta taxa continuï en augment. Anys enrere, una pulmonia o una grip eren malalties mortals però avui en dia es poden tractar sense problemes si el diagnòstic no es complica.

La cura de malalties avança gràcies als tractaments mèdics que avui en dia podem trobar en el mercat. Nombrosos estudis de recerca es desenvolupen per intentar millorar tractaments que ja és comercialitzen o bé per crear-ne de nous que puguin substituir els ja existents i així millorar la seva efectivitat. Tots els tractaments han de ser testats abans de comercialitzar-se ja que cal assegurar la seva eficàcia i seguretat.

Els assajos clínics són els dissenys experimentals que fonamenten la recerca clínica, i els mecanismes a través dels quals es validen els tractaments. Aquests assajos es fonamenten en l'estudi dels efectes que pot generar el tractament d'estudi sobre un grup de persones que representa l'univers d'estudi. Tot assaig haurà d'acabar donant resposta al contrast d'hipòtesi que es postula a continuació:

$$\begin{cases} H_0: \text{El tractament validat no és eficaç} \\ H_1: \text{El tractament validat és eficaç} \end{cases}$$

Mitjançant la informació recollida es podrà resoldre aquest contrast d'hipòtesi postulat gràcies a la utilització de l'estadística, i establirà quina d'elles és afirmativa en base a l'anàlisi realitzat. Caldrà dissenyar l'estudi correctament, definint el contrast d'hipòtesi que es vol testejar en funció de la mida de l'efecte que es vol detectar i calculant la mida mostral necessària per portar-lo a terme.

Per dur a terme aquest procediment existeixen diversos mecanismes sota els quals es fonamenten els assajos clínics. D'entre els més coneguts podem trobar la comparació entre dos grups poblacionals que es diferencien únicament pel tractament que reben, el tractament que fonamenta l'assaig, o bé, el tractament estàndard conegut com placebo.

Cal tenir present que la conclusió final de tot assaig clínic està sotmès a un cert marge d'error, ja que és un disseny experimental. Observem la taula que es mostra a continuació per poder comprendre millor perquè es poden generar aquests errors.

	H0 ES CERTA	H1 ES CERTA
TRIAR H0	No hi ha error	Error de tipo II
TRIAR H1	Error de tipo I	No hi ha error

Taula 1: Possibles errors en contrast d'hipòtesis

Tot estudi conclourà en la tria d'una de les dues hipòtesi plantejades. Pot ser que la hipòtesi escollida en l'estudi realment no sigui la certa. Si escollíssim la hipòtesi nul·la i en realitat la hipòtesi alternativa és la certa, estaríem cometen l'error de tipus II , estadísticament conegut com error de tipus β ; en canvi si escollíssim la hipòtesi alternativa i en realitat la hipòtesi nul·la és la certa estaríem davant de l'error de tipus I, estadísticament conegut com error de tipus α. En tota tipologia d'anàlisi es pot fixar els llindars de α i β , i així determinar el percentatge de marge d'error que es pot assumir.

Imaginem ara que s'està desenvolupant un assaig clínic on es vol estudiar si els individus de l'estudi experimenten o no certs esdeveniments. Suposem que el nou tractament que s'està validant ha estat creat per ser administrat a totes aquelles persones que han patit un ictus. El disseny experimental creat vol donar resposta a les següents preguntes:

- El tractament administrat redueix la mortalitat del conjunt de pacients?
- El tractament administrat resulta eficaç per reduir l'ictus que pateix el pacient?

D'acord amb aquest plantejament, l'estudi en qüestió té diverses variables d'interès. En estudis com el de l'exemple s'introdueix el concepte de variables combinades, conegudes com variables de caràcter binari que prenen valor si i només si algun esdeveniment sota el qual es formula l'assaig s'observa. Si en l'estudi postulat anteriorment algun dels individus mor o presenta un ictus, la variable combinada prendrà valor.

L'objectiu principal d'aquesta tesi es fonamenta en la creació d'un paquet d'R que inclogui un conjunt de funcions d'utilitat en assajos clínics on la variable d'interès és de tipus combinada. En concret es podrà implementar el càlcul d'efectes de tractament,

com l’odds rati, el risc relatiu i la diferència de proporcions, altrament es podrà calcular mides mostrals sota diferents escenaris. Es desenvolupa el contingut teòric i es fonamenta la part pràctica d’aquest treball en base a un article de recerca científic (Gómez & Bofill).

ESTRUCTURA DE LA MEMÒRIA

En el capítol II, s’estudia i es desenvolupa la teoria que hi ha darrere de les variables binàries, per poder entendre exactament la seva simbologia en el món dels assajos clínics. S’estudiaran amb detall algunes mesures estadístiques en assajos clínics on la variable d’interès és de tipus combinada. També en base a aquesta tipologia de variables d’interès veurem com es pot calcular la grandària mostral.

En el capítol III, s'estudien el conjunt de funcions que complementen l'article que s'utilitza com a base per desenvolupar aquest treball. Aquest estudi es realitza per conèixer l’estructura i la funcionalitat de les funcions per tal de poder millorar-les. També es desenvoluparà un conjunt de validacions que s'han realitzat per validar la lògica sota la qual es fonamenta el codi creat.

En el capítol IV veurem com s'ha construït l'entorn i el directori on s'acabarà allotjant tot el codi i tots els arxius requerits que fonamenten el paquet. També es desenvoluparà d'una forma més tècnica com es documentaran tots els arxius requerits, i s'explicarà la metodologia sota la qual es realitzaran les validacions de tota la feina creada.

En el capítol V es dóna un exemple numèric d’algunes de les formulacions teòriques desenvolupades en del capítol II , en concret de totes aquelles que serveixen per anticipar els paràmetres requerits en el càlcul de la grandària mostral.

II. FONAMENTS TEÒRICS

En recerca clínica existeixen múltiples dissenys experimentals sota els qual es poden postular els assajos clínics. Per exemple: podem trobar la comparació de dos tractaments, la comparació de més de dos tractaments on es van eliminant els fàrmacs que no presenten resultats favorables durant la experimentació o alguns dissenys on es comparen diverses fases. En aquest treball considerem assajos clínics que comparen únicament dos tractaments diferents, i suposem que tenim una mostra aleatòria de la població objectiu. Els individus d'aquesta mostra seran aleatòriament assignats al tractament que es vol testejar o al tractament estàndard conegut com placebo. L'esquema que es mostra a continuació plasma el disseny experimental d'interès.

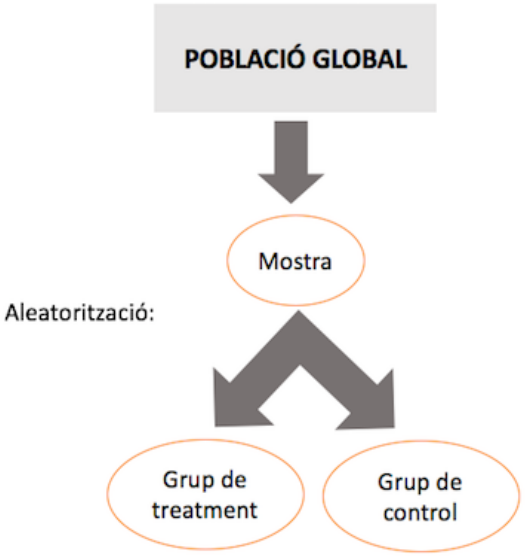

Figura 1:Esquema del disseny experimental

Recuperem el disseny experimental que s'havia postulat com exemple en el capítol I. És suposa que un nou fàrmac resultaria eficaç si s'administra en pacients que han patit un ictus. Per tastar i validar aquest fàrmac, suposem que portem a terme un assaig clínic. S'havia postulat que les variables d'interès sota les quals es fixa el disseny experimental podrien controlar si s'observa algun dels següents esdeveniments:

- El tractament administrat redueix la mortalitat del conjunt de pacients?
- El tractament administrat resulta eficaç per reduir l'ictus que pateix el pacient?

Sota aquesta situació hipotètica, aquests parell d'esdeveniments fonamentarien la variable combinada objecte d'estudi. Denotem E_1 com l'esdeveniment que mesura si l'individu mor i E_2 com l'esdeveniment que mesura si l'individu pateix un ictus. Tot i que es podrien considerar més de dos esdeveniments d'interès en aquest treball en

considerem únicament dos.

Siguin E_1 i E_2 dos esdeveniments diferents entre sí, definim l'esdeveniment compost $E*$ com la unió d'ambdós.

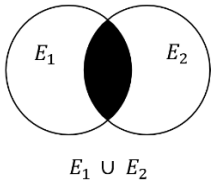

Figura 2:Unió de dos esdeveniments

S'entén $E*$ com l'esdeveniment compost que resulta de la unió de E_1 i E_2. Mesura si en la mostra poblacional d'estudi s'observa almenys un dels dos esdeveniments. Observarem $E*$ si i només si s'observa l'esdeveniment E_1 i/o l'esdeveniment E_2.

Definim X_{ijk} com la variable aleatòria amb distribució Bernoulli que representarà la resposta de l' individu j-ésssim que pertany al grup i-éssim en funció de l'esdeveniment E_k .

$$X_{ijk} = \begin{cases} 1 & si \quad E_k \, s'observa \\ 0 & altrament \end{cases}$$

D'acord amb aquesta postulació podem definir les variables aleatòries i les probabilitats d'observació dels esdeveniments E_1 i E_2 com:

- X_{ij1} = 1 si E_1 s'observa en l'individu *j* que pertany al grup *i*
- X_{ij2} = 1 si E_2 s'observa en l'individu *j* que pertany al grup *i*

$$p_1^i = P\left(X_{ij1} = 1\right) \qquad (1)$$

$$p_2^i = P\left(X_{ij2} = 1\right) \qquad (2)$$

Anàlogament podem postular les probabilitats de no observació d'ambdós esdeveniments en base a la següent formulació:

$$q_1^i = 1 - p_1^{(i)} = P\left(X_{ij1} = 0\right) \quad (3)$$

$$q_2^i = 1 - p_2^{(i)} = P(X_{ij2} = 0) \quad (4)$$

En el cas de l'esdeveniment compost podem definir la variable aleatòria combinada com:

$$X_{ij*} = \begin{cases} 1 & si \quad X_{ij1} + X_{ij2} \geq 1 \\ 0 & si \quad X_{ij1} + X_{ij2} = 0 \end{cases}$$

on:

- X_{ij*} = 1 si $E*$ s'observa , per definició si s'observa E_1 i/o E_2

A partir de la variable aleatòria combinada podem definir la probabilitat $p_*^{(i)}$ d'observació de l'esdeveniment E∗ i la probabilitat $q_*^{(i)}$ de no observació de l'esdeveniment E∗ com:

$$p_*^{(i)} = P(X_{ij*} = 1) \quad (5)$$

$$q_*^{(i)} = 1 - p_*^{(i)} = P(X_{ij*} = 0) \qquad (6)$$

1. Probabilitat de la variable combinada

En aquest apartat veure'm com es calcula la probabilitat d'observar l'esdeveniment compost E∗ coneguda per la unió: $E^* = E_1 \cup E_2$. Siguin p_1^i i p_2^i les probabilitats marginals d'esdevenir un determinat esdeveniment E_k definides teòricament a través de les equacions 1 i 2. Notem *rho* la correlació entre dos esdeveniments, E1 i E2. Aleshores la probabilitat d'observar l'esdeveniment compost E∗ ,depèn de les probabilitats marginals de p_1^i , p_2^i i de la correlació entre els esdeveniments (Bahadur, 1961). La següent formulació representa el càlcul de p∗ coneguts els paràmetres p_1^i , p_2^i i *rho* dels qual depèn.

$$p_*^i = 1 - q_1^i * q_2^i - \rho^i \sqrt{p_1^i * p_2^i * q_1^i * q_2^i} \qquad (7)$$

2. El rol de la correlació

En assajos clínics on la variable d'interès és combinada, resulta interessant conèixer el grau de relació entre els esdeveniments que formen aquesta composició. La correlació

de tota variable combinada es troba acotada de forma intervalica (Prentice, 1988) . El rang inferior d'aquest interval de confiança pren valors negatius acotats entre el 0 i el -1, el rang superior pren valors positius acotats entre 0 i 1. Així doncs, definim l'espai paramètric d'aquesta correlació com:

$$p^i \epsilon \left[m\left(p_1^i, p_2^i\right), M\left(p_1^i, p_2^i\right)\right] \subseteq [-1,1]$$

on:

$$m\left(p_1^i, p_2^i\right) = max\left\{-\sqrt{\frac{p_1^i * p_2^i}{q_1^i * q_2^i}}, -\sqrt{\frac{q_1^i * q_2^i}{p_1^i * p_2^i}}\right\} \qquad (8)$$

$$M\left(p_1^i, p_2^i\right) = min\left\{+\sqrt{\frac{p_1^i * q_2^i}{q_1^i * p_2^i}}, +\sqrt{\frac{p_2^i * q_1^i}{q_2^i * p_1^i}}\right\} \qquad (9)$$

Notem que $m\left(p_1^i, p_2^i\right)$ i $M\left(p_1^i, p_2^i\right)$ corresponen al rang inferior i al rang superior de correlació de tota variable combinada. En aquest treball assumim que la correlació és la mateixa en els dos grups de tractament i, per tant, les cotes de correlació es troben acotades en:

$$\left\{max\left(m(p_1^0, p_2^0), m(p_1^1, p_2^1)\right); min\left(M(p_1^0, p_2^0), M(p_1^1, p_2^1)\right)\right\}$$

3. Mesures estadístiques per quantificar l'efecte d'un fàrmac

Per tal de quantificar l'efecte d'un fàrmac, s'utilitzen diferents mesures estadístiques. Els efectes estadístics resulten àmpliament utilitzats en el camp de l' estadística ja que mesuren la força i/o quantifiquen l'efecte d'un fenomen. Juguen un paper molt important en els assajos clínics. En aquest treball ens centrarem en desenvolupar l'odds rati, el risc relatiu i la diferència de proporcions.

3.1 Odds rati

L'odd és una mesura estadística que representa la raó entre dues probabilitats; la probabilitat d'observar un esdeveniment en un grup i la probabilitat que s'observi el mateix esdeveniment en un grup d'estudi diferent.

$$O_k^{(i)} = \frac{p_k^{(i)}}{q_k^{(i)}} \quad (10)$$

L'odds rati, conegut també com la raó de probabilitat o raó d' oportunitat, és una mesura que es calcula en base la raó de dos odds. L'odds rati es defineix com:

$$OR_k = \frac{O_k^{(1)}}{O_k^{(0)}} = \frac{p_k^{(1)} / q_k^{(1)}}{p_k^{(0)} / q_k^{(0)}}, \quad \text{k =1,2} \quad (11)$$

Denotem OR_1 l'odds rati associat al esdeveniment E1 i anàlogament OR_2 com l'odds rati de l'esdeveniment E2.

Notem que quan $OR < 1$, la probabilitat associada al grup de control d'observar l'esdeveniment d'estudi sempre serà major a la probabilitat associada del grup d'intervenció, és a dir:

$$OR < 1 \rightarrow P_*^0 > P_*^1$$

L'odds rati de la variable combinada es calcula en base a: el valor de OR_k, l'odds rati de l'esdeveniment E_k , p_k^0, com la probabilitat d'observar l'esdeveniment E_k en el grup de control, rho, la correlació de Pearson entre dos esdeveniments $E_1 \, i \, E_2$, i es defineix segona la fórmula següent:

$$OR_* = \frac{\left(\left(1+\frac{OR_1 * p_1^0}{1-p_1^0}\right)\left(1+\frac{OR_2 * p_2^0}{1-p_2^0}\right) - 1 - rho * \sqrt{\frac{OR_1 * OR_2 * p_1^0 * p_2^0}{(1-p_1^0)(1-p_2^0)}}\right) * \left(1+rho * \sqrt{\frac{p_1^0 * p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)}{\left(\left(1+\frac{p_1^0}{1-p_1^0}\right)\left(1+\frac{p_2^0}{1-p_2^0}\right) - 1 - rho * \sqrt{\frac{p_1^0 * p_2^0}{(1-p_1^0)(1-p_2^0)}}\right) * \left(1+rho * \sqrt{\frac{OR_1 * OR_2 * p_1^0 * p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)} \quad (12)$$

3.2 Risc relatiu

El risc relatiu es coneix per representar la proporció de la probabilitat d'observació d'un esdeveniment, és a dir, ens d'on una estimació de com probable és observar un determinat fet. Mesura la força d'associació entre l'exposició dels individus en l'assaig i la suposada observació de la variable d'interès.

Donat un esdeveniment E_K, es defineix el risc relatiu com:

$$RR_k = \frac{p_k^{(1)}}{p_k^{(0)}}, \quad \text{k =1,2} \quad (13)$$

Denotem RR_1 el risc relatiu associat al esdeveniment E1 i anàlogament RR_2 com el risc relatiu de l'esdeveniment E2.

Quan el risc relatiu pren valor inferior a 1, trobem que el tractament actua com a efecte protector davant la possible observació de l'esdeveniment, per tant, la probabilitat

d'observació de E_K associada al grup de control sempre serà superior a la probabilitat d'observació del grup d'intervenció. S'interpreta com:

$$RR < 1 \rightarrow P_*^0 > P_*^1$$

El risc relatiu de la variable combinada depèn dels següents paràmetres: R_k com el risc relatiu de l'esdeveniment E_K , rho com la correlació de Pearson entre els esdeveniments E_1 *i* E_2 , i p_k^0 com la probabilitat d'observació de l'esdeveniment E_K en base als individus que formen part del grup de control, i es defineix com:

$$RR_* = \frac{p_1^0 R_1 + p_2^0 R_2 - p_1^0 p_2^0 R_1 R_2 - rho\sqrt{p_1^0 p_2^0 R_1 R_2 (1-p_1^0 R_1)(1-p_2^0 R_2)}}{1 - q_1^0 q_2^0 - rho\sqrt{p_1^0 p_2^0 q_1^0 q_2^0}} \quad (14)$$

3.3 Diferència de proporcions

Donat E_K podem calcular la diferència de proporcions de l'esdeveniment a través de la següent equació:

$$\delta_k = p_k^{(1)} - p_k^{(0)}, \quad \text{k =1,2} \quad (15)$$

Notem δ_1 com la diferencia de proporcions associada a l'esdeveniment E_1 i de la mateixa manera associat a l'esdeveniment E_2 notem el paràmetre δ_2 que representa la mateixa mesura.

Quan la diferència calculada a partir de l'equació 15 és negativa la proporció de casos positius en el grup de control és superior a la respectiva proporció en el grup d'intervenció.

Podem observar que la diferència de proporcions de l'esdeveniment compost depèn de: la probabilitat $p_k^{(0)}$d'observar E_k i la probabilitat $q_k^{(0)}$de no observar E_k en el grup control, les diferències de proporcions dels esdeveniments individual denotats com δ_k i rho coneguda com la correlacio entre dos esdeveniments . Es defineix aquesta diferència com:

$$\delta_* = \delta_1 q_1^0 + \delta_2 q_2^0 - \delta_1 \delta_2 + \ldots$$

$$rho\sqrt{p_1^{(0)} p_2^{(0)} q_1^{(0)} q_2^{(0)}} - \ldots$$

$$\rho\sqrt{(p_1^{(0)} + \delta_1)(p_2^{(0)} + \delta_2)(q_1^{(0)} - \delta_1)(q_2^{(0)} - \delta_2)} \quad (16)$$

4. Grandària mostral

Un dels punts més importants en el càlcul d'un assaig clínic és el càlcul de la grandària mostral que s'utilitzarà . Si la grandària mostral escollida és massa elevada això suposarà el desenvolupament d'un estudi molt costós i complex, en canvi, si la grandària mostral resulta insuficient estarem davant d'un escenari on augmenten les probabilitats de cometre errors .

Definim la tipologia de contrast d'hipòtesis amb el qual treballem, notant Δ com la magnitud del efecte del tractament d'interès. Definim el contrast:

$$\begin{cases} H_0: \Delta = 0 \\ H_1: \Delta < 0 \end{cases}$$

En la hipòtesis nul·la establim que el tractament no té efecte i que per tant no existeixen diferències entre els dos grups de tractament, altrament en la hipòtesis alternativa postulem que les diferències entre grups són significatives i definim en la hipòtesis la mida de l'efecte que es vol detectar per validar el tractament.

L'adient mida mostral servirà per acotar els riscos de decisions errònies que es poden esdevenir en tot assaig clínic, coneguts com error de tipus I i error de tipus II, els podem trobar descrits en la taula 1. Per una banda estarem davant del error de tipus I quan haguem escollit H1 però en realitat H0 és la correcte, es cataloga aquesta tipologia d'errors com a falsos positius ja que suposem diferències entre els dos grups d'estudi quan en realitat no n'hi ha. D'altre banda, estarem davant del error de tipus II quan haguem escollit H0 però en realitat H1 és la hipòtesis correcte, en aquestes situacions estarem davant d'un fals negatiu ja que postularem una diferencia entre grups quan en realitat no existeix. Alhora de calcular la grandària mostral de l'assaig clínic suposarem que dissenyem un estudi amb un nivell de significació α i una potència $1 - \beta$, caldrà fixar a priori ambdós paràmetres, plasmen el risc que estem disposats a assumir en el disseny.

Per donar resposta al contrast d'hipòtesis es requereix un estadístic. En el gràfic que es mostra a continuació podem observar la distribució del test estadístic sota el qual es resoldrà el contrast postulat.

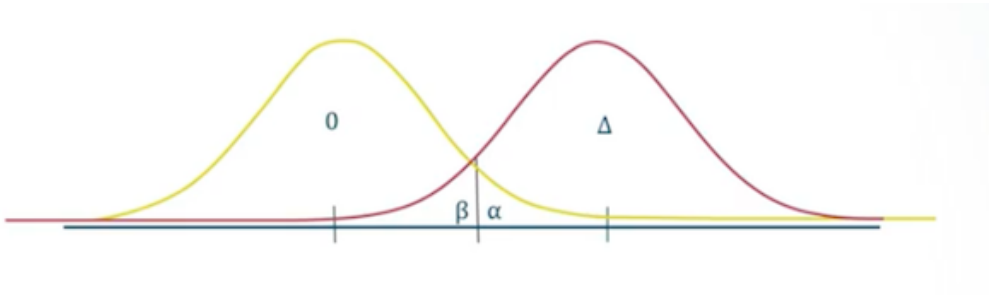

Figura 3: Distribució del test estadístic

En l'eix d'abscisses de la figura 3 s'estableix la resposta del contrast d'hipòtesis. La campana groga dona al tractament que s'està validant un efecte nul establint per tant la hipòtesis nul·la, en canvi, la campana vermella determina que el tractament té un determinat efecte i postula en ella la hipòtesis alternativa. Identificats sota els coeficients de α i β podem trobar els errors de tipus I i II, estan situats al centre de la figura 3. Aquest solapament plasma que la diferència entre la conclusió que prendrà el disseny és molt reduïda.

En aquest treball considerarem tres tipus de contrastos d'hipòtesis per assajos clínics. Estudiarem el càlcul de la mida mostral en cada cas. La taula que es postula a continuació mostra els paràmetres que anticipen els efectes estadístics conjuntament amb els contrastos d'hipòtesis amb els quals ens basarem.

	Paràmetre del efecte	***Hipòtesis nul·la***	***Hipòtesis alternativa***
Diferència de proporcions	$\delta_* = p_*^{(1)} - p_*^{(0)}$	$\delta_* = 0$	$\delta_* < 0$
Risc relatiu	$RR_* = \frac{p_*^{(1)}}{p_*^{(0)}}$	$log(R_*) = 0$	$log(R_*) < 0$
Odds rati	$OR_* = \frac{\frac{p_*^{(1)}}{q_*^{(1)}}}{\frac{p_*^{(0)}}{q_*^{(0)}}}$	$log(OR_*) = 0$	$log(OR_*) < 0$

Taula 2: (Gómez & Bofill)

El nostre objectiu es calcular la grandària mostral pel disseny de l'assaig clínic en funció dels paràmetres dels efectes i els contrastos postulats en la Taula 2. Anticipar aquests valors servirà determinar la mida mostral, petits canvis en la magnitud de l'efecte acabaran provocant grans variacions en les grandàries mostrals calculades.

4.1 Càlcul de la grandària mostral en funció de l'odds rati

En termes d'odds rati la hipòtesis nul·la s'estableix com $H_0 : \log(\mathrm{OR}_*) = 0$ i es compara amb la hipòtesis alternativa $H_1 : \log(\mathrm{OR}_*) < 0$ en la qual és determina que existeix un grau de magnitud del efecte del tractament. Per testejar el contrast d'hipòtesis postulat utilitzem el següent estadístic:

$$W_{*,n} = \frac{\log(\widehat{OR}_*)}{\sqrt{\widehat{Var}(\log(\widehat{OR}_*))}}$$

En l'estadístic $W_{*,n}$ s'estima el valor de l'odds rati de l'esdeveniment compost com $\widehat{OR}_* = \widehat{p_*^{(1)}}\Big/\widehat{q_*^{(1)}} : \widehat{p_*^{(0)}}\Big/\widehat{q_*^{(0)}}$ i es calcula la variància de $log\big(\widehat{\mathrm{OR}}_*\big)$ sota la hipòtesis nul·la o sota la hipòtesis alternativa com:

$$Var_{H_0}(log(\mathrm{OR}_*)) = \frac{8}{n(p_*^0 + p_*^1)(q_*^0 + q_*^1)}$$

$$Var_{H_1}(log(\mathrm{OR}_*)) = \frac{1}{n^0}\left(\frac{1}{p_*^0 q_*^0} + \frac{1}{p_*^1 q_*^1}\right)$$

aquesta variància l'estimem segons:

$$\widehat{Var_{H0}}\big(log\big(\widehat{OR}_*\big)\big) = \frac{8}{n(\widehat{p_*^{(0)}} + \widehat{p_*^{(1)}})(\widehat{q_*^{(0)}} + \widehat{q_*^{(1)}})}$$

$$\widehat{Var_{H1}}\big(log\big(\widehat{OR}_*\big)\big) = \frac{1}{n^0}\left(\frac{1}{\widehat{p_*^{(0)}}\widehat{q_*^{(0)}}} + \frac{1}{\widehat{p_*^{(1)}}\widehat{q_*^{(1)}}}\right)$$

L'estadístic $W_{*,n}$ segueix asimptòticament una distribució normal estàndard. La hipòtesis nul·la es rebutjarà si considerant un nivell de significació α el valor de l'estadístic $W_{*,n} < -Z_\alpha$.Notem Z_α i Z_β com els respectius valors de les desviacions normals estandaritzades d'acord amb els valors dels paràmetres α i β fixats.

Donada la probabilitat p_*^0 l'equació que s'utilitza per calcular la grandària mostral havent estimat la variància de l'estimador sota la hipòtesis nul·la es defineix com:

$$n = \frac{2(Z_\alpha\sqrt{\frac{2}{\overline{p_*^1}*\overline{q_*^1}}}+Z_\beta\sqrt{\frac{(q_*^0+p_*^0\,\mathrm{OR}_*)^2}{p_*^0 q_*^0\,\mathrm{OR}_*}+\frac{1}{p_*^0 q_*^0}})^2}{\log(\mathrm{OR}_*)^2} \quad (15)$$

On:

$\overline{p_*^1} = \frac{p_*^0+p_*^1}{2}$

$\overline{q_*^1} = \frac{q_*^0+q_*^1}{2}$

Altrament, si la variància de l'estimador ha estat calculada en base la hipòtesis alternativa s'utilitza la següent equació:

$$n = \frac{2(Z_\alpha+Z_\beta)^2*\left(\frac{(q_*^0+p_*^0\,\mathrm{OR}_*)^2}{p_*^0 q_*^0\,\mathrm{OR}_*}+\frac{1}{p_*^0 q_*^0}\right)}{\log(\mathrm{OR}_*)^2} \quad (16)$$

4.2 Càlcul de la grandària mostral en funció del risc relatiu

En termes del risc relatiu la hipòtesis nul·la s'estableix com $H_0 : \log(R_*) = 0$ i es compara amb la hipòtesis alternativa $H_1 : \log(R_*) < 0$ en la qual és determina l'efecte del tractament. L'estadístic que utilitzem per testejar la significació del risc relatiu R_* és:

$$Z_{*,n} = \frac{\log(\widehat{R_*})}{\sqrt{\widehat{Var}(\log(\widehat{R_*}))}}$$

S'estima el valor del risc relatiu $\widehat{R_*}$ com $\widehat{R_*} = \widehat{p_*^{(1)}}\Big/\widehat{p_*^{(0)}}$. Notem que la variància de $\log(\widehat{R_*})$ de l'estadístic es pot calcular sota la hipòtesis nul·la i alternativa com:

$$Var_{H_0}(log(\,R_*)) = \frac{2}{n^{(0)}} * \frac{q_*^0+q_*^1}{p_*^0+p_*^1}$$

$$Var_{H_1}(log(\,\mathrm{R}_*)) = \frac{1}{n^{(0)}}\left(\frac{q_*^0+q_*^1}{p_*^0+p_*^1} + \frac{q_*^0}{p_*^0}\right)$$

i per tant podem estimar-la segons:

$$\widehat{Var_{H0}}\left(log\left(\,\widehat{R_*}\right)\right) = \frac{2}{n^{(0)}} * \frac{\widehat{q_*^{(0)}}+\widehat{q_*^{(1)}}}{\widehat{p_*^{(0)}}+\widehat{p_*^{(1)}}}$$

$$\widehat{Var_{H1}}\left(\,log\left(\,\widehat{R_*}\right)\right) = \frac{1}{n^{(0)}}\left(\frac{\widehat{q_*^{(0)}}+\widehat{q_*^{(1)}}}{\widehat{p_*^{(0)}}+\widehat{p_*^{(1)}}} + \frac{\widehat{q_*^{(0)}}}{\widehat{p_*^{(0)}}}\right)$$

L'estadístic $Z_{*,n}$ segueix asimptòticament una distribució normal estàndard. La hipòtesis nul·la es rebutjarà si considerant un nivell de significació α el valor de l'estadístic $Z_{*,n} < -Z_\alpha$. Notem Z_α i Z_β com els respectius valors de les desviacions normals estandarditzades d'acord amb els valors dels paràmetres α i β fixats.

Donada la probabilitat p_*^0 l'equació que s'utilitza per calcular la grandària mostral havent estimat la variància de l'estimador sota la hipòtesis nul·la es defineix com:

$$n = \frac{2(Z_\alpha\sqrt{\frac{2*\overline{q_*}}{\overline{p_*}}}+Z_\beta\sqrt{\frac{(1-p_*^0\,\mathrm{R}_*)^2}{p_*^0\,\mathrm{R}_*}+\frac{q_*^0}{p_*^0}})^2}{\log(\,\mathrm{R}_*)^2} \quad (17)$$

On:

$\overline{p_*^1} = \frac{p_*^0+p_*^1}{2}$

$\overline{q_*^1} = \frac{q_*^0+q_*^1}{2}$

Altrament, si la variància de l'estimador ha estat calculada en base la hipòtesis alternativa s'utilitza la següent equació:

$$n = \frac{2(Z_\alpha+Z_\beta)^2*\left(\frac{1-p_*^0\,\mathrm{R}_*{}^2}{p_*^0\,\mathrm{R}_*}+\frac{q_*^0}{p_*^0}\right)}{\log(\,\mathrm{R}_*)^2} \quad (18)$$

4.3 Càlcul de la grandària mostral en funció de la diferència de proporcions

La hipòtesis nul·la en termes de diferència de proporcions s'estableix com $H_0 : \delta_* = 0$ i es compara amb la hipòtesis alternativa $H_1 : \delta_* < 0$ en la qual és determina que existeixen diferències de proporcions entre grups, on aquesta diferencia es defineix com $\delta_* = p_*^{(1)} - p_*^{(0)}$. Coneixent que $p_*^{(i)}$ es pot estimar com:

$$\widehat{p_*^{(i)}} = \frac{1}{n^{(i)}} \sum_{j=1}^{n^{(i)}} X_{ij*}$$

Definim l'estadístic que s'utilitza per testejar la significació de la diferència de proporcions com:

$$T_{*,n} = \frac{\widehat{p_*^{(1)}} - \widehat{p_*^{(0)}}}{\sqrt{\widehat{Var}\left(\widehat{p_*^{(1)}} - \widehat{p_*^{(0)}}\right)}}$$

Notem que la variància de $\left(\widehat{p_*^{(1)}} - \widehat{p_*^{(0)}}\right)$ de l'estadístic $T_{*,n}$ és pot calculat sota la hipòtesis nul·la o sota la hipòtesis alternativa com:

$$Var_{H_0}\left(p_*^{(1)} - p_*^{(0)}\right) = \frac{(p_*^0 + p_*^1)(q_*^0 + q_*^1)}{2n^{(0)}}$$

$$Var_{H_1}\left(p_*^{(1)} - p_*^{(0)}\right) = \frac{(p_*^0 q_*^0) + (p_*^1 q_*^1)}{n^{(0)}}$$

i per tant podem estimar-la segons:

$$\widehat{Var_{H0}}\left(\widehat{p_*^{(1)}} - \widehat{p_*^{(0)}}\right) = \frac{(\widehat{p_*^{(0)}} + \widehat{p_*^{(1)}})(\widehat{q_*^{(0)}} + \widehat{q_*^{(1)}})}{2n^{(0)}}$$

$$\widehat{Var_{H1}}\left(\widehat{p_*^{(1)}} - \widehat{p_*^{(0)}}\right) = \frac{\widehat{p_*^{(0)}}\widehat{q_*^{(0)}} + \widehat{p_*^{(1)}}\widehat{q_*^{(1)}}}{n^{(0)}}$$

L'estadístic $T_{*,n}$ segueix asimptòticament una distribució normal estàndard. La hipòtesis nul·la es rebutjarà si considerant un nivell de significació α el valor de l'estadístic $T_{*,n} < -Z_\alpha$.Notem Z_α i Z_β com els respectius valors de les desviacions normals estandarditzades d'acord amb els valors dels paràmetres α i β fixats.

Donada la probabilitat p_*^0 l'equació que s'utilitza per calcular la grandària mostral havent estimat la variància de l'estimador sota la hipòtesis nul·la es defineix com:

$$n = \frac{2(Z_\alpha\sqrt{2*\overline{q_*}*\overline{p_*}}+Z_\beta\sqrt{p_*^0*\, q_*^0+(p_*^0+\delta_*)(q_*^0-\delta_*)})^2}{{\delta_*}^2} \quad (19)$$

On:

$\overline{p_*^1} = \frac{p_*^0+p_*^1}{2}$

$\overline{q_*^1} = \frac{q_*^0+q_*^1}{2}$

Altrament, si la variància de l'estimador ha estat calculada en base a la hipòtesis alternativa s'utilitza la següent equació:

$$n = \frac{2(Z_\alpha+Z_\beta)^2*\,(p_*^0*\, q_*^0+(p_*^0+\delta_*)(q_*^0-\delta_*))}{{\delta_*}^2} \quad (20)$$

III. ALGORITMICA DE LES FUNCIONS

En aquest capítol ens centrarem en desenvolupar l'estructura final i tots els canvis realitzats sobre les funcions que complementaven l'article que s'ha agafat com a referència. Es començarà estudiant el format inicial de totes les funcions i s'argumentarà el perquè dels canvis sobre el codi que s'ha cregut oportú realitzar. Finalment, es validaran les funcions en construcció per tal d'evitar errors o mancances en totes elles. Totes les funcions i el codi emprat el podem trobar complert en els annexos de la tesi.

1. Funcions de partida

El coneixement de totes les funcions de partida és un factor imprescindible per poder avaluar quines millores i extensions es poden fer. Observem la taula que es mostra a continuació on es detallen els resultats que s'obtindrien de l'execució de les funcions d'estudi.

FUNCIÓ	RESULTAT OBTINGUT
Bahadur.composite	Probabilitat d'unió de dos esdeveniments
Correlation.min.function	Rang inferior de l'interval de correlació de E*
Correlation.max.function	Rang superior de l'interval de correlació de E*
Diff.pcomp	Diferència de proporcions de E*
RR.pcomp	Risc relatiu de E*
OR.composite.function	Odds relatiu de E*
SampleSize.CBE.OR	Grandària mostral calculada en termes de l'odds rati
SampleSize.OR	Grandària mostral calculada en termes de l'odds rati
SampleSize.CBE.RR	Grandària mostral calculada en termes del risc relatiu
SampleSize.RR	Grandària mostral calculada en termes del risc relatiu
SampleSize.CBE.Diff	Grandària mostral calculada en termes de la diferència de proporcions
SampleSize.Diff	Grandària mostral calculada en termes de la diferència de proporcions

Taula 3: Funcions de partida

Procedim a descriure l'estructura interna de totes les funcions mencionades en la taula 3, agrupant-les totes elles per similitud funcional.

1. **Probabilitat d'unió de dos esdeveniments**

Siguin p1 i p2 les probabilitats d'observació dels esdeveniment E1 i E2, coneguda *rho* com la correlació de Pearson entre dos esdeveniments, la funció Bahadur.composite calcula la probabilitat d'unió de dos esdeveniments interpretada com E^* = $E1 \cup E2$, en base a l'equació 7 en funció dels paràmetres $p1$, $p2$ i *rho*.

2. **Correlació de l'esdeveniment compost**

Trobem dues funcions que s'utilitzen per calcular l'interval de correlació d'un esdeveniment binari compost. Tot i trobar-se per separat, les dues funcions s'alimenten dels mateixos paràmetres d'entrada. Definim pk com la probabilitat d'observació de l'esdeveniment Ek.

- Rang inferior

La crida de la funció que calcula la cota inferior d'aquest interval es coneix com *corr.min*, es fonamenta mitjançant l'equació 8 i s'alimenta de pk.

- Rang superior

La funció que calcula el límit superior d'aquest interval es coneix com *corr.max* , requereix el paràmetre pk i es fonamenta mitjançant l'equació 9.

3. **Efectes de l'esdeveniment compost**

En el conjunt global hi ha funcions que s'utilitzen per calcular els efectes estadístics sobre els esdeveniments compostos binaris. Per tal de descriure-les totes, les segmentarem en dos grups, agrupant-les per similitud estructural, en base als paràmetres d'entrada de les funcions d'estudi. Per una banda estudiarem l'odds rati, i d'altre banda conjuntament el risc relatiu i la diferència de proporcions.

- Odds rati

Sigui p0.Ek la probabilitat d'esdevenir l'esdeveniment Ek en funció dels individus que formen part del grup de control, *rho* la correlació de Pearson entre dos esdeveniments i ORk l'odds rati del respectiu esdeveniment Ek. Trobem una funció que permet calcular l'odds rati de qualsevol esdeveniment compost binari en funció dels paràmetres descrits.

Aquesta funció es coneix sota la nomenclatura de *OR.composite.function* i s'explica a través de l'equació 12.

- Diferència de proporcions i risc relatiu

Existeixen un parell de funcions útils per estudiar el risc relatiu i la diferència de proporcions de qualsevol esdeveniment compost binari. Es formulen a partir de la probabilitat d'unió d'esdeveniments interpretada com $E_1 \cup E_2$. Aquest valor es pot obtenir a partir de la funció *Bahadur.composite* descrita anteriorment.

Definim els paràmetres p0.E1 i p0.E2 com les probabilitats d'observar l'esdeveniment Ek en funció dels individus que formen part del grup de control i del grup d'intervenció respectivament i rho la correlació de pearson de entre dos esdeveniments.

La funció *RR.pcomp* ens permetrà calcular el valor del risc relatiu del esdeveniment compost d'acord amb l'equació 14 i la funció *Diff.pcomp* és fomentarà de l'equació 16 per calcular la diferència de proporcions de la variable combinada.

4. Grandària mostral de l'esdeveniment compost

Finalment, podem trobar funcions que s'utilitzen per calcular les adients grandàries mostrals per assajos clínics amb variable combinada. En concret podem trobar-ne sis. La grandària mostral tal i com l'hem desenvolupada en aquest treball té en compte la magnitud dels efectes estadístics per tal de poder calcular-se. En el conjunt d'estudi trobem dues funcions per cada efecte estadístic postulat; odds rati , risc relatiu i diferència de proporcions.

Definim p0.Ek com la probabilitat que es produeixi l'esdeveniment Ek en base als individus que formen part del grup de control, *rho* com la correlació entre dos esdeveniments, els coeficients α i β com les probabilitats d'error de tipus I i II que estem disposats a assumir en l'assaig respectivament i *Unpooled* com la variable qualitativa que defineix la tipologia de variància utilitzada. Els paràmetres descrits anteriorment és requereixen en totes les parelles de funcions del conjunt d'estudi.

- Funcions que es fonamenten de l'odds rati

En el conjunt que s'estudia trobem la funció *SampleSize.CBE.OR* que permet calcular la grandària mostral de l'esdeveniment compost en funció de la magnitud de l'efecte de l'odds rati. Aquesta funció depèn de l'odds rati de l'esdeveniment Ek conegut com ORk i dels següents paràmetres definits a priori: p0.Ek, α, β i *Unpooled*.

Per substitució calcula les probabilitats p1.Ek d'observació de l'esdeveniment Ek a partir de l'equació 11. En funció de tots els paràmetres d'entrada i aquells que s'han calculat

dins la funció, es calcula la probabilitat d'unió de la variable combinada pel grup de control i pel grup d'intervenció en funció de l'equació 7 , així com el valor de l'odds rati de la variable combinada a través de la funció *OR.composite.function.*

Seguidament en el seu procediment podem trobar la crida de la funció *SampleSize.OR* . S'alimenta del valor de l'odds rati i del valor de la probabilitat d'observació de l'esdeveniment Ek de la variable combinada descrits anteriorment, així com dels paràmetres: α, β i $Unpooled$.

La funció *SampleSize.OR* finalitzarà retornant el valor de la grandària mostral requerit d'acord amb les equacions 15 i 16, diferenciant les dues equacions per la tipologia de variància que s'utilitza. Aquesta sortida representarà també la sortida de la funció *SampleSize.CBE.OR*.

- Funcions que es fonamenten del risc relatiu

La funció *SampleSize.CBE.RR* que es troba en el conjunt d'estudi permet calcular la grandària mostral de l'esdeveniment compost en funció del valor del risc relatiu. Aquesta funció depèn de RRk conegut com el risc relatiu de l'esdeveniment Ek i dels següents paràmetres definits a priori: p0.Ek, α, β i $Unpooled$.

Per substitució calcula les probabilitats p1.Ek d'observació de l'esdeveniment Ek a partir de l'equació 13. Es calcula la probabilitat d'unió de la variable combinada pel grup de control i pel grup d'intervenció en funció de l'equació 7 i el risc relatiu de la variable combinada mitjançant la funció *RR.pcomp* en funció dels paràmetres d'entrada i els valors calculats dins la funció.

Posteriorment podem trobar la crida de la funció *SampleSize.RR* . La crida d'aquesta funció requereix el valor dels paràmetres α, β i $Unpooled$, així com la probabilitat d'observació de l'esdeveniment compost en base al grup de control calculada anteriorment. El valor que retornarà aquesta funció coincidirà amb la sortida de *SampleSize.CBE.RR* .

Es coneix que *SampleSize.RR* finalitzarà retornant el valor de la grandària mostral en funció de la tipologia de variància establerta i d'acord amb les equacions 17 i 18.

- Funcions que es fonamenten de la diferència de proporcions

En funció de la diferència de proporcions com a efecte la funció *SampleSize.CBE.Diff* permet calcular la grandària mostral de l'esdeveniment compost. Aquesta funció depèn de les diferències dels esdeveniments Ek i dels següents paràmetres definits a priori: p0.Ek, α, β i $Unpooled$.

A partir de l'equació 15 via substitució es calculen les probabilitats p1.Ek d'observació de l'esdeveniment Ek. També s'extreu la probabilitat d'unió de la variable combinada pel grup de control i pel grup d'intervenció en funció de l'equació 7 i utilitzant la funció *Diff.pcomp* es permet calcular la diferència de proporcions de l'esdeveniment compost.

Dins d'aquest programa podem trobar la crida de la funció *SampleSize.Diff* que retornarà la grandària mostral que es requereix en funció de les equacions 19 i 20. La crida d'aquesta funció requereix el valor dels paràmetres α, β i $Unpooled$, i addicionalment el valor de la probabilitat d'observació i la diferència de proporcions de la variable combinada.

2. Millores de les funcions

Un cop conegudes les funcions que complementen l'article de base del projecte, s'han pogut observar diversos aspectes ineficients en les funcions, així com alguns filtres mancants.

En cap de les funcions hi ha restriccions per als paràmetres d'entrada, fet que es considera important i es creu que s'ha d'incorporar en totes les funcions que acabaran formant part del paquet. S'han de validar tots els espais paramètrics per tal d'evitar errors de càlcul ja que existeixen acotacions pels valors que representen els paràmetres d'entrada. Per exemple, es recorda que totes les probabilitats han d'estar acotades entre 0 i 1, el valor de la correlació de l'esdeveniment compost ha d'estar acotat en el correcte interval, etc.

Existeixen dependències entre les funcions, fet que justifica l'existència de funcions que només s'utilitzen dins l'execució. Tal i com hem pogut observar en la taula 3 hi ha un conjunt de funcions que només s'utilitzen en la crida a d'altres funcions. La funció que crida a una altre funció i la funció que es crida tenen la mateixa sortida. Per tenir un major control de totes les funcions que acabaran formant part del paquet i tenint en compte que totes les funcions han de passar un filtre analític abans de ser validades, fet que comporta un risc que es pot minimitzar reduint el nombre total de funcions, es decideix unificar totes les funcions que sigui possible dins d'un criteri pactat.

S'acorda agrupar totes les funcions que sigui possible per semblança i funcionalitat. Aquesta decisió ha estat pressa d'acord amb l'argument que s'acaba de postular i després de conèixer l'estructura de totes les funcions originals, fet que ha permès observar les equivalències entre funcions i així garantir la millor agrupació entre elles.

Es decideix canviar el nom d'algunes funcions ja que es considera que no tots els noms són clarificadors, ni es relacionen directament amb la funcionalitat de la funció, fet que es vol plasmar.

En la taula 4 és defineix breument, el significat dels paràmetres d'entrada que formaran part de les funcions. S'han canviat algunes d'aquestes entrades per tal de tenir una unificació i clarificació de tots els paràmetres de les funcions. Interpretem tots els paràmetres en funció dels dos esdeveniments E1 i E2.

PARÀMETRE	SIGNIFICAT
p_e1	Probabilitat d'observar l'esdeveniment E_1
p_e2	Probabilitat d'observar l'esdeveniment E_2
rho	Correlació entre E_1 i E_2
p0_e1	Probabilitat d'observar l'esdeveniment E_1 formant part del grup de control
p0_e2	Probabilitat d'observar l'esdeveniment E_2 formant part del grup de control
p1_e1	Probabilitat d'observar l'esdeveniment E_1 formant part del grup d'intervenció
p1_e2	Probabilitat d'observar l'esdeveniment E_2 formant part del grup d'intervenció
Type_e1	Efecte estadístic de l'esdeveniment E_1
Eff_e1	Valor de l'efecte de l'esdeveniment E_1
Type_e2	Efecte estadístic de l'esdeveniment E_2
Eff_e1	Valor de l'efecte de l'esdeveniment E_2
Effect_ce	Tipologia d'efecte estadístic
Alpha	Nivell de significació α
Beta	Nivell de significació β
Unpooled	Tipologia de variància utilitzada per calcular la grandària mostral

Taula 4: Nomenclatura de variables

3. Estructura final

En aquest apartat es descriuen les funcions que acabaran formant part del paquet, en total en seran quatre, en base a totes les postulacions argumentades anteriorment. Les noves funcions s'anomenen per blocs. Es descriu la seva composició i també es postulen tots els algorismes que fonamenten les funcions d'estudi.

- Probabilitat d'unió de E∗

S'ha mantingut l'estructura d'aquesta funció respecte a la seva original, ja que no existeix equivalència amb qualssevol de les altres funcions d'interès. S'ha incorporat la validació de les probabilitats, sempre han d'estar entre 0 i 1, la correlació donada entre els dos esdeveniments ha d'acotar-se dins l'interval que caracteritza les variables combinades. Així doncs, aquesta funció al igual que la seva predecessora acabarà tenint únicament tres paràmetres d'entrada: dos d'ells relacionats amb la probabilitat pk d'observar Ek i *rho* com la correlació entre esdeveniments.

A continuació es mostra el nom i l'algorisme de la funció que permet calcular la probabilitat d'unió.

Nom de la funció: *prob_ce*

Algorisme 1 Probabilitat d'unió de E∗

Sigui

p_e1 Probabilitat d'observar l'esdeveniment E1

p_e2 Probabilitat d'observar l'esdeveniment E2

rho Correlació entre E1 i E2

Procediment principal

If p_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E1 ha d'estar entre 0 i 1

else if p_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E2 ha d'estar entre 0 i 1

else if rho$\notin [m(p_1,p_2), M(p_1,p_2)]$ **then**

Stop La correlació ha d'estar acotada en el correcte interval

else then

Càlcul de la probabilitat d'unió de l'esdeveniment compost

end if

Fi

- Correlació de de E_*

Tot i les postulacions argumentades anteriorment, s'ha decidit no unificar en una sola les funcions que s'utilitzen per calcular la correlació de la variable combinada, ja que en algunes funcions es criden les funcions de correlació per tal de validar espais paramètrics. D'acord amb aquesta postulació existiran dues funcions que s'utilitzaran per calcular l'interval de confiança on s'acota la correlació de la variable combinada. S'ha validat que totes les probabilitats que s'utilitzen com a paràmetres d'entrada estesin acotades entre 0 i 1.

Seguidament es mostren els noms i els algorismes de les funcions que permeten calcular l'interval de correlació.

Nom de la funció: *lower_corr*

Algorisme 2 Rang inferior de E_*

Sigui

p_e1 Probabilitat d'observar l'esdeveniment E1

p_e2 Probabilitat d'observar l'esdeveniment E2

Procediment principal

If p_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E1 ha d'estar entre 0 i 1

else if p_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E2 ha d'estar entre 0 i 1

else then

Càlcul del rang inferior de correlació

end if

Fi

Nom de la funció: *upper_corr*

Algorisme 3 Rang superior de E∗

Sigui

p_e1 Probabilitat d'observar l'esdeveniment E1

p_e2 Probabilitat d'observar l'esdeveniment E2

Procediment principal

If p_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E1 ha d'estar entre 0 i 1

else if p_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E2 ha d'estar entre 0 i 1

else then

Càlcul del rang superior de correlació

end if

Fi

- Efectes de de E∗

Un cop conegudes les funcions originals utilitzades per calcular diversos efectes, es decideix unificar-les totes elles en una única funció. Aquesta funció permetrà calcular l'odds rati, el risc relatiu i la diferència de proporcions d'un esdeveniment compost. Unifica tres de les funcions originals que es coneixen sota la nomenclatura de:

- Or.composite
- Diff.pcomp
- Rr.pcomp

Tot i que els paràmetres d'entrada no són iguals en les tres funcions de referència, ja que la funció que calcula l'odds rati de l'esdeveniment compost requereix els valors dels odds rati dels esdeveniments diferents com a paràmetres d'entrada, després de tota la teoria postulada, s'ha pogut observar que aquest valor es pot calcular igualment, per tant existeix equivalència entre funcions.

Com a paràmetres d'entrada de la nova funció, trobem les probabilitats que es produeixi un determinat esdeveniment Ek per als individus que pertanyen al grup de control i aquells que reben el tractament que es testa, *rho* com la correlació entre esdeveniments.

Entre els paràmetres d'entrada també trobarem una variable qualitativa que acabarà indicant la tipologia d'efecte que es desitja calcular pel esdeveniment compost binari. Per defecte, si no s'especifica l'efecte desitjat, la funció retorna la diferència de proporcions.

S'han validat tots els paràmetres d'entrada requerits en l'execució d'aquesta funció. En la sortida de la funció es poden veure els dos efectes dels esdeveniments diferents entre sí i l'efecte de l'esdeveniment compost. S'ha considerat que la visualització dels tres efectes resulta interessant a nivell comparatiu. Seguidament podem veure un exemple de la sortida d'aquesta funció:

```
  Effect1   Effect2 EffectCE
32.86047 0.6825397 3.927383
```

Exemple 1:Sortida funció effect_ce

Aquesta nova funció que agrupa antigues funcions seguirà l'algorisme i es coneixerà sota la nomenclatura que es postula a continuació:

Nom de la funció: *effect_ce*

Algorisme 4 Efecte estadístic de E_*

Sigui

p0_e1 Probabilitat d'observar l'esdeveniment E1 pertanyen al grup de control

p0_e2 Probabilitat d'observar l'esdeveniment E2 pertanyen al grup de control

p1_e1 Probabilitat d'observar l'esdeveniment E1 pertanyen al grup d' intervenció

p1_e2 Probabilitat d'observar l'esdeveniment E2 pertanyen al grup d' intervenció

rho Correlació entre E1 i E2

effect_ce Tipologia d'efecte estadístic que es desitja calcular

Procediment principal

If p0_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_1 pels individus del grup de control ha d'estar entre 0 i 1

else if p0_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_2 pels individus del grup de control ha d'estar entre 0 i 1

else if p1_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_1 pels individus del grup d'intervenció ha d'estar entre 0 i 1

else if p1_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_2 pels individus del grup d'intervenció ha d'estar entre 0 i 1

else if rho$\notin \left[max\left(m(p_1^0, p_2^0), m(p_1^1, p_2^1)\right), min\left(M(p_1^0, p_2^0), M(p_1^0, p_2^0)\right)\right]$ **then**

Stop La correlació ha d'estar acotada en el correcte interval

else if effect_ce **not in** (' rr ',' or ',' diff ') **then**

Stop S'ha d'escollir entre odds rati, risc relatiu i diferència de proporcions

else then

if effect_ce **in** ('diff') **then**

Càlcul de l'efecte estadístic en funció de la diferència de proporcions

else if effect_ce **in** ('rr') **then**

Càlcul de l'efecte estadístic en funció del risc relatiu

else if effect_ce **in** ('or') **then**

Càlcul de l'efecte estadístic en funció de l'odds rati

end if

end if

Fi

- Grandària mostral de $E*$

S'ha decidit unificar les funcions de les quals es partia en una sola per tal de consolidar les postulacions argumentades anteriorment. Les funcions que s'han unificat són les següents:

- SampleSize.CBE.OR
- SampleSize.OR

- SampleSize.CBE.RR
- SampleSize.RR
- SampleSize.CBE.Diff
- SampleSize.Diff

En aquesta nova funció comptem amb els següents paràmetres d'entrada: les probabilitats de l'esdeveniment combinat associades al grup de control, la tipologia i el valor de l'efecte l'esdeveniment Ek , els coeficients alpha i beta, la tipologia de variància sota la qual es vol calcular la grandària mostral (pot ser agrupada o desagrupada).

Aquest plantejament de la funció permet donar a l'usuari una major llibertat alhora d'establir els paràmetres d'entrada de la funció ja que compte amb diferents opcions alhora de cridar-la. Altrament resulta interessant introduir els efectes estadístics com a paràmetres d'entrada ja que es coneix que la grandària mostral depèn dels seus valors.

Tot i que els paràmetres d'entrada entre totes les funcions descrites anteriorment no són idèntics, s'ha pogut trobar l'equivalència després de l'argumentació teòrica desenvolupada i tot l'estudi realitzat. Els paràmetres sota els quals es treballa es poden acabar obtenint entre sí.

Abans de calcular qualsevol grandària mostral, la mateixa funció valida tots els paràmetres d'entrada per validar tots els camps paramètrics i evitar errors. Per defecte, es calcula la grandària mostral en base a la diferència de proporcions, amb un llindar d' $\alpha\ i\ \beta$ fixat a 0.05 i 0.2 respectivament, i sota una variància desagrupada.

L'algorisme i la nomenclatura que fonamenten la funció que s'està descrivint es presenta a continuació:

Nom de la funció: *sample_size_ce*

Algorisme 5 Grandària mostral de $E*$

Sigui

p0_e1 Probabilitat d'observar l'esdeveniment E_1 pertanyen al grup de control

p0_e2 Probabilitat d'observar l'esdeveniment E_2 pertanyen al grup de control

type_e1 Tipologia d'efecte en funció de l'esdeveniment E1

eff_e1 Valor de l'efecte en funció de l'esdeveniment E1

type_e2 Tipologia d'efecte en funció de l'esdeveniment E2

eff_e2 Valor de l'efecte en funció de l'esdeveniment E2

effect_ce Tipologia d'efecte estadístic que s'agafa per calcular la grandària

mostral

rho Correlació entre E_1 i E_2

alpha Nivell de significació α

beta nivell de significació β

unpooled Tipologia de variància utilitzada

Procediment principal

If p0_e1 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_1 pels individus del grup de control ha d'estar entre 0 i 1

else if p0_e2 $\notin [0,1]$ **then**

Stop La probabilitat d'observar l'esdeveniment E_2 pels individus del grup de control ha d'estar entre 0 i 1

else if type_e1 **not in** (' rr ',' or ',' diff ') **then**

Stop S'ha d'escollir entre odds rati, risc relatiu i diferència de proporcions

else if (type_e1 = "diff" and eff_e1 > 0) or (type_e1 = "rr" and eff_e1 $\notin [0,1]$) or (type_e1 = "or" and eff_e1 $\notin [0,1]$) **then**

Stop L'efecte de l'esdeveniment E1 no es correcte

else if type_e2 **not in** (' rr ',' or ',' diff ') **then**

Stop S'ha d'escollir entre odds rati, risc relatiu i diferència de proporcions

else if (type_e2 = "diff" and eff_e2 > 0) or (type_2 = "rr" and eff_e2 $\notin [0,1]$) or (type_e2 = "or" and eff_e2 $\notin [0,1]$) **then**

Stop L'efecte de l'esdeveniment E2 no és correcte

else if effect_ce **not in** (' rr ',' or ',' diff ') **then**

Stop S'ha d'escollir entre odds rati, risc relatiu i diferència de proporcions

else if rho$\notin [m(p_1^0, p_2^0), M(p_1^0, p_2^0)]$ **then**

Stop La correlació ha d'estar acotada en el correcte interval

else if alpha $\notin [0,1]$ **then**

Stop El valor de alpha ha d'estar entre 0 i 1

else if beta $\notin [0,1]$ **then**

Stop El valor de beta ha d'estar entre 0 i 1

else if unpooled **not in** ('pooled','unpooled') **then**

Stop S'ha d'escollir entre variància agrupada i desagrupada

End

If type_e1 **in** ('or') **then**

Càlcul de p1_e1

else if type_e1 **in** ('rr) **then**

Càlcul de p1_e1

else if type_e1 **in** ('diff) **then**

Càlcul de p1_e1

End if

if type_e2 **in** ('or') **then**

Càlcul de p1_e1

else if type_e2 **in** ('rr) **then**

Càlcul de p1_e1

else if type_e2 **in** ('diff) **then**

Càlcul de p1_e1

end if

Càlcul de la probabilitat de l'esdeveniment compost pels individus del grup de control

Càlcul de la probabilitat de l'esdeveniment compost pels individus del grup d'intervenció

if effect_ce **in** ('rr) **then**

Càlcul del risc relatiu de l'esdeveniment compost

if unpooled **in** (‘ unpooled variance ’) **then**

Càlcul de la grandària mostral en funció del risc relatiu i tenint en compte una variància desagrupada

else then

Càlcul de la grandària mostral en funció del risc relatiu i tenint en compte una variància agrupada

end if

else if effect_ce **in** (‘or) **then**

Càlcul de l’odds rati de l’esdeveniment compost

if unpooled **in** (‘ unpooled variance ’) **then**

Càlcul de la grandària mostral en funció de l’odds rati i tenint en compte una variància desagrupada

else then

Càlcul de la grandària mostral en funció de l’odds rati i tenint en compte una variància agrupada

end if

else then

Càlcul de la diferència de proporcions de l’esdeveniment compost

if unpooled **in** (‘ unpooled variance ’) **then**

Càlcul de la grandària mostral en funció de la diferència de proporcions i tenint en compte una variància desagrupada

else then

Càlcul de la grandària mostral en funció de la diferència de proporcions i tenint en compte una variància agrupada

end if

end if

end if

Fi

4. Validació de l'estructura final

En aquest apartat es desenvolupen tots els testos complementaris realitzats per la validació de totes les funcions que han experimentat modificacions. Es considera un aspecte molt important a tenir en compte ja que la seva execució permet detectar possibles errors i mancances en l'estructura.

Aquest conjunt de validacions s'han pogut realitzar gràcies a la funció de l' R *testfile* de la llibreria *testthat*. Utilitzant la crida d'aquesta funció, s'aconsegueix validar el codi, és a dir, proporcionar mecanismes per tal d'assegurar que el resultat que proporcionen les funcions és l'esperat, així doncs permet detectar addicionalment errors. En el capítol 4 és desenvolupa més àmpliament els mecanismes que porta associada aquesta funció de forma tècnica en l'àmbit de l' R, aquí només en centrem en descriure com s'ha utilitzat aquesta funció en el codi construït.

Per a totes les funcions, s'ha utilitzat la mateixa metodologia. S'han generat variables de tipus numèric que representen els valors obtinguts en la crida de les diferents funcions, havent modificat els paràmetres d'entrada.

Seguidament es desenvoluparan algunes de les crides de funcions que s'han creat per validar el codi que fonamenta el paquet. En l'annex podem trobar l'accés a tot el codi creat

- Validació A

Si es crida alguna funció sense algun paràmetre i aquest paràmetre no està establert per defecte, ha de retornar un missatge d'error.

```
Sigui :
funcio_estudi <- function(a,b,c,d){
#Càlculs requerits
}
expect_error(funcio_estudi( a , b , c))
```

Es mostra un exemple suposant una funció d'estudi. Aquesta funció llegeix quatre paràmetres d'entrada: a, b, c i d. Si algun d'aquests paràmetres falta en l'execució de la mateixa funció, s'espera que la funció no es pugui executar. Per tant, la funció *expecterror* ha de detectar que hi ha un error en la funció i per tant no ha de permetre la seva execució.

- Validació B

Si es crida alguna funció que conté uns paràmetres d'entrada inapropiats, s'ha d'obtenir el missatge de warning.

```
Sigui :
    funcio_estudi <- function(a,b,c,d){
      #Càlculs requerits
    }
```

on els paràmetres entrada han de ser inferiors a 1

```
  expect_that(funcio_estudi(3,0.4,0.4,0.5), throws_error("Algun
valor no es correcte"))
```

S'espera que la funció *expectthat* detecti un error i acabi retornant un KO ja que un dels paràmetres d'entrada de la funció d'estudi no compleix la condició requerida, els valors han de ser inferior a 1, i per tant existeix un error.

- Validació C

A partir de les variables generades en les crides de funcions, es valida que algunes de les variables valguin el mateix, ja que tenen els mateixos paràmetres d'entrada. Com que no totes aquestes variables compleixen aquest requisit, es valida també que les variables que es fonamenten de diferents paràmetres no obtinguin el mateix resultat. Per tal de validar aquest conjunt de validacions de coherència en totes les funcions d'estudi, s'ha creat una funció que s'anomena *condició* que permet validar de forma lògica si els resultats obtinguts per diferentes vies són coherents o no. En la validació s'utilitzen les funcions *expectequal* o bé *expectfalse*. Seguidament podem veure un exemple del codi que fonamenta aquest conjunt creat:

```
condition <- function(a,b,c){
    i <- 3
    if(c == "T"){
      expect_true(a[i,] == b[i,])
    }else if(c == "F"){
      expect_false(a[i,] == b[i,])
    }
  }
```

Crides de la funció de validació:

```
condition(a,b,"T")
condition(a,c,"F")
```

- Validació D

Es coneix que totes les funcions creades pel paquet han de retornar un valor numèric, per tant es valida que totes les variables generades siguin de caràcter numèric. S'utilitza la funció *expectthat* .

```
expect_that(funció_estudi, is_a("numeric"))
```

Tot el codi que s'ha construït per validar el conjunt de funcions es troba en l'annex.

IV. FONAMENTS COMPUTACIONALS

Per a la creació d'un paquet informàtic és molt important tenir en compte el codi que engloba les funcions i s'ha de donar una gran èmfasi a la documentació que es genera per tal de facilitar la feina a altres usuaris externs un cop la llibreria sigui compartida. Ajudarà a la correcta interpretació de la feina realitzada.

En aquest capítol IV, es desenvoluparan tots els passos que s'han realitzat per crear l'entorn en l'R on s'allotjarà el paquet. Així doncs, també s'explicarà on s'organitza el codi que recull el conjunt de funcions creades i com s'han creat tots els arxius que documenten i suporten aquest conjunt de funcions.

També es vol plasmar l'adaptació del projecte als requeriments del CRAN ja que un dels objectius principals postulat en aquest treball es basa en l'adaptació del paquet en el repositori citat. Tots els arxius creats i avaluats han d'estar en anglès, és per aquest motiu que creem els arxius en aquesta llengua.

1. Creació de l'entorn

Hi ha una gran quantitat d'articles i manuals que expliquen com ha de ser la creació d'un paquet dins l' R. En l'estudi d'aquests arxius, es pot observar que no existeix una única metodologia que permeti aquesta creació, existeixen diverses vies per a fer-ho. Per començar s'ha agafat com a referència un article de recerca científic (Leisch, 2009).

Per tal d'iniciar la preparació, caldrà obrir l' R tal i com s'acostuma a fer normalment. Dins del programa R, trobem l'opció de creació d'un nou projecte. Tenir un espai propi de treball per a cada projecte que es desitja desenvolupar dóna a l'usuari un major control independent, ja que no s'intercanvia codi de diferent interès d'estudi. Cliquem i comencem a crear aquest nou projecte on s'emmagatzemarà tota la informació que necessitem.

L'R permet la creació de nous projectes però també té en compte que és una eina computacional preparada per generar nous paquets que posteriorment s'instal·laran al CRAN o s'utilitzaran en altres entorns entre d'altres funcions que té, per tant a l'hora de crear el nou projecte, creem directament un projecte que està habilitat per allotjar paquets. És a dir, el propi R dóna flexibilitat per escollir la tipologia de projecte que es vol crear. Seguidament podem visualitzar la diversitat de nous projectes que hi ha en l' R per escollir.

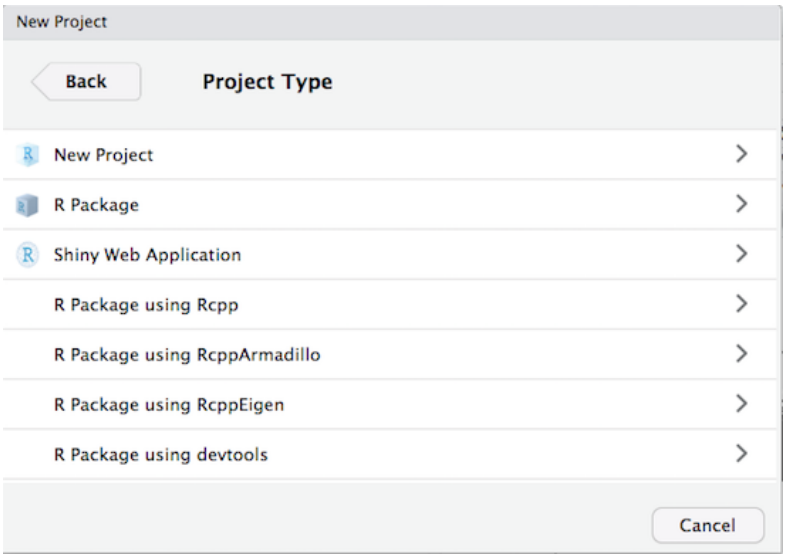

Figura 4: Tria del projecte

Triem l'opció *R package* que ens permet generar un paquet en un projecte de l' R.

Posteriorment, ens demana escollir el directori on s'allotjarà el nostre projecte i el nom sota el qual ens volem referir al projecte. Dit en altres paraules, aquest directori serà una carpeta on s'aniran guardant tots els arxius que es generaran, ens demana la ruta on volem tenir aquesta carpeta.

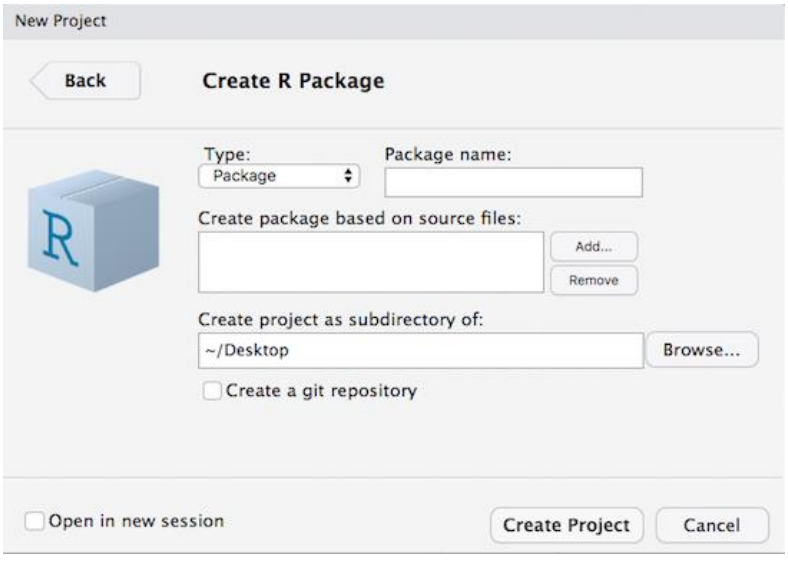

Figura 5: Configuració inicial

Ens referirem al nostre projecte sota les sigles de CBE, que provenen de *Composite binari endpoints*, esdeveniments compostos binaris en anglès. El nostre directori de moment serà l'escriptori, tot i que la carpeta d'estudi es podrà moure de ruta sense problema, fet que facilitarà la correcció i validació del projecte.

2. Preparació de l'entorn

Un cop disposem del propi entorn, el següent pas serà la instal·lació d'alguns paquets disponibles al CRAN que ens ajudaran en la construcció del paquet:

- Paquet Devtools

La instal·lació d'aquest paquet és completament necessària ja que permet compilar i construir paquets. Permet realitzar tasques comunes necessàries en l'execució d'algunes funcions, entre d'altres la càrrega del paquet gràcies a la funció *load all*. El nou projecte, format per un conjunt d'arxius que contenen el codi R, requereix d'aquest paquet per a la seva compilació.

- Paquet Roxygen2

Tal i com s'ha mencionat anteriorment, la documentació resulta molt necessària en la creació del paquet, ja que les funcions s'emmagatzemen en un projecte amb el principal objectiu de poder ser compartides i per tant l'usuari que les utilitzi les interpretarà seguint la documentació publicada. La instal·lació d'aquest paquet permet documentar les funcions fàcilment i tota la informació restant requerida en el paquet, i permet incorporar tot tipus de comentaris que simplifiquen aquesta tasca.

Posteriorment a la instal·lació dels dos paquets descrits anteriorment, s'ha executat en l'entorn la funció *has devel* del propi paquet devtools per poder validar que estem treballant amb la versió adient de l' R. Si la funció retorna TRUE la versió de l' R estarà preparada per començar a treballar, en canvi, si rebem FALSE caldrà instal·lar la nova versió de l' R.

3. Creació de l'estructura de l'entorn

Un cop ja tenim l'entorn a punt podem començar amb la preparació de tots els arxius que es necessiten per el nostre projecte. Seguidament es descriu la metodologia que s'ha seguit alhora de crear tots els arxius requerits.

3.1. Descripció

L'arxiu *description* conté informació molt important a tenir en compte, tant en el paquet en construcció com en tots els paquets creats, ja que conté la informació clau. De fet, guanya importància a mesura que es van reactualitzant noves versions de qualsevol paquet creat. Resulta un arxiu obligatori en tot paquet.

A continuació es descriu breument cadascun dels estaments requerits per a la descripció i en l'annex podem trobar la pròpia descripció.

- *Package*

Nom sota el qual ens referim al projecte d'estudi, en el nostre cas utilitzem les sigles CBE que es fonamenten en el terme de *Composite binary endpoints*, esdeveniments binaris compostos en angles.

- *Type*

Tipologia de projecte en construcció, en el nostre cas estem creant un paquet informàtic.

- *Title*

Breu descripció del paquet que no pot superar un nombre fixat de caràcters. En el nostre cas, definim que el paquet es construeix a partir d'un conjunt de funcions que calculen mesures estadístiques per als esdeveniments binaris compostos.

- *Version*

A part de representar un comptador en base a les variacions que ha experimentat el paquet respecte la seva versió oficial, s'utilitza la versió de qualsevol paquet per transmetre les modificacions que s'han realitzat versió per versió, via codificació numèrica. En el nostre cas, definim 0.0.0.9000 la nostra primera versió del paquet, ja que aquesta codificació indica que el paquet està en desenvolupament.

- *Author*

Els autors del paquet són: Raquel Rovira, Marta Bofill i Jordi Cortés.

- *Maintainer*

Contacte per mantenir consultes sobre el paquet. Posteriorment, si es modifica i es crea una nova versió d'aquest paquet, es podrà modificar sense problemes. Quan és defineixen els autors del projecte anàlogament es defineix quin d'ells agafarà aquest rol.

- *Description*

La descripció es més detallada que el títol, poden haver-hi vàries frases però la mida ideal no hauria de superar el paràgraf. En aquest apartat es dóna una breu descripció de les funcions.

- *Lazydata*

Facilita la incorporació de dades en el paquet, assignem TRUE per facilitar la incorporació de dades en el paquet que estem creant.

- *License*

Assignem la llicència GPL-3 per tal que el nostre paquet pugui ser compartit amb la resta d'usuaris. Conté una única restricció: cada cop que es comparteixi el projecte, també s'ha de compartir la llicència.

3.2. Documentació de les funcions

Documentar les funcions és un dels aspectes claus que s'ha de realitzar amb molta cura, ja que tal i com s'ha exposat anteriorment, els usuaris externs que vulguin utilitzar aquest conjunt de funcions es basaran en aquesta documentació a l'hora de conèixer la seva funcionalitat.

Dins del directori hi ha una carpeta buida que s'anomena R on s'han d'emmagatzemar aquest conjunt de funcions. En aquesta carpeta hi ha tants arxius com funcions finals hem obtingut amb el nom de les respectives funcions, per tant en aquesta carpeta tenim la següent estructura:

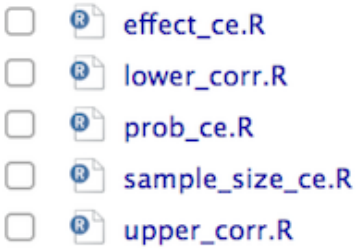

Figura 6: Estructura carpeta R

Gràcies al paquet de *roxygen2* instal·lat anteriorment, s'ha facilitat la documentació de totes les funcions del paquet.

Mitjançant la plantilla que es mostra a continuació, plantilla que conté anotacions per veure clarament que requereix cada ítem, s'ha pogut documentar fàcilment el conjunt de funcions. La documentació de totes les funcions es pot trobar en l'annex, aquí es mostra únicament la plantilla que s'ha seguit per tal de poder descriure la metodologia seguida.

```
#'  TITOL DE LA FUNCIÓ
#'  Breu descripci de la funció
#'  @param a parametre d'entrada
#'  @param b parametre d'entada
#'  @return c valor retornat
#'  @description descripció de la funcio
#'  @details Detalls de la funcio
#'  @example exemple de la funcio
#'  @export
exemple <- function(a,b){
#Realitzacio dels calculs que requereix la funcio
return(c)
}
```

3.3. Carpeta Man

Alhora de crear el conjunt d'arxius que s'allotgen en la carpeta *man* s'utilitza la documentació creada per les funcions. Mitjançant el paquet *roxygen* instal·lat prèviament en els primers processos, quan es compila el paquet tota la documentació creada per documentar les funcions es transforma i acaba configurant els arxius de la carpeta *man* on s'estableix el format que R acabarà visualitzant. Observem el codi que es presenta a continuació.

```
\name{Nom de la funció}
\alias{Alias de la fució}
\title{Títol de la funció}
\usage{
exemple(a, b, c)
}
\arguments{
\item{a}{Paràmetre d'entrada de la funció}
\item{b}{Paràmetre d'entrada de la funció}
\item{c}{Paràmetre d'entrada de la funció}
}
\value{
Valor que retorna la funció d'exemple
}
\description{
Descripció de la funció
}
\details{
Detalls de la funció
}
\examples{
exemple(a,b,c)
}
```

Aquest codi plasma l'estructura interna de tots els arxius que podem trobar en la carpeta *man*. El codi postulat pren relació amb l'exemple de l'apartat 3.2. Aquest format es crea automàticament. En el codi annex podem trobar els exemples de visualització de les funcions creades.

3.4. Namespace

Dins del CRAN, entorn on s'allotja una gran quantitat de paquets i llibreries, es probable trobar vàries funcions que reben el mateix nom. Aquest fet pot originar un conflicte, ja que pot trobar-se la situació en la qual es vulgui cridar una funció d'una certa llibreria A, però que en realitat s'estigui cridant una funció de la llibreria B, ja que són funcions que es coneixen amb el mateix nom. Per evitar aquesta tipologia d'errors s'acostuma a cridar una funció mencionant el paquet al qual pertany a través de la següent crida: *pck:function*.

Aquest arxiu conegut com *namespace* ajuda a minimitzar aquesta tipologia d'errors ja que permet associar les funcions que han estat creades al nostre paquet. Dins d'aquest arxiu doncs, es definiran totes les funcions que es volen exportar i compartir amb els

altres usuaris externs. Totes aquelles funcions que no siguin exportades seran internes i només es podran utilitzar dins del propi paquet.

Per tant, via aquest arxiu podem exportar funcions creades per nosaltres mateixos per a què altres usuaris que no treballen amb el nostre projecte les puguin utilitzar-les, mitjançant la funció *export*, o bé, importar funcions que pertanyen a altres llibreries del CRAN per tal de poder utilitzar-les localment dins del nostre projecte a través de la funció *import*.

Per al nostre projecte, aquest arxiu únicament exportarà les funcions que s'han creat, ja que no s'ha utilitzat en cap cas cap funció que pertany a altres llibreries. El codi generat per aquest arxiu es troba a continuació.

```
export(prob _ ce)
export(lower _ corr)
export(upper _ corr)
export(effect _ ce)
export(sample _ size _ ce)
```

4. Validació de l'estructura

Cal tenir present que a mesura que es generen tots els arxius necessaris per a la creació del paquet, poden generar-se també errors funcionals o errors en l'estructura creada. Per aquest motiu, és molt important validar tot el que s'està generant paral·lelament a la seva construcció ja que permet minimitzar els errors que es puguin ocasionar.

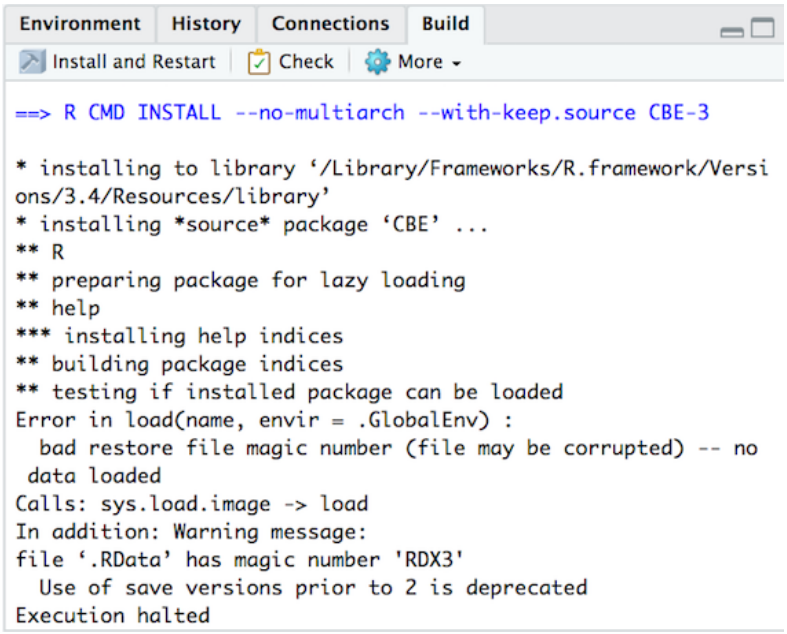

Figura 7: Sortida de la validació

En la imatge superior, podem observar un mecanisme que permet instal·lar i anar

validant el projecte que s'està construint paral·lelament a la feina que s'està realitzant per construir-lo. Gràcies al paquet instal·lat anteriorment, *devtools*, es pot realitzar aquesta validació sistemàticament.

Aquesta validació va testejant tots els arxius que s'han generat i valida també la instal·lació del propi paquet per tal d'assegurar que es compleixen tots els requisits del CRAN. Es validen punt per punt tots els arxius creats en el projecte i descrits anteriorment. A més a més també permet verificar aspectes interns del propi projecte. A continuació es mencionen tots els aspectes que es validen.

- La instal·lació complerta del paquet
- Descripció del paquet
- *Namespace* del paquet
- Funcions del paquet
- Documentació de les funcions del paquet

Aquest conjunt a validar pot finalitzar en qualsevol dels següents estats:

- Estat d'error : s'ha identificat algun error en algun dels arxius que s'han validat.
- Estat d'alerta : s'ha identificat algun arxiu que podria acabar generant un error, per tant resultaria interessant revisar el seu contingut.
- Estat correcte : tots els arxius són vàlids i no requereixen modificació.

Si algun dels arxius validats finalitza en error o alerta, cal modifica'l i tornar a validar el seu contingut fins el punt d'assolir un estat correcte per a tots els arxius creats.

5. Validació del paquet

Independentment de la validesa dels arxius creats, és important validar la lògica de totes les funcions, ja que pot ser que una funció s'acabi aprovant per la seva correcta funcionalitat i estructura però que en el fons no realitzi la tasca que desitgem per mancances en la lògica de la funció.

Per validar aquest punt ens serà de gran utilitat la llibreria *testthat* ja que conté un conjunt de funcions que permeten la creació d'una sèrie de testos de validació. Cal mencionar que aquestes funcions no retornen un valor numèric o una cadena de caràcters com estem acostumats a observar, sinó que retornen un OK o un KO. Si la lògica que s'ha introduït com a paràmetres d'entrada es compleix, rebrem un OK, en cas contrari rebrem un KO.

I com podem generar un test utilitzant aquestes funcions? Doncs bé, primer cal que coneguem algunes de les funcions que conté aquesta llibreria per poder entendre la seva funcionalitat. Seguidament definim algunes d'aquestes funcions que utilitzarem a través d'exemples.

- Funció *expect equal* i *expect true*

Amb la crida d'aquestes funcions s'espera comprovar que els paràmetres que contenen en les entrades són iguals.

```
expect_true(2, 2) #A
expect_true(2, 4) #B
expect_equal(2, 1+1) #C
```

Siguin A, B i C tres crides de funcions amb diferents paràmetres d'entrada, s'espera que la crida de A i C retorni un OK ja que els paràmetres d'entrada són iguals o equivalents. D'altra banda, la crida de la funció B ha de generar un KO ja que 2 és diferent a 4.

- Funció *expect false*

Amb la crida d'aquestes funcions s'espera comprovar que els paràmetres que contenen en les entrades són diferents.

```
expect_false(86,86) #A
expect_false(68,86) #B
```

Siguin A i B dos crides de la funció *expect false* s'espera que la crida A retorni un KO ja que 86 és igual a 86, però la crida B hauria de retornar OK ja que 86 és diferent de 68, els dos paràmetres d'entrada són diferents.

- Funció *expect error*

Amb la crida d'aquestes funcions s'espera localitzar errors comesos en la crida de funcions.

```
suma <- function(a,b){
 return(a+b)
}
expect_error(suma(2)) #A
expect_error(suma(2+3)) #B
```

Sigui una funció creada que retorna la suma dels paràmetres que rep per entrada. S'espera que la crida A retorni un error KO ja que en la funció falta un paràmetre d'entrada, però la crida B s'hauria d'executar sense problemes i per tant generar un OK.

Després d'entendre la funcionalitat de les funcions, ja podem centrar-nos en conèixer com es duen a terme els tests. De fet, l'execució de crides independents ja resulta un test, ja que s'està testejant una lògica que es desitja comprovar. Però, seguidament observarem com es crea un test que engloba un conjunt de crides, ja que es considera quan s'engloben crides es té un major control de la validació del codi que s'està testejant.

A continuació, veiem un test que valida lògicament un conjunt fixat. Per crear aquest test s'han agafat algunes crides descrites anteriorment que esdevindrien un OK. És a dir, tenim un test d'exemple on tota la lògica és correcte. No hem incorporat cap KO per poder visualitzar la sortida de la crida de la funció ideal a aconseguir.

```
context("test-CBEval1")
 test_that("test-CBEval1␣works", {
expect_equal(2, 1+1) #C suma <- function(a,b){
return(a+b) }
expect_error(suma(2)) #A
expect_error(suma(2+3)) #B
}
```

Mitjançant la funció *test file* podem executar els testos construïts i així validar tota la lògica creada a través de la sortida obtinguda.

6. Utilització del paquet

El GitHub, entre d'altres funcions, és coneix com un recurs informàtic que s'utilitza com a repositori de codi, de forma pública o privada. S'utilitza per controlar i modificar codis que fonamenten projectes que es troben en la xarxa. Facilita així la descarrega i la utilització del codi que fonamenta cada projecte.

El paquet creat, un cop validat que compleix tots els requisits del CRAN, s'allotja en el repositori del GitHub per facilitar que diversos navegants el puguin utilitzar. El podem trobar en el següent enllaç:

https://github.com/CompARE-Composite/CompARE-package

En aquest mateix enllaç podem trobar una breu descripció on s'explica breument les funcions que fonamenten el paquet i les instruccions que cal seguir per poder instal·lar el paquet des de R.

V: CAS PRÀCTIC

En aquesta secció, es donarà un exemple numèric de la teoria desenvolupada en el capítol II en referència a la probabilitat d'unió de dos esdeveniments, la correlació de la variable combinada i les mesures dels efectes estadístics. Es suposen els valors d'alguns paràmetres que són desconeguts per poder obtenir resultats. Es vol donar un exemple numèric dels paràmetres que es requereixen anticipadament en el càlcul de la grandària mostral.

Observem la taula 5 on suposem els valors que prenen les probabilitats d'observació dels esdeveniments E_k en funció del grup d'estudi al qual pertanyen els individus. D' acord amb el disseny que s'està desenvolupant la probabilitat associada al grup control d'observació de l'esdeveniment E_k sempre serà superior a la probabilitat associada al grup d'intervenció.

PARÀMETRE	VALOR
p_1^0	0.46
p_1^1	0.3
p_2^0	0.43
p_2^1	0.2

Taula 5: Paràmetres d'observació donats

Un altre paràmetre que es requereix per dur a terme aquesta secció és el valor de la correlació de Pearson entre els esdeveniments E_1 i E_2. No es pot realitzar el càlcul d'aquest paràmetre, resulta desconegut, tot i que podem calcular l'interval on s'acota. Assumint que la correlació és la mateixa en ambdós grups, definim l'interval de la correlació definides les probabilitats p_k^icom:

$$rho\ \epsilon\ \left[max\left(m(p_1^0, p_2^0), m(p_1^1, p_2^1)\right), min\left(M(p_1^0, p_2^0), M(p_1^1, p_2^1)\right)\right] \subseteq [-1,1]$$

A través de les equacions 8 i 9, conjuntament amb els paràmetres de la taula 5 s'obté l'interval on s'acota la correlació:

$$rho\ \epsilon\ [-0.32; 0.76]$$

Es decideix postular tres possibles escenaris sobre els quals podem postular un valor pel paràmetre de la correlació.

i. La correlació és mínima i equival al rang inferior de l'interval

$$rho = -0.32$$

ii. La correlació és nul·la, val zero.

$$rho = 0$$

iii. La correlació és máxima i equival al rang superior de l'interval

$$rho = 0.76$$

En tots els càlculs on es requereixi el paràmetre de la correlació de Pearson, és calcularà el resultat de l'equació en base als tres escenaris postulats. S'obtindran per a cada cas tres resultats diferents.

Calculem les probabilitats d'unió de l'esdeveniment compost pels individus que formen part del grup control i del grup d'intervenció mitjançant l'equació 7. Observem la següent taula:

PROBABILITAT D'UNIÓ	$\boldsymbol{p_*^1}$	$\boldsymbol{p_*^0}$
$\boldsymbol{rho}$ **= - 0.32**	0.49	0.76
$\boldsymbol{rho}$ **= 0**	0.44	0.69
$\boldsymbol{rho}$ **= 0.76**	0.30	0.51

Taula 6: Càlculs de la probabilitat d'unió de l'esdeveniment compost

Observant els resultats de la taula 6 podem concloure que la probabilitat associada a la variable combinada és més forta quan el valor de la correlació entre esdeveniments s'aproxima al rang inferior de correlació. En els tres casos trobem que la probabilitat d'unió de l'esdeveniment compost en el grup control és superior a la del grup d'intervenció.

Anticipats els paràmetres de la taula 5, es calcula l'odds rati, el risc relatiu i la diferència de proporcions dels esdeveniments E_1 i E_2 ; d'acord amb les equacions 11, 13 i 15. En la taula 7 podem observar els resultats obtinguts.

EFFECTES	**Esdeveniment E_1**	**Esdeveniment E_2**
OR_K	0.50	0.33
RR_K	0.65	0.46
$\boldsymbol{\delta_k}$	-0.16	-0.23

Taula 7: Càlculs dels efectes dels esdeveniments E_1 i E_2

Observant els resultats dels efectes postulats en la taula 7 podem destacar com els efectes associats al esdeveniment E_1 en valor absolut són superiors en els tres casos als efectes de l'esdeveniment E_2.

A través de la taula 7 podem conèixer els valors dels efectes dels esdeveniments E_1 i E_2

que requereixen ser anticipats en el càlcul de l'odds rati, el risc relatiu i la diferència de proporcions de l'esdeveniment compost. Utilitzant també els valors de la taula 5 i suposant els tres escenaris per a la correlació, calculem les respectives mesures estadístiques associades a $E*$. Utilitzarem les equacions 12 , 14 i 16.

EFECTES $E*$	**OR_K**	**RR_K**	δ_k
rho**= -0.32**	0.31	0.64	-0.28
rho**= 0**	0.34	0.63	-0.25
rho**= 0.76**	0.39	0.59	-0.21

Taula 8: Càlculs dels efectes de la variable composta

A partir de la taula 8 podem observar com els valors dels odds rati i la diferencia de proporcions de la variable combinada augmenten en valor absolut quan el coeficient de correlació també augmenta. El risc relatiu calculat sobre la variable combinada té un comportament diferent, com més petit sigui el paràmetre de correlació major serà el valor de l'efecte.

VI. DISCUSSIÓ

Amb la realització d'aquest treball he pogut aprendre nous conceptes i noves eines que s'utilitzen en entorns computacionals.

Per una banda, el desenvolupament teòric realitzat en aquest estudi ha permès observar un conjunt d'avantatges i limitacions dels dissenys experimentals que utilitzen variables combinades. La tipologia de variable estudiada permet associar múltiples esdeveniments en una única variable d'interès. Els esdeveniments primaris amb poca significació o aquells que es caracteritzen per tenir una taxa d'observació molt reduïda es poden incorporar en els assajos clínics gràcies a la utilització de les variables combinades. Els esdeveniments primaris que composen la variable combinada han de tenir una importància clínica significativa i molt similar, ja que en base a la unió de totes elles es podria evidenciar una desigualtat entre variables primàries si no es complís aquesta característica. L'heterogeneïtat entre variables acabaria confonent els resultats obtinguts en l'assaig.

D'altra banda, s'ha pogut assolir l'objectiu principal que fonamentava la tesi que s'ha presentat. Gràcies als coneixements adquirits en el desenvolupament teòric del treball, s'ha pogut entendre la funcionalitat d'un conjunt de funcions i anàlogament s'han pogut aplicar modificacions en totes elles, finalitzant el procés amb la creació d'un paquet en l'R. La feina realitzada resulta de gran utilitat en el camp de la recerca clínica, ja que permet calcular diverses mesures estadístiques d'interès.
Aquest paquet creat pren molta relació amb la plataforma web del CompARE, ja que ambdós es poden utilitzar en la fase de planificació dels assajos clínics que compten amb variables combinades. De fet, tota la informació que s'ha agafat de referència per desenvolupar el treball prové del mateix projecte que la plataforma del CompARE.

VII. VALORACIÓ PERSONAL

Treballar en empreses on l'arquitectura i la computació de les dades representa un gran motor econòmic i un factor molt important en la presa de decisions, va despertar en mi una gran motivació. Volia conèixer i no parar d'aprendre què permet realitzar aquest camp i tota la potència que porta associada. Cada vegada resulta més comú sentir a parlar de la revolució de les dades.
Els objectius sota els quals s'ha postulat aquesta tesi es poden relacionar directament amb la computació de les dades. Des d'un primer moment vaig tenir molt clar que volia relacionar la temàtica del treball de fi de grau amb la meva principal motivació en el món laboral.

Cada dia durant la meva jornada laboral, estic acostumada a treballar amb Oracle, un famós programa del món empresarial. Crec que la realització d'aquesta tesi on he hagut de treballar amb l'R m'ha ajudat a ampliar la meva visió en el camp de la programació. M'agradaria continuar coneixent diversos programes informàtics que permeten treballar amb dades ja que n'existeixen múltiples. Amb aquest treball he pogut conèixer coses noves i també he pogut aprendre noves metodologies amb les quals no estava acostumada a treballar.

Durant la realització del treball he pogut veure tot el que es podia implementar. De fet, una de les principals dificultats que m'he trobat ha estat limitar el treball ja que durant la seva realització m'han sorgit moltes idees a desenvolupar, però l'espai temporal amb el qual comptava ha resultat ser una limitació.
Una de les coses que m'hagués agradat fer un cop coneguts tots els fonaments teòrics que sustenten les variables combinades, hauria estat suposar un conjunt de dades per tal de poder utilitzar-les en la crida de les funcions i així poder graficar els resultats, ja que seria addicionalment una eina interpretativa més clara.

VIII. REFERÈNCIES

- Bahadur, R. (1961). *A representation of the joint distribution of responses to n dichotomous items.* Palo Alto, CA: Stanford University Press.

- Gómez, L., & Bofill, M. (2017). *Selection of composite binary endpoints in clinical trials.* Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya.

- Gómez, L., & Bofill, M. (2018). *A new approach for sizing trials with composite binary endpoints using anticipated marginal values and accounting for the correlation between components.* Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya.

- Leisch, F. (2009). *Creating R packages: A Tutorial.* Department of Statistics, Ludwig-Maximilians-Universität München.

- Prentice, R. (1988). *Correlated binary regression with covarites specific to each binary observation.* Biometrics.

- Wickham, H. (2011). *testthat: Get Started with Testing.* Department of Statistics, Race University.

IX. ANNEXOS

1. Annex 1: Justificació capítol V

En aquest primer annex podem trobar el desenvolupament de tots els càlculs realitzats per obtenir els resultats numèrics del capítol V en base a les equacions del capítol II.

I. Cota inferior de l'interval de correlació de la variable combinada

$$Cota\ inferior = max\left(m(p_1^0, p_2^0), m(p_1^1, p_2^1)\right) = \ldots$$

$$max\left(max\left\{-\sqrt{\frac{p_1^0 * p_2^0}{q_1^0 * q_2^0}}, -\sqrt{\frac{q_1^0 * q_2^0}{p_1^0 * p_2^0}}\right\}, max\left\{-\sqrt{\frac{p_1^1 * p_2^1}{q_1^1 * q_2^1}}, -\sqrt{\frac{q_1^1 * q_2^1}{p_1^1 * p_2^1}}\right\}\right) = \ldots$$

$$max\left(max\left\{-\sqrt{\frac{0.46*0.43}{(1-0.46)*(1-0.43)}}, -\sqrt{\frac{(1-0.46)*(1-0.43)}{0.46*0.43}}\right\}, max\left\{-\sqrt{\frac{0.3*0.2}{(1-0.3)*(1-0.2)}}, -\sqrt{\frac{(1-0.3)*(1-0.2)}{0.3*0.2}}\right\}\right) = \ldots$$

$$max(max\{-0.80, -1.24\}, max\{-0.32, -3.05\}) = \ldots$$

$$max(-0.80, -0.32) = \textit{-0.32}$$

II. Cota superior de l'interval de correlació de la variable combinada

$$Cota\ superior = min\left(M(p_1^0, p_2^0), M(p_1^1, p_2^1)\right) = \ldots$$

$$min\left(min\left\{+\sqrt{\frac{p_1^1 * q_2^1}{q_1^1 * p_2^1}}, +\sqrt{\frac{p_2^1 * q_1^1}{q_2^1 * p_1^1}}\right\}, min\left\{+\sqrt{\frac{p_1^1 * q_2^1}{q_1^1 * p_2^1}}, +\sqrt{\frac{p_2^1 * q_1^1}{q_2^1 * p_1^1}}\right\}\right) = \ldots$$

$$min\left(min\left\{+\sqrt{\frac{0.46*(1-0.43)}{(1-0.46)*0.43}}, +\sqrt{\frac{0.43*(1-0.46)}{(1-0.43)*0.46}}\right\}, min\left\{+\sqrt{\frac{0.3*(1-0.2)}{(1-0.3)*0.2}}, +\sqrt{\frac{(1-0.3)*0.2}{0.3*(1-0.2)}}\right\}\right) = \ldots$$

$$min(min\{+1.06, +0.94\}, min\{+1.30, +0.76\}) = \ldots$$

$$min(+0.94, +0.76) = \textit{0.76}$$

III. Probabilitat d'unió: Grup intervenció, correlació mínima

$$p_*^1 = 1 - q_1^1 * q_2^1 - rho\sqrt{p_1^1 * p_2^1 * q_1^1 * q_2^1} = \ldots$$

$$1 - (1 - 0.3) * (1 - 0.2) - (-0.32) * \sqrt{0.3 * 0.2 * (1 - 0.3) * (1 - 0.2)} = \ldots$$

$$\textit{0.44 - (-0.05) = 0.49}$$

IV. Probabilitat d'unió: Grup intervenció, correlació nul·la

$$p_*^1 = 1 - q_1^1 * q_2^1 - rho\sqrt{p_1^1 * p_2^1 * q_1^1 * q_2^1} = \ldots$$

$$1 - (1 - 0.3) * (1 - 0.2) - (0) * \sqrt{0.3 * 0.2 * (1 - 0.3) * (1 - 0.2)} = \textit{0.44}$$

V. Probabilitat d'unió: Grup intervenció, correlació màxima

$$p_*^1 = 1 - q_1^1 * q_2^1 - rho\sqrt{p_1^1 * p_2^1 * q_1^1 * q_2^1} = \ldots$$

$$1 - (1 - 0.3) * (1 - 0.2) - (0.76) * \sqrt{0.3 * 0.2 * (1 - 0.3) * (1 - 0.2)} = \ldots$$

$$0.44 - (0.14) = 0.3$$

VI. Probabilitat d'unió: Grup control, correlació mínima

$$p_*^0 = 1 - q_1^0 * q_2^0 - rho\sqrt{p_1^0 * p_2^0 * q_1^0 * q_2^0} = \ldots$$

$$1 - (1 - 0.46) * (1 - 0.43) - (-0.32) * \sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} = \ldots$$

$$0.69 - (-0.07) = 0.76$$

VII. Probabilitat d'unió: Grup control, correlació nul·la

$$p_*^0 = 1 - q_1^0 * q_2^0 - rho\sqrt{p_1^0 * p_2^0 * q_1^0 * q_2^0} = \ldots$$

$$1 - (1 - 0.46) * (1 - 0.43) - (0) * \sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} = 0.69$$

VIII. Probabilitat d'unió: Grup control, correlació màxima

$$p_*^0 = 1 - q_1^0 * q_2^0 - rho\sqrt{p_1^0 * p_2^0 * q_1^0 * q_2^0} = \ldots$$

$$1 - (1 - 0.46) * (1 - 0.43) - (0.76) * \sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} = \ldots$$

$$0.69 - (0.18) = 0.51$$

IX. Odds rati de l'esdeveniment E_1 i E_2

$$OR_1 = \frac{p_1^{(1)} / q_1^{(1)}}{p_1^{(0)} / q_1^{(0)}} = \frac{0.3/(1-0.3)}{0.46/(1-0.46)} = \frac{0.42}{0.85} = 0.50$$

$$OR_2 = \frac{p_2^{(1)} / q_2^{(1)}}{p_2^{(0)} / q_2^{(0)}} = \frac{0.2/(1-0.2)}{0.43/(1-0.43)} = \frac{0.25}{0.75} = 0.33$$

X. Risc relatiu de l'esdeveniment E_1 i E_2

$$RR_1 = \frac{p_1^{(1)}}{p_1^{(0)}} = \frac{0.3}{0.46} = 0.65$$

$$RR_2 = \frac{p_2^{(1)}}{p_2^{(0)}} = \frac{0.2}{0.43} = 0.46$$

XI. Diferència de proporcions de l'esdeveniment E_1 i E_2

$$\delta_1 = p_1^{(1)} - p_1^{(0)} = 0.3 - 0.46 = -0.16$$

$$\delta_1 = p_2^{(1)} - p_2^{(0)} = 0.2 - 0.43 = -0.23$$

XII. Odds rati de l'esdeveniment E_*, correlació mínima

$$OR_* = \frac{\left(\left(1+\frac{OR_1*p_1^0}{1-p_1^0}\right)*\left(1+\frac{OR_2*p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)}{\left(\left(1+\frac{p_1^0}{1-p_1^0}\right)*\left(1+\frac{p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)} = \ldots$$

$$\frac{\left(\left(1+\frac{0.50*0.46}{1-0.46}\right)*\left(1+\frac{0.33*0.43}{1-0.43}\right)-1-(-0.32)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(-0.32)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)}{\left(\left(1+\frac{0.46}{1-0.46}\right)*\left(1+\frac{0.43}{1-0.43}\right)-1-(-0.32)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(-0.32)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)} = \ldots$$

$$\frac{\left((1.42)*(1.24)-1-(-0.32)*0.18\right)*(1+(-0.32)*0.45)}{\left((1.85)*(1.75)-1-(-0.32)*0.45\right)*(1+(-0.32)*0.18)} = \frac{0.70}{2.24} = 0.31$$

XIII. Odds rati de l'esdeveniment E_*, correlació nul·la

$$OR_* = \frac{\left(\left(1+\frac{OR_1*p_1^0}{1-p_1^0}\right)*\left(1+\frac{OR_2*p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)}{\left(\left(1+\frac{p_1^0}{1-p_1^0}\right)*\left(1+\frac{p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)} = \ldots$$

$$\frac{\left(\left(1+\frac{0.50*0.46}{1-0.46}\right)*\left(1+\frac{0.33*0.43}{1-0.43}\right)-1-(0)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(0)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)}{\left(\left(1+\frac{0.46}{1-0.46}\right)*\left(1+\frac{0.43}{1-0.43}\right)-1-(0)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(0)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)} = \ldots$$

$$\frac{\left((1.42)*(1.24)-1-(0)*0.18\right)*(1+(0)*0.45)}{\left((1.85)*(1.75)-1-(0)*0.45\right)*(1+(0)*0.18)} = \frac{0.76}{2.23} = 0.34$$

XIV. Odds rati de l'esdeveniment E∗, correlació màxima

$$OR_* = \frac{\left(\left(1+\frac{OR_1*p_1^0}{1-p_1^0}\right)*\left(1+\frac{OR_2*p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)}{\left(\left(1+\frac{p_1^0}{1-p_1^0}\right)*\left(1+\frac{p_2^0}{1-p_2^0}\right)-1-rho*\sqrt{\frac{p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)*\left(1+rho*\sqrt{\frac{OR_1*OR_2*p_1^0*p_2^0}{(1-p_1^0)(1-p_2^0)}}\right)}=\ldots$$

$$\frac{\left(\left(1+\frac{0.50*0.46}{1-0.46}\right)*\left(1+\frac{0.33*0.43}{1-0.43}\right)-1-(0.76)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(0.76)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)}{\left(\left(1+\frac{0.46}{1-0.46}\right)*\left(1+\frac{0.43}{1-0.43}\right)-1-(0.76)*\sqrt{\frac{0.46*0.43}{(1-0.46)(1-0.43)}}\right)*\left(1+(0.76)*\sqrt{\frac{0.50*0.33*0.46*0.43}{(1-0.46)(1-0.43)}}\right)}=\ldots$$

$$=\frac{\left((1.42)*(1.24)-1-(0.76)*0.18\right)*(1+(0.76)*0.45)}{\left((1.85)*(1.75)-1-(0.76)*0.45\right)*(1+(0.76)*0.18)}=\frac{0.83}{2.15}=0.39$$

XV. Risc relatiu de l'esdeveniment E∗, correlació mínima

$$RR_* = \frac{p_1^0R_1+p_2^0R_2-p_1^0p_2^0R_1R_2-rho\sqrt{p_1^0p_2^0R_1R_2(1-p_1^0R_1)(1-p_2^0R_2)}}{1-q_1^0q_2^0-rho\sqrt{p_1^0p_2^0q_1^0q_2^0}}=\ldots$$

$$\frac{0.46*0.65+0.43*0.46-0.46*0.43*0.65*0.46-(-0.32)\sqrt{0.46*0.43*0.65*0.46*(1-0.46*0.65)*(1-0.43*0.46)}}{1-(1-0.46)*(1-0.43)-(-0.32)\sqrt{0.46*0.45*(1-0.46)*(1-0.45)}}=\ldots$$

$$\frac{0.44-(-0.05)}{0.69-(-0.07)}=0.64$$

XVI. Risc relatiu de l'esdeveniment E∗, correlació nul·la

$$RR_* = \frac{p_1^0R_1+p_2^0R_2-p_1^0p_2^0R_1R_2-rho\sqrt{p_1^0p_2^0R_1R_2(1-p_1^0R_1)(1-p_2^0R_2)}}{1-q_1^0q_2^0-rho\sqrt{p_1^0p_2^0q_1^0q_2^0}}=\ldots$$

$$\frac{0.46*0.65+0.43*0.46-0.46*0.43*0.65*0.46-(0)\sqrt{0.46*0.43*0.65*0.46*(1-0.46*0.65)*(1-0.43*0.46)}}{1-(1-0.46)*(1-0.43)-(0)\sqrt{0.46*0.45*(1-0.46)*(1-0.45)}}=\ldots$$

$$\frac{0.44}{0.69}=0.63$$

XVII. Risc relatiu de l'esdeveniment E∗, correlació màxima

$$RR_* = \frac{p_1^0R_1+p_2^0R_2-p_1^0p_2^0R_1R_2-rho\sqrt{p_1^0p_2^0R_1R_2(1-p_1^0R_1)(1-p_2^0R_2)}}{1-q_1^0q_2^0-rho\sqrt{p_1^0p_2^0q_1^0q_2^0}}=\ldots$$

$$\frac{0.46*0.65+0.43*0.46-0.46*0.43*0.65*0.46-(0.76)\sqrt{0.46*0.43*0.65*0.46*(1-0.46*0.65)*(1-0.43*0.46)}}{1-(1-0.46)*(1-0.43)-(0.76)\sqrt{0.46*0.45*(1-0.46)*(1-0.45)}}=\ldots$$

$$\frac{0.44-0.14}{0.69-0.18}=0.59$$

XVIII. Diferència de proporcions de l'esdeveniment E∗, correlació mínima

$$\delta_* = \delta_1 q_1^0 + \delta_2 q_2^0 \; - \delta_1\delta_2 \; + \rho\sqrt{p_1^{(0)} p_2^{(0)} q_1^{(0)} q_2^{(0)}} \; - rho\sqrt{(p_1^{(0)} + \delta_1)(p_2^{(0)} + \delta_2)(q_1^{(0)} - \delta_1)(q_2^{(0)} - \delta_2)} = \ldots$$

$$\textit{(-0.16)*}(1 - 0.46) + (-0.23) * \textit{(1-0.43)} - (-0.16 * -0.23) + \ldots$$

$$(-0.32)\sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} - \ldots$$

$$(-0.32)\sqrt{\big(0.46 + (-0.16)\big)\big(0.43 + (-0.23)\big)\big((1 - 0.46) - (-0.16)\big)\big((1 - 0.43) - (-0.23)\big)} = \ldots$$

$$\textit{-0.25+ (-0.08) +0.05 = -0.28}$$

XIX. Diferència de proporcions de l'esdeveniment E∗, correlació nul·la

$$\delta_* = \delta_1 q_1^0 + \delta_2 q_2^0 \; - \delta_1\delta_2 \; + \rho\sqrt{p_1^{(0)} p_2^{(0)} q_1^{(0)} q_2^{(0)}} \; - rho\sqrt{(p_1^{(0)} + \delta_1)(p_2^{(0)} + \delta_2)(q_1^{(0)} - \delta_1)(q_2^{(0)} - \delta_2)} = \ldots$$

$$\textit{(-0.16)*}(1 - 0.46) + (-0.23) * \textit{(1-0.43)} - (-0.16 * -0.23) + \ldots$$

$$(0)\sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} - \ldots$$

$$(0)\sqrt{(0.46 + (-0.16))(0.43 + (-0.23))((1 - 0.46) - (-0.16))((1 - 0.43) - (-0.23))} = \textit{-0.25}$$

XX. Diferència de proporcions de l'esdeveniment E∗, correlació m

$$\delta_* = \delta_1 q_1^0 + \delta_2 q_2^0 \; - \delta_1\delta_2 \; + \rho\sqrt{p_1^{(0)} p_2^{(0)} q_1^{(0)} q_2^{(0)}} \; - rho\sqrt{(p_1^{(0)} + \delta_1)(p_2^{(0)} + \delta_2)(q_1^{(0)} - \delta_1)(q_2^{(0)} - \delta_2)} = \ldots$$

$$\textit{(-0.16)*}(1 - 0.46) + (-0.23) * \textit{(1-0.43)} - (-0.16 * -0.23) + \ldots$$

$$(0.76)\sqrt{0.46 * 0.43 * (1 - 0.46) * (1 - 0.43)} - \ldots$$

$$(0.76)\sqrt{(0.46 + (-0.16))(0.43 + (-0.23))((1 - 0.46) - (-0.16))((1 - 0.43) - (-0.23))} =$$

$$\textit{-0.25 + 0.18 - 0.14= - 0.21}$$

2. Annex 2 : Codi font i visualització dels arxius

A. Funcions de partida

Es mostren les funcions de partida en base a les quals s'ha desenvolupar tota la part pràctica del treball i s'han postulat totes les millores realitzades.

Annex B.1 Funció *Bahadur composite*

```
Bahadur.composite<- function(p1, p2, rho){
  p.composite<- 1- (1-p1)*(1-p2)*( 1+ rho*sqrt(p1*p2/((1-p1)*(1-p2)) ))
  return(p.composite)
}
```

Annex B.2 Funció *diff.pcomp*

```
diff.pcomp <- function(p0.RE, p0.AE, p1.RE, p1.AE, rho){
  diff = Bahadur.composite(p1.RE, p1.AE, rho)-Bahadur.composite(p0.RE, p0.AE,
 rho)
  return(diff)
}
```

Annex B.3 Funció *RR.pcomp*

```
RR.pcomp <- function(p0.RE, p0.AE, p1.RE, p1.AE, rho){
  RR = Bahadur.composite(p1.RE, p1.AE, rho)/Bahadur.composite(p0.RE, p0.AE,
 rho)

  return(RR)
}
```

Annex B.4 Funció *OR.composite.function*

```
OR.composite.function = function(p1.0, p2.0, OR1, OR2, rho){
  O10= p1.0/(1-p1.0)
  O20= p2.0/(1-p2.0)

  OR= ((O10*OR1+1)*(O20*OR2+1)-1-
 rho*sqrt(OR1*OR2*O10*O20))*(1+rho*sqrt(O10*O20))/
    (((1+O10)*(1+O20)-1-rho*sqrt(O10*O20))*(1+rho*sqrt(OR1*OR2*O10*O20)))
  return(OR)
}
```

Annex B.5 Funció *correlation.min.function*

```
correlation.min.function = function(p1, p2){
  rho.min <- max(  -sqrt(p1*p2/((1-p1)*(1-p2))), -sqrt(((1-p1)*(1-
 p2))/(p1*p2)  ) )
  return(rho.min)
}
```

Annex B.6 Funció *correlation.max.function*

```
correlation.max.function = function(p1, p2){
  rho.max <- min(  sqrt( ( p1/ (1-p1) )/(p2/(1-p2))), sqrt((p2/(1-p2))/(p1/(1-
 p1))))
  return(rho.max)
```

```
}
```

Annex B.7 Funció *SampleSize.CBE.OR*

```
SampleSize.CBE.OR <- function(p1.0, p2.0, OR1, OR2, rho, alpha=0.05, beta=0.2,
 Unpooled="Unpooled Variance"){

  p1.1= (OR1*p1.0/(1-p1.0))/(1+(OR1*p1.0/(1-p1.0)))
  p2.1 = (OR2*p2.0/(1-p2.0))/(1+(OR2*p2.0/(1-p2.0)))

  p0.CBE = 1- (1-p1.0)*(1-p2.0)*( 1+ rho*sqrt(p1.0*p2.0/((1-p1.0)*(1-p2.0)) ))
  p1.CBE = 1- (1-p1.1)*(1-p2.1)*( 1+ rho*sqrt(p1.1*p2.1/((1-p1.1)*(1-p2.1)) ))

  OR.CBE = OR.composite.function(p1.0, p2.0, OR1, OR2, rho)

  n =  SampleSize.OR(p0.CBE, OR.CBE, alpha, beta, Unpooled)

  return (n)
}
```

Annex B.8 Funció *SampleSize.OR*

```
SampleSize.OR <- function(p0, OR, alpha=0.05, beta=0.2, Unpooled="Unpooled
 Variance"){
  z.alpha <- qnorm(1-alpha,0,1)
  z.beta <-  qnorm(1-beta,0,1)

  p1 = (OR*p0/(1-p0))/(1+(OR*p0/(1-p0)))

  if(Unpooled=="Unpooled Variance"){
    # sample size per group
    n1 = ((z.alpha+z.beta)/(log(OR)))^2*( 1/(p0*(1-p0)) + 1/(p1*(1-p1)) )
  }else{
    p = (p1 + p0)/2
    # sample size per group
    n1 = ((z.alpha* sqrt(2/(p*(1-p))) + z.beta* sqrt(1/(p0*(1-p0)) + 1/(p1*(1-
p1))))/(log(OR)))^2
  }

  n = 2*n1

  return(n)
}
```

Annex B.9 Funció *SampleSize.CBE.RR*

```
SampleSize.CBE.RR <- function(p1.0, p2.0, R1, R2, rho, alpha=0.05, beta=0.2,
 Unpooled="Unpooled Variance"){

  p1.1=  R1*p1.0
  p2.1 = R2*p2.0

  p0.CBE = 1- (1-p1.0)*(1-p2.0)*( 1+ rho*sqrt(p1.0*p2.0/((1-p1.0)*(1-p2.0))
))
  RR.CBE <- RR.pcomp(p1.0, p2.0, p1.1, p2.1, rho)

  n = SampleSize.RR(p0.CBE, RR.CBE, alpha, beta, Unpooled)

  return (n)
}
```

Annex B.10 Funció *SampleSize.RR*

```
SampleSize.RR <- function(p0, R, alpha=0.05, beta=0.2, Unpooled="Unpooled
 Variance"){
  z.alpha <- qnorm(1-alpha,0,1)
```

```
  z.beta <-  qnorm(1-beta,0,1)


  if(Unpooled=="Unpooled Variance"){
    # sample size per group
    n1 = ((z.alpha+z.beta)/(log(R)))^2*( (1-R*p0)/(R*p0) + (1-p0)/p0 )
  }else{
    p1= R*p0
    p = (p1 + p0)/2

    # sample size per group
    n1 = ( (z.alpha* sqrt(2*(1-p)/p) + z.beta* sqrt( (1-R*p0)/(R*p0) + (1-
p0)/p0) )/(log(R)) )^2
  }

  n = 2*n1

  return(n)
}
```

Annex B.11 Funció *SampleSize.CBE.Diff*

```
SampleSize.CBE.Diff <- function(p1.0, p2.0, d1, d2, rho, alpha=0.05, beta=0.2,
 Unpooled="Unpooled Variance"){

  p1.1= d1+p1.0
  p2.1 = d2+p2.0

  p0.CBE = 1- (1-p1.0)*(1-p2.0)*( 1+ rho*sqrt(p1.0*p2.0/((1-p1.0)*(1-p2.0))
))
  d.CBE <- diff.pcomp(p1.0, p2.0, p1.1, p2.1, rho)

  n = SampleSize.Diff(p0.CBE,d.CBE,alpha,beta,Unpooled)

  return (n)
}
```

Annex B.12 Funció *SampleSize.Diff*

```
SampleSize.Diff <- function(p0, d, alpha=0.05, beta=0.2, Unpooled="Unpooled
 Variance"){
  z.alpha <- qnorm(1-alpha,0,1)
  z.beta <-  qnorm(1-beta,0,1)

  if(Unpooled=="Unpooled Variance"){
    # sample size per group
    n1 = ((z.alpha+z.beta)/d)^2*( p0*(1-p0) + (d+p0)*(1-p0-d))
  }else{
    p1 = p0 + d
    p = (p1 + p0)/2
    # sample size per group
    n1 = ((z.alpha* sqrt(2*p*(1-p)) +  z.beta* sqrt( p0*(1-p0) + (d+p0)*(1-p0-
d)))/d)^2
  }
  n = 2*n1

  return(n)
}
```

B. Estructura final

S'adjunten els arxius que s'han creat per poder construïr el paquet, conjuntament amb la documentació creada.

Annex B.13 Funció *prob_ce*

```
#' PROBABILITY OF UNION OF TWO EVENTS
#'
#'
#' @description This function is used to calculate the probability of the
#' union of two events, E1 and E2,
#' knowing in advance the probability of occurrence of each of the two events
#' by separating
#' them on the basis of the study population, and in the other band knowing
#' the Pearson coefficient that correlates the events.
#'
#'
#' @param p_e1 numeric parameter, probability of the event E1
#' @param p_e2 numeric parameter, probability of the event E2
#' @param rho numeric parameter, Pearson correlation between E1 i E2
#'
#' @export
#'
#' @return Returns the probability of a union of two events
#' @details Returns a numeric value. The input parameters representing the
#' probability of the events
#' taking place are limited between 0 and 1, without including both values.
#' Pearson's correlation
#' must be within the confidence interval that allows the combined variable
#' according to the probabilities given.
#' To calculate this confidence interval you can use lower_corr and upper_corr
#' functions that you can find in this package.
#'
prob_ce <- function(p_e1, p_e2, rho){
  if(p_e1 < 0 || p_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p_e2 < 0 || p_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else if(rho <= lower_corr(p_e1,p_e2)  ||  rho >= upper_corr(p_e1,p_e2)){
    stop("The correlations of events must be in the correct interval")
  }else{
    prob_ce <- 1- (1-p_e1)*(1-p_e2)*( 1+ rho*sqrt(p_e1*p_e2/((1-p_e1)*(1-
p_e2))))
    return(prob_ce)
  }
}
```

Annex B.14 Funció *Lower_corr*

```
#' LOWER LIMIT FOR PEARSON CORRELATION
#'
#'
#' @description Knowing that the correlation of a combined variable is
#' presented in an intervalic way,
#' and has been calculated from the probability of occurrence of the
#' respective events.
#' This function allows to calculate the minimum correlation
#' that this typology of variables can experience.
#'
#'
#' @param p_e1 numeric parameter, probability of the event E1
#' @param p_e2 numeric parameter, probability of the event E2
#'
#' @export
#'
#' @return Returns the lower correlation threshold that the binominal compost
#' event can reach
```

```
#' @details lower_corr returns a numeric value negated bounded between -1 and
#' 0.
#' The probabilities of the occurrence of events must be defined by the open
#' interval of (0.1).
#'
lower_corr <- function(p_e1,p_e2){
  if(p_e1 < 0 || p_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p_e2 < 0 || p_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else{
    lower_corr <- max(  -sqrt(p_e1*p_e2/((1-p_e1)*(1-p_e2))), -sqrt(((1-
p_e1)*(1-p_e2))/(p_e1*p_e2)  ) )
    return(lower_corr)
  }
}
```

Annex B.15 Funció *Upper_corr*

```
#' UPPER LIMIT FOR PEARSON CORRELATION
#'
#' @description Knowing that the correlation of a combined variable is
#' presented in an intervalic manner,
#' and calculated from the probability of occurrence of the respective events.
#' This function allows to calculate the minimum correlation that this
#' typology of variables can experience.
#'
#'
#' @param p_e1 numeric parameter, probability of the event E1
#' @param p_e2 numeric parameter, probability of the event E2
#'
#' @export
#'
#' @return Returns the upper correlation threshold that can be set by the
#' compost binarius event.
#' @details upper_corr returns a numeric value negated bounded between 0 and
#' 1.
#' The probabilities of the occurrence of events must be defined by the open
#' interval of (0.1).
#'
upper_corr <- function(p_e1,p_e2){
  if(p_e1 < 0 || p_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p_e2 < 0 || p_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else{
    upper_corr <- min(  sqrt( ( p_e1/ (1-p_e1) )/( p_e2/ (1-p_e2) ) ),
sqrt((p_e2/(1-p_e2))/(p_e1/(1-p_e1)) ) )
    return(upper_corr)
  }
}
```

Annex B.16 Funció *effect_ce*

```
#' STATISTICAL EFFECTS
#'
#' @description This function calculates various statistical measures for
#' combined variables, in particular some measures
#' such as rati odds, risk ratio or the difference in the proportions of two
#' groups in relation
#' to the binarius compost event.
```

```
#'
#'
#' @param p0_e1 numeric parameter, probability of occurrence E1 by the control
#' group
#' @param p0_e2 numeric parameter, probability of occurrence E2 by the control
#' group
#' @param p1_e1 numeric parameter, probability of occurrence E1 by the
#' intervention group
#' @param p1_e2 numeric parameter, probability of occurrence E2 by the
#' intervention group
#' @param rho numeric parameter, Pearson correlation between E1 i E2
#' @param effect_ce character, specifies the type of statistical measure that
#' is calculated
#'
#' @export
#'
#' @return Returns the desired effect of the composite binary event and the
#' effects of the events
#' @details The input parameters representing the probability of the events
#' taking place are limited between 0 and 1, without including both values.
#' Pearson's correlation
#' must be within the confidence interval that allows the combined variable
#' according to the probabilities given.
#' To calculate this confidence interval you can use lower_corr and upper_corr
#' functions that you can find in this package.
#' For defect, if you don't specify the type of effect you want to obtain,
#' it calculates the difference in proportions.
#'
effect_ce <- function(p0_e1, p0_e2, p1_e1, p1_e2, rho, effect_ce = "diff"){
  if(p0_e1 < 0 || p0_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p0_e2 < 0 || p0_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else if(p1_e1 < 0 || p1_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p1_e2 < 0 || p1_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else if(rho <= max(c(lower_corr(p0_e1,p0_e2),lower_corr(p1_e1,p1_e2)))  ||
rho >= max(c(upper_corr(p0_e1,p0_e2),upper_corr(p1_e1,p1_e2)))){
    stop("The correlations of events must be in the correct interval")
  }else if(effect_ce != "rr" && effect_ce != "diff" && effect_ce != "or"){
    stop("You have to choose between odds ratio, relative risk or difference
in proportions")
  }

    if(effect_ce == "diff"){
      diff_e1 = p1_e1 - p0_e1
      diff_e2 = p1_e2 - p0_e2
      effect = prob_ce(p1_e1,p1_e2,rho) - prob_ce(p0_e1,p0_e2,rho)
      effect_out <- data.frame(diff_e1,diff_e2,effect)
    }else if(effect_ce == "rr"){
      rr_e1 = p1_e1 / p0_e1
      rr_e2 = p1_e2 / p0_e2
      effect = prob_ce(p1_e1,p1_e2,rho)/prob_ce(p0_e1,p0_e2,rho)
      effect_out = data.frame(rr_e1,rr_e2,effect)
    }else if(effect_ce == "or"){
      O10= p0_e1/(1-p0_e1)
      O20= p0_e2/(1-p0_e2)
      or_e1 = (p1_e1/(1-p1_e1))/(p0_e1/(1-p0_e1))
      or_e2 = (p1_e2/(1-p1_e2))/(p0_e2/(1-p0_e2))
      effect = ((O10*or_e1+1)*(O20*or_e2+1)-1-
rho*sqrt(or_e1*or_e2*O10*O20))*(1+rho*sqrt(O10*O20))/
        (((1+O10)*(1+O20)-1-
rho*sqrt(O10*O20))*(1+rho*sqrt(or_e1*or_e2*O10*O20)))
```

```
      effect_out = data.frame(or_e1,or_e2,effect)
    }
    colnames(effect_out) <- c("Effect E1","Effect E2","Effect CE")
    return(effect_out)
}
```

Annex B.17 Funció *sample_size_ce*

```
#' SAMPLE SIZE FOR COMPOSITE BINARY ENDPOINT
#'
#' @description This function calculates the value of the sample size
#' according to its input parameters.
#'
#' @param p0_e1 numeric parameter, probability of occurrence E1 by the control
#' group
#' @param p0_e2 numeric parameter, probability of occurrence E2 by the control
#' group
#' @param type_e1 Effect of the event E1
#' @param eff_e1 numeric parameter, effect of the event E1
#' @param type_e2 Effect of the event E2
#' @param eff_e2 numeric parameter, effect of the event E2
#' @param effect_ce Effect to measure the sample size
#' @param rho numeric parameter, correlation of pearson between two events E1
#' and E2
#' @param alpha level of confidencea alpha
#' @param beta level of confidence beta
#' @param unpooled Variance class used in the calculation of the sample size
#'
#' @export
#'
#' @return Return the sample size for composite binary endpoints based on the
#' anticipated values of the composite components
#' and the association between them in terms of Pearson's correlation.
#' @details The entry parameters representing the probability of the events
#' taking place are limited between 0 and 1,
#'  without including both values. The values of the primary events should be
#' calculated giving a higher probability of observation to the control group.
#'  Pearson's correlation must be within the confidence interval allowed
#'  by the combined variable according to the probabilities given. To
#' calculate this confidence  interval you can use lower_corr and upper_corr
#' functions
#'  that you can find in this package. By default, if the type of effect
#'  with which the resulting large sample is to be associated is not
#' specified,
#'  the difference in proportions is used. When specifying the levels of alpha
#' and beta meaning,
#'  they may not be higher than 1, as well as the following type of variance
#' that is desired to be used
#'  in the calculation of the sample size.

sample_size_ce <- function(p0_e1,p0_e2,type_e1,eff_e1,type_e2,eff_e2,effect_ce
= "diff",rho, alpha = 0.05, beta = 0.2, unpooled = "unpooled Variance"){

  if(p0_e1 < 0 || p0_e1 > 1){
    stop("The probability of observing the event E1 (p_e1) must be number
between 0 and 1")
  }else if(p0_e2 < 0 || p0_e2 > 1){
    stop("The probability of observing the event E2 (p_e2) must be number
between 0 and 1")
  }else if(type_e1 != "diff" && type_e1 != "rr" && type_e1 != "or"){
    stop("You have to choose between odds ratio, relative risk or difference
in proportions")
  }else if((type_e1 == "diff" && eff_e1 > 0) || (type_e1 == "or" && (eff_e1 <
0 || eff_e1 > 1)) || (type_e1 == "rr" && (eff_e1 < 0 || eff_e1 > 1))){
    stop("The effect of the event E1 is not right")
  }else if(type_e2 != "diff" && type_e2 != "rr" && type_e2 != "or"){
```

```
    stop("You have to choose between odds ratio, relative risk or difference
in proportions")
  }else if((type_e2 == "diff" && eff_e2 > 0) || (type_e2 == "or" && (eff_e2 <
0 || eff_e2 > 1)) || (type_e2 == "rr" && (eff_e2 < 0 || eff_e2 > 1))){
    stop("The effect of the event E2 is not right")
  }else if(effect_ce != "diff" && effect_ce != "rr" && effect_ce != "or"){
    stop("You have to choose between odds ratio, relative risk or difference
in proportions")
   }else if(rho <= lower_corr(p0_e1,p0_e2)  ||  rho >=
upper_corr(p0_e1,p0_e2)){
    stop("The correlations of events must be in the correct interval")
  }else if( 0 > alpha || alpha > 1){
    stop("Alpha value must be number between 0 and 1")
  }else if( 0 > beta || beta > 1){
    stop("Beta value must be number between 0 and 1")
  }else if(unpooled != "unpooled Variance" && unpooled != "Pooled Variance"){
    stop("You must choose between pooled and unpooled variance")
  }

  #Per substitució de l'equació
  if(type_e1 == "or"){
    p1_e1= (eff_e2*p0_e1/(1-p0_e1))/(1+(eff_e2*p0_e1/(1-p0_e1)))
  }else if(type_e1 == "rr"){
    p1_e1 = eff_e1 * p0_e1
  }else if(type_e1 == "diff"){
    p1_e1 = eff_e1 + p0_e1
  }

   if(type_e2 == "or"){
    p1_e2 = (eff_e2*p0_e2/(1-p0_e2))/(1+(eff_e2*p0_e2/(1-p0_e2)))
  }else if(type_e2 == "rr"){
    p1_e2 = eff_e2 * p0_e2
  }else if(type_e2 == "diff"){
    p1_e2 = eff_e2 + p0_e2
  }

  p0_CBE = 1- (1-p0_e1)*(1-p0_e2)*( 1+ rho*sqrt(p0_e1*p0_e2/((1-p0_e1)*(1-
p0_e2)) ))
  p1_CBE = 1- (1-p1_e1)*(1-p1_e2)*( 1+ rho*sqrt(p1_e1*p1_e2/((1-p1_e1)*(1-
p1_e2)) ))


  if(effect_ce == "rr"){
    rr_CBE <- effect_ce(p0_e1, p0_e2, p1_e1, p1_e2, rho, type = "rr")[1,3]
    if(unpooled =="unpooled Variance"){
      samp = 2*(((qnorm(1-alpha,0,1)+qnorm(1-beta,0,1))/(log(rr_CBE)))^2*( (1-
rr_CBE*p0_CBE)/(rr_CBE*p0_CBE) + (1-p0_CBE)/p0_CBE))
    }else if(unpooled =="Variance"){
      p = (rr_CBE*p0_CBE + p0_CBE)/2
      # sample size per group
      samp = 2*(( (qnorm(1-alpha,0,1)* sqrt(2*(1-p)/p) + qnorm(1-beta,0,1)*
sqrt( (1-rr_CBE*p0_CBE)/(rr_CBE*p0_CBE) + (1-p0_CBE)/p0_CBE) )/(log(rr_CBE))
)^2  )
    }
  }else if(effect_ce == "or"){
    or_CBE = effect_ce(p0_e1, p0_e2, p1_e1, p1_e2, rho, type = "or")[1,3]
    p1 = (or_CBE*p0_CBE/(1-p0_CBE))/(1+(or_CBE*p0_CBE/(1-p0_CBE)))
    if(unpooled=="unpooled Variance"){
      samp = ((qnorm(1-alpha,0,1)+qnorm(1-beta,0,1))/(log(or_CBE)))^2*(
1/(p0_CBE*(1-p0_CBE)) + 1/(p1*(1-p1)) )
    }else{
      samp = ((qnorm(1-alpha,0,1)* sqrt(2/(((p1 + p0_CBE)/2)*(1-((p1 +
p0_CBE)/2)))) + qnorm(1-beta,0,1)* sqrt(1/(p0_CBE*(1-p0_CBE)) + 1/(p1*(1-
p1))))/(log(or_CBE)))^2
    }
  }else{
    diff_CBE <- effect_ce(p0_e1, p0_e2, p1_e1, p1_e2, rho, type = "rr")[1,3]
    if(unpooled=="unpooled Variance"){
```

```
      samp = ((qnorm(1-alpha,0,1) +qnorm(1-beta,0,1))/diff_CBE)^2*( p0_CBE*(1-
p0_CBE) + (diff_CBE+p0_CBE)*(1-p0_CBE-diff_CBE))
    }else if(unpooled=="Variance"){
      p = (diff_CBE + 2 * p0_CBE)/2
      # sample size per group
      samp = ((qnorm(1-alpha,0,1)* sqrt(2*p*(1-p)) +  qnorm(1-beta,0,1)* sqrt(
p0_CBE*(1-p0_CBE) + (diff_CBE+p0_CBE)*(1-p0_CBE-diff_CBE)))/diff.CBE)^2
    }

  }
  return(samp)

}
```

Annex B.18 *Description*

```
Package: CBE
Title: Set of functions that calculate measures of the composite binary
endpoints
Version: 0.0.0.9000
Authors@R: c(
    person("Marta", "Bofill", email = "marta.bofill.roig@upc.edu", role =
    "cre"),
    person("Raquel", "Rovira", email = "rovirasalvat.raquel@gmail.com", role =
    "aut"),
    person("Jordi", "Cortes", email = "jordi.cortes-martinez@upc.edu", role =
    "aut"))
Description: It includes a set of functions of great use in studies where the
study variable is combined. It has mainly been designed to calculate the large
sample required in any assay with combined variables. In addition, it allows to
calculate different statistical measures of great use in the clinical field as
the odds rati, the relative risk or the difference in proportions, as well as
the confidence interval on the correlation is established of the combined
variable. Some numerical values, such as the probability of the union of two
primary events or the confidence bands on which
the Pearson correlation of all variable combined is established, can also be
calculated using functions of the same package.
License: GPL-3
Encoding: UTF-8
LazyData: true
RoxygenNote: 6.1.1
```

Annex B.19 *Namespace*

```
export(prob_ce)
export(lower_corr)
export(upper_corr)
export(effect_ce)
export(sample_size_ce)
```

C. Validació de l'estructura

El codi generat per validar les funcions creades, d'acord amb els arguments postulats al quart i al cinquè apartat dels capítols III i IV, el podem trobar en la carpeta *testthat* del següent enllaç del repositori del GitHub:

https://github.com/CompARE-Composite/CompARE-package/tree/master/CompARE-bin/tests

S'executa a partir de l'arxiu *testthat.R* que trobem en el mateix enllaç.

D. Visualització de les funcions

Annex B.20 Visualització *prob_ce*

prob_ce {CBE} R Documentation

PROBABILITY OF UNION OF TWO EVENTS

Description

This function is used to calculate the probability of the union of two events, E1 and E2, knowing in advance the probability of occurrence of each of the two events by separating them on the basis of the study population, and in the other band knowing the Pearson coefficient that correlates the events.

Usage

```
prob_ce(p_e1, p_e2, rho)
```

Arguments

p_e1	numeric parameter, probability of the event E1
p_e2	numeric parameter, probability of the event E2
rho	numeric parameter, Pearson correlation between E1 i E2

Details

Returns a numeric value. The input parameters representing the probability of the events taking place are limited between 0 and 1, without including both values. Pearson's correlation must be within the confidence interval that allows the combined variable according to the probabilities given. To calculate this confidence interval you can use lower_corr and upper_corr functions that you can find in this package.

Value

Returns the probability of a union of two events

Annex B.21 Visualització *Lower_corr*

lower_corr {CBE} R Documentation

LOWER LIMIT FOR PEARSON CORRELATION

Description

Knowing that the correlation of a combined variable is presented in an intervalic way, and has been calculated from the probability of occurrence of the respective events. This function allows to calculate the minimum correlation that this typology of variables can experience.

Usage

```
lower_corr(p_e1, p_e2)
```

Arguments

p_e1	numeric parameter, probability of the event E1
p_e2	numeric parameter, probability of the event E2

Details

lower_corr returns a numeric value negated bounded between -1 and 0. The probabilities of the occurrence of events must be defined by the open interval of (0.1).

Value

Returns the lower correlation threshold that the binominal compost event can reach

Annex B.22 Visualització *Upper_corr*

upper_corr {CBE} R Documentation

UPPER LIMIT FOR PEARSON CORRELATION

Description

Knowing that the correlation of a combined variable is presented in an intervalic manner, and calculated from the probability of occurrence of the respective events. This function allows to calculate the minimum correlation that this typology of variables can experience.

Usage

```
upper_corr(p_e1, p_e2)
```

Arguments

p_e1	numeric parameter, probability of the event E1
p_e2	numeric parameter, probability of the event E2

Details

upper_corr returns a numeric value negated bounded between 0 and 1. The probabilities of the occurrence of events must be defined by the open interval of (0.1).

Value

Returns the upper correlation threshold that can be set by the compost binarius event.

Annex B.23 Visualització *effect_ce*

effect_ce {CBE} R Documentation

STATISTICAL EFFECTS

Description

This function calculates various statistical measures for combined variables, in particular some measures such as rati odds, risk ratio or the difference in the proportions of two groups in relation to the binarius compost event.

Usage

```
effect_ce(p0_e1, p0_e2, p1_e1, p1_e2, rho, effect_ce = "diff")
```

Arguments

p0_e1	numeric parameter, probability of occurrence E1 by the control group
p0_e2	numeric parameter, probability of occurrence E2 by the control group
p1_e1	numeric parameter, probability of occurrence E1 by the intervention group
p1_e2	numeric parameter, probability of occurrence E2 by the intervention group
rho	numeric parameter, Pearson correlation between E1 i E2
effect_ce	character, specifies the type of statistical measure that is calculated

Details

The input parameters representing the probability of the events taking place are limited between 0 and 1, without including both values. Pearson's correlation must be within the confidence interval that allows the combined variable according to the probabilities given. To calculate this confidence interval you can use lower_corr and upper_corr functions that you can find in this package. For defect, if you don't specify the type of effect you want to obtain, it calculates the difference in proportions.

Value

Returns the desired effect of the composite binary event and the effects of the events

Annex B.24 Visualització *sample_size_ce*

sample_size_ce {CBE} R Documentation

SAMPLE SIZE FOR COMPOSITE BINARY ENDPOINT

Description

This function calculates the value of the sample size according to its input parameters.

Usage

```
sample_size_ce(p0_e1, p0_e2, type_e1, eff_e1, type_e2, eff_e2,
  effect_ce = "diff", rho, alpha = 0.05, beta = 0.2,
  unpooled = "unpooled Variance")
```

Arguments

`p0_e1`	numeric parameter, probability of occurrence E1 by the control group
`p0_e2`	numeric parameter, probability of occurrence E2 by the control group
`type_e1`	Effect of the event E1
`eff_e1`	numeric parameter, effect of the event E1
`type_e2`	Effect of the event E2
`eff_e2`	numeric parameter, effect of the event E2
`effect_ce`	Effect to measure the sample size
`rho`	numeric parameter, correlation of pearson between two events E1 and E2
`alpha`	level of confidencea alpha
`beta`	level of confidence beta
`unpooled`	Variance class used in the calculation of the sample size

Details

The entry parameters representing the probability of the events taking place are limited between 0 and 1, without including both values. The values of the primary events should be calculated giving a higher probability of observation to the control group. Pearson's correlation must be within the confidence interval allowed by the combined variable according to the probabilities given. To calculate this confidence interval you can use lower_corr and upper_corr functions that you can find in this package. By default, If the type of effect with which the resulting large sample is to be associated is not specified, the difference in proportions is used. When specifying the levels of alpha and beta meaning, they may not be higher than 1, as well as the following type of variance that is desired to be used in the calculation of the sample size.

Value

Return the sample size for composite binary endpoints based on the anticipated values of the composite components and the association between them in terms of Pearson's correlation.