

## Grau en Estadística

**Títol:** Anàlisi de l'evolució del pes segons diferents grups de dietes mitjançant la tècnica dels models lineals mixtos amb mesures repetides.

**Autor:** Carlota Ruiz de Porras Rubio

**Director:** Marta Cubedo Culleré i Antonio Miñarro Alonso

**Departament:** Departament de Genètica, Microbiologia i Estadística

**Convocatòria:** 2n SEMESTRE curs 2020-2021



UNIVERSITAT DE  
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat de Matemàtiques i Estadística



# *Agraïments*

*Als meus directors, Marta Cubedo Culleré i Antonio Miñarro Alonso,  
per l'ajuda, dedicació i contribució durant tot aquest camí.*

*A la meva família per l'amor i confiança incondicional.*

*I als amics i companys que he conegut en aquests últims quatre anys.*

*A tots, gràcies per acompanyar-me!*



## *Resum*

En moltes de les branques que componen el món científic, és freqüent l'observació i estudi de variables al llarg del temps en un conjunt de subjectes, en diferents situacions experimentals depenent de l'àrea d'investigació. En aquest Treball de Final de Grau es presenta una anàlisi basada en el temps d'evolució dels pesos de rates classificades en grups segons la dieta amb la qual estan alimentades, per a comprovar si es detecten diferències entre les corbes que modelitzen el pes al llarg del temps en funció del tipus de dieta. En aquest estudi, s'espera observar diferències entre els tres grups de rates en funció de la dieta amb la qual han estat alimentades; dieta Control, dieta hiperlipídica de blat de moro o dieta hiperlipídica d'oli d'oliva. A causa de la naturalesa d'aquest tipus de dades, els models clàssics no acostumen a ser els indicats per a elaborar una correcta anàlisi de les dades. Per això, s'han d'emprar models que sí que tinguin presents aquestes característiques. En aquest cas, es realitzarà l'anàlisi d'aquestes dades mitjançant la tècnica dels models lineals mixtos amb mesures repetides. Els models mixtos són una generalització del model lineal de regressió clàssic, contemplant la possible existència d'observacions correlacionades o amb variabilitat heterogènia, vinculades a la presència de factors aleatoris. La nomenclatura models mixtos es refereix, precisament, al fet que l'anàlisi ha de bregar amb efectes fixos i aleatoris simultàniament.

**Keywords** – Models mixtos. Mesures repetides. Anàlisi longitudinal. Dietes hiperlipídiques. Dietes en rates. Heteroscedasticitat residual. Correlació. Paquet nlme.

## **CLASSIFICACIÓ AMS**

62P10 - Applications to biology and medical sciences

62F03 - Hypothesis testing

62J10 - Analysis of variance and covariance

## Analysis of the weight evolution according to different groups of diets using the mixed linear models technique with repeated measures.

### *Abstract*

In many of the branches that make up the scientific world, it is common to observe and study variables over time in a set of characters, in different experimental situations regarding the area of research. This Final Degree Project exposes an analysis based on the time evolution of the weights of rats classified according to the diet which they are fed, in order to check if differences are detected between the curves that model the weight along time, depending on the type of diet. In this study it is expected to observe differences between the three groups of rats depending on the diet which they have been fed; Control diet, hyperlipidic corn diet or hyperlipidic olive oil diet. Due to the nature of this type of data, the classical models are not the best indicated to elaborate a correct analysis of the data. Therefore, models that face these characteristics must be used. In this specific case, the analysis of this data will be performed using the technique of mixed linear models with repeated measurements. The mixed models are a generalization of the linear regression model, observing the possible existence of correlated observations or with a heterogeneous variability, linked with random variables. Mixed models are referred specifically, as the fact that the analysis clashes mixed and random effects simultaneously.

**Keywords** — Mixed models. Repeated measures. Longitudinal analysis. Hyperlipidic diets. Diets in rats. Residuals heteroscedasticity. Correlation. nlme package.



# Índex de continguts

*Resum*

*Abstract*

<b>1. INTRODUCCIÓ</b>	9
1.1. Motivació	9
1.2. Objectius i hipòtesi	9
1.3. Tipus de dietes	9
1.3.1. Lípids i dietes hiperlipídiques	10
1.3.2. Dieta normolipídica	11
1.4. Models lineals mixtos	11
1.4.1. Mesures repetides	12
<b>2. METODOLOGIA</b>	14
2.1. Els models lineals mixtos amb mesures repetides	14
2.2. Procedència de les dades	15
2.3. Programes informàtics utilitzats	15
2.3.1. Què és R?	16
2.3.2. RStudio	16
2.3.3. R packages utilitzats	16
<b>3. ANÀLISI DE LES DADES I RESULTATS</b>	18
3.1. Descriptiva	18
3.2. Anàlisi dels possibles models	22
3.2.1. Model 1: model amb Dies com a covariable i simetria composada	22
3.2.2. Model 2: model amb Dies com a factor fix i simetria composada	24
3.2.3. Model 3.1: model amb Dies com a covariable i correlació autoregressiva de primer ordre	25
3.2.4. Model 4.1: model amb Dia com a factor fix i correlació autoregressiva de primer ordre	27
3.2.5. Model 3.2: model amb Dies com a covariable, correlació autoregressiva de primer ordre i variable Pes aplicant logaritmes	29



3.2.6.	Model 4.2: model amb Dia com a factor fix, correlació autoregressiva de primer ordre i variable Pes aplicant logaritmes.....	30
3.2.7.	Model 5: model amb Dia com a variable continua, aplicant-hi logaritmes i simetria composta ..	32
3.2.8.	Model 6: model amb Dia com a variable continua, aplicant-hi logaritmes i correlació autoregressiva de primer ordre.....	33
3.3.	Resultats amb el model triat .....	35
4.	<b>CONCLUSIONS</b> .....	40
5.	<b>BIBLIOGRAFIA</b> .....	41
5.1.	Referències .....	41
5.2.	Pàgines web consultades .....	42
6.	<b>ANNEXOS</b> .....	45
6.1.	Codi d'R .....	45
6.1.1.	Descriptiva .....	45
6.1.2.	Anàlisi dels possible models .....	48
6.1.3.	Resultats amb el model triat .....	50

# 1. INTRODUCCIÓ

## 1.1. Motivació

Una de les raons principals per escollir aquest treball va ser ampliar els meus coneixements envers els models estadístics; eines útils i necessàries per aproximar-se a la realitat o fer prediccions a partir d'informació extreta de dades, utilitzant fórmules i equacions matemàtiques. Durant el grau s'imparteixen dues assignatures obligatòries sobre aquest àmbit de l'estadística; Models Lineals i Models Lineals Generalitzats, les quals considero fonamentals, molt interessants i enriquidores acadèmicament. Per aquest motiu, vaig decidir acceptar la proposta d'aquest tema, atès que obtindria nous coneixements i experiència sobre models, concretament dels Models Lineals Mixtos amb Mesures Repetides, que no havia estudiat prèviament en la carrera.

D'altra banda, en un futur m'agradaria enfocar la meva formació acadèmica cap a les branques de Bioestadística i Bioinformàtica; la primera disciplina és una especialització/àrea de l'estadística vinculada a la biologia, medicina, epidemiologia, salut pública, etc. En canvi, la segona disciplina permet aplicar eines informàtiques a l'estudi i la gestió de dades biològiques, de salut, etc. (Pérez., 2016). Així doncs, tenir l'oportunitat de fer el Treball Final de Grau sobre aquest tema, és una primera aproximació a aquest món en el qual m'agradaria desenvolupar-me professionalment.

## 1.2. Objectius i hipòtesi

En aquest treball, a partir de les dades de l'evolució en el temps dels pesos de rates classificades en grups segons la dieta amb la qual estan alimentades, es vol veure si es detecten diferències entre les corbes que modelitzen el pes al llarg del temps en funció de la dieta. Per això, realitzem un assaig controlat en el qual comparem els tres grups d'individus (classificats segon la dieta que han rebut: dieta Control, dieta hiperlipídica de blat de moro o dieta hiperlipídica d'oli d'oliva) entre ells. Cada individu s'ha observat en 31 instants de temps diferents (dies des del naixement de la rata, els quals fan referència a l'edat que tenen) que són iguals per als 57 individus.

Per tant, la hipòtesi d'aquest treball és la següent: s'esperen observar diferències entre els tres grups de rates dependent de la dieta amb la qual han estat alimentades (dieta Control, dieta hiperlipídica de blat de moro o dieta hiperlipídica d'oli d'oliva).

## 1.3. Tipus de dietes

En els següents apartats s'explicaran i analitzaran els tres tipus de dietes per a poder així, entendre millor aquest estudi.

### 1.3.1. Lípids i dietes hiperlipídiques

Els lípids<sup>1</sup> (greixos i olis) estan constituïts per triglicèrids (glicerol unit a 3 AG<sup>2</sup>: saturats – AGS, monoinsaturats – AGMI o poliinsaturats – AGPI), àcids grassos lliures, hidrocarburs, fosfolípids (glicerol + 2 AG + fosfat + alcohol), esterols (colesterol – animal, fitosterols – vegetal), tocoferols (activitat vitamina E) i carotenoides (precursora vitamina A) (Linder., 1988).

Els triglicèrids són el representant lípid més abundant en els aliments i en l'organisme (Olivia., 2019). Els àcids grassos que componen els triglicèrids poden ser classificats segons la longitud de cadena (curta: entre 4-6, mitjana: entre 8-12, llarga: entre 14-20 o molt llarga: entre 22 o més), segons el nombre d'insaturacions (saturats: 0, monoinsaturats: 1 o poliinsaturats: 2 o més) i, també, segons la forma (cis o trans) (Linder., 1988).

Les funcions generals dels lípids són:

- Funció energètica (AG, 9 kcal/g): durant la fase d'absorció dels aliments, aquesta energia s'emmagatzema en el teixit adipós. En canvi, en la fase de post absorció els lípids són la font d'energia per excel·lència (exceptuat els teixits glucodependents). El fetge utilitza els lípids per a obtenir energia però, també els transforma de manera que puguin ser nous substrats d'energia utilitzats tant per músculs com pel cor (Linder., 1988).
- Funció reguladora i/o protectora: el teixit adipós subcutani realitza una funció termoreguladora (Linder., 1988).
- Altres: funció de reserva energètica (TAGs<sup>3</sup>), estructural (colesterol, fosfolípids i esfingolípids) i funció biocatalitzadora (vitamines i hormones) (Linder., 1988).

L'alt valor energètic que proporcionen els greixos (lípids) fa que se'ls relacioni directament amb l'increment de reserves de greix corporal i de pes (Olivia., 2019).

L'obesitat és defineix com una acumulació anormal o excessiva de greix que pot ser perjudicial per a la salut (Hernández., 2018), ja que pot desencadenar malalties com: la diabetis, hipertensió (pressió arterial alta), atacs cardíacs a causa de malaltia cardíaca coronària, insuficiència cardíaca i accident cerebrovascular (Martins, *et al.*, 2015), problemes ossis i articulars, apnea del son, problemes del fetge, entre altres.

<sup>1</sup> m BIOQUÍMICA Nom genèric de les substàncies constituïdes per un àcid alifàtic de cadena mitjana i nombre parell d'àtoms de carboni (de 4 a 30 àtoms), lligat per un enllaç èster, de vegades amida o èter, amb composts que tenen una funció alcohol o més (esfingosina, alcohols grassos superiors, etc.).

<sup>2</sup> AG fa referència als Àcids Grassos, una molècula orgànica lipídica formada per una llarga cadena hidrocarbonada lineal, normalment formada per un nombre parell d'àtoms de carboni (16-18 àtoms de carboni, generalment), a l'extrem de la qual hi ha un grup carboxil (-COOH).

<sup>3</sup> TAGs fa referència als triglicèrids.

La major part dels casos amb obesitat són deguts a diversos factors. Es reconeixen factors metabòlics, genètics, ambientals i endocrins, tot i que, la causa principal és la sobrealimentació i l'obesitat exògena<sup>4</sup>. A més, trastorns psicològics també poden influir (Hernández., 2018).

Per a poder tractar l'obesitat, estudis com (Villagrán, *et al.*, 2020) o (Martins, *et al.*, 2015) s'enfoquen en comprendre la relació que hi ha entre aquesta i la ingesta dels greixos. Concretament, l'estudi de (Villagrán, *et al.*, 2020) consisteix a conèixer la regulació de l'emmagatzemament i remoció de triglicèrids (TAGs) en el teixit adipós i la seva implicació en l'obesitat.

A partir de subministrar dietes hiperlipídiques en estudis amb animals, des de la dècada dels 40 (Olivia., 2019), s'estudien tant les condicions que generen l'obesitat com les conseqüències d'aquesta malaltia (comentades anteriorment). Gràcies a la ingesta d'aquestes dietes s'aconsegueix augmentar tant el pes com el greix corporal de l'individu al qual se li subministra, ja que estan generades amb elevats continguts de lípids i, per tant, també d'energia.

Tot i això, les investigacions amb dietes hiperlipídiques també s'apliquen, per exemple, en estudis del desenvolupament cognitiu, capacitat reproductora (Hernández., 2018), etc. (Olivia., 2019). S'ha observat que en rosegadors consumir grans quantitats de greixos i olis afavoreix a l'aparició de tumors en les mamelles i, també, afecta la capacitat reproductora i al manteniment durant el període perinatal (Olivia., 2019).

### 1.3.2. Dieta normolipídica

La dieta normolipídica és aquella que busca tant l'equilibri nutricional com el calòric entre què ingerim i el que cremem. Per tant, aquest tipus de dieta té com a objectiu mantenir el pes i greix corporal adequats de cada individu. A diferència de les dietes hiperlipídiques, ni s'espera ni es busca observar un augment desmesurat del pes de l'individu què pugui comportar problemes de salut, com ara l'obesitat.

## 1.4. Models lineals mixtos

Els models lineals mixtos (*Linear Mixed Models* = LMM ) són una generalització dels models lineals generals i permeten millorar la qualitat de l'anàlisi de factors d'efectes fixos i aleatoris (Riobóo, *et al.*, 2017). També, són models que contempen la possibilitat d'obtenir dades correlacionades i variabilitat heterogènia (Fox, *et al.*, 2015).

---

<sup>4</sup> Obesitat deguda a una ingesta calòrica superior a la necessària per a cobrir les necessitats metabòliques de l'organisme.

Podem referir-nos als LMM com: models multinivell, models jeràrquics o niats, models per a dades longitudinals, models per a mesures repetides, models amb coeficients aleatoris, entre d'altres (Fox, *et al.*, 2015).

En aquest cas, parlarem i estudiarem els models lineals mixtos amb mesures repetides, atès que són els més adequats per ajustar les dades amb les quals treballem.

#### 1.4.1. *Mesures repetides*

Les mesures repetides són útils en aquells estudis/situacions on la variable resposta s'analitza en més d'una ocasió (al llarg del temps) i, molt probablement, en diferents situacions experimentals (segons el sexe, diferents tractaments, diferents dietes, etc.).

En estudis d'àmbits com la medicina, biologia, psicologia, educació, etc. és freqüent analitzar un conjunt de dades que presenten l'evolució d'alguna variable al llarg del temps per a cada individu en diferents situacions experimentals. L'objectiu de l'estudi, doncs, és comparar els diferents grups mitjançant la comparació de cadascuna de les seves corbes i, això, s'aconseguirà a partir d'un model matemàtic que expressi la variable resposta en funció del temps (en la majoria dels casos la funció no és lineal). En aquest treball, per exemple, les diferents situacions experimentals fan referència als tipus de dietes que se subministren a les rates i el Pes fa referència a la variable obtinguda al llarg del temps.

Per tant, les mesures repetides involucren el registre de la variable resposta o dependent sota diverses condicions. En un context experimental, aquestes condicions poden ser: diferents tractaments o ocasions de mesures abans, durant o després de la intervenció. En un context observacional, les mesures es registren en diferents intervals temporals. En tots dos contextos quan l'interès és estudiar la resposta a través del temps, el disseny es concep com a longitudinal (Vargas., 2009). Així doncs, podrem obtenir una corba que ens indicarà com es comporta la variable resposta en qüestió en cada individu i, també, en els grups experimentals.

Els estudis amb dades longitudinals s'utilitzen, principalment, per tres raons: incrementen la sensibilitat considerant la variació dins de cada unitat experimental, estudien canvis a través del temps, perquè els mesuraments repetits proporcionen informació sobre la tendència en el temps de la variable resposta sota diferents condicions de tractament i ús eficient d'unitats experimentals, ja que en general es requereix menys unitats experimentals (Vargas., 2009).

Tanmateix, les mesures repetides també s'utilitzen en estudis on a cada individu se li subministra les diferents situacions experimentals deixant, però, un període de temps entre cada dosi diferent (Pascual, *et al.*, 1996). En aquest cas, no és tan important com evoluciona cada tractament al llarg del temps, sinó com ho fa en cada individu particular.

Alguns dels desavantatges de l'anàlisi de mesures repetides són:

- Les comparacions entre tractaments assignats a grups de subjectes, amb el mateix tractament al llarg de tot l'estudi, es fan de manera menys òptima (Ruiz de Villa., 2004).
- Normalment observem heterogeneïtat entre les mesures repetides i, es reflecteix en una heteroscedasticitat (variància diferent al llarg del temps). Això ocorre, principalment, quan les corbes de creixement, de les diferents unitats experimentals, parteixen pràcticament del mateix punt en el temps inicial i, a mesura que avança el temps es van dispersant cada cop més (Ruiz de Villa., 2004).

En canvi, alguns dels avantatges de les mesures repetides són:

- Reducció del nombre de subjectes (sobretot quan en l'estudi tots els subjectes reben el mateix tractament) (Pascual, *et al.*, 1996).
- Avaluen òptimament els canvis entre unitats experimentals i, dins d'aquestes, al llarg del temps (Ruiz de Villa., 2004).
- Les comparacions entre tractaments assignats a cada unitat experimental es fan amb més precisió gràcies a la correlació positiva que hi ha entre les observacions d'una mateixa unitat experimental (Ruiz de Villa., 2004).

Aquest Treball Final de Grau es divideix en tres grans apartats; una primera part on s'inclou la introducció (la motivació, els objectius, explicació dels tipus de dietes, etc.), una segona part on s'explica amb detall tant el marc teòric (la metodologia emprada, les tècniques estadístiques que s'han fet servir, etc.) com el pràctic (descripció de les dades, anàlisi dels models, comparació i selecció del millor model, etc.) i, per últim, una tercera part on s'inclouen les conclusions.

## 2. METODOLOGIA

### 2.1. Els models lineals mixtos amb mesures repetides

En aquest apartat es descriuran de manera teòrica els models lineals mixtos amb mesures repetides.

El model d'efectes mixtos estén el model clàssic a situacions on els supòsits d'homogeneïtat i independència no són necessaris (Vallejo, *et al.*, 2010), i a situacions on els paràmetres del model poden ser tant fixos com variables aleatòries. El model mixt per analitzar dades longitudinals va ser proposat per Laird i Ware (1982) i, en ell s'assumeix que existeix una relació entre el vector d'observacions  $\mathbf{y}_i$  del subjecte o grup  $i$  i les covariables per mitjà de la següent expressió matricial:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \quad (1)$$

on

- $\mathbf{y}_i$  és el vector ( $k_i \times 1$ ) resposta per a les observacions del  $i$ -èsim grup.
- $\mathbf{X}_i$  és la matriu ( $k_i \times p$ ) de covariables de disseny que relaciona el vector de paràmetres  $\boldsymbol{\beta}$  amb la variable resposta  $\mathbf{y}$ .
- $\boldsymbol{\beta}$  és el vector ( $p \times 1$ ) dels paràmetres dels efectes fixos.
- $\mathbf{Z}_i$  és la matriu ( $k_i \times q$ ) de disseny dels efectes aleatoris que relaciona al vector  $\mathbf{b}$  amb la variable resposta  $\mathbf{y}$ .
- $\mathbf{b}_i$  és un vector ( $q \times 1$ ) d'efectes aleatoris desconeguts específic per a cada participant.
- $\mathbf{e}_i$  és un vector ( $k_i \times 1$ ) de paràmetres desconeguts (els errors), els quals no necessiten ser homogenis ni tampoc independents.

El vector  $\mathbf{b}_i$  és anomenat efecte aleatori perquè aquest canvia la mitjana d'individu a individu i té com a funció millorar l'ajust general donat per l'element  $\mathbf{X}_i\boldsymbol{\beta}$  en agregar la quantitat  $\mathbf{Z}_i\mathbf{b}_i$ .

Els supòsits respecte a les distribucions dels vectors aleatoris del model anterior (1) impliquen que  $\mathbf{e}_i$  i  $\mathbf{b}_i$  són mútuament independents i que estan distribuïts, respectivament, com

$$\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i) \text{ i} \quad (2)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}). \quad (3)$$

En aquest cas,  $\mathbf{R}_i$  és una matriu ( $k_i \times k_i$ ) de covariància pels errors i  $\mathbf{G}$  és una matriu ( $q \times q$ ) de covariància per als efectes aleatoris (Vallejo, *et al.*, 2010).

Aquests supòsits impliquen que, marginalment, el vector resposta  $\mathbf{y}_i$  es distribueix segons una normal i independentment amb mitjana  $\mathbf{X}_i\boldsymbol{\beta}$  i matriu de covariància  $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$ .

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i(\boldsymbol{\theta})) \quad (4)$$

On  $\mathbf{V}_i(\boldsymbol{\theta})$  és una matriu normalment definida per blocs (nivells o individus) (Fox, *et al.*, 2015).

Segons el tipus de dades que estudiem/analitzem, els LMM permeten que  $\mathbf{R}_i$  tingui diferents estructures: sense estructura, simetria composta, AR (1), versions heterogènies, etc. (Fox, *et al.*, 2015).

En el cas que  $\mathbf{V}_i(\boldsymbol{\theta})$  sigui conegut, aleshores les estimacions estàndard per a  $\boldsymbol{\beta}$  i  $\mathbf{b}_i$  són l'estimador de mínim quadrats generalitzats

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i(\boldsymbol{\theta})^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i(\boldsymbol{\theta})^{-1} \mathbf{y}_i \quad (5)$$

i el predictor

$$\hat{\mathbf{b}}_i = \mathbf{G}\mathbf{Z}_i' \mathbf{V}_i(\boldsymbol{\theta})^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}), \quad (6)$$

on  $N$  és el nombre total d'individus del conjunt de dades. Tot i això, els components de la matriu  $\mathbf{V}_i(\boldsymbol{\theta})$  solen ser, generalment, desconeguts. Per tant, abans de poder fer les estimacions de  $\boldsymbol{\beta}$  (5) i  $\mathbf{b}_i$  (6), s'haurà d'estimar la matriu de covariància  $\mathbf{V}_i(\boldsymbol{\theta})$  (Vallejo, *et al.*, 2010).

## 2.2. Procedència de les dades

Les dades emprades per ajustar el model lineal mixt amb mesures repetides han estat proporcionades pels dos tutors. Aquestes dades formen part d'un altre estudi experimental, de la seva recerca, analitzat prèviament.

## 2.3. Programes informàtics utilitzats

El *software* empleat per analitzar les dades ha estat l'R versió 3.6.3. A continuació es comentarà, breument, en què consisteix el llenguatge de programació R, l'RStudio i, per



últim, alguns dels paquets d'R utilitzats per a poder analitzar les dades i extreure'n conclusions.

### 2.3.1. Què és R?

R és un llenguatge de programació enfocat a l'anàlisi estadístic i al desenvolupament de gràfics. R compila i funciona amb sistemes com Windows, Mac OS X i nombroses plataformes UNIX<sup>5</sup> (com Linux). Per a la majoria de plataformes, R es distribueix en format binari per a facilitar la seva instal·lació (Verzani., 2011).

Actualment, aquest *software* és útil per a resoldre un gran ventall de problemes. Tot i això, encara té possibilitats i oportunitats per a desenvolupar-se i seguir creixent. A més a més, R permet la integració amb codi escrit en llenguatge C, C++, Java, etc., per a tasques computacionalment complexes o eines avantatjoses per a altres llenguatges (Verzani., 2011).

### 2.3.2. RStudio

RStudio és un IDE<sup>6</sup> (*Integrated Development Environment*) per al llenguatge de programació R. Aquest, és un projecte què combina diversos components del *software* R (*console*, *source editing*, gràfics, etc.) en un únic *workbench*. RStudio està dissenyat per a facilitar l'aprenentatge a nous usuaris d'R, a més d'oferir eines d'alta productivitat als usuaris més experts. Aquest programa també es pot utilitzar en totes les grans plataformes, com Windows, Mac OS X i Linux (Allaire., 2012).

### 2.3.3. R packages utilitzats

A continuació es comentaran, breument, quatre dels *packages*<sup>7</sup> d'R utilitzats per a visualitzar (gràfics) i analitzar les dades amb les quals treballem.

En primer lloc, el paquet **reshape2()** ha permès transformar les dades en un format “llarg”. És a dir, vam passar de representar un individu per fila a una mesura per fila, per tant, un individu va passar a ocupar tantes files com mesures s'havien fet sobre aquest. Un cop es van tenir les dades en format *long*, es va poder procedir a analitzar les dades amb un model amb efectes aleatoris.

---

<sup>5</sup> UNIX és una família de sistemes operatius. És un sistema operatiu portable, multitasca i multiusuari desenvolupat a partir de 1969.

<sup>6</sup> Un Entorn de Desenvolupament Integrat (IDE) és un sistema de *software* per al disseny d'aplicacions que combina eines de desenvolupament comunes en una única Interfície Gràfica d'Usuari (GUI).

<sup>7</sup> Els *packages* d'R són col·leccions de funcions i dades generades per la comunitat. Aquests, augmenten el poder d'R, atès que milloren les funcionalitats ja existents, o n'afegeixen de noves.

En segon lloc, alguns dels gràfics que es mostren en el següent apartat (per exemple: figura 3.1.1.) s'han creat a partir del paquet **ggplot2()**. Aquest, és un sistema de creació de gràfics basat en *The Grammar of Graphics*<sup>8</sup>; és una implementació de la gramàtica dels gràfics en R (Wickham, *et al.*, 2016).

En tercer lloc, en el *software* R hi ha dos paquets que ens permeten ajustar els models mixtos. Tot i això, en aquest treball només s'ha utilitzat un dels dos; el paquet **nlme()**. Aquest, va ser el primer a aparèixer i va ser escrit inicialment per José Pinheiro (Bell Laboratories) i Douglas Bates (University of Wisconsin) (Cayuela., 2010). La funció del paquet que s'ha emprat per ajustar el model mixt amb mesures repetides ha estat **lme()**. La formulació completa del model utilitzant aquesta funció és de la següent forma:

$$lme(fixed = y \sim 1, random = \sim 1 \mid a/b/c)$$

On *fixed* fa referència als factors fixos i *random* als factors aleatoris (Cayuela., 2010).

Per últim, el paquet **emmeans()** ens ha permès comparar, a partir de les funcions **pairs()**, **cld()** i **emtrends()**, les diferents dietes. És a dir, s'ha pogut analitzar si la tendència de les corbes en funció del temps varia segons la dieta.

---

<sup>8</sup> *The Grammar of Graphics* fa referència al llibre escrit per Leland Wilkinson.

### 3. ANÀLISI DE LES DADES I RESULTATS

#### 3.1. Descriptiva

Una primera aproximació a l'anàlisi d'aquestes dades consisteix a realitzar senzills càlculs estadístics i representacions gràfiques. Aquestes representacions poden fer-se individualment per individu o bé amb un gràfic per a cada grup.

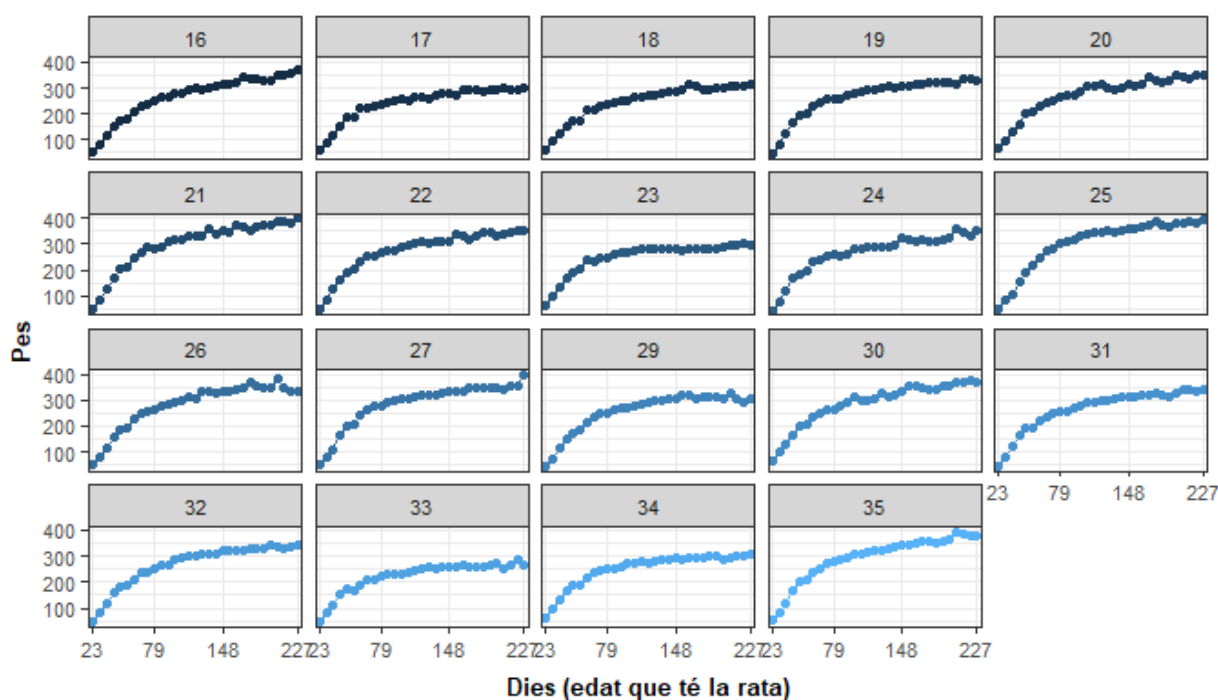


Figura 3.1.1: Representació de les dades segons la dieta Control (normolipídica) per individu.

Taula 3.1.1: Pes mínim, màxim i mitjà (en tres períodes de temps) de les rates amb dieta Control.

Mínim	Primer dia	Meitat de l'estudi	Últim dia	Màxim
44	54.08421	296.8053	341.0158	398.7

A partir de la figura 3.1.1 es pot observar com les corbes dels 19 individus que pertanyen al grup amb dieta Control es comporten de manera similar al llarg del temps.

A més a més, el pes mínim observat en aquest grup d'individus és de 44 grams i el pes màxim no arriba als 400 grams. També, s'ha comprovat que la mitja del pes de les 19 rates durant la meitat de l'estudi era de quasi 300 grams.

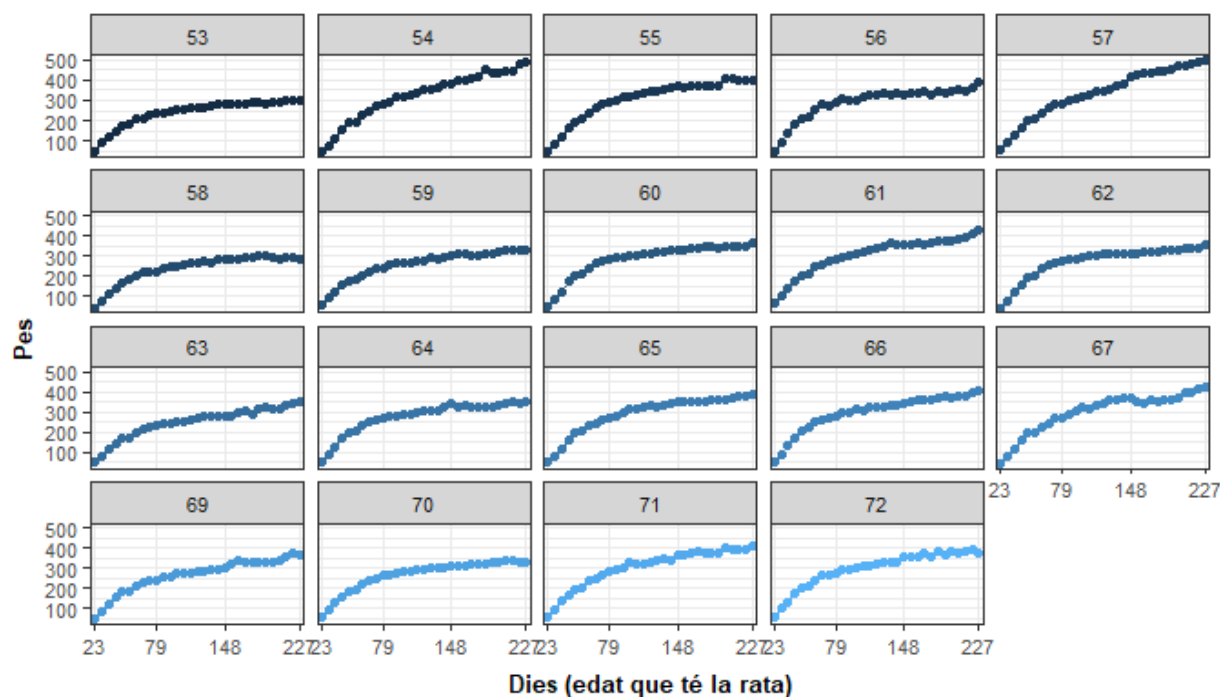


Figura 3.1.2: Representació de les dades segons la dieta hiperlipídica de blat de moro per individu.

Taula 3.1.2: Pes mínim, màxim i mitjà (en tres períodes de temps) de les rates amb dieta hiperlipídica de blat de moro.

Mínim	Primer dia	Meitat de l'estudi	Últim dia	Màxim
43.4	52.01053	310.1474	382.3263	499

D'una banda, a partir de la figura 3.1.2 es pot observar com certes corbes dels individus que pertanyen aquest segon grup es comporten de manera diferent en comparació amb la mitjana del grup al llarg del temps.

D'altra banda, el pes mínim observat en aquest cas és de 43.4 grams i el pes màxim és de 499 grams. A més a més, a diferència del grup de Control, el pes mitjà d'aquestes 19 rates és inferior en el primer dia d'estudi (quan les rates tenen vint-i-tres dies de vida), però tant el pes mitjà observat en la meitat de l'estudi com el de l'últim dia és superior.

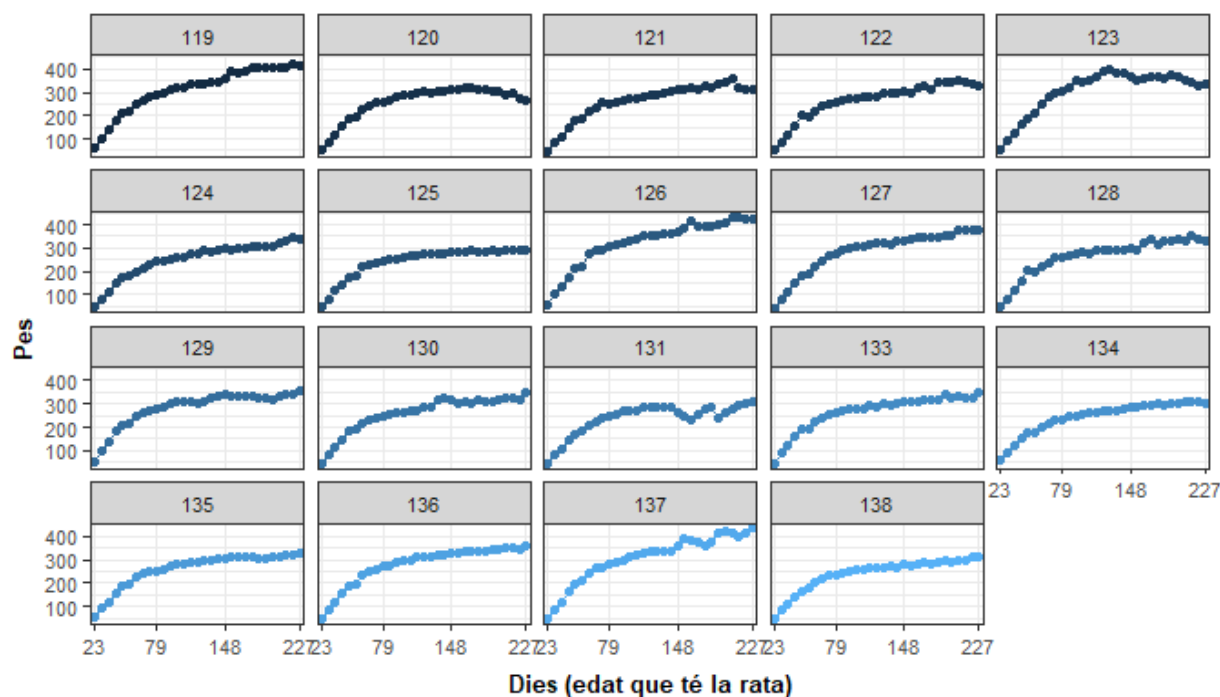


Figura 3.1.3: Representació de les dades segons la dieta hiperlipídica d'oli d'oliva per individu.

Taula 3.1.3: Pes mínim, màxim i mitjà (en tres períodes de temps) de les rates amb dieta hiperlipídica d'oli d'oliva.

Mínim	Primer dia	Meitat de l'estudi	Últim dia	Màxim
42.8	50.10526	301.1605	342.8579	437

En aquest cas, a partir de la figura 3.1.3, també s'han observat com certes corbes de les rates amb dieta hiperlipídica d'oli d'oliva es comporten de manera diferent en comparació amb la mitjana del grup al llarg del temps.

A més a més, tant el pes mínim com el pes mitjà del primer dia observat, en aquests 19 individus, és inferior en comparació amb els dos grups anteriors. Tanmateix, el pes màxim, el pes mitjà en la meitat de l'estudi i en l'últim dia és superior al del grup Control però inferior en comparació amb el grup amb una dieta hiperlipídica de blat de moro.

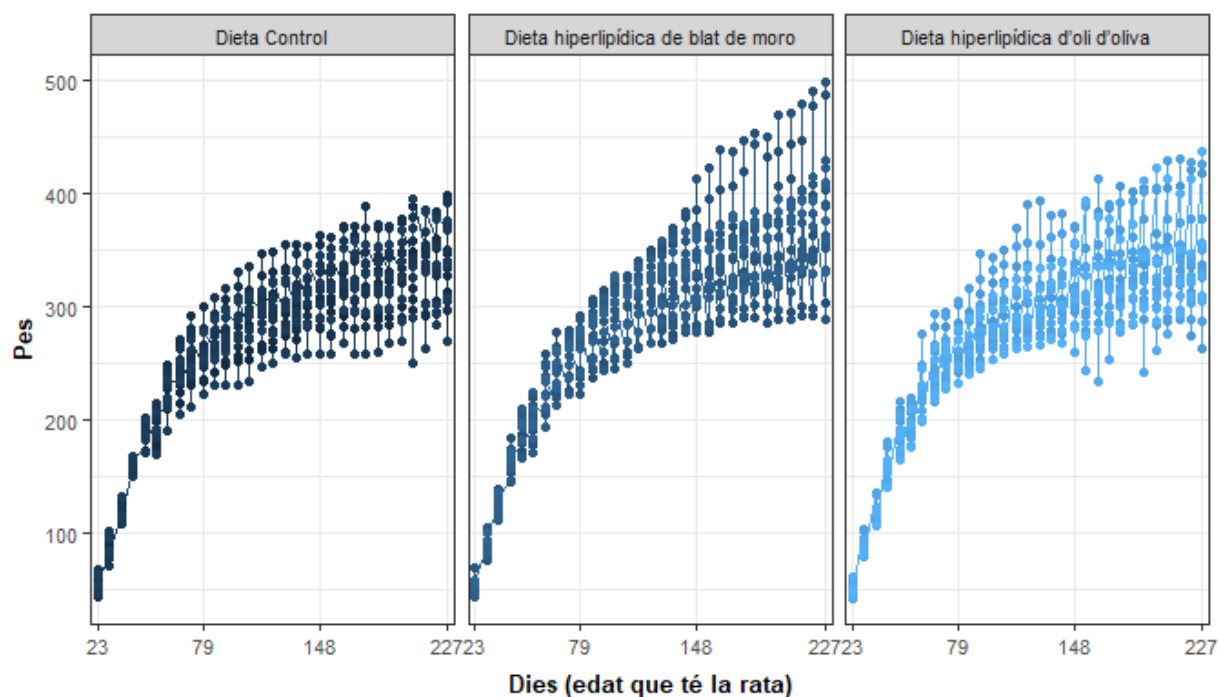


Figura 3.1.4: Representació de les dades segons els tres tipus de dieta.

D'una banda, a partir dels gràfics de la figura 3.1.4 observem com el comportament dels tres grups al llarg del temps no és igual, tot i que s'assemblen; les corbes tenen una orientació similar. Com s'ha comentat prèviament, tant en el grup de la dieta hiperlipídica de blat de moro com en el grup de la dieta hiperlipídica d'oli d'oliva, s'observen diferències en certs individus respecte al comportament mitjà del grup. En canvi, els individus del grup de Control tenen un comportament al llarg dels dies més similar.

D'altra banda, a partir de la figura 3.1.5 observem com el comportament de les corbes dels tres grups segons el pes mitjà al llarg del temps no és igual. És a dir, fins al dia 114 des del naixement de les rates, aproximadament, les corbes es comporten quasi igual. Tot i això, a partir del dia 114, s'observa una clara diferència entre les corbes de la dieta hiperlipídica d'oli d'oliva i la dieta del grup Control, respecte a la dieta hiperlipídica de blat de moro, atès que, aquesta última, presenta un pes mitjà al llarg dels dies més elevat.

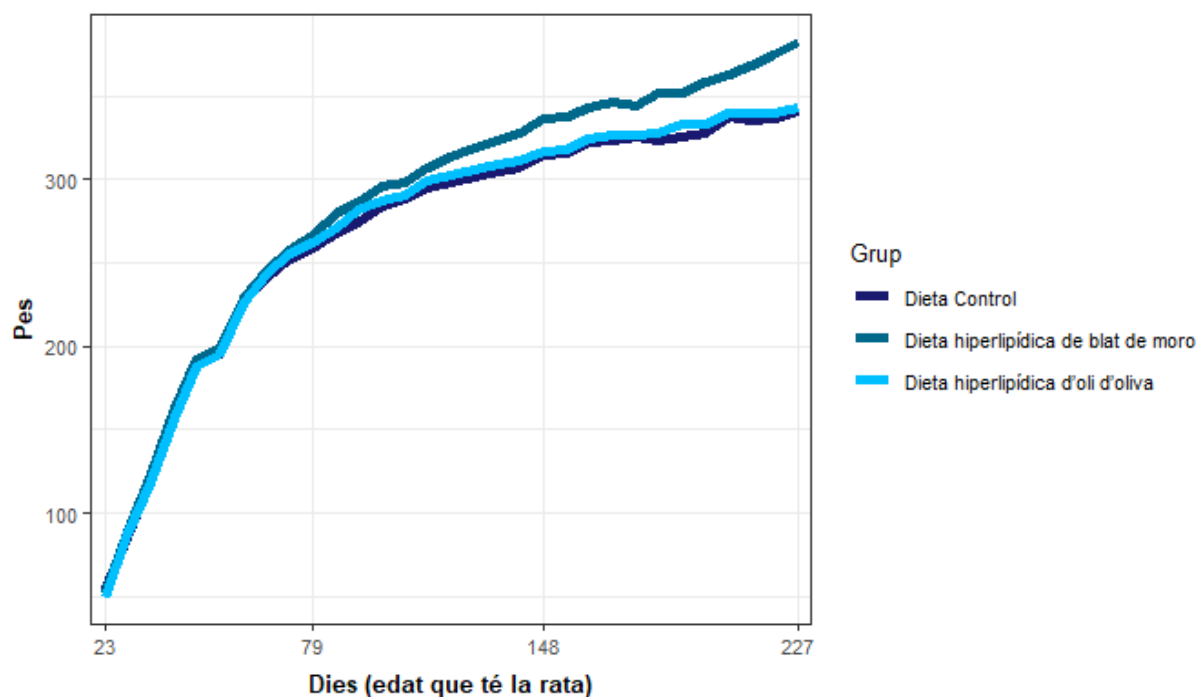


Figura 3.1.5: Representació de les corbes segons el pes mitjà al llarg del temps dels tres tipus de dieta.

## 3.2. Anàlisi dels possibles models

En aquest apartat, s'analitzaran diferents models i es comprovarà quin és el que millor ajusta les dades amb les quals treballem. Comentar que la variable resposta és el Pes de les rates (en grams) en els diferents períodes de temps, la variable Dies fa referència a l'edat que té la rata en dies i la variable Grup fa referència al tipus de dieta que se li ha administrat a l'individu: dieta Control, dieta hiperlipídica de blat de moro i dieta hiperlipídica d'oli d'oliva.

### 3.2.1. Model 1: model amb Dies com a covariable i simetria composta

	numDF	denDF	F-value	p-value
(Intercept)	1	1707	7218.971	<.0001
Dies	1	1707	5159.503	<.0001
Grup	2	54	2.329	0.1071
Dies:Grup	2	1707	12.447	<.0001

Figura 3.2.1.1: Sortida de l'anàlisi ANOVA del model 1.

En primer lloc, a partir de l'anàlisi ANOVA del model 1, observem com la variable Dies és rellevant en el model, atès que presenta un p-valor inferior a 0,05 (nivell de significació fixat). En canvi, la variable Grup no ho és. La interacció d'aquestes dues variables també és rellevant.

En segon lloc, es realitzen un conjunt de gràfics per a poder comprovar si el model ajusta bé les dades. D'una banda, a partir del gràfic de l'esquerra podem identificar si els errors tenen

variància constant; no s'espera observar cap patró i tampoc un augment dels residus amb la resposta, és a dir, esperem que hi hagi homoscedasticitat<sup>9</sup>. D'altra banda, a partir de la funció `qqnorm()`<sup>10</sup> es genera el gràfic de la dreta. Aquest gràfic compara els quantils de les nostres dades amb els quantils teòrics de la distribució normal estàndard,  $N(0, 1)$ . Les dades amb les quals treballem estan representades en l'eix vertical i els quantils teòrics de la distribució Normal en l'eix horitzontal.

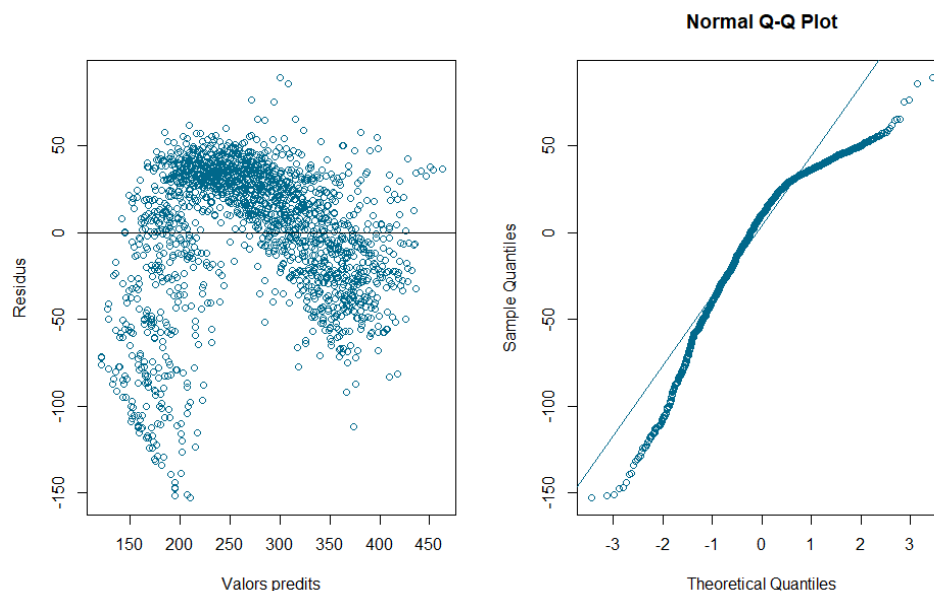


Figura 3.2.1.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 1.

A partir del gràfic de l'esquerra no podem garantir que els errors tinguin variància constant.

Pel que fa al gràfic de la dreta, la no linealitat dels punts podria relacionar-se amb el fet que els errors no es distribueixen segons una Normal i això, ens planteja la necessitat de fer alguna transformació (quadràtica, al cub, etc.) a la variable resposta.

Per últim, es realitza un semivariograma, és a dir, un gràfic que compara la correlació escollida en un model concret amb l'estructura de la correlació teòrica. Els punts fan referència a la correlació escollida i la línia a la teòrica, és a dir, la correlació "real" de les dades.

En aquest cas, els punts no segueixen la forma de la línia horitzontal (figura 3.2.1.3), així doncs, la simetria composta<sup>11</sup> (correlació empírica) s'allunya de la teòrica.

<sup>9</sup> L'homoscedasticitat és una característica d'un model de regressió lineal que implica que la variància dels errors és constant al llarg del temps.

<sup>10</sup> La funció `qqnorm()` es troba dins del paquet `stats()` d'R.

<sup>11</sup> L'estructura de simetria composta fa referència a una correlació uniforme (constant).



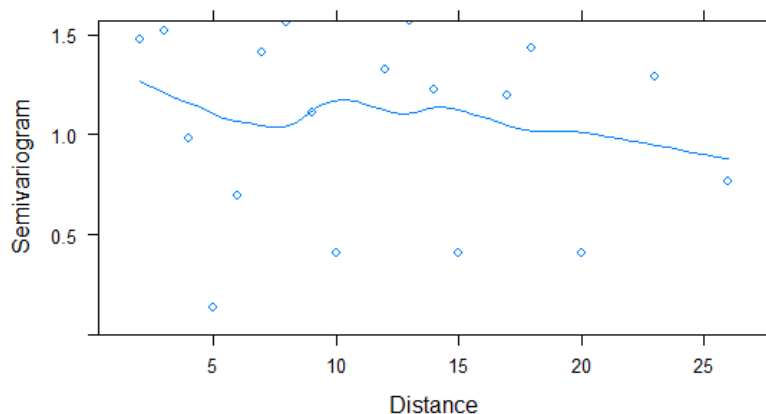


Figura 3.2.1.3: Semivariograma del model 1.

### 3.2.2. Model 2: model amb Dies com a factor fix i simetria composta

	numDF	denDF	F-value	p-value
(Intercept)	1	1620	7218.971	<.0001
factor(Dies)	30	1620	1366.387	<.0001
Grup	2	54	2.329	0.1071
factor(Dies):Grup	60	1620	2.860	<.0001

Figura 3.2.2.1: Sortida de l'anàlisi ANOVA del model 2.

D'una banda, a partir de l'anàlisi ANOVA del model 2 observem com la variable Dies i la interacció són rellevants, atès que presenten un p-valor inferior al nivell de significació fixat del 5%. En canvi, la variable Grup no és rellevant en el model, ja que el seu coeficient presenta un p-valor superior a 0.05.

D'altra banda, a partir dels gràfics de la figura 3.2.2.2, no podem garantir que els errors tinguin variància constant ni tampoc que es distribueixin segons una Normal.

A partir del gràfic de la figura 3.2.2.3, no podem garantir que l'estructura de la correlació empírica (simetria composta) correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal.

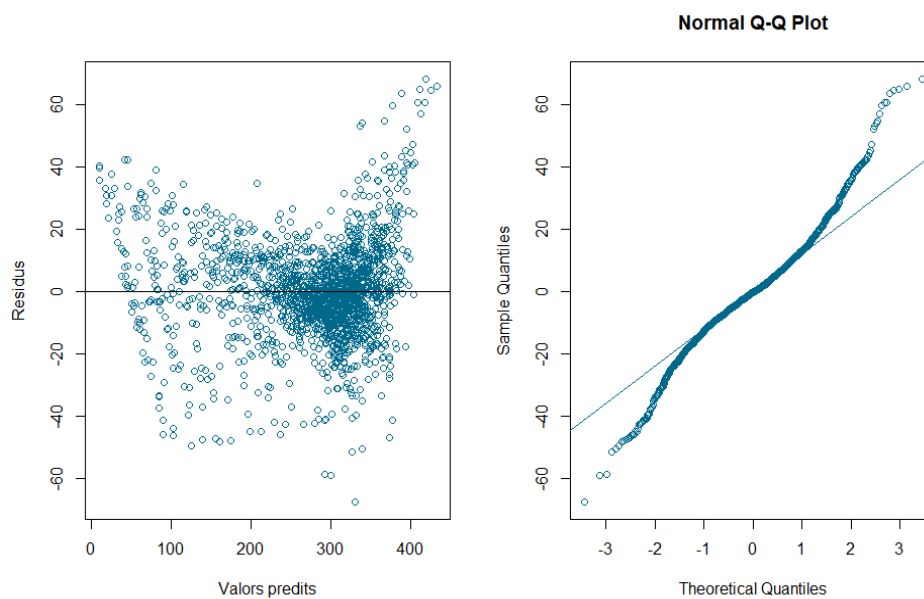


Figura 3.2.2.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 2.

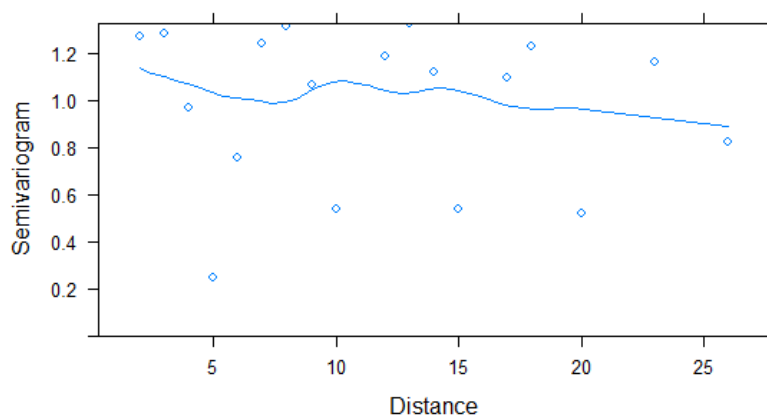


Figura 3.2.2.3: Semivariograma del model 2.

### 3.2.3. Model 3.1: model amb Dies com a covariable i correlació autoregressiva de primer ordre

	numDF	denDF	F-value	p-value
(Intercept)	1	1707	1181.9612	<.0001
Dies	1	1707	1246.1464	<.0001
Grup	2	54	0.8823	0.4197
Dies:Grup	2	1707	2.7340	0.0652

Figura 3.2.3.1: Sortida de l'anàlisi ANOVA del model 3.1.

En primer lloc, a partir de l'anàlisi ANOVA del model 3.1, observem com la variable Dies és rellevant en el model, ja que el seu coeficient presenta un p-valor inferior al nivell de significació fixat del 5%. En canvi, tant la variable Grup com la interacció d'aquesta amb la variable Dies no són rellevants.

En segon lloc, a partir dels dos gràfics de la figura 3.2.3.2, no podem garantir que els errors tinguin variància constant ni tampoc que es distribueixin segons una Normal.

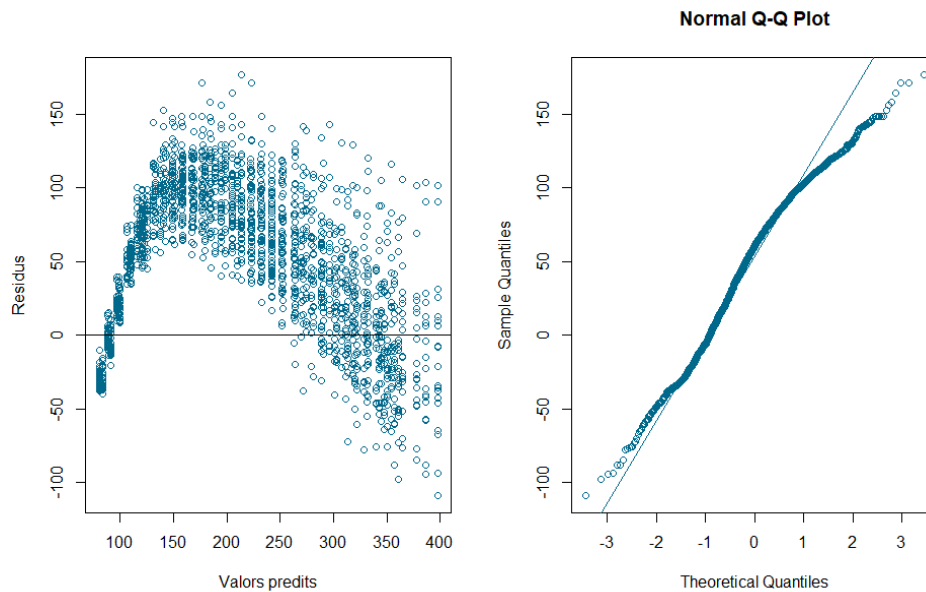


Figura 3.2.3.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 3.1..

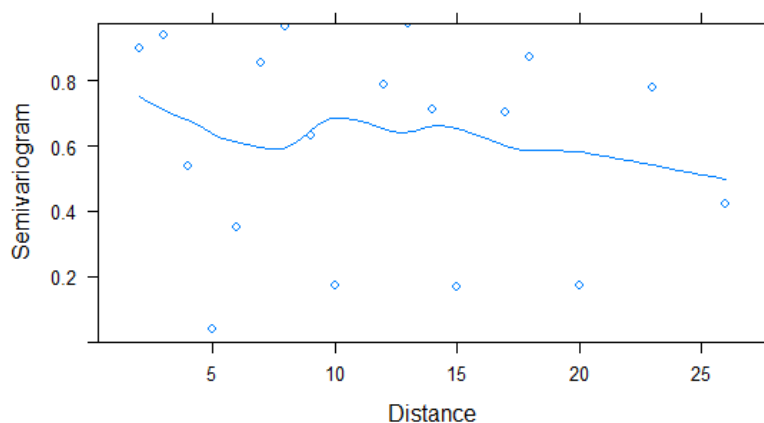


Figura 3.2.3.3: Semivariograma del model 3.1.

En aquest cas, a partir del gràfic de la figura 3.2.3.3, tampoc podem garantir que l'estructura de la correlació autoregressiva d'ordre 1<sup>12</sup> correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal.

<sup>12</sup> La correlació autoregressiva de primer ordre s'utilitza quan s'espera que les correlacions entre observacions estiguin en funció del seu retard en el temps. És a dir, que les observacions més pròximes en el temps tinguin una correlació més alta que les que es troben en temps més llunyans (Vargas., 2009).

### 3.2.4. Model 4.1: model amb Dia com a factor fix i correlació autoregressiva de primer ordre

	numDF	denDF	F-value	p-value
(Intercept)	1	1620	4588.465	<.0001
factor(Dies)	30	1620	251.851	<.0001
Grup	2	54	3.261	0.0460
factor(Dies):Grup	60	1620	1.401	0.0245

Figura 3.2.4.1: Sortida de l'anàlisi ANOVA del model 4.1.

D'una banda, a diferència del model anterior, en el model 4.1 tant la variable Dies com la variable Grup són rellevants en el model, atès que els seus coeficients presenten un p-valor inferior al nivell de significació fixat. A més a més, la interacció d'aquestes dues variables també és rellevant amb un p-valor de 0.0245.

D'altra banda, a partir dels següents dos gràfics (figura 3.2.4.2) no podem garantir que els errors d'aquest model tinguin variància constant ni tampoc que es distribueixin segons una Normal.

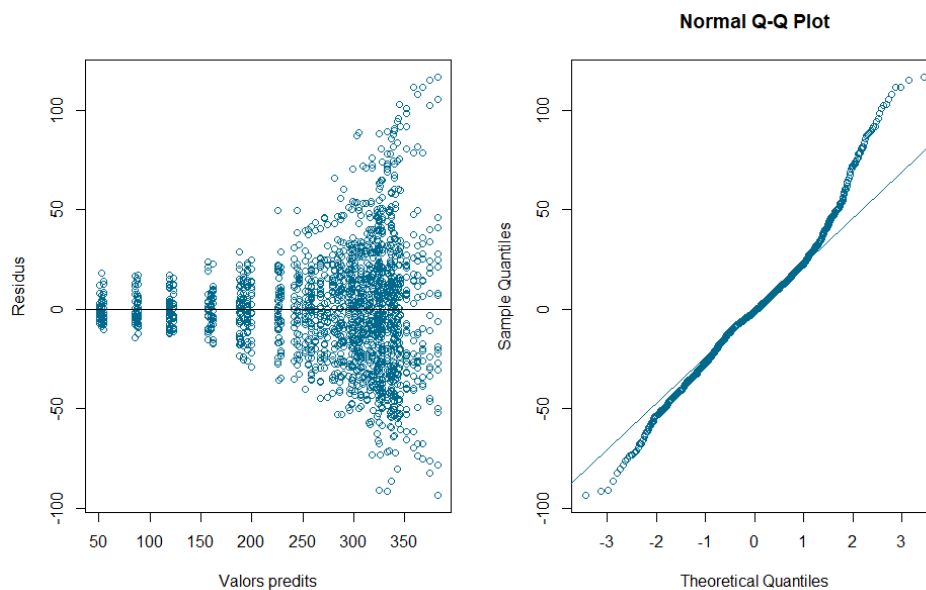


Figura 3.2.4.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 4.1.

En aquest cas, tampoc podem garantir que l'estructura de la correlació autoregressiva de primer ordre correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal (figura 3.2.4.3).

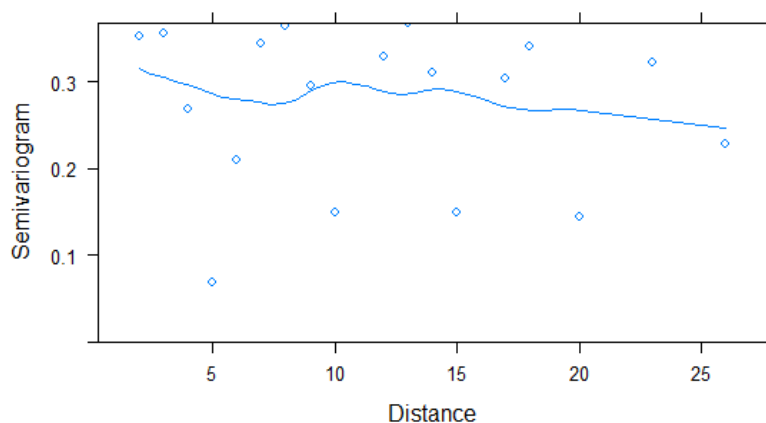


Figura 3.2.4.3: Semivariograma del model 4.1.

Un cop analitzats els quatre models anteriors, s'ha de seleccionar el que millor ajusta les dades.

Taula 3.2.1: Criteris d'informació d'AIC i BIC i test de versemblança dels quatre models anteriors.

Model	1	2	3.1	4.1
<b>AIC</b>	18276.90	14811.21	14437.24	12291.13
<b>BIC</b>	18320.69	15326.39	14486.50	12811.74
<b>LOGLIK</b>	-9130.452	-7310.606	-7209.618	-6049.566

A partir dels resultats anteriors, el model 4.1 (model amb Dia com a factor fix i correlació autoregressiva de primer ordre) és l'òptim segons els criteris d'informació d'AIC i BIC (és el que proporciona un menor valor) i també, el test de versemblança.

Tot i això, cal provar els models transformant la variable resposta (el Pes) aplicant logaritmes per a intentar millorar la distribució dels residus. Concretament farem aquesta transformació en els últims dos models, atès que, l'estructura de simetria composta dels dos primers és molt poc realista. Això és així, perquè treballem amb models de mesures repetides i, a més a més, tenim molts temps i una evolució marcada del pes.

### 3.2.5. Model 3.2: model amb Dies com a covariable, correlació autoregressiva de primer ordre i variable Pes aplicant logaritmes

	numDF	denDF	F-value	p-value
(Intercept)	1	1707	11001.142	<.0001
Dies	1	1707	671.620	<.0001
Grup	2	54	0.137	0.8723
Dies:Grup	2	1707	0.356	0.7003

Figura 3.2.5.1: Sortida de l'anàlisi ANOVA del model 3.2.

En primer lloc, en ajustar aquest model ocorre exactament el mateix que en el model 3.1 anterior. És a dir, observem com la variable Dies és rellevant, ja que el seu coeficient presenta un p-valor inferior al nivell de significació fixat del 5%. En canvi, tant la variable Grup com la interacció no ho són.

En segon lloc, a partir dels gràfics de la figura 3.2.5.2, tampoc podem garantir que els errors tinguin variància constant ni tampoc que es distribueixin segons una Normal. És més, sembla que ha augmentat la no linealitat dels punts respecte al model 3.1.

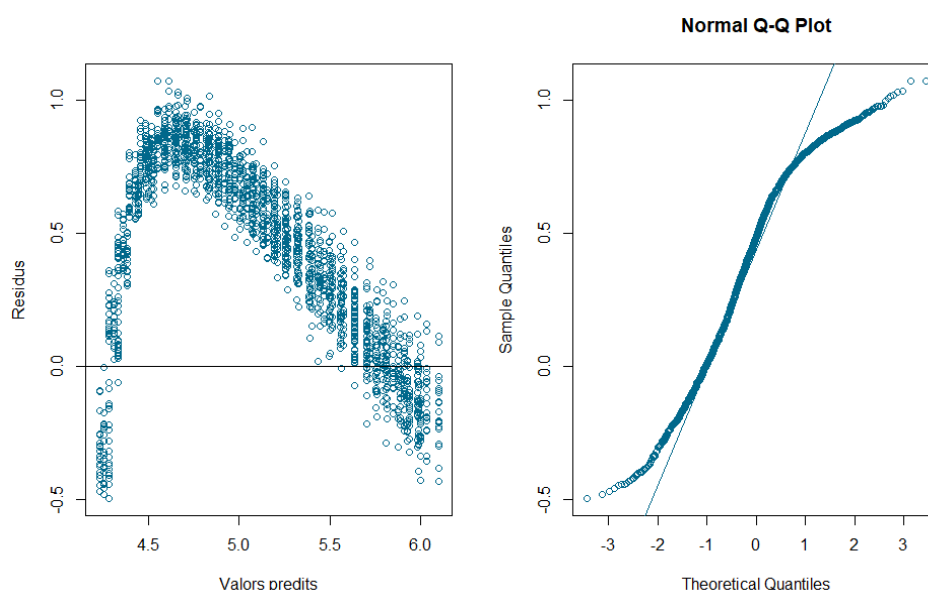


Figura 3.2.5.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 3.2.

Per últim, a partir del gràfic de la figura 3.2.5.3, tampoc podem garantir que l'estructura de la correlació autoregressiva correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal.

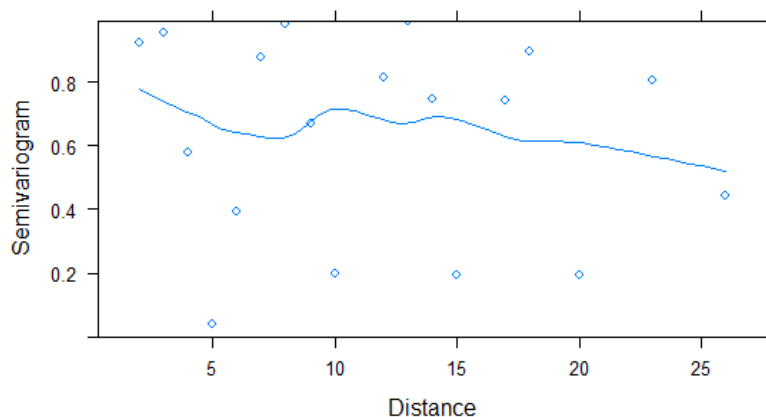


Figura 3.2.5.3: Semivariograma del model 3.2.

### 3.2.6. Model 4.2: model amb Dia com a factor fix, correlació autoregressiva de primer ordre i variable Pes aplicant logaritmes

	numDF	denDF	F-value	p-value
(Intercept)	1	1620	202752.78	<.0001
factor(Dies)	30	1620	1224.62	<.0001
Grup	2	54	2.46	0.0952
factor(Dies):Grup	60	1620	2.68	<.0001

Figura 3.2.6.1: Sortida de l'anàlisi ANOVA del model 4.2.

D'una banda, a diferència del model 4.1, la variable Grup deixa de ser rellevant en aquest nou model, ja que el seu coeficient presenta un p-valor superior al nivell de significació fixat. En canvi, tant la variable Dies com la interacció, ho segueixen sent.

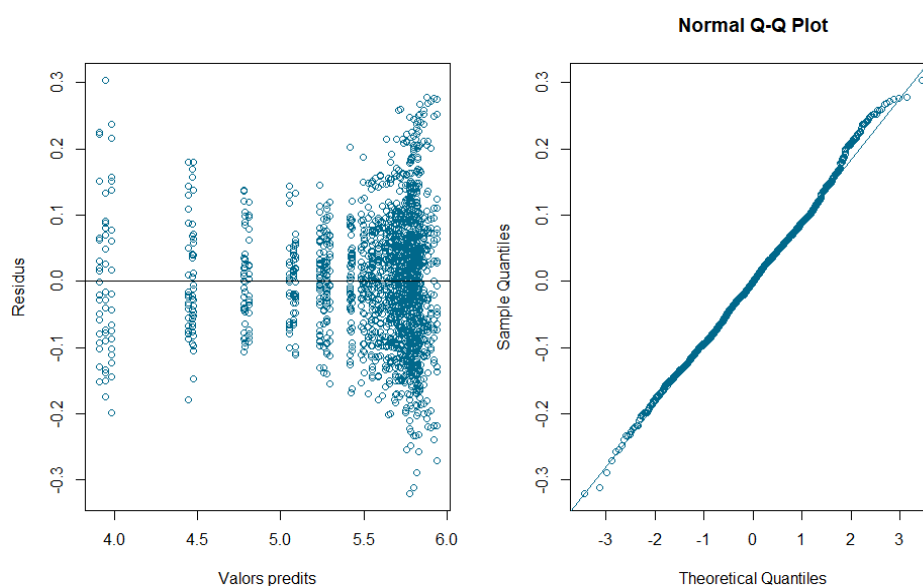


Figura 3.2.6.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 4.2.

D'altra banda, a partir dels dos gràfics anteriors (figura 3.2.6.2) observem una millora considerable respecte els obtinguts en el model 4.1. Els errors d'aquest nou model semblen tenir una variància més constant i, a més, ha millorat la no linealitat dels punts; podríem pensar que els errors es distribueixen segons una Normal.

En aquest cas, tampoc podem garantir que l'estructura de correlació autoregressiva de primer ordre correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal (figura 3.2.6.3).

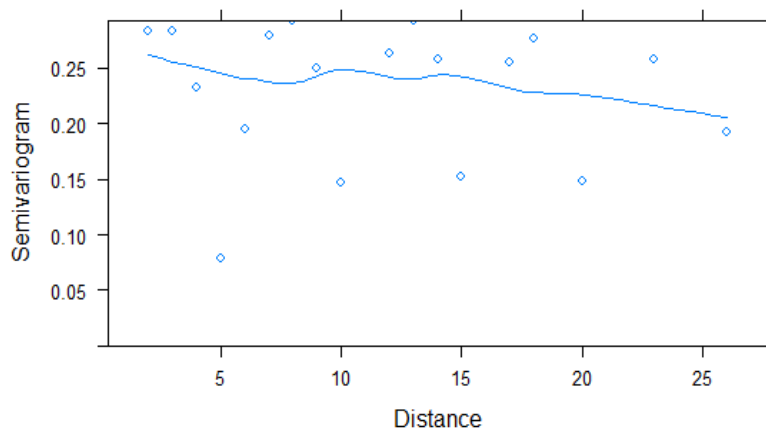


Figura 3.2.6.3: Semivariograma del model 4.2.

Un cop analitzats els dos models anteriors, on s'ha transformat la variable resposta (el Pes) aplicant logaritmes, s'ha de seleccionar el que millor ajusta les dades. A partir dels resultats de la taula 3.2.2, el model 4.2 (model amb Dia com a factor fix i correlació autoregressiva de primer ordre) és l'òptim segons els criteris d'informació d'AIC i BIC i també, el test de versemblança.

Taula 3.2.2: Criteris d'informació d'AIC i BIC i test de versemblança dels quatre models anteriors.

Model	3.2	4.2	3.1	4.1
<b>AIC</b>	-2280.798	-6359.321	14437.24	12291.13
<b>BIC</b>	-2231.535	-5838.715	14486.50	12811.74
<b>LOGLIK</b>	1149.399	3275.66	-7209.618	-6049.566

Tot i això, després de l'anàlisi i la interpretació dels sis models anteriors, ajustarem dos models nous tractant la variable Dies com a contínua i no com a factor fix. Es considera que



la millor manera d'ajustar aquestes dades és a partir de corbes tractant el temps com una variable contínua per a no perdre tanta informació.

Concretament, s'ajustaran dos models nous transformant la variable temps (els Dies) aplicant logaritmes. Tot i que, com s'ha comentat prèviament, l'estructura de simetria composta és molt poc realista en el nostre cas, el primer model s'ajustarà amb aquesta per a poder, així, contrastar i comprovar les diferències respecte al model ajustat amb correlació autoregressiva de primer ordre.

### 3.2.7. Model 5: model amb Dia com a variable continua, aplicant-hi logaritmes i simetria composta

	numDF	denDF	F-value	p-value
(Intercept)	1	1707	7218.971	<.0001
log(Dies)	1	1707	19680.622	<.0001
Grup	2	54	2.329	0.1071
log(Dies):Grup	2	1707	33.064	<.0001

Figura 3.2.7.1: Sortida de l'anàlisi ANOVA del model 5.

En primer lloc, a partir de l'anàlisi ANOVA del model 5, observem com la variable Dies aplicant logaritmes és rellevant en el model, ja que el seu coeficient presenta un p-valor inferior al nivell de significació fixat del 5%. En canvi, la variable Grup no és rellevant a diferència de la interacció que si ho és.

En segon lloc, a partir dels dos gràfics de la figura 3.2.7.2, no podem garantir que els errors tinguin variància constant ni tampoc que es distribueixin segons una Normal.

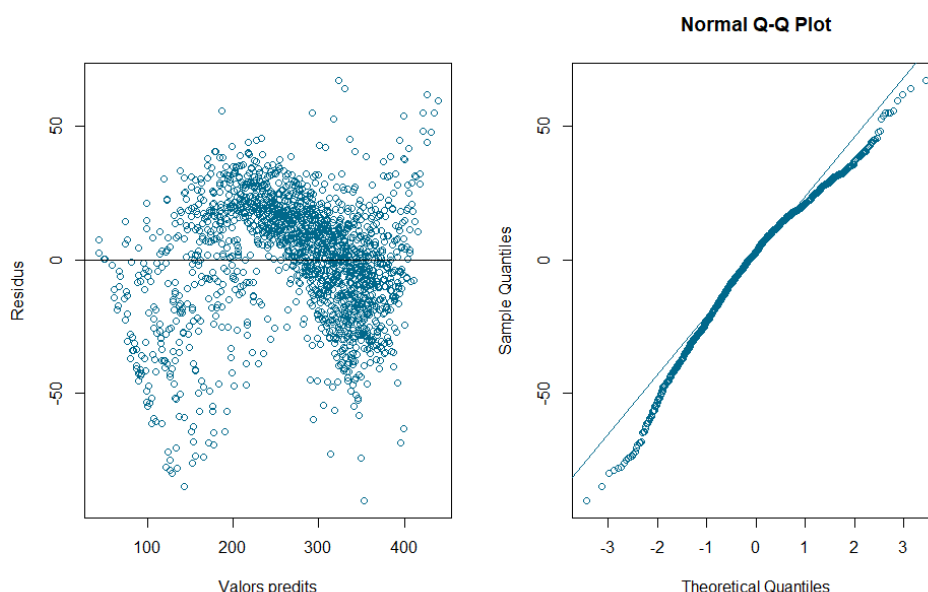


Figura 3.2.7.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 5.

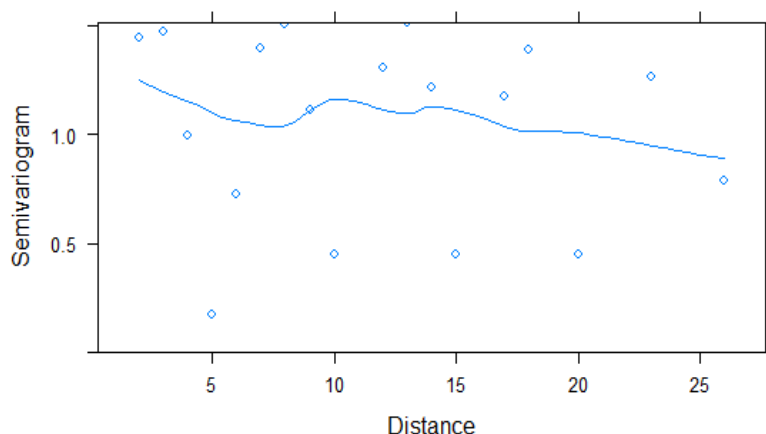


Figura 3.2.7.3: Semivariograma del model 5.

Per últim, a partir del gràfic de la figura 3.2.7.3, no podem garantir que l'estructura de la correlació empírica (simetria composta) correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal.

### 3.2.8. Model 6: model amb Dia com a variable continua, aplicant-hi logaritmes i correlació autoregressiva de primer ordre

	numDF	denDF	F-value	p-value
(Intercept)	1	1707	4282.977	<.0001
log(Dies)	1	1707	2842.138	<.0001
Grup	2	54	2.385	0.1017
log(Dies):Grup	2	1707	5.277	0.0052

Figura 3.2.8.1: Sortida de l'anàlisi ANOVA del model 6.

D'una banda, la variable Grup no és rellevant en aquest nou model, ja que el seu coeficient presenta un p-valor superior al nivell de significació fixat. En canvi, tant la variable Dies com la interacció ho són.

D'altra banda, a partir dels següents dos gràfics (figura 3.2.8.2) observem una millora respecte als obtinguts en el model 5, ja que quasi no s'observa no linealitat en els punts; podríem dir que els errors es distribueixen segons una Normal. Tot i això, no podem garantir que els errors tinguin variància constant.

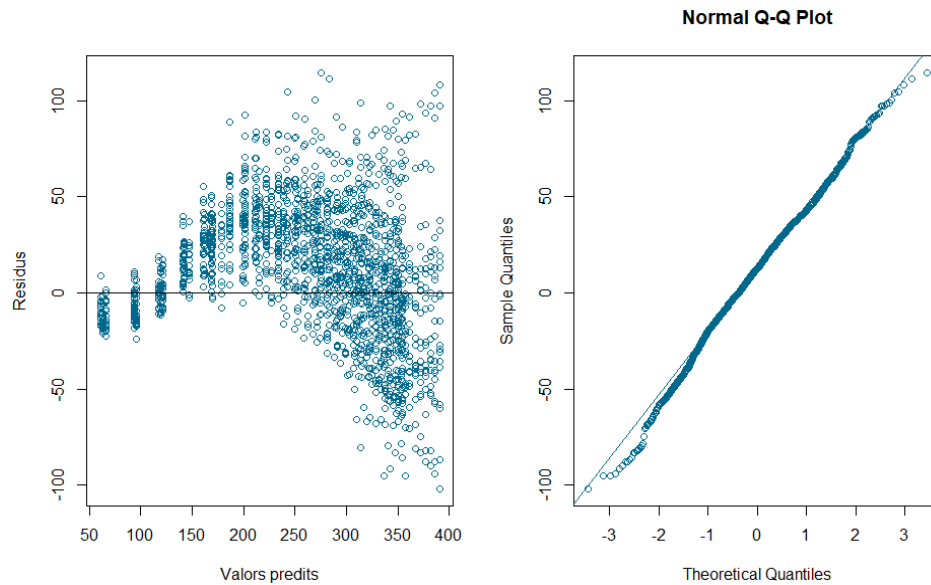


Figura 3.2.8.2: Gràfic dels residus vs els valors predits i gràfic Normal Q-Q Plot del model 6..

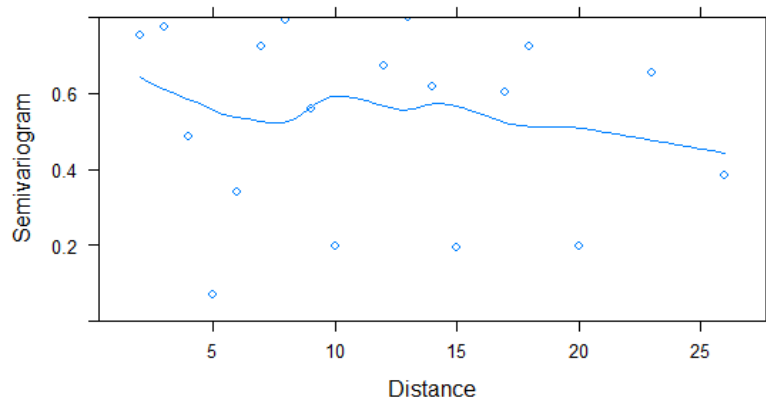


Figura 3.2.8.3: Semivariograma del model 6.

En aquest cas, tampoc podem garantir que l'estructura de la correlació autoregressiva de primer ordre correspongui a l'estructura de la correlació teòrica, ja que els punts no segueixen la forma de la línia horitzontal (figura 3.2.8.3).

Un cop analitzats els dos models anteriors, s'ha d'escollir quin és el que millor ajusta les dades. Comentar que no es compararan amb els anteriors, atès que no és massa adient comparar els AICs, BICs, etc. de dos models en el que en un s'ha fet una transformació logarítmica de la variable temps i en l'altre o no s'ha fet transformació o se n'ha fet una de diferent. Per això, només comparem entre el model 5 i el model 6.

Taula 3.2.3: Criteris d'informació d'AIC i BIC i test de versemblança dels models 5 i 6.

Model	5	6
<b>AIC</b>	16312.89	13301.36
<b>BIC</b>	16362.16	13350.63
<b>LOGLIK</b>	-8147.447	-6641.682

A partir dels resultats de la taula 3.2.3, l'últim model analitzat (el model 6) és l'òptim segons els criteris d'informació d'AIC i BIC (és el que proporciona un menor valor) i també, el test de versemblança.

### 3.3. Resultats amb el model triat

A partir de les anàlisis i resultats anteriors, seleccionem el model amb Dia com a variable continua, aplicant-hi logaritmes i correlació autoregressiva de primer ordre (model 6), com el més adequat, dels analitzats, per ajustar el conjunt de dades. Així doncs, a continuació s'analitzarà l'estimació d'aquest model.

El model seleccionat és de la següent forma:

$$Y = \beta_0 + \beta_1 * \log(Dies) + \beta_2 * GrupM + \beta_3 * GrupO + \beta_4 * \log(Dies) * GrupM + \beta_5 * \log(Dies) * GrupO + \varepsilon$$

El resultat de l'estimació d'aquest model (figura 3.3.1) mostra que  $\hat{\beta}_0 = -327.9153$  i  $\hat{\sigma}_{\beta_0} = 0.008445362$ . La desviació típica residual s'estima com  $\hat{\sigma}_\varepsilon = 33.78739$ . També obtenim que  $\hat{\beta}_1 = 125.7827$ ,  $\hat{\beta}_2 = -62.9926$ ,  $\hat{\beta}_3 = -10.8686$ ,  $\hat{\beta}_4 = 18.3056$  i  $\hat{\beta}_5 = 2.5838$ .

```

Linear mixed-effects model fit by REML
Data: dades_long
      AIC      BIC    logLik
13301.36 13350.63 -6641.682

Random effects:
Formula: ~1 | ID_animal
      (Intercept) Residual
StdDev: 0.008445362 33.78739

Correlation Structure: Continuous AR(1)
Formula: ~log(Dies) | ID_animal
Parameter estimate(s):
      Phi
0.4780977
Fixed effects: Pes ~ log(Dies) * Grup
              Value Std.Error   DF    t-value p-value
(Intercept) -327.9153 19.322950 1707 -16.970251 0.0000
log(Dies)    125.7827  4.312797 1707  29.164994 0.0000
GrupM        -62.9926 27.326778   54  -2.305162 0.0250
Grup0       -10.8686 27.326778   54  -0.397728 0.6924
log(Dies):GrupM  18.3056  6.099215 1707   3.001308 0.0027
log(Dies):Grup0   2.5838  6.099215 1707   0.423634 0.6719
Correlation:
              (Intr) lg(Ds) GrupM  Grup0  l(D):GM
log(Dies)    -0.955
GrupM        -0.707  0.676
Grup0        -0.707  0.676  0.500
log(Dies):GrupM  0.676 -0.707 -0.955 -0.478
log(Dies):Grup0  0.676 -0.707 -0.478 -0.955  0.500

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-3.0148497 -0.2765145  0.3726927  1.0350116  3.3897348

Number of Observations: 1767
Number of Groups: 57

```

Figura 3.3.1: Sortida del summary del model 6.

A més a més, els resultats de l'ANOVA del model 6 (figura 3.2.8.1.) ens indiquen que la interacció entre la variable Dies i Grup és rellevant, atès que el seu coeficient és significatiu. Això implica que segons la dieta que se l'hi subministri a les rates, s'obtenen perfils d'evolució en el temps diferents (com bé surt representat en la figura 3.1.5). És a dir, la tendència de les corbes en funció del temps varia segons la dieta. Així doncs, a continuació s'analitzarà entre quines d'aquestes tres dietes s'observen les diferències.

En primer lloc, es grafiquen els *Estimated Marginal Means* de les diferents dietes. En el gràfic següent (figura 3.3.2), les barres representen els intervals de confiança per als EMMs<sup>13</sup> i, les fletxes vermelles s'utilitzen per a comparar els diferents grups entre ells. Si una fletxa d'una mitjana se superposa amb una fletxa d'un altre grup, implica que la diferència entre aquests dos grups no és significativa.

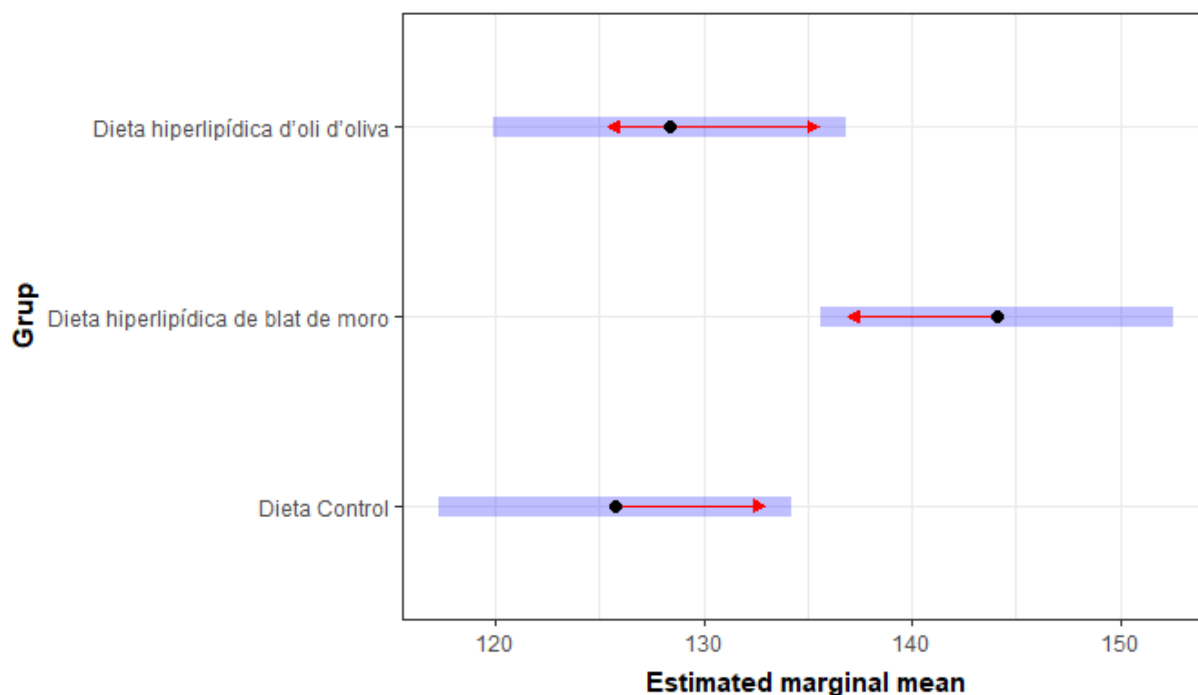


Figura 3.3.2: Valor mitjà estimat per cada nivell del factor dieta.

Per tant, a partir del gràfic anterior, s'observen diferències entre la dieta hiperlipídica de blat de moro i les altres dues. En canvi, no se'n detecten entre la dieta Control i la dieta hiperlipídica d'oli d'oliva, atès que les fletxes vermelles d'aquests dos grups se superposen.

Aquestes diferències també poden ser detectades a partir de les dues figures següents. D'una banda, en la figura 3.3.3, en fer les comparacions dos a dos entre les tres dietes, obtenim un p-valor superior al nivell de significació fixat en comparar la dieta Control i la dieta hiperlipídica d'oli d'oliva. Per tant, no podem garantir que hi hagi diferències significatives entre aquestes dues. En canvi, sí que s'observen diferències entre aquestes dues dietes i la dieta hiperlipídica de blat de moro, atès que obtenim dos p-valors inferiors a 0.05.

D'altra banda, a partir de la figura 3.3.4, obtenim els mateixos resultats. En observar l'última columna anomenada *group*, comprovem que la dieta Control i la dieta hiperlipídica d'oli d'oliva han obtingut el mateix número (un 1) i, això implica que no s'observen diferències entre aquestes dues dietes.

<sup>13</sup> EMMs fa referència als *Estimated Marginal Means*.

contrast	estimate	SE	df	t.ratio	p.value
C - M	-18.31	6.1	1707	-3.001	0.0077
C - O	-2.58	6.1	1707	-0.424	0.9058
M - O	15.72	6.1	1707	2.578	0.0271

Degrees-of-freedom method: containment

P value adjustment: tukey method for comparing a family of 3 estimates

Figura 3.3.3: comparacions dos a dos entre les dietes.

Grup	log(Dies).trend	SE	df	lower.CL	upper.CL	.group
C	126	4.31	1707	117	134	1
O	128	4.31	1707	120	137	1
M	144	4.31	1707	136	153	2

Degrees-of-freedom method: containment

Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 3 estimates

significance level used: alpha = 0.05

Figura 3.3.4: Grups resultants de les comparacions múltiples.

En resum, es detecten diferències entre la dieta hiperlipídica de blat de moro i les altres dues, però no se'n detecten entre la dieta Control i la dieta hiperlipídica d'oli d'oliva.

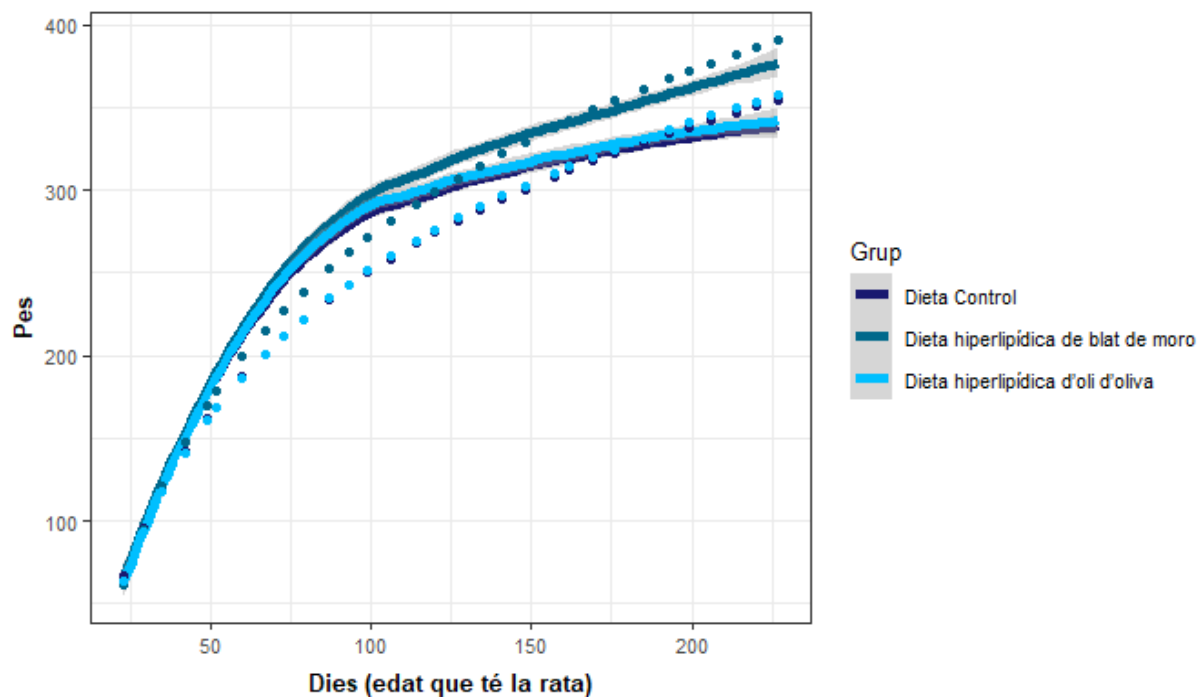


Figura 3.3.5: Representació de les corbes segons el pes mitjà al llarg del temps dels tres tipus de dieta i les prediccions obtingudes a partir del model 6.

Per últim, es representen gràficament les corbes segons el pes mitjà al llarg del temps dels tres tipus de dieta i les prediccions obtingudes a partir del model 6. Com podem observar en la figura 3.3.5, fins al dia 50 i a partir del dia 150, aproximadament, el model genera bones prediccions. Tot i això, entre aquests dos dies les prediccions no són del tot bones (fixar-nos en els punts del gràfic). Això pot estar relacionat amb la variància dels residus que, com hem vist prèviament (figura 3.2.8.2), no és constant.



## 4. CONCLUSIONS

Al llarg d'aquest treball, s'ha intentat donar una visió de què són i per què s'utilitzen els models lineals mixtos (LMM, en anglès). D'una banda, quan en certes situacions els models de regressió clàssics no són adients per elaborar una correcta anàlisi de les dades, s'han de tenir en compte i emprar els LMM. Aquests, són una generalització dels models lineals i permeten incorporar factors d'efectes fixos, aleatoris i covariables. També contemplen la possibilitat d'obtenir dades correlacionades i heteroscedasticitat.

D'altra banda, els models lineals d'efectes mixtos són una bona eina per modelitzar experiments amb dades longitudinals (en les quals s'estudia la variació de la resposta al llarg del temps; interessa conèixer el model que relaciona la resposta amb el temps) perquè permeten l'anàlisi tant de la variació entre unitats com la variació dins de les unitats. Addicionalment, els LMM permeten modelar diferents estructures de correlació (simetria composta, auto-regressiva, sense especificar, etc.) entre les mesures temporals, en funció del tipus de dades que s'estudien.

No obstant això, els models lineals mixtos són una àrea de recerca puntera on encara queden preguntes per aclarir, per exemple, com seleccionar l'estructura de correlacions més correcta.

En aquest projecte, hem treballat amb mesures repetides i, s'ha pogut comprovar que són d'utilitat en situacions on la variable resposta s'analitza en més d'una ocasió i en diferents situacions experimentals. A més a més, quan l'interès d'anàlisi en les mesures repetides és l'estudi de la resposta a través del temps, el disseny es concep com a longitudinal. En quant a les dades amb les quals s'ha treballat, les diferents situacions experimentals fan referència als tipus de dietes que se subministren a les rates i el pes fa referència a la variable obtinguda al llarg del temps.

Per últim, gràcies a modelar un LMM s'han pogut detectar diferències entre les corbes que modelitzen el pes al llarg del temps en funció de la dieta; dieta Control, dieta hiperlipídica de blat de moro o dieta hiperlipídica d'oli d'oliva. En primer lloc, a partir dels resultats de l'ANOVA del model seleccionat, s'ha comprovat que segons la dieta que se li subministri a les rates, s'obtenen perfils d'evolució en el temps diferents. A partir de gràfics i anàlisis estadístics, s'han observat diferències entre la dieta hiperlipídica de blat de moro i les altres dues. Per contra, no se n'han detectat entre la dieta Control i la dieta hiperlipídica d'oli d'oliva.

En resum, els models lineals mixtos són una eina que ens permet extreure conclusions sobre experiments que, per la naturalesa de les dades, no poden ser analitzats a partir dels models clàssics. A més, s'ha comprovat que existeixen diferències entre els tres tipus de dietes.

## 5. BIBLIOGRAFIA

### 5.1. Referències

Allaire, J. (2012). RStudio: integrated development environment for R. *Boston, MA*, 770, 394.

Cayuela, L. (2010). *Modelos lineales mixtos en R*.

Fox, J. Weisberg, S. (2015). “Apèndix: MODELOS LINEALES GENERALIZADOS MIXTOS ALGUNOS CASOS PRACTICOS”. *An R Companion to Applied Regression*, segona edició.

Hernández Martínez, I. (2018). Efecto de la dieta hiperlipídica en la toxicidad reproductiva inducida por fluoruro en ratas Wistar macho. URI:<http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/handle/231104/2425>

Linder, M.C. *Nutrición: aspectos bioquímicos, metabólicos y clínicos*. Pamplona: EUNSA; 1988.

Martins, F., Campos, D. H. S., Pagan, L. U., Martinez, P. F., Okoshi, K., Okoshi, M. P., ... & Oliveira-Junior, S. A. D. (2015). High-fat diet promotes cardiac remodeling in an experimental model of obesity. *Arquivos brasileiros de cardiologia*, 105(5), 479-486. DOI:<https://doi.org/10.5935/abc.20150095>

Oliva Lorenzo, L. (2019). *Efecto de la proporción de los componentes de la dieta en distintos aspectos del metabolismo de la rata en un contexto de obesidad* (Doctoral dissertation, Universitat de Barcelona). URL:<http://hdl.handle.net/10803/667435>

Pascual, J., Frías, D., & García, F. (1996). *Manual de psicología experimental. Metodología de investigación*. Libro Entero. Primera Edición. Editorial Ariel, SA Barcelona.

Pérez Porto, J. Gardey, A. (2016). *Definición de bioestadística*. Definición.de. Recuperat de <https://definicion.de/bioestadistica/>

Pérez Porto, J. Merino, M. (2016). *Definición de bioinformática*. Definición.de. Recuperat de <https://definicion.de/bioinformatica/>

Riobóo, L. D., & Pacheco, M. P. (2017). Modelos lineales generales y mixtos en la caracterización de la variable calificación, Ingeniería Agroindustrial, UNI-NORTE. *Nexo Revista Científica*, 30(2), 84-95.

Ruiz de Villa, M. D. C. (2004). Análisis de medidas repetidas. *Med. clín (Ed. impr.)*, 51-58.

Vallejo, G., Fernández, M. P., Tuero, E., & Rojas, P. E. L. (2010). Análisis de medidas repetidas usando métodos de remuestreo. *Anales de Psicología/Annals of Psychology*, 26(2), 400-409.

Vargas, A. (2009). Modelos Lineales Mixtos con estructura de correlación en el Análisis de Datos Longitudinales. Un caso aplicado. In *Anales Científicos* (Vol. 70, No. 3, pp. 15-24).

Verzani, J. (2011). *Getting started with RStudio*. "O'Reilly Media, Inc."

Villagrán, M., Troncoso-Pantoja, C., Leiva, A. M., Martínez-Sanguinetti, M. A., Petermann-Rocha, F., & Celis-Morales, C. (2020). Dinámica del recambio de lípidos y sus implicancias en la obesidad durante el ciclo vital. *Revista chilena de nutrición*, 47(4), 692-693. DOI:<https://dx.doi.org/10.4067/S0717-75182020000400692>

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create Elegant Data Visualisations Using the Grammar of Graphics*. Version, 2(1), 1-189.

## 5.2. Pàgines web consultades

Angélica M. Argote Vargas y Pedro A. Velasquez-Vasconez. (2018, 1 febrer). *Análisis de distribución normal - datos angélica*. RPubs. Recuperat de <https://rpubs.com/PAVelasquezVasconez/354989>

Barajas, F. H. (2020, 18 febrero). *2 Modelos Mixtos / Modelos Mixtos con R*. GitHub. Recuperat de [https://fhernanb.github.io/libro\\_modelos\\_mixtos/mod-mix.html](https://fhernanb.github.io/libro_modelos_mixtos/mod-mix.html)

Colaboradores de Wikipedia. (2020, 11 gener). *Variograma*. Wikipedia, la enciclopedia libre. Recuperat de <https://es.wikipedia.org/wiki/Variograma#:~:text=El%20variograma%20o%20semivariograma%20es,otro%20punto%20a%20diferentes%20distancias>.

*Data visualization with ggplot2*. (2017). Data Carpentry. Recuperat de <https://erdavenport.github.io/R-ecology-lesson/05-visualization-ggplot2.html>

*El concepto de IDE*. RedHat. Recuperat de <https://www.redhat.com/es/topics/middleware/what-is-ide>

Gabri. (2018, 16 maig). *¿Cómo funciona el semivariograma en la interpolación?* El blog de franz. Recuperat de <https://acolita.com/como-funciona-semivariograma-interpolacion/>

*geom\_dotplot() with median line*. Amazon AWS. Recuperat de [http://rstudio-pubs-static.s3.amazonaws.com/1406\\_947a49f2d7914dad8b0fd050a9df9858.html](http://rstudio-pubs-static.s3.amazonaws.com/1406_947a49f2d7914dad8b0fd050a9df9858.html)

*ggplot time series plotting: group by dates*. (2015, 6 octubre). Stack Overflow. Recuperat de <https://stackoverflow.com/questions/32976820/ggplot-time-series-plotting-group-by-dates>

*ggplot2*. Ggplot2 - Tidyverse. Recuperat de <https://ggplot2.tidyverse.org/>

Jgrubalcaba (2016, 23 febrer). *Cosas que conviene saber al usar AIC, DIC y otros criterios de información*. Modelización y Estadística en Biología. Recuperat de <https://jgrubalcaba.wordpress.com/2016/02/21/cosas-que-conviene-saber-al-usar-aic-dic-y-otros-criterios-de-informacion/>

Lípid. En *Enciclopèdia.cat El cercador de referència en català*. Recuperat de <https://www.enciclopedia.cat/ec-gec-0119486.xml>

*Modelos estadísticos. Consultorías y asesorías | TodoEstadística*. (2018, 13 novembre). Todo Estadística. Recuperat de <https://todoestadistica.com/modelos-estadisticos/>

*Modelos estadísticos*. IBM. Recuperat de <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=nodes-statistical-models>

*Modelos lineales mixtos*. IBM. Recuperat de [https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=SSLVMB\\_25.0.0/spss/advanced/idh\\_mixl.html](https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=SSLVMB_25.0.0/spss/advanced/idh_mixl.html)

Mohamad. (2016, 19 octubre). *Apéndice B: Criterio de Información Akaike (AIC)*. Centro de ayuda. Recuperat de <https://support.numxl.com/hc/es/articles/215531083-Ap%C3%A9ndice-B-Criterio-de-Informaci%C3%B3n-Akaike-AIC->

mpalanco. (2014, 27 octubre). *Gráfico Q-Q en R*. Nube de datos. Recuperat de <https://nubededatos.blogspot.com/2014/10/grafico-q-q-en-r.html>

OBESIDAD. ULPGC. Recuperat de <http://www2.ulpgc.es/hege/almacen/download/6/6655/Obesidad.pdf>

*R Packages: A Beginner's Guide*. (2019). DataCamp. Recuperat de [https://www.datacamp.com/community/tutorials/r-packages-guide?utm\\_source=adwords\\_ppc&utm\\_campaignid=12492439676&utm\\_adgroupid=122563405841&utm\\_device=c&utm\\_keyword=load%20library%20r&utm\\_matchtype=b&utm\\_network=g&utm\\_adposition=&utm\\_creative=504158801983&utm\\_targetid=aud-392016246653:kwd-385495694086&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=9049190&gclid=CjwKCAjwqvyFBhB7EiwAER786WiXBGnkS-csbuhwmUa-SArY3IH05nCNHYVG4dDp3y2vwZ-tAdJ67hoC\\_cYQAvD\\_BwE](https://www.datacamp.com/community/tutorials/r-packages-guide?utm_source=adwords_ppc&utm_campaignid=12492439676&utm_adgroupid=122563405841&utm_device=c&utm_keyword=load%20library%20r&utm_matchtype=b&utm_network=g&utm_adposition=&utm_creative=504158801983&utm_targetid=aud-392016246653:kwd-385495694086&utm_loc_interest_ms=&utm_loc_physical_ms=9049190&gclid=CjwKCAjwqvyFBhB7EiwAER786WiXBGnkS-csbuhwmUa-SArY3IH05nCNHYVG4dDp3y2vwZ-tAdJ67hoC_cYQAvD_BwE)

*Riesgos de la obesidad para la salud*. MedlinePlus. Recuperat de <https://medlineplus.gov/spanish/ency/patientinstructions/000348.htm>

San Segundo, F. (2013, 20 juliol). *QQ-plots y funciones de distribución empírica en R*. PostData. Recuperat de <https://fernandosansegundo.wordpress.com/2013/07/20/qq-plots-y-funciones-de-distribucion-empirica-en-r/>

Serra, A. (2020, 18 maig). *Los tres tipos de dieta; hipocalórica, normocalórica e hipercalórica*. La Bolsa del Corredor. Recuperat de <https://www.sport.es/labolsadelcorredor/tipos-de-dieta-hipocalorica-normocalorica-e-hipercalorica/>

*stat\_smooth* function. RDocumentation. Recuperat de [https://www.rdocumentation.org/packages/ggplot2/versions/0.9.0/topics/stat\\_smooth](https://www.rdocumentation.org/packages/ggplot2/versions/0.9.0/topics/stat_smooth)

Statistics Guides with Dr Paul Christiansen. (2020, 30 juliol). *R Tutorial: Linear mixed-effects models part 1- Repeated measures ANOVA* [Vídeo]. YouTube. Recuperat de <https://www.youtube.com/watch?v=AWInLxpiZuA&t=7s>

Statistics Guides with Dr Paul Christiansen. (2020b, octubre 7). *R Tutorial: Linear mixed-effects models part 2- Repeated measures ANOVA with multiple IVs* [Vídeo]. YouTube. Recuperat de <https://www.youtube.com/watch?v=YsD8b5KYdMw>

Statistics Guides with Dr Paul Christiansen. (2020c, octubre 8). *R Tutorial: Linear mixed-effects models part 3- Mixed ANOVA* [Vídeo]. YouTube. Recuperat de [https://www.youtube.com/watch?v=xOy6ddw-q\\_A](https://www.youtube.com/watch?v=xOy6ddw-q_A)

Torres-Saavedra, P. A. (2018, 21 mayo). *Modelos Estadísticos Avanzados*. GitHub. Recuperat de <https://pegasus.uprm.edu/%7Epedro.torres/book/modelos-lineales-mixtos-mlm.html>

Trujillo, D. (2017). *MODELOS MIXTOS: ANÁLISIS DE MEDIDAS REPETIDAS – PROC MIXED*. [10.13140/RG.2.2.22394.75202](https://doi.org/10.13140/RG.2.2.22394.75202)

*Unix y Linux*. Curso Unix. Recuperat de [https://bioinf.comav.upv.es/courses/unix/unix\\_intro.html#unix](https://bioinf.comav.upv.es/courses/unix/unix_intro.html#unix)

Wikipedia contributors. (2020, 26 desembre). *Geoestadística*. Viquipèdia, l'enciclopèdia lliure. Recuperat de [https://ca.wikipedia.org/wiki/Geoestad%C3%ADstica#Semivariograma,\\_variograma](https://ca.wikipedia.org/wiki/Geoestad%C3%ADstica#Semivariograma,_variograma)

## 6. ANNEXOS

### 6.1. Codi d'R

#### 6.1.1. Descriptiva

```
dades_long<-reshape(dades, varying = list(colnames(dades[3:33])), v.names
= "Pes", timevar = "Dies", idvar = c("ID_animal","Grup"), direction =
"long", times = c(
23,29,35,42,49,52,60,67,73,79,87,93,99,106,114,120,127,134,141,148,157,162
,169,176,185,193,199,206,214,220,227))
dades_long$ID_animal<-as.factor(dades_long$ID_animal)

dieta_C<- subset(dades, dades$Grup=="C")
dddd_tici0 <- dieta_C[c(1,3:33)]
d_melted_tici0_C <- melt(dddd_tici0, id.vars=c("ID_animal"))

ggplot(d_melted_tici0_C, aes(x= variable, y = value,
group=factor(ID_animal), colour = ID_animal)) +
  geom_line() +
  geom_point() +
  facet_wrap(~ ID_animal) +
  theme_bw() +
  theme(legend.position="none") +
  scale_x_discrete(breaks=c("23", "79", "148", "227")) +
  #theme(axis.text.x = element_text(angle = 45, hjust = 1))
  theme (text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",
hjust = 0)) +
  theme(strip.background = element_rect(fill="#d6d6d6"))

taula1<- data.frame(min(d_melted_tici0_C$value), mean(dieta_C$`23`),
(mean(dieta_C$`114`)+ mean(dieta_C$`120`))/2, mean(dieta_C$`227`),
max(d_melted_tici0_C$value))
colnames(taula1)<- c("Mínim", "Primer dia", "Meitat de l'estudi", "Últim
dia", "Màxim")

kbl(taula1, booktabs = T, row.names = FALSE, align =
c("l","c","c","c","c")) %>%
kable_styling(font_size = 10, latex_options = "hold_position", position =
"center") %>%
row_spec(0, bold = T, color = "black", background = "#d6d6d6") %>%
column_spec(3, width = "3.75cm") %>%
column_spec(c(1:2,4:5), width = "2cm")

dieta_M<- subset(dades, dades$Grup=="M")
dddd_tici0 <- dieta_M[c(1,3:33)]
d_melted_tici0_M <- melt(dddd_tici0, id.vars=c("ID_animal"))
```

```

ggplot(d_melted_tici0_M, aes(variable, value, group=factor(ID_animal),
colour = ID_animal)) +
  geom_line() +
  geom_point() +
  facet_wrap(~ ID_animal) +
  theme_bw() +
  theme(legend.position="none") +
  scale_x_discrete(breaks=c("23", "79", "148", "227")) +
  #theme(axis.text.x = element_text(angle = 45, hjust = 1))
  theme (text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",
hjust = 0)) +
  theme(strip.background = element_rect(fill="#d6d6d6"))

taula2<- data.frame(min(d_melted_tici0_M$value), mean(dieta_M$`23`),
(mean(dieta_M$`114`)+ mean(dieta_M$`120`))/2, mean(dieta_M$`227`),
max(d_melted_tici0_M$value))
colnames(taula2)<- c("Mínim", "Primer dia", "Meitat de l'estudi", "Últim
dia", "Màxim")

kbl(taula2, booktabs = T, row.names = FALSE, align =
c("l","c","c","c","c")) %>%
kable_styling(font_size = 10, latex_options = "hold_position", position =
"center") %>%
row_spec(0, bold = T, color = "black", background = "#d6d6d6") %>%
column_spec(3, width = "3.7cm") %>%
column_spec(c(1:2,4:5), width = "2.75cm")

dieta_0<- subset(dades, dales$Grup=="0")
dddd_tici0 <- dieta_0[c(1,3:33)]
d_melted_tici0_0 <- melt(dddd_tici0, id.vars=c("ID_animal"))

ggplot(d_melted_tici0_0, aes(variable, value, group=factor(ID_animal),
colour = ID_animal)) +
  geom_line() +
  geom_point() +
  facet_wrap(~ ID_animal) +
  theme_bw() +
  theme(legend.position="none") +
  scale_x_discrete(breaks=c("23", "79", "148", "227")) +
  #theme(axis.text.x = element_text(angle = 45, hjust = 1))
  theme (text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",

```



```

hjust = 0)) +
  theme(strip.background = element_rect(fill="#d6d6d6"))

taula3<- data.frame(min(d_melted_tici0_0$value), mean(dieta_0$`23`),
  (mean(dieta_0$`114`) + mean(dieta_0$`120`))/2, mean(dieta_0$`227`),
  max(d_melted_tici0_0$value))
colnames(taula3)<- c("Mínim", "Primer dia", "Meitat de l'estudi", "Últim
dia", "Màxim")

kbl(taula3, booktabs = T, row.names = FALSE, align =
c("l","c","c","c","c")) %>%
kable_styling(font_size = 10, latex_options = "hold_position", position =
"center") %>%
row_spec(0, bold = T, color = "black", background = "#d6d6d6") %>%
column_spec(3, width = "4.05cm") %>%
column_spec(c(1:2,4:5), width = "2.5cm")

d_melted_tici0_TOTAL <- melt(dades, id.vars=c("ID_animal","Grup"))
d_melted_tici0_TOTAL$Grup <- revalue(d_melted_tici0_TOTAL$Grup,
c("C"="Dieta Control","M"="Dieta hiperlipídica de blat de moro","O" =
"Dieta hiperlipídica d'oli d'oliva"))

ggplot(d_melted_tici0_TOTAL, aes(variable, value, group=factor(Grup),
colour = ID_animal)) +
  geom_line() +
  geom_point() +
  facet_wrap(~ Grup) +
  theme_bw() +
  theme(legend.position="none") +
  scale_x_discrete(breaks=c("23","79","148","227")) +
  theme(text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",
hjust = 0)) +
  theme(strip.background = element_rect(fill="#d6d6d6"))

ggplot(d_melted_tici0_TOTAL, aes(variable, value, group=factor(Grup),
colour = Grup)) +
  stat_summary(fun.y=mean,geom="line",lwd=2,aes(group=factor(Grup)))+
  scale_colour_manual(values =
c('midnightblue','deepskyblue4','deepskyblue1')) +
  theme_bw() +
  scale_x_discrete(breaks=c("23","79","148","227")) +
  theme(text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",
hjust = 0))

```



### 6.1.2. Anàlisi dels possible models

```

model1<-lme(Pes~Dies*Grup,random=~1|ID_animal,dades_long)
#summary(model1)
anova(model1)

par(mfrow=c(1,2))
plot(fitted(model1),residuals(model1),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model1), col="deepskyblue4")
qqline(resid(model1), col="deepskyblue4")

plot(Variogram(model1))

model2<-lme(Pes~factor(Dies)*Grup,random=~1|ID_animal,dades_long)
#summary(model2)
anova(model2)

par(mfrow=c(1,2))
plot(fitted(model2),residuals(model2),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model2), col="deepskyblue4")
qqline(resid(model2), col="deepskyblue4")

plot(Variogram(model2))

model3<-lme(Pes~Dies*Grup,random=~1|ID_animal,correlation=corCAR1()
,dades_long)
#summary(model3)
anova(model3)

par(mfrow=c(1,2))
plot(fitted(model3),residuals(model3),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model3), col="deepskyblue4")
qqline(resid(model3), col="deepskyblue4")

plot(Variogram(model3))

model4<-lme(Pes~factor(Dies)*Grup,random=~1|ID_animal,correlation=corAR1()
,dades_long)
#summary(model4)
anova(model4)

par(mfrow=c(1,2))
plot(fitted(model4),residuals(model4),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model4), col="deepskyblue4")
qqline(resid(model4), col="deepskyblue4")

plot(Variogram(model4))

```

```

model_anova<-anova(model1,model2,model3,model4)
model_anova$Model

model_anova$AIC

model_anova$BIC

model_anova$logLik

model3_2<-lme(log(Pes)~Dies*Grup,random=~1|ID_animal,correlation=corCAR1(),
,dades_long)
#summary(model3_2)
anova(model3_2)

par(mfrow=c(1,2))
plot(fitted(model3_2),residuals(model3_2),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model3_2), col="deepskyblue4")
qqline(resid(model3_2), col="deepskyblue4")

plot(Variogram(model3_2))

model4_2<-
lme(log(Pes)~factor(Dies)*Grup,random=~1|ID_animal,correlation=corAR1(),
,dades_long)
#summary(model4_2)
anova(model4_2)

par(mfrow=c(1,2))
plot(fitted(model4_2),residuals(model4_2),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model4_2), col="deepskyblue4")
qqline(resid(model4_2), col="deepskyblue4")

plot(Variogram(model4_2))

AIC(model3_2); AIC(model4_2)

BIC(model3_2); BIC(model4_2)

logLik(model3_2); logLik(model4_2)

model5<-
lme(Pes~log(Dies)*Grup,random=~1|ID_animal,correlation=corCompSymm(),dades
_long)
anova(model5)

#summary(model5)

par(mfrow=c(1,2))
plot(fitted(model5),residuals(model5),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model5), col="deepskyblue4")
qqline(resid(model5), col="deepskyblue4")

```

```

plot(Variogram(model5))

model6<-
lme(Pes~log(Dies)*Grup,random=~1|ID_animal,correlation=corCAR1(form=~log(D
ies)|ID_animal), dades_long)
anova(model6)

par(mfrow=c(1,2))
plot(fitted(model6),residuals(model6),xlab="Valors
predits",ylab="Residus", col="deepskyblue4")
abline(h=0)
qqnorm(resid(model6), col="deepskyblue4")
qqline(resid(model6), col="deepskyblue4")

plot(Variogram(model6))

AIC(model5); AIC(model6)

BIC(model5); BIC(model6)

logLik(model5); logLik(model6)

```

### 6.1.3. Resultats amb el model triat

```

summary(model6)

model6.emt2<-emtrends(model6,~Grup,var="log(Dies)")
cld(model6.emt)

pairs = pairs(model6.emt2)
pairs

dades_long$Grup <- revalue(dades_long$Grup, c("C"="Dieta
Control","M"="Dieta hiperlipídica de blat de moro","O" = "Dieta
hiperlipídica d'oli d'oliva"))
model6_2<-
lme(Pes~log(Dies)*Grup,random=~1|ID_animal,correlation=corCAR1(form=~log(D
ies)|ID_animal), dades_long)

ggplot(data=dades_long, aes(x=Dies,colour=Grup)) +
  geom_smooth(aes(y=Pes,group=Grup), geom="line", lwd=2) +
  scale_colour_manual(values =
c('midnightblue','deepskyblue4','deepskyblue1')) +
  theme_bw() +
  theme (text = element_text(size= 10)) +
  labs(x = "Dies (edat que té la rata)", y = "Pes") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=1.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(plot.caption = element_text(color = "black", face = "italic",
hjust = 0)) +
  geom_point(aes(x=Dies,y=pred2,colour=Grup))

```

```
model6.emt<-emtrends(model6_2,~Grup,var="log(Dies)")

plot(model6.emt,comparisons = TRUE) +
  theme_bw() +
  labs(y = "Grup", x = "Estimated marginal mean") +
  theme(axis.title.x = element_text(face="bold", vjust=-0.5,
colour="black", size=rel(1), lineheight = 1)) +
  theme(axis.title.y = element_text(face="bold", vjust=2, colour="black",
size=rel(1), lineheight = 1))

#emmip(model6_2,~Grup,var="Log(Dies)",CIs=T) + theme_bw()
```