



FACULTAD DE MATEMÁTICAS

TRABAJO FIN DE GRADO:

**MODELIZACIÓN ESTADÍSTICA DE DATOS
LONGITUDINALES**

Autor: Sebastián León Isorna

Supervisado por:
D. Rafael Pino Mejías

Índice

1. Resumen	2
2. Abstract	3
3. Estructura y Ejemplos	4
3.1. Características y Ventajas	4
3.2. Estructura de los Datos Longitudinales	5
3.3. Formatos y Operaciones	16
4. Modelización de datos longitudinales para variables continuas	18
4.1. Introducción	18
4.2. Modelización	18
4.3. Fuentes de variación en los datos longitudinales	19
4.4. Estructuras de medias y covarianzas	20
4.5. Modelos más habituales para la estructura de la Matriz de Covarianzas	22
5. Análisis de la varianza de Medidas Repetidas	26
5.1. Consideraciones Previas	26
5.2. Estructura del Modelo	26
5.3. Otras Notaciones Alternativas	29
5.4. Contrastes de Hipótesis para cuestiones de interés	31
5.5. Estructura Matricial para Contrastes de Hipótesis	33
5.6. Análisis de la Varianza	36
5.7. Hipótesis de Simetría Compuesta	38
5.8. Hipótesis y Contrastes Especializados dentro de las unidades	39
5.9. Ajustes cuando no se cumple la Hipótesis de Simetría Compuesta	43
6. Implementación de Ejemplos con R y SPSS	45
6.1. Ejemplo 1	45
6.2. Ejemplo 2	47
6.3. Ejemplo 3	51
6.4. Ejemplo 4	53
7. Referencias	58

1. Resumen

Este Trabajo de Fin de Grado se basa en el estudio de los modelos estadísticos aplicados a los denominados **datos longitudinales**. Estos conjuntos de datos se obtienen mediante la toma de información, repetida durante determinados períodos de tiempo, de una serie de variables que se desean estudiar en un determinado grupo de individuos o unidades (animales, vegetales, empresas, etc). En el ámbito económico se les suele llamar **datos panel**.

Se prestará especial interés en los datos con variable respuesta de tipo continuo, estudiando su modelización según las posibles fuentes de variación que puedan aparecer y aportando las estructuras de medias y covarianzas obtenidas. Finalmente se detallarán una serie de contrastes de hipótesis para cuestiones de interés referentes a esta tipo de datos.

Se tendrá como soporte práctico una serie de ejemplos , expuestos en el capítulo introductorio, que emplean este tipo de datos. Se les aplicará las técnicas que se estudiarán utilizando principalmente el software **R** y el software **IBM SPSS Statistics** como herramienta de apoyo.

2. Abstract

This degree final's study is based on researching about statistical models applied to **longitudinal data**. These data sets are created taking information over the time about a series of variables of interests, inside a group of individual units (animals, vegetables, business, etc). Normally, longitudinal data are called **panel data** if the main objective of the research is inside an economical ambit.

The study is focused on continous variable, whose modeling will depend on all the different souces of variation. Thus will be detailed mean and covariance structure and all the hypotheses and tests of interests. Softwares **R** and **IBM SPSS Statistics** are essential tools to develop all the theoretical content over differents practical examples which will be exposed.

3. Estructura y Ejemplos

3.1. Características y Ventajas

El estudio de datos longitudinales es útil tanto para la comprensión de la evolución anterior como para obtener predicciones del comportamiento futuro sobre las variables en estudio. Los primeros indicios del uso de los datos panel datan a finales de los años treinta y comienzo de los cuarenta, introducidos por Paul Felix Lazarsfeld (Lazarsfeld and Fiske, 1938; Lazarsfeld, 1940) con el objetivo de estudiar la opinión pública. Fue tal el potencial encontrado en este tipo de datos que comenzaron a crearse organizaciones y sociedades dedicadas al estudio de los datos longitudinales. Son de gran utilidad en campos tales como la política (modelar el comportamiento de los votantes en las elecciones), la agricultura (modelar los niveles de producción) o la medicina (modelar el efecto de determinados medicamentos). Una formalización de este tipo de estudio sería:

- Se considera la variable X a estudiar.
- Se consideran n posibles unidades donde se medirá la variable (individuos, poblaciones o terrenos).
- Se consideran p periodos de tiempo durante los que se mide la variable a estudio.

Así, se observa el carácter multivariante de la variable a estudiar y se trabaja con la consiguiente matriz de datos:

$$\mathbf{X} = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \vdots \\ X_{n1} & \cdots & X_{np} \end{pmatrix}$$

Para estos datos no solo interesan las cuestiones de interés habituales, como la evolución de la respuesta media, sino que también se pueden explicar otros puntos: cambios en la respuesta media según tratamientos, cambios en el tiempo y la relación entre tiempos y tratamientos.

Si bien el término **longitudinal** sugiere que los datos se han recogido a lo largo del tiempo, los modelos y métodos desarrollados en este campo son aplicables en otras situaciones. Por ejemplo:

- En un grupo de pacientes, cada uno recibe un medicamento contra la hipertensión con diferentes dosis, y se mide la tensión sanguínea. Aquí cada paciente es medido de forma repetida para diferentes dosis.

- Para un conjunto de árboles, se mide el diámetro de cada uno en diferentes puntos del tronco, por lo que las medidas se toman de forma repetida a lo largo de distintas posiciones.

Por tanto, aunque se use el término de datos longitudinales, estas técnicas son aplicables en general a datos resultantes de un diseño de medidas repetidas.

3.2. Estructura de los Datos Longitudinales

A continuación se aporta una serie de ejemplos para comprender mejor y visualizar este tipo de datos utilizando el software **R** (IBM Corp, 2013):

Ejemplo 1: se estudia la inflamabilidad (medida en m^3/ha) de 16 parcelas del Parque Natural de Doñana recogida durante 6 meses consecutivos. Los incendios forestales pueden ser el mayor problema para unos de los parajes más ricos respecto a biodiversidad y ecosistemas naturales de nuestro país, pero ¿Es Doñana igualmente frágil frente a los incendios durante todo el año? Se estudiará este ejemplo para obtener las respuestas.

Este estudio ayudará a prevenir los grandes incendios producidos en Doñana o tomar medidas para minimizar los efectos de los mismos.

Se presentan ahora los datos a estudiar junto a su representación gráfica:

```
> datosInfla<- read.table("Inflamabilidad.txt",header=T);datosInfla
      X0      X1      X2      X3      X4      X5
1  4276.2  5360.6  5895.5  5279.9  5989.8  7000.3
2  13782.9 15381.7 16143.7 17064.8 17896.3 18319.2
3  14176.4 15898.1 18582.9 20350.8 21020.2 22272.2
4   3273.5  4355.6  4894.2  4283.6  4989.9  5999.8
5   6416.4  6974.7  8649.1  8621.9  8691.2  9851.5
6  14784.0 16375.2 17141.4 18062.8 18893.2 19319.4
7   5593.8  5748.8  6134.2  7195.1  7655.2  8135.5
8  13927.2 14657.3 19177.4 19456.2 20518.4 22912.1
9   20719.6 25571.4 26132.7 26178.8 29880.8 31903.4
10  19718.9 24575.5 25128.3 25174.6 28879.9 30908.5
11   5415.7  5975.8  7642.7  7618.2  7691.2  8847.6
12  12927.5 13653.8 18184.2 18462.4 19520.7 21908.3
13  13177.1 14902.1 17580.6 19357.6 20018.6 21268.7
14  14538.6 19688.6 20341.4 24366.3 26324.7 27164.8
15   4594.5  4749.4  5136.7  6193.3  6659.8  7137.0
16  15539.5 20683.1 21340.7 25367.5 27329.5 28162.9
> matplot(t(datosInfla),type="l",xlab="Mes",ylab="Inflamb.",main="Inflamabilidad Doñana")
> tiempo1<- 1:ncol(datosInfla)
> lines(tiempo1)
> grid()
```

Inflamabilidad Doñana

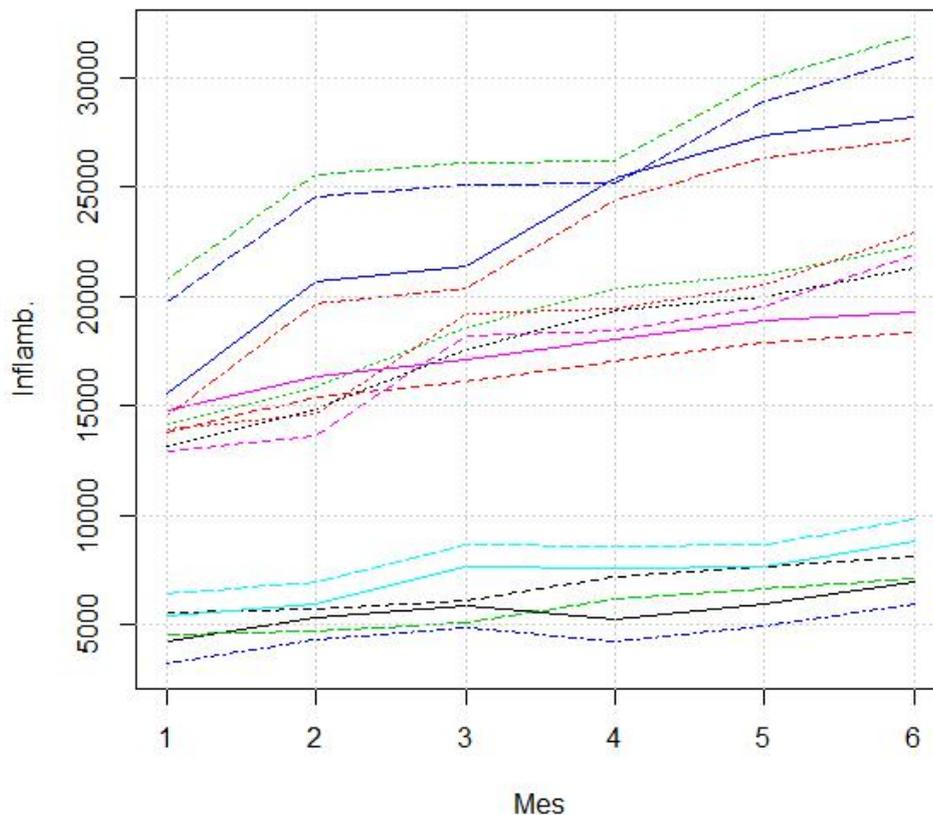


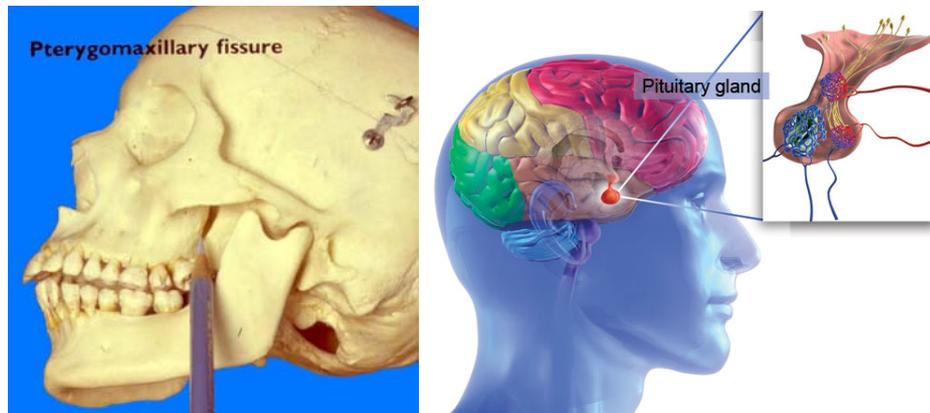
Figura 1: Inflamabilidad

En la Figura 1 podemos ver el desarrollo del nivel de inflamabilidad a lo largo de los 6 meses. La tendencia general muestra un aumento de la inflamabilidad durante estos 6 meses. Con la medición referente al primer mes se pueden distinguir tres grupos:

- 1 **Terrenos de alta inflamabilidad:** son los terrenos 9º y 10º.
- 2 **Terrenos de baja inflamabilidad:** son los terrenos 1º,4º,5º,7º,11º y 15º.
- 3 **Terrenos de inflamabilidad media:** son los restantes terrenos en estudio.

Cabe destacar el comportamiento de los terrenos 16º y 14º, que con el paso del tiempo, pasan de ser terrenos de inflamabilidad media a ser altamente inflamables (puede deberse al aumento de fuentes externas que aumenten la probabilidad de un incendio).

Ejemplo 2: estudio odontológico realizado en 22 niños (11 chicos y 11 chicas). Se midió la distancia en milímetros desde la glándula pituitaria hasta la fisura pterigomaxilar a las edades de 8,10,12 y 14 años (The orthodontic study data of Potthoff and Roy, 1964). Se presentan dos imágenes para entender el experimento:



Será necesario instalar y cargar el paquete "mice" (Stef van Buuren, Karin Groothuis-Oudshoorn, 2011) para la representación en R:

```
> install.packages("mice")
> library(mice)
```

Se lee ahora la correspondiente base de datos que se encontraba en la librería cargada y se prosigue con la representación gráfica.

```
> potthoffroy<-potthoffroy[1:22,]
> potthoffroy
  id sex  d8 d10 d12 d14
1  1  F 21.0 20.0 21.5 23.0
2  2  F 21.0 21.5 24.0 25.5
3  3  F 20.5 24.0 24.5 26.0
4  4  F 23.5 24.5 25.0 26.5
5  5  F 21.5 23.0 22.5 23.5
6  6  F 20.0 21.0 21.0 22.5
7  7  F 21.5 22.5 23.0 25.0
8  8  F 23.0 23.0 23.5 24.0
9  9  F 20.0 21.0 22.0 21.5
10 10  F 16.5 19.0 19.0 19.5
11 11  F 24.5 25.0 28.0 28.0
12 12  M 26.0 25.0 29.0 31.0
13 13  M 21.5 22.5 23.0 26.5
14 14  M 23.0 22.5 24.0 27.5
15 15  M 25.5 27.5 26.5 27.0
16 16  M 20.0 23.5 22.5 26.0
17 17  M 24.5 25.5 27.0 28.5
18 18  M 22.0 22.0 24.5 26.5
19 19  M 24.0 21.5 24.5 25.5
20 20  M 23.0 20.5 31.0 26.0
21 21  M 27.5 28.0 31.0 31.5
22 22  M 23.0 23.0 23.5 25.0
```

```

> tiempo2<-c(8,10,12,14)
> plot(tiempo2,potthoffroy[1,-1:-2],type="l",xlim=c(8,14),ylim=c(15,35),
+ xlab="Años",ylab="Distancia media en m.m.",lty=(1:2)[potthoffroy$sex[1]],
+ main="Estudio Odontológico de Potthoff y Roy (1964)",
+ col=c("green","blue")[potthoffroy$sex[1]])
> for (i in 2:nrow(potthoffroy))
+   lines(tiempo2,potthoffroy[i,-1:-2],lty=(1:2)[potthoffroy$sex[i]],
+         col=c("green","blue")[potthoffroy$sex[i]])
> legend("topleft",col=c("green","blue"),lty=1:2,lwd=2,
+ legend=c("Chicas","Chicos"))
> grid()

```

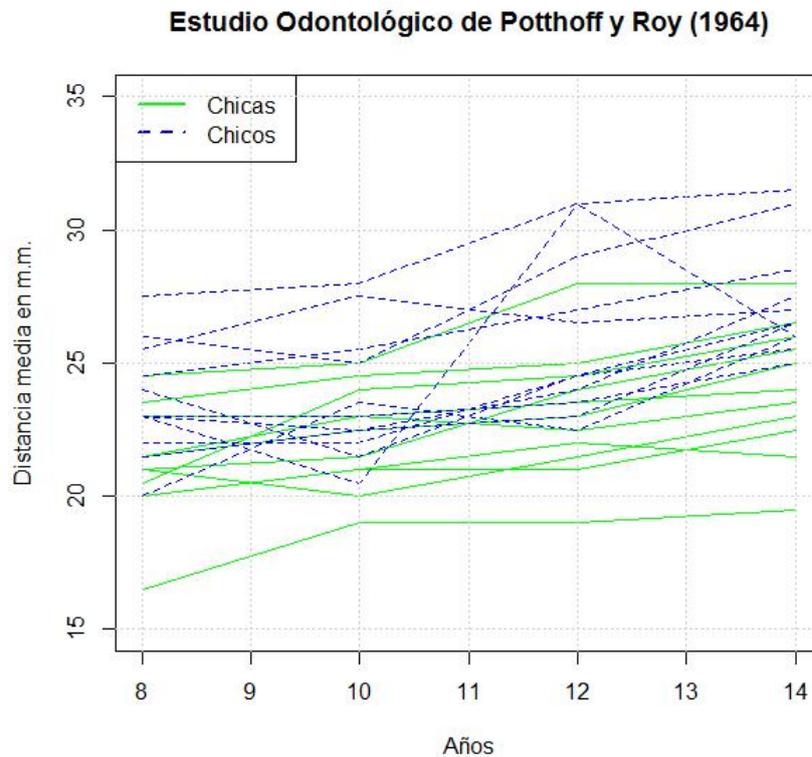


Figura 2: Odontológico

Los objetivos principales del estudio eran los siguientes:

1. Determinar cuál es la mayor distancia a lo largo del tiempo, la de los chicos o la de las chicas.
2. Determinar si la tasa de cambio a lo largo del tiempo es similar en chicos y chicas.

A la vista de la gráfica, se destacan los siguientes resultados respecto al desarrollo de las distancias a lo largo de los 6 años:

- 1 Todas las trayectorias de distancias en función de la edad son similares, pero varían en sus características específicas entre los niños. Destacar dos casos no usuales: un chico cuyo patrón de distancias es ligeramente superior a los de los otros niños y una chica que presenta distancias mucho menores que el resto.
- 2 La tendencia general es el aumento de la distancia con el paso de los años (aunque con varias trayectorias distintas: estrictos crecimientos, pequeños decrecimientos, etc).
- 3 Las distancias en chicos parecen ser mayores en mayor parte que en las chicas (aunque no es cierto uniformemente).
- 4 La razón de cambio a lo largo de las edades es similar en ambos grupos de chicos y chicas.

Ejemplo 3: estudio para determinar el efecto de una dieta de vitamina E sobre el crecimiento de conejillos de indias (Crowder and Hand, 1990). Se tomaron 15 conejillos y se les proporcionó una sustancia inhibidora del crecimiento al comienzo de la primera semana, se midieron sus pesos (en gramos) al final de las semanas 1^a, 3^a y 4^a, y al comienzo de la semana 5^a. Los conejillos fueron asignados al azar a los 3 grupos, 5 a cada uno.

A partir de aquí comienza la terapia de vitamina E: un grupo no consumió ninguna dosis de vitamina, otro consumió una baja dosis y el restante consumió una gran dosis de la misma, definiendo así tres grupos o tratamientos. Finalmente, se midieron a finales de las semanas 5^a, 6^a y 7^a sus respectivos pesos. Se muestra ahora la base de datos y se representan gráficamente los distintos pesos según el grupo al que pertenezcan. Se utiliza el fichero de texto **GuineanPigs** donde se encuentran recogidos los datos:

```
> datapigs<-read.table("GuineanPigs.txt",header=T)
> datapigs
  ID Grupo P1*Sem P3*Sem P4*Sem P5*Sem P6*Sem P7*Sem
1   1     1   455   460   510   504   436   466
2   2     1   467   565   610   596   542   587
3   3     1   445   530   580   597   582   619
4   4     1   485   542   594   583   611   612
5   5     1   480   500   550   528   562   576
6   6     2   514   560   565   524   552   597
7   7     2   440   480   536   484   567   569
8   8     2   495   570   569   585   576   677
9   9     2   520   590   610   637   671   702
10  10    2   503   555   591   605   649   675
11  11    3   496   560   622   622   632   670
12  12    3   498   540   589   557   568   609
13  13    3   478   510   568   555   576   605
14  14    3   545   565   580   601   633   649
15  15    3   472   498   540   524   532   583
> tiempo3<-c(1,3,4,5,6,7)
> plot(tiempo3,datapigs[1,-1:-2],type="l",xlim=c(0,8),ylim=c(400,750),
+ xlab="Semanas",ylab="Peso",col="blue",
+ main="Desarrollo pesos en el primer grupo")
> for (i in 2:5)
+   lines(tiempo3,datapigs[i,-1:-2],type="l",col="blue")
> grid()
```

Desarrollo pesos en el primer grupo

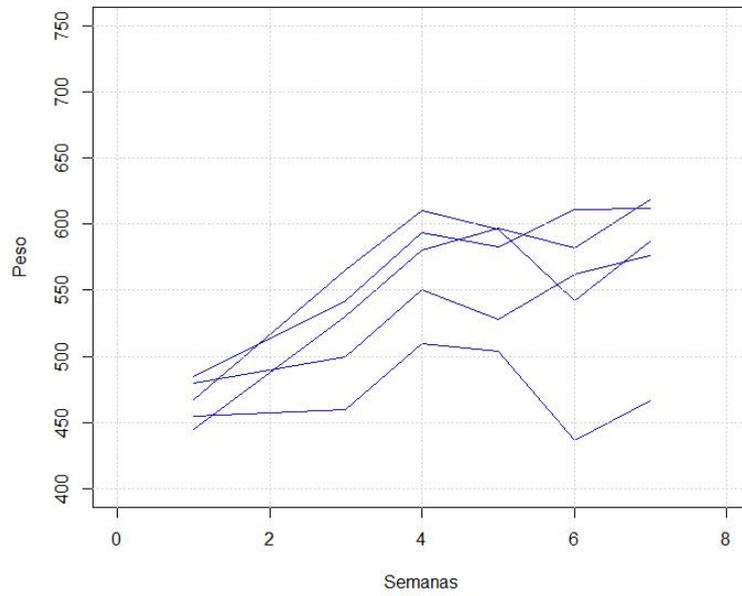


Figura 3: Conejillos 1

```
> plot(tiempo3, datapigs[6, -1:-2], type="l", xlim=c(0, 8), ylim=c(400, 750),  
+ xlab="Semanas", ylab="Peso", col="green",  
+ main="Desarrollo pesos en el segundo grupo")  
> for (i in 7:10)  
+   lines(tiempo3, datapigs[i, -1:-2], type="l", col="green")  
> grid()
```

Desarrollo pesos en el segundo grupo

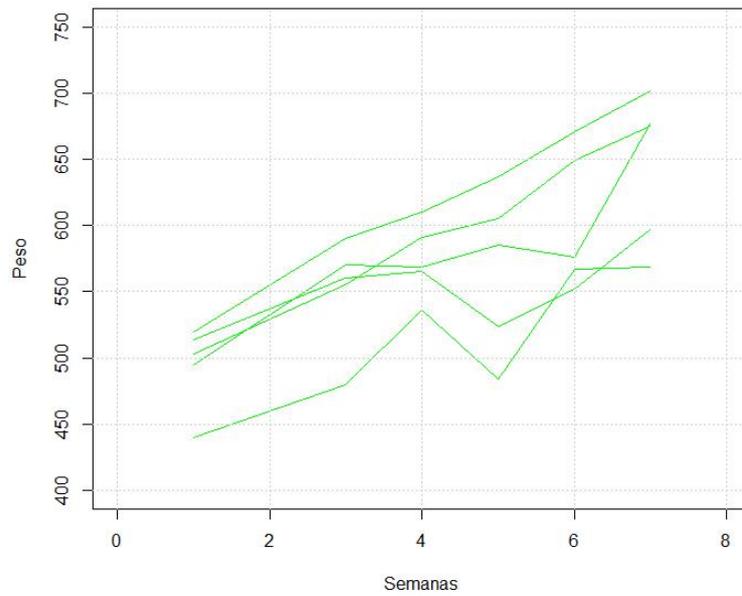


Figura 4: Conejillos 2

```

> plot(tiempo3,datapigs[11,-1:-2],type="l",xlim=c(0,8),ylim=c(400,750),
+ xlab="Semanas",ylab="Peso",col="red",
+ main="Desarrollo pesos en el tercer grupo")
> for (i in 12:15)
+   lines(tiempo3,datapigs[i,-1:-2],type="l",col="red")
> grid()

```

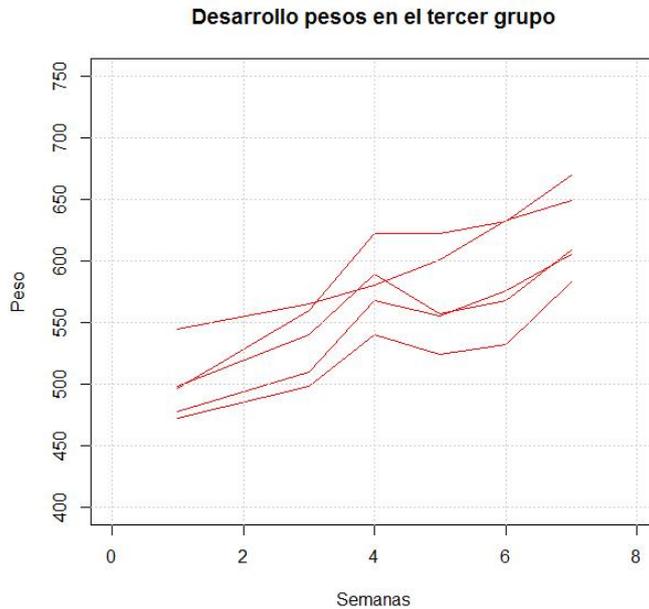


Figura 5: Conejillos 3

A continuación se representan los perfiles medios de cada grupo:

```

> datapigs1<-datapigs[1:5,-1:-2]
> datapigs2<-datapigs[6:10,-1:-2]
> datapigs3<-datapigs[11:15,-1:-2]
> mediapigs1<-apply(datapigs1,2,mean)
> mediapigs2<-apply(datapigs2,2,mean)
> mediapigs3<-apply(datapigs3,2,mean)
> plot(tiempo3,mediapigs1,lwd=2,col="blue",type="l",
+ ,xlim=c(0,8),ylim=c(400,750),
+ xlab="Semanas",ylab="Pesos Medios"
+ ,main="Perfiles de los tres grupos")
> lines(tiempo3,mediapigs2,lwd=2,col="green")
> lines(tiempo3,mediapigs3,lwd=2,col="red")
> legend("topleft",col=c("blue","green","red"),lty=1,lwd=2,
+ legend=c("Grupo 1","Grupo 2","Grupo 3"))
> grid()

```

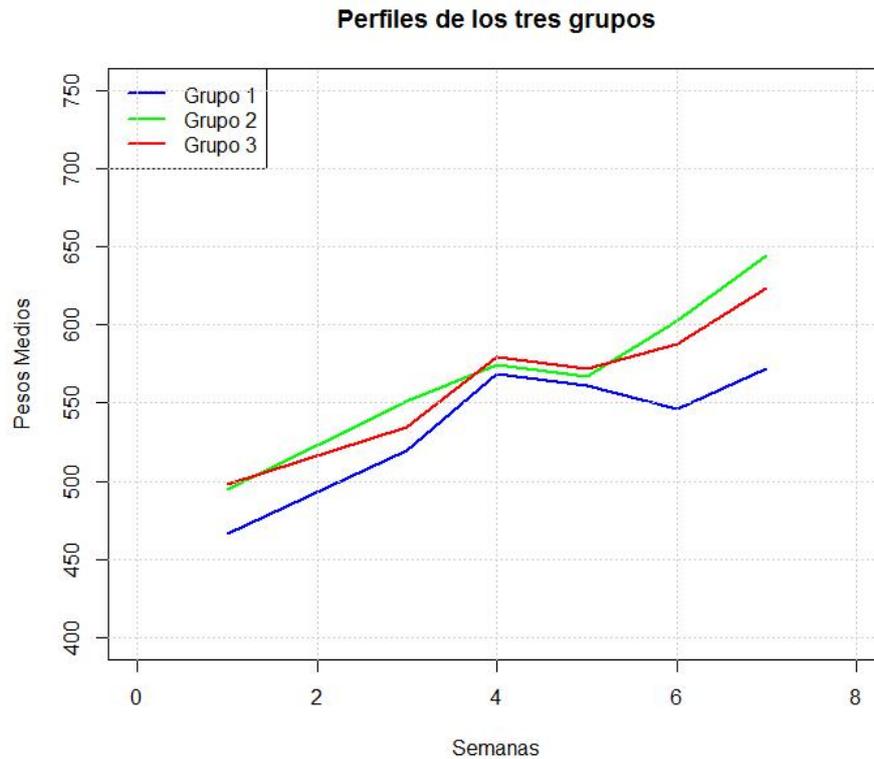


Figura 6: Perfiles Medios

El principal objetivo del estudio era determinar si los patrones de crecimiento diferían entre los tres grupos. A la vista de las distintas gráficas, se destacan los siguientes resultados respecto al desarrollo del peso en los distintos grupos a lo largo de las 7 semanas:

- 1 La mayor parte de las trayectorias individuales de cada uno de los conejillos de indias parecen aumentar globalmente durante el período de estudio (exceptuando un caso perteneciente al grupo que no consumió vitamina E). Dentro de un mismo grupo, los diferentes conejillos tienen trayectorias diferentes. En varios casos, a partir de la cuarta medición (finales de la 5ª semana de estudio, comienzo de tratamiento de vitamina E), comienza a observarse un aumento de peso.
- 2 Los conejillos del grupo de dosis nula presentan unos desarrollos de pesos menores que en los dos restantes grupos.
- 3 No está claro si la tasa de cambio en el peso corporal en promedio es similar o diferente a través de los grupos de dosis. De hecho, no está claro si el patrón seguido (individualmente o en promedio) es una línea recta. Por tanto, la tasa de cambio no puede ser constante. Esto es debido al comenzar la terapia de vitamina E a mediados del estudio.

Ejemplo 4: estudio sobre el crecimiento de dos genotipos de soja (Colleen Hudak, aparece en Davidian and Giltinan 1995,pág 7). Se compararon los patrones de crecimiento de dos genotipos de soja (una variedad comercial que se denotará como genotipo **F** y otra variedad de tipo experimental denominada **P**). Se recogieron los datos entre los años 1988 y 1990, plantando cada año 8 parcelas con el primer genotipo y otras 8 con el restante.

Durante el transcurso de la temporada de siembra, se tomaron muestras de cada parcela cada semana (comenzando por lo general a partir de la segunda semana de siembra), seleccionando 6 plantas al azar de cada parcela. Se tomaron las hojas de dichas plantas, se mezclaron entre ellas y se pesaron, obteniéndose así un peso promedio de por hoja de cada planta (pasará a denominarse **g**, medido en kilogramos). Ahora se representan los pesos promedios de cada genotipo durante el años 1989. Serán necesarias las siguientes librerías y comandos descritos a continuación (Douglas Bates, Martin Maechler and Ben Bolker, 2014):

```
> install.packages("MEMSS")
> library(MEMSS)
> data(Soybean)
```

Debido a la larga extensión de los datos, se mostrarán las tres primeras medidas :

```
> Soybean[1:3,]
  Plot Variety Year Time weight
1 1988F1      F 1988  14  0.106
2 1988F1      F 1988  21  0.261
3 1988F1      F 1988  28  0.666
```

Se muestran las mediciones respectivas a las 3 primeras mediciones del año 1988 correspondientes al genotipo comercial (en total se realizaron 10 para este genotipo) referentes a los días 14, 21, 28, 35,42 ,49 ,56 ,63 ,70 y 77 después de la plantación, junto a los respectivos pesos promedios de las hojas de las plantas. Se representan ahora los pesos promedios de las 16 parcelas medidos durante el año 1989 (dónde se tomaron 8 mediciones anuales), junto a los perfiles medios de ambos genotipos.

```
> dataSoybean1989<-Soybean[157:284,]
> tiempo4<-c(14,20,27,34,41,55,69,84)
> dataframeSoy2<-data.frame(dataSoybean1989[,-1:-4])
> dataframeSoy<-data.frame(dataSoybean1989[1:8,-1:-4],dataSoybean1989[9:16,-1:-4],
+ dataSoybean1989[17:24,-1:-4],dataSoybean1989[25:32,-1:-4],dataSoybean1989[33:40,-1:-4],
+ dataSoybean1989[41:48,-1:-4],dataSoybean1989[49:56,-1:-4],dataSoybean1989[57:64,-1:-4],
+ dataSoybean1989[65:72,-1:-4],dataSoybean1989[73:80,-1:-4],dataSoybean1989[81:88,-1:-4],
+ dataSoybean1989[89:96,-1:-4],dataSoybean1989[97:104,-1:-4],dataSoybean1989[105:112,-1:-4],
+ dataSoybean1989[113:120,-1:-4],dataSoybean1989[121:128,-1:-4])
>
> #Mediante estos comandos, se seleccionan las mediciones respectivas al año 1989,
> #tomados en 8 instantes temporales en los que se estudia el crecimiento durante
> #ese año.
```

```

> #Seguidamente, se consideran los perfiles medios y se realiza la representación
> #gráfica correspondiente
>
> mediaSoy1<-apply(dataframeSoy[,1:8],1,mean)
> mediaSoy2<-apply(dataframeSoy[,9:16],1,mean)
> plot(tiempo4,dataframeSoy[,1],type="l",xlim=c(0,85),ylim=c(0,25),
+ xlab="Días tras la siembra",ylab="Pesos promedio",lty=1,
+ main="Evolución pesos hojas para ambos genotipos, 1989",col="blue")
> lines(tiempo4,mediaSoy1,lwd=3,col="blue")
> lines(tiempo4,mediaSoy2,lwd=3,col="green")
> for (i in 2:8)
+   lines(tiempo4,dataframeSoy[,i],lty=1,col="blue")
> for (i in 9:16)
+   lines(tiempo4,dataframeSoy[,i],lty=1,col="green")
> legend("topleft",col=c("green","blue","green","blue"),lty=1,lwd=c(1,1,3,3),
+ legend=c("Gen.Experimental","Gen.Comercial","P.Medio F","P.Medio P"))
> grid()

```

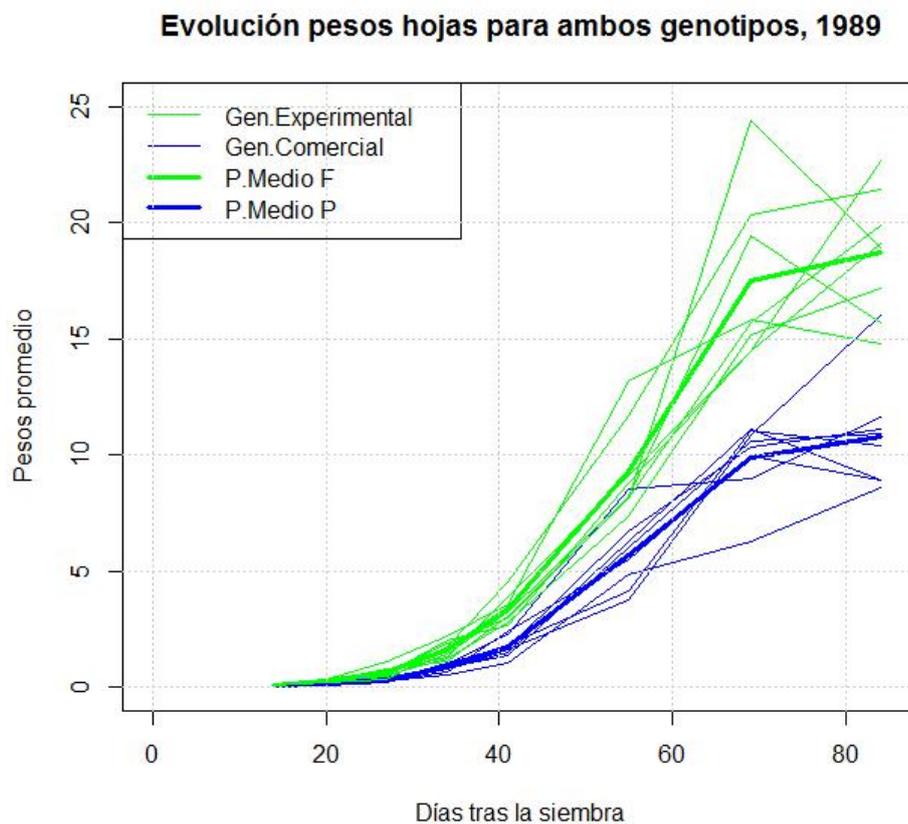


Figura 7: Genotipos

Las líneas verdes representan al genotipo experimental y las líneas azules al genotipo comercial. Con un grosor mayor de color se representan los perfiles medios. Se obtienen a partir de la representación las siguientes conclusiones:

- ① En general, el aumento de pesos en las hojas es pequeño considerando las primeras 3-4 mediciones, en cambio a partir de la 4ª medición, el aumento del peso promedio comienza a hacerse más notable.
- ② Conforme se acerca el final de la cosecha, se observa que el peso comienza a estabilizarse (como es lógico, ya que la planta de soja no crece indefinidamente).
- ③ En general, las trayectorias de crecimiento no siguen una línea recta, como en los 2 ejemplos anteriores, con una aparente tasa de cambio constante a lo largo de los períodos de medida. Por tanto, para caracterizar las diferencias de crecimiento, será necesario algo más además de la simple comparación de la tasa de cambio a lo largo de la temporada.

De hecho, los investigadores se dieron cuenta que el patrón de crecimiento no sería algo tan simple como una línea recta. Sabían que el peso promedio tendía a estabilizarse al llegar el fin de la temporada, por lo tanto se realizó una declaración más precisa de los principales objetivos:

1. Comparar los pesos promedios de ambos genotipos al llegar al final de la temporada.
2. Comparar los notables aumentos del peso producidos a mitad de la temporada.
3. Comparar el aparente peso promedio al inicio de la temporada.

Si se observa la figura anterior, se ve claramente como el peso promedio del genotipo experimental es mayor que el peso referente al genotipo comercial (aproximadamente en unos 8 gramos), mientras que al comienzo de temporada ambos pesos promedios son bastante parecidos.

Añadir además que las condiciones climáticas de los respectivos años del período de estudio fueron considerablemente diferentes (el año 1988 se consideró excepcionalmente seco, 1989 fue excepcionalmente húmedo y 1990 fue relativamente normal). Por lo tanto, la comparación de los patrones de crecimiento a través de los diferentes patrones climáticos, así como la forma en que los patrones climáticos afectaron la comparación de las características de crecimiento entre los genotipos, fue también de interés.

Se han expuesto ejemplos dónde la variable de interés es de tipo **continuo**, existen muchos otros dónde esta variable es de tipo discreto, aunque no se estudiarán durante este trabajo.

3.3. Formatos y Operaciones

Antes de explicar las posibles operaciones con datos de tipo longitudinal, expon-dremos un ejemplo de las dos diferentes formas de organizar los datos panel:

Formato Amplio: es el formato tradicional con el que se organizan los datos longitudinales. Cada unidad estadística ocupa una fila en la matriz de datos y todas las medidas de las posibles variables de estudio a lo largo del tiempo se sitúan en la fila de la unidad correspondiente.

Las bases de datos de los ejemplos 1,2 y 3 siguen este formato.

Formato Largo: es el formato moderno para organizar los datos longitudinales. Los datos se organizan de manera que tengamos asignadas a cada una de las unidades estadísticas tantas filas como mediciones en el tiempo se hayan realizado (en el ejemplo 4, se presenta este formato).

Se ejecuta ahora los comandos en R para obtener nuestra matriz de datos anterior en este nuevo formato (debido a la larga extensión de los mismos se mostrarán sólo los datos referentes a las dos primeras parcelas) :

```
> l<-reshape(datosInfla,varying=c("X0","X1","X2","X3","X4","X5"),
+ v.names="Inflam.",timevar="Meses",times=c("0","1","2","3","4","5"),
+ new.row.names=1:96,direction="long")
> l.sort<- l[order(l$id),]
> l.sort[1:12,]
  Meses Inflam. id
1      0  4276.2  1
17     1  5360.6  1
33     2  5895.5  1
49     3  5279.9  1
65     4  5989.8  1
81     5  7000.3  1
2      0 13782.9  2
18     1 15381.7  2
34     2 16143.7  2
50     3 17064.8  2
66     4 17896.3  2
82     5 18319.2  2
```

Para finalizar la introducción se exponen las dos operaciones básicas que nos permiten operar con datos longitudinales:

Fusionar Datos: consiste en crear un nuevo conjunto de datos a partir de dos que al menos tiene una variable en común. Se realiza ahora un ejemplo con R:

```
> datos1<-data.frame(Pais=c("España","Francia","Italia"),
+ Año=c(2013,2013,2013),
+ PIB=c(1022988,2059852,1560024),
+ Porc.Parados=c(25.6,10.2,12.6));datos1
  País Año    PIB Porc.Parados
1 España 2013 1022988         25.6
2 Francia 2013 2059852         10.2
3 Italia 2013 1560024         12.6
```

(a) Datos 1

```
> datos2<-data.frame(Pais=c("España","Francia","Italia"),
+ Año=c(2013,2013,2013),
+ Deuda=c(996181,1925292,2069216),
+ IPC=c(0.3,0.7,0.7));datos2
  País Año  Deuda IPC
1 España 2013  996181 0.3
2 Francia 2013 1925292 0.7
3 Italia  2013 2069216 0.7
```

(b) Datos 2

Se realiza ahora la mezcla de datos mediante la instrucción:

```
> datos3<-merge(datos1,datos2,by=c("País","Año"));datos3
  País Año  PIB Porc.Parados  Deuda IPC
1 España 2013 1022988      25.6  996181 0.3
2 Francia 2013 2059852      10.2 1925292 0.7
3 Italia  2013 1560024      12.6 2069216 0.7
```

Añadir Datos: consiste en crear un nuevo conjunto de datos a partir de dos que poseen exactamente las mismas variables pero ambos realizados en instantes de tiempo distintos. Se da como ejemplo:

```
> datos4<-data.frame(Pais=c("España","Francia","Italia"),
+ Año=c(2012,2012,2012),
+ PIB=c(1029002,2032297,1566912),
+ Deuda=c(884731,1841027,1989473));datos4
  País Año  PIB  Deuda
1 España 2012 1029002  884731
2 Francia 2012 2032297 1841027
3 Italia  2012 1566912 1989473
```

(c) Datos 4

```
> datos5<-data.frame(Pais=c("España","Francia","Italia"),
+ Año=c(2013,2013,2013),
+ PIB=c(1022988,2059852,1560024),
+ Deuda=c(996181,1925292,2069216));datos5
  País Año  PIB  Deuda
1 España 2013 1022988  996181
2 Francia 2013 2059852 1925292
3 Italia  2013 1560024 2069216
```

(d) Datos 5

Finalmente se adjuntan los datos:

```
> datos6<-rbind(datos4,datos5);datos6
  País Año  PIB  Deuda
1 España 2012 1029002  884731
2 Francia 2012 2032297 1841027
3 Italia  2012 1566912 1989473
4 España 2013 1022988  996181
5 Francia 2013 2059852 1925292
6 Italia  2013 1560024 2069216
```

4. Modelización de datos longitudinales para variables continuas

4.1. Introducción

Se expondrá a continuación un modelo estadístico básico para datos longitudinales, suponiendo siempre que se tiene una **base de datos balanceada** (todas las unidades en estudio poseen mediciones repetidas a lo largo de n períodos de tiempo y los q grupos de estudio a tratar presentan el mismo número de unidades cada uno). Se expondrán 3 situaciones distintas:

- Una muestra aleatoria de una única población.
- Las unidades provienen de distintas poblaciones.
- Se dispone de una característica adicional de las unidades (edad, peso, etc).

4.2. Modelización

Se supone que la variable de interés Y se ha medido en n instantes de tiempo: t_1, \dots, t_n . Correspondiendo así una variable aleatoria Y_j para cada uno de los instantes de tiempo y generándose por consiguiente el vector aleatorio:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Con media : $\boldsymbol{\mu} = E(\mathbf{Y})$ (vector de las medias $\mu_j = E(Y_j), j = 1, \dots, n$).

Y matriz de varianzas-covarianzas: $\boldsymbol{\Sigma} = \text{var}(\mathbf{Y})$.

Nota: como convenio se utilizará el subíndice i para referirnos a las unidades de estudio, el subíndice j para los distintos instantes de tiempo y el valor m para el número total de unidades en estudio.

Un conjunto de datos longitudinales viene dado por $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ vectores ($n \times 1$) todos cumpliendo las mismas propiedades que el anteriormente descrito. Considerando el conjunto de todas las variables aleatorias, se genera una matriz $[Y_{ij}]$ de tamaño ($m \times n$) (m unidades por n instantes de tiempo).

Proposición 1:

- $\text{cov}(Y_{ij}, Y_{ik}) \neq 0$ para $j \neq k = 1, \dots, n$.
- $\text{cov}(Y_{ij}, Y_{lk}) = 0$ para $j = k$ y $j \neq k$.

Un posible modelo estadístico es el siguiente, donde cada vector aleatorio \mathbf{Y}_i viene expresado como una media global más un vector de errores aleatorios:

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i \quad i = 1, \dots, m \quad (1)$$

Donde

- $\boldsymbol{\mu} = E(\mathbf{Y}_i)$ media global (o perfil medio)
- Los $\boldsymbol{\epsilon}_i$ son las desviaciones de los vectores aleatorios: $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in})'$, donde cada ϵ_{ij} , $j = 1, \dots, n$ representa cómo la variable Y_{ij} se desvía de su media μ_j , cumpliendo $E(\epsilon_{ij}) = 0$ y presentando todos ellos correlación entre sí.
- $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $var(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ y los vectores $\boldsymbol{\epsilon}_i$ son independientes entre sí.

Proposición 2: si la variable respuesta estudiada es continua (lo es siempre en este trabajo), asumiendo que las variables Y_{ij} y las desviaciones ϵ_{ij} siguen distribuciones normales y $\boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, se cumple la siguiente distribución para los vectores \mathbf{Y}_i (todos independientes entre sí)

$$\mathbf{Y}_i \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

En el caso de estudiar un total de q poblaciones, cada una de ellas puede presentar un vector media distinto $\boldsymbol{\mu}_l$, $l = 1, \dots, q$ mientras que la matriz de varianzas-covarianzas puede ser constante $\boldsymbol{\Sigma}_l = \boldsymbol{\Sigma}$, $\forall l$ (bajo hipótesis de homocedasticidad). Los elementos de la diagonal de dicha matriz pueden ser distintos, implicando varianzas distintas en diferentes instantes de tiempo. Se tiene entonces el siguiente modelo para q grupos diferentes:

- $\mathbf{Y}_i = \boldsymbol{\mu}_{l(i)} + \boldsymbol{\epsilon}_i$, $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $var(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ $l(i)$ = grupo del i -ésimo individuo

Este modelo será utilizado en los siguientes apartados.

4.3. Fuentes de variación en los datos longitudinales

Se consideran dos tipos de fuentes de variación:

- Variación entre las unidades de estudio
- Variación dentro de las unidades de estudio

Para representar estos conceptos, será necesario retomar el modelo clásico dado para la i -ésima unidad en estudio: $\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i$. La modelización para el valor de la variable Y para esta unidad en un instante t_j vendrá dada por la siguiente igualdad:

$$Y_{ij} = \mu_j + \epsilon_{ij} = \mu_j + b_{ij} + e_{ij} = \mu_j + b_{ij} + e_{1ij} + e_{2ij}$$

Donde

- b_{ij} es una desviación que representa la variación entre unidades en el instante t_j . Se supone $E(b_{ij}) = 0$.
- e_{1ij} representa la desviación adicional debido a las fluctuaciones sobre la tendencia (representada mediante la recta de regresión) dentro de las unidades. $E(e_{1ij}) = 0$.
- e_{2ij} es la desviación producida debido a las medidas de error (dentro de las unidades). $E(e_{2ij}) = 0$.

La suma de estas tres desviaciones da lugar a la desviación total respecto a la media μ_j .

4.4. Estructuras de medias y covarianzas

Consideramos el modelo (1) para una sola población:

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = 0, \quad \text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$$

A continuación, se exponen una serie de términos conocidos en el ámbito de la inferencia multivariante. La estimación del vector media poblacional viene dada por el vector media muestral:

$$\bullet \bar{\mathbf{Y}} = \begin{pmatrix} \bar{Y}_{.1} \\ \bar{Y}_{.2} \\ \vdots \\ \bar{Y}_{.n} \end{pmatrix} \quad \text{para} \quad \bar{Y}_{.j} = \frac{\sum_{i=1}^m Y_{ij}}{m}.$$

Así, $\bar{\mathbf{Y}}$ es un estimador insesgado para $\boldsymbol{\mu}$: $E(\bar{\mathbf{Y}}) = \boldsymbol{\mu}$.

- Para la estimación de la matriz de varianzas-covarianzas $\boldsymbol{\Sigma}$ obtendremos estimadores para las distintas varianzas y covarianzas:
 1. Los elementos de la diagonal de $\boldsymbol{\Sigma}$ son las varianzas σ_j^2 de las distribuciones Y_j . Se estiman insesgadamente mediante las cuasi-varianzas muestrales:

$$S_j^2 = \frac{\sum_{i=1}^m (Y_{ij} - \bar{Y}_{.j})^2}{m - 1}$$

2. Los elementos que no se encuentran en la diagonal son las covarianzas entre las distintas distribuciones Y_j e Y_k , para $j \neq k$:

$$\sigma_{jk} = E\{(Y_j - \mu_j)(Y_k - \mu_k)\}$$

Se estiman insesgadamente mediante la cuasi-covarianza muestral:

$$S_{jk} = \frac{\sum_{i=1}^m (Y_{ij} - \bar{Y}_{.j})(Y_{ik} - \bar{Y}_{.k})}{m - 1}$$

De esta forma, se estima la matriz de cuasi-covarianzas por la matriz de varianzas-covarianzas muestrales:

$$\hat{\Sigma} = \frac{\sum_{i=1}^m (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})'}{m - 1}$$

- Para estimar la matriz de correlaciones poblacional $\mathbf{\Gamma}$, se aplica el coeficiente de correlación muestral entre las observaciones en los instantes t_j y t_k , que estiman al coeficiente de correlación poblacional:

$$\rho_{jk} = \frac{\sigma_{jk}}{\sigma_j \sigma_k}$$

- Matriz de correlaciones muestrales: estimación de la matriz de correlaciones $\mathbf{\Gamma}$:

$$\hat{\mathbf{\Gamma}} = (\Gamma_{jk}) = \begin{cases} 1 & \text{para } j = k \\ \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}\hat{\sigma}_{kk}}} & \text{para } j \neq k \end{cases}$$

En el caso de q poblaciones distintas, con medias $\boldsymbol{\mu}_l$, $l = 1, \dots, q$ las medias respectivas para cada población. Se toma r_l unidades para cada población ($m = r_1 + \dots + r_q$).

- Cada una de las medias $\boldsymbol{\mu}_l$ se estima mediante la media poblacional $\bar{\mathbf{Y}}_{.l}$, y utilizando la fórmula explicada anteriormente se consigue la estimación en cada población $\hat{\Sigma}_l$ de la matriz poblacional Σ_l .
- Σ se estima de la siguiente forma:

$$\hat{\Sigma} = \begin{cases} \frac{(r_1 - 1)\hat{\Sigma}_1 + \dots + (r_q - 1)\hat{\Sigma}_q}{m - q} & \text{para } r_l \text{ distintos} \\ (1/q)(\hat{\Sigma}_1 + \dots + \hat{\Sigma}_q) & \text{para } r_l \text{ iguales} \end{cases}$$

4.5. Modelos más habituales para la estructura de la Matriz de Covarianzas

Modelo No Estructurado

Utilizado en situaciones en donde no hay evidencia de un aparente patrón sistemático de varianza y correlación. En estos casos, se dice que la matriz de covarianzas sigue un modelo no estructurado: n varianzas distintas (una para cada instante de tiempo), y $n(n-1)/2$ covarianzas *distintas* para cada par de instantes temporales.

Esto conlleva a la estimación de un gran número de parámetros descriptores de variación, más aún cuando se estudian q poblaciones en vez de una (el número de parámetros a estimar sería q veces mayor). Suponer este modelo cuando existe alguna estructura implica un uso ineficiente de los datos disponibles.

Modelos de Simetría Compuesta

Son modelos de covarianza aplicados cuando la correlación entre las observaciones en cualquier instante t_i y t_j son similares, si bien las varianzas en diferentes instantes de tiempo pueden ser distintas.

Suponiendo a ρ como el parámetro que representa la correlación común para cualesquiera dos instantes de tiempo ($-1 < \rho < 1$), se ilustra la forma de la matriz de correlaciones para 5 instantes de tiempo (de aquí en adelante se aplicará siempre los ejemplos para 5 instantes de tiempo, aunque siempre podrá generalizarse para cualquier n) :

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

Esta estructura se suele conocer como Modelo de Simetría Compuesta o de Correlaciones Intercambiables, donde el último término hace hincapié en la igualdad de correlación aún cambiando cuales quiera dos instantes de tiempo por otros dos. Se describe ahora la estructura de la matriz de covarianzas aplicando la fórmula del coeficiente de correlación poblacional (seguimos asumiendo $n = 5$)

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 & \rho\sigma_1\sigma_5 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 & \rho\sigma_2\sigma_5 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & \rho\sigma_3\sigma_5 \\ \rho\sigma_1\sigma_4 & \rho\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ \rho\sigma_1\sigma_5 & \rho\sigma_2\sigma_5 & \rho\sigma_3\sigma_5 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}$$

Esta matriz se suele llamar **Matriz de Covarianzas Heterogénea de Simetría Compuesta**

Nota: en ciertas ocasiones, puede considerarse que la varianza poblacional se mantiene constante durante todos los instantes de tiempo: $\sigma_j^2 = \sigma^2$ para $j = 1, \dots, n$, obteniéndose:

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2 \Gamma$$

Nota: los dos modelos explicados anteriormente no explotan el hecho de que las observaciones están recogidas a lo largo del tiempo, sin tener en cuenta la ordenación temporal de las variables.

Modelo de Varianzas con Autocorrelaciones de Orden 1

Es de esperar, debido a las fluctuaciones dentro de las unidades, que las correlaciones sean “más fuertes” para dos observaciones tomadas en instantes de tiempo cercanos, mientras que en observaciones más alejadas en el tiempo esta correlación vaya disminuyendo.

El modelo que explica las situaciones donde dos observaciones adyacentes o consecutivas en el tiempo mantienen correlación dentro de una determinada unidad y no mantiene ninguna correlación con cualquiera otra observación que no sea consecutiva o anterior a ella, se denomina modelo de covarianza con autocorrelación de orden uno. Se representa ahora la estructura de la matriz de correlaciones para este modelo (se sigue aplicando para 5 instantes de tiempo y ρ sigue siendo la correlación entre dos observaciones entre dos instantes de tiempo distintos, $-1 < \rho < 1$) :

$$\Gamma = \begin{pmatrix} 1 & \rho & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 \\ 0 & \rho & 1 & \rho & 0 \\ 0 & 0 & \rho & 1 & \rho \\ 0 & 0 & 0 & \rho & 1 \end{pmatrix}$$

Esta matriz es comúnmente denominada **matriz de Toeplitz en bandas**. Destacar la mayor eficiencia de este modelo cuando todas las observaciones están separadas por un mismo intervalo temporal. La matriz de varianzas-covarianzas correspondiente es:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & 0 & 0 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & 0 \\ 0 & 0 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ 0 & 0 & 0 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}$$

Al asumirse la igualdad de varianzas poblacionales en todos los instantes de tiempo:

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2\Gamma$$

Nota: esta estructura puede ampliarse a la existencia de correlación entre observaciones que estén separadas por dos o más intervalos temporales.

Modelo de Varianzas con estructura Autoregresiva de orden 1

Este modelo tiene en cuenta que la magnitud de la correlación entre dos observaciones decaen con el tiempo, considerando observaciones igualmente espaciadas en el tiempo para cada una de las unidades de estudio. Notaremos a estos espacios temporales fijos como sigue:

$$|t_j - t_{j+1}| = d, \quad \text{para } j = 1, \dots, n - 1$$

Suponiendo las mismas características sobre ρ y n que en los ejemplos anteriores, la matriz de correlaciones para este modelo seguiría la siguiente estructura:

$$\Gamma = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

De esta forma, se obtienen las siguientes matrices de varianzas-covarianzas para varianzas poblacionales distintas e iguales respectivamente:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 & \rho^4\sigma_1\sigma_5 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 & \rho^3\sigma_2\sigma_5 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & \rho^2\sigma_3\sigma_5 \\ \rho^3\sigma_1\sigma_4 & \rho^2\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ \rho^4\sigma_1\sigma_5 & \rho^3\sigma_2\sigma_5 & \rho^2\sigma_3\sigma_5 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 & \rho^4\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^4\sigma^2 & \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2\Gamma$$

Modelo estructural de Markov

Los modelos de autocorrelación de orden 1 y autoregresivo de orden 1 se centran en observaciones espaciadas por intervalos temporales iguales. En cambio, en el modelo que se explica a continuación los distintos instantes de tiempo t_1, \dots, t_n no están necesariamente separados por la misma longitud de tiempo.

Se tienen entonces longitudes temporales entre cualesquiera instantes t_j y t_k para todo $j, k = 1, \dots, n$

$$d_{jk} = |t_j - t_k|$$

Para este modelo, es necesario $\rho \geq 0$, dado la forma de la matriz de correlaciones para el mismo:

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} & \rho^{d_{15}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} & \rho^{d_{25}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 & \rho^{d_{34}} & \rho^{d_{35}} \\ \rho^{d_{14}} & \rho^{d_{24}} & \rho^{d_{34}} & 1 & \rho^{d_{45}} \\ \rho^{d_{15}} & \rho^{d_{25}} & \rho^{d_{35}} & \rho^{d_{45}} & 1 \end{pmatrix}$$

- No es posible el caso $\rho < 0$ ya que podría variar las correlaciones drásticamente de valores negativos para observaciones cercanas en el tiempo y positivas para observaciones alejadas en el tiempo (lo cual no tiene ningún sentido)

Finalmente se representa la matriz de varianzas-covarianzas asumiendo varianzas poblacionales iguales:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{15}} \\ \sigma^2 \rho^{d_{12}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{25}} \\ \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{23}} & \sigma^2 & \sigma^2 \rho^{d_{34}} & \sigma^2 \rho^{d_{35}} \\ \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{34}} & \sigma^2 & \sigma^2 \rho^{d_{45}} \\ \sigma^2 \rho^{d_{15}} & \sigma^2 \rho^{d_{25}} & \sigma^2 \rho^{d_{35}} & \sigma^2 \rho^{d_{45}} & \sigma^2 \end{pmatrix}$$

Teniendo únicamente como parámetros de este modelo a σ^2 y ρ , para cualquier número de instantes de tiempo n .

5. Análisis de la varianza de Medidas Repetidas

5.1. Consideraciones Previas

En este capítulo se supone un diseño experimental balanceado y además se asumen las siguientes características:

1. Se realiza una suposición muy específica sobre las covarianzas en los vectores de datos, que no siempre se cumple para los datos longitudinales. Esto puede conllevar a inferencias erróneas.
2. Se presenta una representación simplista de la media de un vector de datos. Aunque suele escogerse por defecto debido a su familiaridad con las técnicas de análisis de la varianza.
3. Las unidades son asignadas al azar en $q \geq 1$ grupos de tratamiento.

Fuentes de Variación:

- Existe una variación aleatoria en la población de unidades, por ejemplo, la variación biológica existente en los seres vivos. Se identifica como **variación entre los individuos** (unidades).
- Existe otra variación aleatoria debido a las fluctuaciones dentro de cada unidad y las medidas de error. Es denominada como **variación dentro de los individuos** (unidades).

5.2. Estructura del Modelo

Se da una nueva notación a la utilizada en la modelización del apartado anterior

Y_{hlj} = observación de la h -ésima unidad en el l -ésimo grupo para el instante j -ésimo

- $h = 1, \dots, r_l$ donde r_l denota el número de unidades en el l -ésimo grupo (cumpliendo $m = \sum_{l=1}^q r_l$)
- $l = 1, \dots, q$ los distintos grupos
- $j = 1, \dots, n$ los distintos instantes de tiempo

Obteniéndose el siguiente modelo:

$$Y_{hlj} = \mu + \tau_l + b_{hl} + \gamma_j + (\tau\gamma)_{lj} + e_{hlj} \quad (2)$$

Donde

- μ es la media poblacional.
- τ_l es la desviación sobre la media muestral producida al encontrarse una unidad en el l -ésimo grupo.
- b_{hl} es un efecto aleatorio, con esperanza nula, que representa la desviación producida al medir la variable Y en una h -ésima unidad particular en el l -ésimo grupo. Es una variación producida **entre unidades**, por su distribución mediante una distribución de probabilidad.
- γ_j es la desviación asociada al j -ésimo instante
- $(\tau\gamma)_{lj}$ es el efecto de interacción para el l -ésimo grupo en el j -ésimo instante, que conlleva a una desviación adicional.
- e_{hlj} es una desviación aleatoria, con esperanza nula, correspondiente a la variación producida **dentro de las unidades**. Aunque a veces también se denomina como error aleatorio.

Nota: aunque en el modelo aparecen parámetros asociados con cada instante de tiempo, estos valores de tiempo no aparecen explícitamente en el modelo.

A continuación se presenta una expresión alternativa del modelo, utilizando las siguientes propiedades de los parámetros representados:

1. Al tener b_{hl} y e_{hkj} esperanza nula, se cumple:

$$E(Y_{hlj}) = \mu_{lj} = \mu + \tau_l + \gamma_j + (\tau\gamma)_{lj}$$

Representando ésta la media para el l -ésimo grupo en el j -ésimo instante de tiempo.

2. La suma de los parámetros restantes da como resultado todos los efectos aleatorios de desviación (que denominaremos como ϵ).

Al considerar siempre en este estudio una variable respuesta de tipo continuo, se considera una distribución normal para la forma en que las distintas fuentes de variación afectan a la respuesta. Se asumirá por tanto:

- Los distintos $b_{hj} \sim N(0, \sigma_b^2)$ y son independientes.
- Los distintos $e_{hlj} \sim N(0, \sigma^2)$ y son independientes (aunque debemos tener en cuenta los instantes de tiempo cercanos entre sí, ya que en ese caso podrían aparecer correlaciones significativas).
- Todos los b_{hj} y e_{hlj} se asumen como mutuamente independientes.

Nos centramos ahora en la estructura en forma vectorial. Consideramos los datos para una unidad particular en un determinado grupo:

$$\begin{pmatrix} Y_{hl1} \\ Y_{hl2} \\ \vdots \\ Y_{hln} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} + \tau_l + \gamma_1 + (\tau\gamma)_{l1} \\ \boldsymbol{\mu} + \tau_l + \gamma_2 + (\tau\gamma)_{l2} \\ \vdots \\ \boldsymbol{\mu} + \tau_l + \gamma_n + (\tau\gamma)_{ln} \end{pmatrix} + \begin{pmatrix} b_{hl} \\ b_{hl} \\ \vdots \\ b_{hl} \end{pmatrix} + \begin{pmatrix} e_{hl1} \\ e_{hl2} \\ \vdots \\ e_{hln} \end{pmatrix}$$

$$\mathbf{Y}_{hl} = \boldsymbol{\mu}_l + \mathbf{1}b_l + \mathbf{e}_{hl} \quad (3)$$

Al aplicar las propiedades de los parámetros explicadas anteriormente:

$$\mathbf{Y}_{hl} = \boldsymbol{\mu}_l + \boldsymbol{\epsilon}_{hl} \quad (4)$$

Propiedades:

- $E(\mathbf{Y}_{hl}) = \boldsymbol{\mu}_l$
- $var(Y_{hlj}) = var(b_{hl}) + var(e_{hlj}) + 2cov(b_{hl}, e_{hlj}) = \sigma_b^2 + \sigma_e^2$
- Las componentes Y_{hlj} se distribuyen normalmente (al hacerlo los parámetros b_{hl} y e_{hlj}), y por tanto los vectores \mathbf{Y}_{hl} siguen una distribución normal multivariante de orden n con vector de medias $\boldsymbol{\mu}_l$.
- Para observaciones de unidades distintas y instantes temporales y grupos distintos, se cumple:

$$cov(Y_{hlj}, Y_{h'l'j'}) = 0$$

Como conclusión se obtiene que cualquier par de vectores de datos procedentes de unidades de estudio distintas, independientemente de que se encuentren en el mismo grupo o no, son independientes bajo este modelo.

- Considerando el caso para dos observaciones de una misma unidad:

$$\text{cov}(Y_{hlj}, Y_{hlj'}) = \sigma_b^2$$

Utilizando estas propiedades, se obtiene la siguiente matriz de varianzas-covarianzas para cada \mathbf{Y}_{hl} :

$$\text{var}(\mathbf{Y}_{hl}) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 + \sigma_e^2 \end{pmatrix}$$

O alternativamente como:

$$\boldsymbol{\Sigma} = \sigma_b^2 \mathbf{J}_n + \sigma_e^2 \mathbf{I}_n$$

Con \mathbf{I}_n matriz identidad de orden n y $\mathbf{J}_n = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ de orden n . De hecho, el

modelo explicado es un modelo de simetría compuesta (caracterizado en el capítulo anterior):

1. $\rho = \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_e^2)}$, es una correlación positiva que es la misma para cualesquiera dos elementos muy alejados temporalmente o no para \mathbf{Y}_{hl} .
2. Además, los elementos diagonales de $\boldsymbol{\Sigma}$ son iguales, conllevando a la igualdad de varianzas de cada uno de los elementos de \mathbf{Y}_{hl} .

5.3. Otras Notaciones Alternativas

Se tiene el inconveniente de que el modelo anterior implica una restrictiva y posiblemente poco realista suposición acerca de la correlación entre las observaciones sobre la misma unidad en el tiempo. Una solución es considerar una serie de notaciones alternativas que detallaremos a continuación.

Primera Notación

Nuestro subíndice i ahora indica el valor para una unidad dada determinada únicamente por los valores de h y l ($i = 1, \dots, m$), utilizándose de ahora en adelante los siguientes conceptos:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} \text{ Denota a los vectores } \mathbf{Y}_{hl} \text{ anteriores}$$

Para trabajar con las medias se utiliza la siguiente matriz:

$$M = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{q1} & \mu_{q2} & \cdots & \mu_{qn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}'_1 \\ \vdots \\ \boldsymbol{\mu}'_q \end{pmatrix}$$

Si además se añade la siguiente variable indicadora:

$$a_{il} = \begin{cases} 1 & \text{si } \textit{la unidad "i" pertenece al grupo "l"} \\ 0 & \text{c.c} \end{cases}$$

Y las agrupamos en vectores correspondientes a cada una de las unidades:

$$\mathbf{a}'_i = (a_{i1}, a_{i2}, \dots, a_{iq})$$

Se obtiene una nueva estructura para las variables respuesta, ahora clasificadas por unidades:

$$\mathbf{Y}'_i = \mathbf{a}'_i M + \mathbf{1}' b_i + \mathbf{e}_i \quad (5)$$

O análogamente (por propiedades de las variables de error)

$$\mathbf{Y}'_i = \mathbf{a}'_i M + \boldsymbol{\epsilon}_i \quad (6)$$

Segunda Notación

Otra manera de expresar el modelo es definiendo un vector que llamaremos $\boldsymbol{\beta}$, dónde se colocan todos los parámetros a estudiar según el número de grupos e instantes temporales (se representa un ejemplo para dos grupos y dos instantes de tiempo):

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ \gamma_2 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{12} \\ (\tau\gamma)_{21} \\ (\tau\gamma)_{22} \end{pmatrix}$$

Y considerando los vectores $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$, junto a las matrices de ceros y unos \mathbf{X}_i , se obtiene:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}b_i + \boldsymbol{\epsilon}_i \quad (7)$$

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i \quad (8)$$

Se debe remarcar que para que la matriz \mathbf{X}_i sea de rango completo, para todos los distintos grupos e instantes de tiempo, son necesarias las siguientes suposiciones:

- $\sum_{l=1}^q r_l = 0$
- $\sum_{j=1}^n \gamma_j = 0$
- $\sum_{l=1}^q (\tau\gamma)_{lj} = 0 = \sum_{j=1}^n (\tau\gamma)_{lj} \quad \forall l, j$

5.4. Contrastes de Hipótesis para cuestiones de interés

En este apartado se va a considerar la primera de las notaciones alternativas descritas en el apartado anterior ($\mathbf{Y}'_i = \mathbf{a}'_i M + \boldsymbol{\epsilon}_i$), considerando las siguientes suposiciones:

- $\mathbf{Y}_i \sim N_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$
- $\boldsymbol{\Sigma} = \sigma_b^2 \mathbf{J}_n + \sigma_e^2 \mathbf{I}_n$ (Estructura de covarianza de simetría compuesta)
- $\sum_{l=1}^q \tau_l = 0$
- $\sum_{j=1}^n \gamma_j = 0$
- $\sum_{l=1}^q (\tau\gamma)_{lj} = 0 = \sum_{j=1}^n (\tau\gamma)_{lj} \quad \forall l, j$

Una primera cuestión de interés es la posible interacción entre los grupos y el tiempo, es decir ¿El cambio medio en la respuesta a lo largo del tiempo depende del grupo? Se estudia mediante la siguiente expresión:

$$\mu_{lj} = \mu + \tau_l + \gamma_j + (\tau\gamma)_{lj}$$

En el caso de estudiar dos grupos distintos (denominamos 1 y 2), la diferencia media entre ambos en el momento j es:

$$\mu_{1j} - \mu_{2j} = (\tau_1 - \tau_2) + (\tau\gamma)_{1j} - (\tau\gamma)_{2j}$$

Posibilitando los $(\tau\gamma)_{lj}$ la dependencia del factor tiempo

Ahora nos encontramos en disposición de poder describir el denominado **contraste de paralelismo de perfiles**:

$$\begin{cases} H_0 : & \text{perfiles paralelos} \\ H_1 : & \text{c.c.} \end{cases}$$

La hipótesis nula en este caso, conllevaría a que los términos $(\tau\gamma)_{lj}$ (que son los que producen que la diferencia entre ambos grupos pueda ser distinta en diferentes instantes de tiempo) sean iguales en todos los instantes j .

Si a ello le unimos el supuesto $\sum_{l=1}^q (\tau\gamma)_{lj} = 0 = \sum_{j=1}^n (\tau\gamma)_{lj}$, realmente se está estudiando el contraste:

$$\begin{cases} H_0 : & (\tau\gamma)_{lj} = 0 \quad \forall l, j \\ H_1 : & \text{Al menos un } (\tau\gamma)_{lj} \neq 0 \end{cases}$$

Tras comprobar el paralelismo de perfiles, la cuestión se centra en verificar si los perfiles coinciden. Se muestra este estudio para el caso de estudiar dos grupos ($q = 2$) en tres instantes temporales ($n = 3$).

Considerando el grupo l' , la media aritmética de las medias para los 3 instantes temporales se define como:

$$n^{-1}(\mu_{l'1} + \mu_{l'2} + \mu_{l'3}) = \mu + \tau_l + n^{-1}(\gamma_1 + \gamma_2 + \gamma_3) + n^{-1}\{(\tau\gamma)_{l'1} + (\tau\gamma)_{l'2} + (\tau\gamma)_{l'3}\}$$

Tomándose ahora la diferencia entre estas medias de dos grupos distintos (denominados $l = 1$ y $l = 2$):

$$\tau_1 - \tau_2 + n^{-1} \sum_{j=1}^n (\tau\gamma)_{1j} - n^{-1} \sum_{j=1}^n (\tau\gamma)_{2j}$$

Y al aplicar el supuesto dado al comienzo de la sección $\sum_{j=1}^n (\tau\gamma)_{lj} = 0$, se obtiene el **contraste sobre el efecto para dos grupos (tratamiento)**:

$$\begin{cases} H_0 : & \tau_1 = \tau_2 = 0 \\ H_1 : & \tau_1 \neq 0 \text{ ó } \tau_2 \neq 0 \end{cases}$$

Generalizando para cualquier número de grupos e instantes temporales:

$$\begin{cases} H_0 : & \tau_1 = \dots = \tau_q = 0 \\ H_1 : & \text{Al menos un } \tau_l \neq 0 \end{cases}$$

Por último, se estudia si la respuesta media es constante a lo largo del tiempo. Para ello se trabaja con el promedio de medias respuestas entre los grupos para un instante de tiempo j' :

$$q^{-1} \sum_{l=1}^q \mu_{lj'} = \gamma_{j'} + q^{-1} \sum_{l=1}^q \tau_l + q^{-1} \sum_{l=1}^q (\tau\gamma)_{lj'}$$

De nuevo, al aplicar los supuestos $\sum_{l=1}^q \tau_l = 0$ y $\sum_{l=1}^q (\tau\gamma)_{lj} = 0$, se obtiene el **contraste sobre el efecto del tiempo**:

$$\begin{cases} H_0 : & \gamma_1 = \dots = \gamma_n = 0 \\ H_1 : & \text{Al menos un } \gamma_j \neq 0 \end{cases}$$

5.5. Estructura Matricial para Contrastes de Hipótesis

Todos los contrastes explicados en el apartado anterior pueden expresarse de forma matricial, de manera que ayude a trabajar más directamente con los datos o incluso se faciliten los cálculos a llevar a cabo. Se están asumiendo todas las suposiciones dadas al comienzo de la sección anterior.

Para ello se consideran:

1. La matriz \mathbf{M} , cuyas filas son los vectores media correspondientes a los distintos tratamientos (matriz $(q \times n)$ por tanto).
2. \mathbf{C} una matriz $(c \times q)$ de rango completo, con $c \leq q$ (se recuerda que q representa el número total de grupos).

3. \mathbf{U} una matriz ($n \times u$) de rango completo, con $u \leq n$ (se recuerda que n representa el número total de instantes de tiempo).

Con esto, la hipótesis nula para los distintos contrastes se expresará de la forma:

$$H_0 : \mathbf{CMU} = \mathbf{0}$$

Dependiendo de qué contraste se quiera estudiar y el número de grupos e instantes temporales en estudio, se tomará una determinada estructura de las matrices \mathbf{C} y \mathbf{U} . A continuación se exponen las distintas estructuras posibles considerándose $q = 2$ y $n = 3$.

Contraste de Paralelismo

Se pueden considerar las matrices:

$$\mathbf{C} = (1 \quad -1)$$

$$\mathbf{U} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Obteniéndose así:

$$\begin{aligned} \mathbf{CMU} &= (1 \quad -1) \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \\ &= (\mu_{11} - \mu_{21} - \mu_{12} + \mu_{22}, \quad \mu_{12} - \mu_{22} - \mu_{13} + \mu_{23}) \end{aligned}$$

Aplicando la igualdad $\mu_{lj} = \mu + \tau_l + \gamma_j + (\tau\gamma)_{lj}$ y al operar con los distintos factores que se obtienen al sustituir en el vector obtenido;

$$\mathbf{CMU} = ((\tau\gamma)_{11} - (\tau\gamma)_{21} - (\tau\gamma)_{12} + (\tau\gamma)_{22}, \quad (\tau\gamma)_{12} - (\tau\gamma)_{22} - (\tau\gamma)_{13} + (\tau\gamma)_{23})$$

La hipótesis inicial para este contraste realizada teóricamente (sección anterior), se considera que todos los $(\tau\gamma)_{lj}$ son nulos. Por tanto se obtiene el resultado buscado.

Contraste sobre el efecto entre grupos

Se consideran las matrices:

$$\mathbf{C} = (1 \quad -1)$$

$$\mathbf{U} = \begin{pmatrix} 1/n \\ 1/n \\ 1/n \end{pmatrix}$$

Así:

$$\begin{aligned} \mathbf{CM} &= (1 \quad -1) \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} = (\mu_{11} - \mu_{21}, \quad \mu_{12} - \mu_{22}, \quad \mu_{13} - \mu_{23}) = \\ &(\tau_1 - \tau_2 + (\tau\gamma)_{11} - (\tau\gamma)_{21}, \quad \tau_1 - \tau_2 + (\tau\gamma)_{12} - (\tau\gamma)_{22}, \quad \tau_1 - \tau_2 + (\tau\gamma)_{13} - (\tau\gamma)_{23}) \end{aligned}$$

Aplicando la igualdad de los μ_{ij} .

Si ahora multiplicamos por la matriz \mathbf{U} :

$$\mathbf{CMU} = \tau_1 - \tau_2 + n^{-1} \sum_{j=1}^n (\tau\gamma)_{1j} - n^{-1} \sum_{j=1}^n j = 1^n (\tau\gamma)_{2j}$$

Al anularse por supuestos los dos sumatorios, para que $\mathbf{CMU} = 0$ se cumple la hipótesis nula dada para este contraste en la sección teórica. Para casos generales de q y n , se pueden dar matrices \mathbf{C} y \mathbf{U} que lleven a la misma conclusión sin ningún problema.

Contraste sobre el efecto del tiempo

Se considerar las matrices:

$$\begin{aligned} \mathbf{C} &= (1/q \quad 1/q) \\ \mathbf{U} &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

Así:

$$\begin{aligned} \mathbf{MU} &= \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \mu_{11} - \mu_{12} & \mu_{12} - \mu_{13} \\ \mu_{21} - \mu_{22} & \mu_{22} - \mu_{23} \end{pmatrix} = \\ &\begin{pmatrix} \gamma_1 - \gamma_2 + (\tau\gamma)_{11} - (\tau\gamma)_{12}, & \gamma_2 - \gamma_3 + (\tau\gamma)_{12} - (\tau\gamma)_{13} \\ \gamma_1 - \gamma_2 + (\tau\gamma)_{21} - (\tau\gamma)_{22}, & \gamma_2 - \gamma_3 + (\tau\gamma)_{22} - (\tau\gamma)_{23} \end{pmatrix} \end{aligned}$$

Al aplicarse de nuevo la igualdad de los μ_{ij} .

Realizando el producto con \mathbf{C} (en este caso $q = 2$), y aplicándose los supuestos iniciales:

$$\mathbf{CMU} = (\gamma_1 - \gamma_2, \quad \gamma_2 - \gamma_3)$$

Para que se cumpla $\mathbf{CMU} = 0$, debe cumplirse la igualdad entre los γ_n y coincidiendo por tanto con el resultado teórico obtenido.

5.6. Análisis de la Varianza

Se realiza en esta sección el análisis de la varianza, bajo el supuesto del modelo (2) explicado al comienzo del capítulo:

$$Y_{hlj} = \mu + \tau_l + b_{hl} + \gamma_j + (\tau\gamma)_{lj} + e_{hlj}$$

Además de suponer las hipótesis de normalidad. Se necesitan los siguientes estadísticos para obtener la Tabla de Análisis de la Varianza correspondiente.

Definición 1:

- Promedio de la muestra a lo largo del tiempo para la h -ésima unidad en el l -ésimo grupo: $\bar{Y}_{hl\cdot} = n^{-1} \sum_{j=1}^n Y_{hlj}$
- Promedio de la muestra en el j -ésimo instante de tiempo dentro del h -ésimo grupo para todas las unidades: $\bar{Y}_{\cdot lj} = r_l^{-1} \sum_{h=1}^{r_l} Y_{hlj}$
- Promedio de la muestra para todas las observaciones en el l -ésimo grupo: $\bar{Y}_{\cdot l\cdot} = (r_l n)^{-1} \sum_{h=1}^{r_l} \sum_{j=1}^n Y_{hlj}$
- Promedio de la muestra para todas las observaciones en el j -ésimo instante: $\bar{Y}_{\cdot\cdot j} = (m)^{-1} \sum_{l=1}^q \sum_{h=1}^{r_l} Y_{hlj}$
- Tasa Media de todas las mn observaciones: \bar{Y}_{\dots}
- Suma de cuadrados en los grupos: $SS_G = \sum_{l=1}^q nr_l (\bar{Y}_{\cdot l\cdot} - \bar{Y}_{\dots})^2$
- Suma de cuadrados en el tiempo: $SS_T = \sum_{l=1}^q nrl (\bar{Y}_{\cdot l\cdot} - \bar{Y}_{\dots})^2$
- Suma de cuadrados en los grupos y el tiempo:
 $SS_{GT} = \sum_{j=1}^n \sum_{l=1}^q r_l (\bar{Y}_{\cdot lj} - \bar{Y}_{\dots})^2 - SS_T - SS_G$
- Suma de cuadrados total de las unidades: $SS_{Tot,U} = n \sum_{l=1}^q \sum_{h=1}^{r_l} (\bar{Y}_{hl\cdot} - \bar{Y}_{\dots})^2$
- Suma de cuadrados total: $SS_{Tot,all} = n \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n (\bar{Y}_{hlj} - \bar{Y}_{\dots})^2$

Ahora se está en disposición de representar la tabla ANOVA para el modelo expuesto anteriormente:

F.Variación	SS	DF	SS/DF	F-Ratios
Entre los grupos	SS_G	$q - 1$	MS_G	$F_G = MS_G/MS_{EU}$
Tiempo	SS_T	$n - 1$	MS_T	$F_T = MS_T/MS_E$
Grupo x Tiempo	SS_{GT}	$(q - 1)(n - 1)$	MS_{GT}	$F_{GT} = MS_{GT}/MS_E$
Error entre las unidades	$SS_{Tot,U} - SS_G$	$m - q$	MS_{EU}	
Error dentro de las unidades	SS_E	$(m - q)(n - 1)$	MS_E	
Total	$SS_{Tot,all}$	$nm - 1$		

Cuadro 1: Tabla ANOVA

Notación:

SS: suma de cuadrados

DF: grados de libertad

MS: cuadrados medios

Proposición 3: bajo los supuestos:

$$\mathbf{Y}_i \sim N_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$$
$$\boldsymbol{\Sigma} = \sigma_b^2 \mathbf{J}_n + \sigma_e^2 \mathbf{I}_n$$

los cuadrados medios son (cálculos realizados en sección 3.3 de Crowder and Hand, 1990):

F.Variación	MS	Cuadrados Medios Esperados
Entre los grupos	MS_G	$\sigma_e^2 + n\sigma_b^2 + n \sum_{l=1}^q r_l \tau_l^2 / (q-1)$
Tiempo	MS_T	$\sigma_e^2 + m \sum_{j=1}^n \gamma_j^2 / (n-1)$
Grupo x Tiempo	MS_{GT}	$\sigma_e^2 + \sum_{l=1}^q r_l \sum_{j=1}^n (\tau\gamma)_{lj}^2 / (q-1)(n-1)$
Error entre las unidades	MS_{EU}	$\sigma_e^2 + n\sigma_b^2$
Error dentro de las unidades	MS_E	σ_e^2

Cuadro 2: Cuadrados Medios

Esta estimación permite intuir la distribución seguida por las distintas *F-Ratios* según la fuente de variación que se esté estudiando. Cada una de estas distribuciones viene notoriamente afectada por los resultados obtenidos al realizar los distintos contrastes de hipótesis planteados en la sección 4 de este capítulo:

Test de Paralelismo de Perfiles

$$\begin{cases} H_0 : & (\tau\gamma)_{lj} = 0 \quad \forall l, j \\ H_1 : & \text{Al menos un } (\tau\gamma)_{lj} \neq 0 \end{cases}$$

Se rechaza H_0 al nivel de significación α si:

$$F_{GT} > \mathcal{F}_{(q-1)(n-1), (n-1)(m-q), 1-\alpha}$$

Test de Efecto Principal de Grupos

$$\begin{cases} H_0 : & \tau_1 = \dots = \tau_q = 0 \\ H_1 : & \text{Al menos un } \tau_l \neq 0 \end{cases}$$

Rechazando H_0 al nivel de significación α si:

$$F_G > \mathcal{F}_{q-1, m-q, 1-\alpha}$$

Test de Efecto Principal de Tiempo

$$\begin{cases} H_0 : & \gamma_1 = \dots = \gamma_n = 0 \\ H_1 : & \text{Al menos un } \gamma_j \neq 0 \end{cases}$$

Rechazando H_0 al nivel de significación α si:

$$F_G > \mathcal{F}_{n-1, (n-1)(m-q), 1-\alpha}$$

5.7. Hipótesis de Simetría Compuesta

En todo lo explicado en este capítulo, el supuesto de estudiar una matriz de varianzas-covarianzas Σ con estructura de **simetría compuesta** ha sido estrictamente necesaria para poder desarrollar todo nuestro estudio. En esta sección se comprobará que solo es necesario que la matriz posea una estructura especial, denominada de **tipo H**.

Definición 2: una matriz sigue una estructura de tipo H cuando sigue la siguiente estructura:

$$\Sigma = \begin{pmatrix} \lambda + 2\alpha_1 & \alpha_1 + \alpha_2 & \cdots & \alpha_1 + \alpha_n \\ \alpha_2 + \alpha_1 & \lambda + 2\alpha_2 & \cdots & \alpha_2 + \alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n + \alpha_1 & \alpha_n + \alpha_2 & \cdots & \lambda + 2\alpha_n \end{pmatrix}$$

Proposición 4:

*Los vectores \mathbf{Y}_i siguen distribuciones normales multivariantes y sus matrices Σ de son de tipo **H**.*

↓

Los test dados en el apartado anterior son válidos.

En esta nueva estructura se basa el contraste que se explica a continuación.

Contraste de Esfericidad de Mauchy (1940)

Se estudia la igualdad entre las las varianzas de las diferencias entre todos los posibles pares de grupos. Si se rechaza la hipótesis nula, las F-Ratios obtenidas en el Análisis de la Varianza deben estudiarse con precaución y realizarse una serie de modificaciones

$$\begin{cases} H_0 : \Sigma \text{ es de tipo } H \\ H_1 : \quad \quad \text{c.c} \end{cases}$$

Se utiliza un estadístico que sigue una distribución chi-cuadrado con $(n-2)(n+1)/2$ grados de libertad (Todo basado en los estudios realizados en Vonesh and Chinchilli (1997,p. 85)). Este test presenta una serie de inconvenientes:

- No es efectivo al presentarse un bajo número de unidades en cada grupo.
- Pueden obtenerse resultados erróneos al aplicar el test en el caso de que los vectores de datos no sigan una distribución normal multivariante.

5.8. Hipótesis y Contrastes Especializados dentro de las unidades

En esta sección se estudiará con más detalle una serie de contrastes que ayudan al estudio de efectos dentro de las unidades (principalmente el efecto del factor tiempo).

Definición 3: sea \mathbf{c} vector $(n \times 1)$ y $\boldsymbol{\mu}$ vector de medias $(n \times 1)$, se denomina a la combinación lineal:

$$\mathbf{c}'\boldsymbol{\mu} = \boldsymbol{\mu}'\mathbf{c}$$

como valor contraste si los elementos de \mathbf{c} vienen dados de manera que todos sumen cero.

En la sección 5 de este capítulo, se ha estudiado cómo expresar los distintos Contrastes de Hipótesis a estudiar aplicando la notación matricial **CMU** (considerando q grupos y n instantes de tiempo). Ahora, se reutilizarán estas matrices para estudiar varios tipos de contrastes específicos, cada uno de ellos utilizados según el tipo de estudio que se quiera realizar. Partimos de que se van a realizar principalmente los siguientes tipos de estudio:

Tipo I:

- Estudiar si la forma en que la media difiere entre dos instantes de tiempo es diferente para distintos grupos.
- Estudiar si hay una diferencia en media entre dos instantes de tiempo, promediados entre los grupos.

Cada uno de estos aspectos proporcionará información sobre si las características en estudio cambian con el tiempo y de qué forma lo hacen. Aplicando la notación matricial:

- La matriz \mathbf{M} , cuyas filas son los vectores media correspondientes a los distintos tratamientos (matriz $(q \times n)$ por tanto).
- \mathbf{U} matriz $(n \times n - 1)$, que para este tipo de contraste tendrá la siguiente estructura:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 \\ 0 & \cdots & 0 & -1 \end{pmatrix}$$

Definiéndose el producto MU como **transformación de perfiles de medias dentro de las unidades**.

Tipo II:

Se centra en saber cómo la media en cada periodo de tiempo difiere de lo que ocurre sobre todo en momentos posteriores. Esto puede ayudar a entender en qué instante de tiempo comienza a producirse algún cambio en la tendencia (si este fuese el caso).

- Se sigue considerando la misma matriz \mathbf{M} , con las mismas dimensiones e idéntica definición al contraste anterior.
- La matriz \mathbf{U} , de idénticas dimensiones, pero en este caso de estructura distinta, que definiremos como **matriz de transformación de Helmert**:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1/(n-1) & 1 & 0 & \cdots & 0 \\ -1/(n-1) & -1/(n-2) & 1 & \cdots & 0 \\ \vdots & \vdots & -1/(n-3) & \vdots & \vdots \\ -1/(n-1) & -1/(n-2) & \vdots & \cdots & 1 \\ -1/(n-1) & -1/(n-2) & -1/(n-3) & \cdots & -1 \end{pmatrix}$$

Realizándose así una comparación entre cada media frente a a la media aritmética de las medias en instantes de tiempo posteriores.

Nota: Tomando $\mathbf{C} = (1, 1)$ en el caso concreto de $n = 3$, los vectores contraste conllevan ambos (Tipo I y Tipo II) a la igualdad

$$\gamma_1 = \gamma_2 = \gamma_3$$

Por tanto, se demuestra que la elección de la matriz de contrastes \mathbf{U} no afecta a la hora de verificar los principales efectos tiempo-grupo por la interacción del tiempo (que se estudiarán en los denominados **Contrastes Generales**). Sí es importante y determinante esta elección de la matriz \mathbf{U} cuando se está interesado en estudiar "por trozos" esos efectos generales, llegando a resultados más particulares (estudiados en los **Contrastes Separados**).

Contrastes Generales

Consideraremos dos tipos, según la matriz \mathbf{C} (matriz de dimensiones $(q-1) \times q$) que se tome:

- Si se considera la matriz:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}$$

Entonces, al multiplicar por el producto \mathbf{MU} , se estudia cómo los distintos contrastes definidos en \mathbf{U} difieren entre los grupos. También se observa cómo difiere la media en todos los grupos en instantes distintos, y por tanto es un efecto grupo-tiempo.

- Si en cambio se considera la matriz

$$\mathbf{C} = (1/q, 1/q, \dots, 1/q)$$

Tras multiplicar por \mathbf{MU} , se obtiene una matriz $1 \times (n-1)$, cada una de las columnas corresponden a la **media aritmética** de cada uno de los contraste respecto a los grupos (constituyendo todas ellas un efecto principal de tiempo). Así, se produce una comparación entre la media en un instante j y el resto de períodos temporales distintos a j .

Contrastes Separados

Como su nombre indica, nos centraremos ahora en estudiar los contrastes obtenidos mediante la matriz \mathbf{U} cada uno de ellos por separado. Se aplicará la k -ésima columna de la matriz \mathbf{U} como \mathbf{c}_k , $k = 1, \dots, n-1$, de la siguiente forma:

- Se toman los vectores de datos por cada unidad en estudio (\mathbf{Y}_{hl}) y se realiza el producto

$$\mathbf{c}'_k \mathbf{Y}_{hl} = \mathbf{Y}'_{hl} \mathbf{c}_k$$

Así, el valor obtenido representa el valor contraste para esa unidad.

Propiedades:

- Para estudiar la variación de los valores contraste a lo largo de los distintos grupos, es suficiente con realizar un análisis de la varianza de un factor (igualdad de todas las medias en los grupos como hipótesis nula).
- Para contrastar si este valor es de promedio cero a lo largo de los grupos, se aplica una prueba t de Student a la media general de los valores contraste de los grupos.
- La estructura de simetría compuesta no es estrictamente necesaria para la validez de los contrastes, sino que la matriz de varianzas-covarianzas $\mathbf{\Sigma}$ debe ser constante para todas las unidades. Así:

$$var(\mathbf{c}'_k \mathbf{Y}_{hl}) = \mathbf{c}'_k \mathbf{\Sigma} \mathbf{c}_k$$

Según la matriz \mathbf{U} que se utilice para el contraste, puede cumplirse

$$\mathbf{c}'_i \mathbf{c}_j = 0$$

para cualquier par de columnas de \mathbf{U} . Entonces, diremos que se está realizando un **contraste ortogonal**. Aunque para trabajar con estos contrastes, será necesaria una modificación de los mismos, debido a que los distintos valores $\mathbf{c}' \mathbf{Y}_{hl}$ pueden presentar escalas distintas.

Este hecho se produce al observar que la suma de cuadrados de los elementos de las distintas columnas es distinta. Para acabar con este inconveniente, se multiplicarán los elementos de cada una de las columnas por un determinado valor (igual o distinto para cada una de ellas), que conlleve a que la suma de cuadrados de elementos sea igual a 1 para todas las columnas (normalización). Al tener todos los valores contraste en una misma escala, se dice que se ha realizado un **contraste ortonormal**.

Nota: aunque la propiedad de ser ortonormal es una propiedad deseable, no es estrictamente necesaria, ya que aunque los contrastes tengan escalas distintas, las F-Ratios obtenidas en las distintas tablas ANOVA son las mismas.

En cambio, si el contraste realizado no es ortogonal, la interpretación de los contrastes por separado sería más complicada, ya que las salidas de cada uno de ellos por separado estarían relacionadas.

Como vimos en el primer capítulo, una característica común de los datos longitudinales es que cada unidad en estudio parece mostrar una trayectoria de tiempo. De hecho, a la vista de todos los ejemplos, es razonable pensar en el comportamiento de dicha trayectoria como una función polinómica. Por tanto, en las distintas situaciones que nos podamos encontrar, sería ventajoso poder resolver las cuestiones siguientes respecto a la respuesta media en el tiempo (promedio a través de y entre los grupos):

- ¿Qué tipo de tendencia nos encontramos (línea recta, cuadrática, cúbica, etc)?
- ¿La tendencia difiere entre los grupos?

Hay un tipo particular de contraste que se centra en este tema, cuyos coeficientes se denominan **coeficientes polinomiales ortogonales**. Para estudios realizados en n instantes temporales, es posible encontrarse hasta un comportamiento polinomial de grado $n - 1$ para la trayectoria a estudiar. Un estudio más detallado de este tipo de contraste se saldría de nuestro trabajo, por tanto continuamos con la siguiente sección.

5.9. Ajustes cuando no se cumple la Hipótesis de Simetría Compuesta

Se usa este tipo de contraste cuando no se tiene seguridad absoluta de la estructura de tipo **H** de la matriz Σ , y por tanto no presenta la estructura de simetría compuesta que estamos considerando desde el comienzo del capítulo.

Esto conlleva a que las pruebas F de Snedecor usuales para tiempo y grupo-tiempo no sean válidas. Se explican a continuación una serie de ajustes para hacer válidas las pruebas.

Proposición 5: sea U una matriz ($n \times n - 1$) con la estructura explicada en la sección 5, Σ la matriz de varianzas-covarianzas y

$$\epsilon = \frac{tr^2(U'\Sigma U)}{(n - 1)tr(U'\Sigma U U'\Sigma U)}$$

Entonces:

$$\epsilon = 1 \Leftrightarrow \Sigma \text{ es de tipo } H$$

Este valor ϵ se utilizará para disminuir los grados de libertad en el numerador y denominador de las distintas F Ratios (pasando por ejemplo de a y b grados de libertad respectivamente, a ser ϵa y ϵb).

Aunque se tiene un inconveniente, y es que ϵ no es conocido, ya que depende de la matriz Σ y ésta se desconoce. Como solución, se han realizado estimaciones de Σ , se comentan dos de los ajustes realizados y comúnmente utilizados:

- Ajuste de Greenhouse-Geisser (1959)
- Ajuste de Huynh-Feldt (1976): ajuste menos severo que el anterior

Ambos ajustes se realizarán con los softwares R y SPSS en el siguiente capítulo si fuesen necesarios.

6. Implementación de Ejemplos con R y SPSS

En este último capítulo el objetivo principal es aplicar un análisis de la varianza de medidas repetidas en los distintos ejemplos dados en el Capítulo 1. Para ello, se aplicarán los procedimientos presentados al final del Capítulo 5.

Se estudiarán todos los ejemplos a fondo utilizando R, mientras que en **IBM SPSS Statistics**(IBM Corp, 2013) se mostrará la hoja de comandos en los tres primeros y se estudiará el procedimiento detalladamente en el último (ejemplo más completo).

Será necesario instalar y cargar el paquete "ez" (Michael A. Lawrence , 2013) para la implementación en R:

```
> install.packages("ez")
> library("ez")
```

La principal función a aplicar será el comando "ezANOVA", junto a un nivel de significación $\alpha = 0,05$. Será necesario transformar algunas bases de datos del conjunto de ejemplos al formato largo, ya que es el utilizado para este comando.

6.1. Ejemplo 1

Se muestra la lista de variables y el conjunto de datos en SPSS:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Infla0	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada
2	Infla1	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada
3	Infla2	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada
4	Infla3	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada
5	Infla4	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada
6	Infla5	Numérico	5	1		Ninguna	Ninguna	8	☰ Derecha	📏 Escala	📄 Entrada

	Infla0	Infla1	Infla2	Infla3	Infla4	Infla5
1	4276,2	5360,6	5895,5	5279,9	5989,8	7000,3
2	13782,9	15381,7	16143,7	17064,8	17896,3	18319,2
3	14176,4	15898,1	18582,9	20350,8	21020,2	22272,2
4	3273,5	4355,6	4894,2	4283,6	4989,9	5999,8
5	6416,4	6974,7	8649,1	8621,9	8691,2	9851,5
6	14784,0	16375,2	17141,4	18062,8	18893,2	19319,4
7	5593,8	5748,8	6134,2	7195,1	7655,2	8135,5
8	13927,2	14657,3	19177,4	19456,2	20518,4	22912,1
9	20719,6	25571,4	26132,7	26178,8	29880,8	31903,4
10	19718,9	24575,5	25128,3	25174,6	28879,9	30908,5
11	5415,7	5975,8	7642,7	7618,2	7691,2	8847,6
12	12927,5	13653,8	18184,2	18462,4	19520,7	21908,3
13	13177,1	14902,1	17580,6	19357,6	20018,6	21268,7
14	14538,6	19688,6	20341,4	24366,3	26324,7	27164,8
15	4594,5	4749,4	5136,7	6193,3	6659,8	7137,0
16	15539,5	20683,1	21340,7	25367,5	27329,5	28162,9

Y el consiguiente archivo de comandos SPSS, obteniéndose los siguientes resultados:

```

1
2 DATASET ACTIVATE Conjunto_de_datos1.
3 GLM Infla0 Infla1 Infla2 Infla3 Infla4 Infla5
4 /WSFACTOR=Tiempo 6 Polynomial
5 /METHOD=SSTYPE(3)
6 /PRINT=HOMOGENEITY
7 /CRITERIA=ALPHA(.05)
8 /WSDESIGN=Tiempo.
9

```

Prueba de esfericidad de Mauchly^b

Medida:MEASURE_1

Efecto intra-sujetos	W de Mauchly	Chi-cuadrado aprox.	gl	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Límite-inferior
Tiempo	,000	128,012	14	,000	,284	,305	,200

Contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad.

a. Puede usarse para corregir los grados de libertad en las pruebas de significación promediadas. Las pruebas corregidas se muestran en la tabla Pruebas de los efectos inter-sujetos.

b. Diseño: Intersección
Diseño intra-sujetos: Tiempo

Pruebas de efectos intra-sujetos.

Medida:MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Tiempo	Esfericidad asumida	4,811E8	5	96213727,35	39,692	,000
	Greenhouse-Geisser	4,811E8	1,418	3,393E8	39,692	,000
	Huynh-Feldt	4,811E8	1,523	3,159E8	39,692	,000
	Límite-inferior	4,811E8	1,000	4,811E8	39,692	,000
Error(Tiempo)	Esfericidad asumida	1,818E8	75	2423994,749		
	Greenhouse-Geisser	1,818E8	21,265	8549389,322		
	Huynh-Feldt	1,818E8	22,841	7959368,117		
	Límite-inferior	1,818E8	15,000	12119973,74		

La instrucción aplicada en R y los consiguientes resultados:

```

> #En l.sort se encuentran los datos de este ejemplo en el formato adecuado
> #para aplicar este comando
> ezANOVA(l.sort,dv=Inflam.,wid=id,within=Meses)
Aviso: Converting "id" to factor for ANOVA.
Aviso: Converting "Meses" to factor for ANOVA.
$ANOVA
  Effect DFn DFd      F      p p<.05      ges
2  Meses   5   75 39.69222 9.716526e-20 * 0.0817125

$`Mauchly's Test for Sphericity`
  Effect      W      p p<.05
2  Meses 5.703043e-05 5.254123e-20 *

$`Sphericity Corrections`
  Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF] p[HF]<.05
2  Meses 0.2835284 4.478947e-07 * 0.3045461 1.883316e-07 *

```

Ante todo, destacar la coincidencia esperada entre los resultados obtenidos por ambos softwares. Ambos nos llevan a las siguientes conclusiones:

Por el Contraste de Esfericidad de Mauchly, hay suficiente evidencia estadística para rechazar la hipótesis nula sobre la estructura de tipo H de Σ .

Se realizan los correspondientes ajustes de Greenhouse-Geisser y Huynh-Feldt, y ambos producen estadísticos significativos.

Esto conlleva a rechazar la hipótesis nula sobre la igualdad de $\epsilon = 1$ (para ϵ definido en la Proposición 5), y por la misma proposición, se rechaza la estructura de tipo H de Σ . La tabla ANOVA muestra una diferencia significativa entre las unidades a lo largo del tiempo, refutándose la idea prevista en la introducción, de que en los primeros meses del año, la inflamabilidad es menor y conforme avanza el año, ésta va aumentando.

6.2. Ejemplo 2

SPSS:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Sexo	N Numérico	1	0		{0, Chico}...	Ninguna	8	Derecha	Nominal	Entrada
2	Distanc8	N Numérico	2	1		Ninguna	Ninguna	8	Derecha	Escala	Entrada
3	Distanc10	N Numérico	2	1		Ninguna	Ninguna	8	Derecha	Escala	Entrada
4	Distanc12	N Numérico	2	1		Ninguna	Ninguna	8	Derecha	Escala	Entrada
5	Distanc14	N Numérico	2	1		Ninguna	Ninguna	8	Derecha	Escala	Entrada

	Sexo	Distanc8	Distanc10	Distanc12	Distanc14
1	Chica	21,0	20,0	21,5	23,0
2	Chica	21,0	21,5	24,0	25,5
3	Chica	20,5	24,0	24,5	26,0
4	Chica	23,5	24,5	25,0	26,5
5	Chica	21,5	23,0	22,5	23,5
6	Chica	20,0	21,0	21,0	22,5
7	Chica	21,5	22,5	23,0	25,0
8	Chica	23,0	23,0	23,5	24,0
9	Chica	20,0	21,0	22,0	21,5
10	Chica	16,5	19,0	19,0	19,5
11	Chica	24,5	25,0	28,0	28,0
12	Chico	26,0	25,0	29,0	31,0
13	Chico	21,5	22,5	23,0	26,5
14	Chico	23,0	22,5	24,0	27,5
15	Chico	25,5	27,5	26,5	27,0
16	Chico	20,0	23,5	22,5	26,0
17	Chico	24,5	25,5	27,0	28,5
18	Chico	22,0	22,0	24,5	26,5
19	Chico	24,0	21,5	24,5	25,5
20	Chico	23,0	20,5	31,0	26,0
21	Chico	27,5	28,0	31,0	31,5
22	Chico	23,0	23,0	23,5	25,0

```

1
2 DATASET ACTIVATE Conjunto_de_datos1.
3 GLM Distanc8 Distanc10 Distanc12 Distanc14 BY Sexo
4 /WSFACTOR=Tiempo 4 Polynomial
5 /METHOD=SSTYPE(3)
6 /PRINT=HOMOGENEITY
7 /CRITERIA=ALPHA(.05)
8 /WSDESIGN=Tiempo
9 /DESIGN=Sexo.
10

```

Se obtienen los siguientes resultados:

Prueba de esfericidad de Mauchly^b

Medida:MEASURE_1

Efecto intra-sujetos	W de Mauchly	Chi-cuadrado aprox.	gl	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Límite-inferior
Tiempo	,599	9,583	5	,088	,733	,869	,333

Contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad.

a. Puede usarse para corregir los grados de libertad en las pruebas de significación promediadas. Las pruebas corregidas se muestran en la tabla Pruebas de los efectos inter-sujetos.

b. Diseño: Intersección + Sexo
Diseño intra-sujetos: Tiempo

Pruebas de efectos intra-sujetos.

Medida:MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Tiempo	Esfericidad asumida	149,940	3	49,980	32,551	,000
	Greenhouse-Geisser	149,940	2,199	68,182	32,551	,000
	Huynh-Feldt	149,940	2,606	57,547	32,551	,000
	Límite-inferior	149,940	1,000	149,940	32,551	,000
Tiempo * Sexo	Esfericidad asumida	9,372	3	3,124	2,035	,119
	Greenhouse-Geisser	9,372	2,199	4,262	2,035	,139
	Huynh-Feldt	9,372	2,606	3,597	2,035	,128
	Límite-inferior	9,372	1,000	9,372	2,035	,169
Error(Tiempo)	Esfericidad asumida	92,125	60	1,535		
	Greenhouse-Geisser	92,125	43,982	2,095		
	Huynh-Feldt	92,125	52,111	1,768		
	Límite-inferior	92,125	20,000	4,606		

Pruebas de los efectos inter-sujetos

Medida:MEASURE_1
Variable transformada:Promedio

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Intersección	50376,480	1	50376,480	2858,148	,000
Sexo	143,821	1	143,821	8,160	,010
Error	352,511	20	17,626		

R:

```
> #Antetodo, debe cambiarse el formato de los datos
> l2<-reshape(potthoffroy,varying=c("d8","d10","d12","d14"),v.names="Distancia",
+ idvar="sex",timevar="Mediciones",times=c("8 años","10 años","12 años","14 años"),
+ new.row.names=1:108,direction="long")
> l.sort2<- l2[order(l2$id),]
>
> #Se comprueba con la siguiente orden que los datos se encuentran en el formato adecuado
> l.sort2[1:12,]
  id sex Mediciones Distancia
1   1  F      8 años    21.0
23  1  F     10 años    20.0
45  1  F     12 años    21.5
67  1  F     14 años    23.0
2   2  F      8 años    21.0
24  2  F     10 años    21.5
46  2  F     12 años    24.0
68  2  F     14 años    25.5
3   3  F      8 años    20.5
25  3  F     10 años    24.0
47  3  F     12 años    24.5
69  3  F     14 años    26.0
```

Se realiza el análisis propio a continuación:

```
> ezANOVA(l.sort2,dv=Distancia,wid=id,within=Mediciones,between=sex)
Aviso: Converting "id" to factor for ANOVA.
Aviso: Converting "Mediciones" to factor for ANOVA.
$ANOVA
      Effect DFn DFd      F      p p<.05      ges
2      sex      1  20  8.159795 9.754269e-03 * 0.24440346
3  Mediciones      3  60 32.551499 1.288152e-12 * 0.25217998
4 sex:Mediciones      3  60  2.034661 1.185833e-01  0.02064313

$`Mauchly's Test for Sphericity`
      Effect      W      p p<.05
3  Mediciones 0.5993849 0.08835012
4 sex:Mediciones 0.5993849 0.08835012

$`Sphericity Corrections`
      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF]
3  Mediciones 0.7330415 8.131422e-10 * 0.8273303 8.307893e-11
4 sex:Mediciones 0.7330415 1.386828e-01  0.8273303 1.312884e-01
p[HF]<.05
3      *
4
```

De nuevo, ambos softwares son coincidentes: en este caso, no hay evidencia suficiente para rechazar la hipótesis nula de Simetría Compuesta.

La tabla ANOVA muestra una diferencia significativa entre niños y niñas: como se pudo observar en la gráfica, la distancia marcada por los niños es mayor en términos generales que la de las niñas.

También se muestra una diferencia significativa a lo largo del tiempo, ya que la distancia, al comenzar las mediciones, se encontraban en un rango de 16mm y 28mm, y al pasar los 6 años, se encuentran en el rango de 20mm y 32mm.

Finalmente, la tabla ANOVA muestra una interacción entre el sexo y el transcurso del tiempo débilmente significativa, lo cuál nos indica que tanto los niños como las niñas siguen un mismo patrón a lo largo del tiempo refiriéndonos a la evolución de la distancia desde la glándula pituitaria hasta la fisura pterigomaxilar.

Contraste Separado: aplicaremos sobre este ejemplo este tipo de contraste explicado en la sección 5.8, dónde el principal objetivo será comprobar si el paso del tiempo afecta por igual a ambos grupos en estudio (chicas y chicos). Para ello se utilizará únicamente R en este caso.

Se considerará en este caso **la matriz de transformación de Helmert** para 2 grupos y 4 intervalos temporales:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -1/3 & -1/2 & 1 \\ -1/3 & -1/2 & -1 \end{pmatrix}$$

Se aplicará la versión normalizada de esta matriz directamente a la matriz de datos, obteniendo un conjunto de variables transformadas definidas como sigue:

```
> v1<-sqrt(0.75)*(potthoffroy[,3]-potthoffroy[,4]/3-potthoffroy[,5]/3-potthoffroy[,6]/3)
> v2<-sqrt(2/3)*(potthoffroy[,4]-potthoffroy[,5]/2-potthoffroy[,6]/2)
> v3<-sqrt(1/2)*(potthoffroy[,5]-potthoffroy[,6])
```

Finalmente, se agruparán estas variables en una matriz y se realizará un contraste multivariante aplicado a esta nueva matriz dentro de los dos grupos, chicas y chicos. Será necesario instalar el paquete "ICSNP" (Klaus Nordhausen, Seija Sirkia, Hannu Oja and David E. Tyler, 2012)

```
> install.packages("ICSNP")
> library(ICSNP)
> sex<-potthoffroy[,2]
> DContrast<-data.frame(v1,v2,v3)
> Z<- as.matrix(DContrast);Z
      v1      v2      v3
[1,] -0.4330127 -1.8371173 -1.0606602
[2,] -2.3094011 -2.6536139 -1.0606602
[3,] -3.7527767 -1.0206207 -1.0606602
[4,] -1.5877132 -1.0206207 -1.0606602
[5,] -1.2990381  0.0000000 -0.7071068
[6,] -1.2990381 -0.6123724 -1.0606602
[7,] -1.7320508 -1.2247449 -1.4142136
[8,] -0.4330127 -0.6123724 -0.3535534
[9,] -1.2990381 -0.6123724  0.3535534
[10,] -2.3094011 -0.2041241 -0.3535534
[11,] -2.1650635 -2.4494897  0.0000000
[12,] -2.0207259 -4.0824829 -1.4142136
[13,] -2.1650635 -1.8371173 -2.4748737
[14,] -1.4433757 -2.6536139 -2.4748737
[15,] -1.2990381  0.6123724 -0.3535534
[16,] -3.4641016 -0.6123724 -2.4748737
[17,] -2.1650635 -1.8371173 -1.0606602
[18,] -2.0207259 -2.8577380 -1.4142136
[19,]  0.1443376 -2.8577380 -0.7071068
[20,] -2.4537386 -6.5319726  3.5355339
[21,] -2.3094011 -2.6536139 -0.3535534
[22,] -0.7216878 -1.0206207 -1.0606602
```

```

> #Se realiza el contraste
> HotellingsT2(Z ~ sex)

Hotelling's two sample T2-test

data: Z by sex
T.2 = 2.1488, df1 = 3, df2 = 18, p-value = 0.1296
alternative hypothesis: true location difference is not equal to c(0,0,0)

```

Por tanto, no hay evidencia estadística suficiente para rechazar la hipótesis de igualdad de medias en ambos grupos. Por tanto, se confirma que el efecto del paso del tiempo en ambos grupos es el mismo.

6.3. Ejemplo 3

SPSS:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Grupo	Númérico	8	2		{1,00, Grup...	Ninguna	8	≡ Derecha	🎯 Nominal	👉 Entrada
2	Peso1	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
3	Peso2	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
4	Peso3	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
5	Peso4	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
6	Peso5	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
7	Peso6	Númérico	3	0		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada

	Grupo	Peso1	Peso2	Peso3	Peso4	Peso5	Peso6
1	Grupo 1	455	460	510	504	436	466
2	Grupo 1	467	565	610	596	542	587
3	Grupo 1	445	530	580	597	582	619
4	Grupo 1	485	542	594	583	611	612
5	Grupo 1	480	500	550	528	562	576
6	Grupo 2	514	560	565	524	552	597
7	Grupo 2	440	480	536	484	567	569
8	Grupo 2	495	570	569	585	576	677
9	Grupo 2	520	590	610	637	671	702
10	Grupo 2	503	555	591	605	649	675
11	Grupo 3	496	560	622	622	632	670
12	Grupo 3	498	540	589	557	568	609
13	Grupo 3	478	510	568	555	576	605
14	Grupo 3	545	565	580	601	633	649
15	Grupo 3	472	498	540	524	532	583

```

1
2 DATASET ACTIVATE Conjunto_de_datos1.
3 GLM Peso1 Peso2 Peso3 Peso4 Peso5 Peso6 BY Grupo
4 /WSFACTOR=Tiempo 6 Polynomial
5 /METHOD=SSTYPE(3)
6 /PRINT=HOMOGENEITY
7 /CRITERIA=ALPHA(.05)
8 /WSDESIGN=Tiempo
9 /DESIGN=Grupo.
10

```

Obteniéndose los resultados:

Prueba de esfericidad de Mauchly^b

Medida: MEASURE_1

Efecto intra-sujetos	W de Mauchly	Chi-cuadrado aprox.	gl	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Límite-inferior
Tiempo	,054	29,390	14	,011	,486	,719	,200

Contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad.

a. Puede usarse para corregir los grados de libertad en las pruebas de significación promediadas. Las pruebas corregidas se muestran en la tabla Pruebas de los efectos inter-sujetos.

b. Diseño: Intersección + Grupo
Diseño intra-sujetos: Tiempo

Pruebas de efectos intra-sujetos.

Medida: MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Tiempo	Esfericidad asumida	142554,500	5	28510,900	52,550	,000
	Greenhouse-Geisser	142554,500	2,428	58715,980	52,550	,000
	Huynh-Feldt	142554,500	3,596	39646,452	52,550	,000
	Límite-inferior	142554,500	1,000	142554,500	52,550	,000
Tiempo * Grupo	Esfericidad asumida	9762,733	10	976,273	1,799	,080
	Greenhouse-Geisser	9762,733	4,856	2010,559	1,799	,146
	Huynh-Feldt	9762,733	7,191	1357,578	1,799	,110
	Límite-inferior	9762,733	2,000	4881,367	1,799	,207
Error(Tiempo)	Esfericidad asumida	32552,600	60	542,543		
	Greenhouse-Geisser	32552,600	29,134	1117,326		
	Huynh-Feldt	32552,600	43,148	754,445		
	Límite-inferior	32552,600	12,000	2712,717		

Pruebas de los efectos inter-sujetos

Medida: MEASURE_1
Variable transformada: Promedio

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Intersección	28133352,90	1	28133352,90	3201,999	,000
Grupo	18548,067	2	9274,033	1,056	,378
Error	105434,200	12	8786,183		

R:

```
> #Se realiza el cambio de formato necesario
> l3<-reshape(datapigs,varying=c("P1*Sem","P3*Sem","P4*Sem","P5*Sem","P6*Sem","P7*Sem"),
+ v.names="Peso",idvar="Grupo",timevar="Semanas",times=c("1*","3*","4*","5*","6*","7*"),
+ new.row.names=1:90,direction="long")
> l.sort3<- l3[order(l3$ID),]
>
> #Se introduce la siguiente orden para tomar Grupo como factor
> l.sort3$Grupo<- factor(l.sort3$Grupo)
>
> #El análisis consiguiente por tanto es:
> ezANOVA(l.sort3,dv=Peso,wid=ID,within=Semanas,between=Grupo)
Aviso: Converting "ID" to factor for ANOVA.
Aviso: Converting "Semanas" to factor for ANOVA.
$ANOVA
      Effect DFn DFd      F      p p<.05      ges
2      Grupo   2  12  1.055525 3.782088e-01 0.11849160
3     Semanas   5  60 52.550457 1.169141e-20 * 0.50814087
4 Grupo:Semanas 10  60  1.799438 8.014404e-02 0.06607624

$`Mauchly's Test for Sphericity`
      Effect      W      p p<.05
3     Semanas 0.05448352 0.01091271 *
4 Grupo:Semanas 0.05448352 0.01091271 *

$`Sphericity Corrections`
      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF]
3     Semanas 0.4855731 4.257871e-11 * 0.6176731 1.461416e-13
4 Grupo:Semanas 0.4855731 1.457120e-01 0.6176731 1.242842e-01
p[HF]<.05
3 *
4
```

Coincidentemente con ambos softwares, el Test de Mauchly nos da evidencia suficiente para rechazar la estructura de tipo \mathbf{H} para Σ .

Las correspondientes correcciones nos muestran que existe diferencia de pesos lo largo del tiempo, mientras que el aumento de peso de los conejillos a lo largo de las semanas es independiente del grupo en el que se encuentre.

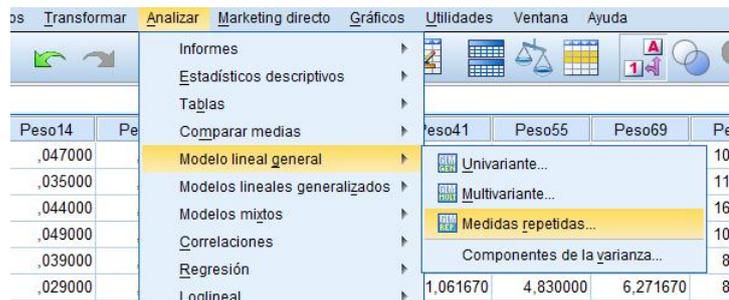
6.4. Ejemplo 4

SPSS: para este ejemplo se detallarán todos los pasos a realizar

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Variedad	Númérico	8	2		{,00, F}...	Ninguna	8	≡ Derecha	🎯 Nominal	👉 Entrada
2	Peso14	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
3	Peso20	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
4	Peso27	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
5	Peso34	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
6	Peso41	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
7	Peso55	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
8	Peso69	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada
9	Peso84	Númérico	8	6		Ninguna	Ninguna	8	≡ Derecha	📏 Escala	👉 Entrada

	Variedad	Peso14	Peso20	Peso27	Peso34	Peso41	Peso55	Peso69	Peso84
1	F	,047000	,136000	,272833	,983333	1,796667	4,117140	11,037140	10,403330
2	F	,035000	,112000	,268000	1,053330	1,715000	6,728330	10,313330	11,088330
3	F	,044000	,146000	,360500	,775000	1,641667	3,748330	10,875000	16,027500
4	F	,049000	,097000	,227500	,833333	1,354000	5,988330	10,585000	10,911430
5	F	,039000	,121000	,271667	,658333	2,390000	5,483333	9,936670	8,888330
6	F	,029000	,092000	,251000	,541667	1,061670	4,830000	6,271670	8,583330
7	F	,041000	,145000	,405333	,760000	1,493333	6,248330	11,111700	8,880000
8	F	,038000	,127000	,340167	,994000	2,281670	8,506000	8,973330	11,632000
9	P	,070000	,208000	,600833	1,901667	2,965000	8,143330	15,700000	19,888300
10	P	,068000	,252000	,580500	2,014000	2,673333	8,800000	14,485000	22,643300
11	P	,077000	,224000	,341333	1,740000	4,524000	11,656700	20,316700	21,402000
12	P	,099000	,198000	,491167	1,205000	2,800000	7,363330	15,141670	17,186670
13	P	,068000	,232000	,436000	1,443333	3,260000	8,208330	19,406700	15,661670
14	P	,063000	,209000	,740833	1,268330	3,890000	9,092000	14,490000	19,095000
15	P	,133000	,303000	1,117000	2,180000	3,536667	13,160000	15,790000	14,775000
16	P	,081000	,255000	,764833	1,093330	3,051670	8,105000	24,424000	18,918300

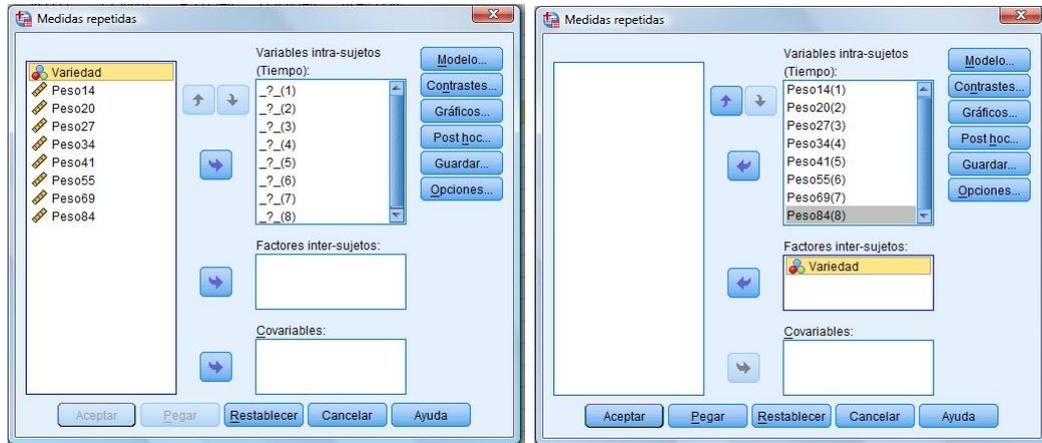
A continuación, se aplica la instrucción de análisis de la varianza de medidas repetidas:



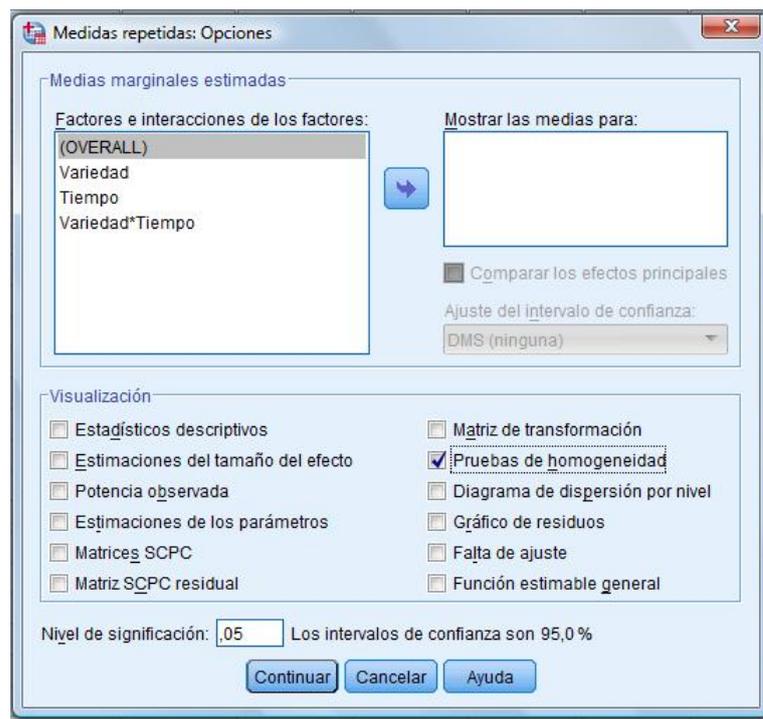
Al seleccionar, realizaremos la siguiente selección del tiempo como factor intra-sujeto:



La siguiente pantalla se centra en la selección de un factor inter-sujetos, que será en este caso la variedad de genotipo del que se trate:



Es esencial pedir las pruebas de homocedasticidad en la ventana de Opciones:



Finalmente se selecciona Aceptar y se obtienen los resultados del estudio.

Prueba de esfericidad de Mauchly^b

Medida:MEASURE_1

Efecto intra-sujetos	W de Mauchly	Chi-cuadrado aprox.	gl	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Límite-inferior
Tiempo	,000	222,796	27	,000	,370	,494	,143

Contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad.

a. Puede usarse para corregir los grados de libertad en las pruebas de significación promediadas. Las pruebas corregidas se muestran en la tabla Pruebas de los efectos inter-sujetos.

b. Diseño: Intersección + Variedad
Diseño intra-sujetos: Tiempo

Pruebas de efectos intra-sujetos.

Medida:MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Tiempo	Esfericidad asumida	4244,391	7	606,342	274,038	,000
	Greenhouse-Geisser	4244,391	2,593	1637,150	274,038	,000
	Huynh-Feldt	4244,391	3,461	1226,361	274,038	,000
	Límite-inferior	4244,391	1,000	4244,391	274,038	,000
Tiempo * Variedad	Esfericidad asumida	303,279	7	43,326	19,581	,000
	Greenhouse-Geisser	303,279	2,593	116,981	19,581	,000
	Huynh-Feldt	303,279	3,461	87,628	19,581	,000
	Límite-inferior	303,279	1,000	303,279	19,581	,001
Error(Tiempo)	Esfericidad asumida	216,837	98	2,213		
	Greenhouse-Geisser	216,837	36,296	5,974		
	Huynh-Feldt	216,837	48,453	4,475		
	Límite-inferior	216,837	14,000	15,488		

Pruebas de los efectos inter-sujetos

Medida:MEASURE_1
Variable transformada:Promedio

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Intersección	3262,378	1	3262,378	1356,600	,000
Variedad	241,494	1	241,494	100,421	,000
Error	33,667	14	2,405		

R:

```
> Soybean<-Soybean[157:284,-3]
> l.sort4<- Soybean[order(Soybean$Plot),]
> l.sort4$Time<- factor(l.sort4$Time)
> ezANOVA(l.sort4,dv=weight,wid=Plot,within=Time,between=Variety)
Aviso: You have removed one or more Ss from the analysis. Refactoring "Plot" for ANOVA.
$ANOVA
      Effect DFn DFd      F      p p<.05      ges
2      Variety  1  14 100.4209 9.101472e-08 * 0.4908433
3          Time  7  98 274.0378 2.168623e-61 * 0.9442692
4 Variety:Time  7  98  19.5811 3.576611e-16 * 0.5476491

$`Mauchly's Test for Sphericity`
      Effect      W      p p<.05
3          Time 3.557692e-09 1.094327e-31 *
4 Variety:Time 3.557692e-09 1.094327e-31 *

$`Sphericity Corrections`
      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF] p[HF]<.05
3          Time 0.370364 3.838637e-24 * 0.4619565 1.419390e-29 *
4 Variety:Time 0.370364 2.909866e-07 * 0.4619565 1.430878e-08 *
```

Coincidentemente de nuevo para SPSS y R, el Test de Mauchly produce un estadístico significativo, al igual que los ajustes de Greenhouse-Geisser y Huynh-Feldt.

Esto indica claramente que existe diferencia en el crecimiento a lo largo del transcurso de las semanas. Además, el crecimiento de la soja depende de el tipo de soja que se trate (en la Figura 7 se observa claramente que la soja del genotipo experimental experimenta un mayor crecimiento que la soja del genotipo comercial).

7. Referencias

Bates,D. , Maechler,M. and Bolker,B. (2014) *MEMSS: Data sets from Mixed-effects Models in S* (R package version 0.9-2)
URL <http://CRAN.R-project.org/package=MEMSS>

Crowder, M.J. and Hand, D.J. (1990) *Analysis of Repeated Measures*.
London: Chapman and Hall/CRC Press.

Davidian, M. (2005) ST 732. *Applied Longitudinal Data Analysis*, Lecture Notes.
North Carolina State University.

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall/CRC Press.

IBM Corp (2013) *IBM SPSS Statistics for Windows, Version 22.0*.
Armonk, NY: IBM Corp.

Lazarsfeld, P.F. (1940) Panel Studies. *Public Opinion Quarterly*.

Lazarsfeld, P.F. and Fiske, M. (1938) The Panel a a New Tool for Measuring Opinion. *Public Opinion Quarterly*.

Nordhausen,K. , Sirkia,S. , Oja,H. and Tyler,D.E. (2012) *ICSNP: Tools for Multivariate Nonparametrics*(R package version 1.0-9)
URL <http://CRAN.R-project.org/package=ICSNP>.

Potthoff, R.F. and Roy, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313-326

R Core Team (2015) *R: A language and environment for statistical computing*.
R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011) *mice: Multivariate Imputation by Chained Formations in R*. *Journal of Statistical Software* 45(3), 1-67.
URL <http://www.jstatsoft.org/v45/i03/>.

Vonesh, E. and Chinchilli, V.M. (1997) *Linear and Nolinear Models for the Analysis of Repeated Measurements* 9. CRC Press