

PAC6

Maria Lucas Gascón

2024-01-17

Problema 1

Cargamos los datos del problema.

```
# Seteamos el directorio de trabajo
setwd("D:/Antiguos estudios/MASTER2/Sem3/Diseño/PAC6/PAC6")

# Importamos Los datos
data <- read.csv("dades_problema1_pec6.csv", dec = ",", header = TRUE, sep = ";")

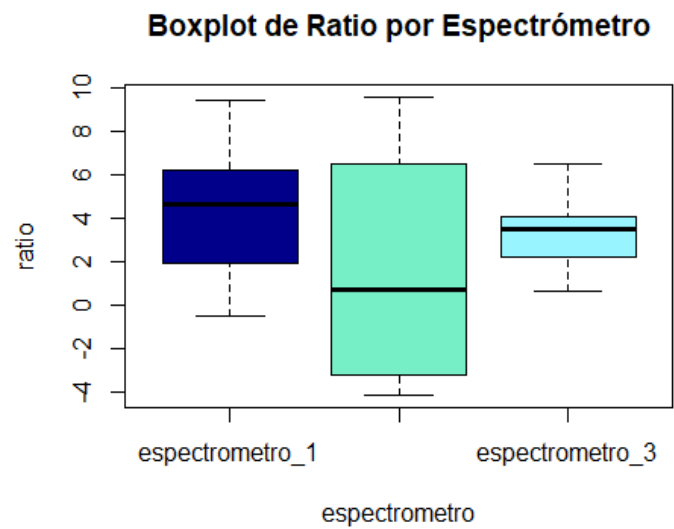
# Marcamos Los factores
data$espectrometro <- as.factor(data$espectrometro)
data$suelo <- as.factor(data$suelo)
data$dia <- factor(data$dia, levels = c(1, 2, 3), labels = c("dia1", "dia2", "dia3"))
```

1.1 Análisis descriptivo de los datos

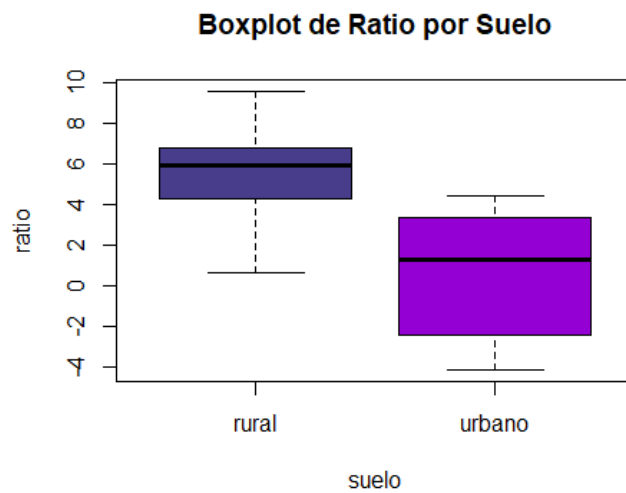
```
# Primero examinamos la variable ratio
summary(data)
```

##	ratio	espectrometro	suelo	dia
##	Min. :-4.1356	espectrometro_1:12	rural :18	dia1:12
##	1st Qu.: 0.9882	espectrometro_2:12	urbano:18	dia2:12
##	Median : 3.5102	espectrometro_3:12		dia3:12
##	Mean : 3.1112			
##	3rd Qu.: 5.7919			
##	Max. : 9.5738			

```
# Box plot de ratio por espectrometro
boxplot(ratio ~ espectrometro, data = data, col = c("blue4", "aquamarine2", "cadetblue1"), main = "Boxplot de Ratio por Espectrómetro")
```

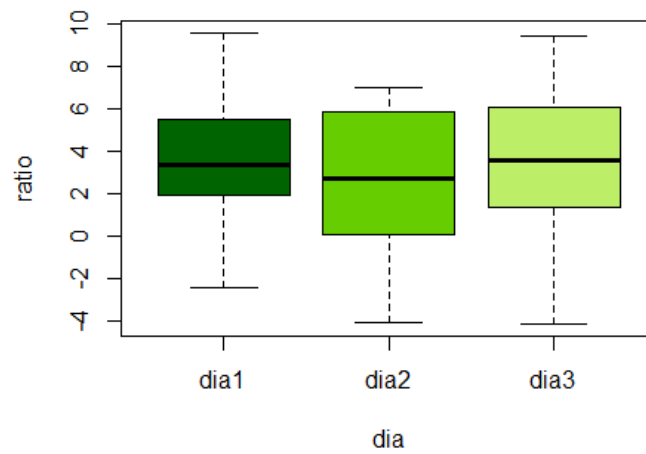


```
# Box plot de ratio por suelo
boxplot(ratio ~ suelo, data = data, col = c("darkslateblue", "darkviolet"),
main = "Boxplot de Ratio por Suelo")
```



```
# Box plot de ratio por dia
boxplot(ratio ~ dia, data = data, col = c("darkgreen", "chartreuse3",
"darkolivegreen2"), main = "Boxplot de Ratio por Dia")
```

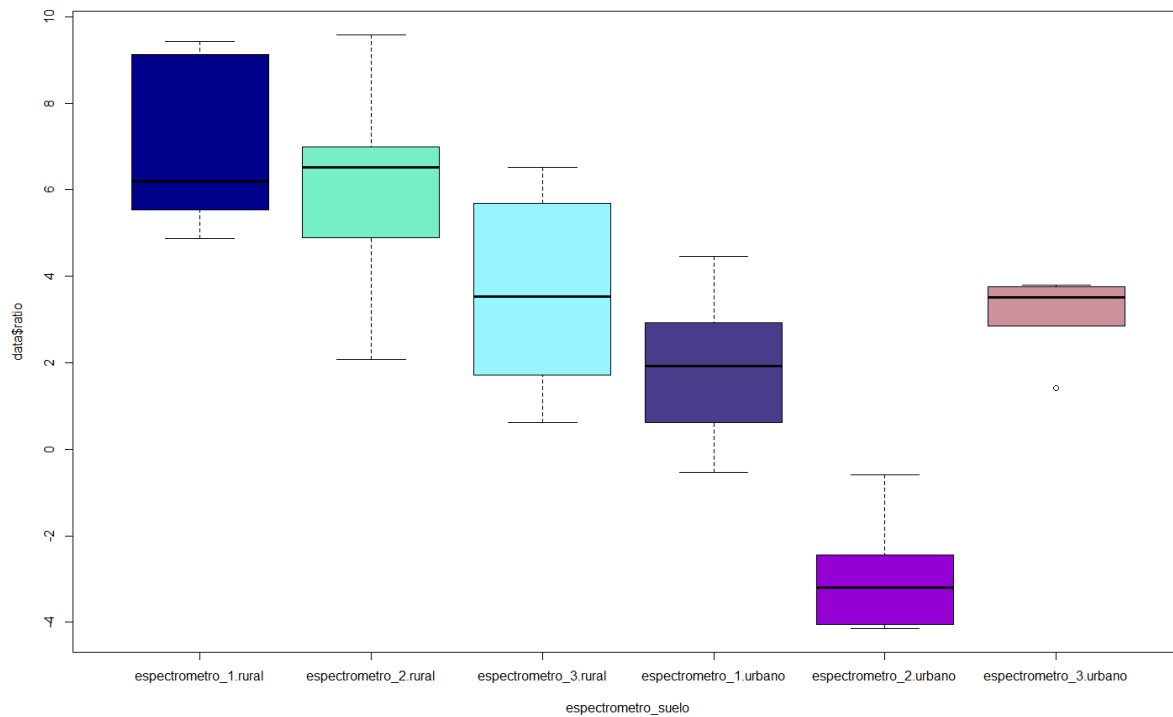
Boxplot de Ratio por Dia



```
# Create an interaction factor of espectrometro and suelo
espectrometro_suelo <- interaction(data$espectrometro, data$suelo)
```

```
# Box plot of ratio by espectrometro and suelo
boxplot(data$ratio ~ espectrometro_suelo, col = c("blue4", "aquamarine2",
"cadetblue1", "darkslateblue", "darkviolet", "pink3"),
main = "Boxplot de Ratio por Espectrómetro y Suelo")
```

Boxplot de Ratio por Espectrómetro y Suelo



En el análisis del boxplot correspondiente al “Ratio por Espectrómetro”, se observa una ligera disparidad en las mediciones entre los espectrómetros. Esta discrepancia se evidencia mediante la diferencia en la altura de las cajas y la variación en la media de las mediciones, destacada por una línea negra en cada caja.

En cuanto al boxplot relativo al “Ratio por Suelo”, se destaca que el suelo rural presenta, en general, un ratio de 14N a 15N superior al observado en el suelo urbano. Esta diferencia se refleja tanto en la altura como en el ancho de las cajas, indicando que el suelo rural exhibe un ratio más elevado en comparación con el suelo urbano.

Adicionalmente, se observa una notable consistencia en las mediciones diarias, ya que muestran una variación mínima.

Al combinar los datos de espectrómetro y suelo en un único gráfico, se puede evaluar la interacción entre ambos factores. Se aprecia que, para el suelo urbano, las mediciones de los espectrómetros presentan una mayor disparidad en comparación con el suelo rural, donde las mediciones son más uniformes. Específicamente, se destaca que el espectrómetro 2 en el suelo urbano proporciona mediciones notablemente más bajas.

1.2 Modelo lineal

En este problema, se busca modelar la proporción de 14N a 15N en dos tipos de suelos (urbano, rural) utilizando tres espectrómetros distintos. Se realizaron mediciones en 18 muestras de cada tipo de suelo. Cada muestra se asignó a uno de los tres espectrómetros al azar, y para cada combinación de suelo y espectrómetro, se eligieron 3 días al azar para realizar las mediciones. El objetivo es analizar la influencia de los factores suelo, espectrómetro y día, así como sus interacciones.

El modelo lineal puede expresarse de la siguiente manera:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + C_{k(ij)} + \epsilon_{ijk}$$

- Y_{ijk} es la medida de la proporción de 14N a 15N para la k-ésima observación en el suelo i, espectrómetro j y día k.
- μ es la media global.
- α_i es el efecto del i-ésimo nivel del factor suelo. La suma de todos los efectos del factor suelo debe ser cero: $\sum_{i=1}^2 \alpha_i = 0$
- β_j es el efecto del j-ésimo nivel del factor espectrómetro. La suma de todos los efectos del factor espectrómetro debe ser cero: $\sum_{j=1}^3 \beta_j = 0$
- $(\alpha\beta)_{ij}$ es la interacción entre el suelo i y el espectrómetro j. La suma de las interacciones debe ser cero: $\sum \alpha\beta_{ij} = 0$
- $C_{k(ij)}$ es el efecto del k-ésimo nivel del factor día anidado en la combinación suelo i y espectrómetro j. Como es un efecto aleatorio, no hay restricciones específicas, pero se asume que sigue una distribución normal con media cero: $C_{k(ij)} \sim N(0, \sigma_C^2)$

- ϵ_{ijk} es el error aleatorio asociado con la k-ésima observación en la combinación suelo i, espectrómetro j y día k. Se asume que sigue una distribución normal con media cero: $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$

1.3 Hipótesis de interés

Para el factor suelo:

- $H_0: \alpha_1 = \alpha_2 = 0$. No hay diferencia significativa en la proporción de 14N a 15N entre los suelos urbanos y rurales.
- H_1 : Al menos un $\alpha_i \neq 0$. Hay al menos una diferencia significativa en la proporción de 14N a 15N entre los suelos urbanos y rurales.

Para el factor espectrómetro:

- $H_0: \beta_1 = \beta_2 = \beta_3 = 0$. No hay diferencia significativa en la proporción de 14N a 15N entre los espectrómetros.
- H_1 : Al menos un $\beta_j \neq 0$. Hay al menos una diferencia significativa en la proporción de 14N a 15N entre los tres espectrómetros distintos.

Para la interacción suelo-espectrómetro:

- $H_0: (\alpha\beta)_{ij} = 0$ para todos los i, j . No hay interacción significativa entre el tipo de suelo y el espectrómetro.
- H_1 : Existe al menos una $(\alpha\beta)_{ij} \neq 0$. Hay al menos una interacción significativa entre el tipo de suelo y el espectrómetro.

Para el efecto aleatorio del día:

- $H_0: \sigma_{C(AB)}^2 = 0$. No hay variabilidad significativa asociada con los días anidados en la combinación de suelo y espectrómetro.
- $H_1: \sigma_{C(AB)}^2 > 0$. Existe variabilidad significativa asociada con los días anidados en la combinación de suelo y espectrómetro.

Para el error aleatorio:

- $H_0: \sigma_\epsilon^2 = 0$. No hay variabilidad significativa no explicada por los factores considerados en el modelo.
- $H_1: \sigma_\epsilon^2 > 0$. Existe variabilidad significativa no explicada por los factores considerados en el modelo.

1.4 Estudiar las hipótesis

Definimos el modelo

```
library(lme4)
```

```
library(car)
```

```
modelo <- lmer(ratio ~ suelo + espectrometro + suelo:espectrometro + (1 |  
dia:suelo:espectrometro), data = data)
```

```
summary(modelo)

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ratio ~ suelo + espectrometro + suelo:espectrometro + (1 |
## dia:suelo:espectrometro)
## Data: data
##
## REML criterion at convergence: 128.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.8173 -0.5379  0.0471  0.4748  1.8520
##
## Random effects:
## Groups              Name              Variance Std.Dev.
## dia:suelo:espectrometro (Intercept) 2.071    1.439
## Residual                    1.841    1.357
## Number of obs: 36, groups: dia:suelo:espectrometro, 18
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                   6.8943    0.9985    6.904
## suelourbano                   -5.0119    1.4122   -3.549
## espectrometroespectrometro_2  -0.8031    1.4122   -0.569
## espectrometroespectrometro_3  -3.2959    1.4122   -2.334
## suelourbano:espectrometroespectrometro_2 -4.0167    1.9971   -2.011
## suelourbano:espectrometroespectrometro_3  4.5518    1.9971    2.279
##
## Correlation of Fixed Effects:
##              (Intr) sulrbn espc_2 espc_3 slr:_2
## suelourbano -0.707
## espctrmtr_2 -0.707  0.500
## espctrmtr_3 -0.707  0.500  0.500
## slrbn:spc_2  0.500 -0.707 -0.707 -0.354
## slrbn:spc_3  0.500 -0.707 -0.354 -0.707  0.500

# Intervalos de confianza
library(broom)

confint(modelo)

## Computing profile confidence intervals ...

##              2.5 %      97.5 %
## .sig01          0.000000  1.8099950
## .sigma          1.010479  1.9580075
## (Intercept)      5.207159  8.5814518
## suelourbano      -7.397879 -2.6259082
## espectrometroespectrometro_2 -3.189038  1.5829331
```

```
## espectrometroespectrometro_3          -5.681922 -0.9099516
## suelo urbano:espectrometroespectrometro_2 -7.391029 -0.6424435
## suelo urbano:espectrometroespectrometro_3  1.177523  7.9261089
```

En la síntesis del modelo, no se evidencian los factores que resultan significativos. No obstante, es posible calcular los valores de confianza mediante el empleo de un valor crítico de la distribución t de Student. En esta distribución, un nivel de confianza del 0.05 implicaría un valor crítico de 1.96.

El intervalo de confianza se formula como el estimado del coeficiente más o menos 1.96 veces el error estándar del coeficiente. Para mayor conveniencia, este cálculo se realiza de forma automática mediante el paquete “broom”.

Al examinar los intervalos de confianza, se constata que todos los factores son estadísticamente significativos, a excepción del espectrómetro 2, ya que su intervalo de confianza incluye el valor 0. Por otro lado, el intercepto, el suelo urbano, el espectrómetro 3 y las interacciones suelo urbano:espectrómetro 2 y suelo urbano:espectrómetro 3, exhiben diferencias significativas en comparación con la referencia establecida, que corresponde al suelo rural y el espectrómetro 1.

```
# Testamos Las hipótesis
anova_modelo <- Anova(modelo, type = "III", test = "Chisq")
print(anova_modelo)

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: ratio
##              Chisq Df Pr(>Chisq)
## (Intercept)    47.6697  1  5.044e-12 ***
## suelo          12.5961  1  0.0003866 ***
## espectrometro    5.9247  2  0.0516970 .
## suelo:espectrometro 18.4324  2  9.942e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La especificación de la ANOVA de “tipo III” implica la atribución de variabilidad a cada término, teniendo en cuenta todos los demás términos, incluidas las interacciones que involucran dicho término. Esta aproximación resulta útil cuando se presentan interacciones y se busca evaluar la contribución de cada variable a la variabilidad explicada, considerando las otras variables y sus interacciones.

Cabe destacar que, dado que no se han verificado las suposiciones de normalidad y homocedasticidad, se prefiere utilizar la prueba de Chi-cuadrado en lugar de la prueba F.

En cuanto a los resultados, se observa que tanto el suelo como la interacción suelo:espectrofotómetro son significativas (valor de $p < 0.05$), mientras que el espectrofotómetro por sí mismo no lo es (valor de $p > 0.05$). A continuación, se detalla la interpretación de estos hallazgos:

- Suelo significativo: Se constatan diferencias estadísticamente significativas en la proporción de 14N a 15N entre los suelos urbanos y rurales (valor de $p < 0.05$, rechazo de H_0), después de considerar los demás efectos en el modelo. En resumen, existe evidencia estadística que respalda la afirmación de que el tipo de suelo tiene un impacto significativo en la proporción de nitrógeno.
- Espectrómetro no significativo: No se dispone de suficiente evidencia estadística para sostener que existen diferencias en la proporción de 14N a 15N entre los distintos espectrofotómetros (valor de $p > 0.05$, aceptación de H_0), después de tomar en cuenta los otros efectos. En resumen, las variaciones observadas en la proporción de nitrógeno no pueden atribuirse de manera significativa a las diferencias entre los espectrofotómetros utilizados. Es posible que la no significancia se deba a que el espectrofotómetro 3 no difiere, a pesar de que el 2 sí lo hace según el intervalo de confianza.
- Interacción suelo-espectrómetro significativa: Se evidencia que la influencia del espectrofotómetro en la proporción de nitrógeno depende del tipo de suelo (valor de $p < 0.05$, rechazo de H_0). En otras palabras, la diferencia en la proporción de 14N a 15N entre los espectrofotómetros puede variar según si el suelo es urbano o rural. Esta interacción indica que el efecto del espectrofotómetro no es constante en todos los suelos y viceversa.

En resumen, los resultados sugieren que el tipo de suelo y la interacción entre el tipo de suelo y el espectrofotómetro son factores significativos en la variabilidad de la proporción de 14N a 15N, mientras que el espectrofotómetro por sí mismo no tiene un efecto significativo en esta proporción, aunque este patrón parece no aplicar al espectrofotómetro 2.

1.5 Comparaciones múltiples

Dado que el factor suelo consta únicamente de dos niveles, carece de fundamento llevar a cabo una comparación múltiple. A pesar de que el espectrómetro no muestra significancia de manera general, hemos observado que el espectrómetro 2 presenta un significado estadístico; procederemos a verificar estos resultados. Por último, examinaremos también la interacción suelo:espectrómetro.

Para la interacción:

```
# install.packages("emmeans")
library(emmeans)

library(knitr)

# Realizamos las comparaciones múltiples
emmeans_result <- emmeans(modelo, pairwise ~ suelo:spectrometro)

# Ajustamos p-valores usando la corrección de Bonferroni
adjusted_emmeans <- summary(emmeans_result, infer = c(TRUE, TRUE), adjust =
"bonferroni")
```


Imprimimos los valores ajustados

```
print(adjusted_emmeans)
```

```
## $emmeans
```

##	suelo	espectrometro	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
##	rural	espectrometro_1	6.89	0.999	12	3.74620	10.042	6.904	0.0001
##	urbano	espectrometro_1	1.88	0.999	12	-1.26569	5.031	1.885	0.5031
##	rural	espectrometro_2	6.09	0.999	12	2.94315	9.239	6.100	0.0003
##	urbano	espectrometro_2	-2.94	0.999	12	-6.08548	0.211	-2.942	0.0740
##	rural	espectrometro_3	3.60	0.999	12	0.45026	6.746	3.604	0.0217
##	urbano	espectrometro_3	3.14	0.999	12	-0.00981	6.286	3.143	0.0509

```
##
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## Confidence level used: 0.95
```

```
## Conf-level adjustment: bonferroni method for 6 estimates
```

```
## P value adjustment: bonferroni method for 6 tests
```

```
##
```

```
## $contrasts
```

##	contrast	estimate	SE	df	lower.CL
##	rural espectrometro_1 - urbano espectrometro_1	5.012	1.41	12	-0.141
##	rural espectrometro_1 - rural espectrometro_2	0.803	1.41	12	-4.350
##	rural espectrometro_1 - urbano espectrometro_2	9.832	1.41	12	4.679
##	rural espectrometro_1 - rural espectrometro_3	3.296	1.41	12	-1.857
##	rural espectrometro_1 - urbano espectrometro_3	3.756	1.41	12	-1.397
##	urbano espectrometro_1 - rural espectrometro_2	-4.209	1.41	12	-9.362
##	urbano espectrometro_1 - urbano espectrometro_2	4.820	1.41	12	-0.333
##	urbano espectrometro_1 - rural espectrometro_3	-1.716	1.41	12	-6.869
##	urbano espectrometro_1 - urbano espectrometro_3	-1.256	1.41	12	-6.409
##	rural espectrometro_2 - urbano espectrometro_2	9.029	1.41	12	3.876
##	rural espectrometro_2 - rural espectrometro_3	2.493	1.41	12	-2.660
##	rural espectrometro_2 - urbano espectrometro_3	2.953	1.41	12	-2.200
##	urbano espectrometro_2 - rural espectrometro_3	-6.536	1.41	12	-11.689
##	urbano espectrometro_2 - urbano espectrometro_3	-6.076	1.41	12	-11.228
##	rural espectrometro_3 - urbano espectrometro_3	0.460	1.41	12	-4.693

```
## upper.CL t.ratio p.value
```

```
## 10.165 3.549 0.0600
```

```
## 5.956 0.569 1.0000
```

```
## 14.985 6.962 0.0002
```

```
## 8.449 2.334 0.5670
```

```
## 8.909 2.660 0.3120
```

```
## 0.944 -2.980 0.1721
```

```
## 9.973 3.413 0.0772
```

```
## 3.437 -1.215 1.0000
```

```
## 3.897 -0.889 1.0000
```

```
## 14.181 6.393 0.0005
```

```
## 7.646 1.765 1.0000
```

```
## 8.106 2.091 0.8769
```

```
## -1.383 -4.628 0.0087
```

```
## -0.923 -4.302 0.0154
```

```
##      5.613   0.326   1.0000
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: bonferroni method for 15 estimates
## P value adjustment: bonferroni method for 15 tests
```

Como podemos ver, las siguientes combinaciones resultan significativas:

```
# Extraemos Los contrastes y p-valores
contrast_df <- as.data.frame(adjusted_emmeans$contrasts)

# Filtramos por p-valores < 0.05
significant_contrasts <- contrast_df[contrast_df$p.value < 0.05,
c("contrast", "p.value")]

# Formateamos La tabla para imprimir
formatted_table <- data.frame(
  Contrast = significant_contrasts$contrast,
  P_Value = sprintf("%.4f", significant_contrasts$p.value)
)

# Imprimimos La tabla
library(knitr)
kable(formatted_table, col.names = c("Contraste", "P-Valor"), format =
"pandoc")
```

Contraste	P-Valor
rural espectrometro_1 - urbano espectrometro_2	0.0002
rural espectrometro_2 - urbano espectrometro_2	0.0005
urbano espectrometro_2 - rural espectrometro_3	0.0087
urbano espectrometro_2 - urbano espectrometro_3	0.0154

Eso significa que:

- Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo rural es diferente cuando se utiliza el espectrómetro_1 en comparación con el suelo urbano cuando se utiliza el espectrómetro_2.
- Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo rural es diferente cuando se utiliza el espectrómetro_2 en comparación con el suelo urbano cuando también se utiliza el espectrómetro_2.
- Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo urbano es diferente cuando se utiliza el espectrómetro_2 en comparación con el suelo rural cuando se utiliza el espectrómetro_3.

- Hay evidencia significativa para afirmar que la proporción de nitrógeno en el suelo urbano es diferente cuando se utiliza el espectrómetro_2 en comparación con el mismo suelo urbano cuando se utiliza el espectrómetro_3.

Estos valores de p bajos indican que, tras ajustar para múltiples comparaciones, las diferencias observadas en estas combinaciones específicas son estadísticamente significativas. Para las demás combinaciones, no se detectan diferencias significativas.

Es relevante subrayar la importancia de realizar correcciones para mitigar el error de tipo 1, especialmente dado el gran número de comparaciones efectuadas. En este caso, se ha implementado la corrección de Bonferroni.

Para el espectrómetro:

```
# Cargar el paquete
library(emmeans)
```

```
# Calcular las medias marginales ajustadas (LSMeans)
lsmeans_modelo <- emmeans(modelo, ~espectrometro)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
# Comparaciones múltiples entre niveles de espectrómetros
comparaciones <- pairs(lsmeans_modelo)
```

```
# Ver los resultados
summary(comparaciones)
```

```
## contrast estimate SE df t.ratio p.value
## espectrometro_1 - espectrometro_2 2.81 0.999 12 2.816 0.0385
## espectrometro_1 - espectrometro_3 1.02 0.999 12 1.022 0.5780
## espectrometro_2 - espectrometro_3 -1.79 0.999 12 -1.794 0.2130
##
## Results are averaged over the levels of: suelo
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Se ratifica la existencia de una disparidad significativa entre el espectrómetro 2 y el espectrómetro 1, dado que se obtiene un valor de 0.03, lo cual conduce al rechazo de la hipótesis nula (H_0) que sostiene que no existen diferencias entre ambos. Por otro lado, se constata la ausencia de discrepancias entre el espectrómetro 1 y el espectrómetro 3, así como entre el espectrómetro 2 y el espectrómetro 3 (valor de $p > 0.05$). Este resultado ya se reflejaba en el resumen del modelo, aunque la comparación entre el espectrómetro 2 y el espectrómetro 3 no estaba detallada.

Problema 2

Cargamos los datos del problema.

```
# Seteamos el directorio de trabajo
setwd("D:/Antiguos estudios/MASTER2/Sem3/Diseño/PAC6/PAC6")

# Importamos Los datos
data <- read.csv("dades_problema2_pec6.csv", dec = ",", header = TRUE, sep =
";")

# Marcamos Los factores
data$dieta <- as.factor(data$dieta)
```

2.1 Análisis descriptivo de los datos

```
# Realiza un resumen descriptivo
summary(data)
```

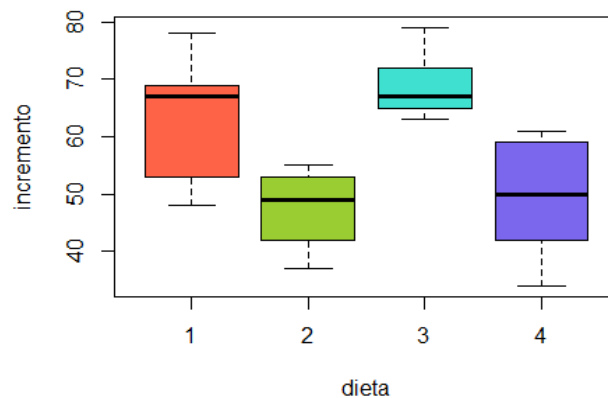
```
##  dieta      consumo      incremento
##  1:5   Min.   :35.00   Min.   :34.00
##  2:5   1st Qu.:42.75   1st Qu.:48.75
##  3:5   Median :47.00   Median :57.00
##  4:5   Mean   :46.25   Mean   :57.15
##       3rd Qu.:51.00   3rd Qu.:67.00
##       Max.   :55.00   Max.   :79.00
```

```
# Gráfico relación consumo-incremento
```

```
color_dieta = c("tomato1", "olivedrab3", "turquoise", "slateblue2")
```

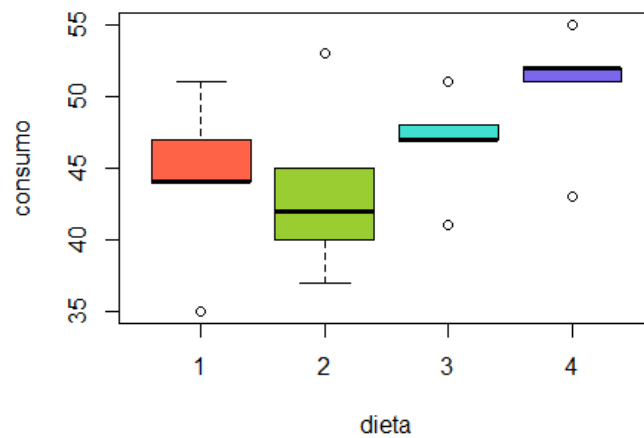
```
# Boxplot incremento por dieta
```

```
boxplot(incremento ~ dieta, data = data, col = color_dieta)
```



```
# Boxplot consumo por dieta
```

```
boxplot(consumo ~ dieta, data = data, col = color_dieta)
```

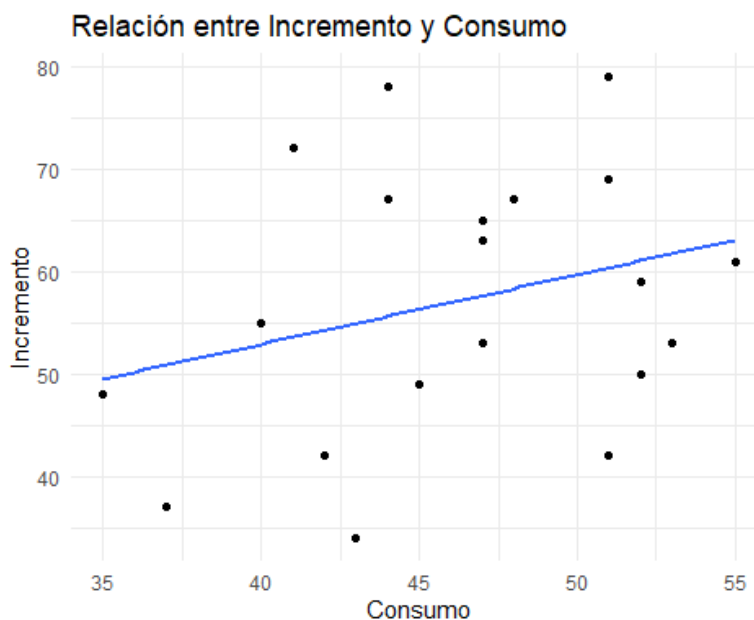


```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

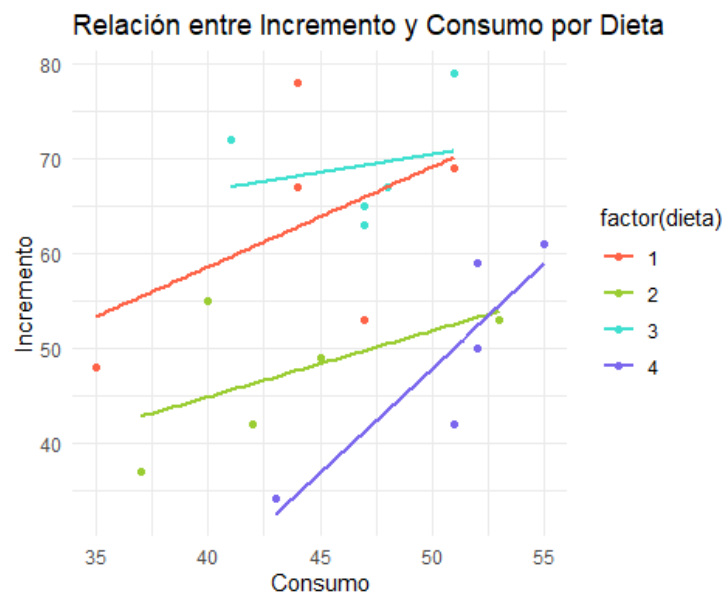
# Gráfico de dispersión Incremento vs Consumo sin dieta
ggplot(data, aes(x = consumo, y = incremento)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # Añade la línea de tendencia
  # general
  labs(title = "Relación entre Incremento y Consumo",
        x = "Consumo",
        y = "Incremento") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



```
# Gráfico de dispersión Incremento vs Consumo por dieta
ggplot(data, aes(x = consumo, y = incremento, color = factor(dieta))) +
  scale_color_manual(values = color_dieta) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # Añade la línea de tendencia +
  labs(title = "Relación entre Incremento y Consumo por Dieta",
        x = "Consumo",
        y = "Incremento") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



Se evidencia una distribución equilibrada con 5 animales por cada tipo de dieta, lo que configura un diseño balanceado.

El análisis del primer boxplot, que representa la relación entre el incremento de peso y la dieta, sugiere variaciones significativas entre los grupos. Esto se infiere al observar diferencias en las alturas de las cajas y la línea central, que representa la media de cada grupo. Además, las distintas alturas y longitudes de las líneas de dispersión indican variabilidades divergentes entre los grupos.

En contraste, al examinar el segundo boxplot que relaciona el consumo de calorías con la dieta, se puede inferir si existen disparidades en el consumo calórico entre los grupos. Efectivamente, se observan diferencias notables entre los grupos.

Por otro lado, el gráfico que ilustra la relación entre el aumento de peso y el consumo calórico revela una correlación positiva general. A mayor consumo de calorías, se evidencia un mayor incremento de peso, y viceversa.

Al analizar esta relación de manera específica para cada dieta, se mantiene la correlación positiva, pero se aprecian pendientes distintas para cada tipo de dieta. Destaca que la dieta 4 exhibe la correlación positiva más pronunciada.

En conjunto, estos hallazgos contribuyen a una comprensión más profunda de las relaciones entre las variables, lo que facilita el ajuste preciso de los modelos de regresión.

2.2 Modelo Lineal

Ajustando según el consumo calórico

El modelo lineal, teniendo en cuenta el consumo calórico, puede especificarse como:

$$y_{ij} = \mu + \alpha_i + \beta X_{ij} + \epsilon_{ij}$$

Donde:

- y_{ij} es el aumento de peso del animal j bajo la dieta i .
- μ es el intercepto común para todas las dietas.
- α_i es el intercepto específico para la dieta i .
- β es la pendiente de regresión, que representa el efecto del consumo calórico en el aumento de peso.
- X_{ij} es el consumo calórico del animal j bajo la dieta i .
- ϵ_{ij} son errores experimentales aleatorios con distribución normal, media 0 y varianza σ^2 . ($\epsilon_{ij} \sim N(0, \sigma^2)$).

Suposiciones adicionales:

- $\sum_i^a \alpha_i = 0$
- El coeficiente de regresión β es el mismo para todos los grupos de tratamiento (dietas).
- Las dietas no influyen sobre el consumo calórico (covariante X)

El modelo covariante propone utilizar una línea de regresión para cada tratamiento. Esta línea representa cómo la variable de interés cambia en relación con la covariable. Todas las líneas son similares en términos de la dirección de cambio, es decir, comparten la misma “pendiente” (β). Esto significa que todos los tratamientos responden de manera similar a la covariable.

Sin embargo, hay diferencias entre los tratamientos, y estas diferencias se reflejan en los términos independientes o “ordenadas al origen” ($\mu + \alpha_i$). Cada tratamiento tiene su propia ordenada al origen (α_i), que se suma a una ordenada al origen común (μ). Esto significa que aunque las líneas son paralelas, pueden tener alturas diferentes en el eje vertical, indicando que cada tratamiento puede tener un nivel inicial diferente en la variable de interés.

En resumen, el modelo utiliza líneas de regresión paralelas para representar cómo los diferentes tratamientos responden a la covariable, con la flexibilidad de tener diferentes niveles iniciales debido a los términos α_i .

Sin ajustar por consumo calórico

El modelo lineal, sin tener en cuenta el consumo calórico, puede especificarse como:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Donde:

- y_{ij} es el aumento de peso del animal j bajo la dieta i .
- μ es el efecto medio global.
- α_i es el efecto adicional de la dieta i en comparación con el promedio global.
- ϵ_{ij} son errores experimentales aleatorios con distribución normal, media 0 y varianza σ^2 . ($\epsilon_{ij} \sim N(0, \sigma^2)$).

En este caso, el modelo lineal es mucho más simple porque no se tiene en cuenta la covariable consumo. Es por ello que este modelo simple describe cómo las observaciones (y_{ij}) se componen del nivel medio común (μ), la contribución única de cada tratamiento (α_i), y la variabilidad no explicada (ϵ_{ij}).

2.3 Efecto de la dieta

Ajustando el consumo

Ajustar el modelo lineal

```
modelo_con <- lm(incremento ~ dieta + consumo, data = data)
```

```
summary(modelo_con)
```

```
##
## Call:
## lm(formula = incremento ~ dieta + consumo, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.994  -4.639  -2.314   5.756  15.214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.7416    18.2280   0.864  0.40141
## dieta2       -14.9446     5.2558  -2.843  0.01233 *
## dieta3         3.4201     5.3499   0.639  0.53228
## dieta4       -20.6428     5.8478  -3.530  0.00303 **
## consumo         1.0692     0.4038   2.648  0.01826 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.294 on 15 degrees of freedom
## Multiple R-squared:  0.6798, Adjusted R-squared:  0.5944
## F-statistic: 7.96 on 4 and 15 DF, p-value: 0.001192
```


En el caso de variables dummy que representan categorías (como las dietas), el término constante (o intercepto) es la estimación del valor medio de la variable de respuesta para la categoría de referencia. En este caso, la dieta 1 es la categoría de referencia. Entonces, el término constante representará la estimación del aumento de peso medio para la dieta 1. Las otras variables dummy (dieta 2, dieta 3 y dieta 4) indicarán cómo difieren las otras dietas en comparación con la dieta 1. En resumen, se puede considerar que la dieta 1 es equivalente al intercepto en este contexto específico.

Una vez explicado esto, podemos ver en el resumen del modelo que la estimación para la dieta 1 (intercepto) es de 15.74, y para la dieta 2 es de -14.94. Esto quiere decir que la dieta 1 hace que el peso incremente más que con la dieta 2. No debemos confundir que el estimado de la dieta 2 sea negativo con que disminuya el peso, si no que aumenta pero en menor medida que con la dieta 1, que es con la que se está comparando. Esto concuerda con la exploración inicial de los datos, en la que la dieta 1 mostraba mayor incremento de peso que la dieta 2.

Sin ajustar el consumo

Ajustar el modelo lineal

```
modelo_sin <- lm(incremento ~ dieta, data = data)
```

Resumen del modelo

```
summary(modelo_sin)
```

```
##
```

```
## Call:
```

```
## lm(formula = incremento ~ dieta, data = data)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -15.20  -6.45   1.30   6.45  15.00
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   63.000      4.351  14.480 1.3e-10 ***
```

```
## dieta2       -15.800      6.153  -2.568  0.0206 *
```

```
## dieta3        6.200      6.153   1.008  0.3286
```

```
## dieta4       -13.800      6.153  -2.243  0.0394 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 9.729 on 16 degrees of freedom
```

```
## Multiple R-squared:  0.5301, Adjusted R-squared:  0.4419
```

```
## F-statistic: 6.016 on 3 and 16 DF, p-value: 0.006055
```

En este caso, el estimado para la dieta 1 es de 63, mientras que para la dieta 2 es de -15.8. Como mencionamos previamente, esto indica que la dieta 1 está asociada con un mayor aumento de peso en comparación con la dieta 2.

Al comparar el estimado de la dieta 1 ajustando el modelo con y sin la covariable, notamos una diferencia sustancial en la estimación (15.74 vs 63). Esta disparidad probablemente se debe a la omisión del efecto del consumo calórico en el modelo sin covariable.

Como hemos observado en la exploración de datos, existe una relación positiva entre la covariable (consumo) y la variable de respuesta. Al no incluir esta covariable en el modelo, parte de la variabilidad debida al consumo se atribuye erróneamente a la variable dieta. Esto se refleja en un estimado más alto para la dieta 1, que sirve como referencia.

Al incorporar el consumo como una covariable en el modelo, estamos intentando aislar el efecto específico de la dieta en el aumento de peso, controlando o ajustando por las diferencias en el consumo entre las dietas. Este enfoque proporciona una estimación más precisa del efecto de la dieta, eliminando el posible sesgo en la relación entre la covariable y la variable de respuesta.

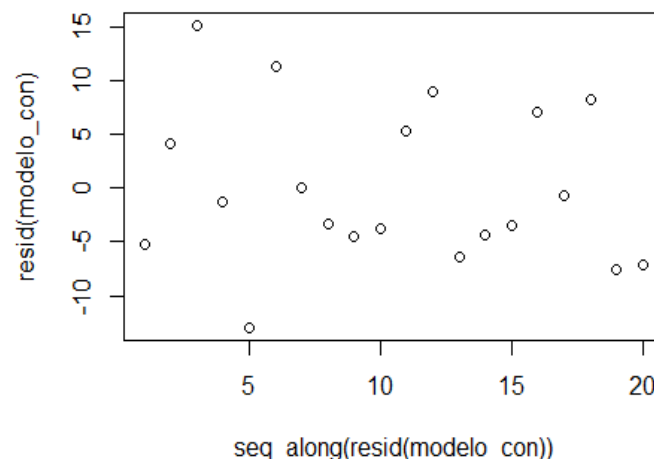
En resumen, al incluir el consumo como covariable, se mejora la capacidad del modelo para discernir entre los efectos de la dieta y el consumo en el aumento de peso, lo que resulta en estimaciones más ajustadas y potencialmente más precisas.

2.4 Diagnóstico de suposiciones

Independencia de los residuos

Gráfico de residuos vs. orden de observaciones

```
plot(resid(modelo_con) ~ seq_along(resid(modelo_con)))
```



```
library(lmtest)
```

Prueba de Durbin-Watson para autocorrelación de Los residuos

```
dwtest(modelo_con)
```

```
##
```

```
## Durbin-Watson test
```

```
##
```

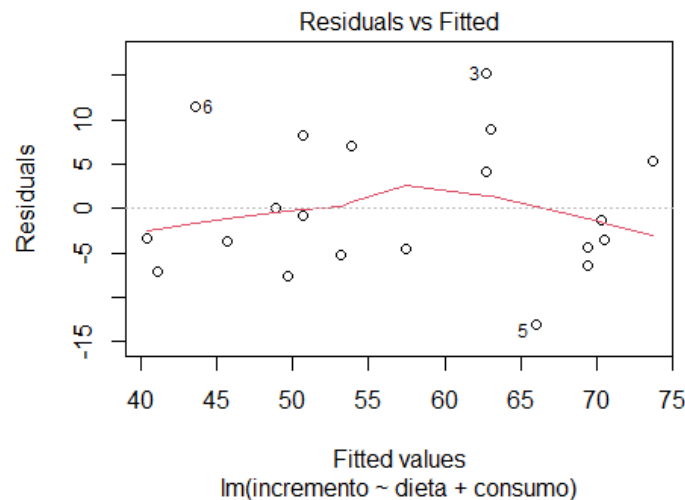
```
## data: modelo_con
```

```
## DW = 2.1341, p-value = 0.3188
## alternative hypothesis: true autocorrelation is greater than 0
```

En la representación gráfica, se observa la ausencia de patrones evidentes en los residuos. Este hallazgo se corrobora mediante la aplicación del test de Durbin-Watson, el cual indica la inexistencia de correlaciones significativas entre los residuos (valor de $p > 0.05$, aceptamos H_0).

Homocedasticidad

```
plot(modelo_con, which=1)
```



Prueba de Breusch-Pagan para homocedasticidad

```
library(lmtest)
bptest(modelo_con)

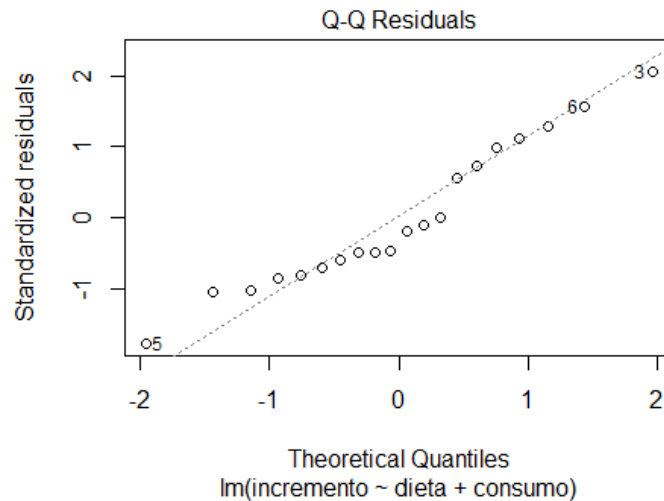
##
## studentized Breusch-Pagan test
##
## data:  modelo_con
## BP = 2.9601, df = 4, p-value = 0.5645
```

La dispersión de los residuos es constante a medida que cambian los valores ajustados. Vemos una dispersión uniforme de puntos alrededor de cero, indicando homocedasticidad.

Por otro lado, la prueba de Breusch-Pagan testa la H_0 de homocedasticidad. Como obtenemos un pvalor alto (mayor que 0.05) sugiere que no hay suficiente evidencia para rechazar la hipótesis nula, indicando homocedasticidad.

Normalidad

```
plot(modelo_con, which=2)
```



```
shapiro.test(modelo_con$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo_con$residuals
## W = 0.95099, p-value = 0.3824
```

Tanto a partir del QQ-Plot (puntos alrededor de la diagonal), como el test de Shapiro-Wilks (p-valor > 0,05), concluimos que los residuos son normales.

Interacción tratamiento-consumo

```
modelo_interaccion <- lm(incremento ~ consumo * dieta, data = data)
Anova(modelo_interaccion)
```

```
## Anova Table (Type II tests)
##
## Response: incremento
##
```

	Sum Sq	Df	F value	Pr(>F)	
consumo	482.42	1	6.5688	0.02487	*
dieta	1923.83	3	8.7318	0.00241	**
consumo:dieta	150.68	3	0.6839	0.57879	
Residuals	881.30	12			

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se advierte que la interacción entre el consumo y la dieta no alcanza significancia estadística.

La observación de estos parámetros confirma que el modelo satisface las suposiciones necesarias, permitiendo así la continuación del análisis sin inconvenientes.

2.5 Diferencias entre dietas

Para determinar entre qué dietas hay diferencias significativas, se deben realizar pruebas de comparaciones múltiples.

```
# Instalar y cargar el paquete multcomp
# install.packages("multcomp")
library(multcomp)

# Realizar comparaciones múltiples con el método de Tukey
comp <- glht(modelo_con, linfct = mcp(dieta = "Tukey"))

# Resumen de las comparaciones con ajuste de pvalores
resumen_comp <- summary(comp)
pvalores_ajustados <- p.adjust(resumen_comp$test$pvalues, method =
"bonferroni")
resumen_comp$test$pvalues_ajustados <- pvalores_ajustados

# Mostrar el resumen con pvalores ajustados
print(resumen_comp)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = incremento ~ dieta + consumo, data = data)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## 2 - 1 == 0  -14.945      5.256  -2.843  0.05334 .
## 3 - 1 == 0   3.420      5.350   0.639  0.91728
## 4 - 1 == 0  -20.643      5.848  -3.530  0.01429 *
## 3 - 2 == 0   18.365      5.423   3.387  0.01894 *
## 4 - 2 == 0   -5.698      5.998  -0.950  0.77807
## 4 - 3 == 0  -24.063      5.466  -4.403  0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Se efectuaron comparaciones múltiples mediante el método de Tukey, y tras ajustar los valores de p mediante el procedimiento de Bonferroni, se constata la presencia de diferencias estadísticamente significativas entre las dietas 4 y 1, entre las dietas 3 y 2, y, finalmente, entre las dietas 4 y 3.

Problema 3

Cargamos los datos del problema.

```
# Seteamos el directorio de trabajo
setwd("D:/Antiguos estudios/MASTER2/Sem3/Diseño/PAC6/PAC6")

# Importamos Los datos
data <- read.csv("dades_problema3_pec6.csv", dec = ".", header = TRUE, sep =
",")

# Marcamos Los factores
data$tratamiento <- as.factor(data$tratamiento)
data$paciente <- as.factor(data$paciente)
```

3.1 Análisis descriptivo

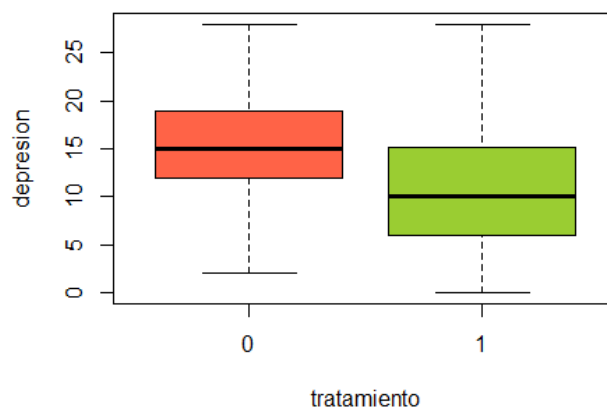
```
# Realiza un resumen descriptivo
summary(data)
```

```
## tratamiento    paciente    depresion      tiempo
## 0:175          1          : 7    Min.   : 0.00    Min.   :0
## 1:175          2          : 7    1st Qu.: 8.00    1st Qu.:1
##              3          : 7    Median :13.00    Median :3
##              4          : 7    Mean   :13.21    Mean   :3
##              5          : 7    3rd Qu.:18.00    3rd Qu.:5
##              6          : 7    Max.   :28.00    Max.   :6
##              (Other):308
```

```
color_tratamiento = c("tomato1", "olivedrab3")
```

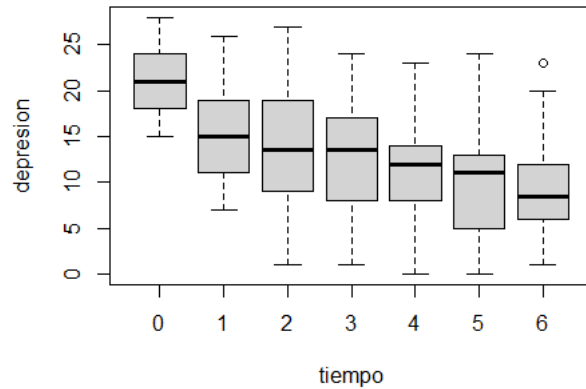
```
# Boxplot depresión por dieta
```

```
boxplot(depresion ~ tratamiento, data = data, col = color_tratamiento)
```



```
# Boxplot depresión por tiempo
```

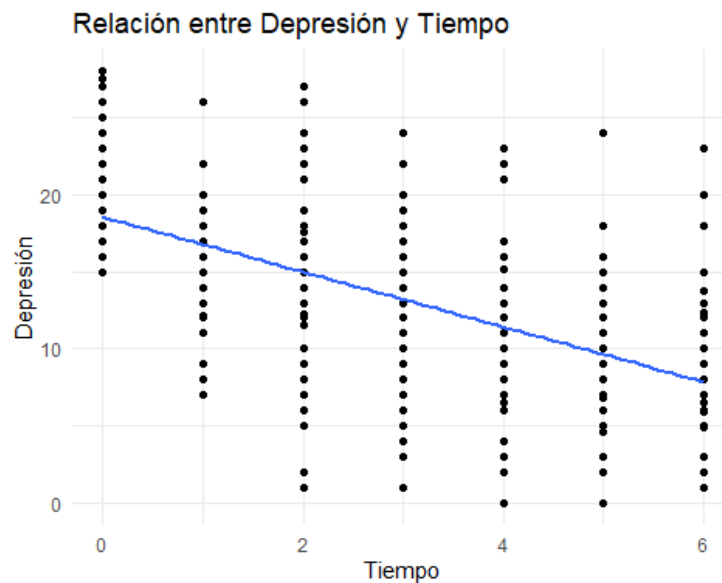
```
boxplot(depresion ~ tiempo, data = data)
```



```
library(ggplot2)

# Gráfico de dispersión depresión vs tiempo
ggplot(data, aes(x = tiempo, y = depression)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # Añade la línea de tendencia
  general
  labs(title = "Relación entre Depresión y Tiempo",
       x = "Tiempo",
       y = "Depresión") +
  theme_minimal()

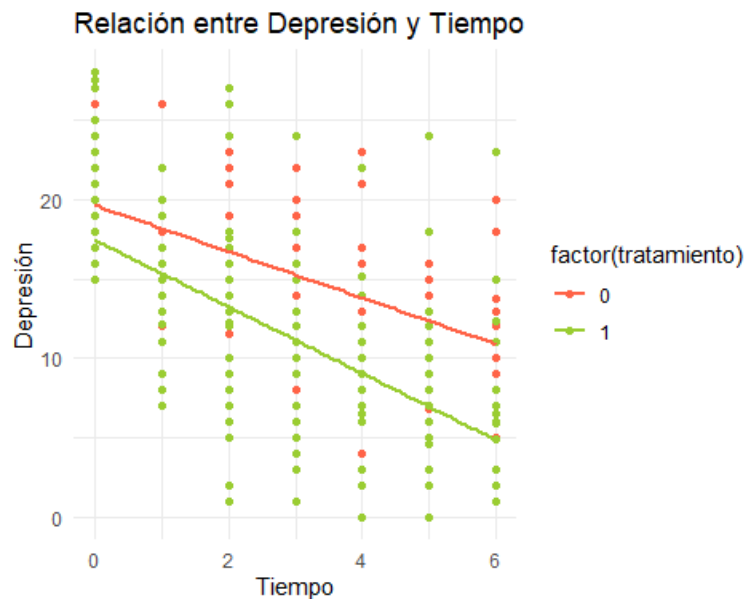
## `geom_smooth()` using formula = 'y ~ x'
```



```
# Gráfico de dispersión depresión vs tiempo con tratamiento
ggplot(data, aes(x = tiempo, y = depression, color = factor(tratamiento))) +
  scale_color_manual(values = color_tratamiento) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) + # Añade la línea de tendencia +
```

```
labs(title = "Relación entre Depresión y Tiempo",
     x = "Tiempo",
     y = "Depresión") +
theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Con base en la observación del diagrama de cajas, se sugiere que el tratamiento 1 exhibe una disminución en la prevalencia de la depresión en comparación con el tratamiento 0.

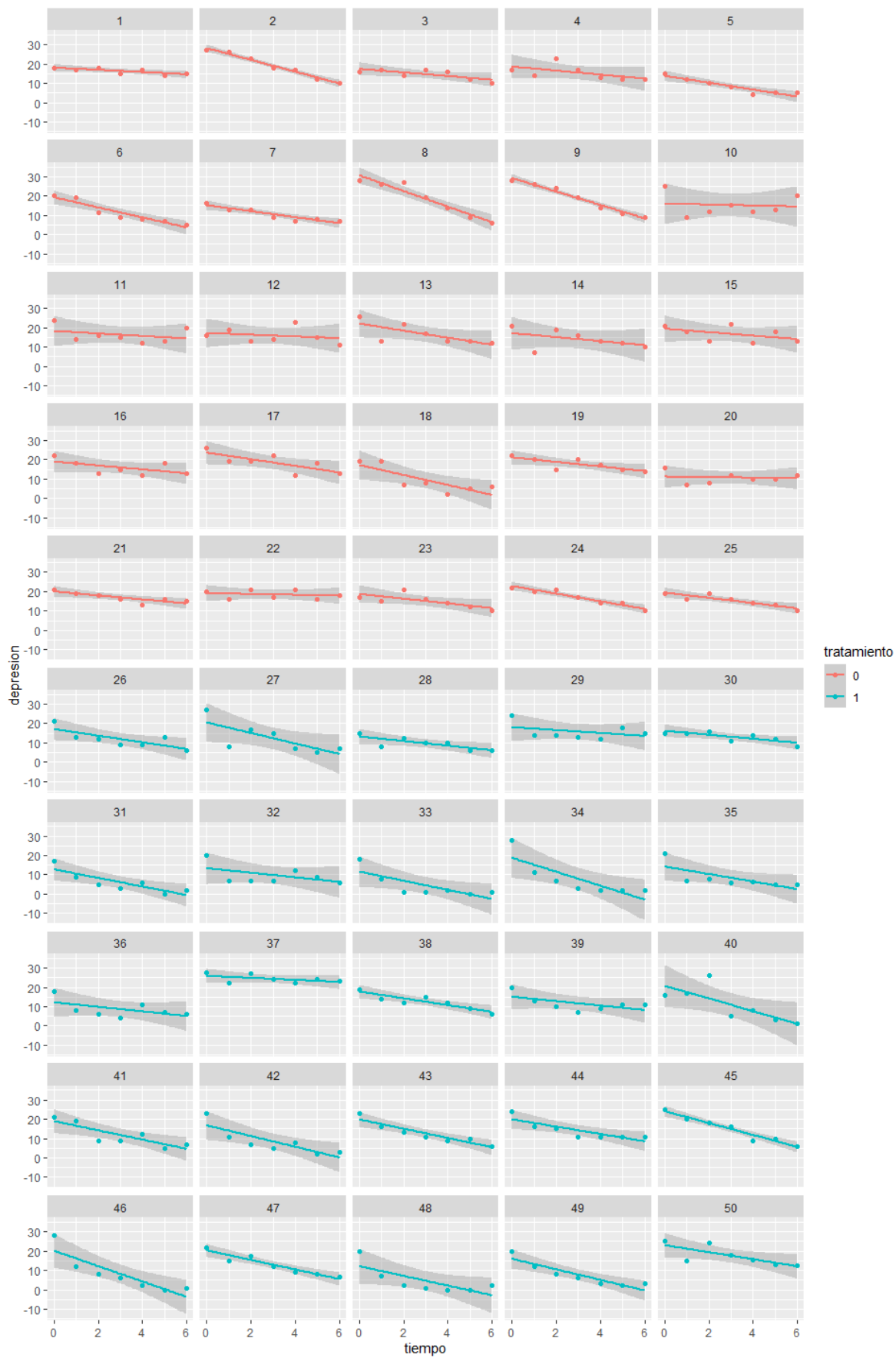
En relación al segundo diagrama de cajas, se aprecia una aparente reducción en la depresión posparto conforme transcurre el tiempo, manifestada por una tendencia descendente en las medianas representadas por las cajas.

Este hallazgo se corrobora al analizar el gráfico de dispersión, donde la línea de tendencia general, identificada en azul, presenta una disminución gradual a lo largo del período de tiempo estudiado.

Al segmentar los resultados según el tipo de tratamiento, se sugiere que el tratamiento exhibe una pendiente descendente ligeramente más pronunciada que la del tratamiento 0.

```
ggplot(data, aes(x = tiempo, y = depresion), color = tratamiento) +
  geom_smooth(method = "lm", level = 0.95) +
  geom_point() +
  facet_wrap(~ paciente, nrow = 10, ncol = 5) +
  theme(
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Finalmente, al examinar la evolución temporal de la depresión por paciente, se evidencia que algunos individuos experimentan una reducción marcada, mientras que en otros casos la evolución es mínima, ilustrada por líneas de tendencia notablemente planas (por ejemplo, paciente 2 vs. paciente 10). Es relevante destacar que en ningún caso se observa una tendencia ascendente, indicativa de un aumento en la prevalencia de la depresión con el paso del tiempo.

3.2 Modelo lineal

$$Y_{ij} = \beta_0 + \beta_1 * tratamiento_i + \beta_2 * tiempo_{ij} + \beta_3 * (tratamiento_i \times tiempo_{ij}) + u_{i0} + u_{il} * tiempo_{ij} + \epsilon_{ij}$$

Donde:

- Y_{ij} es la medida de depresión para la i-ésima paciente en la j-ésima visita.
- $tratamiento_i$ es la variable indicadora para el tratamiento (1 si es activo, 0 si es placebo). Es decir, la "i" en $tratamiento_i$ no indica el índice del tratamiento si no el número de paciente, el propio " $tratamiento_i$ " es indicador de tratamiento.
- $tiempo_{ij}$ es el tiempo de la j-ésima visita para la i-ésima paciente.
- β_0 es la intercepción fija del modelo.
- β_1 es el efecto fijo del tratamiento.
- β_2 es el efecto fijo del tiempo.
- β_3 representa la interacción entre el tratamiento y el tiempo.
- u_{i0} es el término aleatorio que modela las diferencias individuales en las intercepciones.
- u_{il} es el término aleatorio que modela las diferencias individuales en las pendientes respecto al tiempo.
- ϵ_{ij} es el término de error.

Cabe destacar que $i=50$ porque hay 50 pacientes y $j=7$ porque se mide la depresión en 7 visitas.

Este modelo puede expresarse en notación matricial de la siguiente manera:

$$Y = X\beta + Zu + \epsilon$$

Donde:

- Y : Matriz (50×7) que contiene las medidas de depresión para cada paciente en cada visita.
- X : Matriz (350×4) de diseño para los efectos fijos, que incluye una columna de unos para la intercepción, $tratamiento_i$, $tiempo_{ij}$, y $tratamiento_i \times tiempo_{ij}$.
- β : Vector (4×1) que contiene los coeficientes fijos $\beta_0, \beta_1, \beta_2, \beta_3$.
- Z : Matriz (350×2) de regresión de efectos aleatorios que contiene las variables aleatorias u_{i0} y $u_{il} \times tiempo_{ij}$.
- u : Vector (2×1) que contiene los términos aleatorios u_{i0} y u_{il} .

- ϵ : Matriz $(50 \times 7)(50 \times 7)$ que contiene los términos de error ϵ_{ij}

La matriz de efecto fijos X sería construída de la siguiente manera:

$$X = \begin{bmatrix} 1 & \text{tratamiento}_1 & \text{tiempo}_{11} & \text{tratamiento}_1 \times \text{tiempo}_{11} \\ 1 & \text{tratamiento}_2 & \text{tiempo}_{21} & \text{tratamiento}_2 \times \text{tiempo}_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{tratamiento}_{50} & \text{tiempo}_{57} & \text{tratamiento}_{50} \times \text{tiempo}_{57} \end{bmatrix}$$

La matriz de efectos aleatorios Z sería construída de esta otra manera:

$$Z = \begin{bmatrix} 1 & \text{tiempo}_{11} \\ 1 & \text{tiempo}_{21} \\ \vdots & \vdots \\ 1 & \text{tiempo}_{57} \end{bmatrix}$$

Adicionalmente, los términos aleatorios u_{i0} y u_{il} se asume tienen una distribución conjunta normal con media cero y una matriz de covarianza. La asunción de normalidad y la estructura de covarianza permiten modelar la variabilidad entre los pacientes de una manera más flexible y realista. Estos términos aleatorios capturan las diferencias individuales en las intercepciones (u_{i0}) y las pendientes respecto al tiempo (u_{il}) entre los pacientes.

La matriz de covarianza para los términos aleatorios se puede representar como:

$$\text{Cov}(u) = \begin{bmatrix} \sigma_{u0}^2 & \text{Cov}(u_{i0}, u_{il}) \\ \text{Cov}(u_{il}, u_{i0}) & \sigma_{ul}^2 \end{bmatrix}$$

Donde:

- σ_{u0}^2 es la varianza de u_{i0} .
- σ_{ul}^2 es la varianza de u_{il} .
- $\text{Cov}(u_{i0}, u_{il})$ y $\text{Cov}(u_{il}, u_{i0})$ son las covarianzas entre u_{i0} y u_{il} .

3.3 Implementación del modelo

Intercepciones aleatorias

Con el propósito de ajustar el presente modelo, se ha definido que la variable dependiente es la depresión, mientras que las variables independientes son el tratamiento y el tiempo. La inclusión del término “tratamiento * tiempo” se justifica por la posibilidad de que existan efectos diferenciados del tiempo según el tipo de tratamiento, dando cabida a la

interacción entre ambas variables. Además, la especificación “(1|paciente)” indica que se está modelando un intercepto aleatorio para cada paciente, con el fin de tener en consideración la variabilidad existente entre los distintos pacientes en el análisis.

```
# install.packages("lme4", dependencies=TRUE)
# install.packages("Matrix", dependencies=TRUE)
library(lme4)

# Define el modelo con intercepciones aleatorias
modelo1 <- lmer(depresion ~ tratamiento * tiempo + (1|paciente), data = data)

# Muestra un resumen del modelo
summary(modelo1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: depresion ~ tratamiento * tiempo + (1 | paciente)
## Data: data
##
## REML criterion at convergence: 2022.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6485 -0.6353 -0.0297  0.5900  3.4933
##
## Random effects:
##  Groups   Name      Variance Std.Dev.
## paciente (Intercept) 11.56    3.400
## Residual          14.58    3.818
## Number of obs: 350, groups:  paciente, 50
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    19.6417    0.8563  22.939
## tratamiento1    -2.1815    1.2109  -1.802
## tiempo         -1.4595    0.1443 -10.114
## tratamiento1:tiempo -0.6413    0.2041  -3.142
##
## Correlation of Fixed Effects:
##              (Intr) trtmn1 tiempo
## tratamiento1 -0.707
## tiempo       -0.506  0.357
## trtmnt1:tmp  0.357 -0.506 -0.707
```

Intercepciones y pendientes aleatorias

Para llevar a cabo la especificación de este modelo, se ha designado la variable de respuesta como la medida de depresión, considerando tanto el tratamiento como el tiempo como variables independientes. La inclusión del término “tratamiento * tiempo” se justifica al incorporar la interacción entre tratamiento y tiempo, considerando la posibilidad de que existan efectos temporales diferenciados dependiendo del tipo de tratamiento.

Adicionalmente, la especificación “(tiempo | paciente)” establece la modelización de una intercepción aleatoria y una pendiente aleatoria para cada paciente. Este término implica que cada individuo podría presentar una intercepción (efecto inicial) y una pendiente (efecto en el tiempo) propias con respecto a la variable de respuesta.

Cabe señalar que se ha optado por no incluir variaciones aleatorias en la variable tratamiento (tratamiento | paciente). Esta decisión se fundamenta en la consideración de que permitiría que la relación entre la depresión y el tratamiento varíe aleatoriamente entre pacientes. En el contexto de la problemática abordada, donde se está evaluando un fármaco, no resulta deseable que los pacientes reaccionen de manera diversa al tratamiento. Se parte del supuesto de que el fármaco, al ser considerado para su comercialización, debería inducir un efecto uniforme en todos los individuos, aunque se reconoce que la respuesta de cada paciente puede diferir a lo largo del tiempo.

```
library(lme4)

# Define el modelo con intercepciones y pendientes aleatorias
modelo2 <- lmer(depresion ~ tratamiento * tiempo + (tiempo | paciente), data
= data)

# Muestra un resumen del modelo
summary(modelo2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: depresion ~ tratamiento * tiempo + (tiempo | paciente)
##   Data: data
##
## REML criterion at convergence: 2008.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7720 -0.6034 -0.0424  0.5988  3.5554
##
## Random effects:
##   Groups      Name              Variance Std.Dev. Corr
##   paciente (Intercept) 11.9724   3.4601
##               tiempo      0.4736   0.6882  -0.31
##   Residual              12.4404   3.5271
## Number of obs: 350, groups:  paciente, 50
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    19.6417    0.8426  23.312
## tratamiento1   -2.1815    1.1916  -1.831
## tiempo         -1.4595    0.1916  -7.617
## tratamiento1:tiempo -0.6413    0.2710  -2.367
##
## Correlation of Fixed Effects:
##              (Intr) trtmn1 tiempo
```

```
## tratamient1 -0.707
## tiempo      -0.511  0.361
## trtmnt1:tmp  0.361 -0.511 -0.707
```

Comparación de modelos

Comparación de modelos usando likelihood ratio test

```
anova(modelo1, modelo2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
```

```
## Models:
```

```
## modelo1: depresion ~ tratamiento * tiempo + (1 | paciente)
```

```
## modelo2: depresion ~ tratamiento * tiempo + (tiempo | paciente)
```

```
##      npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
```

```
## modelo1      6 2033.0 2056.1 -1010.5    2021.0
```

```
## modelo2      8 2023.7 2054.6 -1003.9    2007.7 13.242  2  0.001332 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Según la prueba de razón de verosimilitud, se observa una diferencia significativa entre los dos modelos evaluados (valor de $p < 0.05$). En virtud de este hallazgo, se concluye que el modelo más complejo es preferible, ya que demuestra una mejor capacidad para explicar la variabilidad inherente a la variable de respuesta.

Asimismo, esta prueba proporciona los valores del Criterio de Información de Akaike (AIC) y del Criterio de Información Bayesiano (BIC). Estas medidas evalúan la calidad del ajuste de un modelo, considerando simultáneamente su complejidad. De manera general, se consideran más deseables aquellos modelos que presentan valores más bajos de AIC y BIC.

El AIC penaliza la complejidad del modelo de manera menos severa que el BIC. Un valor más bajo de AIC indica un mejor equilibrio entre ajuste y parsimonia.

En contraste, el BIC penaliza la complejidad del modelo de forma más estricta que el AIC. Una disminución en el valor de BIC sugiere un mejor balance entre el ajuste del modelo y su complejidad.

Con base en esta información, se concluye que el modelo que incorpora intercepciones y pendientes aleatorias es preferible al modelo que únicamente considera intercepciones aleatorias.

3.4 Factores significativos

```
summary(modelo1)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: depresion ~ tratamiento * tiempo + (1 | paciente)
```

```
## Data: data
```

```
##
```

```
## REML criterion at convergence: 2022.6
```

```
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6485 -0.6353 -0.0297  0.5900  3.4933
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
## paciente (Intercept) 11.56      3.400
## Residual              14.58      3.818
## Number of obs: 350, groups: paciente, 50
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      19.6417     0.8563  22.939
## tratamiento1      -2.1815     1.2109  -1.802
## tiempo            -1.4595     0.1443 -10.114
## tratamiento1:tiempo -0.6413     0.2041  -3.142
##
## Correlation of Fixed Effects:
##              (Intr) trtmn1 tiempo
## tratamient1 -0.707
## tiempo      -0.506  0.357
## trtmnt1:tmp  0.357 -0.506 -0.707
```

Una vez más, debemos calcular los intervalos de confianza para determinar la significancia de los factores.

```
# install.packages("broom")
library(broom)

# Intervalos de confianza
confint(modelo1)

## Computing profile confidence intervals ...

##              2.5 %      97.5 %
## .sig01          2.641929  4.2261757
## .sigma          3.519829  4.1311528
## (Intercept)     17.971968 21.3114031
## tratamiento1    -4.542851  0.1798227
## tiempo          -1.742244 -1.1766700
## tratamiento1:tiempo -1.041207 -0.2413643
```

Los factores que adquieren significancia se identifican a través de la observación de intervalos de confianza que no incorporan el valor 0.

Se constata que únicamente el intervalo de confianza correspondiente al tratamiento 1 incluye el valor 0. Esto indica que los restantes factores (tiempo e interacción tiempo:tratamiento) son estadísticamente significativos, mientras que el tratamiento no lo es.

Este hallazgo plantea perspectivas poco alentadoras para el estudio clínico, ya que sugiere que el tratamiento no ejerce un efecto discernible sobre la depresión. En cambio, parece que la disminución de la depresión ocurre de forma natural a lo largo del tiempo en las pacientes, independientemente de la administración del fármaco.

No obstante, la presencia de una interacción significativa sugiere que la relación entre el tratamiento y la respuesta de depresión experimenta cambios significativos a lo largo del tiempo. En otras palabras, el impacto del tratamiento podría variar en diferentes momentos del estudio. Esta dinámica podría ser objeto de un examen más detenido: es posible que el tratamiento no surta un efecto inmediato, sino que requiera un período de tiempo para manifestarse. O bien, podría indicar que el tratamiento ejerce un efecto destacado en ciertos momentos específicos (por ejemplo, períodos de tiempo particulares después de la administración), pero no en otros. Alternativamente, el tratamiento podría influir de manera dispar en la respuesta de depresión en distintos momentos del estudio, y estas variaciones a lo largo del tiempo podrían amalgamarse de tal manera que el efecto general del tratamiento no alcance significancia.

Para obtener mayor claridad sobre estos efectos, se podrían llevar a cabo análisis exploratorios y post hoc adicionales.

3.5 Modelo estimado en las pacientes 1 y 26

```
# Coeficientes fijos del modelo para todos los pacientes
coeficientes_fijos <- fixef(modelo1)
```

```
# Efectos aleatorios para los pacientes 1 y 26
efectos_aleatorios <- ranef(modelo1)$paciente
efecto_paciente1 <- efectos_aleatorios[1, ]
efecto_paciente26 <- efectos_aleatorios[26, ]
```

```
# Coeficientes para el paciente 1 y 26
coef_paciente1 <- coeficientes_fijos + efecto_paciente1
coef_paciente26 <- coeficientes_fijos + efecto_paciente26
```

```
# Impresión de coeficientes
```

```
cat("Coeficientes fijos para todos los pacientes:")
```

```
## Coeficientes fijos para todos los pacientes:
```

```
coeficientes_fijos
```

```
##           (Intercept)           tratamiento1           tiempo
tratamiento1:tiempo
##           19.6416857           -2.1815143           -1.4594571           -
0.6412857
```

```
cat("\n")
```

```
cat("Efectos aleatorios para el paciente 1:")
```

```
## Efectos aleatorios para el paciente 1:
```



```

coef(modelo1)$paciente["1",]

## (Intercept) tratamiento1 tiempo tratamiento1:tiempo
## 1 20.50805 -2.181514 -1.459457 -0.6412857

cat("\n")

cat("Efectos aleatorios para el paciente 26:")

## Efectos aleatorios para el paciente 26:

coef(modelo1)$paciente["26",]

## (Intercept) tratamiento1 tiempo tratamiento1:tiempo
## 26 20.23418 -2.181514 -1.459457 -0.6412857

cat("\n")

cat("Coeficientes aleatorios para el paciente 1:")

## Coeficientes aleatorios para el paciente 1:

coef_paciente1

## (Intercept) tratamiento1 tiempo
tratamiento1:tiempo
## 20.5080543 -1.3151457 -0.5930886
0.2250829

cat("\n")

cat("Coeficientes para el paciente 26:")

## Coeficientes para el paciente 26:

coef_paciente26

## (Intercept) tratamiento1 tiempo
tratamiento1:tiempo
## 20.23417878 -1.58902122 -0.86696408 -
0.04879265

cat("Coeficientes para el paciente 1:")

## Coeficientes para el paciente 1:

coef_paciente1

## (Intercept) tratamiento1 tiempo
tratamiento1:tiempo
## 20.5080543 -1.3151457 -0.5930886
0.2250829

cat("\n")

```

```
cat("Coeficientes para el paciente 26:")
## Coeficientes para el paciente 26:
coef_paciente26
##           (Intercept)          tratamiento1          tiempo
tratamiento1:tiempo
##           20.23417878          -1.58902122          -0.86696408          -
0.04879265
```

Al extraer los coeficientes fijos sabemos que:

- $\beta_0 = 19.64$
- $\beta_1 = -2.18$
- $\beta_2 = -0.59$
- $\beta_3 = 0.22$

Si examinamos estos coeficientes comprendemos como varía la depresión según cada factor.

En cuánto a los efectos aleatorios para los pacientes 1 y 26, estos nos dan la información de cómo difieren del promedio en términos de intercepción y pendiente. Los coeficientes específicos para cada paciente combinan estos efectos aleatorios con los coeficientes fijos para proporcionar estimaciones específicas para esos pacientes.

Con los coeficientes específicos de cada paciente podemos escribir sus modelos específicos, para el paciente 1:

$$Y_{1j} = 20.50 + -1.31 * tratamiento_{1j} + -0.59 * tiempo_{1j} + 0.22 * (tratamiento_{1j} \times tiempo_{1j}) + u_{10} + u_{11} * tiempo_{1j} + \epsilon_{1j}$$

Para el paciente 26:

$$Y_{26j} = 20.23 + -1.58 * tratamiento_{26j} + -0.86 * tiempo_{26j} + -0.04 * (tratamiento_{26j} \times tiempo_{26j}) + u_{260} + u_{261} * tiempo_{26j} + \epsilon_{26j}$$

La intercepción inicial es ligeramente mayor para el paciente 1 que para el paciente 26, indicando que estaba inicialmente más deprimida.

El paciente 26 parece tener un efecto de tratamiento más negativo en comparación con el paciente 1. Esto indica que le hace más efecto el fármaco al paciente 26, ya que le influye de mayor manera en la variable respuesta.

Ambos pacientes muestran una disminución en la medida de depresión a lo largo del tiempo, pero el paciente 26 tiene una pendiente más pronunciada.

La interacción entre tratamiento y tiempo tiene un impacto positivo para el paciente 1 y un impacto apenas negativo para el paciente 26.

Los términos aleatorios representan las diferencias individuales en las intercepciones y en las pendientes respecto al tiempo .

Los valores específicos de estos términos aleatorios indicarán cómo cada paciente difiere del promedio en términos de su intercepción y pendiente respecto al tiempo.