

Regresión, ANOVA y ANCOVA

Tiempos de coagulación

En cierto estudio, 15 animales fueron asignados al azar a tres tipos de dieta distintos. Tras varias semanas siguiendo la dieta, se midió el tiempo de coagulación de la sangre de cada animal (en segundos). Haz un análisis ANOVA seguido de un análisis post-hoc para averiguar si la dieta influye en los tiempos de coagulación.

En este ejercicio hay una primera dificultad: el csv no está en el formato adecuado, por lo que hay que arreglarlo:

```
# mmmmm... formato erróneo
coag_diet = read_csv("data_reg/coag_diet.csv", show_col_types = FALSE)
print(coag_diet)

## # A tibble: 3 x 6
##   ...1      V1      V2      V3      V4      V5
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 A      60.0  59.3  58.0  59.6  62.0
## 2 B      61.6  60.2  61.6  61.9  61.9
## 3 C      62.5  58.6  62.4  62.8  61.3

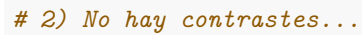
# Extrae los números del fichero como una matriz y usa la primera columna
# como "rownames"
coag_diet_data = as.matrix(coag_diet[, 2:ncol(coag_diet)])
rownames(coag_diet_data) = coag_diet$...1
colnames(coag_diet_data) = NULL

# Esto crear un data.frame correcto pero repite código. ¿Podrías hacerlo mejor?
coag_diet_long = rbind(
  data.frame(
    class = "A",
    coagulation = coag_diet_data["A", ]
  ),
  data.frame(
    class = "B",
    coagulation = coag_diet_data["B", ]
  ),
  data.frame(
    class = "C",
    coagulation = coag_diet_data["C", ]
  )
)
print(head(coag_diet_long))

##   class coagulation
## 1    A    59.96384
## 2    A    59.30779
```

Seguimos ahora los pasos habituales:

```
ggplot(coag_diet_long, aes(x=class, y = coagulation)) + geom_boxplot()
```



```
coag_lm = lm(coagulation ~ class, coag_diet_long)
coag_aov = Anova(coag_lm, type = 3)
```

```
print(coag_aov)
```

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

```
effectsize(coag_aov)
```

```
## For one-way between subjects designs, partial eta squared is equivalent  
## to eta squared. Returning eta squared.
```

```
## # Effect Size for ANOVA
```

```
##
```

```
## Parameter | Eta2 | 95% CI
```

```
## -----
```

```
## class | 0.30 | [0.00, 1.00]
```

```
##
```

```
## - One-sided CIs: upper bound fixed at [1.00].
```

El efecto de la dieta no es estadísticamente significativo, lo que revela que no hay suficiente evidencia apoyando que los tres tipos de dietas sean distintos (si bien el tamaño del efecto es grande, $\eta^2 = 0.29$). Compara esta conclusión con la sugerencia del gráfico exploratorio. ¿A qué se debe esta diferencia? (Fíjate en el tamaño de las muestras y los outliers.)

Al no ser significativo el test ANOVA, no es necesario proceder a hacer el análisis post-hoc.

¿Influye el sexo en los salarios?

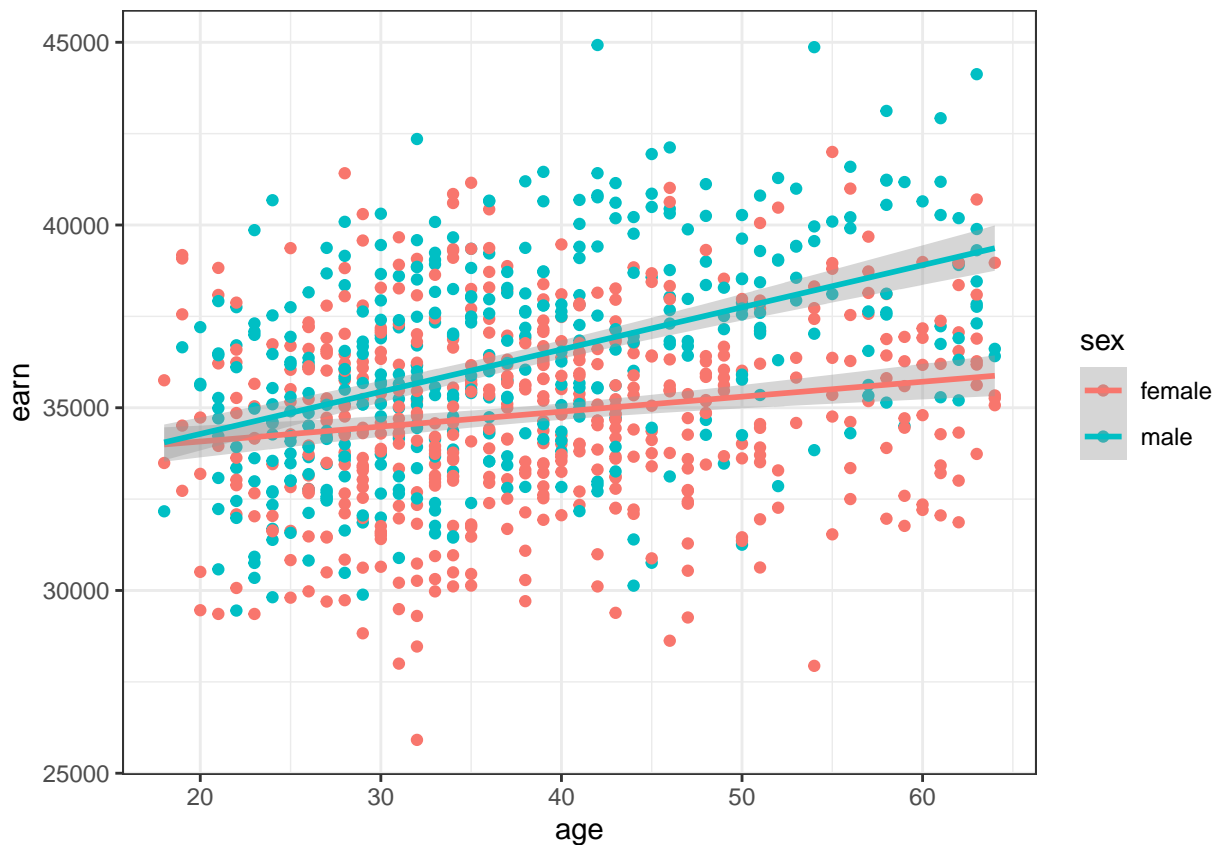
Cierto estudio económico quería evaluar 1) si hombres y mujeres ganan lo mismo y 2) si su progresión salarial es similar a lo largo de los años. Datos en “earnings.csv”. Emplea la visualización para decidir si existen interacciones entre la edad y el sexo. A continuación, realiza el análisis estadístico habitual.

```
earnings <- read_csv("data_reg/earnings.csv", show_col_types = FALSE)
```

```
# 1) Visualización: como una de las variables es continua, intentamos visualizar  
# la interacción usando un scatterplot
```

```
ggplot(earnings, aes(x = age, y = earn, col=sex)) +  
  geom_point() +  
  geom_smooth(method = "lm")
```

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



```
# Si hay interacción entre las variables age y sex.
```

```
# 2) No hay contrastes
```

```
# 3) Regresión (No tiene sentido ANOVA: ¿por qué?)
```

```
earn_model <- lm(earn ~ age * sex, earnings)
```

```
# 4) Analizar resultados
```

```
summary(earn_model)
```

```
##
## Call:
## lm(formula = earn ~ age * sex, data = earnings)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8654.2 -1677.8   -33.2   1731.2   8100.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33258.26    407.42   81.630 < 2e-16 ***
## age           40.89      10.14    4.031 6.01e-05 ***
## sexmale     -1284.86     595.97   -2.156  0.0314 *
## age:sexmale    74.60      15.00    4.972 7.89e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2551 on 921 degrees of freedom
## Multiple R-squared:  0.1803, Adjusted R-squared:  0.1777
## F-statistic: 67.55 on 3 and 921 DF,  p-value: < 2.2e-16
```

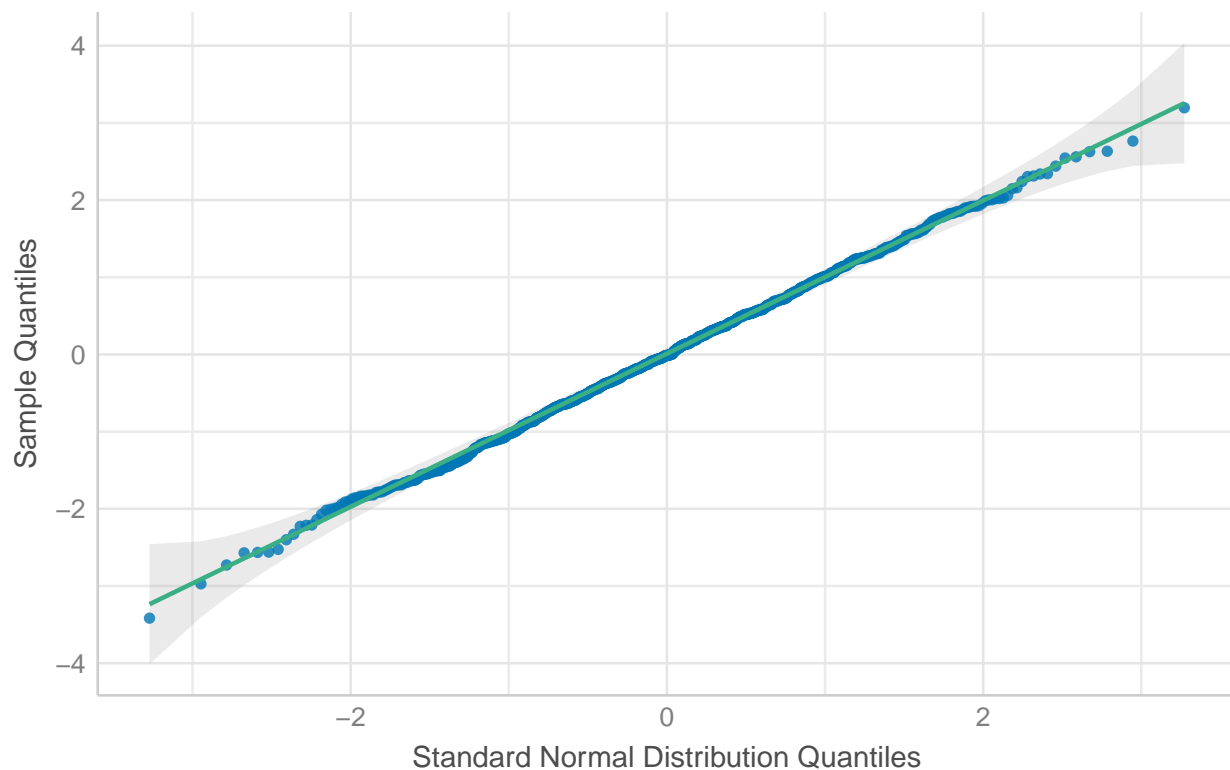
Como las interacciones son significativas no debemos intentar interpretar los coeficientes principales. El efecto de edad \times sexo [masculino] es estadísticamente significativo y positivo ($p < .001$) con un valor esperado de 74.60 (si bien el intervalo (45.2, 104.1) es igualmente plausible al 95% de confianza). Esto quiere decir que, por cada año adicional, un hombre ganará 74.60 dólares más que una mujer (las mujeres incrementan su salario a un ritmo de 40.9 dólares al año, mientras que los hombres incrementan su salario a un ritmo de 115.5 dólares al año).

Estas conclusiones son sólo válidas si las asunciones del modelo son correctas:

```
# 5) Comprobar asunciones: todo OK
plot(check_normality(earn_model), type = "qq")
```

Normality of Residuals

Dots should fall along the line



Joyzepam

(Danielle Navarro, 0.6.1) Te has visto involucrado en un ensayo clínico en el que se está probando un nuevo medicamento antidepresivo llamado Joyzepam. Con el fin de construir una prueba de la eficacia del fármaco, el estudio implica la administración de tres fármacos separados. Uno es un placebo y el otro es un medicamento antidepresivo/ansiolítico existente llamado Anxifree. Se recluta a 18 participantes con depresión moderada a severa para su prueba inicial. Debido a que los medicamentos a veces se administran junto con la terapia psicológica, el estudio incluye a 9 personas que se someten a terapia cognitiva conductual (CBT) y 9 que no la reciben. A los participantes se les asigna aleatoriamente a un tratamiento, de modo que haya 3 personas con CBT y 3 personas sin terapia asignadas a cada uno de los 3 medicamentos. Un psicólogo evalúa el estado de ánimo de cada persona después de 3 meses de tratamiento con cada medicamento y se evalúa la mejora general en el estado de ánimo de cada persona. Evalúa si hay evidencia para concluir (usando un nivel de

significación del 5%)....

1. ... si hay diferencias entre las drogas y el placebo.
2. ... si hay diferencias entre anxifree y joyzepam.
3. ... si hay diferencias entre la terapia CBT y no recibir terapia.

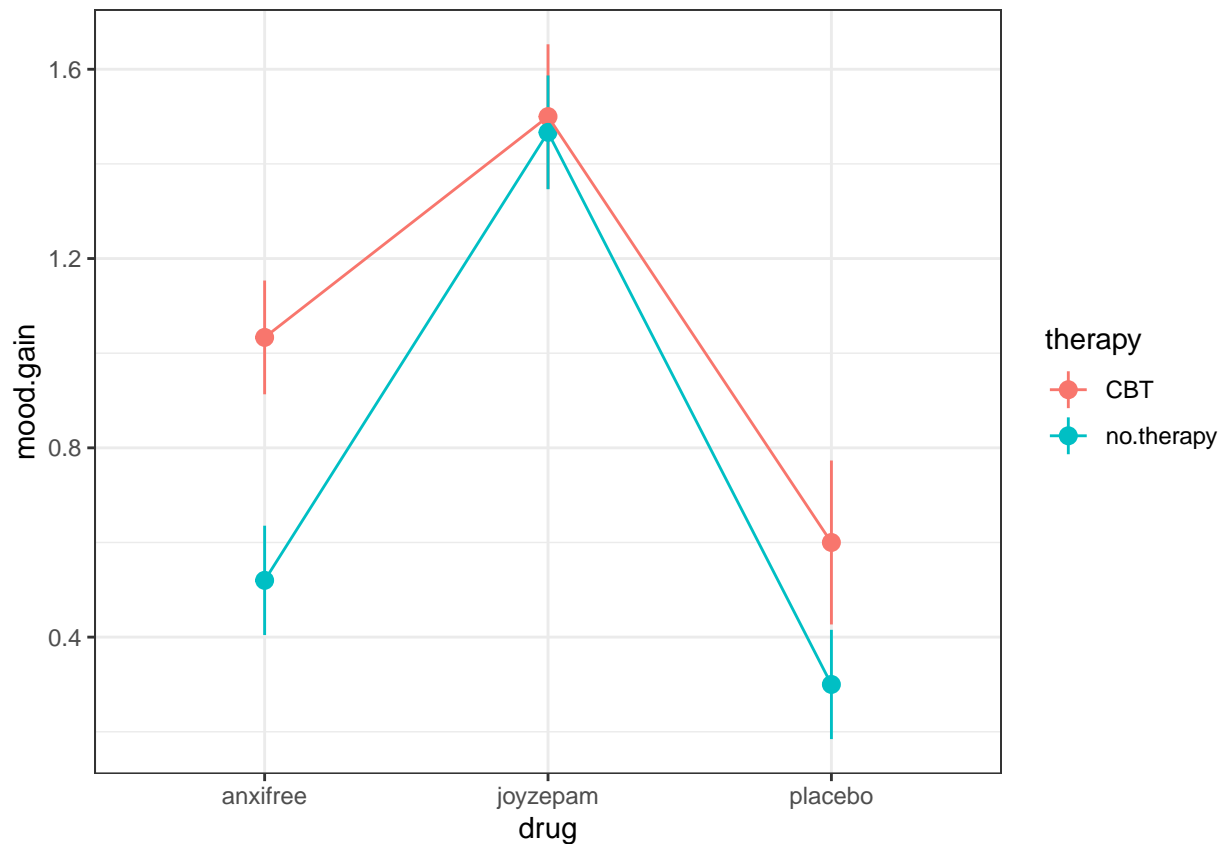
Datos en "joyzepam.csv".

```
joyzepam <- read_tsv("data_reg/joyzepam.csv", show_col_types = FALSE)
joyzepam$drug <- as.factor(joyzepam$drug)
joyzepam$therapy <- as.factor(joyzepam$therapy)

# 1) Visualización: como las dos variables son discretas, empleamos stat_summary
# para visualizar la interacción
ggplot(joyzepam, aes(x = drug, col = therapy, y = mood.gain)) +
  stat_summary() +
  stat_summary(geom = "line", aes(group = therapy))
```

No summary function supplied, defaulting to `mean_se()`

No summary function supplied, defaulting to `mean_se()`



No está claro si hay interacción o no, dados los errores (líneas verticales).
Por si acaso, incluiremos la interacción en el modelo.

```
# 2) Contrastes
contrasts(joyzepam$drug) <- get_contrasts_coding(
  rbind(
```

```

    " drugs Vs placebo" = c(0.5, 0.5, -1),
    " Anxifree Vs joyzepam" = c(1, -1, 0)
  )
)

# 3) Analisis de regresión (dado que tenemos contrastes, no es necesario hacer ANOVA)
joyzepam_lm <- lm(mood.gain ~ drug * therapy, joyzepam)

# 4) Analizar contrastes
summary(joyzepam_lm)

```

```

##
## Call:
## lm(formula = mood.gain ~ drug * therapy, data = joyzepam)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3000 -0.1917  0.0000  0.1917  0.3000
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   1.04444    0.07778  13.429
## drug drugs Vs placebo          0.66667    0.16499   4.041
## drug Anxifree Vs joyzepam      -0.46667    0.19052  -2.449
## therapyno.therapy             -0.28222    0.10999  -2.566
## drug drugs Vs placebo:therapyno.therapy  0.02667    0.23333   0.114
## drug Anxifree Vs joyzepam:therapyno.therapy -0.48000    0.26943  -1.782
##                                Pr(>|t|)
## (Intercept)                   1.37e-08 ***
## drug drugs Vs placebo          0.00164 **
## drug Anxifree Vs joyzepam      0.03062 *
## therapyno.therapy             0.02474 *
## drug drugs Vs placebo:therapyno.therapy  0.91090
## drug Anxifree Vs joyzepam:therapyno.therapy 0.10013
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2333 on 12 degrees of freedom
## Multiple R-squared:  0.8559, Adjusted R-squared:  0.7958
## F-statistic: 14.25 on 5 and 12 DF,  p-value: 0.0001079

```

```

confint(joyzepam_lm)

##                                2.5 %    97.5 %
## (Intercept)                   0.8749812  1.21390766
## drug drugs Vs placebo          0.3071809  1.02615244
## drug Anxifree Vs joyzepam      -0.8817651 -0.05156825
## therapyno.therapy             -0.5218794 -0.04256504
## drug drugs Vs placebo:therapyno.therapy -0.4817230  0.53505633
## drug Anxifree Vs joyzepam:therapyno.therapy -1.0670378  0.10703781

```

Dado que las interacciones no son significativas podemos interpretar los resultados principales (¡fíjate que no es necesario volver a crear un nuevo modelo sin las interacciones!).

El primer contraste revela que la diferencia entre las drogas y el placebo es significativa (p-valor:0.0016). Las

drogas mejoran, en termino medio, 0.67 puntos el estado anímico de los pacientes (si bien el intervalo (0.31, 1.03) es igualmente plausible al 95% de confianza).

El segundo contraste indica que la diferencia entre las dos drogas también es significativa (p-valor 0.031), siendo el joyzepam ligeramente mejor. Este mejora en 0.47 puntos los resultados de anxifree (aunque la mejora puede estar también entre 0.05 y 0.88, al 95% de confianza).

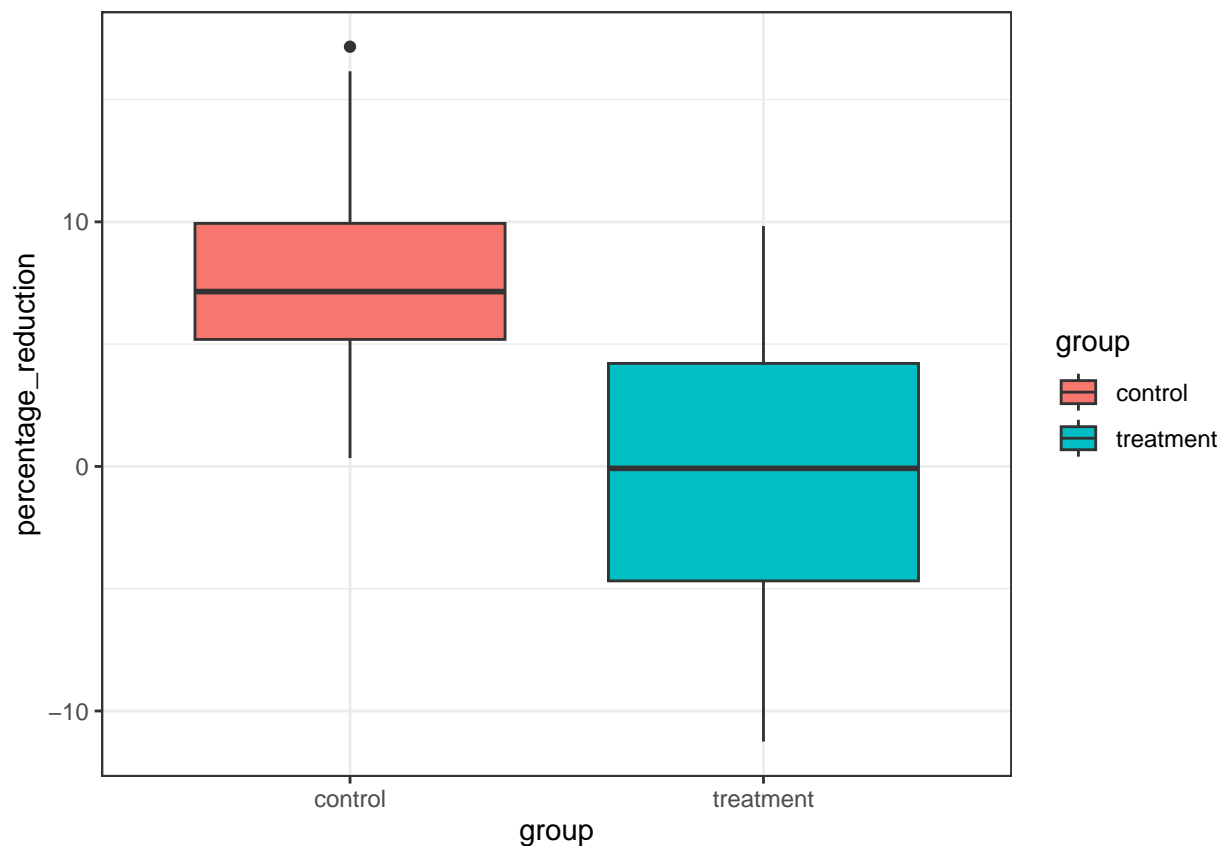
Finalmente, la terapia CBT es preferible a la no terapia (p-valor: 0.025), ya que incrementa en 0.28 puntos el nivel anímico de los pacientes (si bien el rango (0.04, 0.52) es igualmente plausible al 95%).

Colesterol

En “cholesterol.csv” se recogen los datos de un estudio para evaluar los efectos de cierta medicación en el colesterol. Para comprobar su efectividad, se dividió a los participantes en el estudio en dos grupos: control (que reciben una medicación falsa a la que nos referiremos como placebo) y tratamiento (que reciben el fármaco real). Realiza un Análisis ANOVA + Posthoc para comprobar si existe una dependencia entre la reducción en el colesterol y el tratamiento recibido (ignora el resto de variables en el data.frame).

```
cholesterol = read_csv("data_reg/cholesterol.csv", show_col_types = FALSE)
cholesterol$group = as.factor(cholesterol$group)

# 1) Visualización
ggplot(cholesterol, aes(x = group, y = percentage_reduction, fill = group)) +
  geom_boxplot()
```



```
# 2) Codificar contrastes (No hay)
```



```
# 3) Modelos lm y Anova
```

```
chol_lm = lm(percentage_reduction ~ group, cholesterol)
chol_aov = Anova(chol_lm)
```

```
# 4) Analizar contrastes o Anova + posthoc
```

```
print(chol_aov)
```

```
## Anova Table (Type II tests)
```

```
##
```

```
## Response: percentage_reduction
```

```
##           Sum Sq Df F value    Pr(>F)
```

```
## group      645.83  1  22.665 2.795e-05 ***
```

```
## Residuals 1082.78 38
```

```
## ---
```

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

```
print(effectsize(chol_aov))
```

```
## For one-way between subjects designs, partial eta squared is equivalent
##   to eta squared. Returning eta squared.
```

```
## # Effect Size for ANOVA
```

```
##
```

```
## Parameter | Eta2 |          95% CI
```

```
## -----
```

```
## group      | 0.37 | [0.18, 1.00]
```

```
##
```

```
## - One-sided CIs: upper bound fixed at [1.00].
```

El ANOVA sugiere que el efecto de la medicación es estadísticamente significativo $p < .001$; $\text{Eta}^2 = 0,37$). La diferencia en la reducción de colesterol es

```
# No se requiere ajuste de p-valores (¿Por qué?)
```

```
# Este resultado también podría obtenerse de summary(chol_lm) (¿Por qué?)
```

```
pairs(emmeans(chol_lm, "group"), infer = c(TRUE, TRUE), level = 0.95)
```

```
## contrast          estimate    SE df lower.CL upper.CL t.ratio p.value
```

```
## control - treatment      8.04 1.69 38      4.62      11.5   4.761 <.0001
```

```
##
```

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

Es decir, los pacientes de control pierden, en media, un 8.05% más que los pacientes que reciben el tratamiento (si bien el intervalo (-11.5%, -4.62%) es igualmente plausible, al 95% de confianza). Es decir, ¡El medicamento es totalmente ineficiente!

```
# 5) Verificar asunciones del modelo
```

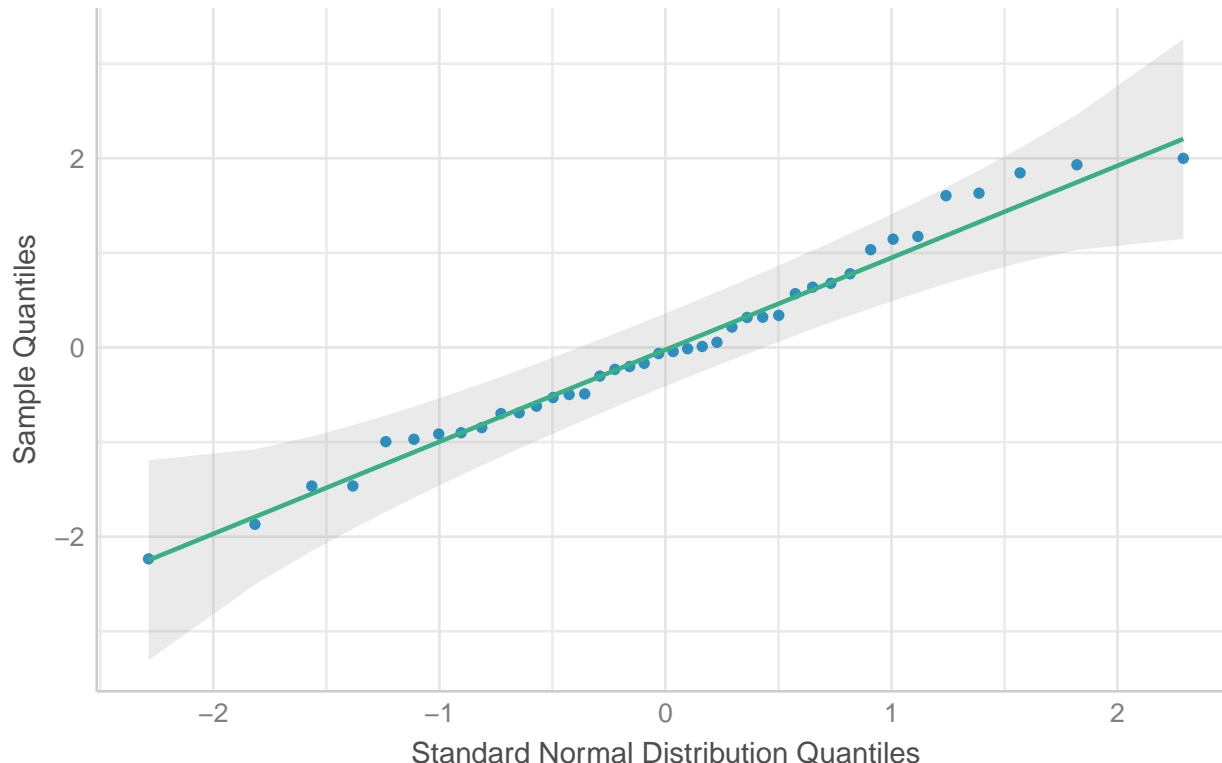
```
check_homogeneity(chol_lm)
```

```
## OK: There is not clear evidence for different variances across groups (Bartlett Test, p = 0.158).
```

```
plot(check_normality(chol_lm), type = "qq")
```

Normality of Residuals

Dots should fall along the line

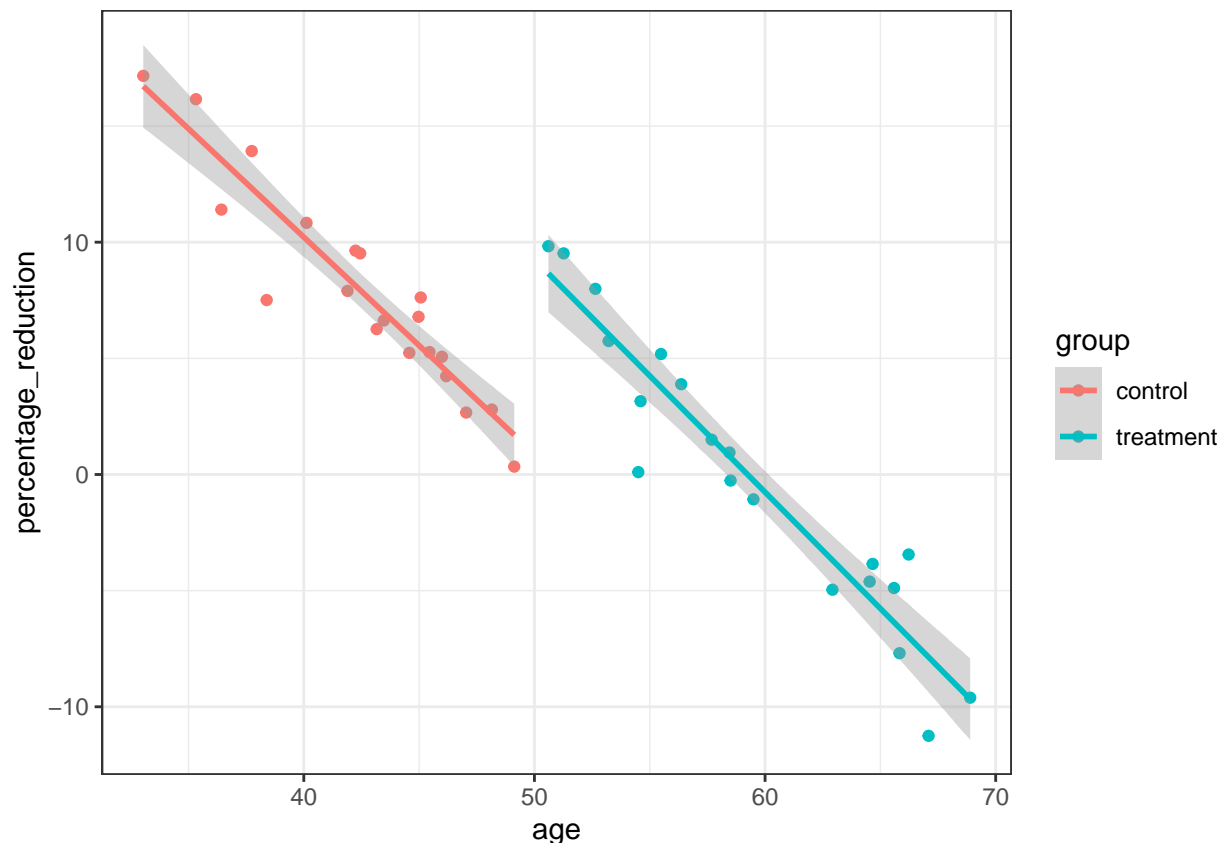


Colesterol (II)

A pesar de que las asunciones del modelo parecen verificarse, el análisis del ejercicio anterior es incorrecto, ya que se ha ignorado la influencia de la variable edad en los resultados. De hecho, el estudio está mal planteado porque las personas pertenecientes al grupo de tratamiento son mayores que las personas pertenecientes al grupo de control (idealmente, su distribución debería ser similar).

1. Haz un gráfico para demostrar que la reducción en colesterol depende de la edad y del grupo. En base a este gráfico, ¿hay interacciones entre la edad y el grupo?
2. Repite el análisis ajustando tanto por la edad como por el tratamiento recibido (recuerda que a este tipo de análisis, donde ajustamos por una variable continua y otra categórica, se le conoce como **ANCOVA**). Para una misma edad, ¿existen diferencias entre ambos grupos?

```
## Warning in geom_smooth(method = "lm", confidence = FALSE): Ignoring unknown
## parameters: `confidence`
## `geom_smooth()` using formula = 'y ~ x'
```



nuevo, el efecto del grupo es significativo y grande ($p < .001$, $\eta^2 = 0.37$), si bien esta vez se obtiene que el grupo de tratamiento reduce, en media, un 8.0% más su colesterol que el grupo de control (si bien el intervalo (4.62%, 11.45%) es igualmente plausible). ¡Fíjate que las conclusiones son opuestas a las del ejercicio anterior! Esto demuestra la importancia de la visualización para elegir qué variables entran en el modelo.

Actividad sexual y esperanza de vida

Los datos de este ejercicio provienen de un estudio sobre la actividad sexual y la esperanza de vida de moscas macho realizado por Partridge y Farquhar (1981). 125 moscas de la fruta macho se dividieron al azar en cinco grupos de 25 cada uno y se registró la longevidad de las moscas en días. El significado de cada uno de los grupos es el siguiente:

- **isolated** = la mosca se mantuvo aislada.
- **one** = la mosca macho se mantuvo con una mosca preñada (las moscas preñadas no se aparean).
- **many** = la mosca macho se mantuvo con 8 moscas preñadas (las moscas preñadas no se aparean).
- **low** = la mosca macho cohabitó con una mosca virgen nueva cada día.
- **high** = la mosca macho cohabitó con 8 moscas vírgenes nuevas cada día.

Adicionalmente, se midió la longitud del tórax de cada macho, ya que se sabía que esto afectaba la longevidad. El propósito del análisis es determinar las diferencias en las esperanzas de vida entre los cinco grupos, si las hay. Datos en “fruitfly.csv”

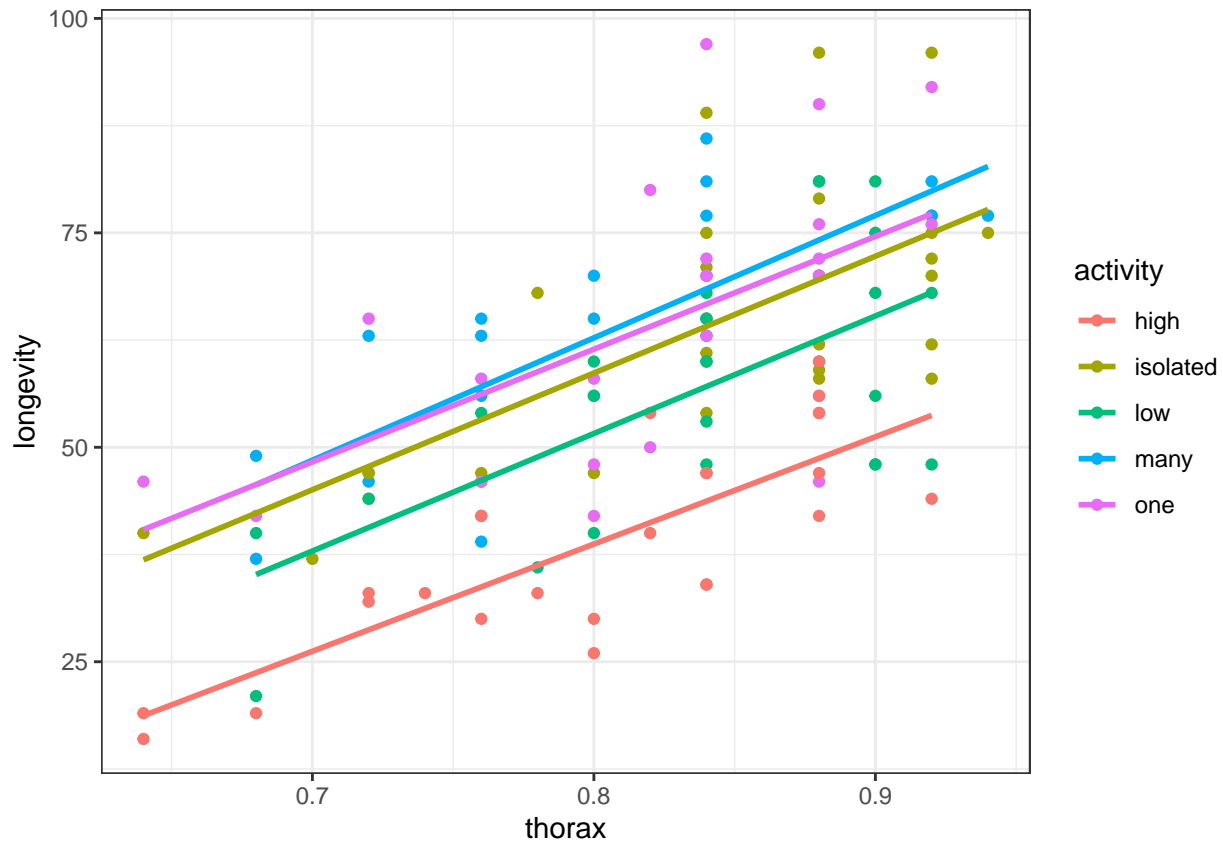
```
fruitfly = read_tsv("data_reg/fruitfly.txt", show_col_types = FALSE)
fruitfly$activity = as.factor(fruitfly$activity)

# 1) Visualización

ggplot(fruitfly, aes(x = thorax, y = longevity, col = activity)) +
```

```
geom_point() +
geom_smooth(method = "lm", se = FALSE)
```

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



La gráfica revela que no es necesario emplear interacciones.

2) Codificar contrastes

No hay

3) lm y Anova

```
fly_lm = lm(longevity ~ thorax + activity, fruitfly) # Análisis tipo ANCOVA
```

```
fly_aov = Anova(fly_lm, type = 3)
```

4) Test Anova + posthoc

```
print(fly_aov)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: longevity
```

```
##          Sum Sq Df F value    Pr(>F)
## (Intercept) 4853.8  1  43.697 1.168e-09 ***
## thorax      12368.4  1 111.348 < 2.2e-16 ***
## activity     9634.6  4  21.684 1.974e-13 ***
## Residuals   13107.3 118
```

```
## ---
```

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

```
print(effectsize(fly_aov))
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter | Eta2 (partial) |          95% CI
## -----
## thorax    |          0.49 | [0.38, 1.00]
## activity  |          0.42 | [0.30, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

El efecto de la actividad es estadísticamente significativo y grande ($p < .001$; Eta^2 (parcial) = 0.42).

```
conditional_means = emmeans(fly_lm, "activity")
pairs(conditional_means, adjust = "bonferroni", infer = c(TRUE, TRUE))
```

```
## contrast      estimate    SE  df lower.CL upper.CL t.ratio p.value
## high - isolated -20.00 3.02 118 -28.63 -11.38 -6.632 <.0001
## high - low      -12.99 3.02 118 -21.63 -4.35 -4.302 0.0004
## high - many     -24.14 3.02 118 -32.77 -15.51 -8.005 <.0001
## high - one      -22.64 3.00 118 -31.22 -14.06 -7.550 <.0001
## isolated - low   7.01 2.98 118 -1.51 15.54 2.353 0.0207
## isolated - many -4.14 3.03 118 -12.80 4.52 -1.367 1.0000
## isolated - one  -2.64 2.98 118 -11.17 5.90 -0.884 1.0000
## low - many      -11.15 3.03 118 -19.82 -2.49 -3.683 0.0035
## low - one       -9.65 2.98 118 -18.19 -1.11 -3.234 0.0159
## many - one       1.50 3.02 118 -7.13 10.13 0.498 1.0000
##
## Confidence level used: 0.95
## Conf-level adjustment: bonferroni method for 10 estimates
## P value adjustment: bonferroni method for 10 tests
```

Las comparaciones dos a dos revelan que los grupos **isolated**, **one** y **many** son estadísticamente equivalentes. En general, estos grupos representan a moscas sin actividad sexual, por lo que parece razonable que no existan diferencias entre estos (si bien podemos concluir que la longevidad de las moscas macho no se ve afectada por la presencia de una mosca hembra si no hay actividad sexual).

Las comparaciones **low** con los grupos de “sin-sexo” son más difíciles de interpretar si nos fijamos solo en los p-valores, ya que hay dos diferencias son significativas (con **many** y **one**) y otra que no lo es (con **isolated**). Si en cambio nos centramos en los intervalos de confianza, observamos que los resultados sugieren que, para un mismo tamaño de tórax, el grupo **low** tiene una menor esperanza de vida que los grupos “sin-sexo”, si bien en el caso de **isolated** no podemos descartar que la diferencia sea nula (ya que la diferencia $\mu_{\text{isolated}} - \mu_{\text{low}} \in (-1.51, 15.54)$ al 95% de confianza).

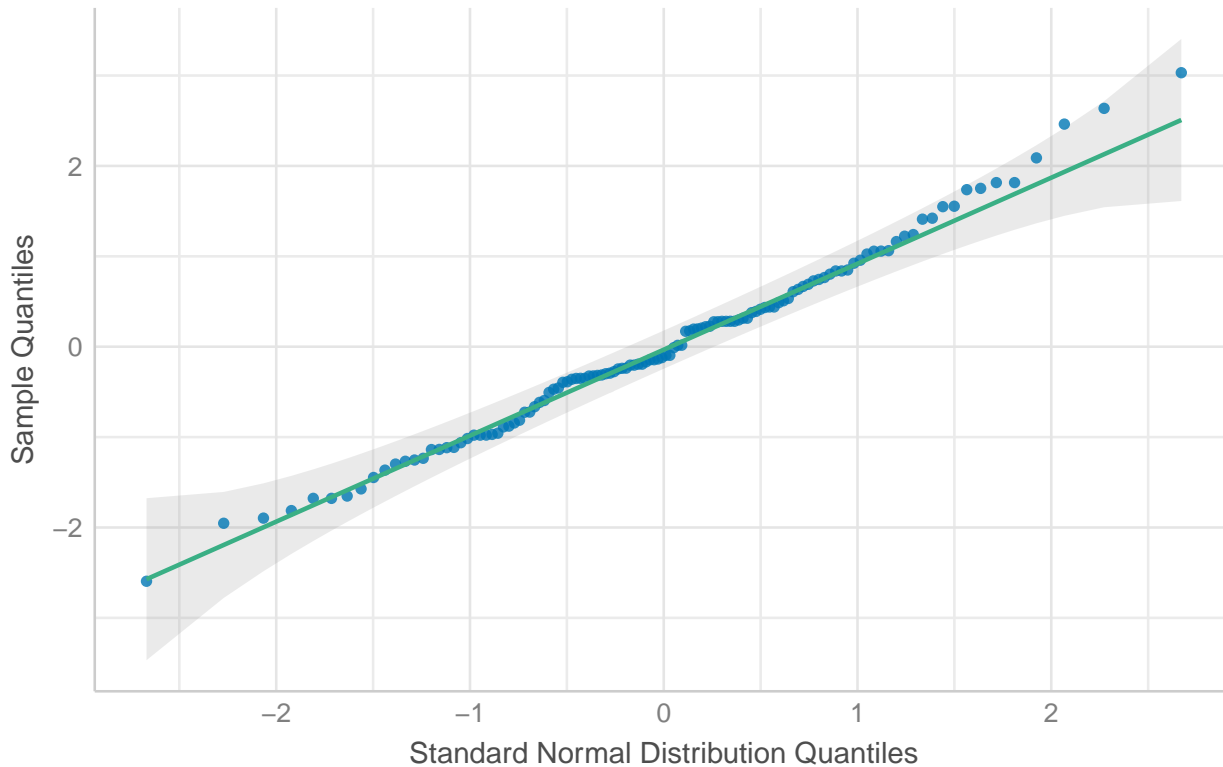
Finalmente, el grupo **high** muestra claras diferencias con el resto de grupos. Podemos afirmar que las moscas del grupo **high**, para un mismo tamaño de tórax, tienen una menor esperanza de vida (siendo la diferencia media máxima de unos 24 días, con el grupo **many**).

Todas estas conclusiones están sujetas a que las asunciones del modelo sean correctas...

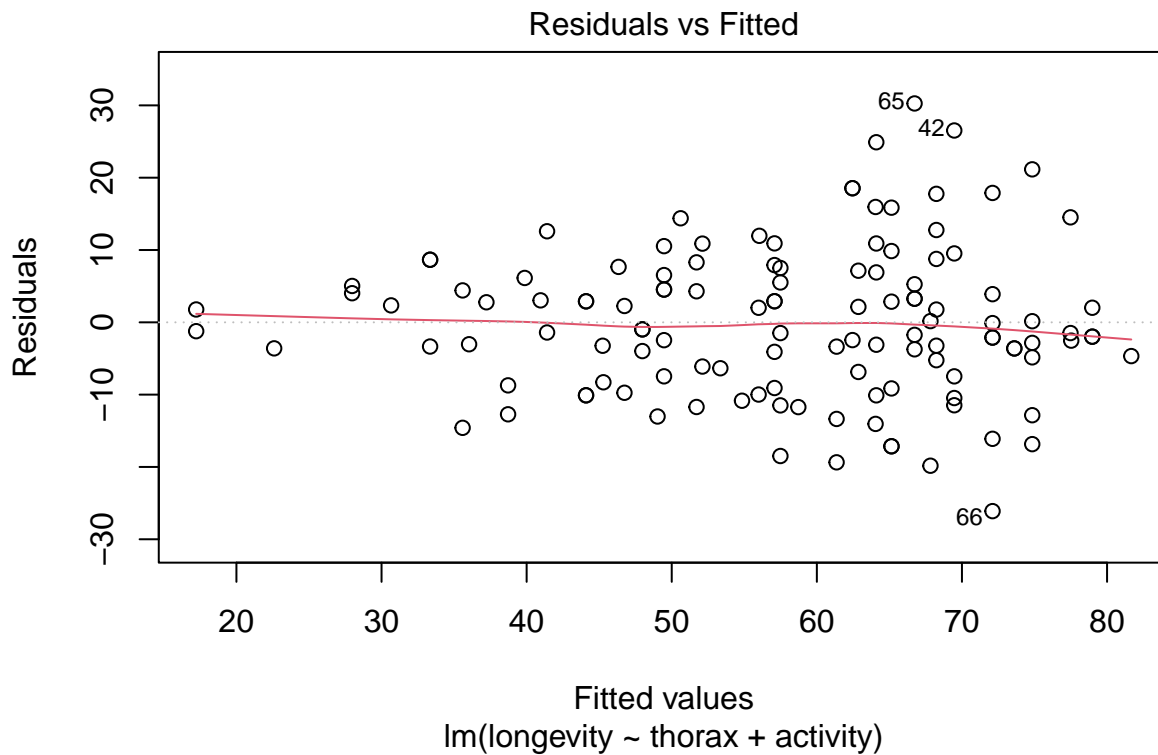
```
# 5) Chequeo de asunciones
plot(check_normality(fly_lm), type = "qq")
```

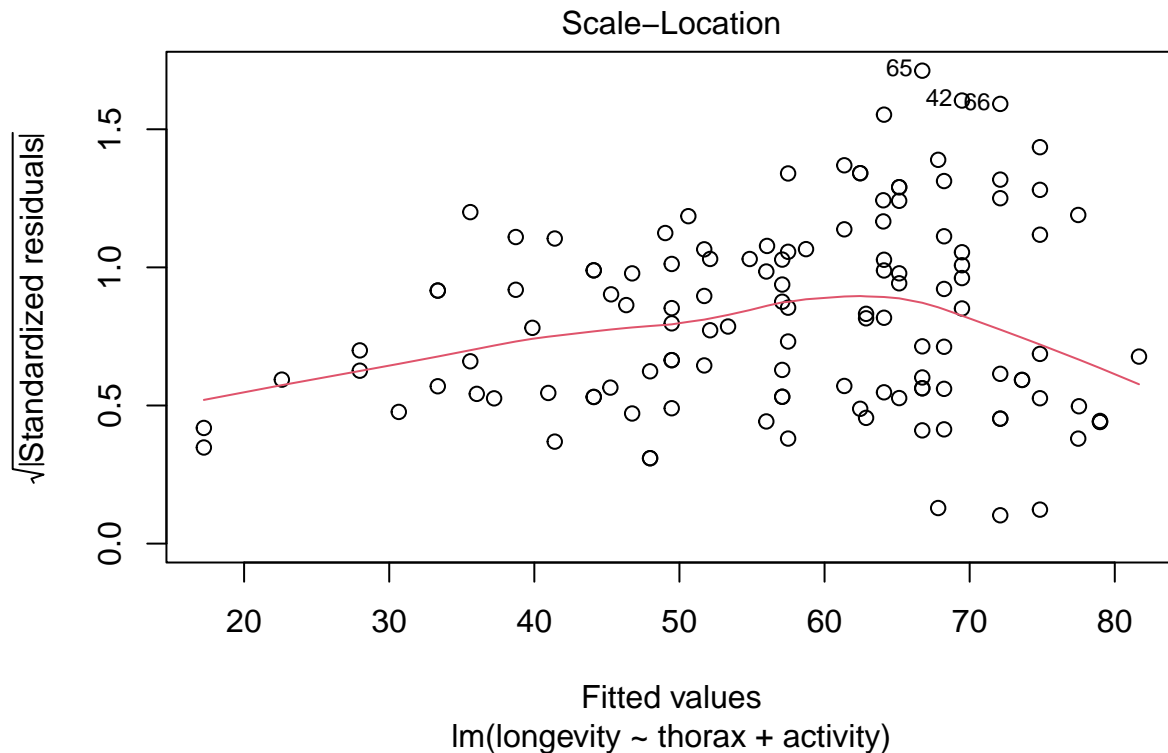
Normality of Residuals

Dots should fall along the line



```
plot(fly_lm, ask = FALSE, which = c(1, 3))
```





Los últimos dos plots sugieren que la asunción de que la varianza es homogénea puede no cumplirse (la zona derecha tiene más dispersión que la izquierda). Aunque no es grave, sería recomendable repetir el análisis aplicando la transformación logarítmica a `longevity`.

Actividad sexual y esperanza de vida (II)

Repite el ejercicio anterior si, en vez de querer comparar todos los grupos entre sí, quisiésemos comprobar las siguientes preguntas:

1. ¿Hay diferencias entre los grupos “sin-sexo” y los grupos “con-sexo”.
2. ¿Hay diferencias entre el grupo totalmente aislado y los grupos con moscas preñadas?
3. ¿Hay diferencias entre los grupos de una sola mosca preñada y el grupo de 8 moscas preñadas?
4. ¿Hay diferencias entre el grupo de “poco sexo” y el grupo de “mucho sexo”?

```
# ...

# 2) contrastes
contrasts(fruitfly$activity) = get_contrasts_coding(
  rbind(
    " Sin Vs Con" = c(-1/2, 1/3, -1/2, 1/3, 1/3),
    " Aislado Vs Pren" = c(0, 1, 0, -1/2, -1/2),
    " 1 Pren Vs 8 Pren" = c(0, 0, 0, -1, 1),
    " Poco Vs Mucho" = c(-1, 0, 1, 0, 0)
  )
)

# 3) Regresión (no es necesario ANOVA porque no vamos a hacer post-hoc)
fly_lm = lm(longevity ~ thorax + activity, fruitfly)

# 4) Análisis de los contrastes
```

```
summary(fly_lm)
```

```
##
## Call:
## lm(formula = longevity ~ thorax + activity, data = fruitfly)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.108  -7.014  -1.101   6.234  30.265
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -52.798     10.512  -5.023 1.83e-06 ***
## thorax          134.341     12.731  10.552 < 2e-16 ***
## activity Sin Vs Con    15.768     1.931   8.166 4.01e-13 ***
## activity Aislado Vs Pren  -3.388     2.600  -1.303   0.195
## activity 1 Pren Vs 8 Pren  -1.502     3.016  -0.498   0.620
## activity Poco Vs Mucho    12.989     3.019   4.302 3.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.54 on 118 degrees of freedom
## Multiple R-squared:  0.6527, Adjusted R-squared:  0.638
## F-statistic: 44.36 on 5 and 118 DF,  p-value: < 2.2e-16
```

```
confint(fly_lm)
```

```
##              2.5 %      97.5 %
## (Intercept)   -73.614683 -31.981223
## thorax        109.130197 159.552553
## activity Sin Vs Con    11.943996 19.591887
## activity Aislado Vs Pren  -8.535705  1.759865
## activity 1 Pren Vs 8 Pren  -7.475023  4.471945
## activity Poco Vs Mucho    7.009965 18.967563
```

No existen diferencias significativas entre los grupos “Aislado” y “Preñados”, ni entre el grupo de “una preñada” Vs “8 preñadas”. Para un mismo tamaño de tórax, las moscas sin actividad sexual viven una media de 15 días más que las moscas con actividad (si bien el intervalo (11.9, 19.6) días es igualmente plausible, con una confianza del 95%). Para un mismo tamaño de tórax, las moscas con poca actividad sexual viven de media 13 días más que las moscas con mucha actividad sexual (si bien el intervalo (7, 19) es igualmente plausible, con un nivel de confianza del 95%).

¿Es buena idea fumar durante el embarazo?

En temas anteriores se exploró el impacto de fumar durante el embarazo en los pesos de los bebés al nacer. Un enfoque más realista para modelar los pesos de los bebés es considerar todas las variables potencialmente influyentes. Algunas variables de interés son: la duración del embarazo en semanas (weeks), la edad de la madre en años (mage), el sexo del bebé (sex), el tabaquismo de la madre (habit) y el número de visitas al hospital (visits) durante el embarazo. Usa `openintro::births14` para analizar el impacto de estas variables en el peso del bebé. Interpreta los resultados y evalúa la bondad del modelo.

```
births = openintro::births14 # ?openintro:births14 para el significado de las variables
births = births[, c("weight", "weeks", "mage", "sex", "habit", "visits")]
# Elimina los NAs presentes en el data.frame
```



```
births = births[complete.cases(births), ]
```

```
# 1) Visualización
```

```
GGally::ggpairs(births)
```

```
## Registered S3 method overwritten by 'GGally':
```

```
##   method from
```

```
##   +.gg      ggplot2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

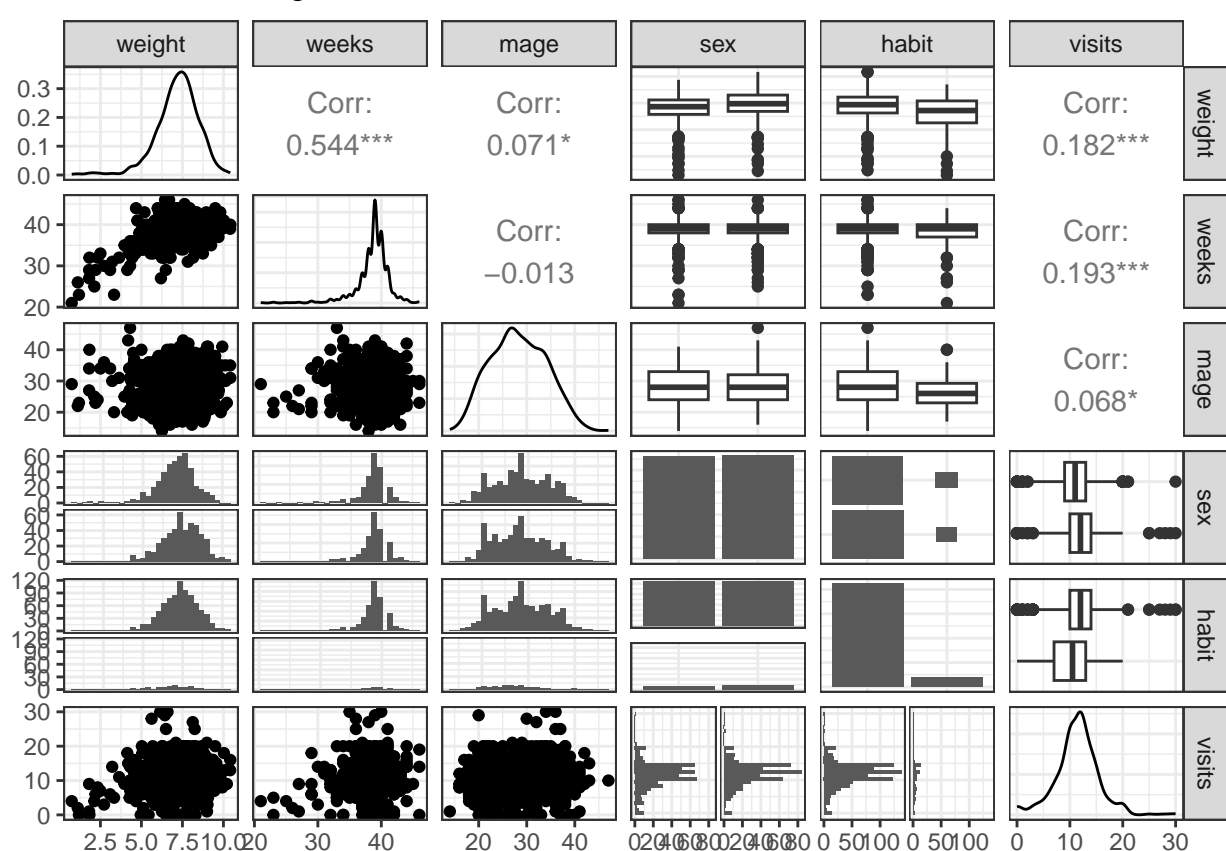
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# 2) No hay contrastes
```

```
# 3) Regresión (no es necesario ANOVA, ¿por qué?)
```

```
model = lm(weight ~ weeks + mage + sex + habit + visits, data = births)
```

```
# 4) Resultados, incluyendo tamaño del efecto.
```

```
print(summary(model))
```

```
##
```

```
## Call:
```

```

## lm(formula = weight ~ weeks + mage + sex + habit + visits, data = births)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4370 -0.7154  0.0115  0.6996  3.7213
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.818059   0.567153  -6.732 2.94e-11 ***
## weeks        0.264509   0.013971  18.933 < 2e-16 ***
## mage         0.015518   0.006129   2.532 0.011511 *
## sexmale      0.369806   0.069732   5.303 1.43e-07 ***
## habit smoker -0.429281   0.125985  -3.407 0.000684 ***
## visits       0.018222   0.008740   2.085 0.037341 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.058 on 920 degrees of freedom
## Multiple R-squared:  0.3351, Adjusted R-squared:  0.3315
## F-statistic: 92.75 on 5 and 920 DF,  p-value: < 2.2e-16
print(eta_squared(model))

## # Effect Size for ANOVA (Type I)
##
## Parameter | Eta2 (partial) |      95% CI
## -----
## weeks      |          0.31 | [0.27, 1.00]
## mage       |       9.06e-03 | [0.00, 1.00]
## sex        |          0.03 | [0.01, 1.00]
## habit      |          0.01 | [0.00, 1.00]
## visits     |       4.70e-03 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].

# 5) Chequeo de la bondad del modelo
is_norm = check_normality(model)
plot(is_norm, type="qq", detrend=TRUE)

```

Normality of Residuals

Dots should fall along the line

