TFM

Javier Pérez

Invalid Date

Tabla de contenidos

Listado de Figuras

Listado de Tablas

Capítulo 1

Ìndice

Capítulo 2

Análisis de la varianza de un factor.

Siguiendo a Lawson (2018, 353-54) vamos a realizar un análisis de la varianza de un factor usando sólo la pregunta 18 que es una valoración general de subtitulado.

2.1 Preparación de datos

TODO: Comentar y poner en bonito la tabla

```
# A tibble: 174 x 24
     Row Group Test   User LastTry               Q01   Q02   Q03   Q04   Q05
   <dbl> <chr> <chr> <dbl> <dttm>              <dbl> <dbl> <dbl> <dbl> <dbl>
 1     5 A     01        4 2022-04-13 14:21:55     2     2     2     2     2
 2     4 A     02        4 2022-04-13 14:23:38     2     2     2     2     2
 3    47 B     01       33 2022-04-28 07:52:08     3     3     3     3     1
 4    27 B     02       33 2022-04-28 08:00:20     3     3     3     3     3
 5    85 A     01       35 2022-04-14 02:53:36     4     4     4     4     4
 6    49 A     02       35 2022-04-14 02:58:19     4     3     4     4     3
 7    23 A     01       38 2022-04-15 14:19:38     4     4     4     4     4
 8    15 A     02       38 2022-04-15 14:28:36     0     4     4     4     0
 9    68 A     01       59 2022-04-18 19:12:43     4     4     4     4     4
10    37 A     02       59 2022-04-18 19:22:01     3     2     3     3     1
# ... with 164 more rows, and 14 more variables: Q06 <dbl>, Q07 <dbl>,
#   Q08 <dbl>, Q09 <dbl>, Q10 <dbl>, Q11 <dbl>, Q12 <dbl>, Q13 <dbl>,
#   Q14 <dbl>, Q15 <dbl>, Q16 <dbl>, Q17 <dbl>, Q18 <dbl>, Rows <int>
```

Disponemos de estos datos:

`Test`

```
Group 01 02
    A 43 43
    B 44 44
```

Vamos a cambiar la nomenclatura para adaptarla a Lawson (2018):

- El grupo A pasará a llamarse AB.
- El grupo B pasará a llamarse BA.
- Los estudiantes se denominarán sujetos.
- Los test 01 y 02 tratamientos.
- Se introduce una variable periodo.

Los valores del test de Likert se desplazarán para que tengan valores más lógicos:

- 0 = No sé / No constesto
- 1 = Muy en desacuerdo
- 2 = En desacuerdo
- 3 = Neutral
- 4 = De acuerdo
- 5 = Muy de acuerdo

Finalmente la tabla se pasará a formato largo.

```
# A tibble: 3,132 x 6
   Group Period Treat Subject Question Response
   <fct> <fct>  <fct> <fct>   <chr>       <dbl>
 1 AB    1      A     4       Q01             3
 2 AB    1      A     4       Q02             3
 3 AB    1      A     4       Q03             3
 4 AB    1      A     4       Q04             3
 5 AB    1      A     4       Q05             3
 6 AB    1      A     4       Q06             3
 7 AB    1      A     4       Q07             3
 8 AB    1      A     4       Q08             3
 9 AB    1      A     4       Q09             3
10 AB    1      A     4       Q10             3
# ... with 3,122 more rows
```

El análisis de la varianza, ANOVA, se debe realizar una variable de respuesta cuantitativo (TODO: incluir supuestos de ANOVA). El test de Likert tiene una escala ordinal y, por lo tanto, ANOVA no es una técnica adecuada. De todas formas vamos a realizar un análisis de la varianza para la pregunta 18 que es una valoración general del subtitulado.

Vamos a seguir el proceso descrito en Meier (2022) para realizar el ANOVA.

2.2 Análisis de la varianza de un factor.

Ajuste del modelo.

El factor va a ser el tratamiento con dos niveles A y B [1].

El modelo que vamos a ajustar es:

$$Y_{ij} \sim N(\mu_i, \sigma^2), \text{ independientes} \tag{2.1}$$

Los valores observados y_{ij} corresponden a las respuestas a la pregunta 18 del sujeto j-ésimo que recibió el i-ésimo tratamiento (A o B) y siguen una distribución normal con media μ_i y varianza σ^2. Es decir, que estamos suponiendo que los dos tratamientos tienen la misma varianza pero pueden tener distinta media.

El modelo (Ecuación 2.1) también se puede expresar:

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2.2}$$

Con ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$. En esta ecuación simplemente hemos separado el término determinista μ_i del estocástico ϵ_{ij}.

Una reparametrización alternativa es:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \tag{2.3}$$

En este caso, estamos considerando que existe un efecto fijo, μ, y que cada factor tiene una desviación, α_i, sobre el ese nivel fijo. Así $\displaystyle\sum_{i=1}^{g} \alpha_i = 0$. Lo que en nuestro, en el que sólo hay dos niveles ($g = 2$), implica que $\alpha_A + \alpha_B = 0$. En R se puede elegir uno (Ecuación 2.2) u otro (Ecuación 2.2) tipo de parametrización.

Podemos visualizar que los valores de respuesta cada nivel de tratamiento está claramente separado:

Finalmente ajustamos el modelo y mostramos los coeficientes:

[1] Para este ejemplo en el que sólo hay dos niveles de factor, podríamos haber utilizado un t-test de comparación de medias.

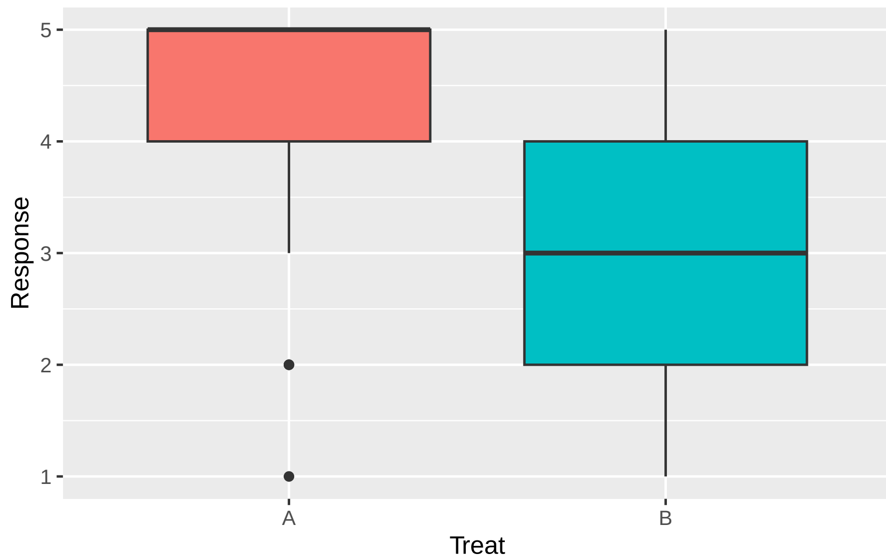

Figura 2.1: Resumen de las respuestas a la pregunta 18 en cada nivel de tratamiento.

```
fit.q18 <- aov(Response ~ Treat, data = df18)
coef.q18 <- coef(fit.q18)
coef.q18
```

```
(Intercept)      TreatB
   4.459770   -1.574713
```

Por defecto R elige como nivel de referencia del factor el A por ser menor alfabéticamente y el término de intercepción se corresponde con este valor, así $\mu_A = 4.46$ y el nivel del tratamiento B está como diferencias sobre el de referencia. Por lo tanto, $\mu_B = 2.89$.

Alternativamente podemos obtener las medias de cada nivel de esta forma:

```
predict(fit.q18, newdata = data.frame(Treat = c("A", "B")))
```

```
       1        2
4.459770 2.885057
```

O con la librería `emmeans`, que también nos proporciona el intervalo de confianza con el 95%:

```r
library(emmeans)
emmeans(fit.q18, specs = ~Treat)
```

```
 Treat emmean   SE  df lower.CL upper.CL
 A       4.46 0.11 172     4.24     4.68
 B       2.89 0.11 172     2.67     3.10

Confidence level used: 0.95
```

Con R, podemos obtener los valores correspondientes a la segunda parametrización (Ecuación 2.3) del modelo:

```r
options(contrasts = c("contr.sum", "contr.poly"))
fit.q18.2 <- aov(Response ~ Treat, data = df18)
coef.q18.2 <- coef(fit.q18.2)
coef.q18.2
```

```
(Intercept)      Treat1
  3.6724138   0.7873563
```

```r
options(contrasts = c("contr.treatment", "contr.poly"))
```

Vemos que cambian tanto los valores como el esquema de nombrado. Ahora el término de intercepción se corresponde con la media global ($\mu = 3.67$) y `Treat1` es la diferencia del nivel de factor 1 con esa media ($\alpha_A = 0.79$), como la suma de todos los niveles tiene que ser 0, $\alpha_B = -0.79$.

Tests estadísticos.

En ANOVA el contraste de hipótesis habitual es contrastar si las medias de los niveles de un factor son iguales o hay alguna diferente:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = ... = \mu_g \\ H_A &: \mu_k \neq \mu_l \text{ para al menos un par } k \neq l \end{aligned}$$

Los resultados de F-test son significativos y permiten rechazar la hipótesis nula de que los dos tratamientos son iguales, es decir que podemos concluir que el subtitulado de los vídeos es percibido por los estudiantes con diferente:

```r
summary(fit.q18)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
Treat         1  107.9  107.87   101.7 <2e-16 ***
Residuals   172  182.5    1.06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos calcular el estadístico F usando un test *drop*1 que consiste en ajustar el modelo con y sin variables predictoras y comparar los resultados.

```r
drop1(fit.q18, test = "F")
```

```
Single term deletions

Model:
Response ~ Treat
       Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>              182.46 12.261
Treat   1    107.87 290.33 91.080  101.68 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Una tercera forma obtener los mismos resultados es comparar mediante un F-test el modelo con un factor con el modelo nulo:

```r
fit.q18.single <- aov(Response ~ 1, data = df18)
anova(fit.q18.single, fit.q18)
```

```
Analysis of Variance Table

Model 1: Response ~ 1
Model 2: Response ~ Treat
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    173 290.33
2    172 182.46  1    107.87 101.68 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos obtener las significación estadística y los intervalos de confianza de los $\alpha_i's$:

```r
summary.lm(fit.q18)
```

```
Call:
aov(formula = Response ~ Treat, data = df18)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4598 -0.8851  0.1149  0.5402  2.1149

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.4598     0.1104   40.39   <2e-16 ***
TreatB       -1.5747     0.1562  -10.08   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.03 on 172 degrees of freedom
Multiple R-squared:  0.3715,    Adjusted R-squared:  0.3679
F-statistic: 101.7 on 1 and 172 DF,  p-value: < 2.2e-16
```

```
confint(fit.q18)
```

```
                2.5 %    97.5 %
(Intercept)  4.241811  4.677729
TreatB      -1.882953 -1.266472
```

Comprobación de los supuestos del modelo.

La inferencia estadística solo es válida si se cumplen las siguientes premisas:

- Los errores son independientes.
- Los errores están distribuidos normalmente.
- La varianza del error es constante.
- Los errores tienen una media de cero.

La independencia de los errores se consigue aleatorizando el experimento. Habría que hacer comprobaciones estadísticas de la representatividad de la muestra que no se abordan en este trabajo por no ser objeto del mismo. En cualquier caso, en un estudio cruzado los errores no son independientes y trataremos este problema más adelante.

Análisis de residuos

No observamos directamente los errores, ϵ_{ij}, sino una estimación suya que denominamos residuos:

$$r_{ij} = y_{ij} - \hat{\mu}_i.$$

Para comprobar la normalidad de los residuos es habitual utilizar un gráfico **QQ-plot** que compara los percentiles de los residuos obtenidos de tras ajustar el modelo con los que resultarían de un distribución normal.

```
qqPlot(fit.q18, distribution = "norm")
```

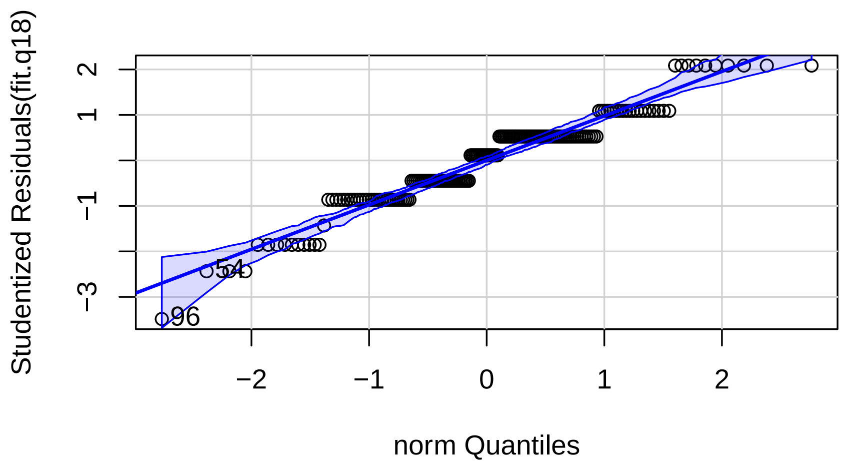

Figura 2.2: QQ-plot para comprobar la normalidad de los residuos.

```
[1] 54 96
```

Como era de esperar, una variable de respuesta ordinal no va a producir residuos con distribución normal.

Si los residuos no tienen una distribución normal, el resto de test que hagamos carecen de sentido ya parten de la premisa de que los residuos son normales. De todas formas realizaremos algunos de estos test. Por ejemplo, para comprobar si los residuos tienen una varianza constante, podemos utilizar el **Tukey-Anscombe plot (TA-plot)**.

```
plot(fit.q18, which = 1)
```

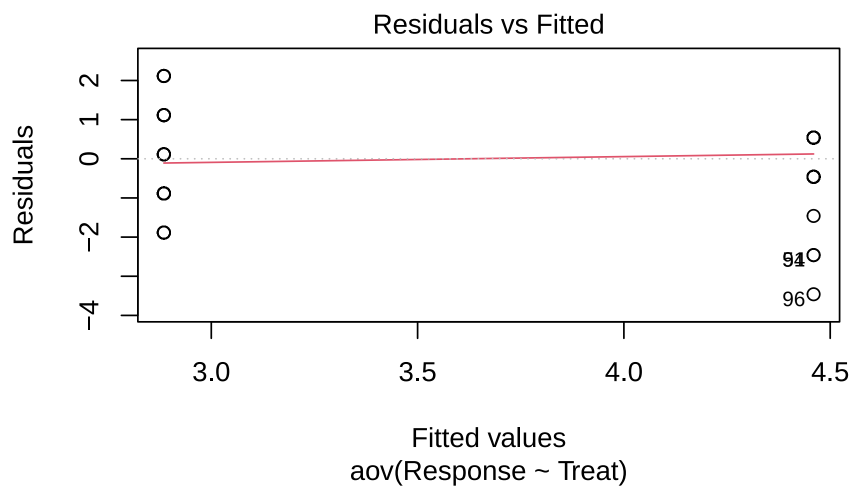

Figura 2.3: TA-plot para comprobar la varianza constante de los residuos.

Constatamos que los residuos tienden a crecer ligeramente.

Aproximaciones no paramétricas.

Dado que la variable de respuesta es ordinal y no sigue una distribución normal, podemos hacer un test no paramétrico como es el **test de Kruskal-Wallis**.

```r
fit.q18.kw <- kruskal.test(Response ~ Treat, data = df18)
fit.q18.kw
```

```
	Kruskal-Wallis rank sum test

data:  Response by Treat
Kruskal-Wallis chi-squared = 65.658, df = 1, p-value = 5.364e-16
```

Este test produce significación estadística de la diferencia de medias.

Capítulo 3

Análisis de la varianza de dos factores.

Vamos a realizar el análisis pero ahora con dos factores. Hay varias posibilidades. Entre ellas vamos a seleccionar como factores el tratamiento (subtitulado) y las preguntas. Es decir, nos planteamos si hay diferencias entre las preguntas en los distintos tratamientos. Lógicamente esperamos que así sea, pero en el gráfico de interacción (Figura 3.1) constatamos algunas cuestiones interesantes. Esperaríamos encontrar que uno de los tratamientos debería tener puntuaciones elevadas y el otro bajas ya que una actividad corresponde a un test sobre un vídeo con subtítulos correctos y en la otra hay deficiencias en el subtitulado. Esto es lo que se puede apreciar en el gráfico. Sin embargo en algunas preguntas esto no sucede.:

- En las preguntas 4 y 13 ambos tratamientos tienen puntuaciones altas.
- En la pregunta 16 ambos tratamientos tienen puntuaciones medias y parecidas.
- Las preguntas 15, 16 y 17 en el tratamiento A tienen un valor muy inferior a su media. Lo mismo sucede con las preguntas 5 y 9 de tratamiento B.
- Las oscilaciones en el tratamiento B son mayores que en el tratamiento A.

Podemos observar que las preguntas 14, 15 y 17 y sobre todo la 16 tienen una gran cantidad de respuestas "No sé / No Contesto".

TODO: Probablemente los dos gráficos anteriores tienen más sentido en el EDA. Revisar.

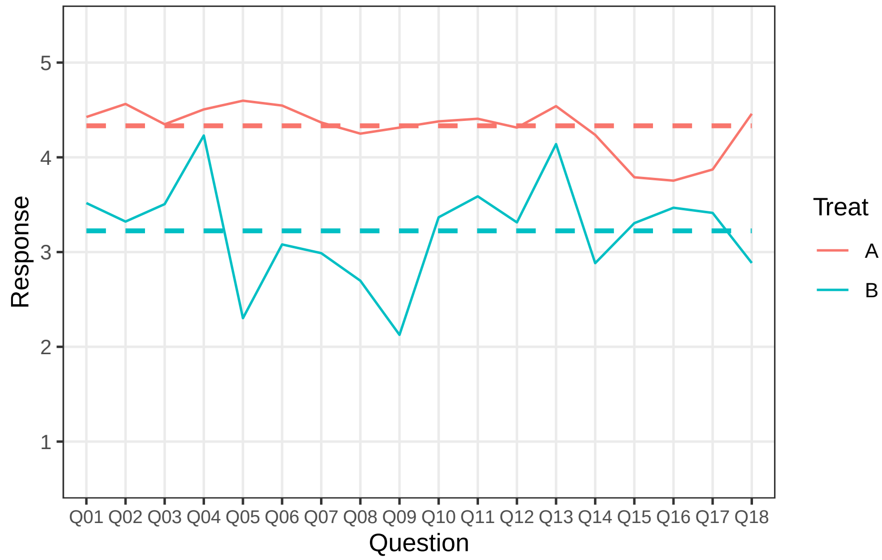

Figura 3.1: Gráficos de interacción tratamiento y pregunta (valor medio de las respuestas).

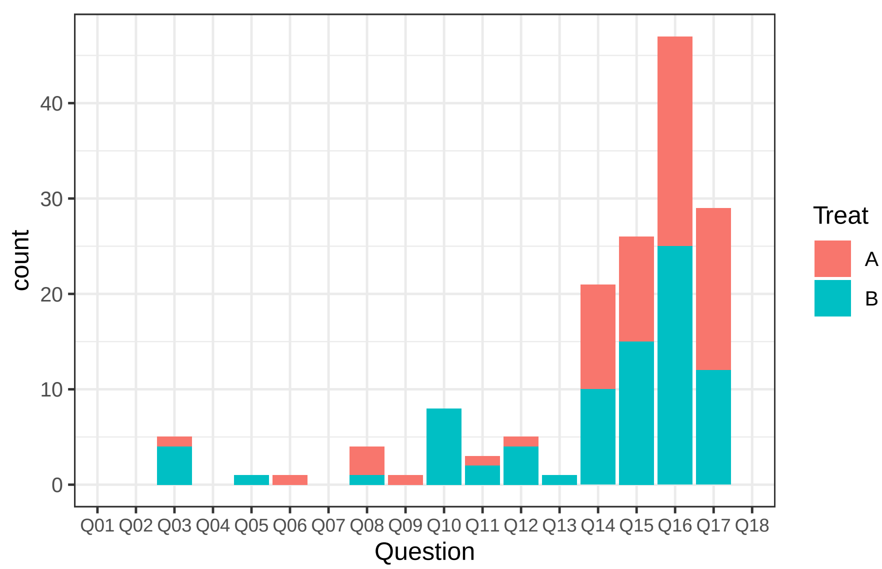

Figura 3.2: Contestación a preguntas con valor No sé / No Contesto por tratamiento.

References

Lawson, John E. 2018. *Design and Analysis of Experiments*. 1st ed. Chapman; Hall/CRC. https://www.taylorfrancis.com/books/mono/10.1201/b17883/design-analysis-experiments-john-lawson.

Meier, Lukas. 2022. *ANOVA and mixed models: A short introduction using R*. Chapman; Hall/CRC. https://www.taylorfrancis.com/books/mono/10.1201/9781003146216/anova-mixed-models-lukas-meier.