TMF

Javier Pérez

Invalid Date

Tabla de contenidos

Tabla de contenidos **i**

1 Ìndice **1**

2 Análisis de la varianza pregunta 18 **3**

2.1 Preparación de datos 3

2.2 Análisis de la varianza de un factor. 5

2.3 Contraste de hipótesis. 7

References **9**

Capítulo 1

Ìndice

Capítulo 2

Análisis de la varianza pregunta 18

Siguiendo a Lawson (2018, 353-54) vamos a realizar un análisis de la varianza.

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
   <chr>  <dbl> <chr>   <dbl> <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 sobre datos cuantitativos (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.

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}.

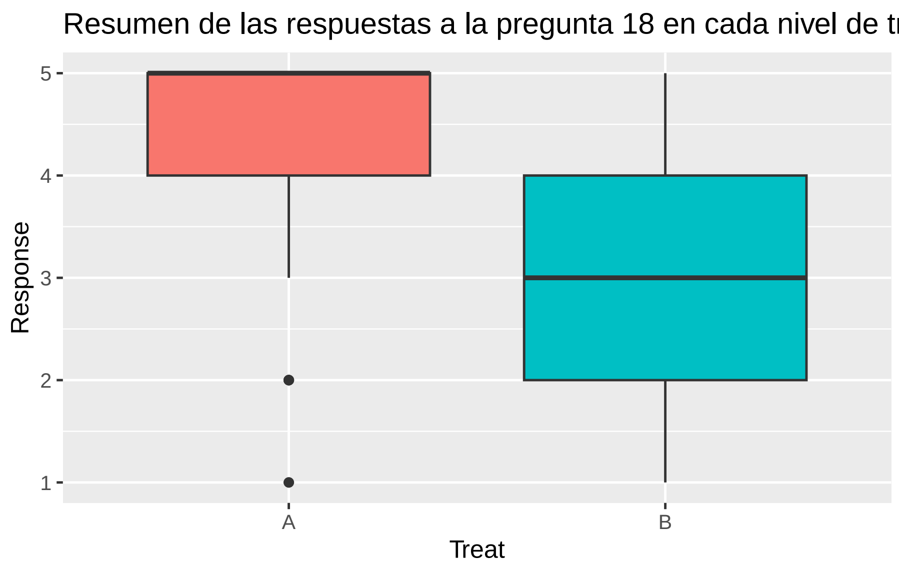
Una reparametrización alternativa es:

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

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 $\sum_{i=1}^{g} \alpha_i = 0$. 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:

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



Finalmente ajustamos el modelo y mostramos los coeficientes:

```
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
```

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, $\alpha_B = -0.79$.

2.3 Contraste de hipótesis.

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

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

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.