



Universidad Nacional de Educación a Distancia  
Escuela Técnica Superior de Informática  
Máster en Ingeniería y Ciencia de Datos

## **Trabajo Fin de Máster**

**Utilización de técnicas multivariantes para  
el estudio del aprendizaje de la mejora de  
la accesibilidad en el subtitulado de vídeos**

Autor: Javier Pérez Arteaga

Directores: Emilio Letón Molina

Jorge Pérez Martín

Fecha de realización: 2023-10-03

---

This document is reproducible thanks to:

- L<sup>A</sup>T<sub>E</sub>X and its class memoir (<http://www.ctan.org/pkg/memoir>).
- R (<http://www.r-project.org/>) and RStudio (<http://www.rstudio.com/>)
- bookdown (<http://bookdown.org/>) and memoirR (<https://ericmarcon.github.io/memoiR/>)



Name of the owner of the logo

<http://www.company.com>

# RESUMEN

TODO: Incluir un resumen del trabajo.



## AGRADECIMIENTOS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed malesuada nulla augue, ac facilisis risus pretium a. Ut bibendum risus id ex fermentum, at accumsan erat vulputate. In hac habitasse platea dictumst. Sed lobortis est a enim bibendum, ac pulvinar nulla aliquam. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque efficitur justo id suscipit pretium. Proin iaculis sit amet nibh vel euismod. Aenean tincidunt faucibus ex, non vehicula ipsum tristique in. Fusce vel tincidunt lectus, vel rutrum nisi. Suspendisse malesuada lectus ac enim vehicula rhoncus. Nullam convallis justo in bibendum eleifend.

Phasellus vitae magna nec mi sagittis luctus vitae eu augue. Donec scelerisque laoreet arcu, eget tempor mi ultricies vel. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vestibulum at blandit ex. Vestibulum eu sagittis mauris. In hac habitasse platea dictumst. Duis eget ante vel lacus sollicitudin convallis quis eu velit. Sed auctor sem non nisi hendrerit, vel tincidunt tortor bibendum.

# ÍNDICE

<b>Resumen</b>	<b>iii</b>
<b>Agradecimientos</b>	<b>v</b>
<b>Índice</b>	<b>vi</b>
<b>Índice de cuadros</b>	<b>ix</b>
<b>Índice de figuras</b>	<b>xi</b>
<b>1 Introducción</b>	<b>1</b>
<b>2 Marco teórico y estado del arte</b>	<b>3</b>
2.1 Características del diseño del experimento . . . . .	3
2.2 Técnicas estadísticas utilizadas . . . . .	5
2.3 Métodos exploratorios . . . . .	5
2.4 Modelos lineales generalizados . . . . .	6
<b>3 Metodología/Materiales y métodos</b>	<b>11</b>
3.1 Descripción de la experiencia . . . . .	11
3.2 Participantes . . . . .	14
3.3 Ficheros suministrados . . . . .	14
3.4 Preprocesado . . . . .	15
3.5 Variables del modelo. . . . .	16
<b>4 Modelado estadístico</b>	<b>19</b>
4.1 Análisis inicial . . . . .	19
4.2 Comparación con <i>Odds Ratio</i> . . . . .	27
4.3 Regresión Logística . . . . .	28
4.4 Regresión Ordinal . . . . .	29
<b>5 Resultados</b>	<b>35</b>
<b>6 Discusión y conclusiones</b>	<b>39</b>
6.1 Comparación con <i>Odds Ratio</i> . . . . .	39
<b>Referencias</b>	<b>41</b>

<b>Apéndices</b>	<b>43</b>
<b>A Efecto secuencia e interacción tratamiento vs. periodo.</b>	<b>43</b>
A.1 Preparación. . . . .	43
A.2 Análisis con un solo factor (tratamiento). . . . .	44
A.3 Análisis con un dos factores (tratamiento y periodo). . . . .	46
A.4 Factor secuencia. . . . .	47





## ÍNDICE DE CUADROS

2.1	Tabla de contingencia . . . . .	6
2.2	Tabla de contingencia . . . . .	6
3.1	Niveles de los items de la escala de Likert. . . . .	13
3.2	Items de la escala de Likert. . . . .	13
3.3	Tablas de contingencia de la información socioeconómica de los estudiantes. . . . .	14
3.8	Descripción de las variables más importantes. . . . .	16
3.9	Muestra del dataframe preparado para el modelado estadístico en formato largo. . . . .	17
4.1	Tiempos de realización de la segunda actividad de duración inferior a 2 minutos. . . . .	20
4.2	Test en los que todas las preguntas se contestan el mismo valor de respuesta. . . . .	20
4.3	Los 5 test con más respuestas ‘No sé/No contesto’ . . . . .	21
4.4	Estudiantes que tienen diferencias en sus respuestas muy alejadas de la tendencia de su grupo. . . . .	23
4.5	Resumen de frecuencias de respuesta. . . . .	23
4.6	Probabilidades de respuesta para el modelo ordinal $\text{Response} \sim \text{Treat}$ . . . . .	31
5.1	Relación de cada pregunta con el índice alpha de Cronbach. . . . .	35
5.4	Tablas de contingencia. . . . .	36
5.5	Valores del contraste de hipótesis $\chi^2$ . . . . .	36
5.6	$\text{LogOR} \sim \text{Treat} + \text{Seq} + \text{Response}_1$ . . . . .	36
5.7	$\text{LogOR} \sim \text{Treat} + \text{Period} + \text{Response}_1$ . . . . .	37
A.1	Ajuste del modelo $\text{Response} \sim \text{Treat}$ con contrasts treatment. . . . .	44
A.2	Ajuste del modelo $\text{Response} \sim \text{Treat} * \text{Period}$ con contrasts treatment. . . . .	46
A.3	Ajuste del modelo $\text{Response} \sim \text{Treat} + \text{Period} + \text{Seq}$ con contrasts treatment. . . . .	47
A.4	Ajuste del modelo $\text{Response} \sim \text{Treat} + \text{Period} + \text{Seq}$ con contrasts sum. . . . .	48
A.5	Ajuste del modelo $\text{Response} \sim \text{Treat} * \text{Period}$ con contrasts sum. . . . .	48



## ÍNDICE DE FIGURAS

2.1	Función latente en una regresión ordinal acumulativa. . . . .	9
4.1	Estudiantes asignados a cada grupo. . . . .	19
4.2	Número de respuestas diferentes en un mismo test. . . . .	20
4.3	Número de respuestas diferentes entre los test para cada estudiante. .	21
4.4	Frecuencias absolutas de las diferencias en las respuestas entre test por estudiante y grupo. . . . .	22
4.5	Frecuencias relativas de las respuestas al test. . . . .	24
4.6	Frecuencias relativas de las respuestas por pregunta. . . . .	25
4.7	Preguntas ordenadas por valoración. . . . .	26
4.8	Frecuencia de preguntas que mejoran por estudiante entre subtitula- dos ( $A > B$ ) . . . . .	27
4.9	Frecuencia de preguntas que mejoran por estudiante entre subtitula- dos (positive A vs negative B) . . . . .	27
5.1	OR ~ Treat + Period + Response . . . . .	37



CAPÍTULO



# INTRODUCCIÓN



## MARCO TEÓRICO Y ESTADO DEL ARTE

### 2.1 Características del diseño del experimento

El objetivo del estudio es responder a la pregunta de investigación:

**i** Pregunta de investigación

¿Son los estudiantes de un curso de creación de materiales accesibles capaces de encontrar diferencias en la calidad del subtítulo de un vídeo?

El diseño del experimento es completamente aleatorizado, de respuesta ordinal, cruzado  $AB/BA$  y doble ciego. Es decir, que la asignación de los estudiantes a cada grupo fue aleatoria; cada grupo vio los vídeos en orden inverso; los estudiantes no conocían a priori qué vídeo estaban viendo en cada momento y tampoco se disponía de esta información en el momento de realizar el modelado estadístico.

Un diseño **completamente aleatorizado** (Lawson 2015, pp. 18) «garantiza la validez del experimento contra sesgos causados por otras variables ocultas. Cuando las unidades experimentales se asignan aleatoriamente a los niveles de factor de tratamiento, se puede realizar una prueba exacta de la hipótesis de que el efecto del tratamiento es cero utilizando una prueba de aleatorización».

En este trabajo se estudiarán las diferencias existentes entre los dos niveles de subtítulo a través de las respuestas de los alumnos a las escalas de Likert de cada uno de los vídeos. Para ello se propondrán modelos estadísticos adecuados al diseño del experimento. La primera cuestión que debemos abordar es que el diseño sea cruzado. Siguiendo a Senn (2022), para que el ensayo sea de tipo cruzado no sería suficiente intercambiar las secuencias sino que debe ser objeto

del ensayo el estudio de las diferencias entre los tratamientos individuales que componen las secuencias. Los principales problemas de un diseño cruzado son el abandono, drop-out, de alguno de los participantes y la interacción entre el tratamiento y el periodo o carry-over. Además, el análisis estadístico es más complicado y particularmente cuando la respuesta es ordinal y hay más de dos tratamientos. En la misma línea, Lui (2016) afirma que «el objetivo principal de un diseño cruzado es estudiar la diferencia entre tratamientos individuales (en lugar de la diferencia entre secuencias de tratamiento). Debido a que cada paciente sirve como su propio control, el diseño cruzado es una alternativa útil al diseño de grupos paralelos para aumentar la potencia».

En un diseño cruzado debemos preocuparnos de la existencia de los efectos periodo y secuencia (o carry-over). El **efecto periodo** aplicado al experimento del subtítulo, se producirá si las respuestas del segundo periodo están influidas por haber realizado la primera actividad de subtítulo. El **efecto secuencia** se producirá si las respuestas fueran diferentes cuando se realizan en un orden que cuando se realizan en el otro.

La segunda cuestión de relevancia es que las respuestas a los ítems de una escala de Likert son de **tipo ordinal**. Los test estadísticos ANOVA o MANOVA presuponen que la variable de respuesta es cuantitativa y con distribución normal. Tratar las respuestas a una escala de Likert como si fueran cuantitativas no es correcto por las siguientes razones:

- Los niveles de respuesta no son necesariamente equidistantes: la distancia entre un par de opciones de respuesta puede no ser la misma para todos los pares de opciones de respuesta. Por ejemplo, la diferencia entre «Muy en desacuerdo» y «En desacuerdo» y la diferencia entre «De acuerdo» y «Muy de acuerdo» es de un nivel, pero psicológicamente puede ser percibida de forma diferente por cada sujeto.
- La distribución de las respuestas ordinales puede ser no normal. En particular esto sucederá si hay muchas respuestas en los extremos del cuestionario.
- Las varianzas de las variables no observadas que subyacen a las variables ordinales observadas pueden diferir entre grupos, tratamientos, periodos, etc.

En Liddell y Kruschke (2018) se han analizado los problemas potenciales de tratar datos ordinales como si fueran cuantitativos constatando que se pueden presentar las siguientes situaciones:

- Se pueden encontrar diferencias significativas entre grupos cuando no las hay: error tipo I.
- Se pueden obviar diferencias cuando en realidad sí existen: error tipo II.
- Incluso se pueden invertir los efectos de un tratamiento.
- También puede malinterpretarse la interacción entre factores.



Otro factor que hay que tener en cuenta es que, al tratarse de un diseño cruzado, es de **medidas repetidas** ya que cada sujeto realiza dos veces el test, uno con cada vídeo y que, por lo tanto, las respuestas a cada test de un mismo sujeto no son independientes. Además, tampoco podemos considerar independientes los ítems que componen el test ya que los ítems pretenden medir la misma variable latente: la calidad del subtitulado.

## 2.2 Técnicas estadísticas utilizadas

En los siguientes apartados se proponen distintas técnicas que pretenden responder al objetivo del trabajo teniendo en cuenta las peculiaridades del diseño del experimento comentadas en el apartado anterior.

### Correlación entre preguntas con el alfa de Cronbach

Normalmente las preguntas de un cuestionario pretenden medir una variable que está oculta o latente. En nuestro caso es la calidad del subtitulado. Las respuestas a estas preguntas relacionadas deben ser consistentes internamente, es decir, las respuestas deben correlacionarse fuerte y positivamente.

Un índice que se utiliza habitualmente para medir la consistencia interna de un cuestionario es el coeficiente alfa de Cronbach (ver Schweinberger 2020). Se define de esta forma:

$$\alpha = \frac{N}{N-1} \left( 1 - \frac{\sum_{i=1}^N s_i^2}{s^2} \right) \quad (2.1)$$

Donde:

- $\alpha$  es el coeficiente alfa de Cronbach.
- $N$  es el número de ítems de la escala de Likert.
- $s_i^2$  es la varianza de la puntuación del ítem  $i$ .
- $s^2$  es la varianza total de las puntuaciones de todos los ítems.

Valores cercanos 1 indican una fuerte correlación en las respuestas y se admite que las preguntas del cuestionario están midiendo la misma variable latente.

## 2.3 Métodos exploratorios

En esta sección se aplicarán técnicas estadísticas que se basan en tablas de contingencia. Una descripción teórica de este tipo de técnicas se pueden encontrar en Agresti (2018). Un tratamiento aplicado y basado en gráficos, que será el enfoque que seguiremos en este trabajo, es realizado en Friendly et al. (2015).

### Asociación de variables con la prueba de homogeneidad $\chi^2$

La prueba de homogeneidad  $\chi^2$  (ver Emilio y Molanes-López [s.f.](#)) está enmarcada en el esquema *Nominal*  $\leftarrow$  *Nominal* y contrasta la hipótesis  $H_0$  de que no hay diferencias entre grupos frente a  $H_1$  de que existen diferencias. Dada una tabla de contingencia (ver Tabla 2.1). El valor del estadístico se calcula:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - c_{ij})^2}{c_{ij}}$$

que sigue una  $\chi^2_{(r-1)(c-1)}$ , donde  $c_{ij}$  son las frecuencias esperadas bajo  $H_0$  y que se calculan:

$$c_{ij} = \frac{r_i c_j}{n}$$

Cuadro 2.1: Tabla de contingencia

	X = 1	X = 2	X = 3	
Y = 1	$n_{11}$	$n_{12}$	$n_{13}$	$r_1$
Y = 2	$n_{21}$	$n_{22}$	$n_{23}$	$r_2$
	$c_1$	$c_2$	$c_3$	$n$

### Comparación con *Odds Ratio*

Dada una tabla de contingencia 2 X 2 (ver Tabla 2.2), el *oddsratio* poblacional (*OR*) se define como el cociente de las probabilidades complementarias (ver Emilio y Molanes-López ([ibíd.](#))). Dado que los factores *Treat*, *Period* y *Seq* tienen todos 2 niveles, podemos contrastar si hay interacción entre ellos para cada nivel de respuesta. Para ello contrastamos  $H_0$  de que  $OR = 1$  frente a  $H_1$  ( $OR \neq 1$ ) <sup>1</sup>.

Cuadro 2.2: Tabla de contingencia

	X = 1	X = 0
Y = 1	$n_{11}$	$n_{12}$
Y = 0	$n_{21}$	$n_{22}$

## 2.4 Modelos lineales generalizados

Los modelos lineales generalizados (*GLM*) son modelos en los que la variable respuesta no es normal. Para especificar un *GLM* son necesarios tres componentes (ver Agresti 2018, pp. 66-67):

<sup>1</sup>Ver Emilio y Molanes-López ([s.f.](#)) para el procedimiento del contraste de hipótesis mediante *log OR*.

- Un componente aleatorio: será una distribución de probabilidad de la familia exponencial que se asume que sigue la variable respuesta,  $Y$ .
- Un componente lineal de predictores.

$$\alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

- Una función de enlace  $g$  que relaciona  $\mu = E(Y)$  con los predictores, de tal forma que:

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

En GLM se realiza maximizando la función de verosimilitud (*MLE*). Es decir, que los coeficientes del modelo son aquellos que maximizan la probabilidad de los datos.

## Regresión Logística

La regresión logística (ver [ibíd.](#), pp. 68-69) es un caso particular de *GLM* donde la variable respuesta,  $Y$ , es Bernoulli o Binomial. Es decir, que  $Y$  toma valores 0 ó 1. En una función de Bernoulli de parámetro  $\pi$ ,  $E[Y] = P(Y = 1) = \pi$ . Necesitamos una función que «mapee» los valores que puede tomar el componente lineal de rango  $(-\infty, +\infty)$  a los valores que puede tomar  $\pi$  en el rango  $(0, 1)$ . Una función que puede hacer esto es la función *logit*:

$$\text{logit}(Y = 1) = \log \left[ \frac{P(Y = 1)}{1 - P(Y = 1)} \right] = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

por lo que:

$$P(Y = 1) = \frac{\exp^{\alpha + \beta x}}{1 + \exp^{\alpha + \beta x}}$$

Para interpretar los coeficientes, podemos reescribir la ecuación (ver Friendly et al. [2015](#), p. 260):

$$\frac{P(Y = 1 | X = x)}{1 - P(Y = 1 | X = x)} = e^{\alpha} e^{\beta x}$$

Para una unidad de incremento de  $x$ :

$$\frac{P(Y = 1 | X = x + 1)}{1 - P(Y = 1 | X = x + 1)} = e^{\alpha} e^{\beta(x+1)}$$

Dividiendo la segunda ecuación entre la primera:

$$\frac{\frac{P(Y=1|X=x+1)}{1-P(Y=1|X=x+1)}}{\frac{P(Y=1|X=x)}{1-P(Y=1|X=x)}} = \frac{e^{\alpha} e^{\beta(x+1)}}{e^{\alpha} e^{\beta(x)}} = e^{\beta}$$

Es decir:

- $\beta$  es el *log OR* asociado a una unidad de incremento de  $x$ .
- $\alpha$  es el *log odds* cuando  $x = 0$ .

El contraste de hipótesis para los coeficientes  $\beta$ :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

se puede realizar con el test de Wald:

$$W = \frac{\hat{\beta}_j - 0}{se(\hat{\beta}_j)} \sim N(0, 1)$$

o con el test de razón de verosimilitudes (LRT):

$$\begin{aligned} \Lambda &= -2 \log \frac{L(\widehat{reduced})}{L(\widehat{full})} \\ &= -2 \log L(\widehat{reduced}) + 2 \log L(\widehat{full}) \sim \chi_r^2 \text{ (donde } r \text{ es el número de betas} = 0) \end{aligned}$$

Para comparar modelos no anidados, se puede usar el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC), que se definen:

$$AIC : -2 \log L + 2p$$

$$BIC : -2 \log L + p \log(n)$$

, donde  $L$  es el valor de máxima verosimilitud.

### Regresión Ordinal

Las respuestas a los ítems de una escala de Likert son ordinales. La Regresión Ordinal es clase de *GLM* que comparte muchas similitudes con la regresión logística (ver Sección 2.4). Según Bürkner y Vuorre (2019) hay tres clases de Regresión Ordinal:

- Regresión Ordinal Acumulativa.
- Regresión Ordinal Secuencial.
- Regresión Ordinal Adyacente.

Nos centraremos en la primera ya que es la más habitual y adecuada para nuestro caso (ver [ibíd.](#), pp. 23-24) <sup>2</sup>. El modelo acumulativo, CM, presupone que la variable ordinal observada,  $Y$ , proviene de la categorización de una variable

---

<sup>2</sup>Las regresiones ordinales secuencial y adyacente presuponen que para alcanzar un nivel se ha tenido que pasar previamente por los anteriores, que no es nuestro caso.

latente (no observada) continua,  $\tilde{Y}$ . Hay  $K$  umbrales  $\tau_k$  que particionan  $\tilde{Y}$  en  $K + 1$  categorías ordenadas observables (ver Figura 2.1). Si asumimos que  $\tilde{Y}$  tiene una cierta distribución (por ejemplo, normal) con distribución acumulada  $F$ , se puede calcular la probabilidad de que  $Y$  sea la categoría  $k$  de esta forma:

$$Pr(Y = k) = F(\tau_k) - F(\tau_{k-1})$$

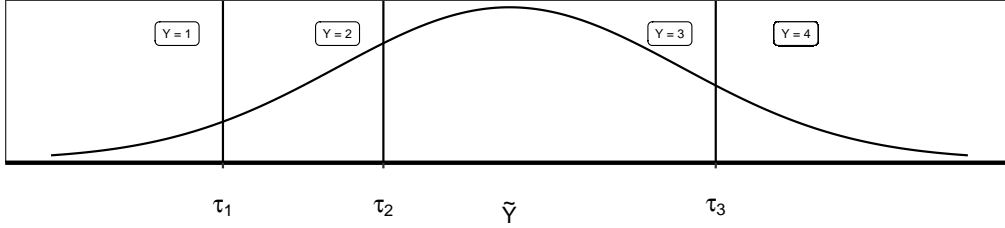


Figura 2.1: Función latente en una regresión ordinal acumulativa.

Por ejemplo en la Figura 2.1,

$$Pr(Y = 2) = F(\tau_2) - F(\tau_1)$$

Si suponemos que  $\tilde{Y}$  tiene una relación lineal los predictores:

$$\tilde{Y} = \eta + \epsilon = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

y que los errores son  $N(0, \sigma^2)$ . Entonces la función de probabilidad acumulada de los errores tendrá la misma forma que la de  $\tilde{Y}$ :

$$Pr(\epsilon \leq z) = F(z)$$

Y podremos calcular la distribución de probabilidad acumulada de  $Y$ :

$$Pr(Y \leq k | \eta) = Pr(\tilde{Y} \leq \tau_k | \eta) = Pr(\eta + \epsilon \leq \tau_k) = Pr(\epsilon \leq \tau_k - \eta) = F(\tau_k - \eta)$$

Por lo que asumiendo la normalidad de los errores:

$$Pr(Y = k) = \Phi(\tau_k - \eta) - \Phi(\tau_{k-1} - \eta)$$

Donde hay que estimar los umbrales y los coeficientes de regresión. La función anterior es la conocida como la función de enlace probit. Otra función de enlace popular es la función logit. Es la que usaremos en este trabajo por ser más fácil su interpretación<sup>3</sup>. Con esta función de enlace la interpretación de los

<sup>3</sup>En la práctica los coeficientes estimados con las funciones de enlace probit y logit suelen similares.

coeficientes es parecida a la de los coeficientes de la regresión logística. Se parte del supuesto de que el *logit* de la función de probabilidad es lineal:

$$\text{logit}[P(Y \leq k)] = \tau_k - \eta = \tau_k - (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

En ese caso, se puede demostrar fácilmente que, por ejemplo:

$$\frac{\frac{\Pr(Y \leq k | \eta)}{\Pr(Y > k | \eta)}}{\frac{\Pr(Y \leq k+1 | \eta)}{\Pr(Y > k+1 | \eta)}} = \exp(\tau_k - \tau_{k+1})$$

Y que <sup>4</sup>:

$$\frac{\frac{\Pr(Y \leq k | x_i = 1)}{\Pr(Y > k | x_i = 1)}}{\frac{\Pr(Y \leq k | x_i = 0)}{\Pr(Y > k | x_i = 0)}} = \exp(-\beta_i)$$

o, equivalentemente:

$$\frac{\frac{\Pr(Y > k | x_i = x+1)}{\Pr(Y \leq k | x_i = x+1)}}{\frac{\Pr(Y > k | x_i = x)}{\Pr(Y \leq k | x_i = x)}} = \exp(\beta_i)$$

Es decir, que  $\exp(\beta_i)$  es el *OR* (cambio en *odds*) de que la variable respuesta esté por encima de una determinada categoría versus estar por debajo de ella para una unidad de incremento del predictor  $x_i$ . Este modelo se denomina proporcional ya que cada predictor se asume que tiene los mismos efectos sobre todas las categorías de la variable de respuesta ordinal (ver Liu 2022). Un valor del coeficiente  $\beta_i$  positivo indica que la relación entre el predictor  $x_i$  y la función de *logit* es positiva y, por lo tanto, se incrementa la posibilidad de un mayor valor de la variable respuesta. Como veremos, esta suposición se puede relajar y permitir que los coeficientes de todos o de algunos de los predictores sean diferentes para cada pareja consecutiva de valores de respuesta. Tendríamos entonces más parámetros a estimar con una interpretación más compleja.

---

<sup>4</sup>En la Sección 4.4 se demuestra esta fórmula.

## MÉTODOLÓGIA/MATERIALES Y MÉTODOS

### 3.1 Descripción de la experiencia

#### Marco de la experiencia

Los datos se recogieron de una actividad de subtítulo que se propuso a los estudiantes de la edición 2022 del curso MOOC Materiales digitales accesibles de la UNED Abierta. Este curso pertenece al Canal Fundación ONCE y según se recoge en la propia página Web del curso sus objetivos son:

- Reconocer y abordar los desafíos a los que se enfrentan las personas con discapacidad (por ejemplo los estudiantes) al usar los documentos electrónicos, adquirir conciencia y experiencia.
- Obtener una mejor comprensión de la accesibilidad como un asunto de derechos civiles y desarrollar los conocimientos y habilidades necesarios para diseñar recursos de aprendizaje que promuevan ambientes de aprendizaje inclusivos.
- Valorar cómo los documentos accesibles benefician a todas las personas, incluyendo a las que tienen y a las que no tienen discapacidad, a través de una mayor facilidad de uso y la interoperabilidad de los materiales basados en la web.
- Tomar conciencia de que cómo los autores pueden no solo eliminar barreras, sino evitar crearlas en primer lugar.
- Adquirir autosuficiencia en la producción de contenidos accesibles y en la identificación de problemas de accesibilidad.

- Adquirir, como autores, estrategias para la producción sostenible de material digital, como la modularidad.

Está destinado «a todas aquellas personas que desean participar en el desarrollo de soluciones de diseño éticas y creativas, que quieran escribir o gestionar contenidos electrónicos accesibles, desde páginas web o aplicaciones a libros electrónicos»ePub”.

El curso tiene cuatro módulos:

- Introducción.
- Accesibilidad de material multimedia.
- Accesibilidad de texto digitales.
- Materiales digitales en la práctica.

#### **Actividad de subtitulado**

La actividad de subtitulado es voluntaria y sin influencia en la calificación final del alumno ni el material al que el tiene acceso. Se realiza en el módulo «Accesibilidad del material multimedia». En este mismo módulo, y antes de proponerles la actividad de subtitulado, los alumnos previamente completaron las secciones «Accesibilidad de la información sonora» y «Accesibilidad de la información visual», con lo cual ya tienen conocimiento sobre creación de vídeos y subtítulos accesibles.

La actividad consistió en ver dos vídeos idénticos y que solo se diferencian en la calidad del subtitulado de 43 segundos de duración. Los subtítulos de uno de los vídeos se realizaron (ver Pérez Martín et al. 2021; Molanes-López et al. 2021) siguiendo la guía Web Content Accessibility Guidelines 2.1 (WCAG 2.1) del W3C (World Wide Web Consortium). El otro vídeo tenía un subtitulado similar pero se introdujeron pequeñas deficiencias, algunas de ellas inapreciables para alguien que carezca de conocimientos sobre accesibilidad. El orden de los vídeos es aleatorio, de tal forma que una cohorte de alumnos vio primero el vídeo bien subtitulado y luego el mal subtitulado y la otra lo hizo al revés. Después de ver cada uno de los vídeos, los alumnos respondieron a una escala de Likert de 5 niveles y 18 ítems. Los 18 ítems de Likert responden a los criterios de la norma UNE 153010 (ver AENOR 2012). El diseño de experimento es doble ciego: es decir, a los alumnos no se les informó de si estaban viendo el vídeo con mejor o con peor calidad de subtitulado. Esta información tampoco se conoce en el momento de realizar este trabajo ya que los vídeos tienen identificaciones genéricas que no contienen ninguna indicación del tipo de subtitulado del vídeo<sup>1</sup>.

---

<sup>1</sup>En la respuesta a cada ítem, el alumno puede añadir comentarios. Éstos han sido eliminados del estudio para que no filtren información referente al tipo de subtitulado que el alumno cree estar contestando.



En la Tabla 3.1 se muestran los 5 niveles de cada uno de los ítems de la escala de Likert utilizados para valorar el subtitulado<sup>2</sup>. En la Tabla 3.2 se muestran los 18 ítems de la escala de Likert que se propuso a los alumnos para que evaluaran cada uno de los vídeos.

Cuadro 3.1: Niveles de los ítems de la escala de Likert.

values	levels
0	No sé / No contesto
1	Muy en desacuerdo
2	En desacuerdo
3	Neutral
4	De acuerdo
5	Muy de acuerdo

Cuadro 3.2: Ítems de la escala de Likert.

Item	Texto
Q01	La posición de los subtítulos.
Q02	El número de líneas por subtítulo.
Q03	La disposición del texto respecto a la caja donde se muestran los subtítulos.
Q04	El contraste entre los caracteres y el fondo.
Q05	La corrección ortográfica y gramatical.
Q06	La literalidad.
Q07	La identificación de los personajes.
Q08	La asignación de líneas a los personajes en los diálogos.
Q09	La descripción de efectos sonoros.
Q10	La sincronización de las entradas y salidas de los subtítulos.
Q11	La velocidad de exposición de los subtítulos.
Q12	El máximo número de caracteres por línea.
Q13	La legibilidad de la tipografía.
Q14	La separación en líneas diferentes de sintagmas nominales, verbales y preposicionales.
Q15	La utilización de puntos suspensivos.
Q16	La escritura de los números.
Q17	Las incorrecciones en el habla.
Q18	Los subtítulos del vídeo cumplen en general con los requisitos de accesibilidad.

<sup>2</sup>En la codificación original los valores asignados a cada respuesta eran diferentes: la opción «No sé / No contesto» se codificó con 5 y las demás opciones con una unidad menos que la mostrada. En este trabajo se ha hecho una rotación para asignar valores más usuales en la literatura científica sobre el tema.

Cuadro 3.3: Tablas de contingencia de la información socioeconómica de los estudiantes.

gender	Freq
f	46
m	19
NA	22

Estudiantes por sexo.

year_of_birth	Freq
None	22
NA	1

Estudiantes con valor nulo en el campo  
año de nacimiento.

level_of_education	Freq
a	25
b	8
hs	2
m	15
other	2
p	10
NA	25

Estudiantes por nivel educativo.

level_of_knowledge	Freq
4	1
6	2
7	15
8	22
9	20
10	16
NA	11

Estudiantes en función del número de  
preguntas acertadas en el test de  
conocimiento sobre accesibilidad.

## 3.2 Participantes

Los datos personales de los estudiantes se suministraron anonimizados para evitar conocer su identidad. De acuerdo con nuestro compromiso ético, del estudio se han eliminado a aquellos estudiantes que, a pesar de haber realizado la actividad, no dieron su consentimiento para que sus datos se utilizaran en estudios científicos. Tras este proceso, se dispone de 198 cuestionarios correspondientes a 111 alumnos. Hay 24 estudiantes que sólo realizaron el primero de los test. Como la muestra es suficientemente amplia, se ha decidido eliminar estos test quedando 87 estudiantes participantes. En la Tabla 3.3 se muestran las tablas de contingencia de algunos de los datos que los estudiantes voluntariamente facilitaron en el cuestionario inicial del curso.

## 3.3 Ficheros suministrados

Se dispuso de los siguientes ficheros csv:

- El fichero `grade` contiene el identificador de estudiante y el grupo al que pertenece (campo `cohort`).

- El fichero `abo` es la información socioeconómica que voluntariamente ha aportado el estudiante: sexo, año nacimiento, nivel de estudios, ocupación.
- El fichero `conoc` contiene el test de evaluación inicial de conocimientos del estudiante.
- El fichero `exp` es la evaluación del curso realizada por cada estudiante.
- El fichero `acc` contiene la información sobre las necesidades/preferencias de accesibilidad que tiene el estudiante.
- Los ficheros `test1` y `test2` son las repuestas al test de Likert sobre la calidad del subtítulo del primer y del segundo vídeo realizado por cada grupo respectivamente.

## 3.4 Preprocesado

En esta sección se describen las transformaciones realizadas con los ficheros suministrados:

- Se lee el fichero de perfil del usuario. El número de fila con el que el usuario aparece en el fichero se utilizará como identificador del usuario para mantener la trazabilidad y comprobar que las transformaciones realizadas son correctas.
- Se eliminan los datos de los estudiantes que aún habiendo realizado la actividad, no han dado su consentimiento para participar en el estudio.
- El valor del campo `cohort` se sustituye por una letra, *A* o *B*, en función del grupo asignado. En el momento de realizar este proceso se desconoce qué vídeo vio primero cada cohorte.
- Se lee el fichero `profile` y se añade información sobre el sexo, el año de nacimiento y el nivel de estudios.
- Se lee el fichero `conoc` y se calcula cuántas preguntas acertó cada usuario en el test de evaluación de conocimientos previos. Se añade esta información al perfil del usuario.
- Se leen los ficheros de test y se procesan. Se utiliza el nombre del fichero (`test1` o `test2`) para saber de qué vídeo se está respondiendo el test <sup>3</sup>.
- Se seleccionan las preguntas que contienen las respuestas y se renombran para que sea más fácil saber de qué pregunta se trata <sup>4</sup>. Se convierte el campo `LastTry`, que contiene la fecha y hora de realización del test, a formato fecha y hora.

---

<sup>3</sup>Se reitera que en el momento de realizar este proceso se desconoce si el vídeo es el correctamente subtítulo o el otro. La única información que se almacena es si se está respondiendo al vídeo que se vio primero.

<sup>4</sup>En los ficheros suministrados la respuesta a cada pregunta ocupa varios campos. Se selecciona en cada pregunta el que contiene el valor de la respuesta y se convierte a numérico.

- Se realizan algunas comprobaciones como la ausencia de valores nulos en las variables más relevantes o que no existan inconsistencias ni errores de procesado.
- Se eliminan los comentarios y se graban en fichero aparte para que no revelen información que podría descubrir el tipo de subtítulo que piensa que está evaluando el estudiante.
- Se almacenan los resultados de los test preprocesados en un fichero csv.

### 3.5 Variables del modelo.

En la Tabla 3.8 se describen las características más relevantes de las principales variables que se utilizarán en el modelado y en el análisis estadístico.

Cuadro 3.8: Descripción de las variables más importantes.

Nombre	Descripción	Tipo	Valores
Response	Respuesta a las preguntas del test.	Factor ordenado	De 0 a 5 <sup>1</sup>
Level	Valoración de la respuesta.	Factor ordenado	Negative, Neutral, Positive <sup>2</sup>
Treat	Subtítulos	Factor	A o B <sup>3</sup>
Period	Periodo	Factor	1 ó 2 <sup>4</sup>
Seq	Secuencia de aplicación de los tratamientos.	Factor	AB o BA
Subject	Identificación del estudiante	Factor	Numérico
Question	Número de la pregunta	Factor	Q01, Q02, ..., Q18 <sup>5</sup>

<sup>1</sup>Se ha hecho una rotación sobre los valores originales. 0 = No sé, 1 = Muy en desacuerdo, ..., 5 Muy de acuerdo.

<sup>2</sup>Positive cuando Response sea 4 ó 5, Negative cuando sea 1 ó 2 y Neutral para 3.

<sup>3</sup>No se conoce si el tratamiento A es el subtítulo bueno o lo es el B.

<sup>4</sup>1 para el primer vídeo visto y 2 el segundo.

<sup>5</sup>Se ha reorganizado de tal forma que Q18, que es la pregunta resumen, sea el valor primero y de referencia.

Partiendo del dataframe que se construyó en el preprocesado (ver Sección 3.4) construimos el dataframe que usaremos a partir de este momento. Las operaciones principales que se han realizado han sido:

- Renombrar las variables (ver Tabla 3.8).
- Eliminar del estudio los usuarios que solo han realizado uno de los test.
- Transformar las variables que lo requieran en factores. La pregunta 18 se usará como referencia en el factor Question.
- Rotar los valores de respuesta para que «No sé / No contesto» tenga valor 0 y el resto de 1 a 5 desde «Muy en desacuerdo», 1, hasta «Muy de acuerdo», 5.
- Crear el factor Level con los niveles negative, neutral y positive dependiendo de si la respuesta es 1 ó 2, 3, 4 ó 5 respectivamente.
- Transformar el dataframe de formato ancho a largo: los ficheros de respuestas se suministran en formato ancho. Es decir, que cada fila es un test

que contiene 18 columnas para las respuestas a cada pregunta. Los nombres de las columnas son  $Q01$ ,  $Q02$ , ...,  $Q18$  y tendrán valores de 0 a 5 con las respuestas. La mayoría de los paquetes de R que vamos a usar requieren que los datos estén en formato largo. Esto que quiere decir que cada fila tendrá una única respuesta por lo que habrá únicamente dos columnas, *Question* y *Response*. En la primera se almacenará el identificador de la pregunta ( $Q01$ ,  $Q02$ , ...,  $Q18$ ) y en la segunda el valor de la respuesta (de 0 a 5). De esta forma, un test pasará de ocupar una fila y 18 columnas en el formato ancho a 18 filas y dos columnas en el largo.

Se crean dos dataframes:

- `df_all` contiene en formato largo todas las respuestas a los test.
- `df_clean` tiene la misma estructura que `df_all` pero en él se han eliminado las respuestas «No sé / No contesto».

`df_all` se utilizará cuando se traten las respuestas como categóricas y, por lo tanto, como no ordenadas. `df_clean` se utilizará cuando se traten las respuestas como ordenadas y por ello no contiene las respuestas con valor «No sé / No contesto».

La estructura de estos dataframes es la siguiente:

```
tibble [2,980 x 7] (S3: tbl_df/tbl/data.frame)
 $ Seq      : Factor w/ 2 levels "AB","BA": 1 1 1 1 1 1 1 1 1 1 ...
 $ Period   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
 $ Treat    : Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
 $ Subject  : Factor w/ 87 levels "4","33","35",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Question : Factor w/ 18 levels "Q18","Q01","Q02",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ Response : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<...: 3 3 3 3 3 3 3 3 3 3 ...
 $ Level    : Ord.factor w/ 3 levels "Negative"<"Neutral"<...: 2 2 2 2 2 2 2 2 2 2..
```

En la Tabla 3.9 se muestran algunos ejemplos datos.

Cuadro 3.9: Muestra del dataframe preparado para el modelado estadístico en formato largo.

Seq	Period	Treat	Subject	Question	Response	Level
BA	1	B	871	Q05	5	Positive
BA	1	B	371	Q03	2	Negative
BA	1	B	749	Q02	4	Positive
BA	1	B	749	Q13	4	Positive
AB	1	A	380	Q07	4	Positive
BA	2	A	901	Q04	5	Positive
BA	1	B	110	Q06	5	Positive
AB	1	A	339	Q02	5	Positive
AB	1	A	893	Q08	2	Negative
BA	1	B	220	Q03	3	Neutral



## MODELADO ESTADÍSTICO

### 4.1 Análisis inicial

Como se explica en la Tabla 3.8, al subtítulo le denominamos tratamiento y a sus niveles (correcto e incorrecto) los hemos llamado  $A$  y  $B$  sin hacer ninguna conjetura de cual de los dos es el subtítulo correcto. El grupo con secuencia  $AB$  será el que primero vio el vídeo con subtítulo  $A$  y luego el  $B$ . Análogamente, el grupo con secuencia  $BA$  vio los vídeos en orden inverso. Recuérdese que el nivel 0 de respuesta se corresponde con «No sé / No contesto» (ver Tabla 3.1). Tras eliminar los test de los usuarios que no dieron su consentimiento para participar en el estudio y los de los que no realizaron el segundo test, las dos cohortes están equilibradas (ver Figura 4.1).

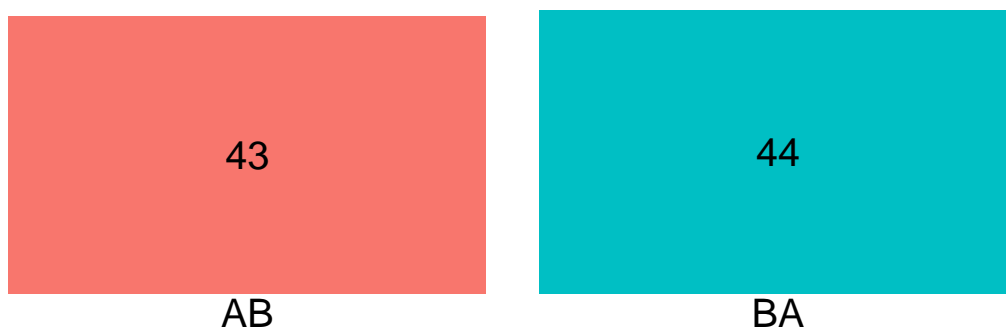


Figura 4.1: Estudiantes asignados a cada grupo.

### Análisis de la calidad de los datos

En esta sección buscamos si hay test que tienen valores de respuesta que puedan resultar anómalos. En los test no se ha observado ningún valor nulo ni erróneo.

El campo LastTry contiene la fecha y hora de realización del test. Con esta información podemos conocer el tiempo que empleó cada estudiante entre actividades. La Tabla 4.1 muestra que hay algunos test que se hicieron demasiado rápido <sup>1</sup>.

Cuadro 4.1: Tiempos de realización de la segunda actividad de duración inferior a 2 minutos.

Minutes
0.93
1.3
1.7
1.72
1.78
1.97

La Figura 4.2 muestra que hay 28 test en los que el estudiante contestó a todas las preguntas usando únicamente 2 respuestas diferentes. Además hay 13 test en los que se contestaron todas las preguntas con 1 respuesta.

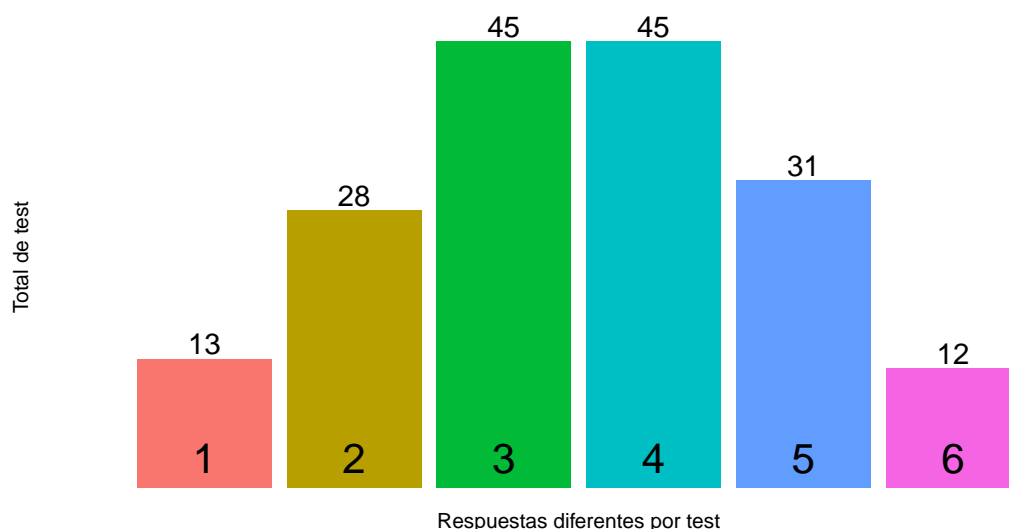


Figura 4.2: Número de respuestas diferentes en un mismo test.

La tabla Tabla 4.2 muestra los test de respuesta única y el valor de esa respuesta.

Cuadro 4.2: Test en los que todas las preguntas se contestan el mismo valor de respuesta.

Response	Seq	Test
2	AB	01
2	AB	02

<sup>1</sup>Hay que tener en cuenta que la duración de vídeo es de algo más de 40 segundos y que los estudiantes tienen que contestar un test de 18 ítems.



---

3	BA	01
3	BA	02
3	BA	02
3	BA	02
4	AB	01
4	AB	01
4	AB	02
4	BA	01
4	BA	02
4	BA	02
4	BA	02

---

La Figura 4.3 presenta la distribución de la cantidad de respuestas cuyo valor cambia entre los dos test que realiza cada estudiante.

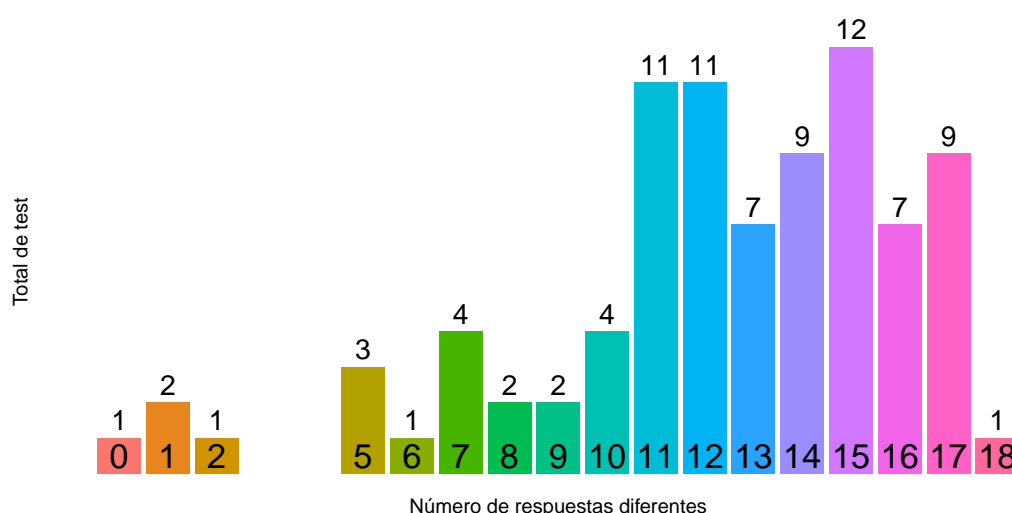


Figura 4.3: Número de respuestas diferentes entre los test para cada estudiante.

Tan solo 1 estudiante respondió a todas las preguntas con el mismo valor en los dos test. Por otro lado, no hay test que tengan un número excesivo de contestaciones «No sé/No contesto» (ver Tabla 4.3).

Cuadro 4.3: Los 5 test con más respuestas ‘No sé/No contesto’

---

Test	Total respuesta por test
01	5
01	5
02	5
02	5
01	4

---

Vemos que algunos test tienen valores que no parecen muy razonables. Por ejemplo, no parece razonable realizar la actividad en menos de 2 minutos. Se observa que en algunos test hay poca variabilidad. Sin embargo, no son muchos los test con estas características así que se ha decidido mantener estos datos a pesar

de que se pueda dudar de si en ellos los estudiantes contestaron con la debida atención y diligencia.

### Comparación de los tratamientos $A$ y $B$ entre grupos.

La Figura 4.4 presenta una forma de comparar los dos test realizados por los estudiantes. Para cada estudiante se comparó pregunta a pregunta sus dos test y se contabilizó la diferencia entre el número de preguntas en que la puntuación en el segundo vídeo fue superior y en las que lo fue inferior (las que no variaron de puntuación no se consideraron). En el eje  $x$  se muestra la diferencia entre preguntas. Cantidades negativas indican que hay más respuestas en el segundo de los test que han empeorado respecto al primero de las que han mejorado. En el eje  $y$  se representa el número de estudiantes para cada diferencia. Esta frecuencia se representa en negativo cuando la diferencia es negativa <sup>2</sup>. Esto es una forma de evaluar si el estudiante valoró mejor o no el segundo vídeo que el primero.

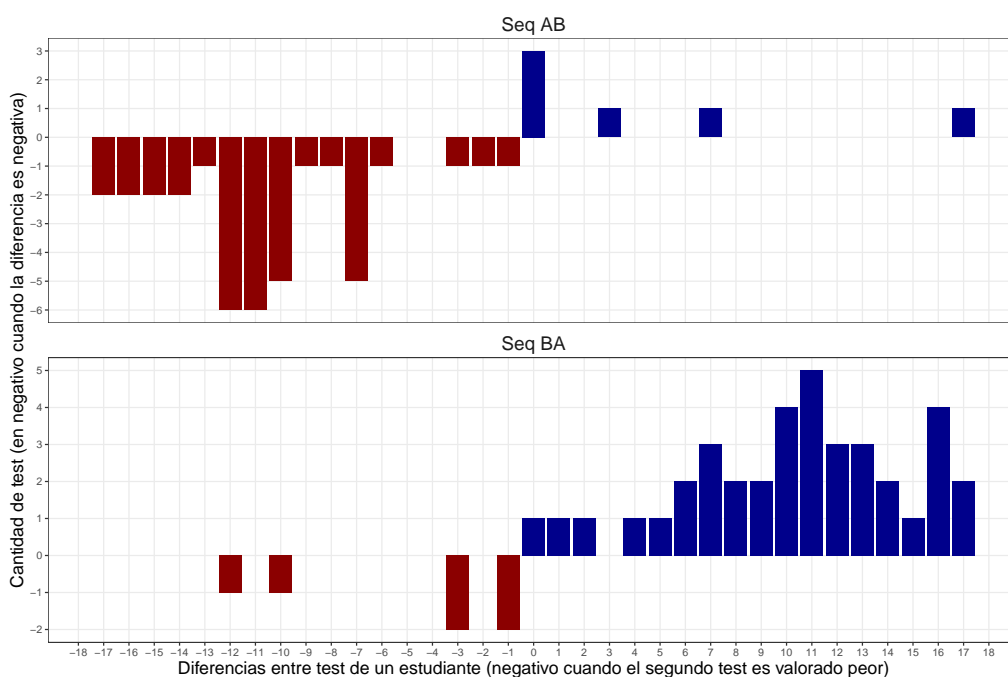


Figura 4.4: Frecuencias absolutas de las diferencias en las respuestas entre test por estudiante y grupo.

Vemos que en el grupo  $AB$  las diferencias tienden a ser negativas y en el  $BA$  positivas. Esto estaría indicando que los estudiantes valoran mejor el subtitulado de nivel  $A$  en ambas secuencias. Por ello es esperable que las respuestas de los estudiantes del grupo  $AB$  hayan empeorado y que las diferencias sean negativas y que lo contrario haya sucedido con las del grupo  $BA$ . La diferencia más frecuente en el grupo  $AB$  es 12 y en el grupo  $BA$  este valor es 11.

<sup>2</sup>En la comparación se han omitido aquellas respuestas en las que el estudiante contestó «No sé/No contesto» en la pregunta correspondiente de uno de los test.

Resulta llamativo que haya estudiantes cuyas contestaciones estén tan alejadas de la tendencia de su grupo. En la Tabla 4.4 se muestran los tiempos que han transcurrido entre la realización de los test de aquellos estudiantes cuyas respuestas difieren de forma importante de su grupo. Se observa que casi todos son tiempos entre actividades muy cortos. En cualquier caso y, como no son muchos, se ha decidido no eliminarlos y realizar el análisis con ellos.

Cuadro 4.4: Estudiantes que tienen diferencias en sus respuestas muy alejadas de la tendencia de su grupo.

Seq	Diff	Minutes
AB	17	1.3
AB	7	3.33
BA	-10	50345.95
BA	-12	1.7

En la Figura 4.5 representamos la frecuencia relativa del valor de respuesta para cada grupo y test en todas la preguntas <sup>3</sup>. Esta es otra forma de comparar los niveles de subtitulado.

Cuadro 4.5: Resumen de frecuencias de respuesta.

Seq	Period	Treat	0	Response				
				1	2	3	4	5
AB	1	A	39	2	25	71	203	434
AB	2	B	43	87	185	121	172	166
BA	1	B	40	76	174	127	237	138
BA	2	A	30	2	30	64	345	321

La Figura 4.5 muestra algunas cuestiones interesantes:

- El tratamiento (subtitulado) con nivel A presenta claramente mayores valores de respuesta que el B como ya habíamos visto (ver Figura 4.4).
- En general los dos grupos muestran bastante acuerdo en el subtitulado en ambos niveles: En el nivel de tratamiento A los dos grupos tienen una frecuencia relativa similar de respuestas positivas (valores 4 y 5). El grupo AB tiene un 82% de respuestas positivas frente a un 84% el grupo BA. No obstante, el grupo AB tiene más respuestas con valor 5 que el grupo BA (56% frente a 41%). La valoración es también similar entre grupos en el nivel de tratamiento B: el grupo AB tiene 44% de respuestas positivas y 47% el grupo BA. Las valoraciones negativas (1, 2), la neutra (3) y la “No sé / No contesto” (0) son también muy similares.
- Las respuestas son similares entre periodos aunque ligeramente más negativas en el segundo. Así un 65% de las respuestas son positivas en el primer periodo frente a un 64% en el segundo.

<sup>3</sup>En el Tabla 4.5 se presenta la misma información con los valores absolutos.

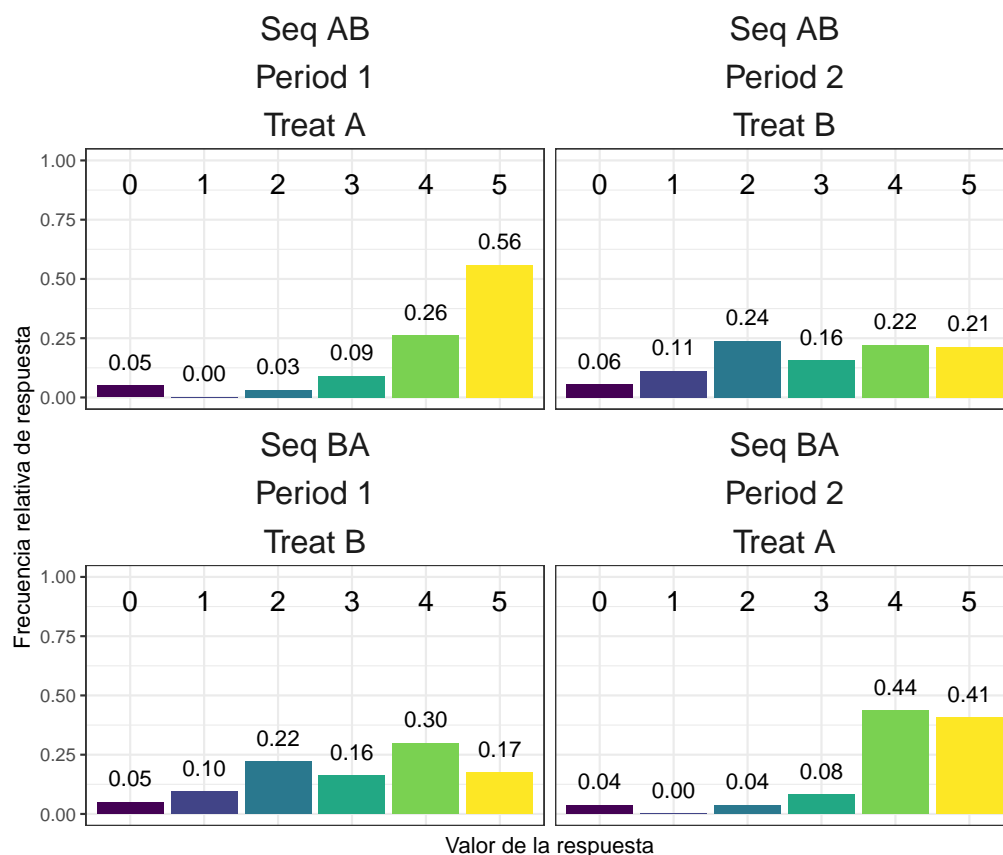


Figura 4.5: Frecuencias relativas de las respuestas al test.

El análisis marginalizado de tratamiento, secuencia y periodo tiene estos resultados referidos a las preguntas con contestación positiva (4, 5):

- El tratamiento *A* tiene un 83% marginalizado de respuestas positivas frente al 46% del tratamiento *B*.
- El periodo 1 tiene un 65% marginalizado de respuestas positivas frente al 64% del periodo 2.
- Finalmente, la secuencia *AB* tiene un 63% de respuestas positivas frente al 66% de la secuencia *BA*.

### Análisis de las preguntas.

El gráfico Figura 4.6 muestra la frecuencia relativa por grupo y por test de las preguntas clasificadas por niveles de respuesta, considerando que:

- Los niveles 1 y 2 se consideran valoraciones negativas.
- El nivel 3 se considera neutro.
- Los niveles 4 y 5 se consideran positivos.
- El nivel 0 («No sé / No contesto») se excluye en este análisis.

Se muestra en primer lugar la pregunta 18 por ser una valoración global del subtitulado y que resume la opinión que sobre el mismo tiene el estudiante. Volvemos a constatar que el subtitulado A es mejor valorado por los estudiantes, pero ahora vemos que en las 18 preguntas ambos grupos tienen más puntuaciones positivas y menos negativas en el subtitulado A que el B. También volvemos a encontrar que los dos grupos valoran de forma muy similar los dos niveles de subtitulado en todas las preguntas. En el nivel de subtitulado A las preguntas Q15, Q16 y Q17 obtienen relativamente peores valoraciones (consultar la Tabla 3.2 para ver los valores) y estas son similares en ambos subtitulados. Hay algunas preguntas que son valoradas de forma positiva incluso en el nivel de subtitulado B (por ejemplo Q04 o Q13) y que, por lo tanto, su valoración es similar en ambos subtitulados. Por último, las preguntas Q05 y Q09 (también la Q14 pero solo para el grupo BA) tienen una valoración muy negativa en el nivel de subtitulado B.



Figura 4.6: Frecuencias relativas de las respuestas por pregunta.

La figura Figura 4.7 clasifica la preguntas por valoración y permite constatar lo que ya habíamos visto en el párrafo anterior con mayor comodidad.

En la Figura 4.8 se muestra para cada pregunta la proporción de estudiantes que han valorado mejor el subtitulado A que el B <sup>4</sup>. Se comprueba que la mayoría de

<sup>4</sup>Se han eliminado las preguntas en las que una de las dos respuestas del usuario ha sido «No sé / No contesto». La comparación realizada es respuesta en  $A > B$  frente a  $A \geq B$ .

#### 4. MODELADO ESTADÍSTICO

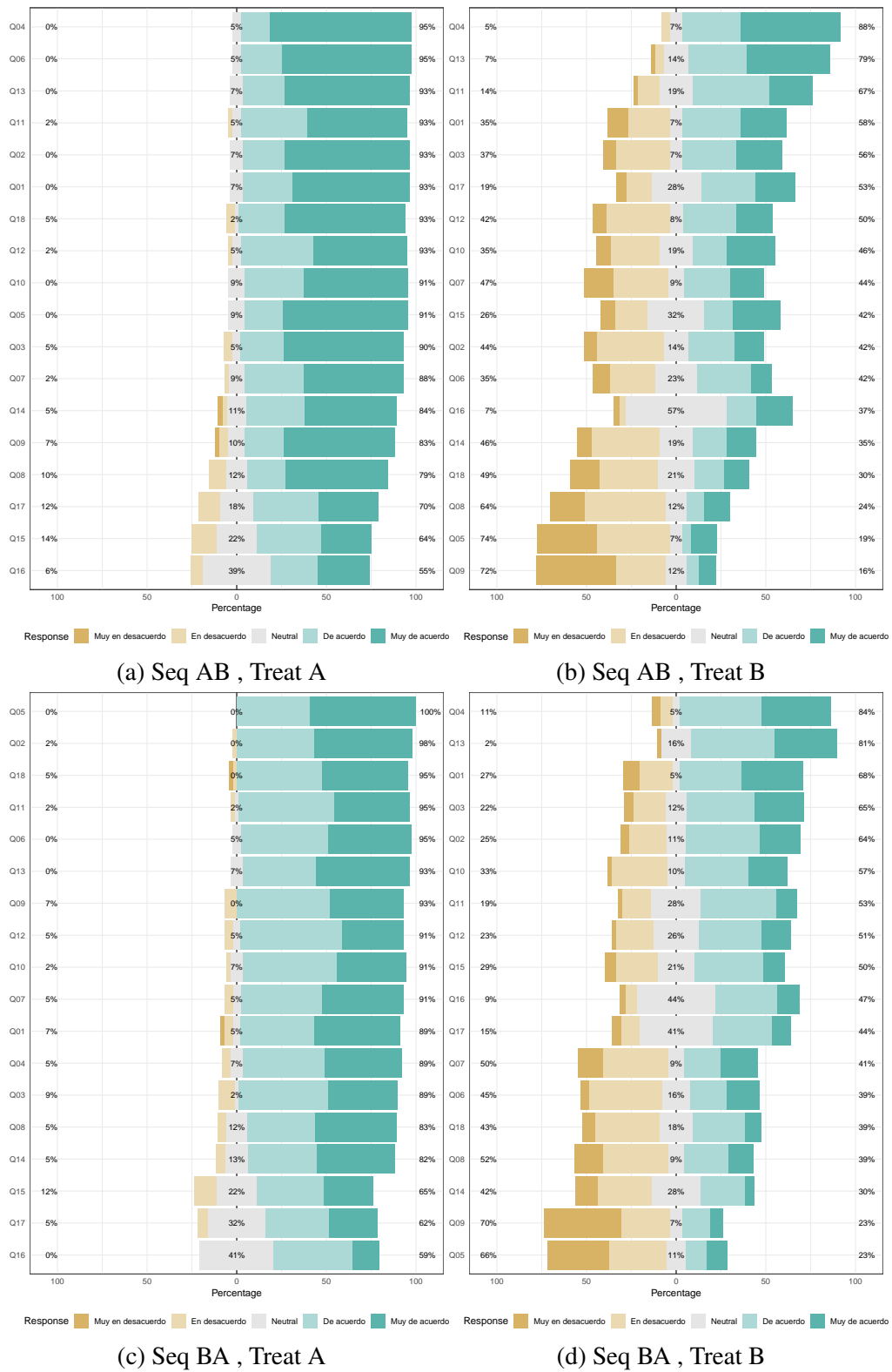


Figura 4.7: Preguntas ordenadas por valoración.

las preguntas superan el 50%, lo que indica que los estudiantes valoran mejor el subtítulo A.

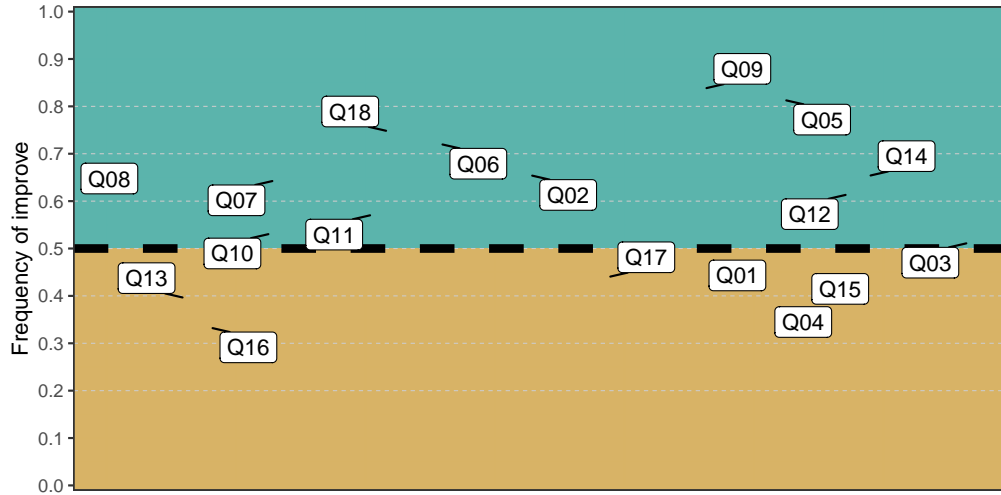


Figura 4.8: Frecuencia de preguntas que mejoran por estudiante entre subtítulos (A > B)

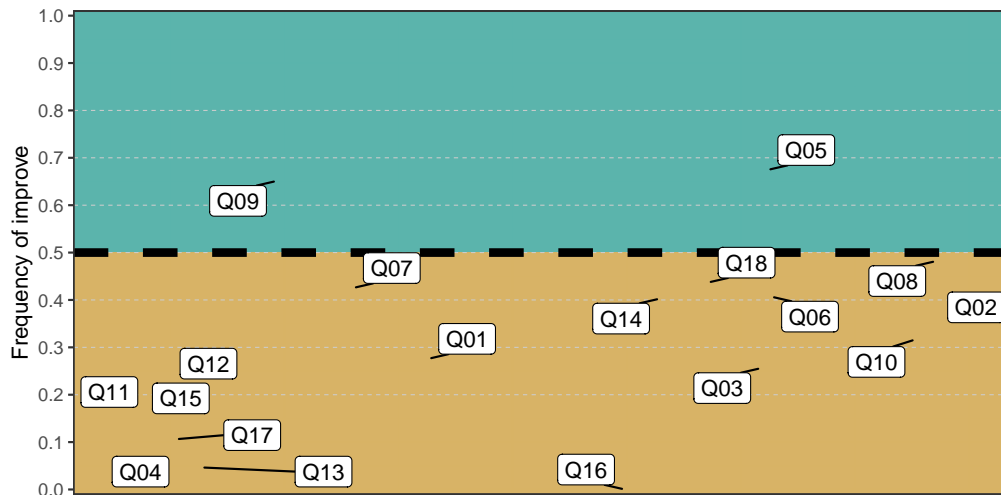


Figura 4.9: Frecuencia de preguntas que mejoran por estudiante entre subtítulos (positive A vs negative B)

En la Figura 4.9 se hace una comparación más exigente ya que ahora se muestra la frecuencia con la que los estudiantes cambian el nivel del subtítulo: de haber valorado el subtítulo positivamente (4 ó 5) en A a hacerlo de forma negativa en B (1 ó 2). En esta ocasión solo las preguntas Q05, Q09 superan el 50%.

## 4.2 Comparación con *Odds Ratio*

En la sección Sección 2.3 se explicó el fundamento teórico de esta técnica. Aquí se expone como se puede aplicar al diseño de experimento que se está analizando.

En la sección de resultados se comentarán e interpretarán los valores producidos por este test. Podemos contrastar la hipótesis de que el *OR* entre subtítulos y grupos para cada nivel de respuesta presenta diferencias significativas entre los dos grupos. Este *OR* se define:

$$OR_{(Treat,Seq|Response=r)} = \frac{\frac{P(Treat=A|Seq=AB,Response=r)}{P(Treat=B|Seq=AB,Response=r)}}{\frac{P(Treat=A|Seq=BA,Response=r)}{P(Treat=B|Seq=BA,Response=r)}} \quad (4.1)$$

Si los *odds* son similares en cada nivel de respuesta, podemos aceptar que la hipótesis nula de que los grupos responden de forma similar a cada nivel de subtítulo. Se hará un test similar pero entre subtítulo y periodos.

### 4.3 Regresión Logística

En la Sección 2.4 se presentó el fundamento teórico de la Regresión Ordinal Acumulativa. En esta sección se justifica el uso de esta técnica y se ajustan y comparan varios modelos. La variable respuesta se compone de 5 valores ordenados. Esto impide usar directamente la Regresión Logística. No obstante, podemos comparar la respuesta que cada estudiante dio a cada uno de los subtítulos y comprobar si ha mejorado o no. Esto producirá una variable de respuesta binaria que permitirá el uso de la regresión logística. Por el contrario reducirá la cantidad de datos disponibles a la mitad e impedirá analizar el efecto periodo. Se ha transformado el dataframe de tal forma que si un usuario valoró una pregunta mejor en el subtítulo A que en el B, se consigna 1 en la variable respuesta, si empeoró o puntuó igual se consigna 0. Si en un0 de los test valora una pregunta con «No sé / No contesto», se elimina esa pregunta.

Se ajusta el modelo con la secuencia como predictor. Se constata que el coeficiente del intercepto es positivo y significativo. El intercepto es el *log odds* de mejorar la valoración en A sobre B respecto a empeorar la valoración. Sin embargo, la secuencia no resulta significativa y además añadirla apenas reduce la «deviance», por lo que el modelo nulo sin predictores resulta más parsimonioso. Se podrían añadir como predictores las preguntas y los estudiantes. No se hace aquí y se pospone la discusión de esto hasta el uso de modelos mixtos.

```
Call:
glm(formula = Improve ~ 1 + Seq, data = df_improve)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5935  -0.5811   0.4065   0.4189   0.4189

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.59353    0.01848  32.125  <2e-16 ***
SeqBA       -0.01245    0.02587  -0.481    0.63
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2426952)
```



```
Null deviance: 351.72 on 1450 degrees of freedom
Residual deviance: 351.67 on 1449 degrees of freedom
AIC: 2067.2
```

```
Number of Fisher Scoring iterations: 2
```

## 4.4 Regresión Ordinal

En la Sección 2.4 se presentó el fundamento teórico de la Regresión Ordinal Acumulativa. En esta sección se ajustan algunos modelos y se interpretan los resultados.

### Ajuste del modelo ordinal `Response ~ Treat`

Existen varios paquetes en R que permiten ajustar una regresión ordinal logística. El más popular es el paquete `Ordinal` (Christensen 2022). El paquete `VGAM` (Yee 2023) es más flexible y potente. Otra posibilidad es usar la función `polr` del paquete `MASS` (Venables y Ripley 2002). Finalmente la función `orm` del paquete `rms` también permite hacerlo (ver Harrell 2015). En este trabajo usaremos el paquete `Ordinal` por permitir también incluir efectos aleatorios que utilizaremos en un apartado posterior. Comenzamos con un modelo simple que tiene como único predictor el nivel de subtitulado por ser la variable objetivo de nuestro modelo:

$$\text{logit}(P(\text{Response}_i \leq k)) = \tau_k - \beta_1 \text{Treat}_i,$$

```
clm_treat <-
  clm(
    Response ~ Treat,
    data = df_clean, link = "logit"
  )
summary(clm_treat)
```

```
formula: Response ~ Treat
data:    df_clean

link threshold nobs logLik  AIC      niter max.grad cond.H
logit flexible 2980 -3966.11 7942.21 5(0)  1.64e-10 3.1e+01

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
TreatB  -1.7206    0.0731  -23.54  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
      Estimate Std. Error z value
1|2  -3.97230    0.09678 -41.045
2|3  -2.45446    0.06812 -36.029
3|4  -1.66453    0.05936 -28.042
4|5  -0.10547    0.04946  -2.132
```

El método `summary()` muestra la información resumen. Para su interpretación vamos a seguir Christensen (2018). El número de condición Hessiano es in-

ferior a  $10^4$  lo que es indicativo de que no hay problemas de optimización <sup>5</sup>. La sección de coeficientes es la más importante. Se muestra la estimación de parámetros, el error estándar y la significación estadística de acuerdo al test de Wald <sup>6</sup>. Comprobamos que el valor es claramente significativo. Es decir, que los estudiantes han valorado de forma diferente la calidad del subtítulo en ambos vídeos. El estimador de máxima verosimilitud del coeficiente TreatB es -1.72. Siguiendo la deducción de Bruin (2011) podemos, por ejemplo, hacer la siguiente interpretación del significado de este coeficiente referido a dos niveles consecutivos de respuesta:

$$\begin{aligned}\text{logit}[P(Y \leq 1)] &= -3.97 - (-1.72x_1) \\ \text{logit}[P(Y \leq 2)] &= -2.45 - (-1.72x_1)\end{aligned}$$

Por lo tanto los *odds* serían:

$$\begin{aligned}\frac{P(Y \leq 1 | x_1 = B)}{P(Y > 1 | x_1 = B)} &= \exp(-3.97)/\exp(-1.72) \\ \frac{P(Y \leq 1 | x_1 = A)}{P(Y > 1 | x_1 = A)} &= \exp(-3.97) \\ \frac{P(Y \leq 2 | x_1 = B)}{P(Y > 2 | x_1 = B)} &= \exp(-2.45)/\exp(-1.72) \\ \frac{P(Y \leq 2 | x_1 = A)}{P(Y > 2 | x_1 = A)} &= \exp(-2.45)\end{aligned}$$

Y los *OR*:

$$\begin{aligned}\frac{P(Y \leq 1|x_1 = B)}{P(Y > 1|x_1 = B)} / \frac{P(Y \leq 1|x_1 = A)}{P(Y > 1|x_1 = A)} &= 1/\exp(-1.72) = 5.59 \\ \frac{P(Y \leq 2|x_1 = B)}{P(Y > 2|x_1 = B)} / \frac{P(Y \leq 2|x_1 = A)}{P(Y > 2|x_1 = A)} &= 1/\exp(-1.72) = 5.59\end{aligned}$$

Se comprueba que el *OR* es equivalente en todos los niveles de respuesta al cuestionario. Esta es una de las suposiciones de la regresión ordinal acumulativa. El *odds* de respuesta al cuestionario entre los niveles inferiores y superiores a uno dado,  $k$ , es 5.59 veces en el subtítulo  $B$  que en el  $A$ . Esto indica que el subtítulo  $B$  es percibido por los estudiantes como de peor calidad que el subtítulo  $A$ . Concretamente, el coeficiente  $\beta$  para Treat es el log odds de observar una mejor respuesta en una pregunta del test es 5.59 veces superior en el nivel de subtítulo  $A$  que en el  $B$ . Aunque no suele ser de interés la interpretación de los coeficientes de los umbrales (Threshold coefficients), se pueden utilizar para estimar las probabilidades de respuesta. Por ejemplo, para el nivel de subtítulo  $B$ :

---

<sup>5</sup>El número de condición de Hessiano es una medida de la curvatura de una función en un punto. Si el número de condición de Hessiano es grande, la función es muy sensible a pequeñas perturbaciones y puede ser difícil de optimizar.

<sup>6</sup>El test de Wald es un contraste de hipótesis estadístico en el que se evalúa si el valor estimado es cero suponiendo que  $W = \left( \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \right)^2 \sim \chi^2$ .

$$\begin{aligned}
\text{logit}[P(Y \leq 1)] &= -3.97 - (-1.72) = -2.25 \\
\text{odds}(P(Y \leq 1)) &= \exp(\text{logit}[P(Y \leq 1)]) = 0.11 \\
P(Y \leq 1) &= \frac{\exp(-2.25)}{1 + \exp(-2.25)} = 0.10 \\
P(Y \leq 2) &= \frac{\exp(-0.73)}{1 + \exp(-0.73)} = 0.32 \\
P(Y = 2) &= P(Y \leq 2) - P(Y \leq 1) = 0.23
\end{aligned}$$

Para el subtitulado A no se tiene en cuenta el coeficiente *TreatB* ya que el valor  $x_1$  es cero:

$$\begin{aligned}
\text{logit}[P(Y \leq 1)] &= -3.97 \\
\text{odds}(P(Y \leq 1)) &= \exp(\text{logit}[P(Y \leq 1)]) = 0.02 \\
P(Y \leq 1) &= \frac{\exp(-3.97)}{1 + \exp(-3.97)} = 0.02
\end{aligned}$$

En Tabla 4.6 se muestran las probabilidades para ambos niveles de subtitulado y todos los posibles valores de respuesta.

Cuadro 4.6: Probabilidades de respuesta para el modelo ordinal Response ~ Treat

	1	2	3	4	5
A	0.018	0.061	0.08	0.315	0.526
B	0.095	0.229	0.19	0.320	0.166

### Ajuste del modelo ordinal Response ~ Treat + Period

Para saber si existe un efecto periodo, añadimos como predictor la variable Period.

$$\text{logit}(P(\text{Response}_i \leq k)) = \tau_k - \beta_1 \text{Treat}_i - \beta_2 \text{Period}_i$$

```

clm_treat_period <-
  clm(
    Response ~ Treat + Period,
    data = df_clean, link = "logit"
  )
summary(clm_treat_period)

```

```

formula: Response ~ Treat + Period
data:    df_clean

```

```

link threshold nobs logLik  AIC      niter max.grad cond.H
logit flexible 2980 -3957.88 7927.76 5(0)  1.94e-10 4.1e+01

```

Coefficients:

```

      Estimate Std. Error z value Pr(>|z|)
TreatB -1.74090    0.07339  -23.72 < 2e-16 ***
Period2 -0.27560    0.06805   -4.05 5.12e-05 ***

```

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

Threshold coefficients:
      Estimate Std. Error z value
1|2 -4.13085    0.10507 -39.314
2|3 -2.60905    0.07872 -33.143
3|4 -1.81652    0.07073 -25.681
4|5 -0.25187    0.06153  -4.093
```

Vemos que ambos coeficientes son significativos y con signo negativo. Un signo negativo en el efecto periodo está asociado con que la valoración del subtítulo empeora en el segundo periodo independientemente de si se trata del subtítulo correcto o incorrecto. Aplicando el mismo razonamiento del apartado anterior, el *OR* del efecto periodo es  $1/\exp(-0.28) = 1.32$ . Lo que quiere decir que una vez controlado el efecto principal del tratamiento, el subtítulo en el segundo periodo es valorado como de inferior calidad que en el primero. Esto estaría indicando que los estudiantes son más exigentes con el subtítulo en la segunda actividad independientemente de su calidad real.

### Ajuste del modelo ordinal `Response ~ Treat * Period`

Añadimos al modelo la interacción entre subtítulo y periodo. Esta interacción corresponde al efecto secuencia. Se puede demostrar que los modelos `Response ~ Treat*Period` y `Response ~ Treat + Period + Seq` ver [A](#).

```
options(contrasts = rep("contr.treatment", 2))
clm_treat.period <-
  clm(
    Response ~ Treat * Period,
    data = df_clean, link = "logit"
  )
summary(clm_treat.period)
```

formula: `Response ~ Treat * Period`  
data: `df_clean`

```
link threshold nobis logLik AIC niter max.grad cond.H
logit flexible 2980 -3953.01 7920.03 5(0) 2.14e-10 8.1e+01
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
TreatB	-1.96046	0.10229	-19.166	< 2e-16 ***
Period2	-0.49179	0.09744	-5.047	4.49e-07 ***
TreatB:Period2	0.42510	0.13638	3.117	0.00183 **

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

Threshold coefficients:
      Estimate Std. Error z value
1|2 -4.24649    0.11182 -37.977
2|3 -2.72823    0.08821 -30.928
3|4 -1.93833    0.08167 -23.732
4|5 -0.36991    0.07308  -5.062
```

Vemos que los tres coeficientes son significativos. El principal efecto es el nivel de subtítulo obteniendo mejores puntuaciones el nivel A; el efecto periodo

es negativo por lo que el primer periodo obtiene mejores puntuaciones; por último, el efecto secuencia es positivo pero de menor valor absoluto que el efecto periodo. Esto quiere decir que el subtitulado de nivel *B* en el periodo 2 (secuencia *AB*), tiene un efecto periodo inferior que el subtitulado *A* en el mismo periodo. Matemáticamente:

$$\begin{aligned} \text{logit}[P(Y \leq 1 \mid \text{Treat} = A, \text{Period} = 1)] &= -4.25 \\ \text{logit}[P(Y \leq 1 \mid \text{Treat} = B, \text{Period} = 1)] &= -4.25 - (-1.96) \\ \text{logit}[P(Y \leq 1 \mid \text{Treat} = A, \text{Period} = 2)] &= -4.25 - (-0.49) \\ \text{logit}[P(Y \leq 1 \mid \text{Treat} = B, \text{Period} = 2)] &= -4.25 - (-1.96 - 0.49 + 0.43) \end{aligned}$$

En definitiva, en el nivel de subtitulado *B* apenas encontramos diferencias entre periodos, sin embargo, en el nivel de subtitulado *A* existe un efecto periodo cuyo valor en logits es -0.49. Es decir, que la valoración del subtitulado de nivel *A* es inferior en el segundo periodo que en el primero.

## Elección del modelo ordinal mediante el test de razón de verosimilitud

Al ser los tres modelos anidados, podemos compararlos con la prueba de razón de verosimilitud. Comprobamos que el tercer modelo (el que incorpora la interacción entre los subtítulos y el periodo) reduce significativamente el logaritmo de la función de verosimilitud y, por lo tanto, debe ser aceptado:

```
anova(clm_treat, clm_treat_period, clm_treat.period)
```

Likelihood ratio tests of cumulative link models:

	formula:	link:	threshold:
clm_treat	Response ~ Treat	logit	flexible
clm_treat_period	Response ~ Treat + Period	logit	flexible
clm_treat.period	Response ~ Treat * Period	logit	flexible

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
clm_treat	5	7942.2	-3966.1			
clm_treat_period	6	7927.8	-3957.9	16.448	1	5e-05 ***
clm_treat.period	7	7920.0	-3953.0	9.738	1	0.001805 **

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

## Comprobación de las hipótesis del modelo

La principal hipótesis de un modelo de regresión logística ordinal proporcional acumulativa es que los coeficientes son iguales entre cualesquiera dos niveles de respuestas correlativos. Se han propuesto diversas fórmulas para comprobar esta hipótesis. El paquete *Ordinal* dispone de la función `nominal_test()` que lo que hace es realizar un test de razón de verosimilitud para cada predictor ajustando un modelo en el que se ha relajado la condición de proporcionalidad.

Se constata que el test resulta significativo para `Treat` y para `Treat:Period`, por lo que para estas dos variables no se puede asumir que los coeficientes estimados se mantengan constantes en todos los niveles de respuesta.

```
nominal_test(clm_treat.period)

Tests of nominal effects

formula: Response ~ Treat * Period
      Df logLik   AIC    LRT Pr(>Chi)
<none>      -3953.0 7920.0
Treat      3 -3904.4 7828.9  97.172  <2e-16 ***
Period     3 -3951.4 7922.7   3.307   0.3467
Treat:Period 9 -3884.8 7801.6 136.408  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Lo que procede es ajustar el modelo relajando la constante de proporcionalidad de esas variables. Se ha realizado esto utilizando la función `vglm` del paquete `VGAM`. Vemos que ahora hay cuatro coeficientes para cada una de las variables `Treat` y `Treat:Period` <sup>7</sup>.

```
vglm_treat.period <- vglm(
  Response ~ Treat * Period,
  VGAM::cumulative(link = "logit", parallel = F ~ Treat + Treat:Period, reverse = T),
  data = df_clean
)
coef(vglm_treat.period) %>% data.frame()

      .
(Intercept):1      6.2295614
(Intercept):2      3.5001281
(Intercept):3      2.2121438
(Intercept):4      0.2998788
TreatB:1           -4.0441016
TreatB:2           -2.8029889
TreatB:3           -2.2174630
TreatB:4           -1.7926200
Period2            -0.5345919
TreatB:Period2:1    0.3509228
TreatB:Period2:2    0.3607009
TreatB:Period2:3    0.3891474
TreatB:Period2:4    0.8024952
```

---

<sup>7</sup>Los umbrales tienen los mismos valores pero de signo contrario debido a diferencias en la parametrización del modelo en cada función utilizada.

## RESULTADOS

En este capítulo se comentarán los resultados de las técnicas estadísticas (ver Sección 2.2) y de los modelos propuestos (ver ?@sec-modelos).

### Correlación entre preguntas con el alfa de Cronbach.

El coeficiente alfa de Cronbach (ver Sección 2.2) de la escala de Likert es 0.921 que indica una muy buena correlación entre las respuestas a todas las preguntas. Este valor apenas se ve alterado si se elimina una de las preguntas. Esto nos permite concluir que todos los ítems pertenecen a la misma escala de Likert.

En la Tabla 5.1 mostramos las preguntas que más contribuyen al índice alfa de Cronbach. Es interesante que la pregunta Q18, que es la valoración general del cuestionario, sea la que mejor contribución tiene al índice.

Cuadro 5.1: Relación de cada pregunta con el índice alfa de Cronbach.

(a) Variables con mayor asociación.

Q18	Q05	Q06	Q09	Q08	Q07	Q10	Q12	Q02
0.86	0.81	0.79	0.79	0.77	0.75	0.73	0.72	0.71

(a) Variables con menor asociación.

Q03	Q14	Q01	Q11	Q15	Q13	Q04	Q16	Q17
0.68	0.66	0.65	0.64	0.56	0.51	0.46	0.44	0.42

Cuadro 5.4: Tablas de contingencia.

(a) ~ Response + Treat			(b) ~ Response + Period		(c) ~ Response + Seq	
Response	A	B	1	2	AB	BA
1	4	163	78	89	89	78
2	55	359	199	215	210	204
3	135	248	198	185	192	191
4	548	409	440	517	375	582
5	755	304	572	487	600	459

### Asociación de variables con la prueba de homogeneidad $\chi^2$

En la Sección 2.3 se describe el marco teórico de esta prueba no paramétrica. Se ha contrastado la existencia de asociación entre la variable respuesta *Response* y cada una de las tres variables más importantes de nuestro modelo (*Treat*, *Period* y *Seq*). En la Tabla 5.4 se muestran las tablas de contingencia. En la Tabla 5.5 se muestra que todos los contrastes son significativos, con lo que se rechaza la hipótesis nula de homogeneidad. Todas las variables explicativas tienen influencia en la variable respuesta.

Cuadro 5.5: Valores del contraste de hipótesis  $\chi^2$ 

	~ Response + Treat	~ Response + Period	~ Response + Seq
$\chi^2$	621.497	15.039	64.904
<i>p</i> -value	$4.5778215 \times 10^{-132}$	0.0101975	$1.1733569 \times 10^{-12}$
<i>df</i>	5	5	5

### Comparación con *Odds Ratio*

En Sección 2.3 y en Sección 4.2 se exponen el marco teórico y la fundamentación de los contrastes de hipótesis realizados respectivamente. El contraste de hipótesis del *log OR* del subtítulo para cada grupo no produce significación estadística en ningún nivel de respuesta, por lo que según esta prueba estadística la secuencia de subtítulo no influiría en la respuesta de los estudiantes (ver Tabla 5.6). Es decir, de acuerdo a este la secuencia no influye en la valoración de los subtítulos.

Cuadro 5.6: *LogOR* ~ Treat + Seq + Response\_1

Response	Estimate	Std. Error	z value	Pr(> z )
No sé / No contesto	0.190	0.327	0.580	0.562
Muy en desacuerdo	-0.135	1.012	-0.134	0.894
En desacuerdo	-0.244	0.291	-0.838	0.402
Neutral	0.152	0.214	0.711	0.477
De acuerdo	-0.210	0.134	-1.570	0.116



Muy de acuerdo	0.117	0.137	0.855	0.393
----------------	-------	-------	-------	-------

Sin embargo, si realizamos este contraste entre subtítulos y periodos podemos constatar la existencia de un efecto periodo de signo contrario para las preguntas 4 y 5 (ver Tabla 5.7). La razón de que se produzca este efecto periodo es que algunas de las respuestas de valoración 5 en ambos niveles de subtítulo y grupos en el primer periodo se convierten en valoración 4 en el segundo periodo. Esto se puede ver más claramente en la Figura 5.1.

Cuadro 5.7:  $\text{LogOR} \sim \text{Treat} + \text{Period} + \text{Response\_1}$

Response	Estimate	Std. Error	z value	Pr(> z )
No sé / No contesto	0.335	0.327	1.022	0.307
Muy en desacuerdo	0.135	1.012	0.134	0.894
En desacuerdo	-0.121	0.291	-0.416	0.677
Neutral	0.055	0.214	0.259	0.796
De acuerdo	-0.851	0.134	-6.367	0.000
Muy de acuerdo	0.486	0.137	3.557	0.000

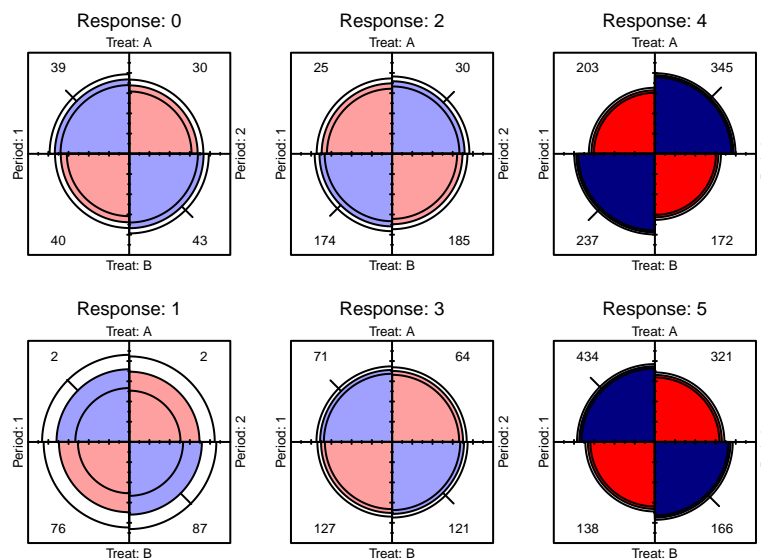


Figura 5.1:  $\text{OR} \sim \text{Treat} + \text{Period} + \text{Response}$



## DISCUSIÓN Y CONCLUSIONES

### 6.1 Comparación con *Odds Ratio*

En la Sección 5 se constató que hay más proporción de respuestas 5 en el primer periodo que en el segundo en ambos niveles de subtitulado. Esto estaría indicando que los estudiantes de ambos grupos prestaron más atención o fueron más exigentes en el segundo visionado y decidieron no otorgar la puntuación máxima incluso en algunos ítems al subtitulado correcto. Que el efecto periodo sea contrario en dos preguntas (ver @Figura 5.1) no debe sorprendernos en este diseño de experimento, ya que un test es un juego de suma cero: la valoraciones que se ganan o se pierden en un nivel de respuesta necesariamente provocan que el resto de niveles pierdan o ganen respectivamente la misma cantidad. En cualquier caso, vemos que el efecto periodo es cuantitativa y cualitativamente pequeño. Al afectar solo al intercambio de valoraciones entre los niveles 4 y 5, y ser las dos positivas, es simplemente una pequeña corrección en la valoración del subtitulado.



## REFERENCIAS

- AENOR (2012). *UNE 153010 Subtitulado para personas sordas y personas con discapacidad auditiva*. Asociación Española de Normalización y Certificación (vid. pág. 12).
- Agresti, A. (oct. de 2018). *An introduction to categorical data analysis, 3rd Edition*. URL: <https://www.wiley.com/en-us/An+Introduction+to+Categorical+Data+Analysis%2C+3rd+Edition-p-9781119405283> (vid. págs. 5-7).
- Bruin, J. (2011). *How do I interpret the coefficients in an ordinal logistic regression in R*. URL: <https://stats.oarc.ucla.edu/r/faq/ologit-coefficients> (vid. pág. 30).
- Bürkner, P.-C. y M. Vuorre (feb. de 2019). «Ordinal Regression Models in Psychology: A Tutorial». En: *Advances in Methods and Practices in Psychological Science* 2, pág. 251524591882319. DOI: [10.1177/2515245918823199](https://doi.org/10.1177/2515245918823199) (vid. pág. 8).
- Christensen, R. H. B. (2022). *ordinal—Regression Models for Ordinal Data*. R package version 2022.11-16. <https://CRAN.R-project.org/package=ordinal> (vid. pág. 29).
- Christensen, R. H. B. (2018). «Cumulative Link Models for Ordinal Regression with the R Package ordinal». En: (vid. pág. 29).
- Emilio, L. y E. M. Molanes-López (s.f.). *Principales técnicas estadísticas*. ver 3 2021-2022. 2021 (vid. pág. 6).
- Friendly, M., D. Meyer y A. Zeileis (dic. de 2015). *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*, págs. 1-525. DOI: [10.1201/b19022](https://doi.org/10.1201/b19022) (vid. págs. 5, 7).
- Harrell, F. (ene. de 2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. DOI: [10.1007/978-3-319-19425-7](https://doi.org/10.1007/978-3-319-19425-7) (vid. pág. 29).
- Lawson, J. (2015). Ed. por Chapman y Hall/CRC. DOI: [10.1201/b17883](https://doi.org/10.1201/b17883). URL: <https://www.taylorfrancis.com/books/mono/10.1201/b17883/design-analysis-experiments-john-lawson> (vid. pág. 3).
- Liddell, T. M. y J. K. Kruschke (2018). «Analyzing ordinal data with metric models: What could possibly go wrong?». En: *Journal of Experimental Social Psychology* 79, págs. 328-348. DOI: [10.1016/j.jesp.2018.08.009](https://doi.org/10.1016/j.jesp.2018.08.009). URL: <https://www.sciencedirect.com/science/article/pii/S0022103117307746> (vid. pág. 4).
- Liu, X. (abr. de 2022). *Categorical Data Analysis and Multilevel Modeling Using R*. Ed. por S. P. Ltd. (vid. pág. 10).

- Lui, K.-J. (ago. de 2016). *Crossover Designs: Testing, Estimation, and Sample Size*. DOI: [10.1002/9781119114710](https://doi.org/10.1002/9781119114710) (vid. pág. 4).
- Molanes-López, E. M., A. Rodríguez-Ascaso, E. Letón y J. Pérez-Martín (2021). «Assessment of Video Accessibility by Students of a MOOC on Digital Materials for All». En: *IEEE Access* 9, págs. 72357-72367. DOI: [10.1109/ACCESS.2021.3079199](https://doi.org/10.1109/ACCESS.2021.3079199) (vid. pág. 12).
- Pérez Martín, J., A. Rodríguez-Ascaso y E. Molanes-López (nov. de 2021). «Quality of the captions produced by students of an accessibility MOOC using a semi-automatic tool». En: *Universal Access in the Information Society* 20. DOI: [10.1007/s10209-020-00740-9](https://doi.org/10.1007/s10209-020-00740-9) (vid. pág. 12).
- Schweinberger, M. (2020). *Questionnaires and Surveys: Analyses with R*. 2020/12/11. <https://slcladal.github.io/survey.html>. The University of Queensland, Australia. School of Languages y Cultures. Brisbane (vid. pág. 5).
- Senn, S. (2022). Ed. por L. John Wiley. DOI: [10.1002/0470854596](https://doi.org/10.1002/0470854596) (vid. pág. 3).
- Venables, W. N. y B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer. URL: <https://www.stats.ox.ac.uk/pub/MASS4/> (vid. pág. 29).
- Yee, T. W. (2023). *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.1-8. URL: <https://CRAN.R-project.org/package=VGAM> (vid. pág. 29).



## EFECTO SECUENCIA E INTERACCIÓN TRATAMIENTO VS. PERIODO.

Vamos a demostrar que el efecto secuencia es equivalente a la interacción de los factores tratamiento y periodo.

### A.1 Preparación.

Partimos del siguiente conjunto de datos generado aleatoriamente <sup>1</sup>:

```
set.seed(100)
n <- 1000
df <- data.frame(
  Response = rnorm(n),
  Treat = as.factor(sample(c("A", "B"), n, replace = TRUE)),
  Period = as.factor(sample(c(1, 2), n, replace = TRUE))
)

df$Seq <- as.factor(
  ifelse(
    df$Period == 1 & df$Treat == "A" | df$Period == 2 & df$Treat == "B",
    "AB",
    "BA"
  )
)

head(df, 10)
```

	Response	Treat	Period	Seq
1	-0.50219235	B	2	AB
2	0.13153117	A	1	AB

<sup>1</sup>Obsérvese que se la variable Response en esta simulación es cuantitativa y no ordinal. Se ha realizado de esta forma para poder usar un ajuste de mínimos cuadrados en lugar de una regresión ordinal para facilitar el cálculo y su interpretación.

## A. EFECTO SECUENCIA E INTERACCIÓN TRATAMIENTO VS. PERIODO.

3	-0.07891709	A	2	BA
4	0.88678481	A	2	BA
5	0.11697127	A	1	AB
6	0.31863009	A	2	BA
7	-0.58179068	A	2	BA
8	0.71453271	A	1	AB
9	-0.82525943	B	2	AB
10	-0.35986213	B	1	BA

Calculamos las medias por cada nivel de factor y combinaciones de niveles que utilizaremos luego en la interpretación de los coeficientes de los modelos

```
M <- mean(df$Response) # 1 media de respuesta global

# 2 medias de respuesta para tratamientos A y B
mTreat <- with(df, tapply(Response, Treat, mean))

# 2 medias de respuesta para periodos 1 y 2
mPeriod <- with(df, tapply(Response, Period, mean))

# 2 medias de respuesta para secuencias AB y BA
mSeq <- with(df, tapply(Response, Seq, mean))

# 4 medias de respuesta para las cuatro combinaciones de tratamiento y periodo
m2 <- with(df, tapply(Response, list(Treat, Period), mean))

dTreat <- diff(mTreat) # diferencia de medias entre tratamientos A y B

dPeriod <- diff(mPeriod) # diferencia de medias entre periodos 1 y 2

d2 <- diff(m2) # diferencias entre niveles de tratamiento en cada nivel de periodo
```

## A.2 Análisis con un solo factor (tratamiento).

```
l1 <- lm(Response ~ Treat, df)
data.frame(t(coef(l1))) %>% gt()
```

Cuadro A.1: Ajuste del modelo  $\text{Response} \sim \text{Treat}$  con contrasts treatment.

X.Intercept.	TreatB
0.03624217	-0.03966751

Vemos que el intercepto es la media de la respuesta en el nivel de tratamiento A:

```
mTreat[1]
```

```
      A
0.03624217
```

Que la pendiente (parámetro  $\text{TreatB}$ ) es la diferencia entre las medias tratamientos:

```
dTreat
```



```

      B
-0.03966751

```

Por ello, para conocer el efecto del tratamiento en el nivel  $B$  hay que sumar intercepto y pendiente:

```

coef(l1)[[1]] + coef(l1)[[2]] - mTreat[[2]]

[1] 1.214306e-16

```

Esto es así ya que por defecto R utiliza el contraste conocido como codificación de tratamiento:

```

contr.treatment(2)

      2
1 0
2 1

```

Podemos ver la matriz ampliada añadiendo el intercepto, que siempre será una columna de 1's:

```

model.matrix(~Treat, expand.grid(Treat = c("A", "B")))

      (Intercept) TreatB
1             1      0
2             1      1
attr(,"assign")
[1] 0 1
attr(,"contrasts")
attr(,"contrasts")$Treat
[1] "contr.treatment"

```

Cada fila representa el nivel del tratamiento (fila 1 nivel  $A$  y fila 2 nivel  $B$ ) y las columnas representan los parámetros del modelo. Los valores son los niveles de tratamiento (0 ó 1). Para obtener el significado de cada parámetro, multiplicamos el valor del contraste por el parámetro. Así:

- De la primera fila obtenemos que el efecto del tratamiento  $A$  es el intercepto:  $A = 1 \cdot \text{Intercept} + 0 \cdot \text{TreatB}$ .
- De la segunda fila obtenemos que el valor del parámetro  $\text{TreatB}$  es la diferencia de los niveles de tratamiento.  $B = 1 \cdot \text{Intercept} + 1 \cdot \text{TreatB} \Rightarrow \text{TreatB} = B - \text{Intercept}$ .

Esto quiere decir que existe una variable para codificar el efecto tratamiento, y esta variable tiene el valor 0 para el nivel  $A$  por ser el de referencia y 1 para el nivel  $B$ . La pendiente se codifica como la diferencia del efecto de los dos niveles ( $B - A$ ).

### A.3 Análisis con un dos factores (tratamiento y periodo).

```
l2 <- lm(Response ~ Treat * Period, df)
data.frame(t(coef(l2))) %>% gt()
```

Cuadro A.2: Ajuste del modelo  $\text{Response} \sim \text{Treat} * \text{Period}$  con contrasts treatment.

X.Intercept.	TreatB	Period2	TreatB.Period2
0.04138614	-0.1076137	-0.01125933	0.1343517

Vemos que el intercepto es la media del tratamiento A en el periodo 1 por ser estos los valores que R usa como referencia <sup>2</sup>:

```
m2["A", "1"]
```

```
[1] 0.04138614
```

El parámetro *TreatB* es la diferencia de medias entre los tratamientos en el periodo 1:

```
m2["B", "1"] - m2["A", "1"]
```

```
[1] -0.1076137
```

El parámetro *Period2* es la diferencia de medias entre los periodos en el nivel de tratamiento A:

```
m2["A", "2"] - m2["A", "1"]
```

```
[1] -0.01125933
```

Finalmente, *TreatB : Period2* es la diferencia entre el segundo periodo y el primero del nivel de tratamiento B menos la diferencia entre periodos del nivel de tratamiento A:

```
m2["B", "2"] - m2["B", "1"] - (m2["A", "2"] - m2["A", "1"])
```

```
[1] 0.1343517
```

La matriz de contraste nos permite razonar por qué esto es así:

```
model.matrix(~ Treat * Period, expand.grid(Treat = c("A", "B"), Period = c("1", "2")))
```

<sup>2</sup>R utiliza como valor de referencia el nivel más bajo de factor.

```

      (Intercept) TreatB Period2 TreatB:Period2
1             1      0      0             0
2             1      1      0             0
3             1      0      1             0
4             1      1      1             1
attr(,"assign")
[1] 0 1 2 3
attr(,"contrasts")
attr(,"contrasts")$Treat
[1] "contr.treatment"

attr(,"contrasts")$Period
[1] "contr.treatment"

```

- La primera fila es el intercepto y corresponde con el tratamiento *A* y el periodo 1.
- La segunda fila es el efecto del tratamiento *B* en el periodo 1 y se calcula con la suma del intercepto y el parámetro *TreatB*. Luego *TreatB* es la diferencia del efecto de los tratamientos en el periodo 1.
- Análogamente con la tercera fila concluimos que *Period2* es la deferencia entre periodos para el tratamiento *A*.
- Finalmente, la cuarta fila, es el tratamiento *B* en el periodo 2 y, por lo tanto, *Treat2 : Period2* es la diferencia el nivel *B* de tratamiento y el periodo 2 y el nivel de tratamiento *A* en el periodo 1, menos la diferencia de niveles de tratamiento para el periodo 1 y menos la diferencia de periodos para el tratamiento *A*.

Obsérvese que antes hemos calculado de forma diferente *TreatB : Period2*. Podemos aplicar la fórmula anterior y comprobar que produce el mismo resultado:

```

m2["B", "2"] - m2["A", "1"] - (m2["B", "1"] - m2["A", "1"]) - (m2["A", "2"] - m2["A", "1"])

[1] 0.1343517

```

## A.4 Factor secuencia.

Vamos a incorporar la secuencia como factor para ver si es equivalente a la interacción entre periodo y tratamiento. En caso de serlo los coeficientes del modelo ajustado deberían coincidir. Sin embargo vemos que los modelos l2 (Tabla A.2) y l3 (Tabla A.3) tienen distintos coeficientes.

```

l3 <- lm(Response ~ Treat + Period + Seq, df)
data.frame(t(coef(l3))) %>% gt()

```

Cuadro A.3: Ajuste del modelo  $\text{Response} \sim \text{Treat} + \text{Period} + \text{Seq}$  con contrasts treatment.

X.Intercept.	TreatB	Period2	SeqBA
0.04138614	-0.04043786	0.05591654	-0.06717587

## A. EFECTO SECUENCIA E INTERACCIÓN TRATAMIENTO VS. PERIODO.

Los coeficientes no coinciden debido a que estamos usando el contraste con codificación de tratamientos. Pero si cambiamos a codificación de sumas:

```
options(contrasts = rep("contr.sum", 2))
```

Y volvemos a ajustar los modelos que ya usarán el contraste `suma`, podemos comprobar que ahora tienen los mismos coeficientes y el coeficiente `Seq1` del modelo que incorpora el efecto secuencia (Tabla A.4) es igual que el coeficiente `Treat1 : Period1` del modelo que incorpora la interacción entre tratamiento y periodo (Tabla A.5). Obsérvese que los nombres de los coeficientes han cambiado respecto al contraste de tratamiento. Esto sucede porque la interpretación de los coeficientes varía como se explica a continuación.

```
l4 <- lm(Response ~ Treat + Period + Seq, df)
data.frame(t(coef(l4))) %>% gt()
```

Cuadro A.4: Ajuste del modelo  $\text{Response} \sim \text{Treat} + \text{Period} + \text{Seq}$  con contrasts sum.

X.Intercept.	Treat1	Period1	Seq1
0.01553755	0.02021893	-0.02795827	0.03358794

```
l5 <- lm(Response ~ Treat * Period, df)
data.frame(t(coef(l5))) %>% gt()
```

Cuadro A.5: Ajuste del modelo  $\text{Response} \sim \text{Treat} * \text{Period}$  con contrasts sum.

X.Intercept.	Treat1	Period1	Treat1.Period1
0.01553755	0.02021893	-0.02795827	0.03358794

La interpretación de los contrastes es diferente. Para explicarlo, mostramos la matriz de contraste:

```
model.matrix(~ Treat * Period, expand.grid(Treat = c("A", "B"), Period = c("1", "2")))
```

```
(Intercept) Treat1 Period1 Treat1:Period1
1          1      1      1              1
2          1     -1      1             -1
3          1      1     -1             -1
4          1     -1     -1              1
attr(,"assign")
[1] 0 1 2 3
attr(,"contrasts")
attr(,"contrasts")$Treat
[1] "contr.sum"

attr(,"contrasts")$Period
[1] "contr.sum"
```

Vemos que ahora los niveles son 1 y -1 <sup>3</sup> en vez de 0 y 1 que se utilizan en el contraste de tratamiento. La interpretación es la siguiente:

- El interceptor es la media de la media de cada uno de los niveles de factor. ¿Por qué?. El interceptor es el valor de la variable de respuesta cuando todas las variables explicativas valen 0. Esto sucede en la media de la variable de respuesta ya que cero es el valor que está en la mitad de +1 y -1. Podemos comprobar que la media global coincide con el interceptor del modelo l4 (Tabla A.4):

```
mean(m2)
```

```
[1] 0.01553755
```

- El coeficiente *Treat1* es la mitad la diferencia de la media entre niveles de tratamiento ( $TreatA - TreatB$ ). La media de cada tratamiento se calcula como la media del tratamiento en cada periodo.

```
-diff(apply(m2, 1, mean)) / 2
```

```

      B
0.02021893

```

Otra forma de entender el coeficiente *Treat1* es como la cuarta parte de la diferencia de los efectos de los tratamientos en cada periodo.

```
(m2["A", "1"] + m2["A", "2"] - (m2["B", "1"] + m2["B", "2"])) / 4
```

```
[1] 0.02021893
```

- El coeficiente *Period1* es la mitad la diferencia de la media entre periodos ( $Period1 - Period2$ ). La media entre periodos se calcula como la media del periodo para cada tratamiento.

```
-diff(apply(m2, 2, mean)) / 2
```

```

      2
-0.02795827

```

Otra forma de entender el coeficiente *Period1* es como la cuarta parte de la diferencia de los efectos del periodo en cada tratamiento.

---

<sup>3</sup>El nivel de referencia del factor tendrá valor 1 y el otro -1. Por ejemplo, en la variable *Treat*, A tendrá +1 y B tendrá valor -1.

## A. EFECTO SECUENCIA E INTERACCIÓN TRATAMIENTO VS. PERIODO.

```
(m2["A", "1"] + m2["B", "1"] - (m2["A", "2"] + m2["B", "2"])) / 4
```

```
[1] -0.02795827
```

- El coeficiente  $Treat1 : Period1$  es el coeficiente  $Treat1$  menos la mitad de la diferencia de la media entre tratamientos para el periodo 2 ( $TreatA - TreatB$ ):

```
-diff(apply(m2, 1, mean)) / 2 + diff(m2[, "2"]) / 2
```

```
B
```

```
0.03358794
```

```
coef(l5)[2] + diff(m2[, "2"]) / 2
```

```
Treat1
```

```
0.03358794
```

El coeficiente  $Treat1 : Period1$  también se puede calcular como  $Period1$  menos la mitad de la diferencia de la media entre periodos para el para el tratamiento  $B$  ( $Period1 - Period2$ ):

```
-diff(apply(m2, 2, mean)) / 2 + diff(m2["B", ]) / 2
```

```
2
```

```
0.03358794
```

```
coef(l5)[3] + diff(m2["B", ]) / 2
```

```
Period1
```

```
0.03358794
```

Un tercera forma de interpretar el coeficiente  $Treat1 : Period1$  es como la cuarta parte de la suma de la diferencia cruzada del efecto de cada tratamiento en cada periodo:

```
(m2["A", "1"] - m2["A", "2"] + m2["B", "2"] - m2["B", "1"]) / 4
```

```
[1] 0.03358794
```

O reorganizando los términos de otra forma, sería la cuarta parte de la suma de la diferencia cruzada del efecto de cada periodo en cada tratamiento:

```
(m2["B", "2"] - m2["A", "2"] + m2["A", "1"] - m2["B", "1"]) / 4
```

```
[1] 0.03358794
```

- Podemos obtener el coeficiente  $TreatB$  del modelo  $l2$  (Tabla A.2) como  $-2 \cdot (Treat1 + Treat1 : Period1)$ :

```
-2 * (coef(l5)["Treat1"] + coef(l5)["Treat1:Period1"])
```

```
Treat1  
-0.1076137
```

- Análogamente el coeficiente *Period2* del modelo *l2* (Tabla A.2) se obtiene  $-2 \cdot (\textit{Period1} + \textit{Treat1} : \textit{Period1})$ :

```
-2 * (coef(l5)["Period1"] + coef(l5)["Treat1:Period1"])
```

```
Period1  
-0.01125933
```

- El coeficiente *TreatB : Period2* se obtiene como  $4 \cdot \textit{Treat1} : \textit{Period1}$ :

```
4 * (coef(l5)["Treat1:Period1"])
```

```
Treat1:Period1  
0.1343517
```

