



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^AT_EX 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

Resumen	iii
Agradecimientos	v
Índice	vi
Índice de cuadros	vii
Índice de figuras	ix
1 Introducción	1
2 Motivación	3
2.1 Propuesta y objetivos	3
2.2 Estructura del documento	3
3 Estado del arte	5
4 Materiales y métodos	7
5 Métodos.	9
5.1 Fuente de datos.	9
5.2 Características del diseño del experimento.	11
5.3 Objetivo.	12
5.4 Preprocesamiento.	12
5.5 Variables relevantes.	14
6 Variables del modelo	15
7 Resultados	17
8 Conclusiones y trabajo futuro	19
Referencias	21

Apéndices	23
A Preprocesado de los ficheros suministrados.	23

ÍNDICE DE CUADROS

5.1	Niveles de los items de la escala de Likert.	10
5.2	Items de la escala de Likert.	10
5.3	Descripción de las variables más importantes	14

ÍNDICE DE FIGURAS

CAPÍTULO



INTRODUCCIÓN

MOTIVACIÓN

2.1 Propuesta y objetivos

2.2 Estructura del documento

CAPÍTULO



ESTADO DEL ARTE

MATERIALES Y MÉTODOS

MÉTODOS.

5.1 Fuente de datos.

Los datos proceden de la edición de 2022 del curso MOOC Materiales digitales accesibles de la UNED. Concretamente a los estudiantes matriculados se les propuso que realizaran una actividad voluntaria consistente en evaluar la calidad del subtitulado de dos vídeos. Los vídeos eran idénticos y se diferenciaban únicamente en la calidad del subtitulado. 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 inapreciables para alguien que carezca de conocimientos sobre accesibilidad. Los estudiantes fueron clasificados en dos grupos. Al primer grupo se le presentó primero el vídeo correctamente subtitulado y luego el otro. El segundo grupo realizó la actividad cruzada: primero evaluó el vídeo mal subtitulado y luego el bien subtitulado. Tras ver cada uno de los vídeos, los estudiantes tuvieron la oportunidad de valorar la calidad del subtitulado realizando un test en escala de Likert de 18 ítems y 5 niveles cada ítem ¹. Los 18 ítems de Likert pretenden asegurar los criterios de la norma UNE 153010 (ver AENOR 2012).

En la Tabla 5.1 se muestran los 5 niveles de cada uno de los ítems de la escala de Likert:

¹Para una descripción sobre cómo se debe realizar una escala de Likert consultar Guerra et al. (2016).

5. MÉTODOS.

Cuadro 5.1: Niveles de los ítems de la escala de Likert.

levels
No sé / No contesto
Muy en desacuerdo
En desacuerdo
Neutral
De acuerdo
Muy de acuerdo

En la Tabla 5.2 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 5.2: 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.

Los datos personales de los estudiantes se suministraron anonimizados para evitar ninguna referencia a su identidad. Del estudio se han eliminado a aquellos estudiantes que, a pesar de haber realizado la actividad, no dieron su autorización para que sus datos se utilizaran en un estudio científicos.

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 accesibilidad que utiliza 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.

5.2 Características del diseño del experimento.

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 análisis estadístico de los datos.

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».

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».

Las respuestas a un test de Likert se realizan en escala ordinal. No es adecuado realizar operaciones aritméticas para calcular medias con este tipo de datos. Pero ellos los test estadísticos para analizar el efecto de un tratamiento con respuesta

continúa como son *ANOVA* y *t-test* no son adecuados con datos ordinales. Según la investigación de Liddell y Kruschke (2018) ajustar datos ordinales con modelos cuantitativos puede producir los siguientes problemas:

- 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.

Una opción es tratar los datos ordinales como si se tratara de datos categóricos y utilizar técnicas no paramétricas como el test de *Kruskal – Wallis*. El problema de este tipo de técnicas es que ignoran que los datos tienen una escala y, en el caso particular del diseño que nos ocupa se trata de datos longitudinales, es decir, que se toman varias medidas de cada sujeto y, por lo tanto, los datos no son independientes. Agresti (2010) expone un catálogo de técnicas para analizar datos categóricos y ordinales.

5.3 Objetivo.

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

Son los estudiantes de un curso de accesibilidad capaces de encontrar los errores en el subtítulo de un vídeo. Para ello se propondrán diversos test y modelos estadísticos que tengan en consideración las características que se han comentado en el diseño del experimento (ver Sección 5.2). Particularmente se tendrá en cuenta que se trata de un diseño cruzado con variable respuesta ordinal y variables explicativas longitudinales.

5.4 Preprocesamiento.

Partiendo de los ficheros suministrados (ver Sección 5.2), se realiza el siguiente preprocesado (para ver el código ejecutado consultar Apéndice A):

1. 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.
2. Se eliminan del estudio a los estudiantes que aún habiendo realizado la actividad, no han dado su consentimiento para participar en el estudio.
3. El valor del campo *cohort* se sustituye por una letra *A* o *B* en función del grupo asignado. En este momento se desconoce qué vídeo vio primero cada grupo.

4. Se lee el fichero `profile` y se añade a los usuarios información sobre el sexo, el año de nacimiento y el nivel de estudios.
5. 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.
6. 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 ².
7. Se seleccionan las preguntas que contienen las respuestas y se renombran para que sea más fácil saber de qué pregunta se trata ³. Se convierte el campo `LastTry`, que contiene la fecha y hora de realización del test, a formato fecha y hora.
8. Se realizan algunas comprobaciones como la ausencia de valores nulos en la variables más relevantes o que no existan inconsistencias ni errores de procesado.
9. 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.
10. Se almacenan los resultados de los test preprocesado en un fichero `csv`.

²Se reitera que en este momento 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 o al que se vio después.

³En los ficheros suministrados la respuesta a cada pregunta ocupa varios campos y se selecciona en cada pregunta el que contiene el valor de la respuesta y se convierte a numérico.

5.5 Variables relevantes.

En la Tabla 5.3 se describen las características más relevantes de las principales variables que se utilizarán en el modelado y en el análisis estadístico.

Cuadro 5.3: Descripción de las variables más importantes

Nombre	Desc	Tipo	Valores
Response	Respuesta a las preguntas del test.	Factor ordenado	De 0 a 5 ¹
Treat	Subtítulos	Factor	A o B ²
Period	Periodo	Factor	1 ó 2 ³
Seq	Secuencia de aplicación de los tratamientos.	Factor	AB o BA
Subject	Identificación del estudiante	Factor	Númerico
Question	Número de la pregunta	Factor	Q01, Q02, ..., Q18 ⁴
Cluster	Grupo de la pregunta	Factor	1, 2, ó 3 ⁵

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

²No se conoce si el tratamiento A es el subtítulo bueno o lo es el B.

³1 para el primer vídeo visto y 2 el segundo.

⁴Se ha reorganizado de tal forma que Q18\$, que es la pregunta resumen, sea el valor primero y de referencia.

⁵Se aplicará una técnica estadística de agrupamiento para agregar las preguntas.

VARIABLES DEL MODELO

CAPÍTULO



RESULTADOS

CONCLUSIONES Y TRABAJO FUTURO

REFERENCIAS

- AENOR (2012). *UNE 153010 Subtitulado para personas sordas y personas con discapacidad auditiva*. Asociación Española de Normalización y Certificación.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. DOI: [10 . 1002 / 9780470594001](https://doi.org/10.1002/9780470594001).
- Guerra, A., T. Gidel y E. Vezzetti (mayo de 2016). «Toward a common procedure using likert and likert-type scales in small groups comparative design observations». En.
- 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>.
- 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>.
- Lui, K.-J. (ago. de 2016). *Crossover Designs: Testing, Estimation, and Sample Size*. DOI: [10.1002/9781119114710](https://doi.org/10.1002/9781119114710).
- 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).
- 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).
- Senn, S. (2022). Ed. por L. John Wiley. DOI: [10.1002/0470854596](https://doi.org/10.1002/0470854596).



PREPROCESADO DE LOS FICHEROS SUMINISTRADOS.

Este es el código en R con el que se transforman los ficheros que se suministran (ver Sección 5.2).

```
library(readr)
library(purrr)
library(dplyr)
library(magrittr)
library(stringr)
library(forcats)
library(testit)
library(tidyr)

##### GRADE #####
## Usuarios que no quieren participar
no_want_users <- read_lines("data/original/ids_a_eliminar.txt")

# Leemos todos los archivos de grade CSV
grade_files <- list.files(
  "data/original", pattern = ".*grade.*.csv", full.names = TRUE
)

grade_df <- map_dfr(
  grade_files, ~ read_delim(., delim = ";", show_col_types = FALSE) %>%
    # Añadimos el número de fila para mantener la trazabilidad
    mutate(Userid = row_number() + 1) %>%
    # Movemos las columnas de identificación de fila a la primera posición
    relocate(Userid, .before = 2) %>%
    # Renombramos las columnas para que empiecen con mayúsculas
    rename_with(~ str_to_title(.), everything()) %>%
    # Renombramos para que sea más fácil procesar el campo Cohort Name
```

A. PREPROCESADO DE LOS FICHEROS SUMINISTRADOS.

```
    rename("Cohort" = "Cohort Name") %>%
    # Eliminamos valores nulos y los que no quieren participar
    filter(!is.na(Cohort) & !Username %in% no_want_users)
  )

assert("Comprobamos que no hay usuarios duplicados", grade_df %>%
  nrow() == grade_df %>%
  distinct(Username) %>%
  nrow())

# Creamos un tibble que tiene un campo con letras en lugar del valor de Cohorte
(groups <- grade_df %>%
  distinct(Cohort) %>%
  arrange(Cohort) %>%
  mutate(Group = LETTERS[1:n()])))

# Unimos los tibbles para asignar en grupo como letra en lugar de la cohorte
grade_df <- left_join(grade_df, groups) %>% dplyr::select(Username, Userid, Group)

##### PROFILE #####
profile_files <- list.files(
  "data/original", pattern = ".*student_profile.*.csv", full.names = TRUE
)

profile_df <- map_dfr(
  profile_files, ~ read_delim(.x, delim = ";", show_col_types = FALSE)
)

grade_df <- left_join(
  grade_df, profile_df %>% dplyr::select(-cohort), by = join_by(Username == username)
)

##### CONOC #####
conoc_files <- list.files(
  "data/original", pattern = ".*conoc.*.csv", full.names = TRUE)

conoc_df <- map_dfr(
  conoc_files, ~ read_delim(.x, delim = ";", show_col_types = FALSE)
)

conoc_df <- conoc_df %>%
  filter(Tries == 1) %>%
  rowwise() %>%
  mutate(
    level_of_knowledge =
      sum(c_across(starts_with(paste("Q", 1:10, "C", sep = ""))) == "correct")
  ) %>%
  dplyr::select(User, level_of_knowledge)

grade_df <- left_join(grade_df, conoc_df, by = join_by(Username == User))

##### TEST #####
# Leemos todos los archivos de test CSV
```

```

test_files <- list.files(
  "data/original", pattern = ".*test.*.csv", full.names = TRUE
)

# Leer todos los archivos de test y los combinamos en un dataframe
test_df <- map_dfr(
  test_files, ~ read_delim(.x, delim = ";", show_col_types = FALSE) %>%
    # Añadimos un número de fila para mantener la trazabilidad
    mutate(Row = row_number() + 1) %>%
    # Añadimos la columna del número de test
    mutate(Test = sprintf("%02d", as.integer(str_extract(.x, "(?<=test)\\d+")))) %>%
    # Movemos las columnas de identificación de test y fila a la primera posición
    relocate(c(Test, Row), .before = 2)
) %>%
  # eliminamos los usuarios que no quieren participar
  filter(!User %in% no_want_users)

num_questions <- 18

# Nombre de los campos que contienen las respuestas al test
questions_original <- paste(
  "Q", seq(from = 1, by = 2, length.out = num_questions), "R", sep = ""
)

# Nombre de los campos que contienen las respuestas al test
comments_original <- paste(
  "Q", seq(from = 2, by = 2, length.out = num_questions - 1), "R", sep = ""
)

# Nombre de los campos que se usarán para renombrar los campos de respuesta al test
questions <- sprintf("Q%02d", seq(from = 1, by = 1, length.out = num_questions))
comments <- sprintf("C%02d", seq(from = 1, by = 1, length.out = num_questions - 1))
columns <- c(
  "Row", "Test", "User", "LastTry", questions_original, comments_original
)

# Procesamos el dataframe
# Con este operador del paquete magrittr hacemos las transformaciones in situ
test_df %<>%
  # Eliminamos las filas que no contienen información
  filter(Tries > 0) %>%
  # Convertimos LastTry a formato fecha
  mutate(LastTry = strptime(LastTry, format = "%Y-%m-%dT%H:%M:%SZ")) %>%
  # Seleccionamos las columnas que nos interesan
  dplyr::select(all_of(columns)) %>%
  # Extraemos la puntuación numérica de la pregunta
  mutate(across(questions_original, ~ if_else(
    startsWith(.x, "choice_"), as.integer(str_extract(.x, "\\d+")), NA_integer_)
  )) %>%
  # Renombramos los respuestas para que sean secuenciales
  rename(
    setNames(questions_original, questions),
    setNames(comments_original, comments)
  ) %>%

```

A. PREPROCESADO DE LOS FICHEROS SUMINISTRADOS.

```
# nos aseguramos de que el orden filas es el mismo que el de los ficheros.
arrange("Test", "Row")

# Guardamos el número de filas para posterior comprobación
n_test <- test_df %>% nrow()

# Unimos los dataframes para tener el grupo y el UserID secuencial
test_df <- inner_join(
  test_df, grade_df, by = join_by(User == Username)
) %>% relocate(Group, .before = 2)

# Cambiamos los valores del campo User por los del UserID
test_df %<>%
  mutate(User = Userid) %>%
  dplyr::select(-Userid) %>%
  arrange(User, Test) # Ordenamos por usuario y test

##### CHECKS #####
assert(
  "Comprobamos que no hay preguntas duplicadas en el dataframe de test",
  n_test == test_df %>%
    distinct(Group, Test, User) %>%
    nrow()
)

assert(
  "Comprobamos que no hay valores nulos",
  test_df %>%
    dplyr::select(
      -c(comments, year_of_birth, gender, level_of_education, level_of_knowledge)
    ) %>% filter(if_any(everything(), is.na)) %>% nrow() == 0)

assert(
  "Comprobamos que no hay respuestas con valores incorrectos",
  sum(sort(unique(unlist(
    test_df %>% dplyr::select(all_of(questions))
  ))) == 0:5) == 6)

comments_df <- test_df %>%
  pivot_longer(
    cols = starts_with(c("Q", "C")),
    names_to = c(".value", "Question"),
    names_pattern = "(Q|C)(.*)" %>%
    rename(Response = Q, Comment = C) %>%
    filter(!is.na(Comment) & grepl("[a-zA-Z]", Comment)) %>%
    dplyr::select(Test, Row, Group, User, Question, Response, Comment) %>%
    arrange(Test, Group, Response, Row)

write_csv(comments_df, "../data/preprocess/comments.csv")
```

```
##### SAVE TO FILE #####  
write_csv(  
  test_df %>% dplyr::select(-all_of(comments)), "./data/preprocess/test.csv"  
)
```


Abstract English abstract, on the last page.

This is a bookdown template based on LaTeX memoir class.

Keywords Keyword in English, As a list.