

Tarea 2

Francisco Herrera Barajas, fherrera110@alumno.uned.es

20-04-2025

Ejercicio 1

Importa la base de datos Datos_Tarea2.txt a R Studio. Calcula los estadísticos descriptivos de la variable dependiente confianza y de las variables independientes c_loun_pc_une_2022 y eisced, así como sus histogramas. Para el resto de los análisis, se asume que estas variables se distribuyen normalmente. En el documento Word, incluya un pantallazo de estos resultados, incluyendo su histograma, y su interpretación.

Cargamos las librerías necesarias

```
library(psych)
library(tidyr)
library(reshape2)
library(lme4)
library(lmerTest)
library(emmeans)
library(pbkrtest)
library(afex)
library(ggplot2)
library(dplyr)
```

Cargamos la base de datos y comprobamos si hay NA

```
datos <- read.table("Datos_Tarea2.txt")

nNA <- sum(is.na(data.frame(datos)))
print(paste("Número de Na: ", nNA))
```

```
## [1] "Número de Na: 0"
```

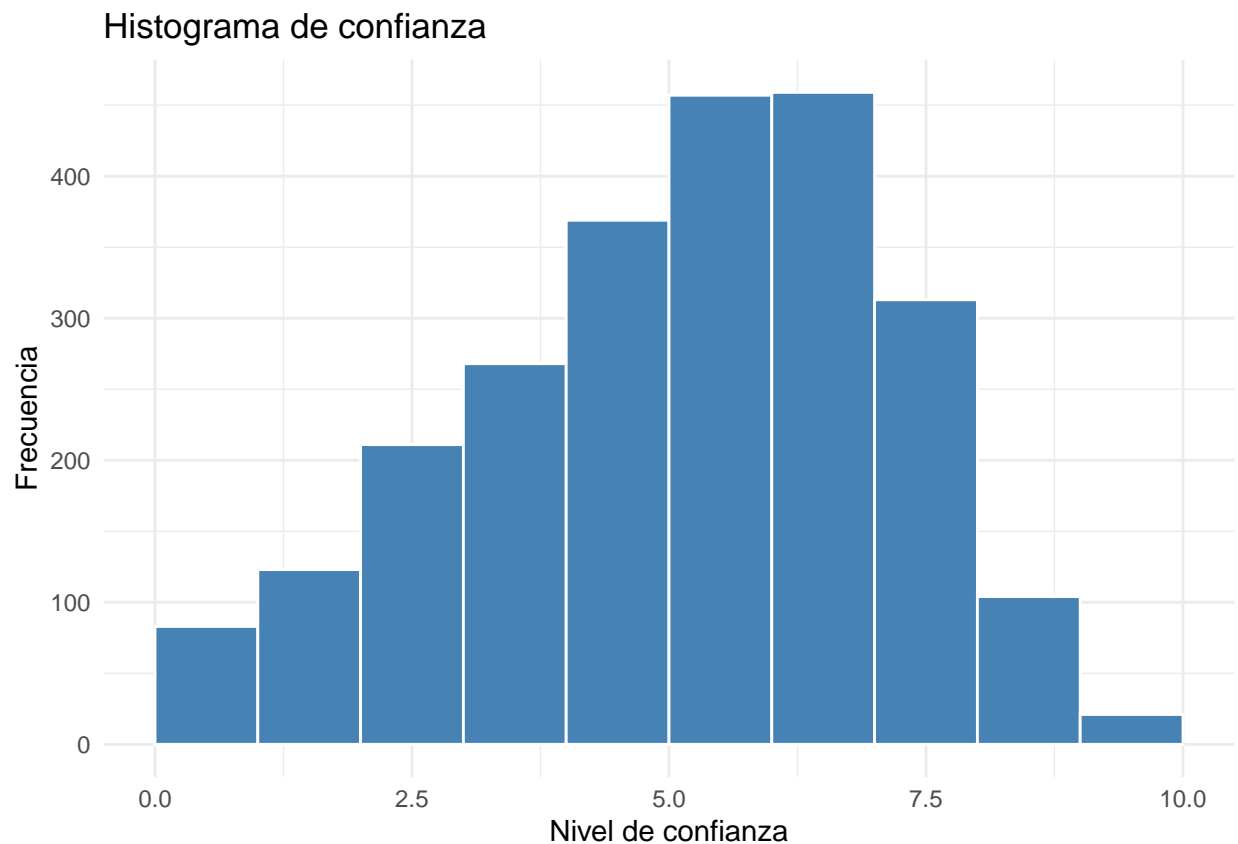
Estadísticos descriptivos

Variable Confianza

```
describe(datos$confianza)
```

```
##      vars      n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 2408 5.16 2.02   5.38   5.26 2.04   0  10   10 -0.4   -0.42 0.04
```

```
ggplot(datos, aes(x = confianza)) +
  geom_histogram(binwidth = 1, boundary = 0, fill = "steelblue", color = "white") +
  labs(title = "Histograma de confianza", x = "Nivel de confianza", y = "Frecuencia") +
  theme_minimal()
```



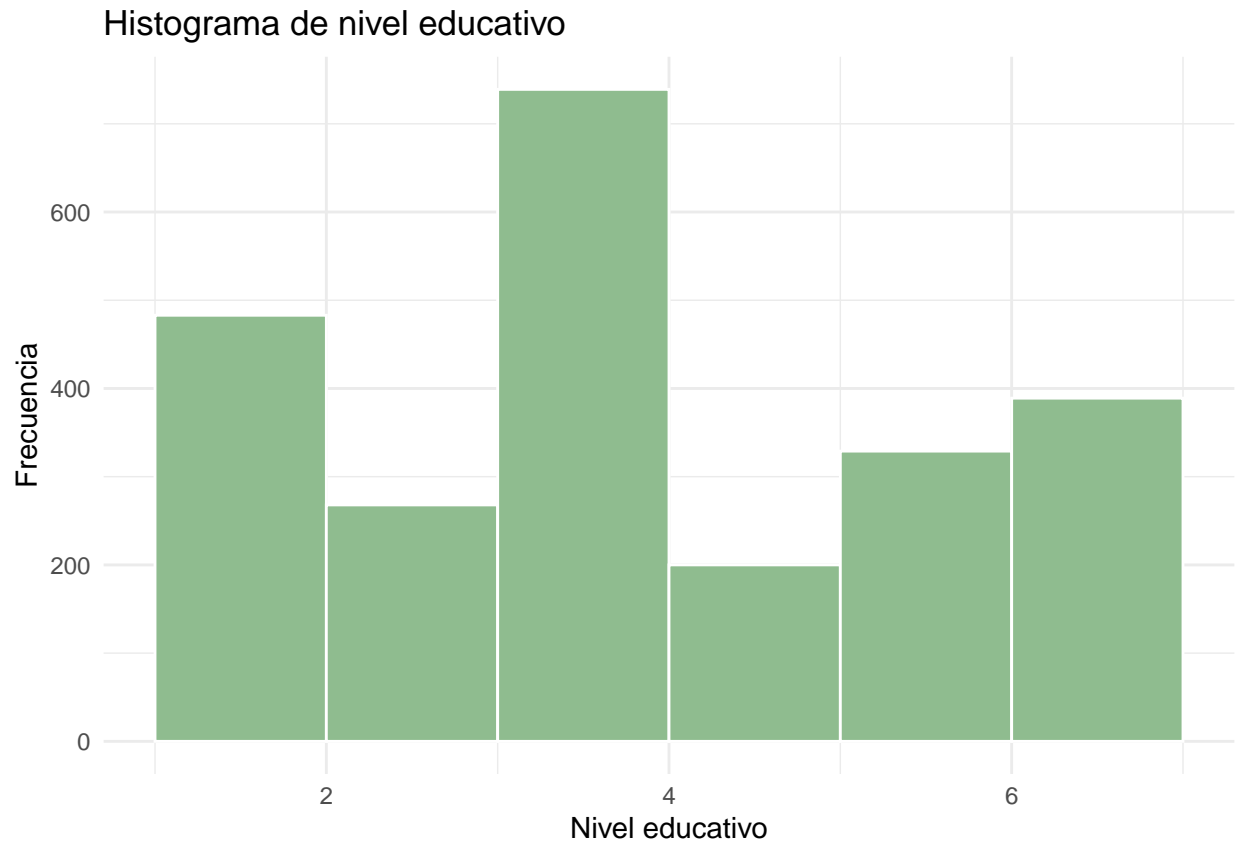
Esta variable confianza es un promedio de una escala del 1 al 10, en ella se refleja la confianza de los ciudadanos de varios países europeos en varios aspectos (Políticos, policías, sistema legal, UE, científicos etc). La media es 5.16 lo que indica una aprobado raspado en mucho de estos aspectos. Por otro lado, el skew -0,4, nos muestra que la distribución es ligeramente asimétrica, muchos de los valores se agrupan a la derecha de la distribución. La dispersión no es muy alta en torno a 2 puntuaciones respecto a la media. Finalmente, el error típico es bastante bajo, lo que indica una alta precisión en la estimación de la media muestral. Esto nos permite confiar en que la media observada es un buen estimador de la media poblacional.

Variable Nivel educativo

```
describe(datos$eisced)
```

```
##      vars      n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 2408 4.27 1.79      4    4.29 1.48   1  7    6 0.02    -0.98 0.04
```

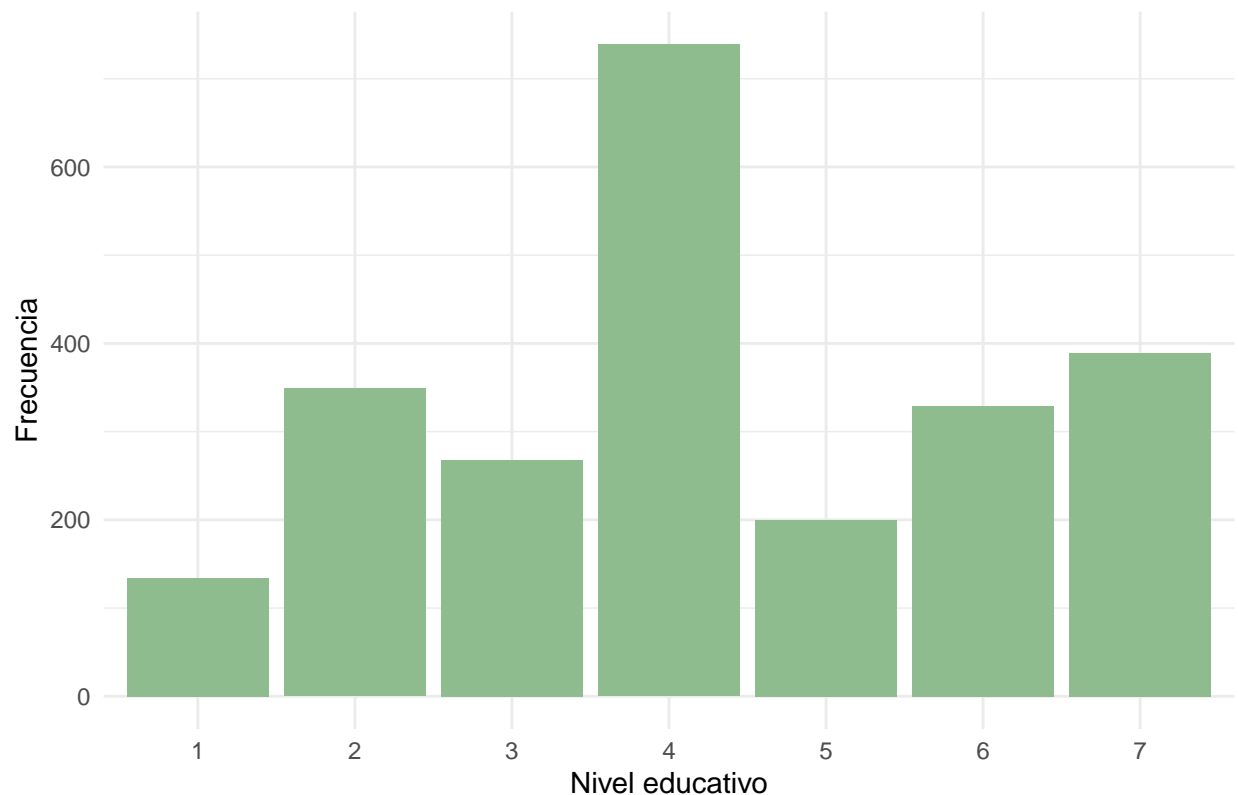
```
ggplot(datos, aes(x = eisced)) +
  geom_histogram(binwidth = 1, boundary = 0, fill = "darkseagreen", color = "white") +
  labs(title = "Histograma de nivel educativo", x = "Nivel educativo", y = "Frecuencia") +
  theme_minimal()
```



```
# Convertir en factor
datos$eisced <- factor(datos$eisced, levels = sort(unique(datos$eisced)))

ggplot(datos, aes(x = eisced)) +
  geom_bar(fill = "darkseagreen") +
  labs(title = "Gráfico de barras del nivel educativo",
       x = "Nivel educativo",
       y = "Frecuencia") +
  theme_minimal()
```

Gráfico de barras del nivel educativo



La variable nivel educativo es una variable ordinal del 1 al 7, por ello aunque la he gráfico primero como un histograma, con el objetivo de visualizar correctamente esta variable, he decido realizar un diagrama de barras. Si nos fijamos en los estadísticos la media esta en el nivel educativo 4 el cual corresponde con un título de bachillerato o de técnico de formación profesional. En estos descriptivos destaca la curtosis -0,98, esta nos indica que la forma de la distribución tiende a estar más aplanada que una curva normal, asimismo, teniendo menos datos concentrados alrededor de la media.

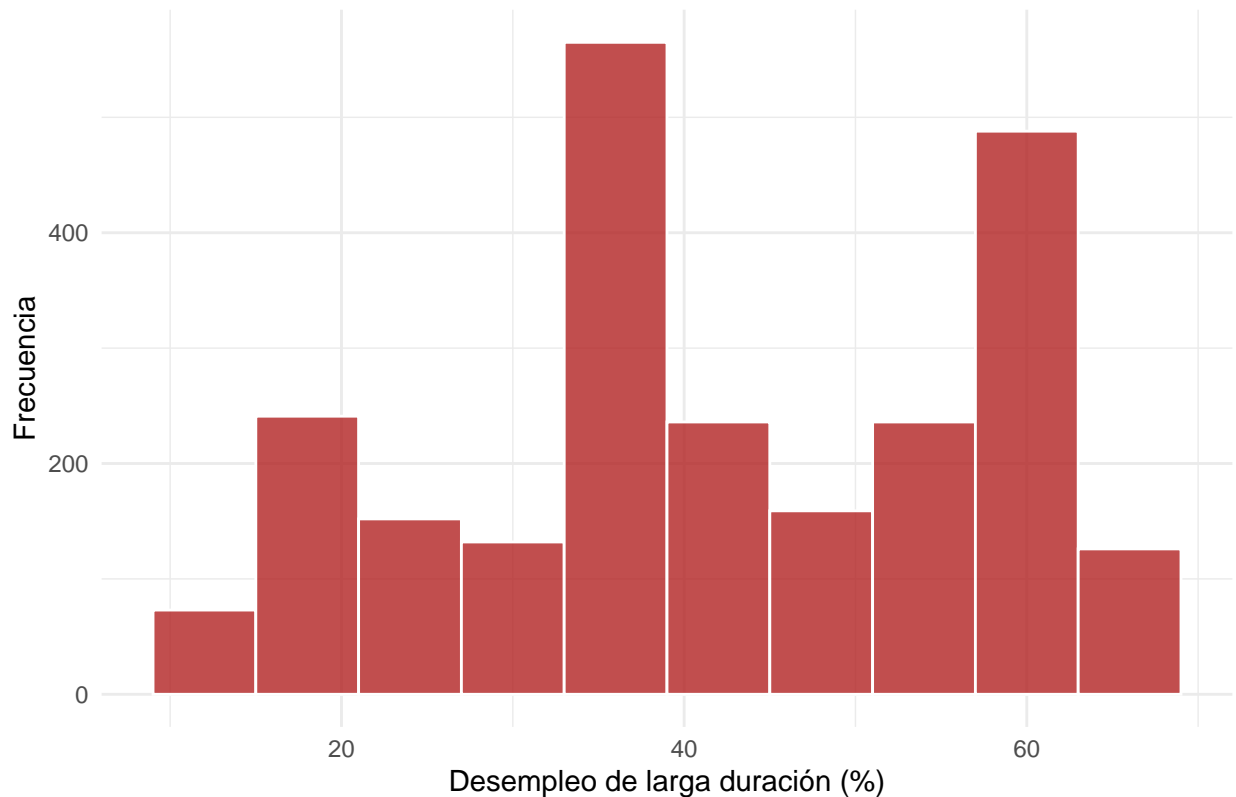
Variable Desempleo de larga duración

```
describe(datos$c_loun_pc_une_2022)
```

```
##   vars    n mean   sd median trimmed  mad  min  max range  skew kurtosis
## X1     1 2408 41.49 15.26   39.1   41.72 22.09 12.5 66.5   54 -0.05   -1.08
##      se
## X1 0.31
```

```
ggplot(datos, aes(x = c_loun_pc_une_2022)) +
  geom_histogram(binwidth = 6, fill = "firebrick", color = "white", alpha = 0.8) +
  labs(
    title = "Histograma del desempleo de larga duración",
    x = "Desempleo de larga duración (%)",
    y = "Frecuencia"
  ) +
  theme_minimal()
```

Histograma del desempleo de larga duración



Esta variable entiendo que mide el porcentaje de desempleados de larga duración entre los desempleados totales en distintos países europeos. La media es 41,49% lo que nos indica que un alto porcentaje de los desempleados europeos son desempleados de larga duración. La desviación típica de los datos es de 15,26, esto nos indica una gran diferencia entre los distintos países europeos. Por otro lado, la curtosis es de -1,11 lo que muestra que es bastante plana la distribución, esto nos indica que los valores se encuentran muy dispersos.

Ejercicio 2.

Ajusta un modelo multinivel nulo (llamado modelo_0), en el que la variable dependiente sea la confianza en las instituciones (confianza), y el efecto aleatorio de segundo nivel sea la variable país (cntry). Realiza las siguientes acciones: -Interpreta la intersección -Halla el coeficiente de correlación intraclase e interprétalo En el documento Word, incluya un pantallazo de estos resultados, y su interpretación).

Coefficientes Modelo 0

```
modelo_0<-lmer(confianza ~ 1 + (1 | cntry), data = datos)
summary(modelo_0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: confianza ~ 1 + (1 | cntry)
```

```
## Data: datos
##
## REML criterion at convergence: 9681
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.0797 -0.6353  0.0869  0.6983  3.1626
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   cntry    (Intercept) 0.9838   0.9919
##   Residual                3.1809   1.7835
## Number of obs: 2408, groups:  cntry, 16
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    5.2722     0.2509 14.9911   21.02 1.55e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La intersección o intercepto es de 5,2722 (SE = 0,25 , $p < 0,0001$). Esto representa el nivel medio de confianza en un país promedio, que sea significativo lo único que indica es que la media es distinta de 0.

CCI Modelo 0

```
var_comp_0 <- as.data.frame(VarCorr(modelo_0))
var_entre_paises_0 <- var_comp_0$vcov[1] # Varianza del intercepto aleatorio
var_residual_0 <- var_comp_0$vcov[2]    # Varianza residual

cci_0 <- var_entre_paises_0 / (var_entre_paises_0 + var_residual_0)
cci_0
```

```
## [1] 0.2362274
```

El coeficiente de correlación intraclase (ICC) fue de 0.236, lo que indica que aproximadamente el 23% de la varianza en los niveles de confianza puede atribuirse a diferencias entre países. El 77% restante se debe a diferencias entre personas dentro de un país.

Ejercicio 3

Ajusta un nuevo modelo (llamado modelo_1), en el que se incluya una variable explicativa de segundo nivel referente a los países: Porcentaje de desempleo de larga duración (c_loun_pc_une_2022). Ten en cuenta que esta variable tiene que centrarse previamente. Realiza las siguientes acciones: -Interpreta los resultados: la intersección, el efecto de la variable Porcentaje de desempleo de larga duración, el coeficiente de correlación intraclase, y el porcentaje en el que se ha reducido la varianza entre países e intra-países gracias a la incorporación de esta variable en el modelo. -Compara el desajuste de este modelo con respecto al modelo nulo. En el documento Word, incluya un pantallazo de estos resultados, y su interpretación

Coeficientes Modelo 1

```
# Centramos la variable
datos <- datos %>%
  mutate(c_loun_pc_une_2022_c = c_loun_pc_une_2022 - mean(c_loun_pc_une_2022, na.rm = TRUE))

# Ajustamos el modelo
modelo_1 <- lmer(confianza ~ c_loun_pc_une_2022_c + (1 | cntry), data = datos)
summary(modelo_1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: confianza ~ c_loun_pc_une_2022_c + (1 | cntry)
## Data: datos
##
## REML criterion at convergence: 9677.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.0870 -0.6430  0.0938  0.6978  3.1795
##
## Random effects:
## Groups   Name            Variance Std.Dev.
## cntry    (Intercept)  0.5265     0.7256
## Residual                3.1808     1.7835
## Number of obs: 2408, groups: cntry, 16
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)    5.12759    0.18933 14.05892  27.083 1.56e-13 ***
## c_loun_pc_une_2022_c -0.04436    0.01205 14.21173  -3.682  0.00241 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## c_l__2022_  0.204
```

En este modelo fijandonos en el bloque de fixed effects podemos ver el resultado de los coeficientes de la recta. El intercepto representa el nivel medio de confianza cuando el porcentaje de desempleo de larga duración está en la media europea. Este significativo pero esto no es importante porque solo nos dice que es distinto de 0. Asimismo, si nos fijamos en el coeficiente del desempleo encontramos que también es significativo y su valor es negativo. Esto lo podemos interpretar como que por cada punto porcentual de desempleo más el nivel de confianza se reduce 0.44 en los distintos países. Es decir, a mayor desempleo estructural, menor confianza ciudadana.

CCI Modelo 1

```
var_comp_1 <- as.data.frame(VarCorr(modelo_1))
var_entre_paises_1 <- var_comp_1$vcov[1]
```

```
var_residual_1 <- var_comp_1$vcov[2]

cci_1 <- var_entre_paises_1 / (var_entre_paises_1 + var_residual_1)
cci_1
```

```
## [1] 0.1420176
```

El CCI es igual al 0,142. Este resultado nos indica que en este modelo solo el 14% de la varianza se explica por la diferencia entre países. El resto de la varianza se explica por factores individuales dentro de cada país.

Reducción de la varianza (efecto del predictor)

```
reduc_1 <- (var_entre_paises_0 - var_entre_paises_1) / var_entre_paises_0
reduc_1
```

```
## [1] 0.4648405
```

Esto significa que la varianza entre países se reduce en un 46% al incorporar en el modelo la variable desempleo de larga duración. Por tanto, esta variable explica casi la mitad de las diferencias entre países en los niveles de confianza ciudadana.

Comparación Modelo 0 y Modelo 1

```
anova(modelo_0, modelo_1)
```

```
## refitting model(s) with ML (instead of REML)

## Data: datos
## Models:
## modelo_0: confianza ~ 1 + (1 | cntry)
## modelo_1: confianza ~ c_loun_pc_une_2022_c + (1 | cntry)
##          npar    AIC    BIC logLik -2*log(L)  Chisq Df Pr(>Chisq)
## modelo_0     3 9686.1 9703.4 -4840.0   9680.1
## modelo_1     4 9677.2 9700.4 -4834.6   9669.2 10.869  1 0.0009779 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La comparación de modelos indica que el modelo que incluye el porcentaje de desempleo de larga duración (modelo_1) ajusta significativamente mejor a los datos que el modelo nulo (modelo_0). Esto confirma que la inclusión de la variable desempleo de larga duración mejora la capacidad predictiva del modelo de explicar los diferentes niveles de confianza entre países.

Ejercicio 4

Ajusta un nuevo modelo (llamado `modelo_2`), en el que aparte de la Porcentaje de desempleo de larga duración (`c_loun_pc_une_2022`) también incluya la variable explicativa de primer nivel (referente a los participantes) nivel educativo (`eisced`). Ten en cuenta que la variable nivel educativo tiene que centrarse previamente. Realiza las siguientes acciones:

-Interpreta los resultados: la intersección, el efecto de las variables independientes, el coeficiente de correlación intraclase, y el porcentaje en el que se ha reducido la varianza entre países e intra-países gracias a la incorporación de la variable nivel educativo (`eisced`). -Compara el desajuste de este modelo con respecto al modelo uno.

En el documento Word, incluya un pantallazo de estos resultados, y su interpretación

Coeficientes

```
datos <- datos %>%
  mutate(eisced = as.numeric(as.character(eisced)), # Convertir la variable en numérica
         eisced_c = eisced - mean(eisced, na.rm = TRUE))

modelo_2 <- lmer(confianza ~ eisced + c_loun_pc_une_2022_c + (1 | cntry), data = datos)
summary(modelo_2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: confianza ~ eisced + c_loun_pc_une_2022_c + (1 | cntry)
## Data: datos
##
## REML criterion at convergence: 9621.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2494 -0.6386  0.0727  0.6692  3.2223
##
## Random effects:
## Groups   Name            Variance Std.Dev.
## cntry    (Intercept)  0.5175    0.7194
## Residual                3.1001    1.7607
## Number of obs: 2408, groups: cntry, 16
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    4.42422    0.20741  20.96507  21.331 1.07e-15 ***
## eisced          0.16509    0.02072 2398.87788   7.967 2.48e-15 ***
## c_loun_pc_une_2022_c -0.04202    0.01195  14.24101  -3.517 0.00334 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) eisced
## eisced       -0.426
## c_l__2022_    0.174  0.025
```

En este modelo 2 el intercepto representa el nivel medio de confianza cuando tanto el desempeño de larga duración y el nivel educativo esta en su media. Siguiendo la interpretación de los modelos anteriores, el hecho de que el intercepto sea estadísticamente significativo no tiene un valor interpretativo relevante, ya que simplemente indica que la media de confianza es distinta de cero. La interpretación del coeficiente de desempleo de larga duración sería la misma que en el anterior modelo. Sin embargo, si que sería interesante interpretar la nueva variable añadida, el nivel educativo. El coeficiente asociado a esta variable es positivo y significativo, lo que indica que, por cada punto que aumenta la variable nivel educativo, del 1 al 7, los individuos tienen mayor confianza en sus instituciones, la política, el sistema legal, la ciencia, etc. En resumen, a mayor nivel educativo, mayor confianza.

CCI Modelo 2

```
var_comp_2 <- as.data.frame(VarCorr(modelo_2))
var_entre_paises_2 <- var_comp_2$vcov[1]
var_residual_2 <- var_comp_2$vcov[2]

cci_2 <- var_entre_paises_2 / (var_entre_paises_2 + var_residual_2)
cci_2
```

```
## [1] 0.1430526
```

El CCI es igual al 0,143. Este resultado es muy parecido al de modelo anterior, repitiendo la anterior explicación nos indica que en este modelo solo el 14% de la varianza se explica por la diferencia entre países. El resto de la varianza se explica por factores individuales dentro de cada país.

Reducción de la varianza (efecto del predictor)

```
reduc_2 <- (var_entre_paises_1 - var_entre_paises_2) / var_entre_paises_1
reduc_2
```

```
## [1] 0.01706728
```

Al añadir el nivel educativo individual al modelo, la varianza entre países se reduce tan solo en un 1.7% adicional respecto al modelo 1. Esto muestra que esta nueva variable aporta muy poco a explicar las diferencias entre países en los niveles de confianza. Pese a ello la variable si que es relevante a nivel individual, ya que explica varianza en el primer nivel del modelo, es decir, las personas.

Comparación Modelo 1 y 2

```
anova(modelo_1, modelo_2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: datos
## Models:
## modelo_1: confianza ~ c_loun_pc_une_2022_c + (1 | cntry)
```

```
## modelo_2: confianza ~ eisced + c_loun_pc_une_2022_c + (1 | cntry)
##           npar    AIC    BIC logLik -2*log(L)  Chisq Df Pr(>Chisq)
## modelo_1      4 9677.2 9700.4 -4834.6    9669.2
## modelo_2      5 9616.5 9645.4 -4803.3    9606.5 62.715  1 2.388e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con estos resultados podemos afirmar que la inclusión de la variable nivel educativo mejora significativamente la capacidad predictiva del modelo.

Ejercicio 5

Ajusta un último modelo, llamado modelo 3, en el que añadas pendientes aleatorias para el coeficiente nivel educativo, es decir, un modelo multinivel en el que se cuantifique la variabilidad entre países de la pendiente que relaciona el nivel educativo con la confianza en las instituciones. No hace falta interpretar los coeficientes del modelo en este caso, sólo hay que realizar una prueba de razón de verosimilitudes que permita concluir si la variabilidad entre pendientes es estadísticamente relevante o no. Realiza las siguientes acciones: - Interpreta el resultado de esta prueba de razón de verosimilitudes y explica las implicaciones prácticas de este resultado. En el documento Word, incluya un pantallazo de estos resultados, y su interpretación.

Comparación Modelo 2 y Modelo 3

```
modelo_3 <- lmer(confianza ~ c_loun_pc_une_2022_c + eisced_c + (1 + eisced_c | cntry), data = datos)
anova(modelo_2, modelo_3)
```

```
## refitting model(s) with ML (instead of REML)

## Data: datos
## Models:
## modelo_2: confianza ~ eisced + c_loun_pc_une_2022_c + (1 | cntry)
## modelo_3: confianza ~ c_loun_pc_une_2022_c + eisced_c + (1 + eisced_c | cntry)
##           npar    AIC    BIC logLik -2*log(L)  Chisq Df Pr(>Chisq)
## modelo_2      5 9616.5 9645.4 -4803.3    9606.5
## modelo_3      7 9620.5 9661.0 -4803.2    9606.5 0.0174  2    0.9913
```

La comparación entre estos dos modelos no resulta significativa, lo que indica que la inclusión de pendientes aleatorias para el nivel educativo no mejora el poder explicativo del modelo. Esto sugiere que el efecto del nivel educativo sobre la confianza puede considerarse constante entre los países analizados, ya que no se observa una variabilidad significativa en dicha pendiente a nivel país.