

# Evaluación curso R

Laura Romero Muñoz

## Entrega

- Recuerda completar tu **nombre en el apartado author**.
- Además del código, no olvides completar las respuestas a las preguntas indicadas en negrita.
- Si tienes dudas/dificultades, **puedes contactar con los profesores**.
- Deadline:
- Puedes realizar la entrega en el mail: **constantino.garciana@ceu.es**.
- La entrega consistirá en el fichero que se genera al hacer Knit (un fichero **html** o **pdf**).

## Apuestas de adolescentes en UK

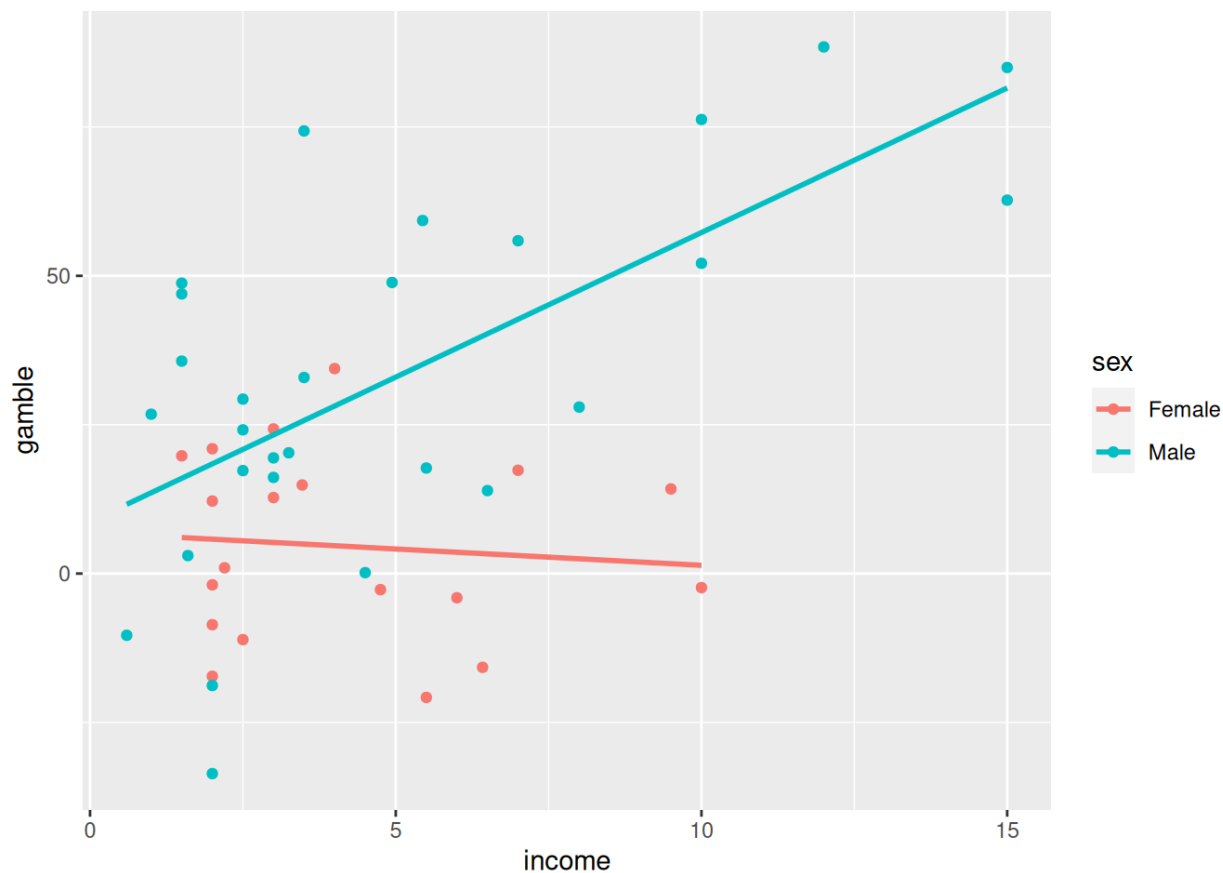
El conjunto de datos `teengamb.csv` contiene datos sobre las tasas de juego entre los adolescentes en Gran Bretaña, su género y estatus socioeconómico. Una pregunta que nos puede interesar es **si los ingresos del adolescente y su sexo influyen en la cantidad de dinero apostado (céntrate solo en las variables `income`, `sex` y `gamble`)** ... Sigue los siguientes pasos para crear un ANCOVA...

### 1a) Carga los datos...

```
# Carga los datos
teengamb = read.csv("teengamb.csv")
```

### 1b) Visualiza los datos...

...para valorar si el modelo debe incluir interacciones. Para ello, escribe el código que genera una gráfica similar a la siguiente:



```
library("tidyverse") # carga ggplot
```

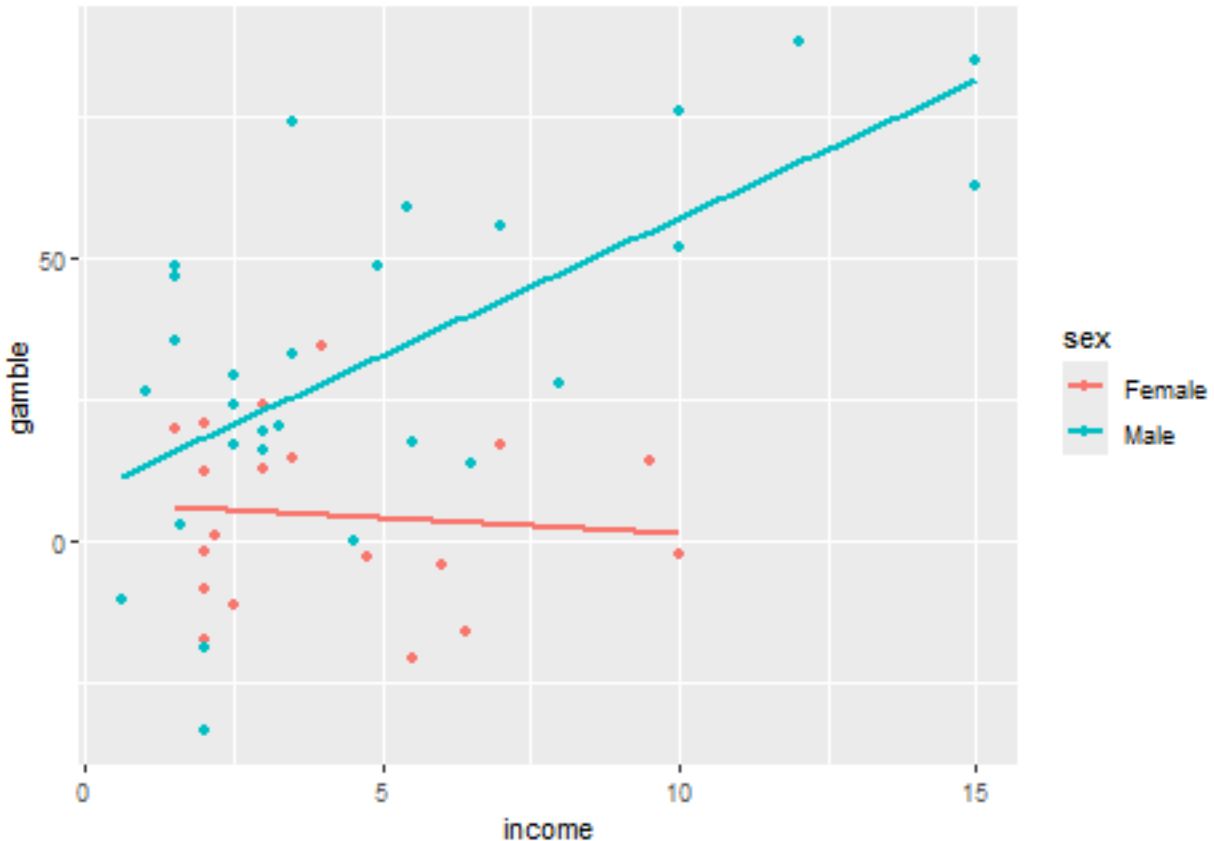
```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.2      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# completa el siguiente código
```

```
ggplot(teengamb, aes(x = income, col = sex, y = gamble)) +
  geom_point(aes(x = income, y = gamble)) + # añade la geometría adecuada para los puntos
  geom_smooth(method = "lm", se = FALSE) # añade las rectas de regresión
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



La gráfica anterior apoya que hay interacciones entre las variables `sex` e `income`. **Explica los motivos brevemente** (Pista: ¿son las rectas paralelas?) Las rectas no son paralelas, lo que sugiere la existencia de una interacción entre las variables, puesto que la pendiente entre `income` y `gamble` no es la misma para los dos sexos.

## 2) Crea el modelo...

... y obtén los intervalos de confianza para los coeficientes y su significación.

```
# Crea un modelo con interacciones en base a tu conclusión del apartado anterior
teengamb$sex <- as.factor(teengamb$sex)
gamb_model <- lm(gamble ~ income * sex , data = teengamb)
# o gamb_model <- lm(gamble ~ income * sex, data = teengamb)

# Obtén p-valores e intervalos de confianza. Usa summary y confint
summary(gamb_model)
```

```
##
## Call:
## lm(formula = gamble ~ income * sex, data = teengamb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -52.025 -15.479  -3.559  14.022  48.625
##
```

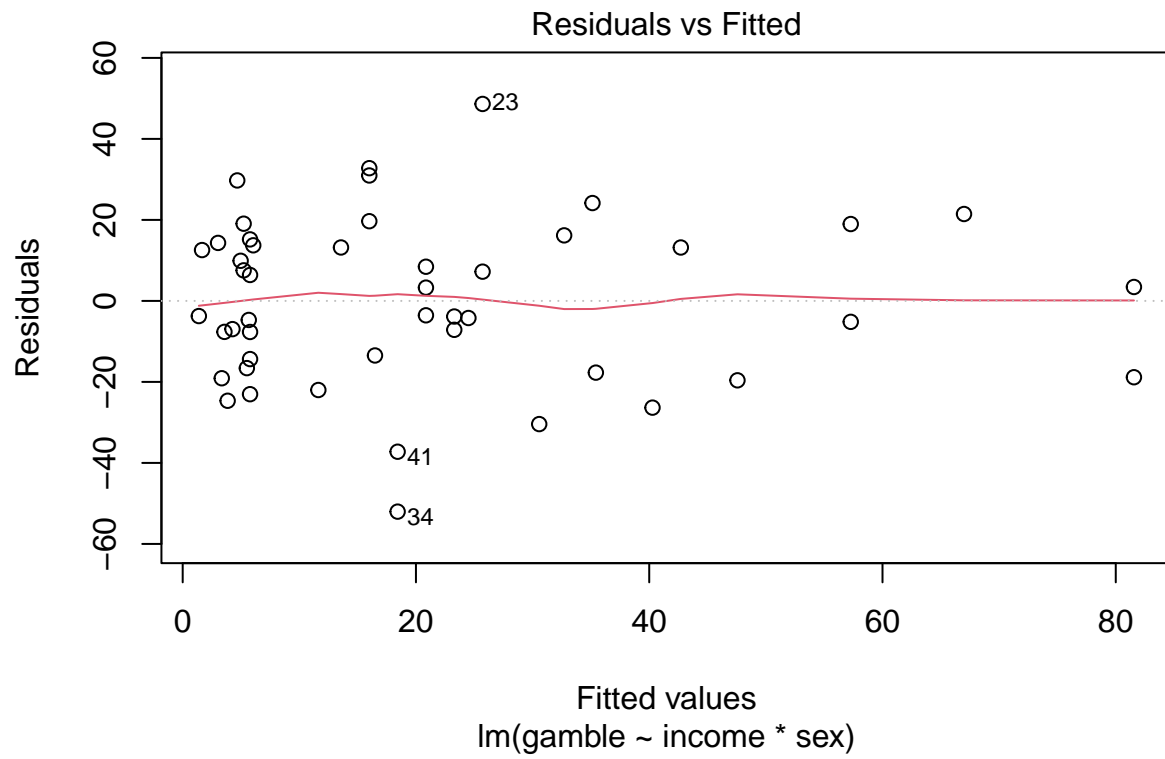
```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.8743     9.2446   0.744  0.4612
## income        -0.5489     1.9024  -0.289  0.7743
## sexMale         1.8427    11.1946   0.165  0.8700
## income:sexMale  5.4050     2.1435   2.522  0.0155 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.97 on 43 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.4764
## F-statistic: 14.95 on 3 and 43 DF,  p-value: 8.279e-07
```

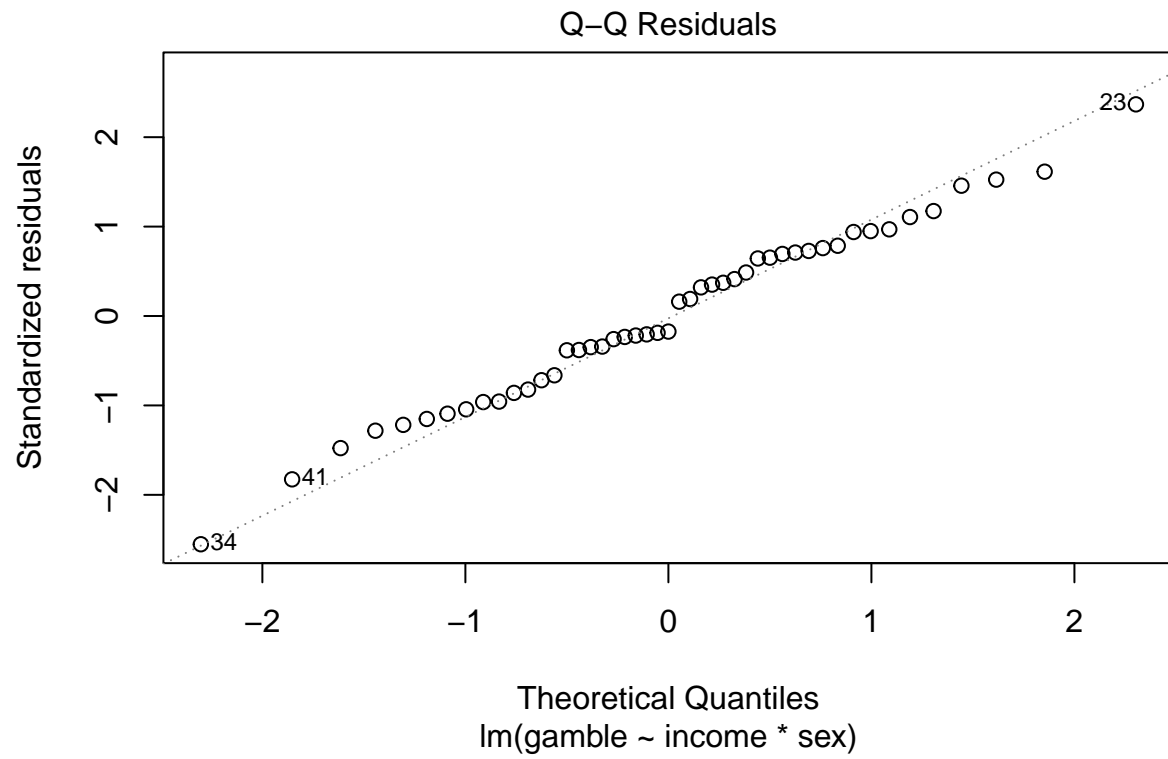
```
confint(gamb_model)
```

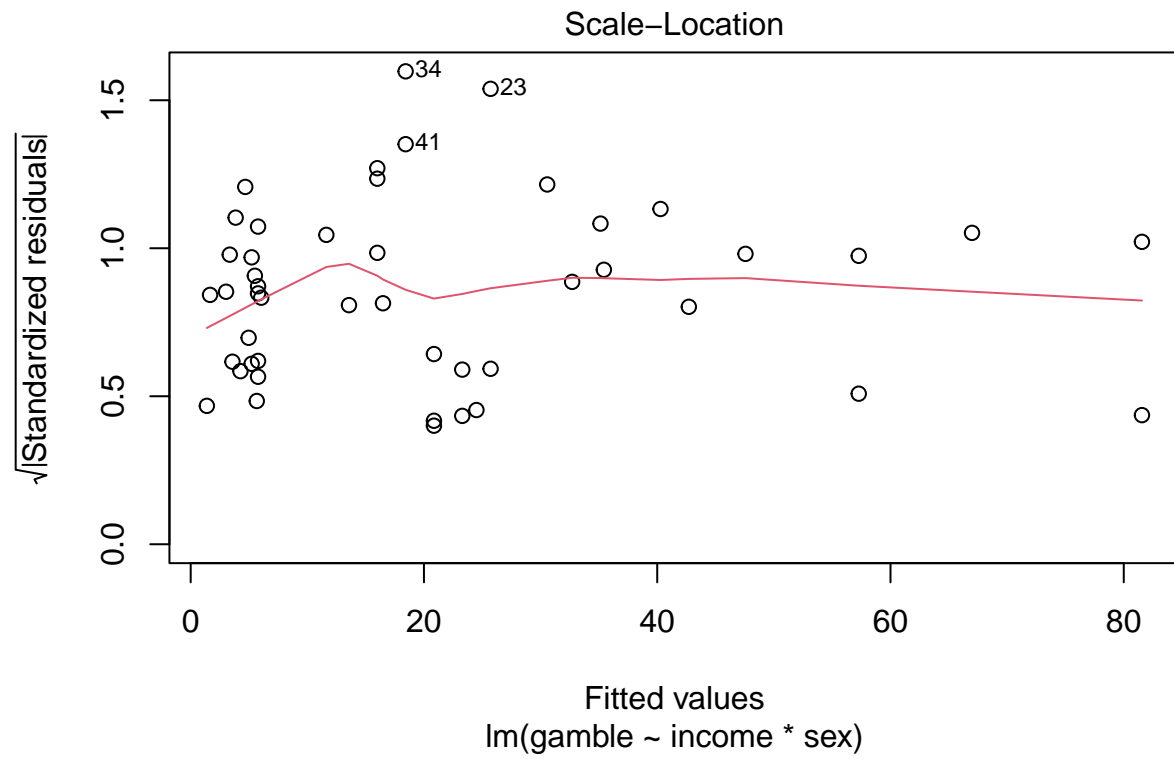
```
##           2.5 %    97.5 %
## (Intercept) -11.769152 25.517839
## income      -4.385455  3.287728
## sexMale     -20.733425 24.418773
## income:sexMale  1.082260  9.727751
```

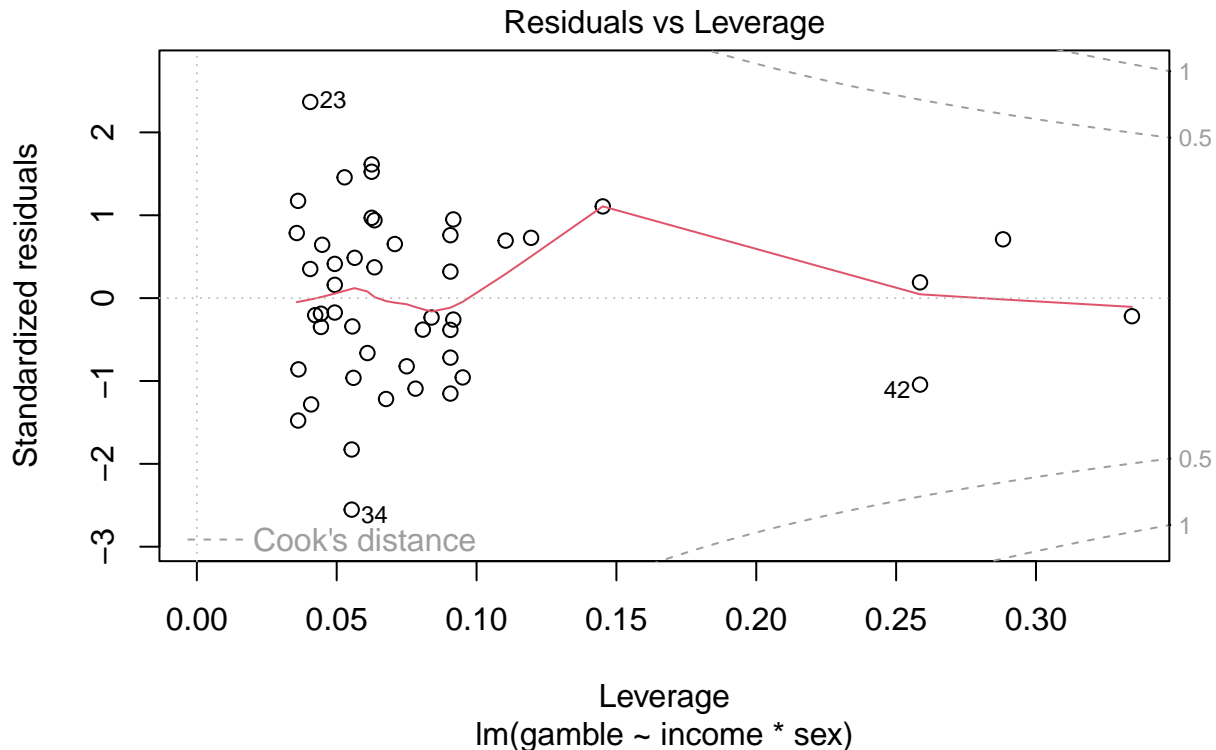
### 3) Valora si las asunciones del modelo se cumplen

```
# Interpretar las 4 gráficas siguientes para decidir si las asunciones del modelo
# se cumplen
plot(gamb_model, ask = FALSE)
```









**¿Es el modelo correcto?:** Sí el modelo es correcto. En la gráfica residual vs fitted los puntos están alrededor del 0 sin formar una curva o un embudo, por lo que tenemos linealidad y homocedasticidad. En la gráfica Q-Q residuals los puntos siguen la línea diagonal, por lo que siguen una distribución normal. En la gráfica Scale-Location vemos una nube de puntos más o menos horizontal, por lo que existe homogeneidad de la varianza. Y, por último, en la gráfica Residuals vs Leverage no hay ningún punto que pase las líneas en forma de herradura, por lo que no hay ningún punto influyente que pueda afectar al modelo.

#### 4) Interpreta los coeficientes y escribe tus conclusiones.

De la tabla de p-valores, podemos concluir que para las mujeres:

$$gambling = 6.87 - 0.5489 * income$$

Mientras que para hombres:

$$gambling = 8.72 + 4.8561 * income.$$

Fíjate que el coeficiente -0.5489 no es significativo, mientras que el salto en las pendientes entre mujeres y hombres es de 5.4050 y sí es significativo.

El modelo indica que los ingresos de las mujeres no influyen en la cantidad de dinero apostado, puesto que su pendiente es muy baja y cercana a 0 (-0.5489). En cambio, en los hombres la pendiente sube a +4.8561, lo que muestra que a mayor ingreso, mayor cantidad de apuestas. Además, la diferencia entre estas dos pendientes (5.4050) es significativa, o que confirma la interacción entre ingreso y sexo.

**¿Cuáles de las siguientes conclusiones son correctas? (puede haber varias)**

- a) Los hombres y mujeres adolescentes apuestan de la misma forma.



- b) No evidencia suficiente de que las mujeres apuesten de forma diferente según sus ingresos. Correcta, porque la pendiente para las mujeres no es significativa, por lo que no hay evidencia de que el ingreso afecte a sus apuestas.
- c) Para los hombres: a mayor nivel de ingresos, mayor cantidad apostada. Correcta, porque al haber una interacción significativa entre los ingresos y el sexo, y dado que la pendiente de las mujeres no es significativa, podemos concluir que en hombres sí hay un efecto del ingreso.
- d) La diferencia entre las pendientes de hombres y mujeres no es significativa.

tinytex::install\_tinytex()