

Evaluación curso R

PHUONG LINH, TA

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: **viernes 16 de junio, 23:59**.
- Puedes realizar la entrega en el mail: **constantino.garciama@ceu.es** (**mailto:constantino.garciama@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...

```
data(teengamb, package='faraway')
head(teengamb)
```

```
##   sex status income verbal gamble
## 1   1     51   2.00      8    0.0
## 2   1     28   2.50      8    0.0
## 3   1     37   2.00      6    0.0
## 4   1     28   7.00      4    7.3
## 5   1     65   2.00      8   19.6
## 6   1     61   3.47      6    0.1
```

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:



```
if (!require(ggplot2)) install.packages("ggplot2")
```

```
## Loading required package: ggplot2
```

```
# completa el siguiente código
teengamb['sex'][teengamb['sex'] == 0] <- 'Male'
teengamb['sex'][teengamb['sex'] == 1] <- 'Female'

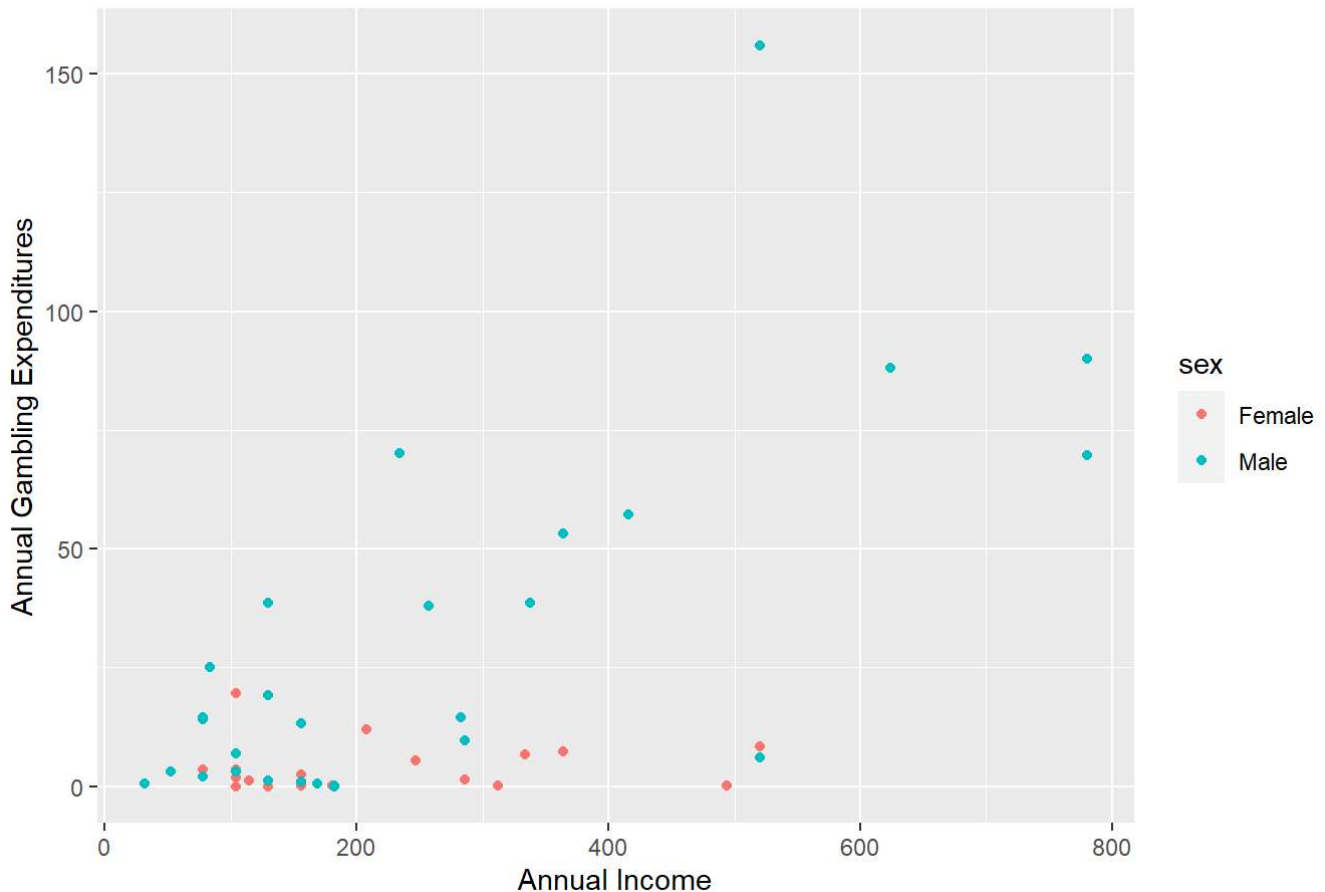
teengamb['income'] <- teengamb['income'] * 52

gamb_male <- subset(teengamb, sex == "Male")
gamb_female <- subset(teengamb, sex == "Female")

p <- ggplot(aes(x = income, y = gamble), data = teengamb)

p + geom_point(aes(group = sex, colour = sex)) +
  xlab("Annual Income") +
  ylab("Annual Gambling Expenditures") +
  ggtitle("British Teen Gambling Survey Study") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))
```

British Teen Gambling Survey Study



```
model <- lm(teengamb$gamble ~ teengamb$income)

model_male <- lm(gamb_male$gamble ~ gamb_male$income)

model_female <- lm(gamb_female$gamble ~ gamb_female$income)

coef_model <- coef(model); coef_model
```

```
##      (Intercept) teengamb$income
##      -6.3245590      0.1061632
```

```
coef_male <- coef(model_male); coef_male
```

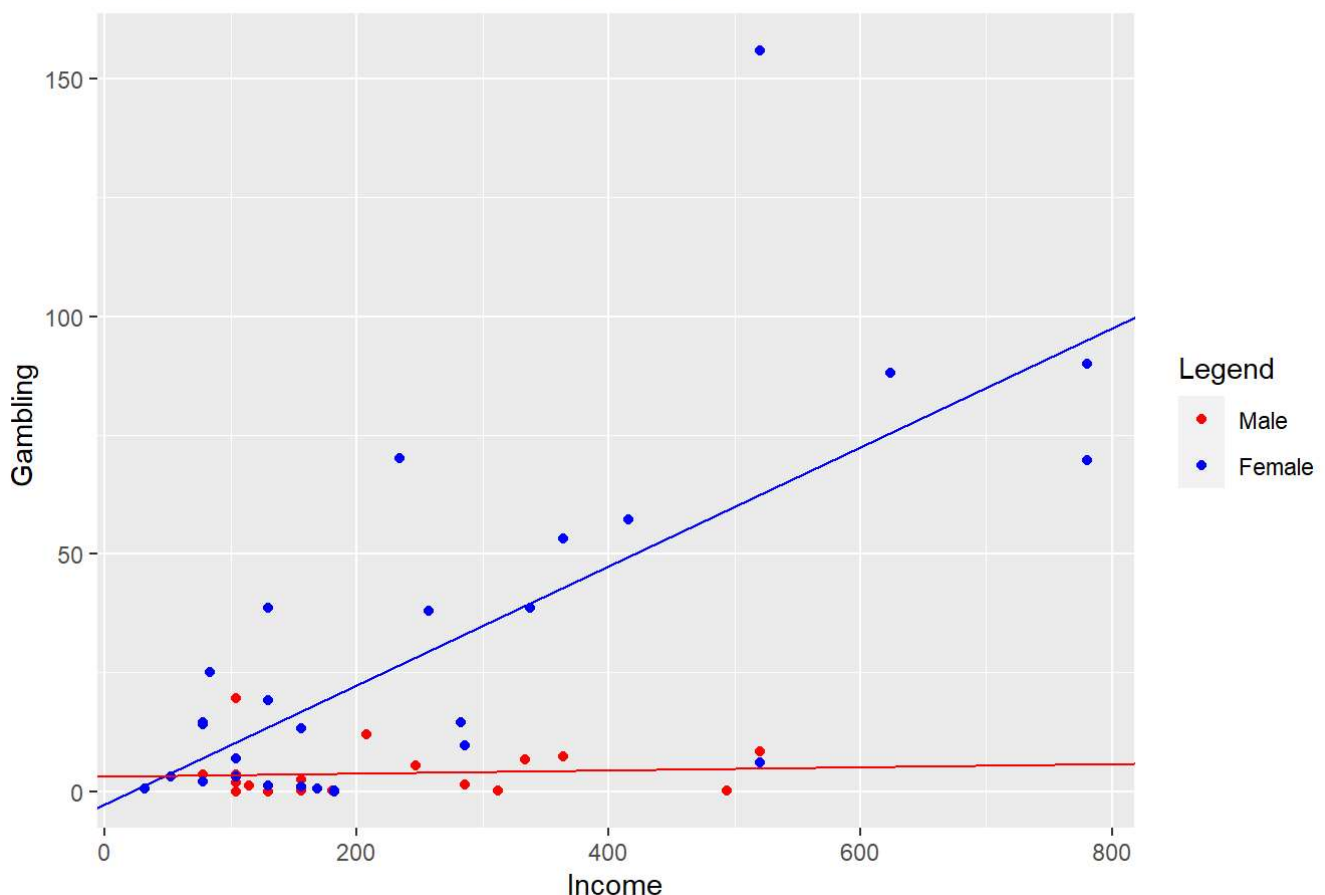
```
##      (Intercept) gamb_male$income
##      -2.6596291      0.1253485
```

```
coef_female <- coef(model_female); coef_female
```

```
##      (Intercept) gamb_female$income
##      3.139973694      0.003363799
```

```
p + geom_point(aes(group = sex, colour = sex)) +
  geom_abline(slope = coef_male[2], intercept = coef_male[1], colour = 'blue') +
  geom_abline(slope = coef_female[2], intercept = coef_female[1], colour = 'red') +
  xlab("Income") +
  ylab("Gambling ") +
  scale_colour_manual(values = c("red","blue"),
                      labels = c("Male", "Female"),
                      name = "Legend") +
  ggtitle("British Teen Gambling Survey Study") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))
```

British Teen Gambling Survey Study



La gráfica anterior apoya que hay interacciones entre las variables `sex` e `income`. Las rectas no son paralelas, es decir, si hay interacciones entre los datos. Los gastos de juego para las mujeres son menores que los gastos de juego de los hombres. Los hombres aumentan su juego en función de sus ingresos semanales, mientras que las mujeres no tienden a jugar ni muestran una tendencia de aumentar el juego con el aumento de los ingresos.

2) Crea el modelo...

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

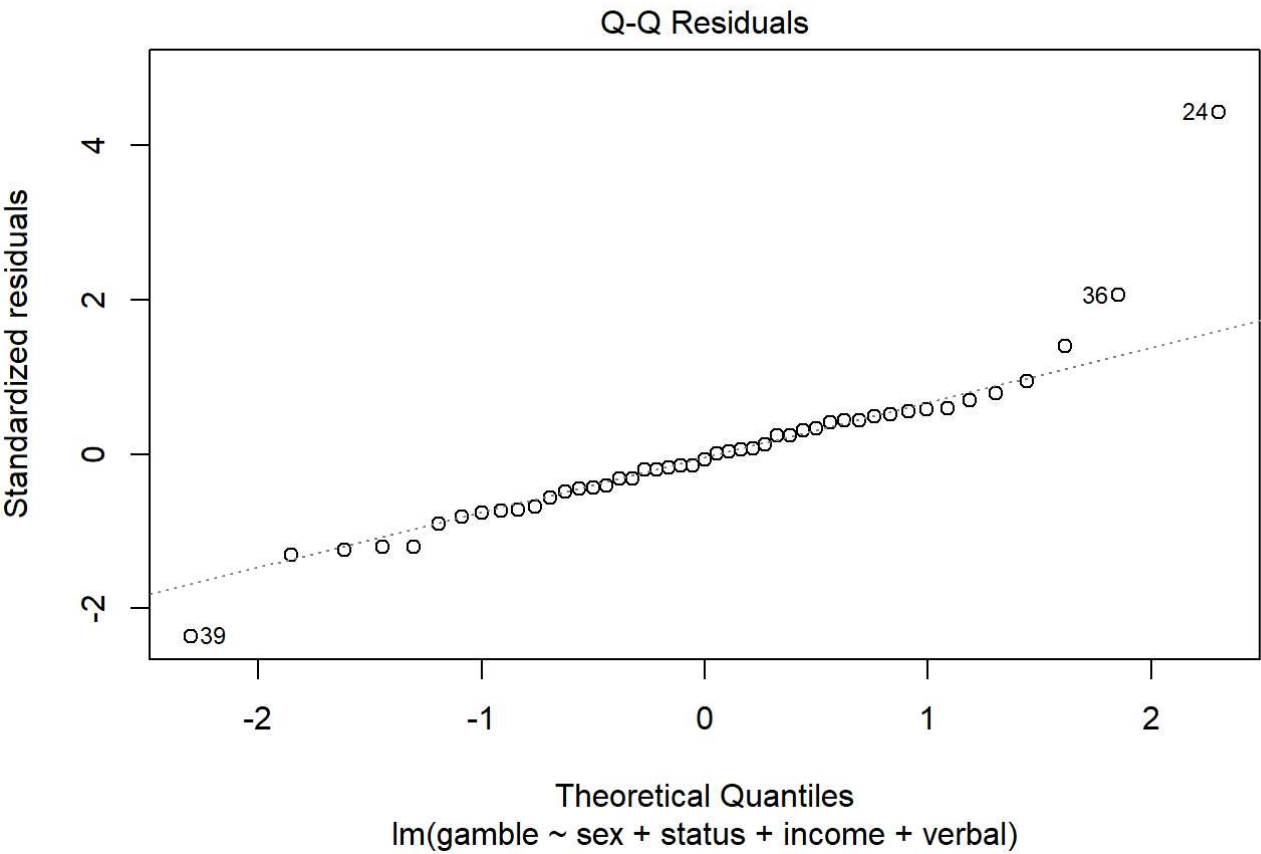
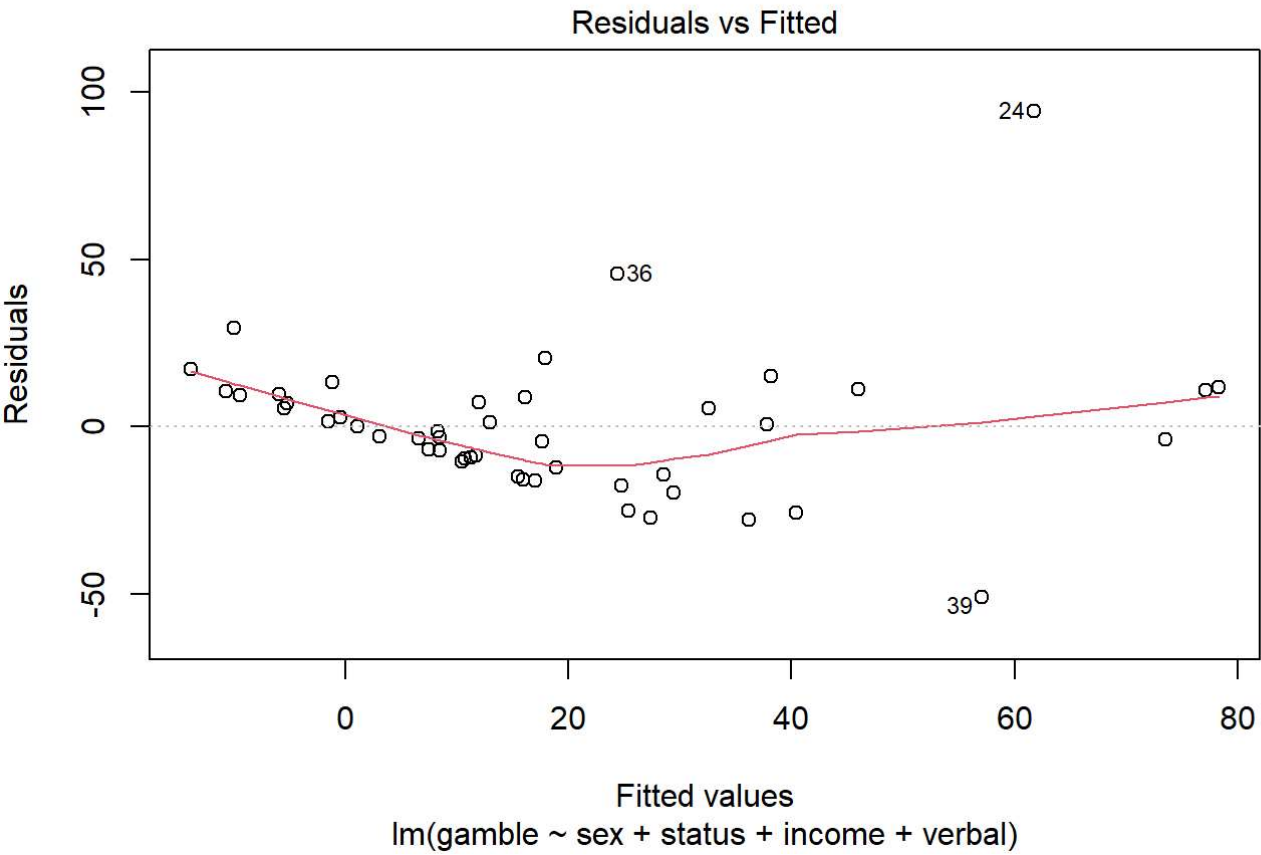
```
# Crea un modelo con interacciones en base a tu conclusión del apartado anterior
gamb_model <- lm(gamble ~ sex+status+income+verbal, data=teengamb)
summary(gamb_model)
```

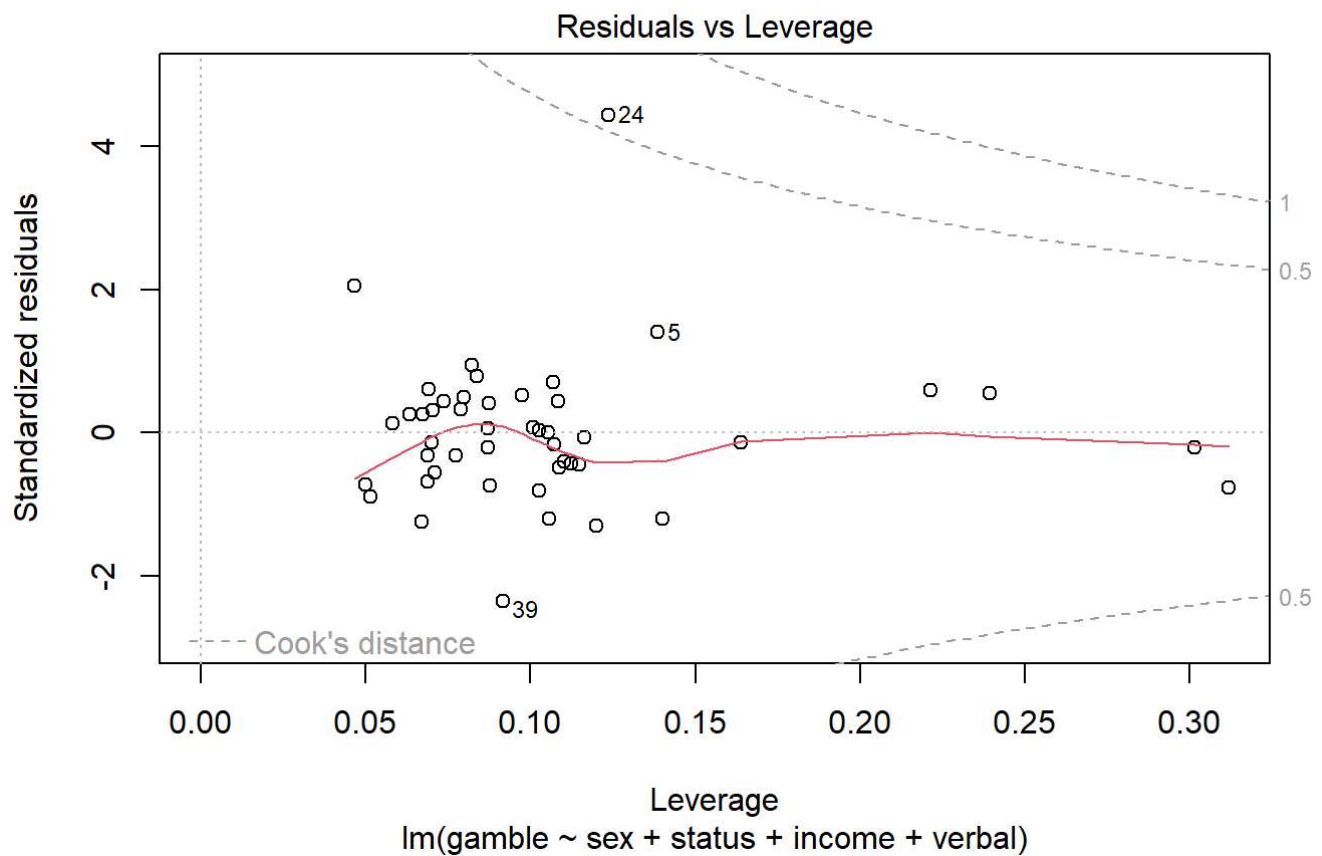
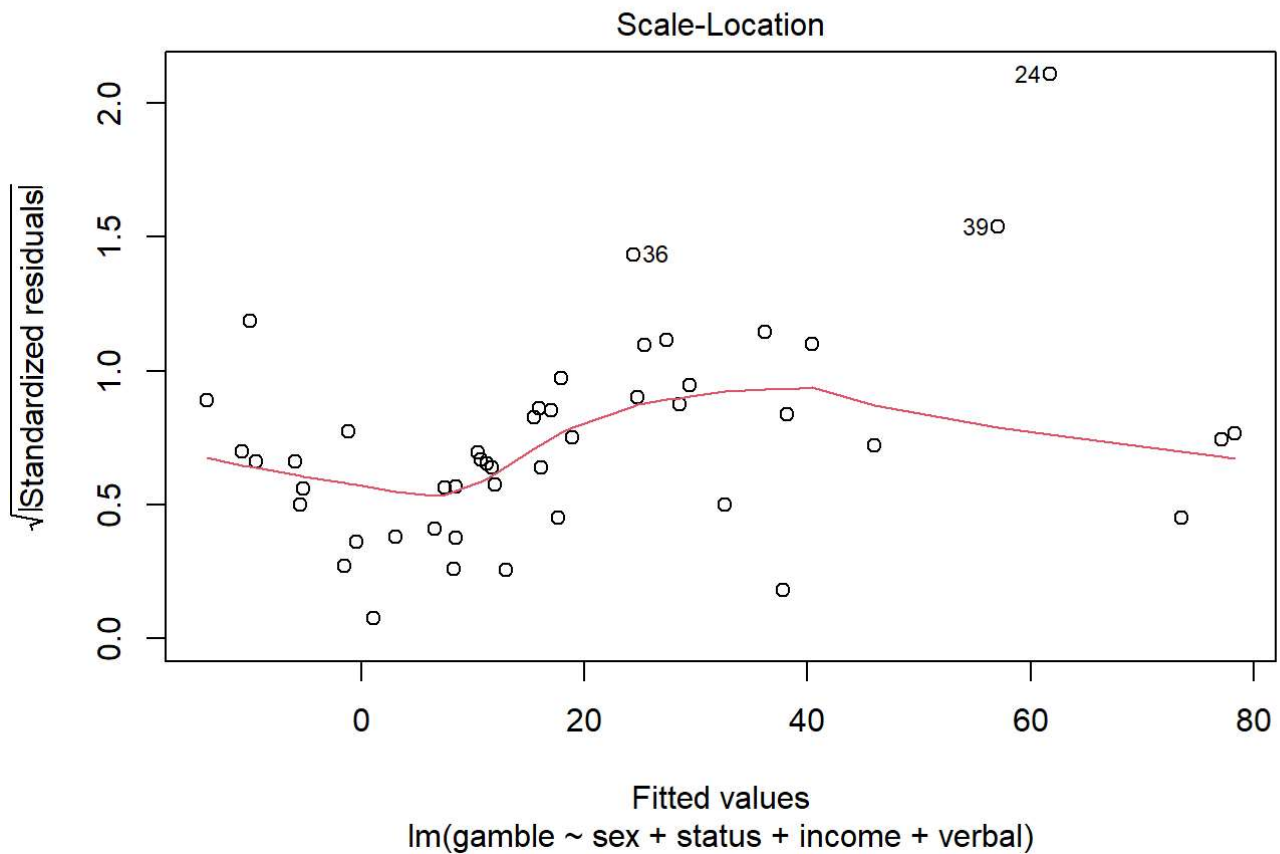
```
##
## Call:
## lm(formula = gamble ~ sex + status + income + verbal, data = teengamb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.082 -11.320  -1.451   9.452  94.252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.43732    14.75429   0.030   0.9765
## sexMale       22.11833     8.21111   2.694   0.0101 *
## status        0.05223     0.28111   0.186   0.8535
## income        0.09542     0.01972   4.839 1.79e-05 ***
## verbal       -2.95949     2.17215  -1.362   0.1803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.69 on 42 degrees of freedom
## Multiple R-squared:  0.5267, Adjusted R-squared:  0.4816
## F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06
```

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

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?: Del gráfico anterior podemos notar que no hay un patrón lineal particular ya que muchos puntos en el centro con una línea hacia abajo están un poco dispersos del modelo ajustado. El modelo resulta ser un buen ajuste.

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.

¿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.
- c. Para los hombres: a mayor nivel de ingresos, mayor cantidad apostada.
- d. La diferencia entre las pendientes de hombres y mujeres no es significativa.

Las afirmaciones son b) No evidencia suficiente de que las mujeres apuesten de forma diferente según sus ingresos y c) Para los hombres: a mayor nivel de ingresos, mayor cantidad apostada.