

PEC6 Otoño 2025. Solución

UOC

En esta actividad no está permitido el uso de herramientas de inteligencia artificial. En el plan docente y en el sitio web sobre integridad académica y plagio de la UOC encontrarán información sobre qué se considera conducta irregular en la evaluación y las consecuencias que puede tener.

La PEC se basará en el archivo “notasfitxer2.csv”. El archivo contiene los resultados de la evaluación de las dos primeras PECs de estudiantes de esta asignatura en un semestre anterior.

Contiene las variables:

1. CPEC1 Puntuación primer cuestionario.
2. CPEC2 Puntuación segundo cuestionario.
3. RPEC1 Puntuación primera prueba de R.
4. RPEC2 Puntuación segunda prueba de R.

Debéis importar los datos y crear un conjunto de datos con el nombre de **notes**. Por ejemplo, con el comando:

```
notes<-read.table(file='notesfitxer2.csv',header=TRUE,sep=';',dec='.')
```

Os puede ser útil consultar el siguiente material:

1. Módulo 10 Regresión lineal simple de las notas de estudio
2. Tema 1 de regresión lineal del módulo 5 de los Manuales de R
3. Actividades Resueltas del Reto 5 (Regresión lineal)

NOMBRE:

Solución

PEC6

Pregunta 1 (100%)

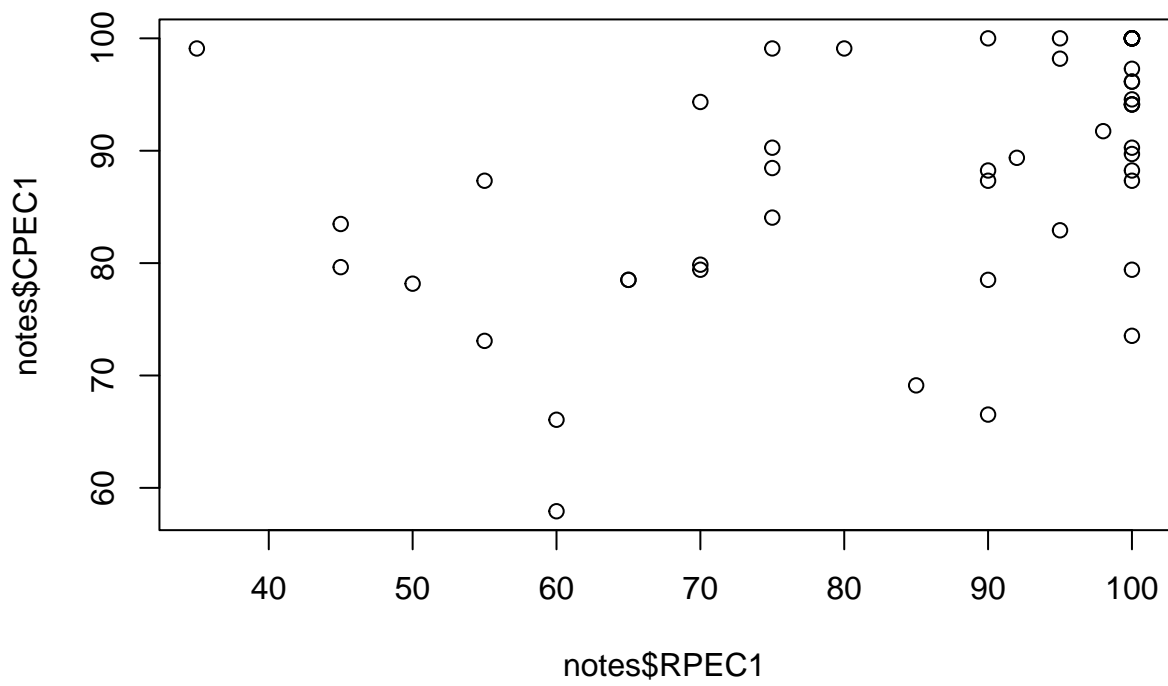
Queremos saber si hay relación lineal entre las notas del cuestionario y de la parte de R de la PEC1. También queremos ver si hay relación entre las notas de la prueba de R de la primera y de la segunda PEC.

a) (10%) En el ámbito de la evaluación académica, se propone utilizar la nota de una prueba anterior (“RPEC1”) para intentar predecir la nota de una prueba posterior (“RPEC2”). Razonad la elección de la variable dependiente (explicada / endógena) y la variable explicativa (independiente / exógena) en este contexto.

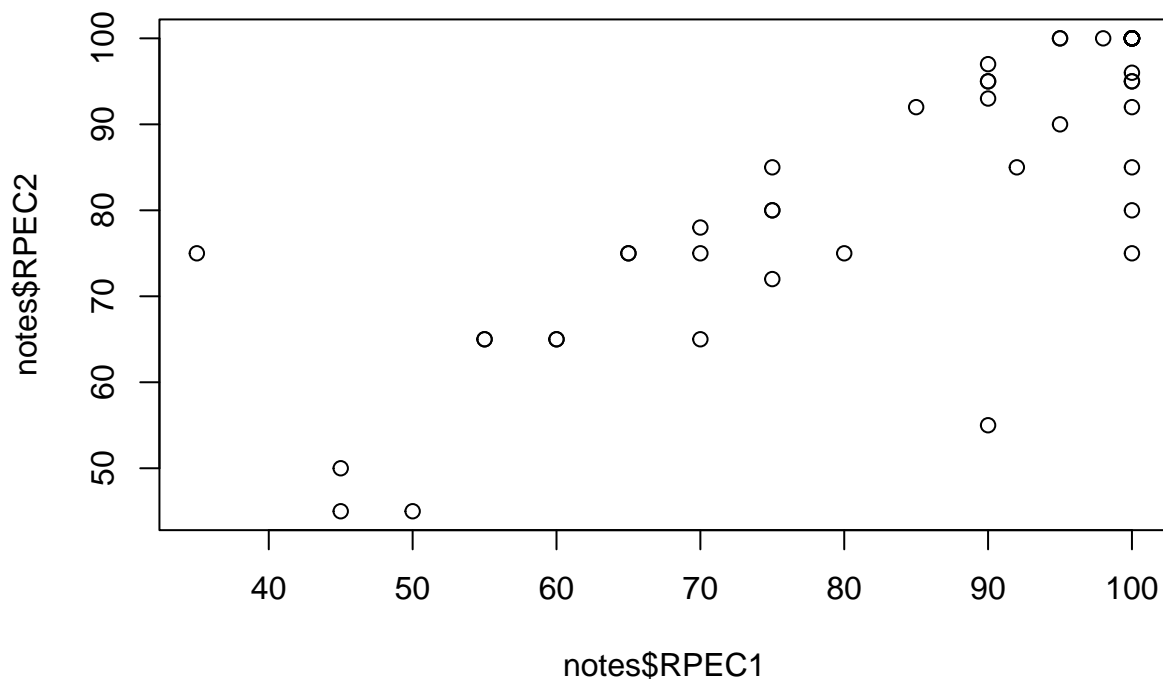
El modelo de regresión lineal simple busca construir un modelo para explicar la relación entre dos variables, principalmente con el propósito de predecir los valores de una variable a partir de los valores de la otra. En el contexto académico, la lógica causal se basa en el orden temporal: la nota obtenida en una prueba anterior “RPEC1” se utiliza como causa o factor explicativo (variable explicativa X) de una nota posterior “RPEC2”. No tendría sentido utilizar un resultado futuro para explicar un resultado ya obtenido.

b) (10%) Haced el diagrama de dispersión de las variables correspondientes a la variable “CPEC1” (en el eje de ordenadas) en función de la variable “RPEC1” (en el eje de abscisas). Luego haced el mismo diagrama de dispersión para la variable “RPEC2” en función de la “RPEC1”. Comentad los gráficos.

```
plot(notes$RPEC1,notes$CPEC1)
```



```
plot(notes$RPEC1,notes$RPEC2)
```



Observamos que los gráficos se parecen bastante, aunque el segundo parece ajustarse más a un modelo lineal.

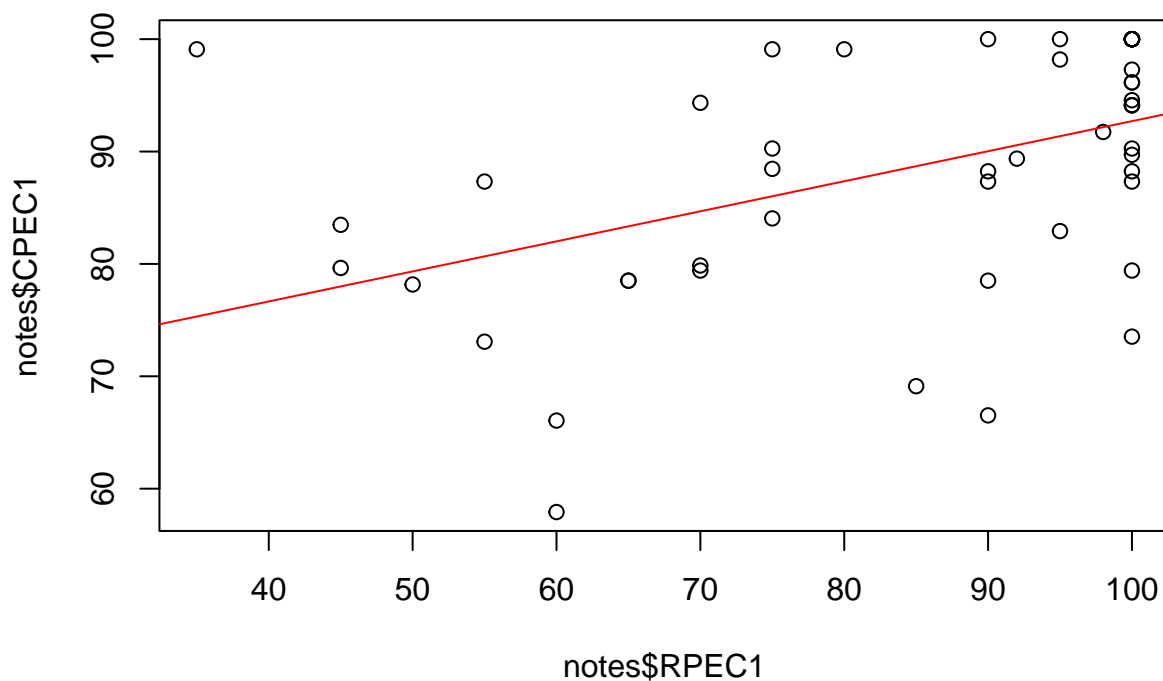
c) (20%) Calculad con R las rectas de regresión correspondientes a los gráficos anteriores. Haced los diagramas de dispersión, añadiendo las rectas de regresión. Dad explícitamente las rectas de regresión.

```
summary(lm(notes$CPEC1~notes$RPEC1))
```

```
##
## Call:
## lm(formula = notes$CPEC1 ~ notes$RPEC1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.088  -4.828   1.416   7.296  23.777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  65.96337    6.36196  10.368 5.97e-14 ***
## notes$RPEC1   0.26741    0.07313   3.656 0.000624 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.666 on 49 degrees of freedom
## Multiple R-squared:  0.2144, Adjusted R-squared:  0.1983
## F-statistic: 13.37 on 1 and 49 DF,  p-value: 0.0006237

plot(notes$RPEC1,notes$CPEC1)
abline(lm(notes$CPEC1~notes$RPEC1),col="red")
```

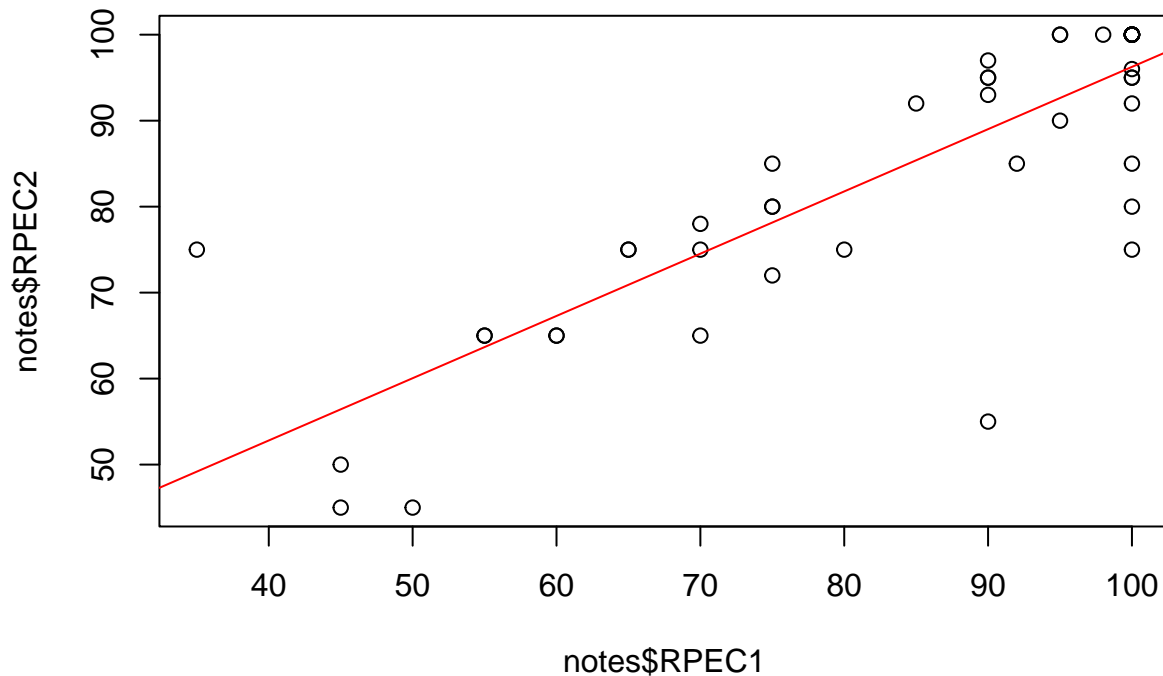


```
summary(lm(notes$RPEC2~notes$RPEC1))

##
## Call:
## lm(formula = notes$RPEC2 ~ notes$RPEC1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.013  -2.461   3.745   3.745  25.817
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.83674     5.91051   4.033 0.000192 ***
```

```
## notes$RPEC1  0.72418    0.06794  10.659 2.32e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.98 on 49 degrees of freedom
## Multiple R-squared:  0.6987, Adjusted R-squared:  0.6925
## F-statistic: 113.6 on 1 and 49 DF,  p-value: 2.322e-14

plot(notes$RPEC1,notes$RPEC2)
abline(lm(notes$RPEC2~notes$RPEC1),col="red")
```



Así, las rectas de regresión son,

$$CPEC1 = 65.963 + 0.267 \cdot RPEC1$$

y

$$RPEC2 = 23.836 + 0.724 \cdot RPEC1$$

d) (20%) ¿Qué variable podemos explicar mejor a partir de la variable “RPEC1”?

Cuando queremos explicar el resultado del primer cuestionario a partir de la primera prueba de R obtenemos que $R^2 = 0.2144$ que nos dice que el modelo explica un 21.44 % de la

variabilidad de esta variable. Cuando queremos explicar la parte de R de la segunda PAC, tenemos $R^2 = 0.6987$ que nos dice que el modelo explica un 69.87 % de la variabilidad. Por lo tanto, podemos explicar mejor la variable “RPEC2”.

e) (10%) Para el mejor de los dos modelos anteriores, interpretad el significado práctico de los parámetros.

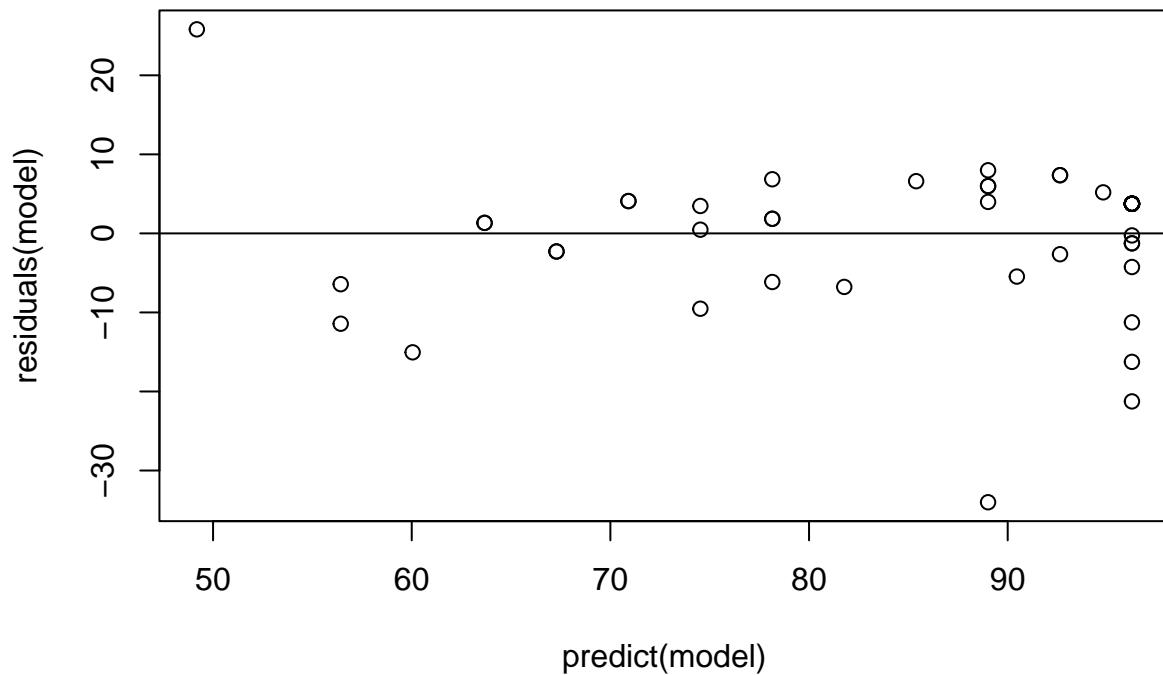
Interpretación de los Parámetros:

Pendiente: La pendiente representa la estimación del incremento que experimenta la variable explicada “RPEC2” cuando la variable explicativa “RPEC1” aumenta en una unidad. Por cada punto adicional en RPEC1, se espera que la nota RPEC2 aumente en 0.724 puntos.

Ordenada al origen: Este parámetro representa la estimación del valor de Y cuando X es igual a cero. En el contexto de las notas de evaluación, la ordenada al origen no siempre tiene una interpretación práctica. En este caso, representaría la nota “RPEC2” estimada para un estudiante que tiene una puntuación de 0 en “RPEC1”. Aunque matemáticamente es posible, su interpretación práctica es limitada o nula.

f) (15%) Haced el plot de los residuos frente a las predicciones para el mejor de los dos modelos anteriores. Comentad el gráfico y discutid si podemos considerar que es un buen modelo.

```
model<-lm(notes$RPEC2~notes$RPEC1)
plot(predict(model), residuals(model))
abline(a=0, b=0)
```



Observamos una nube de puntos que parece bastante aleatoria. No se detecta ninguna relación concreta.

g) (15%) Queremos hacer contrastes de hipótesis con un nivel de significación del 0.05 sobre los coeficientes de los dos modelos que hemos estudiado en el apartado c). ¿Hay algún coeficiente no significativo? Razonad la respuesta.

Todos los coeficientes son significativos, ya que todos los p-valores que nos da R son menores que 0.05.