

PEC4 Otoño 2025 - Solución

UOC

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

Esta PEC se basará en los datos de las ventas de videojuegos en todo el mundo para títulos lanzados entre 1971 y 2024 (los datos se han obtenido de la web *Maven Analytics* que proporciona un “Data Playground”). Hay incluida información como la puntuación de la crítica, el género, la consola y más.

La base de datos contiene:

1. *title* = variable cualitativa que indica el título del juego
2. *console* = variable cualitativa que indica la consola para la cual se lanzó el juego
3. *genre* = variable cualitativa que indica el género del juego
4. *critic_score* = puntuación de Metacritic (sobre 10)
5. *total_sales* = ventas globales de copias en millones

Para importar los datos podéis usar las siguientes instrucciones:

```
datos <- read.csv("DATA.csv", header = TRUE, encoding = "UTF-8",
                   stringsAsFactors = TRUE)
datos <- datos[, -1]
```

Os puede ser útil consultar el siguiente material:

1. Módulo Intervalos de confianza.
2. Actividades Resueltas del Reto 3 (Intervalos de confianza).
3. Procurad usar las funciones propias de R para hacer los cálculos a no ser que se indique lo contrario.

Hay que entregar la práctica en formato “.pdf”. **Solo** se aceptaran PECs con este formato.

NOMBRE:

PAC4

Una vez importados los datos, con la misma base de datos y suponiendo que los datos corresponden a una muestra, responded a las siguientes cuestiones, indicando **tanto las instrucciones de R como los resultados numéricos obtenidos y los correspondientes comentarios**.

Pregunta-1 (50%)

Queremos estudiar ahora las ventas globales de videojuegos y relacionarlas con alguna consola.

- a) (10%) Encontrad un intervalo de confianza para las ventas globales de los videojuegos con un nivel de confianza del 95%.
- b) (20%) Encontrad ahora un intervalo de confianza para las ventas globales de los videojuegos del género *Sports* también con un nivel de confianza del 95%.
- c) (20%) En base a los intervalos obtenidos, ¿qué conclusiones podemos extraer sobre las ventas globales de los videojuegos del género *Sports* en relación al total de ventas de videojuegos?

Solución

- a) (10%) Calculamos

```
options(scipen=999) # Desactiva la notació científica
attach(datos)
tvendes<-t.test(total_sales,conf.level = 0.95)
tvendes

## 
##   One Sample t-test
##
## data: total_sales
## t = 33.621, df = 4125, p-value < 0.000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.6942398 0.7802197
## sample estimates:
## mean of x
## 0.7372298
```

y el intervalo de confianza pedido será: (0.6942, 0.7802).

- b) (20%) Calculamos

```

tvendesSports<-t.test(total_sales[datos$genre=="Sports"] ,
conf.level = 0.95)
tvendesSports

##
## One Sample t-test
##
## data: total_sales[datos$genre == "Sports"]
## t = 15.725, df = 537, p-value < 0.0000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.7890866 1.0143706
## sample estimates:
## mean of x
## 0.9017286

```

y el intervalo de confianza pedido será: (0.7891, 1.0144).

- c) (20%) Claramente las ventas globales para *Sports* son más elevadas que las ventas globales. Concretamente el extremo inferior del intervalo para *Sports* es un 14% superior al de las ventas globales totales y el extremo superior es un 30% superior al de las ventas globales totales. Además los intervalos son disjuntos por poco pero aun así denota ventas más altas del género *Sports*.

Pregunta-2 (50%)

Queremos estudiar ahora las proporciones de juegos correspondientes a diferentes consolas.

- a) (20%) Calculad un intervalo de confianza del 85% para la proporción de videojuegos de la consola (*PS2*) mediante la función *prop.test* con la opción *correct=FALSE*.
- b) (10%) Calculad un intervalo de confianza del 85% para la proporción de videojuegos de la consola (*PS3*) mediante la función *prop.test* con la opción *correct=FALSE*.
- c) (10%) En base en los apartados anteriores, ¿podemos decir que las proporciones son diferentes? Razonad vuestra respuesta. Comparad también los intervalos obtenidos.
- d) (10%) ¿Qué nivel de confianza haría que el intervalo de confianza de la proporción de *PS2* tuviera una longitud que fuese la mitad del actual? Este apartado lo tenéis que hacer usando las fórmulas de las notas de estudio, concretamente el “Módulo: Intervalos de confianza” del Reto 3.

Solución

- a) (20%) Para usar la instrucción *prop.test*, primero tenemos que calcular la medida de la muestra y el número de observaciones de la muestra correspondientes a videojuegos de *PS2*:

```

len<-nrow(datos)
len

## [1] 4126

len2<-sum(console == "PS2")
len2

## [1] 299

prop2<-prop.test(len2, len, alternative='two.sided', p=0.5,
conf.level=0.85,correct=FALSE)
prop2

##
##   1-sample proportions test without continuity correction
##
## data: len2 out of len, null probability 0.5
## X-squared = 3016.7, df = 1, p-value < 0.0000000000000022
## alternative hypothesis: true p is not equal to 0.5
## 85 percent confidence interval:
##  0.06686918 0.07849462
## sample estimates:
##           p
## 0.07246728

```

El intervalo de confianza pedido será: (0.0669, 0.0785).

- b) (10%) Como el apartado anterior pero con la *PS3*

```

len<-nrow(datos)
len

## [1] 4126

len3<-sum(console=="PS3")
len3

## [1] 471

prop3<-prop.test(len3, len, alternative='two.sided', p=0.5,
conf.level=0.85,correct=FALSE)
prop3

##
##   1-sample proportions test without continuity correction
##
## data: len3 out of len, null probability 0.5
## X-squared = 2457.1, df = 1, p-value < 0.0000000000000022
## alternative hypothesis: true p is not equal to 0.5

```

```

## 85 percent confidence interval:
## 0.1072204 0.1214753
## sample estimates:
## p
## 0.1141541

```

El intervalo de confianza pedido será: (0.1072, 0.1215).

- c) (10%) En este caso los intervalos son disjuntos y por tanto podemos concluir que las proporciones de estos dos tipos de videojuegos son diferentes, y más alta la proporción de $PS3$.

- d) (10%) Queremos que el margen de error sea $\frac{0.0784946 - 0.0668692}{4} = 0.0029064$. Como $z_{\alpha/2} = \frac{ME}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ y

```

ME<- (prop2$conf.int[2] - prop2$conf.int[1])/4
phS<-len2/len
phS

```

```

## [1] 0.07246728
zalpha2S<-ME/sqrt((phS*(1-phS))/len)
zalpha2S

```

```

## [1] 0.720076
alphaS<-1-2*(pnorm(zalpha2S,lower.tail=FALSE))
alphaS

```

```

## [1] 0.5285218

```

tenemos que $z_{\alpha/2} = \frac{0.0029064}{\sqrt{\frac{0.0724673(1-0.0724673)}{4126}}} = 0.720076$ con lo cual el nivel de confianza tiene que ser $1 - \alpha = 52.852182\%$. Lo podemos comprobar volviendo a hacer el intervalo de confianza

```

propS2<-prop.test(len2, len, alternative='two.sided', p=0.5,
  conf.level=alphaS, correct=FALSE)
propS2

```

```

##
## 1-sample proportions test without continuity correction
##
## data: len2 out of len, null probability 0.5
## X-squared = 3016.7, df = 1, p-value < 0.0000000000000022
## alternative hypothesis: true p is not equal to 0.5
## 52.85218 percent confidence interval:
## 0.06961433 0.07542767
## sample estimates:
##
```

```
##          p  
## 0.07246728
```

donde vemos que el margen de error ahora es 0.0029 tal como queríamos.