

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 aceptarán PECs con este formato.

NOMBRE: José Carlos López Henestrosa

PEC4

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%.

```
# 1) Calcular intervalo de confianza sobre la columna de ventas
resultado <- t.test(datos$total_sales, conf.level = 0.95)

# 2) Mostrar intervalo
resultado$conf.int

## [1] 0.6942398 0.7802197
## attr(,"conf.level")
## [1] 0.95
```

- 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%.

```
# 1) Obtener juegos del género "Sports"
datos_sports <- subset(datos, genre == "Sports")

# 2) Calcular el intervalo de confianza sobre la columna de ventas
resultado_sports <- t.test(datos_sports$total_sales, conf.level = 0.95)

# 3) Mostrar intervalo
resultado_sports$conf.int

## [1] 0.7890866 1.0143706
## attr(,"conf.level")
## [1] 0.95
```

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?

La conclusión principal a nivel estadístico es que los intervalos no se solapan. El límite superior del intervalo de ventas globales (0.780) es inferior al límite inferior del intervalo de ventas del género *Sports* (0.789). Esto nos indica, con un nivel de confianza del 95%, que los videojuegos del género *Sports* tienen unas ventas medias **superiores** a la media general del mercado.

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*.

```
# 1) Calcular el número de éxitos (juegos de PS2) y el total de observaciones
x_ps2 <- sum(datos$console == "PS2")
n_total <- nrow(datos)

# 2) Calcular el intervalo de confianza
# correct = FALSE => no se aplica la corrección de continuidad de Yates
resultado_ps2 <- prop.test(x_ps2, n_total, conf.level = 0.85, correct = FALSE)

# 3) Mostrar intervalo
resultado_ps2$conf.int

## [1] 0.06686918 0.07849462
## attr(,"conf.level")
## [1] 0.85
```

- 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*.

```
# 1) Contar el número de éxitos (juegos de PS3)
x_ps3 <- sum(datos$console == "PS3")
n_total <- nrow(datos)

# 2) Calcular el intervalo de confianza
resultado_ps3 <- prop.test(x_ps3, n_total, conf.level = 0.85, correct = FALSE)

# 3) Mostrar intervalo
resultado_ps3$conf.int

## [1] 0.1072204 0.1214753
## attr(,"conf.level")
## [1] 0.85
```

c) (10%) En base a los apartados anteriores, ¿podemos decir que las proporciones son diferentes? Razonad vuestra respuesta. Comparad también los intervalos obtenidos.

Vamos a comparar directamente los intervalos de confianza del 85% que hemos calculado en los dos apartados anteriores:

1. Intervalo de confianza para PS2: (0.0669, 0.0785)
2. Intervalo de confianza para PS3: (0.1072, 0.1215)

La regla general para comparar dos parámetros mediante sus intervalos de confianza es observar si existe solapamiento entre ellos. Como podemos ver, $0.0785 < 0.1072$, por lo que existe una brecha clara entre ambos intervalos. No se solapan en absoluto.

Dado que los intervalos son disjuntos (no comparten ningún valor), podemos concluir con un nivel de confianza del 85% (y de hecho, con uno mucho mayor también) que la proporción de videojuegos de PS3 es significativamente mayor que la de PS2 en esta base de datos.

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.

```
# 1) Calcular el Z actual correspondiente al 85%
alpha_old <- 1 - 0.85
z_old <- qnorm(1 - alpha_old/2)

# 2) Calcular el nuevo Z (la mitad del anterior)
z_new <- z_old / 2

# 3) Calcular el nuevo nivel de confianza asociado a ese Z
# La función pnorm nos da la probabilidad acumulada
conf_level_new <- (2 * pnorm(z_new)) - 1

# 4) Mostrar resultados
cat("Z actual:", z_old)

## Z actual: 1.439531
cat("Z nuevo:", z_new)

## Z nuevo: 0.7197657
cat("Nuevo nivel de confianza:", conf_level_new)

## Nuevo nivel de confianza: 0.5283308
```