

# PEC5 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 del Reto 4:

1. Módulos: Contraste de hipótesis y Contraste de dos muestras.
2. Actividades Resueltas del Reto 4 (Contraste de hipótesis y Contraste de dos muestras).
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”.

# NOMBRE:

## PEC5

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

### Pregunta-1 (40%)

- a) (20%) Queremos saber si podemos considerar que el valor medio de la variable “total\_sales” en el conjunto total de videojuegos es igual a 0.75 millones de copias o es inferior. Haced el contraste de hipótesis adecuado con un nivel de significación del 3%, usando únicamente las instrucciones de R que hacen directamente este contraste de hipótesis.

#### Solución:

```
ttest1<-t.test(datos$total_sales,
mu = 0.75,
alternative = "less",
conf.level = 0.97)
ttest1

##
##   One Sample t-test
##
## data:  datos$total_sales
## t = -0.58238, df = 4125, p-value = 0.2802
## alternative hypothesis: true mean is less than 0.75
## 97 percent confidence interval:
##      -Inf 0.7784824
## sample estimates:
## mean of x
## 0.7372298

pv_1<-ttest1$p.value
```

Dado que el p-valor (0.2801707) es superior al nivel de significación (0.03), aceptamos la hipótesis nula y concluimos que la media de la variable “total\_sales” podría ser 0.75.

Esta conclusión es coherente con el intervalo de confianza obtenido,  $(-\infty, 0.7784824)$ , puesto que el valor 0.75 sí se encuentra dentro de este intervalo.

b) (20%) Ahora haced el mismo contraste de hipótesis mediante las fórmulas de los módulos, indicando:

- 1) Las hipótesis nula y alternativa
- 2) El estadístico de contraste.
- 3) El p-valor
- 4) Vuestra decisión y conclusión aplicadas a este caso considerado.

### Solución:

1) Hipótesis nula:  $H_0 : \mu = \mu_0 = 0.75$ , contra la hipótesis alternativa:  $H_1 : \mu < \mu_0 = 0.75$

2) Estadístico de contraste:  $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ , donde:

```
num_observaciones <- nrow(datos)
media <- mean(datos$total_sales, na.rm = TRUE)
desviacion_tipica <- sd(datos$total_sales, na.rm = TRUE)
```

Así pues,  $n = 4126$ ,  $\bar{X} = 0.7372298$  y  $S = 1.4084966$ . Y el estadístico de contraste vale:

```
t <- (media - 0.75) / (desviacion_tipica / sqrt(num_observaciones))
```

$$t = \frac{0.7372298 - 0.75}{1.4084966 / \sqrt{4126}} = -0.5823818$$

3) Dado que la hipótesis alternativa es unilateral, el p-valor será:

```
pvalor <- pt(t, num_observaciones - 1)
```

$$p - \text{valor} = 0.2801707$$

4) Dado que el p-valor ( $p - \text{valor} = 0.2801707$ ) es superior al nivel de significación ( $\alpha = 0.03$ ), no rechazamos la hipótesis nula y concluimos que no podemos decir que el valor medio de las ventas es inferior a 0.75. Observamos que todos los valores obtenidos son iguales a los que ha dado la salida de R en el apartado anterior.

## Pregunta-2 (20%)

Contrastad con un nivel de significación del 5% si la media de las puntuaciones del “critic\_score” es la misma para la “PS4” que para la “XOne”. Indicad qué tipo de contraste tenemos que hacer y las hipótesis nula y alternativa. Usad la instrucción de R que hace este tipo de contraste y, a partir de la salida de R, indicad el p-valor y la conclusión a que llegáis. Supondremos que las varianzas poblacionales son iguales.

### Solución

Hay que hacer un contraste de diferencia de medias para distribuciones normales con varianzas poblacionales desconocidas pero iguales. La hipótesis nula es:  $H_0 : \bar{c}s_{PS4} = \bar{c}s_{XOne}$  y la hipótesis alternativa es:  $\bar{c}s_{PS4} \neq \bar{c}s_{XOne}$ .

```
puntuaciones_PS4 <- datos$critic_score[datos$console == "PS4"]
puntuaciones_XOne <- datos$critic_score[datos$console == "XOne"]
t2<-t.test(puntuaciones_PS4,
puntuaciones_XOne,
alternative = "two.sided",
mu = 0,
var.equal = TRUE,
conf.level = 0.95)
t2p<-t2$p.value
t2

##
## Two Sample t-test
##
## data: puntuaciones_PS4 and puntuaciones_XOne
## t = -1.2131, df = 162, p-value = 0.2269
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.6692741 0.1598991
## sample estimates:
## mean of x mean of y
## 7.650000 7.904687

t2p

## [1] 0.2268571
```

Dado que el  $p - valor = 0.2268571$  es más grande que el nivel de significación  $\alpha = 0.05$ , aceptamos la hipótesis nula y concluimos que las puntuaciones “critic\_score” son iguales para la “PS4” y para la “XOne”.

### Pregunta-3 (20%)

Contrastad con un nivel de significación del 5% si la proporción de videojuegos del género “Fighting” es igual al 7%. Indicad qué tipo de contraste tenemos que hacer y las hipótesis nula y alternativa. Usad la instrucción de R que hace este tipo de contraste y, a partir de la salida de R, indicad el p-valor y la conclusión a la que llegáis.

#### Solución:

En esta pregunta se nos pide hacer un contraste de hipótesis sobre la proporción de videojuegos del género “Fighting”. La hipótesis nula es:  $H_0 : p = 0.07$  y la hipótesis alternativa es:  $H_1 : p \neq p_0 = 0.07$ .

```
# Contamos el número total de videojuegos
n_total <- nrow(datos)

# Contamos el número de videojuegos del género "Fighting"
n_fighting <- sum(datos$genre == "Fighting", na.rm = TRUE)
```

Así pues, el número total de filas es de 4126 y el número de videojuegos del género “Fighting” es de 219. Ahora ya podemos hacer con R el contraste de hipótesis:

```
# Código de R para el test sobre la proporción, para contrastar
# si la proporción de
# videojuegos de género "Fighting" es = 0.07
ptest_1<-prop.test(n_fighting,
  n_total,
  p = 0.07,
  alternative = "two.sided",
  conf.level = 0.95)
ptest_1
```

```
##
## 1-sample proportions test with continuity correction
##
## data:  n_fighting out of n_total, null probability 0.07
## X-squared = 17.89, df = 1, p-value = 2.341e-05
## alternative hypothesis: true p is not equal to 0.07
## 95 percent confidence interval:
##  0.04652993 0.06047227
## sample estimates:
##           p
## 0.05307804
```

```
pvalor_1<-ptest_1$p.value
pvalor_1
```

```
## [1] 2.340651e-05
```

Dado que el  $p - valor = 2.3406514 \times 10^{-5}$  es más pequeño que el nivel de significación  $\alpha = 0.05$ , rechazamos la hipótesis nula y concluimos que la proporción de videojuegos del género “Fighting” no es del 7%.

## Pregunta-4 (20%)

Queremos saber, con un nivel de significación del 5%, si la proporción de videojuegos del género “Role-Playing” disponibles para la consola “PS2” es superior a la proporción del mismo género para la consola “PS3”. Indicad qué tipo de contraste tenemos que hacer y las hipótesis nula y alternativa. Usad la instrucción de R que hace este tipo de contraste y a partir de la salida de R, indicad el p-valor y la conclusión a la que llegáis.

### Solución:

Tenemos que hacer un contraste de hipótesis sobre la diferencia de proporciones. La hipótesis nula es:  $H_0 : p_{PS2} = p_{PS3}$  y la hipótesis alternativa es:  $H_1 : p_{PS2} > p_{PS3}$ . Lo primero que tenemos que hacer es calcular el número de videojuegos de la consola “PS2” y de la consola “PS3”, así como el número de videojuegos del género “Role-Playing” para cada una de estas dos consolas.

```
# Contamos el número de videojuegos para cada consola
n_PS2 <- sum(datos$console == "PS2", na.rm = TRUE)
n_PS3 <- sum(datos$console == "PS3", na.rm = TRUE)

# Contamos el número de videojuegos del género "Role-playing" para cada consola
role_PS2 <- sum(datos$console == "PS2" & datos$genre == "Role-Playing",
na.rm = TRUE)
role_PS3 <- sum(datos$console == "PS3" & datos$genre == "Role-Playing",
na.rm = TRUE)
n_PS2
```

```
## [1] 299
```

```
n_PS3
```

```
## [1] 471
```

```
role_PS2
```

```
## [1] 61
```

```
role_PS3
```

```
## [1] 41
```

Así pues, el número de videojuegos para la consola “PS2” es de 299 y para la consola “PS3” es de 471. El número de videojuegos de género “Role-Playing” para la consola “PS2” es de 61 y para la consola “PS3” es de 41. Finalmente, pedimos a R hacer el contraste para la diferencia de proporciones:

```
# Test de proporciones
```

```
p3<-prop.test(c(role_PS2,role_PS3),  
  c(n_PS2, n_PS3),  
  alternative ="greater")  
p3
```

```
##  
## 2-sample test for equality of proportions with continuity correction  
##  
## data: c(role_PS2, role_PS3) out of c(n_PS2, n_PS3)  
## X-squared = 20.767, df = 1, p-value = 2.594e-06  
## alternative hypothesis: greater  
## 95 percent confidence interval:  
## 0.07034532 1.00000000  
## sample estimates:  
## prop 1 prop 2  
## 0.20401338 0.08704883
```

```
p3pv<-p3$p.value  
p3pv
```

```
## [1] 2.59356e-06
```

Dado que el  $p - valor = 2.5935604 \times 10^{-6}$  es más pequeño que el nivel de significación  $\alpha = 0.05$  tenemos evidencias para rechazar la hipótesis nula y concluimos que la proporción de videojuegos de género “Role-Playing” para la consola “PS2” es superior a esta proporción en la consola “PS3”. Observamos que el intervalo de confianza para la diferencia de proporciones,  $(0.0703453, 1)$  no contiene el cero, por lo tanto, está totalmente de acuerdo con la conclusión a la que hemos llegado.