

# PEC5 Otoño 2025

## 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-1.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: José Carlos López Henestrosa

## PEC5

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 (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.

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

##
## 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
```

El p-value resultante es 0,2802, el cual es mayor que el nivel de significación (0,03), por lo que no rechazamos la hipótesis nula y concluimos que la media es significativamente superior a 0,75.

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.

```
# 1) Calcular estadísticos descriptivos básicos
variable <- datos$total_sales
```

```

n <- length(variable)           # Tamaño de la muestra
media_muestral <- mean(variable) #  $\bar{x}$  barra
desv_std <- sd(variable)         # s (cuasidesviación típica)
mu0 <- 0.75                     # Valor de referencia  $H_0$ 

alpha <- 0.03                   # Nivel de significación (3%)

# 2) Calcular estadístico de contraste ( $t_{obs}$ )
error_estandar <- desv_std / sqrt(n)
t_obs <- (media_muestral - mu0) / error_estandar

# 3) Calcular P-valor
## Como  $H_1$  es "menor que" (<), buscamos el área a
## la izquierda del estadístico (cola inferior)
p_valor <- pt(t_obs, df = n - 1)

# 4) Imprimir resultados
cat("1. Hipótesis:  $H_0: \mu = 0.75$  vs  $H_1: \mu < 0.75$ \n")

## 1. Hipótesis:  $H_0: \mu = 0.75$  vs  $H_1: \mu < 0.75$ 
cat("2. Estadístico de contraste (t):", t_obs, "\n")

## 2. Estadístico de contraste (t): -0.5823818
cat("3. P-valor:", p_valor, "\n")

## 3. P-valor: 0.2801707
cat("    Nivel de significación (alpha):", alpha, "\n")

##    Nivel de significación (alpha): 0.03

# 5) Imprimir decisión
if (p_valor < alpha) {
  cat("4. Decisión: RECHAZAMOS la hipótesis nula.\n")
} else {
  cat("4. Decisión: NO rechazamos la hipótesis nula.\n")
}

## 4. Decisión: NO rechazamos la hipótesis nula.

```

Con un nivel de significación del 3%, no tenemos evidencia estadística suficiente para afirmar que el valor medio de las ventas totales sea inferior a 0,75 millones. Aunque la media muestral (0,737) es matemáticamente menor que 0,75, la diferencia no es significativa y podría deberse al azar. Por tanto, asumimos que la media se mantiene en torno a 0,75.

## 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.

Para resolver este caso, debemos realizar un contraste de comparación de medias para dos muestras independientes, ya que la variable es cuantitativa.

```
# 1) Filtrar las puntuaciones por consola
scores_ps4 <- subset(datos, console == "PS4")$critic_score
scores_xone <- subset(datos, console == "XOne")$critic_score

# 2) Realizar el t-test para dos muestras independientes con varianzas iguales
#   var.equal = TRUE: indicamos la suposición del enunciado
#   conf.level = 0.95: corresponde al nivel de significación del 5%
t.test(scores_ps4, scores_xone, var.equal = TRUE, conf.level = 0.95)

##
##  Two Sample t-test
##
## data:  scores_ps4 and scores_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
```

Como el p-valor (0,280) es mayor que 0,05 (nivel de significación), no rechazamos la hipótesis nula. Esto significa que, con un nivel de confianza del 95%, no hay evidencia estadística suficiente para afirmar que existe una diferencia significativa entre la puntuación media de la crítica para los juegos de PS4 y los de Xbox One. Consideramos que las medias son estadísticamente iguales.

### 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.

Debemos realizar un test de una proporción, ya que la variable a analizar es cualitativa, no cuantitativa.

```
# 1) Calcular el número de éxitos (juegos del género Fighting) y el total
n <- nrow(datos)
x <- sum(datos$genre == "Fighting")

# 2) Realizar el test de proporción
#   x: número de éxitos
#   n: número total de observaciones
#   p: proporción a contrastar (0.07)
#   conf.level: 0.95 (para un nivel de significación del 5%)
prop.test(x, n, p = 0.07, alternative = "two.sided", conf.level = 0.95)
```

```
##
## 1-sample proportions test with continuity correction
##
## data: x out of n, 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
```

El p-valor es, aproximadamente, 0,00002. Como el p-valor es menor que el nivel de significación (0,05), rechazamos la hipótesis nula. Con un nivel de confianza del 95%, tenemos evidencia estadística suficiente para afirmar que la proporción de videojuegos del género “Fighting” NO es igual al 7%. Los datos sugieren que la proporción real es significativamente inferior (alrededor del 5,3%).

## 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.

Debemos realizar un test de una proporción, ya que la variable a analizar es cualitativa, no cuantitativa.

```
# 1) Obtener conteos para PS2
datos_ps2 <- subset(datos, console == "PS2")
n_ps2 <- nrow(datos_ps2) # Total juegos PS2
x_ps2 <- sum(datos_ps2$genre == "Role-Playing") # Juegos RPG en PS2

# 2) Obtener conteos para PS3
datos_ps3 <- subset(datos, console == "PS3")
n_ps3 <- nrow(datos_ps3) # Total juegos PS3
x_ps3 <- sum(datos_ps3$genre == "Role-Playing") # Juegos RPG en PS3

# 3) Realizar el test de comparación de proporciones
# x: vector con los éxitos c(x_ps2, x_ps3)
# n: vector con los totales c(n_ps2, n_ps3)
# alternative: "greater" porque H1 es p_ps2 > p_ps3
prop.test(
  x = c(x_ps2, x_ps3),
  n = c(n_ps2, n_ps3),
  alternative = "greater",
  conf.level = 0.95
)
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: c(x_ps2, x_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
```

El p-valor es, aproximadamente, 0,0000015. Como el p-valor es mucho menor que el nivel de significación (0,05), rechazamos la hipótesis nula. Con un nivel de confianza del 95%, existe

evidencia estadística muy fuerte para afirmar que la proporción de videojuegos del género “Role-Playing” en la consola PS2 es significativamente superior a la proporción en la consola PS3.