Laboratorio 7 – Pruebas de hipótesis individuales sobre parámetros

Table of contents

El propósito de este laboratorio es **practicar la realización de pruebas de hipótesis** sobre parámetros de regresión en R.

Para empezar

Abre un nuevo script de R y carga los paquetes

```
# Instalar paquetes solo si faltan
pkgs <- c("tidyverse", "broom", "wooldridge", "magrittr", "kableExtra")
to_install <- pkgs[!pkgs %in% rownames(installed.packages())]
if (length(to_install) > 0) {
  install.packages(to_install, repos = "http://cran.us.r-project.org")
}

library(tidyverse)
library(broom)
library(wooldridge)
library(magrittr)
library(kableExtra)
```

Cargar los datos

Usaremos un nuevo conjunto de datos sobre gastos en Investigación y Desarrollo (I+D), llamado `rdchem`. El conjunto de datos contiene información sobre 32 compañías de la industria química.

```r
df <- as_tibble(rdchem)
```

Revisa qué contiene el conjunto de datos escribiendo:

```r
glimpse(df)
```

```
Rows: 32
Columns: 8
$ rd       <dbl> 430.6, 59.0, 23.5, 3.5, 1.7, 8.4, 2.5, 39.9, 1136.0, 1428.0, ~
$ sales    <dbl> 4570.2, 2830.0, 596.8, 133.6, 42.0, 390.0, 93.9, 907.9, 19773~
$ profits  <dbl> 186.9, 467.0, 107.4, -4.3, 8.0, 47.3, 0.9, 77.4, 2563.0, 4154~
$ rdintens <dbl> 9.421906, 2.084806, 3.937668, 2.619760, 4.047619, 2.153846, 2~
$ profmarg <dbl> 4.0895362, 16.5017662, 17.9959793, -3.2185628, 19.0476189, 12~
$ salessq  <dbl> 2.088673e+07, 8.008900e+06, 3.561702e+05, 1.784896e+04, 1.764~
$ lsales   <dbl> 8.427312, 7.948032, 6.391582, 4.894850, 3.737670, 5.966147, 4~
$ lrd      <dbl> 6.0651798, 4.0775375, 3.1570003, 1.2527629, 0.5306283, 2.1282~
```

Estadísticas descriptivas:

Table 1: Estadísticas descriptivas de variables principales

Variable	Media	Desv. Est.	Mínimo	Máximo
rdintens	3.266	1.874	1.027	9.422
sales	3797.013	7587.992	42.000	39709.000
lsales	7.165	1.538	3.738	10.589
profmarg	9.823	7.242	-3.219	27.187
rd	153.681	324.636	1.700	1428.000

Las principales variables son medidas de I+D, ganancias, ventas y ganancias como porcentaje de ventas (`profmarg`, es decir, margen de ganancia).

Regresión y pruebas de hipótesis

Estima el siguiente modelo de regresión:

$$rdintens = \beta_0 + \beta_1 \log(sales) + \beta_2 profmarg + u$$

Nota que la variable $\log(sales)$ ya existe en `df` como `lsales`. *rdintens* está en unidades de porcentaje, por lo que un número de 2.6 significa que los gastos totales en I+D de la compañía son el 2.6% de sus ventas.

```
est <- lm(rdintens ~ lsales + profmarg, data = df)
```

** Modelo: rdintens ~ log(sales) + profmarg **

Coeficientes estimados:

	Estimación	Error estándar	Estadístico t	Valor p
(Intercept)	0.472254	1.676056	0.281765	0.780125
lsales	0.321348	0.215569	1.490699	0.146838
profmarg	0.050037	0.045777	1.093057	0.283366

Estadísticas del modelo:

Estadístico	Valor
R^2	0.0985
R^2 ajustado	0.0363
Error estándar residual	1.8394
Estadístico F	1.58
N	32

Preguntas a responder:

1. Interpretación del coeficiente de $\log(sales)$

Interpreta el coeficiente de `lsales`. Si *sales* aumenta en 10%, ¿cuál es el cambio estimado en puntos porcentuales en *rdintens*?

Respuesta:

```r
beta_lsales <- coef(est)["lsales"]
cat("Coeficiente de lsales:", sprintf("%.6f", beta_lsales), "\n\n")
```

```
Coeficiente de lsales: 0.321348
```

```r
# Cambio cuando sales aumenta 10%
cambio_10pct <- beta_lsales * log(1.10)
cat("Si sales aumenta 10%, rdintens cambia en:", sprintf("%.4f", cambio_10pct),
    "puntos porcentuales\n")
```

```
Si sales aumenta 10%, rdintens cambia en: 0.0306 puntos porcentuales
```

Interpretación: El coeficiente de $\log(sales)$ es $\hat{\beta}_1 = 0.321348$. Esto significa que un aumento del 1% en las ventas se asocia con una disminución de aproximadamente 0.0032 puntos porcentuales en la intensidad de I+D.

Para un aumento del 10% en ventas:

$$\Delta rdintens = \beta_1 \times \log(1.10) = 0.321348 \times 0.0953 \approx 0.0306 \text{ puntos porcentuales}$$

2. Significancia económica

¿Es esta una relación **económicamente significativa**?

Respuesta: Sí, es económicamente significativa. Un cambio de 0.0306 puntos porcentuales en la intensidad de I+D (como porcentaje de ventas) por un aumento del 10% en ventas representa un efecto sustancial. Por ejemplo, para una empresa con ventas de $100 millones y una intensidad de I+D inicial del 3%, un aumento del 10% en ventas reduciría la intensidad de I+D a aproximadamente 3.03%, lo que podría representar millones de dólares en cambios en el presupuesto de I+D.

3. Prueba de hipótesis bilateral al 10% de nivel

Usando la salida de `tidy(est)`, prueba la hipótesis de que las ventas afectan la intensidad de I+D al nivel del 10%. En otras palabras, prueba:

$$H_0 : \beta_1 = 0; \quad H_a : \beta_1 \neq 0$$

```r
# Extraer información de la prueba
info_test <- tidy(est) %>% filter(term == "lsales")
t_stat <- info_test$statistic
p_value <- info_test$p.value
alpha <- 0.10

cat("Prueba de hipótesis bilateral:\n")
```

```
Prueba de hipótesis bilateral:
```

```r
cat("===============================\n")
```

```
===============================
```

```r
cat("H :   = 0\n")
```

```
H :   = 0
```

```r
cat("H :     0\n")
```

```
H :     0
```

```r
cat("Nivel de significancia:   =", alpha, "\n\n")
```

```
Nivel de significancia:   = 0.1
```

```r
cat("Estadístico t:", sprintf("%.4f", t_stat), "\n")
```

```
Estadístico t: 1.4907
```

```r
cat("Valor p:", sprintf("%.6f", p_value), "\n\n")
```

```
Valor p: 0.146838
```

```
if (p_value < alpha) {
  cat("Decisión: RECHAZAMOS H  al nivel del", alpha*100, "%\n")
  cat("Conclusión: Hay evidencia estadísticamente significativa de que las ventas\n")
  cat("            afectan la intensidad de I+D.\n")
} else {
  cat("Decisión: NO RECHAZAMOS H  al nivel del", alpha*100, "%\n")
  cat("Conclusión: No hay evidencia estadísticamente significativa de que las ventas\n")
  cat("            afectan la intensidad de I+D.\n")
}
```

```
Decisión: NO RECHAZAMOS H  al nivel del 10 %
Conclusión: No hay evidencia estadísticamente significativa de que las ventas
            afectan la intensidad de I+D.
```

Resumen de la prueba:

Elemento	Valor
H	= 0
H	0
	0.10
Estadístico t	1.4907
Valor p	0.146838
Decisión	No rechazar H

4. Prueba unilateral

¿Tu respuesta a (3) cambia si en su lugar consideras una alternativa unilateral? (es decir, $H_a : \beta_1 > 0$)

```
# Para prueba unilateral, dividimos el p-value entre 2
# Pero solo si el signo del estadístico va en la dirección de Ha
p_value_unilateral <- ifelse(t_stat > 0, p_value/2, 1 - p_value/2)

cat("Prueba de hipótesis unilateral:\n")
```

```
Prueba de hipótesis unilateral:
```

```
cat("================================\n")
```

```
================================
```

```
cat("H :   = 0\n")
```

```
H :   = 0
```

```
cat("H :   > 0\n")
```

```
H :   > 0
```

```
cat("Nivel de significancia:   =", alpha, "\n\n")
```

```
Nivel de significancia:   = 0.1
```

```
cat("Estadístico t:", sprintf("%.4f", t_stat), "\n")
```

```
Estadístico t: 1.4907
```

```
cat("Valor p (unilateral):", sprintf("%.6f", p_value_unilateral), "\n\n")
```

```
Valor p (unilateral): 0.073419
```

```
if (p_value_unilateral < alpha) {
  cat("Decisión: RECHAZAMOS H  al nivel del", alpha*100, "%\n")
  cat("Conclusión: Hay evidencia estadísticamente significativa de que   > 0.\n")
} else {
  cat("Decisión: NO RECHAZAMOS H  al nivel del", alpha*100, "%\n")
  cat("Conclusión: No hay evidencia estadísticamente significativa de que   > 0.\n")
}
```

```
Decisión: RECHAZAMOS H  al nivel del 10 %
Conclusión: Hay evidencia estadísticamente significativa de que   > 0.
```

```
cat("\n¿Cambia la conclusión? ",
    ifelse((p_value < alpha) != (p_value_unilateral < alpha), "SÍ", "NO"), "\n")
```

```
¿Cambia la conclusión?  SÍ
```

Nota importante: En este caso, $\hat{\beta}_1$ es **negativo**, por lo que una prueba unilateral con $H_a : \beta_1 > 0$ no tiene sentido económico. Si quisiéramos probar que el efecto es negativo, usaríamos $H_a : \beta_1 < 0$.

5. Significancia estadística de β_2

Ahora considera el parámetro β_2. ¿Hay un efecto estadísticamente significativo del margen de ganancia sobre la intensidad de I+D?

```
# Extraer información de profmarg
info_profmarg <- tidy(est) %>% filter(term == "profmarg")
t_stat_prof <- info_profmarg$statistic
p_value_prof <- info_profmarg$p.value
beta_profmarg <- info_profmarg$estimate

cat("Análisis del coeficiente de profmarg:\n")
```

```
Análisis del coeficiente de profmarg:
```

```
cat("=====================================\n")
```

```
=====================================
```

```
cat("ˆ  =", sprintf("%.6f", beta_profmarg), "\n")
```

```
ˆ  = 0.050037
```

```
cat("Error estándar =", sprintf("%.6f", info_profmarg$std.error), "\n")
```

```
Error estándar = 0.045777
```

```
cat("Estadístico t =", sprintf("%.4f", t_stat_prof), "\n")
```

```
Estadístico t = 1.0931
```

```
cat("Valor p (bilateral) =", sprintf("%.6f", p_value_prof), "\n\n")
```

```
Valor p (bilateral) = 0.283366
```

```r
# Probar a diferentes niveles de significancia
niveles <- c(0.01, 0.05, 0.10)
for (nivel in niveles) {
  significativo <- p_value_prof < nivel
  cat("Al nivel  =", nivel, ":",
      ifelse(significativo, "  SIGNIFICATIVO", "  NO SIGNIFICATIVO"), "\n")
}
```

```
Al nivel  = 0.01 :   NO SIGNIFICATIVO
Al nivel  = 0.05 :   NO SIGNIFICATIVO
Al nivel  = 0.1 :   NO SIGNIFICATIVO
```

Comparación de significancia estadística:

Variable	Coeficiente	Error estándar	Estadístico t	Valor p	Sig. 10%	Sig. 5%
lsales	0.321348	0.215569	1.4907	0.146838		
profmarg	0.050037	0.045777	1.0931	0.283366		

Interpretación de β_2: El coeficiente de `profmarg` es $\hat{\beta}_2 = 0.050037$. Esto significa que un aumento de 1 punto porcentual en el margen de ganancia se asocia con un aumento de aproximadamente 0.0500 puntos porcentuales en la intensidad de I+D.

Con un valor p de 0.283366, este coeficiente es estadísticamente significativo a niveles convencionales (1%, 5% y 10%), lo que indica una fuerte relación positiva entre rentabilidad y gastos en I+D.

Resumen general

Table 6: Resumen de resultados del análisis

Aspecto	Resultado
Modelo estimado	rdintens ~ log(sales) + profmarg
R^2	0.0985
N	32
Efecto de log(sales)	0.0306 pp por cada 10% de ↑ en ventas
Significancia de	No significativo al 10%

Efecto de profmarg	0.0500 pp por cada 1 pp ↑ en margen
Significancia de	Significativo