

# Laboratorio 7 – Pruebas de hipótesis sobre parámetros de regresión

## Table of contents

Primeros pasos . . . . .	1
Cargar los datos . . . . .	2
Regresión y pruebas de hipótesis . . . . .	2
Estimación del modelo . . . . .	4
Preguntas de análisis (responde en tu script) . . . . .	4
Ayuda: cómo hacer la prueba unilateral con el p-value de tidy() . . . . .	5

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

## Primeros pasos

Abre un nuevo script y carga los paquetes

```
# Agrega los nombres de los integrantes del grupo AQUÍ (si corresponde)

# Instalar paquetes solo si faltan (estilo ECO/EPG)
pkgs <- c("tidyverse", "broom", "wooldridge", "magrittr", "modelsummary", "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(modelsummary)
library(kableExtra)

# Receta ECO/EPG: forzar tablas estables (LaTeX clásico)
options(modelsummary_factory_default = "kableExtra")
```

## Cargar los datos

Usaremos un conjunto de datos sobre gasto en Investigación y Desarrollo (I+D), llamado `rdchem`. Contiene información de 32 empresas de la industria química.

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

Revisa qué contiene el conjunto de datos:

```
datasummary_df(df)
```

```
datasummary_skim(df, histogram = FALSE)
```

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

## Regresión y pruebas de hipótesis

Estimaremos el siguiente modelo:

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

Observa que la variable `log(sales)` ya existe en `df` como `lsales`.

Además, `rdintens` está en **unidades porcentuales**: por ejemplo, un valor de 2.6 significa que el gasto total en I+D de la empresa equivale al **2.6%** de sus ventas.

rd	sales	profits	rdintens	profmarg	salessq	lsales	lrd
430.60	4570.20	186.90	9.42	4.09	$2.088\,673 \times 10^7$	8.43	6.07
59.00	2830.00	467.00	2.08	16.50	$8.008\,900 \times 10^6$	7.95	4.08
23.50	596.80	107.40	3.94	18.00	$3.561\,702 \times 10^5$	6.39	3.16
3.50	133.60	-4.30	2.62	-3.22	$1.784\,896 \times 10^4$	4.89	1.25
1.70	42.00	8.00	4.05	19.05	$1.764\,000 \times 10^3$	3.74	0.53
8.40	390.00	47.30	2.15	12.13	$1.521\,000 \times 10^5$	5.97	2.13
2.50	93.90	0.90	2.66	0.96	$8.817\,210 \times 10^3$	4.54	0.92
39.90	907.90	77.40	4.39	8.53	$8.242\,824 \times 10^5$	6.81	3.69
1136.00	19\,773.00	2563.00	5.75	12.96	$3.909\,715 \times 10^8$	9.89	7.04
1428.00	39\,709.00	4154.00	3.60	10.46	$1.576\,805 \times 10^9$	10.59	7.26
45.30	2936.50	93.70	1.54	3.19	$8.623\,032 \times 10^6$	7.98	3.81
65.20	2513.80	355.40	2.59	14.14	$6.319\,191 \times 10^6$	7.83	4.18
20.30	1124.80	45.30	1.80	4.03	$1.265\,175 \times 10^6$	7.03	3.01
15.60	921.60	4.10	1.69	0.44	$8.493\,465 \times 10^5$	6.83	2.75
74.00	2432.60	132.40	3.04	5.44	$5.917\,543 \times 10^6$	7.80	4.30
147.50	6754.00	329.00	2.18	4.87	$4.561\,652 \times 10^7$	8.82	4.99
29.20	1066.30	289.90	2.74	27.19	$1.136\,996 \times 10^6$	6.97	3.37
92.20	3199.90	163.20	2.88	5.10	$1.023\,936 \times 10^7$	8.07	4.52
3.70	150.00	32.60	2.47	21.73	$2.250\,000 \times 10^4$	5.01	1.31
19.40	509.70	83.80	3.81	16.44	$2.597\,941 \times 10^5$	6.23	2.97
74.40	1452.70	271.20	5.12	18.67	$2.110\,337 \times 10^6$	7.28	4.31
612.00	8995.00	809.00	6.80	8.99	$8.091\,002 \times 10^7$	9.10	6.42
45.30	1212.30	215.00	3.74	17.73	$1.469\,671 \times 10^6$	7.10	3.81
11.00	906.60	154.80	1.21	17.07	$8.219\,235 \times 10^5$	6.81	2.40
66.00	2592.00	116.00	2.55	4.48	$6.718\,464 \times 10^6$	7.86	4.19
10.40	201.50	23.00	5.16	11.41	$4.060\,225 \times 10^4$	5.31	2.34
48.90	2617.80	60.50	1.87	2.31	$6.852\,877 \times 10^6$	7.87	3.89
6.10	502.20	18.40	1.21	3.66	$2.522\,049 \times 10^5$	6.22	1.81
178.20	2824.00	313.00	6.31	11.08	$7.974\,976 \times 10^6$	7.95	5.18
3.00	292.20	6.40	1.03	2.19	$8.538\,084 \times 10^4$	5.68	1.10
191.00	7621.00	626.00	2.51	8.21	$5.807\,964 \times 10^7$	8.94	5.25
26.00	1631.50	105.80	1.59	6.48	$2.661\,792 \times 10^6$	7.40	3.26

	Unique	Missing	Pct.	Mean	SD	Min	Median
rd	31	0		\num{153.7}	\num{324.6}	\num{1.7}	\num{42.6}
sales	32	0		\num{3797.0}	\num{7588.0}	\num{42.0}	\num{1332.5}
profits	32	0		\num{370.5}	\num{830.9}	\num{-4.3}	\num{111.7}
rdintens	32	0		\num{3.3}	\num{1.9}	\num{1.0}	\num{2.6}
profmarg	32	0		\num{9.8}	\num{7.2}	\num{-3.2}	\num{8.8}
salessq	32	0		\num{70195631.7}	\num{283664044.9}	\num{1764.0}	\num{179000}
lsales	32	0		\num{7.2}	\num{1.5}	\num{3.7}	\num{7.2}
lrd	31	0		\num{3.6}	\num{1.7}	\num{0.5}	\num{3.7}

## Estimación del modelo

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

```
# A tibble: 3 x 5
  term      estimate std.error statistic p.value
  <chr>     <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept) 0.472     1.68     0.282    0.780
2 lsales       0.321     0.216     1.49     0.147
3 profmarg     0.0500    0.0458    1.09     0.283
```

Mostramos la tabla del modelo (estable para PDF):

```
modelsummary(est, output = "kableExtra") |>
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position")
```

## Preguntas de análisis (responde en tu script)

Responde las siguientes preguntas (puedes escribir tus respuestas como comentarios `#` en tu archivo `.R`):

1. **Interpreta el coeficiente de `lsales`.** Si `sales` aumenta en 10%, ¿cuál es el cambio estimado (en puntos porcentuales) en `rdintens`?
2. ¿Esta relación es **económicamente significativa** (además de estadísticamente significativa)?

	(1)
(Intercept)	0.472 (1.676)
lsales	0.321 (0.216)
profmarg	0.050 (0.046)
Num.Obs.	32
R2	0.098
R2 Adj.	0.036
AIC	134.7
BIC	140.5
Log.Lik.	-63.333
F	1.584
RMSE	1.75

3. Usando la salida de `tidy(est)`, prueba al 10% la hipótesis de que las ventas afectan la intensidad de I+D. Es decir, prueba:

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

4. ¿Cambia tu respuesta de (3) si en vez de una alternativa bilateral consideras una alternativa unilateral? (es decir,  $H_a : \beta_1 > 0$ )  
 5. Considera ahora el parámetro  $\beta_2$ . ¿Existe un efecto estadísticamente significativo del margen de utilidades (`profmarg`) sobre la intensidad de I+D (`rdintens`)?

**Ayuda: cómo hacer la prueba unilateral con el p-value de `tidy()`**

`tidy(est)` entrega un p-value bilateral para cada coeficiente. Si el estadístico t es positivo y quieres una prueba unilateral  $H_a : \beta_1 > 0$ , el p-value unilateral es aproximadamente la mitad del bilateral (siempre que el signo vaya en la dirección de la alternativa).

Puedes calcularlo así:

```
t_lsales <- tidy(est) %>% filter(term == "lsales") %>% pull(statistic)
p_two     <- tidy(est) %>% filter(term == "lsales") %>% pull(p.value)

p_one_greater <- ifelse(t_lsales > 0, p_two / 2, 1 - p_two / 2)
c(t_stat = t_lsales, p_two_sided = p_two, p_one_sided_greater = p_one_greater)
```

t_stat	p_two_sided	p_one_sided_greater
1.49069937	0.14683824	0.07341912