

Unidad 4: Introducción a la inferencia estadística

Análisis gráfico, correlación y regresión

Nombre: Farias, Gustavo

Comisión: M2025-13

Matrícula: 101662

Repositorio GitHub:

<https://github.com/Lucenear/UTN-TUPaD-TPs/tree/main/Probabilidad%20y%20Estadistica>

Consignas:

Contexto del caso

Una empresa de desarrollo de software educativo recolectó datos sobre la cantidad de horas de capacitación técnica que recibieron sus programadores durante el último semestre, y su rendimiento medido por cantidad de incidencias resueltas en tiempo y forma.

Se busca determinar si existe una relación estadística entre ambas variables, y si es posible ajustar un modelo de regresión que permita hacer predicciones sobre el rendimiento en función de la capacitación.

Tabla de datos observados

Empleado	Horas de capacitación	Incidencias resueltas
A	10	6
B	12	8
C	8	4
D	14	9
E	6	3
F	16	10
G	9	5
H	13	8

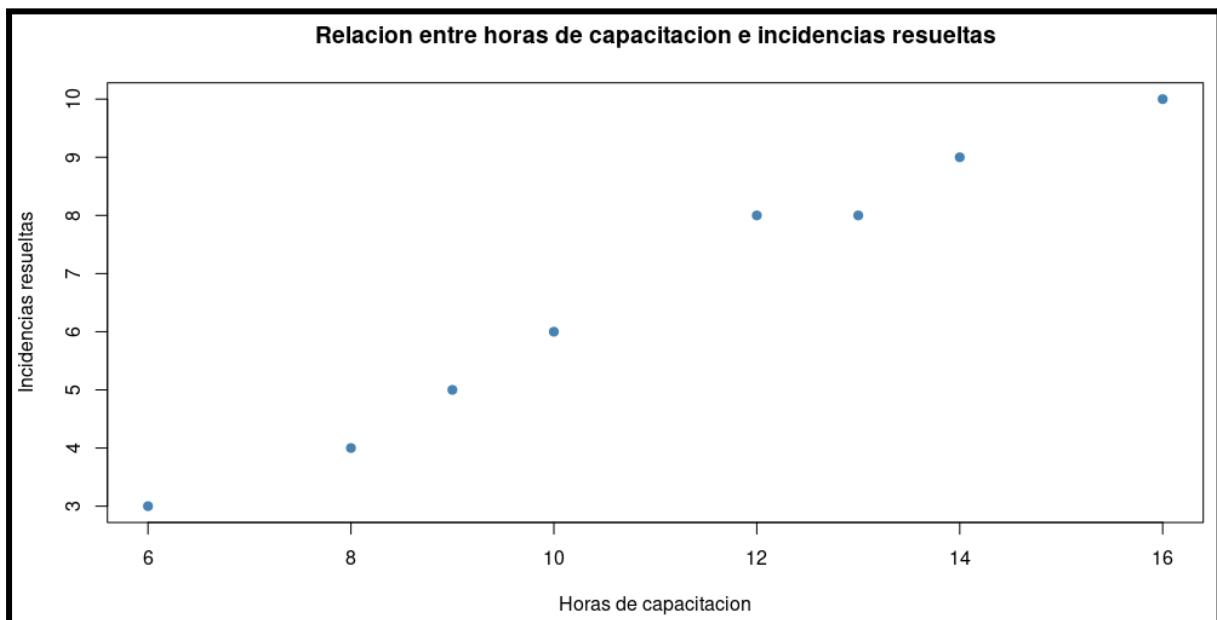
1. Realizá un diagrama de dispersión de los datos presentados. ¿Qué tipo de relación visualizás entre las variables?
2. Calculá el coeficiente de correlación de Pearson. Interpretá su valor en función de la fuerza y dirección de la relación.
3. Calculá la recta de regresión lineal que mejor ajusta los datos. Escribí su ecuación e interpretá los coeficientes.
4. ¿Qué valor de incidencias resueltas se esperaría para un empleado que realizó 11 horas de capacitación? Usá el modelo de regresión para responder.
5. ¿Creés que este modelo es adecuado para tomar decisiones en la empresa? Justificá tu respuesta haciendo referencia a los conceptos de correlación, regresión y coeficiente de determinación (R^2).

Respuestas:

1) Diagrama de dispersión realizado en R:

```
# Cargo los datos
horas <- c(10, 12, 8, 14, 6, 16, 9, 13)
incidencias <- c(6, 8, 4, 9, 3, 10, 5, 8)

# Diagrama de dispersion
plot(horas, incidencias,
     main = "Relacion entre horas de capacitacion e incidencias resueltas",
     xlab = "Horas de capacitacion",
     ylab = "Incidencias resueltas",
     pch = 19,
     col = "steelblue")
```



El gráfico muestra una tendencia creciente: a medida que aumentan las horas de capacitación, también aumenta el número de incidencias resueltas. Los puntos están bien alineados, simulando una línea recta, sin mucha dispersión, por lo cual infiero que hay una relación lineal positiva fuerte entre las variables.

2) Coeficiente de correlacion de Pearson:

```
### EJERCICIO 2 ###
# Calculo correlacion de Pearson
r <- cor(horas, incidencias, method = "pearson")
r
```




```
> r  
[1] 0.991453
```

El coeficiente de correlación es $r = 0.991453$, lo que indica una relación lineal positiva muy fuerte entre las horas de capacitación y el número de incidencias resueltas.

Es decir, a más capacitación, más incidencias resueltas, por lo cual la relación es casi perfectamente lineal.

3) Recta de regresión lineal:

```
### EJERCICIO 3 ###  
# Ajusto modelo de regresion lineal  
modelo <- lm(incidencias ~ horas)  
  
# Muestro resumen  
summary(modelo)  
  
# Extraigo coeficientes  
coef(modelo)
```

```
Call:  
lm(formula = incidencias ~ horas)  
  
Residuals:  
    Min       1Q   Median       3Q      Max   
-0.39423 -0.18910 -0.00962  0.12500  0.63141  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) -1.55449    0.45677  -3.403   0.0144 *  
horas        0.74359    0.03995  18.615 1.55e-06 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.3528 on 6 degrees of freedom  
Multiple R-squared:  0.983,    Adjusted R-squared:  0.9801  
F-statistic: 346.5 on 1 and 6 DF, p-value: 1.551e-06  
  
>  
> # Extraigo coeficientes  
> coef(modelo)  
(Intercept)      horas   
-1.5544872    0.7435897
```

La recta de regresión captura una relación lineal positiva muy fuerte entre la capacitación y el rendimiento. Aunque el intercepto carece de sentido práctico, la pendiente es estadísticamente significativa, ya que por cada hora extra de

capacitación técnica, el número esperado de incidencias resueltas aumenta en aproximadamente 0.74

4) Predicción con 11 horas de capacitación técnica:

```
### EJERCICIO 4 ###  
#Prediccion con 11 horas de capacitacion  
predict(modelo, newdata = data.frame(horas = 11))
```

```
> ### EJERCICIO 4 ###  
> #Prediccion con 11 horas de capacitacion  
> predict(modelo, newdata = data.frame(horas = 11))  
      1  
6.625
```

Para un empleado que realizó 11 horas de capacitación, el modelo predice que resolverá aproximadamente 6.63 incidencias, lo que en promedio daría como resultado promedio, entre 6 y 7 incidencias resueltas.

5) Modelo adecuado para resolver decisiones empresariales?

Si, considero que este modelo es adecuado para apoyar decisiones empresariales, como:

- Planificar inversiones en capacitación
- Establecer metas de rendimiento por niveles de formación
- Estimar el impacto de aumentar las horas de entrenamiento