Guía de estudio - Regresión Lineal

 ¡Hola! Te damos la bienvenida a esta nueva guía de estudio.

¿En qué consiste esta guía?

La siguiente guía de estudio tiene como objetivo profundizar contenidos adicionales, además de recordar y repasar los temas que hemos tratado en clase y que pueden ser de ayuda en tu estudio y análisis de los modelos regresión lineal .

¡Vamos con todo!

Tabla de contenidos

¡Comencemos!

Covarianza y correlación lineal

Covarianza

La covarianza es una medida estadística que indica la relación entre dos variables aleatorias. En otras palabras, mide cómo cambian juntas dos variables en comparación con sus valores medios. Si los valores de ambas variables tienden a aumentar o disminuir juntos, la covarianza será positiva. Si los valores de una variable tienden a aumentar cuando los valores de la otra disminuyen, la covarianza será negativa. Si no hay un patrón discernible en la relación entre las variables, la covarianza será cercana a cero.

La fórmula general para calcular la covarianza entre dos conjuntos de datos X e Y es la siguiente

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{n-1}$$

Donde:

- n es el número de observaciones en los conjuntos de datos.
- x_i e y_i son los valores respectivos de los conjuntos X e Y
- $\bar{x}$ e $\bar{y}$ son las medias aritméticas de los conjuntos

En esta fórmula, se calcula la diferencia entre cada valor de X y su media, y lo mismo para los valores de Y. Luego, se multiplican estas diferencias para cada observación y se suman. Finalmente, se divide entre n - 1 en lugar de n para ajustar la covarianza muestral, ya que esto proporciona un estimado menos sesgado de la covarianza en comparación con el uso de n en el denominador.

La covarianza por sí sola puede ser difícil de interpretar porque su magnitud depende de las unidades en las que se miden las variables. Para tener una idea más clara de la relación entre las variables, a menudo se utiliza el **coeficiente de correlación**, que es una normalización de la covarianza y varía entre -1 y 1. Así se elimina la influencia de las unidades y proporciona una medida estandarizada de la relación entre las variables.

El **coeficiente de correlación** es una medida estandarizada que cuantifica la relación entre dos variables en términos de su magnitud y dirección. El coeficiente de correlación más comúnmente utilizado es el **coeficiente de correlación de Pearson**, denotado como **r**. Se calcula utilizando la siguiente fórmula:

$$r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sigma_x \sigma_y}$$

En esta fórmula, se calcula la covarianza entre X e Y en el numerador, y en el denominador se calculan las desviaciones estándar respectivas para estandarizar la medida. El resultado es un valor r que oscila entre -1 y 1. Los significados de los valores de r son los siguientes:

- r=1 indica una correlación positiva perfecta, lo que significa que las variables aumentan juntas en una relación lineal.
- r=−1 indica una correlación negativa perfecta, lo que significa que una variable aumenta mientras la otra disminuye en una relación lineal.
- r cercano a 0 indica una correlación lineal débil o inexistente entre las variables.

Es importante señalar que el coeficiente de correlación de Pearson solo mide relaciones lineales entre variables. Si la relación entre las variables no es lineal, el coeficiente de correlación puede subestimar o no capturar la verdadera relación entre ellas. En tales casos, otros tipos de correlación, como el coeficiente de correlación de Spearman o el coeficiente de correlación de Kendall, podrían ser más apropiados.

Regresión lineal

La regresión lineal es un método estadístico utilizado para modelar la relación entre una variable independiente (o predictor) y una variable dependiente. En otras palabras, busca establecer una relación lineal que permita predecir o explicar el comportamiento de la variable dependiente en función de la variable independiente. Es uno de los métodos más simples y ampliamente utilizados en estadística y análisis de datos.

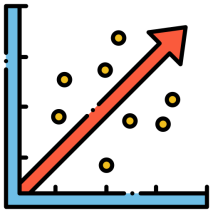
En un modelo de regresión lineal, se asume que existe una relación lineal aproximada entre las dos variables. La forma general de una regresión lineal es:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon$$

Donde:

- y es la variable dependiente que se intenta predecir o explicar,
- x es la variable independiente o predictor,
- β_0 es el intercepto o la ordenada al origen,
- β_1 es la pendiente que representa el cambio en y por un cambio unitario en x , y
- ε es el término de error, que captura las diferencias entre los valores observados y los valores predichos por el modelo.

El objetivo de la regresión lineal es encontrar los valores de β_0 y β_1 que minimicen la suma de los cuadrados de los errores (la distancia vertical entre los puntos reales y la línea de regresión). Esto se conoce como **método de los mínimos cuadrados.**



La regresión lineal se utiliza en diversas áreas, como la economía, la biología, la ingeniería, las ciencias sociales y más, para analizar relaciones y predecir valores. Además de la regresión lineal simple (con una variable independiente), también existe la regresión lineal múltiple, que involucra varias variables independientes.

¿Por qué es importante la regresión lineal?

La regresión lineal es importante por varias razones en el campo de la estadística y el análisis de datos:

- **Modelado de Relaciones:** La regresión lineal permite modelar relaciones entre variables, lo que es esencial para comprender cómo una variable afecta o se relaciona con otra. Esta capacidad es útil en áreas como la investigación científica, la economía, la medicina y más, donde es esencial entender cómo ciertas variables influyen en otras.
- **Predicciones:** La regresión lineal se utiliza para predecir valores futuros basados en datos históricos. Por ejemplo, puedes usar un modelo de regresión lineal para predecir ventas futuras, tasas de crecimiento, demanda de productos, entre otros.
- **Análisis de Causa y Efecto:** La regresión lineal puede ayudar a identificar y cuantificar la relación causal entre variables. Esto es especialmente útil para tomar decisiones informadas y desarrollar estrategias basadas en datos.
- **Simplificación de Datos Complejos:** Cuando se tienen múltiples variables, la regresión lineal puede ayudar a resumir la relación entre ellas en una ecuación simple. Esto hace que los datos sean más comprensibles y fáciles de comunicar.
- **Validación de Hipótesis:** La regresión lineal puede ser utilizada para probar hipótesis y responder preguntas específicas sobre cómo las variables están relacionadas entre sí.
- **Control de Calidad y Optimización:** En el ámbito industrial y de la ingeniería, la regresión lineal puede ser utilizada para optimizar procesos y controlar la calidad, identificando variables que influyen en los resultados y ajustando factores para obtener mejores resultados.
- **Fundamento para Modelos Más Complejos:** La regresión lineal es un punto de partida para modelos más complejos, como la regresión lineal múltiple, regresión logística y otros métodos de análisis de regresión.
- **Toma de Decisiones Basada en Datos:** La regresión lineal proporciona una herramienta objetiva para tomar decisiones informadas, basadas en patrones y relaciones observadas en los datos.

Parámetros de la regresión lineal

La interpretación de la relación entre variables en la regresión lineal depende de los coeficientes de regresión y de sus valores p asociados. Aquí tienes una guía general para interpretar esta relación:

Intercepto (β_0): El intercepto representa el valor de la variable dependiente (y) cuando todas las variables independientes (x) son iguales a cero. Sin embargo, en muchos casos, la interpretación del intercepto puede carecer de sentido práctico si las variables independientes no pueden ser cero en la realidad.

Coeficientes de Variables Independientes ($\beta_1, \beta_2, \beta_3, ...\beta_i$): Estos coeficientes muestran cómo un cambio unitario en una variable independiente afecta a la variable dependiente (y), manteniendo todas las demás variables constantes. Aquí hay algunas interpretaciones comunes para los coeficientes:

- Si el coeficiente (β) es positivo, un aumento en la variable independiente está asociado con un aumento en la variable dependiente.
- Si el coeficiente (β) es negativo, un aumento en la variable independiente está asociado con una disminución en la variable dependiente.
- La magnitud del coeficiente (β) indica cuánto cambia la variable dependiente por un cambio unitario en la variable independiente.

Valores p de los Coeficientes: Los valores p asociados con los coeficientes indican si los coeficientes son estadísticamente significativos. Un valor p bajo (generalmente menos de 0.05) sugiere que el coeficiente es significativamente diferente de cero, lo que significa que la variable independiente tiene un efecto significativo en la variable dependiente.

Coeficientes de Interacción: En casos de regresión lineal múltiple, donde hay interacciones entre variables independientes, los coeficientes de interacción indican cómo el efecto de una variable independiente puede cambiar dependiendo del valor de otra variable independiente.

Supuestos del modelo de la regresión lineal

El modelo de regresión lineal se basa en varias suposiciones que deben cumplirse para que los resultados y las interpretaciones sean válidos y confiables. Estas suposiciones son esenciales para garantizar que el modelo capture adecuadamente las relaciones entre las variables y produzca estimaciones precisas. Aquí están las suposiciones clave del modelo de regresión lineal:

- **Linealidad:** La relación entre las variables independientes y la variable dependiente debe ser aproximadamente lineal. Esto significa que los cambios en las variables

independientes deben estar asociados con cambios proporcionales en la variable dependiente.

- **Homocedasticidad:** La varianza de los errores debe ser constante en todos los niveles de las variables independientes. En otras palabras, los errores no deben aumentar ni disminuir a medida que cambian los valores de las variables independientes.

- **Independencia de Errores:** Los errores (residuos) del modelo deben ser independientes entre sí. Esto significa que los valores de los errores en una observación no deben estar correlacionados con los valores de los errores en otras observaciones.

- **Normalidad de Errores:** Los errores deben seguir una distribución normal con una media de cero. Esto es importante para realizar inferencias estadísticas y pruebas de hipótesis sobre los coeficientes del modelo.

- **No Colinealidad:** Las variables independientes no deben estar altamente correlacionadas entre sí. La colinealidad puede dificultar la identificación individual de los efectos de cada variable independiente sobre la variable dependiente.

- **Ausencia de Multicolinealidad Grave:** La multicolinealidad ocurre cuando hay una alta correlación entre dos o más variables independientes. Puede hacer que sea difícil determinar las contribuciones individuales de las variables al modelo.

- **Homogeneidad de Varianza:** La varianza de los errores debe ser constante en todos los niveles de las variables independientes. Esto se relaciona con la homocedasticidad, pero también aborda la heteroscedasticidad.

Pruebas diagnósticas de los supuestos de regresión

Existen varias pruebas y métodos diagnósticos que se utilizan para comprobar si se cumplen las suposiciones del modelo de regresión. Estas pruebas pueden ayudar a evaluar la validez y confiabilidad de los resultados del modelo. Aquí hay algunas pruebas diagnósticas comunes:

- **Gráfico de Residuos vs. Valores Ajustados:** Este gráfico muestra los residuos (diferencias entre los valores observados y los valores predichos) en función de los valores ajustados. Si los residuos están distribuidos aleatoriamente alrededor de cero y no muestran ningún patrón, esto sugiere que se cumple la homocedasticidad y la linealidad.

- **Gráfico QQ (Quantile-Quantile) de Residuos:** Este gráfico compara los cuantiles de los residuos con los cuantiles de una distribución teórica normal. Si los residuos se ajustan a una línea recta en el gráfico QQ, esto indica normalidad en los errores.

- **Gráfico de Histograma de Residuos:** Un histograma de los residuos puede ayudar a verificar la normalidad de los errores. Si los residuos siguen una distribución aproximadamente normal, el histograma debería ser similar a una campana.

- **Prueba de Jarque-Bera:** Esta prueba estadística evalúa si los residuos tienen una distribución normal. Un valor p alto sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad.

- **Prueba de Breusch-Pagan / Prueba de White:** Estas pruebas se utilizan para evaluar la homocedasticidad. Si los valores p son altos, no hay suficiente evidencia para rechazar la hipótesis nula de homocedasticidad.

- **Factor de Inflación de la Varianza (VIF):** El VIF mide la multicolinealidad entre las variables independientes. Valores de VIF muy altos (generalmente por encima de 10) indican que podría haber problemas de multicolinealidad.

- **Gráfico de Residuos Parciales:** Este gráfico muestra los residuos parciales en función de una variable independiente específica. Puede ayudar a detectar relaciones no lineales o influencias excesivas de una variable.

- **Prueba de Durbin-Watson:** Esta prueba se utiliza para evaluar la autocorrelación de los residuos. Un valor cercano a 2 sugiere ausencia de autocorrelación.

- **Prueba de Shapiro-Wilk:** Esta prueba evalúa la normalidad de los residuos. Un valor p alto indica que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad.

Implementación en Python con Statsmodels

En Python, statsmodels es una herramienta poderosa para realizar análisis estadísticos y modelado en Python. Antes de comenzar, asegúrate de que tienes statsmodels instalado en tu entorno de Python. Puedes instalarlo usando pip:

```
pip install statsmodels
```

Con statsmodels ya instalado, podemos importarlo desde nuestro ide

```
import numpy as np #tambien importamos numpy
import pandas as pd #tambien importamos Pandas
import statsmodels.api as sm
```

Simularemos ahora algunos datos utilizando un modelo lineal con “ruido”, es decir, con un cierto margen de error.

```
# Crear datos simulados
np.random.seed(42)
X =np.random.rand(100, 1) #Variable independiente
y =2*X + 1 + np.random.randn(100, 1)  # Variable dependiente con ruido
data = pd.DataFrame({'X': X.flatten(), 'y': y.flatten()})
data.head(5)
```

	X	y
0	0.374540	1.836127
1	0.950714	2.602421
2	0.731994	2.555749
3	0.598658	0.209748
4	0.156019	1.092365

Podemos añadir el parámetro constante a la matriz de característica y ajustar un modelo mediante OLS. Con el método **summary()** podemos ver las características del modelo entrenar.El resumen del modelo te proporcionará información sobre los coeficientes, estadísticas de ajuste y p-values.

```
# Añadir una constante a la matriz de características
X = sm.add_constant(X)

# Crear el modelo OLS (Ordinary Least Squares)
model = sm.OLS(y, X).fit()

# Imprimir un resumen del modelo
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.205
Model:                            OLS   Adj. R-squared:                  0.197
Method:                 Least Squares   F-statistic:                     25.25
Date:                Sun, 27 Aug 2023   Prob (F-statistic):           2.26e-06
Time:                        22:15:47   Log-Likelihood:                -131.15
No. Observations:                 100   AIC:                             266.3
Df Residuals:                      98   BIC:                             271.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.2151      0.170      7.136      0.000       0.877       1.553
x1             1.5402      0.306      5.025      0.000       0.932       2.148
==============================================================================
Omnibus:                        0.900   Durbin-Watson:                   2.285
Prob(Omnibus):                  0.638   Jarque-Bera (JB):                0.808
Skew:                           0.217   Prob(JB):                        0.668
Kurtosis:                       2.929   Cond. No.                         4.18
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctl
```

¡Realizamos un test a nuestro modelo ! Una vez que has ajustado el modelo, puedes realizar predicciones utilizando los datos de entrada:

```
# Crear nuevos datos para predecir
new_X = np.array([[1], [2]])
new_X = sm.add_constant(new_X)

# Realizar predicciones
predictions = model.predict(new_X)
print(predictions)
```

```
[2.75532293 4.2955497 ]
```

Podemos, además, graficar el modelo de regresión lineal y su respuesta .

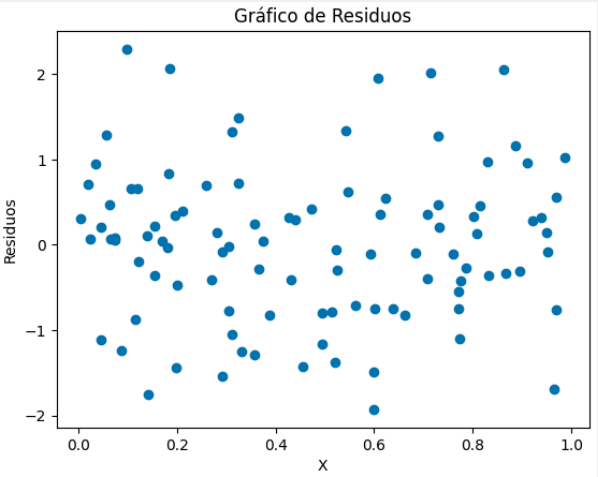
```
import matplotlib.pyplot as plt

# Crear un rango de valores para X
x_range = np.linspace(data['X'].min(), data['X'].max(), 100)
x_range = sm.add_constant(x_range)
```

Es importante realizar diagnósticos del modelo para verificar si se cumplen las suposiciones del modelo. Statsmodels proporciona herramientas para ello, como los gráficos de residuos:

```
# Graficar los residuos
import matplotlib.pyplot as plt

residuals = model.resid
plt.scatter(data['X'], residuals)
plt.xlabel('X')
plt.ylabel('Residuos')
plt.title('Gráfico de Residuos')
plt.show()
```



Interpretación de los estadísticos del modelo de regresión

Los estadísticos del modelo de regresión son medidas que proporcionan información importante sobre la calidad del ajuste del modelo a los datos y la significancia de las relaciones entre las variables. Aquí hay una interpretación general de algunos de los estadísticos clave:

- **R-cuadrado:** Este estadístico varía entre 0 y 1, y representa la proporción de la variabilidad total en la variable dependiente que es explicada por el modelo. Un R cuadrado alto indica que el modelo se ajusta bien a los datos y puede explicar una gran parte de la variabilidad observada.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{total}}}$$

Donde SS_{res} es la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por la línea de regresión, y SS_{total} es la suma de los cuadrados de las diferencias entre los valores observados y la media de los valores observados.

- **Estadístico F y Valor p:** La prueba F se utiliza para evaluar si el modelo en su conjunto es significativo. Un valor alto del estadístico F con un valor p bajo sugiere que al menos una de las variables independientes tiene un efecto significativo en la variable dependiente. En otras palabras, el modelo en su conjunto es significativo.

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

- **Log Likelihood, AIC y BIC:** Estos estadísticos se utilizan para comparar modelos alternativos y evaluar cuál es el mejor ajuste. El Log Likelihood mide la probabilidad

de que los datos observados sean generados por el modelo. El AIC (Criterio de Información de Akaike) y el BIC (Criterio de Información Bayesiana) penalizan los modelos más complejos para evitar el sobreajuste.

- **Coeficientes (β_i)**: Estos coeficientes representan las pendientes de las variables independientes y muestran cómo un cambio unitario en la variable independiente afecta a la variable dependiente, manteniendo las otras variables constantes. Si un coeficiente es positivo, un aumento en la variable independiente está asociado con un aumento en la variable dependiente.

- **Valor p de los Coeficientes:** Los valores p asociados con los coeficientes indican si los coeficientes son estadísticamente significativos. Un valor p bajo (generalmente menos de 0.05) sugiere que el coeficiente es significativamente diferente de cero, lo que significa que la variable independiente tiene un efecto significativo en la variable dependiente.

- **Término de Error (ε):** Este término captura las diferencias entre los valores observados y los valores predichos por el modelo. Un término de error pequeño sugiere que el modelo se ajusta bien a los datos, mientras que un término de error grande indica que el modelo podría no estar capturando todas las relaciones.

La interpretación de estos estadísticos es crucial para evaluar la calidad y significancia del modelo de regresión. Cada estadístico ofrece una perspectiva diferente sobre cómo el modelo se ajusta a los datos y si las relaciones entre las variables son estadísticamente significativas.

Significancia del modelo de regresión

La significancia del modelo de regresión se refiere a la evaluación de si el modelo en su conjunto es estadísticamente significativo para explicar la variabilidad en los datos. En otras palabras, se busca determinar si el modelo de regresión proporciona una explicación válida y útil para las relaciones entre las variables independientes y la variable dependiente.

Para evaluar la significancia del modelo de regresión, se utiliza el estadístico F y su correspondiente valor p. Aquí está cómo se realiza este proceso:

- **Estadístico F:** El estadístico F se calcula como la proporción de la variabilidad explicada por el modelo (la suma de cuadrados de la regresión) dividida por la variabilidad no explicada (la suma de cuadrados del error).

- **Valor p:** Después de calcular el estadístico F, se busca el valor p correspondiente utilizando la distribución F. El valor p indica la probabilidad de obtener un estadístico F igual o más extremo que el calculado bajo la hipótesis nula de que el modelo no es significativo.

- **Interpretación:** Si el valor p es menor que un umbral de significancia predefinido (por ejemplo, 0.05), se rechaza la hipótesis nula y se concluye que el modelo en su

conjunto es significativo. Esto significa que al menos una de las variables independientes tiene un efecto significativo en la variable dependiente.

La significancia del modelo es crucial para determinar si el esfuerzo de construir un modelo de regresión tiene sentido desde una perspectiva estadística. Un modelo significativo indica que al menos una de las variables independientes está relacionada de manera estadísticamente significativa con la variable dependiente, lo que proporciona una base sólida para el análisis y la interpretación de los resultados.

Métricas de evaluación del modelo

Las métricas de evaluación del modelo son medidas utilizadas para cuantificar y evaluar el rendimiento de un modelo de regresión. Estas métricas permiten determinar qué tan bien se ajusta el modelo a los datos y cómo se comporta en la predicción de valores futuros. Aquí hay algunas métricas comunes utilizadas en la evaluación de modelos de regresión:

- **Error Medio Absoluto (MAE):** El MAE calcula el promedio de las diferencias absolutas entre los valores predichos y los valores reales. Es una medida simple y fácil de interpretar. Cuanto menor sea el MAE, mejor será el ajuste del modelo.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Error Cuadrático Medio (MSE):** El MSE calcula el promedio de los cuadrados de las diferencias entre los valores predichos y los valores reales. Pone más peso en los errores más grandes. Un MSE más bajo indica un mejor ajuste.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Raíz del Error Cuadrático Medio (RMSE):** El RMSE es simplemente la raíz cuadrada del MSE. Al estar en las mismas unidades que la variable dependiente, es más fácil de interpretar.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Es importante seleccionar las métricas adecuadas según el contexto y los objetivos del análisis. Algunas métricas se centran en la precisión de las predicciones, mientras que otras evalúan la capacidad del modelo para explicar la variabilidad. Es recomendable utilizar varias métricas en conjunto para obtener una comprensión más completa del rendimiento del modelo.

¡Manos a la obra! - Analizando los sueldos

Considerando la base de datos **earnings.csv**, aplica lo aprendido analizando correlaciones entre las variables y modelos de regresión entre ellas.

Paradoja de Simpson

La paradoja de Simpson es un fenómeno en estadísticas en el que una tendencia o relación que aparece en varios grupos de datos puede desaparecer o incluso revertirse cuando se combinan esos grupos. En otras palabras, la dirección de una relación puede cambiar cuando se considera el efecto de una tercera variable. Esto puede llevar a interpretaciones engañosas si no se tienen en cuenta todas las variables relevantes.

La paradoja de Simpson suele ocurrir cuando hay variables de confusión presentes en los datos. Una variable de confusión es una variable que afecta tanto a la variable independiente como a la dependiente y puede distorsionar la relación entre ellas. Cuando se ignoran o no se controlan las variables de confusión, los resultados pueden llevar a conclusiones erróneas.

Un clásico ejemplo de esta paradoja es un caso de una universidad en dónde se presentó una demanda por discriminación contra las mujeres que habían solicitado su ingreso al postgrado.

Los resultados a simple vista mostraban que los hombres tenían mayor posibilidad de ser admitidos que las mujeres.

	Solicitudes	**Admisiones**
Hombres	8442	**44%**
Mujeres	4321	35%

Pero al examinar los datos de los departamentos de forma individual, se encontró que no existía un sesgo contra las mujeres, sino que se visualizó que en la mayoría de los casos se presentaba un pequeño sesgo a favor de las mujeres:

Departamento	Hombres		Mujeres	
	Solicitudes	Admisiones	Solicitudes	Admisiones
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	272	6%	341	**7%**

La Paradoja de Simpson resalta la importancia de considerar todas las variables relevantes y las posibles relaciones entre ellas al realizar análisis estadísticos. Ignorar variables de confusión puede llevar a conclusiones incorrectas y subraya la necesidad de un enfoque más completo y cuidadoso en el análisis de datos.

Preguntas de proceso

Reflexiona:

- ¿En qué áreas puede ser importante la regresión lineal?
- ¿Has visto casos en que incorrectamente se asocie correlación con causalidad? ¿Cuáles?
- Si el coeficiente de correlación es cercano a cero, ¿será posible que el modelo de regresión sea adecuado? ¿Qué aspectos crees que fallarían?