BAAA, Vol. 66, 2024

Boletín de la
Asociación Argentina de Astronomía

# Ejercicios Practicos: Guia 3

Camila Sofia Beatriz Hormaeche
[1] *Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba*

Received: ... / Accepted: ...


**Resumen** /

*Keywords* /

## 1. Introducción

Resolución de ejercicios prácticos de la guia 3.

## 2. Ejercicio 6: Ajuste por cuadrados mínimos, regresión lineal.

### Objetivo

El objetivo del ejercicio es deducir las fórmulas para los coeficientes del ajuste lineal a un conjunto de puntos $(x_n, y_n)$ utilizando el método de mínimos cuadrados. Además, se busca discutir las diferencias entre asumir errores en los valores de $y$ (variable dependiente) y en los valores de $x$ (variable independiente).

### 2.1. Resolución

Asumimos que los errores están en $y$ y que siguen una distribución gaussiana:

$$P(y_n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_n - y(x_n,\phi))^2}{2\sigma_n^2}}.$$

La probabilidad conjunta de todos los datos es:

$$P(\text{datos}|\text{modelo}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_n - y(x_n,\phi))^2}{2\sigma_n^2}}.$$

Maximizar esta probabilidad es equivalente a minimizar -logaritmo(probabilidad conjunta):

$$s = \sum_{n=1}^{N} \frac{(y_n - y(x_n,\phi))^2}{2\sigma_n^2}.$$

Si asumimos que $\sigma_n = 1$ para todos los $n$, obtenemos la fórmula clásica de **mínimos cuadrados**:

$$s = \sum_{n=1}^{N} (y_n - (\phi_0 + \phi_1 x_n))^2.$$

Para minimizar $s$, derivamos respecto a $\phi_0$ y $\phi_1$, debido a que se quiere encontrar el punto critico:

$$\frac{\partial s}{\partial \phi_0} = 0, \quad \frac{\partial s}{\partial \phi_1} = 0.$$

Las ecuaciones quedan de la siguiente manera:

$$\sum_{n=1}^{N} y_n = N\phi_0 + \phi_1 \sum_{n=1}^{N} x_n,$$

$$\sum_{n=1}^{N} y_n x_n = \phi_0 \sum_{n=1}^{N} x_n + \phi_1 \sum_{n=1}^{N} x_n^2.$$

Resolviendo este sistema, obtenemos las fórmulas de los coeficientes:

$$\phi_1 = \frac{\sum_{n=1}^{N} y_n x_n - \frac{1}{N} \sum_{n=1}^{N} y_n \sum_{n=1}^{N} x_n}{\sum_{n=1}^{N} x_n^2 - \frac{1}{N} \left(\sum_{n=1}^{N} x_n\right)^2},$$

$$\phi_0 = \frac{1}{N} \sum_{n=1}^{N} y_n - \phi_1 \frac{1}{N} \sum_{n=1}^{N} x_n.$$

Todo esto es suponiendo que los errores son en y, la cual es la variable dependiente, recordemos que los datos o la variable con mayor error va en el eje y, y la de menor error (errores despreciables) en el eje x.

Si en lugar de errores en $y$ asumimos errores gaussianos en $x$, el procedimiento es análogo:

$$s = \sum_{n=1}^{N} \frac{(x_n - x(y_n,\phi))^2}{2\sigma_n^2}.$$

En este caso, usamos la relación inversa $x(y)$ para reformular el problema. Si $y = \phi_0 + \phi_1 x$, entonces:

$$x = \frac{y - \phi_0}{\phi_1}.$$

Las ecuaciones normales resultantes para este caso llevan a un ajuste similar, pero con los roles de $x$ e $y$ intercambiados.

### Conclusiones

- Cuando asumimos errores en $y$, el método clásico de mínimos cuadrados es directo y ampliamente utilizado. - Si asumimos errores en $x$, los cálculos requieren usar la relación inversa $x(y)$, lo que puede ser menos común. - Si hay errores significativos en ambas variables, es necesario usar métodos más avanzados, como el ajuste de errores totales. Mínimo Cuadrado

## Ejercicio 7: Demostración de Insesgadez del Estimador Combinado

El objetivo de este ejercicio es demostrar que si tenemos dos estimadores insesgados $\theta_1$ y $\theta_2$ para un parámetro $\theta$,

✉ *Contact* / camila.hormaeche.980@mi.unc.edu.ar

y $\alpha$ es una constante, entonces el estimador combinado:

$$\theta = \alpha\theta_1 + (1-\alpha)\theta_2,$$

también es un estimador insesgado para $\theta$.

Definición de estimadores insesgados: Por la definición de insesgadez, tenemos que:

$$\mathbb{E}[\theta_1] = \theta \quad \text{y} \quad \mathbb{E}[\theta_2] = \theta.$$

Esto significa que los estimadores $\theta_1$ y $\theta_2$ son insesgados.

Estimador combinado: Definimos el estimador combinado $\theta$ como:

$$\theta = \alpha\theta_1 + (1-\alpha)\theta_2,$$

donde $\alpha$ es una constante.

Cálculo del valor esperado del estimador combinado: Para comprobar si $\theta$ es insesgado, calculamos su valor esperado:

$$\mathbb{E}[\theta] = \mathbb{E}[\alpha\theta_1 + (1-\alpha)\theta_2].$$

Utilizando la propiedad de linealidad de la esperanza:

$$\mathbb{E}[\theta] = \alpha\mathbb{E}[\theta_1] + (1-\alpha)\mathbb{E}[\theta_2].$$

Sustituyendo los valores de $\mathbb{E}[\theta_1]$ y $\mathbb{E}[\theta_2]$:

$$\mathbb{E}[\theta] = \alpha\theta + (1-\alpha)\theta.$$

Factorizando $\theta$:

$$\mathbb{E}[\theta] = (\alpha + 1 - \alpha)\theta = \theta.$$

### Conclusiones

Hemos demostrado que el estimador combinado $\theta = \alpha\theta_1 + (1-\alpha)\theta_2$ es insesgado para el parámetro $\theta$, ya que $\mathbb{E}[\theta] = \theta$. UNLP Estimacion insesgada

## 3. Ejercicio 8:Estimación de $\lambda$ para una distribución Poisson

El objetivo de este ejercicio es utilizar el método de máxima verosimilitud para encontrar un estimador del parámetro $\lambda$ de una población Poisson, dado una muestra aleatoria de tamaño $n$.

Dada una muestra aleatoria $x_1, x_2, \ldots, x_n$ de una población Poisson con parámetro $\lambda$, la función de probabilidad de la distribución de Poisson es:

$$P(X = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

La función de verosimilitud $L(\lambda)$ es el producto de estas probabilidades individuales para cada $x_i$ en la muestra:

$$L(\lambda; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(X = x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Para simplificar el cálculo, tomamos el logaritmo natural de la función de verosimilitud:

$$\ln L(\lambda; x_1, x_2, \ldots, x_n) = \ln\left(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)$$

Usando propiedades de los logaritmos, esto se convierte en:

$$\begin{aligned}
\ln L(\lambda; x_1, x_2, \ldots, x_n) &= \sum_{i=1}^{n} \ln\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right) \\
&= \sum_{i=1}^{n} \left(\ln(\lambda^{x_i}) + \ln(e^{-\lambda}) - \ln(x_i!)\right) \\
&= \sum_{i=1}^{n} (x_i \ln\lambda - \lambda - \ln(x_i!))
\end{aligned}$$

Para encontrar el estimador de máxima verosimilitud, derivamos la log-verosimilitud respecto a $\lambda$ y la igualamos a cero:

$$\begin{aligned}
\frac{\partial \ln L}{\partial \lambda} &= \frac{\partial}{\partial \lambda}\left(\sum_{i=1}^{n} (x_i \ln\lambda - \lambda - \ln(x_i!))\right) \\
&= \sum_{i=1}^{n} \left(\frac{x_i}{\lambda} - 1\right) \\
&= \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n
\end{aligned}$$

Igualamos a cero para maximizar:

$$\frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

$$\sum_{i=1}^{n} x_i = n\lambda$$

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$$

## 4. Conclusión

El estimador de máxima verosimilitud para el parámetro $\lambda$ es la media de los datos de la muestra:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

## 5. Ejercicio 9: Intervalo de confianza

Se sabe que la vida en horas de un foco de 100 watts de cierta marca tiene una distribución aproximadamente normal con desviación estándar de 30 horas. Para una muestra al azar de 50 focos, se obtuvo que la vida media fue de 1550 horas. El objetivo es construir un intervalo de confianza del 95% para el verdadero promedio de vida de estos focos. Los datos proporsionados por el ejercicio son los siguientes:

- Desviación estándar de la población, $\sigma = 30$ horas.
- Tamaño de la muestra, $n = 50$ focos.
- Media muestral, $\bar{x} = 1550$ horas.
- Nivel de confianza del 95%, lo que implica un valor de $\alpha = 0.05$.

El intervalo de confianza para la media de una población con varianza conocida se calcula utilizando la fórmula:

$$IC = \bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

Donde:

- $\bar{x}$ es la media muestral,
- $Z_{\alpha/2}$ es el valor crítico de la distribución normal estándar correspondiente a un nivel de confianza del 95%,
- $\sigma$ es la desviación estándar de la población,
- $n$ es el tamaño de la muestra.

El nivel de confianza es del 95%, por lo que $\alpha = 0.05$. Esto implica que el valor crítico $Z_{\alpha/2}$ corresponde a un área de 0.025 en cada cola de la distribución normal estándar. El valor crítico $Z_{\alpha/2}$ es 1.96.

Sustituimos los valores proporcionados en la fórmula del intervalo de confianza:

$$IC = 1550 \pm 1.96 \times \frac{30}{\sqrt{50}}$$

Primero calculamos el error estándar:

$$\frac{30}{\sqrt{50}} = \frac{30}{7.071} \approx 4.243$$

Ahora, multiplicamos por el valor crítico $Z_{\alpha/2} = 1.96$:

$$1.96 \times 4.243 \approx 8.31$$

Por lo tanto, el intervalo de confianza es:

$$IC = 1550 \pm 8.31$$

El intervalo de confianza es:

$$IC = [1550 - 8.31, 1550 + 8.31] = [1541.69, 1558.31]$$

## Conclusión

El intervalo de confianza del 95% para el verdadero promedio de vida de los focos es de 1541.69 horas a 1558.31 horas. Esto significa que estamos un 95% seguros de que el verdadero promedio de vida de los focos de esta marca se encuentra dentro de este rango.

# 6. Ejercicio 10: Prueba de Hipótesis

En este ejercicio se dan los siguientes datos sobre el número de cigarrillos fumados al día por un grupo de 10 fumadores:

$$5, 10, 3, 4, 5, 8, 20, 4, 1, 10$$

Se busca realizar una prueba de hipótesis para verificar si la media del número de cigarrillos al día es menor que 10. La hipótesis nula y alternativa se plantean de la siguiente manera:

$$H_0 : \mu = 10 \quad \text{vs} \quad H_1 : \mu < 10$$

Donde:

- $H_0$: La media de cigarrillos al día es 10.
- $H_1$: La media de cigarrillos al día es menor que 10.

El nivel de significancia es del 95%, es decir, $\alpha = 0.05$.

Se calcula la media muestral $\bar{x}$:

$$\bar{x} = \frac{5 + 10 + 3 + 4 + 5 + 8 + 20 + 4 + 1 + 10}{10} = \frac{70}{10} = 7$$

La media muestral es $\bar{x} = 7$.

Sabemos que la fórmula para el valor de $z$ es:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Donde:

- $\bar{x} = 7$ (media muestral),
- $\mu_0 = 10$ (media bajo la hipótesis nula),
- $\sigma = 1.2$ (desviación estándar de la población),
- $n = 10$ (tamaño de la muestra).

Sustituyendo los valores:

$$z = \frac{7 - 10}{\frac{1.2}{\sqrt{10}}} = \frac{-3}{\frac{1.2}{3.162}} = \frac{-3}{0.379} \approx -7.92$$

El valor calculado de $z$ es $-7.92$.

Para un nivel de significancia $\alpha = 0.05$, el valor crítico de $z$ es:

$$z_\alpha = -1.645$$

Se compara el valor calculado de $z$ con el valor crítico. Si el valor calculado de $z$ es menor que $z_\alpha$, rechazamos $H_0$. En este caso:

$$z = -7.92 \quad \text{y} \quad z_\alpha = -1.645$$

Como se obtienen los siguientes valores $-7.92 < -1.645$, rechazamos la hipótesis nula .

## 6.1. Conclusión

Se rechaza la hipótesis nula $H_0$ al nivel de significancia del 95%. Esto significa que hay suficiente evidencia para concluir que la media del número de cigarros fumados al día es menor que 10.

# References