Estadística

Aylén Avila

2022

Para construir sistemas de inteligencia artificial (IA), es necesario el desarrollo de estrategias para la toma de decisiones.

Estas estrategias requieren de:

- Estimación de parámetros,
- evaluación de alternativas,
- elección entre posibilidades…

Cada tarea es un problema estadístico en su esencia, y es por ello que las herramientas estadísticas son útiles para construir sistemas de IA.

La ciencia de los datos, el aprendizaje automático y la IA se basan en la estadística para proporcionar información valiosa para operar de manera más eficiente y eficaz.

Intervalo para la media poblacional con desvío poblacional conocido

Probabilidad y Estadística
IA 3.1, FB12 y FB22

Aylén Avila
2022

Estimación de parámetros

Estimación de parámetros

θ : parámetro a estimar (μ_X, p, $\sigma^2{}_X$)

$\hat{\theta}$: estimador del parámetro ($\overline{X}$, h(A), $S^2{}_x$)

$$E(\hat{\theta}) = \theta \text{ y } V(\hat{\theta}) \xrightarrow[n \to +\infty]{} 0$$

Estimación de parámetros

intervalo de confianza

Se tendrá un par de valores a y b (con $a < b$), tal que el intervalo $(a; b)$ cubra al verdadero valor del parámetro, con un determinado nivel de confianza $(1 - \alpha)$.

Las principales ventajas de la estimación por parámetros son la precisión y la confianza.

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01

Definir el parámetro a estimar

$$\theta = \mu_X$$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01

Definir el parámetro a estimar

$$\theta = \mu_X$$

02

Elegir un buen estimador

$$\hat{\theta} = \hat{\mu}_X = \bar{X}$$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01

Definir el parámetro a estimar

$\theta = \mu_X$

02

Elegir un buen estimador

$\hat{\theta} = \hat{\mu}_X = \bar{X}$

03

Fijar el grado de confianza

$1 - \alpha = 0.95$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01

Definir el parámetro a estimar

$\theta = \mu_X$

02

Elegir un buen estimador

$\hat{\theta} = \hat{\mu}_X = \bar{X}$

03

Fijar el grado de confianza

$1 - \alpha = 0.95$

04

Definir una nueva variable

$Z = \bar{X} - \mu_X / (\sigma_X / \sqrt{n})$

$\sim N(0; 1)$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01
Definir el parámetro a estimar

$\theta = \mu_X$

02
Elegir un buen estimador

$\hat{\theta} = \hat{\mu}_X = \bar{X}$

03
Fijar el grado de confianza

$1 - \alpha = 0.95$

04
Definir una nueva variable

$Z = \bar{X} - \mu_X / (\sigma_X / \sqrt{n})$

$\sim N(0; 1)$

05
Elegir dos valores tales que

$P\,(y_{\alpha/2} \leq Y \leq y_{1-\alpha/2}) = 1 - \alpha$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01
Definir el parámetro a estimar

$\theta = \mu_X$

02
Elegir un buen estimador

$\hat{\theta} = \hat{\mu}_X = \bar{X}$

03
Fijar el grado de confianza

$1 - \alpha = 0.95$

04
Definir una nueva variable

$Z = \bar{X} - \mu_X / (\sigma_X / \sqrt{n})$

$\sim N(0; 1)$

05
Elegir dos valores tales que

$P\,(y_{\alpha/2} \leq Y \leq y_{1\text{-}\alpha/2}) = 1 - \alpha$

$$\text{IC} = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma_X}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma_X}{\sqrt{n}}\right)$$

Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

01

Definir el parámetro a estimar

$\theta = \mu_X$

02

Elegir un buen estimador

$\hat{\theta} = \hat{\mu}_X = \bar{X}$

03

Fijar el grado de confianza

$1 - \alpha = 0.95$

04

Definir una nueva variable

$Z = \bar{X} - \mu_X / (\sigma_X / \sqrt{n})$

$\sim N(0; 1)$

05

Elegir dos valores tales que

$P\,(y_{\alpha/2} \leq Y \leq y_{1-\alpha/2}) = 1 - \alpha$

$$\text{IC} = \left(\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}\right)$$

06

Obtener una muestra, a partir de la cual se calcula el intervalo de confianza

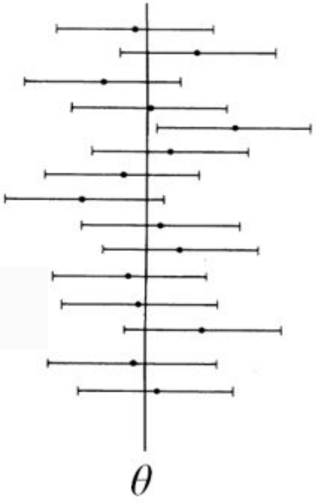
Construir un intervalo de confianza al 95% para el contenido de Cinc en ciertos materiales (σ = 5 ppm, n=100).

Tomamos la muestra o accedemos a la base de datos con la información necesaria y nos vamos a R

https://github.com/ImoPupato/Estadistica-Ejemplo-1

Con una confianza del 95% el intervalo (148.64 ppm; 150.60 ppm) contiene al verdadero valor promedio de la concentración de Cinc en la población de pacientes con cierta enfermedad.

Con una confianza del 95% el intervalo (148.64 ppm; 150.60 ppm) contiene al verdadero valor promedio de la concentración de Cinc en la población de pacientes con cierta enfermedad.



Intepretación

Si se extrajeran reiteradas muestras, todas del mismo tamaño n, y para cada una de ellas se calculara el correspondiente intervalo de $(1-\alpha)\cdot 100$ % de confianza, esperaríamos que el $(1-\alpha)\cdot 100$ % de los intervalos **cubran** al verdadero valor del parámetro y el $\alpha\cdot 100$ % restante, no

¿Consultas?

avilaylen.w@gmail.com

https://github.com/ImoPupato/Estadistica-Ejemplo-1

Estimación de parámetros

intervalo de confianza

X,variable aleatoria tal que $X \sim N(\mu_X; \sigma_X)$ con σ_X conocido

Estimar mediante un intervalo de $(1 - \alpha) \cdot 100$ % de confianza para el promedio poblacional

1. Definir el parámetro a estimar: $\theta = \mu_X$
2. Seleccionar un buen estimador: $\hat{\theta} = \hat{\mu}_X = \bar{X}$
3. Fijar el grado de confianza: $1 - \alpha$.
4. Definir una variable Y que contenga al parámetro a estimar, al estimador y cuya distribución de probabilidad sea conocida ($f(y)$).

Como $X \sim N(\mu_X; \sigma_X)$, sabemos que $\bar{X} \sim N(\mu_X; \sigma_X/\sqrt{n})$.

Definimos entonces $Z = \bar{X} - \mu_X /(\sigma_X/\sqrt{n}) \sim N(0; 1)$ que cumple con los requisitos exigidos para Y .

5. Sobre el eje de variación de Y , elegir dos valores tales que $P\ (y_{\alpha/2} \leq Y \leq y_{1-\alpha/2}) = 1 - \alpha$
6. Obtener una muestra, a partir de la cual se calcula el intervalo de confianza (a;b)
7. Concluir.