Vídeos docentes sobre

Probabilidad y Teoría de la Decisión

Contraste de hipótesis para una distribución normal

F. J. Díez Vegas
Dpto. Inteligencia Artificial. UNED

fjdiez@dia.uned.es
www.ia.uned.es/~fjdiez

Segundo ejemplo: distribución poblacional normal

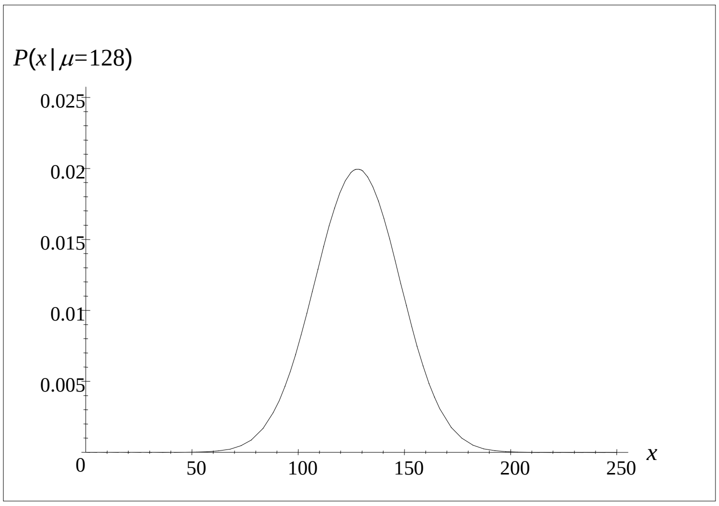
- **Objetivo:** estudiar la relación entre la vitamina X y el síndrome Y
- **Hecho:** la concentración de X en la sangre de personas sanas es 128 μg/cm^3 (desviación estándar 20 μg/cm^3)
- **Datos:** análisis de sangre de 25 pacientes con el síndrome Y
  El promedio de la concentración de X es 117 μg/cm^3
- Pregunta: ¿Hay asociación entre el déficit de vitamina X y la presencia del síndrome Y?

(Sólo queremos averiguar si hay **asociación**. No pretendemos demostrar **relación causal.**)

Contraste de hipótesis
(para este ejemplo)

1. Hipótesis experimental: H_E = "μ < 128"
2. Hipótesis nula: H_0 = "μ ≥ 128"
3. Tipo de distribución poblacional: normal (gaussiana)
4. Distribución poblacional, en el límite de la hipótesis nula:

$$P(x \mid \mu = 128) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{20\sqrt{2\pi}}\, e^{-\frac{(x-128)^2}{2\times 20^2}}$$



Distribución (poblacional) del nivel de X en sangre, X

para el valor límite de la hipótesis nula

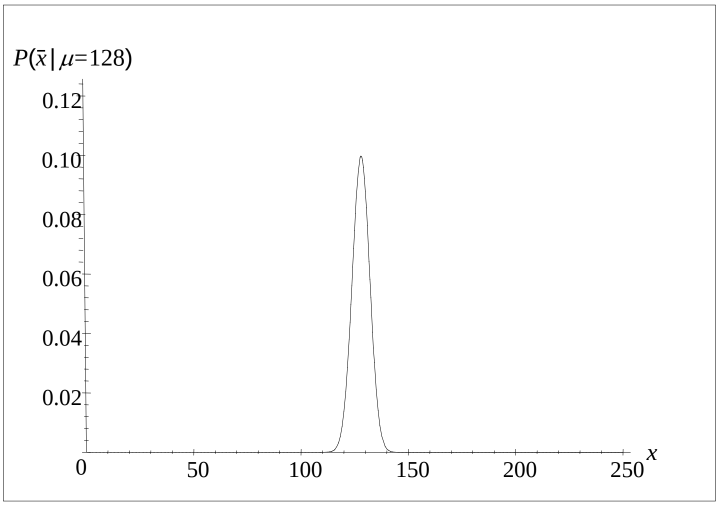
5. Tamaño de la muestra: $n = 25$

6. Distribución muestral para el estadístico $\overline{X}$ (promedio del nivel de X): distribución normal (gaussiana), con

$$\mu' = \mu = 128 \qquad \sigma' = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4$$

es decir,

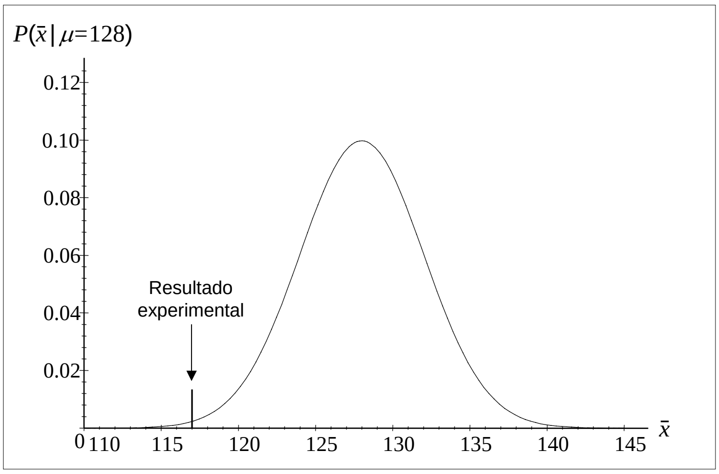
$$P(\bar{x} \mid \mu = 128) = \frac{1}{\sigma'\sqrt{2\pi}}\, e^{-\frac{(\bar{x}-\mu')^2}{2\sigma'^2}} = \frac{1}{4\sqrt{2\pi}}\, e^{-\frac{(\bar{x}-128)^2}{2\times 4^2}}$$



Distribución (muestral) de $\overline{X}$, el promedio de X

para el valor límite de la hipótesis nula

Vista ampliada



Distribución (muestral) de $\bar{X}$, el promedio de X

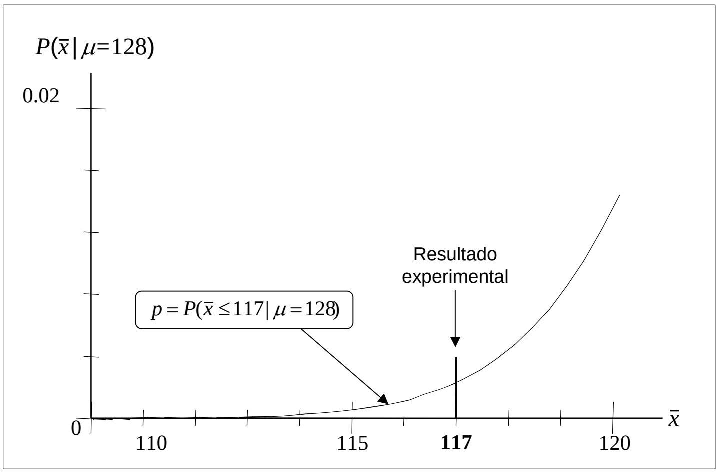
para el valor límite de la hipótesis nula

7. Resultado experimental: $\bar{x}_{exp} = 117$

8. Valor p

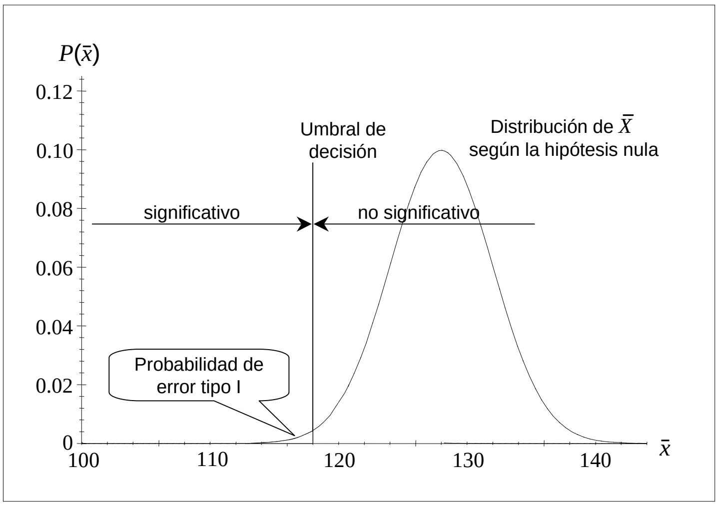
$$p = P(\bar{x} \leq 117 \,|\, \mu = 128) = \frac{1}{4\sqrt{2\pi}} \int_{-\infty}^{117} e^{-\frac{(\bar{x}-128)^2}{2\times 4^2}}\, d\bar{x} = 0'003$$

Vista ampliada

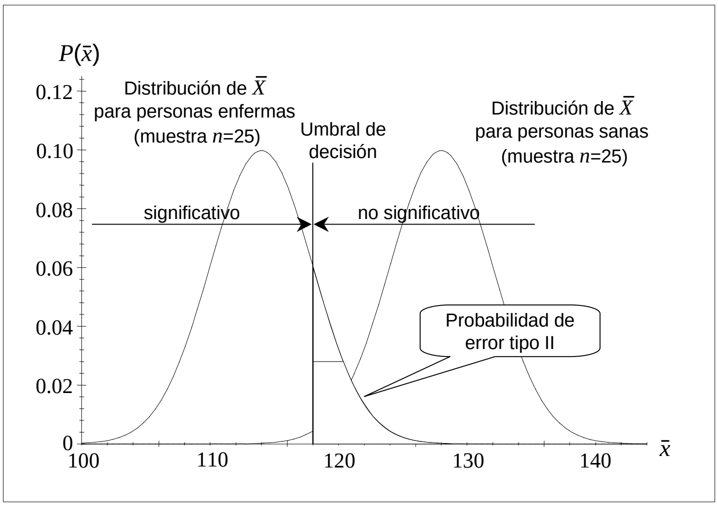


p = área bajo la curva

Probabilidad de error tipo I (α)

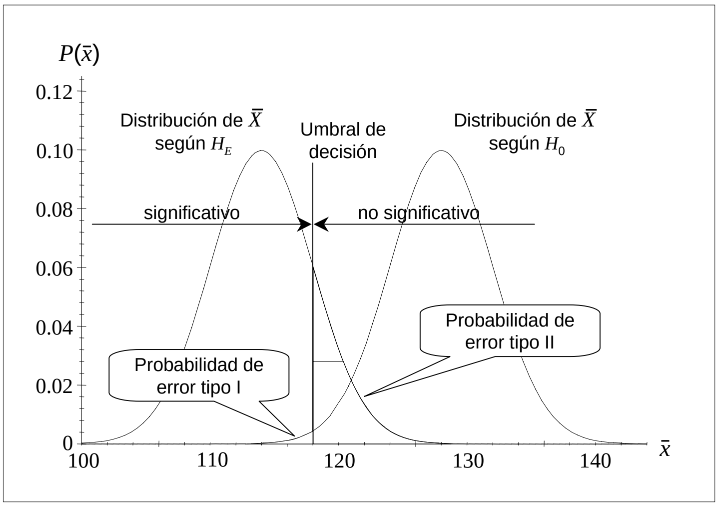


Probabilidad de error tipo II (β)

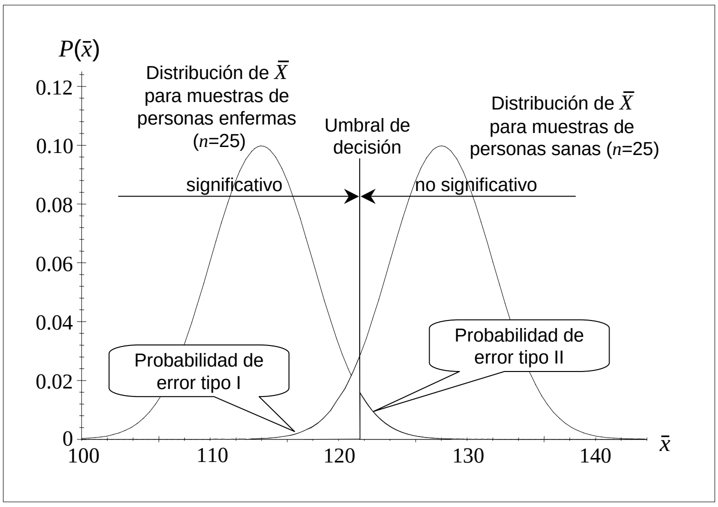


<u>Nota:</u> Sólo podríamos calcular la probabilidad de error tipo II si conociéramos la distribución de $\bar{X}$ para personas enfermas (que es precisamente lo que no conocemos)

Ambos tipos de error son posibles



Compromiso entre ambos tipos de error



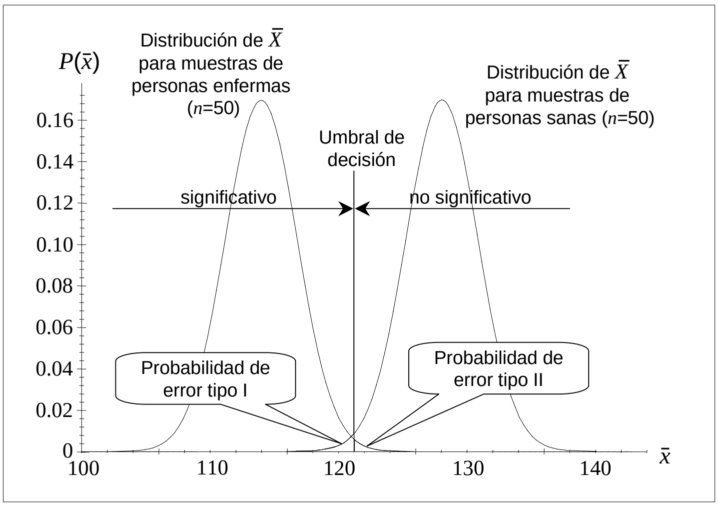
Al intentar eliminar el error tipo II aumenta la probabilidad de error tipo I

Tipos de error

	H_E cierta (H_0 falsa)	H_E falsa (H_0 cierta)
Significativo ($p < \alpha$): concluimos H_E	Acierto: nuevo conocimiento	Error tipo I (α): conclusión errónea
No significativo ($p > \alpha$): no concluimos <u>nada</u>	Error tipo II (β): no se concluye nada	Acierto: no se concluye nada

¿Es posible disminuir a la vez
la probabilidad de error tipo I
y la probabilidad de error tipo II?

Solución: aumentar el tamaño de la muestra



Así conseguimos dismunuir simultáneamente las probabilidades de los dos tipos de error