

Pruebas de significancia estadística: z-test y t-test

Introducción

- ▶ Durante los experimentos generalmente estamos comparando nuestros resultados contra algún valor base
- ▶ Una vez que logramos superar dicho valor base la pregunta que debe surgir es:
¿La mejora alcanzada es significativa?
- ▶ Donde por significativa vamos a entender que la mejora es consecuencia de los métodos propuestos y no del azar
 - GARANTÍA DE CALIDAD EN LAS PUBLICACIONES

Prueba de significancia estadística

▶ Componentes básicos:

- Valor de nuestro experimento **pA**
- Valor base **pB**
- Hipótesis nula H_0 (No existe diferencia entre pA y pB)
- Cálculo estadístico
 - Prueba estadística: Z-test, T-test, Wilcoxon, McNemar, ...
 - Valor de significancia p (comúnmente con un $\alpha = 0.05$, lo que significa que tenemos un 95% de seguridad de no cometer un error tipo I – rechazar incorrectamente H_0)

▶ Resultado

- Si $|pA - pB| > \text{error_estándar} * \text{seguridad_definida}$ entonces rechazamos H_0

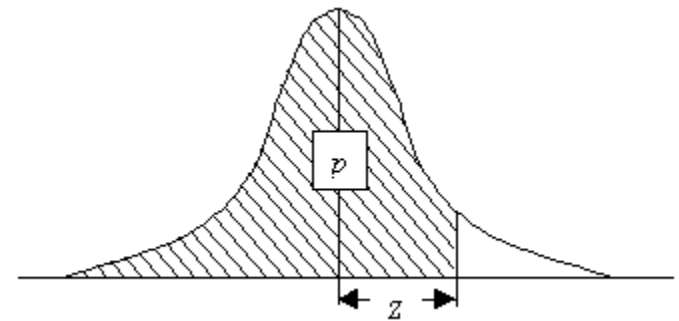
Z-test para comparar dos proporciones

- ▶ Error estándar: $error_estándar = \sqrt{\frac{2a(1-a)}{n}}$

donde $a = \frac{pA + pB}{2}$ y n es el número de elementos en el conjunto de prueba

- ▶ Valor de significancia p (distribución normal)

Seguridad	α	Seguridad definida $Z_{1-\alpha/2}$
95%	0.05	1.960
99%	0.01	2.576



Ejemplo

- ▶ Validación de respuestas (medida-F)
 - $n=2286$
 - $pA = 0.75$
 - $pB = 0.61$
 - $H_0 =$ No existe diferencia entre pA y pB

 - $|pA-pB| = 0.14$
 - $Error_est\acute{a}ndar = 0.01379$
 - $Seguridad_definida = 1.960$

 - Como $0.14 > 0.027$ entonces podemos rechazar H_0

¿Qué prueba estadística es mejor?

Dietterich, T. G., (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 10 (7) 1895-1924

