

# COMPARACIÓN DE 2 SISTEMAS INFORMÁTICOS (usando mediciones)

Objetivo: Comparar una métrica (de prestaciones) de 2 sistemas

Problema: Las medidas (u observaciones) están distorsionadas por ruido

Se necesitan técnicas para determinar

si las diferencias observadas en las métricas de prestaciones de dos sistemas

- ┌ Son debidas al ruido de las medidas
- └ Son debidas a diferencias significativas entre los sistemas

Las técnicas se basan en la “Prueba de media cero”

Consiste en usar intervalos de confianza para comprobar si la media de un conjunto de observaciones es significativamente diferente de cero

Limitaciones de estas técnicas

- 1) Sólo sirven para comparar dos sistemas
- 2) Las cargas de trabajo deben ser las mismas o muy similares

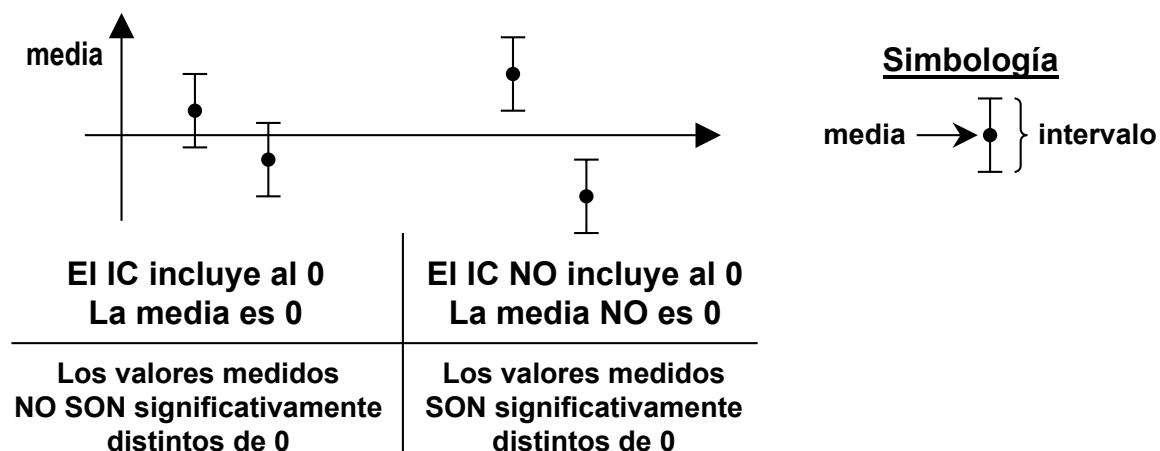
Ejemplos

- Comparar las prestaciones de un servidor antes-y-después de ...
- Comparar las prestaciones de dos servidores usando un conjunto de benchmarks



## COMPARACIÓN DE SISTEMAS INFORMÁTICOS Intervalo para comprobar si la media es cero

Consiste en calcular un intervalo de confianza para la media de las observaciones y comprobar si el intervalo incluye al cero (diferencias)



Para comprobar si la media de unas observaciones es igual a un valor dado A  
Calcular un intervalo de confianza para la media de las observaciones  
y comprobar si el intervalo incluye al valor A



# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Intervalo para Observaciones Emparejadas

### LAS OBSERVACIONES ESTÁN EMPAREJADAS

Si se realizan n experimentos

- Cada uno de ellos, con 2 sistemas A y B
- Y hay correspondencia uno-a-uno  $\longrightarrow$   $\left\{ \begin{array}{l} \text{Entre la } i\text{-ésima prueba del sistema A} \\ \text{Y la } i\text{-ésima prueba del sistema B} \end{array} \right.$

### LAS OBSERVACIONES ESTÁN DESEMPAREJADAS

Si no hay una correspondencia entre las dos muestras

### PASOS PARA COMPROBAR MUESTRAS DE OBSERVACIONES EMPAREJADAS

- 1 Obtener la muestra diferencia de las dos muestras disponibles (calculando las diferencias entre las observaciones de cada pareja)
- 2 Construir un intervalo de confianza para la muestra diferencia
- 3 Si el intervalo incluye al 0, ENTONCES las prestaciones de los sistemas NO SON significativamente diferentes

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Intervalo para Observaciones Desemparejadas (1)

Se dispone de 2 muestras de  $\left\{ \begin{array}{l} \text{Na observaciones del sistema A} \\ \text{Nb observaciones del sistema B} \end{array} \right.$

No hay correspondencia alguna entre las i-ésimas observaciones en las dos muestras

### PASOS PARA COMPROBAR MUESTRAS DE OBSERVACIONES DESEMPAREJADAS

- 1 Calcular las medias de las muestras

$$\bar{x}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} x_{ia} \quad \bar{x}_b = \frac{1}{n_b} \sum_{i=1}^{n_b} x_{ib}$$

- 2 Calcular las desviaciones estándar de las muestras

$$s_a = \sqrt{\frac{\left( \sum_{i=1}^{n_a} x_{ia}^2 \right) - n_a \bar{x}_a^2}{n_a - 1}} \quad s_b = \sqrt{\frac{\left( \sum_{i=1}^{n_b} x_{ib}^2 \right) - n_b \bar{x}_b^2}{n_b - 1}}$$

- 3 Calcular la diferencia de las medias

$$\bar{x}_a - \bar{x}_b$$

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Intervalo para Observaciones Desemparejadas (2)

### PASOS PARA COMPROBAR MUESTRAS DE OBSERVACIONES DESEMPAREJADAS

- 4 Calcular la desviación estándar de la diferencia de las medias

$$s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

- 5 Calcular el número efectivo de grados de libertad

$$v = \frac{\left( \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b} \right)^2}{\frac{1}{n_a + 1} \left( \frac{s_a^2}{n_a} \right)^2 + \frac{1}{n_b + 1} \left( \frac{s_b^2}{n_b} \right)^2} - 2$$

- 6 Calcular el intervalo de confianza de la diferencia de las medias

$$(\bar{x}_a - \bar{x}_b) \pm t_{[1-\alpha/2; v]} s$$

$$t_{[1-\alpha/2; v]}$$

Es el cuantil  $(1-\alpha/2)$  de una distribución t de Student con  $v$  grados de libertad

Si el intervalo de confianza ...

Incluye al 0, la diferencia no es significativa con un nivel de confianza del  $100(1-\alpha)\%$

NO incluye al 0, el signo de la diferencia de las medias indica el sistema que es mejor

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Técnica de contraste de hipótesis: introducción

**Se parte de una hipótesis**  $\begin{cases} H_0: \text{Hipótesis nula (que se desea contrastar)} \\ H_1: \text{Hipótesis alternativa} \end{cases}$

**Contraste de hipótesis:** Proceso mediante el que se decide cuál de las hipótesis es correcta y se acepta; la otra se rechaza

**Se basa en:** Un estadístico  $E$  que se calcula a partir de una muestra

### Idea en que se fundamenta el contraste de hipótesis

Si se obtiene un valor muy improbable del estadístico  $E$  cuando la hipótesis es cierta  
**ENTONCES** La muestra usada es muy rara / La hipótesis nula es falsa

### Los rangos de valores del estadístico

Improbables  $\xrightarrow{\text{DEFINEN}}$  La región de rechazo o crítica (RR ó RC)  
 Probables  $\xrightarrow{\text{DEFINEN}}$  La región de aceptación (RA)

**Prueba**  $\begin{cases} \text{Rechazar } H_0 \text{ si } E \in \text{RR o RC; Aceptar } H_1 \\ \text{Aceptar } H_0 \text{ si } E \in \text{RA; Rechazar } H_1 \end{cases}$

### Errores cometidos

	Acepto	Rechazo
$H_0$ es cierta	Bien	Error Tipo I
$H_0$ es falsa	Error Tipo II	Bien

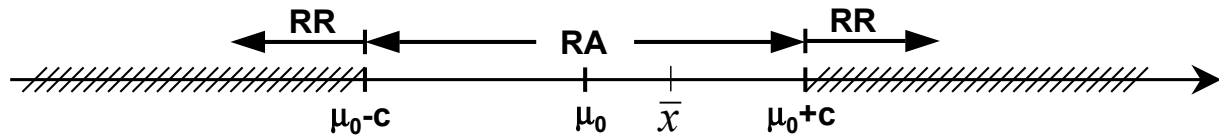
$\alpha$ : probabilidad de que ocurra un error tipo I

$\beta$ : probabilidad de que ocurra un error tipo II

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Contraste de hipótesis de (=)

**Para contrastar la hipótesis**  $\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu\neq\mu_0 \end{cases}$  La región de aceptación es un intervalo alrededor de  $\mu_0$



Para que el error de tipo I tenga una probabilidad  $\alpha$  de producirse cuando  $H_0$  es cierta (cuando realmente  $\mu=\mu_0$ ) debe verificarse que:  $\Pr\{\bar{x} \in RR\} = \Pr\{|\bar{x} - \mu_0| > c\} = \alpha$

SI  $X \rightarrow N(\mu, \sigma)$  ENTONCES  $\frac{\bar{x} - \mu}{s / \sqrt{n}} \rightarrow t_{n-1}$   
 (distribución de la población) (distribución del estadístico de la media muestral)

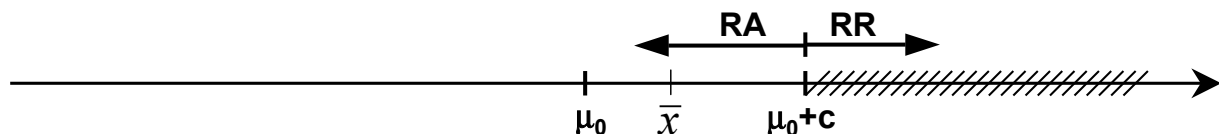
$$\Pr\{|\bar{x} - \mu_0| > c\} = \Pr\{|t_{n-1}| > c \frac{\sqrt{n}}{s}\} = \alpha \longrightarrow c = t_{1-\alpha/2; n-1} \frac{s}{\sqrt{n}}$$

**Prueba**  $\begin{cases} \text{Aceptar } H_0 (\mu=\mu_0) \text{ si } |\bar{x} - \mu_0| \leq t_{1-\alpha/2; n-1} s / \sqrt{n} \\ \text{Rechazar } H_0 (\mu\neq\mu_0) \text{ si } |\bar{x} - \mu_0| > t_{1-\alpha/2; n-1} s / \sqrt{n} \end{cases}$

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Contraste de hipótesis de ( $\leq$ )

**Para contrastar la hipótesis**  $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$  La región de aceptación y la de rechazo son:



Para que el error de tipo I tenga una probabilidad  $\alpha$  de producirse cuando  $H_0$  es cierta (cuando realmente  $\mu \leq \mu_0$ ) debe verificarse que:  $\Pr\{\bar{x} \in RR\} = \Pr\{\bar{x} - \mu_0 > c\} = \alpha$

SI  $X \rightarrow N(\mu, \sigma)$  ENTONCES  $\frac{\bar{x} - \mu}{s / \sqrt{n}} \rightarrow t_{n-1}$   
 (distribución de la población) (distribución del estadístico de la media muestral)

$$\Pr\{\bar{x} - \mu_0 > c\} = \Pr\{t_{n-1} > c \frac{\sqrt{n}}{s}\} = \alpha \longrightarrow c = t_{1-\alpha; n-1} \frac{s}{\sqrt{n}}$$

**Prueba**  $\begin{cases} \text{Aceptar } H_0 (\mu \leq \mu_0) \text{ si } \bar{x} - \mu_0 \leq t_{1-\alpha; n-1} s / \sqrt{n} \\ \text{Rechazar } H_0 (\mu > \mu_0) \text{ si } \bar{x} - \mu_0 > t_{1-\alpha; n-1} s / \sqrt{n} \end{cases}$

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Concepto de p-valor de un Contraste de hipótesis

**El resultado de una prueba de hipótesis =  $F(\alpha)$**

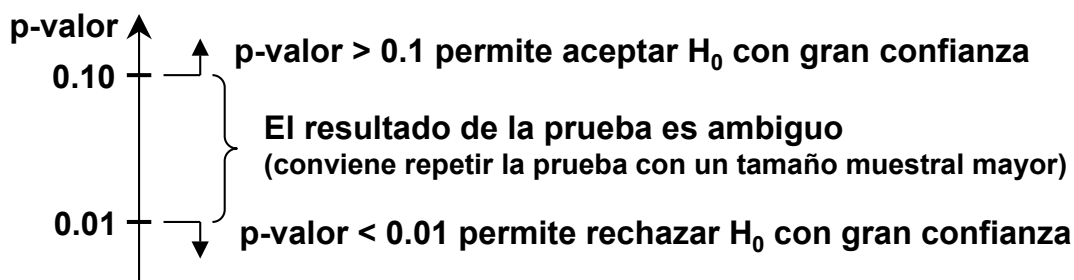
En algunos problemas  $\alpha$  está perfectamente definido

En otros problemas NO está claramente definido el valor de  $\alpha$  que hay que usar

**Solución:** Adjuntar a la decisión de rechazar  $H_0$  el riesgo que ello comporta

**p-valor de una prueba:** es el mínimo  $\alpha$  con el que puede rechazarse  $H_0$   
(para la muestra utilizada)

El contexto del problema permite valorar si el p-valor permite con claridad  
Aceptar/Rechazar  $H_0$  ... PERO EN GENERAL ...



# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Contraste de hipótesis sobre la media de una población normal

Se aplica a la muestra diferencia de **observaciones emparejadas**

Es la prueba para aceptar / rechazar la hipótesis de que la media de la población  $\mu$  es  $= \geq \leq \mu_0$

	$\sigma$ conocida	$\sigma$ desconocida	
$H_0: \mu = \mu_0$ si	$ \bar{x} - \mu_0  \leq z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$	$ \bar{x} - \mu_0  \leq t_{1-\alpha/2; n-1} \frac{S}{\sqrt{n}}$	
$H_0: \mu \leq \mu_0$ si	$\bar{x} - \mu_0 \leq z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$	$\bar{x} - \mu_0 \leq t_{1-\alpha; n-1} \frac{S}{\sqrt{n}}$	
$H_0: \mu \geq \mu_0$ si	$\bar{x} - \mu_0 \geq z_{\alpha} \frac{\sigma}{\sqrt{n}}$	$\bar{x} - \mu_0 \geq t_{\alpha; n-1} \frac{S}{\sqrt{n}}$	

Si la muestra es grande ( $n > 30$ ) no hace falta que la distribución de la población sea  $N(\mu, \sigma)$

Para  $n$  grande la distribución  $t \rightarrow N$ : Calcular los cuantiles  $z$  de todas las pruebas con la  $N$

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Contraste de hipótesis sobre la diferencia de medias de dos poblaciones normales independientes

Se aplica a dos muestras de **observaciones DESemparejadas**

Es la prueba para aceptar / rechazar la hipótesis de que la media de la población  $\mu_a$  es  $= \geq \leq \mu_b$

	$\sigma_a$ y $\sigma_b$ conocidas	$\sigma_a$ y $\sigma_b$ desconocidas =	$\sigma_a$ y $\sigma_b$ desconocidas $\neq$
$H_0: \mu_a = \mu_b$ si	$ \bar{x}_a - \bar{x}_b  \leq z_{1-\alpha/2} \sigma_m$	$ \bar{x}_a - \bar{x}_b  \leq t_{1-\alpha/2; n_a+n_b-2} S_{eq}$	$ \bar{x}_a - \bar{x}_b  \leq t_{1-\alpha/2; \nu} S_m$
$H_0: \mu_a \leq \mu_b$ si	$\bar{x}_a - \bar{x}_b \leq z_{1-\alpha} \sigma_m$	$\bar{x}_a - \bar{x}_b \leq t_{1-\alpha; n_a+n_b-2} S_{eq}$	$\bar{x}_a - \bar{x}_b \leq t_{1-\alpha; \nu} S_m$
$H_0: \mu_a \geq \mu_b$ si	$\bar{x}_a - \bar{x}_b \geq z_{\alpha} \sigma_m$	$\bar{x}_a - \bar{x}_b \geq t_{\alpha; n_a+n_b-2} S_{eq}$	$\bar{x}_a - \bar{x}_b \geq t_{\alpha; \nu} S_m$

$$\sigma_m = \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}$$

$$S_{eq} = \sqrt{\frac{(n_a-1)s_a^2 + (n_b-1)s_b^2}{n_a+n_b-2}} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$$

$$S_m = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

Si  $(n_a \approx n_b) > 30$  se asegura la normalidad de las medias muestrales  
 Usar siempre la distribución N para calcular cuantiles z  
 Si  $\sigma_a$  y  $\sigma_b$  son conocidas usar  $\sigma_m$   
 Si  $\sigma_a$  y  $\sigma_b$  son desconocidas usar  $s_m$

$$\nu = \frac{\left( \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b} \right)^2}{\frac{1}{n_a-1} \left( \frac{s_a^2}{n_a} \right)^2 + \frac{1}{n_b-1} \left( \frac{s_b^2}{n_b} \right)^2}$$



Área de Arquitectura y Tecnología de Computadores  
 Dpto. de Informática de la Universidad de Oviedo

Sistemas Informáticos Físicos

11

# COMPARACIÓN DE SISTEMAS INFORMÁTICOS

## Herramientas y funciones Excel para contraste de hipótesis

Las herramientas de Excel trabajan siempre con dos muestras

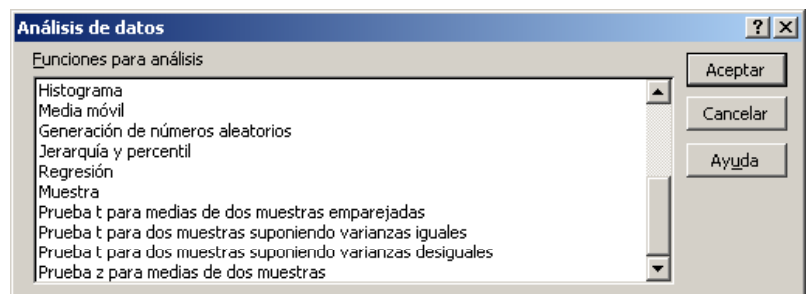
¿Son grandes (>30 obs)? → Usar Prueba z

¿Son pequeñas (<30 obs)? → Usar Pruebas t para → { Emparejadas  
 Desemparejadas } { Varianzas =  
 Varianzas  $\neq$

Las funciones de Excel son

PRUEBA.Z(datos;media;desviación)

PRUEBA.T(datosA;datosB;colas;tipo)



Observar los resultados de la prueba:

- Comparar: Estadístico t con Valor crítico de t (dos colas) para aceptar/rechazar la hipótesis
- Evaluar: P(T<=t) dos colas para valorar el riesgo del rechazar la hipótesis

NOTA: Debería denominarse P(|T|>t) y es el p-valor



Área de Arquitectura y Tecnología de Computadores  
 Dpto. de Informática de la Universidad de Oviedo

Sistemas Informáticos Físicos

12