



# Tests de permutaciones y tests de aleatorización

**Mètodes no paramètrics i de remostratge**

Grau en Estadística UB-UPC

Jordi Ocaña Rebull

**Departament d'Estadística**

Facultat de Biologia








## Qüestions legals

---







Donat el caràcter i la finalitat exclusivament docent i eminentment il·lustrativa de les explicacions a classe d'aquesta presentació, l'autor s'acull a l'article 32 de la Llei de propietat intel·lectual vigent respecte de l'ús parcial d'obres alienes com ara imatges, gràfics o altre material contingudes en les diferents diapositives

## Contenido:

---

-  Tests de permutaciones como test condicional
-  Tests de aleatorización
-  Tests de permutaciones y tests de aleatorización aproximados o de Monte Carlo
-  Estimación del p-valor
-  Determinación del número de permutaciones aleatorias
-  Algunos tests concretos
-  Tests de permutaciones e intervalos de confianza

## Referencias con valor histórico

-  Eden, T. and Yates, F. (1933). On the validity of Fisher's  $z$  test when applied to an actual example of non-normal data. *Journal of Agricultural Science* **23**, 6-16
-  Fisher, R.A. (1935). *The Design of Experiments*. (9th Edition, Macmillan. [ISBN 0028446909](#), 1971)
-  Pitman, E.J.G. (1937). Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society, Ser. B* **4**, 119-130
-  Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220
-  Marriott, F.H.C. (1979). Barnard's Monte Carlo tests: how many simulations? *Appl. Stat.* **28**, 75-77
-  Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Statist.* **28**, 181–187

## Distribución muestral condicionada a la muestra

- Recordemos que estadístico ordinal es suficiente (y completo) para  $F$  (y  $g(F)$ )
- En concreto, si indicamos  $\{\mathbf{X}\} = (X_{(1)}, \dots, X_{(n)})$ ,

$$p(\mathbf{x} | \{\mathbf{x}\}) = \begin{cases} \frac{1}{n!} & \text{si } \text{ord}(\mathbf{x}) = \{\mathbf{x}\} \\ 0 & \text{en caso contrario} \end{cases}$$

- ¿Cómo se puede aprovechar este hecho en el contraste de hipótesis?: condicionar a los datos de la muestra, conduce a “tests condicionales a la muestra”

## Tests de permutaciones planteamiento

📄  $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$  muestra observada

📄 Estadístico  $t$  mide alejamiento de  $H_0$ .

Rechazamos  $H_0$  si

$$t(\tilde{\mathbf{x}}) \geq c_\alpha, \quad \Pr_{H_0} \{t(\mathbf{X}) \geq c_\alpha\} \leq \alpha$$

o, equivalentemente, rechazaremos  $H_0$  si

$$\text{p-valor} = \Pr_{H_0} \{t(\mathbf{X}) \geq t(\tilde{\mathbf{x}})\} \leq \alpha$$

📄 Si  $H_0$  implica *distribución única*  $F$  para toda la muestra, forma “libre de la distribución” de determinar  $c_\alpha$  ó  $\Pr_{H_0} \{t(\mathbf{X}) \geq t(\tilde{\mathbf{x}})\}$  es condicionar al estadístico ordinal

## Test de permutaciones como test condicionado a $\{\tilde{\mathbf{x}}\}$

Enumeración de permutaciones de los datos para las que valor de  $t$  supera el observado

$$P_{H_0} [t(\mathbf{X}) \geq t(\tilde{\mathbf{x}}) | \{\tilde{\mathbf{x}}\}] = \frac{\#\{\mathbf{x}_p : t(\mathbf{x}_p) \geq t(\tilde{\mathbf{x}})\}}{n!}$$

p-valor o región crítica condicionados a la obtención previa de la muestra, no tenemos “la tabla” del test a priori  $\Rightarrow$  algunas objeciones filosóficas (poco justificadas)

Bajo muestreo aleatorio simple, extensión real del test compatible con valor  $\alpha$  nominal

# Tests de aleatorización

---

 Otro enfoque conceptual, pero con las mismas consecuencias a nivel práctico:

- Muestreo aleatorio es infrecuente en la investigación experimental
- Asignación aleatoria de tratamientos (sobre muestra, normalmente, no aleatoria) permite testar la validez de hipótesis, validez restringida a la muestra
- Inferencia sobre la población basada en argumentos no estadísticos



# Tests de permutaciones/ aleatorización en la práctica

---

- ☞ Verdadero problema práctico cuando se plantea un test de permutaciones: evaluar el estadístico de test para **todas** las permutaciones
- ☞ Insoluble en los años 30 del siglo XX.  
Realizable actualmente para tamaños muestrales no muy grandes, pero puede llegar a ser un problema de computación muy intensiva

## Ejemplo: test de permutaciones de comparación de dos grupos indep.

- 📄 Muestra procedente de dos tratamientos,  $n_1$  y  $n_2$  observaciones respect., total  $n = n_1 + n_2$   
$$\mathbf{y} = (\mathbf{y}_1; \mathbf{y}_2) = (y_{11}, \dots, y_{1n_1}; y_{21}, \dots, y_{2n_2})$$
- 📄  $H_0$ : provienen de la misma distribución. (Si interpretación de diferencias o de igualdad en términos de parámetro de localización será necesario suponer idéntica dispersión)
- 📄  $H_0$ :  $F$  común a las dos muestras: todas las permutaciones de los datos son equiprobables

## El mismo problema, punto de vista de los tests de aleatorización

📄 *Muestra* (seguramente *no aleatoria*) sometida a *asignación aleatoria* del tratamiento 1 a  $n_1$  individuos y del tratamiento 2 al resto ( $n_2$ ):

$$\mathbf{y} = (\mathbf{y}_1; \mathbf{y}_2) = (y_{11}, \dots, y_{1n_1}; y_{21}, \dots, y_{2n_2})$$

📄 Si  $H_0$  “el tratamiento no tiene efecto” cierta, es indiferente el tratamiento asignado a cada individuo, tratamientos son simples “etiquetas” asignadas al azar

📄 Entonces la muestra actual es una más de las  $n!$  permutaciones posibles de los datos

## (Eden y Yates, 1933) Tests de permutaciones - aleatorización *aproximados*

---

Principal dificultad: imposible (a nivel práctico) enumerar *todas* las permutaciones posibles si tamaño muestral no es pequeño

Enfoque de remuestreo de Monte Carlo:

- Generar una *muestra* aleatoria *muy grande* (pero mucho menor que el tamaño de la verdadera población) de las permutaciones posibles.
- A partir de ella, estimación (muy precisa) del “verdadero” p-valor y decisión basada en p-valor estimado

# Esquema del procedimiento de Montecarlo

$$\tilde{\mathbf{x}} \mapsto t(\tilde{\mathbf{x}}) = \tilde{t}$$

Generación de  $N$  permutaciones aleatorias

El valor del estadístico sobre la muestra permutada, ¿supera el valor calculado sobre la muestra original?

Cálculo del estadístico sobre cada muestra permutada

Estimación del p-valor

$$\mathbf{x}_{i_1} \mapsto t_{i_1} \mapsto I_{\{t_{i_1} \geq \tilde{t}\}}$$

$$\mathbf{x}_{i_2} \mapsto t_{i_2} \mapsto I_{\{t_{i_2} \geq \tilde{t}\}}$$

$\vdots$

$$\mathbf{x}_{i_N} \mapsto t_{i_N} \mapsto I_{\{t_{i_N} \geq \tilde{t}\}}$$

$$\hat{P}_{H_0} = \frac{1}{N} \sum_{p=1}^N I_{\{t_{i_p} \geq \tilde{t}\}}$$

Criterio de decisión:

Si  $\hat{P}_{H_0} \leq \alpha \Rightarrow$  rechazamos  $H_0$

## Estimación del p-valor

El estimador  $\hat{P}_{H_0}$  es insesgado:

$$E_{\mathbf{i}_p} \left( \hat{P}_{H_0} \right) = P_{H_0} \left[ t \left( \mathbf{X}_{\mathbf{i}_p} \right) \geq t \left( \tilde{\mathbf{x}} \right) \right]$$

Pero preferible el siguiente estimador sesgado,  
(Dwass, 1957):

$$\tilde{P}_{H_0} = \frac{\left( \sum_{p=1}^N I_{\{t_{\mathbf{i}_p} \geq \tilde{t}\}} \right) + 1}{N + 1}$$

Este estimador asegura un test aproximado de extensión  $\alpha$ , para todo  $N$

$$\Pr_{\mathbf{i}_p} \left( \tilde{P}_{H_0} \leq \alpha \right) \leq \alpha$$

## Determinación del número $N$ de permutaciones aleatorias (I)

Estimador  $\tilde{P}_{H_0}$  siempre asegura extensión  $\alpha$ , por pequeño que sea  $N$ , no que p-valor estimado cercano al real, ni potencia alta.

Rechazamos  $H_0$  si  $\tilde{t}$  forma parte del grupo de los  $m = (N + 1)\alpha$  valores mayores  $t_{i_p}$ :

$$\frac{\left( \sum_{p=1}^N I_{\{t_{i_p} \geq \tilde{t}\}} \right) + 1}{N + 1} \leq \alpha \Rightarrow \text{rechazamos } H_0$$

Depende de qué hipótesis es cierta

Curva de potencias en función de parámetro  $p$ :

$$p = \Pr_{i_p} \left[ \tilde{t} \leq t, t \in \{t_{i_p}\} \mid t \text{ escogido al azar} \right]$$

## Determinación del número $N$ de permutaciones aleatorias (II)

La verdadera probabilidad de rechazar  $H_0$ , es decir, la función de potencia del test de Montecarlo es (Marriott, 1979):

$$\beta_N(p) = \sum_{r=0}^{m-1} \binom{N}{r} p^r (1-p)^{N-r} \quad \text{donde } p = P_{H_0} \left[ t(\mathbf{X}_{i_p}) \geq t(\tilde{\mathbf{x}}) \right]$$

Si  $p$  es pequeño,  $\beta_N(p)$  debería ser grande, si  $p$  grande  $\beta_N(p)$  pequeño. Concretamente:

$$\beta_N(p) \leq \alpha \quad \text{si } H_0 \text{ es cierta}$$

$$\beta_N(p) \geq \alpha \quad \text{si } H_0 \text{ es falsa}$$

$\beta_N(p)$ : tabular o aprox.  $N(\mu=Np, \sigma^2=Np(1-p))$



# Determinación del número $N$ de permutaciones aleatorias (y III)

**Table 3.1** Probabilities ( $P$ ) of a significant result for tests as they depend on the number of randomizations ( $N$ ), and the significance level ( $p$ ) that would be given by the full randomization distribution

(a) Test at the 5% level of significance ( $\alpha = 0.05$ )

$m^*$	$N + 1$	$p$				
		0.1	0.075	0.05	0.025	0.01
1	20	0.135	0.227	0.377	0.618	0.826
2	40	0.088	0.199	0.413	0.745	0.942
5	100	0.025	0.128	0.445	0.897	0.997
10	200	0.004	0.065	0.461	0.971	1.000
50	1000	0.000	0.001	0.474	1.000	1.000

(b) Test at the 1% level of significance ( $\alpha = 0.01$ )

$m^\dagger$	$N + 1$	$p$				
		0.02	0.015	0.01	0.005	0.001
1	100	0.135	0.224	0.370	0.609	0.906
2	200	0.091	0.199	0.407	0.738	0.983
5	500	0.028	0.131	0.441	0.892	1.000
10	1000	0.005	0.069	0.459	0.969	1.000
50	5000	0.000	0.001	0.472	1.000	1.000




$^\dagger m = \alpha(N + 1)$ .

Tabla en Manly, BFJ. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall/CRC, 2007

# Algunos tests concretos.

## “Ideas guía” generales

---

-   $H_0$  entendida como restricción sobre mecanismo de generación de datos
-  Aleatorizar según dicha restricción pero preservando todas las demás características del mecanismo generador de datos asumido
-  Por lo tanto, en general *no* se permutarán libremente *todos* los datos, dependerá de cada diseño experimental y de cada hipótesis nula

# Algunos tests concretos.


## Comparación de dos muestras independientes


---

 Datos:

$$\mathbf{y} = (\mathbf{y}_1; \mathbf{y}_2) = (y_{11}, \dots, y_{1n_1}; y_{21}, \dots, y_{2n_2})$$

 Situación e hipótesis ya descritas antes

 Bajo  $H_0$ , permutar libremente los valores de la muestra (o equivalentemente, permutar las “etiquetas” de “grupo 1” y “grupo 2”)

 Calcular sobre cada permutación un estadístico adecuado al problema, p.e.  $t$  de Student

# Algunos tests concretos.

## Dos muestras independientes (II)

---

- Para la mayoría de estadísticos con sentido en esta situación, no importa el orden concreto de valores en cada grupo si no qué valores quedan en cada grupo
- Simplificación importante: enumerar las posibles combinaciones, no permutaciones:

$$\binom{n}{n_1} = \frac{n!}{n_1!n_2!} \ll n!$$

# Algunos tests concretos.

## Dos muestras independientes (y III)

---

- 📄 En un test de permutaciones, lo que importa es en qué orden quedan los valores del estadístico
- 📄 Segunda simplificación importante: simplificar al máximo el estadístico, eliminar cálculos innecesarios respetando el orden
- 📄 P.e.  $t$  de Student  $\Leftrightarrow$  simple diferencia de medias muestrales  $\Leftrightarrow$  suma de valores del primer grupo (si  $n_1 = n_2$ )

# Algunos tests concretos.

## Datos apareados

Si comparamos dos muestras apareadas: permutaremos dentro de cada par  $(x_i, y_i)$  y calcularemos el estadístico de test  $t$  para cada permutación ( $2^n$  posibles)

Si  $t$  basado en  $D = X - Y$ , asignar al azar signo  $-$  ó  $+$  a cada  $|d|$ :  
 $d_{pi} = (-1)^y |d_i|$ ,  $y \sim b(1, 0.5)$

Permutar libremente los  $2n$  valores destruiría la estructura de los datos

$X$	$Y$	$D$
$x_1$	$y_1$	$d_1$
$x_2$	$y_2$	$d_2$
$\vdots$	$\vdots$	$\vdots$
$x_n$	$y_n$	$d_n$

# Algunos tests concretos.

## Dependencia entre variables aleatorias

Si no hay relación entre la variable independiente  $X$  y la dependiente  $Y$ , indiferente que  $x$  se asocie con cada  $y$

La distribución de aleatorización se formará mediante todos los pares en los que  $x$  se mantiene fijo e  $y$  se permuta:  $(x_i, y_{pi})$  (o viceversa)





Sobre la muestra original y cada muestra permutada: estadístico que mida el tipo de asociación conjeturado, p.e.  $r$  de Pearson.

$X$	$Y$
$x_1$	$y_1$
$x_2$	$y_2$
$\vdots$	$\vdots$
$x_n$	$y_n$

# Algunos tests concretos.

## Test de Mantel de asociación entre matrices de distancias


---

-  Test de Mantel de asociación entre matrices de distancias (o similitud, correlación,...)
-  Aplicabilidad muy general (mucho más de la que parece)
-  Dos matrices de distancias calculadas sobre las mismas unidades de estudio  $\Omega = \{1, 2, \dots, n\}$ 
  - p.e. una de distancias genéticas y otra de geográficas, sobre las mismas poblaciones
-  ¿Existe asociación significativa entre ambas formas de medir la distancia?

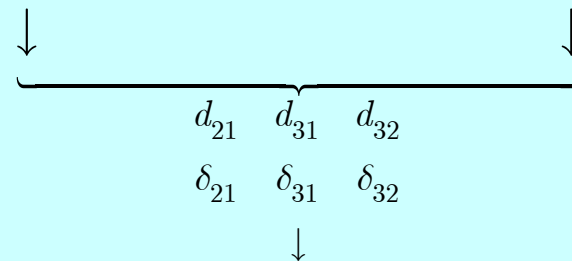


# Algunos tests concretos.

## Test de Mantel (II)

 Sobre los valores no triviales (p.e. triangular inferior) evaluamos algún estadístico de asociación

	1	2	3		1	2	3
1	0				1	0	
2	$d_{21}$	0			2	$\delta_{21}$	0
3	$d_{31}$	$d_{32}$	0		3	$\delta_{31}$	$\delta_{32}$



$$\tilde{t} = \sum_{ij} d_{ij} \delta_{ij}$$

## Algunos tests concretos.

### Test de Mantel (III)

---

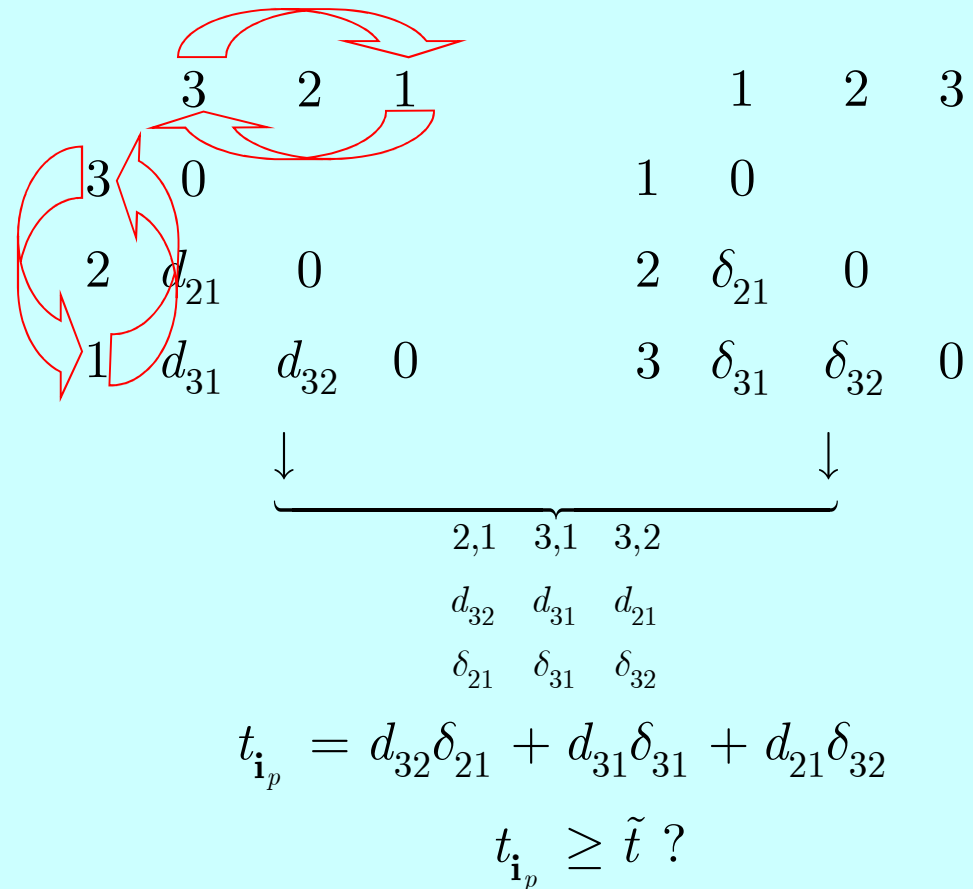
- Si no hay asociación entre ambas formas de medir la distancia, el valor observado  $\tilde{t}$  será uno más entre los obtenidos aleatorizando *el conjunto de las unidades sobre las que se evalúan las distancias* (¡no aleatorizando las distancias!) en una de las matrices, fija la otra
- Repetidamente aleatorizaremos (permutaremos) los elementos de  $\Omega$  y calcularemos  $t$  sobre los valores de distancia que les corresponderían
- Se estimará la significación de la forma habitual

# Algunos tests concretos.

## Test de Mantel (y IV)

Una permutación concreta: se han intercambiado las unidades 1 y 3

Las distancias  $d$  se mantienen pero ahora corresponden a pares distintos



# Algunos tests concretos.

## ANOVA para un factor (one-way)

Factor  $A$  con  $a$  niveles

$n_i$  observaciones  $Y_{ij}$  para el nivel  $i$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, n_i$ .  $n = n_1 + \dots + n_a$  datos:

$$\left( \left( Y_{11}, \dots, Y_{1n_1} \right); \left( Y_{21}, \dots, Y_{2n_2} \right); \dots; \left( Y_{a1}, \dots, Y_{an_a} \right) \right)$$


Hipótesis de partida:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a$$

$$H_1 : \mu_i \neq \mu_{i'} \text{ para algún par } 1 \leq i \neq i' \leq a$$


# Algunos tests concretos.

## ANOVA para un factor (II)

 Bajo enfoque paramétrico normal se calcularía el estadístico  $F$ :

$$F = \frac{MS_A}{MS_R} = \frac{SS_A / (a - 1)}{SS_R / (n - a)}$$

$$SS_A = \sum_{j=1}^a n_i (\bar{Y}_{i\cdot} - \bar{Y}_{..})^2 \quad SS_R = \sum_{j=1}^a \sum_{i=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$$

 Si  $H_0$  es cierta,  $F \sim \mathbf{F}(a - 1, n - a)$ . De ahí obtendríamos el p-valor, etc

## Algunos tests concretos.

### ANOVA para un factor (III)

---

Si  $H_0$  afirma que todos los datos proceden de una misma distribución (es decir, no podemos obviar la suposición de igualdad de dispersiones), podríamos obtener todas las posibles permutaciones de los datos (o una muestra aleatoria grande) y sobre cada una de ellas obtener el estadístico  $F$ . Comparándolo con el valor de  $F$  obtenido sobre la muestra original obtendríamos el p-valor.

## Algunos tests concretos.

### ANOVA para un factor (y IV)

Se obtendría exactamente el mismo resultado, pero más deprisa, con el estadístico:

$$t = \sum_{i=1}^a \frac{\{Y_{i\cdot}\}^2}{n_i} \quad \text{donde} \quad Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$$

Justificación, indicaciones:

$$SS_T = \sum_{j=1}^a \sum_{i=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = SS_A + SS_R$$

$$F = \frac{SS_A / (a - 1)}{SS_R / (n - a)} = \frac{SS_A / (a - 1)}{(SS_T - SS_A) / (n - a)}$$

## Resultado general en inferencia estadística (no solo tests de perm.)

Supongamos que para cada  $\theta_0 \in \Theta$  existe un test  $\phi_{\theta_0}$ , de nivel de significación  $\alpha$ , para  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta \neq \theta_0$ .

Dada una muestra  $\mathbf{y}$ , la región:

$$\begin{aligned} & \left\{ \theta_0 \in \Theta : \phi_{\theta_0}(\mathbf{y}) \text{ no rechaza } H_0 \text{ bajo nivel } \alpha \right\} \\ &= \left\{ \theta_0 \in \Theta : p_{\theta_0}(\mathbf{y}) > \alpha \right\} \text{ (si } p \text{ representa el} \\ & \text{p-valor), define una región de confianza de} \\ & \text{nivel } 1 - \alpha \text{ para el parámetro } \theta \end{aligned}$$



## Resultado general en inferencia estadística (y II)

---

Si el parámetro es escalar y la función  $p$  del p-valor suficientemente regular, se tiene un intervalo de confianza

En este caso, resultado también válido para intervalos de confianza unilaterales:

$$\text{Si } H_1 : \theta > \theta_0, \quad \{\theta_0 \in \Theta : p_{\theta_0}(\mathbf{y}) > \alpha\} = [\theta_L, +\infty]$$

$$\text{Si } H_1 : \theta < \theta_0, \quad \{\theta_0 \in \Theta : p_{\theta_0}(\mathbf{y}) > \alpha\} = [-\infty, \theta_U]$$

## Tests de permutaciones e intervalos de confianza

---

- 📄 Impresión (cierta) de que la metodología de permutaciones solo permite obtener tests
- 📄 Resultado anterior permitiría obtener intervalos de confianza asociados
- 📄 Procedimiento muy de computación intensiva:
  - Escaneado para conjunto amplio de valores de  $\theta_0$  y obtención de los correspondientes p-valores
  - Interpolación para obtener en qué valores del parámetro el p-valor iguala  $\alpha$  (pasa de no rechazar a rechazar o al contrario) para obtener el intervalo