# Mètodes basats en rangs
## tècniques concretes: test de Wilcoxon “rangs amb signe”, dades aparellades

**Mètodes no paramètrics i de remostratge**
Grau interuniversitari en Estadística UB – UPC

*Prof. Jordi Ocaña Rebull*
*Departament d’Estadística, Universitat de Barcelona*

- Adequada per a **comparar** paràmetres de **localització** per **dades aparellades**

- $\mathbf{Y} = \begin{matrix} Y_{11} & Y_{12} & \dots & Y_{1n} \\ Y_{21} & Y_{22} & & Y_{2n} \end{matrix}$ mostra aleatòria

- $(Y_{1j}, Y_{2j})$ obtingudes sota 2 condicions diferents pel mateix subjecte o bloc $j$ → variables dependents amb distribució contínua univariant

# Test de Wilcoxon dels signes-rangs
## planteig – condicions de validesa

- Definim $D_j = Y_{1j} - Y_{2j}$ (o $Y_{2j} - Y_{1j}$ sempre que posteriorment actuem coherentment amb aquesta elecció), $\mathbf{D} = (D_1, \ldots, D_n)$
- Suposarem que $D$ és contínua i simètrica al voltant de la seva mediana $\delta$
- Igualtat de distribucions d'$Y_1$ i $Y_2$ (potser llevat de mesura de localització) implica distribució de $D$ simètrica (però també pot ser-ho en condicions més generals)

# Test de Wilcoxon dels signes-rangs
## planteig – condicions de validesa

$$H_0:\ \delta = 0$$

| | | |
|---|---|---|
| $H_1:\ \delta \neq 0$ | $H_1:\ \delta > 0$ | $H_1:\ \delta < 0$ |
| (bilateral) | (unilateral) | (unilateral) |

# Hipòtesis nul·la i alternativa

# Test dels signes

- Si només comptéssim quantes vegades una diferència és positiva o negativa:

$$S^+ = \sum_{i=1}^{n} I_{\{D_i > 0\}} \quad \text{o} \quad S^- = \sum_{i=1}^{n} I_{\{D_i < 0\}}$$

- Sota $H_0$ $S^+$ (o $S^-$) seguiria una distribució binomial de paràmetres $p$ = ½ i $n$
- Aquesta és la idea del "test dels signes", menys exigent quant a requeriments
- El test de Wilcoxon aprofita més informació (a més del signe, el rang)

- **R** = ($R_1$,…,$R_n$) rangs de les diferències en valor absolut $|D|$
- Suma de rangs de diferències positives i suma de rangs de diferències negatives:

$$R^+ = \sum_{i=1}^{n} R_i I_{\{D_i > 0\}} \qquad R^- = \sum_{i=1}^{n} R_i I_{\{D_i < 0\}}$$

# Test de Wilcoxon dels signes-rangs
## procediment

- Estadístic de Wilcoxon:

$$V = R^{+}$$

- $R^{+} + R^{-} = n(n+1)/2$, novament, per comoditat (brevetat de la taula) s'acostuma a tabular l'estadístic:

$$T = \min\left\{R^{+}, R^{-}\right\}$$

# Estadístic de test

# Criteri de test, procediment "a ma"

$$H_0:\ \delta = 0$$

| $H_1:\ \delta \neq 0$ | $H_1:\ \delta > 0$ | $H_1:\ \delta < 0$ |
|---|---|---|
| es rebutja $H_0$ si: | es rebutja $H_0$ si: | es rebutja $H_0$ si: |
| $T \leq t_\alpha(n)$ | $T \leq t^*_\alpha(n)$ | $T \leq t^*_\alpha(n)$ |
| | i $R^+ > R^-$ | i $R^+ < R^-$ |

$t_\alpha(n)$ valor crític a taula per prova bilateral

$t^*_\alpha(n)$ valor crític a taula per prova unilateral

per nivell de significació $\alpha$ i mida mostral $n$

- Si $H_0$ és certa:
  - (conseqüència de les propietats bàsiques de rangs)

$$E(V) = \frac{n(n+1)}{4}$$

$$\mathrm{var}(V) = \frac{n(n+1)(2n+1)}{24}$$

# Esperança i variància de l'estadístic de Wilcoxon si $H_0$ és certa

- Per mides mostrals “grans”
  - (si l’aproximació pel Teorema central del límit es considera prou vàlida; a la pràctica per *n* fora de la taula)

$$Z \approx N(0,\, 1)$$

$$Z = \frac{V - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

# Aproximació normal

$$H_0: \ \delta = 0$$

| | | |
|---|---|---|
| $H_1: \ \delta \neq 0$<br>es rebutja $H_0$ si:<br>$\lvert Z \rvert \geq z_\alpha$ | $H_1: \ \delta > 0$<br>es rebutja $H_0$ si:<br>$Z \geq z_{2\alpha}$ | $H_1: \ \delta < 0$<br>es rebutja $H_0$ si:<br>$Z \leq -z_{2\alpha}$ |

$z_p$ valor crític >0 a taula $N(0,1)$ per prova bilateral per nivell de significació $p$

# Criteri de test per l'aproximació normal

# Cas d'empats

- En teoria (variables aleatòries contínues...) no hi pot haver empats
- A la pràctica n'hi ha moltes vegades
- Dos tipus d'empats possibles:
  1. $Y_{1j} = Y_{2j} \Rightarrow D_j = 0$
  2. $|D_j| = |D_{j'}|$ per $D_j \neq 0$, $D_{j'} \neq 0$, $j \neq j'$
- Cas "*1*" més problemàtic, no solució clara. Més habitual: ignorar, mida real < $n$
- Pel cas "*2*" procedirem de la forma habitual, amb els rangs mitjans

# Empats però cap $D_j = 0$

- Criteri de decisió pel test de rangs exacte (amb taula) és l'habitual (aproximat)
- Aproximació normal: correcció per variància de *V*:

$$\frac{n(n+1)(2n+1)}{24} - \frac{\sum_{i=1}^{s}\left(t_i^3 - t_i\right)}{48}$$

$s =$ nombre de sèries de valors empatats

$t_i =$ llargada de sèrie $i$ de valors empatats

- Independentment de la presència d'empats, alguns autors recomanen la correcció per continuïtat:

$$Z = \frac{\left| V - \frac{n(n+1)}{4} \right| - \frac{1}{2}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

- No hi ha unanimitat que això representi cap millora

# Correcció per continuïtat

- En test de Wilcoxon, interpretació habitual (i adequada) de $\delta$: mediana de $D = Y_{2j} - Y_{1j}$
- "med" de diferència ≠ diferència de "med"
- Possibles estimadors de $\delta$:

$$\tilde{\delta} = \text{med}(\mathbf{D})$$

$$\hat{\delta} = \text{med}\left(\left\{\frac{1}{2}\left(D_j + D_k\right)\right\}_{1 \le j \le k \le n}\right)$$

- $\hat{\delta}$ directament associat a test Wilcoxon
  - (màx. balanç entre rangs positius i negatius si "centrem" les diferències: $D_j - \hat{\delta}$ )

# Estimació puntual de $\delta$

- Ordenem les $n$ diferències $D = Y_{2j} - Y_{1j}$ (NO rangs, NO $|D|$): $D_{(1)}$, $D_{(2)}$, ..., $D_{(n)}$
- IC de nivell $1 - \alpha$: $[D_{(\lambda)}, D_{(\upsilon)}]$, on:

$$\upsilon^* = \frac{n+1}{2} + \frac{1}{2} z_\alpha \sqrt{n}$$

$$\upsilon = \begin{cases} \upsilon^* & \text{si } \upsilon^* \text{ és enter} \\ \left[\upsilon^* + 1\right] & \text{en cas contrari} \end{cases} \qquad \lambda = n - \upsilon + 1$$

$$z_\alpha \text{ valor t.q. } \Pr\left(\left|Z\right| \le z_\alpha\right) = 1 - \alpha,\ Z \sim N\left(0,1\right)$$

# Interval de confiança per $\delta$