Exemple, test de Wilcoxon, avec $n = 4$

- La loi de T^+ est discrète. Au maximum, il y a 2^n cas à considérer. Par exemple, dans le cas $n = 4$:

Configuration	1	2	3	4	Total
1	0	0	0	0	0
2	0	0	0	1	4
3	0	0	1	0	3
4	0	0	1	1	7
5	0	1	0	0	2
6	0	1	0	1	6
7	0	1	1	0	5
8	0	1	1	1	9
9	1	0	0	0	1
10	1	0	0	1	5
11	1	0	1	0	4
12	1	0	1	1	8
13	1	1	0	0	3
14	1	1	0	1	7
15	1	1	1	0	6
16	1	1	1	1	10

Loi de T^+, $n = 4$

- Ainsi, en résumé, la loi de T^+ lorsque $n = 4$ est de la forme:

$T^+ = t^+$	0	1	2	3	4	5	6	7	8	9	10
$P(T^+ = t^+)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

- La loi de T^+ est symétrique; ceci est vrai en général. Ainsi, on aurait pu se limiter à $\frac{1}{2}2^n = 2^{n-1}$ cas.

Région critique du test de rang signé de Wilcoxon

- On donne la région critique du test sur la médiane $H_0 : \kappa_{1/2} = m_0$ versus les alternatives unilatérales et bilatérale.
- $H_1 : \kappa_{1/2} > m_0$; Sous H_1, il y a plus de chances que $X_i > m_0$ et ainsi la statistique T^+ devrait être grande. On rejette H_0 si $T^+ \geq q_{1-\alpha}$;
- $H_1 : \kappa_{1/2} < m_0$; Sous H_1, il y a plus de chances que $X_i < m_0$ et ainsi la statistique T^+ devrait être petite. On rejette H_0 si $T^+ \leq q_{\alpha}$;
- $H_1 : \kappa_{1/2} \neq m_0$; On rejette H_0 si $T^+ \leq q_{\alpha/2}$ ou encore si $T^+ \geq q_{1-\alpha/2}$.

Recherche des valeurs critiques

- Les valeurs critiques sont fournies dans la Table H du livre de Dickinson Gibbons et Chakraborti (pp. 567-570) pour $n = 2, \ldots, 15$.
- Pour $n > 15$, on peut invoquer une approximation normale.
- Il est recommandé de faire la correction pour la continuité.
- On a le résultat suivant sous l'hypothèse nulle H_0:

$$T^+ \approx \mathcal{N}\left\{\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right\}.$$

Exemple

- Considérons le premier cas, $H_1 : \kappa_{1/2} > m_0$.
- On a alors:

$$P(T^+ \geq q_{1-\alpha}) = P(T^+ \geq q_{1-\alpha} - 1/2).$$

- Après standardisation, on a alors:

$$\frac{q_{1-\alpha} - 0.5 - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = 1.645$$

- Ainsi le point critique est alors:

$$q_{1-\alpha} = \frac{1}{2} + \frac{n(n+1)}{4} + 1.645\sqrt{\frac{n(n+1)(2n+1)}{24}}.$$

Traitement des zéros et des ex aequo

- Nous avons supposé que l'échantillon provenait d'une population avec F continue (et symétrique);
- La probabilité que $X_i = m_0$ est zéro;
- De même, la probabilité que $X_i = X_j$, $i \neq j$ est zéro;
- Compte tenu des arrondis, il arrive que $X_i = m_0$ ou encore il peut y avoir des égalités.

Traitement des zéros et des ex aequo (suite)

- Pour les écarts D_i nuls, on fait comme pour le test des signes;
- On les retire de l'échantillon, et on réduit d'autant la taille de l'échantillon.
- S'il y a des ex aequo parmi les $|D_i|$, plusieurs solutions sont possibles.

Traitement des ex aequo

- Supposons que $X_1, \ldots, X_n$ est iid selon une loi F.
- On rappelle que l'on a définit la fonction rang comme:

$$rang(X_i) = n^{-1} \sum_{j=1}^{n} I(X_j \leq X_i).$$

- Lorsque la distribution sous-jacente F est continue, on a alors que $rang(X_i)$ est une distribution uniforme discrète sur $\{1, \ldots, n\}$.
- D'un point de vue théorique, il n'y a pas de problèmes à utiliser la définition précédente.
- Cependant, même lorsque F est continue, il peut arriver des problèmes pratiques.

Traitement des ex aequo (suite)

- Dans un échantillon de taille $n = 15$, on obtient les mesures suivantes:

3.584	3.024	4.004	4.032	3.948	3.008
3.360	3.164	3.444	3.808	3.864	3.696
3.360	3.920	3.192			

- À trois décimales, on note déjà un ex aequo. Si on arrondit à une décimale, plusieurs groupes d'ex aequo surviennent:

3.6	3.0	4.0	4.0	3.9	3.0
3.4	3.2	3.4	3.8	3.9	3.7
3.4	3.9	3.2			

- Les valeurs ordonnées sont alors:

3.0	3.0	3.2	3.2	3.4	3.4
3.4	3.6	3.7	3.8	3.9	3.9
3.9	4.0	4.0			

- Les rangs selon la définition donnerait: {2, 2, 4, 4, 7, 7, 7, 8, 9, 10, 13, 13, 13, 15, 15 } avec une somme plus grande que $n(n+1)/2 = 120$.

Traitement des ex aequo (suite)

- Une façon de procéder est de calculer les rangs moyens.
- Utiliser les rangs moyens revient à calculer la moyenne des rangs des ex aequo sous l'hypothèse que les observations sont légèrement différentes.

X_i	3.0	3.0	3.2	3.2	3.4	3.4	3.4	3.6	3.7	3.8
$r(X_i)$	1	2	3	4	5	6	7	8	9	10
$\bar{r}(X_i)$	1.5	1.5	3.5	3.5	6	6	6	8	9	10
X_i	3.9	3.9	3.9	4.0	4.0					
$r(X_i)$	11	12	13	14	15					
$\bar{r}(X_i)$	12	12	12	14.5	14.5					

- Avec cette solution la somme des rangs est bien $n(n+1)/2 = 120$.
- Ainsi le rang moyen n'est pas changé.

Traitement des ex aequo (suite)

- L'utilisation des rangs moyens a pour effet de réduire la variance des statistiques de rang.
- Si n est petit et que la distribution exacte est utilisée, on suggère de ne pas corriger.
- Si n est plus grand et que l'approximation normale est utilisée, des ajustements sont nécessaires. Il est suggéré de considérer:

$$V_{corr}(T^+) = \frac{n(n+1)(2n+1)}{24} - \frac{1}{48}\sum_{l=1}^{k} t_l(t_l^2 - 1),$$

lorsqu'il y a k groupes d'ex aequo de taille t_l, $l = 1, \ldots, k$.

Exemple

- On dispose encore de l'échantillon suivant:

3.6	3.0	4.0	4.0	3.9	3.0
3.4	3.2	3.4	3.8	3.9	3.7
3.4	3.9	3.2			

- On désire tester:

$$H_0 : \quad \kappa_{1/2} = 3.5,$$
$$H_1 : \quad \kappa_{1/2} \neq 3.5.$$

- On calcule les différences D_i et on obtient

D_i	0.1	-0.5	0.5	0.5	0.4	-0.5	-0.1	-0.3	-0.1	0.3
$\lvert D_i \rvert$	0.1	0.5	0.5	0.5	0.4	0.5	0.1	0.3	0.1	0.3
$r(\lvert D_i \rvert)$	1	12	13	14	9	15	2	6	3	7
$\bar{r}(\lvert D_i \rvert)$	2.5	13.5	13.5	13.5	10	13.5	2.5	7	2.5	7

D_i	0.4	0.2	-0.1	0.4	-0.3
$\lvert D_i \rvert$	0.4	0.2	0.1	0.4	0.3
$r(\lvert D_i \rvert)$	10	5	4	11	8
$\bar{r}(\lvert D_i \rvert)$	10	5	2.5	10	7

Exemple (suite)

- Dans l'exemple il n'y a pas de zéros, mais quatre groupes d'ex aequo.
- $t_1 = 4$; $t_2 = 3$; $t_3 = 3$ et finalement $t_4 = 4$.
- La statistique est

$$T^+ = 2.5 + 13.5 + 13.5 + 10 + 7 + 10 + 5 + 10 = 71.5;$$

- La variance corrigée est:

$$\begin{aligned} V_{corr}(T^+) &= \frac{n(n+1)(2n+1)}{24} - \frac{1}{48}\sum_{l=1}^{4} t_l(t_l^2 - 1), \\ &= 310 - 3.5 = 306.5; \end{aligned}$$

Exemple (suite)

- La région critique est telle que:

$$\begin{aligned} \alpha/2 &= P(T^+ \leq q_{\alpha/2}), \\ 1-\alpha/2 &= P(T^+ \geq q_{1-\alpha/2}), \end{aligned}$$

- On applique la correction de continuité

$$\begin{aligned} \alpha/2 &= P(T^+ \leq q_{\alpha/2} + 1/2), \\ 1-\alpha/2 &= P(T^+ \geq q_{1-\alpha/2} - 1/2), \end{aligned}$$

- On trouve:

$$\begin{aligned} q_{\alpha/2} &= \frac{n(n+1)}{4} - \frac{1}{2} - z_{1-\alpha/2}\sqrt{V_{corr}(T^+)} = 25.19, \\ q_{1-\alpha/2} &= \frac{n(n+1)}{4} + \frac{1}{2} + z_{1-\alpha/2}\sqrt{V_{corr}(T^+)} = 94.81 \end{aligned}$$

- Puisque $T^+ = 71.5$, on ne rejette pas H_0 au niveau 5%.