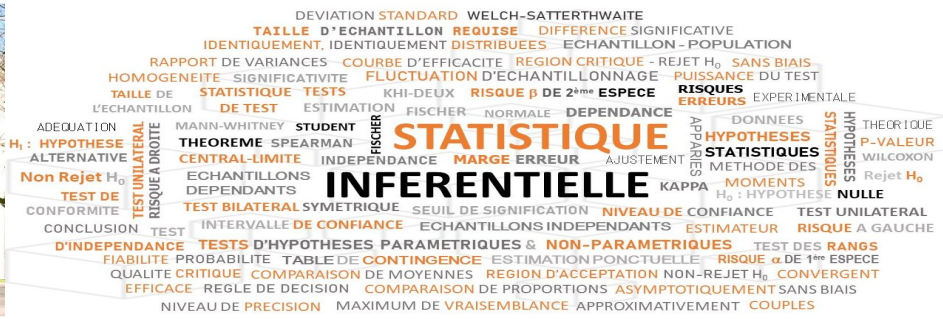
UFR de Sciences économiques et de Gestion

Université Lumière Lyon 2

UNIVERSITÉ DE LYON





UFR SEG L3-S6

Statistique Inférentielle

Support de cours (3/3)

Tests non-paramétriques

Rafik Abdesselam

Courriel : rafik.abdesselam@univ-lyon2.fr
Web : http://perso.univ-lyon2.fr/~rabdesse/fr/
Support pédagogique : http://perso.univ-lyon2.fr/~rabdesse/Documents/

Tests non-paramétriques

- On qualifie de non-paramétriques "distribution free" les tests statistiques qui sont construits à partir d'une fonction des observations sur un échantillon aléatoire, fonction dont la loi de probabilité ne dépend pas de la connaissance de la distribution de la population-mère.

- La validité des tests non-paramètriques dépend seulement d'un nombre très restreint de conditions d'application (échantillons considérés doivent être aléatoires et simples) beaucoup moins contraignantes que celles requises pour la mise en oeuvre des tests paramètriques (distribution normale de la population-mère ou échantillon de grande taille).

- Un test non-paramètrique présente quelques avantages :
  - son application est relativement facile et rapide,
  - s'applique à des échantillons de petites tailles,
  - s'applique à des caractères qualitatifs, à des grandeurs de mesure, à des ratios, à des rangs de classement, etc.

Tests non-paramétriques

- On distinguera principalement les deux familles suivantes :
  1. Test du Khi-deux de Pearson :
     - ★ Test d'indépendance dans un tableau de contingence.
     - ★ Test d'homogénéité de plusieurs populations.
     - ★ Test d'ajustement ou d'adéquation entre deux distributions.
  2. Tests appliqués aux rangs, aux signes et à la concordance :
     - ★ Test de la somme des rangs (Wilcoxon et Mann-Withney)
     - ★ Test de la somme des rangs des différences positives (Wilcoxon)
     - ★ Test d'indépendance de rangs de Spearman
     - ★ Test d'indépendance d'accord ou de concordance Kappa de Cohen
- Une caractéristique essentielle des méthodes non-paramétriques est leur relative simplicité et rapidité des calculs.
- Remplacer les valeurs observées par des indicateurs ou des rangs provoque évidemment une certaine perte d'information. De ce fait, les tests non-paramétriques sont généralement moins puissants que les tests paramétriques, beaucoup plus robustes.

Les différents tests du Khi-deux

- Le test de Pearson communément appelé test du khi-deux est un outil statistique qui permet de vérifier la concordance entre une distribution expérimentale et une distribution théorique.
- Il s'agit là de 3 tests statistiques non paramétriques du khi-deux dont les objectifs sont totalement différents
- La même quantité χ^2 du test de Pearson est utilisée dans les trois cas bien distincts :
  - 1.1 Test d'indépendance entre 2 caractères.
  - 1.2 Test d'homogénéité.
  - 1.3 Test d'ajustement ou d'adéquation ou de conformité.

Les différents tests du Khi-deux

- 1.1 Test d'indépendance du Khi-deux

  Est-il plausible d'affirmer que deux variables observées sont indépendantes ou au contraire elles présentent un certain degré d'association ?

- 1.2 Test d'homogénéité du Khi-deux On compare k populations entre elles pour en vérifier l'homogénéité suivant les modalités d'une variable particulière.

  Peut-on affirmer que les proportions d'individus dans chaque modalité de la variable étudiée sont les mêmes dans les k populations ou au contraire elles ne sont pas identiques dans les k populations pour au moins une modalité ?

- 1.3 Test de conformité - ajustement du Khi-deux
  On vérifie la qualité de l'ajustement entre une distribution théorique et une distribution expérimentale.

  Est-il plausible d'affirmer que la répartition des observations d'un échantillon selon les modalités d'une variable puisse s'apparenter à une répartition théorique connue (loi binomiale, loi de poisson, loi normale, etc.) ?

1.1 Test d'indépendance du khi-deux

- Le test du khi-deux est fréquemment utilisé pour tester si deux caractères qualitatifs ou quantitatifs (répartis en classes), observés dans une population sont indépendants ou si, au contraire, ils sont dépendants : présentent un certain degré d'association (liaison).
- Principe général du test :
  1. Un échantillon aléatoire de taille n est prélevé d'une population et est observé selon deux caractères X à p modalités et Y à q modalités.
  2. La répartition des n observations suivant les modalités croisées des deux caractères se présente sous la forme d'un tableau à double entrée appelé tableau de contingence.
  3. Il s'agit par la suite de tester, à l'aide du khi-deux de Pearson, si les deux caractères sont indépendants ou non.

Démarche du Test d'indépendance du χ^2

- Pour rechercher l'indépendance entre la variable ligne et la variable colonne d'un tableau de contingence, on va comparer la distribution statistique observée dans l'échantillon, à une distribution théorique.
- Cette distribution théorique est celle qu'on doit avoir si les 2 variables sont indépendantes, c'est-à-dire sous l'hypothèse nulle H_0.
- On veut savoir si les écarts entre ces deux distributions sont imputables aux fluctuations d'échantillonnage, ou si au contraire, les écarts sont trop importants pour que l'on puisse accepter l'hypothèse H_0.
- Conclusion : répondre à la question : *l'expérience contredit-elle l'hypothèse nulle H_0 d'indépendance que l'on suppose vraie ?*

Tableau de contingence

- Tableau des effectifs observés :

	y_1	...	y_k	...	y_q	Total ligne
x_1	n_{11}	...	n_{1k}	...	n_{1q}	$n_{1.} = \sum_k n_{1k}$
$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
x_j	n_{j1}	...	n_{jk}	...	n_{jq}	$n_{j.}$
$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
x_p	n_{p1}	...	n_{pk}	...	n_{pq}	$n_{p.}$
Total colonne	$n_{.1} = \sum_j n_{jk}$	...	$n_{.k}$	...	$n_{.q}$	$n = n_{..} = \sum_j \sum_k n_{jk}$

- Les hypothèses statistiques peuvent s'énoncer ainsi :

$$\begin{cases} H_0 : \text{les caractères X et Y sont indépendants} \\ H_1 : \text{les caractères X et Y sont dépendants} \end{cases}$$

Indépendance en probabilité de 2 événements A et B :

$$P(A \cap B) = P(A) \times P(B) \Rightarrow \frac{n_{AB}}{n} = \frac{n_A}{n}\frac{n_B}{n} \Rightarrow n_{AB} = \frac{n_A n_B}{n}$$

- Sous l'hypothèse nulle H_0 : indépendance des deux caractères, on a,
$p_{jk} = p_{j.}p_{.k} \ \forall (j = 1, p \, et \, k = 1, q)$ (probabilités conjointes $p_{jk} = \frac{n_{jk}}{n}$).

- l'estimation des effectifs théoriques s'obtient en répartissant la taille de l'échantillon n dans les proportions obtenues selon les estimations des probabilités conjointes (indépendance en probabilité) :

$$\widehat{p}_{jk} = \frac{n_{j.}}{n}\frac{n_{.k}}{n} = \frac{\widehat{n}_{jk}}{n} \Rightarrow \widehat{n}_{jk} = \frac{n_{j.}n_{.k}}{n}$$

Test d'indépendance du khi-deux

- Pour comparer les répartitions théorique et observée, on calcule, sous l'hypothèse nulle H_0 la quantité :

$$k_0 = \sum_j^p \sum_k^q \frac{(n_{jk} - \widehat{n}_{jk})^2}{\widehat{n}_{jk}}$$

laquelle sous H_0 est distribuée selon la loi du khi-deux $\chi^2_{\nu=(p-1)(q-1)\ d.d.l.}$ pour le risque d'erreur α choisi.

- Décision et conclusion du test statistique :

L'hypothèse nulle H_0 d'indépendance est rejetée, au niveau α, si $k_0 \geq \chi^2_\alpha$. Le test statistique d'indépendance du khi-deux est toujours unilatéral risque à droite.

Conditions d'application

- On considère que l'approximation par la loi du Khi-deux est satisfaisante si $n \geq 30$ et si tous les effectifs théoriques sont supérieurs ou égaux à 5.
- Si les effectifs des classes extrêmes sont inférieurs à 5, il faut procéder à des regroupements de classes ou à la correction de Yates.
- Le choix des classes, quand il est possible, doit être tel que les effectifs théoriques soient proches les uns des autres, quand ils ne sont pas tous égaux.
- Le nombre k des classes, lorsqu'il est à définir, doit être à la fois assez grand pour perdre le moins d'information possible, et assez petit pour que les effectifs des classes ne soient pas trop réduits. Empiriquement, le nombre k = $[\sqrt{n}]$ (partie entière de la racine carrée de n) paraît convenable.

Correction de Yates

- Lorsque les effectifs théoriques sous l'hypothèse nulle H_0 : Indépendance sont inférieures à 5, on peut corriger la statistique de test pour prendre en compte cette situation.
  On considère que l'approximation par la loi du Khi-deux est satisfaisante si $n \geq 30$ et si tous les effectifs théoriques sont supérieurs ou égaux à 5.
- Attention il faut néanmoins que toutes les fréquences théoriques soient supérieures à 3.
- La correction de Yates est une correction de continuité de la statistique de test, elle est modifiée de la façon suivante :

$$\sum_{j=1}^{p} \sum_{k=1}^{q} \frac{(|n_{jk} - \widehat{n}_{jk}| - 0.5)^2}{\widehat{n}_{jk}} \hookrightarrow \chi^2_{(p-1)(q-1)\ d.d.l.}$$

Exemple d'application 1

Test d'indépendance : Niveau de satisfaction et Type de fonds.

- Un conseiller financier d'un établissement bancaire a effectué une étude auprès d'un échantillon de 210 clients pour déterminer s'il existait un lien entre le type de fonds (Actions / Obligations) qu'elle propose et le niveau de satisfaction de ses clients (Elevé / Bas). Les résultats sont présentés dans le tableau suivant :

	Actions	Obligations	Total
Élevé	75	63	138
Bas	25	47	72
Total	100	110	210

Au seuil de signification $\alpha = 5\%$, peut-on conclure quil existe une relation entre ces deux variables ?

1. Hypothèses statistiques
2. Seuil de signification - Conditions d'application du test - Nombre de degrés de liberté
3. Statistique de test
4. Calcul de la valeur de la statistique de test k_0 sous l'hypothèse nulle H_0
5. Règle de décision et conclusion

Exemple d'application 1 - Solution

1. Hypothèses statistiques $\begin{cases} H_0 : \textit{le type de fonds et le niveau de satisfaction sont indpendants} \\ H_1 : \textit{le type de fonds et le niveau de satisfaction sont liés} \end{cases}$

	Actions	Obligations	Total
Élevé	75 $65.71 = \frac{138x100}{210}$	63 $72.29 = \frac{138x110}{210}$	138
Bas	25 $34.29 = \frac{72x100}{210}$	47 $37.71 = \frac{72x110}{210}$	72
Total	100	110	210

Tableaux des effectifs observés - nij et des effectifs théoriques - $\widehat{nij}$

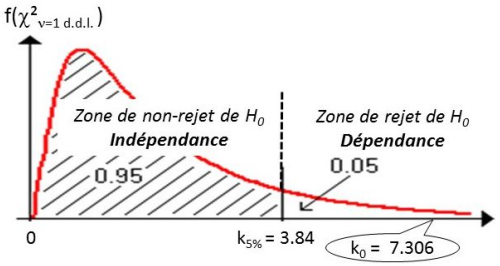
2. Seuil de signification - Conditions d'application du test - Degrés de liberté : $\alpha = 5\%$. Echantillon aléatoire de taille n = 210 observé selon deux caractères qualitatifs à p = 2 et q = 2 modalités. $(p-1)(q-1) = 1\, d.d.l.$

3. Statistique de test : $\sum_{j=1}^{p=2} \sum_{k=1}^{q=2} \frac{(n_{jk} - \widehat{n}_{jk})^2}{\widehat{n}_{jk}} \hookrightarrow \chi^2_{1\, d.d.l.}$

4. Calcul de la statistique de test sous l'hypothèse nulle H_0 : Indépendance

$$k_0 = \sum_{j=1}^{p=2} \sum_{k=1}^{q=2} \frac{(n_{jk} - \widehat{n}_{jk})^2}{\widehat{n}_{jk}} = 7.3061 = \sum_{i=1}^{pq=4} \frac{(O_i - T_i)^2}{T_i}$$

O_i et T_i étant les effectifs Observé et Théorique dans la cellule i.



fractile du $\chi^2_{1\, d.d.l.}$: $k_{\alpha=5\%} = 3.84$.

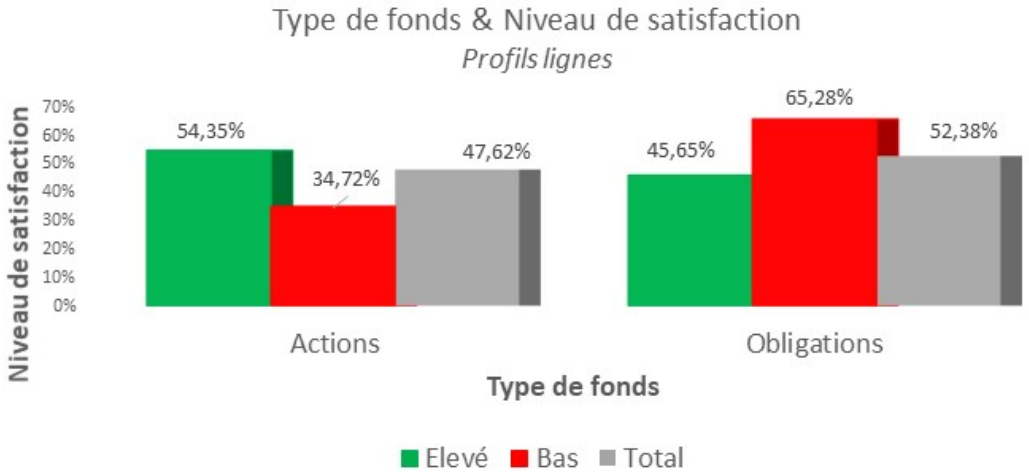
5. Décision et conclusion : La valeur $k_0 = 7.3061$ appartient à la zone de rejet de H_0. En effet, $k_0 = 7.3061 > k_{\alpha=5\%} = 3.84$. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, qu'il y a un lien entre le niveau de satisfaction des clients et le type de fonds.

Statistiques descriptives

Profils Lignes & Colonnes

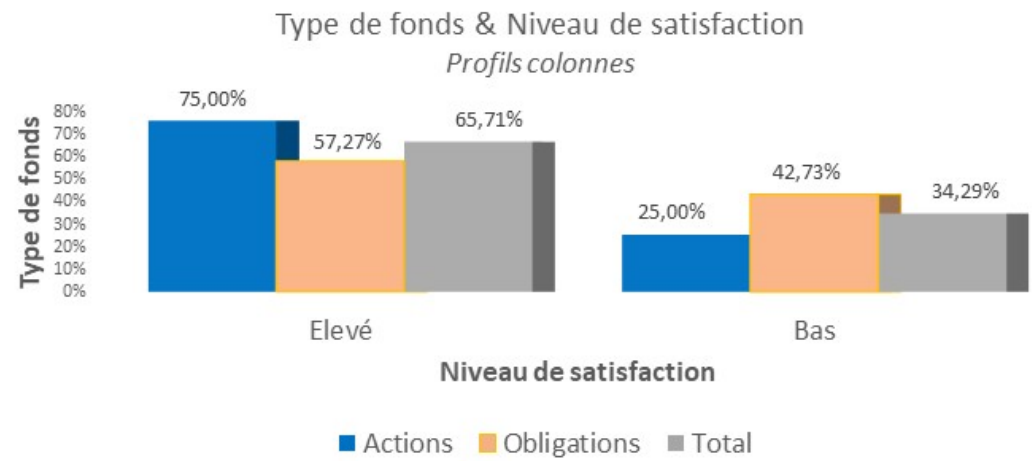
Profils-Lignes

	Actions	Obligations	Total
Elevé	54,35%	45,65%	100,00%
Bas	34,72%	65,28%	100,00%
Total	47,62%	52,38%	100,00%



Profils-Colonnes

	Actions	Obligations	Total
Elevé	75,00%	57,27%	65,71%
Bas	25,00%	42,73%	34,29%
Total	100,00%	100,00%	100,00%



Test d'homogénéité du khi-deux

- Dans le test précédent, d'indépendance du Khi-deux, on prélève un échantillon de taille n d'une population que l'on observe selon deux caractères X_p et Y_q respectivement à p et q modalités - classes.
- Dans le cas où l'on veut tester l'homogénéité de plusieurs populations, on prélève au hasard de k populations , k échantillons indépendants de tailles respectives $n_1, n_2, \ldots, n_k$. Chacun des k échantillons est ensuite observé selon le même et unique caractère X_r à r modalités - classes :

	Echantillons					
Modalités	1^{er}	$\cdots$	j^{eme}	$\cdots$	k^{eme}	Total ligne
X_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1k}	$n_{1.}$
X_2	n_{21}	$\cdots$	n_{2j}	$\cdots$	n_{2k}	$n_{2.}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
X_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ik}	$n_{i.}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
X_r	n_{r1}	$\cdots$	n_{rj}	$\cdots$	n_{rk}	$n_{r.}$
	n_1	$\cdots$	nj	$\cdots$	n_k	n_T

Démarche du Test d'homogénéité du χ^2

- Il ne s'agit plus ici de vérifier l'indépendance entre deux caractères mais plutôt de comparer les k populations entre elles pour vérifier leur homogénéité suivant un caractère X_r à r modalités.
- Les hypothèses statistiques peuvent s'énoncer ainsi :

$$\begin{cases} H_0 : \textit{Les k populations sont homogènes suivant le caractre } X_r \text{ observé} \\ H_1 : \textit{Les k populations ne sont pas homogènes suivant le caractre } X_r \text{ observé} \end{cases}$$

On a sous l'hypothèse nulle H_0 d'homogénéité des k populations :

$$p_{i1} = p_{i2} = \ldots = p_{ij} = \ldots = p_{ik} = p_{i.} \quad \text{pour tout i = 1,r}$$

où, $p_{i.}$ représente la proportion associée à la $i^{ème}$ modalité que l'on suppose identique dans les k populations.

On estime $p_{i.}$ par $\widehat{p_{i.}} = \frac{n_{i.}}{n_T}$

donc toujours sous H_0 , on en déduit les effectifs théoriques :

$$\widehat{n_{ij}} = n_j p_{ij} = n_j \widehat{p_{i.}} = \frac{n_j n_{i.}}{n_T}$$

Test d'homogénéité du Khi-deux

- Pour comparer les répartitions théorique et observée, on calcule, sous l'hypothèse nulle H_0 la quantité :

$$k_0 = \sum_i^r \sum_j^k \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

Remarque : C'est la même expression et la même quantité que dans le cas d'un tableau de contingence.

Laquelle sous l'hypothèse H_0 : d'homogénéité des k populations, est distribuée selon une loi du khi-deux à
$\nu = rk - (r-1) - k = (r-1)(k-1)$ degrés de liberté.

- Décision et conclusion du test statistique :

L'hypothèse nulle H_0 d'homogénéité est rejetée, au seuil de signification α, si $k_0 \geq \chi^2_{table,\nu,\alpha}$. Le test statistique d'homogénéité du khi-deux est toujours unilatéral risque à droite.

Test d'homogénéité du Khi-deux

- Remarques :
  - Si les k populations sont homogènes suivant le caractère X_r observé, cela revient à confondre les k populations en une seule population.

  - Si toutefois les calculs de ce test sont identiques à ceux d'un test d'indépendance, le raisonnement sous-jacent dans chaque cas est différent et les hypothèses statistiques qui sont éprouvées dans chaque test sont différentes.

Exemple d'application 2

- Test d'homogénéité de 2 populations.

Le service marketing d'une entreprise a effectué un pré-test pour étudier la réaction en ce qui a trait à une nouvelle version de son produit. 200 consommateurs ont été choisis au hasard respectivement dans 2 régions différentes. On a remis à chaque consommateur les trois produits (ayant une couleur différente mais de version non identifiée). A l'issue d'une période d'essai d'un mois, les préférences se présentent comme suit :

Version & Région	Région 1	Région 2
Nouvelle	90	105
Actuelle	50	60
Ancienne	60	35
	$n_1 = 200$	$n_2 = 200$

a) Peut-on conclure, au seuil $\alpha = 1\%$ puis $\alpha = 5\%$, que les consommateurs de chaque région apprécient de manière identique les diverses versions du produit ?

1. Hypothèses statistiques :
2. Seuil de signification - Conditions d'application du test - Degrés de liberté :
3. Statistique de test :
4. Calcul de la statistique de test k_0 sous l'hypothèse nulle H_0 :
5. Règle de décision et conclusion :

b) Localiser les disparités au seuil de signification $\alpha = 5\%$.
c) Quel conseil peut-on donner au responsable Marketing qui veut lancer la nouvelle version du produit ?
d) Peut-on parler d'indépendance entre le caractère "versions du produit" et le caractère "Localité - régions", au seuil de signification $\alpha = 5\%$? $\alpha = 1\%$?

Exemple d'application 2 - Solution 1/5

1. Hypothèses statistiques

$$\begin{cases} H_0 : \textit{la préférence du produit suivant les 3 versions se répartit de façon identique dans les 2 régions.} \\ (\textit{comportement homogène des 2 régions en ce qui concerne la préférence des versions du produit}). \\ H_1 : \textit{la préférence ne se répartit pas de façon identique dans les 2 régions.} \end{cases}$$

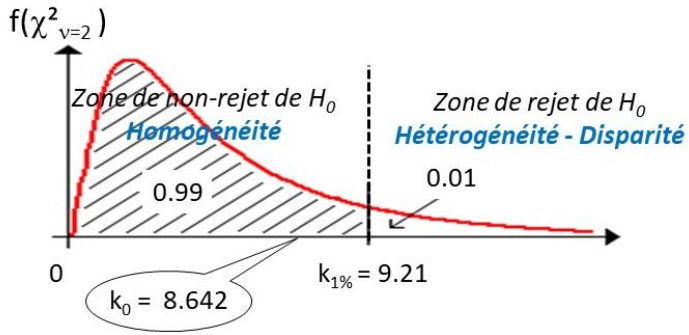
2. Seuil de signification - Conditions d'application du test - Degrés de liberté : $\alpha = 1\%$ $k = 2$ échantillon aléatoire indépendants de taille n = 200 observés selon un caractère qualitatif "version" à $r = 3$ modalités. Nombre de degrés de liberté : $\nu = (r-1)(k-1) = 2\, d.d.l.$

3. Statistique de test : $\sum_{i=1}^{r=3} \sum_{j=1}^{k=2} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}} \hookrightarrow \chi^2_{2\, d.d.l.}$

4. Calcul de la statistique du $\chi^2_{calculé}$ sous l'hypothèse nulle H_0 : Homogénéité

$$k_0 = \frac{(90-97.5)^2}{97.5} + \cdots + \frac{(35-47.5)^2}{47.5} = 8.642$$

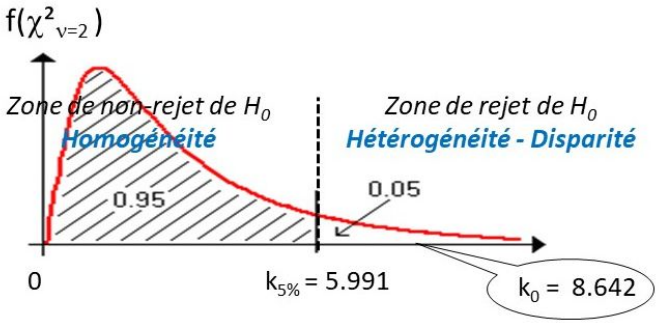
5. Décision et conclusion $\alpha = 1\%$: La valeur k_0 appartient à la zone de non-rejet de H_0. En effet, $k_0 = 8.642 < \chi^2_{2;1\%} = 9.210$. Au seuil de signification $\alpha = 1\%$, on peut conclure à un comportement homogène des populations des 2 régions en ce qui concerne la préférence des versions du produit.



fractile du $\chi^2_{2\, d.d.l.}$: $k_{\alpha=1\%} = 9.21$.

Exemple d'application 2 - Solution 2/5

Décision et conclusion $\alpha = 5\%$: La valeur k_0 appartient à la zone de rejet de H_0.
En effet, $k_0 = 8.642 > \chi^2_{2;5\%} = 5.991$. Au seuil de signification $\alpha = 5\%$, on peut conclure que les 2 régions n'ont pas un comportement homogène en ce qui concerne la préférence du produit.



fractile du $\chi^2_{2\ d.d.l.}$: $k_{\alpha=5\%} = 5.991$.

- Poursuivons cette analyse au seuil $\alpha = 5\%$ vu que les populations ne sont pas homogènes afin de localiser et d'expliquer la nature de(s) disparité(s).

- Localiser les disparités revient à déterminer pour quelle(s) modalité(s) "versions du produit", il y a une différence significative entre les 2 régions.

$$\begin{cases} H_0 : \text{les 2 populations sont homogènes} \Leftrightarrow p_{11} = p_{12} \text{ et } p_{21} = p_{22} \text{ et } p_{31} = p_{32} \\ H_1 : \text{les 2 populations ne sont pas homogènes} \Leftrightarrow p_{11} \neq p_{12} \text{ et/ou } p_{21} \neq p_{22} \text{ et/ou } p_{31} \neq p_{32} \end{cases}$$

Exemple d'application 2 - Solution 3/5

- A l'aide des résultats de l'enquête, estimons les probabilités de préférence dans chaque région.

- Utilisons le test de comparaison de deux proportions

  Comparaisons de proportions - Statistique de test :

  $n_1 = n_2 = 200$

  $\frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \hookrightarrow N(0;1)$ ou $\frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}} \hookrightarrow N(0;1)$ avec $\widehat{p} = \frac{n_1 \widehat{p}_1 + n_2 \widehat{p}_2}{n_1 + n_2}$

Version	Proportion Région 1 $\widehat{p}_{i1}$ $i = 1,3$	Proportion Région 2 $\widehat{p}_{i2}$ $i = 1,3$	Estimation commune $\widehat{p}$ $\widehat{p} = \frac{n_1 \widehat{p}_{i1} + n_2 \widehat{p}_{i2}}{n_1 + n_2}$	Ecart-type de la dfférence des proportions $\sqrt{\widehat{p}\widehat{q}(\frac{1}{n_1} + \frac{1}{n_2})}$
1 - Nouvelle	$\widehat{p}_{11} = \frac{90}{200} = 0.45$	$\widehat{p}_{12} = \frac{105}{200} = 0.525$	0.4875	0.04997
2 - Actuelle	$\widehat{p}_{21} = \frac{50}{200} = 0.25$	$\widehat{p}_{22} = \frac{60}{200} = 0.30$	0.275	0.0446
3 - Ancienne	$\widehat{p}_{31} = \frac{60}{200} = 0.30$	$\widehat{p}_{32} = \frac{35}{200} = 0.175$	0.2375	0.0425

Exemple d'application 2 - Solution 4/5

- Tableau récapitulatif

Version	Hypothèse H_0 :	u_0 valeur de la Statistique de test sous H_0	Non-Rejet de H_0 $-1.96 \leq u_0 \leq 1.96$	Rejet de H_0 $u_0 < -1.96$ ou $u_0 > 1.96$
1 - Nouvelle	$\widehat{p}_{11} = \widehat{p}_{12}$	-1.5	✓	
2 - Actuelle	$\widehat{p}_{21} = \widehat{p}_{22}$	-1.121	✓	
3 - Ancienne	$\widehat{p}_{31} = \widehat{p}_{32}$	2.94		✓

- Conclusion : la disparité est attribuable à la $3^{ème}$ modalité "Version ancienne", où, l'écart entre les proportions de chaque région s'avère significatif au seuil $\alpha = 5\%$, on peut considérer que $p_{31} \neq p_{32}$.
- Conseil : l'entreprise pourrait mettre en oeuvre une campagne de publicité plus dynamique dans la région 1 pour essayer de s'attribuer ou de convaincre les ménagères qui utilisent encore l'ancienne version du produit ($\widehat{p}_{31} = 30\% > \widehat{p}_{32} = 17.50\%$).

Exemple d'application 2 - Solution 5/5

- Au seuil $\alpha = 5\%$, on a conclu que les 2 régions (populations) ne sont pas homogènes donc le test d'indépendance entre le caractère "Version" et "Localité" ne peut pas et ne doit pas s'appliquer.
- Au seuil $\alpha = 1\%$, les 2 régions (populations) sont homogènes en ce qui concerne la préférence des versions du produit. On peut appliquer le test d'indépendance.
- Pour un même risque d'erreur α, l'homogénéité $\Rightarrow$ l'indépendance.

1.3 Test d'ajustement entre 2 distributions

- Ce test permet de juger de la qualité de l'ajustement ou de l'adéquation d'une distribution observée (expérimentale) à une distribution théorique connue (spécifiée) : Uniforme, Poisson, Binomiale, Normale, etc.
- D'une façon générale, on dispose d'un échantillon aléatoire simple $(x_1, x_2.., x_n)$ prélevé d'une population de taille infinie et dont on ne connaît pas *a priori* la loi de probabilité.
- Ces n observations sont réparties selon les k valeurs possibles de la variable observée X si celle-ci est discrète ou selon une répartition en k classes si X est discontinue.

Répartition observée					
Classe : Résultat x_i	C_1	C_2	$\cdots C_i \cdots$	C_k	Total
Effectif observé : n_i (i=1,k)	n_1	n_2	$\cdots n_i \cdots$	n_k	n

Question : Est-il plausible d'affirmer que la répartition observée puisse s'apparenter à une répartition théorique spécifiée (loi uniforme, loi de Poisson, loi binomiale, loi normale, etc.) pour la variable X étudiée ?

1.3 Test d'ajustement entre 2 distributions

- Les hypothèses statistiques peuvent s'énoncer ainsi :

$$\begin{cases} H_0 : \text{les observations suivent la loi théorique spécifiée} \\ H_1 : \text{les observations ne suivent pas la loi théorique spécifiée} \end{cases}$$

- Sous l'hypothèse nulle H_0 , on peut construire une répartition idéale des n observations en ayant recours aux probabilités tabulées (connues) du modèle théorique spécifié.

On obtient alors la répartition théorique suivante :

Répartition théorique Classe : Résultat x_i Effectif théorique : $\widehat{n}_i$ (i=1,k)	C_1 np_1	C_2 np_2	$\cdots C_i \cdots$ $\cdots np_i \cdots$	C_k np_k	Total n

où, p_i est la probabilité pour que X prenne sa valeur dans la classe C_i . En effet, l'effectif d'une classe C_i suit une loi binomiale B(n , p_i) d'espérance mathématique l'effectif théorique : $\widehat{n}_i = np_i$.

1.3 Test d'ajustement entre 2 distributions

- Pour comparer les répartitions théorique et observée, on calcule, sous l'hypothèse nulle H_0 la quantité :

$$k_0 = \sum_{i=1}^{k} \frac{(n_i - \widehat{n}_i)^2}{\widehat{n}_i}$$

laquelle sous H_0 est distribuée selon la loi du khi-deux $\chi^2_{\nu d.d.l.}$, noté $\chi^2_{table,\nu,\alpha}$ pour le risque d'erreur α choisi.

- Décision et conclusion du test statistique : l'hypothèse nulle H_0 d'indépendance est rejetée, au niveau α, si $k_0 \geq \chi^2_{table,\nu,\alpha}$
Le test statistique d'ajustement du khi-deux est toujours unilatéral, risque à droite.

Test d'ajustement entre 2 distributions

- Remarques 1 :
  1. Si sous l'hypothèse nulle H_0, la distribution théorique est complètement spécifiée alors le nombre de degrés de liberté : $\nu = k - 1$ où, k étant le nombre de classes.
  2. Le plus souvent la distribution théorique sous H_0 n'est pas entièrement spécifiée, les paramètres doivent être estimés, dans ce cas, le nombre de degrés de libertés : $\nu = k - 1 - r$ où, r est le nombre de paramètres estimés.

Exemples de lois théoriques spécifiées sous H_0 :

$X \to B(n; p)$; p paramètre inconnu estimé, n étant fixé $\Rightarrow r = 1$

$X \to P(\lambda)$; λ paramètre inconnu estimé $\Rightarrow r = 1$

$X \to N(m; \sigma^2)$; m et/ou σ^2 paramètres inconnus $\Rightarrow r = 1$ ou 2.

Loi de probabilité	Paramètres nécessaires pour calculer les probabilités	Estimation
B(n ; p)	p (la taille de l'échantillon n est fixée)	$\widehat{p} = \frac{\overline{x}}{n}$
$P(\lambda)$	λ	$\widehat{\lambda} = \overline{x} = s^2$
$N(m; \sigma^2)$	m et σ^2	$\overline{x}$ et s^2 ou s^{*2}

Test d'ajustement entre 2 distributions

- Remarques 2 :
  1. Les seules conditions d'application qui sont requises pour effectuer ce test sont :
     - Echantillon prélevé au hasard de la population
     - La taille de l'échantillon doit être suffisamment importante de sorte que les effectifs théoriques $n_i \geq 5$ pour i = 1, k .
  2. Sachant que les k classes sont liées par la relation :
     $$\sum_{i=1,k} \widehat{n_i} = \sum_{i=1,k} n_i = n,$$
     la quantité $\chi^2_{calculé}$ peut être calculée plus facilement :
     $$k_0 = \sum_{i=1,k} \frac{n_i^2}{\widehat{n_i}} - n$$

Exemple d'application 3

- Test d'ajustement à une loi théorique "Uniforme discrète - Equiprobabilité".
- Le dé est-il bien équilibré ? On jette 120 fois un dé et on a enregistré les résultats obtenus dans le tableau ci-dessous :

Résultat	1	2	3	4	5	6	Total
Effectif observé	14	16	28	30	18	14	120

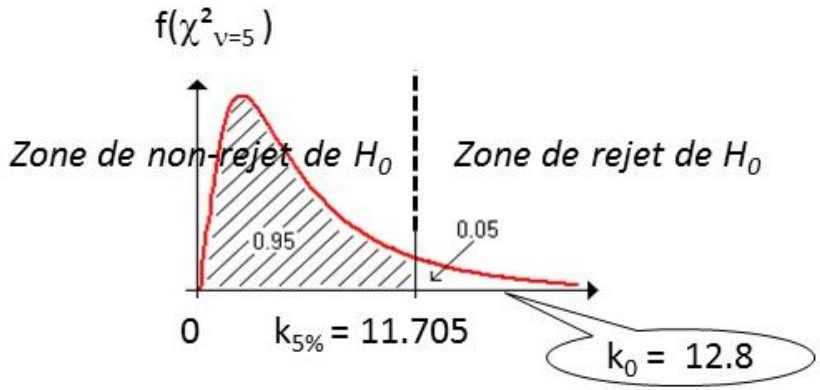
Peut-on conclure, au seuil de signification $\alpha = 5\%$, que le dé est bien équilibré ?

1. Hypothèses statistiques :
2. Seuil de signification - Conditions d'application du test - Degré de liberté :
3. Statistique de test :
4. Calcul de la statistique de test sous l'hypothèse nulle H_0 :
5. Règle de décision et conclusion :

Exemple d'application 3 - Solution

Résultat	1	2	3	4	5	6	Total
Effectif observé n_i	14	16	28	30	18	14	120
Fréquence théorique	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1
Effectif théorique $\widehat{n_i}$	20	20	20	20	20	20	120

1. Hypothèses statistiques $\begin{cases} H_0 : \textit{Le dé est équilibré, répartition uniforme des résultats} \\ H_1 : \textit{Le dé n'est pas équilibré, il est pipé} \end{cases}$
2. Seuil de signification - Conditions d'application du test - Degré de liberté $\alpha = 5\%$ Un échantillon aléatoire de taille n = 120 observé selon un caractère qualitatif à $k = 6$ modalités - classes. $\nu = (k-1) = 5\, d.d.l.$, aucun paramètre estimé
3. Statistique de test : $\sum_{i=1}^{k=6} \frac{(n_i - \widehat{n_i})^2}{\widehat{n_i}} \hookrightarrow \chi^2_{5\, d.d.l.}$
4. Calcul de la statistique de test sous l'hypothèse nulle H_0 : Le dé est équilibré - Répartition uniforme - Equiprobabilité : $k_0 = \sum_{i=1}^{k=6} \frac{(n_i - \widehat{n_i})^2}{\widehat{n_i}} = \frac{(14-20)^2}{20} + \frac{(16-20)^2}{20} + \cdots + \frac{(14-20)^2}{20} = 12.8$



5. Décision et conclusion : fractile de la loi du khi-deux à $\nu = 5$ d.d.l. cf.table: $\chi^2_{\nu=5;\alpha=5\%} = 11.705$.
La valeur k_0 appartient à la zone de Rejet de H_0. En effet, $k_0 = 12.8 > \chi^2_{5;5\%} = 11.705$. Au seuil de signification $\alpha = 5\%$, le dé semble plutôt "pipé" et semble afficher une prépondérence pour les valeurs 3 et 4.

Exemple d'application 4

- Test d'ajustement à une loi théorique "Normale".
- Essai de fiabilité de 70 composants électroniques très sensibles. On a noté la durée de vie (en jours) jusqu'à défaillance.

Classe de jours : x_i	$[22-26[$	$[26-30[$	$]30-34[$	$[34-38[$	$[38-42[$	Total
effectif observé : n_i	8	14	22	20	6	70

a) Estimer les paramètres inconnus m et σ^2 : la moyenne et la variance de la durée de vie des composants.

b) Peut-on admettre, au seuil de signification $\alpha = 5\%$, que la durée de vie des composants suit une loi normale ?

1. Hypothèses statistiques :
2. Seuil de signification - Conditions d'application du test - Degré de liberté :
3. Statistique de test :
4. Calcul de la statistique de test sous l'hypothèse nulle H_0 :
5. Règle de décision et conclusion :

Exemple d'application 4 - Solution 1/3

1. a) Estimations ponctuelles des paramètres inconnus m et σ^2 :

Centre de classe de jours : c_i	24	28	32	36	40	Total
effectif observé : n_i	8	14	22	20	6	70

Calculs : $\overline{x}_{70} = 32$; $s^2_{70} = 4.53^2 \Rightarrow s^{*2}_{70} = 4.56^2$: r = 2 paramètres estimés

2. b) Loi théorique d'ajustement : $X \rightarrow N(m = 32\,; \sigma^2 = 4.56^2 \Leftrightarrow U = \frac{(x-32)}{4.56} \rightarrow N(0\,;1)$

Hypothèses statistiques

$\begin{cases} H_0 : \textit{La durée de vie des composants suit une loi normale} \text{N}(\,32\,;\,4.56^2\,) \\ H_1 : \textit{La durée de vie des composants ne suit pas une loi normale} \text{N}(\,32\,;\,4.56^2\,) \end{cases}$.

$x_1 \leq X < x_2$	$u_1 = \frac{(x_1-32)}{4.56}$	$u_2 = \frac{(x_2-32)}{4.56}$	$p_i = P(u_1 \leq U < u_2)$	$\widehat{n}_i = np_i$	n_i
$X < 22$		-2.19	0.014	0.98	0
$22 \leq X < 26$	-2.19	-1.32	0.079	5.53	8
$26 \leq X < 30$	-1.32	-0.44	0.237	16.59	14
$30 \leq X < 34$	-0.44	0.44	0.340	23.80	22
$24 \leq X < 38$	0.44	1.32	0.237	16.59	20
$38 \leq X < 42$	1.32	2.19	0.079	5.53	6
$X \geq 42$	2.19		0.014	0.98	0

Les calculs :

$p_1 = P(X < 22) = P(U < -2.19) = \Phi(-2.19) = 1 - \Phi(2.19) = 1 - 0.9857 = 0.014$

$p_2 = P(22 \leq X < 26) = P(-2.19 \leq U < -1.32) = \Phi(-1.32) - \Phi(-2.19) = 0.079$

etc.

$p_i = P(x_1 \leq X < x_2) = P(u_1 \leq U < u_2) = \Phi(u_2) - \Phi(u_1)$

$p_7 = P(X \geq 42) = P(U \geq 2.19) = 1 - \Phi(2.19) = 1 - 0.9857 = 0.014$

Exemple d'application 4 - Solution 2/3

Classe x_i	$]-\infty;22[$	$[22-26[$	$[26-30[$	$]30-34[$	$[34-38[$	$[38-42[$	$[42+\infty[$
Observé n_i	0	8	14	22	20	6	0
Théorique $\widehat{n_i}$	0.98	5.53	16.59	23.80	16.59	5.53	0.98

On regroupe les classes à faible effectif (< 5) $\Rightarrow k = 5$ *classes*

Classe de jours : x_i	$]-\infty;26[$	$[26-30[$	$]30-34[$	$[34-38[$	$[38;+\infty[$	Total
effectif observé : n_i	8	14	22	20	6	70
effectif théorique : $\widehat{n_i}$	6.51	16.59	23.80	16.59	6.51	70

1. Seuil de signification - Conditions d'application du test - Degré de liberté $\alpha = 5\%$ Un échantillon aléatoire de taille n = 70 observé selon un caractère qualitatif à $k = 5$ modalités - classes. $(k-1-r) = 2\, d.d.l.$, avec $r = 2$ paramètres estimés.
2. Statistique de test : $\sum_{i=1}^{k=5} \frac{(n_i-\widehat{n}_i)^2}{\widehat{n}_i} \hookrightarrow \chi^2_{2\, d.d.l.}$
3. Calcul de la valeur de la statistique de test k_0 sous l'hypothèse nulle H_0 : La durée de vie des composants suit une loi normale $k_0 = 1.622$
4. Décision et conclusion : fractile de la loi du $\chi^2_{table,\nu=2,\alpha=5\%}$ (cf. table): $\chi^2_{2;\alpha=5\%} = 5.99$. La valeur k_0 appartient à la zone de Non-Rejet de H_0.
   En effet, $k_0 = 1.622 < \chi^2_{2;5\%} = 5.99$. Au seuil de signification $\alpha = 5\%$, on peut conclure que la durée de vie des composants suit une loi normale de moyenne 32 jours avec un écart-type de 4.56 jours.

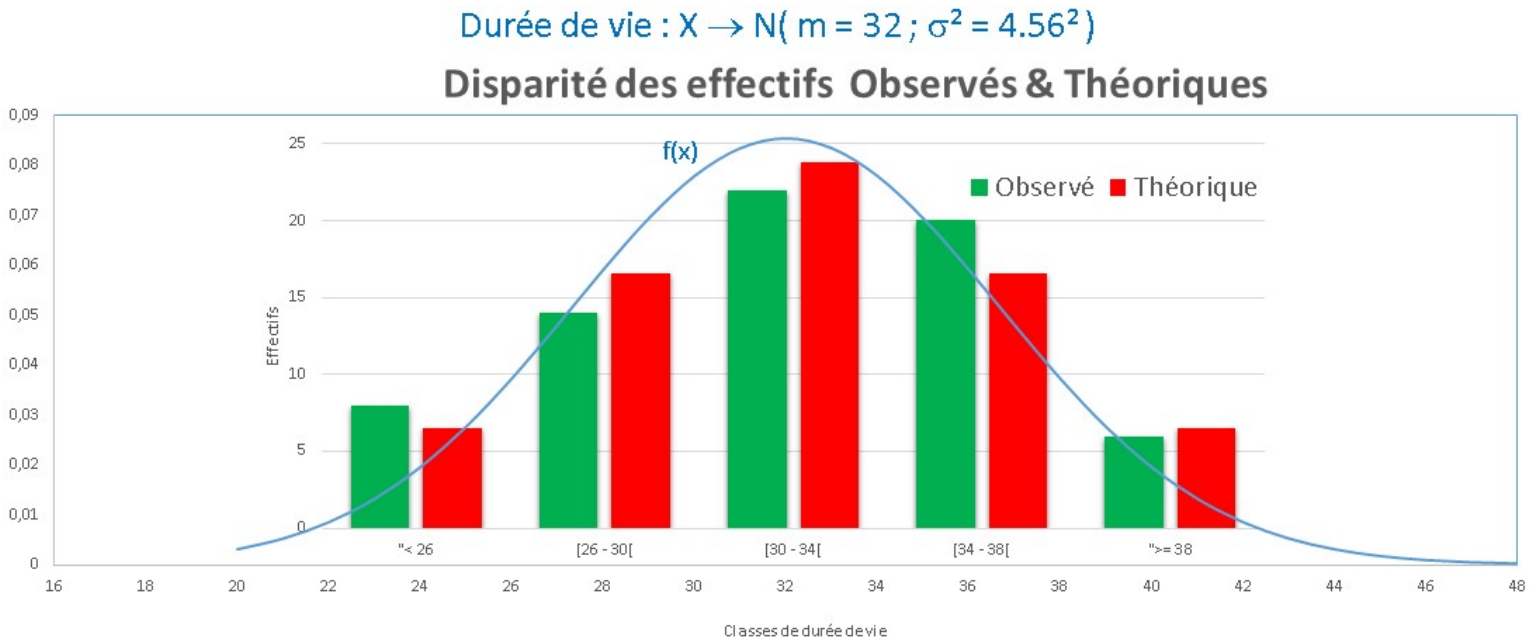
Exemple d'application 4 - Solution 3/3

- Le rejet de H_0 peut amener à 3 situations :

1. La distribution de la durée de vie ne suit pas une loi normale.
2. La moyenne de la durée de vie n'est pas de 32 jours.
3. L'écart-type de la durée de vie n'est pas égal à 4.56 jours.

Les conclusions 2) et 3) devraient se vérifier par les tests appropriés sur une moyenne et sur une variance.

- En pratique, il est plus fréquent de préciser uniquemenet la forme du modèle sans spécifier des valeurs particulières pour les paramètres du modèle.



Résumé des différents tests du Khi-deux

- Dans les trois cas précédents, pour comparer les répartitions théorique et observée, on calcule, sous l'hypothèse nulle H_0 la quantité : $k_0 = \sum_j^p \sum_k^q \frac{(n_{jk} - \widehat{n}_{jk})^2}{\widehat{n}_{jk}}$

  1. Test d'indépendance - H_0 : les variables X_p et Y_q sont indépendantes.

     Distribuée sous H_0 selon une loi du $\chi^2_{(p-1)(q-1)\, d.d.l.}$.
  2. Test d'homogénéité - H_0 :

     Distribuée sous H_0 selon une loi du $\chi^2_{(r-1)(k-1)\, d.d.l.}$
  3. Test d'ajustement - H_0 :

     Distribuée sous H_0 selon une loi du $\chi^2_{(k-1-r)\, d.d.l.}$.

- Décision et conclusion du test statistique :

  L'hypothèse nulle H_0 est rejetée, au niveau α choisi, si $k_0 \geq \chi^2_\alpha$. Le test statistique est toujours unilatéral risque à droite.

2 Tests de Wilcoxon et Mann-Withney

- Cas de 2 échantillons indépendants.
- Le test de rangs de Wicoxon et Mann-Whitney consiste à définir une règle de décision concernant la validité de l'hypothèse relative à l'identité des distributions d'une variable mesurable dans deux populations P_x et P_y. Cela revient à tester si deux populations ont la même loi de probabilité ou non.
- Conditions d'utilisation :
  1. 2 échantillons aléatoires, prélevés de façon indépendante dans deux populations P_x et P_y d'où sont prélevées respectivement m et n $(m \leq n)$ observations indépendantes.
  2. La variable étudiée peut être continue ou non.

  L'idée est que si on prélève 2 échantillons indépendants $(x_1, x_2, , x_m)$ et $(y_1, y_2, , y_n)$ dans 2 populations de même loi, les observations des 2 échantillons doivent être harmonieusement mélangées si on les regroupe.

2.1 Test de Wilcoxon - Echantillons indépendants

1. On classe alors par ordre croissant les observations de la réunion des deux échantillons : $(x_1, x_2, , x_m, y_1, y_2, , y_n)$.
2. On calcule ensuite la somme des rangs R_i (i = 1, m) occupés par les x_i, plus petit groupe dans ce classement $(m \leq n)$. Cette somme sera notée $W_x = \sum_{i=1}^{m} R_i$ (statistique de Wilcoxon).

Hypothèses statistiques : on fait généralement un test bilatéral,

$$\begin{cases} H_0 : \text{les deux distributions sont identiques.} \\ H_1 : \text{les deux distributions ne sont pas identiques.} \end{cases}$$

Sous l'hypothèse nulle H_0 (distributions identiques), la statistique de Wilcoxon ne dépend que des tailles m et n des deux échantillons, puisqu'il s'agit de dénombrer les sommes de m entiers choisis parmi les $m + n$ premiers entiers.

- On peut également calculer la somme des rangs R_j (j = 1, n) occupés par les y_j : cette statistique de test est notée $W_y = \sum_{j=1}^{n} R_j$.

Test de Wilcoxon - Echantillons indépendants

- Remarque 1 :
  - Pour de petites valeurs raisonnables de m et de n, les valeurs de la statistique W_x de Wilcoxon sont tabulées (cf. table de Wicoxon).
  - Si $(n+m \geq 20)$, la statistique W_x de Wilcoxon sous H_0 suit approximativement une loi normale de façon satisfaisante dès que m et n sont supérieurs à 10, de paramètres :

$$E(W_x) = \frac{m(m+n+1)}{2} \quad \text{et} \quad V(W_x) = \frac{n.m(m+n+1)}{12}.$$

$$U = \frac{W_x - E(W_x)}{\sqrt{V(W_x)}} \rightarrow N(0;1)$$

- Remarque 2 : En principe, pour des observations de distributions continues, la possibilité d'avoir deux valeurs égales a une probabilité nulle. Cependant, des ex-aequo viennent souvent des arrondis de mesures, les rangs de ces ex-aequo sont alors remplacés par le rang moyen des rangs qu'ils auraient eus si ils avaient été distincts.

Tables de Wilcoxon - Echantillons indépendants

Test bilatéral

Tests Unilatéraux

m	$\alpha = 5\%$	$\alpha = 1\%$		m	$\alpha = 5\%$	$\alpha = 1\%$
6	0	*		6	2	*
7	2	*		7	2	*
8	3	0		8	5	*
9	5	1		9	8	2
10	8	3		10	10	4
11	10	5		11	13	7
12	13	9	$\widetilde{m}$: la médiane de $X - Y$.	12	17	9
13	17	9	W_X la somme des rangs des différences positives.	13	21	12
14	21	12		14	25	15
15	25	15	Valeur critique $w*$ tel que : $P(W_X \leq w*)$.	15	30	19
16	29	19	m : taille de l'échantillon de X, le plus petit échantillon.	16	35	23
17	34	23	n : taille de l'échantillon de Y, $(m \leq n)$.	17	41	27
18	40	27	$*$ indique que le test ne peut être significatif.	18	47	32
19	46	32		19	53	37
20	52	37		20	60	43
21	59	43		21	68	48
22	66	49		22	75	53
23	73	55		23	83	61
24	81	61		24	92	68
25	89	68		25	101	76

Test unilatéral "risque à gauche"

$$\begin{cases} H_0 : \widetilde{m} = 0 \\ H_1 : \widetilde{m} < 0. \end{cases}$$

Rejet de H_0 si $W_x \leq w*$

Test bilatéral

$$\begin{cases} H_0 : \widetilde{m} = 0 \\ H_1 : \widetilde{m} \neq 0. \end{cases}$$

Rejet de H_0 si $W_x \leq w*$

Test unilatéral "risque à droite"

$$\begin{cases} H_0 : \widetilde{m} = 0 \\ H_1 : \widetilde{m} > 0. \end{cases}$$

Rejet de H_0 si $W_x > w*$

2.2 Test U de Mann-Whitney - Echantillons indépendants

- Le test de rangs couramment utilisé et appelé test U de Mann-Whitney. Il est équivalent au test W de Wicoxon.
- On peut montrer le lien entre les statistiques de test W_x et W_y de Wilcoxon avec celles U_x et U_y de Mann-Whitney :

$$U_x = W_x - \frac{m(m+1)}{2} \qquad \text{et} \qquad U_y = W_y - \frac{n(n+1)}{2}.$$

- La statistique U_x, pour les échantillons indépendants $(x_1, x_2, , x_m)$ et $(y_1, y_2, , y_n)$, correspond au nombre de couples d'observations (x_i, y_j) pour lesquels $x_i > y_j$ (i = 1,m et j = 1,n).

$$U_x = \sum_{i=1}^{m}(R_i - i) = \sum_{i=1}^{m} R_i - \sum_{i=1}^{m} i = W_x - \frac{m(m+1)}{2}$$

- Dans le cas de la statistique W_x de Wicoxon, on vérifie que les valeurs du premier échantillon sont plus souvent plus petites que celles du second échantillon.
- Les deux statistiques de test W_x et U_x sont donc strictement équivalentes.

Test U de Mann-Whitney - Echantillons indépendants

- Quelques propriétés des statistiques U_x et U_y
  - $U_x + U_y = m \times n$
  - $U_y = m \times n - U_x$
  - U_x prend ses valeurs entre 0 et $m \times n$
  - Sous H_0 : $E(U_x) = E(U_y) = \frac{m \times n}{2}$
  - Variable de décision : $U = min(U_x, U_y) \leq \frac{m \times n}{2}$

Décision et interprétation du test U

- Le test de Mann-Whitney est l'équivalent non-paramétrique du test de Student pour 2 échantillons indépendants. Les hypothèses statistiques portent sur les médianes :

$$H_0 : \widetilde{m_1} = \widetilde{m_2} \quad \text{contre} \quad H_1 : \widetilde{m_1} \neq \widetilde{m_2}$$

- Dans le cas de petits échantillons, on utilisera la table U de Mann-Whitney, avec un risque d'erreur α choisi, on calcule la statistique $U = min(U_x, U_y)$, puis on détermine la valeur critique cf. Table U :

1) Test bilatéral symétrique : $u_{\frac{\alpha}{2}} \rightarrow$ Rejet de H_0 Si $U \leq u_{\frac{\alpha}{2}}$

2) Test unilatéral - risque à gauche : $u_\alpha \rightarrow$ Rejet de H_0 Si $U \leq u_\alpha$

3) Test unilatéral - risque à droite : $u_\alpha \rightarrow$ Rejet de H_0 Si $U > u_\alpha$

Décision et interprétation du test U

- Dans le cas de grands échantillons $m+n \geq 20$, on peut montrer également que la statistique U_x sous H_0 suit approximativement une loi normale de paramètres :

$$\mathrm{E}(U_x) = E(W_x) - \frac{m(n+1)}{2} = \frac{nm}{2} \qquad \text{et} \qquad V(U_x) = V(W_x) = \frac{n.m(m+n+1)}{12}.$$

$$U = \frac{U_x - E(U_x)}{\sqrt{V(U_x)}} = \frac{U_x - \frac{nm}{2}}{\sqrt{\frac{n.m(m+n+1)}{12}}} \rightarrow N(0;1)$$

Dans le cas d'un test bilatéral symétrique :

Si $|U| \leq u_{\frac{\alpha}{2}}$ alors Non-rejet de H_0, sinon Rejet de H_0.

Où, $u_{\frac{\alpha}{2}}$ désigne le fractile de la loi normale centrée réduite au seuil de signification α choisi.

Table U du test de Mann-Whitney

Echantillons indépendants - Valeurs seuils ou critiques

Test bilatéral : $\alpha = 5\%$: Valeurs critiques $\rightarrow$ $u_{\frac{\alpha}{2}} = u_{2.5\%}$ et $u_{1-\frac{\alpha}{2}} = u_{97.5\%}$

Test unilatéral $\alpha = 2.5\%$ "Risque à gauche" : Valeur critique $\rightarrow$: $u_{\alpha} = u_{2.5\%}$

Test unilatéral $\alpha = 2.5\%$ "Risque à droite" : Valeur critique $\rightarrow$: $u_{1-\alpha} = u_{97.5\%}$

m \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	*	*	*	*	*	*	0	0	0	0	1	1	1	1	1	2	2	2	2
3		*	*	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5				2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6					5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7						8	10	12	14	16	18	20	22	24	26	28	30	32	34
8							13	15	17	19	22	24	26	29	31	34	36	38	41
9								17	20	23	26	28	31	34	37	39	42	45	48
10									23	26	29	33	36	39	42	45	48	52	55
11										30	33	37	40	44	47	51	55	58	62
12											37	41	45	49	53	57	61	65	69
13												45	50	54	59	63	67	72	76
14													55	59	64	69	74	78	83
15														64	70	75	80	85	90
16															75	81	86	92	98
17																87	93	99	105
18																	99	106	112
19																		113	119
20																			127

m : taille du plus petit échantillon (m < n) ; $*$ indique que le test ne peut être significatif au niveau $\alpha = 5\%$

Conclusion : Rejet de H_0 si la plus petite valeur des deux statistiques (test bilatéral) ou celle qui est appropriée (test unilatéral), est supérieure ou égale à la valeur lue dans la table.

Exemple d'application 5

- Le responsable des ressources humaines (RH) a fait passer un test d'aptitude à 7 hommes et 8 femmes, les notes se présentent comme suit :

Hommes	8.5	12	9.5	7.5	10	12.5	10.5	
Femmes	8.0	9	8	11.5	7	12	11	13

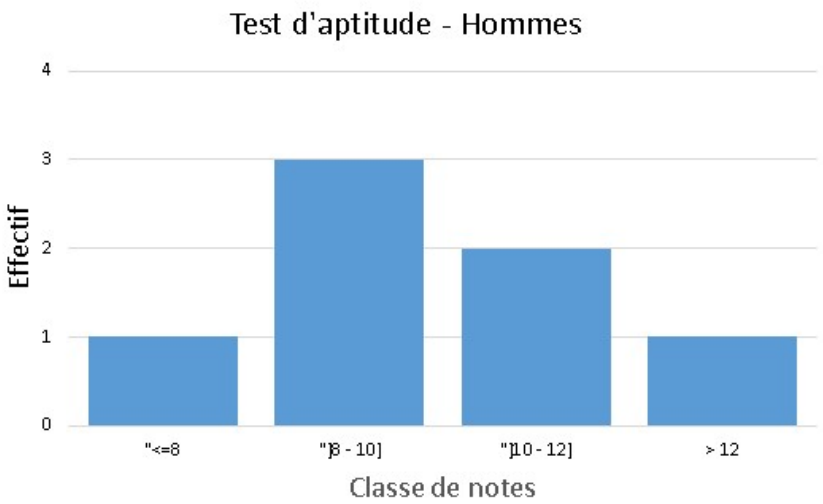
Peut-on affirmer, avec un risque d'erreur $\alpha = 5\%$, qu'il y a une différence réelle entre la performance des hommes et celle des femmes ?

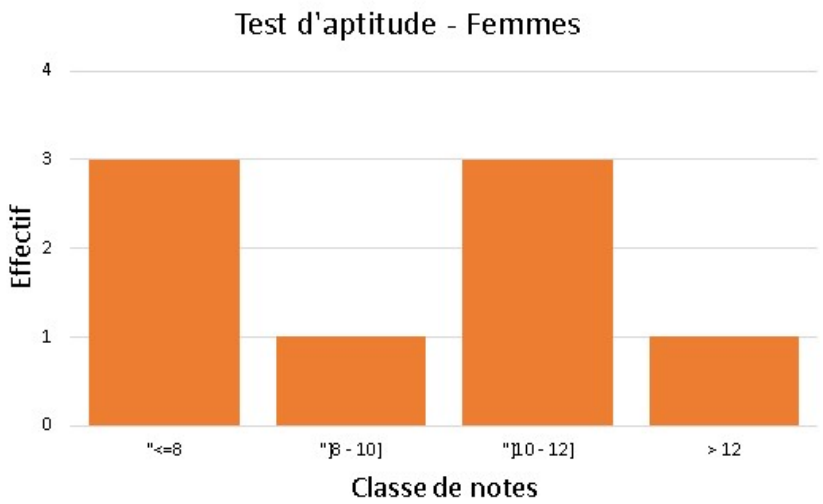
Hypothèses statistiques : on établit un test bilatéral symétrique,

$$\begin{cases} H_0 : \text{Les distributions des performances sont identiques} \\ H_1 : \text{Les distributions des performances sont différentes.} \end{cases}$$

Exemple d'application 5 - Solution 1/3

- Distributions des notes du test d'aptitude





Exemple d'application 5 - Solution 2/3

Solution - Test de Wilcoxon

X : Hommes	8.5	12	9.5	7.5	10	12.5	10.5		m = 7
Y : Femmes	8.0	9	8	11.5	7	12	11	13	n = 8
Rang X : R_i	5	12.5	7	2	8	14	9		$\sum_{i=1}^{7} R_i = 57.5$
Rang Y : R_j	3.5	6	3.5	11	1	12.5	10	15	$\sum_{j=1}^{8} R_j = 62.5$

$$\begin{cases} H_0 : \text{Les distributions des performances sont identiques} \\ H_1 : \text{Les distributions des performances sont différentes.} \end{cases}$$

- Statistique de test de Wilcoxon : $m = 7$ et $n = 8$; $\alpha = 5\%$; $W_x = \sum_{i=1}^{7} R_i = 57.5$

  Valeur critique - Test bilatéral : $w^* = 2$ m = 7 *et* $\alpha = 5\%$ cf. table Wilcoxon

- Règle de décision : Rejet de H_0 si $W_x \leq w*$

  Conclusion : $W_x = 57.5 > w* = 2 \Rightarrow$ Non-Rejet de H_0 : Les performances des hommes et des femmes ne sont pas différentes.

Exemple d'application 5 - Solution 3/3

Solution - Test U de Mann-Whitney

X : Hommes	8.5	12	9.5	7.5	10	12.5	10.5		m = 7
Y : Femmes	8.0	9	8	11.5	7	12	11	13	n = 8
Rang X : R_i	5	12.5	7	2	8	14	9		$\sum_{i=1}^{7} R_i = 57.5$
Rang Y : R_j	3.5	6	3.5	11	1	12.5	10	15	$\sum_{j=1}^{8} R_j = 62.5$

$$\begin{cases} H_0 : \text{Les distributions des performances sont identiques} \\ H_1 : \text{Les distributions des performances sont différentes.} \end{cases}$$

- Test U de Mann-Whitney : $m = 7$ et $n = 8$; $\alpha = 5\%$:
  On calcule les valeurs des 2 statistiques de test :
  $$U_x = W_x - \frac{m(m+1)}{2} = 57.5 - \frac{7\times 8}{2} = 29.5 \quad \text{et} \quad U_y = W_y - \frac{n(n+1)}{2} = 62.5 - \frac{8\times 9}{2} = 26.5$$
  La variable de décision ou observée : $u_{obs} = min(U_x, U_y) = 26.5$ est la plus petite valeur des 2 statistiques de test qui doit être comparée à la valeur seuil ou critique de la table.
- Test bilatéral $\alpha = 5\%$ - Sous l'hypothèse nulle H_0, il existe une valeur critique $u_{\frac{\alpha}{2}} = u_{2.5\%} = 10$. cf. Table U de Mann-Whitney avec $m = 7$ et $n = 8$.
- Règle de décision du test : Rejet de H_0 si $u_{obs} = min(U_x; U_y) \leq u_{\frac{\alpha}{2}} = u_{2.5\%}$.
  Conclusion : $u_{obs} = 26.5 > u_{2.5\%} = 10$; $u_{obs} = 26.5$ appartient à la zone de Non-Rejet de H_0. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que les performances des hommes et des femmes sont identiques.

3 Test de Wilcoxon - Echantillons dépendants

- Cas de 2 échantillons dépendants (appariés).
- Ce test de la somme des rangs des différences positives de Wicoxon, couramment appelé "Test des signes" d'observations appariées consiste à définir une règle de décision concernant la validité de l'hypothèse relative à l'identité des distributions de 2 variables observées sur des échantillons appariés.
- Ce test permet de savoir si les médianes de 2 échantillons appariés sont comparables. Soit les individus sont identiques dans les deux échantillons, soit ils sont appariés en fonction de variables externes.
- Soit (x_1, y_1), (x_2, y_2), , (x_n, y_n) l'ensemble des n couples d'observations. On veut savoir si les deux distributions sont identiques ou pas.

Test de Wilcoxon - Echantillons dépendants

- Conditions d'utilisation :
  1. Un échantillon aléatoire de n paires d'observations indépendantes.
  2. La grandeur observée sur chaque observation doit être telle que l'on puisse lui attribuer un ordre dans chaque paire d'observations.

- Application du test :
  1. On calcule toutes les différences $d_i = x_i - y_i$ pour i = 1,n que l'on classe ensuite par ordre croissant en valeur absolue.
  2. On détermine ensuite :
     T^+ : la somme des rangs des différences positives,
     T^- : la somme des rangs des différences négatives.

Test de Wilcoxon - Echantillons dépendants

- Remarques :

  - Toutes les différences nulles $d_i = (x_i - y_i) = 0$ doivent être supprimées et la valeur de n est réduite en conséquence.

  - Aux différences égales, on attribue le rang moyen des rangs qui auraient été attribués si elles étaient distinctes.

  - $T^+ + T^- = \frac{n(n+1)}{2}$ et sous H_0, $T^+ = T^- = \frac{n(n+1)}{4}$ (n excluant les différences nulles).

- Hypothèses statistiques : on établit généralement un test bilatéral,

$$\begin{cases} H_0 : \text{Les deux distributions sont identiques} \\ H_1 : \text{Les deux distributions sont différentes.} \end{cases}$$

Test de Wilcoxon - Echantillons dépendants

- Hypothèses statistiques :

$$\begin{cases} H_0 : \text{Les deux distributions sont identiques} \\ H_1 : \text{Les deux distributions sont différentes.} \end{cases}$$

- Décision et interprétation du test : pour un risque d'erreur α donné,

- Si $n < 20$, on utilise la table du T de Wilcoxon (échantillons appariés) pour déterminer la valeur critique T_α au niveau de signification α donné.

$$\text{Rejet de } H_0 \text{ si } \; T = min(T^+, T^-) \leq T_\alpha.$$

- Si $n \geq 20$, sous H_0, la statistique de test $T = min(T^+, T^-)$ de Wilcoxon suit approximativement une loi normale de moyenne $E(T) = \frac{n(n+1)}{4}$ et de variance $V(T) = \frac{n(n+1)(2n+1)}{24}$.

$$\text{Rejet de } H_0 \text{ si } \; \mid U = \frac{T - E(T)}{\sqrt{V(T)}} = \frac{(T - \frac{n(n+1)}{4})}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \mid > u_{\frac{\alpha}{2}}$$

Où, $u_{\frac{\alpha}{2}}$ est le fractile de la loi normale centrée réduite $U \to N(0,1)$ au seuil de signification α choisi.

Table Statistique de Wilcoxon - Echantillons dépendants

Valeurs critiques du T de Wilcoxon - Echantillons appariés.

α	5%	2.5%	1.0%	0.5%
α^*	10%	5%	2.0%	1.0%
n				
6	2	0		
7	2	2	0	
8	5	3	2	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	10	97
13	21	17	13	90
14	25	21	16	12
15	30	25	20	15
16	35	29	24	19
17	41	34	28	23
18	47	40	33	27
19	53	46	38	32
20	60	52	43	37
21	68	59	49	43
22	75	66	56	49
23	83	73	62	55
24	92	81	69	61
25	101	89	77	68

n : nombre de différences non nulles
α : Niveau de signification (test unilatéral)
α^* : Niveau de signification (test bilatéral)

Hypothèses statistiques : Test bilatéral symétrique
H_0 : les échantillons ont des distributions identiques
H_1 : les échantillons ont des distributions différentes.

Exemple : n = 12 , $\alpha^* = 5\%$ Test bilatéral
Rejet de H_0 si : $T = min(T^+, T^-) \leq T_{\alpha^*} = 13$

Exemple d'application 6

- Comparaison de 2 méthodes d'apprentissage.

N° couple	Méthode 1	Méthode 2
1	14	12
2	13.5	14.5
3	11	12.5
4	10	10
5	9	9.5
6	14	12.5
7	7	11
8	15.5	12

Peut-on conclure, au seuil de signification $\alpha = 5\%$, qu'il y a une différence significative entre les deux méthodes d'apprentissage ?

Hypothèses statistiques :

$$\begin{cases} H_0 : \text{les deux méthodes d'apprentissage sont équivalentes} \\ H_1 : \text{les deux méthodes d'apprentissage sont différentes.} \end{cases}$$

Exemple d'application 6 - Solution

N° couple	Méthode 1 x_i	Méthode 2 y_i	Différence $d_i = x_i - y_i$	Rang $\lvert d_i \rvert$	Rang +	Rang -
1	14	12	2	5	5	
2	13.5	14.5	-1	2		2
3	11	12.5	-1.5	3.5		3.5
4	10	10	0			
5	9	9.5	-0.5	1		1
6	14	12.5	1.5	3.5	3.5	
7	7	11	-4	7		7
8	15.5	12	3.5	6	6	
Total					$T^+ = 14.5$	$T^- = 13.5$

Remarques :

- La différence du couple n°4 est nulle, ce couple d'observations est supprimé, il reste donc $n = 7$ couples
- Pour $n = 7$, $T^+ + T^- = 28 = \frac{n(n+1)}{2}$

Valeur critique pour $n = 7$ et $\alpha = 5\%$, test bilatéral : $T_\alpha = 2$ (cf. table Wilcoxon échantillons appariés.)

$T = min(T^+, T^-) = min(14.5; 13.5) = 13.5 > T_\alpha = 2 \Rightarrow$ Non Rejet de H_0. On peut donc considérer avec un risque d'erreur $\alpha = 5\%$, que les deux méthodes d'apprentissages sont équivalentes.

4 Test de corrélation de rangs de Spearman

- Test relatif au cas de 2 échantillons appariés $(x_1, x_2, \ldots, x_n)$ et $(y_1, y_2, \ldots, y_n)$. Les couples (x_i, y_i) doivent être indépendants les uns des autres.
- Le but est de déceler une éventuelle dépendance "monotone" ou indépendance entre les deux variables X et Y.
- Ce test consiste à calculer la corrélation qui peut exister entre 2 échantillons de données de rangs. Il mesure donc un degré d'association ou de dépendance non linéaire, à la différence du coefficient de corrélation linéaire de Bravais-Pearson,
- Le coefficient de corrélation de rangs de Spearman, noté R_S, prend en compte les "rangs" des variables et non les "valeurs" numériques des 2 variables, comme cela est le cas du coefficient de corrélation linéaire de Bravais-Pearson.

Application du test de rangs de Spearman

1. A la série des couples de valeurs $(x_1, y_1), \ldots, (x_n, y_n)$ est associée la série des couples de rangs $(s_1, q_1), \ldots, (s_n, q_n)$.
   - s_i étant le rang de l'observation x_i suivant l'ordre décroissant dans l'échantillon $(x_1, x_2, \ldots, x_n)$.
   - q_i étant le rang de l'observation y_i suivant l'ordre décroissant dans l'échantillon $(y_1, y_2, \ldots, y_n)$.
2. On calcule les écarts des rangs de chaque paire d'observations :
   $$d_i = s_i - q_i \quad \text{(i=1, n)},$$
3. Puis la somme des carrés des écarts : $\sum_{i=1}^{n} d_i^2$.
4. La valeur du coefficient de corrélation de rangs de Spearman est bornée : $$-1 \leq \quad r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2 \quad \leq +1 \ .$$

   La seule hypothèse nulle H_0 que l'on peut tester avec R_S est que les 2 variables concernées sont mutuellement indépendantes.

- Remarque : lorsqu'il n'y a pas d'ex-aequo, le coefficient de rangs de Spearman est égal au coefficient de corrélation linéaire de Bravais-Pearson calculé entre les séries de rangs : $r_S = r(S, Q)$

Test de corrélation de rangs de Spearman

- Hypothèses statistiques :

$$\begin{cases} H_0 : \text{les deux variables sont indépendantes} \\ H_1 : \text{les deux variables sont dépendantes.} \end{cases}$$

Sous l'hypothèse H_0, la statistique exacte de R_S ne dépend que de n, puisqu'il s'agit d'énumérer les $n!$ permutations équiprobables des rangs.

- Statistiques de test : pour un risque d'erreur α donné,

- Si $n \leq 10$: la distribution de R_S est tabulée (cf. table statistique des valeurs critiques r_S^*).

- Si $n \geq 30$: Sous l'hypothèse H_0, la distribution de R_S suit approximativement une loi normale de paramètres : $E(R_S) = 0$ et $V(R_S) = \frac{1}{(n-1)}$,

$$\text{Statistique de test : } R_S\sqrt{(n-1)} \to N(0,1)$$

- Si $11 \leq n < 30$: Sous l'hypothèse H_0, la distribution de R_S suit approximativement une loi de Student à $(n-2)$ degrès de liberté.

$$\text{Statistique de test : } R_S\sqrt{\frac{(n-2)}{(1-R_S^2)}} \to T_{n-2\ d.d.l.}$$

Décision et interprétation du test de Spearman

- Hypothèses statistiques :

H_0 Indépendance

contre

H_1 : Dépendance négative (test unilatéral risque à gauche)

H_1 : Dépendance quelconque

H_1 : Dépendance positivee (test unilatéral risque à droite)

Rejet de l'hypothèse nulle H_0, au niveau de signification $\alpha = 5\%$ si :

	$(n \leq 10)$ Table de Spearman	$(11 \leq n < 30)$ Table de Student T_{n-2}	$(n > 30)$ Table N(0 , 1)
H_1	$r_S < -r_S^*$	$r_S\sqrt{\frac{(n-2)}{(1-r_S^2)}} < -t_\alpha$	$r_S\sqrt{(n-1)} < -u_\alpha$
H_1	$\mid r_S \mid > r_S^*$	$\mid r_S\sqrt{\frac{(n-2)}{(1-r_S^2)}} \mid > t_{\frac{\alpha}{2}}$	$\mid r_S\sqrt{(n-1)} \mid > u_{\frac{\alpha}{2}}$
H_1	$r_S > r_S^*$	$r_S\sqrt{\frac{(n-2)}{(1-r_S^2)}} > t_\alpha$	$r_S\sqrt{(n-1)} > u_\alpha$

Exemple d'application 7.1

- On a demandé à 2 experts d'évaluer 10 projets en attribuant une note de 0 à 20, les notes se présentent comme suit :

Note Expert 1 : X	16	15	12	9	13	10	8	10	10	7
Note Expert 2 : Y	9	8	9	10	7	15	13	14	12	11

Peut-on affirmer, avec un risque d'erreur $\alpha = 5\%$, qu'il existe une corrélation "négative" entre les 2 évaluations ?

Hypothèses statistiques : on établit un test unilatéral risque à gauche,

$$\begin{cases} H_0 : R_S = 0 & \text{Les évaluations sont indépendante} \\ H_1 : R_S < 0 & \text{Les évaluations sont dépendantes "négativement - opposition".} \end{cases}$$

Exemple d'application 7.1 - Solution

Note Expert 1 : X	16	15	12	9	13	10	8	10	10	7
Note Expert 2 : Y	9	8	9	10	7	15	13	14	12	11
Rang X : s_i	1	2	4	8	3	6	9	6	6	10
Rang Y : q_i	7.5	9	7.5	6	10	1	3	2	4	5
Ecart : d_i	-6.5	-7	-3.5	2	-7	5	6	4	2	5
$Ecart^2$: d_i^2	42.25	49	12.25	4	49	25	36	16	4	25

- La somme des carrés des écarts : $\sum_{i=1}^{n} d_i^2 = 262.5$
- Coefficient de corrélation de rangs de Spearman :
  $r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2 = 1 - \frac{6}{10\times 99} \times 262.5 = -0.591$.
- Valeur critique - Cf. Table statistique de Spearman pour $n = 10$; $\alpha = 5\%$; $r_S^* = -0.564$

  Conclusion : avec un risque $\alpha = 5\%$, $r_S = -0.591 < r_S^* = -0.564 \Rightarrow$ Rejet de H_0. On peut donc conclure qu'il existe bien une dépendance "négative" des notations. Les 2 experts évaluent les projets de manière différente "opposition".

Exemple d'application 7.2

- On veut étudier s'il existe une corrélation entre la qualité de l'air dans les villes et le nombre de maladies pulmonaires qui y sont détectées.

  Pour ce faire, on a effectué un classement de 11 villes sur la base de ces 2 critères.

Ville	A	B	C	D	E	F	G	H	I	J	K
Rang - Qualité de l'air	4	7	9	1	2	10	3	5	6	8	11
Rang - Nbre de Maladies	5	4	7	3	1	11	2	10	8	6	9

Peut-on affirmer, avec un risque d'erreur $\alpha = 5\%$, qu'il existe une corrélation entre les 2 critères ?

Hypothèses statistiques : on établit un test bilatéral symétrique,

$$\begin{cases} H_0 : R_S = 0 & \text{Indépendance - Pas de corrélation} \\ H_1 : R_S \neq 0 & \text{Dépendance - Corrélation significative.} \end{cases}$$

Exemple d'application 7.2 - Solution

Ville	A	B	C	D	E	F	G	H	I	J	K
Rang s_i - Qualité de l'air	4	7	9	1	2	10	3	5	6	8	11
Rang q_i - Nombre de Maladies	5	4	7	3	1	11	2	10	8	6	9
Ecart : $d_i = s_i - q_i$	-1	3	2	2	1	-1	1	-5	-2	2	2
$Ecart^2$: d_i^2	1	9	4	4	1	1	1	25	4	4	4

- La somme des carrés des écarts : $\sum_{i=1}^{n} d_i^2 = 58$
- Coefficient de corrélation de rangs de Spearman :
  $r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2 = 1 - \frac{6}{11\times 120} \times 58 = 0.736$.
- Valeur critique - Cf. Table statistique de Spearman pour $n = 11$; $\alpha = 5\%$; $r_S^* = \pm 0.618$

  Conclusion : avec un risque $\alpha = 5\%$, $r_S = 0.736 \notin [-0.618, 0.618] \Rightarrow$ Rejet de H_0. On peut donc conclure qu'il existe bien une corrélation entre la qualité de l'air et le nombre de maladies pulmonaires.
- Remarque : La valeur $r_S = 0.736$ du coefficient de corélation suggère une association positive "attraction" importante entre les 2 variables. $H_1 : R_S > 0$ Corrélation positive.
  En effet, pour un test unilatéral risque $\alpha = 5\%$ à droite, $r_S^* = 0.536$ (cf. table) et comme $r_S = 0.736 > r_S^* = 0.536 \Rightarrow$ Rejet de H_0. On peut donc conclure que la qualité de l'air et le nombre de maladies pulmonaires sont corrélés positivement.
- Pour $n = 11 (\geq 10)$, on peut également utiliser l'approximation par une loi de Student $T_{n-2=9 \, d.d.l.}$.

Table du coefficient de corrélation R_S

Valeurs critiques r_S* - Test bilatéral symétrique - $P(\mid R_S \mid > r_S^*) = \alpha$.

n	$\alpha = 5\%$	$\alpha = 2\%$
5	1.000	1.000
6	0.886	0.943
7	0.786	0.893
8	0.738	0.833
9	0.700	0.783
10	0.648	0.745
11	0.618	0.709
12	0.587	0.671
13	0.560	0.648
14	0.538	0.622
15	0.521	0.604
16	0.503	0.582
17	0.485	0.566
18	0.472	0.550
19	0.460	0.535
20	0.447	0.520
21	0.435	0.508
22	0.425	0.496
23	0.415	0.486
24	0.406	0.476
25	0.398	0.466
26	0.390	0.457
27	0.382	0.448
28	0.375	0.440
29	0.368	0.433
30	0.362	0.425

n	Taille des échantillons
α	Niveau de signification
Type de test	Bilatéral

Exemple - Hypothèses statistiques :	Test bilatéral symétrique
$H_0 : R_S = 0$	Indépendance - Pas de corrélation
$H_1 : R_S \neq 0$	Dépendance - Corrélation.

Pour $n = 10$ $\alpha = 5\%$ $r_S^* = \pm 0.648$	Rejet de H_0 si $R_S \notin [-0.648, 0.648]$

Table du coefficient de corrélation R_S

Valeurs critiques r_S* - Test unilatéral - $P(R_S < -r_S^*) = P(R_S > +r_S^*) = \alpha$.

n	$\alpha = 5\%$	$\alpha = 1\%$
4	1.000	
5	0.900	1.000
6	0.829	0.943
7	0.714	0.893
8	0.643	0.833
9	0.600	0.783
10	0.564	0.745
11	0.536	0.709
12	0.506	0.671
13	0.484	0.648
14	0.456	0.622
15	0.443	0.604
16	0.425	0.582
17	0.414	0.566
18	0.399	0.550
19	0.391	0.35
20	0.377	0.520
21	0.370	0.508
22	0.359	0.496
23	0.353	0.486
24	0.343	0.476
25	0.337	0.466
26	0.329	0.457
27	0.324	0.448
28	0.317	0.440
29	0.312	0.433
30	0.306	0.425

n	Taille des échantillons
α	Niveau de signification
Type de tests	Unilatéraux

Exemple - Hypothèses statistiques :	Test Unilatéral risque à droite
$H_0 : R_S = 0$	Indépendance - Pas de corrélation
$H_1 : R_S > 0$	Dépendance - Corrélation positive "attraction".

Pour $n = 10$ $\alpha = 5\%$	Rejet de H_0 si $R_S > r_S^* = +0.564$

5 Test de Kappa

- Le coefficient Kappa de Cohen permet d'évaluer l'accord entre $K \geq 2$ jugements ou techniques lorsque les réponses sont qualitatives.

- Le Coefficient κ de concordance de Kappa mesure le degré d'accord entre $K = 2$ juges. Il est défini d'une façon générale par :

$$\kappa = \frac{\Pi_o - \Pi_e}{1 - \Pi_e} \quad \text{avec} \quad \begin{cases} \Pi_o : \text{proportion de concordance observée} \\ \Pi_e : \text{proportion de concordance aléatoire} \end{cases}$$

- Dans le cas de deux $K = 2$ juges, les données peuvent être représentées par une table de contingence $I \times I$, où I est le nombre de classes dans lesquelles les n objets peuvent être classés par les deux juges. Il est alors estimé par :

$$\widehat{\kappa} = \frac{\widehat{P}_o - \widehat{P}_e}{1 - \widehat{P}_e} \quad \text{avec} \quad \begin{cases} \widehat{P}_o = \sum_{i=1}^{I} p_{ii} = \sum_{i=1}^{I} \frac{n_{ii}}{n} \\ \widehat{P}_e = \sum_{i=1}^{I} p_{i.} p_{.i} = \sum_{i=1}^{I} \frac{n_{i.} n_{.i}}{n^2}. \end{cases}$$

- Nous formulons ensuite, l'hypothèse nulle $H_0 : \kappa = 0$: indépendance d'accord ou de concordance.

Test de Kappa de Cohen

- La concordance est d'autant plus élevée que la valeur κ tend vers $+1$, parfaite ou maximale si $\kappa = 1$. Il est égal à -1 dans le cas d'une discordance parfaite.
- Interprétation - Degré d'accord et valeur de Kappa

Accord	Excellent	Bon	Modéré	Médiocre	Mauvais	Très mauvais
$\widehat{\kappa}$	> 0.81	$[0.80 - 0.61]$	$[0.60 - 0.41]$	$[0.40 - 0.21]$	$[0.20 - 0.0]$	< 0.0

- Avec des fréquences marginales constantes et sous l'hypothèse que les deux juges classent d'une façon complètement arbitraire, le coefficient $\widehat{\kappa}$ suit approximativement une loi normale :

$$\text{Moyenne nulle : } \mathrm{E}(\widehat{\kappa}) = 0 \quad ; \quad \text{Variance estimée : } V(\widehat{\kappa}) = \frac{\widehat{P}_o(1-\widehat{P}_o)}{n(1-\widehat{P}_e)^2}.$$

$$\text{Statistique de test sous } H_0 : U = \frac{\widehat{\kappa} - E(\widehat{\kappa})}{\sqrt{V(\widehat{\kappa})}} = \widehat{\kappa}\sqrt{\frac{n(1-\widehat{P}_e)^2}{\widehat{P}_o(1-\widehat{P}_o)}} \to N(0,1)$$

- Le test du coefficient Kappa est un test non paramétrique d'échantillons appariés sur données binaires. Il est l'équivalent du coefficient de corrélation de rangs de Spearman sur données continues.

Test de Kappa de Cohen

- Tableau de contingence 2×2

		Evaluateur 1		
	Réponse	Oui	Non	Total
Evaluateur 2	Oui	a	b	$n_{1.}$
	Non	c	d	$n_{2.}$
	total	$n_{.1}$	$n_{.2}$	n

- Dans le cas particulier où la variable qualitative est à 2 modalités de réponses : Tableau de contingence 2×2, le coefficient de kappa peut se calculer de la façon suivante : $\widehat{\kappa} = \frac{2 \times (a \times d - b \times c)}{n_{1.} \times n_{.2} + n_{2.} \times n_{.1}}$
- La présentation des résultats sous cette forme montre que les deux évaluateurs sont en accord pour a + d réponses avec a réponses positives concordantes et d réponses négatives concordantes, et en désaccord pour c + b réponses.
- La valeur du coefficient de kappa est indépendante de la taille de l'échantillon n, elle reste la même si on multiplie toutes les cellules du tableau par une constante, mais sa signification statistique sous l'hypothèse nulle H_0 sera plus grande.

Exemple d'application 8.1 - Contrôle Antidopage

- Résultats de 2 laboratoires indépendants de l'Agence Mondiale Antidopage (AMA) selon la variable Contrôle antidopage (Positif, Négatif).
- Anonymes et sous scellés, les prélèvements urinaires des athlètes sont envoyés au département des analyses de l'Agence Française de Lutte contre le Dopage (AFLD) à Châtenay-Malabry en région Île-de-France et dans un autre laboratoire accrédité par l'AMA.
- Tableau de contingence 2×2 - Répartition de 200 prélèvements d'athlètes choisis au hasard pour le contrôle antidopage lors d'une compétition sportive internationale.

		Laboratoire AFLD		
	Contrôle	Négatif	Positif	Total
Autre laboratoire AMA	Négatif	a = 195	b = 1	$n_{1.} = 196$
	Positif	c = 1	d = 3	$n_{2.} = 4$
	total	$n_{.1} = 196$	$n_{.2} = 4$	n = 200

$$\widehat{\kappa} = \frac{2\times(a\times d - b\times c)}{n_{1.}\times n_{.2} + n_{2.}\times n_{.1}} = \frac{2\times(195\times 3 - 1\times 1)}{196\times 4 + 4\times 196} = 0.745$$

Cette valeur indique un bon accord entre les 2 laboratoires d'analyses de l'AMA

Exemple d'application 8.2 - Vins de Bordeaux

- Deux oenologues jugent la qualité du vin de 34 années (1924 - 1957).

  La question ici, est de savoir si ces deux oenologues classent ces vins de la même façon selon la variable qualitative à 3 modalités de réponse: Qualité du vin (Bonne, Moyenne, Médiocre).

Année	1924	1925	1926	⋯	⋯	⋯	⋯	⋯	1956	1957
Juge 1	Bonne	Moyenne	Médiocre	⋯	⋯	⋯	⋯	⋯	Moyenne	Bonne
Juge 2	Bonne	Médiocre	Moyenne	⋯	⋯	⋯	⋯	⋯	Bonne	Bonne

- Le tableau de contingence suivant donne la répartition des 34 vins selon les classements des deux oenologues.

Juge 1 / Juge 2	Bonne	Moyenne	Médiocre	Total
Bonne	8	1	0	9
Moyenne	2	10	1	13
Médiocre	0	1	11	12
Total	10	12	12	34

$$\widehat{\kappa} = \frac{\widehat{P}_o - \widehat{P}_e}{1-\widehat{P}_e} = 0.778 \Rightarrow \text{Bon degré d'accord ou de concordance des jugements}$$

avec $\begin{cases} \widehat{P}_o = \sum_{i=1}^{3} p_{ii} = \sum_{i=1}^{3} \frac{n_{ii}}{n} = \frac{29}{34} = 0.853 \\ \widehat{P}_e = \sum_{i=1}^{3} p_{i.}p_{.i} = \sum_{i=1}^{3} \frac{n_{i.}n_{.i}}{n^2} = \frac{9\times10+13\times12+12\times12}{34^2} = 0.337. \end{cases}$

Exemple d'application 8.3 - Paris sportif

- Deux joueurs parient sur les résultats des 10 rencontres de la $17^{ème}$ journée de championnat de la ligue 1 de football. Les 2 joueurs ont-ils des pronostics identiques ? La variable qualitative à réponse unique :

  Pronostic (1 : Domicile, N : Match nul, 2 : Extérieur).

1 : Equipe à domicile 2 : Equipe à l'extérieur	Lyon Rennes	Bordeaux Monaco	PSG Nice	⋯ ⋯	⋯ ⋯	⋯ ⋯	⋯ ⋯	⋯ ⋯	Bastia Metz	Dijon Marseille
Joueur 1	1	1	N	⋯	⋯	⋯	⋯	⋯	1	N
Joueur 2	N	2	1	⋯	⋯	⋯	⋯	⋯	1	N

- Le tableau de contingence suivant les pronostics des deux parieurs.

Joueur 1 \ Joueur 2	1	N	2	Total
1	2	2	1	5
N	1	1	0	2
2	2	0	1	3
Total	5	3	2	10

$\widehat{\kappa} = \frac{\widehat{P}_o - \widehat{P}_e}{1-\widehat{P}_e} = 0.048 \Rightarrow$ Mauvais accord - Indépendance de concordance des pronostics

avec $\widehat{P}_o = \sum_{i=1}^{3} p_{ii} = \sum_{i=1}^{3} \frac{n_{ii}}{n} = 0.40$ et $\widehat{P}_e = \sum_{i=1}^{3} p_{i.} p_{.i} = \sum_{i=1}^{3} \frac{n_{i.} n_{.i}}{n^2} = 0.37.$