

STATISTIQUES APPLIQUÉES À LA BIOLOGIE : LES TESTS NON PARAMÉTRIQUES.

Alexandre Popier

L3 BBTE, Université du Maine, Le Mans

Semestre 1

1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

PARAMÉTRIQUE ET NON PARAMÉTRIQUE :

- Mathématiques :
 - Paramétrique : porte sur un paramètre (proportion, moyenne, variance, etc.), c'est-à-dire un nombre (ou un vecteur).
Tests de signe : paramétriques (portent sur la médiane).
 - Non paramétrique : porte sur une densité, une loi, etc.
 χ^2 et Kolmogorov : non paramétriques.
- Biologie :
 - Non paramétrique : sans condition sur la loi des données ou libre de distribution.

POINT COMMUN :

- ▶ Test avec une **erreur de seconde espèce pas (ou mal) contrôlée !**
- ▶ Raison : **hypothèse alternative trop vaste.**

TABLEAU RÉSUMÉ.

Test			Conditions
Proportion	Comparaison	P	$n \leq 30$: loi binomiale $n \geq 30$: $\approx \mathcal{N}(0, 1)$ $n \geq 30$: $\approx \mathcal{N}(0, 1)$
Moyenne ou variance (inclus comparaison)		P	$n \geq 30$: $\approx \mathcal{N}(0, 1)$
Indépendance ou homogénéité	χ^2	NP	effectifs attendus E supérieurs à 5
Adéquation	χ^2 K-S	NP	$E \geq 5$, loi discrète $n \geq 8$, loi continue
Médiane (inclus comparaison)	signe ou Wilcoxon	NP	
Comparaison sur $k > 2$ échantillons	Kruskal-Wallis	NP	

1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

EXEMPLE 1 : DERNIER CHIFFRE DES POIDS.

QUESTIONNAIRE DE SANTÉ : sous estimation et arrondi du poids réel.

- 80 personnes interrogées indépendamment.

Dernier chiffre	Effectif	Dernier chiffre	Effectif
0	35	5	24
1	0	6	1
2	2	7	4
3	1	8	7
4	4	9	2

- En principe les 10 chiffres sont équiprobables.

DÉFINITION

Un *test d'adéquation* est utilisé pour tester que la distribution des fréquences observées est en adéquation avec (conforme à) une distribution donnée.

EXEMPLE 2 : GÉNOTYPE DES ENFANTS.

90 enfants sélectionnés aléatoirement.

Génotype	Effectif observé	Effectif attendu
AA	22	$90 \times 0,25 = 22,5$
Aa	55	$90 \times 0,50 = 45,5$
aa	13	$90 \times 0,25 = 22,5$

- AA : enfant sain.
- Aa : enfant porteur.
- aa : enfant malade.

EXEMPLE 3 : VACCIN DE SALK.

- Contre la polyomélite.
- Étude faite en 1954 à Chicago.
- Tableau de contingence suivant :

	Polio paralysante ?	
	Oui	Non
Groupe traitement vaccin de Salk	33	200712
Groupe placebo	115	201114

DÉFINITION

Un *test d'indépendance* teste l'hypothèse nulle qu'il n'y a pas d'association entre la variable ligne et la variable en colonne dans un tableau de contingence.

EXEMPLE 4 : SEXE ET RÉPONSE AUX SONDAGES.

SONDAGE

- avec question posée :
« L'avortement est un sujet privé qui devrait être laissé à la décision des femmes sans intervention du gouvernement. »
- Sondes hommes interrogent 800 hommes.
- Sondes femmes interrogent 400 hommes.

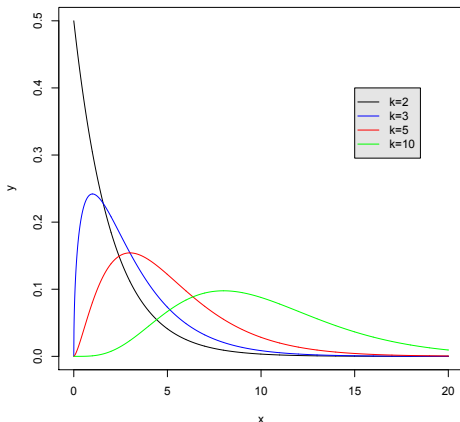
	Sexe du sondeur	
	Homme	Femme
Hommes qui sont d'accord	560	308
Hommes qui sont en désaccord	240	92

QUESTION : la réponse du sondé dépend-elle du sexe du sondeur ?
Ici on parle parfois d'homogénéité : les différentes populations ont-elles les mêmes proportions que certaines caractéristiques ?

BASE COMMUNE : LA LOI DU χ^2 .

La loi du chi-deux à k degrés de liberté est donnée par la densité :

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, \quad x > 0.$$



1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

QUELQUES NOTATIONS ET HYPOTHÈSES.

EFFECTIFS OBSERVÉS ET ATTENDUS :

- 1 O : les effectifs observés des résultats.
- 2 E : les effectifs attendus.
- 3 k : nombre de catégories pour les résultats.
- 4 n : nombre total de résultats.

REMARQUE : effectifs attendus calculés comme suit : multiplication du nombre total par la probabilité de la catégorie.

CONDITIONS À REMPLIR :

- Les données sont sélectionnées aléatoirement.
- Les données d'échantillon sont des comptages des effectifs pour chacune des différentes catégories.
- Pour chaque catégorie **l'effectif attendu est au moins 5** (aucune condition sur l'effectif observé).

STATISTIQUE DE TEST ET CONCLUSION.

STATISTIQUE DE TEST :

$$\chi^2 = \sum \frac{(O - E)^2}{E}.$$

- χ^2 : loi du chi deux à $k - 1$ degrés de liberté (k : nombre de catégories).
- Tests toujours unilatéraux à droite.

CONCLUSION :

- Grande adéquation entre les valeurs : petite valeur de χ^2 ou grande P-valeur.
- Seuil à lire dans la table ou via R en utilisant `qchisq`.

ERREUR DE SECONDE ESPÈCE : dans tous les tests du χ^2 , on ne la contrôle pas ! Accepter H_0 est une faute !

EXEMPLE 1 : POIDS.

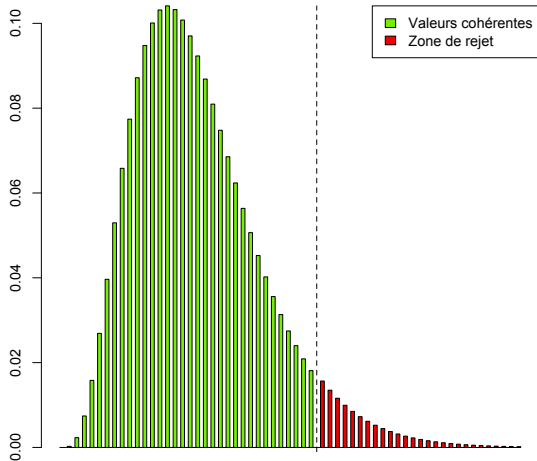
TEST : on pense que les gens trichent et arrondissent leur poids.

- $H_0 = \{p_0 = p_1 = \dots = p_9\}$: pas d'arrondi, chiffres équiprobables.
- H_1 : au moins une des probabilités est différente des autres.
- Niveau $\alpha = 5\%$: seuil pour $k - 1 = 9$: **16,92**
(`qchisq(0.05, 9, lower.tail=F)`).
- $n = 80$ personnes interrogées.

Dernier chiffre	Effectif observé O	Effectif attendu E	$\frac{(O - E)^2}{E}$
0	35	8	91,125
1	0	8	8
2	2	8	4,5
⋮	⋮	⋮	⋮
Totaux	80	80	$\chi^2 = 156,6$

CONCLUSION : $\chi^2 > 16,92$: on rejette H_0 .

EXEMPLE 1 : POIDS.



EXEMPLE 2 : GÉNOTYPE.

TRAITEMENT SOUS R pour un niveau de 1%.

```
O<-c(22, 55, 13)
E<-90*c(0.25,0.5,0.25)
chideux<-sum((O-E)^2/E)
chideux
seuil<-qchisq(0.01,2,lower.tail=F)
seuil
Pvaleur<-pchisq(chideux,2,lower.tail=F)
Pvaleur
p<-E
chisq.test(O,p=p, rescale.p=T)
```

QU'EN CONCLUEZ-VOUS ?

1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

RAPPEL : EXEMPLE 3.

- Contre la poliomyélite.
- Étude faite en 1954 à Chicago.
- Tableau de contingence suivant :

	Polio paralysante ?	
	Oui	Non
Groupe traitement vaccin de Salk	33	200712
Groupe placebo	115	201114

TEST D'INDÉPENDANCE :

- H_0 : variables de ligne et de colonne indépendantes.
- H_1 : variables de ligne et de colonne dépendantes.

CONDITIONS À REMPLIR :

- Les données sont sélectionnées aléatoirement.
- Les données d'échantillon sont des comptages des effectifs.
- Pour chaque catégorie l'effectif attendu est au moins 5 (aucune condition sur l'effectif observé).

FRÉQUENCES MARGINALES ET CONDITIONNELLES.

DÉFINITION

La somme des effectifs de chaque ligne (ou colonne) donne les **effectifs marginaux**. Ils permettent de décrire la distribution de la variable placée en ligne (ou en colonne).

En divisant les effectifs marginaux par l'effectif total du tableau (somme de toutes les cellules), on obtient les **fréquences marginales**.

RETOUR SUR L'EXEMPLE 3 :

	Polio paralysante ?		Effectifs	Fréquences
	Oui	Non		
Traitement	33	200712	200745	49,9%
Placebo	115	201114	201229	50,1%
Effectifs	148	401826	401974	
Fréquences	0,037%	99,963%		100%

DÉFINITION

L'étude séparée des lignes du tableau s'appelle *étude conditionnelle*. On calculera les fréquences pour chaque ligne. On parle de *fréquences conditionnelles des lignes*.

RETOUR SUR L'EXEMPLE 3 : en pourcentage

	Polio paralysante ?		Total
	Oui	Non	
Traitement	0,016	99,984	100
Placebo	0,029	99,971	100

DÉFINITION

L'indépendance est la situation extrême dans une table de contingence où il n'existe aucune relation entre les deux variables.

Les fréquences conditionnelles des lignes sont alors toutes identiques et égales à la fréquence marginale des colonnes.

PROPRIÉTÉ

Dans le cas d'indépendance, l'effectif de chaque cellule se retrouve en utilisant uniquement les effectifs marginaux :

$$\text{Effectif de la cellule } (i, j) = \frac{\text{total ligne } i \times \text{total colonne } j}{\text{total tableau}}.$$

INDÉPENDANCE STATISTIQUE.

- Pour chaque cellule, on calcule cette valeur.
- Elle est considérée comme l'effectif attendu sous l'hypothèse d'indépendance.

RETOUR SUR L'EXEMPLE 3 : tableau des effectifs attendus E

	Polio paralysante ?	
	Oui	Non
Traitement	$\frac{200745 \times 148}{401974} = 73,91$	200671,09
Placebo	74,089	201154,91

TEST D'INDÉPENDANCE DU χ^2 .

DÉFINITION

La *statistique du χ^2 d'indépendance de Pearson* consiste à faire

$$\chi^2 = \sum \frac{(O - E)^2}{E}.$$

χ^2 suit la loi du chi deux à $(r - 1)(c - 1)$ degrés de liberté où r est le nombre de lignes, c celui de colonnes.

Plus sa valeur est importante, plus les deux variables sont liées.

Valeurs numériques :

- $$\chi^2 = \frac{(33 - 73,91)^2}{73,91} + 0,008 + \dots + 0,008 = 45,252.$$
- Seuil critique 3,841 (`qchisq(0.05, 1, lower.tail=F)`).
- **Conclusion** : recevoir le vaccin n'est pas indépendant d'avoir la polio.

RETOUR SUR L'EXEMPLE DE LA POLIO.

CODE R :

```
O<-c(33,200712,115,201114)
Obs<-matrix(O,2,2,byrow=T)
colS<-colSums(Obs)
ligS<-rowSums(Obs)
tot<-sum(colS)
E<-c(ligS[1]*colS[1],ligS[1]*colS[2],...
      ...ligS[2]*colS[1],ligS[2]*colS[2])/tot
chideux<-sum((O-E)2/E)
chideux
seuil<-qchisq(0.05,1,lower.tail=F)
seuil
Pvaleur<-pchisq(chideux,1,lower.tail=F)
Pvaleur
chisq.test(Obs,correct=F)
```

À faire à titre d'entraînement !

QUESTION : détecte-t-on des lésions cérébrales chez les sportifs subissant régulièrement des chocs à la tête par contact ?

ÉTUDE sur le football américain et le soccer par IRM (1997) :

Groupes	Oui	Non
Foot. am.	7	11
Soccer	11	4
Contrôles	5	15

RECHERCHE d'une association entre activité et risque de dysfonctionnement.

RAPPEL EXEMPLE 4 : SEXE ET RÉPONSE AUX SONDAGES.

SONDAGE

	Sexe du sondeur	
	Homme	Femme
Hommes qui sont d'accord	560	308
Hommes qui sont en désaccord	240	92

Test : indépendance entre réponse et sexe du sondeur.

Autre formulation : **homogénéité**

- H_0 : proportions de réponses en accord/désaccord identiques pour les sujets sondés par des hommes et pour les sujets sondés par des femmes.
- H_1 : proportions différentes.

MISE EN PLACE DU TEST.

CALCULS : identiques au test d'indépendance !

EXERCICE :

- 1 Retrouver les effectifs attendus (entre parenthèses)

	Sexe du sondeur	
	Homme	Femme
Accord	560 ($E = 578,67$)	308 (289,33)
Désaccord	240 (221,33)	92 (110,67)

- 2 En déduire la valeur de χ^2 . Quelle est la valeur seuil ?

CONCLUSION : Les réponses et le sexe du sondeur sont dépendants, donc les hommes semblent influencés par le sexe du sondeur.

1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

UN AUTRE TEST D'ADÉQUATION.

TEST D'ADÉQUATION : $H_0 = \{\mathcal{L} = \mathcal{L}_0\}$ vs. $H_1 = \{\mathcal{L} \neq \mathcal{L}_0\}$.

- \mathcal{L} : loi des données.
- \mathcal{L}_0 : loi fixée.
- Exemple : $\mathcal{L}_0 =$ normale $\mathcal{N}(2, 4)$.

TEST DU CHI DEUX : possible pour les lois discrètes

- La loi \mathcal{L}_0 spécifique un nombre fini de valeurs avec une probabilité pour chaque valeur.

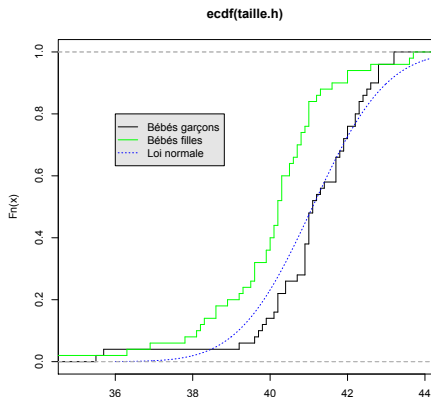
TEST DE KOLMOGOROV-SMIRNOV :

- Complémentaire car valable pour une loi \mathcal{L}_0 avec densité.
- Programmé en R via la fonction `ks.test`.

EXEMPLE.

PÉRIMÈTRE CRANIEN des nouveaux-nés garçons.

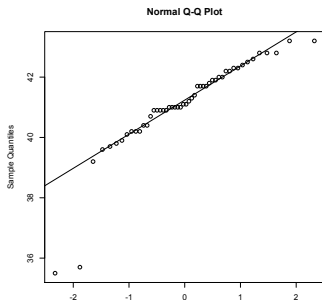
- Récupérer sur UMTICE le fichier MHED.txt et l'importer dans R.
- Calculer la moyenne m et l'écart-type s de cet échantillon.
- Tester si cet échantillon suit une loi gaussienne de paramètres m et s^2 .



EXEMPLE.

EXTENSION :

- Ce test permet aussi de tester l'égalité en loi de deux échantillons. L'illustrer avec les données pour les filles via le fichier FHED.
- **Test de normalité** : utiliser plutôt le test de Shapiro-Wilk (`shapiro.test`).
 - "Efficace" même avec de petits échantillons.
 - Peut être complété par un QQ-plot (`qqnorm`).



- 1 TEST DU CHI DEUX (χ^2)
 - Test d'adéquation
 - Test d'indépendance
- 2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK
- 3 TESTS BASÉS SUR LES SIGNES ET LES RANGS
 - Test des rangs signés de Wilcoxon pour données appariées
 - Test de Wilcoxon avec échantillons indépendants
 - Test de Kruskal-Wallis

- 1 TEST DU CHI DEUX (χ^2)
 - Test d'adéquation
 - Test d'indépendance
- 2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK
- 3 TESTS BASÉS SUR LES SIGNES ET LES RANGS
 - Test des rangs signés de Wilcoxon pour données appariées
 - Test de Wilcoxon avec échantillons indépendants
 - Test de Kruskal-Wallis

DÉFINITION (TEST DES RANGS SIGNÉS)

Test non paramétrique qui utilise les rangs des données appariées et qui teste

- H_0 : les données appariées ont des différences qui viennent d'une population dont la médiane est nulle.
- H_1 : les données appariées ont des différences qui viennent d'une population dont la médiane n'est pas égale à 0.

Peut aussi être utilisé pour tester l'affirmation qu'un échantillon vient d'une population avec une médiane spécifique.

REMARQUE : existe en une version plus simple (test des signes) qui ne tient pas compte de l'amplitude des différences.

ÉTAPES DE CALCUL :

- 1 Calculer la différence pour chaque paire de données.
- 2 Ignorer les signes des différences, trier les différences par ordre croissant et remplacer les différences par leur rang. En cas d'égalité, affecter la moyenne des rangs.
- 3 Affecter à chaque rang le signe de la différence.
- 4 Calculer la somme des valeurs absolues des rangs négatifs et la somme des rangs positifs.

NOTATIONS :

- T : plus petite des deux sommes.
- n : nombre de paires avec une différence non nulle.

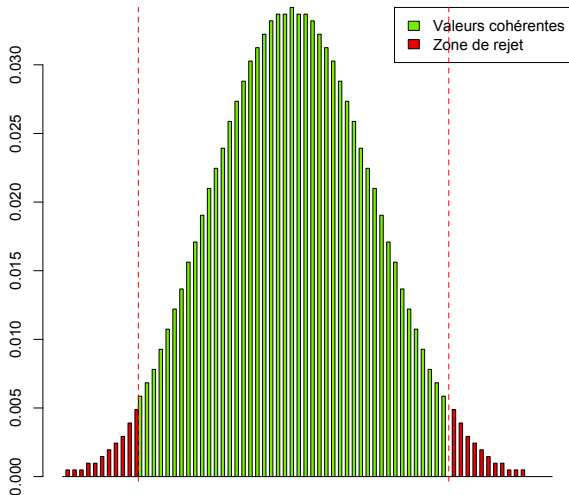
STATISTIQUE :

- $n \leq 30$: T (loi tabulée ou `dsignrank`).

- $n > 30$: $Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ (à peu près $\mathcal{N}(0, 1)$).

STATISTIQUE DE TEST.

$n = 11$, seuils 11 et 55.



EXEMPLE 1 : RENDEMENT DU MAÏS.

TEST sur la différence de rendement de graines de maïs.

- Données appariées.
- H_0 : pas de différence, autant de plus que de moins. Médiane des différences nulle.
- H_1 : différence, médiane non nulle.

EXEMPLE 1 : RENDEMENT DU MAÏS.

Rendement du maïs pour différentes graines				
Ordinaires	Séchées	Différence	Rang	Rang signé
1903	2009	-106	10	-10
1935	1915	20	1	1
1910	2011	-101	9	-9
2496	2463	33	3	3
2108	2180	-72	8	-8
1961	1925	36	4	4
2060	2122	-62	6	-6
1444	1482	-38	5	-5
1612	1542	70	7	7
1316	1443	-127	11	-11
1511	1535	-24	2	-2

EXEMPLE 1 : RENDEMENT DU MAÏS.

- Somme rangs positifs = 15, somme des rangs négatifs = 51 (valeur absolue).
- $T = 15$, $n = 11$.
- Seuils = 11 et 55.
- **Conclusion** : $11 < T < 55$, pas de différence.

CODE EN R : à partir des données déjà importées sur R.

```
diffabs<-abs(diff)
r<-rank(diffabs)
si<-sign(diff)
rs<-r*si
T<-min(sum(rs[rs>0]),-sum(rs[rs<0]))
seuils<-c(qsignrank(0.025,11),qsignrank(0.025,11,lower
Pvaleur<-2*psignrank(T,11)
wilcox.test(rend.ord,rend.sec,paired=T)
```


EXEMPLE 2.

MÉDIANE SUR UNE POPULATION.

EXEMPLE

Le jeu de données BTEMP.txt inclut des températures relevées chez des adultes (en degrés Fahrenheit). Utilisez ces 106 températures prises à midi pour tester l'affirmation que la médiane est inférieure à 37°C.

- Pour convertir les degrés Fahrenheit en Celsius, soustraire 32 et diviser par 1,8.
- Attention : test unilatéral.

1 TEST DU CHI DEUX (χ^2)

- Test d'adéquation
- Test d'indépendance

2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK

3 TESTS BASÉS SUR LES SIGNES ET LES RANGS

- Test des rangs signés de Wilcoxon pour données appariées
- Test de Wilcoxon avec échantillons indépendants
- Test de Kruskal-Wallis

TEST DE WILCOXON AVEC ÉCHANTILLONS INDÉPENDANTS.

QUELQUES REMARQUES.

- Appelé aussi **TEST DES SOMMES DES RANGS**.
- Équivalent au **TEST DE MANN-WHITNEY**.
- Test plus puissant que Kolmogorov-Smirnov ou signes.

IDÉE DU TEST :

- Rassembler les deux échantillons en un seul.
- Calculer les rangs du premier échantillon dans ce regroupement.

NOTATIONS :

- n_1 : taille de l'échantillon 1.
- n_2 : taille de l'échantillon 2.
- $R = R_1$: somme des rangs de l'échantillon 1.
- R_2 : somme des rangs de l'échantillon 2.

STATISTIQUE DE TEST.

Donnée par

- $R - \frac{n_1(n_1 + 1)}{2}$: loi de Wilcoxon (`dwilcox`).
- pour $n_1 \geq 8$ et $n_2 \geq 8$:

$$Z = \frac{R - m_R}{s_R}, \quad m_R = \frac{n_1(n_1 + n_2 + 1)}{2}, \quad s_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}.$$

Sous H_0 , approximation gaussienne $\mathcal{N}(0, 1)$.

TEST DE MANN-WHITNEY.

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}, \quad U = n_1 n_2 - \frac{n_1(n_1 + 1)}{2} - R.$$

EXEMPLE : IMC HOMME-FEMME.

À EFFECTUER SOUS R.

- Récupérer les données IMC homme et femme (fichiers HBIM.txt et FBMI.txt) sur UMTICE.

- Entrer ce qui suit

```
imc.f<-t(imc.f)
imc.h<-t(imc.h)
n1<-length(imc.f)
n2<-length(imc.h)
imc<-c(imc.f, imc.h)
r<-rank(imc)
R<-sum(r[seq(1,n1,1)])
R2<-sum(r[seq(n1+1,n1+n2,1)])
mR<-n1*(n1+n2+1)/2
sR<-sqrt((n1*n2*(n1+n2+1))/12)
```

EXEMPLE : IMC HOMME-FEMME.

À EFFECTUER SOUS R.

- Récupérer les données IMC homme et femme (fichiers HBIM.txt et FBMI.txt) sur UMTICE.
- Entrer ce qui suit
- Puis faites le test

```
RR<-R-n1*(n1+1)/2 #statistique de wilcox.test
seuils<-c(qwilcox(0.025,n1,n2),...
          ...qwilcox(0.025,n1,n2,lower.tail=F))
Pvaleur<-pwilcox(RR,n1,n2)...
          ...+pwilcox(n1*n2-RR,n1,n2,lower.tail=F)
Z<-(R-mR)/sR
Pvaleur<-2*pnorm(abs(Z),0,1,lower.tail=F)
wilcox.test(imc.f,imc.h,correct=F)
```

- 1 TEST DU CHI DEUX (χ^2)
 - Test d'adéquation
 - Test d'indépendance
- 2 TESTS D'ADÉQUATION : KOLMOGOROV-SMIRNOV ET SHAPIRO-WILK
- 3 TESTS BASÉS SUR LES SIGNES ET LES RANGS
 - Test des rangs signés de Wilcoxon pour données appariées
 - Test de Wilcoxon avec échantillons indépendants
 - Test de Kruskal-Wallis

DÉFINITION

Parfois appelé *test H*, ce test est un test non paramétrique qui utilise les rangs de données d'échantillon de trois populations indépendantes ou plus. Il est utilisé pour tester

- l'hypothèse nulle H_0 que les échantillons indépendants proviennent de populations avec des médianes égales ;
- l'alternative H_1 est l'affirmation qu'au moins une des médianes est différente des autres.

REMARQUES :

- Extension de l'ANOVA qui suppose des données gaussiennes.
- Approximation par un χ^2 de paramètre $k-1$, k est le nombre d'échantillons.
- Chaque échantillon a au moins 5 observations.

STATISTIQUE DE TEST.

PROCÉDURE :

- 1 Combiner tous les échantillons en un seul et affecter un rang à chaque valeur.
- 2 Pour chaque échantillon, calculer la taille d'échantillon et la somme des rangs.
- 3 Calculer H .

NOTATIONS :

- N : nombre total d'observations tous échantillons combinés.
- k : nombre d'échantillons.
- n_i : taille, R_i : somme des rangs pour l'échantillon i .

STATISTIQUE :

$$H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1).$$

- Test bilatéral à droite.
- $H \sim \chi^2(k-1)$.

EXEMPLE : TAILLE DES PEUPLIERS.

ÉTAPES :

- ➊ Récupérer les données du fichier `peupliers.txt`
 - 4 colonnes contenant les poids des peupliers suivant traitement reçu : sans, irrigué, fertilisé, irrigué et fertilisé.
 - Par colonne, 20 données : 5 pour le site 1 (riche, humide), l'année 1, 5 pour le site 2 (sablonneux, sec), l'année 2, 5 pour le site 1, l'année 2 et 5 pour le site 2, l'année 2.
- ➋ Calculer H en utilisant toutes les données. Vérifier que $H = 15,43$ avec un seuil de 7,81 au niveau 5%. Conclure.
- ➌ Calculer H sur les données pour le site 1 l'année 1. Conclure
- ➍ Calculer H sur les données pour le site 2 l'année 1. Conclure

Comparer vos résultats à ceux donnés par `kruskal.test`.