

Tests statistiques

Définitions et principes généraux des tests statistiques

Présentation générale

Un **test d'hypothèse** est une **règle de décision**. La décision est un pari et comporte toujours des **risques d'erreur**. A partir des **données d'un échantillon** (observations), on doit **rejeter ou non une hypothèse** statistique, c'est à dire une hypothèse portant sur la nature d'une ou plusieurs distributions, ou sur les paramètres qui leur sont attachés.

Quelles hypothèses ?

La démarche débute par l'établissement d'une hypothèse de travail. C'est cette hypothèse qui sera mise à l'épreuve, testée.

Cette hypothèse s'appelle **l'hypothèse nulle** notée H_0

Le choix de H_0 est particulier : il est fonction de ce que l'on souhaite invalider (rejeter). On supposera que H_0 est vraie et qu'une valeur observée peu probable sous H_0 contredit H_0 . Il y a une analogie avec le raisonnement par l'absurde en mathématiques.

Il existe plusieurs types d'hypothèses.

Dans le cas où l'on veut comparer une espérance mathématique μ avec une valeur théorique μ_0 , on peut faire des hypothèses telles que : $\mu = \mu_0$, $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$

La première hypothèse « $\mu = \mu_0$ » s'appelle **hypothèse simple**, les autres s'appellent **hypothèses composites**.

Nous n'aborderons que le cas où seules les hypothèses simples peuvent être testées contre **l'hypothèse composite complémentaire**, notée H_1 .

Un test conduit à **rejeter** ou à **ne pas rejeter** une hypothèse simple H_0 .

Quels risques ?

Réalité ↙	H_0 vraie	H_0 fausse
Non rejet de H_0	Correct	Manque de puissance (risque de deuxième espèce)
Rejet de H_0	Rejet à tort (risque de première espèce)	Correct

Le **risque de première espèce** est noté α .

Son interprétation est la suivante: une conclusion de rejet de l'hypothèse nulle étant un pari, **le risque que l'on prend en formulant ce pari doit être inférieur ou égal à α** , sinon, on ne le prend pas et on ne rejette pas l'hypothèse nulle.

Le **risque de deuxième espèce** est noté β . La **puissance** est définie comme $\pi = 1 - \beta$

On ne peut pas maîtriser simultanément les deux risques. Le premier est généralement considéré comme le plus lourd de conséquence et c'est ce **risque de rejet à tort** qui sera pris comme **critère de décision** à l'issue de la procédure de test.

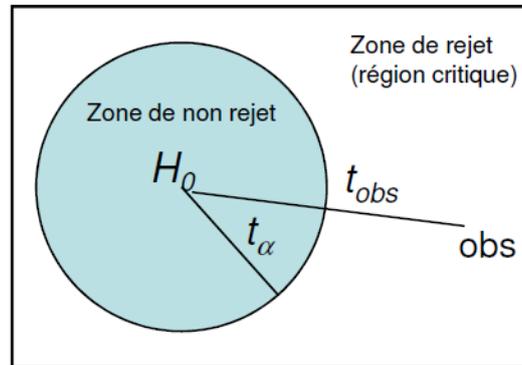
Analogies

Test statistique	Raisonnement par l'absurde	Décision de justice	Démarche scientifique
H_0	H^c	Le suspect est innocent	Th_a
H_1	H	Le suspect est coupable	Th_n
α	0	Probabilité de condamner un innocent	
β		Probabilité de relâcher un coupable	
π	Habilité du mathématicien	Habilité de l'accusation	

Méthodologie générale

Le principe est le suivant : on se place dans un espace « mathématique » abstrait (adapté au problème étudié) et on représente par deux points dans cet espace, d'une part l'hypothèse H_0 , d'autre part l'échantillon observé. On définit un « écart » entre ces deux points, t_{obs} , et on fixe un risque d'erreur α . A ce risque correspond un écart critique t_α .

Si $t_{obs} > t_\alpha$, on rejette l'hypothèse H_0 , sinon, on ne la rejette pas.



L'exécution d'un test est codifiée en trois étapes.

1^{ère} étape

Devant une situation expérimentale et des données recueillies, on doit choisir :

- Une hypothèse nulle H_0
- Un type de test statistique
- Un risque de première espèce α

Le choix du type de test statistique implique le choix de la statistique de test T (qui est une variable aléatoire permettant de déterminer t_{obs}) et de la loi de cette variable sous H_0 (permettant de déterminer t_α). Il est donc nécessaire de connaître (même approximativement) la loi de probabilité de la statistique de test T lorsque l'hypothèse H_0 est vraie.

2^{ème} étape

Il faut déterminer les valeurs dont la comparaison guidera la conclusion:

- t_{obs} est obtenu à partir des valeurs observées sur l'échantillon suivant une formule propre au test choisi : t_{obs} est la valeur de la statistique de test T calculée sur l'échantillon observé
- t_α est lu dans une table statistique spécifique au test choisi

3^{ème} étape

Conclure :

- Si t_{obs} est inférieur à t_α : on ne rejette pas H_0
- Si t_{obs} est supérieur à t_α : on rejette H_0

Niveau de signification

Le choix de α est facultatif. Si l'on n'a pas choisi de risque d'erreur, après avoir calculé t_{obs} , on cherche dans la table la valeur de α telle que t_{obs} et t_α soient égaux. On obtient ainsi le risque exact d'erreur (= niveau de signification) pour une décision de rejet (**p-value**).

Le niveau de signification de la statistique observée t_{obs} est la plus petite valeur p de α conduisant au rejet de l'hypothèse H_0 :

$$p = Prob (T \geq t_\alpha | H_0 \text{ vraie})$$

Ainsi, on rejette l'hypothèse H_0 au risque α lorsque $p \leq \alpha$.

Test unilatéral et test bilatéral

Pour une même hypothèse H_0 , il est possible de considérer deux situations permettant de définir un type d'hypothèse alternative H_1 .

Test bilatéral : on veut savoir si deux paramètres sont statistiquement différents ou non, sans s'occuper de savoir lequel est supérieur à l'autre. Le risque d'erreur doit être réparti de part et d'autre afin de tenir compte du signe de la différence.

Test unilatéral : on veut savoir si deux paramètres sont statistiquement différents ou non, mais ici le sens de la différence est important. Le risque d'erreur n'existe que d'un côté car la différence testée n'est que dans un sens.

Tests paramétriques et non paramétriques

Les **tests paramétriques** reposent sur des hypothèses concernant les distributions des variables étudiées chez les individus. La connaissance de cette distribution permet de calculer la statistique de test t_{obs} . Les tests paramétriques comparent des paramètres entre eux, en nécessitant certaines conditions sur la distribution de la variable.

Les **tests non paramétriques** ne dépendent pas de paramètres et leur application ne repose pas sur des conditions concernant les distributions des variables étudiées chez les individus. Ils sont moins puissants que les tests paramétriques, on ne les emploie que si les conditions d'applications des autres méthodes ne sont pas respectées.

Tests paramétriques pour variables quantitatives

Comparaison d'une moyenne observée à une valeur de référence

- **Test de Student** (pour $n < 30$)

On considère une population, sur laquelle est défini un caractère numérique distribué selon une **loi normale** $N(\mu, \sigma)$

On dispose d'un échantillon de taille n , sur lequel on estime l'espérance μ par la moyenne observée \bar{x} et l'écart-type σ par s .

L'hypothèse nulle H_0 à tester est **$H_0: \mu = \mu_0$** , μ_0 étant une valeur donnée.

La statistique T suit sous H_0 une loi de Student à $(n-1)$ ddl

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \qquad t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

t_α est lu dans la table de la loi de Student à $n-1$ degrés de liberté.

- **Test de l'écart réduit** ($n \geq 30$)

Pour n grand ($n > 30$), il est raisonnable de considérer σ comme égal à la valeur estimée. Dans ce cas (ou quand la variance est connue), on peut utiliser comme statistique de test :

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

t_α est lu dans la table de l'écart réduit

Remarque : La loi de Student converge vers la loi normale centrée réduite lorsque n tend vers l'infini. Ce test n'est valable que si la variable aléatoire observée est de loi normale. On peut néanmoins l'utiliser pour une loi quelconque si $n \geq 30$.

Comparaison de deux moyennes observées sur des échantillons indépendants

- **Test de Student** (pour $n < 30$)

Deux populations de **même variance** σ^2 et de moyenne μ_1 et μ_2 . La distribution de la variable, dans chacune des populations, suit une **loi normale**.

L'hypothèse nulle H_0 à tester est **$H_0 : \mu_1 = \mu_2$**

On note m_1 et m_2 les estimations des moyennes μ_1 et μ_2 , s_1^2 et s_2^2 celles des variances dans les deux échantillons.

$$t_{obs} = \frac{|m_1 - m_2|}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \quad \text{avec} \quad s^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$$

t_α est lu dans la table de la loi de Student à $(n_1 + n_2 - 2)$ ddl

- **Test de l'écart réduit** ($n \geq 30$)

Cas où n_1 et n_2 sont tous les deux grands (≥ 30)

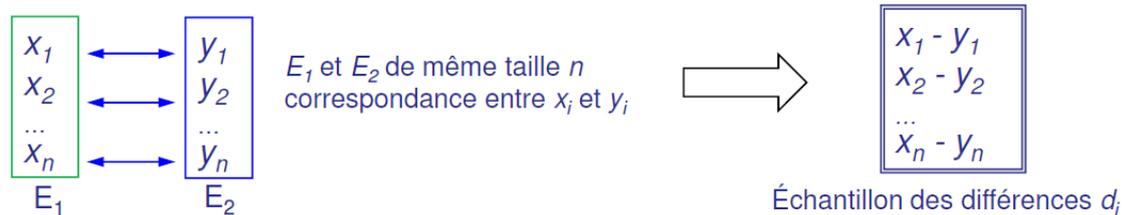
L'hypothèse nulle H_0 à tester est **$H_0 : \mu_1 = \mu_2$**

$$t_{obs} = \frac{|m_1 - m_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t_α est lu dans la table de l'écart réduit

Comparaison de deux moyennes observées sur deux séries appariées

On veut comparer deux séries d'une variable quantitative provenant d'échantillons de même taille et chaque observation d'un échantillon est liée à une observation homologue de l'autre échantillon. Chaque couple de valeur constitue une paire.



Formulation des hypothèses

- H_0 : échantillon des différences (moyenne d) est extrait au hasard d'une population de moyenne des différences $\delta = 0$
- H_1 : $\delta \neq 0$

On calcule sur l'échantillon

- La moyenne des différences
- L'écart-type de la différence moyenne

On est ramené au problème de la comparaison d'un échantillon avec une différence moyenne observée d à une population dont la différence moyenne δ vaut 0.

(Comparaison moyenne théorique / moyenne observée)

Comparaison de deux variances : Test F

On veut comparer les variances de 2 séries de variables quantitatives, on veut vérifier les conditions d'application de certains tests paramétriques qui exigent que les variances soient identiques (test de Student).

Les distributions doivent être **normales** dans les deux populations d'où proviennent les deux échantillons.

On teste le rapport des deux variances s_1^2 et s_2^2 (s_1^2 étant la plus élevée)

Sous H_0 , ce rapport est proche de 1 et les fluctuations d'échantillonnage suivent une loi de Fisher à (n_1-1) et (n_2-1) ddl.

$$t_{obs} = \frac{s_1^2}{s_2^2}$$

Comparaison de plusieurs moyennes observées (échantillons indépendants) Test de l'analyse de la variance ANOVA

⇒ Comparaison globale de plusieurs moyennes

Tester l'existence d'une liaison entre une variable qualitative à k classes et une variable quantitative X

Conditions :

- Variable X gaussienne dans la population
- X même variance dans chacune des k populations

Test pour variables qualitatives

Comparaison de deux pourcentages observés (échantillons indépendants)

- Test du χ^2 d'homogénéité

Tableau des effectifs observés

	Caractère présent	Caractère absent	Totaux	Pourcentages
Echantillon 1	a	b	n_1	$p_1 = a/n_1$
Echantillon 2	c	d	n_2	$p_2 = c/n_2$
Totaux	$t_1 = a + c$	$t_2 = b + d$	N	

Tableau des effectifs théoriques (« attendus ») sous H_0 : même pourcentage dans les deux populations

	Caractère présent	Caractère absent
Echantillon 1	$n_1 = t_1/N$	$n_1 = t_2/N$
Echantillon 2	$n_2 = t_1/N$	$n_2 = t_2/N$

La statistique de test est **une distance entre les deux tableaux**.

O_{ij} est le nombre observé dans la case ij

E_{ij} est le nombre attendu sous H_0 de la case ij

Sous H_0 , χ^2_{obs} suit asymptotiquement une loi du χ^2 à 1 ddl

Quand χ^2_{obs} est grand, on rejette H_0

L'utilisation de ce test impose que le nombre minimal des sujets dans le tableau théorique soit supérieur ou égal à 5.

- **Test de l'écart réduit**

Les proportions des deux populations peuvent être estimées par $p_1 = a/n_1$ et $p_2 = c/n_2$

Sous H_0 , ces deux proportions sont égales avec :

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Et sous les conditions $n_1 p \geq 5$, $n_1(1-p) \geq 5$, $n_2 p \geq 5$ et $n_2(1-p) \geq 5$, alors :

$$t_{obs} = \frac{|p_1 - p_2|}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

t_α lu dans la table de la loi $\mathcal{N}(0,1)$

Remarque : ce test se généralise au cas de la comparaison d'un pourcentage observé p à un pourcentage théorique P .

Si $np \geq 5$ et $n(1-p) \geq 5$:

$$t_{obs} = \frac{|p - P|}{\sqrt{\frac{p(1-p)}{n}}}$$

t_α lu dans la table de la loi $\mathcal{N}(0,1)$

- **Test exact de Fischer**

Les tests du χ^2 et de l'écart réduit ne peuvent être réalisés que sous certaines conditions relatives à la taille de l'effectif. Lorsque ces conditions ne sont pas remplies, on peut employer le test exact de Fisher.

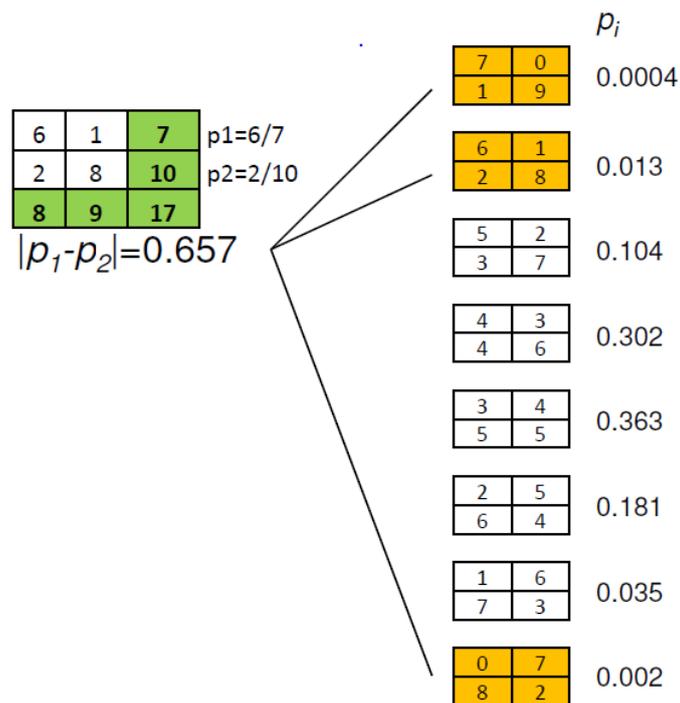
La méthode exacte de Fisher consiste à calculer, si H_0 est vraie, la probabilité d'avoir observé une configuration donnant un écart au moins aussi grand que l'écart observé entre les deux pourcentages que l'on compare. Si cette probabilité est inférieure au risque α , on rejette H_0 .

La probabilité de chaque configuration i se calcule par :

$$p_i = \frac{n_1! n_2! t_1! t_2!}{a! b! c! d! N!}$$

Exemple :

Parmi l'ensemble des configurations possibles (tableaux à marges égales au tableau observé), seuls 3 tableaux présentent une différence au moins aussi grande.



$$p = 0.0004 + 0.013 + 0.002 = 0.0154 = 1,54\%$$

La probabilité d'avoir observé une configuration donnant un écart au moins aussi grand que 65.7% est inférieure à 5%. On rejette H_0 .

Comparaison d'une distribution observée à une distribution théorique

- χ^2 d'ajustement

On désire comparer une distribution observée à une distribution théorique pour une variable qualitative répartie en r classes.

Variable	Distribution théorique (en %)	Echantillon
A_1	f_1	O_1 $E_1 = f_1 \cdot N$
...
A_r	f_r	r $E_r = f_r \cdot N$
Total	100%	N

Tous les effectifs théoriques doivent être supérieurs ou égaux à 5.

$$\chi^2_{obs} = t_{obs} = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i}$$

Sous H_0 , χ^2_{obs} suit asymptotiquement une loi du χ^2 à $(r-1)$ ddl

Comparaison de plusieurs distributions observées (échantillons indépendants)

- χ^2 d'homogénéité

On désire comparer les distributions observées entre k échantillons d'une variable à r classes.

Variable	Echantillon 1	Echantillon 2	...	Echantillon k	Total
A_1	O_{11} / E_{11}	O_{12} / E_{12}	...	O_{1k} / E_{1k}	t_1
...
A_r	O_{r1} / E_{r1}	O_{r2} / E_{r2}	...	O_{rk} / E_{rk}	t_r
Total	n_1	n_2	...	n_k	N

Tous les effectifs théoriques doivent être supérieurs ou égaux à 5.

$$\chi^2_{obs} = t_{obs} = \sum_{i,j=1}^{k,r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sous H_0 , χ^2_{obs} suit asymptotiquement une loi du χ^2 à $(r-1)(k-1)$ ddl.

Comparaison de plusieurs distributions observées (échantillons appariés)

- χ^2 de Mac Nemar

On désire comparer deux pourcentages observés sur deux échantillons appariés. Les paires peuvent être concordantes ou discordantes.

Echantillon 1	Echantillon 2	Nombre de paires
+	+	e
-	+	f
+	-	g
-	-	h

Echantillon 1		Echantillon 2	
Caractère présent	Caractère absent	Caractère présent	Caractère absent
e	f	Echantillon 1	Echantillon 2
g	h	Echantillon 1	Echantillon 2

$$\chi^2_{obs} \text{ Mac Nemar} = \frac{(f - g)^2}{f + g}$$

Suit asymptotiquement une loi du χ^2 à 1 ddl.

La condition d'application de ce test est que le nombre de paires discordantes $f+g$ doit être supérieur ou égal à 10.

Test de la relation entre deux variables qualitatives

- χ^2 d'indépendance

On désire tester le lien (H_0 : l'indépendance) entre deux variables à k et r classes.

Variable	B ₁	B ₂	...	B _k	Total
A ₁	O ₁₁ / E ₁₁	O ₁₂ / E ₁₂	...	O _{1k} / E _{1k}	t ₁
...
A _r	O _{r1} / E _{r1}	O _{r2} / E _{r2}	...	O _{rk} / E _{rk}	t _r
Total	n ₁	n ₂	...	n _k	N

Tous les effectifs théoriques doivent être supérieurs ou égaux à 5.

$$\chi^2_{obs} = t_{obs} = \sum_{i,j=1}^{k,r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sous H_0 , χ^2_{obs} suit asymptotiquement une loi du χ^2 à $(r-1)(k-1)$ ddl.

Test non paramétriques

Introduction

Un test paramétrique requiert un modèle à fortes contraintes (normalité des distributions, égalité des variances). Ces hypothèses sont d'autant plus difficiles à vérifier que les effectifs étudiés sont plus réduits.

Un test non paramétrique est un test dont le modèle ne précise pas les conditions que doivent remplir les paramètres de la population dont a été extrait l'échantillon. Les tests non paramétriques ne font aucune hypothèse quant à la loi de distribution de la variable mesurée.

Les statistiques utilisées sont souvent **basées sur les rangs** et nécessitent l'utilisation de **tables spécifiques**. Bien qu'il existe un grand nombre de tests non paramétriques, nous allons étudier trois tests : le test de Mann-Whitney, le test de Wilcoxon et le test des signes.

La problématique

La question est de savoir si deux distributions I_1 et I_2 sont identiques (sans rien supposer sur les lois).

On veut tester $H_0 : I_1 = I_2$ versus $H_1 : I_1 \neq I_2$

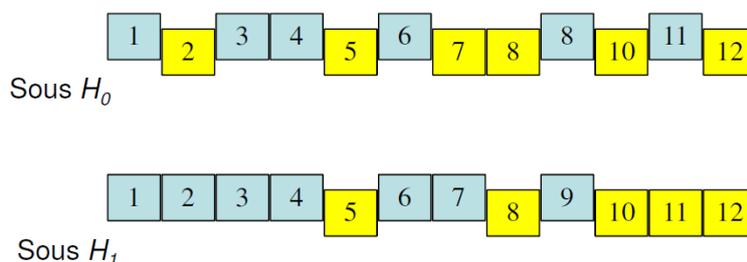
Si I_1 et I_2 sont issues de deux échantillons indépendants, on utilisera le test de Mann-Whitney.

Si I_1 et I_2 sont issues d'un même échantillon (données appariées), on utilisera le test de Wilcoxon ou le test des signes.

Les rangs

Sous l'hypothèse H_0 , les individus devraient être rangés de façon aléatoire, les valeurs d'une série alternant avec celles de l'autre.

Sous H_1 , si les valeurs d'une série à comparer sont en moyenne plus élevées, leurs rangs après classement sont en moyenne plus élevés.



Le test de Mann-Whitney

On a observé pour une variable **deux échantillons indépendants** (les tailles respectives des échantillons sont notées n_1 et n_2).

On ordonne les $n_1 + n_2$ observations par ordre croissant.

On calcule la somme des rangs pour chacun des deux échantillons, S_{rg1} et S_{rg2} .

On calcule :

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - S_{rg1}$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - S_{rg2}$$

$$U = \min(U_1, U_2)$$

Sous H_0 et pour n_1 et/ou n_2 grands, on admet que U suit une loi normale de paramètres :

$$E(U) = \frac{n_1 \cdot n_2}{2} \quad \text{et} \quad \text{Var}(U) = \frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}$$

La réalisation du test se fait donc en comparant la valeur de la statistique de test U à la valeur critique de la loi normale.

Pour des échantillons de petites tailles, on utilise des tables spécifiques.

Il existe **plusieurs méthodes de traitement des ex aequo**, la plus simple consiste à affecter le rang moyen, le reste de la procédure demeure identique.

Le test de Wilcoxon

On a observé deux variables pour un seul échantillon. La question est de savoir si les deux distributions **appariées** sont identiques.

L'échantillon de taille n est $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

On calcule $d_i = x_i - y_i$, les valeurs des différences entre les observations appariées. Le test de Wilcoxon va tester l'hypothèse que la loi des différences est symétrique autour de 0. On classe les valeurs absolues $|d_i|$ non nulles par ordre croissant. Les différences nulles sont ignorées, on soustrait de la taille de l'échantillon toutes les paires qui ne présentent pas de différence ($d_i = 0$). On note m le nombre de paires qui restent.

On calcule S_{rg+} et S_{rg-} la somme des rangs des différences positives et négatives.

On calcule $S = \min(S_{rg+}, S_{rg-})$.

Sous H_0 et pour m grand, on admet que S suit une loi normale de paramètres :

$$E(S) = \frac{m \cdot (m + 1)}{4} \quad \text{et} \quad \text{Var}(S) = \frac{m \cdot (m + 1) \cdot (2m + 1)}{24}$$

La réalisation du test se fait donc en comparant la valeur de la statistique de test S à la valeur critique de la loi normale.

Pour des échantillons de petites tailles, on utilise des tables spécifiques.

Il existe **plusieurs méthodes de traitement des ex aequo**, la plus simple consiste à affecter le rang moyen, la variance doit alors être corrigée, le reste de la procédure demeure identique. **La table peut être utilisée.**

Le test des signes

Il sert à tester si une population tend à avoir des valeurs plus grandes que celles d'une autre population. Il doit son nom au fait qu'il utilise des signes « + » ou « - » au lieu de valeurs quantitatives. Il est donc applicable même quand la série de mesures est **ordinaire**. Ce test est dû à Arbuthnott (1710) qui étudiait la liste des naissances à Londres sur une période de 82 années (« + » est l'évènement « plus de garçons nés que de filles », « - » l'évènement inverse). Il constata qu'il obtenait 82 fois « + ».

Soient n paires d'observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Pour chaque paire, on fait une comparaison suivant la grandeur relative de x_i par rapport à y_i :

- « + » si $x_i < y_i$
- « - » si $x_i > y_i$
- « = » si $x_i = y_i$

On soustrait de la taille de l'échantillon le nombre de signes « = ». On note m le nombre de paires restantes.

On comptabilise le nombre de « + » et de « - », on désigne par T le plus petit des deux.

Dans le cas bilatéral, on teste l'hypothèse $H_0 : P(X < Y) = P(X > Y)$ versus $H_1 : P(X < Y) \neq P(X > Y)$.

Sous H_0 , la probabilité que X soit inférieur à Y est égale à $1/2$

Le test des signes est donc un test binomial de paramètres m et $p = 1/2$

Quand m est grand, on peut utiliser comme approximation de la loi binomiale la loi normale de paramètres :

$$\mu = \frac{1}{2}m \quad \text{et} \quad \sigma = \frac{1}{2}\sqrt{m}$$

Pour m petit, on utilise la table de la fonction de répartition de la loi binomiale de paramètres m et $p = 1/2$ ou on calcule la probabilité exacte.