DEUXIEME PARTIE :
LES TESTS D'HYPOTHESES

« La raison d'être des statistiques, c'est de vous donner raison»
Abe Burrow

Table des matières

Un test d'hypothèses permet de prendre une décision au vu des résultats d'une expérience ou d'observations d'un phénomène. Par exemple

- Contrôle qualité : Décider si une machine est à réparer ou remplacer en fonction du nombre de pièces défectueuses produites.
- Essais thérapeutiques : Décider si un nouveau traitement est efficace au vu des résultats de son expérimentation sur des malades.
- Sondage : Décider de l'impact d'une campagne à l'issu d'un sondage effectué sur la population.

Utiliser une méthode statistique pour prendre une décision permet de gérer le risque dû au caractère aléatoire des données. Un test d'hypothèse donne non seulement une règle de décision mais aussi le risque de se tromper.

1. GENERALITES

1.1. Les hypothèses

Dans chaque cas, le problème consiste à trancher, au vu d'observations, entre une hypothèse appelée *hypothèse nulle*, notée H_0, et une autre hypothèse dite *hypothèse alternative*, notée H_1. On suppose bien entendu qu'une et une seule des deux hypothèses est vraie. Un test d'hypothèses est alors une méthodologie statistique qui permet de choisir entre ces deux hypothèses.

Un test d'hypothèses s'effectue à partir d'un échantillon de variables aléatoires i.i.d de loi dépendant d'un paramètre θ (par exemple une espérance pour la loi normale). Les hypothèses se ramènent alors à tester des valeurs pour le paramètre θ. On distingue deux types d'hypothèses :

- Les *hypothèses simples* qui permettent de tester une valeur fixée : $\theta=\theta_0$
- Les *hypothèses composites* qui permettent de tester un ensemble de valeurs : $\theta<\theta_0$ ou $\theta>\theta_0$ (*test unilatéral*) ou $\theta\neq\theta_0$ (*test bilatéral*)

1.2. Les risques associés

Pour chacun des deux choix possibles, correspond une erreur.

- L'erreur dite de *première espèce* qui consiste à choisir H_1 alors que H_0 est vraie.
- L'erreur dite de *deuxième espèce* qui consiste à choisir H_0 alors que H_1 est vraie.

Les conséquences de ces deux erreurs peuvent être d'importances diverses. Par exemple, en contrôle qualité, si on décide à tort que la machine est défectueuse, on engagera des frais inutiles de réparation, si en revanche, on décide à tort que la machine est en bon état de marche, on produira des pièces de mauvaise qualité avec les répercussions que cela entraine sur le client voire même la sécurité. Les erreurs sont dites de première et deuxième espèce car, en général, l'une des deux erreurs est plus grave que l'autre.

Exemple

Prenons l'exemple classique des faiseurs de pluie (Saporta). Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel de pluie dans la Beauce en millimètres par an suit une loi normale $N(600,100^2)$. Des entrepreneurs prétendaient pouvoir augmenter de 50mm le niveau moyen de pluie par insémination des nuages au moyen d'iodure d'argent. Le procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

Les agriculteurs, à qui les faiseurs de pluie avaient proposé ce procédé forcément onéreux, pouvaient envisager les deux hypothèses suivantes :

$$\begin{cases} H_0 : \text{le procédé est sans effet} \\ H_1 : \text{le procédé augmente réellement le niveau moyen de pluie de 50mm} \end{cases}$$

L'erreur de première espèce consiste alors à acheter un procédé onéreux et inefficace. L'erreur de deuxième espèce revient à passer à côté d'un moyen d'augmenter sa production agricole.

A chaque décision correspond une erreur mais aussi une probabilité de décider juste ou d'avoir tort.

- Le *niveau de signification* du test, noté α, est la probabilité de l'erreur de première espèce, *i.e* la probabilité de rejeter H_0 à tort.
- La probabilité de l'erreur de deuxième espèce est noté β. La *puissance* du test, donnée par 1-β, est la probabilité de décider H_1 à raison.

Décision \ Vérité	H_0	H_1
H_0	1-α	β
H_1	α	1-β

L'idéal serait de diminuer simultanément les deux risques mais cela est impossible. En pratique, on considère que l'une des deux erreurs est plus grave que l'autre et on évite que cette erreur se produise. On choisit alors H_0 et H_1 de façon à ce que l'erreur que l'on considère la plus grave soit l'erreur de première espèce et on fixe α de façon à maitriser cette erreur. Plus l'erreur est grave, plus le niveau sera petit. On ne contrôle pas le risque de deuxième espèce mais on peut le calculer de façon à vérifier si le test est acceptable.

Dans le cas des faiseurs de pluie, il est clair que le risque le plus grave pour les agriculteurs était d'acheter un procédé onéreux et inefficace. Ils choisirent de le fixer à α=5%.

1.3. La variable de décision

Un test d'hypothèse à pour objectif d'établir une *règle de décision*. Celle-ci permet de choisir entre H_0 et H_1 au vu des observations $x_1,\dots,x_n$, sous la contrainte que la probabilité du risque de première espèce est égale à α fixé. L'idée est de conclure que H_0 est fausse s'il est très peu probable d'observer $x_1,\dots,x_n$ quand H_0 est vraie.

La règle de décision s'établit, non pas à partir des observations $x_1,\dots,x_n$ directement, mais à partir d'une caractéristique

estimée sur l'échantillon, en général une valeur moyenne (μ), une variance (σ^2), etc.... La variable aléatoire permettant d'estimer cette caractéristique est appelée la *variable de décision* ou *statistique* du test (en général, un estimateur usuel).
Une fois la variable de décision fixée, il est nécessaire de déterminer sa loi de probabilité.

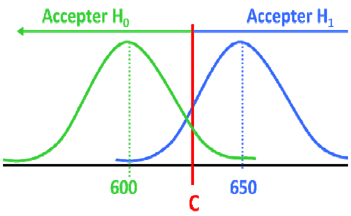
1.4. La région critique

On appelle *région critique*, notée W, l'ensemble des valeurs des observations $x_1,\ldots,x_n$ pour lesquelles on rejettera H_0. Cette région dépend du niveau α et est déterminée indépendamment des observations effectuées. Ensuite, si les observations appartiennent à W, on rejette H_0, sinon on garde H_0.

Exemple Reprenons le cas des faiseurs de pluie. Si μ désigne l'espérance mathématique de la variable aléatoire X représentant le niveau annuel de pluie, alors les hypothèses peuvent se modéliser par

$$\begin{cases} H_0 : \mu=\mu_0=600 \text{ mm} \\ H_1 : \mu=\mu_1=650 \text{ mm} \end{cases}$$

Pour tester une espérance μ, on choisit naturellement la moyenne $\overline{X}$ comme variable de décision. L'échantillon est gaussien de variance connue avec n=9 donc $\overline{X}$ suit une loi normale $N(\mu,100^2/9)$.



Si l'hypothèse H_0 est vraie alors $\mu=600$ et si l'hypothèse H_1 est vraie alors $\mu=650$, d'où le graphique ci-contre.
On en déduit que la région critique (région d'acceptation de H_1) est définie par $W=\{\overline{X}>C\}$.

1.5. Le seuil de décision

La valeur du seuil de décision C est calculée en fonction de α, c'est-à-dire telle que

$$\alpha=P(W \mid H_0 \text{ est vraie}).$$

Si on suppose que H_0 est vraie, alors on connaît la loi de variable de décision. On peut donc résoudre l'équation ci-dessus.

Exemple Dans le cas des faiseurs de pluie, sous l'hypothèse H_0, on connaît la loi de $\overline{X}$ puisque $\mu=\mu_0=600$. D'où

$$0{,}05=P(\overline{X}>C \mid H_0 \text{ est vraie}) \Rightarrow C=622 \text{ mm}.$$

On adopte alors la règle de décision suivante :

- Si $\overline{x}$ (moyenne calculée sur l'échantillon) est supérieure à 622 mm alors on accepte l'hypothèse H_1, *i.e* les agriculteurs investissent dans le procédé avec 5% de risque de se faire avoir.
- Si $\overline{x}$ est inférieure à 622 mm, alors faute de preuve, on garde l'hypothèse H_0, *i.e* les agriculteurs n'investissent pas.

La construction du test se fait sans tenir compte des valeurs observées sur l'échantillon. Ces valeurs sont utilisées uniquement à la fin pour prendre la décision mais ne servent en aucun cas à établir la règle de décision.

1.6. Puissance du test

Pour connaître le risque de se tromper dans le deuxième cas, il suffit de calculer l'erreur de deuxième espèce ou bien la puissance du test,

$$1-\beta=P(W\ |H_1 \text{ est vraie}).$$

Cela suppose de connaître la loi de $\overline{X}$ sous l'hypothèse H_1, ce qui est le cas ici puisque $\mu_1=650$. On trouve alors $1-\beta=0.98$. L'erreur de deuxième espèce est finalement de 2%.

$\alpha=P(W\ \|H_0$ est vraie)	$\beta=P(\overline{W}\ \|H_1$ est vraie)
$1-\alpha=P(\overline{W}\ \|H_0$ est vraie)	$1-\beta=P(W\ \|H_1$ est vraie)

Dans le cas où l'hypothèse H_1 est composite, la loi de la variable de décision sous l'hypothèse H_1 est inconnue puisque le paramètre θ_1 n'est pas fixé mais peut prendre une infinité de valeur. Il est alors impossible de calculer $1-\beta$ ou β. On définit alors la fonction puissance qui aux valeurs possibles de θ_1 associe la puissance correspondante,

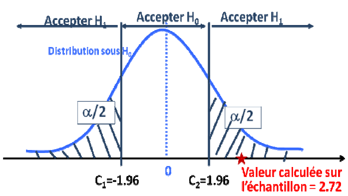
$$\theta \mapsto 1-\beta(\theta).$$

1.7. La p-valeur

La p-valeur est une probabilité donnant une information complémentaire au risque de première espèce α. Connaissant la valeur prise par la variable de décision sur l'échantillon, la p-valeur indique le niveau α minimum du test à partir duquel l'hypothèse H_0 est rejetée.

Si la p-valeur est inférieure à la valeur du seuil préalablement défini (traditionnellement 5 % ou 1 %), on rejette l'hypothèse nulle.

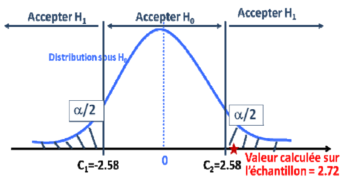
Prenons l'exemple d'une hypothèse H_0 sous laquelle la variable de décision Z suit une loi N(0,1). Supposons que la valeur prise par Z sur l'échantillon soit 2.72. Construisons un test bilatéral.



- Si on fixe un niveau de 5%, on obtient la règle de décision,

  $$H_0 \text{ est rejetée} \Leftrightarrow z \notin [-1.96, 1.96].$$

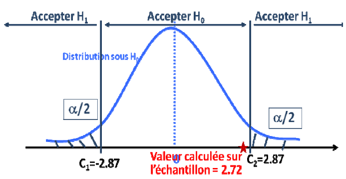
  Avec z=2.72, l'hypothèse H_0 est donc rejetée avec 5% de risque de se tromper.



- Si on fixe un niveau de 1%, on obtient la règle de décision,

  $$H_0 \text{ est rejetée} \Leftrightarrow z \notin [-2.58, 2.58].$$

  Avec z=2.72, 'hypothèse H_0 est donc rejetée avec 1% de risque de se tromper.



- Si on fixe un niveau de 0.5%, on obtient la règle de décision,

  $$H_0 \text{ est rejetée} \Leftrightarrow z \notin [-2.87, 2.87].$$

  Avec z=2.72, l'hypothèse H_0 n'est pas rejetée.

Dans cet exemple, la p-valeur se situe entre 1% et 0.5%. Elle est en fait égale à 0.653%.

1.8. Méthodologie

[(*)] Il est conseillé d'expliciter les erreurs de première et deuxième espèces pour s'assurer du bon choix entre H_0 et H_1

<u>Construction du test</u>

1. Choix de H_0 et H_1 [(*)]
2. Détermination de la variable de décision **et de sa loi**
3. Allure de la région critique en fonction de H_1
4. Calcul de la région critique en fonction de α
5. Calcul éventuel de la puissance 1-β
6. Règle de décision avec risques associés

<u>Prise de décision</u>

a. Choisir entre H_0 et H_1, soit en comparant la valeur de la statistique du test calculée sur l'échantillon avec le(s) seuil(s), soit à partir de la p-valeur.
b. Définir le risque associé

2. TESTS SUR LES CARACTERISTIQUES USUELLES

2.1. Test sur la moyenne

Soit un échantillon $X_1,\dots,X_n$ tel que $E(X)=\mu$ et $var(X)=\sigma^2$. On souhaite tester les hypothèses

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu=\mu_1\neq\mu_0 \end{cases}$$

La moyenne $\overline{X}$ est un estimateur naturel de μ. Pour un échantillon assez grand, on approche la loi de l'estimateur grâce au T.C.L. par une loi normale $N(\mu,\sigma^2/n)$. La statistique du test (variable de décision) est

$$Z=\sqrt{n}\,\frac{\overline{X}-\mu}{\sigma}.$$

Si n est suffisamment grand, on considère donc qu'elle suit une loi normale N(0,1).

Remarque On sait que

$$Z=\sqrt{n}\,\frac{\overline{X}-\mu}{S^*}$$

converge vers une loi N(0,1) quand n tend vers +∞. Donc si σ^2 n'est pas connu, on peut le remplacer par s^{*2} si l'échantillon est assez grand.

Remarque Dans le cas où l'échantillon est gaussien, la loi est exacte et non approchée (normale centrée-réduite si σ^2 est connue et Student à n-1 d.d.l. sinon). On peut donc travailler sur des échantillons de petites tailles.

Exemple Un fabricant produit des lunettes d'un certain modèle dont la longueur en millimètre X suit une loi normale. L'espérance et l'écart-type de cette longueur sont spécifiés dans le cahier des charges à 130 mm et 0.48 mm respectivement. On prélève un échantillon de 9 faces de lunettes et on souhaite tester les hypothèses

$$\begin{cases} H_0 : \mu=\mu_0=130 \text{ mm} & \text{(lunettes conformes au cahier des charges)} \\ H_1 : \mu=\mu_1=130.5 \text{ mm} & \text{(lunettes non conformes au cahier des charges)} \end{cases}$$

avec un niveau de confiance de 99%.

Taille d'un échantillon

On remarque qu'un test d'hypothèses est un problème à

4 paramètres : α, β, C et n

2 équations : $\alpha=P(W|H_0)$ et $1-\beta=P(W|H_1)$

Il faut donc fixer deux paramètres pour résoudre le système. Il y a deux cas de figure

- Les observations ont déjà été effectuées, on a alors α et n fixés (maitrise du risque de première espèce).
- Les observations n'ont pas été effectuées et on fixe α et β (maitrise des deux risques) et on détermine la taille de l'échantillon en conséquence.

Exemple Dans le cas de l'exemple précédent, il faut prélever 15 lunettes pour avoir un risque de deuxième espèce β=5%.

2.2. Test sur une proportion

On souhaite tester les hypothèses

$$\begin{cases} H_0 : p=p_0 \\ H_1 : p=p_1\neq p_0 \end{cases}$$

où p désigne une proportion. La fréquence empirique F_n est un estimateur naturel de p. Si l'échantillon est suffisamment grand, on peut approcher la loi de F_n grâce au TCL par une loi normale $N(p,p(1-p)/n)$.

La variable de décision (ou statistique) du test est

$$Z=\sqrt{n}\frac{F_n-p}{\sqrt{p(1-p)}}.$$

Si n est assez grand, on considère qu'elle suit une loi N(0,1).

Exemple Un parti politique souhaite vérifier si la parité est respectée. Sur un échantillon de 100 adhérents, on compte 45 femmes et 55 hommes. Peut-on affirmer que la parité est respectée avec un risque de 5% de se tromper?

2.3. Test sur une variance dans le cas gaussien

Supposons que l'échantillon soit gaussien avec X de loi $N(\mu,\sigma^2)$.

On souhaite tester les hypothèses

$$\begin{cases} H_0 : \sigma^2={\sigma^2}_0 \\ H_1 : \sigma^2={\sigma^2}_1\neq {\sigma^2}_0 \end{cases}$$

Dans le cas où μ est inconnu, la variance empirique S^{*2} est un estimateur naturel de σ^2 . La variable de décision (statistique) du test est

$$\frac{n-1}{\sigma^2}S^{*2}\sim\chi^2_{n-1}.$$

Le théorème de Neyman-Pearson est défini au paragraphe suivant

Le graphe de la loi du χ^2_{n-1} ne permet pas de déterminer l'allure de la région critique. Il est nécessaire d'appliquer le théorème de Neyman-Pearson qui permet d'affirmer dans ce cas que la région critique est de la forme :

$$W=\{S^{*2}>C\} \Leftrightarrow {\sigma^2}_1> {\sigma^2}_0$$
$$W=\{S^{*2}<C\} \Leftrightarrow {\sigma^2}_1<{\sigma^2}_0$$

Exemple Dans la fabrication de comprimés effervescents, il est prévu que le dosage de bicarbonate de sodium suit une loi normale. On a prélevé un échantillon de 25 comprimés et mesuré un écart-type empirique $s^* = 12.1$ mg. On souhaite tester si la production est homogène avec les hypothèses suivantes

$$\begin{cases} H_0 : \sigma^2=\sigma^2_0=10^2 \\ H_1 : \sigma^2=\sigma^2_1>\sigma^2_0 \end{cases}$$

avec un niveau de confiance de 5%.

3. TESTS DU CHI-DEUX

3.1. Test d'adéquation Un test d'adéquation ou test d'ajustement permet de déterminer si un échantillon suit une loi usuelle donnée. Les hypothèses sont donc

$$\begin{cases} H_0 : \text{la variable suit la loi de probabilité} \\ H_1 : \text{la variable ne suit pas la loi de probabilité} \end{cases}$$

Il est possible d'utiliser ce test dans le cas d'une loi continue dont les données sont regroupées en classes (a_i,a_{i+1}). Dans ce cas la probabilité p_i de la classe (a_i,a_{i+1}) est $p_i = P(X \in (a_i, a_{i+1})) = F(a_{i+1}) - F(a_i)$ Cependant, il est préférable d'utiliser un test adapaté.

Supposons que la loi usuelle est discrète prenant les valeurs x_i avec la probabilité p_i, i=1,…,k. Le test du chi-deux consiste à comparer pour chaque valeur x_i,

- son effectif théorique : $n\times p_i$
- et son effectif empirique, N_i=nombre d'observations prenant la valeur x_i.

On pose alors

$$D_n = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}.$$

Dans le cas où la loi usuelle est une loi discrète uniforme, tous les x_i sont équiprobables, donc $p_i=1/k$.

D_n représente l'écart entre la distribution théorique et la distribution observée. Plus cet écart est grand et moins il y de chance que la distribution de l'échantillon soit la même que la distribution théorique. Donc la région critique est de la forme $W=\{D_n>C\}$ et la règle de décision est

- Si $d_n<C$ alors je considère que l'échantillon suit la loi usuelle
- Si $d_n>C$ alors j'en déduis qu'il ne suit pas la loi usuelle avec un risque α de me tromper

On peut montrer que sous l'hypothèse H_0, D_n suit approximativement une loi χ^2_{k-1}.

- Cette approximation est valable si n est assez grand et p_i pas trop petit avec la règle empirique **$np_i \geq 5$**. Si cette règle n'est pas respectée à cause d'un p_i trop petit, il faut alors regrouper le x_i avec une valeur voisine.
- Si la loi de l'échantillon (sous l'hypothèse H_0) dépend de *r*

paramètres et que ces paramètres ont été estimés au préalable sur l'échantillon, alors D_n suit une loi χ^2_{k-r-1}. Par exemple, si on souhaite tester si un échantillon suit une loi de Poisson et que le paramètre λ est estimé sur ce même échantillon, alors D_n suit une loi χ^2_{k-1-1}.

Exemple On lance 60 fois de suite un dé. A partir des effectifs observés du tableau ci-dessous, on peut en conclure que le dé est équilibré.

x_i	1	2	3	4	5	6
N_i	11	8	9	12	7	13

3.2. Test d'indépendance

Considérons deux variables qualitatives (ou quantitatives regroupées en classe) X et Y avec respectivement *r* et *s* modalités. Le test d'indépendance du khi-deux permet de tester l'indépendance de ces variables. Notons n_{ij} le nombre d'individus de l'échantillon présentant simultanément la modalité i de la variable X et la modalité j de la variable Y.
Soit p_{ij} la probabilité théorique de cette modalité. Notons $p_{i.}$ et $p_{.j}$ les probabilités marginales,

$$p_{i.} = \sum_{j=1}^{s} p_{ij} \text{ et } p_{.j} = \sum_{i=1}^{r} p_{ij} .$$

S'il y a indépendance alors la loi conjointe est égale au produit des lois. Nous allons donc effectuer le test suivant

$$\begin{cases} H_0 : p_{ij}=p_{i.}\times p_{j.} & \text{(indépendance)} \\ H_1 : p_{ij}\neq p_{i.}\times p_{j.} & \text{(pas indépendance)} \end{cases}$$

Pour ce faire, on utilise la variable de décision,

$$D_n = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - n_{i.}n_{.j}/n\right)^2}{n_{i.}n_{.j}/n} .$$

La région critique de ce test est de la forme $W=\{D_n>C\}$.

Sous l'hypothèse H_0, on peut approcher la loi de D_n par une loi $\chi^2_{(r-1)(s-1)}$.

Exemple Le proviseur d'un lycée désire comparer le taux de réussite au baccalauréat des trois sections générales. Les résultats obtenus dans son lycée sont dans le tableau suivant.

	L	ES	S	Total
Réussite	41	59	54	154
Echec	21	36	75	132
Total	62	95	129	286

Le proviseur peut donc conclure avec un risque de 5% que les résultats au baccalauréat dans son établissement dépendent de la section.

4. TESTS DE COMPARAISON D'ECHANTILLONS

4.1. Test de comparaison de deux échantillons

Il s'agit ici de répondre à la question:

« Deux échantillons ont-ils été prélevés dans la même population? »
ou encore
« L'écart entre deux échantillons est-il dû au hasard ou est-il significatif? »

En pratique, on se contente de comparer les échantillons au travers de valeurs remarquables telles que la moyenne, variance, ….

Test de de Student

Le test de Student consiste à faire la différence entre les deux moyennes (proportions) et de tester si la différence est nulle

$$H_0 : \mu_1=\mu_2 \qquad\qquad H_0 : \mu_1-\mu_2=0$$
$$H_1 : \mu_1\neq\mu_2 \qquad \Leftrightarrow \qquad H_1 : \mu_1-\mu_2\neq 0$$

- Si les échantillons sont de grande taille, alors sous l'hypothèse H_0, la statistique du test

$$T=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Le test de Student est robuste et supporte quelques écarts aux conditions d'application.

suit une loi N(0,1), où n_1 (resp. n_2) est la taille de l'échantillon 1 (resp. échantillon 2) et σ_1^2 (resp. σ_2^2) est la variance de la population 1 (resp. population 2).

- Si les échantillon sont petits mais s'ils sont gaussiens de <u>même variance</u>, alors sous l'hypothèse H_0, la statistique du test

$$T=\frac{\overline{X}_1-\overline{X}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \text{ où } S=\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}$$

suit une loi de Student à n_1+n_2-2 degrés de liberté.

Plus la différence est loin de 0 et moins il y a de chance que l'hypothèse H_0 soit vraie, donc la région critique est de la forme W={T>C}.

Exemple

Une étude menée sur 9 garçons et 12 filles montre que le QI moyen des garçons est 107 et celui des filles est 112. On sait par ailleurs que le QI suit une loi normale d'écart-type 15. Avec un test à 5%, on peut dire qu'il n'y a pas de différence significative entre le QI des filles et celui des garçons.

Test non paramétrique

Si les hypothèses du test de Student ne sont pas respectées, on pourra se tourner vers le test de Mann-Whitney. Celui-ci consiste à *de Mann-Whitney* déterminer si les deux échantillons se « positionnent » de la même façon. C'est-à-dire que pour chaque valeur de l'échantillon 1, on regarde le nombre de valeurs de l'échantillon 2 inférieurs à cette valeur, et *vice-versa*. Ce test n'impose ainsi aucune contrainte sur la distribution *a priori* des données. On dit que c'est un test non paramétrique. En général, les tests non paramétriques sont moins puissants que les tests paramétriques.

Test de Fisher Dans le cas d'échantillons gaussiens le test d'égalité des variances se fait à l'aide du test de Fisher,

$$H_0 : \sigma_1=\sigma_2$$
$$H_1 : \sigma_1\neq\sigma_2$$

Sous l'hypothèse H_0, la statistique du test

$$F=\frac{S_1^2}{S_2^2}$$

suit une loi de Fisher $F_{n_1-1;n_2-1}$. La région critique est de la forme W={F>C}.

Exemple Un courtier rapporte que le taux de rendement moyen pour un échantillon d'actions de 10 compagnies pétrolières est de 12.6% avec un écart-type de 3.9%. Le taux de rendement moyen des actions de 8 compagnies de service est de 10.9% avec un écart-type de 3.5%. Avec un risque de 5% peut-on conclure que les actions des compagnies pétrolières sont plus volatiles que celles de compagnies de service?

4.2. Test de comparaison de k>2 échantillons

Là encore, on suppose disposer de k échantillons aléatoires et indépendants d'erreurs. Le test statistique permet de statuer sur le fait que tous les échantillons sont issus d'une même population, c'est-à-dire qu'il n'y pas de différence significative entre les échantillons. Dans le cas contraire, on pourra conclure qu'il existe au moins un échantillon différent des autres mais sans savoir lequel ni s'il y en a plusieurs. Pour obtenir cette précision, il faudra faire des tests deux à deux (à noter que les résultats des tests deux à deux et le résultat du test global peuvent être contradictoires).

Considérons 7 groupes d'observations tirées indépendamment d'une même population statistique. Il faudrait réaliser 7 (7 – 1)/2 = 21 tests pour comparer toutes les paires de groupes. Chaque test étant réalisé au niveau $\alpha = 0,05$, on a, dans chaque cas, 5 chances sur 100 de rejeter H_0 même si H_0 est vraie. La probabilité de rejeter H_0 au moins une fois à tord au cours de 21 tests est 0,66 !

Test paramétrique de l'ANOVA

De même que précédemment, si on suppose que pour chaque échantillon, les données suivent une loi normale de variance constante, alors on pourra utiliser le test paramétrique de l'ANOVA (*ANalysis Of Variance*) à un facteur. Celui-ci consiste à comparer simultanément les k moyennes des échantillons, c'est-à-dire à tester si $\mu_1=\mu_2=\ldots=\mu_k$, à l'aide de leur valeur approchée respective $\overline{y}_1$, $\overline{y}_2, \ldots, \overline{y}_k$. Dans le cas contraire on pourra dire qu'au moins une de ces moyennes est significativement différente des autres sont pouvoir préciser la ou lesquelles.

Test non paramétrique de Kruskall-Wallis

Si les hypothèses de normalité et de variance constante ne sont pas respectées, on préférera le test non paramétrique de Kruskall-Wallis. Celui-ci est basé sur le même principe des rangs que le test de Mann-Whitney.