La théorie des tests

La démarche scientifique dans les sciences expérimentale est la suivante :
Tout d'abord on fait une hypothèse. Celle-ci est plus ou moins plausible, selon les dons ou la chance de celui qui l'a fait.
Ensuite, on définit une expérience permettant de tester cette hypothèse. Si les résultats de l'expérience mènent à rejeter l'hypothèse, on n'en conclut pas pour autant qu'elle est exacte ; mais on l'accepte provisoirement, en attendant par exemple qu'une expérience plus puissante permette de la remettre en cause.
Le test statistique épouse parfaitement cette démarche scientifique. La description de "la" réalité en statistiques se fait à l'aide de *variables*. On se pose souvent la question de comparer ces variables, de tester si elles sont égales ou différentes, de savoir si on peut considérer qu'elles correspondent ou non à une même population [sous-jacente], si elles correspondent à une distribution donnée, si elles sont conformes à un modèle précis etc. sachant que ces variables et leurs données ne correspondent qu'à un échantillon de valeurs.
Un test statistique est donc un mécanisme qui permet , à partir d'échantillons, de

- clarifier et définir le cadre rigoureux de ces études,
- fournir un formalisme précis pour toutes les situations,
- savoir si les différences mises en jeu sont importantes ("significatives" pour un seuil donné) ou non.

1 Définitions

1.1 Hypothèse nulle

C'est l'hypothèse dont on cherche à savoir si elle peut être rejetée grâce aux observations dont on dispose. Elle est noté H_0.

Cette hypothèse doit être solidement établie et facile à être identifiée. C'est l'hypothèse à laquelle on tient particulièrement. Elle correspond généralement à une situation de *statu quo*.

Note historique : Pourquoi "hypothèse nulle" ?
Dans les années 1930, le statisticien anglais Fisher mit au point la méthodologie statistique permettant d'évaluer si deux traitements avaient des efficacités différentes. Pour cela il fallait rejeter l'hypothèse que la différence de leur efficacité est *nulle*. D'où l'expression *hypothèse nulle*.

1.2 Hypothèse alternative

C'est l'hypothèse, noté H_1, qui est en concurence avec l'hypothèse nullle. Elle est différente de H_0 qui est souvent égale à son contraire. C'est souvent l'hypothèse à démontrer.

1.3 Risque de première espèce

C'est la probabilité α que l'on a de rejeter l'hypothèse nulle H_0 avec le test statistique employé, quand cette hypothèse nulle est vraie.
La valeur seuil de α communément admise pour rejeter H_0 est de 0.05. Cette valeur est arbitraire. Certains auteurs expliquent ce choix par le travail historique de *Fisher* en 1925. Il proposait dans ses tables des valeurs particulières de α et notamment celle de $0,05$. De nombreux livres de statistiques ne donnent que des tables valables uniquement pour $\alpha = 0.05$, ignorant d'autres possibilités. D'autres

auteurs insistent sur l'utilisation fréquente de cette valeur dans les études de contrôle de qualité où l'accent est mis sur la performance de l'outil de décision lors de tests fréquemment répétés.

1.4 Zone de rejet de l'hypothèse nulle

C'est l'ensemble W des valeurs expérimentales pour lesquelles on décide à l'avance que l'on déclarera l'hypothèse nulle fausse(que l'on rejettera l'hypothèse nulle). Sa détermination pratique se fait en écrivant :

$$P(W \mid H_0) = \alpha.$$

La région d'acceptation est donc son complémentaire $\overline{W}$ et l'on a donc

$$P(\overline{W} \mid H_0) = 1 - \alpha.$$

1.5 Degré de signification (p-value)

Le paramètre p est égal à la probabilité d'observer des résultats au moins aussi en désaccord avec l'hypothèse nulle H_0 que ceux qui ont été obtenus. Rappelons que le risque α est tel que

$$P_{H_0}(rejet \text{ de } H_0) = \alpha.$$

Une fois les données recueillies, la valeur prise par la statistique de test sera calculée, et la réponse sera binaire : rejet ou non de H_0. On préfère souvent garder l'information contenue dans la valeur de la statistique de test, en retournant le seuil limite auquel elle aurait été rejetée, compte tenu de l'observation. Prenons l'exemple (fréquent) d'une hypothèse H_0 sous laquelle la statistique de test T suit la loi normale $N(0, 1)$. La règle de rejet pour le test bilatéral de seuil 0.05 est :

$$\text{Rejet de } H_0 \Longleftrightarrow T \notin [-1.96, 1.96].$$

Supposons que la valeur prise par T soit 2.72. L'hypothèse H_0 sera donc rejetée. Mais elle serait également rejetée au seuil 0.01. En fait elle serait rejetée pour n'importe quel seuil supérieur à 0.00653, ce qui est un renseignement plus précis qu'une simple réponse binaire.
Nous définissons la p-valeur (p-value, valeur critique) comme suit :

Définition Soit H_0 l'hypothèse nulle, T la statistique de test et F_0 sa fonction de répartition sous l'hypothèse H_0. On suppose que F_0 est continue, alors

- Pour un test bilatéral (rejet des valeurs trop écartées) la p-valeur d'une valeur t prise par T est :

$$p(t) = \begin{cases} 2F_0(t), & \text{si } F_0(t) < 0.5, \\ 2(1 - F_0(t)) & \text{si } F_0(t) \geq 0.5. \end{cases}$$

- Pour un test unilatéral à droite (rejet des valeurs trop grandes) la p-valeur d'une valeur t prise par T est :

$$p(t) = 1 - F_0(t).$$

- Pour un test unilatéral à gauche (rejet des valeurs trop petites) la p-valeur d'une valeur prise par est :

$$p(t) = F_0(t).$$

Cependant calculer une p-valeur pour un test bilatéral est assez artificiel. Au vu de la valeur prise par T, on aura tendance à effectuer plutôt un test unilatéral visant à décider si la valeur observée est trop grande ou trop petite. Pour une statistique de test suivant la loi , la valeur 2.72 est clairement à droite de la distribution. Le problème ne se pose plus de savoir si elle est trop petite, mais plutôt si elle est significativement trop grande. En pratique, pour une statistique de test de fonction de répartition F_0 sous H_0, on définira souvent la p-valeur de la valeur t par :

$$p(t) = min\{F_0(t), 1 - F_0(t)\}.$$

La connaissance de la p-valeur rend inutile le calcul préalable de la région de rejet : si $p(t)$ est la p-valeur d'une observation t sous l'hypothèse H_0, on obtient un test de seuil α par la règle de rejet :

$$\text{Rejet de } H_0 \Longleftrightarrow p(T) < \alpha$$

Dans le cas continu, ceci revient à remplacer la statistique T par $F_0(T)$ ou $1 - F_0(T)$. Sous l'hypothèse H_0, ces deux statistiques suivent la loi uniforme $U(0,1)$.
Quand la statistique de test est discrète, il faut inclure la valeur observée dans l'intervalle dont on calcule la probabilité. Pour un test unilatéral à gauche, cela n'induit pas de changement : $F_0(t)$ est la probabilité que T soit inférieure ou égale à t. Pour un test unilatéral à droite sur une variable à valeurs dans N (le cas le plus fréquent) il faudra calculer $1 - F_0(t-1)$. Supposons par exemple que la loi de T soit la loi binomiale $B(100, 0.5)$, la p-valeur de 60 est la probabilité que T soit supérieure ou égale à 60, à savoir :

$$1 - F_{B(100,0.5)}(59) = 0.0359.$$

1.6 Risque de deuxième espèce et puissance d'un test

- Le risque de deuxième espèce est la probabilité β de se tromper quand on accepte H_0.
- La puissance d'un test est la probabilité $1 - \beta$ de rejeter l'hypothèse nulle H_0 alors que celle ci est fausse. Elle dépend de l'hypothèse alternative considérée et de la taille de l'échantillon.

Tableau récapitulatif

Hypothèses	Décision	
	H_0 acceptée	H_0 rejetée
Si H_0 est vraie	Décision correcte. Probabilité =1-α = **seuil de confiance**	Erreur de **première espèce**. Probabilité $=\alpha$
Si H_0 est fausse	Erreur de **deuxième espèce**. Probabilité $=\beta$	Décision correcte. Probabilité $1 - \beta$= **puissance du test**

1.7 Comment procéder pour effectuer un test

Les différentes étapes doivent donc être suivies pour tester une hypothèse :

1. définir l'hypothèse nulle H_0 à contrôler,

2. choisir un test statistique ou une statistique pour contrôler H_0
3. définir la distribution de la statistique sous l'hypothèse "H_0 est réalisée",
4. définir le niveau de signification du test ou région critique notée α,
5. calculer, à partir des données fournies par l'échantillon, la valeur de la statistique
6. prendre une décision concernant l'hypothèse posée et faire une interprétation.

Remarque :

Les probabiltés α et β varient en sens contraire. α étant fixé, il faut choisir une variable de décision, cette variable doit apporter le maximum d'information sur le problème posé et dont la loi sera différente suivant que H_0 ou H_1 est vraie. Il faut que sa loi soit entièrement connue au moins si H_0 est vraie.

1.8 Familles de tests

Il existe deux grandes familles de tests :

- **Les tests paramétriques**

  On suppose que la variable X suit une loi connue de distribution $F(x, \theta)$ mais que le paramètre θ est inconnu. On fait l'hypothèse H_0 telle que $\theta = \theta_0$ à partir de l'échantillon de taille n et l'on essaie de savoir si H_0 est vraie ou non.
  Dans la plupart des cas, ces tests sont appliqués (utilisés) lorsque les variables sont quantitatives, lorsqu'on a "suffisamment" de valeurs ou lorsque l'hypothèse de normalité sous-jacente aux données est vérifiées (ou supposée être vérifiée). Ces tests sont dits "puissants".

- **Les tests non paramétriques**

  Ils sont valables quelle que soit la loi de la variable aléatoire étudiée et donc vrai lorsque'on ignore tout de cette loi. Pour cela, on fait l'hypothèse H_0 que X suit la loi de distribution $F_0(x, \theta)$ et l'on fait le test pour savoir si X suit effectivement cette distribution F_0 ou non.

  * *Avantages des tests non paramétriques*

  1) Leur emploi se justifie lorsque les conditions d'applications des autres méthodes ne sont pas satisfaites, même après d'éventuelles transformations de variables.

  2) Les probabilités des résultats de la plupart des tests non paramétriques sont des probabilités exactes quelle que soit la forme de la distribution de la population dont est tiré l'échantillon.

  3) Pour des échantillons de taille très faible jusqu'à N = 6, la seule possibilité est l'utilisation d'un test non paramétrique, sauf si la nature exacte de la distribution de la population est précisément connue. Ceci permet une diminution du coût ou du temps nécessaire à la collecte des informations.

  4) Il existe des tests non paramétriques permettant de traiter des échantillons composés à partir d'observations provenant de populations différentes. De telles données ne peuvent être

traitées par les tests paramétriques sans faire des hypothèses irréalistes.

5) Seuls des tests non paramétriques existent qui permettent le traitement de données qualitatives : soit exprimées en rangs ou en plus ou moins (échelle ordinale), soit nominales.

6) Les tests non paramétriques sont plus faciles à apprendre et à appliquer que les tests paramétriques. Leur relative simplicité résulte souvent du remplacement des valeurs observées soit par des variables alternatives, indiquant l'appartenance à l'une ou à l'autre classe d'observation, soit par les rangs, c'est-à-dire les numéros d'ordre des valeurs observées rangées par ordre croissant. C'est ainsi que la médiane est généralement préférée à la moyenne, comme paramètre de position.

* *Inconvénients des tests non paramétriques*

1) Les tests paramétriques, quand leurs conditions sont remplies, sont plus puissants que les tests non paramétriques.
2) Un second inconvénient réside dans la difficulté à trouver la description des tests et de leurs tables de valeurs significatives, surtout en langue française. Heureusement, les niveaux de significativité sont donnés directement par les logiciels statistiques courants.

2 Tests paramétriques

2.1 Tests de conformité

Les tests de conformité sont destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée ou représentatif de cette population, vis-à-vis d'un paramètre comme la moyenne, la variance ou la fréquence observée. Ceci implique que la loi théorique du paramètre est connue au niveau de la population.

2.1.1 Comparaison d'une fréquence expérimentale et d'une fréquence théorique

– **Position du problème**

Dans une population P, on étudie un caractère statistique à deux modalités A et $\overline{A}$. Chaque individu présente ou non la modalité A. Soit p la fréquence (ou la proportion, ou le pourcentage) d'apparition de A dans la population, et f la fréquence d'apparition de A dans un échantillon de taille n. Le problème est de savoir si l'échantillon est représentatif de la population, c'est à dire si la différence entre f et p est explicable par les aléas dus à l'échantillonnage.
On suppose que $n \geq 30$, $np, n(1-p) \geq 5$, et on note F la variable aléatoire qui prend la valeur f sur chaque échantillon.

– **Mise en place du test**

Elle utilise la variable (de décision) $U = \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}}$ qui suit sensiblement la loi normale centrée et réduite. On note $u = \frac{f-p}{\sqrt{\frac{p(1-p)}{n}}}$ la valeur de la variable U obtenue à partir des observations,
et u_β le quantile d'ordre β de la loi $N(0,1)$. On supposera que le risque de première espéce est de α.

- *Cas d'un test bilatéral*

  H_0 : la fréquence observée f est conforme à la fréquence théorique p.
  H_1 : la différence entre f et p est trop importante pour être expliquable par les fluctuations de l'échantillonnage.

  Décision : si $-u_{1-\frac{\alpha}{2}} \leq u \leq u_{1-\frac{\alpha}{2}}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

- *Cas d'un test unilatéral*

  - $H_0 : f \geq p$, la fréquence f observée sur l'échantillon est *a priori* supérieure à la fréquence théorique p.
    $H_1 : f < p$.

    Décision : si $u \geq -u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

  - $H_0 : f \leq p$, la fréquence f observée sur l'échantillon est *a priori* inférieure à la fréquence théorique p.
    $H_1 : f > p$.

    Décision : si $u \leq u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

- **Exemple**

  Une anomalie génétique touche en France 1/1000 des individus. On a constaté dans une région donnée : 57 personnes atteintes sur 50000 naissances. Cette région est-elle représentative de la France entière ?

  Réponse : La fréquence de l'anomalie génétique est $p = 0,0010$ en France. Soit f la fréquence dans la région étudiée. La fréquence observée dans la région étudiée est $\hat{f} = \frac{57}{50000} = 0,00114$. Nous avons à tester (bilatéralement) $H_0 : f = p$ contre $H_1 : f \neq p$.

  Pour conclure, nous avons :

  $u_{obs} = \frac{\frac{k}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.00114 - 0.001}{\sqrt{\frac{0.001 \times 0.999}{50000}}} = 0.99$

  - Statégie 1 : avec un risque d'erreur $\alpha = 0,05$, $u_{0.975} = 1,96$. Comme $u_{obs} \leq u_{0.975}$ et donc H_0 ne peut être rejetée

  - Statégie 2 : $u_{obs} = 0,99$ correspond à une probabilité critique $p_{obs} = 0,32$. Comme $p_{obs} \geq 0,05$ alors qu'elle est vraie est très élevé. On accepte donc l'hypothèse H_0.

  La région considérée est représentative de la France entière en ce qui concerne la fréquence de cette anomalie génétique.

2.1.2 Comparaison d'une moyenne expérimentale à une moyenne théorique μ_0

– **Position du problème**

Soit X une variable aléatoire définie sur une population P avec $E(X) = \mu$ et $V(X) = \sigma^2$. On dispose d'un échantillon de n observations de X. Sur cet échantillon la moyenne et la variance sont estimées par

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

On désire savoir si l'échantillon est représentatif de la population, c'est à dire si la différence observée entre μ et $\overline{X}$ est explicable par les fluctuations d'échantillonnage.

– **Mise en place du test**

Elle fait intervenir la loi de la variable $U = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$

– ***Cas d'un grand échantillon*** ($n > 30$)
La variable $U = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ suit sensiblement une $N(0,1)$.
On note $u = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ la valeur de la variable U déterminée à partir de l'échantillon et u_β le quantile d'ordre β de la loi $N(0,1)$. On supposera que le risque de première espéce est de α.

  – *Cas d'un test bilatéral*

  H_0 l'échantillon est extrait au hasard de la population ; sa moyenne μ est conforme à la moyenne μ_0 fixée,
  H_1 :μ n'est pas conforme à μ_0.

  Décision : si $-u_{1-\frac{\alpha}{2}} \leq u \leq u_{1-\frac{\alpha}{2}}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

  – *Cas d'un test unilatéral*

  – H_0 : La moyenne de l'échantillon est supérieure ou égale à μ_0
  H_1 : La moyenne de l'échantillon est inférieure à μ_0

  Décision : Si $u \geq -u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

  – H_0 : La moyenne de l'échantion est inférieure ou égale à μ_0
  H_1 : La moyenne de l'échantion est supérieure à μ_0

  Décision : si $u \leq u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

- ***Cas d'un petit échantillon*** $(n < 30)$ ***et d'une population gaussienne***
  La variable $T = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ suit sensiblement une loi de Student à $n-1$ degrés de liberté. On note $t = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ la valeur de T déterminée à partir de l'échantillon. Soit $\lambda = n-1$. On note $t_\beta(\lambda)$ le quantile d'ordre β de la loi de Student à λ degrés de liberté. On supposera que le risque de première espéce est de α.

  - *Cas d'un test bilatéral*

    H_0 l'échantillon est extrait au hasard de la population ; sa moyenne $\overline{X}$ est conforme à la moyenne μ_0 de la population,
    H_1 :$\overline{X}$ n'est pas conforme à μ_0.

    Décision : si $-t_{1-\frac{\alpha}{2}}(\lambda) \leq t \leq t_{1-\frac{\alpha}{2}}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

  - *Cas d'un test unilatéral*

    - H_0 : La moyenne de l'échantillon est supérieure ou égale à μ_0
      H_1 : La moyenne de l'échantion est inférieure à μ_0

      Décision : si $t \geq -t_{1-\alpha}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

    - H_0 : La moyenne de l'échantillon est inférieure ou égale à μ_0
      H_1 : La moyenne de l'échantion est supérieure à μ_0

      Décision : si $t \leq t_{1-\alpha}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

- **Exemple**

  La glycémie d'une population suit une loi normale d'espérance $\mu_0 = 1g/l$ et d'écart-type $\sigma_0 = 0,1g/l$. On relève les glycémies chez 9 patients. On trouve $\bar{x} = 1,12g/l$. Cet échantillon est-il représentatif de la population ?

  Réponse : Nous avons à tester (bilatéralement) $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

  Pour conclure, nous avons : $u_{obs} = \frac{1.12-1}{\sqrt{\frac{0.1^2}{9}}} = 3.6$

  - Statégie 1 : avec un risque d'erreur $\alpha = 0,05$, $u_{0.975} = 1,96$. Comme $u_{obs} \geq u_{0.975}$, donc H_0 ne peut être acceptée

  L'échantillon de 9 patients n'est pas représentatif de la population en ce qui concerne la glycémie. Le traitement reçu a déplacé le taux de glycémie par rapport à la norme nationale.

2.1.3 Comparaison d'une variance expérimentale à une variance théorique σ_o^2

- **Position du problème**

  Soit X une variable aléatoire définie sur une population P avec $E(X) = \mu$ et $V(X) = \sigma^2$. On suppose qu'on dispose d'un échantillon de n observations indépendantes de X. On note $\overline{X}$ et S^2 la moyenne et la variance estimées de ces observations. Le problème est de tester si la différence constatée entre σ^2 et σ_0^2, une variance donnée, est explicable par les fluctuations d'échantillonnage.

- **Mise en place du test**
  Elle fait intervenir la loi de la variable $K^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2}$

  - ***Cas ou*** $n \leq 31$

    On suppose que X suit une $N(\mu, \sigma^2)$. La variable K^2 suit une loi du χ^2 à $\lambda = n - 1$ degrés de liberté.
    On note $\chi^2 = \frac{n-1}{\sigma_0^2} s^2$ la valeur de K^2 obtenue à partir des observations, $\chi_\beta(\lambda)$ le quantile d'ordre β de la variable K^2. On supposera en plus que le risque de première espéce est de α.

    - *Cas d'un test bilatéral*

      H_0 : la variabilité de l'échantillon est conforme à la variabilité théorique ; $H_0 : \sigma^2 = \sigma_0^2$,
      H_1 : σ^2 n'est pas conforme à σ_0^2; $H_1 : \sigma^2 \neq \sigma_0^2$,

      Décision : Si $\chi^2_{\alpha/2}(\lambda) \leq \chi^2 \leq \chi^2_{1-\alpha/2}(\lambda)$, l'hypothèse H_0 ne peut être rejetée ; sinon H_0 est écartée avec une probabilité α de se tromper.

    - *Cas d'un test unilatéral*

      - $H_0 : \sigma^2 \leq \sigma_0^2$,
        $H_1 : \sigma^2 > \sigma_0^2$.

        Décision : Si $\chi^2 \leq \chi^2_{1-\alpha}(\lambda)$, l'hypothèse H_0 ne peut être rejetée ; sinon H_0 est écartée avec une probabilité α de se tromper.

      - $H_0 : \sigma^2 \geq \sigma_0^2$,
        $H_1 : \sigma^2 < \sigma_0^2$.

        Décision : Si $\chi^2 \geq \chi^2_\alpha(\lambda)$, l'hypothèse H_0 ne peut être rejetée ; sinon H_0 est écartée avec une probabilité α de se tromper.

  - ***Cas ou*** $n > 31$

    Les tables du K^2 s'arrêtent habituellement à $\nu =$ 30 degrés de liberté. Dans ce cas, on procède à une approximation de cette loi et on montre que : $U = \sqrt{2K^2} - \sqrt{2\nu - 1}$ suit sensiblement une $N(O, 1)$, ν étant le nombre de degrés de liberté.

On calcule la valeur prise par U en utilisant les observations, et on note :

$$u = \sqrt{\frac{2(n-1)S^2}{\sigma_0^2}} - \sqrt{2n-3}$$

Soit $u_{1-\alpha/2}$ le quantile d'ordre $(1-\alpha/2)$ de la loi $N(0,1)$.

Les conclusions sur les tests se font en utilisant les quantiles u_β de la loi normale à la la place des quantiles $\chi^2_\beta(\lambda)$ de la loi du Khi^2.

– **Exemple**

On étudie un processus de fabrication, pour lequel la variable de qualité X est la contenance d'un flacon (en cl) normalement distribuée. On a mesuré 5 valeurs : 101, 103, 99, 102 et 101. On veut savoir si la variabilité du processus de production a augmenté, ce qui induit une baisse de la qualité. On admet qu'en théorie le processus de fabrication a une variabilité constante et connue, $\sigma_0 = 1$. Doit-on régler le processus de fabrication ?

Réponse : on doit tester $H_0 : \sigma^2 \leq 1$, contre $H_1 : \sigma^2 > 1$.
On a $s = 1.48$, $n = 5$, $\frac{(n-1)s^2}{\sigma_0^2} = \chi^2 = 8.8$ et $\chi^2_{0.95}(4) = 9.49$. Comme $\chi^2 \leq \chi^2_{0.95}(4)$, on ne rejette pas H_0.

2.2 Tests d'homogénéité

2.2.1 Comparaison de deux fréquences expérimentales

– **Position du problème**

Dans deux populations P_1 et P_2, on étudie un caractère statistique à deux modalités A et $\overline{A}$. Chaque individu présente ou non la modalité A. Les fréquences d'apparition de A dans les populations P_1 et P_2 sont les nombres (inconnus) p_1 et p_2. De P_1 et P_2, on extrait deux échantillons E_1 et E_2 de tailles n_1, et n_2, dans lesquels les fréquences d'apparition de A sont f_1 et f_2. Le problème est de savoir si la différence observée entre f_1 et f_2 est significative, ou au contraire explicable par les fluctuations d'échantionnage.
On suppose que :
$n_1 \geq 30$, $n_2 \geq 30$, $n_1 f_1 \geq 5$, $n_1(1-f_1) \geq 5$, $n_2 f_2 \geq 5$, $n_2(1-f_2) \geq 5$.
On note F_1 et F_2 les variables aléatoires qui prennent les valeurs f_1 et f_2.

– **Mise en place du test**

On note $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$. La variable suivante est à la base de la mise au point du test, on pose $U = \frac{F_1 - F_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$.
Sous les conditions présentées au dessus, la variable U est normalement distribuée. On note $\hat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ la valeur de p et $u = \frac{f_1 - f_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ la valeur de U obtenues à partir des données.
On rappelle que u_β le quantile d'ordre β de la loi $N(0,1)$. On supposera que le risque de première espéce est de α.

– *Cas d'un test bilatéral*

H_0 : les fréquences "vraies" (théoriques) sont égales, $p_1 = p_2$,
H_1 : les fréquences "vraies"' sont différentes, $p_1 = p_2$.

Décision : si $-u_{1-\frac{\alpha}{2}} \leq u \leq u_{1-\frac{\alpha}{2}}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

– *Cas d'un test unilatéral*

  – $H_0 : p_1 \geq p_2$, la fréquence de P_1 est plus grande que la fréquence de P_2.
  $H_1 : p_1 < p_2$.

  Décision : si $u \geq -u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

  – $H_0 : p_1 \leq p_2$, la fréquence de P_1 est plus petite que la fréquence de P_2.
  $H_1 : p_1 > p_2$.

  Décision : si $u \leq u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

– **Exemple**

Pour tester l'efficacité d'un vaccin antigrippal on soumet 300 personnes à une expérience : - sur 100 personnes non vaccinées, 32 sont atteintes par la grippe, - sur 200 personnes vaccinées, 50 sont atteintes par la grippe, Ce résultat permet-il d'apprécier l'efficacité du vaccin ? On a le tableau suivant :

	grippé	non grippé	taille
non vacciné	32	68	100
Vacciné	50	150	200
total	82	218	300

Réponse : on calcule les valeurs f_1 et f_2 qui sont les proportions des grippés des deux échantillons. Nous avons $f_1 = 32/100$, $f_2 = 50/200 = 25/100$, $\hat{p} = 82/300 = 41/150 = 0.273333333333$ et $u = 1.2824$. Avec un risque d'erreur $\alpha = 0,05$, $u_{0.975} = 1,96$. Pour un test bilatéral, comme $u \leq u_{0.975}$, nous concluons que les deux échantillons ne sont pas significativement différents au seuil de 5% c'est à dire que le vaccin n'est pas efficace et donc H_0 ne peut être rejetée.

2.2.2 Comparaison de deux moyennes expérimentales dans le cas d'échantillons indépendants

– **Position du problème**

Dans deux populations P_1 et P_2, on étudie un caractère X. On note :
- μ_1 et σ_1^2 l'espérance et la variance de X dans la population P_1.

- μ_2 et σ_2^2 l'espérance et la variance de X dans la population P_2.
Les moyennes μ_1 et μ_2 sont inconnues. Les variances σ_1^2 et σ_2^2 peuvent être connues ou inconnues. On estime ces paramètres à partir de deux échantillons indépendants E_1 et E_2, de tailles n_1 et n_2 tirés dans les populations P_1 et P_2 :
- μ_1 est estimé par $\overline{X}_1$ et σ_1^2 est estimé par s_1^2 à partir des données de E_1,
- μ_2 est estimé par $\overline{X}_2$ et σ_2^2 est estimé par s_2^2 à partir des données de E_2.

Le problème est de savoir si la différence entre les moyennes expérimentales $\overline{X}_1$ et $\overline{X}_2$ est significative, ou au contraire explicable par les fluctuations d'échantillonnage.

– ***Cas où σ_1 et σ_2 sont connues***

  On suppose que la distribution dans les deux échantillons est normale ou que les échantillons sont suffisamment grands . Dans ce cas la variable $U = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ suit une $N(0, 1)$.

  – **Mise en place du test**

    On note $u = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, la valeur de U déterminée à partir des échantillons. Soit u_β le quantile d'ordre β de la loi $N(0, 1)$, et α le risque de première espèce.

    – *Cas d'un test bilatéral*

      $H_0 : \mu_1 = \mu_2$, les deux populations sont homogènes,
      $H_1 : \mu_1 \neq \mu_2$.

      Décision : si $-u_{1-\frac{\alpha}{2}} \leq u \leq u_{1-\frac{\alpha}{2}}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

    – *Cas d'un test unilatéral*
      $H_0 : \mu_1 \geq \mu_2$,
      $H_1 : \mu_1 < \mu_2$.

      Décision : si $u \geq -u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.
      $H_0 : \mu_1 \leq \mu_2$,
      $H_1 : \mu_1 > \mu_2$.

      Décision : si $u \leq u_{1-\alpha}$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

– ***Cas où σ_1 et σ_2 sont inconnues mais égales***

  On suppose que la distribution dans P_1 est $N(\mu_1, \sigma)$ et que la distribution dans P_2 est $N(\mu_2, \sigma)$. Un test préalable d'égalité des variances ne rejette pas l'hypothèse H_0. Dans ce cas la variable

$T = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ suit une loi se Student à $\lambda = n_1 + n_2 - 2$ degrés de liberté, où

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}{n_1+n_2-2}}$$

- **Mise en place du test**

  On note $t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, la valeur de T déterminée à partir des échantillons. Soit $t_\beta(\lambda)$ le quantile d'ordre β de la loi de Student à λ degrés de liberté, et α le risque de première espèce.

  - *Cas d'un test bilatéral*

    $H_0 : \mu_1 = \mu_2$, les deux populations sont homogènes,
    $H_1 : \mu_1 \neq \mu_2$.

    Décision : si $-t_{1-\frac{\alpha}{2}}(\lambda) \leq t \leq t_{1-\frac{\alpha}{2}}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

  - *Cas d'un test unilatéral*

    $H_0 : \mu_1 \geq \mu_2$,
    $H_1 : \mu_1 < \mu_2$.

    Décision : si $t \geq -t_{1-\alpha}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

    $H_0 : \mu_1 \leq \mu_2$,
    $H_1 : \mu_1 > \mu_2$.

    Décision : si $t \leq t_{1-\alpha}(\lambda)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

- **Exemple :** voir plus loin.

- ***Cas où σ_1 et σ_2 sont inconnues et inégales (test d'Aspin-Welch)***

Si un test préalable d'égalité des variances rejette l'hypothèse H_0. La variable $T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ suit une loi Student à m degrés de liberté. Le paramètre m est à déterminer en utilisant la formule suivante :

$$\frac{1}{m} = \frac{c^2}{n_1-1} + \frac{(1-c)^2}{n_2-1}$$

avec

$$c = \frac{s_1^2/n_1}{(s_1^2/n_1)+(s_2^2/n_2)}, \qquad s_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1i}-\bar{x_1})^2}{n_1-1} \text{ et } s_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2i}-\bar{x_2})^2}{n_2-1}$$

– **Mise en place du test**

On note $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, la valeur de T déterminée à partir des échantillons. Soit $t_\beta(m)$ le quantile d'ordre β de la loi de Student à m degrès de liberté, et α le risque de première espèce.

– *Cas d'un test bilatéral*

$H_0 : \mu_1 = \mu_2$, les deux populations sont homogènes,
$H_1 : \mu_1 \neq \mu_2$.

Décision : si $-t_{1-\frac{\alpha}{2}}(m) \leq t \leq t_{1-\frac{\alpha}{2}}(m)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec une probabilité α de se tromper.

– *Cas d'un test unilatéral*

$H_0 : \mu_1 \geq \mu_2$,
$H_1 : \mu_1 < \mu_2$.

Décision : si $t \geq -t_{1-\alpha}(m)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

$H_0 : \mu_1 \leq \mu_2$,
$H_1 : \mu_1 > \mu_2$.

Décision : si $t \leq t_{1-\alpha}(m)$, l'hypothèse H_0 ne peut être rejetée, sinon on écarte H_0 avec un risque égal à α.

– **Exemple**

Soit deux échantillons de taille $n_1 = 13$ et $n_2 = 7$ sur lesquels les moyennes et variances suivantes ont été observées : $\bar{x}_1 = 2.7$ $s_1^2 = 1.2$ $\bar{x}_2 = 4.6$ et $s_2^2 = 5.2$. On suppose que l'égalité des variances n'a pas été retenue à la suite d'un test adéquat. Peut on accepter l'égalité des moyennes ?
Réponse : $t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}} = \frac{4.6-2.7}{\sqrt{\frac{5.2}{7} + \frac{1.2}{13}}} = 2.08$, $\frac{1}{m} = \frac{0.097}{0.698} = 0.13896$, $m \approx 8$. $t_{0.975}(8) = 2.306$.
Comme $t < t_{0.975}(8)$, on ne met pas en évidence de différence entre les deux moyennes.

2.2.3 Comparaison de deux moyennes expérimentales dans le cas d'échantillons appariés

– **Position du problème**

Deux échantillons E_1 et E_2 sont appariés lorsque chaque valeur X_{1i} de E_1 est associée à une valeur X_{2i} de E_2 (appariés = associé par paires). Par exemple E_1 peut être un groupe de malades avant traitement et E_2 le groupe des mêmes malades après traitement. Ces deux

échantillons ont donc la même taille n.
Le problème est de savoir si la différence entre les moyennes $\overline{X}_1$ et $\overline{X}_2$ des échantillons est explicable par les fluctuations d'échantillonnage.

- **Mise en place du test**

  On calcule les n différences

$$D_i = X_{1i} - X_{2i}$$

  L'échantillon $D_1, D_2, ..., D_n$ a pour moyenne $\overline{D}$ et pour écart-type estimé S.
  Le critère de comparaison (statistique du test) utilise la variable

$$\frac{\overline{\mathsf{D}} \quad - \quad \overline{}}{S/\sqrt{n}}.$$

  On est ainsi ramené à la comparaison d'une moyenne expérimentale $\overline{d}$ à une moyenne fixée $\mu_0 = 0$.

  Décision :

  - Si $n > 30$, on utilise une loi normale.
  - Si $n \leq 30$ et si la population est gaussienne, on utilise une loi de Student à $n-1$ degrés de liberté.

- **Remarque**

  Si $n \leq 30$ et si les lois de X dans P_1 et P_2 ne sont pas connues, on peut utiliser le test non paramétrique (voir plus loin) de *Wilcoxon.*

- **Exemple**

  On mesure la fréquence cardiaque de 5 sujets avant et après un effort physique. Les données sont résumées dans le tableau suivant : Y'a-t-il un effet effort sur la fréquence cardiaque.

Sujet	Fréquence avant effort	Fréquence après effort	Différence
1	90	98	+8
2	76	77	+1
3	80	88	+8
4	87	90	+3
5	89	83	-6

Solution : On suppose que la variable D_i est normale. Les calculs donnent $\bar{d} = 2.8$, $s_D = 5.8$. Pour un risque $\alpha = 0.05$, $t_{0.975} = 2.776$, $t = \frac{2.8}{\frac{5.8}{\sqrt{5}}} = 1.07$. Comme $t < t_{0.975}(4)$, on ne rejette pas H_0 c'est à dire l'absence de l'effet effort sur la fréquence cardiaque. cette conclusion surprenante, vient probablement de l'individu 5.

2.2.4 Comparaison de deux variances expérimentales

– **Position du problème**

Soient deux populations P_1 et P_2 et une variable aléatoire X étudiée sur les deux populations. On note $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ les moyennes et variances de la variable X sur les populations P_1 et P_2. On suppose que l'on dispose de deux échantillons E_1 et E_2 de tailles n_1 et n_2 de X mesurées sur les deux populations. On calcule les moyennes $\overline{x}_1$ et $\overline{x}_2$, estimateurs des moyennes μ_1 et μ_2, ainsi que s_1^2 et s_2^2 estimateurs sans biais de σ_1^2 et σ_2^2.
Les échantillons E_1 et E_2 sont supposés être indépendants. On désire savoir si la différence observée sur S_1^2 et S_2^2 est explicable ou tout simplement dûe aux fluctuations d'échantillonnage.

– **Mise en place du test**

Si les deux populations sont gaussiennes, la variable

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

suit une loi de *Fisher-Snédécor* à $(n_1 - 1, n_2 - 1) = (\nu_1, \nu_2)$ degrés de liberté.

Soit $f = \frac{s_1^2}{s_2^2}$ la valeur de F obtenue à partir des observations, α désigne le risque de première espèce. On note $F_{1-\alpha/2}(\nu_1, \nu_2)$ le quantile d'ordre $1-\alpha/2$ de la loi de Fisher-Snédécor à (ν_1, ν_2) degrés de liberté.

On suppose que la numérotation des échantillons conduit à porter au numérateur la plus forte des variances estimées.

On évoque seulement le test bilatéral dans cette section.

$H_0 : \sigma_1^2 = \sigma_2^2$,
$H_1 : \sigma_1{}^2 = \sigma^2{}_2$.

Décision : si $f \leq F_{1-\alpha/2}(\nu_1, \nu_2)$, l'hypothèse H_0 ne peut être rejetée, sinon H_0 est écartée avec une probabilité α de se tromper.

– **Exemple**
Nous cherchons à comparer deux méthodes de séchage du mais. On suppose que les taux de séchage sont normalement distribués. Les données sont résumées dans le tableau suivant :

Taux de séchage avec préchauffage (en %)	Taux de séchage sans préchauffage (en %)
16	20
12	10
22	21
14	10
19	12

Comparer ces deux méthodes en supposant que les deux échantillons sont indépendants (non appariés).

Solution : les calculs intermédiaires donnent $\bar{x}_1 = 16.6$, $s_1 = 3.975$, $\bar{x}_2 = 14.6$, $s_2 = 5.459$

- Test d'égalité des variances : $f = \frac{s_2^2}{s_1^2} = 1.89$, $f_{0.975}(4,4) = 9.6$. Comme $f < f_{0.975}(4,4)$, l'hypothèse d'égalité des variances ne peut être rejetée.
- Test d'égalité des moyennes : $s = 4.775$, $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{5}+\frac{1}{5}}} = 0.662$, $t_{0.975}(5+5-2) = 2.306$. Comme $t < t_{0.975}(8)$, On accepte l'hypothèse d'égalité des moyennes.

3 Les tests non paramétriques

3.1 Les tests du χ^2

3.1.1 Test de conformité : ajustement à une loi théorique

- **Position du problème**

On définit sur une population k événements $E_1, E_2, ..., E_k$ formant un système complet d'événements. Dans le modèle théorique, les probabilités de ces événemements sont $p_1, p_2, ..., p_k$.
Sur un échantillon de taille n les *effectifs observés* des événements sont $O_1, O_2, ..., O_k$. On cherche à savoir si la distribution observée est ajustable à la distribution théorique.
Pour pouvoir confronter les observations et le modèle théorique, on calcule les *effectifs théoriques* dits *effectifs calculés* : $C_i = np_i$ (qui ne sont pas necessairement des entiers) et on compare les O_i et les C_i.

Evénements	Effectifs observés	Effectifs théoriques
E_1	O_1	np_1
E_2	O_2	np_2
...	...	...
E_k	0_k	np_k

- **Mise en place du test**

La variable de décision K est la variable qui prend sur chaque échantillon de taille n la valeur :

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(O_i - C_i)^2}{C_i}.$$

Les hypothèses à tester sont :

- H_0 : la distribution observée est conforme à la distribution théorique choisie,
- H_1 : la distribution observée est différente de la distribution théorique.

Décision :

Sous l'hypothèse H_0, la variable K suit une loi du χ^2 à $\nu = k - 1 - r$ degrés de liberté, où r est le nombre de paramètres qu'il faut éventuellement estimer pour connaître la loi théorique.

On note α le risque de première espèce. A l'aide de la table de la fonction de répartition de la loi de K, on cherche $\chi^2_{1-\alpha}(\nu)$ tel que : $P(K \leq \chi^2_{1-\alpha}) = 1 - \alpha$,

- si $\hat{\chi}^2 \geq \chi^2_{1-\alpha}(\nu)$, on écarte l'hypothèse H_0 avec une probabilité α de se tromper,
- si $\hat{\chi}^2 < \chi^2_{1-\alpha}(\nu)$, on ne peut pas écarter l'hypothèse H_0.

- **Exemple**

On a effectué le croisement de balsamines blanches avec des balsamines pourpres. Les fleurs de la première génération sont toutes pourpres. On obtient en deuxième génération 4 catégories avec les fleurs suivantes :

Couleur	pourpre	rose	blanc lavande	blanc
Effectif	1790	547	548	213

Peut-on accepter l'hypothèse de répartition Mendelienne $(9/16, 3/16, 3/16, 1/16)$ au risque de première espèce de 5% ?

Solution : Il s'agit d'ajuster une répartition observée à une répartition théorique. Les hypoythèses à tester sont :
H_0 : la distribution observée est ajustable par la distribution Mendélienne
H_1 : la distribution observée est différente de la distribution Mendélienne.
Sous l'hypothèses H_0, nous avons le tableau récapitulatif suivant :

Couleur	pourpre	rose	blanc lavande	blanc	**totaux**
O_i	1790	547	548	213	**3098=n**
p_i	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	**1**
$C_i = np_i$	1742.625	580.875	580.875	193.625	**3098**

$\hat{\chi}^2 = \sum_{i=1}^{4} \frac{(O_i - C_i)^2}{C_i} = 7.06$, $\nu = 4 - 1$ et $\chi^2_{1-\alpha}(\nu) = 7.81$.

Comme $7.06 < 7.81$, l'hypothèse H_0 ne peut être rejetée.

- **Remarque**

On exige en général que $C_{ij} \geq 5$ pour tout i et pour tout j. Si ce n'est pas le cas, on fait des regroupements entre les A_i.

3.1.2 Test d'homogénéité : comparaison de plusieurs échantillons

- **Position du problème**

Sur une population P, un caractère peut prendre k valeurs $A_1, ..., A_k$ (ou k modalités, ou k classes). On dispose de l échantillons $E_1, E_2, ..., E_l$ pouvant provenir de P. Pour tout $i \in \{1, ..., k\}$ pour tout $j \in (1, .., l)$, on connaît O_{ij} effectif observé de la valeur A_i dans l'échantillon E_j et C_{ij} l'effectif théorique correspondant.

Evénements \ Echantillons	E_1	E_2	...	E_j	...	E_l
A_1	O_{11}	O_{12}	...	O_{1j}	...	O_{1l}
A_2	O_{21}	O_{22}	...	O_{2j}	...	O_{2l}
...	...	...	...	...	...	...
A_i	O_{i1}	O_{i2}	...	O_{ij}	...	O_{il}
...	...	...	...	...	...	...
A_k	O_{k1}	O_{k2}	...	O_{kj}	...	O_{kl}

On note l'effectif total des échantillons

$$N = \sum_{j=1}^{l} \sum_{i=1}^{k} O_{ij}.$$

On cherche à savoir si ces échantillons proviennent de la même population.

– **Mise en place du test**

La variable de décision K est la variable qui prend sur chaque échantillon de taille N la valeur :

$$\hat{\chi}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(O_{ij} - C_{ij})^2}{C_{ij}}.$$

Les hypothèses à tester sont :

- H_0 : les échantillos sont issus de la même population P,
- H_1 : les échantillons proviennent de populations différentes.

Décision :

Sous H_0, la variable K suit une loi du χ^2 à $\nu = (k-1)(l-1)$ degrés de liberté.
On note α le risque de première espèce. A l'aide de la table de la fonction de répartition de la loi de K, on cherche $\chi^2_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi de K à ν degrés de liberté.

On commence par calculer des effectifs théoriques C_{ij} sous l'hypothèse H_0.
Les l échantillons sont réunis en un seul échantillon de taille N, et la probabilité de A_i est égale à :

$$p_i = \frac{\sum_{j=1}^{l} O_{ij}}{N} = \frac{S_i}{N}.$$

L'effectif calculé de la classe A_i pour l'échantillon E_j est alors :

$$C_{ij} = p_i(\sum_{i=1}^{k} O_{ij}) = p_i T_j = \frac{S_i T_j}{N}.$$

- Si $\hat{\chi}^2 \geq \chi^2_{1-\alpha}$, on écarte l'hypothèse H_0 avec une probabilité α de se tromper,
- Si $\hat{\chi}^2 < \chi^2_{1-\alpha}$, on ne peut pas écarter l'hypothèse H_0.

– **Exemple**

Dans la comparaison du taux d'occupation d'un matériel coûteux pour un mois d'hiver (janvier) et un mois d'été (juillet), on dispose de deux échantillons, l'un de 300 observations instantanées en janvier et l'autre de 200 observations instantanées en juillet. Les données sont résumées dans le tableau suivant :

	janvier	juillet
occupation	240	150
inoccupation	60	50

Peut-on considérer que le taux d'occupation de ce matériel est le même en janvier et en juillet avec un risque de 5% ?

Solution : il s'agit de comparer les distributions sur deux échantillons. Les hypothèses à tester sont :

H_0 : le taux d'occupation est le même en janvier qu'en juillet
H_1 : les taux d'occupation sont différents.

Soit le tableau récapitulatif suivant (les effectifs théoriques sont présentés entre parenthèses) :

échantillons\événements	occupation	inoccupation	totaux
janvier	240 (234)	60 (66)	300
juillet	150 (156)	50 (44)	200
totaux	390	110	500

$\hat{\chi}^2 = \sum_i \sum_j \frac{(O_{ij}-C_{ij})^2}{C_{ij}} \approx 1.75$, $\nu = (2-1)(2-1)$ et $\chi^2_{1-\alpha}(\nu) = 3.84$

Comme $1.75 < 3.84$, l'hypothèse H_0 ne peut être rejetée.

– **Remarque**

On exige en général que $C_{ij} \geq 5$ pour tout i et pour tout j. Si ce n'est pas le cas, on fait des regroupements entre les A_i.

3.1.3 Test d'indépendance de deux caractères

– **Position du problème**

Dans une population P, chaque individu possède deux caractères qualitatifs A et B ayant les modalités respectives $A_1, ..., A_k$ et $B_1, ..., B_l$.
Pour tout $i \in \{1, ..., k\}$ pour tout $j \in \{1, .., l\}$, O_{ij} désigne le nombre d'individus présentant les modalités A_i et B_j.
La taille de l'échantillon étudié est donc $N = \sum_{j=1}^{l} \sum_{i=1}^{k} O_{ij}$. Les effectifs marginaux sont

$T_j = \sum_{i=1}^{k} O_{ij}$ et $S_i = \sum_{j=1}^{l} O_{ij}$.

Les données sont résumées dans le tableau de contingence suivant :

Caractère A \ Caractère B	B_1	B_2	...	B_j	...	B_l
A_1	O_{11}	O_{12}	...	O_{1j}	...	O_{1l}
A_2	O_{21}	O_{22}	...	O_{2j}	...	O_{2l}
...	...	...	...	...	...	...
A_2	O_{i1}	O_{i2}	...	O_{ij}	...	O_{il}
...	...	...	...	...	...	...
A_k	O_{k1}	O_{k2}	...	O_{kj}	...	O_{kl}

On cherche à savoir si les caractères A et B sont indépendants.

– **Mise en place du test**

On note C_{ij} le nombre théorique d'individus ayant les modalités A_i et B_j, et on note K la variable aléatoire prenant sur chaque échantillon de taille N la valeur :

$$\hat{\chi}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(O_{ij} - C_{ij})^2}{C_{ij}}.$$

Les hypothèses à tester sont :

– H_0 : les caractères A et B sont indépendants,
– H_1 : les caractères A et B sont dépendants.

Décision :

Sous H_0, la variable K suit une loi du χ^2 à $\nu = (k-1)(l-1)$ degrés de liberté. On note α le risque de première espèce. A l'aide de la table de la fonction de répartition de la loi de K, on cherche $\chi^2_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi de K à ν degrés de liberté.

On commence par calculer des effectifs théoriques C_{ij} sous l'hypothèse H_0. C'est l'effectif des individus présentants les modalités A_i et B_j si l'hypothèse H_0 était vérifiée.

Sous H_0, les événements A_i et B_j sont indépendants et on a :

$$P(A_i \cap B_j) = P(A_i) \times P(B_j) = \frac{C_{ij}}{N} = \frac{S_i}{N}\frac{T_j}{N}.$$

On a donc : $C_{ij} = \frac{S_i T_j}{N}$. La suite du test se fait comme dans le cas précédent :
– si $\hat{\chi}^2 \geq \chi^2_{1-\alpha}$, on écarte l'hypothèse H_0 : les deux caractères ne sont pas indépendants,
– si $\hat{\chi}^2 < \chi^2_{1-\alpha}$, on ne peut pas écarter H_0 : les deux caractères sont indépendants.

– **Exemple**

A la suite d'un même traitement, on a observé 40 bons résultats chez 70 malades jeunes et 50 bons résultats chez 100 malades âgés. Peut on dire au risque 10% qu'il existe une liaison entre l'âge du malade et l'effet du traitement ?

Solution : les hypothèses à tester sont :
H_0 : les effet du traitement sont indépendants de l'âge du malade,
H_1 : le traitement dépend de l'âge du malade.

âge\résultat	bons	mauvais	totaux
jeunes	40 (37.06)	30 (32.94)	70
âgés	50 (52.94)	50 (47.06)	100
totaux	90	80	170

$\hat{\chi}^2 = \sum_i \sum_j \frac{(O_{ij}-C_{ij})^2}{C_{ij}} \approx 0.84$, $\nu = (2-1)(2-1)$ et $\chi^2_{1-\alpha}(\nu) = 2.71$

Comme $0.84 < 2.71$, l'hypothèse H_0 ne peut être rejetée. Il n'y'a pas de liaison entre l'âge du malade et l'effet du traitement.

– **Remarques**

On doit avoir là encore :$\forall i, \forall j$, $C_{ij} \geq 5$. Sinon on regroupe certaines modalités de caractères.

3.2 Test Mann-Whitney

– **Position du problème**

On dispose de deux échantillons E_A et E_B, indépendants et non-exhaustifs, issus de deux populations A et B, de tailles respectives n_A et n_B. On veut comparer les distributions (continues ou non) des deux populations, par conséquent comparer les deux moyennes μ_A et μ_B de ces deux distributions.

– **Mise en place du test**

On classe par ordre croissant l'ensemble des valeurs des deux échantillons. On effecte à chaque valeur de $E_A \cup E_B$ son rang dans ce classement. S'il y'a des ex-aequo, on attribue à chacun un rang moyen égal à la moyenne des rangs qu'ils occupent (par exemple, s'il y'a deux quatrièmes ex-aequo, on attribue à chacun d'eux le rang 4.5).

Soit R_A et R_B la somme des rangs obtenus de chaque série en ordonnant les $(n_A + n_B)$ observation, et

$$U_A = n_A n_B + \frac{n_A(n_A+1)}{2} - R_A \qquad \text{et} \qquad U_B = n_A n_B + \frac{n_B(n_B+1)}{2} - R_B.$$

On pose $U = min(U_A, U_B)$. α désigne le risque de première espèce, et $U_\alpha(n_A, n_B)$ une valeur critique telle que définie ci-dessous :

– si $min(n_A, n_B) > 8$, $U_\alpha(n_A, n_B) = \frac{n_A n_B}{2} - u_p \sqrt{\frac{n_A n_B (n_A+n_B+1)}{12}}$
  où u_p est le fractile d'ordre p de $N(0,1)$, et p prend la valeur $(1-\alpha/2)$ ou bien $(1-\alpha)$ selon

qu'il s'agisse d'un test bilatéral ou unilatéral.

– si $max(n_A, n_B) \leq 8$, les valeurs critiques sont données par la table adéquate.

– *Test bilatéral*

  – $H_0 : (A) = (B) \qquad [\mu_A = \mu_B]$
  – $H_1 : (A) \neq (B) \qquad [\mu_A \neq \mu_B]$

  Décision :

  si $U > U_\alpha(n_A, n_B)$, on accepte H_0, sinon on rejette H_0.

– *Test unilatéral*

  – $H_0 : (A) = (B) \qquad [\mu_A = \mu_B]$
  – $H_1 : (A) > (B) \qquad [\mu_A > \mu_B]$.

Décision :

si $U > U_\alpha(n_A, n_B)$, on accepte H_0, sinon on rejette H_0.

– **Exemple**

Deux groupes, A et B, de 10 étudiants, formés à des méthodes pédagogiques différentes ont subi le même examen. A l'issue de cet examen, les notes attribuées aux deux groupes sont les suivantes :

groupe A : 0, 3, 4, 5, 7, 8, 8, 12, 15, 17
groupe B : 1, 6, 10, 11, 13, 14, 15, 18, 19, 20

Au vu des résultats, aux risques 5% et 1%, ces méthodes sont elles différentes ?

Solution : les hypothèses à tester sont :
H_0 : les deux méthodes sont similaires,
H_1 : les méthodes sont différentes.

comme nous ne disposons pas d'hypothèse de normalité sur les données, nous allons utiliser le test (non paramétrique) de Mann-Witney. Dans le tableau suivant, nous avons remplacer les notes par leur classement.

rg A	1		3	4	5		7	8.5	8.5			12			15.5	17			
rg B		2				6				10	11		13	14	15.5		18	19	20

$U_A = 10 \times 10 + \frac{10 \times 11}{2} - 81.5 = 73.5$, $U_B = 10 \times 10 + \frac{10 \times 11}{2} - 128.5 = 26.5$, $U = 26.5$
$U_{0.05}(10, 10) = 23$, $U_{0.01}(10, 10) = 16$. Comme, dans les deux cas, $U > U_\alpha(10, 10)$, l'hypothèse H_0 ne peut être rejetée.

3.3 Test Wilcoxon

- **Position du problème**

  On mesure deux variables continues X et Y sur une même unité statistique ou sur deux unités statistiques appariées ; par exemple le rendement de deux variétés de blé cultivées sur la même parcelle ou sur des parcelles adjacentes, ou l'effet de deux traitements T_1 et T_2 sur le même sujet, mais à des instants différents (on suppose dans ce cas que le premier traitement a un effet de courte durée et que le deuxième est effectué après que l'effet du premier ait disparu). On dispose de n paires de mesures (X_i, Y_i). L'hypothèse nulle H_0 concerne l'égalité des deux distributions $(X) = (Y)$ et par conséquent l'égalité des moyennes μ_X et μ_Y de ces deux distributions.

- **Mise en place du test**

  On note $d_i = X_i - Y_i$. On classe par ordre croissant les $\mid d_i \mid$ et on affecte à chaque différence son rang dans ce classement. Chaque fois que la différence est nulle, on diminue d'une unité le nombre n de paires considérées. S'il y'a des ex-aequo, on attribue à chacun un rang égal à la moyenne des rangs qu'ils occupent.
  On note N le nombre des différences non nulles.

  On calcule : R^+ somme des rangs des différences positives et R^- somme des rangs des différences négatives. On vérifie que :

  $$R^+ + R^- = \frac{N(N+1)}{2}.$$

  On note

  $$R = min(R^+, R^-)$$

  Décision :

  - Test bilatéral

    On désigne par α le risque de première espèce, et $R_\alpha(N)$ une valeur critique telle que définie ci-dessous (où u est le fractile de la loi N(0,1)) :

    1) si $N > 20$, $R_\alpha(N) = \frac{N(N+1)}{4} - u_{1-\alpha/2}\sqrt{\frac{N(N+1)(2N+1)}{24}}$,
    2) Si $N \leq 20$, les valeurs critiques sont données par la table adéquate.

    - $H_0 : (X) = (Y) \quad [\mu_X = \mu_Y]$,
    - $H_1 : (X) \neq (Y) \quad [\mu_X \neq \mu_Y]$.

    Si $R > R_\alpha(N)$, on accepte H_0, sinon on rejette H_0.

  - Tests unilatéraux

    On désigne par α le risque de première espèce, et $R_\alpha(N)$ une valeur critique telle que définie ci-dessous :

1) si $N > 20$, $R_\alpha(N) = \frac{N(N+1)}{4} - u_{1-\alpha}\sqrt{\frac{N(N+1)(2N+1)}{24}}$,
2) si $N \leq 20$, les valeurs critiques sont données par la table adéquate.

1. Test unilatéral à droite
   - $H_0 : (X) = (Y)$ $[\mu_X = \mu_Y]$,
   - $H_1 : (X) > (Y)$ $[\mu_X > \mu_Y]$.

   Si $R^- > R_\alpha(N)$, on accepte H_0, sinon on rejette H_0.

2. Test unilatéral à gauche
   - $H_0 : (X) = (Y)$ $[\mu_X = \mu_Y]$,
   - $H_1 : (X) < (Y)$ $[\mu_X < \mu_Y]$.

   Si $R^+ > R_\alpha(N)$, on accepte H_0, sinon on rejette H_0.

– **Exemple**

Dans le cadre d'une expertise clinique de validation d'un médicament M, on administre à 10 malades, successivement à chacun et dans un ordre tiré au sort, le médicament M et une même dose d'un médicament de référence R. Les effets de ces substances sur chacun des 10 malades sont :

M	5	4	2	3	4	3	8	5	4	5
R	6	3	3	1	1	3	4	2	5	7

Peut on dire que les médicaments M et R ont un effet significativement différents au risque de 5% ?

Solution : les hypothèses à tester sont

- H_0 : il n'y'a pas de différence significative entre les effets moyens des médicaments,
- H_1 : les effets des médicaments sont différents.

Les échantillons étant de petites tailles et aucune hypothèse de normalité n'est faite sur les données, on utilise le test de Wilcoxon. Pour se faire, soit le tableau suivant :

M - R	-1	1	-1	-1	2	0	-2	3	3	4
rang moyen	2.5	2.5	2.5	2.5	5.5		5.5	7.5	7.5	9

$R^+ = 2.5 + 5.5 + 7.5 + 7.5 + 9 = 32$, $R^- = 2.5 + 2.5 + 2.5 + 5.5 = 13$, $R = min(R^+, R^-) = 13$.

D'après la table, nous avons :

- pour $\alpha = 0.05$, $R_\alpha = 6$,
- pour $\alpha = 0.01$, $R_\alpha = 2$

Comme $R > R_\alpha$, on ne peut donc pas rejeter H_0 : Il n'y'a pas de différence significative entre l'action du médicament testé et celle du médicament de référence.

3.4 Test de Kruskal et Wallis

- **Position du problème**

On dispose de k échantillons, indépendants et non-exhaustifs, $E_1, E_2, ..., E_k$ d'effectifs $n_1, n_2, ..., n_k$; prélevés dans k populations $A_1, A_2, ..., A_k$. On désire tester l'identitité des distributions dans les k populations.

- **Mise en place du test**

On classe par ordre croissant l'ensemble des valeurs de ces k échantillons. Puis on détermine le rang de chaque valeur, de la même manière que dans les tests précédents s'il y'a des ex-aequo.

Pour chaque échantillon E_i, on note R_i la somme des rangs des valeurs de cet échantillon.

On pose $N = \sum_{i=1}^{k} n_i$. A partir des données, on calcule :

$$H = \frac{12}{N(N+1)} \sum_{i=1} \frac{R_i^2}{n_i} - 3(N+1)$$

Cette valeur est divisée par $1 - \frac{\sum_i T_i}{N^3 - N}$ où $T = (t-1)t(t+1)$ pour chaque groupe de t rangs ex-aequo.

Décision :

Les hypothèses à tester sont :

- $H_0 : (A_1) = (A_2) = ... = (A_k)$
- $H_1 : \mu_{A_i} \neq \mu_{Aj}$ pour au moins un couple i, j

ou

- $H_0 : \mu_{A_1} = \mu_{A_2} = ... = \mu_{A_k}$
- $H_1 : \mu_{A_i} \neq \mu_{Aj}$ pour au moins un couple i, j

On désigne par α le risque de première espèce, et χ^2 la variable du Khi-deux à $\nu = k - 1$ degrés de liberté.

Si $H < \chi^2_{1-\alpha}(k-1)$, on accepte H_0, sinon on rejette H_0.

- **Exemple**

Un test d'effort est proposé à trois groupes de personnes. Les scores sont relevés dans le tableau suivant :

Groupe\individu	1	2	3	4	5	6
groupe 1	44	44	54	32	21	28
groupe 2	70	77	48	64	71	75
groupe 3	80	76	34	80	73	80

Ces trois groupes sont-ils homogènes ?.

Solution : les scores sont remplacés dans le tableau suivant par leur classement :

Groupe\individu	1	2	3	4	5	6	moyenne
groupe 1	5.5	5.5	8	3	1	2	4.17
groupe 2	10	15	7	9	11	13	10.83
groupe 3	17	14	4	17	12	17	13.50

A partir des classements, nous avous :

- $H = 9.73$,
- $H = 9.78$, en tenant compte de la correction due aux ex-aequo.

D'autre part $\nu = k - 1 = 3 - 1 = 2$, $\chi^2_{1-\alpha}(\nu) = \chi^2_{0.95}(2) = 5.99$.

Comme $H > \chi^2_{0.95}(2)$, l'hétérogénéité globale est significative au seuil 5%.

3.5 Coefficient de corrélation de rang de Spearman

- **Position du problème**
  Sur une population, on considère deux variables X et Y. On dispose généralement de n couples (X_i, Y_i) de valeurs de (X, Y). Nous cherchons à savoir s'il y'a corrélation entre X et Y sans rien connaitre sur les distributions de X et Y.
  - *Définition*

  On range par ordre croissant, séparément, les valeurs $\{x_1, x_2, ..., x_n\}$ et $\{y_1, y_2, ..., y_n\}$. On remplace alors chaque x_i par son rang x_i^r et chaque y_i par son rang y_i^r. En cas d'ex-aequo, on procède comme dans les tests précédents.

  On appelle coefficient de corrélation de rang de Spearman le nombre r_s égal au coefficient de corrélation calculé à partir de x_i^r et de y_i^r.

- Mise en place du test Les hypothèses à tester sont :

  - H_0 : absence de corrélation entre X et Y,
  - H_1 : X et Y sont corrélées.

  On pose $d_i = x_i^r - y^r{}_i$, et on calcule

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}.$$

Soit R_s la variable aléatoire qui prend la valeur r_s à l'issue de l'expérience aléatoire.

1. $n \geq 10$.

   Si H_0 est vraie, la variable aléatoire :

   $$T = \frac{R_s\sqrt{n-2}}{\sqrt{1-R_s^2}}$$

   suit approximativement la loi de student à $n-2$ degrés de liberté.
   On calcule donc

   $$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}.$$

   On désigne par α le risque de première espèce du test, et $t_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi de Student à $n-2$ degrés de liberté.

   Si ρ_s désigne le coefficient de corrélation de rang de Spearman au niveau des populations, l'hypothèse nulle H_0 que l'on va tester s'écrit : $\rho_s = 0$

   - si $-t_{1-\alpha/2} < t < t_{1-\alpha/2}$, H_0 ne peut pas être rejetée au risque α.
   - si $\mid t \mid \geq \mid t_{1-\alpha/2} \mid$, H_0 est rejetée au risque α. On conclut que X et Y sont liées, sans savoir de quelle façon.

2. $n < 10$.

   On ne peut plus utiliser l'approximation précédente mais une table appropriée donne, en fonction de n et de α, la valeur r_α telle que, sous H_0, on ait

   $$P(\mid R_s \mid > r_\alpha) = \alpha$$

   On rejette donc l'hypothèse nulle si $\mid r_s \mid > r_\alpha$.

– **Exemple**

On a étudié l'inhibition de la cholinestérase par une série de composés organophosphorés. Pour chaque composé, on a déterminé :

- le pouvoir inhibiteur, exprimé par la constante de formation K du complexe enzyme-composé,
- la lipophilie, exprimée par le coefficient de partage P du composé entre l'eau et l'octanol.

Les valeurs obtenues sont les suivantes : y'a-t-il une corrélation significative entre l'action inhi-

log K	2.27	2.44	2.46	2.56	3.08	3.23	3.27	3.32	3.71
log P	0.089	-0.67	0.021	0.66	0.82	1.88	2.53	2.39	1.67

bitrice et la lipophilie ?

Solution : les hypoyhèses à tester sont :

- H_0 : il n'y'a pas de corrélation entre log(K) et log(p),
- log(K) et log(p) sont corrélés.

Déterminons les rangs des valeurs après les avoir rangées par ordre croissant :

rang(log K)	1	2	3	4	5	6	7	8	9
rang(log P)	3	1	2	4	5	7	9	8	6

$r_s = 1 - \frac{6}{9(81-1)}[(-1)^2 + (1^2 + ... + 3^2] \approx 0.83$. D'après la table, nous avons
- pour $\alpha = 0.05$, $r_{0.05} = 0.68$ - pour $\alpha = 0.01$, $r_{0.01} = 0.82$.

Comme $r_s \notin]-r_\alpha, r_\alpha]$, on rejette donc H_0 au risque 5% et 1% : la corrélation est significative.

Remarque :

Dans certaines références statistiques, on procède ainsi :
Sous l'hypothèse d'indépendance entre les deux variables, on peut montrer que

$$E(R_s) = 0 \text{ et} V(R_s) = \frac{1}{n-1}.$$

Si n est grand ($n \geq 30$), la variable $U = R_s\sqrt{n-1}$ suit approximativement une $N(0,1)$.

On note $u = r_s\sqrt{n-1}$ la valeur de U obtenue à partir des observations. On retrouve un test équivalent à un test de moyenne de loi normale.

Dans le cas d'un test bilatéral, avec un risque α, la règle de décision est la suivante :
Si $\mid u \mid < u_{1-\alpha/2}$, on ne rejette pas H_0.

Pour $n < 30$, on utilise les tables usuelles.