

## **Terminale S - Probabilités et statistiques - Estimation**

*Thème d'aide personnalisée : « Prise de décision lors de la comparaison de deux proportions (par exemple lors d'un essai thérapeutique) »*

### **1. Principe des tests d'hypothèse**

Les tests statistiques ont comme objectif l'aide à la décision et la validation d'hypothèses. Ils sont utilisés chaque fois que l'on veut comparer un paramètre à une valeur de référence ou plusieurs paramètres entre eux. Les paramètres peuvent être quantitatifs (moyenne, médiane, ...) ou qualitatifs (pourcentages). Ces paramètres sont issus d'échantillons statistiques qui vont servir à émettre des conclusions sur la population d'où ils sont issus.

Dans le problème du test statistique, l'échantillon sert à trancher entre deux hypothèses en concurrence concernant cette population.

#### **1.1 Estimation et test**

Premier exemple : on dispose d'un échantillon de 400 adolescents dont on a mesuré la taille en cm. On peut se demander si cet échantillon, dont la moyenne observée  $m$  est égale à 172,23 cm et l'écart-type observé est égal à 2,50 cm est représentatif de la population générale où la taille moyenne est  $\mu_0 = 171,33$  cm et l'écart-type  $\sigma$ . Une autre façon de formuler la question est de se demander si les 400 valeurs observées sont compatibles avec l'hypothèse  $\mu = 171,33$  cm.

Deuxième exemple : ***L'essai thérapeutique est une procédure qui vise à comparer des traitements d'une même maladie afin de sélectionner le plus performant.***

Considérons un essai comparant deux médicaments antiulcéreux A et B ; 150 patients ont été répartis par tirage au sort en deux groupes de 75 : l'un recevant A, l'autre recevant B. A l'issue d'un mois et demi de traitement, les patients sont évalués quant au succès du traitement (cicatrisation complète) ou à son échec (non cicatrisation complète).

De la même façon que précédemment, on peut estimer à partir de chacun des deux échantillons, les fréquences théoriques de guérison dans la population des patients prenant le traitement A et dans la population des patients prenant le traitement B.

Deux cas de figure peuvent se présenter :

- a. Le pourcentage de guérison est « légèrement » supérieur avec A qu'avec B mais comment interpréter cette différence ? Existe-t-elle réellement (différence liée au traitement) ou est-elle due au tirage au sort, c'est-à-dire au hasard ?
- b. Le pourcentage observé est « très nettement » supérieur avec A qu'avec B. Comme précédemment, on peut se poser les mêmes questions mais intuitivement, on est peu enclin à croire que la différence soit due au tirage au sort.

#### **1.2 Principe des tests statistiques**

Si notre échantillon d'adolescents est bien représentatif de la population générale, on s'attend à trouver sur notre échantillon une moyenne de tailles « proche » de 171,33 cm. Mais il faut définir

ce qu'on entend par « proche » car on sait que la moyenne observée fluctue d'un échantillon à l'autre autour de cette valeur. Dans notre exemple, on observera dans 95% des cas, une moyenne de tailles comprise dans l'intervalle  $\left[ \mu - 1,96 \sqrt{\frac{s^2}{n}}; \mu + 1,96 \sqrt{\frac{s^2}{n}} \right]$ ,

soit  $\left[ 171,33 - 1,96 \sqrt{\frac{2,5^2}{400}}; 171,33 + 1,96 \sqrt{\frac{2,5^2}{400}} \right]$ , soit encore  $[171,09; 171,58]$ .

Observer une moyenne supérieure à 171,58 ou inférieure à 171,09 est donc peu probable si l'hypothèse  $\mu = 171,33$  est exacte ; cela arrive dans moins de 5% des cas.

On considèrera, dans ce cas, que l'hypothèse doit être rejetée.

**Le principe des tests statistiques revient à formaliser le raisonnement selon les étapes suivantes :**

- 1. Formuler une hypothèse**
- 2. En déduire ce que devraient être les observations si cette hypothèse est vraie**
- 3. Vérifier si les observations sont conformes à ce qu'on attend sous l'hypothèse en question**
- 4. Conclure par un rejet ou un non rejet de l'hypothèse initiale**

### **1.2.1. Formulation des hypothèses**

On est donc amené à énoncer deux hypothèses : **l'hypothèse  $H_0$  appelée hypothèse nulle et  $H_1$  appelée hypothèse alternative.**

Dans notre exemple sur les tailles des 400 adolescents, l'hypothèse nulle  $H_0$  se formule ainsi : « notre échantillon d'adolescents est représentatif de la population générale des adolescents concernant la taille » ( $\mu = \mu_{H_0}$  soit  $\mu = 171,33$ ).

L'hypothèse alternative  $H_1$  est l'hypothèse que l'on propose en cas de rejet de  $H_0$ . Elle se formule ainsi : « notre échantillon d'adolescents n'est pas représentatif de la population générale des adolescents concernant la taille » ( $\mu \neq \mu_{H_0}$  soit  $\mu \neq 171,33$ ).

### **1.2.2. Déterminer ce que devraient être les observations si $H_0$ était vraie**

Si  $H_0$  est vraie, les fluctuations d'échantillonnage de la quantité  $\varepsilon = \frac{|\mu_0 - \mu_{H_0}|}{\sqrt{\frac{\sigma^2}{n}}}$  autour de 0 suivent

une loi normale centrée réduite (sous réserve que l'effectif de l'échantillon soit supérieur ou égal à 30). Ainsi, sous  $H_0$ ,  $\varepsilon$  ne dépasse (en valeur absolue) la valeur 1,96 que dans 5% des cas.

### **1.2.3. Vérifier si les observations sont conformes aux valeurs attendues sous l'hypothèse nulle**

Ceci revient à calculer sur l'échantillon la quantité  $\varepsilon_0 = \frac{|m_0 - \mu_{H_0}|}{\sqrt{\frac{s^2}{n}}}$  ( $\varepsilon_0$  est donc une quantité

observée) et à déterminer si la probabilité que  $\varepsilon$  soit supérieur ou égal en valeur absolue à la quantité observée  $\varepsilon_0$  si  $H_0$  est vraie.

Si cette probabilité est « petite », deux possibilités existent :

- Soit  $H_0$  est vraie et l'échantillon en question a peu de chances d'être observé car  $m_0$  est éloigné de la valeur attendue  $\mu_{H_0}$  à la suite de fluctuations d'échantillonnage.

- Soit  $H_0$  est fausse.

Il faut trancher entre ces deux hypothèses et on préférera rejeter  $H_0$  si l'échantillon observé est trop peu probable.

Le choix de la règle de décision revient à quantifier la probabilité « trop petite ».

#### 1.2.4. Choisir la règle de décision

Deux décisions sont possibles :

- Soit on rejette  $H_0$  car les observations faites sur l'échantillon sont incompatibles avec cette hypothèse, c'est-à-dire que la moyenne observée sur l'échantillon est trop « éloignée » de la moyenne théorique sous  $H_0$ .
- Soit on ne rejette pas  $H_0$  car les observations faites sur l'échantillon sont compatibles avec cette hypothèse, c'est-à-dire que la moyenne observée sur l'échantillon peut-être expliquée par les fluctuations d'échantillonnage ; on ne peut cependant pas affirmer qu'elle est vraie.

**La règle de décision d'un test statistique doit donc permettre, en prenant un risque acceptable, de trancher entre le rejet et le non rejet de  $H_0$  sur la base des observations faites sur l'échantillon.**

Le principe est de fixer une valeur seuil que la quantité calculée  $\varepsilon_0$  devra au moins égaler pour que l'on décide de rejeter  $H_0$ , et qu'elle ne devra pas dépasser pour que l'on décide de ne pas rejeter  $H_0$ .

Si $\varepsilon_0 \geq \text{seuil}$	rejet de $H_0$
Si $\varepsilon_0 < \text{seuil}$	non rejet de $H_0$

## 2. Seuil de signification, risque de première espèce

Quelle que soit la décision que l'on prenne, elle comporte un risque d'erreur :

Si on rejette  $H_0$ , on risque de la rejeter alors qu'elle est vraie.

On minimise ce risque lorsqu'on choisit le seuil suffisamment grand ; ce risque fixé a priori est appelé **risque de première espèce** et noté  $\alpha$ . Il est généralement fixé à 5%, ce qui signifie que l'on a 5% de chances de se tromper en rejetant  $H_0$ . Sur la table de la loi centrée réduite, valable pour les grands échantillons, cette probabilité de 5% correspond à une valeur seuil de 1,96.

*Si l'on calcule  $\varepsilon_0$  dans notre exemple sur les tailles de notre échantillon de 400 adolescents dont la moyenne observée est 172,23 cm et l'écart-type 2,50 cm, on obtient :*

$$\varepsilon_0 = \frac{|172,23 - 171,33|}{\sqrt{\frac{2,50^2}{400}}} \quad \text{soit } \varepsilon_0 \cong 7,2.$$

*Cette valeur est très supérieure à la valeur seuil 1,96. On peut donc rejeter  $H_0$  avec sécurité.*

<b><math>\alpha</math> est appelé risque de première espèce ou seuil de signification</b>
---

## 3. Degré de signification

La règle de décision édictée ci-dessus nous permet de rejeter ou non  $H_0$  mais ne nous permet pas de donner un poids à la décision prise ; on peut juste dire qu'en cas de rejet de  $H_0$ , nous avons moins de 5% de chances de nous tromper.

Prenons l'exemple d'un deuxième échantillon dont la taille moyenne observée est 171,60 cm et l'écart-type 2,50 cm. On obtient pour cet échantillon :  $\varepsilon_0 = \frac{|171,60 - 171,33|}{\sqrt{\frac{2,50^2}{400}}} \text{ soit } \varepsilon_0 \cong 2,16.$

La quantité calculée est encore supérieure au seuil pour 5% de 1,96 mais beaucoup moins que pour l'échantillon précédent.

Si l'on se réfère à la table de la loi centrée réduite, on constate que **le degré de signification**, noté  $p$ , permet de quantifier la force de notre décision. Il est égal à la probabilité d'observer, si  $H_0$  est vraie, un écart à la valeur du paramètre fixé par  $H_0$ , au moins aussi grand que celui constaté sur l'échantillon.

Dans nos deux exemples, nous rejetons  $H_0$  mais dans le premier cas, nous l'aurions également rejeté si nous avions fixé  $\alpha$  à 1% alors que dans le deuxième nous ne l'aurions pas rejeté à 1%.

On admet les règles suivantes :

**Lorsque  $p \leq 0,1\%$ , on dit que le test est hautement significatif**

**Lorsque  $0,1\% < p \leq 1\%$ , on dit que le test est très significatif**

**Lorsque  $1\% < p \leq 5\%$ , on dit que le test est significatif**

**Lorsque  $p > 5\%$ , on dit que le test n'est pas significatif**

*Dans notre premier cas de figure, le test est hautement significatif ; dans le deuxième, il est significatif.*

**Le seuil de signification  $\alpha$  est fixé a priori, c'est de lui dont dépend la décision finale de rejeter ou non  $H_0$ . Il reste le même pour tous les échantillons testés.**

**Le degré de signification  $p$  est propre à un échantillon et nous permet d'accorder une plus ou moins grande confiance à la décision prise.**

**On conclut au rejet de  $H_0$  à partir des observations faites sur un échantillon si  $p < \alpha$ .**

#### **4. Comparaison de deux pourcentages - Test de l'écart-réduit**

En médecine, l'analyse de variables qualitatives, en particulier sous la forme d'un caractère à deux classes (malade ou non, guéri ou non, complication présente ou absente, ...) est très fréquente.

Exemples :

- *Le pourcentage de patients guéris (taux de guérison) après administration du traitement A est-il meilleur que celui obtenu après administration du traitement B ?*

- Le pourcentage de complications varie-t-il en fonction du mode d'intervention chirurgical pour une pathologie donnée ?

Il s'agit alors de comparer des pourcentages observés dans deux ou plusieurs échantillons mais aussi de comparer un pourcentage observé à une valeur de référence.

#### 4.1. Comparaison d'un pourcentage observé à un pourcentage théorique

Exemple :

La proportion considérée comme habituelle de nouveau-nés prématurés dans la population considérée est de 4%. Parmi les 170 femmes de plus de 35 ans ayant accouché dans une maternité donnée, on observe 16 naissances prématurées (soit 9,4%). Ce résultat est-il différent de sa valeur de référence ?

Le problème posé est le suivant : on observe un pourcentage  $p_0$  (9,4%) sur un échantillon et on veut savoir s'il diffère d'une valeur de référence comme dans la population notée  $P_{H_0}$  (4%).

Autrement dit, l'échantillon est-il représentatif de la population de référence dans laquelle la fréquence théorique  $P$  du caractère prématuré est égale à  $P_{H_0}$  ?

La valeur observée sur l'échantillon (9,4%) est-elle compatible avec les fluctuations d'échantillonnage autour de la valeur  $P_{H_0}$  (4%) ou bien la différence entre 9,4% et 4% est-elle trop élevée pour être due au hasard ?

Pour répondre à cette question, on a recours au **test de l'écart-réduit**.

**Première étape du test** : spécifier l'hypothèse nulle  $H_0$  et l'hypothèse alternative  $H_1$ .

- $H_0$ : L'échantillon est représentatif d'une population de référence dans laquelle le caractère étudié a une fréquence théorique  $P = P_{H_0}$ .
- $H_1$ : test bilatéral :  $P \neq P_{H_0}$  (dans l'échantillon, le taux de prématurés diffère de celui de la population).  
test unilatéral :  $P < P_{H_0}$  ou  $P > P_{H_0}$  (dans l'échantillon, la population de femmes de plus de 35 ans connue pour avoir un taux de prématurité supérieur ou inférieur à la population d'ensemble).

Les hypothèses portent toujours sur la valeur du pourcentage dans la population (valeur théorique) et non sur une valeur observée.

**Sous  $H_0$** , si  $n$  est assez grand ( $n \geq 30$ ), si  $nP_{H_0} \geq 5$  et  $n(1 - P_{H_0}) \geq 5$ , on peut assimiler la distribution de la variable aléatoire  $P_0$  (pourcentage observée sur un échantillon) à une loi

normale de moyenne  $P_{H_0}$  et d'écart-type  $\sqrt{\frac{P_{H_0}(1-P_{H_0})}{n}}$ .

On appelle **écart-réduit**  $\varepsilon$  la quantité  $\frac{|p_0 - P_{H_0}|}{\sqrt{\frac{P_{H_0}(1-P_{H_0})}{n}}}$ ,  $|p_0 - P_{H_0}|$  est l'écart effectivement observé

entre le pourcentage observé dans l'échantillon et la valeur théorique connue.

La distribution de la variable centrée réduite, appelée écart-réduit  $\varepsilon$ , suit une loi normale  $\mathcal{N}(0,1)$ .

Sous cette distribution, cette quantité ne dépasse 1,96 en valeur absolue que dans 5% des cas.

**Deuxième étape** : vérifier les conditions de validité du test ( $nP_{H_0} \geq 5$  et  $n(1 - P_{H_0}) \geq 5$ ).

**Troisième étape** : confronter les valeurs observées aux valeurs attendues sous  $H_0$  en calculant l'écart-réduit  $\varepsilon$ .

La lecture de la table de la loi normale permet de trouver que la probabilité que la valeur  $\varepsilon$  soit supérieure ou égale à la valeur absolue de l'écart-réduit observé.

$$P \left( \varepsilon \geq \frac{|p_0 - P_{H_0}|}{\sqrt{\frac{P_{H_0}(1 - P_{H_0})}{n}}} \right).$$

Si cette probabilité est « petite », on peut dire que la différence  $|p_0 - P_{H_0}|$  est un événement peu probable sous l'hypothèse nulle.

**On ne rejette pas  $H_0$  si le degré de signification** (probabilité lue dans la table, caractéristique d'un échantillon donné) **est supérieur au seuil de probabilité choisi (en général 5%) et on rejette  $H_0$  dans le cas contraire.**

Application sur l'exemple cité plus haut :

$$P_{H_0} = 0,04 \quad n = 170 \quad p_0 = 0,094$$

Hypothèses :

- $H_0$  : l'échantillon est représentatif d'une population dans laquelle le caractère étudié a une fréquence théorique  $P = 0,04$ .
- $H_1$  :  $P \neq 0,04$  (test bilatéral)

Conditions de validité :

$$nP_{H_0} = 170 \times 0,04 = 6,8 \text{ et } n(1 - P_{H_0}) = 170 \times 0,96 = 163,2 \text{ sont supérieurs ou égaux à } 5.$$

On peut considérer que la distribution de la variable centrée réduite  $\varepsilon$  suit une loi normale  $\mathcal{N}(0,1)$ .

On calcule l'écart-réduit observé :  $\varepsilon_0 = \frac{|0,094 - 0,04|}{\sqrt{\frac{0,04(1 - 0,04)}{170}}} \quad \varepsilon_0 \cong 3,59.$

On recherche dans la table la probabilité suivante :  $P(\varepsilon \geq 3,59) \cong 10^{-3}.$

On conclut :

Le degré de signification ( $p = 10^{-3}$ ) étant inférieur à 0,05 ou la valeur de l'écart-réduit observé (3,59) étant supérieure à 1,96, valeur lue pour  $\alpha = 0,05$  (test bilatéral), on rejette l'hypothèse nulle au seuil de 5%. La différence est hautement significative. Le risque d'observer un écart aussi grand sous l'hypothèse nulle est inférieur à 0,001.

On peut conclure que le pourcentage de prématurés chez les femmes de plus de 35 ans dans la maternité concernée est différent de 4% avec  $p = 0,001$ .

**Remarque** : dans le cas d'un **test unilatéral** :

Dans l'exemple de l'essai thérapeutique du médicament  $A$  (médicament dont on cherche à montrer l'efficacité) contre le médicament  $B$  (placebo),  $A$  ne peut pas être inférieure en efficacité à  $B$  qui ne contient pas de substance active et l'hypothèse alternative n'a pas à tenir compte de

l'éventualité  $P_A < P_B$  ; on testera donc  $H_0 : P_A = P_B$  (les deux traitements ont la même efficacité) contre  $H_1 : P_A > P_B$  (le traitement A est plus efficace que le traitement B).

Dans l'exemple de la taille des adolescents, la formulation unilatérale conduit à exprimer l'hypothèse alternative  $H_1 : \mu > \mu_{H_0}$  ou  $H_1 : \mu < \mu_{H_0}$ .

Si la moyenne observée dans l'échantillon vaut  $m_0 = 171,60$ , nous avons déjà calculé  $\varepsilon_0 = 2,16$ . Seule va changer la règle de décision.

Dans la cas bilatéral, la règle de décision était : si  $|\varepsilon| \geq u_\alpha$ , on rejette  $H_0$  avec un risque  $\alpha$ .

Dans le cas unilatéral, la règle de décision devient : si  $|\varepsilon| \geq u_\alpha$ , on rejette  $H_0$  avec un risque  $\frac{\alpha}{2}$  ou si  $|\varepsilon| \geq u_{2\alpha}$ , on rejette  $H_0$  avec un risque  $\alpha$ .

Pour le même risque  $\alpha$ , c'est la valeur seuil qui change.

#### 4.2. Comparaison de deux pourcentages observés

Le cas le plus fréquent est celui où l'on dispose de deux échantillons d'effectifs  $n_1$  et  $n_2$ .

Dans le premier échantillon, un caractère a un pourcentage observé  $p_1$  et dans le second, le même caractère a un pourcentage observé  $p_2$ , différent de  $p_1$ .

La question consiste à savoir si la différence observée entre  $p_1$  et  $p_2$  est suffisamment petite pour que l'on puisse admettre que ces deux échantillons sont extraits d'une même population et que cette différence n'est due qu'à des fluctuations d'échantillonnage.

**Première étape** : spécifier l'hypothèse nulle  $H_0$  et l'hypothèse alternative  $H_1$ .

- **$H_0$** : les deux échantillons sont tirés au hasard d'une même population où la fréquence théorique du caractère est  $P$  ( $P_1 = P_2 = P$ ). Ici, la fréquence théorique est inconnue.
- **$H_1$**  : les deux échantillons ne sont pas tirés au hasard d'une même population (chacun a été tiré au hasard d'une population différente ou ils ont été tirés au sort d'une même population mais pas au hasard).

$P_1 \neq P_2$  test bilatéral

$P_1 > P_2$  ou  $P_1 < P_2$  test unilatéral

$P_1$  et  $P_2$  sont les pourcentages théoriques des populations d'où on extrait les échantillons.

**Deuxième étape** : estimer  $P$  et vérifier les conditions de validité.

Sous l'hypothèse  $H_0$ , la meilleure estimation de la fréquence théorique  $P$  est :  $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

(moyenne pondérée de  $p_1$  et  $p_2$ ).

Les conditions de validité sont :  $n_1 P \geq 5$  et  $n_1 (1 - P) \geq 5$ ,  $n_2 P \geq 5$  et  $n_2 (1 - P) \geq 5$ .

**Troisième étape** : sous  $H_0$  :

$P_1$  suit une loi normale de moyenne  $P$  et d'écart-type  $\sqrt{\frac{P(1-P)}{n_1}}$

$P_2$  suit une loi normale de moyenne  $P$  et d'écart-type  $\sqrt{\frac{P(1-P)}{n_2}}$

$P_1 - P_2$  suit une loi normale de moyenne 0 et d'écart-type  $\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

On calcule l'écart réduit observé :  $\varepsilon_0 = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

La lecture de la table permet de trouver la probabilité que, si  $H_0$  est vraie,  $\varepsilon$  soit supérieur ou égal à la valeur absolue de l'écart-réduit observé :  $P(\varepsilon) \geq \left| \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$ .

Si cette probabilité est inférieure à 5%, on rejette  $H_0$  au seuil de 5% et on accepte  $H_1$ . On dit que la différence  $p_1 - p_2$  est significative au seuil de 5%. Pour quantifier la force de la décision, on peut alors donner le degré de signification  $p$ .

Exemple : Dans un groupe de 200 malades, on a constitué par tirage au sort un échantillon soumis à un nouveau traitement A et un échantillon soumis au traitement classique B.

On a :

Traitement A :  $n_A = 102$  et 20 échecs soit  $p_A = 0,196$ .

Traitement B :  $n_B = 98$  et 29 échecs soit  $p_B = 0,296$ .

Les traitements A et B ont-ils le même taux d'échec ?

$H_0 : P_A = P_B = P$

L'hypothèse nulle est celle de l'égalité des pourcentages d'échecs quel que soit le traitement.  $P$  est le pourcentage d'échecs dans la population sous  $H_0$ .

$H_1 : P_A \neq P_B$  test bilatéral

On calcule  $P$  :  $P = \frac{102 \times 0,196 + 98 \times 0,296}{200} = 0,245$ .

On vérifie les **conditions de validité** :  $102 \times 0,245 = 24,99$  ;  $102 \times 0,755 = 77,01$  ;  $98 \times 0,245 = 24,01$  et  $98 \times 0,755 = 73,99$ . Tous ces résultats sont bien supérieurs à 5.

On calcule l'**écart-réduit observé** :  $\varepsilon_0 = \frac{|0,196 - 0,296|}{\sqrt{0,245(1-0,245)\left(\frac{1}{102} + \frac{1}{98}\right)}}$  soit  $\varepsilon_0 \cong 1,64$ .

La différence n'est pas significative bien qu'elle puisse paraître importante (environ 20% contre environ 30%). On ne rejette pas l'hypothèse nulle. On ne peut pas dire que les deux traitements donnent des résultats différents.

Si l'on avait observé les mêmes proportions avec des effectifs dix fois plus importants dans chaque groupe, l'écart-réduit serait égal à 5,19 et la différence serait alors significative. Dans ce cas, les deux groupes ayant été tirés au sort depuis la même population et ne différant que par le traitement, la différence entre les deux taux d'échec aurait été attribué au nouveau traitement.

Le test est d'autant plus **puissant** qu'il porte sur des effectifs importants.

Sources : [www.spc.univ-lyon1.fr](http://www.spc.univ-lyon1.fr)

[cybertim.timone.univ-mrs.fr](http://cybertim.timone.univ-mrs.fr)

[www.med.univ-montp1.fr](http://www.med.univ-montp1.fr)

[www.chups-jussieu.fr](http://www.chups-jussieu.fr)



## Exercices d'application

### Exercice 1 :

#### **Exemple de comparaison d'un pourcentage à une valeur de référence (test de la première loi de Mendel)**

Soit un couple  $[A, a]$  de gènes où  $A$  est le dominant. Le croisement d'hétérozygotes fournit des génotypes  $Aa, aA, aa$  et  $AA$ . Les génotypes  $Aa, aA$  et  $AA$  présentent le phénotype  $[A]$  tandis que le génotype  $aa$  présente le phénotype  $[a]$ .

Lorsque la première loi de Mendel s'applique, les proportions attendues de  $[A]$  et de  $[a]$  dans la descendance sont respectivement  $\frac{3}{4}$  et  $\frac{1}{4}$ .

Dans une descendance de 40 individus, on observe 27 individus  $[A]$ .

On veut tester si ces données permettent de rejeter la première loi de Mendel avec un risque de première espèce égal à 5%.

### Exercice 2 :

#### **Exemple de comparaison de deux proportions**

Deux médicaments  $A$  et  $B$  en concurrence ont été administrés à deux groupes distincts de 50 personnes atteintes de la même maladie. Les patients sont suivis de façon absolument identique. Après 15 jours, on remarque que 44 personnes parmi les 50 traitées par  $A$  ont guéri alors qu'il y en a 40 sur les 50 traitées par  $B$ .

On veut tester au risque de 5% si les deux médicaments sont significativement différents.

### Exercice 3 :

#### **Exemple de comparaison d'une proportion à une valeur de référence**

Sachant que 20% des personnes vaccinées par la formule  $F$  d'un vaccin présentent des allergies ou troubles secondaires, un laboratoire pharmaceutique propose une formule améliorée de ce vaccin  $F_a$  et espère diminuer le taux d'allergies et de troubles secondaires. Sur un échantillon de 400 personnes prises au hasard et ayant opté pour la formule  $F_a$ , on observe 60 cas d'allergies ou troubles secondaires.

On veut tester au risque  $\alpha = 0,05$  puis au risque  $\alpha = 0,01$  si la formule améliorée apporte réellement quelque chose de positif.

## Éléments de correction

### Exercice 1 :

$$H_0 : P = \frac{3}{4} \text{ et } H_1 : P \neq \frac{3}{4}$$

$$\varepsilon_0 = \frac{\left| \frac{27}{40} - \frac{3}{4} \right|}{\sqrt{\frac{3}{4} \times \frac{1}{4} \times \frac{1}{40}}} = 1,095$$

Dans la table de  $\mathcal{N}(0, 1)$ , on lit pour  $\alpha = 0,05$  : seuil = 1,96

$\varepsilon_0 < 1,96$  donc on ne rejette pas  $H_0$ .

**Conclusion : on ne rejette pas la première loi de Mendel au risque 5%.**

### Exercice 2 :

$$H_0 : P_1 = P_2 \text{ et } H_1 : P_1 \neq P_2$$

$$P_1 = \frac{44}{50} = 0,88 \text{ et } P_2 = \frac{40}{50} = 0,80$$

$$P = \frac{50 \times 0,88 + 50 \times 0,80}{100} = 0,84$$

$$\varepsilon_0 = \frac{|0,88 - 0,80|}{\sqrt{0,84 \times (1 - 0,84) \times \left(\frac{1}{50} + \frac{1}{50}\right)}}$$

Soit  $\varepsilon_0 \cong 1,09$

Dans la table de  $\mathcal{N}(0, 1)$ , on lit pour  $\alpha = 0,05$  : seuil = 1,96

$\varepsilon_0 < 1,96$  donc on ne rejette pas  $H_0$  au risque 5%.

**Conclusion : les deux médicaments ne sont pas significativement différents au risque 5%.**

### Exercice 3 :

$$H_0 : P = 0,20 \text{ et } H_1 : P < 0,20$$

$$\varepsilon_0 = \frac{\frac{60}{400} - 0,20}{\sqrt{\frac{0,20 \times (1 - 0,20)}{400}}} = -2,5$$

Pour  $\alpha = 0,05$ , on lit (et on calcule) dans la table : -1,645

Pour  $\alpha = 0,01$ , on lit (et on calcule) dans la table : -2,33

$\varepsilon_0$  est inférieur à ces deux valeurs donc dans les deux cas on rejette  $H_0$ .

**Conclusion : le taux d'allergie ou troubles secondaires est diminué par la formule  $F_a$  du vaccin aux risques de 5% et de 1%.**