# Exemples de tests

## Tests de validité d'hypothèse relatif à un paramètre

### Caractère qualitatif : test de validité d'hypothèse relatif à une fréquence

Dans une population, on étudie la fréquence $p$ d'un caractère qualitatif. Les hypothèses concernent $p$. La fréquence du caractère dans un échantillon de taille $n$ peut être considérée comme la réalisation d'une variable aléatoire $F$.

- On prend $H_0$ : "$p = p_0$" comme hypothèse nulle où $p_0$ est une valeur donnée.
- L'hypothèse alternative $H_1$ s'exprime sous l'une des formes suivantes : "$p \neq p_0$", "$p < p_0$" ou "$p > p_0$" (parfois "$p = p_1$" où $p_1$ est donnée). La première forme conduit à un test bilatéral, les autres à des tests unilatéraux.
- La variable de décision peut prendre une des formes suivantes :
  - ∗ $T = nF$ dont la loi est la loi binomiale $\mathcal{B}(n\,;p_0)$ sous l'hypothèse $H_0$.
  - ∗ $T = \dfrac{F - p_0}{\sqrt{\dfrac{p_0\,(1-p_0)}{n}}}$ dont on approche la loi, sous l'hypothèse $H_0$, par la loi normale $\mathcal{N}(0\,;1)$ lorsque

    $n \geq 30$, $n\,p_0 \geq 15$ et $n\,p_0\,(1-p_0) > 5$
- Sous l'hypothèse $H_0$, le fait que $T$ prenne des valeurs "très éloignées" de 0 est rare.

***Exemple 1***

Le fabricant d'un médicament breveté affirme qu'il est efficace à 90 % pour guérir une allergie en 8 heures.

**1°) -** Élaborer un test permettant de décider au niveau de confiance de 95 % si l'affirmation du fabricant est légitime.

**2°) -** Dans un échantillon de 200 personnes atteintes par cette allergie, on en a guéri 160 avec le médicament. Quelle conclusion peut-on en tirer ?

### Caractère quantitatif : test de validité d'hypothèse relatif à une moyenne

Le paramètre étudié est la moyenne $\mu$ d'un caractère quantitatif dans une population.

Les hypothèses concernent $\mu$ et on note $\sigma$ l'écart-type de la population.

La moyenne du caractère dans un échantillon de taille $n$ peut être considérée comme la réalisation d'une variable aléatoire $\overline{X}$.

- On prend $H_0$ : "$\mu = \mu_0$" comme hypothèse nulle où $\mu_0$ est une valeur donnée.
- L'hypothèse alternative $H_1$ s'exprime sous l'une des formes suivantes : "$\mu \neq \mu_0$", "$\mu < \mu_0$" ou "$\mu > \mu_0$" (parfois "$\mu = \mu_1$" où $\mu_1$ est donnée).
- La variable de décision $T$ prend différentes formes :
  - ∗ <u>Si σ est connu</u> : $T = \dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ et
    - sous l'hypothèse $H_0$, la loi de $T$ est loi normale $\mathcal{N}(0\,;1)$ si le caractère est distribué normalement dans la population ;
    - sous l'hypothèse $H_0$, la loi de $T$ est approchée par la loi normale $\mathcal{N}(0\,;1)$ si $n \geq 30$.

* <u>Si σ est inconnu</u> : $T = \dfrac{\overline{X} - \mu_0}{\widehat{S}/\sqrt{n}}$ où $\widehat{S}$ est la variable aléatoire qui, à chaque échantillon de taille $n$, associe son écart-type corrigé et
    - sous l'hypothèse $H_0$, la loi de $T$ est loi de Student à $n - 1$ degrés de liberté si le caractère est distribué normalement dans la population ;
    - sous l'hypothèse $H_0$, la loi de $T$ est approchée par la loi normale $\mathcal{N}(0 ; 1)$ si $n \geq 30$.

- Sous l'hypothèse $H_0$, le fait que $T$ prenne des valeurs "très éloignées" de 0 est rare.

***Exemple 2***

Un agriculteur plantait jusqu'à présent une variété A de blé pour laquelle il obtenait un rendement moyen de 59 quintaux à l'hectare. Un fournisseur lui propose une nouvelle variété B qu'il teste sur 6 parcelles.

**1°) -** En admettant que le rendement se distribue normalement, élaborer un test permettant de savoir au vu des rendements des 6 parcelles si le rendement de la nouvelle variété B est meilleur que celui de la variété A au seuil de confiance de 1 %.

**2°) -** L'agriculteur obtient sur les 6 parcelles un rendement moyen de 61 quintaux à l'hectare avec un écart-type de 3,8 quintaux à l'hectare. Doit-il adopter la variété B ?

***Exemple 3***

Une machine fabrique des pièces mécaniques en séries. Elle a été réglée pour que le diamètre de celles-ci soit égal à 21 mm. Naturellement, une certaine variabilité est inévitable.

**1° a) -** Élaborer un test permettant de décider, au seuil de confiance de 5 %, au vu d'un échantillon de 100 pièces si le réglage de la machine peut encore être considéré comme correct.

**b°) -** Sur un échantillon de 100 pièces, on a observé pour ce diamètre une valeur moyenne de 21,2 mm avec un écart-type de 0,6 mm.

**2°) -** Un autre échantillon donnant un diamètre moyen de 21,1 mm avec un écart-type de 0,54 mm a permis d'accepter le réglage de la machine au seuil de 5 %. Que peut-on en déduire pour la taille de l'échantillon ?

## Tests de comparaison de populations

### Caractère qualitatif : comparaison de deux proportions

Dans deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, on étudie un caractère qualitatif ayant pour fréquences respectives $p_1$ et $p_2$. On veut savoir, au vu de deux échantillons des deux populations, s'il existe une différence significative entre $p_1$ et $p_2$. La fréquence $f_i$ du caractère dans un échantillon de taille $n_i$ de la population $\mathcal{P}_i$ peut être considérée comme la réalisation d'une variable aléatoire $F_i$ pour $i \in \{1 ; 2\}$. $F_1$ et $F_2$ sont <u>indépendantes</u>.

- On prend comme hypothèse nulle $H_0$ : "$p_1 = p_2$".
- L'hypothèse alternative $H_1$ s'exprime sous l'une des formes suivantes : "$p_1 \neq p_2$" ou "$p_1 < p_2$" ou "$p_1 > p_2$".
- La variable de décision est $T = \dfrac{F_1 - F_2}{\sqrt{F(1-F)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ avec $F = \dfrac{n_1 F_1 + n_2 F_2}{n_1 + n_2}$ Sous l'hypothèse $H_0$, on approche la loi de $T$ par la loi normale $\mathcal{N}(0 ; 1)$ lorsque $n_1$ et $n_2$ sont supérieurs à 30, que $n_1 p_1$ et $n_2 p_1$ sont supérieurs 15 et que $n_1 p_1 (1 - p_1)$ et $n_2 p_1 (1 - p_1)$ sont supérieurs à 5 ; pour les calculs, on estime $p_1$ par

$\dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ où $f_1$ et $f_2$ sont les fréquences observées des caractères dans les échantillons de tailles respectives $n_1$ et $n_2$ dans respectivement les populations $\mathcal{P}_1$ et $\mathcal{P}_2$.

- Sous l'hypothèse $H_0$, le fait que $T$ prenne des valeurs "très éloignées" de 0 est rare.

***Exemple 4***

On veut savoir si les résultats d'une thérapeutique sont les mêmes suivant que le malade a ou non de la fièvre.

**1°) -** Élaborer un test permettant d'en décider, au seuil de confiance de 5 %, au vu d'échantillons de 200 malades fiévreux et de 100 malades non fiévreux.

**2°) -** Parmi 200 malades fiévreux, 72 % sont guéris par le traitement. Parmi 100 malades non fiévreux, 88 % sont guéris. Que peut-on en conclure ?

**Caractère quantitatif**

**Comparaison de deux variances**

Dans deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, on étudie un caractère quantitatif ayant pour variances respectives $\sigma_1^2$ et $\sigma_2^2$. On veut savoir, au vu d'un échantillon de chacune des populations, s'il y a une différence significative entre $\sigma_1^2$ et $\sigma_2^2$. L'écart-type corrigé du caractère dans un échantillon de taille $n_i$ de la population $\mathcal{P}_i$ peut être considérée comme la réalisation d'une variable aléatoire $\widehat{S_i}$ pour i $\in \{1 ; 2\}$. $\widehat{S_1}$ et $\widehat{S_2}$ sont <u>indépendantes</u>.

- On prend comme hypothèse nulle $H_0$ : "$\sigma_1 = \sigma_2$".
- L'hypothèse alternative $H_1$ s'exprime sous la forme : "$\sigma_1 \neq \sigma_2$" ou $\sigma_1 < \sigma_2$" ou "$\sigma_1 > \sigma_2$".
- La variable de décision est $T$ peut prendre plusieurs formes :
  - $T$ est l'une des deux variables aléatoires $\dfrac{\widehat{S_1}^2}{\widehat{S_2}^2}$ ou $\dfrac{\widehat{S_2}^2}{\widehat{S_1}^2}$. Sous l'hypothèse $H_0$, ces variables aléatoires suivent les lois F de Snédécor à respectivement $(n_1 - 1 ; n_2 - 1)$ et $(n_2 - 1 ; n_1 - 1)$ degrés de liberté si le caractère est distribué normalement dans les deux populations ;
  - $T = \dfrac{\widehat{S_1}^2 - \widehat{S_2}^2}{\sqrt{2}\sqrt{\dfrac{\widehat{S_1}^4}{n_1} + \dfrac{\widehat{S_2}^4}{n_2}}}$ dont la loi est approchée, sous l'hypothèse $H_0$, par la loi normale $\mathcal{N}(0 ; 1)$ lorsque $n_1$ et $n_2$ sont supérieurs ou égaux à 30.
- Sous l'hypothèse $H_0$, le fait que $T$ prenne des valeurs "très éloignées" de 1 ou de 0 selon le cas est rare.

***Exemple 5***

A la suite de la publication des résultats du baccalauréat dans deux académies A et B, on prélève

**1°) -** On admet la normalité de la distribution des résultats. Élaborer un test permettant de décider au vu des variances d'un échantillon de 25 résultats dans l'académie A et d'un échantillon de 28 élèves dans l'académie B si les résultats sont plus dispersés dans l'académie A que dans l'académie B au seuil de confiance de 5 %.

**2°) -** Les variances respectives des totaux de points des élèves des échantillons sont $\widehat{s_A}^2 = 13{,}6$ et $\widehat{s_B}^2 = 9{,}5$. Peut-on admettre que les résultats sont plus dispersés dans l'académie A que dans l'académie B ?

**Comparaison de deux moyennes**

Dans deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, on étudie un caractère quantitatif ayant pour moyennes respectives $\mu_1$ et $\mu_2$ et pour écarts-types respectifs $\sigma_1$ et $\sigma_2$. On veut savoir, au vu d'un échantillon de chacune des populations, s'il y a une différence significative entre $\mu_1$ et $q_2$.

La moyenne du caractère dans un échantillon de taille $n_i$ de la population $\mathcal{P}_i$ peut être considérée comme la réalisation d'une variable aléatoire $\overline{X_i}$ pour $i \in \{1 ; 2\}$. $\overline{X_1}$ et $\overline{X_2}$ sont <u>indépendantes</u>.

Pour $i \in \{1 ; 2\}$, $\widehat{S_i}$ est la variable aléatoire qui, à chaque échantillon de $\mathcal{P}_i$, associe son écart-type corrigé.

- On prend comme hypothèse nulle $H_0$ : "$\mu_1 = \mu_2$".
- L'hypothèse alternative $H_1$ s'exprime sous l'une des formes suivantes "$\mu_1 \neq \mu_2$", "$\mu_1 < \mu_2$" ou "$\mu_1 > \mu_2$".
- La variable de décision $T$ prend différentes formes :
  - ∗ <u>Si $\sigma_1$ et $\sigma_2$ sont connus</u> : $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ et
    - sous l'hypothèse $H_0$, la loi de $T$ est loi normale $\mathcal{N}(0 ; 1)$ si le caractère est distribué normalement dans les deux populations ;
    - sous l'hypothèse $H_0$, la loi de $T$ est approchée par la loi normale $\mathcal{N}(0 ; 1)$ si $n_1$ et $n_2$ sont supérieurs ou égaux à 30.
  - ∗ <u>Si $\sigma_1$ et $\sigma_2$ sont inconnus mais égaux</u> : $T = \dfrac{\overline{X_1} - \overline{X_2}}{\widehat{S}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ où $\widehat{S} = \sqrt{\dfrac{(n_1 - 1)\,\widehat{S_1}^2 + (n_2 - 1)\,\widehat{S_2}^2}{n_1 + n_2 - 2}}$ et
    - sous l'hypothèse $H_0$, la loi de $T$ est loi de Student à $n_1 + n_2 - 2$ degrés de liberté si le caractère est distribué normalement dans les deux populations ;
    - sous l'hypothèse $H_0$, la loi de $T$ est approchée par la loi normale $\mathcal{N}(0 ; 1)$ si le caractère est distribué normalement dans les deux populations et si $n_1 + n_2 - 2 \geq 30$ ;
    - sous l'hypothèse $H_0$, la loi de $T$ est approchée par la loi normale $\mathcal{N}(0 ; 1)$ si $n_1$ et $n_2$ sont supérieurs ou égaux à 30.
  - ∗ <u>Si $\sigma_1$ et $\sigma_2$ sont inconnus mais différents</u> : $T = \dfrac{\overline{X_1} - \overline{X_1}}{\sqrt{\dfrac{\widehat{S_1}^2}{n_1} + \dfrac{\widehat{S_2}^2}{n_2}}}$ dont la loi est approchée, sous l'hypothèse $H_0$, par la loi normale $\mathcal{N}(0 ; 1)$ si $n_1$ et $n_2$ sont supérieurs ou égaux à 30.
- Sous l'hypothèse $H_0$, le fait que $T$ prenne des valeurs "très éloignées" de 0 est rare.

***Exemple 6***

Une étude portant sur 1 000 cadres supérieurs (600 hommes et 400 femmes) révèle que la moyenne des salaires masculins est de 3 800 €, tandis que la moyenne des salaires féminins est de 3 500 € avec des écarts-types respectifs de 610 € et 760 €.

Au seuil de confiance de 5 %, la différence des salaires moyens des deux groupes est-elle imputable au seul hasard d'échantillonnage ?