

# Chapitre 6

## Tests statistiques paramétriques usuels

### Chap 6.

1. Introduction
2. Tests paramétriques usuels à partir d'un échantillon
3. Notion de puissance de test
4. Tests de comparaison

## 1. Introduction

On est souvent amenés à prendre des décisions sur la base d'un ensemble d'observations. On fera un choix entre deux propositions contradictoires: les **hypothèses  $H_0$  et  $H_1$  (nulle et alternative respectivement)**, qui dépendent des paramètres théorique de la population.

Pour trancher, nous devons nous baser sur les obs. disponibles, i.e. les échantillons, avec 2 types d'erreurs possibles:

- **Risque de 1ere espèce  $\alpha = P(H_0 \text{ rejetée} / H_0 \text{ vraie})$**   
Fixé à une faible valeur (5% ou 1% généralement)
- **Risque de 2e espèce  $\beta = P(H_0 \text{ non rejetée} / H_1 \text{ vraie})$**
- **$1 - \beta = \text{puissance du test} = P(H_1 \text{ acceptée} / H_1 \text{ vraie})$**

*Si  $H_0$  rejetée, on dit que le test est significatif, le risque de cette décision,  $\alpha$ , étant connu. Risque  $\beta$  souvent inconnu: on ne peut alors pas dire qu'on accepte  $H_0$  (test non significatif).*

Décision prise	Vérité	
	$H_0$	$H_1$
$H_0$	$1 - \alpha$	$\beta$
$H_1$	$\alpha$	$1 - \beta$

**Variable de décision:** variable aléatoire dont on connaît la loi de probabilité; au moins si  $H_0$  est vraie.

**Région critique:** ensemble des valeurs de la variable aléatoire de décision qui permettent d'écarter  $H_0$  au profit de  $H_1$ .

### Déroulement d'un test:

1. **Choix de  $H_0$  et  $H_1$**
2. **Choix de la variable de décision (statistique du test)**  
(choisie de façon à apporter le max d'info sur le pb posé et sa loi de probabilité sera différente selon que l'hypothèse nulle ou alternative est vraie.)
3. **Choix de  $\alpha$  petit (typiquement 1% ou 5%)**
4. **Calcul de la région critique en fonction de  $\alpha$  et de la statistique du test**
5. **Calcul de la valeur observée de la variable de décision**
6. **Comparaison et conclusion: Rejet de l'hypothèse si valeur calculée  $\in$  région critique**

**Test paramétrique:** son objet est de tester une **hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire** (e.g. moyenne, variance) suivant une loi spécifiée ou pas.

Test simple:  $H: \theta = \theta_0$ , où  $\theta_0$  est une valeur isolée de  $\theta$ .

Test composite:  $H: \theta \neq \theta_0, \theta > \theta_0$  ou  $\theta < \theta_0$

Test bilatérale: région critique est séparée en 2 régions distinctes (on ne se soucie pas du signe);

Test unilatéral: région critique ne correspond qu'à une seule région connexe de l'espace des valeurs de la variable (on se soucie du signe).

## 2. Tests paramétriques usuels à partir d'un échantillon

### Chap 6.

1. Introduction
2. Tests paramétriques usuels à partir d'un échantillon
3. Notion de puissance de test
4. Tests de comparaison

### 2.1) Comparaison d'une moyenne à une moyenne théorique donnée

On considère une variable aléatoire  $X \sim N(\mu, \sigma)$  avec  $\mu$  inconnue.

Note: si  $X$  n'est pas Gaussienne mais que l'échantillon est suff. grand ( $n > 30$ ); le théorème central limite (TCL) s'applique

=> méthodes indiquées pour une variable Gaussienne sont applicables.

$$\frac{(\bar{X} - \mu)}{\sigma / \sqrt{n}} \xrightarrow{n \rightarrow \infty} N(0,1) \quad \text{ou} \quad \bar{X} \xrightarrow{n \rightarrow \infty} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Test unilatéral:  $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

$\sigma$  connu,  $\mu$  estimé par la moyenne empirique et:  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

La statistique du test est donc:  $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$  et  $Z \sim N(0,1)$  si  $H_0$  vraie (cad si  $\mu = \mu_0$ ).

Région critique:  $\alpha = P(\bar{X} \geq x_{\text{lim}} / H_0 \text{ vraie}) = P(\bar{X} \geq x_{\text{lim}} / \mu = \mu_0) = P\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \geq \frac{x_{\text{lim}} - \mu}{\sigma / \sqrt{n}} / \mu = \mu_0\right)$

$$= P\left(Z \geq \frac{x_{\text{lim}} - \mu_0}{\sigma / \sqrt{n}}\right) \longrightarrow \text{Quantile } e_{1-\alpha} \text{ t.q.} \quad \left| \begin{array}{l} P(Z \geq Z_{\text{lim}}) = \alpha \text{ (} H_0 \text{ rejetée)} \\ P(Z \leq e_{1-\alpha}) = 1 - \alpha \text{ (} H_0 \text{ non rejetée)} \end{array} \right.$$

On peut ici utiliser la table de la loi normale  $N(0,1)$   
Et  $H_0$  est rejetée si

$$\bar{X} \geq x_{\text{lim}} \quad \text{soit} \quad \bar{X} \geq \mu_0 + e_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

Test unilatéral:  $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

$\sigma$  inconnu,  $\mu$  estimé par la moyenne empirique et:

La statistique du test est alors:  $T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$  avec  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$   
(démonstration: cf chap. 4)

$H_0$  vraie, variable  $T$  suit une loi de Student à  $n-1$  degrés de liberté.

On suit le même raisonnement que précédemment et  $H_0$  est rejetée si

$$\bar{X} \geq x_{\text{lim}} \quad \text{soit} \quad \bar{X} \geq \mu_0 + t_{n-1, 1-\alpha} \frac{S}{\sqrt{n}}$$

Remarque: même raisonnement pour tester le test unilatéral:  $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

$H_0$  est rejetée si

$$\bar{X} \leq x_{\text{lim}} \quad \text{soit} \quad \bar{X} \leq \mu_0 - e_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad \text{Dans le cas d'une variance connue}$$

$$\bar{X} \leq x_{\text{lim}} \quad \text{soit} \quad \bar{X} \leq \mu_0 - t_{n-1, 1-\alpha} \frac{S}{\sqrt{n}} \quad \text{Dans le cas d'une variance inconnue}$$

Note: si  $n > 30$ , grand échantillon  $\Rightarrow$  variable de décision =  $Z$  avec  $\sigma \approx S$  (estimation)

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

$\sigma$  connu,  $\mu$  estimé par la moyenne empirique et:  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

La statistique du test est donc:  $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$  et  $Z \sim N(0,1)$  si  $H_0$  vraie.

Région critique:  $\alpha = P(\bar{X} \geq x_{\text{sup}} \text{ ou } \bar{X} \leq x_{\text{inf}} / H_0 \text{ vraie}) = P(\bar{X} \geq x_{\text{sup}} \text{ ou } \bar{X} \leq x_{\text{inf}} / \mu = \mu_0)$   
 $= P\left(Z \geq \frac{x_{\text{sup}} - \mu_0}{\sigma / \sqrt{n}} \text{ ou } Z \leq \frac{x_{\text{inf}} - \mu_0}{\sigma / \sqrt{n}}\right)$

ou

$$\alpha = P\left(|Z| \geq \frac{x_{\text{lim}} - \mu_0}{\sigma / \sqrt{n}}\right) \rightarrow \text{Quantile } e_{1-\alpha/2} \text{ t.q. } P(|Z| \leq e_{1-\alpha/2}) = 1 - \alpha$$

$$P(-u \leq Z \leq u) = 1 - \alpha$$

$$P(Z \leq u) - (1 - P(Z \leq u)) = 1 - \alpha$$

$$P(Z \leq u) = 1 - \frac{\alpha}{2}$$

Si on connaît  $\sigma$ ,  $H_0$  est donc rejetée si:  $|Z| \geq e_{1-\frac{\alpha}{2}}$

ou si

$$\bar{X} \geq \mu_0 + e_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{ou} \quad \bar{X} \leq \mu_0 - e_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Test bilatéral:  $H_0: \mu = \mu_0$   
 $H_1: \mu \neq \mu_0$

$\sigma$  inconnu,  $\mu$  estimé par la moyenne empirique et:

La statistique du test est donc:  $T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$  avec  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

T suit une loi de Student à n-1 degrés de liberté.

Région critique:

$$\alpha = P\left(|Z| \geq t_{n-1, 1-\frac{\alpha}{2}}\right) \Rightarrow x_{\text{lim}} = \mu_0 \pm t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

Si on ne connaît pas  $\sigma$ ,  $H_0$  est donc rejetée avec un risque  $\alpha$  si:

$$|Z| \geq t_{n-1, 1-\frac{\alpha}{2}}$$

ou si

$$\bar{X} \geq \mu_0 + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \quad \text{ou} \quad \bar{X} \leq \mu_0 - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

## 2.2) Comparaison d'une proportion observée à une proportion théorique donnée

Nous considérons ici une **variable aléatoire X** prenant des valeurs binaires (0 ou 1 par exemple) et **suit une loi de Bernoulli de paramètre  $\pi$**  (fréquence empirique).

On observe un échantillon de  $n$  variables aléatoires indépendantes.

**Si  $n\pi > 5$ , alors  $X \sim N(\pi, \pi(1-\pi))$ .**

**Test bilatéral:**  $\begin{cases} H_0: \pi = \pi_0 \\ H_1: \pi \neq \pi_0 \end{cases}$

La statistique du test est alors:  $Z = \frac{\bar{X} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$  avec  $\bar{X} = p_0$  fréquence observée

**$H_0$  vraie si  $Z$  suit une loi normale  $N(0,1)$  et on rejette  $H_0$  si:**

$$|Z| \geq e_{1-\frac{\alpha}{2}}$$

**Test unilatéral:**  $\begin{cases} H_0: \pi = \pi_0 \\ H_1: \pi > \pi_0 \text{ (resp. } \pi < \pi_0) \end{cases}$

**On rejette  $H_0$  si:  $Z \geq e_{1-\alpha}$  (resp.  $Z \leq e_{1-\alpha}$ )**

## 2.3) Comparaison d'une variance à une variance théorique

Nous considérons ici des **variables aléatoires X** suivant une loi normale  $N(\mu, \sigma^2)$ .

**Moyenne connue (peu fréquent):**  $\begin{cases} H_0: \sigma = \sigma_0 \\ H_1: \sigma > \sigma_0 \end{cases}$

**La statistique du test est alors:**  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

*Note:  $1/n$  et non  $1/(n-1)$  car on veut tester si notre valeur obs est égale à la variance théorique.*

**H0 vraie si**  $\frac{nS^2}{\sigma^2} \rightarrow \chi_n^2$

Note: 
$$\chi_n^2 = \sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} = \frac{n}{\sigma^2} \sum_{i=1}^n \frac{(X_i - \mu)^2}{n} = \frac{nS^2}{\sigma^2}$$

**Risque de 1ere espèce – région critique:**

$$\alpha = P\left(S^2 \geq k_{\text{lim}} / \sigma = \sigma_0\right) = P\left(\frac{nS^2}{\sigma_0^2} \geq \chi_{n,1-\alpha}^2\right)$$

**On rejette H0 si:**  $S^2 \geq \frac{\sigma_0^2}{n} \chi_{n,1-\alpha}^2$



Moyenne inconnue (plus fréquent):  $\begin{cases} H_0: \sigma = \sigma_0 \\ H_1: \sigma > \sigma_0 \end{cases}$

La statistique du test est alors:  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$

H0 vraie si  $\frac{(n-1)S^2}{\sigma^2} \rightarrow \chi_{n-1}^2$  *Note: d° de liberté en moins car on utilise les obs pour calculer  $S^2$ .*

Risque de 1ere espèce  
- région critique:  $\alpha = P\left(S^2 \geq k_{\text{lim}} / \sigma = \sigma_0\right) = P\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \chi_{n-1, 1-\alpha}^2\right)$

On rejette H0 si:  $S^2 \geq \frac{\sigma_0^2}{n-1} \chi_{n-1, 1-\alpha}^2$

Cas bilatéral:

Même méthode mais on considère la région critique sur les 2 bords de la courbe de Chi2.

$$S^2 \leq \frac{\sigma_0^2}{n-1} \chi_{n-1, \frac{\alpha}{2}}^2 \quad \text{et} \quad S^2 \geq \frac{\sigma_0^2}{n-1} \chi_{n-1, 1-\frac{\alpha}{2}}^2$$

### 3. Notion de puissance de test

#### Chap 6.

1. Introduction
2. Tests paramétriques usuels à partir d'un échantillon
3. Notion de puissance de test
4. Tests de comparaison

$$\text{Puissance du test} = P(H_1 \text{ acceptée} / H_1 \text{ vraie}) = 1 - \beta$$

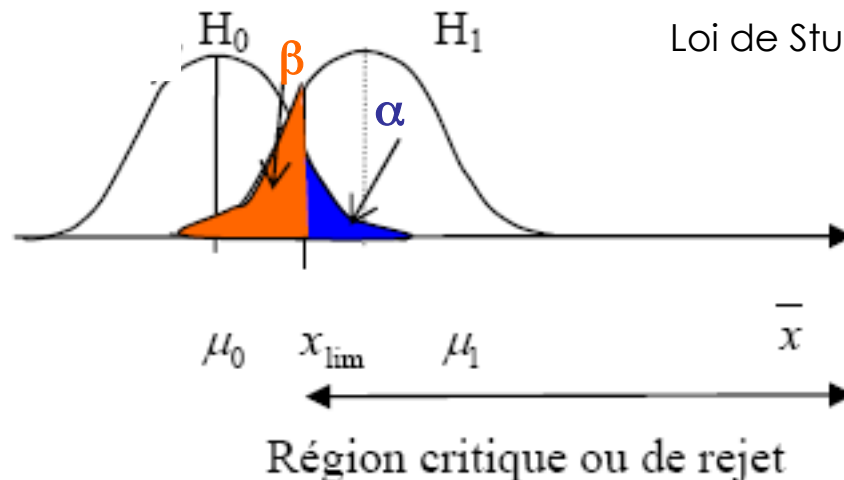
Si nous procédons à un test du type:  $H_0: \mu = \mu_0$  et  $H_1: \mu > \mu_0$ , c'est-à-dire avec  $H_1$  composite, le risque  $\beta$  ne pourra pas être calculé explicitement et donc on ne pourra pas calculer la puissance du test.

Seul cas pour lequel nous pouvons calculer  $\beta$  et donc la puissance de test: cas où  $H_1$  est fixé.

On testera par exemple la moyenne d'une variable Gaussienne avec:

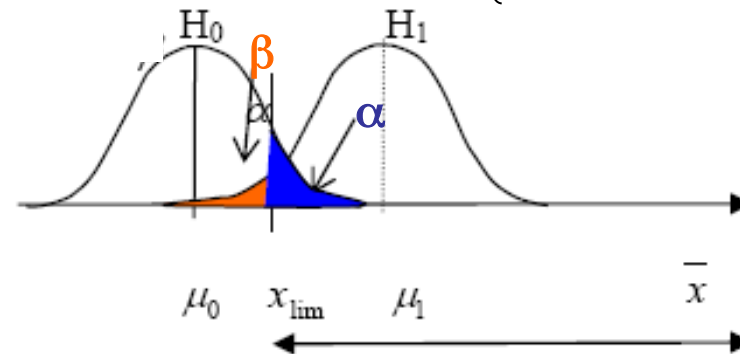
$$H_0: \mu = \mu_0 \text{ et } H_1: \mu = \mu_1.$$

$$\beta = P(H_0 \text{ refusée} / H_1 \text{ vraie}) = P(\bar{X} \geq x_{\text{lim}} / \mu = \mu_1) = P\left(T' \geq \frac{x_{\text{lim}} - \mu_1}{S / \sqrt{n}}\right)$$



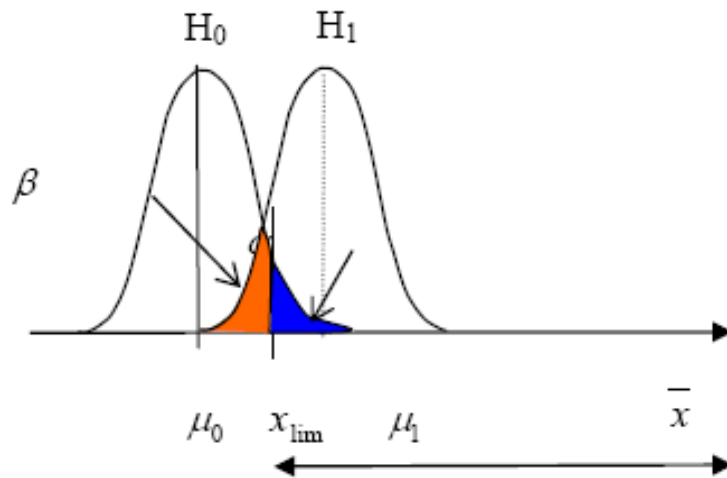
Loi de Student à  $n-1$  d° de liberté

↗ différence entre les moyennes de test  $\mu_0$  et  $\mu_1$ , même risque  $\alpha$



risque  $\beta$  ↘ (-) de risque de se tromper en ne rejetant pas  $H_0$ .  
puissance du test ↗

↗ risque  $\alpha$  de se tromper en rejetant  $H_0$  pour la même différence entre les moyennes de test



risque  $\beta$  ↘ (-) de risque de se tromper en ne rejetant pas  $H_0$ .  
puissance du test ↗

**Uniquement si**  
**taille de l'échantillon  $n$  ↗**

Propriété utile pour déterminer la taille minimale d'un échantillon pour détecter des différences entre des moyennes théoriques avec un faible risque d'erreur.

Région critique ou de rejet

## 4. Tests de comparaison d'échantillons

### Chap 6.

1. Introduction
2. Tests paramétriques usuels à partir d'un échantillon
3. Notion de puissance de test
4. Tests de comparaison

Jusqu'à présent, nous avons considéré la différence entre les propriétés d'un échantillon pour décider de la validité d'hypothèses portant sur des paramètres théoriques.

Les tests statistiques permettent aussi de répondre à des questions concernant les différences et similitudes entre différents échantillons.

Plus particulièrement, on cherchera ici à répondre à la question suivante:

*si nous considérons deux échantillons différents de taille respectivement  $n_x$  et  $n_y$ , prélevés indépendamment l'un de l'autre, peut-on affirmer qu'ils sont issus de la même population (hypothèse nulle) ou qu'ils proviennent chacun de deux populations distinctes ?*

## 4.1) Comparaison de 2 moyennes observées pour des grands échantillons

Considérons 2 variables aléatoires X et Y suivant des lois qcq, avec

$E(X)=\mu_X$ ,  $\text{var}(X)=\sigma_X^2$  et  $E(Y)=\mu_Y$  et  $\text{var}(Y)=\sigma_Y^2$

- 1er échantillon:  $n_X$  observations de X, avec  $n_X > 30$ .
- 2e échantillon:  $n_Y$  observations de Y, avec  $n_Y > 30$ .

**Variable de décision:**  $D = \bar{X} - \bar{Y}$

**Hypothèses**  $\begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$  **ou**  $\begin{cases} H_0 : \mu_X - \mu_Y = 0 \\ H_1 : \mu_X - \mu_Y \neq 0 \end{cases}$

**La statistique du test est donc:**  $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}$  **et si  $H_0$  est vrai alors:**  $Z \sim N(0,1)$

**Le risque de 1ere espèce étant fixé ( $\alpha=5\%$  ou  $1\%$ ), on peut déduire la région critique:**

$$\alpha = P\left(|Z| \geq e_{1-\frac{\alpha}{2}}\right)$$

*Note: c'est un test bilatéral si on ne s'intéresse pas au sens de la différence entre les moyennes.*

**On rejette  $H_0$  si:**  $|Z| \geq e_{1-\frac{\alpha}{2}}$  **ou**  $|\bar{X} - \bar{Y}| \geq e_{1-\frac{\alpha}{2}} \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}$

*Remarque: si on veut tester:  $H_0 : \mu_X - \mu_Y = 0$  ;  $H_1 : \mu_X - \mu_Y > 0$*

*On rejette  $H_0$  si:  $\bar{X} - \bar{Y} \geq e_{1-\alpha} \sqrt{S_X^2/n_X + S_Y^2/n_Y}$*

## 4.2) Comparaison de 2 moyennes observées pour des échantillons Gaussiens de taille qcq et de même variance

Considérons 2 variables aléatoires  $X$  et  $Y$  suivant des lois qcq, avec  $E(X)=\mu_X$ ,  $\text{var}(X)=\sigma^2$  et  $E(Y)=\mu_Y$  et  $\text{var}(Y)=\sigma^2$

- 1<sup>er</sup> échantillon:  $n_X$  observations de  $X$ , avec  $n_X$  qcq.
- 2<sup>e</sup> échantillon:  $n_Y$  observations de  $Y$ , avec  $n_Y$  qcq.

**Variable de décision:**  $D = \bar{X} - \bar{Y}$

**Hypothèses**  $\begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$  **ou**  $\begin{cases} H_0 : \mu_X - \mu_Y = 0 \\ H_1 : \mu_X - \mu_Y \neq 0 \end{cases}$

**La statistique du test est donc:**  $T = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$  **avec**  $S^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$

et si  $H_0$  est vrai alors  $T$  suit une loi de Student à  $n_X + n_Y - 2$  d° de liberté.

**Le risque de 1<sup>ere</sup> espèce étant fixé ( $\alpha=5\%$  ou  $1\%$ ), on peut déduire la région critique:**  $\alpha = P\left(|T| \geq t_{n_X + n_Y - 2; 1 - \frac{\alpha}{2}}\right)$  *Note: c'est un test bilatéral si on ne s'intéresse pas au sens de la différence entre les moyennes.*

**On rejette  $H_0$  si:**  $|T| \geq t_{n_X + n_Y - 2; 1 - \frac{\alpha}{2}}$  **ou**  $|\bar{X} - \bar{Y}| \geq t_{n_X + n_Y - 2; 1 - \frac{\alpha}{2}} S \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$

*Remarque: ce test n'est valable que si les variances théoriques sont égales, il est donc nécessaire de tester en 1<sup>er</sup> lieu cette hypothèse!*

### 4.3) Comparaison de 2 proportions observées

Considérons 2 variables aléatoires X et Y suivant des lois de Bernoulli de paramètres théoriques  $\pi_X$  et  $\pi_Y$ .

- 1er échantillon:  $n_X$  observations de X, avec une prop. obs.  $p_X$ ;
- 2<sup>e</sup> échantillon:  $n_Y$  observations de Y, avec une prop. obs.  $p_Y$ .

Bernoulli:  $E(X) = \pi_X$  ;  $\text{var}(X) = \pi_X(1 - \pi_X)$ , même chose pour Y.

**Variable de décision:**  $D = \bar{X} - \bar{Y}$

$$\text{Hypothèses} \quad \begin{cases} H_0 : \pi_X = \pi_Y \\ H_1 : \pi_X \neq \pi_Y \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \pi_X - \pi_Y = 0 \\ H_1 : \pi_X - \pi_Y \neq 0 \end{cases}$$

**La statistique du test est donc:** 
$$Z = \frac{p_X - p_Y}{\sqrt{p_0(1-p_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \quad \text{où} \quad p_0 = \frac{n_X p_X + n_Y p_Y}{n_X + n_Y}$$

et si  $H_0$  est vrai alors Z suit une loi normale  $N(0,1)$ .

**Le risque de 1<sup>ere</sup> espèce étant fixé ( $\alpha=5\%$  ou  $1\%$ ), on peut déduire la région critique:**

$$\alpha = P\left(|Z| \geq e_{1-\frac{\alpha}{2}}\right)$$

*Note: c'est un test bilatéral si on ne s'intéresse pas au sens de la différence entre les moyennes.*

**On rejette  $H_0$  si:** 
$$|p_X - p_Y| \geq e_{1-\frac{\alpha}{2}} \sqrt{p_0(1-p_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}$$

*Remarque: ce test n'est valable que si les proportions observées vérifient:*

$$n_X p_0 \geq 5 ; n_Y p_0 \geq 5 ; n_X(1-p_0) \geq 5 ; n_Y(1-p_0) \geq 5$$

#### 4.4) Comparaison de moyennes observées pour deux échantillons appariés.

On considère alors les mêmes individus dans **deux expériences différentes**.

e.g.: test médicaments sur les mêmes patients.

$$\text{Hypothèses } \begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu_X - \mu_Y = 0 \\ H_1 : \mu_X - \mu_Y \neq 0 \end{cases}$$

**Variable de décision:** les observations sont faites sur les mêmes individus et ne sont donc pas indépendantes. Dans ce cas on considère des paires d'observations et leur différence:

$$d_i = X_{1,i} - X_{2,i}$$

La statistique du test est alors:

$$T = \frac{\bar{d}}{S_d / \sqrt{n}} \quad \text{où} \quad \bar{d} = \frac{1}{n} \sum_{i=1}^n (x_{1,i} - x_{2,i}) \quad \text{et} \quad S_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$$

et si  $H_0$  est vrai alors  $T$  suit une loi de Student à  $n-1$  d° de liberté.

Le risque de 1ere espèce étant fixé ( $\alpha=5\%$  ou  $1\%$ ), on peut déduire la région critique:  $\alpha = P\left(|T| \geq t_{n-1; 1-\frac{\alpha}{2}}\right)$

*Note: c'est un test bilatéral si on ne s'intéresse pas au sens de la différence entre les moyennes.*

On rejette  $H_0$  si:

$$|T| \geq t_{n-1; 1-\frac{\alpha}{2}}$$



## 4.5) Comparaison des variances de deux échantillons

Considérons 2 variables aléatoires X et Y Gaussiennes:

- 1er échantillon:  $n_X$  observations de  $X \sim N(\mu_X, \sigma_X^2)$
- 2<sup>e</sup> échantillon:  $n_Y$  observations de  $Y \sim N(\mu_Y, \sigma_Y^2)$ .

$$\text{Hypothèses} \begin{cases} H_0 : \sigma_X = \sigma_Y \\ H_1 : \sigma_X \neq \sigma_Y \end{cases}$$

Variable de décision = rapport des deux variances  $K = \frac{S_X^2}{S_Y^2}$

$$\text{Or} \quad (n_X - 1) \frac{S_X^2}{\sigma_X^2} \sim \chi_{n_X-1}^2 \quad \text{et} \quad (n_Y - 1) \frac{S_Y^2}{\sigma_Y^2} \sim \chi_{n_Y-1}^2$$

La statistique du test est donc une loi de Fisher-Snedecor.

Si  $H_0$  est vrai alors K suit une loi de Fisher-Snedecor à  $n_X-1$  et  $n_Y-1$  d° liberté.

$$\text{On rejette } H_0 \text{ si: } \frac{S_X^2}{S_Y^2} \geq F_{n_X-1, n_Y-1; 1-\frac{\alpha}{2}} \quad \text{ou} \quad \frac{S_X^2}{S_Y^2} \leq F_{n_X-1, n_Y-1; \frac{\alpha}{2}} \quad \text{test bilatéral}$$

### Remarque

Une méthode approchée utilisée parfois pour éviter de faire un test bilatéral est d'imposer que le numérateur de la statistique soit la variance numériquement la plus grande, le test peut se réduire alors à un test unilatéral. Il faut alors faire attention à choisir la loi de Fisher correspondante, c'est à dire que le 1er degré de liberté sera celui du numérateur (variance empirique la plus forte) et le 2ème celui du dénominateur.