

Introduction aux tests statistiques

R. Paegelow - MF Beclin

Statistiques et Analyses de Données - D04

Décembre 2024

Tests statistiques

Tests pour une population

Tests pour deux populations

Tests usuels à connaître

Objectif : Prendre une décision relative à une question générale du type :

- *Nouveau médicament plus performant ?*
- *Espérances de vie différentes selon les groupes ?*
- *Loi de la v.a. normale ?...*

~> **questions où la réponse est *oui ou non* !**

Tests statistiques : méthodologie pour répondre à ce type de question sur la base d'observations liées au phénomène d'intérêt.

Test statistique : principes de base

1. Confrontation de deux hypothèses *exclusives* : l'hypothèse dite *hypothèse nulle* H_0 et l'hypothèse dite *hypothèse alternative* H_1 ;
2. H_0 est *favorisée* dans la mesure où l'approche va *supposer cette hypothèse comme étant vraie* ;
3. 0 l'issue d'un test statistique, la décision prise est
 - *garder* H_0 : le test est alors dit *non-significatif* ;
 - *rejeter* H_0 *au profit de* H_1 : le test est alors dit *significatif*.
4. pour chaque décision un type d'erreur à considérer :

Risque de 1ère espèce α : $\alpha = \mathbb{P}(\text{rejet } H_0 | H_0 \text{ vraie})$

Risque de seconde espèce β : $\beta = \mathbb{P}(\text{garder } H_0 | H_1 \text{ vraie})$

Deux grandes catégories de tests :

- **Paramétrique** : tests relatifs à des *paramètres de lois de probabilité sous certaines conditions*
 - tests sur les espérances ;
 - tests sur les variances ;
 - tests sur les proportions.
- **Non-paramétrique** : tests relatifs
 - à *des paramètres généraux*, non-spécifiques de lois particulières : médiane, quartiles ,... ;
 - à la *loi d'une v.a.* d'intérêt : test *d'ajustement à une loi* ;
 - à *l'indépendance entre v.a.* : test *d'indépendance*

Exemple 1.1 : Buveurs de soupe

On cherche à tester l'hypothèse "*manger de la soupe quotidiennement fait grandir*". On suppose que la taille de la population générale est de $\mu_0 = 170$ cm et que les individus sont répartis selon une loi normale $\mathcal{N}(170, 16)$. On suppose alors que la population des personnes mangeant de la soupe régulièrement ont une taille également répartie selon une loi Normale $\mathcal{N}(\mu, 16)$. On dispose d'un échantillon i.i.d $\sim \mathcal{N}(\mu, \sigma^2)$ d'individus buveurs de soupe. La question est alors : Est-ce que $\mu > 170$? Ce qui se traduit en test statistique :

$$H_0 : \mu = 170$$

$$H_1 : \mu \neq 170$$

Exemple 1.2 : Fumeurs ou Fumeuses ?

On se pose la question suivante : pour la population jeunes adultes, fumer est-il dépendant du sexe ?

On pose alors les hypothèses suivantes :

H_0 : les variables "sexe" et "fumeur" sont indépendantes

H_1 : les variables "sexe" et "fumeur" sont dépendantes

Démarche générale :

1. Formuler les hypothèses H_0 et H_1 ;
2. Supposer H_0 vérifiée et *caractériser la région attendue pour les résultats d'expérimentation "sous H_0 "* ;
3. Calculer sur la base des observations *le résultat d'expérimentation* et regarder s'il est *conforme ou en contradiction* par rapport à l'attendu sous H_0 ;
4. Conclure quant au *rejet ou non de H_0* en quantifiant la *confiance que l'on accorde à cette décision*.

Tout repose sur la quantification de la compatibilité du résultat issu des observations en regard de l'hypothèse H_0 faite.

La démarche d'un test est de type sciences expérimentales : regarder dans quelle mesure les résultats d'expérience permettent ou non de rejeter une hypothèse *faite a priori*.

Pour tous tests statistiques, H_0 est une hypothèse "favorisée" puisque supposée vérifiée. $\hookrightarrow H_0$ non rejetée au profit de H_1 ne signifie pas pour autant H_0 valide : une expérience plus précise permettrait peut-être de la rejeter !

Pour tout test statistique, par rapport à la question posée, seul le rejet de H_0 est informatif ! L'hypothèse à démontrer devrait donc en pratique correspondre à H_1 (mais hélas pas toujours possible !)

Remarque 1.1 :

Pour la justice, on est innocent tant que l'on "ne prouve pas" le contraire ! être coupable (rejet de H_0) suppose donc une "démonstration" de la culpabilité : *les faits permettent de contredire ce qui est supposé* (pour autant, tous ceux "déclarés innocents" le sont-ils vraiment ??)

En pratique, pour réaliser un test statistique, il faut être capable d'explicitier (de façon exacte ou approchée) la loi sous H_0 d'une v.a. "caractéristique" appelée **statistique de test**.

- Loi de l'écart entre des effectifs attendus sous H_0 et des effectifs observés.
- Loi de l'estimateur d'un paramètre sous H_0 .
- Loi du nombre de succès attendu sous H_0 .

On reprend l'exemple "Buveur de soupe"

Exemple 1.3 : Buveurs de Soupe

$$H_0 : \mu = 170$$

$$H_1 : \mu \neq 170$$

On peut estimer μ par l'estimateur de la moyenne empirique $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$. avec $\hat{\mu} \sim \mathcal{N}(\mu, \frac{16}{n})$. Ainsi sous l'hypothèse H_0 , $\frac{\hat{\mu} - \mu_0}{\frac{16}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$. La statistique de test est donc $\frac{\hat{\mu} - \mu_0}{\frac{16}{\sqrt{n}}}$.

Les types d'erreur associés à la prise de décision :

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

$$\alpha = P(\text{rejet } H_0 \mid H_0 \text{ vraie})$$

$$1 - \alpha = P(\text{garder } H_0 \mid H_0 \text{ vraie}).$$

$$\beta = P(\text{garder } H_0 \mid H_1 \text{ vraie})$$

$$1 - \beta = P(\text{rejeter } H_0 \mid H_1 \text{ vraie}).$$

Exemple 1.4 : Buveurs de Soupe

On a alors $\alpha = \mathbb{P}\left(\left|\frac{\hat{\mu}-\mu_0}{\frac{16}{\sqrt{n}}}\right| \geq T_{lim} | H_0 \text{ vraie} \right)$. D'après la table des quantile de la loi normal $T_{lim} = 1,96$ pour $\alpha = 0.05$. Si $\left|\frac{\hat{\mu}-\mu_0}{\frac{16}{\sqrt{n}}}\right| > 1,96$ on rejettera l'hypothèse H_0 , sinon on ne la rejette pas.

On peut également calculer la p-valeur : Soit t la valeur obtenue pour $\frac{\hat{\mu}-\mu_0}{\frac{16}{\sqrt{n}}}$, on calcule $\mathbb{P}\left(\left|\frac{\hat{\mu}-\mu_0}{\frac{16}{\sqrt{n}}}\right| > t | H_0 \text{ vraie} \right)$ et on compare cette probabilité qu'on appelle p-valeur avec la valeur de α choisi.

Quels risques de première et seconde espèces ?

- le risque de 1ère espèce α est fixé

↪ se contrôle très facilement !

Valeur $\leftrightarrow$ conséquences d'un rejet à tort de H_0 ... *valeur canonique : 5%.*

- risque β (ou puissance $1 - \beta$) plus difficile à contrôler car en relation directe avec H_1

↪ évalué uniquement pour une H_1 complètement spécifiée !

Quels risques de première et seconde espèces ?

Approche classique : on se **fixe α** .

Approche moderne : exploiter l'ordinateur !

α n'est pas fixé a priori et on utilise une notion de **probabilité critique (p-value)**.

$$\text{probabilité critique (p-value)} = \mathbb{P}(T \geq T_{\text{Observé}} | H_0 \text{ vraie})$$

Remarque :

si on **diminue α** , rejet de H_0 uniquement si “violemment contredite” par les résultats d'expérience ...

en contrepartie, résultats peu vraisemblables vis à vis de H_0 non pris en compte pour un rejet, ce qui augmente le risque de garder H_0 à tort ... i.e. **on augmente β et donc on diminue la puissance** !

H_0 : les variables *sexe* et *fumeur* sont indépendantes ?

Pour cela, on réalise une étude et on obtient les résultats suivants :

Sexe	Fumeur ?		n_{i+}
	Fu	Nfu	
H	70	80	150
F	50	160	210
n_{+j}	120	240	360

$$\hat{P}(H \text{ et } Fu) = 70/360; \hat{P}(F \text{ et } Fu) = 50/360; \hat{P}(H) = 150/360; \\ \hat{P}(Fu) = 120/360 \dots$$

En d'autres termes :

$$\hat{P}(\text{modalité } i \text{ Sexe et modalité } j \text{ Fumeur}) = \frac{n_{ij}}{n} \\ \hat{P}(\text{modalité } i \text{ Sexe}) = \frac{n_{i+}}{n}; \hat{P}(\text{modalité } j \text{ Fumeur}) = \frac{n_{+j}}{n}$$

Un exemple plus simple

Si les variables *Sexe* et *Fumeur* sont **indépendantes**, on sait que

$$P(H \text{ et } Fu) = P(H) \times P(Fu)$$

et on s'attend à

$$\hat{P}(H \text{ et } Fu) \approx \hat{P}(H) \times \hat{P}(Fu)$$

c'est-à-dire :

$$\frac{n_{ij}}{n} \approx \frac{n_{i+}}{n} \times \frac{n_{+j}}{n} = \frac{\frac{n_{i+}n_{+j}}{n}}{n}$$

d'où

$$\boxed{n_{ij} \approx \frac{n_{i+}n_{+j}}{n}} \text{ dans le cas indépendant.}$$

Un exemple plus simple

Plus intuitivement, on a $\hat{P}(\text{Fu})=120/360=33\%$ donc on s'attend sous indépendance à avoir la même répartition chez les hommes et chez les femmes.

Comme on a 150 hommes on s'attend à $150 * 0.33 = 50$ hommes fumeurs et 100 non-fumeurs.

Comme on a 210 femmes on s'attend à $210 * 0.33 = 70$ femmes fumeuses et 140 non-fumeuses.

Autrement dit on s'attend à observer les effectifs du tableau suivant aussi appelés effectifs théoriques.

Sexe	Fumeur ?		n_{i+}
	Fu	Nfu	
H	50	100	150
F	70	140	210
n_{+j}	120	240	360

Ecart à l'indépendance

	y_1	y_2	$\dots$	y_j	$\dots$	y_J	
x_1	$\frac{n_{1+}n_{+1}}{n}$			$\dots$		$\frac{n_{1+}n_{+J}}{n}$	n_{1+}
x_2							$\vdots$
$\vdots$							$\vdots$
x_i	$\vdots$	$\dots$		$\frac{n_{i+}n_{+j}}{n}$		$\vdots$	n_{i+}
$\vdots$							$\vdots$
x_I	$\frac{n_{I+}n_{+1}}{n}$			$\dots$		$\frac{n_{I+}n_{+J}}{n}$	n_{I+}
	n_{+1}	n_{+2}	$\dots$	n_{+j}	$\dots$	n_{+J}	n

Table 1: Tableau théorique sous l'hypothèse d'indépendance

Le χ^2 de Pearson est un indicateur, une mesure qui nous permettra de quantifier l'écart entre le tableau de données et le tableau théorique sous l'hypothèse d'indépendance.

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2}{\frac{n_{i+}n_{+j}}{n}}$$

Cette quantité est comprise entre 0 et n . Peut-on s'attendre à une forte ou à une faible valeur du χ^2 dans le cas de l'indépendance ?

Loi du χ^2 et loi de Student

Définition 1.1 : Loi du χ^2

Soit $X_1, \dots, X_n$ des variables gaussiennes $\mathcal{N}(0, 1)$ alors la variable $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$

Définition 1.2 : Loi de Student

Soit X une variable gaussienne $\mathcal{N}(0, 1)$ et $Y \sim \chi^2(n)$ alors la variable $\frac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de liberté $\mathcal{T}(n)$

On retrouvera ces lois avec les estimateurs de la régression linéaire.

Un exemple

Sous H_0 : les variables sont indépendantes, on s'attend à un *faible écart entre les effectifs "théoriques" $\frac{n_{i+}n_{+j}}{n}$ et les effectifs observés n_{ij}*

$\leadsto$ **quantifier à partir de quelle valeur cette quantité χ^2 doit être considérée comme trop grande !!**

Si cette quantité χ^2 est "trop grande", on rejettera H_0 ; si au contraire cette quantité χ^2 ne va pas au-delà des fluctuations échantillonnales attendues sous H_0 , on gardera H_0 .

Outil probabiliste : il est possible de montrer que

$$\chi^2 = \sum \frac{(Eff. Obs. - Eff. Theo.)^2}{Eff. Theo.} \sim \chi^2((I-1)(J-1))$$

Ici, $(I-1)(J-1) = 1$;

d'après table $\chi^2(1)$, $P(\chi^2(1) > 3.84) = 5\%$, donc

sous H_0 on n'a que 5% de chances d'observer un écart ≥ 3.84

Un exemple

Après calculs, $\chi^2 = 20.57 > 3.84$, intervalle de probabilité 95% sous H_0
donc

on rejette $H_0 \dots$

Application R :

```
> chisq.test(matable,correct=F)
```

Pearson's Chi-squared test

```
data: es
```

```
X-squared = 20.5714, df = 1, p-value = 5.745e-06
```

```
>
```

```
> chisq.test(matable,correct=F)$expected
```

```
    [,1] [,2]
```

```
[1,]   50  100
```

```
[2,]   70  140
```

Outline

Tests statistiques

Tests pour une population

Tests pour deux populations

Tests usuels à connaître

Tests paramétriques pour une population

Tests pour une espérance

Soit $(X_1, \dots, X_n)$ n -échantillon de v.a. d'espérance μ (inconnue). Soit μ_0 une valeur fixée.

Tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$

Ici on rejetera H_0 si la différence $\mu - \mu_0$ est “trop petite” ou “trop grande” : le test est dit *bilatéral*.

μ est inconnue ; estimateur : $\hat{\mu} \equiv \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

$\rightsquigarrow$ **caractériser les fluctuations échantillonnales de $\hat{\mu} - \mu_0$ sous H_0**

C'est la cas "Buveur de Soupe"

Tests de comparaison pour une espérance

- Cas 1 : On suppose $X_i \sim N(\mu; \sigma^2)$, σ^2 connue.

Loi de l'estimateur sous H_0 : $\hat{\mu} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$, d'où

$$P\left(-z_{1-\alpha/2} \leq \frac{\hat{\mu} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2} | H_0\right) = 1 - \alpha$$

$\leadsto$ Au seuil de confiance $1 - \alpha$, garder H_0 si

$$\frac{\hat{\mu} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]$$

ou encore

Au risque α , rejet de H_0 si

$$\frac{\hat{\mu} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{1-\alpha/2} \text{ ou } \frac{\hat{\mu} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\alpha/2}$$

Tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$

- Cas 1bis : On suppose $X_i \sim N(\mu; \sigma^2)$, σ^2 **inconnue**.

On estime σ^2 par $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2$;

Sous H_0 , $\frac{\hat{\mu} - \mu_0}{\frac{S}{\sqrt{n}}} \sim T_{(n-1)}$

On en déduit :

Au seuil de confiance $1 - \alpha$, garder H_0 si

$$\frac{\hat{\mu} - \mu_0}{\frac{S}{\sqrt{n}}} \in [-t_{1-\alpha/2;(n-1)}, t_{1-\alpha/2;(n-1)}]$$

ou encore

Au risque α , rejet de H_0 si

$$\frac{\hat{\mu} - \mu_0}{\frac{S}{\sqrt{n}}} < -t_{1-\alpha/2;(n-1)} \text{ ou } \frac{\hat{\mu} - \mu_0}{\frac{S}{\sqrt{n}}} > t_{1-\alpha/2;(n-1)}$$

Tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$

- Cas 2 : On suppose X_i de loi quelconque.
 - si n est suffisamment grand, les résultats précédents (cas 1bis) restent valides (TCL). On aura par ailleurs $t_{\alpha;(n-1)} \equiv z_\alpha$;
 - si n petit (et absence de normalité) pas de résultats généraux
 - $\rightsquigarrow$ revenir aux lois exactes de l'estimateur sous H_0 .
 - $\rightsquigarrow$ tests non-paramétriques.

Illustration de la puissance

Test $H_0 : \mu=0$ contre $H_1 : \mu=2$

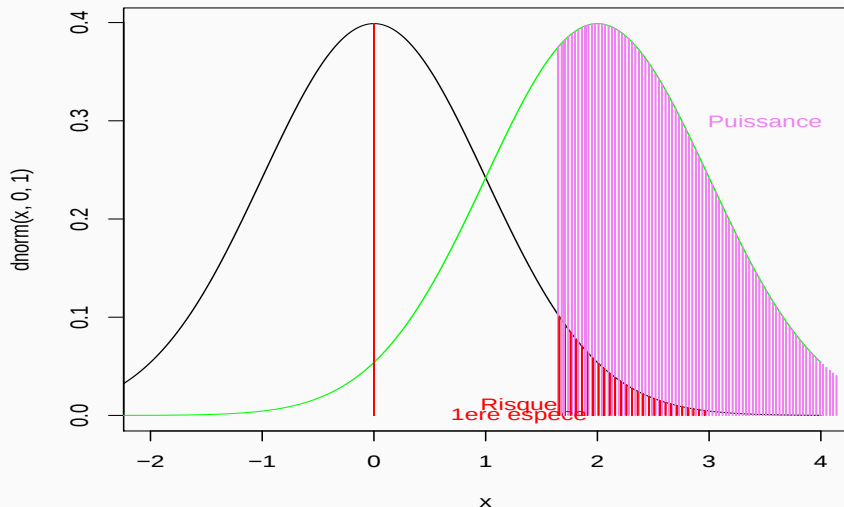
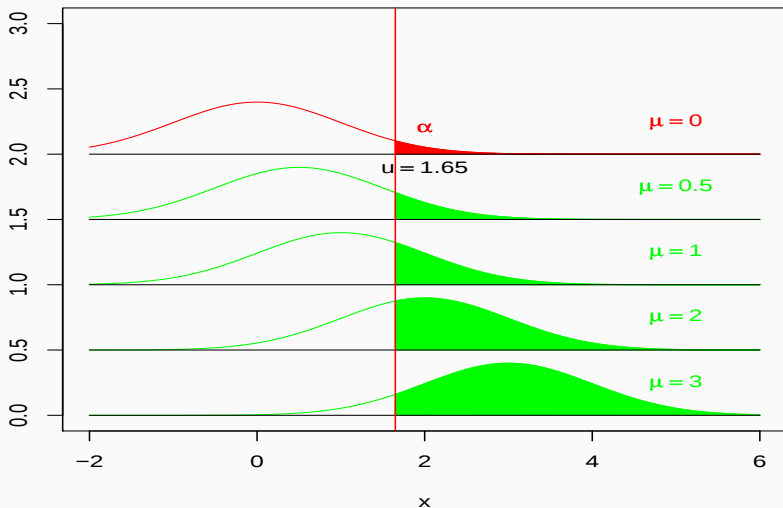


Illustration de la puissance

Puissance selon contre-hypothèses



Tests de comparaison pour une fréquence

Tests pour une fréquence

Tester $H_0 : p = p_0$ contre $H_1 : p \neq p_0$, p_0 fixé

- On suppose **grand échantillon** ($np_0 > 5$ et $n(1 - p_0) > 5$)

Estimateur de p : $p_n =$ fréquence échantillonnale.

$$\rightsquigarrow \text{ sous } H_0, \frac{p_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

Par le même raisonnement que précédemment, on en déduit

Rejet de H_0 au risque α si $\frac{p_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \notin [-z_{1-\alpha/2}, z_{1-\alpha/2}]$

- si **n petit ?** utiliser la loi exacte : on sait que sous H_0 $np_n \sim B(n, p_0)$.

Outline

Tests statistiques

Tests pour une population

Tests pour deux populations

Tests usuels à connaître

Comparaison de deux variances

Soient deux populations P_1 et P_2 .

- σ_1^2 et σ_2^2 sont les variances (inconnues) dans les populations P_1 et P_2 ;
- E_1 et E_2 deux échantillons de tailles n_1 , n_2 tirés de P_1 et P_2 ;
- S_1^2 et S_2^2 sont les estimateurs sur E_1 et E_2 :

$$S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_{1i} - \hat{\mu}_1)^2 \qquad S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (X_{2i} - \hat{\mu}_2)^2$$

Y'a-t-il une différence significative entre s_1^2 et s_2^2 ?

tester $H_0 \sigma_1^2 = \sigma_2^2$ contre $H_1 \sigma_1^2 \neq \sigma_2^2$

Comparaison de deux variances

Mise en place du test

Uniquement pour des $(X_{1i})_{i=1,\dots,n_1}$ et des $(X_{2i})_{i=1,\dots,n_2}$ *normales*!

Penser à le tester!!!

$$(n_1 - 1) \frac{S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1) ; (n_2 - 1) \frac{S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

d'où

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

et *sous* H_0

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Comparaison de deux variances

Convention : choisir la numérotation des populations de telle façon que $s_1^2/s_2^2 > 1$

Pour un test bilatéral :

$\leadsto$ Rejet de H_0 au risque α si $\frac{S_1^2}{S_2^2} > F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$

Remarque : Et si plus de 2 échantillons ? On utilise alors le test de Bartlett !

Outline

Tests statistiques

Tests pour une population

Tests pour deux populations

Tests usuels à connaître

Tests à connaître

Tests paramétriques les observations sont supposées suivre un modèle gaussien ou l'échantillon est de suffisamment grande taille pour accepter la normalité asymptotique par le théorème de la limite centrale. Un échantillon

- Comparaison de la moyenne de l'échantillon à une valeur théorique lorsque la variance est supposée connue (Gauss)
- Comparaison de la moyenne de l'échantillon à une valeur théorique lorsque la variance est inconnue et estimée (Student)
- Comparer une proportion à une valeur théorique Deux échantillons indépendants

Deux échantillons

- Comparaison de deux moyennes (variances égales ou échantillon suffisamment grand, Student)
- Comparaison de deux variances (Fisher)
- Comparaison de deux proportions (χ^2)

Tests statistiques élémentaires :

Deux échantillons appariés : Le même échantillon est observé à deux instants différents ou dans deux conditions différentes (Student apparié)

Tests d'adéquation :

- Comparaison de deux distributions (chi-deux)
- Normalité d'une distribution (Kolmogorov, Shapiro Wilks)

Tests non-paramétriques (pas à connaître)

Dans le cas : petit échantillon et distribution non gaussienne (cf. tests d'adéquation) Deux échantillons indépendants : Comparaison de deux médianes (Mann-Whitney) Deux échantillons appariés : Test de Wilcoxon sur les différences