



Université Frères Mentouri Constantine 1
Faculté des sciences de la nature et de la vie

Département de Biologie Appliquée

Master1
Hygiène Hospitalière et Santé

Année Universitaire : 2019/2020

Cours de Bio-Statistique

Dr. Habiba BOUHALLOUF

8. Tests de comparaison

8.1 Introduction

Dans ce chapitre, on s'intéresse aux lois de distribution théoriques principales utilisées pour les tests statistiques de comparaison :

- La loi normale centrée réduite Z.
- La loi de Student T.
- La loi de Fisher F.
- La loi de χ^2 .

8.1.1 Comment choisir un test statistique ?

Le choix d'un test statistique dépend de plusieurs facteurs qu'on doit préalablement identifier :

La nature des variables à comparer : vérifier si

- la variable est quantitative (continue ou discrète).
- la variable est qualitative (binaire, nominale à plusieurs classes ou ordinale).

Les grandeurs étudiées : vérifier si :

- la grandeur est une moyenne.
- la grandeur est une variance.
- la grandeur est un pourcentage.
- la grandeur est un effectif.
- ...

La nature du problème : vérifier si on doit

- comparer un échantillon à une population de référence.
- comparer deux échantillons.
- comparer plusieurs échantillons.

Le type de séries comparées : vérifier si :

- les séries sont appariées.
- les séries sont indépendantes.

La taille des échantillons : vérifier si :

- l'échantillon a un grand nombre d'individus.
- l'échantillon a un petit nombre d'individus.

Les conditions d'application des tests : vérifier s'il s'agit d'une

- normalité des distributions associées à la population d'où est issu l'échantillon.
- égalité des variances.
- taille minimale des échantillons.

8.1.2 Principaux tests de comparaison et domaine d'application

Les tests de comparaison servent à comparer les moyennes, les variances, les pourcentages, ... des différentes distributions.

Leur principe consiste à poser une hypothèse nulle H_0 et de décider de l'accepter si l'égalité entre les paramètres est vérifiée ou de la rejeter si elle n'est pas vérifiée au profit d'une hypothèse alternative H_1 .

Pour effectuer le calcul du test, on choisit selon la nature de la comparaison la loi de distribution à suivre :

1. **La loi normale centrée réduite (Z)** sert à comparer :

- deux moyennes.
- deux moyenne observées de deux séries appariées.
- une moyenne observée à une moyenne théorique.

2. **La loi de Student (T)** sert à comparer :

- deux moyennes.
- deux moyenne observées de deux séries appariées.
- une moyenne observée à une moyenne théorique.

3. **La loi de Fisher (F)** sert à comparer :

- deux variances.
- plusieurs moyennes.
- deux pourcentage.

4. **La loi (χ^2)** sert à comparer :

- une distribution observée à une distribution théorique.
- plusieurs distributions.
- plusieurs pourcentages.

8.2 Test Z de l'écart réduit

8.2.1 Comparer une moyenne observée à une moyenne théorique

Conditions

Soit la variable quantitative x d'un échantillon de **grande taille** ($n \geq 30$) ayant la moyenne \bar{x} et l'écart type σ .

On cherche à décider si la moyenne de l'échantillon \bar{x} et la moyenne d'une population de référence μ sont significativement différentes. On teste alors, au risque de α :

- * l'hypothèse nulle $H_0 : \bar{x} = \mu$.
- * l'hypothèse alternative H_1 bilatérale : $\bar{x} \neq \mu$.
- * l'hypothèse alternative H_1 unilatérale : $\bar{x} < \mu$ ou $\bar{x} > \mu$.

Calcul

Pour identifier la région de rejet ou d'acceptation de l'hypothèse nulle, on calcule et on compare la quantité :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (8.1)$$

avec $z_{\frac{\alpha}{2}}$ pour l'hypothèse bilatérale ou avec z_{α} si l'hypothèse est unilatérale.

Si on prend $\alpha = 5\%$: $z_{\frac{\alpha}{2}} = 1,96$ et $z_{\alpha} = 1,65$.

Décision

H_1	Z	H_0	Décision
bilatérale	$ z < z_{\frac{\alpha}{2}}$	acceptée	\bar{x} n'est pas significativement différente de μ
bilatérale	$ z \geq z_{\frac{\alpha}{2}}$	rejetée	\bar{x} est significativement différente de μ
unilatérale	$z < z_{\alpha}$	acceptée	\bar{x} n'est pas significativement supérieure (ou inférieure) à μ
unilatérale	$z \geq z_{\alpha}$	rejetée	\bar{x} est significativement supérieure (ou inférieure) à μ

On trouve les valeurs de z_{α} et $z_{\frac{\alpha}{2}}$ dans la table de la loi Z de la loi normale centrée réduite.

8.2.2 Comparer deux moyennes

Conditions

Maintenant on veut comparer deux moyennes \bar{x}_1, \bar{x}_2 de deux échantillons indépendants de tailles supérieures à 30 ($n_1 \geq 30, n_2 \geq 30$). Sachant les écart types (σ_1, σ_2), On pose au risque de α , les hypothèses nulle et alternative :

- * $H_0 : \mu_1 = \mu_2$.
- * H_1 bilatérale : $\mu_1 \neq \mu_2$.
- * H_1 unilatérale : $\mu_1 < \mu_2$ ou $\mu_1 > \mu_2$.

où μ_1, μ_2 sont les moyennes inconnues des deux populations d'où sont issus nos échantillons.

Calcul

On calcul la quantité z à partir de la formule suivante :

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (8.2)$$

Décision

H_1	Z	H_0	Décision
bilatérale	$ z < z_{\frac{\alpha}{2}}$	acceptée	\bar{x}_1 n'est pas significativement différente de \bar{x}_2
bilatérale	$ z \geq z_{\frac{\alpha}{2}}$	rejetée	\bar{x}_1 est significativement différente de \bar{x}_2
unilatérale	$z < z_{\alpha}$	acceptée	\bar{x}_1 n'est pas significativement supérieure (ou inférieure) à \bar{x}_2
unilatérale	$z \geq z_{\alpha}$	rejetée	\bar{x}_1 est significativement supérieure (ou inférieure) à \bar{x}_2

On trouve les valeurs de z_{α} et $z_{\frac{\alpha}{2}}$ dans la table de la loi Z de la loi normale centrée réduite.

8.2.3 Comparer deux moyennes pour deux séries appariées

Conditions

Dans ce cas, nous avons un seul échantillon de taille ($n \geq 30$). On désire comparer deux valeurs qui appartiennent à deux séries de valeurs dites appariées d'une même grandeur (ici la moyenne) observée chez un individu. Chaque couple de mesures constitue une paire $((x_i, y_i), i = 1, \bar{n})$.

Le principe est de construire des paires de mesure puis calculer la différence observée pour chacune $d_i = |x_i - y_i|$.

Pour cela on pose les hypothèses nulle et alternative en fonction des moyennes calculées à partir des différences trouvées (\bar{x}_d et σ_d) :

$$* H_0 : \bar{x}_d = 0.$$

$$* H_1 \text{ bilatérale : } \bar{x}_d \neq 0.$$

$$* H_1 \text{ unilatérale : } \bar{x}_d < 0 \text{ ou } \bar{x}_d > 0.$$

où

$$\bar{x} = \sum d_i \quad (8.3)$$

Calcul

On calcul la quantité z à partir de la formule suivante :

$$z = \frac{\bar{x}_d - 0}{\sqrt{\frac{\sigma_d^2}{n}}} \quad (8.4)$$

où

$$\sigma_d^2 = \frac{1}{n-1} \left(\sum d_i^2 - \frac{1}{n} (\sum d_i)^2 \right) \quad (8.5)$$

Décision

H_1	Z	H_0	Décision
bilatérale	$ z < z_{\frac{\alpha}{2}}$	acceptée	les moyennes des séries ne sont pas significativement différentes
bilatérale	$ z \geq z_{\frac{\alpha}{2}}$	rejetée	les moyennes des séries sont significativement différentes
unilatérale	$z < z_{\alpha}$	acceptée	les moyennes des 2 séries sont significativement différentes
unilatérale	$z \geq z_{\alpha}$	rejetée	la moyenne d'une série est significativement supérieure (ou inférieure) à l'autre

On trouve les valeurs de z_{α} et $z_{\frac{\alpha}{2}}$ dans la table de la loi Z de la loi normale centrée réduite.

8.3 Test T de Student

8.3.1 Comparer une moyenne observée à une moyenne théorique

Conditions

On a vu qu'on peut utiliser le test Z pour comparer une moyenne observée à une moyenne connue dans une population de référence lorsqu'il s'agit d'un échantillon ayant une taille supérieures ou égale à 30.

Pour les petits échantillons ($n < 30$), on utilise plutôt le test de Student.

En effet, soient \bar{x} et σ la moyenne observée et l'écart type de l'échantillon respectivement, et μ la moyenne de la population de référence.

On au risque de ($\alpha = 10\%$), pose les hypothèses nulle et alternatives au risque de ($\alpha = 10\%$) :

$$* H_0 : \bar{x} = \mu.$$

$$* H_1 \text{ bilatérale} : \bar{x} \neq \mu.$$

$$* H_1 \text{ unilatérale} : \bar{x} < \mu \text{ ou } \bar{x} > \mu.$$

Calcul

de la même façon, on calcule la valeur de t du test de Student :

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (8.6)$$

Si l'hypothèse nulle ($\bar{x} = \mu$) est vraie (au risque de $\alpha = 10\%$), la quantité t suit une loi T de Student à $(n - 1)$ degrés de liberté (**ddl**).

Décision

H_1	T	H_0	Décision
bilatérale	$ t < t_{\frac{\alpha}{2}}$	acceptée	\bar{x} n'est pas significativement différente de μ
bilatérale	$ t \geq z_{\frac{\alpha}{2}}$	rejetée	\bar{x} est significativement différente de μ
unilatérale	$t < z_{\alpha}$	acceptée	\bar{x} n'est pas significativement supérieure (ou inférieure) à μ
unilatérale	$t \geq z_{\alpha}$	rejetée	\bar{x} est significativement supérieure (ou inférieure) à μ

On trouve les valeurs de $t_{5\%}$ et $t_{10\%}$ dans la table de la loi de Student.

8.3.2 Comparer deux moyennes

Conditions

On choisit le test de Student aussi lorsqu'on désire comparer deux moyennes observées dans deux échantillons de petites tailles ($n_1 < 30$, $n_2 < 30$).

En effet, on pose les hypothèses :

$$* H_0 : \mu_1 = \mu_2.$$

$$* H_1 \text{ bilatérale} : \mu_1 \neq \mu_2.$$

$$* H_1 \text{ unilatérale} : \mu_1 < \mu_2 \text{ ou } \mu_1 > \mu_2.$$

où μ_1, μ_2 sont les moyennes inconnues des deux populations d'où sont issus nos échantillons.

Calcul

On calcule la quantité z à partir de la formule suivante :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (8.7)$$

avec un nombre de degrés de liberté ($ddl = n_1 + n_2 - 2$).

Décision

H_1	T	H_0	Décision
bilatérale	$ t < t_{\frac{\alpha}{2}}$	acceptée	\bar{x}_1 n'est pas significativement différente de \bar{x}_2
bilatérale	$ t \geq t_{\frac{\alpha}{2}}$	rejetée	\bar{x}_1 est significativement différente de \bar{x}_2
unilatérale	$t < t_{\alpha}$	acceptée	\bar{x}_1 n'est pas significativement supérieure (ou inférieure) à \bar{x}_2
unilatérale	$t \geq t_{\alpha}$	rejetée	\bar{x}_1 est significativement supérieure (ou inférieure) à \bar{x}_2

On trouve les valeurs de $t_{5\%}$ et $t_{10\%}$ dans la table de la loi de Student.

8.3.3 Comparer deux moyennes pour deux séries appariées**Conditions**

Ce cas est similaire à celui où on utilise le test Z pour comparer deux moyennes de deux séries appariées observées dans un seul échantillon. Ce dernier doit avoir une taille inférieure à 30. En effet :

- * $H_0 : \bar{x}_d = 0$.
- * H_1 bilatérale : $\bar{x}_d \neq 0$.
- * H_1 unilatérale : $\bar{x}_d < 0$ ou $\bar{x}_d > 0$.

où μ_1, μ_2 sont les moyennes inconnues des deux populations d'où sont issus nos échantillons.

Calcul

On calcul la quantité t à partir de la formule suivante :

$$t = \frac{\bar{x}_d - 0}{\sqrt{\frac{\sigma_d^2}{n}}} \quad (8.8)$$

où \bar{x}_d et σ_d sont la moyenne des différences des moyennes (respectivement des écart types) des paires :

$$\bar{x} = \sum d_i,$$

$$\sigma_d^2 = \frac{1}{n-1} \left(\sum d_i^2 - \frac{1}{n} (\sum d_i)^2 \right)$$

Le nombre de degrés de liberté est : $dll = n - 1$.

Décision

H_1	T	H_0	Décision
bilatérale	$ t < t_{\frac{\alpha}{2}}$	acceptée	les moyennes des séries ne sont pas significativement différentes
bilatérale	$ t \geq t_{\frac{\alpha}{2}}$	rejetée	les moyennes des séries sont significativement différentes
unilatérale	$t < t_{\alpha}$	acceptée	les moyennes des 2 séries sont significativement différentes
unilatérale	$t \geq t_{\alpha}$	rejetée	la moyenne d'une série est significativement supérieure (ou inférieure) à l'autre

On trouve les valeurs de $t_{5\%}$ et $t_{10\%}$ dans la table de la loi de Student.

8.4 Test F de Fisher**8.4.1 Comparer deux variances****Conditions**

On utilise le test de Fisher lorsqu'il s'agit de comparer deux variances σ_1^2 et σ_2^2 de deux séries indépendantes de variables quantitatives quelque soit la taille des échantillons (n_1 et n_2).

Pour cela soient au risque de $\alpha = 5\%$, les hypothèses :

- * $H_0 : s_1^2 = s_2^2$.
- * H_1 bilatérale : $s_1^2 \neq s_2^2$.
- * H_1 unilatérale : $s_1^2 < s_2^2$ ou $s_1^2 > s_2^2$.

où s_1 et s_2 sont respectivement les variances inconnues des deux populations d'où sont issus les échantillons.

Calcul

Soit σ_1^2 la variance la plus élevée, on définit la quantité F par :

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad (8.9)$$

avec : ($dll_1 = k_1 = n_1 - 1$), ($dll_2 = k_2 = n_2 - 1$) et ($\alpha = 5\%$).

Décision

H_1	F	H_0	Décision
bilatérale	$ F < F_{\frac{\alpha}{2}}$	acceptée	σ_1^2 n'est pas significativement différente de σ_2^2
bilatérale	$ F \geq F_{\frac{\alpha}{2}}$	rejetée	σ_1^2 est significativement différente de σ_2^2
unilatérale	$F < F_{\alpha}$	acceptée	σ_1^2 n'est pas significativement différente de σ_2^2
unilatérale	$F \geq F_{\alpha}$	rejetée	σ_1^2 est significativement supérieure à σ_2^2

On trouve les valeurs de $F_{5\%}$ et $F_{2,5\%}$ dans la table de la loi de Fisher.

8.4.2 Comparer plusieurs moyennes

Conditions

On choisit le test de Fisher également lorsqu'on désire comparer les moyennes observées lors d'une expérience d'une variable quantitative pour plusieurs échantillons. Pour cela les distributions des populations d'où proviennent les échantillons doivent suivre la loi normale et avoir la même variance.

En effet, au risque de ($\alpha = 5\%$), les hypothèses nulle et alternatives sont :

- * H_0 : Les moyennes ne sont pas significativement différentes.
- * H_1 : Les moyennes sont significativement différentes.

Alors, si les séries étudiées sont indépendantes, on teste le rapport entre deux variances :

1. La variance entre individus de chaque série (variance résiduelle) (σ_r^2).
2. La variance entre les séries étudiées (variance générale) (σ_g^2).

où

$$\sigma_g^2 = \frac{\sum \frac{x_i^2}{n_i} - \frac{x_g^2}{N}}{c - 1}, \quad \sigma_r^2 = \frac{\sum x^2 - \sum \frac{x_i^2}{n_i}}{N - c} \quad (8.10)$$

avec :

- * x_i : est la somme des valeurs observées sur chaque série.
- * x_g : est la somme des valeurs observées sur toutes les séries.
- * $\sum x^2$: est la somme des carrés des valeurs observées sur toutes les séries.
- * n : est le nombre des effectifs de chaque série.
- * N : est le nombre des effectifs de toutes les séries.
- * c : est le nombre de séries à comparer.

Calcul

on teste le rapport suivant :

$$F = \frac{\sigma_g^2}{\sigma_r^2} \quad (8.11)$$

avec : ($dll_1 = k_1 = c - 1$), ($dll_2 = k_2 = N - c$) et ($\alpha = 5\%$).

Décision

H_1	F	H_0	Décision
bilatérale	$F < F_\alpha$	acceptée	Les moyennes ne sont pas significativement différentes
bilatérale	$F \geq F_\alpha$	rejetée	Les moyennes sont significativement différentes

8.5 Test de χ^2

Il est fréquent d'avoir à comparer deux grandeurs caractérisées par des variables qualitatives comme en épidémiologie par exemple (malade / non malade).

On compare, Le plus souvent des distributions ou des pourcentages observés dans deux échantillons ou plus, mais parfois on est tombé dans des situations où on doit comparer un pourcentage observé dans un échantillon à un pourcentage théorique (de référence) pour savoir si la différence entre eux est due aux fluctuations d'échantillonnage ou elle correspond à une différence réelle.

Pour tous ça, si tous les effectifs théoriques (de référence) sont supérieurs ou égaux à 5, on utilise le test de comparaison de χ^2 (chi-2).

8.5.1 Comparer une distribution observée à une distribution théorique (Test de χ^2 d'ajustement)**Conditions**

Les effectifs observés sont généralement différents des effectifs théoriques, donc on teste la conformité entre la distribution expérimentale et la distribution théorique. Ce genre de test se fait en utilisant "Test de χ^2 " d'ajustement (ou de conformité) à condition d'avoir un nombre d'effectifs supérieur ou égal à 5.

On considère un échantillon de taille n extrait au hasard d'une population partagée en classes, chaque classe est une modalité de réalisation du caractère. Et on attribut les effectifs : $f_{o_1}, f_{o_2}, f_{o_3}, \dots, f_{o_k}$, tels que :

$$n = \sum_{i=1}^k f_{o_i} \quad (8.12)$$

où f_{o_i} sont appelés "effectifs observés".

D'autre part, on attribut à ces classes les probabilités : $P_1, P_2, P_3, \dots, P_k$. P_i représente la probabilité de présence de la modalité du caractère étudié dans la classe i ($i = \overline{1, k}$) avec :

$$\sum_{i=1}^k P_i = 1 \quad (8.13)$$

Donc, on peut définir les effectifs théoriques en fonction des probabilités ainsi :

$$f_{t_i} = nP_i \quad (8.14)$$

où

$$n = \sum_{i=1}^k f_{t_i} \quad (8.15)$$

On peut finalement construire le tableau suivant :

Classe	1	2	...	i	...	k	Total
Effectifs Observés (f_{o_i})	f_{o_1}	f_{o_2}	...	f_{o_i}	...	f_{o_k}	n
Effectifs Théoriques (f_{t_i})	f_{t_1}	f_{t_2}	...	f_{t_i}	...	f_{t_k}	n

On pose alors les hypothèses suivantes au risque de ($\alpha = 5\%$ ou $\alpha = 1\%$) :

- * H_0 : "Il y a une conformité entre la distribution étudiée (expérimentale) et la distribution théorique".
- * H_1 : "Il y a une différence significative entre la distribution étudiée (expérimentale) et la distribution théorique".

Calcul

Pour décider si on accepte ou on rejette l'hypothèse H_0 , on calcule la statistique de χ^2 :

$$\chi^2_c = \sum_{i=1}^k \frac{(f_{o_i} - f_{t_i})^2}{f_{t_i}} \quad (8.16)$$

χ^2_c doit être comparée avec χ^2_α qui peut être obtenue par la table de χ^2 en fonction de $\nu = k - 1$ (k est le nombre de classes étudiées).

Décision

H_1	χ^2	H_0	Décision
biilatérale	$\chi^2_c < \chi^2_\alpha$	acceptée	Les distributions ne sont pas significativement différentes
biilatérale	$\chi^2_c \geq \chi^2_\alpha$	rejetée	Les distributions sont significativement différentes

8.5.2 Comparer plusieurs pourcentages (Test de χ^2 d'homogénéité)

Conditions

On applique le test de χ^2 d'homogénéité pour comparer plusieurs distributions ou plusieurs pourcentages (pour une variable qualitative à plusieurs classes). Comme nous avons vu dans la section précédente, on va s'intéresser aux effectifs plutôt que les pourcentages. Par conséquent, on obtient le tableau suivant :

Échantillon → Classe ↓	éch(1)	éch(2)	...	éch(j)
1	$f_{o_{11}} (f_{t_{11}})$	$f_{o_{12}} (f_{t_{12}})$...	$f_{o_{1j}} (f_{t_{1j}})$
2	$f_{o_{21}} (f_{t_{21}})$	$f_{o_{22}} (f_{t_{22}})$...	$f_{o_{2j}} (f_{t_{2j}})$
3	$f_{o_{31}} (f_{t_{31}})$	$f_{o_{32}} (f_{t_{32}})$...	$f_{o_{3j}} (f_{t_{3j}})$
⋮	⋮	⋮	⋮	⋮
i	$f_{o_{i1}} (f_{t_{i1}})$	$f_{o_{i2}} (f_{t_{i2}})$...	$f_{o_{ij}} (f_{t_{ij}})$

f_{oij} (f_{tij}) sont respectivement les effectifs observés (les effectifs théoriques) de la classe (modalité du caractère) (i) de l'échantillon (j).

Tous les effectifs théoriques doivent être supérieurs ou égaux à 5 pour pouvoir appliquer la loi de χ^2 .

Au risque de ($\alpha = 5\%$), les hypothèses nulle et alternative sont données par :

- * H_0 : "Il n' y a pas une différence significative entre les pourcentages (distributions).
- * H_1 : "Il y a pas une différence significative entre les pourcentages (distributions).

Calcul

Pour décider si on accepte ou on rejette l'hypothèse H_0 au risque de ($\alpha = 5\%$), on calcule χ^2_c :

$$\chi^2_c = \sum_{i=1}^k \sum_{j=1}^c \frac{(f_{oij} - f_{tij})^2}{f_{tij}} \quad (8.17)$$

χ^2_c doit être comparée avec χ^2_α qui peut être obtenue par la table de χ^2 à l'aide du nombre de degrés de liberté : $dll = (c - 1)(k - 1)$ (k est le nombre de classes étudiées, c est le nombre d'échantillon).

Décision

H_1	χ^2	H_0	Décision
biilatérale	$\chi^2_c < \chi^2_\alpha$	acceptée	Les pourcentage (les distributions) ne sont pas significativement différentes
biilatérale	$\chi^2_c \geq \chi^2_\alpha$	rejetée	Les pourcentage (les distributions) sont significativement différentes

