Unité Fondamentale : Statistique 2 1ère année
Travaux Dirigés
ENSAT

J. Gergaud

septembre 2003

Table des matières

Chapitre 1

Ajustement de données à une loi théorique

1 Problème

On a une série statistique $y_1, y_2, \ldots, y_n$ qui est un échantillon simple et aléatoire extrait d'une population générale $\mathcal{P}$. On désire savoir si on peut supposer que la loi du caractère Y étudié est une loi théorique fixé. Soit $F(y)$ la fonction de répartition de Y. Nous pouvons alors formaliser notre problème sous la forme d'un test statistique :
Hypothèse nulle : $H_0 : F = F_0$
Hypothèse alternative $H_1 : F \neq F_0$

Nous allons voir ici trois méthodes. Deux méthodes basées sur des tests statistiques et une méthode graphique.

2 Test du χ^2

2.1 Principe

Le test du χ^2 peut être utilisé que la loi soit discrète ou continue :

- Si elle est discrète nous supposerons que la variable possède q modalités et nous travaillerons avec le tableau des fréquences absolues :

Modalités	Mod_1	Mod_2	$\cdots\cdots\cdots$	Mod_{q-1}	Mod_q
Fréquences	n_1	n_2	$\cdots\cdots\cdots$	n_{q-1}	n_q

- Si elle est continue, nous supposerons que l'on a défini q classes et nous travaillerons avec le tableau des fréquences absolues suivant :

classes	$]-\infty, x_1]$	$]x_1, x_2]$	$\cdots\cdots\cdots$	$]x_{q-2}, x_{q-1}]$	$]x_{q-1}, +\infty[$
Fréquences	n_1	n_2	$\cdots\cdots\cdots$	n_{q-1}	n_q

Le test du Khi-2 est alors basé sur la comparaison des fréquences absolues observées n_i aux fréquences théoriques n'_i que nous "devrions" avoir si la loi de Y suivait la loi F_0. La statistique utilisée ici quantifiant cette "distance" est définie par :

$$\begin{array}{rcl} D^2 : \mathcal{P}^n & \longrightarrow & \mathbf{R} \\ E_n & \longmapsto & D^2 = \displaystyle\sum_{i=1}^{q} \frac{(n_i - n'_i)^2}{n'_i} \end{array}$$

On démontre alors que lorsque n tend vers $+\infty$ la loi de D^2 suit une loi du χ^2 à $q-1$ degrés de liberté (on "perd" un ddl car on a la relation $n = \sum_{i=1}^{q} n_i$). En pratique, pour que l'approximation soit correcte il suffit que toutes les **fréquences théoriques** n'_i soit supérieures ou égales à 5. Dans le cas contraire, on procède à des regroupements. Il faut aussi que le nombre de classes soit suffisant (au moins une dizaine).

On en déduit le test : on rejettera l'hypothèse nulle H_0 au risque α si la valeur observée de la statistique D^2_{obs} est supérieure à la valeur χ^2_{crit} définie par : $P(\chi^2 < \chi^2_{crit}) = 1 - \alpha$.

Remarque. Lorsque la variable aléatoire Y est continue on peut aussi écrire la statistique ci-dessus de la façon suivante :

$$D^2 = n \sum_{i=1}^{q} \frac{(f_i - f'_i)^2}{f'_i}$$

où

- $f_i = n_i/n$ est l'estimation de la probabilité $P(Y \in]x_{i-1}, x_i])$
- $f_i' = n_i'/n = P(Y \in]x_{i-1}, x_i])$

Par conséquent on peut interpréter la quantité ci-dessus comme une grandeur exprimant la "distance" entre l'histogramme et la fonction de densité théorique (fig. 1.1). Par suite une grande valeur de cette quantité exprime que l'échantillon observé est "loin" d'un échantillon provenant d'une loi normale.

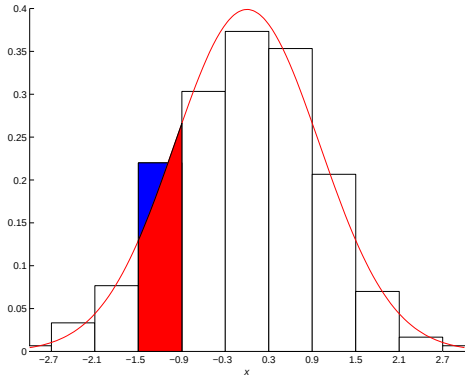

FIG. 1.1 – $f_i - f_i'$ = écart entre l'histogramme et la fonction de densité théorique.

2.2 Application au test d'adéquation à une famille de lois

Dans la pratique, seul la forme de la loi est spécifiée, loi de Poisson, loi normale, ... Mais on ignore certains paramètres qui seront donc estimés. La seule différence avec ce qui a été fait au paragraphe précédent est qu'à chaque fois que l'on estime un paramètre il faut enlever un ddl à la loi du χ^2. Ainsi en pratique :

- pour un test d'adéquation à une loi de Poisson, le ddl sera de $q - 2$,
- pour un test d'adéquation à une loi normale, le ddl sera de $q - 3$.

2.3 Mode opératoire pour le cas de la loi normale

(i) Se donner le risque de première espèce α.

(ii) Répartir les observations en classes.

(iii) Calculer les estimations de μ et de σ (sur les données groupées), soient $\bar{x}$ et $\hat{\sigma}$.

(iv) Calculer les limites des classes correspondant à la loi normale centrée réduite

$$u_i = \frac{(x_i - \bar{x})}{\hat{\sigma}}$$

(v) Rechercher les $\phi(u_i) = P(U < u_i)$ dans la table de la loi normale centrée réduite.

(vi) Calculer les effectifs théoriques : $n_i' = n(\phi(u_i) - \phi(u_{i-1}))$.

(vii) Calculer

$$\chi^2_{obs} = \sum_{i=1}^{q} \frac{(n_i - n_i')^2}{n_i'}$$

(viii) Rechercher la valeur du χ^2_{crit} dans la table. Cette valeur est définie par

$$P(\chi^2 \leq \chi^2_{crit}) = 1 - \alpha$$

.

(ix) Comparer χ^2_{obs} et χ^2_{crit}.
Si $\chi^2_{obs} < \chi^2_{crit}$ alors on accepte l'hypothèse de normalité au risque α.
Si $\chi^2_{obs} \geq \chi^2_{crit}$ alors on rejette l'hypothèse de normalité au risque α.

2.4 Calcul préliminaire

limites de classes x_i	limites de classes u_i	$\phi(u_i)$	probabilités d'appartenance à la classe i	effectifs théoriques $n'_i = P_i \times n$	effectifs observés n_i	écarts
					χ^2_{obs}	

Conclusion

Si $\chi^2_{obs} < \chi^2_{crit}$ alors on accepte l'hypothèse de normalité au risque α.
Si $\chi^2_{obs} \geq \chi^2_{crit}$ alors on rejette l'hypothèse de normalité au risque α.

3 Test de Lilliefors

3.1 Principe

Le test du de Kolmogorov est basé sur la comparaison des fréquences relatives cumulées empiriques et théoriques. On considère la statistique suivante :

$$\begin{array}{rcl} K : \mathcal{P}^n & \longrightarrow & \mathbf{R} \\ E_n & \longmapsto & K = \left\{ \begin{array}{l} Max(Max(|F_i - F_0(x_i)|, |F_{i-1} - F_0(x_i)|) \\ i = 1, \ldots, n \end{array} \right. \end{array}$$

Les valeurs critiques ont été tabulées pour les distributions parfaitement connues. Pour le cas d'une loi normale de paramètre μ et σ inconnus on a une bonne approximation des valeurs critiques par les formules suivantes [1] :

risque α de 0,05	risque α de 0,01
$0,886/\sqrt{n+1,5}$	$1,031/\sqrt{n+1,5}$

Remarque. Ici cette quantité est une mesure de l'écart entre les fonctions de répartitions empirique et théorique.

3.2 Mode opératoire pour le cas d'une loi normale

(i) Se donner le risque de première espèce α.

(ii) Calculer les estimations de μ et de σ (sur les données initiales ou groupées), soient $\bar{x}$ et $\hat{\sigma}$.

(iii) Centrer et réduire les données

$$u_i = \frac{(x_i - \bar{x})}{\hat{\sigma}}$$

[1] voir Dagnélie Tome 2 page 71

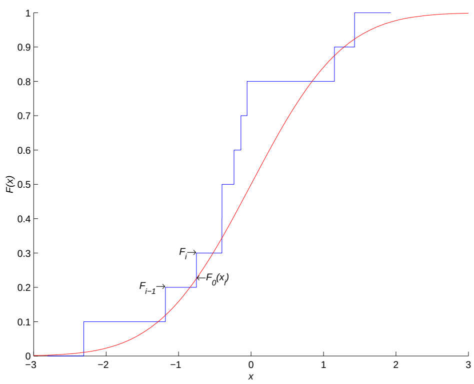

FIG. 1.2 – $F_i - F_0(x_i)$ = écart entre les fréquences cumulées relatives et la fonction de répartition.

(iv) Rechercher les $\phi(u_i)$ dans la table de la loi normale centrée réduite.

(v) Calculer les fréquences cumulées relatives $F_i = \sum_{k=1}^{i} n_k/n$.

(vi) Calculer

$$K = \begin{cases} Max(Max|F_i - \phi(u_i)|, |F_{i-1} - \phi(u_i)|)) \\ i = 1, \dots, n \end{cases}$$

(vii) Calculer la valeur du K critique K_{crit}.

(viii) Comparer K_{obs} et K_{crit}.
Si $K_{obs} < K_{crit}$ alors on accepte l'hypothèse de normalité au risque α.
Si $K_{obs} \geq K_{crit}$ alors on rejette l'hypothèse de normalité au risque α.

3.3 Calcul préliminaire

limites de classes x_i	limites de classes u_i	$\phi(u_i)$	fréquences cumulées relatives F_i	$Max(\lvert F_i - \phi(u_i)\rvert, \lvert F_{i-1} - \phi(u_i)\rvert)$
			K_{obs}	

Conclusion

Si $K_{obs} < K_{crit}$ alors on accepte l'hypothèse de normalité au risque α.
Si $K_{obs} \geq K_{crit}$ alors on rejette l'hypothèse de normalité au risque α.

4 Remarques

– Le test de Kolmogorov ne peut pas être utilisé pour les lois discrètes.

5 Application

5.1 Exemple 1

Distribution des rapports diamétraux de 815 feuilles de tabac (Institut Expérimental de Bergerac). Variété Paraguay P 19 [2]

classes	effectifs
1,45-1,55	0
1,55-1,65	4
1,65-1,75	9
1,75-1,85	31
1,85-1,95	75
1,95-2,05	183
2,05-2,15	204
2,15-2,25	157
2,25-2,35	97
2,35-2,45	40
2,45-2,55	12
2,55-2,65	3
2,65-2,75	0

(i) Réaliser le test du χ^2 pour tester la normalité des données.

5.2 Exemple 2

Dans une étude sur les mécanismes de détoxication du brochet du nord, on a procédé au dosage du DDT et de ces dérivés DDD et DDE contenus dans l'organisme d'individus capturés dans la rivière Richelieu (prov. du Quebec). Les résultats, exprimés en milligrammes par litre, des analyses effectuées sur les brochets de 3 ans figurent dans le tableau ci-dessous[3]

0,285	0,295	0,321	0,254	0,359	0,361	0,362
0,364	0,373	0,382	0,403	0,407	0,413	

(i) Testez l'adéquation de ces données à une loi normale.

5.3 Exemple 3

Le tableau suivant donne les fréquences observées du nombre de cellules de levure des 400 carrés dans lesquels est divisé $1mm^2$ de milieu[4]

[2] Données recueillies dans "Analyse Statistique" de E. Morice et F. Chartier, Paris, 1954.
[3] Données provenant de l'ouvrage de B. Scherrer, Biostatistique,Ed. G. Morin, 1984, page 390.
[4] Exemple provenant du livre de M. Lamotte, Introduction à la biologie quantitative, Ed. Masson, 1948, page 90 (données de Student)

Nombre de cellules	Fréquences observées
0	0
1	20
2	43
3	53
4	86
5	70
6	54
7	37
8	18
9	10
10	5
11	2
12	0
13	0
14	0
15	0
Total	400

(i) Réaliser le test d'adéquation de ces données à une loi de Poisson.

Chapitre 2

Comparaison de 2 variances

1 Problème

On a deux séries statistiques $y_{11}, y_{12}, \ldots, y_{1n_1}$ et $y_{21}, y_{22}, \ldots, y_{2n_2}$ qui sont deux échantillons simples et indépendants provenant de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$ et on désire savoir si les variances des populations parentes σ_1^2 et σ_2^2 sont égales.

2 Test

2.1 Postulats

(i) Les variables aléatoires Y_{ij}, $i = 1, 2$ et $j = 1, \ldots, n_i$ ont pour variance σ_i^2.

(ii) Les variables aléatoires Y_{ij}, $i = 1, 2$ et $j = 1, \ldots, n_i$ ont pour loi $\mathcal{N}(\mu_i, \sigma_i^2)$ ou ($n_1 > 30$ et $n_2 > 30$).

(iii) Les variables aléatoires $(Y_{ij})_{ij}$, $i = 1, 2$ et $j = 1, \ldots, n_i$ sont indépendantes.

2.2 Hypothèses

Hypothèse nulle H_0 **:** $\sigma_1 = \sigma_2$.
Hypothèse alternative H_1 **:** $\sigma_1 \neq \sigma_2$ pour un test bilatéral.

2.3 Principe

On définie la statistique :

$$\begin{array}{rcl} F : \mathcal{P}^{n_1} \times \mathcal{P}^{n_2} & \longrightarrow & \mathbf{R} \\ (E_{n_1}, E_{n_2}) & \longmapsto & F = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \end{array}$$

où

$$\hat{\sigma}_i^2 = \frac{SCE_i}{n_i - 1}$$

On démontre que **si $\mathbf{H_0}$ est vraie alors** F suit une loi de Fisher à $(n_1 - 1, n_2 - 1)$ ddl.

Remarque. Dans le cas où le test est bilatéral on rejettera l'hypothèse nulle si la valeur de F_{obs} est éloignée de 1, c'est-à-dire si F_{obs} est "très proche de 0" ou "très grand". Le règle de décision sera donc :

- Si $F_{\alpha/2} < F_{obs} < F_{1-\alpha/2}$ alors on accepte l'hypothèse nulle au risque α ;
- Si $F_{obs} < F_{\alpha/2}$ ou $F_{obs} > F_{1-\alpha/2}$ alors on rejette l'hypothèse nulle au risque α.

Or si F suit une loi de Fisher à $(\nu_1, \nu_2)ddl$ alors $1/F$ suit une loi de Fisher à $(\nu_2, \nu_1)ddl$. Par suite la double inégalité $F_{\alpha/2} < F_{obs} < F_{1-\alpha/2}$ est équivalente à la procédure :

- Prendre $F'_{obs} = Max(F_{obs}, 1/F_{obs})$
- Comparer F'_{obs} à $F_{1-\alpha/2}$ (avec les bons ddl).

2.4 Mode opératoire

Pour un test bilatéral.

(i) Se donner le risque de première espèce α.

(ii) Calculer F_{obs} (Prendre celui des rapports supérieur à 1).

(iii) Rechercher F_{crit} dans la table. Cette valeur est donnée par :

$$P(F \leq F_{crit}) = 1 - \alpha/2$$

(iv) Comparer F_{obs} et F_{crit}
Si $F_{obs} < F_{crit}$ alors on accepte l'hypothèse d'égalité des variances au risque α.
Si $F_{obs} \geq F_{crit}$ alors on rejette l'hypothèse d'égalité des variances au risque α.

3 Application

3.1 Exemple

Lors d'un TP de chimie, 13 étudiants dosent la teneur en arsenic d'une certaine solution 1, tandis que 13 autres dosent celle d'une solution 2 ; voici les résultats : [1]

Solution 1	3.14	3.19	2.98	3.26	3.01	3.3	3.04	3.19	3.22	3.32	3.4	3.18	3.16
Solution 2	3.16	3.13	3.17	2.68	3.06	2.91	3.37	3.08	2.73				

(i) Quels graphiques feriez-vous ?

(ii) Tester l'égalité des variances.

3.2 Hypothèses

Hypothèse nulle H_0 :
Hypothèse alternative H_1 :

Conclusion

[1] L'exemple provient de Bernard Prum, "Modèle linéaire compraison de groupes et régression", ed. Les ditions INSERM, 1996. page 19

Chapitre 3

Comparaison de 2 moyennes

Echantillons indépendants

1 Problème

On a deux séries statistiques $y_{11}, y_{12}, \ldots, y_{1n_1}$ et $y_{21}, y_{22}, \ldots, y_{2n_2}$ qui sont deux échantillons simples et indépendants provenant de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$ et on désire savoir si les moyennes des populations parentes μ_1 et μ_2 sont égales.

2 Test

2.1 Postulats

(i) Les variables aléatoires Y_{ij}, $i = 1, 2$ et $j = 1, \ldots, n_i$ ont la même variance σ^2.

(ii) Les variables aléatoires Y_{ij}, $i = 1, 2$ et $j = 1, \ldots, n_i$ ont pour loi $\mathcal{N}(\mu_i, \sigma^2)$ ou ($n_1 > 30$ et $n_2 > 30$).

(iii) Les variables aléatoires Y_{ij}, $i = 1, 2$ et $j = 1, \ldots, n_i$ sont indépendantes.

2.2 Hypothèses

Hypothèse nulle H_0 **:** $\mu_1 = \mu_2$.
Hypothèse alternative H_1 **:** $\mu_1 < \mu_2$ pour un test unilatéral à gauche
Hypothèse alternative H_1 **:** $\mu_1 > \mu_2$ pour un test unilatéral à droite
Hypothèse alternative H_1 **:** $\mu_1 \neq \mu_2$ pour un test bilatéral.

2.3 Principe

On définie la statistique :

$$\begin{array}{rcl} T : \mathcal{P}^{n_1} \times \mathcal{P}^{n_2} & \longrightarrow & \mathbf{R} \\ (E_{n_1}, E_{n_2}) & \longmapsto & T = \dfrac{\bar{y}_1 - \bar{y}_2}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \end{array}$$

où

$$\hat{\sigma}^2 = \frac{SCE_1 + SCE_2}{n_1 + n_2 - 2}$$

On démontre que **si** H_0 **est vraie alors** T suit une loi de Student à $\nu = n_1 + n_2 - 2$ ddl.

2.4 Mode opératoire

(i) Se donner le risque de première espèce α.

(ii) Calculer T_{obs}

(iii) Rechercher T_{crit} dans la table. Cette valeur est donnée par :

(a) $P(T \leq T_{crit}) = 1 - \alpha$ pour un test unilatéral à droite.

(b) $P(T \leq T_{crit}) = \alpha$ pour un test unilatéral à gauche.

(c) $P(T \leq T_{crit}) = 1 - \alpha/2$ pour un test bilatéral.

(iv) Comparer T_{obs} et T_{crit}

(a) Pour un test unilatéral à gauche :
Si $T_{obs} > T_{crit}$ alors on accepte l'hypothèse d'égalité des moyennes au risque α.
Si $T_{obs} \leq T_{crit}$ alors on rejette l'hypothèse d'égalité des moyennes au risque α.

(b) Pour un test unilatéral à droite :
Si $T_{obs} < T_{crit}$ alors on accepte l'hypothèse d'égalité des moyennes au risque α.
Si $T_{obs} \geq T_{crit}$ alors on rejette l'hypothèse d'égalité des moyennes au risque α.

(c) Pour un test bilatéral :
Si $|T_{obs}| < T_{crit}$ alors on accepte l'hypothèse d'égalité des moyennes au risque α.
Si $|T_{obs}| \geq T_{crit}$ alors on rejette l'hypothèse d'égalité des moyennes au risque α.

3 Application

3.1 Exemple

Lors d'un TP de chimie, 13 étudiants dosent la teneur en arsenic d'une certaine solution 1, tandis que 9 autres dosent celle d'une solution 2 ; voici les résultats : [1]

Solution 1	3.14	3.19	2.98	3.26	3.01	3.3	3.04	3.19	3.22	3.32	3.4	3.18	3.16
Solution 2	3.16	3.13	3.17	2.68	3.06	2.91	3.37	3.08	2.73				

(i) Quels graphiques feriez-vous ?

(ii) Tester l'égalité des moyennes.

3.2 Hypothèses

Hypothèse nulle H_0 :
Hypothèse alternative H_1 :

Conclusion

[1] L'exemple provient de Bernard Prum, "Modèle linéaire compraison de groupes et régression", ed. Les ditions INSERM, 1996. page 19

Echantillon appariés

4 Problème

Le problème est toujours de comparer deux moyennes mais ici les deux échantillons sont associés par paires, c'est-à-dire que pour chaque individu pris au hasard j nous aurons deux observations y_{1j} et y_{2j} correspondant aux deux caractères Y_1 et Y_2. Par suite nous aurons donc un même nombre de mesures pour les deux caractères : $y_{11}, y_{12}, \ldots, y_{1n}$ et $y_{21}, y_{22}, \ldots, y_{2n}$.

5 Test

5.1 Postulats

(i) Les variables aléatoires $D_j = Y_{1j} - Y_{2j}$, $j = 1, \ldots, n$ sont de loi $\mathcal{N}(\delta, \sigma_D^2)$ ou $n > 30$.

(ii) Les variables aléatoires $D_j = Y_{1j} - Y_{2j}$, $j = 1, \ldots, n$ sont indépendantes.

5.2 Hypothèses

Hypothèse nulle H_0 **:** $\delta = \mu_1 - \mu_2 = 0$.
Hypothèse alternative H_1 **:** $\mu_1 - \mu_2 < 0$ pour un test unilatéral à gauche.
Hypothèse alternative H_1 **:** $\mu_1 - \mu_2 > 0$ pour un test unilatéral à droite.

Hypothèse alternative H_1 **:** $\mu_1 - \mu_2 \neq 0$ pour un test bilatéral.

5.3 Principe

On note $d_j = y_{1j} - y_{2j}$, on définie alors la statistique :

$$\begin{array}{rcl} T : \mathcal{P}^n & \longrightarrow & \mathbf{R} \\ E_n & \longmapsto & T = \dfrac{\bar{d}}{\frac{\hat{\sigma}_D}{\sqrt{n}}} \end{array}$$

où

$$\hat{\sigma}_D^2 = \frac{SCE(d)}{n-1}$$

est l'estimation de la variance des différences.

On démontre que **si** H_0 **est vraie alors** T suit une loi de Student à $\nu = n - 1$ ddl.

5.4 Mode opératoire

(i) Se donner le risque de première espèce α.

(ii) Calculer T_{obs}

(iii) Rechercher T_{crit} dans la table. Cette valeur est donnée par :

(a) $P(T \leq T_{crit}) = 1 - \alpha$ pour un test unilatéral.

(b) $P(T \leq T_{crit}) = 1 - \alpha/2$ pour un test bilatéral.

(iv) Comparer T_{obs} et T_{crit}

(a) Pour un test unilatéral :
Si $T_{obs} < T_{crit}$ alors on accepte l'hypothèse d'égalité des moyennes au risque α.
Si $T_{obs} \geq T_{crit}$ alors on rejette l'hypothèse d'égalité des moyennes au risque α.

(b) Pour un test bilatéral :
Si $|T_{obs}| < T_{crit}$ alors on accepte l'hypothèse d'égalité des moyennes au risque α.
Si $|T_{obs}| \geq T_{crit}$ alors on rejette l'hypothèse d'égalité des moyennes au risque α.

6 Application

6.1 Exemple

On veut savoir si deux préparations virales produisent des effets différents sur une variété de tabac. L'expérience est réalisée sur la deuxième feuille de 8 plantes. On coupe la feuille en 2 et on soumet chaque moitié à l'action d'une souche virale. La mesure de la virulence est données par le nombre de lésions localisées apparaissant sur chaque demi-feuille. Les données recueilliés sont données ci-dessous[2]

Paires	Préparation 1	Préparation 2
1	31	18
2	20	17
3	18	14
4	17	11
5	9	10
6	8	7
7	10	5
8	7	6

(i) Quels graphiques feriez-vous ?

(ii) Tester l'égalité des moyennes.

6.2 Hypothèses

Hypothèse nulle H_0 :
Hypothèse alternative H_1 :

Conclusion

[2]Ref. Snedecor et Cochran, Méthodes statistiques, Paris.

Chapitre 4

Notion de puissance d'un test

1 Introduction

Nous allons ici voir sur un exemple simple la notion de puissance d'un test. Cette notion est très importante puisque c'est elle qui justifie souvent l'emploi d'un test par rapport à un autre lorsque l'on a le choix et que c'est elle qui nous permet de déterminer le nombre de mesures nécessaires pour mettre en évidence les différences, si elles existent, que l'expérimentateur désire constater.

1.1 Rappels

On considère un test statistique et donc une hypothèse nulle H_0 et une hypothèse alternative H_1.

Définition 1.1.1 (Risque de première espèce). On appelle risque de première espèce la probabilité de rejeter l'hypothèse nulle quand elle est vraie, ou encore c'est la probabilité de conclure à une différence qui en réalité n'existe pas et qui n'est due qu'au hasard.

$$\alpha = P_{H_0}(\text{Rejeter } H_0)$$

Définition 1.1.2 (Risque de deuxième espèce). On appelle risque de deuxième espèce la probabilité d'accepter l'hypothèse nulle quand elle est fausse, ou encore c'est la probabilité de ne pas déceler une différence qui existe.

$$\beta = P_{H_1}(\text{Accepter } H_0)$$

Définition 1.1.3 (Puissance). On appelle puissance d'un test la quantité $1 - \beta$

1.2 Application

On considère le cas du test de Student unilatéral à droite pour des échantillons indépendants de mêmes tailles n. Nous avons donc deux série $y_{11}, \ldots, y_{1n}$ et $y_{21}, \ldots, y_{2n}$ provenant de populations normales de même variance σ^2 et nous voulons tester l'hypothèse nulle
$\mathrm{H}_0 : \delta = \mu_1 - \mu_2 = 0$ contre l'hypothèse alternative $\mathrm{H}_1 : \delta = \mu_1 - \mu_2 > 0$.

Théorème 1.2.1. *Pour l'exemple précédent nous avons la relation suivante :*

$$n = 2\frac{\hat{\sigma}^2}{\delta^2}(t_{1-\beta} + t_{1-\alpha})^2 \tag{4.1}$$

Démonstration
On démontre que la statistique suivante :

$$T_\delta = \frac{\bar{y}_1 - \bar{y}_2 - \delta}{\hat{\sigma}\sqrt{\frac{2}{n}}}$$

avec

$$\hat{\sigma}^2 = \frac{SCE_1 + SCE_2}{2n - 2}$$

suit une loi se Student à $2n - 2$ ddl.

Nous avons alors

$$\alpha = P_{H_0}(T_0 > t_{1-\alpha})$$

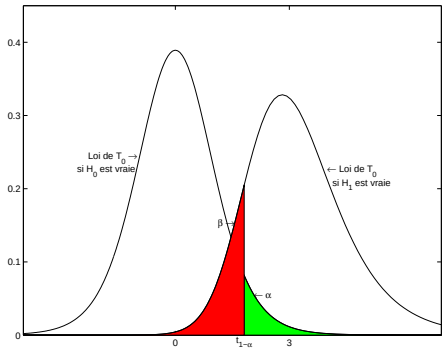

FIG. 4.1 – Visualisation des risques α et β.

et

$$\begin{aligned}\beta &= P_{H_1}(T_0 < t_{1-\alpha}) \\ &= P_{H_1}(T_0 - \frac{\delta}{\hat{\sigma}\sqrt{\frac{2}{n}}} < t_{1-\alpha} - \frac{\delta}{\hat{\sigma}\sqrt{\frac{2}{n}}}) \\ &= P_{H_1}(T_\delta < t_{1-\alpha} - \frac{\delta}{\hat{\sigma}\sqrt{\frac{2}{n}}})\end{aligned}$$

Par suite nous avons :

$$t_\beta = t_{1-\alpha} - \frac{\delta}{\hat{\sigma}\sqrt{\frac{2}{n}}}$$

d'où le résultat. □

Remarque 1.2.2. L'équation (4.1) est non linéaire en n car n apparaît implicitement dans les valeurs $t_{1-\alpha}$ et t_β, mais pour n grand on peut pratiquement remplacer ces valeurs par $u_{1-\alpha}$ et u_β.

Théorème 1.2.3. *Dans le cas d'un test bilatéral on obtient :*

$$n \simeq 2\frac{\hat{\sigma}^2}{\delta^2}(t_{1-\beta} + t_{1-\alpha/2})^2$$

Remarque 1.2.4. Les tables donnent en fait les courbes de n pour α et β fixés en fonction de $\Delta/cv = \delta/\hat{\sigma}$ avec : $\Delta = (\delta/\bar{y}) \times 100\%$ et $cv = (\hat{\sigma}/\bar{y}) \times 100\%$.

2 Exercice

Un expérimentateur veut comparer 2 régimes alimentaires sur des bovins. La différence qu'il voudrait mettre en évidence entre les deux moyennes est de 90g/j sur le poids moyen journalier (GPMJ) alors qu'il s'attend à un résultat moyen de l'ordre de 1200g/j.

Une bonne estimation du cœfficient de variation que l'on peut espérer sur le GPMJ lui a été donnée par une expérience antérieure dont la durée (8 mois) est analogue à celle prévue $\hat{cv} = \frac{\hat{\sigma}}{\bar{y}} \times 100 = 10\%$.

Il souhaite que son essai ait 9 chances sur 10 de déceler la différence de 90g/j si celle-ci existe. Le risque de première espèce choisi est $\alpha = 0,05$ et il s'intéresse à la différence quel que soit le signe de celle-ci.

(i) En fait pour des raisons pratiques, l'expérimentateur ne peut disposer que de 12 répétitions. Quelle est la différence qu'il peut espérer mettre en évidence ? (Tous les autres paramètres du problème restant constants).

(ii) En considérant que l'expérimentateur ne peut toujours disposer que de 12 animaux par traitement et que la différence intéressante à déceler est de 90g/j, quelle est la probabilité de "sortir" cette différence si elle existe ?

(iii) Combien faut-il de bovins par traitement pour résoudre le problème initial ?

Chapitre 5

Test d'indépendance

1 Problème

On étudie sur une population $\mathcal{P}$ deux caractères A et B quantitatifs ou qualitatifs. On désire savoir si ces deux caractères sont indépendants ou non.[1]

2 Test

2.1 Données

Sur un échantillon simple et aléatoire de n individus de la population on mesure 2 caractères A et B répartis respectivement en p et q modalités (ou classes). On obtient donc un tableau de contingence :

$A:B$	B_1	B_2	$\dots$	B_j	$\dots$	B_q	$Totaux$
A_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	n_{1q}	$n_{1.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	n_{iq}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_p	n_{p1}	n_{p2}	$\dots$	n_{pj}	$\dots$	n_{pq}	$n_{p.}$
$Totaux$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.j}$	$\dots$	$n_{.q}$	$n_{..}$

2.2 Hypothèse nulle

H_0 **:** Les deux caractères A et B sont indépendants.

2.3 Principe

On calcule la statistique :

$$\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{n_{ij}^2}{nij'} - n$$

avec

$$n'_{ij} = \frac{n_{i.}n_{.j}}{n}$$

On démontre que :
Si $20 \leq n \leq 40$ et si $n'_{ij} \geq 5$ pour tout i, j
ou si $n > 40$ et $n'_{ij} \geq 5$ sauf pour quelques classes
alors la statistique ci-dessus suit approximativement une loi du χ^2 à $(p-1)(q-1)$ ddl.

Remarque. n'_{ij} représente en fait l'effectif théorique le plus probable si l'hypothèse nulle est vraie. La quantité ci-dessus peut donc être vue comme une mesure de l'écart entre les effectifs obtenus et les effectifs théoriques.

[1] La procédure s'applique dans les mêmes conditions et avec les même notations au cas du **test d'homogénéité** permettant de vérifier la validité de l'hypothèse selon laquelle p distributions réparties en q classes ou modalités sont égales

2.4 Mode opératoire

Calcul préliminaire

$A:B$	B_1	B_2	...	B_j	...	B_q	$Totaux$
A_1	n_{11} n'_{11}	n_{12} n'_{12}	...	n_{1j} n'_{1j}	...	n_{1q} n'_{1q}	$n_{1.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1} n'_{i1}	n_{i2} n'_{i2}	...	n_{ij} n'_{ij}	...	n_{iq} n'_{iq}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_p	n_{p1} n'_{p1}	n_{p2} n'_{p2}	...	n_{pj} n'_{pj}	...	n_{pq} n'_{pq}	$n_{p.}$
$Totaux$	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.q}$	$n_{..}$

(i) Se donner le risque de première espèce α.

(ii) Calculer les effectifs théoriques n'_{ij}.

(iii) Calculer χ^2_{obs}

(iv) Rechercher la valeur du χ^2_{crit} dans la table. Cette valeur est définie par

$$P(\chi^2 \leq \chi^2_{crit}) = 1 - \alpha$$

.

(v) Comparer χ^2_{obs} et χ^2_{crit}.
Si $\chi^2_{obs} < \chi^2_{crit}$ alors on accepte l'hypothèse d'indépendance au risque α.
Si $\chi^2_{obs} \geq \chi^2_{crit}$ alors on rejette l'hypothèse d'indépendance au risque α.

3 Application

3.1 Exemple 1

On a extrait d'une population de blé (variété PUSA 12), un échantillon de 412 individus classés suivant 2 caractères : la longueur de l'épi et le nombre de grains par épi. On a obtenu les données suivantes[2]

		nombre de grains par épi			
		10-20	20-30	30-40	40-50
longueur de l'épi	5-7	12	3	1	1
	7-9	11	56	8	2
	9-11	3	91	115	10
	11-13	1	3	55	27
	13-15	1	2	4	6

(i) Quels graphiques feriez-vous ?

(ii) Tester l'indépendance des deux caractères.

3.2 Exemple 2

Pour une élection où il y a trois candidats, on désire savoir si les femmes et les hommes ont le même comportement. C'est-à-dire si les populations des hommes et des femmes sont homogènes pour ce critère. On réalise pour cela un sondage sur 200 hommes et 100 femmes et on a obtenu le tableau suivant :

[2]Données recueillies dans "Méthodes estadisticos para investigadores agricolas" de V.G. Panse et P.V. Sukhatme, 1969.

Candidats :Populations	Hommes	Femmes
1	68	22
2	51	24
3	81	54

(i) Quels graphiques feriez-vous ?

(ii) Tester l'homogénéité des deux populations.

Chapitre 6

Tests "non paramétriques"

1 Introduction

On appelle tests non paramétriques, les tests où les postulats nécessaires quant aux distributions des variables que l'on manipule sont très faibles[1]. Les postulats d'indépendance sont eux toujours supposés vérifiés. Nous avons en fait déjà rencontré ce type de test (test du χ^2 par exemple).

Nous verrons ici les tests de comparaisons de deux distributions. Il s'agira donc du test correspondant au test de Student dans le cas de lois normales. Nous verrons donc ici aussi deux cas : le cas d'échantillons indépendants et le cas d'échantillons appariés. Nous ne donnerons les résultats que dans le cas de tests bilatéraux.

2 Remarques

D'une façon générale nous pouvons faire les remarques suivantes entre les tests non paramétriques et paramétriques :

(i) Le paramètre de position est la médiane au lieu de la moyenne.

(ii) Le paramètre de dispersion est l'amplitude ou l'écart interquartiles au lieu de l'écart-type.

(iii) Les tests non paramétriques sont moins puissants que les tests paramétriques ; il faut plus d'observations pour avoir la même puissance par rapport à un test paramétrique.

(iv) Les calculs sont plus simples pour les tests non paramétriques.

(v) D'une façon générale les tests non paramétriques utilisent les rangs des observations.

3 Echantillons indépendants

3.1 Problème

On a deux séries statistiques $y_{11}, y_{12}, \ldots, y_{1n_1}$ et $y_{21}, y_{22}, \ldots, y_{2n_2}$ qui sont deux échantillons simples et indépendants provenant de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$ et on désire savoir si les deux distributions sont égales.

3.2 Test de la médiane

Hypothèse nulle

H_0 : les distributions sont identiques.

Principe

Il s'agit tout simplement d'un test d'homogénéité. On travaille donc sur le tableau de contingence suivant :

	$nb\ d'obs. \leq \tilde{x}$	$nb\ d'obs. > \tilde{x}$	$Totaux$
E_1	n_{11}	n_{12}	$n_{1.}$
E_2	n_{21}	n_{22}	$n_{2.}$
$Totaux$	$n_{.1}$	$n_{.2}$	$n_{..}$

[1] En anglais on dit "distribution-free" ce qui est plus parlant

et la statistique est donc définie par :

$$\begin{aligned} \chi^2 : \mathcal{P}^{n_1} \times \mathcal{P}^{n_2} &\longrightarrow \mathbf{R} \\ (E_{n_1}, E_{n_2}) &\longmapsto \chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}} = \frac{n_{..}(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \end{aligned}$$

qui suit une loi du χ^2 à 1 ddl si l'hypothèse H_0 est vraie.

Mode opératoire

Voir test d'indépendance
Remarque. On peut bien évidemment faire plus de 2 classes en utilisant les quartiles par exemple.

3.3 Test des rangs–Test de Mann et Whitney

Hypothèse nulle

H_0 : les distributions sont identiques.

Principe

La statistique utilisée est la somme des rangs des observations d'un des deux échantillons. Mann et Whitney ont étudié la distribution obtenue lorsque l'hypothèse nulle est vraie et ont tabulé les valeurs critiques.

Mode opératoire

(i) Se donner le risque de première espèce α.

(ii) Classer les observations par ordre croissant.

(iii) Déterminer le rang de chaque observation. Pour les ex-æquo on détermine un rang moyen, par exemple :

E_1	19		26			31		...
E_2		23		26	28		32	...
Rangs	1	2	3,5	3,5	5	6	7	...

(iv) Calculer la somme des rangs de chaque échantillon S_1 et S_2. On remarquera que $S_1 + S_2 = (n_1 + n_2)(n_1 + n_2 + 1)/2$.

(v) Aller rechercher dans la table l'intervalle $I =]S_{min}, S_{max}[$.

(vi) Si S_1 est dans I alors on accepte l'hypothèse nulle au risque α.
Si S_1 n'est pas dans I alors on rejette l'hypothèse nulle au risque α.

3.4 Application

Lors d'une étude sur le ver du maïs on a compté le nombre d'œufs de ce parasite trouvés sur 20 plants. Ceux-ci ont été choisis dans des emplacements pris au hasard (on suppose que le champ est homogène). On divise l'échantillon en 2 catégories de même effectif en fonction de la hauteur des plants :
- L'échantillon E_1 correspond à des plants inférieurs à 60 cm.
- L'échantillon E_2 correspond à des plants supérieurs à 60 cm.

On a obtenu les résultats suivants[2] :

	Nombre d'œufs									
E_1	0	14	18	0	31	0	0	0	11	0
E_2	37	42	12	32	105	84	15	47	51	65

(i) Quels graphiques feriez-vous ?

(ii) Le nombre d'œufs est-il significativement différent en fonction de la hauteur des plants ?

[2]Données provenant de G.W. Snedecor, W.G. Cochran, Méthodes statistiques, Paris, 1971

4 Echantillons appariés

4.1 Problème

On désire toujours savoir si les 2 distributions sont identiques mais les deux échantillons sont ici associés par paires, c'est-à-dire que pour chaque individu pris au hasard i nous aurons deux observations $y_{1,j}$ et $y_{2,j}$ correspondant aux deux caractères Y_1 et Y_2. Par suite nous aurons donc un même nombre de mesures pour les deux caractères : $y_{11}, y_{12}, \ldots, y_{1n}$ et $y_{21}, y_{22}, \ldots, y_{2n}$.

4.2 Test des signes

Hypothèse nulle

H_0 : les distributions sont identiques.

Postulats

Les grandeurs observées sur chaque individu doivent être telles qu'il soit possible d'attribuer un ordre dans chaque paire.

Principe

On s'intéresse au signe des différences sur chaque paire et on considère alors l'hypothèse nulle suivante :
$H_0 : P(+) = P(-) = 1/2$.
et l'hypothèse alternative H_1 est donc $P(+) \neq P(-)$ (test bilatéral)
où $P(+)$ est la probabilité d'observer une différence positive
et $P(-)$ esr la probabilité d'observer une différence négative.
On définit alors la statistique suivante :

$$\begin{array}{rcl} N : \mathcal{P}^n & \longrightarrow & \mathbf{R} \\ E_n & \longmapsto & N = \text{nombre de différences positives} \end{array}$$

Si H_0 **est vraie alors** N suit une loi binomiale de paramètre $(n, 1/2)$

Mode opératoire

(i) Se donner un risque de première espèce α.

(ii) Calculer $N(+)$ et $N(-)$ les nombres de différences positives et négatives (S'il y a des différences nulles alors il faut supprimer ces observations et donc diminuer n en conséquence).

(iii) Calculer $N_{obs} = Min(N(+), N(-))$.

(iv) Calculer

$$p = P(N \leq N_{obs}) = \left(\frac{1}{2}\right)^n \sum_{i=0}^{N_{obs}} C_n^i$$

(v) Comparer p et $\alpha/2$
Si $p \leq \alpha/2$ alors on rejette l'hypothèse nulle au risque α
Si $p > \alpha/2$ alors on accepte l'hypothèse nulle au risque α.

4.3 Test des rangs de Wilcoxon

Hypothèse nulle

H_0 : les distributions sont identiques.

Postulats

Les variables doivent être continues.

Principe

La statistique utilisée est la somme des rangs des différences positives (ou négatives).

$$\begin{array}{rcl} T : \mathcal{P}^n & \longrightarrow & \mathbf{R} \\ E_n & \longmapsto & T = \text{somme des rangs des différences positives} \end{array}$$

Wilcoxon a étudié la loi de T et a tabulé les valeurs critiques.

Mode opératoire

(i) Se donner un risque de première espèce α.

(ii) Calculer $T(+)$ et $T(-)$ les sommes des rangs des différences positives et négatives (on vérifie que $T(+) + T(-) = n(n+1)/2$) (S'il y a des différences nulles alors il faut supprimer ces observations et donc diminuer n en conséquence).

(iii) Calculer $T_{obs} = Min(T(+), T(-))$.

(iv) Chercher dans la table la valeur critique T_{crit}

(v) Comparer T_{obs} et T_{crit}
Si $T_{obs} < T_{crit}$ alors on rejette l'hypothèse nulle au risque α
Si $T_{obs} \geq T_{crit}$ alors on accepte l'hypothèse nulle au risque α.

5 Application

On étudie l'effet d'une substance hypotensive sur la pression artérielle systolique. Chaque malade reçoit successivement :

(i) Un traitement 1

(ii) Un traitement 2

A la fin de chaque période du traitement on mesure la pression artérielle du sujet. On a obtenu les données suivantes :

malade	après traitement 1	après traitement 2
01	17	16
02	15	11
03	15	12
04	13	13
05	12	14
06	17	11
07	15	13
08	16	13
09	19	17
10	11	10

(i) Quels graphiques feriez-vous ?

(ii) Question : Les traitements ont-ils des effets différents ?

Chapitre 7

Régression linéaire simple : cas avec terme constant

1 Problème

[1]On désire prédire la hauteur (en feet) d'un arbre d'une essence donnée à partir de la connaissance de son diamètre à 1m30 (en inches). Pour cela on a collecté les données suivantes :

Diamètres	2.3	2.5	2.6	3.1	3.4	3.7	3.9	4.0	4.1	4.1
Hauters	7	8	4	4	6	6	12	8	5	7
Diamètres	4.2	4.4	4.7	5.1	5.5	5.8	6.2	6.9	6.9	7.3
Hauteurs	8	7	9	10	13	7	11	11	16	14

2 Calculs

2.1 Calculs préliminaires

	y	x	xy	y^2	x^2
1	y_1	x_1	x_1y_1	y_1^2	x_1^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	y_i	x_i	x_iy_i	y_i^2	x_i^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_n	x_ny_n	y_n^2	x_n^2
Totaux	$Y_.$	$X_.$	$\sum_i x_iy_i$	$\sum_i y_i^2$	$\sum_i x_i^2$
Moyennes	$\bar{y}_.$	$\bar{x}_.$			

2.2 Tableau d'analyse de la variance

Source de variation	SCE	ddl	carrés moyens	Statistique F	
(1) Totale	$\sum_i (y_i - \bar{y})^2$ $= \sum_i y_i^2 - c_y$	$n-1$			
(2) Régression	$\sum_i (\hat{y}_i - \bar{y})^2$ $= \hat{\beta}_1^2 (\sum_i x_i^2 - c_x)$	1	$CM_{y(x)}$	$F_{obs} = \frac{CM_{y(x)}}{CM_R}$	F_{crit}
Résiduelle	(1)-(2)	$n-2$	$CM_R = \frac{SCE_R}{n-2}$		

avec : $c_x = X_.\bar{x}_.$ et $c_y = Y_.\bar{y}_.$

$$\hat{\beta}_1 = \frac{SPE}{SCE_x} = \frac{\sum_i (x_i - \bar{x}_.)(y_i - \bar{y}_.)}{\sum_i (x_i - \bar{x}_.)^2} = \frac{\sum_i x_iy_i - n\bar{x}_.\bar{y}_.}{\sum_i x_i^2 - c_x}$$

3 Travail demandé

Sous Excel

[1]Données provenant de BRUCE et SCHUMACHER Forest mensuration - Mc Graw-Hill book company, ine - 1950 - 3e édition

(i) entrer les données ;
(ii) faire le graphe de points ;
(iii) réaliser les calculs préliminaires ;
(iv) calculer la droite de régression de y en x ;
(v) réaliser l'analyse de la variance ;
(vi) calculer tous les résidus ;
(vii) calculer la moyenne des résidus et la somme des carrés des résidus ;
(viii) Donner $\hat{\sigma}^2$ et calculer l'intervalle de confiance de σ^2.
(ix) Calculer les résidus standardisés
(x) donner l'intervalle de confiance à 95% de β_0 et de β_1 ;
(xi) calculer l'intervalle de confiance 95% de la prévision et de la prévision de la moyenne pour $x = 5$;
(xii) calculer les quantités :

$$\left(\frac{\hat{\beta}_1}{\hat{\sigma}_{B_1}}\right)^2$$

et

$$(t_{1-\alpha/2})^2$$

(xiii) Commentaires.

Remarque 3.0.1. On utilisera pour les tests et les intervalles de confiance les fonctions LOI.F, LOI.STUDENT, LOI.STUDENT.INVERSE, KHI2.INVERSE, et INVERSEMAT.

Remarque 3.0.2. On utilisera une feuille de calcul pour les calculs préliminaires, une feuille de calcul pour le graphique des points, l'équation de la droite de régression et le tableau d'analyse de la variance et une feuille de calcul pour l'analyse des résidus.

Chapitre 8

Exercises

Exercice 8.1. Mendel dans ses expériences sur les pois a obtenu pour une génération F_2 les données suivantes :

graines lisses	5474
graines ridées	1850

On désire savoir si le caractère étudié suit bien le ratio 3 : 1 pour les phénotipes dominant et récessif.

(i) Quelle est la variable aléatoire étudiée ?

(ii) Quels graphiques feriez-vous ?

(iii) Quels sont les hypothèses nulle et alternative

(iv) Quels tests peut-on réaliser ?

(v) Application numérique.

Exercice 8.2. Pour tester deux méthodes de dosages A et B on a effectué dans une solution homogène 20 prélèvements. Pour 10 prélèvements on a effectué le dosage avec la méthode A et pour les 10 autres avec la méthode B. Les résultats suivants ont été obtenus :

Méthode A	$59.6^{(16)}$	$50.8^{(10)}$	$62.4^{(11)}$	$97.2^{(2)}$	$61.3^{(20)}$	$56.5^{(18)}$	$55.7^{(9)}$	$69.5^{(15)}$	$57.6^{(1)}$	$53.6^{(5)}$
Méthode B	$74.9^{(12)}$	$78.3^{(19)}$	$80.4^{(14)}$	$58.7^{(6)}$	$68.1^{(8)}$	$64.7^{(3)}$	$66.5^{(7)}$	$73.5^{(4)}$	$81.0^{(17)}$	$73.7^{(13)}$

On donne :

$$\bar{y}_{1.} = 62.42 \quad \bar{y}_{2.} = 71.98 \quad \sum_{j=1}^{10} y_{1j}^2 = 40548 \quad \sum_{j=1}^{10} y_{2j}^2 = 52292$$

Les exposants entre parenthèses indiquent l'ordre dans lequel les essais ont été effectués.

(i) Quels sont les variables aléatoires étudiées ?

(ii) Quels graphiques feriez vous ?

(iii) Quelles questions poseriez-vous au chimiste pour savoir comment analyser ces données et quel test feriez-vous ?

(iv) A quoi peuvent servir les indications sur l'ordre dans la collecte des données ?

(v) Donner les hypothèses nulle et alternative.

(vi) Réaliser le test.

Exercice 8.3. On a fait une numération globulaire à 10 malades atteints d'une maladie M et on a obtenu les résultats suivants en nombre de globules blancs par mm^3 :

7400	8600	7900	9600	7100	7700	9400	8400	7300	9100

On suppose que le nombre de globules blancs par mm^3 suit une loi normale d'espérance μ et de variance σ^2

(i) Quelle est la variable aléatoire étudiée ?

(ii) Quel graphique feriez-vous ?

(iii) Sachant que $\mu = 7400$ pour les individus sains, quel test feriez-vous ? (on donnera les postulats et les hypothèses nulle et alternative).

(iv) Réaliser le test.

(v) Calculez l'intervalle de confiance à 95% pour l'espérance μ.

(vi) Commentaire

Exercice 8.4. L'eau destinée à l'alimentation humaine doit être dépourvue de toute contamination bactérienne. Afin de déterminer le degré de pollution bactérienne des eaux, on a effectué des prélèvements dans les 6 bassins hydrographique de France et recherché la présence ou l'absence de salmonelles (germes pathogènes plus résistants que les coliformes aux traitements classiques de chloration et d'ozonisation). Les résultats obtenus à 119 points de prélèvements, répartis dans les 6 bassins hydrographiques, sont les suivants :

Pollution :Bassins	Artois-Picardie	Rhin-Meuse	Loire-Bretagne	Adour-Garonne	Rhône-Méditérranée	Seine-Normandie	Total France
Présence de salmonelles	8	8	10	6	11	16	59
Absence de salmonelles	5	2	11	2	17	23	60
Total	13	10	21	8	28	39	119

(i) Que désire-t-on savoir dans ce type d'expérience ?

(ii) Quelles sont les variables aléatoires étudiées ?

(iii) Quels graphiques feriez-vous ?

(iv) Quel test feriez-vous ?

(v) Réalisez ce test