République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA
FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE
DÉPARTEMENT DE Biologie

Chapitre 01
Tests des Comparaisons :

Par
Prof : CHALA ADEL

2021-2022

Je dédie ce travail............
A mes parents ils m'ont tous,
avec leurs moyens, soutenu et donné
la force d'aller toujours
plus loin.
A ma chère femme Houda.
A l'esprit du professeur Bahlali Seid

Table des matières

Introduction

En 3ème année, les étudiants de Sciences de la Nature et de la Vie se voient proposer des formations spécialisées nécessitant des connaissances et un savoir-faire statistiques qui ne peuvent être acquis en tronc commun. D'autre part, certains étudiants, sans envisager a priori une spécialisation statistique, peuvent désirer acquérir une formation approfondie en méthodes statistiques.

Cette formation est particulièrement appréciée pour un débouché professionnel dans les domaines de l'expérimentation, et préparation de Master et post Doctoral.

Chaque méthode statistique est motivée par une présentation de problèmes concrets, par des utilisateurs dans différents domaines : agronomie, écologie, génétique, médecine, ...

Les connaissances acquises concernent l'estimation des paramètres, les tests statistiques (validité du modèle, effet des variables explicatives), la prévision et la sélection de variables dans le cadre du modèle linéaire (régression simple, analyse de la variance à plusieurs facteurs, analyse de la covariance).

Choix du modèle en fonction du type de données, structuration des données, traitement statistique et informatique (logiciel SPSS) pour les modèles de régression multiple et d'analyse de la variance ou de la covariance, ainsi que pour des extensions de ce modèle (ACP, AFC).

Ce module apporte une formation solide en statistique inférentielle directement exploitable dans de nombreux Masters Végétale ou Master Biologie Moléculaires, ou options de Poste Doctoral. Il donne des compétences indispensables pour la collecte et le traitement de données expérimentales. En cela, il constitue un pré requis important pour des formations en génétique, écologie et en sciences de l'environnement (Master Biologie)

Mots clés : Statistique inférentielle, modèle linéaire (régression, analyse de la variance et de la covariance), modèles mixtes (effet aléatoire), sélection de variables

Chapitre 1

Introduction aux théories des probabilités et statistique :

Lorsqu'on veut étudier les données relatives aux caractéristiques d'un ensemble d'individus ou d'objets il est difficile d'observer toutes les données lorsque leurs nombres sont élevés. Au lieu d'examiner l'ensemble qu'on appelle population on examine un nombre restreint qu'on appelle échantillon, pour être représentatif l'échantillon doit être pris au hasard (une population peut-être finie ou infinie).

Population : C'est l'ensemble sur lequel porte l'étude statistique.

Individus : Les éléments de cet ensemble.

Echantillon : Est un sous-ensemble de la population.

Caractère : C'est le trait (ou propriété) choisi pour l'étude statistique.

Modalités : Les différentes positions que peut prendre un caractère. Usage on numérote les modalités de 1 à k la modalité numéro i est notée C_i

Effectifs : Lorsque la population est répartie sur les différentes modalités nous obtenons pour chacune d'elles un nombre c'est le nombre des individus ayant cette modalité. On note habituellement n_i l'effectif correspondant à la modalité C_i : les fréquence absolu.

Fréquence relative : Par définition c'est le rapport entre n_i et N, où N est la somme totale des individus.

Nous allons ainsi adopter les définitions suivantes :

*Un caractère est dit quantitatif quand ses différentes modalités sont mesurables par des nombres qui en indiquent l'intensité.

*Un caractère est dit qualitatif quand ses différentes modalités ne peuvent être désignées que par leurs qualités.

*Une variable statistique est dite discrète lorsque ses modalités ne peuvent être que des nombres isolés.

*Une variable statistique est dite continue quand elle peut prendre n'importe quelle valeur dans un intervalle donné.

* **Le mode** : c'est la valeur la plus fréquente.

* **La médiane** : C'est la valeur de la variable statistique qui partage la population en deux populations d'effectifs égaux.

* **Les quartiles :** Comme on a définie la médiane on peut définir des paramètres qui la répartissent en quarts.

* **La moyenne arithmétique** : est égale par définition $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$.

b/ Caractéristiques de dispersion :

* **L'étendue :** C'est la longueur de l'intervalle sur lequel se disperse la variable.

***L'écart-interquartiles** : C'est la différence entre les deux quartiles Q_1 et Q_3.

* **La variance :** C'est la caractéristique qui est réellement utilisée pour mesurer la dispersion :

Chapitre 2

La théorie du test statistique

2.1 Test de confirmité :

Les tests de confirmité sont destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée ou représentatif de cette population, vis-à-vis d'un paramètre comme la moyenne, la variance ou la fréquence observée. Ceci implique que la loi théorique du paramètre est connue au niveau de la population.

2.1.1 Comparaison d'une répartition observée à une répartition théorique "Test du χ^2".

La répartition théorique ayant été choisie, il est naturel de se demander si elle représente bien la répartition expérimentale ; Si elle lui est bien confirme. La vérification de la confirmité de la répartition théorique choisie à la répartition expérimentale donnée est faite au moyen du test du χ^2.

Test du χ^2:

Le test du χ^2 se fait selon les étapes suivantes :

1/ On pose l'hypothèse nulle : $\mathbf{H_0}$" Il y a un confirmité (ou concordance) entre la répartition théorique et la répartition expérimentale".

2/ On calcul la quantité suivante :

$$\chi^2_{obs} = \sum_{i=1}^{k} \frac{(O_i - C_i)^2}{C_i},$$

où : O_i : effectifs observés.

C_i : L'effectifs calculées.

k : c'est le nombre de modalités.

3/ Conclusion : Etant donnée un seuil de signification α, On utilise alors la table de $\chi^2_{(k-1)}$. On le note χ^2_α. On applique ensuite la règle de décision suivante :

a) Si $\chi^2_{obs} < \chi^2_\alpha$, l'hypothèse $\mathbf{H_0}$ est retenue c'est à dire que : la distribution observée est conforme à la distribution théorique.

b) Si $\chi^2_{obs} \geq \chi^2_\alpha$ On dit alors que $\mathbf{H_0}$ est rejetée.

Remarque importante : Le test du χ^2 (Khy-deux) ne peut être utilisé que si tous les effectifs calculés sont suffisamment grands.

Exemple N°: 01

Prenons un dosage biologique, qui peut être normale, faible ou fort selon qu'il se situe entre deux bornes, est inferieur à la plus petite, ou supérieur à la plus grande, avec $k = 3$ modalités. On veut tester le fait que 90% des gens ont un dosage normal. alors que 5% l'ont faible et 5% l'ont fort. Pour cela, on tire au hasard 100 sujets et on constate que, sur les 100 dosages 76 sont normaux, 10 faibles et 14 forts. Quelle sera la conclusion ?

Solution Exemple N°: 01

On remarque ici que $k = 3$ et $N = 100$ et $\alpha = 05\%$.

1) **L'hypothèse nulle** H₀ = Il y a une confirmité entre la répartition observée et la répartition calculée

	Normale	Faible	Fort	Totale
Proportion Calculée	0,9	0,05	0,05	1
C_i	90	05	05	100
O_i	76	10	14	100

2) **Calcul de** χ^2 :

$$\chi^2_{obs} = \sum_{1}^{3} \frac{(O_i - C_i)^2}{C_i} = \frac{(76-90)^2}{90} + \frac{(10-5)^2}{5} + \frac{(14-5)^2}{5} = 23,378.$$

De plus

$$\chi^2_\alpha = \chi^2_{(k-1,1-\alpha)} = \chi^2_{(2,0,5)} = 5,991.$$

3) **Conclusion :** Il est évident que $\chi^2_{obs} > \chi^2_\alpha$. on rejette H_0, alors pas de confirmité entre les dosage théoriques et observés.

Exemple N°: 02

On a croisé deux races de plantes différentes par deux caractères A et B. La première génération est homogène, la seconde génération fait apparaître 4 types de plantes, dont les phénotypes sont notés AB, Ab, aB, et ab.

Si les caractères se transmettent selon les lois de Mendel, les proportions théoriques des 4 phénotypes sont 9/16, 3/16, 3/16 et 1/16.

Dans une expérience, un échantillon de 160 plantes a donné :

AB	Ab	aB	ab
100	18	24	18

Cette répartition est-elle conforme aux lois de Mendel au seuil de signification de 5% ?

Solution Exemple N°: 02

1) Hypothèse nulle : $\mathbf{H_0}$ = (La répartition observée est conforme aux loi de Mendel), pour calculer χ^2_{obs} on établit le tableau :

Phénotype	AB	Ab	aB	ab	totale
Proportion théorique	9/16	3/16	3/16	1/16	1
Effectif calculé	90	30	30	10	160
Effectif observé	100	18	24	18	160

2) Le calcul : On obtient alors

$$\chi^2_{obs} = \sum_{1}^{3} \frac{(O_i - C_i)^2}{C_i} = \frac{(10)^2}{90} + \frac{(-12)^2}{30} + \frac{(-6)^2}{30} + \frac{(8)^2}{10} = 13,51$$

Sur la table de Khi-deux on lit $\chi^2_\alpha = 7,815$.

3) Conclusion : Alors comme conclusion on peut dire que $\mathbf{H}_0$ doit être rejeter au seuil de signification 05%.

Exemple N°: 03

Les résultats des épreuves d'un examen à l'échelle nationale sont : 60% de reçus, 25% admissibles (admis à passer les épreuves orales) et 15% éliminés.

Un établissement présente 160 élèves et obtient 75 reçus, 53 admissibles et 32 élimines.

Y a t-il confirmité entre ces résultats et ceux valables à l'échelle nationale ?($\alpha = 0;01$).

Solution Exemple N°: 03

1) Hypothèse nulle : $\mathbf{H_0}$" Les résultats obtenues sont conformes a ceux valables à l'échelle nationale",

2) Calcul : On établit le tableau qui permet de calculer la quantité χ^2_{obs}

	O_i	p_i	$C_i = np_i$	$(O_i - C_i)^2$	$\frac{(O_i-C_i)^2}{C_i}$
Reçus	75	0,60	96	441	4,593
Admissible	53	0,25	40	169	4,225
Eliminé	32	0,15	24	64	2,666
Total	160	1	160		11,484

3) Conclusion : Comme le nombre k de modalité est égale à 3, alors $\chi^2_\alpha = 5,99$. d'où au seuil de signification de 05% il convient de rejeter $\mathbf{H_0}$:

Mais au seuil de signification de 0,1% il y a donc confirmité entre ces résultats et ceux valables a l'échelle nationale.

2.1.2 Comparaison d'un pourcentage observé a un pourcentage théorique :

On extrait au hasard dans une population un échantillon de taille n , soit p le pourcentage du caractère A, le problème qui se pose alors est :

-La divergence constatée entre p et p_0 peut-elle être expliquée uniquement par les fluctuation d'échantionnage ou bien les résultats expérimentaux sont-ils en contradiction avec les valeurs théoriques p_0 :

Si n est assez grand et p_0 est très approche a 0 ou 1 $(np = 05)$ la comparaison entre p et p_0 calculée (théorique) est basée sur l'écart-réduit :

$$\xi_{obs} = \frac{|p - p_0|}{\sqrt{\frac{p_0 q_0}{n}}}.$$

Au seuil de signification α.

*Si $\xi_{obs} \leq \xi_\alpha = 1,96$ La différence n'est pas significative.

*Si $\xi_{obs} > \xi_\alpha = 1,96$ La différence est significative.

Exemple N°: 04

Dans une population qui comporte autant de garçons que de filles, une maladie a frappé 08 filles et 02 garçons.

Cette maladie frappe t-elle davantage les filles ?

Solution Exemple N°: 04

1) Hypothèse nulle : Il s'agit de savoir $\mathbf{H}_0$ "$p_0 = 0,50$ est admissible au vu du pourcentage $p = \frac{8}{10} = 0,8$", de plus il est clair que $np_0 = 10 \times 0,5 = 5 \geq 05$.

2) Le calcul : Alors on peut appliquer la formule du l'écart-réduit suivante

$$\xi_{obs} = \frac{|p - p_0|}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{|0,80 - 0,50|}{\sqrt{\frac{0,50 \times 0,50}{10}}} = 2,37 < 1,96.$$

3) Conclusion : Alors la différence n'est pas significative, autrement dit que la maladie frappe autant les filles que les garçons au risque 05%.

Exemple N°: 05

Une race de souris présente des tumeurs spontanées avec un taux parfaitement connu, soit $p = 20\%$. Dans une expérience portant sur 100 souris, soumises à un certain traitement, on observe 34 cancers. On demande maintenant si la différence entre p_0 et p est significative.

Solution Exemple N°: 05

On peut appliquer la formule du l'écart-réduit suivante

$$\xi_{obs} = \frac{|p - p_0|}{\sqrt{\frac{pp_0}{n}}} = \frac{|0,34 - 0,20|}{\sqrt{\frac{0,80 \times 0,20}{100}}} = 3,50 > 1,96.$$

Donc le taux de souàris présentant des tumeurs spontanées dans l'échantillon diffère significativement du taux de la population.

Remarque : On peut traiter le problème de comparaison par le test du Khi-deux

Résultats	Cancer	Pas de cancer	Totale
Proportion théorique	20%	80%	1
Effectif calculé	20	80	100
Effectif observé	34	66	100

Maintenant on peut utiliser la formule du l'écart-réduit suivante

$$\begin{aligned}
\chi^2_{obs} &= \frac{(O_1 - np_1)^2}{np_1} + \frac{(O_2 - np_2)^2}{np_2} = \frac{(34-20)^2}{20} + \frac{(66-80)^2}{80} = 12,25. \\
\text{avec } O_1 + O_2 &= 34 + 66 = 100, \\
\text{de plus } p &= 0,20 \text{ et } q = 0,80., \quad k - 1 = 2 - 1 = 1, \\
\text{tel que } \chi^2_\alpha &= 3,841.
\end{aligned}$$

Alors il est clair que $\chi^2_{obs} > \chi^2_\alpha$.

2.1.3 Comparaison d'une moyenne observée a une moyenne calculée :

Soit X une variable aléatoire dont la moyenne est m et l'écart-type σ sont connus. Soit $n \geq 30$ (le cas de grands échantillons). Nous avons déjà vu que $\overline{X}$ suit la loi normale :

$$\overline{X} \backsimeq \mathcal{N}\left(m, \frac{\sigma_{obs}}{\sqrt{n}}\right).$$

La quantité $\mathcal{T} = \frac{\overline{X}-m}{\frac{\sigma_{obs}}{\sqrt{n}}}$ suit la loi normale d'espérance 0 et l'écart-type 1.

Pour un taux de risque de 05% on a :

Si $|\mathcal{T}| < 1,96$: Alors l'écart n'est pas significatif.

Remarque : Si σ est inconnue, pour cela on utilise l'estimation par $\frac{\sigma_{échantillon}}{\sqrt{n-1}}$.

Exemple N°: 06

On a prélevé un échantillon de 100 paquets de tabac dans la production d'une machine à paqueter. La mesure du poids de ces paquets a donné une moyenne de $36g$. On demande si la moyenne observée est compatible avec l'hypothèse que la machine fabrique « en moyenne » des paquets de 40 g avec un écart-type de 18 g au risque de 5%.

Solusion d'exemple N°: 06

1) Hypothèse nulle : H$_0$: (la moyenne calculé est compatible avec la moyenne observé).

2) Le calcul : Soit X une variable aléatoire dont la moyenne $m = 40$ et l'écart-type $\sigma = 18$ sont connus où X représente le poids des paquets, alors $\overline{X}_{100} \backsimeq \mathcal{N}\left(40, \frac{18}{\sqrt{100}}\right)$:

$$|\mathcal{T}| = \frac{\left|\overline{X} - m\right|}{\frac{\sigma_{obs}}{\sqrt{n}}} = \frac{|36 - 40|}{\frac{18}{\sqrt{100}}} = 2,22.$$

3) Conclusion : D'où $|\mathcal{T}| > 1,96$. On peut donc conclure que m diffère significativement 40 g au risque 5%.

2.2 Test de homogénéité :

Les tests d'homogénéité ou d'égalité destinés à comparer deux populations à l'aide d'un nombre équivalent d'échantillons sont les plus couramment utilisés. Dans ce cas la loi théorique du paramètre étudié (par exemple p, m, s^2) est inconnue au niveau des deux populations étudiées.

Position de problème :

Soient deux échantillons pris dans deux endroits différents. Peut-on considérer que ces deux échantillons proviennent de la même population ou de deux population différentes.

Le principe de la comparaison consiste à estimer qu'il n'y a pas de différence significative entre les deux populations dont sont issus ces deux échantillons ($\mathbf{H_0}$)

Alors d'après le résultat du test :

-Si $\mathbf{H_0}$ doit être rejetée, cela signifie que les deux populations sont différentes.

-Si au contraire $\mathbf{H_0}$ doit être accepter, il y a deux explications possibles.

1/ Les deux populations sont réellement différentes, mais la taille des échantillons est insuffisante pour pouvoir mettre cette différence en évidence.

2/ Les deux populations sont effectivement semblables pour le caractère étudié.

2.2.1 Test des comparaison des moyennes.

Etant donnés deux échantillons des tailles respectivement n_1 et n_2 de moyenne respectivement $\overline{X}_1 = \frac{1}{n_1}\sum_1^{n_1} X_1^1$ et $\overline{X}_2 = \frac{1}{n_2}\sum_1^{n_2} X_1^2$. Le problème consiste a comparer les moyennes de ces deux échantillons.

Doit-on attribuer au hasard la différence $D = \overline{X}_1 - \overline{X}_2$ des moyennes des deux échantillons, ou au contraire doit-on la considérer comme signification.

Pour cela on étudie d'abord l'intersection des intervalles de confiance pour m_1 et pour m_2 :

Le cas des grands échantillons n_1 et $n_2 \geq 30$:

1/ Hypothèse nulle $\mathbf{H}_0 =$ (Les deux échantillon proviennent d'une même population).

Si $\mathbf{H}_1 =$ (Les deux échantillon ne proviennent pas d'une même population c'est à dire que $m_1 \neq m_2$).

2/ Le calcul : On calcul la quatitée suivante

$$\xi = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

3/ Conclusion : Au seuil de sécurité 95%

" $\xi > 1,96$ (pratiquement 02) alors on rejette $\mathbf{H_0}$.

" $\xi < 1,96$ (pratiquement 02) alors on rejette $\mathbf{H_1}$.

Le cas des petits échantillons $\mathbf{n_1}$ et $\mathbf{n_2} < 30$:

1/ Hypothèse nulle : On pose l'hypothèse nulle :

$\mathbf{H_0}$= (Les deux échantillons proviennent d'une même population)

2) Le calcul : On montre qu'une bonne estimation de σ^2 est fournie par la quantité suivante appelée la variance commune :

$$\begin{aligned} S^2 &= \frac{\sum_1^{n_1}\left(X_i-\overline{X}_1\right)^2+\sum_1^{n_2}\left(X_i-\overline{X}_2\right)^2}{n_1+n_2-2} \\ &= \frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2-2}. \end{aligned}$$

Au lieu de l'expression de l'écart-réduit on utilise le critère de Student

$$\mathcal{D}=\frac{\left|\overline{X}_1-\overline{X}_2\right|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}.$$

3) Conclusion : On compare $\mathcal{D}$ avec t_α de d,d,l (n_1+n_2-2) :

1/Si $\mathcal{D} < t_\alpha$: La différence entre les deux échantillons n'est pas significative.

2/Si $\mathcal{D} > t_\alpha$: Les deux échantillons n'appartiennent pas à la même population.

Exemple N°: 08

Dans des études d'anesthésie, voulant comparer l'effet de deux somnifères, on a noté les durées de sommeil qui ont suivi les injections d'une dose bien définie. Les durées étant exprimées en minutes :

Somnifère 01	170	175	187	180	190	165	175	174	173	181		
Somnifère 02	155	160	164	150	160	159	154	156	160	167	153	158

Que peut-on dire pour cette comparaison ?

Solution d'Exemple N°: 08

1/ Hypothèse nulle : On pose l'hypothèse nulle :

$\mathbf{H_0}$= (Les deux échantillons proviennent d'une même population)

2) Indication des calculs :

$\overline{X}_1 = 177$, et $\overline{X}_2 = 158$, de plus n_1, $n_2 < 30$. Alors la variance commune

$$S^2=\frac{\sum_1^{n_1}\left(X_i-\overline{X}_1\right)^2+\sum_1^{n_2}\left(X_i-\overline{X}_2\right)^2}{n_1+n_2-2}=38,4.$$

Au lieu de l'expression de l'écart-réduit on utilise le critère de Student

$$\mathcal{D}=\frac{\left|\overline{X}_1-\overline{X}_2\right|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{|177-158|}{\sqrt{38,4}\sqrt{\frac{1}{10}+\frac{1}{12}}}=7,2.$$

3) Conclusion : On compare $\mathcal{D}$ avec $t_{0,05}$ de d,d,l $(10+12-2)$ égale à $2,086$, donc il est évident que $\mathcal{D} > t_{\alpha}$: d'où les deux échantillons n'appartiennent pas à la même population, On peut conclure que les deux somnifères ont des effets réellement différents ; le premier provoquant des sommeils de plus longue durée que le deuxième.

Exemple N°: 09

Pour mettre en évidence l'effet éventuel de l'absorption d'un médicament sur le rythme cardiaque, on forme deux groupes, de 100 sujets chacun, par tirage au sort parmi les malades traités par ce médicament :

Au premier groupe, on n'administre pas le médicament, mais un placebo ; Au deuxième groupe on administre le médicament. Les moyennes et variances estimées sur chacun des groupes sont :

$$\begin{aligned} m_y &= 80 \quad s_y^2 = 5 \quad \text{Pour le rythme cardiaque } Y \text{ du groupe témoin,} \\ m_{y'} &= 81 \quad s_{y'}^2 = 3 \quad \text{Pour le rythme cardiaque } Y' \text{ du groupe traité.} \end{aligned}$$

Faire le test bilatéral de H_0 $(EY = EY')$ contre H_1 $(EY \neq EY')$:avec un degré de signification 5%.

Solution d'exemple N°: 09

On pose m_1 la moyenne de la population 01 et m_2 la moyenne de la population 02

1) Hypothèse nulle : H_0 : { Les deux populations sont homogènes}

2) On calcul de taux d'ecart reduit : On utilise l'expression de l'écart-réduit

$$\xi = \frac{|80-81|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{|80-81|}{\sqrt{\frac{5}{100}+\frac{3}{100}}} = 3,943.$$

3) Conclusion : On compare ξ avec $1,96$, donc il est évident que $\xi > 1,96$: d'où les deux échantillons n'appartiennent pas à la même population.

Exemple N°: 10

Un chercheur a fait l'étude sur deux échantillons de souris qu'il a capturé en deux endroits différents. Il a obtenu les résultats suivants :

Echantillon 01	Echantillon 02
$n_1 = 50$	$n_2 = 50$
$\overline{x}_1 = 51g$	$\overline{x}_2 = 48g$
$\sigma^2_{ech_1} = 256$	$\sigma^2_{ech2} = 144$

Ces souris peuvent-elles appartenir à la même population au seuil de confiance de 95% ?.

Solution d'exemple N°: 10

On pose m_1 la moyenne de la population 01 et m_2 la moyenne de la population 02, alors l'application numérique nous donne :

On procède donc au test de l'écart-réduit :

1) L'hypothèse nulle : H$_0$: $m_1 = m_2$.

2) Le calcul : La valeur de l'écart-réduit est :

$$\begin{aligned} \xi &= \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{|51 - 48|}{\sqrt{\frac{256}{50} + \frac{144}{50}}} = 1,060. \end{aligned}$$

3/ Conclusion : Il est clair que $\xi > 1,96$. Donc on rejette H0 ; il y a donc une différence significative entre les deux moyennes. Par conséquent ces souris appartiennent à deux populations différentes.

Exemple N°: 11

Une étude est réalisee en vue de comparer l'efficacité de deux fertilisants sur la croissance des plantes. On mesure la hanteur de deux lots de plantes, chacun avec un fertilisant différent. Bien sûr, nous avons cultivé la même espèce dans des conditions environnementales identiques (ensoleillement, apports d'eau, température...). Les données relevées sont les suivantes :

Fertilisant I		Fertilisant II	
48,2	52,0	52,3	58,0
54,6	55,2	57,4	59,8
58,3	49,1	55,6	54,8
47,8	49,9	53,2	
51,4	52,6	51,3	

Nous désirons savoir s'il existe une différence significative entre les deux types de fertilisants, à un seuil de signification de 5 %.

Solution d'exemple N°: 11

On pose m_1 la moyenne de la population 01 et m_2 la moyenne de la population 02

1) Hypothèse nulle : H$_0$: { Les deux populations sont homogènes}

2) Le calcul :

Pour l'échantillon I : $n_1 = 10, \quad \overline{X}_1 = 51,910, \quad \sigma_1 = 3,370.$

Pour l'échantillon II : $n_2 = 08, \quad \overline{X}_2 = 55,300, \quad \sigma_2 = 2,969.$

On calcul tous d'abord la variance pondurée

$$\begin{aligned} S^2 &= \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2} = \frac{10 \times (3,370)^2 + 8 \times (2,969)^2}{10 + 8 - 2} \\ &= 11,505 \end{aligned}$$

Alors test de Student nous donne

$$\mathcal{D} = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|51,910 - 55,300|}{\sqrt{11,505}\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2,832.$$

3) Conclusion : On compare $\mathcal{D}$ avec $t_{0,05}$ de d,d,l $(10 + 8 - 2) = 2,120$, donc il est évident que $\mathcal{D} > t_\alpha$: d'où les deux échantillons n'appartiennent pas à la même population, alors on accepte H_0, et il y'a l'homogénéité entre les deux fertilisants.

Exemple N°: 12

Le pH (degré d'acidité) a été mesuré dans deux types de solutions chimiques A et B. Dans la solution A, six mesures ont été faites, avec un pH moyen de $7,52$ et un écart-type estimé de $0,024$. Dans la solution B, cinq mesures ont été faites, avec un pH moyen de $7,49$ et un écart-type estimé de $0,032$.

Déterminer si, au seuil de signification de $0,05$, les deux solutions ont des pH différents.

Solution d'exemple N°: 12

On pose m_1 la moyenne de la population 01 et m_2 la moyenne de la population 02

1) Hypothèse nulle : H_0 : $\{$ Les deux populations sont homogènes$\}$.

2) Le calcul : Pour l'échantillon I : $n_1 = 06, \quad \overline{X}_1 = 7,520, \quad \sigma_1 = 0,024.$

Pour l'échantillon I : $n_2 = 05, \quad \overline{X}_2 = 7,490, \quad \sigma_2 = 0,032.$

On calcul l'écart de Student.

$$S^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2} = 0,00095$$

$$\begin{aligned}\mathcal{D} &= \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\\ &= \frac{|7,520-7,490|}{\sqrt{0,00095}\sqrt{\frac{1}{6}+\frac{1}{5}}} = 0,166\end{aligned}$$

3) Conclusion : On compare $\mathcal{D}$ avec t_α de d,d,l (n_1+n_2-2), avec $t_\alpha = t_9^{0,05} = 2,262$. On accepte H_0 : les deux solutions ont même ph significativement.

Chapitre 3

Exercices sur Tests des hypothèses

$\boxed{\textbf{Exercice N°: 01}}$

Partant de races pures, un sélectionneur a croisé des mufliers ivoires avec des mufliers rouges. Il a obtenu en F1 des mufliers pâles puis en F2 ; après autofécondation des plantes de la génération F1 : 22 mufliers rouges, 52 mufliers pâles, et 23 mufliers ivoires.

La couleur des fleurs est-elle gérée par un couple d'allèles ?

$\boxed{\textbf{Exercice N°: 02}}$

On admet que la coloration de l'iris, chez l'homme, est déterminée par un couple d'allèles. La diversité des gènes complique l'étude de la transmission de ce caractère ; on sait cependant que la coloration bleue est récessive.

Le père de Monsieur Dupont et le père de Madame Dupont ont les yeux bleus. Monsieur et Madame Dupont n'ont pas les yeux bleus ; étant hétérozygotes, s'ils attendent un enfant, la probabilité pour qu'il ait les yeux bleus est $1/4$. Sur cinq enfant, le nombre d'enfants aux yeux bleus qu'il peuvent avoir obéit à une loi binomiale $B\left(5, \frac{1}{4}\right)$.

On a classé selon le nombre d'enfants aux yeux bleus qu'elles contiennent 1024 familles de 5 enfants et dont les

parents ont le même génotype que Monsieur et Madame Dupont.

Soit Y le nombre d'enfants aux yeux bleus d'une telle famille.

On se propose de tester $H_0 : Y \backsim B\left(5, \frac{1}{4}\right)$ contre $H_1 : Y \nsim B\left(5, \frac{1}{4}\right)$

Nombre d'enfants aux yeux bleus	0	1	2	3	4	5
Probabilité (sous H_0)	$\frac{243}{1024}$	$\frac{405}{1024}$	$\frac{270}{1024}$	$\frac{90}{1024}$	$\frac{15}{1024}$	$\frac{1}{1024}$
Effectif théorique	243	405	270	90	15	1
Nombre observé de familles	252	410	265	87	10	0

Le tableau ci-dessus résume l'ensemble des résultats expérimentaux et théorique.

Répondre au question posée.

Exercice N°: 03

Un éleveur de poulets possède deux races de coqs génétiquement distinctes : A et B. Afin de savoir s'il est plus avantageux pour lui d'utiliser comme reproducteurs des coqs de l'une ou de l'autre race. Il sépare un lot de 72 poules en deux lots de 36, accouple les 36 poules du premier lot avec le coq de la race A et les poules du second lot avec le coq de la race B. L'un des poulets né de chaque accouplement est pesé à l'âge de 8 semaines (ce poulet est choisi par tirage au sort parmi ceux de la même couvée). Les résultats observées sont données dans le tableau ci-dessous :

	Coq de la race A	Coq de la race B
Nombre de poulets...	36	36
Somme des poids des poulets(Gramme)	27 720	25 200
Variance observé des poids des poulets(Gramme)2	1 880	2 120

Pour savoir s'il existe une différence entre les résultats obtenus avec les deux coqs, on est conduit à mettre en œuvre un test d'hypothèse :

1/ Préciser les hypothèses en présence (hypothèse nulle, hypothèse alternative) l'écart utilisé pour faire le test et se distribution pour hypothèse nulle.

2/ Montrer que la différence des résultats obtenus avec les deux coqs est significative.

Exercice N°: 04

Onze volontaires ont accepté de suivre un traitement qui peut éventuellement modifier la viscosité sanguine. Les résultats avant et après traitement sont les suivants :

Individu	Valeur avant traitement	Valeur après traitement
1	2,40	2,45
2	2,60	2,55
3	2,55	2,55
4	2,85	2,40
5	3,15	2,85
6	3,15	2,90
7	2,15	2,00
8	2,70	2,40
9	2,75	2,60
10	2,45	2,40
11	2,65	2,30

Les viscosités avant et après traitement diffèrent-elles statistiquement ? Le traitement a-t'il eu un effet ? (On fixe le risque à 5 %).

Exercice N°: 05

Soit le croissement de deux souches de drosophiles différentes par trois caractères (a, b, c). On montre, en génétique mendélienne que si ces trois caractères sont portés par trois paires de chromosomes différentes et si l'on a "a^+" dominant par rapport à "a" et "b^+" dominant par rapport à "b" et "c^+" dominant par rapport à "c" on obtient, en théorie, dans le cas général, les proportions indiquées dans le tableau ci-dessous.

En fait on a obtenu sur 383 drosophiles examinées les résultats reportés dans le tableau :

Phénotypes	Proportions théoriques	Effectifs observés
(a^+, b^+, c^+)	27/64	142
(a^+, b^+, c)	9/64	74
(a^+, b, c^+)	9/64	49
(a, b^+, c^+)	9/64	43
(a, b^+, c)	3/64	28
(a, b, c^+)	3/64	24
(a^+, b, c)	3/64	13
(a, b, c)	1/64	10

1 / Quel test choisit ?

2/ Quelle est l'objectif pour cette expérience ?.

3 / Calculer les effectifs théoriques de drosophiles de chaque phénotype.

4 / Formuler les hypothèses H_0 et H_1.

5 / Interpréter les résultats du test .

6 / Quelle en est la conclusion ?

Exercice N°: 06 **Masse de sachets de médicaments.**

Afin de contrôler un lot de fabrication d'un médicament divisé en sachets, on a prélevé un échantillon aléatoire de 15 sachets que l'on a pesés. 1°/ Comparer, au risque a = 5 % et a = 1 %, la masse moyenne du lot à la valeur donnée par la norme de fabrication : 1, 50 g :

— dans le cas où l'hypothèse alternative est :

" la masse moyenne du lot est différente de 1, 50 g "

— dans le cas où l'hypothèse alternative est :

" la masse moyenne du lot est supérieure à 1, 50 g "

N.B. La somme observée des masses des sachets est de 23, 25 grammes et la somme de leurs carrés est 36, 1690.

2°/ Reprendre la question précédente dans le cas où la fabrication est considérée comme normale et où l'écart-type de la masse d'un sachet vaut 0, 095 gramme. On se placera dans le cas d'un test bilatéral et dans le cas d'un test unilatéral. On calculera, dans chaque cas, le risque de dire que la masse moyenne du lot correspond à la norme du fabricant, alors qu'elle en diffère, en réalité, de 0,1 gramme.

Exercice N°: 07 **Capacité vitale et vapeurs nocives.**

On a mesuré la capacité vitale de 100 sujets sains : on a observé parmi eux 26 sujets ayant une capacité vitale inférieure à 4,15 litres. Sur 100 sujets exposés pendant plus de 5 ans à des vapeurs nocives, on a observé 40 sujets ayant une capacité vitale supérieure ou égale à 4,15 litres. La proportion de sujets ayant une capacité vitale inférieure à 4,15 litres diffère-t-elle de façon significative chez les sujets exposés et chez les sujets sains pour un risque a fixé à 5 % ?

Bibliographie

[1] G. Saporta ; Probabilités, analyse des données et statistique. Technip, Paris, 1990.

[2] B. Escofier, J. Pagés ; Analyse factorielles simple et multiple. Dunod, Paris, 1998.

[3] Avner Bar-Hen : Cours de DEUG Probabilités et Statistiques, Université Aix-Marseille III, 2002–2003.

[4] Admane, O., Hoang Ky., Ouakli N : Statistique (cours et exercices) Pour les étudiants du tranc commun bio-médical., OPU 1998.

[5] M Vilain : Méthodes expérimentales en agronomie- Pratique et analyse. Edition Tec et Doc. 1999.