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 05 :

Le 31/05/2020

Par
Dr : CHALA ADEL

BioStatistiques

2019-2020

Je dédie ce travail. . . . .

A mes parents ils m'ont tous,
avec leurs moyens, soutenu et donné
la force d'aller toujours
plus loin.

Table des matières

Chapitre 1

La théorie du test statistique

1.1 Test de conformité :

Les tests de conformité 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.

1.1.1 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 conforme. La vérification de la conformité 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 conformité (ou concordance) entre la répartition théorique et la répartition expérimentale"

2/ On calcule la quantité suivante :

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

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

b) 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.

Remarque importante : Le test du χ^2 ne peut être utilisé que si tous les effectifs calculés sont suffisamment grands :

1.1.2 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 $\xi_{obs} \leq \xi_\alpha$ la différence n'est pas significative.

*Si $\xi_{obs} > \xi_\alpha$ la différence est significative.

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

$$\chi^2_{obs} = \frac{(O_1 - np_1)^2}{np_1} + \frac{(O_2 - np_2)^2}{np_2} = \frac{(34-20)^2}{20} + \frac{(66-80)^2}{80} = 12,25.$$

avec $O_1 + O_2 = 34 + 66 = 100$,

de plus $p = 0,20$ et $q = 0,80.$, $k - 1 = 2 - 1 = 1$,

tel que $\chi^2_\alpha = 3,841$.

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

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

Soit X une variable aléatoire dont la moyenne 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 on utilise l'estimation par $\frac{\sigma_{\acute{e}chantillon}}{\sqrt{n-1}}$.

1.2 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 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.

1.2.1 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 :

$$m_1 \in \left[\overline{X}_1 - u_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}, \overline{X}_1 + u_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}\right],$$

$$m_2 \in \left[\overline{X}_2 - u_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}, \overline{X}_2 + u_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}\right].$$

Le cas des petits échantillons :

$$m_1 \in \left[\overline{X}_1 - t_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}, \overline{X}_1 + t_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}\right],$$

$$m_2 \in \left[\overline{X}_2 - t_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}, \overline{X}_2 + t_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}\right].$$

Avec t_α valeur lue sur la table de Student de $(n-1)$ degré de liberté.

Alors ces trois cas peuvent se présenter :

1) $I_c(m_1) \cap I_c(m_2) = \Phi$: Dans ce cas on conclure qu'il y a une différence significative entre les deux moyennes.

2) $I_c(m_1) \cap I_c(m_2) \neq \Phi$, avec $\overline{X}_1 \in I_c(m_2)$ et $\overline{X}_2 \in I_c(m_1)$: Dans ce cas on peut conclure que la différence entre les moyennes des deux populations n'est pas significative au taux de risque considéré.

3) $I_c(m_1) \cap I_c(m_2) \neq \Phi$, avec $\overline{X}_1 \notin I_c(m_2)$ et $\overline{X}_2 \notin I_c(m_1)$: Dans ce cas pour pouvoir conclure si les moyennes des deux populations sont sembles ou non, on procède au test de comparaison de moyennes (Test de l'écart-réduit).

Test de comparaison des moyennes :

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

1/ Si $\mathbf{H}_0 =$ (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/ On calcule :

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

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 $\mathbf{H_1}$.

Solution :

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

On pose l'hypothèse nulle :

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

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

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

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 :

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 :

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

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

Conclusion : On compare $\mathcal{D}$ avec $t_{0{,}05}$ de d,d,l $(10+12-2) = 2,09$, 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.La théorie du test statistique

1.3 Test de conformité :

Les tests de conformité 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.

1.3.1 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 conforme. La vérification de la conformité 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 conformité (ou concordance) entre la répartition théorique et la répartition expérimentale"

2/ On calcule la quantité suivante :

$$\chi^2_{obs} = \sum_1^k \frac{(O_i - C_i)^2}{C_i},$$

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

b) 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.

Remarque importante : Le test du χ^2 ne peut être utilisé que si tous les effectifs calculés sont suffisamment grands :

1.3.2 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 $\xi_{obs} \leq \xi_\alpha$ la différence n'est pas significative.

*Si $\xi_{obs} > \xi_\alpha$ la différence est significative.

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

$$\chi^2_{obs} = \frac{(O_1 - np_1)^2}{np_1} + \frac{(O_2 - np_2)^2}{np_2} = \frac{(34-20)^2}{20} + \frac{(66-80)^2}{80} = 12,25.$$

avec $O_1 + O_2 = 34 + 66 = 100,$

de plus $p = 0,20$ et $q = 0,80.,\ \ k - 1 = 2 - 1 = 1,$

tel que $\chi^2_\alpha = 3,841.$

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

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

Soit X une variable aléatoire dont la moyenne 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 :

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

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 on utilise l'estimation par $\frac{\sigma_{échantillon}}{\sqrt{n-1}}$.

1.4 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 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.

1.4.1 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 :

$$m_1 \in \left[\overline{X}_1 - u_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}, \overline{X}_1 + u_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}\right],$$

$$m_2 \in \left[\overline{X}_2 - u_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}, \overline{X}_2 + u_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}\right].$$

Le cas des petits échantillons :

$$m_1 \in \left[\overline{X}_1 - t_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}, \overline{X}_1 + t_\alpha \frac{\sigma_1}{\sqrt{n_1 - 1}}\right],$$

$$m_2 \in \left[\overline{X}_2 - t_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}, \overline{X}_2 + t_\alpha \frac{\sigma_2}{\sqrt{n_2 - 1}}\right].$$

Avec t_α valeur lue sur la table de Student de $(n-1)$ degré de liberté.

Alors ces trois cas peuvent se présenter :

1) $I_c(m_1) \cap I_c(m_2) = \Phi$: Dans ce cas on conclure qu'il y a une différence significative entre les deux moyennes.

2) $I_c(m_1) \cap I_c(m_2) \neq \Phi$, avec $\overline{X}_1 \in I_c(m_2)$ et $\overline{X}_2 \in I_c(m_1)$: Dans ce cas on peut conclure que la différence entre les moyennes des deux populations n'est pas significative au taux de risque considéré.

3) $I_c(m_1) \cap I_c(m_2) \neq \Phi$, avec $\overline{X}_1 \notin I_c(m_2)$ et $\overline{X}_2 \notin I_c(m_1)$: Dans ce cas pour pouvoir conclure si les moyennes des deux populations sont sembles ou non, on procède au test de comparaison de moyennes (Test de l'écart-réduit).

Test de comparaison des moyennes :

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

1/ Si $\mathbf{H}_0$ = (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/ On calcule :

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

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 $\mathbf{H_1}$.

Solution :

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

On pose l'hypothèse nulle :

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

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

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

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 :

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 :

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

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

Conclusion : On compare $\mathcal{D}$ avec $t_{0,05}$ de d,d,l $(10 + 12 - 2) = 2,09$, 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.