- FICHE n° 2 Tests paramétriques à deux échantillons et tests du chi-deux

On prendra pour tous les tests un risque $\alpha = 5\%$ lorsqu'il n'est pas précisé. Lorsque le test est significatif, on calculera son degré de signification.

Exercice 1
J'aimerai vérifier qu'un des deux sujet du DST du 10 octobre n'a pas été significativement mieux réussi que l'autre. Faites un test pour essayer de montrer au risque 5%, que les deux sujets on été réussi de manière différentes (donc pour montrer que l'espérance de la note des sujets A diffère de celle des sujets B). On a un première échantillon $X_1, \ldots, X_{n_1}$ de taille $n_1 = 21$ d'étudiants qui ont passé le sujet A et un deuxième échantillon $X_1, \ldots, X_{n_1}$ de taille $n_2 = 21$ d'étudiants qui ont passé le sujet B. On suppose les X_i iid avec $X_i \mathcal{N}(\mu_1, \sigma^2)$, les Y_i iid $Y_i \mathcal{N}(\mu_2, \sigma^2)$ et les X_i et Y_i indépendants. (On notera que la variance des X_i σ^2 est supposée être la même que celle des Y_i). On donne :

$$\bar{x} = 13.9\,;\ \bar{y} = 12.6\,;\ \sum_{i=1}^{n_1} x_i^2 = 4494.812,\,;\ \sum_{j=1}^{n_2} y_j^2 = 3832.125.$$

Pour aller plus loin : calculer la $p - valeur$.

Exercice 2
On étudie l'activité d'un enzyme sérique, en fonction de différents facteurs dans l'espèce humaine. Les résultats sont exprimés en unités internationales par litre de sérum. Chez deux groupes de femmes, enceintes $(y_j)_{1 \le j \le 13}$ ou non $(x_i)_{1 \le i \le 15}$, on obtient les résultats suivants :

non enceinte	1,5	1,6	1,4	2,9	2,2	1,8	2,7	1,9	2,2	2,8	2,1	1,8	3,7	1,8	2,1
enceinte	4,2	5,5	4,6	5,4	3,9	5,4	2,7	3,9	4,1	4,1	4,6	3,9	3,5		

On admettra l'hypothèse de la normalité et d'égalité des variances de la variable ≪ activité de l'enzyme ≫ dans les deux populations. La grossesse a-t-elle une influence significative sur l'activité de l'enzyme ?

1. Proposer une méthode de travail permettant de répondre à cette question ; préciser toutes les hypothèses requises.
2. Mettre en oeuvre cette méthode et conclure pour un risque $\alpha = 1\%$. On donne $\sum_i x_i = 32{,}5$; $\sum_i x_i^2 = 75{,}83$; $\sum_j y_j = 55{,}8$; $\sum_j y_j^2 = 247{,}32$.

Exercice 3
Dans une pisciculture (branche de l'aquaculture qui étudie l'élevage des poissons), on désire comparer l'effet de deux régimes alimentaires (régime A et régime B) sur la croissance d'une espèce de poisson. Pour cela on prélève de manière indépendante deux lots de poissons. Le premier noté $X_1, \ldots, X_{n_1}$ de taille $n_1 = 180$ issu du régime A et le second noté $Y_1, \ldots, Y_{n_2}$ de taille $n_2 = 100$ issu du régime B. On mesure ensuite la longueur des poissons sur les deux échantillons. Les résultats obtenus sont les suivants :

$$\bar{x} = 21\,;\ \bar{y} = 21{,}4\,;\ \sum_{i=1}^{n_1} x_i^2 = 80\,000\,;\ \sum_{j=1}^{n_2} y_j^2 = 46\,000.$$

Une différence de régime alimentaire affecte-t-elle significativement la croissance des poissons au seuil $\alpha = 5\%$?

Exercice 4
Dans le cadre d'études sur la qualité sanitaire des laits, on veut comparer la teneur d'un pesticide, le lindane, dans les laits biologiques (LAIBIO) et les laits non biologiques dits conventionnels (LAICO). Dans ce but, des échantillons de deux types de laits ont été envoyés à un laboratoire d'analyses. Les résultats observés (en ppb) sont les suivants :

N^0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
LAICO	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,2	0,2	0,2	0,3	0,2	0,3
LAIBIO	0	0	0	0	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1		

Teneur en lindane dans les laits biologiques et conventionnels

Nous considérons dans la suite que les échantillons sont gaussiens et que les variances sont égales au sein des deux échantillons.

1. Faire un test bilatéral de comparaison des teneurs moyennes en lindane (au niveau 5%).
2. Est-ce que l'hypothèse d'égalité des variances vous semble effectivement vérifiée ici ?

Exercice 5

Dans un échantillon de 300 personnes, prélevé dans la population d'une ville A, il y en a 36 qui fument au moins deux paquets de cigarettes par jour. Dans une autre ville B et pour un échantillon de 100 personnes, on trouve une proportion de 8% de personnes qui fument au moins deux paquets de cigarettes par jour. On veut tester (H_0) :≪ il n'y a aucune différence entre les deux villes ≫ contre (H_1) :≪ il y a un plus grand pourcentage de personnes qui fument au moins deux paquets de cigarettes par jour dans la ville A que dans la ville B ≫.

1. Quelles sont les variables qui modélisent le problème (donner leur loi) ? On les note X et Y. On note également p_A (resp. p_B) désigne la probabilité qu'un individu de la ville A (resp. B) fume au moins deux paquets de cigarettes par jour.
2. Quel test allez-vous faire ?
3. Quelles statistiques utilisez-vous pour estimer p_A et p_B ?
4. Faire le test de (H_0) contre (H_1).
5. Que vaut la $p-$valeur ?

Exercice 6

Un nouveau vaccin contre le paludisme est expérimenté auprès de la population d'une ville Africaine. On choisit deux échantillons A et B de 200 personnes chacun. On injecte le vaccin aux individus de l'échantillon A et un placebo à ceux de l'échantillon B. Au bout d'un an on constate que 40 personnes de l'échantillon A ont des accès de paludisme et 80 de l'échantillon B. On veut montrer que le vaccin contre le paludisme est efficace, faire un test au risque 5%.

Exercice 7

Un entomologiste (l'entomologie est la branche de la zoologie qui étudie les insectes) qui étudie les araignées lynx vertes affirme que le thorax des araignées femelles semble en moyenne plus long que celui des araignées mâles. On suppose que le longueur du thorax des araignées femelles suit une loi d'espérance μ_F et d'écart-type σ_F et que la longueur du thorax des araignées mâles suit une loi d'espérance μ_M et d'écart-type σ_M. Deux échantillons aléatoires indépendants de tailles respectives $n_1 = 61$ (pour les femelles) et $n_2 = 72$ (pour les mâles) sont collectés et ont conduit aux résultats suivants :

$$\bar{x}_F = 6{,}1\,;\ s_F^2 = 2{,}6 \ \text{ et } \ \bar{x}_M = 5{,}4\,;\ s_M^2 = 1{,}82.$$

1. Au risque $\alpha = 5\%$, l'affirmation de l'entomologiste est-elle vérifée ?
2. L'entomologiste a encore conservé les araignées durant 3 mois et il a observé que 51 des 61 femelles et 58 des 72 mâles sont mortes. Peut-on conclure que le taux de mortalité des femelles est plus élevé que celui des mâles au risque $\alpha = 5\%$?

Exercice 8

Afin d'étudier s'il existe une liaison entre le port de la ceinture de sécurité et le degré de gravité des blessures, la Sécurité Routière a prélevé un échantillon aléatoire de 10 779 conducteurs ayant subi un accident. Les résultats sont les suivants :

Nature des blessures	Port de la ceinture	Pas de ceinture
Fatales	1	43
Graves	4	98
Blessures sérieuses	25	330
Pas ou peu de blessures	1229	9049

Le port de la ceinture a t-il une influence significative sur la nature des blessures ? (on effectuera le test au niveau $\alpha = 10\%$)

Exercice 9
On souhaite comparer les résultats de l'évolution d'une maladie M, à la suite de l'emploi de l'un ou l'autre des traitements A et B. Pour cela on a relevé l'évolution de la maladie sur un échantillon de malades traités avec le traitement A et un échantillon de malades traités avec le traitement B. Les valeurs possibles de l'évolution de la maladie d'un malade sont réparties en 4 classes : guérison, amélioration, état stationnaire, décès. On a obtenu la répartition suivante sur les deux échantillons de malades traités avec A et B :

	Guérison	Amélioration	État stationnaire	Décès
A	280	209	110	55
B	220	90	88	40

1. Estimer la probabilité que l'évolution de la maladie d'un malade soit la guérison, l'amélioration, l'état stationnaire ou le décès sous l'hypothèse que la répartition de l'évolution de la maladie est la même avec les deux traitements A et B.
2. Peut-on considérer que l'évolution de la maladie est la même avec chacun des deux traitements A et B ? (on fera le test au niveau $\alpha = 10\%$).

Exercice 10
On désire savoir si un dosage sanguin dans une population bien définie est différent chez les sujets atteints d'une certaine maladie. On pratique à cet effet des dosages sur 30 malades et 50 témoins. On supposera que les deux échantillons sont indépendants. Les résultats sont les suivants :

Dosage (centre de classes)	Nombre de témoins	Nombre de malades
[5,15[	1	0
[15,25[	2	0
[25,35[	5	2
[35,45[	10	3
[45,55[	20	5
[55,65[	5	10
[65,75[	4	7
[75,85[	2	2
[85,95[	1	1

1. Répondre à la question « Les sujets atteints de la maladie ont-ils un dosage significativement différent de celui des témoins ? » en effectuant un test de comparaison d'espérances.
2. On répartit désormais les valeurs de la variable quantitative « dosage » en deux classes : dosage strictement inférieur à 45 kg et dosage supérieur à 45 kg.
   (a) Répondre à la question 1. en effectuant un test du χ^2 d'indépendance.
   (b) Comparer les résultats obtenus avec ceux du test de comparaison d'espérances. Commenter.
3. Répondre à la même question en effectuant un test de comparaison de proportions. Comparer les résultats obtenus avec ceux du test du χ^2 d'indépendance.