

Statistiques décisionnelles : TD 10

Pour cette feuille de TD, j'ai été piocher dans les cours de statistiques des biologistes à l'ENS de Lyon entre 2000 et 2010. Ces exercices sont volontairement "mal" formulés et vous devrez faire preuve d'autonomie.

1 Tests Bonus

J'introduis ici deux nouveaux tests dont je parlerai en cours. Ce sont des tests classiques qui font partie du bagage en base en statistique.

Test de comparaison des proportions : Si on souhaite comparer deux proportions avec de grand effectif (au moins 50 dans chaque groupe), on peut appliquer un test asymptotique. On a deux échantillons indépendants (X_1, \dots, X_n) et (Y_1, \dots, Y_m) où les $X_i \sim \text{Ber}(p_1)$ et $Y_j \sim \text{Ber}(p_2)$. On veut tester $H_0 : p_1 = p_2$ contre $H_1 : p_1 \leq p_2$ ou $p_1 < p_2$ ou $p_1 > p_2$. Dans ce cas, on fait l'approximation

$$S = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n} + \frac{\bar{Y}_m(1-\bar{Y}_m)}{m}}} \sim \mathcal{N}(0, 1)$$

et on construit la zone de rejet à partir des quantiles de la loi normale.

Test d'indépendance de Pearson Si on dispose de deux échantillons appariés (X_1, \dots, X_n) et (Y_1, \dots, Y_n) , tels que (X_1, Y_1) forme un vecteur gaussien. Alors on peut construire un test pour l'indépendance de X_1 et Y_1 avec

$$H_0 : X_1 \perp\!\!\!\perp Y_1 \quad \text{contre} \quad H_1 : X_1 \text{ et } Y_1 \text{ ne sont pas indépendants.}$$

La corrélation empirique

$$R = \frac{\text{Cov}_n(X, Y)}{\sqrt{V_n^X V_n^Y}} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2 \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$$

satisfait, sous $H_0 : D = \sqrt{n-1} \frac{R}{\sqrt{1-R^2}} \sim t(n-2)$ tandis qu'elle prend de grandes valeurs sous H_1 (positives ou négatives). On peut faire des versions unilatérales de ce test on se demandant si les variables X_1 et Y_1 sont corrélées positivement ou négativement. Les seuils de rejet sont des quantiles de lois de Student.

2 Exercices

Exercice 1. On a mesuré la longueur de l'aile et le poids de 10 drosophiles mâles (L en 10^{-2} mm et P en 10^{-5} g).

L	261	266	268	273	265	269	270	267	259	271
P	155	158	142	145	147	164	143	156	144	162

Conclusion ?

Exercice 2. L'enthalpie ou chaleur latente de la fonte de glace a été mesurée par deux méthodes différentes (cal/g) (Natrella 1963)

A	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05
suite A	80.03	80.02	80	80.02					
B	80.02	79.94	79.98	79.97	80.03	79.95	79.97		

Existe-t-il une différence entre ces méthodes ?

Exercice 3. On veut comparer l'action de 2 antibiotiques A et B en faisant une expérience sur des souris atteintes d'une maladie mortelle sans traitement connu. On administre ces antibiotiques respectivement à deux groupes de 200 souris chacun. Dans le groupe soumis à l'antibiotique A, 10% des souris survivent alors que dans l'autre groupe, 20% des souris survivent.

Peut-on en conclure quelque chose ?

Exercice 4. Sur 100 étudiantes et étudiants ayant suivi les séances de TD au cours de l'année, le nombre de reçus à l'examen a été de 75. Sur 100 étudiantes et étudiants n'ayant pas été en séances de TD, le nombre de reçus a été de 53.

Que pouvez-vous en conclure ?

Exercice 5. Dans une classe de 24 élèves, on range sans les mesurer les enfants par taille.

GFFFFFGGFGFFFFGGFGFFFFGFGF

Conclusion ?

Exercice 6. Dans certaines races de vers (*Bombyx mori*) les œufs sont de couleurs différentes (jaune ou brun) selon leur sexe. On relève les sexes des œufs au fur et à mesure de la ponte d'une femelle (40 œufs au total) :

MFFMMMFMFFFMMFMFFMFFMFMFMFMFFMFMFMFMFFM

Le rang de ponte est-il lié au sexe ?

Exercice 7. Le nombre de découvertes scientifiques a été dénombré chaque année entre 1860 et 1959.

Nombre de découvertes	0	1	2	3	4	5	6	7	8	9	10	12
Nombre d'années	9	12	26	20	12	6	6	4	1	1	1	1

Les grandes découvertes scientifiques se produisent-elles de manière aléatoire au cours du temps ?

Exercice 8. Le groupe sanguin de chaque personne a été déterminé dans trois échantillons : le premier provenant de France, le second de Roumanie et le troisième du Proche-Orient.

	France	Roumanie	Proche-Orient
A	54	45	33
B	14	14	34
AB	6	8	12
O	51	31	33

Que pouvez-vous conclure de ces données ? Avez-vous une piste d'explication ?