Université d'Aix-Marseille 2013–2014

Mathématiques pour la Biologie II

FORMULAIRE.

Notations. Dans ce formulaire, les estimateurs (liés à l'échantillon) sont notées avec des minuscules :
– Moyenne empirique : $\bar{x}$,
– Écart-type empirique modifié : s.
Les paramètres liées à la population globale sont notés avec des lettres grecques :
– Moyenne : μ,
– Écart-type : σ (et variance σ^2).
Les Variables Aléatoires liées aux échantillons virtuels sont notées avec des majuscules :
– Moyenne empirique : $\bar{X}$,
– Écart-type empirique modifié : S.
S'il y a deux échantillons, on ajoute des indices 1 et 2.

Conditions préalables.

Les tests demandent toujours que les échantillons soient bien fabriqués (choix des sujets fait de manière indépendante et représentative), et s'il y a plusieurs échantillons, qu'ils soient "indépendants" entre eux.

Les tests sur les moyennes et variances sont valables sous des hypothèses de normalité : il faut supposer que la (ou les) variables considérées sur la (ou les) populations globales ont une distribution normale.

Les tests sur les fréquences et les tests du χ^2 demandent des échantillons suffisamment grand ($n \geq 30$).

I- Tests sur les moyennes.

Comparaison d'une moyenne μ à une valeur donnée μ_0.

S'il y a égalité, la statistique de test et sa loi sont :

$$Z = \sqrt{n}\frac{\bar{X} - \mu_0}{S} \sim \mathcal{S}_{n-1}, \quad \text{Student à } n-1 \text{ d.d.l.}$$

Comparaison de deux moyennes, avec écarts-type théoriques égaux.

– On définit la variance empirique modifié de la réunion des deux échantillons :

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2},$$

– Si les moyennes sont égales, la statistique de test et sa loi sont :

$$Z = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\frac{(\bar{X}_1 - \bar{X}_2)}{S} \sim \mathcal{S}_{n_1+n_2-2}$$

– Si $n_1 + n_2 - 2 \geq 30$, on peut remplacer $\mathcal{S}_{n_1+n_2-2}$ par la loi $\mathcal{N}(0,1)$.

Comparaison de moyennes ; cas des échantillons appariés. On utilise ce test quand on veut comparer les moyennes de deux grandeurs mesurées sur un même échantillon. Attention, il faut connaître les données complètes de l'expérience, c-à-d les résultats x_i et y_i pour chaque individu. Il suffit alors :
– de calculer les écarts $d_i = x_i - y_i$ pour chaque individus,
– d'appliquer aux différences d_i le test de comparaison de moyenne à une valeur théorique donnée (souvent 0).

II- Test sur les variances (ou écart-types).

Comparaison d'un écart-type à une valeur σ donnée.

S'il y a égalité, la statistique de test et sa loi sont :

$$Z = (n-1)\frac{S'^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \text{la loi du } \chi^2 \text{à } (n-1) \text{ d.d.l. .}$$

Comparaison de deux variances (ou écarts-type)

Si les deux variances sont égales, la statistique de test et sa loi sont :

$$Z = \frac{S'^2_x}{S'^2_y} \sim \mathcal{F}_{(n_x-1, n_y-1)}, \text{ la loi de Fisher-Snedecor de paramètre } (n_x-1, n_y-1).$$

III- Tests sur les proportions.

Comparaison d'une proportion à une valeur donnée.

- Conditions supplémentaires : $nf \geq 5$, $n(1-f) \geq 5$,
- S'il y a bien égalité, la statistique de test, et sa loi (approximative)

$$\sqrt{n}\frac{\bar{F} - f}{\sqrt{f(1-f)}} \sim \mathcal{N}(0,1).$$

Comparaison de deux proportions (Test d'égalité).

- Conditions supplémentaires : $n_1\bar{F}_1 \geq 5$, $n_1(1-\bar{F}_1) \geq 5$, et idem pour le second échantillon.
- On pose $F = \dfrac{n_1\bar{F}_1 + n_2\bar{F}_2}{n_1+n_2}$.
- Si les proportions sont égales (à F), la statistique de test et sa loi (approximative) sont alors

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\frac{\bar{F}_1 - \bar{F}_2}{\sqrt{F(1-F)}} \sim \mathcal{N}(0,1).$$

IV- Tests du χ^2.

Test du χ^2 d'adéquation.

- On dispose des effectifs observés N_i^{obs} pour $i = 1, \dots, k$.
- On fabrique grâce aux données de l'énoncé les effectifs *théoriques* (ou attendus) N_i^{th} pour $i = 1, \dots, k$.
- On vérifie les conditions $N = \sum N_i^{obs} = \sum N_i^{th} \geq 30$ et tous les $N_i^{th} \geq 5$.
- S'il y a adéquation, la statistique de test et sa loi (approximative) sont alors

$$Z = \sum_{i=1}^{k} \frac{(N_i - N_i^{th})^2}{N_i^{th}} \sim \chi^2_{k-1}, \quad \text{la loi du } \chi^2 \text{ à } (k-1) \text{ d.d.l. .}$$

Test du χ^2 d'indépendance.

- On dispose du tableau des effectifs observés $N_{i,j}^{obs}$, avec k^l lignes et k^c colonnes.
- On fabrique grâce aux données de l'énoncé les effectifs *théoriques* (ou attendus) $N_{i,j}^{th}$

$$N_{i,j}^{th} = \frac{N_i^c\, N_j^l}{N}, \quad \text{avec } N^c \text{ (resp } N^l\text{) les totaux par colonnes (resp. lignes).}$$

- On vérifie les conditions $N = \sum N_{i,j}^{obs} = \sum N_{i,j}^{th} \geq 30$ et tous les $N_i^{th} \geq 5$.
- S'il y a indépendance, la statistique de test et sa loi (approximative) sont alors

$$Z = \sum_{i=1}^{k^l}\sum_{j=1}^{k^c} \frac{(N_{i,j} - N_{i,j}^{th})^2}{N_{i,j}^{th}} \sim \chi^2_{\nu}, \quad \text{la loi du } \chi^2 \text{ à } \nu = (k^c-1)(k^l-1) \text{ d.d.l. .}$$