# Formulaire pour les tests de conformité et d'homogénéité

Biostatistiques L2

Université Claude Bernard Lyon 1 – France

## 1 Tests de conformité – comparaison à une valeur théorique

### 1.1 Tests de conformité de la moyenne

On étudie une variable quantitative $X$ et on cherche à savoir si les observations (un échantillon de taille $n$, de moyenne observée $\bar{x}$ et de variance observée $s^2$, provenant d'une population de moyenne $\mu$ et de variance $\sigma^2$) concordent avec une loi théorique de moyenne $\mu_0$. On a :

$$
\begin{aligned}
H_0 :& \mu = \mu_0 \\
H_1 :& \mu \neq \mu_0
\end{aligned}
$$

On supposera toujours que la variable $X$ est distribuée suivant une loi normale. Si ce n'est pas le cas, on aura recours aux statistiques non paramétriques (programme de L3). On peut répondre à la question posée soit en étudiant si $\mu_0$ est dans l'intervalle de confiance de $\mu$, qui fournit une plage de valeurs possibles pour $\mu$ avec un niveau de confiance associé, soit par un test statistique.

Les statistiques de tests et intervalles de confiance à employer sont résumés dans la table suivante :

| Variance $\sigma^2$ | Effectif | Intervalle de confiance | Statistique | Seuil |
|---|---|---|---|---|
| Connue | quelconque | $\mu \in \left[\bar{x} \pm \epsilon_\alpha \sqrt{\frac{\sigma^2}{n}}\right]$ | $\epsilon_{obs} = \dfrac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ | $\epsilon_\alpha$ |
| Estimée : $\hat{\sigma}^2 = \frac{n}{n-1}s^2$ | $n < 30$ | $\mu \in \left[\bar{x} \pm t_{\alpha,n-1} \sqrt{\frac{\hat{\sigma}^2}{n}}\right]$ | $t_{obs} = \left\lvert \dfrac{\bar{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \right\rvert$ | $t_{\alpha,n-1}$ |
| | $n \geq 30$ | $\mu \in \left[\bar{x} \pm \epsilon_\alpha \sqrt{\frac{\hat{\sigma}^2}{n}}\right]$ | $\epsilon_{obs} = \left\lvert \dfrac{\bar{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \right\rvert$ | $t_{\alpha,n-1} \to \epsilon_\alpha$ |

## 1.2 Test de conformité d'une proportion

On étudie une variable $X$, qui représente le nombre $k$ de succès parmi $n$ tirages (par exemple, le nombre $k$ de filles sur $n = 100$ naissances). On note $\hat{p} = \frac{k}{n}$ et on veut vérifier si les observations concordent avec une loi théorique de probabilité de succès $p_0$ (par exemple, $p_0 = 0.5$ pour des naissances équilibrées). La loi théorique de $X$ est une loi binomiale de probabilité $p$ pour $n$ tirages ; on peut raisonner en étudiant si $p_0$ est dans l'intervalle de confiance de $p$, ou par un test.

La formule de l'intervalle de confiance autour de $p$, la proportion dans la population de laquelle vient l'échantillon, est :

$$p \in \left[ \hat{p} \pm \epsilon_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

Dans le cas du test, les hypothèses testées sont :

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

Afin de pouvoir utiliser une loi de référence standard, ce que la loi binomiale ne permet pas de faire, on utilise la convergence de la loi binomiale vers la loi normale, et il faut donc toujours que $n > 30$, avec $np_0 > 5$ et $n(1-p_0) > 5$. La statistique à calculer est alors :

$$\epsilon_{obs} = \left| \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right|,$$

et la valeur seuil est $\epsilon_\alpha$.

Attention, dans cette dernière formule, on raisonne sous l'hypothèse nulle du test, et les termes sous la racine sont des $p_0$ (supposés égaux à $p$ sous $H_0$) et pas des $\hat{p}$ comme dans l'intervalle de confiance.

# 2 Tests d'homogénéité – comparaison de deux valeurs observées

## 2.1 Tests d'homogénéité de deux moyennes

On étudie deux variables $X$ et $Y$ et on cherche à savoir si les observations (deux échantillon de taille $n_X$ et $n_Y$, de moyennes observées $\bar{x}$ et $\bar{y}$ et de variances observées $s_X^2$ et $s_Y^2$, issues respectivement d'une population de moyenne $\mu_X$ et de variance $\sigma_X^2$ et d'une population de moyenne $\mu_Y$ et de variance $\sigma_Y^2$), proviennent de la même loi théorique de moyenne $\mu$. Les hypothèse sont donc :

$$
\begin{aligned}
H_0 :& \mu_X = \mu_Y \\
H_1 :& \mu_X \neq \mu_Y
\end{aligned}
$$

Il faut toujours que les observations dans les deux échantillons soient indépendantes. Les statistiques des tests sont résumés dans la table suivante :

<table>
<tr><th>Variances $\sigma_X^2$ et $\sigma_Y^2$</th><th>Effectifs</th><th>Homoscédasticité</th><th>Statistique</th><th>Seuil</th></tr>
<tr><td>Connues</td><td>quelconques</td><td>non nécessaire</td><td>$\epsilon_{obs} = \left| \dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \right|$</td><td>$\epsilon_\alpha$</td></tr>
<tr><td rowspan="3">Estimées<br>(test de Fisher,<br>voir plus bas)</td><td>$n_X \geq 30$<br>et $n_Y \geq 30$</td><td>non nécessaire</td><td>$\epsilon_{obs} = \left| \dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{\sigma}_X^2}{n_X} + \frac{\hat{\sigma}_Y^2}{n_Y}}} \right|$</td><td>$t_{\alpha, n_X+n_Y-2} \to \epsilon_\alpha$</td></tr>
<tr><td rowspan="2">$n_X < 30$<br>ou $n_Y < 30$</td><td>$\sigma_X^2 = \sigma_Y^2$ et<br>$\hat{\sigma}^2 = \frac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y - 2}$</td><td>$t_{obs} = \left| \dfrac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n_X} + \frac{1}{n_Y} \right)}} \right|$</td><td>$t_{\alpha, n_X+n_Y-2}$</td></tr>
<tr><td>$\sigma_X^2 \neq \sigma_Y^2$</td><td>statistiques non paramétriques</td><td></td></tr>
</table>

**Le test de Fisher** sert à vérifier l'égalité de deux variances (homoscédascticité). Les hypothèses du test sont :

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

La statistique du test est :

$$F_{obs} = \frac{\frac{n_X s_X^2}{n_X - 1}}{\frac{n_Y s_Y^2}{n_Y - 1}},$$

si $s_X^2 > s_Y^2$, et l'inverse sinon. La valeur seuil de ce test est $F_{\alpha, n_X - 1, n_Y - 1}$.

## 2.2 Tests d'homogénéité de deux proportions

On étudie deux variables $X$ et $Y$, et on cherche à savoir si les observations ($k_X$ succès parmi $n_X$ tirages et $k_Y$ succès parmi $n_Y$ tirages, avec $\hat{p}_X = \frac{k_X}{n_X}$ et $\hat{p}_Y = \frac{k_Y}{n_Y}$) proviennent de la même loi théorique, une loi binomiale de probabilité de succès $p$. Ici aussi, on utilise la convergence de la loi binomiale vers la loi normale. On calcule d'abord la probabilité moyenne pour estimer la variance de la loi sous-jacente :

$$\hat{p} = \frac{k_X + k_Y}{n_X + n_Y}$$

Ensuite on vérifie que l'on peut utiliser la convergence de la loi binomiale vers la loi normale. Il faut donc toujours que :

- $n_X > 30$ et $n_Y > 30$
- $n_X \hat{p} > 5, n_Y \hat{p} > 5$
- $n_X(1 - \hat{p}) > 5, n_Y(1 - \hat{p}) > 5$

La statistique est alors :

$$\epsilon_{obs} = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}\,(1 - \hat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}},$$

et la valeur seuil est $\epsilon_\alpha$.