L2 – Sociologie - Statistiques
Contrôle 1 - Octobre 2017–1018 - Corrigé

Exercice 1. On mène une enquête sur les troubles de la vue liés à l'activité professionnelle. Dans le cadre de cette enquête, on veut estimer la proportion p d'employés de l'administration qui portent des lunettes. Sur un échantillon de 900 employés (hommes ou femmes) de l'administration, on observe que 581 portent des lunettes. En déduire une estimation de la proportion p avec une confiance de 95%.

Il s'agit d'un exercice d'*estimation de proportion* (comme il est dit dans l'énoncé), *ce n'est pas un test d'ajustement*. On ne doit donc pas procéder à la formulation de deux hypothèses. La méthode d'estimation de proportion est résumée à la page 13 du formulaire, il suffit de la suivre point par point.

1. On calcule la proportion expérimentale : $p_{\mathrm{e}} = \frac{581}{900} = 0,645$ (et donc $q_{\mathrm{e}} = 1 - 0.645 = 0,355$).
2. On s'assure que l'échantillon est assez grand, pour pouvoir procéder à l'estimation : on a $n = 900 > 30$, $p_{\mathrm{e}} \times n = 581 > 5$, $q_{\mathrm{e}} \times n = 319 > 5$.
3. Puique on demande de travailler avec une confiance de 95%, on pose $\alpha = 1 - 0.95 = 0.05$. On détermine z_α tel que $\varphi(z_\alpha) = \frac{1-\alpha}{2}$. Ce nombre z_α est donné au bas de la page 4 du formulaire. On trouve $z_\alpha = 1.96$.
4. On en déduit $a_\alpha = z_\alpha \sqrt{\frac{p_{\mathrm{e}} q_{\mathrm{e}}}{n}} = 1.96 \times \sqrt{\frac{581 \times 319}{900}} = 0.0312$. Ainsi $p \in I_\alpha = [0.645 - 0.0312, 0.645 + 0.031] = [0.614, 0.677]$.

Exercice 2. Dans le cadre de la même enquête, on étudie l'âge moyen auquel les salariés obtiennent leur premier arrêt de travail. On étudie un échantillon de 150 employés du secteur tertiaire et de 280 employés du secteur agricole et industriel. On obtient les résultats suivants :

Age du premier arrêt de travail	$[20,30[$	$[30,40[$	$[40,50[$	$[50,60[$
Effectifs (tertiaire)	7	9	30	104
Effectifs (agricole et industriel)	30	40	70	140

Il s'agit cette fois d'un exercice d'*estimation de moyenne*. La marche à suivre est résumée en page 14 du formulaire.

1. Estimer l'âge moyen du premier arrêt de travail des employés secteur tertiaire avec une confiance de 95%.
   A l'aide d'un calcul direct sur la ligne du tableau qui donne les effectifs du secteur tertiaire, on obtient : $m_{\mathrm{e}} = 50.4$, $s_{\mathrm{e}} = 8.052$. Afin de pouvoir procéder à l'estimation de moyenne, on s'assure qu'on a bien un grand échantillon. On a $n = 150 > 30$.
   Avec $\alpha = 0.05$, on a à nouveau $z_\alpha = 1.96$, et donc $a_\alpha = z_\alpha \frac{s_{\mathrm{e}}}{\sqrt{n-1}} = 1.96 \times \frac{8.052}{\sqrt{149}} = 1.293$.
   Donc $I_\alpha^{\mathrm{ter}} = [50.4 - 1.293, 50.4 + 1.293] = [49,107, 51.693]$.
2. Estimer l'âge moyen du premier arrêt de travail des employés secteur agricole et industriel avec une confiance de 95%.
   A l'aide d'un calcul direct sur la ligne du tableau qui donne les effectifs du secteur agricole et industriel, on obtient : $m_{\mathrm{e}} = 46.429$, $s_{\mathrm{e}} = 10.252$. Puisque $n = 280 > 30$, on a bien un grand échantillon.
   Avec $\alpha = 0.05$, on a comme précédemment $z_\alpha = 1.96$ et $a_\alpha = 1.96 \times \frac{10.252}{\sqrt{279}} = 1.203$.
   Donc $I_\alpha^{\mathrm{agr}} = [46.429 - 1.203, 46.429 + 1.203] = [45.226, 47, 632]$.
3. Peut-on en déduire que l'âge moyen du premier arrêt de travail des employés du secteur agricole et industriel est inférieur à celui des employés du secteur tertiaire ?
   Comme le deux intervalles ne se chevauchent pas, on peut estimer que la moyenne d'âge du premier arrêt de travail dans le tertiaire est supérieure à la moyenne d'âge du premier arrêt de travail dans le secteur agricole et industriel.

Exercice 3. Les résultats d'études précédentes indiquaient que 45% des adultes français étaient satisfaits de l'enseignement public. Une nouvelle étude, menée sur un échantillon de 900 adultes, observe que 315 d'entre eux sont satisfaits l'enseignement public.

1. Peut-on conclure de cette nouvelle étude, à l'aide d'un test d'ajustement de proportion, que les adultes français sont moins satisfaits de l'enseignement public qu'ils ne l'étaient auparavant ?
   Comme il est dit dans l'énoncé, on procède cette fois à un *test d'ajustement de proportion*.
   (a) Formulation des hypothèses.
   H_0 : les adultes français sont aussi satisfaits de l'enseignement public qu'ils l'étaient avant la nouvelle étude.

H_1 : ils sont moins satisfaits qu'avant (il s'agit donc d'un test unilatéral).
Afin de formuler ces hypothèses en termes statistiques, on désigne par p la proportion (inconnue) d'adultes français satisfaits de l'enseignement public. On a alors :

$$H_0 : p = p_0 = 0.45; \;\; H_1 : p < p_0.$$

(b) Statistique du test.
On désigne par P_n la proportion de français satisfaits de l'enseignement public sur un échantillon *aléatoire* de $n = 900$ individus. Il s'agit d'un grand échantillon. Enn effet : $n = 900 > 30$, $p_0 \times n = 0.45 \times 900 = 405 > 5$, et $q_0 \times n = 0.55 \times 900 = 495 > 5$. Sous l'hypothèse nulle, on a donc :

$$P_n \hookrightarrow \mathcal{N}\left(p_0, \sqrt{\frac{p_0 q_0}{n}}\right), \text{ c'est à dire } P_n \hookrightarrow \mathcal{N}\left(0.45, 0.017\right).$$

(c) Détermination de la région critique.
On se donne le niveau d'erreur $\alpha = 0.05 = 5\%$. La région critique (ou région de rejet) est de la forme $K_\alpha = \{P_n \leq p_\alpha\}$, avec $\mathbb{P}\left[K_\alpha; \mathcal{N}\left(0.45, 0.017\right)\right] = 0.05$. On trouve, en utilisant l'*inverse de la loi normale :* $p_\alpha = 0.422$ et donc $K_\alpha = \{P_n \leq 0.422\}$.

(d) Décision du test.
On observe que $p_e = \frac{315}{900} = 0.35 \in K_\alpha$. Donc, avec un niveau d'erreur de 5%, on accepte l'hypothèse H_1 : les adultes français sont moins satisfaits de l'enseignement public qu'ils ne l'étaient auparavant.

2. Calculer la puissance $\eta\left(0.33\right)$. On rappelle que cette puissance est la probabilité d'accepter l'hypothèse H_1 si la proportion réelle d'adultes insatisfaits dans la population est de 33%. On calcule donc la probabilité de la région critique K_α, mais avec une loi normale dans laquelle on a remplacée la proportion théorique p_0 par la proportion réelle 0.33.
On a ainsi $\eta\left(0.33\right) = \mathbb{P}\left[P_n \leq 0.422; \mathcal{N}\left(0.33, \sqrt{\frac{0.33 \times (1-0.33)}{900}}\right)\right] \sim 1$. Si la proportion réelle d'adultes insatisfaits est de 33%, le test a 100% de chances d'accepter (à juste titre) l'hypothèse H_1.