TD de statistiques. Tests d'homogénéité- Corrigés

Exercice 1 (Exemple page). Le gérant d'un grand bar parisien s'approvisionne depuis longtemps chez le même fournisseur pour une certaine boisson gazeuse, livrée dans des bouteilles d'une contenance normale de 33cl. Or, différentes remarques de clients l'amènent à se poser des questions sur la fiabilité de son fournisseur. En effet, si celui-ci affirme livrer des bouteilles qui contiennent en moyenne 33cl de boisson, avec un écart-type de $0,5$cl, le gérant, après avoir testé 100 bouteilles, parvient à une contenance moyenne de 32,7cl. De plus, en discutant avec un voisin, autre gérant de bar, il apprend que celui-ci reçoit des bouteilles contenant en moyenne 32,9cl d'une boisson tout à fait comparable (résultat trouvé après l'examen d'un lot de 100 bouteilles livrées par un fournisseur qui donne un écart-type de 0,4cl pour sa fabrication).

La différence de contenance des bouteilles livrées aux deux gérants est-elle le signe d'un moins bon service du premier fournisseur, ou est-elle simplement imputable au hasard de l'échantillonnage ?

Corrigé On étudie deux populations, les bouteilles fournies par les fournisseurs des deux bars. La variable étudiée est quantitative, c'est la contenance des bouteilles. Notons μ_1 la contenance moyenne des bouteilles du premier fournisseur, $\sigma_1 = 0,5$cl l'écart-type de cette contenance. L'échantillon prélevé par le premier barman a une taille $n_1 = 100$, et la contenance moyenne sur l'échantillon de ces bouteilles vaut $m_1 = 32,7$cl.

Pour le fournisseur du deuxième bar, on note μ_2 la contenance moyenne des bouteilles, $\sigma_2 = 0,4$cl, et l'échantillon examiné a une taille $n_2 = 100$ et une contenanc moyenne $m_2 = 32,9$.

On va tester l'hypothèse que les deux fournisseurs livrent des produits équivalents. (On aurait pu tester l'hypothèse que $\mu_1 = 33$cl ce qui a déjà été fait plus tot).

Etape 1 : L'hypothèse H_0 est : $\mu_1 = \mu_2$ soit encore $\mu_1 - \mu_2 = 0$. L'hypothèse alternative est $\mu_1 \neq \mu_2$

Etape 2 : on choisit un seuil de signification de $1 - \alpha = 95\%$ et un seuil de risque $\alpha = 5\%$.

Etape 3 : **Rappelons qu'on fait les calculs en supposant H_0 vraie.** les échantillons sont grands ($n_1 = n_2 = 100 > 30$) et les écarts-type σ_1 et σ_2 supposés connus. Donc la statistique théorique, encore appelée valeur seuil, vaut $S_{th} = Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$. La statistique calculée vaut

$$S_{cal} = \frac{m_1 - m_2}{\sqrt{\frac{(\sigma_1)^2}{n_2} + \frac{(\sigma_2)^2}{n_2}}} = -\frac{0,2}{\sqrt{\frac{0,25}{100} + \frac{0,16}{100}}} = -\frac{2}{\sqrt{41}} \simeq -3,12$$

Etape 4 : Décision. On a $|S_{cal}| = +3,12 > 1,96 = S_{th}$ donc on rejette H_0 au seuil de risque de 5%. On a 5% de chances de se tromper (erreur de première espèce).

Exercice 2 (Premier exercice d'entrainement). Une chaine de restaurants possède deux restaurants dans des villes comparables A et B. Le restaurant de la ville A a dépensé une somme plus importante en publicité que celui de la ville B. La direction de la chaîne aimerait savoir si cette publicité a entraîné une fréquentation différente du restaurant de la ville A. Sur une durée de 3 mois (90 jours), le chiffre d'affaires quotidien fut en moyenne de 1500 euros dans la ville A, avec un écart-type de 110 euros, et en moyenne de 1400 euros dans la ville B, avec un écart-type de 100 euros. On suppose que les chiffres d'affaires réalisés dans les restaurants sont distribués normalement.

Corrigé On veut savoir si les chiffres d'affaires moyens des restaurants A et B sont comparables. Pour chacun des deux restaurants on étudie une variable quantitative, le chiffre d'affaire, sur une « population » statistique qui est l'ensemble des jours d'ouverture des restaurants. On voudrait vérifier si ces variables sont en moyenne les mêmes sur chaque échantillon.

Pour le premier restaurant on note μ_1 le chiffre d'affaires moyen (inconnu), σ_1 son écart-type (inconnu), $n_1 = 90$ la taille de l'échantillon de jours étudiés. Sur cet échantillon le chiffre d'affaires moyen vaut $m_1 = 1500$ euros avec un écart-type $s_1 = 110$ euros. Pour le deuxième restaurant, on note μ_2 le chiffre d'affaires moyen (inconnu), σ_2 son écart-type (inconnu), $n_2 = 90$ la taille de l'échantillon de jours étudiés. Sur cet échantillon le chiffre d'affaires moyen vaut $m_2 = 1400$ euros avec un écart-type $s_2 = 100$ euros.

Etape 1 : L'hypothèse H_0 est : $\mu_1 = \mu_2$ soit encore $\mu_1 - \mu_2 = 0$. L'hypothèse alternative est $\mu_1 \neq \mu_2$

Etape 2 : on choisit un seuil de signification de $1 - \alpha = 95\%$ et un seuil de risque $\alpha = 5\%$.

Etape 3 : **Rappelons qu'on fait les calculs en supposant H_0 vraie.** les échantillons sont grands ($n_1 = n_2 = 90 > 30$) et les écarts-type σ_1 et σ_2 sont inconnus. Donc la statistique théorique, encore appelée valeur seuil, vaut $S_{th} = Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$. La statistique calculée vaut

$$S_{cal} = \frac{m_1 - m_2}{\sqrt{\frac{(s_1)^2}{n_1-1} + \frac{(s_2)^2}{n_2-1}}} = \frac{100}{\sqrt{\frac{12100}{89} + \frac{10000}{89}}} = 100\sqrt{\frac{89}{22100}} = 6,35$$

Etape 4 : Décision. On a $|S_{cal}| = +6,35 > 1,96 = S_{th}$ donc on rejette H_0 au seuil de risque de 5%. On a 5% de chances de se tromper (erreur de première espèce).

Exercice 3 (Deuxième exercice d'entrainement)**.** Deux échantillons de taille 10 ont été extraits de deux populations normales de variances inconnues mais supposées égales. Dans le premier échantillon, on a calculé pour un certain caractère une moyenne de 164 et un écart-type de 3,4, tandis que les mêmes calculs ont donné une moyenne de 170 et un écart-type de 2,8 dans le deuxième échantillon. Comparer les deux populations.

Corrigé On veut savoir si la variable quantitative étudiée a même moyenne dans chacune des deux populations. Dans la première population, la moyenne μ_1 de la variable est inconnue et son écart-type σ_1 aussi. L'échantillon a taille $n_1 = 10$, et on a trouvé dans cet échantillon une moyenne $m_1 = 164$ et un écart-type $s_1 = 3,4$. Dans la deuxième population, la moyenne μ_2 de la variable est inconnue et son écart-type σ_1 aussi mais $\sigma_1 = \sigma_2$. De plus, l'échantillon a taille $n_2 = 10$, et on a trouvé dans cet échantillon une moyenne $m_2 = 170$ et un écart-type $s_2 = 2,8$. Les deux échantillons sont petits, et les écarts-type inconnus mais supposés égaux, et la variable suit une loi normale.

Etape 1 : L'hypothèse H_0 est : $\mu_1 = \mu_2$ soit encore $\mu_1 - \mu_2 = 0$. L'hypothèse alternative est $\mu_1 \neq \mu_2$

Etape 2 : on choisit un seuil de signification de $1 - \alpha = 95\%$ et un seuil de risque $\alpha = 5\%$.

Etape 3 : **Rappelons qu'on fait les calculs en supposant H_0 vraie.** Les échantillons sont petits. les écarts-type inconnus mais supposés égaux, et la variable suit une loi normale. Donc la statistique théorique vaut $S_{th} = t_{\alpha}^{(n_1-1+n_2-1)}/2 = t_{0,975}^{(18)} = 2,1009$ (loi de Student à $n_1 + n_2 - 2 = 18$ degrés de liberté). La statistique calculée vaut

$$S_{cal} = \frac{m_1 - m_2}{\sqrt{\frac{n_1(s_1)^2+n_2(s_2)^2}{n_1+n_2-2}}} \times \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -\frac{6}{\sqrt{\frac{10}{18} \times ((3,4)^2 + (2,8)^2)}}\sqrt{\frac{10}{2}} = -4,09$$

Etape 4 : Décision. On a $|S_{cal}| = +4,09 > 2,10 = S_{th}$ donc on rejette H_0 au seuil de risque de 5%. On a 5% de chances de se tromper (erreur de première espèce).

Exercice 4 (Quatrième exercice d'entrainement)**.** Dans le but de tester l'impact de deux campagnes publicitaires différentes on isole deux groupes de personnes témoins : un groupe (1) qui visionnera uniquement la campagne télévisée et un groupe (2) auquel seront proposées la campagne télévisée et la campagne d'affichage (panneaux publicitaires). On s'intéresse au nombre de personnes qui achètent le produit à l'issue de l'étude. Dans le groupe 1, sur 30 personnes 10 achètent le produit. Dans le groupe 2, sur 40 personnes 17 achètent le produit.

Tester l'hypothèse que les fréquences d'achat sont égales au seuil de 5%.

Corrigé La variable étudiée est une variable qualitative : pour un individu donné, on regarde s'il achète ou non le produit. On s'intéresse donc à la proportion de gens qui achètent, parmi deux populations. La première population est celle des gens qui visionnent uniquement la campagne télé, tandis que la deuxième population est celle des gens qui voient la campagne télé et la campagne d'achat. Le premier échantillon a une taille $n_1 = 30$ et la proportion de gens qui achètent est $p_1 = \frac{10}{30} = 0,33$. Le deuxième échantillon a une taille $n_2 = 40$ et la proportion de gens qui achètent vaut $p_2 = \frac{17}{40}$.

Etape 1 : L'hypothèse H_0 est : $\pi_1 = \pi_2$ soit encore $\pi_1 - \pi_2 = 0$. L'hypothèse alternative est $\pi_1 \neq \pi_2$

Etape 2 : on choisit un seuil de signification de $1 - \alpha = 95\%$ et un seuil de risque $\alpha = 5\%$.

Etape 3 : **Rappelons qu'on fait les calculs en supposant H_0 vraie.** Les échantillons sont grands $n_1 = 30$ et $n_2 = 40$. La statistique théorique vaut $S_{th} = Z_{0,975} = 1,96$. La statistique calculée vaut

$$S_{cal} = \frac{p_1 - p_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

avec $f = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{10+17}{70} \simeq 0,3857$ d'où

$$S_{cal} = \frac{0,3333 - 0,425}{\sqrt{0,3857 \times 0,6143 \times (\frac{1}{30} + \frac{1}{40})}} = -0,78.$$

Étape 4 : $|S_{cal}| = 0,78 < 1,96 = S_{th}$ donc on ne peut pas rejeter H_0 au seuil de risque de 5%. On ne connait pas la probabilité de se tromper (erreur de deuxième espèce).