

Corrigé-type de Série N°4 de Statistiques Appliquées
Inférence statistique

Exercice 1

Lors d'une enquête sur la durée de sommeil des enfants de 2 à 3 ans dans une wilaya en Algérie, on a trouvé une moyenne du temps de sommeil par nuit égal à 10.2 heures dans un groupe de 40 enfants. L'écart-type s est de 2.1 heures. La moyenne attendue de sommeil est de 11.7 heures chez les enfants de cet âge.

Les enfants examinés dorment-ils autant que ceux de la population ? ($\alpha = 0.05$).

Solution

C'est un test de comparaison d'une moyenne expérimentale $\bar{x} = 10.2$ heures à une valeur théorique $\mu = 11.7$ heures qui est la moyenne connue de la population.

Grand échantillon $n = 40$; $\bar{x} = 10.2$ heures ; $\mu = 11.7$ heures ; Population non finie : N inconnue. On considère que c'est un tirage avec remise, car la taille de la population N est non finie.

Le problème est de savoir si l'échantillon est représentatif de la population (i.e. la différence constatée entre μ et \bar{x} est explicable par les fluctuations d'échantillonnage).

1^{ère} Etape : C'est le test bilatéral (égal contre inégal) qu'il faut utiliser ici (car on a dans les données le terme « autant que » qui veut dire la même chose).

Hypothèse nulle (H_0) : la moyenne observée ($\bar{x} = 10.2$ heures) du temps de sommeil est conforme à la moyenne de la population ($\mu = 11.7$ heures), ou encore que l'échantillon soit extrait au hasard de la population (c'est-à-dire qu'il n'y a pas de différence entre les moyennes, et on dit que les enfants du groupe dorment autant que ceux de la population). *Contre*

Hypothèse alternative (H_1) : Le temps moyen de sommeil observé n'est pas le même que celui de toute la population.

($\bar{x} = 10.2 \neq \mu = 11.7$ heures) : ce sera donc un test bilatéral.

2^{ème} Etape :

Le risque choisi par hypothèse est de $\alpha = 5\%$ et comme on a affaire à un grand échantillon, on considère la loi normale et la table n° 2 donne le critère seuil du test bilatéral qui est : $z_{0.05} = 1.96$

3^{ème} Etape :

Règle de Décision : On conserve H_0 au risque de 5% si la valeur de la statistique de test observée T_0 appartient à l'intervalle d'acceptation $IA_{0.05}$ de (H_0) du test bilatéral qui est :

$$IA_{0.05}(H_0) = \left[Z_{\frac{\alpha}{2}} ; Z_{(1-\frac{\alpha}{2})} \right] = [-1.96 ; +1.96]$$

4^{ème} Etape : Calcul de la statistique de test T_0 où $T_0 = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_p}$ où σ_p est estimé par s (écart-type corrigé), donc : $T_0 = T_{\text{cal}} = T_{\text{obs}} = \frac{\sqrt{40}(10.2 - 11.7)}{2.1} = -4.517539515 \cong -4.52$

5^{ème} Etape :

Décision statistique : Comme la statistique de test observée (calculée) n'appartient pas à l'intervalle d'acceptation de $H_0 : T_0 = -4.52 \notin [z_{0.025} ; z_{0.975}] = [-1.96 ; +1.96]$

On rejette H_0 (l'échantillon n'est pas conforme à la population).

Décision pratique : les enfants examinés ont une durée de sommeil différente de ceux de toute la population.

Si la question dans le même problème se pose de la manière suivante :

Question : Les enfants examinés ont-ils une durée de sommeil inférieure à la durée de sommeil des enfants de la population d'étude ? ($\alpha = 0.05$).

La réponse à cette question sera comme suit :

1^{ère} étape : choix des deux hypothèses ; $H_0 : \bar{x} = \mu$ contre $H_1 : \bar{x} < \mu$ test unilatéral gauche.

2^{ème} Etape :

Le risque choisi est $\alpha = 5\%$ et comme on a affaire à un grand échantillon, on considère la loi normale et la table n° 2 donne le critère seuil du test unilatéral : $z_{0.05} = -1.645$

3^{ème} Etape :

Règle de Décision : On conserve H_0 au risque de $\alpha = 5\%$ si la valeur de la statistique de test observée T_0 appartient à l'intervalle d'acceptation $IA_{0.05}(H_0)$ du test unilatéral qui est :

$$IA_{0.05}(H_0) = [z_{\alpha} ; +\infty[= [-1.645 ; +\infty[$$

4^{ème} Etape : Calcul de la statistique de test observée T_0 . C'est le même calcul que précédemment et $T_0 = -4.5175$

5^{ème} Etape :

Décision statistique : Comme la statistique de test observée (calculée) n'appartient pas à l'intervalle d'acceptation de $H_0 : T_0 = -4.518 \notin [z_{0.05} ; +\infty[= [-1.645 ; +\infty[$

On rejette H_0 (l'échantillon n'est pas conforme à la population).

Décision pratique : les enfants examinés ont une durée de sommeil inférieure à la durée de sommeil de ceux de toute la population.

Exercice 2

On teste deux traitements anti-cancéreux A et B sur deux populations de patients P_A et P_B (de même taille $n_A = n_B = 50$). L'efficacité d'un traitement est évaluée par l'éventuelle diminution de la taille de la lésion tumorale, estimée par l'imagerie médicale, après un an de traitement. Pour la population soumise au traitement A, on observe une diminution de la taille des tumeurs dans 27 cas sur 50 et pour le traitement B, dans 18 cas.

Peut-on conclure à une différence d'effet des deux traitements (au seuil de 5%) ?

Peut-on conclure que le traitement A est plus efficace que le traitement B ($\alpha = 5\%$) ?

Solution :

Question 1 :

- 1) Formulation des hypothèses : test bilatéral.
 $H_0 : \mathbf{p}_A = \mathbf{p}_B$ contre $H_1 : \mathbf{p}_A \neq \mathbf{p}_B$. (Egal contre inégal).
- 2) Le seuil critique : le risque $\alpha = 0.05$; la table de la loi normale n° 2 donne le critère du test bilatéral $z_{0.05} = 1.96$ et l'intervalle d'acceptation de H_0 est $[- 1.96 ; + 1.96]$
Si la statistique de test $T_o \in [- 1.96 ; + 1.96] \Rightarrow$ on ne peut pas rejeter H_0 .
Les différentes conditions d'application ($n * p > 5$ et $n * q > 5$) sont toutes vérifiées.
Les deux fréquences observées : $f_1 = \frac{27}{50} = 0.54$ et $f_2 = \frac{18}{50} = 0.36 \Rightarrow$
La fréquence commune sera $f_c = \frac{n_1 * f_1 + n_2 * f_2}{n_1 + n_2} = \frac{27 + 18}{50 + 50} = 0.45 \Rightarrow$
$$\sigma_c = \sqrt{f_c(1 - f_c) \left(\frac{1}{n_A} + \frac{1}{n_B} \right)} = \sqrt{0.45(1 - 0.45) \left(\frac{1}{50} + \frac{1}{50} \right)} = 0.099498743 \Rightarrow$$
- 3) La statistique de test observée : $T_o = \frac{f_1 - f_2}{\sigma_c} = \frac{0.54 - 0.36}{0.099498743} = 1.8090689 \cong 1.81$
- 4) Décision : comme la valeur de la statistique de test : $T_o = 1,81 \in [- 1.96 ; + 1.96] \Rightarrow$ on ne peut pas rejeter H_0 .
 \Rightarrow Pas de différence d'effet entre les 2 traitements anti-cancéreux.

Question 2 : test unilatéral droit

- 1) Formulation des hypothèses : $H_0 : \mathbf{p}_A = \mathbf{p}_B$ et $H_1 : \mathbf{p}_A > \mathbf{p}_B$ (Egal contre supérieur)
- 2) Le seuil critique : Le risque à considérer est $\alpha = 0.05$; la table n° 1 de la loi normale donne le critère seuil $u_{1-\alpha} = u_{0.95} = 1.645$
Si la statistique de test $\in]1.645 ; + \infty[\Rightarrow$ on rejette H_0 .
- 3) La statistique de test observée : $T_o = \frac{0.54 - 0.36}{0.099498743} = 1.8090689 \cong 1.809$
- 4) Décision : comme la valeur de la statistique de test :
 $T_o = 1,809 \in]1.645 ; + \infty[\Rightarrow$ on rejette H_0 .
 \Rightarrow Il y a une différence significative entre les 2 traitements anti-cancéreux de sorte que le traitement A s'est avéré le plus efficace.

Exercice 3

Des études en psychologie du développement ont montré qu'à l'âge de 12 mois, 50% des bébés « normaux » marchent. On souhaite mener une étude sur les retards de développement des bébés prématurés. On teste l'hypothèse que les bébés prématurés marchent plus tardivement que les bébés normaux.

On observe une population de 80 bébés prématurés. A 12 mois, 35 de ces 80 bébés marchent.

Faut-il réaliser un test unilatéral ou bilatéral ?

Peut-on, au seuil de signification de 5 %, valider l'hypothèse de recherche ?

Solution :

Comparaison d'une fréquence expérimentale $f = \frac{35}{80} = 0.4375$ à une proportion théorique $p = 0.5$ (Conformité d'une fréquence expérimentale à une proportion théorique).

On dispose d'un grand échantillon de taille $n = 80$ individus.

La proportion $p = 0.5$ et le produit $n * p = 80 * 0.5 = 40 > 5$; $n = 80 > 30$ on peut approximer la binomiale $B(n ; p) = B(80 ; 0.5)$ par la loi normale $N(n * p ; \sqrt{n * p * q})$.

Il faut utiliser un test unilatéral gauche : (égal contre inférieur)

Lire unilatéral car il s'agit de tester si les prématurés marchent plus tardivement.

Hypothèse nulle (H_0) : $f = p$ (la différence est non significative) ; **contre**

Hypothèse alternative (H_1) : $f < p$ (les prématurés marchent tardivement que les normaux).

On choisit un risque $\alpha = 5\%$ et de la table de la loi normale, on lit le critère : le seuil gauche est $z_{0.05} = -1.645$; la zone d'acceptation de H_0 est égale à l'intervalle $[-1.645 ; +\infty[$.

Règle de décision : On conserve H_0 au risque de 5% si la valeur de T_0 appartient à l'intervalle d'acceptation de (H_0).

$$\text{Calcul de la statistique de test } T_{\text{obs}} = T_0 = \frac{f - p}{\sqrt{\frac{p * q}{n}}} = \frac{0.4375 - 0.5}{\sqrt{\frac{0.5 * 0.5}{80}}} = -1.118034 \cong -1.118$$

Décision : Comme la valeur $T_{\text{obs}} = -1.118 \notin$ la région de rejet de $H_0 =]-\infty ; -1.645[$, alors on conserve donc H_0 (et on rejette H_1) avec un risque seuil de : $\alpha = 5\%$.

Exercice 4

On désire comparer la pression artérielle diastolique d'un groupe de sujets sains et d'un groupe de sujets atteints de drépanocytose (hémoglobinopathie). Voici les résultats obtenus :

	Effectif	Pression artérielle diastolique moyenne (mm Hg)	Variance s^2
Sujets sains (A)	88	70.1	10.8
Sujets drépanocytaires (B)	85	61.8	6.9

- 1) La pression artérielle est-elle différente chez les sujets drépanocytaires ?
- 2) La pression artérielle est-elle plus basse chez les sujets drépanocytaires ? ($\alpha = 0.01$)

Solution : Test d'homogénéité

Comparaison de deux moyennes μ_1 et μ_2 à partir de deux échantillons indépendants :

- 1) Si on doit déduire de ces indications que la pression artérielle du groupe A et celle du groupe B sont différentes, alors on utilise le **Test bilatéral** (égal contre inégal)
 - *Hypothèse nulle :* (H_0) : $\mu_1 = \mu_2$ pas de différence significative ; **contre**
 - *Hypothèse alternative :* (H_1) : $\mu_1 \neq \mu_2$ ou existence d'une différence significative.

Puisque les échantillons sont grands $n_1 = 88 > 30$ et $n_2 = 85 > 30$, on considère la loi normale.

Ce test est basé sur la statistique de test observé :

$$T_{\text{obs}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ qui suit une loi normale } N(0,1) \text{ sous } H_0.$$

Comme le risque n'est pas donné aux hypothèses, alors on choisit un risque universel de $\alpha = 5\%$; et de la table de l'écart-réduit de la loi normale on tire le critère seuil $z_{0,05} = 1.96$

La région critique de rejet de (H_0) du test bilatéral au risque $\alpha = 5\%$ est :

$$RC_{5\%} =] - \infty ; z_{0,025} [\cup] z_{0,975} ; + \infty [=] - \infty ; - 1.96 [\cup] + 1.96 ; + \infty [.$$

Calcul de la statistique de test observé $T_o = T_{obs}$ avec la formule suivante $\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

- L'écart type corrigé de X_1 dans P_1 est $s_1 = \sqrt{10.8}$ et $\bar{X}_1 = 70.1$; $n_1 = 88$
- L'écart type corrigé de X_2 dans P_2 est $s_2 = \sqrt{6.9}$ et $\bar{X}_2 = 61.8$; $n_2 = 85$

La valeur de la statistique de test T est : $t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{70.1 - 61.8}{\sqrt{\frac{10.8}{88} + \frac{6.9}{85}}} = 18.38084586 \cong 18.38$

Règle de décision :

Si cette valeur de la statistique observé T_{obs} , appartient à la région critique de rejet de H_0 , on conserve H_1 au risque 5 %.

Décision : Comme la valeur T_{obs} appartient à la zone de rejet de H_0 , on ne peut accepter l'hypothèse nulle de l'homogénéité des sujets sains et des sujets drépanocytaires pour ce qui est de la moyenne de la pression artérielle (c'est-à-dire on est confiant à 95% que les pressions artérielles diastoliques des deux groupes sont significativement différentes).

- 2) Si on a à déduire de ces indications que la pression artérielle diastolique est plus basse chez les sujets drépanocytaires, donc **Test unilatéral** droit (égal contre inégal)
- Hypothèse nulle $H_0 : \mu_1 = \mu_2$ pas de différence significative ; **contre**
 - Hypothèse alternative $H_1 : \mu_1 > \mu_2$ la pression chez les drépanocytaires est plus basse.

Avec un risque de $\alpha = 1\%$ donné aux hypothèses, la table de l'écart-réduit de la loi normale nous donne le critère seuil $z_{0,99} = 2.362$

La région critique de rejet de (H_0) du test unilatéral au risque $\alpha = 1\%$ est :

$$RC_{1\%} =] z_{0,99} ; + \infty [=] + 2.362 ; + \infty [.$$

- Comme la statistique de test $T_o = 18.38$ appartient à la région critique $RC_{1\%}$ alors on ne peut pas accepter l'hypothèse nulle.
- **Décision pratique :** la pression artérielle chez les drépanocytaires est plus basse.