UNIVERSITE DE BATNA_2 **FACULTE DE MEDECINE**

Corrigé type : SERIE DE TD N° 8 *en BIOSTATISTIQUE 2019/2020*

Exercices sur l'estimation

Exercice 1

L'indice de masse corporelle est calculé en divisant le poids d'une personne par le carré de sa taille, et est utilisé comme mesure du seuil dans laquelle la personne n'est pas en surpoids. Supposons que la distribution de l'indice de masse corporelle pour les hommes ait un écart-type de $\sigma = 3$ kg/m^2, et nous souhaitons estimer la moyenne μ en utilisant un échantillon de taille $n = 49$.

Trouvez la probabilité que l'erreur absolue dans cette estimation ne dépasserait pas 1 kg/m2.

Solution

C'est un grand échantillon de taille $n = 49 > 30$, donc c'est la loi normale qui sera utilisée.

L'intervalle de confiance de la moyenne μ au risque $\alpha = 0.05$ est donné par la formule :

$[\bar{x} - u_\alpha \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_\alpha \frac{\sigma}{\sqrt{n}}]$ avec $\sigma = 3$ kg/m² et $n = 49$. Où le symbole $u_\alpha = \bar{z}_\alpha$ du cours.

L'erreur absolue maximale dans l'estimation de μ par $\bar{x}$ est égale à $u_\alpha \frac{\sigma}{\sqrt{n}}$.

Pour que cette erreur absolue ne dépasse pas 1 kg/m², on doit avoir $u_\alpha \frac{\sigma}{\sqrt{n}} \leq 1$.

$u_\alpha \frac{\sigma}{\sqrt{n}} = 1 \Rightarrow u_\alpha = z_{1-\frac{\alpha}{2}} = \frac{\sqrt{n}}{\sigma} = \frac{\sqrt{49}}{3} = 2.33$;

En utilisant les tables n° 1, ou n° 2 on trouve le risque $\alpha = 0.02$ pour la valeur de 2.33

La probabilité que l'erreur absolue ne dépasserait pas 1 kg/m² est $\mathbf{1 - \alpha = 0.98}$

Exercice 2

Les blessures auto déclarées parmi les gauchers et les droitiers ont été comparées dans une enquête auprès de 1 900 étudiants. 90 étudiants parmi les 180 gauchers, et 645 étudiants parmi les 1720 droitiers ont signalé au moins une blessure au cours de la même période.

Calculez l'intervalle de confiance à 95% pour la proportion d'élèves ayant subi au moins une blessure, pour chacune des deux sous-populations d'élèves gauchers et droitiers.

Solution

Les deux échantillons sont grands et donc c'est la loi normale qu'on utilisera.

I) La sous-population des gauchers : la fréquence auto déclarée chez les gauchers est estimée ponctuellement par $\mathbf{f = \frac{90}{180} = 0.5}$ et estimée par intervalle de confiance (95 %)

$IC_{0.95} = [f \pm u_\alpha \frac{\sqrt{f(1-f)}}{\sqrt{n}}] = [0.5 \pm 1.96 \frac{\sqrt{0.5 * 0.5}}{\sqrt{180}}] = [0.5 \pm 0.073044887] \Rightarrow$

$IC_{0.95} = [0.426955112 ; 0.573044887] \approx$ **[0.427 ; 0.573]**

II) La sous-population des droitiers : la fréquence auto déclarée chez les droitiers est estimée ponctuellement par $f = \frac{645}{1720} = 0.375$ et estimée par intervalle de confiance.

$IC_{0.95} = [f \pm u_\alpha \frac{\sqrt{f\,(1-f)}}{\sqrt{n}}] = [0.375 \pm 1.96 \frac{\sqrt{(0.375)(0.625)}}{\sqrt{1720}}] = [0.375 \pm 0.02287955] \Rightarrow$

$IC_{0.95} = [0.352120449\ ;\ 0.39787955] \approx$ **[0.352 ; 0.398]**

Où les symboles f et u_α représentent respectivement les symboles $\tilde{p}$ et $\bar{z}_\alpha$ du cours.

Les conditions d'applicabilité de ces formules sont satisfaites : $n * p > 5$ et $n\,p\,(1 - p) > 5$.

Exercice 3

Le poids à la naissance, obtenu à partir des accouchements sur une longue période de temps dans un certain hôpital, montre une moyenne $\mu = 3175$ g et un écart type $\sigma = 584$ g.

Calculer la probabilité que le poids moyen m à la naissance d'un échantillon de 35 nourrissons se situe entre 3033 et 3317 g.

Solution

L'intervalle de pari (ou de fluctuation) pour la moyenne $\bar{x}$ de l'échantillon de taille n = 35 est donné par hypothèse : $\tilde{I} = [3033\ ;\ 3317]$. On doit calculer le niveau de confiance $(1 - \alpha)$.

Le centre de cet intervalle est son milieu $\frac{3033\,+\,3317}{2} = 3175$. Cet intervalle de pari est centré sur la moyenne $\mu = 3175$. L'erreur absolue donnée par la formule $u_\alpha \frac{\sigma}{\sqrt{n}}$ est égale à la demi-longueur de l'intervalle de pari $u_\alpha \frac{\sigma}{\sqrt{n}} = \frac{3317\,-\,3033}{2} = 142 \Rightarrow$

$u_\alpha = 142 \frac{\sqrt{n}}{\sigma} = \frac{142\sqrt{35}}{584} = 1.4385 \approx 1.44$

La lecture de la table n° 1 pour $u_\alpha = 1.44$ donne le quantile d'ordre $1 - \frac{\alpha}{2} = 0.9251 \Rightarrow \alpha = 0.1498$ $\approx$ **15 %**. De même en utilisant la table n° 2 pour la valeur 1.44 on trouve $\alpha = 0.15 =$ **15 %**.

Par conséquent le niveau de confiance est $\mathbf{1 - \alpha = 85\ \%}$.

Note : Le symbole u_α est le même que le symbole du cours $\bar{z}_\alpha$.

Exercice 4

Une étude a été menée pour déterminer si les céréales, au son d'avoine, contribuent à abaisser le taux de cholestérol sérique. Quatorze hommes choisis au hasard ont été placés sur un régime comprenant du son d'avoine (ou des flocons de maïs) : après deux semaines, leurs taux de cholestérol des lipoprotéines de basse densité (LDL) ont été enregistrés. Chaque homme est ensuite passé à un régime alternatif. Après une deuxième période de deux semaines, le taux de cholestérol LDL de chaque homme a été enregistré à nouveau. Les données sont présentées dans le tableau suivant.

Sujet	LDL (mmol/L)	
	Son d'avoine	Flocons d'avoine
1	4.61	3.84
2	6.42	5.57
3	5.40	5.85
4	4.54	4.8
5	3.98	3.68
6	3.82	2.96
7	5.01	4.41
8	4.34	3.72
9	3.8	3.49

10	4.56	3.84
11	5.35	5.25
12	3.89	3.73
13	2.25	1.84
14	4.24	4.14

Calculer l'intervalle de confiance à 95% pour la moyenne des différences des taux de cholestérol LDL (mmol/L: son d'avoine – flocons d'avoine).

On suppose dans cet exercice que cette différence suit la loi normale.

Solution

Le tableau suivant donne les différences des taux de cholestérol LDL :
(mmol/L: différence d = son d'avoine – flocons d'avoine).

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.77	0.85	– 0.45	– 0.26	0.3	0.86	0.6	0.62	0.31	0.72	0.1	0.16	0.41	0.1

Calcul de $\bar{x}$; et de s = écart-type de l'échantillon.

La calculatrice donne : $\bar{x} = 0.363571428$; $s = 0.405455245$

On a à traiter un petit échantillon de n = 14 < 30.

Donc c'est la loi de Student qui va être utilisée avec les degrés de liberté ddl = n – 1 = 13.

L'intervalle de confiance est donnée par la formule : I.C. = $[\bar{x} \pm t_\alpha \frac{s}{\sqrt{n}}]$ avec $t_{\alpha(n-1)}) = 2.16$

I.C. = $[0.363571428 \pm 2.16\ \frac{0.405455245}{\sqrt{14}}]$ = $[0.363571428 \pm 0.23406294]$ =

$[0.129508487 ; 0.597634368] \approx$ **[0.130 ; 0.598].**

Exercice 5

Une enquête concernant la santé, et portant sur 3000 adolescents d'un certain pays européen de 12 à 20 ans, a dénombré 570 adolescents ayant pris un psychotrope au cours des 12 mois précédant l'enquête. Parmi les 1 400 filles, 378 ont pris un psychotrope.

1. Donner une estimation ponctuelle de la fréquence de consommation de psychotropes chez les adolescents de ce pays.
2. Estimer par intervalle de confiance à 95 % la fréquence de consommation de psychotropes chez les adolescents de ce pays.
3. Même questions chez les filles, puis les garçons.
4. Quelle devrait être la taille de l'échantillon pour que la marge d'erreur à 95 % dans l'estimation de la fréquence de consommation de psychotropes chez les adolescents de ce pays soit inférieure à 1 %, en supposant que la fréquence observée de consommation de psychotropes n'est pas modifiée.

Solution

1) P = {Adolescents de 12 à 20 ans}.
La variable X = consommation de psychotropes.
Cette variable X est qualitative dichotomique (oui et non).

p = proportion de consommateurs de psychotropes dans P = fréquence de consommation de psychotropes dans P, p inconnue dans la population P.

Grand échantillon de taille n (= 3 000 > 30) de X issu de P. La loi normale sera utilisée.

L'estimation ponctuelle de p est donnée par la fréquence observée f = $\frac{570}{3\ 000}$ = 19 %

⇒ La proportion d'adolescents de 12 à 20 ans consommateurs de psychotropes est estimée à **19 %**.

2) L'estimation par intervalle de confiance à 95 % (au risque 5 %) de p dans P est donnée par la formule du cours comme suit :

$IC_{95\%}(p) = [f \pm z_{1-\alpha/2} * \sqrt{\frac{f(1-f)}{n}}] = [0.19 \pm 1.96 * \sqrt{\frac{0.19(1-0.19)}{3\ 000}}] =$

$[0.19 \pm 0.0140383]$ = **[0.176 ; 0.204]**

Où $z_{1-\alpha/2} = z_{0.975} = 1.96$ est le quantile d'ordre 0.975 de la loi N (0 ; 1)

Cette approximation est justifiée puisque toutes les conditions d'application de l'estimation sont satisfaites.

3) FILLES : P_1 = {Adolescentes de 12 à 20 ans, de sexe féminin}.

La variable X = consommation de psychotropes.

Cette variable X est qualitative dichotomique (oui et non).

p_1 = proportion de consommatrices de psychotropes dans P_1 = fréquence de consommation de psychotropes dans P_1, p_1 inconnue dans la population P_1.

Echantillon de taille n_1 = 1 400 de X issu de P_1. C'est un grand échantillon et la loi normale utilisée.

L'estimation ponctuelle de p_1 est donnée par la fréquence observée $f_1 = \frac{378}{1\ 400} = \mathbf{0.27 = 27}$ % ⇒

La proportion d'adolescentes de 12 à 20 ans consommatrices de psychotropes est estimée à 27 %.

L'estimation par intervalle de confiance à 95 % (au risque 5 %) de p_1 dans P_1 s'écrit :

$IC_{95\%}(p_1) = [f_1 \pm z_{1-\alpha/2} * \sqrt{\frac{f_1(1-f_1)}{n}}] = [0.27 \pm 1.96 * \sqrt{\frac{0.27(1-0.27)}{1\ 400}}] =$

$[0.27 \pm 0.023256018]$ = **[0.247 ; 0.293]**

Où $z_{1-\alpha/2} = z_{0.975} = 1.96$ est le quantile d'ordre 0.975 de la loi N (0 ; 1).

GARÇONS : P_2 = {Adolescents de 12 à 20 ans, de sexe masculin}.

La variable X = consommation de psychotropes.

Cette variable X est qualitative dichotomique (oui et non).

p_2 = proportion de consommateurs de psychotropes dans P_2 = fréquence de consommation de psychotropes dans P_2, p_2 inconnue dans la population P_2.

Grand échantillon de taille n_2 = 3 000 – 1 400 = 1 600 de X issu de P_2

L'estimation ponctuelle de p_2 est donnée par la fréquence observée $f_2 = \frac{570-378}{1\ 600}$ = **0.125 = 12.5 %**

⇒ La proportion d'adolescents de 12 à 20 ans consommateurs de psycho est estimée à 12.5 %.

L'estimation par intervalle de confiance à 95 % (au risque 5 %) de p_2 dans P_2 s'écrit :

$IC_{95\%}(p_2) = [f_2 \pm z_{1-\alpha/2} * \sqrt{\frac{f_2(1-f_2)}{n}}] = [0.125 \pm 1.96 * \sqrt{\frac{0.125(1-0.125)}{1\ 600}}] =$

$[0.125 \pm 0.016297565]$ = **[0.109 ; 0.141]**

Où $z_{1-\alpha/2} = z_{0.975} = 1.96$ est le quantile d'ordre 0.975 de la loi N (0 ; 1)

4) La précision de l'intervalle de confiance à 95 % est donnée par sa demi-longueur. Pour une taille d'échantillon n = 3 000, la demi-longueur de l'intervalle $IC_{95\%}(p)$ est de 1.4 % (voir question précédente) ; pour obtenir une demi-longueur plus faible, de 1 %, il faudra donc plus de 3 000

sujets. Pour n inconnu, **f = 0.19 et α = 5 %** connus, la demi-longueur de l'intervalle $IC_{95\%}$ (p) s'écrit comme suit : $z_{1-\alpha/2} * \sqrt{\frac{f(1-f)}{n}} = 1.96 * \sqrt{\frac{0.19 * 0.81}{n}}$. On cherche n tel que :

$1.96 * \sqrt{\frac{0.19 * 0.81}{n}} \leq 0.01$ c'est-à-dire $1.96 * \sqrt{\frac{0.19 * 0.81}{0.0001}} = 76.89097736 \leq \sqrt{n}$ d'où $n \geq 5912.2224$

$\Rightarrow$ On choisira donc une taille d'échantillon au moins égale à **5 913** pour que la demi-longueur de l'intervalle de confiance à 95 % soit inférieure à 1 %.

Note : Les symboles à utiliser comme ceux du cours sont : $u_\alpha = \bar{z}_\alpha = z_{1-\frac{\alpha}{2}}$ f = $\tilde{p}$ $t_\alpha = \bar{t}_\alpha$

Exercice 6

Un échantillon de 80 stimulateurs cardiaques étudié a donné les résultats suivants :
La moyenne est : $\bar{x} = 0.31$ et l'écart-type est s = 0.015

1) Donner un intervalle de confiance à 0.95 % pour la moyenne des stimulateurs cardiaques.
2) Quelle sera la taille n de l'échantillon si l'erreur absolue est inférieure à 0.001 ?

Solution

1) On est en présence d'un grand échantillon n = 80 > 30. La loi normale sera utilisée.
   La formule donnant l'I.C. = $[\bar{x} - u_\alpha \frac{s}{\sqrt{n}} ; \bar{x} + u_\alpha \frac{s}{\sqrt{n}}] \Rightarrow$
   $P[\bar{X} - u_\alpha \frac{S}{\sqrt{n}} < \mu < \bar{X} + u_\alpha \frac{S}{\sqrt{n}}] = 1 - \alpha = 0.95$. Avec les données de l'hypothèse :

   $\bar{x} = 0.31$ et s = 0.015 sont des observations de $\bar{X}$ et S sur un échantillon de n = 80, et de la table n° 1 on tire $u_\alpha = 1.96$ alors l'intervalle de confiance demandé sera :

   $I.C. = [0.31 - 1.96 \frac{0.015}{\sqrt{80}} ; 0.31 + 1.96 \frac{0.015}{\sqrt{80}}] = [0.31 \pm 0.003287] =$

   $[0.306713 ; 0.313287] \approx$ ***[0.307 ; 0.313]***

2) L'erreur absolue $u_\alpha \frac{s}{\sqrt{n}}$ doit être < 0.001 $\Rightarrow 1.96 \frac{0.015}{\sqrt{n}} < 0.001 \Rightarrow n > 29.4^2 = 864.36$
   On prend un échantillon de ***n = 865*** stimulateurs pour avoir 1 erreur absolue qui ne dépasserait pas 0.001

Note : Le symbole u_α est donné dans le cours comme $\bar{z}_\alpha$