!mUNIVERSITE DE BATNA_2 **FACULTE DE MEDECINE**

SERIE DE TD N° 7 **DE BIOSTATISTIQUES**

1ère Année Médecine, 2022/2023

TD sur la théorie d'échantillonnage

Exercice n° 01 :

Dans une population de 5 sujets, on étudie le caractère X associé au poids de chacun des objets, les poids mesurés sont : 2.50 kg 2.53 kg 2.60 kg 2.62kg 2.70 kg

1) Déterminer la moyenne et la variance de la distribution parente.
2) Former tous les k échantillons de taille n = 3 et extraits de la population-mère (les différents **tirages sont exhaustifs**).
3) Calculer $(\mu_{\bar{X}})$ et $(\sigma_{\bar{X}})$ de la distribution d'échantillonnage des moyennes.
4) Vérifier que $(\mu_{\bar{X}}) = \mu$ et $(\sigma_{\bar{X}}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ où N est la taille de la population mère.
5) Mêmes questions avec des tirages **non exhaustifs** et avec n = 2 au lieu de 3.

Solution de l'exercice n° 1

En faisant rentrer les données dans la calculatrice, on tire les résultats voulus suivants :

1) Les paramètres de la population-mère : $\mu = 2.59$ $\sigma = 0.070427267$ $\sigma^2 = 0.00496$
2) En décomposant cette population de 5 sujets en différents échantillons de taille n = 3 (avec des **tirages exhaustifs** ou sans remise), on trouve le nombre k d'échantillons, c'est-à-dire : $\mathbf{k = C_5^3 = 10}$ **échantillons**.

   Tous les k échantillons de taille 3 sont les suivants : D.E. des moyennes

Ech 1	2.50	2.53	2.60	2.543333333
Ech 2	2.50	2.53	2.62	2.55
Ech 3	2.50	2.53	2.70	2.576666667
Ech 4	2.50	2.60	2.62	2.573333333
Ech 5	2.50	2.60	2.70	2.6
Ech 6	2.50	2.62	2.70	2.606666667
Ech 7	2.53	2.60	2.62	2.583333333
Ech 8	2.53	2.60	2.70	2.61
Ech 9	2.53	2.62	2.70	2.616666667
Ech 10	2.60	2.62	2.70	2.64

3) La distribution d'échantillonnage des moyennes des 10 Echantillons se trouve à la dernière colonne du tableau. La calculatrice donne $(\boldsymbol{\mu_{\bar{x}}}) = \mathbf{2.59}$ $(\boldsymbol{\sigma_{\bar{x}}}) = \mathbf{0.028751811}$
4) Vérification : $\mu = \mathbf{2.59} = (\mu_{\bar{x}})$ $(\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{0.070427267}{\sqrt{3}} * \sqrt{\frac{5-3}{5-1}} = \mathbf{0.028751811}$

   Donc on confirme que les relations du cours sont vérifiées.

   Mêmes questions et un **tirage non exhaustif** ou avec remise et n = 2 au lieu de n = 3.
5) Avec ce type de tirage, le nombre k d'échantillons sera : $k = R_N^n = R_5^2 = 5^2 =$ **25 échantillons**.

   Les paramètres de la population-mère μ et σ sont calculés en 1ère question :

   $\mu = 2.59$ $\sigma = 0.070427267$ $\sigma^2 = 0.00496$

En décomposant cette population de 5 sujets en échantillons de taille n = 2 (avec des **tirages exhaustifs** ou sans remise), on trouve C_5^2 = 10 échantillons.

Tous les échantillons de taille 2 sont les suivants : D.E. des moyennes

Ech 1	2.50	2.53	2.515
Ech 2	2.50	2.60	2.55
Ech 3	2.50	2.62	2.56
Ech 4	2.50	2.70	2.6
Ech 5	2.53	2.60	2.565
Ech 6	2.53	2.62	2.575
Ech 7	2.53	2.70	2.615
Ech 8	2.60	2.62	2.61
Ech 9	2.60	2.70	2.65
Ech 10	2.62	2.70	2.66

La distribution d'échantillonnage des moyennes de ces 10 échantillons se trouve à la dernière colonne du tableau ci-dessus. La calculatrice donne $(\boldsymbol{\mu_{\bar{x}}})$ **= 2.59** $(\boldsymbol{\sigma_{\bar{x}}})$ **= 0.043127717**

Vérification : μ = **2.59** = $(\mu_{\bar{x}})$ $(\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{0.070427267}{\sqrt{2}} * \sqrt{\frac{5-2}{5-1}}$ = **0.043127717**

Donc on confirme que les relations du cours sont vérifiées.

En décomposant cette population de 5 sujets en différents échantillons de taille n = 2 (avec des **tirages non exhaustifs** ou avec remise), on trouve au total : $\mathbf{k = 5^2 = 25}$ **échantillons**.

La répartition de ces 25 échantillons (10 échantillons ci-dessus et 15 échantillons ci-dessous)

Les 15 échantillons restants de taille n = 2 et comportant l'inverse des 10 premiers échantillons et les cinq échantillons avec répétition des sujets sont les suivants :

N° Ech			Moyenne	N° Ech			Moyenne
Ech 11	2.53	2.50	2.515	Ech 19	2.70	2.60	2.65
Ech 12	2.60	2.50	2.55	Ech 20	2.70	2.62	2.66
Ech 13	2.62	2.50	2.56	Ech 21	2.5	2.5	2.5
Ech 14	2.70	2.50	2.6	Ech 22	2.53	2.53	2.53
Ech 15	2.60	2.53	2.565	Ech 23	2.6	2.6	2.6
Ech 16	2.62	2.53	2.575	Ech 24	2.62	2.62	2.62
Ech 17	2.70	2.53	2.615	Ech 25	2.7	2.7	2.7
Ech 18	2.62	2.60	2.61				$\mu_{\bar{x}}$ **= 2.59**

La Distribution d'Echantillonnage des moyennes de tous ces 25 échantillons, donne :

$(\boldsymbol{\mu_{\bar{x}}})$ **= 2.59** $(\boldsymbol{\sigma_{\bar{x}}})$ **= 0.049799598**

Vérification μ = **2.59** = $(\mu_{\bar{x}})$ et $(\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}} = \frac{0.070427267}{\sqrt{2}}$ = **0.049799598**

Donc on confirme également que les relations du cours sont vérifiées.

Exercice n° 02 :

Sur 10 personnes vaccinées contre le covid–19, 8 ont des effets indésirables, on considère tous les échantillons de tailles 9 extraits de cette population.

1) Quelle est le nombre k d'échantillons en considérant des tirages exhaustifs ?
2) Quelle est la distribution d'échantillonnage des fréquences F d'apparition de l'événement : « *Effet indésirable* » ?
3) En supposant que la distribution mère suit une Loi Binomiale, vérifier que : $\mu_F = p$ et

$\sigma_F = \sqrt{\frac{p * q}{n}} * \sqrt{\frac{N-n}{N-1}}$ où p est la probabilité élémentaire pour une personne de ressentir des effets indésirables.

Solution de l'exercice n° 2

1) Le nombre k d'échantillons en considérant que les tirages sont exhaustifs est donné par la formule de combinaison simple : k = C_{10}^{9} = 10 échantillons différents.
2) La distribution d'échantillonnage des fréquences F d'apparition de l'événement « *effets indésirables* » est donnée par :

   Le paramètre d'intérêt est la proportion **p** dans la population.

   L'estimateur de **p** : F = $\frac{X}{n}$ où la variable aléatoire X suit la Loi Binomiale B (n ; p)

   $\Rightarrow Z = \frac{F-p}{\sqrt{pq/n}} \sim$ N (0 ; 1) si **n ∗ p** et **n ∗ q** > **5**

En supposant que la distribution mère suit une ***loi Binomiale*** comme il a été bien précisé dans les données par hypothèse, on doit vérifier que $\mu_F = p$ et que $\sigma_F = \sqrt{\frac{p*q}{n}} * \sqrt{\frac{N-n}{N-1}}$

Alors en prenant des individus au hasard de la population, on a ce qui suit : ou bien ils ont des effets indésirables avec une fréquence de 80 % ou bien ils n'en ont pas avec une fréquence de 20% ⇒ on prendra **p = 0.8**

Population : Ind_1 Ind_2 Ind_3 Ind_4 Ind_5 Ind_6 Ind_7 Ind_8 Non_9 Non_{10}

Individus de la population : $Ind_1, \ldots\ldots, Ind_8$ ont des effets indésirables les 2 derniers *NON*.

Les différents échantillons (Ech_1 à Ech_10) extraits de la population sont les suivants :

Ech_1 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_7	Ind_8	$Ind_9 \Rightarrow f_1 = 8/9$
Ech_2 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_7	Ind_8	$Non_{10} \Rightarrow f_2 = 8/9$
Ech_3 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_7	Non_9	$Non_{10} \Rightarrow f_3 = 7/9$
Ech_4 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_8	Non_9	$Non_{10} \Rightarrow f_4 = 7/9$
Ech_5 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_7	Ind_8	Non_9	$Non_{10} \Rightarrow f_5 = 7/9$
Ech_6 :	Ind_1	Ind_2	Ind_3	Ind_4	Ind_6	Ind_7	Ind_8	Non_9	$Non_{10} \Rightarrow f_6 = 7/9$
Ech_7 :	Ind_1	Ind_2	Ind_3	Ind_5	Ind_6	Ind_7	Ind_8	Non_9	$Non_{10} \Rightarrow f_7 = 7/9$
Ech_8 :	Ind_1	Ind_2	Ind_4	Ind_5	Ind_6	Ind_7	Ind_8	Non_9	$Non_{10} \Rightarrow f_8 = 7/9$
Ech_9 :	Ind_1	Ind_3	Ind_4	Ind_5	Ind_6	Ind_{77}	Ind_8	Non_9	$Non_{10} \Rightarrow f_9 = 7/9$
Ech_{10}:	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_7	Ind_8	Non_9	$Non_{10} \Rightarrow f_{10} = 7/9$

On va calculer la moyenne et la variance de cette distribution d'échantillonnage des fréquences {8/9 ; 8/9 ; 7/9 ; 7/9 ; 7/9 ; 7/9 ; 7/9 ; 7/9 ; 7/9 ; 7/9} ci-dessus.

$$\mu_F = \sum_{i=1}^{10} f_i = \frac{\frac{8}{9}+\frac{8}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}+\frac{7}{9}}{10} = \frac{\frac{72}{9}}{10} = \frac{8}{10} = \mathbf{0.8}$$

$$\sigma_F^2 = \frac{\left(\frac{8}{9}\right)^2+\left(\frac{8}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2+\left(\frac{7}{9}\right)^2}{10} - (0.8)^2 = 0.00197531$$

$$\sigma_F = \sqrt{0.00197531} = 0.044444444 \cong \mathbf{0.044}$$

3) Comme la distribution mère suit une Loi Binomiale B (n ; p), alors la vérification des formules de la distribution d'échantillonnage dans le cas d'un tirage exhaustif donne :

   $\boldsymbol{\mu_F} = \mathbf{0.8 = p}$ $\quad\quad$ $\boldsymbol{\sigma_F} = \sqrt{\frac{p*q}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0.8*0.2}{9}}\sqrt{\frac{10-9}{10-1}} = 0.044444444 \cong \mathbf{0.044}$

   Donc les formules données au cours sont bien vérifiées.

Exercice n° 03 :

Après la correction du concours de résidanat (comportant un grand nombre de candidats) on constate que les notes ont pour moyenne 12 et pour écart-type 3. On se propose de prélever des échantillons aléatoires non exhaustifs de 100 participants.

1) Quelle est la probabilité d'avoir la moyenne des notes de ces 100 candidats supérieure à 12.5 ?
2) Quelle est la probabilité d'avoir la moyenne des notes de ces 100 candidats comprise entre 12.5 et 12.9 ?

Solution de l'exercice n° 3

On applique le théorème de la limite centrale. Taille n de l'échantillon = 100, ç'est-à-dire que n > 30 donc c'est un grand échantillon. La variable aléatoire X suit la loi normale quelconque

N ($\mu = \bar{x} = 12$; $\mu_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{10} = 0.3$) = **N (12 ; 0.3)**

Pour ce calcul, on transforme la variable aléatoire quelconque X en variable aléatoire tabulée Z qui est de référence.

1) P (X > 12.5) = P (Z > $\frac{12.5 - 12}{0.3}$ = 1.667) = 1 – P (Z < 1.667) = 1 – 0.9525 = **0.0475**
2) P (12.5 < X < 12.9) = P ($\frac{12.5 - 12}{0.3}$ < Z < $\frac{12.9 - 12}{0.3}$) = Π(3) – Π(1.667) = 0.9987 – 0.9525 = **0.0462** ***Ces probabilités sont calculées en utilisant la table statistique de la loi normale n° 1***.

Exercice n° 04 :

Dans une population, on constate qu'il nait 52% de garçons. On prélève des échantillons aléatoires non exhaustifs de 400 nouveaux nés ?

1) Quelle est la probabilité dans un tel échantillon, un pourcentage de garçons compris entre 50 % et 54 % ?
2) Quelle est la probabilité d'avoir la moyenne d'un tel échantillon, un pourcentage de filles inférieur à 45% ?

Solution de l'exercice n° 4

On étudie un caractère qualitatif (G : être un garçon) de proportion p = 0,52 (donc le caractère F : être une fille a la proportion de 0.48) dans la population-mère. On tire un échantillon de taille 400. On suppose que l'effectif de la population mère est suffisamment important pour que le tirage de l'échantillon puisse être assimilé à un tirage avec remise (non exhaustif).

1) Soit G la variable aléatoire qui à chaque échantillon de 400 nouveau-nés, associe le pourcentage de Garçons. C'est un grand échantillon car : $n = 400 \geq 30$

   Les conditions (données dans le cours) d'approximer une loi binomiale par une loi normale sont les suivantes :

   $n * p = 400 * 0{,}52 = 208 \geq 5$ et $n * q = 400 \times 0{,}48 = 192 \geq 5$, donc la loi de G est approchée (ou approximée) par la loi normale quelconque :

   $N\left(p\ ;\ \sqrt{\frac{p * q}{n}}\right) = N\left(0.52\ ;\ \sqrt{\frac{0.52 * 0.48}{400}}\right)$ ou encore N ($\mu = 0.52$; $\sigma = 0.024979991$) ;

   Alors on considèrera dans les calculs : **N (0.52 ; 0.025)**

   La probabilité d'avoir un pourcentage de garçons compris entre : 50 % et 54 %, dans un échantillon de taille 400, est donnée par la probabilité suivante (sachant que G est une variable aléatoire quelconque et Z est une variable aléatoire tabulée) :

   $\mathbf{P\ (0.5 < G < 0.54)} \cong P\left(\frac{0.5 - 0.52}{0.025} < Z < \frac{0.54 - 0.52}{0.025}\right) \Rightarrow$

   $P(-0.8 < Z < +0.8) = \Pi(0.8) - \Pi(-0.8) = 2 * \Pi(0.8) - 1 = 2 * 0.7881 - 1 = \mathbf{0.5762}$

2) La probabilité d'avoir un pourcentage de filles inférieur à 45 % est synonyme de la probabilité d'avoir un pourcentage de garçons supérieur à 0,55.

   C'est-à-dire : $P(F \leq 0{,}45) = P(G \geq 0{,}55) \cong P\left(Z \geq \frac{0.55 - 0.52}{0.025} = 1.2\right) \Rightarrow$

   $P(Z \geq 1.2) = 1 - P(Z < 1.2) = 1 - 0.8849 = \mathbf{0.1151}$

   On peut considérer également la variable aléatoire F : être une fille de proportion 0.48 et donc cette variable aléatoire F va suivre une loi normale quelconque

   N ($\mu = 0.52$; $\sigma = 0.024979991 \approx \mathbf{0.025}$).

   $P(F < 0{,}45) = P\left(Z < \frac{0.45 - 0.48}{0.025} = -1.2\right) \Rightarrow$

   $P(Z < -1.2) = 1 - P(Z < 1.2) = 1 - 0.8849 = \mathbf{0.1151}$

Remarque

Les symboles < et ≤ ont la même signification pour une variable aléatoire continue.