Durée : 2h Téléphone portable et documents interdits. Calculatrice autorisée.

Les résultats numériques doivent être justifiés en détaillant les calculs. Vous devez donner pour chaque question une **phrase de conclusion en Français**.

EXERCICE 1. On suppose que l'âge auquel apparaissent les premiers mots de vocabulaire chez l'enfant suit la loi normale de moyenne 12 mois et d'écart-type $2,5$ mois.

1.1) Identifier la population, la variable, son type et son/ses paramètre(s).

1.2) Quelle est la proportion d'enfants pour lesquels les premiers mots apparaissent avant 9 mois ?

1.3) Déterminer l'âge au-dessus duquel 2% des enfants prononcent leurs premiers mots.

CORRECTION.

1.1)

- *Population $\mathcal{P}$: { Enfants }.*
- *Variable quantitative X = "âge d'apparition (en mois) des premiers mots de vocabulaire".*
- *2 paramètres connus : moyenne $\mu = 12$ et écart-type $\sigma = 2,5$.*

1.2) *Commençons par rappeler que $Z = (X - 12)/2,5$ suit la loi normale centrée/réduite. On cherche à calculer*

$$\begin{aligned} P(X \leq 9) &= P\left(\frac{X-12}{2,5} \leq \frac{9-12}{2,5}\right) \\ &= P(Z \leq -1,2) \\ &= F(-1,2) = 1 - F(1,2) = 1 - 0,8849 = 0,1151. \end{aligned}$$

Chez environ $11,5\%$ des enfants les premiers mots apparaissent avant 9 mois.

1.3) *On recherche le quantile d'ordre $0,98$, noté $q_{0,98}$, pour la variable X. D'après la formule du cours,*

$$q_{0,98} = \sigma \times z_{0,98} + \mu, \text{ soit } q_{0,98} = 2,5 \times z_{0,98} + 12,$$

où $z_{0,98}$ est le quantile de la loi normale centrée/réduite. D'après la table, il vaut environ $2,05$, et donc $q_{0,98} = 17,125$.

Chez 2% des enfants, les premiers mots apparaissent après $17,1$ mois.

EXERCICE 2. Dans le cadre d'une étude sur la santé au travail, on a interrogé au hasard 500 salariés de différents secteurs et de différentes régions de France. 145 d'entre eux déclarent avoir déjà subi un harcèlement moral au travail.

2.1) Identifier la population, la variable, son type et son/ses paramètre(s).

2.2) Donner une estimation ponctuelle de la proportion de salariés ayant déjà subi un harcèlement moral au travail.

2.3) Donner une estimation de cette proportion par un intervalle de confiance à 90%.

2.4) Si avec les mêmes données on calculait un intervalle de confiance à 95%, serait-il plus grand ou plus petit que celui trouvé à la question précédente ? (justifier sans calcul.)

CORRECTION.

2.1)

- *Population $\mathcal{P}$: { Salariés en France }.*
- *Variable qualitative X = "a déjà subi un harcèlement moral"*
- *1 paramètre : proportion de la modalité "oui".*

2.2) *On estime p par la fréquence observée de "oui" $f = 145/500 = 0,29$.*

2.3) *On a tout d'abord $n = 500 \geq 30$. L'estimation par intervalle à 90% est donnée par*

$$\begin{aligned}
\mathrm{IC}_{0,90}(p) &= \left[f \pm z_{0,95} \frac{\sqrt{f(1-f)}}{\sqrt{n}} \right], \\
&= \left[0,29 \pm 1,645 \frac{\sqrt{0,29 \times 0,71}}{\sqrt{500}} \right], \\
&= [0,29 \pm 0,033] = [0,257; 0,323].
\end{aligned}$$

Il nous reste à vérifier a posteriori *les conditions sur f_i, f_s. Par exemple, $nf_i = 500 \times 0,257 = 128,5$, les trois autres sont également vérifiées.*

L'estimation de p par intervalle à 90% est donc l'intervalle $[0,257; 0,323]$.

2.4) *Dans l'intervalle obtenu à la question précédente, il suffirait de remplacer $1,645$ par $z_{0,975} = 1,96$ qui est plus grand. On obtiendrait donc un intervalle plus grand.*

EXERCICE 3. En vue de réaliser un programme de rééducation, des chercheurs ont soumis un questionnaire de neuropsychologie cognitive à 150 enfants dyslexiques tirés au sort. Le questionnaire comporte 20 questions et les chercheurs ont recueilli pour chaque enfant dyslexique le nombre x_i de bonnes réponses. Les résultats ainsi récoltés sont tels que :

$$\sum x_i = 1502, \qquad \sum x_i^2 = 19486.$$

3.1) Identifier la population, la variable, son type et son/ses paramètre(s).

3.2) Donner une estimation ponctuelle du nombre moyen de bonnes réponses dans la population étudiée.

3.3) Donner une estimation ponctuelle de l'écart-type de la variable.

3.4) Estimer le nombre moyen de bonnes réponses dans la population par un intervalle de confiance au niveau 99%.

3.5) Quelle est la marge d'erreur dans l'estimation du nombre moyen de bonnes réponses au niveau 99% ?

CORRECTION.

3.1)

- *Population $\mathcal{P}$: { Enfants dyslexiques }.*
- *Variable quantitative X = "Nombre de bonnes réponses au questionnaire"*
- *2 paramètres inconnus : moyenne μ et écart-type σ.*

3.2) *On estime la moyenne μ par la moyenne observée $\bar{x} = 1502/150 \approx 10,01$.*

3.3) *Commençons par calculer la variance observée :*

$$s^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \frac{19486}{150} - (10,01)^2 = 29,7.$$

La variance corrigée vaut donc $s^{\star 2} = \frac{n}{n-1} s^2 = \frac{150}{149} 29,7 \approx 29,9$.

Finalement, on estime l'écart-type σ par l'écart-type corrigé $s^\star = \sqrt{s^{\star 2}} \approx 5,47$.

3.4) *Puisque $n = 150 \geq 30$, l'estimation par intervalle à 99% est donnée par*

$$\begin{aligned}
\mathrm{IC}_{0,99}(\mu) &= \left[\bar{x} \pm z_{0,995} \frac{s^\star}{\sqrt{n}} \right], \\
&= \left[10,01 \pm 2,57 \frac{5,47}{\sqrt{150}} \right], \\
&= [10,01 \pm 1,15] = [8,86; 11,16],
\end{aligned}$$

on trouve en effet dans la table que $z_{0,995} \approx 2,57$.

Donc l'estimation de μ *par intervalle à* 99% *est l'intervalle* $[8,86; 11,16]$.

3.5) *La marge d'erreur est la demi-longueur de l'intervalle obtenu à la question précedente, elle vaut donc* $1,15$.

EXERCICE 4. L'inventaire de Padoue est un questionnaire portant sur les troubles obsessionnels du comportement (TOC). Chez les adultes dépressifs, le score obtenu à ce questionnaire a pour moyenne 84 avec un écart-type de 35. Des chercheurs s'intéressent alors aux scores moyens observés dans les échantillons de taille 75.

4.1) Identifier la population, la variable, son type et son/ses paramètre(s).

4.2) Caractériser la distribution de la moyenne empirique du score à l'inventaire de Padoue sur les échantillons de taille 75 (forme et valeur(s) de son/ses paramètre(s)).

4.3) Quelle est la probabilité d'observer sur un échantillon de taille 75 un score moyen inférieur à 90 ?

4.4) En dessous de quelle valeur se trouvent 95 % des scores moyens observés sur les échantillons de taille 75 ?

4.5) Au dessus de quelle valeur se trouvent 95 % des scores moyens observés sur les échantillons de taille 75 ?

4.6) Pour quelle proportion d'échantillons observe-t-on un score moyen compris entre les deux valeurs déterminées aux questions 4.4 et 4.5 ?

CORRECTION.

4.1)

- *Population* $\mathcal{P}$: { *Adultes dépressifs* }.
- *Variable quantitative* X = *"Score à l'inventaire de Padoue"*
- 2 *paramètres connus : moyenne* $\mu = 84$ *et écart-type* $\sigma = 35$.

4.2) *On s'intéresse à la moyenne empirique* $\overline{X}_n$ *obtenue sur un échantillon tiré au sort de taille* $n = 75$. *D'après le cours, puisque* $n \geq 30$,

$$\overline{X}_n \overset{approx.}{\sim} \mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right) = \mathcal{N}\left(84; \frac{35}{\sqrt{75}}\right) = \mathcal{N}(84; 4,04).$$

Donc la forme de $\overline{X}_n$ *est la loi normale, sa moyenne est* 84 *et son écart-type est* $4,04$ *(soit une variance de* $16,3$*).*

4.3) *La variable* $Z = (\overline{X}_n - 84)/4,04$ *suit la loi normale centrée/réduite. On cherche à calculer*

$$\begin{aligned} P(\overline{X}_n \leq 90) &= P\left(\frac{\overline{X}_n - 84}{4,04} \leq \frac{90-84}{4,04}\right) \\ &= P(Z \leq 1,49) = F(1,49) = 0,9319. \end{aligned}$$

Environ 93% *des échantillons de taille* $n = 75$ *donnent un score moyen inférieur ou égal à* 90.

4.4) *On cherche le quantile d'ordre* $0,95$, *noté* $q_{0,95}$ *pour une loi normale* $\mathcal{N}(84; 4,04)$. *D'après la formule du cours, il se calcule de la façon suivante :*

$$q_{0,95} = 4,04 \times z_{0,95} + 84.$$

On trouve dans la table que $z_{0,95} \approx 1,645$, *donc* $q_{0,95} \approx 90,6$.

Donc 95% *des échantillons de taille* $n = 75$ *donnent un score moyen inférieur ou égal à* $90,6$ *(c'est bien sûr très proche de la réponse obtenue à la question précédente).*

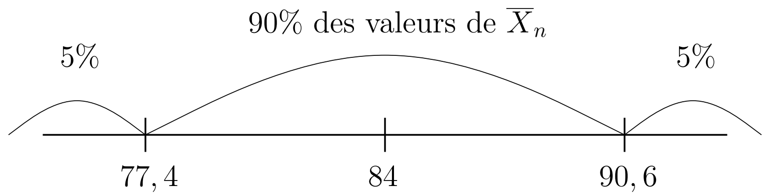
4.5) *On cherche le quantile d'ordre* $0,05$, *il s'obtient de façon similaire :*

$$q_{0,05} = 4,04 \times z_{0,05} + 84 = 4,04 \times (-z_{0,95}) + 84,$$

(on a appliqué la formule $z_{1-\alpha} = -z_\alpha$ *avec* $\alpha = 0,95$*). On trouve* $q_{0,05} \approx 77,4$.

Donc 95% *des échantillons de taille* $n = 75$ *donnent un score moyen supérieur ou égal à* $77,4$.

4.6) *Faisons un schéma qui résume les deux questions précédentes :*



Nous avons montré que 5% *des valeurs de* $\overline{X}_n$ *sont en-dessous de* $77,4$*, et que* 5% *sont au-dessus de* $90,6$*.*

On en déduit que $100 - 5 - 5 = 90\%$ *des échantillons donnent une moyenne empirique comprise entre* $77,4$ *et* $90,6$*.*