Essayez de faire ce devoir **dans les conditions de l'examen** : uniquement avec calculatrice, formulaire et table.

EXERCICE 1. La Journée d'Appel Pour la Défense (JAPD) est l'occasion de mesurer le flux de jeunes en situation d'illetrisme. Chaque jeune Français de 16 ans doit remplir un QCM de 60 questions pour évaluer sa compréhension d'un texte simple.

1) On s'intéresse tout d'abord au nombre moyen d'erreurs dans la population. Sur 200 jeunes tirés au sort, on relève le nombre total d'erreurs :

$$\sum x_i = 810, \quad \sum x_i^2 = 8211.$$

1.1) Identifier la population, l'échantillon, la variable et son/ses paramètres.

1.2) Donner une estimation ponctuelle du nombre moyen d'erreurs dans la population.

1.3) Donner une estimation ponctuelle de l'écart-type du nombre d'erreurs dans la population.

1.4) Donner une estimation par intervalle de confiance à 90% du nombre moyen d'erreurs dans la population.

2) Parmi ces 200 jeunes, 21 ont fait plus de douze erreurs, ce qui les caractérise comme ayant "de graves difficultés à la compréhension de la langue écrite".

2.1) Identifier la population, l'échantillon, la variable et son/ses paramètres.

2.2) Donner une estimation par intervalle de confiance à 99% de la proportion de jeunes ayant de graves difficultés.

CORRECTION. ***1.1)*** *La population est donnée par $\mathcal{P} = \{$ Jeunes Français de 16 ans $\}$, on observe un échantillon de taille $n = 200$ issu de $\mathcal{P}$ (il est fixé pour tout l'exercice).*
La variable mesurée sur chaque individu est $X =$"Nombre d'erreurs au QCM". Il s'agit d'une variable quantitative, elle a donc 2 paramètres : le nombre moyen d'erreurs μ et la variance du nombre d'erreurs σ^2, ils sont inconnus.

1.2) *On calcule $\bar{x} = \frac{\sum x_i}{n} = \frac{810}{200} = 4,05$. On estime donc μ par la moyenne observée $\bar{x} = 4,05$.*

1.3) *On commence par calculer la variance observée*

$$s^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{8211}{200} - 4,05^2 = 24,6525.$$

La variance corrigée est donc $s^{\star 2} = \frac{200}{199} \times 24,6525 \approx 24,78$. Finalement, on estime l'écart-type σ par l'écart-type corrigé $s^\star = \sqrt{s^{\star 2}} = \sqrt{24,78} \approx 4,98$.

1.4) *Puisque $n = 200 \geq 30$, on est dans le cadre de l'approximation par la loi normale. Donc le formulaire assure que l'estimation par intervalle de confiance à 90% de la moyenne est donnée par*

$$\mathrm{IC}_{0,90}(\mu) = \left[\bar{x} \pm z_{0,95} \frac{s^\star}{\sqrt{n}}\right].$$

Ici, $\bar{x} = 4,05$, $s^\star = 4,98$ et on trouve dans la table[1] *que $z_{0,95} \approx 1,645$. On obtient*

$$\mathrm{IC}_{0,90}(\mu) = \left[4,05 \pm 1,645 \frac{4,98}{\sqrt{200}}\right] = [4,05 \pm 0,58] = [3,47; 4,63].$$

Donc l'estimation de μ par intervalle de confiance à 90% est l'intervalle $[3,47; 4,63]$.

1. Ici, puisque $F(1,64) = 0.9495$ et $F(1,65) = 0.9505$, on pourrait prendre comme valeurs pour $z_{0,95}$ 1.64,1.65 ou 1.645.

2.1) *La population et l'échantillon restent identiques. La variable est maintenant Y ="a de graves difficultés à la compréhension de la langue écrite", c'est une variable qualitative dont le seul paramètre (inconnu) est p, la proportion de la modalité "oui" dans $\mathcal{P}$.*

2.2) *On est toujours bien sûr dans le cadre de l'approximation par la loi normale. Le formulaire assure que l'estimation par intervalle de confiance à 99% de p est donnée, sous réserves de conditions à vérifier* a posteriori, *par*

$$\mathrm{IC}_{0,99}(p) = \left[f \pm z_{0,995} \frac{\sqrt{f(1-f)}}{\sqrt{n}} \right].$$

Calculons d'abord f, la fréquence observée de "oui", $f = 21/200 = 0,105$. Dans la table, on trouve que $z_{0,995} \approx 2,575$. On trouve donc

$$\mathrm{IC}_{0,99}(p) = \left[0,105 \pm 2,575 \frac{\sqrt{0,105 \times 0,895}}{\sqrt{200}} \right] = [0,105 \pm 0,056] = [0,049; 0,161].$$

On peut maintenant vérifier les conditions[2] : $f_i = 0,049$ et $f_s = 0,161$.

$$\begin{cases} nf_i = 200 \times 0,049 = 9,8 \geq 5, & n(1-f_i) = 200 \times 0,951 = 190,2 \geq 5, \\ nf_s = 200 \times 0,161 = 32,1 \geq 5, & n(1-f_s) = 200 \times 0.839 = 167,8 \geq 5. \end{cases}$$

Donc l'estimation de p par intervalle de confiance à 99% est l'intervalle $[0,049; 0,161]$. On peut aussi écrire la réponse en pourcentage, cela donne l'intervalle $[4,9\%; 16,1\%]$.

EXERCICE 2. On dit qu'un individu est "fumeur régulier" s'il fume au moins une cigarette par jour. En 2010, 29% des Français de 15 à 75 ans sont des fumeurs réguliers.

1) Identifier la population et la variable étudiée.

2) On tire au sort 80 Français entre 15 et 75 ans, quelle est la probabilité que plus de 25 d'entre eux soient des fumeurs réguliers ?

CORRECTION. ***1)*** *La population est $\mathcal{P}$ = { Français de 15 à 75 ans }. La variable mesurée est X ="fumeur régulier", il s'agit d'une variable qualitative dont les modalités sont {oui,non}.*

2) *Soit p la proportion de la modalité "oui" dans la population, $p = 0,29$.*

Introduisons F_n, qui est la fréquence empirique de "oui" dans un échantillon de taille $n = 80$ tiré au sort dans $\mathcal{P}$. On cherche à calculer $P(F_n \geq \frac{25}{80}) = P(F_n \geq 0,3125)$. Puisque

$$n = 80 \geq 30, \qquad np = 80 \times 0,29 = 23,2 \geq 5, \qquad n(1-p) = 80 \times 0,71 = 56,8 \geq 5,$$

le formulaire assure que F_n suit approximativement une loi normale :

$$F_n \overset{approx.}{\sim} \mathcal{N}\left(p; \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = \mathcal{N}\left(0,29; \frac{\sqrt{0,29 \times 0,71}}{\sqrt{80}}\right) = \mathcal{N}(0,29; 0,0507).$$

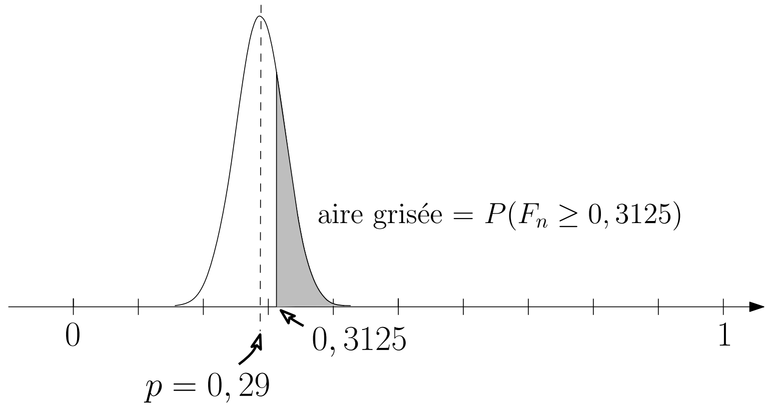
On peut donc faire le calcul en "centrant et réduisant" F_n :

$$\begin{aligned} P(F_n \geq \frac{25}{80}) = P(F_n \geq 0,3125) &= P\left(\frac{F_n - 0,29}{0,0507} \geq \frac{0,3125 - 0,29}{0,0507}\right) \\ &= P(Z \geq 0,44) \\ &= 1 - F(0,44) = 1 - 0,67 = 0,33. \end{aligned}$$

On a donc environ 33% de chance de tomber sur un échantillon avec plus de 25 fumeurs.

2. À l'examen, vous devez vérifier les 4 conditions avec votre calculatrice mais vous êtes autorisé à n'en écrire qu'une seule, en précisant par exemple "de même, les 3 autres conditions sont vérifiées".

Si on représente la distribution de F_n, nous obtenons :



Ce dessin confirme que la probabilité est plus faible que $0,50$: on considère des échantillons "atypiques" : avec une proportion de fumeurs plus élevée que dans $\mathcal{P}$.

EXERCICE 3. Dans une classe d'âge, le QI dit *standard* suit la loi normale de paramètres $\mu = 100$ et $\sigma = 15$.

1) Quelle est la proportion d'enfants de cette classe d'âge ayant un QI inférieur à 95 ?

2) D'après la page Wikipedia sur le Quotient Intellectuel,

> *L'écart type à 15 est arbitraire, mais il correspond à un écart probable de 10, ce qui veut dire qu'entre un QI de 90 et de 110, il y a environ 50% de la population.*

Pouvez-vous expliquer ce 50% ?

3) Les $2,5\%$ des enfants de cette classe d'âge qui ont le QI le plus faible ont un retard mental sévère. Quel est le seuil en-dessous duquel se situent leurs QI ?

4) On mesure le QI de 50 enfants de cette classe d'âge, on observe dans cet échantillon une moyenne de 120. Que peut-on penser de cet échantillon ?

CORRECTION. ***1)*** *Si l'on note X le QI standard dans la population des enfants de la classe d'âge considérée, alors $X \sim \mathcal{N}(100, 15)$. On cherche ici $P(X \leq 95)$. Il suffit de "centrer et réduire" :*

$$\begin{aligned} P(X \leq 95) &= P\left(\frac{X-100}{15} \leq \frac{95-100}{15}\right) \\ &= P(Z \leq -0,33) \\ &= F(-0,33) = 1 - F(0,33) \qquad \text{(rappelons que } F(-x) = 1 - F(x)\text{)} \\ &= 1 - 0,6293 = 0,3707. \end{aligned}$$

Donc environ 37% des enfants ont un QI inférieur ou égal à 95.

2) *Nous allons calculer $P(90 \leq X \leq 110)$, pour vérifier l'affirmation de Wikipedia.*

$$\begin{aligned} P(90 \leq X \leq 110) &= P\left(\frac{90-100}{15} \leq \frac{X-100}{15} \leq \frac{110-100}{15}\right) \\ &= P(-0,67 \leq Z \leq 0,67) = F(0,67) - F(-0,67) \\ &= F(0,67) - (1 - F(0,67)) \\ &= 0,7486 - (1 - 0,7486) = 0,4972. \end{aligned}$$

Ceci confirme qu'environ 50% des enfants de la classe d'âge ont un QI entre 90 et 110.

3) *On cherche le quantile $q_{0,025}$ pour une loi normale $\mathcal{N}(100; 15)$. D'après le formulaire on l'obtient en calculant*

$$q_{0,025} = \mu + z_{0,025} \times \sigma = 100 + z_{0,025} \times 15,$$

où $z_{0,025}$ désigne le quantile d'ordre $0,025$ pour une loi normale centrée réduite. Pour trouver ce dernier, on utilise le fait que $z_{0,025} = -z_{0,975}$, et dans la table on trouve $z_{0,975} \approx 1,96$.

Donc $q_{0,025} = 100 - 1,96 \times 15 = 70,6$, ce qui signifie que $2,5\%$ des enfants ont un QI inférieur à $70,6$.

4) *C'est plus élevé que la moyenne dans la population qui est 100. Pour savoir si c'est anormal, on va chercher la proportion d'échantillons de taille $n = 50$ pour lesquels le QI moyen $\bar{X}_n$ est supérieur à 120. Puisque $n \geq 30$, le formulaire assure que $\bar{X}_n$ suit une loi normale de moyenne $\mu = 100$ et d'écart-type $\sigma/\sqrt{n} = 15/\sqrt{50} = 2,12$. On calcule ensuite*

$$\begin{aligned} P(\bar{X}_n \geq 120) &= P\left(\frac{\bar{X}_n - 100}{2,12} \geq \frac{120-100}{2,12}\right) \\ &= P(Z \geq 9,43) = 1 - F(9,43). \end{aligned}$$

Mais $9,43$ "sort" largement de la table, ce qui signifie que quasiment 100% des valeurs de Z sont inférieures à $9,43$. Donc $P(\bar{X}_n \geq 120) = 1 - 1 = 0$.

Ce calcul montre qu'en tirant au sort un échantillon de 50 enfants uniformément dans la population, il est quasiment impossible de tomber sur un QI moyen aussi élevé. Cet échantillon n'a sûrement pas été tiré au sort uniformément.