EX 6.1 Solution de le workbook. fiche 6 (1)

EX 6.3

Masse d'une pièce métallique X aléatoire

Modèle : $X \sim \mathcal{N}(\mu, \sigma^2)$ μ son espérance (ou sa moyenne) est inconnue
σ^2 sa variance est inconnue.

pour un échantillon de X de taille $n = 40$ on a observé : $x_1, \dots, x_{40}$

tq $\sum_{i=1}^{40} x_i = 2204$ et $\sum_{i=1}^{40} x_i^2 = 121721$.

1) D'après le cours l'intervalle de confiance pour σ^2 de niveau de confiance $1-\alpha$ est donné par :

$$IC(S_n^2, \alpha) = \left[\frac{n S_n^2}{z_{n-1, 1-\alpha/2}} ; \frac{n S_n^2}{z_{n-1, \alpha/2}} \right]$$

L'intervalle proposé ici est $I = \left[\frac{n S_n^2}{58.1} ; \frac{n S^2}{23.65} \right]$

$I = IC(S_n^2, \alpha)$ pour α tel que $z_{39, 1-\frac{\alpha}{2}} = 58.1$

et $z_{39, \alpha/2} = 23.65$

or dans la table page 4, donnant les quantiles de la variable du χ^2_ν (ν entier) on lit en ligne $\nu = 39$

$z_{38, 0.975} = 56.9$

$z_{40, 0.975} = 59.34$

$\Rightarrow z_{39, 0.975} \in [56.9 ; 59.34]$

et $z_{39, 0.975} \simeq 58.1$.

{ Le calcul avec R : qchisq (0.975, 39) = 58.12.

donc $1 - \frac{\alpha}{2} = 0.975$ d'où $\alpha = 2(1 - 0.975) = 5\%$.

On vérifie également que

$z_{39, 2.5\%} \in [z_{38, 2.5\%} ; z_{40, 2.5\%}]$
$\in [22.88 ; 24.43]$

et $23.65 \in [22.88 ; 24.43]$

On peut calculer exactement $z_{39, 2.5\%}$ avec R ou la calculette et on obtient

$\boxed{z_{39, 2.5\%} = \text{qchisq}(0.025, 39) = 23.65}$

$$I = \left[\frac{nS^2}{58{,}1} ; \frac{nS^2}{23.65}\right] = IC(S_n^2, 5\%)$$ est donc bien l'intervalle de confiance pour σ^2 de niveau de confiance $1-\alpha = 95\%$.

A.N : $s_n^2 = \frac{1}{n}\sum x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2 = \frac{121721}{40} - \left(\frac{2204}{40}\right)^2$

$s_n^2 = 7.015$ et $\boxed{IC(s_{40}^2, 5\%) = [4.8296 ; 11.8647]}$

2) Comme σ^2 est supposé inconnu, d'après le cours l'intervalle de confiance pour μ est donné par, niv $1-\alpha$

$$IC(\bar{X}_n, \alpha) = \left[\bar{X}_n - \frac{S'}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} , \bar{X}_n + \frac{S'}{\sqrt{n}} t_{n-1, 1-\alpha/2}\right]$$

A.N : On calcule d'abord $s'^2 = \frac{n}{n-1} s^2 = \frac{40}{39} \cdot 7.015$

puis $s' = \sqrt{s'^2} =$

Ensuite on cherche dans la table page 3 le quantile d'ordre $p = 1 - \frac{\alpha}{2}$ avec $\alpha = 1\%$ pour l'IC, de niveau 99%

avec $\alpha = 5\%$ pour l'IC de niveau 95%

On lit $t_{38, 0.995} = 2.7116$

$t_{40, 0.995} = 2.7045$

d'où $t_{39, 0.995} \in [2.7045, 2.7116]$

prenons le milieu de l'intervalle pour approcher $t_{39, 0.995}$ soit

$t_{39, 0.995} \simeq 2.708$ $\left\{ \text{Avec R: } t_{39, 0.995} = qt(0.995, 39) = 2.7079. \right.$

Donc l'approximation 2.708 obtenue en lisant la table est de très bonne qualité.

Pour $t_{39, 0.975}$ on procède de même :

$2.0211 = t_{40. 0.975} < t_{39, 0.975} < t_{38, 0.975} = 2.0244$

$\Rightarrow t_{39, 0.975} \simeq \frac{2.0244 + 2.0211}{2} = 2.0228$

Donc pour $\alpha = 1\%$ on utilise $t_{39, 0.995} = 2.708$

et pour $\alpha = 5\%$ on utilise $t_{39, 0.975} = 2.023$.

Comme $s'^2 = 7.1949$, $\bar{x} = 55.1$, $n = 40$

on a

$$IC(55.1, 1\%) = \left[55.1 \mp \sqrt{\frac{7.1949}{40}} \cdot 2{,}706\right]$$

$$\boxed{IC(55.1, 1\%) = [53.95 , 56.25]}$$

et

$$IC(55.1, 5\%) = \left[55.1 \mp \sqrt{\frac{7.1949}{40}} \cdot 2{,}706\right]$$

$$\boxed{IC(55.1, 5\%) = [54.24 , 55.96]}$$

L'intervalle de niveau de confiance 95% est plus precis que celui de niveau 99%, mais en contre partie il est moins fiable (puisque 95% < 99%).

3) Le fabricant annonce que $\sigma = 2.7$ soit $\sigma^2 = 7.29$.

$\sigma^2 \in IC(s^2_{40}, 5\%)$ donc on peut en conclure que l'annonce du fabricant est correcte.

4) Reprenons les calculs d'IC sur μ avec $\underline{\sigma^2 \text{ connu } \sigma^2 = 7.29}$

D'après le cours l'IC de niveau de confiance $1-\alpha$, pour le paramètre inconnu μ est donné par

$$IC(\bar{X}_n, \alpha) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} , \bar{X}_n + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right]$$

pour calculer les intervalles aux niveaux 99% et 95% il faut utiliser

$u_{1-5\%/2} = u_{0.975}$ et $u_{1-1\%/2} = u_{0.995}$

Or on lit page 2 des tables $u_{0.975} = 1.96$ et $u_{0.995} = 2.5758$

d'où A.N : $IC(55.1, 1\%) = \left[55.1 \pm \sqrt{\frac{7.29}{40}}\ 2.5758\right]$

$\boxed{IC(55.1, 1\%) = [54 , 56.2]}$ et $IC(55.1 , 5\%) = \left[55.1 \mp \sqrt{\frac{7.29}{40}}\ 1.96\right]$

$$\boxed{IC(55.1, 5\%) = [54.26 , 55.94]}$$

que ce soit au niveau de confiance 99% ou 95% l'intervalle de confiance calculé avec σ connu est plus precis que celui calculé avec σ inconnu.
C'est normal car devoir estimer σ lorsqu'il est inconnu ajoute de l'alea et donc diminue la precision de l'intervalle de confiance obtenu.

EX6.5

1). a) l'intervalle de fluctuation est deterministe et depend de la taille de l'echantillon et des paramètres du modèle, qui sont donc supposés connus.
Ensuite on regarde si l'echantillon prelevé produit une moyenne empirique $\in$ IF. Si c'est le cas on dira que l'échantillon est conforme au modele posé au niv de confiance $1-\alpha$.

b) l'intervalle de confiance est aléatoire et centré en la moyenne empirique de l'échantillon prelevé.
Il ne depend que des données et du niveau choisi.
Si un fabricant affirme que $\begin{cases} \sigma^2 = \sigma_0^2 \\ \mu = \mu_0 \\ \text{ou } p = p_0 \end{cases}$ alors on s'assurera qu'il dit vrai en verifiant que

μ (resp. σ^2, resp. p) appartient à l'intervalle de confiance (comme cela a eté fait dans l'ex précédent 6.3 q3).

2. a) Dans le cas d'un sondage la proportion de votes attendue pour un candidat est p inconnue et estimée par $\hat{p} = \bar{x}_n = f_n$ pour une estimation ponctuelle ou par $\left[\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, u_{1-\alpha/2}\right]$ pour une estimation par intervalle de confiance de niveau 1 − α.

⚠ pour le problème sur p il faut avoir un échantillon de taille suffisamment grande à savoir tel que $n\hat{p} > 10$ et $n(1-\hat{p}) > 10$.

A.N : $n = 1000$, $\hat{p} = 51\%$ (fréquence des intentions de votes pour le candidat)

$u_{0.975} = 1.96$.

$\boxed{IC(\hat{p}, \alpha) \ni p \text{ avec une probabilité de } 95\%}$.

le calcul de l'IC fournit ici :

$$\left[0.51 \mp \sqrt{\frac{0.51 \quad 0.49}{1000}}\; 1.96\right] = \boxed{[47{,}9\% \; ; \; 54{,}1\%] = IC(\hat{p}, 5\%)}$$

Remarquez que l'amplitude est de presque 6% soit une précision de ±3%. Donc pour un sondage sur 1000 personnes une augmentation (ou diminution) d'intentions de votes de 3% n'est pas significative au niveau de confiance de 95%.

Attention donc aux interprétations des sondages politiques où ⊖ de 3% d'écart entre 2 candidats (ou entre 2 sondages pour le m̄ candidat) n'est pas significatif d'une progression (ou l'inverse) du candidat.

b) le vote a eu lieu et le candidat a obtenu $p = 51\%$ de votes.

Dans un echantillon de taille $n = 1000$ on sait que $F_n = \bar{X}_n$ a une proba ~~de~~ $1-\alpha$ de se trouver dans $IF(p, \alpha) = \left[p \pm \sqrt{\frac{p(1-p)}{n}}\, u_{1-\alpha/2}\right]$ si $np > 10$ et $n(1-p) > 10$.

A.N : on a bien $1000 . 51\% > 10$ et $1000 . 49\% > 10$

et $IF(51\%, 5\%) = \left[0.51 \pm \sqrt{\frac{0.51 . 0.49}{1000}}\ 1.96\right]$

Par exemple : $= [47{,}9\% ; 54{,}1\%]$.

Si dans un bureau de vote comptant 1000 votant le candidat a obtenu 45% des voix. Comme $45\% \notin IF$ On pourra dire que dans ce bureau, le score obtenu par le candidat n'est pas conforme à celui qu'il a obtenu parmi tous les votants

EX6.4

Soit p le pourcentage de reussite inconnu d'une operation chirurgicale.

Sur $n = 400$ operations on a observé 49 échecs soit 400-49 reussites.

Notons X la variable parente de l'echantillon $X_1 \dots, X_n$ observé

Modèle : $X \sim B(p)$ et on veut estimer $E(X) = p$ inconnu

Estimateur de $E(X)$: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = F_n$

Intervalle de conf de niveau approx $1-\alpha$:

$$IC(\bar{X}_n, \alpha) = IC(F_n, \alpha) = \left[F_n \mp \sqrt{\frac{F_n(1-F_n)}{n}}\, u_{1-\alpha}\right]$$

si $nf_n > 10$ et $n(1-f_n) > 10$.

Pour l'échantillon observé on obtient

$$\hat{\mu} = \hat{p} = \bar{x}_n = f_n = \frac{400-49}{400} = 87,75\%$$

et $IC(f_n, 5\%) = \left[0.8775 \mp \sqrt{\frac{0.8775(1-0.8775)}{400}} \cdot 1.96\right]$

puisque $u_{0.975} = 1.96$

et $n\hat{f}_n = 400 \times 0.8775 > 10$

et $n(1-\hat{f}_n) = 400 \times 0.1225 > 10$.

$$\boxed{IC(f_n, 5\%) = [84,53\%\ ,\ 90,96\%]}$$

2) Quelle taille d'échantillon doit-on prendre pour que $\bar{X}_n = F_n \in$ à un intervalle de confiance de niveau 95% et de précision $\pm 1\%$?

On cherche n tel que la largeur de l'intervalle soit $\leq 2\%$: soit n tel que

$$2\sqrt{\frac{\hat{f}_n(1-\hat{f}_n)}{n}}\, u_{0.975} \leq 0.02.$$

On admettra que $\hat{f}_n = 87,75\% = \hat{p}$, quel que soit la taille de l'échantillon.

et on cherche ensuite n tq $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, 1.96 \leq 0.02$

soit $\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)^2 \leq \left(\frac{0.02}{2 \cdot 1.96}\right)^2$

$$\Leftrightarrow \quad \frac{\hat{p}(1-\hat{p})}{n} \leq \frac{10^{-4}}{(1.96)^2}$$

$$\Leftrightarrow \quad n \geq (1.96)^2\, \hat{p}(1-\hat{p}) \cdot 10^4$$

$$\Leftrightarrow \quad n \geq (1.96)^2\, 0.8775(1-0.8775) \cdot 10^4$$

$$\Leftrightarrow \quad n \geq 4129,5$$

Donc pour avoir une précision de $\pm 1\%$ sur l'IC à 95%, la clinique aurait du observer 87,75% de réussite sur 4130 opérations

3) ** hors programme mais pour aller plus loin

On cherche N tel que ~~la proba~~ de nombre d'opérations réussies soit ~~dep~~ $400-N$ avec une proba de 99%.

soit $\mathbb{P}(\sum X_i \geq 400-N) = 99\%$.

$$\Leftrightarrow \quad \mathbb{P}(\underbrace{400-\sum X_i}_{\text{nbre d'operations ratées}} \leq N) = 99\%.$$

$$\Leftrightarrow \quad \mathbb{P}(\text{nbre d'operations ratees} > N) = 1\%.$$

$$\mathbb{P}(\sum X_i \geq 400-N) = 1-\mathbb{P}(\sum X_i < 400-N) = 99\%.$$

$$\Leftrightarrow \quad \mathbb{P}(\sum X_i < 400-N) = 1\%.$$

Rappelons que comme $np > 10$ et $n(1-p) > 10$.

$\sum X_i \sim B(n,p)$ peut être approchée par une $\mathcal{N}(np, np(1-p))$

Donc $\mathbb{P}(\sum X_i < 400-N) \simeq \mathbb{P}(Y < 400-N)$ pour $Y \sim \mathcal{N}(np, np(1-p))$

Soit $Z = \dfrac{Y-np}{\sqrt{np(1-p)}}$ la centrée réduite de Y et $W = \dfrac{Y-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}} \sim \mathcal{N}(0,1)$

$$\mathbb{P}(Y < 400-N) = \mathbb{P}\left(Z < \frac{400-N-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}}\right) \overset{(*)}{=} \phi\left(\frac{400-N-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}}\right)$$

et on cherche donc N tq

$$\phi\left(\frac{400-N-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}}\right) = 10^{-2} \Leftrightarrow \frac{400-N-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}} = \phi^{-1}(10^{-2})$$

$$\Leftrightarrow \quad 400-N-n\hat{p} = \sqrt{n\hat{p}(1-\hat{p})}\,\phi^{-1}(10^{-2})$$

$$\Leftrightarrow \quad N = 400-n\hat{p}-\sqrt{n\hat{p}(1-\hat{p})}\,\phi^{-1}(10^{-2})$$

$$\Leftrightarrow \quad N \simeq 64,25$$

Donc max $N = 64$ operations ratees sur 400. sinon l'assureur n'assure pas la clinique.

(*) $\mathbb{P}(Z<t) \simeq \mathbb{P}(W<t)$ car W suit approx une $\mathcal{N}(0,1)$

Resultat à admettre