Saïd El Melhaoui

Échantillonnage & Estimation

Notes de Cours, Exercices
& Examens Corrigés

Edition 2026

Avant-propos

Le présent ouvrage est destiné à tous les étudiants abordant des études d'économie et de gestion dans les facultés de droit ou les grandes écoles. Son contenu concerne le module intitulé «Echantillonnage et Estimations» selon la nouvelle architecture pédagogique de la filière des Sciences Économiques. Il s'agit d'un recueil d'exercices corrigés, qui ont été expérimentés dans les séances de TD à à la faculté des Sciences Juridiques, Économiques et Sociales d'Oujda et à l'École National de Commerce et de Gestion.

L'ouvrage contient des notes de cours qui résument les différents concepts et formules étudiée dans le cours. Un QCM est ensuite donné pour faciliter l'assimilation des différents concepts. Enfin, un certain nombre d'exercices sont proposés ainsi que leurs corrigés détaillés.

Nous incitons le lecteur à chercher patiemment la solution des différents exercices sans céder à la tentation de recourir hâtivement aux corrigés proposés. L'auteur remercie par avance le lecteur pour les suggestions, les corrections ou les critiques qu'il voudra lui adresser sur l'adresse e-mail s1.elmelhaoui@ump.ac.ma.

Table des matières

I. Notes de Cours

`Ne dites pas « J'ai trouvé la vérité », mais plutôt`
`« J'ai trouvé une vérité ».      KHALIL GIBRAN, LE PROPHÈTE`

1. Théorèmes Asymptotiques (Rappel)

(a) ***Propriétés des opérateurs Espérance et Variance.*** Soit X et Y deux v.a. et a et b deux nombres réels, alors

(i) $E[aX + bY] = aE[X] + bE[Y]$.

(ii) $Var(aX \pm bY) = a^2Var(X) + b^2Var(Y) \pm 2ab\,Cov(X,Y)$.

(iii) Si en plus, X et Y sont <u>indépendantes</u>, alors,

$$Var(X+Y) = Var(X)+Var(Y) \text{ et } Var(X-Y) = Var(X)+Var(Y).$$

(b) ***Théorèmes d'addition.*** Soit X et Y deux v.a. <u>indépendantes</u>

i. Si $X \sim \mathcal{B}(n_1, p)$ et $Y \sim \mathcal{B}(n_2, p)$ alors $X + Y \sim \mathcal{B}(n_1 + n_2, p)$.

ii. Si $X \sim \mathcal{P}(\lambda_1)$ et $Y \sim \mathcal{P}(\lambda_2)$ alors $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

iii. Si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ alors

$$aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2).$$

iv. Si $X \sim \chi^2_{n_1}$ et $Y \sim \chi^2_{n_2}$ alors $X + Y \sim \chi^2_{n_1+n_2}$.

(c) ***Théorèmes asymptotiques.***

(i) ***Inégalité de Markov.*** Soit X une v. a. réelle supposée presque sûrement positive ($P(X \geq 0) = 1$). Alors

$$\forall \alpha > 0, \qquad P(X \geqslant \alpha) \leqslant \frac{E[X]}{\alpha}.$$

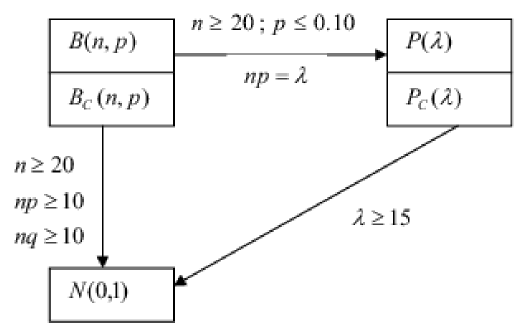
(ii) ***Théorème de Bienaymé-Tchébicheff.*** Soit X une variable aléatoire de moyenne μ et de variance σ^2

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

(iii) ***Loi des grand nombres.*** Soient $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d., d'espérance μ, alors $\overline{X}$ tend en probabilité vers μ lorsque $n \longrightarrow +\infty$:

$$\forall \varepsilon > 0, \quad \lim_{n \to +\infty} P(|\overline{X} - \mu| > \varepsilon) = 0.$$

(iv) **Quelques approximations**. $B_c(n,p)$ et $P_c(\lambda)$ désignent resp. la loi binomiale et celle de Poisson centrées réduites :



(v) ***Théorème Central Limite (TCL).*** Soient $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes identiquement distribuées (i.i.d.), d'espérance μ et de variance $\sigma^2 < \infty$. Posant $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

On a

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1).$$

N. B. *Le* **TCL** *est généralement formulé pour une moyenne* $\overline{X}$, *mais on peut le formuler pour une somme :* $X_T = \sum_{i=1}^{n} X_i$ *:*

$$\frac{X_T - n\mu}{\sigma\sqrt{n}} \approx \mathcal{N}(0,1).$$

2. Échantillonnage

A. Échantillonnage aléatoire simple

Lors de ***l'échantillonnage aléatoire simple*** l'échantillon est sélectionné de façon à ce tous les échantillons possibles de tailles fixées n ont la même probabilité d'être choisi.

1. ***Échantillon Simple Avec Remise (ESAR).*** On choisi successivement d'une façon aléatoire n unités de la population, une fois qu'un élément a été inclus dans l'échantillon, il est remis dans la population et peut être choisi plusieurs fois.

   (a) **Estimation de la moyenne**. X une v.a., d'espérance μ sur une population $\mathcal{P}$ et $X_1, \ldots, X_n$ un ESAR.

Estimateur de μ	**Espérance**	**Variance (Erreur)**
$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$E[\overline{X}] = \mu$	$Var(\overline{X}) = \sigma^2/n$

Distribution d'échantillonnage	
$X \sim \mathcal{N}(\mu, \sigma^2)$	$X \sim \mathcal{L}oi(\mu, \sigma^2)$ et $n \geqslant 30$
$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$	$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1)$

   (b) **Estimation de la proportion**. Soit p la proportion d'un certain caractère dans une population $\mathcal{P}$. Soit $X_1, \ldots, X_n$ un ESAR tel que $X_i = 1$ si la ième unité ait le caractère et $X_i = 0$ sinon.

Estimateur de p	**Espérance**	**Variance**
$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$E[\bar{p}] = p$	$Var(\bar{p}) = p(1-p)/n$

Distribution d'échantillonnage $n \geqslant 30$ ou $n \geqslant 20,\quad np \geqslant 10$ et $n(1-p) \geqslant 10$
$\frac{\bar{p} - p}{\sqrt{p(1-p)/n}} \approx \mathcal{N}(0,1)$

2. ***Échantillon Simple Sans Remise (ESSR)***. On choisi successivement d'une façon aléatoire n unités de la population, une fois qu'un élément a été inclus dans l'échantillon, il est retiré de la population et ne peut être choisi une autre fois.
   Soit $X_1, \dots, X_n$ un ESSR issu d'une population de taille N $(n \leq N)$.

Estimateur	**Espérance**	**Variance**
$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$	$E[\overline{X}] = \mu$	$Var(\overline{X}) = \left(\frac{N-n}{N-1}\right) \frac{\sigma^2}{n}$

Distribution d'échantillonnage $n/N \leq 5\%$ et $n \geqslant 30$
$\frac{\overline{X} - \mu}{Var(\overline{X})} \approx \mathcal{N}(0, 1)$

N. B. *La proportion $f = n/N$ est dite le* **taux de sondage**. *Lorsque $n \geq 30$ et $f \leq 5\%$ on peut assimiler un ESSR à un ESAR.*

B. Autres types d'échantillonnage

- ***Échantillonnage stratifié.*** Échantillonnage aléatoire dans lequel la population est tout d'abord divisée en strates et un échantillon aléatoire simple est ensuite sélectionné parmi chaque strate.
  Utilisé lorsque on distingue plusieurs groupes homogènes : variances internes petites et variances externes grandes.
- ***Échantillonnage par grappes.*** Échantillonnage aléatoire dans lequel la population est tout d'abord divisée en grappes et un échantillon aléatoire simple de grappes est ensuite sélectionné ; tous les unités des grappes choisi sont sélectionnées.
  Utilisé lorsque on observe plusieurs groupes hétérogènes : variances internes grandes et variances externes petites.
- ***Échantillonnage de commodité.*** Échantillonnage non aléatoire dans lequel les éléments de l'échantillon sont sélectionnés en fonction de leurs commodité (disponibilité, facilité).
- ***Échantillonnage subjectif.*** Échantillonnage non aléatoire dans lequel les éléments de l'échantillon sont sélectionnés en fonction des

croyances de la personne qui fait l'étude : **Sondage par la méthode des quotas**.

3. Estimation

A. Méthodes d'estimation

1. ***Méthode des moments.*** Les estimateurs $\hat{\theta}_1, \ldots, \hat{\theta}_K$ des paramètres $\theta_1, \ldots, \theta_K$ sont déterminés en résolvant le système de K équations à K inconnus :

$$\begin{array}{rcl} \mu_1(\theta_1, \ldots, \theta_K) & = & m_1 \\ \mu_2(\theta_1, \ldots, \theta_K) & = & m_2 \\ .... & = & ... \\ \mu_K(\theta_1, \ldots, \theta_K) & = & m_K \end{array}$$

où les $\mu_k = E(X^k)$ sont les moments simples de X et les m_k sont les moments empiriques simples :

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k, \quad k = 1, \ldots, K.$$

2. ***Méthode du Maximum de Vraisemblance.*** Soient $X_1, ..., X_n$ un i.i.d. de loi P_θ dans le cas discret ou de densité f_θ dans le cas continu :
   - ***Fonction de vraisemblance.*** C'est la probabilité de réalisation de l'échantillon :
     - Cas discret : $L(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} P_\theta[X_i = x_i]$.
     - Cas continu : $L(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i)$.
   - **L'estimateur du Maximum de Vraisemblance** est la valeur $\hat{\theta} \in$ choisit dans Θ qui assure à l'échantillon observée la plus grande '*vraisemblance*' :

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

   Si $L(\theta)$ est différentiable et possède au point $\hat{\theta}$ un maximum sur Θ alors $\hat{\theta}$ est solution de ***l'équation du M. V.*** :

$$\frac{dL(\theta)}{d\theta} \mid_{\theta=\hat{\theta}} = 0.$$

N. B. *Si $\hat{\theta}$ correspond au maximum de $L(\theta)$, il correspond aussi au maximum de la* **fonction de log-vraisemblance** $L^*(\theta) = \log(L(\theta))$.

B. Propriétés des estimateurs

1. Un estimateur $\hat{\theta}$ de θ est ***sans biais*** si $B(\hat{\theta}) = E[\hat{\theta}] - \theta = 0$.
2. Un estimateur $\hat{\theta}$ de θ est **asymptotiquement sans biais** si
$$\lim_{n \mapsto \infty} E(\hat{\theta}_n) = \theta.$$
3. Soit $\hat{\theta}_n$ un estimateur de θ, il est ***convergent*** si
$$\lim_{n \mapsto +\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1.$$
Un estimateur sans biais $\hat{\theta}$ de θ, tel que $V(\hat{\theta}) \overset{\longmapsto}{n} 0$ est un estimateur convergent.
4. ***Efficacité relative et absolue d'un estimateur.***
   - Le ***MSE*** (Mean Squared Error) d'un estimateur $\hat{\theta}$ d'un paramètre θ est :
   $$\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = V(\hat{\theta}) + B^2(\hat{\theta})$$
   N. B. *Si l'estimateur est sans biais alors*
   $$MSE(\hat{\theta}) = V(\hat{\theta}).$$
   - Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont des estimateurs quelconque de θ, ***l'efficacité relative*** de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$ est définie par
   $$E_r(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{MSE}(\hat{\theta}_2)}{\text{MSE}(\hat{\theta}_1)}.$$
   - $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si
   $$E_r(\hat{\theta}_1, \hat{\theta}_2) \geq 1.$$
   - Un estimateur $\hat{\theta}$ de θ est ***efficace*** si ça variance est minimale ; c.à.d. elle atteint la *borne de Cramer-Rao* :
   $$Var(\hat{\theta}) = [I(\theta)]^{-1}$$
   où $I(\theta) = E[(\frac{dL^*(\theta)}{d\theta})^2] = E[\frac{-d^2L^*(\theta)}{d\theta^2}]$ est *l'information de Fisher.*

C. Estimation par intervalles de confiance

Intervalle de confiance pour la moyenne

Soit $X \sim \mathcal{L}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$. μ est le ***paramètre d'intérêt*** et σ^2 est un ***paramètre de nuisance***. Pour un échantillon ESAR $\{X_1, X_2, \ldots, X_n\}$ de X, soit $\hat{\mu} = \overline{X}$ un estimateur de μ.

1. **Cas des petits échantillons ($n < 30$).** On suppose que X suit une loi **normale**.
   - **Cas où σ est connu.**

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

   - **Cas où σ est inconnu.**

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]$$

   N. B. *La variance empirique :*

$$s = \frac{1}{n}\sum_{i}^{n}(X_i - \overline{X})^2$$

   est biaisé, un estimateur sans biais est

$$S^2 = \frac{1}{n-1}\sum_{i}(X_i - \overline{X})^2.$$

   Remarquons que $S/\sqrt{n} = s/\sqrt{n-1}$

2. **Cas des grands échantillons ($n \geq 30$).** La loi de X est supposée quelconque.
   - **Cas où σ est connu.**

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

   - **Cas où σ est inconnu.**

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Intervalle de confiance pour la variance

Soit $X \sim \mathcal{L}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$. σ^2 est le ***paramètre d'intérêt*** et μ est un ***paramètre de nuisance***. Pour un échantillon ESAR $\{X_1, X_2, \ldots, X_n\}$ de X, soit $\hat{\sigma}^2$ un estimateur de σ^2.

1. **Cas des petits échantillons ($n < 30$).** On suppose que X suit une loi **normale**.
   - **Cas où μ est connu.** On prend $\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \mu)^2$ l'intervalle de confiance est :

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{n\hat{\sigma}^2}{\chi^2_{n;1-\alpha/2}}, \frac{n\hat{\sigma}^2}{\chi^2_{n;\alpha/2}}\right].$$

   - **Cas où μ est inconnu.** On prend $\hat{\sigma}^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}\right].$$

2. **Cas des grands échantillons ($n \geq 50$).** La loi de X est supposée quelconque et μ connu ou inconnu :

$$IC_{1-\alpha}(\sigma^2) = \left[S^2 - z_{1-\alpha/2}\sqrt{(\hat{\mu}_4 - S^4)/n}, S^2 + z_{1-\alpha/2}\sqrt{(\hat{\mu}_4 - S^4)/n}\right],$$

où

$$\hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4.$$

N. B. *Puisque n est assez grand on a $\overline{X} \xrightarrow{n} \mu$, $S^2 \simeq s^2 \xrightarrow{n} \sigma^2$ et $\hat{\mu}_4 \xrightarrow{n} E\left[(X-\mu)^4\right] = \mu_4$.*

Intervalle de confiance pour la proportion

Soit X une v.a. qui suit une loi de Bernoulli $B(p)$ où $0 < p < 1$ est le paramètre d'intérêt. Soit un échantillon ESAR $\{X_1, X_2, \ldots, X_n\}$ de X dont la taille $n \geq 30$ (cas des grands échantillons). L'intervalle de confiance de p est :

$$IC_{1-\alpha}(p) = \left[\bar{p} - z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right].$$

où

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} X_{i=1}^{n} = \text{ Fréquence empirique du succès.}$$

Précision d'une estimation par intervalle de confiance

- ***L'incertitude absolue*** d'un intervalle de confiance bilatérale (ℓ_1, ℓ_2) est son rayon :
$$I_a = \frac{\ell_2 - \ell_1}{2} = \text{ Rayon de I.C..}$$
- ***L'incertitude relative*** d'un intervalle de confiance bilatérale (ℓ_1, ℓ_2) est :
$$I_r = \frac{I_a}{\hat{\mu}} = \frac{\ell_2 - \ell_1}{\ell_2 + \ell_1}.$$
- L'incertitude relative ou absolue est proportionnelle à l'écart-type-population σ et à au degré de confiance $1-\alpha$ et inversement proportionnelle à la racine de la taille échantillon n.
- Afin de contrôler l'incertitude d'estimation de la moyenne μ par $\Delta\mu$; si σ est supposée connu ; il faut que $n \geq (z_{1-\alpha/2}\sigma/\Delta\mu)^2$.
- Afin de contrôler l'incertitude d'estimation de la proportion p par Δp il faut que $n \geq (z_{1-\alpha/2}/2\Delta p)^2$.

4.Tests d'hypothèses

A. Méthodologie des tests

- Réaliser un test d'hypothèse paramétrique consiste à émettre une hypothèse dite ***hypothèse nulle*** sur un paramètre population θ et à décider d'accepter ou de refuser cette hypothèse à partir d'un échantillon.
- L'hypothèse H_1 alternative à H_0 est dite la ***contre hypothèse***, selon H_1 on peut définir un test bilatéral ou unilatéral :

Test unilatéral gauche	Test bilatéral	Test unilatéral droite
$\begin{cases} H_0 : & \theta \geq \theta_0 \\ H_1 : & \theta < \theta_0 \end{cases}$	$\begin{cases} H_0 : & \theta = \theta_0 \\ H_1 : & \theta \neq \theta_0 \end{cases}$	$\begin{cases} H_0 : & \theta \leq \theta_0 \\ H_1 : & \theta > \theta_0 \end{cases}$

- Un test est une statistique qui prend deux valeurs : 1 lorsque on accepte une hypothèse H_0 : $\overline{RH_0}$; et 0 lorsque on refuse H_0 : RH_0.

- Le test est basé sur une statistique T dite ***statistique-test*** dont on établit la loi sous l'hypothèse H_0.
- La ***règle de décision*** est issue de la comparaison de la valeur calculée T_{cal} de la statistique T aux quantiles extrêmes de la loi hypothétique sous H_0 de T.
- Dans l'approche de **Neyman-Pearson** on fixe à priori à $\alpha := 1\%,\ 5\%,\ 10\%$ comme ***erreur de première espèce*** qui est 'la probabilité de refuser à tort une bonne hypothèse H_0' et on essaye de minimiser **l'erreur de deuxième espèce** : β = 'la probabilité d'accepter à tort une mauvaise hypothèse H_0'
- $\alpha = P(RH_0|H_0)$ est appelée aussi le **niveau du test** et $1-\beta = P(RH_0|H_1)$ est dite ***la puissance du test***.

B. Tests pour une moyenne

Hypothèses		Règle	La stat. test T et sa loi sous H_0		
H_0	H_1	RH_0 si	$n < 30$ & X Normale		$n \geqslant 30$
			σ connu	σ inconnu	
$\mu \leq \mu_0$	$\mu > \mu_0$	$T > q_{1-\alpha}$	$T = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$T = \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$T = \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\|T\| > q_{1-\alpha}$	$T \sim \mathcal{N}(0,1)$	$T \sim t_{n-1}$	$T \approx \mathcal{N}(0,1)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$T < -q_{1-\alpha}$			

q_α est le quantile d'ordre α de la loi de la statistique test correspondant au problème de test.

C. Tests pour une variance

Hypothèses		Règle	La stat. test T et sa loi sous H_0	
H_0	H_1	RH_0 si	$n < 50$ & X Normale	$n \geqslant 50$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$T > q_{1-\alpha}$	$T = \dfrac{(n-1)\,S^2}{\sigma_0^2}$	$T = \dfrac{S^2-\sigma_0^2}{\sqrt{(\hat{\mu}_4 - S^4)/n}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$T \notin [q_{\alpha/2}, q_{1-\alpha/2}]$	$T \sim \chi^2_{n-1}$	$T \approx \mathcal{N}(0,1)$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$T < q_\alpha$		

N. B. *Dans le cas des petit échantillon, si μ est spécifiée alors on substitue* *à S^2 la statistique $\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \mu)^2$ et on utilise la loi χ^2 à n degré de liberté.*

D. Tests pour une proportion

Hypothèses		Règle	La stat. test T et sa loi sous H_0
H_0	H_1	RH_0 si	$n \geq 30$
$p \leq p_0$	$p > p_0$	$T > q_{1-\alpha}$	$T = \dfrac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
$p = p_0$	$p \neq p_0$	$\lvert T\rvert > q_{1-\alpha/2}$	$T \approx \mathcal{N}(0,1)$
$p \geq p_0$	$p < p_0$	$T < -q_{1-\alpha}$	

N. B. *Les tests restent valables pour les petits échantillons sous les conditions du Théorème de Moivre-Laplace. Dans le cas échéant il faut utiliser la loi exacte de $\overline{p}$ c.à.d. la loi binomiale.*

E. Test du Chi-deux

E.1. Test d'ajustement

Ayant observé une variable aléatoire X sur un échantillon de taille n, on test si la distribution des observations $(A_j, n_j)_{1\leq j\leq J}$ indique que X est distribué suivant une distribution théorique fixée telle que $p_{j0} = P(X \in A_j)$. Avec A_j est un caractère (classe, valeur, modalité) dont n_j est son effectif observé et $n_j^* = np_{j0}$ est son effectif théorique.

Hypothèse		La stat. test D et sa loi sous H_0	RH_0 si
H_0	H_1		
$\forall 1 \leq j \leq J$ $p_j = p_{j0}$	$\exists$ au moins un $p_j \neq p_{j0}$	$\displaystyle\sum_{j=1}^{J} \frac{(n_j - n_j^*)^2}{n_j^*} \approx \chi^2_{J-1}$	$D > \chi^2_{J-1;1-\alpha}$

N.B. La condition d'utilisation est que : $n \geq 30$, $\forall j,\quad n_j^* \geq 1$ et au moins 80% des n_j^* sont supérieurs ou égaux à 5.

E.2. Test d'indépendance

Afin de tester l'indépendance entre deux variables nominales X et Y, on observe un échantillon $(x_i, y_i)_{1\leq i\leq n}$ de taille n. On résume les données dans un tableau de contingence (n_{jk}) ; $1 \leq j \leq J$ $1 \leq k \leq K$. L'effectif total est $n = \sum_j \sum_k n_{jk}$ et Les effectifs théoriques, sous l'hypothèse d'indépendance sont $n^*_{jk} = n_{j.}n_{.k}/n$.

Hypothèse		La stat. test D	RH_0 si
H_0	H_1		
Indépendance	Dépendance	$\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{jk}-n^*_{jk})^2}{n^*_{jk}}$	$D > \chi^2_{(J-1)(K-1);1-\alpha}$

N.B. La condition d'utilisation est que : $n \geq 30$, $n^*_{jk} \geq 1$ et au moins 80% des n^*_j sont supérieurs ou égaux à 5.

E.3. Test d'homogénéité

Test d'homogénéité consiste à tester si deux échantillons se trouvent tirés de la même population et de là régis par la même loi. On l'utilise par exemple lorsque : on a à comparer entre deux distribution de consommation des individus de deux CSP, on a à comparer avant et après un événement, l'application d'une méthode ou d'un traitement...Le principe de l'élaboration du test est le même que celui de l'indépendance, en effet d'homogénéité est une forme d'indépendance.

Une banque utilise dans deux agences deux méthodes différentes d'accueil des clients, dans la première elle utilise une file d'attente classique et dans l'autre une file d'attente électronique. La satisfaction de deux échantillons de clients est résumé dans le tableau suivant :

Degré de satisfaction	Très satisfait	satisfait	pas satisfait
Accueil classique	280	210	110
Accueil électronique	220	90	90

Les deux techniques d'accueil ont il le même effet ?

II. Exercices

Échantillonnage

Exercice 1

Afin d'estimer leurs moyennes respectives, on échantillonne deux populations. On utilise un ESAR de taille n_1 pour la population P_1, qui présente un écart type égal à σ_1. Pour la population P_2 dont l'écart type vaut $\sigma_2 = 2\,\sigma_1$, on prend un ESAR de taille $n_2 = 2\,n_1$. Pour quel échantillon l'estimation de la moyenne de la population est-elle plus précise ?

Exercice 2

Des briques sont fabriquées selon une technique qui assure un poids normalement distribué d'espérance 1.6 kg et d'écart type 30 g. On prélève un échantillon aléatoire simple avec remise, de moyenne $\bar{X}$.

1. Quel est l'écart type de $\bar{X}$, lorsque sa taille vaut 50 et lorsqu'elle vaut 100 ?
2. Quelle est la loi de $\bar{X}$?
3. à partir de quelle taille de l'échantillon, on aura plus de 95% de chance, pour que la moyenne $\bar{X}$ s'écartera-t-elle de sa valeur espérée de moins de 20 g ?
4. Supposons qu'on prélève deux échantillons indépendants I et II de tailles respectives 50 et 100 briques. On considère leurs deux moyennes respectives $\bar{X}_I$ et $\bar{X}_{II}$. Quelle est la loi de $\bar{X}_I - \bar{X}_{II}$? Déduisez la probabilité que les moyennes des deux échantillons diffèrent de plus de 10 g ?

Exercice 3

Selon la magazine *USA Today*, le nombre moyen des jours par an passés sur les routes pour un représentant commercial est égal à 115. L'écart type est de 60 jours par an. Supposez que ces résultats soient associés à la population des représentants commerciaux et qu'un échantillon de 50 représentants soit sélectionné.

1. Quelle est la valeur de l'écart type de la moyenne ?
2. Quelle est la probabilité que la moyenne d'échantillon soit supérieur à 115 jours par an ?
3. Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 5 jours de la moyenne de la population ?
4. Quelle serait la probabilité de la question (3) si la taille d'échantillon était 100 ?

Exercice 4

Trois firmes ont des inventaires différents par leur taille. La firme A a une population de 2000 pièces, La firme B a une population de 5000 pièces et la firme C a une population de 10 000 pièces. L'écart type de la population pour le coût d'une pièce est de $\sigma = 144$.

1. En utilisant le facteur d'exhaustivité, calculez l'erreur type pour chacune des trois firmes, étant donné que les échantillons sélectionnés sont de type ESSR et de taille 50.
2. Quelle est la probabilité que pour chaque firme, la moyenne d'échantillon $\bar{X}$ s'écarte au plus de ± 25 de la moyenne de la population μ.

Exercice 5

Une série de fabrications ne peut pas être expédiée si un échantillon de 100 pièces contient plus de 5% de pièces défectueuses. Si une série de fabrication a une proportion de pièces défectueuses égale à 10%, quelle est la probabilité que cette série de fabrication soit expédiée ?

Estimation

Exercice 6

On s'interesse à la non réussite des entreprises au cours de l'année 2014. Notons p la probabilité de faillite d'une entreprise en cette année 2014. Pour étudier ce paramètre p, on a prélevé un échantillon aléatoire simple avec remise de n entreprises en activité au début de l'année 2006, et on a relevé pour chaque entreprise si oui ou non une faillite avait eu lieu au cours de l'année.

1. Déterminez l'estimateur de p obtenu par : (a) la méthode des moments ; (b) par la méthode du maximum de vraisemblance.
2. Recherchez les propriétés de ces estimateurs (biais, convergence, efficacité).

Exercice 7

L'institut national de statistique s'est intéressé au nombre moyen μ d'accidents à un certain endroit, pendant les heures de pointes. Pour étudier ce paramètre μ, une étude a été menée pendant n jours sur une route entre 8h et 9h ; le nombre d'accident X a été relevé chaque jour entre ces heures, ce qui donne lieu aux observations $X1, ..., Xn$.

1. Déterminez l'estimateur de μ obtenu par (a) la méthode des moments ; (b) la méthode du maximum de vraisemblance
2. Recherchez les propriétés de ces estimateurs (biais, convergence, efficacité).

Exercice 8

Soit un échantillon aléatoire simple de taille n, $(X1, ..., Xn)$, distribué suivant une loi géométrique $\mathcal{G}(p)$. Déterminez, à partir de cet échantillon, les estimateurs de p obtenus par la méthode des moments et par la méthode du maximum de vraisemblance.
Rappel : Si $X \sim \mathcal{G}(p)$ alors :

$$P[X = x_i] = p\,(1-p)^{x_i} \quad ; \quad x_i = 1, 2, \dots$$

Exercice 9

En 1897, l'économiste suisse Vilfredo Pareto (1848-1923), professeur d'économie politique à l'université de Lausanne, eut l'idée de modéliser la loi des revenus en postulant que le nombre relatif de personnes dont le revenu dépasse une valeur x est inversement proportionnel à une puissance de x. La définition suivante fut adoptée : une variable aléatoire X, absolument continue, suit une loi de Pareto de paramètres α et θ si sa densité est

$$f(x;\alpha,\theta) = kx^{-\alpha}\mathbb{I}_{[\theta,+\infty[}(x), \qquad (\theta > 0, \alpha > 1).$$

On suppose ici que α est connu. Estimez θ par la méthode du maximum de vraisemblance.

Exercice 10

Soit un échantillon aléatoire simple de taille n, $(X1, ..., Xn)$, distribué suivant une loi exponentiel de paramètre $\theta > 0$:

$$f(x,\theta) := \begin{cases} \frac{1}{\theta}e^{-x/\theta}, & x \geq 0\,; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminez l'estimateur de θ obtenus par la méthode des moments
2. Déterminez l'estimateur de θ obtenus par la méthode du maximum de vraisemblance.
3. Déterminez les propriétés de ces estimateurs.

Exercice 11

Soit une population normale de moyenne μ inconnue et d'écart-type $\sigma = 1.5$. On y a prélevé un échantillon aléatoire simple d'effectif $n = 9$. Les observations sont les suivantes :

24.3 ; 24.7 ; 23.2 ; 23.8 ; 24.5 ; 26.0 ; 25.4 ; 24.8 ; 23.8.

Trouvez un intervalle de confiance pour μ au niveau de confiance 95%.

Exercice 12

Soient deux populations normales, la première de moyenne μ_1 inconnue et d'écart-type $\sigma_1 = 1.5$ et la seconde de moyenne μ_2 inconnue et d'écart-type $\sigma_2 = 3$. Un premier échantillon aléatoire simple d'effectif $n_1 = 9$ est extrait de la population 1 et on observe $\bar{x}_1 = 24.5$. Un second échantillon aléatoire simple d'effectif $n_2 = 12$ est extrait de la population 2 et on observe $\bar{x}_2 = 23.5$. On suppose ces deux échantillons sont indépendants.
Trouvez un intervalle de confiance pour la différence $\mu_1 - \mu_2$ au niveau de confiance 95%.

Exercice 13

La société Butler County Bank désire estimer l'écart entre le solde moyen des cartes de crédit dans deux de ses succursales. Les statistiques descriptives des échantillons aléatoires indépendants des clients ayant une carte de crédit sont présentés dans le tableau ci-dessous.

Succursale 1	Succursale 2
$n_1 = 32$	$n_2 = 36$
$\bar{x}_1 = 500$	$\bar{x}_2 = 375$
$S_1 = 150$	$S_2 = 130$

1. Construisez un estimateur de l'écart entre les deux soldes moyens ;
2. Construisez un intervalle de confiance au niveau de confiance 99% pour l'écart entre les deux soldes moyens.
3. Déterminez l'incertitude absolu de l'estimation ainsi que son incertitude relative.

Exercice 14

Un comptable pense que les problèmes de liquidité d'une entreprise sont une conséquence directe de l'encaissement lent des comptes fournisseurs. Le comptable prétend qu'au moins 70% des comptes fournisseurs datent de plus de deux mois. Un échantillon de 120 comptes fournisseurs a révélé que 88 dataient de plus de deux mois.

1. Quelle est l'estimation ponctuelle de la proportion p de comptes fournisseurs datant de plus de deux mois.
2. Donnez un intervalle de confiance à 99% pour la vraie proportion p de comptes fournisseurs datant de plus de deux mois.
3. Déterminez la taille minimum n de l'échantillon à prélevé pour que l'incertitude absolue de l'estimation (au niveau de confiance 99%) soit inférieur à 0.03.

Exercice 15

Supposons que l'on prenne un échantillon aléatoire de 100 comptes clients d'une chaîne de grands magasins et que l'on trouve que le solde moyen débiteur est de 74 dh.

1. Si l'on sait que l'écart type pour tous les comptes est de 86 dh, trouvez l'intervalle de confiance à 95% pour la moyenne μ, des soldes de la population.

2. S'il y a en tout 243000 comptes, déterminez l'intervalle de confiance à 95% pour l'espérance du solde totale. Expliquez brièvement au vice-président la signification de la réponse
3. Supposons que le vice-président, qui demeure sceptique, entreprenne une vérification de tous les soldes débiteurs et qu'il trouve un total de 19 714 000. Que diriez-vous ?

Exercice 16

Une entreprise mène une politique de formation du personnel. Au cours de l'année l'ensemble de ces 200 employés ont suivi une même formation intitulée «FC». Le directeur s'intéresse à l'ancienneté de ces employés ainsi qu'a leur satisfaction de la FC.

1. Le directeur a demandé à un nombre n de ces employés choisis au hasard avec remise (ESAR) leurs ancienneté dans l'entreprise : la moyenne échantillon est $\bar{x} = 68$ mois et l'écart-type échantillon $S = 24$ mois. Déterminer les intervalles de confiance à 90% de l'ancienneté moyenne des employés dans les cas suivants $n = 50$ et $n = 9$.
2. Le directeur a demandé à 50 de ces employés s'ils étaient satisfait de la FC ; 40 se sont déclarés satisfaits.

   a) Supposons que les employés sont choisis au hasard et avec remise. Déterminer l'intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.

   b) Supposons maintenant que les employés sont choisis au hasard et sans remise.
      - i – Dans ces conditions peut on affirmer que L'ESSR donneras approximativement les mêmes résultats que l'ESAR ? pourquoi ?
      - ii – Proposer un intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.

Tests paramétriques

Exercice 17

Un fabricant de GSM s'inquiète de l'autonomie en veille de ses appareils. Il fait prélever un échantillon de 12 GSM, et on relève les autonomies suivantes (en heures) :

$$75 \quad 77 \quad 79 \quad 80 \quad 77 \quad 76 \quad 80 \quad 81 \quad 83 \quad 72 \quad 80 \quad 76$$

1. Déterminer un intervalle de confiance à 95% de μ : l'autonomie moyenne des GSM produits, sachant que l'autonomie de ceux-ci peut être considérée comme distribuée normalement.
2. à partir de l'échantillon de départ, testez l'hypothèse $H_0 : \mu \geq 80$ contre l'hypothèse $H_1 : \mu < 80$, au seuil de signification $\alpha = 5\%$,
3. Calculez la puissance du test sous la contre-hypothèse $H_1 : \mu = 78$.

Exercice 18

Soit une population normale de moyenne μ inconnue et d'écart-type $\sigma = 1.5$ On y a prélevé un échantillon aléatoire simple d'effectif $n = 9$. Les observations sont les suivantes :
24.3 ; 24.7 ; 23.2 ; 23.8 ; 24.5 ; 26.0 ; 25.4 ; 24.8 ; 23.8.

1. Effectuez le test bilatéral $H_0 : \mu = 22.8$ contre $H_1 : \mu \neq 22.8$ au niveau 1%.
2. Effectuez le test unilatéral $H_0 : \mu \leq 22.8$ contre $H_1 : \mu > 22.8$ au niveau 1%.
3. Effectuez le test unilatéral $H_0 : \mu \leq 22.8$ contre $H_1 : \mu > 22.8$ au niveau 1% si on suppose σ inconnu.

Exercice 19

Le bureau d'analyse économique du ministère américain du commerce a indiqué que le revenu annuel moyen d'un habitant de Caroline du Nord était de 18688 dollars (USA Today, 24 août 1995). Un chercheur de l'état de Caroline du Sud veut tester les hypothèses $H_0 : \mu = 18688$ contre $H_1 : \mu \neq 18688$, où μ correspond au revenu annuel moyen d'un habitant de Caroline du Sud. Un échantillon de 400 habitants de Caroline du Sud fournit un revenu annuel moyen de 16860 dollars et un écart-type de 14624 dollars.

1. Quelle est la conclusion appropriée si on fixe le risque de première espèce à 5% ?

2. Trouvez un intervalle de confiance pour μ au niveau de confiance 95%.

Exercice 20

Lors d'un test de qualité sur deux publicités télévisées, chaque publicité a été diffusée six fois au cours d'une semaine. La semaine suivante, une enquête téléphonique a été effectuée pour identifier les personnes qui avaient vu les publicités. On demandait alors aux individus qui avaient vu les publicités de décrire leur slogan principal. On a obtenu les résultats suivants.

Publicité	Nombre de personnes qui ont vu la publicité	Nombre personnes qui qui se souviennent du slogan
A	150	63
B	200	60

1. Effectuez un test au niveau 5% pour vérifier que la publicité A a été mieux perçue que la publicité B
2. Construisez un intervalle de confiance à 95% pour estimer l'écart entre les proportions des deux populations ;

Exercice 21

Considérons une population normale. On a obtenu, pour un échantillon aléatoire simple de taille 20, $S^2 = 10$.

1. Testez l'hypothèse $H_0 : \sigma^2 = 12$ contre $H_1 : \sigma^2 \neq 12$ au niveau $\alpha = 5\%$.
2. Testez l'hypothèse $H_0 : \sigma^2 = 8$ contre $H_1 : \sigma^2 < 8$ au niveau $\alpha = 1\%$.
3. Testez l'hypothèse $H_0 : \sigma^2 = 12$ contre $H_1 : \sigma^2 > 12$ au niveau $\alpha = 0.5\%$.

Exercice 22

Dans le cadre d'une étude de marketing, on analyse la consommation de lessive des ménages. Sur la base d'un échantillon aléatoire simple de 100 ménages, on a observé une consommation moyenne de 14 kilos avec un écart-type S de 2 kilos. D'autre part, 30%, des personnes interrogées accordaient leur préférence à un produit sans phosphates.

1. Construisez un intervalle de confiance à 95% pour l'espérance de la quantité consommée de lessive.

2. Un fabricant affirme que 25% des ménages préfèrent la lessive sans phosphates alors que le consultant responsable de l'étude estime qu'il y en a plus que 25%. Réalisez un test pour départager ces deux opinions (Utiliser un risque de première espèce $\alpha = 5\%$)

Exercice 23

Les individus qui ont rempli leur déclaration de revenus de 1994 avant le 31 mars 1995 ont été remboursé en moyenne 1056 dollars (USA Today, 5 avril 1995). Considérez la population des individus de dernières minutes qui envoient leur déclaration au cours des cinq derniers jours du délai de paiement (entre le 10 et le 15 avril).

1. Un chercheur a suggéré que l'une des raisons pour lesquelles certains individus attendent les cinq derniers jours pour remplir leur déclaration est qu'en moyenne, ces individus ont un remboursement inférieur à celui des personnes qui remplissent leur déclaration relativement tôt. Formulez les hypothèses appropriées de sorte à ce que le rejet de H_0 confirme les dires du chercheur.
2. Le remboursement moyen d'un échantillon de 400 individus qui ont rempli leur déclaration entre le 10 et le 15 avril, était de 910 dollars et l'écart-type de 1600 dollars. Au niveau 5%, quelle est votre conclusion ?

Exercice 24

On dispose de 10 prix de clôture d'une action A et de 16 prix de clôture d'une action B. Pour chacune des deux séries de prix, on a mesuré la moyenne et l'écart type correspondant. On a eu pour le premier échantillon respectivement 1.5, 2.99 ; et pour le second échantillon respectivement : 3.4, 1.96.

1. Construisez un intervalle de confiance de 95% pour chacune des prix moyens des deux actions.
2. Au niveau 5%, ces données indiquent-elles que l'action B est plus chère que l'action A ?
3. Construisez un intervalle de confiance de 95% pour chacune des variances des prix des deux actions.
4. Au niveau 5%, ces données indiquent-elles que la variance du prix de l'action A est supérieur à 9 ?

5. Au niveau 5%, ces données indiquent-elles que le prix de la variance du prix de l'action B est inférieure à 9 ?

Exercice 25

Une maladie contagieuse inconnue touchant les bovins vient de se déclarer au Maroc et vous êtes chargé(e) par le ministère de la santé d'organiser la campagne de dépistage de cette maladie. Cette maladie se caractérise par un taux moyen élevé d'un certain virus dans l'organisme de l'animal.

Le ministère vous a imposé d'effectuer 3 mesures du taux de ce virus par animal, et votre travail consiste à élaborer un test statistique permettant, en fonction de ces 3 mesures, de déterminer si un animal est malade ou non.

Une première étude sur ce virus a été réalisée. Elle indique que le taux observé lors d'une mesure peut être considéré comme une variable aléatoire X suivant une loi normale $\mathcal{N}(\mu, 4)$, où l'espérance μ inconnue est le taux moyen du virus dans l'organisme de l'animal (μ est exprimée en pourcentage). L'étude indique également qu'on peut considérer différentes mesures effectuées sur un même animal comme indépendantes.

1. L'étude indique qu'il n'y a que deux valeurs possibles de μ : soit $\mu = 20$ et l'animal est sain, soit $\mu = 30$ et l'animal est malade. Quelle hypothèse de base choisissez-vous ? Pourquoi ?

   Vous choisissez un seuil $\alpha = 5\%$. Que représente ce seuil ? Que représente la puissance d'un test ?

3. Formulez un test, au seuil α.

4. Déterminez la puissance de ce test.

5. Une seconde étude, plus poussée, sur ce virus vient d'être réalisée. Elle indique que chez un animal sain, le taux moyen de ce virus peut prendre toute valeur strictement inférieure à 30%. Cette étude indique également que lors qu'un animal contracte la maladie, ce taux monte à 30%, puis reste constant jusqu'à la mort de l'animal.
   La prise en compte de cette nouvelle étude vous conduit-elle à modifier votre test ? Pourquoi ?

Tests du Chi-deux

Exercice 26

1. Au niveau $\alpha = 5\%$, le caractère 'taille' mesuré sur 1000 individus peut-il être ajusté par une loi normale ?

X : taille en cm	< 155	$[155 - 165[$	[165-175[	[175-185[	>185
Effectifs	1	40	470	447	42

Exercice 27

Une étude sur le niveau scolaire des électeurs et leur appartenance politique a fourni les informations suivantes.

Niveau scolaire	**Appartenance politique**		
	Gauche	Droite	Centre
Lycée	40	20	10
Baccalauréat	30	35	15
Universitaire	30	45	25

1. Au niveau $\alpha = 10\%$, déterminez si l'appartenance politique est indépendante du niveau scolaire des électeurs.

Exercice 28

Lors d'un test de qualité sur deux publicités télévisées, chaque publicité a été diffusée six fois au cours d'une semaine. La semaine suivante, une enquête téléphonique a été effectuée pour identifier les personnes qui avaient vu les publicités. On demandait alors aux individus qui avaient vu les publicités de décrire leur slogan principal. On a obtenu les résultats suivants.

Publicité	Nombre de personnes qui ont vu la publicité	Nombre personnes qui qui se souviennent du slogan
A	150	63
B	200	60

1. Effectuez un test au niveau 5% pour vérifier si les publicités A et B sont de même qualité.

III. Examens

Exercice 29 ENCGO 2007

Dans une grande ville, on prélevé un échantillon aléatoire de 1000 personnes auxquelles on demande si elles sont favorables à l'interdiction de fumer dans les restaurants. 670 personnes se déclarent favorables à cet interdiction.

1. Donnez un intervalle de confiance du pourcentage de personnes favorables à l'interdiction, pour un risque α fixé à 1% (degré de confiance 99%).
2. Quelle devrait être la taille de l'échantillon si l'on souhaite connaître le pourcentage de gens favorables à l'interdiction, avec une précision égale à 3% (avec toujours le risque fixé à $\alpha = 1\%$) ?
3. À partir de l'échantillon de départ, testez l'hypothèse $H_0 : p \geq 0.7$ contre l'hypothèse $H_1 : p < 0.7$, au seuil de signification $\alpha = 1\%$ (où p désigne la proportion de gens favorables à l'interdiction dans la population de la ville).

Exercice 30 ENCGO 2007

Pour déterminer l'âge moyen de ses clients, une grande entreprise de confection pour hommes prélève un échantillon aléatoire de 50 clients et trouve $\bar{x} = 36$. On suppose que la variance $\sigma^2 = 144$.

1. Trouver un intervalle de confiance à 95% pour l'âge moyen μ, de l'ensemble de ses clients.
2. Supposer que pour le même seuil de confiance (95%) on veuille réduire la longueur de l'intervalle, de façon précise, à 2 années. Quelle doit être alors la taille de l'échantillon ?

3. Un client supplémentaire a été choisi au hasard et l'on trouve qu'il est âgé de 18 ans. Cette valeur est-elle surprenante ? Justifier.

Exercice 31 ENCGO, juillet 2007

On a noté le nombre quotidien d'accidents X pendant une période de 100 jours. Les observations sont les suivantes :

x_j	0	1	2	3	4	5
n_j	41	32	17	7	2	1

Au niveau de probabilité de 0.05, doit-on rejeter les hypothèses suivantes :

1. $P[X = 0] = 0.35$, $P[X = 1] = 0.30$, $P[X = 2] = 0.20$, $P[X = 3] = 0.10$, $P[X > 4] = 0.05$?
2. X suit une loi de Poisson de paramètre $\lambda = 0.9$?

Exercice 32 ENCGO 2008

Un sondage USA Today a révélé que sur un échantillon de 596 répondants, 83 étaient prêts à travailler moins pour un salaire moins important afin d'avoir plus de loisirs personnels. Notons p la vraie proportion de travailleurs prêts à moins travailler et à être moins payés pour avoir plus de loisirs personnels.

1. Donnez un intervalle de confiance à 95% pour la vraie proportion p.
2. Testez l'hypothèse H_0 :$p \leq 12\%$ contre H_1 :$p > 12\%$ au niveau $\alpha = 0.05$.

Exercice 33 FSJESO 2010

1. On considère un échantillon $(X_1, ..., X_n)$ de loi $\mathcal{N}(\mu, 1)$, où μ est inconnu. On estime μ par la moyenne empirique de l'échantillon $\hat{\mu} = \bar{X}$.
   a) Déterminez la loi de $\hat{\mu}$.
   b) Calculez la vraisemblance et la log-vraisemblance du paramètre μ. Montrez que l'estimateur $\hat{\mu}$ est efficace.
2. Soient $\bar{X}_1$ la moyenne d'un échantillon aléatoire de taille n d'une population normale avec moyenne μ, et variance σ_1^2, et $\bar{X}_2$ la moyenne d'un échantillon aléatoire de taille n d'une population normale avec moyenne μ, et variance σ_2^2. On suppose que les deux échantillons sont indépendants.

a) Montrez que $\bar{X}_\omega = \omega \bar{X}_1 + (1-\omega)\bar{X}_2$ est un estimateur sans biais de μ.
b) Montrez que la variance de $\bar{X}_\omega$ est minimale lorsque

$$\omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Exercice 34 FSJESO 2011

La moyenne, au niveau de la population, du prix d'une nouvelle maison est de 750 500 Dhs. Supposez que l'écart type de la population soit 50 000 Dhs et qu'un échantillon de 100 nouvelle maisons soit sélectionné.

1. Déterminez la distribution d'échantillonnage du prix moyen des nouvelles maisons de l'échantillon.
2. Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de 10 000 Dhs de la moyenne de la population.
3. Pour estimer le prix moyen de la population avec un écart maximum de 10 000 Dhs, avec une certitude de 99%, que recommanderiez vous ?.

Exercice 35 FSJESO 2011

Le loyer mensuel d'un appartement avec deux chambres dans une ville particulière est en moyenne μ. Supposez que nous voulions tester $H_0 \ : \ \mu = 2\,500$ contre $H_a \ : \ \mu \neq 2\,500$. Un échantillon de 36 appartements avec deux chambres est sélectionné. La moyenne d'échantillon est égale à 2 650 Dhs et l'écart type de l'échantillon à 154 Dhs.

1. Construisez un intervalle de confiance à 95% pour la moyenne de la population.
2. Effectuez le test d'hypothèse au seuil de signification de 5%.
3. Calculez la puissance de ce test sous l'alternative $H_1 : \mu = 2600$ (la probabilité de refuser H_0 sachant que $\mu = 2\,600$).

Exercice 36 FSJESO 2012

En février 1995, le coût d'un voyage aller-retour par avion suivait une loi normale de moyenne 258 dollars (USA Today). Un échantillon aléatoire de 15 billets aller-retour, vendu au cours du mois de mars, a fourni les données suivantes sur les prix : 310, 260, 265, 255, 300, 310, 230, 250, 265, 280, 290, 240, 285, 250, 260.

1. Quel est l'estimation du prix moyen d'un billet aller-retour en mars ?
2. Construisez un intervalle de confiance à 95% du prix moyen d'un billet au mois de mars.
3. Au niveau $\alpha = 5\%$, testez si le prix moyen d'un billet aller-retour a augmenté en mars.

Exercice 37 FSJESO 2014

Supposons que deux économistes estiment respectivement μ à l'aide des deux estimateurs sans biais et indépendants $\hat{\theta}_1$ et $\hat{\theta}_2$. En plus on sait que l'écart type de $\hat{\theta}_2$ est trois fois plus grand que celui de $\hat{\theta}_1$. Des combinaison des deux estimateurs sont proposés afin d'estimer μ d'une meilleure façon :

(i) $\hat{\alpha}_1 = \frac{1}{2}\,\hat{\theta}_1 + \frac{1}{2}\,\hat{\theta}_2$ (moyenne simple) ;

(ii) $\hat{\alpha}_2 = \frac{3}{4}\,\hat{\theta}_1 + \frac{1}{4}\,\hat{\theta}_2$ (moyenne pondérée) ;

(iii) $\hat{\alpha}_3 = 1\,\hat{\theta}_1 + 0\,\hat{\theta}_2$ (l'estimateur le moins précis est négligé).

1. Quels sont parmi ces estimateurs qui sont sans biais ?
2. Quel est le meilleur estimateur est le plus mauvais ? Justifier votre réponse.

Exercice 38 FSJESO 2014

Dans un échantillon aléatoire de 1500 Américains, la proportion favorable à la décriminalisation de la marijuana (type de drogue) a baissé de 52% en 1980 à 46% en 1985.

1. Construisez un intervalle de confiance à 95% pour la proportion de la population favorable à la marijuana pour chacune des années.
2. Trouvez un intervalle de confiance à 95% pour la variation de cette proportion entre 1980 et 1985.

Exercice 39 FSJESO 2015

Cochez la bonne réponse (bonne réponse +1,5pt, mauvaise réponse -0,5pt)

1. Le sondage stratifié est efficace lorsque les variances internes des strates sont

   ☐ importantes ☐ faibles ☐ nulles ☐ égaux

2. On prélève un échantillon aléatoire de 1000 jeunes diplômés de l'UMP, 330 jeunes déclarent qu'ils sont en chômage. L'intervalle de confiance ; à 95% ; du pourcentage des chômeurs parmi les diplômés de l'UMP est

   ☐$[30\%, 35.9\%]$ ☐$[33\%, 36\%]$ ☐$[27\%, 33\%]$ ☐$[36\%, 39\%]$

3. Un échantillon de 50 appartements situées à Oujda est sélectionné. La moyenne du loyer de ces appartements est égale à 1850 Dhs et leurs écart-type est égal à 230 Dhs.
   a) L'intervalle de confiance ; à 95% ; du loyer moyen est
      ☐$[1586, 1641]$ ☐$[1786, 1914]$ ☐$[1850, 1943]$ ☐$[1760, 1980]$
   b) Supposons que le loyer suit la loi normale, la probabilité de trouver un appartement de loyer inférieur à 1600 Dhs est
      ☐5%] ☐10% ☐14% ☐25%
   c) Pour être confiant à 95%, que le loyer moyen estimé ne s'éloigne du loyer moyen réel que par 50 Dhs il faut donc prélever un échantillon de taille supérieure à
      ☐55 ☐66 ☐72 ☐82

Exercice 40 FSJESO 2016

QCM. Cochez la bonne réponse (bonne réponse +0,5pt, mauvaise réponse-0,5pt)

1. La moyenne d'un grand échantillon converge en probabilité vers la moyenne de la population

   ☐ Vrai ☐ Faux

2. Les résultats d'un recensement sont plus précis que ceux d'un sondage

   ☐ Vrai ☐ Faux

3. Un sondage est plus coûteux qu'un recensement

   ☐ Vrai ☐ Faux

4. Pour un échantillonnage simple avec remise on a $V(\bar{X}) = \sigma^2/n$

   ☐ Vrai ☐ Faux

5. Un sondage stratifié est adéquat pour une population composé de différents groupes homogènes

   ☐ Vrai ☐ Faux

6. Un sondage par grappes est plus adéquat q'un sondage stratifié

   ☐ Vrai ☐ Faux

7. Un estimateur sans biais est celui de moyenne nulle

   ☐ Vrai ☐ Faux

8. Un estimateur qui a la variance minimale est l'estimateur le plus efficace

   ☐ Vrai ☐ Faux

9. Plus la population est homogène (de variance petite), plus les fluctuations de la moyenne de l'échantillon sont importantes

   ☐ Vrai ☐ Faux

10. Plus la taille de l'échantillon est importante, plus les fluctuations de sa moyenne sont importantes

    ☐ Vrai ☐ Faux

Exercice 41 FSJESO 2016

QCM. Cochez la bonne réponse (bonne réponse +1,5pt, mauvaise réponse -0,5pt).

N. B. Dans toute la suite nous fixons l'erreur (niveau) à $\alpha = 5\%$.

1. Soit X une v.a. suivant la loi de poisson de paramètre λ et $X_1, \ldots, X_n$ un ESAR. L'estimateur du maximum de vraisemblance de λ est :

   ☐ $\overline{X}$ ☐ $1/\overline{X}$ ☐ S^2 ☐ $1/S^2$

2. On prélève un échantillon aléatoire simple de 200 clients d'une société de transport, 175 d'entre eux se déclarent satisfait du service. Un responsable déclare que : 'le taux de satisfaction p parmi tous les clients dépasse 83%'.

   (a) L'estimation ponctuelle du taux de satisfaction est

   ☐ 175% ☐ 75% ☐ 87.5% ☐ 89%

   (b) L'intervalle de confiance du pourcentage des clients satisfaits est :

   ☐ $[86.8\%, 88.2\%]$ ☐ $[87\%, 95\%]$ ☐ $[83.7\%, 91.3\%]$ ☐ $[82.3\%, 92.1\%]$

   (c) Pour vérifier son hypothèse, le responsable doit considérer le problème du test suivant :

☐ $\begin{cases} H_0: & p \geq 0.83 \\ H_0: & p < 0.83, \end{cases}$ ☐ $\begin{cases} H_0: & p \leq 0.83 \\ H_0: & p > 0.83, \end{cases}$

☐ $\begin{cases} H_0: & p = 0.83 \\ H_0: & p \neq 0.83, \end{cases}$ ☐ $\begin{cases} H_0: & p \geq 0.83 \\ H_0: & p \leq 0.83, \end{cases}$

(d) La statistique-test calculée est :

☐ $T = 1.69$ ☐ $T = -1.61$ ☐ $T = 1.67$ ☐ $T = 1.63$

(e) La décision prise est :

☐ RH_0 ☐ $\overline{RH_0}$ ☐ RH_1 ☐ Autre

3. On s'intéresse au prix d'une action qu'on suppose normalement distribué. On observe le prix de cette action pendant 20 jours. Le prix moyen observé est 12.5 Dhs et l'écart-type observé est égal à 1.8 Dhs.

(a) L'estimation par intervalle de confiance du prix moyen de l'action est :

☐ $[11.80, 13.19]$ ☐ $[10.66, 12.39]$ ☐ $[9.5, 12.33]$ ☐ $[13.6, 19.64]$

(b) L'incertitude relative de l'estimation ci-dessus est :

☐ 10% ☐ 5% ☐ 2% ☐ 5.56%

(c) L'intervalle de confiance de l'écart-type du prix de l'action est :

☐ $[1.37, 2.63]$ ☐ $[1.4, 2.74]$ ☐ $[-2.5, -1.9]$ ☐ $[1.6, 5.9]$

(d) On va observer le prix de l'action sur une période dépassant 30 jours. Afin de s'assurer que le prix moyen estimé ne s'éloigne du prix moyen réel que par moins de 0.5 Dhs, il faut observer le prix de l'action pendant au moins :

☐ 35 jours ☐ 40 jours ☐ 50 jours ☐ 90 jours

Exercice 42 FSJESO Normale 2017

QCM : Cochez la bonne réponse (Une bonne réponse +0,5pt, une mauvaise réponse -0,5pt)

1. La valeur $\bar{x}$ prise par $\bar{X}$ est une variable aléatoire

☐ Vrai ☐ Faux

2. Une stratification consiste à diviser la population en sous populations homogènes

☐ Vrai ☐ Faux

3. Plus la taille de l'échantillon est grande, plus l'erreur-type de la moyenne de l'échantillon est petite

☐ Vrai ☐ Faux

4. Pour un échantillon simple sans remise on a $V(\bar{X}) = \sqrt{\frac{N-n}{N-1}\frac{\sigma^2}{n}}$

☐ Vrai ☐ Faux

5. La distribution de la moyenne d'un échantillon ESAR normale est normale

☐ Vrai ☐ Faux

6. Un sondage par grappes est utilisé lorsque la population est composée de différents groupes intérieurement homogènes

☐ Vrai ☐ Faux

7. Un estimateur sans biais est convergent si et seulement si sa variance converge vers zéro

☐ Vrai ☐ Faux

8. Plus le MSE est faible plus l'estimateur est moins efficace

☐ Vrai ☐ Faux

9. Un E.S.B. $\hat{\theta}_n$ de θ de variance $Var(\hat{\theta}_n) = 1/I_n(\theta)$, où $I_n(\theta)$ est l'information de Fisher. L'estimateur $\hat{\theta}_n$ est dit efficace

☐ Vrai ☐ Faux

10. Pour tester H_0 : X et Y sont dépendantes contre H_0 : X et Y sont indépendantes, on utilise le test de khi-deux

☐ Vrai ☐ Faux

Exercice 43 FSJESO Normale 2017

QCM : Cochez la bonne réponse (Une bonne réponse +1,5pt, une mauvaise réponse -0,5pt).

1. Une banque se demande si elle n'aurait pas accordé trop de prêts immobiliers à des clients pas assez solvables. Pour se rassurer, elle voudrait avoir une majoration du taux de défaillance des prêts en cours. Elle tire alors au hasard 200 clients dans son portefeuille et fait analyser leur situation. Sur ces 200 cas, seuls 8 présentent un risque sérieux de défaillance.

(a) Le taux moyen de défaillance est

☐ 4% ☐ 6% ☐ 8% ☐ 10%

(b) Un intervalle de confiance à 99% du taux réel de défaillance est

☐ $[1,82\%; 9,12\%]$ ☐ $[0,43\%; 7,57\%]$
☐ $[1,12\%; 8,56\%]$ ☐ $[3,11\%; 12,65\%]$

2. On veut estimer le coût moyen μ_A de séjour d'un patient dans une clinique A. On sait que le coût par jour d'un patient est 1200 DH. Pour un échantillon de 200 patients on a observé une durée de séjour moyenne de 4,6 jours avec un écart-type de 3,8 jours.

(a) L'intervalle de confiance à 95% pour le coût moyen d'un patient est

☐ $[1560; 8520]$ ☐ $[3363; 8345]$ ☐ $[4476; 7362]$ ☐ $[4888; 6152]$

(b) Une autre clinique B, dont le coût par jour d'un patient est toujours 1200 DH, réalise la même étude, le calcul sur un échantillon de 300 patients, a donné une durée de séjour moyenne de 5 jours avec un écart-type de 2,5 jours. On se demande si le coût moyen de séjour μ_B dans la clinique B est plus élevé que le coût moyen de séjour μ_A dans la clinique A.

(c) Le problème du test qu'on doit poser est :

☐ $H_0 : \mu_B = \mu_A; \quad H_1 : \mu_B < \mu_A$ ☐ $H_0 : \mu_B < \mu_A; \quad H_1 : \mu_B \geq \mu_A$
☐ $H_0 : \mu_B \leq \mu_A; \quad H_1 : \mu_B > \mu_A$ ☐ $H_0 : \mu_B > \mu_A; \quad H_1 : \mu_B \leq \mu_A$

(d) Pour un niveau de signification $\alpha = 0,05$, la règle de décision sera :

☐ Rejeter H_0 si $\bar{X}_B - \bar{X}_A < 0$ ☐ Rejeter H_0 si $\bar{X}_B - \bar{X}_A \geq 1200$
☐ Rejeter H_0 si $\bar{X}_B - \bar{X}_A > 600$ Rejeter H_0 si $\bar{X}_B - \bar{X}_A \leq 600$

(e) La conclusion du test, au seuil 5%, est :

☐ Il n ya pas de différence entre les coûts par patients dans les deux cliniques
☐ Le coût par patient dans la clinique B est nettement supérieur que celui dans la clinique A
☐ On ne peut confirmer que le coût par patient dans la clinique B est supérieur à celui de la clinique A
☐ On ne peut rien conclure

Exercice 44 FSJESO Normale 2017

QCM : Cochez la bonne réponse (Une bonne réponse +1,5pt, une mauvaise réponse -0,5pt) 1.

Afin d'estimer $\mu = E(X)$, où X est la consommation mensuel finale des ménages par tête, le HCP en 2014 a sélectionné un échantillon (ESAR) de taille ménages $n = 16000$ ménages. La moyenne d'échantillon est notée $\bar{X}$; sa valeur calculée est $\bar{x} = 1325$ DHS ; l'écart-type ajusté de l'échantillon est $S = 530$ DHS.

1. La distribution d'échantillonnage de $\bar{X}$ est

   ☐ $\frac{\bar{X}-\mu}{0{,}033} \approx \mathcal{N}(0,1)$ ☐ $\frac{\bar{X}-\mu}{4{,}19} \approx t_{15999}$
   ☐ $\frac{\bar{X}-\mu}{4{,}19} \approx \mathcal{N}(0,1)$ ☐ $\bar{X} \approx \mathcal{N}(0,1)$

2. La probabilité que la moyenne estimée s'écarte de la valeur réelle μ de plus de 10 DH est :

   ☐ 5% ☐ $1,74\%$ ☐ $2,5\%$ ☐ $0,87\%$

3. Afin de s'assurer à 99% que la moyenne estimée ne s'écarte de la valeur réelle que de 10 DH au plus, il faut sélectionner un échantillon dont la taille est égale au moins à :

   ☐ 12000 ☐ 18640 ☐ 16000 ☐ 19460

4. On sélectionne du premier échantillon un sous échantillon ESSR de taille $n' = 3000$. Notons par $\bar{X}'$ la moyenne du deuxième échantillon. L'écart-type de $\bar{X}'$ est :

   ☐ $\sigma_{\bar{X}'} = 8,72$ ☐ $\sigma_{\bar{X}'} = 4,19$ ☐ $\sigma_{\bar{X}'} = 9,67$ ☐ $\sigma_{\bar{X}'} = 530$

Exercice 45 FSJESO Rattrapage 2017

QCM. Cochez la bonne réponse (bonne réponse +3 pt, mauvaise réponse -1 pt)

À la veille d'un référendum sur l'adoption d'une nouvelle constitution, on a prélevé un échantillon aléatoire de 1000 votants auxquels on a demandé s'ils vont voter par OUI ou par NON. 520 personnes se déclarent favorables au OUI.

1. Pour une confiance fixé à 95%, l'intervalle de confiance du pourcentage p des personnes favorables au OUI, est :

   ☐ $[45\%, 59\%]$ ☐ $[47.2\%, 56.8\%]$
   ☐ $[48.9\%, 55.1\%]$ ☐ $[53.9\%, 55.1\%]$

2. Si l'on souhaite connaître le pourcentage de gens favorables au OUI, avec une précision inférieure ou égale à 2%, la taille de l'échantillon doit être supérieure à :

☐ 900 ☐ 1204 ☐ 2401 ☐ 3631

3. Au niveau $\alpha = 5\%$, le test de l'hypothèse $H_0 : p \leqslant 0.5$ contre l'hypothèse
$H_1 : p > 0.5$ donne :

☐ Victoire du OUI ☐ Rien ne permet d'écarter la victoire du NON
☐ Victoire du NON ☐ Rien ne permet d'écarter la victoire du OUI

Exercice 46 FSJESO Rattrapage 2017

QCM. Cochez la bonne réponse (bonne réponse +2 pt, mauvaise réponse -1 pt)

Une entreprise souhaite lancer un nouveau produit et elle veut tester la dépendance entre l'intention d'achat du produit et l'âge du client. Ainsi, elle demande à 400 clients potentiels, s'ils sont prêts à acheter ce nouveau produit. Les résultats en fonction de l'âge des personnes interrogées sont les suivants :

	Intention d'achat du produit		
Âge	Oui	Non	Abstention
<30 ans	65	27	8
de 30 à 45 ans	50	19	11
de 45 à 60 ans	35	24	11
$\geq$ 60 ans	50	80	20

1. La statistique utilisée pour tester l'indépendance suit une loi de :

☐ Chi-deux à 12 degrés de liberté ☐ Chi-deux à 6 degrés de liberté
☐ Student à 3 degrés de liberté

2. Au niveau $\alpha = 5\%$, peut t-on affirmer que l'intention d'achat est **dépendante** de l'âge du client ?

☐ Oui ☐ Non ☐ On ne peut conclure.

IV. Solutions

Solution de l'Exercice 1

L'erreur d'échantillonnage du premier échantillon est

$$S_1 = \frac{\sigma_1}{\sqrt{n_1}},$$

tandis que l'erreur d'échantillonnage du deuxième échantillon est

$$S_2 = \frac{\sigma_2}{\sqrt{n_2}} = \frac{2\sigma_1}{\sqrt{2n_1}} = \sqrt{2}S_1.$$

Ainsi,

$$S_2 = \sqrt{2}S_1 > S_1.$$

Conclusion. *L'estimation de la moyenne de la première population est plus précise que celui de la deuxième population.*

Solution de l'Exercice 2

1. L'écart type de $\bar{X}$, lorsque la taille de l'échantillon vaut 50 est

$$S = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{50}} = \mathbf{4.24}.$$

L'écart type de $\bar{X}$, lorsque la taille de l'échantillon vaut 100 est

$$S = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{100}} = \mathbf{3}.$$

2. $\bar{\mathbf{X}} \sim \mathcal{N}(\mathbf{1600}, \sigma/\sqrt{\mathbf{n}})$.
3. Afin d'assurer, à 95%, que $\bar{X}$ s'écarte de 1600 g de moins de 20 g, il faut que

$$\begin{aligned} & P(|\bar{X} - 1600| < 20) \geq 0.95 \\ \Leftrightarrow \; & P(|\bar{X} - 1600| > 20) < 0.05 \\ \Leftrightarrow \; & P(|\bar{X} - 1600| > k\,S) < 1/k^2 \qquad (IBT) \end{aligned}$$

avec $kS = 20$ et $1/k^2 = 0.05$. Ainsi,

$$\begin{aligned} & k = \sqrt{1/0.05} \approx 4.47 \\ \Rightarrow \; & S = \sigma/\sqrt{n} = 20/k = 20/4.47 \\ \Rightarrow \; & \sqrt{n} = \frac{30 \times 4.47}{20} = 6.70 \\ \Rightarrow \; & n = 6.70^2 = \mathbf{44.89} \end{aligned}$$

Conclusion. *On est, à 95%, confiant qu'avec un échantillon de taille supérieure ou égale à 45, l'écart entre $\bar{X}$ et sa valeur espérée est au plus 20 g.*

4. On a $\mathrm{E}[\bar{X}_I - \bar{X}_{II}] = 1600 - 1600 = 0$ et comme $\bar{X}_I$ et $\bar{X}_{II}$ sont indépendants alors

$$\mathrm{Var}(\bar{X}_I - \bar{X}_{II}) = \mathrm{Var}(\bar{X}_I) + \mathrm{Var}(\bar{X}_{II}) \approx 4.24^2 + 3^2 \approx 26.98.$$

En plus, $\bar{X}_I$ et $\bar{X}_{II}$ sont normales et par suite

$$\bar{X}_I - \bar{X}_{II} \sim \mathcal{N}(\mathbf{0}, \mathbf{26.98}).$$

Ainsi,

$$\begin{aligned} P(|\bar{X}_I - \bar{X}_{II}| > 10) &= P\left(\frac{|\bar{X}_I - \bar{X}_{II}|}{\sqrt{26.98}} > \frac{10}{\sqrt{26.98}}\right) \\ &= P(|Z| > 1.92) \\ &= 2(1 - F_Z(1.92)) \\ &= \mathbf{0.0548}. \end{aligned}$$

Conclusion. *La probabilité que les moyennes des deux échantillons diffèrent de plus de **10** g est **5.48**%.*

Solution de l'Exercice 3

1. La valeur de l'écart type de la moyenne est
$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 60/\sqrt{50} \approx \mathbf{8.485}.$$

2. Comme $n = 50 \geq 30$, il s'agit du cas des grands échantillons donc
$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - 115}{8.485} \approx \mathcal{N}(0, 1).$$
Ainsi, la probabilité que la moyenne d'échantillon soit supérieur à 115 jours par an est
$$P(\bar{X} \geq 115) = P(\frac{\bar{X} - 115}{8.485} \geq \frac{115 - 115}{8.485}) = 1 - P(Z \leq 0) = \mathbf{50}\%.$$

3. La probabilité que la moyenne d'échantillon s'écarte au plus de ± 5 jours de la moyenne de la population est
$$\begin{aligned} P(|\bar{X} - 115| \geq 5) &= P(|Z| \leq 5/8.485) \\ &= P(|Z| \leq 0.59) \\ &= 2F_Z(0.59) - 1 = \mathbf{44.48}\%. \end{aligned}$$

4. Si $n = 100$ alors $\sigma_{\bar{X}} = 60/\sqrt{100} \approx \mathbf{6}$ ainsi La probabilité que la moyenne d'échantillon s'écarte au plus de ± 5 jours de la moyenne de la population est
$$\begin{aligned} P(|\bar{X} - 115| \leq 5) &= P(|Z| > \frac{5}{6}) = 1 - P(|Z| \leq 0.83) \\ &= 2 - 2F_Z(0.83) = \mathbf{40.66}\%. \end{aligned}$$

Solution de l'Exercice 4

1. L'erreur type pour la firme A, étant donné que les échantillons sélectionnés sont de type ESSR et de taille 50 est :
$$\begin{aligned} \sigma_{\bar{X}}^{(A)} &= \sqrt{\frac{N - n}{N - 1} . \frac{\sigma^2}{n}} \\ &= \sqrt{\frac{2000 - 50}{2000 - 1} . \frac{144^2}{50}} \\ &\simeq \mathbf{20.11}. \end{aligned}$$

Par la même façon on trouve,

$$\begin{aligned} \sigma_{\bar{X}}^{(B)} &\simeq \mathbf{20.26}; \\ \sigma_{\bar{X}}^{(C)} &\simeq \mathbf{20.31}. \end{aligned}$$

2. Pour la firme A, la probabilité que la moyenne d'échantillon $\bar{x}$ s'écarte au plus de ± 25 de la moyenne de la population μ est :

$$\begin{aligned} P(|\bar{X} - \mu| \leq 25) &= P(|\bar{X} - \mu|/20.11 \leq 25/20.11) \\ &= P(|Z| \leq 1.24) \\ &= 2F_Z(1.24) - 1 \simeq \mathbf{78.50\%} \end{aligned}$$

Par la même façon on trouve, pour les firme B et C, respectivement :

$$\begin{aligned} 2F_Z(1.2339) - 1 &\simeq \mathbf{78.27\%} \\ 2F_Z(1.2309) - 1 &\simeq \mathbf{78.16\%} \end{aligned}$$

Conclusion. *En conclusion, même pour des tailles de population différentes, l'ESSR donne pour une taille d'échantillon fixé approximativement les mêmes performances.*

Solution de l'Exercice 5

Dans un processus continu de fabrication, on peut supposer toujours que la taille population est assez grande par rapport à la taille échantillon, ainsi on suppose qu'on a un ESAR avec $n = 100 > 30$. Par conséquent, La distribution d'échantillonnage de $\bar{p}$ est :

$$\frac{\bar{p} - 0.1}{\sqrt{0.1(1-0.1)/100}} = \frac{\bar{p} - 0.1}{0.03} \approx \mathcal{N}(0, 1).$$

Ainsi,

$$\begin{aligned}
P(\bar{p} > 0.05) &\approx P(\frac{\bar{p} - 0.1}{0.03} > \frac{\bar{0}.05 - 0.1}{0.03}) \\
&\approx P(Z > -1.67) \\
&= 1 - P(Z < -1.67) \\
&= 1 - F_Z(-1.67) \\
&= F_Z(1.67) \\
&= \mathbf{0.9525}.
\end{aligned}$$

Conclusion. *Une série de fabrication qui contient* 10% *de pièces défectueuses à* $1 - 0.9525 = 4.75\%$ *de chance d'échapper à ce test de contrôle.*

Solution de l'Exercice 6

On s'interesse à la non réussite des entreprises au cours de l'année 2006. Notons p la probabilité de faillite d'une entreprise en cette année 2006. Pour étudier ce paramètre p, on a prélevé un échantillon aléatoire simple avec remise de n entreprises en activité au début de l'année 2006, et on a relevé pour chaque entreprise si oui $(X = 1)$ ou non $(X = 0)$ une faillite avait eu lieu au cours de l'année.
$X \sim B(1,p)$ où p est le paramètre d'intérêt. Donc $P[X = 1] = p$ et $P[X = 0] = 1 - p = q$. Et on a

$$P[X = x_i] = p^{x_i}(1-p)^{1-x_i}; \quad x_i \in \{0,1\}$$

1. Déterminez l'estimateur de p obtenu
   (a) par la méthode des moments

$$E(X) = p \text{ et } \Rightarrow \hat{p} = \bar{X}.$$

   (b) par la méthode du maximum de vraisemblance

$$\begin{aligned}
L(x_1,...,x_n;p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}
\end{aligned}$$

$$L^*(p) = \ln L(x_1, ..., x_n; p) = \sum_{i=1}^{n} x_i \ln p + (n - \sum_{i=1}^{n} x_i) \ln(1-p)$$

$$\frac{\partial L^*}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{l-p} = \frac{\sum_{i=1}^n x_i - np}{p(1-p)}$$

$$\frac{\partial L*}{\partial p} = 0 \Leftrightarrow \sum_{i=1}^{n} x_i = np \Leftrightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

$$\frac{\partial^2 L*}{\partial p^2} = \frac{-\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2} < 0$$

(évident car x_i ne prend que les valeurs 0 et 1).

2. Recherchez les propriétés de ces estimateurs (biais, convergence,efficacité).
(i)

$$E[\hat{p}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} np = p.$$

Donc, $\hat{p}$ est un estimateur sans biais.
(ii)

$$Var[\hat{p}] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \overset{n \mapsto \infty}{\rightarrow} 0.$$

Donc, p est un estimateur convergent.
(iii)

$$\begin{aligned} I(p) &= E[-\frac{\partial^2 L*}{\partial p^2}] \\ &= \frac{\sum_{i=1}^n E[X_i]}{p^2} + \frac{n - \sum_{i=1}^n E[X_i]}{(1-p)^2} \\ &= \frac{np}{p^2} + \frac{n-np}{(1-p)^2} \\ &= \frac{n}{p} + \frac{n}{(1-p)} \\ &= \frac{n}{p(1-p)} \end{aligned}$$

Et comme $I(p) = (Var[\hat{p}])^{-1}$, p est efficace.

Solution de l'Exercice 7

L'institut national de statistique s'est intéressé au nombre moyen μ d'accidents à un certain endroit, pendant les heures de pointes. Pour étudier ce paramètre μ une étude a été menée pendant n jours sur une route entre 8h et 9h ; le nombre d'accident X a été relevé chaque jour entre ces heures, ce qui donne lieu aux observations $X_1, \ldots, X_n$.

La variable est le nombre d'événements qui se produit dans un laps de temps elle suit donc la loi de poisson c. à. d. $X \sim \mathcal{P}(\mu)$ et donc

$$P[X = x_i] = \frac{e^{-\mu}\mu^{x_i}}{x_i!} \quad i = 0, 1, 2,$$

1. . Déterminons l'estimateur de μ obtenu

   (a) par la méthode des moments

   $$\mu_1' = E(X) = \mu \Rightarrow \hat{\mu} = \bar{X}.$$

   (b) par la méthode du maximum de vraisemblance

   $$\begin{aligned} L(x_1, ..., x_n; \mu) &= \prod_{i=1}^{n} \frac{e^{-\mu}\mu^{x_i}}{x_i!} \\ &= \frac{e^{-n\mu}\mu^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \end{aligned}$$

   $$L^*(\mu) = -n\mu + \sum_{i=1}^{n} x_i \ln \mu - \sum_{i=1}^{n} \ln x_i!$$

   $$\frac{\partial L^*}{\partial \mu} = -n + \frac{\sum_{i=1}^n x_i}{\mu}$$

   $$\frac{\partial L^*}{\partial \mu} = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

   Cette solution (point critique) est bien un maximum car

   $$\frac{\partial^2 L^*}{\partial \mu^2} = \frac{-\sum_{i=1}^n x_i}{\mu^2} < 0$$

2. . Recherchez les propriétés de ces estimateurs

   i)

   $$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} n\mu = \mu.$$

Donc, $\hat{\mu}$ est un estimateur sans biais.
ii)

$$Var[\hat{\mu}] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2} n\mu = \frac{\mu}{n} \overset{n\mapsto\infty}{\to} 0.$$

Donc, $\hat{\mu}$ est un estimateur convergent.
iii)

$$\begin{aligned} I(\mu) &= E[-\frac{\partial^2 L^*}{\partial \mu^2}] \\ &= \frac{\sum_{i=1}^{n} E[X_i]}{\mu^2} \\ &= \frac{n\mu}{\mu^2} \\ &= \frac{n}{\mu} \end{aligned}$$

Et comme $I(\mu) = (Var[\hat{\mu}])^{-1}$, μ est efficace.

Solution de l'Exercice 8

Soit un échantillon aléatoire simple de taille n, $(X_1, ..., X_n)$, distribué suivant une loi géométrique (p).
Rappel : Si $X \sim \mathcal{G}(p)$ alors $P[X = x_i] = p(1-p)^{x_i} \quad x_i = 1, 2, \ldots$
Déterminons, à partir de cet échantillon, les estimateurs de p.
(a)Par la méthode des moments

$$E[X] = q/p \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = (1-\hat{p})/\hat{p} \Leftrightarrow \hat{p} = \frac{1}{\bar{X}+1}.$$

(b) Par la méthode du maximum de vraisemblance

$$\begin{aligned} L(x_1, ..., x_n; p) &= \prod_{i=1}^{n} p(1-p)^{x_i} \\ &= p^n (1-p)^{\sum_{i=1}^{n} x_i} \end{aligned}$$

$$L^*(p) = n\ln p + (\sum_{i=1}^{n} x_i)\ln(1-p)$$

$$\frac{\partial L^*}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^n x_i}{1-p} = \frac{n - np - p\sum_{i=1}^n x_i}{p(1-p)}$$
$$\frac{\partial L*}{\partial p} = 0 \Leftrightarrow \hat{p} = \frac{n}{n + \sum_{i=1}^n X_i} = \frac{1}{\bar{X}+1}.$$

Cette solution (point critique) est bien un maximum car

$$\frac{\partial^2 L*}{\partial p^2} = -\frac{n}{p^2} - \frac{\sum_{i=1}^n x_i}{(1-p)^2} < 0.$$

Solution de l'Exercice 9

En 1897, l'économiste suisse Vilfredo Pareto (1848-1923), professeur d'économie politique à l'université de Lausanne, eut l'idée de modéliser la loi des revenus en postulant que le nombre relatif de personnes dont le revenu dépasse une valeur x est inversement proportionnel à une puissance de x. La définition suivante fut adoptée : une variable aléatoire X, absolument continue, suit une loi de Pareto de paramètres α et θ si sa densité est

$$f(x;\alpha,\theta) = kx^{-\alpha}\mathbb{I}_{[\theta,+\infty[}(x), \qquad (\theta > 0, \alpha > 1).$$

On suppose ici que α est connu. Estimons θ par la méthode du maximum de vraisemblance.

$$\begin{aligned}\int f(x;\alpha,\theta)dx = 1 \quad &\Leftrightarrow \quad k = \frac{1}{\int_\theta^{+\infty} x^{-\alpha}dx}\\ &\Leftrightarrow \quad k = \frac{-\alpha+1}{[x^{-\alpha+1}]_\theta^{+\infty}}\\ &\Leftrightarrow \quad k = (\alpha-1)\theta^{\alpha-1}.\end{aligned}$$

Ainsi,

$$f(x;\alpha,\theta) = (\alpha-1)\theta^{\alpha-1}x^{-\alpha}\mathbb{I}_{[\theta,+\infty[}(x).$$

$$\begin{aligned}L(x_1,...,x_n;\theta) &= \prod_{i=1}^n (\alpha-1)\theta^{\alpha-1}x_i^{-\alpha}\mathbb{I}_{[\theta,+\infty[}(x_i)\\ &= (\alpha-1)^n\theta^{n(\alpha-1)}\left(\prod_{i=1}^n x_i\right)^{-\alpha}\prod_{i=1}^n \mathbb{I}_{[\theta,+\infty[}(x_i)\end{aligned}$$

$$\mathbb{I}_{[\theta,+\infty[}(x_i) = \begin{cases} 1, & \text{si } x_i \geq \theta; \\ 0, & \text{sinon.} \end{cases}$$

$$\begin{aligned} \prod_{i=1}^{n} \mathbb{I}_{[\theta,+\infty[}(x_i) &= \begin{cases} 1, & \text{si tous les } x_i \geq \theta; \\ 0, & \text{sinon.} \end{cases} \\ &= \begin{cases} 1, & \text{si } \min\limits_{1\leq i\leq n} x_i \geq \theta; \\ 0, & \text{sinon.} \end{cases} \\ &= \mathbb{I}_{[\theta,+\infty[}(\min_{1\leq i\leq n} x_i) \end{aligned}$$

Donc

$$L(1, ..., x_n; \theta) = (\alpha - 1)^n \left(\prod_{i=1}^{n} x_i\right)^{-\alpha} \theta^{n(\alpha-1)} \mathbb{I}_{[\theta,+\infty[}(\min_{1\leq i\leq n} x_i).$$

$(\alpha-1)^n \left(\prod_{i=1}^{n} x_i\right)^{-\alpha}$ est une constante positive indépendante de θ. $\theta^{n(\alpha-1)}$ est une fonction croissante de θ. Ainsi, pour maximiser la vraisemblance, il faut choisir θ le plus grand possible mais $\mathbb{I}_{[\theta,+\infty[}(\min\limits_{1\leq i\leq n} x_i)$ est nul si $\min_{1\leq i\leq n} x_i < \theta$. Donc la plus grande valeur de θ telle que la fonction de vraisemblance ne soit pas nulle est $\min\limits_{1\leq i\leq n} x_i$, ainsi

$$\hat{\theta} = \min_{1\leq i\leq n} X_i.$$

Solution de l'Exercice 10

1. Déterminons l'estimateur de θ obtenus par la méthode des moments

$$\begin{aligned} E(X) &= \frac{1}{\theta} \int_0^{+\infty} x e^{-x/\theta} dx \\ &= [-x e^{-x/\theta}]_0^{+\infty} + \int_0^{+\infty} e^{-x/\theta} dx \text{ (int. par parties)} \\ &= 0 + [-\theta\, e^{-x/\theta}]_0^{+\infty} \\ &= -\theta\, (e^{-\infty/\theta} - e^{-0/\theta}) = -\theta\, (0 - 1) \\ &= \theta \end{aligned}$$

ainsi

$$E(X) = \theta \Rightarrow \hat{\theta} = \bar{\mathbf{X}}.$$

2. Pour un échantillon où tous les X_i sont strictement positifs la vraisemblance du paramètre θ est

$$\begin{aligned} L(X_1, ..., X_n; \theta) &= \prod_{i=1}^{n} f(X_i) \\ &= \frac{1}{\theta^n} * \exp\left(-\sum_{i=1}^{n} X_i/\theta\right). \end{aligned}$$

La log-vraisemblance du paramètre θ est

$$L^*(\theta) = \ln L(X_1, ..., X_n; \theta) = -n \ln \theta - \sum_{i=1}^{n} (X_i/\theta).$$

La fonction L^* ; et par suite la fonction L ; admet un maximum au point $\hat{\theta}$ si et seulement

$$\begin{cases} \dfrac{\partial L^*}{\partial \theta}(\hat{\theta}) = 0, \\ \dfrac{\partial^2 L^*}{\partial \theta^2}(\hat{\theta}) < 0 \end{cases}$$

$$\frac{\partial L^*}{\partial \theta}(\hat{\theta}) = 0 \Leftrightarrow -n/\hat{\theta} + \sum_{i=1}^{n} (X_i/\hat{\theta}^2) = 0 \Leftrightarrow \hat{\theta} = \bar{X}.$$

$$\frac{\partial^2 L*}{\partial \theta^2}(\hat{\theta}) = n/\hat{\theta}^2 - 2\, n\, \bar{X}/\hat{\theta}^3 = -n/\hat{\theta}^2 < 0.$$

Par conséquent l'estimateur de maximum de vraisemblance de θ est égale à l'estimateur obtenu par la méthode des moments :

$$\hat{\theta}^{(\mathbf{n})} = \bar{\mathbf{X}}.$$

3. L'estimateur $\hat{\theta}$ est **sans biais** puisque $E(\hat{\theta}) = \theta$ en effet :

$$E(\hat{\theta}) = E(\bar{X}) = E(X) = \theta.$$

L'estimateur $\hat{\theta}^{(n)}$ est **convergent** puisque :

$$V(\hat{\theta}) = V(\bar{X}) = V(\bar{X}) = \frac{V(X)}{n} = \frac{\theta^2}{n} \overset{n \mapsto +\infty}{\longrightarrow} 0.$$

N. B.

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{1}{\theta}\int_0^{+\infty} x^2 e^{-x/\theta} dx - \theta^2 \\ &= 2\theta^2 - \theta^2 = \theta^2. \end{aligned}$$

Solution de l'Exercice 11

Déterminons un intervalle de confiance pour μ au niveau de confiance 95%. On sait bien que $X \sim N(\mu, \sigma^2)$ donc

$$\bar{X} \sim N(\mu, \sigma^2/n),$$

par suite,

$$\begin{aligned} P[-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}] &= 1-\alpha \\ P[-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq 1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[-\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \end{aligned}$$

Donc,

$$\begin{aligned} I.C_{0.95}(\mu) &= [\bar{x}-1.96\frac{\sigma}{\sqrt{n}}; \bar{x}+1.96\frac{\sigma}{\sqrt{n}}] \\ &= [24.5 - 1.96 * 1.5/3; 24.5 + 1.96 * 1.5/3] \\ &= \mathbf{[23.52\,;\,25.48]} \end{aligned}$$

Solution de l'Exercice 12

Soient deux populations normales, la première de moyenne μ_1 inconnue et d'écart-type $\sigma_1 = 1.5$ et la seconde de moyenne μ_2 inconnue et d'écart-type $\sigma_2 = 3$. Un premier échantillon aléatoire simple d'effectif $n_1 = 9$ est extrait de la population 1 et on observe $\bar{x}_1 = 24.5$. Un second échantillon aléatoire simple d'effectif $n_2 = 12$ est extrait de la population 2 et on observe

$\bar{x}_2 = 23.5$. On suppose ces deux échantillons sont indépendants.
Trouvez un intervalle de confiance pour la différence $\mu_1 - \mu_2$ au niveau de confiance 95%.
Cas deux populations normales avec σ_1 et σ_2 connus.
On a

$$\bar{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1}) \text{ et } \bar{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

donc,

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

Ainsi,

$$P\left[-z_{1-\alpha/2} \leq \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{1-\alpha/2}\right] = 1 - \alpha$$

$$P\left[\bar{X}_1 - \bar{X}_2 - 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 0.95$$

Donc,

$$\begin{aligned} I.C_{0.95}(\mu_1 - \mu_2) &= \left[(\bar{x}_1 - \bar{x}_2) - 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{x}_1 - \bar{x}_2) + 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \\ &= \left[(24.5 - 23.5) - 1.96 * \sqrt{\frac{1.5^2}{9} + \frac{3^2}{12}}; (24.5 - 23.5) + 1.96 * \sqrt{\frac{1.5^2}{9} + \frac{3^2}{12}}\right] \\ &= \mathbf{[-0.96\ ;\ 2.96]} \end{aligned}$$

Solution de l'Exercice 13

La société Butler County Bank désire estimer l'écart entre le solde moyen des cartes de crédit dans deux de ses succursales. Les statistiques descriptives des échantillons aléatoires indépendants des clients ayant une carte de crédit sont présentés dans le tableau ci-dessous.

Succursale 1	Succursale 2
$n_1 = 32$	$n_2 = 36$
$\bar{x}_1 = 500$	$\bar{x}_2 = 375$
$S_1 = 150$	$S_2 = 130$

1. Construisons un estimateur de l'écart entre les deux soldes moyens ;
   L'estimateur de $\mu_1 - \mu_2$ est :

$$\widehat{\mu_1 - \mu_2} = \hat{\mu}_1 - \hat{\mu}_2 = \boldsymbol{\bar{X}_1 - \bar{X}_2},$$

   une estimation de la différence des moyens est donc $\bar{x}_1 - \bar{x}_2 = \mathbf{125}$.

2. Construisons un intervalle de confiance au niveau de confiance 99% pour l'écart entre les deux soldes moyens.
   Cas n_1 et n_2 supérieurs à 30 avec σ_1 et σ_2 inconnus. Ainsi,

$$\bar{X}_1 - \bar{X}_2 \approx N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

   Par conséquent,

$$\begin{aligned} P\left[-z_{1-\alpha/2} \leq \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \leq z_{1-\alpha/2}\right] &= 1 - \alpha \\ \Leftrightarrow P\left[\bar{X}_1 - \bar{X}_2 - 1.96\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \right. &\leq \\ \left. \bar{X}_1 - \bar{X}_2 + 2.58\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] &= 0.99 \end{aligned}$$

   Donc,

$$\begin{aligned} I.C_{0.99}(\mu_1 - \mu_2) &= \left[(\bar{x}_1 - \bar{x}_2) \pm 2.58\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] \\ &= \left[125 \pm 2.58 * \sqrt{\frac{150^2}{32} + \frac{130^2}{36}}\right] \\ &= \mathbf{[36.66\,,\ 213.34]} \end{aligned}$$

3. Déterminons l'incertitude absolue d'estimation

$$EA = \frac{213.34 - 36.66}{2} = \mathbf{88.34},$$

ainsi l'incertitude relative de l'estimation est

$$ER = \frac{EA}{\hat{\mu}} = \frac{88.34}{125} = \mathbf{70.7}\%.$$

Remarquons que cette l'incertitude relative d'estimation est assez grande ce qui indique que l'estimation est assez imprécise !

Solution de l'Exercice 14

1. L'estimation ponctuelle de la proportion p de comptes fournisseurs datant de plus de deux mois est $\hat{p}$ la proportion de comptes fournisseurs datant de plus de deux mois dans l'échantillon :

$$\bar{p} = 88/120 = 0.73.$$

2. Puisque $n = 120 \geq > 30$, on est dans le cas d'un grand échantillon :

$$\begin{aligned} & \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \approx N(0,1) \\ \Rightarrow \quad & P[-z_{1-\alpha/2} \leq \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \leq z_{1-\alpha/2}] = 1 - \alpha \\ \Rightarrow \quad & P[\hat{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58 \leq p \leq \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58] = 0.99 \end{aligned}$$

Donc

$$\begin{aligned} I_{0.99}(p) &= [\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58; \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58] \\ &= [0.73 - 2.58 * 0.04; 0.73 + 2.58 * 0.04] = [\mathbf{0.63}; \mathbf{0.84}]. \end{aligned}$$

Soit I_a l'incertitude absolue (le rayon de l'intervalle de confiance).

$$\begin{aligned} I_a &\leq 0.03 \\ \sqrt{\frac{p(1-p)}{n}} z_{1-\alpha/2} &\leq 0.03 \\ n &\geq \frac{z^2_{1-\alpha/2}}{4 * (0.03)^2} \quad \text{car } p(1-p) \leq 1/4, \forall p \\ n &\geq \mathbf{1}849. \end{aligned}$$

Solution de l'Exercice 15

Supposons que l'on prenne un échantillon aléatoire de 100 comptes clients d'une chaîne de grands magasins et que l'on trouve que le solde moyen débiteur est de 74 dh.

1. Si l'on sait que l'écart type pour tous les comptes est de 86 dh, trouvons l'intervalle de confiance à 95% pour la moyenne μ, des soldes de la population.
   Cas $n > 30$ et $\sigma = 86$ connu. On sait bien que :
$$\bar{x} \sim N(\mu, \sigma^2/n)$$
   Donc,
$$\begin{aligned} P[-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}] &= 1-\alpha \\ P[-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq 1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[-\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \end{aligned}$$
   Donc,
$$\begin{aligned} I.C_{0.95}(\mu) &= [\bar{x}-1.96\frac{\sigma}{\sqrt{n}}; \bar{x}+1.96\frac{\sigma}{\sqrt{n}}] \\ &= [74-1.96*84/10; 74+1.96*84/10] \\ &= \mathbf{[57.54\ ;\ 90.45]} \end{aligned}$$

2. S'il y a en tout 243 000 comptes, déterminons l'intervalle de confiance à 95% pour l'espérance du solde totale.
   La taille de la population est $N = 243\ 000$, la somme des soldes débiteurs est $\sum_{i=1}^{N} x_i = N * \mu$.
$$P(57.54 \leq \mu \leq 90.45) = 0.95 \Rightarrow P(N*57.54 \leq N*\mu \leq N*90.45) = 0.95,$$
   par conséquent $I.C_{0.95}$ pour le solde débiteur totale est [13 982 220 ; 21 979 350].
   Rapport du manager : D'après l'étude de 100 comptes, on peut être confiant à 95%, qu'on doit à nos clients un montant pas moins de 13 982 220 dh mais pas plus de 21 979 350 dh.

3. Le vice-président, qui demeure sceptique, entreprenne une vérification de tous les soldes débiteurs et trouve un total de 19 714 000.
ça confirme le rapport Mr le vice président !

Solution de l'Exercice 16

1. Soit X la v.a égale à l'ancienneté d'un employé.
   — Le cas $n = 50$ est le cas des grands échantillons alors
   $$\mathrm{I}_{0.9}(\mu) = \left[\bar{x} \pm z_{0.95}\frac{S}{\sqrt{n}}\right] = [\mathbf{62.43}\ ;\ \mathbf{73.56}].$$
   ($z_{0.95} = 1.64$ est le quantile d'ordre 0.95 d'une normale standard).
   — Le cas $n = 9$ est le cas des petits échantillons alors en supposant que $X \sim \mathcal{N}(\mu, \sigma^2)$ on a
   $$\mathrm{I}_{0.9}(\mu) = \left[\bar{x} \pm t_{8;0.95}\frac{S}{\sqrt{n}}\right] = [\mathbf{53.12}\ ;\ \mathbf{82.88}].$$
   ($t_{8;0.95} = 1.86$ est le quantile d'ordre 0.95 de la loi de Student à 8 dll).
2. La fréquence empirique des employés satisfaits de la formation est $\bar{p} = 40/50 = 0.8$
   a) On a un grand ESAR de taille $n = 50 \geq 30$, alors
   $$\mathrm{I}_{0.9}(p) = \left[\bar{p} \pm z_{0.95}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = [\mathbf{70.72}\%, \mathbf{89.28}\%].$$
   b) Supposons maintenant que les employés sont choisis au hasard et sans remise.
      i – Non car le taux de sondage est $f = 50/200 > 5\%$.
      ii – L'erreur type dans le cas d'un ESSR est
      $$\sqrt{\frac{200-50}{200-1}} * \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.8681 * 0.0566 = \mathbf{0.0491}.$$
      Si on suppose que le $\bar{P}$ reste normale, l'intervalle de confiance à 90% de la proportion des employés satisfaits de la formation est
      $$\mathrm{I}_{0.9}(p) \approx [\bar{p} \pm z_{0.95} * 0.0491] \approx [\mathbf{72}\%\ ;\ \mathbf{88}\%].$$

Solution de l'Exercice 17

1. La taille d'échantillon est $n = 12$, il s'agit du cas des petits échantillons, avec l'hypothèse de la normalité, par conséquent l'intervalle de confiance à 95% de μ est :

$$\mathrm{I}_{0.95}(\mu) = \left[\bar{X} \pm t_{11;0.975}\frac{S}{\sqrt{n}}\right].$$

La table des quantiles de Student indique que $t_{11;0.975} = 2.201$ calcul des statistiques $\bar{x}$, S donne :

$$\bar{x} = \frac{75+77+79+80+77+76+80+81+83+72+80+76}{12} = 78;$$

$$S = \sqrt{12/11}\left(\frac{75^2+77^2+\ldots,+76^2}{12} - 78^2\right)^{1/2} = 3.05;$$

Ainsi on a

$$\mathrm{I}_{0.95}(\mu) = \left[78 \pm 2.201\frac{3.05}{\sqrt{12}}\right] = [\mathbf{76.06}; \mathbf{79.93}].$$

2. L'erreur de première espèce étant fixée à $\alpha = 5\%$, effectuons le test unilatéral :

$$\begin{cases} H_0, & \mu \geq 80; \\ H_1, & \mu < 80. \end{cases}$$

Statistique de test :

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}};$$

loi de T sous H_0 :

$$T \approx t_{11};$$

Règle de décision :

$$RH_0 \text{ si } T_{\text{cal}} < -t_{11;1-\alpha};$$

Décision $\boldsymbol{RH_0}$ car

$$T_{\text{cal}} = \frac{78-80}{3.05/\sqrt{12}} = -2.271 < -t_{11;0.95} = -1.796.$$

3. Calculons la puissance du test ; la probabilité de rejeter l'hypothèse de base alors qu'une hypothèse alternative est vraie :

$$\begin{aligned} P(RH_0|H_1) &= P(RH_0|\mu = 78) \\ &= P(T < -t_{11;1-\alpha}|\mu = 78) \\ &= P\left(\frac{\bar{X}-80}{3.05/\sqrt{12}} < -1.796 \,\middle|\, \frac{\bar{X}-78}{3.05/\sqrt{12}} \sim t_{11}\right) \\ &= P(t_{11} \leq 0.4756) \\ &= F_t(0.4756) \approx \mathbf{67.82\%}. \end{aligned}$$

Solution de l'Exercice 18

La population est normale de moyenne μ inconnue et d'écart-type connu : $\sigma = 1.5$, la taille de l'échantillon est $n = 9$ et ça moyenne arithmétique observée est :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{9} x_i = 24.5.$$

1. Effectuons le test bilatéral $H_0 : \mu = 22.8$ contre $H_1 : \mu \neq 22.8$ au niveau 1%.
   Problème de test : $\begin{cases} H_0, & \mu = 22.8; \\ H_1, & \mu \neq 22.8. \end{cases}$
   Statistique de test :
   $$T = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{24.5-22.8}{1.5/3} = 3.4.$$
   Loi de T sous H_0 : $T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $|T| > z_{1-\alpha/2} = z_{0.995} = 2.58 \Rightarrow \boldsymbol{RH_0}$ car $3.4 > 2.58$.
2. Effectuons le test unilatéral $H_0 : \mu \leq 22.8$ contre $H_1 : \mu > 22.8$ au niveau 1%. Problème de test : $\begin{cases} H_0, & \mu = 22.8; \\ H_1, & \mu > 22.8. \end{cases}$
   Statistique de test :
   $$T = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{24.5-22.8}{1.5/3} = 3.4.$$

Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
Règle de décision : RH_O si $T > z_{1-\alpha} = z_{0.99} = 2.33 \Rightarrow \boldsymbol{RH_0}$ car $3.4 > 2.33$.

3. Effectuons le test unilatéral $H_0 : \mu \leq 22.8$ contre $H_1 : \mu > 22.8$ au niveau 1% si on suppose σ inconnu. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - \bar{x})^2 = 0.74 \Rightarrow S = 0.86$.
   Problème de test : $\begin{cases} H_0, & \mu \leq 22.8; \\ H_1, & \mu > 22.8. \end{cases}$
   Statistique de test :
   $$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{24.5 - 22.8}{0.86/\sqrt{9}} = 5.93.$$
   Loi de T sous $H_0 : T \sim t_{n-1} = t_8$.
   Règle de décision : RH_O si $T > t_{8,1-\alpha} = t_{8,0.99} = 2.897 \Rightarrow \boldsymbol{RH_0}$ car $5.93 > 2.897$.

Solution de l'Exercice 19

Le bureau d'analyse économique du ministère américain du commerce a indiqué que le revenu annuel moyen d'un habitant de Caroline du Nord était de 18 688 dollars (USA Today, 24 août 1995). Un chercheur de l'état de Caroline du Sud veut tester les hypothèses $H_0 : \mu = 18\,688$ contre $H_1 : \mu \neq 18\ 688$, où μ correspond au revenu annuel moyen d'un habitant de Caroline du Sud. Un échantillon de 400 habitants de Caroline du Sud fournit un revenu annuel moyen de 16 860 dollars et un écart-type de 14 624 dollars.

1. Cas : $n > 30$ et σ est inconnu.
   Problème de test : $\begin{cases} H_0, & \mu = 18688; \\ H_1, & \mu \neq 18688. \end{cases}$
   Statistique de test :
   $$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{16860 - 18688}{14624/\sqrt{400}} = -2.5.$$
   Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $|T| > z_{1-\alpha/2} = z_{0.975} = 1.96 \Rightarrow \boldsymbol{RH_0}$ car $|-2.5| = 2.5 > 1.96$.

2. Trouvons un intervalle de confiance pour μ au niveau de confiance 95%.

$$
\begin{aligned}
& P\left[-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}\right] = 1-\alpha \\
\Leftrightarrow\quad & P\left[\bar{X} - 1.96\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{S}{\sqrt{n}}\right] = 0.95.
\end{aligned}
$$

Donc,

$$
\begin{aligned}
IC(0.95) &= \left[\bar{X} - 1.96\frac{S}{\sqrt{n}}, \bar{X} + 1.96\frac{S}{\sqrt{n}}\right] \\
&= \left[16860 - 1.96\frac{14624}{\sqrt{400}}, 16860 + 1.96\frac{14624}{\sqrt{400}}\right] \\
&= \mathbf{[15426.84, 18293.15]}.
\end{aligned}
$$

Solution de l'Exercice 20

1. Effectuons un test au niveau 5% pour vérifier que la publicité A a été mieux perçue que la publicité B. On a : $\hat{p}_1 = 63/150 = 0.42$, $\hat{p}_2 = 60/200 = 0.3$ et $\hat{p} = \frac{63+60}{150+200} = 0.3514$.
   Problème de test : $\begin{cases} H_0, & p_1 = p_2; \\ H_1, & p_1 > p_2. \end{cases}$
   Statistique de test :

$$
T = \frac{\hat{p}_1 - \hat{p}_2}{\hat{p}(1-\hat{p}) * (\frac{1}{n_1} + \frac{1}{n_2})} \Rightarrow T_{\text{cal}} = \frac{0.42 - 0.30}{0.35 * 0.65 * (1/150 + 1/200)} = 2.33.
$$

   Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $T > z_{1-\alpha} = z_{0.95} = 1.64 \Rightarrow \boldsymbol{RH_0}$ car $2.33 > 1.64$.

2. Construisons un intervalle de confiance à 95% pour l'écart entre les

deux proportions.

$$P\left[-z_{1-\alpha/2} \leq \frac{(p_1-p_2)-(\hat{p}_1-\hat{p}_2)}{\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n_2}}} \leq z_{1-\alpha/2}\right] = 1-\alpha$$

$$P\Big[(\hat{p}_1-\hat{p}_2)-1.96*\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n_2}} \leq (p_1-p_2)$$

$$\leq (\hat{p}_1-\hat{p}_2)+1.96*\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n_2}}\Big] = 0.95$$

Donc, $IC(0.95) = [0.12 \pm 1.96*\sqrt{\frac{0.42*0.58}{150}+\frac{0.3*0.7}{200}}] = \mathbf{[0.019\ ;\ 221]}$.

Solution de l'Exercice 21

Considérons une population normale. On a obtenu, pour un échantillon aléatoire simple de taille 20, $S^2 = 10$.

1. Au niveau $\alpha = 5\%$, considérons le problème de test :
$$\begin{cases} H_0, & \sigma^2 = 12; \\ H_1, & \sigma^2 \neq 12. \end{cases};$$
la statistique de test :
$$T = \frac{(n-1)S^2}{\sigma_0^2} \Rightarrow T_{\text{cal}} = \frac{19*10}{12} \approx 15.83;$$
loi de T sous H_0
$$T \sim \chi^2_{n-1} = \chi^2_{19};$$
Règle de décision : RH_O si
$$T \notin [\chi^2_{19,\alpha/2}, \chi^2_{19,1-\alpha/2}] = [\chi^2_{19,0.025}, \chi^2_{19,0.975}] = [8.91, 32.85]$$
$\Rightarrow \overline{\boldsymbol{RH_0}}$ car $15.83 \in [8.91, 32.85]$.
2. Testons l'hypothèse $H_0 : \sigma^2 = 8$ contre $H_1 : \sigma^2 < 8$ au niveau $\alpha = 1\%$.
Problème de test : $\begin{cases} H_0, & \sigma^2 = 8; \\ H_1, & \sigma^2 < 8. \end{cases}$
Statistique de test :
$$T = \frac{(n-1)S^2}{\sigma_0^2} \Rightarrow T_{\text{cal}} = \frac{19*10}{8} = 23.75.$$

Loi de T sous $H_0 : T \sim \chi^2_{19}$.
Règle de décision : RH_O si $T < \chi^2_{19,\alpha} = \chi^2_{19,0.01} = 7.63 \Rightarrow \overline{\boldsymbol{RH_0}}$ car $23.75 > 7.63$.

3. Testons l'hypothèse $H_0 : \sigma^2 = 12$ contre $H_1 : \sigma^2 > 12$ au niveau $\alpha = 0.5\%$.
   Problème de test : $\begin{cases} H_0, & \sigma^2 = 12; \\ H_1, & \sigma^2 > 12. \end{cases}$
   Statistique de test :
$$T = \frac{(n-1)S^2}{\sigma_0^2} \Rightarrow T_{\text{cal}} = \frac{19 * 10}{12} = 15.83.$$
   Loi de T sous $H_0 : T \sim \chi^2_{19}$.
   Règle de décision : RH_O si $T > \chi^2_{19,1-\alpha} = \chi^2_{19,0.995} = 38.58 \Rightarrow \overline{\boldsymbol{RH_0}}$ car $15.83 < 38.58.63$.

Solution de l'Exercice 22

1. Construisons un intervalle de confiance à 95% pour μ l'espérance de la quantité consommée de lessive.
   Cas : $n > 30$ et σ est inconnu ! Le TCL donne :
$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \mathcal{N}(0,1)$$
   .
$$\begin{aligned} & P\left[-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}\right] = 1 - \alpha \\ \Leftrightarrow \quad & P\left[\bar{X} - 1.96\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{S}{\sqrt{n}}\right] = 0.95. \end{aligned}$$
   Donc,
$$\begin{aligned} IC_{0.95}(\mu) &= \left[\bar{X} - 1.96\frac{S}{\sqrt{n}}, \bar{X} + 1.96\frac{S}{\sqrt{n}}\right] \\ &= \left[14 - 1.96\frac{2}{\sqrt{100}}, 14 + 1.96\frac{2}{\sqrt{100}}\right] \\ &= \mathbf{[13.6, 14.39]}. \end{aligned}$$

2. Un fabricant affirme que 25% des ménages préfèrent la lessive sans phosphates alors que le consultant responsable de l'étude estime qu'il y en a plus que 25%. Réalisez un test pour départager ces deux opinions (Utiliser un risque de première espèce $\alpha = 5\%$).
   Problème de test : $\begin{cases} H_0, & p = 0.25; \\ H_1, & p > 0.25. \end{cases}$
   Statistique de test :
   $$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \Rightarrow T_{\text{cal}} = \frac{0.30 - 0.25}{\sqrt{0.25 * 0.75/100}} \approx 1.155.$$
   Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $T > z_{1-\alpha} = z_{0.95} = 1.64 \Rightarrow$ ***NRH*$_0$** car $1.155 < 1.64$.
   Conclusion : Avec un risque d'erreur limité à 5%, rien ne permet de rejeter l'affirmation du fabricant qui stipule que 25% des ménages préfèrent la lessive sans phosphates.

Solution de l'Exercice 23

1. Un chercheur a suggéré que l'une des raisons pour lesquelles certains individus attendent les cinq derniers jours pour remplir leur déclaration est qu'en moyenne, ces individus ont un remboursement inférieur à celui des personnes qui remplissent leur déclaration relativement tôt. Posons μ Le remboursement moyen des 'individus de dernières minutes'. Problème de test : $\begin{cases} H_0, & \mu = 1056; \\ H_1, & \mu < 1056. \end{cases}$

2. Le remboursement moyen d'un échantillon de 400 individus qui ont rempli leur déclaration entre le 10 et le 15 avril, était de 910 dollars et l'écart-type de 1600 dollars.
   Cas $n > 30$ et σ inconnu. Statistique de test :
   $$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{910 - 1056}{1600/\sqrt{400}} \approx -1.82.$$
   Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $T < z_\alpha = z_{0.05} = -1.64 \Rightarrow$ ***RH*$_0$** car $-1.82 < -1.64$.
   Conclusion On effet, la suggestion du chercheur est envisageable.

Solution de l'Exercice 24

1. Les tailles d'échantillons sont $n_A = 10$ et $n_B = 16$, il s'agit du cas des petits échantillons, avec l'hypothèse de la normalité, par conséquent l'intervalle de confiance à 95% de μ_A et μ_B respectivement sont :

$$\mathrm{I}_{0.95}(\mu_A) = \left[\bar{x}_A \pm t_{9;0.975}\frac{S_A}{\sqrt{10}}\right] = [\mathbf{-0.64}; \mathbf{3.64}],$$

$$\mathrm{I}_{0.95}(\mu_B) = \left[\bar{x}_B \pm t_{15;0.975}\frac{S_B}{\sqrt{16}}\right] = [\mathbf{2.35}; \mathbf{4.44}].$$

2. Considérons $\hat{\mu}$ un estimateur de μ l'écart entre les deux prix moyens :

$$\widehat{\mu} = \widehat{\mu_B - \mu_A} = \hat{\mu}_B - \hat{\mu}_A = \boldsymbol{\bar{X}_B - \bar{X}_A}.$$

Effectuons le test unilatéral :

$$\begin{cases} H_0, & \mu_B \leq \mu_A; \\ H_1, & \mu_B > \mu_A. \end{cases} \Leftrightarrow \begin{cases} H_0, & \mu \leq 0; \\ H_1, & \mu > 0. \end{cases}$$

Statistique de test :

$$T = \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\frac{S_B^2}{n_B} + \frac{S_A^2}{n_A}}}$$

loi de T sous H_0 :

$$T \sim t_{n_A+n_B-2};$$

Règle de décision :

$$RH_0 \text{ si } T_{\text{cal}} > t_{24;0.95} = 1.711.$$

La décision est $\boldsymbol{RH_0}$ car $T_{\text{cal}} = \dfrac{3.4 - 1.5}{\sqrt{1.96^2/16 + 2.99^2/10}} = 1.78 > 1.711$.

3. L'intervalle de confiance à 95% des variances des prix des deux actions sont respectivement :

$$\mathrm{I}_{0.95}(\sigma_A^2) = \left[\frac{(n_A - 1)S_A^2}{\chi^2_{n_A-1;0.975}}, \frac{(n_A - 1)S_A^2}{\chi^2_{n_A-1;0.025}}\right] = \mathbf{[4.23; 29.8]};$$

$$\mathrm{I}_{0.95}(\sigma_B^2) = \left[\frac{(n_B - 1)S_B^2}{\chi^2_{n_B-1;0.975}}, \frac{(n_B - 1)S_B^2}{\chi^2_{n_B-1;0.025}}\right] = \mathbf{[2.1, 9.2]}.$$

4. Au niveau 5%, effectuons le test unilatéral :

$$\begin{cases} H_0, & \sigma_A^2 \leq 9; \\ H_1, & \sigma_A^2 > 9. \end{cases}$$

Statistique de test :

$$T = \frac{(n_A - 1)S_A^2}{\sigma_0^2} = \frac{9S_A^2}{9}$$

loi de T sous H_0 :

$$T \sim \chi^2_{n_A - 1};$$

Règle de décision :

$$RH_0 \text{ si } T_{\text{cal}} > \chi^2_{9;0.95} = 16.92.$$

Décision : ne pas rejeter H_0 car $T_{\text{cal}} = 8.94 < 16.92$.
En conclusion, on ne peut affirmer avec une certitude de 95%, que les données indiquent que la variance du prix de l'action A est supérieur à 9.

5. Au niveau 5%, effectuons le test unilatéral :

$$\begin{cases} H_0, & \sigma_B^2 \geq 9; \\ H_1, & \sigma_B^2 < 9. \end{cases}$$

Statistique de test :

$$T = \frac{(n_B - 1)S_B^2}{\sigma_0} = \frac{15S_B^2}{9}$$

loi de T sous H_0 :

$$T \sim \chi^2_{n_B - 1};$$

Règle de décision :

$$RH_0 \text{ si } T_{\text{cal}} < \chi^2_{15;0.05} = 7.261.$$

Décision : rejeter H_0 car $T_{\text{cal}} = 6.4 < 7.261$.
En conclusion, on peut affirmer avec une certitude de 95%, que les données indiquent que la variance du prix de l'action B est inférieur à 9.

Solution de l'Exercice 25

1. L'étude indique qu'il n'y a que deux valeurs possibles de μ : soit $\mu = 20$ et l'animal est sain, soit $\mu = 30$ et l'animal est malade.

   On veut tester l'hypothèse S : "*l'animal est sain*" contre l'hypothèse M : "*l'animal est malade*". D'après la première étude, S :"$\mu = 20$" et M :="$\mu = 30$". Il est clair que pour le ministère de la santé, l'erreur la plus grave consiste à "*laisser passser*" un animal malade. On considère donc M comme hypothèse de base et S comme hypothèse alternative.

2. Soit un seuil $\alpha = 5\%$. Le seuil du test est la probabilité de rejeter l'hypothèse de base alors qu'elle est vraie. Ici, c'est donc la probabilité de laisser passer un animal sain. L'autre erreur possible consiste à accepter M alors que S est vraie. La probabilité de commettre cette seconde erreur est $1 - \beta$, où β est la puissance du test. En effet, la puissance du test est par définition la probabilité de rejeter l'hypothèse de base alors que l'hypothèse alternative est vraie.
3. Formulons un test, au seuil 5%.
   Cas normale et σ connu. Statistique de test :
   $$T = \frac{\bar{X} - 30}{2/\sqrt{3}}.$$
   où $\bar{X} = \frac{X_1+X_2+X_3}{3}$ est la moyenne des trois mesure éffectués sur l'animlal.
   Loi de T sous $H_0 : T \sim \mathcal{N}(0,1)$.
   Règle de décision : RH_O si $T < z_{0.05} = -1.64 \Leftrightarrow RH_O$ si $\bar{X} < 28.10\%$.

4. La puissance de ce test est
   $$\begin{aligned} P(RH_0|H_1) &= P(RH_0|S) \\ &= P(\bar{X} \leq 28.10|\mu = 20) \\ &= P\left(\frac{\bar{X}-20}{2/\sqrt{3}} \leq \frac{28.10-20}{2/\sqrt{3}} \,\Big|\, \frac{\bar{X}-20}{2/\sqrt{3}} \sim \mathcal{N}(0,1)\right) \\ &= P(\mathcal{N}(0,1) \leq 7.02) \\ &= F(7.02) \approx \mathbf{1}. \end{aligned}$$
   Le test est donc très puissant.

5. D'après la seconde étude, on a toujours M :"$\mu = 30$", mais par contre S devient S :"$\mu < 30$", qui est une hypothèse multiple. Dans la question 3, on voit que la forme de la région de rejet ne dépend pas de la valeur de μ prise pour l'hypothèse simple S, pourvu que cette valeur soit inférieure à 30. On en déduit que le test précédent est le test uniformément le plus puissant de M contre S. On ne modifie donc pas notre test

Solution de l'Exercice 26

1. Ayant observer la variable X : 'taille' sur 1000 individus, on se demande si cette variable peut être ajusté par une loi normale. X est groupée dans $J = 5$ classes. Utilisons le test d'ajustement du Chi-deux.
   Le problème de test est

$$\begin{cases} H_0 : & X \sim \mathcal{N}(\mu, \sigma^2) \\ H_1, & X \nsim \mathcal{N}(\mu, \sigma^2). \end{cases}$$

   Les paramètres μ et σ^2 de la loi normale étant inconnus, on doit les estimer, respectivement, à partir des caractéristiques de l'échantillon $\bar{x}$ et s. Puis on calcul les probabilités théoriques sous H_0 :

$$p_j = P(X \in I_j) = F_Z(X_j^+) - F_Z(X_j^-),$$

   ensuite on déduit les effectifs théoriques $n_j^* = n * p_j$. Tout calcul fait on a

Classes	$\mathbf{n_j}$	$\mathbf{x_j}$	$\mathbf{n_j\, x_j}$	$\mathbf{n_j\, x_j^2}$	$\mathbf{p_{j0}}$	$\mathbf{n_j^*}$	$\mathbf{\frac{(n_j-n_j^*)^2}{n_j^*}}$
<160	1	160	160	25600	0,0053	5,337	3,52
[160,165[	40	162.5	6500	1056250	0,0402	40,173	0,0007
[165,175[	470	170	79900	13583000	0,4689	468,95	0,0023
[175,185[	447	180	80460	14482800	0,4465	446,55	0,0004
>185	42	185	7770	1437450	0,0390	38,99	0,2326
Somme	1000		174785	30583525	1	1000	3,76

   Vérification des hypothèses de travail : $n = 1000 \geq 30$, Tous les effectifs théoriques $n^* = n * p_{j0}$ sont supérieurs ou égaux à 1 et tous (plus de 80%) sont supérieurs à 5.
   Estimation des paramètres de la loi théorique : $\bar{x} = 174785/1000 =$

$174,79$, $s = \sqrt{(\sum n_j x_j^2 - n\bar{x}^2)/(n-1)} = 5,79$.
Calcul de la statistique-test :

$$D = \sum_{j=1}^{J} \frac{(n_j - n_j^*)^2}{n_j^*} = 3.76.$$

Exécution de la règle de décision : le degré de liberté de la statistique-test est $(J-3) = 2$; car on a estimer deux paramètres. Puisque

$$D = 3.76 < \chi_{2;0.95} = 5.99,$$

alors rien ne permet de refuser H_0.
En conclusion, au niveau d'erreur 5%, on peut 'accepter' que la taille des individus suit une loi normale.

Solution de l'Exercice 27

1. Il s'agit de tester l'indépendance de deux caractères qualitatifs : X le niveau scolaire à $J = 3$ modalités et Y l'appartenance politique à $K = 3$ modalités. Utilisons donc le test d'indépendance du Chi-deux. Le problème de test est

$$\begin{cases} H_0 : & \text{indépendance} \\ H_1, & \text{dépendance.} \end{cases}$$

On calcul les effectifs marginaux $n_{j.}$, $n_{.k}$ et les effectifs théoriques sous H_0 :

$$n_{jk}^* = \frac{n_{j.} * n_{.k}}{n}.$$

Tout calcul fait on a

Niveau scolaire	**Appartenance politique**			$n_{.k}$
	Gauche	Droite	Centre	
Lycée	28	28	14	70
Baccalauréat	32	32	16	80
Universitaire	40	40	20	100
$n_{j.}$	100	100	50	250

Le degré de liberté de la statistique-test est $(J-1)*(K-1) = 2*2 = 4$ et la statistique-test calculée est :

$$D = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{jk} - n_{jk}^*)^2}{n_{jk}^*} = \frac{(40-28)^2}{28} + \ldots + \frac{(25-20)^2}{20} = 13.41$$

ainsi, $D = 13.41 > \chi_{4;0.9} = 7.78$, ce qui implique qu'il faut refuser H_0. En conclusion au niveau de confiance 90% on peut affirmer qu'il y a dépendance entre le niveau scolaire et l'appartenance politique.

N. B. *Même au niveau de confiance 99%, on peut affirmer qu'il y a dépendance entre le niveau scolaire et l'appartenance politique, puisque* $D = 13.41 > \chi_{4;0.99} = 13.27$

Solution de l'Exercice 28

1. Il s'agit de tester l'homogénéité de la proportion des individus qui se se souviennent du slogan de la pub sur deux échantillons : ceux qui ont suivi la pub A et ceux qui ont suivi la pub B. Posons p_A et p_B, respectivement, les probabilités qu'une personne se souviennent du slogan de la pub A et la pub B.
Le problème de test est

$$\begin{cases} H_0 : & p_A = p_B \\ H_1, & p_A \neq= p_B. \end{cases}$$

Utilisons donc le test d'l'homogénéité du Chi-deux. Pour cela dressons le tableau de contingence. Désignons par **Oui** la modalité : se se souvenir du slogan et par **Non** la modalité complémentaire.

Publicité	**OUI**	**NON**	$n_{.k}$
A	63	87	150
B	60	140	200
$n_{j.}$	123	227	350

On calcul les effectifs théoriques sous H_0 :

$$n^*_{jk} = \frac{n_{j.} * n_{.k}}{n}.$$

Tout calcul fait on a

Publicité	**OUI**	**NON**	$n_{.k}$
A	52.71	97.28	150
B	70.28	129.71	200
$n_{j.}$	123	227	350

Le degré de liberté de la statistique-test est $(J-1)*(K-1) = 1*1 = 1$ et la statistique-test calculée est :

$$D = \sum_{j=1}^{2}\sum_{k=1}^{2} \frac{(n_{jk} - n^*_{jk})^2}{n^*_{jk}} = \frac{(63 - 52.71)^2}{52.71} + \ldots + \frac{(140 - 129.71)^2}{129.71} = 5.42$$

ainsi, $D = 5.42 > \chi_{1;0.95} = 3.841$, ce qui implique qu'il faut refuser H_0. En conclusion au niveau de confiance 95% on peut affirmer qu'il y a une différence de qualité entre les deux publicités.

Solution de l'Exercice 29

1. On a un grand échantillon de de taille $n = 1000 >> 30$, et la fréquence dans l'échantillon des favorables à l'interdiction est $\bar{p} = 670/1000 = 0.67$ alors

$$\begin{aligned} \mathrm{I}_{0.99}(p) &= \left[\bar{p} \pm z_{1-\alpha/2}\sqrt{\bar{p}(1-\bar{p})/n}\right] \\ &= \left[0.67 \pm 2.5758\sqrt{0.67 * 0.3/1000}\right] \\ &= [\mathbf{63.17\%}, \mathbf{70.83\%}]. \end{aligned}$$

2. La taille de l'échantillon doit vérifier :

$$\begin{aligned} z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} &\leq 0.03 \\ \Leftrightarrow z_{1-\alpha/2}^2 \frac{p(1-p)}{0.03^2} &\leq n \end{aligned}$$

Puisque $\forall p \in [0,1]$, $p * (1-p) \leq 1/4$ alors il suffit que :

$$n \geq \frac{z_{1-\alpha/2}^2}{4 * 0.03^2} \simeq \mathbf{1843}$$

3. Le problème de test est unilateral

$$\begin{cases} H_0 : p \geq 0.7, & ; \\ H_1 : p < 0.7, & . \end{cases}$$

On est dans le cas des grands échantillons, on rejette H_0 si

$$T = \frac{\bar{p} - 0.7}{\sqrt{0.7 * 0.3/1000}} < z_{0.01}$$

Puisque le calcul de la statistique donne $T = -2.07 > -2.32$ alors rien ne permet de refuser l'hypothèse nulle.

Solution de l'Exercice 30

Pour déterminer l'âge moyen de ses clients, une grande entreprise de confection pour hommes prélève un échantillon aléatoire de 50 clients et trouve $\bar{x} = 36$. On suppose que la variance des âge de ses clients est $\sigma^2 = 144$.

1. La taille de l'échantillon étant $n = 50 \geq 30$ alors on est dans le cas des grands échantillons par conséquent l'intervalle de confiance à 95% pour l'âge moyen μ de l'ensemble des clients est

$$\mathrm{I}_{0.95}(\mu) = \left[\bar{x} \pm z_{1-\alpha/2}\sigma/\sqrt{n}\right] = \left[36 \pm 1.96\sigma/\sqrt{50}\right] = [\mathbf{32.67}\ ;\ \mathbf{39.33}].$$

2. La longueur de l'intervalle est réduit à 2 années implique que

$$2 * z_{(1-\alpha)/2}\sigma/\sqrt{n} \leq 2 \Leftrightarrow \sqrt{n} \geq 1.96 * 12 \Leftrightarrow n \geq 553.19.$$

Donc, il faut prendre un échantillon de taille minimum égale à $\mathbf{n} = \mathbf{554}$.

3. Cette valeur n'est pas surprenante car si on prend l'estimation la moins favorable c. à. d. celle par défaut $\hat{\mu} = 32.67$ de μ on a :

$$\begin{aligned} X \sim \mathcal{N}(32.67, 144) \quad &\Rightarrow \quad P(X \leq 18) \approx P(Z \leq (18 - 32.67)/12) \\ &\Rightarrow \quad P(X \leq 18) \approx \mathbf{11.12}\% >> \mathbf{0}\%. \end{aligned}$$

Solution de l'Exercice 31

1. Au niveau de probabilité de 0.05, doit-on rejeter les hypothèses suivantes : $P[X = 0] = 0.35$, $P[X = 1] = 0.30$, $P[X = 2] = 0.20$, $P[X = 3] = 0.10$, $P[X > 4] = 0.05$?

x_j	0	1	2	3	≥ 4
n_j	41	32	17	7	3
$n * p_j^0$	100*0.35=35	30	20	10	5
$(n_j - np_j^0)^2$	36	4	9	9	4
$\frac{(n_j - np_j^0)^2}{np_j^0}$	1.03	0.13	0.45	0.9	0.8

Vérification des hypothèses de travail : $n = 100 > 30$, Tous les effectifs théoriques $n * p_j^0$ sont supérieurs ou égaux à 1 et tous (plus de 80%) d'entre eux sont supérieurs à 5.

Statistique de test : $D = \sum_{j=1}^{5} = \frac{(n_j - np_j^0)^2}{np_j^0} = 3.31$.
Loi de D sous H_0 : $D \sim \chi^2_{J-1} = \chi^2_4$.
Règle de comportement : $RH0$ si $D > \chi^2_{4;0.95} = 9.49 \Rightarrow \boldsymbol{NRH_0}$ car $3.31 < 9.49$.

2. X suit une loi de Poisson de paramètre 0.9 ?

x_j	0	1	2	3	4	5	6
n_j	41	32	17	7	2	1	0
p_j^0	0.4066	0.3659	0.1646	0.0494	0.0112	0.002	0.0003
np_j^0	40.66	36.59	16.46	4.94	1.12	0.2	0.03

Vérification des hypothèses de travail : $n = 100 > 30$, Tous les effectifs théoriques np_j^0 sont strictement positifs mais moins de 80% d'entre eux sont supérieurs à 5, d'où la nécessité de regrouper les dernières classes et considérer la classe ≥ 3 avec effectif 10 et effectif théorique 6.29.
Statistique de test : $D = \sum_{j=1}^{4} = \frac{(n_j - np_j^0)^2}{np_j^0} = 2.78$.
Loi de D sous H_0 : $T \sim \chi^2_{J-1} = \chi^2_3$.
Règle de comportement : $RH0$ si $D > \chi^2_{3;0.95} = 7.81 \Rightarrow \boldsymbol{NRH_0}$ car $2.78 < 7.81$.

Solution de l'Exercice 32

1. On a un grand échantillon de de taille $n = 596 >> 30$, et la fréquence des personnes favorables à la réduction du temps de travail est $\bar{p} = 83/596 = 0.1393$ alors

$$\begin{aligned} \mathrm{I}_{0.95}(p) &= \left[\bar{p} \pm z_{0.975}\sqrt{\bar{p}(1-\bar{p})/n}\right] \\ &= \left[0.1393 \pm 1.96\sqrt{0.1393 * 0.8607/596}\right] \\ &= [\mathbf{11.15\%}, \mathbf{16.71\%}]. \end{aligned}$$

2. Le problème de test est unilateral

$$\begin{cases} H_0 : p \leq 0.12, & ; \\ H_1 : p > 0.12, & . \end{cases}$$

On est dans le cas des grands échantillons, on rejette H_0 si

$$T = \frac{\bar{P} - 0.12}{\sqrt{0.12 * 0.88/596}} > z_{0.95} = 1.6945$$

Puisque le calcul de la statistique donne $T_{\text{cal}} = 1.45 < 1.6945$ alors rien ne permet de refuser l'hypothèse nulle.

Solution de l'Exercice 33

1. On considère un échantillon $(X_1, ..., X_n)$ de loi $\mathcal{N}(\mu, 1)$, où μ est inconnu. On estime μ par la moyenne empirique de l'échantillon

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

a) La loi de $\hat{\mu}$ est normale comme somme de n variables normales, de moyenne $E[\hat{\mu}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} * n * \mu = \mu$ et de variance $V[\hat{\mu}] = \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] = \frac{1}{n} * n * 1 = 1$; en résumé $\boldsymbol{\hat{\mu} \sim \mathcal{N}(\mu, 1)}$.

b) La vraisemblance du paramètre μ est

$$\begin{aligned} L(x_1, ..., x_n; \mu) &= \prod_{i=1}^{n} f(x_i) \\ &= \frac{1}{(\sqrt{2\pi})^n} * e^{\sum_{i=1}^{n} -(x_i - \mu)^2/2}. \end{aligned}$$

La log-vraisemblance du paramètre μ est

$$L^*(\mu) = \ln L(x_1, ..., x_n; \mu) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Montrons que cet estimateur est efficace.
On a $Var[\hat{\mu}] = 1/n$, $\frac{\partial L*}{\partial \mu} = \sum_{i=1}^{n} (x_i - \mu) \Rightarrow \frac{\partial^2 L*}{\partial \mu^2} = -n$ et

$$\begin{aligned} I(\mu) &= E[-\frac{\partial^2 L*}{\partial \mu^2}] \\ &= E[n] \\ &= n. \end{aligned}$$

Comme $I(\mu) = (Var[\hat{\mu}])^{-1}$ et $\hat{\mu}$ est un estimateur sans biais ; en effet $E[\hat{\mu}] = \mu$; alors $\hat{\mu}$ est efficace.

2. Soient $\bar{X}_1$ la moyenne d'un échantillon aléatoire de taille n d'une population normale avec moyenne μ, et variance σ_1^2, et $\bar{X}_2$ la moyenne d'un échantillon aléatoire de taille n d'une population normale avec moyenne μ, et variance σ_2^2. On suppose que les deux échantillons sont indépendants.

a) On a

$$E[\bar{X}_\omega] = \omega E[\bar{X}_1] + (1-\omega)E[\bar{X}_2] = \omega\mu + (1-\omega)\mu = \mu,$$

d'où $\bar{X}_\omega$est un estimateur sans biais de μ.

b) Minimisons la la variance de de $\bar{X}_\omega$:

$$\begin{aligned} f(\omega) &= V(\bar{X}_\omega) = \omega^2 V(\bar{X}_1) + (1-\omega)^2 V(\bar{X}_1) \\ &= \omega^2\sigma_1^2/n + (1-\omega)^2\sigma_2^2/n \\ &= \omega^2((\sigma_1^2+\sigma_2^2)/n) - (2\sigma_2^2/n)\omega + \sigma_2^2/n \end{aligned}$$

$$f'(\omega) = (2(\sigma_1^2+\sigma_2^2)/n)\omega - 2\sigma_2^2/n = 0 \Leftrightarrow \omega = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2};$$

Puisque la fonction f est concave ($f" > 0$), alors elle admet un minimum absolu au point critique

$$\omega = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

Solution de l'Exercice 34

1. Comme $n = 100 >> 30$, il s'agit du cas des grands échantillons donc d'après le TCL :

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 750\,000}{5000} \approx \mathcal{N}(0,1).$$

2. Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de 10 000 Dhs de la moyenne de la population.
3. La probabilité que la moyenne d'échantillon s'écarte au plus de 10 000 Dhs de la moyenne population est

$$\begin{aligned} P(|\bar{X} - 750\,000| \leq 10\,000) &= P(|Z| \leq 10000/5000) \\ &= P(|Z| \leq 2) \\ &= 2F_Z(2) - 1 = 2 * 0.9792 - 1 = \mathbf{95.84}\%. \end{aligned}$$

4. Pour estimer le prix moyen de la population avec un écart maximum de 10 000 Dhs, avec une certitude de 99% il faut que

$$\begin{aligned} n &> (z_{1-\alpha/2}\sigma/\Delta)^2 \\ &> (z_{0.995} 50\,000/10\,000)^2 \\ &> 165.12. \end{aligned}$$

En conclusion il faut sélectionner un échantillon de taille supérieure ou égale à 166.

Solution de l'Exercice 35

1. Comme $n = 36 > 30$, il s'agit du cas des grands échantillons ; avec σ inconnu. L'intervalle de confiance à 95% pour la moyenne de la population est

$$I_{0.95}(\mu) = \left[\bar{X} \pm z_{0.975}\frac{S}{\sqrt{n}}\right] = \mathbf{[2599.69, 2700.31]}.$$

2. Effectuez le test d'hypothèse au seuil de signification de 5%. Au niveau $\alpha = 5\%$, testons si le prix moyen d'un billet a augmenté en mars.
Problème de test : $\begin{cases} H_0, & \mu = 2500; \\ H_1, & \mu \neq 2500. \end{cases}$
Statistique de test :

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{2650 - 2500}{154/\sqrt{36}} = 5.84$$

Loi de T sous H_0 : $T \approx \mathcal{N}(0,1)$.
Règle de décision : RH_O si $|T_{\text{cal}}| > z_{1-\alpha/2} = z_{0.975} = 1.96 \Rightarrow \boldsymbol{RH_0}$ car $5.84 > 1.96$.

3. Calculons la puissance du test sous l'alternative $H_1 : \mu = 2600$: la probabilité de rejeter l'hypothèse H_0 alors qu'une hypothèse alternative

$H_1 : \mu = 2600$ est vraie :

$$
\begin{aligned}
P(RH_0|H_1) &= P(RH_0|\mu = 2600) \\
&= P(|T| > 1.96|\mu = 2600) \\
&= P\left(|\frac{\bar{X} - 2500}{25.66}| > 1.96 \left| \frac{\bar{X} - 2600}{25.66} \approx \mathcal{N}(0,1)\right.\right) \\
&= 1 - P\left(|Z + \frac{2600 - 2500}{25.66}| < 1.96\right) \\
&= 1 - P\left(-5.86 < Z < -1.94\right) \approx 1 - F_Z(-1.94) \\
&= F_Z(1.94) \approx \mathbf{97.38\%}.
\end{aligned}
$$

Solution de l'Exercice 36

1. L'estimation du prix moyen d'un billet aller-retour en mars est :

$$\bar{x} = \frac{310 + \ldots, 260}{15} = \mathbf{270\ \$}$$

.

2. Cas des petits échantillons car $n = 15 < 30$; σ est inconnu : il faut supposer que le prix d'un billet $X \sim \mathcal{N}(\mu, \sigma^2)$.
   L'intervalle de confiance dans ce cas 95% du prix moyen est

$$I_{0.95}(\mu) = \left[\bar{X} \pm t_{14;0.975} \frac{S}{\sqrt{n}}\right] = \mathbf{[256.3; 283.7]}.$$

   Notons que L'écart-type empirique est $S = \sqrt{S^2} = \sqrt{(16/15) * 573.33} = 24.73$ et le quantile de la loi de Student est $t_{14;0.975} = 2.145$

3. Au niveau $\alpha = 5\%$, testons si le prix moyen d'un billet a augmenté en mars.
   Problème de test : $\begin{cases} H_0, & \mu = 258; \\ H_1, & \mu > 258. \end{cases}$
   Statistique de test :

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \Rightarrow T_{\text{cal}} = \frac{270 - 258}{24.73/\sqrt{15}} = 1.88.$$

   Loi de T sous H_0 : $T \sim t_{n-1} = t_{14}$.
   Règle de décision : RH_O si $T > t_{14,1-\alpha} = t_{8,0.95} = 1.761 \Rightarrow \boldsymbol{RH_0}$ car $1.88 > 1.761$.

Solution de l'Exercice 37

1. Les trois estimateurs sont sans biais car leurs espérances respectives sont égales à μ, en effet :

$$\begin{aligned} E(\hat{\alpha}_1) &= E\left(\frac{1}{2}\,\hat{\theta}_1 + \frac{1}{2}\,\hat{\theta}_2\right) \\ &= \frac{1}{2}E(\hat{\theta}_1) + \frac{1}{2}E(\hat{\theta}_2) \\ &= \frac{1}{2}\mu + \frac{1}{2}\mu = \mu \end{aligned}$$

$$\begin{aligned} E(\hat{\alpha}_2) &= E\left(\frac{3}{4}\,\hat{\theta}_1 + \frac{1}{4}\,\hat{\theta}_2\right) \\ &= \frac{3}{4}E(\hat{\theta}_1) + \frac{1}{4}E(\hat{\theta}_2) \\ &= \frac{1}{4}\mu + \frac{3}{4}\mu = \mu \end{aligned}$$

$$E(\hat{\alpha}_3) = E(\hat{\theta}_1) = \mu$$

2. Comme les trois estimateurs sont sans biais, alors leurs précisions sont les inverses de leurs variances, comparons donc leurs variances :

$$\begin{aligned} V(\hat{\alpha}_1) &= V\left(\frac{1}{2}\,\hat{\theta}_1 + \frac{1}{2}\,\hat{\theta}_2\right) \\ &= \frac{1}{2^2}V(\theta_1) + \frac{1}{2^2}E(\hat{\theta}_2) \quad \text{(vu l'indépendance)} \\ &= \frac{1}{4}V(\hat{\theta}_1) + \frac{9}{4}E(\hat{\theta}_1) \quad \text{(vu que } V(\hat{\theta}_2) = 3^2 * V(\hat{\theta}_1)) \\ &= \frac{10}{4}V(\theta_1) = 2.5\,V(\theta_1) \end{aligned}$$

$$\begin{aligned} V(\hat{\alpha}_2) &= V\left(\frac{3}{4}\,\hat{\theta}_1 + \frac{1}{4}\,\hat{\theta}_2\right) \\ &= \frac{3^2}{4^2}V(\theta_1) + \frac{1}{4^2}E(\hat{\theta}_2) \\ &= \frac{9}{16}V(\hat{\theta}_1) + \frac{9}{16}V(\hat{\theta}_1) \\ &= \frac{18}{16}V(\theta_1) = 1.25\,V(\theta_1). \end{aligned}$$

$$V(\hat{\alpha}_3) = V\left(\hat{\theta}_1\right).$$

Ainsi

$$V(\hat{\alpha}_3) < V(\hat{\alpha}_2) < V(\hat{\alpha}_1),$$

le meilleur estimateur est $\hat{\alpha}_3$, le plus mauvais est $\hat{\alpha}_1$.

Solution de l'Exercice 38

1. Posons p_1 et p_2 les proportions respectives de la population favorable à la décriminalisation da marijuana pour les années 1980 et 1985. Puisque $n = 1500 >> 30$, on est dans le cas des grands échantillons :

$$\begin{aligned} I_{0.95}(p_1) &= \left[\bar{p}_1 \pm \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n}}\,z_{0.975}\right] \\ &= \left[0.52 \pm 1.96 * \sqrt{\frac{0.52 * 0.48}{1500}}\right] = \mathbf{[49.47\%, 54.53\%]}. \end{aligned}$$

$$\begin{aligned} I_{0.95}(p_2) &= \left[\bar{p}_2 \pm \sqrt{\frac{\bar{p}_2(1-\bar{p}_2)}{n}}\,z_{0.975}\right] \\ &= \left[0.46 \pm 1.96 * \sqrt{\frac{0.46 * 0.54}{1500}}\right] = \mathbf{[43.48\%, 48.52\%]}. \end{aligned}$$

2. Construisons un intervalle de confiance à 95% pour l'écart $p_2 - p_1$ entre les deux proportions.

$$P\left[-z_{1-\alpha/2} \leq \frac{(p_2-p_1)-(\hat{p}_2-\hat{p}_1)}{\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n}}} \leq z_{1-\alpha/2}\right] = 1-\alpha$$

$$P\Bigg[(\hat{p}_2-\hat{p}_1) - 1.96*\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n}} \leq (p_2-p_1)$$
$$\leq (\hat{p}_2-\hat{p}_1) + 1.96*\sqrt{\frac{\hat{p}_1*(1-\hat{p}_1)}{n}+\frac{\hat{p}_2*(1-\hat{p}_2)}{n}}\Bigg] = 0.95$$

Donc,

$$\begin{aligned} IC_{0.95}(p_2-p_1) &= \left[(0.46-0.52) \pm 1.96*\sqrt{\frac{0.46*0.54+0.52+0.48}{1500}}\right] \\ &= \mathbf{[-9.57\%, -2.43\%]}. \end{aligned}$$

Solution de l'Exercice 39

1. Le sondage stratifié est efficace lorsque les variances internes des strates sont

   ☐ importantes ☒ faibles ☐ nulles ☐ égaux

2. On prélève un échantillon aléatoire de 1000 jeunes diplômés de l'UMP, 330 jeunes déclarent qu'ils sont en chômage. L'intervalle de confiance ; à 95% ; du pourcentage des chômeurs parmi les diplômés de l'UMP est

   ☒$[30\%, 35.9\%]$ ☐$[33\%, 36\%]$ ☐$[27\%, 33\%]$ ☐$[36\%, 39\%]$

3. Un échantillon de 50 appartements situées à Oujda est sélectionné. La moyenne du loyer de ces appartements est égale à 1850 Dhs et leurs écart-type est égal à 230 Dhs.
   a) L'intervalle de confiance ; à 95% ; du loyer moyen est

      ☐$[1586, 1641]$ ☒$[1786, 1914]$ ☐$[1850, 1943]$ ☐$[1760, 1980]$

   b) Supposons que le loyer suit la loi normale, la probabilité de trouver un appartement de loyer inférieur à 1600 Dhs est

      ☐5%] ☐10% ☒14% ☐25%

c) Pour être confiant à 95%, que le loyer moyen estimé ne s'éloigne du loyer moyen réel que par 50 Dhs il faut donc prélever un échantillon de taille supérieure à

☐55 ☐66 ☐72 ☒82

Solution de l'Exercice 40

QCM. Les bonnes réponses sont cochées :

1. La moyenne d'un grand échantillon converge en probabilité vers la moyenne de la population

   ☒ Vrai ☐ Faux

2. Les résultats d'un recensement sont plus précis que ceux d'un sondage

   ☒ Vrai ☐ Faux

3. Un sondage est plus coûteux qu'un recensement

   ☐ Vrai ☒ Faux

4. Pour un échantillonnage simple avec remise on a $V(\bar{X}) = \sigma^2/n$

   ☒ Vrai ☐ Faux

5. Un sondage stratifié est adéquat pour une population composé de différents groupes homogènes

   ☒ Vrai ☐ Faux

6. Un sondage par grappes est plus adéquat q'un sondage stratifié

   ☐ Vrai ☒ Faux

7. Un estimateur sans biais est celui de moyenne nulle

   ☐ Vrai ☒ Faux

8. Un estimateur qui a la variance minimale est l'estimateur le plus efficace

   ☒ Vrai ☐ Faux

9. Plus la population est homogène (de variance petite), plus les fluctuations de la moyenne de l'échantillon sont importantes

   ☐ Vrai ☒ Faux

10. Plus la taille de l'échantillon est importante, plus les fluctuations de sa moyenne sont importantes

☐ Vrai ☒ Faux

Solution de l'Exercice 41

1. Soit X une v.a. suivant la loi de poisson de paramètre λ et $X_1, \ldots, X_n$ un ESAR. L'estimateur du maximum de vraisemblance de λ est :

   ☒ $\overline{X}$ ☐ $1/\overline{X}$ ☐ S^2 ☐ $1/S^2$

2. On prélève un échantillon aléatoire simple de 200 clients d'une société de transport, 175 d'entre eux se déclarent satisfait du service. Un responsable déclare que : 'le taux de satisfaction p parmi tous les clients dépasse 83%'.

   (a) L'estimation ponctuelle du taux de satisfaction est

   ☐ 175% ☐ 75% ☒ 87.5% ☐ 89%

   (b) L'intervalle de confiance du pourcentage des clients satisfaits est :

   ☐ $[86.8\%, 88.2\%]$ ☐ $[87\%, 95\%]$ ☐ $[83.7\%, 91.3\%]$ ☒ $[82.3\%, 92.1\%]$

   (c) Pour vérifier son hypothèse, le responsable doit considérer le problème du test suivant :

   ☐ $\begin{cases} H_0 : & p \geq 0.83 \\ H_0 : & p < 0.83, \end{cases}$ ☒ $\begin{cases} H_0 : & p \leq 0.83 \\ H_0 : & p > 0.83, \end{cases}$

   ☐ $\begin{cases} H_0 : & p = 0.83 \\ H_0 : & p \neq 0.83, \end{cases}$ ☐ $\begin{cases} H_0 : & p \geq 0.83 \\ H_0 : & p \leq 0.83, \end{cases}$

   (d) La statistique-test calculée est :

   ☒ $T = 1.69$ ☐ $T = -1.61$ ☐ $T = 1.67$ ☐ $T = 1.63$

   (e) La décision prise est :

   ☒ RH_0 ☐ $\overline{RH_0}$ ☐ RH_1 ☐ Autre

3. On s'intéresse au prix d'une action qu'on suppose normalement distribué. On observe le prix de cette action pendant 20 jours. Le prix moyen observé est 12.5 Dhs et l'écart-type observé est égal à 1.8 Dhs.

   (a) L'estimation par intervalle de confiance du prix moyen de l'action est :

   ☒ $[11.8, 13.19]$ ☐ $[10.66, 12.39]$ ☐ $[9.5, 12.33]$ ☐ $[13.6, 19.64]$

   (b) L'incertitude relative de l'estimation ci-dessus est :

☐ 10% ☐ 5% ☐ 2% ☒ 5.56%

(c) L'intervalle de confiance de l'écart-type du prix de l'action est :

☒ $[1.37, 2.63]$ ☐ $[1.4, 2.74]$ ☐ $[-2.5, -1.9]$ ☐ $[1.6, 5.9]$

(d) On va observer le prix de l'action sur une période dépassant 30 jours. Afin de s'assurer que le prix moyen estimé ne s'éloigne du prix moyen réel que par moins de 0.5 Dhs, il faut observer le prix de l'action pendant au moins :

☐ 35 jours ☐ 40 jours ☒ 50 jours ☐ 90 jours

Solution de l'Exercice 42

1. La valeur $\bar{x}$ prise par $\bar{X}$ est une variable aléatoire

   ☐ Vrai ☒ Faux

2. Une stratification consiste à diviser la population en sous populations homogènes

   ☒ Vrai ☐ Faux

3. Plus la taille de l'échantillon est grande, plus l'erreur-type de la moyenne de l'échantillon est petite

   ☒ Vrai ☐ Faux

4. Pour un échantillon simple sans remise on a $V(\bar{X}) = \sqrt{\frac{N-n}{N-1}} \times \frac{\sigma^2}{n}$

   ☐ Vrai ☒ Faux

5. La distribution de la moyenne d'un échantillon ESAR normale est normale

   ☒ Vrai ☐ Faux

6. Un sondage par grappes est utilisé lorsque la population est composée de différents groupes intérieurement homogènes

   ☐ Vrai ☒ Faux

7. Un estimateur sans biais est convergent si et seulement si sa variance converge vers zéro

   ☒ Vrai ☐ Faux

8. Plus le MSE est faible plus l'estimateur est moins efficace

☐ Vrai ☒ Faux

9. Un E.S.B. $\hat{\theta}_n$ de θ de variance $Var(\hat{\theta}_n) = 1/I_n(\theta)$, où $I_n(\theta)$ est l'information de Fisher. L'estimateur $\hat{\theta}_n$ est dit efficace

☒ Vrai ☐ Faux

10. Pour tester H_0 : X et Y sont dépendantes contre H_0 : X et Y sont indépendantes, on utilise le test de khi-deux

☐ Vrai ☒ Faux

Solution de l'Exercice 43

1. Une banque se demande si elle n'aurait pas accordé trop de prêts immobiliers à des clients pas assez solvables. Pour se rassurer, elle voudrait avoir une majoration du taux de défaillance des prêts en cours. Elle tire alors au hasard 200 clients dans son portefeuille et fait analyser leur situation. Sur ces 200 cas, seuls 8 présentent un risque sérieux de défaillance.
   (a) Le taux moyen de défaillance est

   ☒ 4% ☐ 6% ☐ 8% ☐ 10%

   (b) Un intervalle de confiance à 99% du taux réel de défaillance est

   ☐ $[1,82\%; 9,12\%]$ ☒ $[0,43\%; 7,57\%]$
   ☐ $[1,12\%; 8,56\%]$ ☐ $[3,11\%; 12,65\%]$

2. On veut estimer le coût moyen μ_A de séjour d'un patient dans une clinique A. On sait que le coût par jour d'un patient est 1200 DH. Pour un échantillon de 200 patients on a observé une durée de séjour moyenne de 4,6 jours avec un écart-type de 3,8 jours.
   (a) L'intervalle de confiance à 95% pour le coût moyen d'un patient est

   ☐ $[1560; 8520]$ ☐ $[3363; 8345]$ ☐ $[4476; 7362]$ ☒ $[4888; 6152]$

   (b) Une autre clinique B, dont le coût par jour d'un patient est toujours 1200 DH, réalise la même étude, le calcul sur un échantillon de 300 patients, a donné une durée de séjour moyenne de 5 jours avec un écart-type de 2,5 jours. On se demande si le coût moyen de séjour μ_B dans la clinique B est plus élevé que le coût moyen de séjour μ_A dans la clinique A.

(c) Le problème du test qu'on doit poser est :

☐ $H_0 : \mu_B = \mu_A;\quad H_1 : \mu_B < \mu_A$ ☐ $H_0 : \mu_B < \mu_A;\quad H_1 : \mu_B \geq \mu_A$
☒ $H_0 : \mu_B \leq \mu_A;\quad H_1 : \mu_B > \mu_A$ ☐ $H_0 : \mu_B > \mu_A;\quad H_1 : \mu_B \leq \mu_A$

(d) Pour un niveau de signification $\alpha = 0,05$, la règle de décision sera :

☐ Rejeter H_0 si $\bar{X}_B - \bar{X}_A < 0$ ☐ Rejeter H_0 si $\bar{X}_B - \bar{X}_A \geq 1200$
☒ Rejeter H_0 si $\bar{X}_B - \bar{X}_A > 600$ Rejeter H_0 si $\bar{X}_B - \bar{X}_A \leq 600$

(e) La conclusion du test, au seuil 5%, est :

☐ Il n ya pas de différence entre les coûts par patients dans les deux cliniques
☐ Le coût par patient dans la clinique B est nettement supérieur que celui dans la clinique A
☒ On ne peut confirmer que le coût par patient dans la clinique B est supérieur à celui de la clinique A
☐ On ne peut rien conclure

Solution de l'Exercice 44

Afin d'estimer $\mu = E(X)$, où X est la consommation mensuel finale des ménages par tête, le HCP en 2014 a sélectionné un échantillon (ESAR) de taille ménages $n = 16000$ ménages. La moyenne d'échantillon est notée $\bar{X}$; sa valeur calculée est $\bar{x} = 1325$ DHS ; l'écart-type ajusté de l'échantillon est $S = 530$ DHS.

1. La distribution d'échantillonnage de $\bar{X}$ est

   ☐ $\frac{\bar{X}-\mu}{0{,}033} \approx \mathcal{N}(0,1)$ ☐ $\frac{\bar{X}-\mu}{4{,}19} \approx t_{15999}$
   ☒ $\frac{\bar{X}-\mu}{4{,}19} \approx \mathcal{N}(0,1)$ ☐ $\bar{X} \approx \mathcal{N}(0,1)$

2. La probabilité que la moyenne estimée s'écarte de la valeur réelle μ de plus de 10 DH est :

   ☐ 5% ☐ 1,74% ☐ 2,5% ☒ 0,87%

3. Afin de s'assurer à 99% que la moyenne estimée ne s'écarte de la valeur réelle que de 10 DH au plus, il faut sélectionner un échantillon dont la taille est égale au moins à :

   ☐ 12000 ☒ 18640 ☐ 16000 ☐ 19460

4. On sélectionne du premier échantillon un sous échantillon ESSR de taille $n' = 3000$. Notons par $\bar{X}'$ la moyenne du deuxième échantillon. L'écart-type de $\bar{X}'$ est :

☐ $\sigma_{\bar{X}'} = 8,72$ ☐ $\sigma_{\bar{X}'} = 4,19$ ☒ $\sigma_{\bar{X}'} = 9,67$ ☐ $\sigma_{\bar{X}'} = 530$

Solution de l'Exercice 45

À la veille d'un référendum sur l'adoption d'une nouvelle constitution, on a prélevé un échantillon aléatoire de 1000 votants auxquels on a demandé s'ils vont voter par OUI ou par NON. 520 personnes se déclarent favorables au OUI.

1. Pour une confiance fixé à 95%, l'intervalle de confiance du pourcentage p des personnes favorables au OUI, est :

☐ [45%, 59%] ☐ [47.2%, 56.8%] ☒ [48.9%, 55.1%] ☐ [53.9%, 55.1%]

2. Si l'on souhaite connaître le pourcentage de gens favorables au OUI, avec une précision inférieure ou égale à 2%, la taille de l'échantillon doit être supérieure à :

☐ 900 ☐ 1204 ☒ 2401 ☐ 3631

3. Au niveau $\alpha = 5\%$, le test de l'hypothèse $H_0 : p \leqslant 0.5$ contre l'hypothèse
$H_1 : p > 0.5$ donne :

☐ Victoire du OUI ☒ Rien ne permet d'écarter la victoire du NON
☐ Victoire du NON ☐ Rien ne permet d'écarter la victoire du OUI

Solution de l'Exercice 46

Une entreprise souhaite lancer un nouveau produit et elle veut tester la dépendance entre l'intention d'achat du produit et l'âge du client. Ainsi, elle demande à 400 clients potentiels, s'ils sont prêts à acheter ce nouveau produit. Les résultats en fonction de l'âge des personnes interrogées sont les suivants :

Âge	**Intention d'achat du produit** Oui	Non	Abstention
<30 ans	65	27	8
de 30 à 45 ans	50	19	11
de 45 à 60 ans	35	24	11
$\geq$ 60 ans	50	80	20

1. La statistique utilisée pour tester l'indépendance suit une loi de :

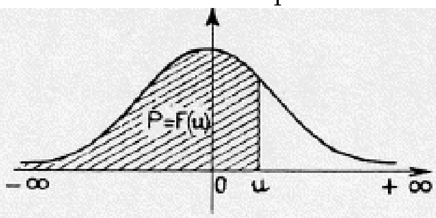
   ☐ Chi-deux à 12 degrés de liberté ☒ Chi-deux à 6 degrés de liberté
   ☐ Student à 3 degrés de liberté

2. Au niveau $\alpha = 5\%$, peut t-on affirmer que l'intention d'achat est **dépendante** de l'âge du client ?

   ☒ Oui ☐ Non ☐ On ne peut conclure.

Tables statistiques

Figure 1 – Fonction de répartition de la loi normale

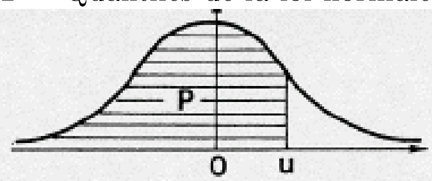


u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de u

u	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
F(u)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

FIGURE 2 – Quantiles de la loi normale



P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,99
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537	0,98
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	0,97
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,96
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,95
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,94
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,93
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,92
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,91
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	0,90
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,89
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,88
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,87
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,86
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407	1,0364	0,85
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,84
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,83
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,82
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,81
0,19	0,8779	0,8742	0,8705	0,8669	0,8633	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,80
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,79
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,78
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,77
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,76
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,75
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,74
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,73
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,72
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,71
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,70
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4959	0,69
0,31	0,4959	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,68
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,67
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,66
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,65
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,64
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,63
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,62
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,61
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,60
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,59
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,58
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,57
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,56
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,55
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,54
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,53
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,52
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,51
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,50
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

Grandes valeurs de u

P	0,9999	0,99999	0,999999	0,9999999	0,99999999	0,999999999
u	3,7190	4,2649	4,7534	5,1993	5,6120	5,9978

N.B. Si $P < 0.5$, u est négatif.

FIGURE 3 – Quantiles de la loi Chi-deux à ν degrés de liberté

	Probabilité p									
ddl v	**0,005**	**0,010**	**0,025**	**0,050**	**0,100**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672

FIGURE 4 – Quantiles de la loi de Student à ν degrés de liberté

	Probabilité p					
ddl v	0,9	0,95	0,975	0,99	0,995	0,999
1	3,078	6,314	12,706	31,821	63,657	318,309
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385

Références

1. Anderson et al. (2004). Statistique pour Économie et Gestion. *Deboeck*.
2. Pupion, P.. (2008) Statistiques pour La Gestion. 2 Eds. *Dunod*.
3. Wonnacott, T., et Wonnacott, R. (1991). Statistique. 4 Eds. *Economica*.

Biographie de l'Auteur

Saïd EL MELHAOUI est Docteur en Modélisation, Probabilités et Statistique Mathématique, diplômé en 2003 de la Faculté des Sciences d'Oujda (Maroc).
Fort d'une riche expérience pédagogique de huit années dans l'enseignement secondaire, il a rejoint l'enseignement supérieur en 2007. Il exerce actuellement en tant qu'enseignant-chercheur au sein du Département des Sciences Économiques de la Faculté des Sciences Juridiques, Économiques et Sociales Oujda (FSJESO).
Spécialiste de la statistique, de l'économétrie et de la Data Science, il contribue activement à la recherche scientifique à travers de nombreuses publications et ouvrages de référence. Ses travaux visent à vulgariser et approfondir l'usage des outils mathématiques, notamment la statistique dans l'analyse économique moderne.

Pour plus de détails sur ses travaux, consultez son : profil ResearchGate