MTH2302D - TD 9

Vincent Perreault

POLYTECHNIQUE MONTRÉAL

Ex. 1 - Énoncé

Soit $\hat{\theta}_1$ et $\hat{\theta}_2$ des estimateurs du paramètre θ dont on sait que $E(\hat{\theta}_1) = \theta$, $E(\hat{\theta}_2) = \theta/2$, $V(\hat{\theta}_1) = 10$ et $V(\hat{\theta}_2) = 4$. Lequel de ces estimateurs est le meilleur ? Dans quel sens ?

- $EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$

Ex. 1 - Réponse

$$E(\hat{\theta}_1) = \theta, \quad Var(\hat{\theta}_1) = 10$$
$$E(\hat{\theta}_2) = \theta/2, \quad Var(\hat{\theta}_2) = 4$$

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

Quel estimateur est le meilleur?

$$EQM(\hat{\theta}_1) = Var(\hat{\theta}_1) + (E(\hat{\theta}_1) - \theta)^2 = 10 + 0^2 = 10$$

$$EQM(\hat{\theta}_2) = Var(\hat{\theta}_2) + (E(\hat{\theta}_2) - \theta)^2 = 4 + \frac{\theta^2}{4}$$

$\hat{\theta}_1$ est meilleur si

$$10 \leq 4 + \frac{\theta^2}{4}$$
$$24 \leq \theta^2$$
$$\theta \geq \sqrt{24}$$

sinon $\hat{\theta}_2$ est meilleur.

Ex. 2 - Énoncé

Le service du génie industriel veut estimer le temps moyen requis pour assembler une carte de circuit imprimé. On considère que le temps requis pour assembler une carte suit une loi normale avec un écart-type de 0,45 minute.

a) Quelle devrait être la taille de l'échantillon si l'on veut être confiant à 95 % que l'erreur d'estimation sera inférieure à 0,25 minute ?

b) Si on avait voulu une marge d'erreur deux fois plus petite, aurait-il fallu que l'échantillon soit deux fois plus grand ?

c) Si on avait utilisé un niveau de confiance de 90 %, aurait-il fallu utiliser un échantillon plus grand ou plus petit que celui en a) pour une même marge d'erreur ?

- Si $X \sim \mathrm{N}(\mu, \sigma^2)$ avec σ^2 connue, alors $\mu \in \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Ex. 2 - Réponse - a) - c)

Quelle taille n pour erreur $\leq$0.25 pour I.C. 95% ?

$X \sim \mathrm{N}(\mu, 0.45^2)$

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq E$$

$$n \geq \Big(\underbrace{z_{\alpha/2}}_{z_{0.05/2}}\frac{\sigma}{E}\Big)^2 = \left(1.96 \cdot \frac{0.45}{0.25}\right)^2 = 12.447$$

$$n \geq 13$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ et σ^2 connue

$\Rightarrow \mu \in \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Pour un E 2x plus petit, n doit-il être 2x plus grand?

Non, n aurait dû être 4x plus grand.

Si $1 - \alpha = 90\%$, n doit-il être $>$ pour un même E?

Non, plus petit car $z_{0.10/2} = 1.645 < 1.96 = z_{0.05/2}$.

Ex. 3 - Énoncé

Soit des ampoules de 75 W dont on sait que la durée de vie, en heures, obéit approximativement à une loi normale, avec un écart-type σ de 25 heures. Les 20 ampoules d'un échantillon aléatoire ont duré en moyenne $\overline{x} = 1014$ heures.

a) Établissez, pour la durée de vie moyenne des ampoules, un intervalle de confiance à 95 %.

b) Supposons qu'à partir de cet échantillon, on a construit, pour la durée de vie moyenne des ampoules, un intervalle de confiance d'une longueur totale correspondant à 8 heures. Quel est le niveau de confiance de cet intervalle ?

c) Pour une expérience future, on veut s'assurer à 95 % que l'erreur commise en estimant la durée de vie moyenne des ampoules sera inférieure à 5 heures. Quelle devra être la taille de l'échantillon utilisé ?

- Si $X \sim \mathrm{N}(\mu, \sigma^2)$ avec σ^2 connue, alors $\mu \in \bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

Ex. 3 - Réponse - a), b)

Déterminer l'I.C. 95% de μ.

$X_i \sim \mathrm{N}(\mu, 25^2), \quad i = 1, 2, ..., 20$

$\bar{x} = 1014$

$$\mu \in \bar{x} \pm \underbrace{z_{\alpha/2}}_{z_{0.05/2}} \frac{\sigma}{\sqrt{n}} = 1014 \pm 1.96 \cdot \frac{25}{\sqrt{20}}$$

$$\approx 1014 \pm 10.957 = [1003.04, 1024.96]$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ et σ^2 connue

$\Rightarrow \mu \in \bar{X} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

Quel est $1 - \alpha$ si on veut un intervalle de longueur 8?

$$2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = L$$

$$z_{\alpha/2} = \frac{L\sqrt{n}}{2\sigma} = \frac{8\sqrt{20}}{2 \cdot 25} \approx 0.716$$

$$\Rightarrow 1 - \alpha/2 = 0.76424 \qquad \Rightarrow 1 - \alpha = 0.52848$$

Ex. 3 - Réponse - c)

Quelle taille n pour erreur $\leq$5 pour I.C. 95% ?

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq E$$

$$n \geq \Bigg(\underbrace{z_{\alpha/2}}_{z_{0.05/2}}\frac{\sigma}{E}\Bigg)^2 = \left(1.96 \cdot \frac{25}{5}\right)^2 = 96.04$$

$$n \geq 97$$

$X_i \sim \mathrm{N}(\mu, 25^2), \quad i = 1, 2, ..., 20$

$\bar{x} = 1014$

$X \sim \mathrm{N}(\mu, \sigma^2)$ et σ^2 connue

$\Rightarrow \mu \in \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Ex. 4 - Énoncé

Le diamètre des tiges d'acier produites par deux extrudeuses fait l'objet d'une étude. On tire deux échantillons aléatoires, l'un de taille $n_1 = 15$ et l'autre de taille $n_2 = 18$. La moyenne et la variance du premier échantillon sont respectivement $\overline{x}_1 = 8{,}73$ et $s_1^2 = 0{,}30$ et celles du second, $\overline{x}_2 = 8{,}68$ et $s_2^2 = 0{,}34$.

a) Établissez pour σ_1^2/σ_2^2 un intervalle de confiance :

i) à 90 % ;

ii) à 95 % (et comparez-le à celui que vous avez obtenu en i).

b) Y a-t-il des raisons de croire que les variances sont significativement différentes ?

c) En supposant que $\sigma_1^2 = \sigma_2^2$ et que cette variable suit une loi normale, établissez un intervalle de confiance à 95 % pour la différence entre les diamètres moyens.

- Si $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$, alors $\dfrac{\sigma_X^2}{\sigma_Y^2} \in \left[\dfrac{S_X^2}{S_Y^2} F_{1-\alpha/2;n_Y-1,n_X-1}, \dfrac{S_X^2}{S_Y^2} F_{\alpha/2;n_Y-1,n_X-1}\right]$

- $F_{1-\alpha;u,v} = \dfrac{1}{F_{\alpha;u,v}}$

- Si $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ et $\sigma_X^2 = \sigma_Y^2$, alors $\mu_X - \mu_Y \in \bar{X} - \bar{Y} \pm t_{\alpha/2;n_X+n_Y-2} S_p \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}$

  où $S_p = \sqrt{\frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}}$

Ex. 4 - Réponse - a)

Déterminer l'I.C. 90 et 95% de σ_1^2/σ_2^2.

$\bar{x}_1 = 8.73, \quad s_1^2 = 0.30, \quad n_1 = 15$

$\bar{x}_2 = 8.68, \quad s_2^2 = 0.34, \quad n_2 = 18$

$$\frac{\sigma_X^2}{\sigma_Y^2} \in \left[\frac{S_X^2}{S_Y^2} F_{1-\alpha/2;n_Y-1,n_X-1}, \frac{S_X^2}{S_Y^2} F_{\alpha/2;n_Y-1,n_X-1}\right]$$

$$= \left[\frac{S_X^2}{S_Y^2} \frac{1}{\underbrace{F_{\alpha/2;n_Y-1,n_X-1}}_{F_{0.1/2;17,14}}}, \frac{S_X^2}{S_Y^2} \underbrace{F_{\alpha/2;n_Y-1,n_X-1}}_{F_{0.1/2;17,14}}\right]$$

$$= \left[\frac{0.30}{0.34} \frac{1}{2.329}, \frac{0.30}{0.34} 2.329\right] \approx [0.379, 2.055]$$

$$\frac{\sigma_X^2}{\sigma_Y^2} \in \left[\frac{S_X^2}{S_Y^2} \frac{1}{\underbrace{F_{\alpha/2;n_Y-1,n_X-1}}_{F_{0.05/2;17,14}}}, \frac{S_X^2}{S_Y^2} \underbrace{F_{\alpha/2;n_Y-1,n_X-1}}_{F_{0.05/2;17,14}}\right]$$

$$= \left[\frac{0.30}{0.34} \frac{1}{2.753}, \frac{0.30}{0.34} 2.753\right] \approx [0.321, 2.429]$$

$X \sim \mathrm{N}(\mu_X, \sigma_X^2), \; Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$

$\Rightarrow \frac{\sigma_X^2}{\sigma_Y^2} \in$

$\left[\frac{S_X^2}{S_Y^2} F_{1-\alpha/2;n_Y-1,n_X-1}, \frac{S_X^2}{S_Y^2} F_{\alpha/2;n_Y-1,n_X-1}\right]$

$$F_{1-\alpha;u,v} = \frac{1}{F_{\alpha;u,v}}$$

$X \sim \mathrm{N}(\mu_X, \sigma^2), \; Y \sim \mathrm{N}(\mu_Y, \sigma^2),$

$\Rightarrow \mu_X - \mu_Y \in$

$$\bar{X} - \bar{Y} \pm t_{\alpha/2;n_X+n_Y-2} S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$

$$S_p = \sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}}$$

Ex. 4 - Réponse - b)

Les variances sont-elles significativement différentes?

"Puisque les deux intervalles contiennent la valeur 1, on peut conclure que les variances ne diffèrent pas significativement."

$$\bar{x}_1 = 8.73, \quad s_1^2 = 0.30, \quad n_1 = 15$$
$$\bar{x}_2 = 8.68, \quad s_2^2 = 0.34, \quad n_2 = 18$$

$$X \sim \mathrm{N}(\mu_X, \sigma_X^2),\ Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$$
$$\Rightarrow \frac{\sigma_X^2}{\sigma_Y^2} \in$$
$$\left[\frac{S_X^2}{S_Y^2} F_{1-\alpha/2; n_Y-1, n_X-1}, \frac{S_X^2}{S_Y^2} F_{\alpha/2; n_Y-1, n_X-1}\right]$$
$$F_{1-\alpha; u,v} = \frac{1}{F_{\alpha; u,v}}$$

$$X \sim \mathrm{N}(\mu_X, \sigma^2),\ Y \sim \mathrm{N}(\mu_Y, \sigma^2),$$
$$\Rightarrow \mu_X - \mu_Y \in$$
$$\bar{X} - \bar{Y} \pm t_{\alpha/2; n_X+n_Y-2} S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$
$$S_p = \sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}}$$

Ex. 4 - Réponse - c)

$\bar{x_1} = 8.73, \quad s_1^2 = 0.30, \quad n_1 = 15$

$\bar{x_2} = 8.68, \quad s_2^2 = 0.34, \quad n_2 = 18$

Déterminer l'I.C. 95% de $\mu_1 - \mu_2$ si $\sigma_1^2 = \sigma_2^2$.

$$\mu_1 - \mu_2 \in \bar{x_1} - \bar{x_2} \pm \underbrace{t_{\alpha/2;n_1+n_2-2}}_{t_{0.05/2;31}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 8.73 - 8.68 \pm 2.0423 \sqrt{\frac{14 \cdot 0.30 + 17 \cdot 0.34}{31}} \sqrt{\frac{1}{15} + \frac{1}{18}}$$

$$\approx 0.05 \pm 0.405 = [-0.355, 0.455]$$

$X \sim \mathrm{N}(\mu_X, \sigma_X^2),\ Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$

$\Rightarrow \frac{\sigma_X^2}{\sigma_Y^2} \in$

$\left[\frac{S_X^2}{S_Y^2} F_{1-\alpha/2;n_Y-1,n_X-1}, \frac{S_X^2}{S_Y^2} F_{\alpha/2;n_Y-1,n} \right.$

$F_{1-\alpha;u,v} = \frac{1}{F_{\alpha;u,v}}$

$X \sim \mathrm{N}(\mu_X, \sigma^2),\ Y \sim \mathrm{N}(\mu_Y, \sigma^2),$

$\Rightarrow \mu_X - \mu_Y \in$

$\bar{X} - \bar{Y} \pm t_{\alpha/2;n_X+n_Y-2} S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$

$S_p = \sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}}$

Ex. 5 - Énoncé

Soit 400 automobilistes choisis au hasard, dont 48 n'assurent pas leur véhicule contre les dommages.

a) Établissez un intervalle de confiance à 95 % pour la proportion d'automobilistes qui n'assurent pas leur véhicule contre les dommages.

b) Si 100 automobilistes avaient affirmé ne pas assurer leur véhicule plutôt que 48, déterminez (sans faire le calcul complet) si la marge d'erreur aurait été supérieure ou inférieure à celle calculée en a).

c) En supposant que la proportion d'assurés est complètement inconnue avant le sondage, quelle devrait être la taille de l'échantillon pour qu'on puisse affirmer avec 95 % de confiance que l'erreur d'estimation de la proportion en cause est inférieure à 0,03 ?

- Si $X \sim \text{Bernouilli}(p)$,

  alors $p \in \hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

  où $\hat{p} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

Ex. 5 - Réponse - a), b)

Déterminer l'I.C. 95% de p.

48 non assurés / 400 automobilistes

$$p \in \hat{p} \pm \underbrace{z_{\alpha/2}}_{z_{0.05/2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{48}{400} \pm 1.96\sqrt{\frac{\frac{48}{400}\left(1-\frac{48}{400}\right)}{400}}$$

$$\approx 0.12 \pm 0.0318 = [0.088, 0.152]$$

$X \sim \text{Bernouilli}(p)$

$\Rightarrow p \in \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

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

La marge d'erreur est-elle supérieure si plutôt $100/400$?

$$\frac{48}{400}\left(1-\frac{48}{400}\right) = 0.1056 < 0.1875 = \frac{100}{400}\left(1-\frac{100}{400}\right)$$

$\Rightarrow$ *Oui.*

Ex. 5 - Réponse - c)

Quelle taille n pour erreur < 0.03 de l'I.C. 95% de p
(avant de récolter les données)?

48 non assurés / 400 automobilistes

$$z_{0.05/2}\sqrt{\frac{0.5(1-0.5)}{n}} \leq E$$

$$n \geq \left(z_{0.05/2}\frac{0.5}{E}\right) = \left(1.96\frac{0.5}{0.03}\right) \approx 1067.11$$

$$\Rightarrow n \geq 1068$$

$X \sim \text{Bernouilli}(p)$

$$\Rightarrow p \in \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

Ex. 6 - Énoncé

Exercice n° 6 Un composant d'un certain type fonctionne pendant un temps (durée de fonctionnement) de moyenne 1000 heures et d'écart type 200 heures avant d'être remplacé par un composant neuf du même type.

a) Quelle est la probabilité qu'au plus 62 composants soient utilisés en 60 000 heures ? Justifier votre démarche.

b) On envisage d'utiliser un nouveau modèle de composant à la place du modèle actuel. Des tests en laboratoire ont permis d'obtenir les données suivantes sur les durées de fonctionnement (en heures) d'un échantillon de composants de ce nouveau modèle (ces données se trouvent à la colonne 1 du fichier `Td9.csv`, disponible sur le site du cours).

1196,51	1125,01	836,11	917,81	1472,01	1281,54	755,02
1007,18	811,32	1045,84	957,06	871,35	1009,72	1103,36
1175,78	1260,91	1037,83	1156,91	1281,90	1101,20	906,61
928,87	1042,12	632,10	1121,96	1165,09	1141,98	1029,10
1308,58	1125,14	986,33	1223,59	1167,92	910,87	1062,98

Remarque : on a $\overline{x} = 1061{,}646$; $s = 172{,}677$.

On suppose que ces données proviennent d'une population de distribution normale (i.e. gaussienne).

1.b) Calculer un intervalle de confiance pour la moyenne de la durée de fonctionnement du nouveau modèle de composant au niveau de confiance de 95%. Interpréter brièvement le résultat. [Vous pouvez vérifier le résultat à l'aide d'un logiciel tel que `R`]

2.b) Calculer un intervalle de confiance pour la variance de la durée de fonctionnement du nouveau modèle de composant au niveau de confiance de 95%. Interpréter brièvement le résultat.

3.b) Calculer un intervalle de prévision pour la prochaine observation de la durée de fonctionnement du nouveau modèle de composant au niveau de confiance de 90%. Interpréter brièvement le résultat.

- Si X_i i.i.d., $i = 1, 2, \ldots, n$ et n grand, alors $\sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu_X, n\sigma^2)$
- Si $X \sim \mathrm{N}(\mu, \sigma^2)$, alors $\mu \in \bar{X} \pm t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}$
- Si $X \sim \mathrm{N}(\mu, \sigma^2)$, alors $\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}} \right]$
- Si $X \sim \mathrm{N}(\mu, \sigma^2)$, alors $X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1} S \sqrt{1 + \frac{1}{n}}$

Ex. 6 - Réponse - a)

Déterminer P(au plus 62 utilisés en 60 000 heures).

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}(nE(X), nVar(X)) = \mathrm{N}(62 \cdot 10000, 62 \cdot 200^2)$$

P(au plus 62 utilisés en 60 000 heures)

$= 1 - P$(au moins 63 utilisés en 60 000 heures)

$= 1 - P$(62 utilisés ont duré $<$ 60 000 heures)

$$= 1 - P\left(\sum_{i=1}^{62} X_i < 60000\right) = 1 - \Phi\left(\frac{60000 - 62 \cdot 10000}{\sqrt{62 \cdot 200^2}}\right)$$

$= 1 - \Phi(-1.27) = \Phi(1.27) \approx 0.89796$

vieux: $E(X) = 1000$, $Var(X) = 200^2$

nouveau: $X \sim \mathrm{N}(\mu, \sigma^2)$, $\quad n = 35$

$\bar{x} = 1061.646$, $\quad s = 172.677$

X_i i.i.d., n grand

$\Rightarrow \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu_X, n\sigma^2)$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow \mu \in \bar{X} \pm t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

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

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1} S \sqrt{1 + \frac{1}{n}}$

Ex. 6 - Réponse - 1.b)

Déterminer l'I.C. 95% de μ_{nouveau}.

$$\mu \in \bar{x} \pm \underbrace{t_{\alpha/2;n-1}}_{t_{0.05/2;34}} \frac{s}{\sqrt{n}} = 1061.66 \pm 2.0317 \frac{172.677}{\sqrt{35}}$$

$$= 1061.66 \pm 59.300 = [1002.36, 1120.96]$$

Ce nouveau modèle est meilleur (avec 95 % de chance) car la moyenne de l'ancien est plus basse que l'I.C. du nouveau.

vieux: $E(X) = 1000$, $Var(X) = 200^2$

nouveau: $X \sim \mathrm{N}(\mu, \sigma^2)$, $\quad n = 35$

$\bar{x} = 1061.646$, $\quad s = 172.677$

X_i i.i.d., n grand

$\Rightarrow \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu_X, n\sigma^2)$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow \mu \in \bar{X} \pm t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

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

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1} S \sqrt{1 + \frac{1}{n}}$

Ex. 6 - Réponse - 2.b)

Déterminer l'I.C. 95% de $\sigma^2_{\text{nouveau}}$.

$$\sigma^2 \in \left[\underbrace{\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}}_{\chi^2_{0.05/2;34}}, \underbrace{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}}_{\chi^2_{1-0.05/2;34}}\right] = \left[\frac{34 \cdot 172.677^2}{59.13}, \frac{34 \cdot 172.677^2}{20.61}\right]$$

$$\approx [17145.10, 49189.22]$$

Ce nouveau modèle est similaire à l'ancien (40 000) puisqu'il est compris dans l'I.C. du nouveau.

vieux: $E(X) = 1000$, $Var(X) = 200^2$

nouveau: $X \sim \mathrm{N}(\mu, \sigma^2)$, $n = 35$

$\bar{x} = 1061.646$, $s = 172.677$

X_i i.i.d., n grand

$$\Rightarrow \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu_X, n\sigma^2)$$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$$\Rightarrow \mu \in \bar{X} \pm t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}$$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

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

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$$\Rightarrow X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1} S\sqrt{1 + \frac{1}{n}}$$

Ex. 6 - Réponse - 3.b)

Déterminer l'intervalle de prévision de 90% de la prochaine observation pour le nouveau modèle.

$$X_{n+1} \in \bar{X} \pm \underbrace{t_{\alpha/2;n-1}}_{t_{0.1/2;34}} S\sqrt{1+\frac{1}{n}}$$

$$= 1061.646 \pm 1.6906 \cdot 172.677\sqrt{1+\frac{1}{35}}$$

$$\approx 1061.66 \pm 296.069 = [765.59, 1357.73]$$

Ça a du sens : sur les 35 analysés, seulement 2 étaient en dessous de cet intervalle et seulement 1 au dessus ($32/35 \approx 0.91$).

vieux: $E(X) = 1000$, $Var(X) = 200^2$

nouveau: $X \sim \mathrm{N}(\mu, \sigma^2)$, $\quad n = 35$

$\bar{x} = 1061.646$, $\quad s = 172.677$

X_i i.i.d., n grand

$\Rightarrow \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu_X, n\sigma^2)$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow \mu \in \bar{X} \pm t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}$

$X \sim \mathrm{N}(\mu, \sigma^2)$,

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

$X \sim \mathrm{N}(\mu, \sigma^2)$,

$\Rightarrow X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1} S\sqrt{1+\frac{1}{n}}$

Ex. 7 - Énoncé

Exercice nº 7 Des ingénieurs travaillent au développement d'un procédé de fabrication. Afin d'en améliorer la fiabilité, ils s'intéressent à la proportion p d'unités non conformes produites par le procédé. Au stade actuel du développement du procédé, les ingénieurs ont établi que la proportion p ne dépasse pas 20%. Dénotons par X_n le nombre d'unités non conformes trouvées dans un échantillon de taille n de la production du procédé.

a) Proposez un estimateur ponctuel pour p en vous basant sur un échantillon de taille n. Calculez son erreur quadratique moyenne (EQM).

b) Quelle taille d'échantillon n devrait-on utiliser afin de s'assurer que l'erreur d'estimation commise soit inférieure à 4% pour 95% des échantillons possibles ?

c) Indépendamment de la réponse en b), il fut décidé en définitive qu'un échantillon de taille 250 sera choisi. Supposons que 41 des 250 unités d'un échantillon convenablement recueilli ont été trouvées défectueuses. Donnez un intervalle de confiance pour p au niveau de confiance 95% et interprétez le résultat.

- $EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$
- Si $X \sim \mathrm{B}(n, p)$, alors $E(X) = np, \quad Var(X) = np(1-p)$
- Si $X \sim$ Bernouilli(p), alors $p \in \hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ où $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}X_n$ où $X_n \sim \mathrm{B}(n,p)$

Ex. 7 - Réponse - a)

Proposer un $\hat{p}$ et calculer son EQM.

$$\hat{p} = \tfrac{1}{n} X_n \text{ où } X_n \sim \mathrm{B}(n,p)$$

$$E(\hat{p}) = E\left(\frac{1}{n}X_n\right) = \frac{1}{n}E(X_n) = \frac{1}{n}np = p$$

$$Var(\hat{p}) = Var\left(\frac{1}{n}X_n\right) = \frac{1}{n^2}Var(X_n) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

$$EQM(\hat{p}) = Var(\hat{p}) + (E(\hat{p}) - p)^2 = \frac{p(1-p)}{n}$$

proportion $p < 0.2$ des non conformes

X_n : somme des non conformes parmi n

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

$X \sim \mathrm{B}(n,p)$
$\Rightarrow E(X) = np$,
$Var(X) = np(1-p)$

$X \sim$ Bernouilli(p)
$$\Rightarrow p \in \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

Ex. 7 - Réponse - b)

Quelle taille n pour erreur $\leq$0.04 pour I.C. 95% ?

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{\alpha/2}\sqrt{\frac{0.2(1-0.2)}{n}} \leq E$$

$$n \geq (\underbrace{z_{\alpha/2}}_{z_{0.05/2}})^2 \frac{0.2(1-0.2)}{E^2}$$

$$n \geq 1.96^2 \cdot \frac{0.2 \cdot 0.8}{0.04^2} = 384.16$$

$$n \geq 385$$

proportion $p < 0.2$ des non conformes

X_n : somme des non conformes parmi n

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

$X \sim \mathrm{B}(n, p)$
$\Rightarrow E(X) = np$,
$Var(X) = np(1-p)$

$X \sim$ Bernouilli(p)

$$\Rightarrow p \in \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

Ex. 7 - Réponse - c)

Déterminer l'I.C. 95% de p avec 41/250 défectueux?

$$p \in \hat{p} \pm \underbrace{z_{\alpha/2}}_{z_{0.05/2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{41}{250} \pm 1.96\sqrt{\frac{\frac{41}{250}\left(1-\frac{41}{250}\right)}{250}}$$

$$\approx 0.164 \pm 0.0459 = [0.118, 0.210]$$

Cohérent avec le max établi par les ingénieurs car l'I.C. contient 0.2.

Alors, avec 95% de chance, on a $p \in [0.118, 0.2]$.

proportion $p < 0.2$ des non conformes

X_n : somme des non conformes parmi n

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

$X \sim \mathrm{B}(n, p)$
$\Rightarrow E(X) = np$,
$Var(X) = np(1-p)$

$X \sim \mathrm{Bernouilli}(p)$

$$\Rightarrow p \in \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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