STT-20694: Probabilités pour ingénieurs

SOLUTION DE L'EXAMEN NUMÉRO 2

Lundi le 16 décembre 2002

NUMÉRO 1. On suppose que la loi normale avec moyenne 100 et avec écart-type 16 est un bon modèle pour la distribution des quotients intellectuels dans la population adulte au Canada.

(a) [2 points] Quel est le 67^e centile de cette distribution?

SOLUTION:

$$\begin{aligned}
&\mathbb{P}[X \leq q_{0.67}] = 0.67 \\
&\mathbb{P}\left[\frac{X-100}{16} \leq \frac{q_{0.67}-100}{16}\right] = 0.67 \\
&\Phi\left(\frac{q_{0.67}-100}{16}\right) = 0.67 \\
&\frac{q_{0.67}-100}{16} = 0.44 \\
&q_{0.67} = 100 + (16 \times 0.44) \\
&q_{0.67} = 107.04
\end{aligned}$$

(b) [2 points] Dans cette population, quelle est la proportion d'individus ayant un quotient intellectuel entre 92 et 112?

SOLUTION:

$$\begin{aligned}
\mathbb{P}[92 \leq X \leq 112] &= \mathbb{P}\left[\frac{92-100}{16} \leq \frac{X-100}{16} \leq \frac{112-100}{16}\right] \\
&= \Phi\left(\frac{112-100}{16}\right) - \Phi\left(\frac{92-100}{16}\right) \\
&= \Phi(0.75) - \Phi(-0.50) \\
&= 0.7734 - 0.3085 \\
&= 0.4649
\end{aligned}$$

(c) [2 points] On choisit 25 individus au hasard dans cette population. Quelle est la probabilité qu'il y ait, parmi ces 25 individus, au moins un individu avec un quotient intellectuel supérieur à 132?

SOLUTION:

$$
\begin{aligned}
\mathbb{P}[N \geq 1] &= 1 - \mathbb{P}[N = 0] \\
&= 1 - \left(\mathbb{P}[X \leq 132]\right)^{25} \\
&= 1 - \left(\Phi\left(\frac{132 - 100}{16}\right)\right)^{25} \\
&= 1 - \left(\Phi(2)\right)^{25} \\
&= 1 - (0.9772)^{25} \\
&= 1 - 0.5618 \\
&= 0.4382
\end{aligned}
$$

(d) [2 points] On choisit 25 individus au hasard dans cette population. Quelle est la probabilité que la moyenne des quotients intellectuels de ces 25 individus soit supérieure à 108?

SOLUTION:

$$
\begin{aligned}
\mathbb{P}[\overline{X} > 108] &= 1 - \mathbb{P}[\overline{X} \leq 108] \\
&= 1 - \mathbb{P}\left[\frac{\overline{X} - 100}{16/\sqrt{25}} \leq \frac{108 - 100}{16/\sqrt{25}}\right] \\
&= 1 - \Phi\left(\frac{108 - 100}{16/\sqrt{25}}\right) \\
&= 1 - \Phi(2.50) \\
&= 1 - 0.9938 \\
&= 0.0062
\end{aligned}
$$

NUMÉRO 2. On suppose que le processus de Poisson avec intensité $\lambda = 6$ par heure est un bon modèle pour décrire les passages d'automobiles du Canada vers les États-Unis au poste frontalier américain situé sur la route qui relie Whitehorse (Yukon) et Fairbanks (Alaska). Autrement dit, on suppose que les temps entre les passages successifs d'automobiles allant du Canada vers les États-Unis à ce poste frontalier sont des variables aléatoires exponentielle(6) indépendantes les unes des autres.

(a) [2 points] Quelle est la probabilité que durant la prochaine heure exactement 4 automobiles franchiront ce poste frontalier en direction des États-Unis?

SOLUTION: Si X dénote le nombre d'automobiles qui franchiront ce poste frontalier en direction des États-Unis durant la prochaine heure, alors on a

$$X \sim \text{Poisson}(6).$$

On obtient donc

$$\mathbb{P}[X = 4] = \mathrm{e}^{-6}\,\frac{6^4}{4!} = 0.1339$$

(b) [2 points] Les douaniers américains inspectent une automobile sur dix. Ils viennent tout juste d'inpecter une automobile. Ils vont donc laisser passer les 9 prochaines automobiles sans les inspecter, puis ils vont inspecter la suivante. Et ainsi de suite. Calculez l'espérance et l'écart-type du temps qui s'écoule entre 2 inspections successives.

SOLUTION: Le temps total V qui s'écoule entre deux inspections successives est la somme de 10 variables aléatoires exponentielle(6) indépendantes les unes des autres:

$$V = T_1 + T_2 + \cdots + T_{10}.$$

On obtient donc

$$\begin{aligned}
\mu = \mathbb{E}[V] &= \mathbb{E}[T_1] + \cdots + \mathbb{E}[T_{10}] = \frac{1}{6} + \cdots + \frac{1}{6} = \frac{10}{6} = 1.6667 \\
\sigma^2 = \mathbb{V}\mathrm{ar}[V] &= \mathbb{V}\mathrm{ar}[T_1] + \cdots + \mathbb{V}\mathrm{ar}[T_{10}] = \frac{1}{36} + \cdots + \frac{1}{36} = \frac{10}{36} = 0.2778 \\
\sigma &= \sqrt{0.2778} = 0.5270
\end{aligned}$$

Autre approche: La variable V ci-haut est une gamma(10,6). On obtient donc directement

$$\mu = \frac{10}{6} = 1.6667 \quad \text{et} \quad \sigma = \sqrt{\frac{10}{36}} = 0.5270.$$

(c) [2 points] Quel est l'espérance et l'écart-type du nombre d'automobiles qui franchiront ce poste frontalier (en direction des États-Unis) durant les prochaines 24 heures?

SOLUTION: Si Y dénote le nombre d'automobiles qui franchiront ce poste frontalier en direction des États-Unis durant les prochaines 24 heures, alors

$$Y \sim \text{ Poisson}(24 \times 6) = \text{ Poisson}(144).$$

On obtient donc

$$\mu = \mathbb{E}[Y] = 144 \quad \text{ et } \quad \sigma = \sqrt{\mathbb{V}\text{ar}[Y]} = \sqrt{144} = 12.$$

(d) [2 points] Calculez une approximation pour la probabilité qu'il y ait au moins 150 automobiles qui franchissent ce poste frontalier (en direction des États-Unis) durant les prochaines 24 heures.

SOLUTION: Puisque

$$Y \sim \text{ Poisson}(144),$$

on a

$$Y \approx N(144, 144).$$

Donc, avec correction pour la continuité, on obtient

$$\begin{aligned} \mathbb{P}[Y \geq 150] &= 1 - \mathbb{P}[Y \leq 149] \\ &\approx 1 - \Phi\left(\frac{149.5 - 144}{\sqrt{144}}\right) \\ &= 1 - \Phi(0.4583) \\ &\approx 1 - \Phi(0.46) \\ &= 1 - 0.6772 \\ &= 0.3228 \end{aligned}$$

Numéro 3. La densité de probabilité suivante est un bon modèle pour décrire la distribution du temps X (exprimé en minutes) nécessaire pour répondre à un client au service à la clientèle d'un grand magasin à rayon:

$$f_\theta(x) = \begin{cases} \dfrac{2\ x^3}{\theta^4}\ \mathrm{e}^{-x^2/\theta^2} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

La moyenne et la variance de cette distribution sont

$$\mu = \frac{3\sqrt{\pi}}{4}\,\theta \qquad \text{et} \qquad \sigma^2 = \left(2 - \frac{9\pi}{16}\right)\theta^2$$

On dispose d'un échantillon aléatoire de taille n à partir duquel on désire estimer le paramètre θ.

(a) [4 points] Obtenez l'estimateur du maximum de la vraisemblance pour θ.

Solution.

$$L(\theta) = \frac{2^n\ \left(\prod_{i=1}^n x_i^3\right)}{\theta^{4n}}\ \mathrm{e}^{-\sum_{i=1}^n x_i^2/\theta^2}$$

$$\mathcal{L}(\theta) = n\ln(2) + 3\sum_{i=1}^n \ln(x_i) - 4n\ln(\theta) - \frac{\sum_{i=1}^n x_i^2}{\theta^2}$$

$$\frac{d}{d\theta}\mathcal{L}(\theta) = -\frac{4n}{\theta} + \frac{2\sum_{i=1}^n x_i^2}{\theta^3}$$

On résoud l'équation

$$\frac{d}{d\theta}\mathcal{L}(\theta) = 0$$

et on obtient notre estimateur:

$$\hat{\theta}_{\text{MAX}} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{2n}}.$$

(b) [4 points] Avec le modèle donné à la page précédente, l'équation *"moyenne théorique = moyenne échantillonnale"* prend la forme $3\sqrt{\pi}\theta/4 = \overline{X}$. La méthode des moments nous donne donc l'estimateur

$$\hat{\theta}_{\text{MOM}} = \frac{4}{3\sqrt{\pi}}\,\overline{X}.$$

On obtient un échantillon de 20 temps de réponse. Voici ces 20 temps:

2.50	2.85	3.70	2.45	0.65
1.50	1.55	2.30	3.75	3.00
1.85	4.15	3.80	3.55	2.95
2.60	3.60	3.95	4.50	2.25

La moyenne échantillonnale est 2.873 minutes et l'écart-type échantillonnal est 1.013 minutes. La méthode des moments nous donne donc l'estimation

$$\hat{\theta}_{\text{MOM}} = \frac{4}{3\sqrt{\pi}}\,\overline{x} = \frac{4}{3\sqrt{\pi}}\,2.873 = 2.161 \text{ (minutes)}.$$

Quelle est l'erreur type associée à cette estimation?

Solution. L'estimateur $\hat{\theta}_{\text{MOM}}$ est sans biais. Son erreur quadratique moyenne est donc égal à sa variance. Son erreur type est donc égal à son écart-type. Donc

$$\begin{aligned}
\text{erreur type} &= \sqrt{\mathbb{V}\text{ar}[\hat{\theta}_{\text{MOM}}]} = \sqrt{\mathbb{V}\text{ar}\left[\frac{4}{3\sqrt{\pi}}\,\overline{X}\right]} \\
&= \sqrt{\left(\frac{4}{3\sqrt{\pi}}\right)^2 \mathbb{V}\text{ar}\left[\overline{X}\right]} = \sqrt{\left(\frac{4}{3\sqrt{\pi}}\right)^2 \frac{\sigma^2}{n}} \\
&= \frac{4}{3\sqrt{\pi}}\sqrt{\frac{\left(2-\frac{9\pi}{16}\right)\theta^2}{20}} = \frac{4}{3\sqrt{\pi}}\sqrt{\frac{\left(2-\frac{9\pi}{16}\right)}{20}}\;\theta \\
&\approx \frac{4}{3\sqrt{\pi}}\sqrt{\frac{\left(2-\frac{9\pi}{16}\right)}{20}}\;\hat{\theta}_{\text{MOM}} \\
&= \frac{4}{3\sqrt{\pi}}\sqrt{\frac{\left(2-\frac{9\pi}{16}\right)}{20}}\;(2.161) \\
&= 0.1754
\end{aligned}$$

NUMÉRO 4 (a). Un biologiste a mesuré les longueurs de 310 perchaudes attrapées dans le Chenal du Moine en août 2002. Le tableau suivant résume ces 310 mesures:

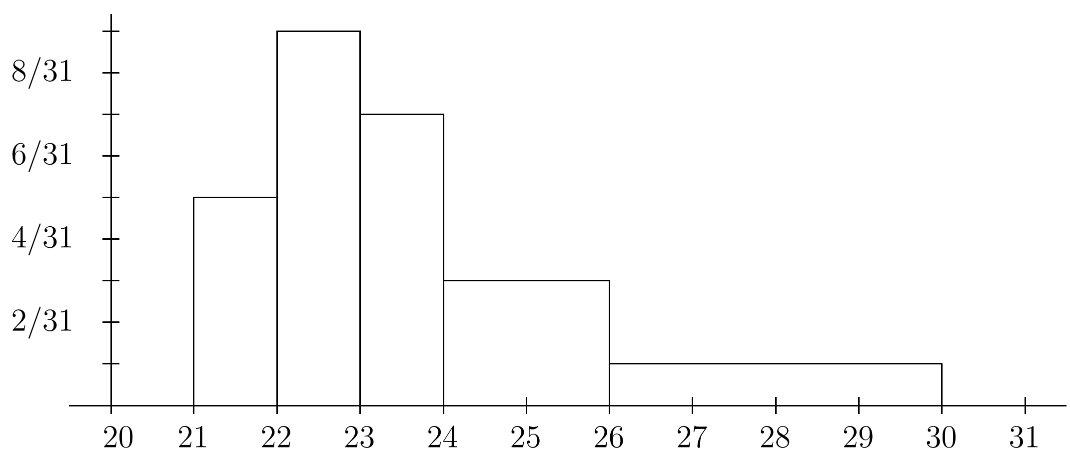
Longueur	Nombre de perchaudes
entre 21.0 et 22.0 cm	50
entre 22.0 et 23.0 cm	90
entre 23.0 et 24.0 cm	70
entre 24.0 et 26.0 cm	60
entre 26.0 et 30.0 cm	40

(i) [2 points] Calculez une approximation pour le 10^e centile de la distribution des longueurs des perchaudes.

RÉPONSE: 21.62.

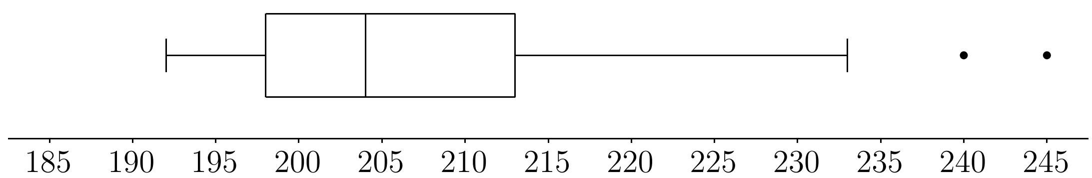
(ii) [2 points] À partir de l'axe horizontal gradué reproduit ci-bas, tracez l'histogramme de ces 310 longueurs.

SOLUTION:



Numéro 4 (b). Chaque bonne réponse vaut 1 point. Aucune justification nécessaire.

Le diagramme en boîte suivant représente la distribution des poids (exprimés en livres) des 92 joueurs qui se sont présentés au camps d'entrainement de l'équipe de football du Rouge et Or en juin dernier.



(i) La somme des poids des 3 joueurs les plus pesants est

■ inférieure à 720 livres.

□ égale à 720 livres.

□ supérieure à 720 livres.

(ii) La moyenne de ces 92 poids est 204 livres.

□ Vrai.

■ Faux.

□ Peut-être que oui, peut-être que non.

(iii) Quel est l'écart inter-quartile de ces 92 poids?

□ 5

□ 10

■ 15

□ 20

(iv) La loi normale est-elle un bon modèle pour décrire cette distribution de poids?

□ Oui, bien sûr.

□ Oui, mais il faut d'abord soustraire la moyenne et diviser par l'écart-type.

■ Non, pas du tout.

□ Il faudrait voir l'histogramme pour répondre à cette question.

NUMÉRO 5. (a) [2 points]
On calcule un intervalle de confiance de niveau 95% pour une différence de moyenne théorique $\mu_1 - \mu_2$. On utilise la formule

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2, n_1+n_2-2}\ S_p\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Si nos tailles d'échantillons sont $n_1 = 8$ et $n_2 = 11$ et si on obtient une variance échantillonnale combinée égal à 6.77, quel sera la longueur de notre intervalle de confiance?

SOLUTION:

$$\begin{aligned} L &= 2\ t_{\alpha/2, n_1+n_2-2}\ S_p\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ &= 2\ t_{0.025,17}\ \sqrt{6.77}\ \sqrt{\frac{1}{8} + \frac{1}{11}} \\ &= 2\ (2.110)\ \sqrt{6.77}\ \sqrt{\frac{1}{8} + \frac{1}{11}} \\ &= 5.102 \end{aligned}$$

NUMÉRO 5. (b) [3 points]
On veut comparer deux distributions normales. Nos tailles d'échantillons sont $n_1 = 20$ et $n_2 = 15$. Notre rapport de variances échantillonnales est 6.57. Plus précisément,

$$\frac{\textit{variance de l'échantillon provenant de la population 1}}{\textit{variance de l'échantillon provenant de la population 2}} = 6.57.$$

Calculez in intervalle de confiance de niveau 90% pour le rapport des variances théoriques, c'est-à-dire le rapport

$$\frac{\textit{variance de la distribution 1}}{\textit{variance de la distribution 2}}.$$

SOLUTION:

$$\begin{aligned}
\left[\frac{1}{F_{\alpha/2,n_1-1,n_2-1}}\,\frac{S_1^2}{S_2^2}\,,\;\frac{1}{F_{1-\alpha/2,n_1-1,n_2-1}}\,\frac{S_1^2}{S_2^2}\right] &= \left[\frac{1}{F_{0.05,19,14}}\;6.57\,,\;\frac{1}{F_{0.95,19,14}}\;6.57\right] \\
&= \left[\frac{1}{F_{0.05,19,14}}\;6.57\,,\;F_{0.05,14,19}\;6.57\right] \\
&= \left[\frac{1}{2.40}\;6.57\,,\;(2.256)\;6.57\right] \\
&= [2.74\,,\;14.82]
\end{aligned}$$

NUMÉRO 5. (c) [3 points]
Lorsqu'on estime une différence de proportions $p_A - p_B$ par une différence de proportions échantillonnales $\hat{p}_A - \hat{p}_B$ calculées à partir d'échantillons indépendants de même taille $n = n_A = n_B$, quelle doit être cette taille d'échantillon n si on veut être 95% certain que l'erreur $|(\hat{p}_A - \hat{p}_B) - (p_A - p_B)|$ soit au plus 0.02?

SOLUTION: On veut que la demi-longueur de l'intervalle de confiance de niveau 0.95 pour $p_1 - p_2$ soit au plus 0.02. Or, on sait que cette demi-longueur $L/2$ satisfait

$$\frac{L}{2} = z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq z_{\alpha/2}\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}.$$

Puisque $n_1 = n_2 = n$ et $\alpha = 0.05$, on a donc

$$\frac{L}{2} \leq z_{0.025}\sqrt{\frac{1}{4n} + \frac{1}{4n}} = \frac{1.96}{\sqrt{2n}}.$$

Il suffit donc de prendre n assez grand pour que

$$\frac{1.96}{\sqrt{2n}} = 0.02.$$

On résoud et on obtient $n = 4\,802$.

Numéro 6.

Questions à choix multiples. Chaque bonne réponse vaut 2 points. Aucune justification nécessaire.

(a) Dans le cas discret, la fonction de log-vraisemblance $\mathcal{L}(\theta)$

- ☐ est toujours positive (c'est-à-dire $\mathcal{L}(\theta) \geq 0$ pour tout θ).
- ■ est toujours négative (c'est-à-dire $\mathcal{L}(\theta) \leq 0$ pour tout θ).
- ☐ peut être positive pour certaines valeurs de θ et négative pour d'autres valeurs de θ.

(b) La densité de probabilité suivante est un cas particulier de la loi de Weibull:

$$f(x) = \begin{cases} 3x^2 \mathrm{e}^{-x^3} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

Sa fonction de répartition est donnée par

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-x^3} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

Si on veut générer une variable aléatoire X ayant cette distribution de Weibull, il suffit de générer (à l'aide d'un générateur de nombres aléatoires) une variable aléatoire U avec distribution uniforme sur l'intervalle $(0, 1)$ et de prendre

- ☐ $X = 1 - \mathrm{e}^{-U^3}$.
- ☐ $X = 3U^2\mathrm{e}^{-U^3}$.
- ☐ $X = (\ln(1-U))^{1/3}$.
- ☐ $X = (\ln(1-U))^{-1/3}$.
- ■ $X = \left(\ln\left(\frac{1}{1-U}\right)\right)^{1/3}$.

(c) En général, lorsqu'on utilise une moyenne échantillonnale pour estimer une moyenne théorique, si on veut doubler la précision de l'estimation (c'est-à-dire réduire de moitié l'erreur type), il faut

- ☐ réduire de moitié la taille de l'échantillon.
- ☐ doubler la taille de l'échantillon.
- ☐ tripler la taille de l'échantillon.
- ■ quadrupler la taille de l'échantillon.

(d) Le biologiste du numéro 4 (a) a calculé la moyenne des longueurs de ses 310 perchaudes. Cette moyenne est égale à 23.60 cm. Il a également calculé l'écart-type de ces 310 longueurs. Cet écart-type est égal à 1.90 cm. Notre biologiste estime donc à 23.60 cm la longueur moyenne des perchaudes que les pêcheurs attrapaient dans le Chenal du Moine en août 2002. Quelle est l'erreur type associée à cette estimation?

- ☐ 1.90 cm
- ☐ 0.006 cm
- ■ 0.108 cm
- ☐ 0.212 cm

[On suppose que les 310 perchaudes que notre biologiste a attrapées constituent un échantillon aléatoire issu de la population des perchaudes que les pêcheurs attrapaient durant le mois d'août 2002 dans le Chenal du Moine.

(e) Notre biologiste décide de calculer un intervalle de confiance de niveau 99% pour la longueur moyenne des perchaudes qu'on attrapait dans le Chenal du Moine en août 2002. Il utilise la formule

$$\left[23.60 - z_{0.005}\frac{1.90}{\sqrt{n}}\,,\, 23.60 + z_{0.005}\frac{1.90}{\sqrt{n}}\right]$$

Une seule des six affirmations suivantes est juste. Laquelle?

- ■ L'intervalle obtenu par notre biologiste est valide.
- ☐ L'intervalle obtenu par notre biologiste n'est pas valide parce que l'histogramme obtenu au numéro 4 (a) montre très bien que la loi normale n'est pas un bon modèle pour la distribution des longueurs des perchaudes.
- ☐ L'intervalle obtenu par notre biologiste n'est pas valide parce qu'il fallait utiliser $z_{0.01}$ plutôt que $z_{0.005}$.
- ☐ L'intervalle obtenu par notre biologiste n'est pas valide parce qu'il fallait utiliser l'écart-type théorique σ plutôt que l'écart-type échantillonnal s.
- ☐ L'intervalle obtenu par notre biologiste n'est pas valide parce que la taille de notre échantillon n'est pas suffisamment grande.
- ☐ L'intervalle obtenu par notre biologiste n'est pas valide parce que notre biologiste travaille pour Environnement Canada.