L1-S4-SV&SVN 2006/2007

Examen du 15 mai 2007

CORRIGÉ
de l'Examen de Mathématiques (Probabilités et Statistiques)

Exercice 1 :

1. On tient compte du fait que la première classe est 3 fois plus longue que les autres. La valeur lui correspondant dans le tableau étant 12 on la divisera par 3 et on placera dans l'histogramme l'hauteur du premier rectangle à 4=12/3. Les autres hauteurs seront celles indiquées dans le tableau sur la ligne $X = 0$. L'histogramme est donc :

2. On "discretise" la variable Z en prenant pour Z chaque fois la valeur de la moitié de chaque classe. Notons par N la population totale (dans notre cas = 500). Les 2 dernières lignes/colonnes du tableau nous aideront à calculer la moyenne, respectivement la variance de X et Y. On a :

$X \backslash Y'$	-4	-2	-1	0	1	2	$n_{i\cdot}$	$n_{i\cdot}x_i$	$n_{i\cdot}x_i^2$
-2	22	8					30	-60	120
-1	36	41	20	13			110	-110	110
0	12	36	68	50	22	22	210	0	0
1			12	26	32	30	100	100	100
2				6	14	20	40	80	160
3,25					2	8	10	32,5	105,625
$n_{\cdot j}$	70	85	100	95	70	80	500	42,5	595,625
$n_{\cdot j}y_j$	-280	-170	-100	0	70	160	-320		
$n_{\cdot j}y_j^2$	1120	340	100	0	70	320	1950		

3.

$$\begin{aligned}
\overline{Y} &= \frac{1}{N}\sum_{j=1}^{6} n_{\cdot j}y_j = \frac{-320}{500} = -0,64 \quad \text{et} \\
\overline{X} &= \frac{1}{N}\sum_{j=1}^{6} n_{i\cdot}x_i = \frac{42,5}{500} = 0,085 \\
V(Y) &= \frac{1}{N}\sum_{j=1}^{6} n_{\cdot j}(y_j-\overline{Y})^2 = \overline{Y^2}-\overline{Y}^2 \\
&= 1950/500-(-0,64)^2 = 3,49 \\
V(X) &= \frac{1}{N}\sum_{j=1}^{6} n_{i\cdot}(x_i-\overline{X})^2 = \overline{X^2}-\overline{X}^2 \\
&= 595,625/500-(0,085)^2 = 1,18 \\
\sigma(Y) &= \sqrt{V(Y)} = 1,868 \\
\sigma(X) &= \sqrt{V(X)} = 1,086.
\end{aligned}$$

4. On a : $\overline{Z} = \overline{5Y+42,5} = 5\overline{Y}+42,5 = 39,3$ et $V(Z) = V(5Y+42,5) = 5^2V(Y) = 87,25$ d'où $\sigma(Z) = 9,34$.

5. Le nuage statistique :

6. $\text{cov}(X,Y) = \overline{X \cdot Y} - \overline{X} \cdot \overline{Y}$. Mais

$$\overline{XY} = \frac{1}{N} \sum_{i,j=1}^{6} n_{ij} x_i y_j.$$

Finalement on obtient $\overline{XY} = 733,5/500$, d'où $\text{cov}(X,Y) = 1,52$.

7. Le coefficient de corrélation $r = \dfrac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)} = 0,70$, donc il est assez mauvais, ce qui est visible aussi sur le nuage statistique.

8. Soient a le coefficient de la droite d'ajustement de y en x et a' le coefficient de la droite d'ajustement de x en y. On a : $a = \dfrac{\text{cov}(X,Y)}{V(X)} = \dfrac{1,52}{3,49} = 0,435$ et $a' = \dfrac{\text{cov}(X,Y)}{V(Y)} = \dfrac{1,52}{1,18} = 1,288$.

9. La droite d'ajustement de y en x par la méthode des moindres carrés a pour équation : $y - \overline{Y} = a(x - \overline{X})$ donc $y = 0,435x - 0,677$.
La droite d'ajustement de x en y par la méthode des moindres carrés a pour équation : $x - \overline{X} = a'(y - \overline{Y})$ donc $x = 1,288y + 0,909$.

Exercice 2 :

1. OUI. $n_e = 83 > 30$.

2. NON. On connaît seulement l'écart-type "experimental", c'est-à-dire celui de l'echantillon, qu'on notera par σ_e.

3. $\sigma = \sigma_e \sqrt{\dfrac{n_e}{n_e - 1}}$.

4. On sait que bien que la moyenne "theorique" n'est pas connue, par convention elle est considérée égale à la moyenne experimentale m_e. Par le théorème du cours qui correspond au cas présent (voir questions 1 & 2) on a $\overline{X} \hookrightarrow \mathcal{N}\Big(m_e, \dfrac{\sigma}{\sqrt{n_e}}\Big) \equiv \mathcal{N}\Big(m_e, \dfrac{\sigma_e}{\sqrt{n_e - 1}}\Big)$.

5. $T \equiv \dfrac{\overline{X} - m_e}{\dfrac{\sigma}{\sqrt{n_e}}} = \dfrac{\overline{X} - m_e}{\dfrac{\sigma_e}{\sqrt{n_e - 1}}} \hookrightarrow \mathcal{N}(0,1)$.

6. On a par hypothèse $m_e = 52,17$ et $\sigma_e = 1,8$. Donc $\dfrac{\overline{X} - 52,17}{1,8/\sqrt{83-1}} \hookrightarrow \mathcal{N}(0,1)$. Par ailleurs, on doit avoir

$$P\Big(\Big\{\Big|\frac{\overline{X} - 52,17}{1,8/\sqrt{83-1}}\Big| \le t\Big\}\Big) = 2\,\pi(t) - 1 = 0,95$$

où π est la fonction de répartition de la loi normale. Des tables de la loi normale résulte $t = 1,96$, donc on doit avoir $|\overline{X} - 52,17| \le 0,389$, c-à-d l'encadrement

$$51,78 \le \overline{X} \le 52,56.$$

7. Le même raisonnement. On trouve dans les tables $t = 3,31$ et l'encadrement sera alors :

$$51,51 \le \overline{X} \le 52,83.$$

8. Remarquons qu'au risque égal (de 5%) on a obtenu à la question 6 un encadrement de $\overline{X}$ dans un intervalle de longueur $52,56 - 51,78 = 0,78$, donc plus large que l'intervalle de largeur 0,4 mm. Il est clair que l'echantillon devra être encore plus grand que celui de 83 individus. On rèste donc dans le même cadre théorique (c-à-d on peut utiliser la loi normale, etc.).
Remarquons aussi que cette-fois on doit se considérer dans le cas où la valeur de 1,8 est donnée pour un écart-type théorique.
Soit n l'effectif inconnu. On doit avoir donc : $P(\{52,17 - 0,2 \le \overline{X} \le 52,17 + 0,2\}) = P\Big(\Big\{\Big|\dfrac{\overline{X} - 52,17}{1,8/\sqrt{n}}\Big| \le \dfrac{0,4/2}{1,8/\sqrt{n}}\Big\}\Big) = 0,95$.
Mais en utilisant les tables et en tenant compte du fait que $t \mapsto P(\{|Z| \le t\})$ est croissante, on obtient la condition $0,111\sqrt{n} \ge 1,96$, donc

$$n \ge [17,65^2] = 312.$$