Université Paul Valéry
Arts - Lettres - Langues
Sciences humaines & sociales

PSYCHOLOGIE - Licence 3

E53XPM

Statistique pour la psychologie 3

Examen du 25 mai 2007

CORRIGÉ

Exercice 1 :

Partie 1 :

1. Notons les événements :

S : "vote Sarkozy"		Pa	: "habite Paris"
R : "vote Royal"		L	: "habite Lyon"
		T	: "habite Toulouse"
		M	: "habite Montpellier"

   Le tableau 1 présente les 4 distributions conditionnelles de vote dans chacune des villes et le tableau 2 présente la distribution du lieu d'habitation.

   $$P(Pa) = 0.664 \qquad P(L) = 0.142 \quad P(T) = 0.123 \quad P(M) = 0.071$$

   $$\begin{array}{llll} P(S|Pa) = P_{Pa}(S) = 0.502 & P_L(S) = 0.53 & P_T(S) = 0.424 & P_M(S) = 0.448 \\ P_{Pa}(R) = 1 - P_{Pa}(S) = 0.498 & P_L(R) = 0.47 & P_T(R) = 0.576 & P_M(R) = 0.552 \end{array}$$

2. On cherche : $P(S)$ avec Pa, L, T, M formant une partition de Ω.

   $$\begin{aligned} P(S) &= P_{Pa}(S) \times P(Pa) + P_L(S) \times P(L) + P_T(S) \times P(T) + P_M(S) \times P(M) \\ &= (0.502 \times 0.664) + (0.53 \times 0.142) + (0.424 \times 0.123) + (0.448 \times 0.071) \\ &= 0.493 \end{aligned}$$

3. On cherche : $P(R)$
   $P(R) = 1 - P(S) = 1 - 0.493 = 0.507$
4. On cherche : $P_{Pa}(S) = 0.502$ (tableau)
5. On cherche : $P_S(Pa)$
   Formule de Bayes :

   $$\begin{aligned} P(Pa|S) &= \frac{P(S|Pa) \times P(Pa)}{P(S)} \\ &= \frac{0.502 \times 0.664}{0.493} = 0.676 \end{aligned}$$

6. Calculons de même pour les autres villes :

   $$\begin{aligned} P(L|S) &= \frac{P(S|L) \times P(L)}{P(S)} &= \frac{0.53 \times 0.142}{0.493} &= 0.153 \\ P(T|S) &= \frac{P(S|T) \times P(T)}{P(S)} &= \frac{0.424 \times 0.123}{0.493} &= 0.106 \\ P(M|S) &= \frac{P(S|M) \times P(M)}{P(S)} &= \frac{0.448 \times 0.071}{0.493} &= 0.065 \end{aligned}$$

La ville de résidence la plus probable pour un électeur votant *Sarkozy* est donc Paris. On vérifie bien : $0.676 + 0.153 + 0.106 + 0.065 = 1$. Le calcul n'était pour autant pas nécessaire puisque Paris avait déjà une probabilité de 0.676, les autres devant "se partager le reste du gateau" donc la somme faisant 1, elles devaient forcément être inférieures.

7. Calculons pour les 4 villes :

$$\begin{aligned}
P(Pa|R) &= \frac{P(R|L)\times P(L)}{P(R)} &= \frac{0.498\times 0.664}{0.507} &= 0.652\\
P(L|R) &= \frac{P(R|L)\times P(L)}{P(R)} &= \frac{0.47\times 0.142}{0.507} &= 0.132\\
P(T|R) &= \frac{P(R|T)\times P(T)}{P(R)} &= \frac{0.576\times 0.123}{0.507} &= 0.140\\
P(M|R) &= \frac{P(R|M)\times P(M)}{P(R)} &= \frac{0.552\times 0.071}{0.507} &= 0.077
\end{aligned}$$

Paris est donc aussi la ville de résidence la plus probable d'un électeur ayant voté *Royal*. Ce qui n'est pas surprenant vu le déséquilibre de la distribution de la résidence en faveur de Paris.

Partie 2 :

8. L'événement contraire de l'événement A est $\overline{A}$: "4 électeurs au moins parmi ces 10 votent *Sarkozy*" ?

9. $$\begin{aligned}
P(A) &= C_{10}^0\times 0.448^0\times 0.552^{10} + C_{10}^1\times 0.448^1\times 0.552^9 + C_{10}^2\times 0.448^2\times 0.552^8\\
&\quad + C_{10}^3\times 0.448^3\times 0.552^7\\
&= 1\times 0.552^{10} + 10\times 0.448^1\times 0.552^9 + 45\times 0.448^2\times 0.552^8 + 120\times 0.448^3\times 0.552^7\\
&= 0.270
\end{aligned}$$

10. Les électeurs votent indépendamment les uns des autres. Choisis au hasard, ils ont tous la même probabilité de voter pour *Sarkozy*. À chacun, on associe la variable de Bernoulli de probabilité 0.448 qui vaut 1 si vote *Sarkozy* et 0 sinon. X est donc la somme de ces 10 variables aléatoires et compte donc le nombre de personnes parmi les 10 qui ont voté *Sarkozy*.

$$X \sim \mathcal{B}in(10, 0.448)$$

11. $E(X) = 10\times 0.448 = 4.48$. Si on observait un grand nombre de groupes de 10 personnes, la moyenne du nombre de personnes ayant voté *Sarkozy* dans les groupes devrait être proche de 4.48

Exercice 2 :

Partie 1 :

Notons Y_G la variable aléatoire mesurant l'IMC d'un garçon de 10 ans.

$$Y_G \sim \mathcal{N}(16.5, 2^2)$$

Et notons U une variable aléatoire de loi normale centrée réduite.

1. $$\begin{aligned}
P(Y_G > 15.5) &= P\left(\frac{Y_G - 16.5}{2} > \frac{15.5 - 16.5}{2}\right)\\
&= P(U > -0.5)\\
&= 1 - P(U > 0.5)\\
&= 1 - 0.309 = 0.691
\end{aligned}$$

2. $$\begin{aligned}
P(Y_G < 17.5) &= P\left(\frac{Y_G - 16.5}{2} > \frac{17.5 - 16.5}{2}\right)\\
&= P(U < 0.5)\\
&= 1 - 0.309 = 0.691
\end{aligned}$$

3. On cherche l tel que $P(Y_G > l) = 0.40$.
$P\left(\dfrac{Y_G - 16.5}{2} > \dfrac{l - 16.5}{2}\right) = P(U > \dfrac{l-16.5}{2}) = 0.40$
Alors $\dfrac{l-16.5}{2} = 0.2533$. D'où $l = 16.5 + 2 \times 0.2533 = 17$.
4. Étant donné la définition de l'obésité, le pourcentage d'enfants obèses est 3%. La valeur seuil lui correspondant est la valeur s telle que $P(Y_G > s) = 0.03$.
Par la même démarche qu'à la question précédente, on obtient $s = 16.5 + 2 \times 1.8808 = 20.2616$.

Partie 2 :
Notons Y_F la variable aléatoire mesurant l'IMC d'une fille de 10 ans.

$$Y_F \sim \mathcal{N}(16.2, \sigma^2)$$

Et notons U une variable aléatoire de loi normale centrée réduite.

5. On sait que $P(Y_F < 19) = 0.95$.
$$\begin{array}{lrcl} \text{D'où} & P\left(\dfrac{Y_F - 16.2}{\sigma} < \dfrac{19 - 16.2}{\sigma}\right) & = & 0.95 \\ & P(U > \dfrac{19-16.2}{\sigma}) & = & 0.05 \end{array}$$
Ainsi $\dfrac{19-16.2}{\sigma} = 1.6449$. Donc $\sigma = \dfrac{19-16.2}{1.6449} = 1.7022$.
6. $$\begin{array}{rcl} P(Y_F < 17.5) & = & P\left(\dfrac{Y_F - 16.2}{1.7022} < \dfrac{17.5 - 16.2}{1.7022}\right) \\ & = & P(U < 0.7637) \\ & = & 0.777 \end{array}$$
7. On cherche v_1 telle que $P(Y_F > v_1) = 0.80$
$P\left(\dfrac{Y_F - 16.2}{1.7022} > \dfrac{v_1 - 16.2}{1.7022}\right) = P(U > \dfrac{l - 16.2}{1.7022}) = 0.80$
Alors $\dfrac{l - 16.2}{1.7022} = -0.8416$. D'où $l = 16.2 - 1.7022 \times 0.8416 = 14.77$.
8. On cherche v_2 telle que $P(Y_F < v_2) = 0.40$
$P\left(\dfrac{Y_F - 16.2}{1.7022} > \dfrac{v_1 - 16.2}{1.7022}\right) = P(U > \dfrac{l - 16.2}{1.7022}) = 0.60$
Alors $\dfrac{l - 16.2}{1.7022} = -0.2533$. D'où $l = 16.2 - 1.7022 \times 0.2533 = 15.77$.
9. $P(Y_F \in [v_1\,;\,v_2]) = 0.20$ puisqu'il y a 20% de chances en dessous de v_1 et 60% au dessus de v_2.

Partie 3 :
On s'intéresse à une classe d'enfants composée de 12 garçons et 18 filles.

10. Définissons les 2 événements :
G : "être une garçon" et F : être une fille"
Alors tous les enfants étant équiprobables, on a : $P(G) = \frac{12}{30} = 0.4$ et $P(F) = \frac{18}{30} = 0.6$
11. Définisson l'événement I : "son IMC est inférieur à 17.5". Alors par la formule des probabilités totales, on a :
$$\begin{array}{rcl} P(I) & = & P(I|G) \times P(G) + P(I|F) \times P(F) \\ & = & 0.309 \times 0.4 + 0.777 \times 0.6 \\ & = & 0.6825 \end{array}$$
12. 3% puisque c'est le cas dans chacun des groupes, ça l'est aussi dans la population globale des enfants de 10 ans.