ISARA 1ère année Pascale **NEYRAN**

Corrigé épreuve STATISTIQUE 2008

Exercice I (4 points)

Réponses: C / A / C / C / B / C / A / B.

Exercice II (16,5 points)

Partie A : (11 points)

1) Tableau de calculs :

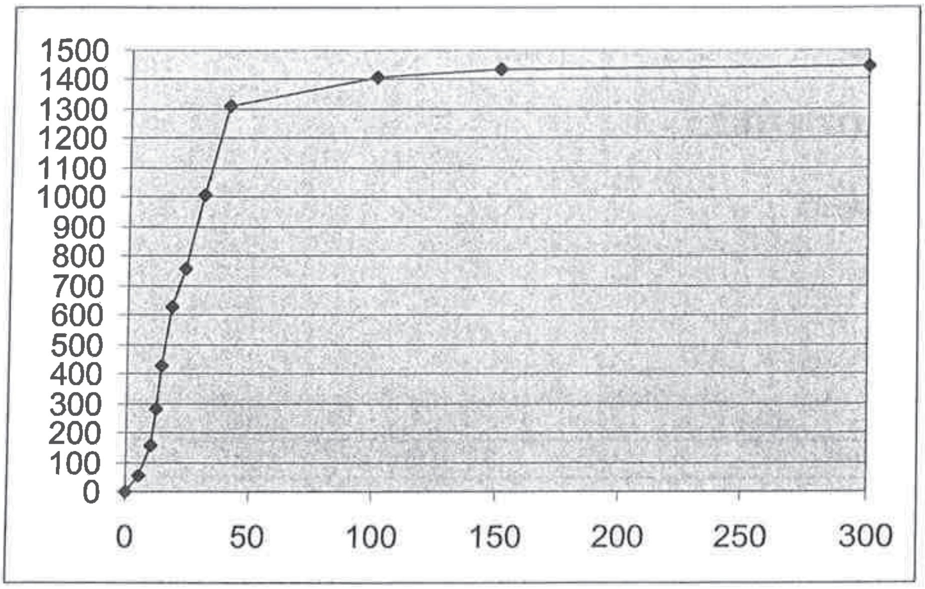
classe	ni	ai	ci	hi	ni^+
[0 ; 5[	56	5	2,5	22,40	56
[5 ; 10[	100	5	2,5	40,00	156
[10 ; 12[	123	2	1	123,00	279
[12 ; 14[	150	2	1	150,00	429
[14 ; 18[	200	4	2	100,00	629
[18 ; 23[	130	5	2,5	52,00	759
[23 ; 30[	250	7	3,5	71,43	1009
[30 ; 40[	300	10	5	60,00	1309
[40 ; 100[	98	60	30	3,27	1407
[100 ; 150[	26	50	25	1,04	1433
[150 ; 300[	12	150	75	0,16	1445

Classe modale	[12 ; 14[car elle correspond à la hauteur hi la plus forte.
Moyenne	39993/1445 = 27,68.
D(4)	$\frac{4}{10}\times 1446 = 578{,}40$
Q_1	$Q_{(1)} = 361{,}50 \Rightarrow Q_1 = 13{,}10$ $\frac{Q_1 - 12}{14-12} = \frac{361{,}50-279}{429-279}$
Médiane	Le rang de la me = 723 $\Rightarrow$ me = 21,62 $\frac{Q_2 - 18}{23-18} = \frac{723-629}{759-629}$
SCE	$n\times s^2(x) = 1\ 082\ 174{,}07$
CV	$(27{,}37/27{,}68)\times 100 = 98{,}88 \Rightarrow$ la série est très dispersée.

2) Calcul du coefficient de Yule :

$$s = \frac{(32{,}52 - 21{,}62) - (21{,}62 - 13{,}10)}{(32{,}52 - 13{,}10)} = 0{,}12 \Rightarrow \text{la série est un peu étalée à droite.}$$

3)



La médiane correspond à l'abscisse du point d'ordonnée 723.

__Partie B__ : (5,5 points)

1) Pour pouvoir comparer la dispersion des 2 zones, il faut comparer leur coefficient de variation.
Pour la zone Alpha, il est égal à 98,88.
Pour la zone Béta, il est égal à $\frac{30}{35} \times 100 = 85{,}71$.
La dispersion des surfaces est plus importante dans la zone Alpha.

2) a) La moyenne des surfaces dans la région Gama est égale à :

$$\frac{(1445 \times 27{,}68 + 1430 \times 35)}{(1445 + 1430)} = 31{,}32.$$

La variance intra : $\frac{(1445 \times 27{,}37^2 + 1430 \times 30^2)}{(1445 + 1430)} = 824{,}16.$

La variance inter : $\frac{1445(27{,}68 - 31{,}32)^2 + 1430(35 - 31{,}32)^2}{(1445 + 1430)} = 13{,}40.$

La variance totale est donc égale à 824,16 + 13,40 = 837,56.

b) La dispersion des surfaces dans la région s'explique essentiellement par une forte dispersion des surfaces interne aux deux zones (à raison de 98,40 %). La dispersion des surfaces entre les 2 zones est par contre très faible (1,6 %).
La dispersion intra explique presque intégralement la variance globale.

Exercice III (6 points)

Hypothèse nulle : La répartition des groupes sanguins dans l'échantillon des 200 malades est **conforme** à celle observée dans la population en générale.

Effectifs théoriques :

Groupe sanguin	O	A	B	AB	**total**
Effectifs théoriques	94	86	14	6	200

Critère statistique calculé :	**6,04**
Nombre de degrés de liberté :	**3**
Critère statistique théorique :	**7,81**
Conclusion :	**On accepte** H_0 car si on la refusait, on prendrait un risque de se tromper supérieur à 5 % (compris entre 10 % et 50 %)

Exercice IV (12 points)

Partie A (6 points)

La loi suivie par X est une loi normale de moyenne *m* et d'écart type σ.

1) On obtient un système :

$$P(X < 10) = P(T < \frac{10-m}{\sigma}) = 0,16 \Rightarrow \frac{10-m}{\sigma} = -0,9945$$

$$P(X < 14) = P(T < \frac{14-m}{\sigma}) = 0,63 \Rightarrow \frac{14-m}{\sigma} = 0,3319$$

$$\begin{cases} 10-m = -0,9945\sigma \\ 14-m = 0,3319\sigma \end{cases} \Leftrightarrow \begin{cases} m = 12,999 \approx 13 \\ \sigma = 3,0157 \approx 3,02 \end{cases}$$ on obtient l'écart type en faisant (2) -(1)

2) Le premier quartile vérifie :

$P(X < Q_1) = 0,25 \Rightarrow (Q_1 - 13)/3 = -0,674$ soit $Q_1 =$ **10,98.**

25% des candidats ont obtenu une note inférieure à 10,98.

3) Il faut trouver x tel que $P(13 - x \leq X \leq 13 + x) = 0,95$.

Or $P(T \leq x/3) - P(T < - x/3) = 2P(T\leq x/3) -1 \Rightarrow P(T \leq x/3) = 0,975$.

Dans la table 2 de la loi normale on lit la valeur 1,96 = x/3. D'où l'intervalle centré sur la moyenne **[7,12 ; 18,88].**

Il y a 95% de chances pour que la note d'un membre du personnel soit comprise dans cet intervalle.

Partie B (2,5 points)

1) L(Y) = B(20 ; 0,10). (répétition dans les mêmes conditions d'une expérience à 2 issues)

2) $E(Y) = 20\times0,10 = 2$ et $V(Y) = 20\times0,10\times0,90 = 1,8$ d'où $\sigma(Y) = 1,34$.

3) On cherche $P(Y \leq 5) = 0,9887$.

Partie C (3,5 points)

1) Z suit une loi de Poisson de paramètre 4.

a) $P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0,2381 = 0,7619$.

b) $P(2 \leq Z \leq 7) = P(Z \leq 7) - P(Z < 2) = P(Z \leq 7) - P(Z \leq 1) = 0,9489 - 0,0916 = 0,8573$.

2) U : « nombre d'accidents pour le prochain semestre »

a) Z suit une loi de Poisson de paramètre 4

$U = \sum_{i=1}^{6} Z_i$ où Z_i est la variable aléatoire représentant le nombre d'accidents pour le mois i.

U est donc la somme de 6 variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs égaux à 4 $\Rightarrow$ U suit une loi de Poisson de paramètre 24.

b) La loi de U peut être approchée par un loi normale $N(24 ; \sqrt{24})$ car le paramètre de la loi de Poisson dépasse 18.

$P(U < 30) = P(T < 1,22) = 0,8888$.

Exercice V (5,5 points)

Le centre de gravité a pour coordonnées (226 ; 280).

$$SPE\,(X ; Y) = n\times COV(X ; Y) = 10\times\left(\frac{472000}{10} - 226\times280\right) = -\,160800.$$

$$s^2(X) = \frac{624400}{10} - 226^2 = 11364.$$

La pente de la droite est égale à $= b = \dfrac{COV(X;Y)}{s^2(X)} = \dfrac{-16080}{11364} = -1,4150$ (X et Y varient en sens inverse).

$s^2(Y) = SCE(yi)/n = 230800/10 = 23080$.

On en déduit $r = \dfrac{COV(X;Y)}{s(X)\times s(Y)} = \dfrac{-16080}{\sqrt{11364\times 23080}}$ = -0,9929 ⇒ l'ajustement est d'excellente qualité.

La variance résiduelle $s^2(ei) = (1-r^2)\times s^2(Y) = (1-(-09929)^2)\times 23080 = 326,57$.

Exercice VI (6 points)

Hypothèse nulle : la série peut être ajustée par une loi de Poisson de paramètre 3.

xi	pi	ni' = n×pi
0	0,0498	2,49
1	0,1494	7,47
2	0,224	11,2
3	0,224	11,2
4	0,168	8,4
5	0,1008	5,04
6	0,0504	2,52
7 et plus	0,0336	1,68

Les effectifs théoriques doivent être inférieurs ou égaux à 5, mais on tolère un effectif théorique supérieur à 5 à chaque extrémité. On va donc regrouper les deux dernières lignes.

xi	ni	ni' = n×pi
0	3	2,49
1	7	7,47
2	11	11,2
3	11	11,2
4	8	8,4
5	5	5,04
6 et plus	5	4,2

Critère statistique calculé : 0,3129

Critère statistique théorique : 11,10 pour un ddl de 5 (7-1-1) et un risque de 5%. Si on néglige le nombre de paramètre estimé soit 1, le ddl est égal à 6 et le critère statistique théorique vaut 12,6.

Conclusion : On accepte H_0, car si on la refuse on prend un risque de se tromper supérieur à 5% (risque réel compris entre 99,5 % et 99,9 % pour un ddl de 5 et supérieur à 99,9 % pour un ddl de 6).