Université de Picardie Jules Verne **2017-2018**
IAE Amiens

Licence mention Gestion parcours Management et Marketing Vente - Semestre 5
Statistiques appliquées - Corrigé du partiel du 15/11/2017

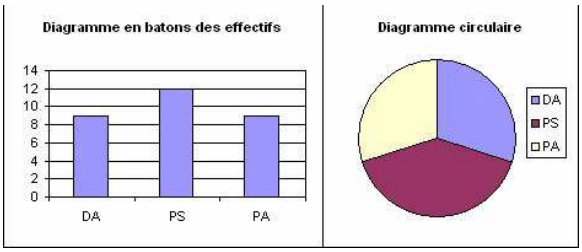
Exercice 1.

1) La population étudiée est l'ensemble des clients du centre commercial. La taille de l'échantillon est $n = 30$. Les variables étudiées sont :

- boisson préférée ; variable qualitative ;
- achat préalable de RC ; variable qualitative ;
- réaction préalable ; variable qualitative ;
- quantité de boisson consommée ; variable quantitative continue.

2) Les données obtenues avec la question Q3 sont données dans le tableau modalités/effectifs et sur les graphiques suivants :

Centre x_i	Effectif n_i
DA	9
PS	12
PA	9
Total	30



Exercice 2.

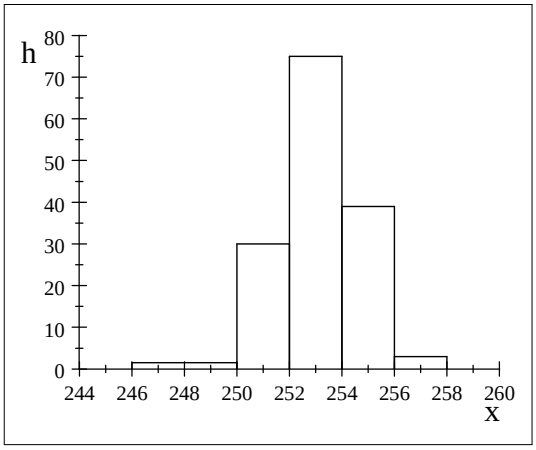
1)

Classe C_i	Centre x_i	Effectif n_i	Ampl. a_i	Haut. $h_i = \frac{n_i}{a_i}$	Fréq. f_i	Freq. Cum. F_i
$]246;250]$	248	6	4	1.5	0.02	0.02
$]250;252]$	251	60	2	30	0.20	0.22
$]252;254]$	253	150	2	75	0.50	0.72
$]254;256]$	255	78	2	39	0.26	0.98
$]256;258]$	257	6	2	3	0.02	1
Total		300			1	

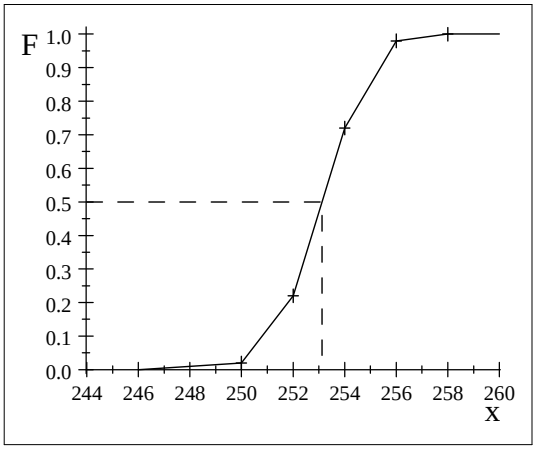
a) La population étudiée est l'ensemble des sachets de café conditionnés par l'entreprise. La variable étudiée est la masse, elle est quantitative continue. La taille de l'échantillon est $n = 300$.

b) Voir histogramme des effectifs ci-dessous. Hauteur des rectangles : $h_i = \frac{n_i}{a_i}$.

c) Calcul des fréquence $f_i = \frac{n_i}{n}$: voir tableau ci-dessus. Calcul des fréquences cumulées $F_i = f_1 + \cdots + f_i$: voir tableau ci-dessus. Polygone des fréquences cumulées : voir ci-dessous. Pour chaque classe C_i, on construit le point d'abscisse la borne de droite de la classe et d'ordonnée F_i.



1) b) Histogramme des effectifs



1) c) Polygone des fréquences cumulées

d) La médiane est la valeur de la variable correspondant à une fréquence cumulée égale à 0.50.
Lecture graphique : environ 253 g.
Formule d'interpolation linéaire : $Me = 252 + (254 - 252) \times \dfrac{0.50 - 0.22}{0.72 - 0.22} = 253.12$ g.
Interprétation : la moitié (50%) des sachets ont une masse inférieure ou égale à 253.12 g, l'autre moitié ayant une masse supérieure à 253.12 g.

e) A l'aide de la calculatrice, en utilisant les centres de classe x_i, on obtient les valeurs suivantes :
Moyenne : $\bar{x} = \frac{1}{n}\sum n_i x_i = 253.1$ g Ecart-type : $s_x \simeq 1.628$ g.

f) Le pourcentage de sachets de l'échantillon de masse inférieure à 250 g est égal à $\dfrac{6}{300} \times 100 = 2$ %, donc est inférieur à 4%. De plus, la moyenne de l'échantillon est comprise entre 252 g et 254 g. Mais l'écart type de l'échantillon est égal à 1.628 g, donc n'est pas inférieur à 1.5 g. La machine doit donc être réglée.

g) Le coefficient d'asymétrie de Fisher est $S \simeq -0,356 < 0$. Cela est cohérent avec l'histogramme précédent qui montre une asymétrie avec classe modale à droite de la moyenne.

2) La variable aléatoire X qui représente la masse d'un sachet suit la loi normale $\mathcal{N}(253; 1.5)$.
Ainsi, $T = \dfrac{X - 253}{1.5}$ suit la loi normale $\mathcal{N}(0;1)$.
a) $P(X \leqslant 250) = P\left(\dfrac{X - 253}{1.5} \leq \dfrac{250 - 253}{1.5}\right) = P(T \leq -2)$
$= \phi(-2) = 1 - \phi(2) = 1 - 0.9772 = 0.0228.$
b) La probabilité qu'un sachet ait masse d'au moins 250 g est :
$P(X \geq 250) = 1 - P(X < 250) = 1 - P(X \leqslant 250) = 1 - 0.0228 = 0.9772.$
c) Sans changer la valeur de l'écart type σ, on cherche μ tel que $P(X \leqslant 250) \leq 0.01$, ce qui équivaut successivement à : $P\left(\dfrac{X - \mu}{1.5} \leq \dfrac{250 - \mu}{1.5}\right) \leq 0.01$, $P\left(T \leq \dfrac{250 - \mu}{1.5}\right) \leq 0.01$,
$\phi\left(\dfrac{250 - \mu}{1.5}\right) \leq 0.01 < 0.5 = \phi(0)$, ce qui induit que $\dfrac{250 - \mu}{1.5} < 0$,
$1 - \phi\left(-\dfrac{250 - \mu}{1.5}\right) \leq 0.01$, $\phi\left(-\dfrac{250 - \mu}{1.5}\right) \geq 0.99 \simeq \phi(2.33)$, $-\dfrac{250 - \mu}{1.5} \geq 2.33$,
$\dfrac{250 - \mu}{1.5} \leq -2.33$, $250 - \mu \leq -2.33 \times 1.5$, $\mu \geq 250 + 2.33 \times 1.5 = 253.495.$
La moyenne μ doit donc être supérieure à 253,5 g.

3) a) On peut considérer que l'on effectue $n = 100$ tirages avec remise (production très importante) d'un sachet dans la population des sachets produits (à deux catégories) contenant une proportion $p = 0.02$ de sachets de masse inférieure à 250 g et une proportion $1 - p = 0.98$ de sachets de masse supérieure à 250 g. La variable aléatoire Y égale au nombre de sachets de masse inférieure à 250 g obtenus suit donc la loi Binomiale $\mathcal{B}(n;p) = \mathcal{B}(100; 0.02)$.

b) En moyenne, le nombre de sachets dont la masse est inférieure à 250 g dans un lot est : $E(Y) = np = 100 \times 0.02 = 2$.

c) La probabilité que tous les sachets aient une masse supérieure à 250 g, c'est-à-dire qu'aucun sachet n'ait une masse inférieure à 250 g, est :
$$P(Y = 0) = \binom{100}{0}(0.02)^0(0.98)^{100-0} \simeq 0.13262.$$
d) Comme $n = 100 \geq 30$, $p = 0.02 \leq 0.1$ et $np = 100 \times 0.02 = 2 \leq 10$, on peut approcher la loi de Y par la loi de Poisson de paramètre $\lambda = 2$.
On a alors $P(Y = 0) \simeq e^{-2}\dfrac{2^0}{0!} \simeq 0.13534$

Exercice 3.

Partie A - Ajustement affine

1) a) Un ajustement affine (droite) peut sembler approprié car les points du nuage sont relativement "alignés". Mais attention, avec le point le plus à droite, le nuage est un peu "incurvé", ce qui peut faire penser à une parabole ou une courbe "exponentielle".

b) Le coefficient de corrélation linéaire entre x et y, arrondi au millième, est $r_{x;y} \simeq 0.925$. Comme $r_{x;y}^2 \simeq 0.8552 \geq \frac{3}{4}$ mais n'est pas très proche de 1, l'ajustement affine est correct et cela confirme la réponse à la question précédente.

2) a) Une équation de la droite de régression de y en x par la méthode des moindres carrés (coefficients arrondis au millième) est $y = 58.342x + 87.623$.

b) Pour $x = 4$ milliers d'euros de frais publicitaires engagés, une estimation de la fréquentation est $y = 58.342 \times 4 + 87.623 \simeq 321$ milliers de clients. Cette estimation est relativement fiable étant donné la qualité de l'ajustement. Mais attention, $x = 4$ sort de la zone d'étude et rien ne dit que le phénomène se prolonge de façon analogue, surtout au vu de la forme "incurvée" du nuage.

c) Pour $x = 5$ milliers d'euros de frais publicitaires engagés, une estimation de la fréquentation est $y = 58.342 \times 5 + 87.623 \simeq 379$ milliers de clients. Ce qui surestime un peu le nombre de clients effectivement observés, à savoir 330 milliers de clients.

3) a) En utilisant ce modèle $y = 58.342x + 87.623$, pour espérer 400 milliers de clients au cours d'un mois, le montant x des frais publicitaires devant être engagés doit vérifier $400 = 58.342x + 87.623$, ce qui donne $x = \frac{400 - 87.623}{58.342} \simeq 5.3542$ milliers d'euros, soit environ 5354 €.

b) Il n'est pas cohérent d'utiliser ce modèle estimer x à partir de y. Il aurait été préférable de déterminer la droite de régression de x en y par la méthode des moindres carrés.

Partie B - Autres ajustements

1) Le meilleur ajustement est celui donne le plus grand R^2. Il s'agit ici du modèle parabolique $y = 33.982x^2 - 55.333x + 173.62$.

2) La courbe représentée en pointillés est visuellement celle qui ajuste au mieux le nuage de points. Son équation est celle donnée au 1). On peut aussi la trouver avec la calculatrice en cherchant les deux courbes exponentielle et parabolique et en visualisant ces courbes sur le nuage de points.

Exercice 4.

1) a) Considérant l'expérience aléatoire $\mathcal{E}$ "choisir un visiteur au hasard", on peut considérer l'espace probabilisé $(\Omega, \mathcal{A}, P)$ suivant :

- Univers : $\Omega = \{\text{visiteurs}\}$; $card\Omega = N$ est inconnu (mais fini) ;
- Ensemble des événements : $\mathcal{A} = \mathcal{P}(\Omega)$ (car Ω est fini)
- Probabilité : P = équiprobabilité sur $(\Omega, \mathcal{A})$

Considérant les événements donnés dans l'énoncé, on a :

$P(S) = \frac{\frac{2}{5}N}{N} = 0.4$, $P(T) = 0.1$ et $P(O) = 0.5$

et $P^S(R) = 0.65$, $P^T(R) = 0.53$ et $P^O(C) = 0.59$.

b) La probabilité que le visiteur utilise un smartphone et quitte le site après avoir visité la première page est : $P(S \cap R) = P(S)P^S(R) = 0.4 \times 0.65 = 0.26$.

c) Utilisant le système complet d'événements de probabilité non nulle $\{S, T, R\}$, la formule des probabilités complètes donne la probabilité que le visiteur quitte le site après avoir visité la première page :

$$P(R) = P(S \cap R) + P(T \cap R) + P(0 \cap R) = P(S)P^S(R) + P(T)P^T(R) + P(O)P^O(R)$$
$$= 0.4 \times 0.65 + 0.1 \times 0.53 + 0.5 \times 0.59 = 0.608.$$

d) La probabilité que le visiteur utilise un ordinateur sachant qu'il a quitté le site après avoir consulté la première page est :

$$P^R(O) = \frac{P(O \cap R)}{P(R)} = \frac{P(O)P^O(R)}{P(R)} = \frac{0.5 \times 0.59}{0.608} = \frac{0.295}{0.608} \simeq 0.4852.$$

2) a) Un visiteur utilise un seul outil à la fois donc les évènements O et S sont incompatibles. De plus, s'il utilise un ordinateur, il n'utilise pas un smartphone donc les évènements O et S ne sont pas indépendants ; ce qui est confirmé par $P(O \cap S) = 0 \neq 0.2 = 0.5 \times 0.4 = P(O)P(S)$.

b) On a $P(O \cap R) = 0.295 \neq 0$ donc $O \cap R \neq \varnothing$ et les événements O et R ne sont pas incompatibles.

On a $P(O)P(R) = 0.5 \times 0.608 = 0.304 \neq 0.295 = P(O \cap R)$ donc les événements O et R ne sont pas indépendants.

Exercice 5.

1) Le temps d'attente X à une caisse, exprimé en minute, est une variable aléatoire qui suit la loi uniforme sur l'intervalle [0; 12]. On a donc :

$$F_X(x) = P(X \leq x) = \begin{cases} 0 \text{ si } x < 0 \\ \frac{x}{12} \text{ si } x \in [0;12] \\ 1 \text{ si } x > 12 \end{cases}, E(X) = \frac{0+12}{2} = 6 \text{ et } V(X) = \frac{(12-0)^2}{12} = 12.$$

a) La probabilité qu'un client attende au moins 5 minutes avant d'être pris en charge est :

$$P(X \geq 5) = 1 - P(X \leq 5) = 1 - F_X(5) = 1 - \frac{5}{12} = \frac{7}{12}.$$

b) La probabilité qu'un client attende entre 2 et 5 minutes avant d'être pris en charge est :

$$P(2 \leq X \leq 5) = F_X(5) - F_X(2) = \frac{5}{12} - \frac{2}{12} = \frac{3}{12} = \frac{1}{4} = 0.25.$$

c) Le temps moyen d'attente à une caisse est $E(X) = 6$ minutes.

d) On cherche le nombre réel x tel que $P(X \leq x) = P(X \geq x)$, ce qui équivaut successivement à $P(X \leq x) = 1 - P(X \leq x)$, $2P(X \leq x) = 1$, $P(X \leq x) = \frac{1}{2}$, $F_X(x) = \frac{1}{2}$, $\frac{x}{12} = \frac{1}{2}$, $x = \frac{12}{2} = 6$ minutes. Il s'agit de la médiane du temps d'attente. On peut remarquer que médiane et moyenne sont ici égales.

2) Le temps d'attente Y à une caisse automatique, exprimé en minute, est une variable aléatoire qui suit la loi exponentielle de paramètre $\theta = 0,5$. On a donc :

$$F_Y(x) = P(Y \leq x) = \begin{cases} 1 - e^{-0.5x} \text{ si } x \geq 0 \\ 0 \text{ si } x < 0 \end{cases}, \quad E(Y) = \frac{1}{0.5} = 2, \quad Var(Y) = \frac{1}{0.5^2} = 4.$$

a) Le temps d'attente moyen à une caisse automatique est $E(Y) = 2$.

b) La probabilité que le temps d'attente à une caisse automatique soit supérieur à 2 minutes est :

$$P(Y \geq 2) = 1 - P(Y \leq 2) = 1 - F_Y(2) = 1 - (1 - e^{-0.5\times 2}) = e^{-1} \simeq 0.3679.$$

c) La probabilité que le temps d'attente à une caisse automatique soit compris entre 30 secondes et 3 minutes est :

$$P(0.5 \leq X \leq 3) = F_X(3) - F_X(0.5) = (1 - e^{-0.5\times 3}) - (1 - e^{-0.5\times 0.5}) \simeq 0.5557.$$

3) a) On peut considérer que l'on répète $n = 10$ fois l'expérience aléatoire "observer une caisse automatique pendant une journée" au cours de laquelle l'événement A "la caisse tombe en panne" a la probabilité $p = 0.1$ d'être réalisé. La variable aléatoire Z, correspondant au nombre de caisses automatiques qui tombent en panne pendant une journée donnée, suit donc la loi Binomiale $\mathcal{B}(n;p) = \mathcal{B}(10;0.1)$.

b) La probabilité qu'aucune caisse automatique ne tombe en panne pendant une journée donnée est :

$$P(Z = 0) = \binom{10}{0}(0.1)^0(0.9)^{10-0} \simeq 0.3487.$$