# Corrigé - Série 1

## Rappels des probabilités et statistique descriptive

### Exercice 1

a) Binomiale(10, 0.25)

b) Géométrique(1/13 983 816)

c) Pascal(5, 0.20)

d) Probablement une loi Poisson($\lambda$), où $\lambda$ est le nombre moyen d'accrochages par an.

e) Exponentielle(100) (temps mesuré en heures)

f) Uniforme(0, 15) (temps mesuré en minutes)

g) Probablement une loi normale

h) Approximativement $N(1000 \times 23.5,\ 1000 \times 0.75)$, i.e. $N(23\,500,\ 750)$

### Exercice 2

Soit $X$ la note obtenue au test. On sait que $X \sim N(485, 30^2)$.

a) On cherche la probabilité $P(X \geq 500)$.

$$P(X \geq 500) = P\left(\frac{X - 485}{30} \geq \frac{500 - 485}{30}\right) = P(Z \geq 0.5) = 0.3085.$$

Ainsi, 30.85% des gens satisfont à l'exigence.

b)

$$P(X > 525|X > 500) = \frac{P(X > 525)}{P(X > 500)} = \frac{P(Z > 1,333)}{0,3085} = 0,2956.$$

c) On cherche $k$ tel que $P(X > k) = 0.20$.

$$P(X > k) = P\left(\frac{X - 485}{30} > \frac{k - 485}{30}\right) = 0.20$$

Il suit que $k = 30(0.841) + 485 = 510.23$

**Exercice 3**

$V$ (km/h) $\sim N(112, 225)$.

a) $P(V > 100) = 0.788$.

b) On cherche $k$ tel que $P(V < k) = 0.40$. On trouve $k = 108.2$

c) $P(|V - 105| > 10) = 1 - P(95 < V < 115) = 0.549$.

d) Notons $V_m$ la vitesse exprimée en m/s.
$P(D > 100) = P(V_m^2 > 12 \times 100) = P(V_m > 34,64)$
Si $V_{km}$ est la vitesse en km/h, alors $V_{km} = V_m \times 3600/1000$,
et il suit que $P(V_m > 34,64) = P(V_m > 124,704) = 0,198$.

On aurait aussi pu calculer la probabilité $P(V_m > 34,64)$ directement en modifiant les paramètres de la loi normale ($V_m \sim N(31.11, 17.36)$).

**Exercice 4**

On pourrait ordonner les valeurs dans chaque tige pour clarifier les idées :

| Tiges | Feuilles |
|---|---|
| 0 | 0, 1, 2, 3, 5, 6, 6, 7, 9, 9 |
| 1 | 0, 0, 2, 3, 3, 3, 3, 5, 7, 9 |
| 2 | 0, 1, 2, 2, 2, 4, 4, 5, 5 |
| 3 | 3, 5, 5, 5, 8, 9 |
| 4 | 4 |
| 5 | 0, 6, 9 |
| 6 | 3 |

a) Le jeu de données contient 40 observations. La médiane est donc

$$\tilde{x} = Q_2 = x_{(0,5 \times n + 0,5)} = x_{(20,5)} = \frac{x_{(20)} + x_{(21)}}{2} = \frac{19 + 20}{2} = 19.5$$

b) La moyenne de cet ensemble de données sera supérieure à la médiane, car la distribution est clairement asymétrique à droite.

c) Le rang centile de 44 est la proportion de données (parmi les autres observations) qui sont strictement inférieures à 44. Puisque 44 est la $36^e$ valeur de l'échantillon ordonné, alors son rang centile est l'entier supérieur à $\frac{36}{40} \times 100 = 90$.

Selon une seconde définition, on aurait aussi pu calculer le rang centile comme l'entier supérieur à $\frac{35,5}{40} \times 100 = 88,75$, soit 89.

d) Le rang cinquième de 44 est calculé en divisant l'échantillon en 5 parties de fréquence à peu près égale, puis en attribuant les rangs de 1 à 5 en commençant par les valeurs les plus élevées. On obtient clairement que 44 est au premier rang cinquième.

## Exercice 5

a)

| | *IMC* | *Taille en mètres* | *Poids en kg* |
|---|---|---|---|
| Moyenne | 22,94266187 | 1,715107914 | 67,44043885 |
| Erreur-type | 0,265016694 | 0,006063438 | 0,801276227 |
| Médiane | 22,008 | 1,71 | 65,95 |
| Mode | 24,418 | 1,69 | 72,2 |
| Écart-type | 4,418711317 | 0,101097707 | 13,35994455 |
| Variance de l'échantillon | 19,5250097 | 0,010220746 | 178,4881184 |
| Kurstosis (Coefficient d'aplatissement) | 7,362717424 | 0,057438594 | 0,170516793 |
| Coefficient d'asymétrie | 1,917730375 | -0,019343838 | 0,70160475 |
| Plage | 36,755 | 0,65 | 59,826 |
| Minimum | 15,581 | 1,36 | 44,5 |
| Maximum | 52,336 | 2,01 | 104,326 |
| Somme | 6378,06 | 476,8 | 18748,442 |
| Nombre d'échantillons | 278 | 278 | 278 |
| Niveau de confiance(95,0%) | 0,5217026 | 0,011936271 | 1,577364372 |

i) La moyenne étant plus grande que la médiane, la distribution de l'IMC sera légèrement asymétrique à droite.

ii) écart-type = $s$, erreur-type = écart-type de la moyenne = $s/\sqrt{n}$.

iii) coefficients d'aplatissement (*kurtosis*) : caractérise la forme de pic ou l'aplatissement relatifs d'une distribution comparée à une distribution normale. Un kurtosis positif indique une distribution relativement pointue, tandis qu'un kurtosis négatif signale une distribution relativement aplatie.
coefficient d'asymétrie : caractérise le degré d'asymétrie d'une distribution par rapport à sa moyenne. Une asymétrie positive indique une distribution unilatérale décalée vers les valeurs les plus positives. Une asymétrie négative indique une distribution unilatérale décalée vers les valeurs les plus négatives.

iv) plage = étendue = $\max(x_i)$ - $\min(x_i)$

v) *Niveau de confiance* est bizarrement représenté dans Excel par la marge d'erreur de l'intervalle de confiance sur la moyenne.

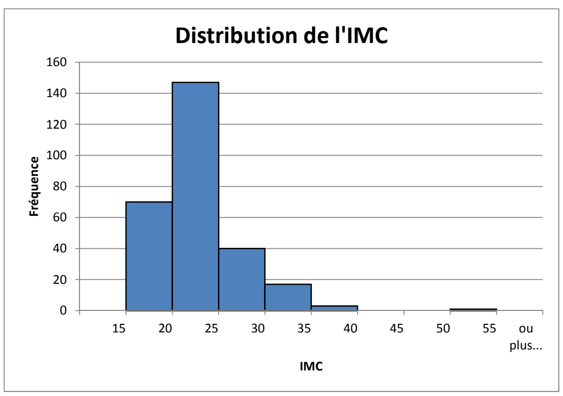
vi) On trouve facilement sur Internet que $IMC = \dfrac{Poids}{Taille^2}$

vii) Non. Notons $Z$ la variable IMC. On sait que $z_i = \dfrac{p_i}{t_i^2}$

On peut en déduire que $\overline{z} = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{p_i}{t_i^2}$ $(= 22,94266$ dans notre exemple).

Cette quantité n'est pas égale à $\dfrac{\overline{p}}{\overline{t}^2} = \dfrac{\frac{1}{n}\sum_{i=1}^{n} p_i}{\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)^2}$ $(= 22,92649$ dans notre exemple.)

b) Vous devez coller les bâtons pour obtenir un histogramme (double-cliquez sur les bâtons pour atteindre l'option.) Il faut aussi justifier les valeurs de l'axe horizontal à droite (Sélectionnez l'axe, puis cliquez avec le bouton de droite).




c) Le poids semble un meilleur prédicteur.