MAT 2779, Introduction à la biostatistique

Devoir 2 - solutionnaire

Total=100 points

Partie I) *Total*=50 points
Exercice 6.1 : (10 points)

(a) $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) = 0{,}2 + 0{,}1 + 0{,}1 = 0{,}4$

(b) $P(X \leq 3) = P(X = 2) + P(X = 3) = 0{,}05 + 0{,}25 = 0{,}30$

(c) $E[X] = \sum x\, f(x) = 4{,}35.$

(d) $V[X] = [\sum x^2\, f(x)] - \mu_X^2 = 20{,}75 - (4{,}35)^2 = 1{,}8275.$

Exercice 6.5 : (10 points) Soit X une variable aléatoire binomiale avec $n = 20$ et $p = 0{,}5$. On veut

$$P(X \geq 17) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20) = 0{,}0013,$$

où

$$P(X = 17) = \left(\begin{array}{c} 20 \\ 17 \end{array}\right) p^{17}(1-p)^3 = 0{,}001087; P(X = 18) = \left(\begin{array}{c} 20 \\ 18 \end{array}\right) p^{18}(1-p)^2 = 0{,}0001811;$$

$$P(X = 19) = \left(\begin{array}{c} 20 \\ 19 \end{array}\right) p^{19}(1-p)^1 = 1{,}91 \times 10^5; P(X = 20) = \left(\begin{array}{c} 20 \\ 20 \end{array}\right) p^{20}(1-p)^0 = 9{,}54 \times 10^{-7}.$$

Exercice 7.8 : (15 points) Soit X le poids d'un ours grizzly (en kg). X suit une loi normale de moyenne $\mu = 357$ et $\sigma = 21$.

(a) $P(X > 420) = 1 - \Phi\left(\frac{420-357}{21}\right) = 1 - \Phi(3) = 1 - 0{,}9987 = 0{,}0013$

(b) $P(X > 300) = 1 - \Phi\left(\frac{300-357}{21}\right) = 1 - \Phi(-2{,}71) = 1 - 0{,}0034 = 0{,}9966$

(c) On veut x_0 tel que $0{,}05 = P(X < x_0) = \Phi\left((x_0 - 357)/21\right)$. Alors,

$$-1{,}645 = \frac{x_0 - 357}{21} \quad \Rightarrow \quad x_0 = 21\,(-1{,}645) + 357 = 322{,}455.$$

(d) On veut y_0 tel que $0{,}75 = P(X < y_0) = \Phi\left((y_0 - 357)/21\right)$. Alors,

$$0{,}675 = \frac{y_0 - 357}{21} \quad \Rightarrow \quad y_0 = 21\,(0{,}675) + 357 = 371{,}175.$$

(e) Soit Y le nombre d'ours grizzly parmi $n = 6$ qui ont un poids inférieur à 300kg. Y suit une loi binomiale avec $n = 6$ et $p = P(X < 300) = 0{,}0034$ (voir la partie (b)). On veut $P(Y \leq 2) = \binom{6}{0} p^0(1-p)^6 + \binom{6}{1} p^1(1-p)^5 + \binom{6}{2} p^2(1-p)^4 = 0{,}9999992 \approx 1.$

Exercice 9.4 : (15 points) Nous allons arranger les valeurs en ordre croissant.

$$4{,}24{,}2;\ 4{,}3;\ 4{,}4;\ 4{,}5;\ 4{,}6;\ 4{,}6;\ 4{,}7;\ 4{,}94{,}95{,}15{,}6$$

(a) On veut la médiane et les quartiles q_1 et q_3. Les rangs pour ces statistiques sont :

	rang
q_1	$(n+1)\,25\% = (12+1)(0{,}25) = 3{,}25$
médiane	$(n+1)\,50\% = (12+1)(0{,}50) = 6{,}5$
q_3	$(n+1)\,75\% = (12+1)(0{,}75) = 9{,}75$

Alors,

$$\text{médiane} = (1-0{,}5)y_6 + (0{,}5)y_7 = (0{,}5)(4{,}6) + (0{,}5)(4{,}6) = 4{,}6;$$

$$q_1 = (1-0{,}25)y_3 + (0{,}25)y_4 = (0{,}75)(4{,}3) + (0{,}25)(4{,}4) = 4{,}325;$$

$$q_3 = (1-0{,}75)y_9 + (0{,}75)y_{10} = (0{,}25)(4{,}9) + (0{,}75)(4{,}9) = 4{,}9.$$

(b) La distance interquartile est $\text{DIQ} = q_3 - q_1 = 4{,}9 - 4{,}325 = 0{,}575$. Calculons

$$\text{clôture inf.} = q_1 - 1{,}5\,\text{DIQ} = 3{,}4625$$

et

$$\text{clôture sup.} = q_3 + 1{,}5\,\text{DIQ} = 5{,}7625.$$

Aucune valeur est à l'extérieur des clôtures, alors il n'y a pas de valeur aberrante.

(c) On veut la moyenne $\overline{x} = \sum_{i=1}^{n} x_i/n = 56/12 = 4{,}6667$ et on veut l'écart type

$$s = \sqrt{\frac{(\sum x_i^2) - (\sum x_i)^2/n}{n-1}} = \sqrt{\frac{263{,}18 - (56)^2/12}{12-1}} = 0{,}40973.$$

Partie II) *Total*=50 points

Utiliser `R` dans la résolution des problèmes suivants. Attacher votre sortie de `R` à votre devoir. La sortie de `R` n'est pas une réponse à la question. Il faut répondre à la question et faire référence à votre sortie de `R`.

(1) **(5 points)**

$$\text{(a)} \quad P(X \geq 3) = 1 - P(X \leq 2) = 1 - F(2) = 0{,}7472$$

$$\text{(b)} \quad P(4 \leq X < 8) = P(X \leq 7) - P(X \leq 3) = F(7) - F(3) = 0{,}49799$$

Calculs avec `R`:

```
> 1-pbinom(2,12,0.3)
[1] 0.7471847
> pbinom(7,12,0.3)-pbinom(3,12,0.3)
[1] 0.4979949
```

(2) **(15 points)**

$$\text{(a)} \quad P(X \leq 9) = F(9) = 0{,}6554; \quad \text{(b)} \quad P(1{,}8 < X < 7{,}8) = F(7{,}8) - F(1{,}8) = 0{,}4615.$$

Calculs avec `R`:

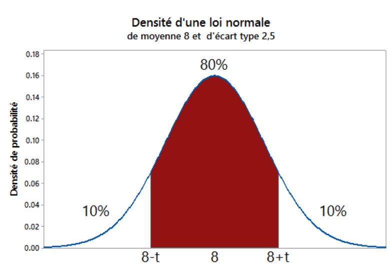
```
> pnorm(9,8,2.5)
[1] 0.6554217
> pnorm(7.8,8,2.5)-pnorm(1.8,8,2.5)
[1] 0.4615495
```

(c) On veut t tel que $P(X > t) = 0{,}2$. Ceci est équivalent à $0,\, 8 = P(X \leq t) = F(t)$. Alors, $t = F^{-1}(0{,}8) = 10{,}10405$.

Calcul avec `R`:

```
> qnorm(0.8,8,2.5)
[1] 10.10405
```

(d) On veut t tel que $P(\mid X - 8 \mid < t) = 0{,}8$. N.B. $\mid X - 8 \mid$ est interpreté comme la distance entre X et 8. Alors, si la distance entre X et 8 est inférieure à t, alors on a $8 - t < X < 8 + t$. Notons que $8 - t$ et $8 + t$ sont équidistants de la moyenne et que $0{,}8 = P(8 - t < X < 8 + t)$. Alors, par la symétrie, $P(X > 8 + t) = P(X < 8 - t) = 0{,}20/2 = 0{,}10$.



Alors, on veut t tel que $0{,}10 = P(X < 8 - t) = F(8 - t)$. Alors, $8 - t = F^{-1}(0{,}1) = 4{,}7961$ et ceci implique que $t = 8 - 4{,}7961 = 3,2039$.

Calcul avec `R`:

```
> qnorm(0.1,8,2.5)
[1] 4.796121
> 8-qnorm(0.1,8,2.5)
[1] 3.203879
```

(2) **(20 points)**

(a) Voici le sommaire à 5 nombres pour chacun de ces groupes.

min	q_1	médiane	moyenne	q_3	max
papous	75,7	78,025	79,7	82,025	84,2
jugulaire	73,1	75,05	76,05	76,3	76,8

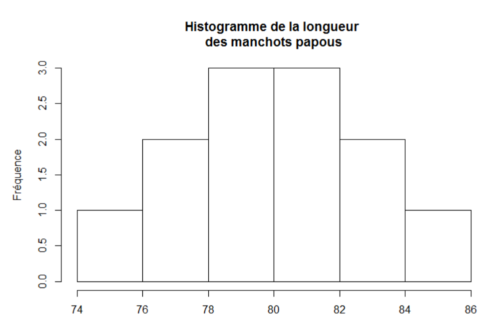
Calcul avec `R`:

```
> quantile(papous,type=6)
    0%    25%    50%    75%   100%
75.700 78.025 79.700 82.025 84.200
> quantile(jugulaire,type=6)
   0%   25%   50%   75%  100%
73.10 75.05 76.05 76.30 76.80
```

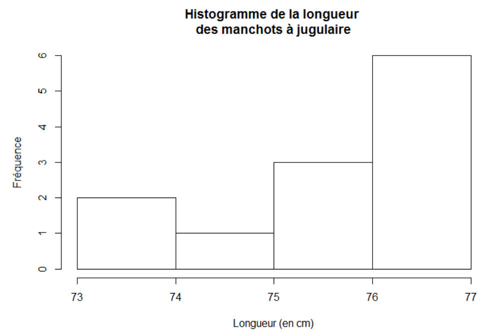
Remarque: J'ai utilisé des quantiles de type 6 avec R. Si vous n'avez utilisé les quantiles de type 6 (qui est acceptable), vous devrez avoir :

```
> quantile(papous)
    0%    25%    50%    75%   100%
75.700 78.075 79.700 81.875 84.200
> quantile(jugulaire)
   0%   25%   50%   75%  100%
73.10 75.35 76.05 76.30 76.80
```

(b) L'histogramme pour la longueur d'un manchot papou est ci-bas. La distribution de la longueur est approximativement symétrique.



L'histogramme pour la longueur d' manchot à jugulaire est ci-bas. La distribution de la longueur a une asymétrie négative.

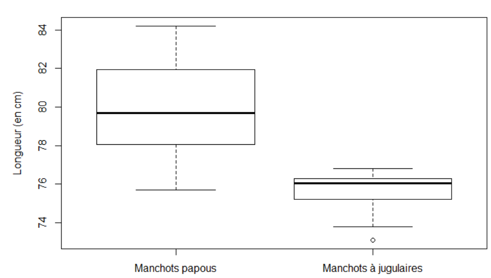


Commandes pour les histogrammes avec `R`:

```
hist(papous,main="Histogramme de la longueur
des manchots papous",ylab="Fréquence",xlab="Longueur (en cm)")
hist(jugulaire,main="Histogramme de la longueur
des manchots à jugulaire",ylab="Fréquence",xlab="Longueur (en cm)")
```

Remarque: C'est correct si vous avez utilisé la fonction `hist` sans les arguments `main`, `xlab` et `ylab`.

(c) Voici les diagrammes à boîte et moustaches comparatifs pour ces deux groupes. Il y a une valeur aberrante dans le groupe des manchots à jugulaire. Il y a un manchot avec une très petite longueur. **Comparaison :** La longueur médiane est plus petite chez les manchots à jugulaire. En outre, la longueur est moins dispersée chez les manchots à jugulaire en comparaison aux manchots papous.



Commande pour le diagrame avec `R`:

```
boxplot(papous,jugulaire,
ylab="Longueur (en cm)",
names=c("Manchots papous","Manchots à jugulaires"))
```

Remarque: C'est correct si vous avez utilisé la fonction `boxplot` sans les arguments `ylab`, `xlab` et `names`.

(3) [**10 points**]Ceci est une transformation linéaire avec un facteur de redimension $a = 9/5$ et un terme de translation $b = -459{,}67$. Seulement le facteur de redimension a un effet sur les mesures de dispersion. Alors, **l'écart type** en degrés F est 1,872 degrés.

Calcul avec R:

```
> 1.04*9/5
[1] 1.872
```

Pour les statistiques dans le sommaire à 5 nombres et la moyenne, il suffit de réexprimer la température en F. On obtient (en degrés F) :

min	q_1	médiane	moyenne	q_3	max
96,71	98,69	99,59	99,77	101,03	102,47

Calcul avec R:

```
> x=c(309.1, 310.2, 310.7, 310.8, 311.5, 312.3)
> x*(9/5)-459.67
[1]  96.71  98.69  99.59  99.77 101.03 102.47
```