Corrigé Test Blanc

Problème 1 Vrai ou Faux (4 points)

Cocher la case qui convient pour répondre Vrai ou Faux.

a) Soit X une variable de distribution normale. Il existe une valeur a telle que $P(X > a) = -0.5080$

☐ VRAI ☐ FAUX

Réponse : Faux. $P(X > a)$ est une probabilité/représente l'aire sous la courbe, qui est nécessairement positive.

b) Soit un échantillon de taille 100, à partir duquel on construit un intervalle de confiance (IC) pour une moyenne. Alors l'IC à 90% est contenu dans l'IC à 95%.

☐ VRAI ☐ FAUX

Réponse : Vrai. Plus le degré de confiance est grand, plus l'IC est grand ($z_{\frac{\alpha}{2}}$ augmente lorsque le risque α diminue). Cas extrême = 100 %

c) Soit une estimation d'une moyenne, faite à partir d'un échantillon de taille 20. L'erreur d'échantillonnage maximale absolue est $\mu - x_{\min})$, où $x_{\min}$ est le plus petit élément de l'échantillon, et μ la moyenne de la population.

☐ VRAI ☐ FAUX

Réponse : Faux. Il se peut qu'elle soit obtenue en prenant le plus grand.

d) Soit un échantillon de taille 100, à partir duquel on construit un IC pour une proportion. Alors un IC à 90% possible est $[-0.056\,; 0.143]$.

☐ VRAI ☐ FAUX

Réponse : Vrai. Une proportion est nécessairement positive, mais l'intervalle peut être plus grand !

Problème 2 Hockey (7 points)

Supposons que le nombre de buts marqués en moyenne par une équipe de hockey sur glace lors des 8 saisons précédentes suive une loi normale de moyenne 100 et d'écart-type 20. Un nouvel attaquant vient d'être engagé moyennant une forte somme d'argent, et l'entraîneur espère que le nombre de buts marqués durant la prochaine saison suivra une loi $\mathcal{N}(120, 400)$.

a) **1 point** Sur quelle valeur devra-t-on effectuer un test ?
Réponse : sur la moyenne de buts marqués par saison

b) **1 point** Formuler les hypothèses nulle et alternative.
Réponse : $H_0 : \quad \mu = 100$ (ou $\mu \leq 100$)
$H_1 : \quad \mu > 100$

c) **1 points** Quelle est la probabilité de commettre une erreur de première espèce, si le seuil de confiance est fixé à 90% ?
Réponse : 0.1.

d) **3 point** Supposons maintenant que la valeur critique soit 110 (buts marqués en moyenne). Quelle est alors la probabilité de ne pas rejeter H_0 si en réalité le nombre de buts marqués suit une loi $\mathcal{N}(120, 400)$?
Réponse : $\beta = P(X < 110 | X \sim \mathcal{N}(120, 400)) = P(z < \frac{110-120}{20}) = 0.3085$. De plus, dans ce cas on a $\beta = \alpha$ (voir p.ex. géométriquement)

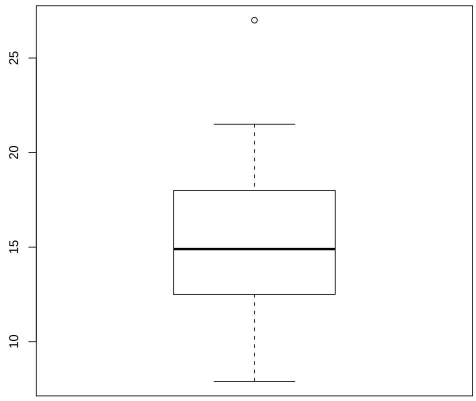
e) **1 point** Supposons ensuite que l'on ait calculé avec un logiciel la p-valeur suivante : p-valeur $= 0.4567$. Quelle va être la conclusion du test ? L'entraîneur a-t-il de quoi se réjouir, ou au contraire doit-il se mordre les doigts ?
Réponse : Comme p-valeur $= 0.4567 > 0.3085 = \alpha$, on ne va pas rejeter H_0. La situation ne va donc pas changer la saison prochaine, et l'argent du transfert a été dépensé pour rien.

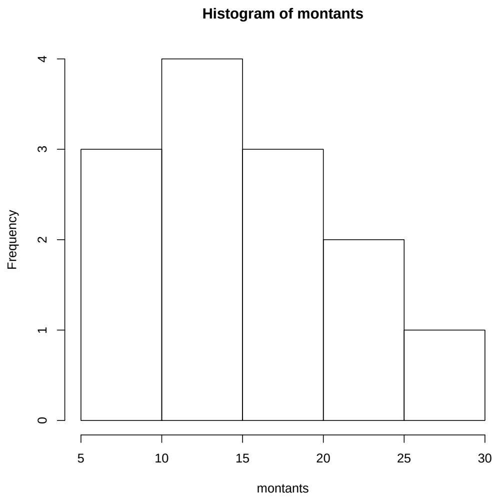
Problème 3 Fast-food (14 points)

Le gérant d'un *fast-food* souhaite estimer le montant moyen dépensé par client lors d'une commande. Pour cela, il a pris un échantillon de 13 commandes et a relevé les montants encaissés suivants :

12.50 21.50 9.90 15.10 13.50 27 16.10 7.90 14.90 9.90 14 18 21.50

a) **4 points** Vérifier si la population semble être distribuée normalement.





Réponse : Dans l'ordre, on a

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$
7.90	9.90	9.90	12.50	13.50	14	14.90	15.10	16.10	18	21.50	21.50	27

Les quartiles sont $q_1 = x_{(4)} = 12.50$, médiane $= q_2 = x_{(7)} = 14.90$, et $q_3 = x_{(10)} = 18$
IQR $= q_3 - q_1 = 18 - 12.5 = 5.5$
L-max moustaches $= 1.5 \cdot 5.5 = 8.25$, et donc moustaches de 7.90 à q_1 et de q_3 à 21.5 (27 = valeur aberrante).
Donc l'exemple construit donne bien un boxplot symétrique autour de la médiane, donc hypothèse de normalité remplie.

```
montants <- c(7.9,9.9,9.9,12.5,13.5,14,14.9,15.1,16.1,18,21.5,21.5,27)
boxplot.stats(montants, coef = 1.5, do.conf = FALSE, do.out = TRUE)
boxplot(montants, coef = 1.5)
hist(montants)
```

b) **4 points** On donne $\sum x_i = 201.8$ et $\sum x_i^2 = 3479.66$. Construire un intervalle de confiance à 90% pour le montant moyen par commande.
Réponse : On calcule $\bar{x} = \frac{201.8}{13} = 15.523$ et $s^2 = \frac{13}{12}(\frac{1}{13} \cdot 3479.66 - (15.523)^2) = 28.927$, d'où $s = 5.378$; avec $t_{0.05;12} = 1.7823$ on a donc
$IC_{90\%} = 15.523 \pm 1.7823 \cdot \frac{5.378}{\sqrt{13}} = [12,86\,;\,18,18]$

```
mean(montants)
sd(montants)
t.test(montants,conf.level=0.9)

        One Sample t-test

data:  montants
t = 10.4066, df = 12, p-value = 2.324e-07
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
```

```
 12.86453 18.18163
sample estimates:
mean of x
 15.52308
```

c) **6 points** Le gérant exige de ses employés qu'ils se chargent de servir les clients le plus rapidement possible tout en fournissant un service de qualité, et dans ce but a mis en place des nouvelles directives. Il estime qu'un temps de service moyen de 3 minutes avec une variation de ± 2 minutes est acceptable.

Laissant de côté la valeur $x_{\max} = 27$ de son échantillon, qu'il considère comme aberrante, le gérant note les différences, en minutes, entre le temps moyen estimé et celui de l'attente effective du client :

$$3 \quad 5 \quad -2 \quad 3 \quad 0 \quad 1 \quad 2 \quad 1 \quad -1 \quad 0 \quad -2.5 \quad -2.5$$

Il calcule ensuite : *médiane* $\tilde{y} = 0.5$; *moyenne* $\bar{y} = 0.58\bar{3}$; *écart type* $s_y = 2.382$

Testez la variabilité de ces temps d'attente avec un degré de confiance de 90%, par rapport aux souhaits du gérant, en supposant que la distribution suit une loi normale. Indiquez toutes les étapes.

```
attente <- c(3,5,-2,3,0,1,2,1,-1,0,-2.5,-2.5)
sigma0 <- 4
pchisq( var(attente)*(length(attente)-1)/sigma0, length(attente)-1, lower.tail= FAL
```

Réponse :

i) **1 point** La valeur de la population d'intérêt est la variation des délais d'attente, transformés en variance.

ii) **1 point** Hypothèses : $H_0 : \quad \sigma^2 \leq 4$
$H_1 : \quad \sigma^2 > 4$

iii) **1 point** Le niveau de signification est fixé à $\alpha = 0.1$

iv) **1 point** La région de rejet du test est l'ensemble des valeurs supérieures à

$$\chi^2_{0.1,11} = 17.275$$

v) **1 point** Comme $s^2 = 5.674$, $\sigma^2 = 4$ et $n - 1 = 11$, la statistique associée à l'échantillon est

$$\chi^2 = \frac{11 \cdot 5.674}{4} = 15.6035$$

vi) **1 point** Comme 15.6035 n'appartient pas à la région critique, l'hypothèse H_0 n'est pas rejetée.

vii) **1 point** Conclusion : les directives ont l'effet voulu, à savoir qu'il est très probable que la variation de temps de service effectif soit tout à fait acceptable aux yeux du gérant.