Attention: Prenez $\alpha = 5\%$ pour tous les tests et intervalles de confiance. $Z_{0,025} = 1,96$ $Z_{0,05} = 1,64$ $t_{0,975;38} = 2,02$ $\chi^2_{1;0.05} = 3,84$.

Question 1 : (10 points)

Complétez le tableau ci-dessous en répondant par Vrai ou Faux pour chacune des affirmations suivantes (dans chacun des cas, fournissez une brève justification) :

1. (**2 pt**) En fiabilité, le taux de défaillance $\lambda(t)$ est une variable aléatoire.

Vrai	
Faux	

Faux, vous tourvez la justification dans le livre.

2. (**2 pt**) Un système avec taux de défaillance $\lambda(t) = \lambda t$ est sans mémoire.

Vrai	
Faux	

Faux, vous tourvez la justification dans le livre.

3. (**2 pt**) Dans un test d'hypothèse on réduit le risque de deuxième espèce lorsqu'on augmente le risque de première espèce.

Vrai	
Faux	

Vrai, vous tourvez la justification dans le livre.

4. (**2 pt**) En statistique descriptive, le diagramme de Tukey (boxplot) ne permet pas toujours de calculer l'étendue des données.

Vrai	
Faux	

Vrai, vous tourvez la justification dans le livre.

5. (**2 pt**) Dans un modèle de régression linéaire simple, si les résidus ne sont pas distribués selon une loi normale, alors le modèle n'est pas significatif.

Vrai	
Faux	

Faux, vous tourvez la justification dans le livre.

Question 2: (30 points)

Vingt informaticiens ont chacun installé soit Linux, soit Windows. Le temps nécessaire (en minutes) à chacun pour l'installation est répertorié dans le tableau suivant.

Linux	126	149	139	145	146	152	162	157	153	139
Windows	160	131	160	150	150	190	164	180	165	127

En considérant que les valeurs sont indépendamment distribuées selon une loi normale de variance $\sigma^2 = 225$.

$\sigma = 15$

a) (**10 pt**) Calculer l'intervalle de confiance de la durée moyenne d'installation de chacun des deux logiciels.
μ_L : (137,50 156,09)
μ_W: (148,4 166,99)

σ^2 connue

293,6

$\overline{W} = 157,7$

$\overline{L} = 146,8$

Linux:

$$146,8 \pm Z_{\alpha/2} \cdot \frac{15}{\sqrt{10}} = 146,8 \pm 9,3 = \mu_L : [137,5 ; 156,1]$$

Windows

$$157,7 \pm 1,96 \frac{15}{\sqrt{10}} = 157,7 \pm 9,3 = \mu_W : [148,4 ; 167]$$

Pour μ_L : on utilise 146,8 · 2

b) (**10 pts**) Un informaticien installe successivement un de ces systèmes d'exploitation pour deux clients A et B. Calculer l'intervalle de confiance dans le cas où

b.1) A et B souhaitent installer Linux.
$X_{AL} + X_{BL} \sim N(2 \times \mu_L, 2 \times 225)$
$X_{AL} + X_{BL}$: (252,02 335,17)

$Z \sim N(2 \cdot \mu_L ; 450)$

$\sigma = 21,21$

$293,6 \pm 1,96 \cdot \frac{21,21}{\sqrt{1}} = [252,02 ; 335,17$

b.2) A et B souhaitent installer Windows.
$X_{AW} + X_{BW} \sim N(2 \times \mu_W, 2 \times 225)$
$X_{AW} + X_{BW}$: (273,82 356,98)

$315,4 \pm 1,96 \cdot \frac{21,21}{1} = [273,82 ; 356,97]$

b.3) A souhaite installer Linux et B souhaite installer Windows.
$X_{AL} + X_{BW} \sim N(\mu_L + \mu_W, 2 \times 225)$
$X_{AL} + X_{BW}$: (262,92 346,07)

$304,5 \pm 1,96 \cdot \frac{21,21}{\sqrt{1}} = [262,93 ; 346,0716]$

Question 2 (suite)

c) (**10 pt**) Déterminer par un test statistique si l'installation de Linux est significativement plus rapide que l'installation de Windows.

$$H_0 : \mu_L \leq \mu_W$$

$$H_1 : \mu_L > \mu_W$$

~~R. H_0 si $Z_0 = \frac{\overline{X}_L - \overline{X}_W}{\sqrt{225 \times (1/10 + 1/10)}} > 1.64$~~

~~$Z_0 = -1,62$ alors on accept H_0~~

$$T_0 = \frac{\overline{L} - \overline{W}}{S_p \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$S_p^2 = \frac{9(S_L^2) + 9(S_W^2)}{18}$$

$$= \frac{9 \cdot 225 + 9 \cdot 225}{18}$$

$$= 225$$

$$S_p = \sqrt{225} = 15$$

$$T_0 = \frac{-110,9}{6,71}$$

$$= -16,53$$

rejette H_0 si

$-16,53 > t_{0,05;18}$ 1,764

Faux

On ne rejette pas

[illegible]!

Question 3: (20 points)

Y a-t-il un lien entre l'intelligence et la taille du cerveau ? Une étude de Willerman et al. (1991) a étudié cette question en mesurant le QI et la taille du cerveau (TC par résonance magnétique) de 40 individus. On suppose le modèle de régression

$$\mathrm{QI}_i = \alpha + \beta \times \mathrm{TC}_i + \varepsilon_i, i = 1, \ldots, 40.$$

Sachant que l'estimation par la méthode des moindres carrés donne $\hat{\alpha} = 1,74$, $\hat{\beta} = 0,00012$, $\hat{\sigma} = 20,99$, $\sqrt{S_{xx}} = 457152$ et $\overline{x} = 9,5 \times 10^5$. On suppose que les ε_i sont indépendamment distribuées selon une loi normale.

a) (**10 pts**) La taille du cerveau de Samuel est 10^6. Donner l'estimation de son QI et son intervalle de confiance.
$QI = 1,74 + 0,00012 \times 10^6 = 121,74$
$QI \pm t_{n-2;\alpha/2}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}} = (79,85 \quad 163,63).$

$\hat{QI} = \alpha + \beta \hat{T}$
$= 1,74 + 0,00012 \cdot 10^6$
$= 121,74$

$121,74 \pm$

Question 3 (suite)

b) (**10 pts**) Le QI dépend-il de la taille du cerveau ?
oui $H_0 : \beta = 0$
$H_1 : \beta \neq 0$
$T_0 = \frac{\hat{\beta}-0}{\hat{\sigma}/\sqrt{S_{xx}}} = 2.61 > 1.96$

Question 4: (30 points)

Au total 50936 enfants sont nés aux Etats Unis entre 1998 et 2002. Le tableau suivant contient les données sur la séquence des sexes des enfants nés entre 1998 et 2002 dans les familles avec exactement deux enfants.

Fille-Fille	5 844
Garçon-Fille	6 628
Fille-Garçon	6 451
Garçon-Garçon	6 545
Total	25 468

a) (**10 pt**) Donner une estimation $\hat{p}$ de la probabilité qu'un nouveau né soit une fille.

$\hat{p} = \frac{5844 \times 2 + 6628 + 6451}{50936} = 0,4862$

$$\frac{2 \cdot 5844 + 6628 + 6451}{50\,936} = 48,6\%$$

Question 4 (suite)

b) (**30 pt**) Si le sexe du premier enfant est indépendant du sexe du deuxième enfant, donner l'estimation de la distribution du nombre de filles pour une famille de trois enfants.
$X = \text{Bin}(3, \hat{p})$
$\Pr(X = 0) = (1 - \hat{p})^3 = 0,13564$
$\Pr(X = 1) = 3 \times \hat{p}(1 - \hat{p})^2 = 0,3851$
$\Pr(X = 2) = 3 \times \hat{p}^2(1 - \hat{p}) = 0,3644$
$\Pr(X = 3) = \hat{p}^3 = 0,1150$

X : nb de filles pour une famille de 3 enfants

$X \sim \text{Bin}(3, ,4468)$

Question 4 (suite)

c) (**10 pts**) Tester si le sexe du premier enfant est indépendant du sexe du deuxième enfant.

[redacted]

[redacted]

[redacted]

Question 5 : (10 points)

Le système d'exploitation (OS) 32 bits est capable de faire le calcul sur une variable x_i, si $-10^{10} < x_i < 10^{10}$. Vous avez une taille d'échantillon $n = 10^6$ et les valeurs de x_i sont positives et autour de 10^8, et en plus $\max(\mathrm{x_i}) < 10^{16}$. Comment pouvez vous calculer la moyenne d'échantillon $\overline{X}$?

$\overline{X} = \frac{1}{10^6} \sum_{i=1}^{10^6} X_i = \sum_{i=1}^{10^6} \frac{1}{10^6} X_i$