POLYTECHNIQUE MONTRÉAL

Questionnaire
Examen final

MTH2302D

Sigle du cours

Identification de l'étudiant(e)		
Nom : Nollet	**Prénom :** Charles	
Signature : Charles Nollet	**Matricule :** 1848802	**Groupe :** 03

Réservé	
1.	1 /6
2.	5.5 /6
3.	4.75 /8
4.	11,5 /12
5.	11,75 /14
6.	2 /4
TOTAL	36,50 /50

Sigle et titre du cours	*Groupe*	*Trimestre*
MTH2302D Probabilités et statistique	**TOUS**	**Automne 2017**
Professeur	***Local***	***Téléphone***
Luc Adjengue	**A-520.33**	**4475**

Jour	*Date*	*Durée*	*Heures*
Mercredi	**20 décembre 2017**	**2h30**	**9h30 à 12h00**

Documentation	*Calculatrice*	
☐ Aucune	☐ Aucune	**Les cellulaires, agendas électroniques ou téléavertisseurs sont interdits.**
☐ Toute	☐ Toutes	
☒ Voir directives particulières	☒ Non programmable	

Directives particulières

1. Ne détachez aucune feuille de ce questionnaire.
2. Les 3 dernières pages constituent une annexe (ne pas détacher).
3. Documentation permise : **une feuille 8,5x11 recto-verso.**
4. Chaque réponse doit être complète et accompagnée d'une justification.

Par souci d'équité envers tous les étudiants, le professeur ne répondra à aucune question durant cet examen. Si vous estimez que vous ne pouvez pas répondre à une question pour diverses raisons (données manquantes, données erronées, etc.), veuillez le justifier (maximum 2 lignes) puis passez à la question suivante.

Important

Cet examen contient **6** questions sur un total de **16** pages
(excluant cette page et l'annexe)

La pondération de cet examen est de **50 %**

Vous devez répondre sur : ☒ le questionnaire ☐ le cahier ☐ les deux

Vous devez remettre le questionnaire : ☒ oui ☐ non

L'étudiant doit honorer l'engagement pris lors de la signature du code de conduite.

Question n° 1 : (6 points)

1

Soit $X_1, X_2, \ldots, X_{100}$ un échantillon aléatoire d'une population X de loi normale de moyenne $\mu = 1$ et de variance $\sigma^2 = 1$, c'est-à-dire $X \sim N(1,\ 1)$.

a) Calculer $P\left(1 \leq \overline{X} \leq 1{,}05\right)$.

1.00

$$P\left(\frac{1-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{1{,}05-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P(0 \leq Z \leq 0{,}5) = P(Z \leq 0{,}5) - P(Z \leq 0)$$

$$= \phi(0{,}5) - \phi(0)$$

$$= 0{,}69146 - 0{,}5 =$$

$$P(1 \leq \bar{X} \leq 1{,}05) = 0{,}19146$$

b) On définit deux variables T et U par

$$T = \sum_{i=1}^{6} X_i \quad \text{et} \quad U = \sum_{i=7}^{9} X_i.$$

Calculer $P(T > U)$.

$$P(T > U) = P(\mu_T > \mu_U)$$

$$\mu_T = \frac{\sum^{6} x_i}{6} \qquad \mu_U = \frac{\sum_{7} x_i}{3}$$

Question n° 1 : (suite)

c) On définit $Y = \sum_{i=1}^{100}(X_i - 1)^2$.

1.c) Donner le nom de la loi exacte de Y ainsi que les valeurs de ses paramètres.

2.c) Donner le nom de la loi approximative de Y selon le théorème central limite ainsi que les valeurs de ses paramètres.

Admettons que $Y \sim U(-0,5\,;\,0,5)$ car je ne sais pas c'est quoi la loi exacte

Attention à ce type de rédaction

$E(Y_i) = 0 = \frac{a+b}{2}$

$\sigma^2(Y_i) = \frac{1}{12}$ $n = 100$

$X \sim N(E(x)\,;\,V(x))$ $E(X) = \sum^{100} EY_i = 100 \cdot 0 = 0$

$\sigma^2(X) = \text{var}\left(\sum^{100} Y_i\right) = \frac{100}{12} = \frac{25}{3}$

$X \sim N\left(0, \frac{25}{3}\right)$

d) On définit deux variables G et H par

$$G = \sum_{i=1}^{6} (X_i - 1)^2 \quad \text{et} \quad H = \sum_{i=7}^{9} (X_i - 1)^2.$$

Déterminer la valeur de la constante m telle que $P(G > mH) = 0{,}01$.

Question n° 2 : (6 points)

Le tableau ci-dessous présente la distribution de fréquence du nombre de défauts X sur le fini d'une pièce observés dans un échantillon aléatoire de 100 pièces d'une production.

Nombre de défauts (x_i)	0	1	2	3
Nombre de pièces (O_i)	15	45	25	15

En utilisant le test d'ajustement du Khi-deux, on veut tester l'hypothèse selon laquelle la variable X est distribuée selon une loi de Poisson.

a) (1 point) Formuler les hypothèses H_0 et H_1 du test à effectuer.

$H_0 : X \sim Poi(c)$
$H_1 : X \not\sim Poi(c)$ ✓

b) (5 points) Effectuer le test et conclure au seuil $\alpha = 5\%$.

Rappel : si $X \sim$ Poisson(c), alors $P(X = x) = \dfrac{c^x e^{-c}}{x!}$, pour $x = 0, 1, \ldots$

On rejette H_0 si $\chi_0^2 > \chi^2_{K-p-1}$

somme des pieces $= n = 100$

si $X \sim Poi(c)$ alors $E(X) = c$.

$\bar{X} = \dfrac{15 \cdot 0 + 45 \cdot 1 + 25 \cdot 2 + 15 \cdot 3}{100} = 1,4$ ✓

$c = \bar{X} = 1,4$.

$$\chi_0^2 = \sum^4 \frac{(O_i - E_i)^2}{E_i}$$

$P(x=0) = \dfrac{1,4^0 \cdot e^{-1,4}}{0!} = 0,2466$
$P(x=1) = 0,3452$
$P(x=2) = 0,2417$
$P(x=3) = 0,1128$ ✓

ND	0	1	2	3
NP	15	45	25	15
p_i	0,2466	0,3452	0,2417	0,1128
E_i	24,66	34,52	24,17	11,28 ✓

$\chi_0^2 = \cancel{7,33}$ -0,5

$\chi^2_{K-p-1} = \chi^2_{4-1-1} = \chi^2_2 = 5,99$ ✓

On rejette H_0 car $\chi_0^2 > \chi^2_{K-p-1}$ ✓

$X \not\sim Poi(c)$ ✓

Question n° 2 : (suite)

R

Question n° 3 : (8 points) 4,75

Le service du contrôle de la qualité veut étudier (estimer et tester) la proportion actuelle d'unités non conformes d'une production. Pour cette étude, le service compte utiliser un échantillon de n unités prélevées au hasard de la production.

a) (2 points) Quelle doit être la plus petite valeur de n pour que, avec un niveau de confiance de 95%, l'erreur d'estimation soit inférieure à 0,045 ?

$\alpha = 0{,}05$ erreur d'estimation $= |\bar{x} - \mu| \leq 0{,}045$

Erreur max $= 0{,}045 = \dfrac{z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}}$

1,5

$n \geq \left(\dfrac{z_{\frac{\alpha}{2}} \cdot \sigma}{0{,}045}\right)^2 = \left(\dfrac{1{,}96}{0{,}045}\right)^2 \cdot \sigma^2 = 1897 \cdot \sigma^2$

Dans le passé, la proportion d'unités non conformes de la production était de 10%. Le service du contrôle de la qualité veut effectuer un test statistique, de seuil $\alpha = 5\%$, afin de vérifier si cette proportion est à présent plus élevée.

b) (1 point) Définir le paramètre d'intérêt et formuler les hypothèses H_0 et H_1 du test à effectuer.

$H_0 : p = 0{,}10$ ~~paramètre d'intérêt : n~~

0,75

$H_1 : p > 0{,}10$

Question nº 3 : (suite)

c) **(3 points)** Avec un échantillon de 400 unités prélevées de la production, on obtient une valeur-p (ou «*p-value*») de 0,0024 pour le test des hypothèses définies en b).

- Quelle doit être la conclusion du test ?

On rejette H_0 car p-value $< \alpha$

Z_0 a de faibles probabilités d'être $< Z$

1.5

- Quel est le nombre d'unités non conformes observées dans cet échantillon ?

Il y a plus de 40 unités non conformes car la proportion d'unités non conforme est plus grande que 10%.

d) **(2 points)** Supposons que la proportion réelle d'unités non conformes soit actuellement de 11% et qu'on utilise un échantillon aléatoire de 500 unités pour le test. Quelle serait alors la probabilité de conclure à tort que la proportion d'unités non conformes actuelle n'est pas plus élevée que dans le passé ?

$$P\left(\text{accepter } H_0 / H_0 \text{ est fausse}\right) = \beta$$

$$\beta = \phi\left(\frac{P_0 - P - Z_{\frac{\alpha}{2}}\sqrt{\frac{P_0(1-P_0)}{n}}}{\sqrt{\frac{P(1-P)}{n}}}\right) = \phi\left(\frac{0,1 - 0,11 - 1,96 \cdot 0,01342}{0,01399}\right)$$

$$= \phi(-2,5944) = 1 - \phi(2,5944)$$

$$= 1 - 0,9952$$

$$= \boxed{0,0048}$$

1

Question nº 4 : (12 points) 11,5

Dans un réseau informatique, un ingénieur mesure les délais de communication (en millisecondes) entre deux ordinateurs avec un protocole standard (population X_1) et avec un protocole expérimental (population X_2). L'ingénieur prélève un échantillon aléatoire de taille $n_1 = 16$ de la population X_1 et un échantillon aléatoire de taille $n_2 = 9$ de la population X_2. On suppose que les deux échantillons sont indépendants. Les modèles utilisés et les résultats obtenus sont décrits dans le tableau suivant.

Population X_i	Taille n_i	Moyenne $\bar{x}_i$	Variance s_i^2
$X_1 \sim N(\mu_1,\ \sigma_1^2)$	16	52,4	0,37
$X_2 \sim N(\mu_2,\ \sigma_2^2)$	9	50,3	2,18

***Important** : formuler les hypothèses H_0 et H_1 pour chaque test.*

a) **(2 points)** Donner un intervalle de confiance de niveau 90% pour le paramètre μ_2.

2

$X_2 \sim N(\mu_2, \sigma_2^2)$: $\bar{X}_2 = 50,3$ $\quad S_2^2 = 2,18$ $\quad n_2 = 9$

$S_2 = 1,48$ $\quad \sqrt{n_2} = 3$

$$IDC(\mu_2) = \bar{X} \pm t_{\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}}$$

$$50,3 \pm 1,86 \cdot \frac{1,48}{3}$$

$$[49,3824 ; 51,2176]$$

Question n° 4 : (suite)

b) (2 points) Donner un intervalle de prévision de niveau 90% pour une dixième mesure de délai du protocole expérimental.

1,8

$\hookrightarrow x_2$

$$X_{n+1} \in \bar{X} \pm t_{n-1}(\tfrac{\alpha}{2}) \sqrt{S^2(1+\tfrac{1}{n})}$$

$$X_{n+1} \in 50,3 \pm 1,86 \cdot \cancel{2,42}$$

$$X_{n+1} \in [45,795 \; ; 54,805]$$

c) (2 points) Peut-on rejeter l'affirmation selon laquelle $\sigma_1^2 = 0{,}5$? Donner la conclusion d'un test d'hypothèse au seuil $\alpha = 0{,}05$.

2

$H_0 : \sigma_1^2 = 0,5$

$H_1 : \sigma_1^2 \neq 0,5$

On rejette H_0 si. $\chi_0^2 > \chi^2_{\frac{\alpha}{2};n-1}$ ou $\chi_0^2 < \chi^2_{1-\frac{\alpha}{2};n-1}$

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(15)\cdot 0,37}{0,5} = 11,1$$

$\chi^2_{n-1}(\frac{\alpha}{2}) = 27,49$ $\qquad \chi_0^2 < 27,49$

$\chi^2_{n-1}(1-\frac{\alpha}{2}) = 6,27$. $\qquad \chi_0^2 > 6,27$.

donc Non on ne peut pas rejeter H_0

Question nº 4 : (suite)

d) **(3 points)** Peut-on dire que les variances σ_1^2 et σ_2^2 sont différentes ? Donner la conclusion d'un test d'hypothèse au seuil $\alpha = 0{,}05$.

3

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$F_0 = \frac{S_1^2}{S_2^2}$

on rejette H_0 si $F_0 < F_{1-\frac{\alpha}{2};\, n_1-1;\, n_2-1}$ ou $F_0 > F_{\frac{\alpha}{2};\, n_1-1;\, n_2-1}$.

$F_0 = \frac{0{,}37}{2{,}18} = 0{,}1697$

$F_{1-\frac{\alpha}{2};\, n_1-1;\, n_2-1} = F_{0{,}975;\,15;\,8} = \frac{1}{F_{0{,}025;\,8;\,15}} = \frac{1}{3{,}20} = 0{,}3125$

$F_0 < 0{,}3125$
On rejette H_0.

$F_{\frac{\alpha}{2};\, n_1-1;\, n_2-1} = F_{0{,}025;\,15;\,8} = 4{,}10$.

Oui on peut dire que σ_1^2 et σ_2^2 sont différentes car on a rejeté l'hypothèse qu'elles sont égales.

Question n° 4 : (suite)

e) **(3 points)** Peut-on affirmer qu'en moyenne les délais du protocole expérimental sont plus courts que ceux du protocole standard ? Donner la conclusion d'un test d'hypothèse au seuil $\alpha = 0{,}05$.

$\sigma_1^2 \neq \sigma_2^2$ 3

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_2 < \mu_1$ donc $\mu_1 > \mu_2$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$v = \left(\frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} \right) - 2$$

On rejette Ho si

$T_0 > t_{\alpha ; v}$

$$T_0 = \frac{52{,}4 - 50{,}3}{\sqrt{\frac{0{,}37}{16} + \frac{2{,}18}{9}}} = 4{,}07673$$

$v = 9{,}937$

$t_{\alpha ; 9{,}937} \approx t_{0{,}05 ; 10} = 1{,}81$

On rejette Ho car $4{,}077 > 1{,}81$

Donc oui on peut affirmer qu'en moyenne les délais du protocole expérimental sont plus courts que ceux du protocole standard.

Question nº 5 : (14 points)

Pour des raisons de santé publique, on cherche à déterminer le lien entre la concentration d'ozone O_3 dans l'air d'une journée et la température à midi de la même journée. Pour ce faire, on envisage un modèle de régression linéaire simple d'équation

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

où X représente la température à midi (en °C) et Y, la concentration d'ozone O_3 dans l'air (en microgrammes par millilitre), β_0 et β_1 sont des paramètres et ε, une erreur aléatoire que l'on suppose de loi $N(0,\ \sigma^2)$.

Les données recueillies pour un échantillon de 12 jours sont présentées dans le tableau suivant.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	10,0	23,8	16,3	27,2	7,1	25,1	27,5	19,4	15,0	19,8	32,2	20,7
y_i	71,3	115,4	76,8	113,8	81,6	115,4	125,0	83,6	84,5	75,2	136,8	102,8

L'ajustement du modèle aux 12 observations a permis d'obtenir les résultats partiels suivants :

$$\sum_{i=1}^{12} x_i = 244{,}10\ ; \quad \sum_{i=1}^{12} x_i^2 = 5\,567{,}37\ ; \quad \sum_{i=1}^{12} x_i y_i = 25\,631{,}34\ ;$$
$$\sum_{i=1}^{12} y_i = 1\,182{,}20\ ; \quad \sum_{i=1}^{12} y_i^2 = 121\,916{,}60.$$

a) **(4 points)** Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$. Donner l'équation de la droite des moindres carrés.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\bar{X} = \frac{\sum x_i}{12} = 20{,}34$$
$$\bar{Y} = \frac{\sum y_i}{12} = 98{,}517$$

$$= \frac{1585{,}31}{602{,}78} = 2{,}63$$

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X} = 98{,}517 - 2{,}63 \cdot 20{,}34 = 45{,}02$$

$$Y = 45{,}02 + 2{,}63X + \varepsilon$$

Question nº 5 : (suite)

b) **(3 points)** Compléter le tableau d'analyse de la variance ci-dessous. Quelle conclusion peut-on en tirer au seuil $\alpha = 0{,}05$?

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	F_0
Régression	4169,37	1	4169,37	32,552
Erreur	1280,83	10	128,083	
Totale	5 450,197	11		

$SS_R = \hat{\beta}_1 \cdot S_{xy} = \hat{\beta}_1 \cdot \sum x_i y_i - n\bar{x}\bar{y} = 2{,}63 \cdot 25631{,}34 - 12 \cdot 20{,}34 \cdot 98{,}517 =$
$4\,169{,}37$

$F_0 = \frac{MS_R}{MS_E}$

$SS_E = SS_T - SS_R = 1280{,}83$

$H_0 : \beta_1 = 0$ $\quad F_{\alpha;1;n-2} = F_{0{,}05;1;10} = 4{,}67$

$H_1 : \beta_1 \neq 0$ $\quad F_0 > 4{,}67$ donc on rejette H_0

-0,75 hyp

On peut en conclure que la regression n'est pas impossible

c) **(2 points)** Donner une estimation ponctuelle de la variance σ^2.

$S^2 = m_2' - (m_1')^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = 50{,}3$

$\frac{0}{2}$

Question n° 5 : (suite)

d) **(3 points)** À un niveau de 95%, donner un intervalle de prévision pour la concentration d'ozone O_3 d'une journée dont la température à midi est de 20 °C.

$$IP(20) \in \left(\hat{Y} \pm t_{\frac{\alpha}{2};n-2} \cdot \sqrt{MS_E\left[1+\frac{1}{n}+\frac{(20-\bar{x})^2}{S_{xx}}\right]}\right)$$

$$\hat{Y} = 45,02 + 2,63 \cdot 20 = 97,62$$

$$t_{\frac{\alpha}{2};n-2} = 2,23$$

$$\sqrt{MS_E\left[1+\frac{1}{n}+\frac{(20-\bar{x})^2}{S_{xx}}\right]} = \sqrt{128,083\left(1+\frac{1}{12}+\frac{-0,34}{602,78}\right)} = 11,77$$

$$IP(20) \in [71,37 \; ; \; 123,87]$$

e) **(2 points)** Calculer le coefficient de détermination R^2 et interpréter ce résultat.

$$R^2 = \frac{SS_R}{SS_T} = \frac{4169,37}{5450,197} = 0,765$$

R^2 varie entre 0 et 1 et désigne à quel point fiable notre régression est.

Ici avec $R^2 = 0,765$ notre régression est relativement bonne

Question n° 6 : (4 points)

2

Cette question est constituée de deux parties a) et b) indépendantes l'une de l'autre.

a) **(2 points)** Un système est constitué de deux composants placés en série qui fonctionnent indépendamment l'un de l'autre. De plus, un troisième composant est utilisé en redondance passive, précisément :

- il est mis en fonction pour remplacer le premier composant actif qui tombe en panne et,
- il fonctionne indépendamment du composant encore en fonction.

On suppose que la durée de vie de chacun des trois composants est distribuée selon une loi exponentielle Exp(2).

1.a) Déterminer la durée de vie moyenne du système.

0

$\lambda = 2 \quad \mu = 0,5$

$\tau = \sum^{2} R_i(t)$

circuit en série → $R_1(t) = \exp\left(-\int 2e^{-2t}\,dx\right)$

$\exp\left(-2\int e^{-2t}\,dt\right)$

$\exp\left(2\lambda e^{-2t}\right)$

$\exp\left(4e^{-2t}\right)$

circuit en parallèle → $R_2(t) = 1 - \sum^{k-1} C_i^2 \cdot R_i(t)^i\,[1 - R_i(t)]^{n-i}$

2.a) Donner le nom et les valeurs des paramètres de la loi de probabilité qui modélise la durée de vie du système.

0

Question n° 6 : (suite)

b) **(2 points)** Un commerce possède deux caisses modélisées par une file d'attente $M/M/1$. La première caisse sert à payer les achats : les clients s'y présentent au taux $\lambda_1 = 45$ clients/heure et ils sont servis au taux $\mu_1 = 50$ clients/heure. La deuxième caisse sert à traiter les retours et échanges : les clients s'y présentent au taux $\lambda_2 = 4$ clients/heure et ils sont servis au taux $\mu_2 = 5$ clients/heure. On suppose que les deux files fonctionnent indépendamment l'une de l'autre.

1.b) Déterminer le nombre moyen de clients dans le système (aux deux caisses) à l'équilibre.

$\lambda_1 = 45 \qquad \mu_1 = 50$

$\lambda_2 = 4 \qquad \mu_2 = 5$

$$\bar{N} = \bar{N}_Q + \bar{N}_S = \sum^2 \frac{\varphi}{1-\varphi} = \sum^2 \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} = \frac{\frac{\lambda_1}{\mu_1}}{1-\frac{\lambda_1}{\mu_1}} + \frac{\frac{\lambda_2}{\mu_2}}{1-\frac{\lambda_2}{\mu_2}} = 9 + 4 = \boxed{13}$$

2.b) Déterminer la probabilité que le système soit vide à l'équilibre.

$P_1(X=0) = (1-\varphi)\cdot\varphi^n = 0,1$

$P_2(X=0) = (1-\varphi)\cdot\varphi^n = 0,2$

$P_1(0)\cdot P_2(0) = P(0) = 0,02$