POLYTECHNIQUE MONTRÉAL

Questionnaire
Examen final

MTH2302D

Sigle du cours

COPIE

Identification de l'étudiant(e)		
Nom :	Prénom :	
Signature :	Matricule :	Groupe :

Réservé	
1.	0.5 /6
2.	4 /6
3.	2.25 /8
4.	6 /12
5.	10,50 /14
6.	1,5 /4
TOTAL	24.75 /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)

0.5

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 $X \sim N(1,\ 1)$.

$z = \frac{\text{borne} - \mu}{\sigma}$

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

0.5

$P(1 \leq x_i \leq 1{,}05) = P(x_i \leq 1{,}05) - P(x \leq 1) = \Phi\left(\frac{1{,}05-1}{1}\right) - \Phi\left(\frac{1-1}{1}\right) = \Phi(0{,}05) - 1/2$

$= 0{,}51994 - 0{,}5 = 0{,}01994$

$\bar{x} = \frac{x_1 + x_2 + \ldots + x_{100}}{100}$, $P(1 \leq \bar{x} \leq 1{,}05) = \boxed{0{,}01994}$

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 $P(T > U)$.

$T = \underbrace{x_1 + x_2 + \ldots + x_6}_{6}$ $\qquad$ $U = \underbrace{x_7 + x_{14} + \ldots + x_{63}}_{9}$

$P(T > U) = 0{,}5$ car x_i provient toujours de la même population ...

$\boxed{0{,}5}$

Question nº 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.

$X \sim N(1,1) \quad \mu = 1$

$Y = \sum_{i=1}^{100}(x_i - 1)^2 = (x_i - \underbrace{1}_{\mu})^2 + (x_2 - 1)^2 + \ldots = \sum_{i=1}^{100} \sigma_i^2$

Cela correspont à la loi de student

$Y \sim t_{1;1}$

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

La loi de Student

$\alpha = 1$

$k = 1$

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 $P(G > mH) = 0{,}01$.

$X \sim N(1,1)$

si on a $P(G > H) = 0{,}5$ (→ en b)) alors

$\frac{1}{m} = \frac{0{,}01}{0{,}5} = 0{,}02 \Rightarrow m = 50$

Question nº 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: $X \sim \text{Poisson}(c)$ $\quad X \sim \text{Poi}(1)$

H_1: $X \not\sim \text{Poisson}(c)$ ✓ $\quad X \not\sim \text{Poi}(1)$ } avec ~~c = 1~~

(↳ "ne suit pas")

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

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

$i = \{0, 1, 2, 3\}$

$E_0 = P(x=0) \cdot 100 \text{ pièces} = \dfrac{1^0 \cdot e^{-1}}{0!} \cdot 100 = e^{-1} \cdot 100 \approx \cancel{36}$

$E_1 = P(x=1) \cdot 100 \text{ pièces} = \dfrac{1^1 \cdot e^{-1}}{1!} \cdot 100 = e^{-1} \cdot 100 \approx \cancel{36}$

$E_2 = P(x=2) \cdot 100 \text{ pièces} = \dfrac{1^2 e^{-1}}{2!} \cdot 100 \approx \cancel{18}$

$E_3 = P(x=3) \cdot 100 \text{ pièces} = \dfrac{1^3 e^{-1}}{3!} \cdot 100 \approx \cancel{6}$ (-1

x_i	0	1	2	3
O_i	15	45	25	15
E_i	36	36	18	6

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

0. $\dfrac{(O_1 - E_1)^2}{E_1} = \dfrac{(15-36)^2}{36} = 12{,}25$

1. $\dfrac{(O_2 - E_2)^2}{E_2} = \dfrac{(45-36)^2}{36} = 2{,}25$

2. $\dfrac{(25-18)^2}{18} = 2{,}72$

3. $13{,}5$

$\chi_0^2 = 12{,}25 + 2{,}25 + 2{,}72 + 13{,}5 = \cancel{30{,}72}$

Question n° 2 : (suite)

$\chi^2_{\alpha,\nu}$ $\alpha = 5\% = 0,05$

$\nu = k - p - 1 = 4_{\text{classes}} - 1_{\text{paramètre [illegible]}} - 1 = 2$ ✓

$\chi^2_{0,05;\,2} = 5,99$ (table) ✓

puisque $\chi_0^2 > \chi^2_{0,05;\,2}$, on doit rejeter H_0 ✓

X ne suit pas une loi de poisson(1) pour un $\alpha = 5\%$. ✓

R

Question n° 3 : (8 points)

2,25

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 ?

$1-\alpha = 95\% \Rightarrow \boxed{97{,}5\%}$ $\qquad z_{\alpha/2} \Rightarrow \Phi(z_{\alpha/2}) \qquad z_{\frac{\alpha}{2}} = 1{,}96 \qquad \Phi(1{,}96) = 0{,}975$

$\frac{\alpha}{2} = 2{,}5\%$
$= 0{,}025$

$E < 0{,}045$

$$n = \left(\frac{z_{\alpha/2}}{2E}\right)^2 = \left\lceil\left(\frac{1{,}96}{2 \cdot 0{,}045}\right)^2\right\rceil = 475$$

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 : la proportion p est plus élevée que 10%. ~~($p > 0.1$)~~

H_1 : la proportion p n'est pas plus élevée que 10%. ~~($p \leq 0.1$)~~

le paramètre d'intérêt est la proportion p.

0,25

Question n° 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 ?

puisque $pv = 0{,}0024 < 0{,}05 = \alpha$, on peut accepter H_0.

rejet de H_0

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

proportion 10% $= \frac{40}{400}$ puisque le p-valeur est assez bas, la quantité d'unités non conformes doit être pas mal plus haute que $\frac{40}{400}$.

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 ?

$\beta(11\%) = \Phi\Big(z_\alpha +$

Question n° 4 : (12 points)

6

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.

1 −

$\Phi(z_{\alpha/2}) = 95\% \Rightarrow z_{\alpha/2} = 1,65$

$\cancel{z_{\alpha/2}} \frac{\sigma}{n} = \frac{1,65 \cdot \sqrt{2,18}}{9} = 0,2707$

$\bar{x} + 0,2707 = 50,5707$

$\bar{x} - 0,2707 = 49,7293$

intervalle de confiance de niveau 90% : $[50,57 ; 49,73]$

Question n^o 4 : (suite)

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

$Z_{\alpha/2}$ est pareil

1

$\bar{X} \pm \cancel{Z_{\alpha/2}} \sqrt{\sigma^2(1+\frac{1}{n})} = 1{,}65 \cdot \sqrt{2.18(1+\frac{1}{9})}$

$= 2{,}57$

$\bar{x} + 2{,}57 = 52{,}57$

$\bar{x} - 2{,}57 = 47{,}43$

intervalle de prévision : $[47{,}43;\ 52{,}57]$

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 $\alpha = 0{,}05$.

$\to \sigma_0^2 = 0{,}5$

1,5

$\boxed{\begin{array}{l} H_0 : \sigma_1^2 = 0{,}5 \\ H_1 : \sigma_1^2 \neq 0{,}5 \end{array}}$

$n = 16 < 40 \Rightarrow \chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$

$\chi_0^2 = \frac{(16-1)\,0{,}37}{0{,}5} = 11{,}1$

$\chi^2_{1-\alpha/2;n-1} = \chi^2_{0{,}975;15} = 6{,}27$

$\to 6{,}27 < 11{,}1$

$\chi_0^2 > \chi^2_{1-\alpha/2;n-1}$

$\chi^2_{\alpha/2;n-1}$

on ne peut donc pas rejeter $H_0 : \sigma_1^2 = 0{,}5$

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 $\alpha = 0{,}05$.

2

$$\boxed{\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}}$$

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{0{,}37}{2.18} = 0{,}17$$

H_1 vraie si $F_0 < F_{1-\alpha;(n_1-1);n_2-1} \Rightarrow F_{0.025;16-1;9-1} = F_{0.025;15;8} = 4{,}10$

$0{,}17 < 4{,}10$ $F_{0{,}975;15,8}$?

donc oui, on peut dire que H_1 est vraie $(\sigma_1^2 \neq \sigma_2^2)$

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 $\alpha = 0{,}05$.

0,5

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$ $\Rightarrow Z_0 = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ et on veut savoir si $Z_0 >$

$$Z_0 = \frac{52{,}4 - 50{,}3}{\sqrt{\frac{0{,}37}{16} + \frac{2{,}18}{9}}} = 4{,}077$$

$Z_\alpha \Rightarrow \Phi(z_\alpha) = 0{,}05 \Rightarrow \Phi(-z_\alpha) = 1 - 0{,}05 \Rightarrow -z_\alpha = 1{,}65$

$z_\alpha = -1{,}65$

$4{,}077 > \cancel{-1{,}65}$

$Z_0 > Z_\alpha$

On peut donc accepter H_1 qui est que $\mu_1 > \mu_2$

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 ^oC) 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.

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ $\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

$\bar{x} = \frac{1}{n}\sum x_i = \frac{244{,}10}{12} = 20{,}34$

$\bar{y} = \frac{1}{n}\sum y_i = \frac{1182{,}2}{12} = 98{,}52$

$S_{xy} = 25\,631{,}34 - 12 \cdot 20{,}34 \cdot 98{,}52 = 1585{,}39$

$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 5\,567{,}37 - 12 \cdot 20{,}34^2 = 602{,}78$

$\hat{\beta}_1 = \frac{1585{,}39}{602{,}78} = \boxed{2{,}63}$

$\hat{\beta}_0 = \bar{y} - 2{,}63\,\bar{x}$

$= 98{,}52 - 2{,}63 \cdot 20{,}34$

$= \boxed{45{,}02}$

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

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 $\alpha = 0{,}05$?

H_0 : la quantité d'ozone varie avec la te[mpérature] linéairement
H_1 : pas H_0 : ne varient pas ensemble

$SS_R = \frac{S_{xy}^2}{S_{xx}} = \frac{1585{,}87^2}{602{,}78}$

$= 4\,169{,}78$

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	F_0
Régression	4 169,782	1	4 169,782	32,57
Erreur	1 280,415	10	128,0415	
Totale	5 450,197			

-0,5

$SS_E = \sum_{i=1}^{n} E_i^2$

avec $E_i = \hat{y}_i - y_i$

$5\,450{,}197 - 4\,169{,}78$

$MS_I = \frac{SS_I}{2-1} = \frac{4\,169{,}782}{1} = 4169{,}782$

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

$MS_E = \frac{SS_E}{n-2} = \frac{1280{,}415}{12-2} = 128{,}0415$

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

Puisque $F_0 = 32{,}57 > 4{,}67 = F_{\alpha,1,n-2}$, nous pouvons rejeter H_0 et dire que l'ozone et la température ne présente pas de linéarité.

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

$\sigma^2 = \frac{SS_E}{n-2} = MS_E = 128{,}0415$

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.

$\bar{x} \pm \cancel{z}_{\alpha/2} \sqrt{\sigma^2\left(1+\frac{1}{n}\right)} = \cancel{1,}96 \sqrt{128,0415\left(1+\frac{1}{12}\right)}$

$\pm 23,084$

$\bar{x} + 23,084 = 121,6$

$\bar{x} - 23,084 = 75,436$

intervalle :

$[\, 75,4 ; 123,6 \,]$

-1

OK

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

-2

Question n° 6 : (4 points)

1,5

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

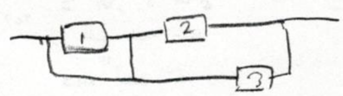
a) **(2 points)** Un système est constitué de deux composants placés en série qui fonctionnent indépendamment l'un de l'autre. De plus, un troisième composant est utilisé en redondance passive, précisément :

- 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.

0,5



$T_{total} = \mathrm{Min}\{T_1 + T_3, \; T_2 + T_3\}$

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

~~exponentielle~~

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.

$\bar{N} = \frac{\rho}{1-\rho}$ $\quad \rho_1 = \frac{\lambda_1}{\mu_1} = \frac{45}{50} = 0,9$ $\quad \rho_2 = \frac{\lambda_2}{\mu_2} = \frac{4}{5} = 0,$

$$\bar{N}_{tot} = \underbrace{\bar{N}_1 + \bar{N}_2}_{\text{car les 2 caisses sont indépendantes}} = \frac{0,9}{1-0,9} + \frac{0,8}{1-0,8} = 9 + 4 = \boxed{13}$$ ✓

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

Pour que le système soit vide à ~~l'équilibre~~, $\lim_{t \to \infty} N(t) = 0$