

Questionnaire
Examen final

MTH2302D

Sigle du cours

Identification de l'étudiant(e)			
Nom :			
Prénom :			
Signature :	Matricule :	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		
<input type="checkbox"/> Aucune <input type="checkbox"/> Toute <input checked="" type="checkbox"/> Voir directives particulières	<input type="checkbox"/> Aucune <input type="checkbox"/> Toutes <input checked="" type="checkbox"/> Non programmable	Les cellulaires, agendas électroniques ou téléavertisseurs sont interdits.	

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

Question n° 1 : (6 points)

(1)

Soit X_1, X_2, \dots, 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 $P(1 \leq \bar{X} \leq 1,05)$.

$$\begin{aligned} P\left(\frac{1-\mu}{\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sqrt{n}} < \frac{1,05-\mu}{\sqrt{n}}\right) &= P(0 \leq Z \leq 0,5) = P(Z \leq 0,5) \\ &\quad - P(Z \leq 0) \\ &= \Phi(0,5) - \Phi(0) \\ &= 0,69146 - 0,5 = \\ &\quad \boxed{P(1 \leq \bar{X} \leq 1,05) = 0,19146} \end{aligned}$$

- 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 $P(T > U)$.

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

~~$$\mu_T = \frac{\sum_{i=1}^6 X_i}{6} \quad \mu_U = \frac{\sum_{i=7}^9 X_i}{3}$$~~

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 $Y \sim U(0,5; 0,5)$

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

$$\text{Var}(Y_i) = \frac{1}{12} \quad n = 100$$

$$X \sim N(E(x), \text{Var}(x))$$

$$E(x) = \sum_{i=1}^{100} E(Y_i) = 100 \cdot 0 = 0$$

$$\text{Var}(x) = \text{var}(\sum_{i=1}^{100} Y_i) = \frac{100}{12} = \frac{25}{3}$$

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

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.

$$H_0: X \sim \text{Poi}(c) \quad \checkmark$$

$$H_1: X \neq \text{Poi}(c) \quad \checkmark$$

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

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

$$\text{On rejette } H_0 \text{ si } \chi^2_o > \chi^2_{k-p-1} \quad \text{si } X \sim \text{Poi}(c) \text{ alors } E(X) = c.$$

somme des pièces = $n = 100$

$$\chi^2_o = \sum_{i=0}^4 \frac{(O_i - E_i)^2}{E_i} \quad P(X=0) = \frac{1,4 \cdot e^{-1,4}}{0!} = 0,2466 \quad c = \bar{x} = 1,4$$

$$P(X=1) = 0,3452$$

$$P(X=2) = 0,2417$$

$$P(X=3) = 0,1128 \quad \checkmark$$

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

$$\chi^2_o = 7,33 \quad (0,5)$$

$$\chi^2_{k-p-1} = \chi^2_{4-1-1} = \chi^2_2 = 5,99 \quad \checkmark$$

On rejette H_0 car $\chi^2_o > \chi^2_{k-p-1} \quad \checkmark$

$$X \neq \text{Poi}(c) \quad \checkmark$$

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?

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

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

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

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.

$$H_0: p = 0,10$$

$$H_1: p > 0,10$$

paramètre d'intérêt: p

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 ?

on rejette H_0 car $p\text{-value} < \alpha$

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

- 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 ?

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

$$\begin{aligned}\beta &= \phi \left(\frac{p_0 - p_0 \Phi \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right) = \phi \left(\frac{0,1 - 0,11 - 1,96 \cdot 0,01342}{0,01399} \right) \\ &= \phi(-2,5944) = 1 - \phi(2,5944) \\ &= 1 - 0,9952 \\ &= 0,0048.\end{aligned}$$

Question n° 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 $X_2 \sim N(\mu_2, \sigma_2^2) \quad \bar{X}_2 = 50,3 \quad S_2^2 = 2,18 \quad n_2 = 9$
 $S_2 = 1,48 \quad \sqrt{n_2} = 3$

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

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

$$[49,3824 ; 51,2176] \quad \checkmark$$

Question n° 4 : (suite)

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

$$1,5 \quad \xrightarrow{\quad} x_2$$

$$x_{n+1} \in \bar{x} \pm t_{n-1}(\frac{\alpha}{2}) \sqrt{s^2(1+\frac{1}{n})}$$

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

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

- 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

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

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

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(15) \cdot 0,37}{0,5} = 11,1 \quad \text{on rejette } H_0 \text{ si } \chi_0^2 > \chi_{\alpha/2; n-1}^2$$

$$\chi_{n-1}^2(\frac{\alpha}{2}) = 27,49$$

$$\chi_0^2 < 27,49$$

$$\chi_{n-1}^2(1-\frac{\alpha}{2}) = 6,27$$

$$\chi_0^2 > 6,27$$

donc Non on ne peut pas rejeter H_0

Question n° 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

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

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

$$F_0 = \frac{S_1^2}{S_2^2} \quad \begin{array}{l} \text{on rejette } H_0 \text{ si} \\ F_0 < F_{1-\frac{\alpha}{2}}; n_1-1; n_2-1 \\ \text{ou } F_0 > F_{\frac{\alpha}{2}}; n_1-1; n_2-1 \end{array}$$

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

$$F_{1-\frac{\alpha}{2}}; n_1-1; n_2-1 = F_{0,975}; 15; 8 = \frac{1}{F_{0,025}; 8; 15} = \frac{1}{3,20} = 0,3125 \leftarrow F_0 < 0,3125$$

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

Oui on peut dire que σ_1^2 et σ_2^2 sont différentes car on a rejetté 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 $\alpha = 0,05$.

$$\bar{Y}_1^2 \neq \bar{Y}_2^2$$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_2 < \mu_1 \text{ donc } \mu_1 > \mu_2$$

On rejette H_0 si

$$T_0 > t_{\alpha}$$

$$T_0 = \frac{57,4 - 50,3}{\sqrt{\frac{0,37}{16} + \frac{2,18}{9}}} = 4,07673$$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$V = \left(\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}} \right) - 2 \checkmark$$

$$V = 9,937$$

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

On rejette H_0 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 n° 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 $^{\circ}C$) 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 :

$$\sum_{i=1}^{12} x_i = 244,10 ; \quad \sum_{i=1}^{12} x_i^2 = 5567,37 ; \quad \sum_{i=1}^{12} x_i y_i = 25631,34 ;$$

$$\sum_{i=1}^{12} y_i = 1182,20 ; \quad \sum_{i=1}^{12} y_i^2 = 121916,60.$$

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

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

$$= \frac{1585,31}{607,78} = 2,63$$

$$\bar{x} = \frac{\sum x_i}{12} = 20,34$$

$$\bar{y} = \frac{\sum y_i}{12} = 98,517$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 98,517 - 2,63 \cdot 20,34 = 45,02$$

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

Question n° 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	5450,197	11		

$$SS_R = \hat{\beta}_1 \cdot S_{xy} = \hat{\beta}_1 \cdot \sum x_i y_i - n \bar{x} \bar{y} = 2,63 \cdot 25631,34 - 12 \cdot 20,34 \cdot 98,517 = 4169,37$$

$$SS_E = SST - SS_R = 1280,83$$

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

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

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

(10,71)
hyp

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

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

$$S^2 = m'_2 - (m'_1)^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = 50,23$$

$\frac{0}{2}$

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.

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

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

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

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

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

- 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 $\text{Exp}(2)$.

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

$$\begin{aligned}
 & \lambda = 2 \quad \mu = 0,5 \\
 & Z = \sum R_i(t) \\
 & \text{circuit en parallèle} \\
 & R_1(t) = \exp(-\int 2e^{-2t} dt) \\
 & \exp(-2 \int e^{-2t} dt) \\
 & \exp(2t e^{-2t}) \\
 & \exp(4e^{-2t})
 \end{aligned}$$

$$\begin{aligned}
 & R_2(t) = 1 - \prod_{i=1}^{k-1} [1 - R_i(t)]^{n_i}
 \end{aligned}$$

- 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.

$$\left\{ \begin{array}{ll} \lambda_1 = 45 & \mu_1 = 50 \\ \lambda_2 = 4 & \mu_2 = 5 \end{array} \right.$$

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

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

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

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

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