Pratique pour l'examen final

STATISTIQUES POUR LES SCIENCES

MAT4681-020 **Mardi 12 décembre 2023** **Automne 2023**

***Question* 1 (5 points)**

En 2010, le résultat moyen obtenu à un examen de mathématiques par l'ensemble des étudiants de 5e secondaire en Ontario a été de 60 sur 100. En 2020, on a fait passer le même examen à 51 étudiants de 5e secondaire de l'Ontario, choisis au hasard. Le résultat moyen obtenu par ces 51 étudiants à l'examen est de 58 sur 100, avec un écart-type de 4,75 sur 100. Peut-on conclure que le résultat moyen des étudiants à cet examen en 2020 est supérieur au résultat moyen obtenu en 2010? En utilisant un seuil de signification de 5%, déterminez la valeur observée de la statistique (t) du test approprié, et concluez par C1 ou C2, avec:

- C1 : on peut conclure avec confiance que le résultat moyen des étudiants à cet examen en 2020 est supérieur au résultat moyen obtenu en 2010;
- C2 : on ne peut pas conclure avec confiance que le résultat moyen des étudiants à cet examen en 2020 est supérieur au résultat moyen obtenu en 2010.

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
t = -3,007 et C_1	t= -3,007 et C_2	t= -3,407 et C_1	t= -3,407 et C_2	t= 3,007 et C_1	t= 3,007 et C_2	t= 3,407 et C_1	t= 3,407 et C_2

***Question* 2 (5 points)**

Voici quelques statistiques descriptives relatives à une étude sur la consommation d'essence des principaux modèles de voitures sur le marché. L'étude a été réalisée à l'aide d'un échantillon aléatoire de 41 voitures.

	Consommation sans traction avant (Variable X)	Consommation avec traction avant (Variable Y)
Effectif	16	25
Moyenne échant.	12,383	10,688
Variance corrigée	1,190	2,386

On souhaite tester si la consommation moyenne d'essence en ville des voitures équipées d'une traction avant est différente de celle des autres voitures. Pour un seuil de α = 5 %, donnez les bornes (a, b) de l'intervalle de confiance de la moyenne de X - Y, et concluez par C1 ou C2 :

- C1 : on peut conclure avec confiance que la consommation moyenne d'essence en ville des voitures équipées d'une traction avant est différente de celle des autres voitures;
- C2 : on ne peut pas conclure avec confiance que la consommation moyenne d'essence en ville des voitures équipées d'une traction avant est différente de celle des autres voitures.

A	B	C	D	E	F	G	H
a = 0,796 b = 2,593 et C_1	a = 0,796 b = 2,593 et C_2	a =0,942 b = 0,968 et C_1	a =0,942 b = 0,968 et C_2	a = 0,912 b = 0,988 et C_1	a = 0,912 b = 0,988 et C_2	a = 1,722 b = 2,418 et C_1	a = 1,722 b = 2,418 et C_2

***Question* 3 (5 points)**

On a recueilli un échantillon de 17 utilisateurs de Twitter, et on a noté leur nombre de gazouillis (*tweets*) au cours de la journée d'hier. Voici les données recueillies:

16	13	13	16	2	14	1	3	12	29	6	1	11	16	11	4	12

On veut tester si la proportion d'individus tweetant plus de 13 fois par jour est supérieure à 0,30. En utilisant une approximation basée sur la normale, déterminez la valeur critique $\boldsymbol{c_\alpha}$ de la statistique du test au seuil de α = 1 %, puis conclure par C1 ou C2, avec:

– C_1 : le test est significatif au seuil de α = 1 %;
– C_2 : le test n'est pas significatif au seuil α = 1 %.

A	B	C	D	E	F	G	H
c_α = 1,684 et C_1	c_α = 1,684 et C_2	c_α= 1,960 et C_1	c_α= 1,960 et C_2	c_α= 2,021 et C_1	c_α=2,021 et C_2	c_α= 2,326 et C_1	c_α= 2,326 et C_2

***Question* 4 (5 points)**

On a choisi 499 étudiants au hasard parmi tous les étudiants de niveau universitaires au Canada, dont 274 sont des femmes et 225 sont des hommes. Parmi ces 499 étudiants, 144 femmes sont inscrites à temps plein et 144 hommes sont inscrits à temps plein. On veut tester si la proportion p_x des étudiantes qui sont inscrites à temps plein est différente de la proportion p_y des étudiants qui sont inscrits à temps plein. En utilisant une approximation basée sur la loi normale, donnez les bornes (a, b) de l'intervalle de confiance de $p_x - p_y$ de niveau 98 %, puis concluez par C1 ou C2, avec:

- C1 : le test est significatif au seuil de α = 2 %;
- C2 : le test n'est significatif au seuil de α = 2 %.

A	B	C	D	E	F	G	H
a = -0,216 b = -0,012 et C_1	a = -0,216 b = -0,012 et C_2	a = -0,256 b = -0,062 et C_1	a = -0,256 b = -0,062 et C_2	a = 0,012 b = 0,216 et C_1	a = 0,012 b = 0,216 et C_2	a = 0,062 b = 0,256 et C_1	a = 0,062 b = 0,256 et C_2

Contexte pour les Questions 5 et 6

Voici les statistiques descriptives relatives à une étude sur la consommation d'essence des principaux modèles de voitures sur le marché. L'étude a été réalisée à l'aide d'un échantillon aléatoire de 41 voitures.

	Consommation en ville (Variable X)	Consommation sur l'autoroute (Variable Y)	Différence de consommation (Variable X-Y)
Effectif	41	41	41
Moyenne	11,123	7,623	3,50
Variance	2,618	0,884	0,849

***Question* 5 (5 points)**

On veut tester si, en moyenne, la consommation d'essence d'une voiture en ville est supérieure à 10,50. Au seuil de α =5 %, déterminez le point critique d'ordre α ($\boldsymbol{c_\alpha}$) du test approprié, puis concluez par C1 ou C2, avec:

– C_1 : en moyenne, la consommation d'essence est égale à 10,50;
– C_2 : en moyenne, la consommation d'essence est supérieure à 10,50.

A	B	C	D	E	F	G	H
c_α=1,684 et C_1	c_α=1,684 et C_2	c_α=1,960 et C_1	c_α=1,960 et C_2	c_α=2,021 et C_1	c_α=2,021 et C_2	c_α=2,326 et C_1	c_α=2,326 et C_2

***Question* 6 (5 points)**

On souhaite maintenant vérifier si la consommation moyenne d'essence d'une voiture en ville est significativement différente de sa consommation moyenne sur l'autoroute. Au seuil de α = 5%, déterminez un intervalle de confiance (a, b) de la moyenne de X-Y, puis concluez par C1 ou C2:

– C_1 : le test est significatif au seuil de α = 5%;
– C_2 : le test n'est pas significatif au seuil α = 5%.

A	B	C	D	E	F	G	H
a = 3,209 b = 3,790 et C_1	a = 3,209 b = 3,790 et C_2	a = 3,409 b = 3,990 et C_1	a = 3,409 b = 3,990 et C_2	a = 3,609 b = 3,790 et C_1	a = 3,609 b = 3,790 et C_2	a = 3,809 b = 4,190 et C_1	a = 3,809 b = 4,190 et C_2

Contexte pour les questions 7 et 8

Une boutique de thé possède trois succursales situées à Montréal, Laval et Longueuil. On veut estimer le total des ventes (soit la somme de toutes les factures) durant le mois de mai 2020 par un échantillon stratifié selon les trois succursales. Pour cela, on prélève un échantillon de factures dans chacune des succursales, en mai 2020. On se sert également des informations relatives à une étude réalisée en mai 2019. Voici les résultats :

	Succursale			
	Montréal	**Laval**	**Longueuil**	**Total**
Nombre de factures en mai 2020	3 000	1 800	1 200	6 000
Nombre de factures dans l'échantillon	300	180	120	600
Moyenne des factures de l'échantillon	30	40	60	
Écart type corrigé des factures de l'échantillon	2	3	4	
Écart type corrigé de toutes les factures de **mai 2019**	3	3	8	

Question 7 (5 points)

Estimez le total des ventes de la boutique de thé au cours du mois de mai 2020.

A	B	C	D	E	F	G	H
234 000	239 000	244 000	249 000	254 000	259 000	264 000	269 000

Question 8 (5 points)

Déterminez le nombre de factures à prélever dans la succursale de Laval en mai 2020 selon l'allocation optimale d'un échantillon de taille 600.
Note : utilisez les écarts-types corrigés de toutes les factures de mai 2019 pour estimer la taille de l'échantillon à prélever dans la succursale de Laval en mai 2020.

A	B	C	D	E	F	G	H
115	120	125	130	135	140	145	150

***Question* 9 (5 points)**

On veut comparer l'indice de masse corporelle (IMC) des Ontariens à celui des Québécois. Pour ce faire, on a observé l'IMC de 1200 Ontariens qu'on a répartis en trois classes. Voici leur distribution en effectif, ainsi que la distribution, en fréquence, de l'IMC de plusieurs millions de Québécois selon les mêmes classes:

Classes des IMC	Ontariens (effectif)	Québécois (fréquence)
30 et plus (obésité)	312	0,22
25 à 29.9 (embonpoint)	720	0,62
8.5 à 24.9 (normal)	168	0,16

On veut savoir si la distribution des IMC des Ontariens diffère de celle des Québécois. Si le niveau du test est de 5%, déterminez la valeur observée de la statistique du test approprié (χ^2), et concluez par C1 ou C2, avec :

- C1: on peut conclure qu'il y a une différence significative;
- C2: on ne peut pas conclure qu'il y a une différence significative.

A	B	C	D	E	F	G	H
χ^2 = 11,50 et C1	χ^2 = 11,50 et C2	χ^2 = 12,50 et C1	χ^2 = 12,50 et C2	χ^2 = 13,50 et C1	χ^2 = 13,50 et C2	χ^2 = 14,50 et C1	χ^2 = 14,50 et C2

***Question* 10 (5 points)**

On veut comparer l'indice de masse corporelle (IMC) des Ontariens à celui des Québécois. Pour ce faire, on a observé l'IMC de 1200 Ontariens et l'IMC de 1200 Québécois, qu'on a répartis en trois classes. Voici la distribution (en fréquence) des IMC:

Classes des IMC	**Ontariens (effectif)**	**Québécois (fréquence)**
30 et plus (obésité)	0,20	0,26
25 à 29.9 (embonpoint)	0,70	0,62
8.5 à 24.9 (normal)	0,10	0,12

On veut savoir si l'IMC varie en fonction de la région d'habitation. Si le niveau du test est de 5%, déterminez la valeur observée de la statistique du test approprié (χ^2), et concluez par C1 ou C2, avec:

- C1: on peut conclure qu'il n'y a pas de relation entre l'IMC et la région d'habitation;
- C2: on ne peut pas conclure qu'il y a une relation entre l'IMC et la région d'habitation.

A	B	C	D	E	F	G	H
χ^2 = 14,39 et C1	χ^2 = 14,39 et C2	χ^2= 15,39 et C1	χ^2 = 15,39 et C2	χ^2 = 16,39 et C1	χ^2 = 16,39 et C2	χ^2 = 17,39 et C1	χ^2 = 17,39 et C2

Question 12 (5 points)

Des chercheurs se demandent quel est le mode de transport du futur selon les citoyens de Montréal et Québec. Ils veulent déterminer si le mode de transport considéré comme celui du futur est distribué selon les mêmes proportions chez les Montréalais et chez les Québécois. On a sondé 150 citoyens sur le sujet et voici les résultats :

	Mode de transport considéré comme celui du futur			
Ville	Voiture	Transport en commun	Vélo	Total
Montréal	25	20	15	60
Québec	35	40	15	90
Total	60	60	30	150

Y a-t-il une différence entre les Montréalais et les Québécois quant au mode de transport considéré comme celui du futur? Si le niveau du test est de 5 %, déterminez la statistique du test approprié ($\chi 2$) et concluez par C1 ou C2, où

- C1 : on peut conclure avec confiance qu'il y a une différence entre les Montréalais et les Québécois quant au mode de transport considéré comme celui du futur.
- C2 : on ne peut pas conclure avec confiance qu'il y a une différence entre les Montréalais et les Québécois quant au mode de transport considéré comme celui du futur.

A) χ^2 = 2,431 et C_1	**B)** χ^2 = 2,431 et C2	**C)** χ^2 = 4,431 et C_1	**D)** χ^2 = 4,431 et C_2
E) χ^2 = 6,431 et C_1	**F)** χ^2 = 6,431 et C_2	**G)** χ^2 = 8,431 et C_1	**H)** χ^2 = 8,431 et C_2

Question 13 (5 points)

On cherche à déterminer la relation linéaire entre le nombre de personnes assises à une table de restaurant (X) et le pourboire (en pourcentage de la facture totale) donné (Y), en supposant qu'il y a une seule facture par table. Voici les données qu'on a prélevées sur six tables :

	Table *i*					
	1	2	3	4	5	6
x_i : Nombre de personnes	3	3	6	9	12	12
y_i : Pourboire en pourcentage	12	22	14	13	14	9

$$\sum_{i=1}^{n} x_i = 45 \qquad \sum_{i=1}^{n} y_i = 84 \qquad \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = -51$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 85{,}5 \qquad \sum_{i=1}^{n} (y_i - \bar{y})^2 = 94$$

Déterminez, selon le modèle de régression simple, de quel pourcentage le pourboire diminue en moyenne lorsque quatre personnes s'ajoutent à la table.

A) 1,79	**B)** 1,99	**C)** 2,19	**D)** 2,39
E) 2,59	**F)** 2,79	**G)** 2,99	**H)** 3,19

Question 14 (5 points)

Une compagnie de location de voitures veut expliquer la valeur marchande d'une voiture par son kilométrage. Elle s'intéresse à un modèle de voiture et à une année de fabrication en particulier. La compagnie observe donc le kilométrage et le prix de vente au cours du dernier mois de 100 voitures correspondant à ses critères. Voici les résultats :

Moyennes		Écarts types corrigés	
Kilométrage	Prix	Kilométrage	Prix
45 000	35 000	15 000	9 000
Corrélation entre le km et le prix : -0,25			

Déterminez un intervalle de confiance à 95 % pour la pente de la droite de régression de la population (fournir seulement la borne supérieure pour la réponse ci-dessous) et déterminez, à l'aide de cet intervalle, si la relation linéaire observée entre le kilométrage et le prix de vente est significative. Concluez par C1 ou C2, où

- C1 : on peut conclure avec confiance qu'il y a une relation linéaire significative entre le kilométrage et le prix de vente.
- C2 : on ne peut pas conclure avec confiance qu'il y a une relation linéaire significative entre le kilométrage et le prix de vente.

A) - 0,083 et C_1	**B) -** 0,083 et C_2	**C) -** 0,033 et C1	**D) -** 0,033 et C_2
E) 0,013 et C_1	**F)** 0,013 et C_2	**G)** 0,063 et C_1	**H)** 0,063 et C_2

Question 15 (5 points)

Une population de commerces de détail est répartie en 3 strates. Le tableau suivant présente les tailles de ces strates ainsi qu'une estimation des écarts types (corrigés) de la variable « nombre d'employés ».

	Strate		
	Strate 1	*Strate* 2	*Strate* 3
Nombre d'unités dans la strate	8 000	5 000	2 000
Écart-type corrigé de la strate	25	40	120

On compte tirer un échantillon stratifié de taille 720 afin d'estimer le nombre moyen d'employés. Estimez l'écart-type de l'estimateur lorsqu'on utilise l'allocation proportionnelle.

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
1,9195	2,0295	2,1295	2,2295	2,3295	2,4295	2,5295	2,6295

Question 16 (5 points)

Un employeur se demande s'il y a un lien entre les salaires (Y) et le nombre d'enfants (X) des employés dans son entreprise. Il a observé ces données pour 100 employés. Voici quelques statistiques sur l'échantillon (n = 100).

	X	Y
Somme	185	4 800 000
Variance corrigée	0,0625	144 000 000
Covariance corrigée entre x et y	2400	

Testez l'hypothèse selon laquelle il n'y a pas de relation linéaire entre les salaires et le nombre d'enfants des employés, au niveau de signification de 5%. Déterminez la statistique du test (Z) et concluez par C1 ou C2, où

- C1: on peut conclure avec confiance qu'il y a une relation linéaire entre les salaires et le nombre d'enfants des employés.
- C2: on ne peut pas conclure avec confiance qu'il y a une relation linéaire entre les salaires et le nombre d'enfants des employés.

A	B	C	D	E	F	G	H
Z = 12,20 et C_1	Z = 12,20 et C_2	Z= 13,20 et C_1	Z = 13,20 et C_2	Z = 14,20 et C_1	Z = 14,20 et C_2	Z = 15,20 et C_1	Z = 15,20 et C_2

Question 17 (5 points)

On désire tester l'hypothèse qu'un certain candidat dans une élection est aussi populaire chez les électeurs francophones que chez les anglophones. On désigne par p_1 (respectivement p_2) la proportion des électeurs francophones (respectivement anglophones) qui appuient le candidat en question. On a interrogé 200 électeurs francophones et 100 électeurs anglophones pris au hasard. Parmi ces électeurs, on a dénombré 76 et 52 intentions de votes, respectivement chez les francophones et les anglophones.

Au seuil de α =5 %, déterminez le point critique d'ordre α ($\boldsymbol{c_\alpha}$) du test approprié, puis concluez par C1 ou C2, avec:

– C_1 : on peut conclure avec confiance que les deux proportions sont égales;
– C_2 : on ne peut pas conclure avec confiance que les deux proportions sont égales.

A	B	C	D	E	F	G	H
c_α= 1.64 et C_1	c_α= 1.64 et C_2	c_α= 1.96 et C_1	c_α= 1.96 et C_2	c_α=2,02 et C_1	c_α=2,02 et C_2	c_α=2,32 et C_1	c_α=2,32 et C_2

Question 18 (5 points)

Au cours d'une analyse récente effectuée par une compagnie d'assurance, on a recueilli des données sur 500 accidents selon le jour de la semaine. Voici la distribution observée :

	Jour							
	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche	Total
Nombre d'accidents	88	72	85	75	80	47	53	500

En théorie, 20 % des accidents surviennent de façon uniforme durant la fin de semaine (soit 10 % des accidents le samedi et 10 % le dimanche), et l'autre 80 % des accidents surviennent de façon uniforme durant la semaine (16 % le lundi, 16 % le mardi, …, 16 % le vendredi). On veut tester si les données appuient la théorie. Si le niveau du test est de 5 %, déterminez la statistique du test approprié (χ^2) et concluez par C_1 ou C_2, où

- C_1 : on peut conclure que les données n'appuient pas la théorie.
- C_2 : les données ne nous permettent pas de rejeter la théorie.

A) χ^2 = 2,585 et C_1	**B)** χ^2 = 2,585 et C2	**C)** χ^2 = 3,585 et C_1	**D)** χ^2 = 3,585 et C_2
E) χ^2 = 4,585 et C_1	**F)** χ^2 = 4,585 et C_2	**G)** χ^2 = 5,585 et C_1	**H)** χ^2 = 5,585 et C_2

Question 19 (5 points)

Un sociologue s'intéresse à la relation entre la couleur de la peau et la mobilité professionnelle. Il prélève un échantillon de 95 personnes de peau pâle, 170 de peau moyenne, et 85 personnes de peau brunes. Il construit une mesure de mobilité à l'aide de quoi il classifie ses sujets selon leur mobilité. Voici les résultats**.**

	Couleur de la peau			
Mobilité	Pâle	Moyenne	Brune	Total
Grande	35	84	51	170
Faible	60	86	34	180
Total	95	170	85	350

Y a-t-il une relation entre la couleur de la peau et la mobilité professionnelle? Si le niveau du test est de 5 %, déterminez la statistique du test approprié (χ^2) et concluez par C1 ou C2, où

- C1 : on peut conclure avec confiance qu'il y a un impact de la mobilité professionnelle sur la couleur de la peau;
- C2 : on ne peut pas conclure avec confiance qu'il y a un impact de la mobilité professionnelle sur la couleur de la peau.

A) χ^2 = 6,725 et C_1	**B)** χ^2 = 6,725 et C2	**C)** χ^2 = 7,725 et C_1	**D)** χ^2 = 7,725 et C_2
E) χ^2 = 8,725 et C_1	**F)** χ^2 = 8,725et C_2	**G)** χ^2 = 9,725 et C_1	**H)** χ^2 = 9,725 et C_2

Question 20 (5 points)

On s'intéresse à la proportion *p* de pommes non gâtées dans un lot de pommes. Un échantillon de 360 pommes contient 30 pommes gâtées. En se basant sur l'approche avec la loi normale, donnez la longueur de l'intervalle de confiance de *p* à 90%.

A	B	C	D	E	F	G	H
0,507	0,407	0,307	0,207	0,107	0,087	0,067	0,047