



Université d'Ottawa • University of Ottawa

Faculté des sciences
Mathématiques et de statistique

Faculty of Science
Mathematics and Statistics

MAT 2779 : Examen Final

Cahier-Questionnaire

Professeur : Gilles Lamothe

15 décembre 2015

Durée : 3 heures

Il est interdit de se servir de téléphones cellulaires, de dispositifs électroniques non autorisés ou de notes de cours (à moins qu'il s'agisse d'un examen à livre ouvert). Les téléphones et les dispositifs doivent être fermés et rangés dans votre sac : vous ne pouvez pas les laisser dans vos poches ou sur votre personne. Sinon, on pourrait vous demander de quitter l'examen immédiatement et des allégations de fraude scolaire pourraient être déposées dont le résultat pourrait être un 0 (zéro) pour l'examen.

- C'est un examen à livre fermé. Deux feuilles (à deux côtés) sont permises.
- Seule une calculatrice non-programmable et non-graphique est permise.
- Il y a trois parties : 12 questions à choix multiples, 10 questions Vrai/Faux et 3 questions à courtes réponses.
- L'examen est corrigé sur 55 points.

585, av. King-Edward Ottawa (Ontario) K1N 6N5 Canada
585 King Edward Avenue Ottawa, Ontario K1N 6N5 Canada
(613) 562-5864 • Téléc./Fax (613) 562-5776
Courriel/Email: uomaths@science.uottawa.ca

Questions à choix multiples (2 points/question pour un total de 24 points)

S'il vous plaît entrer vos choix dans le cahier-réponse.

1. Environ 12 % des femmes ont un cancer du sein. Supposons que 5% des femmes atteintes de cancer du sein ont des antécédents familiaux de cancer du sein (soit une mère ou une sœur avec un cancer du sein), alors que seulement 2% des femmes sans cancer du sein ont des antécédents familiaux de cancer du sein. Étant donné qu'une femme choisie au hasard a des antécédents familiaux de cancer du sein, calculer la probabilité que cette femme souffre d'un cancer du sein.

A) 0,122 B) 0,0236 C) 0,006 D) 0,254 E) 0,746

2. La prévalence de la maladie d'Alzheimer a été étudiée. Le tableau suivant présente la prévalence de la maladie d'Alzheimer (en pourcentage) selon le groupe d'âge et le sexe. Par exemple, la probabilité qu'un homme dans le groupe d'âge 80-84 ait la maladie d'Alzheimer est 8,6%.

Groupe d'âge	Homme	Femme
65-69	1,6	0,0
70-74	0,0	2,2
75-79	4,9	2,3
80-84	8,6	7,8
85+	35,0	27,9

On sélectionne au hasard un homme du groupe d'âge 75-79, une femme du groupe d'âge 75-79 et une femme du groupe d'âge 80-84. Quelle est la probabilité qu'au moins un des trois individus ait la maladie d'Alzheimer ?

A) 0,857 B) 0,999 C) 0,143 D) 0,255 E) 0,145

3. Le poids moyen d'un nouveau bébé normal né dans l'Amérique du Nord est environ 120 onces. Supposons que nous voulons tester l'hypothèse que les mères avec un statut socioéconomique (SSE) faible auront des bébés avec un poids moyen inférieur. Formuler des hypothèses et expliquer quand une erreur de première ou de deuxième espèce sera commise en choisissant la déclaration correcte dans la liste ci-dessous. (Une seule déclaration est correcte.) Soit μ le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible.

A) $H_0 : \mu = 120$ versus $H_1 : \mu < 120$. Une erreur de deuxième espèce est commise lorsque nous concluons que le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible est plus petit que le poids moyen d'un nouveau-né normal, alors que dans la réalité ce n'est pas le cas.

B) $H_0 : \mu = 120$ versus $H_1 : \mu < 120$. Une erreur de première espèce est commise lorsque nous concluons que le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible est plus petit que le poids moyen d'un nouveau-né normal, alors que dans la réalité ce n'est pas le cas.

C) $H_0 : \mu = 120$ versus $H_1 : \mu > 120$. Une erreur de première espèce est commise lorsque nous concluons que le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible est plus petit que le poids moyen d'un nouveau-né normal, alors que dans la réalité ce n'est pas le cas.

D) $H_0 : \mu = 120$ versus $H_1 : \mu > 120$. Une erreur de deuxième espèce est commise lorsque nous concluons que le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible est semblable au poids moyen d'un nouveau-né normal, alors que dans la réalité ce n'est pas le cas.

E) $H_0 : \mu = 120$ versus $H_1 : \mu \neq 120$. Une erreur de première espèce est commise lorsque nous concluons que le poids moyen d'un nouveau bébé né avec une mère de statut socioéconomique faible est semblable au poids moyen d'un nouveau-né normal, alors que dans la réalité ce n'est pas le cas.

4. L'anémie ferriprive est un problème de santé nutritionnel important au Canada. Supposons que l'apport en fer quotidienne moyenne parmi une grande population de jeunes garçons de toutes les strates de revenu est 14,44 mg. Une évaluation a été effectuée sur l'alimentation d'un échantillon de jeunes garçons dont les familles étaient au-dessous du seuil de pauvreté. Nous tenons à vérifier l'hypothèse que l'apport moyen en fer chez le groupe à faible revenu est différente de celle de la population générale. Pour chaque garçon dans l'échantillon, nous avons mesuré son apport quotidien de fer et nous avons attribué les valeurs à la variable x dans R. Voici la sortie du test.

```
> t.test(x,mu=14.44)
      One Sample t-test
data:  x
t = -2.917, df = 50, p-value = 0.005
alternative hypothesis: true mean is not equal to 14.44
95 percent confidence interval:
 11.19634 13.80366
sample estimates:
mean of x
 12.50
```

Déterminer l'écart type de l'échantillon s pour cet échantillon de jeunes garçons. En outre, en supposant que l'apport quotidien de fer est normalement distribué, que pouvons-nous conclure de cette étude à un niveau de signification de $\alpha = 0,05$?

- (A) $s = 4,75$; Nous pouvons conclure que l'apport moyen en fer pour ce groupe à faible revenu est différente de celle de la population générale.
B) $s = 0,66$; Nous ne pouvons pas conclure que l'apport moyen en fer pour ce groupe à faible revenu est différente de celle de la population générale.
C) $s = 33,93$; Nous pouvons conclure que l'apport moyen en fer pour ce groupe à faible revenu est différente de celle de la population générale.
D) $s = 4,75$; Nous ne pouvons pas conclure que l'apport moyen en fer pour ce groupe à faible revenu est différente de celle de la population générale.
E) $s = 0,66$; Nous pouvons conclure que l'apport moyen en fer pour ce groupe à faible revenu est différente de celle de la population générale.

5. Plusieurs discussions ont apparu dans la littérature médicale ces dernières années sur le rôle de l'alimentation dans le développement de la maladie cardiaque. Les niveaux de cholestérol d'un groupe de personnes sur un régime macrobiotique sont mesurés. Parmi 20 d'entre eux, de 20 à 39 ans, le niveau moyen de cholestérol fut mesuré à 175 mg/dL avec écart type de 35 mg/dL. En supposant que la population est normalement distribuée, calculer un intervalle de confiance à 90% pour le vrai niveau moyen de cholestérol chez ce groupe.

- A) [164,6 ; 185,4] B) [114,5 ; 235,5] C) [167,2 ; 182,8]
D) [171,9 ; 178,0] E) [161,5 ; 188,5]

6. Considérons un échantillon aléatoire X_1, X_2, \dots, X_{22} d'une population normale de moyenne $\mu = 47$. Soient \bar{X} et S la moyenne de l'échantillon et l'écart type de l'échantillon, respectivement. Déterminer c tel que

$$P\left(\frac{\bar{X} - 47}{S/\sqrt{22}} \leq c\right) = 0,90.$$

- A) $c = 1,323$ B) $c = 1,282$ C) $c = 1,321$ D) $c = 1,721$ E) $c = -1,323$

7. En 2004, 90% des naissances au Canada furent des naissances à terme, c.-à-d. ils ont eu la livraison dans la semaine 37 ou plus tard. Si nous choisissons 5 des naissances canadiennes en 2004, quelle est la probabilité qu'exactly 2 d'entre eux étaient des naissances à terme ?

- A) 0,0729 B) 0,9271 C) 0,0081 D) 0,2617 E) 0,00405 F) 0,0523

8. Il est connu que 25% des femmes mariées ont une pression artérielle élevée. On soupçonne que la prévalence de l'hypertension artérielle est plus faible pour les femmes célibataires. Pour tester $H_0 : p = 25\%$ contre $H_1 : p < 25\%$, où p est la prévalence de l'hypertension artérielle chez les femmes célibataires, nous avons cueilli un échantillon de 100 femmes célibataires. Nous avons constaté que 20 femmes de l'échantillon ont une pression artérielle élevée. Calculer la valeur observée de la statistique du test pour tester ces hypothèses. Donnez la valeur P et votre conclusion à $\alpha = 5\%$.

- A) $z_0 = -1,41$, valeur $P=0,0793$; les preuves que p est plus petite que 25% sont significatives
- B) $z_0 = -1,15$, valeur $P=0,1251$; les preuves que p est plus petite que 25% ne sont pas significatives
- C) $z_0 = -1,25$, valeur $P=0,1056$; les preuves que p est plus petite que 25% ne sont pas significatives
- D) $z_0 = -1,62$, valeur $P=0,0526$; les preuves que p est plus petite que 25% ne sont pas significatives
- E) $z_0 = -1,04$, valeur $P=0,1492$; les preuves que p est plus petite que 25% ne sont pas significatives
- F) $z_0 = 1,62$, valeur $P=0,0526$; les preuves que p est plus petite que 25% ne sont pas significatives

9. Supposons que l'on sait que dans une certaine grande population la longueur crânienne est normalement distribuée avec une moyenne de $\mu=185,6$ mm et un écart type de $\sigma=12,7$ mm. Quel est la probabilité qu'un échantillon aléatoire de taille 10 de cette population aura une moyenne plus grande que 190?

- A) 0,8643 B) 0,1357 C) 0,1867 D) 0,2547 E) 0,00205 F) 0,5234

10. Cardosi et al. ont effectué une étude rétrospective de 4 ans sur 102 femmes qui subissent une hystérectomie radicale pour le cancer du col. Une infection des voies urinaires associées aux cathéters a été observée chez 12 des sujets. Voici les nombres de jours postopératoires jusqu'au diagnostic de l'infection pour chaque sujet qui ont eu une infection.

16, 10, 49, 15, 6, 15, 8, 19, 11, 22, 13, 17

Est-ce qu'il y a des valeurs aberrantes dans cet échantillon de taille $n = 12$?

- A) Seulement 49 et 22 sont des valeurs aberrantes
 - B) Seulement 49, 22 et 6 sont des valeurs aberrantes
 - C) Seulement 49, 22, 6, and 8 sont des valeurs aberrantes
 - D) Seulement 49 est une valeur aberrante
 - E) Seulement 6 est une valeur aberrante
 - F) Il n'y a pas de valeurs aberrantes.
11. Considérons la sortie de R suivante :

```
> pbinom(15,100,0.25)
[1] 0.01108327 ←
> pbinom(16,100,0.25)
[1] 0.02111062
> pbinom(17,100,0.25)
[1] 0.03762626
> pbinom(30,100,0.25)
[1] 0.8962128
> pbinom(31,100,0.25)
[1] 0.9306511
> pbinom(32,100,0.25)
[1] 0.9554037
```

Soit X une variable aléatoire binomiale avec $n = 100$ et $p = 0,25$.
Calculer $P(16 \leq X \leq 31)$.

- A) 0,9196 B) 0,9095 C) 0,9343 D) 0,9443 E) 0,9177 F) 0,9598

12. Considérons une étude portant sur 1465 patients à risque de maladie coronarienne (MC) qui sont classés comme ayant la maladie ou non, selon les résultats de la coronarographie. Cet examen permet de visualiser la coronaire et est considéré comme le *gold-standard*. Indépendamment des résultats de la coronarographie, les patients sont interrogés sur l'existence de douleurs à la poitrine qui est le diagnostic que nous aimerions évaluer en termes de sa performance pour identifier la MC. Les résultats de cette étude sont résumés dans le tableau suivant.

Présence de douleurs à la poitrine	Présence de la MC		Total
	Oui	Non	
Positif	969	245	1214
Négatif	54	197	251
Total	1023	442	1465

Calculer la spécificité et la sensibilité de la présence de douleurs à la poitrine comme un test diagnostic pour identifier la MC.

- A) spécificité = $969/1023$; sensibilité = $197/442$
- B) spécificité = $197/251$; sensibilité = $969/1214$
- C) spécificité = $197/1465$; sensibilité = $969/1465$
- D) spécificité = $245/442$; sensibilité = $54/1023$
- E) spécificité = $197/245$; sensibilité = $54/969$
- F) spécificité = $197/442$; sensibilité = $969/1023$ ←

Questions Vrai ou Faux (0,5 point/question pour un total de 5 points)

Utilisez vos connaissances pour identifier la véracité des affirmations suivantes. Si vous croyez qu'un énoncé est vrai, encercler VRAI dans le cahier-réponse; si vous pensez qu'un énoncé est faux ou il n'y a pas suffisamment d'informations pour déterminer la véracité, encercler FAUX dans le cahier-réponse.

1. L'ensemble d'échantillonnage d'une expérience aléatoire est l'ensemble de tous les résultats possibles de cette expérience.
2. Soient A et B deux événements. Alors, $P(A \cup B) = P(A) + P(B)$.
3. Dans le test d'hypothèses, si une hypothèse nulle est rejetée au niveau de signification $\alpha = 0,05$, elle sera également rejetée au niveau de signification $\alpha = 0,01$.
4. Soient A et B deux événements. $A' \cap B$ et $A \cap B'$ sont mutuellement exclusifs.
5. On cueille un échantillon de taille $n = 1500$ d'une population de moyenne μ . Un intervalle de confiance à 95% pour μ est $[-1,5; 2,2]$. Alors, ceci veut dire que 95% des valeurs dans l'échantillon sont dans l'intervalle $[-1,5; 2,2]$.
6. Supposons que X suit une loi normale de moyenne 3 et de variance 4. Alors, $P(X < 2) < P(X \leq 2)$.

7. Soient X et Y deux variables aléatoires indépendantes. Alors,
 $V[X - Y] = V[X] + V[Y]$.
8. Dans les tests d'hypothèses, les hypothèses nulle et alternative doivent être formulées avant de regarder aux données.
9. Soient A et B deux événements avec $P(A) > 0$ et $P(B) > 0$. Alors,
 $P(A|B) + P(A^c|B) = 1$.
10. Supposons que X suit une loi normale de moyenne 3 et d'écart type 2. Dans \mathbf{R} , la commande `1 - qnorm(3, mean=3, sd=2)` calcul la probabilité suivante : $P(X > 3)$.

Questions à réponses courtes (Total de 26 points)

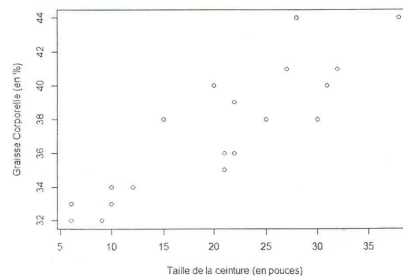
S'il vous plaît entrer vos réponses complètes dans le cahier-réponse.

Question 1 [6 points] : Il est difficile de déterminer avec précision le pourcentage de graisse corporelle d'une personne sans l'immersion dans l'eau. Les chercheurs, qui espèrent trouver des moyens de faire une bonne estimation, ont immergé $n = 20$ sujets masculins pour mesurer leur graisse corporelle (en %). Ils ont également mesuré la taille de la ceinture (en pouces). La corrélation entre la graisse corporelle et la taille de la ceinture est 0,887. Utiliser y pour la graisse corporelle (en %) et x pour la taille de la ceinture (en pouces). Nous avons calculer quelques sommes.

$$\sum_{i=1}^{20} (x_i - \bar{x})^2 = 276,95; \quad \sum_{i=1}^{20} (y_i - \bar{y})^2 = 1737,75$$

$$\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = 615,25; \quad \sum_{i=1}^{20} y_i = 395; \quad \text{et} \quad \sum_{i=1}^{20} x_i = 741.$$

- (a) Voici un nuage de points pour la graisse corporelle (en %) contre la taille de la ceinture (en pouces). Décrire l'association entre les deux variables. (Décrire la forme, l'orientation et l'intensité.)

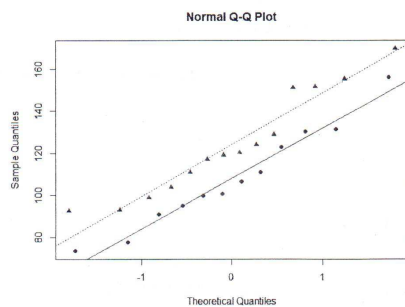


- (b) Déterminer la droite des moindres carrés pour décrire la graisse corporelle (en %) comme une fonction linéaire de la taille de la ceinture (en pouces).
- (c) Utiliser le modèle linéaire de la partie (c), pour prévoir le pourcentage de graisse corporelle pour un homme avec une ceinture de 37 pouces.

Question 2 : [13 points] Pour étudier le lien entre le calcium intracellulaire et la pression artérielle, les chercheurs ont mesuré le calcium libre entre la concentration de calcium intracellulaire dans les plaquettes sanguines de 12 personnes avec une pression artérielle normale et 14 personnes avec une pression artérielle élevée. Voici un tableau avec quelques statistiques descriptives.

Pression artérielle	Calcium plaquettaire (nM)		
	n	Moyenne	Écart type
Normale	12	104,35	23,94
Élevée	14	124,45	24,68

Voici des diagrammes quantile-quantile superimposés pour ces données.



Soient μ_1 et μ_2 la concentration moyenne de calcium dans les plaquettes sanguines pour une personne ayant une pression artérielle normale et pour une personne ayant une pression artérielle élevée, respectivement.

- (a) Nous croyons que la concentration de calcium moyenne diffère entre ces deux groupes. Formuler une hypothèse nulle et l'hypothèse alternative à tester cette affirmation.

- (b) Référez aux diagrammes quantile-quantile. On utilise ce diagramme pour vérifier quelles conditions sous-jacentes concernant les populations? Quelles sont vos observations?
- (c) Calculer la valeur observée de la statistique du test des hypothèses de la partie (a). Donner la valeur P correspondante (ou un intervalle de valeurs pour cette valeur P).
- (d) À un niveau de signification de 5%, quelles sont les conclusions du test des hypothèses que vous avez formulées dans la partie (a).
- (e) Donner un intervalle de confiance à 95% pour $\mu_1 - \mu_2$.

Question 3 : [7 points] Les effets de l'entraînement physique sur le taux de triglycérides ont été étudiés en utilisant 10 sujets. Voici les mesures avant et après le régime d'entraînement (en milligrammes de triglycérides par 100 millilitres de sang).

Sujet	Avant	Après	Changement = Après - Avant
1	68	95	27
2	76	90	14
3	94	85	-9
4	73	59	-14
5	37	48	11
6	130	121	-9
7	77	136	59
8	24	65	41
9	99	131	32
10	110	104	-6

Supposons que le changement dans le niveau de triglycérides est normalement distribué.

- (a) Déterminer un intervalle de confiance pour le changement moyen dans le niveaux de triglycérides.
- (b) Basé sur l'intervalle de confiance de la partie (a), est-ce que nous sommes très confiants que le taux de triglycérides a changé en moyenne? Si oui, quel est la direction du changement?