5 Étude du lien entre deux variables statistiques

5.1 Distribution conjointe

Q.5.1

Répartition de 383 cégépiens selon le niveau de scolarité de la mère et leur intention d'aller à l'université.

Scolarité de la mère	Intention		Total
	Oui	Non	
Primaire	23	16	39
Secondaire	107	91	198
Collégial	61	44	105
Universitaire	35	6	41
Total	226	157	383

a) Quelle proportion des répondants ont une mère ayant fait des études universitaires ?

b) Quelle proportion de ces cégépiens ont une mère ayant un niveau de scolarité primaire et ont l'intention d'aller à l'université ?

c) Quelle proportion des cégépiens n'ayant pas l'intention d'aller à l'université ont une mère ayant un niveau de scolarité collégial ?

d) Parmi les cégépiens dont la mère a un niveau de scolarité collégial, quelle est la propotion de ceux qui n'ont pas l'intention d'aller à l'université ?

e) Parmi les cégépiens ayant l'intention d'aller à l'université, quelle est la proportion de ceux dont la mère à un niveau de scolarité collégial ou plus ?

f) Quelle proportion des répondants veulent aller à l'université et ont une mère dont le niveau de scolarité ne dépasse pas le secondaire ?

5.2 Test d'indépendance

Q.5.2

D'après les données de la question 5.1, faire un test d'indépendance au seuil de signification 5%, calculer le coefficient de contingence C et interpréter les résultats.

Q.5.3

Dans un échantillon de 150 personnes prélevé dans une région du pays, on a demandé aux gens leur opinion sur l'avortement libre et la religion à laquelle ils adhéraient.

Religion	Opinion			Total
	Pour	Contre	Indécis	
Catholique	31	28	17	76
Protestant	5	25	22	52
Autre	6	8	8	22
Total	42	61	47	150

Peut-on affirmer que l'opinion sur l'avortement est indépendante de la religion ? (Seuil de signification : 5%)

Q.5.4

La pression artérielle a été mesurée chez 300 rats atteints d'une maladie. Le tableau suivant résume la répartition des rats selon leur pression et la conséquence éventuelle de la maladie.

Pression	Conséquence		Total
	Meurt	Survit	
Élevée	136	44	180
Normale	64	56	120
Total	200	100	300

a) Quel est le taux de mortalité parmi les rats ayant une pression artérielle élevée ?

b) Quel est le taux de mortalité parmi les rats ayant une pression artérielle normale ?

c) Peut-on conclure que la pression artérielle élevée améne les rats atteints à mourir ?

Q.5.5

Avant de tirer des conclusions hâtives, on décide de réxaminer les données de l'échantillon de la question 6.4 en séparant les rats selon leur âge (jeune ou vieux).

Répartition des jeunes rats.

Pression	Conséquence		Total
	Meurt	Survit	
Élevée	8	12	20
Normale	32	48	80
Total	40	60	100

Répartition des vieux rats.

Pression	Conséquence		Total
	Meurt	Survit	
Élevée	128	32	160
Normale	32	8	40
Total	160	40	200

a) Quel est le taux de mortalité parmi les jeunes rats ayant une pression artérielle élevée ? normale ?

b) Quel est le taux de mortalité parmi les vieux rats ayant une pression artérielle élevée ? normale ?

c) Que pouvons-nous conclure ?

Q.5.6

On résume dans un tableau de contigence l'opinion concernant l'avortement libre selon l'âge des répondants.

Âge	Opinion: Pour	Contre	Indécis	Total
[15,25[	24	18	8	50
[25,35[	41	16	13	70
[35,45[	33	20	27	80
[45,55[	11	12	7	30
[55,65[	23	32	5	60
[65,95[	12	21	7	40
Total	144	119	67	330

Peut-on affirmer, au niveau de signification $\alpha = 0{,}01$, que l'opinion concernant l'avortement libre est indépendante de l'âge des répondants ?

Q.5.7

Une expérience est menée pour déterminer si un vaccin est efficace pour prémunir les jeunes enfants contre une certaine maladie. On a constitué un groupe expérimental de 105 enfants auxquels on a donné le vaccin et un groupe témoin de 75 enfants qui ont reçu un placebo. On a suivi ces deux groupes pendant cinq ans pour voir si cette maladie se manifesterait au cours de cette période. Voici les résultats obtenus :

Groupe	Apparition de la maladie: Oui	Non	Total
Expérimental	71	34	105
Témoin	37	38	75
Total	108	72	180

Cette expérience démontre-t-elle l'efficacité du vaccin pour prémunir contre cette maladie ? Faire un test au seuil de 1%.

5.3 Test d'ajustement

Q.5.8

Un épicier prétend que les cinq marques de tomates en boîtes sont également populaires auprès de ses clients. Un échantillon de 125 d'entre eux ont exprimé leur préférence et voici le tableau obtenu.

Marque	A	B	C	D	E
Nombre de clients	34	19	32	22	18

Est-ce que l'opinion de l'épicier est vraie ? Faire un test d'ajustement avec un seuil de 1%.

Q.5.9

Les résultats à un test de QI administré à un échantillon de 250 personnes sont présentés dans ce tableau.

Résultats	Nombre de personnes
Moins de 70	10
[70, 85[	32
[85, 100[	80
[100, 115[	80
[115, 130[	40
130 et plus	8
Total	250

Peut-on dire que le QI de l'ensemble des personnes de la population suit une loi normale d'espérance 100 et d'écart-type 15 ? Faire un test d'ajustement avec un niveau de 5

Q.5.10

On répète 200 fois l'expérience aléatoire consistant à lancer 5 pièces de monnaies d'un coup. Soit X : le nombre de faces obtenues parmi les 5 pièces à un lancer. Voici les fréquences observées pour chacune des réalisations de X.

Nombre de faces (x_i)	0	1	2	3	4	5	Total
Nombre de lancers	18	51	62	34	18	17	200

Au seuil 5%, peut-on affirmer que la variable aléatoire X suit une loi binomiale ?

Q.5.11

On prélève un échantillon de 200 tiges produites par une machine et on mesure le diamètre. On obtient un diamètre moyen de 1.198 cm et un écart-type de 0,046 cm. Peut-on dire que la distribution du diamètre de toutes les tiges produites suit un modèle normal ? Faire un test au seuil de 0,05.

Diamètre (cm)	Nombre de tiges
Moins de 1,12	10
[1,12; 1,15[	20
[1,15; 1,18[	46
[1,18; 1,21[	54
[1,21; 1,24[	37
[1,24; 1,27[	24
1,27 et plus	12
Total	200

5.4 Régression linéaire

Q.5.12

On veut étudier le lien entre le poids d'une personne et le taux d'alcool dans le sang. Pour ce faire, on demande à sept personnes de consommer trois bières en une heure. Voici les résultats obtenus :

X : Poids (kg)	52	57	60	66	70	82	91
Y : Taux d'alcool (mg/100ml)	131	120	112	105	84	71	65

a) Tracer le diagramme de dispersion.

b) Calculer le coefficient de détermination et interpréter.

c) Calculer le coefficient de corrélation et interpréter.

d) Donner l'équation de la droite de régression.

e) Au Québec, il est illégal de conduire avec un taux d'alcool supérieur à 80 mg/100 ml. En utilisant le modèle, déterminer si une personne de 60 kg peut prendre le volant après avoir consommé trois bières en une heure.

f) D'après le modèle, à partir de quel poids serait-il illégal de conduire si on a consommé trois bières en une heure ?

Q.5.13

Soit une étude portant sur la concentration de mercure sous phase gazeuze présente à l'intérieur d'un édifice à bureaux situé à ville Saint-Laurent et sur la température extérieure.

Échantillon	Température (°C)	Concentration en mercure (mg/mm^3)
1	25	43,9
2	18	31,7
3	19	30,1
4	22	45,2
5	20	41,4
6	21	45,4
7	13	22,3
8	22	38,2
9	14	22,9
10	12	20,1
Moyenne	18,6	34,12
Écart-type	4,3256	9,9971

a) Quelle proportion de la variation de la concentration de mercure est-elle expliquée par la variation de la température extérieure ?

b) Calculer le coefficient de corrélation et interpréter.

c) Donner l'équation de la droite de régression.

d) Estimer à l'aide du modèle le taux de mercure que l'on devrait observer pour une température de 15°C.

5.5 Exercices récapitulatifs

Q.5.14

Répartition des étudiants ayant réussi un cours de sociologie à l'université, selon la note et le sexe.

Note	Sexe: Masculin	Sexe: Féminin	Total
A	3	3	6
B	12	6	18
C	15	9	24
D	10	2	12
Total	40	20	60

a) Quel est le nombre de personnes ayant eu la note C et étant de sexe féminin ?

b) Quel est le nombre de personnes ayant eu la note B ou étant de sexe masculin ?

c) Parmi les étudiants ayant eu la note C ou D, quel est le pourcentage de personnes de sexe masculin ?

d) Parmi les étudiants de sexe féminin, quel est le pourcentage ayant eu la note B ?

Q.5.15

Répartition de 1500 conducteurs selon l'âge et l'implication dans un accident au cours des 12 derniers mois.

Âge	Accident: Oui	Accident: Non	Total
16 à 24 ans	14	135	149
25 à 44 ans	26	538	564
45 à 64 ans	19	570	589
65 ans et plus	6	192	198
Total	65	1435	1500

a) Les données obtenues permettent-elles de conclure qu'il ya un lien entre l'âge du conducteur et le risque d'accident ? Tester cette hypothèse au seuil de 1%.

b) S'il y a lieu, calculer le coefficient de contingence.

Q.5.16

Dans un échantillon de 2 000 jeunes de 25 ans, on a noté le niveau de scolarité et le revenu des parents. On résume les réponses recueillies dans le tableau de contingence suivant où les revenus sont en milliers de dollars :

	Prim.	Sec.	Coll.	Univ.	Total
[0,20[	43	137	96	24	300
[20,40[	107	380	310	103	900
[40,60[	89	281	246	84	700
[60,80[	21	42	28	9	100
Total	260	840	680	220	2 000

Peut-on affirmer, au seuil de signification de 2,5%, que la scolarité est indépendante du revenu des parents ?

Q.5.17

Pour étudier l'achalandage à une station de service, on note chaque quart d'heure le nombre de clients. Voici les données recueillies au cours des deux dernières semaines.

Nombre de clients au quart d'heure	Nombre de quarts d'heure
0	3
1	10
2	30
3	45
4	39
5	32
6	20
7	13
8	5
9	2
10	1
Total	200

Au seuil 1%, peut-on affirmer que le nombre de clients au quart d'heure obéit à une loi de Poisson ?

Q.5.18

Un chercheur veut comparer la taille d'un homme d'âge adulte avec celle de son père. Il choisit un échantillon de dix couples pères-fils pour lesquels il recueille les données suivantes :

X : Taille du père (cm)	Y : Taille du fils (cm)
176	177
187	183
165	171
172	176
181	180
177	177
179	182
185	182
168	174
170	169
$\bar{x} = 176$ $s_x = 7{,}2572$	$\bar{y} = 177{,}1$ $s_y = 4{,}7714$

a) Calculer le coefficient de corrélation et interpréter.

b) Quelle proportion de la variation de la taille d'un fils peut être attribuée à la variation de la taille du père ?

c) Donner l'équation de la droite de régression.

d) Estimer à l'aide du modèle la taille du fils d'un homme mesurant 175 cm.

Réponses aux exercices

R.5.1

a) $\dfrac{41}{383}$ c) $\dfrac{44}{157}$ e) $\dfrac{96}{226}$

b) $\dfrac{23}{383}$ d) $\dfrac{44}{105}$ f) $\dfrac{130}{383}$

R.5.2 $\chi^2_{obs} = 13{,}84$; $\nu = 3$; $\chi^2_c = 7{,}82$
On rejette $\mathcal{H}_0$: L'intention d'aller ou non à l'université et la scolarité de la mère sont dépendantes. $C = 0{,}19$.

R.5.3 $\chi^2_{obs} = 15{,}92 > 9{,}49$. Rejeter $\mathcal{H}_0$, donc les variables sont dépendantes.

R.5.4

a) 75,56% b) 53,33% c) C'est possible.

R.5.5

a) 40% dans les deux cas. b) 80% dans les deux cas.

c) Parmi les jeunes rats, le taux de mortalité est le même quelque soit la pression. Il en va de même pour les vieux rats. Ces résultats semblent indiquer qu'il n'y a pas de causalité entre la pression artérielle et la mortalité, mais que c'est plutôt l'âge qui agit en même temps sur ces deux variables.

R.5.6 $\chi^2_{obs} = 32{,}72$; $\chi^2_c = 23{,}21$. Rejeter $\mathcal{H}_0$.
Donc les variables sont dépendantes.

R.5.7 $\chi^2_{obs} = 6{,}0952$; $\chi^2_c = 6{,}64$. Non rejet de $\mathcal{H}_0$.
On ne peut pas conclure que le vaccin soit efficace.

R.5.8 $\chi^2_{obs} = 8{,}96$; $\chi^2_c = 13{,}28$. Non rejet de $\mathcal{H}_0$.
L'épicier a raison.

R.5.9 $\chi^2_{obs} = 6{,}097$; $\chi^2_c = 7{,}81$. Non rejet de $\mathcal{H}_0$.
Les données semblent suivre la loi normale mentionnée.

R.5.10 $\chi^2_{obs} = 71{,}64$; $\chi^2_c = 11{,}07$. Rejet de $\mathcal{H}_0$.
La variable alátoire ne suit pas une loi binomiale de paramètres $n = 5$ et $p = 1/2$.

R.5.11 $\chi^2_{obs} = 2{,}52$; $\chi^2_c = 9{,}49$. Non rejet de $\mathcal{H}_0$.
Les données semblent suivre une loi normale.

R.5.12

a) Laissé à l'étudiant.

b) 0,94.
94% de la variation du taux d'alcool est expliquée par la variation du poids.

c) -0,97.
Plus le poids augmente, plus le taux d'alcool diminue. Lien très fort.

d) $y = -1{,}75x + 217{,}80$

e) Non, car son taux d'alcool serait de 112,8 mg/100 ml.

f) 78,74 kg

R.5.13

a) 86,31%

b) 0,9290.
Plus la température est élevée, plus la concentration en mercure augmente. Lien fort.

c) $y = 2{,}1471x - 5{,}817$

d) 26,39 mg/mm^3

R.5.14

a) 9 b) 46 c) 69,4% d) 30%

R.5.15

a) $\chi^2_{obs} = 11{,}96$; $\chi^2_c = 11{,}35$. Rejet de $\mathcal{H}_0$.
Les variables sont dépendantes.

b) $C = 0{,}088$

R.5.16 $\chi^2_{obs} = 13{,}12$; $\chi^2_c = 19{,}02$. Accepter $\mathcal{H}_0$.
Les variables sont indépendantes.

R.5.17 $\chi^2_{obs} = 3{,}09$; $\chi^2_c = 20{,}09$. Non rejet de $\mathcal{H}_0$.
La variable alátoire suit une loi de Poisson.

R.5.18

a) 0,9081.
Plus le père en grand, plus le fils l'est aussi. Lien très fort.

b) 82,46%

c) $y = 0{,}597x + 72{,}02$

d) 176,5 cm.