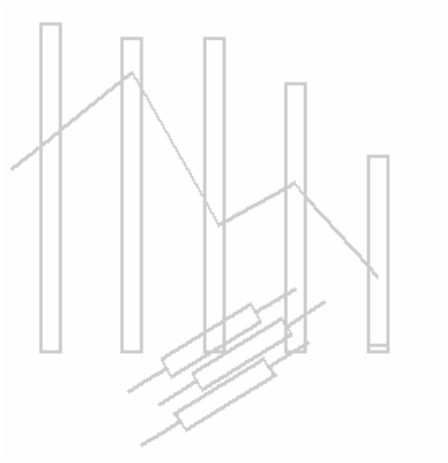




MAT4104

Statistiques 2

Mesure et collecte de données



Prétest A

Dimension 1**5 points**

Une municipalité au nord de Montréal désire implanter un parc en bordure d'une rivière. Toutefois la superficie ne permet pas que ce dernier ait plusieurs fonctions. Deux options ont alors été avancées, soit : un parc de villégiature (banc, sentier pédestre ...) ou un parc à fonction récréative (terrain de soccer et de football). La municipalité désire connaître l'opinion de ses concitoyens sur ce sujet.

Afin de connaître cette opinion, doit-on avoir recours à un sondage, un recensement ou une enquête? Justifier votre réponse.

Dimension 2**5 points**

La directrice d'une école secondaire de 1005 étudiants désire connaître l'opinion de ses élèves concernant les heures du dîner, afin de mieux établir l'horaire pour l'année prochaine.

Quelles devraient être les caractéristiques de l'échantillon pour qu'il soit représentatif de la population?

Dimension 3**5 points**

Une firme de sondage a effectué une étude portant sur la qualité du service après vente d'une entreprise. Elle a sélectionné, parmi la liste de ses clients, tous ceux dont le nom commençait par les lettres A, B, C ou D et qui habitaient la région immédiate de Montréal. Cette entreprise possède une liste de 119 572 clients répartis au Québec et dans certaines villes de la côte est des États-Unis. L'échantillon comprenait 256 clients pour une marge d'erreur de $\pm 2,5\%$.

Quel(s) élément(s) peut ou peuvent être source de biais?

Dimension 4**5 points**

Récemment une firme de sondage a interviewé 1600 personnes pour connaître le taux de satisfaction attribuable à l'implantation du virage à droite.

Les résultats obtenus sont les suivants :

CHOIX	RÉSULTATS
Pour	758
Contre	347
Indécis	495

La marge d'erreur est de $\pm 2,5\%$.

Déterminer dans quel intervalle se situe le résultat « **Pour** » si on ne tient pas compte des indécis.

Dimension 5**5 points**

Une firme d'experts en sondage croit qu'un échantillon de 2500 personnes serait satisfaisant pour connaître l'opinion du public face à la mise en marché d'un nouveau produit nettoyant. Si la marge d'erreur voulue est $\pm 2\%$ et que la population visée est de 7 millions, croyez-vous que la marge d'erreur soit compatible avec l'échantillon?

Donner les détails de la solution.

Dimension 6**5 points**

Soit la distribution suivante représentant le nombre d'employés dans la fonction publique selon leur groupe d'âge au 31 mars 2000.

Nombre d'effectifs dans la fonction publique âgé entre 30 et 65 ans

GRUPE D'ÂGE	FRÉQUENCE
[30,35[2000
[35,40[4000
[40,45[5500
[45,50[8000
[50,55[7000
[55,60[3000
[60,65[500

Source : SAGIP, 31 mars 2000

a) Calculer la médiane de cette distribution. Donner les détails de la solution.

b) Quelle en est la classe modale?

Dimension 7**10 points**

La distribution suivante représente les résultats obtenus à une épreuve de mathématique.

Résultats de l'épreuve :

5	2-9
6	9
7	2-3-3-6-7
8	0-1-2-2-5-6-7
9	2-2-3-5-5-5
10	0-0

a) Calculer la moyenne de cette distribution. Donner les détails de la solution.

b) Calculer sa médiane. Donner les détails de la solution.

c) Trouver le mode de la distribution.

d) Quel est son étendue? Donner les détails de la solution.

Dimension 8**5 points**

Les données suivantes représentent le nombre d'étudiants inscrits mensuellement à un centre d'éducation des adultes.

161 228 243 250 255 258 271 275 276 276 285 287 292 292 303 315

Calculer le rang centile de la valeur 271. Donner les détails de la solution.

Dimension 9**5 points**

Les résultats obtenus en sciences physiques lors d'une épreuve trimestrielle :

014 – 034- 038- 046- 051- 051- 052- 053- 055- 061- 064- 065- 066- 067- 068096-
[52 données]
097- 097- 097- 098- 098- 098- 098- 100- 100- 100

Quel résultat correspond à un rang centile de 93%?

Dimension 10**10 points**

Voici la liste des prix, en milliers de dollars, des voitures les plus populaires au Québec.

1	5-6-6-7-7-8-9-
2	2-2-3-5-5-7-7-8-8-9
3	0-0-4-7-7-9
4	2-3-4-6
5	5

Construire le diagramme de quartiles représentant cette distribution. Donner les détails de la solution.

Dimension 11**5 points**

Le tableau suivant représente la répartition du salaire annuel des joueurs des Sénateurs d'Ottawa en juin 2002.

Salaires en milliers\$US	Nombre de joueurs	Cumulatif
[100 - 200 [3	3
[200 - 300 [2	5
[300 - 400 [4	9
[400 - 500 [4	13
[500 - 600 [7	20
[600 - 700 [1	21
[700 - 800 [1	22
[800 - 900 [1	23
[900 - 1000 [14	37
	37	

Sources : Ottawa Senators stats for a better team, Sept2003

Vrai ou faux :

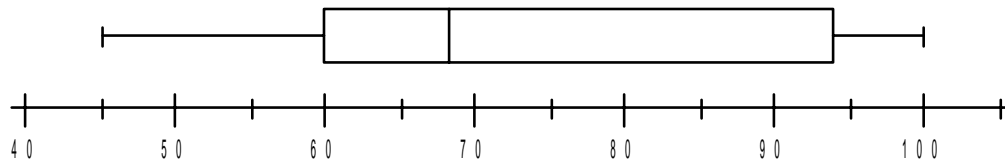
a) Le mode est la mesure de tendance centrale la plus appropriée. Justifier.

b) L'étendue de cette distribution est de 1 000 000\$US. Justifier.

c) La moyenne est de 636 486 \$US. Justifier.

Dimension 12**5 points**

La classe de Frédérick a eu un examen de mathématique la semaine dernière. Afin de leur faire travailler certains concepts statistiques, leur professeur leur a présenté les résultats sous forme de diagramme de quartiles.

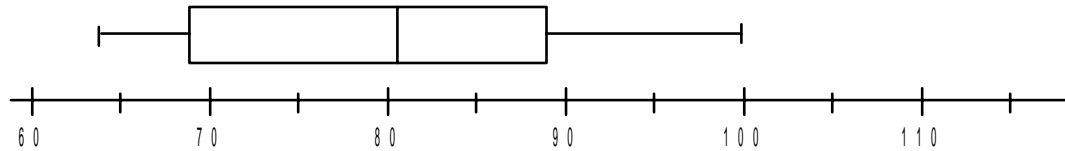


Voici une série d'affirmations. Déterminer lesquelles sont vraies et lesquelles sont fausses.
Attention! il est possible que l'affirmation proposée ne puisse être répondue.

- a) Un seul étudiant a eu un résultat de 100%.
- b) La moyenne est de 68%.
- c) 25% des étudiants n'ont pas obtenu la note de passage, soit 60%.
- d) La moitié des élèves se situent entre 60 et 94%.
- e) Plus du $\frac{3}{4}$ des étudiants ont réussi cet examen.

Dimension 13**10 points**

Le diagramme de quartiles suivant représente le nombre de décès causé par un incendie pour l'ensemble du Québec (1992 à 2001).



Source : Direction de la sécurité incendie, Ministère de la sécurité publique.

Déterminer si les affirmations suivantes sont vraies ou fausses. Justifier votre réponse par une explication ou un calcul.

- Le rang centile se rattachant à 85 décès est supérieur à 50%.
- On peut supposer que les valeurs situées entre 64 et 69 décès appartiennent au 1^{er} rang cinquième.
- Le rang centile de 75 décès est inférieur à 25%.

Dimension 14**10 points**

Le diagramme suivant représente les résultats obtenus à un examen de mathématique.

CLASSE A		CLASSE B
	3	4
3	5	
0-1-1-7-7-9	6	0-0-2-7-9
2-3-3-3-4-5-7-8-8	7	2-2-2-4-5-5-9-9-9
1-1-1-8-8-9	8	1-1-3-5-6-6
1-3	9	2-5-6-7
	10	0

Les affirmations suivantes sont-elles vraies ou fausses?

- a) La moyenne obtenue dans la classe B est plus grande que dans la classe A. Justifier votre réponse.
- b) La médiane de la classe B est plus élevée que dans l'autre classe. Justifier votre réponse.

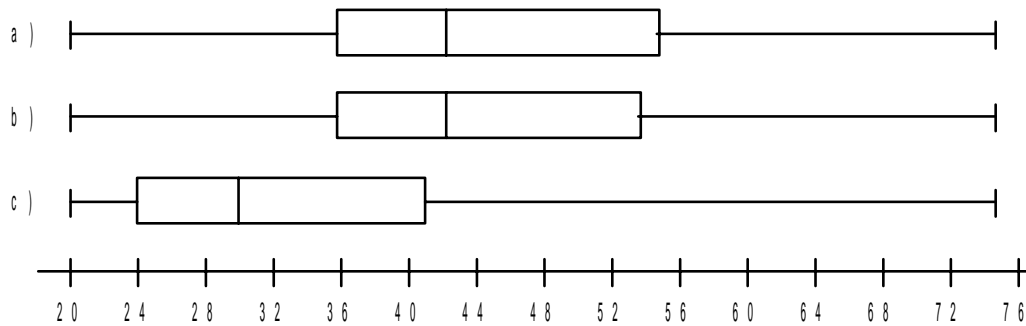
Dimension 15

10 points

Soit les deux distributions suivantes représentant les salaires annuels en milliers de dollars, des employés de deux firmes informatiques.

Associer chaque distribution au bon diagramme de quartiles et à la bonne affirmation.

COMPAGNIE A		COMPAGNIE B
0	2	0-0-0-1-2-2-3-3-4-4-4-4-5-5-5-5-6-6-6-8-9
5-5-5-3-3-3-3-2-2-1 8-8-8-6-6-6-6-6-6	3	2-3-4-6-7-8-9-9
8-8-7-7-6-6-3-1	4	0-1-1-2-5-5-7
8-6-6-6-4-4-3-2-1	5	0-7
7-4-0	6	5-8
5-1	7	2-5



Affirmations

- 1) la moyenne de cette distribution est de 35,2.
- 2) 50 % des valeurs sont en deçà de 42 000,00 \$ de salaire.
- 3) le rang centile de 47 000,00 \$ est de 83%.

	COMPAGNIE A	COMPAGNIE B
Diagramme de quartiles		
Affirmation		

ANNEXE

FORMULAIRE

Formules	Liste des symboles
<p>Taille de l'échantillon et marge d'erreur</p> $n = \frac{0,9604}{E^2}$	<p>n : nombre de données E : marge d'erreur en pourcentage</p>
<p>Médiane d'une distribution donnée en classes</p> $M_d = l_i + \frac{r}{f} \times e$	<p>M_d : médiane l_i : limite inférieure de la classe médiane r : rang de la médiane dans la classe médiane f : fréquence de la classe médiane e : étendue de la classe médiane</p>
<p>Moyenne d'une distribution donnée en classes</p> $\bar{x} = \frac{\sum f_i \times m_i}{n}$	<p>\bar{x} : moyenne \sum : somme de f_i : fréquence de chaque classe m_i : milieu de chaque classe n : nombre total de données</p>
<p>Rang cinquième</p> $R_5 = 5 \times \frac{N_{>} + \frac{N_{=}}{2}}{N_t}$	<p>R_5 : rang cinquième $N_{>}$: nombre de données de valeur supérieure $N_{=}$: nombre de données de valeur égale N_t : nombre total de données</p>
<p>Rang centile</p> $R_{100} = 100 \times \frac{N_{<} + \frac{N_{=}}{2}}{N_t}$	<p>R_{100} : rang centile $N_{<}$: nombre de données de valeur inférieure $N_{=}$: nombre de données de valeur égale N_t : nombre total de données</p>