

TP 1 : statistiques descriptives et règles d'association

Un jeune entrepreneur songe à la possibilité d'ouvrir une boutique sur Côte-des-Neiges. Sa boutique s'adressera principalement aux étudiants et prévoit vendre des vêtements et souliers pour femmes et pour hommes. Il vous contacte afin de l'aider à mieux cerner son projet.

Vous l'informez que vous avez sondé des étudiants de HEC Montréal à propos des principales marques de vêtements et souliers qu'ils achètent régulièrement. Les données de ce sondage se trouvent dans l'onglet « **Données_Association_HEC** » du fichier « **30-650-Données_TP1.xlsx** »

En plus, vous disposez d'un sondage mené à la même période auprès d'un échantillon aléatoire simple de 166 étudiants de l'Université de Montréal pour évaluer, entre autres, leur attachement à la marque de souliers qu'ils achètent. Les données de ce sondage sont disponibles dans l'onglet « **Données_sondage_UM**ⁱ » du même fichier Excel « **30-650-Données_TP1.xlsx** ».

Vous expliquez à l'entrepreneur que l'analyse de ces deux échantillons peut révéler des informations intéressantes pour évaluer le profil de la clientèle potentielle de la boutique.

Partie I : Sondage HEC

1. Veuillez fournir les statistiques descriptives relatives aux variables : sexe, âge et mode d'achat.

Étiquettes de lignes	Nombre de Sexe
Femme	108
Homme	75
Total général	183

Nombre de Sexe	Étiquettes de colonnes			
Étiquettes de lignes	En ligne	En magasin	No Answer	Total général
Femme	6	101	1	108
Homme	8	67		75
Total général	14	168	1	183

Étiquettes de lignes	Nombre de Mode_Achat
En ligne	14
En magasin	168
No Answer	1
Total général	183

<i>Âge</i>	
Moyenne	20.2185792
Erreur-type	0.1181912
Médiane	20
Mode	19
Écart-type	1.59886091
Variance de l'échantillon	2.55635621
Kurstosis (Coefficient d'aplatissement)	1.81904455
Coefficient d'asymétrie	1.53720797
Plage	8
Minimum	18
Maximum	26
Somme	3700
Nombre d'échantillons	183

2. Évaluez la popularité par sexe des marques de vêtements/Souliers chez les étudiants de HEC Montréal afin de recommander 5 marques à présenter en boutique. Commentez vos résultats.

Étiquettes de lignes	Nombre de Sexe
Femme	108
Homme	75
Total général	183

Valeurs	Femme	% de femmes	Valeurs	H	% hommes
H&M	93	86	Nike	67	89
Nike	88	81	Calvin Klein	59	79
Zara	77	71	H&M	52	69
Converse	75	69	American Eagle	45	60
Calvin Klein	68	63	Zara	42	56
American Eagle	66	61	Levi's	39	52
Old Navy/Gap/Banana Republic	51	47	Converse	26	35
Guess	46	43	Puma	25	33
Levi's	42	39	Old Navy/Gap/Banana Republic	24	32
Puma	31	29	Guess	20	27
Abercrombie&Fitch	22	20	Abercrombie&Fitch	18	24
Diesel	7	6	Diesel	16	21

Sportswear	6	6	Sportswear	8	11
Designal	4	4	Designal	4	5
Kaporal	1	1	Kaporal	1	1

Les marques les plus populaires et les moins populaires demeurent semblables chez les hommes et chez les femmes, mais pas identiques. **H&M, Nike, Zara et Calvin Klein** sont parmi les 5 marques les plus populaires chez les hommes et les femmes. En termes de proportions d'hommes et de femmes qui achètent chaque marque, American Eagle semble également populaire chez les hommes que chez les femmes.

On pourrait proposer ces 5 marques.

- Calculez et interprétez, dans le contexte du problème, le lift de la règle **Kaporal → Diesel pour tous les répondants au sondage HEC**. Commentez, s'il y a lieu, les implications managériales de cette règle.

	Diesel
Diesel	
Nike	1.08
Sportswear	1.70
Converse	0.87
Puma	0.85
Guess	1.08
Calvin Klein	1.19
Kaporal	7.96
H&M	1.15
Levi's	1.47
Zara	1.20
Abercrombie&Fitch	1.39
Designal	2.98
Old Navy/Gap/Banana	
Republic	0.85
American Eagle	0.86

Le tableau ci-dessus est extrait du tableau « lift » du gabarit « règles d'association ».

Le lift en question est très grand, mais on se rappelle (question précédente) que le nombre de répondants ayant choisi Kaporal est très petit. Donc le support de Kaporal est très petit (il est de 1.09%). Donc les conséquences managériales du grand lift sont négligeables.

- Calculez et interprétez, dans le contexte du problème, le lift de la règle **Levi's → Guess pour tous les répondants au sondage HEC**. Commentez, s'il y a lieu, les implications managériales

de cette règle.

	Diesel	Nike	Sportswear	Converse	Puma	Guess
Diesel						
Nike	1.08					
Sportswear	1.70	1.01				
Converse	0.87	1.02	1.29			
Puma	0.85	1.14	1.63	1.10		
Guess	1.08	1.06	1.58	1.24	1.24	
Calvin Klein	1.19	1.04	1.13	1.08	1.21	1.11
Kaporal	7.96	0.59	6.54	1.81	1.63	0.00
H&M	1.15	1.03	0.99	1.07	0.95	1.09
Levi's	1.47	1.06	1.45	1.12	1.13	0.92

Le tableau ci-dessus est extrait du tableau « lift » du gabarit « règles d'association ». Le lift **Levis → Guess** est inférieur à 1, donc l'association est négative. Ces marques semblent substituables pour les répondants.

5. Une stratégie serait d'offrir des marques qu'un consommateur de Calvin Klein (CK) risque également d'acheter. On cherche donc les marques (X) pour lesquelles le lift(CK=>X) est le plus grand. Les tableaux suivants sont extraits du tableau « lift » du gabarit « règles d'association » :

	Diesel	Nike	Sportswear	Converse	Puma	Guess	Calvin Klein
Diesel							
Nike	1.08						
Sportswear	1.70	1.01					
Converse	0.87	1.02	1.29				
Puma	0.85	1.14	1.63	1.10			
Guess	1.08	1.06	1.58	1.24	1.24		
Calvin Klein	1.19	1.04	1.13	1.08	1.21	1.11	
Kaporal	7.96	0.59	6.54	1.81	1.63	0.00	1.44
H&M	1.15	1.03	0.99	1.07	0.95	1.09	1.04
Levi's	1.47	1.06	1.45	1.12	1.13	0.92	1.14
Zara	1.20	1.03	1.21	1.19	0.91	1.12	1.11
Abercrombie&Fitch	1.39	1.15	0.98	0.95	1.23	1.18	1.19
Designal	2.98	1.03	0.00	1.36	1.23	1.04	1.44
Old Navy/Gap/Banana Republic							
Republic	0.85	1.09	1.74	1.23	1.18	1.07	1.10
American Eagle	0.86	1.01	1.30	1.04	1.15	1.00	1.06

Les trois plus grands lifts sont pour les marques Kaporal, Designal et Puma. Par contre Designal et Kaporal n'ont été mentionnées que par très peu de répondants (4.37% et 1.09%), donc on devrait s'intéresser à des marques plus populaires. Diesel et Abercrombie et Fitch (A&F) semblent plus intéressantes (les supports de Puma, de Diesel et de A&F et des règles entre ces marques et CK sont tous intéressants). Comme Nike est une marque très populaire et le lift est assez bon avec CK, on pourrait aussi concevoir de la recommander.

Partie II : mode d'achat

L'entrepreneur compte créer un site web pour, d'une part, promouvoir sa boutique sur les réseaux sociaux et, d'autre part, vendre ses articles en ligne.

6. Comparez le mode d'achat des étudiants de HEC Montréal à celui des étudiants de l'Université de Montréal.

UdeM :

Étiquettes de lignes	Fréquence absolue	Fréquence relative
Magasins spécialisés	152	92%
Grande surface	14	8%
Total général	166	100%

HEC :

Étiquettes de lignes	Nombre de Mode_Achat
En ligne	14
En magasin	168
No Answer	1
Total général	183

Il faut noter que les répondants UdeM pouvaient répondre « sur internet » mais qu'aucun répondant n'a choisi cette option.

Considérant les options « magasins spécialisés » et « grandes surfaces » (UdeM) comme étant équivalentes à l'option « en magasin » (HEC), on peut dire que le mode d'achat des répondants est dans les deux cas principalement en magasin.

7. En vous basant sur les résultats de la question précédente, recommanderiez-vous à l'entrepreneur le mode de vente en ligne?

Nous n'avons que 2 sondages d'étudiants, un avec 167 répondants et l'autre avec 183 répondants. Ces répondants n'utilisent que très peu internet pour acheter des vêtements. Si ces répondants sont typiques de notre clientèle, le sondage nous porte à croire qu'une boutique en ligne n'amènerait pas beaucoup ventes en ligne. Par contre ceci ne mesure pas l'impact éventuel d'une boutique en ligne sur les ventes en magasin (les clients pourraient vouloir « magasiner » en ligne mais acheter en magasin). Il faut voir également quels sont les coûts de création et de maintien d'une boutique en ligne; s'ils sont négligeables, l'option peut demeurer intéressante.

8. Utilisez les commentaires fournis par les étudiants de HEC Montréal pour expliquer le manque de popularité de l'achat de chaussures en ligne. Attention! Vous devez **quantifier** vos affirmations autant que faire se peut.

On peut tenter de regrouper les commentaires en catégories et faire ressortir le type de commentaires le plus courant. Par exemple :

Étiquettes de lignes	Nombre de Commentaire - Type
Possibilité de toucher / essayer	46
Autre malaise par rapport à internet	13
Autre raison	2
Facilité d'achats en magasin	1
Rabais en magasin	1
(vide)	
Total général	63

Partie III : Attachement à la marque

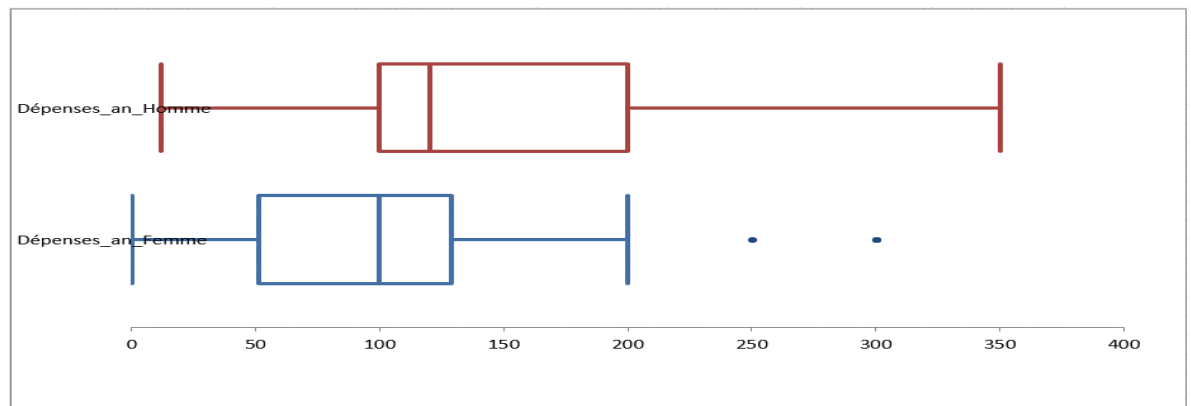
Pour répondre aux questions de cette partie du rapport, vous avez besoin des données dans l'onglet « **Données_Sondage_UM** ». Notez que pour les variables mesurées sur une échelle ordinale, un score de **4 et plus** signifie un résultat satisfaisant à très satisfaisant ou important à très important.

9. Comparez la distribution des dépenses annuelles moyennes en souliers des hommes et des femmes (fournissez au moins 4 indicateurs pertinents et deux graphiques appropriés).

	Dépenses_an_Femme	Dépenses_an_Homme
Moyenne	101.9	134.32
Médiane	100	120
Mode	100	100
Écart-type	66.09	71.29

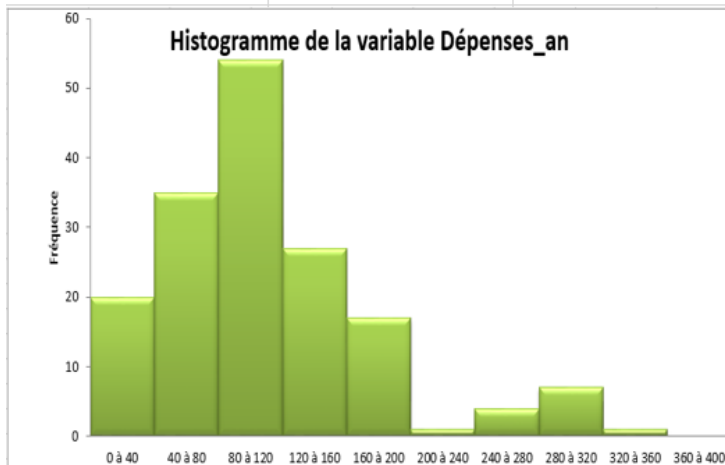
Variance de l'échantillon	4368.01	5081.98
Coefficient d'asymétrie	1.21	0.91
Plage	300	338
Minimum	0	12
Maximum	300	350
Somme	9984.32	9133.66
Nombre d'échantillons	98	68

10. Discutez de la symétrie des distributions des dépenses annuelles moyennes chez les hommes et les femmes.



Dépenses par an

10 classes

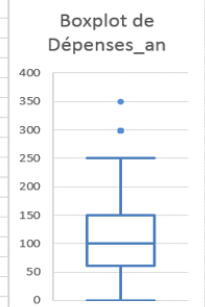


Quantiles	Femme	Homme
Min	- \$	12 \$
5%	20 \$	50 \$
25%	51 \$	100 \$
50%	100 \$	120 \$
75%	129 \$	200 \$
95%	250 \$	283 \$
Max	300 \$	350 \$

Statistiques descriptives							
Variable	3	Taille d'échantillon	166				
Sujet	1						
Mode Achat	2	Mesures de centralité		Quantiles			
Dépenses_an	3	Moyenne	115.168554	Min	0		
Fréquence_an	4	Médiane	100	5%	23.7475		
Durée_vie	5			25%	61.25		
Importance_marqu	6	Mesures de dispersion		50%	100		
Usage	7	Écart-type	69.9103493	75%	150		
Prix	8	Étendue	350	95%	250		
Qualité	9	Étendue inter-quartile	88.75	Max	350		
Intention d'achat	10						
Fidélité	11	Mesures de géométrie		70%	130		
Sexe	12	Asymétrie	1.03279663				
Age	13	Aplatissement	1.02910668				
Revenu	14						
	15	Probabilités empiriques					
		Valeur d'intérêt :	80				
		P(Dépenses_an < 80) =	0.31325301				
		P(Dépenses_an <= 80) =	0.3313253				
		P(Dépenses_an > 80) =	0.6686747				
		P(Dépenses_an >= 80) =	0.68674699				

Boxplot de Dépenses_an

Statistic	Value
Min	0
Q1	60
Median	100
Q3	150
Max	250
Outliers	300, 350



11. Combien de répondants déclarent acheter des souliers au moins une fois par an?
Comment se comparent les hommes et les femmes à cet égard?

Nombre d'achat par année (H+F)	Nombre de Sujet	%	% cumulée
0	2	1.20%	1.20%
0.16	1	0.60%	1.81%
0.2	3	1.81%	3.61%
0.25	1	0.60%	4.22%
0.33	5	3.01%	7.23%
0.4	6	3.61%	10.84%
0.5	27	16.27%	27.11%
0.66	2	1.20%	28.31%
1	87	52.41%	80.72%
1.5	8	4.82%	85.54%
2	22	13.25%	98.80%
2.5	1	0.60%	99.40%
3	1	0.60%	100.00%
Total général	166	100%	

71.69% des répondants déclarent acheter des souliers au moins une fois par an.

Fréquence d'achat_F	NB	%
0	2	2.0%
0.2	3	3.1%
0.33	5	5.1%
0.4	4	4.1%
0.5	19	19.4%
1	49	50.0%
1.5	6	6.1%
2	10	10.2%
Total général	98	100.0%

des répondantes déclarent acheter des souliers au 66.3% moins une fois par an

Fréquence d'achat H	NB	%
0.16	1	1.5%
0.25	1	1.5%
0.4	2	2.9%
0.5	8	11.8%
0.66	2	2.9%
1	38	55.9%
1.5	2	2.9%
2	12	17.6%
2.5	1	1.5%
3	1	1.5%
Total général	68	

des répondants déclarent acheter des 79.4% souliers au moins une fois par an

12. Quelle est la proportion de femmes qui achètent des chaussures à rabais parmi les femmes?

Nombre de Sujet	Étiquettes de colonnes		
Étiquettes de lignes	Homme	Femme	Total général
Sans escompte	50.00%	45.92%	47.59%
Avec escompte	50.00%	54.08%	52.41%
Total général	100.00%	100.00%	100.00%

13. Quelle proportion d'hommes achètent des chaussures à rabais parmi les répondants qui cherchent les aubaines?

Nombre de Sujet	Étiquettes de colonnes		
Étiquettes de lignes	Homme	Femme	Total général
Sans escompte	43.04%	56.96%	100.00%
Avec escompte	39.08%	60.92%	100.00%
Total général	40.96%	59.04%	100.00%

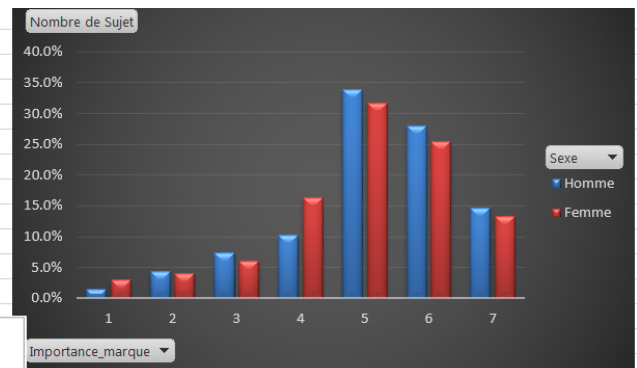
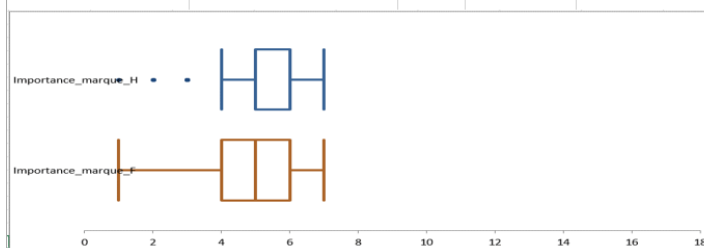
14. Si on tire au hasard un individu de l'échantillon, quelles sont les chances que cet individu soit une femme qui achète des chaussures sans escompte?

Nombre de Sujet	Étiquettes de colonnes		
Étiquettes de lignes	Homme	Femme	Total général
Sans escompte	20.48%	27.11%	47.59%
Avec escompte	20.48%	31.93%	52.41%
Total général	40.96%	59.04%	100.00%

15. Qui accorde plus d'importance à la marque des chaussures : les hommes ou les femmes? Justifiez votre réponse.

Notons que l'importance à la marque est donnée selon une échelle de 1: Aucune importance à 7: Très grande importance.

Importance de la marque			
Nombre de Sujet	Étiquettes de colonnes		
Étiquettes de lignes	Homme	Femme	Total général
1	1.5%	3.1%	2.4%
2	4.4%	4.1%	4.2%
3	7.4%	6.1%	6.6%
4	10.3%	16.3%	13.9%
5	33.8%	31.6%	32.5%
6	27.9%	25.5%	26.5%
7	14.7%	13.3%	13.9%
Total général	100.0%	100.0%	100.0%



16. Qu'en est-il pour la fidélité à une seule marque? Justifier votre réponse.

Notons que les répondants doivent dire s'ils sont fidèles à une seule marque de souliers en répondants sur une échelle allant de 1 à 7. 1: Pas du tout... 7: Tout à fait.

