Suite de series des exercices de TD

Sciences de données; groupe ScFo 11.1

Ce document recense l'ensemble des questions non résolues pendant les heures de TD. Cela suit de très près la chronologie des points abordés courant les heures de TD.

1 Analys descriptive

4. Quel est le pourcentage de ménages ayant 2 voitures? Combien de ménages n'ont pas de voiture ? Quel est le pourcentage de ménages ayant au moins 1 voiture ? Combien de ménages ont au plus 2 voitures ?

Solution: Le pourcentage de ménages ayant 2 voitures est de 36.66 %. Il y a 6 ménages sans voiture sur les 30 Il y a $24 * 100/30 = 80\%$ des ménages qui ont au moins une voiture. Il y a 26 ménages sur 30 qui ont au plus 2 voitures (donc 0, 1 ou 2).

2 Autour des sommes géométriques

2. On organise un tournoi de tennis, pour lequel 32 joueurs sont inscrits. Le tournoi s'effectue en seizièmes, huitièmes, quarts, demis et finale. Combien de matchs sont nécessaires pour désigner le vainqueur ?

Solution: Il y a 16 seizièmes, 8 huitièmes, 4 quarts, 2 demis et 1 finale donc le nombre de matchs nécessaires pour désigner le vainqueur est leur somme, soit $16 + 8 + 4 + 2 + 1 = 31$.

3. Imaginons que l'on ait 32 sprinteurs dont on veut trouver le meilleur. On propose la procédure suivante : ils effectuent une première course et le dernier est éliminé. Puis une deuxième et le dernier est éliminé, etc. Le vainqueur de la dernière course (à 2 coureurs donc) est déclaré meilleur sprinteur. Combien de courses sont nécessaires pour désigner le vainqueur ? Comparer au résultat de la question précédente.

Solution : On élimine 1 sprinteur à chaque course et il faut tous les éliminer sauf 1, donc il faut faire 31 courses. Dans la question précédente, puisque chaque match de tennis éliminait exactement un joueur et qu'on voulait tous les éliminer sauf 1, c'était exactement la même chose.

4. On reprend le tournoi de tennis à 32 joueurs. Combien y a-t-il de déroulements possibles du tournoi, sachant que la place des joueurs sur la feuille de match est fixée ?

Solution : Le nombre S de déroulements possibles du tournoi est

$$S = 2^{16}2^{8}2^{4}2^{2}2^{1} = 2^{31}$$

3 Application de la décomposition de la variance

Solution : On prend la moyenne de chaque intervalle comme valeur de temps, le calcul donne :

$$\overline{x}_F = \frac{1}{n_F}\sum_{j=1}^{n_F} x_{F,j} = (80 + 18 \times 140 + 230 \times 71)/90 = 210.33.$$

Pour la variance (empirique) :

$$\frac{1}{n_F}\sum_{j=1}^{n_F}(\overline{x}_F - x_{F,j})^2 = ((80-210.33)^2+(18\times(140-210.33)^2)+(230-210.33)^2\times 71)/90 = 1483.22.$$

2. Retrouver, par le calcul, que la part de variabilité du temps consacrée aux tâches domestiques qui est expliquée par le sexe est de $71,92\%$. Vous utiliserez les éléments de calculs adéquats fournis dans le tableau ci-dessous.

Temps consacré aux tâches domestiques	Moyenne	Variance	Ecart-type
Femme	210.3	1483.22	38.51
Homme	75.07	1874.25	43.29
Total	129.42	6114.40	78.19

Solution :

$$n_F(\overline{x}_F - \overline{\overline{x}})^2 + n_H(\overline{x}_H - \overline{\overline{x}})^2 = 90 \times (210.33 - 129.42)^2 + 134 \times (75.07 - 129.42)^2 = 985004.1$$

Puis : la dispersion globale est $6114.40 \times 224 = 1369626$ donc on voit que : $985004.1 \times 100/1369626 = 71.91774$ c'est à dire que la dispersion due au sexe est de 71.92%.

3. Interpréter dans ce cadre la formule de décomposition de la variance.

Solution : On voit donc que un peu plus de 28% de la variance est due à la variabilité au sein du groupe des hommes et au sein du groupe des femmes. Le sexe semble avoir un réel impact sur le temps consacré aux tâches ménagères.

1.5 Tables de contingence

1.5.1 Éducation

On se propose de décrire l'évolution de l'éducation des enfants d'une génération à l'autre. L'étude porte sur un échantillon d'individus âgés de 25 à 40 ans mariés et ayant des enfants. On dispose pour chaque individu d'un couple d'observations : (le type d'éducation que cet individu donne à ses enfants ; le type d'éducation que cet individu a reçu quand il était enfant). On considère ainsi deux variables :

Variable x : *Education actuelle*, le type d'éducation que cet individu donne à ses enfants.

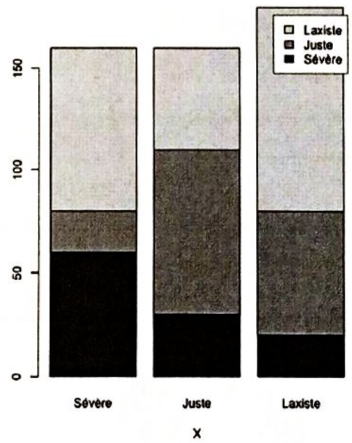
Variable y : *Education d'origine*, le type d'éducation que cet individu a reçu quand il était enfant.

On dispose du tableau de contingence en effectifs de ces deux variables :

x \ y	Sévère	Juste	Laxiste
Sévère	60	30	20
Juste	20	80	60
Laxiste	80	50	100

1. Donner la population étudiée, l'unité statistique, les variables étudiées et leur type. Quels diagrammes sont adaptés pour la représentation graphique de telles variables ?

Solution : Nous étudions des variables qualitative sur le type d'éducation de parents et d'enfants. Il s'agit de deux variables qui prennent chacune 3 modalités "Sévère" Juste", "Laxiste". On pourra représenter ces variables par des diagrammes circulaires ou en barres par exemple :



2. Déterminer le tableau de contingence en fréquences des deux variables étudiées et donner les fréquences marginales.

Solution : Effectifs marginaux

y / x	Sévère	Juste	Laxiste	n_X
Sévère	60	30	20	110
Juste	20	80	60	160
Laxiste	80	50	100	230
n_Y	160	160	180	500

Fréquences marginales (en tout il y a $n = 500$ observations)

y / x	Sévère	Juste	Laxiste	f_X
Sévère	12%	6%	4%	22%
Juste	4%	16%	12%	32%
Laxiste	16%	10%	20%	46%
f_Y	32%	32%	36%	100%

3. Quelle est la proportion d'individus qui ont reçu une éducation laxiste *et* éduquent pareillement leurs enfants ?

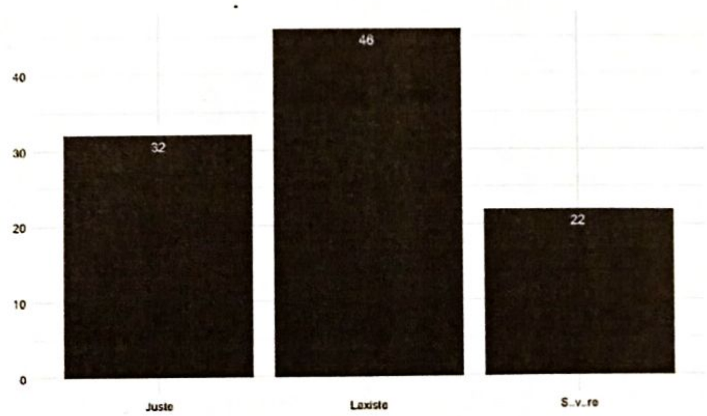
Solution : 20%

4. Donner la proportion d'individus ayant reçu une éducation laxiste qui éduquent ainsi leurs enfants.

Solution : 55.56%

5. Représenter par un diagramme en barres les fréquences marginales des modalités de la variable *Éducation actuelle.*

Solution :



6. Donner les fréquences conditionnelles de la variable *Éducation d'origine* sachant la variable *Éducation actuelle.*

Solution :

$f_{y\|x}$	Sévère	Juste	Laxiste	Total
Sévère	0.545	0.273	0.182	1
Juste	0.125	0.500	0.375	1
Laxiste	0.348	0.217	0.435	1

7. Représenter par des diagrammes en barres les fréquences conditionnelles de la variable *Education actuelle* sachant la variable *Education d'origine.*

Solution :

$f_{x\|y}$	Sévère	Juste	Laxiste
Sévère	0.375	0.188	0.111
Juste	0.125	0.500	0.333
Laxiste	0.500	0.312	0.556
Total	1	1	1