

Evaluation Individuelle TP 2018-2019 : Solution Sujet 2

NOM :

GPE :

Les données utilisées dans le sujet sont dans le fichier `mtcars` et les notations décrites dans `AideExam.pdf` sont à utiliser impérativement ainsi que les notations suivantes :

Notations :

- En 1973 un véhicule est considéré comme économique lorsque `mpg > 20` et au contraire gourmand lorsque `mpg ≤ 20`. En 2018 une étude a montré que la proportion de véhicules économiques était $p_0 = 0.60$ et en 2017 on a montré que la proportion de véhicules gourmands était $p_1 = 0.35$.
- On notera p la proportion de véhicules économiques en 1973 et p' celle de véhicules gourmands en 1973. Ces deux paramètres sont inconnus.

Questions :

1. (3pts) Remplir le tableau suivant donnant les estimations sans biais de p et p' ainsi que les intervalles de confiance de niveau 90% symétriques et calculés par la procédure `prop.test`:

Paramètre	taille éch.	e.s.b.	Borne Inf IC à 90%	Borne Sup IC à 90%
p	32	0,4375	0,3038	0,5810
p'	32	0,5625	0,4190	0,6962

2. (5pts) On veut savoir si la proportion de véhicules gourmands était la même en 1973 qu'en 2017.

(a) Quel test faites vous ? test no : **3 : test sur une proportion**

(b) Poser les hypothèses du test :

$$\mathcal{H}_0 : p' = p_1 \qquad \mathcal{H}_1 : p' \neq p_1$$

(c) Quelles conditions doit on vérifier si on utilise le test asymptotique (celui programmé sur les calculatrices) :

$np_1 = 11,2 > 10$ et $n(1 - p_1) = 20,8 > 10$ on pourrait donc utiliser le test asymptotique

(d) Donner la ligne de commande R permettant de réaliser le test avec `prop.test` et sans correction de continuité:

```
prop.test(sum(mpg<=20),length(mpg<=20),p=p1,conf.level=0.9,correct=F)
```

(e) Que vaut la p-valeur du test et que décide-t-on pour $\alpha = 5\%$?

$$p - val = 1,2\% \quad \text{on décide } \mathcal{H}_1 : p' \neq p_1 \quad \text{car } 5\% > p\text{-valeur}$$

(f) Quel test unilatéral suggère cette conclusion ? Donner la commande R permettant de le réaliser: **Unilatéral supérieur (c. à d. $\mathcal{H}_1 : p' > p_1$)**:

```
prop.test(sum(mpg<=20),length(mpg<=20),p=p1,correct=F,alternative ="greater")
```

- (g) Conclusion littérale de ce dernier test : **On peut conclure de façon statistiquement significative (avec risque sup à 0,6%) que la proportion de véhicule gourmands a baissé entre 1973 et 2017.**

3. (8pts) On souhaite à présent savoir si il y a un lien entre nombre de cylindres et nombre de vitesses dans les moteurs de 1973. On notera X la variable aléatoire : nombre de cylindres et Y nombre de vitesses

- (a) Calculer les répartitions de X conditionnellement à Y , compléter le tableau suivant et indiquer la commande R utilisée: `prop.table(table(cyl,gear),2)`

cond. Y	3	4	5
rep. X			
4	0.0667	0.6667	0.40
6	0.1333	0.3333	0.2
8	0.8	0	0.4
total	1	1	1

- (b) Représenter les trois distributions conditionnelles avec des diagrammes en barres sur un même graphique. Donner la commande R :

`barplot(prop.table(table(cyl,gear),2),beside=T) ...`

Interpréter : **Au dessus de 3 la répartition du nb de cylindres pour les moteurs à 3 carburateurs, au dessus de 5 pour les moteurs à 5 carburateurs et au dessus du 7 pour ceux à 7 carb. : elles sont très différentes les unes des autres. Il y a manifestement un lien entre les deux variables étudiées.**

- (c) On veut vérifier l'interprétation précédemment donnée avec un test. Quel test faites-vous ? Test no : **7 : test d'indépendance du chi-deux**
- (d) Compléter

\mathcal{H}_0 : X et Y indépendantes

\mathcal{H}_1 : X et Y dépendantes

- (e) Compléter le tableau des effectifs **attendus** si \mathcal{H}_0 vraie et indiquer la commande R permettant de les obtenir :

`chisq.test(table(cyl,gear))$expected`

Y	3	4	5	total
X				
4	5.15625	4.125	1.71875	11
6	3.28125	2.625	1.09375	7
8	6.56250	5.250	2.18750	14
total	15	12	5	32

- (f) Donner la p-valeur : $pval = 0,12\%$ et la conclusion littérale de ce test : **On peut conclure avec un faible risque de se tromper (> à 0,12%) que le nombre de cylindres et le nombre de vitesses sont liés.**