
Evaluation Individuelle TP 2018-2019 : Solution Sujet 4

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 :

- Un véhicule sera dit économique lorsque `mpg`>20 et dit gourmand lorsque `mpg`<=20
- On notera X la variable `qsec` sur la sous-population des véhicules économiques et Y sur la sous-population des véhicules gourmands.
- Les moyennes et variances de X et Y sont inconnues et seront notées resp. μ_X, μ_Y, σ_X^2 et σ_Y^2 .
- On notera Z la variable indicatrice de boîte manuelle pour les véhicules de 1973 et p son espérance.
- On sait qu'en 2018 la répartition entre boîtes manuelles et boîtes automatiques sur tous les véhicules du marché est connue et donnée par 3/4 de manuelles et 1/4 d'automatiques.

Questions :

1. (3pts) Construire les échantillons de X et Y et les affecter à `x` et `y`. Remplir le tableau suivant donnant les estimations sans biais de σ_X^2 et σ_Y^2 ainsi que les intervalles de confiance de niveau 90% (on supposera que X et Y suivent des lois normales):

Paramètre	taille éch.	e.s.b.	Borne Inf IC à 90%	Borne Sup IC à 90%
σ_X^2	14	2,9698	1,7255	6,5527
σ_Y^2	18	2,1809	1,3439	4,2754

2. (7pts) On veut savoir si en moyenne, qu'un véhicule soit économique ou non, a un effet sur `qsec`.

- (a) Donner la ligne de commande R permettant de réaliser un graphique où figurent les boxplot des deux échantillons : `boxplot(x,y)`

Interpréter : **dispersions semblables mais médiane de x plus grande que celle de y, pas d'indiv hors norme : les moyennes sont peut-être différentes, à tester**

- (b) Quel test faites-vous ? test no : **5 : comparaison de moyennes avec 2 éch indep**
(c) Poser les hypothèses du test :

$$\mathcal{H}_0 : \mu_X = \mu_Y \qquad \mathcal{H}_1 : \mu_X \neq \mu_Y$$

- (d) Quelles conditions doivent satisfaire X et Y pour pouvoir utiliser ce test (modèle) ?
:
 X et Y suivent des lois normales et sont de même variance

- (e) Donner la ligne de commande R permettant de réaliser le test :
`t.test(x,y,var.equal=T)`

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

$$p - val = 0,49\% \quad \text{on} \quad \text{décide} \quad \mathcal{H}_1 : \mu_X \neq \mu_Y \quad \text{car} \quad 5\% > pval$$

- (g) Quel test unilatéral suggère cette conclusion ? Donner la commande R permettant de le réaliser: **Unilatéral supérieur (puisque $\hat{\mu}_X > \hat{\mu}_Y$)**
`t.test(x,y,var.equal=T,alternative = "greater")`

- (h) Conclusion littérale de ce dernier test : **On concluera avec un risque de se tromper faible ($> 0,24\%$) qu'en moyenne qsec est plus élevée chez les véhicules économiques que chez les véhicules gourmands.**

3. (2pts) Le test précédent a été réalisé sous condition d'égalité des variances de X et Y . Faire le test d'égalité des variances pour vous assurer que c'est une condition raisonnablement satisfaite. Indiquer la commande R :

.....

et votre conclusion :

4. (4pts) On souhaite à présent savoir si la répartition de la variable type de boîte de vitesse en 1973 est la même que celle bien connue de 2018.

- (a) Calculer la répartition observée de **am** dans l'échantillon tiré parmi les véhicules de 1973. Compléter le tableau suivant :

modalités de am	0	1	total
freq. obs.	0,5938	0,4062	1

- (b) Représenter sur un même graphique et côte à côte la répartition observée et la répartition connue pour les véhicules de 2018. Donner la commande R:
`barplot(rbind(prop.table(table(am)),c(3/4,1/4)),beside=T)`

Interpréter : **La répartition observée (en noir) semble assez différente de la répartition connue en 2018 (en gris). Il n'y a probablement pas adéquation, à vérifier par un test.**

- (c) On veut vérifier l'interprétation précédemment donnée avec un test. Quel test faites-vous ? Test no : **8 : test d'adéquation du chi-deux**

- (d) Compléter

$$\mathcal{H}_0 : \quad \text{la répart de am en 1973 est } (3/4, 1/4)$$

$$\mathcal{H}_1 : \quad \text{la répart de am en 1973 est différente de } (3/4, 1/4)$$

- (e) Indiquer la commande R permettant de faire le test:
`chisq.test(table(am),p=c(3/4,1/4))`

- (f) Donner la p-valeur : $pval = 4,12\%$ et la conclusion littérale de ce test : **On peut conclure avec un faible risque de se tromper ($>4,12\%$) que la répartition du type de boîte de vitesse n'est pas la même en 1973 qu'en 2018**