Université Grenoble Alpes **Frédérique Leblanc**
L2/STA351 – TP

Evaluation Individuelle TP 2018-2019 : Solution Sujet 5

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 :

- On notera Y la variable ayant donné l'échantillon `disp` de moyenne μ_Y et variance σ_Y^2 inconnues et X la variable poids du véhicule (ech. `wt`) de moyenne et variance notées μ_X et σ_X^2.
- On pose $X_1 = 100(X - 1)$ et $X_2 = 100X$ d'espérances notées resp. μ_1, μ_2.

Questions :

1. L'échantillon de Y est disponible dans `disp`. Construire l'échantillon de X_1 et l' affecter à `x1`. Indiquer la commande R exécutée pour créer `x1`:
   `x1<-100*(wt-1)`

2. (3pts) Remplir le tableau suivant donnant les estimations sans biais de μ_Y et μ_1 ainsi que les intervalles de confiance de niveau 90% (X_1 et Y seront supposées normales):

Paramètre	taille éch.	e.s.b.	Borne Inf IC à 90%	Borne Sup IC à 90%
μ_Y	32	230,7	193,6	267,9
μ_1	32	221,7	192,4	251,1

3. (7pts) On veut savoir si la moyenne de Y est semblable ou pas à celle de X_1.

   (a) Proposer un graphique permettant de visualiser grossièrement les répartitions de Y et X_1 dans une même fenêtre. Indiquer la commande R :
   `boxplot(disp,x1)`
   Interpréter ce graphique : **des médianes semblables mais des dispersions assez différentes, qqs indiv hors norm en x1 donc les moyennes sont probablement égales (à vérifier avec un test)**

   (b) Représenter la répartition de la variable $D = Y - X_1$ et y ajouter la courbe de la densité d'une loi normale dont on choisira les paramètres en fonction de l'échantillon observé :
   Commande R pour la répartition (sans détails) :
   `hist(disp-x1,prob=T)`
   Commande R pour l'ajout de la densité normale : `curve(dnorm(x,mean(disp-x1),sd(disp-x1)),col`
   Interpréter : **le modèle normal est raisonnable**

   (c) Quel test faites vous pour répondre au problème posé ? test no : **4 : comp de moyennes avec 2 echs appariés**

(d) Poser les hypothèses du test :

$$\mathcal{H}_0 : \quad \mu_Y = \mu_1 \qquad\qquad \mathcal{H}_1 : \quad \mu_Y \neq \mu_1$$

(e) Sous quelle hypothèse de modélisation peut-on faire ce test ? :
$Y - X_1$ **suit une loi normale**

(f) Donner la ligne de commande R permettant de réaliser le test :
`t.test(disp,x1,paired=T)`

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

$$p - val = 38,9\% \qquad \text{on} \qquad \text{décide} \quad \mathcal{H}_0 : \mu_Y = \mu_1 \quad \text{car} \qquad 5\% < pval$$

4. (6pts) On souhaite à présent savoir si il y a un lien entre nombre de cylindres et nombre de carburateurs dans les moteurs de 1973. On notera X la variable aléatoire : nombre de cylindres et Y nombre de vitesses (ech. `gear`)

(a) Calculer les effectifs observés pour tous les couples de modalités et compléter le tableau suivant. Indiquer la commande R utilisée pour produire la table de contingence :
`table(cyl,gear)`

Y X	3	4	5	total
4	1	8	2	11
6	2	4	1	7
8	12	0	2	14
total	15	12	5	32

(b) Quel test faites-vous pour répondre au problème posé ? Test no : **8 : test d'indépendance du chi-deux**

(c) Compléter

$$\mathcal{H}_0 : \quad X \;\; et \;\; Y \;\; \text{indépendantes} \qquad\qquad \mathcal{H}_1 : \quad X \;\; et \;\; Y \;\; \text{liées}$$

(d) 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 X	3	4	5	total
4	5,156	4,125	1,719	11
6	3,281	2,625	1,094	7
8	6,562	5,250	2,188	14
total	15	12	5	32

(e) Donner la p-valeur : $pval = 0,12\%$ et la conclusion littérale de ce test :
On peut concluredе façon statistiquement significative (avec un faible risque de se tromper car > 0,12%) que nombre de vitesses et nombres de cylindres sont liés.