UGA
L2 STA401 – TP

Frédérique Leblanc
frederique.leblanc@univ-grenoble-alpes.fr

Exemples d'exercices pour l'évaluation individuelle N2 : Statistiques

Les données analysées ici sont celles proposées dans le fichier `mtcars` du package de base de R. Les questions peuvent être traitées séparément même si leur ordre n'est pas sans logique. Les deux dernières questions portent sur des calculs de probabilités et sont totalement indépendantes des premières qui elles portent sur une analyse statistique des données de `mtcars`.

La durée prévue pour cette liste d'exercices est d'environ 75 minutes et le barême n'est qu'indicatif pour vous permettre de gérer votre temps et serait ramené sur 20. Dans les questions type QCM il peut y avoir plusieurs bonnes réponses parmi celles proposées et toute réponse sélectionnée à tort induira une pénalité.

Notations : Dans ce sujet on se propose d'étudier les variables `disp` qui sera notée Y et `cyl` qui sera notée X.

On notera μ_Y (resp. μ_X) pour la moyenne théorique de Y (resp. X) et σ^2_Y (resp. σ^2_X) pour la variance théorique de Y (resp. X).

1. (4 pts) Indiquez dans le tableau ci-dessous les résumés numériques associés à la variable Y étudiée (à 10^{-2}) :

taille de l'échan n	moyenne emp. $\bar{y}$	var. emp. s^2	var. emp. corrigée s'^2
32	230.7	14880.77	15360.8
1er quart.	second quart.	3ième quart	ecart-type emp. corrigé s'
120.8	196.3	326	123.9

2. (2pts) Pour utiliser les intervalles de confiance ou les tests sur μ_Y, (ceux vus en cours), pour n'importe quelle taille d'échantillon on doit s'assurer que Y suit une loi normale. Sinon en cas de non-normalité de l'échantillon, on ne peut appliquer les méthodes que pour un échantillon assez grand (de taille au moins 30). Quel graphique permet-il de diagnostiquer la normalité ? Indiquer les instructions R permettant de le réaliser et le commenter :

qqnorm(y) ; qq line (y)
alignement au milieu mais pas aux extrémités
⟹ l'échantillon n'est pas issu d'une v.a. Y normale.
pas grave puisque $n_Y \geq 30$. (cadre asympt)

3. (5 pts) Donnez les estimations sans biais de μ_Y et σ^2_Y. et les intervalles de confiance de niveaux 95% et 80% pour le paramètre μ_Y dans le tableau suivant :

est. ss biais de μ_Y	est. ss biais de σ^2_Y
230.7	15360.8
borne inf. IC niv 95% pour μ_Y	borne sup. IC niv 95% pour μ_Y
186.0	275.4
borne inf. IC niv 80% pour μ_Y	borne sup. IC niv 80% pour μ_Y
202.0	259.4

4. (2 pts) On peut décrire la répartition observée de la variable X à l'aide de : `x<-mtcars$cyl` suivi de :

   1. `barplot(x,prob=T)` 2. `pie(table(x))` 3. `barplot(table(x)/length(x))`
   4. `hist(table(x))` 5. `hist(x,prob=T)` 6.`pie(x,freq=T)`

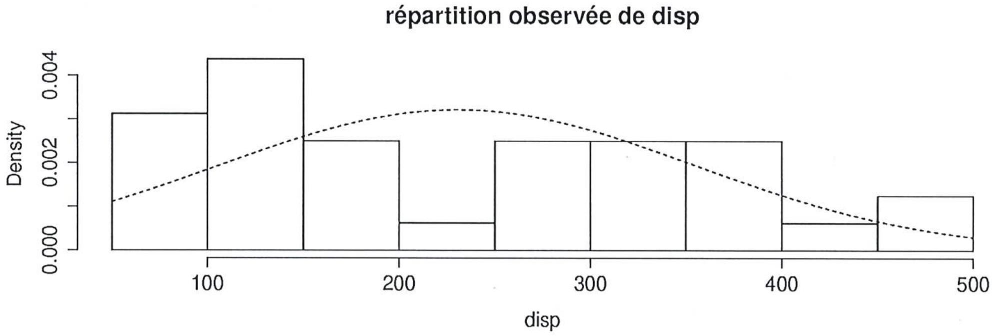
5. (5pts) Pour réaliser le graphique proposé ci-dessous, on exécute la suite de commandes :

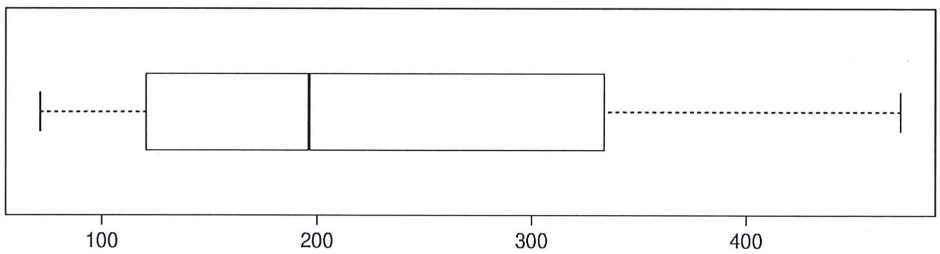
par(mfrow=c(2,1))................................# part.fen.graph.

hist(y, prob=T, main="répart...", xlab="disp")# répart.obs.disp

curve(dnorm(~~x~~ mean(y), sd(y)), add=TRUE, lty=2)# ajout dens.norm.

boxe.plot(y, horizental=TRUE)......................# boxplot horiz.





6. (6,5 pts) On veut effectuer un test d'égalité de la variance de Y à la valeur cible $\sigma_0^2 = 15000$.

Quelles conditions doit on poser sur le modèle décrivant Y et sur la taille de l'échantillon pour pouvoir appliquer le test sur une variance vu en cours ? (1pt)

$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ ou $n \geq 30$.

Préciser le choix des hypothèses du test à mettre en oeuvre (1pt) :

$\mathcal{H}_0$: $\sigma_Y^2 = \sigma_0^2$ .... $\mathcal{H}_1$: $\sigma_Y^2 \neq \sigma_0^2$ ....

Indiquer les instructions R utilisées pour réaliser les calculs (2pts) :

La stat de test est $T = \frac{nS_Y^2}{\sigma_0^2}$ T ← 32 * var(y)/15 000

La p-valeur pval = $\min\left(2P(\chi^2_{n-1} < t_{calc}), 2(P(\chi^2_{n-1} > t_{calc}))\right)$

Donner les valeurs de la statistique et de la p-valeur du test (1pt):

tcalc ← 31*var(y)/15000 ; pval = min(2*pchisq(tcalc, 31), 2*(1-pchisq(tcalc, 31)))

$tcalc =$ ....31,7.... $pval =$ ....85,8 %....

Que conclut-on dans un test de niveau $\alpha = 5\%$ (justifier la réponse) (1,5pts) ?

Comme $\alpha <$ p-val, dans un test de niveau $\alpha = 5\%$ on ne rejette pas $\mathcal{H}_0$: $\sigma_Y^2 = 15000$, et on peut considérer la variance connue et valant $\sigma_Y^2 = 15000$.

7. (7pts) On veut à présent savoir si en moyenne la variable `disp` est conforme ou non à une valeur de référence donnée $\mu_0 = 280$.

On se propose de faire le test (1 seul choix possible 0,5pt)

1. test sur μ_Y avec σ_Y connu	2. test sur σ_Y avec μ_Y inconnu
3. test sur $p = P(Y > 250)$	4. test de comp. entre μ_Y et μ_X
5.test sur μ_Y avec σ_Y inconnu	6. test de comp. entre σ_X et σ_Y

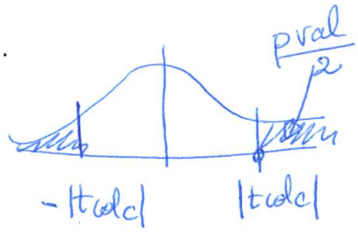
Quelles conditions doit on poser sur le modèle décrivant Y et sur la taille de l'échantillon pour pouvoir appliquer le test sur une variance vu en cours ? (1pt)

$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ (mais pas raisonnable vu question 2)
ou $n_Y \geq 30$ (vérifiée ici)

Préciser le choix des hypothèses du test à mettre en oeuvre (1pt) :

$\mathcal{H}_0$: $\mu_Y = \mu_0$ .... $\mathcal{H}_1$: $\mu_Y \neq \mu_0$ ....

Indiquer les instructions R utilisées pour réaliser les calculs (2pts) :



tcalc = sqrt(32)*(mean(y) - 280)/sd(y)
pval = 2*(1-pt(|tcalc|)) = 2*pt(tcalc)

Donner les valeurs de la statistique et de la p-valeur du test (1pt):

$tcalc =$ ....-2,25.... $pval =$ ....3,17 %....

Conclusion (justifier la réponse) (1,5pts) ?

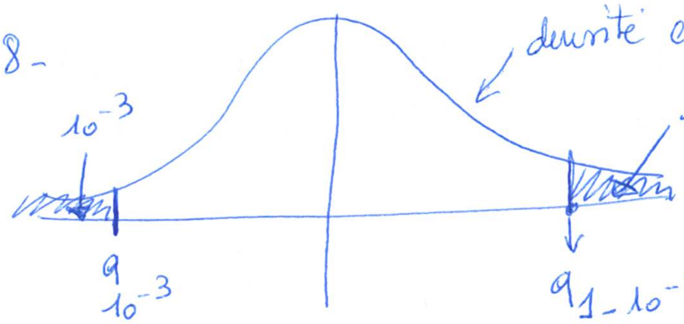
On peut conclure que μ_Y diffère de la valeur 280, avec un risque de se tromper α, dès que $\alpha > 3,17\%$.

8. (3pts) Le quantile d'ordre 10^{-3} de la loi normale centrée et réduite s'obtient avec :
   1. `-qnorm(1-0.001,0,1)` 2. `qnorm(1-0.001,0,1)` 3. `qnorm(0.001,1,1)-1`
   4. `1-qnorm(0.001,0,1)` 5. `qnorm(0.001,0,1)` 6. `1-qnorm(0 .001,1,1)`

9. (3pts) Si X suit la loi normale $\mathcal{N}(0,1)$ alors $P(X \leq 1.9)$ s'obtient avec :
   1. `0.5+pnorm(1.9,0,1)-pnorm(0,0,1)` 2. `pnorm(1.9,0,1)` 3. `1-qnorm(-1.9,0,1)`
   4. `0.5-pnorm(1.9,0,1)+pnorm(0,0,1)` 5. `1-pnorm(-1.9,0,1)` 6. `qnorm(1.9,0,1)`

8 -

densité $\mathcal{N}(0,1)$ symétrique.



$q_{1-10^{-3}} = -q_{10^{-3}}$ (par symétrie de la densité).

9 -

surface $\{\equiv\}$ = $P(X \leq 1{,}9)$ = pnorm(1,9)

= surface rouge + surface bleue