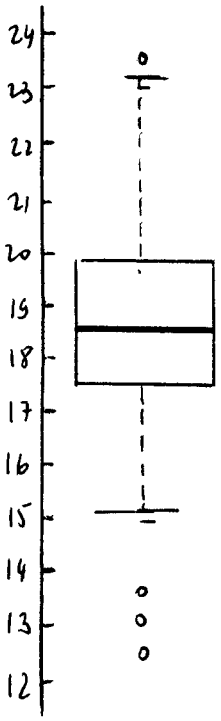
Université de Nice — L1MASS, année 2013-2014
Département de Mathématiques — Statistique
NOM : Date :
PRENOM : Salle et machine :



Contrôle Final
Mardi 6 Mai 13 :15-14 :45
Document autorisé : clé USB avec fichiers .R personnels

Former le vecteur y des tailles de `NW.hnd` du fichier `survey`

```
library(MASS)
data(survey)
Tailles<-survey[, c("Wr.Hnd", "NW.Hnd", "Height")]
Tailles.cc<-Tailles[complete.cases(Tailles),]
Tailles.cc
y<-Tailles.cc$NW.Hnd
```

1. Que signifie "`NW.Hnd`" ? Représenter avec soin, en marge, la boîte à moustaches obtenue avec la commande `boxplot`.

NW.Hnd signifie "non-writing hand"

2. Former le vecteur `y.sort` des valeurs de y triées par ordre croissant. Quelles sont les valeurs Q_1 et Q_3 des 1er et 3ème quartile de l'échantillon y.

```
y.sort= sort(y)
n = length(y)
Q1=y.sort(n/4)      donne Q1 = 17.5
Q3 = y.sort(n*3/4)  donne Q3 = 19.5
```

3. Quelle est la médiane Q_2 et la moyenne m de y ?

```
Q2 = y.sort(n/2)    donne Q2 = 18.5
m = mean(y)         donne m = 18.6726
```

4. Y a-t-il des valeurs exceptionnelles ; si oui, lesquelles ?

On observe 3 valeurs exceptionnelles inférieure et une supérieure
On affiche les valeurs extrêmes de la façon suivante
y.sort[1:8] permet de trouver 12.5, 13.0, 13.0, 13.5, 15.1, ...
y.sort[(n-7):n] permet de trouver ... 23.2, 23.3, 23.5
Les valeurs extrêmes sont 12.5, 13.0, 13.5 et 23.5

5. Former de même les vecteurs x et z des tailles de `Wr.hnd` et `Height` du fichier `survey`. Former un `data.frame` appelé `MesDonnées` regroupant x, y, et z ; quelles sont les trois premières lignes de `MesDonnées`

Commandes :

```
x <- Tailles.cc$Wr.Hnd
z <- Tailles.cc$Height
MesDonnées <- data.frame(x,y,z)
```

Les trois premières lignes de Mes.Données sont

	x	y	z
1	18.5	18.0	173.00
2	19.5	20.5	177.80

6. Quelles est la moyenne et l'écart-type de chacun ce vos échantillons x, y, et z.

mean(x); sd(x); mean(y); et... donnent

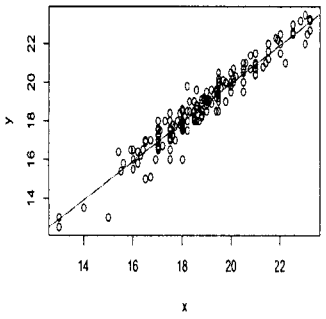
	x	y	z
mean	18.7476	18,6226	172.3827
sd	1.903517	1.980989	9.87125

7. Quelle est la matrice de variance-covariance de vos échantillons x, y, et z (arrondir à 2 chiffre après le point décimal). Précisez quelle commande de R vous permet d'obtenir ce résultat.

cov(MesDonnées) donne

3.62	3.63	11.29
3.63	3.92	11,42
11.29	11.42	97.44

8. Représenter le nuage des x et y et la droite de régression des y sur le x. Représentez le point M ayant pour coordonnées la moyenne des x et la moyenne des y. Donnez les commandes R vous permettant d'obtenir cette représentation.




```
Reg_y_sur_x = lm(y ~ x, data=MesDonnées)
plot(x,y)
abline(coef(Reg_y_sur_x))
points(mean(x), mean(y))
```

9. Quelle est l'équation $y = ax + b$ de la droite de régression des y sur les x. Indiquez comment vous avez trouvé cette équation.

`Coef(Reg_y_sur_x)` donne

(Intercept)	x
-0.1148053	1.0021232

(-0.1148053 : b ; 1.0021232 : a)

10. Donner une autre manière de trouver cette équation de la droite de régression des y sur les x.

Au cours 05 on a vu que

$$a = \frac{cov(x,y)}{Var(x)} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

on peut donc trouver a et b par

```
a = cov(x,y) / var(x)
b = mean(y) - a * mean(x)
```

11. On se demande si un dé est pipé ; on procède à un tirage de 900 coups de ce dés et on obtient les effectifs suivant pour les 6 valeurs possibles. On fait l'hypothèse nulle que le dé n'est pas pipé. Quels sont les effectifs théoriques des 6 classes ? Quelle statistique T du chi2 obtenez-vous pour votre échantillon ? Détaillez ici la valeur qui doit être calculée et donnez le résultat trouvé par R.

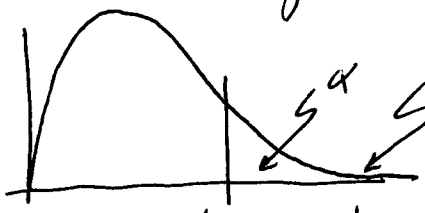
	1	2	3	4	5	6
effectif	143	132	145	165	144	171

```
O = c(143, 132, 145, 165, 144, 171)
n = sum(O)
T = sum((O - n/6)^2 / (n/6))
```

je trouve $\boxed{T = 7.3333}$

12. Rejetez-vous l'hypothèse H_0 ?(Précisez votre seuil)

Le nombre de degrés de liberté, df, est $6-1=5$



t_α = qchisq(1-alpha, df=5)

Pour $\alpha = 5\%$ je trouve $t_{5\%} = 11.0705$

$\alpha = 10\%$ —— $t_{10\%} = 9.236357$

Conclusion : même au seuil $\alpha = 10\%$ on ne rejette pas H_0 puisque $T = 7.3.. < t_{10\%}$
"Le dé n'est pas pipé"

13. On a reçu des plaintes de joueurs pour un autre dé. On a obtenu les effectifs suivant. On fait l'hypothèse nulle que le dé n'est pas pipé. Quels sont les effectifs théoriques des 6 classes ? Peut-on affirmer, au seuil $\alpha = 5\%$, que ce dé est pipé ?

	1	2	3	4	5	6
effectifs	466	502	496	504	469	563

```
O.pipé = c(466, 502, 496, 504, 469, 563)
n = sum(O.pipé)
T.pipé = sum((O.pipé - n/6)^2 / (n/6))
```

je trouve T.pipé = 12,244 > 11.0705 = $t_{5\%}$

Au seuil $\alpha = 5\%$ on peut donc rejeter l'hypothèse nulle

"Le dé est pipé"