Université de Nice — L1MASS, année 2014-2015
Département de Mathématiques — Statistique
NOM : Date :
PRENOM : Salle :

Feuille-réponse 5
Régression et simulation linéaire

1 Reprise du cours 05

Former le vecteur x des tailles `Wr.hnd` du fichier `survey`, comme dans le cours :

```
library(MASS)
data(survey)
survey.cc<-survey[complete.cases(survey), ]
x<-survey.cc$Wr.Hnd
```

et former de même les vecteurs y et z des tailles `NW.Hnd` et `Height`.

1. Former un data.frame appelé `MesDonnées` regroupant x, y, et z ; quelles sont les trois premières lignes de `MesDonnées`

Les trois premières lignes sont

	x	y	z
1	18.5	18.0	173.00
2	19.5	20.5	177.80
3	20.0	20.5	165.00

2. Former le "scatter-plot" des échantillons x, y, et z. Quels sont les nuages qui sont le mieux regroupés autour d'une droite ?

pairs(~ x+y+z, data = MesDonnées) donne 6 nuages de points
les abscisses se lisent en colonnes, les ordonnées en lignes.
Les nuages les mieux regroupés autour d'une droite sont ceux des (x,y) et (y,x)

3. Faire dessiner (`plot`) le nuage des x et y ; que provoque le fait de choisir `pch=17` ? Recommencer avec y et z, dans cet ordre.

Pour pch=17 les points (x_i, y_i) sont représentés par des triangles ▲
plot(x, y, pch=17)
plot(y, z) donne un nuage peu aligné de points représentés par o.

4. Former le modèle de régression linéaire `DteReg_z_sur_y` des z sur les y. Quelle est l'équation $z = ay + b$ que vous obtenez ? Quelle commande avez-vous utilisé.

Regression :
DteReg_z_sur_y = lm(z~y, data = MesDonnées)
Coefficients a et b : on trouve a = 3.172034 et b = 113,058201 au moyen de
coef(DteReg_z_sur_y) en donnant le "y" en dernier et "(intercept)"

5. Ajoutez la droite de régression à votre nuage. Quelle commande utilisez-vous ?

J'utilise la commande

abline(coef(DteReg_z_sur_y))

6. Ajoutez le "centre de gravité" de votre nuage. Appartient-il à la droite de régression ? Quelle commande utilisez-vous ?

Le centre de gravité est bien situé sur la droite de regression.
On en obtient la représentation graphique au moyen de
points(mean(y), mean(z))

7. Former les valeur z2y prédites par votre modèle linéaire. Quelle commande utilisez-vous ?

z2y = fitted(DteReg_z_sur_y)

8. Ajoutez les (y,z2y) à votre dessin. Quelle commande utilisez-vous ?

points(y, z2y, color = "red") (je change de couleurs pour distinguer z2y de z)

J'observe que les points sont bien sur la droite de regression.

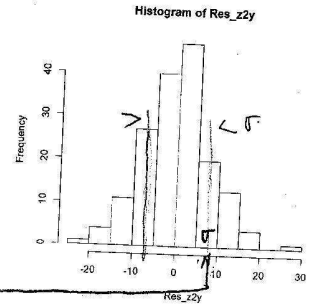
9. Représentez **dans une nouvelle fenêtre** l'histogramme des résidus z-z2y. Esquissez ici le résultat ; commentez. Donnez ici la moyenne et l'écart-type de ces résidus.

J'obtiens une nouvelle fenêtre, additionnelle, par la commande windows()

Je calcule les residus (Res_z2y) par

Res_z2y = z - z2y.

J'obtiens l'histogramme par hist(Res_z2y)



devrait être 0 !!! valeur plus petite que la précision "devant 1"

μ = moyenne = mean(Res_z2y) = 8.45..e-16

σ = ecart-type = sd(Res_z2y) = 7.737373 — signifie $8{,}45 \cdot 10^{-16}$

2 Simulation

1. Formez un échantillon `Residu` de même taille que z, normal, centré, et d'écart-type 8. Commencez par la commande `set.seed(sssmm)` où `sssmm` désigne votre machine. Quelle commande utilisez-vous ensuite ?

set.seed(214 20) # si salle = 214 et machine = 20

Residu = rnorm(length(z), 0, 8)

2. Quelle est la moyenne et l'écart-type de l'échantillon `Residu` ; tracez son histogramme ; commentez.

mean(Residu) donne −0,09036316 ← petit mais différent de 0

sd(Residu) donne 7,247686 ← voisin de 8 mais distinct

3. Posez `zSim=3*y+118+Residu` ; expliquez le nom donné à ce nouvel échantillon.

On avait $z_i = 3{,}17.. y_i + 113{,}05 + Res_z2y_i$ avec sd(Res_z2y) = 7,73 ≈ 8

On fabrique un z simulé ayant des caractéristiques semblables

4. Quelle est l'équation de la droite de régression des zSim sur les y ?

Regression DteReg_zSim_sur_y = lm(zSim ~ y)

Coefficients coef(DteReg_zSim_sur_y)

equation obtenue : zSim = 113,571644 + 2,911272 y

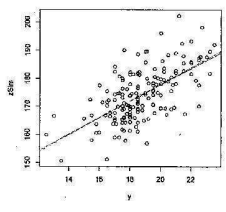
5. Quelle est la moyenne et l'écart-type des résidus de cette régression ?

Calcul des résidus : zSim2y = fitted(DteReg_zSim_sur_y) ; Res_zSim2y = zSim − zSim2y

mean(Res_zSim2y) = 1,861647e-17 ≈ 0

sd(Res_zSim2y) = 7,245563 proche

6. Représentez le nuage des (y,zSim) et la droite de régression obtenue. Ajoutez la droite d'équation $z = 3y + 118$ au moyen de la commande `abline(c(118,3))`. Commentez le résultat obtenu.



plot(y, zSim)

abline(coef(DteReg_zSim_sur_y), col = "blue") # Regression en bleue

abline(c(118, 3), col = "red") # modèle 3y+118 en rouge

La droite de regression et la droite modèle sont très proches et semble se couper au centre du graphe.

TRIAL MODE - a valid license will remove this message. See the keywords property of this PDF for more information.