Université Claude Bernard Lyon 1

M1 APA (Semestre 1) statistique/gestion
Session 1 24 novembre 2008

Corrigé de l'examen CCF2 de statistiques

IMPORTANT : Seuls sont corrigés les exercices 1 à 3, proposés par J. Bastien.

Correction de l'exercice 1.

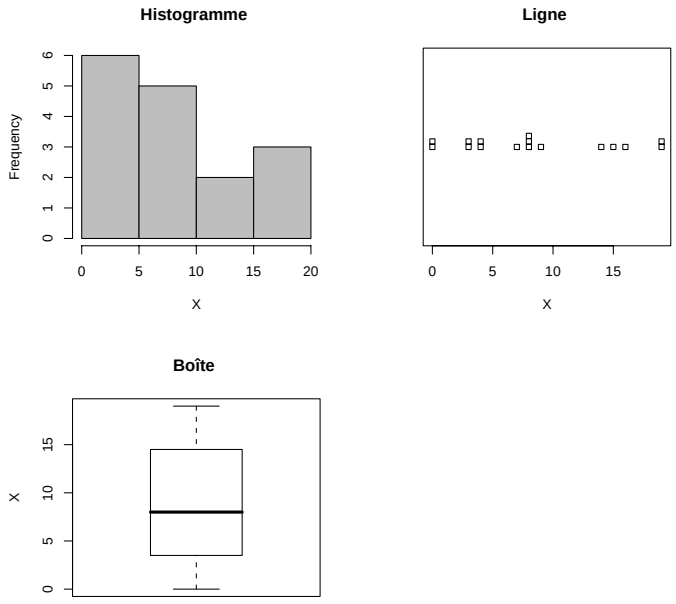
(1) On tape dans R

```
notes <- c(19, 8, 3, 0, 15, 4, 16, 9, 14, 8, 7, 3, 0, 8, 4, 19)
```

- On étudie une variable quantitative (ou numérique) . On procèdera donc comme dans la section 3.4 page 20 du chapitre 3 du document de cours.
- Les différents résultats déterminés par R sont donnés dans le tableau suivant

noms	valeurs
moyenne	8.5625
sd	6.313676
Q_1 (quartile à 25 %)	3.75
médiane	8
Q_3 (quartile à 75 %)	14.25
minimum	0
maximum	19
nombre	16

-

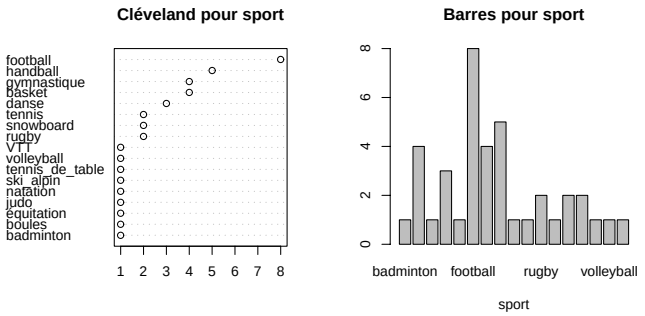


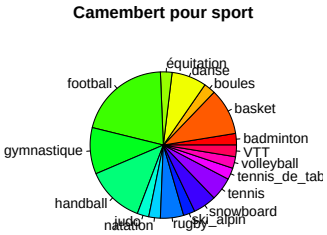
Voir les trois graphiques ci-dessus pour la variable 'notes'. Le nombre de données étant faible (16), l'histogramme et la boîtes à moustaches ne sont pas très pertinents ici.

(2)
- On étudie la variable qualitative (ou catégorielle) 'sport'. On procèdera donc comme dans la section 3.3 page 18 du chapitre 3 du document de cours.
- Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
badminton	1	2.564
basket	4	10.256
boules	1	2.564
danse	3	7.692
équitation	1	2.564
football	8	20.513
gymnastique	4	10.256
handball	5	12.821
judo	1	2.564
natation	1	2.564
rugby	2	5.128
ski_alpin	1	2.564
snowboard	2	5.128
tennis	2	5.128
tennis_de_table	1	2.564
volleyball	1	2.564
VTT	1	2.564

•

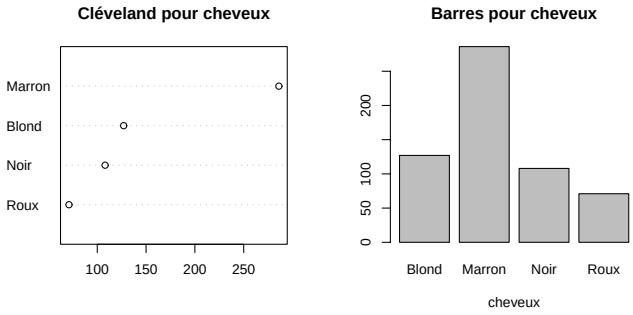


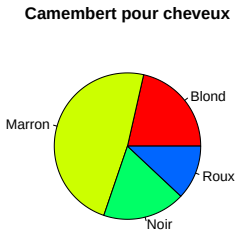


Voir les trois graphiques ci-dessus pour la variable 'sport'. Ici, le nombre de modalités étant élevés (17), le diagramme en barre et le camenbert ne sont pas très lisibles.

Correction de l'exercice 2.

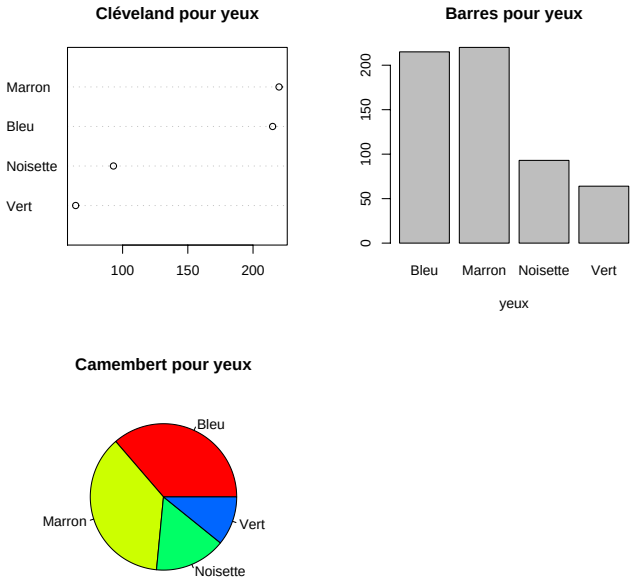
(1) (a) • – On étudie la variable qualitative (ou catégorielle) 'cheveux'.
–





Voir les trois graphiques ci-dessus pour la variable 'cheveux'.

- – On étudie la variable qualitative (ou catégorielle) 'yeux'.
  –



Voir les trois graphiques ci-dessus pour la variable 'yeux'.

(b) • Pour la variable 'cheveux' : Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
Blond	127	21.453
Marron	286	48.311
Noir	108	18.243
Roux	71	11.993

- Pour la variable 'yeux' : Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
Bleu	215	36.318
Marron	220	37.162
Noisette	93	15.709
Vert	64	10.811

(2)
- On étudie le croisement des deux variables qualitatives (ou catégorielles) 'cheveux' et 'yeux'. On procèdera donc comme dans la section 5.6 page 46 du chapitre 5 du document de cours.
- La table de contingence déterminée par R est donnée dans le tableau suivant

	Bleu	Marron	Noisette	Vert
Blond	94	7	10	16
Marron	84	119	54	29
Noir	20	68	15	5
Roux	17	26	14	14

Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
χ^2	138.289842
coefficient de Cramer V	0.279045
taille d'effet w	0.483319
probabilité critique p_c	2.32529e-25

On compare la taille d'effet w=0.483319 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=2.32529e-25 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

significativité pratique	**forte**
significativité statistique	**oui**

- On peut donc affirmer qu'il existe une relation entre les variables 'cheveux' et 'yeux'.

(3) (a)
- On étudie le croisement des deux variables qualitatives (ou catégorielles) 'teinte.cheveux' et 'teinte.yeux'.
- La table de contingence déterminée par R est donnée dans le tableau suivant

	clair	foncé
clair	165	33
foncé	207	187

Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
χ^2	53.516205
coefficient de Cramer V	0.300664
taille d'effet w	0.300664
probabilité critique p_c	2.56468e-13

On compare la taille d'effet w=0.300664 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=2.56468e-13 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

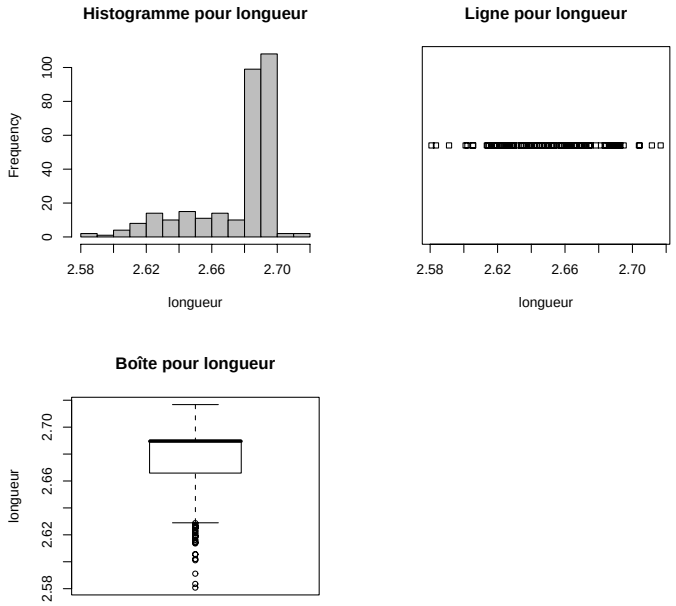
significativité pratique	**forte**
significativité statistique	**oui**

- On peut donc affirmer qu'il existe une relation entre les variables 'teinte.cheveux' et 'teinte.yeux'.

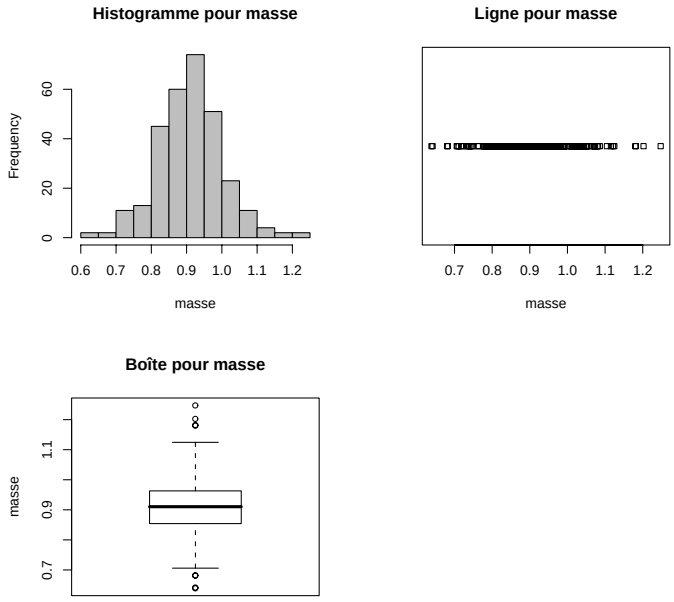
(b) Le χ^2 de la question 3 (égal à 53.5162) est plus faible que celui de la question 2 (égal à 138.2898). Il en est de même pour les taille d'effet (égales respectivement à 0.3007 et 0.4833). Les deux probabilité sont dans les deux cas très petites (égales respectivement à 2.565e-13 et 2.325e-25). Ainsi, il semblerait que la liaison soit "moins forte" pour la question 3 que pour la question 2. Cela provient probablement du fait que la simpification de la classification par teinte (claire ou foncée) soit trop brutale. Il aura fallu introduire des degrès.

Correction de l'exercice 3.

(1) (a) •



Voir les trois graphiques ci-dessus pour la variable 'longueur'.

- 

Voir les trois graphiques ci-dessus pour la variable 'masse'.

(b) • Pour la variable 'longueur' : Les différents résultats déterminés par R sont donnés dans le tableau suivant

noms	valeurs
moyenne	2.676045
sd	0.02526
Q_1 (quartile à 25 %)	2.66622
médiane	2.689515
Q_3 (quartile à 75 %)	2.690491
minimum	2.580788
maximum	2.716749
nombre	300

• Pour la variable 'masse' : Les différents résultats déterminés par R sont donnés dans le tableau suivant

noms	valeurs
moyenne	0.910772
sd	0.092023
Q_1 (quartile à 25 %)	0.854203
médiane	0.910367
Q_3 (quartile à 75 %)	0.962655
minimum	0.638803
maximum	1.247661
nombre	300

Le nombre de données étant importants (300), les lignes de points ne sont pas très lisibles ici.

(2)
- On étudie le croisement de la variable qualitative (ou catégorielle) 'type' et de la variable quantitative (ou numérique) 'longueur'. On procèdera donc comme dans la section 7.2 page 65 du chapitre 7 du document de cours.
- Voir la figure 1 page ci-contre.
- Avec R, on obtient les statistiques par groupes données dans le tableau suivant ;

	moyenne	écart-type (sd)	0%	25%	50%	75%	100%	n
A	2.6480	0.0271	2.5808	2.6284	2.6493	2.6655	2.7167	100
B	2.6901	0.0009	2.6876	2.6894	2.6901	2.6907	2.6923	200

On rappelle que :
- le quartile à 0 % correspond au minimum ;
- le quartile à 25 % correspond à Q_1 ;
- le quartile à 50 % correspond à la médiane ;
- le quartile à 75 % correspond à Q_3 ;
- le quartile à 100 % correspond au maximum.

Les graphiques et les statistiques par groupes montrent une certaine hétérogénéité entre les types. Plus précisément, les longueurs de type 'A' ont une moyenne plus faible que ceux du type 'B' ; l'écart-type du type 'A' est beaucoup plus important que celui du type 'B'.
Confirmons cela grâce à R.
Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.617007
probabilité critique p_c	0

On compare le rapport de corélation RC=0.617007 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=0 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

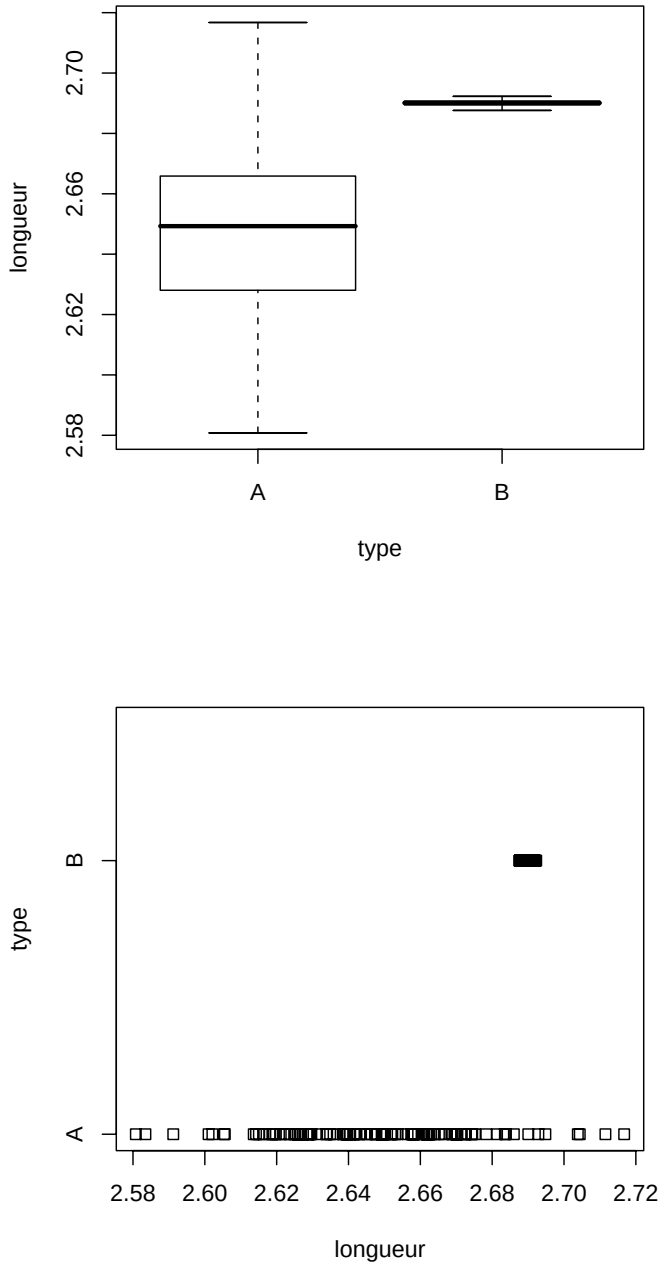


FIG. 1. Les collections de boîte de dispersion et de lignes de point

significativité pratique	**très forte**
significativité statistique	**oui**

- On peut donc affirmer qu'il existe une relation entre les variables 'longueur' et 'type'.

(3) (a) Si 'javelot' désigne le nom de la variable dans laquelle vous avez enregistré le fichier '`javelot.txt`', la première commande de la série suivante

```
indl<-(javelot$longueur>=2.6)&(javelot$longueur<=2.7)
javelot$longueur[indl]
```

renvoie un tableau de booléens, vrais si la longueur des javelot est comprise entre 2.6 et 2.7 et faux sinon. La seconde commande indique donc les longueurs des javelots réglementaires (par rapport à la longueur). Les autres commandes

```
indm<-javelot$masse>=0.8)
javelot$longueur[indm]
inddia<-(javelot$diametre>=0.025)&(javelot$diametre<=0.03)
javelot$longueur[inddia]
```

renvoie de même les longueur des javelots réglementaires (par rapport à la masse et au diamètre).

(b) Pour calculer la moyenne et l'écart-type des longueurs des javelots qui ont une longueur réglementaire, il suffit donc de taper

```
mean(javelot$longueur[indl])
sd(javelot$longueur[indl])
```

On obtient alors

$$m = 2.676524,$$
$$\sigma = 0.023509,$$

(c) On peut introduire un booléen qui permet d'obtenir tous les javelots réglementaires (par rapport à la masse, à la longueur et au diamètre) :

```
indtot <- indl & indm & inddia
```

puis

```
mean(javelot$longueur[indtot])
sd(javelot$longueur[indtot])
```

On obtient alors

$$m = 2.679029,$$
$$\sigma = 0.021688,$$

(d) Il suffit d'utiliser cette fois la numerotation tableau

```
dodo<-javelot[indtot,]
```

ou mieux

```
dodo<-data.frame(javelot[indtot,])
```

Par exemple, on aurait

```
head(dodo)
```

```
  longueur  masse diametre type
1    2.634 1.0225  0.02926    A
3    2.672 0.9671  0.02825    A
4    2.629 0.9091  0.02761    A
5    2.690 0.9543  0.02797    A
6    2.643 0.9037  0.02745    A
7    2.648 0.8596  0.02675    A
```

(4) (a)
- On étudie le croisement des deux variables quantitatives (ou numériques) 'longueur' et 'masse'. On procèdera donc comme dans la section 6.5 page 58 du chapitre 6 du document de cours.
- Voir la figure 2. Sur cette figure, les points semblent bien peu alignés.

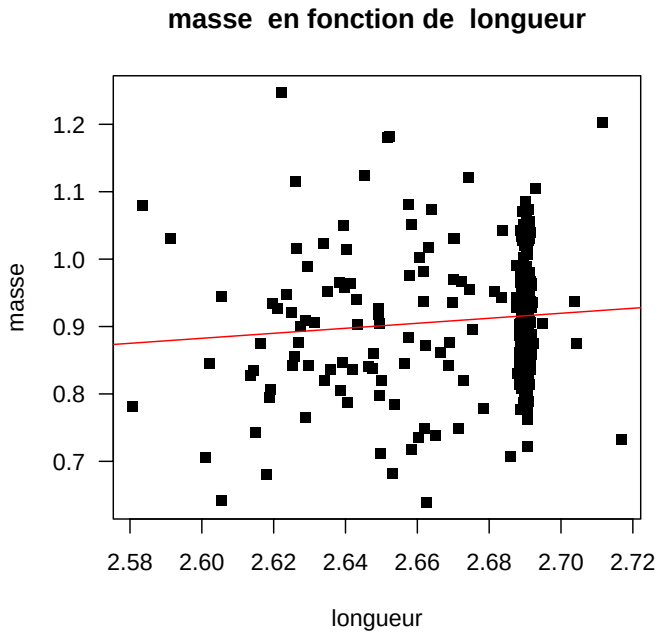


FIG. 2. Le nuage de point et la droite de régression

- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
pente a	0.371278
ordonnée à l'origine b	-0.082784
corrélation linéaire r	0.101914
probabilité critique p_c	0.0779982

On compare la taille d'effet corrélation linéaire r =0.101914 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0.0779982 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison linéaire :

significativité pratique	**moyenne**
significativité statistique	**non**

- On peut donc affirmer qu'il existe une faible relation entre les variables 'longueur' et 'masse'.

(b)
- On étudie le croisement des deux variables quantitatives (ou numériques) 'auxi' et 'masse'.
- Voir la figure 3. Sur cette figure, les points semblent très bien alignés.

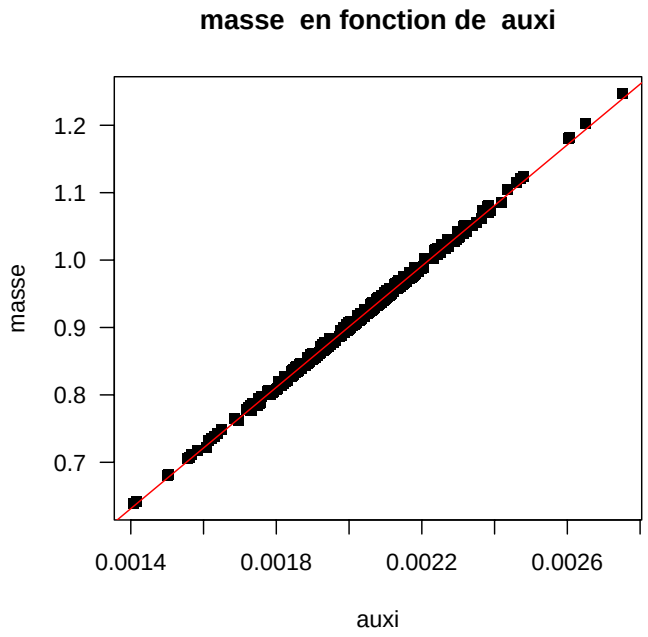


FIG. 3. Le nuage de point et la droite de régression

- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
pente a	449.774013
ordonnée à l'origine b	0.001594
corrélation linéaire r	0.998916
probabilité critique p_c	0

On compare la taille d'effet corrélation linéaire r =0.998916 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison linéaire :

significativité pratique	**très forte**
significativité statistique	**oui**

- On peut donc affirmer il existe une relation entre les variables 'auxi' et 'masse'.

(c) On admet que le nombre défini par

$$\text{rapport} = \frac{\text{masse}}{\text{longueur} \times \text{diamètre}^2} \tag{1}$$

est proportionnel à la masse volumique du matériau qui constitue le javelot. Or le résltat de la question 4b peut s'écrire sous la forme

$$\text{auxi} \approx a \times \text{masse} + b = 449.774013 \times \text{masse} + 0.001594 \approx 449.774013 \times \text{masse}$$

ce qui s'écrit donc

$$\text{longueur} \times \text{diamètre}^2 \approx 449.774013 \times \text{masse}$$

soit encore

$$\frac{\text{masse}}{\text{longueur} \times \text{diamètre}^2} = \frac{1}{449.774013} = 0.002223. \tag{2}$$

Au du de l'équation (1), cela ne fait que traduire que la masse volumique de l'ensemble des javelot est à peu près constante !

Remarque 1. Reprenons la remarque 4.5 page 34 du chapitre 4 du document de cours et tapons :

```
neojavelot <- split(javelot, javelot$type)
class(neojavelot)
names(neojavelot)
neojavelot$A
neojavelot$B
```

Pour chacun des deux data frame crée, faisons une corrélation entre la variable 'auxi' et la variable 'longueur' :

$*$: data frame 'A' :

- On étudie le croisement des deux variables quantitatives (ou numériques) 'auxi' et 'masse'.
- Voir la figure 4 page suivante. Sur cette figure, les points semblent très bien alignés.
- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

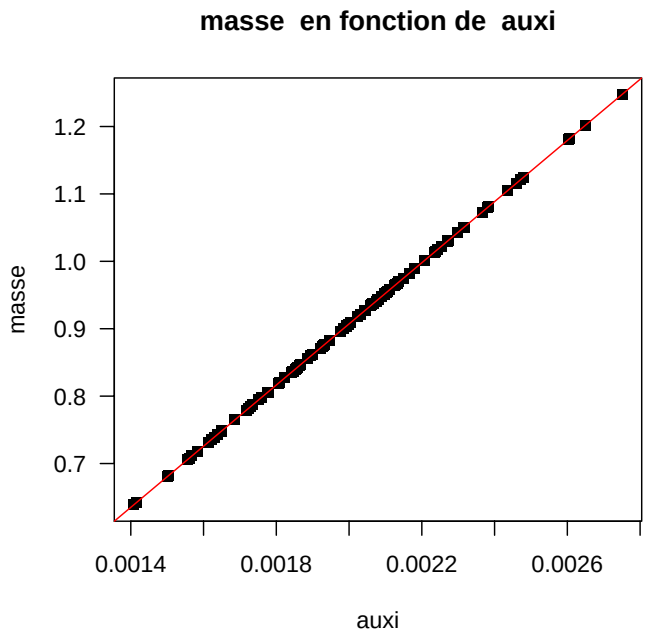


FIG. 4. Le nuage de point et la droite de régression

Noms des indicateurs	Valeurs
pente a	453.578668
ordonnée à l'origine b	0
corrélation linéaire r	1
probabilité critique p_c	0

On compare la taille d'effet corrélation linéaire r =1 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison linéaire :

significativité pratique	**très forte**
significativité statistique	**oui**

- On peut donc affirmer il existe une relation entre les variables 'auxi' et 'masse'.

$*$: data frame 'B' :

- On étudie le croisement des deux variables quantitatives (ou numériques) 'auxi' et 'masse'.
- Voir la figure 5 page ci-contre. Sur cette figure, les points semblent très bien alignés.
- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

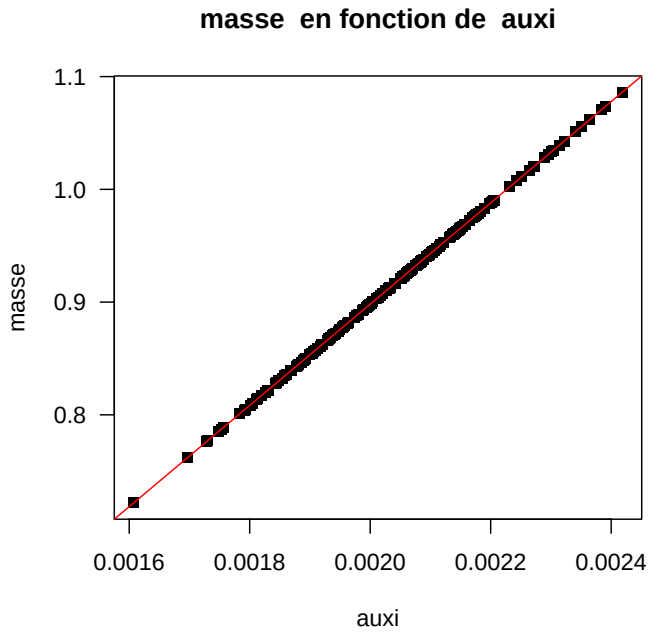


FIG. 5. Le nuage de point et la droite de régression

Noms des indicateurs	Valeurs
pente a	449.08779
ordonnée à l'origine b	0
corrélation linéaire r	1
probabilité critique p_c	0

On compare la taille d'effet corrélation linéaire r =1 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison linéaire :

significativité pratique	**très forte**
significativité statistique	**oui**

• On peut donc affirmer il existe une relation entre les variables 'auxi' et 'masse'.
En fait, les deux ordonnées à l'origine (8.058658e-16 et -1.189897e-15 sont ici, rigoureusement nulles (au zéro machine près) ; il en est de même des deux probabilités critiques (0 et 0) ; enfin, les deux corrélations linéaires (1 et 1) sont rigoureusement égales à 1. Un peu de reflexion nous montre en reprenant l'équation (2) qu'en fait la masse volumique de tout les javelot de type 'A' (respectivement de type 'B') est rigoureusement constante ! Elles sont respectivement proportionnelles à $1/453.578668 = 0.002205$ et $1/449.08779 = 0.002227$.

quelques sites à consulter

- sur les dimensions réglementaires du Javelot :
  - `http://www.zanzisport.com/spip/article.php3?id_article=467`
  - `http://www.edufr.ch/cgafr/FR/formation/TravInterd/javelot.pdf`
- Sur le matériaux constitutif des javelots (duralumin) : `http://fr.wikipedia.org/wiki/Alliages_d'aluminium_pour_corroyage`

Références

[Coh92] J Cohen. A power primer. *Psychological bulletin*, 112(1) :155–159, 1992.