Université Claude Bernard Lyon 1

M2 APA (Semestre 1) — statistique/gestion
Session 1 — 12 janvier 2009

Corrigé de l'examen CT de statistiques

À lire ! !

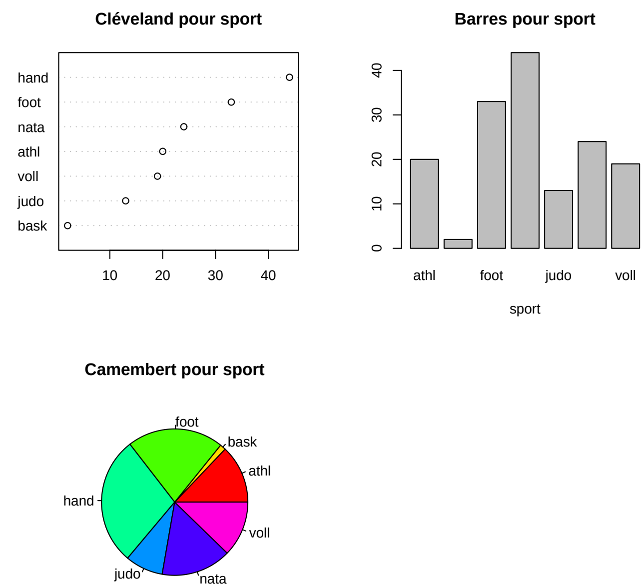
Un mouvement de rétentions de notes s'organise à Lyon I, pour exprimer un certain nombre de mécontentements. Ce mouvement ne se limite ni à la rétention de notes, ni à Lyon et concerne un très grand nombre d'universités et d'UFR en France. Pour en comprendre les fondements (et notamment pourquoi ce mouvement concerne aussi les étudiants), voir : `http://douaalter.lautre.net/mobilisation/index.html`.

Correction de l'exercice 1.

(1)
- On étudie la variable qualitative (ou catégorielle) 'sport'. On procèdera donc comme dans les sections 2.4 page 4 et 2.5 page 5 du chapitre 2 du document de cours.
- Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
athl	20	12.903
bask	2	1.290
foot	33	21.290
hand	44	28.387
judo	13	8.387
nata	24	15.484
voll	19	12.258

-



Voir les trois graphiques ci-dessus pour la variable 'sport'. Ici, le diagramme en barre n'est pas très lisible.

(2) • ∗ On étudie le croisement de la variable qualitative (ou catégorielle) 'sport' et de la variable quantitative (ou numérique) 'tde'. On procèdera donc comme dans les sections 4.2.5 page 25 et 4.2.7 page 29 du chapitre 4 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
athl	174.105	6.275	159.000	170.000	173.750	179.000	183.000	20
bask	183.300	10.889	175.600	179.450	183.300	187.150	191.000	2
foot	175.076	5.780	164.000	170.000	174.000	180.000	185.000	33
hand	177.416	5.324	162.000	173.000	178.000	181.500	188.000	44
judo	173.854	6.414	162.100	170.000	173.000	176.000	186.000	13
nata	177.996	5.894	161.600	176.000	178.150	181.150	191.600	24
voll	178.642	6.537	170.000	173.500	176.700	183.000	192.000	19

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 par groupes montrent une certaine hétérogénéité entre les groupes. Confirmons cela grâce à R.

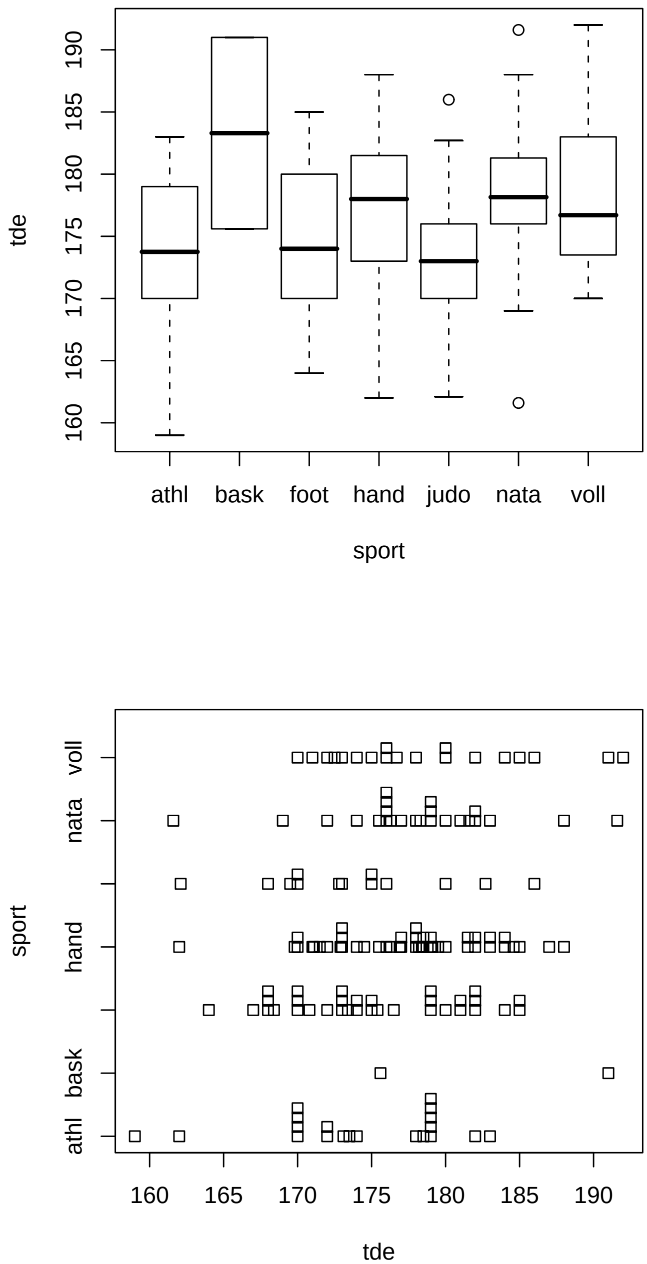


FIG. 1. Les collections de boîte de dispersion et de lignes de point

Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.094272
probabilité critique p_c	0.021457

On compare le rapport de corélation RC=0.094272 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=0.021457 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**

- Il y a donc une liasion entre les deux variables étudiées.
- D'après le graphique précédent et les statistiques par groupes, il semblerait que la moyenne des basketeurs soit plus importante que celle des autre sport. C'est normal, le basket attire les grands, et non l'inverse.

(3) (a)
- On étudie le croisement de la variable qualitative (ou catégorielle) 'pratique.bask' et de la variable quantitative (ou numérique) 'tde'.
- Voir la figure 2 page suivante.
- Avec R, on obtient les statistiques par groupes données dans le tableau suivant ;

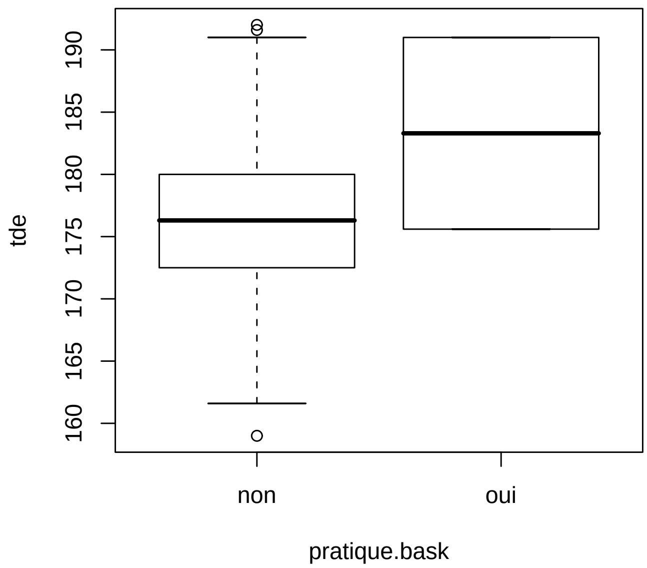
	moyenne	écart-type (sd)	0%	25%	50%	75%	100%	n
non	176.419	6.041	159.000	172.500	176.300	180.000	192.000	153
oui	183.300	10.889	175.600	179.450	183.300	187.150	191.000	2

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 que les basketeurs semblent plus grands que les autres sportifs.
Confirmons cela grâce à R.
Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.01623
probabilité critique p_c	0.114179

On compare le rapport de corélation RC=0.01623 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=0.114179 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :



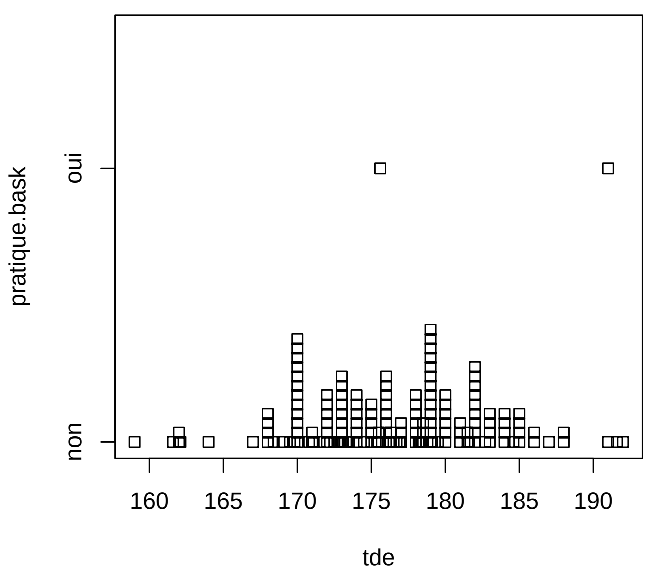


FIG. 2. Les collections de boîte de dispersion et de lignes de point

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

- En fait, la liaison est moyenne et statistiquement non significative, ce qui semble infirmer nos observations précédentes.

(b) Dans la question 2, nous trouvions un RC égal à 0.01623 et ici, il vaut 0.09427 ; de même, dans la question 2, nous trouvions une probabilité critique égale à 0.11418 et ici, elle vaut 0.02146.

Bref, paradoxalement, les deux variables sont moins liées que dans la question 2.

(c) En fait, il y a 2 sportifs qui pratiquent le basket ; ils sont donc très peu nombreux et il n'est pas pertinent de les comparer avec les autres, ce qui explique la conclusion de la question 3b.

Correction de l'exercice 2.

(1) Pour obtenir le nombre de pratiquants par sport, on procéde come dans la section 2.4 page 4 du chapitre 2 du document de cours pour dénombrer les effectifs de la variable 'sport'. On obtient les résultats ci-dessous :

	effectifs
athl	20
bask	2
foot	33
hand	44
judo	13
nata	24
voll	19

(2) (a)
- On étudie le croisement des deux variables quantitatives (ou numériques) 'tde' et 'poids'. On procèdera donc comme dans la section 4.4.5 page 39 du chapitre 4 du document de cours.
- Voir la figure 3 page ci-contre. Sur cette figure, les points semblent à peu près alignés.
- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
pente a	0.816063
ordonnée à l'origine b	-72.228531
corrélation linéaire r	0.625381
probabilité critique p_c	3.35886e-18

On compare la taille d'effet corrélation linéaire r =0.625381 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=3.35886e-18 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**

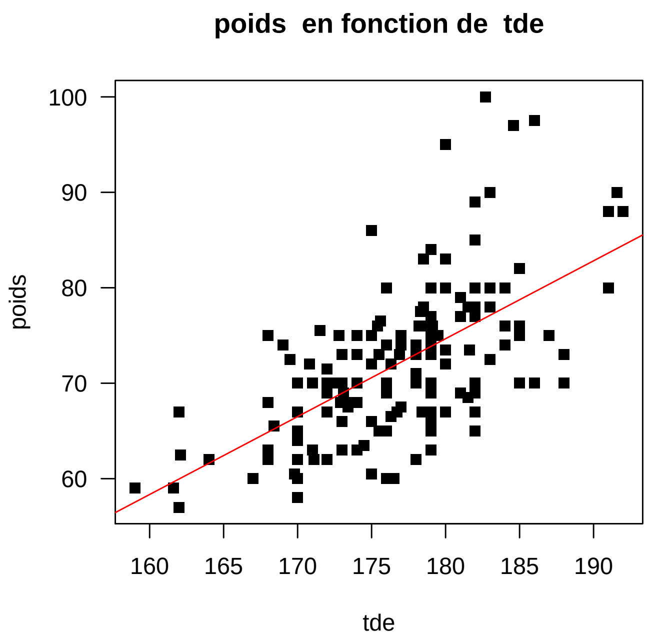


FIG. 3. Le nuage de point et la droite de régression

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

(b)

Pour obtenir le graphique de la figure 4 page suivante, on introduit d'abord deux nouvelles variables qui correspondent aux logarithme du poids et de la taille debout en procédant comme dans l'exercice 3.2 page 12 du chapitre 3. On utilisera la fonction log :

```
log(x)
```

On étudie le croisement des deux variables quantitatives (ou numériques) 'logarithme de taille debout' et 'logarithme de poids'. On procèdera donc comme dans la section 4.4.5 page 39 du chapitre 4 du document de cours.

(c)
- Voir la figure 4 page suivante. Sur cette figure, les points semblent à peu près alignés.
- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
pente a	1.964302
ordonnée à l'origine b	-5.892652
corrélation linéaire r	0.636185
probabilité critique p_c	5.84566e-19

On compare la taille d'effet corrélation linéaire r =0.636185 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=5.84566e-19 à la valeur seuil

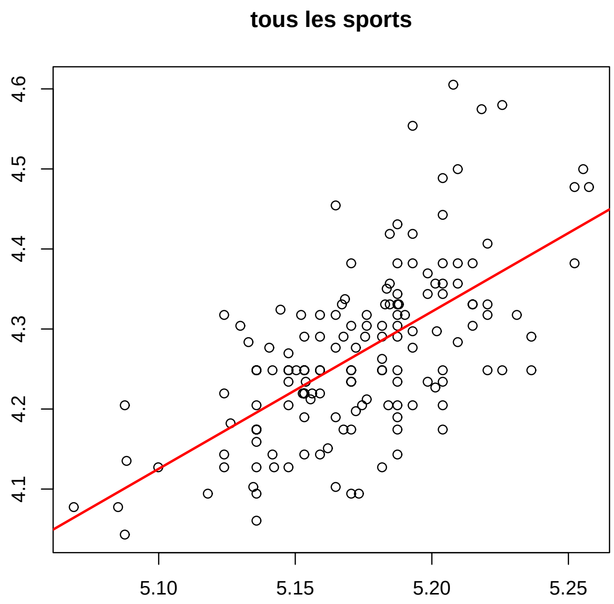


FIG. 4. Le nuage de point et la droite de régression

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 qu'il existe une relation entre les variables 'logarithme de taille debout' et 'logarithme de poids'

(3) Si on se restreint par exemple à l'athlétisme, les logarithmes de la taille et du poids peuvent être obtenus en tapant par exemple (si morphosportbis contient la variable dans laquelle vous avez enregistré le fichier '`morphosportbis.txt`'.

```
log.taille<-log(morphosportbis$tde[morphosportbis$sport=="athl"])
log.poids<-log(morphosportbis$poids[morphosportbis$sport=="athl"])
```

- Voir la figure 5 page suivante. Sur cette figure, les points semblent à peu près alignés.
- Confirmons cela grâce à R.
  Les résultats donnés par R sont les suivants :

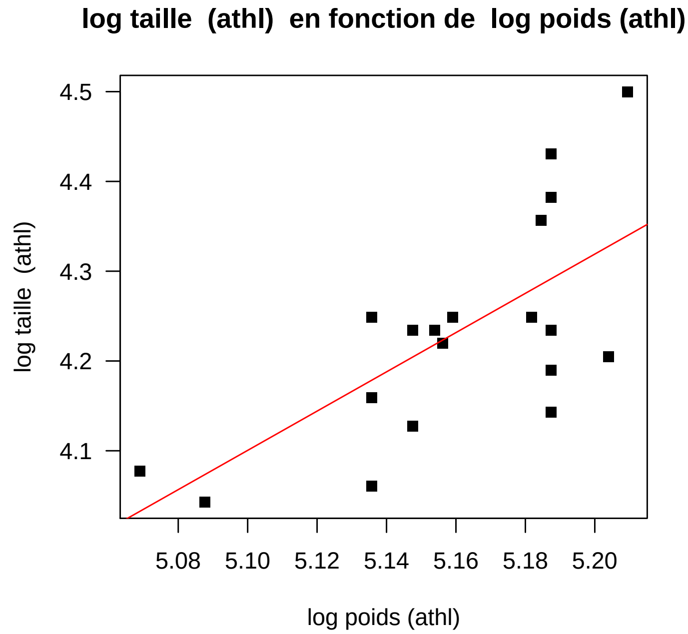


Fig. 5. Le nuage de point et la droite de régression

Noms des indicateurs	Valeurs
pente a	2.186065
ordonnée à l'origine b	-7.048488
corrélation linéaire r	0.675851
probabilité critique p_c	0.00107205

On compare la taille d'effet corrélation linéaire r =0.675851 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0.00107205 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 qu'il existe une relation entre les variables 'log poids (athl)' et 'log taille (athl)'.

Références

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