



Corrigé de l'examen CT de statistiques
--

**IMPORTANT : Seuls sont corrigés les exercices 1 et 2, proposés par J. Bastien.**

**À lire !!**

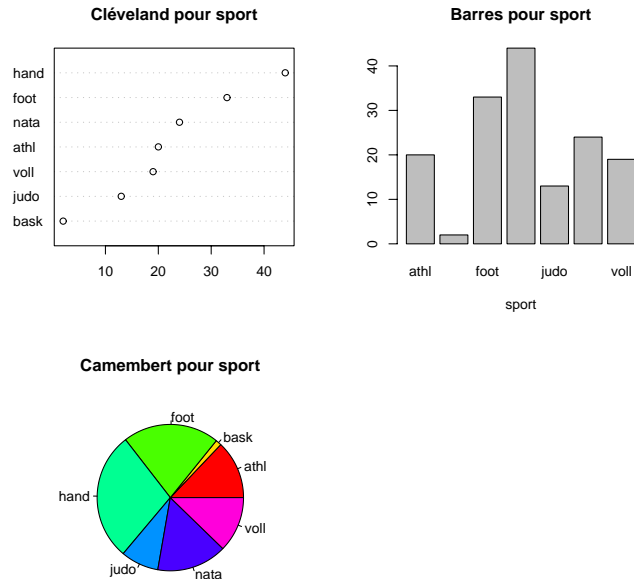
Un mouvement de rétentions de notes s'organise à Lyon I, pour exprimer un certain nombre de mécontentements. Ce mouvement ne se limite ni à la rétention de notes, ni à Lyon et concerne un très grand nombre d'universités et d'UFR en France. Pour en comprendre les fondements (et notamment pourquoi ce mouvement 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 la section 3.3 page 18 du chapitre 3 du document de cours.
  - Les effectifs et les pourcentages déterminés par  $\mathcal{R}$  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 la section 7.2 page 65 du chapitre 7 du document de cours.
- \* Voir la figure 1 page ci-contre.
- \* Avec  $\mathbb{R}$ , 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  $\mathbb{R}$ .

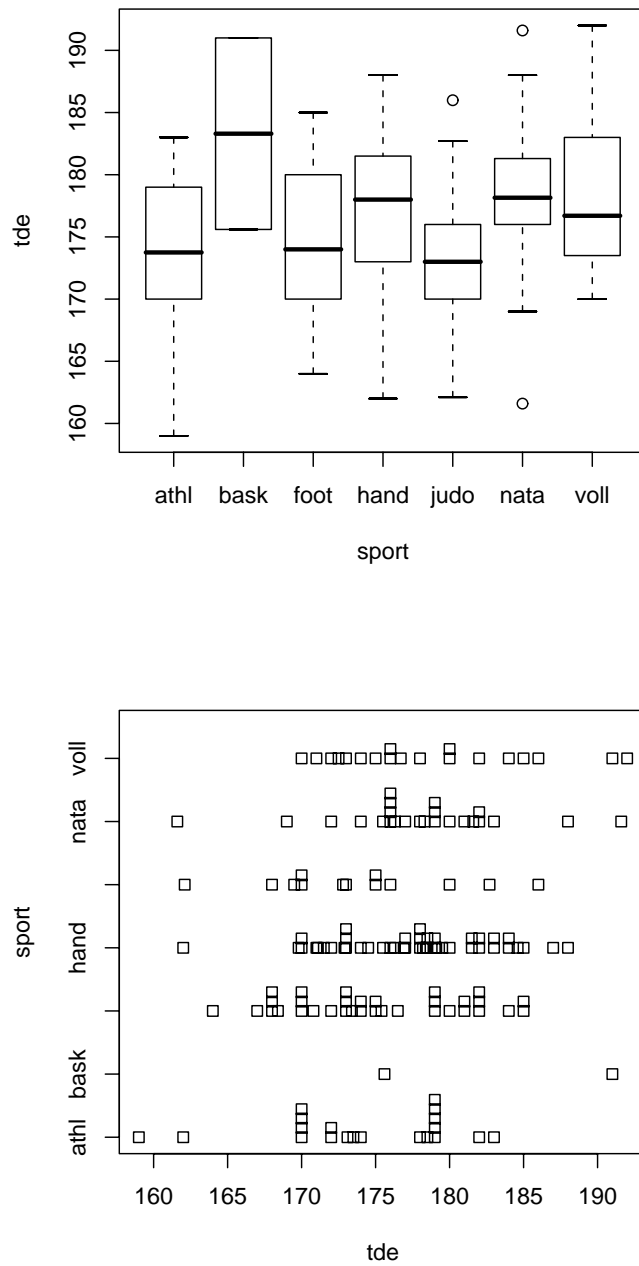


FIG. 1. Les collections de boîte de dispersion et de lignes de point

Les autres résultats donnés par  $\mathbb{R}$  sont les suivants :

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

On compare le rapport de corrélation  $RC=0.094272$  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	<b>forte</b>
significativité statistique	<b>oui</b>

- Il y a donc une liaison 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 autres sports. 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  $\mathbb{R}$ , 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  $\mathbb{R}$ .

Les autres résultats donnés par  $\mathbb{R}$  sont les suivants :

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

On compare le rapport de corrélation  $RC=0.01623$  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 :

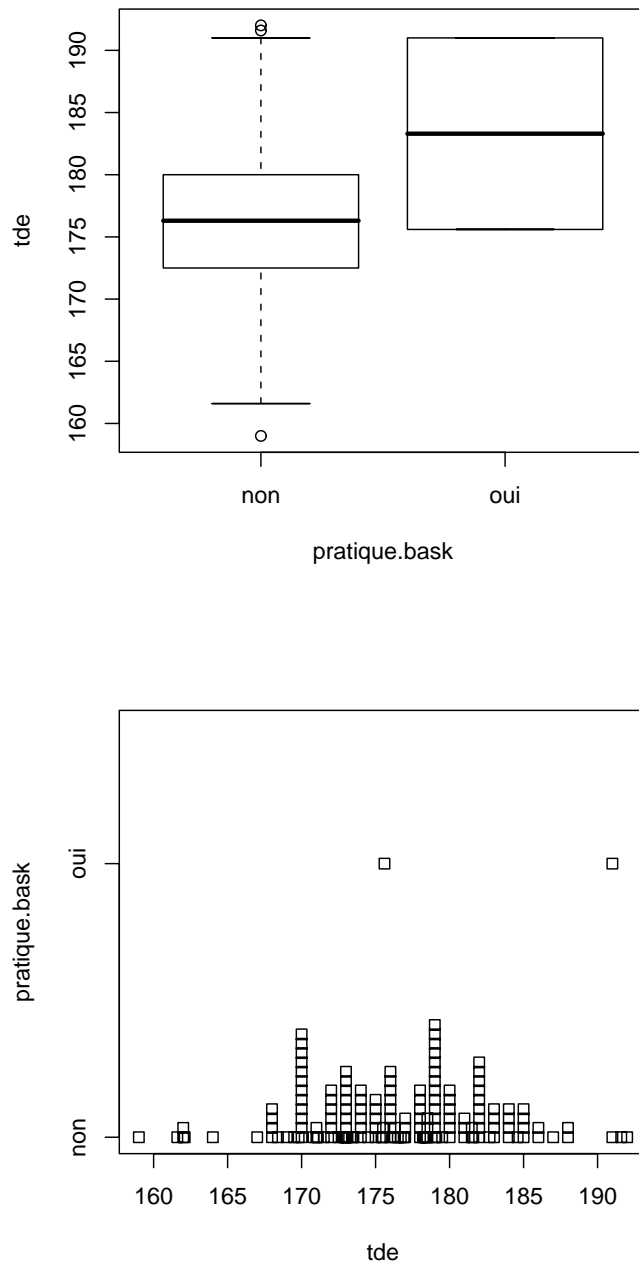


FIG. 2. Les collections de boîte de dispersion et de lignes de point

significativité pratique	<b>moyenne</b>
significativité statistique	<b>non</b>

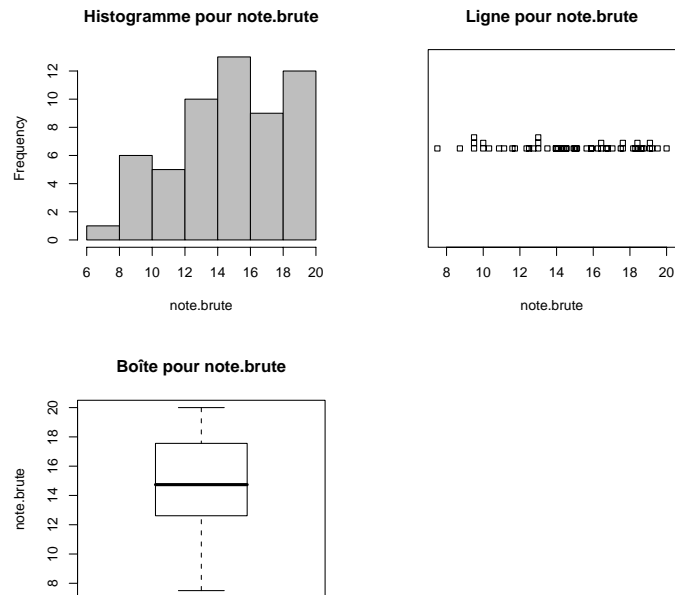
- En fait, la liaison est moyenne et statistiquement non significative, ce qui semble infirmer nos observations précédentes.
- (b) Dans la question 2, nous trouvons un  $RC$  égal à 0.01623 et ici, il vaut 0.09427 ; de même, dans la question 2, nous trouvons une probabilité critique égale à 0.11418 et 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)
- On étudie la variable quantitative (ou numérique) 'note.brute'. On procèdera donc comme dans la section 3.4 page 20 du chapitre 3 du document de cours.
  - Les différents résultats déterminés par  $\mathbb{R}$  sont donnés dans le tableau suivant

noms	valeurs
moyenne	14.729927
sd	3.206721
$Q_1$ (quartile à 25 %)	12.670455
médiane	14.738636
$Q_3$ (quartile à 75 %)	17.528409
minimum	7.5
maximum	20
nombre	56

•



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

- (2)
- On étudie le croisement de la variable qualitative (ou catégorielle) 'groupe' et de la variable quantitative (ou numérique) 'note.brute'.
  - Voir la figure 3.

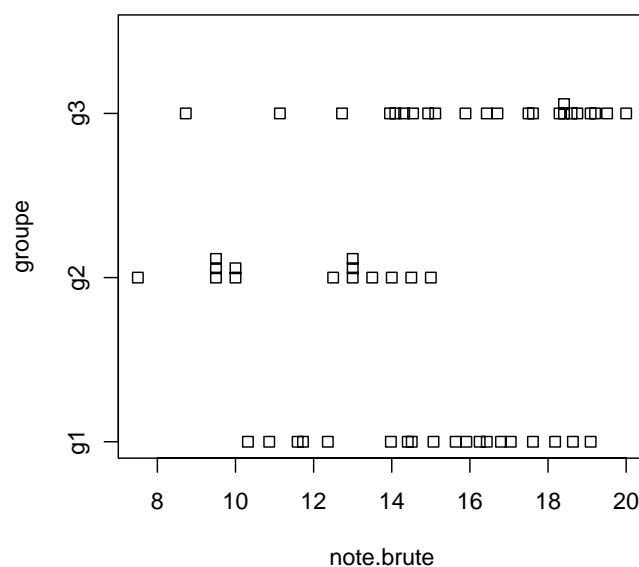
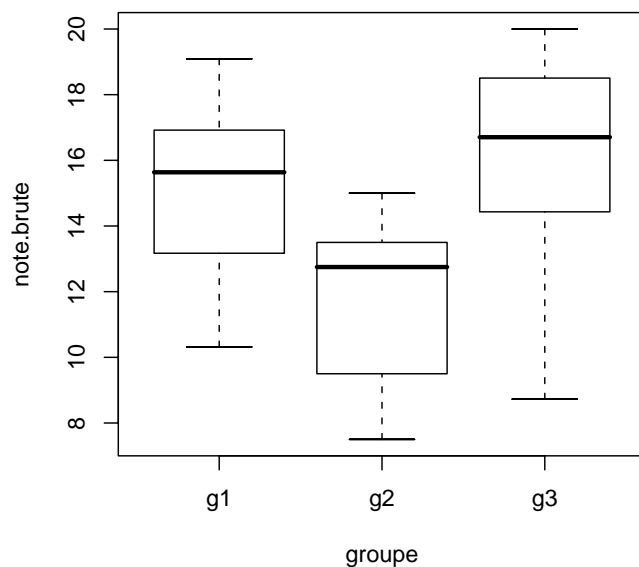


FIG. 3. Les collections de boîte de dispersion et de lignes de point

- Avec  $\mathbb{R}$ , on obtient les statistiques par groupes données dans le tableau suivant ;

	moyenne	écart-type (sd)	0%	25%	50%	75%	100%	n
g1	15.07	2.68	10.32	13.17	15.64	16.92	19.09	19
g2	11.75	2.33	7.50	9.62	12.75	13.38	15.00	14
g3	16.26	2.90	8.73	14.43	16.70	18.51	20.00	23

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 groupes.

Confirmons cela grâce à  $\mathbb{R}$ .

Les autres résultats donnés par  $\mathbb{R}$  sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.318882
probabilité critique $p_c$	3.80533e-05

On compare le rapport de corrélation  $RC=0.318882$  aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique  $p_c=3.80533e-05$  à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

significativité pratique	<b>très forte</b>
significativité statistique	<b>oui</b>

- Il y a donc une liaison entre les deux variables étudiées.
- (3) (a) La moyenne de l'ensemble des notes brutes est égale à 14.7299 et l'écart-type à 3.2067.
- (b) En utilisant par exemple les commandes : ('file138' désigne le nom de la variable dans laquelle vous avez enregistré le fichier 'file138.txt')

```
indg1<-file138$groupe=="g1"
file138$note[indg1]
mean(file138$note[indg1])
sd(file138$note[indg1])
```

on obtient par exemple la moyenne et l'écart-type des notes du groupe g1. On obtient les autres moyennes et écart-types de la même façon. Bref, on obtient les moyennes et écart-types suivants :



groupe	moyenne	écart-type (sd)
g1	14.72993	3.20672
g2	14.72993	3.20672
g3	14.72993	3.20672

On vérifie que ces moyennes et écart-types sont bien égaux à ceux obtenus dans la question 3a.

- (c) • On étudie le croisement de la variable qualitative (ou catégorielle) 'groupe' et de la variable quantitative (ou numérique) 'note'.
- Voir la figure 4 page suivante.
  - Avec  $\mathbb{R}$ , on obtient les statistiques par groupes données dans le tableau suivant ;

	moyenne	écart-type (sd)	0%	25%	50%	75%	100%	n
g1	14.72993	3.20672	9.03247	12.44894	15.40262	16.94071	19.54049	19
g2	14.72993	3.20672	8.89312	11.81152	16.10329	16.96165	19.19337	14
g3	14.72993	3.20672	6.40873	12.71128	15.22225	17.21245	18.86316	23

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 homogénéité entre les groupes (mêmes moyennes, mêmes écart-types).

Confirmons cela grâce à  $\mathbb{R}$ .

Les autres résultats donnés par  $\mathbb{R}$  sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0
probabilité critique $p_c$	1

On compare le rapport de corrélation  $RC=0$  aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique  $p_c=1$  à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

significativité pratique	<b>faible</b>
significativité statistique	<b>non</b>

- Il n'y a donc pas de liaison entre les deux variables étudiées. Attention, l'absence de significativité statistique nous montre que cette absence de liaison est peut-être due au hasard !



- (d) Un problème subsiste, qu'on parfois du mal a appréhender certains étudiants : puisqu'une harmonisation est faite, les groupes à faible moyenne (plus faible que la moyenne générale) voient leur note augmenter et réciproquement les groupes à forte moyenne (plus forte que la moyenne générale) voient leur note diminuer. Cette harmonisation mathématiquement juste est individuellement parfois mal vécue par les individus "extrêmes"; en effet, cette harmonisation, fondée sur des notions de statistiques descriptives, traite nécessairement les étudiants comme un groupe et ne les considère pas comme individus particuliers ! Une autre façon plus juste serait de faire corriger l'ensemble des travaux par un seul même professeur.

Ou alors, pour casser ce système, de la part des étudiants pas contents, d'être unis et de rendre tous un travail excellent (pour obtenir tous la même note 20) ou nul (au sens mathématique du terme : pour obtenir tous la même note 0). Dans ces deux cas, l'écart-type des notes est nul et tout calcul est impossible !

## Références

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