



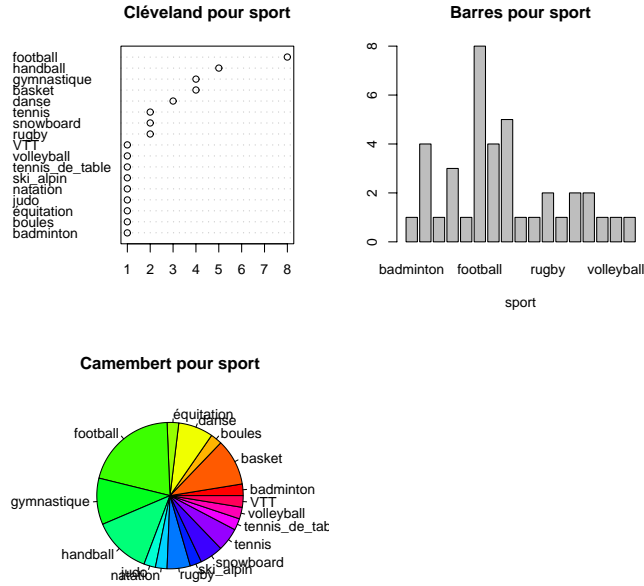
Corrigé de l'examen CCF2 de statistiques

Correction de l'exercice 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 \mathcal{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

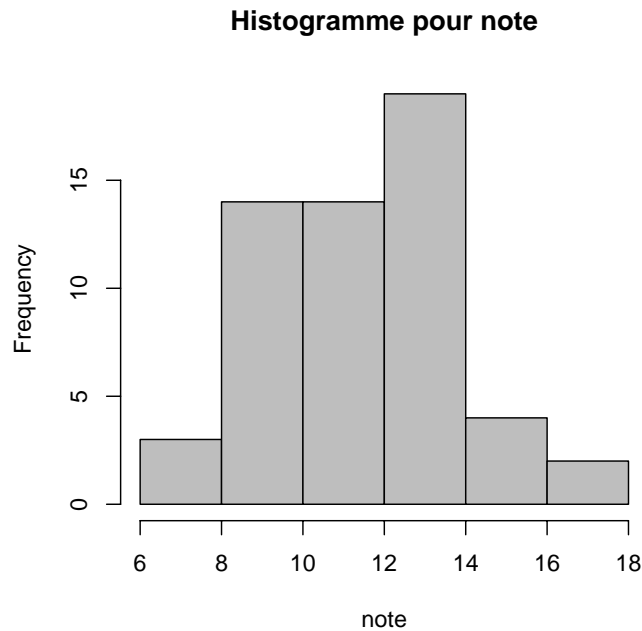
•



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

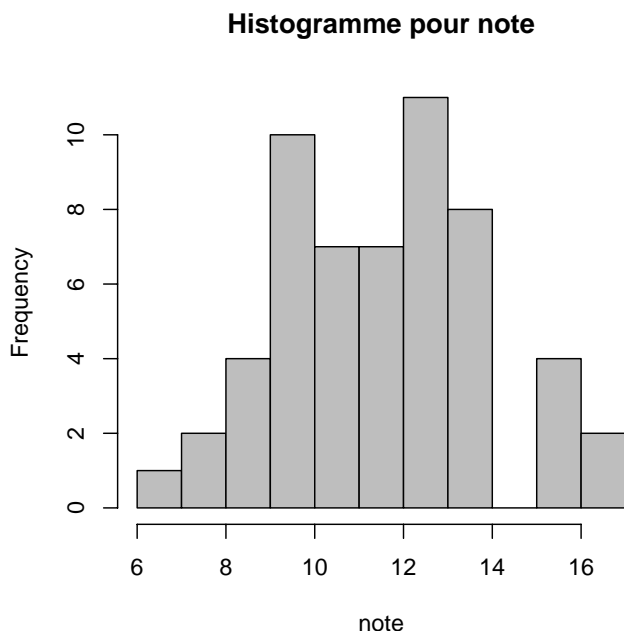
Correction de l'exercice 2.


- (1) On étudie la variable quantitative (ou numérique) 'note'. On procèdera donc comme dans les sections 3.2 page 11 et 3.3 page 12 du chapitre 3 du document de cours.




Voir l'histogramme ci-dessus pour la variable 'note'.

(2)

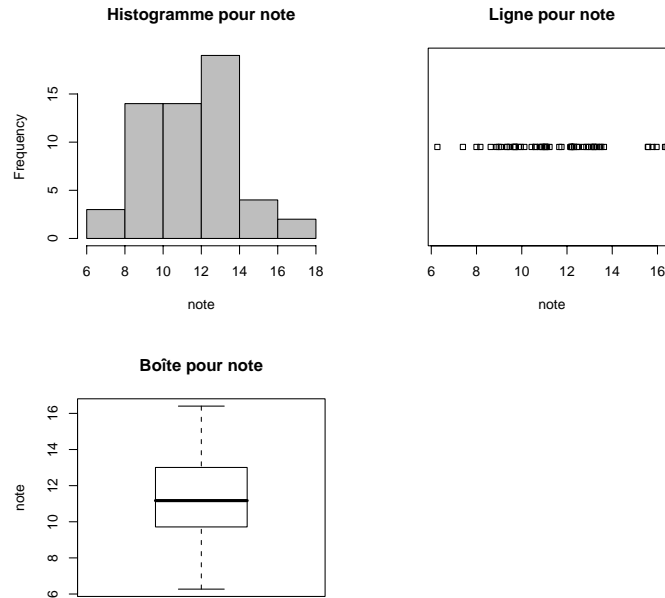


Voir l'histogramme ci-dessus pour la variable 'note' pour 12 classes. On constate, en prenant un nombre de classes plus élevé que  choisit par défaut, qu'il semble apparaître deux pics ; peut-être, observe-t-on une bi-normalité : il se détacherait deux groupes d'étudiants. L'un de moyenne 9 et l'autre de moyenne 13. Peut-être, y'aurait-il un groupe d'étudiants "motivés", opposé à un groupe d'étudiants démotivés.

- (3)
- On étudie la variable quantitative (ou numérique) 'note'. On procèdera donc comme dans la section 3.2 page 11 et 3.3 page 12 du chapitre 3 du document de cours.
 - Les différents résultats déterminés par  sont donnés dans le tableau suivant

noms	valeurs
moyenne	11.508065
sd	2.294819
Q_1 (quartile à 25 %)	9.71661
médiane	11.1719
Q_3 (quartile à 75 %)	12.978262
minimum	6.270278
maximum	16.402698
nombre	56

•



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

Correction de l'exercice 3.

- On étudie le croisement des deux variables quantitatives (ou numériques) 'poids' et 'taille'. On procèdera donc comme dans la section 4.4.5 page 39 du chapitre 4 du document de cours.
- Voir la figure 1 page suivante. Sur cette figure, les points semblent alignés.
- Confirmons cela grâce à \mathcal{R} .

Les résultats donnés par \mathcal{R} sont les suivants :

Noms des indicateurs	Valeurs
pende a	0.832284
ordonnée à l'origine b	118.708569
corrélacion linéaire r	0.891533
probabilité critique p_c	6.50023e-21

On compare la taille d'effet corrélation linéaire $r = 0.891533$ aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique $p_c = 6.50023e-21$ 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 'poids' et 'taille'.

Correction de l'exercice 4.

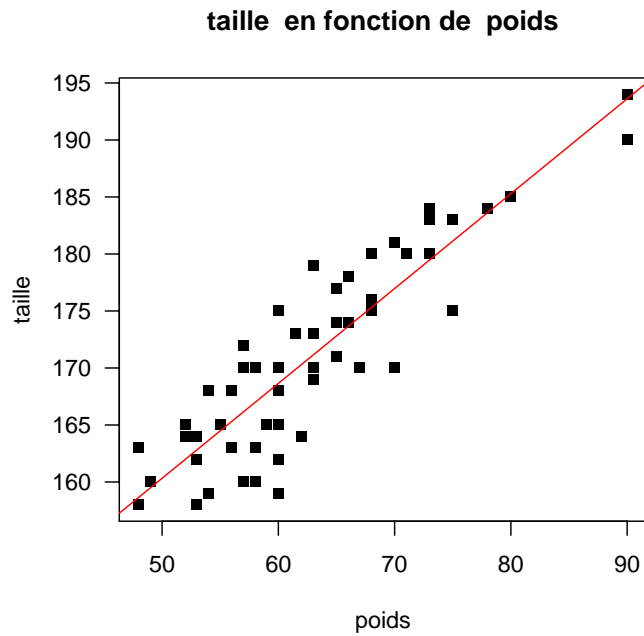
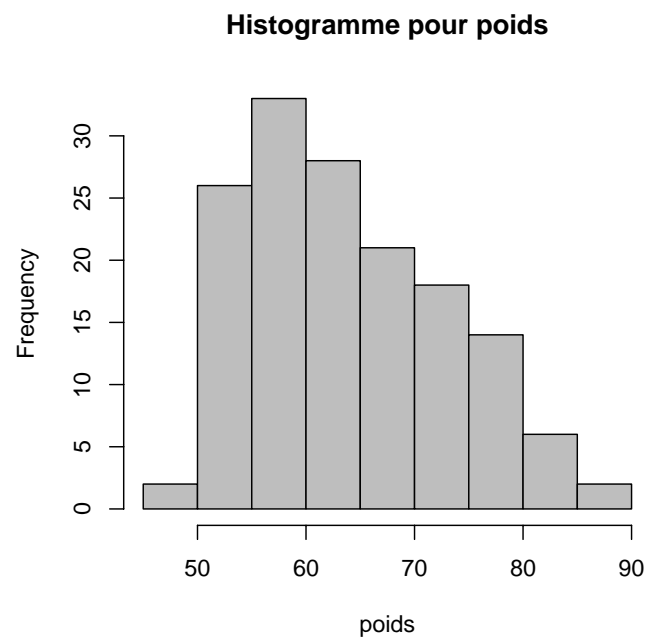


FIG. 1. Le nuage de point et la droite de régression

- (1) (a) On étudie la variable quantitative (ou numérique) 'poids'.
 (b)



Voir l'histogramme ci-dessus pour la variable 'poids'.

(2)

- (a) On étudie la variable qualitative (ou catégorielle) 'factor.poids'.
 (b)
 (c)
 (d) *Question 2b, 2c et 2d corrigées ensemble.*

On étudie le croisement des deux variables qualitatives (ou catégorielles) 'sexe' et 'factor.poids'. On procèdera donc comme dans la section 4.2 page 21 du chapitre 3 du document de cours.

La table de contingence déterminée par \mathcal{R} est donnée dans le tableau suivant

	[45,50)	[50,55)	[55,60)	[60,65)	[65,70)	[70,75)	[75,80)	[80,85)	[85,90)
féminin	0	19	27	21	1	1	0	0	0
masculin	1	1	4	10	22	19	12	10	2

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

Noms des indicateurs	Valeurs
χ^2	97.203759
coefficient de Cramer V	0.805
taille d'effet w	0.805
probabilité critique p_c	1.58983e-17

On compare la taille d'effet $w=0.805$ aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique $p_c=1.58983e-17$ à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

significativité pratique	très forte
significativité statistique	oui

Il semblerait que, chez les femmes, la classe la plus importante soit [55,60) ; au contraire, chez les hommes, les classes les plus importantes soient [65,70) et [70,75) ; le poids semble donc dépendant du sexe. Cela est confirmé par les calculs de \mathcal{R} .

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

- (e) La critique de cette méthode est que l'on a rendu la variable 'poids' factorielle (ou qualitative) en la remplaçant par des classes de poids. Le résultat observé pourrait par exemple dépendre *a priori* de la largeur de classe choisie. La méthode de la question 3 pallie ce problème, puisque l'on traite telle qu'elle la variable 'poids'.

- (3) (a) Ici, contrairement à la question 2, on conserve l'aspect quantitatif de la variable 'poids'.
- (b)
- (c)
- (d) *Question 3b, 3c et 3d corrigées ensemble.*
- On étudie le croisement de la variable qualitative (ou catégorielle) 'sexe' et de la variable quantitative (ou numérique) 'poids'. 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 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
féminin	57.522	4.168	50.000	54.000	57.000	61.000	70.000	69
masculin	70.160	7.925	47.000	65.000	70.000	76.000	88.000	81

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 : les femmes ont l'air plus légères que les hommes.

Confirmons cela grâce à \mathbb{R} .

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

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.489542
probabilité critique p_c	0

On compare le rapport de corrélation $RC=0.489542$ 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 :

significativité pratique	très forte
significativité statistique	oui

- (e) On peut donc affirmer qu'il existe une relation entre les variables 'sexe' et 'poids', ce qui confirme bien les résultats de la question 2.

Références

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

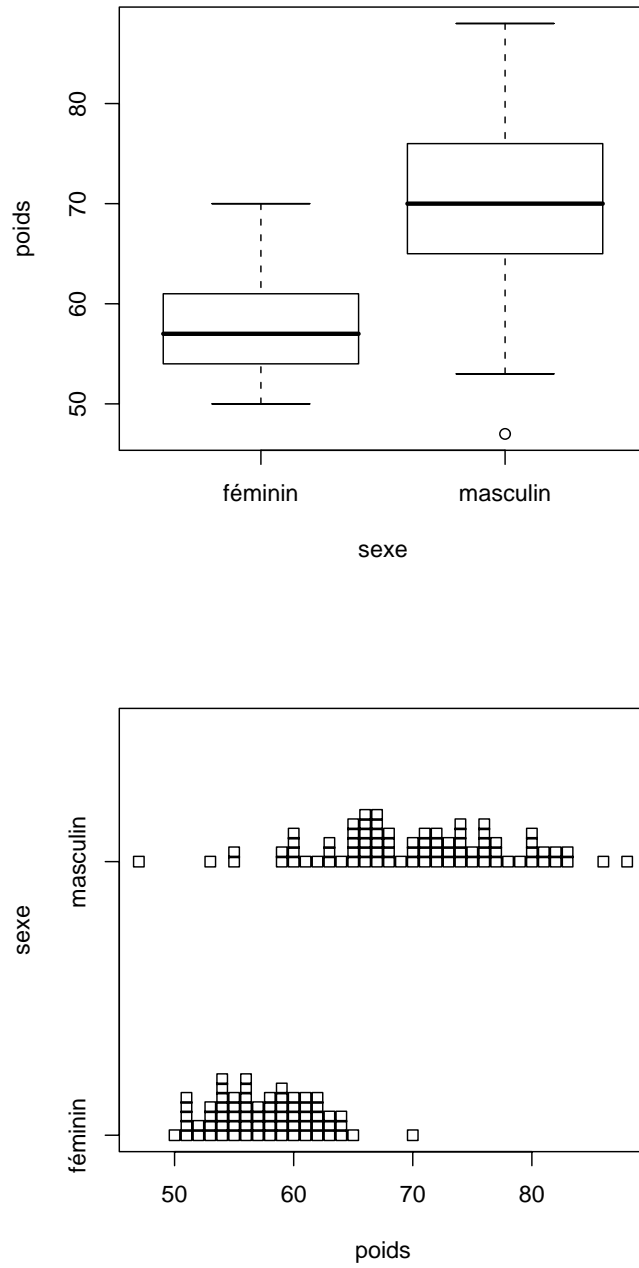


FIG. 2. Les collections de boîte de dispersion et de lignes de point