Université Claude Bernard Lyon 1

M1IGAPAS (Semestre 1) — statistique
Session 1 — 29 Novembre 2012

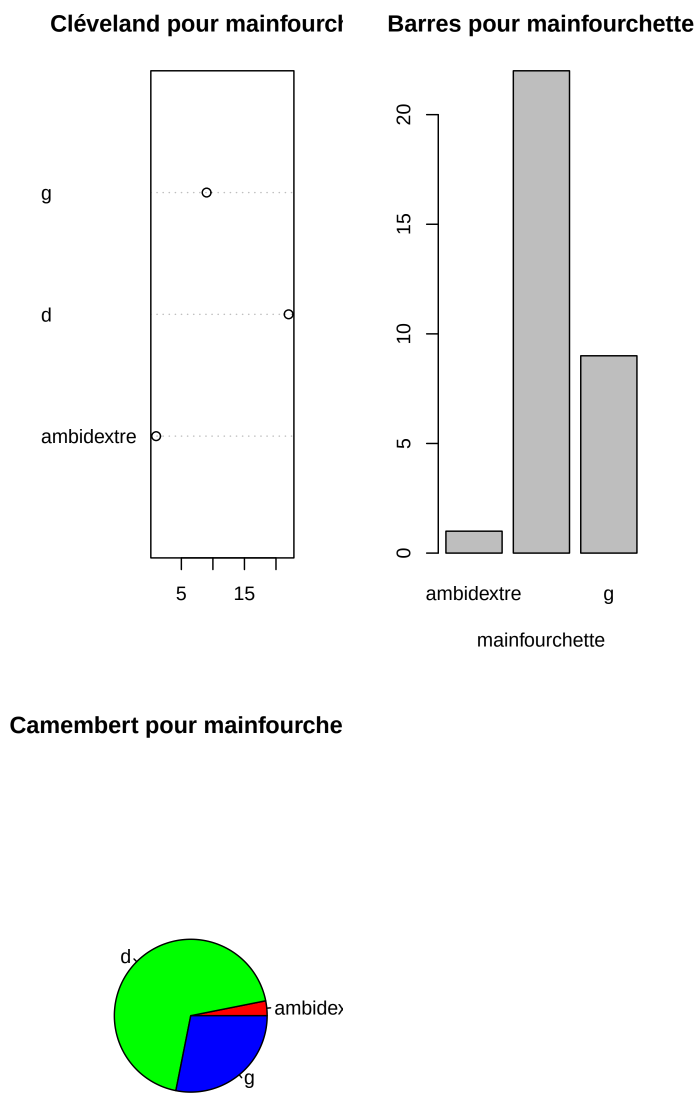
Corrigé de l'examen CCF2 de statistiques

Correction de l'exercice 1.

(1) (a)
- On étudie la variable qualitative (ou catégorielle) '`mainfourchette`'. Pour les manipulations avec R, on renvoie donc à la section 3.3 et aux sections récapitulatives 9.1.1 et 9.1.2 du document de cours.
- Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
ambidextre	1	3.125
g	9	28.125
d	22	68.750

-

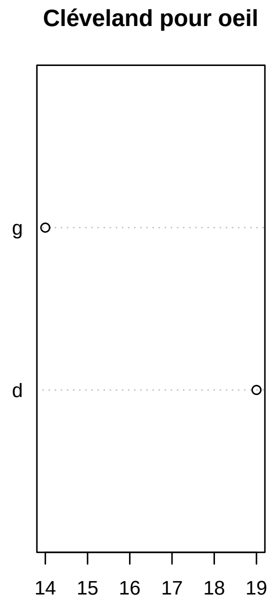


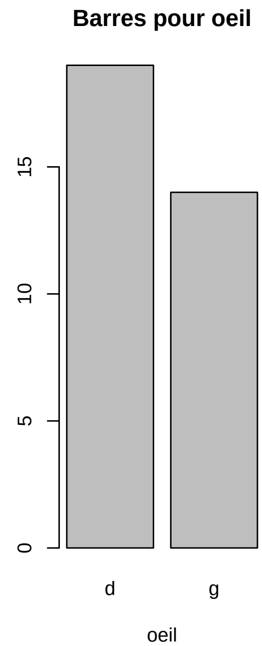
Voir les deux graphiques ci-dessus pour la variable '`mainfourchette`'.

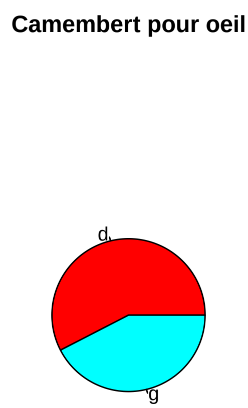
(b)
- On étudie la variable qualitative (ou catégorielle) '`oeil`'.
- Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

-

	effectifs	pourcentages
g	14	42.424
d	19	57.576





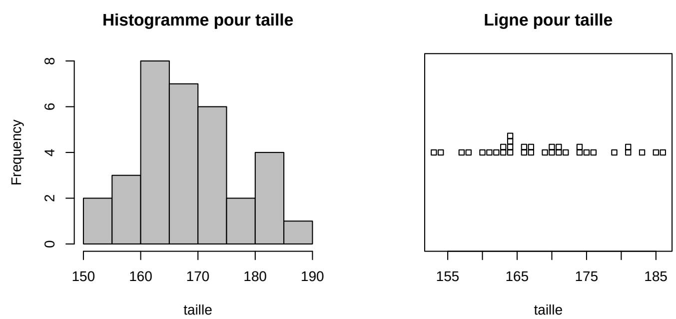


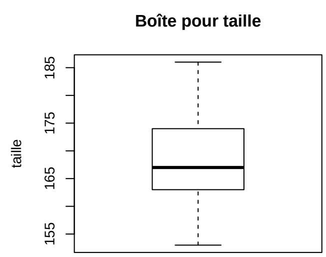
Voir les deux graphiques ci-dessus pour la variable '`oeil`'.

(c) • On étudie la variable quantitative (ou numérique) '`taille`'. Pour les manipulations avec R, on renvoie donc à la section 3.4 et aux sections récapitulatives 9.1.1 et 9.1.3 du document de cours.

- Les différents résultats déterminés par R sont donnés dans le tableau suivant

noms	valeurs
moyenne	168.787879
écart-type	8.695392
Q_1 (quartile à 25 %)	163
médiane	167
Q_3 (quartile à 75 %)	174
minimum	153
maximum	186
nombre	33

- 



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

(2) (a) • On étudie le croisement de la variable qualitative (ou catégorielle) '`mainfourchette`' et de la variable qualitative (ou catégorielle) '`oeil`'. Pour les manipulations avec R, on renvoie donc à la section 5.5 et la section récapitulative 9.2.2 du document de cours.

- La table de contingence déterminée par R est donnée dans le tableau suivant

	d	g
ambidextre	0	1
d	14	8
g	4	5

Les autres résultats donnés par R sont les suivants :

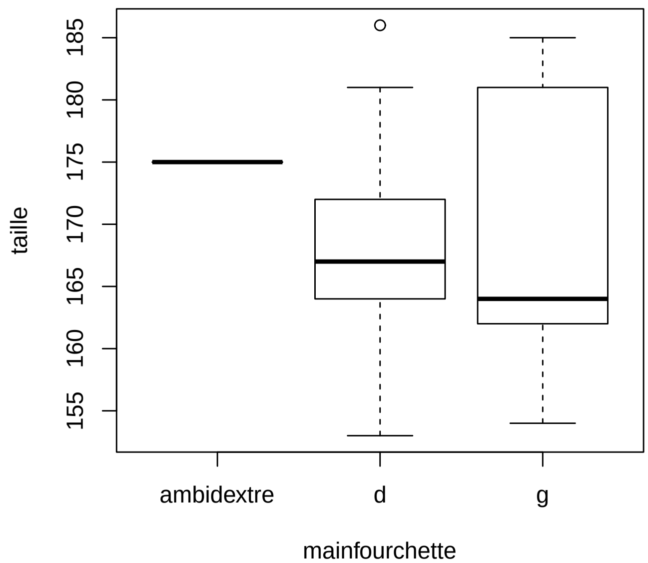
Noms des indicateurs	Valeurs
χ^2	2.283149
coefficient de Cramer V	0.263033
taille d'effet w	0.263033
probabilité critique p_c	0.319316

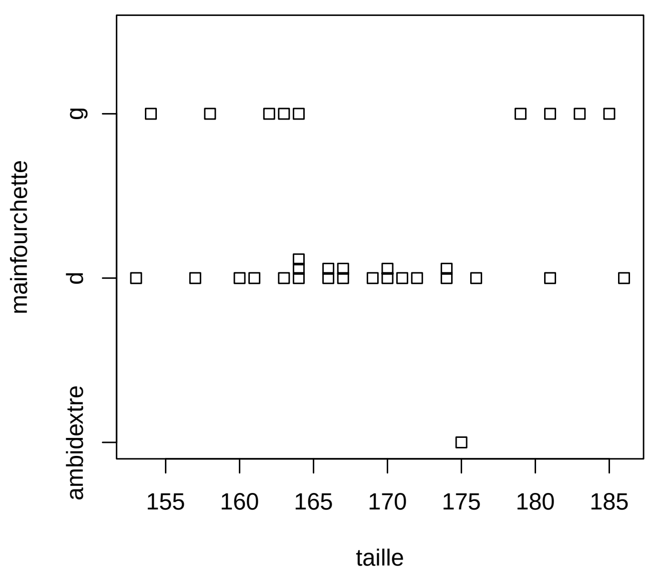
On compare la taille d'effet w=0.263033 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=0.319316 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

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

- On peut donc affirmer qu'il existe une relation moyenne non significative entre les variables 'mainfourchette' et 'oeil'.

(b)
- On étudie le croisement de la variable qualitative (ou catégorielle) '`mainfourchette`' et de la variable quantitative (ou numérique) '`taille`'. Pour les manipulations avec R, on renvoie donc aux sections 6.2 et 6.3 et la section récapitulative 9.2.3 du document de cours.
-





Voir la figure ci-dessous.

- Avec R, on obtient les statistiques par groupes données dans le tableau suivant ;

On rappelle que, dans ce tableau :

- le nombre noté `0%` est le quartile à 0 % (c'est le minimum) ;
- le nombre noté `25%` est le quartile à 25 % (c'est Q_1) ;
- le nombre noté `50%` est le quartile à 50 % (c'est la médiane) ;

	moyenne	écart-type	0%	25%	50%	75%	100%	n
ambidextre	175.00		175.00	175.00	175.00	175.00	175.00	1
d	167.95	7.57	153.00	164.00	167.00	171.75	186.00	22
g	169.89	11.96	154.00	162.00	164.00	181.00	185.00	9

- le nombre noté `75%` est le quartile à 75 % (c'est Q_3) ;
- le nombre noté `100%` est le quartile à 100 % (c'est le maximum).

Les graphiques et les statistiques par groupes montrent une certaine hétérogénéité entre les types.
Confirmons cela grâce à R.
Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.026766
probabilité critique p_c	0.674763

On compare le rapport de corrélation RC=0.026766 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=0.674763 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

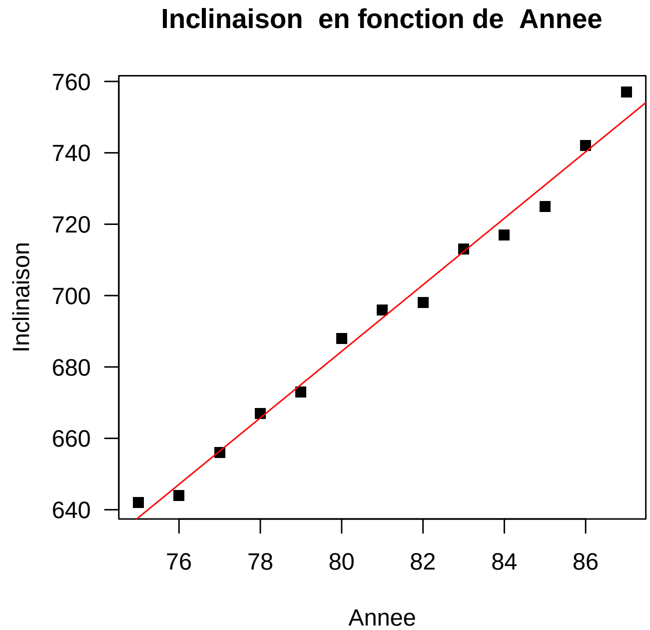
significativité pratique	**moyenne**
significativité statistique	**non**

- On peut donc affirmer qu'il existe une relation moyenne non significative entre les variables '`mainfourchette`' et '`oeil`'.

(c) Oui, on pouvait s'attendre aux résultats observés !

Correction de l'exercice 2.
Cet exercice avait été proposé par Stéphane Champely aux M1PPMR en 2007.

- On étudie le croisement de la variable quantitative (ou numérique) '`Annee`' et de la variable quantitative (ou numérique) '`Inclinaison`'. Pour les manipulations avec R, on renvoie donc à la section 4.5 et la section récapitulative 9.2.1 du document de cours.
- Voir la figure ci-dessous.



Sur cette figure, les points semblent alignés.

- Confirmons cela grâce à R.

Les résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
pente a	9.318681
ordonnée à l'origine b	-61.120879
corrélation linéaire r	0.993972
probabilité critique p_c	6.50337e-12

On compare la valeur absolue de la corrélation linéaire r =0.993972 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=6.50337e-12 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 '`Annee`' et '`Inclinaison`'.

Les ingénieurs ont donc du soucis à se faire ! En effet, l'inclinaison passe de la première à la dernière année de 642 dixièmes de millimètres, (soit 6.42 cm) à 757 dixièmes de millimètres, (soit 7.57 cm) ; le modèle linéaire représente bien la réalité : l'inclinaison passe de donc de

$$a \times Annee + b = 9.318681 \times 75 + (-61.120879) = 637.78022$$

à

$$a \times Annee + b = 9.318681 \times 87 + (-61.120879) = 749.604396$$

et ceci avec une évolution constante.

La question qu'on ne peut résoudre est l'évolution ultérieure de cette inclinaison !

Correction de l'exercice 3.

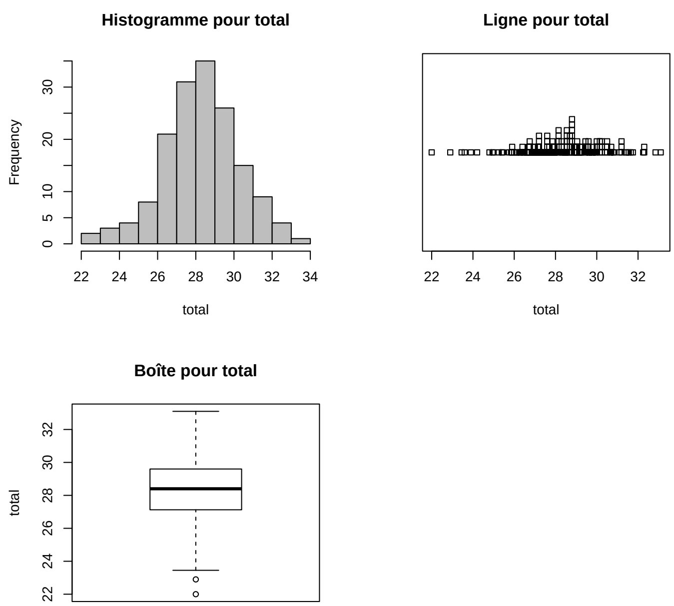
Cet exercice provient du site `http://pbil.univ-lyon1.fr/R/enseignement.html`

Ces données et les questions avaient été par proposées par Maud Tournoud, (INSA 4 ième année BIM, 2003).

(1)
- On étudie la variable quantitative (ou numérique) '`total`'.
- Les différents résultats déterminés par R sont donnés dans le tableau suivant

noms	valeurs
moyenne	28.319
écart-type	1.974
Q_1 (quartile à 25 %)	27.125
médiane	28.4
Q_3 (quartile à 75 %)	29.6
minimum	22
maximum	33.1
nombre	159

-

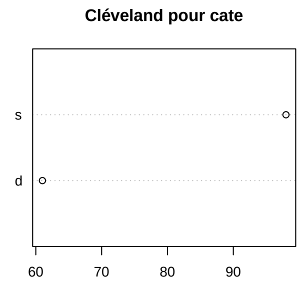


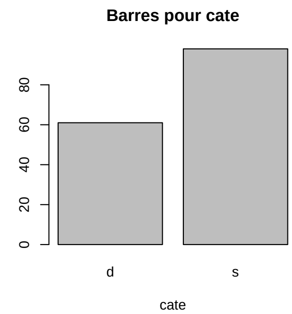
Voir les trois graphiques ci-dessus pour la variable '`total`'.

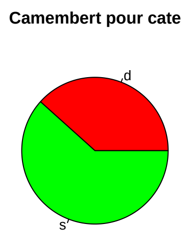
(2)
- On étudie la variable qualitative (ou catégorielle) '`cate`'.
- Les effectifs et les pourcentages déterminés par R sont donnés dans le tableau suivant

	effectifs	pourcentages
d	61	38.365
s	98	61.635

•



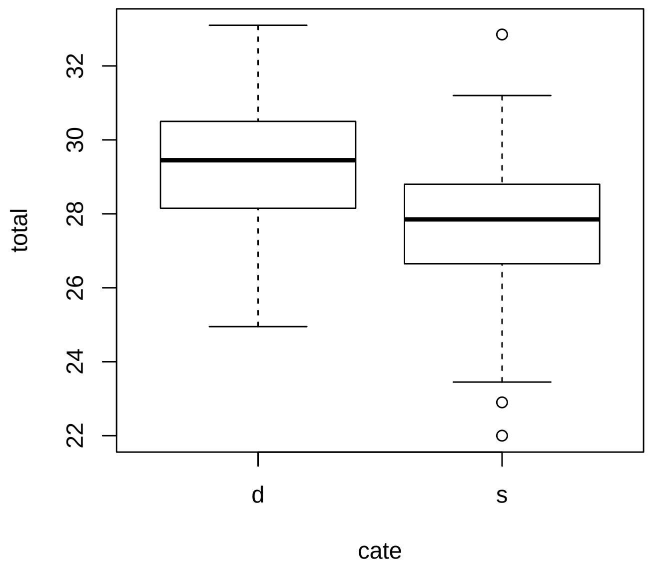


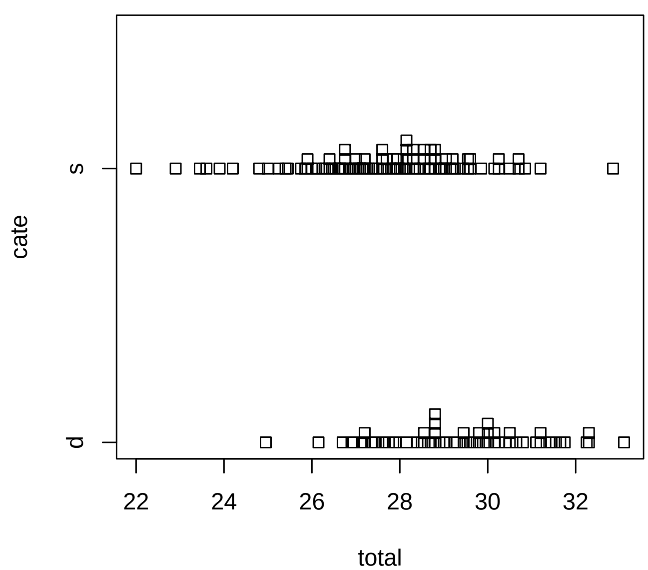


Voir les trois graphiques ci-dessus pour la variable '`cate`'.

(3) (a) • On étudie le croisement de la variable quantitative (ou numérique) '`total`' et de la variable qualitative (ou catégorielle) '`cate`'. Pour les manipulations avec R, on renvoie donc aux sections 6.2 et 6.3 et la section récapitulative 9.2.3 du document de cours.

•





Voir la figure ci-dessous.

- Avec R, on obtient les statistiques par groupes données dans le tableau suivant ;

	moyenne	écart-type	0%	25%	50%	75%	100%	n
d	29.35	1.71	24.95	28.15	29.45	30.50	33.10	61
s	27.68	1.86	22.00	26.66	27.85	28.80	32.85	98

Les graphiques et les statistiques par groupes montrent une certaine hétérogénéité entre les types.
Confirmons cela grâce à R.
Les autres résultats donnés par R sont les suivants :

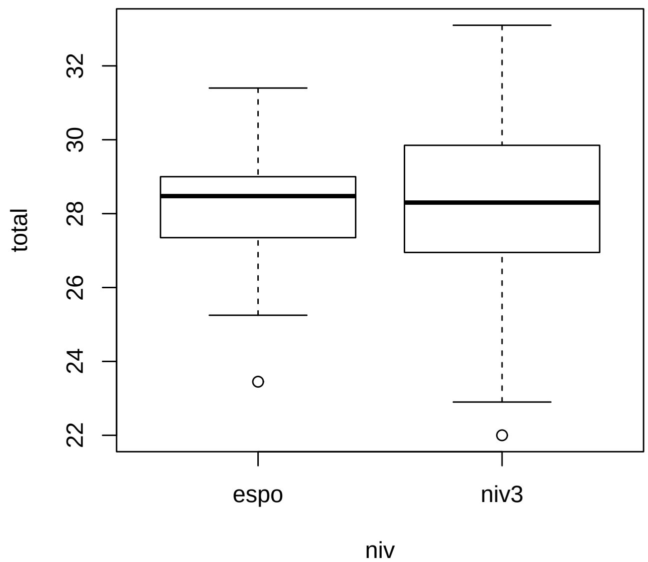
Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.170979
probabilité critique p_c	6.05141e-08

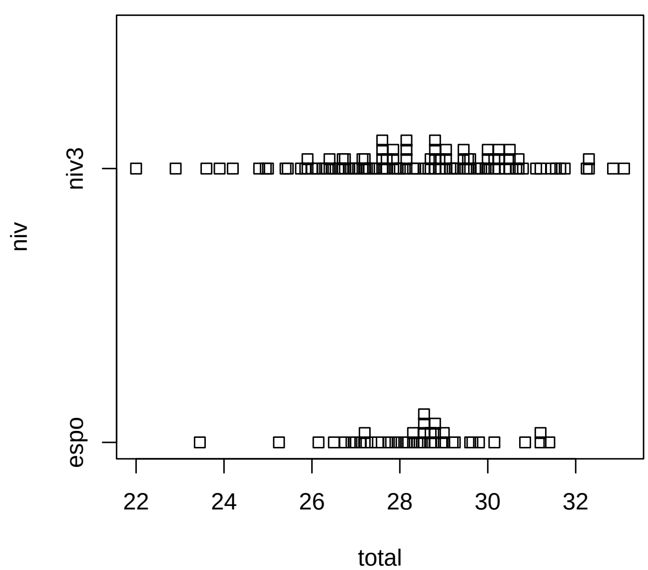
On compare le rapport de corrélation RC=0.170979 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=6.05141e-08 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

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

- On peut donc affirmer qu'il existe une relation entre les variables '`total`' et '`cate`'. Ceci semble logique !

(b) • On étudie le croisement de la variable quantitative (ou numérique) '`total`' et de la variable qualitative (ou catégorielle) '`niv`'.
  •





Voir la figure ci-dessous.

- Avec R, on obtient les statistiques par groupes données dans le tableau suivant ;

	moyenne	écart-type	0%	25%	50%	75%	100%	n
espo	28.31	1.50	23.45	27.39	28.48	28.99	31.40	46
niv3	28.32	2.14	22.00	26.95	28.30	29.85	33.10	113

Les graphiques et les statistiques par groupes montrent peu de différence entre les niveau. Confirmons cela grâce à R.
Les autres résultats donnés par R sont les suivants :

Noms des indicateurs	Valeurs
Rapport de corrélation RC	1.5e-05
probabilité critique p_c	0.96154

On compare le rapport de corrélation RC=1.5e-05 aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique p_c=0.96154 à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison :

significativité pratique	**faible**
significativité statistique	**non**

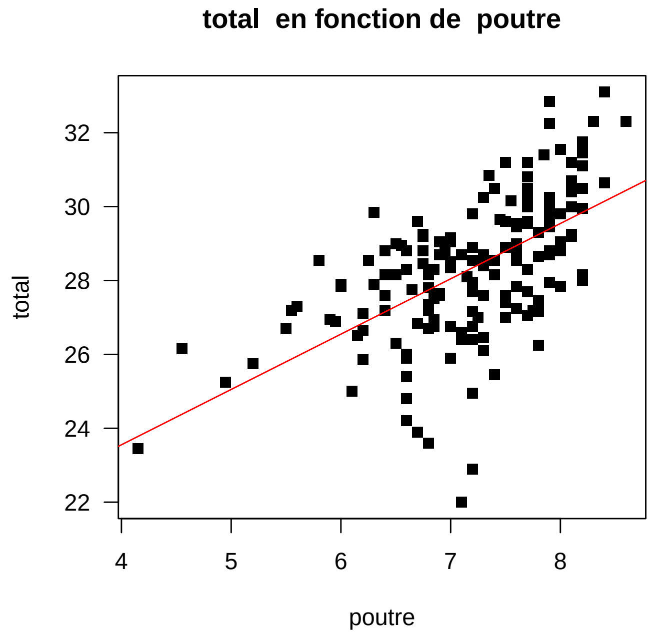
- On peut donc affirmer qu'il n'existe pas de relation entre les variables '`total`' et '`niv`'. Ceci semble un peu paradoxal !

(c) (i) Nous obtenons, agrès par agrès les moyennes et écarts-type suivants :

agrès	moyenne	écart-type
poutre	7.1827	0.7795
barres	7.1154	0.8762
saut	6.9733	0.7176
sol	7.0481	0.7254
total	28.3195	1.974

Nous ne pouvons pas, ici, considérer cette étude comme le croisement d'une variable qualtitative et d'une variable quantitative, mais nous en inspirer. Les écart-types sont du même ordre, excepté celui de la barre, épreuve qui semble être donc plus variée, donc plus selective. Les moyennes sont du même ordre, sauf celle du saut, un peu plus faible que les autres, donc moins bien notée.

(ii)
- On étudie le croisement de la variable quantitative (ou numérique) '`poutre`' et de la variable quantitative (ou numérique) '`total`'.
- Voir la figure ci-dessous.



Sur cette figure, les points semblent alignés.

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

Noms des indicateurs	Valeurs
pente a	1.497487
ordonnée à l'origine b	17.563491
corrélation linéaire r	0.591369
probabilité critique p_c	2.26474e-16

On compare la valeur absolue de la corrélation linéaire r =0.591369 aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique p_c=2.26474e-16 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 '`poutre`' et '`total`'.

De même, si nous étudions successivement la relation entre la note totale et la note obtenue au premier agrès, puis au second, puis au troisième, puis au quatrième, on obtient les résultats suivants :

relation étudiée	pente a	corrélation linéaire r	probabilité critique p_c
'poutre'-'total'	1.4975	0.5914	2.265e-16
'barres'-'total'	1.6364	0.7263	2.387e-27
'saut'-'total'	1.5414	0.5603	1.592e-14
'sol'-'total'	1.7799	0.6541	8.848e-21

Au vu de ces nombres, on peut conclure que dans tous les cas, il y a significativité statistique et que la significativité pratique est très forte. Les corrélations sont similaires. On peut tout de même observer une différence entre les pentes : plus la pente est forte, plus la note de l'agrès joue sur le total. L'ordre, du plus important au moins important, serait donc :
– sol
– barres
– saut
– poutre

En fait, il faudrait, pour être plus rigoureux, procéceder à une analyse par régression multilinéaire, hors programme ! Voir par exemple

`http://fr.wikipedia.org/wiki/Régression_multilinéaire`

Références

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