



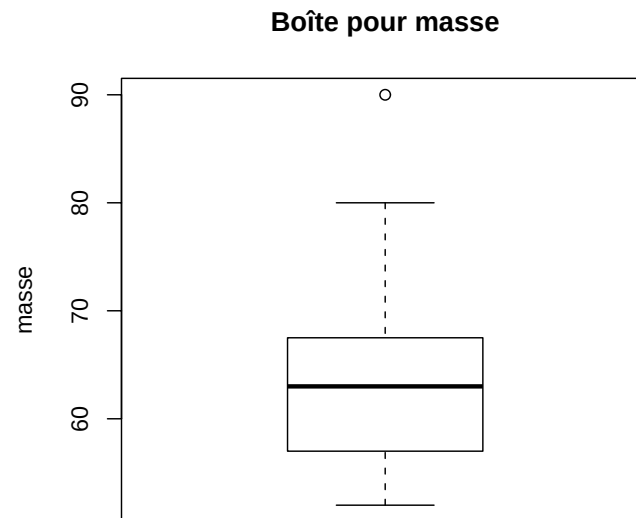
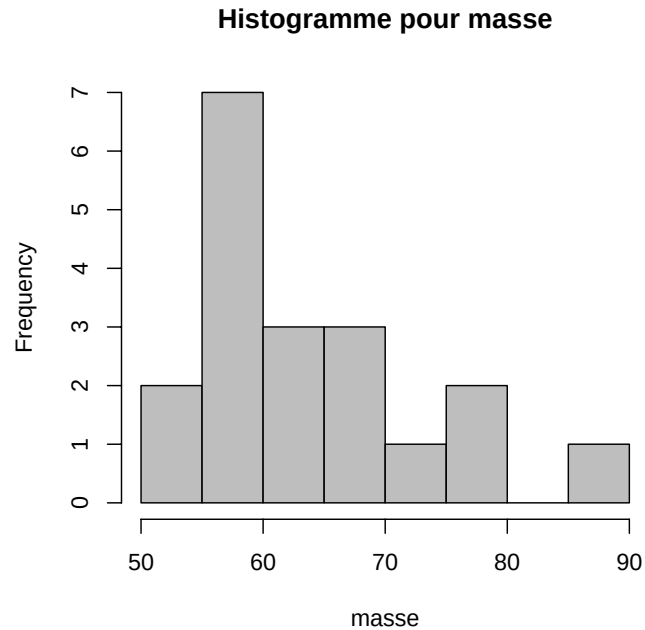
Corrigé de l'examen CCF2 de statistiques
--

Correction de l'exercice 1.

- (1) • On étudie la variable quantitative (ou numérique) 'masse'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 3.4 et aux sections récapitulatives 7.1.1 et 7.1.3 du document de cours.
- Les différents résultats déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

noms	valeurs
moyenne	64.53
écart-type	10.1
Q_1 (quartile à 25 %)	57
médiane	63
Q_3 (quartile à 75 %)	67.5
minimum	52
maximum	90
nombre	19

•



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

- (2) Vu que le nombre de données est "grand" (19 individus), les graphes les plus adaptés sont l'histogramme et la boîte à moustache.
- (3) On les trace avec les commandes

```
M1IGAPASA10data <- read.table("M1IGAPASA10data.txt", h = T)
hist(M1IGAPASA10data)
```

et

```
boxplot(M1IGAPASA10data)
```

- (4) On calcule la moyenne, l'écart-type et les quartiles de cette variable en tapant

```
summary(M1IGAPASA10data)
sd(M1IGAPASA10data)
```

ce qui fournit

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 52.00  57.00  63.00  64.53  67.50  90.00
[1] 10.10153
```

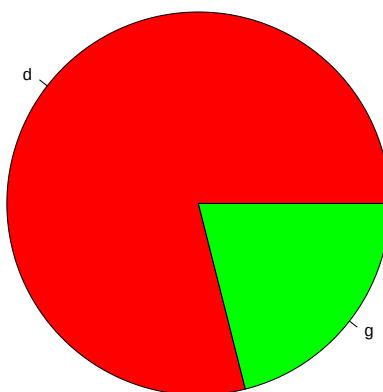
Correction de l'exercice 2.

- (1) (a) • On étudie la variable qualitative (ou catégorielle) 'main.ecriture'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 3.3 et aux sections récapitulatives 7.1.1 et 7.1.2 du document de cours.
- Les effectifs et les pourcentages déterminés par $\mathbb{R}$ (en prenant en compte les éventuelles valeurs non déterminées (NA)) sont donnés dans le tableau suivant

	effectifs	pourcentages
d	15	78.947
g	4	21.053

•

Camembert pour main.ecriture



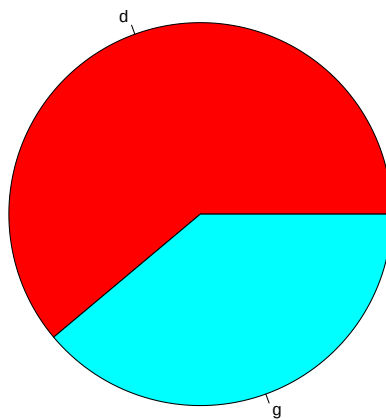
Voir le camembert ci-dessus pour la variable 'main.ecriture'.

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

	effectifs	pourcentages
d	11	61.111
g	7	38.889

•

Camembert pour main.fourchette



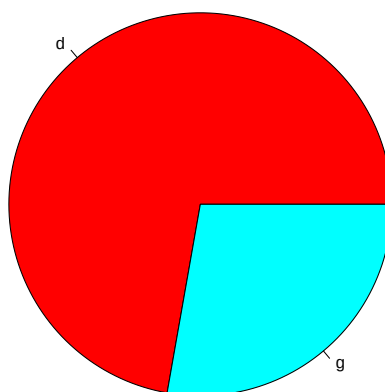
Voir le camembert ci-dessus pour la variable `'main.fourchette'`.

- (c) • On étudie la variable qualitative (ou catégorielle) 'oeil'.
- Les effectifs et les pourcentages déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

	effectifs	pourcentages
d	13	72.222
g	5	27.778

•

Camembert pour oeil



Voir le camembert ci-dessus pour la variable 'oeil'.

Il semblerait que la proportion des droitiers, quelque que soit la latéralité, soit d'environ 2/3. Reste à savoir si ces sont les mêmes personnes qui sont latéralisées ainsi. On passe donc à l'analyse bivariée.

- (2) (a) • On étudie le croisement de la variable qualitative (ou catégorielle) 'main.ecriture' et de la variable qualitative (ou catégorielle) 'main.fourchette'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 5.5 et la section récapitulative 7.2.2 du document de cours.
- La table de contingence déterminée par $\mathbb{R}$ est donnée dans le tableau suivant

	d	g
d	11	3
g	0	4

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

Noms des indicateurs	Valeurs
χ^2	8.081633
coefficient de Cramer V	0.652188
taille d'effet w	0.652188
probabilité critique p_c	0.00447161

On compare la taille d'effet $w=0.652188$ aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique $p_c=0.00447161$ 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

- Au vu de la table de contingence, on peut constater qu'il y a, *a priori* beaucoup de "droitiers-droitiers".
On peut finalement affirmer qu'il existe une relation entre les variables 'main.ecriture' et 'main.fourchette'.
- (b) • On étudie le croisement de la variable qualitative (ou catégorielle) 'main.ecriture' et de la variable qualitative (ou catégorielle) 'oeil'.
- La table de contingence déterminée par $\mathbb{Q}$ est donnée dans le tableau suivant

	d	g
d	12	2
g	1	3

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

Noms des indicateurs	Valeurs
χ^2	5.716484
coefficient de Cramer V	0.548514
taille d'effet w	0.548514
probabilité critique p_c	0.0168064

On compare la taille d'effet $w=0.548514$ aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique $p_c=0.0168064$ 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

- Au vu de la table de contingence, on peut constater qu'il y a, *a priori* beaucoup de "droitiers-droitiers".

On peut finalement affirmer qu'il existe une relation entre les variables 'main.écriture' et 'oeil'.

(c) Comme on pouvait s'y attendre, les trois latéralités sont donc très liées entre elles!

On peut aussi remarquer que la relation entre latéralité écriture et latéralité fourchette est un peu plus élevée que la relation entre latéralité écriture et latéralité œil (par exemple, comparer les coefficients de cramer qui valent respectivement 0.6522 et 0.5485).

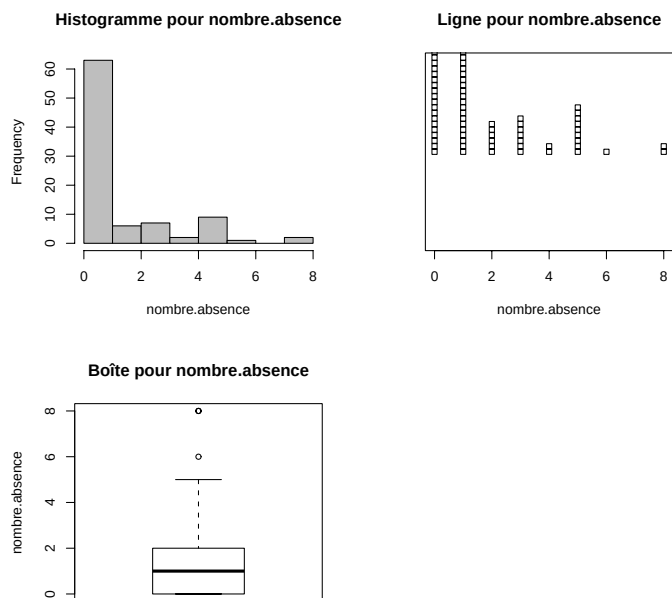
Correction de l'exercice 3.

On étudie le fichier de données 'notesCCF2A09'.

- (1) (a) (i) • On étudie la variable quantitative (ou numérique) 'nombre.absence'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 3.4 et aux sections récapitulatives 7.1.1 et 7.1.3 du document de cours.
- Les différents résultats déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

noms	valeurs
moyenne	1.44
écart-type	1.96
Q_1 (quartile à 25 %)	0
médiane	1
Q_3 (quartile à 75 %)	2
minimum	0
maximum	8
NA	4
nombre	94

•

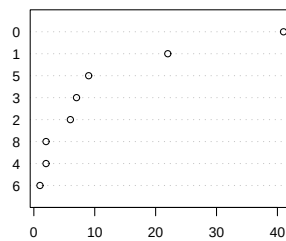
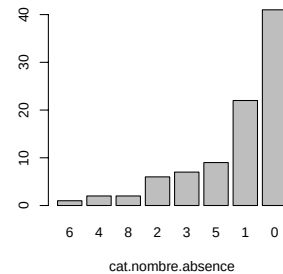


Voir les trois graphiques ci-dessus pour la variable 'nombre.absence'. On constate que le nombre essentiel d'absences est 0 ou 1 !

- (ii) • On étudie la variable qualitative (ou catégorielle) '`cat.nombre.absence`'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 3.3 et aux sections récapitulatives 7.1.1 et 7.1.2 du document de cours.
- Les effectifs et les pourcentages déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

	effectifs	pourcentages
6	1	1.111
4	2	2.222
8	2	2.222
2	6	6.667
3	7	7.778
5	9	10.000
1	22	24.444
0	41	45.556

•

Clévieland pour `cat.nombre.absence`Barres pour `cat.nombre.absence`Camembert pour `cat.nombre.absence`

Voir les trois graphiques ci-dessus pour la variable '`nombre.absence.cat`'. la encore, on constate que le nombre essentiel d'absences est 0 ou 1 !

La donnée '`nombre.absence`' est *a priori* numérique, mais le nombre de ces modalités (nombre de valeurs possibles) est faible. L'avantage de l'avoir transformé en donnée catégorielle est de voir apparaître sur les graphiques et les tableaux, chacune de ces valeurs !

Attention, les proportions calculées dans le tableau ci-dessus ne correspondent pas à la variable '`pourcentage.absence`' évoquée dans la question 1b.

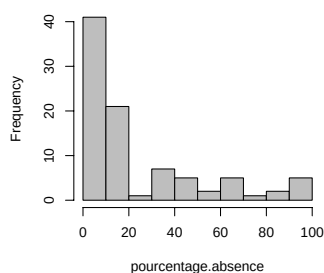
- (b) Si on regarde la variable '`nombre.appel`', on constate que ce nombre est différent selon les groupes de TD. Il est donc important, pour obtenir, des renseignements de même nature pour les étudiants, de regarder le pourcentage de présence. La question 1a n'est donc pas si pertinente !

- On étudie la variable quantitative (ou numérique) '`pourcentage.absence`'.
- Les différents résultats déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

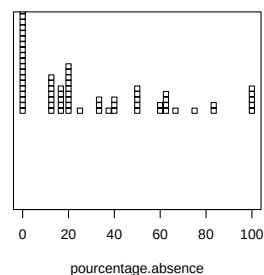
noms	valeurs
moyenne	22.907444
écart-type	29.66142
Q_1 (quartile à 25 %)	0
médiane	12.5
Q_3 (quartile à 75 %)	39.375
minimum	0
maximum	100
NA	4
nombre	94

•

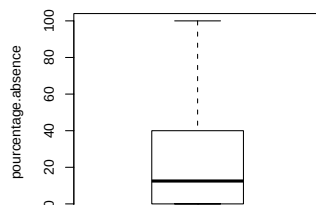
Histogramme pour pourcentage.absence



Ligne pour pourcentage.absence



Boîte pour pourcentage.absence

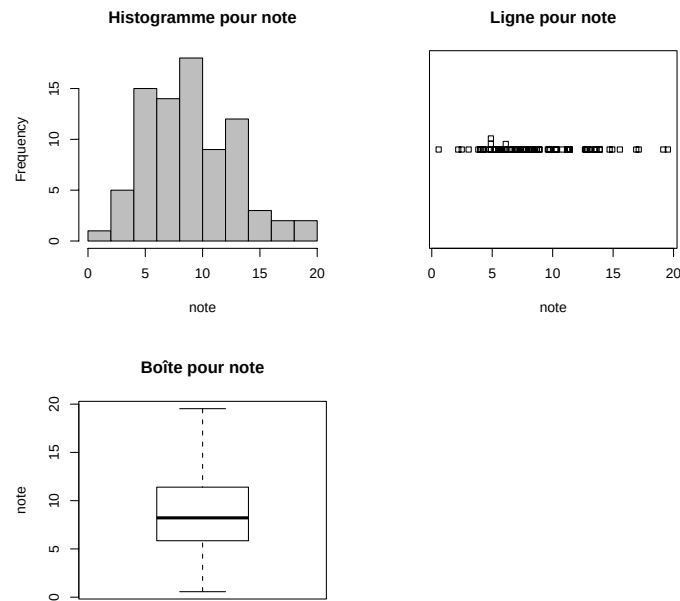


Voir les trois graphiques ci-dessus pour la variable '`pourcentage.absence`'. On constate que les proportions les plus représentées sont comprises entre 0 et 20 % ! Les proportions sont différentes de celles déterminées lors de la question 1(a)ii.

- (2) (a) • On étudie la variable quantitative (ou numérique) '`note`'.
- Les différents résultats déterminés par $\mathbb{R}$ sont donnés dans le tableau suivant

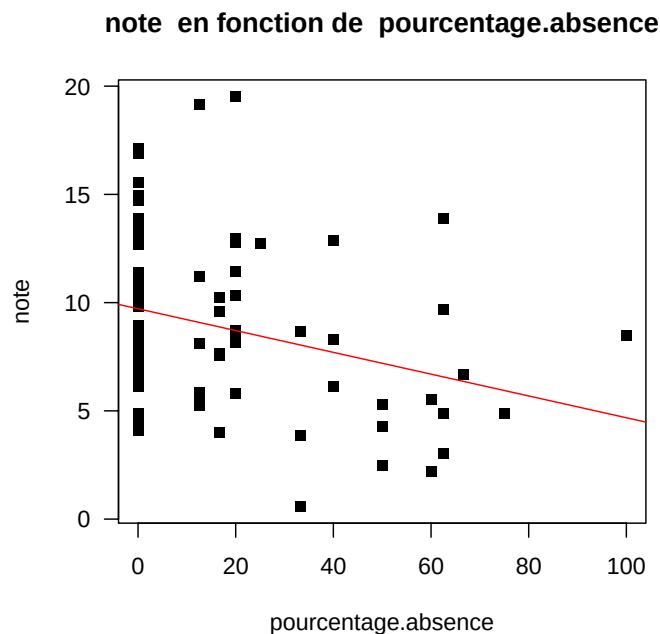
noms	valeurs
moyenne	8.87
écart-type	3.97
Q_1 (quartile à 25 %)	5.85
médiane	8.22
Q_3 (quartile à 75 %)	11.4
minimum	0.57
maximum	19.53
NA	13
nombre	94

•



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

- (b) Au vu des graphiques tracés, on se rend compte que l'ensemble des notes est compris entre 5 et 10.
- (3) • On étudie le croisement de la variable quantitative (ou numérique) 'pourcentage.absence' et de la variable quantitative (ou numérique) 'note'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc à la section 4.5 et la section récapitulative 7.2.1 du document de cours.
- Voir la figure ci-dessous.



Sur cette figure, les points semblent moyennement alignés.

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

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

Noms des indicateurs	Valeurs
pente a	-0.050338
ordonnée à l'origine b	9.716228
corrélation linéaire r	-0.292502
probabilité critique p_c	0.0080534

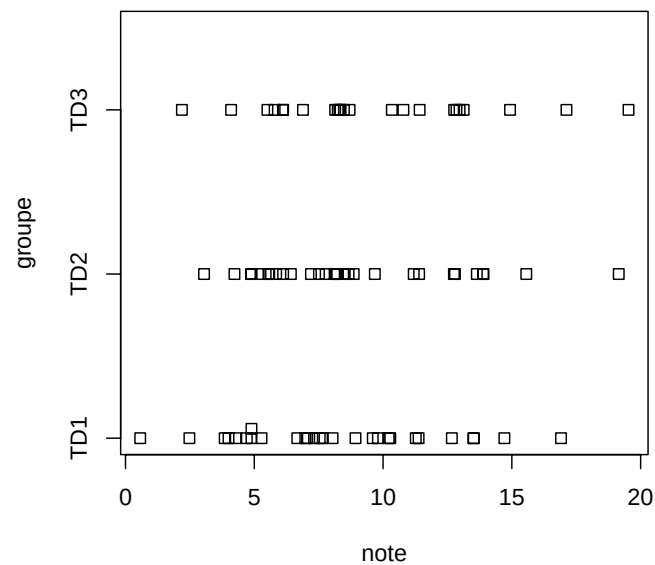
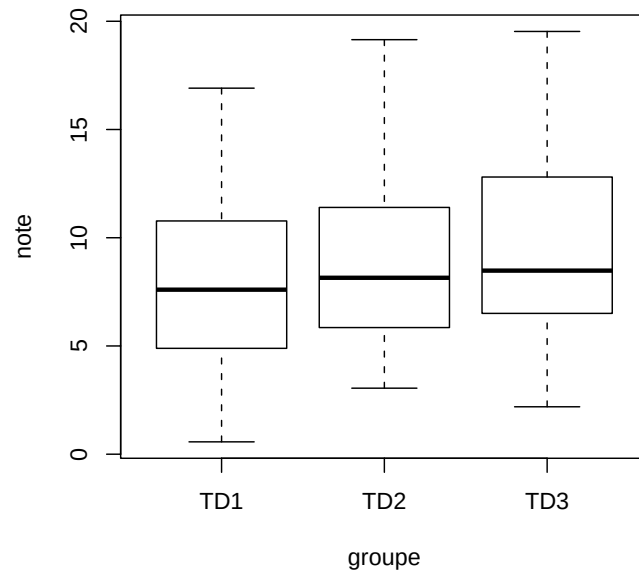
On compare la valeur absolue de la corrélation linéaire $r = -0.292502$ aux seuils de Cohen (0.1, 0.3, 0.5) (voir [Coh92]) et la probabilité critique $p_c = 0.0080534$ à la valeur seuil de la probabilité critique 0.05 et on déduit les résultats suivants sur la significativité de la liaison linéaire :

significativité pratique	moyenne
significativité statistique	oui

- On peut donc affirmer il existe une relation moyenne entre les variables 'pourcentage.absence' et 'note'. Compte tenu de la valeur de la pente (-0.050338), on vient donc de montrer qu'une augmentation de 10 % de la proportion d'absence entraîne (en moyenne) une diminution de la note de $= 10 \times 0.050338 = 0.503381$!

- (4) (a) • On étudie le croisement de la variable qualitative (ou catégorielle) 'groupe' et de la variable quantitative (ou numérique) 'note'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc aux sections 6.2 et 6.3 et la section récapitulative 7.2.3 du document de cours.

•



Voir la figure ci-dessous.

- Avec $\mathbb{R}$, 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) ;

	mean	sd	0%	25%	50%	75%	100%	n	NAs
TD1	8.18	3.94	0.57	4.89	7.60	10.53	16.91	28	3
TD2	8.90	3.80	3.05	5.92	8.15	11.35	19.15	30	4
TD3	9.69	4.21	2.19	6.50	8.48	12.80	19.53	23	2

- 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 a priori que les notes dépendent des 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	0.022805
probabilité critique p_c	0.406692

On compare le rapport de corrélation $RC=0.022805$ aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique $p_c=0.406692$ 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

- (b) On peut donc affirmer qu'il existe une relation moyenne entre les variables 'groupe' et 'note', mais attention au fait que c'est peut-être dû au hasard !

(5) *Question facultative*

- On étudie le croisement de la variable qualitative (ou catégorielle) 'cat.pourcentage.absence' et de la variable qualitative (ou catégorielle) 'cat.note'.
- La table de contingence déterminée par $\mathbb{R}$ est donnée dans le tableau suivant

	FALSE	TRUE
FALSE	4	1
TRUE	5	80

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

Noms des indicateurs	Valeurs
χ^2	28.823529
coefficient de Cramer V	0.553745
taille d'effet w	0.553745
probabilité critique p_c	7.92821e-08

On compare la taille d'effet $w=0.553745$ aux seuils de Cohen (0.1,0.3,0.5) (voir [Coh92]) et la probabilité critique $p_c=7.92821e-08$ 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 'cat.pourcentage.absence' et 'cat.note'. Ainsi, un étudiant qui ne vient pas en TD aura très peu de chance de venir à l'examen, et réciproquement !

- (6) (a) (i) Si on veut croiser la égale à K et la variable 'groupe', il faut d'abord créer une variable numérique, par exemple notée 'coeffbrut', qui vaut $K = 1.1$ pour la groupe 1, $K = 1$ pour la groupe 2, $K = 1.05$ pour la groupe 3. Pour cela, on peut taper (un peu comme dans l'exercice 3.11 page 31) :

```
coeffbrut<-numeric(length(na.omit(notesCCF2A09$groupe)))
coeffbrut[notesCCF2A09$groupe=="TD1"]<-1.1
coeffbrut[notesCCF2A09$groupe=="TD2"]<-1
coeffbrut[notesCCF2A09$groupe=="TD3"]<-1.05
```

Il ne reste plus qu'à croiser !

On obtient un RC égal à

$$RC = 1$$

et une probabilité critique égale à

$$p_c = 0$$

La liaison est donc très forte et statistiquement significative!!!

- (ii) Cela est tout à fait normal, en fait ! Rappelons les formules théoriques ((6.1), page 74 et (6.2), page 74)

$$\underbrace{\sum_{j=1}^G \sum_{i=1}^{n_j} (y_{ij} - M)^2}_{SC_{Totale}} = \underbrace{\sum_{j=1}^G n_j (M_j - M)^2}_{SC_{Inter-Groupes}} + \underbrace{\sum_{j=1}^G \sum_{i=1}^{n_j} (y_{ij} - M_j)^2}_{SC_{Intra-Groupes}}. \quad (1)$$

et

$$RC = \frac{SC_{Inter-Groupes}}{SC_{Totale}}$$

Ici, dans chaque groupe, la valeur de y_{ij} vaut M_j donc, pour tout j , la somme

$$\sum_{i=1}^{n_j} (y_{ij} - M_j)^2$$

est nulle ; ainsi, la seconde somme ($SC_{Intra-Groupes}$) est nulle, la première ($SC_{Inter-Groupes}$) est donc égale au terme de gauche (SC_{Totale}) et donc

$$RC = \frac{1}{1} = 1.$$

- (b) On peut vérifier que sur l'intervalle $[0, 40]$, la valeur de 'coefficient' est égale à K . Sur l'intervalle $[40, 100]$, c'est une fonction affine de a , qui vaut K en 40 et un en 100, car

$$-\frac{K-1}{60}a + K + \frac{K-1}{60} \times 40 = -\frac{K-1}{60} \times 100 + K + \frac{K-1}{60} \times 40 = \frac{-100(K-1) + 60K + 40(K-1)}{60} = 1$$

Son graphe est donc donné par la figure 1.

Ainsi, un élève qui a une absence inférieure à 40% ne verra pas sa note bouger, tandis qu'un élève qui a une absence supérieur à 40% verra sa note multipliée par un nombre décroissant en fonction de son absence. Cela est fait pour ne pas pénaliser les étudiants réguliers !

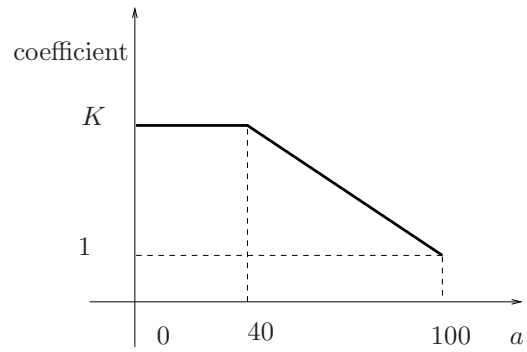
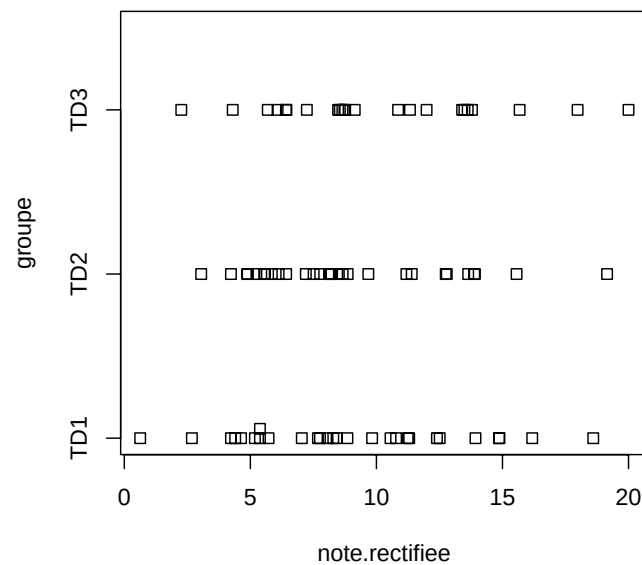
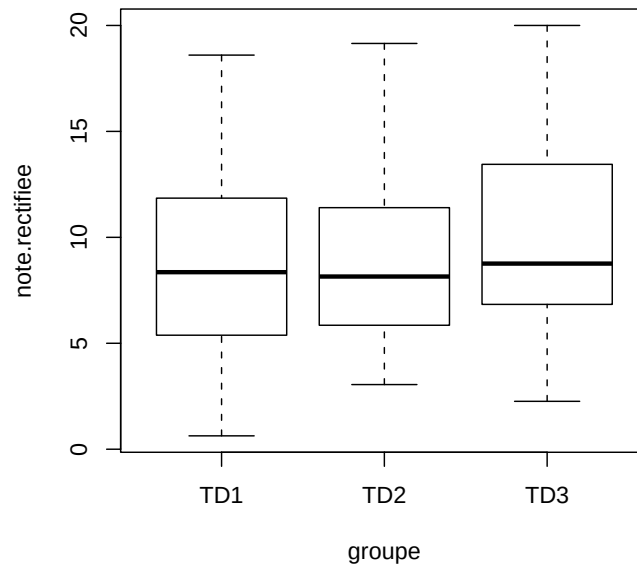


FIG. 1. L'allure de coefficient en fonction de l'absence a .

(c) Cette façon de procéder pour paraître juste (pour le prof!).

(d) • On étudie le croisement de la variable qualitative (ou catégorielle) '**groupe**' et de la variable quantitative (ou numérique) '**note.rectifiée**'. Pour les manipulations avec $\mathbb{R}$, on renvoie donc aux sections 6.2 et 6.3 et la section récapitulative 7.2.3 du document de cours.

•



Voir la figure ci-dessous.

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

Les graphiques et les statistiques par groupes montrent a priori que les notes dépendent des groupes.

Confirmons cela grâce à .

	mean	sd	0%	25%	50%	75%	100%	n	NAs
TD1	8.98	4.35	0.63	5.38	8.36	11.58	18.60	28	3
TD2	8.90	3.80	3.05	5.92	8.15	11.35	19.15	30	4
TD3	10.12	4.38	2.26	6.83	8.76	13.45	20.00	23	2

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

Noms des indicateurs	Valeurs
Rapport de corrélation RC	0.016961
probabilité critique p_c	0.513176

On compare le rapport de corrélation $RC=0.016961$ aux seuils de Cohen (0.01,0.05,0.15) (voir [Coh92]) et la probabilité critique $p_c=0.513176$ 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 entre les variables 'groupe' et 'note. rectifiée', mais attention au fait que c'est peut-être dû au hasard!
- (e) Peu de changement, en fait par rapport à la question 4.

Références

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