



Corrigé de l'examen CCF1 de statistiques
--

IMPORTANT : Corrigé pour les groupe 1 et 2

Correction de l'exercice 1.

Corrigé pour le groupe 1

Non rédigé

Correction de l'exercice 1.

Corrigé pour le groupe 2

Non rédigé

Pour les exercices suivants (2 à 4), voir les fichiers L3ETU1.xls et L3ETU2.xls disponibles sur SPIRAL.
Pour plus de lisibilité tous les tableaux et les figures sont renvoyés à la fin du corrigé, à partir de la page 4.

Correction de l'exercice 2.

On s'intéresse aux données (pour l'instant indépendamment l'une de l'autre) : sexe, poids et taille.
Pour le variable catégorielle sexe, voir le chapitre 3. Pour les variables numériques poids et taille, voir le chapitre 4.

- Dans les tableaux 1 page 4 et 2 page 5, on a indiqué le contenu des tableaux L3ETU1.xls et L3ETU2.xls.
- Avec Rcommander, pour la variable catégorielle sexe, on peut dénombrer les catégories ; pour le groupe 1, on trouve

F M
4 17

Il y a en tout 2 catégories. Les fréquences associées sont :

F M
19.04762 80.95238

soit 19.05 % pour la valeur F et 80.95 % pour la valeur M.

Pour le groupe 2, on trouve

F M
3 22

Il y a en tout 2 catégories. Les fréquences associées sont :

F M

12 88

soit 12 % pour la valeur F et 88 % pour la valeur M.

- Avec Rcommander, pour la variable catégorielle sexe, on peut tracer le camembert. Voir les figures 1 page 6 et 2 page 7. On peut aussi éventuellement tracer un graphe en barre. Voir les figures 3 page 7 et 4 page 8.
- Pour la variable numérique poids ou taille, avec Rcommander, on peut tracer un histogramme, une ligne de point en tapant la ligne de commande relative à `stripchart` du polycopié page 30, c'est-à-dire

```
stripchart(L3ETU1$Poids, method = "stack", main = "ligne de points")
```

ou une boîte de dispersion. Voir les figures 5 page 8 et 6 page 9. On peut déterminer la moyenne, l'écart-type et les quartiles (voir chapitre 4) : on obtient pour le groupe 1 (variable poids) :

```
mean      sd 0% 25% 50% 75% 100%  n
68.66667 10.32634 45  63  71  77   80 21
```

et pour le groupe 2 (variable taille) :

```
mean      sd 0% 25% 50% 75% 100%  n
176.96 6.698756 163 174 177 181 193 25
```

Correction de l'exercice 3.

Corrigé pour le groupe 1

On étudie maintenant la liaison entre la donnée poids et la donnée taille, qui sont toutes les deux numériques. Voir le chapitre 9.

- (1) On peut tracer le nuage de points où l'abscisse est poids (variable explicative) et l'ordonnée est la taille. Voir la figure 7 page 9.

Les points semblent être alignés et on conclue à l'existence d'une relation linéaire entre ces deux données *a priori*.

- (2) Confirmons cela grâce à \mathbb{R} .

En utilisant l'ajustement de modèle, on obtient une pente égale à $p = 0.96$ et une ordonnée à l'origine égale à -101.61 .

Le coefficient de détermination vaut $r^2 = 0.7649 = 76.4874\%$. D'après les tables de Cohen, pour un r^2 supérieur à 0.25, on a donc une forte relation linéaire.

On pouvait aussi calculer le coefficient de corrélation linéaire donné par

$$r = \text{signe}(p) \sqrt{r^2},$$

soit numériquement $r = 0.8746$ et comparer sa valeur absolue aux valeurs de seuil (0.1, 0.3, 0.5) données par Cohen.

La probabilité critique vaut $2.16892900484822e-07$, largement inférieure à 5% et on a donc une relation statistiquement significative.

Correction de l'exercice 3.

Corrigé pour le groupe 2

On étudie maintenant la liaison entre la donnée poids et la donnée sexe ; l'une est catégorielle et l'autre est numérique. Voir le chapitre 5.

- (1) On peut tracer une collection de boîte de dispersion ou de ligne de points (par groupe) : voir figure 8 page 10.

On peut aussi faire des statistiques par groupe :

	mean	sd	0%	25%	50%	75%	100%	n
F	53.25000	6.396614	45	50.25	54	57	60	4
M	72.29412	7.226381	55	69.00	72	80	80	17

Ainsi, *a priori*, le poids des hommes semble être plus élevé que celui des femmes et un peu plus hétérogène (écart type un peu plus grand).

- (2) Confirmons cela grâce à \mathcal{R} . En utilisant l'ajustement de modèle, on obtient un rapport de corrélation RC égal à $0.1999 = 19.9948\%$, supérieur à 0.15 et une probabilité critique égale à 0.0250187677498703 , inférieure à $.05$. On a donc une forte significativité pratique de la liaison et une relation statistiquement significative.

Correction de l'exercice 4.

Croisons maintenant la variable sexe et la variable expérience, toutes les deux catégorielle.

Les tableaux croisés sont indiqués dans les tableaux 3 page 5 et 4 page 5. On obtient

Pearson's Chi-squared test

data: xtable1

X-squared = 1.4852, df = 1, p-value = 0.2230

pour le groupe 1 et

Pearson's Chi-squared test

data: xtable2

X-squared = 0.7108, df = 1, p-value = 0.3992

pour le groupe 2.

On a donc la statistique du chi-carré $X^2 = 1.4852$ pour le groupe 1 et $X^2 = 0.7108$ pour le groupe 2.

On obtient donc une taille d'effet

$$w = \sqrt{\frac{X^2}{n}} = \sqrt{\frac{1.4852}{21}} = 0.265936$$

pour le groupe 1 et

$$w = \sqrt{\frac{X^2}{n}} = \sqrt{\frac{0.7108}{25}} = 0.168623$$

pour le groupe 2. On a donc, a vu des seuils conventionnels, une liaison moyenne.

Pour la probabilité critique, on obtient 0.223 pour le groupe 1 et 0.3992 pour le groupe 2. Elles sont supérieures à 0.05 donc pas de relation statistiquement significative.

On conclut donc que le fait d'avoir suivi un cours de statistiques en début de L3MOS ne dépend pas du sexe.

Ensemble des tableaux et des figures

	Pratique	Sport	Lettre	Nombre	Expérience	Utilité	Sexe	Age	Taille	Poids
1	Oui	VTT	S	1	Non	3	M	24	173	63
2	Oui	Basket	S	7	Oui	3	M	20	167	55
3	Oui	Tennis	Q	5	Non	3	M	22	177	72
4	Oui	Basket	S	8	Oui	3	F	21	166	56
5	Oui	Football	S	3	Oui	4	M	20	177	74
6	Oui	Snowboard	S	8	Non	3	M	23	176	65
7	Oui	Full Contact	S	6	Non	3	M	23	174	71
8	Oui	Aviron	Q	7	Oui	4	M	20	183	65
9	Oui	Escalade	S	2	Non	3	F	37	158	52
10	Oui	Football	Q	7	Non	3	M	24	181	80
11	Oui	Windsurf	S	7	Non	3	M	38	175	77
12	Oui	Danse	Q	2	Oui	3	M	22	185	80
13	Oui	Boxe	S	8	Non	3	M	23	178	72
14	Oui	Badminton	S	8	Oui	3	M	20	187	80
15	Oui	Boxe	Q	3	Non	3	M	22	179	69
16	Oui	Football	S	7	Non	3	M	20	185	71
17	Oui	Badminton	S	8	Non	3	M	21	187	80
18	Oui	Golf	S	5	Oui	3	F	21	153	45
19	Oui	Roller	S	3	Oui	4	F	21	178	60
20	Oui	Tennis	S	4	Oui	3	M	24	188	80
21	Oui	Football	S	10	Oui	4	M	20	185	75

TAB. 1. Le jeu de données L3ETU1 (sans les identifiants ni le groupe)

	Pratique	Sport	Lettre	Nombre	Expérience	Utilité	Sexe	Age	Taille	Poids
1	Oui	Volley	S	1	Non	3	F	20	163	52
2	Oui	Gymnastique	S	6	Non	3	F	21	172	58
3	Oui	Basket	Q	5	Non	3	M	23	170	62
4	Oui	Tennis	S	7	Non	3	M	23	182	70
5	Oui	Handball	S	6	Non	3	M	22	187	93
6	Oui	Football	S	1	Oui	3	M	22	175	65
7	Oui	Aviron	S	8	Oui	3	M	21	175	74
8	Oui	Tennis	Q	8	Non	4	M	20	178	80
9	Oui	Tennis	S	4	Oui	3	M	19	180	60
10	Non		S	7	Non	3	M	24	184	67
11	Oui	Tennis	S	9	Oui	3	M	20	193	70
12	Oui	Football	S	7	Oui	4	M	21	182	70
13	Oui	Rock acrobatique	S	7	Oui	3	M	21	174	70
14	Oui	Athlétisme	S	6	Non	3	M	20	177	67
15	Oui	Rugby	S	7	Oui	3	M	20	170	68
16	Oui	Football	S	6	Oui	3	M	22	175	60
17	Oui	Karaté	Q	7	Oui	4	M	21	178	71
18	Oui	Ski	S	7	Oui	2	F	21	170	69
19	Oui	Basket	Q	6	Oui	3	M	21	179	78
20	Oui	Athlétisme	S	7	Oui	4	M	20	176	71
21	Oui	Football	Q	1	Oui	3	M	21	179	78
22	Oui	Tennis	S	9	Non	3	M	21	175	65
23	Oui	Judo	Q	8	Non	3	M	22	165	73
24	Oui	Football	Q	9	Oui	3	M	22	181	78
25	Oui	Boxe Française	S	8	Non	3	M	20	184	85

TAB. 2. Le jeu de données L3ETU2 (sans les identifiants ni le groupe)

	F	M
Non	1	10
Oui	3	7

TAB. 3. Le tableau croisé expérience/sexe pour le groupe 1

	F	M
Non	2	9
Oui	1	13

TAB. 4. Le tableau croisé expérience/sexe pour le groupe 2

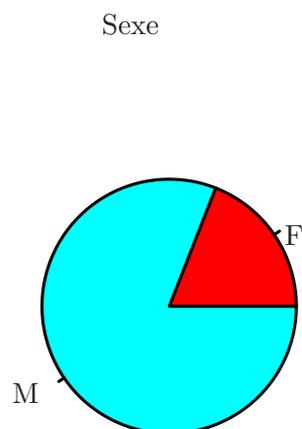


FIG. 1. La camembert de la variable sexe pour le groupe 1

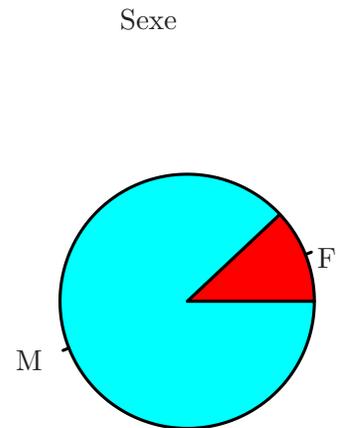


FIG. 2. La camembert de la variable sexe pour le groupe 2



FIG. 3. Le graphe en barre de la variable sexe pour le groupe 1



FIG. 4. Le graphe en barre de la variable sexe pour le groupe 2

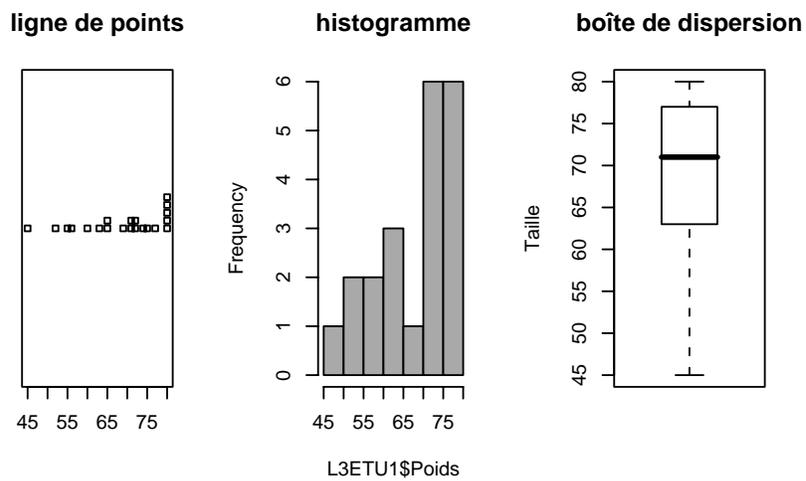


FIG. 5. la ligne de points, la boîte de dispersion et l'histogramme de la variable poids (pour le groupe 1).

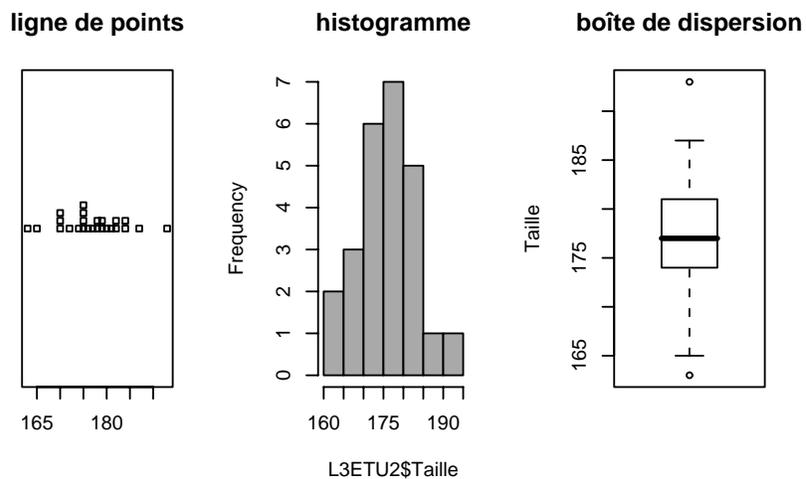


FIG. 6. la ligne de points, la boîte de dispersion et l'histogramme de la variable taille (pour le groupe 2).

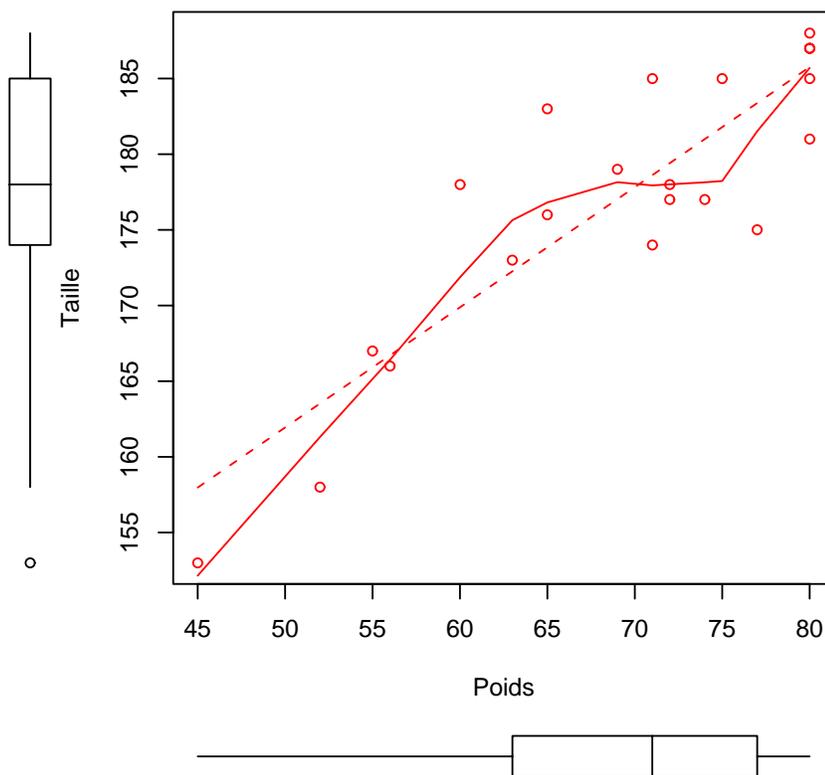


FIG. 7. Le nuage de points taille en fonction de poids (pour le groupe 1).

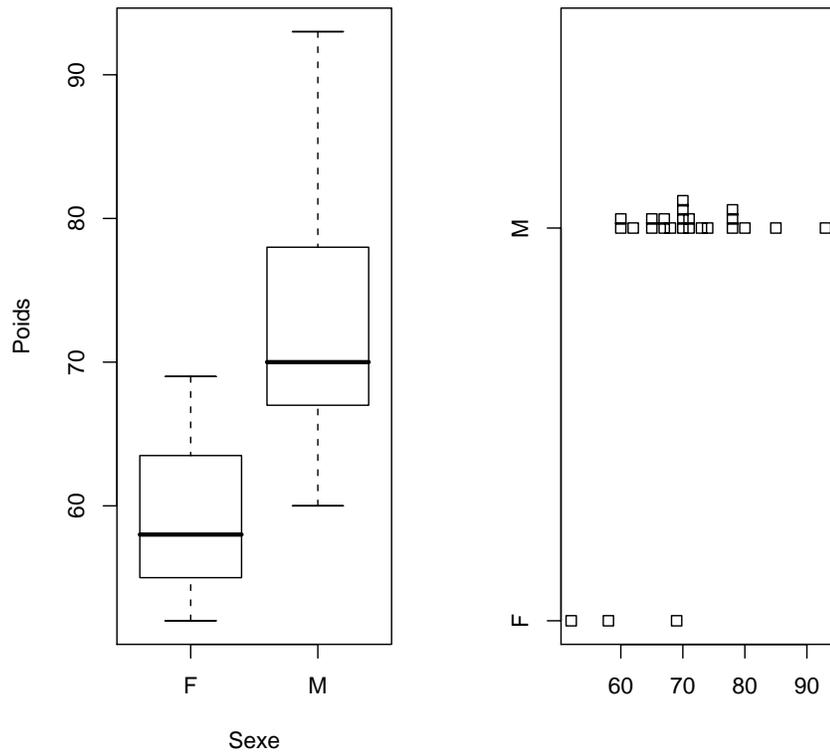


FIG. 8. collection de boîte de dispersion ou de ligne de points (pour le groupe 2).