



Rapport de SY09 : Analyse de données et data mining
UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE

Printemps 2017

Ccile HEIDSIECK et Simon LAURENT

Sujet du rapport :
Statistique descriptive et ACP

Département des étudiants :
Génie Biologique et Génie Informatique

Professeurs :
M. Sylvain ROUSSEAU
M Benjamin QUOST

Table des matières

1	Statistique Descriptive	4
1.1	Notes	4
1.1.1	Analyse 1	4
1.1.2	Analyse 2	5
1.1.3	Annexes	5

Chapitre 1

Statistique Descriptive

1.1 Notes

1.1.1 Analyse 1

Dans cette première partie, nous allons décrire les données du dataset Notes :
Pour l'intégralité des variables, le dataset a un volume de 296.

- nom : variable qualitative nominale, modalités "EtuX" avec X un chiffre ;
- specialit : variable qualitative nominale, modalités GB, GI, GM, GSM, GP, GSU, HuTech, ISS, TC ;
- niveau : variable quantitative discrète, modalités 1..6 ;
- statut : variable qualitative nominale, modalités : change, UTC ;
- dernier diplôme obtenu : variable qualitative nominale, 12 modalités avec certaines en N/A ;
- note mdian, note finale, note totale : variable quantitative continue, modalités entre 0 et 20 ;
- correcteur mdian, correcteur final : variable qualitative nominale, modalités "CorX" avec X un chiffre ;
- resultat : qualitative ordinale, modalités A>B>C>D>E>Fx>F

Les valeurs nulles pour les notes du mdian et du final sont des absences. Quant aux NAs pour la note finale et les résultats sont la conséquence des unions des valeurs nulles. Par ailleurs, on remarque que les étudiants étrangers ne disposent pas de dernier diplôme obtenu et que leur résultat est souvent NA ou Fx ou F.

La question que l'on se pose est "existe-t-il une corrélation entre les notes du mdian et du final?". Le résultat est le suivant 0.3861882. Ceci montre qu'il y

a une faible corrélation entre la réussite d'un médian et d'un final.

1.1.2 Analyse 2

Dans cette seconde partie, nous allons étudier les liens statistiques entre variables. Nous allons donc afficher un ensemble de graphes en annexe et les commenter. Le graphique 1.1 montre que 80% des résultats de SY02 se situent entre 10 et 15. Nous avons également un pic à 7.5. Ce qui correspond à un ensemble de personnes ayant rencontré les mêmes difficultés au sein de l'UV. Le graphique 1.2 démontre clairement que plus le niveau des étudiants est élevé, plus les résultats en SY02 diminuent. De même, avec le graphique 1.3 nous constatons que les étudiants en TC assistant au cours de SY02, ont obtenu de meilleurs résultats. En revanche les ISS ont obtenu des plus mauvais. Sachant que ISS correspond à une fin d'étude et un TC à un début, cela souligne ce que nous avons noté par avant concernant l'influence du niveau sur les notes. Enfin le graphique ?? souligne que les correcteurs 3, 4, 6 et 7 déplacent les résultats de facons significatives.

1.1.3 Annexes

FIGURE 1.1 – densité des notes en SY02

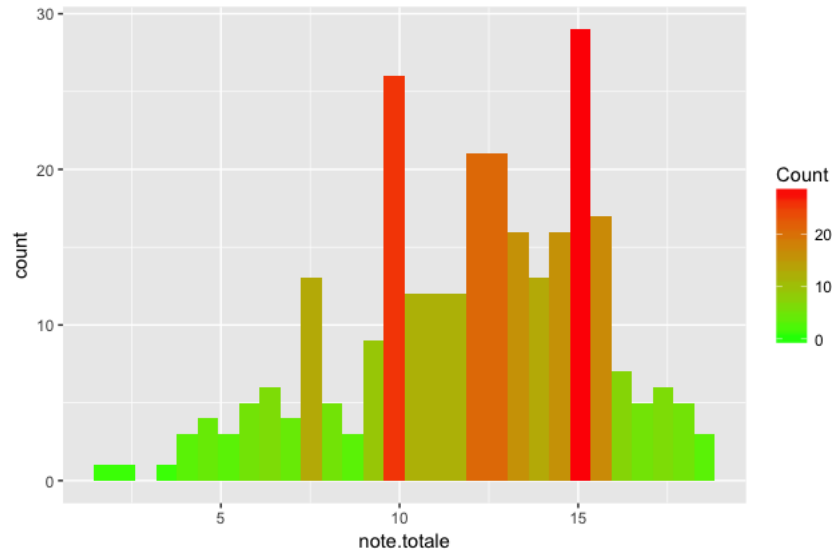


FIGURE 1.2 – influence du niveau sur les notes en SY02

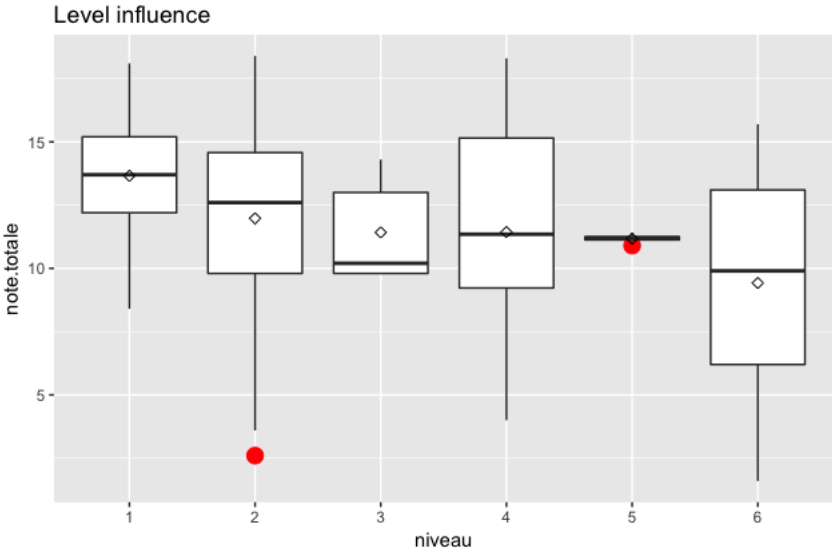


FIGURE 1.3 – influence de la specialit sur la note en SY02

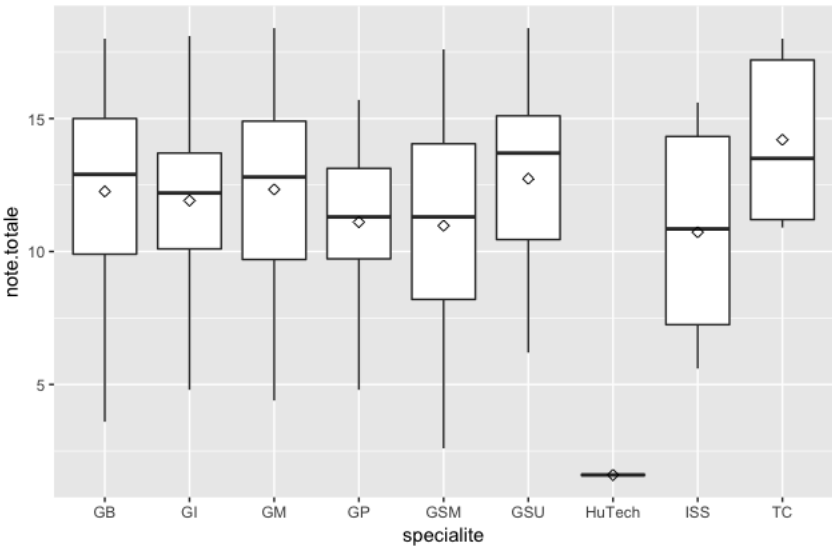


FIGURE 1.4 – influence du correcteur sur la note en SY02

