

## Master 1-Statistique économétrie

### Plans d'expériences-TD 1

Vous êtes chargés de conseiller les décideurs d'établissements publics dans divers domaines. Un directeur d'UFR d'une Université vous contacte pour vous poser la question suivante :

Des étudiants titulaires d'un Bac S sont recrutés pour une formation. La moyenne d'algèbre des étudiants ayant suivi l'option math est-elle significativement différente de celle des étudiants n'ayant pas suivi l'option math ?

Afin de vous éclairer sur la situation, le directeur d'UFR vous envoie un fichier contenant la liste des notes de mathématiques obtenues à l'examen d'algèbre de deuxième année des étudiants ayant suivi l'option mathématiques en terminale ainsi que celles des étudiants n'ayant pas suivi l'option mathématiques (voir fichier txt).

On effectue une ANOVA sur la population étudiée pour tester si la moyenne du groupe d'étudiants ayant suivi l'option math est égale à celle des étudiants n'ayant pas suivi l'option math.

Dans la suite on notera  $Y_{1i}$  la note d'algèbre du  $i$ ème candidat n'ayant pas suivi l'option math, et  $Y_{2i}$  la note d'algèbre du  $i$ ème candidat ayant suivi l'option math. On considère  $X_{ij}$  la variable prenant la valeur zéro si le candidat n'a pas suivi l'option math et 1 sinon.

Partie I :

Question 1 : Quelle est la population étudiée ? Quelles est (sont) la (les) variable(s) considérée(s) ? Quelle en est la nature (qualitative(s), quantitative(s)) ? Quel est le facteur ? La variable expliquée ?

Question 2 : Le modèle sous jacent pour une observation est donné par  $Y_{ij} = m_0 + m_1 X_{ij} + \epsilon_{ij}$ . Ecrire le modèle pour l'ensemble des données la forme vectorielle suivante

$$Y = X\beta + \epsilon,$$

où  $Y$  contient les observations  $Y_{ik}$  (notes d'algèbre). Spécifier la matrice  $X$  et les erreurs  $\epsilon$ . Donner l'interprétation de  $m_0$  et  $m_1 + m_0$ .

Question 3 : Mettre la variable dépendante dans une colonne d'un fichier Excel et codez la variable  $X_{ij}$  dans une deuxième colonne d'un fichier

Excel. Importer les données et programmer les lignes suivantes.

```
proc print data=WORK.anova; run ;
proc plot data=WORK.anova;
plot notes*option='*';
run ;
proc anova data=WORK.anova ;
class option ;
model notes=option ;
run ; quit ;
```

## Partie II

Question 1 : Ecrire un programme SAS IML vous permettant de tester l'égalité entre la moyenne des notes d'algèbre des étudiants ayant suivi l'option math et la moyenne des étudiants ayant suivi une option autre que math.

Question 2 : Premier moyen de vérifier son programme : A partir du cours comment peut-on vérifier les calculs des variances factorielle et résiduelle ?

Question 3 : Deuxième moyen de vérifier son programme : Simuler deux jeux de données vous permettant de vérifier que votre programme vous donne des résultats "logiques". Les lignes ci-dessous permettent de simuler des observations de deux variables centrées, normalement distribuées.

```
data sim;
do i=0 to 100;
X1=normal(1);
X2=normal(3);
output;
end; run;
```

Question 4 : Troisième moyen de vérifier son programme : A partir du cours, quels résultats doit-on attendre pour la statistique de test si l'on multiplie les données par 10 ?

Question 5 : Que conseillez vous au directeur d'UFR ?