

Traitement des données en Psychologie

UV PSY38X2

Présentation du cours 2003/2004

Organisation matérielle

Cours magistral :

mardi 16h-17h - Salle A216

Travaux dirigés :

1 groupe de TD de statistiques.

Gr. 1 : mardi 17h-18h - Salle A216

2 sous-groupes de TD en Informatique.

salle A204 ou A206, 2 h. par quinzaine

sous-gr. A - mercredi 18h-20h - Sem. A

sous-gr. B - mercredi 18h-20h - Sem. B

Monitorat informatique

Contrôle des connaissances : (contrôle continu)

70 % Examen écrit (3 heures)

30 % Evaluation de TD

Bibliographie

- G. Mialaret. Statistiques appliquées aux sciences humaines. PUF
- B. Beaufile. Statistiques appliquées à la psychologie. Tomes 1 et 2. LEXIFAC Bréal
- N. Guéguen. Manuel de statistiques pour psychologues
- D.C. Howell. Méthodes statistiques en sciences humaines
- D. Laveault, J. Grégoire. Introduction aux théories des tests en sciences humaines. De Boeck Université
- J.P. Rossi. La méthode expérimentale en Psychologie
- H. Abdi. Introduction au traitement statistique des données expérimentales. PUG
- P. Rateau. Méthode et statistique expérimentales en sciences humaines. Ellipses

Contenu

Documents fournis :

Copie des transparents en statistiques

Fiches de TD

Sites internet permettant de télécharger ces documents :

Depuis les salles de TD d'informatique :

<http://letsamba.univ-brest.fr/~carpentier/>

Depuis le reste de l'Université (domaine univ-brest.fr) :

<http://infolettres.univ-brest.fr/~carpentier/>

Depuis l'extérieur, ou depuis le domaine univ-brest.fr :

<http://geai.univ-brest.fr/~carpentier/>

Voir aussi :

<http://cours.univ-brest.fr/Discipline/informatique/carpentier/>

N.B. Documents (autres que les fiches de TD d'informatique) au format .pdf lisible par Acrobat Reader

Programmes, contrôle des connaissances, documents plus anciens :

<http://geai.univ-brest.fr/enseignements.html>

F de Fisher. Analyse de variance à un facteur.

Introduction aux plans d'expériences.

Analyse de variance à plusieurs facteurs.

Corrélation linéaire. Régression linéaire.

Introduction à l'analyse des données multidimensionnelles :

Analyse en composantes principales.

Analyse en facteurs principaux.

Analyse factorielle des correspondances.

Analyse des correspondances multiples.

Classification ascendante hiérarchique.

Comparaison de deux variances F de Fisher

Exemple. Deux tests mesurant la même aptitude

– Groupe 1 : 25 sujets. $\bar{x}_1 = 40$; $s_{1c}^2 = 65$

– Groupe 2 : 30 sujets. $\bar{x}_2 = 38$; $s_{2c}^2 = 30$

La précision est-elle la même pour les deux tests ?

Cas général

Deux échantillons de tailles n_1 et n_2 extraits de deux populations. Moyennes égales ou différentes. Distribution normale de la variable dans les populations parentes.

Problème : Les *variances* dans les populations parentes sont-elles égales ?

H_0 : Les variances sont égales

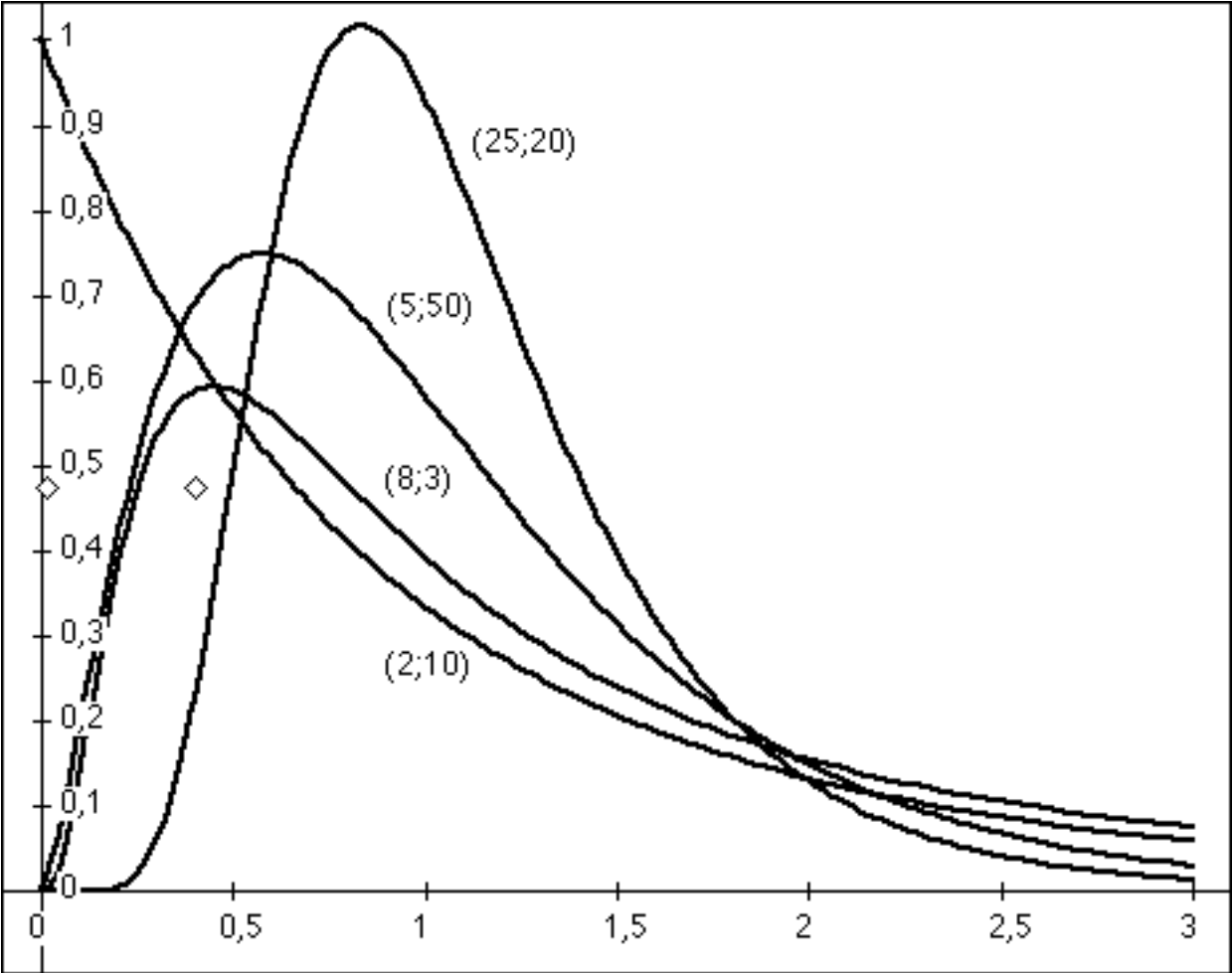
H_1 : La première variance est supérieure à la deuxième.

Statistique de test

$$F = \frac{s_{1,c}^2}{s_{2,c}^2}$$

F suit une **loi de Fisher** à $n_1 - 1$ et $n_2 - 1$ degrés de liberté.

Distributions du F de Fisher



Analyse de Variance à un facteur

Exemple introductif : Test commun à trois groupes d'élèves. Moyennes observées dans les trois groupes : $\bar{x}_1 = 8$, $\bar{x}_2 = 10$, $\bar{x}_3 = 12$.

Question : s'agit-il d'élèves "tirés au hasard" ou de groupes de niveau ?

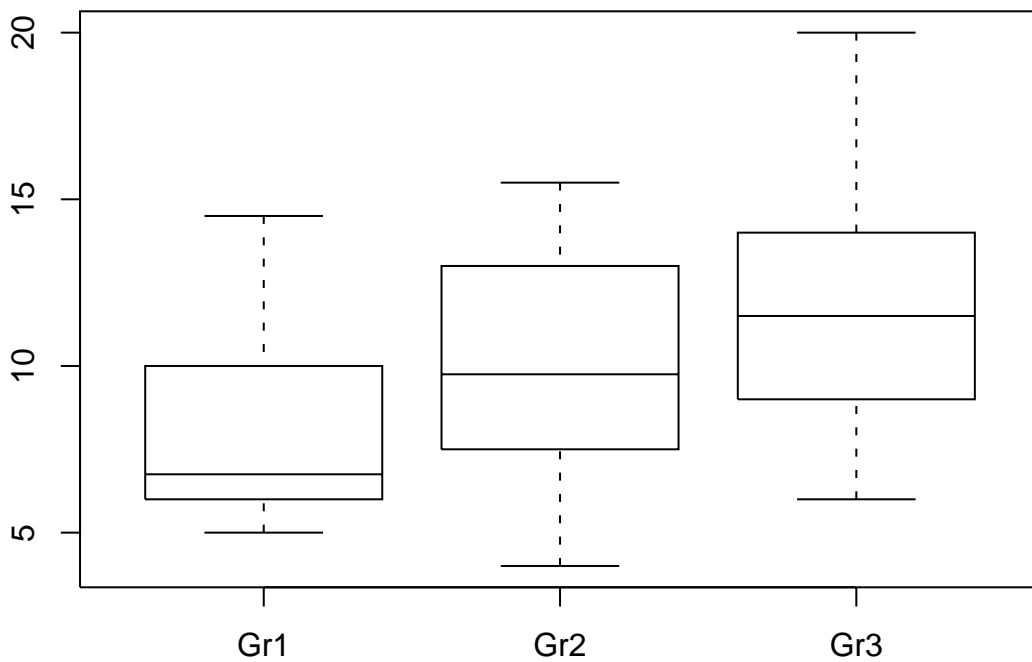
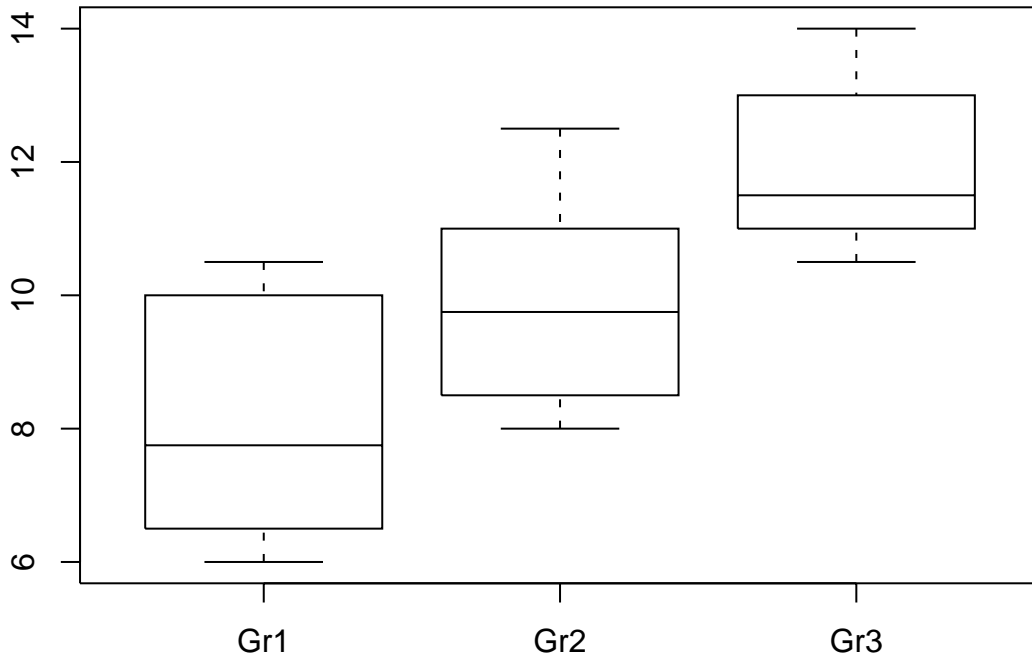
Première situation :

	Gr1	Gr2	Gr3
	6	8	10.5
	6.5	8.5	10.5
	6.5	8.5	11
	7	9	11
	7.5	9.5	11
	8	10	12
	8	11	13
	10	11	13
	10	12	14
	10.5	12.5	14
\bar{x}_i	8	10	12

Deuxième situation :

	Gr1	Gr2	Gr3
	5	4	6
	5.5	5.5	7
	6	7.5	9
	6	9	10
	6.5	9.5	11
	7	10	12
	7.5	11	13
	10	13	14
	12	15	18
	14.5	15.5	20
\bar{x}_i	8	10	12

Boîtes à moustaches pour les deux situations proposées



Démarche utilisée : nous comparons la dispersion des moyennes (8, 10, 12) à la dispersion à l'intérieur de chaque groupe.

Comparer a moyennes sur des groupes indépendants

Plan d'expérience : $\mathcal{S} < \mathcal{A}_a >$

Une variable \mathcal{A} , de modalités A_1, A_2, \dots, A_a définit a groupes indépendants.

Variable dépendante X mesurée sur chaque sujet.

x_{ij} : valeur observée sur le i -ème sujet du groupe j .

Problème : La variable X a-t-elle la même moyenne dans chacune des sous-populations dont les groupes sont issus ?

H_0 : $\mu_1 = \mu_2 = \dots = \mu_a$

H_1 : Les moyennes ne sont pas toutes égales.

Construction de la statistique de test :

Notations :

n_1, n_2, \dots, n_a : effectifs des groupes.

N : effectif total

$T_{.1}, \dots, T_{.a}$: sommes des observations pour chacun des groupes.

$T_{..}$ ou T_G : somme de toutes les observations.

Somme des carrés totale ou variation totale :

$$SC_T = \sum_{i,j} x_{ij}^2 - \frac{T_G^2}{N}$$

Elle se décompose en une variation “intra-groupes” et une variation “inter-groupes” :

$$SC_T = SC_{inter} + SC_{intra} \text{ avec :}$$

$$SC_{inter} = \sum_{j=1}^a \frac{T_{\cdot j}^2}{n_j} - \frac{T_G^2}{N}$$

$$SC_{intra} = \sum_{i,j} x_{ij}^2 - \sum_{j=1}^a \frac{T_{\cdot j}^2}{n_j}$$

Carrés moyens :

$$CM_{inter} = \frac{SC_{inter}}{a - 1} ; \quad CM_{intra} = \frac{SC_{intra}}{N - a}$$

$$\text{Statistique de test : } F = \frac{CM_{inter}}{CM_{intra}}$$

F suit une loi de Fisher à $(a - 1)$ et $(N - a)$ ddl.

Présentation des résultats

Source de variation	SC	ddl	CM	F
\mathcal{A} (inter-groupes)	SC_{inter}	$a - 1$	CM_{inter}	F_{obs}
Résiduelle (intra-gr.)	SC_{intra}	$N - a$	CM_{intra}	
Total	SC_T	$N - 1$		

Organisation des calculs

$i \ j$	1	2	3	Total
1	x_{11}			
...	
$T_{\cdot j}$	$T_{\cdot 1}$			T_G
$T_{\cdot j}^2$				
n_j				N
$\frac{T_{\cdot j}^2}{n_j}$				
$\sum x_{ij}^2$				

Exemple :

15 sujets évaluent 3 couvertures de magazine. Sont-elles équivalentes ?

	C1	C2	C3
	14	16	14
	6	14	16
	12	8	14
	10	8	14
	8	14	12
\bar{x}_i	10	12	14

Calculs

$i \ j$	1	2	3	Total
1	$x_{11} = 14$	16	14	
2	$x_{21} = 6$	14	16	
...	
$T_{\cdot j}$	50	60	70	$T_G = 180$
$T_{\cdot j}^2$	2500	3600	4900	
n_j	5	5	5	$N = 15$
$\frac{T_{\cdot j}^2}{n_j}$	500	720	980	2200
$\sum x_{ij}^2$	540	776	988	2304

$$SC_{inter} = \sum_{j=1}^a \frac{T_{\cdot j}^2}{n_j} - \frac{T_G^2}{N} = 2200 - \frac{180^2}{15} = 40$$

$$SC_{intra} = \sum_{i,j} x_{ij}^2 - \sum_{j=1}^a \frac{T_{\cdot j}^2}{n_j} = 2304 - 2200 = 104$$

$$SC_T = \sum_{i,j} x_{ij}^2 - \frac{T_G^2}{N} = 144$$

$$CM_{inter} = \frac{SC_{inter}}{a - 1} = 20 ; CM_{intra} = \frac{SC_{intra}}{N - a} = 8.67$$

$$F_{obs} = \frac{CM_{inter}}{CM_{intra}} = 2.31$$

F suit une loi de Fisher avec $ddl_1 = a - 1 = 2$ et $ddl_2 = N - a = 12$.

Résultats

Source	Somme carrés	<i>ddl</i>	Carré Moyen	<i>F</i>
<i>C</i>	40	2	20	2.31
Résid.	104	12	8.67	
Total	144	14		

Pour $\alpha=5\%$, $F_{crit} = 3.88$: H_0 est acceptée

Remarques

– SC_{inter} : c'est la somme des carrés (totale) que l'on obtiendrait si toutes les observations d'un groupe étaient égales à la moyenne de ce groupe.

CM_{inter} : variance corrigée de cet ensemble de données.

– SC_{intra} : c'est la somme des carrés (totale) que l'on obtiendrait en "décalant" chaque observation de façon à avoir la même moyenne dans chaque groupe.

CM_{intra} : "moyenne pondérée" des trois variances corrigées ainsi obtenues.

– Hypothèses "a priori" :
distribution normale de X dans chacun des groupes
Egalité des variances dans les trois populations.

– Si 2 groupes, équivaut à un T de Student. $F = T^2$

Pour les deux situations proposées en introduction :

Situation 1

Analysis of Variance Table

Response : x1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	80.000	40.000	17.008	1.659e-05 ***
Residuals	27	63.500	2.352		

Situation 2

Analysis of Variance Table

Response : x2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	80.00	40.00	2.7136	0.08436 .
Residuals	27	398.00	14.74		

Vocabulaire des plans d'expérience

Variable dépendante

– On formule une hypothèse : “telle variable a tel effet sur le comportement des sujets”

– On choisit une **variable dépendante**

Définir une variable mesurable (numérique) caractérisant le comportement du sujet.

Qualités d'une bonne variable dépendante : *pertinence, sensibilité.*

Variables indépendantes ou facteurs

– Recherche des **variables indépendantes** ou **facteurs de variation**

Indépendant : indépendant du sujet, manipulé ou contrôlé par l'expérimentateur

Causes susceptibles d'entraîner une variation de la variable dépendante.

Les facteurs sont des variables nominales ou ordinales. Les valeurs prises par un facteur sont ses *modalités* ou *niveaux*.

Les *Facteurs principaux* sont ceux dont on désire étudier l'effet. Ils sont aussi appelés *facteurs d'intérêt*.

Les *Facteurs secondaires* sont les autres causes susceptibles d'influer sur le comportement des sujets. Deux manières de les prendre en compte :

- Contrôle
- Neutralisation.

Facteur systématique ou fixe : l'ensemble des modalités possibles est fini (et petit). Toutes les modalités sont présentes dans l'expérience.

Facteur aléatoire : l'ensemble des modalités est grand (infini). On choisit alors (par tirage au sort) un ensemble de modalités.

Le facteur sujet : généralement, c'est un facteur aléatoire et secondaire. Il est souvent assimilé à une incertitude sur une mesure.

Facteur étiquette : par exemple, le sexe, ou le milieu socio-culturel.

Groupe contrôle

Souvent, il existe un niveau particulier ou “état nul” de la VI. Un groupe de sujets soumis à ce niveau de la VI constitue un *groupe contrôle*.

Permet notamment de contrôler ou de mettre en évidence d'éventuelles *variables parasites*.

Interaction entre facteurs

Exemple : Mémorisation d'une liste de mots.

VD : nombre de mots mémorisés

Facteur \mathcal{A} : a_1 sujets normaux, a_2 déficit mnésique

Facteur \mathcal{B} : b_1 liste de 12 mots, b_2 30 mots

Moyennes observées sur 4 groupes indépendants :

	b_1	b_2
a_1	10	25
a_2	8	12

Graphe d'interaction :

