# Statistiques paramétriques et non paramétriques

## E.C. PSR73B

## Présentation du cours 2013/2014

### Organisation matérielle

Cours et TD de Statistiques : 24 heures.

Horaire : mercredi 9h15 - 11h30

Contrôle des connaissances :
Examen écrit (2 heures)

### Bibliographie


– D.C. Howell. Méthodes statistiques en sciences humaines De Boeck Université
– P. Rateau, Méthode et statistique expérimentales en sciences humaines, Ellipses
– S. Siegel, N. Castellan, Non-parametric Statistics for the Behavioural Sciences, Mac Graw-Hill, 1988
– N. Gauvrit, Stats pour psycho, De Boeck, 2005
– J. Navarro, L'essentiel de la statistique en psychologie, Editions Ellipses, 2012


### Documents fournis

Transparents du cours de statistiques
Fiches de TD de statistiques.

*Adresse Web*

`http://geai.univ-brest.fr/~carpenti/`





### Contenu

*Statistiques :*

Ce cours vise à présenter les concepts et les procédures de l'analyse statistique des données en psychologie, en insistant sur les aspects méthodologiques : quelle stratégie pour analyser les données ? Quelles sont les méthodes disponibles pour tester telle hypothèse ? Il comporte également des compléments aux méthodes de statistiques descriptives et inférentielles vues en licence :
– Statistiques paramétriques : tests de normalité des distributions parentes ; tests d'homogénéité des variances. Loi de Fisher Snedecor ; analyse de variance.
– Statistiques non paramétriques : test de Kolmogorov-Smirnov ; test de Kruskal-Wallis ; test Q de Cochran ; test de Friedman.
– Compléments sur la corrélation et la régression linéaires à 2 ou plusieurs variables. Alpha de Cronbach. Régression linéaire pas à pas.
– Puissance d'un test : effet calibré, taille d'un effet.





## Tester les conditions d'application d'un test paramétrique

### Conditions d'application du test de Student

Le test de Student est un test paramétrique. Comme tous les tests de ce type, son utilisation est soumise à des conditions d'application ou hypothèses a priori sur la distribution des variables dans les populations de référence.

Rappel : L'application du test de Student (égalité des moyennes) sur deux groupes indépendants suppose :

– La normalité des distributions parentes

– L'égalité des variances (homoscédasticité des résidus)

Problèmes :

– Comment étudier si ces conditions sont respectées ?

– Peut-on s'affranchir de ces conditions ?

### Tests de normalité d'une distribution

Variable numérique $X$ définie sur une population
$(x_i)$ : valeurs observées sur un échantillon de taille $n$
Au vu de cet échantillon : est-il légitime de supposer que la distribution de $X$ dans la population est une loi normale ?

Différents tests proposés : Test de Kolmogorov-Smirnov, test de Lilliefors, test de Shapiro-Wilk.





### Tests de Kolmogorov-Smirnov et de Lilliefors

Echantillon : 8, 9, 9, 10, 10, 10, 11, 13, 14, 14

$H_0$ : $X$ est distribuée selon une loi normale dans la population
$H_1$ : $X$ n'est pas distribuée selon une loi normale.

Construction de la statistique de test :

Moyenne observée : $\overline{x} = 10.8$
Ecart type corrigé : $s_c = 2.15$

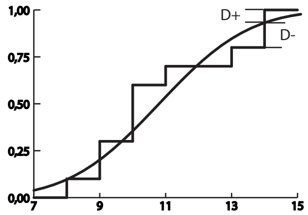
Valeurs centrées réduites : $z_i = \dfrac{x_i - \overline{x}}{s_c}$

| $x_i$ | 8 | 9 | 10 | 11 | 13 | 14 |
|---|---|---|---|---|---|---|
| $z_i$ | $-1.30$ | $-0.84$ | $-0.37$ | 0.09 | 1.02 | 1.49 |

Détermination de la distribution cumulative théorique et calcul des écarts entre distributions cumulatives observée et théorique

| $z_i$ | $F(X < x_i)$ | $F(X \leq x_i)$ | $F(z_i)$ | Ecart - | Ecart + |
|---|---|---|---|---|---|
| $-1.3024$ | 0.0 | 0.1 | 0.0964 | 0.0964 | 0.0036 |
| $-0.8372$ | 0.1 | 0.3 | 0.2012 | 0.1012 | 0.0988 |
| $-0.3721$ | 0.3 | 0.6 | 0.3549 | 0.0549 | 0.2451 |
| 0.0930 | 0.6 | 0.7 | 0.5371 | 0.0629 | 0.1629 |
| 1.0233 | 0.7 | 0.8 | 0.8469 | 0.1469 | 0.0469 |
| 1.4884 | 0.8 | 1.0 | 0.9317 | 0.1317 | 0.0683 |

Comparaison des deux courbes cumulatives :




Maximum des écarts absolus : $D_{obs} = 0.2451$.
Taille de l'échantillon : $n = 10$.

$D_{obs}$ est la valeur observée de la statistique de test pour le test de Kolmogorov-Smirnov et celui de Lilliefors.

Ces deux tests utilisent la même statistique de test, mais des lois statistiques et donc des tables de valeurs critiques différentes.

– Kolmogorov-Smirnov s'utilise lorsque la moyenne et l'écart type de la loi théorique sont connus a priori et sont donc fixés indépendamment de l'échantillon.

– Lilliefors s'utilise lorsque la moyenne et l'écart type de la loi normale théorique sont estimés à partir de l'échantillon.





Dans l'exemple ci-dessus, c'est le test de Lilliefors qui s'applique.
$L_{obs} = 0.2451$
Consultation de la table : pour $\alpha = 5\%$, $L_{crit} = 0.258$.
Conclusion : hypothèse de normalité retenue.

Remarque : ce test est fréquemment utilisé dans les publications de psychologie.

Si les paramètres de la loi théorique avaient été connus a priori, on aurait utilisé le test de Kolmogorov-Smirnov avec les résultats suivants :
Consultation de la table : pour $\alpha = 5\%$, $D_{crit} = 0.41$
Conclusion : hypothèse de normalité retenue.

**Test de Shapiro-Wilk**

Les statisticiens ont proposé un autre test, nettement plus puissant que les deux tests précédents : le test de Shapiro-Wilk.

Le calcul de la valeur observée de la statistique de test et de son niveau de significativité est très fastidieux : on utilisera donc un logiciel de statistiques pour le mettre en oeuvre.

Ainsi, sur l'exemple précédent, Statistica nous indique :

$W = 0.8849$, $p = 0.1485$

et la conclusion demeure identique.





**Sensibilité des tests précédents aux ex aequo.**
**Test de D'Agostino-Pearson**

On observe une variable numérique continue sur un échantillon de sujets. Cependant (cas fréquent en Psychologie), nous ne pouvons pas obtenir des valeurs précises de cette variable. Seules quelques modalités (par exemple 5 à 7 modalités) sont présentes dans les résultats. Les valeurs observées sont donc des arrondis des valeurs réelles et comportent de nombreux ex aequo. Par exemple, nous avons évalué un paramètre continu à partir de réponses sur une échelle de Likert.

La présence d'ex aequo et de "trous" entre les valeurs observées a un effet immédiat sur les tests précédents : il y a de fortes chances que les tests de Lilliefors et de Shapiro-Wilk ne permettent pas d'inférer la normalité de la distribution parente.

Aussi, certains auteurs ont mis au point d'autres tests de normalité, fondés sur l'étude de l'*asymétrie (skewness)* et de l'*aplatissement (kurtosis)* de la distribution.

**Asymétrie**

Pour une série statistique $(x_i)$ définie sur une population finie d'effectif $n$, l'asymétrie $\gamma_1$ est définie par :

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

où $\mu_3 = \dfrac{\sum(x_i - \mu)^3}{n}$ et où $\mu$ et $\sigma$ désignent la moyenne et l'écart type.





Les statisticiens utilisent généralement une formule plus compliquée, qui donne l'estimation $g_1$ de l'asymétrie sur une population à partir d'un échantillon (asymétrie corrigée).

Pour une distribution symétrique et notamment pour une distribution normale, $\gamma_1 = 0$.

**Aplatissement**

Pour une série statistique $(x_i)$ définie sur une population finie d'effectif $n$, l'aplatissement $\gamma_2$ est défini par :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

où $\mu_4 = \dfrac{\sum(x_i - \mu)^4}{n}$ et où $\mu$ et $\sigma$ désignent la moyenne et l'écart type.

De même, une formule dérivée donne l'estimation $g_2$ de l'aplatissement sur une population à partir d'un échantillon (aplatissement corrigé).

Pour une distribution normale, $\gamma_2 = 0$.

**Test de d'Agostino-Pearson**

Deux tests, dont les statistiques $Z_1$ et $Z_2$ suivent approximativement des lois normales centrées réduites, permettent de tester la nullité des paramètres $\gamma_1$ et $\gamma_2$.

Le test de d'Agostino-Pearson pose comme hypothèses :

$H_0$ : Normalité de la distribution parente

$H_1$ : Défaut de normalité de la distribution parente.

Il utilise comme statistique de test : $\chi^2 = Z_1^2 + Z_2^2$, qui suit approximativement une loi du $\chi^2$ à 2 ddl.

*Exemple*

$$(X_i) = (2,4,1,2,3,4,3,2,3,3,3,5,1,4,3,2,\\ 4,2,3,0,4,4,3,3,4,3,0,4,3,1,4,3,\\ 3,3,1,4,3,3,4,3,3,3,3,1,5,2,3,5)$$

Les tests de Lilliefors et de Shapiro-Wilk concluent sur la non-normalité :

Pour Lilliefors : $D = 0.2641$, $p = 4.872 \times 10^{-9}$

Pour Shapiro-Wilk : $W = 0.9018$, $p = 0.0007196$.

En revanche, les tests sur l'asymétrie et l'aplatissement et le test de D'Agostino donnent :

```
Test Results :
  STATISTIC :
    Chi2 | Omnibus : 3.2229
    Z3 | Skewness : -1.7038
    Z4 | Kurtosis : 0.5658
  P VALUE :
    Omnibus Test : 0.1996
    Skewness Test : 0.08843
    Kurtosis Test : 0.5715
```





**Tests d'homogénéité des variances**

Deuxième condition d'application d'un test t de Student : égalité des variances.

Pour vérifier cette condition : tests de Fisher, de Levene, de Brown et Forsythe et le test de Bartlett (que nous n'étudierons pas).

**Test de Fisher et loi de Fisher Snedecor**

**Exemple.** Etude sur la boulimie. Deux groupes de sujets : boulimie simple ou "avec vomissements". Variable dépendante : écart relatif par rapport au poids normal.

| | Simple | Avec vom. |
|---|---|---|
| $\overline{x}_i$ | 4.61 | $-0.83$ |
| $s_{ic}^2$ | 219.04 | 79.21 |
| $n_i$ | 49 | 32 |

*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.



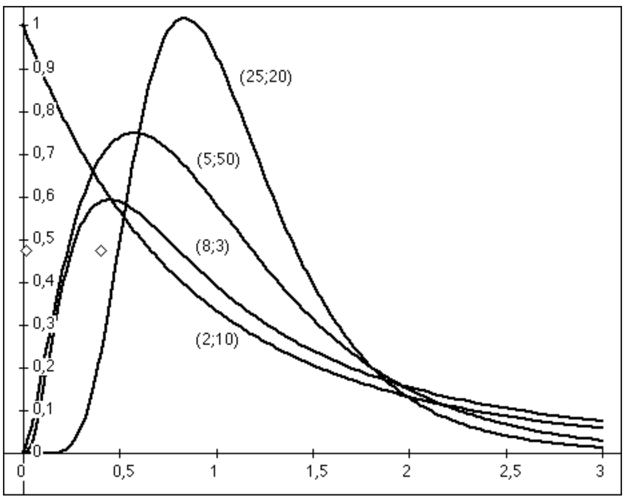


*Statistique de test*

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

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

**Distributions du F de Fisher**




Sur l'exemple considéré : $F_{obs} = \frac{219.04}{79.21} = 2.76$
Pour $\alpha = 5\%$, $ddl_1 = 48$ et $ddl_2 = 31$, $F_{crit} = 1.79$.

On rejette donc l'hypothèse $H_0$.

Inconvénient du test de Fisher : très sensible à un défaut de normalité.





**Test de Levene**

On dispose des valeurs observées $x_{ij}$ d'une variable dépendante $X$ dans 2 *ou plusieurs* groupes.

Au vu de ces valeurs, peut-on admettre l'hypothèse d'égalité des variances dans les différents groupes ($H_0$), ou doit-on rejeter cette hypothèse, et accepter $H_1$ ?

Principe du test :

Dans chaque groupe, on forme la série des *écarts absolus à la moyenne du groupe :* $|x_{ij} - \overline{x}_j|$.

On réalise ensuite un test (bilatéral) de comparaison de moyennes sur ces séries ($t$ de Student ou analyse de variance à un facteur).

**Test de Brown et Forsythe**

Il s'agit d'une version modifiée du test de Levene, dans laquelle on considère les écarts absolus aux médianes, au lieu des moyennes.

Ce test est plus robuste (moins sensible à un défaut de normalité) que celui de Levene.

### Test de Student dans le cas de variances hétérogènes

Pour le test t de Student, il existe des formules (approximatives) à utiliser lorsqu'on ne fait pas l'hypothèse d'égalité des variances.

– La statistique de test est alors :

$$t' = \frac{\overline{x_1} - \overline{x_2}}{E} \quad \text{avec} \quad E^2 = \frac{s_{1c}^2}{n_1} + \frac{s_{2c}^2}{n_2}$$

Cette statistique est identique à celle vue l'an dernier lorsque $n_1 = n_2$.

– Le nombre de degrés de liberté à prendre en compte dépend des variances observées, mais est en général strictement inférieur à $n_1 + n_2 - 2$. On retrouve $dll = n_1 + n_2 - 2$ lorsque $n_1 = n_2$ et $s_{1c} = s_{2c}$.





## Analyse de Variance à un facteur

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

**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 |
| $\overline{x}_i$ | 8 | 10 | 12 |

*Deuxième situation :*

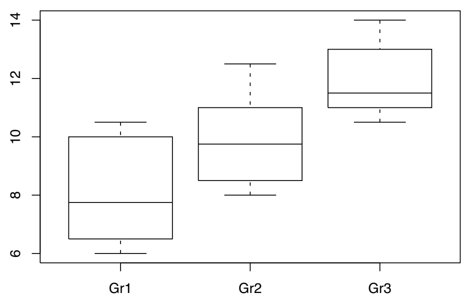
| | Gr1 | Gr2 | Gr3 |
|---|---|---|---|
| | 5 | 4 | 6 |
| | 5.5 | 5.5 | 7 |
| | 6 | 7.5 | 9 |
| | 6 | 9 | 10 |
| | 6.5 | 9.5 | 11 |
| | 7 | 10 | 12 |
| | 7.5 | 11 | 13 |
| | 10 | 13 | 14 |
| | 12 | 15 | 18 |
| | 14.5 | 15.5 | 20 |
| $\overline{x}_i$ | 8 | 10 | 12 |

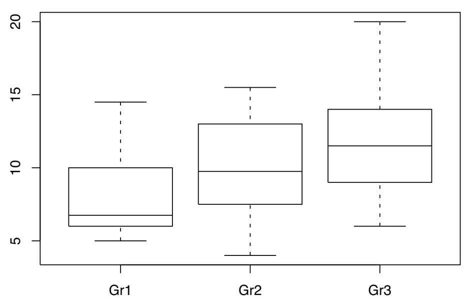




**Démarche utilisée :** nous comparons la dispersion des moyennes (8, 10, 12) à la dispersion à l'intérieur de chaque groupe.

Boîtes à moustaches pour les deux situations proposées








### Comparer $a$ moyennes sur des groupes indépendants

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

Une variable $\mathcal{A}$, de modalités $A_1, A_2, \ldots, A_a$ 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 ?

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

*Hypothèses du test :*

$H_0$ : $\mu_1 = \mu_2 = \ldots = \mu_a$
$H_1$ : Les moyennes ne sont pas toutes égales.

**Exemple :**

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

| | C1 | C2 | C3 | |
|---|---|---|---|---|
| | 13 | 17 | 14 | |
| | 5 | 15 | 16 | |
| | 11 | 9 | 14 | |
| | 9 | 9 | 14 | |
| | 7 | 15 | 12 | |
| $\overline{x}_i$ | 9 | 13 | 14 | 12 |

*Variation (ou somme des carrés) totale :*

$SC_T = (13-12)^2 + (5-12)^2 + \ldots + (12-12)^2 = 174$

*Décomposition de la variation totale :*

Score d'un sujet = Moyenne de son groupe + Ecart

| C1 | C2 | C3 |
|---|---|---|
| 9 | 13 | 14 |
| 9 | 13 | 14 |
| 9 | 13 | 14 |
| 9 | 13 | 14 |
| 9 | 13 | 14 |

| C1 | C2 | C3 |
|---|---|---|
| 4 | 4 | 0 |
| -4 | 2 | 2 |
| 2 | -4 | 0 |
| 0 | -4 | 0 |
| -2 | 2 | -2 |

*Variation (ou somme des carrés) inter-groupes :*

$SC_{inter} = (9-12)^2 + (9-12)^2 + \ldots + (14-12)^2 = 70$

*Variation (ou somme des carrés) intra-groupes :*

$$SC_{intra} = 4^2 + (-4)^2 + \ldots + (-2)^2 = 104$$





*Calcul des carrés moyens :*

$$CM_{inter} = \frac{SC_{inter}}{a-1} = 35 \; ; \; CM_{intra} = \frac{SC_{intra}}{N-a} = 8.67$$

*Statistique de test :*

$$F_{obs} = \frac{CM_{inter}}{CM_{intra}} = 4.04$$

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

**Résultats**

| Source | Somme carrés | $ddl$ | Carré Moyen | $F$ |
|---|---|---|---|---|
| $\mathcal{C}$ | 70 | 2 | 35 | 4.04 |
| Résid. | 104 | 12 | 8.67 | |
| Total | 174 | 14 | | |

Pour $\alpha$=5%, $F_{crit} = 3.88$ : $H_1$ est acceptée





**Formules de calcul pour un calcul "à la main" efficace**

*Construction de la statistique de test :*

*Notations :*
$n_1, n_2, \ldots, n_a$ : effectifs des groupes.
$N$ : effectif total
$T_{\cdot 1}, \ldots, T_{\cdot a}$ : sommes des observations pour chacun des groupes.
$T_{\cdot\cdot}$ 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" :

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

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

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





**Formules de calcul (sans recherche d'efficacité)**

– $SC_T = N\times$ (Variance non corrigée de l'ensemble des observations)

– $SC_{inter} = N\times$ (Variance non corrigée du tableau obtenu en remplaçant chaque observation par la moyenne de son groupe)

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.

– $SC_{intra} = N\times$ (Variance non corrigée du tableau des écarts à la moyenne de chaque groupe)

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.

*Carrés moyens :*

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

*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 | $SC_T$ | $N-1$ | | |

**Remarque**

S'il n'y a que 2 groupes, l'ANOVA équivaut à un $T$ de Student (bilatéral). $F = T^2$





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 | | |





## Après l'analyse de variance : tests post hoc

La statistique $F$ calculée par l'ANOVA est souvent appelée $F$ *omnibus*. Si $H_1$ est retenue, il existe au moins une différence significative entre deux groupes, mais l'ANOVA ne renseigne pas sur les paires de groupes pour lesquelles la différence est significative.

Utiliser le $t$ de Student pour comparer les groupes deux à deux ? Peu correct du point de vue statistique.

Plusieurs tests de comparaison *post hoc* ont été proposés.

Donnons quelques indications sur :
- le test LSD de Fisher
- le test de Bonferroni-Dunn
- le test HSD de Tukey
- le test de Dunnett

**Le test LSD de Fisher**

LSD : least significant difference

Pour comparer le groupe $a$ et le groupe $b$, on calcule une statistique $t$ de Student :

$$t = \frac{\overline{x}_a - \overline{x}_b}{E}$$

où $\overline{x}_a$ et $\overline{x}_b$ sont les moyennes observées dans les deux groupes et dans laquelle l'erreur type $E$ est calculée à partir du carré moyen intra-groupe de l'ANOVA par :

$$E^2 = \text{CM}_{intra}\left(\frac{1}{n_a} + \frac{1}{n_b}\right)$$

.





Sous $H_0$, $t$ suit une loi de Student dont le nombre de ddl est celui de CM$_{intra}$, c'est-à-dire $N-k$ (nombre total d'observations moins nombre de groupes).

Le seuil et/ou le niveau de significativité sont évalués de la même façon que dans le cas d'un test de Student.

Ce test est peu conservateur (risque important de commettre une erreur de type I, c'est-à-dire de conclure sur des différences qui ne sont pas réelles).

*Exemple*

On considère les données suivantes :

| | Gr1 | Gr2 | Gr3 |
|---|---|---|---|
| | 8 | 7 | 4 |
| | 10 | 8 | 8 |
| | 9 | 5 | 7 |
| | 10 | 8 | 5 |
| | 9 | 5 | 7 |
| | | 6 | 6 |
| | | | 4 |
| $n_i$ | 5 | 6 | 7 |
| $\overline{x}_i$ | 9.2 | 6.5 | 5.86 |
| $s_c^2$ | 0.70 | 1.90 | 2.48 |

Le tableau d'ANOVA est le suivant :

| Source | Somme carrés | *ddl* | Carré Moyen | $F$ | p |
|---|---|---|---|---|---|
| Groupe | 34.84 | 2 | 17.42 | 9.92 | .002 |
| Résid. | 27.16 | 15 | 1.81 | | |
| Total | 62.00 | 17 | | | |

Effectuons, par exemple, un test LSD de Fisher pour comparer les moyennes des groupes 1 et 2.

$$\overline{x}_1 - \overline{x}_2 = 2.7$$
$$E^2 = 1.8105\left(\frac{1}{5}+\frac{1}{6}\right) = 0.6639$$
$$t_{obs} = \frac{2.7}{\sqrt{0.6639}} = 3.31$$

On a ici $ddl = 15$. D'où un niveau de significativité $p = 0.004723$. Les deux groupes semblent significativement différents aussi bien au seuil de 5% qu'au seuil de 1%.

**Le test de Bonferroni-Dunn**

Soit $k$ le nombre de groupes. On calcule les statistiques $t$ comme dans le cas du test LSD de Fisher, mais, pour énoncer un résultat au seuil $\alpha$ (seuil "familywise"), on fait les tests individuels au seuil "par comparaison" $\alpha_{PC} = \frac{\alpha}{c}$ où $c$ est le nombre de comparaisons à effectuer. Dans le cas général $c = \frac{k(k-1)}{2}$.

Le test de Bonferroni-Dunn est plus conservatif (moins puissant) que le test LSD de Fisher. Autrement dit, ce test fait courir moins de risques de commettre une erreur de type I (conclure sur des différences non réelles), mais plus de risques de commette une erreur de type II (ne pas voir une différence qui existe).





*Exemple*

On reprend l'exemple précédent.
On a ici $k = 3$ et donc $c = 3$. Pour obtenir un résultat au seuil global $\alpha_{FW} = 1\%$, chaque test de comparaison est fait au seuil $\alpha_{PC} = 0.0033$. On ne trouve pas dans les tables "papier" la valeur critique du $t$ de Student à 15 ddl pour ce seuil, mais on peut l'obtenir à partir de Statistica ou du site

`geai.univ-brest.fr/~carpenti/statistiques/table1.php`

On obtient ainsi : $t_{crit} = 3.49$, et la différence entre les groupes 1 et 2 n'est pas significative selon ce test.

**Le test HSD de Tukey**

HSD : honestly significant difference

Le test HSD de Tukey représente un moyen terme entre les deux tests précédents. Le test proposé au départ par Tukey s'applique à des groupes équilibrés mais une variante pour des groupes non équilibrés a également été développée.

Dans le cas de groupes équilibrés, la statistique calculée $Q$ est égale à la statistique $t$ précédente multipliée par $\sqrt{2}$. Mais, pour tester la significativité du résultat, on utilise la *loi de l'étendue studentisée*, qui prend comme paramètres le nombre de groupes ($k$) et le nombre de degrés de liberté de $CM_{intra}$ (en général $N - k$).





*Exemple*

On reprend l'exemple précédent, sans tenir compte du déséquilibre entre les groupes. Pour la comparaison entre les groupes 1 et 2, on a :

$Q_{obs} = 3.31\sqrt{2} = 4.68$.

Or, pour 3 groupes, 15 ddl et un seuil de 1 %, on obtient comme valeur critique : $Q_{crit} = 4.83$. Ici encore, la différence entre les groupes 1 et 2 n'est pas significative au seuil de 1 %.

**Synthèse sur ces trois tests**

Après une ANOVA qui a conclu sur $H_1$ :
– Le test LSD de Fisher permet de repérer les différences vraisemblablement non significatives ;
– Le test de Bonferroni Dunn permet de repérer les différences vraisemblablement significatives ;
– Le test HSD de Tukey permet d'obtenir un résultat concernant les cas ambigus.





**Comparaison de groupes expérimentaux à un groupe témoin : test de Dunnett**

Lorsqu'il s'agit de comparer un ou plusieurs groupes expérimentaux à un groupe témoin, le nombre de comparaisons n'est plus $\frac{k(k-1)}{2}$ mais $(k-1)$.

Dans ce cas, on peut utiliser le test de Dunnett. La statistique de test est la même que celle du test LSD de Fisher, mais Dunnett a développé des tables spécifiques qui tiennent compte :
– du nombre de ddl (celui de $CM_{intra}$)
– du nombre de groupes
– du type de test (unilatéral, bilatéral) car le sens de la différence fait souvent partie de l'hypothèse de recherche dans ce type de situation.

## Analyse de Variance à plusieurs facteurs

Autres situations couramment rencontrées :

– Plan à mesures répétées : $\mathcal{S}_n * \mathcal{A}$
$n$ sujets observés dans plusieurs conditions.
Résultat attendu : effet du facteur $\mathcal{A}$.

– Plan factoriel à 2 facteurs : plan $\mathcal{S} < \mathcal{A} * \mathcal{B} >$
Etude simultanée de 2 facteurs ; groupes indépendants de sujets pour chaque combinaison de niveaux des facteurs.
Résultats attendus : effets de $\mathcal{A}$, de $\mathcal{B}$, de l'interaction $\mathcal{AB}$.

– Plan à mesures partiellement répétées :
plan $\mathcal{S} < \mathcal{A} > *\mathcal{B}$
Groupes indépendants de sujets correspondant à chaque niveau du facteur $\mathcal{A}$ ; chaque sujet est observé dans plusieurs conditions (niveaux de $\mathcal{B}$).
Résultats attendus : effets de $\mathcal{A}$, de $\mathcal{B}$, de l'interaction $\mathcal{AB}$.

Ces situations seront développées en TD avec Statistica.





## Tests non paramétriques

Paramètres : moyenne, variance, covariance, etc ;

Les tests tels que t de Student, ANOVA, etc sont des tests paramétriques :

– hypothèses relatives à un paramètre des populations parentes
– nécessité d'estimer un ou plusieurs paramètres de la distribution parente à l'aide des échantillons observés
– conditions d'application liées à la forme des distributions parentes

Il existe également des tests *non paramétriques* ou indépendants de toute distribution.

– pas de condition d'application
– peu affectés par la présence d'un ou plusieurs scores extrêmes
– ils ont en général une plus faible puissance que les tests paramétriques correspondants





## Tests non paramétriques sur deux groupes indépendants

Situation envisagée : un plan $\mathcal{S} < \mathcal{A}_2 >$ avec un facteur $\mathcal{A}$ à 2 niveaux définissant deux groupes indépendants et une variable dépendante $X$ ordinale ou numérique

Effectifs des deux groupes : $n_1$ et $n_2$.

### Test de la médiane

*Hypothèses*

$H_0$ : Les deux populations parentes ont même médiane.
$H_1$ : Les deux populations parentes ont des médianes différentes

*Construction de la statistique de test*

On détermine la médiane $M$ de la série obtenue en réunissant les deux échantillons.

On constitue un tableau de contingence en croisant la variable indépendante et la variable dérivée "position par rapport à $M$"

| | Gr 1 | Gr 2 | Ensemble |
|---|---|---|---|
| $\leq M$ | $N_{1\,1}$ | $N_{1\,2}$ | $N_{1\cdot}$ |
| $> M$ | $N_{2\,1}$ | $N_{2\,2}$ | $N_{2\cdot}$ |
| Total | $N_{\cdot 1}$ | $N_{\cdot 2}$ | $N_{\cdot\cdot}$ |

On fait un test du $\chi^2$ sur le tableau obtenu.

Condition d'application (selon Siegel et Castellan) : le nombre total d'observations doit être supérieur à 20.





### Test bilatéral de Kolmogorov-Smirnov

$H_0$ : La distribution de la VD est la même dans les deux populations parentes
$H_1$ : Les distributions sont différentes

*Construction de la statistique de test*

Soient $n_1$ et $n_2$ les tailles des deux échantillons.

On choisit un regroupement en classes : $b_1, b_2, \ldots, b_k$

On construit le tableau des fréquences cumulées dans les deux groupes :

| | Gr 1 | Gr 2 |
|---|---|---|
| $X \leq b_1$ | $F_{11}$ | $F_{12}$ |
| $X \leq b_2$ | $F_{21}$ | $F_{22}$ |
| ... | ... | ... |
| $X \leq b_k$ | $F_{k1}$ | $F_{k2}$ |

On calcule :

$$D = \max |F_{i1} - F_{i2}|$$

Pour $n_1 \leq 25$ et $n_2 \leq 25$, les valeurs critiques de $J = n_1 n_2 D$ sont tabulées.

Pour de grands échantillons ($n_1 > 25$ et $n_2 > 25$), la valeur $D_{crit}$ peut être calculée par la formule :

$$D_{crit} = \lambda \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

où $\lambda$ est une constante dépendant du seuil choisi :

| seuil | 0.10 | 0.05 | 0.01 | 0.001 |
|---|---|---|---|---|
| $\lambda$ | 1.22 | 1.36 | 1.63 | 1.95 |