Statistiques paramétriques et non paramétriques

E.C. PSR73B

Présentation du cours 2010/2011

Organisation matérielle

Cours et TD de Statistiques : 24 heures.

Mardi - 13h45-15h45 - A204

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

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

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

Les tests de Kolmogorov-Smirnov et de Lilliefors 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.

– 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.

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^2_{ic}	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 ?

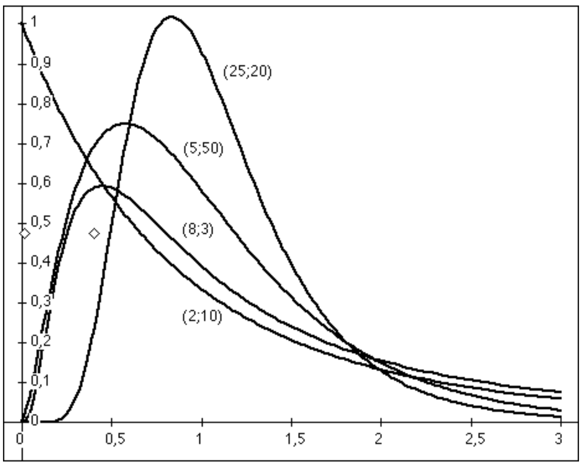
H_0 : Les variances sont égales
H_1 : La première variance est supérieure à la deuxième.

Statistique de test

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

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 :

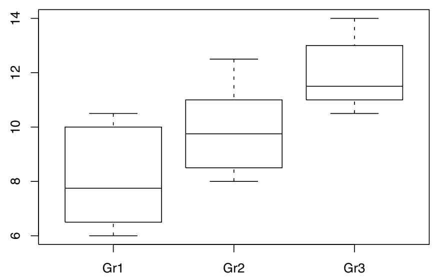
	Gr1	Gr2	Gr3
	6	8	10.5
	6.5	8.5	10.5
	6.5	8.5	11
	7	9	11
	7.5	9.5	11
	8	10	12
	8	11	13
	10	11	13
	10	12	14
	10.5	12.5	14
$\overline{x}_i$	8	10	12

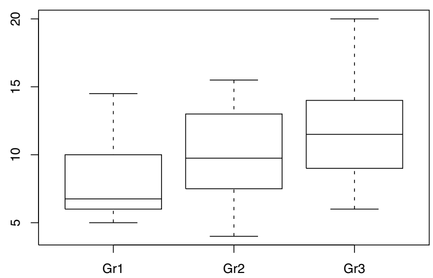
Deuxième situation :

	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 α=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 α (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 χ^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 λ est une constante dépendant du seuil choisi :

seuil	0.10	0.05	0.01	0.001
λ	1.22	1.36	1.63	1.95

Test unilatéral de Kolmogorov-Smirnov

H_0 : La distribution de la VD est la même dans les deux populations parentes
H_1 : L'intensité de la VD est plus forte dans la deuxième population

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 le maximum des différences, ordonnées en fonction du sens du test :

$$D = \max\left[F_{i1} - F_{i2}\right]$$

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 $n_2 > 25$), on utilise l'approximation suivante : sous H_0, la statistique :

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2}$$

suit une loi du χ^2 à 2 ddl.

Exemple

Apprentissage séquentiel par des élèves du 11è grade et des élèves du 7è grade.

Hypothèse : la matière apprise au début de la série est rappelée plus efficacement, mais cet effet est moins prononcé chez les sujets jeunes.

Variable dépendante : pourcentage d'erreurs commises sur la première partie de la série.

On fait ici un test unilatéral. L'hypothèse H_1 est : le pourcentage d'erreurs est significativement plus élevé dans le 2è groupe.

Données :

11è grade	7è grade
35.2	39.1
39.2	41.2
40.9	46.2
38.1	48.4
34.4	48.7
29.1	55.0
41.8	40.6
24.3	52.1
32.4	47.2
	45.2

Découpage en classes et fréquences cumulées :

	11è grade	7è grade	Diff.
$X \leq 28$	0.111	0	0.111
$X \leq 32$	0.222	0	0.222
$X \leq 36$	0.556	0	0.555
$X \leq 40$	0.778	0.1	0.678
$X \leq 44$	1	0.3	0.700
$X \leq 48$	1	0.6	0.400
$X \leq 52$	1	0.8	0.200
$X \leq 56$	1	1	0

On obtient : $D = 0.7$ d'où $J = 9 \times 10 \times 0.7 = 63$.

Au seuil de 1%, la table indique : $J_{crit} = 61$.

On conclut donc sur H_1.

Test des suites de Wald-Wolfowitz

Anderson, Wald et Wolfowitz ont proposé (1943) un test d'auto-corrélation utilisé notamment sur des séries chronologiques. Le test étudié ci-dessous (*test des suites*) en est une version simplifiée.

Certains auteurs distinguent deux tests des suites :

- L'un est un test sur un échantillon, sur lequel sont observées une variable dichotomique et une variable ordinale ou numérique. Il s'agit alors d'étudier si les deux variables sont indépendantes.

- L'autre est un test sur deux échantillons, et permet d'étudier si la VD a les mêmes caractéristiques dans les deux populations parentes. Dans ce cas, en général, seule l'hypothèse H_1 unilatérale à gauche a un sens.

H_0 : L'appartenance d'une observation à l'un ou l'autre groupe est indépendante de son rang.

H_1 bilatérale : L'appartenance d'une observation à l'un ou l'autre groupe dépend de son rang.
H_1 unilatérale à gauche : Etant donné deux observations consécutives, la probabilité qu'elles appartiennent au même groupe est supérieure à celle résultant du hasard.
H_1 unilatérale à droite : L'alternance entre les observations de l'un et l'autre groupe est trop élevée pour être due au hasard.

Méthode : on classe toutes les observations par ordre croissant. On construit un compteur démarrant à 1, et qui augmente d'une unité chaque fois que l'on change de groupe en parcourant la liste ordonnée. On obtient ainsi le nombre de "runs" u.

Exemple : On a fait passer une épreuve à 31 sujets, 14 hommes et 17 femmes. Le protocole des rangs observé est le suivant :

Hommes : 1 2 3 7 8 9 10 13 14 15 23 24 26 27
Femmes : 4 5 6 11 12 16 17 18 19 20 21 22 25 28 29 30 31

Détermination du nombre de "runs" :

```
MMM FFF MMMM FF MMM FFFFFFF MM F MM FFFF
111 222 3333 44 555 6666666 77 8 99 0000
```

Ici : $u = 10$.

Soit n_1 et n_2 les effectifs des deux groupes.

Pour $n_1 \leq 10$ ou $n_2 \leq 10$, on utilise des tables spécialisées.

Pour $n_1 > 10$ et $n_2 > 10$, on utilise l'approximation par une loi normale :

$$\mu = \frac{2n_1n_2}{n_1+n_2} + 1 \quad \sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

$$Z = \frac{u - \mu \pm 0.5}{\sigma}$$

N.B. ± 0.5 est une correction de continuité. Le signe doit être choisi de façon diminuer la valeur de $|Z|$.

Ici : $\mu = 16.35$ $\sigma^2 = 7.35$ $Z = -2.16$

Remarque : Ce test suppose l'absence d'ex aequo.

Test U de Mann-Whitney - Test de Wilcoxon Mann Whitney

Soient θ_1 et θ_2 les médianes de la variable dépendante dans les populations parentes.

H_0 : $\theta_1 = \theta_2$
H_1 : $\theta_1 \neq \theta_2$ (hypothèse bilatérale),
$\theta_1 < \theta_2$ ou $\theta_1 > \theta_2$ (hypothèses unilatérales)

Méthode : On construit le protocole des rangs pour l'ensemble des (n_1+n_2) observations (avec la convention du rang moyen pour les ex aequo).

W_1 : somme des rangs du premier échantillon
W_2 : somme des rangs du deuxième échantillon.

Pour $n_1 \leq 10$ ou $n_2 \leq 10$, on utilise des tables spécialisées.

Si $n_1 > 10$ et $n_2 > 10$, ou si l'un des deux effectifs est supérieur à 12, on utilise l'approximation par une loi normale :

Test U de Mann-Whitney proprement dit (cf. Statistica) :
On calcule :

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - W_1$$

$$U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

$$U = \min(U_1, U_2)$$

$$Z = \frac{U - \frac{n_1n_2}{2}}{E} \quad \text{avec} \quad E^2 = \frac{n_1n_2(n_1+n_2+1)}{12}$$

Variante : On calcule les rangs moyens dans les deux groupes $\overline{R_1}$ et $\overline{R_2}$ puis la statistique :

$$Z = \frac{\overline{R_1} - \overline{R_2}}{E} \text{ avec } E^2 = \frac{(n_1+n_2+1)(n_1+n_2)^2}{12n_1n_2}$$

Remarques.

1. Dans le test de Mann-Whitney, la variable X est supposée continue. La probabilité d'observer des ex aequo est donc *en théorie* nulle. Cependant, certains auteurs ont proposé des formules correctives pour tenir compte des ex aequo. Pour de grands échantillons, la correction proposée revient à multiplier la statistique précédente par $\frac{s}{s'}$ où s est l'écart type de la série $(1, 2, \ldots, n_1+n_2)$ et s' l'écart type des rangs réels. L'effet de la correction est en général peu important.

2. Certains auteurs proposent une formule comportant une correction de continuité.

3. Un test analogue : le test de permutation des rangs de Wilcoxon.

4. Le test de Mann Whitney permet notamment de tester l'égalité des médianes, *à condition que les variances des distributions soient égales*. Un autre test (test des rangs robuste) permet de s'affranchir de cette dernière condition.

Tests non paramétriques sur k groupes indépendants

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

Effectifs des groupes : $n_1, n_2, \ldots, n_k$.

Test de la médiane

Le test de la médiane peut encore être utilisé dans cette situation.

Test de Kruskal-Wallis

Soient $\theta_1, \theta_2, \ldots, \theta_k$ les médianes de la variable dépendante dans les populations parentes.

H_0 : $\theta_1 = \theta_2 = \ldots = \theta_k$
H_1 : Les médianes ne sont pas toutes égales.

Méthode : On construit le protocole des rangs pour l'ensemble des observations.

Soit $\overline{R}_j$ la moyenne des rangs dans le groupe j, N le nombre total d'observations et $\overline{R} = \frac{N+1}{2}$ le rang moyen général.

Statistique de test :

$$K = \frac{12}{N(N+1)} \sum n_j(\overline{R}_j - \overline{R})^2$$

ou

$$K = \left[\frac{12}{N(N+1)} \sum n_j\overline{R}_j^2\right] - 3(N+1)$$

Si le nombre de groupes est supérieur à 3 et le nombre d'observations dans chaque groupe est supérieur à 5, K suit approximativement une loi du χ^2 à $(k-1)ddl$.

Remarques :

1. Si le nombre de groupes est égal à 2, la statistique K est le carré de la statistique Z du test de Wilcoxon Mann et Whitney. Pour de grands échantillons et des tests bilatéraux, les deux tests sont donc équivalents.

2. Comme dans le cas précédent, le test concerne en principe des situations sans ex aequo. La correction proposée pour tenir compte des ex aequo revient à multiplier la statistique K précédente par $\frac{s^2}{s'^2}$ où s^2 est la variance de la série $(1, 2, \ldots, N)$ et s'^2 la variance des rangs réels. L'effet de la correction est en général peu important.

Test post hoc de comparaisons par paires

Dans le cas où le test de Kruskal-Wallis conclut à une différence significative entre les groupes, on peut compléter l'étude en réalisant des tests de comparaisons par paires. Plusieurs tests post hoc ont été proposés. Détaillons le test (analogue à celui de Bonferroni-Dunn) qui est disponible dans Statistica.

Pour comparer le groupe a et le groupe b, on calcule :

$$Z = \frac{\overline{R_a} - \overline{R_b}}{E_{ab}} \text{ avec } E_{ab}^2 = \frac{N(N+1)}{12}\left(\frac{1}{n_a} + \frac{1}{n_b}\right)$$

où N désigne l'effectif total des k groupes et où les rangs moyens $\overline{R_a}$ et $\overline{R_b}$ sont calculés en considérant le protocole des rangs sur l'ensemble des k groupes. La loi suivie est la loi normale.

Pour énoncer un résultat au seuil global (*familywise*) α_{FW}, chaque test individuel est réalisé au seuil $\alpha_{PC} = \frac{\alpha_{FW}}{c}$ avec $c = \frac{k(k-1)}{2}$ (nombre de comparaisons).

Exemple : 12 sujets soumis à 3 conditions différentes. On fait l'hypothèse que les sujets du 3è groupe auront un score significativement supérieur.

Valeurs observées de la variable dépendante :

G_1	G_2	G_3
.994	.795	.940
.872	.884	.979
.349	.816	.949
	.981	.890
		.978

Protocole des rangs et calcul des rangs moyens :

	G_1	G_2	G_3
	12	2	7
	4	5	10
	1	3	8
		11	6
			9
R_j	17	21	40
$\overline{R_j}$	5.67	5.25	8

On obtient ici :

$$K = \frac{12}{12 \times 13}\left[3(5.67)^2 + 4(5.25)^2 + 5(8)^2\right] - 3(12+1)$$

c'est-à-dire $K = 1.51$. Pour $\alpha = 5\%$, la table donne : $K_{crit} = 5.63$. On retient donc H_0.

Exemple de test post hoc (ici inutile, car on a conclu sur H_0) :
Pour un résultat au seuil de 5%, on considère la valeur critique de la loi normale correspondant à $\alpha = 1.67\%$, c'est-à-dire : $Z_c = 2.39$.
G_2 v/s G_3 : $E^2 = 13(\frac{1}{4} + \frac{1}{5}) = 5.85$ d'où $E = 2.42$ et $Z = \frac{8-5.25}{2.42} = 1.14$. Différence non significative.

Tests non paramétriques sur 2 groupes appareillés

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

Effectif de l'échantillon de sujets : n.

Test du χ^2 de Mac Nemar

Il s'applique au cas où la variable X est dichotomique.

Situation générale :

		A_1: $X=1$	$X=0$
A_2	$X=1$	a	c
	$X=0$	b	d

Les paires discordantes sont les observations ($X = 1$ en A_1, $X = 0$ en A_2) et ($X = 0$ en A_1, $X = 1$ en A_2) L'information utile est alors fournie par les effectifs "de discordance" b et c.

Notations
p_1 : fréquence de la combinaison ($X = 1$ en A_1, $X = 0$ en A_2) par rapport à la discordance totale dans la population.
p_2 : fréquence de la combinaison ($X = 0$ en A_1, $X = 1$ en A_2) par rapport à la discordance totale dans la population.

Hypothèses du test
$H_0 : p_1 = p_2 (= 50\%)$
$H_1 : p_1 \neq p_2$

Statistique de test

$$\chi^2 = \frac{(b-c)^2}{b+c}, \quad ddl = 1$$

ou, avec la correction de Yates (petits effectifs) :

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}, \quad ddl = 1$$

Condition d'application : $b + c > 10$.

Remarques.

1. Cette statistique est la distance du χ^2 calculée entre le tableau d'effectifs observés et le tableau d'effectifs théoriques suivant :

		A_1: $X=1$	$X=0$
A_2	$X=1$	a	$\frac{b+c}{2}$
	$X=0$	$\frac{b+c}{2}$	d

2. On peut aussi utiliser la statistique suivante, qui permet éventuellement un test unilatéral :

$$Z = \frac{b - c \pm 1}{\sqrt{b+c}}$$

Z suit la loi normale centrée réduite.
Correction de continuité (± 1) : choisir le signe de façon à diminuer la valeur absolue de la statistique.

3. Sous la forme indiquée dans la remarque 2, il s'agit en fait d'un test des signes sur les protocoles 0-0, 0-1, 1-0, 1-1.

Tests des permutations

Principe : on observe un protocole de différences individuelles d_i. On observe D_+ différences positives et D_- différences négatives, soit $N = D_+ + D_-$ différences non nulles. On élimine les différences nulles.

On imagine tous les protocoles obtenus en affectant arbitrairement le signe "+" ou le signe "−" aux différences absolues $|d_i|$. Les 2^N protocoles ainsi obtenus sont supposés équiprobables.

Pour chaque protocole, on calcule $\sum d_i$. La zone d'acceptation de H_0 est formée des protocoles conduisant à des sommes $\sum d_i$ "proches de 0". La zone de rejet est formée des protocoles conduisant à des valeurs "extrêmes", positives et/ou négatives.

Deux tests basés sur ce principe :
• test des signes : les d_i sont codées $+1$ ou -1
• test de Wilcoxon : les d_i sont les rangs signés.

Test des signes

Un échantillon de sujets, placés dans deux conditions expérimentales différentes : groupes appareillés.

Variable dépendante : ordinale ou numérique.

– protocole du signe des différences individuelles ; modalités : −1, 0, 1
– on élimine les différences nulles

D_+ : nombre de différences positives
D_- : nombre de différences négatives
$N = D_+ + D_-$: nombre total d'observations après élimination des différences nulles.

Hypothèses du test :

H_0 : les différences sont dues au hasard : dans la population parente, la fréquence des différences positives est 50%.

H_1 : Cette fréquence n'est pas 50% (test bilatéral)
ou (tests unilatéraux)
Cette fréquence est inférieure à 50%
Cette fréquence est supérieure à 50%

• *Cas des petits échantillons*

Sous H_0, D_+ suit une *loi binomiale* de paramètres N et $p = 0.5$.

On raisonne en termes de "niveau de significativité".

Par exemple, dans le cas d'un test unilatéral tel que
H_1 : fréquence inférieure à 50%
on calcule la fréquence cumulée $P(X \leq D_+)$ de D_+ pour la loi binomiale $B(N, 0.5)$.
Pour un seuil α donné :
Si $P(X \leq D_+) < \alpha$ on retient H_1
Si $P(X \leq D_+) \geq \alpha$ on retient H_0

• *Cas des grands échantillons : approximation par une loi normale*

$$Z = \frac{2D_+ \pm 1 - N}{\sqrt{N}}$$

Z suit une loi normale centrée réduite.

Correction de continuité (± 1) : choisir le signe de façon à diminuer la valeur absolue de la statistique.

Test de Wilcoxon sur des groupes appareillés
Test T, ou test des rangs signés

C'est le test des permutations appliqué au protocole des rangs signés.

Un échantillon de sujets, placés dans deux conditions expérimentales différentes : groupes appareillés.

Variable dépendante : numérique.

On construit :

– le protocole des effets individuels d_i
– le protocole des valeurs absolues de ces effets $|d_i|$
– le protocole des rangs appliqués aux valeurs absolues, en éliminant les valeurs nulles.

T_+ : somme des rangs des observations tq $d_i > 0$
T_- : somme des rangs des observations tq $d_i < 0$
N = nombre de différences non nulles
$T_m = \min(T_+, T_-)$;
$T_M = \max(T_+, T_-)$

Hypothèses

H_0 : Dans la population parente, les effets individuels positifs et les effets individuels négatifs s'interclassent de manière homogène

H_1 : Les deux classements sont différents (test bilatéral) ou les effets individuels positifs apparaissent plus fréquemment dans les rangs les moins élevés (resp. les plus élevés) (test unilatéral).

• *Cas des petits échantillons*
$N \leq 15$ *: utilisation de tables spécialisées*

On compare T_m aux valeurs critiques indiquées par la table.

• *Cas des grands échantillons*
$N > 15$ *: approximation par une loi normale*

$$Z = \frac{T_+ \pm 0.5 - \frac{N(N+1)}{4}}{E}$$

avec

$$E^2 = \frac{N(N+1)(2N+1)}{24}$$

Z suit une loi normale centrée réduite.

Remarques.

Dans le test de Wilcoxon, la variable X est supposée continue. La probabilité d'observer des ex aequo est donc *en théorie* nulle. Cependant, certains auteurs ont proposé des formules correctives pour tenir compte des ex aequo.

Tests non paramétriques sur k groupes appareillés

Situation envisagée : un plan $\mathcal{S} * \mathcal{A}_k$ avec un facteur $\mathcal{A}$ à k niveaux définissant des groupes appareillés et une variable dépendante X ordinale ou numérique.

Effectif de l'échantillon de sujets : n.

Test Q de Cochran

Il s'applique au cas où la variable X est dichotomique. Par exemple : n sujets sont soumis aux k items d'un test. Résultats possibles : succès/échec. Les items ont-ils le même niveau de difficulté ?

H_0 : Dans la population, la probabilité de la modalité "1" est la même dans toutes les conditions : $\pi_1 = \pi_2 = \ldots = \pi_k$
H_1 : Dans la population, la probabilité de la modalité "1" n'est pas la même dans toutes les conditions.

Protocole observé :

Suj.	A_1	A_2	...	A_k	L_i	L_i^2
s_1	1	1	...	0	L_1	L_1^2
s_2	1	0	...	0	L_2	L_2^2
...						
s_n					L_n	L_n^2
G_j	G_1	G_2	...	G_k	G	

Notations :
L_i : somme sur la ligne i
G_j : somme sur la colonne j
G : somme des G_j avec $j = 1, 2, \ldots, k$
$\overline{G}$: moyenne des G_j

Statistique :

$$Q = \frac{(k-1)\left(k\sum G_j^2 - G^2\right)}{k\sum L_i - \sum L_i^2}$$

Calcul équivalent :

$$Q = \frac{k(k-1)\sum(G_j - \overline{G})^2}{k\sum L_i - \sum L_i^2}$$

Loi suivie par Q :

On élimine les lignes composées exclusivement de "0" ou exclusivement de "1". Soit N le nombre de lignes restantes.
Si $N \geq 4$ et $Nk > 24$, Q suit approximativement une loi du χ^2 à $k-1$ ddl.

Exemple :

Trois ascensions tentées par 5 alpinistes. Les trois ascensions présentent-elles le même niveau de difficulté ?

Suj.	A_1	A_2	A_3	L_i	L_i^2
s_1	1	1	0	2	4
s_2	1	0	1	2	4
s_3	0	0	1	1	1
s_4	0	1	1	2	4
s_5	1	0	1	2	4
G_j	3	2	4	9	

$$Q = \frac{2\left[3\,(9+4+16) - 9^2\right]}{3 \times 9 - 17} = 1.2$$

Remarques

1. Lorsque $k = 2$, le test Q de Cochran est exactement le test du χ^2 de Mac Nemar, sans correction de Yates.

2. On ne trouve pas couramment de tables du test Q de Cochran pour les petites valeurs de N et k. En effet, les valeurs critiques dépendent alors non seulement de N et k, mais aussi de la "signature" correspondant à l'échantillon observé. c'est-à-dire du nombre de lignes ayant pour somme 1, 2, ..., $k-1$.

3. Tests post hoc de comparaison par paires

Lorsque le test de Cochran conclut sur H_1, on peut utiliser un test analogue à celui de Bonferroni Dunn pour faire des comparaisons par paires entre les différentes conditions.

Pour comparer la condition a et la condition b, on calcule :

$$Z = \frac{p_a - p_b}{E}$$

où p_a et p_b sont les proportions de "1" dans chacune des conditions et où E est calculée par :

$$E^2 = \frac{k\sum L_i - \sum L_i^2}{n^2 k(k-1)}$$

La loi suivie est une loi normale.

Pour énoncer un résultat au seuil global (*familywise*) α_{FW}, chaque test individuel est réalisé au seuil $\alpha_{PC} = \frac{\alpha_{FW}}{c}$ avec $c = \frac{k(k-1)}{2}$ (nombre de comparaisons).

Analyse de variance à deux facteurs par les rangs de Friedman

Ce test s'applique au cas où la variable X est ordinale ou numérique.

H_0 : Dans les différentes conditions, les médianes sont égales : $M_1 = M_2 = \ldots = M_k$.
H_1 : Les k médianes ne sont pas toutes égales.

Statistique de test :

On calcule un protocole de rangs *par sujet*.

Notations :

n : nombre de sujets
k : nombre de conditions
R_j : somme des rangs de la colonne j (dans la condition A_j).

La statistique de Friedman est donnée par :

$$F_r = \left[\frac{12}{nk(k+1)}\sum R_j^2\right] - 3n(k+1)$$

ou, de manière équivalente :

$$F_r = \frac{12}{nk(k+1)}\sum\left(R_j - \frac{n(k+1)}{2}\right)^2$$

F_r est tabulée pour les petites valeurs de n et k. Au delà, F_r suit approximativement une loi du χ^2 à $(k-1)$ ddl.

Exemple :

Trois individus statistiques (groupes de sujets) sont soumis à 4 conditions. Y a-t-il une différence significative entre les 4 conditions ?

Protocole observé :

Ind.	Conditions I	II	III	IV
i_1	9	4	1	7
i_2	6	5	2	8
i_3	9	1	2	6

Protocole des rangs par individu statistique :

Ind.	Conditions I	II	III	IV
i_1	4	2	1	3
i_2	3	2	1	4
i_3	4	1	2	3
R_j	11	5	4	10

$$F_r = \frac{12}{3 \times 4 \times (4+1)} \left(11^2 + 5^2 + 4^2 + 10^2\right) - 3 \times 3 \times (4+1)$$

D'où : $F_r = 7.4$

Lecture de la table : Au seuil de 5%, $F_{r,crit} = 7.4$. Il est difficile de conclure sur H_0 ou H_1.

Remarques.

1. Correction possible pour les ex aequo. Mais l'effet est assez peu important. Il s'agit alors de remplacer la somme de variances "théoriques" $\frac{nk(k+1)}{12}$ par la somme des variances réelles des lignes.

2. Si le nombre de conditions est 2 (k=2) et s'il n'y a pas de sujet ayant un score identique dans les deux conditions, le test de Friedman équivaut à l'approximation normale du test du signe. Le test de Friedman apparaît donc comme une extension du test du signe. Il s'agit d'un test relativement peu puissant.

3. Test post hoc de comparaison par paires

Lorsque le test de Friedman conclut sur H_1, on peut utiliser un test analogue à celui de Bonferroni Dunn pour faire des comparaisons par paires entre les différentes conditions.

Pour comparer la condition a et la condition b, on calcule :

$$Z = \frac{R_a - R_b}{E} \quad \text{avec} \quad E^2 = \frac{nk(k+1)}{6}$$

La loi suivie est une loi normale.

Pour énoncer un résultat au seuil global (*familywise*) α_{FW}, chaque test individuel est réalisé au seuil $\alpha_{PC} = \frac{\alpha_{FW}}{c}$ avec $c = \frac{k(k-1)}{2}$ (nombre de comparaisons).

Corrélation et régression linéaires

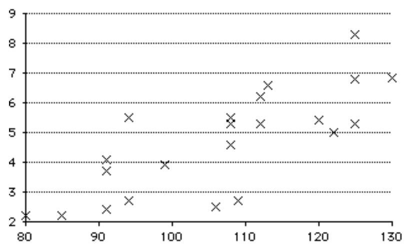
Corrélation linéaire

Situation envisagée : un échantillon de sujets, deux variables numériques observées (ou une variable observée dans deux conditions).

Données :

	X	Y
s_1	x_1	y_1
s_2	x_2	y_2
...	...	...

Nuage de points : points (x_i, y_i)



Covariance des variables X et Y

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

ou

$$Cov(X,Y) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \overline{x}\ \overline{y}$$

Covariance corrigée des variables X et Y

$$Cov_c(X,Y) = \frac{n}{n-1}\, Cov(X,Y)$$

Coefficient de corrélation de Bravais Pearson

On désigne par $s(X)$ et $s(Y)$ les écarts types de X et Y et par $s_c(X)$, $s_c(Y)$ leurs écarts types corrigés.

$$r = \frac{Cov(X,Y)}{s(X)s(Y)} = \frac{Cov_c(X,Y)}{s_c(X)s_c(Y)}$$

Remarques

- r n'est pas une estimation correcte du coefficient de corrélation dans la population. Certains logiciels de statistiques donnent comme estimation :

$$r_{aj} = \sqrt{1 - \frac{(1-r^2)(N-1)}{N-2}}$$

- Il existe des relations non linéaires
- Corrélation n'est pas causalité

Alpha de Cronbach

Dans un questionnaire, un groupe d'items X_1, X_2, ..., X_k (par exemple des scores sur des échelles de Likert) mesure un même aspect du comportement.

Problème : comment mesurer la cohérence de cet ensemble d'items ?

Dans le cas de 2 items : la cohérence est d'autant meilleure que la covariance entre ces items est plus élevée. Le rapport :

$$\alpha = 4\ \frac{Cov(X_1, X_2)}{Var(X_1 + X_2)}$$

– vaut 1 si $X_1 = X_2$
– vaut 0 si X_1 et X_2 sont indépendantes
– est négatif si X_1 et X_2 sont anti-corrélées.

Généralisation :
On introduit $S = X_1 + X_2 + \ldots + X_k$ et on considère le rapport :

$$\alpha = \frac{k}{k-1}\ \frac{2\sum_{i<j} Cov(X_i, X_j)}{Var(S)} = \frac{k}{k-1}\left[1 - \frac{\sum Var(X_i)}{Var(S)}\right]$$

Ce rapport est le coefficient α de Cronbach.

Justification théorique : Pour chaque item, et pour leur somme :

Valeur mesurée = "Vraie valeur" de l'item sur le sujet + erreur aléatoire

La fiabilité (théorique) est :

$$\rho = \frac{\text{Variance(Vraies valeurs)}}{\text{Variance(Valeurs mesurées)}}$$

α est une estimation de la fiabilité de la somme des k items.

Exemple : On a mesuré 3 items (échelle en 5 points) sur 6 sujets.

	X_1	X_2	X_3	S
s1	1	2	2	5
s2	2	1	2	5
s3	2	3	3	8
s4	3	3	5	11
s5	4	5	4	13
s6	5	4	4	13
Var.	2.167	2.000	1.467	13.77

On obtient : $\alpha = \frac{3}{2}\left[1 - \frac{2.167 + 2 + 1.467}{13.77}\right] = 0.886$

Signification de α : on considère généralement que l'on doit avoir $\alpha \geq 0.7$. Mais une valeur trop proche de 1 revèle une pauvreté dans le choix des items.

Significativité du coefficient de corrélation

- Les données (x_i, y_i) constituent un échantillon
- r est une statistique
- ρ : coefficient de corrélation sur la population

H_0 : Indépendance sur la population ; $\rho = 0$
H_1 : $\rho \neq 0$ (bilatéral) ou $\rho > 0$ ou $\rho < 0$ (unilatéral)

Statistique de test

– Petits échantillons : tables spécifiques. $ddl = n - 2$
– Grands échantillons :

$$T = \sqrt{n-2}\ \frac{r}{\sqrt{1-r^2}}$$

T suit une loi de Student à $n-2$ degrés de liberté.

Conditions d'application

Dans la population parente, le couple (X, Y) suit une *loi normale bivariée*, ce qui implique notamment :

– la normalité des distributions marginales de X et Y ;

– la normalité de la distribution de l'une des variables lorsque l'autre variable est fixée ;

– l'égalité des variances des distributions de l'une des variables pour deux valeurs distinctes de l'autre variable.

Corrélation et statistiques non paramétriques : corrélation des rangs de Spearman

Si les données ne vérifient pas les conditions d'application précédentes, ou si les données observées sont elles-mêmes des classements, on pourra travailler sur les protocoles des rangs définis séparément pour chacune des deux variables.

	Rangs X	Rangs Y
s_1	r_1	r'_1
s_2	r_2	r'_2
...	...	...

Le coefficient de corrélation des deux protocoles de rangs est appelé coefficient de corrélation de Spearman, et noté R_s.

Calcul de R_s (en l'absence d'ex aequo)

On calcule les différences individuelles $d_i = r_i - r'_i$, puis

$$R_s = 1 - \frac{6\sum d_i^2}{N(N^2-1)}$$

Significativité du coefficient de corrélation de Spearman

– Pour $N \leq 30$, on utilise en général des tables spécialisées.

– Lorsque $N > 30$:

Certains auteurs (et Statistica) utilisent :

$$t = \sqrt{N-2}\ \frac{R_s}{\sqrt{1-R_s^2}}$$

et une loi de Student à $N-2$ ddl.

D'autres auteurs utilisent la statistique :

$$Z = \sqrt{N-1}\ R_s$$

et une loi normale centrée réduite.

Remarque. Siegel et Castellan donnent “20 ou 25” comme seuil pour les grands échantillons et indiquent que la première statistique est “légèrement meilleure” que la seconde.

Le coefficient τ de Kendall

Avec un protocole comportant N sujets, on a $\frac{N(N-1)}{2}$ paires de sujets.

On examine chaque paire de sujets, et on note si les deux classements comportent une inversion ou non.

s_3	3	6
s_4	5	2

Désaccord, Inversion

s_3	3	4
s_4	5	6

Accord, Pas d'inversion

Le coefficient τ est alors défini par :

$$\tau = \frac{\text{Nb d'accords} - \text{Nb de désaccords}}{\text{Nb de paires}}$$

ou

$$\tau = 1 - \frac{2 \times \text{Nombre d'inversions}}{\text{Nombre de paires}}$$

Significativité du τ de Kendall

H_0 : Indépendance des variables. $\tau = 0$
H_1 : $\tau \neq 0$

Pour $N > 10$, sous H_0, la statistique

$$Z = 3\tau\sqrt{\frac{N(N-1)}{2(2N+5)}}$$

suit une loi normale centrée réduite.

Régression linéaire

Rôle “explicatif” de l'une des variables par rapport à l'autre. Les variations de Y peuvent-elles (au moins en partie) être expliquées par celles de X ? Peuvent-elles être prédites par celles de X ?

Modèle permettant d'estimer Y connaissant X

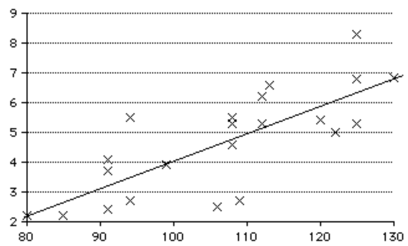
Droite de régression de Y par rapport à X :

La droite de régression de Y par rapport à X a pour équation :

$$y = b_0 + b_1 x$$

avec :

$$b_1 = \frac{Cov(X,Y)}{s^2(X)} \quad ; \quad b_0 = \overline{Y} - b_1\overline{X}$$



Remarques

Si les variables X et Y sont centrées et réduites, l'équation de la droite de régression est :

$$Y = rX$$

On définit le *coefficient de régression standardisé* par :

$$\beta_1 = b_1\frac{s(X)}{s(Y)}$$

Dans le cas de la régression linéaire simple : $\beta_1 = r$.

Comparaison des valeurs observées et des valeurs estimées

Valeurs estimées : $\widehat{y}_i = b_0 + b_1 x_i$: variable $\widehat{Y}$
Erreur (ou résidu) : $e_i = y_i - \widehat{y}_i$: variable E

Les variables $\widehat{Y}$ et E sont indépendantes et on montre que :

$$s^2(Y) = s^2(\hat{Y}) + s^2(E)$$

avec :

$$\frac{s^2(E)}{s^2(Y)} = 1 - r^2 \quad ; \quad \frac{s^2(\hat{Y})}{s^2(Y)} = r^2$$

$s^2(\hat{Y})$: variance *expliquée* (par la variation de X, par le modèle)

$s^2(E)$: variance *perdue* ou *résiduelle*

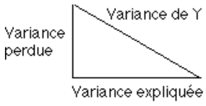
r^2 : part de la variance de Y qui est expliquée par la variance de X. r^2 est appelé *coefficient de détermination*.

Exemple : $r = 0.86$

$r^2 = 0.75$; $1 - r^2 = 0.25$; $\sqrt{1 - r^2} = 0.5$.

• La part de la variance de Y expliquée par la variation de X est de 75%.

• L'écart type des résidus est la moitié de l'écart type de Y.



Interprétation géométrique, du point de vue des variables

Lorsque les variables X et Y sont centrées et réduites, elles peuvent être vues comme des vecteurs OM et ON de norme 1 dans un espace géométrique de dimension n (le nombre d'observations).

Dans cette interprétation :

– La variable $\hat{Y}$ est la projection de Y sur X

– Le coefficient de corrélation r est le cosinus de l'angle (OM, ON).

Test du coefficient de corrélation à l'aide du F de Fisher

Valeurs estimées : $\hat{y}_i = b_0 + b_1 x_i$
Erreur (ou résidu) : $e_i = y_i - \hat{y}_i$

On introduit les sommes de carrés suivantes :

$$SC_{\text{Totale}} = \sum (y_i - \overline{y})^2$$

$$SC_{\text{Régression}} = \sum (\hat{y}_i - \overline{y})^2$$

$$SC_{\text{Résidus}} = \sum (y_i - \hat{y}_i)^2$$

Lien avec le coefficient de corrélation

$r^2 = \dfrac{SC_{\text{Régression}}}{SC_{\text{Totale}}}$ est le *coefficient de détermination*

Tableau d'analyse de variance

Source	SC	ddl	CM	F
Régression	$SC_{\text{Régression}}$	1	CM_{Reg}	F_{obs}
Résiduelle	$SC_{\text{Résidus}}$	$n-2$	CM_{Res}	
Total	SC_{Totale}	$n-1$		

$F_{obs} = \dfrac{CM_{\text{Reg}}}{CM_{\text{Res}}} = (n-2)\dfrac{r^2}{1-r^2}$ suit une loi de Fisher à 1 et $n-2$ ddl.
On retrouve : $F_{obs} = T^2_{obs}$

Estimations de $\hat{y}$ et de y par des intervalles de confiance

Dans la population, le lien entre X et Y suit le modèle mathématique :

$$Y = B_0 + B_1 X + \epsilon$$

où B_0 et B_1 sont des constantes numériques et ϵ est une variable statistique centrée et indépendante de X

A partir de l'échantillon, nous avons calculé $b_0, b_1, e_1, \ldots, e_n$ tels que :

$$y_i = b_0 + b_1 x_i + e_i$$

$$\hat{y}_i = b_0 + b_1 x_i$$

Mais un autre échantillon amènerait d'autres valeurs de ces paramètres : b_0, b_1 et e_i ne sont que des estimations de B_0, B_1 et ϵ_i.

Questions que l'on peut se poser :

– Quelle estimation peut-on donner de la variance de ϵ ?

– On peut voir $\hat{y}_i$ comme une estimation ponctuelle de la moyenne des valeurs de Y sur la population lorsque $X = x_i$. Peut-on déterminer un intervalle de confiance pour cette moyenne ?

– Etant donné une valeur x_i de X, quel intervalle de confiance peut-on donner pour les valeurs de Y correspondantes ?

Estimation de $Var(\epsilon)$:

$$s^2 = \frac{SC_{Res}}{n-2}$$

Estimation par un intervalle de confiance de la moyenne de Y pour X fixé :

Pour une valeur x_p fixée de X, la variance des valeurs estimées $\hat{y}_p$ est estimée par :

$$s^2_{\hat{y}_p} = s^2 \left(\frac{1}{n} + \frac{(x_p - \overline{x})^2}{\sum x_i^2 - n\overline{x}^2} \right)$$

Notons $Moy(Y|X = x_p)$ la moyenne de la variable Y lorsque la variable X est égale à x_p.

Un intervalle de confiance de $Moy(Y|X = x_p)$ avec un degré de confiance $1 - \alpha$ est donné par :

$$\hat{y}_p - t_\alpha s_{\hat{y}_p} \leq Moy(Y|X = x_p) \leq \hat{y}_p + t_\alpha s_{\hat{y}_p}$$

où t_α est la valeur du T de Student à $n-2$ ddl telle que $P(|T| > t_\alpha) = \alpha$

Détermination d'un intervalle de confiance pour les valeurs de Y : intervalle de prévision

$Moy(Y|X = x_p)$ est connue par une estimation ponctuelle ($\widehat{y}_p$) et un intervalle de confiance.

La différence $Y - Moy(Y|X = x_p)$ est le résidu ϵ, dont on peut également donner un intervalle de confiance.

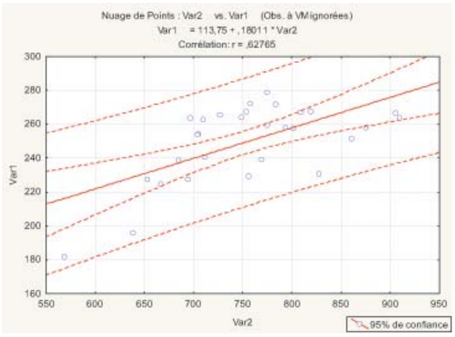
Finalement, on pourra écrire l'intervalle de confiance :

$$\widehat{y}_p - t_\alpha s_{ind} \leq y \leq \widehat{y}_p + t_\alpha s_{ind}$$

avec :

$$s_{ind}^2 = s^2\left(1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{\sum x_i^2 - n\overline{x}^2}\right) = s^2 + s_{\widehat{y}_p}^2$$

Intervalle de confiance et intervalle de prévision apparaissent dans les graphiques réalisés par les logiciels sous forme de "bandes" :



Corrélation et Régression linéaires multiples

Position du problème

Une population (ou un échantillon) sur laquelle on a observé un ensemble de variables numériques.

	X_1	X_2	...	X_p
s_1	x_{11}	x_{12}	...	x_{1p}
...	...	...	...	...

Exemple avec trois variables

X_1 : âge de la mère
X_2 : rang de l'enfant dans la fratrie
X_3 : poids de l'enfant à la naissance
n : nombre d'observations (ici : $n = 200$)

	X_1	X_2	X_3
s_1	26.5	1	2100
...	...	...	...
s_{200}	34.5	2	4500

Nuage de points

Pour trois variables : représentation dans l'espace.

Pour plus de trois variables, détermination des directions de "plus grande dispersion du nuage" : analyse en composantes principales.

Paramètres associés aux données

Matrice des covariances, matrice des corrélations.

Sur l'exemple : coefficients de corrélation des variables prises 2 à 2 :

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array} \begin{bmatrix} X_1 & X_2 & X_3 \\ 1 & 0.60 & 0.24 \\ 0.60 & 1 & 0.28 \\ 0.24 & 0.28 & 1 \end{bmatrix}$$

$r_{xz} = 0.24$ $**$: âge et poids sont corrélés
$r_{yz} = 0.28$ $**$: rang et poids sont corrélés
$r_{xy} = 0.60$ $**$: rang et âge sont fortement corrélés

Régression : "hyperplan" de régression

L'une des variables (Y) est la variable "à prévoir". Les autres ($X_1, X_2, \ldots, X_p$) sont les variables "prédicatives". L'hyperplan de régression a pour équation :

$$Y = b_0 + b_1 X_1 + \ldots + b_p X_p$$

Calcul de b_1, b_2, ..., b_p : ce sont les solutions du système d'équations linéaires :

$$j = 1, 2, \ldots, p \qquad \sum_i \text{Cov}(X_i, X_j)\ b_i = \text{Cov}(Y, X_j)$$

Coefficients de régression standardisés :

Pour $i = 1, 2, \ldots, p$, on pose : $\beta_i = b_i \dfrac{s(X_i)}{s(Y)}$

Ce sont les coefficients que l'on obtiendrait en travaillant sur les variables centrées réduites associées aux variables Y et X_1 à X_p.

Contrairement aux b_i, les coefficients β_i sont comparables entre eux.

Coefficient de corrélation multiple

$\hat{Y}$: valeurs estimées à l'aide de l'équation précédente.

$$R = r_{Y\hat{Y}} = \frac{Cov(Y, \hat{Y})}{s(Y)s(\hat{Y})}$$

Exemple : dans l'exemple proposé, $R = 0.29$

Comme précédemment, R^2 est la part de la variance "expliquée par le modèle".

Test du coefficient de corrélation multiple :

Valeurs estimées : $\widehat{y}_i = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip}$
Erreurs (ou résidus) : $e_i = y_i - \widehat{y}_i$

On introduit les sommes de carrés suivantes :

$$SC_{\text{Totale}} = \sum (y_i - \overline{y})^2$$

$$SC_{\text{Régression}} = \sum (\widehat{y}_i - \overline{y})^2$$

$$SC_{\text{Résidus}} = \sum (y_i - \widehat{y}_i)^2$$

Lorsque l'on a p prédicteurs, le tableau d'analyse de variance devient :

Source	SC	ddl	CM	F
Régression	$SC_{\text{Régression}}$	p	CM_{Reg}	F_{obs}
Résiduelle	$SC_{\text{Résidus}}$	$n-p-1$	CM_{Res}	
Total	SC_{Totale}	$n-1$		

$F_{obs} = \dfrac{CM_{\text{Reg}}}{CM_{\text{Res}}} = \dfrac{n-p-1}{p} \dfrac{R^2}{1-R^2}$ suit une loi de Fisher à p et $n-p-1$ ddl.

Coefficients de corrélation partielle

Ce sont les corrélations obtenues en contrôlant une partie des variables. Par exemple, dans le cas de 3 variables X, Y et Z, pour calculer $r_{yz.x}$:
– On calcule les résidus de la régression de Z par rapport à X
– On calcule les résidus de la régression de Y par rapport à X
– On calcule le coefficient de corrélation entre les deux séries de résidus obtenues.

Sur l'exemple : $r_{yz.x} = 0.18$ $**$: A âge constant, rangs et poids sont corrélés
$r_{xz,y} = 0.09$ NS : A rang constant, pas de corrélation entre âge et poids.

Seul le rang de naissance intervient. L'âge de la mère n'est lié au poids de l'enfant que par le rang de naissance.

Analyse de médiation

L'effet d'une VI sur une VD est-il direct, ou relève-t-il plutôt d'un facteur intermédiaire M ?

– On effectue la régression linéaire de la VD sur la VI :

$$VD = b_0 + b_1 VI$$

Coefficient de régression standardisé : β_1

– On effectue la régression linéaire de la variable de médiation M sur la VI :

$$M = b'_0 + b'_1 VI$$

Coefficient de régression standardisé : β'_1

– On effectue la régression linéaire multiple de la VD sur les deux variables M et VI :

$$VD = b''_0 + b''_1 VI + b''_2 M$$

Coefficients de régression standardisés : β''_1, β''_2

Si β''_2 est significativement différent de 0, et que β''_1 est nettement plus proche de 0 que β_1, il y a médiation (partielle ou totale). En particulier, il y a médiation si β''_1 n'est pas significativement différent de 0 alors que β_1 l'était.