Statistiques paramétriques et non paramétriques Informatique

E.C. PSRS73B et PSRS73C

Présentation du cours 2005/2006

Organisation matérielle

Cours-TD de Statistiques : 24 heures.
Travaux dirigés d'informatique : 12 heures.

Mardi - 13h45-17h - A204

Monitorat informatique

Contrôle des connaissances : (contrôle continu)
EC PSRS73B : Examen écrit (2 heures)
EC PSRS73C : Note de TD

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

Documents fournis

Transparents du cours de statistiques
Fiches de TD de statistiques et d'informatique

Adresses Web

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

Contenu

Statistiques :

Aspects méthodologiques : quelle stratégie pour analyser les données ? Quelles sont les méthodes disponibles pour tester telle hypothèse ?

Compléments aux méthodes de statistiques descriptives et inférentielles vues en licence :
Statistiques paramétriques : effet calibré, taille d'un effet ; test de normalité des distributions parentes ; loi de Fisher Snedecor ; ANOVA.
Statistiques non paramétriques : test de Kolmogorov-Smirnov ; test de Kruskal-Wallis.
Compléments sur la corrélation et la régression linéaires à 2 ou plusieurs variables.

Informatique :

Mise en oeuvre des traitements étudiés à l'aide d'un logiciel de traitement statistique professionnel (Statistica).

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.

Test de Kolmogorov-Smirnov

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	$S_n(z_i)$	$F(z_i)$	Ecart absolu
-1.3024	0.1	0.0964	
-0.8372	0.2	0.2012	
-0.8372	0.3	0.2012	
-0.3721	0.4	0.3549	
-0.3721	0.5	0.3549	
-0.3721	0.6	0.3549	0.2451
0.0930	0.7	0.5371	
1.0233	0.8	0.8469	
1.4884	0.9	0.9317	
1.4884	1	0.9317	

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

Consultation de tables spécialisées : Pour $\alpha = 5\%$, $D_{crit} = 0.41$

Conclusion : L'hypothèse de normalité de X sur la population parente ne peut pas être rejetée.

Test de Lilliefors

L'utilisation du test de Kolmogorov-Smirnov est critiquable lorsque la moyenne et l'écart type dans la population sont estimés à partir de l'échantillon. Lilliefors a proposé un autre test, plus satisfaisant d'un point de vue théorique :
– la statistique de test est la même
– les tables des valeurs critiques sont différentes.

Dans l'exemple ci-dessus : $L_{obs} = 0.2451$ et pour $\alpha = 5\%$, $L_{crit} = 0.258$.

Conclusion : On ne rejette pas l'hypothèse de normalité de X dans la population parente.

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

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'homogénéïté 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 ?

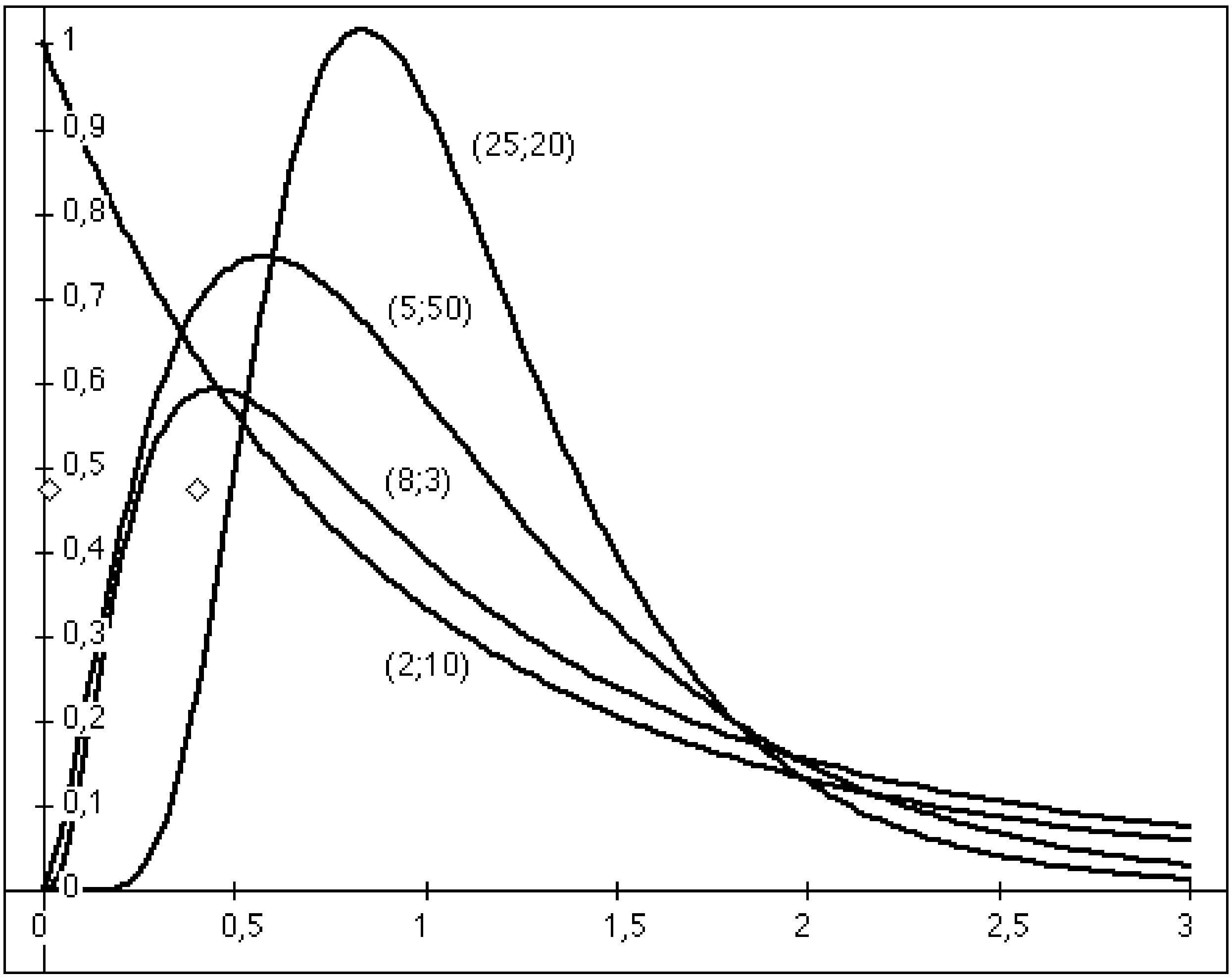
H_0 : Les variances sont égales
H_1 : La première variance est supérieure à la deuxième.

Statistique de test

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

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.

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 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 $s_{1c} = s_{2c}$.

Taille d'un effet – Puissance d'un test

Exemple introductif

On se place dans une situation de comparaison de deux groupes indépendants, avec une VD numérique.

Ces deux groupes sont issus de deux populations et l'on estime que :

– La moyenne de la VD dans la population 1 est 100.
– La moyenne de la VD dans la population 2 est 105.

On se pose des questions telles que :

– La différence entre les deux groupes est-elle grande, facile à mettre en évidence ou au contraire petite, difficile à mettre en évidence ?

– Est-on sûr de pouvoir mettre en évidence une différence entre les deux groupes ?

– Si les effectifs des échantillons sont $n_1 = 25$ et $n_2 = 30$, pourra-t-on mettre en évidence une différence entre les deux groupes ?

– Quelles tailles d'échantillons doit-on choisir pour avoir au moins 80% de chances de mettre en évidence une différence entre les deux groupes ?

Rappel concernant les tests statistiques : erreurs de type I et II

		Hypothèse vraie H_0	H_1
Hypothèse	H_0	$1-\alpha$	β
retenue	H_1	α	$1-\beta$

• α : seuil de significativité. C'est aussi la probabilité de rejeter H_0 alors que H_0 est vraie (risque de première espèce ou risque de commettre une erreur de type I) : risque de voir une différence là où il n'y en a pas.

• β : risque de seconde espèce. C'est la probabilité d'accepter H_0 alors que H_0 est fausse (risque de commettre une erreur de type II) : risque de ne pas mettre en évidence une différence qui, pourtant, existe.

• $1-\beta$: probabilité de détecter correctement un cas où H_0 doit être rejetée. **Puissance du test**.

Comment influer sur la puissance d'un test ?

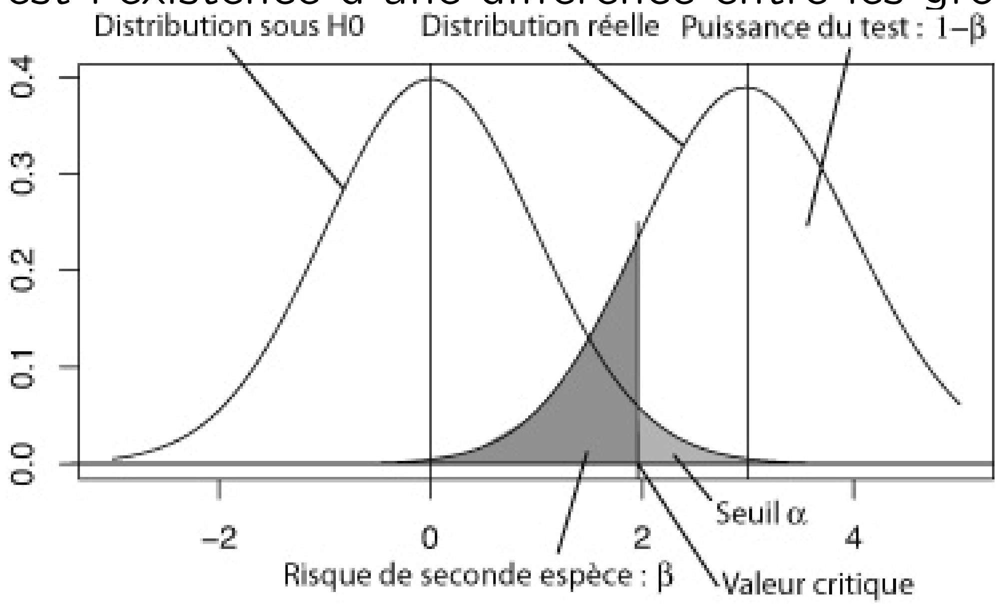
– Changement du seuil

– Choix d'une autre variable dépendante

– Taille des échantillons

Risque β et puissance du test $1-\beta$

Illustration graphique dans le cas où l'hypothèse *vraie* est l'existence d'une différence entre les groupes.



Sous l'hypothèse nulle H_0, la distribution de la statistique de test est celle représentée par la courbe de gauche.

Mais, la *vraie* distribution de la statistique de test est représentée par la courbe de droite.

La probabilité que le test nous permette de conclure correctement est de $1-\beta$.

Taille d'un effet - Effet calibré de Cohen

Deux groupes appareillés - Variable numérique

Moyenne des différences individuelles dans la population : δ.
Ecart type des différences individuelles : σ.

La **taille de l'effet** est définie par :

$$d = \frac{\mid \delta \mid}{\sigma}$$

Estimation de la taille de l'effet à partir d'un échantillon : **effet calibré** observé sur l'échantillon :

$\overline{d}$: moyenne observée sur l'échantillon
s_c : écart type corrigé observé sur l'échantillon.

L'effet calibré est défini par :

$$\text{EC} = \frac{\mid \overline{d} \mid}{s_c}$$

Critère psychométrique :

- EC $< 1/3$: effet faible
- $1/3 \leq$ EC $\leq 2/3$: effet moyen
- EC $> 2/3$: effet important.

Deux groupes indépendants - Variable numérique

Variable dépendante X définie sur deux populations.
Moyennes (sur les populations) : μ_1 et μ_2.
Ecart type (le même pour les deux populations) : σ.

Taille de l'effet :

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

Effet calibré

Paramètres évalués à partir de deux échantillons de tailles n_1 et n_2 :

Moyennes : $\overline{x}_1$, $\overline{x}_2$
Ecarts types (non corrigés) : s_1 et s_2
Ecart type corrigé pondéré s défini par : $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

Effet calibré : $\text{EC} = \dfrac{|\overline{x}_1 - \overline{x}_2|}{s}$

Deux groupes indépendants - Variable dichotomique

Paramètres évalués à partir de deux échantillons de tailles n_1 et n_2 :

Fréquences observées : f_1, f_2
Fréquence pondérée moyenne : f, définie en général par : $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n2}$

Effet calibré : $\text{EC} = \dfrac{|f_1 - f_2|}{\sqrt{f(1-f)}}$

Estimer la puissance d'un test

Exemple

– On se place dans une situation de groupes appareillés.
– On connaît une évaluation de la taille de l'effet.
– Pour une taille d'échantillon donnée, et un seuil donné, quelle estimation peut-on faire de la puissance du test (autrement dit : quelles chances a-t-on de tirer un échantillon permettant de conclure sur H_1) ?

Par exemple :
– on estime que $d = 0.33$
– on choisit $n = 40$ et $\alpha = 5\%$

On remarque que la statistique de test s'écrit :

$$t = \text{EC}\ \sqrt{n}$$

On calcule : $\Delta = d\sqrt{n} = 0.33\sqrt{40} = 2.09$

Par lecture de la table, on a : Puissance $= 0.56$

Estimer la taille d'échantillon requise

Deux groupes appareillés - Variable numérique

– On se place dans une situation de groupes appareillés
– On connaît une évaluation de la taille de l'effet d
– On se donne un seuil donné et une puissance de test
– Quelle taille minimale d'échantillon faut-il choisir ?

On remarque que la statistique de test s'écrit :

$$t = \text{EC}\ \sqrt{n}$$

On introduit : $\Delta = d\sqrt{n}$

Par lecture de la table, on obtient la valeur Δ requise.

On calcule ensuite n de façon que : $\Delta = d\sqrt{n}$.

Exemple : Dans la situation précédente ($d = 0.33$ et $\alpha = 5\%$), quelle taille d'échantillon faut-il choisir pour avoir au moins 4 chances sur 5 de conclure sur H_1 ?

On veut que la puissance du test soit au moins 80%.
Lecture de la table : $\Delta = 2.80$
L'équation : $0.33\sqrt{n} = 2.80$ donne : $n = 72$.

La taille minimale d'échantillon est donc de 72.

Deux groupes indépendants équilibrés

Variable numérique ou dichotomique.

Soit n la taille commune de chacun des deux échantillons. La statistique de test s'écrit alors :

$$t = \text{EC}\ \sqrt{\frac{n}{2}}$$

On raisonne comme précédemment, mais on a alors :

$$\Delta = d\sqrt{\frac{n}{2}}$$

Deux groupes indépendants non équilibrés

Dans ce cas : $\Delta = d\sqrt{\frac{n_h}{2}}$

où n_h est la *moyenne harmonique* des deux tailles d'échantillons définie par :

$$\frac{2}{n_h} = \frac{1}{n_1} + \frac{1}{n_2} \quad \text{ou} \quad n_h = \frac{2n_1n_2}{n_1 + n_2}$$