LES TESTS CLASSIQUES COURANTS

Ce document donne les informations essentielles sur les tests statistiques classiques utiles dans quatre situations : tests sur moyenne ou médiane, tests sur variance, tests sur distributions et tests sur effectfs. Tous les calculs peuvent être effectués avec les logiciels statistiques courants (notamment R : le nom de la fonction correspondante est indiquée dans la marge droite). Pour chacune des quatre situations, le premier test de la liste est celui qui est plus généralement applicable (mais pas forcément le plus performant).

Tests sur moyenne ou médiane

Test de "Student". Comparaison de deux moyennes dont l'une au moins est estimée à partir d'un échantillon. Le test statistique suit une loi de t. H_0 : les deux moyennes sont égales. Dans le cas de deux moyennes observées, les deux variances peuvent être égales ou non, dans ce cas le nombre de degré de liberté sera différent. `t.test`

Contrairement à ce que l'on peut fréquemment lire, il n'est pas nécessaire que les données soient distribuées selon une loi normale. Par contre les moyennes des échantillons doivent suivre une loi normale, ce qui est généralement le cas grâce au théorème de la limite centrale.

Il est intéressant de noter qu'il est relativement facile de calculer la puissance d'un test de "Student" sous une hypothèse alternative définie (c'est-à-dire de calculer la probabilité, notée β, d'accepter H_0 si elle est fausse). Si H_0 est fausse, le test de "Student" suit approximativement une loi de t non-centrale qui au lieu d'être centrée sur zéro, comme une loi de t, est centrée sur une valeur déterminée par $\sqrt{n}(\mu_1 - \mu_2)/\sigma$, où n est la taille de l'échantillon, μ_1 et μ_2 sont les deux (vraies) moyennes et σ l'écart-type de la population. `power.t.test`

Test de Welch. Comparaison de plusieurs moyennes. Ce test est proche du test de celui d'une analyse de variance à un facteur, mais dans le cas du test de Welch aucune hypothèse n'est faite sur l'égalité des variances entre échantillons. Le test statistique suit une loi de F et est égal à celui calculé pour une analyse de variance mais les nombres de degrés de liberté sont différents. H_0 : toutes les moyennes sont égales. `oneway.test`

Test de Quade. La structure des données analysées pas ce test est semblable au cas d'une analyse de variance à deux facteurs sans réplication. On a donc deux variables catégoriques comme prédicteurs et une seule observation pour chaque combinaison de ces prédicteurs. C'est un test non-paramétrique : la réponse peut être non-normale. L'hypothèse testée est qu'un des deux facteurs a un effet sur la réponse en tenant compte d'un possible effet de l'autre facteur. H_0 : les moyennes pour chaque niveau du facteur testé sont égales. Le test statistique suit une loi de F. `quade.test`

Test de Friedman. Ce test est presqu'identique au test de Quade, la différence est que le test statistique suit une loi du χ^2. Le test de Friedman semble moins puissant que le test de Quade. `friedman.test`

Test de Kruskal et Wallis. C'est un test non-paramétrique de comparaison de plusieurs moyennes (similaire donc à une analyse de variance à un facteur). H_0 : toutes les moyennes sont égales. Le test statistique suit une loi du χ^2. `kruskal.test`

Test de Mann et Whitney. Test non-paramétrique de l'égalité de deux médianes observées. Le test statistique suit une loi particulière qui tend vers une loi normale quand l'effectif de l'échantillon est grand (> 50). `wilcox.test`

Test de Wilcoxon. Test non-paramétrique de la symétrie d'une distribution observée. H_0 : la distribution est symétrique par rapport à zéro. Le test statistique suit une loi particulière qui tend vers une loi normale quand l'effectif de l'échantillon est grand (> 50). Le test de Wilcoxon peut être utilisé pour comparer deux échantillons si les observations sont appariées (par paires de chaque variable). Dans ce cas, H_0 : la distribution des différences entre les deux variables est symétrique par rapport à zéro. (C'est un cas particulier du test de Mann et Whitney.) `wilcox.test`

Tests sur variances

Comparaison de deux variances. Si deux échantillons proviennent de deux populations normalement distribuées avec la même variance, le rapport des deux variances estimées suit une loi de F. Ce test est dû à Fisher mais ce n'est pas celui qui est communément appelé "test de Fisher" (*cf.* ci-dessous). `var.test`

Test de Bartlett. Comparaison de plusieurs variances pour des populations normalement distribuées. H_0 : toutes les variances sont égales. Le test statistique suit une loi du χ^2. Ce test est réputé pour être peu robuste (c'est-à-dire que ses propriétés sont mauvaises si les populations ne sont pas normales, en particulier un risque de première espèce élevé). `bartlett.test`

Test de Fligner et Killeen. Test non-paramétrique de comparaison de plusieurs variances. H_0 : toutes les variances sont égales. Le test statistique suit une loi du χ^2. Des simulations ont montré que le test de Fligner et Killeen est un des plus robustes à la non-normalité des observations. `fligner.test`

Tests sur distributions

Test de Kolmogorov et Smirnov. Un test très répandu et assez puissant pour comparer deux distributions dont au moins une est observée. H_0 : les deux distributions sont identiques. Le test statistique, noté D, est simplement basé sur le plus grand écart entre les deux distributions cumulées et suit une loi particulière. Le test de Kolmogorov et Smirnov peut être utilisé pour tester la normalité d'une distribution observée. `ks.test`

Test de Shapiro et Wilk. Teste la normalité d'une distribution observée. H_0 : la distribution suit une loi normale. Ce test, dont la statistique notée W suit une loi particulière, est beaucoup moins puissant que le test de Kolmogorov et Smirnov. `shapiro.test`

Test de Cramér et von Mises. Ce test est utilisé pour comparer une distribution observée à une distribution théorique. Il considère l'ensemble des différences entre distributions cumulées théorique et observée à chaque observation, alors que le test de Kolmogorov et Smirnov ne considère que la plus grande de ces différences : le test de Cramér et von Mises prend donc en compte plus d'information et est réputé être plus puissant. La difficulté vient du fait que la distribution du test statistique, noté W^2, n'est pas déterminée et on doit se reporter à des tables (dont les valeurs critiques ont été déterminées par simulations) qui dépendent de la situation (distribution théorique considérée, paramètres estimés à partir des données). H_0 : la distribution suit la loi considérée.

Test d'Anderson et Darling. Ce test est une modification de celui de Cramér et von Mises qui le rend plus sensible pour les valeurs extrêmes de la distribution (où la densité de probabilité est faible). La distribution du test statistique, noté A^2, n'est également pas déterminée. Tous les autres commentaires faits au sujet du test de Cramér et von Mises s'appliquent ici.

Tests sur effectifs

Test de Pearson. Les données sont composées d'effectifs (donc des entiers) de plusieurs catégories que l'on peut disposer dans un tableau dit de contingence à deux dimensions. Notons c le nombre de colonnes de ce tableau et l son nombre de lignes. Le test statistique suit une loi du χ^2. Deux situations sont couramment rencontrées : `chisq.test`

- comparaison de plusieurs séries d'effectifs observés, H_0 : les probabilités d'être dans une catégorie sont les mêmes pour toutes les séries (parfois appelé test d'homogénéité du tableau de contingence), le nombre de degrés de liberté du test est $(c-1)(l-1)$;
- comparaison d'une série observée à une série théorique (donc $c = 2$), H_0 : les probabilités d'être dans les différentes catégories de la série observée sont égales à celles de la série théorique, le nombre de degrés de liberté du test est $l - 1$ (aucun paramètre de la série théorique n'étant estimé à partir des données).

Le test de Pearson est peu performant (très faible puissance) si au moins un des effectifs du tableau est ≤ 5.

Test de Fisher. Ce test s'applique pour tester l'homogénéité d'un tableau de contingence à deux dimensions mais est beaucoup plus performant que le test de Pearson pour les petits effectifs. Le test de Fisher ne calcule pas de statistique mais directement les probabilités de chacun des événements possibles qui sont en défaveur de l'hypothèse nulle : la somme de ces probabilités est donc le risque de première espèce (probabilité de rejeter H_0 si elle est vraie). Ce test requière donc des calculs relativement intensifs qui deviennent très longs pour des effectifs importants (mais dans ce cas le test de Pearson retrouve ses performances). À noter que la plupart des logiciels proposent le test de Fisher uniquement pour des tableaux avec $c = 2$ et $l = 2$ (R et S-PLUS n'ont pas ces limites mais les calculs deviennent assez longs pour des tableaux étendus). `fisher.test`

Test de McNemar. Teste la symétrie d'un tableau de contingence à deux dimen- `mcnemar.test` sions. H_0 : la probabilité d'être dans la ligne i et la colonne j du tableau est égale à celle d'être dans la ligne j et la colonne i. Le test statistique suit une loi du χ^2.

Test de Cochran, Mantel et Haenszel. Ce test considère un tableau de contingence à trois dimensions. H_0 : les tableaux de contingence des deux premières dimensions sont homogènes indépandemment d'une possible influence du troisième facteur. Le test statistique suit une loi du χ^2. Ce test est valide s'il n'y a pas d'interaction entre les trois facteurs. `mantelhaen.test`

Emmanuel Paradis
paradis@isem.univ-montp2.fr
3 octobre 2004