



Tests non-paramétriques, cadre général.

cours VI, P. Coquillard, 2014

L'ensemble des tests étudiés préalablement (cours I à V) sont des tests dits « paramétriques » car ils nécessitent de calculer en général deux paramètres : moyenne et variance (position et dispersion) Ex : test de Student. Ils supposent que la variable aléatoire étudiée est distribuée selon la loi normale (de paramètres (μ, σ)).

Cette condition est vérifiée dans une très grande majorité de cas en biologie. Mais pas toujours. Il convient d'employer une méthode rationnelle pour choisir le test approprié. Pour cela on considère les critères concernant:

- la façon dont l'échantillon a été réalisé,
- la nature de la population de laquelle a été tiré l'échantillon et
- la nature des mesures réalisées.

A chaque test statistique est associé un modèle et des contraintes de mesure. Ce test n'est alors valide que si les conditions imposées par le modèle et les contraintes sont respectées. Il est difficile de dire si les conditions d'un modèle sont remplies, et le plus souvent nous nous contentons d'admettre qu'elles le sont.

- Moins les exigences imposées par le modèle sont nombreuses et restrictives, plus les conclusions que l'on tire sont générales (et peu fiables).
- Les plus puissants sont ceux qui ont les hypothèses les plus strictes. Si ces hypothèses sont valides, ces tests sont alors les mieux à même de rejeter H_0 quand elle est fautive et de ne pas rejeter H_0 quand elle est vraie.

Exemple : Le test de Student (dit test t , test paramétrique) est l'un des tests parmi les plus puissants. Mais avant d'accepter les conclusions on doit vérifier les conditions suivantes :

1. Les observations doivent être indépendantes. La sélection d'une observation pour un échantillon ne doit pas biaiser les chances de sélectionner une autre observation pour cet échantillon.
2. Les observations doivent être extraites de populations distribuées selon des lois normales.
3. Ces populations doivent avoir la même variance (condition d'homoscédasticité). La validité de cette hypothèse peut être contrôlée à l'aide des tests d'égalité de variance.
4. Les variables doivent être mesurées dans une échelle de mesure autorisant des opérations arithmétiques.

A part l'hypothèse d'homoscédasticité, qui peut être testée, les autres hypothèses sont considérées comme vraies (dans la plupart des cas). Il faut y ajouter le problème de la taille de l'échantillon. Plus il est petit, plus difficile est la vérification de sa normalité éventuelle, moins le test est puissant (quel qu'il soit). En pratique, il existe des tests de Student pour petits échantillons et l'ANOVA est assez robuste.

Si l'on ne peut pas considérer que la loi de X est normale ou bien si les données ne peuvent faire l'objet d'opérations arithmétiques (données qualitatives ou comptages) on adoptera des tests non paramétriques¹. (voir la page http://sites.unice.fr/coquillard/tableau_stats.htm, partie droite de l'arbre). Le cas où la variable est une mesure mais de distribution non normale est particulier (test de randomisation).

Les tests non paramétriques sont des tests de comparaison de moyennes.

1. Si les échantillons peuvent être considérés indépendants, on applique le test de la médiane de **Mann et Whitney** pour 2 échantillons, et de **Kruskal et Wallis** (ANOVA non paramétrique) pour un nombre quelconque d'échantillons.
2. Lorsque on a à faire à deux échantillons appariés (= non indépendants), on applique le test des **rangs signés Wilcoxon**.

Les tests non paramétriques ne nécessitent pas d'estimation de la moyenne et de la variance. En fait, ils n'utilisent même pas les valeurs x_i recueillies dans les échantillons, mais seulement leur rang dans la liste ordonnée de toutes les valeurs.

¹ Mais notez bien qu'un *glm* permet de transformer des données poissoniennes (comptage) au moyen d'une fonction de lien (log), et effectue ensuite une régression sur des données « normales »...