

Principe des tests d'hypothèse

1 ère année médecine 2016/2017

Dr M LAOUSSATI

Plan du cours

- I- Introduction
- II- Définitions
- III- Les différentes étapes de l'analyse statistique
- V- Bibliographie

I- Introduction

La finalité d'une étude statistique est de tirer des conclusions sur la population à partir de l'étude d'un échantillon issu de cette population.

L'échantillon n'est qu'une image ponctuelle, les observations qu'on en tire n'ont aucun intérêt en tant que telles, et n'ont de valeur que si on les extrapole à la population d'où est issu l'échantillon.

Lorsqu'on effectue une comparaison entre deux ou plusieurs séries de données, on observe toujours une différence, plus ou moins grande entre les paramètres mesurés.

Le but du test statistique est de déterminer si la différence observée est simplement due au hasard, c'est à dire aux fluctuations d'échantillonnage, ou si au contraire la différence observée est bien réelle.

L'intérêt majeur d'un test statistique est donc de réaliser une économie énorme de moyens en permettant de déceler des différences sur un nombre réduit d'observations.

II- Définitions

Un test statistique d'hypothèse est définie comme une procédure qui permet, avec un risque d'erreur connu, d'effectuer un choix entre deux hypothèses complémentaires (H_0 et H_1) au vu des observations réalisées sur un échantillon.

On distingue:

- Les tests de comparaison
- Les tests de liaison

a) Les tests de comparaison servent à comparer des séries de données entre elles.

Deux situations sont possibles:

1- Comparer un échantillon observé à une population de référence (théorique).

Exemple: la distribution par sexe observée dans une maternité est-elle conforme à la distribution nationale?.

2- Comparer deux ou plusieurs échantillons entre eux.

Exemple: comparer la fréquence du paludisme dans deux régions d'Afrique.

b) Les tests de liaison:

Dans ce type de problème on se demande s'il existe une liaison entre une ou plusieurs variables étudiées sur un échantillon.

Exemple: Etudier le lien entre la consommation de glace au chocolat et la survenue de la gastroentérite.

Les différentes étapes de l'analyse statistique :

Règle générale:

Quelle que soit la nature d'un test, son principe et sa chronologie sont toujours les mêmes. Il s'agit d'une démarche hypothético-déductive.

1^{ère} étape: Formulation des hypothèses

Formulation littérale puis statistique des hypothèses :

H₀ (hypothèse nulle), et **H₁** (hypothèse alternative)

H₀ : consiste à poser à priori l'hypothèse que les paramètres ou les distributions des populations d'où sont issus les échantillons étudiés sont identiques.

Proposer l'hypothèse nulle, c'est supposer que la différence observée provient seulement des fluctuations d'échantillonnage.

Exemple: comparaison de deux moyennes

$$- H_0 : \mu_1 = \mu_2$$

μ_1 et μ_2 : moyennes inconnues des deux populations d'où sont tirés les échantillons.

Attention !!! Les hypothèses sont toujours posées à partir des populations et non des échantillons.

H₁: hypothèse qui sera retenue au cas où les résultats du test aboutiraient à rejeter H₀.

Selon le type du problème posé, on distingue:

H₁ bilatérale: on ne cherche pas à connaître le sens de la différence.

H₁ unilatérale : on s'intéresse à un sens particulier de l'inégalité de deux paramètres.

Exemple1:

$$- H_0 : \mu_1 = \mu_2$$

$$- H_1 \text{ bilatérale} : \mu_1 \neq \mu_2$$

$$- H_1 \text{ unilatérale} : \mu_1 < \mu_2 \text{ ou } \mu_1 > \mu_2$$

Exemple2: comparaison de deux traitements

$$H_0 : A = B$$

$$H_1 \text{ bilatérale: } A \neq B$$

$$H_1 \text{ unilatérale} : A > B \text{ ou } A < B$$

2^{ème} étape: Le choix du test statistique

Le choix du test statistique : en fonction des données du problème et de la variable(s) étudiée (s):

- La nature de la variable (qualitative ou quantitative)
- Le nombre d'échantillons d'observations: si plus d'un échantillon, sont-ils indépendants ?
- La taille des échantillons: grands ($n \geq 30$) ou petits ($n < 30$)

Exemples:

Nature des 2 variables	Le test statistique approprié
2 qualitatives	Test du Chi-deux
1 qualitative 1 quantitative	Analyse de la variance (test d'ANOVA) Comparaison de deux moyennes (test de l'écart réduit ou test de Student)
2 quantitatives	Test du coefficient de corrélation Test de la pente de la droite de régression

3^{ème} étape: Les conditions de validité

Vérification des conditions de validité (ou conditions d'application) du test statistique.

Exemple: Les conditions d'application d'un test du chi-deux de Pearson : Tous les effectifs théoriques (attendus) doivent être ≥ 5

4^{ème} étape: Calcul de la statistique

Tous les tests d'hypothèse consistent à :

- calculer une quantité mathématique exprimant l'écart entre les paramètres ou les distributions et,
- confronter cette quantité à un modèle de distribution théorique.

Remarque:

Cette quantité (la valeur expérimentale de la statistique) est calculée à partir d'observations faites sur des échantillons.

5^{ème} étape: Choix du risque α

-Le risque d'erreur de première espèce α :

Est la probabilité de trouver une différence à tort c'est à dire de conclure à une différence entre deux paramètres alors qu'elle n'existe pas.

En d'autres termes, c'est le risque de se tromper en rejetant H_0 , si par malheur H_0 était vraie.

Lorsqu'on réalise un test, on se fixe un seuil au-delà duquel on accepte ce risque. Ce seuil est fixé à priori, habituellement choisi à **5 %**.

6^{ème} étape: Conclusion

Conclusion et prise de décision :

-Si la statistique calculée $<$ la statistique tabulaire, au risque $\alpha = 5\% \rightarrow$

H_0 n'est pas rejetée, la différence n'est pas significative \rightarrow **Fin de l'analyse.**

-Si la statistique calculée \geq la statistique tabulaire, au risque $\alpha = 5\% \rightarrow$

H_0 est rejetée, la différence est significative $\rightarrow \dots$

Degré de signification p (p- value):

Lorsque H_0 est rejetée avec un risque $\alpha = 5\%$, on désire aller plus loin et préciser la limite du risque pris.

On appelle degré de signification p, cette limite à posteriori.

Ce degré indique la probabilité d'avoir rejeté H_0 si on avait fixé le degré de 1^{ère} espèce égal à p au lieu de α .

Remarque:

Si le calcul montre un p à 0,001, cela ne change pas la valeur du risque α que l'on s'était fixé ($\alpha = 5\%$) mais signifie que si la règle été plus dure, par **exemple** un risque $\alpha = 0,001$, le test aurait encore été significatif.

A noter:

Si $p \geq 0,05$ \rightarrow la différence entre les paramètres est non significative (**DNS**)

Si $p < 0,05$ \rightarrow la différence entre les paramètres est significative(**DS**)

Exemple:

« L'enquête porte sur un échantillon représentatif de 2 252 enfants

(1 104 filles et 1 148 garçons) âgés de six à 11 ans scolarisés dans

neuf écoles primaires à Bejaia.

La prévalence du surpoids incluant l'obésité est de 13,1 % (16,4 % chez les

filles vs 10 % chez les garçons, **$p < 10^{-4}$**).

Aucun lien significatif n'est retrouvé entre le niveau d'étude du père et le statut pondéral des enfants.

Quant au type d'habitat, le pourcentage du surpoids est plus élevé lorsque le type d'habitat est individuel que collectif (62,6 % vs 46,1 % ; $p < 10^{-4}$).»

Bibliographie

Revue Santé Publique. 2012 /6 - Vol. 24, pages 561 à 571.

Monbet.V. Tests statistiques. Université de Rennes.2009

Huguier.M, Flahault.A. Biostatistiques au quotidien. Edition Elsevier 2003

Ancelle T. Statistique Épidémiologie. Édition 2002

Bouyer.J. Methodes statistiques. Edition INSERM 2000

Dabis.F, Drucker.J, Moren.A. Épidémiologie d'intervention, chapitre V. Edition Arnette 1992

Friant-Michel.P. Les tests statistiques. Université de Nancy.

Campillo-Gimenez.B. Principes généraux des tests d'hypothèse. Université de Rennes.