

LE TRAITEMENT STATISTIQUE

I. Les types de traitements statistiques en fonction des questions-problèmes (QP)

Dans une recherche, on se pose toujours une ou plusieurs questions qu'il s'agira de résoudre. Le chercheur émet alors une solution qu'il croit possible (hypothèse). Une hypothèse (H) peut être considérée comme une réponse possible à la question-problème. D'autre part, toute question implique au moins une variable qu'il s'agit d'identifier ainsi que ses modalités et le type d'échelle correspondant.

Il existe plusieurs types de questions-problèmes qui se différencient plus ou moins nettement mais qui impliquent des traitements statistiques différents:

A. Recherche de description

Ce type de questions invite le chercheur de décrire les caractéristiques d'une ou plusieurs variables sur un échantillon ou une population. Il s'agira de collecter, présenter, résumer les informations jugées nécessaires et adéquates.

Exemples:

QP1: Quel est le poids moyen de mon échantillon ?

Hypothèse: Le poids moyen se situera en dessous d.: 100 kg.

QP2: Quelle est la répartition de mon échantillon au niveau de la nationalité ?

Hypothèse: Il y aurait davantage de Tunisiens, puis des Maghrébins et enfin quelques autres nationalités.

QP3: Quelle est la répartition observée des sujets en fonction de leur poids par rapport au sexe ?

Hypothèse: Il y aura plus d'hommes qui pèseront plus lourds.

B. Recherche d'estimation

Il s'avère souvent impossible d'examiner une population entière. On recourt alors à un échantillon que l'on espère représentatif de cette population.

Le problème de savoir si les caractéristiques d'un échantillon observé correspondent bien aux caractéristiques de la population de référence est un problème d'estimation.

C. Recherche de comparaison

Lorsque dans un échantillon, nous avons des groupes par rapport à une variable, il est légitime de se demander si ces groupes sont similaires ou non. Autrement dit, les résultats obtenus dépendent-ils de la formation des groupes ?

D. Recherche de liaison

Si plusieurs variables différentes sont observés sur un même échantillon, nous cherchons à étudier la relation (la liaison) entre ces variables. Si deux variables s'avèrent reliées, ce sont des variables interdépendantes, c.à.d, l'une peut expliquer l'autre.

II. Les procédures avant tout traitement statistique

Les procédures expérimentales avant tout traitement statistique consiste à :

1. poser une question de recherche
2. émettre une hypothèse
3. élaborer une procédure expérimentale pour tester une hypothèse

La suite de la procédure consiste à :

1. constituer un ou de plusieurs échantillons
2. choisir des procédures statistiques
 - définir l'hypothèse nulle
 - choisir le seuil de signification
 - choisir le test

III. Les mesures à prendre avant tout traitement statistique

Pour choisir un test statistique, il est nécessaire de définir:

1. les types de variables (qualitatives / quantitatives)
2. le nombre de variables
3. la taille de l'échantillon
4. la loi de répartition (normale ou non)
5. les mesures répétées ou non (nombre de facteurs)

IV. Le recueil des données

Les données sont rassemblées dans un tableau à double entrée comprenant autant de lignes que d'unités d'observation indépendantes (les sujets d'étude); autant de colonnes que de "variables" ou de mesures caractérisant chacune des unités d'observation.

Le tableau ainsi constitué se nomme un tableau de codage des données

- Les lignes de du tableau représentent les unités d'observation.

Les observations sont réalisées sur un nombre limité d'unités ou d'individus dans le but de tirer des conclusions applicables à une population entière. Ce nombre limité d'unités d'observation forme l'échantillon.

- Les colonnes du tableau de codage représentent les variables.

Une variable, c'est une quantité ou qualité susceptible de fluctuations ou encore une grandeur à laquelle on peut attribuer plusieurs valeurs différentes. Au minimum, une variable a toujours deux modalités ou deux fluctuations. Ainsi, la variable sexe possède deux modalités: fille et garçon.

Exemple : La variable « degré de concentration » peut être estimée sur une échelle de valeur à cinq degrés; elle a alors cinq modalités (très faible, faible, moyen, fort, très fort).

Les variables placées en colonne dans le tableau de codage seront accompagnées d'une légende de codage reprenant l'ensemble de l'information nécessaire pour pouvoir reconnaître la signification précise de n'importe quel code dans n'importe quel endroit du tableau.

Exemple : La légende mentionne le numéro de la variable, son énoncé explicite et un code pour chaque modalité. V_{10} : sexe, 1= homme, 2 = femme

V. Le choix du test statistique

❖ Pour les liens entre variables qualitatives, le choix du test dépend de la taille de l'échantillon :

- pour un échantillon de taille normale (effectifs calculés > 5) : test de khi 2
- pour un échantillon de taille réduite (effectifs calculés > 3) : test de khi2 corrigé (correction de Yates)
- pour un échantillon de taille très réduite (effectifs calculés < 3) : le « test exact »

❖ **Pour les liens entre une variable qualitative et une variable quantitative le choix du test dépend :**

➤ **s'il s'agit de comparaison de deux moyennes :**

- comparaison de deux moyennes observées
- comparaison d'une moyenne observée à une moyenne théorique

- **si les effectifs sont suffisamment grands ($n > 30$) où la loi de répartition est normale (faire éventuellement un test de normalité) : c'est le test t de Student qui devrait être retenu.**

options du test :

- comparaison d'une moyenne observée et d'une moyenne théorique (one population) ou de deux moyennes observées (two populations)

- mesures appariées (paired) ou non appariées (unpaired)

- comparaison unilatérale (one-tailed) ou bilatérale (two-tailed)

- **Les effectifs sont faibles et la répartition n'est pas normale (faire éventuellement un test de normalité) : ce sont les tests non paramétriques qui devront être retenus.**

Les options :

- **séries non appariées :**

test W de Wilcoxon

test U de Mann et Whitney

test C1 de Fisher-Yates-Terry

- **séries appariées :**

test T de Wilcoxon

➤ **s'il s'agit de comparaison de plusieurs moyennes (> 2)**

si la loi de répartition de probabilité est normale pour la variable mesurée (faire éventuellement un test de normalité) : **c'est l'analyse de la variance (ANOVA) qui devrait être retenue.**

options :

ANOVA à plusieurs facteurs

tests « post-hoc » :

Méthode de Bonferonni (test t) : recommandé pour un usage général, si les comparaisons ne sont pas trop nombreuses.

Méthode de Tukey (test t) : à utiliser lorsque toutes les comparaisons paire par paire sont intéressantes.

Méthode de Dunnet : à utiliser lorsque l'on compare le groupe contrôle avec les autres groupes, mais pas les autres groupes en eux.

Méthode de Sheffé (test F) : à utiliser pour les comparaisons compliquées

- **si la loi de répartition de probabilité n'est pas normale pour la variable mesurée (faire éventuellement un test de normalité) : ce sont les tests non paramétriques qui devront être retenus.**

Les tests de normalité proposées :

Martinez-Iglewicz

(Kolmogonov-Smirnov)

Agontino Omnibus

Agostino-Skewness (si $n > 8$)

Agostino-Kurtosis (si $n > 20$)

❖ Pour les liens entre deux variables quantitatives

- une des deux distributions liées au moins est normale avec une variance constante : c'est le test de corrélation (test de Bravais-Pearson) ou le test de régression qui détermine si la pente est statistiquement significative de 0.
- si aucune des variables liées n'est normale et de variance constante (petits échantillons) : c'est le test non paramétrique de corrélation des rangs (test de Spearman)