Comparaison de deux moyennes

TALHI.R
Service de Biostatistique, Faculté de Médecine d'Oran

Plan du cours

I- Grands échantillons → Test de l'écart réduit (Z = Ɛ)
1-Comparaison d'une moyenne observée à une moyenne théorique
2-Comparaison de deux moyennes observées

II- Petits échantillons → Test de Student (T)
1-Comparaison d'une moyenne observée à une moyenne théorique
2-Comparaison de deux moyennes observées

Test de l'écart réduit

Le test Z sert à comparer des paramètres en testant leurs différences (Δ). La division de Δ par son écart type suit une loi Z normale centrée réduite de moyenne 0 et d'écart type 1

- **Utilisé pour comparer :**
  - Une moyenne observée à une moyenne théorique
  - Deux moyennes observées
  - Deux moyennes de deux séries appariées
  - .....

Comparaison d'une moyenne observée à une moyenne théorique

On compare une moyenne observée dans un échantillon à une moyenne connue dans la population de référence.

- **Hypothèses:**
  - H0 : M = μ
  - H1 : M ≠ μ

μ : moyenne théorique connue de la population de référence
M : moyenne inconnue de la population d'où est issu l'échantillon

- **Conditions d'applications:**

Taille de l'échantillon ≥ 30

- **Calcul de la statistique:**

$$\boxed{z = \frac{|m - \mu|}{\frac{s}{\sqrt{n}}}}$$

– μ : moyenne théorique connue de la population de référence
– m : moyenne observée de l'échantillon
– s : écart type de l'échantillon
– n : effectif

- On fixe le seuil de signification au risque **α = 5%**
- On compare le Z calculé au **Z de la table = 1,96**
- **Conclusion:**

$Z_o < 1,96 \rightarrow H_o$ non rejetée → M n'est pas significativement différente de μ
$Z_o \geq 1,96 \rightarrow H_o$ est rejetée → M diffère significativement de μ

Exemple n°1:

- Lors d'une enquête sur la durée de sommeil des enfants de 2 à 3 ans dans un département français, on a trouvé une moyenne du temps de sommeil par nuit de 10,2 heures dans un groupe de 40 enfants. L'écart type est 2,1 heures.
- La moyenne du temps de sommeil est de 11,7 heures chez les enfants de cet âge.
- La durée de sommeil des enfants de ce département diffère-t-elle du temps de sommeil des enfants de cet âge?

Solution n°1:

1-**H_0** : les enfants de ce département dorment autant que ceux de la population
H_1: la durée de sommeil des enfants de ce département est différente

2-Choix du test → **test Z**

3-Conditions d'applications → **40 > 30**

4- Calcul de la statistique:
$\mathbf{Z_o}$ = (11,7 – 10,2)/(2,1/√ 40) = **4,5**

5- α = 5% et Z tabulaire = 1,96

6- Conclusion → 4,5 > 1,96 →**On rejette H_0** → DS
La population des enfants examinés présente un temps de sommeil **significativement différent** de la population générale.

Comparaison de deux moyennes observées

On veut comparer **les moyennes observées** dans deux échantillons

- **Hypothèses**
  - $H_0 : \mu_1=\mu_2$
  - $H_1 : \mu_1 \neq \mu_2$

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

- **Conditions d'application :**
  effectif de chaque échantillon ≥ 30

- **Calcul de la statistique :**

$$\boxed{z = \frac{|m_1 - m_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

– m_1 et m_2 : moyennes observées des 2 échantillons
– s^2_1 et s^2_2 : variances des 2 échantillons
– n_1 et n_2 : effectifs des 2 échantillons

- On fixe le seuil de signification au risque **α = 5%**
- On compare le Z calculé au **Z de la table = 1,96**
- **Conclusion:**

$Z_o < 1,96 \rightarrow H_o$ non rejetée $\rightarrow \mu_1$ n'est pas significativement différent de μ_2
$Z_o \geq 1,96 \rightarrow H_o$ est rejetée $\rightarrow \mu_1$ diffère significativement de μ_2

Exemple n°2:

- On désire comparer la pression artérielle diastolique d'un groupe de sujets sains et d'un groupe de sujets atteints de drépanocytose. Une étude donne les résultats suivants :

	Effectif (n)	Pression artérielle diastolique	Variance (s^2)
Sujets sains	88	70,1	10,8
Sujets drépanocytaires	85	61,8	6,9

- La pression artérielle des sujets drépanocytaires diffère t- elle de celle des sujets sains ?

Solution n°2:

1-H_0 : les pressions artérielles sont identiques
H_1 : la pression artérielle est différente chez les sujets drépanocytaires

2- Choix du test → test Z

3-Conditions d'application → **85 > 30 et 88 > 30**

4-Calcul de la statistique:
$Z_o = (|70,1 - 61,8|) / \sqrt{(10,8/88) + (6,9/85)} =$ **18,4**

5- On fixe α = 5% , Z tabulaire = 1,96

6- Conclusion → 18,4 > 1,96 → on rejette H_0 → **DS**
La pression artérielle des sujets drépanocytaires est significativement différente de celle des sujets sains.

Test T de Student

Lorsque la taille des échantillons est faible **(n<30)**
le rapport entre les différences de leurs moyennes et l'écart type ne suit pas une loi normale centrée réduite Z → On utilise alors le test T de Student.

• **Le test de Student sert à comparer :**
– Une moyenne observée à une moyenne théorique
– Les moyennes observées de 2 petits échantillons
_........

Conditions d'application :

– Utilisable si petits effectifs
– Mais la distribution de la variable dans les populations doit être normale
– **Et** les populations doivent avoir des variances identiques
  • Soit on le sait
  • Soit on le teste (test F de comparaison de 2 variances)

Utilisation de la table T

- La table de T est plus difficile à utiliser que la table de Z
- Il y a autant de table T que de degré de liberté

- **Calcul du ddl:**
  – Pour 1 échantillon : ddl = n-1
  – Pour 2 échantillons : ddl = (n1 - 1) + (n2 - 1)

comparaison d'une moyenne observée à une moyenne théorique

On compare une moyenne observée dans un échantillon de petite taille à une moyenne connue dans une population de référence

- **Hypothèses:**
  - $H_0 : M=\mu$
  - $H_1 : M\neq \mu$

μ : moyenne théorique connue de la population de référence
M : moyenne inconnue de la population d'où est issu l'échantillon

- **Condition d'application :**

La distribution de la variable doit être supposée normale dans la population d'où est issu l'échantillon

- **Calcul de la statistique:**

$$\boxed{t = \frac{|m-\mu|}{\frac{s}{\sqrt{n}}} \text{ avec ddl} = n - 1}$$

- μ : moyenne théorique connue de la population de référence
- m : moyenne observée de l'échantillon
- s : écart type de l'échantillon
- n : effectif
- ddl : degré de liberté

- On fixe le seuil de signification au risque α **= 5%**
- Calcul du ddl → **ddl = n-1**
- **Conclusion:**

$t_o < t_\alpha$ (tabulaire) → H_o non rejetée → M n'est pas significativement différente de μ

$t_o \geq t_\alpha$ (tabulaire) → H_o est rejetée → M diffère significativement de μ

Exemple n°3:

- Dans un échantillon de 18 sujets suspects d'être atteints de trypanosomiase, on mesure la quantité de protéines dans le liquide céphalorachidien. On trouve dans ce groupe une protéinorachie moyenne de 460 mg/l avec un écart type de 280 mg/l.
  Dans la population générale, la protéinorachie est en moyenne de 300 mg/l.

- On se demande si ce groupe de sujet présente une protéinorachie différente de la normale ?

Solution n°3:

H_0: la protéinorachie des sujets atteints de trypanosomiase ne diffère pas de celle de la population générale

H_1: la protéinorachie des sujets atteints de trypanosomiase est différente de celle de la population

n < 30 → Test T

• **Condition d'application** : on suppose que la protéinorachie est distribuée normalement chez les sujets atteints de trypanosomiase

• T_o = (460 – 300) / (280 / √18) = **2,4 et ddl =17**

• T_α (tabulaire) pour 17 ddl = **2,11**

Conclusion → 2,4 > 2,11 → on rejette H_0 → DS

La protéinorachie des sujets atteints de trypanosomiase est **significativement différente** de celle de la population.

comparaison de 2 moyennes observées

On veut comparer les moyennes dans 2 échantillons de petite taille (au moins un échantillon inférieur à 30).

• **Hypothèses:**

– $H_0 : \mu_1=\mu_2$

– $H_1 : \mu_1 \neq \mu_2$

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

• **Conditions d'application :**

– Les distributions de la variable dans les populations d'où sont tirés les échantillons doivent être normales

– Les variances des deux populations d'où sont tirés les échantillons doivent être égales

• **Calcul de la statistique (T_o):**

Estimation de la **variance commune** aux deux échantillons:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Ecart type de la différence $\Delta = \mu_1 - \mu_2$ par:

$$s_d = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$

$$t = \frac{|m_1 - m_2|}{s_d} \text{ avec ddl} = n_1 + n_2 - 2$$

– m_1 et m_2 : moyennes observées des 2 échantillons

– s^2_1 et s^2_2 : variances des 2 échantillons

– n_1 et n_2 : effectifs des 2 échantillons

– ddl : degré de liberté

• **Conclusion:**

$t_o < t_\alpha$ (tabulaire) → H_o non rejetée → μ_1 n'est pas significativement différent de μ_2

$t_o \geq t_\alpha$ (tabulaire) → H_o est rejetée → μ_1 diffère significativement de μ_2

Exemple n°4:

- On a mesuré un marqueur biologique chez 2 séries de sujets, l'une composée de sujets sains, l'autre de sujets atteints d'hépatite alcoolique. L'étude a trouvé les résultats suivants:

	Effectif (n)	Moyenne du marqueur (g/l)	Ecart type
Sujets sains	15	1.6	0.19
Sujets alcooliques	12	1.4	0.21

- On veut comparer les 2 populations.

Solution n°4:

Ho: la valeur moyenne du marqueur est identique dans les 2 populations
H1: la valeur moyenne du marqueur est différente chez les sujets atteints d'hépatite alcoolique

n < 30 → test de T
Condition d'application : on suppose que :
- le marqueur se distribue normalement dans les 2 populations
- Les variances des 2 populations sont égales

Calcul de la statistique:
$S^2 = [(15\text{-}1) \times (0,19)^2 + (12\text{-}1) \times (0,21)^2] / (15 + 12 - 2) =$ **0,04**
$S_d = \sqrt{(0,04 / 15) + (0,04 / 12)} =$ **0,077**

$t_o = (1,6 - 1,4) / 0,077 =$ **2,60**

ddl = 15+12-2 = 25

T_α (tabulaire) pour 25 ddl = **2,06**

Conclusion:
2,6 > 2,06 : on rejette H_0 → DS
Les malades atteints d'hépatite alcoolique présentent une valeur du marqueur **significativement différente** de celle des sujets sains.

Les références:

Schwartz D. Méthodes statistiques.1992
Bouyer J. Méthodes statistiques.1996
Ancelle T. Statistique Épidémiologie. Édition 2002
Abrouk S. Biostatistique. INSP octobre 2005
Bayat S. Introduction à la Biostatistique. Université de Rennes.2009-2010