Comparaison de deux moyennes

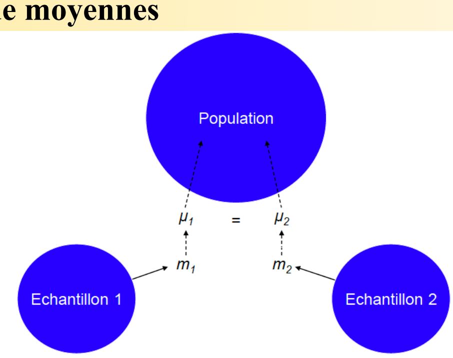
0

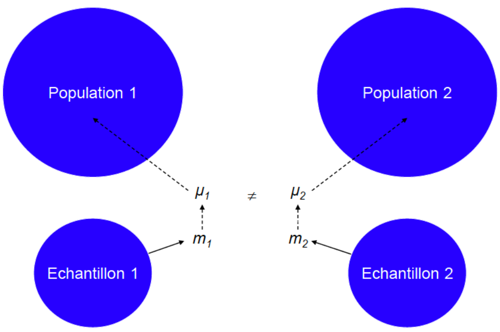
0

Comparaison de moyennes

- *Test d'homogénéité*

H0 : Les deux échantillons proviennent de la même population





H1 : Les deux échantillons proviennent de deux populations distinctes

1

1

Comparaison de moyennes

- ***Test d'homogénéité***

  - Comparaison de moyennes de séries **indépendantes**
  - Echantillons A et B peuvent être d'effectifs différents.

Sujets de l'échantillon A ne sont pas les mêmes que ceux de l'échantillon B.

  - Peut être réalisé à l'aide de trois types de test :
    - ✓ Test z - Loi de Gauss (n > 30)
    - ✓ Test *t* - Loi de Student (n < 30)
    - ✓ Test de Mann-Whitney (n > 10)

✓ **Test *z* - Loi de Gauss (n > 30)**

$$z = \frac{m_A - \mu_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

- Calcul de la variance ➔ $s^2 = \sum_{i=1}^{n} \frac{(x_i - m)^2}{n-1}$

2

2

Comparaison de moyennes

Exemple

Nous souhaitons comparé la concentration en vit D dans le sang entre deux populations :
- non fumeurs (n_1= 97 ; m_1 =23,6 ng/ml ; s_1 =8.3)
- et de fumeurs (n_2= 85 ; m_2= 20,9 ng/ml ; s_2=7.6).

La moyenne de la concentration sanguine en vit D est-elle réellement différente entre population de fumeurs et la population de non-fumeurs ?

Solution

1) Formuler les hypothèses :
- H_0 : La moyenne de la concentration sanguine en vit D n'est pas différente entre les deux populations ($m_1 = m_2$) ;
- H_1 : La moyenne de la concentration sanguine en vit D est différente entre les deux populations ($m_1 \neq m_2$)

2) Calcul de la statistique Z :

$$z = \frac{m_A - \mu_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{23.6 - 20.9}{\sqrt{\frac{(8.3)^2}{97} + \frac{(7.6)^2}{85}}} = 2.29$$

3

3

Comparaison de moyennes

3) Décision :

- ✓ En comparant les quantiles : z = 2,29 est supérieur à la valeur seuil qui est de 1,96.
  En comparant les probabilités : p = 2% (table) est inférieur à la valeur seuil de 5%.

NB. Comme les tables ne donnent pas des p précises, on peut effectuer une commande sur Excel pour trouver la valeur de p précisément.

- ✓ $z_{\text{observé}} > z_{\text{seuil}}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

✓ Test *t* - Loi de Student (n < 30)

La différence : ici ! on calcul une **variance commune** s^2.

$$t = \frac{m_A - \mu_B}{s^2\left(\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right)}$$

- Calcul de la variance ➔ $s^2 = \dfrac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A + n_B) - 2}$

t suit une loi de Student à $(n_1 + n_2) - 2$ degré de liberté

4

4

Comparaison de moyennes

Condition d'utilisation :

- ✓ Distribution des populations soit normale.
- ✓ **Indépendance** des deux échantillons
- ✓ Variance pas trop différente : le rapport variance1/variance2 doit être inférieur à trois.

Exemple

Nous cherchons à comparer le taux du cholestérol LDL (g/l) entre :

- une population non-traité (n_1= 15 ; m_1 =1,81 ng/ml ; s_1 = 0,50)
- une population traité (n_2 = 12 ; m_2= 1,41 ng/ml ; s_2 =0,39).

La moyenne du taux du cholestérol LDL (g/l) est-elle réellement différente entre population non-traité et la population traité ?

Solution

1) Formuler les hypothèses :

- H_0 : La moyenne du taux du cholestérol LDL (g/l) est n'est pas différente entre la population non-traité et la population traité ($m_1 = m_2$) ;
- H_1 : La moyenne du taux du cholestérol LDL (g/l) est différente entre la population non-traité et la population traité ($m_1 \neq m_2$)

5

5

Comparaison de moyennes

2) Calcul de la statistique t :

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - n_B) - 2} = \frac{(15-1)(0.50)^2 + (12-1)(0.39)^2}{(15-12)-2} = 0.21$$

$$t = \frac{m_A - \mu_B}{s^2\left(\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right)} = \frac{1.81 - 1.41}{0.21\left(\sqrt{\frac{1}{15} + \frac{1}{12}}\right)} = 2.27$$

$$ddl = (n_A - n_B) - 2 = (15 - 12) - 2 = 25$$

3) Décision :

- ✓ Lire de la table ddl = 25 avec un α = 0.05 ------- $z = 2.06$
- ✓ En comparant les quantiles $z = 2{,}27$ est supérieur à la valeur seuil qui est de 2,06.

- ✓ $z_{\text{observé}} > z_{\text{seuil}}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

6

6

Comparaison de moyennes

- ***Test sur séries appariées***

  - Comparer des données qui sont « liés ».
  - Sont de même taille ($n_1 = n_2$).
  - Permet de tenir compte du lien qui existe entre ces deux séries

Ex. On parle de données appariées lorsque les deux échantillons considérés sont formés d'individus, pour lesquels on a fait deux mesures d'une même quantité, généralement avec *un écart temporel et/ou après l'occurence d'un évènement.*

- Peut être réalisé à l'aide de trois types de test :
  - ✓ Test z - Loi de Gauss (n > 30)
  - ✓ Test *t* - Loi de Student (n < 30)
  - ✓ Test de Mann-Whitney (n > 10)

✓ **Test *z* - Loi de Gauss (n > 30)**

$$z = \frac{m_d}{\sqrt{\frac{s_d^2}{n}}}$$

✓ **Test *t* - Loi de Student (n < 30)**

Calcule ddl ➔ $t = \frac{m_d}{\sqrt{\frac{s_d^2}{n}}}$

7

7

Comparaison de moyennes

✓ **Test z - Loi de Gauss (n > 30)**

Exemple :
Dans une usine chimique réputée pour avoir utilisé des produits dangereux pour la santé, nous avons étudier le volume globulaire moyen VGM (taille des globules rouges) de 30 travailleurs de cette usine avant et trois mois après embauche. On mesuré :

- Avant embauche, une moyenne VGM $m_1 = 94{,}7\ \mu m^3$
- Après embauche, une moyenne VGM $m_2 = 93{,}6\ \mu m^3$

On vous donne n = 30, la moyenne de cette différence de 1,14 μm^3 et un écart-type de 1,73 μm^3

Solution

1) Formuler les hypothèses :

- H_0 : La moyenne VGM avant embauche n'est pas différente de celle après embauche ($m_1 = m_2$) ;
- H_1 : La moyenne VGM avant embauche est différente de celle après embauche ($m_1 \neq m_2$)

2) Calcul de la statistique z :

$$z = \frac{m_d}{\sqrt{\frac{s_d^2}{n}}} = \frac{1.14}{\sqrt{\frac{(1.73)^2}{30}}} = 3.61$$

3) Décision :

✓ En comparant les quantiles : z = 3,61 est supérieur à la valeur seuil qui est de 1,96.
✓ En comparant les probabilités : p = 0.1% (table) est inférieur à la valeur seuil de 5%.

✓ $z_{\text{observé}} > z_{\text{seuil}}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

8

Comparaison de moyennes

✓ **Test t - Loi de Student (n < 30)**

Exemple :
Si on reprend l'exemple précèdent de l'étudie du volume globulaire moyen VGM, mais cette fois-ci avec un effectif de chaque série de 17.

1) Calcul de la statistique t :

$$z = \frac{m_d}{\sqrt{\frac{s_d^2}{n}}} = \frac{1.14}{\sqrt{\frac{1.73}{17}}} = 2.72$$

un $ddl = 17 - 1 = 18$

3) Décision :

✓ Lire de la table ddl = 32 avec un α = 0.05 ------- t = 2.1098
✓ En comparant les quantiles t = 2,72 est supérieur à la valeur seuil qui est de 2,036.

✓ $t_{\text{observé}} > t_{\text{seuil}}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

9

Comparaison de moyennes

- ***Test de conformité***

  - C'est de comparer une de moyenne observée (m) et de variance s^2, à partir d'un échantillon de taille n à une moyenne <u>théorique</u> ou de <u>référence</u> (μ)
  - L'hypothèse du test :
    - $H_0 : \mu = m$
    - $H_1 : \mu \neq m$

On va chercher est-ce que la différence $\Delta = m - \mu$ est **suffisamment éloignée** de zéro pour qu'on rejette hypothèse nulle H_0.

Pour répondre à cette question deux test peuvent être réalisés :
- ✓ Test z - Loi de Gauss (n > 30)
- ✓ Test *t* - Loi de Student (n < 30)

✓ **Test *z* - Loi de Gauss (n ≥ 30)**

$$z = \frac{m - \mu}{\sqrt{\frac{s^2}{n}}}$$

✓ **Test *t* - Loi de Student (n < 30)**

Calcule ddl ➔ $$t = \frac{m - \mu}{\sqrt{\frac{s^2}{n}}}$$

10

10

Comparaison de moyennes

✓ **Test *z* - Loi de Gauss (n > 30)**

On var reprendre l'exemple précèdent de la concentration en vit D dans le sang : dans un échantillon de taille n = 200, nous avons mesuré une moyenne m =22,3 ng/ml et un écart-type s = 8,3.
Nous voulons savoir est-ce que l'échantillon provient-il d'une population de référence de moyenne μ =20,9 ng/ml ?

Solution
On va chercher si la concentration sanguine en vitamine D de l'échantillon est différente de celle de la population de référence :

1) Formulé les hypothèses :

- H0 : La moyenne de la concentration sanguine en vit D observée n'est pas différente de celle de la population de référence ($m = \mu$) ;

- H1 : La moyenne de la concentration sanguine en vit D observée est différente de celle de la population de référence ($m \neq \mu$)

11

11

Comparaison de moyennes

2) Calcul de la statistique z :

$$z = \frac{m - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{22.3 - 20.9}{\sqrt{\frac{(8.3)^2}{200}}} = 2.38$$

3) Décision :

- ✓ En comparant les quantiles : z = 2,38 est supérieur à la valeur seuil qui est de 1,96.
- ✓ En comparant les probabilités : p = 2% (table) est inférieur à la valeur seuil de 5%.

- ✓ $z_{observé} > z_{seuil}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

Donc la moyenne de la concentration sanguine en vit D observée est différente de celle de la population de référence

12

12

- ✓ **Test *t* - Loi de Student (n < 30)**

Exemple :
Si on reprend l'exemple précèdent de l'étudie du volume globulaire moyen VGM, mais cette fois-ci avec un effectif de chaque série de 17.

1) Calcul de la statistique t :

$$t = \frac{m - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{9.4.7\ - 93.6}{\sqrt{\frac{(1.73)^2}{17}}} = 2.62 \qquad \text{un } ddl = (17 + 17) - 2 = 32$$

3) Décision :

- ✓ Lire de la table ddl = 32 avec un α = 0.05 ------- z = 2.0369
- ✓ En comparant les quantiles z = 2,62 est supérieur à la valeur seuil qui est de 2,036.

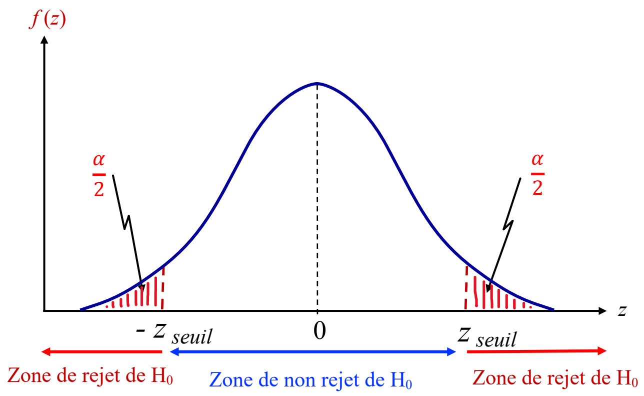
- ✓ $t_{observé} > t_{seuil}$ ➔ On rejet l'hypothèse nulle H_0 et on garde H_1

13

13

Comparaison de moyennes

Test Bilatéral : $H_0 : m = \mu$ **et** $H_1 : m \neq \mu$



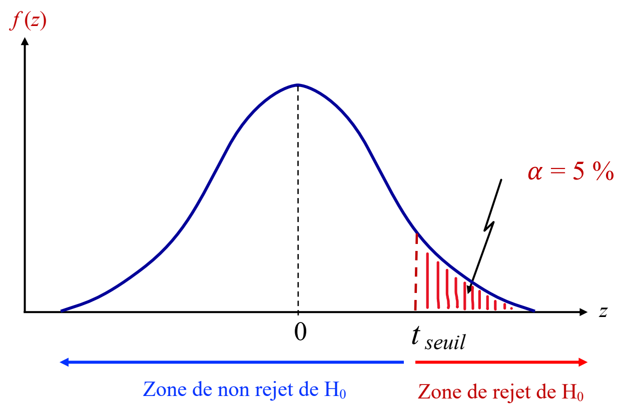
Pour un risque $\alpha = 5$ % ⇨ $z_{seuil} = 1.96$

14

14

Comparaison de moyennes

Unilatéral droit : $H_0 : m = \mu$ et $H_1 : m > \mu$



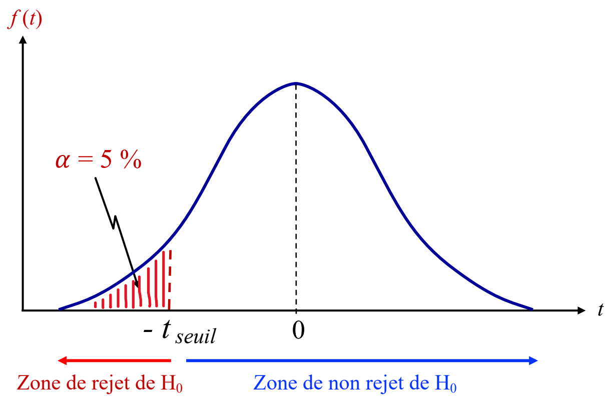
Pour un risque $\alpha = 5$ % ⇨ $z_{seuil} = 1.65$

15

15

Comparaison de moyennes

Unilatéral gauche : $H_0 : m = \mu$ et $H_1 : m < \mu$



Pour un risque $\alpha = 5$ % $\Rightarrow$ $z_{seuil} = - 1.65$