amU
Aix Marseille Université

Test de comparaison de moyennes

Détection des protéines dont l’abondance diffère entre foie et hépatocarcinome

Choix d'un test de comparaison de moyennes

Test d'égalité de moyenne

- Modalités
  - Bilatéral : $H_0: m_1 = m_2$ $H_1: m_1 \neq m_2$
  - Unilatéral :
    - A gauche : $H_0: m_1 \geq m_2$ $H_1: m_1 < m_2$
    - A droite : $H_0: m_1 \leq m_2$ $H_1: m_1 > m_2$
- Principe du test
  - Estimer la différence *d* entre les moyennes des deux population m_1 and m_2
  - Comparer cette différence avec la distribution théorique
- Les variances des deux populations sont généralement inconnues → il faut également les estimer
  - La variance d'une différence est la somme des variances :
    $\sigma_{m1\text{-}m2} = \sigma_{m1} + \sigma_{m2}$
  - La formule pour l'estimation est différente selon qu'on considère que les deux variances sont égales (**hypothèse d'homoscédasticité**) ou pas (**hypothèse d'hétéroscédasticité**)
- La distribution théorique est celle de **Student *(t)***
  - $k = n_1 + n_2 - 2$ degrés de liberté
  - Test bilatéral : risque α partagé entre les deux queues de distribution → on utilise le seuil $t_{1-\alpha/2}$ dans la table de Student
  - Test unilatéral : risque α sur une seule des queues de la distribution *t*

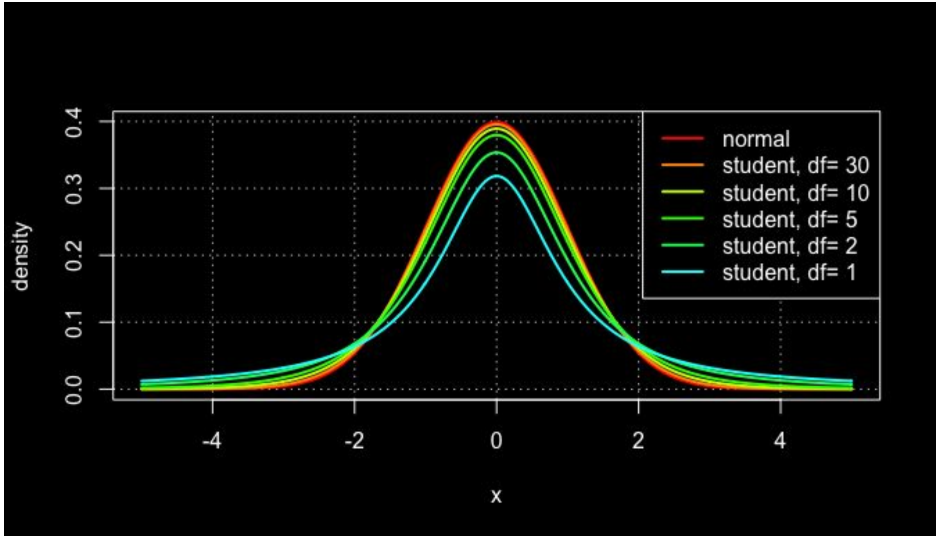
$$t_{obs} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\hat{\sigma}_{m_1 - m_2}}$$

Bilatéral : R(H_0) if $t_{obs} \geq t_{(1-\alpha)/2}$

Unilatéral : R(H_0) if $t_{obs} \geq t_{(1-\alpha)}$

Distribution de Student

- Une famille de courbes caractérisées par 1 paramètre : le nombre de **degrés de liberté** (df)
- La forme varie selon df
- Converge vers une normale quand df $\longrightarrow \infty$



Symboles et formules

Symboles et formules	Description
μ_1, μ_2	Moyennes respectives des populations 1 et 2.
σ_1, σ_2	Ecarts-types respectifs des populations 1 and 2.
N_1, N_2	Tailles (nombre d'individus) des populations 1 et 2.
n_1, n_2	Effectifs (nombre d'individus) des échantillons prélevés sur les populations 1 et 2.
$\bar{x}_1, \bar{x}_2$	Moyennes d'échantillons.
$\delta = \mu_2 - \mu 1$	Différence entre les moyennes des populations.
$d = \hat{\delta} = \hat{\mu_2} - \hat{\mu_1} = \bar{x}_2 - \bar{x}_1$	d = **Taille d'effet**: dans un test de comparaison de moyennes, il s'agit de la différence entre les moyennes d'échantillons, utilisée comme estimateur de δ.
s_1^2, s_2^2	Variances mesurées sur les échantillons.
$\hat{\sigma_p} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$	Ecart-type groupé (*pooled standard deviation*), utilisé comme estimateur de l'écart-type commun des deux populations, en supposant leurs variances égales (hypothèse de travail d'homoscédasticité).
$\hat{\sigma_\delta} = \hat{\sigma_p} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	Erreur standard sur la différence entre moyennes, en supposant que les populations ont la même variance (test de Student).
$t_S = \frac{\hat{\delta}}{\hat{\sigma_\delta}} = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	statistique t de Student
$t_W = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	statistique t de Welch