Introduction

Le test t est un test statistique qui sert à comparer deux moyennes pour voir si la différence entre les deux est statistiquement significative. Il existe trois façons d'appliquer ce test : test t pour échantillons indépendant (plan simple), le test t pour échantillons pairés (plan à mesures répétés) et le test t pour échantillon unique (on compare la moyenne de notre échantillon à la moyenne de la population). Dans ce document, nous allons voir le test t pour échantillons indépendants et nous verrons le test t pour échantillons pairés dans un autre document.

Nous utilisons le test t pour échantillons indépendants si les deux moyennes proviennent deux de groupes distincts. Par exemple, comparer la moyenne à l'examen du groupe A et du groupe B ; comparer l'âge moyen du groupe A et du groupe B, comparer le niveau de bonheur moyen du groupe A et du groupe B, etc. Chaque participant n'appartient qu'à un seul groupe. Si les deux moyenne que vous comparez proviennent du même groupe que vous avez testé deux fois, vous êtes dans un plan à mesures répétées et vous devrez utiliser un test t pour échantillons pairés.

Il est important de mentionner que pour calculer la statistique *t* (le rapport entre les deux moyennes que l'on calcule à l'aide du test t), vous aurez besoins de la variance de chaque groupe que vous allez diviser par la taille des deux groupes. Les deux groupes doivent être de tailles égales. Si les deux groupes ne sont pas de tailles égales, vous allez devoir calculer la variance combinée des deux groupes.

Formules

<u>Si les groupes sont de tailles égales, on calcule la statistique *t*</u>

Moyenne du groupe 1 Moyenne du groupe 2

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Variance du groupe 1 Variance du groupe 2

Taille du groupe 1 Taille du groupe 2

<u>Si les groupes sont de taille inégales, on calcule une variance combiné avec cette formule</u>

Taille de l'échantillon 1

Taille de l'échantillon 2

Somme pondérée des variance

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

Variance de l'échantillon 1

Variance de l'échantillon 2

<u>Puis, on calcule la statistique *t* avec cette formule:</u>

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Variances combinées

*Il s'agit de la même formule que pour les groupes de tailles égales, sauf qu'on utilise les variances combinées à la place de la variance de chaque groupe. Si vos groupes sont de tailles égales, elle donnera le même résultat que la formule de base.

Mise en situation

Miles est musicien de jazz et Johnny est musicien de country. Chacun est convaincu que les musiciens de son style de musique ont un QI plus élevé que les musiciens de l'autre style. Ils font passer un test de QI à tous les musiciens de Jazz et de Country qu'ils connaissent et calcule le QI moyen selon le style de musique.

H: La différence entre nos moyennes est supérieure (ou inférieure) à 0

H_0: La différence entre nos moyennes est de 0

Voici leurs résultats:

	Jazz	**Country**
Moyennes des QI	114	107
Variance	115	110
Tailles du groupe (*n*)	41	32

Calculs

Étape 1: Calculer la variance combiné

Nos groupes ne sont pas égaux, il faut donc calculer la variance combiné. Si vos groupes sont égaux, vous n'avez pas besoin de faire cette étape !

$s_1^2 = 115$ $n_1 = 41$
$s_2^2 = 110$ $n_2 = 32$

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

$$\frac{(41-1)*115+(32-1)*110}{41+32-2}$$

$$\frac{40*115+31*110}{71}$$

$$\frac{4600+3410}{71}$$

$$\frac{8010}{71}$$

$$s_p^2 = 112.82$$

Étape 2: calculer la statistique *t*

$n_1 = 41$

$n_2 = 32$

$\bar{X}_1 = 114$

$\bar{X}_2 = 107$

$s_p^2 = 112.82$

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$t = \frac{(114-107)}{\sqrt{\frac{112.82}{41} + \frac{112.82}{32}}} = \frac{7}{\sqrt{2.75+3.53}} = \frac{7}{\sqrt{6.28}} = \frac{7}{2.51} = 2.79$$

Étape 3: Le degré de liberté

Voici notre formule calculer le degré de liberté avec deux échantillons indépendants:
dl = $n_1 + n_2$ - 2
dl = 41+ 32 - 2
dl = 73 - 2
dl = 71

Étape 4: Table *t*

Comme pour le khi-carré, le test t a une table de distribution, la table t. Normalement, elle se trouve à la fin de votre manuel ou elle vous a été donnée par votre professeur. Vous voyez la table t fournie par Tabachnick et Fidell dans leur manuel (Tabachnick et Fidell, 2021). Vous noterez qu'à partir d'un degré de liberté de 30, on saute directement à 40, puis 60, puis 120, puis l'infini. En fait, à partir d'un certain degré de liberté, la valeur critique change tellement peu que les tables font des sauts. Par exemple, la différence entre la valeur critique avec un degré de liberté de 1 et un degré de liberté de 2 est plus grande qu'entre 2 et 3 et ainsi de suite. D'ailleurs la valeur critique diminue plus entre 2 et 3 qu'entre 60 et 120. Si votre valeur critique ne se trouve pas dans la table que vous utilisez, vous prenez celle au-dessus. Notre valeur critique est de 71, nous prendrons donc celle de 120. Notre seuil de rejet est de .05, notre valeur critique est donc de 1.98.

TABLE C.2 Critical Values of the *t* Distribution for α = .05 and .01, Two-Tailed Test

Degrees of Freedom	.05	.01
1	12.706	63.657
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012
14	2.145	2.977
15	2.131	2.947
16	2.120	2.921
17	2.110	2.898
18	2.101	2.878
19	2.093	2.861
20	2.086	2.845
21	2.080	2.831
22	2.074	2.819
23	2.069	2.807
24	2.064	2.797
25	2.060	2.787
26	2.056	2.779
27	2.052	2.771
28	2.048	2.763
29	2.045	2.756
30	2.042	2.750
40	2.021	2.704
60	2.000	2.660
120	1.980	2.617
∞	1.960	2.576

Source: Adapted from Table 9 in *Biometrika Tables for Statisticians*, vol. 1, 3d ed., edited by E. S. Pearson and H. O. Hartley (New York: Cambridge University Press, 1958).

Si votre statistique *t* est **<u>supérieure</u>** à votre valeur critique, vous pouvez rejeter l'hypothèse nulle.

2.79 > 1.98

C'est le cas pour nous, vous pouvez donc rejeter l'hypothèse nulle et assumer qu'il existe une différence significative entre les QI moyens des deux groupes. Dans notre cas, le QI moyen des musiciens de jazz est significativement supérieur au QI moyen des musiciens de country.
*Note : Il s'agit d'un exemple fictif avec des données totalement inventées en 30 secondes, ne vous sentez pas insulté si vous êtes fan de country.

Signification des symbole

X : Valeur

N : Nombre d'observations total

n : Nombre d'observations d'un groupe

- n_1 : nombre d'observations du groupe 1 ;
- n_2 : nombre d'observations du groupe 2 ;
- etc.

K : Nombre de groupes

Σ : Somme

$\bar{X}$: Moyenne d'un échantillon

μ : Moyenne d'une population

s : Écart-type d'un échantillon

σ : Écart-type d'une population

s^2 : Variance d'un échantillon

σ^2 : Variance d'une population

χ^2 : Khi-carré

A: Fréquence attendue

O: Fréquence observé

L: Ligne

C: Colonne

t : Statistique t (ou score t dans le cas d'une corrélation/régression linéaire)

r : Coefficient de Pearson

Z : Score Z

b : Coefficient de régression

a : Ordonnée à l'origine

$\hat{Y}$: Valeur de la VD qu'on veut prédire l'aide la VI

F : Statistique F

Ouvrages de référence

- Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics: North American Edition* (5th ed.). Sage Edge
- Howell. D.C. (2008). *Méthodes statistiques en sciences humaines* (M. Rogier, V. Yzerbyt, & Y. Bestgen, Trans.). (6th ed.). De boeck. (Original work published 2008)
- Tabachnick, B. G., & Fidell, L. S. (2021). *Using Multivariate Statistics* (7th ed.). Pearson