

Comparaisons de 2 échantillons

Comparaisons de 2 échantillons indépendants

Le test t pour échantillons indépendants (student t test)

Objectif du test

Tester la différence entre les moyennes de deux échantillons indépendants (les deux échantillons sont composés d'éléments non appariés).

Principe

Soit deux échantillons avec n_1 et n_2 éléments respectivement.

Nous voulons savoir si la différence entre \bar{x}_1 et \bar{x}_2 reflète une différence significative des moyennes des populations statistiques dont sont extraits les échantillons, ou si l'écart observé n'est dû qu'aux fluctuations naturelles de l'échantillonnage.

Nous calculerons **une statistique t de Student** à partir des données et nous déterminerons la probabilité de cette valeur à l'aide de la distribution de Student à **$\nu = n_1 + n_2 - 2$ degrés de liberté.**

Cette distribution nous permettra de savoir si la probabilité de rencontrer notre valeur de t sous H_0 est plus grande ou plus petite que notre seuil α . Nous prendrons alors la décision de ne pas rejeter ou de rejeter H_0 .

Règles de décision

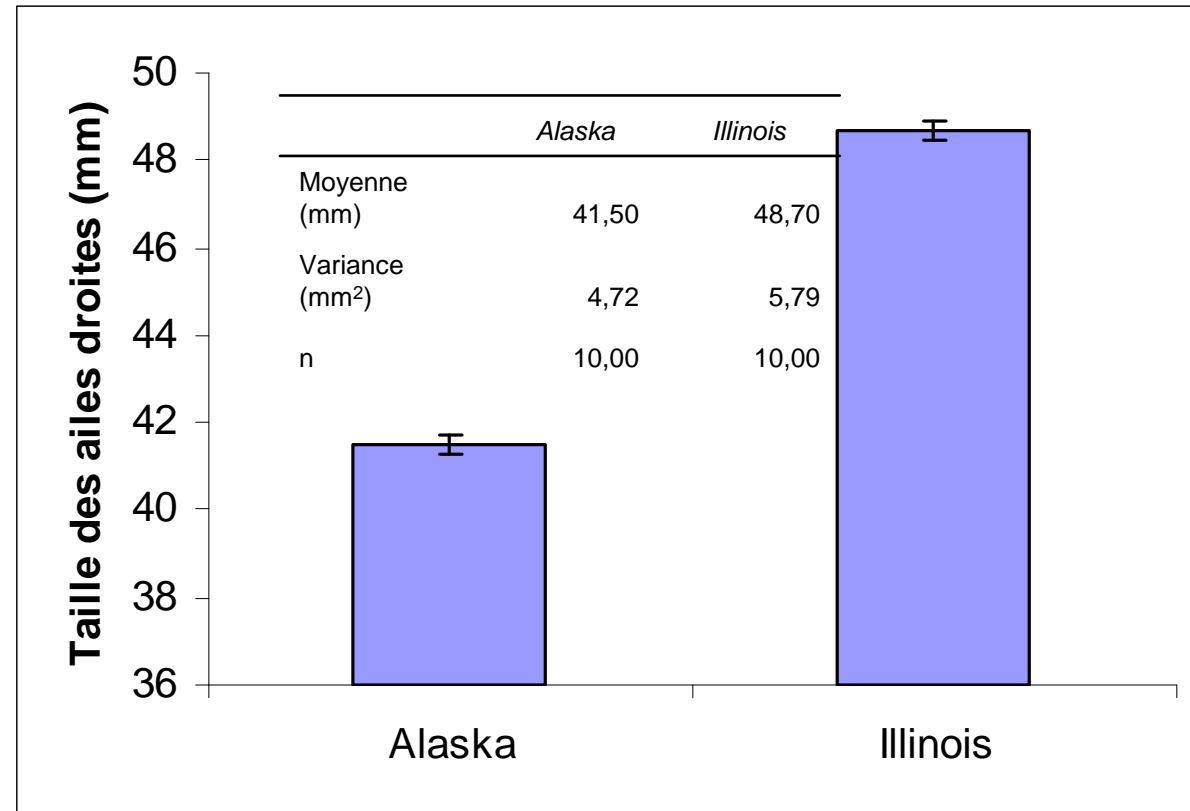
H_0	H_1	Rejet de H_0 si
	$\mu_1 \neq \mu_2$	$ t_{\text{calc}} \geq t_{\alpha/2, \nu} $
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$t_{\text{calc}} \geq t_{\alpha, \nu}$
	$\mu_1 < \mu_2$	$t_{\text{calc}} \leq -t_{\alpha, \nu}$

Robert va à la chasse au *Papilio glaucus*

Longueur des ailes antérieures droites (mm) des mâles de *Papilio glaucus* échantillonnés en Alaska et en Illinois une certaine année.



Alaska	Illinois
42	51
41	48
41	49
37	48
44	47
43	46
43	47
40	47
40	50
44	54



Étape 1: Question biologique

Y a-t-il une différence au niveau de la longueur des ailes des mâles de *Papilio glaucus* en Alaska et en Illinois ? (Le climat de l'Alaska empêche-t-il les papillons de s'y développer autant ?)

Étape 2: Déclaration des hypothèses

H_0 : Il n'y a pas de différence entre la longueur moyenne des ailes des mâles de *Papilio glaucus* d'Alaska et d'Illinois

$$\mu_{\text{Alaska}} = \mu_{\text{Illinois}}$$

H_1 : La longueur moyenne des ailes des mâles de *Papilio glaucus* d'Alaska est plus **petite** que celle des mâles d'Illinois (hypothèse unilatérale).

$$\mu_{\text{Alaska}} < \mu_{\text{Illinois}}$$

Étape 3: Choix du test

Comme n_1 ou n_2 est plus petit que 30, le test statistique utilisé est un test de t de Student où:

$$t_{calc} = \frac{\bar{x}_A - \bar{x}_I}{s_{pd} \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_I}\right)}} \quad \text{où} \quad s_{pd} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_I - 1)s_I^2}{n_A + n_I - 2}}$$

Si n_1 et $n_2 \geq 30$, on utilise le test Z

Étape 4: Conditions d'application

Indépendance des observations.

Normalité des distributions de données des échantillons.

Équivariance (ou homoscédasticité) des échantillons....????

Pour vérifier l'homoscédasticité, le mini-test de F

H_0 : les variances des deux échantillons sont égales : $\sigma^2_1 = \sigma^2_2$

H_1 : Les variances des deux échantillons sont différentes : $\sigma^2_1 \neq \sigma^2_2$

Si les échantillons sont tirés de populations normales (conditions de normalité), le rapport de leurs variances s^2_1/s^2_2 suivra une distribution de F à $v1 = n_1 - 1$ et $v2 = n_2 - 1$ degrés de liberté.

On rejettera H_0 au seuil $\alpha = 0,05$ si $F_{\text{calc}} \geq F_{(\alpha/2; v1, v2)}$

Attention!!! Il faut **toujours** mettre la **variance la plus grande** au **NUMÉRATEUR**!

MINI-TEST de F!

$$H_0: \sigma^2_1 = \sigma^2_2 \quad H_1: \sigma^2_1 \neq \sigma^2_2$$

$$F_{\text{calc}} = s^2_I / s^2_A = 5,79 / 4,72 = 1,227$$

$$F_{(0,025;9,9)} = 4,03$$

Donc $F_{\text{calc}} < F_{(\alpha/2; v_1, v_2)}$ et on ne rejette pas H_0 : il y a équivariance

Étape 5: Distribution de la variable auxiliaire

Si H_0 est vraie, la variable auxiliaire t_{calc} suivra une distribution de Student à $\nu = n_A + n_I - 2 = 10 + 10 - 2 = 18$ degrés de liberté.

Étape 6: Règle de décision

On rejette H_0 au seuil $\alpha = 0,05$ si $t_{\text{calc}} \leq -t_{(\alpha;\nu)} = -t_{(0.05;18)} = -1,734$

Étape 7: Calcul du test

	<i>Alaska</i>	<i>Illinois</i>
Moyenne	41,50	48,70
Variance	4,72	5,79
n	10	10
S_{pd}	5,26	
ν	18,00	
t_{calc}	-7,022	

$$t_{\text{calc}} = \frac{\bar{x}_A - \bar{x}_I}{s_{pd} \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_I}\right)}}$$

où $s_{pd} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_I - 1)s_I^2}{n_A + n_I - 2}}$

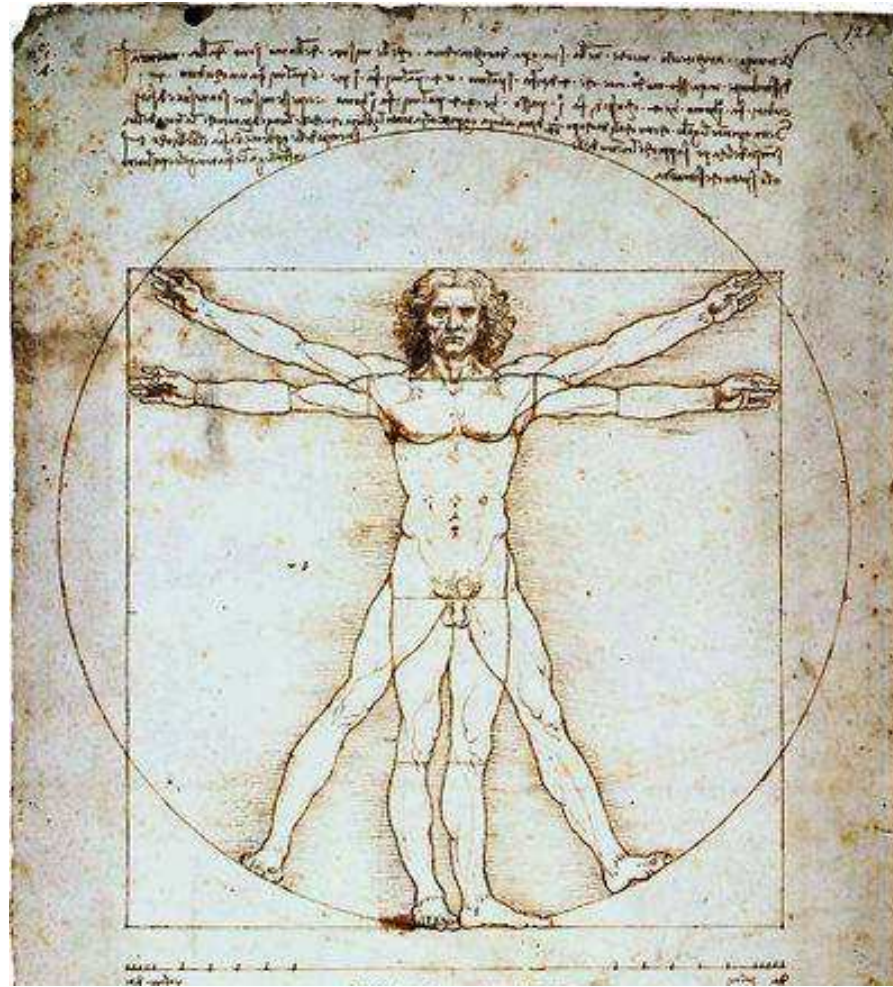
Étape 8: Décision statistique

Puisque $t_{calc} = -7,022 < -t_{(0.05;18)} = -1,734$, on rejette H_0 au seuil $\alpha = 0,05$.

Étape 9: Interprétation biologique

Les papillons mâles ont en moyenne des longueurs **d'ailes antérieures droites plus petites en Alaska qu'en Illinois** car la saison de croissance y est plus courte ce qui défavorise leur croissance.

Comparaison de 2 échantillons appariés



Objectif

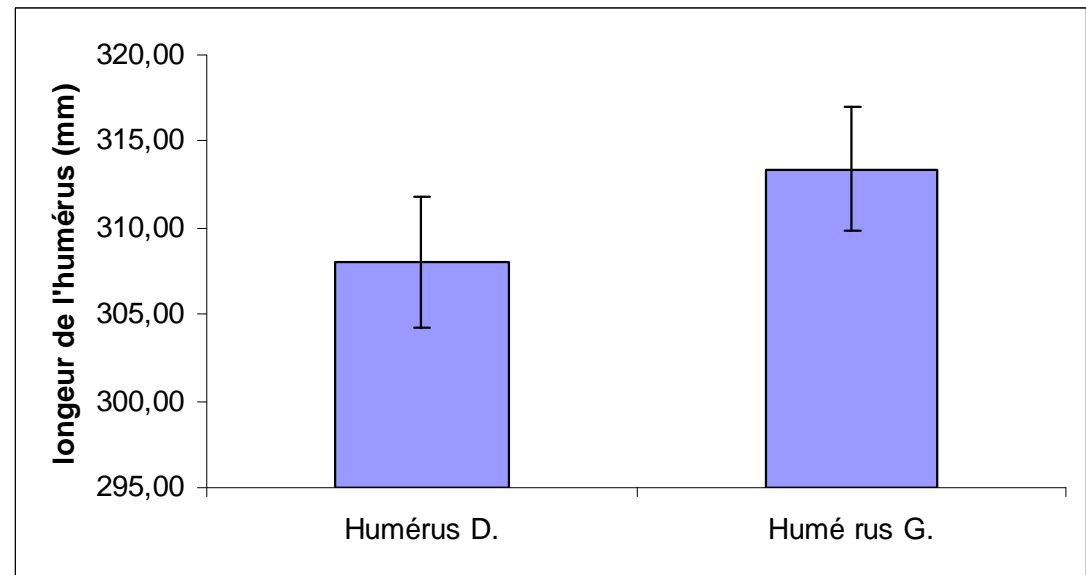
Tester la différence entre les moyennes de 2 échantillons appariés [i.e. les deux échantillons ont les mêmes éléments ou des éléments liés par au moins un critère

Le test t pour échantillons appariés

Principe : analyse des différences d observées pour chaque paire d'observations

Exemple. Comparaison des longueurs moyennes (mm) des humérus droit et gauche de dix squelettes de femmes. D'après Jolicoeur (1998).

# squelette	Humérus D.	Humérus G.
1	311	315
2	302	306
3	301	311
4	322	333
5	312	316
6	285	292
7	305	308
8	310	318
9	328	326
10	304	309



	<i>Humérus D.</i>	<i>Humérus G.</i>
Moyenne	308,00	313,40
Variance	140,44	126,71

Étape 1. Question biologique

Y a-t-il une différence de longueur entre l'humérus droit et l'humérus gauche chez les femmes ?

Étape 2. Déclaration des hypothèses

H_0 : Il n'y a pas de différence entre la longueur moyenne des humérus gauche et droit chez les femmes

$$\mu_D = \mu_G$$

H_1 : Il y a une différence entre la longueur moyenne des humérus gauche et droit chez les femmes

$$\mu_D \neq \mu_G$$

Étape 3. Choix du test

Le test statistique utilisé est un test de t de Student pour échantillons appariés:

$$t_{\bar{d}} = \frac{\bar{d}}{s_{\bar{d}}} \quad \text{où} \quad s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \quad \text{et} \quad \bar{d} = \bar{x}_1 - \bar{x}_2$$

Étape 4. Conditions d'application

Les échantillons sont appariés.

Les différences de longueur suivent une distribution normale.

Étape 5: Distribution de la variable auxiliaire

Si H_0 est vraie, la variable auxiliaire t_d suivra une distribution de Student à $\nu = n - 1 = 10 - 1 = 9$ ddl, où n est le nombre de **différences** de longueur.

Étape 6. Règle de décision

On rejette H_0 au seuil $\alpha = 0,05$ si $|t_d| \geq |t_{(\alpha/2;\nu)}|$ où $t_{(\alpha/2;\nu)} = t_{(0,025;9)} = 2,262$.

Étape 7. Calcul du test

# squelette	Humérus D.	Humérus G.	d
1	311	315	-4
2	302	306	-4
3	301	311	-10
4	322	333	-11
5	312	316	-4
6	285	292	-7
7	305	308	-3
8	310	318	-8
9	328	326	2
10	304	309	-5

$$\bar{d} = -5,4$$

$$s_d = 3,78$$

$$s_{\bar{d}} = 1,19$$

$$t_{\bar{d}} = -4,52$$

Étape 8. Décision statistique

Puisque $|t_d| > |t_{(\alpha;v)}|$, on rejette H_0 au seuil $\alpha = 0,05$.

Étape 9. Interprétation biologique

Les femmes ont des humérus de longueurs différentes car le bras ...à vous d'imaginer la conclusion 😊