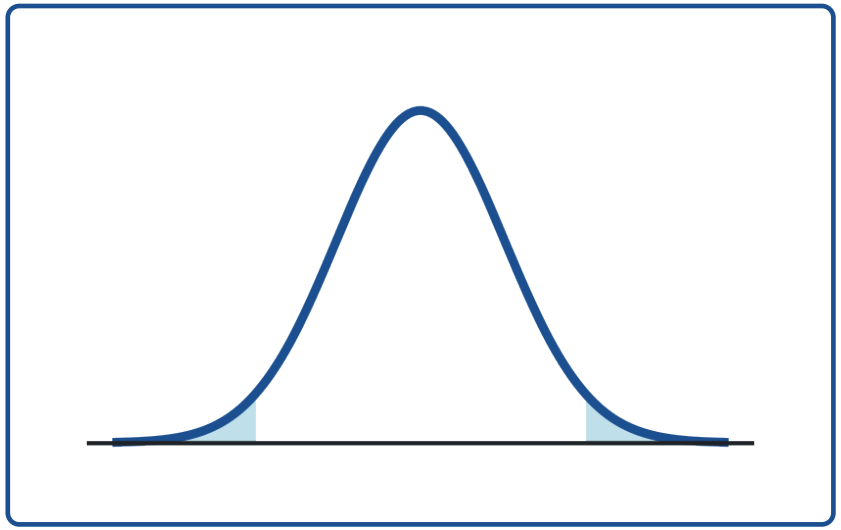
Tests statistiques (2)

Tests paramétriques : test t de Student et test F de Fisher



Plan du cours

01 Test t de Student pour échantillons indépendants

02 Test t de Student pour échantillons dépendants

03 Test F de Fisher, Levene, Bartlett

04 Exemples depuis des articles médicaux

Tests statistiques PARAMÉTRIQUES

- Comparaison de deux **moyennes** :
  - échantillons indépendants, avec variances égales
    - par le *test t de Student pour échantillons indépendants*
  - échantillons dépendants (appariés)
    - par le *test t de Student pour échantillons dépendants*
- Comparaison de deux **variances**
  - par le *test F de Fisher*
- Exemples depuis des articles médicaux, avec explications
- Exemples sur vos données (questionnaire), avec explications

Test t de Student : comparaison de deux moyennes

1. Échantillons indépendants : les observations sont indépendantes d'un groupe à l'autre

a. Comparaison de deux moyennes

- le test de Student (t) avec **inégalité des variances** dans les populations
- le test de Student (t) avec **égalité des variances** dans les populations

2. Échantillons dépendants (appariés) : mesures répétées de la même variable

a. Comparaison de deux moyennes

- le test t de Student pour échantillons dépendants

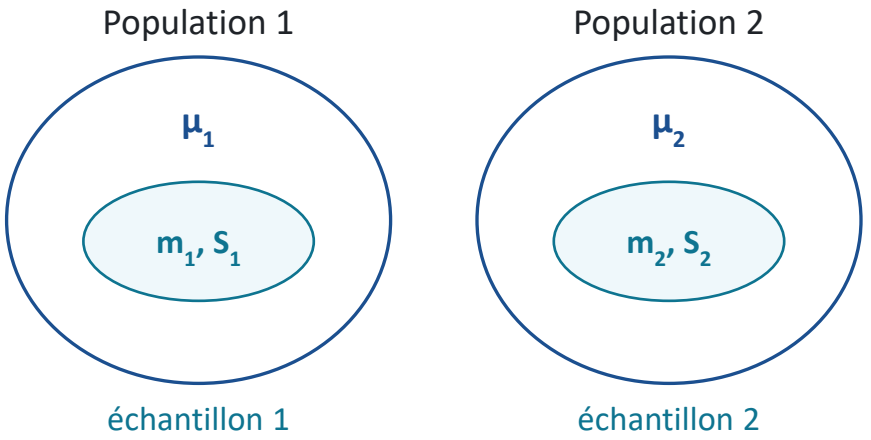
Test t de Student : comparaison de deux moyennes

- On dispose d'une variable qualitative dichotomique qui permet de définir deux groupes (ex. hommes vs femmes).
- On mesure une variable quantitative (X = cholestérol).

m_1 = moyenne de la variable X sur l'échantillon 1

m_2 = moyenne de la variable X sur l'échantillon 2

S_1, S_2 = déviations standard d'échantillonnage



μ_1 = moyenne de la variable X sur la population 1

μ_2 = moyenne de la variable X sur la population 2

Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Objectif

comparaison des moyennes d'une variable quantitative continue (2 échantillons)

Conditions d'application :

- observations indépendantes dans chaque échantillon
- deux échantillons indépendants
- une variable quantitative mesurée sur les 2 groupes (échantillons)
- les données sont normalement distribuées dans les deux populations
- les variances des populations sont inconnues
- le nombre des observations < 30 ou ≥ 30

L'hypothèse nulle (H_0)

Il n'y a pas de **différence significative entre** les **moyennes** des deux **populations** (depuis lesquelles les deux échantillons ont été extraits) (la différence entre les deux moyennes est égale à zéro)

L'hypothèse alternative (H_1) du test de Student bilatéral

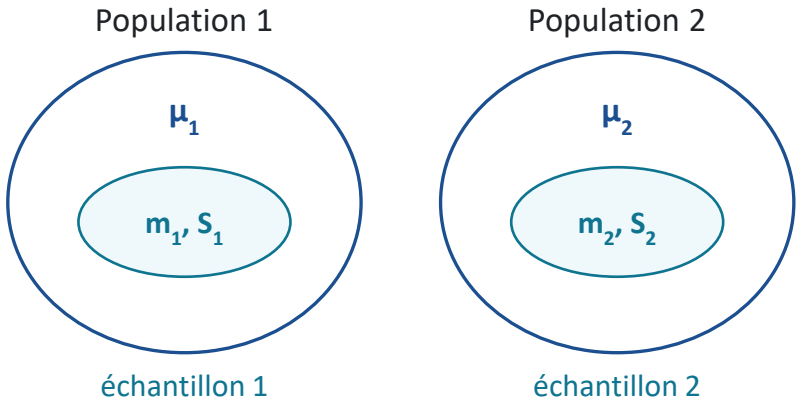
Il y a une **différence significative entre** les **moyennes** des deux **populations** (la différence entre les deux moyennes n'est pas égale à zéro)

Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Étape 1 (test non directionnel)

$H_0 : \mu_1 = \mu_2 \leftrightarrow \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 \neq \mu_2 \leftrightarrow \mu_1 - \mu_2 \neq 0$



Étape 2. Deux possibilités : variances égales ou inégales.

Pour décider, on doit appliquer un autre test (test F de Fisher) pour comparer les variances (voir ce cours).

a) Cas des variances inconnues et inégales

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

n_1, n_2 = les tailles des échantillons

m_1, m_2 = moyennes des échantillons

S_1^2, S_2^2 = les variances d'échantillonnage

Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Étape 2.

b) Cas des variances inconnues et égales :

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

où la variance d'échantillonnage commune est donnée par :

$$S^2 = \frac{(n_1 - 1) \times S_1^2 + (n_2 - 1) \times S_2^2}{n_1 + n_2 - 2}$$

Étape 3.

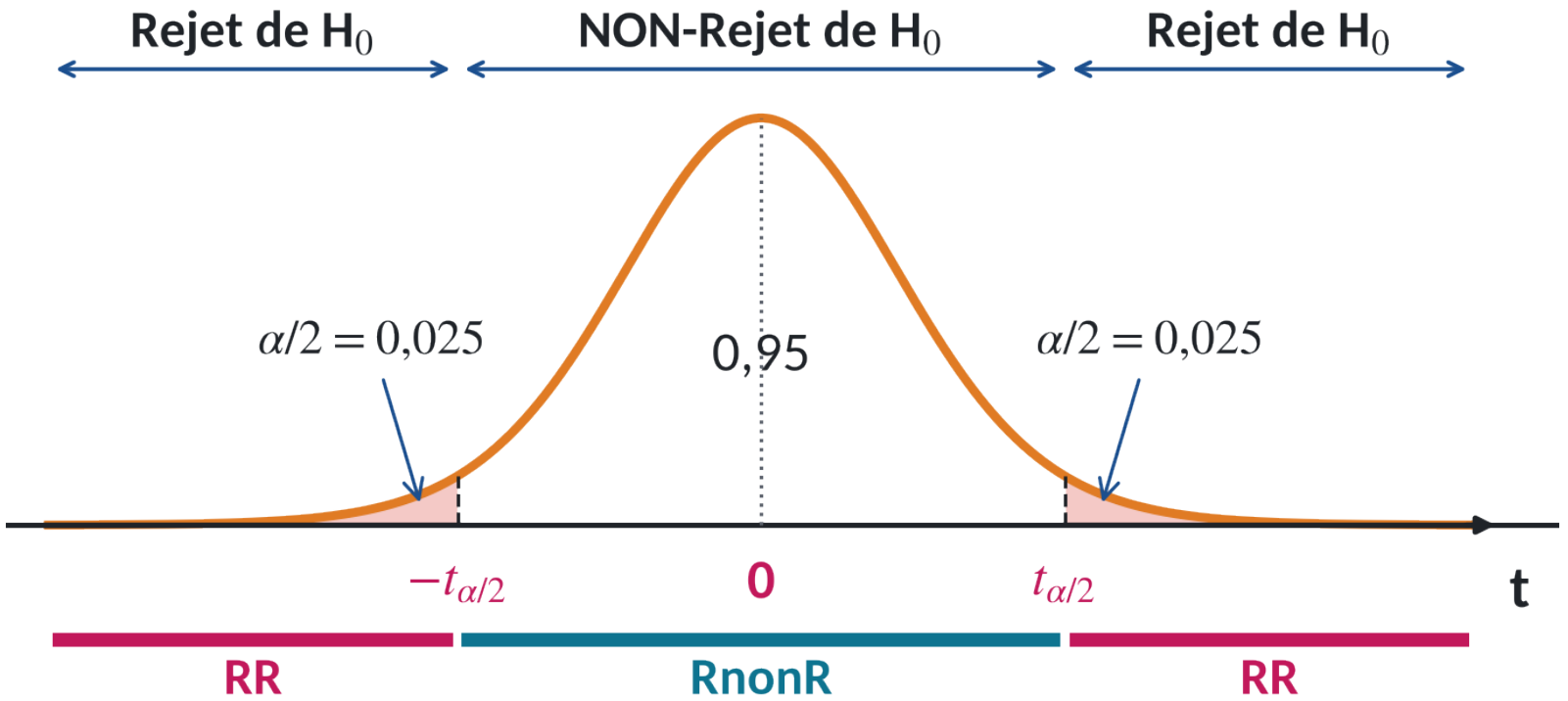
α = **0,05** (ou α = 0,01, α = 0,001)

Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Étape 3. α = **0,05** (ou α = 0,01, α = 0,001)

Étape 4. Région de rejet (RR) = région de rejet de H_0

$$RR = (-\infty;\ -t_{n_1+n_2-2,\ \alpha/2}] \cup [t_{n_1+n_2-2,\ \alpha/2};\ +\infty)$$



$$t = \frac{m_1 - m_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Étape 5 : on calcule la valeur de la statistique du test (notée t_0)

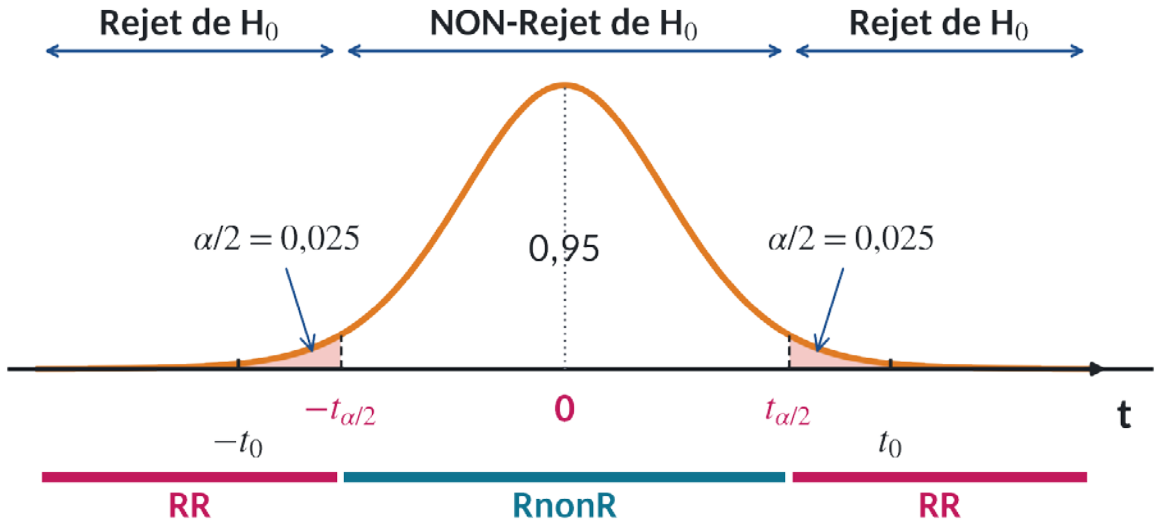
Étape 6. Décision du test à l'aide de la RR

Si $t_0 \in$ RR → on rejette H_0 → en faveur de H_1 (en gros, on accepte H_1)

Si $t_0 \notin$ RR (donc $t_0 \in$ RnonR) → **on NE peut PAS rejeter H_0**, donc notre étude n'a pas fourni de preuves suffisantes pour conclure à l'existence d'une différence significative* entre les moyennes…

↔ on ne peut pas dire qu'il **existe** une **différence significative entre** les **moyennes**…

↔ dans cette étude, on n'a pas trouvé **une différence significative entre les moyennes…**



* « L'absence de preuve n'est pas la preuve de l'absence » : **on n'accepte jamais H_0**

« on n'a pas trouvé » ≠ « il n'y a pas » ou « il n'existe pas »

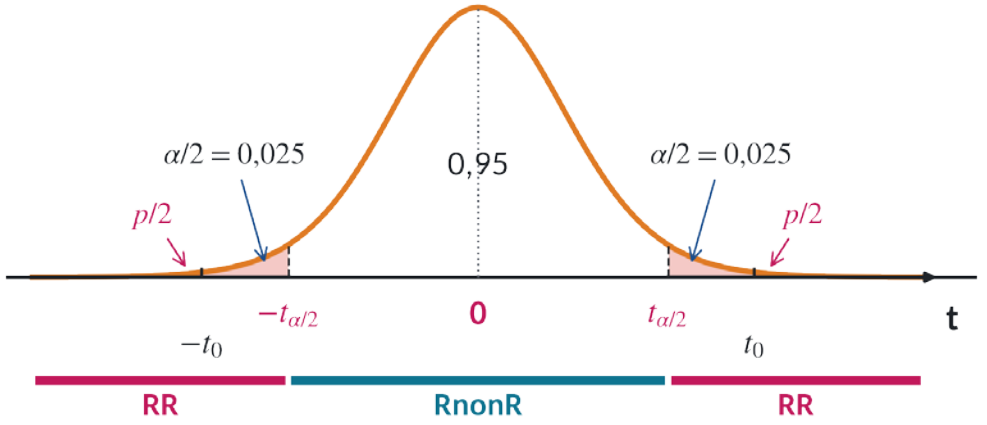
Test t de Student pour échantillons indépendants : comparaison de deux moyennes

Étape 6'. Prendre la décision avec la p-value (p-valeur).

p-value = niveau de signification estimé = probabilité sous H_0 pour que la statistique du test prenne une valeur plus extrême que celle qu'on observe.

Si p-value < α (= 0,05) → **on rejette H_0** → **en faveur de H_1** (en gros, on accepte H_1)

Si p-value ≥ α (= 0,05) → **on NE peut PAS rejeter H_0** → notre étude n'a pas fourni de preuves suffisantes pour conclure à l'existence d'une différence significative* entre les moyennes… ↔ dans cette étude, on n'a pas trouvé **une différence significative entre les moyennes…**



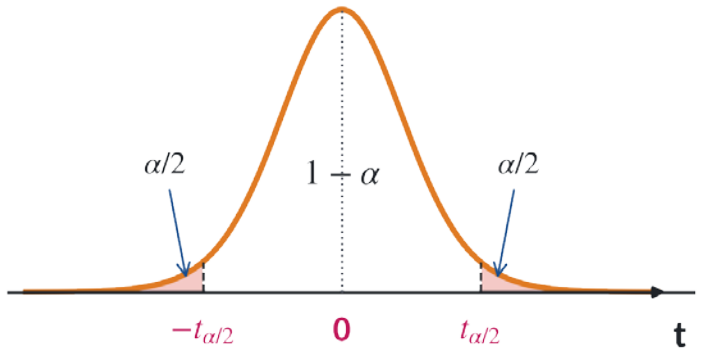
$p = Pr(|T| \geq |t_0| \mid \text{sous } H_0) = 2 \cdot Pr(T \geq |t_0| \mid \text{sous } H_0) = Pr(T \geq |t_0|) + Pr(T \leq -|t_0|)$

Table de la loi de Student

	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	

Source : http://statcalculators.com/students-t-distribution-table/

t	α = 0,05
df = 11	2,20



Test bilatéral (2-tail) :

α = 0,05 et df = 11 → valeur critique

$t_{\alpha/2,\ df} = |t_{0,025;\ 11}| = 2,20$

Exemple : test t de Student, échantillons indépendants avec des variances égales

Question : on se pose la question de savoir si la créatinine moyenne des femmes diabétiques est significativement différente de celle des femmes non diabétiques.

À partir de 2 échantillons de taille : $n_1 = 8$; $n_2 = 9$

Femmes diabétiques :
moyenne du taux de créatinine $m_1 = 15$ mg/L
déviation standard d'échantillonnage $S_1 = 4$ mg/L

Femmes non diabétiques :
moyenne du taux de créatinine $m_2 = 7$ mg/L
déviation standard d'échantillonnage $S_2 = 3$ mg/L

Les données (valeurs de la créatinine) suivent une distribution normale et un test F a suggéré que les variances sont égales.

Au risque de 5 %, peut-on affirmer que la moyenne de la créatinine des femmes diabétiques est significativement différente de celle des femmes non diabétiques ?

Exemple : test t de Student, échantillons indépendants avec des variances égales

Le choix du test : obs. indép. ; taux de créatinine (**mg/L**) = variable quantitative continue ; distribution normale ; **variances des populations inconnues et égales** ⇒ le test t pour des échantillons indépendants avec des variances égales

Étape 1 $H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$ (test bilatéral)

Étape 2.

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

Étape 3. α = 5 % (= 0,05)

Étape 4.

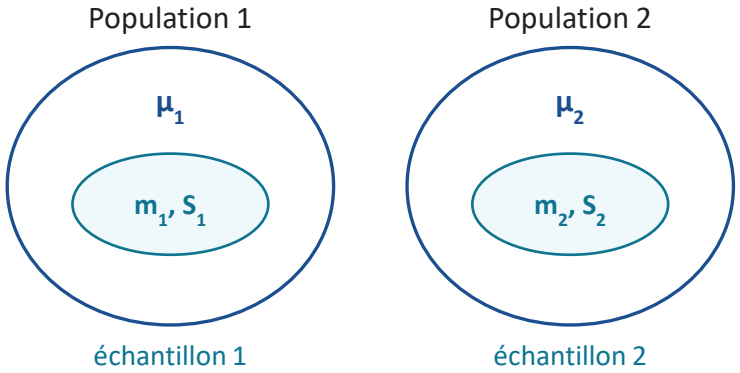
$$RR = (-\infty;\ -2{,}131] \cup [2{,}131;\ +\infty)$$

Étape 5.

$$t_{\text{observé}} = \frac{m_1 - m_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} = \frac{15 - 7}{\sqrt{\frac{12{,}27}{8} + \frac{12{,}27}{9}}}$$

$t_{observé}$ = **4,70**

$$S^2 = \frac{(8-1)\times 16 + (9-1)\times 9}{15} = 12{,}27$$



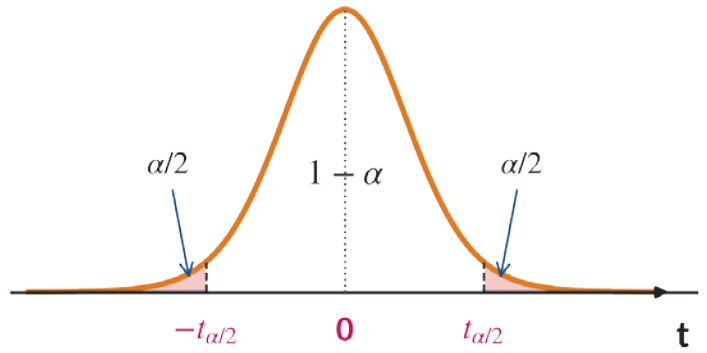
n_1, n_2 = les tailles des échantillons

Table de la loi de Student

	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	

Source : http://statcalculators.com/students-t-distribution-table/

t	α = 0,05
df = 15	2,13



Test bilatéral (2-tail) :

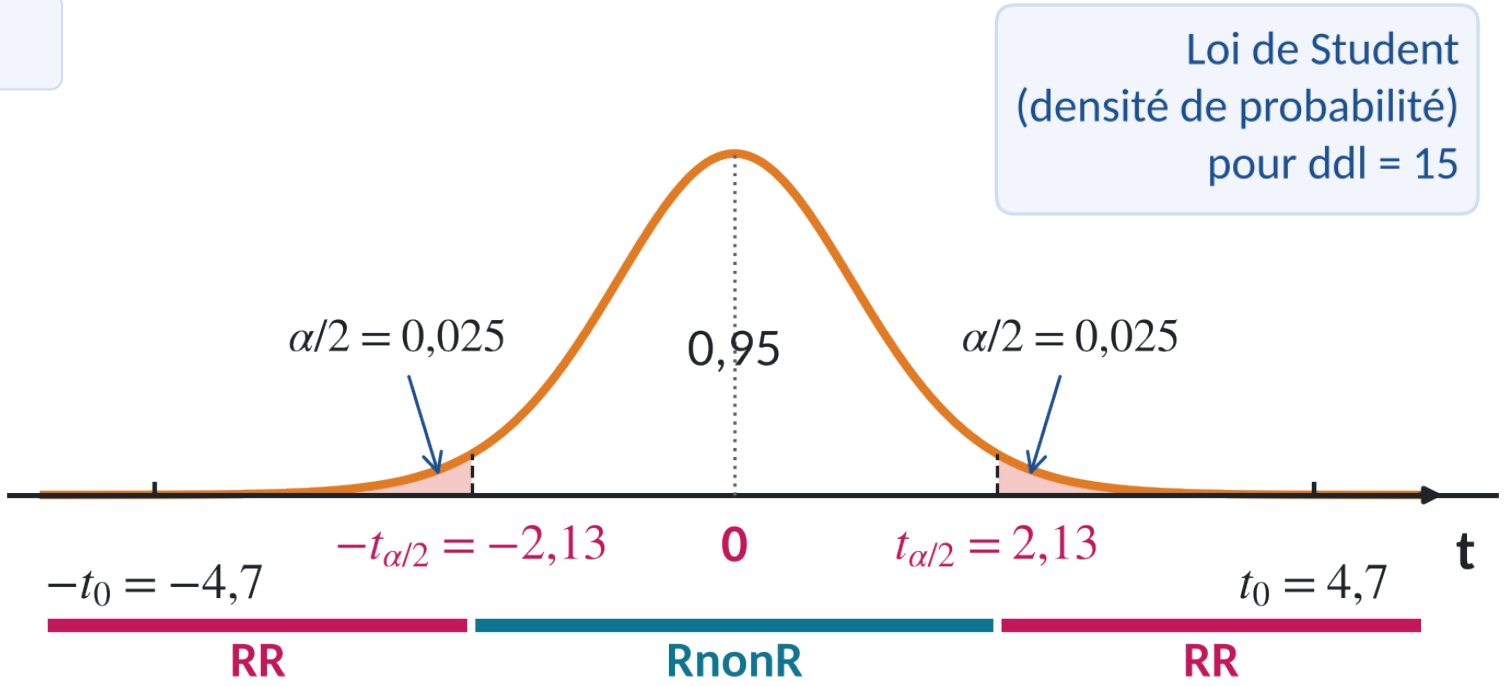
α = 0,05 et df = 15 → valeur critique

$t_{\alpha/2,\ df} = |t_{0,025;\ 15}| = \mathbf{2,13}$

Exemple : test t de Student, échantillons indépendants avec des variances égales

Étape 5.

$t_0 = 4,7$



Étape 6 (décision statistique)

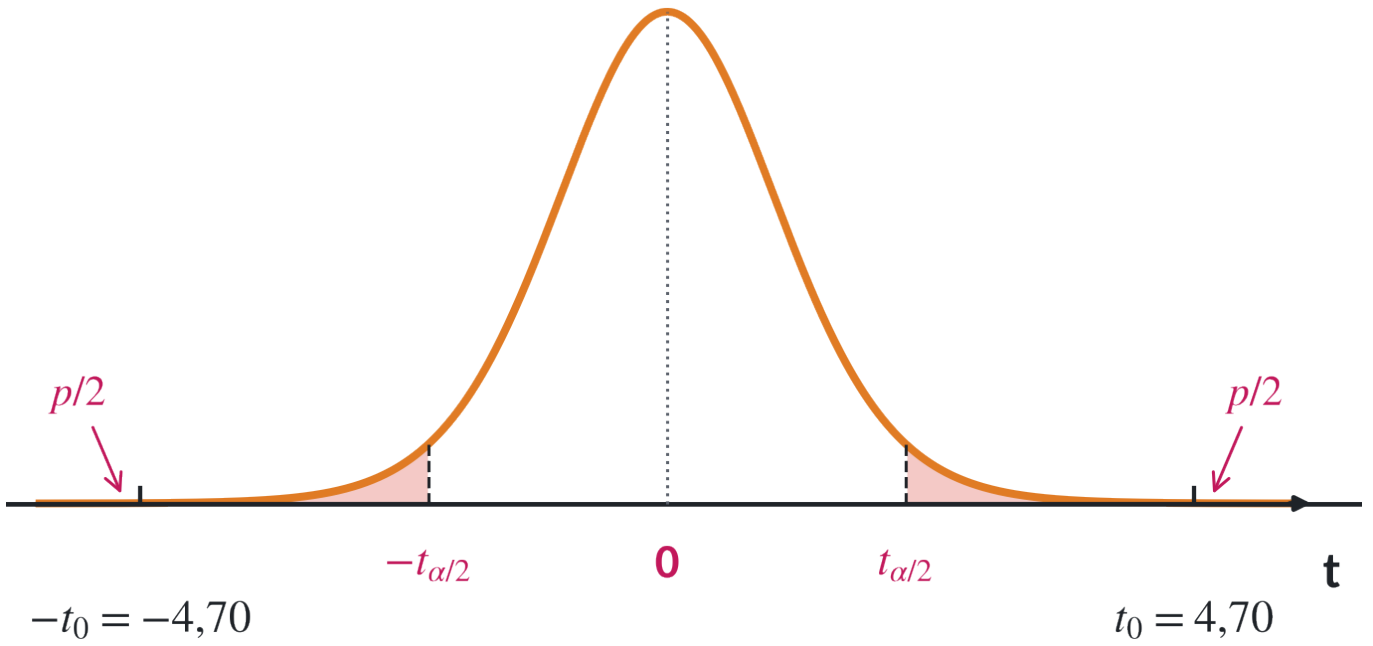
$t_0 = 4,7 \in$ RR : **on rejette H_0** → en faveur de H_1

Au risque d'erreur de 5 %, **il y a** une **différence significative entre** les **moyennes** du **taux de créatinine** pour les **deux populations**.

Exemple : test t de Student, échantillons indépendants avec des variances égales

Étape 6' : décision du test à l'aide de la p-valeur

p = 0,0003 < 0,05 : **on rejette H_0** → en faveur de H_1, donc au risque d'erreur de 5 %, **il y a** une **différence significative entre** les **moyennes** du **taux de créatinine** pour les **deux populations**.



Le test F de Fisher pour des échantillons indépendants

Objectif : comparaison des **variances** d'une variable quantitative continue dans **deux échantillons indépendants**.

Conditions d'application :

- les observations sont indépendantes
- 2 échantillons indépendants
- variable quantitative
- les données sont normalement distribuées (dans les deux populations)

L'hypothèse nulle : H_0

Il n'y a pas de **différence entre** les deux **variances** des deux **populations**

L'hypothèse alternative : H_1 pour le test non directionnel

Il y a une **différence entre** les deux **variances** des deux **populations**

Il y a aussi d'autres tests pour comparer les variances, plus robustes que le test F : le test de **Bartlett** ou le test de **Levene**.

Les étapes du test F de Fisher pour des échantillons indépendants

Étape 1. pour le test bilatéral

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Étape 2. Statistique du test :

$$\mathbf{F} = \begin{cases} \frac{S_2^2}{S_1^2}, & \text{pour } S_2^2 > S_1^2 \\ \frac{S_1^2}{S_2^2}, & \text{pour } S_1^2 > S_2^2 \end{cases}$$

- S_1^2, S_2^2 = les variances d'échantillonnage
- F suit la loi de Fisher avec ($n_2 - 1$, $n_1 - 1$) ou ($n_1 - 1$, $n_2 - 1$) degrés de liberté

Étape 3. α = **0,05** (ou α = 0,01, α = 0,001)

Étape 4. $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ d.d.l.

$$RR = [F_{\nu_1, \nu_2, \alpha/2};\ \infty)$$

- $F_{v1, v2, \alpha}$ se trouve dans la table de la distribution de Fisher ($n_2 - 1$, $n_1 - 1$) ou ($n_1 - 1$, $n_2 - 1$) degrés de liberté (d.d.l.) *(voir cours Variables aléatoires)*

Étape 5. Calculer la valeur de la statistique du test (F_0)

Étape 6. Si $F_0 \in RR$, **on rejette H_0** → H_1 (dans le cas contraire, on NE rejette PAS H_0)

Les étapes du test F de Fisher pour des échantillons indépendants

Étape 6'. Prendre la décision avec la p-value (p-valeur).

p-value = **niveau de signification observé**

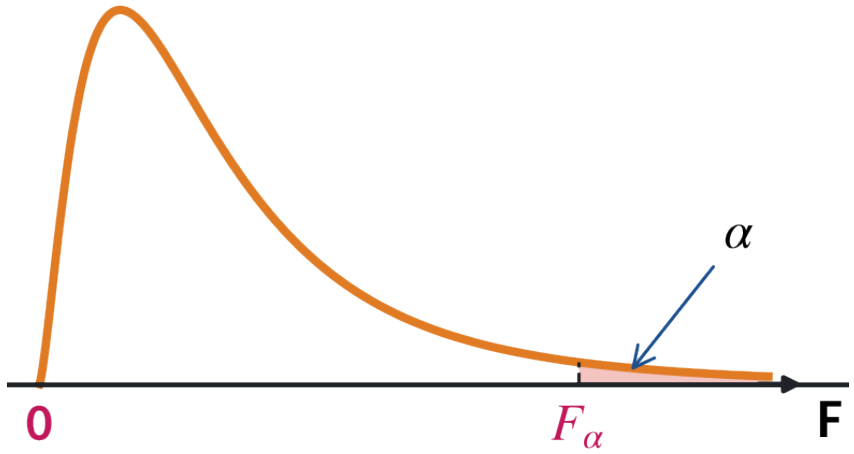
- **Si p-value ≤ α (= 0,05)**, **on rejette H_0** → en faveur de H_1
- **Si p-value > α (= 0,05)** → **on NE peut PAS rejeter H_0**

La loi de Fisher

F	α = 0,05
$df_{numérateur} = \nu_1 = 120$; $df_{dénominateur} = \nu_2 = 60$	**1,47**

F Distribution critical values for P=0.05

▼ Denominator	Numerator DF													
DF	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**15**	**20**	**30**	**60**	**120**	**500**	**1000**
1	161.45	199.50	215.71	224.58	230.16	236.77	241.88	245.95	248.01	250.10	252.20	253.25	254.06	254.19
2	18.513	19.000	19.164	19.247	19.296	19.353	19.396	19.429	19.446	19.462	19.479	19.487	19.494	19.495
3	10.128	9.5522	9.2766	9.1172	9.0135	8.8867	8.7855	8.7028	8.6602	8.6165	8.5720	8.5493	8.5320	8.5292
4	7.7086	6.9443	6.5915	6.3882	6.2560	6.0942	5.9644	5.8579	5.8026	5.7458	5.6877	5.6580	5.6352	5.6317
5	6.6078	5.7862	5.4095	5.1922	5.0504	4.8759	4.7351	4.6187	4.5582	4.4958	4.4314	4.3985	4.3731	4.3691
7	5.5914	4.7375	4.3469	4.1202	3.9715	3.7871	3.6366	3.5108	3.4445	3.3758	3.3043	3.2675	3.2388	3.2344
10	4.9645	4.1028	3.7082	3.4780	3.3259	3.1354	2.9782	2.8450	2.7741	2.6996	2.6210	2.5801	2.5482	2.5430
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7066	2.5437	2.4035	2.3275	2.2467	2.1601	2.1141	2.0776	2.0718
20	4.3512	3.4928	3.0983	2.8660	2.7109	2.5140	2.3479	2.2032	2.1241	2.0391	1.9463	1.8962	1.8563	1.8498
30	4.1709	3.3159	2.9223	2.6896	2.5336	2.3343	2.1646	2.0149	1.9317	1.8408	1.7396	1.6835	1.6376	1.6300
60	4.0012	3.1505	2.7581	2.5252	2.3683	2.1666	1.9927	1.8365	1.7480	1.6492	1.5343	1.4672	1.4093	1.3994
120	3.9201	3.0718	2.6802	2.4473	2.2898	2.0868	1.9104	1.7505	1.6587	1.5544	1.4289	1.3519	1.2804	1.2674
500	3.8601	3.0137	2.6227	2.3898	2.2320	2.0278	1.8496	1.6864	1.5917	1.4820	1.3455	1.2552	1.1586	1.1378
1000	3.8508	3.0047	2.6137	2.3808	2.2230	2.0187	1.8402	1.6765	1.5811	1.4705	1.3318	1.2385	1.1342	1.1096



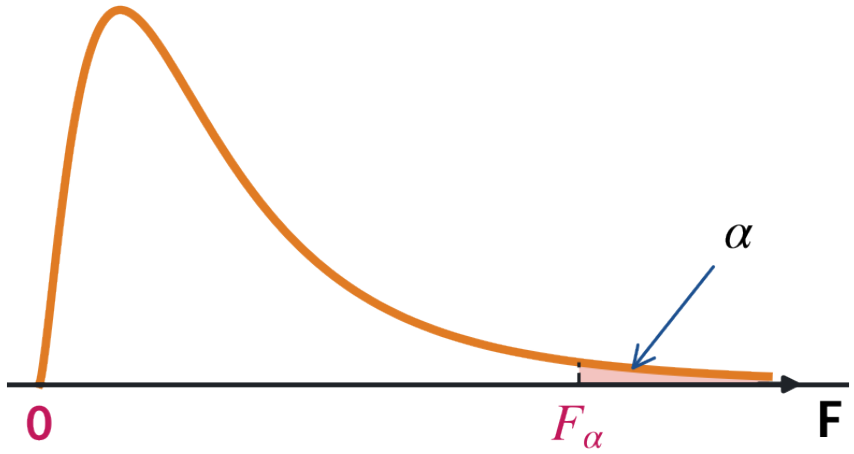
Source : https://www.medcalc.org/manual/f-table.php

La loi de Fisher

F	α = 0,05
$df_{numérateur} = \nu_1 = 4$; $df_{dénominateur} = \nu_2 = 5$	**5,19**

F Distribution critical values for P=0.05

↓ Denominator	Numerator DF													
DF	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**15**	**20**	**30**	**60**	**120**	**500**	**1000**
1	161.45	199.50	215.71	224.58	230.16	236.77	241.88	245.95	248.01	250.10	252.20	253.25	254.06	254.19
2	18.513	19.000	19.164	19.247	19.296	19.353	19.396	19.429	19.446	19.462	19.479	19.487	19.494	19.495
3	10.128	9.5522	9.2766	9.1172	9.0135	8.8867	8.7855	8.7028	8.6602	8.6165	8.5720	8.5493	8.5320	8.5292
4	7.7086	6.9443	6.5915	6.3882	6.2560	6.0942	5.9644	5.8579	5.8026	5.7458	5.6877	5.6580	5.6352	5.6317
5	6.6078	5.7862	5.4095	5.1922	5.0504	4.8759	4.7351	4.6187	4.5582	4.4958	4.4314	4.3985	4.3731	4.3691
7	5.5914	4.7375	4.3469	4.1202	3.9715	3.7871	3.6366	3.5108	3.4445	3.3758	3.3043	3.2675	3.2388	3.2344
10	4.9645	4.1028	3.7082	3.4780	3.3259	3.1354	2.9782	2.8450	2.7741	2.6996	2.6210	2.5801	2.5482	2.5430
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7066	2.5437	2.4035	2.3275	2.2467	2.1601	2.1141	2.0776	2.0718
20	4.3512	3.4928	3.0983	2.8660	2.7109	2.5140	2.3479	2.2032	2.1241	2.0391	1.9463	1.8962	1.8563	1.8498
30	4.1709	3.3159	2.9223	2.6896	2.5336	2.3343	2.1646	2.0149	1.9317	1.8408	1.7396	1.6835	1.6376	1.6300
60	4.0012	3.1505	2.7581	2.5252	2.3683	2.1666	1.9927	1.8365	1.7480	1.6492	1.5343	1.4672	1.4093	1.3994
120	3.9201	3.0718	2.6802	2.4473	2.2898	2.0868	1.9104	1.7505	1.6587	1.5544	1.4289	1.3519	1.2804	1.2674
500	3.8601	3.0137	2.6227	2.3898	2.2320	2.0278	1.8496	1.6864	1.5917	1.4820	1.3455	1.2552	1.1586	1.1378
1000	3.8508	3.0047	2.6137	2.3808	2.2230	2.0187	1.8402	1.6765	1.5811	1.4705	1.3318	1.2385	1.1342	1.1096



Source : https://www.medcalc.org/manual/f-table.php

Le test t de Student pour des échantillons appariés

Objectif : comparaison des moyennes d'une variable quantitative continue de deux échantillons dépendants (appariés)

Conditions d'application :

- observations indépendantes dans chaque échantillon
- échantillons dépendants (appariés)
- variable quantitative
- les données sont normalement distribuées
- les variances des observations sont inconnues
- le nombre des observations < 30 ou ≥ 30

L'hypothèse nulle : H_0

Il n'y a pas de **différence statistiquement significative entre** les deux **populations** du point de vue de la **moyenne** (la moyenne de la différence entre les deux populations est égale à zéro)

L'hypothèse alternative : H_1 pour le test non directionnel

Il y a une **différence statistiquement significative entre** les deux **populations** du point de vue de la **moyenne** (la moyenne de la différence entre les deux populations n'est pas égale à zéro)

Le test t de Student pour des échantillons appariés

Étape 1. pour le test non directionnel

$$H_0 : \bar{d} = 0$$ **la différence = 0**

$$H_1 : \bar{d} \neq 0$$

Étape 2. **Statistique du test**

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

- $n_1 = n_2$ → **n** = taille de l'échantillon (nombre de paires)
- đ = moyenne des différences
- S^2 = la variance d'échantillonnage des différences

Étape 3. α = **0,05** (ou α = 0,01, α = 0,001)

Étape 4. d.d.l. = n − 1

$$RR = (-\infty;\ -t_{n-1,\ \alpha/2}] \cup [t_{n-1,\ \alpha/2};\ +\infty)$$

Étape 5.

la valeur calculée de la statistique du test

Étape 6.

Si $t_0 \in$ RR, **on rejette H_0** → H_1 (dans le cas contraire, on ne peut pas rejeter H_0)

Exemple : test t de Student pour des données appariées

- **Objectif :** un certain médicament X modifie le rythme cardiaque (battements par minute, bpm) chez les personnes diabétiques.
- Sur un échantillon de 10 personnes diabétiques, on mesure le rythme cardiaque (**fréquence cardiaque**) deux fois par individu (avant la prise du médicament et après son administration).
- On suppose que le rythme cardiaque suit une loi normale dans la population dont sont issus ces 10 patients. On a obtenu les résultats suivants :

Patients	1	2	3	4	5	6	7	8	9	10
Avant	73	67	83	52	74	86	68	70	65	66
Après	65	61	46	33	59	62	71	61	65	70

Question de recherche :

Formulation simplifiée : **Y a-t-il** une **différence significative entre** les **fréquences cardiaques moyennes avant** et **après** l'administration du médicament X ?

Formulation plus précise : est-ce que la **moyenne des différences** entre les valeurs du rythme cardiaque avant et après l'administration du médicament est **significativement différente** de 0 ?

Exemple : test t de Student pour des données appariées

Le choix du test : 10 observations indépendantes ; le rythme cardiaque = variable quantitative ; distribution normale ; on compare des données avant et après l'administration du médicament ⇒ le test t pour échantillons dépendants ou appariés

Étape 1. test bilatéral

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} \neq 0$$

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

Étape 2. **Statistique du test :**

- n = taille de l'échantillon ; $\bar{d}$ = moyenne des différences
- S^2 = la variance d'échantillonnage des différences

Patients	1	2	3	4	5	6	7	8	9	10
Avant	73	67	83	52	74	86	68	70	65	66
Après	65	61	46	33	59	62	71	61	65	70
Différences	**8**	**6**	**37**	**19**	**15**	**24**	**-3**	**9**	**0**	**-4**

Exemple : test t de Student pour des données appariées

Différences (d_i)	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
8	−3,1	9,61
6	−5,1	26,01
37	25,9	670,81
19	7,9	62,41
15	3,9	15,21
24	12,9	166,41
−3	−14,1	198,81
9	−2,1	4,41
0	−11,1	123,21
−4	−15,1	228,01
$\bar{d}$ = 11,1	**S^2 =**	**167,2111**
	S =	**12,93101**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$$

$$S = \sqrt{S^2}$$

Exemple : test t de Student pour des données appariées

Étape 3 : $\alpha = 0{,}05$

Étape 4 :

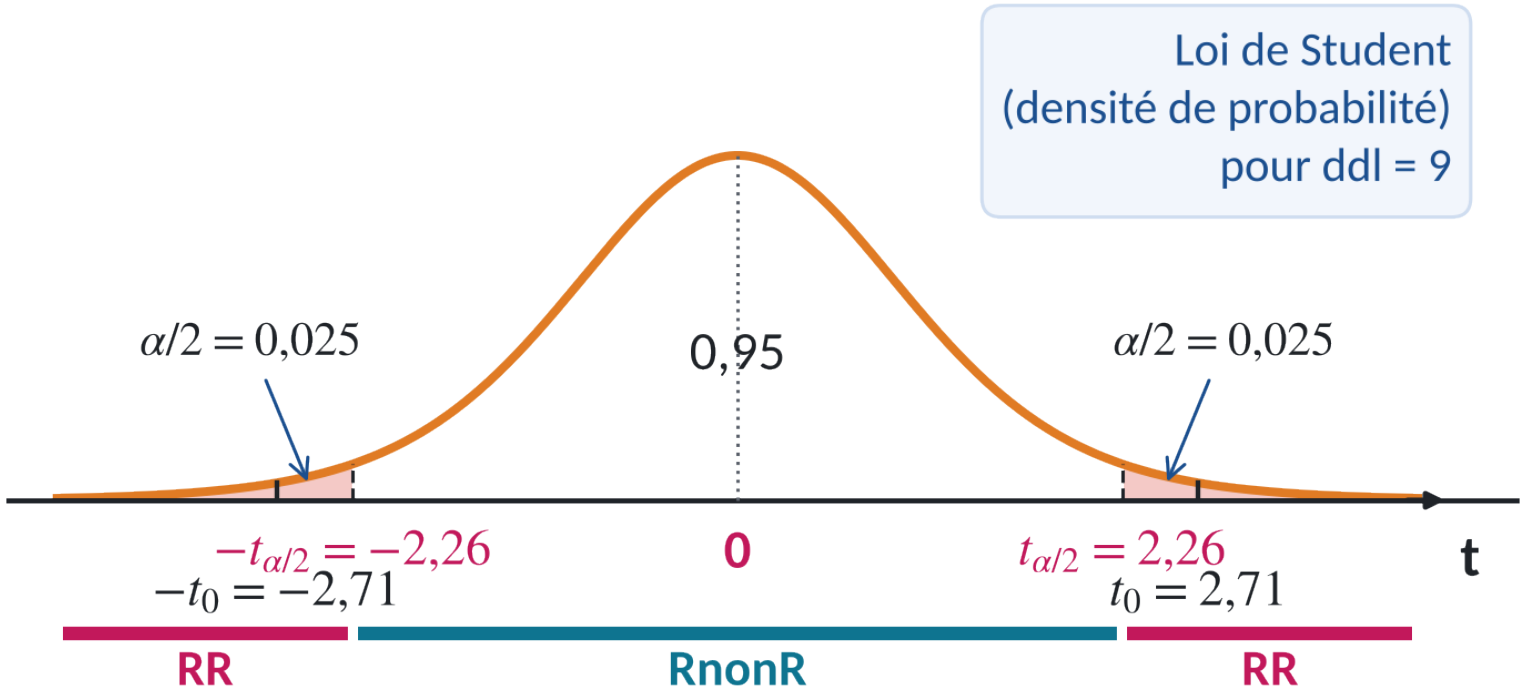
$$RR = (-\infty;\ -t_{9;\ 0{,}025}] \cup [t_{9;\ 0{,}025};\ +\infty)$$
$$RR = (-\infty;\ -2{,}262] \cup [2{,}262;\ +\infty)$$

Étape 5.

$$t_0 = \frac{11{,}1}{12{,}93/\sqrt{10}} = 2{,}71$$

Exemple : test t de Student pour des données appariées

Étape 6.



Décision du test par rapport à la RR : $t_0 \in$ RR, donc **on rejette H_0** → en faveur de H_1

→ la moyenne des différences du rythme cardiaque avant et après l'administration du médicament chez les patients diabétiques est **significativement différente** de 0

↔ **il y a** une **différence significative entre** les **moyennes** du **rythme cardiaque avant** et **après** l'administration du médicament X chez les patients diabétiques.

Exemples d'articles scientifiques avec tests statistiques : test t pour échantillons dépendants (appariés)

article

Changes in Personality and Learning Styles for First Year Medical Students

Nicole J. Borges & Dean X. Parmelee

Medical Science Educator Volume 21 : No. 3. IAMSE http://www.iamse.org/artman/publish/article_611.shtml

Abstract Sixty-five **students** (62 response rate) **completed** the *Neuroticism, Extraversion, Openness to Experience Personality Inventory-Revised and Grasha-Reichmann Student Learning Styles Scale* **at the beginning and end of their first year of medical school.** Paired t-tests showed significant differences for 2 of 5 personality scales and 5 of 6 learning styles scales.

Personality and Learning Styles Scales	Pretest		Posttest			
	Mean	SD	Mean	SD	*t*	*p*
Neuroticism	85.32	20.70	88.91	23.95	-1.647	.104
Extraversion	122.92	19.15	118.12	18.50	3.415	.001*
Openness to Experience	121.25	18.99	121.31	19.11	-.052	.958
Agreeableness	125.29	19.85	126.02	19.19	-.675	.502
Conscientiousness	132.65	15.66	126.40	19.06	4.091	<.001*
Dependent	35.54	4.06	37.12	4.93	-3.110	.003*
Collaborative	37.60	5.72	35.38	5.61	3.652	.001*
Participant	38.25	4.15	36.18	4.32	3.674	<.001*
Avoidant	24.29	5.20	27.52	5.32	-5.511	<.001*
Independent	35.25	3.90	35.00	4.68	.506	.615
Competitive	27.03	5.91	24.98	6.24	2.926	.004*

Table 1. *Results of Paired t-tests for Pre and Post Comparison of Personality and Learning Styles*
**Significant at p < .005 level*

Interprétation du tableau dans l'étude

On observe que : ils comparent différents types de scores de personnalité (p. ex. extraversion…) (dans la première colonne) avec le test : paired t-test (dans le titre), pour voir des différences avant et après la première année (pre and post comparison, dans le titre du tableau) ; pour chaque variable quantitative (score de personnalité), on observe la moyenne (mean), la déviation standard (SD) au début (pre) et à la fin de l'année (post), et le résultat du test t (P).

Énoncez les hypothèses nulle et alternative pour le **test** qui vérifie s'il y a une **différence** statistiquement significative entre les étudiants au **début** et à la **fin** de l'année en ce qui concerne le niveau **moyen** du **style collaboratif d'apprentissage**.

H_0 (hypothèse nulle) : **il n'y a pas** de **différence statistiquement significative entre** les étudiants au **début** de l'année et à la **fin** de l'année en ce qui concerne le niveau **moyen** du **style collaboratif d'apprentissage**

H_1 (hypothèse alternative, test bilatéral) : **il y a** une **différence statistiquement significative entre** les étudiants au **début** de l'année et à la **fin** de l'année en ce qui concerne le niveau **moyen** du **style collaboratif d'apprentissage**

Écrivez le nom du test utilisé pour la comparaison :

test t de Student pour échantillons dépendants/appariés (voir paired t-test dans le tableau).

Interprétation du tableau dans l'étude (suite)

Écrivez la valeur du p du test t (voir la colonne P)

- p = 0,001

Interprétez du point de vue statistique le résultat du test t et argumentez votre réponse

- **Il y a** une **différence statistiquement significative entre** les étudiants au **début** de l'année et à la **fin** de l'année en ce qui concerne le niveau **moyen** du **style collaboratif d'apprentissage**, parce que p = 0,001 < le niveau de signification de 0,05 (on rejette l'hypothèse nulle).

Écrivez les valeurs des moyennes et des écarts-types pour chaque groupe (voir M : moyenne, SD : écart-type/déviation standard)

- avant (pretest dans le tableau) : moyenne = 37,60, écart-type = 5,72
- après (posttest dans le tableau) : moyenne = 35,38, écart-type = 5,61

Écrivez quel groupe a la moyenne du style collaboratif la plus grande

- le groupe avant (voir 37,60 > 35,38)

Exemples de questions

Scénario : regardez le tableau suivant (contenant des résultats partiels) tiré d'un article scientifique médical. Il décrit les caractéristiques de l'embolie pulmonaire chez les patients atteints de COVID-19. Répondez aux questions posées sur les diapositives suivantes :

	All ($n=689$)	No PE ($n=637$)	PE ($n=52$)	p value
Age (years)	67.3 ± 13.2	67.6 ± 13.4	63.8 ± 10.6	0.04
Male gender	487 (69.4)	437 (68.6)	41 (78.8)	0.12
BMI (kg/m^2)	27.2 ± 5.3	27 ± 5.2	29.6 ± 6.3	0.003
Ever smoker	159 (27)	151 (27.7)	8 (18.6)	0.20
Hypertension	398 (56.9)	364 (57.6)	25 (48.1)	0.18
Dyslipidaemia	188 (27.5)	175 (27.7)	13 (25.0)	0.74
Diabetes	157 (23)	144 (22.8)	13 (25.0)	0.72
Heart failure	92 (13.5)	90 (14.2)	2 (3.8)	0.04
Atrial fibrillation	105 (15.4)	102 (16.1)	3 (5.8)	0.05
Coronary artery disease	143 (20.9)	137 (21.7)	6 (11.5)	0.08
COPD	67 (9.8)	64 (10.1)	3 (5.8)	0.31
Chronic kidney disease	127 (18.6)	123 (19.5)	4 (7.7)	0.04
ACEi/ARB therapy	133 (20.6)	123 (20.6)	10 (20.0)	0.91
Oral anticoagulant therapy	90 (14.1)	79 (13.5)	11 (21.6)	0.11
Direct oral anticoagulant	47 (7.4)	40 (6.8)	7 (13.7)	0.07
Vitamin K antagonist	48 (7.5)	43 (7.3)	5 (9.8)	0.52
Statin therapy	176 (27.2)	165 (27.7)	11 (21.6)	0.35
Fever (≥ 37.5 °C)	440 (64.1)	408 (64.3)	32 (62.7)	0.83
Respiratory rate ≥ 22/min	279 (52.0)	253 (50.8)	26 (66.7)	0.06
SBP (mmHg)	129.6 ± 21.5	129.7 ± 21.4	129.2 ± 22.4	0.89
Heart rate (bpm)	86.6 ± 18.1	86.3 ± 18.2	90.7 ± 15.9	0.09
Oxygen saturation (%)	90.5 ± 7.6	90.8 ± 7.2	86.6 ± 10.1	< 0.001
LV ejection fraction (%)	52.5 ± 11.3	52.1 ± 11.7	55.3 ± 8.4	0.12

Data are shown as count (%), mean ± SD or median (interquartile range)

BMI body mass index, *COPD* chronic obstructive pulmonary disease, *ACEi* angiotensin-converting enzyme inhibitor, *ARB* angiotensin receptor blocker, *SBP* systolic blood pressure, *LV* left ventricular

Source :

Ameri P, Inciardi RM, Di Pasquale M, Agostoni P, Bellasi A, Camporotondo R, Canale C, Carubelli V, Carugo S, Catagnano F, Danzi G, Dalla Vecchia L, Giovinazzo S, Gnecchi M, Guazzi M, Iorio A, La Rovere MT, Leonardi S, Maccagni G, Mapelli M, Margonato D, Merlo M, Monzo L, Mortara A, Nuzzi V, Piepoli M, Porto I, Pozzi A, Provenzale G, Sarullo F, Sinagra G, Tedino C, Tomasoni D, Volterrani M, Zaccone G, Lombardi CM, Senni M, Metra M. Pulmonary embolism in patients with COVID-19: characteristics and outcomes in the Cardio-COVID Italy multicenter study. Clin Res Cardiol. 2020 Nov 3:1-9. doi: 10.1007/s00392-020-01766-y.

Lien vers l'article :

https://pubmed.ncbi.nlm.nih.gov/33141251/

Exemples de questions

E1. Lesquelles des affirmations suivantes sont correctes ?

A La variable *Saturation en oxygène (%)* est une variable quantitative

B La variable *Saturation en oxygène (%)* est une variable qualitative

C La variable *Fréquence respiratoire ≥ 22/min* est une variable quantitative

D La variable *Fibrillation atriale* est une variable qualitative

E La variable *Fréquence respiratoire ≥ 22/min* est une variable qualitative

R1. A, D, E

Exemples de questions

E2. Les auteurs ont choisi de montrer :

A la moyenne et l'erreur-type pour la variable *Fréquence cardiaque (bpm)* parce qu'elle suit la distribution gaussienne

B la moyenne et l'écart-type pour la variable *Fréquence cardiaque (bpm)* parce qu'elle suit la distribution gaussienne

C la médiane (intervalle interquartile, IQR) pour la variable *Fréquence cardiaque (bpm)* parce qu'elle ne suit pas la distribution gaussienne

D les fréquences absolue et relative pour la variable *Fréquence respiratoire ≥ 22/min*

E le nombre de cas et les pourcentages pour la variable *Fréquence respiratoire ≥ 22/min*

R2. B, D, E

Exemples de questions

E3. Les valeurs de saturation en oxygène (%) ont été comparées entre les deux groupes de sujets au moyen du :

A test t de Student pour échantillons indépendants

B test t de Student pour échantillons dépendants

C test Z pour échantillons indépendants

R3. A

Exemples de questions

E4. La formulation des hypothèses statistiques concernant la comparaison des valeurs de la saturation en oxygène (SO_2, %) :

A H_0 : **Il n'y a pas** de **différence significative entre** les **moyennes** de SO_2 chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

B H_0 : **Il n'y a pas** de **différence significative entre** les **médianes** de SO_2 chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

C H_1 : **Il y a** une **différence significative entre** les **distributions** de SO_2 chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

D H_1 : **Il y a** une **différence significative entre** les **fréquences** de SO_2 chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

E H_1 : **Il y a** une **différence significative entre** les **moyennes** de SO_2 chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

R4. A, E

Exemples de questions

E5. Interprétation de la p-valeur concernant la comparaison des valeurs de SO_2 :

A. parce que $p < 0,05$, on rejette H_0, donc nous sommes en faveur de H_1

B. parce que $p < 0,05$, on rejette H_1, donc nous sommes en faveur de H_0

C. parce que $p \geq 0,05$, on ne rejette pas H_0, donc on accepte H_0

D. **Il y a** une **différence significative entre** les **moyennes** de **SO_2** chez les patients atteints de COVID-19 **qui ont développé une embolie pulmonaire** et **ceux qui n'ont pas d'embolie pulmonaire**

E. la différence observée entre les moyennes de SO_2 n'est pas significative et elle est due aux fluctuations d'échantillonnage

R5. A, D

Exemples de questions

E6. Interprétation des résultats concernant la comparaison des variances de SO_2 (on considère $\alpha = 0,05$, valeur critique du quantile = 1,45)

A la région de rejet (RR) du test est $(-\infty ; -1,45] \cup [1,45 ; +\infty)$

B la région de rejet du test (RR) est $[1,45 ; +\infty)$

C la région de non-rejet du test (RnonR) est $(0 ; 1,45)$

D la statistique du test utilisé est égale à 1,968

E **Il n'y a pas** de **différence significative entre** les **variances** de **SO_2** sur les **deux populations** car la statistique appartient à la RR

R6. B, C, D

Solution

	NoPE	PE
taille de l'échantillon	$n_1 = 637$	$n_2 = 52$
écart-type (S)	7,2	10,1
variances (S^2)	51,84	102,01

$$\mathbf{F} = \begin{cases} \frac{S_2^2}{S_1^2}, & \text{pour } S_2^2 > S_1^2 \\ \frac{S_1^2}{S_2^2}, & \text{pour } S_1^2 > S_2^2 \end{cases}$$

F suit la loi de Fisher avec (51, 636) d.d.l.

$$RR = [F_{51,\ 636;\ 0,025};\ \infty) = [1,45;\ \infty)$$

Exemples de questions

E7. Interprétation des résultats concernant la comparaison des moyennes de SO_2 (on considère $\alpha = 0,05$, valeur critique du quantile = 2,004)

A la région de rejet (RR) du test est $(-\infty ; -2,004] \cup [2,004 ; +\infty)$

B la région de rejet (RR) du test est $[2,004 ; +\infty)$

C la région de non-rejet du test (RnonR) est $(-2,004 ; +2,004)$

D la statistique du test utilisé est égale à 2,938

E **Il y a** une **différence significative entre** les **moyennes** de **SO_2** sur les **deux populations** car la statistique appartient à la RR

R7. A, C, D, E

Solution

	NoPE	PE
taille de l'échantillon	$n_1 = 637$	$n_2 = 52$
moyennes	90,8	86,6
écart-type (S)	7,2	10,1
variances (S^2)	51,84	102,01

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{90{,}8 - 86{,}6}{\sqrt{\frac{51{,}84}{637} + \frac{102{,}01}{52}}}$$

$t_{\alpha/2}$ = **2,004 (d.d.l. de Welch ≈ 55)**

RR = $(-\infty ; -2{,}004] \cup [2{,}004 ; +\infty)$

$t_{observé} = t_0 = 2{,}938$

Synthèse : réaliser les tests du cours dans Jamovi

Situation / objectif	Test (conditions)	Jamovi : Analyses → menu → analyse	Options à cocher	À lire dans les résultats
Normalité des données (condition)	Graphiques, coefficients, Shapiro-Wilk	**Exploration → Descriptives**	Statistics : **Skewness**, **Kurtosis**, **Shapiro-Wilk** ; Plots : **Histogram**, **Box plot**, **Q-Q** (ou dans le test t : **Assumption Checks → Normality test**, **Q-Q plot**)	W, p ; p ≤ 0,05 → normalité rejetée
Égalité de deux variances (condition)	Test F de Fisher (données normales)	**T-Tests → Independent Samples T-Test**	**Assumption Checks → Homogeneity test** (test de **Levene's**)	F, ddl, p ; p ≤ 0,05 → variances inégales (pas de F de Fisher dans jamovi)
Comparer 2 moyennes, échantillons indépendants, variances égales	Test t de Student (normalité, variances égales)	**T-Tests → Independent Samples T-Test**	Tests : **Student's** ; Dependent Variables + Grouping Variable ; **Mean difference**, **Confidence interval**, **Descriptives**	t, df = $n_1 + n_2 - 2$, p ; p ≤ 0,05 → rejet de H_0
Idem, variances inégales	Test t de Welch	**T-Tests → Independent Samples T-Test**	Tests : **Welch's**	t, df de Welch, p ; p ≤ 0,05 → rejet de H_0
Comparer 2 moyennes, échantillons appariés	Test t apparié (différences normales)	**T-Tests → Paired Samples T-Test**	Paired Variables (avant, après) ; Tests : **Student's** ; **Mean difference**, **Confidence interval**	t, df = n − 1, p ; p ≤ 0,05 → rejet de H_0

Test Z (σ connue) : pas d'analyse dédiée dans jamovi, calcul manuel.
Interface jamovi en anglais (version 2.x), onglet Analyses.