Chapitre 1 : Tests d'hypothèses pour une moyenne

1. Introduction :

1. La machine produisant des comprimés est-elle déréglée ?
2. La fréquence de grippe dans une région est-elle différente de celle d'une autre région ?
3. Le régime alimentaire est-il plus efficace pour les hommes que pour les femmes ?

   - Pour répondre à ces questions en statistique, on utilise une procédure appelée « test d'hypothèses ».
   - Il s'agit de parvenir à une décision statistique sur une population compte tenu des résultats expérimentaux observés sur un échantillon représentatif extrait de la population concernée.

1.1. Estimation ponctuelle :

$$\left\{\begin{matrix} \text{estimation de la} \\ \text{moyenne} \\ \mu \ \text{de la population} \end{matrix}\right\} = m = \bar{x}, \ \text{moyenne de l'éch.}$$

$$\left\{\begin{matrix} \text{estimation de la} \\ \text{variance} \\ \sigma^2 \ \text{de la population} \end{matrix}\right\} = S^2 = \frac{n}{n-1} S_{ech}^2.$$

$$\left\{\begin{matrix} \text{estimation de la} \\ \text{probabilité} \\ p \ \text{de la population} \end{matrix}\right\} = f, \ \text{fréquence observée.}$$

2. Définitions et composantes d'un test formel :

<u>Hypothèse statistique :</u> une affirmation concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population.

Test d'hypothèse : un test d'hypothèse (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base des résultats d'échantillon de faire un choix entre deux hypothèses statistiques.

Hypothèse nulle (H_0) et hypothèse alternative (H_1):

L'hypothèse selon laquelle on fixe à priori un paramètre de la population à une valeur particulière s'appelle l'hypothèse nulle et est notée H_0. N'importe quelle autre hypothèse qui diffère de H_0 s'appelle l'hypothèse alternative (ou contre-hypothèse) et est notée H_1.

C'est l'hypothèse nulle qui est soumise au test et toute la démarche du test s'effectue en considérant cette hypothèse comme vraie.

Statistique du test : c'est une valeur calculée à partir des données de l'échantillon et elle est utilisée dans la prise de décision du rejet ou non de l'hypothèse nulle.

Région critique : la région critique (ou zone de rejet est l'ensemble des valeurs de la statistique qui nous font rejeter H_0.

Niveau de significativité : le niveau de significativité (ou seuil ou risque) noté α est la probabilité que la statistique de test tombe dans la région critique.

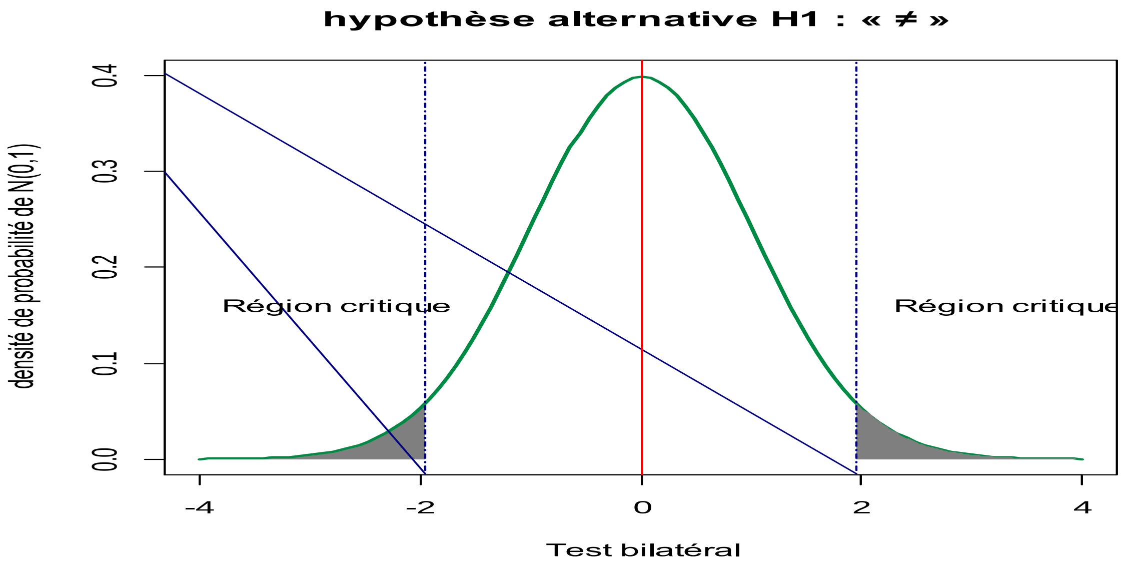
Les choix courants pour α sont $0{,}05$; $0{,}01$; $0{,}1$.

Valeur critique :

- la valeur critique est une valeur qui sépare la région critique (où on rejette H_0) des autres valeurs de la statistique de test.
- Elle dépend de la nature de H_0, de la distribution d' échantillonnage à appliquer et du niveau de significativité α.

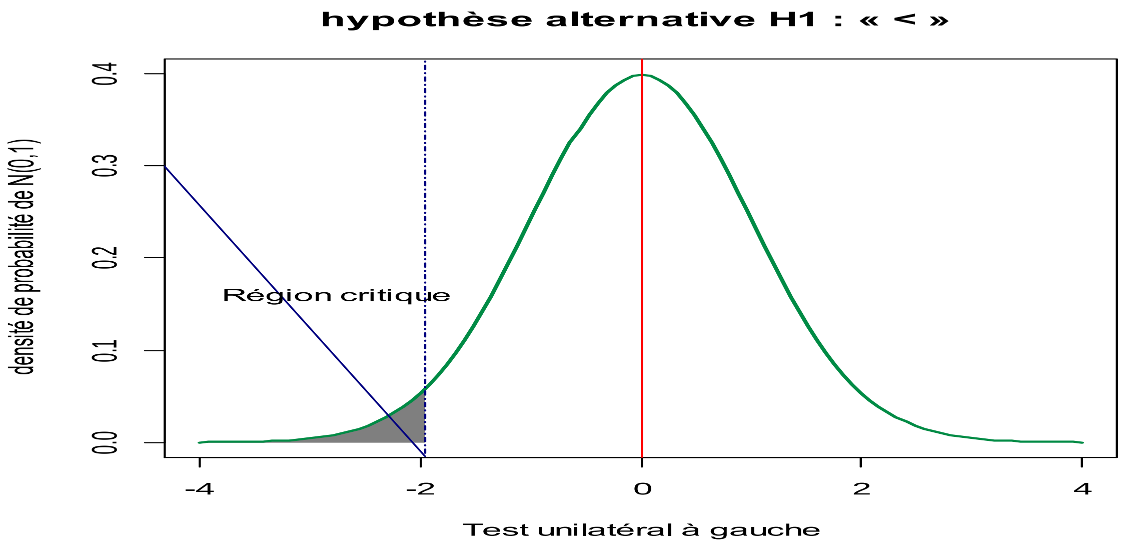
Test bilatéral :

H_1 : « $\neq$ » la région critique est dans les deux régions extrêmes sous la courbe.



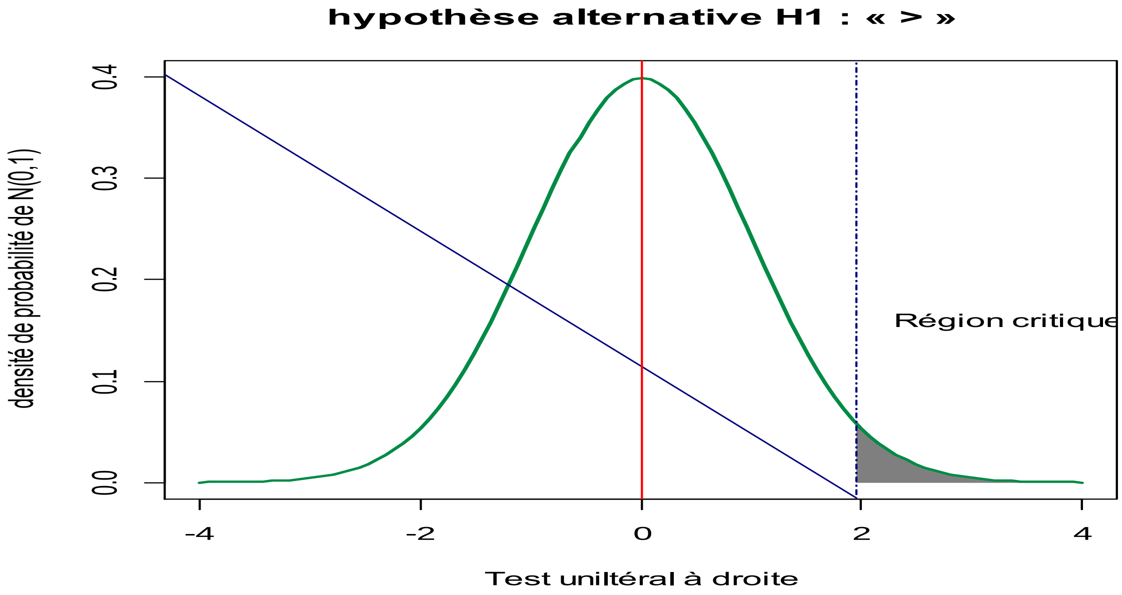
Test unilatéral à gauche :

H_1 : « < » la région critique est dans la région extrême à gauche sous la courbe.



Test unilatéral à droite :

H_1 : « > » la région critique est dans la région extrême à droite sous la courbe.



Décision et conclusion :

- La décision est l'une des deux affirmations :

1. Rejet de H_0.
2. Echec de rejet de H_0.

- Selon la méthode traditionnelle on décide :

1. « Rejet de H_0 » si la statistique de test tombe dans la région critique.
2. « Echec de rejet de H_0 » si la statistique de test ne tombe pas dans la région critique.

Tests paramétriques :

Ce sont des tests en général basés sur la considération de la loi normale. Plus précisément, ce sont des tests pour lesquels la distribution de la variable de décision n'est connue que sous certaines conditions.

Tests non paramétriques :

Ce sont des tests valables quelle que soit la distribution des variables étudiées. Plus exactement, ce sont des tests pour lesquels la distribution de la variable de décision ne dépend pas des paramètres caractéristiques de la population (aucune condition d'application du test).

3. Test d'hypothèses avec un échantillon

3.1. Test d'hypothèse pour une moyenne :

	σ^2 connue	**σ^2 inconnue**
Conditions d'applications du test	• Echantillon aléatoire simple • Une des deux conditions est satisfaite : Population normale ou $n > 30$	• Echantillon aléatoire simple • Une des deux conditions est satisfaite : Population normale ou $n > 30$
Statistique du test	$Z_{calc} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$t_{calc} = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$
Valeurs critiques	Z_{table} de la table de la loi normale	t_{table} de la table de la loi de Student avec $d.d.l = n - 1$

<u>Exemple 1 :</u>

Un échantillon de **106** températures humaines dont la moyenne est **36.78°C**. Supposer que l'échantillon est aléatoire simple, que l'écart type est connu et vaut **0.34**.

Utilisez un niveau de significativité de **0.05** pour tester la croyance commune que la température d'un adulte en bonne santé est **37.0°C.**

<u>Solution :</u>

Les conditions requises sont satisfaites **:**

- L'échantillon est aléatoire simple

- La variance de la population est connue

$$\sigma^2 = (\mathbf{0.34}\ )^2$$

- Une des deux conditions est satisfaite : la population est normalement distribuée ou n>30.

Ici $\boldsymbol{n = 106 > 30}$.

1) Ecriture des hypothèses

$\boldsymbol{H_0 \colon \mu = 37}$

$\boldsymbol{H_1 \colon \mu \neq 37}$

2) Calcul de la statistique de test

$$\boldsymbol{Z_{calc} = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{36.78 - 37}{\frac{0.34}{\sqrt{106}}} = -6.66}$$

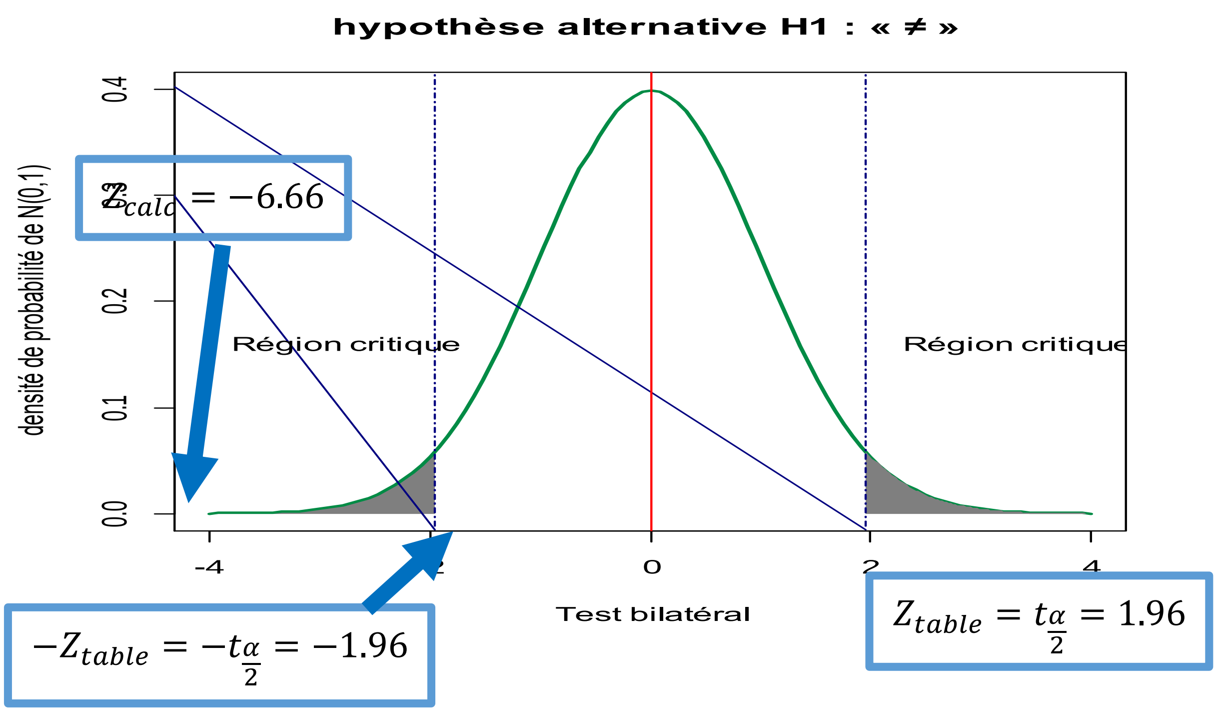
3) Type du test

Il s'agit d'un test bilatéral car

$$\boldsymbol{H_1 \colon \mu \neq 37}$$

On doit comparer la statistique $\boldsymbol{Z_{calc}}$ aux valeurs critiques de la loi normale pour α = 0.05.

4) Détermination de la zone critique



$$F(t_{\alpha/2})=1-\alpha/2 = 0.975 \xrightarrow{la\ table\ \ N(0,1)} t_{\alpha/2} = 1.96$$

$z= -6.66 < -t_{\alpha/2} =-1.96$ d'où le rejet de H_0

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

5) Conclusion

On conclut qu'il y a suffisamment de preuves pour dire que la moyenne des températures diffère de 37.0°C.

<u>Exemple 2 :</u>

On a les données de température de 12 personnes en bonne santé, on a obtenu les températures listées ci-dessous.

36.67 36.39 37.00 37.11 36.67 36.94

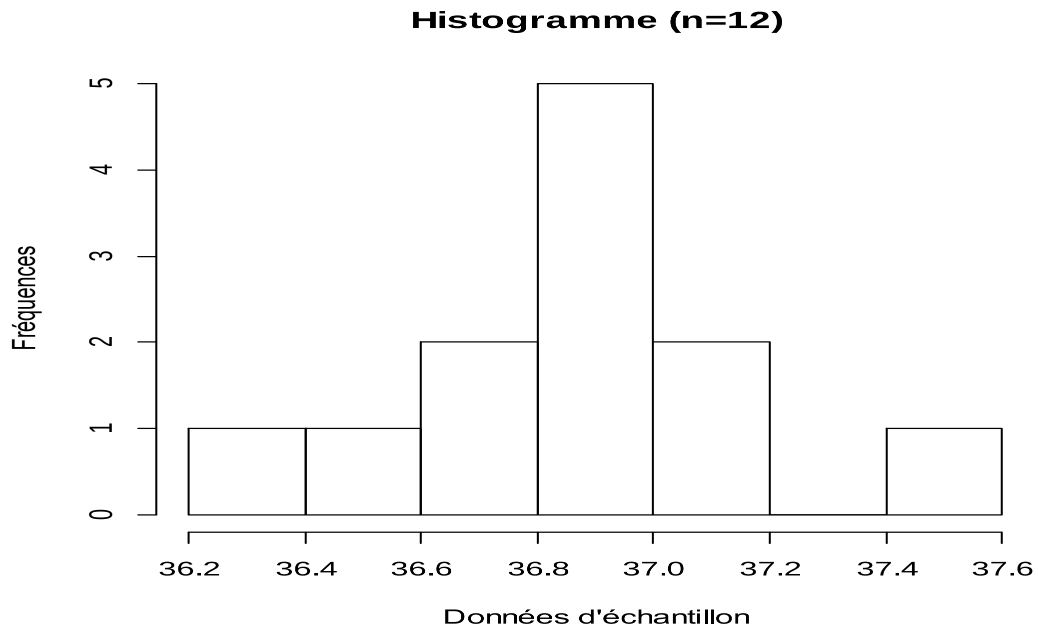
37.00 37.44 36.89 37.06 37.00 36.44

Utiliser un niveau de significativité de 0.05 pour tester l'hypothèse que la moyenne de ces températures est issue d'une population dont la moyenne est inférieure à 37.0°C.

Solution :

- il faut d'abord vérifier que les conditions requises sont satisfaites.
- On a un échantillon aléatoire simple, reste à vérifier la condition de normalité puisque n=12<30.

L'histogramme suivant fourni par le logiciel **R** montre que les données suivent une distribution pas très éloignée de la loi normale.



1) Ecriture des hypothèses

$\boldsymbol{H_0}$: $\boldsymbol{\mu = 37}$

$\boldsymbol{H_1}$: $\boldsymbol{\mu < 37}$

2) Calcul de la statistique de test ($\bar{x}$ et S sont calculés à partir des données de l'échantillon)

$$\boldsymbol{t_{calc} = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{36.88 - 37}{\frac{0.297}{\sqrt{12}}} = -1.4}$$

3) Type du test

Il s'agit d'un test unilatéral à gauche car

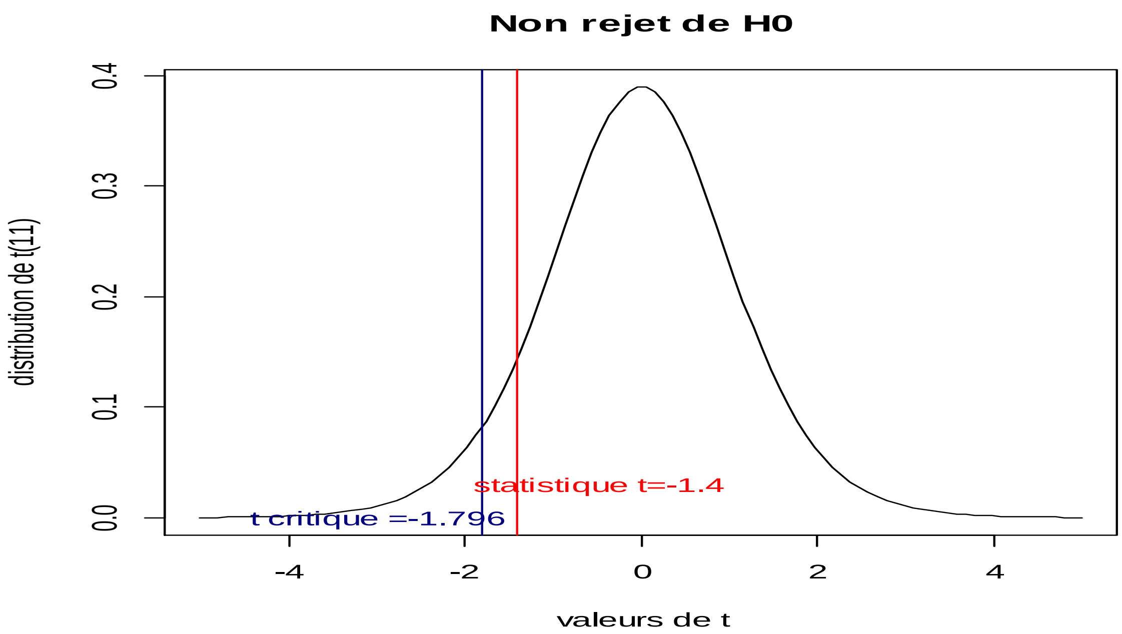
$$\boldsymbol{H_1: \mu < 37}$$

On doit comparer la statistique $\boldsymbol{t_{calc}}$ à la valeur critique de la loi de Student à

$n - 1 = 11$ degrés de liberté pour
α = 0.05 et 2α = 0.1
D'après la table

$$t_{table} = t_{0.1,11} = 1.796$$

4) Détermination de la zone critique



Comme $t_{calc} = -1.4$ n'est pas dans la région critique, on ne peut pas rejeter H_0.

ν \ α	0,900	0,500	0,300	0,200	0,100	0,050	0,020	0,010	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6193
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,1303	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,1286	0,6974	1,0877	1,3634	$\boxed{1,7959}$	2,2010	2,7181	3,1058	4,4370
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178

5) Conclusion

il n'y a pas suffisamment de preuves pour dire que l'échantillon vient d'une population dont la moyenne est inférieur à 37°C.

La moyenne peut être bien inférieure à 37 °C mais les 12 valeurs de l'échantillon ne fournissent pas assez d'éléments pour le confirmer.

3.2. Test d'hypothèse pour une variance :

Conditions d'application du test	• **Echantillon aléatoire simple** • **Population normalement distribuée**
Statistique du test	$\chi^2_{calc} = \frac{(n-1)S^2}{\sigma^2}$
Valeurs critiques	Table du khi-deux avec $d.d.l = n - 1$

<u>Exemple :</u>

Dans un laboratoire d'analyse médicale on effectue le dosage de calcium sérique par une méthode colorimétrique dont l'écart-type sur la mesure est de 1,1 mg/L. Ce dosage suit une loi normale. Après une remise à niveau de l'appareil de mesure, le directeur de ce laboratoire veut vérifier que l'écart-type sur la mesure n'est pas augmenté.

Il fait refaire par un même technicien, dans les mêmes conditions, 32 dosages d'un même prélèvement sérique, il obtient une variance de 1,3975 mg/L.

Pensez-vous que l'écart-type après intervention est supérieur à l'écart-type initial ? (on supposera que la remise à niveau ne modifie pas la valeur moyenne des mesures).

<u>Solution :</u>

- On a un échantillon aléatoire simple.

- La distribution est normale.

Donc les conditions requises pour le test sont vérifiées.

$s_{ech}^2 = 1{,}3975 \Rightarrow S^2 = \frac{n}{n-1} s_{ech}^2 = \frac{32}{31}(1{,}3975) = 1{,}44258$

Et $S = \sqrt{1{,}44258} = 1{,}201$, $\sigma^2 = (1{,}1)^2 = 1{,}21$

1) Ecriture des hypothèses

$\boldsymbol{H_0 : \sigma = 1,1}$

$\boldsymbol{H_1 : \sigma > 1,1}$

2) Calcul de la statistique de test

$$\chi_{calc}^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{31(1{,}44258\,)}{1{,}21} = 36{,}96$$

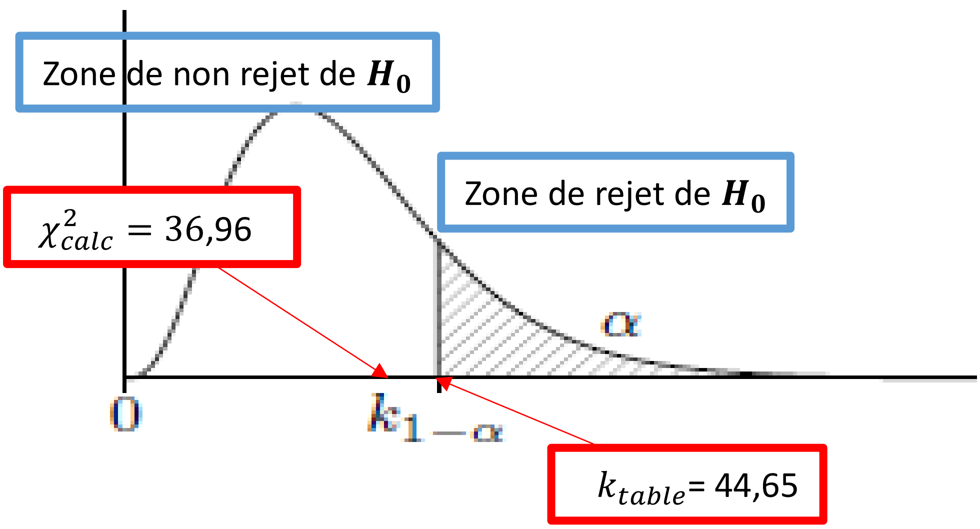
3) Type du test

Il s'agit d'un test unilatéral à droite car

$$\boldsymbol{H_1 : p > 1,1}$$

On doit comparer la statistique χ_{calc}^2 aux valeurs critiques de la loi khi-deux pour α = 0.05

4) Détermination de la zone critique



Sur la table du khi-deux, le d.d.l=31 n'existe pas, donc on doit faire une approximation par une loi normale.

Pour $n > 30$, $U = \sqrt{2\chi^2} - \sqrt{2\nu - 1} \sim \mathcal{N}(0,1)$

Donc $Z_{table} = \sqrt{2k} - \sqrt{2\nu - 1}$

Pour un test unilatéral à droite et α = 0.05, on a $Z_{table} = 1,64 \Rightarrow k = 44,65 > \chi^2_{calc} = 36,96$. D'où le non rejet de $\boldsymbol{H_0}$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

5) Conclusion

On ne peut pas conclure au risque de 5% , que l'écart-type après intervention est supérieur à l'écart-type initial .

4.Test d'hypothèses avec deux échantillons :

Dans la réalité, il y a des situations dans lesquelles il est nécessaire de comparer deux échantillons issues de deux populations afin de conduire des inférences à propos de ces populations.

4.1. Inférence sur deux moyennes : échantillons indépendants

Deux échantillons sont indépendants si les valeurs d'une population ne sont pas liées aux valeurs de l'autre population.

	Variances connues	Variances inconnues	Variances inconnues et égales $\sigma_1^2 = \sigma_2^2$
Conditions d'application du test	• Les deux échantillons sont indépendants. • Les deux échantillons sont aléatoires simples. • Une des deux conditions est satisfaite : les deux échantillons sont grands **<u>ou</u>** ils sont issus des populations possédant des distributions normales.		
Statistique du test	$Z_{calc} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$t_{calc} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$t_{calc} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{S_c^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$
Valeurs critiques	Loi normale	Loi de Student de $d.d.l = \min(n_1 - 1,\ n_2 - 1)$	Loi de Student de $d.d.l = (n_1 - 1) + (n_2 - 1)$

Remarque :

- σ_1^2 et σ_2^2 sont rarement connues en réalité.
- Même quand les valeurs spécifiques des variances ne sont pas connues, s'il est possible de considérer qu'elles ont la même valeur, on peut avoir une estimation de la variance commune :

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

- Pour déterminer si les deux variances sont égales, on utilise un test de comparaison de deux variances qui sera traité dans la suite du chapitre.

<u>Exemple :</u>

Lors d'une expérience à tester l'efficacité de la paroxétine pour traiter la maladie bipolaire, des mesures ont été réalisées sur des sujets en utilisant l'échelle de dépression de Hamilton avec les résultats donnés ci-dessous.

Groupe placebo n=43, $\bar{x}$=21.57 s=3.87

Groupe traité n=33, $\bar{x}$ = 20.38 s=3.91

Utiliser un niveau de significativité de 0.05 pour tester l'affirmation que le groupe traité et le groupe placebo viennent d'une population avec la même moyenne. Interpréter le résultat.

<u>Solution :</u>

On vérifie les conditions d'application du test,

- les deux échantillons sont indépendants et issus d'un tirage aléatoire simple.
- les échantillons sont de taille supérieure à 30 (de grandes tailles)

1) Ecriture des hypothèses :

$\boldsymbol{H_0}$: $\boldsymbol{\mu_1 = \mu_2}$

$\boldsymbol{H_1}$: $\boldsymbol{\mu_1 \neq \mu_2}$

2) Calcul de statistique : (cas de variances inconnues)

$$t_{calc} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{21{,}57 - 20{,}38}{\sqrt{\frac{(3{,}87)^2}{43} + \frac{(3{,}91)^2}{33}}} = 1{,}321$$

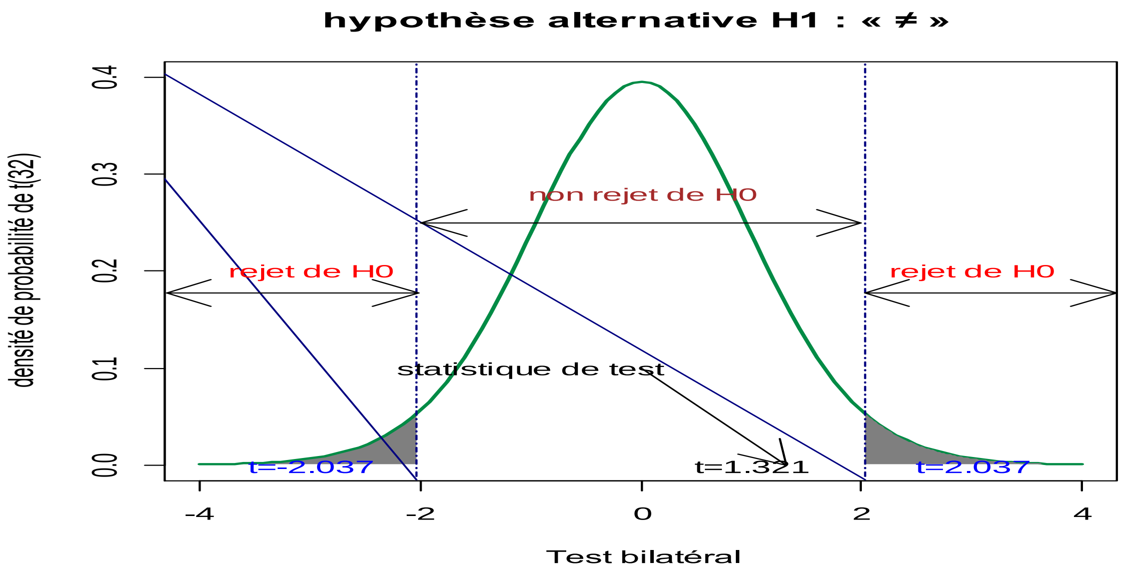
3) Type du test :

Il s'agit d'un test bilatéral car $\boldsymbol{H_1}: \boldsymbol{\mu_1 \neq \mu_2}$

On doit comparer la statistique $\boldsymbol{t_{calc}}$ aux valeurs critiques de la loi de Student pour α = 0.05 et $d.d.l = \min(33 - 1,\ \ 43 - 1) = 32.$

4) Détermination de la zone critique :

$t_{\alpha,\nu} = t_{0.05,32} = 2{,}037 > 1{,}321 \Rightarrow non\,rejet\;deH_0$



5) Conclusion et interprétation :

Au risque de 5%, on ne peut pas rejeter l'affirmation que le groupe traité et le groupe placebo viennent d'une population avec la même moyenne.

Au risque de 5%, on ne peut pas mettre en évidence l'efficacité de la paroxétine pour le traitement de la maladie bipolaire.

4.2. Inférence sur deux moyennes : échantillons appariés

Deux échantillons sont dits appariés lorsqu'ils portent :

- sur les mêmes individus (ex : on mesure le taux de glycémie d'un groupe de sujets diabétiques avant et après traitement).
- sur des individus ayant au moins un caractère semblable (ex : deux groupes de personnes appariés en âge, pour lesquels on dose le cholestérol à un instant donné ; on a donc deux séries de données provenant de deux groupes d'individus différents mais de même âge).

Conditions d'applications :

- Les données sont appariées.
- Les échantillons sont aléatoires simples.
- Une ou les deux conditions sont satisfaite s:
- le nombre de paires est grand

Ou

- les paires de valeurs proviennent de populations dont la distribution est approximativement normale.

Notations :

d: différence individuelle entre les valeurs d'une paire.

μ_d: valeur moyenne des différences pour la population de toutes les paires.

$\bar{d}$: valeur moyenne des différences.

S_d: écart-type des différences pour les données appariées de l'échantillon.

n: nombre de paires.

Statistique du test :

$$t_{calc} = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$

Valeurs critiques :

$t_{critique}$ de table de Student de $d.d.l = n - 1$.

Exemple : « efficacité de l'hypnose pour réduire la douleur»

Une étude a cherché à mesurer l'efficacité de l'hypnose pour réduire la douleur. Les résultats pour les sujets aléatoires sont donnés dans le tableau ci-dessous. Les valeurs concernent des mesures avant et après hypnose.

L'hypnose semble t- elle être un bon traitement pour réduire la douleur ?

Sujet	A	B	C	D	E	F	G	H
Avant	6.6	6.5	9	10.3	11.3	8.1	6.3	11.6
Après	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

Solution :

Les données sont liées par paires car ce sont les mesures prises sur les mêmes individus (avant et après l'hypnose) et on suppose que les échantillons sont issus de populations distribuées normalement.

1) Écriture des hypothèses :

$\boldsymbol{H_0}$: $\boldsymbol{\mu_d = 0}$

$\boldsymbol{H_1}$: $\boldsymbol{\mu_d > 0}$

2) Calcul de statistique :

On introduit une nouvelle variable des différences notée d, les valeurs de d sont:

-0.2 4.1 1.6 1.8 3.2 2.0 2.9 9.6

On calcule :

$$\bar{d} = 3.125 \quad \text{et} \quad S_d = 2.912$$

Donc

$$t_{calc} = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{3.125}{\frac{2.912}{\sqrt{8}}} = 3.035$$

3) Valeur critique et décision :

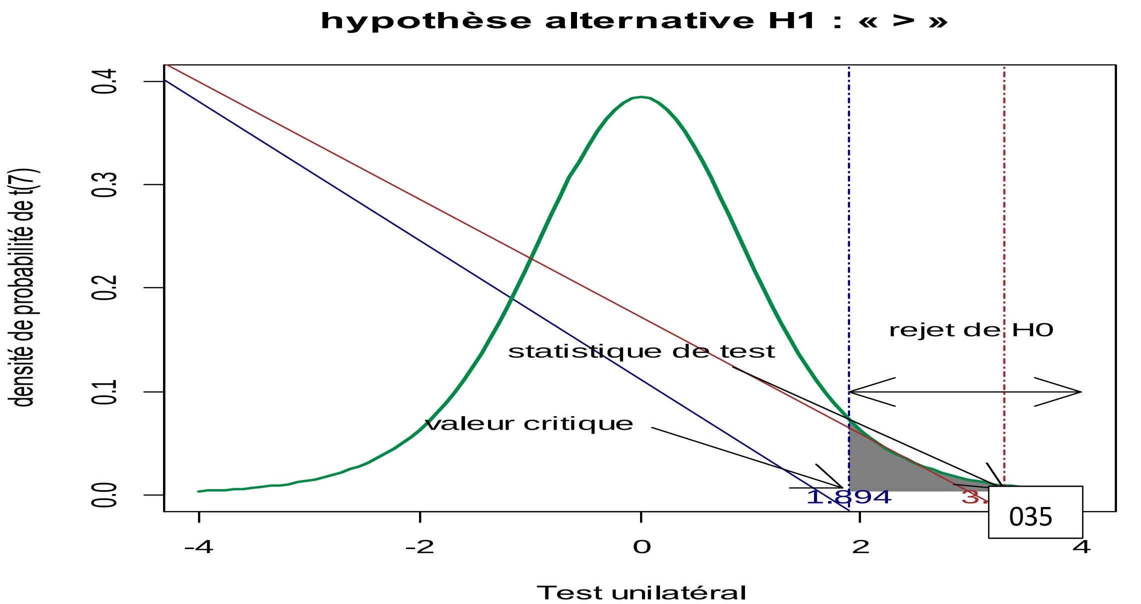
On a un test unilatéral à droite ($\boldsymbol{H_1}$: $\boldsymbol{\mu_d > 0}$).

La valeur critique à partir de la table de Student de $d.d.l = 8 - 1 = 7$ et la colonne $2 * 0.05 = 0.1$

$t_{critique} = t(0.1\ ,\ 7) = 1.894$, donc

$t_{calc} = 3.035 > 1.894$ d' où le rejet de $\boldsymbol{H_0}$

ν \ α	0,900	0,500	0,300	0,200	0,100	0,050	0,020	0,010	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6193
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,1303	0,7111	1,1192	1,4149	$\boxed{1{,}8946}$	2,3646	2,9980	3,4995	5,4079
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,1286	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178



4) Conclusion :

Il y a suffisamment de preuves pour confirmer que les mesures de douleur sont plus basses après hypnose.

L'hypnose semble être un bon traitement pour réduire la douleur.

4.3. Inférence sur deux variances :

Les calculs seront plus simplifiés si nous considérons les deux échantillons de telle façon que S_1^2 est la plus grande des deux variances.

Conditions d'applications :

- Les deux populations sont indépendantes.
- Les deux populations ont une distribution normale (même si elles sont grandes).

Notations :

- S_1^2: la plus grande des variances des deux échantillons.
- n_1: taille de l'échantillon de la plus grande variance.
- Les symboles S_2^2, n_2 sont utilisés pour l'autre échantillon et population.

Statistique du test :

$$F = \frac{S_1^2}{S_2^2} \quad (> 1 \, car \; S_1^2 > S_2^2).$$

Valeurs critiques :

Table de Fisher de degrés de liberté $(n_1 - 1, \; n_2 - 1)$.

<u>Exemple :</u> « Calcium et pression sanguine »

Des données ont été collectées au cours d'une étude sur les suppléments calciques et leurs effets sur la pression sanguine. Un groupe placebo et un groupe calcium ont commencé l'étude par une mesure de pression sanguine.

On a obtenu les résultats suivants

	Effectif	Ecart type
Placebo	n=13	s_1= 9.46
Calcium	n=15	s_2 = 8.469

A un niveau de significativité de 0.05, tester l'affirmation que les deux échantillons sont issus de populations de mêmes écart-type.

<u>Solution :</u>

1) Conditions requises :

Nous vérifions si les conditions sont satisfaites ;

- Les deux échantillons sont indépendants.
- Supposons que les échantillons viennent de populations normales.

2) Écriture des hypothèses :

H_0 : $\boldsymbol{\sigma_1^2 = \sigma_2^2}$

H_1 : $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$

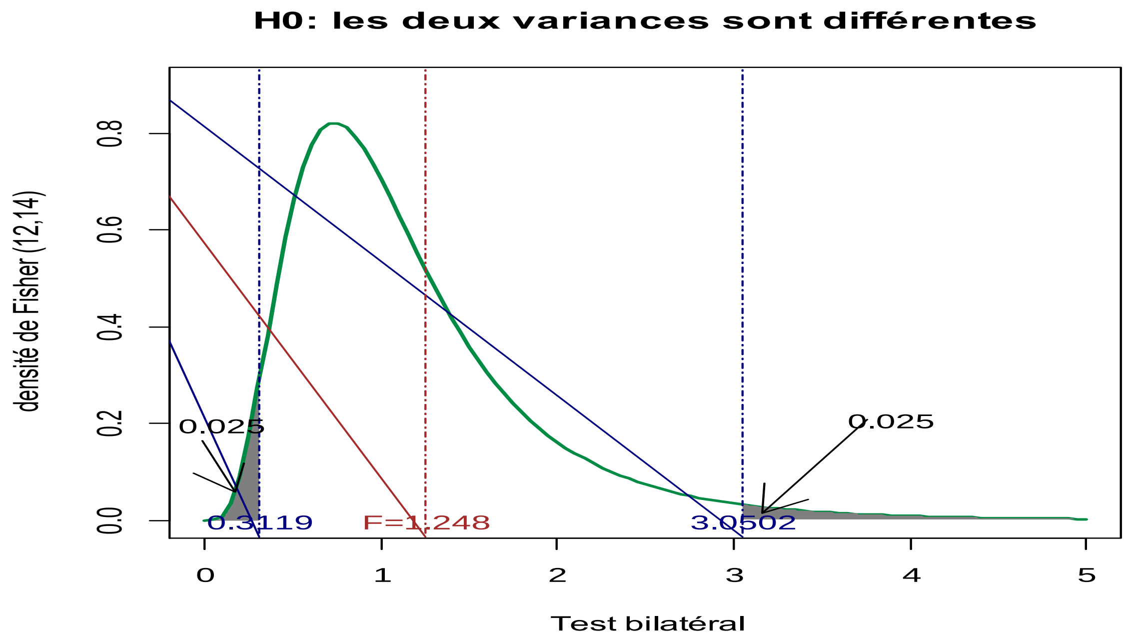
3) Statistique du test :

$$F = \frac{9.46^2}{8.469^2} = 1.248$$

4) Valeur critique :

Il s'agit d'un test bilatéral (**H_1 :** $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$) avec une aire de 0.025=(0.05/2), on compare F à la valeur critique qui se trouve à droite et qui correspond à 3.0502 .

(table de Fisher pour $\alpha = 0.025$, d.d.l.1=12 et d.d.l.2=14).



ν_2 \ ν_1	1	2	3	4	5	6	8	10	15	20	30
1	648	800	864	900	922	937	957	969	985	993	1 001
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,5	14,4	14,3	14,2	14,1
4	12,2	10,6	9,98	9,60	9,36	9,20	8,98	8,84	8,66	8,56	8,46
5	10,0	8,43	7,76	7,39	7,15	6,98	6,76	6,62	6,43	6,33	6,23
6	8,81	7,26	6,60	6,23	5,99	5,82	5,60	5,46	5,27	5,17	5,07
7	8,07	6,54	5,89	5,52	5,29	5,12	4,90	4,76	4,57	4,47	4,36
8	7,57	6,06	5,42	5,05	4,82	4,65	4,43	4,30	4,10	4,00	3,89
9	7,21	5,71	5,08	4,72	4,48	4,32	4,10	3,96	3,77	3,67	3,56
10	6,94	5,46	4,83	4,47	4,24	4,07	3,85	3,72	3,52	3,42	3,31
11	6,72	5,26	4,63	4,28	4,04	3,88	3,66	3,53	3,33	3,23	3,12
12	6,55	5,10	4,47	4,12	3,89	3,73	3,51	3,37	3,18	3,07	2,96
13	6,41	4,97	4,35	4,00	3,77	3,60	3,39	3,25	3,05	2,95	2,84
14	6,30	4,86	4,24	3,89	3,66	3,50	3,29	$\boxed{3,15}$	$\boxed{2,95}$	2,84	2,73

5) Décision :

F < 3.0502, F=1.248 ne se situe pas dans la région critique. Ainsi nous ne pouvons pas rejeter H_0.

6) conclusion :

Il n'y a pas suffisamment de preuves pour rejeter l'hypothèse nulle d'égalité des variances.