

Tests d'hypothèses statistiques paramétriques

I. Test d'hypothèses avec un échantillon :

1.1 Définitions :

- En statistique , une hypothèse est une affirmation ou un énoncé à propos d'une population.
- Un test statistique (ou test de significativité) est une procédure standard pour tester un énoncé à propos d'une propriété d'une population et prendre une décision basée sur les résultats de l'échantillon.

1.2/ Etapes d'un test statistique :

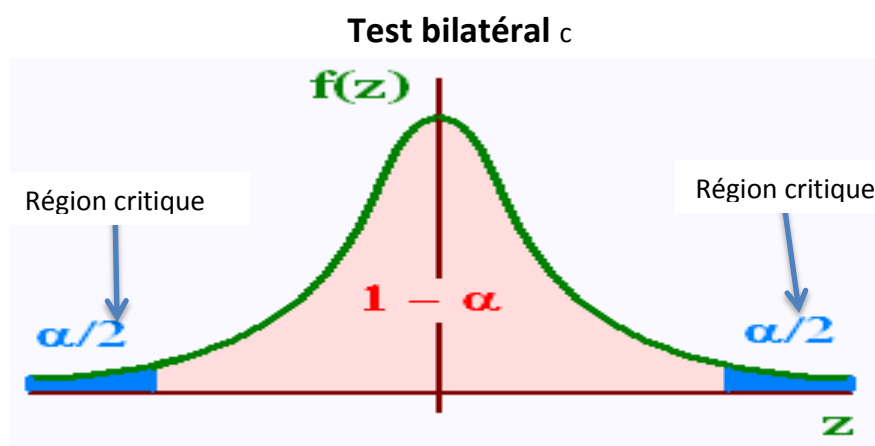
- **L'hypothèse nulle** : (notée H_0) est l'affirmation que la valeur d'un paramètre d'une population est égale à une certaine valeur (exemple : $H_0 : p = 0.50 ; \mu = 1 ; \sigma = 15$)
- **L'hypothèse alternative** : (notée H_1) est l'affirmation que le paramètre a une valeur différente de celle de H_0 , la forme symbolique de H_1 devra utiliser l'un des symboles

« < », « > », « ≠ » (par ex : $H_1 : p < 0.50 ; \mu > 1 ; \sigma < 15$)

On teste H_0 et on arrive à une conclusion que H_0 est acceptée ou rejetée

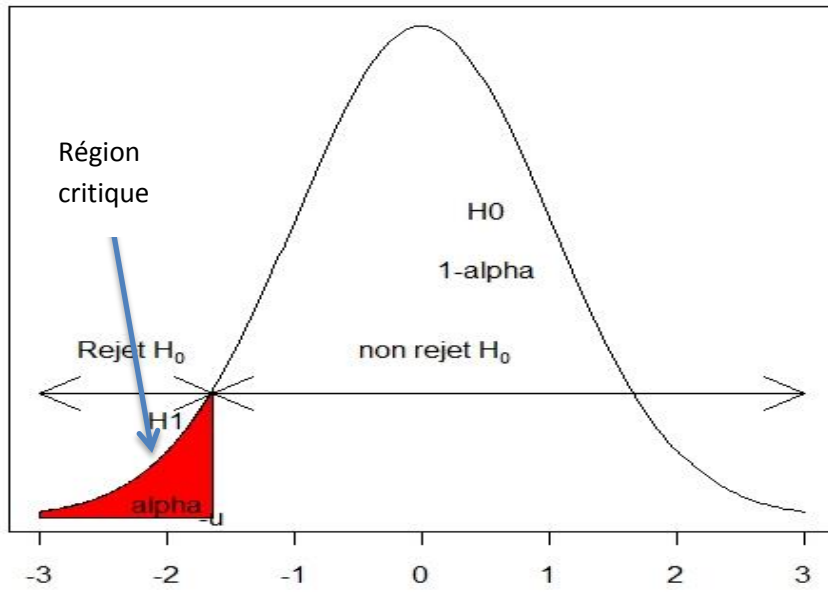
- **La statistique du test** : c'est une valeur calculée à partir des données de l'échantillon et elle est utilisée dans la prise de décision du rejet ou non de l'hypothèse nulle.
- **La région critique , niveau de significativité, valeur critique** :
 - ❖ La région critique (ou zone de rejet) est l'ensemble des valeurs de la statistique du test qui nous font rejeter H_0
 - ❖ Le niveau de significativité (ou le risque noté α) est la probabilité que la statistique du test soit dans la région critique
 - ❖ La valeur critique est une valeur qui sépare la région critique (où on rejette H_0) de la région des autres valeurs de la statistique du test.
 - ❖ Test bilatéral : ($H_1 : \neq$) la région critique est dans les deux régions extrêmes sous la courbe.

Test bilatéral : $H_1 : \neq$



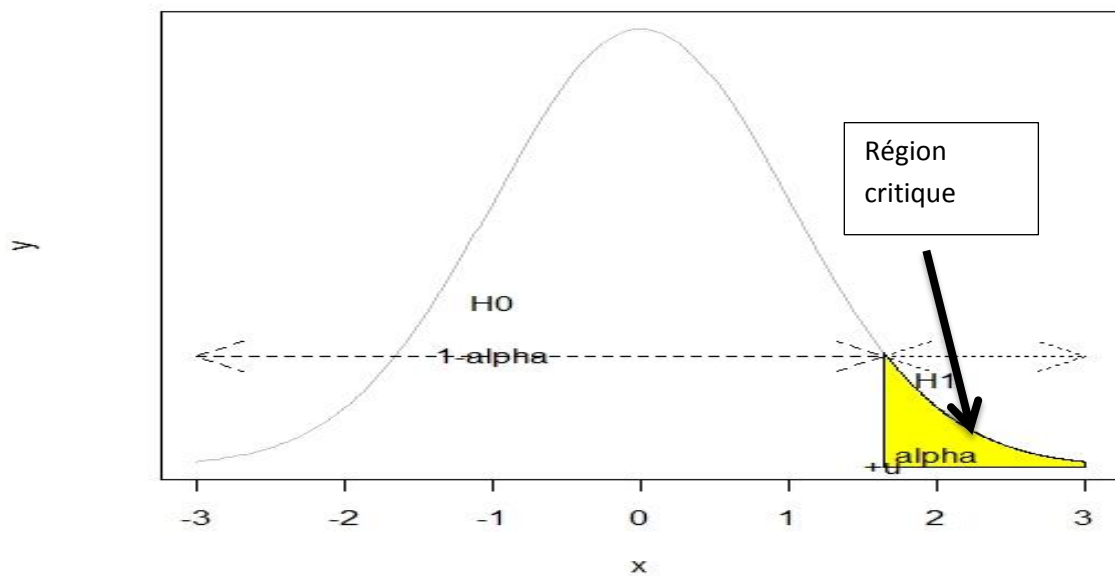
Test unilatéral à gauche : ($H_1 : \ll \gg$)

Test unilatéral à gauche



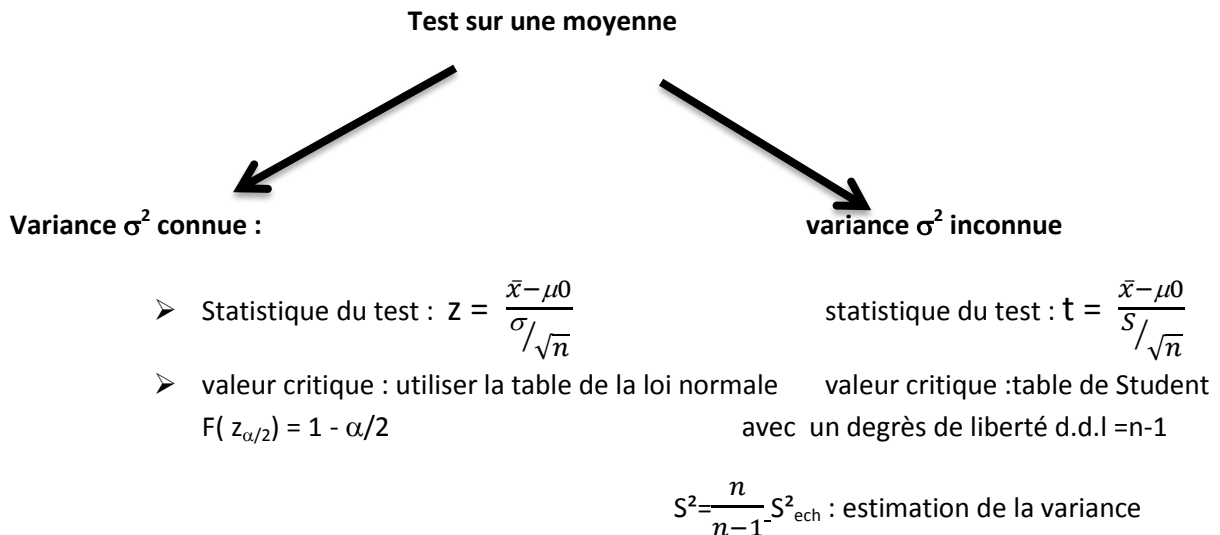
Test unilatéral à droite : ($H_1 : \gg \gg$)

Test unilatéral à droite



1.3/ Test d'hypothèse pour une moyenne :

- Conditions d'application du test :
 - ❖ L'échantillon est aléatoire et simple
 - ❖ Une des deux conditions est satisfaite : la population est distribuée normalement ou bien $n > 30$.
- $H_0 : \mu = \mu_0$



Exemples:

1/ Sur un échantillon de 106 températures humaines dont la moyenne est 36.78°C , on suppose que l'échantillon est aléatoire simple, que l'écart-type σ est connu et vaut 0.34. Avec un risque de 0.05, tester la croyance commune que la température d'un adulte en bonne santé est 37°C

Solution : les conditions du test sont satisfaites

$$H_0 : \mu = 37.0 \quad H_1 : \mu \neq 37.0 \quad (\text{test bilatéral})$$

$$\text{La statistique du test : } Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{36.78 - 37}{0.34 / \sqrt{106}} = -6.66$$

$$\text{Valeur critique : } F(z_{\alpha/2}) = 1 - \alpha/2 = 0.975 \longrightarrow z_{\alpha/2} = 1.96$$

Loi normale

Comme $z = -6.66 < -z_{\alpha/2} = -1.96$ on rejette alors H_0

2/ Le taux normal de glycémie est 1g/l. On dose la glycémie de 17 sujets à jeun, la moyenne est 1.2g/l avec un écart-type de 0.10g/l.

Peut-on dire ,au risque de 5% nque ces sujets sont hyperglycémiques en supposant que le taux de glycémie est distribué selon une loi normale ?

Solution : $H_0 : \mu = 1\text{g/l}$ $H_1 : \mu > 1\text{g/l}$

$$\text{statistique du test : } t = \frac{\bar{x} - \mu_0}{S / \sqrt{n}} = \frac{\bar{x} - \mu_0}{\text{Sech} / \sqrt{n-1}} = \frac{1.2 - 1}{0.10 / \sqrt{16}} = 8$$

valeur critique : la table de Student donne : $t_\alpha = t(n-1 ; \alpha) = t(16 ; 0.10) = 1.74$

comme $t > t_\alpha$ on rejette alors H_0 .

1.4/ test statistique sur une variance ou un écart-type :

Les ingénieurs de qualité veulent s'assurer qu'un produit a une moyenne acceptable mais ils veulent aussi produire des articles de qualité à peu près constante pour qu'il y ait peu de défauts . C'est pour cette raison , dans ce paragraphe, on va présenter une méthode qui teste des affirmations à propos de l'écart-type ou de la variance d'une population.

- Conditions d'application du test :
- ❖ L'échantillon est simple et aléatoire
- ❖ La population doit suivre une distribution normale
- $H_0 : \sigma^2 = \sigma_0^2$
- Statistique du test :

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

- Valeur critique : on utilise la table du khi-deux avec un d.d.l n-1

Exemple :

Une méthode , dite de référence , pour le titrage du phosphore donne pour une confiance de 95% une variance sur la concentration de phosphore à $0.05 \text{ g}^2/\text{l}^2$. Une nouvelle méthode est mise au point , on dose le phosphore dans 20 solutions , la variance des résultats obtenus est $0.06 \text{ g}^2/\text{l}^2$. On suppose que la concentration est distribuée selon une loi normale Peut-on dire qu'au risque de 5% , la précision de la nouvelle méthode est moins bonne que celle de la méthode de référence ?

Solution : $H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$

$$\text{Statistique du test : } X^2 = \frac{n}{\sigma_0^2} S_{\text{ech}}^2 = 20 * 0.06 / 0.05 = 24$$

Valeur critique : sur la table du khi-deux , on lit $X^2_\alpha = X^2(n-1 ; \alpha) = X^2(19 ; 0.05) = 30.14$

Comme $X^2 < X^2_\alpha$ alors on accepte H_0

Conclusion : au risque de 5%, on ne peut pas dire que la précision de la nouvelle méthode est moins bonne que celle de la méthode de référence.

1.5/ Test sur une proportion :

- Conditions requises pour ce test :
- ❖ Les observations d'échantillon proviennent d'un échantillon aléatoire simple
- ❖ Les conditions pour une loi binomiale sont satisfaites (un nombre fixé d'essais indépendants avec une même probabilité et chaque essai n'a que deux résultats : succès et échec
- ❖ $np_0 > 5$, $n(1-p_0) > 5$ sont satisfaites par conséquent la loi binomiale peut être approximée par la loi normale avec $\mu = np$ et $\sigma = \sqrt{npq}$
 - $H_0 : p = p_0$
 - Statistique du test :

$$Z = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad \text{avec } f : \text{ la proportion au niveau de l'échantillon}$$

- Valeur critique : on utilise la table de la loi normale

Exemple :

Mendel affirmait que la proportion de pois à gousse jaune devrait être égale à 25% .Sur un échantillon de 580 nouveaux pois on a obtenu 26.2% de pois à gousse jaune.

On veut tester l'hypothèse $H_0 : p = 0.25$ avec un risque de 0.05

Solution : vérifions d'abord les conditions du test :

1. l'échantillon est simple aléatoire
2. $np_0 = 145 > 5$ et $n(1-p_0) = 445 > 5$ les conditions sont satisfaites

$H_0 : p = 0.25$ $H_1 : p \neq 0.25$

Statistique du test :
$$Z = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.262 - 0.25}{\sqrt{\frac{0.25 * 0.75}{580}}} = 0.67$$

Valeurs critiques : la table de la loi normale nous donne : $F(Z_{\alpha/2}) = 1 - 0.05/2 = 0.975 \rightarrow Z_{\alpha/2} = 1.96$,

- $Z_{\alpha/2} = - 1.96$ comme $- 1.96 < Z < 1.96$ alors on accepte H_0

II. Test d'hypothèses avec deux échantillons :

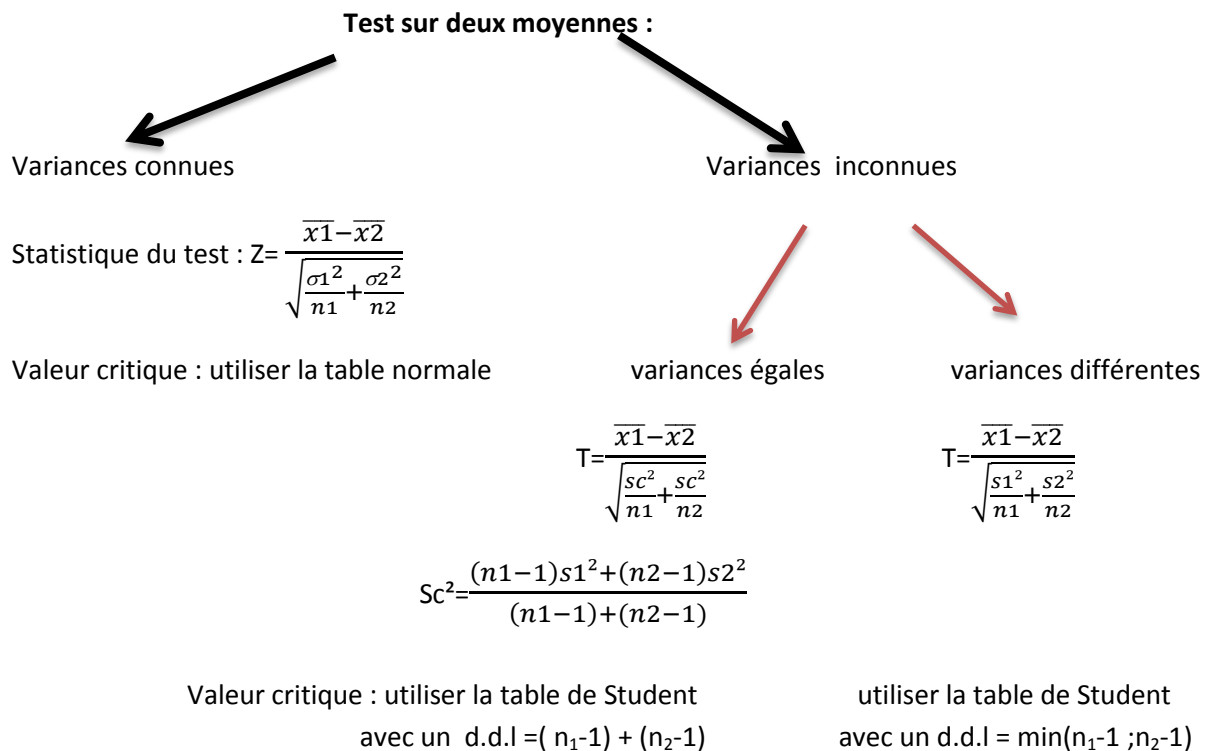
2.1 Test sur deux moyennes : échantillons indépendants :

Définition : _deux échantillons sont indépendants_ si les valeurs d'une population ne sont pas liées aux valeurs de l'autre population.

- Conditions d'application du test :
- ❖ Les deux échantillons sont simples et aléatoires
- ❖ Les deux échantillons sont indépendants
- ❖ Une ou les deux conditions sont satisfaites : les deux échantillons sont grands

$(n_1 > 30, n_2 > 30)$ ou les deux échantillons proviennent de populations possédant des distributions normales.

➤ $H_0 : \mu_1 = \mu_2$



Exemple1 :

On dose une même substance selon deux procédés A et B donnant une variabilité statistique déjà connue grâce à des mesures répétées , préalables, mesurées par $\sigma_A^2 = 105 ; \sigma_B^2 = 26$.

Le problème est de déceler une éventuelle différence entre les résultats des 2 méthodes.

Pour cela , on compare 50 dosages faits avec A et 30 avec B donnant les deux moyennes

$\bar{x}_A = 24.8 ; \bar{x}_B = 27.5$.

Solution : on vérifie les conditions du test :

1. Les 2 échantillons sont indépendants et aléatoires
2. Les échantillons sont de taille supérieure à 30 (grands échantillons)
3. Les variances sont connues

$H_0 : \mu_A = \mu_B ; H_1 : \mu_A \neq \mu_B$

Statistique du test : $Z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{24.8 - 27.5}{\sqrt{\frac{105}{50} + \frac{26}{30}}} = -1.5676$

Valeur critique : utilisons la table normale $N(0,1)$: $F(Z_{\alpha/2}) = 1 - \alpha/2 = 1 - 0.025 = 0.975$ on lit sur la table :

$Z_{\alpha/2} = 1.96$. Comme $Z < Z_{\alpha/2}$ donc on accepte H_0

Conclusion : il n'y a pas suffisamment de preuves pour garantir une différence significative entre les résultats des 2 méthodes A et B.

Exemple 2 :

Pour tester l'efficacité d'un traitement d'une maladie, on a obtenu les résultats suivants :

Groupe placebo : $n_1 = 43$; $\bar{x}_1 = 21.57$; $s_1 = 3.87$

Groupe traité : $n_2 = 33$; $\bar{x}_2 = 20.38$; $s_2 = 3.91$

Utiliser un niveau de significativité de 0.05 pour tester l'hypothèse que les 2 groupes proviennent d'une population avec la même moyenne.

Solution : vérifions les conditions d'application du test :

1. les deux échantillons sont simples et aléatoires
2. les deux échantillons sont de taille supérieure à 30

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Statistique du test :

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{21.57 - 20.38}{\sqrt{\frac{3.87^2}{43} + \frac{3.91^2}{33}}} = 1.321$$

Valeur critique : utilisons la table de Student avec un d.d.l = $\min(n_1 - 1 ; n_2 - 1) = \min(33 - 1 ; 43 - 1) = 32$

$$T(\alpha ; \text{d.d.l}) = T(0.05 ; 32) = 2.037$$

Comme $T < T(0.05 ; 32)$ donc on accepte H_0

Conclusion : il y a suffisamment de preuves pour dire que les moyennes ne sont pas significativement différentes le traitement ne semble pas avoir d'effet significatif

2.2 Test sur deux moyennes : échantillons appariés :

Définition : deux échantillons sont appariés lorsqu'ils portent :

- * Sur les mêmes individus (exemple : un groupe de 20 personnes pour lesquelles on dose le cholestérol à deux temps différents, on a alors deux séries de données provenant d'un même groupe d'individus)
- * Sur des individus ayant au moins un caractère semblable (exemple : deux groupes de personnes appariés en âge pour lesquelles on dose le cholestérol à un temps donné, on a deux séries de données provenant de deux groupes d'individus différents mais de même âge).

➤ Conditions d'application du test :

- ❖ Les échantillons simples aléatoires appariés

- ❖ Une ou les deux conditions sont satisfaites : le nombre de paires est grand ou bien les paires de valeurs proviennent de populations dont la distribution est approximativement normale.

Notations :

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

\bar{d} : valeur moyenne des différences

S_d : écart-type estimé des différences d pour la population à partir de l'échantillon

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

n : le nombre de paires

➤ $H_0 : \mu_d = 0$

➤ Statistique du test : $T = \frac{\bar{d}}{sd/\sqrt{n}}$

➤ valeur critique : utiliser la table de Student avec un d.d.l = $n-1$.

Exemple :

Pour étudier l'action d'un produit sur un paramètre biologique, on a mesuré sur un échantillon de 10 individus la valeur du paramètre avant et après traitement. Les résultats sont :

individu	1	2	3	4	5	6	7	8	9	10
avant	5.33	6.13	5.66	4.50	5.35	6.32	4.24	5.83	6.27	4.86
après	5.32	6.00	5.64	4.59	5.49	6.17	4.11	5.86	6.13	4.68

Le traitement modifie-t-il de façon significative le paramètre biologique ? on suppose que le paramètre biologique suit une loi normale

Solution : vérifions que les conditions du test sont satisfaites :

Les deux séries de valeurs sont aléatoires et appariées liées par paires car ce sont les mesures prises sur les mêmes individus (avant et après traitement) et on suppose que la population est distribuée normalement.

$H_0 : \mu_d = 0$ $H_1 : \mu_d > 0$

On introduit une nouvelle variable des différences notées d

Les valeurs de d sont : 0.01 0.13 0.02 -0.09 -0.14 0.15 0.13 -0.03 0.14 0.18

On calcule : $\bar{d} = 0.05$ $s^2_d = 1.249 * 10^{-2}$

Statistique du test :

$$T = \frac{\bar{d}}{sd/\sqrt{n}} = \frac{0.05}{\sqrt{1.249 \cdot 10^{-2}/10}} = 1.415$$

Valeur critique : utilisons la table de Student avec un d.d.l = 10-1=9 et la colonne 2*0.05=0.10

On a $T_{\alpha} = 2.262$

Comme $T < T_{\alpha}$ on accepte alors H_0

Conclusion : la différence n'est pas significative, le traitement ne modifie pas de façon significative le traitement biologique.

2.3 Test sur deux variances :

- Conditions d'application du test :
 - ❖ Les deux échantillons sont indépendants
 - ❖ Les deux populations ont une distribution normale (une condition importante)
- Notations :
 - S_1^2 : la plus grande des variances des deux échantillons
 - n_1 : la taille de l'échantillon de la plus grande variance
 - σ_1^2 : la variance de la population à partir de laquelle est extrait l'échantillon ayant la plus grande variance.
 - Les symboles S_2^2, n_2, σ_2^2 sont utilisés pour l'autre échantillon et population.
- $H_0 : \sigma_1^2 = \sigma_2^2$
- Statistique du test :
$$F = S_1^2 / S_2^2 \quad (> 1)$$
- Valeur critique : utiliser la table de Fischer avec d.d.l ($n_1 - 1, n_2 - 1$)

Exemple :

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	$n_1 = 13$	$S_1 = 9.46$
Calcium	$n_2 = 15$	$S_2 = 8.469$

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

Solution :

Vérifions que les conditions du test sont satisfaites :

Les deux échantillons sont indépendants.

Les échantillons proviennent de populations normales.

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Statistique du test :

$$F = S_1^2 / S_2^2 = 9.468^2 / 8.469^2 = 1.248$$

Valeurs critiques : on a un test bilatéral , la table de Fischer nous donne :

$$\begin{aligned} F(\alpha/2 ; n_1 - 1, n_2 - 1) &= F(0.05/2 ; 13-1 ; 15-1) \\ &= F(0.025 ; 12 ; 13) = 3.0502 \end{aligned}$$

Comme $F < F_{\alpha/2} = 3.0502$ alors on accepte H_0 .

Conclusion : on peut dire que les échantillons sont issus de populations de même écart-type.

2.4 Test sur deux proportions

- Conditions requises du test :
 - ❖ Les proportions sont issues de deux échantillons simples et aléatoires.
 - ❖ $n_1 p_1 > 5$, $n_1(1 - p_1) > 5$ et $n_2 p_2 > 5$, $n_2(1 - p_2) > 5$

- Notations :
 - n_1 : taille de l'échantillon 1
 - f_1 : proportion de l'échantillon 1
 - p_1 : proportion de la population 1
 - les symboles n_2 , f_2 , p_2 sont utilisés pour l'autre échantillon et population.

- $H_0 : p_1 = p_2$

- Statistique du test :

$$Z = \frac{f_1 - f_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\text{Avec : } P = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

- Valeur critique : utiliser la table de la loi normale.

Exemple :

Pour tester un certain médicament , un essai clinique est réalisé en mode ambulatoire . Deux groupes de 40 sujets chacun , sont constitués par tirage au sort. Au cours de cet essai, une épidémie s'abat sur la ville et atteint 15 sujets du groupe 1 et 9 sujets du groupe 2 . Peut-on dire ,au risque de 5%, que la proportion du groupe 1 est supérieure à celle du groupe 2 ?

Solution :

Echantillon 1 :

$$n_1 = 40$$

$$f_1 = 15/40$$

Echantillon 2 :

$$n_2 = 40$$

$$f_2 = 9/40$$

Vérifions que les conditions du test sont satisfaites :

- Les 2 échantillons sont simples et aléatoires .
- $15 > 5$, $25 > 5$ et $9 > 5$, $31 > 5$

$$H_0 : p_1 = p_2 \quad H_1 : p_1 > p_2$$

On calcule

$$P = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{15 + 9}{40 + 40} = 0.30$$

Statistique du test :

$$Z = \frac{f_1 - f_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 1.46$$

Valeur critique : test unilatéral , on lit sur la table normale avec $\alpha = 0.05$, $Z(0.05) = 1.64$

Comme $Z < Z\alpha = 1.64$ donc on accepte H_0 .

Conclusion : on ne peut pas dire que la proportion du groupe 1 est supérieure à la proportion du groupe 2 .

