LES TESTS D'HYPOTHÈSE

Dr. Meriem Bouhadjar

2020/2021

1. LES TESTS D'HYPOTHÈSE

1. Généralités

Les tests d'hypothèse constituent un autre aspect important de l'inférence statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit :

- On étudie une population dont les éléments possèdent un caractère (mesurable ou qualitatif) et dont la valeur du paramètre relative au caractère étudié est inconnue.
- Une hypothèse est formulée sur la valeur du paramètre : cette formulation résulte de considérations théoriques, pratiques ou encore elle est simplement basée sur un pressentiment.
- On veut porter un jugement sur la base des résultats d'un échantillon prélevé de cette population.

1. Généralités

Il est bien évident que la statistique (c'est-à-dire la variable d'échantillonnage) servant d'estimateur au paramètre de la population ne prendra pas une valeur rigoureusement égale à la valeur théorique proposée dans l'hypothèse. Cette variable aléatoire comporte des fluctuations d'échantillonnage qui sont régies par des distributions connues.

Pour décider si l'hypothèse formulée est supportée ou non par les observations, il faut une méthode qui permettra de conclure si l'écart observé entre la valeur de la statistique obtenue dans l'échantillon et celle du paramètre spécifiée dans l'hypothèse est trop important pour être uniquement imputable au hasard de l'échantillonnage.

1.Généralités

La construction d'un test d'hypothèse consiste en fait à déterminer entre quelles valeurs peut varier la variable aléatoire, en supposant l'hypothèse vraie, sur la seule considération du hasard de l'échantillonnage.
Les distributions d'échantillonnage d'une moyenne, d'une variance et d'une proportion que nous avons traitées dans un chapitre précédent vont être particulièrement utiles dans l'élaboration des tests statistiques.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Hypothèse statistique
Une **hypothèse statistique** est un énoncé (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 de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Hypothèse nulle ($\mathbf{H_0}$) et hypothèse alternative ($\mathbf{H_1}$)

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 l'hypothèse 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.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Dans notre démarche, nous allons établir des règles de décision qui vont nous conduire à l'acceptation ou au rejet de l'hypothèse nulle $H_{\mathbf{0}}$. Toutefois cette décision est fondée sur une information partielle, les résultats d'un échantillon. Il est donc statistiquement impossible de prendre la bonne décision à coup sûr. En pratique, on met en oeuvre une démarche qui nous permettrait, à long terme de rejeter à tort une hypothèse nulle vraie dans une faible proportion de cas. La conclusion qui sera déduite des résultats de l'échantillon aura un caractère probabiliste : on ne pourra prendre une décision qu'en ayant conscience qu'il y a un certain risque qu'elle soit erronée. Ce risque nous est donné par le seuil de signification du test.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Seuil de signification du test

Le risque, consenti à l'avance et que nous notons α de rejeter à tort l'hypothèse nulle H_0 alors qu'elle est vraie, s'appelle le **seuil de signification** du test et s'énonce en probabilité ainsi :

$$\alpha = P(rejeter\ H_0/H_0\ vraie)$$

A ce seuil de signification, on fait correspondre sur la distribution d'échantillonnage de la statistique une région de rejet de l'hypothèse nulle (appelée également région critique). L'aire de cette région correspond à la probabilité α. Si par exemple , on choisit $\alpha = 0.05$, cela signifie que l'on admet d'avance que la variable d'échantillonnage peut prendre, dans 5% des cas, une valeur se situant dans la zone de rejet de H_0, bien que H_0 soit vraie et ceci uniquement d'après le hasard de l'échantillonnage.

Sur la distribution d'échantillonnage correspondra aussi une région complémentaire, dite région d'acceptation de H_0 (ou région de non-rejet) de probabilité $1 - \alpha$.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Remarques :

1. Les seuils de signification les plus utilisés sont $\alpha = 0,05$ et $\alpha = 0,01$, dépendant des conséquences de rejeter à tort l'hypothèse $H_{\mathbf{0}}$.

2. La statistique qui convient pour le test est donc une variable aléatoire dont la valeur observée sera utilisée pour décider du « rejet » ou du « non-rejet » de $H_{\mathbf{0}}$. La distribution d'échantillonnage de cette statistique sera déterminée en supposant que l'hypothèse est $H_{\mathbf{0}}$ vraie.

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Exemple de formulation d'un test :
Supposons que nous affirmions que la valeur d'un paramètre θ d'une population est égale à la valeur θ_0. On s'intéresse au changement possible du paramètre θ dans l'une ou l'autre direction (soit $\theta > \theta_0$ soit $\theta < \theta_0$). On effectue un test bilatéral.
Les hypothèses H_0 et H_1 sont alors :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

Définition des concepts utiles à l'élaboration des tests d'hypothèse

Si, suite aux résultats de l'échantillon, la valeur de la statistique utilisée se situe dans l'intervalle $[\theta_{c_1}, \theta_{c_2}]$, on acceptera H_0 au seuil de signification choisi. Si, au contraire, la valeur obtenue est supérieure à θ_{c_2} ou inférieure à θ_{c_1}, on rejette H_0 et on accepte H_1.
Remarque : Si on s'intéresse au changement du paramètre dans une seule direction, on opte pour **un test unilatéral**, en choisissant comme hypothèse H_1.
soit $\theta > \theta_0$ soit $\theta < \theta_0$. La région critique est alors localisée uniquement à droite ou uniquement à gauche de la région d'acceptation.
Dans un souci de simplification, nous nous intéresserons dans ce cours essentiellement aux tests bilatéraux.

2. Tests peremettant de détertminer si un échantillon appartient a une population donnée

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

Nous voulons déterminer si l'échantillon de taille n dont nous disposons appartient à une population de moyenne μ_0 au seuil de signification α. Nous allons dans tous les tests travailler de la même façon, en procédant en quatre étapes.

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

$1^{\text{ère}}$ étape : formulation des hypothèses

L'échantillon dont nous disposons provient d'une population de moyenne μ. Nous voulons savoir si

$\mu = \mu_0$.

On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

$2^{\text{ème}}$ étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

• On détermine la statistique qui convient pour ce test.. Ici, l'estimateur de la moyenne μ, c'est-à-dire $\overline{X}$, semble tout indiquée.

• On détermine la loi de probabilité de $\overline{X}$ en se plaçant sous l'hypothèse H_0.

Deux cas peuvent se produire :

Premier cas : L'échantillon est de grande taille $(n \geq 30)$ ou bien la population est normale de variance σ_X^2 connue.

$\overline{X}$ suit alors une loi normale de moyenne μ_0 (puisqu'on se place sous H_0) et d'écart-type $\frac{\sigma_X}{\sqrt{n}}$: $\overline{X} \backsim \mathcal{N}\left(\mu_0, \frac{\sigma_X}{\sqrt{n}}\right)$. On pose $Z = \frac{\overline{X}-\mu_0}{\sigma_X/\sqrt{n}}$

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

Z mesure un écart réduit. Z est aussi appelée **fonction discriminante du test**.

$$Z \backsim \mathcal{N}(0,1)$$

Deuxième cas : L'échantillon est de petite taille $(n < 30)$ prélevé au hasard d'une population normale de variance σ_X^2 inconnue.
Dans ce cas **la fonction discriminante du test** sera :

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Ici $T \backsim T_{n-1}$ (loi de Student à $(n-1)$ degrés de liberté).

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

3ème étape : Détermination des valeurs critiques de Z délimitant les zones d'acceptation et de rejet

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.

Il nous faut donc déterminer dans la table la valeur maximale $z_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire

vérifiant : $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$

Tests sur une monnyenne : comparaison d'une moyenne expérimentale a une moyenne théorique dans le cas d'un caractère quantitatif

4ème étape : Calcul de la valeur de Z prise dans l'échantillon et conclusion du test

• On calcule la valeur z_0 prise par Z dans l'échantillon.

• $\rightarrow$ Si la valeur z_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est statistiquement significatif au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.

$\rightarrow$ Si la valeur z_0 se trouve dans la zone d'acceptation, on dira que l'écartréduit observé n'est pas significatif au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

Tests sur une proportion

Nous nous proposons de tester si la proportion p d'éléments dans la population présentant un certain caractère qualitatif peut être ou non considérée comme égale à une valeur hypothétique p_0. Nous disposons pour ce faire de la proportion d'éléments possédant ce caractère dans un échantillon de taille n. Nous allons procéder comme au paragraphe précédent, en quatre étapes.

Tests sur une proportion

1ère étape : formulation des hypothèses

L'échantillon dont nous disposons provient d'une population dont la proportion d'éléments présentant le caractère qualitatif est p. Nous voulons savoir si $p = p_0$.

On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

Tests sur une proportion

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

• On détermine la statistique qui convient pour ce test. Ici, l'estimateur de la proportion p, c'est-à-dire F, semble tout indiquée.

• On détermine la loi de probabilité de F **en se plaçant sous l'hypothèse** H_0. On suppose que l'on dispose d'un grand échantillon $(n \geq 30)$ et que « p n'est pas trop petit » (de manière que l'on ait $np \geq 5$ et $n(1-p) \geq 5$) .

Tests sur une proportion

F suit alors une loi normale de moyenne p_0 (puisqu'on se place sous H_0) et d'écart-type
$\sqrt{\frac{p_0(1-p_0)}{n}}$: $F \backsim \mathcal{N}\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$

On pose $Z = \frac{F-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ Z mesure un écart réduit.

Zest aussi appelée **fonction discriminante du test**. $Z \backsim \mathcal{N}(0,1)$

Tests sur une proportion

3ème étape : Détermination des valeurs critiques de Z délimitant les zones d'acceptation et de rejet

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.

Il nous faut donc déterminer dans la table la valeur maximale $z_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage

au seuil de signification α, c'est-à-dire

vérifiant : $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$

Tests sur une proportion

4[ème] étape : Calcul de la valeur de Z prise dans l'échantillon et conclusion du test

• On calcule la valeur z_0 prise par Z dans l'échantillon.

• $\rightarrow$ Si la valeur z_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est **statistiquement significatif** au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0 .

$\rightarrow$ Si la valeur z_0 se trouve dans la zone d'acceptation, on dira que l'écartréduit observé **n'est pas significatif** au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0 .

3. Risque de première et de deuxième espèce

Définition :
Tous les règles de décision que nous avons déterminées acceptaient un risque α qui était le risque de rejeter à tort l'hypothèse H_0, c'est-à-dire le risque de rejeter l'hypothèse H_0, alors que H_0 est vraie. Ce risque s'appelle aussi **le risque de première espèce.**
La règle de décision du test comporte également un deuxième risque, à savoir de celui de ne pas rejeter l'hypothèse nulle H_0 alors que c'est l'hypothèse H_1 qui est vraie. **C'est le risque de deuxième espèce.**

3. Risque de première et de deuxième espèce

Les deux risques peuvent se définir ainsi :
$\alpha = P\,(\text{rejeter } H_0/H_0 \text{ vrais}) =$ probabilité de comm ettre une erreur de première espèce
$\beta = P\,(\text{ne pas rejeter } H_0/H_0 \text{ vrais}) =$ probabilité de comm ettre une erreur de deuxième espèce
Le risque de première espèce α est choisi à priori. Toutefois le risque de deuxième espèce β dépend de l'hypothèse alternative H_1 et on ne peut le calculer que si on spécifie des valeurs particulières du paramètre dans l'hypothèse H_1 que l'on suppose vraie.

3. Risque de première et de deuxième espèce

Les risques liés aux tests d'hypothèses peuvent se résumer ainsi :
Tableau dans la page Word

3. Risque de première et de deuxième espèce

Remarque : La probabilité complémentaire du risque de deuxième espèce $(1-\beta)$ définit la puissance du test à l'égard de la valeur du paramètre dans l'hypothèse alternative H_1. La puissance du test représente la probabilité de rejeter l'hypothèse nulle H_0 lorsque l'hypothèse vraie est H_1. Plus β est petit, plus le test est puissant.

3. Risque de première et de deuxième espèce

1. Pour un même risque α et une même taille d'échantillon, on constate que, si l'écart entre la valeur du paramètre posée en H_0 et celle supposée dans l'hypothèse vraie H_1 augmente, le risque β diminue.
2. Une réduction du risque de première espèce (de $\alpha = 0.05$ à $\alpha = 0.01$ par exemple) élargit la zone d'acceptation de H_0. Toutefois, le test est accompagné d'une augmentation du risque de deuxième espèce β. On ne peut donc diminuer l'un des risques qu'en consentant à augmenter l'autre.
3. Pour une valeur fixe de α et un σ déterminé, l'augmentation de la taille d'échantillon aura pour effet de donner une meilleure précision puisque
$\sigma\left(\overline{X}\right) = \sigma/\sqrt{n}$
diminue. La zone d'acceptation de H_0 sera alors plus restreinte, conduisant à une diminution du risque β. Le test est alors plus puissant.

4.Tests permettant de déterminer si deux échantillons appartiennent a la même population

Introduction :
Il existe de nombreuses applications qui consistent, par exemple, à comparer deux groupes d'individus en regard d'un caractère quantitatif particulier (poids, taille, rendement scolaire, quotient intellectuel,....) ou à comparer deux procédés de fabrication selon une caractéristique quantitative particulière (résistance à la rupture, poids, diamètre, longueur,...) ou encore de comparer les proportions d'apparition d'un caractère qualitatif de deux populations (proportion de défectueux, proportion de gens favorisant un parti politique,...). Les variables aléatoires qui sont alors utilisées pour effectuer des tests d'hypothèses (ou aussi calculer des intervalles de confiance) sont la *différence des moyennes* d'échantillon, le *quotient des variances* d'échantillon ou la *différence des proportions* d'échantillon.

On étudie un caractère quantitatif

Comparaison de deux moyennes d'échantillon : « test Z »

Nous nous proposons de tester si la moyenne de la première population (μ_1) peut être ou non considérée comme égale à la moyenne de la deuxième population (μ_2).

Nous allons alors comparer les deux moyennes d'échantillon $\overline{x_1}$ et $\overline{x_2}$. Il est évident que si $\overline{x_1}$ et $\overline{x_2}$ diffèrent beaucoup, les deux échantillons n'appartiennent pas la même population. Mais si $\overline{x_1}$ et $\overline{x_2}$ diffèrent peu, il se pose la question de savoir si l'écart $d = \overline{x_1} - \overline{x_2}$ peut être attribué aux hasards de l'échantillonnage. Afin de donner une réponse rigoureuse à cette question, nous procéderons encore en quatre étapes.

Comparaison de deux moyennes d'échantillon : test Z

$1^{\text{ère}}$ étape : formulation des hypothèses

Le premier échantillon dont nous disposons provient d'une population dont la moyenne est μ_1. Le deuxième échantillon dont nous disposons provient d'une population dont la moyenne est μ_2.

Nous voulons savoir si il s'agit de la même population en ce qui concerne les moyennes, c'est-à-dire si $\mu_1 = \mu_2$.

On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Comparaison de deux moyennes d'échantillon : test Z

2$^{\text{ème}}$ étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

• On détermine la statistique qui convient pour ce test. Ici, la différence $D = \overline{X_1} - \overline{X_2}$ des deux moyennes d'échantillon, semble tout indiquée.

• On détermine la loi de probabilité de D **en se plaçant sous l'hypothèse** H_0. On suppose que l'on dispose de grands échantillons $(n_1 \geq 30$ et $n_2 \geq 30)$ et que les deux variances d'échantillon $\sigma_{\overline{X}_1}$ et $\sigma_{\overline{X}_2}$ sont connues.

Comparaison de deux moyennes d'échantillon : test Z

$\Rightarrow \overline{X_1}$ suit alors une loi normale de moyenne μ_1 et d'écart-type $\sigma_{X_1}/\sqrt{n_1}$ que l'on
peut sans problème estimer par $\sigma_{\overline{X}_1}/\sqrt{n_1 - 1}$ car $(n_1 \geq 30)$

$$\overline{X_1} \backsim \mathcal{N}(\mu_1, \frac{\sigma_{\overline{X}_1}}{\sqrt{n_1 - 1}})$$

$\Rightarrow \overline{X_2}$ suit alors une loi normale de moyenne μ_2 et d'écart-type $\sigma_{X_2}/\sqrt{n_2}$ que l'on
peut sans problème estimer par $\sigma_{\overline{X}_2}/\sqrt{n_2 - 1}$ car $(n_2 \geq 30)$

$$\overline{X_2} \backsim \mathcal{N}(\mu_2, \frac{\sigma_{\overline{X}_2}}{\sqrt{n_2 - 1}})$$

Comparaison de deux moyennes d'échantillon : test Z

⇒On en déduit, puisque $\overline{X_1}$ et $\overline{X_2}$ sont indépendantes que $D = \overline{X_1} - \overline{X_2}$ également une loi normale.

$E(D) = E(\overline{X_1}) - E(\overline{X_2}) = \mu_1 - \mu_2 = 0$ puisqu'on se place sous H_0.

$V(D) = V(\overline{X_1}) + V\ (\overline{X_2}) = \frac{\sigma^2_{\overline{X}_1}}{n_1 - 1} + \frac{\sigma^2_{\overline{X}_2}}{n_2 - 1}$ puisque les variables sont indépendantes.

On pose $Z = \frac{D}{\sqrt{\frac{\sigma^2_{\overline{X}_1}}{n_1 - 1} + \frac{\sigma^2_{\overline{X}_2}}{n_2 - 1}}}$ Z mesure un écart réduit.

Z est **la fonction discriminante du test**. $Z \backsim \mathcal{N}(0, 1)$.

Comparaison de deux moyennes d'échantillon : test Z

3ème étape : Détermination des valeurs critiques de Z délimitant les zones d'acceptation et de rejet

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.

Il nous faut donc déterminer dans la table la valeur maximale $z_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage
au seuil de signification α, c'est-à-dire
vérifiant : $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$

Comparaison de deux moyennes d'échantillon : test Z

$4^{\text{ème}}$ étape : Calcul de la valeur de Z prise dans l'échantillon et conclusion du test

• On calcule la valeur z_0 prise par Z dans l'échantillon.

• → Si la valeur z_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est **statistiquement significatif** au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. <u>On rejette H_0</u> .

→ Si la valeur z_0 se trouve dans la zone d'acceptation, on dira que l'écartréduit observé **n'est pas significatif** au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. <u>On accepte H_0</u> .

On étudie caractère qualitatif : comparaison de deux proportions échantillon

Il y a de nombreuses applications (échéances électorales, expérimentations médicales...) où nous devons décider si l'écart observé entre deux proportions échantillonnales est significatif où s'il est attribuable au hasard de l'échantillonnage. Pour répondre à cette question, nous procéderons comme d'habitude en quatre étapes.

1$^{\text{ère}}$ étape : formulation des hypothèses

Le premier échantillon dont nous disposons provient d'une population 1 dont les éléments possèdent un caractère qualitatif dans une proportion inconnue p_1. Le deuxième échantillon dont nous disposons provient d'une population 2 dont les

éléments possèdent le même caractère qualitatif dans une proportion inconnue p_2.

On étudie caractère qualitatif : comparaison de deux proportions échantillon

Nous voulons savoir si il s'agit de la même population en ce qui concerne les proportions, c'est-à-dire si $p_1 = p_2$.
On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :

$$\left[\begin{array}{l} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{array}\right.$$

On étudie caractère qualitatif : comparaison de deux proportions échantillon

2$^{\text{ème}}$ étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

Nous traiterons uniquement le cas où nous sommes en présence de grands échantillons.

• On détermine la statistique qui convient pour ce test. Ici, la différence $D = F_1 - F_2$ des deux proportions d'échantillon, semble tout indiquée, puisque F_1 est un estimateur sans biais de p_1 et F_2 un estimateur sans biais de p_2..

• On détermine la loi de probabilité de D **en se plaçant sous l'hypothèse** H_o.

$\Rightarrow$ F_1 suit alors une loi normale de moyenne p_1 et d'écart-type $\sqrt{\frac{p_1(1-p_1)}{n_1}}$

$\Rightarrow$ De même, F_2 suit alors une loi normale de moyenne p_2 et d'écart-type $\sqrt{\frac{p_2(1-p_2)}{n_2}}$

On étudie caractère qualitatif : comparaison de deux proportions échantillon

$\Rightarrow$ On en déduit, puisque F_1 et F_2 sont indépendantes que $D = F_1 - F_2$ suit également une loi normale.
$E(D) = E(F_1) - E(F_2) = p_1 - p_2 = 0$ puisqu'on se place sous H_0.
$V(D) = V(F_1) + V(F_2) = \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}$ puisque les variables sont indépendantes. Ici, on a posé $p_1 = p_2 = p$ puisque l'on se place sous H_0.

On étudie caractère qualitatif : comparaison de deux proportions échantillon

Mais comment trouver p puisque c'est justement sur p que porte le test ?
Puisque nous raisonnons en supposant l'hypothèse H_0 vraie, on peut considérer que les valeurs de F_1 et F_2 obtenues sur nos échantillons sont des approximations de p. De plus , plus la taille de l'échantillon est grande, meilleure est l'approximation (revoir le chapitre sur les intervalles de confiance).
Nous allons donc pondérer les valeurs observées dans nos échantillons par la taille respective de ces échantillons.
On approchera p dans notre calcul par : $\hat{p} \approx \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

On étudie caractère qualitatif : comparaison de deux proportions échantillon

On pose $Z = \frac{D}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$ Z mesure un écart réduit.

Z est **la fonction discriminante** du test. $Z \backsim \mathcal{N}(0,1)$.

3$^{\text{ème}}$ étape : Détermination des valeurs critiques de Z délimitant les zones d'acceptation et de rejet

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.

Il nous faut donc déterminer dans la table la valeur maximale $z_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage
au seuil de signification α, c'est-à-dire
vérifiant : $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$

On étudie caractère qualitatif : comparaison de deux proportions échantillon

$4^{\text{ème}}$ étape : Calcul de la valeur de Z prise dans l'échantillon et conclusion du test

• On calcule la valeur z_0 prise par Z dans l'échantillon.

• $\rightarrow$ Si la valeur z_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est **statistiquement significatif** au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. <u>On rejette H_0</u> .

$\rightarrow$ Si la valeur z_0 se trouve dans la zone d'acceptation, on dira que l'écartréduit observé **n'est pas significatif** au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. <u>On accepte H_0</u> .