Statistique inférentielle

Tests

A. Godichon-Baggioni

I. Généralités

Principe d'un test

Le principe d'un test consiste à

1. Poser une hypothèse nulle H0 contre une hypothèse alternative H1
2. A l'aide d'un modèle probabiliste, définir une zone de rejet de l'hypothèse nulle H0
3. Regarder si le résultat expérimental est dans cette zone de rejet
4. Si le résultat est dans la zone de rejet, on rejette l'hypothèse H0.
5. Si le résultat n'est pas dans la zone de rejet, on ne rejette pas l'hypothèse H0.

Attention ! La sémantique est importante.

Exemple

Le lancer de pièce : On souhaite savoir si la pièce est équilibrée. On jette 10 fois une pièce et on obtient 9 ″Face″. Un test permet de dire avec un certain risque si cette proportion de ″Face″ est seulement due à la fluctuation d'échantillonnage ou bien si la pièce est truquée.

Etapes d'un test statistique

On dispose de n données $x_1, \ldots, x_n$ qui sont des réalisations de variables qualitatives ou quantitatives et on fixe un risque $\alpha \in (0, 1)$.

- Introduire un modèle probabiliste

  On peut supposer que les x_i sont des réalisations de variables aléatoires indépendantes $X_1, \ldots, X_n$ (ce qui est le cas pour le lancer de pièce)

- Les hypothèses nulles et alternatives : disposant d'un modèle probabiliste, on souhaite vérifier une hypothèse sur ce modèle, appelée hypothèse nulle H0 et on introduit sa négation H1.

  Dans le cas du lancer de pièce, on pose les hypothèses

$$H0 : \text{"}\theta = 0.5\text{"} \quad \text{contre} \quad H1 : \text{"}\theta \neq 0.5\text{"}$$

Etapes d'un test statistique

- La statistique de test et sa loi sous H0 :

  Une fois le modèle probabiliste introduit, il faut disposer d'une variable aléatoire appelée statistique de test, notée Z

  On connaît sa loi sous H0

  Sa loi n'est pas la même sous H1

Z est définie en fonction des variables aléatoires $X_1, \ldots, X_n$, i.e $Z = T(X_1, \ldots, X_n)$ et on note sa réalisation $z_{obs} = T(x_1, \ldots, x_n)$ qui doit être calculable.

Etapes d'un test statistique

- Zone de rejet :

  Connaissant la loi de Z sous H0, on peut déterminer ses valeurs les plus extrêmes, i.e déterminer une zone ZR telle que $\mathbb{P}[Z \in ZR] \leq \alpha$ si H0 est vraie.

- On conclut le test à partir de la valeurs observée z_{obs}

  Si z_{obs} appartient à la zone de rejet, on rejette H0 au risque α.
  Si z_{obs} n'appartient pas à la zone de rejet, on ne rejette pas H0 au risque α.

Exemple

Le lancer de pièce : On teste au risque de 5%

$$H0 : "\theta = 0.5" \quad \text{contre} \quad H1 : "\theta \neq 0.5"$$

On a la statistique de test

$$Z = \sum_{i=1}^{10} X_i \sim \mathcal{B}(10, 0.5) \quad \text{sous H0}$$

On a la zone de rejet

$$ZR_{H0} = \{0, 1\} \cup \{9, 10\}.$$

Ici, $z_{obs} = 9$ et on rejette donc H0.

P-VALUE

Plus le seuil α est petit plus le test est fiable.

$$p - \text{value} = \inf \{\alpha \in (0,1), \text{"On rejette H0"}\}$$

La p-value est la probabilité d'avoir eu un aussi mauvais résultat que z_{obs} sous H0.

- Si $p - \text{value} \geq \alpha$, on ne rejette pas H0.
- Si $p - \text{value} < \alpha$, on rejette H0.

II. Tests sur la moyenne et la variance

Tests sur la moyenne et la variance

Contexte : On considère $x_1, \ldots, x_n$ des réalisations de variables aléatoires indépendantes et identiquement distribuées $X_1, \ldots, X_n$ avec $X_1 \sim \mathcal{N}\left(\mu, \sigma^2\right)$ avec μ et σ^2 inconnus.

Test de conformité de la moyenne : Soit μ_0 une valeurs donnée, on souhaite tester

$$H0 : "\mu = \mu_0" \quad \text{contre} \quad H1 : "\mu \neq \mu_0".$$

Test de conformité de la variance : Soit σ_0^2 une valeurs donnée, on souhaite tester

$$H0 : "\sigma^2 = \sigma_0^2" \quad \text{contre} \quad H1 : "\sigma^2 \neq \sigma_0^2".$$

Test de conformité d'une moyenne

Construction de la statistique de test : On a la statistique de test :

$$Z = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n}$$

Sous H0,

$$\sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n} \sim T_{n-1}$$

où T_{n-1} suit une loi de Student à $n-1$ degrés de liberté, ce qui est faux sous H1.

Test de conformité d'une moyenne

Le test : On teste au risque $\alpha \in (0, 1)$

$$H0 : "\mu = \mu_0" \quad \text{contre} \quad H1 : "\mu \neq \mu_0".$$

On a la statistique de test

$$Z = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n} \sim T_{n-1} \quad \text{sous H0}$$

On définit la zone de rejet par

$$ZR = \left\{|Z| > t_{n-1,1-\alpha/2}\right\}$$

où $t_{n-1,1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à $n-1$ degrés de liberté.
On calcule z_{obs}, si $|z_{obs}| > t_{n-1,1-\alpha/2}$, on rejette H0, et on ne rejette pas H0 sinon.

REMARQUE

Remarque 1 : Faire ce test revient à vérifier que μ_0 est dans l'intervalle de confiance de niveau $1-\alpha$

$$IC_{1-\alpha}(\mu) = \left[\overline{x}_n - t_{n-1,1-\alpha/2}\frac{s_n}{\sqrt{n}}; \overline{x}_n + t_{n-1,1-\alpha/2}\frac{s_n}{\sqrt{n}}\right].$$

Si $\mu_0 \in IC_{1-\alpha}(\mu)$, on ne rejette pas H0 et inversement.

REMARQUE

Remarque 2 : Pour trouver la zone de rejet, cela revient à chercher c_α tel

$$\mathbb{P}_{\mu=\mu_0}\left[\text{"On rejette H0"}\right] = \mathbb{P}_{\mu=\mu_0}\left[\left|\overline{X}_n - \mu_0\right| \geq c_\alpha\right] = \alpha.$$

Remarque

Remarque 3 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté. On a alors

$$p-\text{value} = \mathbb{P}\left[|T| \geq |z_{obs}|\right].$$

TEST D'INÉGALITÉ $\mu \leq \mu_0$

On teste au risque $\alpha \in (0,1)$

$$H0 : "\mu \leq \mu_0" \quad \text{contre} \quad H1 : "\mu > \mu_0"$$

Sous H0, il existe $\mu' \leq \mu_0$ tel que

$$Z(\mu') = \sqrt{n}\frac{\overline{X}_n - \mu'}{S_n} \sim T_{n-1}.$$

On définit la zone de rejet par

$$ZR = \left\{Z\left(\mu_0\right) > t_{n-1,1-\alpha}\right\}.$$

On calcule z_{obs}, si $z_{obz} > t_{n-1,1-\alpha}$, on rejette H0 et inversement.

Remarque

Remarque 1 : Pour trouver la zone de rejet, cela revient à chercher c_α tel que

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu \left[\text{''On rejette H0''}\right] = \sup_{\mu \leq \mu_0} \mathbb{P}_\mu \left[\overline{X}_n \geq c_\alpha\right] = \alpha.$$

Remarque

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté. On a alors

$$p-\text{value} = \mathbb{P}\left[T \geq z_{obs}\right].$$

Test d'inégalité $\mu \geq \mu_0$

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\mu \geq \mu_0" \quad \text{contre} \quad H1 : "\mu < \mu_0"$$

Sous H0, il existe $\mu' \geq \mu_0$ tel que

$$Z(\mu') = \sqrt{n}\frac{\overline{X}_n - \mu'}{S_n} \sim T_{n-1}.$$

On définit la zone de rejet par

$$ZR = \left\{Z\left(\mu_0\right) < -t_{n-1,1-\alpha}\right\}.$$

On calcule z_{obs}, si $z_{obz} < -t_{n-1,1-\alpha}$, on rejette H0 et inversement.

Remarque

Remarque 1 : Pour trouver la zone de rejet, cela revient à chercher c_α tel que

$$\sup_{\mu \geq \mu_0} \mathbb{P}_\mu \left[\text{''On rejette H0''}\right] = \sup_{\mu \geq \mu_0} \mathbb{P}_\mu \left[\overline{X}_n \leq c_\alpha\right] = \alpha.$$

Remarque

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté. On a alors

$$p - \text{value} = \mathbb{P}\left[T \leq z_{obs}\right].$$

Application

A la suite d'un traitement (régime alimentaire) sur une variété de porcs, on prélève un échantillon de 5 individus et on les pèse. On obtient les poids suivants (en kg) :

$$83 \quad 81 \quad 84 \quad 80 \quad 85$$

On suppose que les x_i sont des réalisations de variables aléatoires i.i.d suivant une loi normale. On sait que le poids moyen de cette variété de porcs est de 87.6 kg. Le régime alimentaire a-t-il eu un impact sur le poids moyen ?

APPLICATION

On teste au risque de 5%

$$H0 : "\mu = 87.6" \quad \text{contre} \quad H1 : "\mu \neq 87.6".$$

On a la statistique de test

$$Z = \sqrt{5}\frac{\overline{X}_5 - 87.6}{S_n} \sim T_4 \quad \text{sous H0.}$$

On a donc la zone de rejet

$$ZR = \{|Z| \geq t_{4,0.975}\}.$$

Ici, $t_{4,0.975} = 2.78$. De plus,

$$z_{obs} = \sqrt{5}\frac{\overline{x}_5 - 87.6}{s_5} = \sqrt{5}\frac{82.6 - 87.6}{s_5} = -5.28.$$

On a $|z_{obs}| > 2.78$, et on rejette donc H0.

APPLICATION

Un biologiste affirme que le régime augmente le poids des porcs ! On teste au risque de 5%

$$H0 : "\mu \geq 87.6" \quad \text{contre} \quad H1 : "\mu < 87.6".$$

On a la zone de rejet

$$ZR = \left\{ \sqrt{5}\frac{\overline{X}_5 - 87.6}{S_5} < -t_{4,0.95} \right\}$$

Ici, $t_{4,0.95} = 2.13$ et $z_{obs} = -5.28$. On rejette donc H0.

Test de conformité d'une variance

Soit σ_0^2 connue, on souhaite tester au risque $\alpha \in (0,1)$

$$H0 : "\sigma^2 = \sigma_0^2" \quad \text{contre} \quad H1 : "\sigma^2 \neq \sigma_0^2".$$

Construction de la statistique de test : Comme

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1},$$

on a la statistique de test

$$\frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1} \quad \text{sous H0},$$

ce qui est faux sous H1.

Test de conformité d'une variance

Le test : On teste au risque $\alpha \in (0, 1)$

$$H0 : "\sigma^2 = \sigma_0^2" \quad \text{contre} \quad H1 : "\sigma^2 \neq \sigma_0^2".$$

On a la statistique de test

$$Z = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1} \quad \text{sous H0},$$

On définit la zone de rejet comme

$$ZR = \left\{Z < k_{\alpha/2}\right\} \cup \left\{Z > k_{1-\alpha/2}\right\},$$

où $k_{\alpha/2}$ et $k_{1-\alpha/2}$ sont les quantiles de la loi du Khi-deux à $n-1$ degrés de liberté.
On calcule z_{obs}. Si $z_{obs} < k_{\alpha/2}$ ou $z_{obs} > k_{1-\alpha/2}$, on rejette H0 et inversement.

Remarques

Remarque 1 : Notons que faire ce test revient à vérifier que σ_0^2 appartient à l'intervalle de confiance obtenu à partir des données observées.

Si σ_0^2 appartient à l'intervalle de confiance, on ne rejette pas H0 et inversement.

Remarques

Remarque 2 : Soit Z une variable aléatoire suivant une loi du Khi-deux à $n-1$ degrés de liberté. On a alors

$$p-\text{value} = \min\left\{2\mathbb{P}\left[Z \leq z_{obs}\right], 2\mathbb{P}\left[Z \geq z_{obs}\right]\right\}.$$

Application

On reprend l'exemple des porcs. Le même biologiste pense que si les résultats ne sont pas concluants, c'est dû à une variance qui serait égale à 25.
On va donc "vérifier" si cela est vrai.

APPLICATION

On teste au risque de 5%

$$H0 : "\sigma^2 = 25" \quad \text{contre} \quad H1 : "\sigma^2 \neq 25".$$

On a la statistique de test

$$Z = \frac{4}{25} S_n^2 \sim \chi^2_4 \quad \text{sous H0.}$$

On a la zone de rejet

$$ZR = \{Z < k_{0.025}\} \cup \{Z > k_{0.975}\}.$$

Ici, $k_{0.025} = 0.48$ et $k_{0.975} = 11.14$. De plus

$$z_{obs} = \frac{4}{25} s_n^2 = 0.29$$

Comme $z_{obs} < 0.48$, on rejette H0.

III. Comparaison de moyennes

Tests de comparaison de deux moyennes

Présentation du problème On considère deux jeux de données

- $x_1, \dots, x_p$ qui sont p réalisations d'une variable aléatoire X.
- $y_1, \dots, y_q$ qui sont q réalisations d'une variable aléatoire Y.

On souhaite comparer les moyennes théoriques des variables X et Y.

- On dispose des estimations $\overline{x}_p$ et $\overline{y}_q$.
- Les différences entre ces estimations ne sont-elles dues qu'à la fluctuation d'échantillonnage ?

Tests de comparaison de deux moyennes

Le modèle probabiliste : On fait les hypothèses suivantes :

- Les données $x_1, \ldots, x_p$ sont les réalisations de variables aléatoires $X_1, \ldots, X_p$ indépendantes et de même loi $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$.
- Les données $y_1, \ldots, y_q$ sont les réalisations de variables aléatoires $Y_1, \ldots, Y_q$ indépendantes et de même loi $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$.
- Les deux échantillons sont indépendants et de même variance, i.e $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Application

On souhaite comparer les productions laitières de deux races bovines. On choisit 50 bêtes de chaque race ($p = q = 50$) et on mesure pour chaque bête la production annuelle totale de lait (en tonnes). On obtient les productions moyennes

$$\bar{x}_{50} = 4.680 \quad \text{et} \quad \bar{y}_{50} = 4.690$$

On souhaite donc savoir si cette différence est seulement due à la fluctuation d'échantillonnage.

Test bilatéral

On veut tester au risque α

$$H0 : "\mu_1 = \mu_2" \quad \text{contre} \quad H1 : "\mu_1 \neq \mu_2".$$

Construction de la statistique de test :

Proposition

On a

$$\overline{X}_p - \overline{Y}_q \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{p} + \frac{\sigma^2}{q}\right).$$

Cependant, on ne connait pas σ^2 et on ne peut donc pas construire de statistique de test à partir de cette proposition.

Test d'égalité

On considère S^2 l'estimateur de σ^2 défini par

$$\begin{aligned} S^2 &= \frac{1}{p+q-2}\left(\sum_{i=1}^{p}\left(X_i-\overline{X}_p\right)^2+\sum_{i=1}^{q}\left(Y_i-\overline{Y}_q\right)^2\right) \\ &= \frac{1}{p+q-2}\left((p-1)S_X^2+(q-1)S_Y^2\right), \end{aligned}$$

avec $S_X^2=\frac{1}{p-1}\sum_{i=1}^{p}\left(X_i-\overline{X}_p\right)^2$ et $S_Y^2=\frac{1}{q-1}\sum_{i=1}^{q}\left(Y_i-\overline{Y}_q\right)^2$.

Proposition

1. *On a*
$$\frac{p+q-2}{\sigma^2}S^2 \sim \chi^2_{p+q-2}.$$
2. *Les variables $\overline{X}_p-\overline{Y}_q$ et S^2 sont indépendantes.*

Construction de la statistique de test

Corollaire

On a

$$\frac{\sqrt{pq}}{\sqrt{p+q}}\frac{\left(\overline{X}_p-\overline{Y}_q\right)-(\mu_1-\mu_2)}{S}\sim T_{p+q-2},$$

où T_{p+q-2} suit une loi de Student à $p+q-2$ degrés de liberté.

On obtient donc la statistique de test

$$\frac{\sqrt{pq}}{\sqrt{p+q}}\frac{\overline{X}_p-\overline{Y}_q}{S}\sim T_{p+q-2}\quad \text{sous H0},$$

ce qui est faux sous H1.

Test d'égalité de deux moyennes

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\mu_1 = \mu_2" \quad \text{contre} \quad H1 : "\mu_1 \neq \mu_2".$$

On a la statistique de test

$$Z = \frac{\sqrt{pq}}{\sqrt{p+q}} \frac{\overline{X}_p - \overline{Y}_q}{S} \sim T_{p+q-2} \quad \text{sous H0.}$$

On définit la zone de rejet

$$ZR = \left\{Z, |Z| > t_{p+q-2,1-\alpha/2}\right\},$$

où $t_{p+q-2,1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à $p+q-2$ degrés de liberté. On calcule z_{obs}. Si

$|z_{obs}| > t_{p+q-2,1-\alpha/2}$ on rejette H0 et inversement.

Remarques

Remarque 1 : Cela revient à vérifier que 0 appartient à l'intervalle de confiance de $\mu = \mu_1 - \mu_2$.
Si 0 appartient à cet intervalle de confiance, on ne rejette pas H0 et inversement.

Remarques

Remarque 2 : Pour trouver la zone de rejet, comme sous H0 $\overline{X}_q - \overline{Y}_q$ doit être proche de 0, cela revient à chercher c_α tel que

$$\mathbb{P}_{\mu=0}\left[''\text{On rejette H0}''\right] = \mathbb{P}_{\mu=0}\left(\left|\overline{X}_p - \overline{Y}_q\right| \geq c_\alpha\right] = \alpha.$$

Remarque 3 : Soit T une variable aléatoire suivant une loi de Student à $p+q-2$ degrés de liberté, on a alors

$$p - \text{value} = \mathbb{P}\left[|T| > |z_{obs}|\right].$$

Application : les Iris de Fisher

On considère un jeu de données classique : les Iris de Fisher, où on souhaite reconnaître le type d'Iris à partir de la longueur des sépales.

On note $x_1, \ldots, x_{50}$ les longueurs des Iris de la variété Virginica et $y_1, \ldots, y_{50}$ celle de la variété Versicolor et on obtient les résultats

$$\begin{aligned} \overline{x}_{50} &= 5.94 \qquad s_X = 0.52 \\ \overline{y}_{50} &= 6.59 \qquad s_Y = 0.64. \end{aligned}$$

Application : les Iris de Fisher

On test au risque de 5%

$$H0 : "\mu_1 = \mu_2" \quad \text{contre} \quad H1 : "\mu_1 \neq \mu_2".$$

On a la statistique de test

$$Z = \frac{\sqrt{50 \times 50}}{\sqrt{50 + 50}} \frac{\overline{X}_{50} - \overline{Y}_{50}}{S} \sim T_{98} \quad \text{sous H0.}$$

On a la zone de rejet

$$ZR = \{|Z| > t_{98,0.975}\}.$$

Ici $t_{98,0.975} = 1.98$ et $z_{obs} = -5.63$. On rejette donc H0.

Test d'inégalité $\mu_1 \leq \mu_2$

Le test : On veut tester au risque $\alpha \in (0,1)$

$$H0 : "\mu_1 \leq \mu_2" \quad \text{contre} \quad H1 : "\mu_1 > \mu_2".$$

Sous H0, il existe $\mu' \leq 0$ tel que

$$Z(\mu') = \frac{\sqrt{pq}}{\sqrt{p+q}} \frac{\overline{X}_p - \overline{Y}_q - \mu'}{S} \sim T_{p+q-2}.$$

On définit la zone de rejet par

$$ZR = \left\{ Z(0) > t_{p+q-2,1-\alpha} \right\}$$

où $t_{p+q-2,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Student à $p+q-2$ degrés de liberté.
On calcule z_{obs}. Si $z_{obs} > t_{p+q-2,1-\alpha}$, on rejette H0 et inversement.

REMARQUES

Remarque 1 : Pour trouver la zone de rejet, comme $\mu = \mu_1 - \mu_2$ peut être aussi petit que l'on veut sous H0, trouver la zone de rejet revient donc à chercher c_α tel que sous H0,

$$\sup_{\mu \leq 0} \mathbb{P}_\mu \left[\text{"On rejette H0"}\right] = \sup_{\mu \leq 0} \mathbb{P}_\mu \left[\overline{X}_q - \overline{Y}_q \geq c_\alpha\right] = \alpha.$$

Remarques

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $p+q-2$ degrés de liberté, on a alors

$$p - \text{value} = \mathbb{P}\left[T > z_{obs}\right].$$

TEST D'INÉGALITÉ $\mu_1 \geq \mu_2$

Le test : On veut tester au risque $\alpha \in (0,1)$

$$H0 : "\mu_1 \geq \mu_2" \quad \text{contre} \quad H1 : "\mu_1 < \mu_2".$$

Sous H0, il existe $\mu' \geq 0$ tel que

$$Z(\mu') = \frac{\sqrt{pq}}{\sqrt{p+q}} \frac{\overline{X}_p - \overline{Y}_q - \mu'}{S} \sim T_{p+q-2}.$$

On définit la zone de rejet par

$$ZR = \left\{ Z(0) < -t_{p+q-2,1-\alpha} \right\}$$

où $t_{p+q-2,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Student à $p+q-2$ degrés de liberté.
On calcule z_{obs}. Si $z_{obs} < -t_{p+q-2,1-\alpha}$, on rejette H0 et inversement.

Remarques

Remarque 1 : Pour trouver la zone de rejet, comme $\mu = \mu_1 - \mu_2$ peut être aussi grand que l'on veut sous H0, trouver la zone de rejet revient donc à chercher c_α tel que sous H0,

$$\sup_{\mu \geq 0} \mathbb{P}_\mu \left[\text{''On rejette H0''}\right] = \sup_{\mu \geq 0} \mathbb{P}_\mu \left[\overline{X}_q - \overline{Y}_q \leq c_\alpha\right] = \alpha.$$

Remarques

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $p+q-2$ degrés de liberté, on a alors

$$p - \text{value} = \mathbb{P}\left[T < z_{obs}\right].$$

IV. Tests de Fisher

Test de Fisher

On rappelle que pour les tests de comparaison de moyennes, on a du faire les hypothèses suivantes :

- Les données $x_1, \ldots, x_p$ sont les réalisations de variables aléatoires $X_1, \ldots, X_p$ indépendantes et de même loi $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$.
- Les données $y_1, \ldots, y_q$ sont les réalisations de variables aléatoires $Y_1, \ldots, Y_q$ indépendantes et de même loi $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$.
- Les deux échantillons sont indépendants et de même variance, i.e $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

L'indépendance est souvent donnée par le protocole expérimentale. On veut donc

- Tester si les variables X, Y suivent des loi normales.
- Tester si elles ont la même variance.

Test de Fisher

On veut tester au risque $\alpha \in (0,1)$

$$H0 : "\sigma_1^2 = \sigma_2^2" \quad \text{contre} \quad H1 : "\sigma_1^2 \neq \sigma_2^2".$$

Le cadre probabiliste :

- Les données $x_1, \ldots, x_p$ sont les réalisations de variables aléatoires $X_1, \ldots, X_p$ indépendantes et de même loi $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$.
- Les données $y_1, \ldots, y_q$ sont les réalisations de variables aléatoires $Y_1, \ldots, Y_q$ indépendantes et de même loi $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$.
- Les deux échantillons sont indépendants.

Construction de la statistique de test

On dispose des estimateurs S_X^2, S_Y^2 de σ_1^2, σ_2^2 définis par

$$S_X^2 = \frac{1}{p-1} \sum_{i=1}^{p} \left(X_i - \overline{X}_p\right)^2 \qquad S_Y^2 = \frac{1}{q-1} \sum_{i=1}^{q} \left(Y_i - \overline{Y}_q\right)^2 .$$

La construction du test de Fisher repose sur la définition suivante :

Définition

Soient p, q deux entiers positifs et soient $Z_p \sim \chi_p^2$ et $Z_q \sim \chi_q^2$ deux variables aléatoires indépendantes. Alors

$$\frac{Z_p/p}{Z_q/q} \sim F(p,q)$$

où $F(p,q)$ suit une loi de Fisher à p, q degrés de liberté.

Construction de la statistique de test

Corollaire

On a

$$\frac{\sigma_2^2 S_X^2}{\sigma_1^2 S_Y^2} \sim F(p-1, q-1).$$

On obtient donc la statistique de test

$$\frac{S_X^2}{S_Y^2} \sim F(p-1, q-1) \quad \text{sous H0},$$

ce qui est faux sous H1.

Test de Fisher

Le test : On teste au risque α

$$H0 : "\sigma_1^2 = \sigma_2^2" \quad \text{contre} \quad H1 : "\sigma_1^2 \neq \sigma_2^2".$$

On a la statistique de test

$$Z = \frac{S_X^2}{S_Y^2} \sim F(p-1, q-1) \quad \text{sous H0}$$

On a la zone de rejet

$$ZR = \left\{Z < f_{p-1,q-1,\alpha/2}\right\} \cup \left\{Z > f_{p-1,q-1,1-\alpha/2}\right\},$$

où $f_{p-1,q-1,\alpha/2}$ et $f_{p-1,q-1,1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ de la loi de Fisher à $p-1, q-1$ degrés de liberté.

On calcule z_{obs}. Si z_{obs} appartient à la zone de rejet, on rejette H0 et inversement.

REMARQUES

Remarque 1 : Soit $f_{p,q,\alpha}$ le quantile d'ordre α de la loi de Fisher de paramètres p, q. Alors

$$f_{p,q,1-\alpha} = \frac{1}{f_{p,q,\alpha}},$$

où $f_{p,q,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Fisher de paramètres p, q.

Remarque 2 : Soit F une variable aléatoire suivant une loi de Fisher de paramètres p, q,

$$p - \text{value} = \min\left\{\mathbb{P}\left[F \leq z_{obs}\right], \mathbb{P}\left[F \geq z_{obs}\right]\right\}.$$

Remarques

Remarque 3 : On peut intervertir S_X^2 et S_Y^2, et on obtient alors la statistique de test

$$Z = \frac{S_Y^2}{S_X^2} \sim F(q-1, p-1) \quad \text{sous H0.}$$

Remarque 4 : On voit souvent la statistique de test

$$Z = \frac{\max\left(S_X^2, S_Y^2\right)}{\min\left(S_X^2, S_Y^2\right)}$$

Application

On reprend l'exemple des Iris de Fisher. On a $s_X^2 = 0.52^2$ et $s_Y^2 = 0.64^2$. On teste au risque de 5%

$$H0 : "\sigma_1^2 = \sigma_2^2" \quad \text{contre} \quad H1 : "\sigma_1^2 \neq \sigma_2^2".$$

On a la statistique de test

$$Z = \frac{S_X^2}{S_Y^2} \sim F(49, 49) \quad \text{sous H0}$$

et la zone de rejet

$$ZR = \{Z < 0.57\} \cup \{Z > 1.76\}$$

et $z_{obs} = 0.66$ et on ne rejette donc pas H0.

Test de Shapiro-Wilk

Soient $x_1, \ldots, x_n$ des réalisations de variables aléatoires indépendantes et de même loi $X_1, \ldots, X_n$. Le test de Shapiro-Wilk permet de tester :

H0 :"la variable X est Gaussienne"

contre

H1 :"la variable X n'est pas Gaussienne"

- si la p-value est supérieure au risque α, alors on ne rejette pas H0
- sinon, on rejette H0.

Application : les Iris de Fisher

On reprend l'exemple des Iris, on obtient pour le test de Shapiro-Wilk :

- $p-$ value $= 0.46$ pour les Virginica
- $p-$ value $= 0.26$ pour les Versicolor

Dans les deux cas, on ne peut pas rejeter le caractère gaussien des données observées.

V. Cas apparié

Test de Student dans le cas apparié

On considère ici des réalisations $((x_1, y_1), \ldots, (x_n, y_n))$ d'un couple de variable aléatoire (X, Y) et on souhaite comparer les moyennes des variables aléatoires X, Y.

Le carde probabiliste :

- Les données $(x_1, y_1), \ldots, (x_n, y_n)$ sont les réalisations de couples variables aléatoires $(X_1, Y_1), \ldots, (X_n, Y_n)$ indépendantes et de même loi que (X, Y).
- Les variables X et Y ne sont pas indépendantes.
- La variable aléatoire $X - Y$ suit une loi normale d'espérance $\mu = \mu_1 - \mu_2$ et de variance σ^2.

Construction de la statistique de test

On dispose des estimateurs $\overline{X}_n$ et$\overline{Y}_n$.

Proposition

On a

$$\overline{X}_n - \overline{Y}_n \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{n}\right).$$

En particulier, on a

$$\sqrt{n}\frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sigma} \sim \mathcal{N}(0, 1).$$

Comme σ est inconnu, ce résultat est inexploitable.

Construction de la statistique de test

On considère l'estimateur S^2 de σ^2 défini par

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - Y_i - (\overline{X}_n - \overline{Z}_n) \right)^2$$

Proposition

1. *On a*

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

2. *Les variables $\overline{X}_n - \overline{Y}_n$ et S^2 sont indépendantes.*

CONSTRUCTION DE LA STATISTIQUE DE TEST

Corollaire

On a

$$\sqrt{n}\frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{S} \sim T_{n-1},$$

où T_{n-1} *suit une loi de Student à* $n-1$ *degrés de liberté.*

On obtient donc, pour le test d'égalité, la statistique de test

$$\sqrt{n}\frac{\overline{X}_n - \overline{Y}_n}{S} \sim T_{n-1} \quad \text{sous H0}$$

ce qui est faux sous H1.

Test d'égalité $\mu_1 = \mu_2$

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\mu_1 = \mu_2" \quad \text{contre} \quad H1 : "\mu_1 \neq \mu_2".$$

On a la statistique de test

$$\sqrt{n}\frac{\overline{X}_n - \overline{Y}_n}{S} \sim T_{n-1} \quad \text{sous H0.}$$

On a la zone de rejet

$$ZR = \left\{Z, |Z| > t_{n-1,1-\alpha/2}\right\}$$

où $t_{n-1,1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à $n-1$ degrés de liberté.
Si $|z_{obs}| > t_{n-1,1-\alpha/2}$, on rejette H0 et inversement.

Remarques

Remarque 1 : Faire ce test revient à vérifier que 0 appartient à l'intervalle de confiance de niveau $1-\alpha$ de $\mu = \mu_1 - \mu_2$.
Si 0 appartient à cet intervalle, on ne rejette pas H0 et inversement.

REMARQUES

Remarque 2 : Pour trouver la zone de rejet, comme $\overline{X}_n - \overline{Y}_n$ doit être proche de 0, on cherche c_α tel que sous H0,

$$\mathbb{P}_{\mu=0}\left[\text{"On rejette H0"}\right] = \mathbb{P}_{\mu=0}\left[\left|\overline{X}_n - \overline{Y}_n\right| \geq c_\alpha\right] = \alpha.$$

Remarques

Remarque 3 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté, on a alors

$$p-\text{value} = \mathbb{P}\left[|T| > |z_{obs}|\right].$$

Application

On s'intéresse à un échantillon de 30 matières fécales, que l'on soumet à deux méthodes de spectrométrie pour étudier leur teneur en lutécium radioactif. On obtient

$$\overline{x}_{30} = 120.83 \quad \text{et} \quad \overline{y}_{30} = 119.33$$

- En faisant le test pour les données appariées, on obtient une p-value égale à 0.66, et au risque de 5%, on ne rejette pas H0.
- En faisant le test de Student vu précédemment, on obtient une p-value égale à 0.0031 et on aurait alors rejeté H0.

Test d'inégalité $\mu_1 \leq \mu_2$

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\mu_1 \leq \mu_2" \quad \text{contre} \quad H1 : \mu_1 > \mu_2".$$

Sous H0, il exsite $\mu' \leq 0$ tel que

$$Z(\mu') = \sqrt{n}\frac{\overline{X}_n - \overline{Y}_n - \mu'}{S} \sim T_{n-1}.$$

On définit la zone de rejet par

$$ZR = \{Z(0) > t_{n-1,1-\alpha}\},$$

où $t_{n-1,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Student à $n-1$ degrés de liberté.
On calcule z_{obs}. Si $z_{obs} > t_{n-1,1-\alpha}$ on rejette H0 et inversement.

Remarques

Remarque 1 : Pour trouver la zone de rejet, comme $\mu = \mu_1 - \mu_2$ peut être aussi petit que l'on veut sous H0, on cherche c_α tel que

$$\sup_{\mu \leq 0} \mathbb{P}_\mu \left[\text{"On rejette H0"}\right] = \sup_{\mu \leq 0} \mathbb{P}_\mu \left[\overline{X}_n - \overline{Y}_n \geq c_\alpha\right] = \alpha.$$

Remarques

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté, on a

$$p-\text{value} = \mathbb{P}\left[T \geq z_{obs}\right].$$

TEST D'INÉGALITÉ $\mu_1 \geq \mu_2$

Le test : On teste au risque $\alpha \in (0, 1)$

$$H0 : "\mu_1 \geq \mu_2" \quad \text{contre} \quad H1 : \mu_1 < \mu_2".$$

Sous H0, il exsite $\mu' \geq 0$ tel que

$$Z(\mu') = \sqrt{n}\frac{\overline{X}_n - \overline{Y}_n - \mu'}{S} \sim T_{n-1}.$$

On définit la zone de rejet par

$$ZR = \{Z(0) < -t_{n-1,1-\alpha}\},$$

où $t_{n-1,1-\alpha}$ est le quantile d'ordre α de la loi de Student à $n-1$ degrés de liberté.
On calcule z_{obs}. Si $z_{obs} < -t_{n-1,1-\alpha}$ on rejette H0 et inversement.

Remarques

Remarque 1 : Pour trouver la zone de rejet, comme $\mu = \mu_1 - \mu_2$ peut être aussi grand que l'on veut sous H0, on cherche c_α tel que

$$\sup_{\mu \geq 0} \mathbb{P}_\mu \left[\text{"On rejette H0"}\right] = \sup_{\mu \geq 0} \mathbb{P}_\mu \left[\overline{X}_n - \overline{Y}_n \leq c_\alpha\right] = \alpha.$$

Remarques

Remarque 2 : Soit T une variable aléatoire suivant une loi de Student à $n-1$ degrés de liberté, on a

$$p-\text{value} = \mathbb{P}\left[T \leq z_{obs}\right].$$

VI. Tests du Khi-deux

Test d'indépendance du Khi-deux

Le test d'indépendance du Khi-deux permet de vérifier l'indépendance entre deux variables aléatoires X et$Y$, i.e on veut tester au risque α

$H0$: "X, Y indépendantes" contre $H1$: X, Y pas indépendantes".

Le cadre : On dispose de n données $(x_1, y_1), \ldots, (x_n, y_n)$ qui sont des réalisations de couples de variables aléatoires i.i.d (X_k, Y_k) de même loi que (X, Y) à valeurs dans $\{a_1, \ldots, a_p\} \times \{b_1, \ldots, b_q\}$.

Construction de la statistique de test

Le test du Khi-deux repose sur le fait que les variables aléatoires X et Y sont indépendantes si et seulement si pour tout $\left(a_i, b_j\right) \in \left\{a_1, \ldots, a_p\right\} \times \left\{b_1, \ldots, b_q\right\}$,

$$\mathbb{P}\left[X = a_i \text{ et } Y = b_j\right] = \mathbb{P}\left[X = a_i\right] \mathbb{P}\left[Y = b_j\right].$$

Pour tester si X et Y sont indépendantes, il "suffit" donc de comparer les estimations des probabilités jointes au produit des estimations des probabilités marginales.

Notations

On note :

- $O_{i,j} = \sum_{k=1}^{n} \mathbf{1}_{\{X_k = a_i, Y_k = b_j\}}$, qui désigne le nombre de données pour lesquelles $(X, Y) = \left(a_i, b_j\right)$.
- $O_{i,.} = \sum_{j=1}^{q} O_{i,j}$ qui désigne le nombre de données pour lesquelles $X = a_i$.
- $O_{.,j} = \sum_{i=1}^{p} O_{i,j}$ qui désigne le nombre de données pour lesquelles $Y = b_j$.
- $E_{i,j} = \frac{O_{i,.} \times O_{.,j}}{n}$.

Notons que :

- $O_{i,j}/n$ est un estimateur de $\mathbb{P}\left[X = a_i, Y = b_j\right]$.
- $E_{i,j}/n$ est un estimateur de $\mathbb{P}\left[X = a_i\right] \mathbb{P}\left[Y = b_j\right]$.

Construction de la statistique de test

Sous l'hypothèse H0, les estimations $o_{i,j}/n$ et $e_{i,j}/n$ sont censées être proches. On introduit donc la statistique de test :

$$Z = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}}.$$

Proposition

Si X et Y sont indépendantes,

$$Z \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2_{(p-1)(q-1)}.$$

La proposition précédente est fausse sous H1.

Test d'indépendance du Khi-deux

Le test : On teste au risque $\alpha \in (0,1)$,

$H0$: "X, Y indépendantes" contre $H1$: X, Y pas indépendantes".

On a la statistique de test

$$Z = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}} \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2_{(p-1)(q-1)}.$$

On a la zone de rejet

$$ZR = \left\{Z > k_{(p-1)(q-1),1-\alpha}\right\}$$

où $k_{(p-1)(-1),1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi du Khi-deux à $(p-1)(q-1)$ degrés de liberté.

On calcule z_{obs}. Si $z_{obs} > k_{(p-1)(q-1)}$, on rejette H0 et inversement.

Application :

On étudie l'influence du sexe sur la couleur des cheveux d'un groupe d'élèves. On veut savoir si la couleur des cheveux est indépendante du sexe. Pour cela, on dispose des données suivantes :

Sexe	Blond	Roux	Châtain	Brun	Noir de Jais	Total
Garçon	592	119	849	504	36	2100
Fille	544	97	677	451	14	1783
Total	1136	2126	1526	955	50	3883

On désigne par X la variable couleur des cheveux et par Y la variable sexe.

Application

On teste au risque de 5%

$H0$: "X, Y indépendantes" contre $H1$: X, Y pas indépendantes".

On a la statistique de test

$$Z = \sum_{i=1}^{5} \sum_{j=1}^{2} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi_4^2 \quad \text{sous H0.}$$

On a la zone de rejet

$$ZR = \{Z > 9.50\}\,.$$

On a $z_{obs} = 10.47$ et on rejette donc l'hypothèse H0. A noter que l'on a une p-value égale à 0.033 et on n'aurait donc pas rejeté H0 au risque de 1%.

Test d'adéquation du Khi-deux

Le cadre : On dispose de n données $x_1, \dots, x_n$ qui sont des réalisations de variables aléatoires indépendantes $X_1, \dots, X_n$ et de même loi, à valeurs dans $\{a_1, \dots, a_K\}$. On note P la loi inconnue de la variable aléatoire discrète X et on souhaite savoir si $P = P0$, où $P0$ est une loi connue.

On veut donc tester au risque $\alpha \in (0, 1)$

$H0$: "La loi de X est $P0$" contre $H1$: "La loi de X n'est pas $P0$"

Notations

On note $p_{0,1}, \ldots, p_{0,K}$ les probabilités définissant la loi P0. On note pour tout $k = 1, \ldots, K$:

- E_k l'effectif "observé" pour la modalité a_k

$$E_k = \sum_{i=1}^{n} \mathbf{1}_{\{X_j = a_k\}}$$

- N_k l'effectif théorique pour la modalité a_k sous la loi P0

$$N_k = np_{0,k}.$$

Construction de la statistique de test

La variable E_k/n est un estimateur naturel de $p_k = \mathbb{P}[X = a_k]$.
Sous H0, cet estimateur doit donc converger vers $N_k/n = p_{0,k}$.
On s'intéresse donc à la statistique de test :

$$Q^2 = \sum_{k=1}^{K} \frac{(E_k - N_k)^2}{N_k}.$$

Proposition

Si la loi de X est P0,

$$Q^2 \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2_{K-1}.$$

La proposition précédente est fausse sous H1.

Test d'adéquation du Khi-deux

Le test : On teste au risque $\alpha \in (0,1)$

$H0$: "La loi de X est $P0$" contre $H1$: "La loi de X n'est pas $P0$"

On a la statistique de test

$$Q^2 = \sum_{k=1}^{K} \frac{(E_k - N_k)^2}{N_k} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_{K-1} \quad \text{sous H0}$$

On définit la zone de rejet

$$ZR = \left\{Q^2 > k_{K-1,1-\alpha}\right\},$$

où $k_{K-1,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi du Khi-deux à $K-1$ degrés de liberté.
On calcule q^2_{obs}. Si $q^2_{obs} > k_{K-1,1-\alpha}$, on rejette H0 et inversement.

Application

Un biologiste pense qu'à un age donné :

- 50% des bébés marchent
- 12% ont une ébauche de marche
- 38% ne marchent pas

On fait une étude sur 80 bébés et on obtient :

- 35 de ces bébés ne marchent pas
- 4 ont une ébauche de marche
- 41 ne marchent pas

Application

On souhaite donc vérifier que le biologiste à tort. On résume les données dans le tableau suivant :

	Marche	Ebauche	Ne marche pas	Total
Effectif observé	35	4	41	80
Effectif Théorique	$80\times$ 0.5	$80\times$ 0.12	$80\times$ 0.38	80
"Distance"	0.625	3.267	3.696	7.588

Application

On teste donc au risque 5% l'hypothèse nulle H0 : "X suit la loi P0" contre H1 : "X ne suit pas la loi P0". On a la statistique de test

$$Q^2 = \sum_{k=1}^{3} \frac{(E_k - N_k)^2}{N_k} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2_2 \quad \text{sous H0}$$

On a la zone de rejet

$$ZR = \left\{ Q^2 > 5.99 \right\}.$$

Ici, $q^2_{obs} = 7.5878$ et on rejette donc H0.

VII. Tests asymptotiques

Tests asymptotiques

On considère ici des données $x_1, \ldots, x_n$ qui sont des réalisations de variables aléatoires $X_1, \ldots, X_n$ qui sont indépendantes et de même loi que X. On s'intéresse à l'estimation d'un paramètre θ de la loi de X, et on suppose que l'on a un estimateur $\hat{\theta}_n$ asymptotiquement normal, i.e qu'il existe $\sigma^2 > 0$ tel que

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right).$$

On suppose également que l'on dispose d'un estimateur consistant $\hat{\sigma}_n^2$ de la variance asymptotique σ^2. A l'aide du théorème de Slutsky, on a alors

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, 1\right).$$

TEST D'ÉGALITÉ

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : \text{"}\theta = \theta_0\text{"} \quad \text{contre} \quad H1 : \text{"}\theta \neq \theta_0\text{"}.$$

On a la statistique de test

$$Z = \sqrt{n}\frac{\hat{\theta}_n - \theta_0}{\hat{\sigma}_n} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1) \quad \text{sous H0.}$$

On a la zone de rejet

$$ZR = \left\{|z_{obs}| > q_{1-\alpha/2}\right\}$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite.
On calcule z_{obs}. Si $|z_{obs}| > q_{1-\alpha/2}$, on rejette H0 et inversement.

Remarques

Remarque 1 : Faire ce test revient à vérifier que θ_0 est dans l'intervalle de confiance asymptotique de niveau $1-\alpha$ de θ.

Remarques

Remarque 2 : Pour trouver la zone de rejet, on cherche c_α tel que sous H0

$$\lim_{n\to+\infty} \mathbb{P}_{\theta=\theta_0}\left[\text{"On rejette H0"}\right] = \lim_{n\to+\infty} \mathbb{P}_{\theta=\theta_0}\left[\left|\hat{\theta}_n - \theta\right| > c_\alpha\right] = \alpha.$$

Remarques

Remarque 3 : Soit Z une variable aléatoire suivant une loi normale centrée réduite, alors

$$p - \text{value} = \mathbb{P}\left[|Z| > |z_{obs}|\right].$$

TEST D'INÉGALITÉ $\theta_0 \geq \theta$

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\theta_0 \geq \theta" \quad \text{contre} \quad H1 : "\theta_0 < \theta".$$

Sous H0, il existe $\theta' \leq \theta_0$ tel que

$$Z(\theta') = \sqrt{n}\frac{\hat{\theta}_n - \theta'}{\hat{\sigma}_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On définit la zone de rejet par

$$ZR = \{Z(\theta_0) > q_{1-\alpha}\},$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite.
On calcule z_{obs}. Si $z_{obs} > q_{1-\alpha}$, on rejette H0 et inversement.

Remarques

Remarque : Soit Z une variable aléatoire suivant une loi normale centrée réduite, on a

$$p - \text{value} = \mathbb{P}\left[Z \geq z_{obs}\right].$$

TEST D'INÉGALITÉ $\theta_0 \leq \theta$

Le test : On teste au risque $\alpha \in (0,1)$

$$H0 : "\theta_0 \leq \theta" \quad \text{contre} \quad H1 : "\theta_0 > \theta".$$

Sous H0, il existe $\theta' \geq \theta_0$ tel que

$$Z(\theta') = \sqrt{n}\frac{\hat{\theta}_n - \theta'}{\hat{\sigma}_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On définit la zone de rejet par

$$ZR = \{Z(\theta_0) < -q_{1-\alpha}\},$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite.
On calcule z_{obs}. Si $z_{obs} < -q_{1-\alpha}$, on rejette H0 et inversement.

Remarques

Remarque : Soit Z une variable aléatoire suivant une loi normale centrée réduite, on a

$$p - \text{value} = \mathbb{P}\left[Z \leq z_{obs}\right].$$