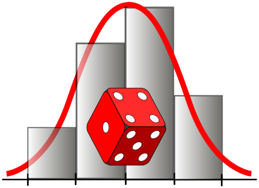
Tests d'hypothèses paramètriques



Module : Techniques d'estimation pour l'ingénieur

Plan

1. Introduction
2. Principe des tests d'hypothèses paramètriques
3. Test d'hypothèses de la :
    - La moyenne
    - La variance
    - La proportion

Test d'hypothèses sur la moyenne

On s'interesse au paramètre moyenne μ, pour une valeur de référence μ_0.
On construit un test d'hypothèses sur la moyenne :

1. Bilatéral
2. Unilatéral à droite
3. Unilatéral à gauche

Test d'hypothèses bilatéral

Étape 1: Choix des deux hypothèses $\mathbf{H}_0$ et $\mathbf{H}_1$

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = \mu_0 \\ \mathbf{H}_1 & : \quad \mu \neq \mu_0 \end{cases}$$

Étape 2: Détermination de la variable de décision et de l'écart réduit.

- **Estimateur de μ** On détermine l'estimateur qui convient pour ce test. L'estimateur de la moyenne μ est

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- **Distribution de $\overline{X}_n$** On passe à déterminer la loi de $\overline{X}_n$:

Test d'hypothèses sur la moyenne

On résume la distribution de la moyenne empirique $\overline{X_n}$ dans ce tableau récapitulatif:

$n \geq 30$ & Population de loi quelconque de moyenne μ et de variance σ^2		
Variance σ^2	$\overline{X}_n$	**Ecart réduit**
connue	$\overline{X}_n \sim \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$	$Z = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
inconnue, on utilise l'estimateur $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$	$\overline{X}_n \sim \mathcal{N}(\mu_0, \frac{s}{\sqrt{n}})$	$Z = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0,1)$

$n < 30$ & Population normale de moyenne μ et de variance σ^2		
Variance σ^2	$\overline{X}_n$	**Ecart réduit**
connue	$\overline{X}_n \sim \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$	$Z = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
inconnue, on utilise l'estimateur $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$	$T = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{n-1}$	$T \sim T_n - 1$ ddl

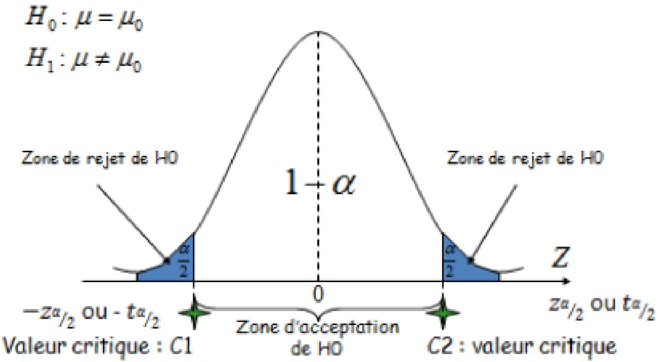
Étape 3: Détermination des régions critiques

On détermine les valeurs critiques de l'écart réduit au seuil de signification α. On impose toujours à la zone d'acceptation de **H**$_0$ que l'écart réduit soit centrée autour de 0.

Ecart réduit	Zone d'acceptation de H_0	Zone de rejet de H_0
$Z = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\mathbb{P}(\lvert Z\rvert < z_{\frac{\alpha}{2}}) = 1-\alpha$	$\mathbb{P}(Z > z_{\frac{\alpha}{2}}) = \mathbb{P}(Z < -z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$
$T = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{n-1}$	$\mathbb{P}(\lvert T\rvert < t_{\frac{\alpha}{2}}) = 1-\alpha$	$\mathbb{P}(T < -t_{\frac{\alpha}{2}}) = \mathbb{P}(T > t_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

Ci-aprés une illustration graphique des zones d'acceptation et de rejet d'un test bilatéral de la moyenne :

Test d'hypothèses sur la moyenne



Test d'hypothèses sur la moyenne

Étape 4: **Prise de décision**

On calcule l'estimation de $\overline{X}_n$ ainsi que la valeur z_0 (ou t_0) prise par l'écart réduit Z (ou T) à partir de l'échantillon et on prend notre décision suite aux valeurs déterminées. Si la valeur prise par l'écart réduit est dans la zone d'acceptation de $\mathbf{H}_0$ alors on accepte $\mathbf{H}_0$ et on rejette $\mathbf{H}_1$ sinon on rejette $\mathbf{H}_0$ et on accepte $\mathbf{H}_1$.

Ecart réduit calculé	$\mathbf{H}_0$	$\mathbf{H}_1$
$-z_{\frac{\alpha}{2}} \leq z_0 \leq z_{\frac{\alpha}{2}}$ (resp. $-t_{\frac{\alpha}{2}} \leq t_0 \leq t_{\frac{\alpha}{2}}$)	**acceptée**	rejetée
$-z_{\frac{\alpha}{2}} \geq z_0$ ou $z_0 \geq z_{\frac{\alpha}{2}}$ (resp. $-t_{\frac{\alpha}{2}} \geq t_0$ ou $t_0 \geq t_{\frac{\alpha}{2}}$)	rejetéee	**acceptée**

Application

Une entreprise utilise un ancien modèle de fabrication d'écrans dont la durée de vie en heures suit une loi normale de moyenne 15000 et d'écart type σ.

Une nouvelle méthode de fabrication est mise en application sur une autre chaine . On examine un échantillon de 100 écrans issus de cette nouvelle chaine, on s'appercoit que la durée de vie moyenne est de 15100.

1. Sachant que $\sigma = 1000$, au seuil de risque de 5%, peut-on dire que le modèle de fabrication est différent de l'ancien en terme de durée de vie moyenne des écrans?
2. L'échantillon utilisé nous donne un écart-type égale à 600, au seuil de risque de 10%, peut-on dire que le modèle de fabrication est différent de l'ancien en terme de durée de vie moyenne des écrans?

Solution

1. **Étape 1:** Choix d'hypothèses

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = 15000 = \mu_0 \\ \mathbf{H}_1 & : \quad \mu \neq 15000 \end{cases}$$

Étape 2: Variable de décision et écart réduit.

On a $n = 100 > 30$, $\sigma = 1000$ connu

$$\overline{X}_n \sim \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}}) = \mathcal{N}(15000, \frac{1000}{\sqrt{100}}) = \mathcal{N}(15000, 100)$$

Par la suite

$$Z = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Étape 3: Régions critiques.

On détermine $z_{\frac{\alpha}{2}}$ à partir du tableau de la loi normale centrée réduite,

$$\alpha = 0.05 \quad \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$$

On calcule la valeur z_0 prise par Z

$$z_0 = \frac{\overline{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{15100 - 15000}{\frac{1000}{\sqrt{100}}} = 1$$

Étape 4: Décision

$$-z_{\frac{\alpha}{2}} < z_0 < z_{\frac{\alpha}{2}}$$

On est dans la zone d'acceptation de $\mathbf{H}_0$, alors le modèle de fabrication n'est pas différent de l'ancien d'une manière significative

2. **Étape 1:** Choix d'hypothèses

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = 15000 = \mu_0 \\ \mathbf{H}_1 & : \quad \mu \neq 15000 \end{cases}$$

Étape 2: Variable de décision et écart réduit

On a $n = 100 > 30$ et σ inconnu mais estimé par $s = 600$

$$\overline{X}_n \sim \mathcal{N}(\mu_0, \frac{s}{\sqrt{n}}) = \mathcal{N}(15000, \frac{600}{\sqrt{100}}) = \mathcal{N}(15000, 60)$$

Par la suite

$$Z = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Test d'hypothèses sur la moyenne

Étape 3: Régions critiques

On détermine $z_{\frac{\alpha}{2}}$ à partir du tableau de la loi normale centrée réduite,

$$\alpha = 0.1 \quad \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.64$$

On calcule la valeur z_0 prise par Z

$$z_0 = \frac{\overline{x}_n - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{15100 - \mu_0}{\frac{600}{\sqrt{100}}} = 1.66$$

Étape 4: Décision

$$z_0 > z_{\frac{\alpha}{2}}$$

On est dans la zone de rejet de $\mathbf{H}_0$ et acceptation de $\mathbf{H}_1$, alors le modèle de fabrication est différent de l'ancien d'une manière significative

Test d'hypothèses unilatéral à droite

Étape 1: **Choix des deux hypothèses $\mathbf{H}_0$ et $\mathbf{H}_1$**

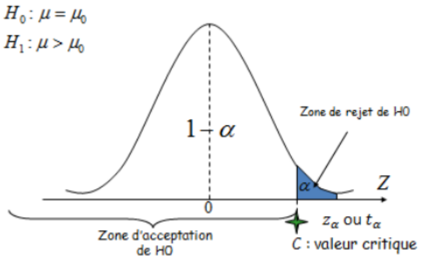
$$\begin{cases} \mathbf{H}_0 & : \quad \mu = \mu_0 \\ \mathbf{H}_1 & : \quad \mu > \mu_0 \end{cases}$$

Étape 2: **Détermination de la variable de décision et de l'écart réduit**

On reprend le même tableau que pour le test bilatéral.

Étape 3: Détermination des régions critiques

On détermine les valeurs critiques de l'écart réduit au seuil de signification α. On détermine alors les valeurs z_α ou t_α.



Ecart réduit	Zone d'acceptation de H_0	Zone de rejet de H_0
$Z = \dfrac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\mathbb{P}(Z < z_\alpha) = 1-\alpha$	$\mathbb{P}(Z > z_\alpha) = \alpha$
$T = \dfrac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{n-1}$	$\mathbb{P}(T < t_\alpha) = 1-\alpha$	$\mathbb{P}(T > t_\alpha) = \alpha$

Test d'hypothèses sur la moyenne

Étape 4: Prise de décision.

On calcule l'estimation de $\overline{X}_n$ ainsi que la valeur z_0 (ou t_0) prise par l'écart réduit Z (ou T) à partir de l'échantillon et on prend notre décision suite aux valeurs déterminées. Si la valeur prise par l'écart réduit est dans la zone d'acceptation de $\mathbf{H}_0$ alors on accepte $\mathbf{H}_0$ et on rejette $\mathbf{H}_1$ sinon on rejette $\mathbf{H}_0$ et on accepte $\mathbf{H}_1$.

Ecart réduit calculé	$\mathbf{H}_0$	$\mathbf{H}_1$
$z_0 \leq z_\alpha$ (resp. $t_0 \leq t_\alpha$)	**acceptée**	rejetée
$z_0 \geq z_\alpha$ (resp. $t_0 \geq t_\alpha$)	rejetée	**acceptée**

Application

On extrait d'une population, dont le caractère masse est distribué selon le modèle normal, un échantillon aléatoire simple de 7 éléments. La moyenne de cet échantillon est $257,9$. L'écart type de la population est supposé inconnu et l'écart type de l'échantillon est égal à 14. Est-il possible, au seuil de 5%, de conclure que la moyenne de la population est supérieure à 250 ?

Solution

Étape 1: Choix d'hypothèses

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = 250 = \mu_0 \\ \mathbf{H}_1 & : \quad \mu > 250 \end{cases}$$

Étape 2: Variable de décision et écart réduit

On a $n = 7 < 30$ et σ inconnu, mais estimé par $s = 14$

$$T = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{(n-1)} = T_6 \quad \text{Student de } (n-1) = 6 \text{ ddl}$$

Test d'hypothèses sur la moyenne

Étape 3: Régions critiques

On détermine t_α à partir du tableau de la loi Student,

$$\alpha = 0.05 \quad \Rightarrow t_\alpha = 1.943$$

On calcule la valeur t_0 prise par T

$$t_0 = \frac{\overline{x}_n - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{257.9 - 250}{\frac{14}{\sqrt{7}}} = 1.49$$

Étape 4: Décision

$$1.49 = t_0 < t_\alpha = 1.943$$

On est dans la zone d'acceptation de **H**$_0$, alors la moyenne de la population n'est pas supérieure à 250 d'une manière significative.

esprit Se former autrement

Test d'hypothèses unilatéral à gauche

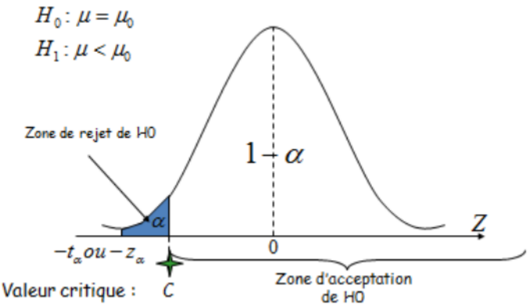
Étape 1: Choix des deux hypothèses $\mathbf{H}_0$ et $\mathbf{H}_1$.

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = \mu_0 \\ \mathbf{H}_1 & : \quad \mu < \mu_0 \end{cases}$$

Étape 2: Détermination de la variable de décision et de l'écart réduit.

On reprend le même tableau que pour le test bilatéral.

Étape 3: Détermination des régions critiques



Ecart réduit	Zone d'acceptation de $\mathbf{H}_0$	Zone de rejet de $\mathbf{H}_0$
$Z = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\mathbb{P}(Z > -z_\alpha) = 1-\alpha$	$\mathbb{P}(Z < -z_\alpha) = \alpha$
$T = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{n-1}$	$\mathbb{P}(T > -t_\alpha) = 1-\alpha$	$\mathbb{P}(T < -t_\alpha) = \alpha$

Test d'hypothèses sur la moyenne

Étape 4: Prise de décision.

On calcule l'estimation de $\overline{X}_n$ ainsi que la valeur z_0 (ou t_0) prise par l'écart réduit Z (ou T) à partir de l'échantillon et on prends notre décision suite aux valeurs déterminées. Si la valeur de l'écart réduit est dans la zone d'acceptation de $\mathbf{H}_0$ alors on accepte $\mathbf{H}_0$ et on rejette $\mathbf{H}_1$ sinon on rejette $\mathbf{H}_0$ et on accepte $\mathbf{H}_1$.

Ecart réduit calculé	$\mathbf{H}_0$	$\mathbf{H}_1$
$z_0 \geq -z_\alpha$ (resp. $t_0 \geq -t_\alpha$)	**acceptée**	rejetée
$z_0 \leq -z_\alpha$ (resp.$t_0 \leq -t_\alpha$)	rejetée	**acceptée**

Application

Dans une entreprise de conditionnement de colis, chaque employé est supposé s'occuper, en moyenne, de 45 colis par jour. Le chef de service soupçonne un employé Mr Slow, de travailler lentement et il effectue quelques mesures à son insu. Il note, sur une période de 10 jours, le nombre de colis qu'il traite quotidiennement. Il obtient les résultats suivants (On supposera que le nombre de colis traités quotidiennement par un employé suit une loi normale):

44	3	45	46	34	39	43	40	44	48

Pour savoir si Mr Slow est plus lent que ses collègues de travail, testez au seuil de signification de 5%, les hypothèses:

$$\begin{cases} \mathbf{H}_0 & : \quad \mu = 45 \\ \mathbf{H}_1 & : \quad \mu < 45 \end{cases}$$

Solution

Étape 1: Choix d'hypothèses

$$\begin{cases} \mathbf{H}_0 & : & \mu = 45 \\ \mathbf{H}_1 & : & \mu < 45 \end{cases}$$

Étape 2: Variable de décision et écart réduit.

On a $n = 10 < 30$ et σ inconnu, mais estimé par s, on doit calculer alors les estimations $\overline{x}_n$ et s^2

$$\begin{aligned} \overline{x}_n &= \tfrac{1}{n} \textstyle\sum_{i=1}^{10} x_i = 38.6 \\ s^2 &= \tfrac{1}{n-1} \textstyle\sum_{i=1}^{10} (x_i - \overline{x}_n)^2 = 172.48 \Rightarrow s = 13.13 \end{aligned}$$

$$T = \frac{\overline{X}_n - \mu_0}{\frac{s}{\sqrt{n}}} \sim T_{(n-1)} = T_9 \quad \text{Student de } (n-1) = 9 \text{ ddl}$$

Test d'hypothèses sur la moyenne

Étape 3: Régions critiques.

On détermine t_α à partir du tableau de la loi Student,

$$\alpha = 0.05 \quad \Rightarrow t_\alpha = 1.833$$

On calcule la valeur t_0 prise par T

$$t_0 = \frac{\overline{x}_n - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{38.6 - 45}{\frac{13.13}{\sqrt{10}}} = -1.54$$

Étape 4: Décision

$$-1.833 = -t_\alpha < t_0 = -1.54$$

On est dans la zone d'acceptation de $\mathbf{H}_0$, alors Mr Slow n'est pas lent que ses collègues de travail d'une manière significative.

esprit Se former autrement