Probabilités et observations économiques

C. Tuleau-Malot

Université de Nice - Sophia Antipolis

Plan

Comparaison de la moyenne à une norme : σ^2 inconnu

On a toujours $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0; 1)$.

Comparaison de la moyenne à une norme : σ^2 inconnu

On a toujours $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0; 1)$.

Mais Z_n dépend d'un paramètre inconnu

Comparaison de la moyenne à une norme : σ^2 inconnu

On a toujours $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0;1)$.

Mais Z_n dépend d'un paramètre inconnu : σ

Comparaison de la moyenne à une norme : σ^2 inconnu

On a toujours $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0; 1)$.

Mais Z_n dépend d'un paramètre inconnu : σ

Donc, on ne peut pas utiliser Z_n pour déterminer la zone de rejet

Comparaison de la moyenne à une norme : σ^2 inconnu

On a toujours $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0; 1)$.

Mais Z_n dépend d'un paramètre inconnu : σ

Donc, on ne peut pas utiliser Z_n pour déterminer la zone de rejet

On utilise la même astuce que dans la construction d'un intervalle de confiance à variance inconnue

Comparaison de la moyenne à une norme : σ^2 inconnu (2)

- on estime σ^2 par $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

Comparaison de la moyenne à une norme : σ^2 inconnu (2)

- on estime σ^2 par $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$
- si on a l'hypothèse de normalité sur X, alors on utilise le fait que $T_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} \sim \mathcal{T}(n-1)$

Comparaison de la moyenne à une norme : σ^2 inconnu (2)

- on estime σ^2 par $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$
- si on a l'hypothèse de normalité sur X, alors on utilise le fait que $T_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} \sim \mathcal{T}(n-1)$
- sans hypothèse de normalité et avec n grand, alors on utilise le fait que $T_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} \sim \mathcal{N}(0;1)$

Comparaison de la moyenne à une norme : test unilatéral

Qu'est ce qui change si on procède non pas à un test bilatéral mais à un test unilatéral?

Comparaison de la moyenne à une norme : test unilatéral

Qu'est ce qui change si on procède non pas à un test bilatéral mais à un test unilatéral?

- test unilatéral supérieur ($\mathcal{H}_0 : \mu \leq \mu_0$)
  La zone de rejet est de la forme $z_n \in]t, +\infty[$ avec t le fractile d'ordre $1 - \alpha$ d'une loi $\mathcal{N}(0; 1)$ ou d'une loi $\mathcal{T}(n-1)$ selon le cadre

Comparaison de la moyenne à une norme : test unilatéral

Qu'est ce qui change si on procède non pas à un test bilatéral mais à un test unilatéral?

- test unilatéral supérieur ($\mathcal{H}_0 : \mu \leq \mu_0$)
  La zone de rejet est de la forme $z_n \in]t, +\infty[$ avec t le fractile d'ordre $1 - \alpha$ d'une loi $\mathcal{N}(0; 1)$ ou d'une loi $\mathcal{T}(n-1)$ selon le cadre

- test unilatéral inférieur ($\mathcal{H}_0 : \mu \geq \mu_0$)
  La zone de rejet est de la forme $z_n \in]-\infty, -t[$ avec t le fractile d'ordre $1 - \alpha$ d'une loi $\mathcal{N}(0; 1)$ ou d'une loi $\mathcal{T}(n-1)$ selon le cadre

Test pour une proportion

Exemple : On sait qu'en 2008, la proportion de fumeurs dans une population donnée était de 38%. Une enquête menée sur 235 personnes donnent 79 fumeurs. Peut-on affirmer que la proportion de fumeurs a diminué?

Test pour une proportion

Exemple : On sait qu'en 2008, la proportion de fumeurs dans une population donnée était de 38%. Une enquête menée sur 235 personnes donnent 79 fumeurs. Peut-on affirmer que la proportion de fumeurs a diminué?

Méthodologie :

- on se fixe $\alpha = 0.05$

Test pour une proportion

Exemple : On sait qu'en 2008, la proportion de fumeurs dans une population donnée était de 38%. Une enquête menée sur 235 personnes donnent 79 fumeurs. Peut-on affirmer que la proportion de fumeurs a diminué?

Méthodologie :

- on se fixe $\alpha = 0.05$
- on choisit les hypothèses :

$$\begin{aligned}\mathcal{H}_0 &: \ p \geq p_0 = 0.38\\ \mathcal{H}_a &: \ p < p_0 = 0.38\end{aligned}$$

Test pour une proportion

Exemple : On sait qu'en 2008, la proportion de fumeurs dans une population donnée était de 38%. Une enquête menée sur 235 personnes donnent 79 fumeurs. Peut-on affirmer que la proportion de fumeurs a diminué?

Méthodologie :

- on se fixe $\alpha = 0.05$
- on choisit les hypothèses :

$$\begin{array}{l} \mathcal{H}_0 : p \geq p_0 = 0.38 \\ \mathcal{H}_a : p < p_0 = 0.38 \end{array}$$

- il faut créer une statistique de test à savoir une variable aléatoire dont on connait la loi

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

calcul de la p-valeur :

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

calcul de la p-valeur :

ici $\hat{p}_n = 79/235$

Test pour une proportion (2)

<u>idée :</u> faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

calcul de la p-valeur :

ici $\hat{p}_n = 79/235$

on calcule $z_n = \sqrt{n}\frac{\hat{p}_n - p_0}{\sqrt{p_0 * (1-p_0)}} = \sqrt{235}\frac{79/235 - 0.38}{\sqrt{0.38*(1-0.38)}} \approx -1.38$

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

calcul de la p-valeur :

ici $\hat{p}_n = 79/235$

on calcule $z_n = \sqrt{n}\frac{\hat{p}_n - p_0}{\sqrt{p_0 * (1-p_0)}} = \sqrt{235}\frac{79/235 - 0.38}{\sqrt{0.38 * (1-0.38)}} \approx -1.38$

On calcule $P(\mathcal{N}(0;1) < z_n)$; on trouve 1-0.9162=0.0838

Test pour une proportion (2)

idée : faire comme on a fait pour la construction de l'intervalle de confiance

$\Rightarrow$ utiliser $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow$ on sait que cette variable suit approximativement une loi normale d'espérance p et de variance $p(1-p)/n$

$\Rightarrow$ sous $\mathcal{H}_0$ avec égalité, $\hat{p}_n$ suit approximativement une loi normale d'espérance p_0 et de variance $p_0(1-p_0)/n$

calcul de la p-valeur :

ici $\hat{p}_n = 79/235$

on calcule $z_n = \sqrt{n}\frac{\hat{p}_n - p_0}{\sqrt{p_0 * (1-p_0)}} = \sqrt{235}\frac{79/235 - 0.38}{\sqrt{0.38 * (1-0.38)}} \approx -1.38$

On calcule $P(\mathcal{N}(0;1) < z_n)$; on trouve 1-0.9162=0.0838

puisque $0.05 < 0.0838$, dans le cas présent on ne rejette pas $\mathcal{H}_0$

Résumé :

- Test sur la moyenne de la population à variance connue $\Rightarrow Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$ (cadre loi normale ou $n > 30$)
  - test bilatéral : zone de rejet $]-\infty; -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}; +\infty[$
  - test unilatéral supérieur : zone de rejet $]z_{1-\alpha}; +\infty[$
  - test unilatéral inférieur : zone de rejet $]-\infty; -z_{1-\alpha}[$

Résumé :

- Test sur la moyenne de la population à variance inconnue $\Rightarrow T = \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n/\sqrt{n}}$
  - on sait que loi normale : $\Rightarrow T \sim T(n-1)$
    - test bilatéral : zone de rejet $]-\infty; -t_{1-\alpha/2}[\cup]t_{1-\alpha/2}; +\infty[$
    - test unilatéral supérieur : zone de rejet $]t_{1-\alpha}; +\infty[$
    - test unilatéral inférieur : zone de rejet $]-\infty; -t_{1-\alpha}[$
  - on ne connaît pas la loi : il faut obligatoirement n grand et alors on dit que T suit approximativement une loi normale centrée réduite
    - test bilatéral : zone de rejet $]-\infty; -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}; +\infty[$
    - test unilatéral supérieur : zone de rejet $]z_{1-\alpha}; +\infty[$
    - test unilatéral inférieur : zone de rejet $]-\infty; -z_{1-\alpha}[$

Résumé :

- Test sur une proportion : $Z = \frac{\bar{p}_n - \varphi_0}{\sqrt{p_0.(1-p_0)}/\sqrt{n}}$ approximativement une loi normale centrée réduite
  - test bilatéral : zone de rejet $]-\infty; -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}; +\infty[$
  - test unilatéral supérieur : zone de rejet $]z_{1-\alpha}; +\infty[$
  - test unilatéral inférieur : zone de rejet $]-\infty; -z_{1-\alpha}[$

Comparaison de deux populations

Exemple : On considère deux centres commerciaux et on veut comparer la dépense moyenne des clients dans chacun des centres.

Comparaison de deux populations

Exemple : On considère deux centres commerciaux et on veut comparer la dépense moyenne des clients dans chacun des centres.

Méthodologie :

- Population 1 avec une moyenne m_1 et une variance σ_1^2
  Echantillon 1 : $(X_1, \ldots, X_n)$ et on considère $X_n = \frac{X_1 + \ldots + X_n}{n}$
- Population 2 avec une moyenne m_2 et une variance σ_2^2
  Echantillon 2 : $(Y_1, \ldots, Y_l)$ et on considère $Y_l = \frac{Y_1 + \ldots + Y_l}{l}$

Comparaison de deux populations

Exemple : On considère deux centres commerciaux et on veut comparer la dépense moyenne des clients dans chacun des centres.

Méthodologie :

- Population 1 avec une moyenne m_1 et une variance σ_1^2
  Echantillon 1 : $(X_1, \ldots, X_n)$ et on considère $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$
- Population 2 avec une moyenne m_2 et une variance σ_2^2
  Echantillon 2 : $(Y_1, \ldots, Y_l)$ et on considère $\bar{Y}_l = \frac{Y_1 + \ldots + Y_l}{l}$

Cadre : 2 possibilités

- on suppose que les deux échantillons sont de loi normale et on connaît les variances
- on ne connaît pas les variances et l et n grands

Comparaison de deux populations (2)

Dans les deux cadres, on a :

- $\bar{X}_n$ suit approximativement $\mathcal{N}(m_1, \sigma_1^2/n)$
- $\bar{Y}_l$ suit approximativement $\mathcal{N}(m_2, \sigma_2^2/l)$

Comparaison de deux populations (2)

Dans les deux cadres, on a :

- $\bar{X}_n$ suit approximativement $\mathcal{N}(m_1, \sigma_1^2/n)$
- $\bar{Y}_l$ suit approximativement $\mathcal{N}(m_2, \sigma_2^2/l)$

Or nous on veut regarder comment est $m_1 - m_2$!

Comparaison de deux populations (2)

Dans les deux cadres, on a :

- $\bar{X}_n$ suit approximativement $\mathcal{N}(m_1, \sigma_1^2/n)$
- $\bar{Y}_l$ suit approximativement $\mathcal{N}(m_2, \sigma_2^2/l)$

Or nous on veut regarder comment est $m_1 - m_2$!
$\Rightarrow$ on considère $\bar{X}_n - \bar{Y}_l$

Comparaison de deux populations (2)

Dans les deux cadres, on a :

- $\bar{X}_n$ suit approximativement $\mathcal{N}(m_1, \sigma_1^2/n)$
- $\bar{Y}_l$ suit approximativement $\mathcal{N}(m_2, \sigma_2^2/l)$

Or nous on veut regarder comment est $m_1 - m_2$!
$\Rightarrow$ on considère $\bar{X}_n - \bar{Y}_l$

$$\bar{X}_n - \bar{Y}_l \sim \mathcal{N}(m_1 - m_2; \sigma_1^2/n + \sigma_2^2/l)$$

Comparaison de deux populations (2)

Dans les deux cadres, on a :

- $\bar{X}_n$ suit approximativement $\mathcal{N}(m_1, \sigma_1^2/n)$
- $\bar{Y}_l$ suit approximativement $\mathcal{N}(m_2, \sigma_2^2/l)$

Or nous on veut regarder comment est $m_1 - m_2$!
$\Rightarrow$ on considère $\bar{X}_n - \bar{Y}_l$

$$\bar{X}_n - \bar{Y}_l \sim \mathcal{N}(m_1 - m_2; \sigma_1^2/n + \sigma_2^2/l)$$

- si σ_1^2 et σ_2^2 connus : il suffit de connaître $m_1 - m_2$ pour connaître précisément la loi
- si σ_1^2 et σ_2^2 inconnus : on les remplace par $\hat{\sigma_1}^2$ et $\hat{\sigma_2}^2$ puis il suffit de connaître $m_1 - m_2$ pour connaître précisément la loi

Comparaison de deux populations (3)

Test :

$$\mathcal{H}_0 : m_1 = m_2$$
$$\mathcal{H}_a : m_1 \neq m_2$$

Comparaison de deux populations (3)

Test :

$$\mathcal{H}_0 : m_1 = m_2$$

$$\mathcal{H}_a : m_1 \neq m_2$$

Exemple : $n = 120$, $l = 150$, $\bar{x}_n = 82$, $\bar{y}_l = 78$, $\hat{\sigma}_1 = 11$, $\hat{\sigma}_2 = 9$ et $\alpha = 0.1$

Comparaison de deux populations (3)

Test :

$$\mathcal{H}_0 : m_1 = m_2$$

$$\mathcal{H}_a : m_1 \neq m_2$$

Exemple :$n = 120$, $l = 150$, $\bar{x}_n = 82$,$\bar{y}_l = 78$, $\hat{\sigma}_1 = 11$, $\hat{\sigma}_2 = 9$ et $\alpha = 0.1$

- on calcule : $Z = \frac{\bar{X}_n - \bar{Y}_l}{\sqrt{\hat{\sigma}_1^2/n + \hat{\sigma}_2^2/l}}$

Comparaison de deux populations (3)

Test :

$$\mathcal{H}_0 : m_1 = m_2$$
$$\mathcal{H}_a : m_1 \neq m_2$$

Exemple : $n = 120$, $l = 150$, $\bar{x}_n = 82$, $\bar{y}_l = 78$, $\hat{\sigma}_1 = 11$, $\hat{\sigma}_2 = 9$ et $\alpha = 0.1$

- on calcule : $Z = \frac{\bar{X}_n - \bar{Y}_l}{\sqrt{\hat{\sigma}_1^2/n + \hat{\sigma}_2^2/l}}$
- $z \simeq 3.21$ et on regarde la p-valeur associée qui est $P(|\mathcal{N}(0;1)| > z)$
- on lit 0.0013 et donc on décide de

Comparaison de deux populations (3)

Test :

$$\mathcal{H}_0 : m_1 = m_2$$
$$\mathcal{H}_a : m_1 \neq m_2$$

Exemple : $n = 120$, $l = 150$, $\bar{x}_n = 82$, $\bar{y}_l = 78$, $\hat{\sigma}_1 = 11$, $\hat{\sigma}_2 = 9$ et $\alpha = 0.1$

- on calcule : $Z = \frac{\bar{X}_n - \bar{Y}_l}{\sqrt{\hat{\sigma}_1^2/n + \hat{\sigma}_2^2/l}}$
- $z \simeq 3.21$ et on regarde la p-valeur associée qui est $P(|\mathcal{N}(0;1)| > z)$
- on lit 0.0013 et donc on décide de rejeter $\mathcal{H}_0$