h e g
Haute école de gestion de Genève
Geneva School of Business Administration

Test d'une proportion
Test d'une variance

Dr. Sacha Varone

heg

Objectif

Savoir effectuer un test

- sur une proportion
- sur une variance

Rappels

Étapes d'un test

1. Spécifier la valeur de la population d'intérêt.
2. Formuler l'hypothèse nulle H_0 et l'hypothèse alternative H_1
3. Choisir le niveau de signification α
4. Déterminer la région critique.
5. Calculer la statistique associée à l'échantillon.
6. Rejeter H_0 si la statistique appartient à la région critique. Ne pas rejeter H_0 dans le cas contraire.
7. Énoncer une conclusion.

Test de μ, σ^2 connu

Population qui suit une loi normale de variance σ^2 connue
Par le TCL, $\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Test de μ, σ^2 connu

Population qui suit une loi normale de variance σ^2 connue
Par le TCL, $\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Par conséquent, pour tester une hypothèse sur μ, on utilisera la statistique

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

où :

$\bar{x}$	=	moyenne de l'échantillon
$z_{\alpha/2}$	=	valeur critique de la distribution normale standard pour un degré de confiance de $1-\alpha$
σ	=	écart type de la population
n	=	taille de l'échantillon

Test de μ, σ^2 inconnu, n grand

Par le TCL, si $n \geq 30$,

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

où

$$\begin{aligned} \bar{x} &= \text{moyenne de l'échantillon} \\ \mu &= \text{moyenne supposée de la population étudiée} \\ s &= \text{écart type de l'échantillon} \\ n &= \text{taille de l'échantillon} \end{aligned}$$

Test de μ, σ^2 inconnu, n petit

Utiliser la distribution de Student, à $n-1$ degrés de liberté

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

où

$\bar{x}$ = moyenne de l'échantillon
μ = moyenne supposée de la population étudiée
s = écart type de l'échantillon
n = taille de l'échantillon

Supposition : POPULATION NORMALEMENT DISTRIBUÉE

Test d'une proportion

Statistique de test

$$z = \frac{\bar{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

où

$\bar{p}$ = proportion de l'échantillon

π = proportion supposée de la population étudiée

n = taille de l'échantillon

Supposition : TAILLE DE L'ÉCHANTILLON SUFFISAMMENT GRANDE

En pratique

$$n\bar{p} \geq 5 \quad \text{et} \quad n(1-\bar{p}) \geq 5$$

Exemple

Un contrôle interne doit être effectué dans une banque pour vérifier que les contrats des hypothèques accordées comportent tous les documents nécessaires. Parfois, un contrat ne comporte pas tous les documents, auquel cas, le dossier n'est pas complet. La banque a octroyé 22500 hypothèques. et exige qu'il n'y ait pas plus d'1% de contrats incomplets. L'équipe chargée du contrôle n'a pas le temps d'examiner les 22500 contrats et décide de contrôler 600 contrats choisis aléatoirement, et fixe le niveau de signification à 0.02. Elle trouve 9 contrats incomplets.

1. La valeur de la population d'intérêt est

Exemple

Un contrôle interne doit être effectué dans une banque pour vérifier que les contrats des hypothèques accordées comportent tous les documents nécessaires. Parfois, un contrat ne comporte pas tous les documents, auquel cas, le dossier n'est pas complet. La banque a octroyé 22500 hypothèques. et exige qu'il n'y ait pas plus d'1% de contrats incomplets. L'équipe chargée du contrôle n'a pas le temps d'examiner les 22500 contrats et décide de contrôler 600 contrats choisis aléatoirement, et fixe le niveau de signification à 0.02. Elle trouve 9 contrats incomplets.

1. La valeur de la population d'intérêt est
   la proportion de contrats incomplets $\pi = 0.01$
   $600 \times 0.01 = 6 \geq 5 \quad \text{et} \quad 600 \times (1 - 0.01) = 594 \geq 5$
   l'utilisation de la statistique z est alors possible.

2. Les hypothèses du test sont

Exemple

Un contrôle interne doit être effectué dans une banque pour vérifier que les contrats des hypothèques accordées comportent tous les documents nécessaires. Parfois, un contrat ne comporte pas tous les documents, auquel cas, le dossier n'est pas complet. La banque a octroyé 22500 hypothèques. et exige qu'il n'y ait pas plus d'1% de contrats incomplets. L'équipe chargée du contrôle n'a pas le temps d'examiner les 22500 contrats et décide de contrôler 600 contrats choisis aléatoirement, et fixe le niveau de signification à 0.02. Elle trouve 9 contrats incomplets.

1. La valeur de la population d'intérêt est
   la proportion de contrats incomplets $\pi = 0.01$
   $600 \times 0.01 = 6 \geq 5$ et $600 \times (1 - 0.01) = 594 \geq 5$
   l'utilisation de la statistique z est alors possible.

2. Les hypothèses du test sont
   $H_0 : \pi \leq 0.01$
   $H_1 : \pi > 0.01$

Exemple (suite)

1. Le niveau de signification est

Exemple (suite)

1. Le niveau de signification est $\alpha = 0.02$
2. La région critique est

Exemple (suite)

1. Le niveau de signification est $\alpha = 0.02$
2. La région critique est $[z_\alpha \approx 2.05; \infty[$
3. La statistique associée à l'échantillon est

heg

Exemple (suite)

1. Le niveau de signification est $\alpha = 0.02$
2. La région critique est $[z_\alpha \approx 2.05; \infty[$
3. La statistique associée à l'échantillon est

$$\bar{p} = 9/600 = 0.015 \quad z = \frac{0.015 - 0.01}{\sqrt{\frac{0.01(1-0.01)}{600}}} = 1.23$$

4. Comme $z = 1.23 < 2.05$, H_0 n'est pas rejetée.
5. Conclusion :

Exemple (suite)

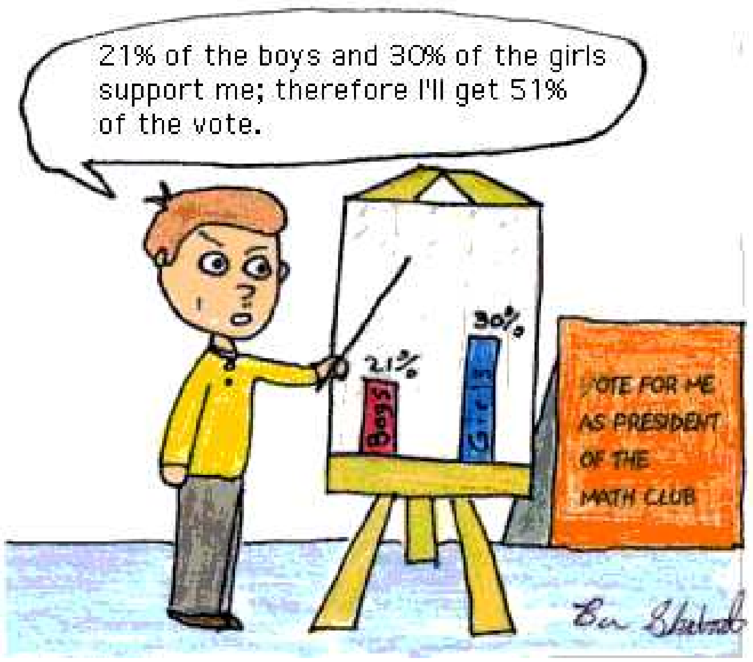
1. Le niveau de signification est $\alpha = 0.02$
2. La région critique est $[z_\alpha \approx 2.05; \infty[$
3. La statistique associée à l'échantillon est

$$\bar{p} = 9/600 = 0.015 \quad z = \frac{0.015 - 0.01}{\sqrt{\frac{0.01(1-0.01)}{600}}} = 1.23$$

4. Comme $z = 1.23 < 2.05$, H_0 n'est pas rejetée.
5. Conclusion :
   L'équipe de contrôle interne peut donc supposer que l'exigence de la banque est satisfaite.

Test unilatéral et bilatéral

- Lors d'un test unilatéral, la valeur critique est déterminée par z_α ou $-z_\alpha$
- Lors d'un test bilatéral, les valeurs critiques sont déterminées par $\pm z_{\alpha/2}$



source : http ://davidmlane.com/ben/cartoons.html

Test d'une variance

Pourquoi sur la variance ?

Intérêt : dispersion des données
$\Rightarrow$ Écart type ?

Pourquoi sur la variance ?

Intérêt : dispersion des données
$\Rightarrow$ Écart type ? Non, pas de test sur l'écart type !
Mais des tests sur la variance existent

Généralement la moyenne de la population n'est pas connue
$\Rightarrow$ Estimer la moyenne grâce à l'échantillon.

Statistique de test

La statistique de test à utiliser suit une loi du χ^2 à $n-1$ degrés de libertés :

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

où

σ^2 = variance supposée de la population
s^2 = variance de l'échantillon
n = taille de l'échantillon

Hypothèse : échantillon aléatoire provenant d'une population dont les éléments sont iid de distribution normale :

Remarque

Le nombre de degrés de liberté n'est pas n mais $n-1$ car la véritable moyenne (de la population) n'est pas connue. Il faut donc l'estimer par la moyenne de l'échantillon.

Exemple

Le système de satellites Galileo est un projet européen prévu pour concurrencer le système GPS américain actuel. Supposons qu'un constructeur de GPS fournisse un appareil utilisant Galileo, et annonce une précision de ± 5 cm. Un test est effectué pour vérifier cette affirmation, en utilisant un échantillon de taille 20, et un niveau de signification de 0.05. La variance de l'échantillon est calculée $s^2 = 0.0108$. Les données de l'échantillon suivent une loi normale, et l'on peut supposer qu'il en est de même pour la population.

1. La valeur de la population d'intérêt est

Exemple

Le système de satellites Galileo est un projet européen prévu pour concurrencer le système GPS américain actuel. Supposons qu'un constructeur de GPS fournisse un appareil utilisant Galileo, et annonce une précision de ± 5 cm. Un test est effectué pour vérifier cette affirmation, en utilisant un échantillon de taille 20, et un niveau de signification de 0.05. La variance de l'échantillon est calculée $s^2 = 0.0108$. Les données de l'échantillon suivent une loi normale, et l'on peut supposer qu'il en est de même pour la population.

1. La valeur de la population d'intérêt est
   la précision de l'appareil, qui est une variance.
2. Les hypothèse du test sont

Exemple

Le système de satellites Galileo est un projet européen prévu pour concurrencer le système GPS américain actuel. Supposons qu'un constructeur de GPS fournisse un appareil utilisant Galileo, et annonce une précision de ± 5 cm. Un test est effectué pour vérifier cette affirmation, en utilisant un échantillon de taille 20, et un niveau de signification de 0.05. La variance de l'échantillon est calculée $s^2 = 0.0108$. Les données de l'échantillon suivent une loi normale, et l'on peut supposer qu'il en est de même pour la population.

1. La valeur de la population d'intérêt est
   la précision de l'appareil, qui est une variance.

2. Les hypothèse du test sont
   $H_0 : \sigma^2 \leq 0.0025$
   $H_1 : \sigma^2 > 0.0025$

Exemple (suite)

3. Le niveau de signification est

Exemple (suite)

3. Le niveau de signification est fixé à $\alpha = 0.05$
4. La région de rejet du test est

Exemple (suite)

3. Le niveau de signification est fixé à $\alpha = 0.05$

4. La région de rejet du test est l'ensemble des valeurs supérieures à

$$\chi^2_{0.05} \approx 30.1435$$

   La valeur limite calculée est celle pour une distribution χ^2 à $20 - 1 = 19$ degrés de liberté, et un niveau de signification de 0.05.

5. La statistique associée à l'échantillon est

Exemple (suite)

3. Le niveau de signification est fixé à $\alpha = 0.05$

4. La région de rejet du test est l'ensemble des valeurs supérieures à
$$\chi^2_{0.05} \approx 30.1435$$
La valeur limite calculée est celle pour une distribution χ^2 à $20 - 1 = 19$ degrés de liberté, et un niveau de signification de 0.05.

5. La statistique associée à l'échantillon est
$$s^2 = 0.0108 \quad \chi^2 = \frac{(20-1)0.0108}{0.0025} = 82.08$$

Exemple (fin)

6. Comme $\chi^2 = 82.08$ appartient à la région critique, l'hypothèse H_0 est rejetée.

7. Conclusion

Exemple (fin)

6. Comme $\chi^2 = 82.08$ appartient à la région critique, l'hypothèse H_0 est rejetée.
7. Conclusion
   Il est donc très peu probable que la précision annoncée soit correcte.