

Statistiques

Chapitre 3 – Tests statistiques (un seul échantillon)

- 3.1 Principe des tests d'hypothèses
- 3.2 Comparaison d'une moyenne à une valeur donnée
- 3.3 Comparaison d'une variance à une valeur donnée
- 3.4 Comparaison d'une proportion à une valeur donnée
- 3.5 Seuil descriptif du test
- 3.6 Risques et courbe d'efficacité
- 3.7 Test d'ajustement – Test du Khi-Deux

Marie-Christine SUHNER

marie-christine.suhner@univ-lorraine.fr

3.1 Principe des tests d'hypothèses

❖ Principe général

- ♦ Soit une population dont les éléments présentent un caractère mesurable (ou dénombrable) et dont la valeur du paramètre, relative au caractère, est inconnue
- ♦ Une hypothèse est formulée sur la valeur du paramètre ; cette formulation repose sur des considérations théoriques, pratiques ou sur un pressentiment
- ♦ On veut porter un jugement sur cette hypothèse, sur la base des résultats d'un échantillon prélevé de cette population
- ♦ La valeur estimée ne sera pas rigoureusement identique à la valeur théorique, elle comporte des fluctuations d'échantillonnage
- ♦ Il s'agit de savoir si l'écart observé entre la valeur observée sur l'échantillon et la valeur de l'hypothèse est dû au hasard de l'échantillonnage ou non

❖ Concepts importants

– Hypothèse statistique

Affirmation concernant les caractéristiques (valeurs des paramètres, forme de la distribution...) d'une population

– Test d'hypothèses

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

– Hypothèse nulle (H0) et hypothèse alternative (H1)

- ♦ H0 : hypothèse selon laquelle on fixe *a priori* un paramètre de la population à une valeur particulière
- ♦ H1 : hypothèse qui diffère de H0
 - Il faut convenir à l'avance à quelle condition l'une ou l'autre des hypothèses sera considérée comme vraisemblable
 - C'est H0 qui est soumise au test et toute la démarche s'effectue en considérant que H0 est vraie
 - Si on rejette H0, on considère que H1 est vraisemblable

3.1 Principe des tests d'hypothèses

❖ Concepts importants (suite)

– Aide à la décision

La décision de favoriser H_0 ou H_1 est basée sur une information partielle, les résultats d'un échantillon

Il est statistiquement impossible de prendre toujours la bonne décision et, à chaque décision, on prend un risque de se tromper

– Seuil de signification du test

Risque noté α , consenti à l'avance, de rejeter à tort H_0 alors qu'elle est vraie

$$\alpha = \Pr(\text{rejeter } H_0 / H_0 \text{ est vraie})$$

– Région de rejet de H_0 (ou région critique)

Au seuil de signification, on fait correspondre sur la distribution d'échantillonnage, une région de rejet de H_0

L'aire de cette région correspond à la probabilité α

– Région de non rejet de H_0

Région complémentaire de probabilité $(1 - \alpha)$

– Prise de décision

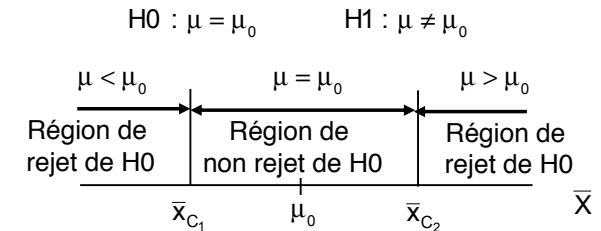
La valeur observée déduite des résultats de l'échantillon appartient, soit à la région de rejet de H_0 (on favorise H_1), soit à la région de non rejet de H_0 (on favorise H_0)

❖ Remarques

- Les seuils de signification les plus utilisés sont $\alpha = 0,05$ et $\alpha = 0,01$
- La statistique qui convient pour le test est une variable aléatoire dont la valeur observée est utilisée pour décider du « Rejet » ou du « Non rejet » de H_0
- La distribution d'échantillonnage de cette statistique est déterminée en supposant que H_0 est vraie

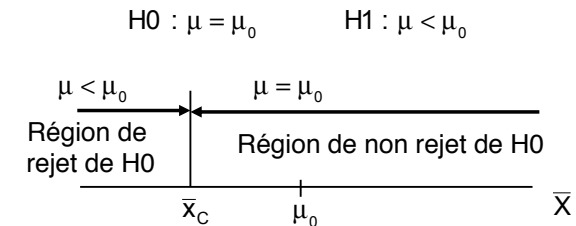
3.1 Type de tests d'hypothèses

❖ Test bilatéral

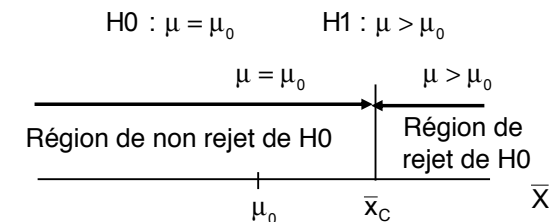


- Si, suite aux résultats de l'échantillon, $\bar{x}_{c1} \leq \bar{X} \leq \bar{x}_{c2}$ on ne peut pas rejeter H_0 au seuil de signification choisi
- Sinon, on rejette H_0 et on favorise H_1
- Les valeurs \bar{x}_{c1} et \bar{x}_{c2} sont appelées valeurs critiques
- Les valeurs critiques peuvent s'exprimer également en termes d'écart réduit

❖ Test unilatéral à gauche



❖ Test unilatéral à droite



❖ Choix du type de test

On choisit l'hypothèse H_1 la plus pertinente à la situation pratique analysée

3.1 Règle de décision Test sur une moyenne

❖ Contexte

$$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$$

- ♦ On considère une distribution normale de variance connue σ^2
- ♦ On prélève au hasard de cette population un échantillon de taille n
- ♦ En supposant H_0 vraie,

$$\bar{X} \rightarrow N\left(\mu_0; \frac{\sigma}{\sqrt{n}}\right) \quad \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \rightarrow \text{LNCR}$$

- ♦ On choisit la valeur de α

❖ Détermination des valeurs critiques

$$\Pr(\text{Non rejet de } H_0 / H_0 \text{ est vraie}) = 1 - \alpha$$

$$\Pr(\bar{X}_{C_1} \leq \bar{X} \leq \bar{X}_{C_2} / H_0 : \mu = \mu_0 \text{ est vraie}) = 1 - \alpha$$

$$\Pr\left(\frac{\bar{X}_{C_1} - \mu_0}{\sigma / \sqrt{n}} \leq \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \leq \frac{\bar{X}_{C_2} - \mu_0}{\sigma / \sqrt{n}}\right) = 1 - \alpha$$

On en déduit les valeurs critiques de \bar{X}

$$\bar{X}_{C_1} = \mu_0 - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \bar{X}_{C_2} = \mu_0 + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

❖ Règles de décision

- ♦ On ne rejette pas H_0 si

$$\bar{X}_{C_1} \leq \bar{X} \leq \bar{X}_{C_2} \quad \text{ou} \quad -u_{1-\alpha/2} \leq U = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \leq u_{1-\alpha/2}$$
- ♦ Si on rejette H_0 , la différence $\bar{X} - \mu_0$ est statistiquement significative et n'est pas compatible avec H_0 : les résultats de l'échantillon sont en contradiction avec H_0
- ♦ Si on ne rejette pas H_0 , l'écart n'est pas significatif, il est imputable aux fluctuations d'échantillonnage

3.1 Tests d'hypothèses Démarche générale

❖ Notion d'écart réduit

$$\text{Ecart réduit} = \frac{\text{Ecart entre la statistique et la valeur en } H_0}{\text{Ecart type de la statistique}}$$

❖ Les étapes de la démarche

1. Formuler les **hypothèses** H_0 et H_1
2. Fixer à l'avance le **seuil de signification** α
3. Préciser les **conditions d'application** du test :
 - forme de la distribution
 - taille d'échantillon
 - paramètre de la population connu ou estimé
4. Spécifier la **variable d'échantillonnage** qui convient
En déduire sa **distribution** d'après les conditions d'application, lorsque H_0 est vraie
5. Définir une **région critique** (ou zone de rejet) en fonction de α
6. Calculer la **valeur numérique** de la variable d'échantillonnage (ou de l'écart réduit), à partir des résultats de l'échantillon
7. Enoncer la règle de décision correspondante :
 - **rejeter H_0** si le résultat expérimental est dans la région critique (ou zone de rejet)
 - **ne pas rejeter H_0** si le résultat expérimental est dans la région de non rejet

3.2 Comparaison d'une moyenne à une valeur donnée *Variance population connue*

1. Hypothèses statistiques

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

$$H_1 : \mu > \mu_0$$

$$H_1 : \mu < \mu_0$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

- ♦ Population normale de variance connue
- ♦ Population quelconque de variance connue avec un grand échantillon ($n \geq 30$)

4. Variable d'échantillonnage et sa distribution

- ♦ Statistique \bar{X}
- ♦ En supposant H_0 vraie et selon les conditions d'application

$$U = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \rightarrow \text{LNCR}$$

5. Région critique et règle de décision

$$\begin{array}{llll} H_0 : \mu = \mu_0 & H_1 : \mu \neq \mu_0 & \text{Rejeter } H_0 \text{ si } U > u_{1-\alpha/2} & \text{ou } U < -u_{1-\alpha/2} \\ & H_1 : \mu > \mu_0 & \text{Rejeter } H_0 \text{ si } U > u_{1-\alpha} & \\ & H_1 : \mu < \mu_0 & \text{Rejeter } H_0 \text{ si } U < -u_{1-\alpha} & \end{array}$$

6. Calcul de l'écart réduit u

$$u \text{ est calculé à partir de } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

7. Décision et conclusion

- ♦ Si la valeur calculée de l'écart réduit se situe dans la zone de rejet, on rejette H_0
- ♦ Si la valeur calculée de l'écart réduit se situe dans la zone de non rejet, on ne rejette pas H_0

3.2 Comparaison d'une moyenne à une valeur donnée *Variance population inconnue*

1. Hypothèses statistiques

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

$$H_1 : \mu > \mu_0$$

$$H_1 : \mu < \mu_0$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

- ♦ Population normale de variance inconnue mais estimée avec un petit échantillon ($n < 30$)
- ♦ Population quelconque de variance inconnue mais estimée avec un grand échantillon ($n \geq 30$)

4. Variable d'échantillonnage et sa distribution

- ♦ Statistique \bar{X}
- ♦ En supposant H_0 vraie et selon les conditions d'application

$$T = \frac{\bar{X} - \mu_0}{s^* / \sqrt{n}} \rightarrow \text{Loi de Student avec } v = n - 1 \text{ DL}$$

5. Région critique et règle de décision

$$\begin{array}{llll} H_0 : \mu = \mu_0 & H_1 : \mu \neq \mu_0 & \text{Rejeter } H_0 \text{ si } T > t_{1-\alpha/2; n-1} & \text{ou } T < -t_{1-\alpha/2; n-1} \\ & H_1 : \mu > \mu_0 & \text{Rejeter } H_0 \text{ si } T > t_{1-\alpha; n-1} & \\ & H_1 : \mu < \mu_0 & \text{Rejeter } H_0 \text{ si } T < -t_{1-\alpha; n-1} & \end{array}$$

6. Calcul de l'écart réduit t

$$t \text{ est calculé à partir de } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ et de } s^{*2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

7. Décision et conclusion

3.3 Comparaison d'une variance à une valeur donnée *Moyenne population connue*

1. Hypothèses statistiques

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

$$H_1 : \sigma^2 > \sigma_0^2$$

$$H_1 : \sigma^2 < \sigma_0^2$$

2. Seuil de signification

α

3. Conditions d'application

Population normale

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$\chi^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \rightarrow \text{Loi du Khi-Deux avec } v = n \text{ DL}$$

5. Région critique et règle de décision

$H_0 : \sigma^2 = \sigma_0^2$	$H_1 : \sigma^2 \neq \sigma_0^2$	Rejeter H_0 si $\chi^2 > \chi_{1-\alpha/2;n}^2$ ou $\chi^2 < \chi_{\alpha/2;n}^2$
	$H_1 : \sigma^2 > \sigma_0^2$	Rejeter H_0 si $\chi^2 > \chi_{1-\alpha;n}^2$
	$H_1 : \sigma^2 < \sigma_0^2$	Rejeter H_0 si $\chi^2 < \chi_{\alpha;n}^2$

6. Calcul de la valeur de χ^2

$$\chi^2 \text{ est calculée à partir de } \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

7. Décision et conclusion

3.3 Comparaison d'une variance à une valeur donnée *Moyenne population inconnue*

1. Hypothèses statistiques

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

$$H_1 : \sigma^2 > \sigma_0^2$$

$$H_1 : \sigma^2 < \sigma_0^2$$

2. Seuil de signification

α

3. Conditions d'application

Population normale

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$\chi^2 = \frac{(n-1) S^{*2}}{\sigma_0^2} \rightarrow \text{Loi du Khi-Deux avec } v = n-1 \text{ DL}$$

5. Région critique et règle de décision

$H_0 : \sigma^2 = \sigma_0^2$	$H_1 : \sigma^2 \neq \sigma_0^2$	Rejeter H_0 si $\chi^2 > \chi_{1-\alpha/2;n-1}^2$ ou $\chi^2 < \chi_{\alpha/2;n-1}^2$
	$H_1 : \sigma^2 > \sigma_0^2$	Rejeter H_0 si $\chi^2 > \chi_{1-\alpha;n-1}^2$
	$H_1 : \sigma^2 < \sigma_0^2$	Rejeter H_0 si $\chi^2 < \chi_{\alpha;n-1}^2$

6. Calcul de la valeur de χ^2

$$\chi^2 \text{ est calculée à partir de } s^{*2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

7. Décision et conclusion

3.4 Comparaison d'une proportion à une valeur donnée

Approximation par la loi normale

1. Hypothèses statistiques

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

$$H_1 : p > p_0$$

$$H_1 : p < p_0$$

2. Seuil de signification

α

3. Conditions d'application

Approximation de la loi binomiale par la loi normale

$$n \cdot p_0 \geq 5 \quad \text{et} \quad n \cdot (1 - p_0) \geq 5$$

4. Variable d'échantillonnage et sa distribution

♦ Statistique \hat{P}

♦ En supposant H_0 vraie et selon les conditions d'application

$$U = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow \text{LNCR}$$

5. Région critique et règle de décision

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

Rejeter H_0 si $U > u_{1-\alpha/2}$ ou $U < -u_{1-\alpha/2}$

$$H_1 : p > p_0$$

Rejeter H_0 si $U > u_{1-\alpha}$

$$H_1 : p < p_0$$

Rejeter H_0 si $U < -u_{1-\alpha}$

6. Calcul de l'écart réduit u

u est calculé à partir de f

7. Décision et conclusion

3.5 Seuil descriptif du test (ou p-value)

❖ Principe

♦ Etant donnés les divers éléments d'un test statistique et la valeur calculée de l'écart réduit (ou de la variable d'échantillonnage), le seuil descriptif du test (noté α_p) est la plus petite valeur du seuil de signification α qui conduirait au rejet de H_0

♦ La décision de rejeter ou pas H_0 s'effectue alors en comparant la valeur α_p au seuil de signification α

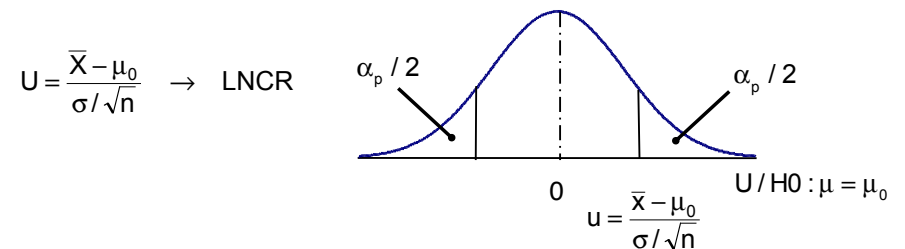
– Si $\alpha_p < \alpha$, on rejette H_0

– Si $\alpha_p \geq \alpha$, on ne peut pas rejeter H_0

❖ Exemple : test sur la moyenne

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$



$$\alpha_p = 2 \Pr \left(U = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} > u = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right)$$

Si $\alpha_p < \alpha$, on rejette H_0

Si $\alpha_p \geq \alpha$, on ne peut pas rejeter H_0

3.6 Risques de première et de deuxième espèce

❖ Types d'erreurs possibles

– Erreur de première espèce

On rejette H_0 alors que H_0 est vraie

– Erreur de deuxième espèce

On ne rejette pas H_0 alors que H_1 est vraie

❖ Risques d'erreur

$\alpha = \Pr(\text{rejeter } H_0 / H_0 \text{ est vraie})$

= Probabilité de commettre une erreur de première espèce

$\beta = \Pr(\text{ne pas rejeter } H_0 / H_1 \text{ est vraie})$

= Probabilité de commettre une erreur de deuxième espèce

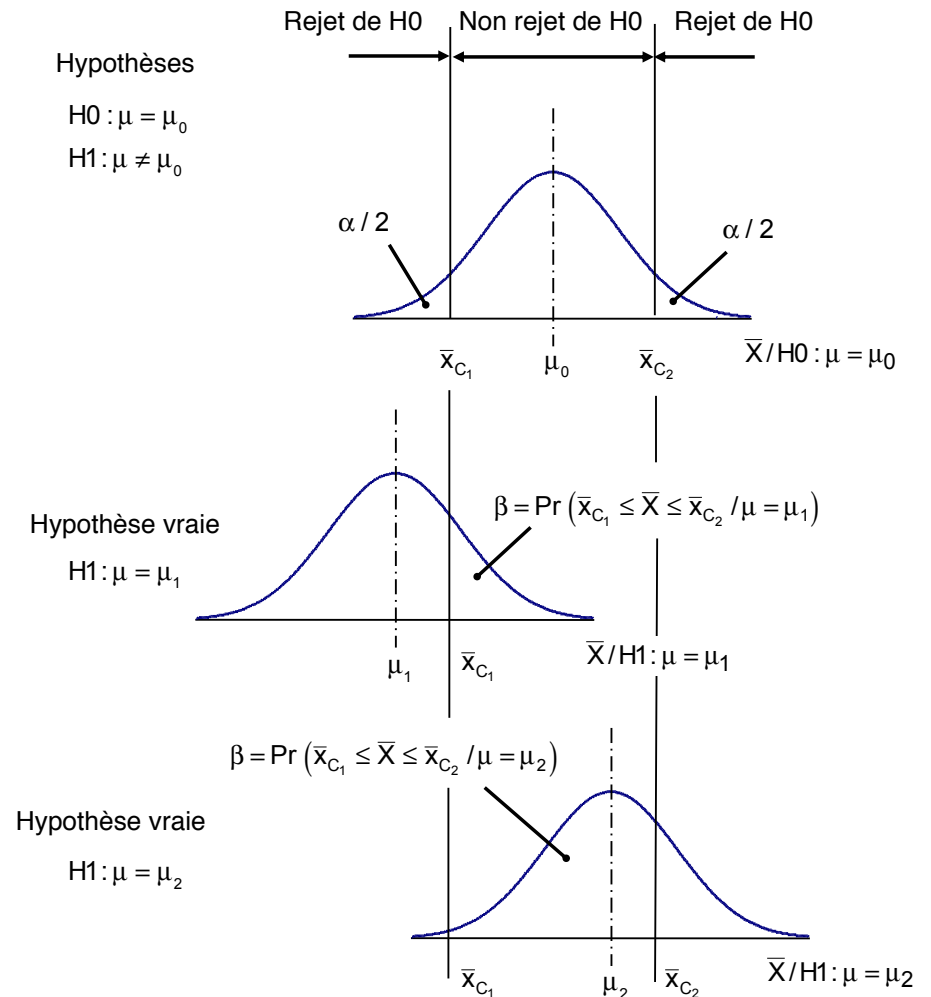
- ♦ Le seuil de signification α (risque de première espèce) est choisi *a priori*
- ♦ Le risque de deuxième espèce β dépend de l'hypothèse alternative H_1 : on ne peut le calculer que si on spécifie des valeurs particulières du paramètre dans H_1 , que l'on suppose vraie

SITUATION REELLE	DECISION	
	H_0 n'est pas rejetée	H_0 est rejetée
H_0 est vraie	Décision correcte Seuil de confiance : $1 - \alpha$	Décision incorrecte Erreur de type I : α
H_1 est vraie	Décision incorrecte Erreur de type II : β	Décision correcte Puissance du test : $1 - \beta$

❖ Puissance du test

Probabilité $(1 - \beta)$ de rejeter H_0 alors que c'est une certaine hypothèse alternative H_1 qui est vraie

3.6 Risques d'erreur sur la distribution d'échantillonnage de la moyenne



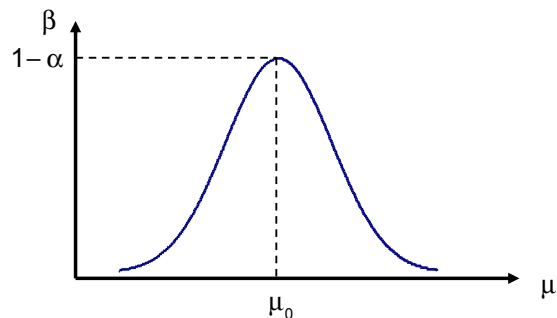
3.6 Courbe d'efficacité

❖ Propriétés des risques α et β

- ♦ Pour un même risque α et une même taille d'échantillon, si l'écart entre la valeur du paramètre en H_0 et celle supposée vraie en H_1 augmente, le risque β diminue
- ♦ Une réduction du risque α élargit la zone de non rejet de H_0 mais augmente le risque β
- ♦ Pour α fixe et σ déterminé, l'augmentation de n donne une meilleure précision : le test est alors plus puissant

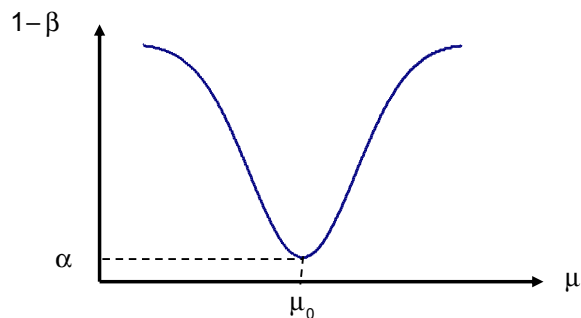
❖ Courbe d'efficacité du test

Graphique de β en fonction des diverses valeurs de μ posées en H_1



❖ Courbe de puissance du test

Graphique de $(1 - \beta)$ en fonction des diverses valeurs de μ posées en H_1



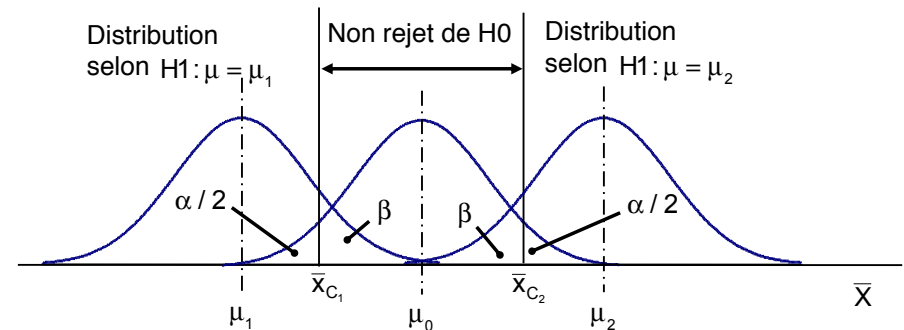
3.6 Test sur une moyenne avec contrôle de α et β

❖ Objectif

- ♦ On veut définir un plan d'échantillonnage qui limite à des valeurs particulières les risques α et β selon les hypothèses posées en H_0 et H_1 : on impose la courbe d'efficacité du test
- ♦ Plan d'échantillonnage :
 - définir la taille de l'échantillon requise
 - établir la règle de décision

❖ Principe

- ♦ Un procédé doit opérer à une valeur centrale μ_0
- ♦ Plan d'échantillonnage requis :
 - Si le procédé opère à μ_0 , on veut un plan d'échantillonnage ayant un risque α de rejeter à tort $H_0 : \mu = \mu_0$ vraie
 - Si le procédé opère à valeur centrale $\mu_1 (\mu_1 < \mu_0)$ ou $\mu_2 (\mu_2 > \mu_0)$, on veut que le risque encouru dans le non rejet de $H_0 : \mu = \mu_0$ soit limité à la valeur β



$$\begin{aligned} \mu_2 - \mu_0 &= (\bar{x}_{c_2} - \mu_0) + (\mu_2 - \bar{x}_{c_2}) \\ \frac{\mu_2 - \mu_0}{\sigma / \sqrt{n}} &= \frac{\bar{x}_{c_2} - \mu_0}{\sigma / \sqrt{n}} + \frac{\mu_2 - \bar{x}_{c_2}}{\sigma / \sqrt{n}} = u_{1-\alpha/2} + |u_\beta| \\ \text{d'où } n &= \frac{\sigma^2 (u_{1-\alpha/2} + |u_\beta|)^2}{(\mu_2 - \mu_0)^2} \end{aligned}$$

On peut alors déterminer les valeurs critiques et la règle de décision 49

3.7 Test d'ajustement

Test du Khi-Deux

❖ Objectif

- ♦ On veut examiner si la distribution des données observées semble s'apparenter à une distribution théorique connue
- ♦ Test du Khi-Deux (ou de Pearson) : vérification de la qualité de l'ajustement entre une distribution théorique et une distribution expérimentale

❖ Principe général du test

On cherche à déterminer si un modèle théorique est susceptible de représenter le comportement probabiliste de la variable observée, en se basant sur les fréquences absolues (effectifs) des résultats obtenus sur l'échantillon

❖ Répartition des données

- ♦ On suppose que les données sont réparties suivant q classes (ou q valeurs) ayant respectivement pour effectifs :

$$n_{o1}, n_{o2}, \dots, n_{oq}$$

$$\sum_{i=1}^q n_{oi} = n$$

- ♦ En admettant comme plausible une distribution théorique particulière, on peut construire une répartition idéale des observations de l'échantillon de taille n
- ♦ On obtient alors les effectifs théoriques :

$$n_{t1} = n \cdot p_1, n_{t2} = n \cdot p_2, \dots, n_{ti} = n \cdot p_i, \dots, n_{tq} = n \cdot p_q$$

où p_i représente la probabilité que le caractère observé prenne une valeur appartenant à la classe i

$$\sum_{i=1}^q n_{ti} = n$$

3.7 Test d'ajustement

Test du Khi-Deux

❖ Mesure de l'ampleur de l'écart

$$\chi^2 = \sum_{i=1}^q \frac{(n_{oi} - n_{ti})^2}{n_{ti}} \rightarrow \text{Loi du Khi-Deux à } v \text{ DL}$$

❖ Détermination du nombre de degrés de liberté

- ♦ Si la distribution théorique est entièrement spécifiée c'est-à-dire aucun paramètre n'est à estimer à l'aide des données de l'échantillon

$$v = q - 1$$

On perd un degré de liberté car $\sum_{i=1}^q n_{oi} = \sum_{i=1}^q n_{ti} = n$

- ♦ S'il faut estimer r paramètres à partir des données de l'échantillon pour caractériser la distribution théorique

$$v = q - 1 - r$$

❖ Conditions d'application

- ♦ En pratique, il faut un nombre suffisant d'observations :
effectifs théoriques de chaque classe supérieurs à 5
- ♦ Parfois, il faut regrouper deux ou plusieurs classes adjacentes pour obtenir des effectifs suffisants par classe
- ♦ q représente le nombre de classes après regroupement dans le calcul des DL

❖ Considérations théoriques

- ♦ Le nombre de données de l'échantillon de taille n susceptible d'appartenir à la classe i est une variable aléatoire binomiale
 - de moyenne $n \cdot p_i$
 - et de variance $n \cdot p_i (1 - p_i) \approx n \cdot p_i$
- ♦ Si $n \cdot p_i \geq 5$ alors la quantité $\frac{n_{oi} - n \cdot p_i}{\sqrt{n \cdot p_i}} \rightarrow \text{LNCR}$
- ♦ La somme du carré de q variables normales centrées réduites suit une loi du Khi-Deux

$$\sum_{i=1}^q \frac{(n_{oi} - n \cdot p_i)^2}{n \cdot p_i} \rightarrow \text{Loi du Khi-Deux}$$

3.7 Test d'ajustement

Test du Khi-Deux

1. Hypothèses statistiques

H0 : les données suivent la distribution théorique spécifiée

H1 : les données ne suivent pas la distribution théorique spécifiée

Calcul des effectifs théoriques

$$n_{t1} = n \cdot p_1, n_{t2} = n \cdot p_2, \dots, n_{ti} = n \cdot p_i, \dots, n_{tq} = n \cdot p_q$$

où p_i représente la probabilité que le caractère observé prenne une valeur appartenant à la classe i

$$\sum_{i=1}^q n_{ti} = n$$

2. Seuil de signification

α

3. Conditions d'application

Effectifs théoriques tous ≥ 5

4. Variable d'échantillonnage et sa distribution

En supposant H0 vraie et selon les conditions d'application

$$\chi^2 = \sum_{i=1}^q \frac{(n_{oi} - n_{ti})^2}{n_{ti}} \rightarrow \text{Loi du Khi-Deux à } v \text{ DL}$$

avec $v = q - 1 - r$

5. Région critique et règle de décision

$$\text{On rejette H0 si } \chi^2 = \sum_{i=1}^q \frac{(n_{oi} - n_{ti})^2}{n_{ti}} > \chi^2_{1-\alpha; v}$$

avec $v = q - 1 - r$

6. Calcul de χ^2

χ^2 est calculé à partir des effectifs observés et des effectifs théoriques

7. Décision et conclusion

Polytech Nancy
3^e année

2020 / 2021

Statistiques

Chapitre 4 – Tests statistiques (plusieurs échantillons)

- 4.1 Principe des tests
- 4.2 Comparaison de deux variances
- 4.3 Comparaison de deux moyennes
- 4.4 Autres tests sur les moyennes
- 4.5 Comparaison de deux proportions
- 4.6 Test d'indépendance – Test du Khi-Deux
- 4.7 Tests d'homogénéité de plusieurs populations – Test du Khi-Deux

Marie-Christine SUHNER

marie-christine.suhner@univ-lorraine.fr

4.1 Tests de comparaison

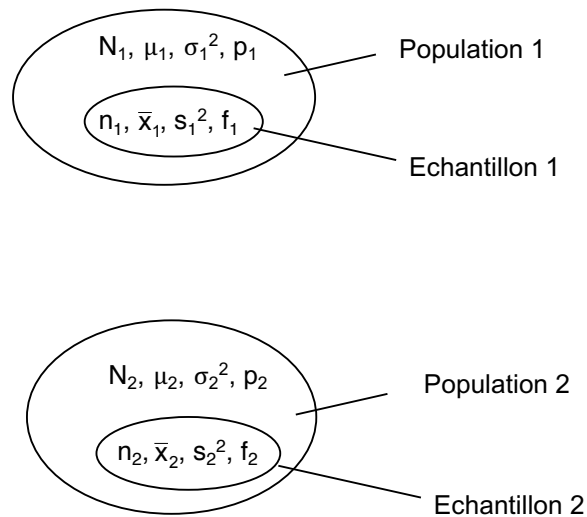
❖ Comparaison de deux populations

- moyenne
- variance
- proportion

❖ Etude des fluctuations d'échantillonnage

- différence de deux moyennes
- rapport de deux variances
- différence de deux proportions

❖ Notations



4.2 Comparaison de deux variances

Moyennes populations connues

1. Hypothèses statistiques

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$H_1: \sigma_1^2 < \sigma_2^2$$

2. Seuil de signification

α

3. Conditions d'application

Populations normales

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$F = \frac{\sum_{i=1}^{n_1} (X_{i1} - \mu_1)^2}{n_1} \rightarrow \text{Loi de Fisher-Snedecor avec } v_1 = n_1 \text{ et } v_2 = n_2 \text{ DL}$$

$$\frac{\sum_{j=1}^{n_2} (X_{j2} - \mu_2)^2}{n_2}$$

5. Région critique et règle de décision

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\text{Rejeter } H_0 \text{ si } F > F_{1-\alpha/2; n_1; n_2} \text{ ou } F < F_{\alpha/2; n_1; n_2}$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$\text{Rejeter } H_0 \text{ si } F > F_{1-\alpha; n_1; n_2}$$

$$H_1: \sigma_1^2 < \sigma_2^2$$

$$\text{Rejeter } H_0 \text{ si } F < F_{\alpha; n_1; n_2}$$

6. Calcul de la valeur de F

F est calculé à partir de

$$\frac{\sum_{i=1}^{n_1} (x_{i1} - \mu_1)^2}{n_1} \text{ et } \frac{\sum_{j=1}^{n_2} (x_{j2} - \mu_2)^2}{n_2}$$

7. Décision et conclusion

4.2 Comparaison de deux variances

Moyennes populations inconnues

1. Hypothèses statistiques

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

$$H_1 : \sigma_1^2 < \sigma_2^2$$

2. Seuil de signification

α

3. Conditions d'application

Populations normales

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$F = \frac{S_1^{*2}}{S_2^{*2}} \rightarrow \text{Loi de Fisher-Snedecor avec } v_1 = n_1 - 1 \text{ et } v_2 = n_2 - 1 \text{ DL}$$

5. Région critique et règle de décision

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{Rejeter } H_0 \text{ si } F > F_{1-\alpha/2; n_1-1; n_2-1} \text{ ou } F < F_{\alpha/2; n_1-1; n_2-1}$$

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad \text{Rejeter } H_0 \text{ si } F > F_{1-\alpha; n_1-1; n_2-1}$$

$$H_1 : \sigma_1^2 < \sigma_2^2 \quad \text{Rejeter } H_0 \text{ si } F < F_{\alpha; n_1-1; n_2-1}$$

6. Calcul de la valeur de F

$$F \text{ est calculé à partir de } s_1^{*2} = \frac{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2}{n_1 - 1} \text{ et } s_2^{*2} = \frac{\sum_{j=1}^{n_2} (x_{j2} - \bar{x}_2)^2}{n_2 - 1}$$

7. Décision et conclusion

4.2 La loi de Fisher-Snedecor

❖ La loi de Fisher-Snedecor (ou Snedecor)

La quantité F est une variable aléatoire continue issue du rapport de deux variables du χ^2 divisées par leurs degrés de liberté respectifs

$$F = \frac{\frac{\chi_1^2}{v_1}}{\frac{\chi_2^2}{v_2}} \rightarrow \text{Loi de Fisher-Snedecor avec } v_1 \text{ et } v_2 \text{ DL}$$

$$E(F) = \frac{v_2}{v_2 - 2} \quad v_2 > 2 \quad \text{et} \quad V(F) = \frac{v_2^2(2v_1 + 2v_2 - 4)}{v_1(v_2 - 2)^2(v_2 - 4)} \quad v_2 > 4$$

❖ Propriétés

- ♦ $F \geq 0$
- ♦ Elle dépend des nombres de degrés de liberté v_1 et v_2
- ♦ La loi présente une asymétrie positive
- ♦ A mesure que v_1 et v_2 augmentent, elle tend vers une loi normale

❖ Remarques

$$F_{1-\alpha; v_1, v_2} \neq F_{1-\alpha; v_2, v_1}$$

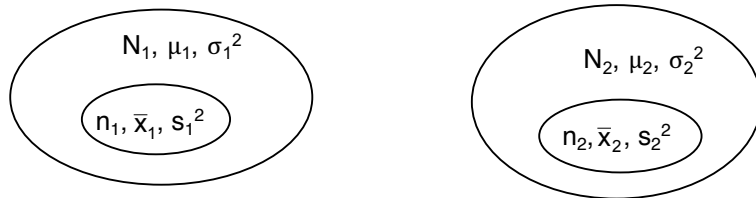
$$F_{\alpha; v_1, v_2} = \frac{1}{F_{1-\alpha; v_2, v_1}} \quad \text{ou} \quad F_{1-\alpha; v_1, v_2} = \frac{1}{F_{\alpha; v_2, v_1}}$$

La loi F est également très utilisée en régression et analyse de variance

4.3 Comparaison de deux moyennes

❖ Besoin de connaître

- la forme des distributions
- la taille des échantillons
- les variances



❖ Tests les plus courants

- ♦ Variances σ_1^2 et σ_2^2 connues
- ♦ Grands échantillons $n_1 \geq 30$, $n_2 \geq 30$ et variances inconnues
- ♦ Populations normales, variances inconnues mais supposées égales $\sigma_1^2 = \sigma_2^2 = \sigma^2$ et un des échantillons (ou les deux) petit(s)

❖ Autres tests

- ♦ Populations normales, variances inconnues mais supposées inégales $\sigma_1^2 \neq \sigma_2^2$ et un des échantillons (ou les deux) petit(s)
- ♦ Données appariées

4.3 Comparaison de deux moyennes

Variances populations connues

1. Hypothèses statistiques

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

- ♦ Populations normales et variances connues
- ♦ Populations quelconques, variances connues et échantillons de grandes tailles ($n_1 \geq 30$, $n_2 \geq 30$)

4. Variable d'échantillonnage et sa distribution

$$\text{Statistique } \bar{X}_1 - \bar{X}_2$$

- ♦ En supposant H_0 vraie et selon les conditions d'application

$$\bar{X}_1 - \bar{X}_2 \rightarrow \text{Loi normale avec } E(\bar{X}_1 - \bar{X}_2) = 0 \text{ et } V(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow \text{LNCR}$$

5. Région critique et règle de décision

$$\begin{array}{llll} H_0 : \mu_1 = \mu_2 & H_1 : \mu_1 \neq \mu_2 & \text{Rejeter } H_0 \text{ si } & U > u_{1-\alpha/2} \quad \text{ou} \quad U < -u_{1-\alpha/2} \\ & H_1 : \mu_1 > \mu_2 & \text{Rejeter } H_0 \text{ si } & U > u_{1-\alpha} \\ & H_1 : \mu_1 < \mu_2 & \text{Rejeter } H_0 \text{ si } & U < -u_{1-\alpha} \end{array}$$

6. Calcul de l'écart réduit u

$$u \text{ est calculé à partir de } \bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} \text{ et } \bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{j2}$$

7. Décision et conclusion

4.3 Comparaison de deux moyennes

*Grands échantillons $n_1 \geq 30$, $n_2 \geq 30$ et
variances populations inconnues*

1. Hypothèses statistiques

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

Grands échantillons et variances σ_1^2 et σ_2^2 estimées

4. Variable d'échantillonnage et sa distribution

♦ Statistique $\bar{X}_1 - \bar{X}_2$

♦ En supposant H_0 vraie et selon les conditions d'application

$$\bar{X}_1 - \bar{X}_2 \rightarrow \text{Loi normale avec } E(\bar{X}_1 - \bar{X}_2) = 0 \text{ et } V(\bar{X}_1 - \bar{X}_2) = \frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}$$

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}} \rightarrow \text{LNCR}$$

5. Région critique et règle de décision

$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	Rejeter H_0 si $U > u_{1-\alpha/2}$ ou $U < -u_{1-\alpha/2}$
	$H_1 : \mu_1 > \mu_2$	Rejeter H_0 si $U > u_{1-\alpha}$
	$H_1 : \mu_1 < \mu_2$	Rejeter H_0 si $U < -u_{1-\alpha}$

6. Calcul de l'écart réduit u

u est calculé à partir de

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} \quad \bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{j2}$$

$$s_1^{*2} = \frac{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2}{n_1 - 1} \quad s_2^{*2} = \frac{\sum_{j=1}^{n_2} (x_{j2} - \bar{x}_2)^2}{n_2 - 1}$$

7. Décision et conclusion

4.3 Comparaison de deux moyennes

*Populations normales, un ou les deux échantillons petits (< 30) et
variances populations inconnues mais supposées égales*

1. Hypothèses statistiques

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

Populations normales, un ou les deux échantillons petits (< 30)
et variances inconnues mais supposées égales $\sigma_1^2 = \sigma_2^2 = \sigma^2$

4. Variable d'échantillonnage et sa distribution

♦ Statistique $\bar{X}_1 - \bar{X}_2$

♦ En supposant H_0 vraie et selon les conditions d'application

$\bar{X}_1 - \bar{X}_2 \rightarrow$ Loi de Student à $v = n_1 + n_2 - 2$ DL avec

$$E(\bar{X}_1 - \bar{X}_2) = 0 \text{ et } V(\bar{X}_1 - \bar{X}_2) = S^{*2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \text{ et}$$

$$S^{*2} = \frac{(n_1 - 1)S_1^{*2} + (n_2 - 1)S_2^{*2}}{n_1 + n_2 - 2}$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow \text{Loi de Student à } v = n_1 + n_2 - 2 \text{ DL}$$

5. Région critique et règle de décision

$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	Rejeter H_0 si $T > t_{1-\alpha/2; n_1+n_2-2}$ ou $T < -t_{1-\alpha/2; n_1+n_2-2}$
	$H_1 : \mu_1 > \mu_2$	Rejeter H_0 si $T > t_{1-\alpha; n_1+n_2-2}$
	$H_1 : \mu_1 < \mu_2$	Rejeter H_0 si $T < -t_{1-\alpha; n_1+n_2-2}$

6. Calcul de l'écart réduit t

t est calculé à partir des moyennes et des variances estimées
sur les échantillons

7. Décision et conclusion

4.4 Autres tests sur les moyennes

Test de Smith-Satterthwaite (ou Aspin-Welch)

Populations normales, un ou les deux échantillons petits (< 30) et variances populations inconnues mais supposées inégales

1. Hypothèses statistiques

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

2. Seuil de signification

α

3. Conditions d'application

Population normales, un ou les deux échantillons petits (< 30) et variances inconnues mais supposées inégales

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$T' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}} \rightarrow \text{Loi de Student à } v = \frac{\left(\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}\right)^2}{\frac{(S_1^{*2}/n_1)^2}{n_1 - 1} + \frac{(S_2^{*2}/n_2)^2}{n_2 - 1}} \text{ DL}$$

v est arrondi au nombre entier inférieur

5. Région critique et règle de décision

$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	Rejeter H_0 si $T' > t_{1-\alpha/2;v}$ ou $T' < -t_{1-\alpha/2;v}$
	$H_1 : \mu_1 > \mu_2$	Rejeter H_0 si $T' > t_{1-\alpha;v}$
	$H_1 : \mu_1 < \mu_2$	Rejeter H_0 si $T' < -t_{1-\alpha;v}$

6. Calcul de l'écart réduit t'

t' est calculé à partir des moyennes et des variances estimées sur les échantillons

7. Décision et conclusion

62

4.4 Autres tests sur les moyennes

Test de comparaison dans le cas de deux données appariées

Echantillons obtenus à partir de la même unité expérimentale (même individu, périodes différentes, traitements différents...)

Pour chaque paire, on calcule $d_i = x_{i1} - x_{i2}$

1. Hypothèses statistiques

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

$$H_1 : \mu_d > 0$$

$$H_1 : \mu_d < 0$$

2. Seuil de signification

α

3. Conditions d'application

- Population normale et $n < 30$
- Population quelconque et $n \geq 30$

4. Variable d'échantillonnage et sa distribution

En supposant H_0 vraie et selon les conditions d'application

$$T = \frac{\bar{D}}{\frac{S_d^*}{\sqrt{n}}} \rightarrow \text{Loi de Student à } v = n - 1 \text{ DL}$$

$$\text{avec } S_d^* = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n - 1}}$$

5. Région critique et règle de décision

$H_0 : \mu_d = 0$	$H_1 : \mu_d \neq 0$	Rejeter H_0 si $T > t_{1-\alpha/2;n-1}$ ou $T < -t_{1-\alpha/2;n-1}$
	$H_1 : \mu_d > 0$	Rejeter H_0 si $T > t_{1-\alpha;n-1}$
	$H_1 : \mu_d < 0$	Rejeter H_0 si $T < -t_{1-\alpha;n-1}$

6. Calcul de l'écart réduit t

t est calculé à partir de la moyenne et de la variance estimée des écarts sur les échantillons

7. Décision et conclusion

63

4.5 Comparaison de deux proportions

Approximation par la loi normale

1. Hypothèses statistiques

$$H_0 : p_1 = p_2 = p$$

$$H_1 : p_1 \neq p_2$$

$$H_1 : p_1 > p_2$$

$$H_1 : p_1 < p_2$$

2. Seuil de signification

$$\alpha$$

3. Conditions d'application

Approximation des lois binomiales par des lois normales

$$n_1 \cdot f_1 \geq 5 \quad \text{et} \quad n_1 \cdot (1 - f_1) \geq 5 \quad n_2 \cdot f_2 \geq 5 \quad \text{et} \quad n_2 \cdot (1 - f_2) \geq 5$$

4. Variable d'échantillonnage et sa distribution

- Statistique $\hat{p}_1 - \hat{p}_2$ $\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$
- En supposant H_0 vraie et selon les conditions d'application

$$\hat{p}_1 - \hat{p}_2 \rightarrow \text{Loi normale avec} \quad E(\hat{p}_1 - \hat{p}_2) = 0 \quad \text{et}$$

$$V(\hat{p}_1 - \hat{p}_2) = \hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$U = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow \text{LNCR}$$

5. Région critique et règle de décision

$$\begin{array}{llll} H_0 : p_1 = p_2 & H_1 : p_1 \neq p_2 & \text{Rejeter } H_0 \text{ si} & U > u_{1-\alpha/2} \quad \text{ou} \quad U < -u_{1-\alpha/2} \\ & H_1 : p_1 > p_2 & \text{Rejeter } H_0 \text{ si} & U > u_{1-\alpha} \\ & H_1 : p_1 < p_2 & \text{Rejeter } H_0 \text{ si} & U < -u_{1-\alpha} \end{array}$$

6. Calcul de l'écart réduit u

u est calculé à partir de f_1 , f_2 et f

7. Décision et conclusion

4.6 Test d'indépendance

Test du Khi-Deux

❖ Etude de la liaison entre deux caractères

- Observation de deux caractères X et Y sur un échantillon de taille n
- Caractères X et Y qualitatifs ou quantitatifs
 - X présente q_A modalités
 - Y présente q_B modalités
- Tableau de contingence : répartition des n observations

X / Y	y_j	y_{q_B}	Marginale de X
x_i	n_{oij}		$n_{oi.}$
x_{q_A}			
Marginale de Y	$n_{o.j}$		$n = n_{o..}$

$$n = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} n_{oij}$$

❖ Hypothèses

H_0 : les deux caractères X et Y sont indépendants

H_1 : les deux caractères X et Y ne sont pas indépendants

❖ Calcul des effectifs théoriques

$$n_{tij} = n \cdot p_{i.} \cdot p_{.j} = n \cdot \frac{n_{oi.}}{n} \cdot \frac{n_{o.j}}{n} = \frac{n_{oi.} \cdot n_{o.j}}{n} \quad n = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} n_{tij} = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} n_{oij}$$

❖ Mesure de l'ampleur de l'écart

$$\chi^2 = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} \frac{(n_{oij} - n_{tij})^2}{n_{tij}} \rightarrow \text{Loi du Khi-Deux à } v = (q_A - 1) \cdot (q_B - 1) \text{ DL}$$

4.6 Test d'indépendance

Test du Khi-Deux

1. Hypothèses statistiques

H0 : les deux caractères X et Y sont indépendants

H1 : les deux caractères X et Y ne sont pas indépendants

Calcul des effectifs théoriques

$$n_{tij} = \frac{n_{oi.} \cdot n_{o.j}}{n} \quad n = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} n_{tij}$$

2. Seuil de signification

α

3. Conditions d'application

Effectifs théoriques tous ≥ 5

4. Variable d'échantillonnage et sa distribution

En supposant H0 vraie et selon les conditions d'application

$$\chi^2 = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} \frac{(n_{oij} - n_{tij})^2}{n_{tij}} \rightarrow \text{Loi du Khi-Deux à } v = (q_A - 1) \cdot (q_B - 1) \text{ DL}$$

5. Région critique et règle de décision

On rejette H0 si
$$\chi^2 = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} \frac{(n_{oij} - n_{tij})^2}{n_{tij}} > \chi^2_{1-\alpha; (q_A-1) \cdot (q_B-1)}$$

6. Calcul de χ^2

χ^2 est calculé à partir des effectifs observés et des effectifs théoriques

7. Décision et conclusion

4.7 Test d'homogénéité de plusieurs populations

Test du Khi-Deux

q échantillons de taille n_1, n_2, \dots, n_q de q populations

Sur chaque échantillon, mesure d'un caractère à k modalités

1. Hypothèses statistiques

H0 : $p_{i1} = p_{i2} = \dots = p_{iq} = p_{i.}$ avec $i = 1, 2, \dots, k$ modalités

H1 : les proportions ne sont pas identiques dans les q populations pour au moins une modalité

Calcul des effectifs théoriques

$$n_{tij} = n_j \cdot p_{i.} = n_j \cdot \frac{\sum_{j=1}^q n_{oij}}{n} = \frac{(\text{Nbre d'obs. dans l'éch. } j) \cdot (\text{Total des obs. pour la modalité } i)}{\text{Total des observations}}$$

2. Seuil de signification

α

3. Conditions d'application

Effectifs théoriques tous ≥ 5

4. Variable d'échantillonnage et sa distribution

En supposant H0 vraie et selon les conditions d'application

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^q \frac{(n_{oij} - n_{tij})^2}{n_{tij}} \rightarrow \text{Loi du Khi-Deux à } v = (k-1) \cdot (q-1) \text{ DL}$$

5. Région critique et règle de décision

On rejette H0 si
$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^q \frac{(n_{oij} - n_{tij})^2}{n_{tij}} > \chi^2_{1-\alpha; (k-1) \cdot (q-1)}$$

6. Calcul de χ^2

7. Décision et conclusion

4.7 Test d'égalité de q proportions

Test du Khi-Deux

Cas particulier du test d'homogénéité : caractère à 2 modalités

Proportion de succès de q populations binomiales identique

1. Hypothèses statistiques

$$H_0 : p_1 = p_2 = \dots = p_q = \bar{p}$$

H1 : les proportions ne sont pas toutes identiques dans les q populations

Calcul des effectifs théoriques

$$\bar{p} = \frac{\sum_{j=1}^q n_{o1j}}{n}$$
$$n_{t1j} = n_j \cdot \bar{p} \quad \text{et} \quad n_{t2j} = n_j \cdot (1 - \bar{p})$$

2. Seuil de signification α

3. Conditions d'application

Effectifs théoriques tous ≥ 5

4. Variable d'échantillonnage et sa distribution

En supposant H0 vraie et selon les conditions d'application

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^q \frac{(n_{oij} - n_{tij})^2}{n_{tij}} \rightarrow \text{Loi du Khi-Deux à } v = q - 1 \text{ DL}$$

5. Région critique et règle de décision

On rejette H0 si

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^q \frac{(n_{oij} - n_{tij})^2}{n_{tij}} > \chi_{1-\alpha; q-1}^2$$

6. Calcul de χ^2

7. Décision et conclusion