Tests d'égalité

Moulay Hachemi

/Juin/2013

Contents

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Abou Bekr Belkaïd Tlemcen
Département de mathématique

Mémoire de fin d'étude
Pour l'obtention du diplôme de licence en mathématique

Option : Statistique et Probabilité

TESTS D'EGALITE DE MOYENNES

Présenté par : MOULAY HACHEMI Rahma Yasmina .
Sous la direction de T.MOURID
Année universitaire :2012/2013

0.1 Remerciements

Je tiens tout d'abord à remercier le bon DIEU de m'avoir aidée à accomplir ce modeste travail.

Par suite je veux exprimer toute ma gratitude à mes professeurs qui m'ont suivie tout le long de mes études et qui m'ont fournie de nombreux conseils.

En particulier Monsieur MOURID pour son aide présieuse dans le choix de ma spécialité ,ainsi pour son aide à la réalisation de ce mémoire.

Enfin je tiens également à remercier ma famille d'avoir penser à moi.

0.2 Introduction

0.2.1 Objectif

Dans plusieurs domaines nous sommes confrontés à des phénomènes complexes et aléatoires où la prise de décision est assez difficile. Les outils de la théorie des tests ont pour objectif d'aider à faire des choix entre différentes alternatives.

De façon générale, il s'agit de décider à partir des observations d'un modèle choisi à priori si on peut choisir entre deux hypothèses.

Réaliser un test statistique consiste à mettre en oeuvre une procédure permettant :

-De confronter une hypothèse avec la réalité, ou plus exactement avec ce que l'on perçoit de la réalité à travers les observations disponibles

-De prendre une décision à la suite de cette confrontation.

La théorie statistique des tests est très bien développée et permet de répondre à ce genre de problèmes.

0.2.2 Problématique

Souvent nous sommes confrontés à tester un médicament s'il est efficace. Aussi nous voulons savoir si des pièces sortant d'une machine sont-elle conformes à un usage . De même si des gènes sont significativement exprimés de manière différente dans un tissu pathologique. Ce sont des questions auquelles des tests statistiques peuvent apporter des réponses adéquates. Cependant nous devons respecter quelques conditions qui sont les suivantes :

a. La question doit être posée de sorte qu'il n'yait que deux réponses possibles:"OUI" ou "NON".

b. Une expérimentation planifiée doit fournir des données relatives à cette question.

c. Les données doivent être considérées comme la réalisation de variables aléatoires décrites par un modèle statistique.

d. La réponse à la question se traduit par l'acceptation ou le rejet d'une hypothèse formulée par le modèle précédent.

0.2.3 Généralités sur les tests

Forme des hypothèses En statistiques des tests nous disposons de deux hypothèses notée H_0 dite hypothèse nulle et d'une hypothèse altèrnative notée H_1.

Par des procédés statistiques et au vu des observations nous devons confirmer l'une des hypothèses précédentes.

De façon générale, les hypothèses H_0 et H_1 ne jouent pas des rôles symétriques et la formulation de H_0 doit respecter certaines règles scientifiques.

H_0 est l'hypothèse comménument admise, sauf si une expèrience vient la refuter. Elle est conservée si les résultat de l'experience ne sont pas concluants. Par analogie avec la justice, H_0 est par exemple la présomption d'innocence pour un individu et des éléments matériels sont requis pour rejeter H_0 et prouver la culpabilité.

Soit X une v.a. suivant un loi paramétrique $P_X = P_\theta$ où le paramétre $\theta \in \Theta$ un espace paramétrique $\subset R$.On observe un échantillon $(X_1,...,X_n)$ de la loi P_θ.

Les hypothèses H_0 et H_1se formulent souvent sous cette forme:

$H_0 : \theta \in \Theta_0$ contre H_1: $\theta \in \Theta_1$

où Θ_0 contre Θ_1 sont des parties de R.

dans le cas où ces ensembles sont des singletons on présente les hypothèses sous la forme suivante:

$H_0 : \theta = \theta_0$ contre H_1: $\theta = \theta_1$

Elles sont appelées hypothèses simples.

Dans d'autres situations les parties Θ_0 et Θ_1 sont de la forme suivante:

H_0 : $\theta = \theta_0$ contre H_1: $\theta < \theta_1$
H_0 : $\theta = \theta_0$ contre H_1: $\theta > \theta_1$
H_0 : $\theta > \theta_0$ contre H_1:$\theta \leq \theta_0$
etc...

Ces hypothèses altèrnatives sont appelées composites.

Principe d'un test

La construction d'un test pour tester une hypothèse H_0 contre H_1 repose sur la détermination d'une région de rejet notée D.

Si les observations (X_1,...,X_n) $\in D$ on rejette H_0.Sinon on accèpte H_0.

Risques -Niveau et Puissance Dans un test d'hypothèses statistiques, il existe deux risques d'erreurs:

Risque de première espèce : c'est rejet de l'hypothèse H_0 alors qu'elle est vraie.

Ce risque est noté α=$\sup_{\theta\in\Theta_0}$ $P_{H_0}(D)$

Risque de deuxième espèce : C'est l'acceptation de l'hypothèse H_0 alors qu'elle est fausse

Ce risque est noté 1- β_θ.

la fonction puissance du test notée β_θ est la probabilitée de rejet de H_0 alors qu'elle est fausse β_θ =$P_\theta(D)$, $\theta \in \Theta_1$

la puissance d'un test mesure donc son aptitude à bien rejeter une hypothèse fausse .

En théorie des tests et plus généralement en statistique inférentielle on cherche à construire des tests uniformément les plus puissants (UPP).

Exemple

Dans cet exemple nous montrons l'importance de la formulation des hypothèses et les risques associés.

Pour la consommation de l'eau potable nous considérons μ la moyenne du niveau de radioactivité en picoucuries par litre.

La valeur μ =5 est considérée comme la valeur critique entre eau potable et non potable.On peut donc tester:

H_0:"$\mu \geq 5$"*contre* H_1:"$\mu < 5$ "

H_0 est "l'eau est toxique" contre H_1 "l'eau est potable"

L'érreur de première espèce a pour conséquences de laisser boire de l'eau toxique

alors que l'erreur de deuxième espèce conduit au rejet de l'eau potable.

la puissance d'un test serait la possibilité du rejet de l'eau toxique.

Démarche d'un test

Après avoir défini l'hypothèse H_0 liée au modèle statitique sous-jacent, la démarche d'un test suit les étapes suivantes:

1- Choix de H_0 et H_1et fixer α
2- Détermination de la région de rejet D en fonction de α et H_0.
3- Calcul de la valeur observée de la statistique de test.
4- Conclusion : rejet ou acceptation de H_0 au risque α.
5 - Cacul de la puissance du test β_θ.

Chapter 1

Tests d'égalité

1.1 Test d'égalité de moyenne

1.1.1 Definitions

Dans cette section , la question posée est de comparer deux moyennes de deux échantillons. On veut tester :

H_0: $\mu_1 = \mu_2$ *contre* H_1: $\mu_1 \neq \mu_2$

c'est à dire les deux échantillons proviennent de la même population contre qu'ils proviennent de populations différentes.

On considére deux variables X_1 et Y_1 gaussiennes de moyennes μ_1 et μ_2 et de même variance σ^2 et on tire deux échantillons d'effectifs n_1 et n_2 indépendants.

$\overline{X}_{n_1}$ et $\overline{Y}_{n_2}$ désigne les moyennes empiriques :

$$\overline{X}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$$

$$\overline{Y}_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$$

on a

$$E(\ \overline{X}_{n_1}) = E(\frac{1}{n_1} \sum_{i=1}^{n_1} X_i) = \frac{1}{n_1} \sum_{i=1}^{n_1} E(X_i) = \frac{1}{n_1} * n_1 * \mu_1 = \mu_1$$

.

$$E(\overline{Y}_{n_2}) = E(\frac{1}{n_2} \sum_{i=1}^{n_2} Y_i) = \frac{1}{n_2} \sum_{i=1}^{n_2} E(Y_i) = \mu_2.$$

$$var(\overline{X}_{n_1}) = var(\frac{1}{n_1} \sum_{i=1}^{n_1} X_i) = \frac{1}{n_1^2} var(\sum_{i=1}^{n_1} X_i) = \frac{1}{n_1^2} \sum_{i=1}^{n_1} var(X_i) = \frac{\sigma^2}{n_1}.$$

$$var(\overline{Y}_{n_2}) = var(\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i) = \frac{\sigma^2}{n_2}.$$

l'écart type est :

$$\sigma(\overline{X}_{n_1}) = \frac{\sigma}{\sqrt{n_1}} \ et\ \sigma(\overline{Y}_{n_2}) = \frac{\sigma}{\sqrt{n_2}}.$$

Pour $n_1 > 30$ et $n_2 > 30$ et par le théorème central limite pour σ connu :

$$Z_{n_1} = \frac{\overline{X}_{n_1} - E(\overline{X}_{n_1})}{\sigma(\overline{X}_{n_1})} \Longrightarrow N(0,1)$$

et encore:

$$\overline{X}_{n_1} \Longrightarrow N(\mu_1, \frac{\sigma^2}{n_1})$$

de même pour $\overline{Y}_{n_2}$

$$\overline{Y}_{n_2} \Longrightarrow N(\mu_2, \frac{\sigma^2}{n_2})$$

Le test d'hypothèse

H_0 :$\mu_1 = \mu_2$ contre H_1 $\mu_1 \neq \mu_2$

est basé sur la loi des différences de moyennes des variables $\overline{X}_{n_1}$ et $\overline{Y}_{n_2}$ qui sont des variables indépendantes car elles proviennent de deux échantillons indépendants.

on a

$$E(\overline{X}_{n_1} - \overline{Y}_{n_2}) = E(\overline{X}_{n_1}) - E(\overline{Y}_{n_2}) = \mu_1 - \mu_2.$$

$$var(\overline{X}_{n_1} - \overline{Y}_{n_2}) = var(\overline{X}_{n_1}) + var(\overline{Y}_{n_2}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}.$$

Donc:

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \Longrightarrow N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$$

et pour n_1 ,$n_2 \geq 30$

$$Z_{(n_1,n_2)} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \Longrightarrow N(0,1)$$

Maintennant si σ est *inconnu* ,dans ce cas on prend n_1=n_2 $= n$ et on le remplace par la variance empirique

$$S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{n-1}$$

Pour la loi de la v a

$$Z_n = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sqrt{\frac{S_n^2}{n} + \frac{S_n^2}{n}}} = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sqrt{\frac{2S_n^2}{n}}} = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{S_n\sqrt{\frac{2}{n}}} = \frac{\boxed{\frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{2}{n}}}}}{\boxed{\sqrt{\frac{\frac{(n-1)S_n^2}{\sigma^2}}{(n-1)}}}}$$

on sait que le numérateur suit la loi N(0,1) et le dénominateur est une variable de χ^2_{n-1} divisée par son degré de liberté.

Par le théorème de Fisher: le numérateur et le dénominateur sont des v a indépendantes et leur rapport définit une nouvelle loi de probabilitée usuelle en statistique,appelée loi de Student à n-1 degré de liberté notée T_{n-1}

D'où

$$Z_n \Longrightarrow T_{n-1}$$

1.1.2 Test d'hypothèse bilatéral

on veut tester si les deux échantillons ont la même moyenne:

$\mathrm{H}_0 : \mu_1 = \mu_2$

$\mathrm{H}_1 : \mu_1 \neq \mu_2$

Expression du test

on considère la fonction test ℓ suivante

$$\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = \begin{cases} 1 & \text{si} \quad |\overline{X}_{n_1} - \overline{Y}_{n_2}\ | > k \\ 0 & \text{si} \quad |\overline{X}_{n_1} - \overline{Y}_{n_2}\ | \leq k \end{cases}$$

Région de rejet

La région de rejet D est définie par :

D=$\{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2})/|\overline{X}_{n_1} - \overline{Y}_{n_2}\ |$>k$\}$

Le principe du test est le suivant :

si $\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = 1 \Leftrightarrow |\overline{X}_{n_1} - \overline{Y}_{n_2}|$>k ; on rejette H_0 sinon on l'accepte.

Soit α un risque fixé $\alpha = P_{H_0}$(D)

Détermination de k

D'aprés ce qui précéde on a:

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \Longrightarrow N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

Sous H_0 nous avons :

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \Longrightarrow N(0, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$$

D'où pour n_1 ,$n_2 \geq 30$

$$Z_{n_1,n_2} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \Longrightarrow N(0,1)$$

$$\alpha = P_{H_0}(D) = P_{H_0}(|\overline{X}_{n_1} - \overline{Y}_{n_2}\ | > k) = P_{H_0}(\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}})$$

$$\alpha = P_{H_0}(|Z_{n_1,n_2}| > \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}) \ .$$

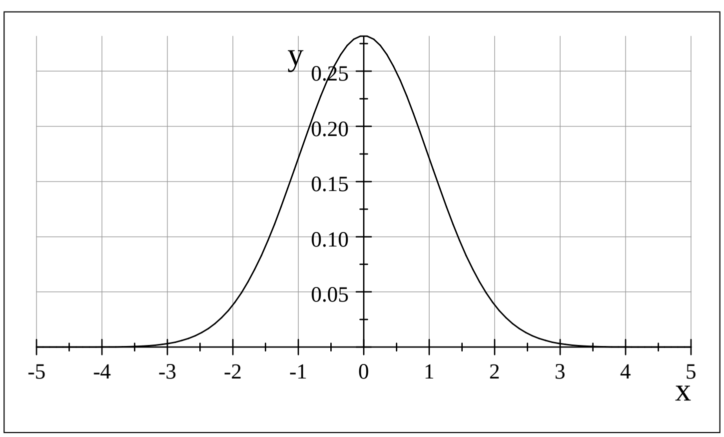
on pose :

$$t_{\alpha/2} = \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

on a aussi

$$1 - \alpha = P_{H_0}(|Z_{n_1,n_2}| < \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}) = P_{H_0}(-t_{\alpha/2} < Z_{n_1,n_2} < t_{\alpha/2})$$

Soit Z$\sim N(0,1)$
On obtient le graphe suivant de la densite de Z $\sim N(0,1)$:



D'où

$$\frac{\alpha}{2} = P_{H_0}(Z > t_{\alpha/2}) \Rightarrow F_Z(t_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

On détermine $t_{\alpha/2}$ à partir de la table de la loi centrée réduite
par suite :

$$k = t_{\alpha/2} * \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

et donc la région de rejet devient :

$$\begin{aligned} D &= \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2})/|\overline{X}_{n_1} - \overline{Y}_{n_2}| > t_{\alpha/2} * \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}\} \\ &= \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2})/\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} > t_{\alpha/2}\} \end{aligned}$$

Règles de décision

on definit les decisions suivantes :
Si $|Z_{n_1,n_2}| > t_{\alpha/2}$,on rejette H_0.
Si $|Z_{n_1,n_2}| < t_{\alpha/2}$, on accepte H_0.

Mise en oeuvre du test

Une valeur z de la variable aléatoire $Z_{n_1,n_2} est$ calculée par

$$z = \frac{\overline{x}_{n_1} - \overline{y}_{n_2}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

On la calcule à partir des deux échantillons et nous utilisons les règles de décision précédentes pour dire si les deux échantillons proviennent de la même population ou non.

Calcul de la fonction puissance du test

Nous commençons par calculer le risque de deuxième espèce:

$$1-\beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \leq t_{\alpha/2}) = P_{H_1}(-t_{\alpha/2} \leq \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \leq t_{\alpha/2})$$

Sous $H_1 : \mu_1 \neq \mu_2$; nous avons d'après le théorème central limite et pour $n_1,n_2 \geq 30$:

$$Z^*_{n_1,n_2} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (\mu_1 - \mu_{2)}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \Longrightarrow N(0,1)$$

Notons $\lambda = \dfrac{(\mu_1 - \mu_{2)}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \dfrac{\delta}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$ $\quad où \quad \delta = \mu_1 - \mu_2$

Donc:

$$1 - \beta_\theta = P_{H_1}(-t_{\alpha/2} - \lambda \leq Z^*_{n_1,n_2} \leq t_{\alpha/2} - \lambda)$$

avec

$$si\ Z \sim N(0,1) \quad alors\ \ 1 - \beta_\theta = F_Z(t_{\alpha/2} - \lambda) - F_Z(-t_{\alpha/2} - \lambda)$$

D'où la fonction puissance:

$$\beta_\theta = \beta(\delta) = 1 - F_Z(t_{\alpha/2} - \lambda) + F_Z(-t_{\alpha/2} - \lambda)$$

Si on pose
u'=$-t_{\alpha/2} - \lambda$, u"=$t_{\alpha/2} - \lambda$
La fonction puissance devient :

$$\beta(\delta) = 1 + F_Z(u') \ - F_Z(u'')$$

qui est l'aire comprise au dessous de la courbe centrée réduite et en dehors de l'intervalle [u',u"].

la longueur de [u',u"] est u"-u'=2t$_{\alpha/2}$=constante.
Si u'<0 $\Rightarrow F_Z(u') = 1 - F_Z(|u'|)$
Si u"<0$\Rightarrow F_Z(u'') = 1 - F_Z(|u''|)$
On conclut que : si u'<0:

$$\beta(\delta) = \begin{cases} 2 - F_Z(|u'|) - F_Z(|u''|) & si \quad u" > 0 \\ 1 - F_Z(|u'|) + F_Z(|u''|) & si \quad u" \leq 0 \end{cases}$$

et si u'>0:

$$\beta(\delta) = \begin{cases} 1 + F_Z(|u'|) - F_Z(|u''|) & si \ \ u" > 0 \\ F_Z(|u'|) + F_Z(|u''|) & si \ \ u" < 0 \end{cases}$$

Exemple 1 :Comparaison de production laitières de deux races bovines

On désire comparer les productions laitières de deux races bovines qui proviennent de la même région :

Pour cela on choisit au hasard et indépendament 50 bêtes de chaque race,et nous disposons ainsi de leurs productions laitières en deux échantillons A et B.

Les rendements moyens de A et B sont : $\overline{x}_{n_1} = 3123l$,$\overline{y}_{n_2} = 3200l$

Question

Est-ce que cette différence entre les deux rendements reflette une certaine information (différence entre les deux races) ou elle est dûe au simple fait du hasard ?

On suppose que dans les deux populations les rendements sont des

v.a normales et de même écart type $\sigma = 0,5l$

Solution Soit A et B les deux échantillons indépendants:
$(X_1, ..., X_{n_1}), (Y_1, ..., Y_{n_2})$ tels que $X_{n_1} \sim N(\mu_1, \sigma^2)$ et $Y_{n_2} \sim N(\mu_2, \sigma^2)$, $\sigma = 0,5l$ et $n_1 = n_2 = n = 100$.

On veut tester les hypothèses suivantes:
$H_0 : \mu_1 = \mu_2$ (absence de différence de rendements entre les deux races)
$H_1 : \mu_1 \neq \mu_2$ (existence de différence de rendements entre les deux races)

1èreméthode : par l'utilisation de la fonction test ℓ Soit le test déterministe de fonction ℓ donnée par :

$$\ell(X_1, ..., X_n, Y_1, ..., Y_n) = \begin{cases} 1 & si \quad |\overline{X}_n - \overline{Y}_n| > k \\ 0 & si \quad |\overline{X}_n - \overline{Y}_n| \leq k \end{cases}$$

C'est le test *de* région de rejet :

$$D = \{(X_1, ..., X_n, Y_1, ..., Y_n) / \frac{|\overline{X}_n - \overline{Y}_n|}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}} > t_{\alpha/2}\} = \{(X_1, ..., X_n, Y_1, ..., Y_n) / \frac{|\overline{X}_n - \overline{Y}_n|}{\sigma\sqrt{\frac{2}{n}}} > t_{\alpha/2}\}$$

Sous H_0 et comme n>30 ,par le théorème central limite
$Z_n = \dfrac{\overline{X}_n - \overline{Y}_n}{\sigma\sqrt{\frac{2}{n}}} \Longrightarrow N(0,1)$, par la table de la loi normale centrée réduite on obtient: $t_{\alpha/2} = 1,96$

Si $Z \sim N(0,1)$
Pour $\alpha = 0,05$, $F_Z(t_{\alpha/2}) = 0,975$
Pour $\alpha = 0,01$,$F_Z(t_{\alpha/2}) = 0,995$, par la table de la loi centrée réduite on a : $t_{\alpha/2} = 2,58$

Pour $\alpha = 0,001$,$F_Z(t_{\alpha/2}) = 0,999$ par la table de la loi centrée réduite $t_{\alpha/2} = 3,29$.

Remarque :
Si on réduit le risque α la région de rejet se rétrécit.

Mise en oeuvre du test Nous calculons à partir des deux échantillons le nombre |z|: $\overline{x_n} = 3123l$, $\overline{y}_n = 3200l$,n=100 ,$\sigma = 0,5l$

$|z| = \dfrac{|\overline{x}_n - \overline{y}_n|}{\sigma\sqrt{\frac{2}{n}}} = \dfrac{|3123 - 3200|}{0,5\sqrt{\frac{2}{100}}} = \dfrac{77}{0,070} = 1,10.$

Pour $\alpha = 0,05$ pour z=1,10 $\leq 1,96.{=}t_{\alpha/2}$.On accepte H_0: Absence de différence de rendement entre les deux races.

Calcul de la fonction puissance $\beta(\delta)$ On rappelle la formule générale de la fonction puissance :

Si u'<0:

$$\beta(\delta) = \begin{cases} 2 - F_Z(|u'|) - F_Z(|u''|) & si \quad u'' > 0 \\ 1 - F_Z(|u'|) + F_Z(|u''|) & si \quad u'' < 0 \end{cases}$$

Si u'>0:

$$\beta(\delta) = \begin{cases} 1 + F_Z(|u'|) - F_Z(|u''|) & si \;\; u" > 0 \\ F_Z(|u'|) + F_Z(|u''|) & si \;\; u" < 0 \end{cases}$$

Dans notre exemple on a :

$\alpha = 0,05$, $t_{\alpha/2} = 1,96, \mu_1 - \mu_2 = 77, \sigma = 0,5$ et $n_1 = n_2 = 100$

$\lambda = \dfrac{(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = 1,1$

$u' = -t_{\alpha/2} - \lambda = -3,6$

$u'' = t_{\alpha/2} - \lambda = 0,86$

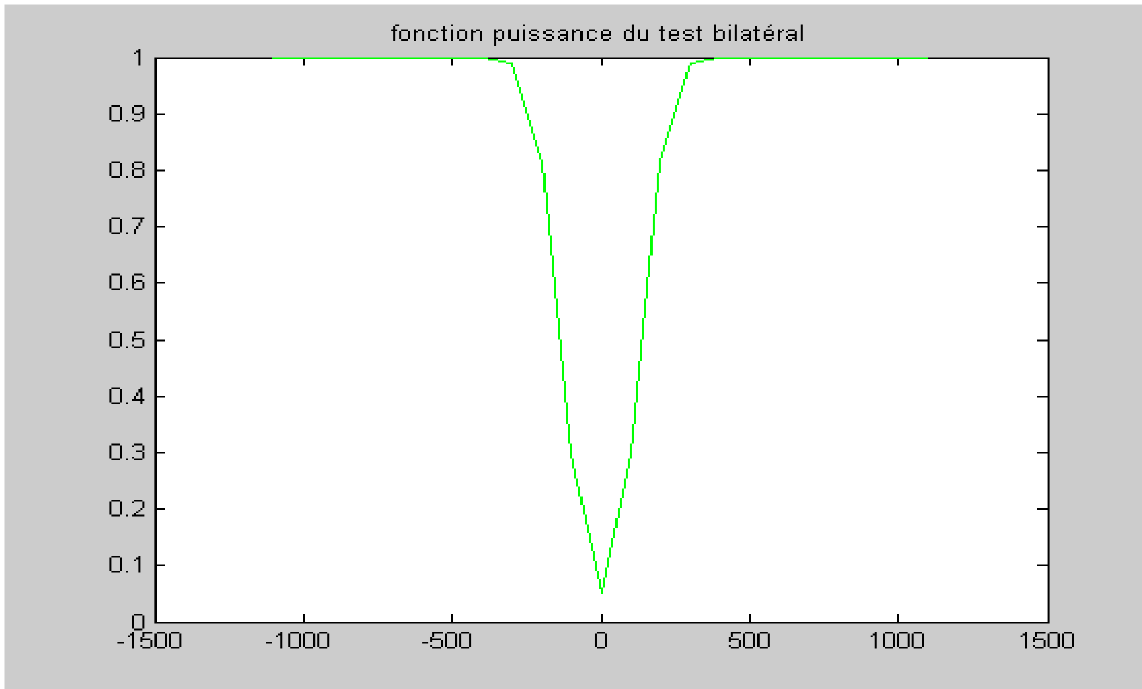
Comme u'<0 et u">0 alors:

$\beta(77) = 2 - F_Z(|-3,6|) - F_Z(|0,86|) = 2 - 0,9995 - 0,8051 = 0,1954.$

Maintenant ,on va étudier les variations de β_θ en fonction de $\delta = |\mu_1 - \mu_2|$, avec n=100,$\alpha = 0,05$ et $\lambda = \dfrac{|\delta|}{0,070}$.

$\|\delta\|$	$\|\lambda\|$	u'	$u"$	$F_Z(\|u'\|)$	$F_Z(\|u''\|)$	$\beta\ (\ \delta)$
0	0	$-1,96$	$1,96$	$0,975$	$0,975$	$0,05$
100	$1,42$	$-3,38$	$0,54$	$0,999$	$0,705$	$0,296$
200	$2,85$	$-4,81$	$-0,89$	$0,999$	$0,813$	$0,814$
300	$4,28$	$-6,24$	$-2,32$	$\simeq 1$	$0,989$	$0,989$
400	$5,71$	$-7,67$	$-3,75$	$\simeq 1$	$0,999$	$0,999$
600	$8,57$	$-10,53$	$-6,61$	$\simeq 1$	$\simeq 1$	$\simeq 1$
700	10	$-11,96$	$-8,04$	$\simeq 1$	$\simeq 1$	$\simeq 1$
800	$11,42$	$-13,38$	$-9,46$	$\simeq 1$	$\simeq 1$	$\simeq 1$
900	$12,85$	$-14,81$	$-10,89$	$\simeq 1$	$\simeq 1$	$\simeq 1$
1000	$14,28$	$-16,24$	$-12,38$	$\simeq 1$	$\simeq 1$	$\simeq 1$
1200	$17,14$	$-19,1$	$-15,18$	$\simeq 1$	$\simeq 1$	$\simeq 1$

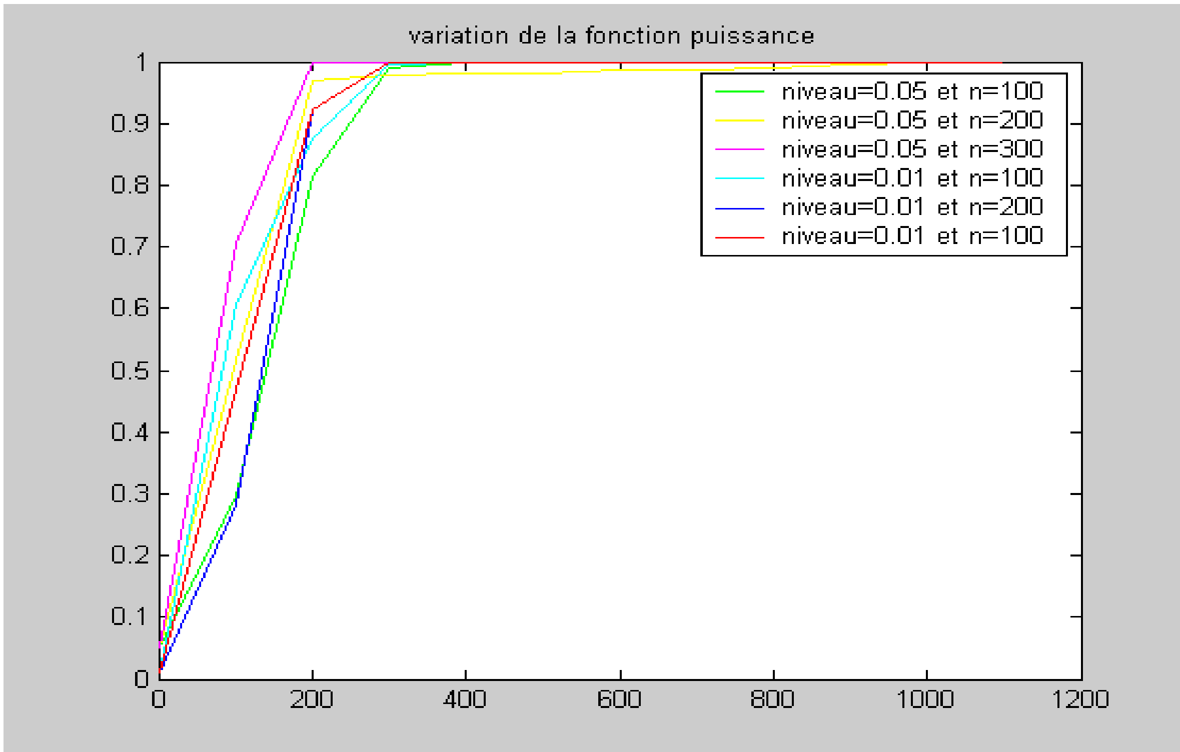
Graphe:



Etude de la variation de $\beta(\delta)$ pour differentes valeurs de de n,α

α	$0,05$			$0,01$		
	n			n		
$\lvert\delta\rvert$	100	200	300	100	200	300
0	$0,05$	$0,05$	$0,05$	$0,01$	$0,01$	$0,01$
100	$0,296$	$0,516$	$0,705$	$0,606$	$0,281$	$0,469$
200	$0,814$	$0,979$	$0,998$	$0,877$	$0,922$	$0,992$
300	$0,989$	$\simeq 1$	$\simeq 1$	$0,9954$	$0,999$	$\simeq 1$
400	$0,999$	$\simeq 1$	$\simeq 1$	$0,991$	$\simeq 1$	$\simeq 1$
500	$0,9999$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
600	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
700	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
800	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
900	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
1000	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$
1100	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$

Graphe:



Remarque
La puissance tend vers 1 quand l'écart entre les moyennes est important.

$2^{ième}$méthode: Test de Neyman-Pearson

nous rappelons le résultat suivant :

Pour $\alpha \in [0,1]$, il existe un test UPP de région critique D:
$D = \{(X_1, ...X_n)/\frac{L(X_1,...X_n,\theta_1)}{L(X_1,...X_n,\theta_0)} > k\}$
où $(X_1, ...X_n)$ est un échantillon de taille n de la variable X dont la loi de probabilité dépend du paramètre θ ,et $L(X_1, ...X_n)$ est la vraisemblance de l'échantillon.

Application :

On a

$(X_1, ..., X_{n_1}), (Y_1, ..., Y_{n_2})$ tels que $X_{n_1} \sim N(\mu_1, \sigma^2)$ et $Y_{n_2} \sim N(\mu_2, \sigma^2)$, $\sigma = 0,5l$ et $n_1 = n_2 = 100$.

On veut tester les hypothèses suivantes:
$H_0 : \mu_1 = \mu_2$
contre
$H_1 : \mu_1 \neq \mu_2$
qui est équivalent à
$H_0 : \mu_1 - \mu_2 = 0$
contre
$H_1 : \mu_1 - \mu_2 \neq 0$

On pose $\theta = \mu_1 - \mu_2$
Les hypothèses deviennent donc :
$H_0 : \theta = 0$
contre
$H_1 : \theta \neq 0$
On considère l'échantillon $Z=(Z_1, Z_2, ..., Z_n) = (X_1-Y_1, X_2 - Y_2, ..., X_{n_1} - Y_{n_2})$, *or* X_1et Y_1 sont indépendantes donc $Z_1 \sim N(\mu_1 - \mu_2, 2\sigma^2)$
ou encore $Z_1 \sim N(\theta, 2\sigma^2)$
L'énoncé devient:
soit l'échantillon:$Z=(Z_1, Z_2, ..., Z_n)$ *avec* $Z_i \sim N(\theta, 2\sigma^2)$
On applique le théorème de Neyman -Pearson

$$L(Z_1, ...Z_n, \theta_0) = f_0(Z_1, ...Z_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}\sqrt{2}} \exp(\frac{-1}{4\sigma^2} z_i^2) = (\frac{1}{\sigma\sqrt{2\pi}\sqrt{2}})^n \exp(\frac{-1}{4\sigma^2} \sum\nolimits_{i=1}^{n} z_i^2)$$

$$L(Z_1, ...Z_n, \theta_1) = f_1(Z_1, ...Z_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}\sqrt{2}} \exp(\frac{-1}{4\sigma^2} (z_i-\theta)^2) = (\frac{1}{\sigma\sqrt{2\pi}\sqrt{2}})^n \exp(\frac{-1}{4\sigma^2} \sum\nolimits_{i=1}^{n} (z_i-\theta)^2)$$

On obtient

$$\frac{f1(Z_1, ...Z_n)}{f_0(Z_1, ...Z_n)} = \exp(\frac{-1}{4\sigma^2}(\sum\nolimits_{i=1}^{n} (z_i-\theta)^2 - \sum\nolimits_{i=1}^{n} z_i^2) = \exp(\frac{-1}{4\sigma^2}(\sum\nolimits_{i=1}^{n} \theta^2 - 2\theta z_i)$$

et par suite

$$\frac{f1(Z_1,...Z_n)}{f_0(Z_1,...Z_n)} = \exp(-\frac{n\theta^2}{4\sigma^2})\exp(\frac{\theta}{2\sigma^2}\sum_{i=1}^{n} z_i)$$

La région de rejet du test de Neyman Pearson est:

D=$\{(Z_1,...Z_n)/\exp(-\dfrac{n\theta^2}{4\sigma^2})\exp(\frac{\theta}{2\sigma^2}\sum_{i=1}^{n} z_i) > k\}$

D=$\{(Z_1,...Z_n)/\sum_{i=1}^{n} z_i > \dfrac{\ln k + \dfrac{n\theta^2}{4\sigma^2}}{\frac{1}{2\sigma^2}}\}$

D=$\{(Z_1,...Z_n)/\overline{Z_n} > \dfrac{\ln k + \frac{n}{4\sigma^2}\theta^2}{\frac{\theta}{2\sigma^2}}\}$

$$D = \{(Z_1,...Z_n)/\overline{Z_n} > \frac{2\sigma^2}{n\theta}(\ln k + \frac{n}{4\sigma^2}\theta^2)\ \}$$

D change en fonction du signe de θ :

$1^{èr}$cas : $\theta > 0$

D$^+ = \{(Z_1,...Z_n)/\ \overline{Z_n} > \frac{2\sigma^2}{n\theta}(\ln k + \frac{n}{4\sigma^2}\theta^2)\} = \{(Z_1,...Z_n)/\ \overline{Z_n} > C'\}$, avec C'=$\frac{2\sigma^2}{n\theta}(\ln k + \frac{n}{4\sigma^2}\theta^2)$ $\theta > 0$.

$2^{ème}$ cas: $\theta < 0$

D$^- = \{(Z_1,...Z_n)/\ \overline{Z_n} < \frac{2\sigma^2}{n\theta}(\ln k + \frac{n}{4\sigma^2}\theta^2)\} = \{(Z_1,...Z_n)/\ \overline{Z_n} < C''\}$,avec C"=$\frac{2\sigma^2}{n\theta}(\ln k + \frac{n}{4\sigma^2}\theta^2)$ $\theta < 0$.On a : $D = D^+ \cup D^-$.Par conséquent

$D = \{(Z_1,...Z_n)/\ \overline{Z_n} > C'$ ou $\overline{Z_n} < C''\}$

On remarque que $C' = -C''$ d'où $D = \{(Z_1,...Z_n)/\ \overline{Z_n} > C'$ ou $\overline{Z_n} < -C'\}$ ou encore

$$D = \{(Z_1,...Z_n)/|\overline{Z_n}| > C'\} = \{(Z_1,...Z_n)/|\overline{X_A} - \overline{X_B}| > C'\}$$

c'est la même région de rejet que le test précédent. Donc le premier test est un test de N.P qui est U.P.P.

1.1.3 Test d'hypothèse unilatéral

On veut tester si les deux échantillons ont la même moyenne ou bien si l'une est supérieur à l'autre:

Soit les deux échantillons suivants:

X=$(X_1,...,X_{n_1})$ et Y=$(Y_1,...Y_{n_2})$

soit le test:

H$_0$: $\mu_1 = \mu_2$

contre

H$_1$: $\mu_1 > \mu_2$

Expression du test

on considère la fonction test ℓ suivante

$$\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = \begin{cases} 1 & si\ \overline{X}_{n_1} - \overline{Y}_{n_2} > k \\ 0 & si\ \overline{X}_{n_1} - \overline{Y}_{n_2} \leqslant k \end{cases}$$

1.1.4

Région de rejet

La région de rejet D est définie par :

$$D = \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2})/\overline{X}_{n_1} - \overline{Y}_{n_2} \ > k\}$$

Le principe du test est le suivant :

si $\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = 1 \Leftrightarrow \overline{X}_{n_1} - \overline{Y}_{n_2} \ > k$; on rejette H_0sinon on l'accepte.

Soit α un risque fixé $\alpha = P_{H_0}$(D)

Détermination de k:

Pour n_1 et $n_2 > 30$

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \Longrightarrow N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

Sous H_0 nous avons :

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \Longrightarrow N(0, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

D'où

$$Z_{n_1,n_2} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Longrightarrow N(0,1)$$

$$\alpha = P_{H_0}(D) = P_{H_0}(\ \overline{X}_{n_1} - \overline{Y}_{n_2} > k) = P_{H_0}(\frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}})$$

Donc:

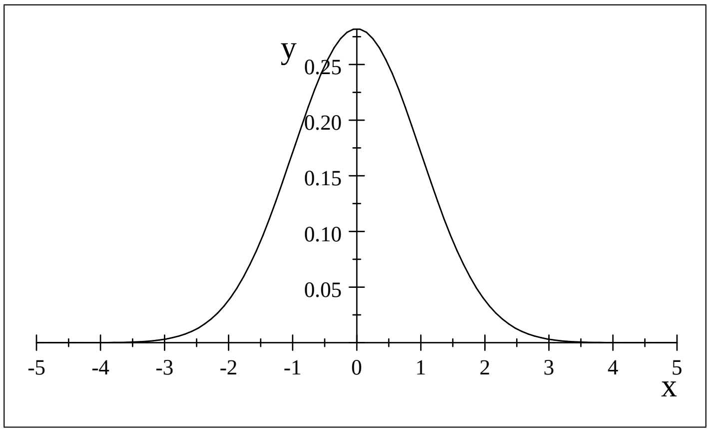
$$\alpha = P_{H_0}(Z_{n_1,n_2}) > \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}) \ .$$

on pose $t_\alpha = \dfrac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$,$si \ Z \sim N(0,1)$

Donc

$$1 - \alpha = P_{H_0}(Z \leq \frac{k}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}) = P_{H_0}(Z \leq t_\alpha)$$

On obtient le graphe suivant :



D'où

$$\alpha = P_{H_0}(Z > t_\alpha) \Rightarrow F_Z(t_\alpha) = 1 - \alpha$$

On détermine t_α à partir de la table de la loi centrée réduite par suite

$$k = t_\alpha * \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

et donc la région de rejet devient:

$$D = \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) / \overline{X}_{n_1} - \overline{Y}_{n_2} > t_\alpha * \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\}$$

$$D = \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) / \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_\alpha\}$$

Règles de décision

on definie les décisions suivantes :
Si $Z_{n_1,n_2} > t_\alpha$,on rejette H$_0$.
Si $Z_{n_1,n_2} \leq t_\alpha$, on accepte H$_0$.

Mise en oeuvre du test

on calcule le nombre $\frac{\overline{x}_{n_1} - \overline{y}_{n_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ à partir des deux échantillons ,nous utilisons les règles de décision précédentes pour conclure.

Calcul de la fonction puissance du test

Nous commençons par calculer le risque de deuxième espèce:

$$1-\beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(\frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_\alpha)$$

Sous H$_1$: $\mu_1 \neq \mu_2$;nous avons d'aprés le théorème central limite:

$$Z^{"}_{n_1,n_2} = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Longrightarrow N(0,1)$$

Notons $\lambda = \dfrac{(\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} et\ \delta = \mu_1 - \mu_2$

Donc: 1-$\beta_\theta = P_{H_1}(Z^{"}_{(n_1,n_2)} \leq t_\alpha - \lambda)$ si Z$\sim N(0,1)$ alors

$$1-\beta_\theta = F_Z(t_\alpha - \lambda)$$

D'où la fonction puissance est

$$\begin{cases} \beta(\delta) = 1 - F_Z(|t_\alpha - \lambda|) & si\ t_\alpha - \lambda > 0 \\ \beta(\delta) = F_Z(|t_\alpha - \lambda|) & si\ t_\alpha - \lambda \leq 0 \end{cases}$$

Exemple 2 : Les faiseurs de pluie

Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel X des pluies en milimètres par an suit une loi normale N(600,100^2).Des entrepreuneurs,surnommés" faiseurs de pluie " prétendaient pouvoir augmenter de 50mm,le niveau moyen de pluie,ceci par insémination des nuages au moyen d'iodure d'argent.Leur procédé fut mis à l'ev ssai entre 2001 et 2009 et on releva les hauteurs de pluies suivantes:

Année	2001	2002	2003	2004	2005	2006	2007	2008	2009
milimètres	510	614	780	512	501	603	788	534	650

Question: Que pouvait -on conclure ?Deux hypothèses s'affrontaient : ou bien l'insémination était sans effet ,ou bien elle augmentait réellement le niveau moyen de pluie de 50mm.

Ces hypothéses se formalisent comme suit:

Si μ désigne la moyenne de v a r égal au niveau annuel de pluie.

H$_0$: $\mu = 600mm$

H$_1$: $\mu = 650mm$

ou encore

H$_0$: $\mu = 600mm$

H$_1$: $\mu > 600mm$

Les agriculteurs hésitaient à opter pour le procédé forcément onéreux des faiseurs de pluie.Ainsi ,il fallait donc que l'éxpérience puisse les convaincre,c'est à dire que les faits observés contredisent nettement l'hypothèse H_0.

Les agriculteurs n'étaient donc décidés à abondoner H_0 qu'en présence de faits expérimentaux traduisant une éventualité improbable de H_0.

Comment décider? il s'agit de tester la moyenne μ .On utilise la moyenne empirique: $\overline{X_n} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ que l'on sait convergente vers μ par la loi forte des grands nombres.

Alors si H_0 est vraie $\overline{X_9} \sim N(600, \frac{100^2}{9})$

Les données relevées indiquent que $\overline{x_9} = 610,2mm$ (la moyenne obtenue sur toutes les années).

Région de rejet:

Pour $\alpha = 0,05$.

D=$\{(X_1, X_2, ..., X_9)/\overline{X_9} > \overline{x}_9)$

sous H_0 : la v a r Z=$\frac{\overline{X_n}-600}{100/9}$ suit la loi normale N(0,1)

$P_{H_0}(\frac{\overline{X_n}-600}{100/9} > \frac{610,2-600}{100/3}) = P_{H_0}(Z > 0,306) = 1 - 0,6406 = 0,354$

Conclusion:

Comme cette probabilité critique est un risque supèriure à $\alpha = 0,05$,alors on accepte l'hypothèse H_0,c'est à dire que les données sont compatibles avec H_0.

Ainsi le procédé des faiseurs de pluie n'a pas d'effet .

Exemple 2

On considère l'exemple de la comparaison de production laitière de deux races bovines:

On rappelle que:

$(X_1, ..., X_{n_1}), (Y_1, ..., Y_{n_2})$ deux échantillons tels que $X_{n_1} \sim N(\mu_1, \sigma^2)$ et $Y_{n_2} \sim N(\mu_2, \sigma^2)$, $\sigma = 0,5l$ et $n_1 = n_2 = 100$.

Par le test unilatéral:

Les hypothèses deviennent:

$H_0 : \mu_1 = \mu_2$ contre $H_1 = \mu_1 > \mu_2$.

On considère la fonction test suivante:

$$\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = \begin{cases} 1 & si\ \overline{X}_{n_1} - \overline{Y}_{n_2} > k \\ 0 & si\ \overline{X}_{n_1} - \overline{Y}_{n_2} \leqslant k \end{cases}$$

Région de rejet

$$D = \{(X_1, ..., X_n, Y_1, ...Y_n)/\frac{\overline{X}_n - \overline{Y}_n}{\sigma\sqrt{\frac{2}{n}}} > t_{\alpha\}}$$

Pour n assez grand $Z_n = \dfrac{\overline{X}_n - \overline{Y}_n}{\sigma\sqrt{\frac{2}{n}}} \Longrightarrow N(0,1)$.Soit Z$\sim N(0,1)$.Pour $\alpha = 0,05$, $F_Z(t_\alpha) = 0,95$,par la table de la loi normal centrée réduite
on obtient: $t_\alpha = 1,6449$

Mise en oeuvre du test On calcule $z = \dfrac{\overline{x}_n - \overline{y}_n}{\sigma\sqrt{\frac{2}{n}}} = 1,1$

1,1<1,64 donc on accepte H_0

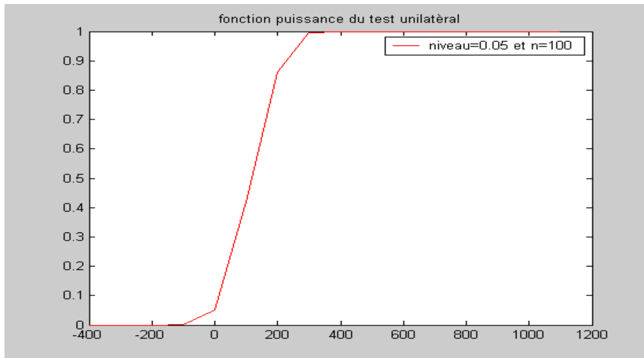
Calcul de la fonction puissance Formule générale:

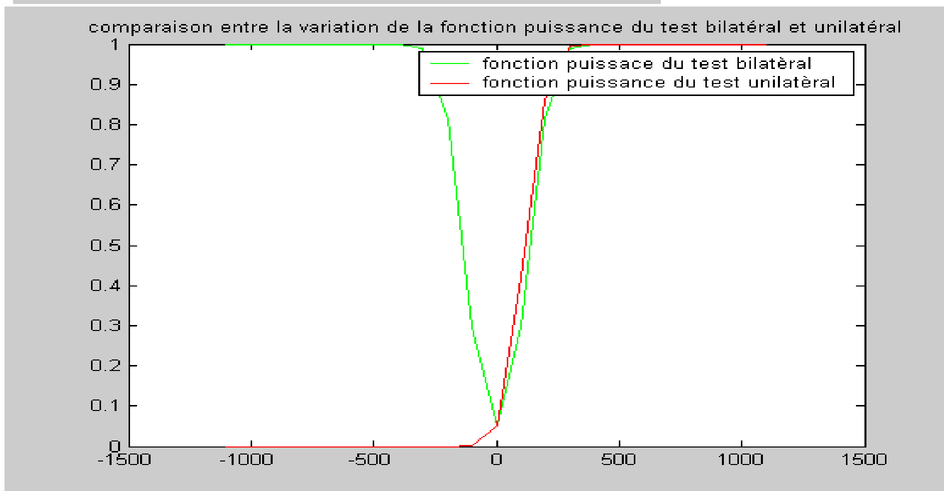
$$\beta(\delta) = \begin{cases} 1 - F_Z|t_\alpha - \lambda| & si \quad t_\alpha - \lambda > 0 \\ F_Z|t_\alpha - \lambda| & si \quad t_\alpha - \lambda < 0 \end{cases}$$

Où $\lambda = \dfrac{\mu_1 - \mu_2}{\sigma\sqrt{\frac{2}{n}}} = 1,1; \delta = \mu_1 - \mu_2 = 77; t_\alpha = 1,6449$
on a $t_\alpha - \lambda = 0,54 > 0, d'où \ \ \beta(77) = 1 - F_Z|0,54| = 1 - 0,7054 = 0,2946$
Remarque:
La puissance du test unilatéral est supérieure à celle du test bilatéral.
Mais est-ce que ce dernier est plus puissant que le premier?ce qu'on va voir à l'aide de l'étude des variations de la fonction puissance du test
en fonction de δ $(n = 100, \alpha = 0,05)$

$\boldsymbol{\delta}$	$\boldsymbol{\lambda}$	$\mathbf{t}_\alpha - \boldsymbol{\lambda}$	$\mathbf{F}_Z\|\mathbf{t}_\alpha - \boldsymbol{\lambda}\|$	$\boldsymbol{\beta}_\theta$
-400	-5,71	7,35	$\simeq \mathbf{1}$	$\simeq \mathbf{0}$
-300	-4,28	5,92	$\simeq \mathbf{1}$	$\simeq \mathbf{0}$
-200	-2,85	4,49	$\simeq \mathbf{1}$	$\simeq \mathbf{0}$
-100	-1,42	3,06	$\mathbf{0,998}$	$\mathbf{0,002}$
0	0	1,64	0,05	$\mathbf{0,05}$
100	1,42	0,22	$\mathbf{0,58}$	$\mathbf{0,42}$
200	2,85	-1,21	$\mathbf{0,886}$	$\mathbf{0,886}$
300	4,28	-2,64	$\mathbf{0,995}$	$\mathbf{0,995}$
400	5,71	-4,07	$\simeq \mathbf{1}$	$\simeq \mathbf{1}$
500	7,14	-5,49	$\simeq \mathbf{1}$	$\simeq \mathbf{1}$
600	8,57	-6,92	$\simeq \mathbf{1}$	$\simeq \mathbf{1}$
700	10	-8,35	$\simeq \mathbf{1}$	$\simeq \mathbf{1}$
800	11,42	-9,77	$\simeq \mathbf{1}$	$\simeq \mathbf{1}$

Graphes





Remarque :
Le test d'hypothèse unilatéral
$H_0 : \mu_1 = \mu_2$
contre
$H_1 : \mu_1 < \mu_2$

On considère la fonction test suivante:

$$\ell(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) = \begin{cases} 0 \text{ si } \overline{Y}_{n_2} \ \text{-}\overline{X}_{n_1} > k \\ 1 si \ \overline{X}_{n_1} - \overline{Y}_{n_2} \leq k \end{cases}$$

La région de rejet D est définie par :

$$D = \{(X_1, ..., X_{n_1}, Y_1, ...Y_{n_2}) / \overline{X}_{n_1} - \overline{Y}_{n_2} \ \leq k\}$$

1.2 Comparaison de deux fréquences

1.2.1 Principe du test

Soit X une variable qualitative prenant deux modalités (succès X=1, échec X=0) observée sur 2 populations et deux échantillons indépendants extraits de ces deux populations.

On fait l'hypothèse que les deux échantillons proviennent de 2 populations dont les probabilités de succès sont identiques.

Le problème est de savoir si la différence entre les deux fréquences observées est réelle ou explicable par les fluctuations d'échantillonnage.

Soit les deux échantillons:

A d'effectif n_A et de fréquence f_A.

B d'effectif n_B et de fréquence f_B.

On note F_A, F_B deux va.

F_A : La va qui à chaque échantillon de taille n_A ,associe la fréquence f_A.

F_B : La va qui à chaque échantillon de taille n_B,associe la fréquence f_B.

1.2.2 Loi de $\mathbf{F}_A$-$\mathbf{F}_B$

Loi de $\mathbf{F}_A$

Soit $(Y_{1_A}, ..., Y_{n_A A})$ défini par :

$$Y_{i_A} = \begin{cases} 1 \quad si \quad X \quad se \quad réalise. \\ 0 \quad si \quad X \quad ne \quad se \quad réalise \quad pas \end{cases}$$

Donc

P($Y_{i_A} = 1) = P(X$ se réalise)=p_A=(proportion ou fréquence du caractére X qui se réalise dans la population)

Alors, p_A=f_A et

$$Y_{i_A} \sim B(p_A).$$

La va $\sum_{i=1}^{n_A} Y_{i_1}$ représente le nombre d'individus ayant le caractère X:

$$\sum_{i=1}^{n_A} Y_{i_A} \sim B(n_A, p_A)$$

car les Y_{i_A} sont iid

Donc

$$\overline{Y}_{n_A} = \frac{1}{n_A}\sum_{i=1}^{n_A} Y_{i_A}$$

représente la va qui a chaque n$_A$ associe la fréquence f$_A$ du caractère X dans l'échantillon.Avec F$_A = \overline{Y}_{n_A}$.

Les va Y$_{i_A}$ sont iid,et

$$E(Y_{i_A}) = p_A, V(Y_{i_A}) = p_A(1-p_A)$$

par le TCL et pour n$_A > 30$

$$\sqrt{n_A}\frac{(\overline{Y}_{n_A} - p_A)}{\sqrt{p_A(1-p_A)}} \Longrightarrow N(0,1)$$

Alors

$$\sqrt{n_A}\frac{(F_A - p_A)}{\sqrt{p_A(1-p_A)}} \Longrightarrow N(0,1)$$

D'où pour n$_A$>30

$$F_A \Longrightarrow N(p_A, \frac{p_A(1-p_A)}{n_A})$$

Loi de F$_B$

De même pour F$_B$ on a pour n$_B$ >30

$$F_B \Longrightarrow N(p_B, \frac{p_B(1-p_B)}{n_B}) avec\ p_B = f_B.$$

Loi de F$_A$−F$_B$ Les va F$_A$ et F$_B$ sont gaussiennes et indépendantes, puisque les échantillons sont indépendants.

Donc F$_A - F_B$ est gaussienne.

$$E(F_A - F_B) = E(F_A) - E(F_B) = p_A - p_B$$

$$V(F_A - F_B) = V(F_A) + V(F_B) = \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}.$$

Finalement pour n$_A$ et n$_B$ sont assez grand

$$F_A - F_B \Longrightarrow N(p_A - p_B, \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B})$$

1.2.3 Test d'hypothèse bilatéral

Hypothèses

Tester $H_0 : p_A = p_B$ contre $H_1 : p_A \neq p_B$

Expression du test

On considère la fonction test ℓ définie par :

$$\ell(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B}) = \begin{cases} 1 & \text{si } |F_A - F_B| > k \\ 0 & \text{si } |F_A - F_B| \leq k \end{cases}$$

La région de rejet D est définie par :

D=$\{(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B})/|F_A - F_B| >k\}$

Le principe du test est le suivant :

si $\ell(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B}) = 1 \Leftrightarrow | F_A - F_B| >k$; on rejette H_0sinon on l'accepte.

Soit α un risque fixé $\alpha = P_{H_0}(D)$

Détermination de k

D'aprés ce qui précéde on a:

$$F_A - F_B \quad \Longrightarrow N(p_A - p_B, \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B})$$

Sous H_0 nous avons :

$$F_A - F_B \Longrightarrow N(0, \frac{p_A(1-p_A)}{n_A} + \frac{p_A(1-p_A)}{n_B})$$

D'où pour n_A, n_B assez grand

$$Z_{n_A,n_B} = \frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}} \Longrightarrow N(0,1)$$

$$\alpha = P_{H_0}(D)=P_{H_0}(|F_A - F_B| > k) = P_{H_0}(\frac{|F_A - F_B|}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}} > \frac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}})$$

$$\alpha = P_{H_0}(|Z_{n_A,n_B}| > \frac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}})$$

on pose $t_{\alpha/2} = \frac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}}$, on a aussi

1-$\alpha = P_{H_0}(|Z_{n_A,n_B}| < t_{\alpha/2}) = P_{H_0}(|Z_{n_A,n_B}| < t_{\alpha/2})$

Si Z$\sim N(0,1)$

$$\frac{\alpha}{2}=P_{H_0}(Z\ >t_{\alpha/2})\Rightarrow F_Z(t_{\alpha/2})=1-\frac{\alpha}{2}$$

On détermine $t_{\alpha/2}$ à partir de la table de la loi centrée réduite
par suite

$$k=t_{\alpha/2}\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}$$

et donc la région de rejet devient:

$$D=\{(X_1,...,X_{n_A},Y_1,...Y_{n_B})/|F_A-F_B|>k\}$$

$$D=\{(X_1,...,X_{n_A},Y_1,...Y_{n_B})/\frac{|F_A-F_B|}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}>t_{\alpha/2}\}$$

Règles de décision

-Si $|Z_{n_A,n_B}\ |>t_{\alpha/2}$,on rejette H$_0$.
-Si $|Z_{n_A,n_B}|<t_{\alpha/2}$, on accépte H$_0$.
Mise en oeuvre du test.
on calcule le nombre z=$\frac{|f_A-f_B|}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}$ à partir des deux échantillons ,nous utilisons les règles de décision précédentes pour dire si les deux échantillons proviennent de la même population ou non.

Remarque

On général p est inconnu on utilise les deux échantillons et on estime p par p̂ donnée par:

$$\hat{p}=\frac{n_AF_A+n_BF_B}{n_A+n_B}$$

En effet

$$\hat{p}=\frac{n_AF_A+n_BF_B}{n_A+n_B}=\frac{n_A\overline{Y}_{n_A}+n_B\overline{Y}_{n_B}}{n_A+n_B}$$

Pour n_A,n_B assez grand
$\overline{Y}_{n_A}\sim E(Y_{i_A})p.s$ par la loi forte des grands nombres
$\overline{Y}_{n_B}\sim E(Y_{i_B})p.s$ par la loi forte des grands nombres
Alors pour n_A,n_B assez grand:

$$\hat{p}\sim\frac{n_Af_A+n_Bf_B}{n_A+n_B}p.s$$

Sous H$_0$: $f_A=f_B=p\ alors\ \hat{p}\sim\frac{n_Af_A+n_Bf_A}{n_A+n_B}=p$

D'où $\hat{p}$ *est* un estimateur convergeant p.s.

L'expression de la fonction puissance reste la même soit pour le test bilatéral ou unilatéral mais avec

$\lambda' = \frac{f_A - f_B}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A}+\frac{1}{n_B})}}$

Calcul de la fonction puissance du test

Nous commençons par calculer le risque de deuxième espèce:

1-$\beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(|F_A - F_B| \leq t_{\alpha/2}) = P_{H_1}(-t_{\alpha/2} \leq F_A - F_B \leq t_{\alpha/2})$

Sous H_1 : $p_A \neq p_B$;nous avons d'aprés le théorème central limite:

$$Z^*_{n_A,n_B} = \frac{F_A - F_B - (p_A - p_B)}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}} \Longrightarrow N(0,1)$$

Notons $\lambda' = \dfrac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$ *et* $\delta = p_A - p_B$

Donc

$$1 - \beta_\theta = P_{H_1}(-t_{\alpha/2} - \lambda' \leq Z^*_{n_A,n_B} \leq t_{\alpha/2} - \lambda')$$

avec Z$\sim N(0,1)$ alors

$$1 - \beta_\theta = F_Z(t_{\alpha/2} - \lambda') - F_Z(-t_{\alpha/2} - \lambda')$$

D'où la fonction puissance:

$$\beta(\delta) = 1 - F_Z(t_{\alpha/2} - \lambda') + F_Z(-t_{\alpha/2} - \lambda')$$

On pose

u'=$-t_{\alpha/2} - \lambda'$, u"=$t_{\alpha/2} - \lambda'$

La fonction puissance devient :

$\beta(\delta) = 1 + F_Z(u') \ - F_Z(u'')$

Si u'<0 $\Rightarrow F_Z(u') = 1 - F_Z(|u'|)$

Si u"$\leq$0$\Rightarrow F_Z(u'') = 1 - F_Z(|u''|)$

On conclut que : si u'<0:

$$\beta(\delta) = \begin{cases} 2 - F_Z(|u'|) - F_Z(|u''|) \ si \ u" > 0 \\ 1 - F_Z(|u'|) + F_Z(|u''|) \ si \ u" \leq 0 \end{cases}$$

et si u'>0:

$$\beta(\delta) = \begin{cases} 1 + F_Z(|u'|) - F_Z(|u''|) \ si \ u" > 0 \\ F_Z(|u'|) + F_Z(|u''|) \qquad si \ u" \leq 0 \end{cases}$$

Exemple 3 : Pouvoirs de détection de deux tests médicaux Soit P une population d'individus qui reprèsente une certaine maladie M. Pour déceler la présence de cette maladie chez un individu on dispose de deux tests médicaux T_A et T_B

on extrait de P un échantillon notée A d'effectif n_A=300 et on applique aux individus le test T_A.

On décèle la présence de la maladie chez 243 sujets.

Indépendament de la première expèrience, on extrait de P un échantillon notée B d'effectif n_B=200 et on applique aux individus le test T_B.

On décèle la présence de la maladie chez 152 sujets.

On se propose de savoir si les deux tests médicaux ont un pouvoir de détection égal.

Pour cela on construit un test bilatéral permettant de répondre à cette interrogation au risque de 5%

On peut se demander :

1) D'après ces résultats expérimentaux,des estimations ponctuelles,notées respectivement f_A et f_B ,des pouvoirs de détection (inconnus) p_A et p_B

de chacun des tersts médicaux T_A et T_B?

2) Si on note F_A la variable aléatoire qui à tout échantillon de taille 300, associe le pouvoir de détection du test T_A dans cet échantillon . Quelle est la loi de F_A?

De même, on note F_B la variable aléatoire qui , à tout échantillon de taille 200, associe le pouvoir de détection du test T_B dans cet échantillon . Quelle est la loi de F_B?

Résolution: 1) Estimation ponctuelle.Dans l'échantillon A le pouvoir de détection de T_A est $f_A = \frac{243}{300}$, Donc $f_A = 0,81$ est une estimation ponctuelle de p_A.De même dans l'échantillon B , le pouvoir de détection de T_B est $f_B = \frac{152}{200}$ donc $f_B = 0,76$ est une estimation ponctuelle de p_B.

2) La variable aléatoire F_A qui , à tout échantillon de taille 300, associe le pouvoir de détection du test T_A dans cet échantillon ,suit la loi $N(p_A, \sqrt{\frac{p_A(1-p_A)}{n_A}})$

La variable aléatoire F_B qui , à tout échantillon de taille 200, associe le pouvoir de détection du test T_B dans cet echantillon ,suit la loi $N(p_B, \sqrt{\frac{p_B(1-p_B)}{n_B}})$

Hypothèses à tester :

H_0 :‹‹ les pouvoirs de détection des deux tests sont égaux ›› ou encore $p_A = p_B$

Contre H_1 : ‹‹ les pouvoirs de détection des deux tests sont différents ›› ou encore $p_A \neq p_B$.C'est un test bilatéral

Région de rejet:
D=$\{(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B})/|F_A - F_B| > k\}$
sous l'hypothèse H_0

$$F_A - F_B \Longrightarrow N(0, \sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})})$$

Par suite

$$Z_{n_A,n_B} = \frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}} \Longrightarrow N(0,1)$$

Donc la région de rejet devient

$$D = \{(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B})/\frac{|F_A - F_B|}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}} > t_{\alpha/2}\}$$

$$D = \{(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B})/|Z_{n_A,n_B}| > t_{\alpha/2}\}$$

Au risque 5%, on rejette H_0 si $|Z_{n_A,n_B}| > 1,96$
Mise en oeuvre du test :
On calcule la valeur de la v.a Z_{n_A,n_B}

$$z = \frac{|f_A - f_B|}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}}$$

p_A et p_B étant inconnus ils sont estimés par

$$p = \frac{n_A f_A + n_B f_B}{n_A + n_B} = 0,79$$

Alors

$$z_{n_A,n_B} = \frac{0,81 - 0,76}{\sqrt{0,79 + 0,21(\frac{1}{300} + \frac{1}{200})}} = 1,34 < 1,96$$

donc on accepte H_0.
Conclusion :
On conclut que les pouvoirs de détection des tests T_1 et T_2 ne sont pas différents au risque 5% de se tromper

Puissance:

Au seuil de 5% on a $t_\alpha = 1,96$ et $\lambda' = 1,34$
Donc $u'' = 0,62 > 0$ et $u' = -3,30 < 0$ alors on est dans le cas où
Si $u' < 0$:

$$\beta_\theta = \begin{cases} 2 - F_Z(|u'|) - F_Z(|u''|) \text{ si } u'' > 0 \\ 1 - F_Z(|u'|) + F_Z(|u''|) \text{ si } u'' \leq 0 \end{cases}$$

Donc

$$\beta_\theta = 2 - F_Z(|-3,30|) - F_Z(|0,62||) = 0,6945$$

1.2.4 Test unilatéral

Hypothèses

tester $H_0 : p_A = p_B$ contre $H_1 : p_A > p_B$

Expression du test

On considère la fonction test :

$$\ell(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B}) = \begin{cases} 1 \text{ si } F_A - F_B > k \\ 0 \text{ si } F_A - F_B \leq k \end{cases}$$

Région de rejet On définit la région de rejet par:

$$D = \{(X_1, ..., X_{n_A}, Y_1, ...Y_{n_B})/F_A - F_B > k\}$$

Détermination de k :

Soit α un niveau fixé : $\alpha = P_{H_0}(D) = P_{H_0}(F_A - F_B > k)$.
On a

$$F_A - F_B \Longrightarrow N(p_A - p_B, \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B})$$

Sous H_0

$$F_A - F_B \Longrightarrow N(0, p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B}))$$

Alors

$$Z_{n_A,n_B} = \frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A} + \frac{1}{n_B})}} \Longrightarrow N(0,1)$$

Par suite

$$\alpha = P_{H_0}\left(\frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}} > \frac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}\right)$$

On pose

$$t_\alpha = \frac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}.D\prime o\grave{u}\ \alpha = P_{H_0}\left(\frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}} > t_\alpha\right)$$

Donc $P_{H_0}(Z > t_\alpha) = \alpha$ et $F_Z(t_\alpha) = 1 - \alpha$ avec Z$\sim N(0,1)$.

Et par la table de la loi normale centrée et réduite on détermine t_α. comme $t_\alpha = \dfrac{k}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}$

Alors

$$k = t_\alpha \sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}$$

La région de rejet D devient :

$$\begin{aligned} D &= \{(X_1,...,X_{n_A},Y_1,...Y_{n_B})/F_A - F_B > k\} \\ D &= \{(X_1,...,X_{n_A},Y_1,...Y_{n_B})/\frac{F_A - F_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}} > t_\alpha\} \end{aligned}$$

Règles de décision On définit les régles de décision suivantes :

Si Z_{n_A,n_B}>t_α on rejette H_0

Si $Z_{n_A,n_B} \leq t_\alpha$ on accepte H_0

Mise en oeuvre du test

On calcule le nombre z=$\dfrac{f_A - f_B}{\sqrt{p_A(1-p_A)(\frac{1}{n_A}+\frac{1}{n_B})}}$

et on le compare à t_α .On utilise les règles de décision pour conclure.

Calcul de la fonction puissance du test D'abord on calcule le risque de deuxième espèce

$$1 - \beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(F_A - F_B < k)$$

Sous $H_1(p_A > p_B)$

$$F_A - F_B \Longrightarrow N(p_A - p_B, \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B})$$

$$Z^*_{n_A,n_B} = \frac{F_A - F_B - (p_A - p_B)}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}} \Longrightarrow N(0,1)$$

On pose

$$\lambda' = \frac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$$

$$1 - \beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(Z \leq t_\alpha - \lambda')$$

D'où

$$1 - \beta_\theta \ = P_{H_1}(Z \leq t_\alpha - \lambda')$$

avec Z$\sim$ $N(0,1)$,

$$1 - \beta_\theta = F_Z(t_\alpha - \lambda').$$

Par suite

$$\beta_\theta = 1 - F_Z(t_\alpha - \lambda').$$

On pose $u' = t_\alpha - \lambda'$
Et donc

$$\beta = \begin{cases} 1 - F_Z(u') & \text{si} \ \ u' > 0 \\ F_Z(|u'|) & \text{si} \ \ u' \leq 0 \end{cases}$$

Exemple 3.Efficacité d'un médicament Un médicament a été expéri-

menté sur 200 malades divisés en deux groupes M$_A$ et M$_B$ choisis indépendament

Le groupe M$_A$ composé de 110 malades a absorbé le médicament étudié.

Le groupe M$_B$ composé de 90 malades a absorbé un placebo (produit d'apparence identique au médicament mais inactif).

Les résultats sont les suivants : 66 malades guéris dans le groupe M$_A$,36 guéris dans le groupe M$_B$.

On voudrait :

1) Calculer le pourcentage de guérison pour chacun des échantillons M$_A$ et M$_B$.

2)En admettant que le phénomène étudié suit une loi normale , construire un test unilatéral permetant d'accepter ou de rejeter l'hypothèse de l'efficacité du médicament ,au risque 5%

Résolution :

1) Dans un échantillon de taille n extrait d'une population P où la fréquence de guérison suit la loi normale N(p,$\sqrt{\frac{p(1-p)}{n}}$) lorsque n $\geq$ 30 (th central limite)

Si p est inconnu, on l'estime par f, fréquence de guérison dans l'échantillon.
Dans M_A, $f_A = \dfrac{66}{100} = 0,6$,dans M_B, $f_B = \dfrac{36}{90} = 0,4$.
2)On construit un test d'homogénéité permettant de comparer les fréquences f_1 et f_2 .

Hypothèses à tester :
H_0 :‹‹ Le médicament est inefficace›› contre H_1 :‹‹ Le médicament est efficace››
Autrement dit H_0 : ‹‹ f_A=f_B›› contre H_1 : ‹‹ $f_A > f_B$››
C'est un test unilatéral.
Région de rejet : notons $F_A = F_{M_A}$, $F_B = F_{M_B}$.

$$D = \{(X_{1_{M_1}}, ..., X_{n_{M_1}}, X_{1_{M_2}}, ..., X_{n_{M_2}})/F_A - F_B > k\}$$

Sous l'hypothèse H_0,la variable aléatoire

$$F_A - F_B \Longrightarrow N(0, \sqrt{p(1-p)(\frac{1}{n_A} + \frac{1}{n_B})})$$

Ici p est inconnu nous estimons p par p̂:

$$\hat{p} = \frac{n_A f_1 + n_B f_2}{n_A + n_B} = \frac{102}{200} = 0,51.$$

Alors, sous l'hypothèse H_0,

$$F_A - F_B \Longrightarrow N(0, \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A} + \frac{1}{n_B})}) = N(0, 0, 71)$$

La va

$$T = \frac{F_A - F_B}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A} + \frac{1}{n_B})}} \Longrightarrow N(0,1)$$

La région de rejet devient

$$D = \{(X_{1_{M_1}}, ..., X_{n_{M_1}}, X_{1_{M_2}}, ..., X_{n_{M_2}})/\frac{F_A - F_B}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A} + \frac{1}{n_B})}} > k\}$$

On pose $t_\alpha = \dfrac{k}{\sqrt{p(1-p)(\frac{1}{n_A} + \frac{1}{n_B})}}$.

D'où

$$D = \{(X_{1_{M_1}}, ..., X_{n_{M_1}}, X_{1_{M_2}}, ..., X_{n_{M_2}})/T > t_\alpha \}$$

Au risque 5%, on a P$(T > t_\alpha) = 0,05$ ie P$(T \leq t_\alpha) = 0,95 \Rightarrow F_T(t_\alpha) = 0,95$.
Par la table de la loi centrée réduite on obtient $t_\alpha = 1,645$.
La région de rejet est :D=$\{$(X$_{1_{M_1}}, ..., X_{n_{M_1}}, X_{1_{M_2}}, ..., X_{n_{M_2}})/T > 1,645 \}$
On a

$$t = \frac{f_A - f_B}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A} + \frac{1}{n_B})}}$$

$$t = \frac{0,6 - 0,4}{\sqrt{0,51 * 0,49 * (\frac{1}{110} + \frac{1}{90})}} = 2,82$$

Donc T=2,82>1,645 et donc nous rejetons H$_0$.
Conclusion :
Nous acceptons l'hypothèse H$_1$:‹‹ Le médicament est efficace›› avec le risque 5%

Puissance:
On a d'aprés les données précédentes:

$$t_\alpha = 1,64 et\ \lambda' = 2,82$$

.

$$u' = t_\alpha - \lambda' = 1,64 - 2,82 = -1,18 < 0$$

La formule générale est :

$$\beta = \begin{cases} 1 - F_Z(u') & \text{si } u' > 0 \\ F_Z(|u'|) & \text{si } u' \leq 0 \end{cases}$$

Alors

$$\beta = F_Z(|u'|) = F_Z(|-1,18|) = F_Z(1,18) = 0,88$$

Conclusion :On a obtenu une trés bonne puissance ,notre test est puissant.

Chapter 2

Test de conformité de moyennes

2.1 Définitions

Soit C un caractère de moyenne m et d'écart type σ connu.On veut comparer sa moyenne m à une valeur μ_0 fixée. Pour cela on considère un échantillon $(X_1, ...X_n)$ de C de taille n.

On sait que chaque X_i est de moyenne m et d'écart type σ.

$\overline{X_n}$ désigne la moyenne empirique de la variable X:

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n_1} X_i$$

$$E(\overline{X}_n) = E(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} * n * m = m$$

.

$$var(\overline{X}_n) = var(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n^2} var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}.$$

L'écart type est :

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

pour n> 30 nous utilisons le théorème central limite et pour σ connu :

$$Z_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sigma(\overline{X_n})} \Longrightarrow N(0, 1)$$

et encore:

$$\overline{X_n} \Longrightarrow N(m, \frac{\sigma^2}{n})$$

Pour μ_0 une valeur numérique fixée ,$\overline{X_n}-\mu_0$ est une gaussienne, déterminons sa moyenne et sa variance:

$$E(\overline{X_n} - \mu_0) = E(\overline{X}) - E(\mu_0) = m - \mu_0.$$

$$var(\overline{X}_n - \mu_0) = var(\overline{X}) + var(\mu_0) = \frac{\sigma^2}{n}.$$

Donc:

$$\overline{X}_n - \mu_0 \Longrightarrow N(m - \mu_0, \frac{\sigma^2}{n})$$

et pour n$\geq$ 30

$$Z_n = \frac{\overline{X}_n - \mu_0 - (m - \mu_0)}{\sqrt{\frac{\sigma^2}{n}}} \Longrightarrow N(0,1)$$

2.2 Test d'hypothèse bilatéral

Hypothèses soit le test:
$H_0 : m = \mu_0$
$H_1 : m \neq \mu_0$

Expression du test on considère la fonction test ℓ suivante

$$\ell(X_1, ..., X_n) = \begin{cases} 1 & \text{si} \quad |\overline{X_n} - \mu_0 \mid > k \\ 0 & \text{si} \quad |\overline{X_n} - \mu_0 \mid \leq k \end{cases}$$

Région de rejet La région de rejet D est définie par :
D=$\{(X_1, ..., X_n)/|\overline{X_n} - \mu_0 \mid$>k$\}$
Le principe du test est le suivant : si $\ell(X_1, ..., X_n) = 1 \Leftrightarrow |\overline{X}_n - \mu_0 \mid$>k ; on rejette H_0 sinon on accepte H_0.

– **Détermination de k:**

Soit α un risque fixé $\alpha = P_{H_0}$(D)=$P_{H_0}(|\overline{X}_n - \mu_0 \mid$>k)
D'aprés ce qui précéde on a :

$$\overline{X}_n - \mu_0 \Longrightarrow N(m - \mu_0, \frac{\sigma^2}{n})$$

Sous H_0 nous avons :

$$\overline{X}_n - \mu_0 \Longrightarrow N(0, \frac{\sigma^2}{n})$$

D'où

$$Z_n = \frac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \Longrightarrow N(0,1)$$

$$\alpha = P_{H_0}(D) = P_{H_0}(|\overline{X}_n - \mu_0| > k) = P_{H_0}(\sqrt{n}\frac{\overline{|X}_n - \mu_0|}{\sigma} > \sqrt{n}\frac{k}{\sigma})$$

$$\alpha = P_{H_0}(|Z_n| > \sqrt{n}\frac{k}{\sigma}) \ .$$

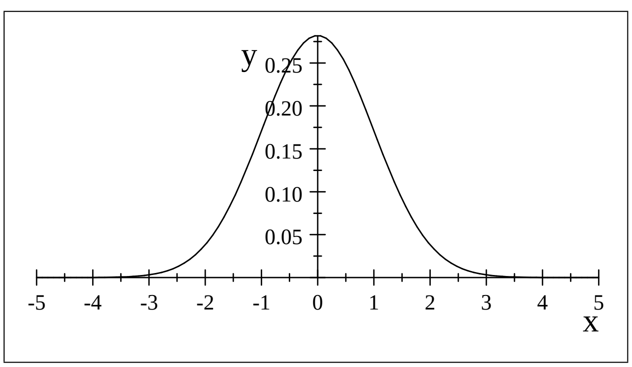
on pose :

$$t_{\alpha/2} = \sqrt{n}\frac{k}{\sigma},$$

On pose Z$\sim N(0,1)$
Par suite

$$1 - \alpha = P_{H_0}(|Z| < \sqrt{n}\frac{k}{\sigma}) = P_{H_0}(-t_{\alpha/2} < Z < t_{\alpha/2})$$

On obtient le graphe suivant :



D'où

$$\frac{\alpha}{2} = P_{H_0}(Z > t_{\alpha/2}) \Rightarrow F_Z(t_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

On détermine $t_{\alpha/2}$ à partir de la table de la loi centrée réduite
par suite

$$k = t_{\alpha/2} * \sqrt{\frac{\sigma^2}{n}}$$

et donc la région de rejet devient:

$$D = \{(X_1, ..., X_n, Y_1, ...Y_n)/|\overline{X}_n - \mu_0| \ > t_{\alpha/2} * \sqrt{\frac{\sigma^2}{n}}\}$$

$$D = \{(X_1, ..., X_n, Y_1, ...Y_n)/\sqrt{n}\frac{|\overline{X}_n - \mu_0|}{\sigma} > t_{\alpha/2}\}$$

$$D = \{(X_1, ..., X_n, Y_1, ...Y_n)/|Z_n| > t_{\alpha/2}\}$$

Règles de décision -Si $|Z_n| > t_{\alpha/2}$,on rejette H_0.
-Si $|Z_n| < t_{\alpha/2}$, on accepte H_0.
Mise en oeuvre du test:
on calcule le nombre z=$\sqrt{n}\frac{|\overline{x}_n - \mu_0|}{\sigma}$ et nous utilisons les règles de décision précédentes pour conclure .

Calcul de la fonction puissance du test Nous commençons par calculer le risque de deuxième espèce:

$$1-\beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(\sqrt{n}\frac{|\overline{X}_n - \mu_0|}{\sigma} \leq t_{\alpha/2}) = P_{H_1}(-t_{\alpha/2} \leq \sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} \leq t_{\alpha/2})$$

Sous $H_1 : \mu_1 \neq \mu_2$;nous avons d'aprés le théorème central limite:

$$Z_n^+ = \frac{\overline{X}_n - \mu_0 - (m - \mu_0)}{\sqrt{\frac{\sigma^2}{n}}} \Longrightarrow N(0,1)$$

Notons $\lambda = \frac{m - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$
Donc:

$$1 - \beta_\theta = P_{H_1}(-t_{\alpha/2} - \lambda \leq Z_n^+ \leq t_{\alpha/2} - \lambda)$$

Si $Z \sim N(0,1)$
Alors

$$1 - \beta_\theta = F_Z(t_{\alpha/2} - \lambda) - F_Z(-t_{\alpha/2} - \lambda)$$

D'où la fonction puissance

$$\beta(\delta) = 1 - F_Z(t_{\alpha/2} - \lambda) + F_Z(-t_{\alpha/2} - \lambda)$$

On pose
u'=$-t_{\alpha/2} - \lambda$, u"=$t_{\alpha/2} - \lambda$
La fonction puissance devient :

$$\beta(\delta) = 1 + F_Z(u') \ - F_Z(u'')$$

Si u'<0 $\Rightarrow F_Z(u') = 1 - F_Z(|u'|)$
Si u"<0$\Rightarrow F_Z(u'') = 1 - F_Z(|u''|)$
On conclut que : si u'<0:

$$\beta(\delta) = \begin{cases} 2 - F_Z(|u'|) - F_Z(|u''|) & si \quad u" > 0 \\ 1 - F_Z(|u'|) + F_Z(|u''|) & si \quad u" \leq 0 \end{cases}$$

et si u'>0:

$$\beta(\delta) = \begin{cases} 1 + F_Z(|u'|) - F_Z(|u''|) & si \;\; u" > 0 \\ F_Z(|u'|) + F_Z(|u''|) & si \;\; u" \leq 0 \end{cases}$$

Exemple 4 Pour un caractère C de loi normale on a un échantillon de 7 éléments.

La moyenne de cet échantillon est 257,9 et l'écart type est supposé connu et égal à14.

Est il possible de conclure ,au seuil 5% que la moyenne de la population est différente de 250?

Résolution :

On a un échantillon de taile 7 : (X_1,...,X_7) avec $X_i \sim N(257.5, 14^2)$.
On prend $\mu_0 = 250$ et $\alpha = 0,05$.
Soit le test d'hypothèse suivant :
H_0 :m=μ_0 contre $H_1 : m \neq \mu_0$.
On considère la fonction test suivante :

$$\ell(X_1, ..., X_7) = \begin{cases} 1 \;\text{ si } |\overline{X}_7 - 250| > k \\ 0 \;\text{ si } |\overline{X}_7 - 250| \leq k \end{cases}$$

D'où

$$D = \{(X_1, ..., X_7)/|\overline{X}_7 - 250| > k\}.$$

Or $\overline{X}_7 - 250 \sim N(m - 250, \frac{14^2}{7})$
Donc :

$$\alpha = P_{H_0}(\frac{|\overline{X}_7 - 250|}{\frac{14}{\sqrt{7}}} > \frac{k}{\frac{14}{\sqrt{7}}}) = P_{H_0}(\sqrt{7}\frac{|\overline{X}_7 - 250|}{14} > \sqrt{7}\frac{k}{14})$$

On pose

$$Z = \sqrt{7}\frac{|\overline{X}_7 - 250|}{14}$$

$$F_Z(\sqrt{7}\frac{k}{14}) = 1 - \frac{\alpha}{2} = 0,975$$

Par la table centée réduite nous obtenons : $\dfrac{k}{14} = \dfrac{1,96}{\sqrt{7}}$,par suite k=10,37

Donc

D=$\{(X_1, ..., X_7)/\ |\overline{X}_7 - 250| > 10,37\}$.

On a |m-250|=|257,5-250|=7,9<10,37 d'où on accepte H_0.

Conclusion:

Au seuil 5% on conclut que la moyenne de la population est différente de 250.

Puissance du test :

$\lambda = \frac{7,9}{14}\sqrt{7} = 1,492;$

$-t_\alpha - \lambda = -1,96 - 1,492 = -3,452;$

$t_\alpha - \lambda = 0,468.$

$$\beta_\theta = 2 - F_Z(|-3,45|) - F_Z(|0,47)\ avecZ \sim N(0,1)$$

$$\beta_\theta = 0,32.$$

2.3 Test d'hypothèse unilatèral

Hypothèses

on veut tester:

H_0 : $m = \mu_0$

contre

H_1 : $m > \mu_0$

Expression du test

on considère la fonction test notée ℓ

$$\ell(X_1, ..., X_n) = \begin{cases} 1 & si\ \overline{X_n} - \mu_0 > k \\ 0 & si\ \overline{X_n} - \mu_0 \leqslant k \end{cases}$$

Région de rejet

La région de rejet D est définie par :

D=$\{(X_1, ..., X_n)/\overline{X_n} - \mu_0$ >k}

– Détermination de k:

Soit α un risque fixé $\alpha = P_{H_0}$(D)=$P_{H_0}(\overline{X}_n - \mu_0 > k\ )$
on a:

$$\overline{X}_n - \mu_0 \Longrightarrow N(m - \mu_0, \frac{\sigma^2}{n})$$

Sous H_0 nous avons :

$$\overline{X}_n - \mu_0 \Longrightarrow N(0, \frac{\sigma^2}{n})$$

D'où pour n assez grand

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

$$\alpha = P_{H_0}(D) = P_{H_0}(\overline{X}_n - \mu_0 > k) = P_{H_0}(\sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} > \sqrt{n}\frac{k}{\sigma})$$

$$\alpha = P_{H_0}(Z_n > \sqrt{n}\frac{k}{\sigma})\ .$$

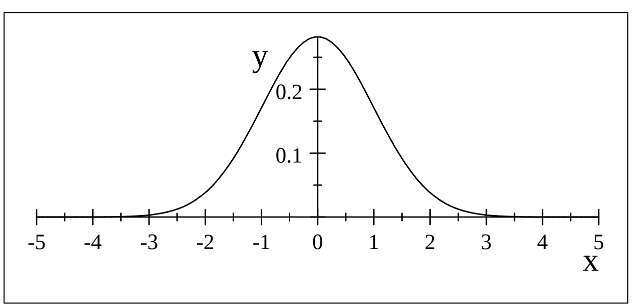
on pose :

$$t_\alpha = \sqrt{n}\frac{k}{\sigma},$$

Si $Z \sim N(0,1)$
Alors

$$1 - \alpha = P_{H_0}(Z < \sqrt{n}\frac{k}{\sigma}) = P_{H_0}(Z < t_\alpha) = F_Z(t_\alpha)$$

On obtient le graphe suivant :



On détermine t_α à partir de la table de la loi centrée réduite
et par suite

$$k = t_\alpha * \frac{\sigma}{\sqrt{n}}$$

et donc la région de rejet devient:

$$D=\{(X_1,...,X_{n_1},Y_1,...Y_{n_2})/\overline{X}_n-\mu_0 \;\; > t_\alpha * \frac{\sigma}{\sqrt{n}}\}$$

$$D=\{(X_1,...,X_{n_1},Y_1,...Y_{n_2})/\sqrt{n}\frac{\overline{X}_n-\mu_0}{\sigma}>t_\alpha\}$$

Règles de décision

-Si $Z_n > t_\alpha$,on rejette H_0.
-Si $Z_n < t_\alpha$, on accépte H_0.

Mise en oeuvre du test:

on calcule le nombre z= $\frac{|\overline{x}_n-\mu_0|}{\sqrt{\frac{\sigma^2}{n}}}$ et nous utilisons les règles de décision précédentes pour conclure .

Calcul de la fonction puissance du test

Nous commençons par calculer le risque de deuxième espèce:

$$1-\beta_\theta=P_{H_1}(\overline{D})=P_{H_1}(\frac{\overline{X}_n-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\leq t_\alpha)=P_{H_1}(\frac{\overline{X}_n-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\leq t_{\alpha/2})$$

Sous H_1 : $m\neq\mu_0$;nous avons d'aprés le théorème central limite:

$$Z_n^+=\frac{\overline{X}_n-\mu_0-(m-\mu_0)}{\sqrt{\frac{\sigma^2}{n}}}\Longrightarrow N(0,1)$$

Notons $\lambda=\frac{m-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}$

Donc:

$$1-\beta_\theta=P_{H_1}(Z_n^+\leq t_\alpha-\lambda)$$

Si $Z\sim N(0,1)$
Alors

$$1-\beta_\theta=F_Z(t_\alpha-\lambda)$$

D'où la fonction puissance:

$$\beta(\delta)=1-F_Z(t_\alpha-\lambda)$$

$$\beta(\delta)=\begin{cases}1-F_Z(t_\alpha-\lambda) & \text{si } t_\alpha-\lambda>0\\ F_Z(|t_\alpha-\lambda|) & \text{si } t_\alpha-\lambda\leq 0\end{cases}$$

Remarque

Dans le test unilatéral précédent on a considéré les hypothèses : $H_0 : m = \mu_0$ contre $H_1 : m > \mu_0$

On pouvait avoir :

$H_0 : m = \mu_0$ contre $H_1 : m < \mu_0$

On considère le test suivant :

$$\ell(X_1, ..., X_n) = \begin{cases} 1 & si\ \overline{X_n} - \mu_0 < -k \\ 0 & si\ \overline{X_n} - \mu_0 \geq -k \end{cases}$$

Région de rejet La région de rejet D est définie par :

$$D = \{(X_1, ..., X_n)/\overline{X_n} - \mu_0 < -k\}$$

– Détermination de k:

Soit α un risque fixé

$$\alpha = P_{H_0}(D) = P_{H_0}(\overline{X}_n - \mu_0 < -k)$$

on a:

$$\overline{X}_n - \mu_0 \Longrightarrow N(m - \mu_0, \frac{\sigma^2}{n})$$

Sous H_0 nous avons :

$$\overline{X}_n - \mu_0 \Longrightarrow N(0, \frac{\sigma^2}{n})$$

Pour n assez grand

$$Z_n = \frac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \Longrightarrow N(0, 1)$$

$$\alpha = P_{H_0}(D) = P_{H_0}(\overline{X}_n - \mu_0 < -k) = P_{H_0}(\sqrt{n}\frac{X_n - \mu_0}{\sigma} < -\frac{\sqrt{n}k}{\sigma})$$

on pose :

$$t_\alpha = \frac{\sqrt{n}k}{\sigma},$$

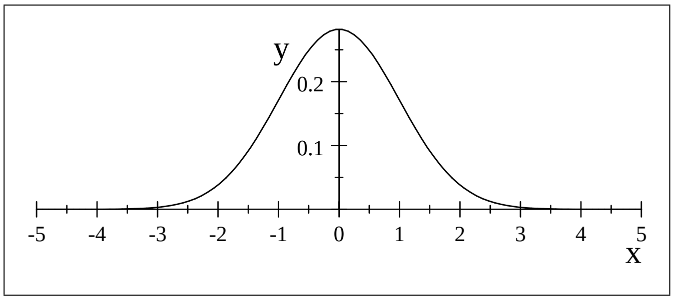
Si Z$\sim N(0, 1)$

$$\alpha = P_{H_0}(Z < -\frac{\sqrt{n}k}{\sigma}) .$$

Par suite

$$\alpha = P_{H_0}(Z < -t_\alpha) = F_Z(-t_\alpha)$$

On obtient le graphe suivant :



On détérmine t_α à partir de la table de la loi centrée réduite
et par suite

$$k = -t_\alpha * \frac{\sigma}{\sqrt{n}}$$

et donc la région de rejet devient:

$$D = \{(X_1, ..., X_n) / \frac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} < -t_\alpha\}$$

Règles de décision

-Si Z_n <-t_α,on rejette H_0.

-Si $Z_n \geq -t_\alpha$, on accépte H_0.

Mise en oeuvre du test:

on calcule le nombre z=$\frac{|\overline{x}_n - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}}$ et nous utilisons les règles de décision précédentes pour conclure .

Calcul de la fonction puissance du test

Nous commençons par calculer le risque de deuxième espèce :

$$1 - \beta_\theta = P_{H_1}(\overline{D}) = P_{H_1}(\frac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > -t_\alpha)$$

Sous H_1 : $m \neq \mu_0$;nous avons d'aprés le théorème central limite:

$$Z_n^+ = \frac{\overline{X}_n - \mu_0 - (m - \mu_0)}{\sqrt{\frac{\sigma^2}{n}}} \Longrightarrow N(0,1)$$

Notons $\lambda = \frac{m - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$

Donc:

$$1-\beta_\theta = P_{H_1}(Z_n^+ > -t_\alpha - \lambda)$$

On pose Z$\sim N(0,1)$

Donc la fonction puissance est :

$$\beta(\delta) = F_Z(-t_\alpha - \lambda)$$

$$\beta(\delta) = \begin{cases} F_Z(-t_\alpha - \lambda) & \text{si -t}_\alpha - \lambda > 0 \\ 1 - F_Z(|-t_\alpha - \lambda|) & \text{si -t}_\alpha - \lambda < 0 \end{cases}$$

Exemple :

Dans un concours d'examen de copies, on pratique une double correction des copies :

deux correcteurs A et B corrigent de façon indépendante les copies.Ainsi nous obtenons pour un échantillon de 50 copies les résultats suivant :

A	B	A	B	A	B	A	B	A	B
17	18	17	14	7	6	2	4	7	8
12	12	18	18	4	6	18	16	16	15
9	6	4	5	12	14	6	7	10	12
11	14	6	4	14	15	13	13	15	17
10	11	15	17	10	10	11	12	9	10
8	9	19	18	9	7	7	6	7	7
6	5	9	12	12	13	17	15	12	11
14	15	10	12	15	17	8	10	14	13
15	17	10	13	7	6	4	4	8	11
3	3	8	6	5	6	12	13	16	17

Question : est ce que les correcteurs A et B corrigent par le même barème ou bien le correcteur A est plus sévère que le correcteur B dans les notes ?

On associe à chaque copie i un couple de notes (X$_i, Y_i$) et i=1,...,50.

On effectue pour chaque couple de notes (X$_i, Y_i$) la différence d$_i = X_i - Y_i$ de ces deux notes de A et B.

On obtient ainsi une nouvelle série statistique :

d_i	-3	-2	-1	0	1	2	3
n$_i$	4	11	14	7	8	5	1

Dont la moyenne $\overline{d} = -0,54$ et l'écart type=1,54.

Si les examinateurs notent de la même façon ,la moyenne de cette série statistique doit être =0

Si l'un note plus sévèrement que l'autre $\overline{d}$ doit être soit supèrieur ou infèrieur à 0

si A est plus sévère que B ie $\overline{d} < 0$

$H_0 : \overline{d} = 0$ contre $H_1 : \overline{d} < 0$

Utilisons le test de conformité de moyenne d'hypothèse unilatéral :

On calcule z=$\dfrac{\overline{d} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{-0,54}{0,218} = -2,48.$

P(Z<-t_α) $= 0,01$

Au risque 1% à l'aide de la table de la loi centrée réduite on obtient $t_\alpha = 2,33.Or$ z=$-2,48. < -2,33$

donc on rejette H_0

Conclusion:

Le correcteur A note plus sévèrement que le correcteur B

Calcul de la fonction puissance:

$\beta(\delta) = F_Z(-t_\alpha - \lambda)$

Avec $\lambda = \dfrac{m - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ et Z $\sim N(0,1)$

$\lambda = \dfrac{m - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} = \dfrac{-0,54}{\sqrt{\frac{1,54}{50}}} = -2,47$

-$t_\alpha - \lambda = -2,33 + 2,48 = 0,15$

$\beta = F_Z(0,15) = 0,556.$

Chapter 3

Conclusion

Un test d'hypothèse est un procédé d'inférence statistique permettant de contrôler (accepter ou rejeter) à partir de l'étude d'un ou de plusieurs échantillons aléatoires la validité des hypothèses relatives à une ou plusieurs populations.

Les tests d'hypothèses font appel à un certain nombre d'hypothèses concernant la nature de la population dont provient l'échantillon étudié (égalité des moyennes ,égalité des fréquences,tests de conformité).

RESUME

Il serait vain de chercher à présenter l'ensemble des tests statistiques,la littérature est trés abondante sur le sujet .Cette vignette introduit les plus couramment calculés par les logiciels statistiques standards (SAS,Minitab,R).Sont concernés les tests à un échatillon,de comparaison de deux échantillons:leurs moyennes,fréquence et enfin les tests de conformités des moyennes.

Mots clés:TCL,va,iid,N(0,1),χ^2_n,T_n

Chapter 4

Bibliographie

1/ D MOUCHIROUD . Outils pour les Mathématiques
2/N KARA TERKI .Tests d'égalité de moyenne
3/ JEAN-PIERRE LECOUTRE .Statistique et probabilité
4/ WIKI-STAT .Tests statistiques élémentaires.