Faculté des Sciences Juridiques
Economiques & Sociales –Fes-

D.TOUIJAR
2010-11

LES TESTS D'HYPOTHESE

SEMESTRE 5

Sections **C** & **D**

Attention !

N'essayez pas de comprendre le cours en lisant tout (e) seul(e) Ce document.

Par contre, je vous recommande vivement d'assister à toutes les séances en espérant mieux cerner le programme de statistique IV

INTRODUCTION GENERALE

- Il y a en statistique deux approches: La description et L’inférence.
- <u>L’inférence statistique</u> , rappelons le, a pour but d’étendre les propriétés de l’échantillon à la population entière et de valider ou de rejeter des hypothèses a priori ou formulées après une étape descriptive.

Les deux principaux champs d'étude de l'**inférence statistique** sont :

- L'estimation de paramètres : déjà traité en S4 (statistique III).

- Les Tests statistiques : qui font l'objet du programme de ce semestre **5**.

PROGRAMME DE CE SEMESTRE

- **Chapitre 1: Les Tests paramétriques**

- **Chapitre 2: Les Tests non paramétriques**

BIBLIOGRAPHIE

Titre	Auteurs	Code
Handbook of PARAMETRIC and NONPARAMETRIC STATISTICAL PROCEDURES	DJ. Sheskin	
Méthodes statistiques	B. Grais	stat22
Introduction à la statistique	J.P Bélisle ;J.Desrosiers	stat20
Eléments de statistiques	J. Desrosiers	stat25

Chapitre 1

Les Tests paramétriques

Introduction

– Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire relatif à la V.A. parente X de loi $\mathscr{L}(\theta)$, où $\theta \in \Theta$ est un paramètre réel inconnu.

– Le semestre précédent, on cherchait à estimer θ. Mais il arrive qu'on ait une idée préconçue sur sa valeur: $\theta = \theta_0$

– On désire alors tester la validité de cette hypothèse, en la confrontant à une hypothèse alternative.

– Cette dernière exprime une tendance différente au sujet du paramètre.

– **Exemple :**

- Est-ce que le taux de chômage au Maroc est p_0 ?
- Est-ce que l’espérance de vie au Maroc est μ_0 ?...ou a augmenté ?

Méthodologie du test d'hypothèse

–On suppose que Θ est partitionné en Θ_0 et Θ_1:

$$\Theta = \Theta_0 \bigcup \Theta_1 \quad et \quad \Theta_0 \bigcap \Theta_1 = \emptyset$$

– Exprimons le fait que $\theta \in \Theta_0$ par l'hypothèse $\boldsymbol{H_0}$

et le fait que $\theta \in \Theta_1$ par $\boldsymbol{H_1}$

- ✓ H_0 : « $\theta \in \Theta_0$ »

- ✓ H_1 : « $\theta \in \Theta_1$ »

Définition: Un test d'hypothèse, est une règle de décision permettant, au vu de la réalisation $(x_1, x_2, \ldots, x_n)$ de l'E.A., de répondre à la question « dans lequel des deux sous ensemble se trouve θ ? »

- H_0 : s'appelle **l'hypothèse nulle**.

- H_1 : s'appelle **l'hypothèse alternative**.

- Si Θ_0 se réduit au seul point θ_0: $\Theta_0 = \{\theta_0\}$

H_0 devient H_0: « $\theta = \theta_0$ » et sera appelée **l'hypothèse simple**.

➢Soit l'hypothèse simple H_0: « $\theta = \theta_0$ »

➢ Si H_1 est telle que H_1: « $\theta > \theta_0$ » ; alors on dit que le test est unilatéral à droite:

$$T.U.D.\begin{cases} H_0 : "\theta = \theta_0" \\ \# \\ H_1 : "\theta > \theta_0" \end{cases}$$

➢ Si H_1 est telle que H_1: « $\theta < \theta_0$ » ; alors on dit que le test est unilatéral à gauche:

$$T.U.G.\begin{cases} H_0 :"\theta = \theta_0" \\ \# \\ H_1 :"\theta < \theta_0" \end{cases}$$

- Si H_1 est telle que H_1: « $\theta \neq \theta_0$ » ;

alors on dit que le test est bilatéral :

$$T.B.\begin{cases} H_0 : "\theta = \theta_0" \\ \# \\ H_1 : "\theta \neq \theta_0" \end{cases}$$

- **Exemple Introductif** :
  - un fabricant de lessive prétend que son produit est efficace dans 75% des cas. Une association de consommateurs désire tester le produit.
  - A notre problème on associe le modèle statistique suivant:

Notons par **X** la V.A. qui donne le résultat de l'utilisation du détachant sur un vêtement:

- X = $\begin{cases} 1 \quad avec \;\; proba = p \quad si\, la\, tâche\, disparait \\ 0 \quad avec \;\; proba = 1 - p \quad si\, la\, tâche\, persiste \end{cases}$

- D'où

- X $\rightsquigarrow \mathscr{B}(\boldsymbol{p})$

- Soit X_1, X_2, …, X_{100} l'E.A. relatif à X

- Alors $F = \dfrac{S_{100}}{100} = \dfrac{X_1 + X_2 + \cdots X_{100}}{100}$ est le

meilleur estimateur de $\boldsymbol{p}$.

- On a alors 2 hypothèses opposées :

- La première est soutenue par le fabricant

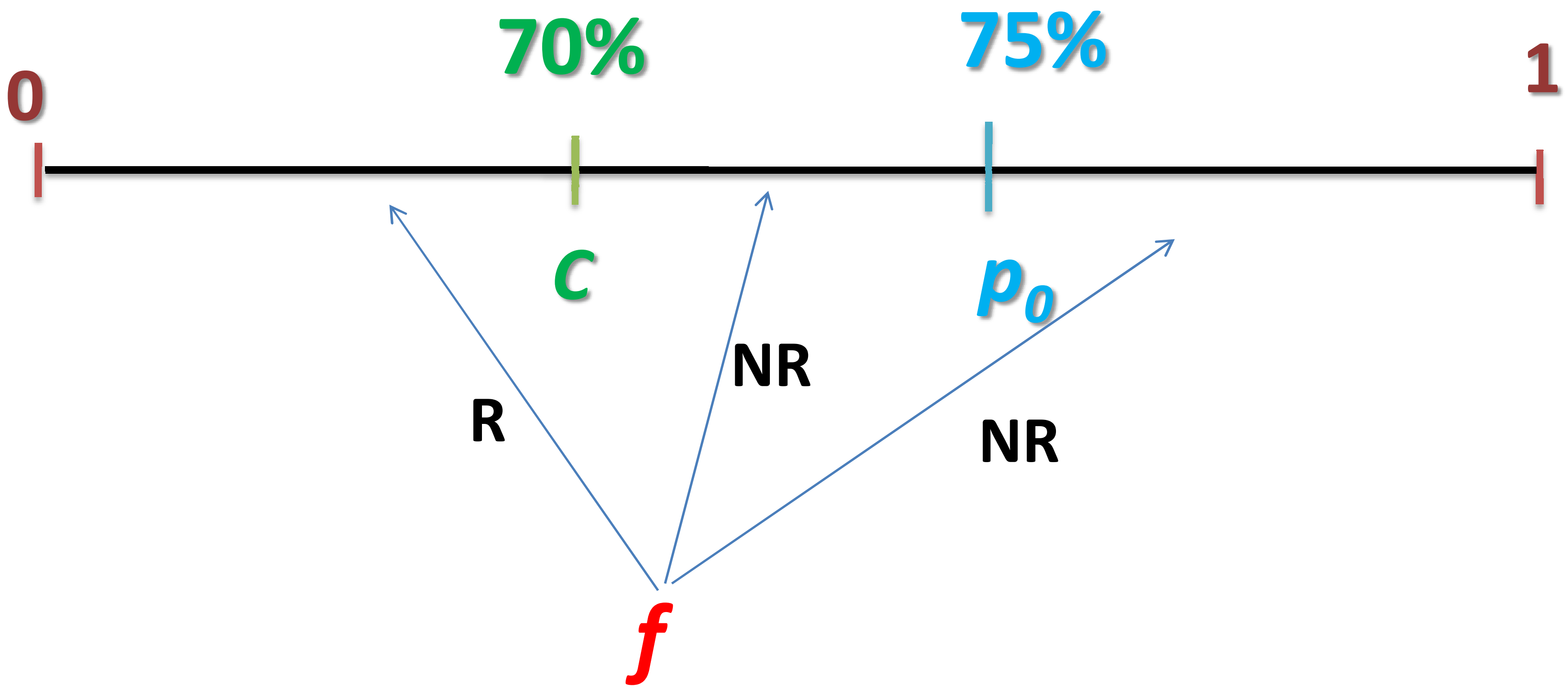
$$H_0 : « p = 0,75 »$$

- Ce que conteste l'association, en lui opposant une hypothèse alternative:

$$H_1 : « p < 0,75 »$$

- Alors avec un peu de <u>bon sens</u>, l'association suit la procédure suivante:

$$\begin{cases} \text{On accepte le produit si } \boldsymbol{f} \geq 70\% \\ \text{donc pas de poursuite judiciaire.} \\ \\ \text{On refuse si } \boldsymbol{f} < 70\% \end{cases}$$



70%
75%
0
1
C
p_0
R
NR
NR
f

- Où $\boldsymbol{f}$ est la proportion des 100 vêtements tachés et testés pour lesquels le produit a été efficace.

- Cette procédure s’appelle « **Règle de décision** ». Et

$c = 70\% = 0{,}70$ s’appelle **valeur critique** du test.

$$\begin{cases} \text{Si } f < c = 0{,}7 \text{ alors } H_0 \text{ est rejetée au profit de } H_1 \\ \text{Si } f \geq c = 0{,}7 \text{ alors } H_0 \text{ n'est pas rejetée.} \end{cases}$$

Remarque :

- $R_0 = [0\ ;\ 0{,}7[$: zone de rejet de H_0

  =Région critique

- $\overline{R}_0 = [0{,}7\ ;\ 1]$: zone de non-rejet de H_0

  =Région d'acceptation

- Cette règle de décision peut conduire à deux types d'erreurs:

- On rejette H_0 alors que H_0 est vraie:

$$\boxed{RH_0/H_0 vraie}$$

- On l'appelle « **erreur de première espèce** »

- On ne rejette pas H_0 alors que H_1 est vraie:

$$\boxed{NRH_0/H_1 vraie;}$$

- c'est « **l'erreur de seconde espèce** ».

- On définit alors la probabilité de commettre l'une ou l'autre erreur:

- 1)-

$$\alpha = P\left(RH_0 / H_0 Vraie\right) = P_0\left(RH_0\right)$$

- C’est le **risque** de première espèce.

- 2)-

$$\beta = P\left(NRH_0 / H_1 Vraie\right) = P_1\left(NRH_0\right)$$

- C’est le **risque** de seconde espèce.

- Définition: On appelle puissance du test la quantité π :

$$\pi = 1 - \beta = P_1(RH_0)$$

- Voici un tableau résumant toutes les situations

Réalité \ Décision	RH_0	NRH_0
H_0 Vraie	Erreur de $1^{ère}$ espèce	Bonne Décision
H_1 Vraie	Bonne Décision	Erreur de $2^{ème}$ espèce

- α est généralement **petit**; c'est le risque que court l'association en lançant, à tort, une action contre le fabricant.

- **Question**: Quel risque court l'association en lançant, à *tort*, une action contre le fabricant ?

- Calculons dans notre cas α :

- Si H_0 est vraie, alors $X \leadsto \mathscr{B}(0,75)$.

- En supposant les cond$^{\theta}$ T.C.L. vérifiées sous H_0 :
- $$Z = \sqrt{n}\,\frac{F - 0{,}75}{\sqrt{0{,}75 \times 0{,}25}} \approx \mathcal{N}(0,\ 1)\,.$$

$$\alpha = P_0\left(F < 0{,}7\right) = P_0\left(Z < 10 \times \frac{0{,}7 - 0{,}75}{0{,}43}\right)$$

$$= P_0\left(Z < -1{,}16\right) = P_0\left(Z > 1{,}16\right) = 12{,}3\%$$

- **Problème:** Peut-on contrôler le risque α, en jouant sur la valeur critique c ?
- **Réponse : Oui**

- **Exemple :** si on pense que prendre un risque de 5% est acceptable, on a :

$$\alpha = 0,05 = P_0\left(RH_0\right) = P_0\left(F < c\right) = P_0\left(Z < \sqrt{n} \times \frac{c - p_0}{\sqrt{p_0 q_0}}\right)$$

$$\Rightarrow \frac{c - 0,75}{0,43} \times 10 = -z_{0,05} = -1,645 \Rightarrow c = p_0 - z_\alpha \sqrt{\frac{p_0 q_0}{n}} = 0,68$$

- A partir du seuil c=68% on rejette $\boldsymbol{H_0}$

Que signifie l'hypothèse nulle?

- \- hypothèse sur la base de laquelle est définie la distribution de probabilité (théorie de la décision):
  - L'hypothèse H_0 entraîne la connaissance de la loi de la *statistique de test* .

Par exemple, si le dé est équilibré, l'hypothèse nulle entraîne que le nombre d'apparition d'un numéro (cinq par exemple) sur n lancés, suit la loi binomiale $\mathscr{B}(n,p_0)$, où p_0 est la probabilité (supposée connue) ici égale à 1/6.

- \- Négation d'une hypothèse de **recherche**.

- \- Spécification d'un paramètre égal à zéro.

L'hypothèse nulle est exacte ($\mu = \mu_0$), l'hypothèse alternative est indéfinie:

  - L'hypothèse nulle H_0 est une hypothèse de non différence [il n'y a pas de différence **significative** entre les échantillons A et B]. Elle est formulée de façon à être rejetée.

- \-La statistique inférentielle procède généralement par le rejet de l'hypothèse nulle.

- \-Un test d'hypothèse constitue donc une sorte de démonstration par l'absurde en probabilité.

- La conclusion finale, une fois que le test a été effectué, est toujours donnée en termes d'hypothèse nulle. Soit qu'on conclu "rejetons H_0 en faveur de H_1" ou qu'on conclu "ne pas rejeter H_0"; nous ne concluons jamais "rejeter H_1 ", ou même "accepter H_1 ".

- Si nous concluons "ne pas rejeter H_0 ", ceci ne signifie pas nécessairement que l'hypothèse nulle est vraie, il suggère seulement qu'il n'y ait pas d'évidence suffisante contre H_0 en faveur de H_1. Rejeter l'hypothèse nulle suggère, alors, que l'hypothèse alternative *peut* être vraie.

ANALYSE UNIVARIEE

Cas d'un seul ECHANTILLON

(Comparaison par rapport à 1 standard (ou fixe))

I- Tests Relatifs à Une Proportion

- **1- Test Unilatéral <u>à gauche</u> pour la proportion**
- *i)* Formulation des hypothèses

$$T.U.G.\begin{cases} H_0 : " p = p_0 " \\ \# \\ H_1 : " p < p_0 " \end{cases}$$

- *ii)* si le T.C.L. est vérifié sous H_0 alors

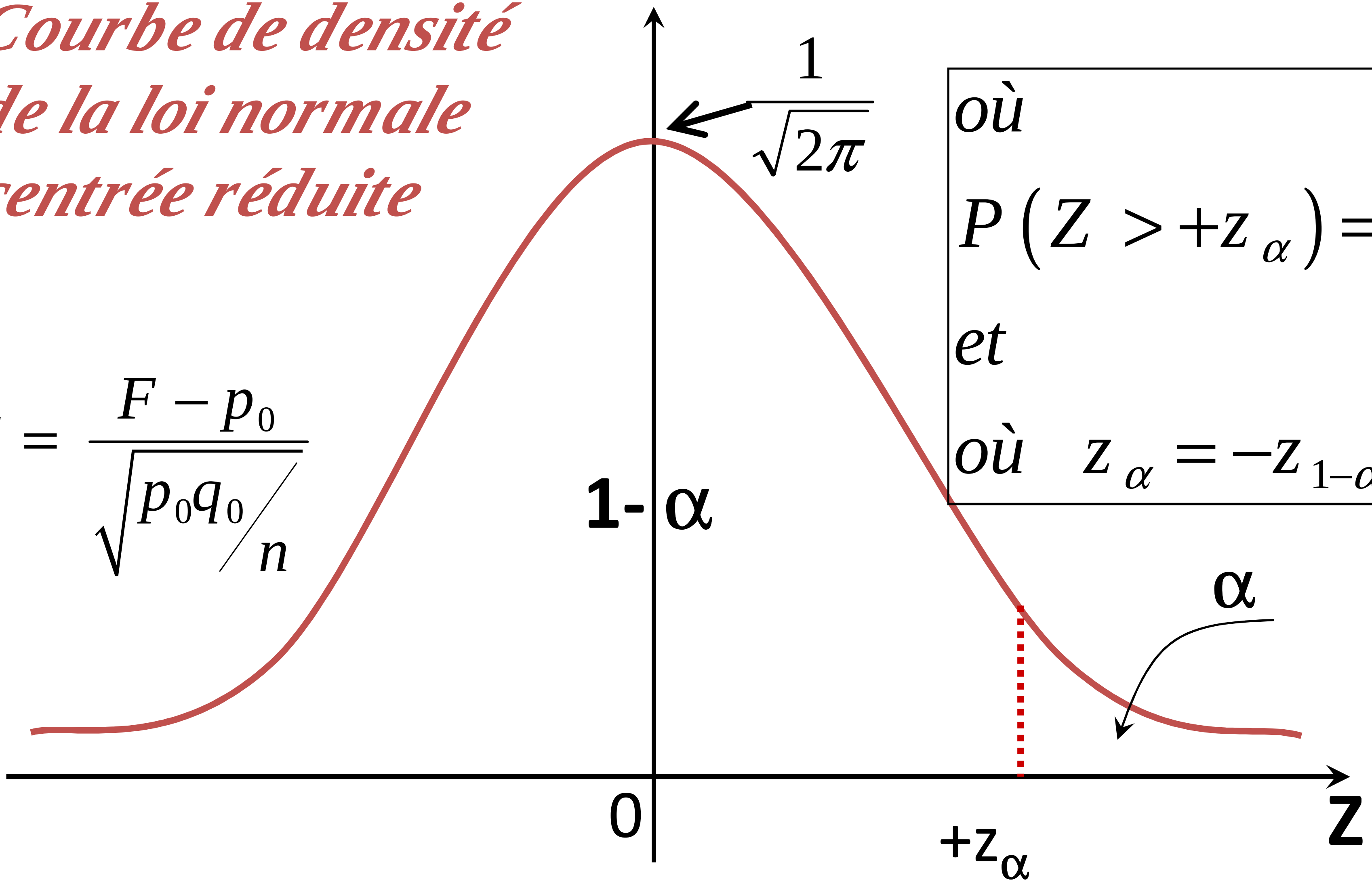
<u>R.D</u>

$$\begin{cases} Si \ \ f < c_\alpha = p_0 - z_\alpha \sqrt{\dfrac{p_0 q_0}{n}} \ \ \text{on rejette } \mathrm{H}_0 \\ Si \ \ f \geq c_\alpha = p_0 - z_\alpha \sqrt{\dfrac{p_0 q_0}{n}} \ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

En effet; commençons d'abord par un petit rappel sur les valeurs critiques :

Courbe de densité de la loi normale centrée réduite

$$Z = \frac{F - p_0}{\sqrt{p_0 q_0 / n}}$$



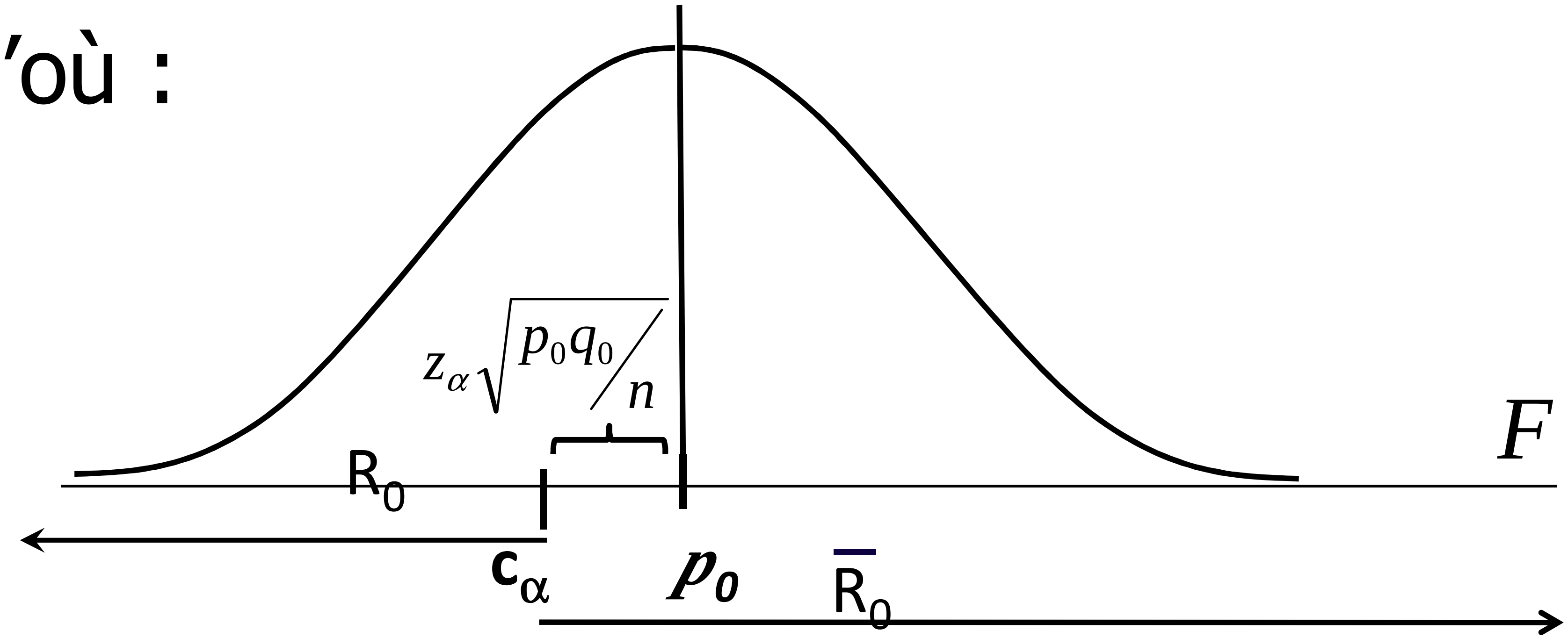
où

$$P\left(Z > +z_\alpha\right) = \alpha$$

et

où $\quad z_\alpha = -z_{1-\alpha}$

D'où :



$$\alpha = P_0\left(RH_0\right) = P_0\left(F < c_\alpha\right) = P_0\left(\frac{F-p_0}{\sqrt{p_0q_0/n}} < \frac{c_\alpha - p_0}{\sqrt{p_0q_0/n}}\right)$$

$$\Rightarrow \frac{c_\alpha - p_0}{\sqrt{p_0q_0/n}} = -z_\alpha \Rightarrow \quad c_\alpha = p_0 - z_\alpha\sqrt{p_0q_0/n}$$

- **Calcul du risque β (pour un risque α fixé):**
  - Puisque l'hypothèse alternative H_1 est composée, alors pour chaque valeur du paramètre p de H_1, on peut calculer un β (respectivement un π). Soit p_1 cette valeur, alors :

$$\beta_1 = P\left(F \geq c_\alpha / p = p_1\right) = P_1\left(F \geq c_\alpha\right)$$

$$\pi_1 = P\left(F < c_\alpha / p = p_1\right) = P_1\left(F < c_\alpha\right)$$

- Revenons à notre exemple et calculons π pour $p = p_1 = 0{,}65$ et α = 5%:

$$\pi_1 = P_1\left(F < c_\alpha\right) = P_1\left(\underbrace{\sqrt{n}\,\frac{F - p_1}{\sqrt{p_1 q_1}}}_{Z_1} < \sqrt{n}\,\frac{c_\alpha - p_1}{\sqrt{p_1 q_1}}\right)$$

- Sous H_1 , Z_1 est normale; or c_α = 0,68 d'où :

$$\pi_1 = P_1\left(Z_1 < 0{,}629\right) = 73{,}5\%$$

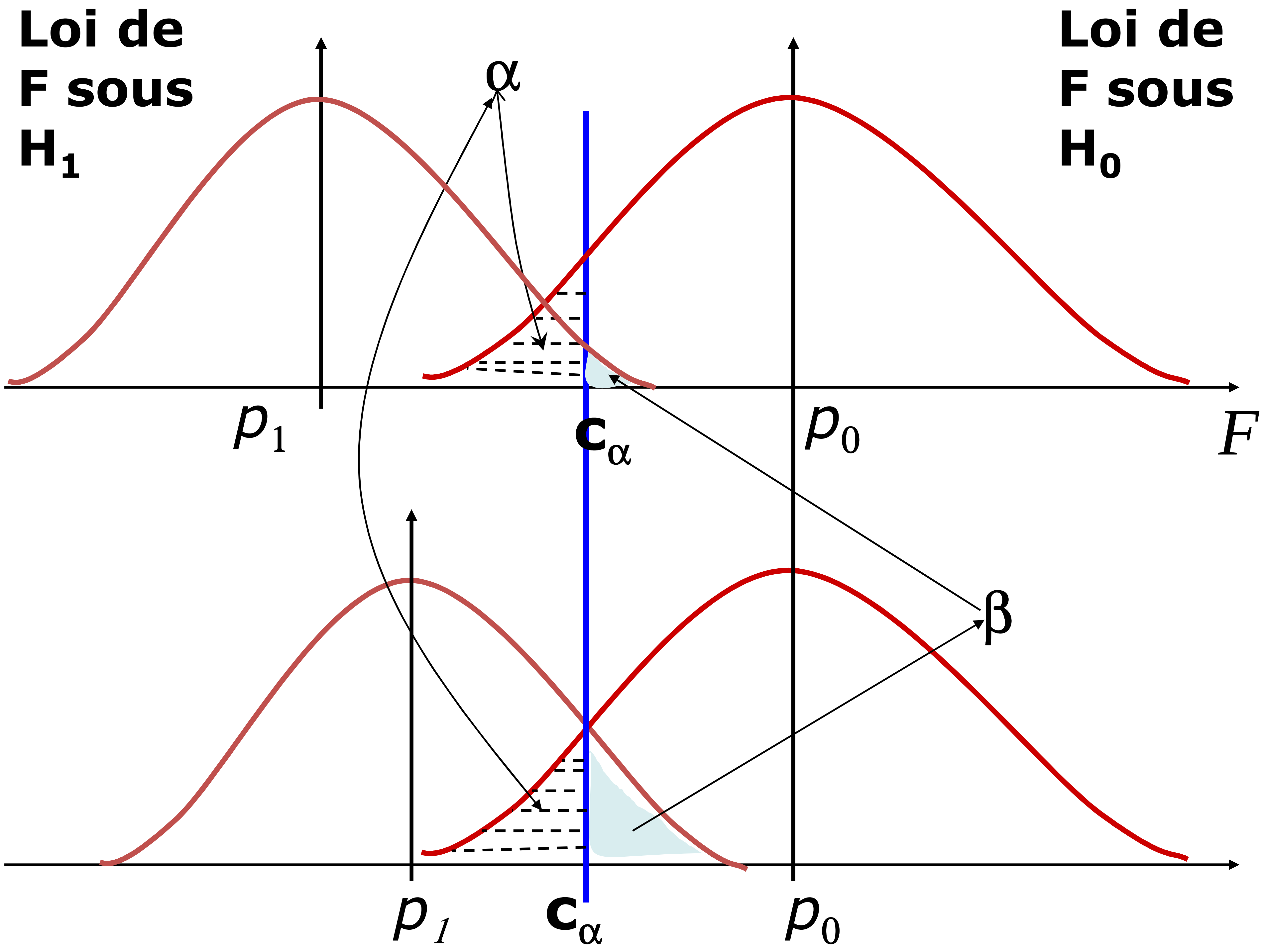
$$\Rightarrow \beta_1 = 26{,}5\%$$

- **Si** $p = p_2 = 0{,}60$:

$$\pi_2 = P_1\left(Z_2 < 1{,}63\right) = 94{,}8\%$$

$$\Rightarrow \beta_2 = 5{,}2\%$$

- **Remarque:** Plus p_1 s'éloigne de p_0 et plus la puissance du test augmente (β↘) et réciproquement. Autrement, lorsque les deux hypothèses H_0 et H_1 sont très différents, la probabilité de commettre l'erreur de second espèce est faible.
- Si α et β sont petits, on dit que le test une bonne règle de décision.


Loi de
F sous
H_1
Loi de
F sous
H_0
α
β
p_1
c_α
p_0
F
p_1
c_α
p_0

- **Remarque :** Les logiciels utilisent souvent

  le niveau de signification α_0 (*p-value*) d'une réalisation f de F : c'est le plus petit α à partir du quel on ne peut plus rejeter H_0 :

$$P_0\left(F < f\right) = \alpha_0$$

Si dans notre exemple on suppose que $f = 0{,}65$

Alors : $$\alpha_0 = P_0\left(F < f\right)$$
$$= P_0\left(Z < -2{,}309\right) = 1\%$$

Le test est donc significatif à 5%

Signe de la région de Rejet

α_0 **P-value**	**Interpretation**
$\alpha_0 < 0{,}01$	very strong evidence against H_0
$0{,}01 \leq \alpha_0 < 0{,}05$	moderate evidence against H_0
$0{,}05 \leq \alpha_0 < 0{,}10$	suggestive evidence against H_0
$0{,}10 \leq \alpha_0$	little or no real evidence against H_0

- **2- Test Unilatéral <u>à droite</u> pour la proportion**
- *i)-F.H.*

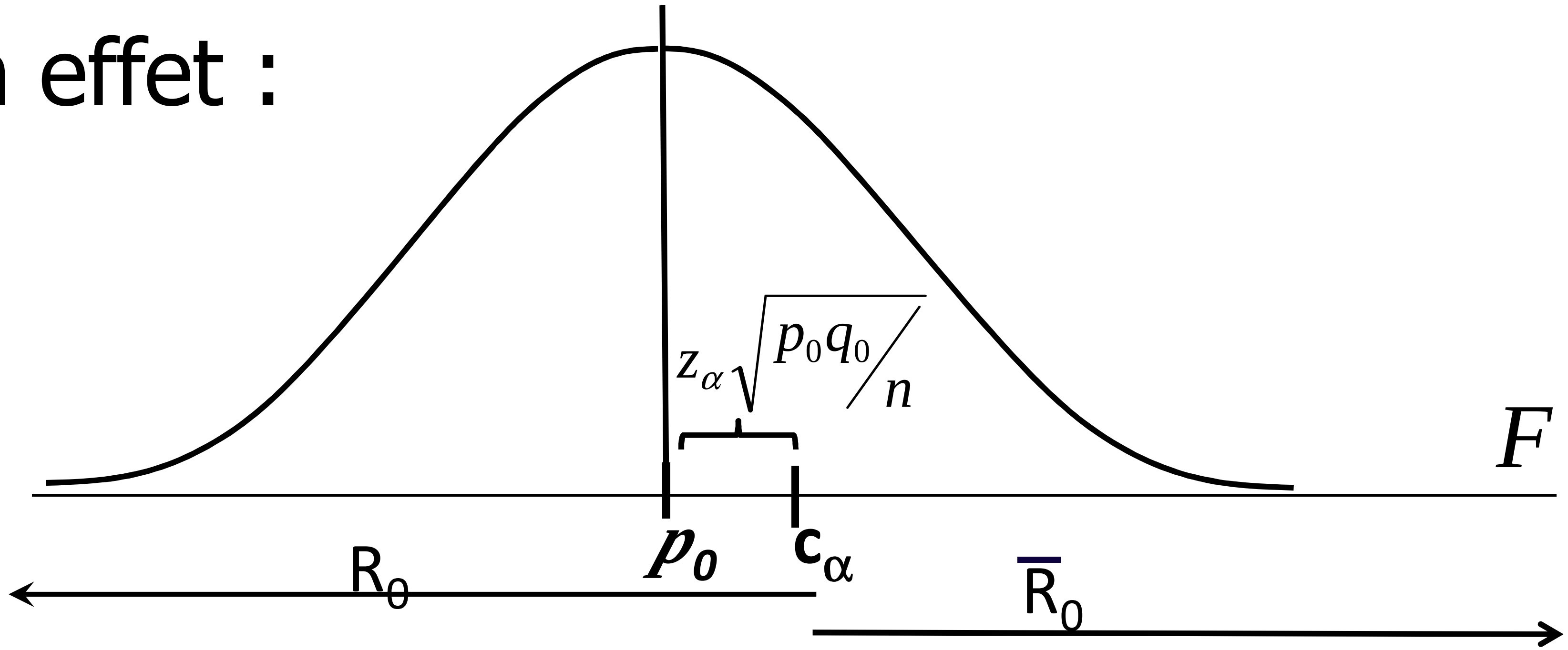
$$T.U.D.\begin{cases} H_0 : "\, p = p_0 \," \\ \# \\ H_1 : "\, p > p_0 \," \end{cases}$$

- *ii)* si le T.C.L. est vérifié sous H_0 alors la règle de décision devient :

R.D

$$\begin{cases} Si \;\; f > c_\alpha = p_0 + z_\alpha \sqrt{\dfrac{p_0 q_0}{n}} \quad \text{on rejette } \mathrm{H}_0 \\ Si \;\; f \le c_\alpha = p_0 + z_\alpha \sqrt{\dfrac{p_0 q_0}{n}} \quad \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

En effet :



$$\alpha = P_0\left(RH_0\right) = P_0\left(F > c_\alpha\right) = P_0\left(\frac{F - p_0}{\sqrt{p_0 q_0 / n}} > \frac{c_\alpha - p_0}{\sqrt{p_0 q_0 / n}}\right)$$

$$\Rightarrow \frac{c_\alpha - p_0}{\sqrt{p_0 q_0 / n}} = +z_\alpha \Rightarrow \quad c_\alpha = p_0 + z_\alpha \sqrt{p_0 q_0 / n}$$

- **Remarque :** La *p-value* ici vaut :

$$\alpha_0 = P_0\left(F > f\right)$$

Si on teste $\boldsymbol{p = 0{,}75}$ # $\boldsymbol{p > 0{,}75}$ et si F vaut $f = 0{,}65$

Alors :

$$\begin{aligned}\alpha_0 &= P_0\left(F > f\right)\\ &= P_0\left(Z > -2{,}309\right) = 99\%\end{aligned}$$

Le test n’est donc pas significatif à 5% ni à 95%

3- Test Bilatéral pour la proportion

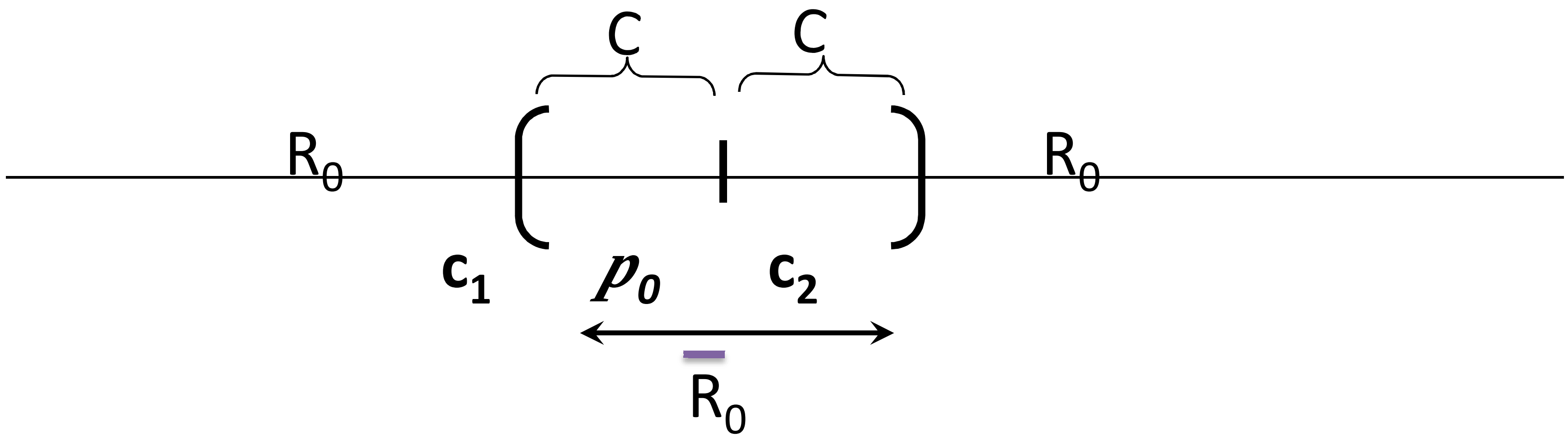
F.H

$$T.B.\begin{cases} H_0 : "p = p_0" \\ \# \\ H_1 : "p \neq p_0" \end{cases}$$

- Si T.C.L. On adopte la Règle de décision suivante :

R.D

$$\begin{cases} Si\ \ f \notin \left[c_1 ; c_2\right] \ \text{ on rejette } \mathrm{H}_0 \\ Si\ \ f \in \left[c_1 ; c_2\right] \ \text{ on ne rejette pas } \mathrm{H}_0 \end{cases}$$



or $$f \in [c_1, c_2] \Leftrightarrow f \in [p_0 - c,\, p_0 + c] \Leftrightarrow |f - p_0| \leq c$$

$$d'où \;\; f \notin [c_1, c_2] \Leftrightarrow |f - p_0| > c$$

$$\alpha = P_0(RH_0) = P_0(|F - p_0| > c) \;=\; P_0\left(\frac{|F - p_0|}{\sqrt{p_0 q_0 / n}} > \frac{c}{\sqrt{p_0 q_0 / n}} \right)$$

$$\Rightarrow \frac{c}{\sqrt{p_0 q_0 / n}} = z_{\alpha/2} \Rightarrow \quad c = z_{\alpha/2} \sqrt{p_0 q_0 / n}$$

- D’où :

R.D

$$\begin{cases} Si\ |f - p_0| > z_{\alpha/2}\sqrt{\dfrac{p_0 q_0}{n}} & \text{on rejette } H_0 \\ Si\ |f - p_0| \leq z_{\alpha/2}\sqrt{\dfrac{p_0 q_0}{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

Remarque Si on teste *p = 0,75* # *p ≠ 0,75* et si F

vaut $f = 0{,}65$, alors la *p-value* vaut ici **:**

$$\alpha_0 = P_0\left(\left|\frac{F - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| > \left|\frac{f - 0{,}75}{\sqrt{\frac{0{,}75 \times 0{,}25}{100}}}\right|\right)$$

$$= P_0\left(\left|Z\right| > +2{,}309\right) = 2\%$$

Le test est donc significatif à **5%** et même à **2,5%**

II- Tests Relatifs à Une Moyenne

- **1- Test Unilatéral <u>à droite</u> pour la moyenne**

- **Question** : est-ce que l'espérance de vie des marocains a augmenté depuis le dernier recensement ?

- Pour répondre à cette question, on doit confronter 2 hypothèses:

• ***i***)-

$$T.U.D.\begin{cases} H_0 :"\mu = \mu_0" \\ \# \\ H_1 :"\mu > \mu_0" \end{cases}$$

• ***ii***)-On adopte ensuite une Règle de décision basée sur la statistique $\overline{X}$ et qui répondra à la question: à partir de quelle réalisation de $\overline{X}$ décidera-t-on du rejet de H_0 pour un risque α ?

$$\begin{cases} Si\ \overline{x} > c \quad \text{on rejette } \mathrm{H}_0 \\ Si\ \overline{x} \le c \quad \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- **1)- Détermination de C en fonction du risque α :**

a)Si σ est connu et (X est normale ou $n \geq 30$)

$$\alpha = P_0\left(RH_0\right) = P_0\left(\overline{X} > c\right) = P_0\left(Z > \sqrt{n} \times \frac{c - \mu_0}{\sigma}\right)$$

$$\Rightarrow \frac{c - \mu_0}{\sigma} \times \sqrt{n} = z_\alpha \Rightarrow c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

D'où la règle de décision :

$$\begin{cases} Si\ \bar{x} > c_{\alpha} = \mu_0 + \mathrm{z}_{\alpha} \dfrac{\sigma}{\sqrt{\mathrm{n}}} & \text{on rejette } \mathrm{H}_0 \\ Si\ \bar{x} \leq c_{\alpha} = \mu_0 + \mathrm{z}_{\alpha} \dfrac{\sigma}{\sqrt{\mathrm{n}}} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

b) Si σ est inconnu et X est normale; on a la règle de décision :

$$\begin{cases} Si\ \bar{x} > c_{\alpha} = \mu_0 + \mathrm{t}_{n-1;\alpha}\dfrac{s}{\sqrt{\mathrm{n}}} & \text{on rejette } \mathrm{H}_0 \\ Si\ \bar{x} \le c_{\alpha} = \mu_0 + \mathrm{t}_{n-1;\alpha}\dfrac{s}{\sqrt{\mathrm{n}}} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

c) Si σ est inconnu et $n \geq 50$; on a la règle de décision

$$\begin{cases} Si\ \bar{x} > c_{\alpha} = \mu_0 + \mathrm{z}_{\alpha}\dfrac{s}{\sqrt{\mathrm{n}}} & \text{on rejette } \mathrm{H}_0 \\ Si\ \bar{x} \le c_{\alpha} = \mu_0 + \mathrm{z}_{\alpha}\dfrac{s}{\sqrt{\mathrm{n}}} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

II- Tests Relatifs à Une Moyenne

- **2- Test Unilatéral <u>à gauche</u> pour la moyenne**

On doit confronter les deux hypothèses suivantes

***i*)-**

$$T.U.G.\begin{cases} H_0 : "\mu = \mu_0" \\ \# \\ H_1 : "\mu < \mu_0" \end{cases}$$

•*ii*)-On adopte ensuite une Règle de décision basée sur la statistique $\overline{X}$ et qui répondra à la question: à partir de quelle réalisation de décidera-t-on du $\overline{X}$ rejet de H_0 ?

$$\begin{cases} Si\ \overline{x} < c \quad \text{on rejette } \mathrm{H}_0 \\ Si\ \overline{x} \geq c \quad \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

•De la même façon que précédemment, on a les règles de décision selon les cas :

a) Si σ est connu et (X est normale ou $n \geq 30$)

$$\begin{cases} Si\ \overline{x} < c_\alpha = \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}} & \text{on rejette } H_0 \\ Si\ \overline{x} \geq c_\alpha = \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

b) Si σ est inconnu et X est normale; on a la règle de décision :

$$\begin{cases} Si\ \overline{x} < c_\alpha = \mu_0 - t_{n-1;\alpha} \dfrac{s}{\sqrt{n}} & \text{on rejette } H_0 \\ Si\ \overline{x} \geq c_\alpha = \mu_0 - t_{n-1;\alpha} \dfrac{s}{\sqrt{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

c) Si σ est inconnu et $n \geq 50$; on a la règle de décision :

$$\begin{cases} Si\ \bar{x} < c_{\alpha} = \mu_0 - z_{\alpha}\dfrac{s}{\sqrt{n}} & \text{on rejette } H_0 \\ Si\ \bar{x} \geq c_{\alpha} = \mu_0 - z_{\alpha}\dfrac{s}{\sqrt{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

II- Tests Relatifs à Une Moyenne

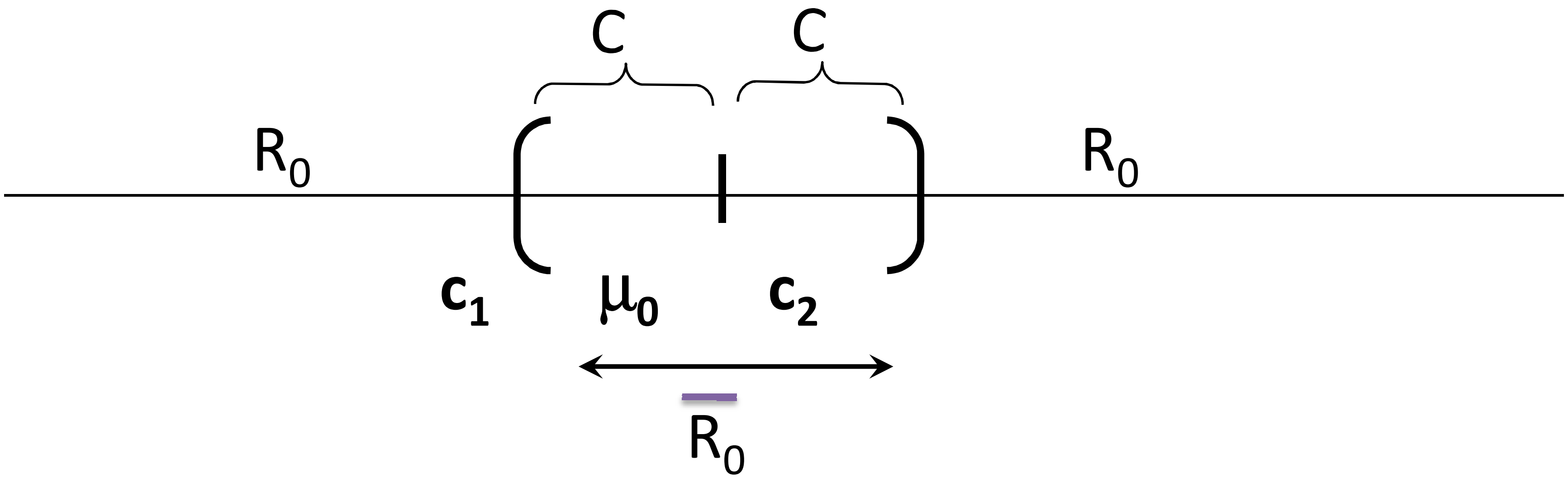
3- Test Bilatéral pour la moyenne

i)-

$$T.B. \begin{cases} H_0 : "\mu = \mu_0" \\ \# \\ H_1 : "\mu \neq \mu_0" \end{cases}$$

ii)-On adopte la Règle de décision suivante :

$$\begin{cases} Si\ \bar{x} \notin [c_1 ; c_2] & \text{on rejette } \mathrm{H}_0 \\ Si\ \bar{x} \in [c_1 ; c_2] & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$



$$\bar{x} \in [c_1 ; c_2] \Leftrightarrow c_1 \leq \bar{x} \leq c_2 \Leftrightarrow \mu_0 - c \leq \bar{x} \leq \mu_0 + c$$

$$\Leftrightarrow -c \leq \bar{x} - \mu_0 \leq +c \Leftrightarrow |\bar{x} - \mu_0| \leq c$$

D'où la règle de décision suivante

$$\begin{cases} Si \, |\bar{x} - \mu_0| > c \ \text{ on rejette } \mathrm{H}_0 \\ Si \ |\bar{x} - \mu_0| \leq c \quad \text{ on ne rejette pas } \mathrm{H}_0 \end{cases}$$

On a les règles de décision selon les cas :

a) Si σ est connu et (X est normale ou $n \geq 30$)

$$\begin{cases} Si \left|\bar{x} - \mu_0\right| > z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \quad \text{on rejette } H_0 \\ \\ Si \ \left|\bar{x} - \mu_0\right| \leq z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \quad \text{on ne rejette pas } H_0 \end{cases}$$

b) Si σ est inconnu et X est normale

$$\begin{cases} Si\ \left|\overline{x}-\mu_0\right| > t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}} & \text{on rejette } H_0 \\ Si\ \left|\overline{x}-\mu_0\right| \leq t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

c) Si σ est inconnu et $n \geq 50$; on a

$$\begin{cases} Si\ \left|\overline{x}-\mu_0\right| > z_{\alpha/2}\dfrac{s}{\sqrt{n}} & \text{on rejette } H_0 \\ Si\ \left|\overline{x}-\mu_0\right| \leq z_{\alpha/2}\dfrac{s}{\sqrt{n}} & \text{on ne rejette pas } H_0 \end{cases}$$

III- Tests Relatifs à Une Variance

- Dans ce paragraphe on supposera la ***normalité*** de la population

- **1- Test Unilatéral <u>à droite</u> pour la variance**

$$T.U.D.\begin{cases} H_0 : "\sigma^2 = \sigma_0^2" \\ \# \\ H_1 : "\sigma^2 > \sigma_0^2" \end{cases}$$

De la même façon que précédemment, on a les règles de décision selon les cas :

a) Si μ est inconnue

La statistique utilisée est S^2; d'où la R.D. est :

$$\begin{cases} Si \;\; s^2 > \chi^2_{n-1;\alpha} \dfrac{\sigma_0^2}{n-1} \text{ on rejette } \mathrm{H}_0 \\ Si \;\; s^2 \leq \chi^2_{n-1;\alpha} \dfrac{\sigma_0^2}{n-1} \text{ on ne rejette pas } \mathrm{H}_0 \end{cases}$$

2- Test Unilatéral <u>à gauche</u> pour la variance

$$T.U.G.\begin{cases} H_0 : "\sigma^2 = \sigma_0^2" \\ \# \\ H_1 : "\sigma^2 < \sigma_0^2" \end{cases}$$

<u>**R.D.** (pour μ inconnue)</u>

$$\begin{cases} Si \;\; s^2 < \chi^2_{n-1;1-\alpha} \dfrac{\sigma_0^2}{n-1} \text{ on rejette } \mathrm{H}_0 \\ Si \;\; s^2 \geq \chi^2_{n-1;1-\alpha} \dfrac{\sigma_0^2}{n-1} \quad \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

3- Test Bilatéral pour la variance

$$T.B.\begin{cases} H_0 : "\sigma^2 = \sigma_0^2" \\ \# \\ H_1 : "\sigma^2 \neq \sigma_0^2" \end{cases}$$

Si μ est inconnue

$$\begin{cases} Si \;\; s^2 \notin \left[\chi^2_{n-1;1-\alpha/2} \dfrac{\sigma_0^2}{n-1} \,;\, \chi^2_{n-1;\alpha/2} \dfrac{\sigma_0^2}{n-1} \right] \\ \text{on rejette } \mathrm{H}_0 \\ Si \;\; s^2 \in \left[\chi^2_{n-1;1-\alpha/2} \dfrac{\sigma_0^2}{n-1} \,;\, \chi^2_{n-1;\alpha/2} \dfrac{\sigma_0^2}{n-1} \right] \\ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

ANALYSE UNIVARIEE

Cas de 2 Echantillons Indépendants

(Comparaison de 2 paramètres)

- **1- Test <u>Bilatéral</u> de comparaison de 2 Variances :**
- <u>On suppose la normalité des 2 populations</u>
- <u>F.H.</u>

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ \# \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \sigma_1^2 / \sigma_2^2 = 1 \\ \# \\ H_1 : \sigma_1^2 / \sigma_2^2 \neq 1 \end{cases}$$

Cas où μ est inconnue

- Rapport critique (sous H_0):

$R_c = \dfrac{S_1^2}{S_2^2}$ Si $s_1^2 > s_2^2$

Sinon $R_c = \dfrac{S_2^2}{S_1^2}$ et <u>F.H.</u> devient alors $\begin{cases} H_0 : & \sigma_2^2 / \sigma_1^2 = 1 \\ H_1 : & \sigma_2^2 / \sigma_1^2 \neq 1 \end{cases}$

- **<u>Loi de</u>** ***R_c*** sous H_0 :

$$R_c \rightsquigarrow \mathscr{F}(\nu_1;\nu_2)$$

- **R.D.**

$$\begin{cases} Si \quad r_c \notin \left[F_{1-\alpha/2}\left(\nu_1;\nu_2\right); F_{\alpha/2}\ \left(\nu_1;\nu_2\right) \right] \\ \text{on rejette } \mathrm{H}_0 \\ Si \quad r_c \in \left[F_{1-\alpha/2}\left(\nu_1;\nu_2\right); F_{\alpha/2}\ \left(\nu_1;\nu_2\right) \right] \\ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

$$où \quad \nu_i = n_i - 1$$

- **2- Test <u>Bilatéral</u> de comparaison de 2 moyennes :**

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ \# \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \ \mu_1 - \mu_2 = 0 \\ \# \\ H_1 : \ \mu_1 - \mu_2 \neq 0 \end{cases}$$

- On propose $\overline{X}_1 - \overline{X}_2$ comme estimateur de $\mu_1 - \mu_2$:

a) Si σ_1 et σ_2 connus et (X_1 normale ou $n_1 \geq 30$) et (X_2 normale ou $n_2 \geq 30$)

R.D.

$$\begin{cases} Si\ \left|\bar{x}_1 - \bar{x}_2\right| > z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \quad \text{on rejette}\ H_0 \\ \\ Si\ \left|\bar{x}_1 - \bar{x}_2\right| \leq z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \quad \text{on ne rejette} \\ \text{pas}\ H_0 \end{cases}$$

b) Si σ_1 et σ_2 inconnus mais égales et X_1 et X_2 normales, alors la R.D. est la suivante :

$$\begin{cases} Si \left|\bar{x}_1 - \bar{x}_2\right| > \mathrm{t}_{\alpha/2;n_1+n_2-2}\hat{s}\sqrt{\frac{1}{\mathrm{n}_1}+\frac{1}{\mathrm{n}_2}} \\ \text{on rejette } \mathrm{H}_0 \\ Si \left|\bar{x}_1 - \bar{x}_2\right| \leq \mathrm{t}_{\alpha/2;n_1+n_2-2}\hat{s}\sqrt{\frac{1}{\mathrm{n}_1}+\frac{1}{\mathrm{n}_2}} \\ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

•Où

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

•**c)** Si σ_1 et σ_2 inconnus et $n_1 \geq 50$, $n_2 \geq 50$
alors la R.D. est la suivante :

$$\begin{cases} Si\ \left|\bar{x}_1 - \bar{x}_2\right| > z_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} \quad \text{on rejette } H_0 \\ \\ Si\ \left|\bar{x}_1 - \bar{x}_2\right| \leq z_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} \quad \text{on ne rejette} \\ \text{pas } H_0 \end{cases}$$

- **3- Test <u>Bilatéral</u> de comparaison de 2 Proportions :**

$$\begin{cases} H_0 : p_1 = p_2 \\ \# \\ H_1 : p_1 \neq p_2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \ p_1 - p_2 = 0 \\ \# \\ H_1 : \ p_1 - p_2 \neq 0 \end{cases}$$

- On pose $p_1=p_2=p$ et on propose comme estimateur de la proportion commune la fréquence commune des 2 échantillons :

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

- **On suppose les conditions T.C.L. vérifiées sous H_0**

- La statistique utilisée est F_1-F_2
- Rapport critique (sous H_0):

$$R_c = \frac{F_1 - F_2}{\sqrt{pq\left(1/n_1 + 1/n_2\right)}}$$

$$\approx \frac{F_1 - F_2}{\sqrt{f\left(1 - f\right)\left(1/n_1 + 1/n_2\right)}}$$

- **R.D.**

$$\begin{cases} Si \quad r_c \notin \left[-z_{\alpha/2} \; ; +z_{\alpha/2} \right] \\ \text{on rejette } \mathrm{H}_0 \\ Si \quad r_c \in \left[-z_{\alpha/2} \; ; +z_{\alpha/2} \right] \\ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- **<u>R.D.</u>** s'écrit aussi :

$$\begin{cases} Si\ \left|f_1 - f_2\right| > +z_{\alpha/2}\sqrt{f\left(1-f\right)\left(1/n_1 + 1/n_2\right)} \\ \text{on rejette } \mathrm{H}_0 \\ Si\ \left|f_1 - f_2\right| \leq +z_{\alpha/2}\sqrt{f\left(1-f\right)\left(1/n_1 + 1/n_2\right)} \\ \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

ANALYSE UNIVARIEE

Cas de 2 Echantillons Dépendants

(Comparaison de 2 moyennes)

V- Test de Comparaison de 2 moyennes : cas de 2 échantillons appariés

- **L'hypothèse testée est comme suit :** est ce que les deux échantillons dépendants proviennent de deux populations de même moyenne ?

- On teste deux produits auprès d'un même échantillon de 5 consommateurs. Chacun des consommateurs donne 2 notes, la première pour le produit A, l'autre pour le produits B.

On obtient le tableau ci-contre :

Produit	A	B
	13,3	14,15
	10,1	11,05
	8,5	9,5
	6,9	7,95
	3,7	4,85
Total	42,5	47,5

- Le problème posé est de savoir si les 2 séries de notes viennent de la même population de moyenne

$$\mu = \mu_A = \mu_B \quad ?$$

Remarque : Obligatoirement les 2 échantillons ont des tailles égales !

- Le test posé est formulé comme suit :

$$\begin{cases} H_0 : \mu_A = \mu_B \\ \# \\ H_1 : \mu_A \neq \mu_B \end{cases} \Leftrightarrow \begin{cases} H_0 :\ \mu_A - \mu_B = 0 \\ \# \\ H_1 :\ \mu_A - \mu_B \neq 0 \end{cases}$$

- On pose $\overline{D} = \overline{X}_A - \overline{X}_B$ la statistique du test afin de tester la différence $\mu_A - \mu_B$

- D'où $T = \dfrac{\overline{D}}{\hat{\sigma}_{\overline{D}}} = \dfrac{\overline{D}}{S_{\overline{D}}} = \dfrac{\overline{D}}{S_D / \sqrt{n}} \rightsquigarrow$ t(n-1)

- Un problème se pose c'est que les 2 échantillons (5 notes pour A et 5 pour B) n'étant plus indépendant (les notes sont probablement corrélées), il n'est plus possible de calculer la variance de D en faisant la somme des deux variances.

- R.D.

$$\begin{cases} Si\ |t| > \mathrm{t}_{\alpha/2;n-1} & \text{on rejette } \mathrm{H}_0 \\ Si\ |t| \le \mathrm{t}_{\alpha/2;n-1} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- A.N. :

Produit	x_A	x_B	$d=x_A - x_B$	d^2
	13,3	14,15	-0,85	0,72
	10,1	11,05	-0,95	0,90
	8,5	9,5	-1	1
	6,9	7,95	-1,05	1,10
	3,7	4,85	-1,15	1,32
Total	**42,5**	**47,5**	**-5**	**5,04**

- A.N. :

$$\overline{d} = \frac{-5}{5} = -1 \quad \overline{d^2} = \frac{5,04}{5} = 1,01$$

$$S_d^2 = 5 * (1,01 - 1) / 4 = 0,0125$$

$$D'où \quad t = \sqrt{5} * \frac{-1}{\sqrt{0,0125}} = -20$$

$$\Rightarrow |t| > t_{\alpha/2;\,4} = 2,78$$

$$\Rightarrow on \ \ rejette \ \ H_0$$

- Voilà ce que donne Xlstat
- sous Excel

t (valeur observée)	**-20,000**
t (valeur critique)	**2,776**
ddl	**4**
p-value bilatérale	**< 0,0001**
Alpha	**0,05**

Conclusion : Au seuil de signification Alpha=0,05 on peut rejeter l'hypothèse nulle d'égalité des moyennes. Autrement dit, la différence entre les moyennes est significative.

ANALYSE BIVARIEE

Cas de deux Variables

I
Test
sur le coefficient
de corrélation
de pearson

Test paramétrique
sur
le coefficient de Corrélation de pearson : ρ

- Dans ce paragraphe, nous allons aborder les questions relatives aux relations entre deux variables quantitatives X et Y en souhaitant montrer que la variable dépendante Y est fonction de la variable indépendante X.

- Pour ce on cherche à calculer le coefficient de corrélation théorique ρ sur la population :

$$\rho = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Un estimateur de ρ sur un E.A. est le coefficient r (vu en S2) :

$$r = \hat{\rho} = \frac{\hat{\operatorname{cov}}(X, Y)}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{\hat{\operatorname{cov}}(X, Y)}{S_X S_Y} \; ; où \;\; \hat{\operatorname{cov}}(x, y) = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

$$et \;\; S_X^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

$$d'où \quad r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}} = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}(\sum x_i)^2\right]\left[\sum y_i^2 - \frac{1}{n}(\sum y_i)^2\right]}}$$

1- Formulation des Hypothèses :

$$\begin{cases} H_0 : \rho = 0 \\ \# \\ H_1 : \rho \neq 0 \end{cases}$$

2- On propose la statistique du test, sous H_0 :

$$T = \frac{r}{S_r} = r\sqrt{\frac{(n-2)}{1-r^2}} \rightsquigarrow t\ (n-2)$$

3- R.D.

$$\begin{cases} Si\ \ |t| > t_{\alpha/2}(n-2)\ \ on\ rejette\ H_0 \\ \\ Si\ \ |t| \leq t_{\alpha/2}(n-2)\ \ on\ ne\ rejette\ pas\ H_0 \end{cases}$$

- Condition d’application :
  - La liaison entre X et Y est supposée linéaire.
  - Les lois de X et Y sont supposées Normales.
  - La loi conditionnelle de Y/X (recip X/Y) est normale.

- **Exemple**
- Le service des études économiques de la société α veut mesurer l'incidence de la modulation de la pression marketing (variable X: explicative) sur la vente de flacons de parfums (variable Y:expliquée). Il enregistre, alors, les ventes y_i(en milliers de flacons) ainsi que les dépenses publicitaires x_i(en milliers de DH)dans 5 zones qui forment un échantillon aléatoire

x_i	5	6	9	12	18
y_i	25	30	35	45	65

•On cherche à étudier la liaison pouvant exister entre les variables aléatoire X et Y. Pour ce, on représente dans un repère orthogonal les points (x_i, y_i) . La forme de ce nuage nous renseigne sur la nature de la liaison entre X et Y et le type de courbe qui ajustera le mieux, c'est est une droite (*ajustement linéaire ou droite de régression*).

i	x_i	y_i	$x_i\ y_i$	x_i^2	y_i^2
1	5	25	125	25	625
2	6	30	180	36	900
3	9	35	315	81	1225
4	12	45	540	144	2025
5	18	65	1170	324	4225
TOTAL	**50**	**200**	**2330**	**610**	**9000**

$$r = \frac{330}{\sqrt{110 \times 1000}} = 0{,}99$$

$$\Rightarrow t = .99\sqrt{\frac{3}{1-.99^{2}}} = 12.16$$

$$Or \quad t_{,025}(3) = 3{,}18 \qquad d'où \ |t| > t_{,025}(3)$$

$$\Rightarrow RH_{0}$$

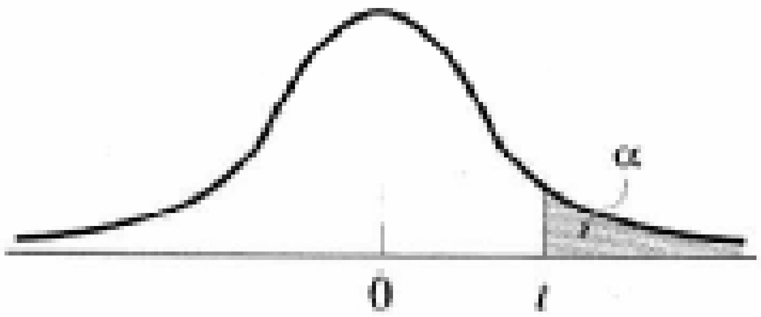
$$Or \ P(|T| > 12{,}16) = 0{,}0012 = 0{,}12\%$$

Conclusion:

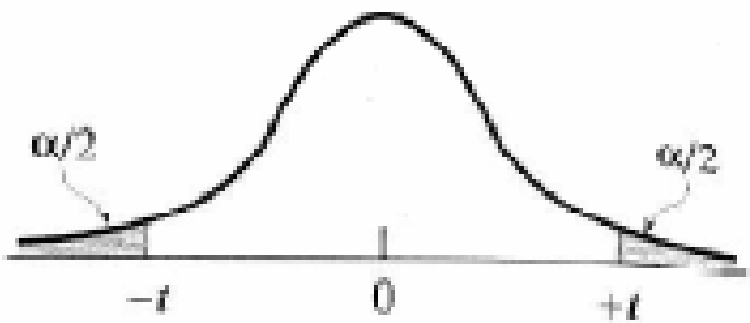
- Le coefficient de corrélation entre les ventes et les dépenses publicitaires est très significativement différent de 0.

Table A2 Table of Student's t Distribution

Two-tailed	.80	.50	.20	.10	.05	.02	.01	.001
One-tailed	.40	.25	.10	.05	.025	.01	.005	.0005
p / *df*	.60	.75	.90	.95	.975	.99	.995	.9995
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.289	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	6.869
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	3.291



Test unilatéral



Test bilatéral

Tests
non paramétriques

Chapitre 2

Introduction

- Les tests statistiques qui imposent des conditions d'application relatives aux paramètres de la population sont appelés des **tests paramétriques.**
- *Un* **test non paramétrique** *est donc un test d'hypothèse pour lequel il n'est pas nécessaire de spécifier la forme de la distribution de la population étudiée.*
- *Les méthodes non paramétriques requièrent peu d'hypothèses concernant la population étudiée. Elles ignorent notamment l'hypothèse classique de la normalité de la population.*
- *Ces tests peuvent être appliqués à de petits échantillons. Ils peuvent s'appliquer à des caractères qualitatifs, à des grandeurs de mesure. On peut les utiliser dans le cas des données incomplètes ou imprécises.*

- **En résumé :**

- **- Quelles sont les conditions d'application des tests non paramétriques ?**

Lorsque:

- - pour la variable quantitative, il n'est pas possible d'émettre certaines hypothèses :
  - <u>- normalité de la distribution ;</u>
  - <u>- égalité des variances, ...</u>
- - L'inférence ne concerne pas un paramètre dans la distribution de population
- - la taille de l'échantillon devient trop faible (hypothèses précédentes invérifiables)

- Le problème qui se pose alors est de savoir quand la distribution d’une population est normale.
- Il existe plusieurs tests statistiques qui permettent de vérifier si des données proviennent d’une population normalement distribuée. Citons par exemple le test de χ^2 et celui de Kolmogorov-Smirnov et son extension test de Lilliefort…
- On peut parfois, se faire une idée quant à la normalité des données en examinant la distribution de ces derniers dans l’échantillon:
  - Les données sont-elles relativement groupées autour d’une tendance centrale ? N’y a-t-il pas une distribution clairement bimodale incompatible avec la propriété de normalité ? N’y a-t-il pas des données aberrantes…?

ANALYSE UNIVARIE

A- Cas d’un seul échantillon

I- Tests d'ajustements de **Khi-deux**

INTRODUCTION

- Le but de ce test est de vérifier, en se basant sur les données d'1 E.A., la concordance entre une distribution observée (empirique) et une distribution théorique. Soit X la V.A. parente relative à un E.A. de taille n. Cet échantillon fournit n observations qu'on réparti en K modalités (dans le cas discret) ou K classes (dans le cas continue), notées C_1, C_2,…, C_K aux quelles on associe les effectifs aléatoires N_1, N_2,…, N_K de

Réalisations n_{o1}, n_{o2},…, n_{ok} ; on a alors les contraintes suivantes :

$$\sum_{i=1}^{k} N_i = n \qquad et \qquad \sum_{i=1}^{k} n_{oi} = n$$

On pense que ces observations proviennent d'une loi théorique L. A fin de tester cette hypothèse, on pose :

$$\begin{cases} H_0 : X \rightsquigarrow L \\ \# \\ H_1 : X \not\rightsquigarrow L \end{cases}$$

- Sous H_0 , la loi L fait correspondre à $C_1, C_2,\dots, C_K$ les probabilités $p_1, p_2,\dots, p_K$ et par suite les effectifs théorique $n_{t1}= np_1, n_{t2}= np_2,\dots, n_{tk}= np_k$, avec : $\sum_{i=1}^{k} n_{ti} = n$

I) Cas où la loi ne dépend pas de paramètres

- Soit la statistique :

$$D^2 = \sum_{i=1}^{k} \frac{\left(N_i - n_{ti}\right)^2}{n_{ti}}$$

- **Théorème :**

– Si $n \to \infty$ $\quad D^2 \rightsquigarrow \chi^2_{k-1}$

– D'où la règle de décision du test :

$$\begin{cases} Si \;\; d^2 > \chi^2_{k-1;\alpha} & \text{on rejette } \mathrm{H}_0 \\ Si \;\; d^2 \leq \chi^2_{k-1;\alpha} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- Où :

$$d^2 = \sum_{i=1}^{k} \frac{(n_{oi} - n_{ti})^2}{n_{ti}}$$

$$= \sum_{i=1}^{k} \frac{(n_{oi})^2}{n_{ti}} - n$$

- Remarque : 1) le ddl (k-1) vient du fait que les N_i sont liés par la relation

$$\sum_{i=1}^{k} N_i = n$$

Remarque : 2) le théorème s'applique si $np_i \geq 5 \ \ \forall i$, quitte à regrouper les classes qui ne vérifient pas la condition.

1) Ajustement à 1 loi Uniforme discrète

Exemple : considérons l'expérience suivante effectuée avec un dé. On désire vérifier si ce dé est réellement bien balancé. Après 96 essais, on obtient les résultats suivants:

Résultats	1	2	3	4	5	6
Fréquences observées	**12**	**18**	**20**	**13**	**10**	**23**

– Quelle conclusion peut-on tirer de cette expérience ? (α= 5%).

Modalité x_i	n_{oi}	p_i	$n_{ti} = np_i$
1	12	1/6	16
2	18	1/6	16
3	20	1/6	16
4	13	1/6	16
5	10	1/6	16
6	23	1/6	16
TOTAL	96	1	96

- On émet deux hypothèses :

$$\begin{cases} H_0 : X \rightsquigarrow \bigcup_{\{1,2,\ldots,6\}} \\ \# \\ H_1 : X \not\rightsquigarrow \bigcup_{\{1,2,\ldots,6\}} \end{cases}$$

- Peut-on rejeter H_0 au risque de 5% ?

$$\begin{cases} Si\ \ d^2 > \chi^2_{5;0,05} \quad \text{on rejette } \mathrm{H}_0 \\ \\ Si\ \ d^2 \leq \chi^2_{5;\alpha} \quad \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- Le calcul de la réalisation d^2 :

x_i	n_{oi}	n_{ti}	$(n_{ti} - n_{oi})^2$	$[(n_{ti} - n_{oi})^2]/n_{ti}$
1	**12**	**16**	**16**	1
2	**18**	**16**	**4**	0,25
3	**20**	**16**	**16**	1
4	**13**	**16**	**9**	0,56
5	**10**	**16**	**36**	2,25
6	**23**	**16**	**49**	3,06
TOTAL	96	96		$d^2 = 8,12$

- Or $d^2 = 8{,}12$ et $\chi^2_{5\,;\,0{,}05} = 11{,}07$

- D'où on ne peut rejeter H_0 au niveau 5%.

II) Cas d'une loi discrète dépendant de paramètres

- Soit la statistique :

$$D^2 = \sum_{i=1}^{k} \frac{\left(N_i - n_{ti}\right)^2}{n_{ti}}$$

- **Théorème :**

– Si $n \to \infty$ $\quad D^2 \rightsquigarrow \chi^2_{k-m-1}$

où m est le nombre de paramètres à estimer.

1) Ajustement à 1 loi de poisson

Exemple : Le nombre d'accidents de travail en une journée pour une période de 100 jours est réparti comme suit :

Modalité x_i	0	1	2	3	4	5	6	**TOTAL**
Effectif n_i	13	27	27	19	9	4	1	**100**

Tester au risque de 5% si les données proviennent d'1 loi de poisson.

- On émet deux hypothèses :

$$\begin{cases} H_0 : X \rightsquigarrow \mathscr{P}(\lambda) \\ \# \\ H_1 : X \not\rightsquigarrow \mathscr{P}(\lambda) \end{cases}$$

- R.D (puisque λ est le seul à estimer) :

$$\begin{cases} Si \;\; d^2 > \chi^2_{k-2;\alpha} & \text{on rejette } \mathrm{H}_0 \\ \\ Si \;\; d^2 \leq \chi^2_{k-2;\alpha} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- Estimation de λ :

$$\hat{\lambda} = \frac{\sum_{i=1}^{k} n_{oi}\ x_i}{n} = \frac{200}{100} = 2$$

$Modalité\ x_i$	n_{oi}	$n_{oi}x_i$
0	**13**	**0**
1	**27**	**27**
2	**27**	**54**
3	**19**	**57**
4	**9**	**36**
5	**4**	**20**
6	**1**	**6**
TOTAL	**100**	**200**

•Le Tableau théorique pour λ=2 :

x_i	p_i	$n_{ti} = n\,p_i$
0	**0,1353**	**13,53**
1	**0,2707**	**27,07**
2	**0,2707**	**27,07**
3	**0,1804**	**18,04**
4	**0,0902**	**9,02**
5	**0,0361**	**3,61**
6	**0,0120**	**1,20**
7	**0,0034**	**0,34**
8	**0,0009**	**0,09**
9	**0,0002**	**0,02**
TOTAL	1	

5,26

K=6

•Le calcul de la réalisation d^2 :

x_i	n_{oi}	n_{ti}	$(n_{ti} - n_{oi})^2$	$[(n_{ti} - n_{oi})^2]/n_{ti}$
0	13	13,53	0,2809	0,0208
1	27	27,07	0,0049	0,0002
2	27	27,07	0,0049	0,0002
3	19	18,04	0,9216	0,0511
4	9	9,02	0,0004	0
5 et +	5	5,26	0,0676	0,0129
TOTAL	100	100		d^2=0,0852

- $d^2 = 0{,}0852 \quad et \; \chi^2_{4;\,0{,}05} = 9{,}49$

$$d^2 \ll \chi^2_{4;5\%}$$

- D’où on ne peut rejeter l’idée que les données proviennent d’une loi de poisson

- **III) Cas d’une loi continue dépendant de paramètres**

1) <u>Ajustement à 1 loi Normale</u>

- **Exemple :** Est ce que les données ci-dessous –représentant la distribution de 685 ménages selon le revenu en centaines de DH- proviennent d’une population normalement distribuée ?
- (α = 5%).

- **Le Tableau observable :**

Classes	n_i	x_i	$n_i x_i$	$n_i x_i^2$
[5 , 15[	**41**	**10**	**410**	**4100**
[15 , 25[	**75**	**20**	**1500**	**3000**
[25 , 35[	**62**	**30**	**1860**	**55800**
[35 , 45[	**226**	**40**	**9040**	**361600**
[45 , 55[	**89**	**50**	**4450**	**222500**
[55 , 65[	**109**	**60**	**6540**	**392400**
[65 , 75[	**83**	**70**	**5810**	**406700**
TOTAL	685	-----	**29610**	**1473100**

- On émet deux hypothèses :

$$\begin{cases} H_0 & X \rightsquigarrow \mathcal{N}(\mu\,,\sigma^2) \\ H_1 & X \not\rightsquigarrow \mathcal{N}(\mu\,,\sigma^2) \end{cases}$$

- Estimation de μ et de σ^2 :

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{7} n_{oi}\ x_i}{n} = \frac{29610}{685} = 43{,}2263\ cdh$$

$$s_e^2 = \frac{\sum_{i=1}^{7} n_{oi}\, x_i^2}{685} - \hat{\mu}^2 = \frac{1473100}{685} - 1868{,}5111$$

$$\cong 282 \Rightarrow \hat{\sigma}^2 = s^2 = \frac{n}{n-1} s_e^2 = 282{,}4123$$

R.D.

$$\begin{cases} Si \;\; d^2 > \chi^2_{k-3;\alpha} & \text{on rejette } \mathrm{H}_0 \\ Si \;\; d^2 \leq \chi^2_{k-3;\alpha} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- Posons :

- $Z = \dfrac{X - \hat{\mu}}{\hat{\sigma}} \rightsquigarrow \mathcal{N}(0\ ,\ 1)$ sous H_0

- **Le Tableau théorique :**

$[z_{i-1} ; z_i[$	p_i	n_{ti}	$(n_{ti}-n_{oi})^2/n_{ti}$
$]-\infty ; -1,68[$	0,0465	31,8525	2,6270
$[-1,68 ; -1,08[$	0,0936	64,1160	1,8476
$[-1,08 ; -0,49[$	0,1720	117,8200	26,4460
$[-0,49 ; 0,11[$	0,2317	158,7145	28,5250
$[0,11 ; 0,70[$	0,2142	146,7270	22,7116
$[0,70 ; 1,30[$	0,1452	99,4620	0,9147
$[1,30 ; +\infty[$	0,0968	66,3080	4,2020
TOTAL	1,0000	685	87,2739

- $d^2 = 87{,}2739$ *et* $\chi^2_{4;\,0{,}05} = 9{,}49$

$$d^2 >> \chi^2_{4;5\%}$$

- D’où on peut rejeter l’idée que les données proviennent d’une loi NORMALE.

II
1-Test de Kolmogorov
-
Smirnov

Le test de *Kolmogorov-Smirnov* est un test d'ajustement à une loi «<u>continue</u>». On extrait un échantillon aléatoire $(X_1, X_2, \ldots, X_n)$ de fonction de répartition observée $\hat{F}$. On pense que cet échantillon provient d'une population distribuée selon une loi $\mathscr{L}$ inconnue de fonction de répartition théorique F_0. Le test se formule comme suit :

$$\begin{cases} H_0 : \hat{F} = F_0 \\ \# \\ H_1 : \hat{F} \neq F_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : X \leadsto \mathscr{L} \\ \# \\ H_1 : X \not\leadsto \mathscr{L} \end{cases}$$

Où :

$\hat{F}(x)$ *est la fonction de répartition empirique*

La règle de décision se base sur l'écart maximal entre les deux fonctions de répartitions; plus cet écart est grand plus on a tendance à rejeter l'hypothèse nulle:

$$D = \max_{i=1,\ldots,n}\left\{ \left|\hat{F}\left(x_{(i)}\right) - F_0\left(x_{(i)}\right)\right| , \left|\hat{F}\left(x_{(i-1)}\right) - F_0\left(x_{(i)}\right)\right|\right\}$$

où $x_{(i)}$ est la $i^{ème}$ réalisation par ordre ↗

Exemple : *Un chercheur entreprend une étude pour évaluer si la distribution de la durée que prend la migraine chez des patients pour répondre à une dose d'un médicament administrée est normal, avec un temps moyen de réponse* de 90 *secondes et un écart type de 35 secondes (c.-à-d.,* $\mathcal{N}(90\,,35^2)$*). On relève chez 7 patients le temps (en secondes)que met la migraine pour disparaître après l'administration du médicament. Après classement par ordre croissant, on obtient : 21, 32, 38, 90, 90, 145, 155.*

Les données proviennent-elles d'une loi normale $\mathcal{N}(90,35^2)$ *?*

F.H.

$$\begin{cases} H_0 : \hat{F} = F_0 \\ \# \\ H_1 : \hat{F} \neq F_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : X \rightsquigarrow \mathcal{N}(90,35^2) \\ \# \\ H_1 : X \not\rightsquigarrow \mathcal{N}(90,35^2) \end{cases}$$

Stat. Utilisée : $D = \max\limits_{i=1,\ldots,n} \left\{ \left| \hat{F}\left(x_{(i)}\right) - F_0\left(x_{(i)}\right) \right| , \left| \hat{F}\left(x_{(i-1)}\right) - F_0\left(x_{(i)}\right) \right| \right\}$

Loi de D : à lire dans la table de K-S

RD.

$$\begin{cases} Si \quad d > d_{\alpha;n} \quad on \quad RH_0 \\ Si \quad d \leq d_{\alpha;n} \quad on \quad NRH_0 \end{cases}$$

A.N. on range les x_i par ordre croissant puis on calcul

$$\hat{F}_i = \frac{N_i}{n} = \frac{\sum_{j=1}^{i} n_j}{n}$$ avec n_j l'effectif de la modalité x_j

x_i	z_i	F_0	$\hat{F}$	$\lvert\hat{F}(x_i)-F_0(x_i)\rvert$	$\lvert\hat{F}(x_{i-1})-F_0(x_i)\rvert$
21	-1,97	0,0244	1/7	0,1184	+0,0244
32	-1,66	0,0485	2/7	0,2373	0,0944
38	-1,49	0,0681	3/7	0,3605 **d’**	0,2176
90	0	0,5000	5/7	0,2143	+0,0714
145	1,57	0,9418	6/7	+0,0846	+0,2275 **d’’**
155	1,86	0,9686	1	0,0314	+0,1114

$0{,}3605 = d = \sup(d', d'') < d_{0,05\,;\,7} = 0{,}483$: on ne peut donc rejeter H_0 (table de Kolmogorov-Smirnov)

Remarques :

- Le test de l’Ajustement de χ^2 est utilisé pour X qualitative, discrète, ou éventuellement continue regroupée en classes
- Le test de Kolmogorov-Smirnov est utilisé pour des lois continues et parfois pour X qualitative ordinale

- Le test de Kolmogorov-Smirnov est meilleur que celui de l’Ajustement de χ^2 lorsque :

a)-la taille de l’E.A. est petite

b)- les effectifs théorique du χ^2 sont faibles.

- Si les paramètres de la loi théorique sont inconnus, on utilise une <u>extension</u> du test de Kolmogorov-Smirnov qui s’appelle le test de Lilliefors

II
2- Test de Lilliefors

Exemple :

➢On reprend l'exemple précédent mais en supposant que les paramètres (μ et σ) sont inconnus.

➢D'abord on les estime, on trouve que $\begin{cases} \hat{\mu} = 81{,}57 \\ \hat{\sigma} = 54{,}13 \end{cases}$

x_i	z_i	F_0	$\hat{F}$	$\left\lvert \hat{F}(x_i) - F_0(x_i) \right\rvert$	$\left\lvert \hat{F}(x_{i-1}) - F_0(x_i) \right\rvert$
21	-1,12	0,1314	1/7	0,0115	+0,1314
32	-0,92	0,1788	2/7	0,1069	+0,0359
38	-0,8	0,2119	3/7	0,2167 d'	0,0739
90	0,16	0,5636	5/7	0,1507	+0,1350
145	1,17	0,8790	6/7	+0,0219	+0,1647 d''
155	1,36	0,9131	1	0,0869	+0,0559

$$RD \begin{cases} Si \quad d > d_{\alpha;n} \quad on \quad RH_0 \\ Si \quad d \le d_{\alpha;n} \quad on \quad NRH_0 \end{cases}$$

$0{,}2167 = d = \sup(d', d'') < d_{0{,}02;7} = 0{,}348$: on ne peut donc rejeter H_0

(Utilisez la table de Lilliefors)

ANALYSE UNIVARIE

A- Cas de deux échantillons indépendants

I
Test du Kolmogorov
-
smirnov

- Il permet de tester si deux échantillons indépendants sont extrait de la même population.
  Il compare deux distributions cumulatives et est concerné par la concordance entre deux séries de valeurs
- On émet deux hypothèses :

$$\begin{cases} H_0 : \quad F_1 = F_2 \\ \# \\ H_1 : \quad F_1 \neq F_2 \end{cases}$$

- Le test est basé sur l'écart maximum entre les deux fonctions de répartitions empiriques :
-

$$D = \max_{i=1,\dots,n} \left\{ \left| F_1(X) - F_2(X) \right| \right\}$$

- Exemple : *Un enseignant a noté deux groupes d'étudiants (aléatoirement choisis) :*
- *Groupe 1 : 11 - 1 - 0 - 4 - 0*
- *Groupe 2 : 11 - 11 - 5 - 8 - 3*

- *Est-ce que les deux populations sont Identiquement distribuées?*

$d=0,6 \leq d_{0,05;5;5}=0,8$

$\Rightarrow NRH_0$

x_i	y_i	F_1	F_2	$\|F_1 - F_2\|$
0-0	-	2/5	0/5	0,4
1	-	3/5	0/5	0,6
-	3	3/5	1/5	0,4
4	-	4/5	1/5	0,6
-	5	4/5	2/5	0,4
-	8	4/5	3/5	0,2
11	11-11	5/5	5/5	0

II
Test U
de
Mann-Whitney

(Test Bilatéral seulement !)

pour le test Unilatéral voir T.D.

- C’est l’équivalent du test t pour la comparaisons de 2 moyennes **(E.A. indépendants)**
- C’est donc un test de comparaison des moyennes lorsque l’hypothèse de normalité ou de l’égalité des variances (pour 2 échantillons de tailles faibles) sont violées.

- On émet deux hypothèses :

- H_0 : Les deux échantillons proviennent d’une population identique ($\textit{même moyenne}$)

- H_1 : Les deux échantillons proviennent de populations différentes

- ***Décision*** : le test de Mann-Whitney est basé sur la statistique $min\,(U_1,U_2)$ où :

$$U_1 = n_1 * n_2 + \frac{n_1 * (n_1 + 1)}{2} - R_1$$

$$et$$

$$U_2 = n_1 * n_2 + \frac{n_2 * (n_2 + 1)}{2} - R_2$$

- où n_1 et n_2 sont les tailles respectives des deux échantillons et R_i est la somme des rangs du $i^{éme}$ échantillon. Les 2 échantillons étant mélangée en un seul puis on range ce dernier par ***ordre croissant***. *A chaque valeur* on affecte un **rang**. Si deux *valeur*s sont égales, on leur assigne le **rang médian.**

- **Si n_1 et $n_2 \leq 20$** : On rejette H_0 quand la valeur calculée est ***inférieure*** à celle lue dans la table de Mann-Withney.

- Remarque:

$$U_1 + U_2 = n_1 \times n_2 \quad et \quad R_1 + R_2 = \frac{(n_1 + n_2) \times (n_1 + n_2 + 1)}{2}$$

- Exemple : On veut tester l'aptitude de 12 femmes et de 12 hommes pour un certain travail administratif. Voici les points obtenus:
- Femmes: 82 87 89 91 91 76 74 70 88 99 61 94
- Hommes: 80 79 92 65 70 84 95 78 81 85 73 70

a) **F.H.**

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

b) Statistique utilisée : $U = \min(U_1, U_2)$

c) Loi de U : table de Mann-Whitney Car $(n_1 ; n_2) \leq 20$

d) R.D.

$$\begin{cases} Si \;\; u < u_{0,05;n_1;n_2} & \text{on rejette } H_0 \\ Si \;\; u \geq u_{0,05;n_1;n_2} & \text{on ne rejette pas } H_0 \end{cases}$$

e) A.N.

On range les 24 valeurs par ordre croissant. On obtient le classement suivant:

61	65	70	70	70	73	74	76	78	79	80	81	82	84	85	87
F	H	F	H	H	H	F	F	H	H	H	H	F	H	H	F

88	89	91	91	92	94	95	99
F	F	F	F	H	F	H	F

On a :

$$n_1 = n_2 = 12 \quad et$$

$$R_1 = 1 + 4 + 7 + 8 + 13 + 16 + 17 + 18 + 19{,}5 + 19{,}5$$

$$+ 22 + 24 = 169$$

$$R_2 = 2 + 4 + 4 + 6 + 9 + 10 + 11 + 12 + 14 + 15 + 21 + 23$$

$$= 131$$

$$\Rightarrow \quad U_1 = 53 \quad et \quad U_2 = 91 \quad \Rightarrow \quad U = 53$$

Or d'après la table de Mann-Withney (bilatéral) :

$$U < U_{0,05;12;12} = 37$$

D'où on ne peut rejeter H_0

Remarque :

- On peut montrer que

$$E(U) = \frac{n_1 \times n_2}{2} \quad et \quad V(U) = \frac{n_1 \times n_2 \times (n_1 + n_2 + 1)}{12}$$

- Si n_1 ; $n_2 > 20$:

$$\frac{U - E(U)}{\sqrt{V(U)}} = \frac{U - \frac{n_1 \times n_2}{2}}{\sqrt{\frac{n_1 \times n_2 \times (n_1 + n_2 + 1)}{12}}} \approx \mathcal{N}(0,1)$$

- Pour notre cas :

$$z = \frac{53 - 72}{\sqrt{300}} = -1{,}097$$

- $|z| \leq z_{0,025} = 1{,}96 \Rightarrow$ On ne peut rejeter H_0

ANALYSE UNIVARIE

B- Cas de deux échantillons dépendants

I

Test T de Wilcoxon

On dispose de deux séries de mesures : $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_n$, de variables X et Y continues, sur les mêmes n individus.

- On classe les n différences $x_i - y_i$ par ordre croissant de leurs valeurs absolues, puis on note les rangs des différences >0 (notées +) dans l'interclassement.

X	Y	$\|D\|=\|X-Y\|$	Signe de la différence	Rang
x_1	y_1	$\|d_1\|$	-	r_i
.	.	.	-	.
.	.	.	-	.
.	.	.	+	.
.	.	.	+	.
.	.	.	-	.
.	.	.	+	.
.	.	.	-	.
.	.	.	-	.
.	.	.	+	.
x_n	y_n	$\|d_n\|$	-	r_n

- **Si H_0 est vraie**, X et Y sont de même loi, chaque différence peut être > 0 ou < 0 avec la probabilité 1/2.
- s'il y a des différences nulles, on enlève les mesures correspondantes.

- On vérifie ensuite si la somme des rangs des différences positives diffère ou non de la somme des rangs des différences négatives. La somme de tous les rangs est égale à la somme des n premiers entiers :

avec n = nombre de paires non nulles

$$\sum R_+ + \sum R_- = \frac{n(n+1)}{2}$$

- L'hypothèse nulle, s’écrit alors :

$$H_0 \; : \; \sum R_+ = \sum R_- = \frac{\frac{n(n+1)}{2}}{2} = \frac{n(n+1)}{4}$$

- Si au contraire, les séries diffèrent, les sommes des rangs positifs et négatifs seront différentes.
- Le test en T consiste à comparer les rangs positifs ou négatifs à la valeur théorique sous H_0.

- Soit $T_+ = \sum R_+$ la somme des rangs positifs et
- $T_- = \sum R_-$ la somme des rangs négatifs, on prend la plus petite des deux valeurs, donc $T = \min(T_+ ; T_-)$, et on lit la valeur théorique de T dans la table, en fonction du nombre "n" de paires non nulles. La différence n'est jamais significative, si "n" est inférieur à 5 (voir table).

- Si T observé >= T théorique, on ne peut rejeter H_0 d’où il n'existe pas de différence significative entre les deux séries.

- Exemple : soit un échantillon de 7 consommateurs qui notent 2 produits : Pour notre cas :

N° cons	Prod A	Prod B	|d|	Signe	rang
1	9	6	3	+	5
2	7	5	2	+	3
3	4	8	4	-	7
4	7,5	9	1,5	-	2
5	8,5	5	3,5	+	6
6	7,5	5	2,5	+	4
7	5	6	1	-	1

- $T_+ = 5+3+6+4=18$ $T_- = 7+2+1=10$ $\Rightarrow T=10$
- Or $T > T_{0,05;7} = 2 \Rightarrow$ on ne peut rejeter H_0

- T tend vers une Loi normale si le nombre de paires est suffisamment grand, c'est à dire > 20 Dans ce cas

- $$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx \mathcal{N}(0,1)$$

ANALYSE BIVARIEE

Cas de deux Variables

I

Test sur le coefficient de corrélation des rangs de spearman

- Si l’une des conditions du test du coefficient de corrélation de pearson n’est pas vérifiée, on utilise alors un test non paramétrique : le coefficient de corrélation des rangs de Spearman, qui utilise le rang des variables et non leurs valeurs.

- Il étudie l’existence d’une liaison entre 2 variables quantitatives.

- On définit x'_i et y'_i les rangs des valeurs observées de X et de Y.
- On définit r_s le coefficient de corrélation des rangs de Spearman.

$$r_S = 1 - \frac{6\sum\left(x'_i - y'_i\right)^2}{n\left(n^2 - 1\right)}$$

- Si $|r_S|$ observé **>=** $r_{S;\alpha}$ théorique (lu sur la table de Spearman), on peut rejeter $H_0 : \rho_S = 0$
- d'où il existe une liaison significative entre les deux variables.

- Si n>10; on peut approximer la loi de la statistique .

$$t = \frac{r_S}{S_r} = r_S \sqrt{\frac{n-2}{1-r_S^2}} \rightsquigarrow t\ (n-2)$$

- Où S_r est l'écart type du coefficient de Spearman :

$$S_r = \sqrt{\frac{1-r_S^2}{n-2}}$$

- R.D.

$$\begin{cases} Si \ \ |t| > t_{\alpha/2}(n-2) \ \ on \ rejette \ H_0 \\ Si \ \ |t| \leq t_{\alpha/2}(n-2) \ \ on \ ne \ rejette \ pas \ H_0 \end{cases}$$

- **Exemple**

Une entreprise désire tester s'il y a un lien entre les statistiques des ventes de ses employés et les cours de perfectionnement offerts à tout son personnel. Un test passé par 9 vendeurs ayant suivi le cours donne un classement de leur aptitude à la vente. En utilisant les statistiques sur les ventes, on obtient alors les classements suivants:

vendeur nº :	1	2	3	4	5	6	7	8	9
rang du test	5	4	6	3	9	1	7	8	2
rang des ventes	9	5	1	4	8	7	2	3	6
$d_i = x'_i - y'_i$	-4	-1	5	-1	1	-6	5	5	-4

- On calcule r_s le coefficient de corrélation des rangs de Spearman.

$$r_S = 1 - \frac{6 \times 146}{9(81-1)} = 1 - \frac{876}{720} = -,22$$

Table A18 Table of Critical Values for Spearman's Rho

	One–tailed level of significance			
	.05	.025	.01	.005
	Two–tailed level of significance			
	.10	.05	.02	.01
n				
4	1.000	–	–	–
5	.900	1.000	1.000	–
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.700	.783	.833
10	.564	.648	.745	.794

- Donc le $r_{S;5\%}$ théorique vaut 0,7

- $|r_S| = 0,22 < 0,7$ (lu sur la table de Spearman), on ne peut rejeter H_0 , d'où la liaison entre les cours de perfectionnement et les ventes n'est pas significative à 5%

II
Test du Khi-deux

B- Test d'indépendance de χ^2

- Soient X et Y deux V.A. qualitatives, sur une même population, de modalités resp^{es}
- X : $$x_1; x_2; \cdots; x_i; \cdots; x_m$$
- Y : $$y_1; y_2; \cdots; y_j; \cdots; y_k$$
- Le tableau suivant résume la situation

1-TABLEAU DE CONTINGENCE

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_k	TOTAL
x_1	n_{11}	n_{12}				n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}				n_{2k}	$n_{2.}$
$\vdots$							$\vdots$
x_i				n_{ij}			$n_{i.}$
$\vdots$							$\vdots$
x_m	n_{m1}	n_{m2}				n_{mk}	$n_{m.}$
TOTAL	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.k}$	$n_{..}$

- On a les relations suivantes :

$$n_{i\bullet} = \sum_{j=1}^{k} n_{ij} \qquad n_{\bullet j} = \sum_{i=1}^{m} n_{ij}$$

$$n_{\bullet\bullet} = n = \sum_{i=1}^{m} \sum_{j=1}^{k} n_{ij} = \sum_{j=1}^{k} \sum_{i=1}^{m} n_{ij}$$

$$p_{ij} = P\left(X = x_i ; Y = y_j\right);$$

$$\text{probabilité conjointe de } (X, Y)$$

$$p_{i\bullet} = P\left(X = x_i\right);$$

probabilité marginale de X

$$p_{\bullet j} = P\left(Y = y_j\right);$$

probabilité marginale de Y

$$\text{Si } X \coprod Y \Rightarrow p_{ij} = p_{i\bullet} \times p_{\bullet j}$$

- On désire savoir si X est indépendante de Y; pour ce faire, on confronte les deux hypothèses suivantes :

- On émet deux hypothèses :

$$\begin{cases} H_0 : \; X \text{ et } Y \quad \text{indépendantes} \\ \# \\ H_1 : \; X \text{ et } Y \quad \text{sont liées} \end{cases}$$

- On remarque : sous H_0

$$p_{ij} = p_{i\bullet} \times p_{\bullet j}$$

Estimons :

$$\begin{cases} p_{i\bullet} & par & \hat{p}_{i\bullet} & = & \dfrac{n_{i\bullet}}{n} \\ p_{\bullet j} & par & \hat{p}_{\bullet j} & = & \dfrac{n_{\bullet j}}{n} \\ p_{ij} & par & \hat{p}_{ij} & = & \dfrac{n_{ij}}{n} \end{cases}$$

Sous H_0

$$n_{ij} = n\hat{p}_{ij} \approx n\hat{p}_{i\bullet} \times \hat{p}_{\bullet j} = \frac{n_{i\bullet} \times n_{\bullet j}}{n} = n_{t\,ij}$$

- Posons

$$D^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(N_{ij} - n_{t\,ij}\right)^2}{n_{t\,ij}}$$

- **Théorème :**

- Si $n \to \infty$ $D^2 \rightsquigarrow \chi^2_{(k-1)(m-1)}$

- <u>R.D</u>

$$\begin{cases} Si \;\; d^2 > \chi^2_{(k-1)(m-1);\alpha} & \text{on rejette } \mathrm{H}_0 \\ Si \;\; d^2 \leq \chi^2_{(k-1)(m-1);\alpha} & \text{on ne rejette pas } \mathrm{H}_0 \end{cases}$$

- **Exemple :** On a relevé les principales défectuosités sur quelques modèles de moto-neiges comme suit :

Y → / X ↓	**Electrique**	**Mécanique**	**Total**
Lynx	105	95	**200**
Renard	100	60	**160**
Total	**205**	**155**	**360**

- Tester au risque de 5% si on a les mêmes types de défectuosités quelque soit le modèle ?

- On émet deux hypothèses :

$$\begin{cases} H_0 : \ X \text{ et } Y \quad \text{indépendantes} \\ \# \\ H_1 : \ X \text{ et } Y \quad \text{sont liées} \end{cases}$$

- R.D.

$$\begin{cases} Si \ \ d^2 > \chi^2_{1;\ 0,05} \quad \text{on rejette H}_0 \\ \\ Si \ \ d^2 \leq \chi^2_{1;\,0,05} \quad \text{on ne rejette pas H}_0 \end{cases}$$

Le calcul de la réalisation d^2 :

	n_{ij}	$n_{t\,ij}=(n_{i.}\,n_{.j}/n)$	$(n_{ij}-n_{t\,ij})^2/n_{t\,ij}$
(1,1)	**105**	**113,8889**	**0,6938**
(1,2)	**95**	**86,1111**	**0,9176**
(2,1)	**100**	**91,1111**	**0,8672**
(2,2)	**60**	**68,8889**	**1,1470**
TOTAL	**360**	**360**	**3,6256**

- Or $\chi^2_{1;\,0,05} = 3,8415$ d'où $d^2 \leq \chi^2_{1;\,0,05}$

- On ne peut donc rejeter l'indépendance