Chapitre III
Comparaison

1) Test d'hypothèse

A partir des données d'un échantillon d'une population Ω on teste si celles-ci sont conformes à un modèle précis, à des paramètres choisis, etc. On pourra aussi tester à partir de deux échantillons de deux populations distinctes, les caractères de deux variables aléatoires, etc.

Cela se fait à travers une analyse qui permet de prendre une décision tout en ayant une estimation du risque d'erreur.

Le cadre est celui d'un problème de décision entre une hypothèse notée H_0, hypothèse nulle, et une autre hypothèse dite H_1.

Ici les hypothèses portent sur les paramètres d'une variable aléatoire.

- On testera par exemple l'hypothèse

$H_0 : \theta = \theta_0$, où θ = paramètre d'une v.a.

Alors $H_1 : \theta \neq \theta_0$: test **bilatéral**.

Ou encore $H_0 : \theta \geq \theta_1$ et $H_1 : \theta < \theta_1$

test **unilatéral**.

- La situation est alors la suivante :

réalité / décision	H_0 vraie	H_0 fausse
ne pas rejeter H_0	Vrai positif	Faux Positif
rejeter H_0	Faux Négatif	Vrai Négatif

VP et VN sont de bonnes décisions.

FN : erreur de 1re espèce.

FP : erreur de 2nde espèce.

FN : On note α la probabilité de rejeter à tort H_0. (7)

α est appelé niveau du test (ou seuil)

FP : β = probabilité d'accepter H_0 à tort.

$1-\beta$ = puissance du test.

Les tests d'hypothèse ne permettent pas d'accepter H_0 mais seulement de rejeter H_0 / ne pas rejeter H_0.

Region critique W

Etant donné un échantillon, celui-ci donne lieu à des observations et à une valeur pour un certain paramètre θ. **Si θ se trouve dans W, on rejette H_0.** Sinon on ne rejette pas H_0.

W dépend de α

$\overline{W}$ = complémentaire de W.

Démarche à suivre.

(1) On formule H_0 et H_1 avec le fait que ce qui est important est le rejet de H_0.

(2) On fixe α. Probabilité de rejeter H_0 à tort.

(3) On détermine la région critique **W**.

(4) On regarde si la quantité issue des observations se trouve dans W.
Ou encore si θ **est dans W.**

(5) Conclusion : **on rejette ou non H_0.**

rejeter H_0 : conclusion forte
ne pas rejeter H_0 : conclusion faible

2) Comparaison d'une moyenne d'un échantillon avec une moyenne donnée.

On dispose d'un échantillon de taille n sur une population Ω.
Soit la variable aléatoire X sur Ω de moyenne μ et d'écart-type σ.

On souhaite tester μ vis à vis d'une valeur donnée μ_0.

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_1 \end{array}\right\} \text{ test bilatéral.}$$

ici $\mu = \theta$

On suppose $X \sim \mathcal{N}(\mu, \sigma)$ avec σ connu.

1er cas

Par hypothèse $\mu = \mu_0$.

Soient $X_1, \dots, X_n$ n v.a.i avec $X_i \sim \mathcal{N}(\mu_0, \sigma)$.

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \quad \text{moyenne empirique.}$$

Alors $\dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.

On fixe **le seuil α.**

$\rightarrow$ $t_{\alpha/2}$ donné par $\mathcal{N}(0,1)$: probabilités bilatérales.

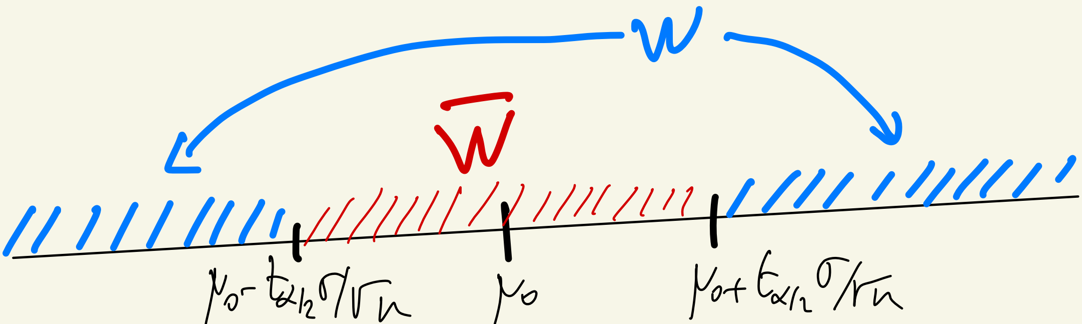
La probabilité que $\left|\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq t_{\alpha/2}$ vaut α.

Ou encore :

$\bar{x}$ est en dehors de $[\mu_0 - t_{\alpha/2}\,\sigma/\sqrt{n} \,;\, \mu_0 + t_{\alpha/2}\,\sigma/\sqrt{n}]$ avec une probabilité α.

Région critique W.

$$\bar{W} = [\mu_0 - t_{\alpha/2}\,\sigma/\sqrt{n} \,;\, \mu_0 + t_{\alpha/2}\,\sigma/\sqrt{n}]$$



Sous H_0, $\bar{x}$ est dans W avec une petite probabilité.

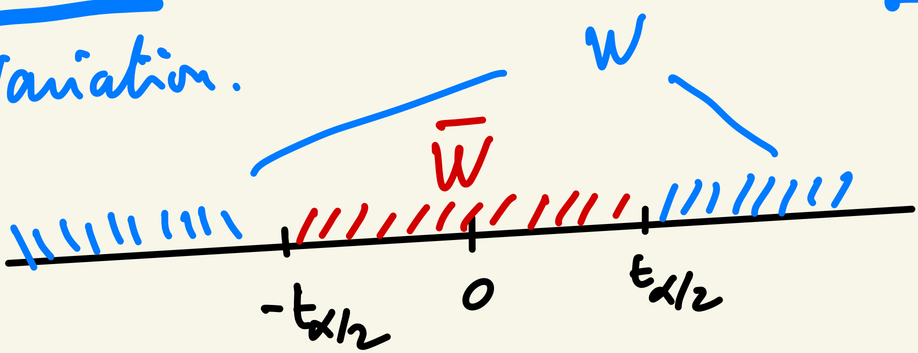
Si x est dans W : on rejette H_0. Risque d'erreur contrôlé par α.

Interprétation. Dans la famille des échantillons de taille n, la proportion de ceux pour lesquels $\bar{x}$ est dans W est égale à α.

Ou encore : il est "rare" que la moyenne $\bar{x}$ d'un échantillon tombe dans W... ainsi, lorsque cela arrive, on part du principe qu'il y a une anomalie et donc que l'hypothèse de départ n'est pas "bonne".

D'où le rejet de H_0.

Variation.



On teste si $\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ appartient à W :

→ Si oui : on rejette H_0.

→ Si non : on ne rejette pas H_0.

2ᵉ cas

8

On suppose $X \sim \mathcal{N}(\mu, \sigma)$ avec σ inconnu

- $H_0 : \mu = \mu_0$; $H_1 : \mu \neq \mu_0$ au seuil α.
- $X_1, \dots, X_n$ n v.a.i telles que $X \sim \mathcal{N}(\mu_0, \sigma)$

Soient $\bar{X} = \dfrac{X_1 + \dots + X_n}{n}$ et $S^2 = (X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2$.

Alors $\dfrac{\bar{X} - \mu_0}{\sqrt{\dfrac{S^2}{n(n-1)}}} \sim St(n-1)$. (vu dans les chapitres précédents)

loi de Student à $n-1$ ddl

- Le test porte sur la quantité $\dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$ où :
    - $\bar{x} = \dfrac{x_1 + \dots + x_n}{n}$.
    - $\hat{s} = \sqrt{\dfrac{1}{n-1}\left((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\right)}$.

- Région $\overline{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$

donné par $St(n-1)$.

On rejette H_0 si $\dfrac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}}$ n'est pas dans W

3e cas

X quelconque, n grand

- $H_0 : \mu = \mu_0 \longrightarrow$ moyenne de X. $\quad H_1 : \mu \neq \mu_0$ au seuil α
- $X_1, \dots, X_n$ n v.a.i avec $X_i \sim X$

On utilise le théorème central limite.

$\bar{X} = \dfrac{X_1 + \dots + X_n}{n}$; $\bar{x}$ = réalisation de $\bar{X}$
= moyenne issue de l'échantillon

Soit $\hat{s} = \sqrt{\dfrac{1}{n-1}(x_1-\bar{x})^2 + \dots + (x_n-\bar{x})^2}$ l'estimation ponctuelle de l'écart-type σ de X.

Alors $\dfrac{\bar{X}-\mu_0}{\hat{s}/\sqrt{n}}$ proche de $\mathcal{N}(0,1)$.

(n grand)

- $\bar{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$.

  ↑ donné par $\mathcal{N}(0,1)$

On rejette H_0 si $\dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$ n'est pas dans $\bar{W}$.

Sinon non rejet de H_0.

Test sur une proportion pour n grand.

$X \sim \mathcal{B}(p)$: loi de Bernoulli.

- $H_0 : p = p_0$; $H_1 : p \neq p_1$ au seuil α.

- Ici $\sigma^2 = p(1-p) = p_0(1-p_0)$ par hypothèse.

Par conséquent $\dfrac{\bar{X} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ proche de $\mathcal{N}(0,1)$

(théorème central limite)

- $\overline{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$

  ↳ donné par $\mathcal{N}(0,1)$.

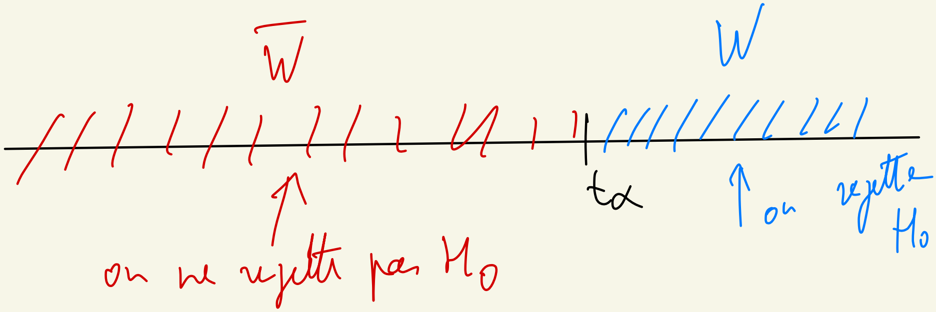
- Si $\dfrac{\bar{x} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ ne tombe pas dans $\overline{W}$, on rejette H_0

---

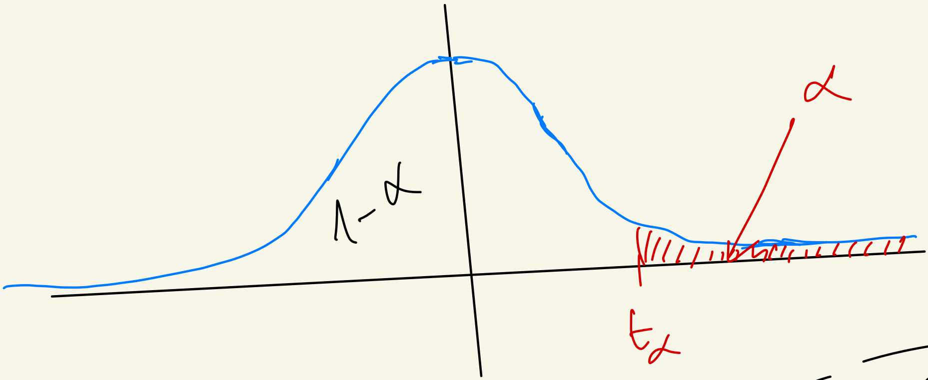
test unilatéral pour une moyenne — n grand.

- $H_0 : \mu \leq \mu_0$; $H_1 : \mu > \mu_0$ au seuil α.

- Région critique : W (à droite)



- t_α est donné par la table de $\mathcal{N}(0,1)$:



x_i : échantillon

- $\bar{x}$ = moyenne de l'échantillon
- $\hat{s} = \sqrt{\frac{1}{n-1}\left((x_1-\bar{x})^2+\dots+(x_n-\bar{x})^2\right)}$ proche de l'écart-type de X

- Par le théorème central limite, on sait que

$$\frac{\bar{X}-\mu}{\hat{s}/\sqrt{n}} \text{ est proche de } \mathcal{N}(0,1).$$ (ici n grand)

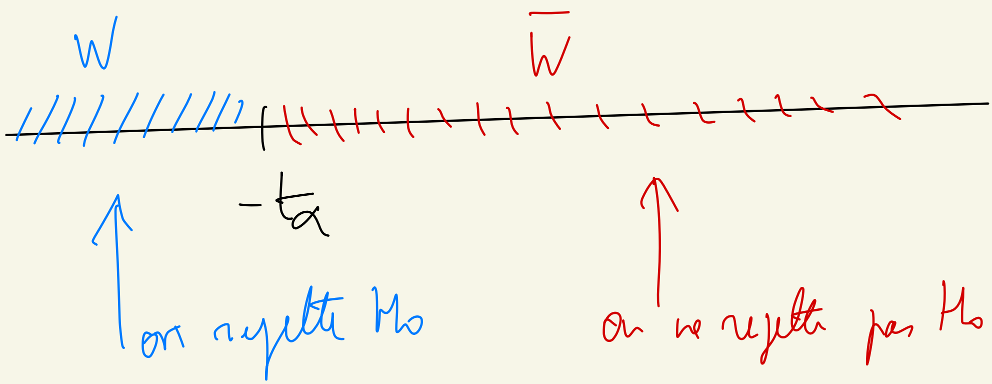
Supposons $\frac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}} \geq t_\alpha$; alors comme $\mu_0 \geq \mu$ il vient :

$$\frac{\bar{x}-\mu}{\hat{s}/\sqrt{n}} \geq t_\alpha .$$

Ainsi cela arrive avec une probabilité plus petite que α donc petite.
"On pense que ce n'est pas possible".
On rejette alors H_0.

Test unilatéral à gauche.

$$H_0 : \mu \geq \mu_0 \quad ; \quad H_1 : \mu < \mu_0$$



14

Exemple. Election entre deux candidats A et B.

p = proportion pour A ; $q = 1-p$ = proportion pour B.

Un sondage sur 800 personnes aboutit à $\bar{p} = 0{,}48$

$\bar{q} = 0{,}52$.

Seuil $\alpha = 5\% = 0{,}05$.

- Peut-on affirmer que $p < 0{,}5$?

$H_0 : p \geq 0{,}5$; $H_1 : p < 0{,}5$. **Ici : $p_0 = 0{,}5$.**

(à gauche)

$\alpha = 0{,}05 \longrightarrow t_\alpha = 1{,}64$.

Région critique $W =]-\infty ; -1{,}64]$.

Le test porte sur : $\dfrac{\bar{p} - 0{,}5}{s/\sqrt{800}}$ avec $s = \sqrt{0{,}48 \times 0{,}52}$.

$\bar{p} = 0{,}48$

$\parallel$

$-1{,}13$: On ne rejette pas H_0

- Que peut-on dire si le sondage porte sur $n = 2000$ avec $\bar{p} = 0{,}48$ et $\bar{q} = 0{,}52$?

ici le test porte sur $\dfrac{0{,}48 - 0{,}50}{s/\sqrt{2000}} \simeq -1{,}79$

On rejette H_0 : $p < 0{,}5$ avec un risque d'erreur de 5%.

3) Comparaison de deux moyennes.

On part de deux populations A et B et de deux échantillons extraits de ces populations. Soient X et Y deux v.a.i. sur chacune de ces populations.

On cherche à comparer μ_A et μ_B.

moyenne de X

moyenne de Y.

σ_A : écart-type de X

σ_B : écart-type de Y

n_A : taille de l'échantillon issu de A

n_B : ——— B

$\bar{X} = \dfrac{X_1 + \cdots + X_{n_A}}{n_A}$ où $X_1, \ldots, X_{n_A}$ n_A v.a.i. $X_i \sim X$.

$\bar{Y} = \dfrac{Y_1 + \cdots + Y_{n_B}}{n_B}$ où $Y_1, \ldots, Y_{n_B}$ n_B v.a.i. $Y_i \sim Y$.

$$\boxed{\text{On pose } \bar{Z} = \bar{X} - \bar{Y}}$$

$\bar{X}$ et $\bar{Y}$ sont indépendantes

Alors $\boxed{\mu_{\bar{Z}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_A - \mu_B \text{ et } Var(\bar{Z}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$.

$H_0 : \mu_A = \mu_B$; $H_1 : \mu_A \neq \mu_B$.

$\boxed{\text{Sous } H_0,\ \mu_{\bar{Z}} = 0}$.

1^er cas : $\boxed{X \sim \mathcal{N}(\mu_A, \sigma_A) ; Y \sim \mathcal{N}(\mu_B, \sigma_B) \text{ avec } \sigma_A \text{ et } \sigma_B \text{ connus}}$

- $\bar{X} \sim \mathcal{N}(\mu_A, \sigma_A/\sqrt{n_A})$ et $\bar{Y} \sim \mathcal{N}(\mu_B, \sigma_B/\sqrt{n_B})$

Alors $\bar{Z} = \bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$.

Sous H_0, $\bar{Z} \sim \mathcal{N}\left(0, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$.

Ou encore :

$$\boxed{\frac{\bar{Z}}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim \mathcal{N}(0,1).}$$

- le test porte sur $\dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \bar{e}$

$\bar{x}$ = moyenne de l'échantillon de A

$\bar{y}$ = moyenne de l'échantillon de B

- **Région critique** $\bar{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$

par $\mathcal{N}(0,1)$.

Si $\bar{e}$ ne tombe pas dans $\bar{W}$, on rejette H_0 au seuil α

2ᵉ cas

$$\boxed{\begin{array}{c}X \text{ et } Y \text{ quelconques} \\ n_A \text{ et } n_B \text{ grands}\end{array}}$$

- On estime ponctuellement σ_A et σ_B par :

$$\hat{\Delta}_A = \sqrt{\frac{1}{n_A - 1}\left[(x_1-\bar{x})^2 + \cdots + (x_{n_A}-\bar{x})^2\right]}$$

(x_i = échantillon de A)

$$\hat{\Delta}_B = \sqrt{\frac{1}{n_B - 1}\left((y_1-\bar{y})^2 + \cdots + (y_{n_B}-\bar{y})^2\right)}$$

(y_i = échantillon de B.)

- Par le théorème central limite.

$\bar{X}$ proche de $\mathcal{N}(\mu_A, \hat{\Delta}_A/\sqrt{n_A})$

et

$\bar{Y}$ proche de $\mathcal{N}(\mu_B, \hat{\Delta}_B/\sqrt{n_B})$

Comme pour le cas précédent, on en déduit que $Z = \bar{X} - \bar{Y}$ est proche de :

$$\mathcal{N}\left(0, \sqrt{\frac{\hat{\Delta}_A^2}{n_A} + \frac{\hat{\Delta}_B^2}{n_B}}\right)$$

- **Région critique** : $\bar{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$

de $\mathcal{N}(0,1)$

$$\boxed{\text{Si } \bar{e} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{\Delta}_A^2}{n_A} + \frac{\hat{\Delta}_B^2}{n_B}}} \text{ ne tombe pas dans } \bar{W}, \text{ on rejette } H_0 \text{ au seuil } \alpha}$$

Lorsque σ_A et σ_B sont inconnus, que n_A ou n_B est petit mais que X et Y suivent des lois normales, il est possible d'effectuer un test. Cela va faire intervenir la loi de Student.

3e cas

$$\boxed{\begin{array}{l} X \sim \mathcal{N}(\mu_A, \sigma) \\ Y \sim \mathcal{N}(\mu_B, \sigma) \\ \text{avec } \sigma \text{ inconnu.} \end{array}}$$

$\sigma_A = \sigma_B$

- $H_0 : \mu_A = \mu_B$; $H_1 : \mu_A \neq \mu_B$.
- Sous H_0, $\bar{Z} = \bar{X} - \bar{Y} \sim \mathcal{N}\left(0, \sigma\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right)$

$$\frac{\bar{Z}}{\sigma\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim \mathcal{N}(0,1).$$

problème : "simplifier" σ

- Soit $S_A^2 = (X_1 - \bar{X})^2 + \cdots + (X_{n_A} - \bar{X})^2$

$S_B^2 = (Y_1 - \bar{Y})^2 + \cdots + (Y_{n_B} - \bar{Y})^2$.

"variances empiriques"

(21)

Alors $\frac{S_A^2}{\sigma^2} \sim \chi^2(n_A - 1)$ et $\frac{S_B^2}{\sigma^2} \sim \chi^2(n_B - 1)$.

Soit $T = S_A^2 + S_B^2$.

Alors $\boxed{\frac{T}{\sigma^2} \sim \chi^2(n_A + n_B - 2)}$.

Ainsi

Student

$$\boxed{\frac{\dfrac{\bar{Z}}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}}{\sqrt{S_A^2 + S_B^2}} \times \sqrt{n_A + n_B - 2} \sim St(n_A + n_B - 2)}$$

• Région critique : $\bar{W} = [-t_{\alpha/2} \,;\, t_{\alpha/2}]$

$St(n_A + n_B - 2)$

Si

$$\bar{t} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \sqrt{\frac{(n_A-1)\hat{s}_A^2 + (n_B-1)\hat{s}_B^2}{n_A+n_B-2}}}$$

ne tombe pas dans $\bar{W}$, on rejette H_0.

où

$$\hat{s}_A^2 = \frac{1}{n_A-1}\left((x_1-\bar{x})^2 + \cdots + (x_{n_A}-\bar{x})^2\right)$$

$$s_B^2 = \frac{1}{n_B-1}\left((y_1-\bar{y})^2 + \cdots + (y_{n_B}-\bar{y})^2\right)$$

22

$j = 400$

$X \sim \mathcal{N}(\mu_A, \sigma_A)$

$Y \sim \mathcal{N}(\mu_B, \sigma_B)$

σ_A et σ_B inconnus

- $H_0 : \mu_A = \mu_B$; $H_1 : \mu_A \neq \mu_B$.
- On effectue un test sur la quantité :

$$\bar{e} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{s}_A^2}{n_A} + \frac{\hat{s}_B^2}{n_B}}}$$

où
$$\hat{s}_A^2 = \frac{1}{n_A - 1}\left((x_1 - \bar{x})^2 + \dots + (x_{n_A} - \bar{x})^2\right)$$
$$\hat{s}_B^2 = \frac{1}{n_B - 1}\left((y_1 - \bar{y})^2 + \dots + (y_{n_A} - \bar{y})^2\right)$$

- Région critique. $\bar{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$

$$\chi \sim St(N)$$

où N est l'entier le plus proche de

$$\frac{\left(\frac{\hat{\Delta}_A^2}{n_A} + \frac{\hat{\Delta}_B^2}{n_B}\right)^2}{\frac{\left(\frac{\hat{\Delta}_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{\hat{\Delta}_B^2}{n_B}\right)^2}{n_B - 1}}.$$

4) Comparaison de deux moyennes : populations appariées

- Soient X et Y deux variables aléatoires sur une population Ω.

- On observe n couples d'échantillons $(x_1, y_1), \dots, (x_n, y_n)$ où $x_i = X(\omega_i)$ et $y_i = Y(\omega'_i)$.

  Ici $\omega_1, \dots, \omega_n$ sont des individus de Ω.
  $\omega'_1, \dots, \omega'_n$

  Eventuellement $\boxed{\omega_i = \omega'_i}$.

- Mathématiquement c'est considérer n couples de v.a. : $(X_1, Y_1), \dots, (X_n, Y_n)$ où :
  - les X_i sont des v.a.i et $X_i \sim X$
  - les Y_i sont des v.a.i et $Y_i \sim Y$
  - X_i et Y_i peuvent ne pas être indépendantes

Exemples.

- Evaluer sur un même individu $\underbrace{\text{la taille}}_{X}$ et $\underbrace{\text{le poids}}_{Y}$

Ici $\omega_i = \omega'_i$

- Pour un couple marié (H–F), évaluer la taille des deux individus.

X : taille de H.

Y : taille de F.

Ici $\omega_i \neq \omega'_i$

- Evaluer la consommation d'essence d'une voiture : $\underbrace{\text{à la sortie}}_{X}$; $\underbrace{\text{après 10 ans}}_{Y}$.

Ici $\omega_i = \omega'_i$

On pose :
- μ_X et μ_Y les moyennes de X et de Y
- $Z_i = X_i - Y_i$. Alors $\mu_{Z_i} = \mu_X - \mu_Y$.
- $\bar{X} = \dfrac{X_1 + \dots + X_n}{n}$ et $\bar{Y} = \dfrac{Y_1 + \dots + Y_n}{n}$
- $\bar{Z} = \dfrac{(X_1 - Y_1) + \dots + (X_n - Y_n)}{n} = \bar{X} - \bar{Y}$.

Les variables aléatoires $Z_1, \dots, Z_n$ sont indépendantes.

- Mais attention on ne connaît pas a priori la variance (ou l'écart-type) de Z_i.

  En effet :

  $$\mathrm{var}(Z_i) = \mathrm{var}(X_i - Y_i) = \mathrm{Var}(X_i) + \mathrm{Var}(Y_i) + 2\,\mathrm{cov}(X_i; Y_i)$$

  Covariance. Peut être non nulle car X et Y non indépendantes.

Notons σ = écart-type de X−Y.

$$\boxed{\sigma = \sigma_{X-Y} = \sigma_Z}$$

Ici on pose $Z = X - Y$

X et Y quelconques
n grand

1 cas

On utilise le théorème central limite.

- On commence par estimer σ.

$$\hat{S}^2 = \frac{1}{n-1}\left((z_1-\bar{z})^2+\dots+(z_n-\bar{z})\right)$$: bon estimateur de $\sigma_Z = \sigma$.

$$\hat{\Delta}^2 = \frac{1}{n-1}\left(\left(x_1-y_1-(\bar{x}-\bar{y})\right)^2+\dots+\left(x_n-y_n-(\bar{x}-\bar{y})\right)^2\right)$$

$$= \frac{1}{n-1}\left(\left(x_1-\bar{x}-(y_1-\bar{y})\right)^2+\dots+\left(x_n-\bar{x}-(y_n-\bar{y})\right)^2\right)$$

Alors $\hat{\Delta}^2$ est une bonne estimation de σ^2.

- Par le théorème central limite,

$\bar{Z}$ proche de $\mathcal{N}(\mu_X-\mu_Y, \sigma/\sqrt{n})$

- Hypothèse $H_0 : \mu_X = \mu_Y$ (test bilatéral)

$H_1 : \mu_X \neq \mu_Y$

Sous H_0, $\bar{Z}$ proche de $\mathcal{N}(0, \hat{\Delta}/\sqrt{n})$

ou encore $\dfrac{\bar{Z}}{\hat{\Delta}/\sqrt{n}}$ proche de $\mathcal{N}(0,1)$.

- **Région critique**

$\overline{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$.

donné par $\mathcal{N}(0,1)$.

Statistique $\bar{c} = \dfrac{\bar{x}-\bar{y}}{\hat{\Delta}/\sqrt{n}}$

avec $\hat{\Delta} = \sqrt{\dfrac{1}{n-1}\left((x_1-y_1-(\bar{x}-\bar{y}))^2 + \dots + (x_n-y_n-(\bar{x}-\bar{y}))^2\right)}$.

Si $\bar{c}$ tombe dans $\overline{W}$, on ne rejette pas H_0

Si $\bar{c}$ ne tombe pas dans $\overline{W}$, on rejette H_0.

$X \sim \mathcal{N}(\mu_x, \sigma_x)$
$Y \sim \mathcal{N}(\mu_y, \sigma_y)$
n petit

2^e cas

- $X_1, \dots, X_n$ n v.a.i. $X_i \sim X$
  $Y_1, \dots, Y_n$ n v.a.i $Y_i \sim X$

  $Z_i = X_i - Y_i$.

  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$; $\bar{Y} = \frac{Y_1 + \dots + Y_n}{n}$; $\bar{Z} = \frac{Z_1 + \dots + Z_n}{n}$

- Alors $Z_i \sim \mathcal{N}(\mu_x - \mu_y, \sigma)$

  où $\sigma = \sigma_Z = \sigma_{X-Y}$.

Ainsi $\bar{Z} \sim \mathcal{N}(\mu_x - \mu_y, \sigma/\sqrt{n})$.

(on rappelle que l'on ne connaît pas a priori σ).

il faut "éliminer" σ.

- Soit $S^2 = (Z_1 - \bar{Z})^2 + \dots + (Z_n - \bar{Z})^2$.

$$\boxed{\begin{array}{l}\text{Alors} \\ \dfrac{\bar{Z} - (\mu_X - \mu_Y)}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \\ \text{et} \\ \dfrac{S^2}{\sigma^2} \sim \chi^2(n-1)\end{array}}$$

Par conséquent :

$$\boxed{\dfrac{\bar{Z} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{S^2}{n(n-1)}}} \sim St(n-1)}$$

loi de Student à $n-1$ ddl.

- $H_0 : \mu_x = \mu_y$; $H_1 : \mu_x \neq \mu_y$

(test bilatéral)

Sous H_0,
$$\frac{\bar{Z}}{\sqrt{\frac{S^2}{n(n-1)}}} \sim St(n-1).$$

- Région critique

$$\bar{W} = [-t_{\alpha/2} ; t_{\alpha/2}]$$

St(n−1)

Soit $\bar{c} = \dfrac{\bar{x}-\bar{y}}{\hat{s}/\sqrt{n}}$

où $\hat{s} = \sqrt{\frac{1}{n-1}\left((z_1-\bar{z})^2 + \dots + (z_n-\bar{z})^2\right)}$.

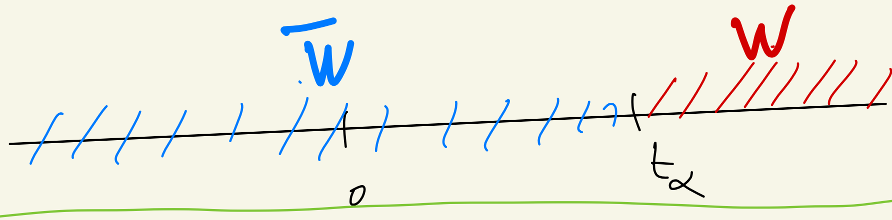
Si $\bar{c}$ ne tombe pas dans $\bar{W}$, on rejette H_0.
Si $\bar{c}$ tombe dans $\bar{W}$, on ne rejette pas H_0.

test unilatéral

1er cas : test à droite

- $H_0 : \mu_x \leq \mu_y$; $H_1 : \mu_x > \mu_y$. (au seuil α)

- $W =]t_\alpha ; +\infty[$; $\overline{W} =]-\infty ; t_\alpha[$.



Ici t_α est donné par $\mathcal{N}(0,1)$ si n est grand, ou par $St(n-1)$ si n est petit.

- la probabilité que $\dfrac{\bar{z} - (\mu_x - \mu_y)}{\hat{s}/\sqrt{n}}$ tombe dans W vaut α.

Or sous H_0, $\mu_x - \mu_y \leq 0$, et donc

$$\frac{\bar{z}}{\hat{s}/\sqrt{n}} \leq \frac{\bar{z}-(\mu_x-\mu_y)}{\hat{s}/\sqrt{n}}.$$

Ainsi « $\frac{\bar{z}}{\hat{s}/\sqrt{n}} \geq t_\alpha$ » arrive avec une probabilité plus petite que α.

Soit $\bar{c} = \frac{\bar{z}}{\hat{s}/\sqrt{n}} = \frac{\bar{x}-\bar{y}}{\hat{s}/\sqrt{n}}$.

Si $\bar{c}$ tombe dans W, on rejette H_0.

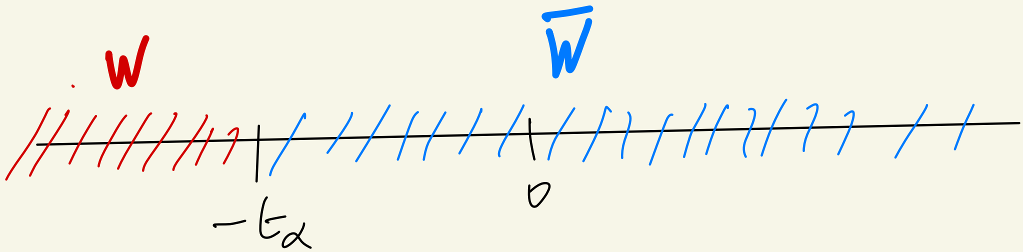
Si c tombe dans $\bar{W}$, on ne rejette pas H_0.

2e cas Test à gauche.

- $H_0 : \mu_X \geq \mu_Y$; $H_1 : \mu_X < \mu_Y$.

(se déduit du 1er cas)

• Région critique (seuil α)



• $\bar{t} = \dfrac{\bar{x}-\bar{y}}{\hat{s}/\sqrt{n}}$

Si $\bar{t}$ tombe dans W, on rejette H_0.

Si $\bar{t}$ tombe dans $\bar{W}$, on ne rejette pas H_0.

EXEMPLE

• Une étude porte sur les salaires H/F à travers 50 couples.

Soit X la v.a. qui indique le salaire de l'homme

Soit Y ———————— la femme.

- $(x_1, y_1) = (2752, 2253) \leadsto z_1 = x_1 - y_1 = 499$

  $(x_2, y_2) = (2893, 1790) \longrightarrow z_2 = 1103$

  $(x_3, y_3) = (1064, 1212) \longrightarrow z_3 = -148$

  $\vdots$

  $(x_{50}, y_{50}) = (1808, 1287) \longrightarrow z_{50} = 521$

X	Y	Z
$\bar{x} = 1941$	$\bar{y} = 1644$	$\bar{z} = 297$
$\sigma_x = 1146$	$\sigma_y = 925$	$\sigma_z = 515$

écarts-types de l'échantillon de taille 50.

1^e test — On compare les moyennes μ_x et μ_y.

- $H_0 : \mu_x = \mu_y$; $\mu_x \neq \mu_y$
- $\alpha = 5\%$; $t_{\alpha/2} = 1{,}96$ (issu de $\mathcal{N}(0,1)$)
  $n \geqslant 30$.
- $\overline{W} = [-1{,}96 ; 1{,}96]$.
- Ici la statistique est la quantité :

$$C = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n}}} \simeq 1{,}42$$

où $\hat{s}_x \simeq \sigma_x$
$\hat{s}_y \simeq \sigma_y$

On observe que C est dans $\overline{W}$:

On ne rejette pas H_0.

2nd test : Test bilatéral sur des échantillons appariés

- $H_0 : \mu_x = \mu_y$; $H_1 : \mu_x \neq \mu_y$.
- Région critique au seuil α : $t_{\alpha/2} = 1,96$.
- Ici $\sigma_z \simeq 515$.

La statistique est la quantité :

$$\tilde{c} = \frac{\bar{z}}{\sigma_z/\sqrt{n}} = \frac{297}{515/\sqrt{50}} \simeq 4,07\ldots$$

On observe que $\tilde{c}$ ne tombe pas dans $\bar{W} = [-1,96 ; 1,96]$.

on rejette H_0 ; on accepte $H_1 : \mu_x \neq \mu_y$.

3° test — Test unilatéral sur des échantillons appariés.

- $H_0 : \mu_X \leq \mu_Y$; $H_1 : \mu_X > \mu_Y$

H F.

- $\alpha = 5\%$; $t_\alpha = 1,64$.

- Région critique.

$$W = [1,64 ; +\infty[$$

- $\bar{c} = \dfrac{\bar{z}}{\sigma_z / \sqrt{n}} \simeq 4,07$

$\bar{c}$ est dans W : On rejette H_0.

On accepte $H_1 : \mu_X > \mu_Y$.

« Dans un couple, les salaires des hommes sont supérieurs à celui des femmes »

- Prenons $\alpha = 0{,}002\,\% = 0{,}00002$.

$$t_\alpha = 4{,}08 \longrightarrow \mathcal{N}(0,1).$$

Ici $\bar{t} = 4{,}07$ tombe dans $\overline{W} = [-t_\alpha\,;\,t_\alpha]$; on ne rejette pas $H_0 : \mu_x \leq \mu_y$.

5) Comparaison de deux variances

Pour le test de comparaison de deux moyennes on a vu l'importance de l'hypothèse $\sigma_A = \sigma_B$ (pour de petits effectifs) quand X et Y suivent des lois normales.

Ici on propose un protocole pour tester l'hypothèse $H_0 : \sigma_A = \sigma_B$; $H_1 : \sigma_A \neq \sigma_B$.

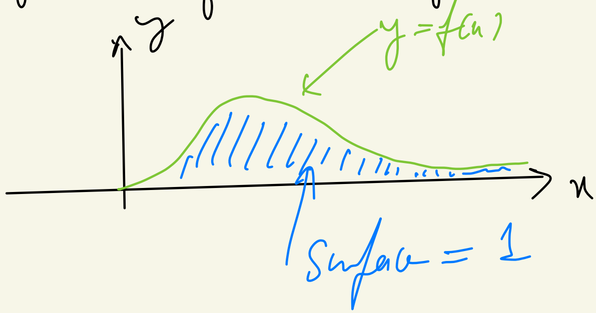
a) La loi de Fisher-Snedecor

- Soient deux entiers $m, n \geqslant 1$.

  On pose

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \dfrac{\Gamma(n+m)}{\Gamma(n/2)\Gamma(m/2)} \left(\dfrac{n}{m}\right)^{n/2} \dfrac{x^{n/2 - 1}}{\left(1 + \frac{n}{m}x\right)^{\frac{n+m}{2}}} \end{cases}$$

- La fonction f est une fonction à densité !



- On note par $F(m,n)$ la loi associée : c'est la loi de Fisher-Snedecor.

- On a $\mu_F = m/_{m-3}$ (pour $m \geqslant 4$) ;

$$\sigma_F = \sqrt{\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}} \quad (\text{pour } m \geqslant 5).$$

Résultats clefs.

① Si $X \sim F(m,n)$ alors $\frac{1}{X} \sim F(n,m)$.

② Si $X_1 \sim \chi^2(m)$ et $X_2 \sim \chi^2(n)$ alors

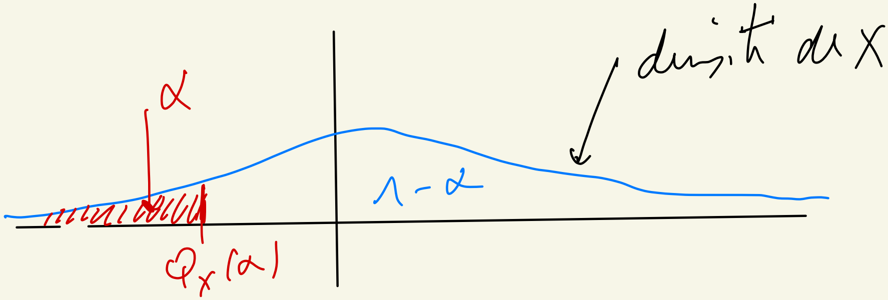
$$\frac{X_1/m}{X_2/n} \sim F(m,n).$$

b) Quantiles

- Pour une loi X on note par $Q_X(\alpha)$ le nombre tel que

$$P(X \leq Q_X(\alpha)) = \alpha.$$

Le nombre $Q_X(\alpha)$ est le **αième-quantile.**



- Ainsi

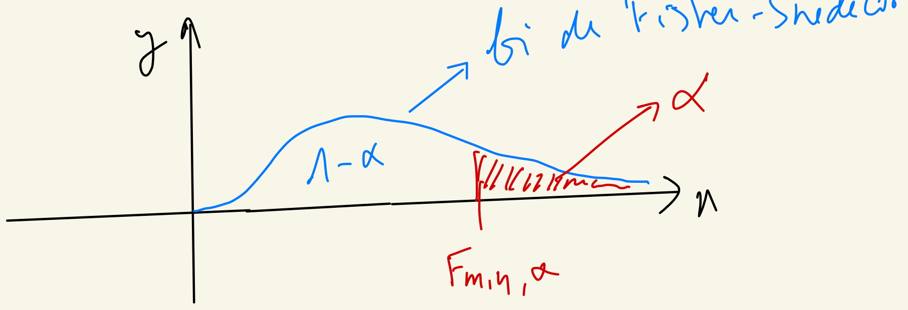
$$P(X > Q_X(1-\alpha)) = 1 - P(X \leq Q_X(1-\alpha))$$
$$= 1 - (1-\alpha) = \alpha$$

- **Résultat**

$$\boxed{Q_X(1-\alpha) = \frac{1}{Q_{1/X}(\alpha)}.}$$ (facile)

- On pose

$$\boxed{F_{m,n,\alpha} = Q_{F_{m,n}}(1-\alpha).}$$ → loi de Fisher-Snedecor



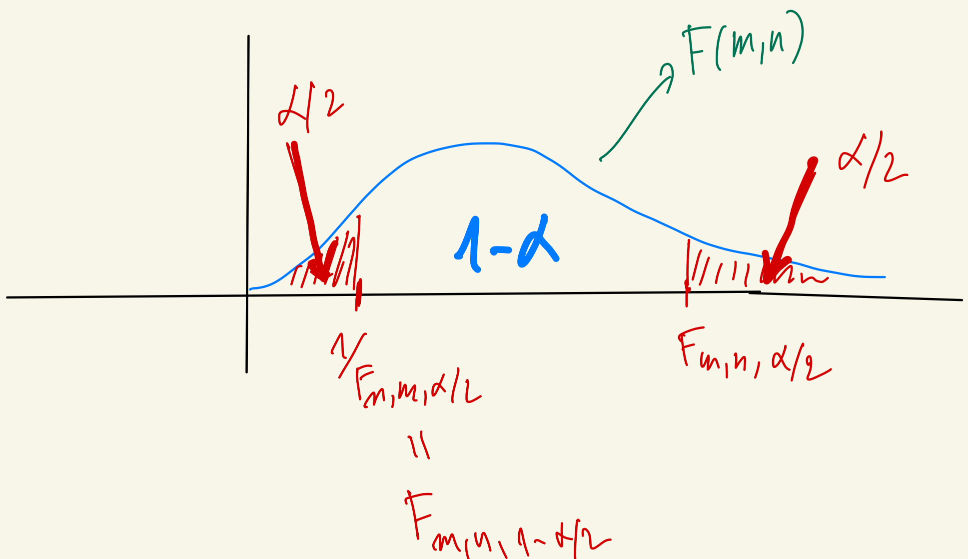
- Ainsi

$$F_{m,n,\alpha} = Q_{F_{m,n}}(1-\alpha) = \frac{1}{Q_{F_{n,m}}(\alpha)} = \frac{1}{F_{n,m,1-\alpha}}.$$

$$\boxed{F_{m,n,\alpha} = \frac{1}{F_{n,m,1-\alpha}}}$$

- Pour α petit, la quantité $F_{m,n,\alpha}$ est plus grande que 1.

  Ainsi $F_{m,n,1-\alpha}$ est plus petite que 1.



- Soit la zone critique

$$\bar{W} = \left[\frac{1}{F_{n_1, m, \alpha/2}} \,; F_{m, n, \alpha/2} \right]$$

Soit le nombre C. Supposons $C \geq 1$.

$$\boxed{\text{Alors } C \text{ tombe dans } \bar{W} \text{ si et seulement si } C \leq F_{m,n,\alpha/2}.}$$

c) test d'égalité sur l'écart-type.

- Soient X et Y deux v.a. **normales** sur A et B.
  On note par σ_A l'écart-type de X,
  σ_B ——— Y.

- Soient $X_1, \dots, X_{n_A}$ n_A v.a.i. avec $X_i \sim X$.
  $Y_1, \dots, Y_{n_B}$ n_B v.a.i avec $Y_i \sim Y$.

- Ici n_A et n_B sont petits.

- On veut tester l'hypothèse H_0 suivante :

$$H_0 : \sigma_A = \sigma_B \quad ; \quad H_1 : \sigma_A \neq \sigma_B .$$

- $X \sim \mathcal{N}(\mu_A, \sigma_A)$; $Y \sim \mathcal{N}(\mu_B, \sigma_B)$

$$S_A^2 = (X_1 - \overline{X})^2 + \dots + (X_{n_A} - \overline{X})^2$$

$$S_B^2 = (Y_1 - \overline{Y})^2 + \dots + (Y_{n_B} - \overline{Y})^2$$

On a vu :

$$\frac{S_A^2}{\sigma_A} \sim \chi^2(n_A - 1)$$

$$\frac{S_B^2}{\sigma_B} \sim \chi^2(n_B - 1)$$

loi du chi-deux.

- Sous H_0, $\sigma_A = \sigma_B$, et ainsi :

$$\boxed{\frac{S_A^2/n_A-1}{S_B^2/n_B-1} \sim F(n_A-1, n_B-1).}$$

(Sous H_0, σ_A et σ_B se simplifient).

- **Région critique au seuil α.**

$$\bar{W} = \left[\frac{1}{F_{n_B-1, n_A-1, \alpha/2}} \; ; \; F_{n_A-1, n_B-1, \alpha/2} \right]$$

- La statistique $\bar{C}$ est la quantité :

$$\boxed{\bar{C} = \frac{\hat{\Delta}_A^2}{\hat{\Delta}_B^2}}$$

$$\hat{\Delta}_A^2 = \frac{1}{n_A-1}\left((x_1-\bar{x})^2 + \dots + (x_{n_A}-\bar{x})^2\right)$$

$$\hat{\Delta}_B^2 = \frac{1}{n_B-1}\left((y_1-\bar{y})^2 + \dots + (y_{n_B}-\bar{y})^2\right)$$

- Supposons $\bar{c} \geq 1$ (ou encore $\hat{\sigma}_A^2 \geq \hat{\sigma}_B^2$).

$\boxed{\text{Alors } \bar{c} \text{ tombe dans } \overline{W} \text{ si et seulement si } \bar{c} \leq F_{n_A-1,\, n_B-1,\, \alpha/2}.}$

Si $\bar{c} < 1$, $\bar{c}$ tombe dans $\overline{W}$ si et seulement si

$$\frac{1}{\bar{c}} \geq F_{n_B-1,\, n_A-1,\, \alpha/2}.$$

- Conclusion.

Si $\bar{c}$ ne tombe pas dans $\overline{W}$, on rejette H_0.
Si $\bar{c}$ tombe dans $\overline{W}$, on ne rejette pas H_0.

<u>Exemple</u> $\alpha = 10\% = 0,1$

$\alpha/2 = 0,05$

$F_{3,5,\alpha/2} = 5,409.$