Chapitre 6 TESTS STATISTIQUES

A. Principe des tests

Partons d'un exemple...Une machine fabrique des tiges d'acier. Si la machine est réglée correctement, l'utilisateur obtient une population de tiges de longueurs moyenne m et d'écart-type σ. On désire savoir si cette machine se dérègle. Ainsi, on prélèvera, à intervalles réguliers, des échantillons pour mesurer la longueur effective des tiges.
Nous faisons alors l'hypothèse H_0 dite hypothèse nulle que la machine est bien réglée. On teste alors cette hypothèse: 2 cas se présentent :

- la machine est bien réglée, on accepte H_0.
- la machine est mal réglée, on rejette H_0 et donc on accepte H_1 dite hypothèse alternative.

Définition : un test statistique est une méthode permettant de prendre une **décision** à partir d'informations fournies par un **échantillon**.
Cette décision dépend donc de l'échantillon. Ainsi qu'elle que soit la décision prise, on court deux sortes de risques :

- le risque dit de 1ère espèce noté α, est la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie en réalité : $\alpha = p(\text{rejeter } H_0 / H_0 \text{ vraie})$
- le risque dit de 2nde espèce noté β, est la probabilité d'accepter l'hypothèse H_0 alors qu'elle est fausse en réalité : $\beta = p(\text{accepter } H_0 / H_0 \text{ fausse})$.

Un test est bon si on arrive à minimiser α et β.

B. Test de comparaison à une valeur standard

1. Position de problème

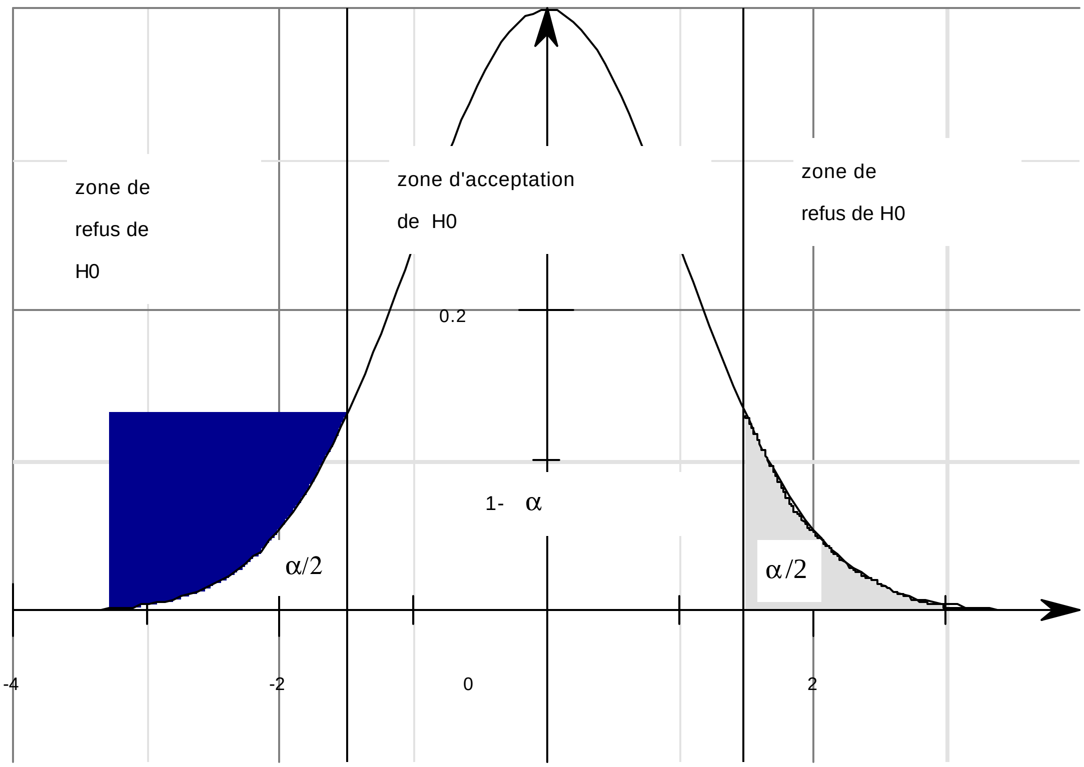
On considère une population P sur laquelle on veut étudier un paramètre γ inconnue associé à un paramètre c. Sur un échantillon de taille n, on obtient γ_e connu. Sur la base de cette valeur observée γ_e, on se propose de comparer la vraie valeur γ à une valeur γ_0 fixée à priori, constituant la valeur standard.

2. Tests relatifs à une moyenne

Soit une population P de grand effectif sur laquelle on étudie un caractère c. La moyenne m de c est inconnue. Sur un échantillon, on a trouvé une moyenne $\bar{x}_e$. On doit tester la moyenne m par rapport à une valeur notée m_0 qui est la valeur standard.

a) Test bilatéral

Soit H_0 : " $m=m_0$ " l'hypothèse nulle et H_1 : " $m \quad m_0$ " l'hypothèse alternative
Soit X la variable aléatoire prenant pour valeurs les moyennes des différents échantillons de taille n   30, alors on sait que X suit une $N(m_0 ; \frac{\sigma}{\sqrt{n}})$, σ étant l'écart-type de la population P.



Il faut donc que X soit telle que $p(m_0 - t_\alpha \frac{\sigma}{\sqrt{n}} < X < m_0 + t_\alpha \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$ d'où en faisant le changement de variable : $T = \dfrac{X - m_0}{\frac{\sigma}{\sqrt{n}}}$ alors T suit une loi normale centrée réduite $N(0,1)$ d'où

$p(-t_\alpha < T < t_\alpha) = 1 - \alpha$ c'est à dire $\pi(t_\alpha) = 1 - \frac{\alpha}{2}$ d'où la règle du test bilatéral :

- On choisit un risque α
- on cherche dans la table de la loi normale centrée réduite $N(0,1)$, t_α tel que $\pi(t_\alpha) = 1 - \frac{\alpha}{2}$
- soit $\bar{x}_e$ la moyenne de l'échantillon de taille n alors

si $\bar{x}_e \in \left[m_0 - t_\alpha \frac{\sigma}{\sqrt{n}} ; m_0 + t_\alpha \frac{\sigma}{\sqrt{n}} \right]$, on accepte H_0 avec le risque α

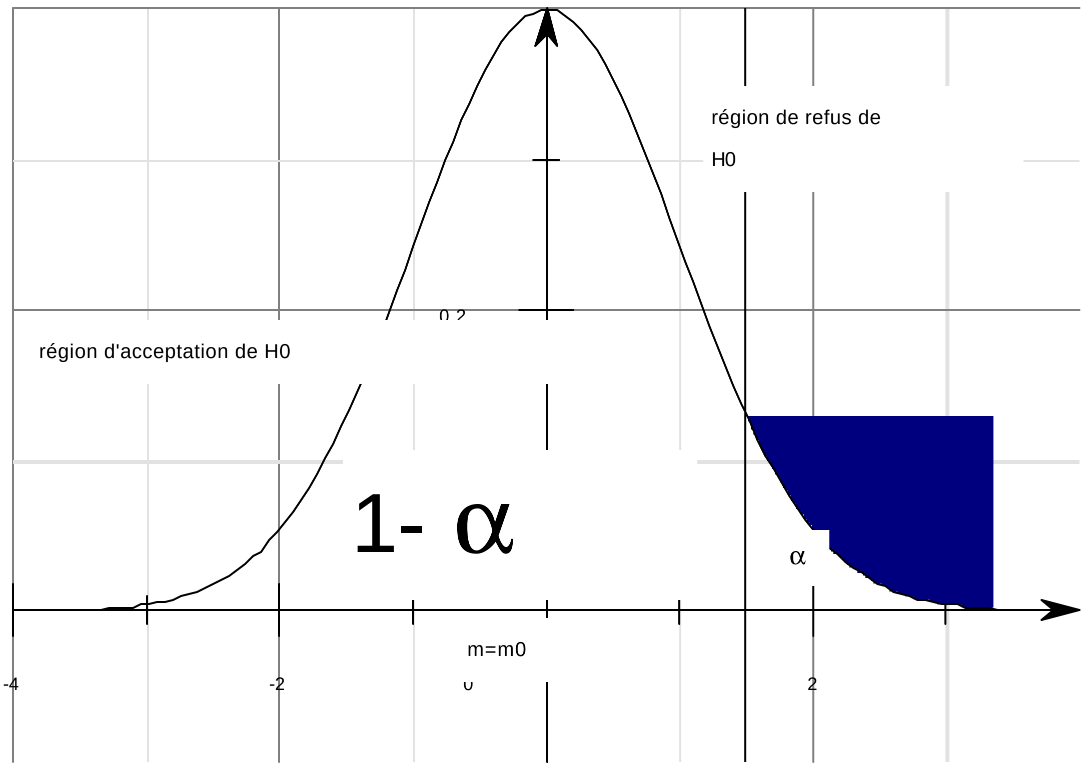
sinon on rejette H_0 et donc on accepte H_1 avec un risque α.

Remarque : dans le cas usuel où α = 5% alors t_α = 1,96 et si α = 1% alors t_α = 2,58.

Si σ est inconnu (ce qui est souvent le cas) alors on prend son estimateur ponctuel $\sqrt{\frac{n}{n-1}}\sigma_e$ où σ_e est l'écart-type de l'échantillon.

b) <u>Tests unilatéraux</u>

Règle du test unilatéral à gauche

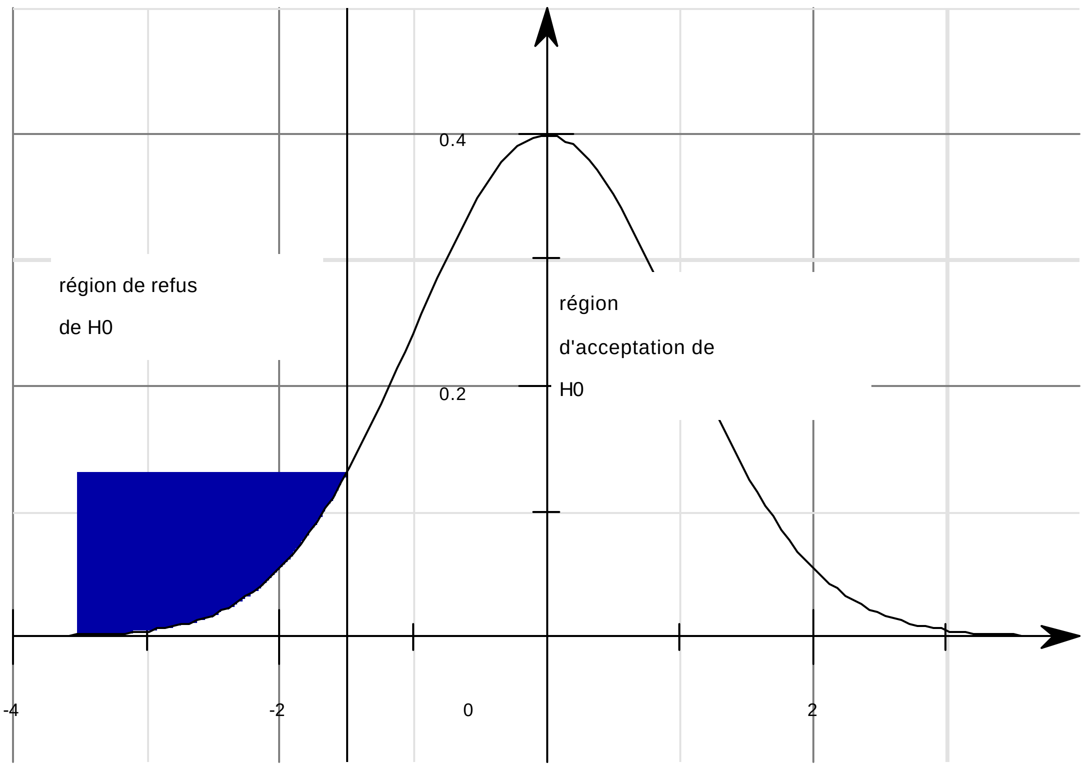


L'hypothèse nulle est H_0 : "m = m_0" et l'hypothèse alternative est H_1 : "m > m_0"
On peut la retrouver par exemple dans le cas d'un test de dépassement d'une norme.

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $N(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha$
- Si $\bar{x}_e \leq m_0 + t_\alpha \frac{\sigma}{\sqrt{n}}$ alors on accepte H_0 sinon on refuse H_0 et donc on accepte H_1

Rem : dans les cas usuels : si α = 5% alors t_α = 1,645 ; si α = 1% alors t_α = 2,33

Règle du test unilatéral à droite



L'hypothèse nulle est : H_0 : "$m = m_0$" et l'hypothèse alternative est : H_1 : "$m < m_0$"
On la retrouve dans les cas de tests de non égalité d'une norme.

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $N(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha$
- Si $\bar{x}_e > m_0 - t_\alpha \dfrac{\sigma}{\sqrt{n}}$ on accepte H_0, sinon on rejette H_0 et donc on accepte H_1

Dans ces deux cas, il est très fréquent qu'on ne connaisse pas σ ; on a alors $\dfrac{\sigma}{\sqrt{n}} = \dfrac{\sigma_e}{\sqrt{n-1}}$

3. Tests relatifs à un fréquence ou un pourcentage

Tous les tests que l'on vient de voir restent valables ; il suffit de remplacer m par p (proportion inconnue dans la population P), $\bar{x}_e$ par f_e (proportion effective sur l'échantillon) et $\dfrac{\sigma}{\sqrt{n}}$ par $\sqrt{\dfrac{f_e(1-f_e)}{n-1}}$

C. Test de comparaison de 2 populations

1. Test de comparaison de 2 moyennes

	Population P1		Population P2	
Caractères Etudiés	C		C	
Moyenne Ecart-type	m_1 σ_1	inconnus	m_2 σ_2	inconnus
	Echantillon e_1		Echantillon e_2	
Taille Moyenne Ecart-type	$n_1 > 30$ $\overline{x}_{e_1}$ σ_{e_1}	connus	$n_2 > 30$ $\overline{x}_{e_2}$ σ_{e_2}	connus

Règle du test de comparaison de 2 moyennes

L'hypothèse nulle est : H_0 : "$m_1 = m_2$" et l'hypothèse alternative est H_1 : "$m_1 <> m_2$"

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $N(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha/2$
- Si $\dfrac{\overline{x}_{e_1} - \overline{x}_{e_2}}{\sqrt{\dfrac{\sigma_{e_1}^2}{n_1 - 1} + \dfrac{\sigma_{e_2}^2}{n_2 - 1}}} \in \left[-t_\alpha ; t_\alpha\right]$ on accepte H_0 sinon on rejette H_0 et donc on accepte H_1
- Si H_0 est acceptée, on dit que la différence $m_1 - m_2$ n'est pas significative au risque α

2. <u>Règle de comparaison de deux pourcentages</u>

	Population P_1	Population P_2
Caractère étudié Pourcentage	C p_1 inconnu	C p_2 inconnu
	Echantillon e_1	Echantillon e_2
Taille Pourcentage	$n_1 > 30$ f_1 connu	$n_2 > 30$ f_2 connu

L'hypothèse nulle est H_0 : "$p_1 = p_2$" et l'hypothèse alternative H_1: "p_1 p_2"

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $N(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha/2$
- Soit $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ alors si $\dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \in \left[-t_\alpha ; t_\alpha\right]$ on accepte H_0 sinon on rejette H_0 et on accepte donc H_1.
- Si H_0 est acceptée, on dit que $p_1 - p_2$ n'est pas significative au risque α.