Chapitre 6

TESTS SUR LA MOYENNE ET LA VARIANCE.

1. TEST SUR LA MOYENNE.

X suit la loi normale N(μ, σ).

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

M : estimateur empirique de la moyenne S^2 : estimateur empirique de la variance

Test de Student :

$$\boxed{T = \frac{M - \mu_0}{S/\sqrt{(n-1)}}}$$

Loi de T sous H_0 : loi de Student de degré de liberté n-1

Région critique :

$$\boxed{]-\infty, \mu_0 - t_\alpha\, s/(n-1)^{1/2}[\cup]\, \mu_0 + t_\alpha\, s/(n-1)^{1/2}, +\infty [}$$

t_α étant défini par $P(|T| > t_\alpha) = \alpha$

Exemple : *l'objectif fixé par les responsables nationaux de l'enseigne Euromarket est un montant moyen des achats égal à 420F. Le directeur commercial s'inquiète du montant moyen observé (316.95F) dans son hypermarché et veut donc vérifier si cette valeur montre effectivement une différence. La variance estimée est égale à s^2 = 42902.472.*

Pour un risque de première espèce de 5% et un degré de liberté égal à 49, on a :

$$t_a = 2.02.$$

La région critique est donc :

$$\boxed{]-\infty, 360.23 [\cup] 479.77, +\infty [}$$

La valeur 316.95 appartient à la région critique. On rejette donc l'hypothèse nulle.

Équivalence : on calcule l'intervalle de confiance de la moyenne (cf. chapitre 5) :

$$IC = [\ 257.173,\ 376.717\ ]$$

La valeur 420 n'appartient pas à l'intervalle de confiance. Elle n'est pas acceptable.

2. Comparaison de moyennes.

$$X_1 = \mathcal{N}(\mu_1, \sigma_1) \quad X_2 : \mathcal{N}(\mu_2, \sigma_2)$$
$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

M_1 et S_1^2 estimateurs empiriques de la moyenne et la variance de X_1
M_2 et S_2^2 estimateurs empiriques de la moyenne et la variance de X_2
n_1 et n_2 : nombres d'observations de X_1 et X_2.

Statistique U :

$$\boxed{U = \frac{M_1 - M_2}{[S_1^2/(n_1-1) + S_2^2/(n_2-1)]^{1/2}}}$$

Loi de U sous H_0 : la loi normale centrée réduite (n_1 et n_2 suffisants, $\sigma_1 = \sigma_2$) .

Région critique :

$$\boxed{RC =]-\infty, -u_\alpha] \cup [u_\alpha, +\infty[}$$

u_α étant défini par $P(|U| > u_\alpha) = \alpha$

Exemple *: la moyenne des achats de l'autre hypermarché (410F) a été calculée sur 100 clients. La variance des achats calculée sur ces 100 clients est égale à 35401.*
On en déduit :

$$T = \frac{316.95 - 410}{[\,42902.47 / 49 + 35401 / 99\,]^{1/2}} = -2.65$$

Pour un risque de première espèce $\alpha = 0.05$, on a $u_\alpha = 1.96$. La valeur observée t appartient à la région critique et on rejette donc l'hypothèse nulle : la différence entre les deux moyennes n'est vraisemblablement pas due uniquement au hasard.

3. Test sur la variance.

X suit la loi normale $N(\mu, \sigma)$.

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 \neq \sigma_0^2$$

M : estimateur empirique de la moyenne
S^2 : estimateur empirique de la variance

Statistique :

$$X^2 = n\, S^2/\sigma_0^2$$

Loi de X^2 sous H_0 : loi du χ^2 de degré de liberté $\nu = n-1$.

Région Critique :

$$\boxed{RC = [0, \chi_\alpha^2] \cup [\chi_{1-\alpha}^2, +\infty[}$$

χ_α^2 et $\chi_{1-\alpha}^2$ étant définies par la relation :

$$P(X^2 < \chi_\alpha^2) = \alpha/2 \qquad P(X^2 > \chi_{1-\alpha}^2) = \alpha/2$$

Exemple *: nous supposons que la loi de probabilité de la v.a. âge est la loi normale (en éliminant les trois clients retraités) et testons l'hypothèse* $H_0 : \sigma^2 = 50$.

La valeur observée sur les 47 clients est $s^2 = 47.86$. *On en déduit :*

$$X^2 = 47 \times 47.86 / 50 = 44.99$$

Nous choisissons comme risque de première espèce $\alpha = 0.05$. *La table donne directement pour le degré de liberté* $\nu = 46$ *:*

$$\chi_\alpha^2 = 29.160 \quad \text{tel que } P(X^2 < \chi_\alpha^2) = 0.025$$
$$\chi_{1-\alpha}^2 = 66.617 \quad \text{tel que } P(X^2 > \chi_{1-\alpha}^2) = 0.025$$

La région critique est : $RC = [0, 29.160] \cup [66.617, +\infty[$. *La valeur observée n'appartient pas à la région critique et on accepte l'hypothèse nulle.*

En déterminant l'ensemble des valeurs σ^2 *telles que l'on accepte l'hypothèse nulle, on retrouvera l'intervalle de confiance déterminé dans le chapitre 5. On pourra examiner aussi la figure 11 du chapitre 5.*

4. Comparaison de variances.

$$X_1 = N(\mu_1, \sigma_1) \qquad X_2 : N(\mu_2, \sigma_2)$$
$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2 \qquad H_1 : \sigma_1^2 \neq \sigma_2$$

M_1 et S_1^2 estimateurs empiriques de la moyenne et la variance de X_1
M_2 et S_2^2 estimateurs empiriques de la moyenne et la variance de X_2
n_1 et n_2 : nombres d'observations de X_1 et X_2.

Statistique :

$$\boxed{F = \frac{n_1 S_1^2 / \sigma_1^2}{n_2 S_2^2 / \sigma_2^2} \times \frac{n_2 - 1}{n_1 - 1}}$$

Loi de F sous H_0 : loi de Fisher de degrés de liberté n_1-1 et n_2-1.

Région Critique :

$$\boxed{RC =]0, f_\alpha[\cup]f_{1-\alpha}, +\infty[}$$

f_α et $f_{1-\alpha}$ étant définies par : $P(F < f_\alpha) = \alpha/2$, $P(F > f_{1-\alpha}) = \alpha/2$

Exemple *: Pour contrôler l'égalité des moyennes des achats des deux hypermarchés, nous avons supposé que les variances étaient égales. Nous le vérifions ci-dessous, en supposant que les lois sont normales . Les variances des achats sont égale à :*

$$s_1^2 = 42902.47 \ (n_1 = 50) \qquad s_2^2 = 35401 \ (n_2 = 100)$$

On en déduit :

$$f = \frac{50 \times 42902.47 / 49}{100 \times 35401 / 99} = 1.2243$$

Nous choisissons comme risque de première espèce $\alpha = 0.02$. Les degrés de liberté sont $\nu_1 = 49$ et $\nu_2 = 99$.La table donne directement $f_{1-\alpha} = 1.73$.

Pour calculer f_α, il faut considérer la v.a. $1/F$, qui suit la loi de Fisher de degrés de liberté $\nu_1 = 99$ et $\nu_2 = 49$. On a :

$$P(1/F > 1/f_\alpha) = 0.01 \Leftrightarrow 1/f_\alpha = 1.82 \Leftrightarrow f_\alpha = 0.549$$

La région critique est donc : $RC =]0, 0.549 [\ ] 1.73, + \infty[$.

La valeur observée n'appartient pas à cette région critique et on accepte l'hypothèse d'égalité des variances.

5. Introduction au risque de seconde espèce et à la fonction puissance.

Définition :

risque de seconde espèce : probabilité d'accepter l'hypothèse nulle quand elle est fausse.

puissance : probabilité de rejeter l'hypothèse nulle quand elle est fausse.

$$\pi = 1 - \beta$$

Puissance du test sur la variance :

Hypothèses:

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 = \sigma_1^2$$

Acceptation de l'hypothèse nulle : $n\, S^2/\sigma_0^2$ n'appartient pas à la région critique.

Risque β de seconde espèce (sous H_1) :

$$\boxed{\beta = P\left(\chi_\alpha^2 < \frac{n\, S^2}{\sigma_0^2} < \chi_{1-\alpha}^2\right)}$$

Calcul de β :

$$\beta = P\left(\chi_\alpha^2 < \frac{n\, S^2}{\sigma_0^2} < \chi_{1-\alpha}^2\right)$$

$$= P\left(\chi_\alpha^2 < \frac{nS^2\sigma_1^2}{\sigma_0^2\sigma_1^2} < \chi_{1-\alpha}^2\right)$$

$$\chi_\alpha^2\sigma_0^2 \qquad nS^2 \qquad \chi_{1-\alpha}^2\sigma_0^2$$

$$= P\left(\frac{\quad}{\sigma_1^2} < \frac{\quad}{\sigma_1^2} < \frac{\quad}{\sigma_1^2}\right)$$

Loi de n S^2 / σ_1^2 sous H_1 : loi du χ^2 de degré de liberté $\nu = n - 1$.

Exemple *: nous donnons ci-dessous la puissance du test sur la variance de l'âge. La valeur testée est fixée à* $\sigma^2 = 50$*, le risque de première espèce à 0.05, et le nombre d'observations est égal à 47.*

La lecture de ce tableau donne le renseignement suivant : la probabilité de rejeter l'hypothèse $\sigma^2 = 50$ *lorsque la vraie valeur est 33.333 est égale à 0.432 pour un risque de première espèce égal à 0.05.*

Rang	*variance vraie*	*puissance*	*Rang*	*variance vraie*	*puissance*
1	*20.0000*	*0.993*	*11*	*64.4444*	*0.263*
2	*24.4444*	*0.915*	*12*	*68.8889*	*0.379*
3	*28.8889*	*0.699*	*13*	*73.3333*	*0.497*
4	*33.3333*	*0.432*	*14*	*77.7778*	*0.606*
5	*37.7778*	*0.227*	*15*	*82.2222*	*0.701*
6	*42.2222*	*0.109*	*16*	*86.6667*	*0.778*
7	*46.6667*	*0.057*	*17*	*91.1111*	*0.839*
8	*51.1111*	*0.053*	*18*	*95.5556*	*0.885*
9	*55.5556*	*0.090*	*19*	*100.0000*	*0.919*
10	*60.0000*	*0.162*			

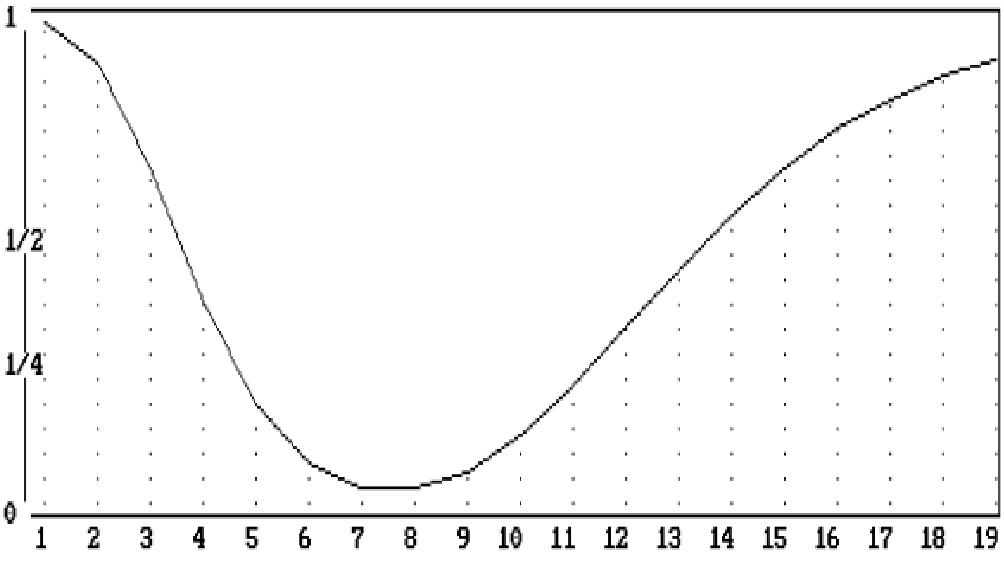

Figure 4.6 : fonction puissance $(\sigma^2 = 50, \nu = 46, \alpha = 0.05)$

C