Chapitre 04 – Notions sur les Tests statistiques

I/ Introduction : On émet une hypothèse $H_0$ concernant une population et ceci sur la base de résultats expérimentaux ; on est alors amené à accepter ou à rejeter l'hypothèse $H_0$ avec un seuil de signification $\alpha$ "assez petit".

Ex On teste un nouveau médicament sur 100 personnes.
on dit qu'il est efficace à un seuil de signification $\alpha = 0,1$
s'il l'est dans $90\% = (1-\alpha)$ des cas.

$\alpha$ est appelé seuil de signification
$(1-\alpha)$ — " — — confiance

l'hypothèse émise $H_0$ est appelée hypothèse nulle
on note $H_1$ l'hypothèse alternative ie
$H_0$ vraie ie $H_1$ fausse
$H_1$ fausse ie $H_1$ vraie.

il y'a 4 cas possibles

1/ $H_0$ est vraie et on a accepté (choisi, validé) $H_0$

2/ $H_0$ est fausse et on a rejeté $H_0$

3/ $H_0$ est vraie et on a rejeté $H_0$

4/ $H_1$ est vraie et on a choisi (accepté, validé) $H_0$

- Seuls les cas 1 et 2 sont justes
- pour le cas 3/ on dit que l'on a commis une erreur de 1ère espèce $\alpha$ (          )
- pour le cas 4/ on dit qu'on a commis une erreur de 2ème espèce $\beta$

$\alpha$ est choisi assez petit seuil de signification

$(1-\alpha)$ est appelé seuil de confiance

$\beta$ est généralement inconnue

$(1-\beta)$ est appelé la puissance du test

on a les tableaux récapitulatifs

| | $H_0$ vraie | $H_0$ fausse |
|---|---|---|
| $H_0$ validée | $1-\alpha$ | $\beta$ |
| $H_0$ invalidée | $\alpha$ | $1-\beta$ |
| | 1 | 1 |

| | $H_0$ vraie | $H_0$ fausse |
|---|---|---|
| $H_0$ validée | Confiance | Erreur de type II |
| $H_0$ invalidée | Erreur de type I | puissance |

## II / Catégories de tests

- Un test est dit d'ajustement si

$$\begin{cases} H_0 = F(x) = F_0(x) \\ H_1 = F(x) \neq F_0(x) \end{cases}$$

où $F(x)$ est la fonction de répartition de la variable échantillonnée (expérimentale) et $F_0(x)$ est la F.d.r d'une V.a connue.

- Un test est dit d'hypothèse simple si on veut choisir entre deux valeurs d'un paramètre $\theta$ ($\theta_0$ ou $\theta_1$).

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

- Un test est dit d'hypothèse multiples si :

a/ $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$ test unilatéral à droite (test de supériorité).

b/ $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$ test unilatéral à gauche (test d'infériorité).

c/ $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$ test bilatéral.

- Un test est dit d'homogénéité si $\begin{cases} H_0 : \theta_0 = \theta_1 \\ H_1 : \theta_0 \neq \theta_1 \end{cases}$

où $\theta_0$ et $\theta_1$ sont deux valeurs d'un même paramètre dans deux populations différentes

• Test d'ajustement $\chi^2$ (khi deux, khi au carré).

Soit un échantillon aléatoire de taille $N$ extrait d'une population et divisé en $k$ classes d'effectifs partiels : $n_1, n_2, \dots, n_k$ et de fréquences (probabilités) $p_1, p_2, \dots, p_k$ théoriques

il s'agit de tester si la V.a. échantillonnée suit une loi

Connue (exp, uniforme, gaussienne ? ...).

$$\begin{cases} H_0 : F(x) = F_0(x) \\ H_1 : F(x) \neq F_0(x). \end{cases}$$

pour cela on utilise le test du $\chi^2$ si $Np_i > 5$ (sinon on regroupe)

. On calcule la quantité : $\displaystyle\sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} = \chi^2$.

• on choisit un seuil de signification $\alpha$ (assez petit).

. On cherche le degré de liberté (généralement d.d.l $= k-1$ car $n_1 + n_2 + \dots + n_k = N$)

• ~~si~~ prend $\chi^2_{k-1}(\alpha)$ de la table du $\chi^2$.

• si $\chi^2 < \chi^2_{k-1}(\alpha)$ nous sommes dans le cas $P(\chi^2 < \chi^2_{k-1}(\alpha)) = 1-\alpha$

on accepte l'hypothèse $H_0$. donc $F_0$ représente valablement $X$

sinon $\chi^2 \geqslant \chi^2_{k-1}(\alpha)$ on rejette $H_0$ et dans ce cas $F_0$ ne représente pas valablement $X$ expérimentale.

Exemple :

On a croisé deux races de plantes, différant par deux caractères A et B. La première génération est homogène ; la seconde génération fait apparaître 4 types de plantes, dont le phénotype est noté AB, Ab, aB, ab.

Si les caractères se transmettent selon les lois de Mendel, les proportions (fréquences) des 4 phénotypes sont : 9/16, 3/16, 3/16, 1/16.

Dans un échantillon de 160 plantes nous avons :

| AB | Ab | aB | ab |
|---|---|---|---|
| 100 | 18 | 24 | 18 |

Cette répartition est-elle conforme aux lois de Mendel au seuil de signification $\alpha = 5\%$ ? utiliser le $\chi^2$.

Sol : $H_0$ : " la répartition est conforme aux lois de Mendel "

| Phénotypes : | AB | Ab | aB | ab | |
|---|---|---|---|---|---|
| proportion thque $p_i$ | 9/16 | 3/16 | 3/16 | 1/16 | 1 |
| Effectif calculé $N.p_i$ | 90 | 30 | 30 | 10 | 160 |
| Effectif observé $n_i$ | 100 | 18 | 24 | 18 | 160 |

$\underline{Np_i > 5}$

$$\chi^2 = \sum \frac{(n_i - Np_i)^2}{Np_i} = \frac{10^2}{90} + \frac{(-12)^2}{30} + \frac{(-6)^2}{30} + \frac{8^2}{10} = 13,51$$

$ddl = k-1$ , $\alpha = 0,05$
$= 3$

table du $\chi^2$

| $\nu = ddl$ \ $\alpha$ | 0,5 | 0,1 | 0,05 |
|---|---|---|---|
| 2 | 1,386 | 4,61 | 5,99 |
| 3 | 2,366 | 6,25 | 7,81 |
| 4 | 3,357 | 7,78 | 9,49 |
| 5 | 4,251 | 9,24 | 11,07 |

$\Rightarrow \chi^2_3(0,05) = 7,81$

d'où $\chi^2 > \chi^2_3(0,05)$

l'hypothèse d'ajustement doit être rejetée.

Notre échantillon n'est pas conforme aux lois de Mendel au seuil de $\alpha = 0,05$.

Exemple 2:

Une machine fabrique des pièces en grand nombre on effectue un contrôle en prelevant 100 échantillons aléatoires de 30 pièces chacun on obtient le tableau.

| Nbre de pièces défectueuses / échantillon : $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|---|
| Nombre de lots Correspondant $n_i$ | 4 | 15 | 24 | 23 | 17 | 9 | 6 | 2 | $\Rightarrow N = 100$ |

1/ Calculer $\overline{X}$ et $Var(X)$ ?

2/ Peut-on ajuster cette répartition par une loi binomiale ? avec un seuil de signification $\alpha = 0,05$

Solution:

1/ $\overline{X} = \sum f_i x_i = \sum \frac{n_i}{100} \cdot x_i = \frac{4}{100} 0 + \frac{15}{100} \cdot 1 + \cdots + \frac{2}{100} \cdot 7 = 2,95.$

$Var(X) = \sum f_i x_i^2 - \overline{X}^2 = \sum \frac{n_i}{100} \cdot x_i^2 - \overline{X}^2 = \frac{4}{100} \cdot 0^2 + \frac{15}{100} \cdot 1^2 + \cdots + \frac{2}{100} \cdot 7^2 - (2,95)^2$

$= 2,59.$

$\sigma_X = \sqrt{2,59} = 1,6$

2/ Trouver les paramètres de la loi binômiale: $B(n,p)$.

$n = 30.$ $p$ ? on prend $p = \frac{\overline{X}}{30} = \frac{2,95}{30} \simeq \frac{3}{30} = 0,1$. $B(30, 0,1)$.

pour $k = 0, 30$ $\quad P(X = k) = C_{30}^{k} p^k (1-p)^{30-k} = P_k$

On obtient alors le tableau :

| $x_i$ | $n_i$ | $p_i$ | $100p_i = Np_i$ | $n_i - Np_i$ |
|---|---|---|---|---|
| 0 | 4 | 0,0424 | 4,24 | −0,24 |
| 1 | 15 | 0,1413 | 14,13 | 0,87 |
| 2 | 24 | 0,2277 | 22,77 | 1,23 |
| 3 | 23 | 0,2361 | 23,61 | −0,61 |
| 4 | 17 | 0,1771 | 17,71 | −0,71 |
| 5 | 9 | 0,1023 | 10,23 | −1,23 |
| 6 | 6 | 0,0474 | 4,74 | 1,26 |
| 7 | 2 | 0,018 | 1,8 | 0,2. |

Pour pouvoir appliquer le test du $\chi^2$ il est nécessaire d'avoir $Np_i > 5$

On obtient le nouveau tableau :

| $x_i$ | $n_i$ | $Np_i$ | $(n_i - Np_i)^2$ |
|---|---|---|---|
| 0 ; 1 | 19 | 18,37 | 0,3969 |
| 2 | 24 | 22,77 | 1,5129 |
| 3 | 23 | 23,61 | 0,3721 |
| 4 | 17 | 17,71 | 0,5041 |
| 5 | 9 | 10,23 | 1,5129 |
| 6 ; 7 | 8 | 6,54 | 2,1316 |

$$\chi^2 = \sum \frac{(n_i - Np_i)^2}{Np_i} = 0,606$$

il y'a $k-2$ degré de liberté $= 6-2 = 4$

$\chi^2_4(0,05) = 9,49$ $\Rightarrow \chi^2 < \chi^2_4(0,05)$ on accepte $H_0$

Exemple 3 :

Les notes de mathématiques d'un groupe d'étudiants lors d'un examen sont réparties selon le tableau :

| Notes : $x$ | $2 \leq x < 4$ | $4 \leq x < 6$ | $6 \leq x < 8$ | $8 \leq x < 10$ | $10 \leq x \leq 12$ |
|---|---|---|---|---|---|
| Nombres d'étudiants : $n_i$ | 2 | 5 | 8 | 14 | 16 |

| Notes | $12 \leq x < 14$ | $14 \leq x < 16$ | $16 \leq x < 18$ |
|---|---|---|---|
| $n_i$ | 12 | 9 | 4 |

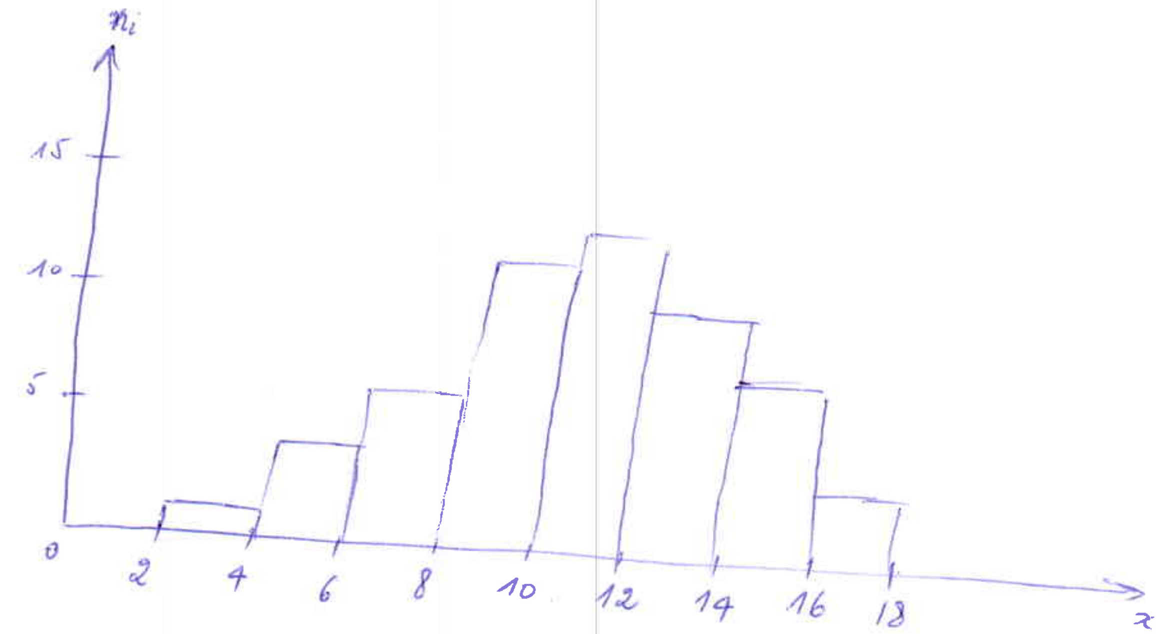
$(N = 70)$

1- Par quelle loi de probabilité on peut procéder à un ajustement de la distribution empirique proposée.

2- Vérifier l'ajustement par un test du $\chi^2$ à un seuil de signification $\alpha = 0,05$.

Solution :

1) Si l'on trace l'histogramme des effectifs partiels on a :




la forme de l'histogramme des fréquences permet d'affirmer qu'on peut proposer un ajustement par une loi normale.

Les paramètres $m, \sigma$ de la loi loi normale sont estimés par :

$\bar{X}$ et $\sigma$

$$\bar{X} = \sum f_i x_i = \sum \frac{n_i}{N} \cdot x_i = 10,69.$$

$$\sigma^2 = \sum f_i x_i^2 - \bar{X}^2 = \sum \frac{n_i}{N} x_i^2 - \bar{X}^2 = 11,9.$$

$$\sigma = 3,45.$$

donc $X$ peut être ajusté par une loi normale de d.d.p

$$f(x) = \frac{1}{3,45\sqrt{2\pi}} \, e^{-\frac{(x-10,69)^2}{2 \cdot 11,9}}$$

$H_0$ : $X$ suit une loi normale $N(10,69,\ 3,45)$

2-

Les probabilités théoriques $P(a \leq x \leq b) = F(b) - F(a) = p_i$.

sont calculés à partir de tables.

| Notes | Effectif $n_i$ | $p_i$ | $Np_i = 70p_i$ |
|---|---|---|---|
| $2 \leq x < 4$ | 2 | 0,0203 | 1,421 |
| $4 \leq x < 6$ | 5 | 0,0607 | 4,249 |
| $6 \leq x < 8$ | 8 | 0,1337 | 9,359 |
| $8 \leq x < 10$ | 14 | 0,2001 | 14,007 |
| $10 \leq x < 12$ | 16 | 0,2273 | 15,911 |
| $12 \leq x < 14$ | 12 | 0,1835 | 12,845 |
| $14 \leq x < 16$ | 9 | 0,1067 | 7,49 |
| $16 \leq x < 18$ | 4 | 0,0448 | 3,136 |

la condition pour appliquer le $\chi^2$ est $Np_i \geq 5$

| Notes | $n_i$ | $p_i$ | $Np_i$ |
|---|---|---|---|
| $2 \leq x < 6$ | 7 | 0,081 | 5,67 |
| $6 \leq x < 8$ | 8 | 0,1337 | 9,359 |
| $8 \leq x < 10$ | 14 | 0,2001 | 14,007 |
| $10 \leq x < 12$ | 16 | 0,2273 | 15,911 |
| $12 \leq x < 14$ | 22 | 0,1835 | 12,845 |
| $14 \leq x < 18$ | 13 | 0,1515 | 10,605 |

$$\chi^2 = \sum \frac{(n_i - Np_i)^2}{Np_i} = 1,1062$$

d.d.l : $\cancel{\nu} = 6 - 1 - 2 = 3$

$1 \rightarrow \sum n_i = N$

$2 \rightarrow m = \overline{x} = \frac{\sum n_i x_i}{N}$

$3 \rightarrow \sigma = \sum \frac{n_i}{N} x_i^2 - \overline{x}^2$

$\alpha = 0{,}05 \rightarrow ddl = 3$ $\qquad \chi_3(0{,}05) \underset{table}{=} 7{,}81$

$$\chi^2 = 1{,}1062 < \chi_3^2(0{,}05) = 7{,}81$$

on accepte $H_0$.

Exemple 4 :

Les résultats des épreuves d'un examen à l'échelle nationale sont : 60% de reçus ; 25% admissibles (admis à l'oral) et 15% éliminés.

Un établissement présente 160 élèves et obtient 75 reçus, 53 admissibles et 32 éliminés,

y'a t-il conformité entre ces résultats et ceux valables à l'échelle nationale ?

Solution :

$H_0$ : " Les résultats obtenus sont conformes à ceux valables à l'échelle nationale ".

| | $n_i$ | $P_i$ | $Np_i$ |
|---|---|---|---|
| Reçus | 75 | 0,6 | 96 |
| admissibles | 53 | 0,25 | 40 |
| Eliminés | 32 | 0,15 | 24 |
| | = 160 | 1 | |

ici ddl = 3-1 = 2

$$\chi^2 = \sum \frac{(n_i - Np_i)^2}{Np_i} = 11,484.$$

la table du $\chi^2$ donne $\chi^2_2(0,01) = 9,21$ , $\chi^2_2(0,001) = 13,82$.

au seuil de signification $\alpha = 0,01$ $H_0$ rejetée

au seuil de signification $\alpha = 0,001$ $H_0$ acceptée.