Faculté de Médecine d'Oran
Laboratoire de Biostatistique

Ajustements à différentes lois de probabilités connues

Année universitaire 2015-2016

Introduction

Le test du khideux ou chi-carré ou encore de khideux de Pearson
Il sert a comparer des distributions, appliqué sur des variables qualitative nominale, ordinale, binaire, quantitative discrète ou continue discrétisée

Introduction

On retrouve essentiellement :
-khideux de conformité ou d'ajustement
-khideux d'homogéneité
-khideux de Mac Nemar pour série appariée
-khideux d'indépendance

Khideux de conformité ou d'ajustement

il sert à comparer une distribution observée sur un échantillon a une distribution connue dans une population ou une distribution théorique(Poison, Binomiale, Normale par exemple) quelque soit le test du χ^2 réalisé.

l'objectif du test est de déterminer si les écarts entre la distribution **des effectifs observés et la distribution des effectifs théoriques est significative ou imputable uniquement aux fluctuations d'échantillonnage.**

Principe du test

Deux cas peuvent se présenter :

- soit la loi de probabilité **est spécifiée** a priori car elle résulte par exemple d'un modèle déterministe tel que la distribution mendélienne des caractères, l'évolution de la taille d'une population, etc.

- soit la loi de probabilité théorique **n'est pas connue** a priori et elle est déduite des caractéristiques statistiques mesurées sur l'échantillon (distribution des fréquences, moyenne et variance).

Application

- L'établissement des distributions théoriques de probabilité se réfèrent aux **lois de probabilité.**
- A chaque modalité ou valeur de la variable aléatoire X,
  les probabilités associées à la loi de probabilité sont calculées ainsi que **les effectifs théoriques attendues sous cette loi :**

	Modalité du caractère A A_1 A_2 ... A_i .... A_k	
Effectif observé n_i	n_1 n_2 .....n_i .... n_k	$n = \sum_{i=1}^{k} n_i$
p_i	p_1 p_2 ......p_i .... p_k	$\sum_{i=1}^{k} p_i = 1$
Effectif théorique $t_i = n * p_i$	t_1 t_2 .....t_i .... t_k	$n = \sum_{i=1}^{k} t_i$

Principe du test

- si n est l'effectif total étudié, **l'effectif théorique** attendu, t^i pour la modalité i de la variable aléatoire X est :

$$\boxed{t_i = n * p_i}$$

- quelque soit l'hypothèse nulle testée, la stratégie est la même pour tous les tests du χ^2.

Ajustements à différentes lois de probabilité connues

Ajustement à la loi binomiale

Application

- Est-ce que la distribution du nombre de filles observées dans 40 fratries de 4 enfants suit une loi binomiale de paramètre B(n, p) ?

fille(s) Xi	0	1	2	3	4
famille(s) Ni	8	8	10	4	10

- **la loi de probabilité théorique n'est pas connue; elle est déduite des caractéristiques statistiques mesurées sur l'échantillon cad distribution des fréquences.**

1 - si notre distribution suit une loi binomiale alors B(n, p) ?
n: nbre d'épreuves p: probabilité du succès
x: valeur de chaque modalité de la variable discrète.

[redacted]

n: 4

p: ?

p: E(x)/n E(x):?

$\boxed{E(X) = np}$

2 - L' espérance mathématique peut être calculée par la moyenne pondérée de ces valeurs, à savoir

$$\overline{X} = \frac{\sum ni\,Xi}{N} = (0+8+12+12+40)\ /40 = 2$$

N: taille de l'échantillon

donc **p = E(x) / n** = 2 / 4 soit **p** = 0,5.

3 - **n: 4, p = 0,5** alors la loi binomiale est : **B(4 ; 0,5).**

Calculons les probabilités correspondantes au valeurs des xi correspondantes.

[redacted]

Xi	**0**	**1**	**2**	**3**	**4**
P(xi)	**0.0625**	**0.2500**	**0.3750**	**0.2500**	**0.0625**

4 - Calculons les effectifs correspondants

$\boxed{t_i = n * p_i}$ taille de l'échantillon=40

OBS_i (Oi)	**8**	**8**	**10**	**4**	**10**
TH_i (ti)	**2.5**	**10**	**15**	**10**	**2.5**

5-Calculons les effectifs corrigés correspondants

Effectifs observes (Oi)	Effectifs théoriques (Ti)
16	12.5
10	15
14	12.5

PROCÉDURE

1- Détermination du risque α

2- Construction de l'hypothèse nulle Ho.

3- Validité :

Tous les ti ≥ 5

4- Calcul du degrés de liberté **d.d.l = ν = k –2** (car les effectifs théoriques sont calculés a partir de l'échantillon)

Détermination de la valeur limite (critique) de la table $\chi^2_{cr} = f(\alpha, \nu)$

5- Variable testée :

$$\chi^2 = \sum \frac{(ti - oi)^2}{t_i}$$

6- Conclusion :

Si $\chi^2 \geq \chi^2_{cr}$ La différence est significative au risque α, alors Ho est rejetée

Si $\chi^2 < \chi^2_{cr}$ La différence n'est pas significative au risque α, alors Ho est retenue.

1- risque α=5%

2-Ho: la distribution observée est conforme à la distribution théorique (binomiale)

H1: la distribution observée n'est pas conforme à la distribution théorique (binomiale)

3- Validité :

tous les ti ne sont pas ≥ 5 (on fait la somme de l'effectif < 5 de la modalité de la variable avec celui la modalité suivante jusqu'à ce que l'on obtienne un effectif ≥ 5)

4- Calcul du degré de liberté

d.d.l = ν = k –2 (car les effectifs théoriques sont calculés a partir de l'échantillon)

ddl = 3-2=1

5- calcul de la variable testée

Effectifs observés (Oi)	Effectifs théoriques (Ti)	$(Ti - Oi)^2 / Ti$
16	12.5	0.98
10	15	1.66
14	12.5	0.18
$\sum Oi = 40$	$\sum Ti = 40$	$\sum (Oi - Ti)^2 / Ti\ \chi^2 =$ 2.82

χ^2 (cal)2.82 < **χ^2 (tab)**3.815 (la valeur maximale lue de la table, au seuil de 5 %, pour 1 ddl ou "degrés de liberté")

6- La différence n'est pas significative donc on peut conclure qu'au seuil de 5 %, l'hypothèse nulle est retenue autrement dit **les données de cette distribution suivent la loi binomiale.**

Ajustement à la loi de Poisson

- On a répertorié dans une usine le nombre d'accidents mineurs subis par le personnel dans une journée de travail sur une période de 200 jours. Ces accidents sont survenus indépendamment les uns des autres. Les résultats sont consignés dans le tableau suivant :

Nombre d'accidents	0	1	2	3	4	5
Nombre de jours	86	82	22	7	2	1

1°/ Quel est le nombre moyen d'accidents par jour ?
2°/ Cette distribution suit-elle la loi de Poisson ?

1-si notre distribution suit une loi de poisson alors P(λ) ?

$$p(x) = \frac{\lambda^{x}}{x!} \cdot e^{-\lambda}$$

e = 2,71 (base du logarithme népérien) et
λ : espérance mathématique E (x), appelée encore en pratique la moyenne.
x: valeur de chaque modalité de la variable discrète.

$$E(x) = V(X) = n\,p = \lambda$$

E(x): λ ?
2-L' espérance mathématique peut être calculée par la moyenne pondérée de ces valeurs, à savoir

$$\overline{X} = \frac{\sum ni\,Xi}{N} = (0 \times 86 + 1 \times 82 + 2 \times 22 + 3 \times 7 + 4 \times 2 + 5 \times 1)/200 = 4/5 = 0{,}8$$

$E(x) = n\,p = \lambda = 0{,}8$ $\quad P(X = xi) = e^{-0{,}8}(0{,}8)^{x}/x!$

3-Calculons les probabilités correspondantes au valeurs des xi correspondantes et les effectifs correspondants.

Nombre d'accidents (Xi)	Nombre de jours (Ni)	p_i théoriques selon Poisson	Ti=npi
0	86	0.449	90
1	82	0.359	72
2	22	0.143	29
3	7	0.038	8
4	2	0.007	1
5	1	0.001	0.2
Total	∑ni=n=200		∑ti=200.

4-calculons les effectifs théoriques corrigés .

Effectifs observes (Oi)	Effectifs théorique (Ti)
86	*90*
82	*72*
22	*29*
10	9.2
∑*Oi* = 200	∑*Ti* = 200

PROCÉDURE

1- Détermination du risque α

2- Construction de l'hypothèse nulle Ho.

3- Validité :

Tous les ti ≥ 5

4- Calcul du degrés de liberté **d.d.l** = ν = **k −2** (car les effectifs théoriques sont calculés a partir de l'échantillon)

Détermination de la valeur limite (critique) de la table $\chi^2_{cr} = f(\alpha, \nu)$

5- Variable testée :

$$\chi^2 = \sum \frac{(t_i - o_i)^2}{t_i}$$

6- Conclusion :

Si $\chi^2 \geq \chi^2_{cr}$ La différence est significative au risque α, alors Ho est rejetée

Si $\chi^2 < \chi^2_{cr}$ La différence n'est pas significative au risque α, alors Ho est retenue.

1- risque α=5%

2-H0: la distribution observée est conforme à la distribution théorique (loi de poisson)
H1: la distribution observée n'est pas conforme à la distribution théorique (loi de poisson)

3- Validité :
tous les ti ne sont pas ≥ 5 (on fait la somme de l'effectif < 5 de la modalité de la variable avec celui la modalité suivante jusqu'à ce que l'on obtienne un effectif ≥ 5)

4- Calcul du degré de liberté
d.d.l = ν = k −2 (car les effectifs théoriques sont calculés a partir de l'échantillon)
d.d.L = $\nu = 4 - 2 = 2$

5-calcul de la variable testée

Effectifs observes (Oi)	Effectifs théorique (Ti)	$(Ti - Oi)^2 / Ti$
86	*90*	*0.178*
82	*72*	*1.390*
22	*29*	*1.690*
10	9.2	*0.070*
$\sum Oi = 200$	$\sum Ti = 200$	$\sum (Ti - Oi)^2 / Ti = 3.328$

6- Détermination du χ^2_t de la table.
$\chi^2_t = f(\alpha, \nu) = f(0,05 ; 2) = 5.991$ (lue dans une table, au seuil de 5%,)

7-Conclusion et prise de décision. :
$\chi^2_c < \chi^2_t \rightarrow$ (3.328 < 5.991).
Ho est retenue , la distribution observée suit une loi de Poisson avec un risque α égal à 5%

Ajustement à la loi normale

- Le tableau suivant donne la répartition d'une population par tranche d' âge

Classe xi	[0,10[	[10,20[	[20,30[	[30,40[	[40,50[	[50,60[	[60,70[	[70,80[
Effectifs ni	18	44	68	54	42	36	16	10

- 1- calculer la moyenne , la variance et l'écart type de la série.
- 2-au risque de 5 %,tester l'hypothèse de normalité

1-Si notre distribution suit une loi normale, la variable *X* suit une loi normale d'espérance μ et d'écart-type σ.
Alors la variable $\boxed{T = \frac{x - \mu}{\sigma}}$ est distribuée selon une loi normale centrée réduite.

Nous pouvons calculer les probabilités pour une loi normale, à l'aide de cette transformation.
μ et σ ?
2- la moyenne est égale à $\overline{X} = \frac{\sum ni\,Xi}{N}$

$$\overline{X} = \frac{10000}{288} = 34{,}72$$

$$S^2 = \frac{\sum ni\,Xi^2}{N} - \overline{X}^2 = 311{,}033 \quad S = 17{,}6$$

Classe xi	Xicc	ni(eff.obs)	ni*xi	ni*xi²
[0,10[	5	18	90	450
[10,20[	15	44	660	9900
[20,30[	25	68	1700	42500
[30,40[	35	54	1890	66150
[40,50[	45	42	1890	85050
[50,60[	55	36	1980	108900
[60,70[	65	16	1040	67600
[70,80[	75	10	750	56250
	Σ	=288	Σ =10000	Σ =436800

3-Calculons les probabilités correspondantes au valeurs des xi correspondantes et les effectifs correspondants.

Classe xi	ni= oi(eff.obs)	$t = \frac{xi-34,72}{17,6}$	Pi	ti=n*p
[0,10[	18	$-1,97 < t < -1,40$ G(t)(0,475; 0,419)	0,056	16,12
[10,20[	44	$-1,40 < t < -0,83$ G(t)(0,419; 0,296)	0,123	35,42
[20,30[	68	$-0,83 < t < -0,26$ G(t)(0,296; 0,102)	0,194	55,87
[30,40[	54	$-0,26 < t < 0,30$ G(t)(0,102; 0,117)	0,219	63,07
[40,50[	42	$0,30 < t < 0,86$ G(t)(0,117; 0,305)	0,188	54,14
[50,60[	36	$0,86 < t < 1,43$ G(t)(0,305; 0,423)	0,118	33,98
[60,70[	16	$1,43 < t < 2,00$ G(t)(0,423; 0,477)	0,054	15,55
[70,80[	10	$2,00 < t < 2,57$ G(t)(0,477; 0,494)	0,017	4,89
Σ	288			Σ =279,04

PROCÉDURE

1- Détermination du risque α

2- Construction de l'hypothèse nulle Ho.

3- Validité :
Tous les $ti \geq 5$

4- Calcul du degrés de liberté **d.d.l** = **ν = k –3** (car les effectifs théoriques sont calculés a partir de l'échantillon)
Détermination de la valeur limite (critique) de la table $\chi^2_{cr} = f(\alpha, \nu)$

5- Variable testée :

$$\chi^2 = \Sigma \frac{(ti - oi)^2}{t_i}$$

6- Conclusion :

Si $\chi^2 \geq \chi^2_{cr}$ La différence est significative au risque α, alors Ho est rejetée

Si $\chi^2 < \chi^2_{cr}$ La différence n'est pas significative au risque α, alors Ho est retenue.

1- risque α=5%

2-H0: la distribution observée est conforme à la distribution théorique (loi normale)

H1: la distribution observée n'est pas conforme à la distribution théorique (loi normale)

3- Validité :

tous les ti ne sont pas ≥ 5 (on fait la somme de l'effectif < 5 de la modalité de la variable avec celui la modalité suivante jusqu'à ce que l'on obtienne un effectif ≥ 5)

4- Calcul du degré de liberté

d.d.l = $\nu = k - 3$ (car les effectifs théoriques sont calculés a partir de l'échantillon)

d.d.L = $\nu = 7 - 3 = 4$

5-calcul de la variable testée

ni=oi(eff.obs)	ti=n*p	$(Ti-oi)^2/ti$
18	16,12	0,219
44	35,42	2,078
68	55,87	2,633
54	63,07	1,304
42	54,14	2,722
36	33,98	0,120
26	20,44	0,272
288		$\Sigma (Oi - Ti)^2 / Ti$ =**9,348**

6- Détermination du χ^2_t de la table.

$\chi^2_t = f(\propto, \nu) = f(0{,}05\ ;\ 4) = 9{,}488$ (lue dans une table, au seuil de 5%,)

7-Conclusion et prise de décision. :

$\chi^2_c < \chi^2_t \rightarrow$ (9,348 < 9,488).

Ho est retenue , la distribution observée suit une loi normale avec un risque α égal à 5%