Feuille n° 4

Exercice 1. *On cherche à savoir si la fréquence d'une maladie est liée au groupe sanguin. Sur 200 malades observés, on a dénombré 104 personnes du groupe O, 76 du groupe A, 18 du groupe B, 2 du groupe AB.*

On admettra que, dans la population générale, la répartition entre les groupes est : 47% pour le groupe O, 43% pour le groupe A, 7% pour le groupe B et 3 % pour le groupe AB.

Peut-on admettre, au niveau 0,10, que la répartition en groupe sanguin est conforme à celle de la population totale?

↳ Fréquences observées

$O_O : \frac{104}{200}$ $\quad O_B : \frac{18}{200}$

$O_A : \frac{76}{200}$ $\quad O_{AB} : \frac{2}{200}$

O : observed
E : expected

↳ Répartition population générale :

$E_O : 0{,}47$ $\quad E_B : 0{,}07$

$E_A : 0{,}43$ $\quad E_{AB} : 0{,}03$

↳ Si Maladie $\perp\!\!\!\perp$ Gpe Sanguin, alors $O_i \simeq E_i \quad \forall i \in \{O, A, B, AB\}$

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Groupe sanguin	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
O	104	94	$104 - 94 = 10$	$10^2 = 100$	$\frac{100}{94} = 1.064$
A	76	86	$76 - 86 = -10$	$(-10)^2 = 100$	$\frac{100}{86} = 1.163$
B	18	14	$18 - 14 = 4$	$4^2 = 16$	$\frac{16}{14} = 1.143$
AB	2	6	$2 - 6 = -4$	$(-4)^2 = 16$	$\frac{16}{6} = 2.667$

$$\chi^2 = 1{,}064 + 1{,}163 + 1{,}143 + 2{,}667 = 6{,}036$$

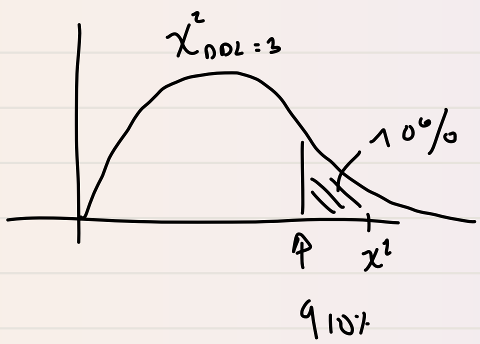
H_0 : « Les répartitions "Gpes sanguins" et "maladie" sont les mêmes dans la population »

H_1 : « Les répartitions "Gpes sanguins" et "maladie" sont $\neq$ dans la population »

$\hookrightarrow$ Quantile 10% pour une variable du χ^2 avec $k-1 = 4-1 = 3$ degrés de liberté :

$$q_{10\%} = 6,251 \quad \text{(lire la table)}$$

$q_{10\%} < \chi^2 \Rightarrow$ On rejette H_0 $\Leftrightarrow$ Gpe sanguins et maladie ne sont pas indépendants



Exercice 2. *Une enquête effectuée auprès du comptoir de 150 coopératives agricoles a permis d'étudier l'arrivée dans le temps des usagers de ces coopératives.*

Pendant l'unité de temps (d'une heure) on a noté :

usagers arrivés	0	1	2	3	4	5	6
Nombre de coopératives	37	46	39	19	5	3	1

≥ 5

⚠ Regrouper les valeurs de queue

(1) *Calculer la moyenne et la variance empirique.*
(2) *Peut-on admettre, au niveau 5%, que la population suit une loi de Poisson?*

$$\overline{X} = \frac{1}{N}\sum_{i=0}^{6} i \times n_i = \frac{37\times 0 + 46\times 1 + \dots + 1\times 6}{150} = 1{,}48 = \lambda$$

↳ Pour un processus de Poisson, $X \sim \mathcal{P}(\lambda)$, $E(X) = \lambda$

↳ H_0 : "Les arrivées suivent une loi de Poisson de paramètre $\lambda = 1{,}48$"

H_1 : "Les arrivées ne suivent pas de loi de Poisson"

Usagers arrivés (X_i)	Effectifs observés (O_i)	Effectifs attendus (E_i)
0	37	37.75
1	46	55.82
2	39	41.31
3	19	20.37
4	5	7.53
5+	4	3.22

Éno

$\mathbb{P}(X = k)$ pour $X \sim \mathcal{P}(\lambda = 1{,}48)$

↳ $$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

X_i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
0	37	37.75	$37 - 37.75 = -0.75$	$(-0.75)^2 = 0.5625$	$\frac{0.5625}{37.75} = 0.0149$
1	46	55.82	$46 - 55.82 = -9.82$	$(-9.82)^2 = 96.43$	$\frac{96.43}{55.82} = 1.728$
2	39	41.31	$39 - 41.31 = -2.31$	$(-2.31)^2 = 5.34$	$\frac{5.34}{41.31} = 0.129$
3	19	20.37	$19 - 20.37 = -1.37$	$(-1.37)^2 = 1.88$	$\frac{1.88}{20.37} = 0.0924$
4	5	7.53	$5 - 7.53 = -2.53$	$(-2.53)^2 = 6.40$	$\frac{6.40}{7.53} = 0.849$
5+	4	3.22	$4 - 3.22 = 0.78$	$(0.78)^2 = 0.61$	$\frac{0.61}{3.22} = 0.190$

6

Somme des contributions

$$\chi^2 = 0.0149 + 1.728 + 0.129 + 0.0924 + 0.849 + 0.190 = 2.088$$

↳ Quantile χ^2 : 5% à $k = (6-1)-1 = 4$ DDL :

$q_{95\%}^{DDL=4} : 9,49 > \chi^2 = 2,088 \Rightarrow$ On ne peut pas Rejeter H_0 !

8.1. **Test de conformité à une loi.**

Soit un échantillon $(X_1, ..., X_n)$ prenant un nombre fini de valeurs $\{r_1, ..., r_k\}$.

Soit des nombres réels strictement positifs $p_1, ..., p_k$ tels que $\sum_{i=1}^k p_i = 1$.

On souhaite tester l'hypothèse $\mathcal{H}_0$: "$\forall i = 1, ..., r,\ \mathbb{P}(X_1 = r_i) = p_i$" contre l'hypothèse $\mathcal{H}_1$ contraire.

La statistique du test est : $\boxed{D_n^2 := \sum_{i=1}^k \frac{(N(i) - np_i)^2}{np_i}}$, où $N(i) := Card\{j = 1, ...n \ : \ X_q = i\}$.

Soit $\alpha \in]0; 1[$. Le domaine de rejet du test au niveau α est : $\left\{D_n^2 \geq t_{(\alpha,k-1)}\right\}$ où $t_{(\alpha,k-1)}$ est l'unique nombre réel tel que $\chi^2_{k-1}(] - \infty; t_{(\alpha,k-1)}]) = 1 - \alpha$.

Ce test s'applique dès que $\min_{i=1,...,k} np_i \geq 5$ (chaque facteur devant contribuer de manière raisonnable à D_n^2).

Ce test peut s'appliquer :

- à une loi discrète quelconque, en regroupant les valeurs de queues dans une même classe.
- à une loi continue en la discrétisant (en considérant par exemple des classes équiprobables). Mais cela peut être maladroit. Pour les lois continues, il existe un autre test que nous ne verrons pas : le test de Kolmogorov-Smirnov.

Si vous devez estimer un ou plusieurs paramètres : si pour déterminer la loi, vous devez estimer ℓ paramètres (tels que l'espérance estimée par la moyenne empirique ou bien la variance estimée par la variance empirique), alors $\mathcal{D}_n$ suivra approximativement un loi $\chi^2_{k-\ell-1}$ (k est le nombre de classe et ℓ le nombre de paramètres estimés).

Par exemple, pour tester si un échantillon est de loi de Poisson, pour déterminer les p_i, il faut connaître le paramètre de la loi de Poisson (que l'on peut estimer par exemple par la moyenne empirique). Il faudra alors retrancher 1 au nombre de degré de liberté de la loi du chi-deux.

Exercice 3. *Une société d'assurances a comptabilisé, parmi ses 500 assurés, ceux qui ont déclaré un (ou plusieurs) sinistres au cours d'une année. Les résultats sont présentés dans le tableau suivant :*

Sinistres déclarés	0	1	2	3	≥ 4
Nombre d'assurés	171	202	80	36	11

Peut-on admettre au niveau 5% que le nombre de sinistres déclarés par un assuré suit une loi de Poisson de paramètre 1?

H_0 : "Les données suivent une loi de Poisson de paramètre $\lambda=1$"

H_1 : "Les données ne suivent pas une loi de Poisson de paramètre $\lambda=1$"

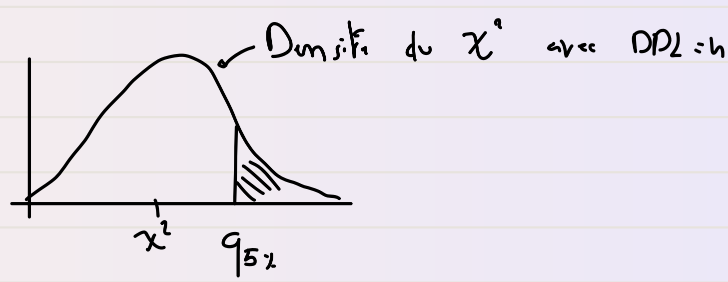
Sinistres (k)	Observé O_k	Attendu E_k	$(O_k - E_k)^2/E_k$
0	171	183.95	$(171 - 183.95)^2/183.95 = 0.912$
1	202	183.95	$\frac{(202-183.95)^2}{183.95} = 1.77$
2	80	91.95	$\frac{(80-91.95)^2}{91.95} = 1.552$
3	36	30.65	$\frac{(36-30.65)^2}{30.65} = 0.934$
≥4	11	9.15	$\frac{(11-9.15)^2}{9.15} = 0.374$

$\chi^2 = 0{,}912 + 1{,}77 + \dots + 0{,}374 = 6{,}184$

$ddl = 5 - 1 = 4$ (⚠ Ici la moyenne $\lambda = 1$ est donnée a priori, ce n'est donc pas un ddl)

↳ Quantile χ^2 à $\alpha = 5 \cdot 10^{-2}$ pour DDL = 4 : $q_{5\%} = 9{,}49$

↳ Ici $q_{5\%} \geqslant \chi^2$ ie



↳ On ne peut pas rejeter H_0.

Exercice 4. *Les résultats de l'évolution d'une maladie sur 1000 personnes ayant suivi l'un ou l'autre des traitements A et B sont résumés dans le tableau ci-dessous :*

	Guérison	Amélioration	Stationnaire	Totaux
Traitement A	280	210	110	600
Traitement B	220	90	90	400
Totaux	500	300	200	1000

Peut-on conclure au niveau $\alpha = 0,05$ que les traitements A et B ont le même effet?

H_0 : « L'évolution de la maladie est indépendante du traitement »

H_1 : « L'évolution de la maladie dépend du traitement »

↳ Tableau des effectifs théoriques (sous H_0)

↳ Si c'était indépendant, « les proba se multiplieraient » :

par exemple :

$\mathbb{P}(\text{Guérison} \cap A)$ devrait être égal à $\mathbb{P}(\text{Guérison}) \times \mathbb{P}(A)$

$$= \frac{500}{1000} \times \frac{600}{1000}$$

Donc le nombre théorique de guérisons avec A devrait être

$$1000 \times \mathbb{P}(\text{Guérison} \cap A) = \frac{5\cancel{0}\cancel{0} \times 6\cancel{0}\cancel{0}}{1\cancel{0}\cancel{0}\cancel{0}} = 300$$

Traitement \ Résultat	Guérison	Amélioration	Stationnaire	Total
Traitement A (600)	$\frac{600\times500}{1000} = 300$	$\frac{600\times300}{1000} = 180$	$\frac{600\times200}{1000} = 120$	600
Traitement B (400)	$\frac{400\times500}{1000} = 200$	$\frac{400\times300}{1000} = 120$	$\frac{400\times200}{1000} = 80$	400
Totaux	**500**	**300**	**200**	**1000**

La statistique du Chi-deux observée est donnée par :

$$\chi^2_{\text{obs}} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Calcul détaillé :

Case	**Observé** O_{ij}	**Attendu** E_{ij}	$(O_{ij} - E_{ij})^2/E_{ij}$
A, Guérison	280	300	$\frac{(280-300)^2}{300} = \frac{400}{300} = 1.333$
A, Amélioration	210	180	$\frac{(210-180)^2}{180} = \frac{900}{180} = 5.000$
A, Stationnaire	110	120	$\frac{(110-120)^2}{120} = \frac{100}{120} = 0.833$
B, Guérison	220	200	$\frac{(220-200)^2}{200} = \frac{400}{200} = 2.000$
B, Amélioration	90	120	$\frac{(90-120)^2}{120} = \frac{900}{120} = 7.500$
B, Stationnaire	90	80	$\frac{(90-80)^2}{80} = \frac{100}{80} = 1.250$

On additionne tout :

$$\chi^2_{\text{obs}} = 1.333 + 5.000 + 0.833 + 2.000 + 7.500 + 1.250 = 17.916$$

$\hookrightarrow$ $DDL = (\#\text{lignes} - 1)(\#\text{colonnes} - 1) = (2-1)(3-1) = 2$

$\hookrightarrow$ $q_{5\%}^{DDL=2} = 5,991$

$\hookrightarrow$ $\chi^2 > q_{5\%}^{DDL=2} \Rightarrow$ on rejette H_0

Exercice 5. *On veut savoir si le temps écoulé depuis la vaccination contre une maladie donnée a ou non une influence sur le degré de gravité de la maladie lorsqu'elle apparaît.*

Pour simplifier, nous ne distinguons que trois degrés de gravité.

Parmi les malades, nous comparons les vaccinés depuis moins de 25 ans et ceux vaccinés depuis plus de 25 ans :

Degré de gravité	*Légère*	*Moyenne*	*Forte*
vaccin $<$ 25 *ans*	43	120	324
vaccin $>$ 25 *ans*	230	347	510

Conclure aux niveaux 0,05 et 0,01.

1

Temps depuis vaccination	Gravité légère	Gravité moyenne	Gravité forte	Total
Vaccin < 25 ans	43	120	324	487
Vaccin > 25 ans	230	347	510	1087
Total	273	467	834	1574

- H_0 : Le degré de gravité est indépendant du temps écoulé depuis la vaccination.
- H_1 : Le degré de gravité dépend du temps écoulé depuis la vaccination.

$$E_{ij} = \frac{\text{(total ligne i)} \times \text{(total colonne j)}}{\text{total général}}$$

Calcul détaillé :

Temps \ Gravité	**Légère**	**Moyenne**	**Forte**	**Total**
Vaccin < 25 ans	$\frac{487\times273}{1574} = 84.49$	$\frac{487\times467}{1574} = 144.50$	$\frac{487\times834}{1574} = 258.01$	487
Vaccin > 25 ans	$\frac{1087\times273}{1574} = 188.51$	$\frac{1087\times467}{1574} = 322.50$	$\frac{1087\times834}{1574} = 575.99$	1087
Total	**273**	**467**	**834**	1574

$$\chi^2_{obs} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Calcul détaillé :

Case	Observé O_{ij}	Théorique E_{ij}	$\frac{(O-E)^2}{E}$
<25 ans, Légère	43	84.49	$\frac{(43-84.49)^2}{84.49} = 20.38$
<25 ans, Moyenne	120	144.50	$\frac{(120-144.50)^2}{144.50} = 4.15$
<25 ans, Forte	324	258.01	$\frac{(324-258.01)^2}{258.01} = 16.89$
>25 ans, Légère	230	188.51	$\frac{(230-188.51)^2}{188.51} = 9.13$
>25 ans, Moyenne	347	322.50	$\frac{(347-322.50)^2}{322.50} = 1.86$
>25 ans, Forte	510	575.99	$\frac{(510-575.99)^2}{575.99} = 7.56$

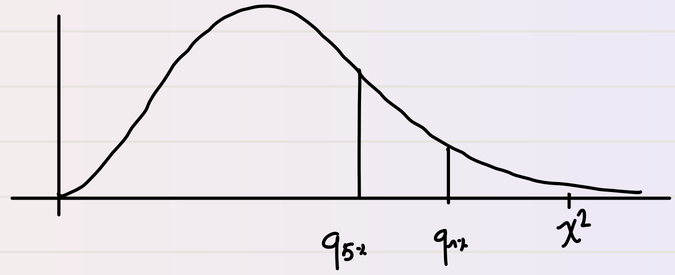
Somme totale :

$$\chi^2_{obs} = 20.38 + 4.15 + 16.89 + 9.13 + 1.86 + 7.56 = 59.97$$

↳ $DDL = (2-1) \times (3-1) = 2$

↳ $q_{5\%}^{DDL=2} = 5{,}991$

$q_{1\%}^{DDL=2} = 9{,}210$



↳ On rejette H_0 dans les 2 cas ! $q_{5\%} < q_{1\%} \ll \chi^2$!

Exercice 6. *On étudie deux caractères et on classe les effectifs observés par couples de valeurs dans le tableau suivant.*

Les deux caractères peuvent-ils être considérés comme indépendants au niveau 0,01?

	0	1	2	≥ 3
0	130	82	68	20
1	75	73	36	16
2	35	25	16	24

Caractère 1 \ Caractère 2	0	1	2	≥3	Totaux
0	130	82	68	20	300
1	75	73	36	16	200
2	35	25	16	24	100
Totaux	240	180	120	60	600

- Hypothèse nulle H_0 : Les deux caractères sont indépendants.
- Hypothèse alternative H_1 : Les deux caractères ne sont pas indépendants.

$$E_{ij} = \frac{(\text{total ligne i}) \times (\text{total colonne j})}{\text{total général}}$$

Calcul détaillé des effectifs théoriques :

Cellule	Calcul	Valeur théorique E_{ij}
(0,0)	$\frac{300\times240}{600}$	120
(0,1)	$\frac{300\times180}{600}$	90
(0,2)	$\frac{300\times120}{600}$	60
(0,≥3)	$\frac{300\times60}{600}$	30
(1,0)	$\frac{200\times240}{600}$	80
(1,1)	$\frac{200\times180}{600}$	60
(1,2)	$\frac{200\times120}{600}$	40
(1,≥3)	$\frac{200\times60}{600}$	20
(2,0)	$\frac{100\times240}{600}$	40
(2,1)	$\frac{100\times180}{600}$	30
(2,2)	$\frac{100\times120}{600}$	20
(2,≥3)	$\frac{100\times60}{600}$	10

$$\chi^2_{\text{obs}} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Calcul détaillé :

Case	Observé (O)	Théorique (E)	$\frac{(O-E)^2}{E}$
(0,0)	130	120	$\frac{(130-120)^2}{120} = 0.833$
(0,1)	82	90	$\frac{(82-90)^2}{90} = 0.711$
(0,2)	68	60	$\frac{(68-60)^2}{60} = 1.067$
(0,≥3)	20	30	$\frac{(20-30)^2}{30} = 3.333$
(1,0)	75	80	$\frac{(75-80)^2}{80} = 0.313$
(1,1)	73	60	$\frac{(73-60)^2}{60} = 2.817$
(1,2)	36	40	$\frac{(36-40)^2}{40} = 0.400$
(1,≥3)	16	20	$\frac{(16-20)^2}{20} = 0.800$
(2,0)	35	40	$\frac{(35-40)^2}{40} = 0.625$
(2,1)	25	30	$\frac{(25-30)^2}{30} = 0.833$
(2,2)	16	20	$\frac{(16-20)^2}{20} = 0.800$
(2,≥3)	24	10	$\frac{(24-10)^2}{10} = 19.600$

Somme totale :

$$\chi^2_{\text{obs}} = 0.833 + 0.711 + 1.067 + 3.333 + 0.313 + 2.817 + 0.400 + 0.800 + 0.625 + 0.833 + 0.800 + 19.600 = 32.132$$

$$DDL = (\#\text{lignes} - 1)(\#\text{colonnes} - 1) = (3-1)(4-1) = 6$$

$$q_{1\%}^{DDL=6} = 16,812$$

$$\hookrightarrow \chi^2 > q_{1\%}^{DDL=6} \Rightarrow \text{On rejette } H_0 \,!$$