U.F.R. S.P.S.E.
Licence de Psychologie
U.E. 341 A. Mollié

<u>CORRIGES DES EXERCICES</u> : Tests du khi-deux d'homogénéité et d'indépendance

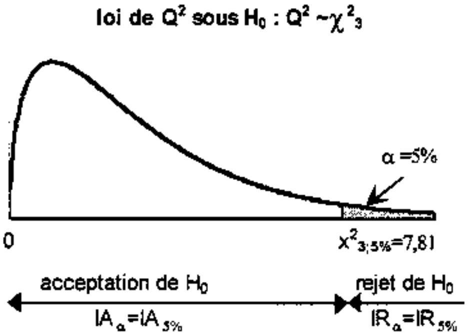
<u>Exercice 1</u>

$\mathcal{P}$={patients d'un hôpital psychiatrique}, X=saison définie sur E={printemps, été, automne, hiver} variable qualitative à 4 modalités, Y= réaction à un certain médicament définie sur F={oui, non} variable qualitative à 2 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0: \text{X et Y indépendantes} \\ H_1: \text{X et Y liées} \end{cases}$ au risque $\alpha=5\%$.

Sur deux échantillons appariés de X et de Y de taille n=490, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X saison	Y réaction oui		non	
printemps	55	(54,9)	64	(64,1)
été	59	(54,9)	60	(64,1)
automne	52	(53)	63	(62)
hiver	60	(63,2)	77	(73,8)



Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 3 ddl car n>30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $\mathcal{R}_{5\%} = \left[x^2_{3;5\%};+\infty\right[= [7,81;+\infty[$ car $x^2_{3;5\%}$ est le quantile d'ordre 0,95 de la loi χ^2_3, et la valeur observée de Q^2 vaut $q^2 = 0,9 \notin \mathcal{R}_{5\%}$. Donc on ne rejette pas H_0 au risque 5% : il n'existe pas de lien entre la saison et la réaction au risque $\alpha=5\%$.

<u>Exercice 2</u>

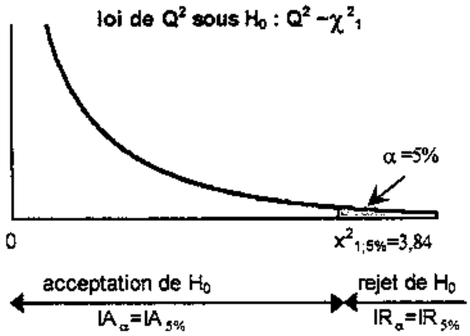
$\mathcal{P}_1$={lycéens du lycée A}, X_1= résultat à l'examen dans le lycée A
$\mathcal{P}_2$={lycéens du lycée B}, X_2= résultat à l'examen dans le lycée B
X_1 et X_2 variables qualitatives à 2 modalités définies sur E={réussite, échec}.

Test du khi-deux d'homogénéité sur 2 populations avec une variable à 2 modalités c'est à dire un test de comparaison de deux proportions sur 2 échantillons indépendants (procédure bilatérale) :

$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ où $\begin{bmatrix} p_1 = \text{proportion de réussite dans le lycée A} \\ p_2 = \text{proportion de réussite dans le lycée B} \end{bmatrix}$

Sur deux échantillons indépendants de tailles $n_1=100$ et $n_2=120$, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X résultat	lycée A		B	
réussite	40	(34,1)	35	(40,9)
échec	60	(65,9)	85	(79,1)



Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 1 ddl car $n=n_1+n_2>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $\mathcal{R}_{5\%} = \left[x^2_{1;5\%};+\infty\right[= [3,84;+\infty[$ car $x^2_{1;5\%}$ est le quantile d'ordre 0,95 de la loi χ^2_1, et la valeur observée de Q^2 : $q^2 = 2,85 \notin \mathcal{R}_{5\%}$. Donc on ne rejette pas H_0 au risque 5% : la proportion de réussite est la même dans les deux lycées au risque $\alpha=5\%$.

La région de rejet du test au risque 10% est $\mathfrak{R}_{10\%} = [x^2_{1;10\%};+\infty[= [2{,}71;+\infty[$ car $x^2_{1;10\%}$ est le quantile d'ordre 0,9 de la loi χ^2_1, et la valeur observée de Q^2 : $q^2 = 2{,}85 \in \mathfrak{R}_{10\%}$. Donc on rejette H_0 au risque 10% : la proportion de réussite est différente dans les deux lycées au risque $\alpha=10\%$.

Correction de Yates : sous H_0, la statistique de test du khi-deux corrigé Q_Y^2 suit approximativement une loi du khi-deux à 1 ddl car $n>30$ et tous les e_{ij} sont supérieurs à 5. La valeur observée de Q_Y^2 : $q^2_Y = 2{,}39$.

$q^2_Y \notin \mathfrak{R}_{5\%}$: on ne rejette pas H_0 au risque 5%.

$q^2_Y \notin \mathfrak{R}_{10\%}$: on ne rejette pas H_0 au risque 10%.

Exercice 3

$\mathcal{P}$={personnes}, X=niveau d'instruction définie sur E={BEPC, BAC, universitaire} variable qualitative à 3 modalités, Y= opinion sur le droit de vote à 16 ans définie sur F={pour, contre} variable qualitative à 2 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0: \text{X et Y indépendantes} \\ H_1: \text{X et Y liées} \end{cases}$ au risque $\alpha=5\%$.

Sur deux échantillons appariés de X et de Y de taille n=250, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

	opinion			
niveau d'instruction	pour		contre	
BEPC	10	(5,1)	15	(19,9)
BAC	21	(21,4)	84	(83,6)
universitaire	20	(24,5)	100	(95,5)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 2 ddl car $n>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $\mathfrak{R}_{5\%} = [x^2_{2;5\%};+\infty[= [5{,}99;+\infty[$ car $x^2_{2;5\%}$ est le quantile d'ordre 0,95 de la loi χ^2_2, et la valeur observée de Q^2 vaut $q^2 = 6{,}95 \in \mathfrak{R}_{5\%}$. Donc on rejette H_0 au risque 5% : il existe un lien entre l'opinion d'une personne et son niveau d'instruction au risque $\alpha=5\%$.

Exercice 4

$\mathcal{P}_1$={étudiants du collège I}, X_1= participation à un club sportif dans le collège I

$\mathcal{P}_2$={étudiants du collège II}, X_2= participation à un club sportif dans le collège II

X_1 et X_2 variables qualitatives à 2 modalités définies sur E={oui, non}.

Test du khi-deux d'homogénéité sur 2 populations avec une variable à 2 modalités c'est à dire un test de comparaison de deux proportions sur 2 échantillons indépendants (procédure bilatérale) :

$$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases} \quad \text{où} \quad \begin{cases} p_1 = \text{taux de participation dans le collège I} \\ p_2 = \text{taux de participation dans le collège II} \end{cases}$$

Sur deux échantillons indépendants de X_1 et X_2 de tailles $n_1=50$ et $n_2=60$, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

	collège			
X participation	I		II	
oui	12	(17,3)	26	(20,7)
non	38	(32,7)	34	(39,3)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 1 ddl car $n=n_1+n_2>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $\mathfrak{R}_{5\%} = [x^2_{1;5\%};+\infty[= [3{,}84;+\infty[$ car $x^2_{1;5\%}$ est le quantile d'ordre 0,95 de la loi χ^2_1, et la valeur observée de Q^2 : $q^2 = 4{,}51 \in \mathfrak{R}_{5\%}$. Donc on rejette pas H_0 au risque 5% : le taux de participation n'est pas le même dans les deux collèges au risque $\alpha=5\%$.

Correction de Yates : sous H_0, la statistique de test du khi-deux corrigé Q_Y^2 suit approximativement une loi du khi-deux à 1 ddl car $n=n_1+n_2>30$ et tous les e_{ij} sont supérieurs à 5. La valeur observée de Q_Y^2 : $q^2_Y = 3{,}69 \notin \mathfrak{R}_{5\%}$. Donc on ne rejette pas H_0 au risque 5%: le taux de participation est le même dans les deux collèges au risque $\alpha=5\%$. Dans ce cas les conclusions diffèrent entre les tests du khi-deux et du khi-deux corrigé.

Exercice 5

$\mathcal{P}$={enfants}, X=note au test de Peabody définie sur E={[60;86[; [86;111[} variable qualitative à 2 modalités, Y= sexe définie sur F={fille, garçon} variable qualitative à 2 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0 : \text{X et Y indépendantes} \\ H_1 : \text{X et Y liées} \end{cases}$ au risque $\alpha=10\%$.

Sur deux échantillons appariés de X et Y de taille n=85, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X note	Y sexe			
	fille		garçon	
[60 - 86[	31	(30,1)	33	(9,9)
[86 - 111[	9	(33,9)	12	(11,1)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 1 ddl car n>30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 10% est $\mathcal{R}_{10\%} = [x^2_{1;10\%};+\infty[= [2{,}71;+\infty[$ car $x^2_{1;10\%}$ est le quantile d'ordre 0,9 de la loi χ^2_1, et la valeur observée de Q^2 : $q^2 = 0{,}21 \notin \mathcal{R}_{10\%}$. Donc on ne rejette pas H_0 au risque 10% : le sexe et la note au test de Peabody sont indépendantes au risque $\alpha=10\%$.

Correction de Yates : sous H_0, la statistique de test du khi-deux corrigé $Q_Y{}^2$ suit approximativement une loi du khi-deux à 1 ddl car n>30 et tous les e_{ij} sont supérieurs à 5. La valeur observée de $Q_Y{}^2$: $q^2_Y = 0{,}04 \notin \mathcal{R}_{10\%}$ donc on ne rejette pas H_0 au risque 10%.

Exercice 6

$\mathcal{P}$={enfants}, X=réussite scolaire définie sur E={bonne, moyenne, médiocre} variable qualitative à 3 modalités, Y= adaptation sociale définie sur F={ bonne, moyenne, mauvaise} variable qualitative à 3 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0 : \text{X et Y indépendantes} \\ H_1 : \text{X et Y liées} \end{cases}$ aux risques $\alpha=5\%$ et $\alpha=1\%$.

Sur deux échantillons appariés de X et de Y de taille n=200, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X réussite scolaire	Y adaptation sociale					
	bonne		moyenne		mauvaise	
bonne	29	(20,4)	18	(20,4)	13	(19,2)
moyenne	22	(25,5)	29	(25,5)	24	(24)
médiocre	17	(22,1)	21	(22,1)	27	(20,8)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 4 ddl car n>30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $\mathcal{R}_{5\%} = [x^2_{4;5\%};+\infty[= [9{,}49;+\infty[$ car $x^2_{4;5\%}$ est le quantile d'ordre 0,95 de la loi χ^2_4, et la valeur observée de Q^2 vaut $q^2 = 9{,}95 \in \mathcal{R}_{5\%}$. Donc on rejette H_0 au risque 5% : il existe un lien entre réussite scolaire et adaptation sociale au risque $\alpha=5\%$.

La région de rejet du test au risque 1% est $\mathcal{R}_{1\%} = [x^2_{4;1\%};+\infty[= [13{,}28;+\infty[$ car $x^2_{4;1\%}$ est le quantile d'ordre 0,99 de la loi χ^2_4, et la valeur observée de Q^2 : $q^2 = 9{,}95 \notin \mathcal{R}_{1\%}$. Donc on ne rejette pas H_0 au risque 1% : réussite scolaire et adaptation sociale sont indépendantes au risque $\alpha=1\%$.

Exercice 7

$\mathcal{P}$={personnes}, X=opinion sur l'avortement définie sur E={pour, contre} variable qualitative à 2 modalités, Y= nombre d'années de scolarité définie sur F={moins de 8 ans, entre 9 et 12 ans, plus de 12 ans} variable qualitative à 3 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0: \text{X et Y indépendantes} \\ H_1: \text{X et Y liées} \end{cases}$ au risque $\alpha=0{,}001$.

Sur deux échantillons appariés de X et de Y de taille n=625, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X scolarité	Y opinion sur l'avortement pour	contre
moins de 8 ans	31 (44,3)	56 (42,7)
entre 9 et 12 ans	171 (177)	177 (171)
plus de 12 ans	116 (96,7)	74 (93,3)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 2 ddl car n>30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 0,001 est $\mathcal{R}_{0{,}001} = [x^2_{2;0{,}001};+\infty[= [13{,}82;+\infty[$ car $x^2_{2;0{,}001}$ est le quantile d'ordre 0,999 de la loi χ^2_2, et la valeur observée de Q^2 : $q^2 = 16{,}38 \in \mathcal{R}_{0{,}001}$. Donc on rejette H_0 au risque 0,001 : il existe un lien entre l'opinion sur l'avortement et la durée de la scolarité au risque $\alpha=0{,}001$.

Exercice 8

$\mathcal{P}_1$={élèves de terminale du département I}, X_1= nature du bac préparé dans le département I
$\mathcal{P}_2$={élèves de terminale du département II}, X_2= nature du bac préparé dans le département II
$\mathcal{P}_3$={élèves de terminale du département III}, X_3= nature du bac préparé dans le département III
X_1, X_2 et X_3 variables qualitatives à 4 modalités définies sur E={A, B, C, D}.

Test du khi-deux d'homogénéité : $\begin{cases} H_0: \text{les 3 populations sont homogènes vis à vis de X} \\ H_1: \text{les 3 populations sont hétérogènes vis à vis de X} \end{cases}$ au risque $\alpha=1\%$

ou $\begin{cases} H_0: \text{X a la même distribution dans les 3 populations} \\ H_1: \text{X a des distributions différentes dans les 3 populations} \end{cases}$

Sur trois échantillons indépendants de X_1, X_2 et X_3, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

X baccalauréat	département I		II		III	
A	11	(32,2)	30	(25,3)	39	(22,5)
B	51	(46,7)	36	(36,7)	29	(32,7)
C	32	(20,9)	14	(16,4)	6	(14,6)
D	46	(40,2)	30	(31,6)	24	(28,2)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 6 ddl car $n=n_1+n_2+n_3=140+110+98>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 1% est $\mathcal{R}_{1\%} = [x^2_{6;1\%};+\infty[= [16{,}81;+\infty[$ car $x^2_{6;1\%}$ est le quantile d'ordre 0,99 de la loi χ^2_6, et la valeur observée de Q^2 : $q^2 = 40{,}55 \in \mathcal{R}_{1\%}$. Donc on rejette H_0 au risque 1% : les élèves de terminale ne se répartissent pas de façon homogène dans les trois départements choisis, quant à la nature du baccalauréat qu'ils préparent, au risque $\alpha=1\%$.

Exercice 9

$\mathcal{P}_1$={sujets du groupe expérimental recevant le somnifère}, X_1= amélioration du sommeil avec le somnifère
$\mathcal{P}_2$={sujets du groupe témoin recevant le placebo}, X_2= amélioration du sommeil avec le placebo
X_1 et X_2 variables qualitatives à 2 modalités définies sur E={oui, non}.

Test du khi-deux d'homogénéité sur 2 populations avec une variable à 2 modalités c'est à dire un test de comparaison de deux proportions sur 2 échantillons indépendants (procédure bilatérale) :

$\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ où $\begin{bmatrix} p_1 = \text{proportion d'amélioration du sommeil dans le groupe expérimental} \\ p_2 = \text{proportion d'amélioration du sommeil dans le groupe témoin} \end{bmatrix}$

Sur deux échantillons indépendants de X_1 et X_2 de tailles $n_1=n_2=50$, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

	groupe			
X amélioration	expérimental		témoin	
oui	30	(27,5)	25	(27,5)
non	20	(22,5)	25	(22,5)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 1 ddl car $n=n_1+n_2>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 1% est $\mathcal{R}_{1\%} = \left[x^2_{1;1\%};+\infty\right[= [6{,}63;+\infty[$ car $x^2_{1;1\%}$ est le quantile d'ordre 0,99 de la loi χ^2_1, et la valeur observée de Q^2 : $q^2 = 1{,}01 \notin \mathcal{R}_{1\%}$. Donc on ne rejette pas H_0 au risque 1% : la proportion d'amélioration du sommeil est la même dans les deux groupes c'est à dire que le somnifère n'est pas efficace, au risque $\alpha = 1\%$.

Correction de Yates : sous H_0, la statistique de test du khi-deux corrigé Q_Y^2 suit approximativement une loi du khi-deux à 1 ddl car $n>30$ et tous les e_{ij} sont supérieurs à 5. La valeur observée de Q_Y^2 : $q_Y^2 = 0{,}65 \notin \mathcal{R}_{1\%}$ donc on ne rejette pas H_0 au risque 1%.

Exercice 10

$\mathcal{P}_1$ = population des cas = {femmes atteintes d'un cancer du sein}, X_1 = nombre d'enfants des cas

$\mathcal{P}_2$ = population des témoins = {femmes indemnes de cancer}, X_2 = nombre d'enfants des témoins

X_1 et X_2 variables qualitatives à 5 modalités définies sur E = {0, 1, 2, 3, plus de 4}.

Test du khi-deux d'homogénéité : $\begin{cases} H_0: \text{les 2 populations sont homogènes vis à vis de X} \\ H_1: \text{les 2 populations sont hétérogènes vis à vis de X} \end{cases}$ au risque $\alpha = 10\%$

ou $\begin{cases} H_0: \text{X a la même distribution dans les 2 populations} \\ H_1: \text{X a des distributions différentes dans les 2 populations} \end{cases}$

Sur deux échantillons indépendants de X_1 et X_2 de tailles $n_1=n_2=918$, les effectifs observés n_{ij} et les effectifs théoriques (attendus) e_{ij} notés entre parenthèses :

	population			
X nombre d'enfants	cas		témoin	
0	159	(138)	117	(138)
1	124	(123,5)	123	(123,5)
2	431	(438)	445	(438)
3	156	(165)	174	(165)
plus de 4	48	(53,5)	59	(53,5)

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 4 ddl car $n=n_1+n_2=2\times918>30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 10% est $\mathcal{R}_{10\%} = \left[x^2_{4;10\%};+\infty\right[= [7{,}78;+\infty[$ car $x^2_{4;10\%}$ est le quantile d'ordre 0,9 de la loi χ^2_4, et la valeur observée de Q^2 : $q^2 = 8{,}73 \in \mathcal{R}_{10\%}$. Donc on rejette H_0 au risque 10% : les nombres d'enfants ne se répartissent pas de façon homogène dans les deux populations, les distributions du nombre d'enfants dans les deux populations sont différentes au risque $\alpha = 10\%$.