Feuille TD 4 : Tests statistiques du χ^2 Correction

Exercice 4.1 Naissance et jour de la semaine

On désire étudier la répartition des naissances suivant le type du jour dans la semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du *National Vital Statistics Report* et concernent les naissances aux USA en 1997.

Naissances	Naturelles	César.	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Naissances	Naturelles	César.	Total
J.O.	60.6%	17.3%	77.9%
W.E.	18.6%	3.5%	22.1%
Total	79.2%	20.8%	100.0%

On définit les deux variables Y, correspondant au type de naissances, et Z, correspondant au jour de la semaine. Ces deux variables peuvent prendre 2 valeurs différentes et on souhaite tester

$$(H_0) : \text{"il y a indépendance entre les deux phénomènes" Vs } (H_1) : (H_0)^c.$$

On utiliser la statistique

$$U_n = n\sum_{i=1}^{2}\sum_{i=1}^{2}\frac{(f_{ij} - f_{\bullet j}f_{i\bullet})^2}{f_{\bullet j}f_{i\bullet}} = \sum_{i=1}^{2}\sum_{i=1}^{2}\frac{(N_{ij} - N_{\bullet j}N_{i\bullet})^2}{N_{\bullet j}N_{i\bullet}}.$$

Ici, on trouve que $U_n \approx 15994$. On doit comparer cette quantité à la distribution $\chi^2((2-1)(2-1)) = \chi^2(1)$. Pour $\alpha = 5\%$, on a $q_{1-\alpha}^{\chi^2(1)} = 3.84$, donc dans ce cas, on rejette l'hypothèse nulle.

Exercice 4.2 Est-il utile de mettre sa ceinture en voiture ?

Pour un échantillon de 10779 conducteurs, on a trouvé 5 blessures graves ou fatales en cas de port de la ceinture et 141 sans port de ceinture, 25 blessures sérieuses avec port de ceinture contre 330 en cas de non port de ceinture et enfin 1229 cas avec pas ou peu de blessure parmi les conducteurs portant la ceinture et 9049 parmi ceux ne portant pas la ceinture.

1. En utilisant un test du χ^2 de niveau 0.95, conclure sur l'efficacité du port de la ceinture de sécurité.

2. Estimez la p-valeur du test. Qu'en concluez-vous sur la significativité du test ?

On est dans le même cadre que l'exercice précédent, i.e. tester l'indépendance entre deux phénomènes. On remplit les tableaux suivants. On a $U_n = 17.81$ à comparer à la distribution $\chi^2(2)$. Le quantile d'ordre $1-\alpha$ pour $\alpha = 5\%$ de cette distribution est 5.99, donc on rejette l'hypothèse d'indépendance car $U_n > 5.99$.

E_{obs}/E_{th}	Port ceinture	Non port ceinture	Totaux
Blessures G ou F	$5/17,05$	$141/128,95$	146
Blessures S	$25/41,46$	$330/313,54$	355
Peu de blessures	$1229/1200.49$	$9049/9077.51$	10278
Totaux	1259	9520	10779

Ecarts normalisés	P.C.	N.P.C	Totaux
BGF	$8,516$	$1,126$	$9,642$
BS	$6,535$	$0,864$	$7,399$
PB	$0,677$	$0,09$	$0,766$
Totaux			$U_n = 17.81$

Exercice 4.3 Pratique sportive et âge

Sur un échantillon de 240 personnes, on a obtenu les effectifs suivants selon l'âge et le sport pratiqué :

Activité sportive\Âge	< 20	$[20, 25]$	$[25, 30]$	> 30	$Total$
Tennis	15	20	15	30	
Natation	15	10	20	25	
Course	20	10	30	30	
$Total$					

1. À l'aide d'un test du χ^2, examiner l'indépendance de ces deux caractères.
2. Estimez la p-valeur du test. Qu'en concluez-vous sur la significativité du test ?

Exercice 4.4

On jette un dé 600 fois afin de voir si le dé est truqué. On obtient les résultats suivants :

face obtenue	1	2	3	4	5	6
nombre de lancers	80	110	98	103	81	128

Avec un test du χ^2 pour les probabilités des faces obtenues, tester l'hypothèse nulle que toutes les probabilités sont égales à 1/6. Rejetez-vous l'hypothèse d'équilibre avec un niveau de confiance de 95% ? Estimez la p-valeur du test. Qu'en concluez-vous sur la significativité du test ?

Nous voulons tester si le dé est équilibré donc

$$(H_0) : "\pi = (1/6, \dots, 1/6)" \text{ Vs } (H_1) : (H_0)^c.$$

On utilise la statistique

$$U_n = n \sum_{i=1}^{r} \frac{(p_j - \hat{p}_j)^2}{p_j},$$

avec $r = 6$ et $p_j = 1/6$.

	1	2	3	4	5	6
Fréquence empirique	0,13333333	0,18333333	0,16333333	0,17166667	0,135	0,21333333
Fréquence théorique	0,16666667	0,16666667	0,16666667	0,16666667	0,16666667	0,16666667
Ecart normalisé	0,00666667	0,00166667	6,6667E-05	0,00015	0,00601667	0,01306667

U_n	16,58

Il faut comparer $U_n = 16.58$ à la distribution $\chi^2(6-1)$. Pour $\alpha = 5\%$, on trouve que le quantile vaut 11.07, on rejette donc l'hypothèse nulle. Concernant la p-valeur, on a $15.09 < 16.58 < 16.75$, donc la p-valeur est entre 1% et 0.5%. Donc on a une très forte présomption contre (H_0).

Exercice 4.5

Dans un supermarché, on relève un vendredi soir le nombre de clients passant à chacune des huits caisses. On obtient les résultats suivants :

Numéro de la caisse	1	2	3	4	5	6	7	8
Nombre de clients	72	70	71	52	45	55	71	48

1. Tester l'hypothèse que la clientèle se répartit uniformément entre les caisses ?
2. Estimez la p-valeur du test. Qu'en concluez-vous sur la significativité du test ?

Exercice 4.6 Attaque d'insectes

Les fruits d'un verger en agriculture biologique ont subi une attaque d'insectes. Sur un lot de quatre cents fruits, on compte le nombre d'insectes contenu dans chaque fruit. On obtient les résultats suivants :

Nombre d'insectes par fruit	0	1	2	3	4	5	6	7	≥ 8
Nombre de fruits	85	138	104	49	15	8	0	1	0

Les observations confirment-elles l'hypothèse selon laquelle le nombre d'insectes présents dans un fruit choisi arbitrairement est une variable aléatoire obéissant à une loi de Poisson ?