Université de Picardie Jules Verne

UFR d'économie et de gestion

Probabilités

Licence 2 - Semestre 3

Exercices d'entrainement **Test du χ^2** **Corrigés**

Correction 1

Population : les employés.
Deux caractères X : sites, à $r = 3$ modalités, et Y : postes, à $s = 3$ modalités.
Echantillon de taille $n = 300$.

H_0 : X et Y sont indépendantes
H_1 : X et Y ne sont pas indépendantes
Test de H_0 contre H_1.

Ce test s'appuie sur la distance D entre les effectifs observés et théoriques : $D = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(N_{i,j} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}}$.

Sous l'hypothèse H_0 d'indépendance de X et Y, les effectifs théoriques sont $\widetilde{\mathbf{n}}_{i,j} = \frac{n_{i,\star} \times n_{\star,j}}{n}$.

	Techniciens	Cadres moyens	Cadres supérieurs	Total
Site 1	105	18	12	135
Site 2	40	7	6	53
Site 3	77	20	15	112
Total	222	45	33	300

$\widetilde{\mathbf{n}}_{i,j}$	Techniciens	Cadres moyens	Cadres supérieurs	$n_{i,\star}$
Site 1	99,9	20,25	14,85	135
Site 2	39,22	7,95	5,83	53
Site 3	82,88	16,8	12,32	112
$n_{\star,j}$	222	45	33	300

La condition $\widetilde{\mathbf{n}}_{i,j} \geq 5$ est vérifiée.

$d_i = \frac{(n_{ij} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}}$	Techniciens	Cadres moyens	Cadres supérieurs
Site 1	0,2604	0,25	0,547
Site 2	0,0155	0,1135	0,005
Site 3	0,4172	0,6095	0,583

On obtient : $d = \sum_{i=1}^{3}\sum_{j=1}^{3} d_i = 2,8$.

On sait que sous l'hypothèse H_0, D suit approximativement la loi de χ^2 à $(r-1)(s-1) = (3-1) \times (3-1) = 4$ degrés de liberté. D'après la table du χ^2, pour $\alpha = 0,05$, on trouve $b_{\max} = 9,488$.

Comme $d < b_{\max}$, on ne rejette pas l'hypothèse H_0 d'indépendance de X et Y à un niveau de confiance de 95%. On peut donc considérer que les sites ont la même répartition en terme de personnel.

Correction 2

On dispose de $r = 2$ échantillons provenant de la population des produits vendus dans les deux villes
On étudie la variable X des différents produits vendus pouvant prendre $s = 5$ modalités.
Question : les deux populations présentent-elles les mêmes résultats?
Test de H_0 : les échantilllons sont issus de la même population contre $H_1 = \overline{H_0}$.
Ce test s'appuie sur la distance D entre les effectifs observés et théoriques : $D = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{i,j} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}}$
Sous l'hypothèse H_0 d'homogénéité des échantillons, les effectifs théoriques sont $\widetilde{\mathbf{n}}_{i,j} = \frac{n_{i\star}\, n_{\star,j}}{n}$.

$n_{i,j}$	Produit 1	Produit 2	Produit 3	Produit 4	Produit 5	$n_{i,\star}$
Ville A	32	16	32	8	8	96
Ville B	16	8	24	8	8	64
$n_{\star,j}$	48	24	56	16	16	160

Les effectifs théoriques sont :

$\widetilde{\mathbf{n}}_{i,j}$	Produit 1	Produit 2	Produit 3	Produit 4	Produit 5	$n_{i,\star}$
Ville A	28,8	14,4	33,6	9,6	9,6	96
Ville B	19,2	9,6	22,4	6,4	6,4	64
$n_{\star,j}$	48	24	56	16	16	160

$d_{i,j}$	Produit 1	Produit 2	Produit 3	Produit 4	Produit 5
Ville A	0,3556	0,1778	0,0762	0,2667	0,2667
Ville B	0,5333	0, 2667	0,1143	0,4	0,4

On obtient : $d = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}} = 2,86.$
On sait que sous l'hypothèse H_0, D suit approximativement la loi de χ^2 à $(r-1)(s-1) = (2-1) \times (5-1) = 4$ degrés de liberté. D'après la table du χ^2, pour $\alpha = 0,05$, on trouve $b_{\max} = 9,488$.
Comme $d \leq b_{\max}$, on ne peut pas rejeter l'hypothèse H_0 d'homogénéité des échantillons à un niveau de confiance de 95% : on peut donc considérer que les deux populations présentent les mêmes répartitions de vente.

Correction 3

Population : les véhicules.
Caractère X : le nombre de pannes à $r = 4$ modalités.
Echantillon $(X_1, ..., X_n)$ de taille $n = 23 + 17 + 11 + 9 = 60$.

1. Le nombre moyen de pannes dans l'échantillon est :
   $\overline{x} = \frac{0 \times 23 + 1 \times 17 + 2 \times 11 + 3 \times 9}{60} = \frac{66}{60} = 1,1,$
   $\overline{x^2} = \frac{0^2 \times 23 + 1^2 \times 17 + 2^2 \times 11 + 3^2 \times 9}{60} = \frac{142}{60} \approx 2,3666667,$
   $s^2 \approx 2,3666667 - 1,1^2 \approx 1,157$ et $s_c^2 \approx 1,177.$

2. On teste H_0 contre H_1 où H_0 est "X suit une loi de Poisson" et $H_1 = \overline{H_0}$ est "X ne suit pas une loi de Poisson".

   Comme $\overline{x}$ et s_c^2 sont très proches, on pouvait effectivement penser à une loi de Poisson.

   On peut alors estimer le paramètre λ à $1,1 (= \overline{x})$.

   Les p_i devant être calculés à l'aide la loi de Poisson, il y a un paramètre à estimer : $k = 1$.

   Les p_i sont obtenus grâce à la table :

   $p_0 = 0,3329$, $p_1 = 0,699 - 0,3329 = 0,3661$, $p_2 = 0,9004 - 0,699 = 0,2014$.

   Pour que le total des probabilités soit bien égal à 1, on considère que l'effectif de la classe "3 pannes" correspond à la classe "3 pannes et plus".

   Et on a $p_3 = 1 - 0,9004 = 0,0996$.

Valeurs x_i	Effectifs observés n_i	p_i	Effectifs théoriques $n \times p_i$	$d_i = \frac{(n_i - n\times p_i)^2}{n\times p_i}$
0	23	$0,3329$	$19,974$	$0,458$
1	17	$0,3661$	$21,996$	$1,222$
2	11	$0,2014$	$12,084$	$0,097$
3 et plus	9	$0,0996$	$5,976$	$1,530$
Total	60	1		$3,207$

   On sait que sous l'hypothèse H_0, D suit approximativement la loi de χ^2 à $r - k - 1 = 4 - 1 - 1 = 2$ degré de liberté.

   On a $d = \sum_{i=1}^{3} d_i \simeq 3,207$. Et, d'après la table du χ^2, pour $\alpha = 0,05$, on trouve $b_{\max} = 5,991$.

   Comme $d \leq b_{\max}$, on ne peut rejeter l'hypothèse H_0, i.e., la conformité à la loi théorique de Poisson : la répartition des pannes annuelles est conforme à une loi de Poisson. En prenant cette décision de non-rejet de H_0, on ne connait pas la probabilité de se tromper (erreur de deuxième espèce).

Correction 4

On teste H_0 contre H_1 où H_0 est "X suit une loi normale" et $H_1 = \overline{H_0}$ est "X ne suit pas une loi normale".

Les p_i devant être calculés à l'aide la loi normale, il y a deux paramètres à estimer (μ et σ) c'est-à-dire $k = 2$.

On a $\overline{x} = \dfrac{6 \times 1,5 + 32 \times 2,5 + 44 \times 4 + 18 \times 7,5}{6 + 31 + 44 + 18} = \dfrac{400}{100} = 4$

$\overline{x^2} = \dfrac{6 \times 1,5^2 + 32 \times 2,5^2 + 44 \times 4^2 + 18 \times 7,5^2}{6 + 31 + 44 + 18} = \dfrac{1930}{100} = 19,3$

$s^2 = 19,3 - 4^2 = 3,3$, $s_c^2 = \dfrac{100}{99} s^2 \approx 3,33$ et $s_c \approx 1,83$.

On peut alors estimer les paramètres μ et σ par 4 et $1,83$.

Il s'agira donc de tester si X suit la loi normale $\mathcal{N}(4; 1,83)$, i.e., si $Z = \dfrac{X - 4}{1,83}$ suit la loi normale $\mathcal{N}(0; 1)$.

Puisque la loi normale est définie sur $\mathbb{R}$, on change la première borne des intervalles en $-\infty$ et la dernière en $+\infty$.

$p(]-\infty;2[) = \varphi\left(\dfrac{2-4}{1,83}\right) \approx \varphi(-1,09) = 1-\varphi(1,09) \approx 1-0,8621 = 0,1379$

$p([2;3[) = \varphi\left(\dfrac{3-4}{1,83}\right) - \varphi\left(\dfrac{2-4}{1,83}\right) \approx \varphi(-0,55) - \varphi(-0,9)$

$= 1-\varphi(0,55) - 0,1379 \approx 1-0,7088-0,1379 = 0,2912-0,1379 = 0,1533$

$p([3;5[) = \varphi\left(\dfrac{5-4}{1,83}\right) - \varphi\left(\dfrac{3-4}{1,83}\right) \approx \varphi(0,55) - \varphi(-0,55) \approx 0,7088-0,2912 = 0,4176$

$p([5;+\infty[) = 1-\varphi\left(\dfrac{5-4}{1,83}\right) \approx 1-\varphi(0,55) \approx 1-0,7088 = 0,2912$

Autonomie (en heures)	$]-\infty;2[$	$[2,3[$	$[3,5[$	$[5;+\infty[$	Total
Nombre de vélos (n_i)	6	32	44	18	100
p_i	0,1379	0,1533	0,4176	0,2912	1
$n \times p_i$	13,79	15,33	41,76	29,12	100
d_i	4,4	18,13	0,12	4,25	26,9

La condition $n \times p_i \geq 5$ est vérifiée.

On a donc $d = \displaystyle\sum_{i=1}^{5} d_i \approx 26,9$.

Pour $\alpha = 0,05$ et $d.d.l. = 4-2-1 = 1$, on trouve $b_{\max} = 3,841$.

Comme $d > b_{\max}$, on rejette l'hypothèse H_0, i.e., la variable ne suit pas une loi normale.

Correction 5

Méthode 1 : Test de comparaison de proportions (voir feuille d'exercice sur le sujet)

On suppose $X_1 \sim \mathscr{B}(p_1)$ et $X_2 \sim \mathscr{B}(p_2)$

On a bien que $n_1 f_1 = 16$, $n_1(1-f_1) = 109$, $n_2 f_2 = 18$ et $n_2(1-f_2) = 197$ sont tous plus grand que 5.

On a $f_1 = \dfrac{16}{125} = 0,128 = 12,8\%$ et $f_1 = \dfrac{18}{215} \approx 0,084 = 8,4\%$

On pose $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \dfrac{16+18}{125+215} = \dfrac{34}{340} = 0,1 = 10\%$.

Et $z_{calcul} = \dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx 1,3122$.

De plus, $z_{critique} = 2,58$

Puisque $|z_{calcul}| < z_{critique}$, on ne rejette pas l'hypothèse $H_0 : p_1 = p_2$ au risque d'erreur α.

Méthode 2 : Test du χ^2

On peut également traiter cette question par un test d'indépendance des deux variables X : modèle de disque dur, à $r = 2$ modalités et Y : Nombre de pannes, à $s = 2$ modalités (pas de panne et au moins une panne).

Echantillon de taille $n = 340$.

H_0 : X et Y sont indépendantes

H_1 : X et Y ne sont pas indépendantes

Test de H_0 contre H_1.

Ce test s'appuie sur la distance D entre les effectifs observés et théoriques : $D = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{i,j} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}}$

Sous l'hypothèse H_0 d'indépendance de X et Y, les effectifs théoriques sont $\widetilde{\mathbf{n}}_{i,j} = \frac{n_{i,\star} \times n_{\star,j}}{n}$.

$n_{i,j}$	pannes	pas de panne	$n_{i,\star}$
Modèle A	16	109	125
Modèle B	18	197	215
$n_{\star,j}$	34	306	340

$\widetilde{\mathbf{n}}_{i,j}$	pannes	pas de panne	$n_{i,\star}$
Modèle A	12,5	112,5	125
Modèle B	21,5	193,5	215
$n_{\star,j}$	34	306	340

La condition $\widetilde{\mathbf{n}}_{i,j} \geq 5$ est vérifiée.

$d_{i,j} = \frac{(n_{ij} - \widetilde{\mathbf{n}}_{i,j})^2}{\widetilde{\mathbf{n}}_{i,j}}$	pannes	pas de panne
Modèle A	0,98	0,1089
Modèle B	0,5698	0,0633

On a $d = \sum_{i=1}^{2} \sum_{j=1}^{2} d_{i,j} = 1,72$.

On sait que sous l'hypothèse H_0, D suit approximativement la loi de χ^2 à $(r-1)(s-1) = (2-1) \times (2-1) = 1$ degré de liberté. D'après la table du χ^2, pour $\alpha = 0,01$, on trouve $b_{\max} = 6,635$.

Comme $d \leq b_{\max}$, on ne rejette pas l'hypothèse H_0 d'indépendance de X et Y.