CORRECTIONS TD 6

Exercice 1

Si on pouvait connaître avec certitude le nombre de chèvres que la société doit mettre à disposition de ses clients pour la tonte de leur jardin, cela permettrait d'anticiper le besoin et donc d'éviter toute rupture. Mais pour cela, il faudrait connaître de la loi suivie par le nombre de chèvres louées par jour.

Il vous donc demandé, à partir des données suivantes, observées sur le derniers mois d'activité, de montrer que X le nombre de chèvres louées chaque jour suit une loi de Poisson dont vous déterminerez le paramètre.

Nombre de chèvres	0-4	5	6	7	8	9	10	11	12-15	Total
Nombre d'observations	10	8	10	14	20	17	13	9	9	n = 110

Solution :

Il faut effectuer un test d'ajustement (d'adéquation de loi) du Chi-2 pour vérifier que les observations suivent bien une loi de Poisson.

1. Il faut dans un premier temps estimer le paramètre de la loi de Poisson. On sait que l'espérance d'une loi de Poisson est égale à son paramètre (en effet, d'après le cours 2, on sait que si $X \sim \mathcal{P}(\lambda)$ alors $\mathbb{E}(X) = \lambda$ et $\mathbb{V}(X) = \lambda$).

   Par ailleurs, on sait qu'un bon estimateur de l'espérance est la moyenne empirique $M = \frac{1}{n}\sum_{i=1}^{n} X_i$. On peut donc estimer le paramètre de la loi de Poisson par

$$\hat{\lambda} = \frac{1}{110}\sum_{i=1}^{110} X_i = \frac{(10 \times 2) + (8 \times 5) + \cdots + (9 \times 11) + (9 \times 13.5)}{110} \simeq 8$$

2. On va maintenant tester (avec le test du Chi-2) que les observations ("nombre de chèvres louées par jour") suivent une loi de Poisson $\mathcal{P}(8)$. L'hypothèse H_0 que l'on souhaite tester est donc

$$H_0 : X \sim \mathcal{P}(8)$$

3. On va maintenant calculer les probabilités théoriques sous l'hypothèse H_0. C'est-à-dire, on considère que H_0 est vraie ($X \sim \mathcal{P}(8)$) et on calcule les probabilités correspondant à chaque événement. En utilisant la table de la loi de Poisson on obtient

Nombre de chèvres	0-4	5	6	7	8	9	10	11	12-15	Total
Probabilité théorique	0.1	0.092	0.122	0.14	0.14	0.124	0.099	0.072	0.111	1

avec notamment $\mathbb{P}(X \in 0,1,2,3,4) = \mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3) + \mathbb{P}(X=4) \simeq 0.1$
et $\mathbb{P}(X \in 12,13,14,15) = \mathbb{P}(X=12) + \mathbb{P}(X=13) + \mathbb{P}(X=14) + \mathbb{P}(X=15) \simeq 0.111$.

4. Une fois qu'on a calculé les probabilités théoriques (p^j_{Theo}) sous H_0, on peut en déduire les effectifs théoriques (n^j_{Theo}) sous l'hypothèse H_0, en effet

$$n^j_{Theo} = p^j_{Theo} \times n$$

.

Nombre de chèvres	0-4	5	6	7	8	9	10	11	12-15	Total
Effectif théorique	11	10.1	13.4	15.4	15.4	13.6	10.9	7.9	12.3	110

Comme les effectifs théoriques sont tous ≥ 5, il n'y a pas de regroupement à faire.

5. On calcule ensuite l'écart ou la "distance" entre les effectifs observés (n^j_{Obs}) et les effectifs théoriques (n^j_{Theo})

$$\chi^2_{Obs} = \sum_{j=1}^{9} \frac{(n^j_{Obs} - n^j_{Theo})^2}{n^j_{Obs}}$$

Nombre de chèvres	0-4	5	6	7	8	9	10	11	12-15	Total
Effectif observé	10	8	10	14	20	17	13	9	9	n = 110
Effectif théorique	11	10.1	13.4	15.4	15.4	13.6	10.9	7.9	12.3	110
Ecart	0.09	0.44	0.86	0.13	1.37	0.85	0.40	0.15	0.89	$\chi^2_{Obs} = 5.18$

La valeur du $\chi^2_{Obs} = 5.18$

6. Pour savoir si on rejette l'hypothèse H_0 on veut savoir si cette distance (entre nos observations et l'hypothèse H_0) est suffisamment grande c'est-à-dire si elle est plus grande qu'un certain seuil.
Ce seuil dépend du risque d'erreur que l'on est prêt à faire α qui est généralement fixé à 5%.
Le seuil dépend également du nombre de degré de liberté qui est $ddl = m - p - 1$

où m représente le nombre de modalités après regroupement et p est le nombre de paramètre à estimer pour caractériser la loi.
Dans notre cas $ddl = 9 - 1 - 1 = 7$.

7. En utilisant ensuite une table de la loi du Chi-2, en prenant le degré de liberté ($ddl = 7$) en ligne et le taux d'erreur ($\alpha = 5\%$) en colonne, on obtient la valeur du $\chi^2_{Theo} = 14.07$.
8. On peut maintenant prendre une décision en comparant χ^2_{Obs} et χ^2_{Theo}. En effet on obtient que
$$\chi^2_{Obs} = 5.18 < 14.07 = \chi^2_{Theo}$$
Donc on ne rejette pas l'hypothèse H_0, c'est-à-dire qu'on ne rejette pas le fait que le nombre de chèvre louées par jour suit une loi de Poisson $\mathcal{P}(8)$.
9. En pratique on peut conclure que le nombre de chèvre louées par jour suit une loi de Poisson $\mathcal{P}(8)$.

<u>Exercice 2</u>

Grâce aux relevés des pisteurs sur l'enneigement des pistes lors de la saison dernière, on a recensé les informations suivantes :

Enneigement en cm	< 10	[10 ;20[	[20 ;30[	[30 ;50[	[50 ;100[	Total
Nombre de pistes	15	22	20	13	5	n=75

On note X le niveau d'enneigement sur les pistes l'année dernière.

1. Calculer la moyenne de l'échantillon
2. Si on considère les observations des dix dernières années on observe un écart-type de 15 cm. On souhaite utiliser cette valeur, on considère donc que l'écart-type est fixé à 15 cm. Par quelle loi pouvez vous modéliser la variable X ?
3. Peut-on considérer que l'enneigement observé l'année dernière suit la loi définie à la question 2, au taux d'erreur $\alpha = 5\%$?

Solution :

1. La moyenne de l'échantillon est
$M = \frac{(15\times5)+(22\times15)+(20\times25)+(13\times40)+(5\times75)}{75} = 24$
2. La quantité d'enneigement est une quantité continue (et qui est unimodale), il est donc naturel de la modéliser par une loi normale. D'après l'estimation de la moyenne sur l'échantillon et comme l'écart-type est fixé à 15 cm on suppose que
$$X \sim \mathcal{N}(24, 15)$$
3. On va faire un test d'ajustement du Chi-2 pour vérifier si les observations suivent bien une loi $\mathcal{N}(24, 15)$.

   (a) L'hypothèse H_0 que l'on souhaite tester est
   $$H_0 : X \sim \mathcal{N}(24, 15)$$

(b) Calcul des probabilités correspondant à chacun des événements sous l'hypothèse $H_0 : X \sim \mathcal{N}(24, 15)$

Enneigement en cm	< 10	[10 ;20[	[20 ;30[	[30 ;50[	> 50	Total
Probabilité théorique	$\mathbb{P}(X < 10) = 0.175$	0.22	0.261	0.302	$\mathbb{P}(X > 50) = 0.042$	1

Soit $U \sim \mathcal{N}(0, 1)$, on a par exemple en utilisant la table de la loi normale $\mathbb{P}(X < 10) = \mathbb{P}(\frac{X-24}{15} < \frac{10-24}{15}) = \mathbb{P}(U < -0.93) = 1 - \mathbb{P}(U < 0.93) = 1 - 0.8228 = 0.1772$
$\mathbb{P}(20 \leq X < 30) = \mathbb{P}(\frac{20-24}{15} \leq \frac{X-24}{15} < \frac{30-24}{15}) = \mathbb{P}(-0.27 \leq U < 0.4) = \mathbb{P}(U < 0.4) - \mathbb{P}(U < -0.27) = \mathbb{P}(U < 0.4) - (1 - \mathbb{P}(U < 0.27)) = 0.6554 - 1 + 0.6064 = 0.2618$

Remarque : les résultats du tableau sont obtenus avec un logiciel de statistique qui utilise une table de loi des arrondis plus précis que la table que vous avez. Ceci qui peut expliquer la légère différence entre les résultats obtenus avec la table et ceux donnés dans le tableau.

(c) Calcul des effectifs théoriques sous l'hypothèse $H_0 : X \sim \mathcal{N}(24, 15)$

Enneigement en cm	< 10	[10 ;20[	[20 ;30[	[30 ;50[	> 50	Total
Effectif théorique	13.1	16.6	19.5	22.7	3.1	75

On regroupe les deux dernières cases car l'effectif de la dernière modalité est ≤ 5.

Enneigement en cm	< 10	[10 ;20[	[20 ;30[	> 30	Total
Effectif théorique	13.1	16.6	19.5	25.8	75

Aucun regroupement n'est nécessaire car les effectifs théoriques sont tous ≥ 5

(d) On calcule ensuite l'écart ou la "distance" entre les effectifs observés (n^j_{Obs}) et les effectifs théoriques (n^j_{Theo})

$$\chi^2_{Obs} = \sum_{j=1}^{4} \frac{(n^j_{Obs} - n^j_{Theo})^2}{n^j_{Obs}}$$

Enneigement en cm	< 10	[10 ;20[	[20 ;30[	> 30	Total
Effectif observé	15	22	20	18	75
Effectif théorique	13.1	16.6	19.5	25.8	75
Ecart	0.28	1.76	0.01	2.36	$\chi^2_{Obs} = 4.41$

La valeur du $\chi^2_{Obs} = 4.41$

(e) D'après l'énoncé $\alpha = 5\%$. Comme il y a 4 modalités après regroupement dans le tableau et qu'on estime un seul paramètre (uniquement la moyenne car l'écart-type est fixé) on a $ddl = 4 - 1 - 1 = 2$.

(f) En utilisant une table de la loi du Chi-2, en prenant le degré de liberté ($ddl = 2$) et le taux d'erreur ($\alpha = 5\%$) , on obtient la valeur du $\chi^2_{Theo} = 5.99$.

(g) On observe que

$$\chi^2_{Obs} = 4.41 < 5.99 = \chi^2_{Theo}$$

Donc on ne rejette l'hypothèse H_0, c'est-à-dire qu'on ne rejette pas le fait que le nombre de chèvre louées par jour suit une loi normale $\mathcal{N}(24, 15)$. En pratique, on considère souvent que X suit bien une loi normale $\mathcal{N}(24, 15)$.

(h) On peut donc conclure que l'enneigement des pistes suit une loi normale $\mathcal{N}(24, 15)$.