Thème Statistique

Exercice 3 (Questions de cours (3 points)).

1. (0,5) - Le nombre d 'enfants par famille.
2. (0,5) - Les familles qui viennent au parc d'attraction.
3. (0,5) - Les nombres entiers
4. (0,5) - Les 100 familles qui viennent le plus souvent. Un biais est possible car on ne prendre pas les familles au hasard.
5. (1) - Diagramme en bâtons.

Exercice 4 (Table de distribution (3,5 points)).

1. (0,5) - discrète.
2. (0,5) - qualitative.
3. (0,5) - 20.
4. (1,5) -

classe	1	2	3
effectif	4	7	9
effectif cumulé	4	11	20
fréquence (%)	20	35	45
fréquence cumulée (%)	20	55	100

5. (0,5) - 2, 2 et 3.

Exercice 5 (Alcool et tabac (6,5 points)).

1. (1) - Moyenne empirique : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Estimation : $\bar{x} = 12.4$ donné par la fonction `mean`.
2. (1) - Variance corrigée : $S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Estimation : $s'^2 = 5.34$ donné par la fonction `var`.
3. (2) - On estime la moyenne d'un échantillon non gaussien. On utilise l'intervalle de confiance asymptotique :
$$\left[\bar{X} - u_{1-\alpha/2}\frac{S'}{\sqrt{n}}, \bar{X} + u_{1-\alpha/2}\frac{S'}{\sqrt{n}}\right],$$
où $\bar{X}$ est la moyenne empirique, S' est l'écart-type corrigé et $u_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$. Ce choix est justifié car l'échantillon est grand : $n = 100 \geq 100$.
4. (1) - Le calcul donne :
$$\left[12.4 - 1.96\frac{2.31}{10}, 12.4 + 1.96\frac{2.31}{10}\right] = [11.95, 12.85].$$
5. (0,5) - L'intervalle est $[11.37, 11.83]$, donné par la commande `t.test(X2)`.
6. (0,5) - Non. Le logiciel utilise les quantiles $t_{n-1;1-\alpha/2}$ de la loi de Student.
7. (0,5) - Celui de la question 5 est moins large que celui de la question 4 : cela s'explique parce que l'échantillon est plus grand. L'intersection des deux intervalles est vide : cela peut être un signe que la consommation est différente entre la population des fumeurs et celle des non fumeurs.

Exercice 6 (Voilà les Dalton (7 points)).

1. (1) - Fréquence empirique. Estimation : $f = 75/900 \approx 8,33\%$.

2. (2) - On estime une proportion. On utilise l'intervalle de confiance asymptotique :

$$\left[F - u_{1-\alpha/2}\frac{\sqrt{F(1-F)}}{\sqrt{n}}, F + u_{1-\alpha/2}\frac{\sqrt{F(1-F)}}{\sqrt{n}}\right],$$

où F est la fréquence empirique et $u_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$. Ce choix est justifié car l'échantillon est grand : $n = 900 \geq 100$.

3. (0,5) - L'intervalle est $[0.067, 0.104]$, donné par la commande `prop.test(75,900)`.

4. (0,5) - Non, comme on l'a vu en TP.

5. (0,5) - $\mathcal{H}_0 : p = p_0$ et $\mathcal{H}_1 : p < p_0$.

6. (1) - Test de conformité d'une proportion avec la statistique

$$T = \sqrt{n}\frac{F - p_0}{\sqrt{p_0(1-p_0)}},$$

et la règle de décision (au niveau de risque α) : rejet si et seulement si $t_{obs} < -u_{1-\alpha}$.

7. (1) - La statistique calculée sur les données est :

$$t_{obs} = \sqrt{900}\frac{\frac{75}{900} - 0.1}{\sqrt{0.1 * 0.9}} \approx -1.67.$$

La p-valeur de notre test est $\alpha^* = F(t_{obs}) = F(-1.67) = 1 - F(1.67) = 1 - 0.953 = 4.7\%$. La p-valeur est inférieure à 5%, donc on rejette $\mathcal{H}_0$: le probabilité d'être daltonien est inférieure à 10%.

8. (0,5) - Le résultat de la dernière question semble en contradiction avec le fait que 10% appartienne à l'intervalle de confiance de la question 3 : cela vient du fait que l'on compare un intervalle de confiance (bilatéral) avec un test unilatéral.