Questions types d'examen

Question 1

On estime la distance (km) séparant le lieu de résidence durant la semaine et l'ULg dans 2 échantillons indépendants

a) → comparaison de deux moyennes indépendantes avec $\sigma_1^2 = \sigma_2^2$

Etudiants possédant une voiture : $\bar{y}_1 = \frac{1}{8} \sum_{i=1}^{8} y_{i,1} = 20,75$ km

$$\hat{\sigma}_1^2 = \frac{1}{8} \sum_{i=1}^{8} y_{i,1}^2 - \bar{y}_1^2 = 61.44 \text{ km}^2$$

$$s_1^2 = \frac{8}{7} \hat{\sigma}_1^2 = 70.21 \text{ km}^2$$

Etudiants sans voiture : $\bar{y}_2 = \frac{1}{16} \sum_{i=1}^{16} y_{i,2} = 15,19$ km

$$\hat{\sigma}_2^2 = \frac{1}{16} \sum_{i=1}^{16} y_{i,2}^2 - \bar{y}_2^2 = 125,53 \text{ km}^2$$

$$s_2^2 = \frac{16}{15} \hat{\sigma}_2^2 = 133,90 \text{ km}^2$$

Distribution a posteriori de $\theta = \mu_1 - \mu_2$

$$(\theta \mid \text{donnees}) \sim t_{n_1+n_2-2}\left(\bar{y}_1 - \bar{y}_2, s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{7(70.21) + 15(133,90)}{8+16-2} = \frac{2499,94}{22} = 113,63$$

donc $(\theta \mid \text{donnees}) \sim t_{22}\left(20.75 - 15.19, 113.63\left(\frac{1}{8} + \frac{1}{16}\right)\right)$

$$\sim t_{22}(5,56 ; 21,31)$$

* Méthode 1 : Test de la plausibilité de $\mu_1 - \mu_2 > 0$

$$P(\mu_1 - \mu_2 > 0 \mid \text{donnees}) = P\left(\frac{\theta - 5.56}{\sqrt{21.31}} > \frac{0 - 5.56}{\sqrt{21.31}}\right) = P(T_{22} > -1.21)$$
$$= P(T_{22} < 1.21) < 0.90$$

Comme $P(T_{22} < 1.21) < 0.90$, l'hypothèse $\mu_1 - \mu_2 > 0$ n'est pas assez plausible. Les étudiants possédant une voiture n'habitent pas plus loin de l'ULg la semaine, en moyenne, que les étudiants qui ne possèdent pas de voiture.

* Méthode 2 : Intervalle de crédibilité à 95% pour $\theta = \mu_1 - \mu_2$

$$\bar{y}_1 - \bar{y}_2 \pm t_{n_1+n_2-2}(0,975)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_p^2}$$

$$= 5.56 \pm 2.0740\sqrt{21.31} = [-4,01\,;\,15,14] \ni 0$$

Comme 0 est une valeur plausible pour $\mu_1 - \mu_2$, on ne peut conclure à une différence de distance entre les deux groupes, comme pour la méthode 1.

b) Non. En effet, l'échantillon est composé uniquement d'étudiants de l'institut des Sciences Humaines. Pour pouvoir étendre les conclusions aux étudiants de l'ULg, il aurait dû être composé d'étudiants choisis aléatoirement dans l'université en général et pas seulement de l'institut des Sciences Humaines

Question 2

On estime la proportion de fumeurs en Sciences Humaines et en Médecine.

a) → comparaison de deux proportions indépendantes

Fumer

	oui	non	
S. H.	8	17	25
M	31	36	67

$\pi_{1|1}$ = Risque de Fumer en S.H.

$\pi_{1|2}$ = Risque de Fumer en M.

On a $\hat{\pi}_{1|1} = 8/25 = 0.32$ et $\hat{\pi}_{1|2} = 31/67 = 0.46$

Méthode 1 : comparaison de deux proportions indépendantes

$(\pi_{1|1} \mid Y_{1|1} = 8) \sim Beta(9, 18) \approx \mathcal{N}(0.32, 0.0087)$ si $\begin{cases} 25(0.32) > 5 \\ 25(0.68) > 5 \end{cases}$ (ok)

$(\pi_{1|2} \mid Y_{1|2} = 31) \sim Beta(32, 37) \approx \mathcal{N}(0.46, 0.0037)$ si $\begin{cases} 67(0.46) > 5 \\ 67(0.54) > 5 \end{cases}$ (ok) donc $(\delta = \pi_{1|1} - \pi_{1|2} \mid \text{données}) \approx \mathcal{N}\left(\hat{\pi}_{1|1} - \hat{\pi}_{1|2}\,,\ \frac{\hat{\pi}_{1|1}(1-\hat{\pi}_{1|1})}{n_1} + \frac{\hat{\pi}_{1|2}(1-\hat{\pi}_{1|2})}{n_2}\right)$

$\approx \mathcal{N}(-0.14, 0.0124)$

a) → Test de la plausibilité de $\delta > 0$

$$P(\delta > 0 \mid \text{données}) = P\left(\frac{\delta + 0.14}{\sqrt{0.0124}} > \frac{0 + 0.14}{\sqrt{0.0124}}\right) = P(Z > 1.28)$$

$$= 1 - P(Z \leq 1.28) = 1 - 0.8997 = 0.1003$$

donc $\delta > 0$ non plausible

→ Test de la plausibilité de $\delta < 0$

$$P(\delta < 0 \mid \text{données}) = P\left(\frac{\delta + 0.14}{\sqrt{0.0124}} < \frac{0 + 0.14}{\sqrt{0.0124}}\right) = P(Z < 1.28) = 0.8997$$

donc $\delta < 0$ non plausible

On ne peut pas conclure à une différence entre la proportion de fumeurs en S.H et en M.

b) Intervalle de crédibilité à 95% pour δ

$$\hat{\delta} \pm 1.96\ SE(\delta) = -0.14 \pm 1.96\sqrt{0.0124} = [-0.36\,;\,0.08]$$

Comme 0 est une valeur plausible pour $\pi_{1|1} - \pi_{1|2}$, on ne peut pas conclure à une différence.

<u>Méthode 2</u> : Estimation du risque relatif.

$RR = \pi_{1|1} / \pi_{1|2}$, $\widehat{RR} = 0.69$, $\ln(\widehat{RR}) = -0.37$

Intervalle de crédibilité à 95% pour $\ln(RR)$

$$\ln(\widehat{RR}) \pm 1.96\sqrt{\frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}}} = -0.37 \pm 1.96\sqrt{0.1023}$$

$= [-0.996\,;\,0.258]$ (0 est une valeur plausible)

Intervalle de crédibilité à 95% pour RR $[e^{-0.996}\,;\,e^{0.258}] = [0.37\,;\,1.29]$

(1 est une valeur plausible)

On ne peut pas conclure à une différence entre les deux groupes

Question 3

30 personnes se présentent par heure.

Soit X = nombre de personnes en 2 minutes $\sim \mathcal{P}(1)$

$$P(X=3) = e^{-1}\frac{1^3}{3!} = \frac{1}{6}e^{-1} = 0.0613$$

$$P(X \geq 3) = 1 - P(X<3) = 1 - (P(X=0) + P(X=1) + P(X=2))$$
$$= 1 - \left(e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!}\right)\right)$$
$$= 1 - e^{-1}\left(1+1+\frac{1}{2}\right) = 1 - \frac{5}{2}e^{-1} = 0.083$$

Question 4

Effet	SC	ddl	MC	F
ord. origine	2,59	1	2,59	
Age	0,18	2	0,090	$\frac{0.090}{0.00381} = 23,6$
Erreur	0,08	21	0.00381	

On teste la plausibilité de $H_0 : \mu_2 = \mu_3 = \mu_4$ où μ_i = concentration en DDT moyenne dans le groupe d'âge i ans.

$F_{obs} = 23,6$

Seuil $= F_{2,21}(0,95) \in [2,99 ; 3,10]$

Comme $23,6 > 2.99$, H_0 n'est pas plausible. La concentration moyenne en DDT varie avec l'âge des brochets

$(\mu_2 - \mu_3 \mid \text{données}) \sim t_{21}\left(0.218 - 0.337, \frac{2}{8} \cdot 0.00381\right) \sim t_{21}(-0.1190, 0.0010)$

$(\mu_2 - \mu_4 \mid \text{données}) \sim t_{21}\left(0.218 - 0.431, \frac{2}{8} \cdot 0.00381\right) \sim t_{21}(-0.2130, 0.0010)$

$(\mu_3 - \mu_4 \mid \text{données}) \sim t_{21}\left(0.337 - 0.431, \frac{2}{8} \cdot 0.00381\right) \sim t_{21}(-0.0940, 0.0010)$

donc IC 95% pour $\mu_2 - \mu_3$: $-0.1190 \pm 2.0800\sqrt{0.0010}$: $[-0.1848, -0.0532]$

$\mu_2 - \mu_4$: $-0.2130 \pm 2.0800\sqrt{0.0010}$: $[-0.2788, -0.1472]$

$\mu_3 - \mu_4$: $-0.0940 \pm 2.0800\sqrt{0.0010}$: $[-0.1598, -0.0282]$

Ainsi, $\mu_2 - \mu_3 < 0$ c'est à dire $\mu_2 < \mu_3$

$\mu_2 - \mu_4 < 0$ $\qquad \mu_2 < \mu_4$

$\mu_3 - \mu_4 < 0$ $\qquad \mu_3 < \mu_4$

Finalement, la concentration en DDT augmente, en moyenne, avec l'âge des brochets.