Kurskod: TAMS65 Provkod: TEN2

Solutions to the exam on 25 March 2020, 14:00-18:00

1 (3 points)

Solution. (1.1) By the Method of Moments, we get $E(X) = \bar{x}$. Since $E(X) = (-1) \times \frac{\theta}{8} + \ldots + 2 \times \frac{5}{8} = \frac{13-4\theta}{8}$ we have

$$\frac{13-4\theta}{8} = \bar{x}, \text{ thus } \hat{\theta}_{MM} = \frac{13-8\bar{x}}{4} = \frac{41}{20},$$

where $\bar{x} = \frac{2\cdot(-1)+3\cdot 0+\ldots+3\cdot 2}{10} = \frac{3}{5}$.
$\hat{\theta}_{MM} = \frac{41}{20} \notin (0,1)$ which means this is not a reasonable point estimate. So it is not necessary to check if the point estimate is unbiased.
(1.2) The likelihood function is

$$L(\theta) = (\frac{\theta}{8})^2 \times (\frac{\theta}{4})^3 \times (\frac{3(1-\theta)}{8})^2 \times (\frac{5}{8})^3 = \theta^5(1-\theta)^2 \frac{3^2 \cdot 5^3}{8^7 \cdot 4^3}.$$

Maximizing $L(\theta)$ is equivalent to maximize $\ln L(\theta)$ where

$$\ln L(\theta) = 5\ln(\theta) + 2\ln(1-\theta) + \ln(\frac{3^2 \cdot 5^3}{8^7 \cdot 4^3}).$$

By taking $\frac{d \ln L(\theta)}{d\theta} = 0$, we get $\frac{5}{\theta} - \frac{2}{1-\theta} = 0$. Therefore

$$\hat{\theta}_{ML} = \frac{5}{7}.$$

Since $\frac{d^2 \ln L(\theta)}{d\theta^2} = -\frac{5}{\theta^2} - \frac{2}{(1-\theta)^2} < 0$. □

2 (3 points)

Solution. (2.1) Model: $X \sim Bin(81054, p_C) : x = 3261$. and $Y \sim Bin(59183, p_I) : y = 5476$.
First $\hat{p}_C = \frac{x}{n_C} = \frac{3261}{81054} \approx 0.040$, then $n_C\hat{p}_C(1-\hat{p}_C) \approx 3130 > 10$.
And $\hat{p}_I = \frac{y}{n_I} = \frac{5476}{59183} \approx 0.093$, then $n_I\hat{p}_I(1-\hat{p}_I) \approx 4969 > 10$.
The sampling distribution is

$$\frac{(P_I - 2P_C) - (p_I - 2p_C)}{\sqrt{p_I(1-p_I)/n_I + 2^2 \cdot p_C(1-p_C)/n_C}} \approx N(0,1)$$

since $\hat{P}_C \approx N(p_C, \sqrt{p_C(1-p_C)/n_C})$, $\hat{P}_I \approx N(p_I, \sqrt{p_I(1-p_I)/n_I})$, and $\hat{P}_C$ and $\hat{P}_I$ are independent.
Therefore, one way to get test statistic as following:

$$TS = \frac{\hat{p}_I - 2\hat{p}_C - (0)}{\sqrt{\hat{p}_I(1-\hat{p}_I)/n_I + 2^2 \cdot \hat{p}_C(1-\hat{p}_C)/n_C}} \approx 7.13$$

and $C = (\lambda_{0.05}, \infty) = (1.645, \infty)$. Thus, $TS \in C$, reject H_0. That is $p_I > 2p_C$.
(2.2) 95% one sided lower bound confidence interval for $p_I - 2p_C$ is

$$\begin{aligned} I_{p_I - 2p_C} &= (a, \infty) \\ &= \left((\hat{p}_I - 2\hat{p}_C) - \lambda_{0.05} \cdot \sqrt{\hat{p}_I(1-\hat{p}_I)/n_I + 2^2 \cdot \hat{p}_C(1-\hat{p}_C)/n_C}, \infty\right) \\ &= (0.01, \infty). \end{aligned}$$

We also can get $p_I > 2p_C$. □

3 (2.5 points)

Solution. Method 1: Consider the differences of paired data: $d_i = y_i - x_i : 2.2, 1, -2.2, 0.6, -2, 2.3$ which gives $\bar{d} = \frac{1}{6}\sum_{i=1}^{6} d_i = 0.317$ and $s^2 = \frac{1}{6-1}\sum_{i=1}^{6}(d_i - \bar{d}_i)^2 \approx 3.686$
Model: Assume $D_i = Y_i - X_i \sim N(\Delta, \sigma), i = 1, \dots, 6$.
The sampling distribution is

$$\frac{\bar{D} - \Delta}{S/\sqrt{n}} \sim t(n-1)$$

Then 95% CI:

$$I_\Delta = \bar{d} \mp t_{0.025}(n-1)\frac{s}{\sqrt{n}}$$

$$= 0.317 \mp (2.57)\frac{\sqrt{3.686}}{\sqrt{6}} = (-1.70, \quad 2.33)$$

$0 \in I_\Delta$, so we can't conclude that there is a systematic difference between the two methods.
Method 2: Make a test on $H_0 : \Delta = 0$ versus $H_0 : \Delta \neq 0$, which will give the same conclusion.

□

4 (3 points)

Solution. H_0 : The age doesn't affect the choice of games.
H_1 : The age affects the choice of games.

	Games			
Group	Nintendo Switch	PS5	Xbox	Total
Group 1	$N_{11} = 30$	$N_{12} = 40$	$N_{13} = 50$	120
Group 2	$N_{21} = 35$	$N_{22} = 55$	$N_{23} = 40$	130
Total	65	95	90	n=250

$$p_{11} = p_{1-row} \cdot p_{1-column} = \frac{120}{250} \cdot \frac{65}{250}, np_{11} \approx 31.2$$
$$p_{12} = p_{1-row} \cdot p_{2-column} = \frac{120}{250} \cdot \frac{95}{250}, np_{12} \approx 45.6$$
$$p_{13} = \frac{120}{250} \cdot \frac{90}{250}, \quad np_{13} \approx 43.2$$
$$p_{21} = \frac{130}{250} \cdot \frac{65}{250}, \quad np_{21} \approx 33.8$$
$$p_{22} = \frac{130}{250} \cdot \frac{95}{250}, \quad np_{22} \approx 49.4$$
$$p_{23} = \frac{130}{250} \cdot \frac{90}{250}, \quad np_{23} \approx 46.8$$

$TS = \frac{(30-31.2)^2}{31.2} + ... + \frac{(40-46.8)^2}{46.8} = 3.47$
$C = (\chi^2_{0.05}((2-1)(3-1)), \infty) = (5.99, \infty)$
If $TS \notin C$, don't reject H_0, i.e. We can't conclude that the age affects the choice of games.

□

5 (3.5 points)

Solution. (5.1) Since population standard deviation σ is related to three samples, the confidence interval for σ^2 is

$$I_{\sigma^2} = \left(\frac{(n_1+n_2+n_3-3)s^2}{\chi^2_{\alpha/2}(n_1+n_2+n_3-3)}, \qquad \frac{(n_1+n_2+n_3-3)s^2}{\chi^2_{1-\alpha/2}(n_1+n_2+n_3-3)}\right) = \left(\frac{(13)s^2}{\chi^2_{0.025}(13)}, \qquad \frac{(13)s^2}{\chi^2_{0.975}(13)}\right)$$
$$= (\frac{(13)3.72}{24.75}, \qquad \frac{(13)3.72}{5.01}) = (1.954, \qquad 9.653).$$

Where $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+(n_3-1)s_3^2}{n_1+n_2+n_3-3} = 3.720$. Therefore the confidence interval for σ is

$$I_\sigma = (\sqrt{1.954}, \qquad \sqrt{9.653}) = (1.40, \qquad 3.11).$$

(5.2) Note that: $\mu_1 - \mu_2 = \mu_2 - \mu_3$ is equivalent to $\mu_1 - 2\mu_2 + \mu_3 = 0$. We also can get that

$$\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3 \sim N(\mu_1 - 2\mu_2 + \mu_3, \sigma\sqrt{\frac{1}{n_1} + \frac{2^2}{n_2} + \frac{1}{n_3}})$$

Thus the sampling distribution is

$$\frac{(\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3) - (\mu_1 - 2\mu_2 + \mu_3)}{\sqrt{\frac{13}{12}}S} \sim t(n_1 + n_2 + n_3 - 3)$$

Method I: We use confidence interval

$$\begin{aligned} I_{\mu_1-2\mu_2+\mu_3} &= (\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3) \mp t_{0.01}(13) \cdot \sqrt{\frac{13}{12}} \cdot s \\ &= 2.5 \mp 2.65 \cdot \sqrt{\frac{13}{12}} \cdot \sqrt{3.720} \approx (-2.82, 7.82). \end{aligned}$$

We can see that $0 \in I_{\mu_1-2\mu_2+\mu_3}$, so it is possible that $\mu_1 - 2\mu_2 + \mu_3 = 0$ with confidence 98%.
Method II: Make a test on
$H_0 : \mu_1 - \mu_2 = \mu_2 - \mu_3$ versus $H_1 : \mu_1 - \mu_2 \neq \mu_2 - \mu_3$
Then we will get the same conclusion. $\square$