Written Assignment-2 Solutions

Question 1

a) *(3 marks :0.5 marks for defining X, 1 marks for identifying parameters in the Binomial distribution, 1 marks for the method, 0.5 for the answer)*

Let X be the number of ambulances have delays out of 100

$\Rightarrow X \sim B(100, 0.63)$

$P(X = 50) = \binom{100}{50} p^{50}(1-p)^{50} = (1.008913445 \times 10^{29}) \times 0.63^{50} \times (1-0.63)^{50} = 0.0024$

b) *(4 marks:0.5 marks for checking conditions to use Normal, 1 marks for identifying parameters in the Normal distribution, 0.5 marks for the continuity correction, 1 mark for the method, 1 mark for the answer)*

$np = 100 \times 0.63 = 63 > 5$, $n(1-p) = 100 \times (1-0.63) = 37 > 5$

Since above two conditions are satisfied, we can use normal approximation.

approximately, $X \sim N(\mu, \sigma^2)$, where $\mu = np = 63$, and $\sigma^2 = np(1-p) = 23.31$

Use a continuity correction with a range of 49.5 to 50.5.

$$\begin{aligned}
P(49.5 < X < 50.5) &= P(\frac{49.5 - 63}{\sqrt{23.31}} < Z < \frac{50.5 - 63}{\sqrt{23.31}}) \\
&= P(Z < -2.589041) - P(Z < -2.796164) \\
&= 0.002226528
\end{aligned}$$

pnorm(-2.589041) - pnorm(-2.796164) = 0.002226528

yes, they are very similar.

c) *(3 marks:1.5 mark for the method, 0.5 marks for using continuity correction, 1 mark for the answer)*

Let X be the number of ambulances have delays out of 100
$\Rightarrow X \sim B(100, 0.63)$
from part b) we have, $X \sim N(\mu, \sigma^2)$ approximately with, $\mu = np = 63$, and $\sigma^2 = np(1-p) = 23.31$
Therefore approximate probability that no more than 25 of them are blocked is (use continuity correction)

$$\begin{aligned}
P(X \leq 25) &= P(X < 25.5) && \textit{use continuity correction} \\
&= P(\frac{X-\mu}{\sigma} < \frac{25.5-63}{\sqrt{23.31}}) \\
&= P(Z < -7.77) \\
&= 3.92 \times 10^{-15} \\
&\approx 0
\end{aligned}$$

Question 2

a) *(6 marks: 1 mark for hypotheses, 1 mark for the pooled variance (or pooled standard deviation), 1 marks fo for the test statistic, 1 mark for finding the critical value or p-value, 1 mark for decision (reject the null hypothesis), 1 mark for the conclusion)*

Let μ_1, μ_2 be the mean fat contents of Joe's and Frank's restaurant, respectively. The test to use here is a two-sample comparison. The hypotheses tested are

$$H_0 : \mu_1 = \mu_2, \qquad H_a : \mu_1 \neq \mu_2.$$

or

$$H_0 : \mu_1 - \mu_2 = 0, \qquad H_a : \mu_1 - \mu_2 \neq 0.$$

From the data, we have that $n_1 = n_2 = 15$, $\bar{y}_1 \approx 37.3867, \bar{y}_2 \approx 40.3667$ and $s_1^2 \approx 4.4827, s_2^2 \approx 4.3581$. The model gives us that the population variance of the two groups are equal, so we calculate the pooled standard deviation

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 4.4204.$$

The test statistic is:

$$t_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{37.3867 - 40.3667}{\sqrt{4.4204}\sqrt{\frac{1}{15} + \frac{1}{15}}} \approx -3.882.$$

The degrees of freedom is $n_1 + n_2 - 2 = 28$, so the p-value is $2 \times P(|t_{28}| > 3.882) = 2 \times 2.88 \times 10^{-4} = 5.76 \times 10^{-4}$ (this value is different if you use $t-$table). Based on a significance level of $\alpha = 0.05$, we reject the null hypothesis. Thus, there is sufficient evidence to show that the mean fat contents are different for the two restaurants.

b) *(2 marks: 1 mark for stating the formula, 0.5 mark for finding $t_{\alpha/2,n_1+n_2-2}$ value, 0.5 mark for the answer)*

We know that $\frac{\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ has a t_{28} distribution and

$$\begin{aligned}
0.99 &= P\left(-2.7633 < \frac{\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < 2.7633\right)\\
&= P\left((\bar{y}_1-\bar{y}_2) - 2.7633 \times s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{y}_1-\bar{y}_2) + 2.7633 \times s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).
\end{aligned}$$

Thus the 99% confidence interval for $\mu_1-\mu_2$ is given as

$$(\bar{y}_1-\bar{y}_2) \pm t_{\alpha/2,\, n_1+n_2-2} \times s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad \text{where}\alpha = 0.01$$

$$\begin{aligned}
&\Rightarrow (37.3867-40.3667) \pm 2.7633 \times \sqrt{4.4204}\sqrt{\frac{1}{15}+\frac{1}{15}}\\
&\Rightarrow [-5.101, -0.858].
\end{aligned}$$

c) *(2 marks: 0.5 for stating No, 1.5 marks for the reason)*

No. As -3 is contained in the confidence interval [-5.101, -0.858] in Part (b), there isn't sufficient evidence to reject the null hypothesis that the difference between the true mean fat contents of Joe's and Frank's restaurants is -3 at $\alpha = 0.01$.

Question 3

a) *(2 marks: : 2 marks for correctly defining the nul and alternative hypotheses)*

Let μ denote the mean caffeine content in Starbucks regular Breakfast Blend. The two hypotheses to be tested are:

$$H_0 : \mu \leq 350, \qquad H_a : \mu > 350.$$

b) *(5 marks: : 2 marks for the test statistic, 1 mark for finding the critical value or p-value, 1 mark for decision (not reject the null hypothesis), 1 mark for the conclusion)*

We have that $\mu_0 = 350$. The test statistic is calculated as

$$t_{obs} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \approx \frac{371.967 - 350}{126.367/\sqrt{6}} \approx 0.426.$$

The test statistic has a $t_{n-1} = t_5$ distribution. There are two ways that we can test the hypothesis:

- We can find the critical value by finding the value q that satisfies $P(t_5 > q) = 0.05$. Using R, we find $q = qt(0.05, 5, lower.tail = F) = 2.015$. Since $t_{obs} < q$, we do not have enough evidence to reject the null hypothesis.Therefore we can conclude that there is no enough evidence to conclude that the mean caffeine content in regular Starbucks Breakfast Blend is greater than 350 mg at $\alpha = 0.05$.
- We can find the p-value by calculating $P(t_5 > t_{obs}) = pt(0.426, 5, lower.tail = F) \approx 0.344$. Since the p-value is greater than 0.05, we do not have enough evidence to reject the null hypothesis. Therefore we can conclude that there is no enough evidence to conclude that the mean caffeine content in regular Starbucks Breakfast Blend is greater than 350 mg at $\alpha = 0.05$.

c) *(3 marks: : 1 mark for stating the formula, 1 mark for finding $t_{\alpha/2,n-1}$ value, 1 mark for the answer)*

We know that $\frac{\bar{y}-\mu}{s/\sqrt{n}}$ has a t_5 distribution and

$$\begin{aligned}
0.95 &= P\left(-2.5706 < \frac{\bar{y}-\mu}{s/\sqrt{n}} < 2.5706\right) \\
&= P\left(-2.5706 \times \frac{s}{\sqrt{n}} < \mu - \bar{y} < 2.5706 \times \frac{s}{\sqrt{n}}\right) \\
&= P\left(\bar{y} - 2.5706 \times \frac{s}{\sqrt{n}} < \mu < \bar{y} + 2.5706 \times \frac{s}{\sqrt{n}}\right).
\end{aligned}$$

Thus the two-sided 95% confidence interval for μ is given as

$$\bar{y} \pm t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}} \qquad \text{where } \alpha = 0.05, n = 6$$

$$\Rightarrow 371.967 \pm 2.5706 \times \frac{126.367}{\sqrt{6}}$$

$$\Rightarrow [239.352, 504.582].$$

Question 4

a) *(5 marks: 2 marks df, 2 marks for sum of squares, 1 mark for mean squares)*

Source of Variation	df	Sum of Square	Mean square
Treatment			
Error			

We calculate $\bar{y}.. = 0.8026$,

$$SS(treatment) = \sum_{i=1}^{k} n_i[\bar{y}_{i.} - \bar{y}..]^2 = 0.1712,$$

and

$$SS(error) = \sum_{i=1}^{k}\sum_{i=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \approx 1.0942.$$

Then, the ANOVA table can be filled out as below:

Source of Variation	df	Sum of Square	Mean square
Treatment	2	0.1712	0.0856
Error	27	1.0942	0.0405

b) *(5 marks: 0.5 for defining population parameters (population means), 1 mark for hypotheses, 1 mark for the test statistic (F_{obs}), 1 mark for finding the critical value or p-value, 0.5 mark for decision (do not reject the null hypothesis), 1 mark for the conclusion)*

Let μ_1, μ_2, μ_3 represent the mean daily calcium intake of group Normal bone density, Osteopenia and Osteoporosis, respectively. The hypotheses we want to tests are

$$H_0 : \mu_1 = \mu_2 = \mu_3, \qquad H_a : \mu_i \neq \mu_j, \text{for some } i, j \in \{1, 2, 3\}.$$

From the ANOVA table above, we can calculate F-ratio as

$$F_{obs} = \frac{0.0856}{0.0405} \approx 2.1136.$$

The degrees of freedoms for the F statistic are 2 and 27, so the p-value is calculated as $P(F_{2,27} > F_{obs}) \approx 0.1404$. At $\alpha = 0.05$ significance level, we do not reject H_0, so there isn't enough evidence that the average daily calcium intake is different for the three groups.

or

The critical value is calculated as $q = qf(0.05, 2, 27, lower.tail = F) = 3.3541$, which is bigger than 2.1136. At $\alpha = 0.05$ significance level, we do not reject H_0, so there isn't enough evidence that the average daily calcium intake is different for the three groups.