Student Name: **Luke Nguyen**
Student ID: **D5850A**

Statistical Methods and Data Analysis (EN.625.603)

Problem Set 8

Question 9.2.12
Suppose that $H_0 : \mu_X = \mu_Y$ is being tested against $H_1 : \mu_X \neq \mu_Y$, where $sigma_x^2$ and σ_Y^2 are known to be 17.6 and 22.9, respectively. If $n = 10$, $m = 20$, $\bar{x} = 81.6$, and $\bar{y} = 79.9$, what $P-$value would be associated with the observed Z ratio?
Solution
Applying **Theorem 9.2.3**, test statistic as follows:

$$\begin{aligned} Z &= \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \\ &= \frac{81.6 - 79.7}{\sqrt{\frac{17.6}{10} + \frac{29.9}{20}}} \\ &\approx 0.9974 \\ \text{P-value} &= 2 \times 0.1611 \\ &= 0.3222 \end{aligned}$$

Question 9.2.15
If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are independent random samples from normal distribution with the same σ^2, prove that their pooled sample variance, S_p^2, is an unbiased estimator for σ^2.
Solution
From **Theorem 9.2.1**, we have:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

Since $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are independent random samples from normal distribution with the same σ^2, we have:

$$\begin{aligned} E(S_p^2) &= \frac{\sum_{i=1}^{n-1} \sigma^2 + \sum_{i=1}^{m-1} \sigma^2}{n+m-2} \\ &= \frac{n\sigma^2 + m\sigma^2 - 2\sigma^2}{n+m-2} \\ \boldsymbol{E(S_p^2)} &= \boldsymbol{\sigma^2} \qquad \text{Q.E.D} \end{aligned}$$

Question 9.2.17

n_i	X	m_i	Y
1	640	12	10
2	80	13	320
3	1280	14	320
4	160	15	320
5	640	16	320
6	640	17	80
7	1280	18	160
8	640	19	10
9	160	20	640
10	320	21	160
11	160	22	320

$H_0 : \mu_X = \mu_Y \qquad H_1 : \mu_X \neq \mu_Y$
$\alpha = 0.05$
$S_X = 428, \quad S_Y = 183$

Solution
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 545.45$
$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i = 241.82$
From **Theorem 9.2.1**, we can find S_p as follows:

$$\begin{aligned}
S_p^2 &= \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \\
&= \frac{(11-1)428^2 + (11-1)183^2}{11+11-2} \\
&= 108336.5 \\
S_P &\approx 329.1 \\
\hat{\theta} &\approx \frac{S_X^2}{S_Y^2} = \frac{428^2}{183^2} = 5.47 \\
13.54 &\approx \frac{(5.47 + \frac{11}{11})^2}{\frac{1}{10}5.47^2 + \frac{1}{10}(\frac{11}{11})^2} \\
v &= 14
\end{aligned}$$

From **Theorem 9.2.2**, the test statistic is as follows:

$$\begin{aligned}
t &= \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \\
&= \frac{(545.45 - 241.82) - 0}{329.1\sqrt{\frac{1}{11} + \frac{1}{11}}} \\
&= 2.164 \\
t &> t_{\alpha/2,df} \\
&> t_{0.05,14} \\
\boldsymbol{t} &> \mathbf{1.7613} \qquad \textbf{Reject } \boldsymbol{H_0}
\end{aligned}$$

Question 9.2.18

For the approximate two-sample t test described in Question 9.2.17, it will be true that

$$v < n + m - 2$$

Why is that a disadvantage for the approximate test? That is, why is it better to use the Theorem 9.2.1 version of the t test, in fact, $\sigma_X^2 = \sigma_Y^2$?

Solution

Because the sample of collected subjects belongs to a normal distribution population.

Question 9.3.3

Solution

(a) We have,

$$\begin{aligned}
n &= 20, \qquad \bar{x} = 3.55 \\
m &= 20, \qquad \bar{y} = 2.10 \\
\sum_1^{20} x_i^2 &= 319.0, \qquad S_X^2 = 1.877 \\
\sum_1^{20} y_i^2 &= 134.0, \qquad S_Y^2 = 1.553 \\
H_0 : \sigma_X^2 &= \sigma_Y^2 \qquad H_1 : \sigma_X^2 \neq \sigma_Y^2 \\
\alpha &= 0.05
\end{aligned}$$

Applying **Theorem 9.3.1**, we have:

$$\begin{aligned}
\frac{S_Y^2}{S_X^2} &= \frac{1.553^2}{1.877^2} = 0.684 \\
&> F_{\alpha/2,m-1,n-1} > F_{0.025,19,19} = 0.406 \\
\boldsymbol{\frac{S_Y^2}{S_X^2}} &> \boldsymbol{F_{\alpha/2,m-1,n-1}} \qquad \textbf{Accept } \boldsymbol{H_0}
\end{aligned}$$

(b)

$$\begin{aligned}
S_p^2 &= \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \\
&= \frac{(20-1)1.877^2 + (20-1)1.553^2}{20+20-2} \\
&= 2.94 \\
S_p &\approx 1.714 \\
t &= \frac{\bar{x} - \bar{y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \\
&= \frac{3.55 - 2.10}{1.714\sqrt{\frac{1}{20} + \frac{1}{20}}} \\
&= 2.675 \\
t &> t_{\alpha/2,38} = 2.0244 \\
\boldsymbol{t} &> \boldsymbol{t_{\alpha/2,df}} \qquad \textbf{Reject } \boldsymbol{H_0}
\end{aligned}$$

Question 9.5.7

Solution We have,

$$
\begin{aligned}
n &= 8, \qquad \bar{x} = 63.25 \\
m &= 6, \qquad \bar{y} = 44.67 \\
\sum_{1}^{8} x_i^2 &= 32966.0, \qquad S_X^2 = 137.367 \\
\sum_{1}^{6} y_i^2 &= 13672.0, \qquad S_Y^2 = 340.277 \\
H_0 : \mu_X = \mu_Y & \qquad H_1 : \mu_X \neq \mu_Y \\
\alpha &= 0.05
\end{aligned}
$$

Applying **Theorem 9.2.2**, we have:

$$
\begin{aligned}
S_p &= \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2 + \sum_1^m (y_i - \bar{y})^2}{n + m - 2}} \\
&= \sqrt{\frac{32966.0 + 13672.0}{8 + 6 - 2}} \\
&\approx 14.897 \\
t &= \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \\
&= \frac{63.25 - 44.67}{14.897 \sqrt{\frac{1}{8} + \frac{1}{6}}} \\
&= 2.3099 \\
t &> t_{\alpha/2,df} = t_{0.025,12} = 2.1788 \\
\boldsymbol{t} &\boldsymbol{> t_{\alpha/2,df}} \qquad \textbf{Reject } \boldsymbol{H_0}
\end{aligned}
$$

Applying **Theorem 9.5.2**, we have:

$$
\begin{aligned}
&= \left(\frac{S_X^2}{S_Y^2} F_{\alpha/2,n-1,m-1}, \frac{S_X^2}{S_Y^2} F_{1-\alpha/2,n-1,m-1} \right) \\
&= \left(\frac{137.367}{340.277} F_{0.025,7,5}, \frac{137.367}{340.277} F_{0.975,7,5} \right) \\
&= \left(\frac{137.367}{340.277} 0.146, \frac{137.367}{340.277} 0.529 \right) \\
&= (0.0589, 0.2136)
\end{aligned}
$$

It is correct to use 9.2.2 because the ratio $\frac{\sigma_X^2}{\sigma_Y^2}$ is within the range of the confidence interval.