Q2

Huiyan Li

4/18/2020

Assignment 2: Question 2

- For both part of the question, given we don't know the exact vaule for σ, therefore, we will replace $Z_{\alpha/2}$ with $T_{\alpha/2,df}$ to approximate

```
# Only change this following line, remove 261 and put your student id
student_id=1001145664
# do not change anything below
set.seed(student_id)
Group1= round(rnorm(15,mean=10,sd=4),2)
Group2= round(rnorm(12,mean=7,sd=4),2)
```

```
group1Size = length(Group1)
group2Size = length(Group2)

group1Mean = mean(Group1)
group2Mean = mean(Group2)

group1Var = var(Group1)
group2Var = var(Group2)

group1SD = sqrt(group1Var)
group2SD = sqrt(group2Var)
```

A) Assuming $\sigma_1^2 = \sigma_2^2$

i) Construct a 95% confidence interval for $\mu_1 - \mu_2$ and interpret

- Given $\sigma_1^2 = \sigma_2^2$, therefore, $df = n + m - 2 = 15 + 12 - 2 = 25$
- Below is the Lower and Upper Confidence Limits equation for equal-variance:
- $P(\overline{X} - \overline{Y} - Z_{\alpha/2}\sqrt{\frac{(m-1)\sigma_1^2+(n-1)\sigma_2^2}{m+n-2}(\frac{1}{m} + \frac{1}{n})} < \mu_1 - \mu_2 < \overline{Y} + Z_{\alpha/2}\sqrt{\frac{(m-1)\sigma_1^2+(n-1)\sigma_2^2}{m+n-2}}(\frac{1}{m} + \frac{1}{n}))$
- $\mu_1 - \mu_2 = \overline{X} - \overline{Y} \pm t_{1-\alpha/2,df}\sqrt{\frac{(m-1)\sigma_1^2+(n-1)\sigma_2^2}{m+n-2}(\frac{1}{m} + \frac{1}{n})}$
- $\mu_1 - \mu_2 = 10.26933 - 4.41 \pm -2.059539\sqrt{\frac{(15-1)11.46239+(12-1)29.94576}{15+12-2}(\frac{1}{15} + \frac{1}{12})}$
- $\mu_1 - \mu_2 = (2.328404, 9.390263)$

```
alpha = 0.05 / 2
dfSize = group1Size + group2Size - 2

lower = group1Mean - group2Mean + qt(alpha,df=dfSize) * sqrt(((((group1Size - 1) * group1Var) +
 ((group2Size - 1) * group2Var)) /(group1Size + group2Size - 2)) * ((1/group1Size) + (1/group2Si
ze)))
upper = group1Mean - group2Mean - qt(alpha,df=dfSize) * sqrt(((((group1Size - 1) * group1Var) +
 ((group2Size - 1) * group2Var)) /(group1Size + group2Size - 2)) * ((1/group1Size) + (1/group2Si
ze)))

#t.test(Group1, Group2, var.equal=TRUE) # t.test by default assume the variances are not equal

lower
```

```
## [1] 2.328404
```

```
upper
```

```
## [1] 9.390263
```

- From this, we can see that $2.328404 < \mu_1 - \mu_2 < 9.390263$
- Therefore, we are highly confident that the true population from group 1 exceeds the true population from group2 by between 2.328404 - 9.390263.

ii) At 5% level of significance, test $H_0 : \mu_1 = \mu_2$ vs $H_\alpha : \mu_1 \neq \mu_2$

- We will reject the Null hypothesis $H_0 : \mu_1 = \mu_2$ if and only if $|T| > t_{1-\alpha/2,df}$, where the T-test statistics value exceeds the critical value
- $CriticalValue = t_{1-\alpha/2,df} = t_{0.975,25} = qt(0.975, df = 25) = 2.059539$
- $\overline{X} - \overline{Y} = t_{1-\alpha/2.df}\sqrt{\frac{(m-1)\sigma_1^2+(n-1)\sigma_2^2}{m+n-2}(\frac{1}{m} + \frac{1}{n})}$
- $T = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{(m-1)\sigma_1^2+(n-1)\sigma_2^2}{m+n-2}(\frac{1}{m}+\frac{1}{n})}} = \frac{10.26933-4.41}{\sqrt{\frac{(15-1)11.46239+(12-1)29.94576}{15+12-2}(\frac{1}{15}+\frac{1}{12})}} = 3.41766025706$

```
t = (group1Mean - group2Mean) / sqrt(((((group1Size - 1) * group1Var) + ((group2Size - 1) * grou
p2Var))/ (group1Size + group2Size - 2)) * (1/group1Size + 1/group2Size))
t
```

```
## [1] 3.417662
```

- $|T| > t_{1-\alpha/2,df} \implies = |3.417662| > 2.059539 \implies True$
- Therefore, we reject the null hypothesis and conclude that the two population means are different at the 0.05 significance level.

B) Assuming $\sigma_1^2 \neq \sigma_2^2$

i) Construct a 90% confidence interval for $\mu_1 - \mu_2$ and interpret

- Given $\sigma_1^2 \neq \sigma_2^2$, therefore, we can assume df as follow:
- $df = \frac{(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})^2}{\frac{(\frac{\sigma_1^2}{m})^2}{m-1}+\frac{(\frac{\sigma_2^2}{n})^2}{n-1}} = \frac{(\frac{11.46239}{15}+\frac{29.94576}{13})^2}{\frac{(\frac{11.46239}{15})^2}{15-1}+\frac{(\frac{29.94576}{12})^2}{15-1}} = 17.48036$
- Below is the Lower and Upper Confidence Limits equation for unequal-variance:
- $P(\overline{X} - \overline{Y} - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \overline{Y} + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}})$
- $\mu_1 - \mu_2 = \overline{X} - \overline{Y} \pm t_{1-\alpha/2,df}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$
- $\mu_1 - \mu_2 = 10.26933 - 4.41 \pm 1.736861\sqrt{\frac{11.46239}{15} + \frac{29.94576}{12}}$
- $\mu_1 - \mu_2 = (2.723523, 8.995144)$

```
alpha = 0.10 / 2

numerator = ((group1Var/group1Size) + (group2Var/group2Size))**2
denominator = ((group1Var/group1Size)**2)/(group1Size - 1) + ((group2Var/group2Size)**2)/ (group
2Size - 1)
dfSize = numerator / denominator

lower = group1Mean - group2Mean + qt(alpha,df=dfSize) * sqrt(group1Var/group1Size + group2Var/gr
oup2Size)
upper = group1Mean - group2Mean - qt(alpha,df=dfSize) * sqrt(group1Var/group1Size + group2Var/gr
oup2Size)

#t.test(Group1, Group2, conf.level = 0.90)

lower
```

```
## [1] 2.723523
```

```
upper
```

```
## [1] 8.995144
```

- From this, we can see that $2.723523 < \mu_1 - \mu_2 < 8.995144$
- Therefore, we are highly confident that the true population from group 1 exceeds the true population from group2 by between 2.723523 - 8.995144

ii) At 10% level of significance, test $H_0 : \mu_1 = \mu_2$ vs $H_\alpha : \mu_1 \neq \mu_2$

- We will reject the Null hypothesis $H_0 : \mu_1 = \mu_2$ if and only if $T > t_{1-\alpha,df}$, where the T-test statistics value exceeds the critical value

- $CriticalValue = t_{1-\alpha/2,df} = t_{0.90,17.48036} = qt(0.90, df = 17.48036) = 1.3319$
- $\overline{X} - \overline{Y} = t_{1-\alpha/2,df}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$
- $T = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} = \frac{10.26933-4.41}{\sqrt{\frac{11.46239}{15}+\frac{29.94576}{12}}} = \frac{5.859333}{1.805447} = 3.245364$

```
t = (group1Mean - group2Mean) / sqrt(group1Var/group1Size + group2Var/group2Size)
t
```

```
## [1] 3.245364
```

- $T > t_{1-\alpha,df} \implies = 3.245364 > 1.3319 \implies True$
- Therefore, we reject the null hypothesis and conclude that the two population means are different at the 0.10 significance level.