HW10

Darnell Chen

2024-12-03

10.1.2

a) Since we want the mean of the two machines to be the same, we thus have $\mu_1 - \mu_2 = 0$. Our two test statistics $\bar{x}_1$ and $\bar{x}_2$ come out to 16.015 and 16.005 respectively. Using these two test statistics, we can find the p-value for our statistical inference:

$$\begin{aligned} z_0 &= \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{16.015 - 16.005}{\sqrt{\frac{0.02^2}{10} + \frac{0.025^2}{10}}} \\ &\approx 0.9877296 \end{aligned}$$

We have a two-sided interval since H_1 is $\mu \neq \mu_0$. Since $P(Z > 0.9877296) \approx 0.1616425$, this means our p-value is 0.3232851, since we have the consider the lower-tail rejection area too.

Since our p-value is greater than our α of 0.05, we will thus fail to reject our null hypothesis, where our $\mu_0 = \mu$ - and the engineer is thus correct.

b) Our 95% CI will be double sided, thus we're looking for:

$$\begin{aligned} (\bar{x}_1 - \bar{x_2}) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= (16.015 - 16.005) + z_{0.025} \sqrt{\frac{0.02^2}{10} + \frac{0.025^2}{10}} \\ &\approx 0.01 \pm 1.96(0.0101242) \\ &\approx [-0.009843,\ 0.029843] \end{aligned}$$

Since 0 is within our CI, we thus fail to reject the null hypothesis, since our interval contains the chance that our mean difference between our two machines is 0 (they have the same net volume).

c) Again, we're looking at a two-sided interval, however, this time we're looking for the rejection region when the true mean is 0.04, since $power = 1 - \beta$, and the acceptance region here will give us β:

$$\begin{aligned} z &= \frac{\bar{x}_1 - \bar{x}_2 - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{16.015 - 16.005 - 0.04}{\sqrt{\frac{0.02^2}{10} + \frac{0.025^2}{10}}} \\ &\approx \frac{0.01 - 0.04}{0.0101242} \\ &\approx -2.963189 \end{aligned}$$

Our rejection region will be the area less than our z-value and greater than $|z|$, $power = 2{\cdot}P(Z < -2.963189) \approx 0.003044698$

```r
machine1 <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)
machine2 <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)

mean(machine1)
```

```
## [1] 16.015
```

```r
mean(machine2)
```

```
## [1] 16.005
```

```r
2*(1-pnorm(0.9877296))
```

```
## [1] 0.3232851
```

```r
2*pnorm(-2.96318878995)
```

```
## [1] 0.003044698
```

10.2.2

a) This is a small sample size with equal variance assumption. So we have:

$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(15-1)(4)+(15-1)(6.25)}{15+15-2} = 5.125$

This means our estimated standard deviation is $S_p = \sqrt{5.125} \approx 2.263846$. Using this, we can get our t_0 for our p-value:

$$\begin{aligned} t_0 &= \frac{\bar{x}_1 - \bar{x}_2 - \Delta}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \\ &= \frac{4.7 - 7.8}{\sqrt{5.125}\sqrt{\frac{1}{15}+\frac{1}{15}}} \\ &\approx -3.7501219 \end{aligned}$$

Note that we set delta to 0 since our null hypothesis is $\mu_1 = \mu_2$, so their difference is 0.

Using our t-table, we find that $t_{0.025,28} = -2.048$. Since our test statistic less less than our critical value, we will thus reject our null hypothesis.

Since this is a two-sided hypothesis, our p-value will be $2 \cdot P(t_{28} < -3.7501219) \approx 0.001$.

Since our p-value is much less than our $\alpha = 0.05$, we will thus reject our null-hypothesis.

b) To find our CI, we will use the following formula:

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}=0.025} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} &= (4.7-7.8) \pm 2.048\sqrt{5.125}\sqrt{\frac{1}{15}+\frac{1}{15}} \\ &\approx -3.02 \pm 1.692958 \\ &\approx [-4.792958, -1.4070417] \end{aligned}$$

Since 0 is not in our CI, we will thus reject the null hypothesis.

10.2.6

a) We have a small sample size with unequal variance, thus our degrees of freedom for our critical value will be:

$$v = \frac{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1-1}+\frac{(\frac{s_2^2}{n_2})^2}{n_2-1}} = \frac{(\frac{12^2}{10}+\frac{22^2}{16})^2}{\frac{(\frac{12^2}{10})^2}{10-1}+\frac{(\frac{22^2}{16})^2}{16-1}} \approx 23.72112877 \approx 24$$

Thus our critical value is $-t_{0.025,24} = -1.711$. We then want the test statistic to compare it to:

$$\begin{aligned} t_0 &= \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{290 - 321}{\sqrt{\frac{12^2}{10} + \frac{22^2}{16}}} \\ &\approx -4.639284 \end{aligned}$$

Since our test statistic is less than our critical value, we will thus reject our null hypothesis.

To prove with our p-value, we know that $P(t_{24} < -4.629284) \approx 0$. Since our p-value is much less than our alpha $\alpha = 0.05$, we thus reject our null hypothesis.

b) We're looking at:

$$\begin{aligned} t_0 &= \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{290 - 321 + 25}{\sqrt{\frac{12^2}{10} + \frac{22^2}{16}}} \\ &\approx -0.897926 \end{aligned}$$

From earlier, we calculated that our critical value is -1.711. Since our observed statistic larger than our critical value, that means it's in the acceptance region. Hence, we fail to reject the null hypothesis here.

c) To find the CI, we are computing:

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm t_{24}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= (290 - 321) \pm 1.711\sqrt{\frac{12^2}{10} + \frac{22^2}{16}} \\ &\approx -31 \pm 11.433014 \\ &\approx [-42.433014, -19.5669858] \end{aligned}$$

Our CI shows that supplier 2 has a higher average impact strength, ranging from 19.5669 to 42.433 foot-pounds higher.

10.4.2

a) Given that the difference in coding time roughly follows our qq-distribution line, we can reasonably assume that the difference in coding time in normally distributed. (WiFi is out so I can't do anderson darling test)

b) To find our standard deviation s_D, we will calculate:

$s_D = \sum_{i=0}^{12} \frac{(x_i - \bar{x})^2}{n-1} \approx 2.9644$

The average of the differences is $\bar{d} = \frac{2}{3}$. Then for our CI, we calculate:

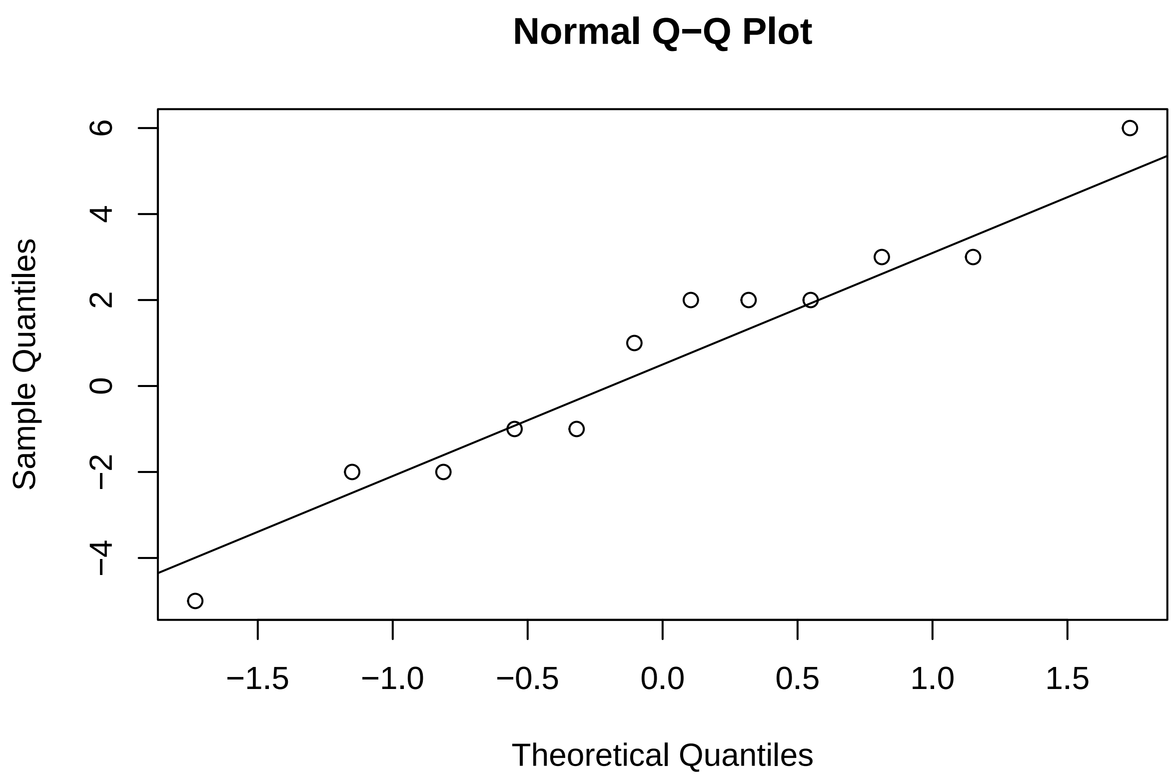
$$\begin{aligned} \bar{d} \pm t_{\frac{\alpha}{2},\ n-1} \cdot \frac{s_D}{\sqrt{n}} &= \frac{2}{3} \pm 2.201 \cdot \frac{2.9644}{\sqrt{12}} \\ &\approx \frac{2}{3} \pm 1.883525 \\ &\approx [-1.21685, 2.55019] \end{aligned}$$

Since the CI leans more towards the positive side, this means that design language 1 is more preferable.

```r
programmer1 <- c(17, 16, 21, 14, 18, 24, 16, 14, 21, 23, 13, 18)
programmer2 <- c(18, 14, 19, 11, 23, 21, 10, 13, 19, 24, 15, 20)

differences = programmer1 - programmer2

qqnorm(differences)
qqline(differences)
```



```r
sd(differences)
```

```
## [1] 2.964436
```

```r
mean(differences)
```

```
## [1] 0.6666667
```

```r
t.test(programmer1, programmer2, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  programmer1 and programmer2
## t = 0.77904, df = 11, p-value = 0.4524
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.216846  2.550179
## sample estimates:
## mean difference
##       0.6666667
```