Math 343 - Homework 1

Preston Duffield
Western Washington University
April 7, 2023

Question 1

a)

I would choose to test if the two population means where equal, ie,

H_0: $\mu_1 = \mu_2$
H_a: $\mu_1 \neq \mu_2$

b)

	Location 1	Location 2
Sample Mean	$\bar{y}_1 = 2980$	$\bar{y}_2 = 3205$
Sample Standard Deviation	$S_1 = 1140$	$S_2 = 963$
Sample Size	$n_1 = 40$	$n_2 = 40$
Sample Variance	$S_1^2 = 1299600$	$S_2^2 = 927369$

At significance level $\alpha = 0.10$. With degrees of freedom $v = (n_1 + n_2) - 2 = 40 + 40 - 2 = 78$.

$$
\begin{aligned}
S_p &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\
&= \frac{(40 - 1) \times 129960 + (40 - 1) \times 927369}{40 + 40 - 2} \\
&= 528664.5
\end{aligned}
$$

$$
\begin{aligned}
t &= \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\
&= \frac{2980 - 3205}{528664.5\sqrt{\frac{1}{40} + \frac{1}{40}}} \\
&\approx -0.0019
\end{aligned}
$$

The Test Statistic $t_{\alpha/2,v} = t_{.05,78} \approx 1.671$. Note that $t < t_{.05,78}$ and is not in the critical region. Therefore, there is not enough statistical evidence to support the population means being unequal, that is, $\mu_1 \neq \mu_2$.

c)

The P-value can be obtained in R using the following code.

Listing 1: Calculating the P-value for a t value

```
# define the t-statistic
t <- -0.0019

# calculate the P-value for a two-tailed t-test
p_value <- 2 * pt(t, df = 78, lower.tail = FALSE)

# print the P-value
p_value
```

The following code shows that the P-value = 1.001511.

d)

A 90% confidence interval for the difference in the true mean distance between the two populations is: interval for the difference in the true mean distance between

$$(y_1 - \bar{y}2) \pm t\frac{\alpha}{2}, 78 * S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\pm 1.671 * 528664.5 * \sqrt{\frac{1}{40} + \frac{1}{40}}$$

$$\pm 197533.8$$

$$(-197785.8, 197308.8)$$

This confidence interval contains zero, which is consistent with conclusion reported above.

Question 2

First we note that,

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0, 1)$$

Then we can derive that,

$$1 - \alpha = P(-t_{\alpha,n-1} \leq \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq t_{\alpha,n-1})$$

$$= P(-t_{\alpha,n-1}\frac{S}{\sqrt{n}} \leq \bar{X} - \mu_0 \leq t_{\alpha,n-1}\frac{S}{\sqrt{n}})$$

$$= P(-\bar{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}} \leq -\mu_0 \leq -\bar{X} + t_{\alpha,n-1}\frac{S}{\sqrt{n}})$$

$$= P(\bar{X} + t_{\alpha,n-1}\frac{S}{\sqrt{n}} \geq \mu_0 \geq \bar{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}})$$

$$= P(\bar{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + t_{\alpha,n-1}\frac{S}{\sqrt{n}})$$

Therefore, the confidence interval for one population mean μ in the case where the population variance σ^2 is unknown can be described as

$$\bar{X} \pm t_{\alpha,n-1}\frac{S}{\sqrt{n}}$$

Question 3

The P-value can be obtained in R using the following code, where t_0 varies.

Listing 2: Calculating the P-value for a t_0 value

```
# Define the t_0 value, degrees of freedom, and tail of the distribution
t_0 <- 2.48
df <- 10
tail <- 2

# Calculate the P-value using the "pt" function
p_val <- pt(t_0, df, tail)

# Print the P-value
print(p_val)
```

a)

When $t_0 = 2.48$, the P-value is 0.637.

b)

When $t_0 = 3.55$, the P-value is 0.869.

c)

When $t_0 = 2.00$, the P-value is 0.478.

Question 4

a)

Test

Null hypothesis H_0: $\mu = 225$
Alternative hypothesis H_1: $\mu > 225$

T-Value	P-Value
0.67	0.257

Figure 1: The output of the 1-sample t test from Minitab.

H_0: The mean repair time is 225 hours.
H_a: The mean repair time exceeds 225 hours.

b/c)

Since the p-value (0.257) is greater than the significance level (0.05), we fail to reject the null hypothesis. That is, there is not enough statistical evidence to support that the mean repair time exceeds 225 hours.

d)

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
16	241.5	98.7	24.7	198.2

μ: population mean of Hours

Figure 2: The Descriptive Statistics of the 1-sample t test from Minitab.

Question 5

a)

b)

c)

d)

e)

Question 6

a)

Listing 3: Calculating the P-value for a t_0 value

```
> t.test(x_1, y = x_2, var.equal = F, conf.level = 0.95,
+        alternative = "greater")

        Welch Two Sample t-test

data:   x_1 and x_2
t = -5.5372, df = 38.097, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -4.037613          Inf
sample estimates:
mean of x mean of y
 3.285714   6.380952
```

There is enough statistical evidence to support that the true difference in means
is greater than 0.

b)

P-value = 0.9999988.

c)

This confidence interval suggests that the true value of the difference in means
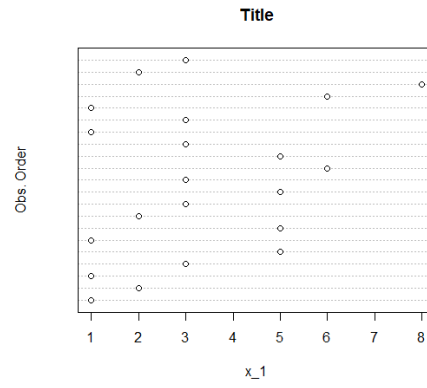is likely to be greater than -4.037613.

d)

Title

Obs. Order

x_1

Figure 3: Dot plot of 10 second cool-down time.
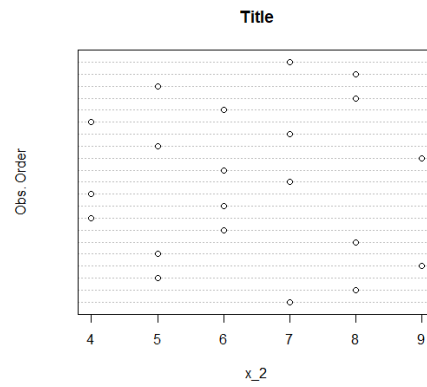
Title

Obs. Order

x_2

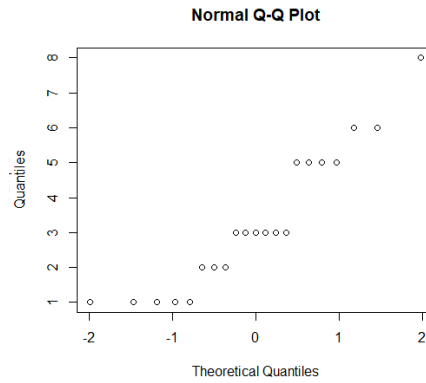Figure 4: Dot plot of 20 second cool-down time.

6

e)

Figure 5: Normal probability plot for 10 second cool-down time.

Listing 4: Shapiro-Wilk normality test for 10 second cool-down time.
```
> shapiro.test(x_1)

        Shapiro-Wilk normality test

data:   x_1
W = 0.89888, p-value = 0.03329
```
The 10 second cool-down time has a p-value of 0.03329, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that the sample data is not normally distributed.
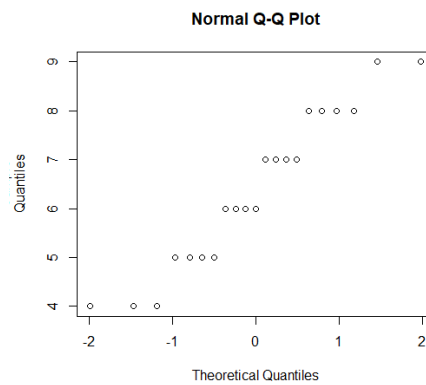

Figure 6: Normal probability plot for 20 second cool-down time.

Listing 5: Shapiro-Wilk normality test for 20 second cool-down time.

```
        Shapiro−Wilk  normality  test

data :    x_2
W =  0.93279 ,  p−value  =  0.1565
```

The 20 second cool-down time has a p-value of 0.1565, which is greater than the significance level of 0.05. Therefore, we reject the alternative hypothesis and conclude that the sample data is normally distributed.

Question 7

We can estimate the sample mean by observing the values at the 50th percentile. This gives us: $\widetilde{\mu} = 98.6$

We can estimate the sample standard deviation by taking the reciprocal of the slope of the best-fit line. This can be approximated as: $\widetilde{\sigma} = 1/0.27 = 3.7$