STAT 523 HW1

Md Muhtasim Billah

1/23/2020

Problem 1.

To obtain information on the corrosion-resistance properties of a certain type of steel conduit, 35 specimens are buried in soil for a 2-year period. The penetration (in mils) for each specimen is then measured, yielding a sample mean penetration of $\bar{x} = 35.7$ and a sample standard deviation of s = 4.2.
A. Suppose the true population standard deviation is $\sigma = 5$. Construct a 90% confidence interval for the true average penetration for this type of steel conduit. Interpret the interval.
B. Now suppose the true standard deviation is unknown. Construct a 90% confidence interval for the true average penetration for this type pf steel conduit. Interpret the interval.

Answer:

A.

Given, Sample size, n = 35 Sample mean, $\bar{x} = 35.7$ Sample standard deviation, s = 4.2 True population standard deviation, $\sigma = 5$ Confidence level = 90%

Find, The confidence interval for the true average penetration of the steel conduit.

Now,

$$\alpha = 1 - 90\% = 1 - 0.9 = 0.1$$

$$\alpha/2 = 0.1/2 = 0.05$$

Assuming a normal distribution, since the true population standard deviation is known, we will use the z-statistic.

From class notes,

$$z_{0.05} = 1.645$$

Using R,

```
qnorm(1-0.05)
```

```
## [1] 1.644854
```

Now, we calculate the cofidence interval (when the population standard deviation is known) with the following formula.

$$\left[\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right]$$

Plugging in the values,

$$\left[35.7 - z_{0.05} \times \frac{5}{\sqrt{35}}, 35.7 + z_{0.05} \times \frac{5}{\sqrt{35}}\right]$$

$$\left[35.7 - 1.645 \times \frac{5}{\sqrt{35}}, 35.7 + 1.645 \times \frac{5}{\sqrt{35}}\right]$$

Calculating the value in R,

```r
35.7-1.645*(5/sqrt(35))
```

```
## [1] 34.30972
```

```r
35.7+1.645*(5/sqrt(35))
```

```
## [1] 37.09028
```

So, the confidence interval (CI) for the true average penetration for this type of steel conduit (with 90% confidence) is,

$$[34.31, 37.09]$$

Interpretaion:

It can be said with 90% cofidence that the true mean of the population will be within this interval.

B.

Assuming a normal distribution, since the true population standard deviation is unknown, we will use the t-statistic.

From Table A5 of formula packet page 22 (FP-22),

$$t_{\alpha/2,n-1} = t_{0.05,35-1} = 1.691$$

Using R,

```r
qt(1-0.05,35-1)
```

```
## [1] 1.690924
```

Now, we calculate the cofidence interval (when the population standard deviation is unknown) with the following formula.

$$\left[\bar{x} - t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}}\right]$$

Plugging in the values,

$$\left[35.7 - t_{0.05,35-1} \times \frac{4.2}{\sqrt{35}}, 35.7 + t_{0.05,35-1} \times \frac{4.2}{\sqrt{35}}\right]$$

$$\left[35.7 - 1.691 \times \frac{4.2}{\sqrt{35}}, 35.7 + 1.691 \times \frac{4.2}{\sqrt{35}}\right]$$

Calculating the value in R,

```r
35.7-1.691*(4.2/sqrt(35))
```

```
## [1] 34.49951
```

```r
35.7+1.691*(4.2/sqrt(35))
```

```
## [1] 36.90049
```

So, the confidence interval (CI) for the true average penetration for this type of steel conduit (with 90% confidence) is,

$$[36.30, 38.70]$$

Interpretaion:

It can be said with 90% cofidence that the true mean of the population will be within this interval.

Problem 2.

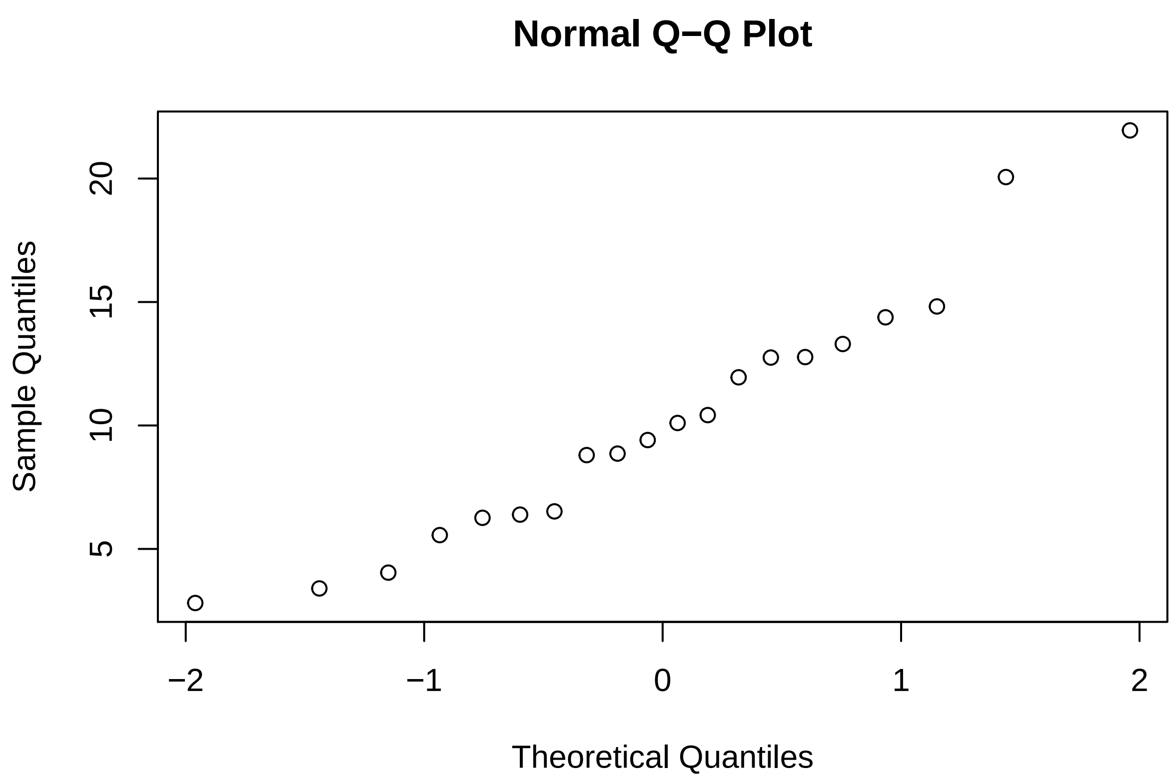
The recommended daily dietary allowance for zinc among males older than age 50 years is 15 mg/day. A study reports the following summary data on intake for a sample of males age 65–74 years: n = 20, $\bar{x}$ = 10.23, and s = 5.17. The scientist wants to know if this data indicates that average daily zinc intake in the population falls below the recommended allowance.
A. The QQ plot is provided on the right and Shapiro-Wilk normality test is provided. Use this to assess if it is plausible to assume that the daily zinc intake is normally distributed. Briefly explain your answer.
B. Carry out the test of $H_0 : \mu = 15$ vs $H_a : \mu < 15$ at $\alpha = 0.05$. Make sure you calculate the test statistic, define the rejection region, and make a decision about the test.

Answer:

A.

The normal probability plot or QQ plot generated in R using the provided information idicates that there exist linearity among the theroretical and sample quantiles. This indicates that the data come from a normal distribution.
Another way to see this is the outcome of the Shapiro-Wilk normality test performed on this dataset. For this test, the null hypothesis was that the data come from a normal distribution. The p-value is way higher than the significance level $\alpha = 0.05$ (assumed) which indicates that the null cannot be rejected. Which means the data come from a normal distribution.

```
x=c(9.41, 20.06, 8.80, 10.10, 14.38, 10.42, 13.30, 4.04, 3.40, 11.95, 12.77, 6.39, 21.95, 14.82, 12.75,
qqnorm(x)
```



```r
shapiro.test(x) # Test of H0: {Data come from a normal distribution.}
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.95112, p-value = 0.3844
```

B.

Step 1: Stating the null and alternative hypothesis

Null hypothesis, $H_0 : \mu = 15$ Alternate hypothesis, $H_a : \mu < 15$

Level of significance, $\alpha = 0.05$

Step 2: Calculate the test statistic

Since the true population standard deviation, σ is unknown, we use the t-statistic.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$t = \frac{10.23 - 15}{5.17/\sqrt{20}}$$

Calculating value using R,

```r
(10.23-15)/(5.17/sqrt(20))
```

```
## [1] -4.126129
```

So, the t-statistic is,

$$t = -4.126$$

Step 3: Find the rejection region (RR) or p-value

Rejection region:

$$RR : t \leq -t_{\alpha,n-1}$$

Now, from FP-22, Table A5,

$$t_{0.05,19} = 1.729$$

Using R,

```r
qt(1-0.05,19)
```

```
## [1] 1.729133
```

Thus, it is evident that,

$$-4.126 \leq -1.729$$

$$t \leq -t_{\alpha,n-1}$$

So, out t-statistic is within the RR.

p-value:

$$P(T_{n-1} \geq |t|)$$

Using R,

```
pt(-4.126,19)
```

```
## [1] 0.0002873021
```

It is clearly evident that,

$$0.0002873021 < 0.05$$

$$p - value < \alpha$$

So, the p-value is much lesser than α.

Step 4: Make a decision

Since, the t-statistic is within the RR and the p-value is smaller than α, it is safe to say the the null hypothesis is rejected and alternate hypothesis is accepted.

The results of the hypotheis testing can also be validated using R.

```
t.test(x,mu=15,alternative="less")
```

```
##
##  One Sample t-test
##
## data:  x
## t = -4.1268, df = 19, p-value = 0.0002868
## alternative hypothesis: true mean is less than 15
## 95 percent confidence interval:
##      -Inf 12.22717
## sample estimates:
## mean of x
##   10.2275
```

Since, same results are obtained from R, it can be safely said that the test was correctly done.

Problem 3

A pollution-control inspector suspected that a riverside community was releasing semi-treated sewage into a river and this, as a consequence, was changing the level of dissolved oxygen of the river. To check this, he drew 45 randomly selected specimens of river water at a location above the town and another 45 specimens below. The sample information for the measured oxygen level by groups are given blow.

	Sample mean	sd	n
Above	4.83	0.175	45
Below	4.55	0.234	45

A. Construct a 90% two-sided confidence interval for the difference between the average dissolved oxygen levels above town and below town. Does the data provide evidence to indicate a difference in the true average dissolved oxygen between locations above and below town?
B. The scientist wants to know if the average oxygen level below town is higher than above town. Run a hypothesis test at significance level $\alpha = 0.05$. Use the 4-step procedure: 1: State the null and alternative hypotheses. 2: Give the test statistic, 3: find the rejection region or the p-value, and 4: make a decision.

Answer:

A.

Given,
For above town, $\bar{x} = 4.83$, $s_1 = 0.175$, m = 45
For below town, $\bar{y} = 4.55$, $s_2 = 0.234$, n = 45

Here, the sample size for two populations is > 30. So, it falls under the Case II and thus we will use the z-statistic.

For 2-sided CI, the formula is as below.

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}}$$

Since, σ_1 and σ_2 are unknown, we will replace them with s_1 and s_2 respectively.

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{{s_1}^2}{m} + \frac{{s_2}^2}{n}}$$

Plugging in the values,

$$(4.83 - 4.55) \pm z_{0.05}\sqrt{\frac{0.175^2}{45} + \frac{0.234^2}{45}}$$

Calculating using R, where $z_{0.05} = 1.645$

```
(4.83-4.55)+(1.645*sqrt(((0.175)^2/45)+((0.234)^2/45)))
```

```
## [1] 0.351654
```

```
(4.83-4.55)-(1.645*sqrt(((0.175)^2/45)+((0.234)^2/45)))
```

```
## [1] 0.208346
```

So, the two sided confidence interval is,

$$[0.35, 0.21]$$

Since, 0 is not within the CI, we can say that the data provides evidence to indicate a difference in the true average dissolved oxygen between locations above and below town.

B.

Given,
Level of significance, $\alpha = 0.05$

We will run a hypothesis test to know if the average oxygen level below town is higher than above town.

Step 1: Stating the null and alternative hypothesis

Null hypothesis, $H_0 : \mu_1 - \mu_2 = 0$.
Alternate hypothesis, $H_a : \mu_1 - \mu_2 < 0$.

Step 2: Calculate the test statistic

Here, the sample size for two populations is > 30. So, it falls under the Case II and thus we will use the z-statistic.

$$z = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}}}$$

Here,

$$\Delta_0 = \mu_1 - \mu_2 = 0$$

Since, σ_1 and σ_2 are unknown, we will replace them with s_1 and s_2 respectively.

$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{{s_1}^2}{m} + \frac{{s_2}^2}{n}}}$$

Pluggin in the values,

$$z = \frac{(4.83 - 4.55)}{\sqrt{\frac{0.175^2}{45} + \frac{0.234^2}{45}}}$$

Calculating value using R,

```
(4.83-4.55)/(sqrt((0.175^2/45)+(0.234^2/45)))
```

```
## [1] 6.428115
```

So, the z-statistic is,

$$z = 6.43$$

Step 3: Find the rejection region (RR) or p-value

Rejection region:

The rejection region criteria is,

$$RR : z \leq -z_\alpha$$

Now, from class notes,

$$z_{0.05} = 1.645$$

Using R,

```
qnorm(1-0.05)
```

```
## [1] 1.644854
```

Thus, it is evident that,

$$6.43 \geq -1.645$$

$$z \geq -z_{\alpha}$$

So, z-statistic is not within the RR.

p-value:

$$P(Z \leq z)$$

Using R,

```
pnorm(1.645)
```

```
## [1] 0.9500151
```

It is clearly evident that,

$$0.95 > 0.05$$

$$p - value > \alpha$$

So, the p-value is greater than α.

Step 4: Make a decision

Since, the z-statistic is not within the RR and the p-value is greater than α, we will retain the null hypothesis and reject the alternate hypothesis. This means that, the average oxygen level below town is not higher than above town.

Comment Summary