Stats assignment

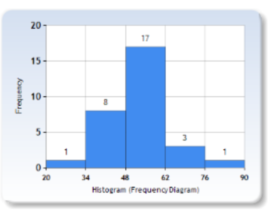
Name : Saba Hajebi

Student ID: 21364577

1) Sampling Process

- I defined my population as the past 200 days and chose a sample size of 30 randomly selected days.
- I tracked the time it takes me to commute for 200 days by using my location history on google maps.
- I wrote down the different times and shuffled the times randomly.
- I then selected 30 random days out of the population.
- My thoughts about the process were that it was relatively simple to gather the data but since I wanted the commute durations to be random and representative to the population, I had to come up with a way to make sure there was no bias in the process of random selection. An irregularity I considered was to be wary of defining a small population that could be impacted by seasonal factors. By picking 200 days I made sure to rule out any biased observations related to the time of year.

2) Data Description



Frequency Table	
Class	Count
20-33	1
34-47	8
48-61	17
62-75	3
76-89	1

Characteristics:

Mean	52.4
Standard Deviation (s)	12.18705
Lowest Score	25
Highest Score	89
Distribution Range	64
Total Number of Scores	30
Lowest Class Value	20
Highest Class Value	89
Class Range	14

Shape: The distribution follows a near pattern of the gaussian distribution, with a concentration of data in the center.

Center: The data seems to be centered around 48-61 minutes.

Spread: the data ranges from 25 minutes to 89 minutes, so the range is 64 minutes.

Appropriateness of:

Z distribution:

- This is appropriate because the standard deviation is known so you can conduct confidence intervals, and when the sample size is n >= 30.

T distribution:

- This isn't appropriate since T distributions are used when standard deviation is not known so you can make use of the standard error when calculating confidence intervals and when the sample size is n < 30.

Both t-test and z-test partake in the different use of distribution to correlate values and make conclusions in terms of hypothesis testing.

I concur that the z distribution is more suitable.

4) Estimation

95% Confidence interval

$$x \pm z\left(\frac{s}{\sqrt{n}}\right)$$

$x = 52.4$
$n = 30$
$\sigma = 12.18$
$CI = 95\%$

$$52.4 + 1.96\left(\frac{12.18}{\sqrt{30}}\right) = 56.758$$

$$52.4 - 1.96\left(\frac{12.18}{\sqrt{30}}\right) = 48.042$$

∴ With 95% confidence, the population mean is in the range [48.042 – 56.758]

97.5% Confidence interval

$$x \pm z\left(\frac{s}{\sqrt{n}}\right)$$

$x = 52.4$
$n = 30$
$\sigma = 12.18$
$CI = 97.5$

$$52.4 + 2.24\left(\frac{12.18}{\sqrt{30}}\right) = 57.384$$

$$52.4 - 2.24\left(\frac{12.18}{\sqrt{30}}\right) = 47.416$$

∴ With 97.5% confidence, the population mean is in the range [57.384 – 47.416]

5) Hypothesis Testing

Hypothesis testing

alpha = 0.025

$H_0 : \mu = 53$
$H_1 : \mu \neq 53$

$$\text{Test statistic} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

mean

median

$$\frac{53 - 52.4}{\frac{12}{\sqrt{30}}} = 0.27$$

$1 - 0.67 = 0.3333 = 33.3\%$

Conclusion: the probability of observing a sample mean as extreme as 52.4 is 33.3% which is above the 0.25% level of significance, therefore we fail to reject the null hypothesis.

6) Sample size

$$\left(\frac{2.58 \times 12}{4.35}\right)^2 = 51$$

To check if it is accurate:

$$2.58^2 \times \left(\frac{12^2}{51}\right) = 18.79$$

$$= \sqrt{18.79} = 4.33 \text{ (half width)}$$

∴ estimated sample is accurate.

6)

Sample Size

The estimated sample size is reliable and accurate as it correlates with our critical value. Our resulting 99% CI (half width) for a proportion will not exceed the proposed half width.