Assignment - 7

Pradyoth Singenahalli Prabhu - 02071847

Problem - 1

According to the Hospital Care cost Institute, the annual expenditure for prescription drugs is $838 per person in the Northeast region of the country. A sample of 60 individuals in the Midwest shows a per person annual expenditure for prescription drugs of $745. Use a population standard deviation of $300 to answer the following questions.

a. Formulate hypotheses for a test to determine whether the sample data support the conclusion that the population annual expenditure for prescription drugs per person is lower in the Midwest than in the Northeast.

- **Null hypothesis (H0):** The population annual expenditure for prescription drugs per person is not lower in the Midwest compared to the Northeast. This can be represented as **μMidwest ≥ μNortheast**.
- **Alternative hypothesis (H1):** The population annual expenditure for prescription drugs per person is lower in the Midwest compared to the Northeast. This can be represented as **μMidwest < μNortheast**.

b. What is the value of the test statistic?

The formula for the test statistic (z) is:

$$z = \frac{(\bar{X} - \mu)}{(\frac{\sigma}{\sqrt{n}})}$$

Where:

- $\bar{X}$ is the sample mean (in this case, **$745**)
- μ is the population mean (in this case, **$838**)
- σ is the population standard deviation (given as **$300**)
- n is the sample size (in this case, **60**)

$$z = \frac{(745 - 838)}{(\frac{300}{\sqrt{60}})}$$

z = **-2.401249675**

c. What is the p-value?

In this case, we would only need to find the p-value for the left tail of the distribution.

For a one-tailed test, the p-value can be calculated using the following Excel formula:

`=NORM.S.DIST(A2, TRUE)` = **0.008169592**

d. At α= 0.01, what is your conclusion?

At an alpha level of 0.01, we compare the p-value to the significance level to make a decision about the null hypothesis.

With a p-value of approximately 0.0082, which is less than the significance level of 0.01, we have enough evidence to **reject** the null hypothesis.

Therefore, at the 0.01 significance level, we can conclude that the population annual expenditure for prescription drugs per person is significantly lower in the Midwest than in the Northeast.

Problem - 2

The United States ranks ninth in the world in per capita chocolate consumption; Forbes reports that the average American eats 9.5 pounds of chocolate annually. Suppose you are curious whether chocolate consumption is higher in Hershey, Pennsylvania, the location of the Hershey Company's corporate headquarters. A sample of 36 individuals from the Hershey area showed a sample mean annual consumption of 10.05 pounds and a standard deviation of s = 1.5 pounds. Using α =.05, do the sample results support the conclusion that mean annual consumption of chocolate is higher in Hershey than it is throughout the United States?

- **Null hypothesis (H0):** The mean annual consumption of chocolate in Hershey is not higher than the national average. ***μ*≤9.5**
- **Alternative hypothesis (H1):** The mean annual consumption of chocolate in Hershey is higher than the national average. ***μ*>9.5**

To conduct the hypothesis test, we'll first calculate the test statistic (z-score) using the formula:

$$z = \frac{(\bar{X} - \mu)}{(\frac{\sigma}{\sqrt{n}})}$$

Where:

- $\bar{X}$ is the sample mean = **10.05** pounds
- μ is the population mean = **9.5** pounds
- σ is the population standard deviation = **1.5** pounds
- n is the sample size = **36**

$$z = \frac{10.05 - 9.5}{\frac{1.5}{\sqrt{36}}} = \frac{0.55}{\frac{1.5}{6}} = \frac{0.55}{0.25} = 2.2$$

Finding p-value `=1-NORM.S.DIST(C2,TRUE)` = **0.013903448**

Since the p-value (0.013903448) is lesser than the significance level (0.05), we **reject** the null hypothesis. Therefore, based on the sample results, there is enough evidence to support the conclusion that the mean annual consumption of chocolate is higher in Hershey, Pennsylvania, than it is throughout the United States.

Problem - 3

Last year, a soft drink manufacturer had 21% of the market. In order to increase their portion of the market, the manufacturer has introduced a new flavor in their soft drinks. A sample of 400 individuals participated in the taste test and 100 indicated that they like the taste. We are interested in determining if more than 21% of the population will like the new soft drink.

a. Set up the null and the alternative hypotheses.

Null hypothesis (H0): The proportion of individuals who like the new soft drink is 21% or less **(p ≤ 0.21)**.
Alternative hypothesis (H1): The proportion of individuals who like the new soft drink is greater than 21% **(p > 0.21)**.

b. Determine the test statistic.

$z = \frac{\hat{p}-p}{SE}$

where SE is Standard Error, defined as:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Given:

- p = 0.21 (hypothesized proportion under the null hypothesis)
- $\hat{p}$ = 0.25 (sample proportion)
- n = 400 (sample size)

$$SE = \sqrt{\frac{0.21(1-0.21)}{400}} = 0.020365412$$

Now substituting SE in z:

$z = \frac{0.25-0.21}{0.020365412} \approx 1.96$

c. Determine the p-value.

Using excel formulae, `=1-NORM.S.DIST(F3,TRUE)` = **0.02475841**

d. Using α = .05, test to determine if more than 21% of the population will like the new soft drink.

And the given p-value of 0.02475841 at a significance level of 0.05, we can conclude that the p-value is less than the significance level. Therefore, we **reject** the null hypothesis.

Problem - 4

Consider the following hypothesis test:

$$H_0 : \mu \leq 38$$
$$H_a : \mu > 38$$

You are given the following information obtained from a random sample of six observations.

X	38	40	42	32	46	42

a) Using α = 0.05, what is the rejection rule?

To find the critical value from the t-distribution table, you can use the T.INV function in Excel:

`=T.INV(0.05, 5)` = **- 2.015048373** (this is for lower tail)

`=T.INV((1-0.05), 5)` = **2.015048373** (this is for upper tail) we choose this because

If the calculated value of t is greater than 2.015, we reject the null hypothesis H_0 in favor of the alternative hypothesis H_a.

b) Determine the standard error of the mean.

The standard error of the mean (SEM) can be calculated using the formula:

$SEM = \frac{s}{\sqrt{n}}$

where,

- s is the sample standard deviation, and
- n is the sample size.

Finding standard deviation,

$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

where,

X = {38, 40, 42, 32, 46, 42}

$\bar{X}$ is sample mean

Finding sample mean:

$\bar{X} = \frac{\sum X}{n}$ = **40**

Substituting sample mean to find standard deviation,

$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$ = **4.732863826**

Now, we can find the standard error of the mean (SEM) using the formula:

$SEM = \frac{s}{\sqrt{n}} = \frac{4.732863826}{6}$ = **1.932183566**

c) Compute the value of the test statistic. What is your conclusion?

To compute the value of the test statistic, we'll use the formula for the t-test:

$t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$

$t = \frac{40-38}{\frac{4.72}{\sqrt{6}}}$ = **1.035098339**

Finding p-value: `=1-T.DIST(1.04,5,1)` = **0.174033097**

Since the p-value (0.174033097) is greater than the significance level (0.05), we do not have enough evidence to reject the null hypothesis. Therefore, we **fail to reject** the null hypothesis. This implies that there is insufficient evidence to conclude that the population mean is greater than 38.

Case Study: Quality Associates, Inc.

Quality Associates, Inc. Quality Associates, Inc., a consulting firm, advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. In one particular application, a client gave Quality Associates a sample of 800 observations taken during a time in which that client's process was operating satisfactorily. The sample standard deviation for these data was 0.21; hence, with so much data, the population standard deviation was assumed to be 0.21. Quality Associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. By analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. When the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. The design specification indicated the mean for the process should be 12. The hypothesis test suggested by Quality Associates follows.

$H_0 : \mu = 12$

$H_a : \mu \neq 12$

Corrective action will be taken any time H_0 is rejected.

The following samples were collected at hourly intervals during the first day of operation of the new statistical process control procedure. These data are available in the data set *Quality*.

Data set *Quality*

Sample 1	Sample 2	Sample 3	Sample 4
11.55	11.62	11.91	12.02
11.62	11.69	11.36	12.02
11.52	11.59	11.75	12.05
11.75	11.82	11.95	12.18
11.90	11.97	12.14	12.11
11.64	11.71	11.72	12.07
11.80	11.87	11.61	12.05
12.03	12.10	11.85	11.64
11.94	12.01	12.16	12.39
11.92	11.99	11.91	11.65
12.13	12.20	12.12	12.11
12.09	12.16	11.61	11.90
11.93	12.00	12.21	12.22
12.21	12.28	11.56	11.88
12.32	12.39	11.95	12.03
11.93	12.00	12.01	12.35
11.85	11.92	12.06	12.09
11.76	11.83	11.76	11.77
12.16	12.23	11.82	12.20
11.77	11.84	12.12	11.79
12.00	12.07	11.60	12.30
12.04	12.11	11.95	12.27
11.98	12.05	11.96	12.29
12.30	12.37	12.22	12.47
12.18	12.25	11.75	12.03
11.97	12.04	11.96	12.17
12.17	12.24	11.95	11.94
11.85	11.92	11.89	11.97
12.30	12.37	11.88	12.23
12.15	12.22	11.93	12.25

Managerial Report

1. Conduct a hypothesis test for each sample at the 0.01 level of significance and determine what action, if any, should be taken. Provide the test statistic and p-value for each test.

The formula for the test statistic (t) in this case is:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

We have to find t value for all 4 samples.

	Sample 1	Sample 2	Sample 3	Sample 4
Mean	11.9586667	12.0286667	11.889	12.0813333
Standard Deviation	0.22035603	0.22035603	0.20717059	0.206109
t	-1.0780571	0.74768476	-2.8951049	2.12133816
p-value	0.28989425	0.46067068	0.00713177	0.04256295

Based on the provided p-values for each sample:

- Sample 1: p-value = **0.28989425**
- Sample 2: p-value = **0.46067068**
- Sample 3: p-value = **0.00713177**
- Sample 4: p-value = **0.04256295**

For Sample 1 and Sample 2, since the p-values (0.28989425 and 0.46067068) are greater than the significance level of 0.01, we fail to reject the null hypothesis.

For Sample 3 and Sample 4, the p-values (0.00713177 and 0.04256295) are less than the significance level, indicating that we reject the null hypothesis for these samples.

Based on this analysis, Quality Associates, Inc. should consider taking action for Sample 3 and Sample 4, as the results suggest a significant deviation from the specified population mean. However, for Sample 1 and Sample 2, no immediate action is necessary as the results do not indicate a significant deviation from the population mean.

2. Compute the standard deviation for each of the four samples. Does the assumption of 0.21 for the population standard deviation appear reasonable?

Sample	1	2	3	4
Standard Deviation	0.2204	0.2204	0.2072	0.2061

Comparing these values to the assumed population standard deviation of 0.21, we find that the calculated standard deviations are very close to the assumed value. This suggests that the assumption of 0.21 for the population standard deviation appears reasonable based on the sample data. The slight variations in the standard deviations are within an acceptable range and do not significantly deviate from the assumed value.

3. Compute limits for the sample mean $\bar{x}$ around μ=12 such that, as long as a new sample mean is within those limits, the process will be considered to be operating satisfactorily. If $\bar{x}$ exceeds the upper limit or if $\bar{x}$ is below the lower limit, corrective action will be taken. These limits are referred to as upper and lower control limits for quality control purposes.

$$UCL = \mu + 3 \times SE$$

$$LCL = \mu - 3 \times SE$$

Upper limit	12.07936	12.14936	12.00247	12.19422
Lower limit	11.83797	11.90797	11.77553	11.96844

4. Discuss the implications of changing the level of significance to a larger value. What mistake or error could increase if the level of significance is increased?

When the level of significance in a hypothesis test is increased, it means that the criteria for rejecting the null hypothesis becomes less strict. This results in a greater likelihood of rejecting the null hypothesis, even when it is true. In the context of the Quality Associates case study, changing the level of significance to a larger value would imply a shift in the standard for accepting or rejecting the hypothesis regarding the mean of the process.

If the level of significance is increased, the probability of committing a Type I error, also known as a false positive, would increase. In this scenario, a Type I error would occur if the null hypothesis is wrongly rejected when it is actually true. Consequently, the Quality Associates might take corrective action based on an incorrect assessment of the process, leading to unnecessary interventions, increased costs, and disruptions in the manufacturing process.

In summary, by increasing the level of significance, the risk of making decisions based on false results is heightened, which could lead to incorrect conclusions and subsequent inappropriate actions being taken. It is essential to strike a balance between the risks of Type I and Type II errors in order to make informed decisions about the manufacturing process.