Assignment - 5

Pradyoth Singenahalli Prabhu - 02071847

Problem - 1

A simple random sample of 8 employees of a corporation provided the following information.

Employee	1	2	3	4	5	6	7
Age	25	32	26	40	50	54	22
Gender	M	M	M	M	F	M	M

a) Determine the point estimate for the average age of all employees. (Excel function)

Average `=AVERAGE(B2:B9)` = 34

b) What is the point estimate for the standard deviation of the population?

Standard Deviation `=STDEV.S(B2:B9)` = 12.57

c) Determine a point estimate for the proportion of all employees who are female.

$\text{Proportion of females} = \frac{\text{Number of females}}{\text{Total number of employees}} = \frac{2}{8} = 0.25$

Problem - 2

The mean annual cost of automobile insurance is $939. Assume that the standard deviation is $245.

a. What is the probability that a simple random sample of automobile insurance policies will have a sample mean within $25 of the population mean for each of the following sample sizes: 30, 50, 100, and 400? (Write Excel function)

To calculate the probability that a simple random sample of automobile insurance policies will yield a sample mean within $25 of the population mean for different sample sizes, the Excel function NORM.DIST, in conjunction with the Z-distribution, is used. The used formula is as follows:

$\text{Standard Error:} = \frac{\text{standard deviation}}{\sqrt{\text{sample size}}}$

For sample size 30:

SE `=245/SQRT(I3)` = 44.73067553

Probability `=2*NORM.DIST(25,0,J3,TRUE) - 1` = 0.423770419

For sample size 50:

SE `=245/SQRT(I4)` = 34.64823228

Probability `=2*NORM.DIST(25,0,J4,TRUE) - 1` = 0.529421142

For sample size 100:

SE `=245/SQRT(I5)` = 24.5

Probability `=2*NORM.DIST(25,0,J5,TRUE) - 1` = 0.692465076

For sample size 400:

SE `=245/SQRT(I6)` = 12.25

Probability `=2*NORM.DIST(25,0,J6,TRUE) - 1` = 0.958730913

Sample Size	Standard Error	Probability
30	44.73067553	0.423770419

Sample Size	Standard Error	Probability
50	34.64823228	0.529421142
100	24.5	0.692465076
400	12.25	0.958730913

b. What is the advantage of larger sample size when attempting to estimate the population means?

Increasing the sample size in statistical analysis offers the advantage of reducing the margin of error, thereby leading to more precise estimates of population parameters. Larger sample sizes contribute to sample means that closely resemble population means, enhancing the accuracy of statistical inferences. By diminishing the margin of error, there is a greater probability that the sample mean will accurately reflect the population mean, thereby instilling greater confidence in the findings. Moreover, larger sample sizes have a higher likelihood of being representative of the entire population, thus reinforcing the validity of any conclusions drawn. Striking a balance between the benefits of larger sample sizes and the necessary resources for their acquisition and analysis is crucial for conducting robust and informed statistical analyses.

Problem - 3

Students of a large university spend an average of $5 a day on lunch. The standard deviation of the expenditure is $3. A simple random sample of 36 students is taken.

a) What is the expected value, standard error, and shape of the sampling distribution of the sample mean?

$\text{Standard Error:} = \frac{\text{standard deviation}}{\sqrt{\text{sample size}}} = \frac{\$3}{\sqrt{36}} = \$0.5.$

The Central Limit Theorem states that the sampling distribution of the sample mean becomes approximately normal, even if the underlying population distribution is not normal, as the sample size grows larger. In this scenario, since the sample size is 36, it satisfies the condition for a large sample size. Consequently, we can infer that the sampling distribution of the sample mean follows an approximately **normal distributio**n. This characteristic is vital as it enables us to make reliable inferences about the population based on the sample data, thus reinforcing the validity and robustness of our statistical analysis.

b) What is the probability that the sample mean will be at least $4? (Excel function)

`=1 - NORM.DIST(4, 5, 0.5, TRUE)` = 0.977249868 ≈ 0.98

c) What is the probability that the sample mean will be at most $5.50? (Excel function)

`=NORM.DIST(5.5, 5, 0.5, TRUE)` = 0.841344746 ≈ 0.84

Problem - 4

In a large university, 20% of the students are business majors. A random sample of 100 students is selected, and their majors are recorded.

a) Compute the standard error of the proportion.

`=SQRT((0.2*(1-0.2))/100)` = 0.04

b) What is the probability that the sample contains at least 12 business majors? (Write Excel function)

`=1-NORM.DIST(0.12,0.2,0.04,TRUE)` = 0.977249868

c) What is the probability that the sample contains less than 15 business majors? (Write Excel function)

`=NORM.DIST(0.15,0.2,0.04, TRUE)` = 0.105649774

d) What is the probability that the sample contains between 12 and 14 business majors? (Write Excel function)

`=NORM.DIST(0.14,0.2,0.04, TRUE) - NORM.DIST(0.12,0.2,0.04, TRUE)` = 0.044057069

Case Study: Marion Dairies

Last year Marion Dairies decided to enter the yogurt market, and it began cautiously by producing, distributing, and marketing a single flavor—a blueberry-flavored yogurt that it calls Blugurt. The company's initial venture into the yogurt market has been very successful; sales of Blugurt are higher than expected, and consumers' ratings of the product have a mean of 80 and a standard deviation of 25 on a 100-point scale for which 100 is the most favorable score and zero is the least favorable score. Experience has also shown Marion Dairies that a consumer who rates one of its products with a score greater than 75 on this scale will consider purchasing the product, and a score of 75 or less indicates the consumer will not consider purchasing the product.

Emboldened by the success and popularity of its blueberry-flavored yogurt, Marion Dairies management is now considering the introduction of a second flavor. Marion's marketing department is pressing to extend the product line through the introduction of a strawberry flavored yogurt that would be called Strawgurt, but senior managers are concerned about whether or not Strawgurt will increase Marion's market share b appealing to potential customers who do not like Blugurt. That is, the goal in offering the new product is to increase Marion's market share rather than cannibalize existing sales of Blugurt. The marketing department has proposed giving tastes of both Blugurt and Strawgurt to a simple random sample of 50 customers and asking each of them to rate the two flavors of yogurt on the 100-point scale. If the mean score given to Blugurt by this sample of consumers is 75 or less, Marion's senior management believes the sample can be used to assess whether Strawgurt will appeal to potential customers who do not like Blugurt.

Prepare a managerial report that addresses the following issues.

a) Calculate the probability the mean score of Blugurt given by the simple random sample of Marion Dairies customers will be 75 or less.

`=NORM.DIST(75, 80, 25 / SQRT(50), TRUE)` = 0.078649604 ≈ 0.079

b) If the Marketing Department increases the sample size to 150, what is the probability the mean score of Blugurt given by the simple random sample of Marion Dairies customers will be 75 or less?

`=NORM.DIST(75, 80, 25 / SQRT(150), TRUE)` = 0.007152939 ≈ 0.007

c) Explain to Marion Dairies senior management why the probability that the mean score of Blugurt for a random sample of Marion Dairies customers will be 75 or less is different for samples of 50 and 150 Marion Dairies customers.

The probability that the mean score of Blugurt for a random sample of Marion Dairies customers will be 75 or less differs between samples of 50 and 150 customers due to the effect of the sample size on the standard error of the mean.

When the sample size increases, the standard error decreases. This implies that with a larger sample size, the calculated mean score of Blugurt becomes more representative of the true population mean. As a result, the probability of obtaining a mean score of 75 or less decreases as the sample size grows.

Conversely, with a smaller sample size, there is greater uncertainty in the estimated mean, which can lead to a higher probability of observing a mean score of 75 or less, even if the true population mean is higher. This is why the probability of obtaining a mean score of 75 or less is comparatively higher for the sample of 50 customers than for the sample of 150 customers.