Lesson 12
Sampling Distributions

Learning Objectives

Upon completion of this lesson you should be able to do the following.

1. Describe the concept known as sampling distribution.

2. Find the mean and variance of a population of sample means from the mean and variance of the original population.

3. Recognize that the distribution of the sample mean is approximately normal if the sample is of adequate size based on the central limit theorem.

4. Recognize the relationship between sample size and a sampling distribution.

Key Words

sampling distribution, sampling distribution of the sample mean, Central Limit Theorem, mean of the sample mean, variance of the sample mean, standard deviation of the sample mean

Concepts

Samples of the same size drawn from the same population result in different values for the various sample statistics, such as the sample mean and sample variance. The population of values for a certain sample statistic that would occur from repeated sampling from the same population has a distribution known as the probability distribution of a sample statistic. This type of probability distribution is known as a **sampling distribution** since the possible values of the sample statistic result from repeated sampling. The sampling distribution of primary concern is the sampling distribution of the sample mean.

Sampling Distribution of the Sample Mean

The **sampling distribution of the sample mean** tells us what values are possible for the sample mean and the likelihood of those values. The population of the sample mean is the set of all possible values that can occur for the sample mean from samples of the same size drawn from the same population. The sampling distribution of the sample mean is the probability distribution for this population of all possible sample mean values. What is the shape of the probability density function for this distribution? What are the values of the mean and variance of this set of all possible sample means? What factors affect the variance or spread of the possible sample mean values?

The **Central Limit Theorem** states that regardless of the shape of the original population distribution, if the sample size is adequately large, then sums and averages resulting from the sample are approximately normally distributed. This means that if the sample size is large enough then the distribution of the sample mean is approximately normal, regardless of the original population shape. The original sampled population does not even have to be continuous; it can even be discrete. How many observations the sample must have to be consider 'adequately large' is arbitarily set at the value 30.

How close to the normal distribution that the sampling distribution of the sample mean is depends on the original population distribution and the sample size. The distribution of the sample mean achieves normality faster if the original population distribution is not highly skewed. If the original population distribution is close to symmetric then the sampling distribution of the sample mean will be reasonably close to the normal distribution even for fairly small samples. As the sample size increases the sampling distribution of the sample mean is closer to the normal distribution.

The Central Limit Theorem is essential to the field of statistics. This theorem indicates the shape of the distribution of the sample mean is approximately normal, even if nothing is known about the shape of the original population, as long as the number of observations is adequate. As long as the data set is big enough then the distribution of the sample mean is known to be approximately normally distributed. This is the foundation for many important statistical processes.

Where is the center of this normal distribution that represents the set of all possible sample means? How much spread does the distribution of the sample mean have? These questions are answered by sampling properties. The average of all of the possible sample mean values is called the **mean of the sample mean** and is equal to the mean of the original population from which the samples were drawn.

The variance, or spread, of the set of all possible sample mean values is less than the variance of the original population. How much less depends on how many observations are included in the samples. The **variance of the sample mean**, which is the variance of the set of all possible sample means, is the variance the original population from which the samples were drawn divided by the sample size. The **standard deviation of the sample mean** is the square root of the variance or can be calculated as the original population standard deviation divided by the square root of the sample size.

The dispersion of the sample means is affected by two factors, the dispersion in the original population and the number of observations in the sample. As the sample size increases the spread of the set of sample mean values decreases. The bigger the sample the tighter the distribution of the sample mean around the population mean. If the sample size is large, so that many numbers go into each sample average, then the set of all averages is not as spread out as it would be for smaller samples.

Information about the sampling distribution of the sample mean will be used to answer questions about the likelihood of certain sample mean values. If the population mean is believed to be a certain value then the distribution of the sample mean is known. Is the value observed for the sample mean a likely value in this set? If not, perhaps the belief about the population mean is not correct.

The following statements summarize the information about distribution of the sample mean.

The Random Variable, the Sample Mean: $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

The Mean of the Sampling Distribution of the Sample Mean: $\mu_{\overline{X}} = \mu$.

The Variance of the Sampling Distribution of the Sample Mean: $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$.

The Standard Deviation of the Sampling Distribution of the Sample Mean: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

If adequately large samples of size n are drawn from a population with a mean μ and a variance σ^2, then the resulting sample mean is normally distributed with a mean μ and a variance $\frac{\sigma^2}{n}$.

Standardization of $\overline{X}$

The random variable the sample mean is normal or at least approximately so by the Central Limit Theorem. Since the sample mean is normally distributed it can be standardized to generate z-scores to use as measures of relative standing for values of the sample mean. Recall that standardization is a process on a normally distributed random variable that changes the mean of the distribution to zero and the standard deviation to one. The mean is change to zero by subtracting off the mean; the standard deviation is change to one by dividing by the standard deviation. To standardize the sample mean consider the following.

To change $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ *to* $Z \sim N(0,1)$, *use* $Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$.

To change $Z \sim N(0,1)$ *to* $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$, *use* $\overline{X} = \mu + Z(\sigma_{\overline{X}}) = \mu + Z\left(\frac{\sigma}{\sqrt{n}}\right)$.

Exercise with Answers

Sampling Distribution

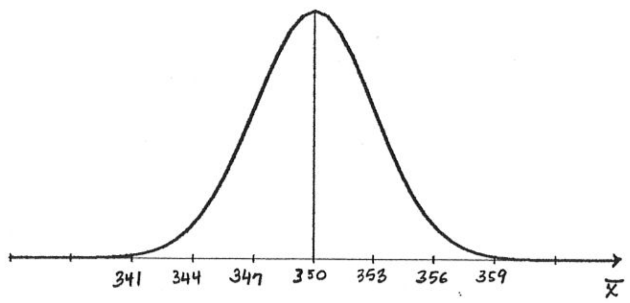
Assume that a random sample with 400 observations was drawn from a population with a mean of $350 and a standard deviation of $60.

1. State the mean and the standard deviation of the sample mean that would result from such a sampling situation. Use the correct symbols.
2. Draw the sampling distribution of the sample mean, label the mean and the points that are one, two and three standard deviations from the mean.
3. What is the probability of observing a sample mean less than the value $355? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
4. What is the chance of observing a sample mean between the values $344 and $354.5? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
5. What is the probability that the sample mean is less than $345.80? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
6. Thirty-three percent of all the possible sample means that could result from the above sampling situation are less than what amount? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
7. Of all the possible sample means from the above sampling situation 43.7% will have values between $350 and what amount? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
8. Two and one-half percent of the sample mean values that could occur from the above sampling situation are more than what value? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
9. If a sample of 400 yielded a mean of 338 would you begin to think that μ, the population mean, was actually less than 350? Why or why not? Answer with complete sentences.

Answers

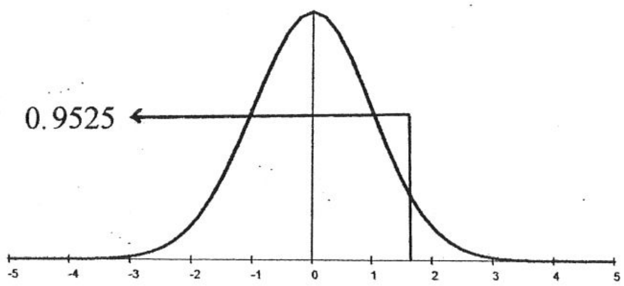
1. $\mu_{\overline{X}} = \mu = 350,\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{400}} = 3.$

2.



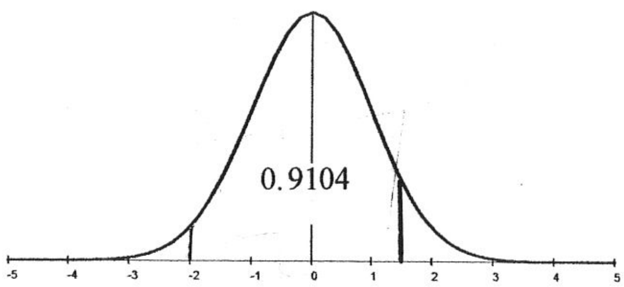
Sampling Distribution of the Sample Mean

3. $P(\overline{X} < 355) = P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{355-350}{\frac{60}{\sqrt{400}}}\right) = P(Z < 1.67) = 0.5 + 0.4525 = 0.9525.$

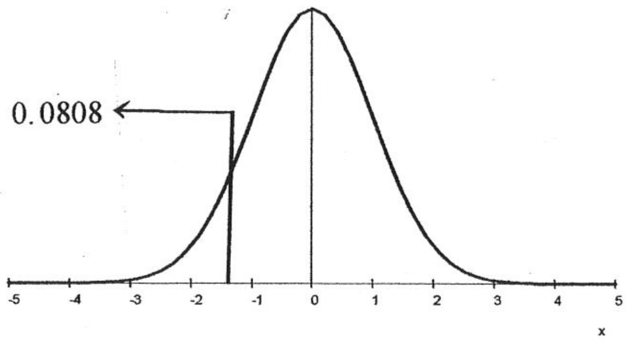


4. $P(344 < \overline{X} < 354.5) = P\left(\frac{344-350}{\frac{60}{\sqrt{400}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{354.5-350}{\frac{60}{\sqrt{400}}}\right) =$

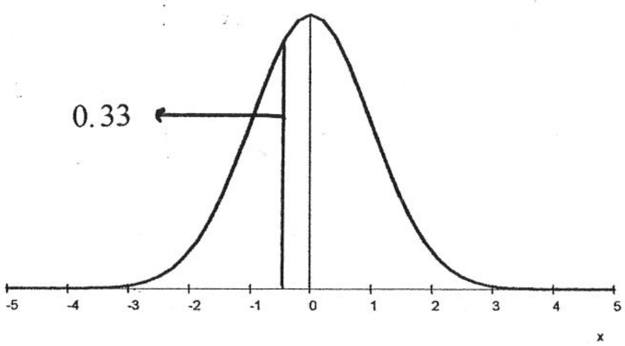
$= P(-2 < Z < 1.5) = 0.4772 + 0.4332 = 0.9104.$



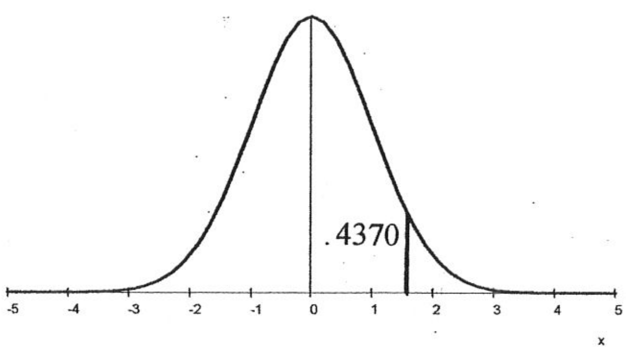
5. $P(\overline{X} < 345.8) = P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{345.8-350}{\frac{60}{\sqrt{400}}}\right) = P(Z < -1.4) = 0.5 - 0.4192 = 0.0808.$



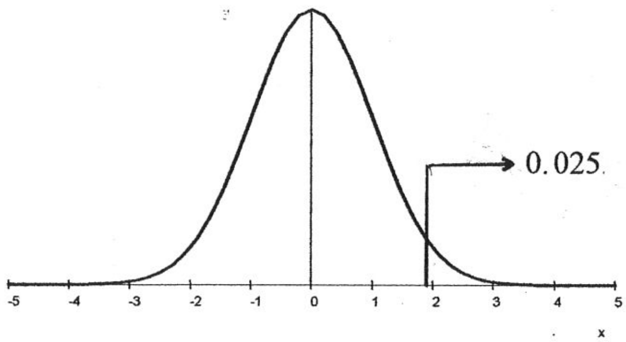
6. Find $\overline{x}_o$, such that, $P(\overline{X} < \overline{x}_o) = 0.33$. First find z_o, such that, $P(Z < z_o) = 0.33$. $P(z_o < Z < 0) = 0.17$, so $z_o = -0.44$. Then $\overline{x}_o = \mu + z_o(\frac{\sigma}{\sqrt{n}}) = 350 + (-0.44)3 = 348.68$



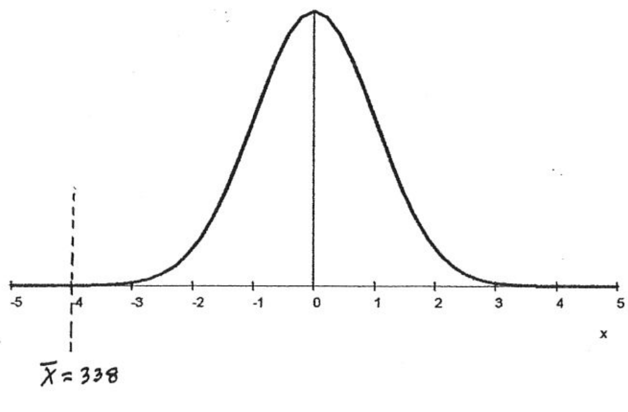
7. Find $\overline{x}_o$, such that, $P(350 < \overline{X} < \overline{x}_o) = 0.4370$. First find z_o, such that, $P(0 < Z < z_o) = 0.4370$, so $z_o = 1.53$. Then $\overline{x}_o = \mu + z_o(\frac{\sigma}{\sqrt{n}}) = 350 + (1.53)3 = 354.59.$



8. Find $\overline{x}_o$, such that, $P(\overline{X} > \overline{x}_o) = 0.025$. First find z_o, such that, $P(Z > z_o) = 0.025$. $P(0 < Z < z_o) = 0.4750$, so $z_o = 1.96$. Then $x_o = \mu + z_o(\frac{\sigma}{\sqrt{n}}) = 350 + 1.96(3) = 355.88$.



9. The sample mean value of 338 is very unlikely if the population mean is 350, since it has a z-score of -4.0, which is a very unlikely z-score. Such an unlikely sample mean would lead to a conclusion that the population mean is significantly less than 350.



Exercise

Sampling Distribution

The monthly utility bill, not including internet or telephone service, for a standard one-bedroom apartment has a mean of \$62 and a standard deviation of \$15. Assume that 100 apartments of this type are randomly sampled.

1. State the mean and the standard deviation of the sample mean that would result from such a sampling situation. Use the correct symbols.
2. Draw the sampling distribution of the sample mean, label the mean and the points that are one, two and three standard deviations from the mean.
3. What is the probability of observing a sample mean less than the value \$59.90? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
4. What is the chance of observing a sample mean between the values \$59.30 and \$63.35? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
5. What is the probability that the sample mean is between \$58.25 and \$61.25? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
6. Thirty-three percent of all the possible sample means that could result from the above sampling situation are less than what amount? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
7. Of all the possible sample means from the above sampling situation 43.7% will have values between what amount and \$62? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
8. Five percent of the sample mean values that could occur from the above sampling situation are less than what value? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
9. If a sample mean of \$56.75 were reported would you begin to think that μ, the mean utility cost of such apartments, was actually less than \$62? Why or why not? Answer with complete sentences.
10. Assume that the sample mean on 100 apartments was \$63.50. What is the chance of observing a sample mean this much or more? Do you think a sample mean of \$63.50 would indicate that the mean monthly utility bill has increased? Why or why not? Answer in complete sentences.
11. The sample mean would have to be more than what amount before you would be convinced that the average monthly utility bills for the population of such apartments had increased from \$62? State your rationale. Answer in complete sentences. Your answer may differ from others' answers.

Exercise

Sampling Distribution

An assembly line manufactures steel rods with an average length of 30 inches and a standard deviation of 0.1 inch. To assure quality control on the length of the rods, 100 rods are randomly sampled each day off of the manufacturing line and the lengths of each of the 100 rods are recorded.

1. State the mean and the standard deviation of the sample mean that would result from such a sampling situation. Use the correct symbols.
2. Draw the sampling distribution of the sample mean, label the mean and the points that are one, two and three standard deviations from the mean.
3. What is the probability of observing a sample mean less than the value 29.985 inches? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
4. What is the chance of observing a sample mean that is further than three standard deviations from the stated population mean of 30 inches? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
5. What is the probability that the sample mean for a given day would be more than 30.012 inches? Give the probability statements for $\overline{X}$, the transformation from $\overline{X}$ to Z, and for the Z variable. Give a graph; shade the area of interest.
6. Ten percent of the time the daily sample means that would result from the above sampling situation are less than what amount? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
7. If the manufacturing process is in control, in terms of the mean length being equal to 30 inches, 0.1% of the daily sample means will be more than what value? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
8. Five percent of the sample means that would occur from the above sampling situation are outside what interval of values? Give probability statements in terms of $\overline{x}_o$ and in terms of z_o. State the equation to calculate $\overline{x}_o$ from z_o. Give a graph; shade the area of interest.
9. If the daily sample mean was 29.94 inches do you think this would provide evidence that the manufacturing process is out of control and not producing rods that are 30 inches on average? Why or why not? Answer with complete sentences.
10. Assume that the sample mean from today's 100 steel rods was 30.005 inches. What is the chance of observing a sample mean this value or greater? Do you think a sample mean of 30.005 would indicate that the mean steel rod length has increased? Why or why not? Answer in complete sentences.
11. The sample mean would have to be more than what value before you would be convinced that the average length of the steel rods being manufactured is actually longer than the 30 inches advertised? State your rationale. Answer in complete sentences. Your answer may differ from others' answers.