14-5

Study Guide

Sample Sets of Data

In statistics, the word **population** refers to an entire set of items or individuals in a group. Rarely will 100% of a population be accessible as a source of data. Therefore, researchers usually select a **random sample** of the population to represent the entire population. Because discrepancies are common in random samples, researchers often take many samples and assume that the sample mean is near its true population mean. The standard deviation of the distribution of the sample means is known as the **standard error of the mean.**

Standard Error of the Mean	If a sample of data has N values and σ is the standard deviation, the standard error of the mean $\sigma_{\bar{x}}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$.

Example 1 **A sample of data has 4500 values and a standard deviation of 12. What is the standard error of the mean?**

For the sample, $N = 4500$. Find $\sigma_{\bar{x}}$.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = \frac{12}{\sqrt{4500}}$, or about 0.179

The standard error of the mean for the set of data is approximately 0.179.

Example 2 **The daily calorie consumption of people in the United States is normally distributed. A team of nutritionists takes a sample of 250 people and records their daily calorie consumption. From this sample, the average daily calorie consumption is 2150 with a standard deviation of 60 calories per day. Determine the interval about the sample mean that has a 1% level of confidence.**

A 1% level of confidence means that there is less than a 1% chance that the true mean differs from the sample mean by a certain amount. A 1% level of condifence is given when $P = 99\%$. When $P = 0.99$, $t = 2.58$.

Find $\sigma_{\bar{x}}$. $\sigma_{\bar{x}} = \frac{60}{\sqrt{250}}$ or about 3.795

Find the range. $\overline{X} = 2150 \pm 2.58(3.795)$

$\overline{X} = 2140.2089$ and 2159.7911

Thus, the probability is 99% that the true mean is within the interval of 2140.2089 calories per day to 2159.7911 calories per day.