

Section 5.4, Sampling Distributions and the Central Limit Theorem

A **sampling distribution** is the probability distribution of a sample statistic when samples of size n are taken randomly from the population repeatedly (typically, a simple random sample is used).

The **sampling distribution of the sample mean** has mean and standard deviation denoted by $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$, respectively. If μ and σ represent the mean and standard deviation for the population, we can calculate $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ by:

$$\begin{aligned}\mu_{\bar{x}} &= \mu \\ \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}}\end{aligned}$$

The standard deviation of the sampling distribution for the mean is often called the **standard error of the mean**.

Example

A measurement from a population has population mean 6 and standard deviation 2. What are the mean and standard error of $\bar{x}$ when $n = 4$? When $n = 100$? When $n = 400$?

For all sample sizes, $\mu_{\bar{x}} = 6$. For $n = 4$, $\sigma_{\bar{x}} = \frac{2}{\sqrt{4}} = 1$. For $n = 100$, $\sigma_{\bar{x}} = \frac{2}{\sqrt{100}} = 0.2$. For $n = 400$, $\sigma_{\bar{x}} = \frac{2}{\sqrt{400}} = 0.1$.

1 Central Limit Theorem

The Central Limit Theorem (CLT) describes the shape of the sampling distribution of the sample mean.

If the population is normally distributed, then the sampling distribution of $\bar{x}$ is normally distributed for any sample size n .

The **Central Limit Theorem (CLT)** says that, regardless of the population distribution (in most cases), if $n \geq 30$, then the sampling distribution of $\bar{x}$ is approximately normal with mean, variance, and standard error given by:

$$\begin{aligned}\mu_{\bar{x}} &= \mu \\ \sigma_{\bar{x}}^2 &= \frac{\sigma^2}{n} \\ \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}}\end{aligned}$$

The larger n is, the better the approximation will be. Also, the closer to normal the population is, the better the approximation will be. If the original distribution is drastically skewed (for example, people's winnings when they play the lottery), you may need a sample size that is larger than 30 to get a good approximation.

Examples

1. For the population of farm workers in New Zealand, suppose that weekly income has a distribution that is skewed right with a mean of $\mu = \$500$ (N.Z. dollars) and a standard deviation of $\sigma = \$160$. A survey of 100 farm workers is taken, including information on their weekly income.

- (a) What are the mean and standard error of the sampling distribution of $\bar{x}$?
The mean is $\mu_{\bar{x}} = 500$, and the standard error is $\sigma_{\bar{x}} = 160/\sqrt{100} = 16$.
- (b) What is the probability that the mean weekly income of these 100 workers is less than \$448?
Using the CLT, we will find the z -score for 448, then use the table to find the probability

$$P(\bar{x} < 448) = P\left(z < \frac{448 - 500}{16}\right) = P(z < -3.25) = 0.0006$$

- (c) What is the probability that the mean weekly income of these 100 workers is between \$480 and \$520?

$$\begin{aligned} P(480 \leq \bar{x} \leq 520) &= P\left(\frac{480 - 500}{16} \leq z \leq \frac{520 - 500}{16}\right) = P(-1.25 \leq z \leq 1.25) \\ &= 0.8944 - 0.1056 = 0.7888 \end{aligned}$$

2. The heights of 18-year-old men are approximately normally distributed with mean of 68 inches and standard deviation of 3 inches.

- (a) What is the probability that an 18-year-old man selected at random is between 67 and 69 inches tall?
With a sample size of 1, we just need that the heights are approximately normal, not the CLT:

$$\begin{aligned} P(67 \leq x \leq 69) &= P\left(\frac{67 - 68}{3} \leq z \leq \frac{69 - 68}{3}\right) = P(-0.33 \leq z \leq 0.33) \\ &= 0.6293 - 0.3707 = 0.2586 \end{aligned}$$

- (b) For a sample of 36 18-year-old men, what is the probability that the average of their heights is between 67 and 69 inches?

For this, we will carry out the same calculation as for the last part, but replace the population standard deviation by the standard error: $\sigma_{\bar{x}} = 3/\sqrt{36} = 0.5$:

$$\begin{aligned} P(67 \leq \bar{x} \leq 69) &= P\left(\frac{67 - 68}{0.5} \leq z \leq \frac{69 - 68}{0.5}\right) = P(-2 \leq z \leq 2) \\ &= 0.9772 - 0.0228 = 0.9544 \end{aligned}$$

Note that since the distribution of heights is approximately normal, we did not need $n \geq 30$ to get probabilities from the standard normal table.

3. For people under 50, the level of glucose in the blood (in milligrams per deciliter of blood) after a 12-hour fast have a standard deviation of 25 and a mean of μ . What is the probability that, for a sample of size 49 readings, the sample mean is within 7 of μ ?

We can calculate the standard error $\sigma_{\bar{x}} = 25/\sqrt{49} = 3.5714$. We are not given the value of the population mean μ , but we know that we want to find $P(-7 \leq \bar{x} - \mu \leq 7) = P(\frac{-7}{3.5714} \leq z \leq \frac{7}{3.5714}) = P(-1.96 \leq z \leq 1.96) = 0.975 - 0.025 = 0.95$