Confidence Intervals

Our goals for today will be to estimate the population mean μ from data when σ is known, and to estimate p from data for a binomial distribution. We do this by calculating something called a **confidence interval**.

A **level c confidence interval** for a parameter is an interval computed from data such that the probability of the parameter being in that interval is c. The most common value of c is 0.95. Each confidence interval will have the form "point estimate" $\pm$ "Margin of Error".

For example, a 95% confidence interval for μ will be a range of values such that we are 95% certain that μ is in that interval.

1 Estimating the Population Mean, μ

Assume the following:

1. A simple random sample of size n was taken from the population.

2. The value of σ is known.

3. If the distribution of the data is normal, the sample size n can be any number. If the distribution is unknown, then we will require $n \geq 30$ (or higher in cases of distinct skew) in order to apply the CLT.

The point estimate for μ is $\bar{x}$. The margin of error is $z_c \cdot \frac{\sigma}{\sqrt{n}}$, where z_c is a z-score such that $P(-z_c < z < z_c) = c$. (Some of these values for z_c are given on the bottom of page A16 in Appendix B of your book.)

Therefore, the level c confidence interval for μ is

$$\bar{x} \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$$

Example
Julia has been jogging over a period of several years. The standard deviation of her times to run 2 miles is $\sigma = 1.80$. A random sample of 90 of her jogging times over the last year gives a sample mean of 15.60 minutes. Find a 0.95 confidence interval for μ and interpret your answer.

We are given $\sigma = 1.80$, $\bar{x} = 15.60$, and $n = 90$. $z_{.95} = 1.96$ from Table 4, so the confidence interval is $15.60 \pm 1.96 \cdot \frac{1.80}{\sqrt{n}} = 15.60 \pm 0.37$. This means that we are 95% certain that the population mean for Julia's run times is between 15.23 and 15.97.

2 Estimating p for a Binomial Distribution

In order to use this technique, we will need to use data with at least 5 successes and at least 5 failures.

If our data set has sample size n and x successes, our point estimate will be $\hat{p} = \frac{x}{n}$. Denote $\hat{q} = 1 - \hat{p}$. The margin of error is $z_c \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$, so the level c confidence interval for p is

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} = \hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example

A random sample of 188 books purchased at a local bookstore showed that 66 of the books were murder mysteries. Find and interpret a 90% confidence interval for the proportion of books sold by this store that are murder mysteries.

First, we need to calculate $\hat{p} = \frac{66}{188} = 0.351$. Then, $\hat{q} = 1 - \hat{p} = .649$. From Table 4, we know that $z_{.90} = 1.645$, so the confidence interval is:

$$0.351 \pm 1.645 \cdot \sqrt{\frac{0.351 \cdot 0.649}{188}} = 0.351 \pm 0.0573.$$

Therefore, we are 90% certain that the true value of p is between 0.294 and 0.407.

3 Finding z_c

Some of these values are given directly in Table 4. When they are not given, you will need to approximate them using the probabilities in the table and using that we need $P(-z_c < z < z_c) = c$. The "quick" way to find z_c is to find the z-score that gives the probability closest to $\frac{c+1}{2}$ in the table. For example, if $z_{.95}$ were not given, we would need to look up the probability $\frac{.95+1}{2} = .975$. This comes from the z-score 1.96, so $z_{.95} = 1.96$ (Note that this agrees with the value they gave).

If we want to find $z_{.97}$, we would need to find the probability $\frac{.97+1}{2} = .985$, which comes from the z-score 2.17, so $z_{.97} = 2.17$.