

Stats: Introduction to Estimation

One area of concern in inferential statistics is the estimation of the population parameter from the sample statistic. It is important to realize the order here. The sample statistic is calculated from the sample data and the population parameter is inferred (or estimated) from this sample statistic. Let me say that again: Statistics are calculated, parameters are estimated.

We talked about problems of obtaining the value of the parameter earlier in the course when we talked about sampling techniques.

Another area of inferential statistics is sample size determination. That is, how large of a sample should be taken to make an accurate estimation. In these cases, the statistics can't be used since the sample hasn't been taken yet.

Point Estimates

There are two types of estimates we will find: Point Estimates and Interval Estimates. The point estimate is the single best value.

A good estimator must satisfy three conditions:

- Unbiased: The expected value of the estimator must be equal to the mean of the parameter
- Consistent: The value of the estimator approaches the value of the parameter as the sample size increases
- Relatively Efficient: The estimator has the smallest variance of all estimators which could be used

Confidence Intervals

The point estimate is going to be different from the population parameter because due to the sampling error, and there is no way to know how close it is to the actual parameter. For this reason, statisticians like to give an interval estimate which is a range of values used to estimate the parameter.

A confidence interval is an interval estimate with a specific level of confidence. A level of confidence is the probability that the interval estimate will contain the parameter. The level of confidence is $1 - \alpha$. $1 - \alpha$ area lies within the confidence interval.

Maximum Error of the Estimate

The maximum error of the estimate is denoted by E and is one-half the width of the confidence interval. The basic confidence interval for a symmetric distribution is set up to be the point estimate minus the maximum error of the estimate is less than the true population parameter which is less than the point estimate plus the maximum error of the estimate. This formula will work for means and proportions because they will use the Z or T distributions which are symmetric. Later, we will talk about variances, which don't use a symmetric distribution, and the formula will be different.

Area in Tails

Since the level of confidence is $1 - \alpha$, the amount in the tails is α . There is a notation in statistics which means the score which has the specified area in the right tail.

Examples:

- $Z(0.05) = 1.645$ (the Z-score which has 0.05 to the right, and 0.4500 between 0 and it)
- $Z(0.10) = 1.282$ (the Z-score which has 0.10 to the right, and 0.4000 between 0 and it).

As a shorthand notation, the $()$ are usually dropped, and the probability written as a subscript. The greek letter alpha is used represent the area in both tails for a confidence interval, and so $\alpha/2$ will be the area in one tail.

Here are some common values

Confidence Level	Area between 0 and z-score	Area in one tail ($\alpha/2$)	z-score
50%	0.2500	0.2500	0.674
80%	0.4000	0.1000	1.282
90%	0.4500	0.0500	1.645
95%	0.4750	0.0250	1.960
98%	0.4900	0.0100	2.326
99%	0.4950	0.0050	2.576

Notice in the above table, that the area between 0 and the z-score is simply one-half of the confidence level. So, if there is a confidence level which isn't given above, all you need to do to find it is divide the confidence level by two, and then look up the area in the inside part of the [Z-table](#) and look up the z-score on the outside.

Also notice - if you look at the [student's t distribution](#), the top row is a level of confidence, and the bottom row is the z-score. In fact, this is where I got the extra digit of accuracy from.

[Table of Contents](#)