

# What is the importance of accuracy in parameter estimation in Stata data analysis?

Authored by  
**stats writer**

June 29, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is the importance of accuracy in parameter estimation in Stata data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159524>

Accuracy in parameter estimation is a crucial aspect of Stata data analysis as it directly impacts the reliability and validity of the results obtained. Parameter estimation involves using statistical techniques to estimate the values of unknown parameters in a given dataset. These estimated parameters are then used to make inferences and draw conclusions about the population from which the data is collected. Therefore, the accuracy of these estimated parameters is essential in ensuring the validity of the conclusions drawn. Inaccurate parameter estimation can lead to incorrect conclusions and potentially misguide decision-making processes. Hence, it is imperative to strive for accuracy in parameter estimation to ensure the credibility and usefulness of Stata data analysis results.

## Accuracy in Parameter Estimation | Stata Data Analysis Examples

### Introduction

**Not all sample size issues are directly related to power. Accuracy in parameter estimation (AIPE) is also a function of sample size, that is, the larger the sample size the smaller the confidence interval for a parameter estimate. Accuracy in parameter estimation allows you to specify the size of confidence interval you want to achieve and, in return, gives you the sample size needed to achieve that confidence interval. As Kelly and Maxwell (2003) state, "The AIPE approach yields precise estimates of population parameters by providing necessary sample sizes in order for the likely widths of confidence intervals to be**

**sufficiently narrow."**

**In this unit we will illustrate how to do an AIPE analysis for a multiple regression model that has two control variables, one categorical research variable and one continuous research variable, with the focus being on the confidence interval for the continuous research variable.**

#### **Description of the Experiment**

**We will be using the same data analysis example that was used in the unit on multiple regression power analysis. In that analysis, a school district is designing a multiple regression study looking at the effect of gender, family income, mother's education and language spoken in the home (3 levels, 2 dummy variables) on the English language proficiency scores of Latino high school students. Mother's education is the primary research variable that measures the number of years that the mother attended school. It is a continuous variable ranging from 4 to 18**

**years.**

**When we ran the power analysis for testing the parameter for mother's education, we came up with sample sizes of 108, 138 and 182 for power values of .7, .8 and .9 respectively. We can check these values against the sample size needed to achieve a researcher specified confidence interval.**

**AIPE**

**To conduct an AIPE analysis we will use the Stata program aipe (search aipe) (see How can I use the search command to search for programs and get additional help? for more information about using search). This program was written by UCLA Academic Technology Services as an implementation of the approach outlined in the Kelley and Maxwell (2003) article.**

**In this analysis we want to determine how large a sample will be needed to have a confidence interval on the regression coefficient that extends 0.1 above and below the point**

**estimate. A confidence interval can be calculated as estimate  $\pm$  margin of error.**

**The margin**

**of error, then, is the half-width width of the confidence interval which in the aipe program is specified using the w option. To use aipe we also need to include r2, the R2 for the full model, r2xx, the R2 for the variable of interest with the other predictor variables, and p the number of variables in the full model.**

**For this analysis, we believe, based on previous research, that the R2 for the full model will be about 0.48 and that the R2 for mother's education with the other predictors will be about 0.4. The total number of predictors in the model is 5 and we want a confidence interval half-width of 0.1 with an alpha level of 0.05.**

**aipe, r2(.48) r2xx(.4) w(.1) p(5) alpha(.05)**

## Accuracy in Parameter Estimation

$p = 5$  -- number of predictor variables in full model

$\alpha = .05$  -- alpha level for confidence interval

$w = .1$  -- confidence interval half-width

$R^2 = .48$  -- R-squared for full model

$R^2_{xx} = .4$  -- R-squared for target predictor with other predictors

$\text{quan} = .8$  -- quantile of chi-square distribution

### AIPE sample size

$N = 339$  -- n needed for CI half-width of  $w = .1$ , 50% of time

$N_m = 366$  -- n needed for CI half-width of  $w = .1$ ,  $100 \times (\text{quan})\% = 100 \times (.8)\% = 80\%$  of time

A confidence interval with a half-width of 0.1 will require either 339

or 366 students, depending on the percent of time that the half-width is likely to occur.

These sample sizes are considerably larger than those from the power analysis for mother's education.

Let's see what making the confidence interval wider does for the sample size. We will rerun

**the analysis using a half-width of 0.15.**

**aipe, r2(.48) r2xx(.4) w(.15) p(5) alpha(.05)**

### **Accuracy in Parameter Estimation**

**p = 5 -- number of predictor variables in full model**

**alpha = .05 -- alpha level for confidence interval**

**w = .15 -- confidence interval half-width**

**R2 = .48 -- R-squared for full model**

**R2xx = .4 -- R-squared for target predictor with other predictors**

**quan = .8 -- quantile of chi-square distribution**

### **AIPE sample size**

**N = 154 -- n needed for CI half-width of w = .15, 50% of time**

**Nm = 174 -- n needed for CI half-width of w = .15, 100\*(quan)% = 100\*(.8)% = 80% of time**

**The required sample sizes are much smaller for a half-width of 0.15 than for 0.1.**

**Based on the accuracy in parameter estimation analyses and taking into consideration that**

the researchers want a very narrow confidence interval half-width of 0.1, we will go with a planned sample size of 340 students. A larger sample size would make the specified confidence interval more likely, but having a half-width of 0.1 50% of the time was considered sufficient by the researchers involved in the study.

See Also

**Multiple Regression Power analysis**

**Kelley, K. and Maxwell, S.E. 2003. Sample Size for Multiple Regression: Obtaining Regression Coefficients That Are Accurate, Not Simply Significant. Psychological Methods, 8(3), 305-321.**