Lab Handout   Lab 5.1: Sampling Distributions

The instructions below are keyed to the lab instructions found on pp. 279-280 of the text. Please use those instructions as well in preparing your report.

Experimental Procedure

1. The macro LAB5_1A will generate as many sets of data from the C+E model as you tell it to. A window will ask you for the number of data sets, the name of the SAS data file where you want the data sets written, the number of observations per data set, and the values of μ and σ^2 that define the C+E model. For this part of the lab, choose μ and σ^2 (You will need to remember these values for step 3 below.), and then use the LAB5_1A macro to generate five data sets of twenty observations each from a normal distribution with your choice of μ and σ^2 as its parameters.

   In SAS/INSIGHT, open the SAS data file you just generated. It will contain five variables, $y1 - y5$, each representing one data set. In this part of the lab, you will create histograms and normal quantile plots for all five variables. You will also compute the parameter estimates $\hat{\mu} = \overline{y}$ and $\hat{\sigma}^2 = s^2$ for all five variables.

   To get all the necessary output, after opening your SAS data file in SAS/INSIGHT, choose *Analyze: Distribution(Y)*. Select all five variables to be analyzed. The resulting distribution window will have a density histogram and the mean and variance (as well as a lot of other stuff) for each variable. To get the normal quantile plots, choose *Curves: QQ Ref line.* You do not have to submit these plots or output with your lab report, but answer the following questions in your report:

   What do the histograms look like? Do they have the same shape, location, and spread? Are the estimates $\hat{\mu} = \overline{y}$ and $\hat{\sigma}^2 = s^2$ consistent with the μ and σ^2 you chose? Do the normal quantile plots show roughly the same trend, and are they linear?

2. Input the values of the five estimates $\hat{\mu} = \overline{y}$, and the five estimates $\hat{\sigma}^2 = s^2$, into a new data file in SAS/INSIGHT and then form histograms of each. The basic idea here is that for different data sets generated from the same C+E model, the parameter estimates will be different. In fact, their values will form a distribution. The distribution formed by the parameter estimates under repeated sampling is called a **sampling distribution** of the estimator. Submit plots of these histograms with your lab report.

3. To get a really good idea of what the sampling distributions of the parameter estimators look like, you'd need to use the estimates from a lot more than five data sets. Using macro LAB5_1B generate 500 data sets each of size 20 from the same C+E model that you chose in step 1. The macro will place the 500 parameter estimates in the SAS data set you name, and will draw histograms of the 500 parameter estimates $\overline{y}$ and the 500 parameter estimates s^2.

4. Use your data analysis skills to summarize the distributions of the 500 parameter estimates $\overline{y}$ and s^2. You will want to use SAS/INSIGHT to obtain appropriate summary measures. Because 500 is a substantial number of observations, these distributions should be good approximations to the sampling distributions of the parameter estimators $\overline{Y}$ and S^2, and your conclusions should be applicable to those sampling distributions as well.