I. Mean square convergence

In this this part, the goal is to illustrate the theoretical result that the Variance of the OLS estimator decreases when the sample size increases.

1 Resampling (6 samples)

2 Estimation

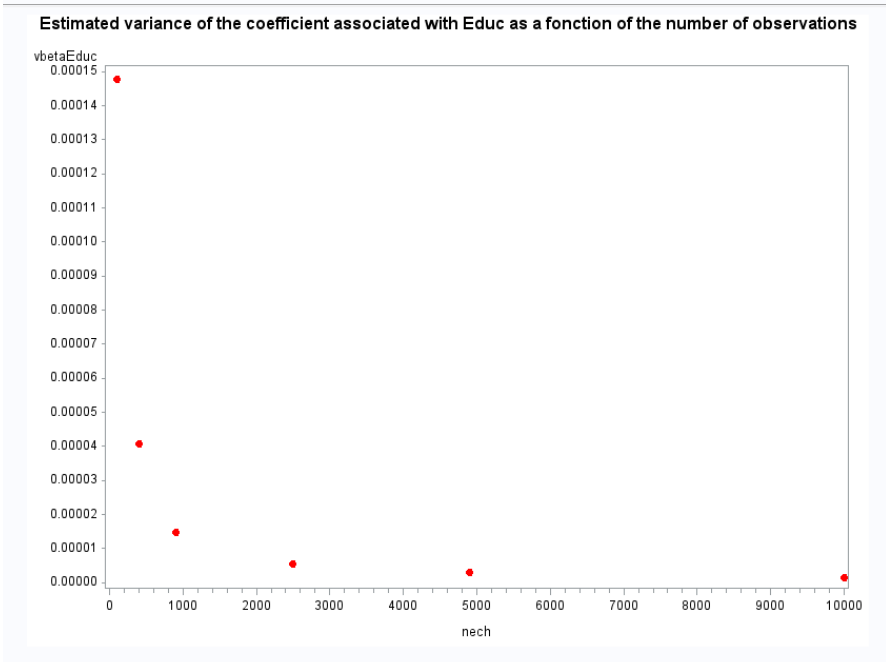
3 Draw Graph

Randomly draw 100 observations from the sample (with replacement). Estimate $\hat{\beta}$.

Repeat for $n = 400, 900, 2500, 4900$ and 10000.

This is Resampling. Resampling methods use a computer to generate a large number of simulated samples of data. Simulated samples are generated by drawing new samples (with replacement) from the sample of data we have. In resampling methods, the researcher DOES NOT know or control the DGP (as it is the case in Monte Carlo simulations), but the goal is learning about the DGP. To this end, patterns in the simulated samples are summarized and described. The results are used to evaluate theoretical and/or statistical properties or assumptions of some estimator.

If we do not use resampling methods, we cannot illustrate this result because we only have one sample and the sample size is fixed. The idea is then to draw several samples of different sizes from the initial sample and to estimate the linear regression model on each simulated sample. For each sample, we will then implement the OLS estimation and store the estimated variance (or the standard error) of each estimated coefficient $\hat{\beta}_k$, for $k = 0, 1, 2$.



II. Bootstraping and Central Limit Theorem.

In this this part, the goal is to illustrate the theoretical result that for a fixed number of bootstrap samples, when the size of the bootstrap samples increases, the distribution of the OLS tends to a normal distribution (Central Limit Theorem)

1 Bootstrap (a type of resampling, draw 100 samples)

2 Estimation

3 Draw Graph (QQ-plot)

1. Implement 100 repetitions of the following steps : (1) randomly draw 100 observations from the sample (with replacement) and (2) re-estimate $\hat{\beta}_b$ for each repetition $b = 1, ., 100$. Plot the density of these estimators (as they are random) and check the normality assumption of this distribution by using a QQplot.
2. Repeat the previous experiment for sample size 900 and 10000. Interpret.

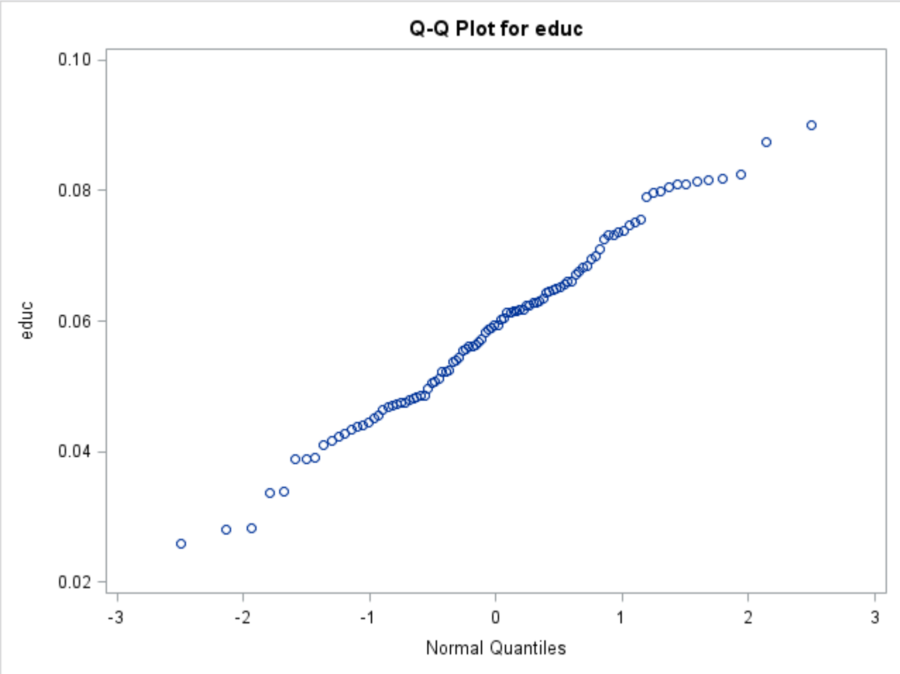
Bootstraping refers to a resampling technique that allows estimation of the sampling distribution of an estimator. In other words, bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from the empirical distribution of the observed data. In the case where observations can be assumed to be from an independent and identically distributed population, this can be implemented by constructing a number of resamples with replacement, of the observed dataset.

One example is you have a statistic whose standard error is very complex to compute, you can use bootstrapping to estimate this standard error and then perform a test, compute CI ... In this case, the principe is to construct B bootstrap samples of the same size using a computer. The bootstrap sample is taken from the original dataset using sampling with replacement. For each of these bootstrap samples, we estimate the statistic of interest, we then have B bootstrap estimation values for our statistic. We now have the bootstrap distribution of our statistic which provides an estimate of the shape of the distribution of our statistic from which we compute the empirical variance (this is the bootstrap estimation of the variance of our statistic) and when taking the square root of this empirical variance, we obtain the bootstrap estimation of the standard error of our statistic. (The method here, described for the standard error, can be applied to almost any other statistic or estimator, such as the mean ...)

QQPLOT of the coefficient associated with Educ

The UNIVARIATE Procedure

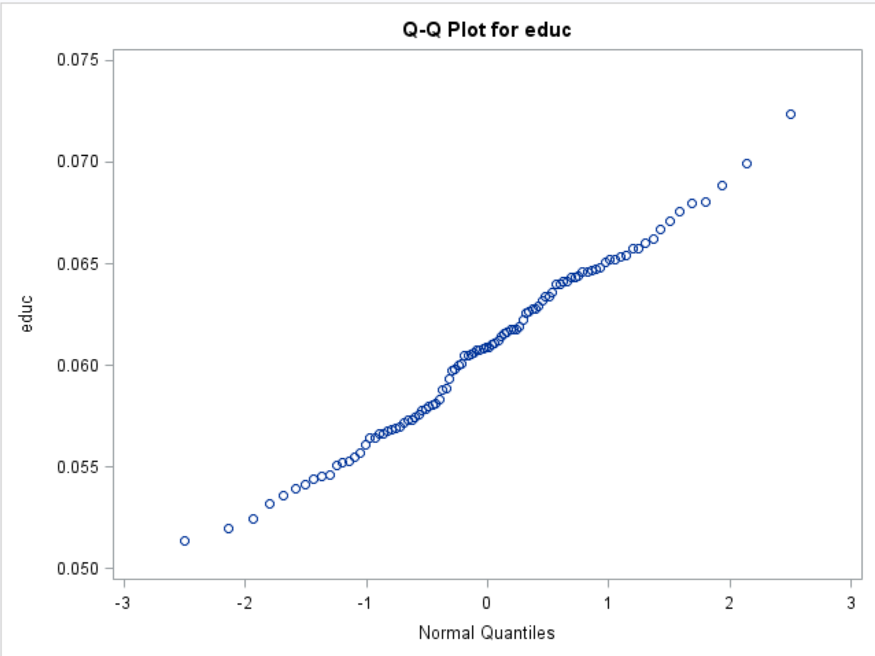
nech=100



QQPLOT of the coefficient associated with Educ

The UNIVARIATE Procedure

nech=900



QQPLOT of the coefficient associated with Educ

The UNIVARIATE Procedure

nech=10000