

Very Short Paper: Stage I

Go find some interesting data. Best would be to collect your own or run an experiment, but data you download from the internet are acceptable too.

Your data should include at least 40 observations of values for at least two variables in a *cross-section* or a panel ($N \geq 40$). Do not use time series data (i.e., X may not be calendar time). The more observations the better.

Report a table of means, which includes for each variable to be used in the analysis, the mean, standard deviation, number of observations, minimum and maximum value. If sampling weights are to be used in analysis, report the weighted means and standard deviations as well. At the bottom of the table you should *document the source*, including enough information to make *replication* of your table possible.

The very short paper will be graded mostly on econometric method, with no grades deducted for even the most ludicrous economic analysis, so feel free to have fun. On the other hand, you will spend many intimate hours with this project, so you may as well design something interesting for yourself.

The table of means is due by email on Thursday January 21, at midnight PST.

Unsure about your choice of dataset? Send me an email and ask, or come to OH, Wed 4pm on [zoom](#).

You can get a jump on this assignment by emailing it to me early: elib@ucsd.edu.

TABLE 3.—DESCRIPTIVE STATISTICS FOR OIL REFINERIES PACE, LRD, AND REGULATORY DATA

	Mean	Standard Deviation
Value of shipments*	1,707,848	2,890,197
Value added	118,772	231,349
Employment	372	500
Air pollution abatement investment	2,096	7,618
Net abatement investment	1,495	7,475
Depreciation of abatement capital	601	1,796
Abatement operating costs	6,586	16,607
Change in abatement operating costs	141	6,951
New regulation adoption dates	0.053	0.369
New regulation compliance dates	0.041	0.267
New increased stringency dates	0.012	0.136
South coast indicator	0.055	0.228
California indicator	0.129	0.335
Texas indicator	0.208	0.406
Louisiana indicator	0.094	0.292

* Thousands of 1991 dollars deflated by the Producer Price Index.

Source: Pollution Abatement Costs and Expenditures microdata.

The sample contains 1,914 observations weighted by PACE sampling weights to represent 2,425 plant-years in the population. Sampled from 1979–1991, excluding 1983 and 1987. Data from 1992 and 1993 were excluded due to errors. Change in operating costs is from year to year and is defined only for plants observed for two consecutive sampled years. Employment is measured in persons.

Note: This example includes the number of observations in the note. The min and max are often excluded in the published version of a paper due to space constraints.

Your first draft should include the number of observations, min, and max, for each variable.