**Abortion rates in the USA**

(Source: Gujarati, D.N., 2011. Econometrics by example.)

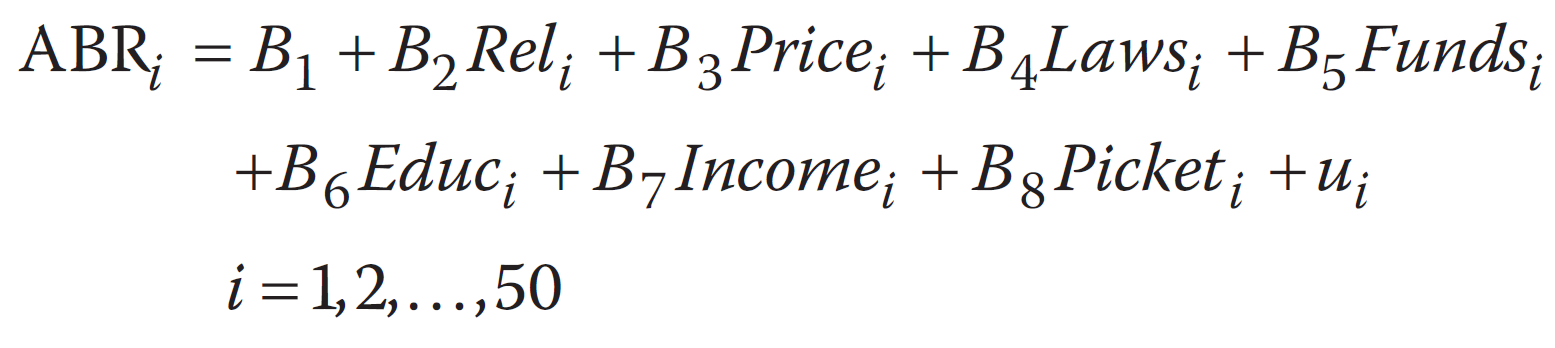
What are the factors that determine the abortion rate across the 50 states in the USA?

**The variables**

* State = name of the state (50 US states).
* ABR=Abortion rate, number of abortions per thousand women aged 15–44 in 1992.
* Religion = the percent of a state’s population that is Catholic, Southern Baptist, Evangelical, or Mormon.
* Price = the average price charged in 1993 in non-hospital facilities for an abortion at 10 weeks with local anesthesia (weighted by the number of abortions performed in 1992).
* Laws = a variable that takes the value of 1 if a state enforces a law that restricts a minor’s access to abortion, 0 otherwise.
* Funds = a variable that takes the value of 1 if state funds are available for use to pay for an abortion under most circumstances, 0 otherwise.
* Educ = the percent of a state’s population that is 25 years or older with a high school degree (or equivalent), 1990.
* Income = disposable income per capita, 1992.
* Picket = the percentage of respondents that reported experiencing picketing with physical contact or blocking of patients.

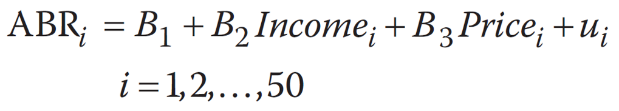
**The model**

As a starting point, we consider the following linear regression model:



A priori, we would expect ABR to be negatively related to religion, price, laws, picket, education, and positively related to fund and income. We assume the error term satisfies the standard classical assumptions, including the assumption of homoscedasticity. Of course, we will do a post-estimation analysis to see if this assumption holds in the present case.

1. Estimate the model and comment on the results
2. Plot a scatter plot of the residuals and a histogram of squared residuals
3. Plot the squared residuals against the estimated abortion rate from the regression model (predicted)
4. To test for heteroscedasticity, we are going to use only Income and Price as independent variables.



* 1. Estimate the above OLS regression and obtain the squared OLS residuals, Ui2.
  2. Regress ui2 on the two regressors included in the model; the squared terms of these regressors, and the pairwise cross-product term of each regressor. The idea here is to see if the squared residuals (a proxy for true squared error term) are related to one or more X variables, their squares and interaction terms. You can choose other regressors also that might have some bearing on the error variance. Now run the regression.
     1. Save R2 from regression (b); call it Raux2, where aux stands for auxiliary, since Eq. (b) is auxiliary to the primary regression (a). The idea behind Eq. (b) is to find out if the squared error term is related to one or more of the regressors, which might indicate that perhaps heteroscedasticity is present in the data.
  3. The null hypothesis here is that the error variance is homoscedastic – that is, all the slope coefficients in Eq. (b) are simultaneously equal to zero. You can use the F statistic from this regression with (k) and (t– k) in the numerator and denominator df, respectively, to test this hypothesis. If the computed F statistic in Eq. (b) is statistically significant, we can reject the hypothesis of homoscedasticity. If it is not, we may not reject the null hypothesis.
  4. Perform White’s Test.