

Solutions: Problem Set #8

(1) The regression output is provided in the other file. Increasing team batting average by, say, 10 points will lead to about 5 more wins per season. Hitting 10 more home runs in a season will lead to about 1.2 more wins in that season. Finally, lowering ERA by 1 point will increase the number of wins by about 17. All of these variables are statistically significant at the 1 percent level.

(1b) Note that when including both an intercept and the FULL set of team dummies, we have a perfect collinearity problem. As a result, one of the coefficient must be dropped. When I performed the regression, STATA dropped the dummy for the rockies.

(1c) A preferable alternative is to simply drop one of the variables yourself rather than have STATA do it for you. In (1c), we drop the intercept and include the full set of team dummies. Note that none of these dummies are individually significant (though jointly, they may be). The point estimate for the yankees, for example, is about 7. For other teams, it is about as large, and for still others, it is negative.

(1d) I should have been a bit more careful about this, but when implementing the test, you need to use the *RSS* version of the test statistic. The reason behind this is that STATA is not smart enough to recognize that even though the intercept has been dropped, it is equivalent to a different model that adds the intercept back in and drops one of the remaining team dummies. If you look carefully at your regression output across the restricted and unrestricted models, the TSS changes when you drop the intercept (which it should not, but STATA reserves something special for “cons” which appears to account for why the TSS changes). Thus, R-squared changes inappropriately and the R-squared version of the test is invalid. (If you implemented this version of the test, however, you should not be penalized).

To implement the test using a method that stata is more familiar with, let's suppose instead that we drop the yankees dummy variable, but keep the intercept in the

model. (As a result, STATA won't freak out). Note, in this case, the intercept in the model is the same as the yankees dummy when dropping the intercept, and the coefficients on all of the other team dummies are interpreted relative to the yankees dummy. The regression model has not changed at all - it is just a matter of coefficient interpretation.

When implementing this test (which we now can do using the R-squareds), we obtain the test statistic

$$\frac{(.8895 - .8594)/29}{(1 - .8895)/(210 - 33)} \approx 1.66.$$

The 5 percent $F_{29,177}$ -statistic is approximately 1.54, leading us to reject the null at the 5 percent level. NOTE: you did not have to calculate the critical value here, since, unfortunately, it was not provided in your book.

2 (a) The point estimate is just $810(.45) = 364.5$. So, in terms of a point prediction, a one dollar increase in the Beer Tax would save about 365 lives. The 95 percent confidence interval is given by

$$364.5 \pm 1.96(810)(.22) = [15.23, 713.8].$$

(b) Based on the results of column 4, if New Jersey lowers its drinking age to 18, the impact on the fatality rate will be .028. For a population of 8.1 million people, the predicted increase in the number of fatalities is

$$810(.028) = 22.7.$$

A 95 percent confidence interval is given by

$$22.7 \pm 810(1.96)(.066) = [-82.1, 127.48].$$

(c) First, note that income is entered in log form, and so a *1 percent increase in real per capita income* results in about 1.81 additional deaths per 10,000 people. Again, applied to a population of 8.1 million people, we obtain the point estimate:

$$810(1.81) = 1466.1,$$

so that the 1 percent increase in real income will lead to an additional 1,466 deaths. The 90 percent confidence interval is:

$$1466.1 \pm 810(1.64)(.47) = [840, 2092].$$

(d) Yes, the time effects should be included. The F tests reject the null that the time effects are zero under each specification of the model.

(e) There are a few different ways to do this. Perhaps the easiest is to create a dummy variable, denoted W_i which indicates if the state is actually a western state. Then, interact this variable with the unemployment rate variable, denoted $unemp_{it}$. Estimate a model of the form:

$$FatalityRate_{it} = \beta_0 + \beta_1 Unemp_{it} + \beta_2 (W_i * Unemp_{it}) + OtherVariables + u_{it}.$$

The “other Variables” will consist of the state dummies, time effects, drinking age, etc. In this specification the marginal effect of unemployment on the fatality rate is $\beta_1 + \beta_2$ if your state is in the west, and β_1 otherwise. A test of $\beta_2 = 0$ will determine if there is variation in response to the unemployment rate in the west. You can conduct this test by simply looking at the t-statistic in the output and corresponding p-value.