Lab 9: Dummy Variables and Data Manipulations

Objectives:
This lab will introduce you to data manipulations in ***Minitab***, especially the creation of ***dummy variables*** and ***interaction variables***. We'll then use one of the most common data transformations in econometrics – creating natural logarithms of variables. You should recall that we did this using Excel in Lab 7.

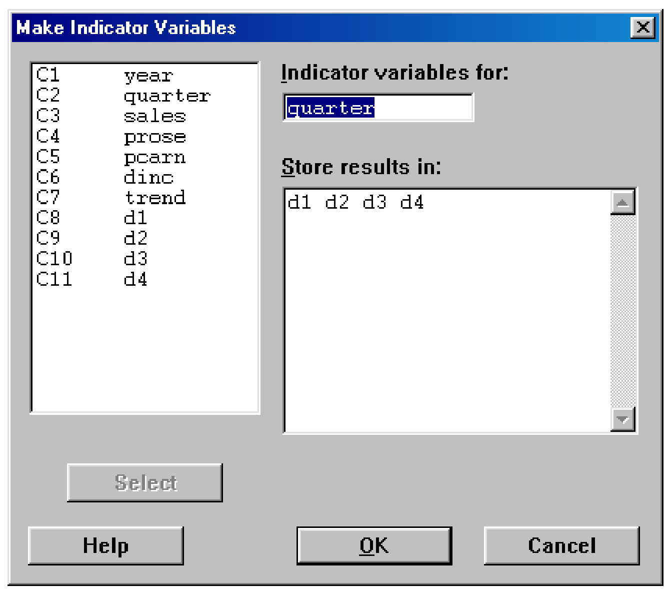
Data: Same data used last week: ***Lab 8 Exercises.xls***.; and ***Roses.dat*** (an ASCII data set). If you saved your Minitab project for Lab 8, you can open Minitab, and then open your Minitab Project.

♦ ***Creating Dummy Variables***.

1. Dummy variables are binary variables (they take on only two values) that indicate whether a condition, characteristic, or quality exists.
2. We can use Minitab to create dummy variables for us for this exercise. First, load the ***Roses*** data. The data are on the server as an ASCII file (we used these in Lab 8). Or, you can copy and paste the data from the Excel spreadsheet. Copy both data and names, but make sure you get the names pasted into the first row in Minitab where the names belong (the cells are shaded beige/gray in the Minitab worksheet).
3. Check the data by **calculating descriptive statistics (means, standard deviations, medians, minimums and maximums) for the variables: *sales, prose, pcarn,*** and ***income***. You should observe the same results that we obtained last week. **Copy these to a *Word* file**.
4. Next we need to create dummy variables for the different quarters. You could simply type the appropriate dummy variables into the worksheet. But there is an easier way: Let Minitab do it! In the menu under **Calc**, you will find **Make Indicator Variables…** When you ask Minitab to ***create indicator variables*** for a variable in your data set, Minitab will check the variable and find out how many different values there are. It only makes sense to create indicator variables for ***categorical variables*** (variables with a limited number of discrete values). The variable ***quarter*** has 4 different values denoting the quarter of the year. Thus, Minitab will create 4 dummy variables, one for each quarter, of the form:

$$d\,j \;=\; \begin{cases} 1 & if\ \ quarter = j\ ; \\ 0 & if\ \ quarter \neq j\ . \end{cases} \qquad for\ \ j = 1,\ 2,\ 3,\ 4\ .$$

Minitab is clever enough to associate ***d1*** with *quarter* = 1, etc. Try it. In the box, **Indicator variables for:**, type the variable name **quarter** (or place the cursor in the box and double-click the variable *quarter*). Then type the names: ***d1 d2 d3 d4***, in the box, **Store results in:** Click **OK** and presto - you have four new dummy variables. My screen capture at the right illustrates what should be typed to create the dummy variables. Check the worksheet after creating the dummy variables. When quarter = 1, d1 = 1, etc.



5. Use ***just one dummy variable*** in a multiple regression model for the demand for roses. **In which quarter would you expect the sales of roses to be highest?** Estimate a regression model that includes the dummy variable as follows:

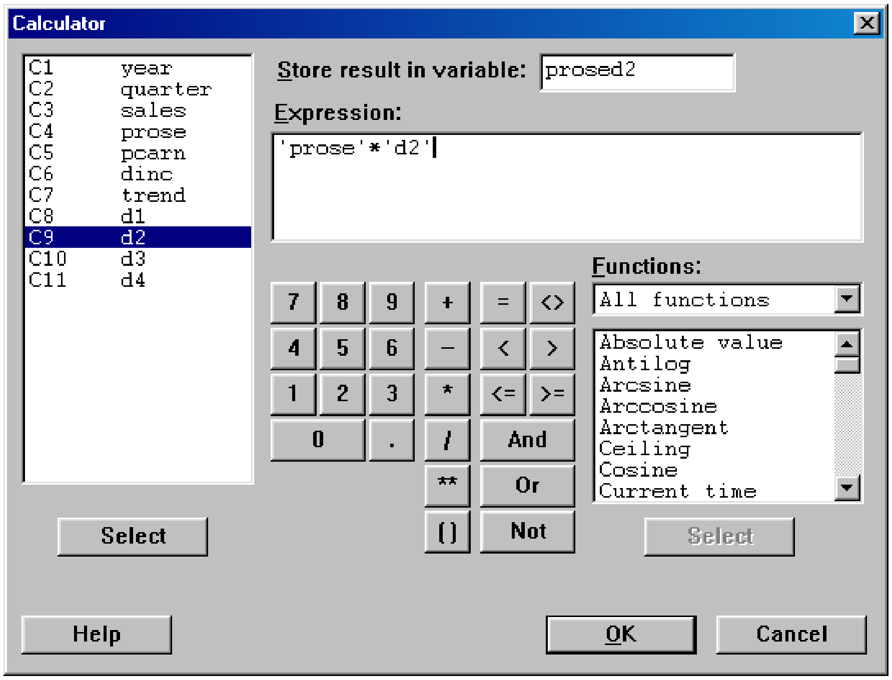
$$Sales_t \;=\; \beta_0 \;+\; \beta_1\, Prose_t \;+\; \beta_2\, Pcarn_t \;+\; \beta_3\, Income_t \;+\; \delta\, d\, j_t \;+\; u_t\,,$$

Copy the results to a Word document, which you will print at the end of the lab. Including a dummy variable "shifts" the intercept for the quarter that you chose. What is the estimated parameter for the dummy variable? That is an estimate of the "shift" in the intercept for that quarter.

♦ ***Using Minitab's Calculator – data manipulations***

1. We use ***interaction variables*** to shift the slope of the regression line for different quarters. For example, you might ask whether the effect of ***Prose*** on ***Sales*** is different in the second quarter. This is the same as asking whether the slope is different in quarter 2 than a common slope for all other quarters, an example of shifting the slope by using dummy variables. A model for the sales of roses with dummy variables included to shift both the intercept and slope is as follows:

$$Sales_t = \beta_0 + \beta_1 Prose_t + \beta_2 Pcarn_t + \beta_3 Income_t + \delta d2_t + \gamma \left(d2_t \cdot Prose_t \right) + u_t$$



To estimate this model, we need to create the interaction term $\left(d2_t \cdot Prose_t \right)$. Again, let Minitab do the work. Choose **Calc** and then **Calculator...** You are presented with boxes for **Store results in variable:** and **Expression.** You need to pick a name for the interaction variable you are about to create (I chose **Prosed2**) and you need to provide the mathematical ***expression***: **Prose * d2** . The screen capture at the right illustrates creating this new variable.

After entering this information for Minitab, just click OK and you will have the appropriate interaction variable in a blink of your eye. To shift both the slope and intercept for quarter 2 in the roses demand model, include ***both the dummy variable and the interaction term in the model as indicated in the PRE above.*** Estimate the multiple regression model specified above and **copy the regression results to your *Word* document**.

2. There are a number of variable transformations that are useful in econometrics. For example, you can use ***Minitab*** to calculate elasticities for every quarter after estimating the demand for roses. To do this, use your ***Minitab*** estimates of the following regression model (we just estimated this model):

$$Sales_t = \beta_0 + \beta_1 Prose_t + \beta_2 Pcarn_t + \beta_3 Dinc_t + \delta d2_t + \gamma Prose d2_t + u_t$$

Copy the estimated sample regression function (highlight the sample regression function from your **Session** window, then hit "***Ctrl c***"), we'll use it in the next step.

3. Choose **Calc** and then **Calculator...** Variables are created by listing new variable names in the **Store results in variable:** box. The equations used to create the new variables are entered in the **Expression:** box. **Create a new variable called *Saleshat*** by pasting the sample regression function you just copied into the **Expression** box. The sample regression function from Minitab is:

$$\hat{Sales}_t = \hat{\beta}_0 + \hat{\beta}_1 Prose_t + \hat{\beta}_2 Pcarn_t + \hat{\beta}_3 Dinc_t + \hat{\delta}\, d2_t + \hat{\gamma}\, \mathrm{Pr}osed2_t \ ,$$

or:

```
sales = 17377 - 3865 prose + 2243 pcarn - 32.3 dinc + 492 d2 + 331 prosed2.
```

Don't forget to *add all the multiplication signs* in the **Expression** box, but the dependent variable name "sales" and the "equal sign" that appear above are NOT entered in the "Expression" box.

4. The formula for the price elasticity of demand for roses is: $\hat{\varepsilon}_p = \left(\frac{\partial \widehat{Sales}}{\partial Prose}\right)\cdot\frac{Prose}{\widehat{Sales}}$

   For the sample regression function estimated here the first part of the elasticity formula, the partial effect, is:

   $$\frac{\partial \widehat{Sales}}{\partial Prose} = \hat{\beta}_1 + \hat{\gamma}\, d2_t = -3865 + 331\cdot d2_t .$$

   You'll need to use the actual numeric values for the estimates as shown above. You should note that we've replaced $\hat{\beta}_1 = -3865$ and $\hat{\gamma} = 331$ in the expression for the partial effect. The dummy variable ***d2**$_i$* will change for different time periods and will affect the estimated partial effect of *Prose* on *Sales*.

   The changing dummy variable ($\boldsymbol{d2_t}$ = 0 or $\boldsymbol{d2_t}$ = 1) will give us two different partial effects for a change in *Prose* depending upon the quarter:

   when $d2_t = 0$: $\hat{\varepsilon}_p = \left(\frac{\partial \widehat{Sales}}{\partial Prose}\right)\cdot\frac{Prose}{\widehat{Sales}} = \hat{\beta}_1\cdot\frac{Prose}{\widehat{Sales}}$; and

   when $d2_t = 1$: $\hat{\varepsilon}_p = \left(\frac{\partial \widehat{Sales}}{\partial Prose}\right)\cdot\frac{Prose}{\widehat{Sales}} = (\hat{\beta}_1 + \hat{\gamma})\cdot\frac{Prose}{\widehat{Sales}}$

   Thus, if the quarter is 1, 3, or 4 the elasticity is calculated using the slope estimate $\hat{\beta}_1 = -3865$. If it is quarter 2, the elasticity is calculated using the slope estimate $\hat{\beta}_1 + \hat{\gamma} = (-3865 + 331) = -3534$.

   To calculate the elasticities for quarter 2 and for the other three quarters we need to choose values for *Sales* and *Prose* at which to calculate the elasticity. We can do two things: (i) calculate the elasticity at the means of the data, or (ii) calculate the elasticity for each time period using the formula:

   $$\hat{\varepsilon}_{p,t} = \left(\frac{\partial \widehat{Sales}}{\partial Prose}\right)\cdot\frac{Prose_t}{\widehat{Sales}_t} = (\hat{\beta}_1 + \hat{\gamma}\cdot d2_t)\cdot\frac{Prose_t}{\widehat{Sales}_t}.$$

5. First, calculate the elasticities at the means of the data. We should use the means for quarter 2, when d2 = 1 and the means for quarters 1,3 and 4, when d2 = 0. Have Minitab calculate the means for quarter 2 and quarters 1, 3 and 4. To do this, return to **Descriptive statistics**, check the box **By variable**, and enter "d2." We only need descriptive statistics for Sales and Prose. Minitab will give you different means for *d2=1* and *d2=0*. **Copy these descriptive statistics to your *Word* document**. You can then calculate elasticities using Excel or your hand calculator, the sample means and your estimates, $\hat{\beta}_1$ *and* $\hat{\gamma}$. We'll leave this calculation to you. Would you expect demand in quarter 2 to be relatively elastic or inelastic when compared to the elasticity for quarters 1, 3 and 4?

6. Next, calculate elasticities for every observation using the **Calculator** and the equation above. Let's all call the elasticity: ***PriceElast***. The dummy variable, *d2*, should appear in your formula and causes ***different slopes and elasticities*** for quarter 2. Once you've calculated elasticities for each year, use Minitab to calculate the mean elasticity for the data set (**Stat**, then **Basic Statistics**, …) and by the variable "d2." **Copy these results to your *Word* document**.

So, does all this make sense? Can you do all this on your own? Let's give it a try.

- ***Create your own Dummy variables.***

Try creating dummy variables on your own. Using **your sample of used 2004 Honda Accords**, create dummy variables for: ***Engine size*** (4 or 6 cylinder engines); ***number of doors*** (4 door vs 2 door cars); and the ***type of transmission***. Because transmission is a character variable, Minitab will assign dummy variable values alphabetically by default. So, if you want the dummy variable AT to represent automatic transmission, you should list AT first in the Minitab box for dummy variable names. I created the following dummy variables for my sample: *AT* – cars with automatic transmissions; *Dr4* – cars with 4 doors; *DX* – a Honda Accord DX model; and *Cyl4* – a car with a 4-cylinder engine. You should be able to accomplish the same.

- ***Create your own Interaction variables.***

Using your Honda data, create an interaction variable or two. What variables might we interact? It is possible that the rate of depreciation is different for 4-cylinder and 6-cylinder engines. Or, maybe the Automatic transmission cars with 4-cylinder engines have lower values because they tend to wear out faster. Create interaction terms for these variables – the latter will be used in a multiple regression model.

- ***Run some regressions.***

Estimate the following regression model for used 2004 Honda Accords:

$$Price_i = \beta_0 + \beta_1 Miles_i + \delta_1 AT_i + \delta_2 Dr4_i + \delta_3 DX_i + \gamma\left(Cyl4_i \cdot AT_i \right) + u_i .$$

Hmmm. I wonder where we could use these results?