V. Extensions of Multiple Regression

- **A. Dummy Variables**
- **B. Scaling Variables**
- **C. Non-Linear Models**
  1. **Basic Idea for Estimation**.
  2. **Models that don't require logs**
  3. **Models that require log transform**.
  4. **Other Items:**
     a. Normality Assumption
     b. Dummy variables in Log Models.

b. Dummy Variables in Log Models.

- Dependent variable is now lnY. Not in the natural units.
- Dummy variable coefficient measures:

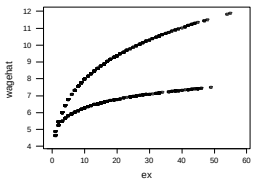
  δ = lnY(D=1) – lnY(D=0)
- Proper interpretation:

  % difference in Y = $(\exp^{\delta} - 1)\cdot 100$

```
lnwage = - 0.921 + 0.988 lned + 0.235 lnex + 0.047 fe
         - 0.124 lnexfe

Predictor        Coef       StDev          T        P
Constant      -0.9212      0.2524      -3.65    0.000
lned          0.98774     0.08731      11.31    0.000
lnex          0.23505     0.03111       7.56    0.000
fe             0.0475      0.1256       0.38    0.706
lnexfe       -0.12383     0.04555      -2.72    0.007

S = 0.4471      R-Sq = 27.4%     R-Sq(adj) = 26.8%
```



V. Extensions of Multiple Regression

A. **Dummy Variables**
B. **Scaling Variables**
C. **Non-Linear Models**

VI. Problems

A. **Multicollinearity**
B. **Heteroskedasticity**
C. **Autocorrelation**

VI. Problems

A. Multicollinearity

1. **Definition**: The presence of ***linear association*** among independent variables.
   - ***Sample Problem*** – the problem lies in your sample data.
   - There is ***no causal relationship*** between X_2 and the other independent variables.

2. **Consequences:**
   - OLS estimators – remain unbiased.
   - Standard errors are inflated.
   - Calculated t-statistics are deflated.
   - t_{calc} falls in the FTR region more often than it should for a false null hypothesis.
   - Can't trust your hypothesis tests.
   - ***P(Type II Error) is high***.

2. Consequences of Multicollinearity

- Multiple regression standard error:

$$s^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2 \sum x^2_{2i}}{\sum x^2_{1i} \sum x^2_{2i} - (\sum x_{1i} x_{2i})^2}$$

- What happens if X_1 and X_2 are strongly and linearly associated. Recall:

$$r_{x_1 x_2} = \frac{\sum x_{1i} x_{2i}}{\sqrt{\sum x^2_{1i}} \sqrt{\sum x^2_{2i}}} \qquad r^2_{x_1 x_2} = \frac{(\sum x_{1i} x_{2i})^2}{\sum x^2_{1i} \sum x^2_{2i}}$$

3. Diagnosis (Multicollinearity)

How can we tell if we have this problem?

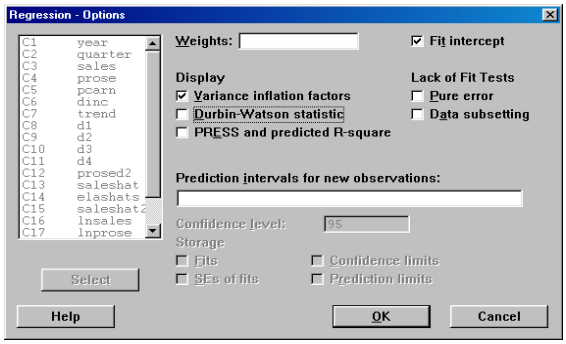
a. **Classic Signs** – look on your printout for the following ***combination***:

3. Diagnosis (Multicollinearity)

b. **Correlation Coefficients** – how strong are the pair-wise correlations between the Xs?

c. **Auxilliary Regressions** – regressions of one X on all the others, i.e.,

d. Variance Inflation Factors: Check the box in Minitab under **Regression Options**



4. Solutions – fixing the problem

- Sample data problem –
- ***Eliminate the offensive variable*** –
- Linear Association –
- ***Use "non-sample" information***.
- ***Data transformations*** –