DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Posc/Uapp 816

INFERENCE FOR REGRESSION

I. AGENDA:
- A. "R-square measure of goodness of fit"
- B. Tests and confidence intervals for regression parameters.
- C. Reading: Agresti and Finlay ***Statistical Methods in the Social Sciences***, 3rd edition, Chapter 9 pages 326 to 333.

II. SUMMARY - ASSESSING GOODNESS OF FIT WITH R^2:
- A. See the notes for Class 10 for definition of R^2
- B. Recall that R^2 can be defined as the portion of the total sum of squares (TSS) "explained" or attributable to the regression model (RegSS).

$$R^2 = RegSS/TSS = \frac{\sum_{i=1}^{N} (\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}$$

- C. Properties:
  1. R^2 varies between 0 and 1.0.
     - i. A value of 0 means no linear correlation.
       1) The variables may be associated; but not linearly.
     - ii. A value of 1.0 suggests "perfect" linear correlation.
       1) The regression constant, β_1 nay be positive or negative.
       2) In Figure 1, for example, R^2 = 1,0 for both sets of data, although one is a negative relationship and the other positive.

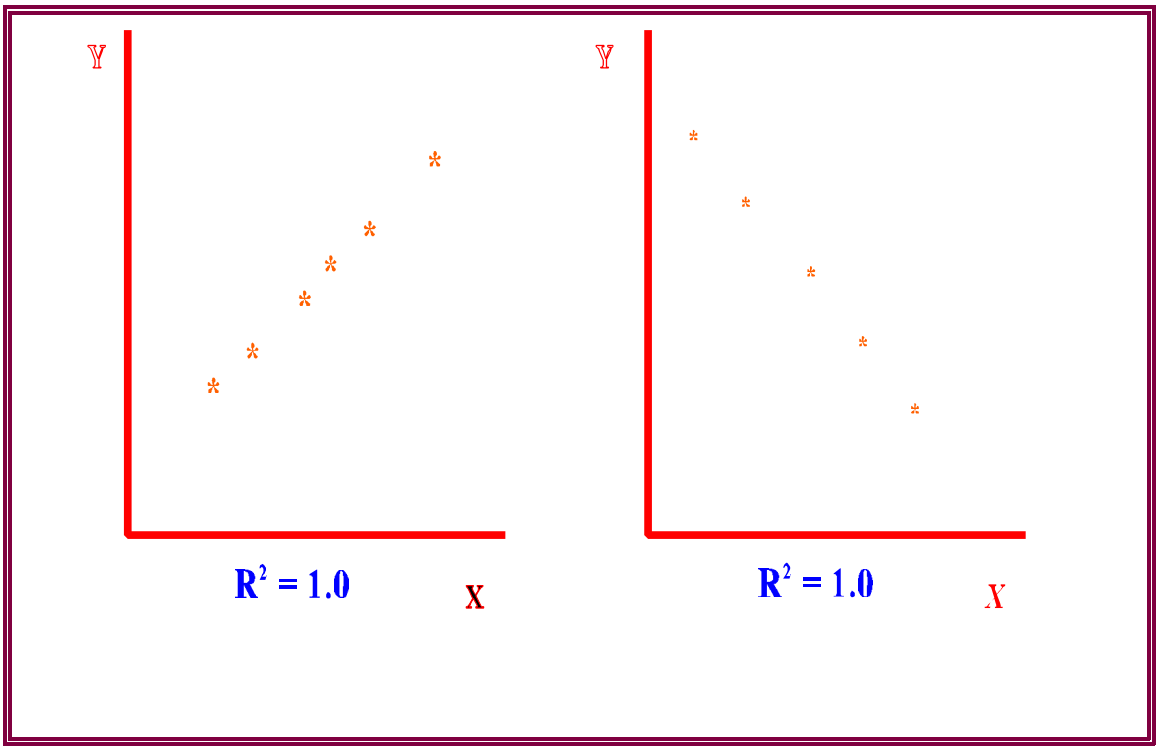

Figure 1: R^2 And Perfect Linearity

D. Although this is a very commonly reported measure, it by itself is not entirely satisfactory. Like OLS in general, it is sensitive to "outlying" values, as is illustrated in Figure 2.

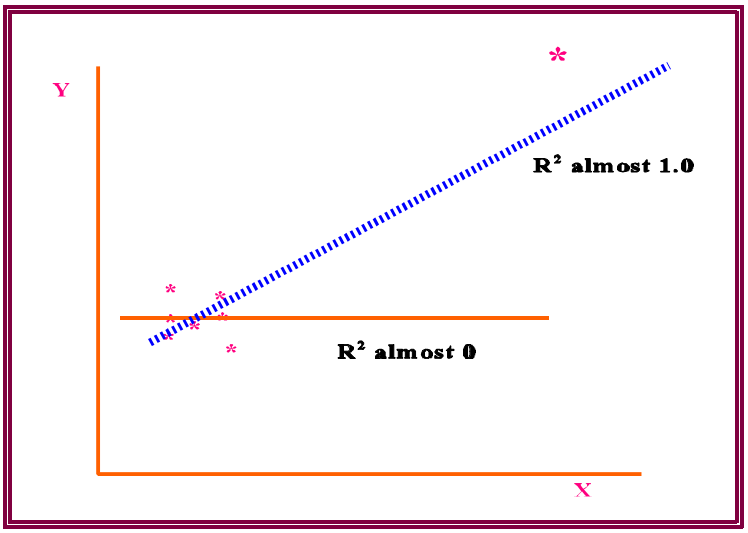

Figure 2: R^2 Can Be Affected by Outliers

1. In this example, the one outlying value will "pull" the least squares line up, thereby creating the impression of a positive relationship. Moreover, R^2 will be large, which is in this case very misleading.
2. Thus, always draw graphs or plots.

E. It is possible, as the definition indicates, to interpret R^2 as the "percent of variation in Y explained by X." But don't take this notion too literally.
1. Hence, values of .8 or .9 suggest a tight "fit" while those closer to 0 suggest a weak (linear) relationship.

F. Example:
1. Here once again is the results from of the mortality and air pollution analysis:

```
Mortality = 919 + 0.412 SO2

59 cases used 1 cases contain missing values

Predictor        Coef       StDev          T        P
Constant      918.671       9.853      93.24    0.000
SO2            0.4117      0.1181       3.49    0.001

S = 57.17       R-Sq = 17.6%     R-Sq(adj) = 16.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       39698       39698     12.15    0.001
Residual Error    57      186295        3268
Total             58      225993
```

2. R^2 = .176 suggests a modest relationship between mortality and air pollution.

III. TEST OF SIGNIFICANCE:

A. Refer to the notes for Class 10. There is no need to repeat them here.

IV. NEXT TIME:

A. Correlation
B. Transformations
C. Multiple regression.

Go to Notes page
Go to Statistics page