

Regression Analysis

Multiple Linear Regression

Nicoleta Serban, Ph.D.

Professor

School of Industrial and Systems Engineering

Model Evaluation and
Multicollinearity



1

About This Lesson



2

1

R²: Coefficient of Determination

A measure that efficiently summarizes how well the Xs can be used to predict Y is R² (called *R-squared* or the *coefficient of determination*):

$$R^2 = 1 - SSE/SST$$

where

$$SSE = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

R² is interpreted as the proportion of total variability in Y that can be explained by the linear regression model.



3

R²: Notation and Terminology

SSE, **SST**, and **SSR** refer to *sum of squared errors*, *sum of squares total*, and *sum of squares for regression*. Unfortunately, the field of statistics abounds in inconsistent terminology and notation.

- **SSE**: sometimes denoted **SS_{error}** or **SS_{err}**, is also known as **RSS** (residual sum of squares) and **SS_{res}** (sum of squared residuals, sometimes **SSR**).
- **SST**: sometimes written as **SS_{total}** or **SS_{tot}**. It is also called total sum of squares and written as **TSS**.
- **SSR**: also called the sum of squares due to regression, and it is sometimes written as **SS_{reg}**. It's also called explained sum of squares (**ESS**). Don't confuse ESS with SSE, and, for R², remember that SSR is SS regression, not SS residuals!



4

Model Evaluation

- **F-test for overall regression**
 - $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$
 - $F_0 = \text{MSR}/\text{MSE} \sim F(p, n-p-1)$
 - $\text{MSR} = \text{SSR}/p$
 - $\text{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
 - $\text{MSE} = \text{SSE}/(n-p-1)$
- **Coefficient of determination**
 - $R^2 = 1 - \text{SSE}/\text{SST} = \text{SSR}/\text{SST}$
 - R^2 increases when additional predictors are added to a model
 - Such increase might not indicate increased explanatory power
- **Adjusted coefficient of determination**
 - Penalizes for more predictors
 - adjusted $R^2 = 1 - (n-1)(1-R^2)/(n-p-1)$



5

Correlation Coefficient

A statistic that efficiently summarizes how well one of the X s is *linearly* related to Y (or to another X) is ρ , the (Pearson) correlation coefficient:

$$\rho_{X_j, Y} = \text{cor}(X_j, Y) = \frac{\sum_{i=1}^n (x_{i,j} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{i,j} - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- Can be used to evaluate the *linear* relationship between the response variable and any of the predicting variables, X_j
 - Useful when looking for transformations of predicting variables
- Can also be used to evaluate correlation between predicting variables
 - Can help detect near linear dependence (multicollinearity)



6

Multicollinearity

Recall that finding the ordinary least squares estimator of $\hat{\beta}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

depends on $X^T X$ being invertible (nonsingular or nondegenerate). From linear algebra, a square matrix is invertible if and only if its columns are linearly independent (i.e., no column is a linear combination of the others).

If that doesn't hold, the ordinary least squares estimator of $\hat{\beta}$ doesn't exist. That's probably due to a specification error where one or more predictors should be eliminated as redundant (e.g., if years and number of rings were included in a model for trees).

Even if the columns of $X^T X$ are linearly independent, some problems might arise if the value of one predictor can be closely estimated from the other predictors. We call this condition *multicollinearity* or *near collinearity*.



7

Multicollinearity

- Indications that near collinearity is present:
 - The estimated coefficients $\hat{\beta}$ are unstable: When the value of one predictor changes slightly, the fitted regression coefficients change dramatically
 - The standard error of $\hat{\beta}$ is artificially large
 - The overall F statistic is significant, but individual t-statistics are not
- Prediction may be affected
 - The relationship to the response may change widely
- Some computational algorithms are sensitive to multicollinearity
- But no inflation or deflation in R^2



8

Multicollinearity Diagnosis

Compute the variance inflation factor (VIF_j) for each predicting variable X_j

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the coefficient of determination for the regression of X_j against all other predicting variables.

What is an acceptable VIF (i.e., multicollinearity is not problematic)?

- $VIF < \max(10, 1/(1 - R_{model}^2))$
 - R_{model}^2 is the coefficient of determination for the original model
 - Rule of thumb only



9

Multicollinearity Diagnosis

Steps:

1. For $j = 1, \dots, p$, regress X_j against all other X_i , $i = 1, \dots, p$, $i \neq j$ (i.e., $X_j \sim X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p$).
2. For each regression run, compute R^2 for that regression (i.e., compute R_j^2 for $j = 1, \dots, p$).
3. For each regression run, compute VIF_j based on the computed R_j^2 for that regression.
4. If any $VIF_j \geq 10$ and it is also $\geq 1/(1 - R_{model}^2)$, where R_{model}^2 is the coefficient of determination for the original model, the test is positive for multicollinearity.

High multicollinearity is not detected if each $VIF_j < \max(10, \frac{1}{1 - R_{model}^2})$.



10

Multicollinearity Interpretation

VIF measures the proportional increase in the variance of $\hat{\beta}_i$ compared to what it would have been if the predicting variables had been completely uncorrelated.

- VIF of 1 (the minimum possible VIF) means the tested predictor is not correlated with the other predictors
- The higher the VIF:
 - The more correlated a predictor is with the other predictors
 - The more the standard error is inflated
 - The larger the confidence interval
 - The less likely it is that a coefficient will be evaluated as statistically significant



11

Summary



12