

## Chapter 3: Learning to Use Regression Analysis

# 1 Steps in Applied Regression Analysis

- Review the literature and develop the theoretical model.
- Specify the model: Select the independent variables and the functional form.
- Hypothesize the expected signs of coefficients.
- Collect the data. Inspect and clean the data.
- Estimate and evaluate the equation.
- Document the results.

## 1.1 Step 1: Review the Literature and Develop the Theoretical Model

One useful resource is Google Scholar: <http://scholar.google.com/>. For example, searching for "restaurant location regression" result in:

Fast food, race/ethnicity, and income A geographic analysis - Find it with OLinks - all 4 versions ■ JP Block, RA Scribner, KB DeSalvo - American Journal of Preventive Medicine, 2004 - Elsevier ... Using multiple regression in SPSS (Graduate Pack 10.0 for ... Locations of alcohol outlets were available in ... of highways may dictate fast food restaurant location. ... Cited by 51 - Related Articles - Web Search

Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low- ■ - Find it with OLinks - all 5 versions ■ HL Burdette,

RC Whitaker - Preventive Medicine, 2004 - Elsevier ... However, using multivariable logistic regression models, the ... linked to that address might not ... that altering fast food restaurant locations without decreasing ... Cited by 44 - Related Articles - Web Search

Pay attention to how many citations a paper gets. The more it is cited, the more influential it is.

## **1.2 Step 2: Specify the Model: Select the Independent Variables and the Functional Form**

The *specification* of a model involves the following components:

- the selection of the dependent variable
- the independent variables and how they should be measured,
- the functional (mathematical) form of the variables, and
- the type of stochastic error term.

### **1.3 Step 3: Hypothesize the Expected Signs of the Coefficients**

The quantity demanded for a consumption good, say coffee, ( $Q_d$ ) is expected to be inversely related to its price ( $P$ ) and the price of a complementary good, say cream, ( $P_c$ ), and positively related to consumer income ( $Y$ ) and the price of a substitute good, say tea, ( $P_s$ ). We express this as a general function as in Equation (3.2) of Studenmund.

### **1.4 Step 4: Collect the Data. Inspect and Clean the Data.**

Pay attention to *units of measurement*. For example, an exchange rate is the price of one currency in terms of another currency. Dollar per euro

was 1.424, and dollar per yen was 0.00864 on October 1, 2007. If you use data for exchange rates for a regression, a computer file may round dollar all per yen data to 0.

In order avoid unexpected problems, we should inspect the data, obtain a printout, and plot (graph) of the data. We should look for outliers. An *outlier* is an observation that lies outside the range of the rest of the observations. Outliers might be a result of data entry errors.

## **1.5 Step 5: Estimate and Evaluate the Equation**

- The overall fit of the estimated equation

- Signs of the estimated coefficients
- Magnitudes of the estimated coefficients

### Step 6: Document the Results

A standard format we adopt in this course is

$$\hat{Y}_i = 103.40 + 6.38 X_i$$

(0.88)

$t = 7.22$

$N = 20, R^2 = 0.74$

For this class format, report  $R^2$  and do not report  $\bar{R}^2$ . We will discuss about the standard error in the parentheses and the  $t$ -value below it in Chapter 5.

Using Regression Analysis to Pick Restaurant Locations

See Section 3.2 of Studenmund.