

# Remaking A Simple Regression Analysis in LaTeX

Tim McGinley

October 31, 2016

## 1 Abstract

The goal of this homework assignment is to reproduce the primary findings (regression, charts, and tables) on pages 59-71 of **An Introduction to Statistical Learning** by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.

## 2 Introduction

Every time a company produces something new, it has a tough decision to make. See, a new product is useless sitting on the shelves, and the only way people will buy it is if they hear about it through word of mouth or, more commonly, advertising. So the question is, how much should a company invest in the advertising for one of their products? This report compares the sales figures for various markets for a single product to their market's TV advertising budget and searches for a connection.

## 3 Data

The data, located in a file called `data/Advertising.csv` contains information about the television budgets, sales figures, and a few unused variables. The entries for the television budgets are in thousands of dollars, while the figures for sales are in thousands of units sold. Overall, 200 markets are represented, and all 200 have extant information for the television budget and sales.

## 4 Methodology

To conduct the analysis, first the data was loaded from `Advertising.csv` and the variables of interest ("TV" for TV ad budget and "Sales" for units sold) were summarized. Summaries can be found in 'hw02/data/eda-output.txt', and histograms of both variables can be found in 'hw02/images/'.

To better understand how these two variables might relate to one another, I performed a simple linear regression analysis with the data. The general idea is that we imagine changes one variableâ in this case TV ad budgetsâ affects the beget changes in another variable, here sales figures. These are called the independent and dependent variables, respectively.

In a simple linear regression, we assume that this relationship can be described roughly by the equation  $S = \beta_0 + \beta_1(TV)$ , where  $\beta_0$  and  $\beta_1$  are random variables determined from the data by a least-squares fit. The function `lm()` determines these variables from the data for us, and this information as well as a summary are located in `data/regression.Rdata`. Scatterplots were constructed using `plot()` and `abline()`.

## 5 Results

Our linear regression analysis produced the results that appear in the following table. The entries in the table and the paragraph thereafter are inline code and will autogenerate if a different `Advertising.csv` is chosen.