Multiple Linear and 1D Regression

David J. Olive
Southern Illinois University
Department of Mathematics
Mailcode 4408
Carbondale, IL 62901-4408
dolive@math.siu.edu

January 4, 2010

Contents

Preface

Regression is the study of the conditional distribution $Y|\boldsymbol{x}$ of the response Y given the $p \times 1$ vector of nontrivial predictors $\boldsymbol{x}$. In a **1D regression model**, Y is conditionally independent of $\boldsymbol{x}$ given a single linear combination $\alpha + \boldsymbol{\beta}^T\boldsymbol{x}$ of the predictors, written

$$Y \perp\!\!\!\perp \boldsymbol{x} | (\alpha + \boldsymbol{\beta}^T\boldsymbol{x}) \text{ or } \mathrm{Y} \perp\!\!\!\perp \boldsymbol{x} | \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}.$$

Many of the most used statistical methods are 1D models, including generalized linear models such as multiple linear regression, logistic regression, and Poisson regression. Single index models, response transformation models and many survival regression models are also included. The class of 1D models offers a unifying framework for these models, and the models can be presented compactly by defining the population model in terms of the sufficient predictor $\mathrm{SP} = \alpha + \boldsymbol{\beta}^T\boldsymbol{x}$ and the estimated model in terms of the estimated sufficient predictor $\mathbf{ESP} = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T\boldsymbol{x}$. In particular, the **response plot** or estimated sufficient summary plot of the ESP versus Y is used to visualize the conditional distribution $Y|(\alpha + \boldsymbol{\beta}^T\boldsymbol{x})$. The residual plot of the ESP versus the residuals is used to visualize the conditional distribution of the residuals given the ESP. The goal of this text is to present the applications of these models in a manner that is accessible to undergraduate and beginning graduate students.

Response plots are heavily used in this text. With the response plot the presentation for the $p > 1$ case is about the same as the $p = 1$ case. Hence the text immediately covers models with $p \geq 1$, rather than spending 100 pages on the $p = 1$ case and then covering multiple regression models with $p \geq 2$.

The literature on multiple linear regression is enormous. See Stigler (1986) and Harter (1974ab, 1975abc, 1976) for history. Draper (2002) is

a good source for more recent literature. Some texts that were "standard" at one time include Wright (1884), Johnson (1892), Comstock (1895), Bartlett (1900), Merriman (1910), Weld (1916), Leland (1921), Ezekial (1930), Bennett and Franklin (1954), Ezekial and Fox (1959) and Brownlee (1965).

Draper and Smith (1966) was a breakthrough because it popularized the use of residual plots, making the earlier texts obsolete. Excellent texts include Chatterjee and Price (1977), Draper and Smith (1998), Fox (2008), Hamilton (1992), Kutner, Nachtsheim, Neter and Li (2005), Montgomery, Peck and Vining (2006), Mosteller and Tukey (1977), Ryan (2009), Sheather (2009) and Weisberg (2005). Cook and Weisberg (1999a) was a breakthrough because of its use of response plots.

Other texts of interest include Abraham and Ledolter (2006), Harrell (2006), Pardoe (2006), Mickey, Dunn and Clark (2004), Cohen, Cohen, West and Aiken (2003), Kleinbaum, Kupper, Muller and Nizam (1997), Mendenhall and Sinich (2003), Vittinghoff, Glidden, Shiblski and McCulloch (2005) and Berk (2003).

The author's hope is that this text's use of the response plot will make other regression texts obsolete much as Draper and Smith (1966) made earlier texts obsolete by using residual plots. The response plot is much more important than a residual plot since 1D regression is the study of the $Y|(\alpha+\boldsymbol{\beta}^T\boldsymbol{x})$, and the response plot is used to visualize this conditional distribution. The response plot emphasizes model goodness of fit and can be used to complement or even replace goodness of fit tests, while the residual plot of the ESP versus the residuals emphasizes model lack of fit. In this text the response plot is used to explain multiple linear regression, logistic regression, Poisson regression, single index models and models for experimental design. The response plot can also be used to explain and complement the ANOVA F and deviance tests for $\boldsymbol{\beta} = \mathbf{0}$.

This text provides an introduction to several of the most used 1D regression models. Chapter 1 reviews the material to be covered in the text and can be skimmed and then referred to as needed. Concepts such as interpretation of coefficients and interactions, goodness and lack of fit diagnostics, and variable selection are all presented in terms of the SP and ESP. The next few chapters present the multiple linear regression model. Then the one and two way ANOVA, logistic and Poisson regression models are easy to learn. Generalized linear models, single index models and general 1D models are

also presented. Several important survival regression models are 1D models, but the sliced survival plot is used instead of the response plot to visualize the model.

The text also uses recent literature to provide answers to the following important questions.

- How can the conditional distribution $Y|(\alpha + \boldsymbol{\beta}^T\boldsymbol{x})$ be visualized?
- How can α and $\boldsymbol{\beta}$ be estimated?
- How can variable selection be performed efficiently?
- How can Y be predicted?
- What happens if a parametric 1D model is unknown or misspecified?

The author's research on 1D regression models includes visualizing the models, outlier detection, and extending least squares software, originally meant for multiple linear regression, to 1D models. Some of the applications in this text using this research are listed below.

- It is shown how to use the response plot to detect outliers and to assess the adequacy of linear models for multiple linear regression and experimental design.
- It is shown how to use the response plot to detect outliers and to assess the adequacy of very general regression models of the form $Y = m(\boldsymbol{x}) + e$.
- A graphical method for selecting a response transformation for linear models is given. Linear models include multiple linear regression and many experimental design models.
- A graphical method for assessing variable selection for the multiple linear regression model is described. It is shown that for submodels I with k predictors, the widely used screen $C_p(I) \leq k$ is too narrow. More good submodels are considered if the screen $C_p(I) \leq \min(2k, p)$ is used.

- Fast methods of variable selection for multiple linear regression, including an all subsets method, are extended to the 1D regression model. Plots for comparing a submodel with the full model after performing variable selection are also given.

- It is shown that least squares partial F tests, originally meant for multiple linear regression, are useful for exploratory purposes for a much larger class of 1D regression models.

- Asymptotically optimal prediction intervals for a future response Y_f are given for general regression models of the form $Y = m(\boldsymbol{x}) + e$ where the errors are iid, unimodal and independent of $\boldsymbol{x}$.

- Rules of thumb for selecting predictor transformations are given.

- The DD plot is a graphical diagnostic for whether the predictor distribution is multivariate normal or from some other elliptically contoured distribution. The DD plot is also useful for detecting outliers in the predictors.

- Graphical aids, including plots for overdispersion, for binomial regression models such as logistic regression are given.

- Graphical aids, including plots for overdispersion, for Poisson regression models such as loglinear regression are given.

- Graphical aids for survival regression models, including the Cox proportional hazards regression model and Weibull regression model, are given.

- Throughout the book there are goodness of fit and lack of fit plots for examining the model. The response plot is especially important.

The website (www.math.siu.edu/olive/regbk.htm) for this book provides 28 data sets for *Arc*, and 40 *R/Splus* programs in the file *regpack.txt*. The students should save the data and program files on a disk. Chapter 17 discusses how to get the data sets and programs into the software, but the commands below will work for *R/Splus*.

Downloading the book's R/Splus functions *regpack.txt* into *R* or *Splus*:

Download *regpack.txt* onto a disk. Enter *R* and wait for the cursor to appear. Then go to the *File* menu and drag down *Source R Code*. A window should appear. Navigate the *Look in* box until it says *3 1/2 Floppy(A:)*. In the *Files of type* box choose *All files(*.*)* and then select *regpack.txt*. The following line should appear in the main *R* window.

```
> source("A:/regpack.txt")
```

If you use *Splus*, the command

```
> source("A:/regpack.txt")
```

will enter the functions into *Splus*. Creating a special workspace for the functions may be useful.

Type *ls()*. The *R/Splus* functions from *regpack.txt* should appear. In *R*, enter the command *q()*. A window asking "*Save workspace image?*" will appear. Click on *No* to remove the functions from the computer (clicking on *Yes* saves the functions on *R*, but you have the functions on your disk).

Similarly, to download the text's *R/Splus* data sets, save *regdata.txt* on a disk and use the following command.

```
 > source("A:/regdata.txt")
```

This text is an introduction to 1D regression models for undergraduates and beginning graduate students, and the prerequisites for this text are linear algebra and a calculus based course in statistics at the level of Hogg and Craig (1995), Hogg and Tanis (2005), Rice (2006), Wackerly, Mendenhall and Scheaffer (2008), or Walpole, Myers, Myers and Ye (2002). The student should be familiar with vectors, matrices, confidence intervals, expectation, variance, the normal distribution and hypothesis testing. This text may not be easy reading for nonmathematical students. Lindsey (2004) and Bowerman and O'Connell (1990) attempt to present regression models to students who have not had calculus or linear algebra. Also see Kachigan (1982, ch. 3–5) and Allison (1999).

This text will help prepare the student for the following courses.

1) Categorical data analysis: Agresti (2002, 2007) and Simonoff (2003).

2) Econometrics: see Greene (2007), Judge, Griffiths, Hill, Lütkepohl and Lee (1985), Kennedy (2008), and Woolridge (2008).

3) Experimental design: see Box, Hunter and Hunter (2005), Cobb (1998), Kirk (1982), Kuehl (1994), Ledolter and Swersey (2007), Maxwell and Delaney(2003), Montgomery (2005) and Oehlert (2000).

4) Exploratory data analysis: this text could be used for a course in exploratory data analysis, but also see Chambers, Cleveland, Kleiner and Tukey (1983) and Tukey (1977).

5) Generalized linear models: this text could be used for a course in generalized linear models, but also see Dobson and Barnett (2008), Fahrmeir and Tutz (2001), Hoffmann (2004), McCullagh and Nelder (1989) and Myers, Montgomery and Vining (2002).

6) Large sample theory for linear and econometric models: see White (1984).

7) Least squares signal processing: see Porat (1993).

8) Linear models: see Christensen (2002), Graybill (2000), Rao (1973), Ravishanker and Dey (2002), Scheffé (1959), Searle (1971) and Seber and Lee (2003).

9) Logistic regression: see Collett (2003) or Hosmer and Lemeshow (2000).

10) Poisson regression: see Cameron and Trivedi (1998) or Winkelmann (2008).

11) Numerical linear algebra: see Gentle (1998), Datta (1995), Golub and Van Loan (1989) or Trefethen and Bau (1997).

12) Regression graphics: see Cook (1998) and Li (2000).

13) Robust statistics: see Olive (2009a).

14) Survival Analysis: see Klein and Moeschberger (2003), Allison (1995), Collett (2003), or Hosmer, Lemeshow and May (2008).

15) Time Series: see Brockwell and Davis (2002), Chatfield (2003), Cryer and Chan (2008) and Shumway and Stoffer (2006).

This text does not give much history of regression, but it should be noted that many of the most important ideas in statistics are due to Fisher, Neyman, E.S. Pearson and K. Pearson. For example, David (2006-7) says that the following terms were due to Fisher: analysis of variance, confounding, consistency, covariance, degrees of freedom, efficiency, factorial design, information, information matrix, interaction, level of significance, likelihood,

location, maximum likelihood, null hypothesis, pivotal quantity, randomization, randomized blocks, sampling distribution, scale, statistic, Student's t, test of significance and variance.

David (2006-7) says that terms due to Neyman and E.S. Pearson include alternative hypothesis, composite hypothesis, likelihood ratio, power, power function, simple hypothesis, size of critical region, test criterion, test of hypotheses, type I and type II errors. Neyman also coined the term confidence interval.

David (2006-7) says that terms due to K. Pearson include bivariate normal, goodness of fit, multiple regression, nonlinear regression, random sampling, skewness, standard deviation, and weighted least squares.

Acknowledgements

This work has been partially supported by NSF grant DMS 0202922 and DMS 0600933. Collaborations with Douglas M. Hawkins and R. Dennis Cook were extremely valuable. I am very grateful to the developers of useful mathematical and statistical techniques and to the developers of computer software and hardware. Cook (1998) and Cook and Weisberg (1999a) influenced this book. Teaching material from this text has been invaluable. Some of the material in this text has been used in two Math 484 multiple linear regression and experimental design courses, two Math 485 categorical data courses, a Math 473 survival analysis course, a Math 583 regression graphics course, a Math 583 experimental design course and a Math 583 robust statistics course. Chapters 1 to 9 were used in a Fall 2009 Math 484 course.

Chapter 1

Introduction

All models are wrong, but some are useful.
Box (1979)

This chapter provides a preview of the book but is presented in a rather abstract setting and will be much easier to follow after the reading the rest of the book. The reader may omit this chapter on first reading and refer back to it as necessary.

In *data analysis*, an investigator is presented with a *problem* and *data* from some *population*. The population might be the collection of all possible outcomes from an experiment while the problem might be predicting a future value of the response variable Y or summarizing the relationship between Y and the $p \times 1$ vector of predictor variables $\boldsymbol{x}$. A **statistical model** is used to provide a useful approximation to some of the important underlying characteristics of the population which generated the data. Many of the most used models for 1D regression, defined below, are families of conditional distributions $Y|\boldsymbol{x} = \boldsymbol{x}_o$ indexed by $\boldsymbol{x} = \boldsymbol{x}_o$. A 1D regression model is a *parametric model* if the conditional distribution is completely specified except for a fixed finite number of parameters, otherwise, the 1D model is a *semiparametric model*.

Definition 1.1. *Regression* investigates how the response variable Y changes with the value of a $p \times 1$ vector $\boldsymbol{x}$ of nontrivial predictors. Often this *conditional distribution* $Y|\boldsymbol{x}$ is described by a *1D regression model*, where Y is conditionally independent of $\boldsymbol{x}$ given $\boldsymbol{\beta}^T\boldsymbol{x}$, written

$$Y \perp\!\!\!\perp \boldsymbol{x} | \boldsymbol{\beta}^T\boldsymbol{x} \quad \text{or} \quad \mathrm{Y} \perp\!\!\!\perp \boldsymbol{x} | (\alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}). \tag{1.1}$$

This class of models is very rich. Generalized linear models (GLMs) are a special case of 1D regression, and an important class of parametric or semiparametric 1D regression models has the form

$$Y_i = g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i, e_i) \tag{1.2}$$

for $i = 1, ..., n$ where g is a bivariate function, $\boldsymbol{\beta}$ is a $p \times 1$ unknown vector of parameters, and e_i is a random error. Often the errors $e_1, ..., e_n$ are **iid** (independent and identically distributed) from a distribution that is known except for a scale parameter. For example, the e_i's might be iid from a normal (Gaussian) distribution with *mean* 0 and unknown *standard deviation* σ. For this Gaussian model, estimation of α, $\boldsymbol{\beta}$ and σ is important for inference and for predicting a new value of the response variable Y_f given a new vector of predictors $\boldsymbol{x}_f$.

Notation. Often the index i will be suppressed. For example, model (1.2) could be written as $Y = g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}, e)$. More accurately, $Y|\boldsymbol{x} = g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}, e)$, but the conditioning on $\boldsymbol{x}$ will often be suppressed.

Many of the most used statistical models are 1D regression models. A *single index model* with additive error uses $g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}, e) = m(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) + e$, and thus

$$Y = m(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) + e. \tag{1.3}$$

An important special case is *multiple linear regression*

$$Y = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e \tag{1.4}$$

where m is the identity function. The *response transformation model* uses

$$g(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}, e) = t^{-1}(\alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e) \tag{1.5}$$

where t^{-1} is a one to one (typically monotone) function. Hence

$$t(Y) = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e. \tag{1.6}$$

Several important *survival models* have this form. In a *1D binary regression model*, the $Y|\boldsymbol{x}$ are independent Bernoulli$[\rho(\alpha + \boldsymbol{\beta}^T \boldsymbol{x})]$ random variables where

$$P(Y = 1|\boldsymbol{x}) \equiv \rho(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) = 1 - P(Y = 0|\boldsymbol{x}) \tag{1.7}$$

In particular, the *logistic regression model* uses

$$\rho(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) = \frac{\exp(\alpha + \boldsymbol{\beta}^T \boldsymbol{x})}{1 + \exp(\alpha + \boldsymbol{\beta}^T \boldsymbol{x})}.$$

In a *1D Poisson regression model*, the $Y|\boldsymbol{x}$ are independent

$$\text{Poisson}[\mu(\alpha + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x})]$$

random variables. In particular, the *loglinear regression model* uses

$$\mu(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}) = \exp(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}). \tag{1.8}$$

In the literature, the response variable is sometimes called the dependent variable while the predictor variables are sometimes called carriers, covariates, explanatory variables, or independent variables. The ith *case* $(Y_i, \boldsymbol{x}_i^T)$ consists of the values of the response variable Y_i and the predictor variables $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,p})$ where p is the number of predictors and $i = 1, ..., n$. The *sample size* n is the number of cases.

Box (1979) warns that "all models are wrong, but some are useful." For example the function g or the error distribution could be misspecified. *Diagnostics* are used to check whether model assumptions such as the form of g and the proposed error distribution are reasonable. Often diagnostics use *residuals* r_i. If m is known, then the single index model (1.3) uses

$$r_i = Y_i - m(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i)$$

where $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ is an estimate of $(\alpha, \boldsymbol{\beta})$.

Exploratory data analysis (EDA) can be used to find useful models when the form of the regression or multivariate model is unknown. For example, suppose g is a monotone function t^{-1} :

$$Y = t^{-1}(\alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e). \tag{1.9}$$

Then the transformation

$$Z = t(Y) = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + e \tag{1.10}$$

follows a multiple linear regression model.

Definition 1.2: If the 1D model (1.1) holds, then $Y \perp\!\!\!\perp \boldsymbol{x} | (a + c\boldsymbol{\beta}^T\boldsymbol{x})$ for any constants a and $c \neq 0$. The quantity $a + c\boldsymbol{\beta}^T\boldsymbol{x}$ is called a *sufficient predictor* (SP), and a sufficient summary plot is a plot of any SP versus Y. An *estimated sufficient predictor* (**ESP**) is $\tilde{\alpha} + \tilde{\boldsymbol{\beta}}^T\boldsymbol{x}$ where $\tilde{\boldsymbol{\beta}}$ is an estimator of $c\boldsymbol{\beta}$ for some nonzero constant c. An *estimated sufficient summary plot* (ESSP) or **response plot** is a plot of any ESP versus Y.

Assume that the data has been collected and that a 1D regression model (1.1) has been fitted. Suppose that the *sufficient predictor*

$$SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} = \alpha + \boldsymbol{\beta}_R^T\boldsymbol{x}_R + \boldsymbol{\beta}_O^T\boldsymbol{x}_O \tag{1.11}$$

where the $r \times 1$ vector $\boldsymbol{x}_R$ consists of the nontrivial predictors in the *reduced model*. Then the investigator will often want to check whether the model is useful and to perform inference. Several things to consider are listed below.

i) Use the response plot (and/or the sufficient summary plot) to explain the 1D regression model to consulting clients, students or researchers.

ii) Goodness of fit: use the response plot to show that the model provides a simple, useful approximation for the relationship between the response variable Y and the nontrivial predictors $\boldsymbol{x}$. The response plot is used to visualize the conditional distribution of $Y|(\alpha + \boldsymbol{\beta}^T\boldsymbol{x})$ when the 1D regression model holds.

iii) Check for lack of fit of the model (eg with a residual plot of the ESP versus the residuals).

iv) Check whether Y is independent of $\boldsymbol{x}$ by testing $H_o : \boldsymbol{\beta} = \mathbf{0}$, that is, check whether the nontrivial predictors $\boldsymbol{x}$ are needed in the model.

v) Test $H_o : \boldsymbol{\beta}_O = \mathbf{0}$, that is, check whether the reduced model can be used instead of the full model.

vi) Use variable selection to find a good submodel.

vii) Estimate the mean function $E(Y_i|\boldsymbol{x}_i) = \mu(\boldsymbol{x}_i) = d_i\tau(\boldsymbol{x}_i)$ or estimate $\tau(\boldsymbol{x}_i)$ where the d_i are known constants.

viii) Predict Y_i given $\boldsymbol{x}_i$.

The field of statistics known as *regression graphics* gives useful results for examining the 1D regression model (1.1) even when it is unknown or

misspecified. The following sections show that the sufficient summary plot is useful for explaining the given 1D model while the response plot can often be used to visualize the conditional distribution of $Y|(\alpha + \boldsymbol{\beta}^T\boldsymbol{x})$. If there is only one predictor x, then the plot of x versus Y is both a sufficient summary plot and a response plot, but generally $\boldsymbol{\beta}$ is unknown and only a response plot can be made. In Definition 1.2, since $\tilde{\alpha}$ can be any constant, $\tilde{\alpha} = 0$ is often used.

1.1 Multiple Linear Regression

Suppose that the response variable Y is quantitative and that at least one predictor variable x_i is quantitative. Then the multiple linear regression (MLR) model is often a very useful model. For the MLR model,

$$Y_i = \alpha + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i = \alpha + \boldsymbol{\beta}^T\boldsymbol{x}_i + e_i \quad (1.12)$$

for $i = 1, \ldots, n$. Here Y_i is the response variable, $\boldsymbol{x}_i$ is a $p \times 1$ vector of nontrivial predictors, α is an unknown constant, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and e_i is a random variable called the error.

The Gaussian or normal MLR model makes the additional assumption that the errors e_i are iid $N(0, \sigma^2)$ random variables. This model can also be written as $Y = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$ where $e \sim N(0, \sigma^2)$, or $Y|\boldsymbol{x} \sim N(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}, \sigma^2)$ or $Y|\boldsymbol{x} \sim N(SP, \sigma^2)$. The normal MLR model is a parametric model since, given $\boldsymbol{x}$, the family of conditional distributions is completely specified by the parameters α, $\boldsymbol{\beta}$ and σ^2. Since $Y|SP \sim N(SP, \sigma^2)$, the conditional mean function $E(Y|SP) \equiv M(SP) = \mu(SP) = SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x}$. The MLR model is discussed in detail in Chapters 2, 3 and 4.

A sufficient summary plot (SSP) of the sufficient predictor $SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x}_i$ versus the response variable Y_i with the mean function added as a visual aid can be useful for describing the multiple linear regression model. This plot can not be used for real data since α and $\boldsymbol{\beta}$ are unknown. To make Figure 1.1, the artificial data used $n = 100$ cases with $k = 5$ nontrivial predictors. The data used $\alpha = -1$, $\boldsymbol{\beta} = (1, 2, 3, 0, 0)^T$, $e_i \sim N(0, 1)$ and $\boldsymbol{x}$ from a multivariate normal distribution $\boldsymbol{x} \sim N_5(\boldsymbol{0}, \boldsymbol{I})$.

In Figure 1.1, notice that the *identity line* with unit slope and zero intercept corresponds to the mean function since the identity line is the line $Y = SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} = \mu(SP) = E(Y|SP)$. The vertical deviation of Y_i

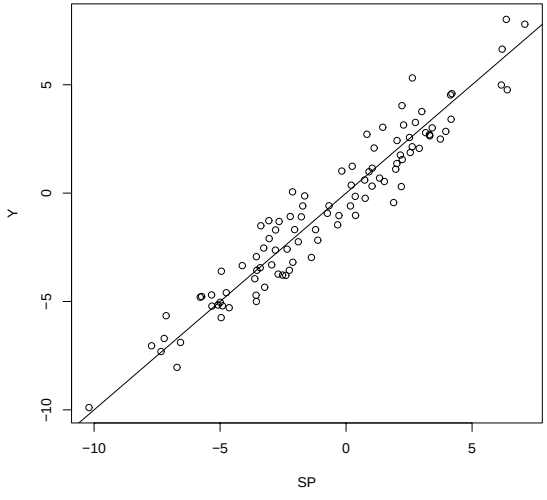

Figure 1.1: SSP for MLR Data

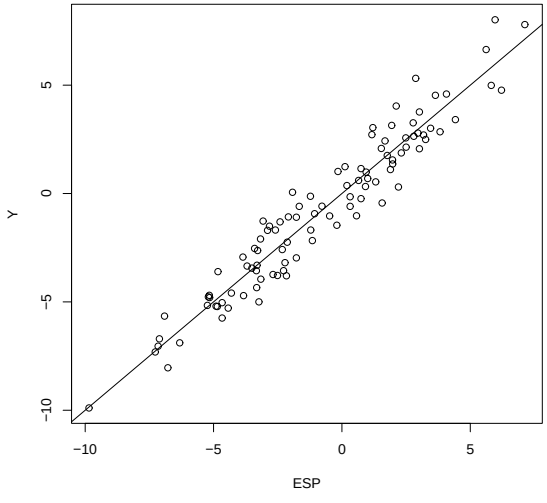

Figure 1.2: ESSP = Response Plot for MLR Data

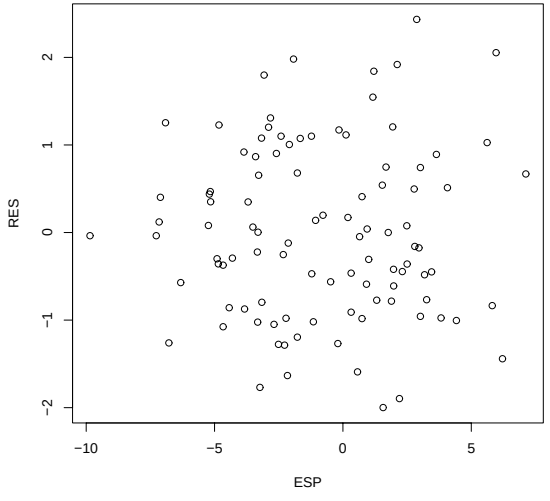

Figure 1.3: Residual Plot for MLR Data

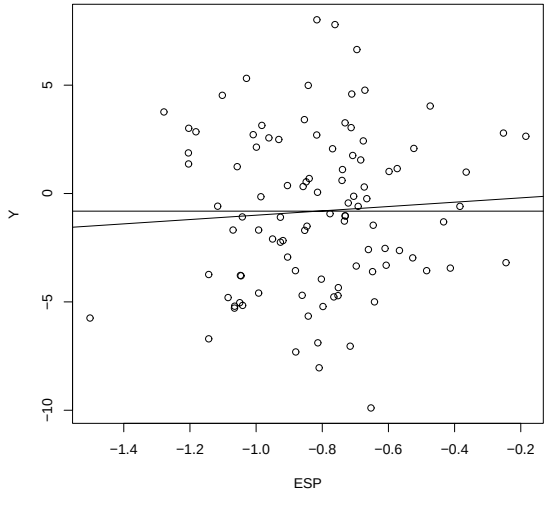

Figure 1.4: Response Plot when Y is Independent of the Predictors