

# Applied Linear Statistical Models Outline

Darshan Patel

December 19, 2019

# Contents

<b>1</b>	<b>Simple Linear Regression</b>	<b>3</b>
1.1	Linear Regression with One Predictor Variable . . . . .	3
1.1.1	Relations between Variables . . . . .	3
1.1.2	Regression Models and their Uses . . . . .	3
1.1.3	Simple Linear Regression Model with Distribution of Error Terms Unspecified	4
1.1.4	Data for Regression Analysis . . . . .	6
1.1.5	Overview of Steps in Regression Analysis . . . . .	6
1.1.6	Estimation of Regression Function . . . . .	7
1.1.7	Estimation of Error Terms Variance $\sigma^2$ . . . . .	10
1.1.8	Normal Error Regression Model . . . . .	11
1.2	Inferences in Regression and Correlation Analysis . . . . .	14
1.2.1	Inferences Concerning $\beta_1$ . . . . .	14
1.2.2	Inferences Concerning $\beta_0$ . . . . .	14
1.2.3	Some Considerations on Making Inferences Concerning $\beta_0$ and $\beta_1$ . . . . .	14
1.2.4	Interval Estimation of $E[Y_h]$ . . . . .	14
1.2.5	Prediction of New Observation . . . . .	14
1.2.6	Confidence Interval for Regression Line . . . . .	14
1.2.7	Analysis of Variance Approach to Regression Analysis . . . . .	14
1.2.8	General Linear Test Approach . . . . .	14
1.2.9	Descriptive Measures of Linear Association between $X$ and $Y$ . . . . .	14
1.2.10	Considerations in Applying Regression Analysis . . . . .	14
1.2.11	Normal Correlation Models . . . . .	14
1.3	Diagnostics and Remedial Measures . . . . .	14
1.3.1	Diagnostics for Predictor Variable . . . . .	14
1.3.2	Residuals . . . . .	14
1.3.3	Diagnostics for Residuals . . . . .	14
1.3.4	Overview for Tests Involving Residuals . . . . .	14
1.3.5	Correlation Test for Normality . . . . .	14
1.3.6	Tests for Constancy of Error Variance . . . . .	14
1.3.7	$F$ Test for Lack of Fit . . . . .	14
1.3.8	Overview of Remedial Measures . . . . .	14
1.3.9	Transformations . . . . .	14
1.3.10	Exploration of Shape of Regression Function . . . . .	14
1.3.11	Case Example - Plutonium . . . . .	14
1.4	Simultaneous Inferences and Other Topics in Regression Analysis . . . . .	14
1.4.1	Joint Estimation of $\beta_0$ and $\beta_1$ . . . . .	14
1.4.2	Simultaneous Estimation of Mean Responses . . . . .	14

1.4.3	Simultaneous Prediction Intervals for New Observations . . . . .	14
1.4.4	Regression through Origin . . . . .	14
1.4.5	Effects of Measurement Errors . . . . .	14
1.4.6	Inverse Predictions . . . . .	14
1.4.7	Choice of $X$ Levels . . . . .	14
1.5	Matrix Approach to Simple Linear Regression Analysis . . . . .	14
1.5.1	Matrices . . . . .	14
1.5.2	Matrix Addition and Subtraction . . . . .	14
1.5.3	Matrix Multiplication . . . . .	14
1.5.4	Special Types of Matrices . . . . .	14
1.5.5	Linear Dependence and Rank of Matrix . . . . .	14
1.5.6	Inverse of a Matrix . . . . .	14
1.5.7	Some Basic Results for Matrices . . . . .	14
1.5.8	Random Vectors and Matrices . . . . .	14
1.5.9	Simple Linear Regression Model in Matrix Terms . . . . .	14
1.5.10	Least Squares Estimation of Regression Parameters . . . . .	14
1.5.11	Fitted Values and Residuals . . . . .	14
1.5.12	Analysis of Variance Results . . . . .	14
1.5.13	Inferences in Regression Analysis . . . . .	14

# Chapter 1

## Simple Linear Regression

### 1.1 Linear Regression with One Predictor Variable

#### 1.1.1 Relations between Variables

- Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that a response or outcome variable can be predicted from the other, or others
- A functional relation between two variables is expressed as follows: if  $X$  denotes the independent variable and  $Y$  the dependent variable, a functional relation is of the form

$$Y = f(X)$$

Given a particular value of  $X$ , the function  $f$  indicates the corresponding value of  $Y$

- A statistical relation, unlike a functional relation, is not a perfect one; in general, the observations for a statistical relation do not fall directly on the curve of relationship
- Statistical relations can be highly useful, even though they do not have the exactitude of a functional relation

#### 1.1.2 Regression Models and their Uses

- A regression model is a formal means of expressing the two essential ingredients of a statistical relation:
  - A tendency of the response variable  $Y$  to vary with the predictor variable  $X$  in a systematic fashion
  - A scattering of points around the curve of statistical relationship
- These two characteristics are embodied in a regression model by postulating that:
  - There is a probability distribution of  $Y$  for each level of  $X$
  - The means of these probability distributions vary in some systematic fashion with  $X$
- The systematic relationship between  $X$  and  $Y$  is called the regression function of  $Y$  on  $X$ ; the graph of the regression function is called the regression curve

- Regression models may differ in the form of the regression function (linear, curvilinear), in the shape of the probability distribution of  $Y$  (symmetrical, skewed), and in other ways
- Regression models may contain more than one predictor variable
- Since reality must be reduced to manageable proportions whenever models are constructed, only a limited number of explanatory or predictor variables can, or should, be included in a regression model for any situation of interest
- A central problem in many exploratory studies is therefore that of choosing, for a regression model, a set of predictor variables that is “good” in some sense for the purposes of the analysis
- The choice of the functional form of the regression relation is tied to the choice of the predictor variables; sometimes, relevant theory may indicate the appropriate functional form
- More frequently, the functional form of the regression relation is not known in advance and must be decided upon empirically once the data have been collected
- Linear or quadratic regression functions are often used as satisfactory first approximations to regression functions of unknown nature
- In formulating a regression model, the coverage is usually restricted to some interval or region of values of the predictor variable(s) which is determined either by the design of the investigation or by the range of data at hand
- Regression analysis serves three major purposes: description, control and prediction
- The existence of a statistical relation between the response variable  $Y$  and the explanatory or predictor variable  $X$  does not imply in any way that  $Y$  depends casually on  $X$

### 1.1.3 Simple Linear Regression Model with Distribution of Error Terms Unspecified

- A basic regression model where there is only one predictor variable and the regression function is linear can be stated as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where

$Y_i$  is the value of the response variable in the  $i$ th trial

$\beta_0$  and  $\beta_1$  are parameters

$X_i$  is a known constant, namely, the value of the predictor variable in the  $i$ th trial

$\varepsilon_i$  is a random error with mean  $E[\varepsilon_i] = 0$  and variance  $\text{Var}[\varepsilon_i] = \sigma^2$ ;  $\varepsilon_i$  and  $\varepsilon_j$  are uncorrelated so that their covariance is zero (i.e.,  $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$  for all  $i, j; i \neq j$ )

$i = 1, \dots, n$

- This regression model is said to be simple, linear in its parameters, and linear in the predictor variable

- Important Features of the Model

1. The response  $Y_i$  in the  $i$ th trial is the sum of two components: (1) the constant term  $\beta_0 + \beta_1 X_i$  and (2) the random term  $\varepsilon_i$ ; hence  $Y_i$  is a random variable
2. Since  $E[\varepsilon_i] = 0$ , then

$$E[Y_i] = E[\beta_0 + \beta_1 X_i + \varepsilon_i] = \beta_0 + \beta_1 X_i + E[\varepsilon_i] = \beta_0 + \beta_1 X_i$$

Thus, the response  $Y_i$ , when the level of  $X$  in the  $i$ th trial is  $X_i$ , comes from a probability distribution whose mean is

$$E[Y_i] = \beta_0 + \beta_1 X_i$$

3. The response  $Y_i$  in the  $i$ th trial exceeds or falls short of the value of the regression function by the error term amount  $\varepsilon_i$
4. The error terms  $\varepsilon_i$  are assumed to have constant variance  $\sigma^2$  and so the responses  $Y_i$  have the same constant variance

$$\text{Var}[Y_i] = \sigma^2$$

- The parameters  $\beta_0$  and  $\beta_1$  in the regression model are called regression coefficients
- The parameter  $\beta_1$  is the slope of the regression line (indicating the change in the mean of the probability distribution of  $Y$  per unit change in  $X$ )
- The parameter  $\beta_0$  is the  $Y$  intercept of the regression line
- When the scope of the model includes  $X = 0$ ,  $\beta_0$  gives the mean of the probability distribution of  $Y$  at  $X = 0$ ; when the scope of the model does not cover  $X = 0$ ,  $\beta_0$  does not have any particular meaning as a separate term in the regression model
- The simple linear regression model can be written equivalently as follows: let  $X_0$  be a constant identically equal to 1, then

$$Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \text{ where } X_0 \equiv 1$$

This version of the model associates an  $X$  variable with each regression coefficient

- An alternative modification is to use for the predictor variable the deviation  $X_i - \bar{X}$  rather than  $X_i$ , then

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_i - \bar{X}) + \beta_1\bar{X} + \varepsilon_i \\ &= (\beta_0 + \beta_1\bar{X}) + \beta_1(X_i - \bar{X}) + \varepsilon_i \\ &= \beta_0^* + \beta_1(X_i - \bar{X}) + \varepsilon_i \end{aligned}$$

Thus this alternative model is

$$Y_i = \beta_0^* + \beta_1(X_i - \bar{X}) + \varepsilon_i$$

where

$$\beta_0^* = \beta_0 + \beta_1\bar{X}$$

### 1.1.4 Data for Regression Analysis

- Data for regression analysis may be obtained from nonexperimental or experimental studies
- Observational data are data obtained from nonexperimental studies; such studies do not control the explanatory or predictor variable(s) of interest
- A major limitation of observational data is that they often do not provide adequate information about cause and effect relationships
- Frequently, it is possible to conduct a controlled experiment to provide data from which the regression parameters can be estimated
- When control over the explanatory variable(s) is exercised through random assignments, the resulting experimental data provide much stronger information about cause and effect relationships than do observational data; the reason is that randomization tends to balance out the effects of any other variables that might affect the response variable
- In the terminology of experimental design, a treatment is the object being measured and the experimental units are the subjects of the study, from whom the treatment is done on and measured; control over the explanatory variable(s) then consists of assigning a treatment to each of the experimental units by means of randomization
- The most basic type of statistical design for making randomized assignments of treatments to experimental units (or vice versa) is the completely randomized design; with this design, the assignments are made completely at random
- This complete randomization provides that all combinations of experimental units assigned to the different treatments are equally likely, which implies that every experimental unit has an equal chance to receive any one of the treatment
- A completely randomized design is particularly useful when the experimental units are quite homogeneous; this design is very flexible; it accommodates any number of treatments and permits different sample sizes for different treatments
- Its chief disadvantage is that, when the experimental units are heterogeneous, this design is not as efficient as some other statistical designs

### 1.1.5 Overview of Steps in Regression Analysis

- The regression models given in the following chapters can be used either for observational data or for experimental data from a completely randomized design (regression analysis can also utilize data from other types of experimental designs, but the regression models presented here will need to be modified)
- Typical Strategy for Regression Analysis
  1. Start
  2. Exploratory data analysis
  3. Develop one or more tentative regression models
  4. Is one or more of the regression models suitable for the data at hand?

- If yes, continue
- If no, revise regression models and/or develop new ones and answer the question again
- 5. Identify most suitable model
- 6. Make inferences on basis of regression model
- 7. Stop

### 1.1.6 Estimation of Regression Function

- The observational or experimental data to be used for estimating the parameters of the regression function consist of observations on the explanatory or predictor variable  $X$  and the corresponding observations on the response variable  $Y$ ; for each trial, there is an  $X$  observation and a  $Y$  observation; denote the  $(X, Y)$  observations for the first trial as  $(X_1, Y_1)$ , for the second trial as  $(X_2, Y_2)$ , and in general for the  $i$ th trial as  $(X_i, Y_i)$ , where  $i = 1, \dots, n$
- For the observations  $(X_i, Y_i)$  for each case, the method of least squares considers the deviation of  $Y_i$  from its expected value  $Y_i - (\beta_0 + \beta_1 X_i)$ ; in particular, the method of least squares requires considering the sum of the  $n$  squared deviations; this criterion is denoted by  $Q$ :

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

According to the method of least squares, the estimators of  $\beta_0$  and  $\beta_1$  are those values  $b_0$  and  $b_1$ , respectively, that minimize the criterion  $Q$  for the given sample observations  $(X_1, Y_1), \dots, (X_n, Y_n)$

- The estimators  $b_0$  and  $b_1$  that satisfy the least squares criterion can be found in two basic ways:
  - Numerical search procedures can be used that evaluate in a systematic fashion the least squares criterion for different estimates  $b_0$  and  $b_1$  until the ones that minimize  $Q$  are found
  - Analytical procedures can often be used to find the values of  $b_0$  and  $b_1$  that minimize  $Q$ ; this is feasible when the regression model is not mathematically complex
- Using the analytical approach, the values  $b_0$  and  $b_1$  that minimize  $Q$  for the simple linear regression model are given by the following simultaneous equations:

$$\begin{aligned} \sum Y_i &= nb_0 + b_1 \sum X_i \\ \sum X_i Y_i &= b_0 \sum X_i + b_1 \sum X_i^2 \end{aligned}$$

These two equations are called normal equations;  $b_0$  and  $b_1$  are called point estimators of  $\beta_0$  and  $\beta_1$  respectively

- The normal equations can be solved simultaneously for  $b_0$  and  $b_1$ :

$$\begin{aligned} b_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ b_0 &= \frac{1}{n} \left( \sum Y_i - b_1 \sum X_i \right) = \bar{Y} - b_1 \bar{X} \end{aligned}$$

where  $\bar{X}$  and  $\bar{Y}$  are the means of the  $X_i$  and  $Y_i$  observations respectively



- Derivation of above result: for given sample observations  $(X_i, Y_i)$ , the quantity  $Q$  is a function of  $\beta_0$  and  $\beta_1$ ; the values of  $\beta_0$  and  $\beta_1$  that minimize  $Q$  can be derived by differentiating  $Q$  with respect to  $\beta_0$  and  $\beta_1$ :

$$\begin{aligned}\frac{\partial Q}{\partial \beta_0} &= -2 \sum (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i)\end{aligned}$$

Setting these partial derivatives to zero, using  $b_0$  and  $b_1$  to denote the particular values of  $\beta_0$  and  $\beta_1$  that minimize  $Q$  and simplifying, the following is obtained

$$\begin{aligned}\sum_{i=1}^n (Y_i - b_0 - b_1 X_i) &= 0 \\ \sum_{i=1}^n X_i (Y_i - b_0 - b_1 X_i) &= 0\end{aligned}$$

Expanding this, the following is true:

$$\begin{aligned}\sum Y_i - nb_0 - b_1 \sum X_i &= 0 \\ \sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 &= 0\end{aligned}$$

By rearranging terms, the normal equations are obtained

- Gauss-Markov Theorem: Under the conditions of the simple linear regression model the least squares estimators  $b_0$  and  $b_1$ , as given above, are unbiased and have minimum variance among all unbiased linear estimators
- This theorem states first that  $b_0$  and  $b_1$  are unbiased estimators and so

$$E[b_0] = \beta_0 \quad E[b_1] = \beta_1$$

so that neither estimator tends to overestimate or underestimate systematically

- Second, the theorem states that the estimators  $b_0$  and  $b_1$  are more precise (o.e., their sampling distributions are less variable) than any other estimators belonging to the class of unbiased estimators that are linear functions of the observations  $Y_1, \dots, Y_n$ ; the estimators  $b_0$  and  $b_1$  are such linear functions of the  $Y_i$
- Given sample estimators  $b_0$  and  $b_1$  of the parameters in the regression function,  $E[Y] = \beta_0 + \beta_1 X$ , the regression function is estimated as

$$\hat{Y} = b_0 + b_1 X$$

where  $\hat{Y}$  is the value of the estimated regression function at the level  $X$  of the predictor variable

- A value of the response variable is called a response while  $E[Y]$  is called the mean response; thus, the mean response stands for the mean of the probability distribution of  $Y$  corresponding to the level  $X$  of the predictor variable

- $\hat{Y}$  is a point estimator of the mean response when the level of the predictor variable is  $X$
- As an extension of the Gauss-Markov Theorem,  $\hat{Y}$  is an unbiased estimator of  $E[Y]$ , with minimum variance in the class of unbiased linear estimators
- Let  $\hat{Y}_i$  be the fitted value for the  $i$ th case

$$\hat{Y}_i = b_0 + b_1 X_i \quad i = 1, \dots, n$$

Thus the fitted value  $\hat{Y}_i$  is to be viewed in distinction to the observed value  $Y_i$

- The  $i$ th residual is the difference between the observed value  $Y_i$  and the corresponding fitted value  $\hat{Y}_i$ ; this residual is denoted by  $e_i$  and is defined as follows:

$$e_i = Y_i - \hat{Y}_i$$

- For the simple linear regression model, the residual  $e_i$  becomes

$$e_i = Y_i - (b_0 + b_1 X_i) = Y_i - b_0 - b_1 X_i$$

- The model error term value  $\varepsilon_i = Y_i - E[Y_i]$  involves the vertical deviation of  $Y_i$  from the unknown true regression line hence is unknown; the residual  $e_i = Y_i - \hat{Y}_i$  is the vertical deviation of  $Y_i$  from the fitted value  $\hat{Y}_i$  on the estimated regression line, and it is known

- Properties of Fitted Regression Line

1. The sum of the residuals is zero

$$\sum_{i=1}^n e_i = 0$$

2. The sum of the squared residuals,  $\sum e_i^2$  is a minimum

3. The sum of the observed values  $Y_i$  equals the sum of the fitted values  $\hat{Y}_i$ :

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$$

4. The sum of the weighted residuals is zero when the residual in the  $i$ th trial is weighted by the level of the predictor variable in the  $i$ th trial:

$$\sum_{i=1}^n X_i e_i = 0$$

5. The sum of the weighted residuals is zero when the residual in the  $i$ th trial is weighted by the fitted value of the response variable for the  $i$ th trial:

$$\sum_{i=1}^n \hat{Y}_i e_i = 0$$

6. The regression line always goes through the point  $(\bar{X}, \bar{Y})$

### 1.1.7 Estimation of Error Terms Variance $\sigma^2$

- The variance  $\sigma^2$  of the error terms  $\varepsilon_i$  in the regression model needs to be estimated to obtain an indication of the variability of the probability distribution of  $Y$
- The variance  $\sigma^2$  of a single population is estimated by the sample variance  $s^2$  as follows:

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}$$

where the sum is called a sum of squares and  $n - 1$  is the degrees of freedom; this number is  $n - 1$  because one degree of freedom is lost by using  $\bar{Y}$  as an estimate of the unknown population mean  $\mu$

- This estimator is the usual sample variance which is an unbiased estimator of the variance  $\sigma^2$  of an infinite population
- The sample variance is often called a mean square because a sum of squares has been divided by the appropriate number of degrees of freedom
- For the regression model, the variance for each observation  $Y_i$  is  $\sigma^2$ , the same as that of each error term  $\varepsilon_i$ ; a sum of squared deviations are needed to be calculated but note that the  $Y_i$  now come from different probability distributions with different means that depend upon the level  $X_i$ ; thus, the deviation of an observation  $Y_i$  must be calculated around its own estimated mean  $\hat{Y}_i$
- The deviations are the residuals

$$Y_i - \hat{Y}_i = e_i$$

and the appropriate sum of squares, denoted by SSE, is

$$\text{SSE} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$$

where SSE stands for error sum of squares or residual sum of squares

- The SSE has  $n - 2$  degrees of freedom because both  $\beta_0$  and  $\beta_1$  had to be estimated in obtaining the estimated means  $\hat{Y}_i$
- The appropriate mean square, denoted by MSE or  $s^2$  is

$$s^2 = \text{MSE} = \frac{\text{SSE}}{n - 2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - 2} = \frac{\sum e_i^2}{n - 2}$$

where MSE stands for error mean square or residual mean square

- The MSE is an unbiased estimator for  $\sigma^2$  for a regression model

$$E[\text{MSE}] = \sigma^2$$

- An estimator of the standard deviation  $\sigma$  is simply  $s = \sqrt{\text{MSE}}$ , the positive square root of the MSE

### 1.1.8 Normal Error Regression Model

- The normal error regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where

$Y_i$  is the observed response in the  $i$ th trial

$X_i$  is a known constant, the level of the predictor variable in the  $i$ th trial

$\beta_0$  and  $\beta_1$  are parameters

$\varepsilon_i$  are independent  $N(0, \sigma^2)$

$i = 1, \dots, n$

- The symbol  $N(0, \sigma^2)$  stands for normally distributed with mean 0 and variance  $\sigma^2$
- The normal error model is the same as the regression model with unspecified error distribution, except this one assumes that the errors  $\varepsilon_i$  are normally distributed
- Since the errors are normally distributed, the assumption of uncorrelatedness of the  $\varepsilon_i$  in the regression model becomes one of independence in the normal error model
- This model implies that the  $Y_i$  are independent normal random variables, with mean  $E[Y_i] = \beta_0 + \beta_1 X_i$  and variance  $\sigma^2$
- The normality assumption for the error term is justifiable in many situations because the error terms frequently represent the effects of factors omitted from the model that affect the response to some extent and that vary at random without reference to the variable  $X$
- A second reason why the normality assumption of the error terms is frequently justifiable is that the estimation and testing procedures are based on the  $t$  distribution and are usually only sensitive to large departures from normality
- When the functional form of the probability distribution of the error terms is specified, estimators of the parameters  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$  can be obtained using the method of maximum likelihood; this method chooses as estimates those values of the parameters that are most consistent with the sample data
- The method of maximum likelihood uses the product of the densities as the measure of consistency of the parameter value with the sample data; the product is called the likelihood value of the parameter value and is denoted by  $L(\cdot)$  where  $\cdot$  is the parameter being estimated; if the value of  $\cdot$  is consistent with the sample data, the densities will be relatively large and so will be the likelihood value; if the value of  $\cdot$  is not consistent with the data, the densities will be small and the product  $L(\cdot)$  will be small
- The method of maximum likelihood chooses as the maximum likelihood estimate that value of  $\cdot$  for which the likelihood value is largest; there are two methods of finding the estimates: by a systematic numerical search or by use of an analytical solution
- The product of the densities viewed as a function of the unknown parameters is called the likelihood function

- In general, the density of an observation  $Y_i$  for the normal error regression model is as follows, utilizing the fact that  $E[Y_i] = \beta_0 + \beta_1 X_i$  and  $\text{Var}[Y_i] = \sigma^2$ :

$$f_i = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma} \right)^2 \right]$$

- The likelihood function for  $n$  observations  $Y_1, \dots, Y_n$  is the product of the individual densities; since the variance  $\sigma^2$  of the error terms is usually unknown, the likelihood function is a function of three parameters  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left[ -\frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1 X_i)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right] \end{aligned}$$

- The values of  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$  that maximize this likelihood function are the maximum likelihood estimators and are denoted by  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\sigma}^2$  respectively; these estimators are calculated analytically and are as follows:

Parameter	Maximum Likelihood Estimator
$\beta_0$	$\hat{\beta}_0 = b_0 = \frac{1}{n} (\sum Y_i - b_1 \sum X_i) = \bar{Y} - b_1 \bar{X}$
$\beta_1$	$\hat{\beta}_1 = b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$
$\sigma^2$	$\hat{\sigma}^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n}$

- Thus, the maximum likelihood estimators of  $\beta_0$  and  $\beta_1$  are the same estimators as those provided by the methods of least squares; the maximum likelihood estimator  $\hat{\sigma}^2$  is biased and ordinarily the unbiased MSE or  $s^2$  is used
- The unbiased MSE or  $s^2$  differs but slightly from the maximum likelihood estimator  $\hat{\sigma}^2$ , especially if  $n$  is not small

$$s^2 = \text{MSE} = \frac{n}{n-2} \hat{\sigma}^2$$

- Since the maximum likelihood estimators of  $\hat{\beta}_0$  and  $\hat{\beta}_1$  are the same as the least squares estimators  $b_0$  and  $b_1$ , they have the properties of all least squares estimators:
  - There are unbiased
  - They have minimum variance among all unbiased linear estimators
- In addition, the maximum likelihood estimators  $b_0$  and  $b_1$  for the normal error regression model have other desirable properties:
  - They are consistent
  - They are sufficient
  - They are minimum variance unbiased; that is, they have minimum variance in the class of all unbiased estimators (linear or otherwise)

- Derivation of maximum likelihood estimators: take partial derivatives of  $L$  with respect to  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$ , equating each of the partials to zero and solving the system of equations obtained; work with  $\log_e L$  rather than  $L$  since both are maximized for the same values of  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$ :

$$\log L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

The partial derivatives are as shown:

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_0} &= \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial \log L}{\partial \beta_1} &= \frac{1}{\sigma^2} \sum X_i (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \end{aligned}$$

Setting these partial derivatives equal to zero and replacing  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$  by the estimators  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\sigma}^2$ , and after some simplifications:

$$\begin{aligned} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0 \\ \sum X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0 \\ \frac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n} &= \hat{\sigma}^2 \end{aligned}$$

The first two equations are identical to the earlier least squares normal equations and the last one is the biased estimator of  $\sigma^2$  as given earlier

## 1.2 Inferences in Regression and Correlation Analysis

- 1.2.1 Inferences Concerning  $\beta_1$
- 1.2.2 Inferences Concerning  $\beta_0$
- 1.2.3 Some Considerations on Making Inferences Concerning  $\beta_0$  and  $\beta_1$
- 1.2.4 Interval Estimation of  $E[Y_h]$
- 1.2.5 Prediction of New Observation
- 1.2.6 Confidence Interval for Regression Line
- 1.2.7 Analysis of Variance Approach to Regression Analysis
- 1.2.8 General Linear Test Approach
- 1.2.9 Descriptive Measures of Linear Association between  $X$  and  $Y$
- 1.2.10 Considerations in Applying Regression Analysis
- 1.2.11 Normal Correlation Models

## 1.3 Diagnostics and Remedial Measures

- 1.3.1 Diagnostics for Predictor Variable
- 1.3.2 Residuals
- 1.3.3 Diagnostics for Residuals
- 1.3.4 Overview for Tests Involving Residuals
- 1.3.5 Correlation Test for Normality
- 1.3.6 Tests for Constancy of Error Variance
- 1.3.7  $F$  Test for Lack of Fit
- 1.3.8 Overview of Remedial Measures
- 1.3.9 Transformations
- 1.3.10 Exploration of Shape of Regression Function
- 1.3.11 Case Example - Plutonium

## 1.4 Simultaneous Inferences and Other Topics in Regression Analysis

- 1.4.1 Joint Estimation of  $\beta_0$  and  $\beta_1$
- 1.4.2 Simultaneous Estimation of Mean Responses
- 1.4.3 Simultaneous Prediction Intervals for New Observations
- 1.4.4 Regression through Origin
- 1.4.5 Effects of Measurement Errors
- 1.4.6 Inverse Predictions
- 1.4.7 Choice of  $X$  Levels