2. TWO-VARIABLE REGRESSION ANALYSIS

In order to understand two-variable regression, consider the data given in Table 2.1.
The data in the below table refer to a total **Population** of 42 families with their weekly income (X) and weekly consumption expenditure (Y).

Table 2.1: Weekly family Expenditure (Y), Baht and Income (X), Baht

	X=Weekly family Income, Baht					
	500	600	700	800	900	1000
	360	376	458	610	600	700
	313	475	422	468	531	679
	322	380	498	575	670	730
Y= Weeky	310	382	560	542	630	591
Family Expenditure	390	390	442	588	544	550
	315	425	440	466	565	620
	390	442	-	461	-	695
	400	-	-	-	-	635
Total	2800	2870	2820	3710	3540	5200
Conditional means of Y, $E(Y\|X)$	350	410	470	530	590	650

Notes -

Table 2.2: Conditional Probabilities $p(Y|X_i)$ for the Weekly Family Income (X) and Expenditure (Y)

	X=Weekly family Income, Baht					
	500	600	700	800	900	1000
	1/8	1/7	1/6	1/7	1/6	1/8
	1/8	1/7	1/6	1/7	1/6	1/8
	1/8	1/7	1/6	1/7	1/6	1/8
Y= Weekly	1/8	1/7	1/6	1/7	1/6	1/8
Family Expenditure	1/8	1/7	1/6	1/7	1/6	1/8
	1/8	1/7	1/6	1/7	1/6	1/8
	1/8	1/7	-	1/7	-	1/8
	1/8	-	-	-	-	1/8
Conditional means of Y, $E(Y|X)$	350	410	470	530	590	650

Notes -

Conditional expected value of weekly consumption expenditure given the income level =X , $E(Y|X)$

Unconditional expected value ,$E(Y)$

Figure 2.1: Conditional Distribution of Expenditure for Various Levels of Income

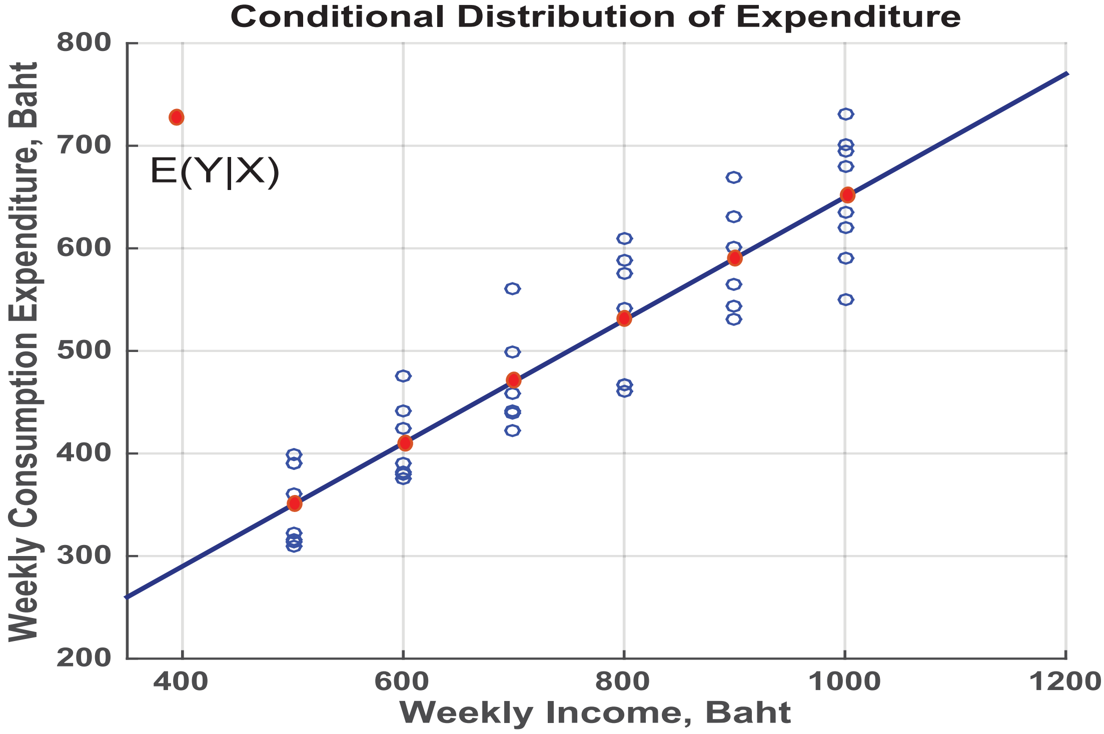
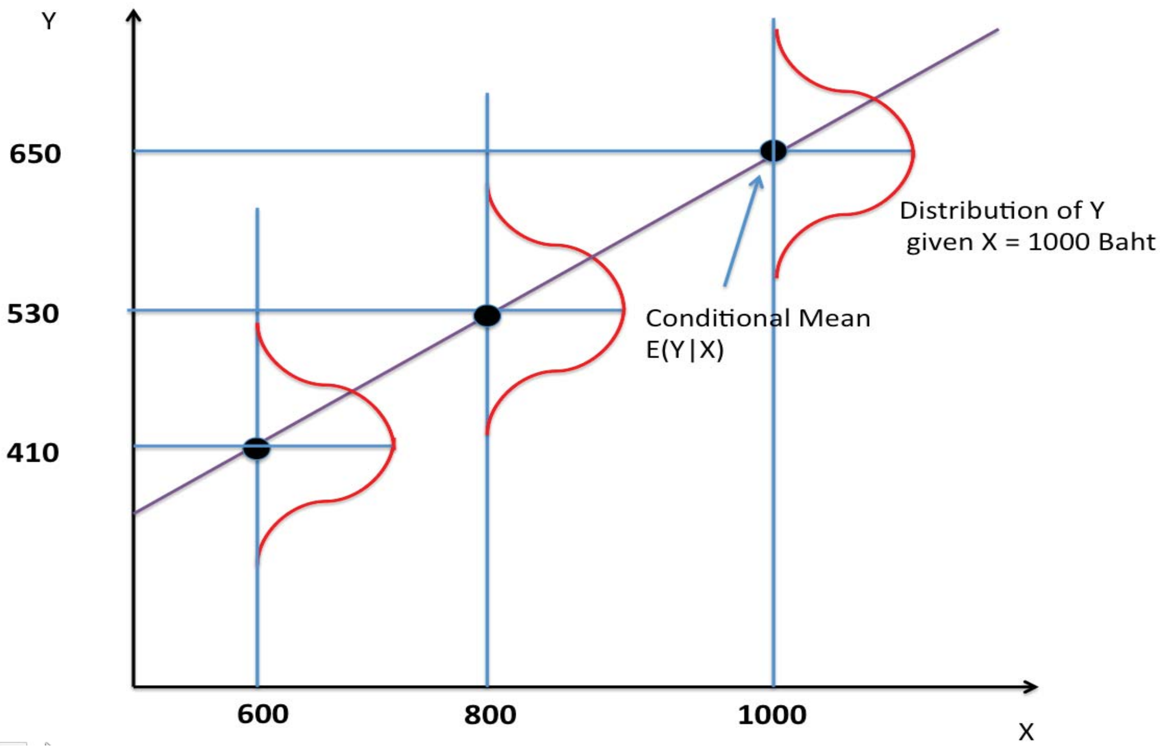


Figure 2.2: Population Regression Line (PRL)



2.1 The Concept of Population Regression Function (PRF)

The population regression function (PRF) can be written as the function of X_i:

2.1.1 What form does the function f(X) assume?

If we assume the PRF $E(Y|X_i)$ is a linear function of X_i, we get

$$E(Y|X_i) = \beta_1 + \beta_2 X_i$$

2.1.2 What is the meaning of the term LINEAR?

LINEARITY in the variables

LINEARITY in the parameters

2.2 Stochastic Specification of PRF

We can write the **deviation** of an individual Y_i around it expected value as follows:

2.2.1 The roles of the stochastic disturbance term

1. Vagueness of theory

2. Unavailability of data

3. Core variables versus peripheral variables

4. Intrinsic randomness in human behavior

5. Poor proxy variable

6. Principle of parsimony

7.Wrong functional form

2.3 The Sample Regression Function (SRF)

As mentioned, in the real situation, we cannot find out all the population of Y values corresponding to the fixed X's. We only have a sample of Y values corresponding to some fixed X's.

Therefore, our goal in this section is to estimate the population regression line (PRF) on the basis of the **SAMPLE INFORMATION.**

As a result, for the fixed X's as given in table 2.1, we only have a randomly selected sample of Y values. For example, table 2.3 and table 2.4 show a random sample from the population of table 2.1

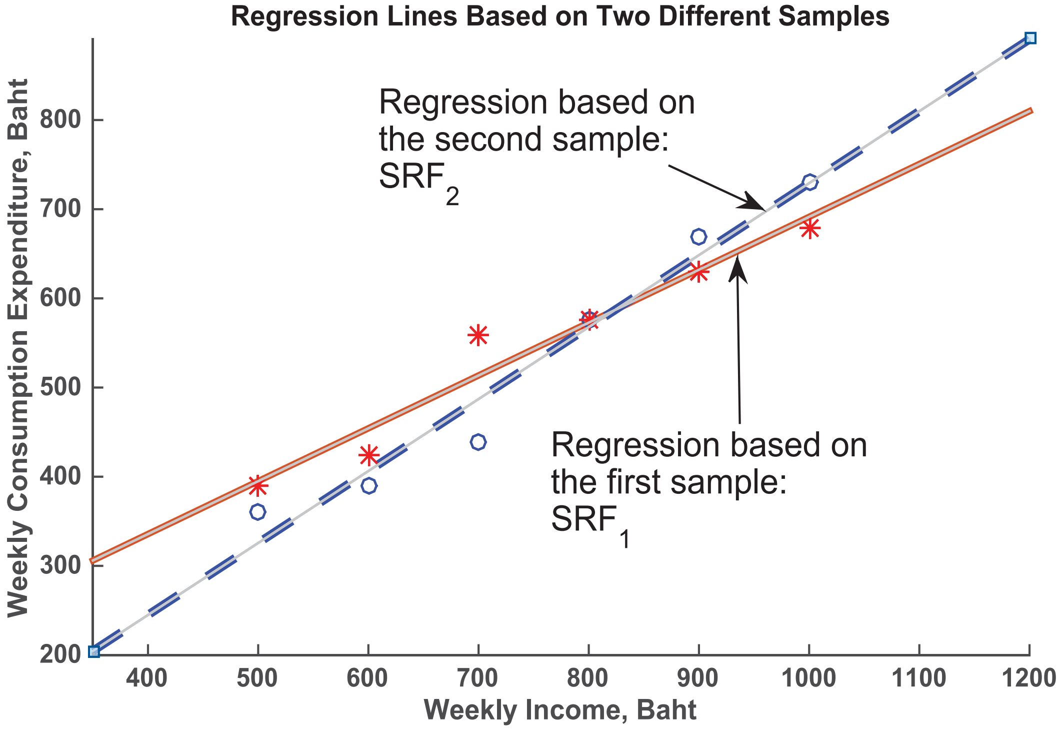
Table 2.3: Random Sample From the Population

X	Y
500	390
600	425
700	560
800	575
900	630
1000	679

Table 2.4: Another Random Sample From the Population

X	Y
500	360
600	390
700	440
800	575
900	670
1000	730

Figure 2.3: Regression lines based on two different samples



The sample regression function (SRF) can be written as:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

where $\hat{Y}$ is read as "Y-hat"

$\hat{Y}_i$ = estimator of $E(Y|X_i)$

$\hat{\beta}_1$=estimator of β_1

$\hat{\beta}_2$=estimator of β_2

We can express the SRF in its stochastic form as follows:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\mu}_i$$

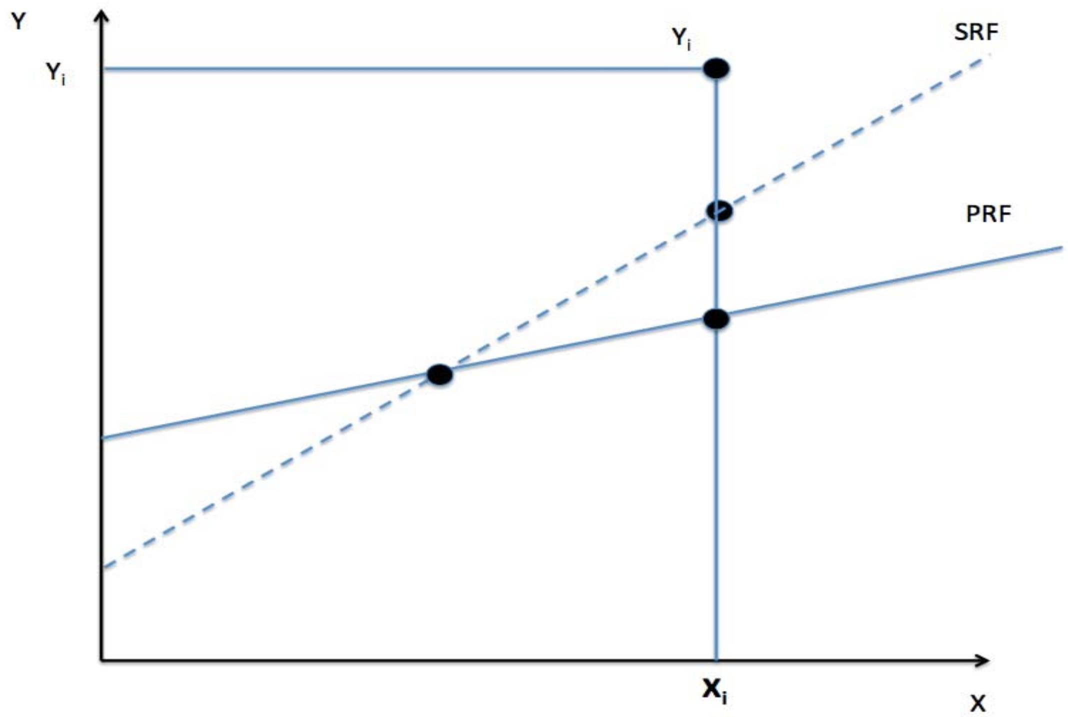
In sum, our ultimate goal is to estimate

the PRF

on the basis of

the SRF

Figure 2.4: Sample and Population Regression Lines



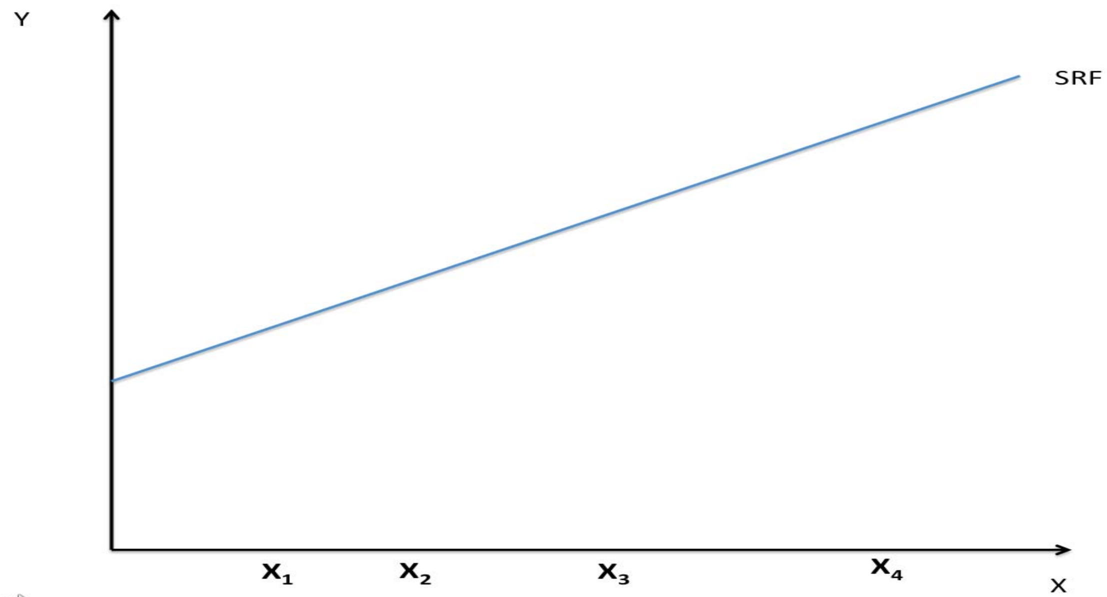
3. REGRESSION:THE PROBLEM OF ESTIMATION

As mentioned in the previous chapter, our main objective is to estimate the population regression function (PRF) based on the basis of the sample regression function (SRF) as accurately as possible.

In this chapter, we are going to discuss the method of estimation:Ordinary Least Squares (OLS)

3.1 The Method of Ordinary Least Squares (OLS)

Figure 3.1: Least-Squares Criterion



3.1.1 The Method to Find Out the Least-Squares Estimators: $\hat{\beta}_1$ and $\hat{\beta}_2$

From the SRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Now, we obtain the **least-squares estimators**:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2} \\ &= \bar{Y} - \hat{\beta}_2 \bar{X} \end{aligned} \tag{3.1}$$

$$\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum Xi \sum Yi}{n \sum X_i^2 - (\sum X_i)^2} \tag{3.2}$$

If we define $\bar{X}$ and $\bar{Y}$ to be the sample means of X and Y. Then:

$$\begin{aligned} x_i &= (X_i - \bar{X}) \\ y_i &= (Y_i - \bar{Y}) \end{aligned} \tag{3.3}$$

We can have the alternative expressions for $\hat{\beta}_2$:

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2} \\ &= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2} \end{aligned} \tag{3.4}$$

Show that

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

EXAMPLE

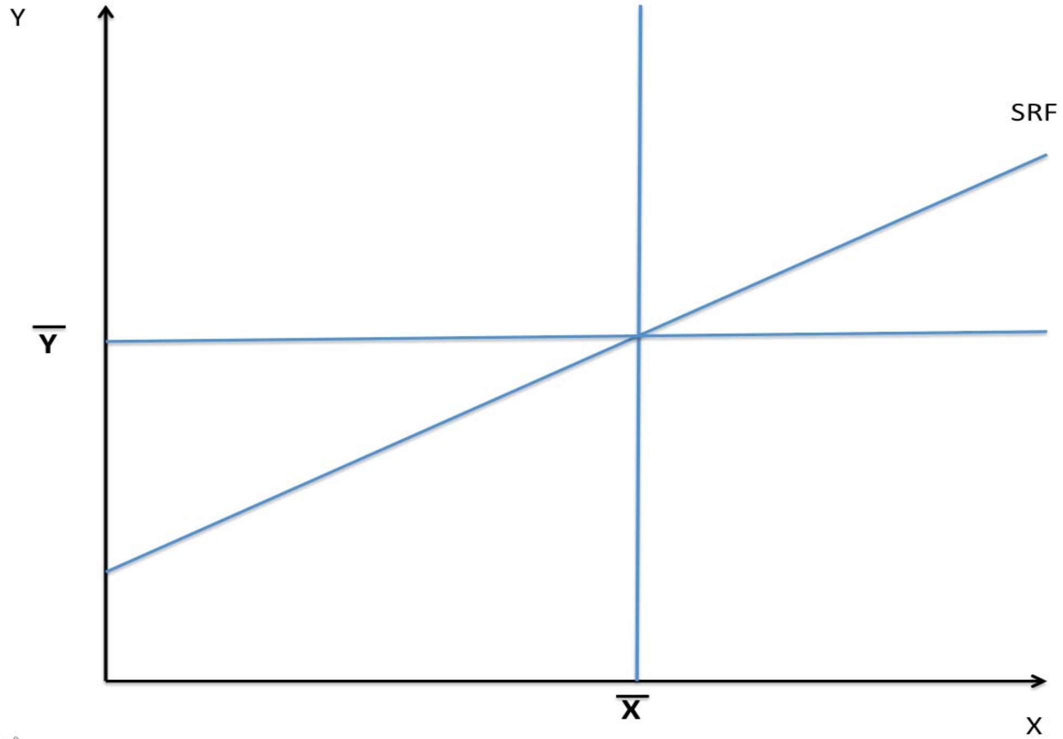
Table 3.1: Random Sample From the Population

X	Y
500	390
600	425
700	560
800	575
900	630
1000	679

Table 3.2: Raw Data Based on the Sample Data on Table 3.1

	(1) Y_i	(2) X_i	(3) Y_iX_i	(4) X_i^2	(5) $x_i = X_i - \bar{X}$	(6) $y_i = Y_i - \bar{Y}$	(7) x_i^2	(8) x_iy_i	(9) $\hat{Y}_i$	(10) $\hat{u}_i = Y_i - \hat{Y}_i$	(11) $\hat{Y}_i\hat{u}_i$
	390	500	195,000	250,000	−250	−153.17	62,500	38,291.67			
	425	600	255,000	360,000	−150	−118.17	22,500	17,725			
	560	700	392,000	490,000	−50	16.83	2,500	−841.67			
	575	800	460,000	640,000	50	31.83	2,500	1,591.67			
	630	900	567,000	810,000	150	86.83	22,500	13,025			
	679	1,000	679,000	1,000,000	250	135.83	62,500	33,958.33			
Sum	3,259	4,500	2,548,000	3,550,000	0	0	175,000	103,750			
Mean	543.17	750	424,666.67	591,666.670	0	0	29,166.67	17,291.67			

Figure 3.2: Sample Regression Line Based on the Data of Table 3.2



3.1.2 The numerical and statistical properties of OLS estimators

1. The OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are expressed solely in terms of the observable (Sample size) and quantities (i.e X and Y).

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2} \\ &= \bar{Y} - \hat{\beta}_2 \bar{X}\end{aligned} \tag{3.5}$$

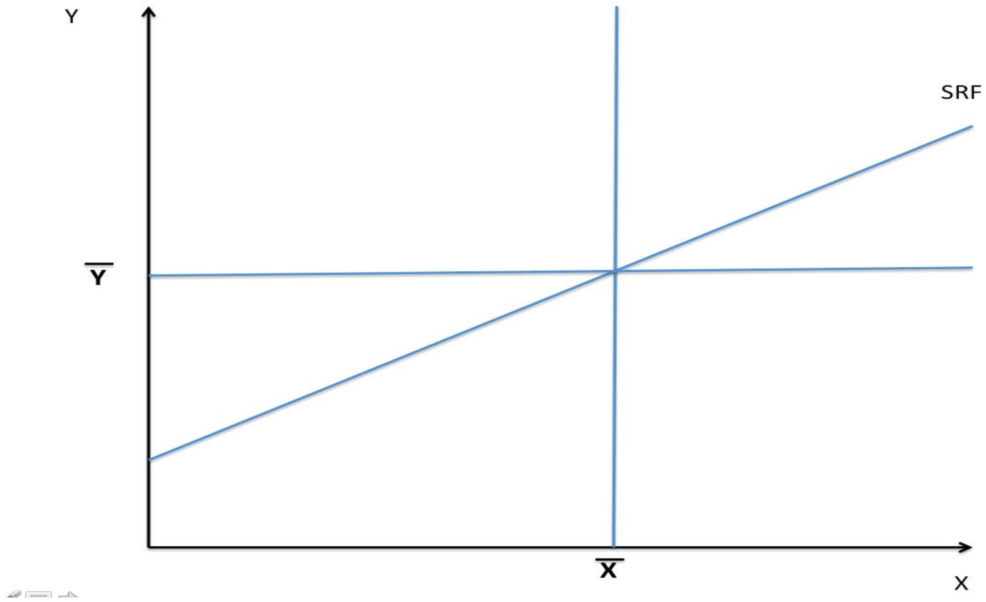
$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum Xi \sum Yi}{n\sum X_i^2 - (\sum X_i)^2} \tag{3.6}$$

2. They are **point estimators**.

3. The regression line has the following properties.

3.1 The sample regression function (SRF) passes through the sample means of Y and X ($\bar{Y}$ and $\bar{X}$).

Figure 3.3: The Sample regression Line Passes through the Sample Mean Values of Y and X



3.2 The mean value of the estimated $Y = \hat{Y}_i$ is equal to the mean value of the actual Y.

3.3. The mean value of the residuals $\hat{u}_i$ is zero.

3.4 The residuals $\hat{u}_i$ are uncorrelated with the predicted Y_i.

3.5 The residuals $\hat{u}_i$ are uncorrelated with X_i.

3.1.3 The Assumptions Underlying the Method of Least Squares

Assumption 1: Linear regression model

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

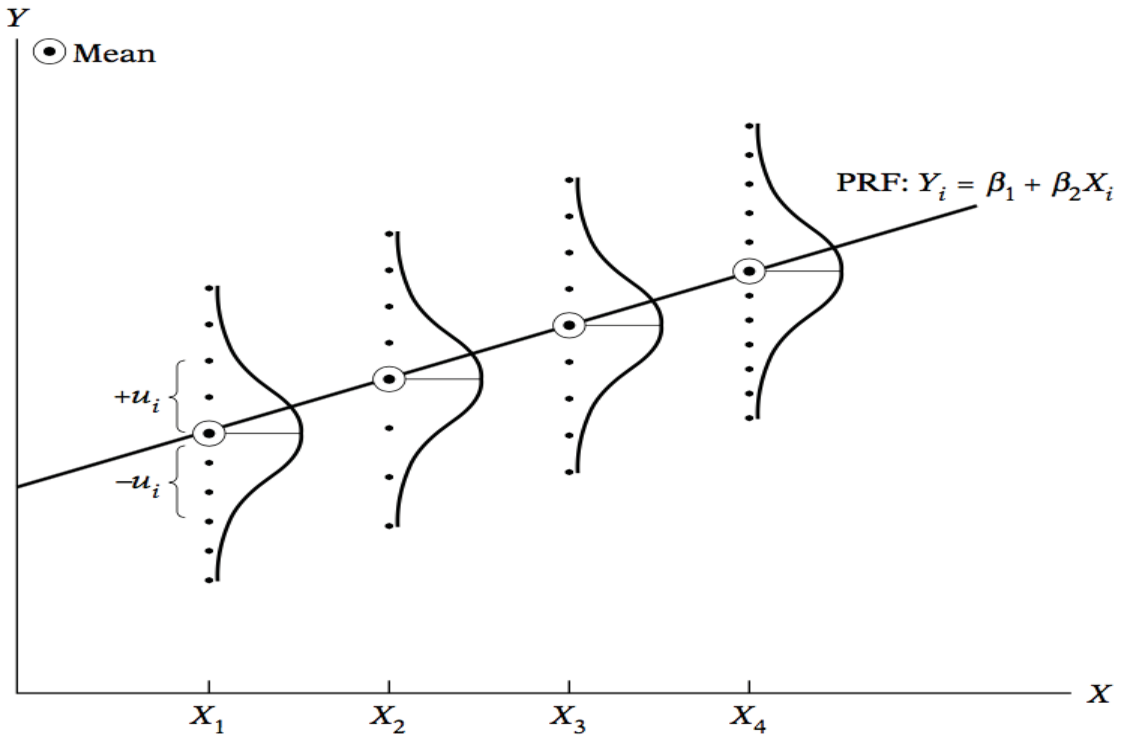
Assumption 2: X values are fixed in repeated sampling

X is assumed to be nonstochastic.

Assumption 3: Zero mean value of disturbance u_i

$$E(u_i|Xi) = 0$$

Figure 3.4: Conditional Distribution of the Disturbances u_i



Assumption 4: Homoscedasticity or Equal Variance of u_i

Figure 3.5: Homoscedasticity

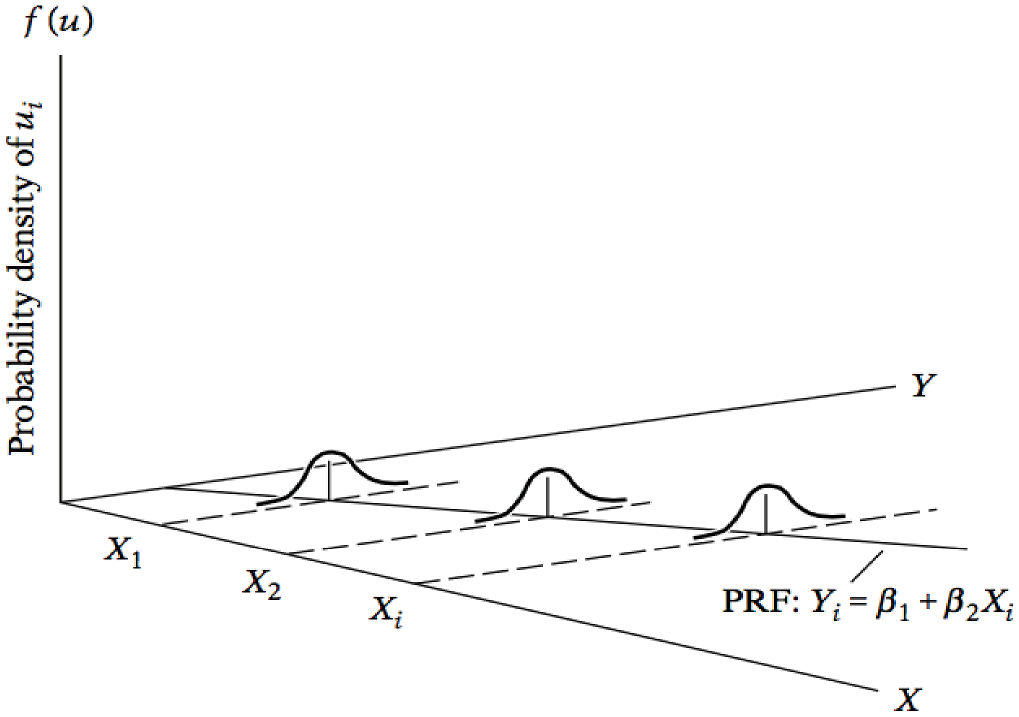
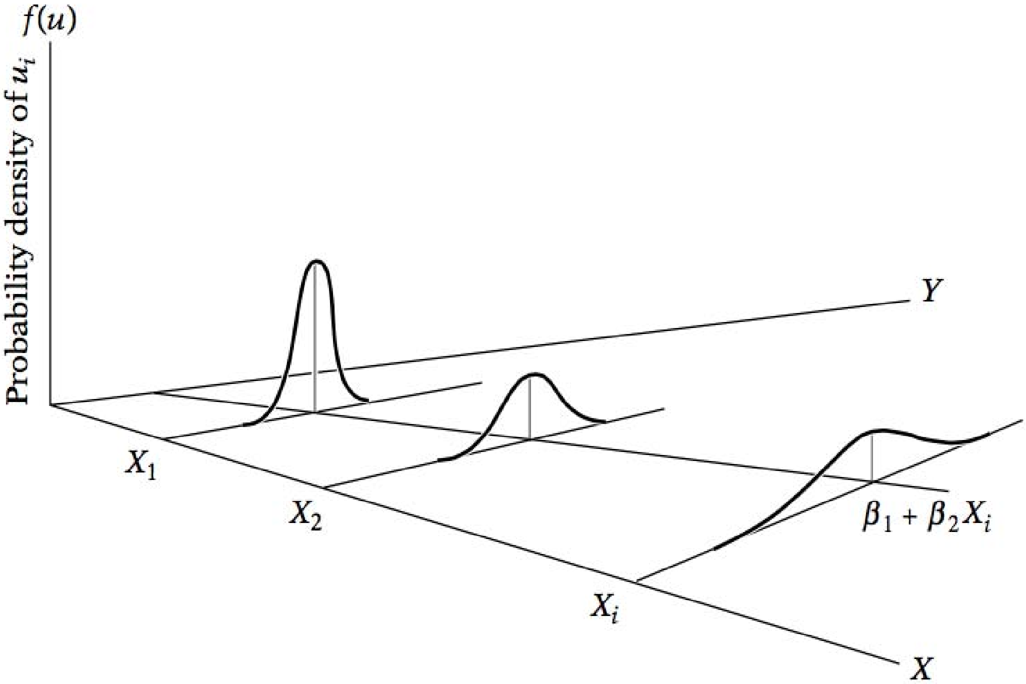


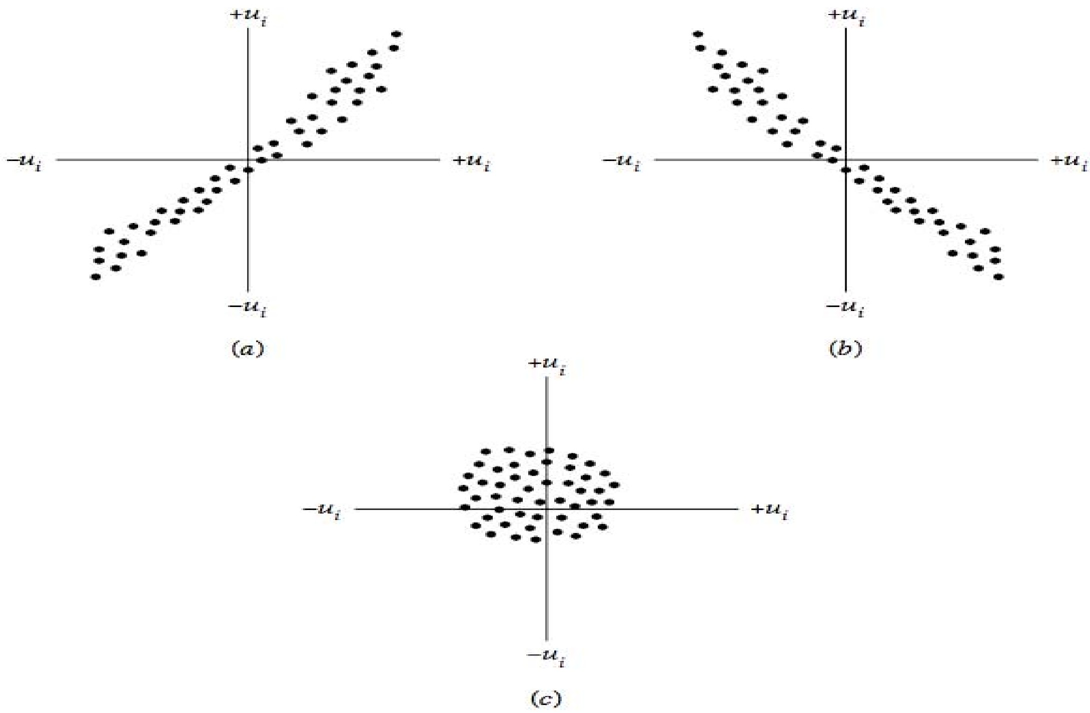
Figure 3.6: Heteroscedasticity



Assumption 5: No Autocorrelation Between the Disturbances

Assumption 6: Zero Covariance Between u_i and X_i

Figure 3.7: Patterns of Correlation Among the disturbances



Assumption 7: The number of observations n must be greater than the number of parameters to be estimated.

Assumption 8:Variability in X values.

Assumption 9: The regression model is correctly specified.

Assumption 10: There is no perfect multicollinearity.

3.1.4 Standard Errors of Least-Squares Estimates

The standard errors of the OLS estimates can be obtained as follows:
We know that

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Yi$$

where

$$k_i = \frac{x_i}{\sum x_i^2}$$

The properties of the weights k_i

1. The k_i are nonstochastic.
2. $\sum k_i = 0$
3. $\sum k_i^2 = \frac{1}{\sum x_i^2}$
4. $\sum k_i x_i = \sum k_i X i = 1$

Since

$$var(\hat{\beta}_2) = E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2$$

First Step

Find the $E(\hat{\beta}_2)$

Second Step

Using the definition of variance

$$var(\hat{\beta}_2) = E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2$$

The covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$

3.1.5 The Least-Square Estimator of σ^2

In sum, the standard errors of the OLS estimators can be obtained as follow:

$$\begin{aligned}\text{var}(\hat{\beta}_2) &= \frac{\sigma^2}{\sum x_i^2} \\ \text{se}(\hat{\beta}_2) &= \frac{\sigma}{\sqrt{\sum x_i^2}}\end{aligned} \tag{3.7}$$

$$\begin{aligned}\text{var}(\hat{\beta}_1) &= \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \\ \text{se}(\hat{\beta}_1) &= \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\sigma\end{aligned} \tag{3.8}$$

We can estimate the σ^2 from the data where the formula for the estimated σ^2 is following :

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-2}$$

where

$$\sum \hat{u_i}^2 = \sum y_i^2 - \hat{\beta_2}^2 \sum x_i^2$$

The alternative expression for computing $\sum \hat{u_i}^2$ is

$$\sum \hat{u_i}^2 = \sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2}$$

The covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ is:

$$\begin{aligned}\text{cov}(\hat{\beta}_1, \hat{\beta}_2) &= -\bar{X}\,\text{var}(\hat{\beta}_2) \\ &= -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right)\end{aligned} \tag{3.9}$$

3.1.6 Properties of Least-Squares Estimators: The Gauss-Markov Theorem

Given the assumptions of the classical linear regression model, the least-square estimators are satisfied the optimum properties which is known as **“The Gauss- Markov Theorem."** To understand this theorem, we need to know the small-sample properties of an estimator first.

The Small-Sample Properties of An Estimator

1. Unbiasedness

An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if the expected value of $\hat{\theta}$ is equal to the true θ

$$E(\hat{\theta}) = \theta$$

Therefore, if the expected value of $\hat{\theta}$ is not equal to the true θ, then the estimator is said to be biased. We can calculate the biased as:

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Figure: Biased and Unbiased Estimators

2. Minimum Variance

$\hat{\theta}_1$ is said to be a minimum variance estimator of θ if the variance of $\hat{\theta}_1$ is smaller that or at most equal to the variance of $\hat{\theta}_2$, which is any other estimator of θ

Figure: Minimum Variance

3.Best Unbiased or Efficient Estimator = property 1+ property 2

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ and the variance of $\hat{\theta}_1$ is smaller that or at most equal to the variance of $\hat{\theta}_2$, then $\hat{\theta}_1$ is **a minimum-variance unbiased estimator or best unbiased estimator.**

4. Linearity

An estimator $\hat{\theta}$ is said to be a linear estimator of θ if it is a linear function of the sample observations. For example:

$$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

Thus, $\bar{X}$ is a linear estimator because it is a linear function of the X values.

Best Linear Unbiased Estimators : BLUE

The estimator $\hat{\theta}$ is called as the Best Linear Unbiased Estimator **BLUE** if it is satisfied the properties 1,2,4 that is $\hat{\theta}$ is linear, is unbiased, and has the minimum variance in the class of all linear unbiased estimators of θ.

Minimum Mean-Square-Error (MSE) Estimator

The MSE measures dispersion around the true value of the parameter. It is defined as:

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

However, the variance of $\hat{\theta}$ measures the dispersion of the distribution of the distribution of $\hat{\theta}$ around its mean or expected value.

$$\text{var}(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2$$

The relationship between the MSE($\hat{\theta}$) and the var($\hat{\theta}$) is as follows:

An estimator $\hat{\beta}_2$ is said to be a best linear unbiased estimator (BLUE) of β_2 if the following hold:

♣ **It is linear**. It is the linear function of a random variable.

♣ **It is unbiased**. That is $E(\hat{\beta}_2)$ is equal to the true value, β_2

♣ **It has the minimum variance in the class of all such linear unbiased estimators**.

Gauss-Markov Theorem: Given the assumptions of the classical linear regression model, the least-squares estimators, in the class of unbiased linear estimators, have minimum variance, that is , they are BLUE.

3.1.7 A measure of goodness of fit: r^2

In this section, we are going to study the goodness of fit of the fitted regression line to a set of data. Let us consider the following example:

Suppose we were to estimate the family expenditure (Y) based on our information from a random sample (as in Table 3.2).

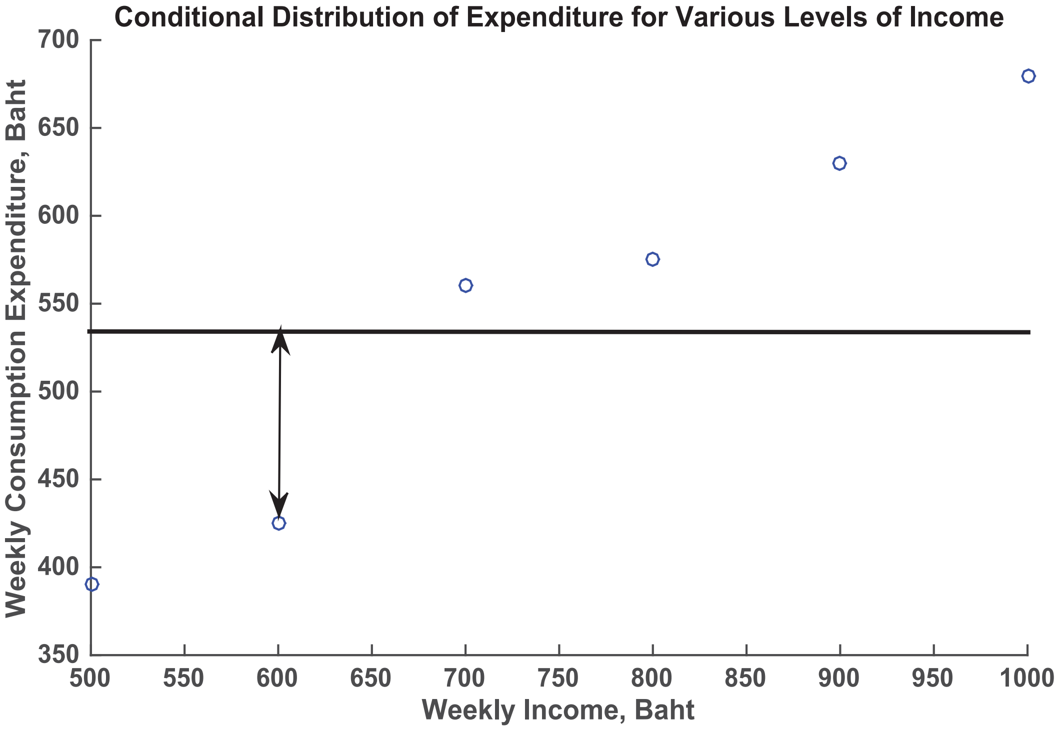
What will happen if we set the estimated Y to be $\bar{Y}$?

Table 3.3: Estimating the expenditure of the household

Family Number (i)	Actual Y_i	Estimate $\hat{Y}_i = \bar{Y}$	**Error in Estimation** $Y_i - \bar{Y}$	**Errors Squared** $(Y_i - \bar{Y})^2$
1	390	543	-153	23460.03
2	425	543	-118	13963.36
3	560	543	17	283.36
4	575	543	32	1013.36
5	630	543	87	7540.03
6	679	543	136	18450.69
Sum	3259	3259	0	64710.83

We can see all this graphically:

Figure 3.8: Graphic Representation



Question: Can we determine the total estimation error for this sample data?

Answer: Yes, we can calculate the total (combined) amount of estimation error for all observations in the sample when **using the mean as the estimate** as following:

$$TSS = \sum (Y_i - \bar{Y})^2$$

It is called the total sum of squares (TSS) which is the total variation of the actual Y values about their sample mean.

Since our objective in estimation is to minimize error (maximize precision), we need to cut down the amount of the estimation error (TSS).

We can achieve this by using information about other variables suspected to be strong predictors (strongly related to) the expenditure of the families.

We now can attempt to estimate the expenditure from the information on the income level of the family, rather than from its own mean.

Table 3.4: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$
1	390	500	-250	-153.17	38291.67	62500
2	425	600	-150	-118.17	17725.00	22500
3	560	700	-50	16.83	-841.67	2500
4	575	800	50	31.83	1591.67	2500
5	630	900	150	86.83	13025.00	22500
6	679	1000	250	135.83	33958.33	62500
Sum	3259	4500	0	0	103750	175000

From the table 8, we can calculate the simple regression as following:

Figure 3.9: Breakdown of the variation of Y_i into two components

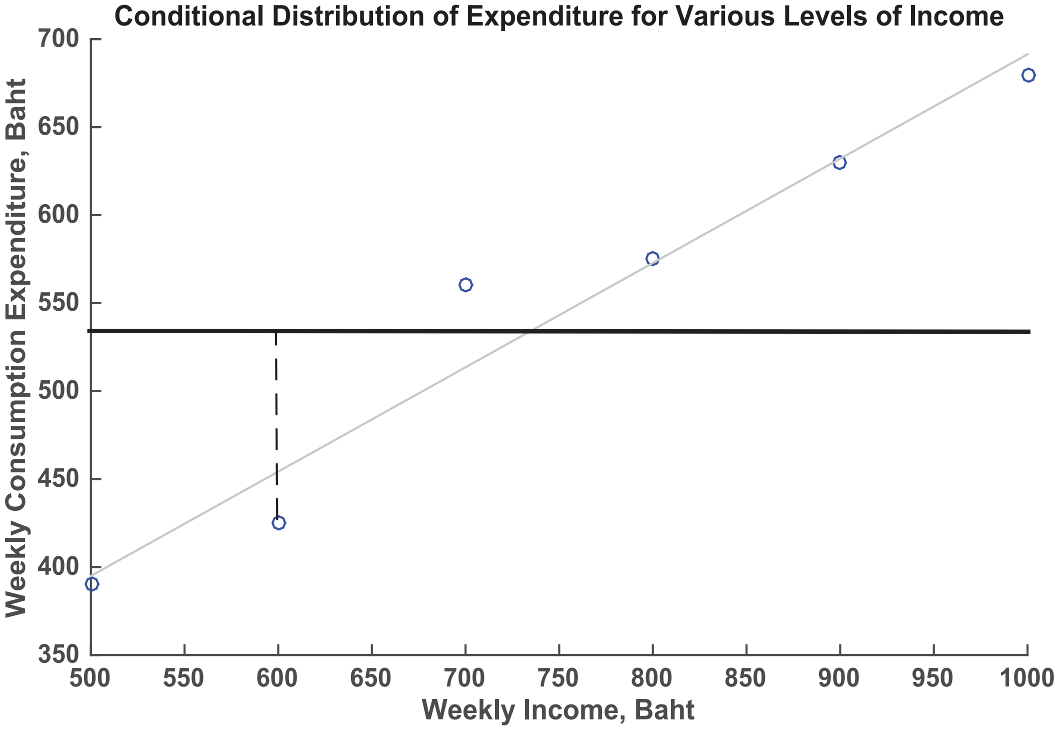


Table 3.5: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$	Residual squared $(Y-\hat{Y})^2$
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90

From the table 9, we can calculate the estimation error we have committed by using the regression line as:

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum \hat{u}_i^2$$

where RSS stands for the residual sum of squares. which is the unexplained variation of the Y values about the regression line.

Total Baseline Error using the mean (SS Total) =

New or Remaining Error (SS Error or SS Residual) =

QUESTION: How much of the original estimation error have we explained away (eliminated) by using the regression model (instead of the mean)?

ANS

QUESTION: What % of estimation error have we explained (eliminated by using the regression model?

ANS

QUESTION: What does the remaining% represent?

ANS
Percent of variation (differences) in expenditures that can be accounted for by: (a) all other potential predictors not included in the model, beyond income levels, and (b) unexplainable random/chance variations.

$$r^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

♣ r^2 is a measure of our success regarding accuracy of our estimation effort.

♣ r^2 = % of estimation error that we have been able to explain away by using the regression model, instead of using the mean.

♣ r^2 indicates how much better we can predict Y from information about Xs, rather than from using its own mean.

♣ r^2 = % of differences (variations) in Y values that is explained by (attributable to) differences in X values.

4. Classical Normal Regression Model (CNLRM)

We know that the classical theory of statistical inference consists of:

1. Estimation

We have covered this topic since we were able to estimate the parameters β_1,β_2, and σ^2 by using the method of OLS.

We also proved that these estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ satisfy several desirable statistical properties, such as unbiasedness, minimum variance, and linearity (BLUE property).

However, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ change their values from sample to sample. The following tables show the two different sets of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ depending on the two different sample data.

Table 4.1: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$	Residual squared $(Y-\hat{Y})^2$
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90

If we use this sample data. We can estimate:

$$\hat{\beta}_1 = 98.524$$

$$\hat{\beta}_2 = 0.593$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-2} = \frac{3201.90}{6-2} = 800.476$$

Table 4.2: Estimating the expenditure of the household with income with another sample data

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y - \hat{Y}$	Residual squared $(Y - \hat{Y})^2$
1	360	500	325.71	64.29	4132.65
2	390	600	406.43	18.57	344.90
3	440	700	487.14	72.86	5308.16
4	575	800	567.86	7.14	51.02
5	670	900	648.57	-18.57	344.90
6	730	1000	729.29	-50.29	2528.65
Sum	3165	4500	0	0	12710.29

If we use this sample data. We can estimate:

$$\hat{\beta}_1 = -77.857$$

$$\hat{\beta}_2 = 0.807$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{12710.29}{6-2} = 3177.571$$

From the example, you can easily see that these estimators are **RANDOM VARIABLES**. Therefore, we need to learn another part of statistical inference which is called **Hypothesis Testing.**

2. Hypothesis Testing

The main objective is to find out how close of $\hat{\beta}_1$ and $\hat{\beta}_2$ to the true β_1 and the true β_2, respectively. Also, we would like to see how close of $\hat{\sigma}^2$ compared to the true σ^2.

To achieve this goal, we need to know the probability distributions of $\hat{\beta}_1$, $\hat{\beta}_2$,and $\hat{\sigma}^2$.
Consider the estimator of β_2:

$$\hat{\beta}_2 = \sum k_i Y_i$$

We can write the above equation as:

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i)$$

From this equation, the probability distribution of $\hat{\beta}_2$ will depend on the assumption made about the probability distribution of u_i

4.1 The Normality Assumption for u_i

In the classical normal linear regression model (CNLRM), we assume that each u_i is distributed normally :

$$u_i \sim N(0, \sigma^2)$$

where

Mean:

$$E(u_i) = 0$$

Variance:

$$E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$$

$$cov(u_i, u_j) = E\left\{[u_i - E(u_i)][u_j - E(u_j)]\right\} = E(u_i U_j) = 0$$

Therefore,

$$u_i \sim N(0, \sigma^2)$$

Also, u_i and u_j are not only uncorrelated but also independently distributed.

we can then write the above equation as:

$$u_i \sim NID(0, \sigma^2)$$

where NID stands for normally and independently distributed.

4.2 Properties of OLS estimators under the normality assumption

1. They are unbiased.

2. They have minimum variance.

3. By 1+2 properties, they are minimum-variance unbiased, or efficient estimators.

4. $\hat{\beta}_1$ is normally distributed with:

$$\text{Mean:} E(\hat{\beta}_1) = \beta_1$$

$$\mathrm{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2$$

Therefore,

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

By the properties of the normal distribution, we can:

5. $\hat{\beta}_2$ is normally distributed with

$$\text{Mean:} E(\hat{\beta}_2) = \beta_2$$

$$\mathrm{var}(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$$

or more compactly

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$$

then we can define the standard normal distribution as

6. $(n-2)(\hat{\sigma}^2/\sigma^2)$ is distributed as the χ^2 (chi-square) distribution with (n-2) df.

7. $(\hat{\beta}_1, \hat{\beta}_2)$ are distributed independently of $\hat{\sigma}^2$

8. $\hat{\beta}_1$ and $\hat{\beta}_2$ have the minimum variance in the entire class of unbiased estimators, whether linear or not.

9. we can find out the probability distribution of Y_i as following: