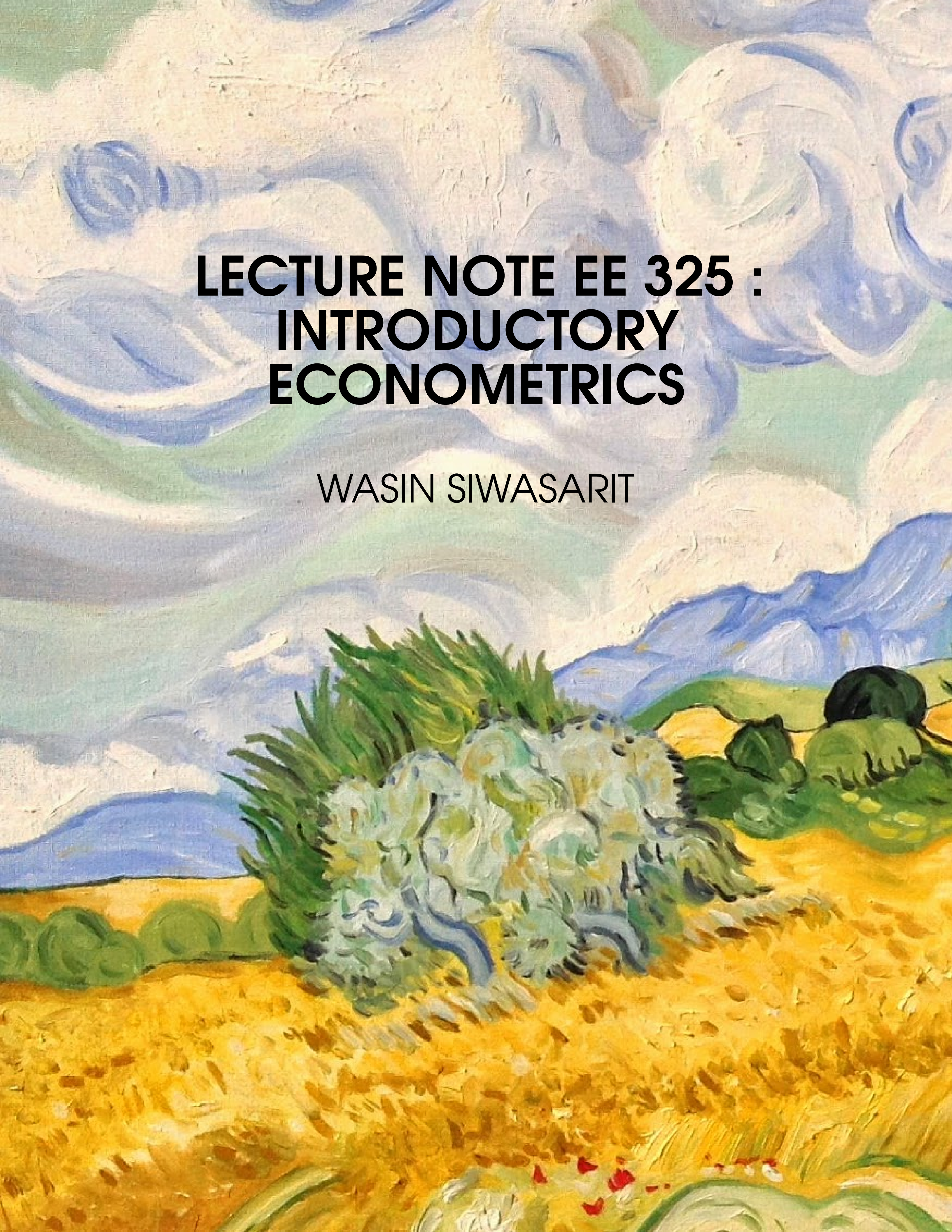


LECTURE NOTE EE 325 : INTRODUCTORY ECONOMETRICS

WASIN SIWASARIT

PROJECT 2016, FACULTY OF ECONOMICS, THAMMASAT UNIVERSITY.

Academic Year 2016

1. Introduction

1.1 Review of Some Statistical Concepts

The notation $\sum$ (sigma), in mathematical term, denotes the **summation**

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{1.1}$$

The noteworthy properties of summation include:

1. $\sum_{i=1}^{n} k = nk$

2. $\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i$, where k is a constant term.

3. $\sum_{i=1}^{n}(a+bx_i) = na + b\sum_{i=1}^{n} x_i$, where a and b are constants.

4. $\sum_{i=1}^{n}(X_i+Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$.

Multiple summation is the summation of variable that is in the form of matrix, shown as,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} = \sum_{i=1}^{n}(x_{i1}+x_{i2}+\ldots+x_{im}) = (x_{11}+x_{21}+\ldots+x_{n1})+(x_{12}+x_{22}+\ldots+x_{n2})+\ldots(x_{1m}+x_{2m}+\ldots+x_{nm}) \tag{1.2}$$

where

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix}_{nXm}$$

The significant properties of multiple summations are:

1. $\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} X_{ij}$

2. $\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j = \sum_{i=1}^{n} X_i \times \sum_{j=1}^{m} Y_j$

3. $\sum_{i=1}^{n}\sum_{j=1}^{m} (X_{ij} + Y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}$

4. $(\sum_{i=1}^{n} X_i)^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_i X_j$

The product operator $\prod$ is defined as:

$$\prod_{i=1}^{n} x_i = x_1 * x_2 \ldots * x_n \tag{1.3}$$

1.2 Experiment

Sample space is the set of all possible results of an experiment. For example, if you toss the coin twice, all feasible outcomes are composed of head twice, head followed by tail, tail followed by head, and tail twice. Let H and T denotes head and tail, respectively. The sample space can be written as,

$$SS = \{HH, HT, TH, TT\}$$

Sample Point is the member of sample space, eg. the event that head occurs twice from tossing a coin twice. Specifically, sample point is,

$$SP = HH \text{ or } HT \text{ or } TH \text{ or } TT$$

Events are the set of specific consequences of the experiment such as the events that head occurs twice. Events are the subset of sample space.

$$A = \text{the event that head occurs twice} = \{HH\}$$

Events are **mutually exclusive**, if the occurrence of one event makes no other events in sample space possible. As an illustration, for the experiment of tossing two coins once, let C be the event that both turn head and D be the event that both turn tail. Since C and D cannot happen at the same time, these two events are said to be mutually exclusive. Another example is the experiment of drawing one card from the standard 52-card deck, let E be the event that the rank of card is King and F be the event that suit of card is Clubs. As the event E and F can occur simultaneously, namely the King of Clubs, the two events are not mutually exclusive.

Events are **collectively exhaustive** if they cover all possible outcomes in the sample space. With the experiment of tossing the coin twice, let A be the event that head appears twice, B be the event that tail appears twice, and C be the event that head and tail each appear once. In this case, A, B and C are collectively exhaustive since all events cover all possible results from sample space; that is, HH, HT, TH and TT.

1.3 Probability and Random Variable

Probability is the possibility that any event will occur, given some specific sample space.

Let A be the event occurring in the given sample space and $P(A)$ be the probability that A will happen. Then, $P(A)$ is defined as;

$$P(A) = \frac{\text{the number of times the event A will occur}}{\text{the number of all possible outcomes in sample space}} \tag{1.4}$$

For instance, to draw one card from the standard 52-card deck, let A be the event that the rank of card is 2. Times the event will occur is 4 and the amount of all possible outcomes is 52; hence, the probability of A is $\frac{4}{52}$ or $\frac{1}{13}$.

Some properties of probability are;

1. $0 \leq P(A) \leq 1$

2. If A, B and C are exhaustive set, then,

   $P(A) + P(B) + P(C) = 1$

3. If A, B and C are mutually exclusive, then,

   $P(A + B + C) = P(A) + P(B) + P(C)$

Suppose that the results of an experiment are in the form of value, the variable, whose value is determined by one of those results, is known as **Random Variable**. Random variable can be either **discrete** or **continuous value**.

For discrete random variable, the example is the sum of the values on the face of two dice, when rolling two dice once. In other word, the obtained sum will range from 2 to 12, and it is impossible to get 2.5 or 3.5.

For continuous random variable, the example is the height of the high-school student, constricted to the range from 160 to 180 centimetres. It can be seen that the value of the height need not be the integers and can take the value of 160.5 or 160.52 centimetres.

These two distinct characteristics of random variable enable us to classify them into different probability density functions, which would be stated in section 1.4.

1.4 Probability Density Function

As the value of random variable depends on an experiment, the **probability density function** would portray the overall image of possible random results. The type of the probability density function relies on the characteristics of the random variable. In this section, many important types are discussed.

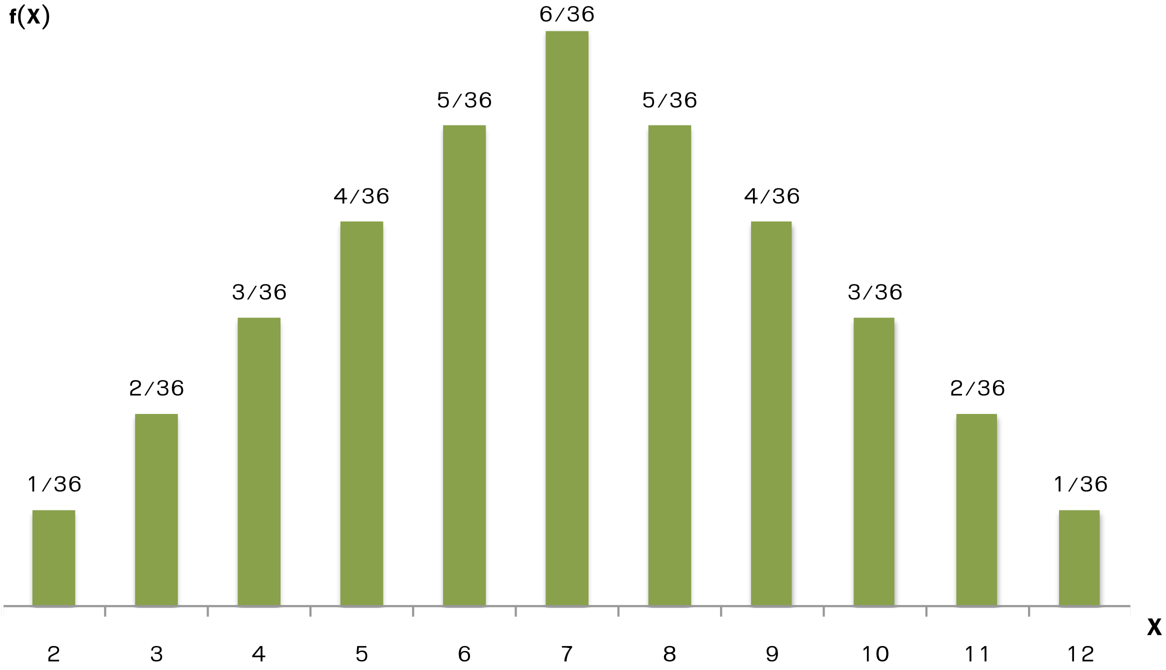
1.4.1 Probability Density Function for Discrete Random Variable

Let X be the discrete random variable with the value $x_1, x_2, \ldots, x_n$ and we get,

$$\begin{aligned} f(x) &= P(X = x_i) \quad \text{for} \quad i = 1,2,\ldots,n \\ f(x) &= 0 \quad \text{for} \quad x \neq x_i \end{aligned}$$

Example: Let X be random variable of the sum of values on the face of two dices. The value might be 2 or 12, that is the value from both rolling round is 1 or 6, respectively. The Figure 1 summarizes all possible results#

Figure 1: Probability Density function of the Sum of Values on the Side of the Dice, Obtained from Rolling the Dice Twice



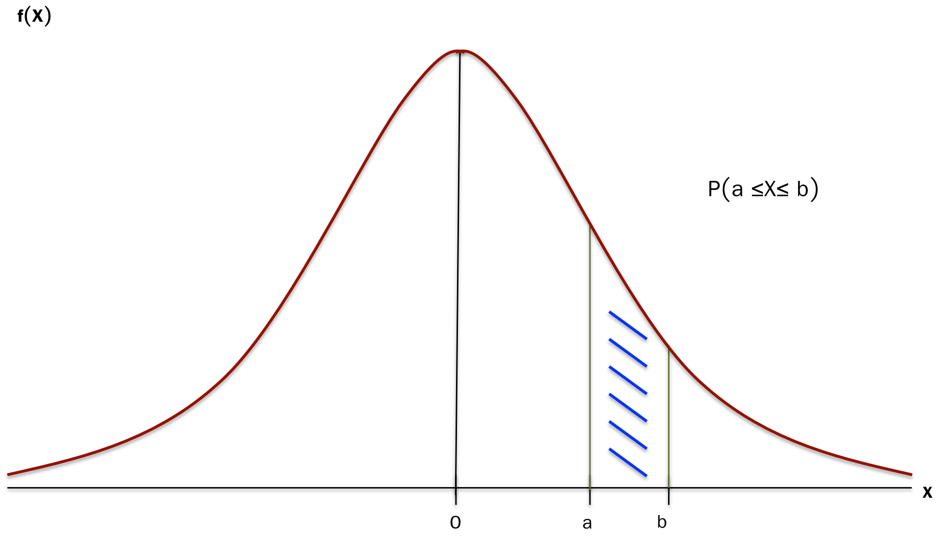
1.4.2 Probability Density Function for Continuous Random Variable

Let X be the continuous random variable. The probability density function of X satisfies the three following conditions.

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\int_{a}^{b} f(x)dx = P(a \leq x \leq b)$

Figure 2 exhibits the probability density function for the continuous random variable, where the area under the curve represents the probability that the variable will lay on that range. Specifically, $P(a \leq X \leq b)$ means the probability that X will take the value between a and b.

Figure 2: Probability Density Function for Continuous Random Variable



1.4.3 Joint Probability Density Function

In this section, only **joint probability density function** for discrete variable is discussed. Let X and Y be discrete random variables. The joint probability density function, identifying the probability that X and Y happen simultaneously, is written as,

$$f(X,Y) = P(X = x \text{ and } Y = y)$$

Example: The following table explains the joint probability density function.

Table 1: The table illustrating the joint probability density function of X and Y

		X		
		-1	0	1
	1	0.11	0.08	0.05
Y	2	0.09	0.05	0.03
	3	0.35	0.07	0.17

According to the table 1, the probability that random variable X will be 0 and random variable Y will be 3 is 0.07 or 7 percent. In mathematical term, it can be written as $f(X = 0, Y = 3) = 0.07$.

1.4.4 Marginal Probability Density Function

The above joint probability density function $f(X,Y)$ shows the joint distribution of two variables. On the other hand, **marginal probability density function** with respect to joint probability function, displays the probability density function of single variable like $f(X)$, $f(Y)$, which can be derived from;

$$\begin{array}{lclll} f(X) & = & \sum_Y f(X,Y) & \text{called} & \text{marginal PDF of X} \\ f(Y) & = & \sum_X f(X,Y) & \text{called} & \text{marginal PDF of Y} \end{array}$$

where $\sum_Y$ or $\sum_X$ means the summation of probability over all values of X and Y respectively.

Example: According to Table 2 above, marginal PDF of X is obtained from

$$\begin{aligned} (X=-1) &= \\ &= \\ &= \\ &= \\ f(X=0) &= \sum_Y f(X=0,Y) \\ &= f(X=0,Y=1)+f(X=0,Y=2)+f(X=0,Y=3) \\ &= 0.08+0.05+0.07 \\ &= 0.20 \\ f(X=1) &= \sum_Y f(X=1,Y) \\ &= f(X=1,Y=1)+f(X=1,Y=2)+f(X=1,Y=3) \\ &= 0.05+0.03+0.17 \\ &= 0.25 \end{aligned}$$

marginal PDF of Y is obtained from

$$\begin{aligned} f(Y=1) &= \\ &= \\ &= \\ &= \\ f(Y=2) &= \sum_X f(X,Y=2) \\ &= f(X=-1,Y=2)+f(X=0,Y=2)+f(X=1,Y=2) \\ &= 0.09+0.05+0.03 \\ &= 0.17 \\ f(Y=3) &= \sum_X f(X,Y=3) \\ &= f(X=-1,Y=3)+f(X=0,Y=3)+f(X=1,Y=3) \\ &= 0.35+0.07+0.17 \\ &= 0.59 \end{aligned}$$

According to the calculation above, the result can be summarized into Table 2.

Table 2 shows joint probability of random variable X and Y

		X			
		-1	0	1	
	1	0.11	0.08	0.05	$f(Y=1)$ =
Y	2	0.09	0.05	0.03	$f(Y=2)$ =
	3	0.35	0.07	0.17	$f(Y=3)$ =
		$f(X=-1)$ =	$f(X=0)$ =	$f(X=1)$ =	$f(X) =$ $f(Y) =$

1.4.5 Conditional Probability Density Function

Conditional probability density function is the probability of one event given that some events have already occurred. The function is written as,

$$f(X|Y) = P(X = x|Y = y)$$

This function can be obtained from the joint probability density function through,

$$f(X|Y) = \frac{f(X,Y)}{f(Y)}$$

Example: According to Table 2, find $f(X=1|Y=2)$ and $f(Y=2|X=0)$

$$\begin{aligned}
(X = 0|Y = 1) &= \\
&= \\
&= \\
&= \\
(Y = 2|X = 0) &= \\
&= \\
&=
\end{aligned}$$

Example: Let event A be tossing the dice once and the point is odd number and B be the tossing the dice once and the point is at least 5. Find the probability that the point coming up is odd given that the point has to be at least 5.

Answer A and B will occur simultaneously if the point from tossing the dice is 5; so, the joint probability of A and B is $\frac{1}{6}$. The probability that B occurs is $\frac{2}{6}$. Hence, the conditional probability of A given B is

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}\#$$

1.4.6 Statistical Independence

Two random variables are **independent** if the resulting value of one variable does not affect the resulting value of the other; namely,

$$f(X,Y) = f(X)f(Y)$$

Example: Consider Mr. Ake's expenditure for a meal and the Miss Somsri's expenditure for a dessert. Given that they do not know each other, the realization of Mr. Ake's expenditure does not imply the realization of Miss Somsri's expenditure. We can, thus, conclude that the expenditures of these two people are independent#

Example: Consider drawing cards sequentially from the standard 52-card deck without putting it back into the deck. Once the first card is drawn, the probability of drawing the second card will be influenced because the amount of cards in the deck is reduced. In this case, it can be concluded that drawing the first and second card are not independent#

1.5 Expectation, Variance, Covariance and Correlation

1.5.1 Mean or Expected Value

Because the value of random variable hinges on the value of random results of experiment which cannot be determined certainly, statisticians have invented the measures of central tendency of the random variable. One of them is **expected value**, indicating the mean of the random variable.

For discrete random variable, the expected value is calculated by;

$$E(X) = \sum_{i=i}^{n} x_i f(x_i) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n)$$

For continuous random variable, the expected value is calculated by,

$$E(X) = \int_a^b x f(x) dx$$

where;

$E(X)$ is the measure of central tendency of random variable, resulting from repeated trial of experiment.

$\sum_{i=i}^{n} x_i f(x_i)$ is the average of random variable weighted by the probability corresponding to each value.

a and b are the lowest and highest constant possible respectively.

Example: Find the expected value of rolling two dice once (Figure 1)

Example:

Crucial properties of expected value include:

1. $E(b) = b$

2. $E(aX + b) = aE(X) + b$

3. $E(XY) = E(X)E(Y)$; given that X and Y are independent

4. $E(g(X)) = \sum_x g(X) f(X)$

where a and b are constant.

Conditional expectation value is the expectation value of random variable under some conditions such as expected value of X conditional on Y or $E(X|Y = 5)$

Let $f(X,Y)$ be the joint probability function of X and Y. The expectation of X conditional on some value of Y is defined as,

For discrete random variable $E(X|Y = y) \quad = \quad \sum_X X_i f(X|Y = y)$

For continuous random variable $E(X|Y = y) \quad = \quad \int_{-\infty}^{\infty} X_i f(X|Y = y)$

Example

1.5.2 Variance

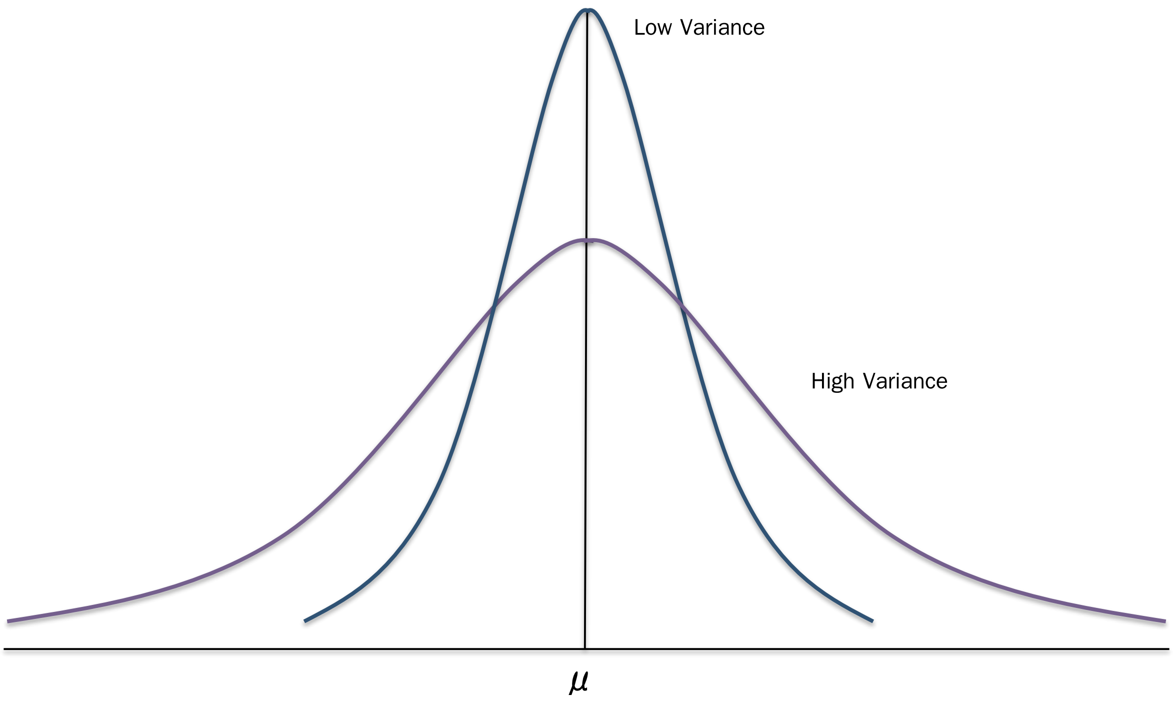
Variance is the measure of dispersion of the value of variable around the expected value. The higher the variance, the more dispersing the random variable (Figure 3). If X is the random variable with expected value μ, we get;

$$Var(X) = \sigma_X^2 = E[X - E(X)]^2 = E(X)^2 - \mu^2 \tag{1.5}$$

From,

$$\begin{aligned} Var(X) &= \sigma_X^2 \\ &= E[X - E(X)]^2 \\ &= E[X^2 - 2XE(X) + (E(X))^2] \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X)^2 - \mu^2 \end{aligned}$$

Figure 3: Distribution of Random Variables with Different Variance



Important properties of expected value include;

1. $Var(b) = 0$

2. $Var(aX+b) = a^2Var(X)$

3. $Var(X \pm Y) = Var(X) + Var(Y)$; given that X and Y are independent

4. $Var(aX \pm bY) = a^2Var(X) + b^2Var(Y)$

where *a* and *b* are constant.

1.5.3 Conditional Variance

The conditional variance of X is given Y = y is defined as following:

$$\begin{aligned} var(X|Y=y) &= E\left\{[X-E(X|Y=y)]^2|Y=y\right\} \\ &= \sum_x [X-E(X|Y=y)]^2 f(x|Y=y) \\ &= \int_{-\infty}^{\infty} [X-E(X|Y=y)]^2 f(x|Y=y)dx \end{aligned} \tag{1.6}$$

Example

Properties of conditional expectation and conditional variance

1.5.4 Covariance

Theorem. Let X and Y be two random variables with means μ_x and μ_y, respectively. Then, we can define the covariance between these two variables as following:

$$cov(X,Y) = E\left\{(X-\mu_x)(Y-\mu_y)\right\} = E(XY) - \mu_x\mu_y \tag{1.7}$$

If X and Y are continuous random variables we can calculate their cov(X,Y):

$$\begin{aligned} cov(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_x)(Y-\mu_y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}XYf(x,y)dxdy - \mu_x\mu_y \end{aligned} \tag{1.8}$$

Properties of Covariance

1. If X and Y are independent, the covariance between X and Y is zero.

Proof:

2. $cov(a+bX, c+dY) = bd * cov(X,Y)$, where a,b,c, and d are constants.

Proof:

Example Suppose the join PDF of random variables X and Y can be represented as in the below table. What is the covariance between X and Y?

			X		
		1	2	3	
Y	1	0.25	0.25	0	$f(Y=1)$ =
	2	0	0.25	0.25	$f(Y=2)$ =
		$f(X=1)$ =	$f(X=2)$ =	$f(X=3)$ =	$f(X)=$ $f(Y)=$

Next, we will turn our attention to seeing how we can apply the covariance to calculate the correlation between the random variables X and Y

1.5.5 Correlation

When we calculate the covariance of X and Y, it reflects the units of both random variables. However, it is useful to have **a dimensionless measure of dependency** by calculating the correlation instead.

Definition Let X and Y be any two random variables (discrete or continuous) with standard deviation σ_X and σ_Y, respectively. The **correlation coefficient** of X and Y, denoted **corr(X,Y)** or ρ_{XY} (the greek letter "rho") is defined as:

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(x,y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Example Suppose the join PDF of random variables X and Y can be represented as in the below table. What is the correlation between X and Y?

		X			
		1	2	3	
	1	0.25	0.25	0	$f(Y=1)$ =0.5
Y	2	0	0.25	0.25	$f(Y=2)$ =0.5
		$f(X=1)$ = 0.25	$f(X=2)$ = 0.5	$f(X=3)$ =0.25	$f(X)=1$ $f(Y)=1$

From the definition, ρ_{XY} is measure of linear association between two random variables. The value of ρ lies between -1 and +1, $-1 \leq \rho_{XY} \leq +1$. We can interpret the value of correlation as:

► If $\rho_{XY} = 1$, then X and Y are perfectly, positively, linearly correlated.
► If $\rho_{XY} = -1$, then X and Y are perfectly, negatively, linearly correlated.
► If $\rho_{XY} = 0$, then X and Y are completely,un-linearly correlated. This means that X and Y may correlated in some other manner i.e. a parabolic manner., but NOT in a linear manner
► If $\rho_{XY} \leq 0$, then X and Y are positively, linearly correlated, but NOT perfectly.
► If $\rho_{XY} \geqq 0$, then X and Y are negatively, linearly correlated, but NOT perfectly.

Theorem. If X and Y are independent random variables, then:

$$corr(X,Y) = cov(X,Y) = 0$$

Example: Let X = the outcome of a fair, black, 6-sided die.
Let Y =outcome of a fair, red, 4-sided die.
What is the covariance of X and Y? What is the correlation of X and Y?

NOTE: The converse of the theorem is NOT NECESSARILY CORRECT!

Example: Let X and Y be two discrete random variables with the following join PDF:

			X		
		0	1	2	
	0	0	0.20	0.10	$f(Y=0)$ =
Y	1	0.20	0.40	0	$f(Y=1)$ =
	2	0.10	0	0	$f(Y=2)$ =
		$f(X=0)$ =	$f(X=1)$ =	$f(X=2)$ =	

What is the correlation between X and Y? And, are X and Y independent?

1.5.6 Variances of Correlated Variables

Let X and Y be two random variables, then

$$
\begin{aligned}
var(X+Y) &= var(X)+var(Y)+2cov(X,Y) \\
&= var(X)+var(Y)+2\rho\sigma_x\sigma_y \\
var(X-Y) &= var(X)+var(Y)-2cov(X,Y) \\
&= var(X)+var(Y)-2\rho\sigma_x\sigma_y
\end{aligned}
\tag{1.9}
$$

The generalized result:
Let $\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$, then the variance of the linear combination $\sum X_i$ is:

$$
\begin{aligned}
var\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} var(X_i) + 2\sum_{i<j}\sum cov(X_i, X_j) \\
&= \sum_{i=1}^{n} var(X_i) + 2\sum_{i<j}\sum \rho_{ij}\sigma_i\sigma_j
\end{aligned}
\tag{1.10}
$$

Example:
what is the var$(X_1 + X_2 + X_3)$?

1.5.7 Higher Moments of Probability Distributions

In the previous subsection, we have already discussed about mean, variance, and covariance as the measures of the first and second moments of univariate and multivariate PDFs. Besides the first two moments, we are occasionally interested in the higher moments such as the third and fourth moments which are normally applied in studying the "Shape" of the distribution. In general, the r^{th} moments about the mean is defined as

$$r^{th}\textbf{moment} : E(X-\mu)^r$$

By the definition of r^{th} moments, we can easily define the third and fourth moments as:

Third moment:

$$E(X-\mu)^3$$

Fourth moment:

$$E(X-\mu)^4$$

We can study the shape of the distribution by calculating **skewness** and **kurtosis**.

SKEWNESS is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

One measure of skewness is defined as:

$$\begin{aligned} S &= \frac{E(X-\mu)^3}{\sigma^3} \\ &= \frac{\text{third moment about the mean}}{\text{cube of the standard deviation}} \end{aligned} \tag{1.11}$$

KURTOSIS is a measure of the peakedness of the probability distribution of a real-valued random variable

We can also measure kurtosis as:

$$\begin{aligned} S &= \frac{E(X-\mu)^4}{\sigma^4} \\ &= \frac{\text{fourth moment about the mean}}{\text{square of the second moment}} \end{aligned} \tag{1.12}$$

♣ **Platykurtic (fat or short-tailed)** $\Longrightarrow$ PDFs with Kurtosis < 3

♣ **Leptokurtic (slim or long-tailed)** $\Longrightarrow$ PDFs with Kurtosis > 3

♣ **Mesokurtic (which is the normal distribution)** $\Longrightarrow$ PDFs with Kurtosis $= 3$

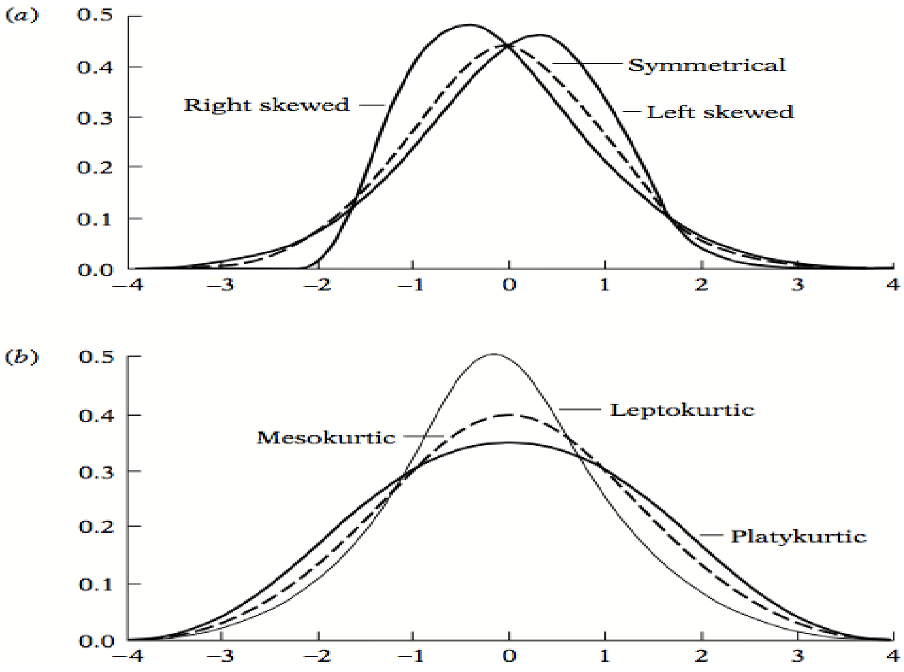


Figure 1.1: (a) Skewness; (b)Kurtosis

1.6 Some important probability distribution

1.6.1 Normal Distribution

A continuous random variable X has a normal distribution with mean μ and variance σ^2 if its probability density function (pdf) is

$$f(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \quad \text{where} \quad -\infty < x < \infty$$

NOTE: The normal distribution can be described by two parameters

- μ = The mean of the distribution.
- σ = The standard deviation of the distribution.

Therefore, changing the values of μ and σ alter the positions and shapes of the distributions.

If X is Normally distributed with mean μ and standard deviation σ, we can write it as:

$$X \sim N(\mu, \sigma^2)$$

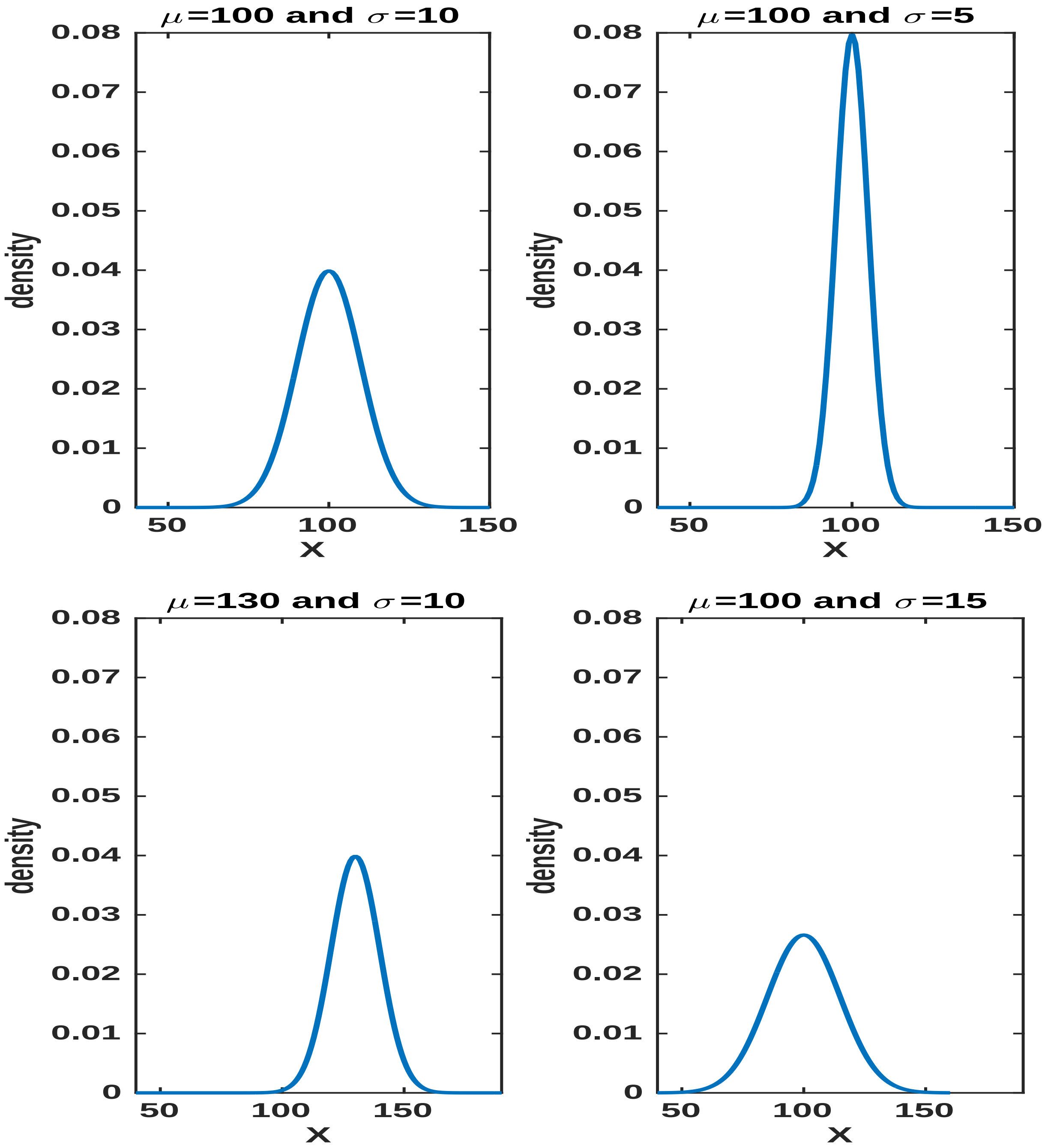


Figure 1.2: Compare the mean and standard deviation of the normal distribution

The properties of the normal distribution.

★ It is symmetrical around its mean value.

★ About 68 percent of the area under the normal distribution lies between the value $\mu \pm \sigma$
About 95 percent of the area under the normal distribution lies between the value $\mu \pm 2\sigma$
About 99.7 percent of the area under the normal distribution lies between the value $\mu \pm 3\sigma$ (as shown in figure 2)

★ We can convert the given normally distributed variable X with mean μ and σ^2 into the standardized normal variable Z by calculating Z where Z can be defined as:

$$Z =$$

With the standardized normal variable Z, we can rewrite the normal pdf as:

$$f(Z) =$$

In sum, you can see that we convert the given normally distributed variable X into the standardized normal variable by:
(i) Subtracting the mean μ
(ii)Dividing by the standard deviation σ

♡ Subtracting the mean re-centers the distribution on zero.

♡ Dividing by the standard deviation re-scales the distribution so it has standard deviation 1.

It should be remarked that it mean value is zero and its variance is unity for any standardized variable.

By convention, we can denote a normally distributed variable X with zero mean and unit variance as

$$X \sim N(0,1)$$

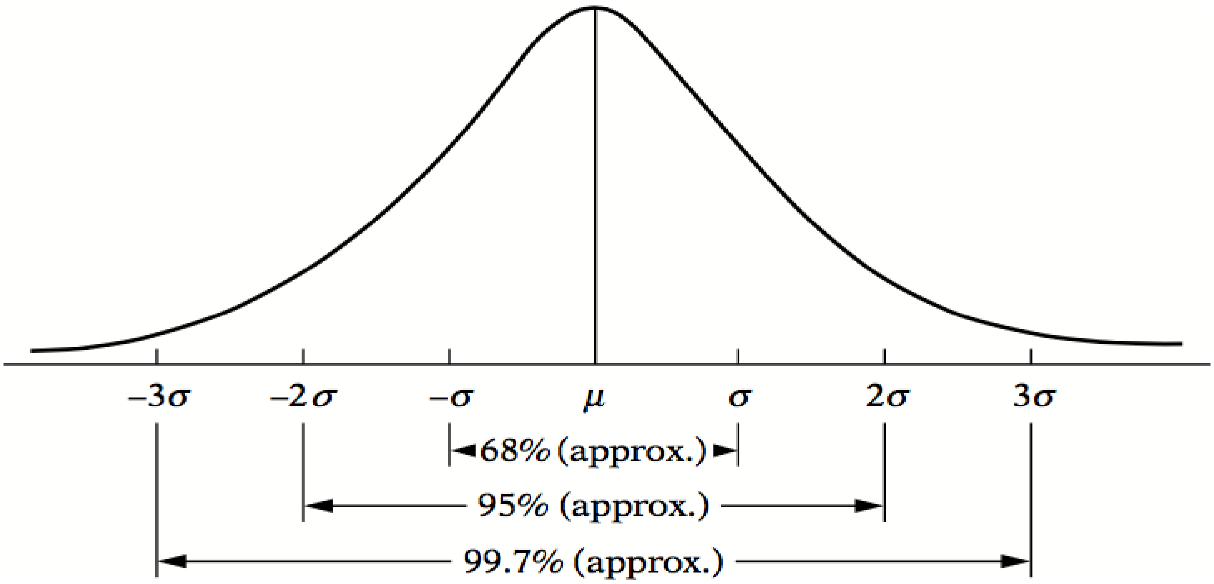


Figure 1.3: Areas under the normal distribution

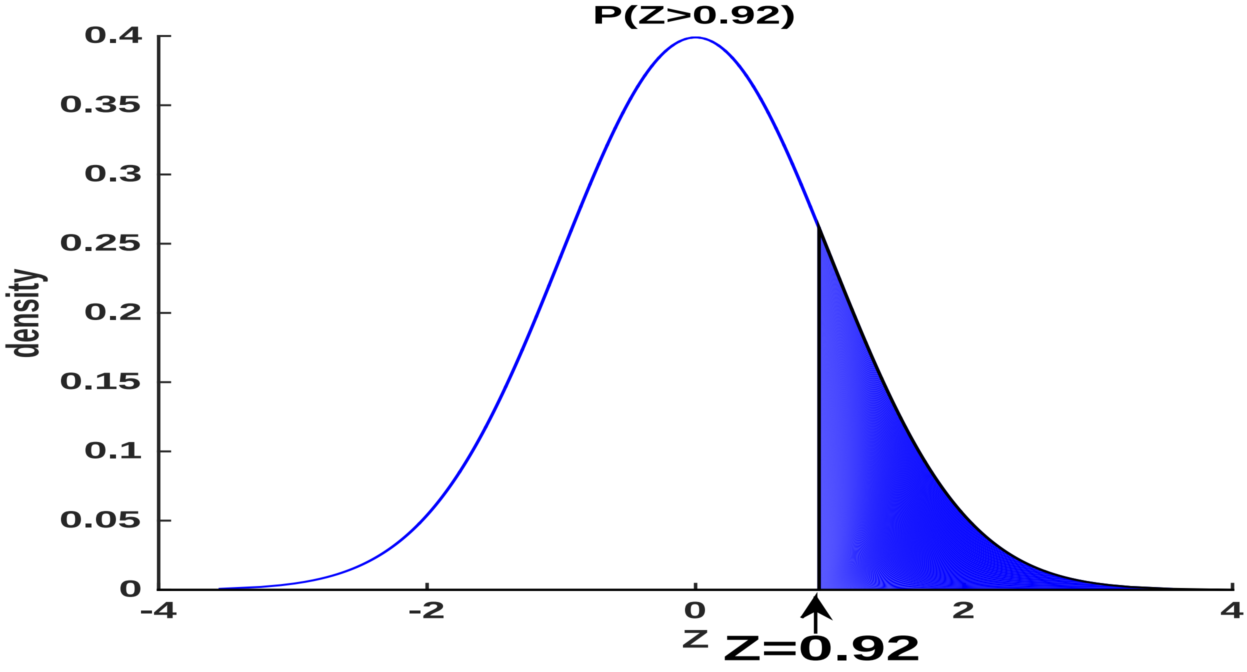


Figure 1.4: If Z $\sim$ N(0,1), the probability that P(Z>0.92)

Example If Z $\sim$ N(0,1) what is P(Z>0.92)?

Example If Z $\sim$ N(0,1) what is P(-0.64<Z<0.43)?

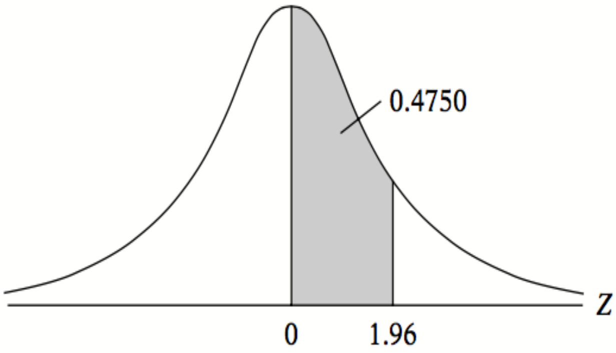
Example If X $\sim$ N(3500,500^2) what is P(X<3100)?

AREAS UNDER THE STANDARDIZED NORMAL DISTRIBUTION

Example

$\Pr(0 \leq Z \leq 1.96) = 0.4750$

$\Pr(Z \geq 1.96) = 0.5 - 0.4750 = 0.025$



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4454	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ and assume that X_1 and X_2 are independent. If we have the linear combination between X_1 and X_2 where we can write it as:

$$Y = aX_1 + bX_2,$$

where a and b are the constant terms. Then

$$Y \sim N\left[(a\mu_1 + b\mu_2), (a^2\sigma_1^2 + b^2\sigma_2^2)\right]$$

In other words, **a linear combination of normally distributed variables is itself normally distributed.**

Central limit theorem Let $X_1, X_2, ..., X_n$ denote n independent random variables and

$$X_i \sim N(\mu, \sigma)$$

Let $\bar{X}$=$\sum \frac{X_i}{n}$, then as n increases indefinitely (i.e, n $\rightarrow \infty$),

The third and fourth moments of the normal distribution:

Third moment: $E(X-\mu)^3 = 0$

Fourth moment: $E(X-\mu)^4 = 3\sigma^4$

1.6.2 The χ^2 (Chi-Square) Distribution

Let $Z_1, Z_2, ..., Z_k$ be **independent standardized normal variables**. Then the quantity

$$Z = \sum_{i=1}^{k} Z_i^2$$

is said to possess the χ^2 with k degree of freedom (df)

Properties of the χ^2 distribution are as follows:

1.The χ^2 distribution is a skewed distribution where the degree of the skewness depending on the df. As the number of df increases, the distribution becomes more symmetrical. For the df excess of 100, the variable

$$\sqrt{2\chi^2} - \sqrt{(2k-1)}$$

can be converted to a standardized normal variable, where k is the df.

2. The mean of the chi-square distribution is k, and its variance is 2k, where k is the df.

3. If Z_1 and $Z_2$are two independent chi-square variables with k_1 and k_2 df, then the sum of $Z_1 + Z_2$ is also a chi-square with $df = k_1 + k2$

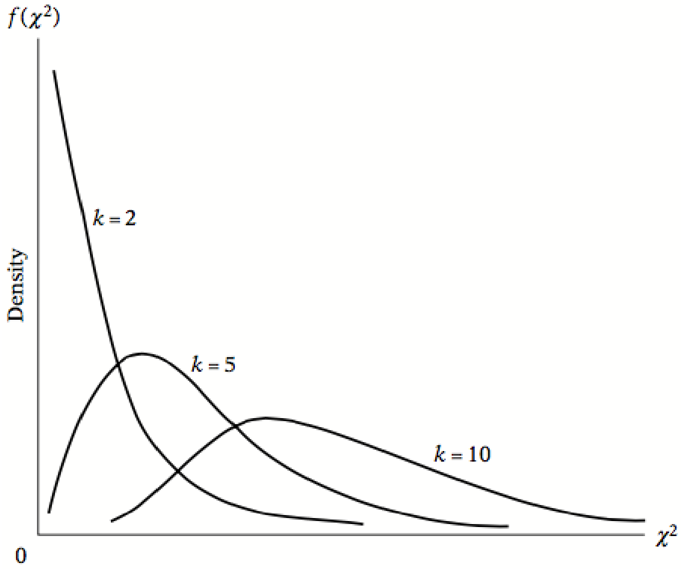


Figure 1.5: Density function of the χ^2 variable

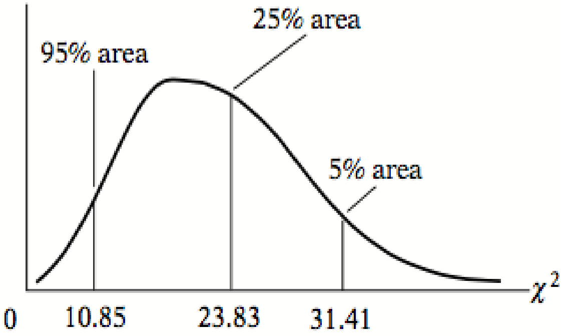
UPPER PERCENTAGE POINTS OF THE χ^2 DISTRIBUTION

Example

$\Pr(\chi^2 > 10.85) = 0.95$

$\Pr(\chi^2 > 23.83) = 0.25$ for df = 20

$\Pr(\chi^2 > 31.41) = 0.05$



Degrees of freedom \ Pr	.995	.990	.975	.950	.900
1	392704×10^{-10}	157088×10^{-9}	982069×10^{-9}	393214×10^{-8}	.0157908
2	.0100251	.0201007	.0506356	.102587	.210720
3	.0717212	.114832	.215795	.351846	.584375
4	.206990	.297110	.484419	.710721	1.063623
5	.411740	.554300	.831211	1.145476	1.61031
6	.675727	.872085	1.237347	1.63539	2.20413
7	.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100*	67.3276	70.0648	74.2219	77.9295	82.3581

*For df greater than 100 the expression $\sqrt{2\chi^2} - \sqrt{(2k-1)} = Z$ follows the standardized normal distribution, where k represents the degrees of freedom.

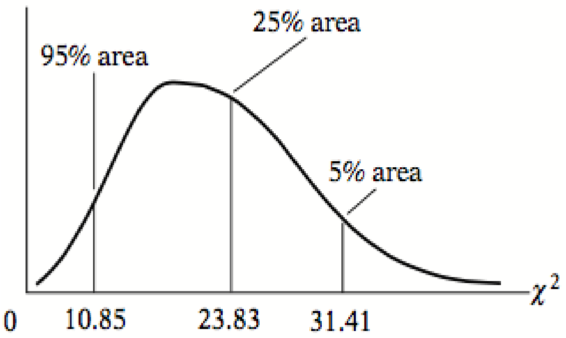
UPPER PERCENTAGE POINTS OF THE χ^2 DISTRIBUTION

Example

$\Pr(\chi^2 > 10.85) = 0.95$

$\Pr(\chi^2 > 23.83) = 0.25$ for df = 20

$\Pr(\chi^2 > 31.41) = 0.05$



.750	.500	.250	.100	.050	.025	.010	.005
.1015308	.454937	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
.575364	1.38629	2.77259	4.60517	5.99147	7.37776	9.21034	10.5966
1.212534	2.36597	4.10835	6.25139	7.81473	9.34840	11.3449	12.8381
1.92255	3.35670	5.38527	7.77944	9.48773	11.1433	13.2767	14.8602
2.67460	4.35146	6.62568	9.23635	11.0705	12.8325	15.0863	16.7496
3.45460	5.34812	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476
4.25485	6.34581	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777
5.07064	7.34412	10.2188	13.3616	15.5073	17.5346	20.0902	21.9550
5.89883	8.34283	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893
6.73720	9.34182	12.5489	15.9871	18.3070	20.4831	23.2093	25.1882
7.58412	10.3410	13.7007	17.2750	19.6751	21.9200	24.7250	26.7569
8.43842	11.3403	14.8454	18.5494	21.0261	23.3367	26.2170	28.2995
9.29906	12.3398	15.9839	19.8119	22.3621	24.7356	27.6883	29.8194
10.1653	13.3393	17.1170	21.0642	23.6848	26.1190	29.1413	31.3193
11.0365	14.3389	18.2451	22.3072	24.9958	27.4884	30.5779	32.8013
11.9122	15.3385	19.3688	23.5418	26.2962	28.8454	31.9999	34.2672
12.7919	16.3381	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185
13.6753	17.3379	21.6049	25.9894	28.8693	31.5264	34.8053	37.1564
14.5620	18.3376	22.7178	27.2036	30.1435	32.8523	36.1908	38.5822
15.4518	19.3374	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968
16.3444	20.3372	24.9348	29.6151	32.6705	35.4789	38.9321	41.4010
17.2396	21.3370	26.0393	30.8133	33.9244	36.7807	40.2894	42.7956
18.1373	22.3369	27.1413	32.0069	35.1725	38.0757	41.6384	44.1813
19.0372	23.3367	28.2412	33.1963	36.4151	39.3641	42.9798	45.5585
19.9393	24.3366	29.3389	34.3816	37.6525	40.6465	44.3141	46.9278
20.8434	25.3364	30.4345	35.5631	38.8852	41.9232	45.6417	48.2899
21.7494	26.3363	31.5284	36.7412	40.1133	43.1944	46.9630	49.6449
22.6572	27.3363	32.6205	37.9159	41.3372	44.4607	48.2782	50.9933
23.5666	28.3362	33.7109	39.0875	42.5569	45.7222	49.5879	52.3356
24.4776	29.3360	34.7998	40.2560	43.7729	46.9792	50.8922	53.6720
33.6603	39.3354	45.6160	51.8050	55.7585	59.3417	63.6907	66.7659
42.9421	49.3349	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900
52.2938	59.3347	66.9814	74.3970	79.0819	83.2976	88.3794	91.9517
61.6983	69.3344	77.5766	85.5271	90.5312	95.0231	100.425	104.215
71.1445	79.3343	88.1303	96.5782	101.879	106.629	112.329	116.321
80.6247	89.3342	98.6499	107.565	113.145	118.136	124.116	128.299
90.1332	99.3341	109.141	118.498	124.342	129.561	135.807	140.169

Source: Abridged from E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 8, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

1.6.3 Student's t Distribution

If Z_1 is a standardized normal variable and Z_2 is the chi-square distribution with k degree of freedom and is distributed independently of Z_1, then the Student's t distribution (t_k) with k degree of freedom can be represented as

$$\begin{aligned} t &= \frac{Z_1}{\sqrt{(Z_2/k)}} \\ &= \frac{Z_1\sqrt{k}}{\sqrt{Z_2}} \end{aligned} \tag{1.13}$$

Properties of the Student's t distribution are as follows:

1. The t distribution is symmetrical, BUT it is flatter than the normal distribution. However, as the df increase, the t distribution is converted to the normal distribution.

2. The mean of the t distribution is zero, and the variance is $\frac{k}{k-2}$

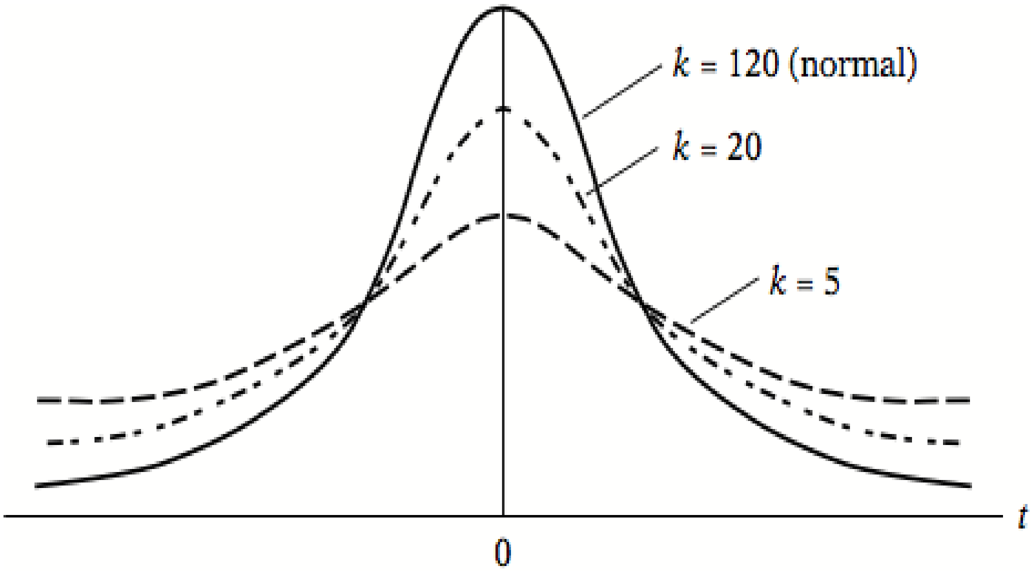


Figure 1.6: Density function of the student's t distribution

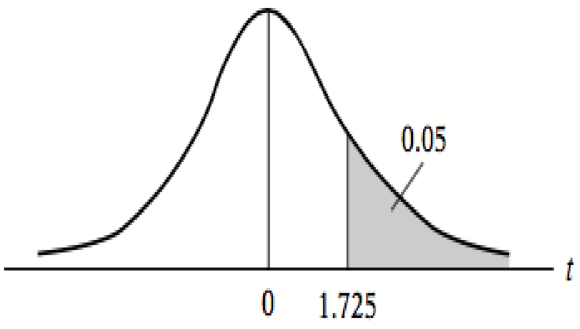
PERCENTAGE POINTS OF THE t DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$



Pr	0.25	0.10	0.05	0.025	0.01	0.005	0.001
df	0.50	0.20	0.10	0.05	0.02	0.010	0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

1.6.4 The F Distribution

If Z_1 and Z_2 are independently distributed chi-square variables with k_1 and k_2 df, respectively, the (Fisher's) F distribution with k_1 and k_2 df can be written as

$$F = \frac{Z_1/k_1}{Z_2/k_2}$$

The F distribution has the following properties:

1. The F distribution is skewed to the right, but as k_1 and k_2 become large, the F distribution is converted to normal distribution.

2.The mean value of ann F-distributed variable is $\frac{k_2}{(k_2-2)}$, and its variance is

$$\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$$

3. The square of a t-distributed random variable with k df is equivalent to an F distribution with 1 and k df.

$$t_k^2 = F_{1,k}$$

4. If the denominator df, k_2, is fairly large, we can get the following relationship

$$k_1 F \sim \chi^2_{k_1}$$

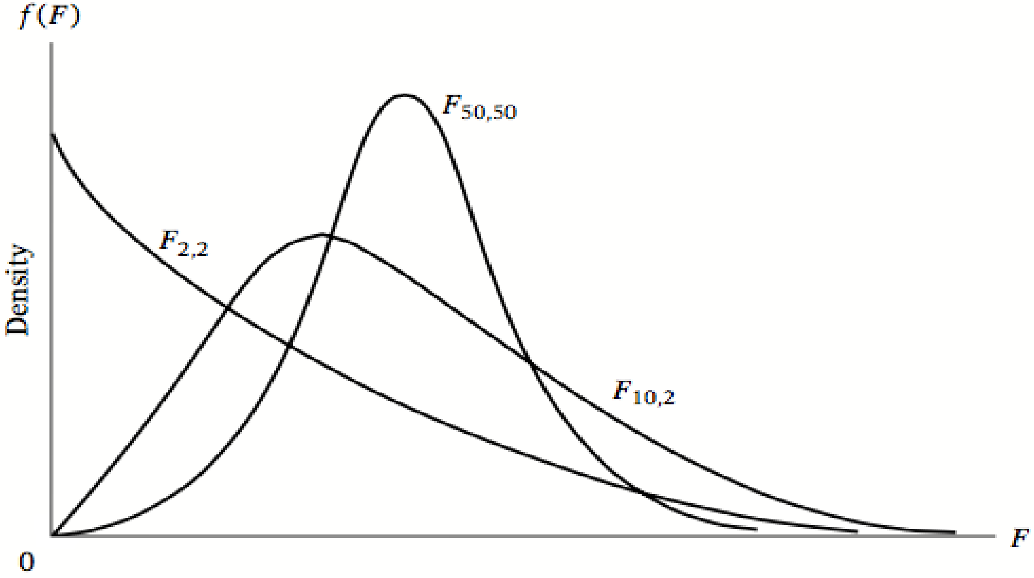


Figure 1.7: Density function of F distribution

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
Y= Weeky Family Expenditure	360	376	458	610	600	700
	313	475	422	468	531	679
	322	380	498	575	670	730
	310	382	560	542	630	591
	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

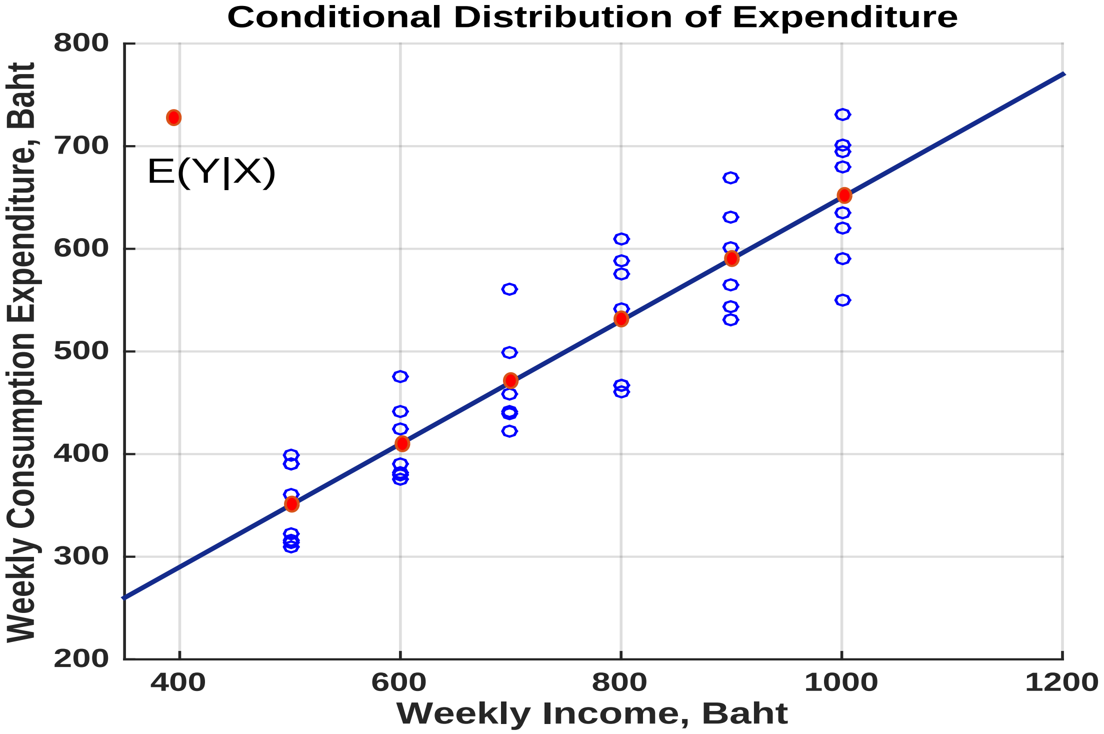
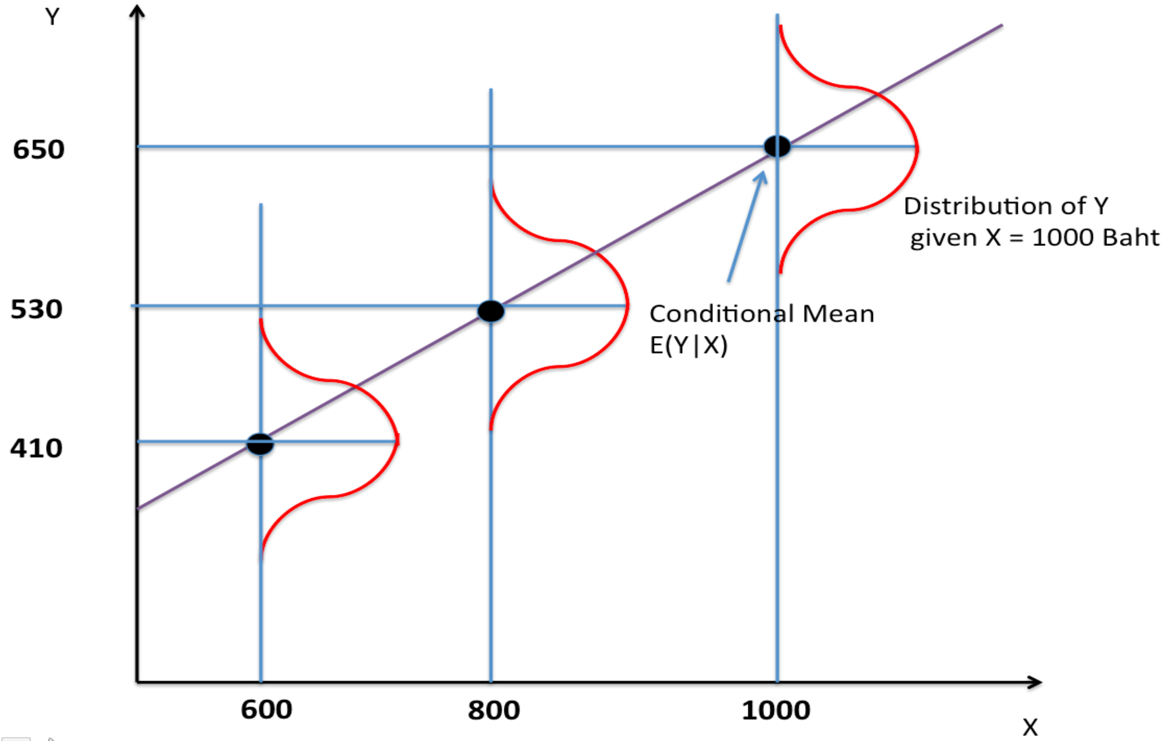


Figure 2.2: Population Regression Line (PRL)



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

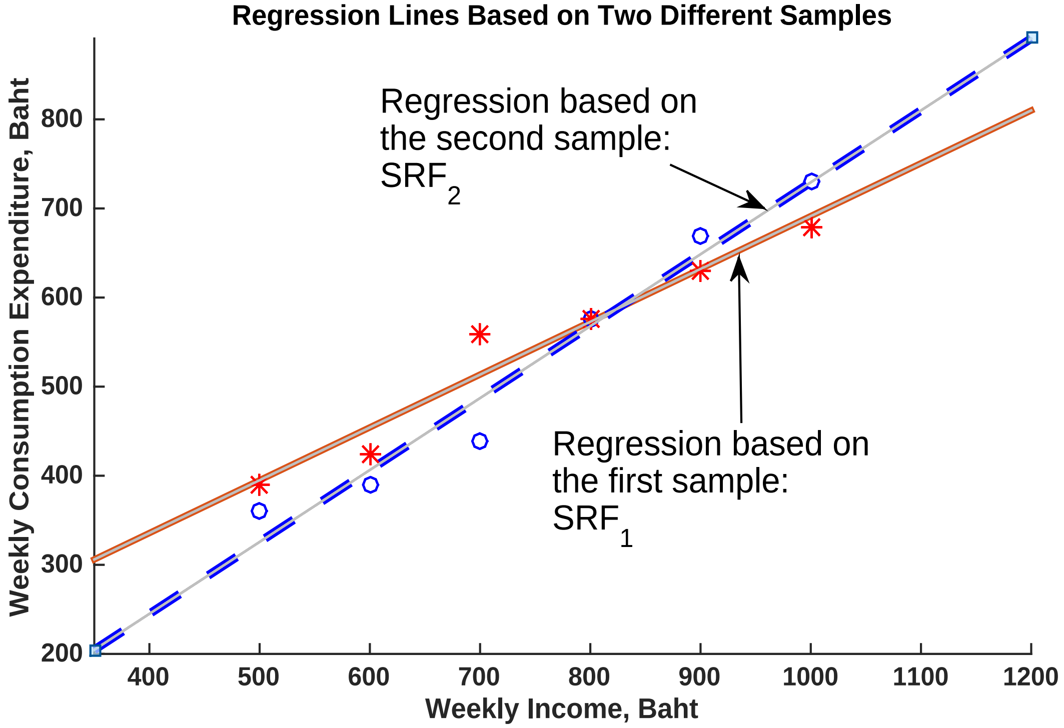
Table 2.3: A Random Sample From the Population

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

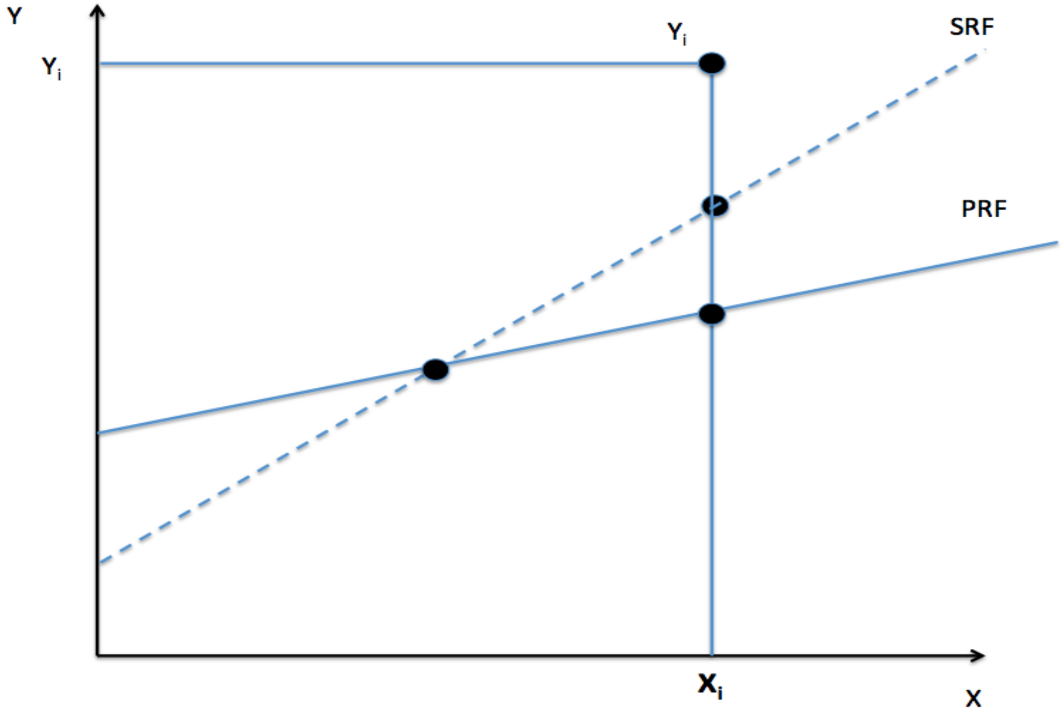
We can express the SRF in its stochastic form as follows:

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

In sum, our ultimate goal is to estimate
the PRF

on the basis of
the SRF

Figure 2.4: Sample and Population Regression Lines



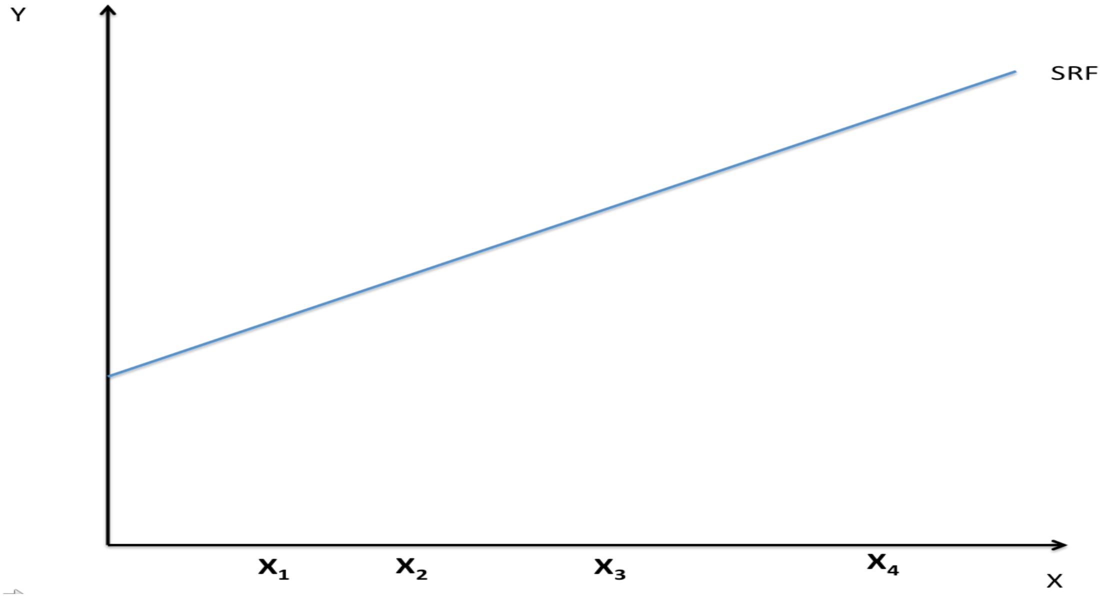
3. REGRESSION:THE PROBLEM OF ESTIMATION

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**:

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

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

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

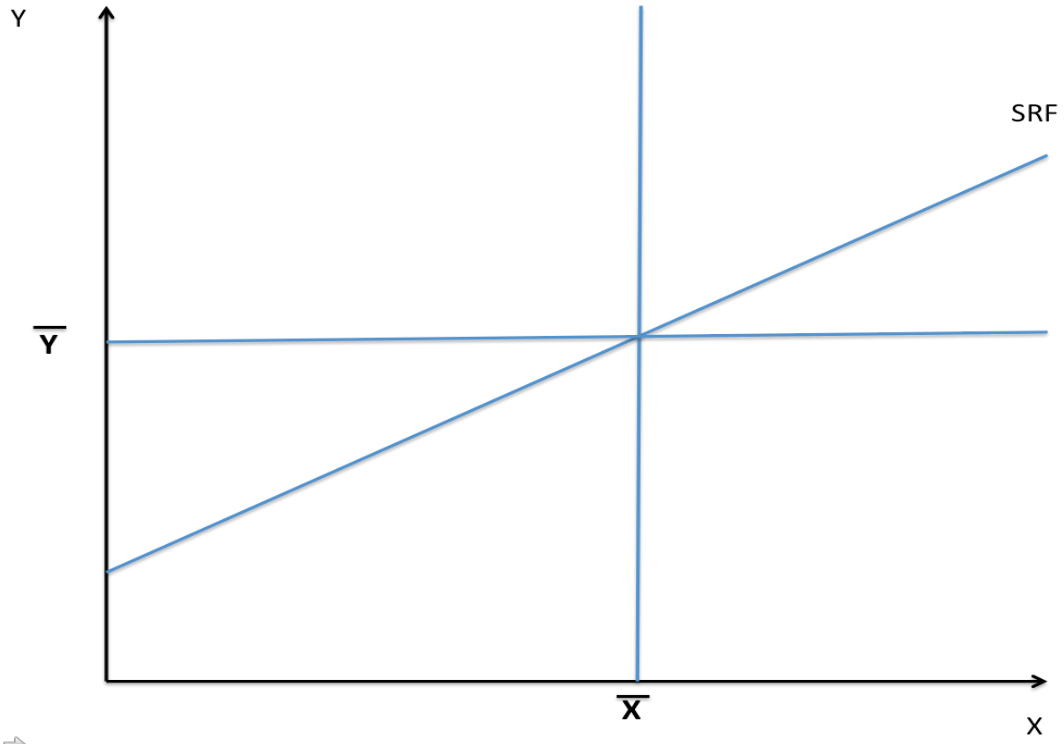
Table 3.1: A Random Sample From the Population

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).

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

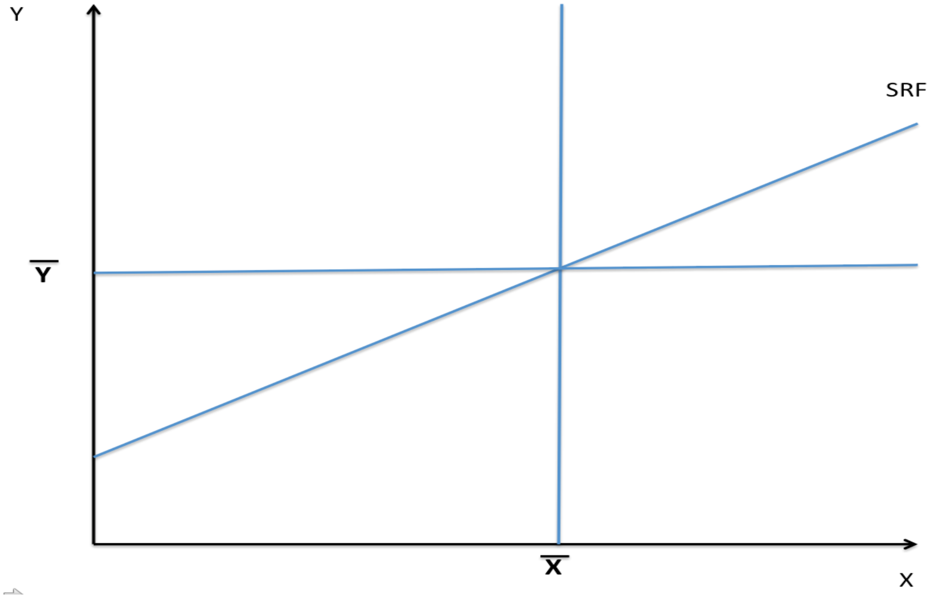
$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum Xi \sum Yi}{n\sum X_i^2 - (\sum X_i)^2} \tag{3.6}$$

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

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

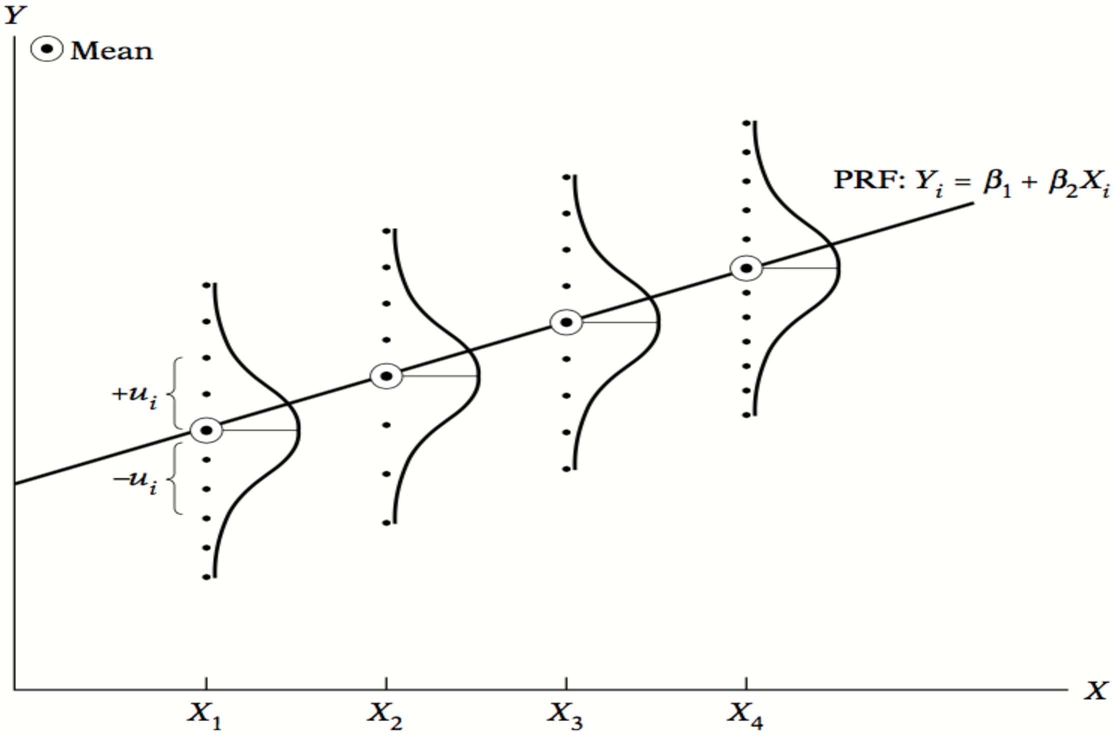
Assumption 2: X values are fixed in repeated sampling

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

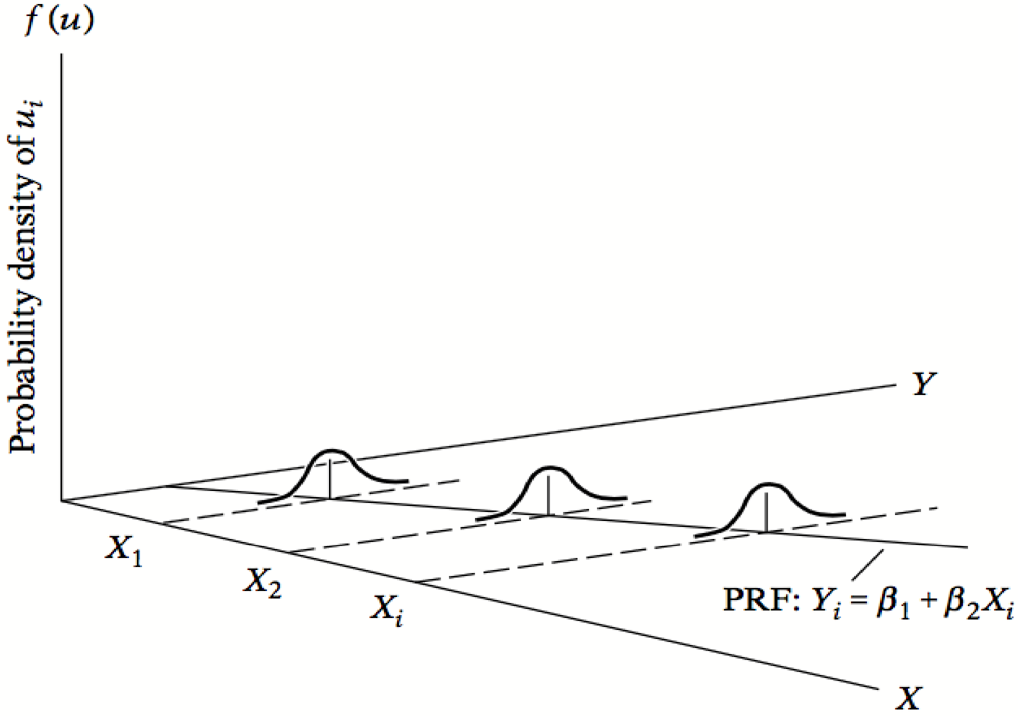
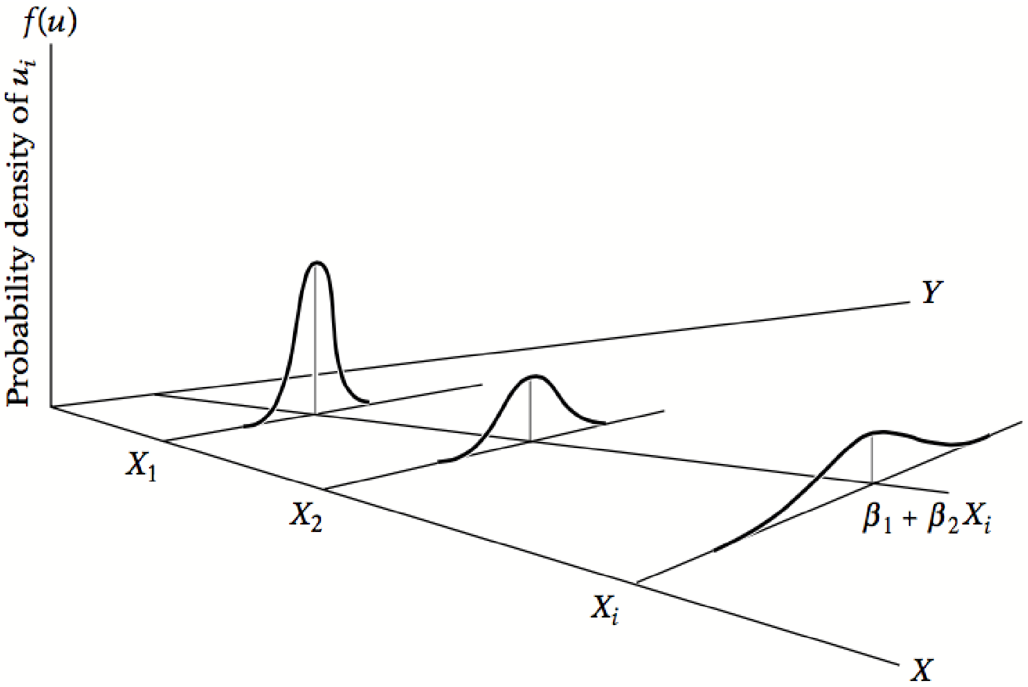


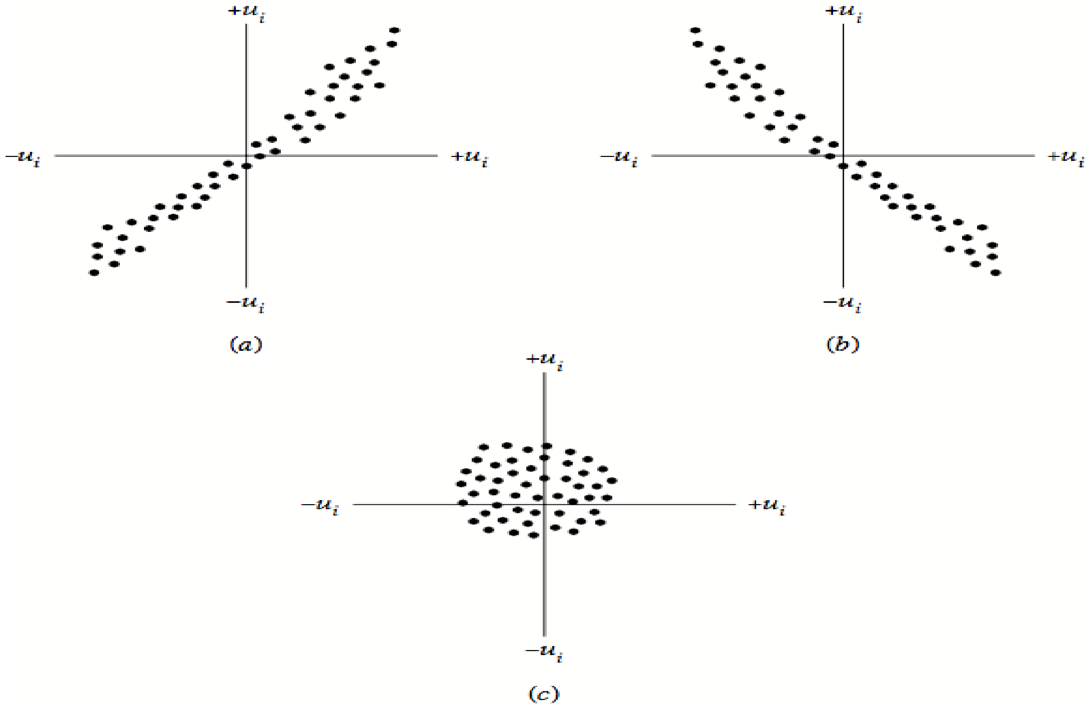
Figure 3.6: Heteroscedasticity



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. $\sum k_i x_i = \sum k_i Xi = 1$

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:

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

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

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:

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

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).

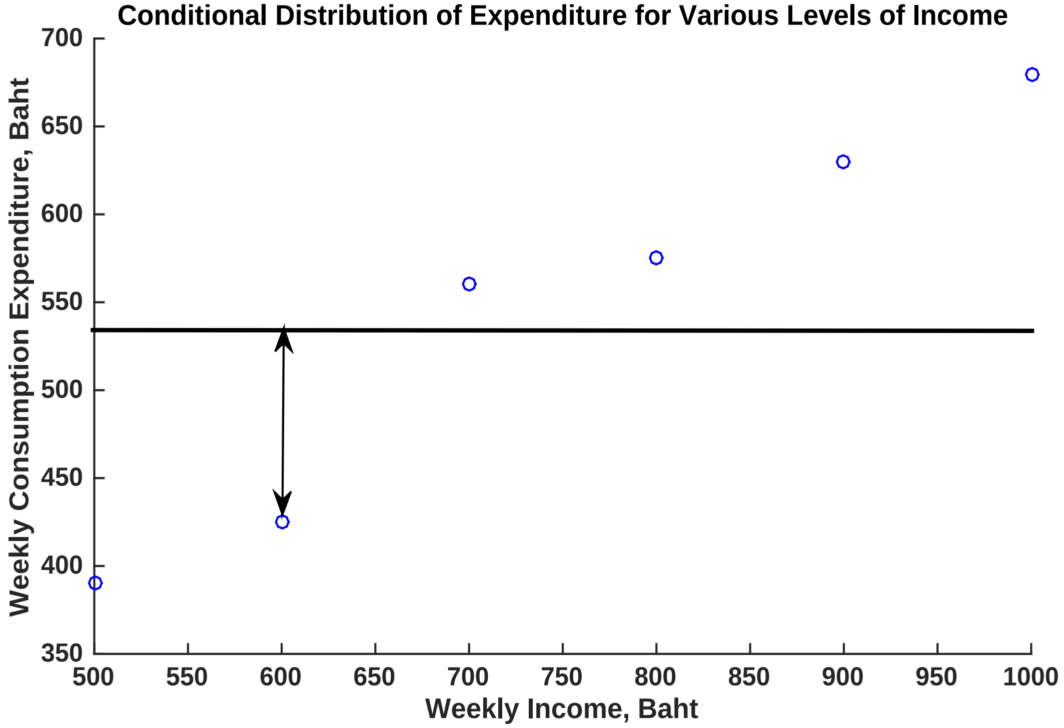
What will happen if we set the estimated Y to be $\bar{Y}$?

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	Income				
	Y_i	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

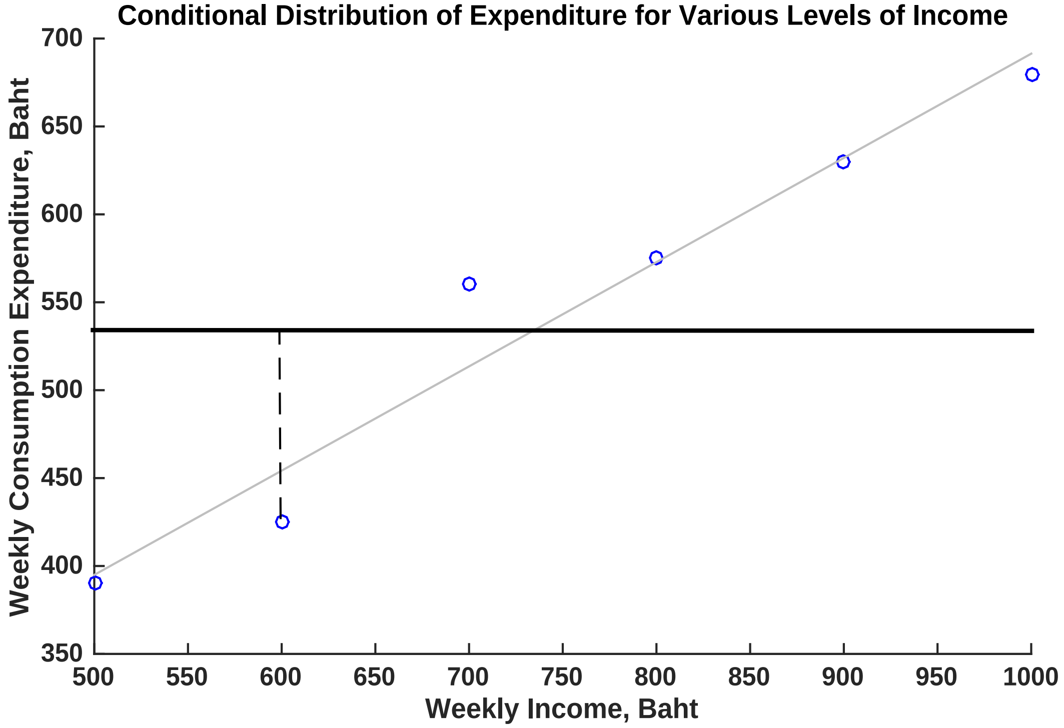


Table 3.5: Estimating the expenditure of the household with income

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:

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

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

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

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

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

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:

5. Interval Estimation and Hypothesis Testing

Interval Estimation

5.1 Confidence Intervals for Regression Coefficients β_1 and β_2

In Sum

A $100(1-\alpha)$ percent **confidence interval** for β_2 can be defined as:

$$\hat{\beta}_2 \pm t_{\alpha/2}\text{se}(\hat{\beta}_2)$$

or

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\text{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\text{se}(\hat{\beta}_2)] = 1-\alpha$$

Analogously, we can define $100(1-\alpha)$ percent **confidence interval** for β_1 as:

$$\hat{\beta}_1 \pm t_{\alpha/2}\text{se}(\hat{\beta}_1)$$

or

$$\Pr[\hat{\beta}_1 - t_{\alpha/2}\text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\text{se}(\hat{\beta}_1)] = 1-\alpha$$

Example

Table 5.1: 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

Table 5.2: 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

Confidence Interval for β_2

Confidence Interval for β_1

5.2 Confidence Interval for σ^2

5.3 Hypothesis Testing: The Confidence-Interval Approach

Based on our sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = 0.6$$

$$H_1 : \beta_2 \neq 0.6$$

5.4 Hypothesis Testing: The Test of Significance Approach

5.4.1 Two-Tail Test

Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = 0.6$$

$$H_1 : \beta_2 \neq 0.6$$

5.4.2 One-Tail Test

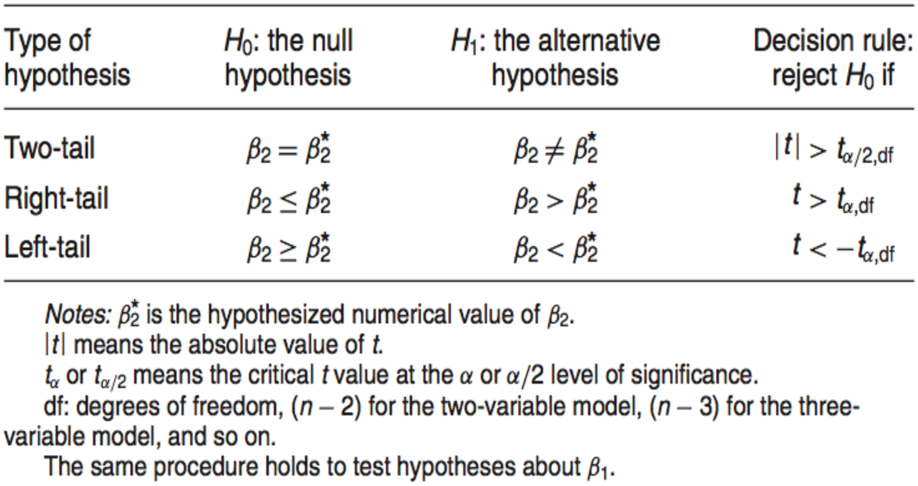
Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 \leq 0.6$$

$$H_1 : \beta_2 > 0.6$$

We can summarize the decision rules for the t test as follow:

Figure 5.1 The t test of Significance: Decision rules



Type of hypothesis	H_0: the null hypothesis	H_1: the alternative hypothesis	Decision rule: reject H_0 if
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2,\text{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\text{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\text{df}}$

Notes: β_2^* is the hypothesized numerical value of β_2.

$\lvert t \rvert$ means the absolute value of t.

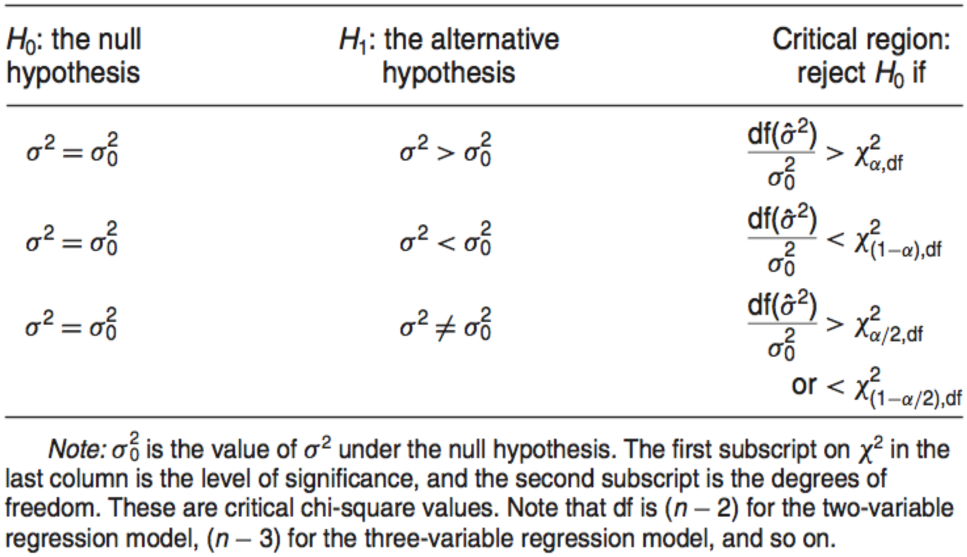
t_α or $t_{\alpha/2}$ means the critical t value at the α or $\alpha/2$ level of significance.

df: degrees of freedom, $(n - 2)$ for the two-variable model, $(n - 3)$ for the three-variable model, and so on.

The same procedure holds to test hypotheses about β_1.

5.4.3 Testing the significance of σ^2: The χ^2 test

Figure 5.2 The χ^2 Test : Decision rules



H_0: the null hypothesis	H_1: the alternative hypothesis	Critical region: reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{df}}$ or $< \chi^2_{(1-\alpha/2),\text{df}}$

Note: σ_0^2 is the value of σ^2 under the null hypothesis. The first subscript on χ^2 in the last column is the level of significance, and the second subscript is the degrees of freedom. These are critical chi-square values. Note that df is $(n-2)$ for the two-variable regression model, $(n-3)$ for the three-variable regression model, and so on.

Why do we say “we cannot reject the null hypothesis?" instead of “We accept the null hypothesis"

The Level of Significance: α

Type I error

Type II error

The Exact Level of Significance: The p Value

5.4.1 Regression Analysis and Analysis of Variance

Table 5.3: ANOVA Table for the two-variable regression model

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

Table 5.4: 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

Table 5.5: 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

Table 5.6: ANOVA Table: Estimating the expenditure of the household with income

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

(After MID-TERM)

5.5 The problem of prediction

Based on our sample data, we have the following sample regression:

We can use the above regression to “Predict” or “Forecast” the future consumption expenditure Y corresponding to some given level of income X

There are two kinds of predictions which are:

[1] **Mean prediction** We will predict the conditional mean value of Y corresponding to a chosen X (i.e X_0)

[2] **Individual Prediction** We will predict an individual Y value corresponding to (i.e X_0)

Mean Prediction

We know that

$$\hat{Y}_0 \sim N(E(\hat{Y}_0), \operatorname{var}(\hat{Y}_0))$$

where

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0$$

and

$$\operatorname{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

If we replace the unknown σ^2 by the unbiased estimator $\hat{\sigma}^2$ we can get

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{se(\hat{Y}_0)}$$

which is the t distribution with n-2 df.

Therefore, we can derive the confidence interval for the true $E(Y_0|X_0)$ as following:

$$Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2}} se(\hat{Y}_0)] = 1 - \alpha$$

Example

Individual Prediction
We can prediction an individual Y value, Y_0 corresponding to a given X value (X_0) :

but the variance in this case is:

$$\operatorname{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

We can show that Y_0 follows the normal distribution:

$$Y_0 \sim N(\hat{Y}_0, \operatorname{var}(Y_0 - \hat{Y}_0))$$

Therefore, we can construct the confidence interval for the Y_0 as well.

From our example:

6. Extensions of The Two-Variable Linear Regression Mode

6.1 Functional Form of regression Models

We will consider the following models:

[1] The log-linear model

[2] Semilog models

[3] Reciprocal models

[4] The logarithmic reciprocal model

The Log-linear Model

The Semilog Models

The Reciprocal Models

The Logarithmic Reciprocal Models

6.2 Regression Through the Origin

In this section, we consider the case that the two-variable PRF assumes the following form:

$$Y_i = \beta_2 X_i + u_i$$

This model is called **the regression through the origin** where the intercept term $\hat{\beta}_1$ is absent from the model.

Example

Since it is the linear regression model, we can apply the Ordinary Least Square (OLS) to estimate the formula for $\hat{\beta}_2$

Let us first write the sample regression function (SRF) as:

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

We would like to minimize

$$\sum \hat{u}_i{}^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2$$

therefore,

$$\boxed{\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}}$$

Now we can find out the variance of $\hat{\beta}_2$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$$

$$\hat{\sigma^2} = \frac{\sum \hat{u}_i{}^2}{n-1}$$

It should be noted that we get the condition $\sum \hat{u}_i X_i = 0$ from the normal equation. However, with the regression through the origin model, we cannot get the condition $\sum \hat{u}_i = 0$.

For the zero-intercept model, r^2 can be negative, whereas for the conventional model it cannot be negative.

Since the conventional r^2 is not appropriate for the regressions that do not contain the intercept, we therefore compute what is known as the **raw** r^2 instead:

$$\text{raw}\, r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

This raw r^2 has its value between 0 and 1, but we cannot directly compare its value to the conventional r^2 value. For this reason, some researchers do not report the r^2 value for zero intercept regression models.

6.2.1 Scaling and Units of Measurements

Consider our old example given in table 18 which refer to weekly family expenditure (Y) and Income (x), in baht.

Table 6.1: Weekly family Expenditure (Y), Baht and Income (X), (Unit:Baht)

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

By using the OLS estimation, we get the following results:

Now, we are interested in changing the units of our data. For example, we would prefer to express our sample data in the unit of 1000 baht. By using the new unit of X and Y, we can report our data in 1000 baht as in the following table.

Table 6.2: Weekly family Expenditure (Y), Baht and Income (X), (Unit: 1000 Baht)

X	Y
0.5	0.360
0.6	0.390
0.7	0.440
0.8	0.575
0.9	0.670
1	0.730

With the new unit, we would like to answer these two questions:
1. Do the units in which the regressand (Y) and regressor/s (X) are measured make any difference in the regression results?
2. If so, what is the sensible course to follow in choosing units of measurement for regression analysis?

To answer these questions, let:

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

where Y is the weekly family expenditure and X is the income, in baht.

Now, let w_1 and w_2 are constants, called the **Scale factors**. For example, in our data, if we need to use the unit of 1000 baht instead, we can directly multiple the original data in table 18 with the scale factors equal to 0.001. In other words, $w_1 = w_2 = \frac{1}{1000} = 0.001$.

Define

$$Y_i^* = w_1 Y_i$$

$$X_i^* = w_1 X_i$$

Now consider the regression using Y_i^* and X_i^* variables:

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

$\hat{u_i^*} =?$

Our target is to find out the relationship between the following pairs:

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$

2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$

3.var$(\hat{\beta}_1)$ and var$(\hat{\beta}_1^*)$

4.var$(\hat{\beta}_2)$ and var$(\hat{\beta}_2^*)$

5.$\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$

6. r_{xy}^2 and $r_{x^*y^*}^2$

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$

2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$

3.var$(\hat{\beta}_1)$ and var$(\hat{\beta}_1^*)$

4.var$(\hat{\beta}_2)$ and var$(\hat{\beta}_2^*)$

5.$\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$

6. r_{xy}^2 and $r_{x^*y^*}^2$

7. Multiple Regression Analysis: The Problem of Analysis

Three-Variable Model: Notation and Assumptions

Let us consider the following three-variable PRF as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

where

Y_i is the dependent variable (regressand)

X_{2i} and X_{3i} are the regressors or the explanatory variables

u_i is the stochastic disturbance term

Remark: the subscript i is denoted the observation i from our sample data.

In case our data are time series, the subscript t will denote the t observation.

β_1 means the average value of Y when X_2 and X_3 are set equal to zero

β_2 and β_3 are called the partial regression coefficients.

We will talk about the meaning of β_1 and β_2 shortly after knowing the assumptions of the classical linear regression model (CLRM)

Under the CLRM, we assume:

1. Zero mean value of u_i

2. No serial correlation

3. Homoscedasticity

4. Zero covariance between u_i and each X variable, or

5. No specification bias or

The model is correctly specified.

6. No exact collinearity between the X variables or

By the above assumptions, we can find out the conditional expectation of Y_i:

The meaning of partial coefficients:

β_2

β_3

7.1 OLS Estimation of the Partial Regression Coefficients

In order to find the OLS estimators, we need to write down the sample regression function (SRF) corresponding to the PRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

From the FOC, we then get the normal equations:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$$

We therefore get:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

Variance and Standard Errors of OLS Estimators

$$var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3{}^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \right] * \sigma^2$$

$$se(\hat{\beta}_1) = +\sqrt{var(\hat{\beta}_1)}$$

$$var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} * \sigma^2$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$se(\hat{\beta}_2) = +\sqrt{var(\hat{\beta}_2)}$$

$$var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} * \sigma^2$$

$$var(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

$$se(\hat{\beta}_3) = +\sqrt{var(\hat{\beta}_2)}$$

$$cov(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1-r_{23}^2)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i{}^2}{n-3}$$

7.2 Properties of OLS Estimators

Properties of OLS Estimators (Cont:)

Properties of OLS Estimators (Cont:)

The Multiple Coefficient of Determination R^2 and the Multiple Coefficient of Correlation R

In this section, we will study how to measure the proportion of the variation in Y explained by the variables X_2 and X_3 jointly. This is the same concept of r^2 that we have learned before.

The quantity that gives this information is known as the **the multiple coefficient of determination** and is denoted by R^2.

To derive R^2, we firstly write down the following equation:

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned} \tag{7.1}$$

where $\hat{Y}_i$ is the estimated value of Y_i from the fitted regression line and is an estimator of true $E(Y_i|X_{2i}, X_{3i})$.

7.1 may be written as

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \\ &= \hat{y}_i + \hat{u}_i \end{aligned} \tag{7.2}$$

Squaring 7.2 on both sides and summing over the sample values, we obtain

$$\boxed{\begin{aligned} \sum y_i^2 &= \sum \hat{y_i}^2 + \hat{u_i}^2 + 2\sum \hat{y}_i \hat{u}_i \\ &= \sum \hat{y_i}^2 + \sum \hat{u_i}^2 \end{aligned} \tag{7.3}}$$

$$
\begin{aligned}
R^2 &= \frac{ESS}{TSS} \\
&= \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}
\end{aligned}
\tag{7.4}
$$

The three-or-more-variable analogue of r is the coefficient of multiple correlation, denoted by R, and it is a measure of the degree of association between Y and all the explanatory variables jointly. Although r can be positive or negative, R is always taken to be positive.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right)$$

7.2.1 R^2 and the Adjusted R^2

It should be noted that he R^2 is a nondecreasing function of the number of explanatory variables. Thus, when the number of regressors increases, R^2 almost invariably increases and never decreases. **In other words, an additional X variable will not decrease R^2!**

To explain this fact, let we write down the definition of R^2 again:

$$\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= 1-\frac{RSS}{TSS} \\ &= 1-\frac{\sum \hat{u}_i{}^2}{\sum y_i^2} \end{aligned} \tag{7.5}$$

Therefore, in comparing two regression models **with the same dependent variable but differing number of X variables**, one should be very wary of choosing the model with the highest R^2.

In light of comparing two R^2 terms, we have to take into account the number of X variables present in the model. To achieve this goal , we can consider the alternative coefficient of determination, which is as follows:

k= the number of parameters in the model including the intercept term.
n= the number of observations in the sample data.

The above equation is known as **the adjusted** R^2, denoted by $\bar{R}^2$. The term adjusted means adjusted for the df associated with the sums of squares entering into 7.5.

We can rewrite the the adjusted R^2 as:

We can also get the equation which shows the relationship between $\bar{R}^2$ and R^2:

Besides R^2 and $\bar{R}^2$ as goodness of fit measures, other criteria are often used to judge the adequacy of a regression model. Two of these are **Akaike's Information criterion and Amemiya's Prediction criteria**, which are used to select between competing models. We will discuss these criteria in greater detail later.

8. Multiple Regression Analysis: The Problem of Inference

In this chapter, we will extend the ideas of interval estimation and hypothesis testing developed there to models involving three or more variables.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

We have already known that if our objective is to do interval estimation and hypothesis testing, we need to assume that the u_i follow the normal distribution with zero mean and constant variance σ^2

With the normality assumption and the CLRM assumptions, we know that:

[1] The OLS estimations of partial regression coefficients are best linear unbiased estimators (BLUE).

[2] The estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are normally distributed with means equal to true β_1,β_2, and β_3 and variances are following:

$$var(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X_3}^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\right] * \sigma^2$$

$$se(\hat{\beta}_1) = +\sqrt{var(\hat{\beta}_1)}$$

$$var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2} * \sigma^2$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$se(\hat{\beta}_2) = +\sqrt{var(\hat{\beta}_2)}$$

$$var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2} * \sigma^2$$

$$var(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

$$se(\hat{\beta}_3) = +\sqrt{var(\hat{\beta}_3)}$$

Moreover, $\frac{(n-3)\hat{\sigma^2}}{\sigma^2}$ follows the χ^2 distribution with n-3 df. We can also show that, if we replace the true σ^2 by its unbiased estimator $\hat{\sigma^2}$ in the computation of the standard errors, we then get

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)}$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)}$$

follows the t distribution with n-3 df.

Example Consider the following regression:

$$\begin{aligned} \log(\widehat{\text{salary}}) &= 4.32 + 0.280 \ \log(\text{sales}) + 0.0174 \ \text{ROE} + 0.00024 \ \text{ROS} \\ se &= (0.32) \ (0.035) \qquad\qquad (0.0041) \qquad\quad (0.00054) \end{aligned} \tag{8.1}$$

$$R^2 = 0.283$$

where
salary =salary of CEO
sales=annual firm sales
ROE=return on equity in percent
ROS =return on firm's stock

Interpreter the partial regression coefficients

Questions What about the statistical significance of the observed results?

For the coefficient of log(sales) of 0.280, Is this coefficient statistically significant different from zero?

For the coefficient of ROE of 0.0174, Is this coefficient statistically significant different from zero?

For the coefficient of ROS of 0.00024, Is this coefficient statistically significant different from zero?

Are these three coefficients statistically significant?
To answer these questions, we have to learn the kinds of hypothesis testing.

8.1 Hypothesis Testing About Individual Regression Coefficients

We can use the t-test to test a hypothesis about any individual partial regression coefficient.

8.1.1 Two-tail test:

Let us postulate that
$H_0: \ \beta_2 = 0$
$H_1: \ \beta_2 \neq 0$

8.1.2 One-tail test:

Let us postulate that
$H_0: \quad \beta_2 \leq 0$
$H_1: \quad \beta_2 > 0$

8.2 Testing The Overall Significance of the Sample Regression

In the previous section, we test the significance of the estimated partial regression coefficients individually, that is under the separate hypothesis that each true population partial regression coefficient was zero. But now we are interested in testing β_2 , β_3 and β_4 are jointly or simultaneously equal to zero. In other words, we would like to test the following hypothesis:

$$H_0 \;\; \beta_2 = \beta_3 = \beta_4 = 0$$

In order to reach this goal, we have to learn the following test.

The Analysis of Variance Approach to Testing the Overall Significance of an Observed Multiple Regression: The F-Test

The joint hypothesis can be tested by the **Analysis of Variance (ANOVA)** which can be demonstrated as follows:

Table 8.1: ANOVA Table for the three-variable regression model

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

Decision Rule Given the k- variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

To test the hypothesis

$$H_0: \quad \beta_2 = \beta_3 = ... = \beta_k = 0$$

(i.e ., all slope coefficients are simultaneously zero) versus

$$H_1 \text{ Not all slope coefficients are simultaneously zero}$$

If $F > F_\alpha(k-1, n-k)$, we reject H_o ; otherwise we cannot reject it, where $F_\alpha(k-1, n-k)$ is the critical F value at the α level of significance and (k-1) numerator df and (n-k) denominator df.

An important Relationship between R^2 and F

Table 8.2: ANOVA Table in Terms of R^2

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

Decision Rule Testing the overall significance of a regression in terms of R^2

Given the k- variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

To test the hypothesis

$$H_0 : \ \beta_2 = \beta_3 = ... = \beta_k = 0$$

(i.e ., all slope coefficients are simultaneously zero) versus

$$H_1 \text{ Not all slope coefficients are simultaneously zero}$$

Compute

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

If $F > F_\alpha(k-1, n-k)$, we reject H_o ; otherwise we cannot reject it, where $F_\alpha(k-1, n-k)$ is the critical F value at the α level of significance and (k-1) numerator df and (n-k) denominator df.

8.3 The "Incremental" or "Marginal" Contribution of an Explanatory Variable

Let consider the following regression:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_i$$

Having run the above regression, let us suppose we decide to add the additional variable, X_{3i}, to the model and obtain the multiple regression as follow:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Comparing between these two regressions, we might need to answer the below questions:

[1]. What are the marginal, or incremental, contribution of X_{3i}, knowing that X_{2i} is already in the model and that it is significantly related to Y_i.

[2]. Is the incremental contribution of X_{3i} statistically significant?

[3]. What is the criterion for adding variables to the model?

By contribution we mean whether the additional of the variable, X_{3i}, to the model increases ESS (and hence R^2) "significantly" in relation to the RSS. This contribution is called **the incremental, or marginal** contribution of an additional variable.

To assess the incremental contribution of X_3 after allowing for the contribution of X_2, we form

$$\begin{aligned} F &= \frac{Q_2/\text{df}}{Q_4/\text{df}} \\ &= \frac{(ESS_{new} - ESS_{old})/\text{number of new regressors}}{RSS_{new}/df(\text{=n-number of parameters in the new model})} \end{aligned} \tag{8.2}$$

Under the normality assumption of u_i and CLRM assumptions, this F value follows the F distribution with 1 and n-number of parameters in the new model.

Table 8.3: ANOVA Table To Assess Incremental Contribution of A Variable(s)

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
ESS due to X_2 alone	$Q_1 = \hat{\alpha}_2{}^2 \sum x_2^2$	1	$\frac{Q_1}{1}$
ESS due to the addition of X_3	$Q_2 = Q_3 - Q_1$	1	$\frac{Q_2}{1}$
ESS due to both X_2,X_3	$Q_3 = \hat{\beta}_2 \sum x_{2i} y_i + \hat{\beta}_3 \sum x_{3i} y_i$	2	$\frac{Q_3}{2}$
RSS	$Q_4 = Q_5 - Q_3$	n-3	$\frac{Q_4}{n-3}$
TSS	$Q_5 = \sum y_i^2$	n-1	

As usual method, we can re write 8.2 in term of R^2 only. Thus the F ratio of 8.2 is equivalent to the following F ratio:

$$
\begin{aligned}
F &= \frac{R^2_{new} - R^2_{old}/\text{df}}{(1 - R^2_{new})/\text{df}} \\
&= \frac{(R^2_{new} - R^2_{old})/\text{number of new regressors}}{1 - R^2_{new}/df(\text{=n-number of parameters in the new model})}
\end{aligned}
\tag{8.3}
$$

This F ratio follows the F distribution with 1 and n-number of parameters in the new model.

Example

Consider the child mortality example. We considered the behavior of child mortality (CM) in relation to per capita GNP (PGNP). There we found that PGNP has a negative impact on CM, as one would expect. Now let us bring in female literacy as measured by the female literacy rate (FLR). A priori, we expect that FLR too will have a negative impact on CM. Our sample consists of 64 countries.

In model 1, we regressed child mortality (CM) on per capita GNP (PGNP) and female literacy rate (FLR).

Model 1:

$$\begin{aligned}\widehat{CM_i} &= 263.6416 - 0.0056PGNP_i - 2.2316FLR_i \\ se &= (11.5932) \quad (0.0019) \qquad\qquad (0.2099) \quad R^2 = 0.7077\end{aligned} \tag{8.4}$$

Now we extend this model to model 2 by including total fertility rate (TFR):

Model 2:

$$\begin{aligned}\widehat{CM_i} &= 168.3067 - 0.00555GNP_i - 1.7680FLR_i + 12.8686TFR_i \\ se &= (32.8916) \quad (0.0018) \qquad\qquad (0.2480) \qquad\qquad (?) \quad R^2 = 0.7474\end{aligned} \tag{8.5}$$

Questions

1.How would you choose between models 1 and 2? Which statistical test would you use to answer this question? Show the necessary calculations.

2. We have not given the standard error of the coefficient of TFR. Can you find it out? (Hint: Recall the relationship between the t and F distributions.

8.4 Testing the Equality of Two Regression Coefficients

Suppose we have the following model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \ldots + \beta_k X_{ki} + u_i$$

We would like to test the hypotheses:

$$H_0: \quad \beta_3 = \beta_4 \text{ or } (\beta_3 - \beta_4) = 0$$
$$H_1: \quad \beta_3 \neq \beta_4 \text{ or } (\beta_3 - \beta_4) \neq 0$$

Under the classical assumptions, it can be shown that:

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{se(\hat{\beta}_3 - \hat{\beta}_4)}$$

where the t follows the t distribution with (n-k) df because the above equation is a k-variable model, where k is the total number of parameters estimated, including the constant term.

The $se(\hat{\beta}_3 - \hat{\beta}_4)$ is calculated from the following formula:

$$se(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{var(\hat{\beta}_3) + var\hat{\beta}_4 - 2cov(\hat{\beta}_3, \hat{\beta}_4)}$$

Example

among other things, you were asked to consider the following demand function for chicken:

$$\begin{aligned}\widehat{\ln Y_t} &= 2.0328 + 0.4515 \ln X_{2t} - 0.3772 \ln X_{3t} \\ se &= (0.1162) \quad (0.0247) \qquad\qquad (0.0635) \quad R^2 = 0.9801\end{aligned} \tag{8.6}$$

where Y = per capita consumption of chicken, lb, X_2 = real disposable per capita income, \$, X_3 = real retail price of chicken per lb.

Question

For the above demand function, how would you test the hypothesis that the income elasticity is equal in value but opposite in sign to the price elasticity of demand? Show the necessary calculations. [Note: $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.00142$. and the sample data = 23 observations]

8.5 Restricted Least Squares: Testing Linear Equality Restriction

In economic theories, the coefficients in a regression model need to satisfy some linear equality restrictions, For example, in microeconomics, consider the Cobb-Douglas production function:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i}$$

where Y=output, X_2 = labor input, and X_3=capital input. We can transform the above equation to be the log form as:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

where $\beta_0 = \ln \beta_1$

Now, if there are the constant returns to scale, economic theory would suggest that

$$\beta_2 + \beta_3 = 1$$

which is an example of a linear equality restriction.
In order to test the above linear equality restriction, we can follow two approaches which are:

[1]. The t-test approach

[2]. The F-test approach: Restricted Least Squares.

First Approach: The t-Test
A test of the hypothesis or restriction can be conducted by the t-test:

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{se(\hat{\beta}_2 + \hat{\beta}_3)}$$

where the t follows the t distribution with (n-k) df for a k-variable model, where k is the total number of parameters estimated, including the constant term. In this case, df=n-3.

The $se(\hat{\beta}_2 + \hat{\beta}_3)$ is calculated from the following formula:

$$se(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{var(\hat{\beta}_2) + var\hat{\beta}_3 + 2cov(\hat{\beta}_2, \hat{\beta}_3)}$$

Example

Consider the Cobb-Douglas production function to the Mexican economy (1955-1974: n=20):

$$\begin{aligned} \widehat{\ln GDP_t} &= -1.6524 + 0.3397 \ln Labor_t + 0.8460 \ln Capital_t \\ t &= (-2.7259) \quad (1.8295) \qquad\qquad (9.0625) \quad R^2 = 0.9951 \;\; RSS_{UR} = 0.0136 \end{aligned} \tag{8.7}$$

where GDP = Real GDP, Millions of 1960 pesos, *Labor* = Employment, Thousands of People, *Capital* = Fixed Capital, Millions of 1960 pesos.

Question

As you can see, the output/labor elasticity is about 0.34 and the output/capital elasticity is about 0.85. If we add these coefficients, we obtain 1.19, suggesting that perhaps the Mexican economy during the stated time period was experiencing increasing returns to scale. However, we do not know if 1.19 is statistically different from 1.

Therefore, we have to test this linear equality restriction.

8.6 The F-Test Approach: Restricted Least Squares

From the Cobb-Douglas production function:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \tag{8.8}$$

if there are the constant returns to scale, economic theory would suggest that

$$\beta_2 + \beta_3 = 1$$

We can rewrite it as:

$$\beta_2 = 1 - \beta_3$$

or

$$\beta_3 = 1 - \beta_2$$

Using either of these equalities, we can eliminate one of the β coefficients. Therefore, we can rewrite the Cobb-Douglas production function as:

$$\ln(Y_i/X_{2i}) = \beta_0 + \beta_3 \ln(X_{3i}/X_{2i}) + u_i \tag{8.9}$$

where $\frac{Y_i}{X_{2i}}$=output/labor ratio
$\frac{X_{3i}}{X_{2i}}$= capital labor ratio.

It should be noted that:
8.8 is known as **unrestricted Least Squares (URLS)**
8.9 is known as **restricted Least Squares (RLS)**

We can compare the unrestricted and restricted least-squares regressions by applying the F-test as follows:

$\sum \hat{U}_{UR}^2$=RSS of the unrestricted regression 8.8

$\sum \hat{U}_R^2$ = RSS of the restricted regression 8.9

m=number of linear restrictions (in this example, we have 1 restriction)

k= number of parameters in the unrestricted regression

n= number of observations

Then, we have

$$\begin{aligned} F &= \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} \\ &= \frac{(\sum \hat{U}_R^2 - \sum \hat{U}_{UR}^2)/m}{\sum \hat{U}_{UR}^2/(n-k)} \end{aligned} \tag{8.10}$$

follows the F-distribution with m, (n-k) df.

We can also rewrite the F-test in terms of R^2 as follows:

$$F = \frac{R_{UR}^2 - R_R^2/\text{m}}{(1-R_{UR}^2)/\text{n-k}} \tag{8.11}$$

Example
Consider the Cobb-Douglas production function to the Mexican economy(1955-1974: n=20):

$$\begin{aligned} \widehat{\ln GDP_t} &= -1.6524 + 0.3397 \ln Labor_t + 0.8460 \ln Capital_t \\ t &= (-2.7259) \quad (1.8295) \qquad\qquad (9.0625) \quad R^2 = 0.9951 \;\; RSS_{UR} = 0.0136 \end{aligned} \tag{8.12}$$

where GDP = Real GDP, Millions of 1960 pesos, *Labor* = Employment, Thousands of People, *Capital* = Fixed Capital, Millions of 1960 pesos.

The restriction of constant return to scale, which gives the following regression:

$$\begin{aligned} \ln\left(\widehat{GDP/Labor}\right)_t &= -0.4947 + 1.0153 \ln\left(Capital/Labor\right)_t \\ t &= (-4.0612) \quad (28.1056) \quad R_R^2 = 0.9777 \;\; RSS_R = 0.0166 \end{aligned} \tag{8.13}$$

8.7 Testing for Structural or Parameter Stability of Regression Models: The Chow Test

Sometime when we estimate the regression model, it may happen that there is a **Structural Change** in the relationship between the regressand Y and the regressors X's, especially the model involving time series data. The structural change may be due to the external forces (i.e the financial crisis of 2007-2008) or due to policy changes (such as the switch from a fixed exchange rate system to a flexible exchange rate system in 1997).

The question is **"How do we figure out that there is a structural change in our sample data?"**

To answer this question, consider the following example.

Based on the sample data, we found out that in 1982 the United State suffers its worst peacetime regression. This event might disturb the relationship between savings and DPI.

To see this effect, we can divide our sample data into two time periods: 1970-1981 (Pre-1982 crisis) and 1982-1995 (Post-1982 crisis).

Therefore we have three possible regressions:

Time period 1970-1981: $Y_t = \beta_1 + \beta_2 X_t + u_1 t$ where $n_1 = 12$

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_2 t$ where $n_2 = 14$

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_t$ where $n = n_1 + n_2 = 26$

For our sample data, we can get the following results:

Time period 1970-1981:

$$\begin{aligned} \hat{Y}_t &= 1.0161 + 0.0803X_t \\ t &= (0.00873) \quad (9.6015) \end{aligned} \tag{8.14}$$

$R^2 = 0.9021 \quad RSS_1 = 1785.032 \quad df = 10$

Time period 1982-1995:

$$\begin{aligned} \hat{Y}_t &= 153.4947 + 0.0148X_t \\ t &= (4.6922) \quad (1.7707) \end{aligned} \tag{8.15}$$

$R^2 = 0.2971 \quad RSS_2 = 10,005.22 \quad df = 12$

Time period 1970-1995:

$$\begin{aligned} \hat{Y}_t &= 62.4226 + 0.0376X_t \\ t &= (4.8917) \quad (8.8937) \end{aligned} \tag{8.16}$$

$R^2 = 0.7672 \quad RSS_3 = 23,248.30 \quad df = 24$

We can apply **the Chow test** to investigate the structural changes that may be caused by differences in the intercept or the slope coefficient or both.

The chow test assumes that:

[1] $u_{1t} \sim N(0, \sigma^2)$ and $u_{2t} \sim N(0, \sigma^2)$

[2] The two error terms u_{1t} and u_{2t} are independently distributed.

Chow Test

$$H_0: \text{ There is no structural change in the model}$$

$$H_1: \text{ There is structural change in the model}$$

Then, we need to construct the F-ratio:

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n_1 + n_2 - 2k)} \tag{8.17}$$

where the F ratio follows the F distribution with k and $(n_1 + n_2 - 2k)$ df in the numerator and denominator, respectively.

We do not reject the null hypothesis of parameter stability (i.e no structural change) if the computed F value does not exceed the critical value F value obtained from the F table.

9. Dummy Variable Regression Models

In the previous chapter, the dependent and independent variables in our multiple regression models have had **quantitative** meaning. For example, the salary of CEO,annual firm sales,return on equity in percent, and return on firm's stock. In each case the magnitude of the variable conveys useful information.

However, in the empirical work, we must also incorporate **qualitative factors** into regression models. The gender or race of an individual, the industry of a firm (manufacturing, retail, and so on), and the region in Thailand where a city is located (north, south, west, and so on) are all considered as the qualitative factors.

9.1 Describing Qualitative Information

Normally, qualitative factors often come in the form of binary information:

Example:

[1] A person is female or male or female.

[2] A firm offers a certain kind of employee pension plan or it does not.

[3] A farm is located nearby the dam or not.

All of these examples, the relevant information can be captured by defining a **binary variable** or a zero-one variable.

In econometrics, binary variables are most commonly called **dummy variables**, although this name is not especially descriptive.

In defining a dummy variable, we must decide which event is assigned the value one and which is assigned the value zero.

Question: Why do we use the the values zero and one to describe qualitative information?

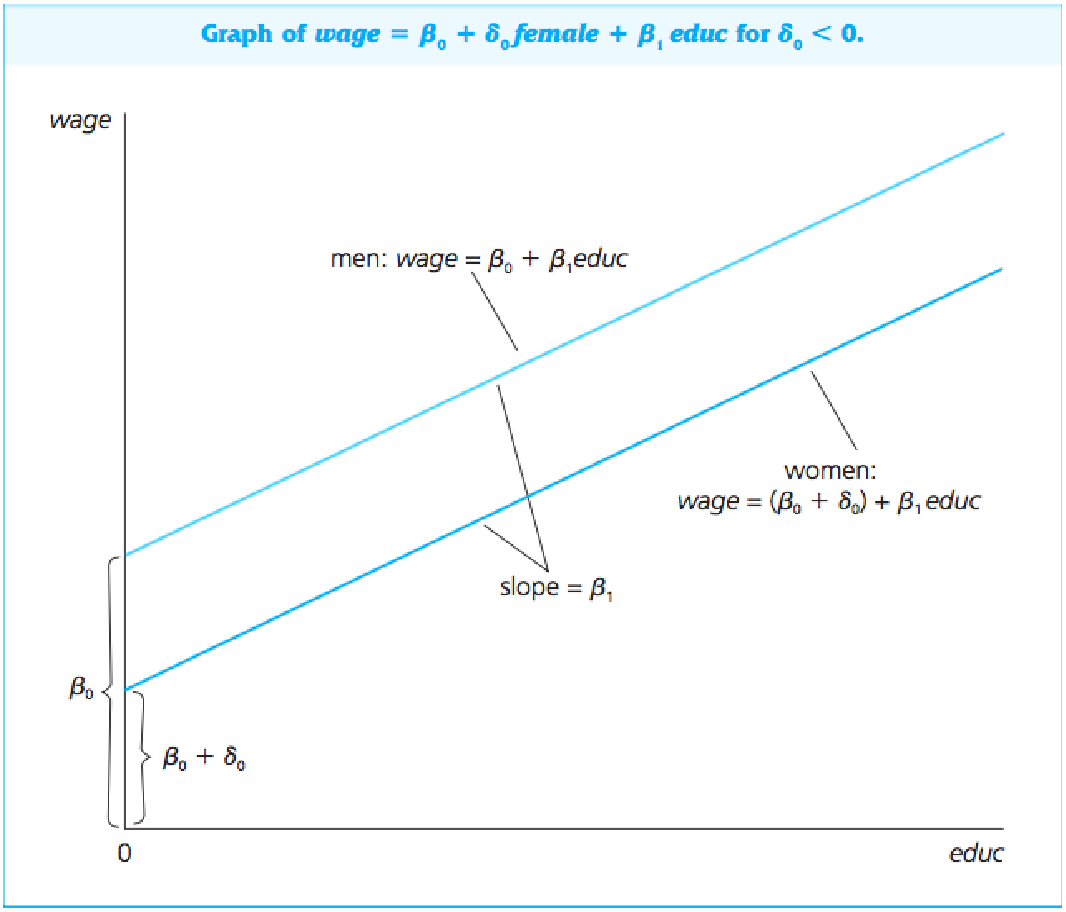
Answer: These values are arbitrary: any two different values would do. The real benefit of capturing qualitative information using zero-one variable is that it leads to regression models where the parameters have very natural interpretations.

9.1.1 A Single Dummy Independent Variable

Suppose we would like to estimate the following simple model of hourly wage determination:

$$wage_i = \beta_0 + \delta_0 \text{female} + \beta_1 \text{edu} + u_i$$

Figure 9.1: Graph of Wage



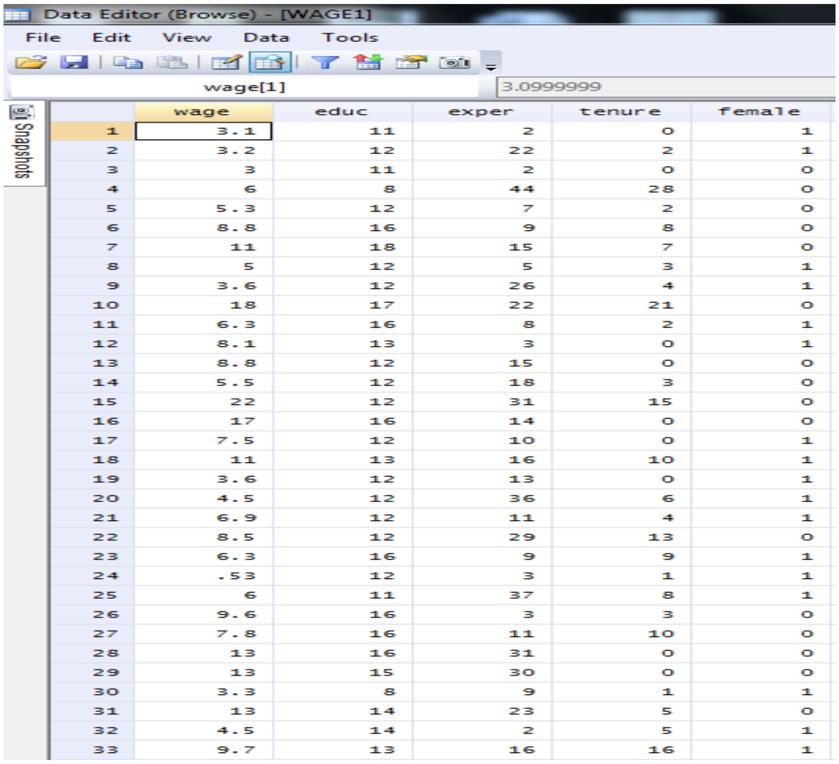
Now, we added more variables into the wage model. Taking males as the base group, a model that controls for experience and tenure in addition to education is

$$wage_i = \beta_0 + \delta_0\text{female} + \beta_1\text{edu} + \beta_2\text{exper} + \beta_3\text{tenure} + u_i$$

If edu, expert, and tenure are all relevant productivity characteristics, the null hypothesis of no difference between men and women (No wage discrimination) is:

In table 9.1, it represents the partial listing of the sample data of wage model. We see that Person 1 is female, Person 2 is female , Person 3 is male, and so on.

Table 9.1: A Partial Listing of the Wage Data.



wage[1] 3.0999999

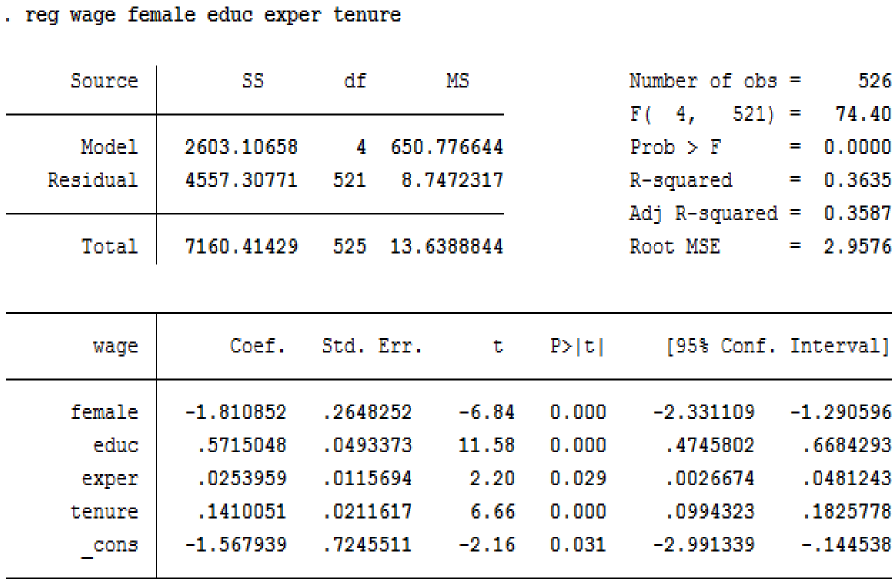
	wage	educ	exper	tenure	female
1	3.1	11	2	0	1
2	3.2	12	22	2	1
3	3	11	2	0	0
4	6	8	44	28	0
5	5.3	12	7	2	0
6	8.8	16	9	8	0
7	11	18	15	7	0
8	5	12	5	3	1
9	3.6	12	26	4	1
10	18	17	22	21	0
11	6.3	16	8	2	1
12	8.1	13	3	0	1
13	8.8	12	15	0	0
14	5.5	12	18	3	0
15	22	12	31	15	0
16	17	16	14	0	0
17	7.5	12	10	0	1
18	11	13	16	10	1
19	3.6	12	13	0	1
20	4.5	12	36	6	1
21	6.9	12	11	4	1
22	8.5	12	29	13	0
23	6.3	16	9	9	1
24	.53	12	3	1	1
25	6	11	37	8	1
26	9.6	16	3	3	0
27	7.8	16	11	10	0
28	13	16	31	0	0
29	13	15	30	0	0
30	3.3	8	9	1	1
31	13	14	23	5	0
32	4.5	14	2	5	1
33	9.7	13	16	16	1

Table 9.2: The command function to estimate the wage model in STATA program

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   12.0   Copyright 1985-2011 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     MP - Parallel Edition            College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

10-student 2-core Stata lab perpetual license:
       Serial number:  50120540791
         Licensed to:  Faculty of Economics
                       Thammasat University

Notes:
      1.  (/v# option or -set maxvar-) 5000 maximum variables
      2.  New update available; type -update all-

. use C:\Users\wasin\AppData\Local\Temp\Temp1_data.zip\WAGE1.DTA

.
```

```
reg  wage female educ
```

Table 9.3: $wage_i = \beta_0 + \delta_0 \textbf{female} + \beta_1 \textbf{edu} + \beta_2 \textbf{exper} + \beta_3 \textbf{tenure} + u_i$



```
. reg wage female educ exper tenure

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  4,   521) =   74.40
       Model |  2603.10658     4  650.776644           Prob > F      =  0.0000
    Residual |  4557.30771   521   8.7472317           R-squared     =  0.3635
-------------+------------------------------           Adj R-squared =  0.3587
       Total |  7160.41429   525  13.6388844           Root MSE      =  2.9576

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -1.810852   .2648252    -6.84   0.000    -2.331109   -1.290596
        educ |   .5715048   .0493373    11.58   0.000     .4745802    .6684293
       exper |   .0253959   .0115694     2.20   0.029     .0026674    .0481243
      tenure |   .1410051   .0211617     6.66   0.000     .0994323    .1825778
       _cons |  -1.567939   .7245511    -2.16   0.031    -2.991339    -.144538
------------------------------------------------------------------------------
```

Example: the Hourly Wage Equation:

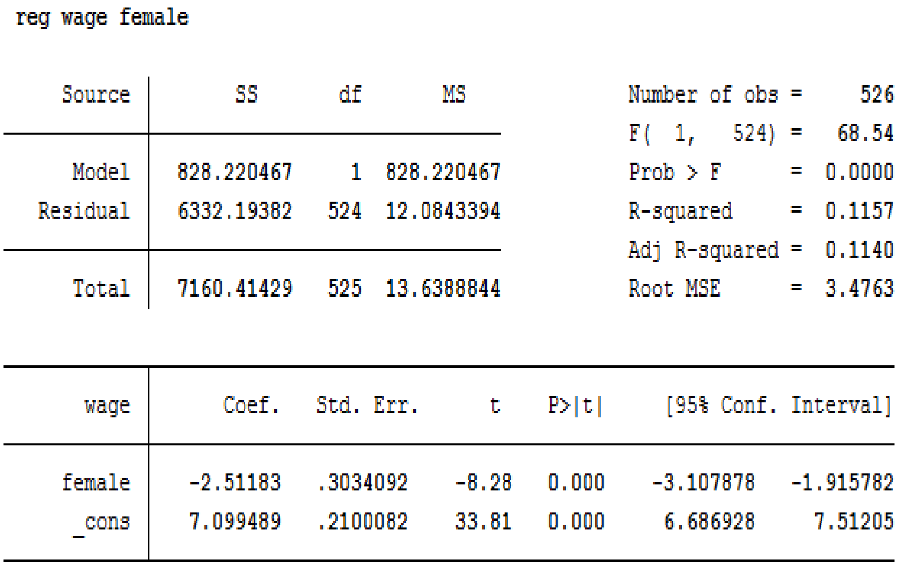
$$
\begin{aligned}
\widehat{\text{wage}} &= -1.5679 - 1.8109 \;\; \text{female} + 0.5715 \;\; \text{edu} + 0.025 \;\; \text{exper} + 0.141 \;\; \text{tenure} \\
&= (0.7246) \qquad (0.2648) \qquad (0.0493) \qquad (0.0116) \qquad (0.0212)
\end{aligned}
\tag{9.1}
$$

$R^2 = 0.3635 \quad n = 526$

Interpret the model:

The intercept:

Table 9.4: $wage_i = \beta_0 + \delta_0 \textbf{female} + u_i$



```
reg wage female

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =   68.54
       Model |  828.220467     1  828.220467           Prob > F      =  0.0000
    Residual |  6332.19382   524  12.0843394           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1140
       Total |  7160.41429   525  13.6388844           Root MSE      =  3.4763

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -2.51183   .3034092    -8.28   0.000    -3.107878   -1.915782
       _cons |   7.099489   .2100082    33.81   0.000     6.686928     7.51205
------------------------------------------------------------------------------
```

The coefficient on female

It is informative to compare the coefficient on female in the above equation to the estimate we get when all other explanatory variables are dropped from the equation:

$$
\begin{aligned}
\widehat{\text{wage}} &= 7.0995 - 2.5118 \ \text{female} \\
se &= (0.2100) \qquad (0.3034)
\end{aligned}
\tag{9.2}
$$

$R^2 = 0.1157 \quad n = 526$

9.2 Interpreting Coefficients on Dummy Explanatory Variables When the Dependent Variable is log(y)

In this section, we will study a model that has the dependent variable appearing in logarithmic form, with one or more dummy variables approving as independent variables.
Question: How do we interpret the dummy variable coefficients in this case?

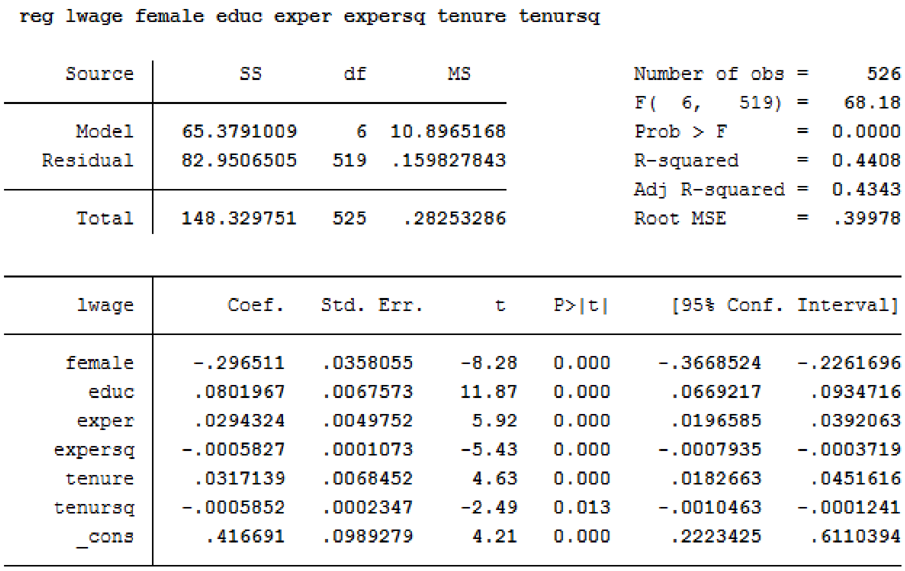
Answer: Not surprisingly, the coefficients have a percentage interpretation.

Let us reestimate the wage equation , using log (wage) as the dependent variable and adding quadratics in expert and tenure:

$$log(wage_i) = \beta_0 + \delta_0 female + \beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

The Stata result is shown in table 9.5.

Table 9.5:



```
 reg lwage female educ exper expersq tenure tenursq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  6,   519) =   68.18
       Model |  65.3791009     6  10.8965168           Prob > F      =  0.0000
    Residual |  82.9506505   519  .159827843           R-squared     =  0.4408
-------------+------------------------------           Adj R-squared =  0.4343
       Total |  148.329751   525   .28253286           Root MSE      =  .39978

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.296511   .0358055    -8.28   0.000    -.3668524   -.2261696
        educ |   .0801967   .0067573    11.87   0.000     .0669217    .0934716
       exper |   .0294324   .0049752     5.92   0.000     .0196585    .0392063
     expersq |  -.0005827   .0001073    -5.43   0.000    -.0007935   -.0003719
      tenure |   .0317139   .0068452     4.63   0.000     .0182663    .0451616
     tenursq |  -.0005852   .0002347    -2.49   0.013    -.0010463   -.0001241
       _cons |    .416691   .0989279     4.21   0.000     .2223425    .6110394
------------------------------------------------------------------------------
```

Example: Log Hourly Wage Equation:

$$\begin{aligned}\widehat{\log(\text{wage})} &= \underset{(0.0989)}{0.4167} - \underset{(0.0358)}{0.2965}\ \text{female} + \underset{(0.0068)}{0.0802}\ \text{edu} + \underset{(0.0050)}{0.0294}\ exper - \underset{(0.0001)}{0.0006}\ exper^2 \\ &\quad + \underset{(0.0068)}{0.0317}\ tenure - \underset{(0.0002)}{0.0006}\ tenure^2 \end{aligned} \tag{9.3}$$

$R^2 = 0.4408 \quad n = 526$

Interpret the model:

The coefficient on female

9.3 Using Dummy Variables for Multiple Categories

We can use several dummy independent variables in the same equation. For example, we could add the dummy variable **married** to the wage model.

The previous model:

$$log(wage_i) = \beta_0 + \delta_0 female + \beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

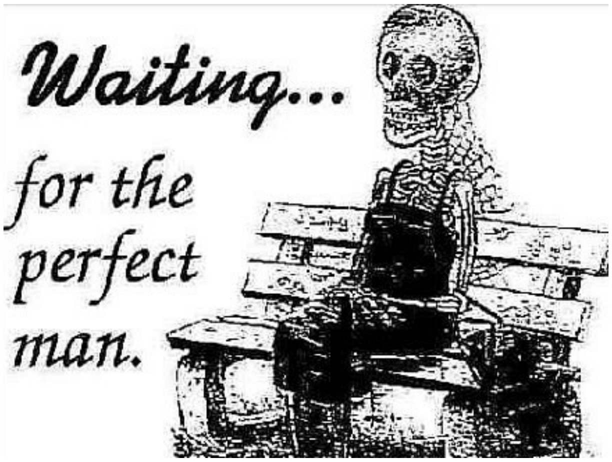
Now, Let us estimate a model that allows for wage differences among four groups:

[1.] Married Men

[2] Married Women

[3] Single Men

[4] Single Women



To do this, we must select a base group:

Now, we need to define dummy variables for each of the remaining groups.

Therefore, our model is:

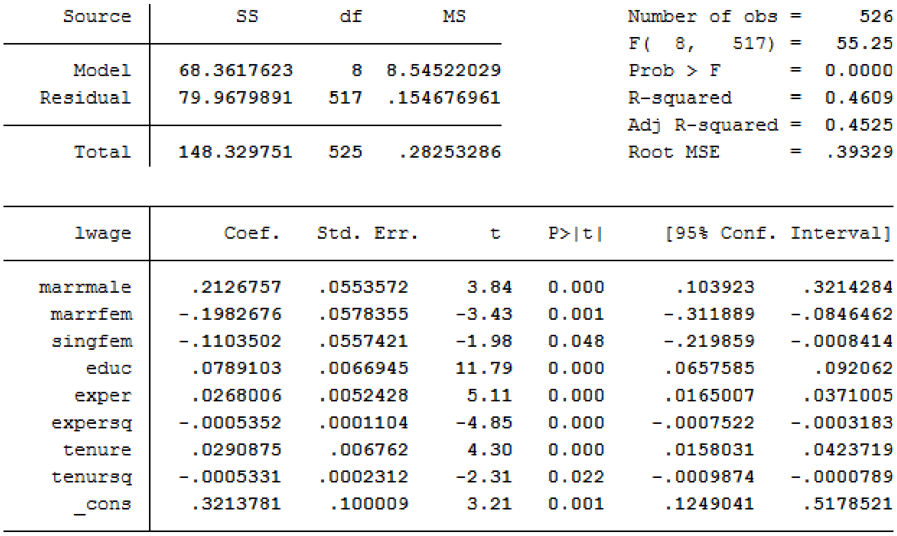
$$log(wage_i) = \beta_0 + \delta_0 marrmale + \delta_1 marrfem + +\delta_2 singfem +$$

$$\beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

We of course drop the dummy variable (female). (Why?)

Table 9.5:

`reg lwage marrmale marrfem singfem educ exper expersq tenure tenursq`



```
      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  8,   517) =   55.25
       Model |  68.3617623     8  8.54522029           Prob > F      =  0.0000
    Residual |  79.9679891   517  .154676961           R-squared     =  0.4609
-------------+------------------------------           Adj R-squared =  0.4525
       Total |  148.329751   525  .28253286            Root MSE      =  .39329

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    marrmale |   .2126757   .0553572     3.84   0.000      .103923    .3214284
     marrfem |  -.1982676   .0578355    -3.43   0.001     -.311889   -.0846462
     singfem |  -.1103502   .0557421    -1.98   0.048     -.219859   -.0008414
        educ |   .0789103   .0066945    11.79   0.000     .0657585     .092062
       exper |   .0268006   .0052428     5.11   0.000     .0165007    .0371005
     expersq |  -.0005352   .0001104    -4.85   0.000    -.0007522   -.0003183
      tenure |   .0290875    .006762     4.30   0.000     .0158031    .0423719
     tenursq |  -.0005331   .0002312    -2.31   0.022    -.0009874   -.0000789
       _cons |   .3213781    .100009     3.21   0.001     .1249041    .5178521
------------------------------------------------------------------------------
```

$$
\begin{array}{rcl}
\widehat{\log(\text{wage})} & = & 0.3214 + 0.2127\ \text{marrmale} - 0.1983\ \text{marrfem} - 0.1104\ singfem \\
 & & (0.1000) \qquad (0.0554) \qquad\qquad (0.0578) \qquad\qquad (0.0557) \\
 & & +0.0789\ edu + 0.0268\ exper - 0.0005\ exper^2 + 0.0291\ tenure - 0.0005\ tenure^2 \\
 & & (0.0067) \qquad (0.0268) \qquad\qquad (0.0001) \qquad\qquad (0.0068) + \qquad\qquad (0.0002)
\end{array}
\tag{9.4}
$$

$R^2 = 0.4609 \quad n = 526$

Interpret the model:

9.3.1 Interactions Involving Dummy Variables

8.5.1 The Interactions Among Dummy Variables:

We can recast the model by adding an **interaction term** between female and married to the model where female and married appear separately. This allows the marriage premium to depend on gender. The estimated model with the female-married interaction term is :

$$\begin{aligned}
\widehat{\log(\text{wage})} &= \underset{(0.100)}{0.321} - \underset{(0.056)}{0.110}\ \text{female} + \underset{(0.055)}{0.213}\ \text{married} \\
&\quad + \underset{(0.072)}{0.301}\ female \cdot married + ...,
\end{aligned} \tag{9.5}$$

8.5.2 The interaction between Dummy Variable/s and Explanatory Variable/s: the Allowing for the Different Slopes

There are also occasions for interacting dummy variables with explanatory variables that are not dummy variables to allow for a **difference in slope.**

To see the interaction between female and edu, we can rewrite the model as follow:

$$wage_i = \beta_0 + \delta_0 \text{female} + \beta_1 \text{edu} + \delta_1 female \cdot edu + u_i$$

Men Group we plug female =0

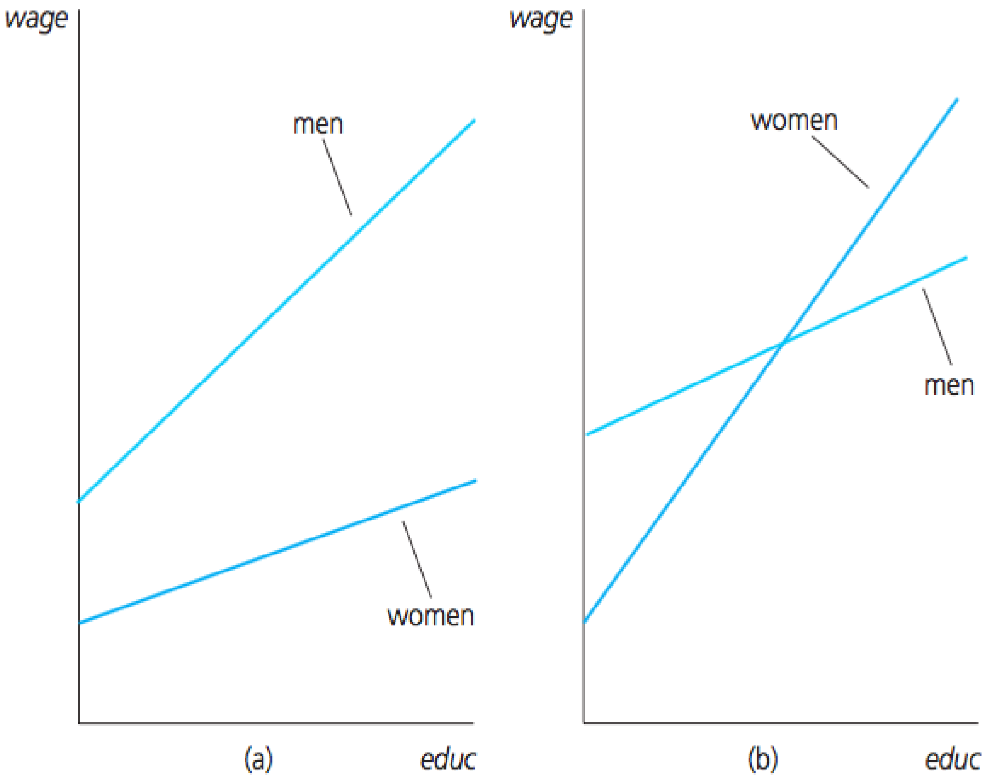
Therefore:

Women Group we plug female =1

Therefore:

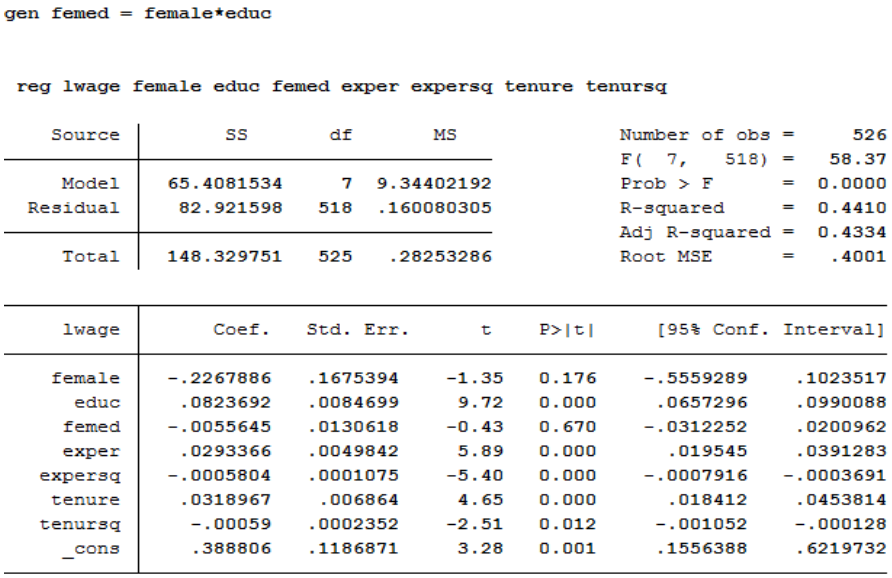
Figure 9.2: Graph of the Wage Model with an Interaction between female and education

Graphs of equation (7.16): (a) $\delta_0 < 0, \delta_1 < 0$; (b) $\delta_0 < 0, \delta_1 > 0$.



Example

Table 9.5:



```
gen femed = female*educ


 reg lwage female educ femed exper expersq tenure tenursq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  7,   518) =   58.37
       Model |  65.4081534     7  9.34402192           Prob > F      =  0.0000
    Residual |   82.921598   518  .160080305           R-squared     =  0.4410
-------------+------------------------------           Adj R-squared =  0.4334
       Total |  148.329751   525  .28253286            Root MSE      =   .4001

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2267886   .1675394    -1.35   0.176    -.5559289    .1023517
        educ |   .0823692   .0084699     9.72   0.000     .0657296    .0990088
       femed |  -.0055645   .0130618    -0.43   0.670    -.0312252    .0200962
       exper |   .0293366   .0049842     5.89   0.000      .019545    .0391283
     expersq |  -.0005804   .0001075    -5.40   0.000    -.0007916   -.0003691
      tenure |   .0318967    .006864     4.65   0.000      .018412    .0453814
     tenursq |    -.00059   .0002352    -2.51   0.012     -.001052    -.000128
       _cons |    .388806   .1186871     3.28   0.001     .1556388    .6219732
------------------------------------------------------------------------------
```

$$
\begin{array}{rcl}
\widehat{\log(\text{wage})} & = & 0.3889 - 0.2268 \ \text{female} + 0.082 \ \text{edu} - 0.0056 \ \textit{female} \cdot \textit{edu} \\
 & & (0.1187) \qquad (0.1675) \qquad\qquad (0.0085) \qquad\qquad (0.0131) \\
 & & +0.0293 \ \textit{exper} - 0.0006 \ \textit{exper}^2 + 0.0319 \ \textit{tenure} - 0.00059 \ \textit{tenure}^2 \\
 & & (0.0050) \qquad\quad (0.0001) \qquad\quad (0.0069) + \qquad\quad (0.0002)
\end{array} \tag{9.6}
$$

$R^2 = 0.4410 \quad n = 526$