TWO-VARIABLE REGRESSION ANALYSIS: SOME BASIC IDEAS

SIMPLE REGRESSION MODEL

- Two variable linear regression model
- Bivariate linear regression model

$$Y = \beta_1 + \beta_2 X + u$$

Y and X are two variables, representing some population

- Explaining Y in terms of X
- Studying how Y varies with changes in X

EXAMPLE

- A total population of 60 families in a hypothetical community and their weekly income (X) and weekly consumption expenditure (Y), both in dollars
- The 60 families are divided into 10 income groups

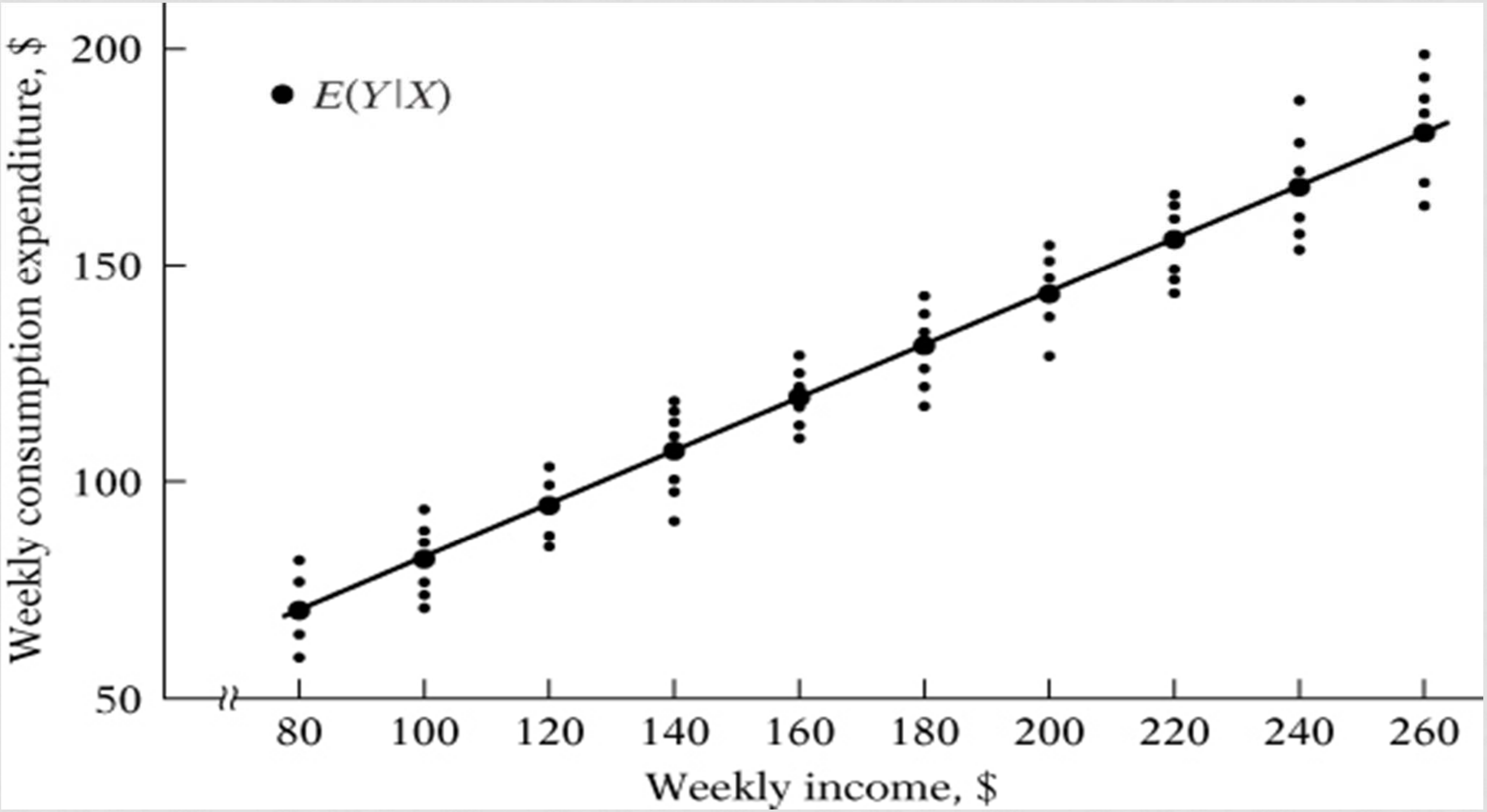
Y \ X	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**
Weekly family consumption expenditure Y, $	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Total	325	462	445	707	678	750	685	1043	966	1211
$E(Y \mid X)$	**65**	**77**	**89**	**101**	**113**	**125**	**137**	**149**	**161**	**173**

CONDITIONAL EXPECTED VALUES

$$E(Y \mid X)$$

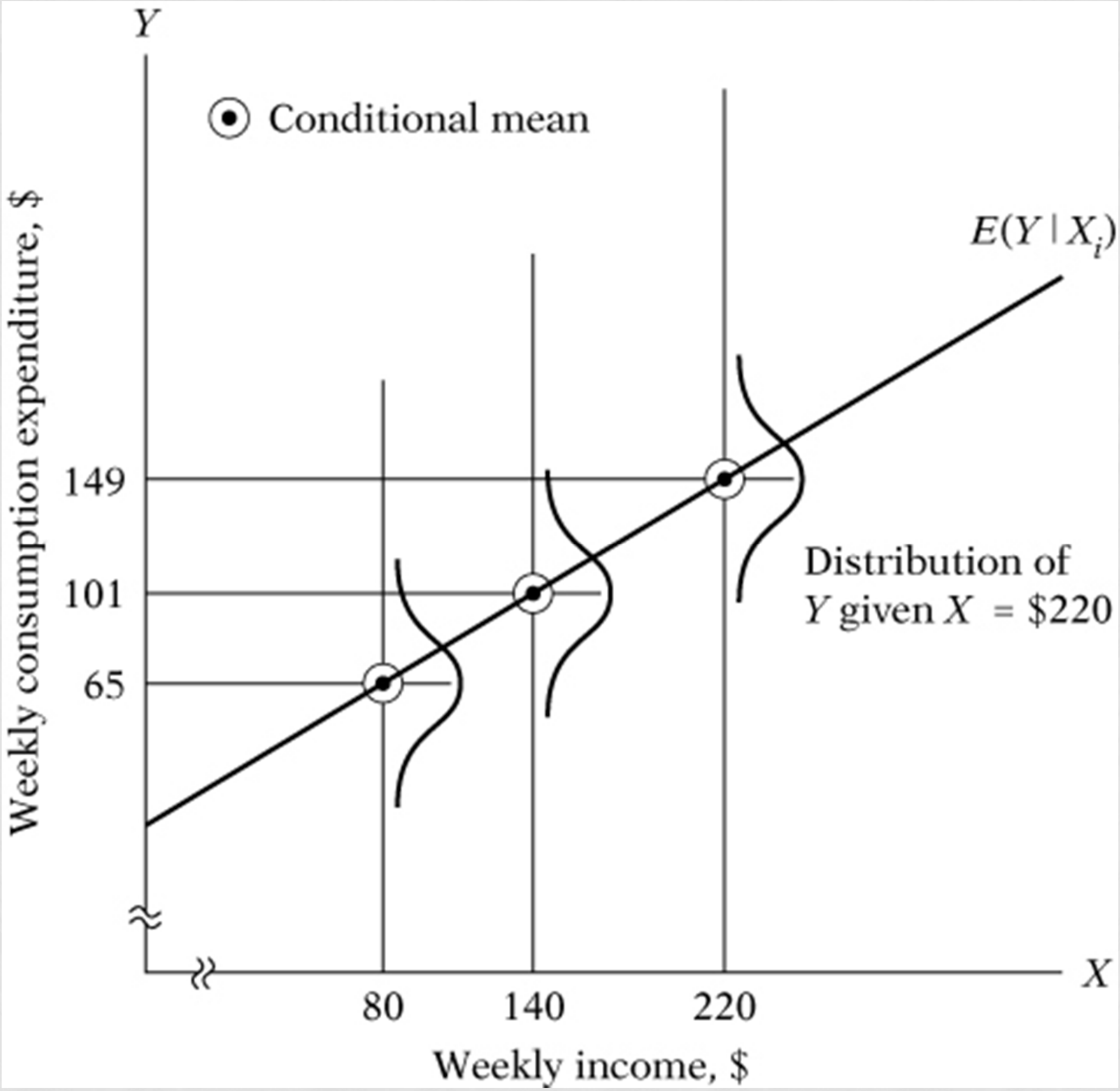
The expected value of Y given the value of X

$P(Y\|X_i)$ \ X	80	100	120	140	160	180	200	220	240	260
Conditional probabilities $P(Y\|X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	-	$\frac{1}{6}$	-	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	-	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	-	-	-	$\frac{1}{7}$	-	-	-	$\frac{1}{7}$	-	$\frac{1}{7}$
$E(Y\|X)$	**65**	**77**	**89**	**101**	**113**	**125**	**137**	**149**	**161**	**173**


E(Y|X)
Weekly consumption expenditure, $
200
150
100
50
80
100
120
140
160
180
200
220
240
260
Weekly income, $

POPULATION REGRESSION LINE (PRL)

- A population regression curve is simply the locus of the conditional means of the dependent variable for the fixed values of the explanatory variable (s)


Y
Conditional mean
E(Y | Xi)
Weekly consumption expenditure, $
149
101
65
Distribution of
Y given X = $220
X
80
140
220
Weekly income, $

POPULATION REGRESSION FUNCTION (PRF)

- Conditional Expectation Function (CEF)
- Population Regression Function (PRF)
- Population Regression (PR)

$$E(Y \mid X_i) = f(X_i)$$

Each conditional mean $E(Y|X_i)$ is a function of X_i, where X_i is a given value of X.

LINEAR POPULATION REGRESSION FUNCTION

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

β_1 is known as intercept

β_2 is known as slope coefficients

LINEAR

- Linear in the Variables

- Linear in the Parameters

LINEAR IN THE VARIABLES

- The regression curve is a straight line
- X's are raised to the first power only

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

LINEAR IN THE PARAMETERS

- The conditional expectation of Y is a linear function of the parameters, the $\beta's$

- May or may not be linear in the variable X

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2$$

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3$$

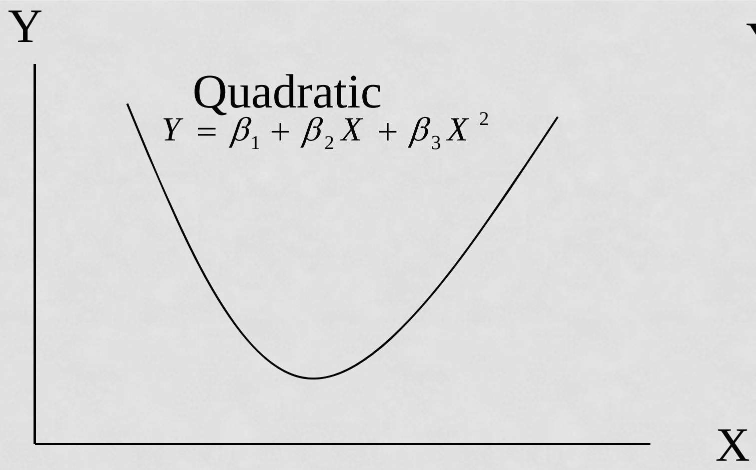
$$Y = e^{\beta_1 + \beta_2 X}$$

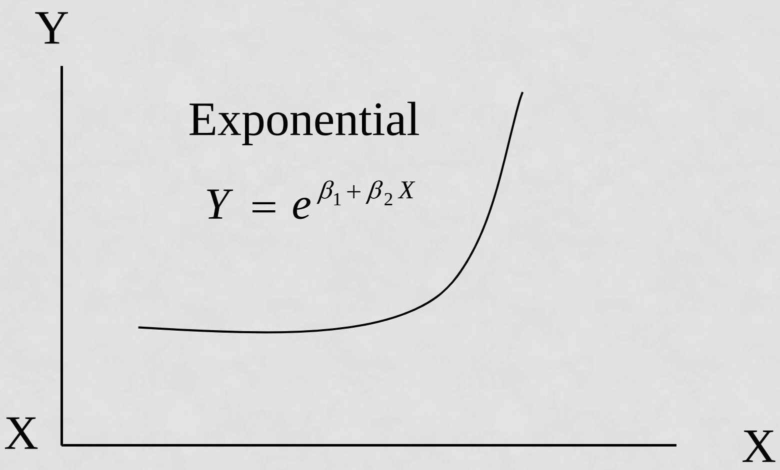
LINEAR REGRESSION

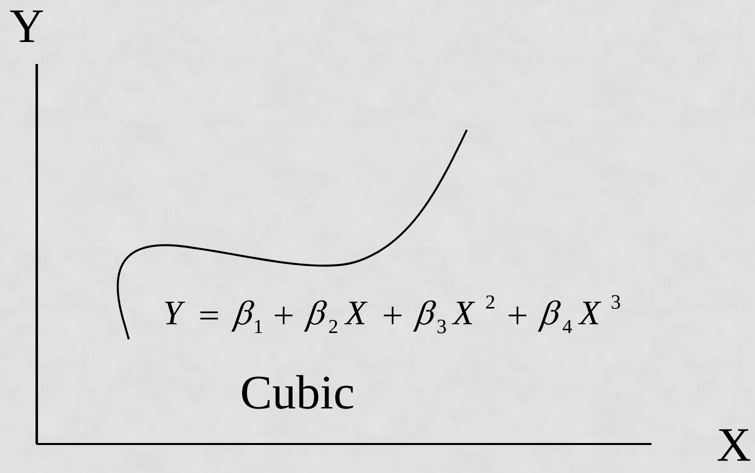
- Linear regression will always mean a regression that is linear in the parameters; the β's are raised to the first power only.

- May or May not be linear in the explanatory variables, the X's

LINEAR IN PARAMETER FUNCTIONS







ARE THE FOLLOWING MODELS LINEAR REGRESSION MODELS?

$$a.\ Y_i = e^{\beta_1 + \beta_2 X_i + u_i}$$

$$b.\ \ln Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i} \right) + u_i$$

$$c.\ Y_i = \beta_1 + \beta_2^3 X_i + u_i$$

$$d. \ln Y_i = \beta_1 + \beta_2 X_i + u_i$$

STOCHASTIC SPECIFICATION OF PRF

The deviation of an individual Y_i around its expected value as follows:

$$u_i = Y_i - E(Y \mid X_i)$$

$$Y_i = E(Y \mid X_i) + u_i$$

STOCHASTIC SPECIFICATION OF PRF

How do we interpret this equation?

$$Y_i = E(Y \mid X_i) + u_i$$

(1) $E(Y \mid X_i)$ is the mean consumption expenditure of all the families with the same level of income- This component is known as the **systematic or deterministic component**

(2) u_i is the random or nonsystematic component

$$E(Y_i \mid X_i) = E[E(Y \mid X_i)] + E(u_i \mid X_i)$$
$$= E(Y \mid X_i) + E(u_i \mid X_i)$$

Since $E(Y_i|X_i)$ is the same thing as $E(Y|X_i)$ implies that

$$E(u_i \mid X_i) = 0$$

THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM

- Vagueness of theory
- Unavailability of data
- Intrinsic randomness in human behavior
- Poor proxy variables
- Wrong functional form

THE SAMPLE REGRESSION FUNCTION (SRF)

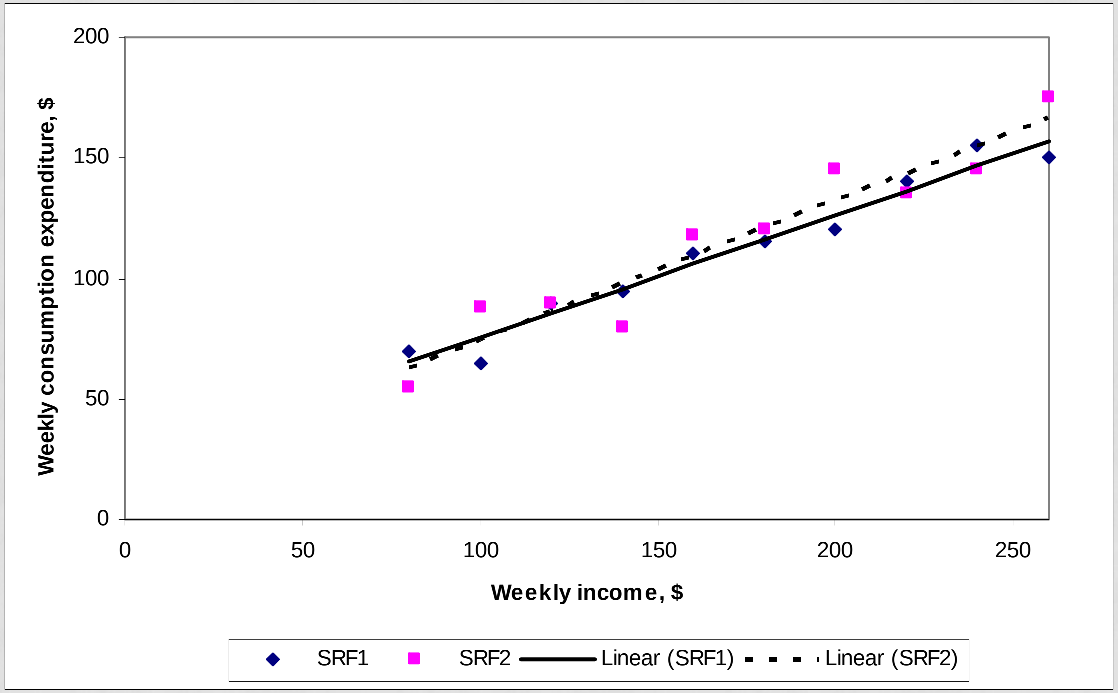
A Random Sample from the Population of Table 2.1

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Another Random Sample from the Population of Table 2.1

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

THE SAMPLE REGRESSION FUNCTION (SRF)



THE SAMPLE REGRESSION FUNCTION (SRF)

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

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

$\hat{\beta}_1$ = estimator of β_1

$\hat{\beta}_2$ = estimator of β_2

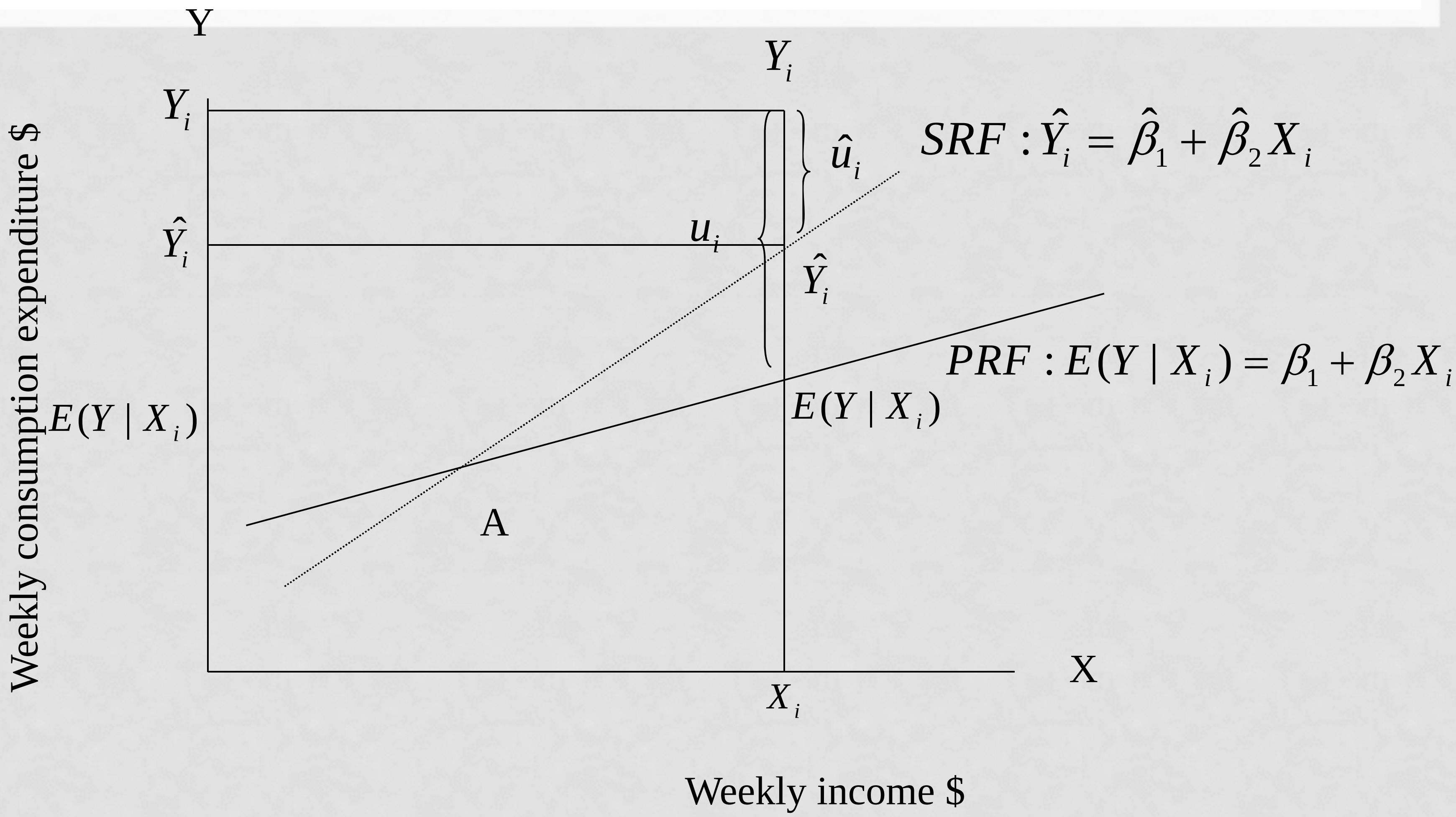
The PRF

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

On the basis of the SRF

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

THE SAMPLE REGRESSION FUNCTION (SRF)



ILLUSTRATIVE EXAMPLES

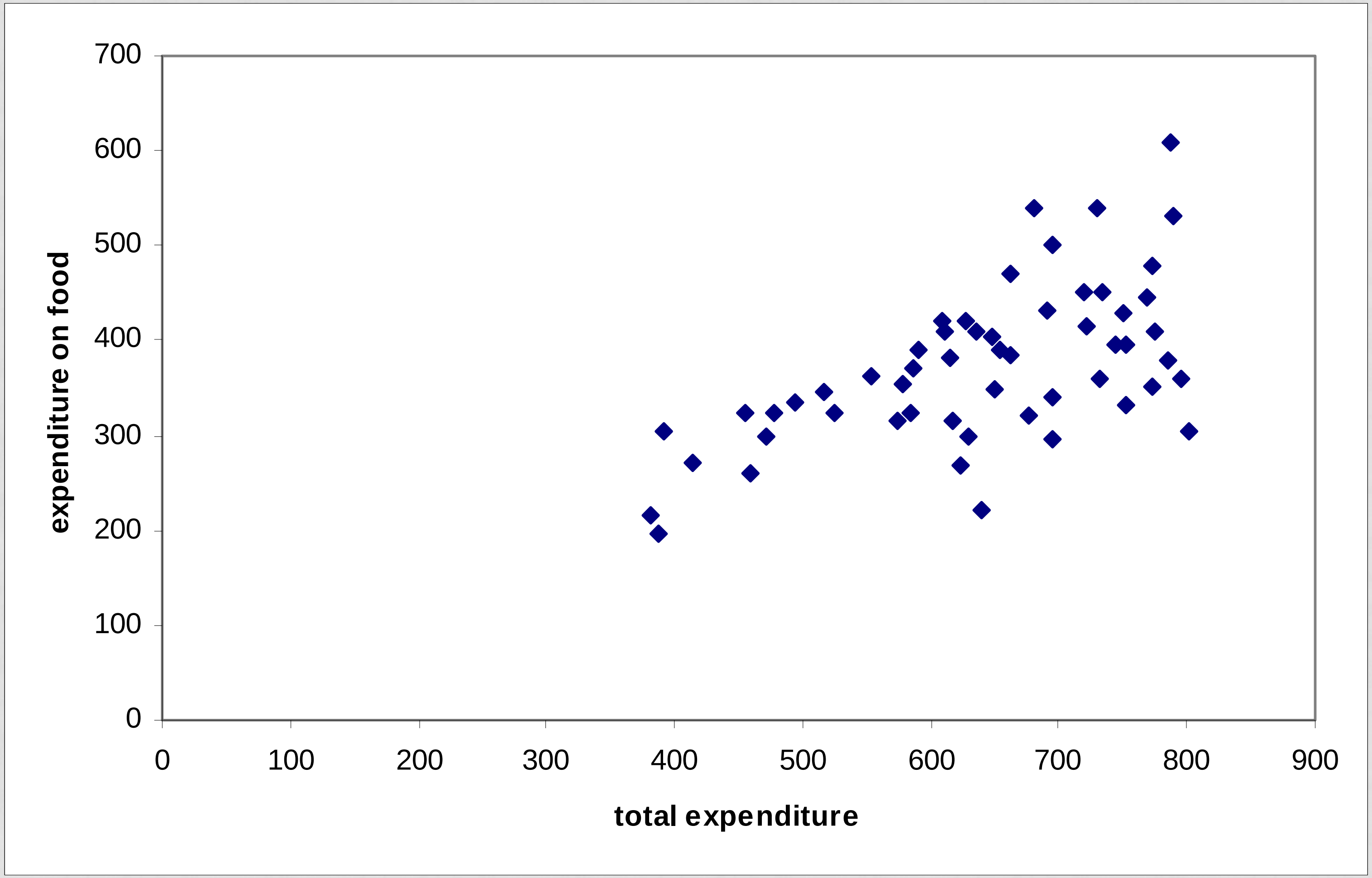
Table 2.8 Data on expenditure on food and total expenditure, measured in rupees, for a sample of 55 rural households from India. (In early 2000, a U.S. dollar was about 40 Indian rupees)

Plot the data, using the vertical axis for expenditure on food and the horizontal axis for total expenditure

TABLE 2.8 **Food and Total Expenditure (Rupees)**

Observation	Food Expenditure	Total Expenditure	Observation	Food Expenditure	Total Expenditure
1	217.0000	382.0000	29	390.0000	655.0000
2	196.0000	388.0000	30	385.0000	662.0000
3	303.0000	391.0000	31	470.0000	663.0000
4	270.0000	415.0000	32	322.0000	677.0000
5	325.0000	456.0000	33	540.0000	680.0000
6	260.0000	460.0000	34	433.0000	690.0000
7	300.0000	472.0000	35	295.0000	695.0000
8	325.0000	478.0000	36	340.0000	695.0000
9	336.0000	494.0000	37	500.0000	695.0000
10	345.0000	516.0000	38	450.0000	720.0000
11	325.0000	525.0000	39	415.0000	721.0000
12	362.0000	554.0000	40	540.0000	730.0000
13	315.0000	575.0000	41	360.0000	731.0000
14	355.0000	579.0000	42	450.0000	733.0000
15	325.0000	585.0000	43	395.0000	745.0000
16	370.0000	586.0000	44	430.0000	751.0000
17	390.0000	590.0000	45	332.0000	752.0000
18	420.0000	608.0000	46	397.0000	752.0000
19	410.0000	610.0000	47	446.0000	769.0000
20	383.0000	616.0000	48	480.0000	773.0000
21	315.0000	618.0000	49	352.0000	773.0000
22	267.0000	623.0000	50	410.0000	775.0000
23	420.0000	627.0000	51	380.0000	785.0000
24	300.0000	630.0000	52	610.0000	788.0000
25	410.0000	635.0000	53	530.0000	790.0000
26	220.0000	640.0000	54	360.0000	795.0000
27	403.0000	648.0000	55	305.0000	801.0000
28	350.0000	650.0000			

Source: Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries,* Routledge, New York, 1998, p. 457.



700
600
500
400
300
200
100
0
expenditure on food
0
100
200
300
400
500
600
700
800
900
total expenditure

- As total expenditure increases, on the average, expenditure on food also increases.

- We would not expect the expenditure on food to increase linearly for ever. Once basic needs are satisfied, people will spend relatively less on food as their income increases. That is, at higher levels of income consumers will have more discretionary income.

SOURCE

Gujarati, D.N. (2009) Basic Econometrics. 5th ed. Singapore, McGraw-Hill.