DATA ANALYSIS, EDITING, CODING

UNIT 5

PRESENTED BY HIMANSHU GOEL

DATA PROCESSING:

The data after collection has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan.

The data processing & analysis is essential in recording information, analyzing the information & communicating analysis.

Importance:

- Check the data accuracy
- Provide better understanding
- Puts into a suitable form
- Helps in decision making
- Makes data transferable
- Readability

Editing:

The raw data is likely to certain no. of errors during the process of recording the information in surveys. By means of editing one tries to eliminate the errors or remove the points of confusion.

Essentials of Editing:

- Completeness
- Accuracy
- Uniformity

Stages of editing:

- Field editing
- Office/ central editing

Coding: Coding is the procedure of classification the answers to a question into meaningful categories. The symbols used to indicate these categories are called codes.

Essentials coding:

- Appropriate to the research problems
- Exhaustive
- Mutually exclusive
- Single dimension
- Code sheets

Procedure of coding:

- Identification of open coding
- Axial Coding
- Selective Coding

Tabulation:

Tabulation is the primary function of data analysis. The data is validated & analysed to generate tables in a client-specified format that helps the researcher to interpret the results of the survey & present it to his/her client.

According to Blair: "Tabulation in its broadest sense is an orderly arrangement of data in columns & rows"

TABLE

A rectangular arrangement of data in which the data are positioned in rows and columns.

PARTS of TABLE

Table 1.

Records of Individual Points For Every Group Activities of BSED BIOSCI 3-1

TITLE up 1

	Activity					
Name of Members	1	2	3	4	5	Total
2. Do**, Jan**	15	25	20	25	15	100
3. Ebo, Jo**	13	23	17	24	14	91
4. Gob, Ma**	14	23	20	23	13	93
6. Jug, Be**	23	19	21	12		
7. Mar, Ma**	22	17	23			

** Add 2 points

- Provides a brief description of the contents of the table.
- It should be concise and include the key elements shown in the table, for example, groups, classifications, variables, etc.
- Always at the top part of the table

Table 1. Records of Individual Points For Every Group Activities of BSED BIOSCI 3-1

TABLE NUMBER

Name of Members	Activity 1	2	3	4	5	Total
	15	25	20	25	15	100
1. Bajado, Joshua	13	25	19	20	13	90
2. Doon, James Marlou **	21				12	67
3. Eboni te,		2	3		14	91
Jo hn Ch ris to p er		1 7			13	93
5. Hallasgo, John Paul	15	24	13	22	15	89
6. Junio, Benidict *		23	19	21	12	75
7. Manrique, Rudive	12	22	17	23	11	85

Note: Black Mark signifies that the student is absent when the activity is held.

* Add 3 points

Table must be numbered properly to avoid confusion

Table 1. Records ... For Every ... Group ... [OSCI 3-1

Heading

Group 1	Activity					
	1	2	3	4	5	Total
...ame of Members	15	5	20	25	5	100
1. B...						
2. D...						
5. H...						
6. Ju...						
7. M...						
No...						

- Keywords that allow you to establish an order to the data.
- Should be written in the singular form
- The first letter of the first word should be capitalized.

Stub

Table Spanner

in a table occ ns, for
percentages, fre al test
ults, means, "N" (numb

Name of Members	15	25	20	25	15	100
1. Ba do, Joshua	13	25	19	20	13	90
2. Do on, James Marlou **	12	22		21	12	67
3. Ebonite, John Christoper	13	23	17	24	14	91
4. Glodo, Mellan	14	23	20	23	13	93
5. Hallasgo, John Paul	15	24	13	22	15	89
6. Junio, Benidict *		23	19	21	12	75
7. Manrique, Rudive	12	22	17	23	11	85

Table Body

Note: Black Mark ■ sign the activity is held.

* Add 3 points

** Add 2 points

- **You may use table notes to explain anything in your table that is not self-explanatory.**

Dividers

Name of Members		25	20	25	15	100
1. Bajado, Joshua		25	19	20	13	90
2. Doon, James Marlou **	12	22		21	12	67
3. Ebonite, John Christoper				24	14	91
4. Glodo, Mellan				23	13	93
5. Hallasgo, John Paul	15	24	13	22	15	89
6. Junio, Benidict *		23	19	21	12	75
7. Manrique, Rudive	12	22	17	23	11	85

Table Notes

Note: Black Mark ■ signifies that the student is absent when the activity is held.

* Add 3 points

** Add 2 points

TYPES Of TABLE

According to Kind of Variables

Textual Table

Table 2. Observations of Catalyst Reactions Under Boiling or Moderate Heat Conditions

Catalyst	Reaction Intensity	Boiling Temp. (y/n)
Organic		
A	none	y
B	high	n
C	low	n
Inorganic		
A	high	y
B	moderate	n
C	low	y

Numerical Table

Table 4. Demographic Composition of White-Tailed Deer Prehunting Populations in North Carolina on a 30,000 Acre Area from 1965-2000

	Males			Females			
Year	Adults	Yearlings	Fawns	Adults	Yearlings	Fawns	Total
1965	307	135	442	1002	265	462	2613
1970	333	222	318	1069	228	332	2458
1975	235	162	260	887	183	271	2325
1980	221	130	450	900	250	462	2502
1985	190	112	320	862	230	360	1998
1990	165	220	289	782	216	234	2413
1995	185	132	476	1041	218	406	2074
2000	155	312	302	911	315	330	2325

Statistical Table

Table 3. ANOVA Table for Two-Way Anylysis of Variance

Source	DF	Mean Square	*F*-Value	Prob > *F*
Between Subjects				
Treatment	2	315.20	5.3	0.003
Error	70	67.90		
Within Subjects				
Time	1	128.30	7.6	0.003
Time x Treatment	2	95.36	5.6	0.006
Error	70	16.30		

According to Number of Variables

One-Way Table

Problem

Twenty first graders were asked which color they liked best - red, green, or blue. Their responses appear below.

red, green, green, blue, red, blue, red, red, blue, red red, blue, red, red, blue, red, red, blue, red, blue, green, green, red

red, green, green, blue, red,
blue, red, red, blue, red red,
blue, red, red, blue, red, blue,
green, green, red

Table. 1 20 First Grader Best Liked Color Among Red, Green, and Blue.

Choice	Red	Green	Blue
Response	10	4	6

Frequency Table

Table. 1

20 First Grader Best Liked Color Among Red, Green, and Blue.			
Choice	Red	Green	Blue
Response	50%	20%	30%

Percentage Table

Table. 1

20 First Grader Best Liked Color Among Red, Green, and Blue.			
Choice	Red	Green	Blue
Response	0.5	0.2	0.3

Proportion Table

Two-Way Table

Consisting of 20 men and 30 women.

Adult	Dance	Sports	TV	Total
Men	2	10	8	20
Women	16	6	8	30
Total	18	16	16	50

	For	Against	No opinion	Total
21 - 40	25	20	5	50
41 - 60	20	35	20	75
Over 60	55	15	5	75
Total	100	70	30	200

A public opinion survey explored the relationship between age and support for increasing the minimum wage. The results are summarized in the two-way table.

In the 21 to 40 age group, what percentage supports increasing the minimum wage?

(A) 12.5%
(B) 20%
(C) 25%
(D) 50%
(E) 75%

Three-Way Table

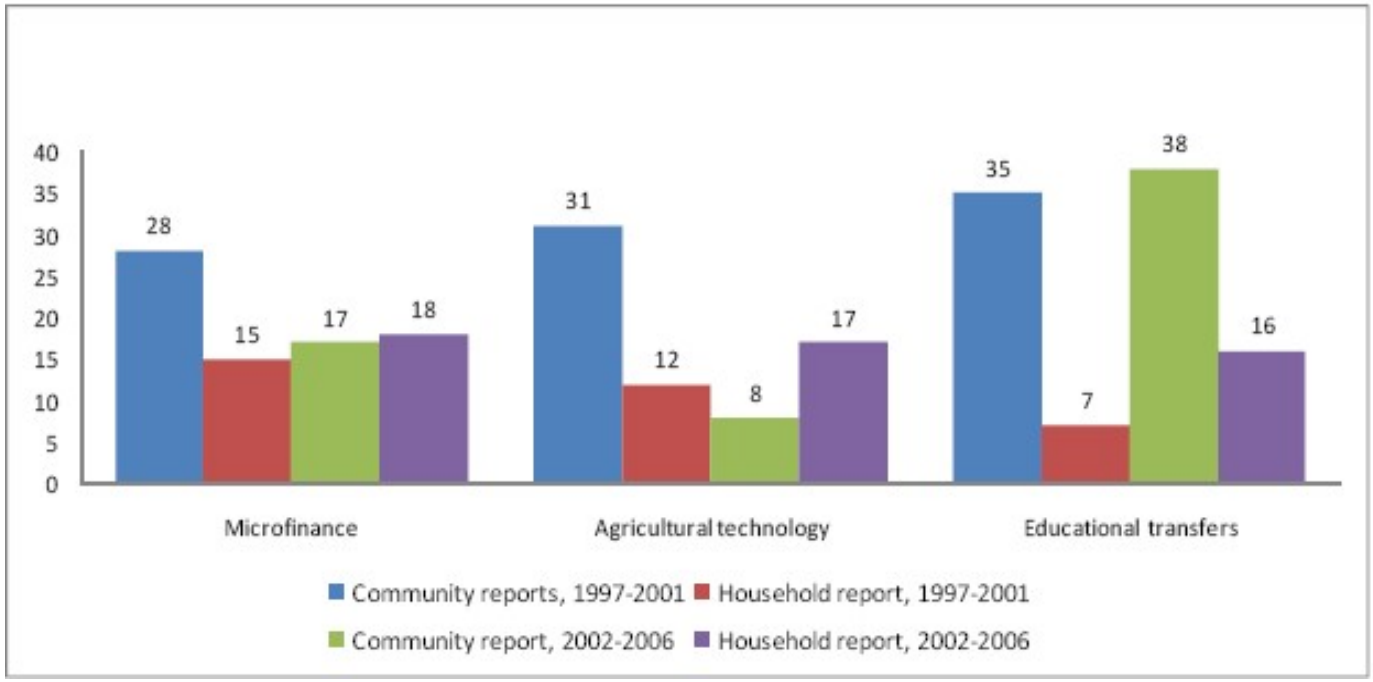
Dept.	Men		Female	
	Rejected	Accepted	Rejected	Accepted
A	313	512	19	89
B	207	353	8	17
C	205	120	391	202
D	278	139	244	131
E	138	53	299	94
F	351	22	317	24

6x2x2

Bar Diagram

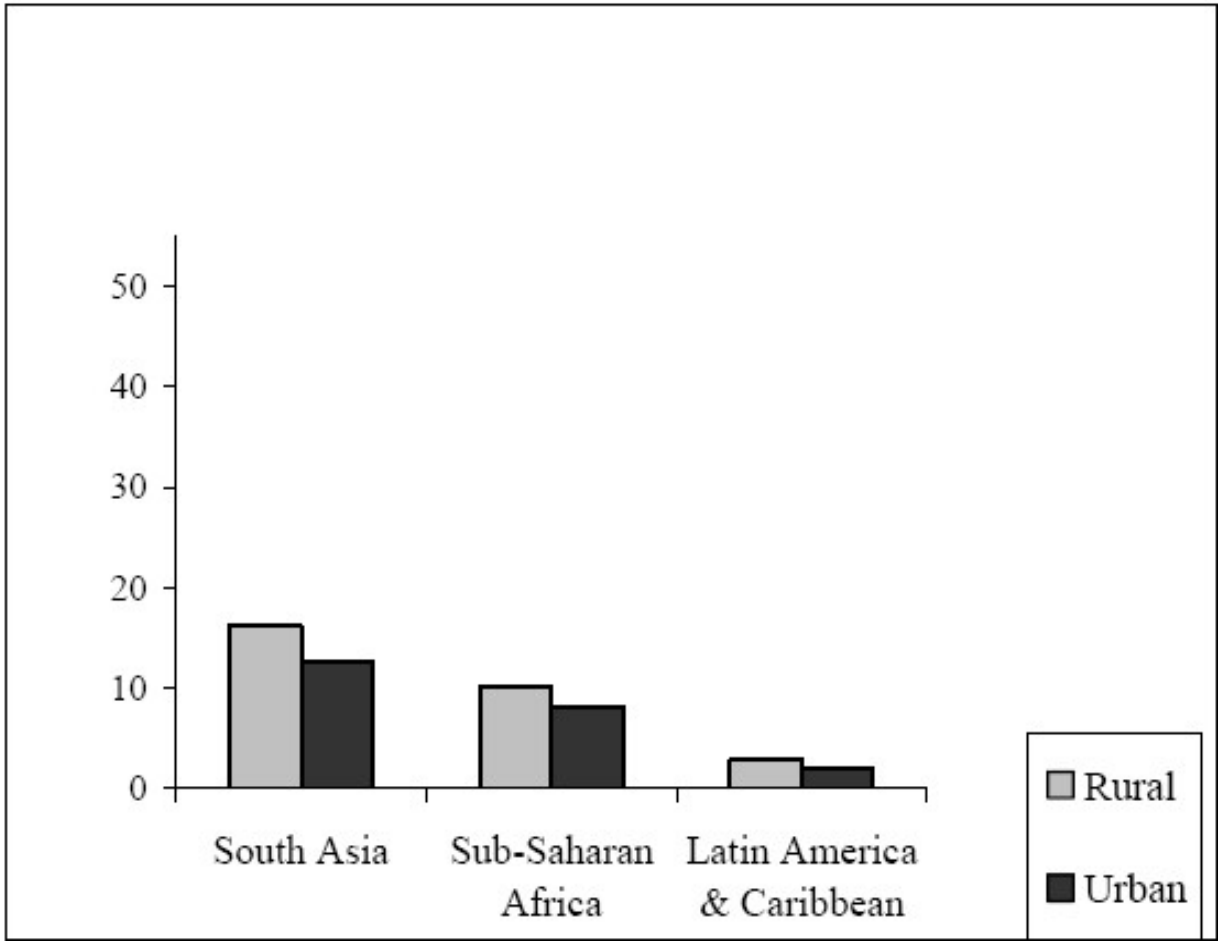
- **A graphical method for depicting qualitative data.**
- **Specify the labels for each of the classes on the horizontal axis.**
- **Scale the vertical axis with reference to frequency, relative frequency, or percent frequency .**
- **Draw bars of fixed width above each class with heights corresponding to the frequency.**
- **Bars are separated to convey the information that each class is a separate category.**

Figure 2. Proportion of households affected by floods, 1997–2001 and 2002–2006



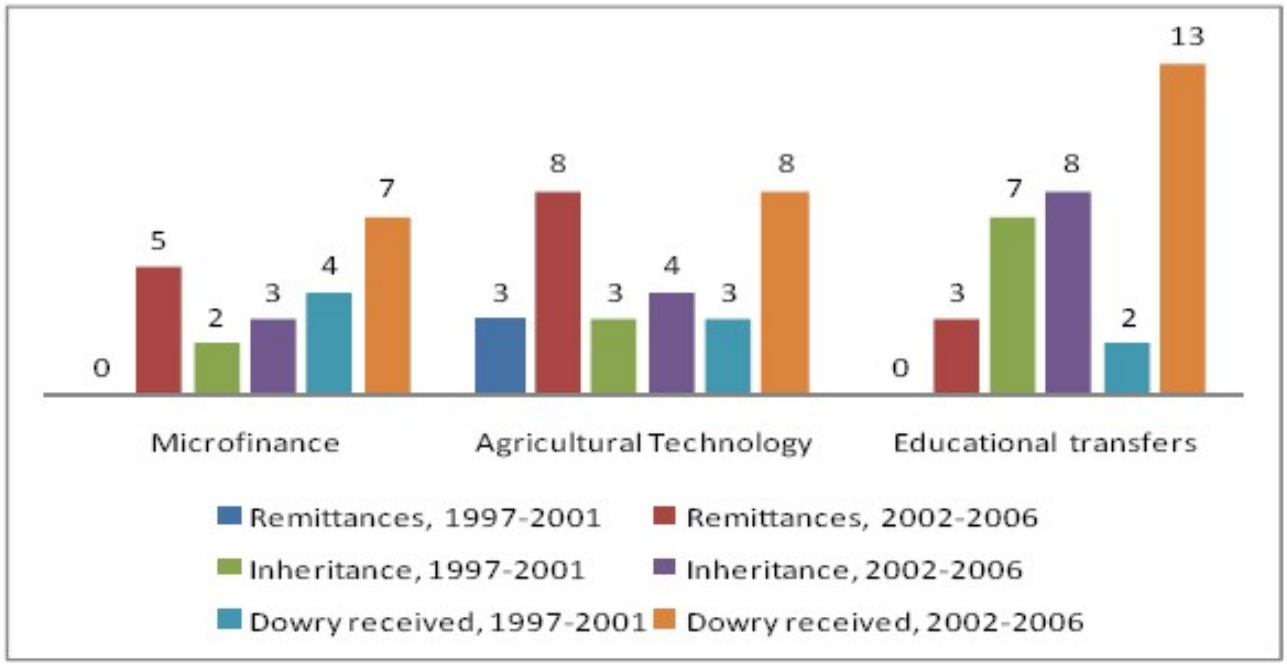
Source: Quisumbing & Baulch, CPRC No. 143

Figure 2—Wasting prevalences across urban and rural areas, by region



Source: IFPRI FAND ,Working paper No 176

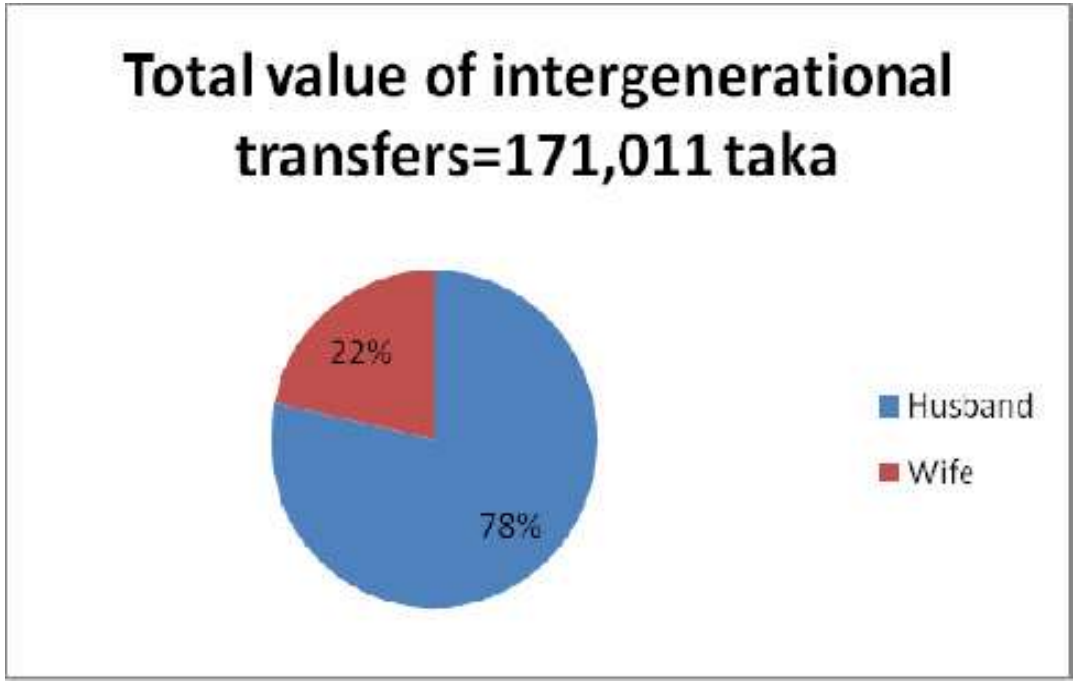
Figure 4. Proportion of households reporting positive events, 1997–2001 and 2002–2006



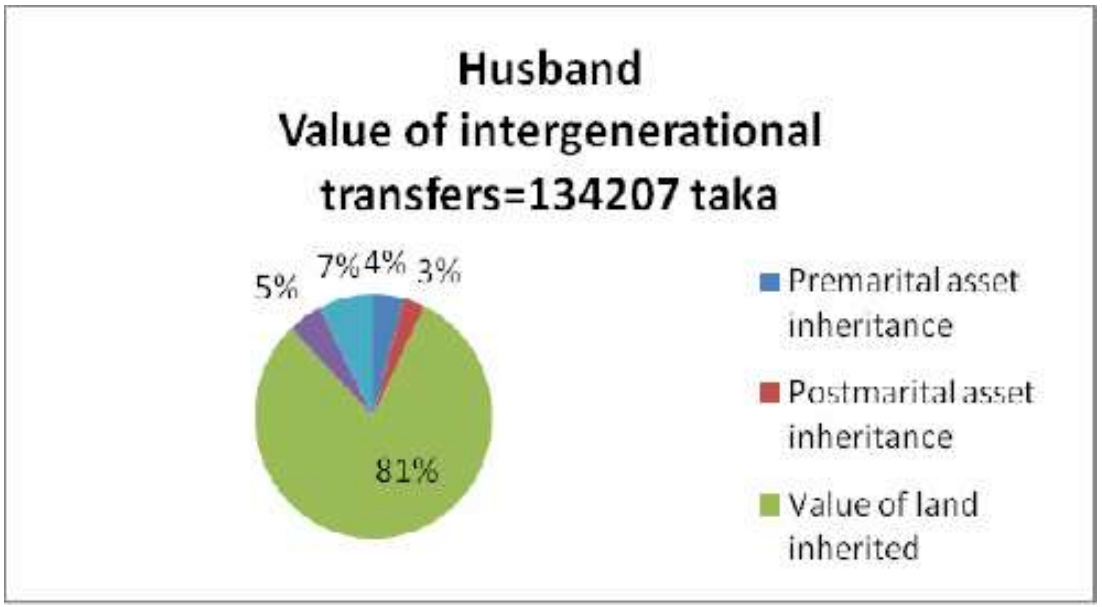
Pie Chart

- **A graphical tool to present relative frequency distributions for qualitative data.**
- **Draw a circle; subdivide the circle into sectors to represent the relative frequency for each class.**
- **For example, a class with a relative frequency of .25 would consume .25(360) = 90 degrees of the circle.**

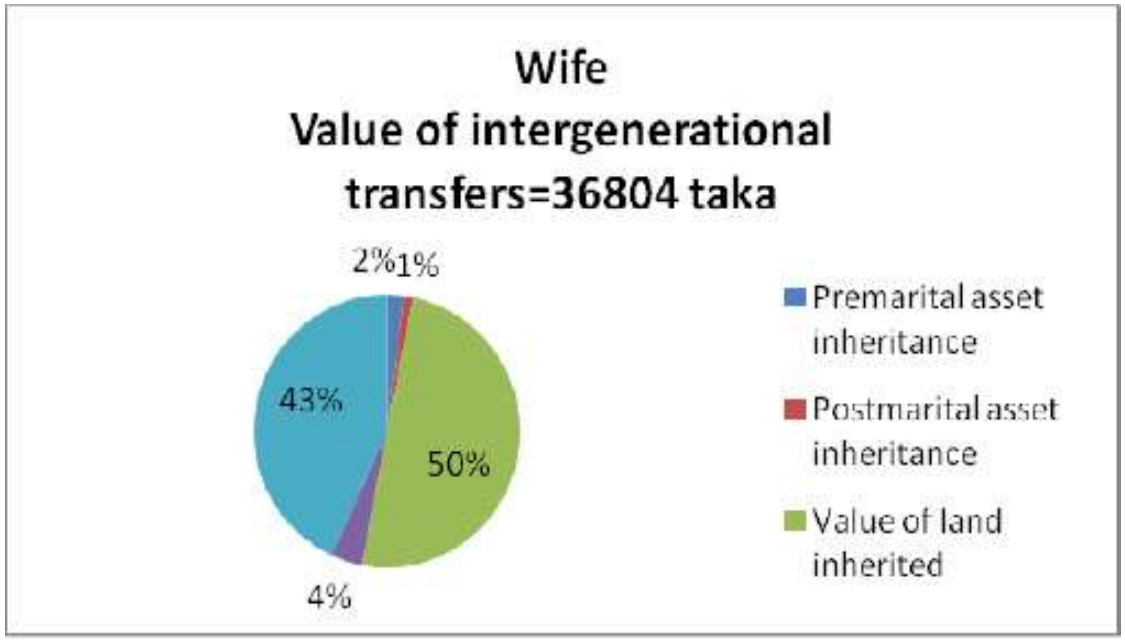
Distribution of intergenerational transfers of husbands and wives, by type of transfer



Source: Quisumbing, CPRCE Working paper No 117



Source: Quisumbing, CPRCE Working paper No 117



Note: All values in 2007 taka.

Source: Quisumbing, CPRCE Working paper No 117

Graphical Presentation of Data

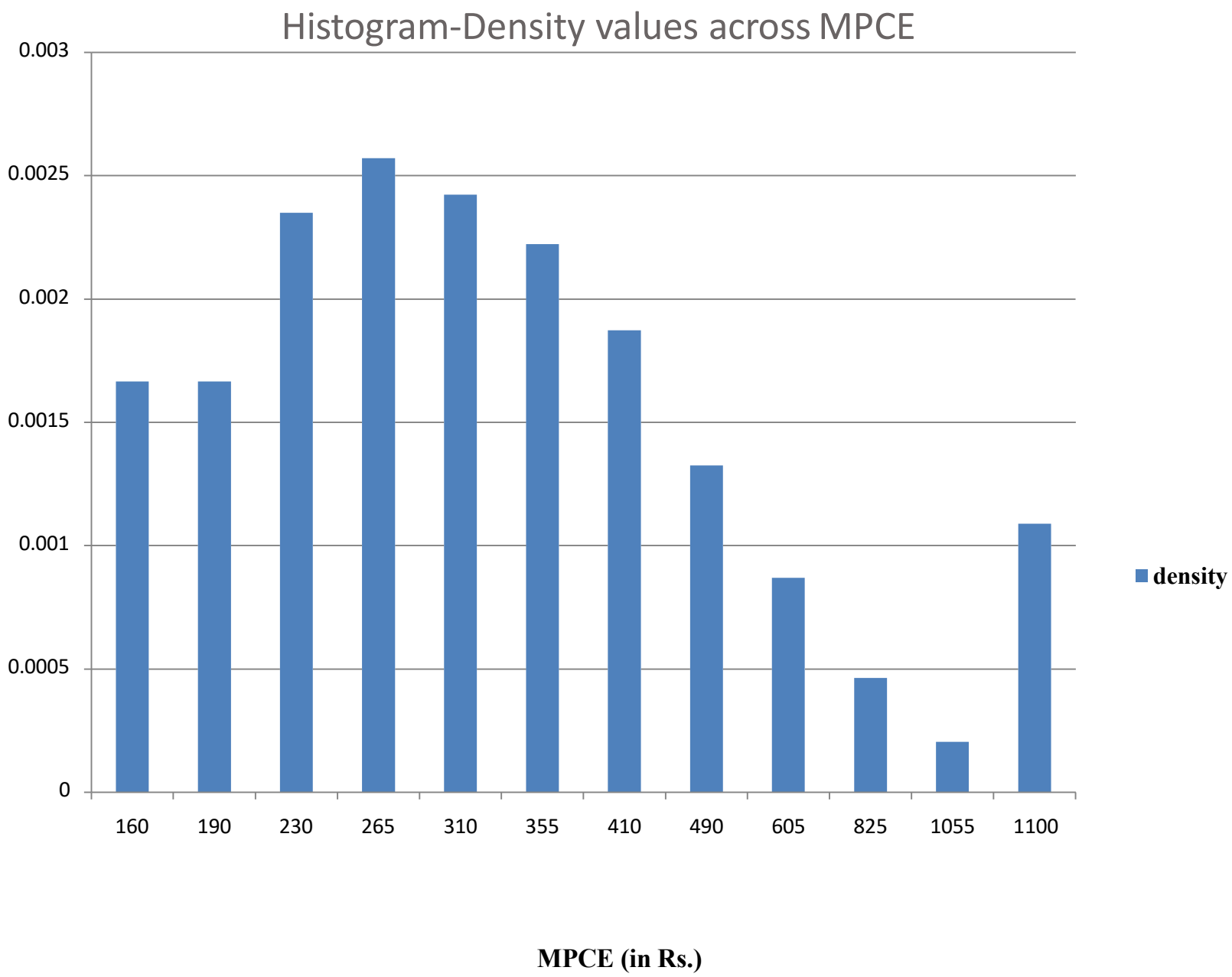
Quantitative Data

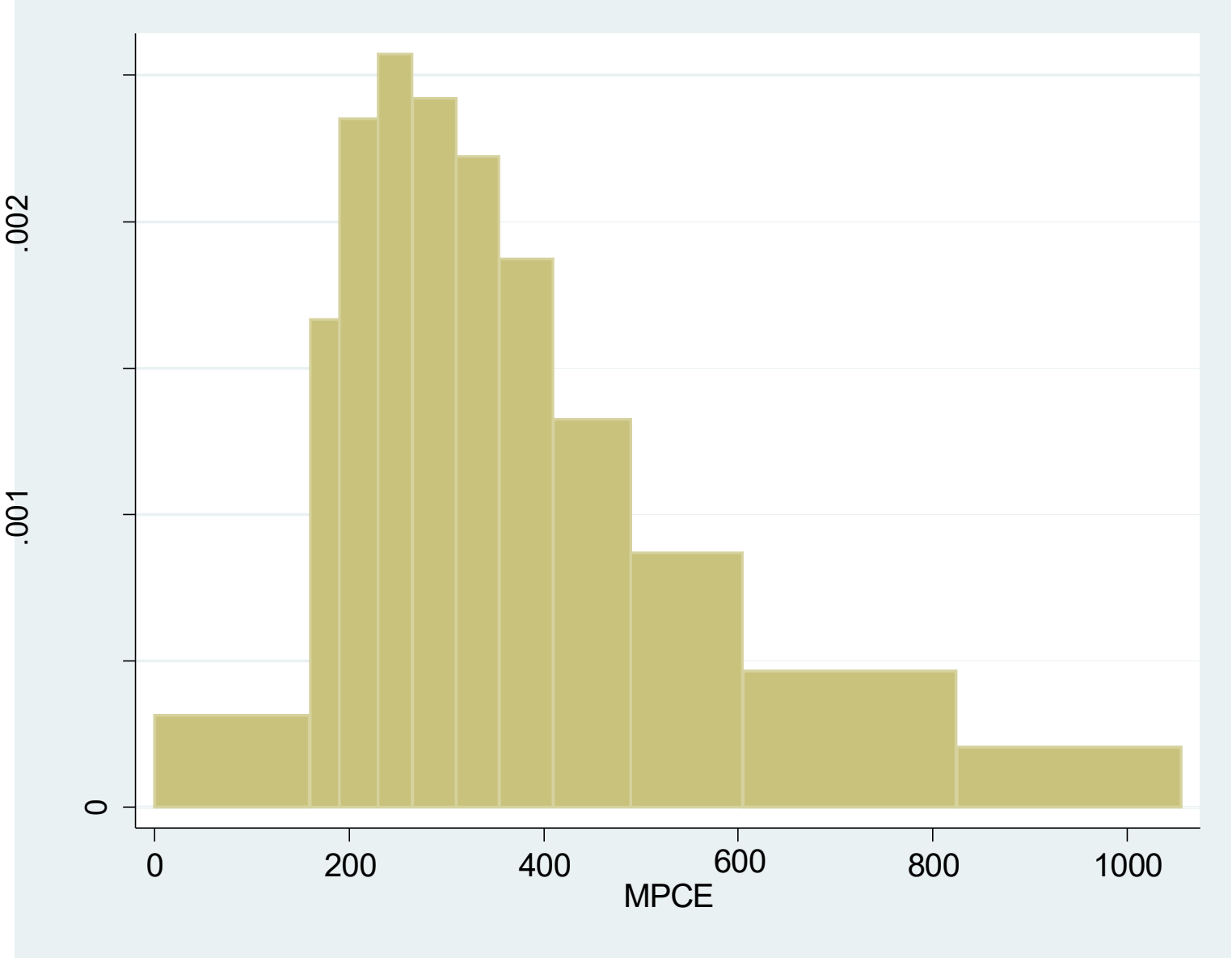
Histogram

- **Measure variable under review on the horizontal axis.**
- **Draw a rectangle above each class interval with its area corresponding to the interval's frequency, relative frequency, or percent frequency; plot frequency density if the class intervals are of unequal width.**
- **Unlike a bar graph, a histogram does not separate between rectangles of adjacent classes.**

The National Sample Survey data on consumer expenditure distribution for urban all-India for 1993-94 is as follows. Draw a (relative frequency) histogram.

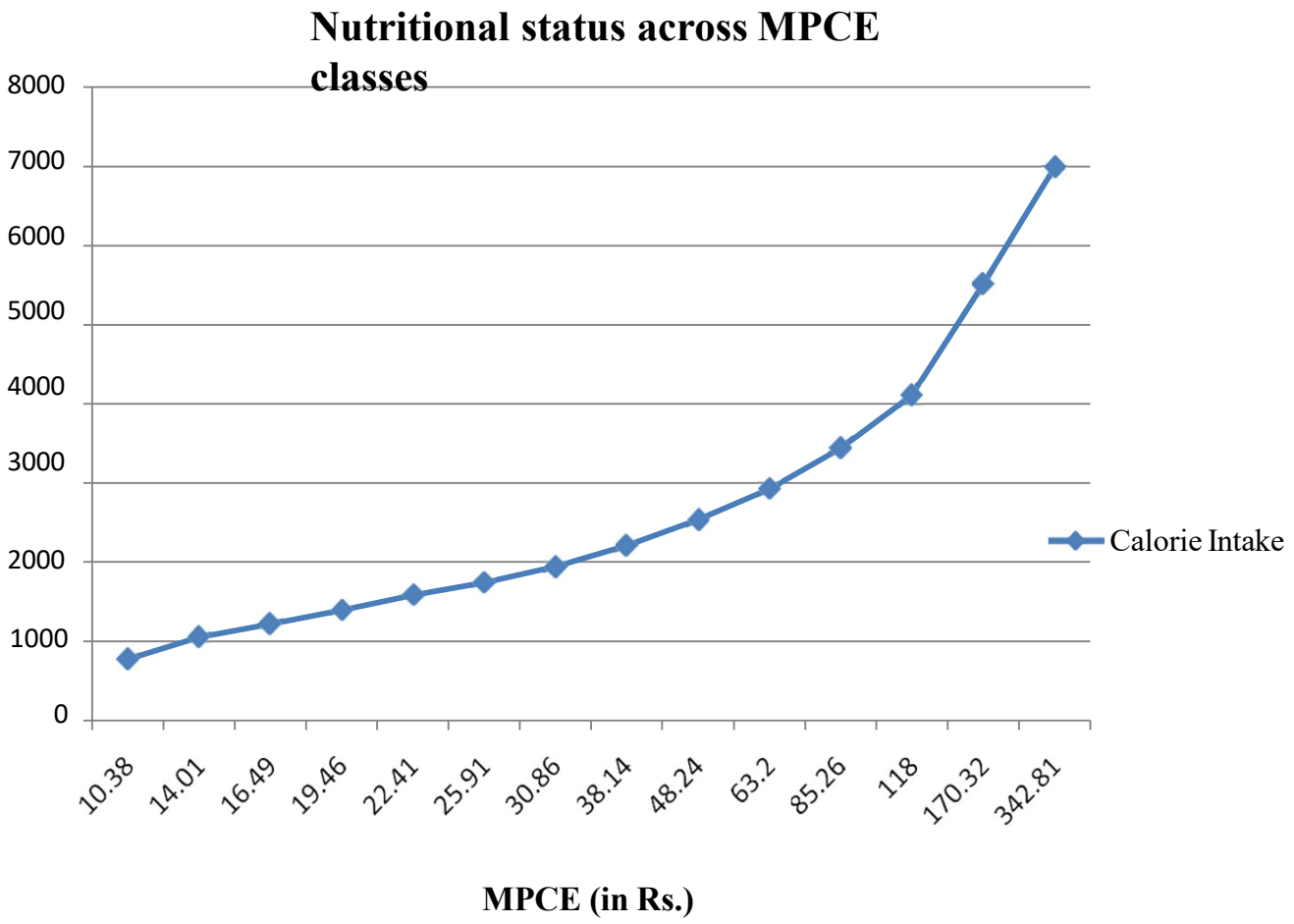
Monthly per capita expenditure class	Per thousand no of persons	Monthly per capita expenditure
< 160	50	132.84
160 - 190	50	175.52
190 - 230	94	210.80
230 - 265	90	247.51
265 - 310	109	286.84
310 - 355	100	331.57
355 - 410	103	380.72
410 - 490	106	447.58
490 - 605	100	543.46
605 - 825	102	698.33
825 - 1055	47	923.38
≥ 1055	49	1643.06
All Classes	1000	458.04

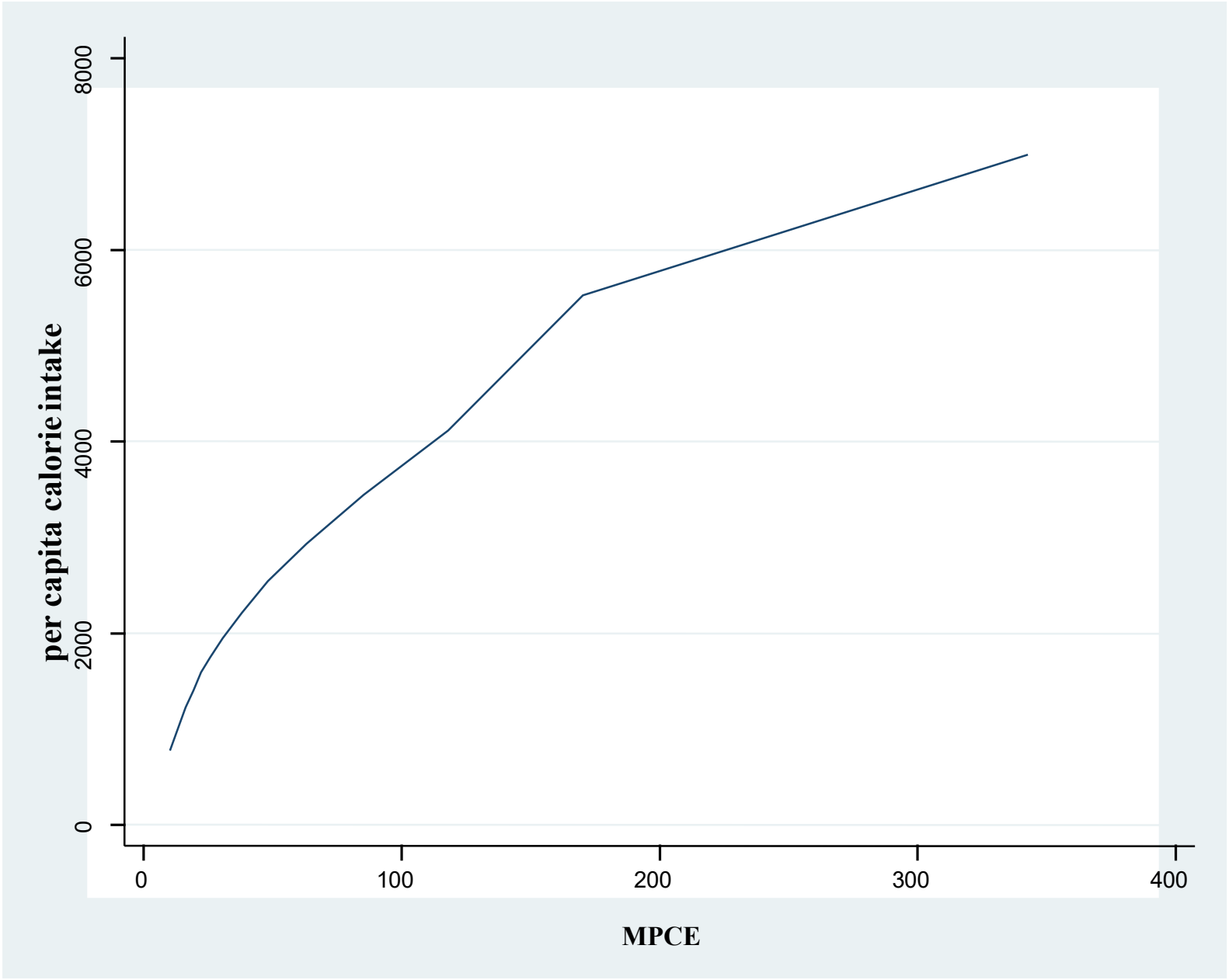

Histogram-Density values across MPCE
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
160
190
230
265
310
355
410
490
605
825
1055
1100
density
MPCE (in Rs.)


.002
.001
0
0
200
400
600
800
1000
MPCE

Estimates of nutritional status of households across per capita monthly expenditure classes in 1972/73 are provided in the Table given below.

Class Intervals	Per Capita Calorie Intake	Percentage of Population	Monthly Per Capita Expenditure	% of Total Expenditure on Food
Poor				
0-13	776	1.67	10.38	81.41
13-15	1055	1.34	14.01	82.73
15-18	1220	3.36	16.49	82.65
18-21	1398	5.13	19.46	82.52
21-24	1586	6.46	22.41	82.41
24-28	1743	10.09	25.91	81.68
28-34	1944	15.58	30.86	81.70
Non-Poor				
34-43	2210	19.07	38.14	78.85
43-55	2538	16.25	48.24	75.56
55-75	2929	11.87	63.20	71.56
75-100	3439	5.31	85.26	66.05
100-150	4110	2.90	118.00	59.42
150-200	5521	0.56	170.32	52.00
200+	6991	0.47	342.81	38.22
All	2266	100.00	43.91	72.81


Nutritional status across MPCE classes
8000
7000
6000
5000
4000
3000
2000
1000
0
10.38
14.01
16.49
19.46
22.41
25.91
30.86
38.14
48.24
63.2
85.26
118
170.32
342.81
Calorie Intake
MPCE (in Rs.)

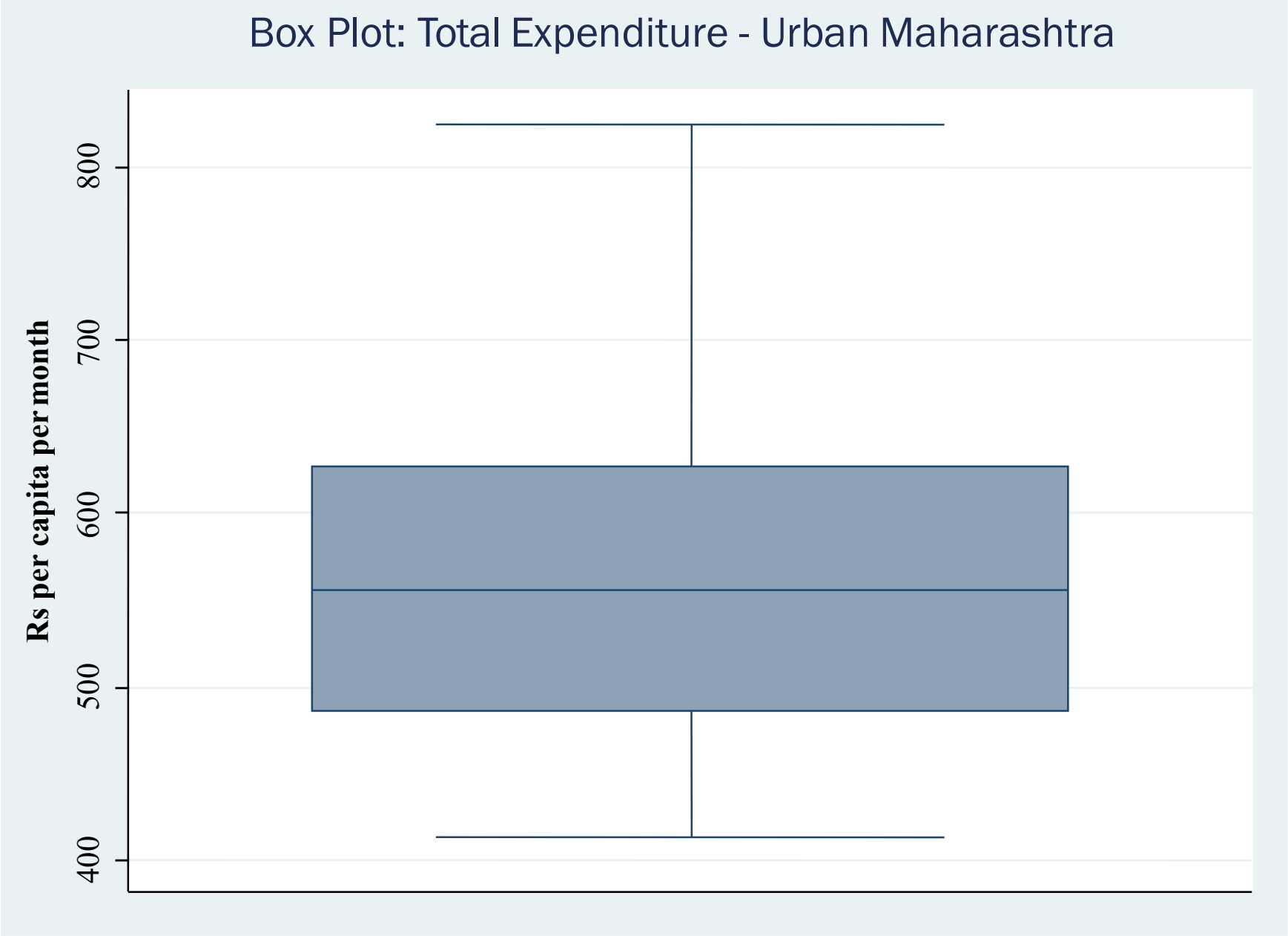

per capita calorie intake
0
2000
4000
6000
8000
0
100
200
300
400
MPCE

Basic data for Engel function : Urban Maharashtra (2004-05)
Source: GoM (pooled state & central samples)

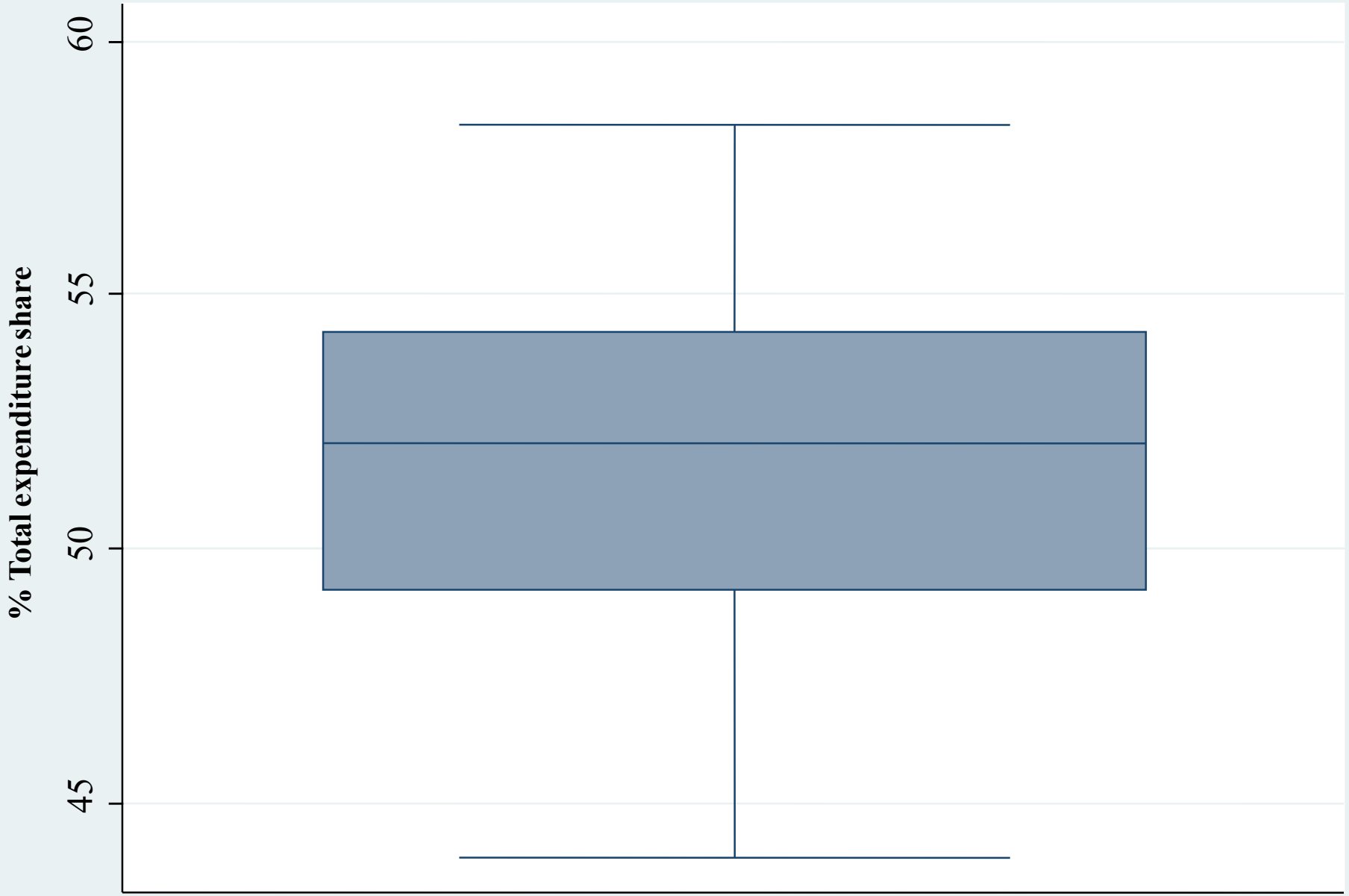
Average monthly per capita expenditure in Rs.				
District	Food	Non-Food	Total	% Share of food
Thane	519.50	787.60	1307.10	39.74
Mumbai	604.19	943.73	1547.91	39.03
Raigad	555.16	777.16	1332.32	41.67
Ratnagiri	483.68	615.09	1098.78	44.02
Sindhudurg	424.12	399.33	823.46	51.50
Konkan Division	572.32	881.47	1453.78	39.37
Pune	514.35	877.43	1391.78	36.96
Solapur	369.10	380.52	749.62	49.24
Satara	574.19	991.98	1566.17	36.66
Kolhapur	411.70	461.48	873.18	47.15
Sangli	363.48	322.46	685.94	52.99
Pune Division	469.98	710.77	1180.74	39.80
Ahmadnagar	400.08	516.52	916.60	43.65
Nandurbar	358.72	435.22	793.94	45.18
Dhule	360.29	393.28	753.57	47.81
Jalgaon	390.79	466.40	857.19	45.59
Nashik	342.12	526.41	868.53	39.39
Nashik Division	368.42	493.61	862.04	42.74

Average monthly per capita expenditure in Rs.				
District	Food	Non-Food	Total	% Share of food
Nanded	286.89	284.71	571.60	50.19
Hingoli	316.84	340.54	657.18	48.21
Parbhani	353.05	486.44	839.49	42.06
Jalna	428.61	618.58	1047.19	40.93
Aurangabad	352.10	520.40	872.50	40.36
Bid	265.25	199.53	464.78	57.07
Latur	332.43	381.34	713.77	46.57
Osmanabad	413.72	839.58	1253.30	33.01
Aurangabad Division	338.01	446.77	784.78	43.07
Buldhana	336.37	467.88	804.15	41.83
Akola	337.53	451.61	789.14	42.77
Washim	322.60	331.10	653.70	49.35
Amravati	311.91	385.64	697.55	44.72
Yavatmal	309.14	455.46	764.60	40.43
Amravati Division	322.68	422.45	745.13	43.31
Wardha	343.06	354.97	698.03	49.15
Nagpur	403.74	637.17	1040.91	38.79
Bhandara	373.63	612.69	986.32	37.88
Gondiya	418.34	638.79	1057.13	39.57
Gadchiroli	334.05	502.59	836.63	39.93
Chandrapur	419.57	780.25	1199.82	34.97
Nagpur Division	401.13	643.77	1044.90	38.39
State	479.80	720.78	1200.58	39.96

Box Whisker plot

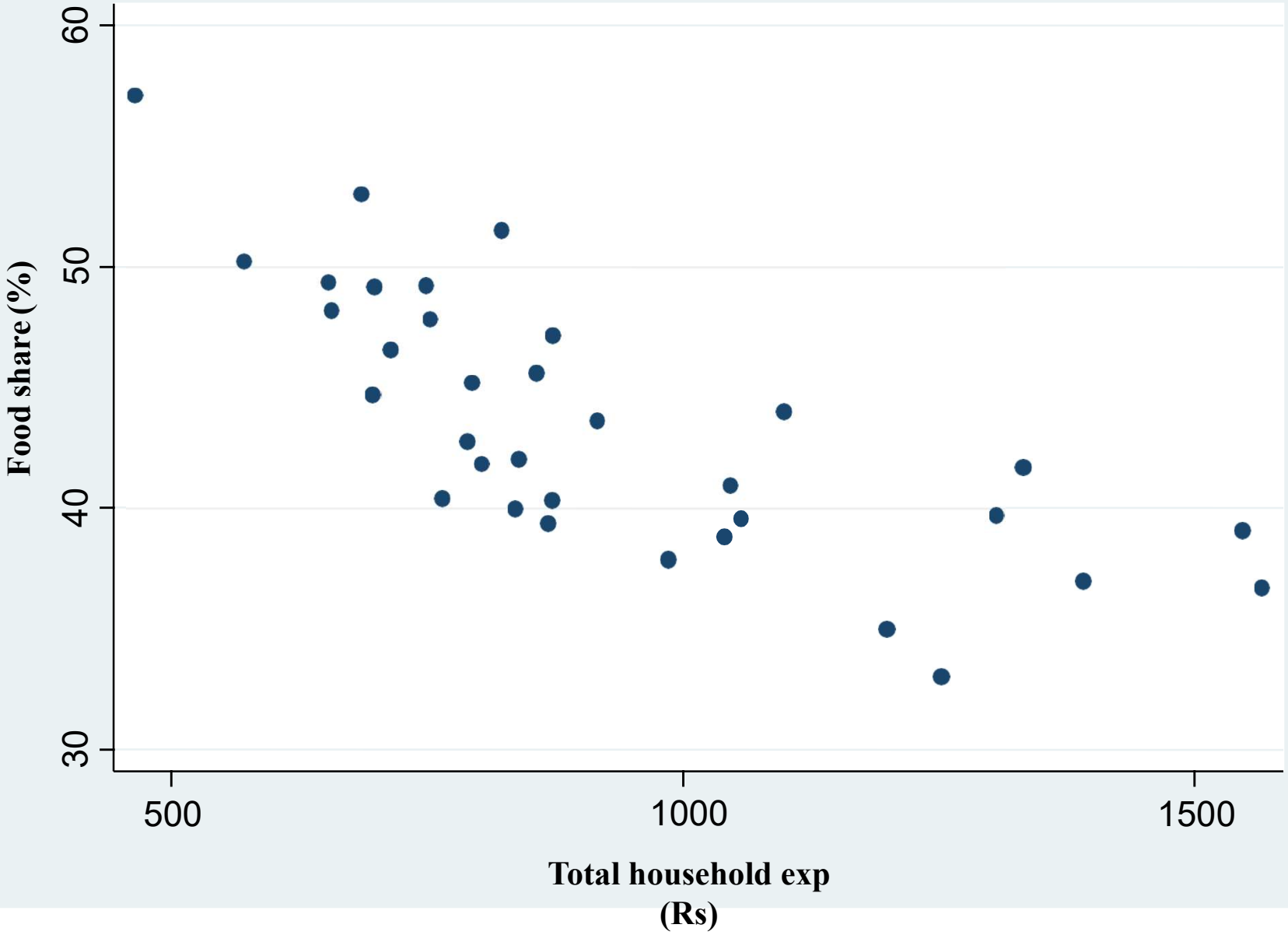


Box Plot: Food Share (%) in Total Expenditure - Urban Maharashtra

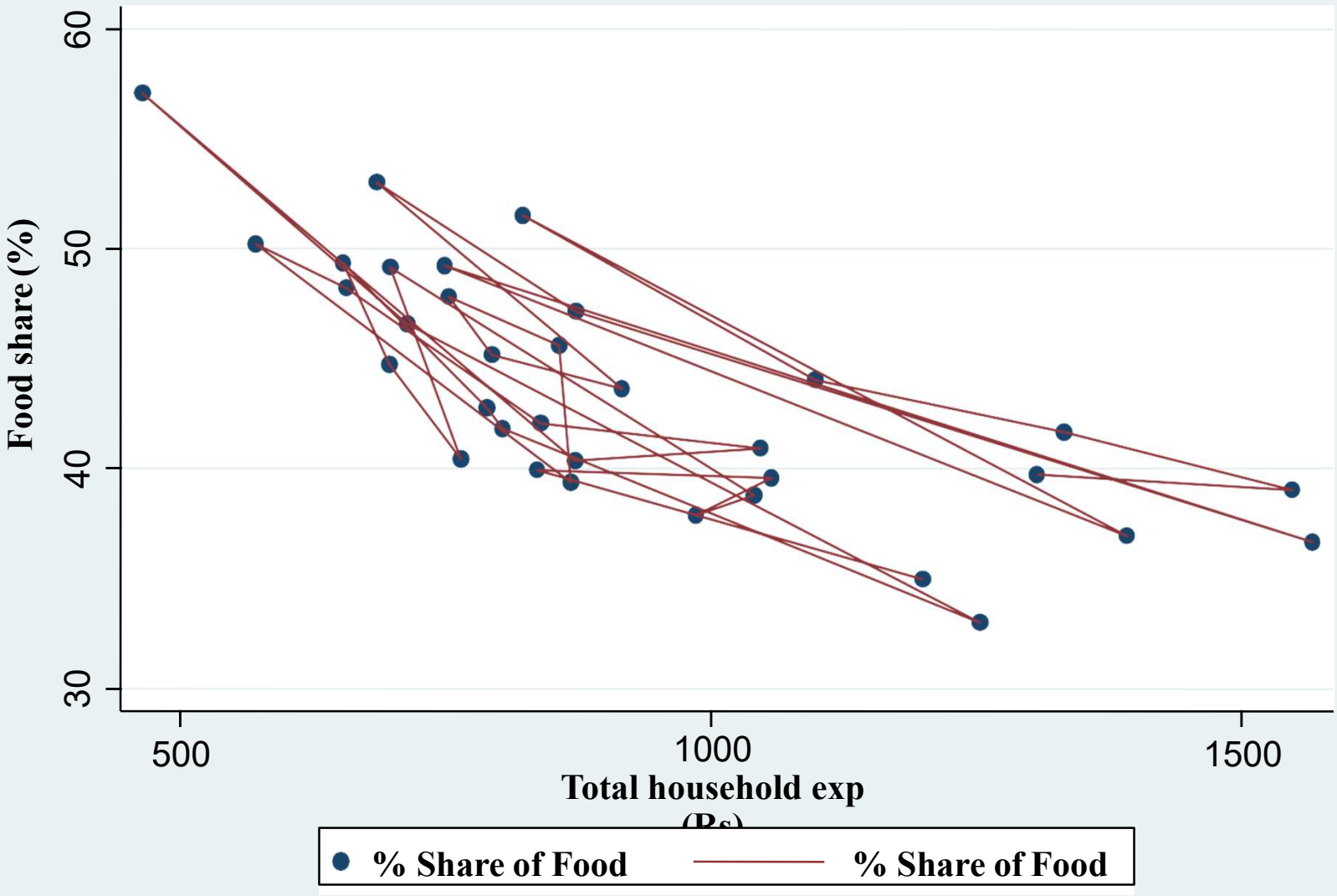


```
. regress Foodshare Total_exps

      Source |       SS       df       MS              Number of obs =      34
-------------+------------------------------           F(  1,    32) =   42.14
       Model |  565.222061     1  565.222061           Prob > F      =  0.0000
    Residual |  429.260208    32  13.4143815           R-squared     =  0.5684
-------------+------------------------------           Adj R-squared =  0.5549
       Total |  994.482269    33  30.1358263           Root MSE      =  3.6626

------------------------------------------------------------------------------
  Foodshare |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
------------+-----------------------------------------------------------------
  Total_exp |  -.0151772   .0023381    -6.49   0.000    -.0199399   -.0104146
      _cons |   57.54894   2.256355    25.51   0.000     52.95289    62.14498
------------------------------------------------------------------------------
```

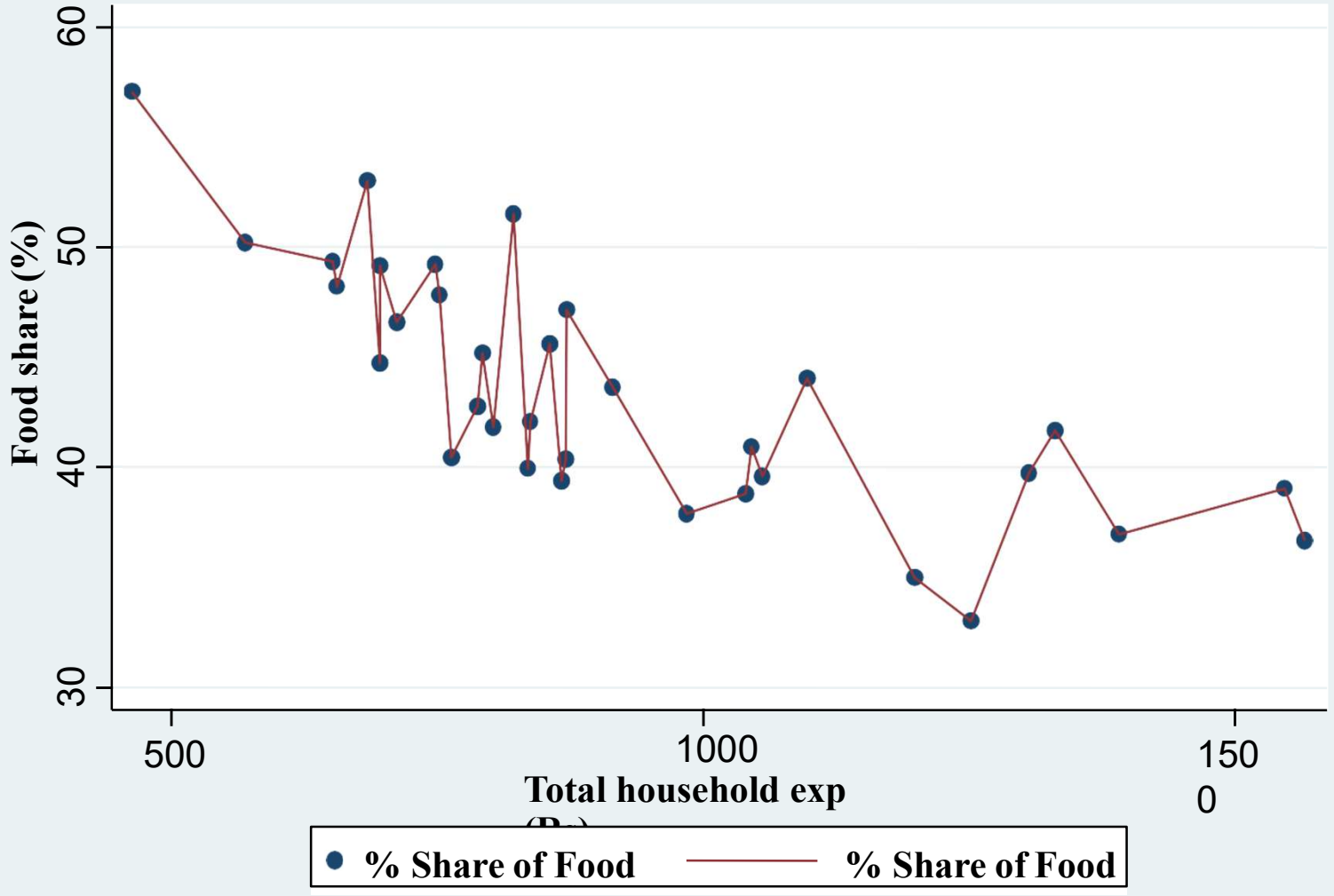
Engel Relation for Food: Urban Maharashtra (2004-05)



Engel Relation for Food: Urban Maharashtra (2004-05)

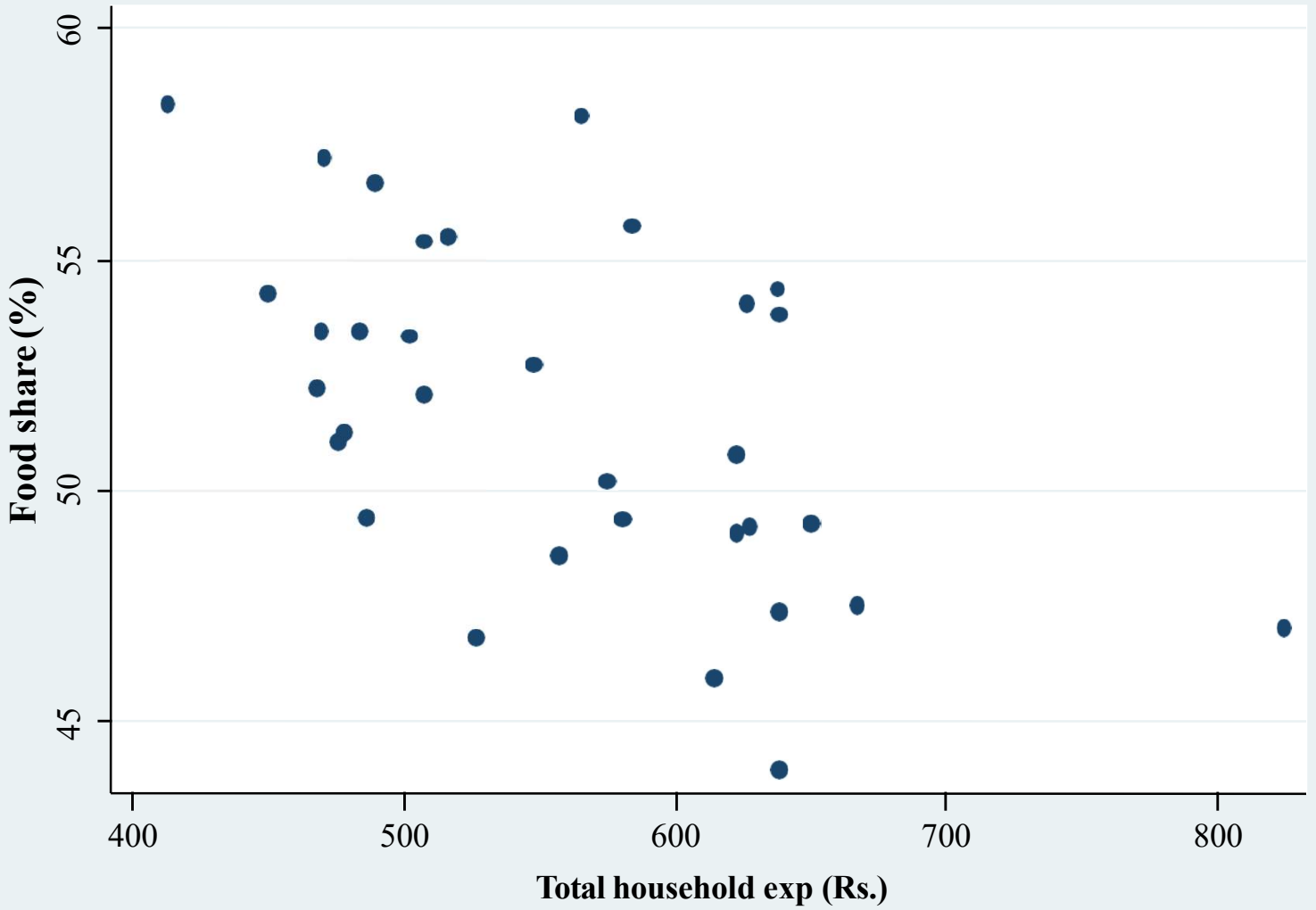


Engel Relation for Food: Urban Maharashtra (2004-05)

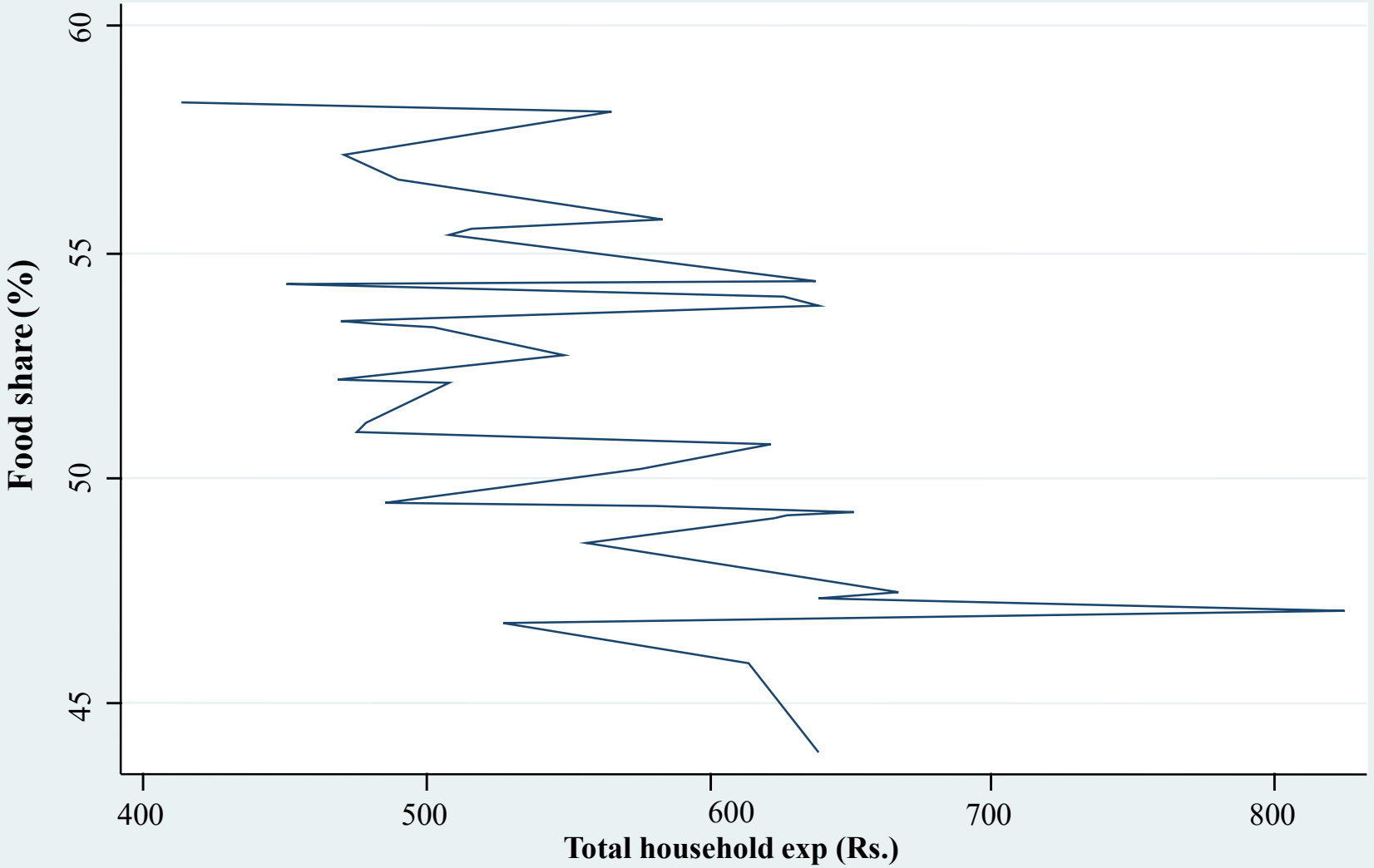


```
. regress shareoffood total

      Source |       SS       df       MS              Number of obs =      33
-------------+------------------------------           F(  1,    31) =   12.56
       Model |  126.172681     1  126.172681           Prob > F      =  0.0013
    Residual |  311.400862    31  10.0451891           R-squared     =  0.2883
-------------+------------------------------           Adj R-squared =  0.2654
       Total |  437.573542    32  13.6741732           Root MSE      =  3.1692

------------------------------------------------------------------------------
 shareoffood |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       total |  -.0231554   .0065334    -3.54   0.001    -.0364799   -.009831
       _cons |   64.68854   3.697611    17.49   0.000     57.14721    72.22987
------------------------------------------------------------------------------
```

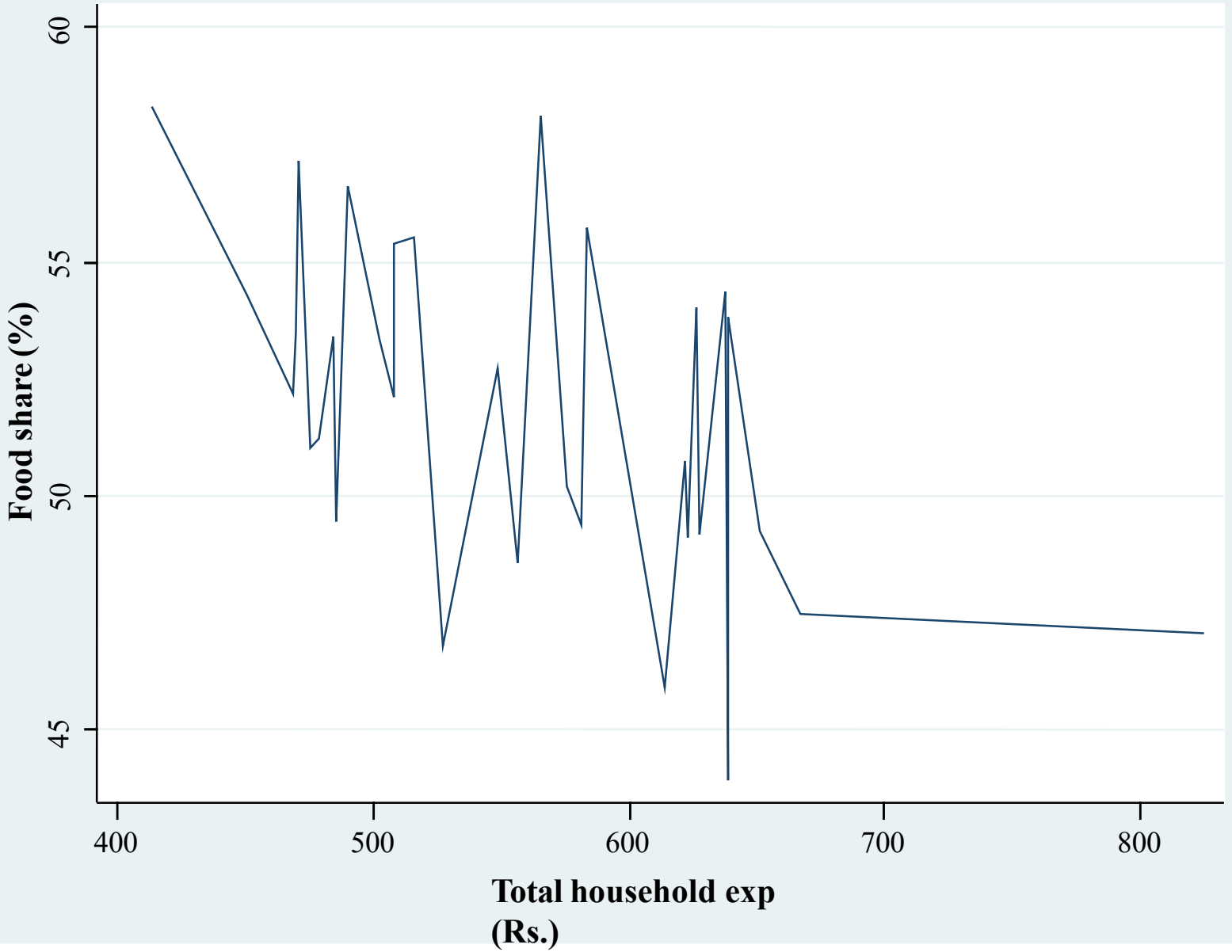
Engel Relation for Food: Rural Maharashtra (2004-05)



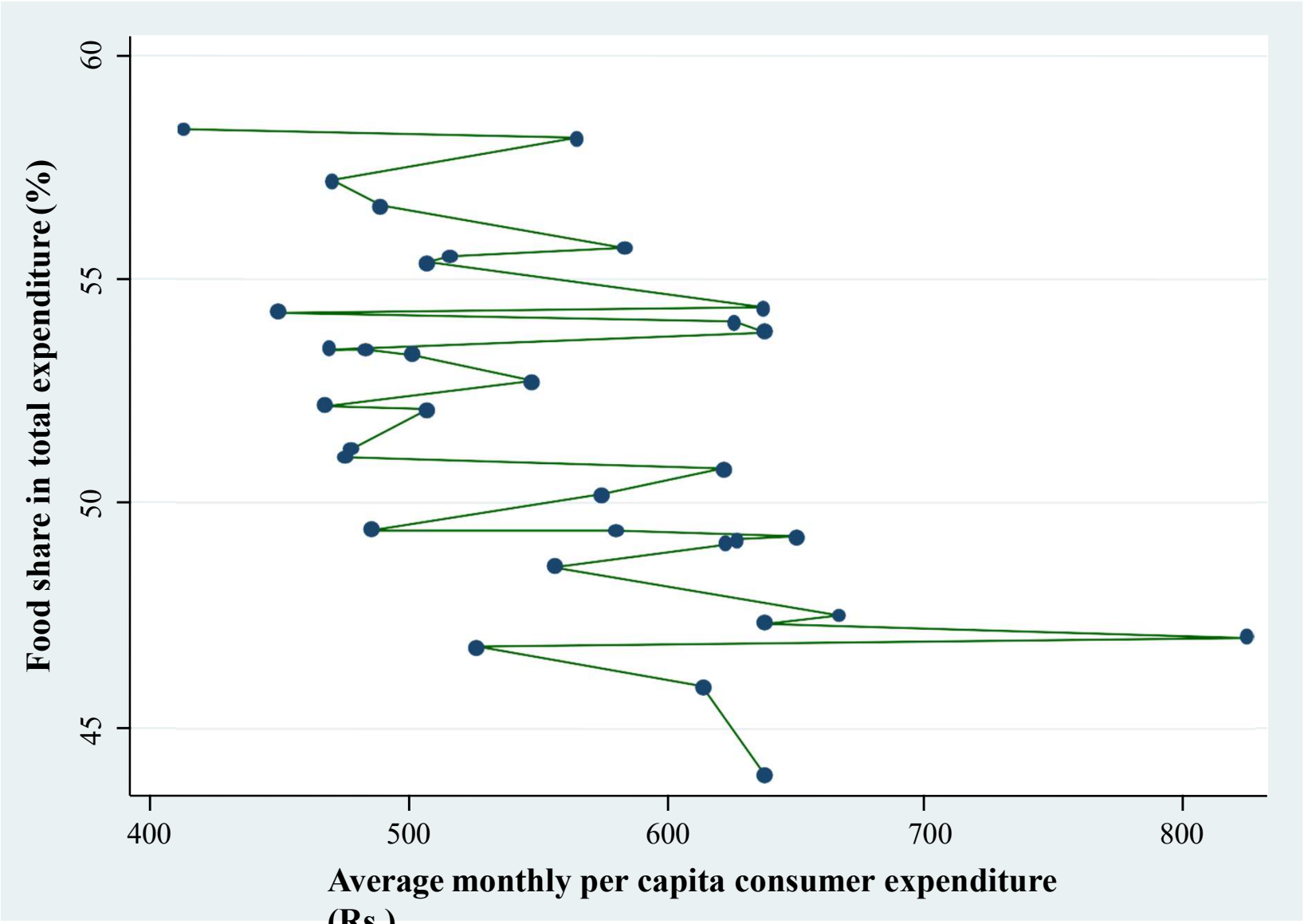
Engel Relation for Food: Rural Maharashtra (2004-05)



Engel Relation for Food: Rural Maharashtra (2004-05)



Engel relation: Rural Maharashtra



Box Whisker Plot: 5 Number Summary

- **Five number summary : Visual representation of the box and whisker plot.**

- **The five number summary consists of :**
  - **The median (2nd quartile)**
  - **The 1st quartile**
  - **The 3rd quartile**
  - **The maximum value in a data set**
  - **The minimum value in a data set**

Box and whisker plot: Steps

- **Step 1 – Estimate the median.**
- **Median: Central value in ordered data set.**

18, 27, 34, 52, 54, 59, 61, 68, 78, 82, 85, 87, 91, 93, 100

68 is the median of this data set.

Constructing a box and whisker plot

- **Step 2 – Estimate the lower quartile.**
- **Lower quartile: Median of the bottom half - data set to the left of 68.**

(18, 27, 34, 52, 54, 59, 61,) 68, 78, 82, 85, 87, 91, 93, 100

52 is the lower quartile

Constructing a box and whisker plot

- **Step 3 – Estimate the upper quartile.**
- **Upper quartile: Median of the top half - data set to the right of 68.**

18, 27, 34, 52, 54, 59, 61, 68, (78, 82, 85, 87, 91, 93, 100)

87 is the upper quartile

Constructing a box and whisker plot

- **Step 4 – Estimate the maximum and minimum values in the set.**
- **The maximum is the largest value in the data set.**
- **The minimum is the smallest value in the data set.**

18, 27, 34, 52, 54, 59, 61, 68, 78, 82, 85, 87, 91, 93, 100

18 is the minimum and 100 is the maximum.

Constructing a box and whisker plot

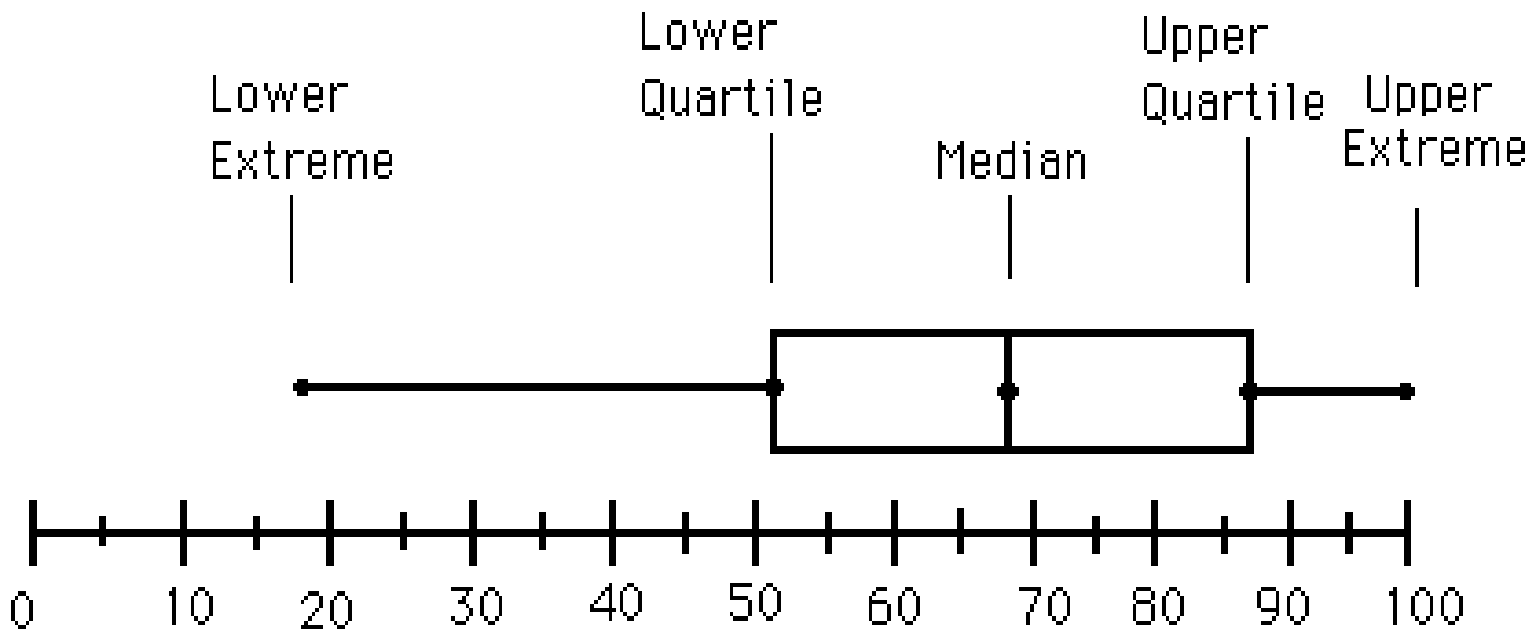
- **Step 5 – Estimate the inter-quartile range (IQR).**
- **IQR: Difference between the upper and lower quartiles.**
  - **Upper Quartile = 87**
  - **Lower Quartile = 52**
  - **87 – 52 = 35**
  - **35 = IQR**

The 5 Number Summary

- **Organize the 5 number summary**
  - **–Median – 68**
  - **–Lower Quartile – 52**
  - **–Upper Quartile – 87**
  - **–Max – 100**
  - **–Min – 18**

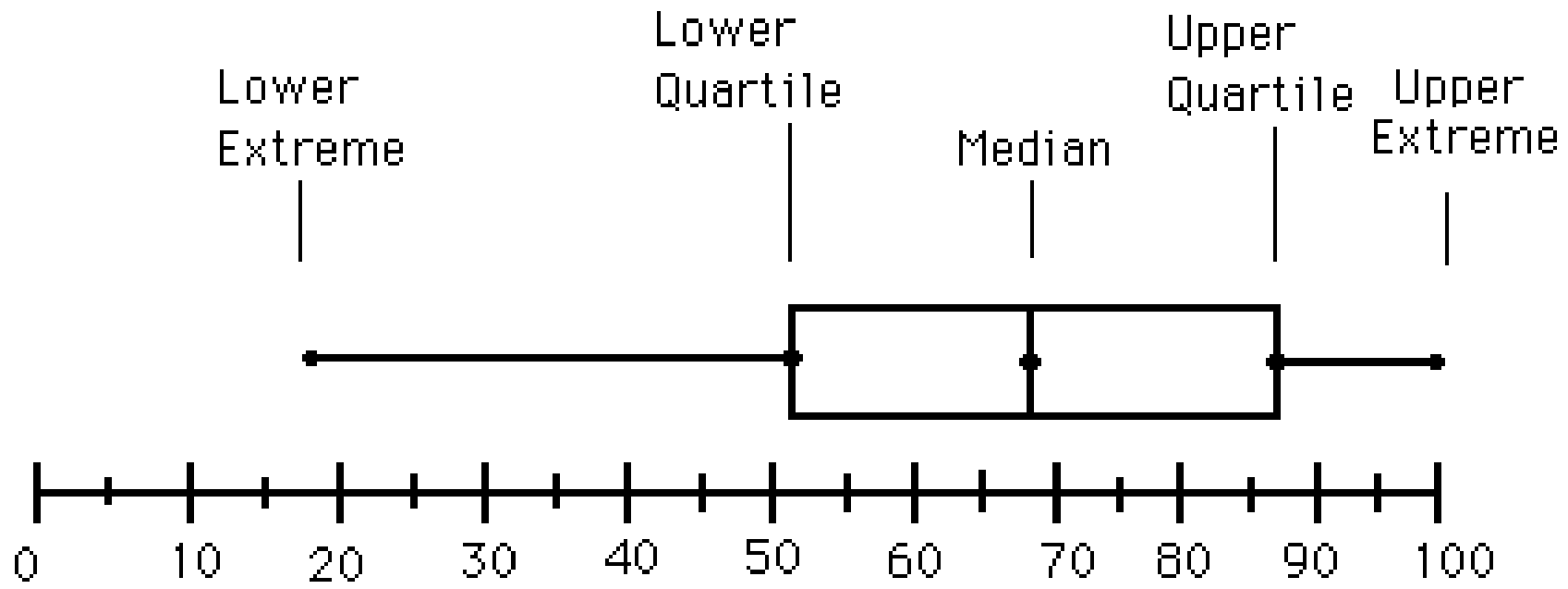
Graphing The Data

- **The Box includes the lower quartile, median, and upper quartile.**
- **The Whiskers extend from the Box to the max and min.**



Analyzing The Graph Slide 18

- **Observations inside the box represent the middle half (50%) of the data.**
- **The line segment inside the box represents the median.**



Research Hypothesis

- Hypothesis is considered as an intelligent guess or prediction, that gives directional to the researcher to answer the research question.

- Hypothesis or Hypotheses are defined as the formal statement of the tentative or expected prediction or explanation of the relationship between two or more variables in a specified population.

- A hypothesis is a formal tentative statement of the expected relationship between two or more variables under study.

- A hypothesis helps to translate the research problem and objective into a clear explanation or prediction of the expected results or outcomes of the study.

- Hypothesis is derived from the research problems, literature review and conceptual framework.

- Hypothesis in a research project logically follow literature review and conceptual framework.

IMPORTANCE OF HYPOTHESIS IN RESEARCH

- Hypotheses enables the researcher to objectively investigate new areas of discovery. Thus , it provides a powerful tool for the advancement of knowledge.
- Hypotheses provides objectivity to the research activity.
- It also provides directions to conduct research such as defining the sources & relevance of data.
- Hypotheses provides clear & specific goals to the researchers. These clear & specific goals provide the investigator with a basis for selecting sample & research procedures to meet these goals.

Count..

- Hypotheses provides link between theories & actual practical research.
- It provides a bridge between theory & reality.
- A hypothesis suggests which type of research is likely to be most appropriate.
- As it is a tentative statement of anticipated results, it guides the researcher towards the direction in which the research should proceed.
- It stimulates the thinking process of researcher as the researcher forms the hypothesis by anticipating the outcome.

Co
unt
...

- It also determines the most appropriate research designs & techniques of data analysis.
- Hypotheses provides understanding to the researchers about what expect from the results of the research study.
- It serves as framework for drawing conclusions of a research study.
- Without hypotheses, research would be like aimless wandering.

CHARACERISTICS OF A GOOD HYPOTHESIS

- **Conceptual clarity:**

Hypothesis should consist of clearly defined & understandable concepts. It should be stated in very terms, the meaning & implication of which cannot be doubted. To facilitate the conceptual clarity, hypothesis can be stated in declarative statement, in present tense.

- **Empirical referents:**

Research must have an ultimate empirical referent. No usable hypothesis can embody moral judgments. A good hypothesis must have empirical basis from the area of enquiry.

Co
unt
...

- **Objectivity:**

Hypothesis must be objective, which facilitates objectivity in data collection & keeps the research activity free from researcher value - judgment.

- **Specificity:**

It should be specific, not general, & should explain the expected relations between variables. For example, regular yoga reduces stress.

Count...

- **Relevant:**

The hypothesis should be relevant to the problem being studied as well as the objectives of the study. Hypothesis must have relevance with theory under test in a research process.

- **Testability:**

Hypothesis should be testable & should not be a moral judgment. It must be directly/indirectly observable & measurable. The researcher can set up a situation that permits one to assess if it is true or false. It must be verifiable. For example, a statement such as 'bad partners produce bad children'. This sort of hypothesis cannot be tested.

Co
unt

- **Consistency:**

A hypothesis should be consistent with an existing body of theories, research findings, & other hypotheses. It should correspond with existing knowledge. ...

- **Simplicity:**

A hypothesis should be formulated in simple & understandable terms. It should require fewer conditions & assumptions.

Co
unt

- **Availability of techniques:**

The researchers must make sure that methods are available for testing their proposed hypotheses

- **Purposiveness:**

The researcher must formulate only purposeful hypotheses, which has relevance with research problem & objectives.

- **Verifiability:**

A good hypothesis can be actually verified in practical terms.

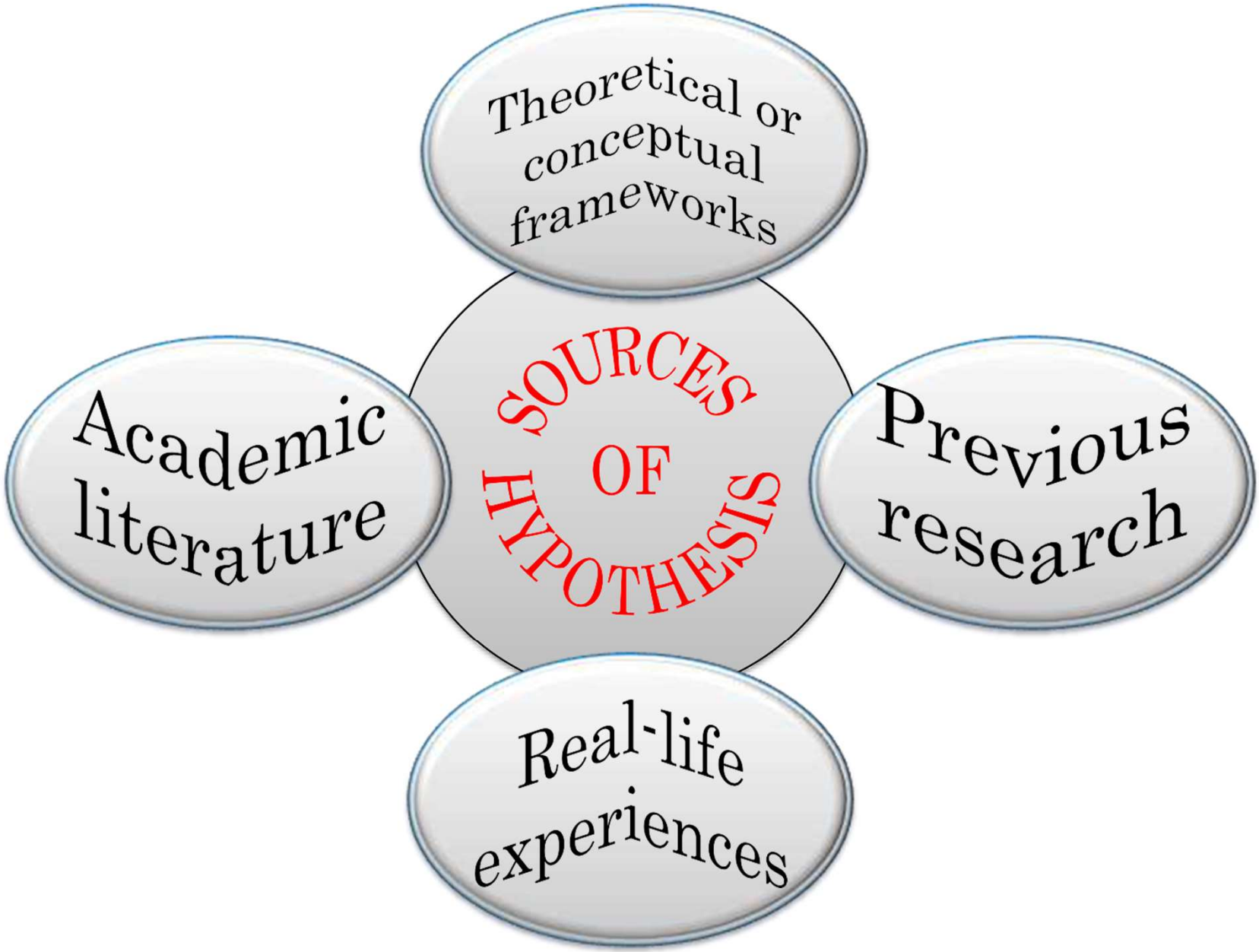
Co
unt

- **Profundity of effect:**

A good hypothesis should have profound effect.. upon a variety of research variables.

- **Economical:**

The expenditure of money & the time can be controlled if the hypotheses underlying the research undertaken is good.


Theoretical or conceptual frameworks
Academic literature
SOURCES OF HYPOTHESIS
Previous research
Real-life experiences

Theoretical or conceptual frameworks:

- The most important sources of hypotheses are theoretical or conceptual frameworks developed for the study.
- Through a deductive approach these hypotheses are drawn from theoretical or conceptual frameworks for testing them.
- For example, Roy's adaptation Model is used in a research study, where a hypothesis can be drawn from a concept of the theoretical mode that 'patient's adaptation to a chronic illness depends on availability of social support for them.'

Previous research:

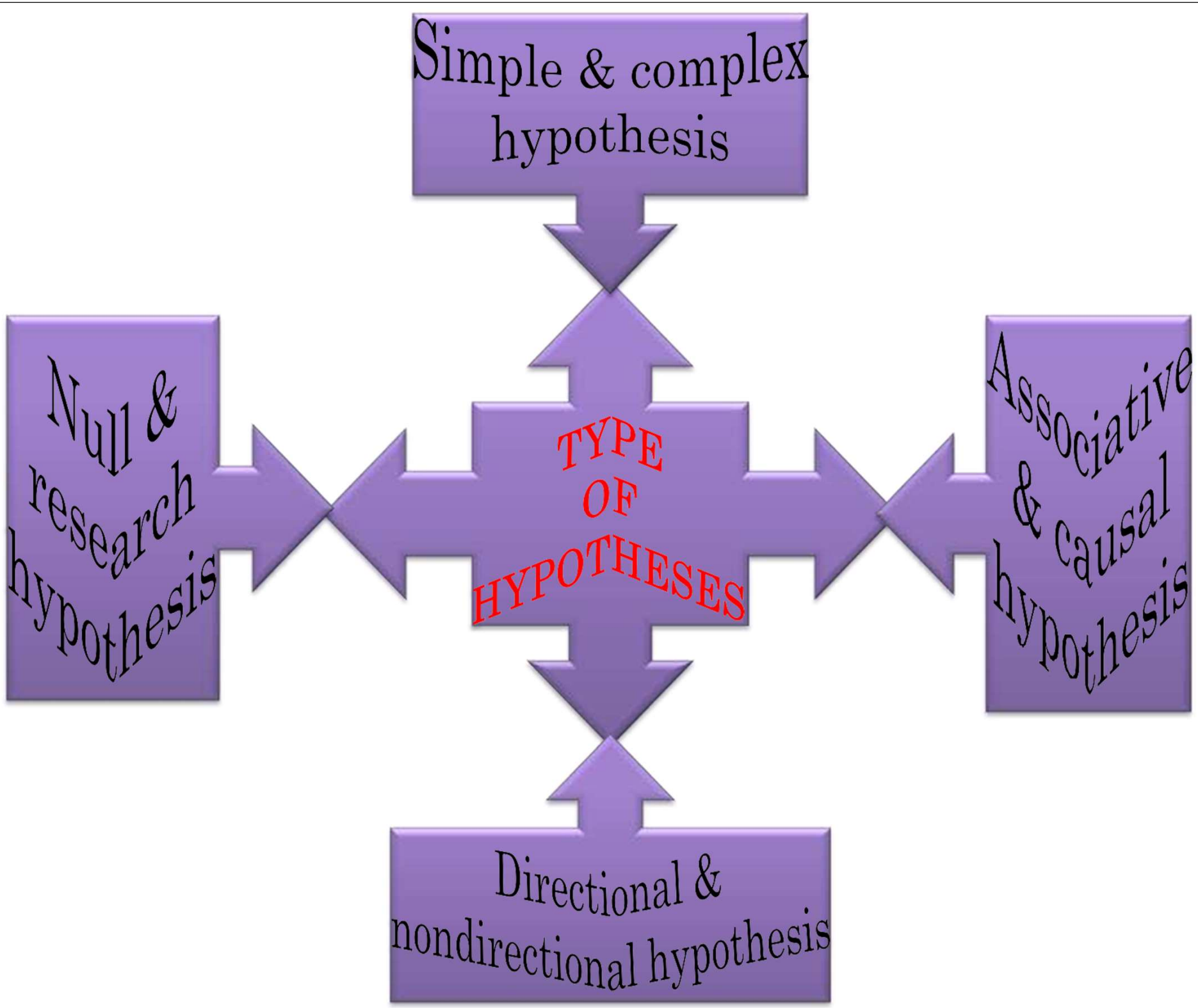
- Findings of the previous studies may be used for framing the hypotheses for another study.
- For example, in a small sample descriptive study, a researcher found that a number of patients admitted with coronary artery disease had increased body mass index.
- In another research study, a researcher may use this finding to formulate a hypothesis as 'Obese patients have increased risk for development of coronary artery disease'.

Real-life experiences:

- Real-life experiences also contribute in the formulation of hypotheses for research studies.
- For example, Newton had a life-changing experience of the falling of an apple & formulated a hypothesis that earth attracts all the mass towards its centre, through several researchers were conducted before generating a law of central gravity.

Academic literature

- Academic literature is based on formal theories, empirical evidences, experiences, observation, & conceptualizations of academicians.
- These literatures may serve as good sources for formulating hypotheses for research studies.


Simple & complex hypothesis
Null & research hypothesis
TYPE OF HYPOTHESES
Associative & causal hypothesis
Directional & nondirectional hypothesis

Simple & complex hypothesis

- **Simple hypothesis:**

✓ It is a statement which reflects the relationship between two variables.

✓ *For example*, 'the lower the level of hemoglobin, the higher is the risk of infection among postpartum women'.

- **Complex hypothesis:**

✓ It is a statement which reflects the relationship between more than two variables.

✓ *For example*, 'satisfaction is higher among patients who are older & dwelling in rural area than those who are younger & dwelling in urban area'.

Associative & causal hypothesis

- **Associative hypothesis:**

✓ It reflects a relationship between variables that occurs or exists in natural settings without manipulation.

✓ This hypothesis is used in correlational research studies

Examples of associative hypothesis	prediction
Communication skills of health care providers & cost of care related to the satisfaction of patients	Predictsrelationship among variables but not the type of relationship

Count…

- **Causal hypothesis:**

✓ It predicts the cause-and-effect relationship between two or more dependent & independent variables in experimental or interventional setting, where independent variable is manipulated by research to examine the effect on the dependent variable.

✓ The causal hypothesis reflects the measurement of dependent variable to examine the effect of dependent variable, which is manipulated by the researcher(s).

✓ *For examples,* prevalence of pin site infection is lower in patients who receive pin site care with hydrogen proxidide as compared to patients who receive the pin site care with Betadine solution.

Directional & nondirectional hypothesis

- **Directional hypothesis:**

✓ It specifies not only the existence, but also the expected direction of the relationship between variables.

✓ Directional hypothesis states the nature of the relationship between two or more variables such as positive, negative, or no relationship.

✓ To express the direction of relationship between variables, the directional terms are used to state the hypothesis such as

positive, negative, less, more, increased, decreased, great er, higher, lower, etc.

✓ *For examples,* 'there is a positive relationship between years of nursing experience & job satisfaction among nurses.

Count…

- **Nondirectional Hypothesis:**

- ✓ It reflects the relationship between two or more variables, but is does not specify the anticipated direction & nature of relationship such as positive or negative.
- ✓ It indicates the existence of relationship between the variables.
- ✓ *For example,* 'there is relationship between years of nursing experience & job satisfaction among nurses.

Null & research hypothesis:

- **Null hypothesis (H_0):**

✓ It is also known as statistical hypothesis & is used for statistical testing & interpretation of statistical outcomes.

✓ It states the existence of no relationship between the independent & dependent variables.

✓ *For example,* 'there is no relationship between smoking & the incidence of coronary artery disease'.

- **Research hypothesis (H_1):**

✓ It states the existence of relationship between two or more variables.

✓ *For examples,* 'there is relationship between smoking & incidence of lung cancer.

Type I Error α error

- **In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H0 is wrongly rejected.**

Type I Error α error

- **For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H0: there is no difference between the two drugs on average. A type I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.**

Type I Error α error

- **The following table gives a summary of possible results of any hypothesis test:**

Type I Error α error

		Decision	
		Reject H_0	**Don't reject H_0**
Truth	**H_0**	**Type I Error**	**Right Decision**
	H_a	**Right Decision**	**Type II Error**

Type I Error α error

- **A type I error is often considered to be more serious, and therefore more important to avoid, than a type II error. The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly;**

Type I Error α error

- **this probability is never 0. This probability of a type I error can be precisely computed as,**
- **P(type I error) = significance level = The exact probability of a type II error is generally unknown.**

Type I Error α error

- **If we do not reject the null hypothesis, it may still be false (a type II error) as the sample may not be big enough to identify the falseness of the null hypothesis (especially if the truth is very close to hypothesis).**

Type I Error α error

- **For any given set of data, type I and type II errors are inversely related; the smaller the risk of one, the higher the risk of the other.**
- **A type I error can also be referred to as an error of the first kind**

Type II Error β error

- **In a hypothesis test, a type II error occurs when the null hypothesis H0, is not rejected when it is in fact false.**

Type I Error α error

- **For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H0: there is no difference between the two drugs on average.**

Type I Error α error

- **A type II error would occur if it was concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different ones.**

Type I Error α error

- **A type II error is frequently due to sample sizes being too small.**
- **The probability of a type II error is symbolised by and written:**
- **P(type II error) = (but is generally unknown).**
- **A type II error can also be referred to as an error of the second kind**

Significance Level

- **The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H0, if it is in fact true.**

Significance Level

- **It is the probability of a type I error and is set by the investigator in relation to the consequences of such an error. That is, we want to make the significance level as small as possible in order to protect the null hypothesis and to prevent, as far as possible, the investigator from inadvertently making false claims.**

Significance Level

- **The significance level is usually denoted by Significance Level = P(type I error) = Usually, the significance level is chosen to be = 0.05 = 5%.**

Significance Level

- **One-sided Test**
- **A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H0 are located entirely in one tail of the probability distribution.**

Significance Level

- **In other words, the critical region for a one-sided test is the set of values less than the critical value of the test, or the set of values greater than the critical value of the test.**
- **A one-sided test is also referred to as a one-tailed test of significance**

Significance Level

- **The choice between a one-sided and a two-sided test is determined by the purpose of the investigation or prior reasons for using a one-sided test**

Significance Level

- **Example**
- **Suppose we wanted to test a manufacturers claim that there are, on average, 50 matches in a box. We could set up the following hypotheses**
- Ho:**μ** =50 as against
- Ha: **μ** <50 or
- Ha:**μ** >50

Significance Level

- **Either of these two alternative hypotheses would lead to a one-sided test. Presumably, we would want to test the null hypothesis against the first alternative hypothesis since it would be useful to know if there is likely to be less than 50 matches, on average, in a box (no one would complain if they get the correct number of matches in a box or more).**

Significance Level

- **Yet another alternative hypothesis could be tested against the same null, leading this time to a two-sided test:**
- Ho:**μ** =50 as against
- Ha: **μ ≠** 50

Significance Level

- **That is, nothing specific can be said about the average number of matches in a box; only that, if we could reject the null hypothesis in our test, we would know that the average number of matches in a box is likely to be less than or greater than 50.**

Two-Sided Test

- **A two-sided test is a statistical hypothesis test in which the values for which** we can reject the **null hypothesis, H0 are located in both tails of the probability distribution.**

Two-Sided Test

- **In other words, the critical region for a two-sided test is the set of values less than a first critical value of the test and the set of values greater than a second critical value of the test**
- **A two-sided test is also referred to as a two-tailed test of significance.**

Two-Sided Test

- **The choice between a one-sided test and a two-sided test is determined by the purpose of the investigation or prior reasons for using a one-sided test.**

Level of significance and confidence

- Significance means the percentage risk to reject a null hypothesis when it is true and it is denoted by α. Generally taken as 1%, 5%, 10%
- $(1 - \alpha)$ is the confidence interval in which the null hypothesis will exist when it is true.

Designation	Risk α	Confidence $1-\alpha$	Description
Supercritical	0.001 0.1%	0.999 99.9%	More than $100 million (Large loss of life, e.g. nuclear disaster
Critical	0.01 1%	0.99 99%	Less than $100 million (A few lives lost)
Important	0.05 5%	0.95 95%	Less than $100 thousand (No lives lost, injuries occur)
Moderate	0.10 10%	0.90 90%	Less than $500 (No injuries occur)

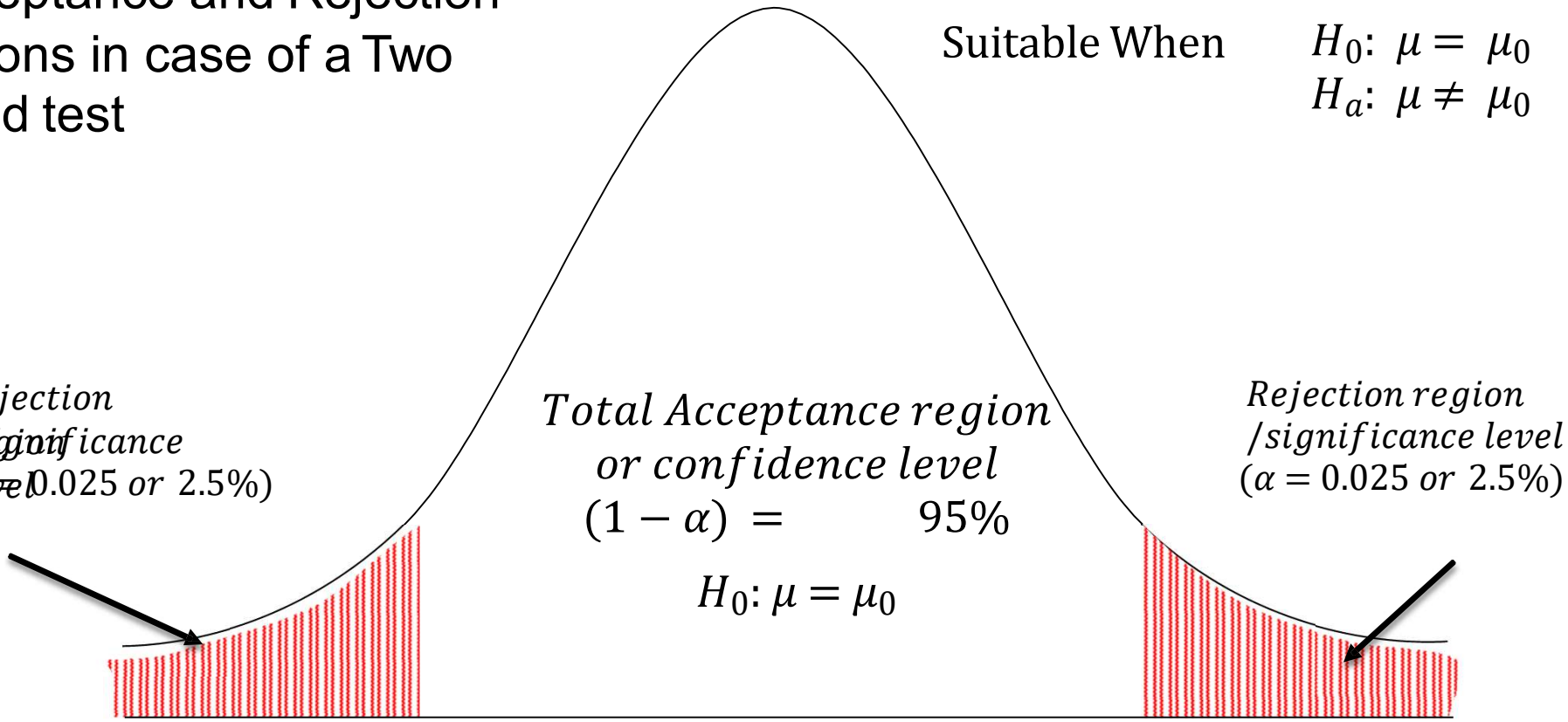
Type I and Type II Error

Situation	Decision	
	Accept Null	Reject Null
Null is true	Correct	Type I error (α *error*)
Null is false	Type II error (β *error*)	Correct

Two tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Two tailed test

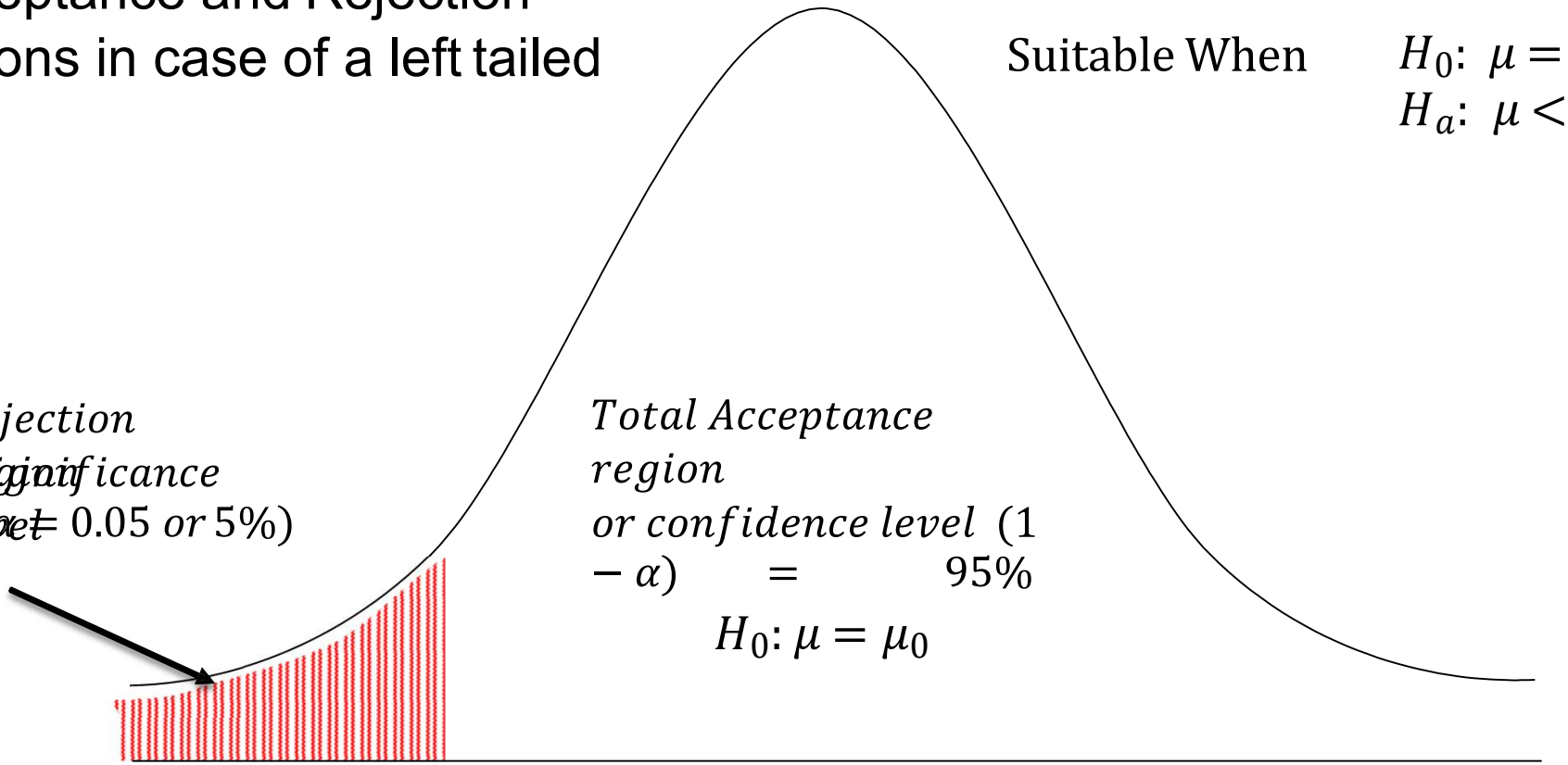
Suitable When $H_0: \mu = \mu_0$
$H_a: \mu \neq \mu_0$



Left tailed test @ 5% Significance level

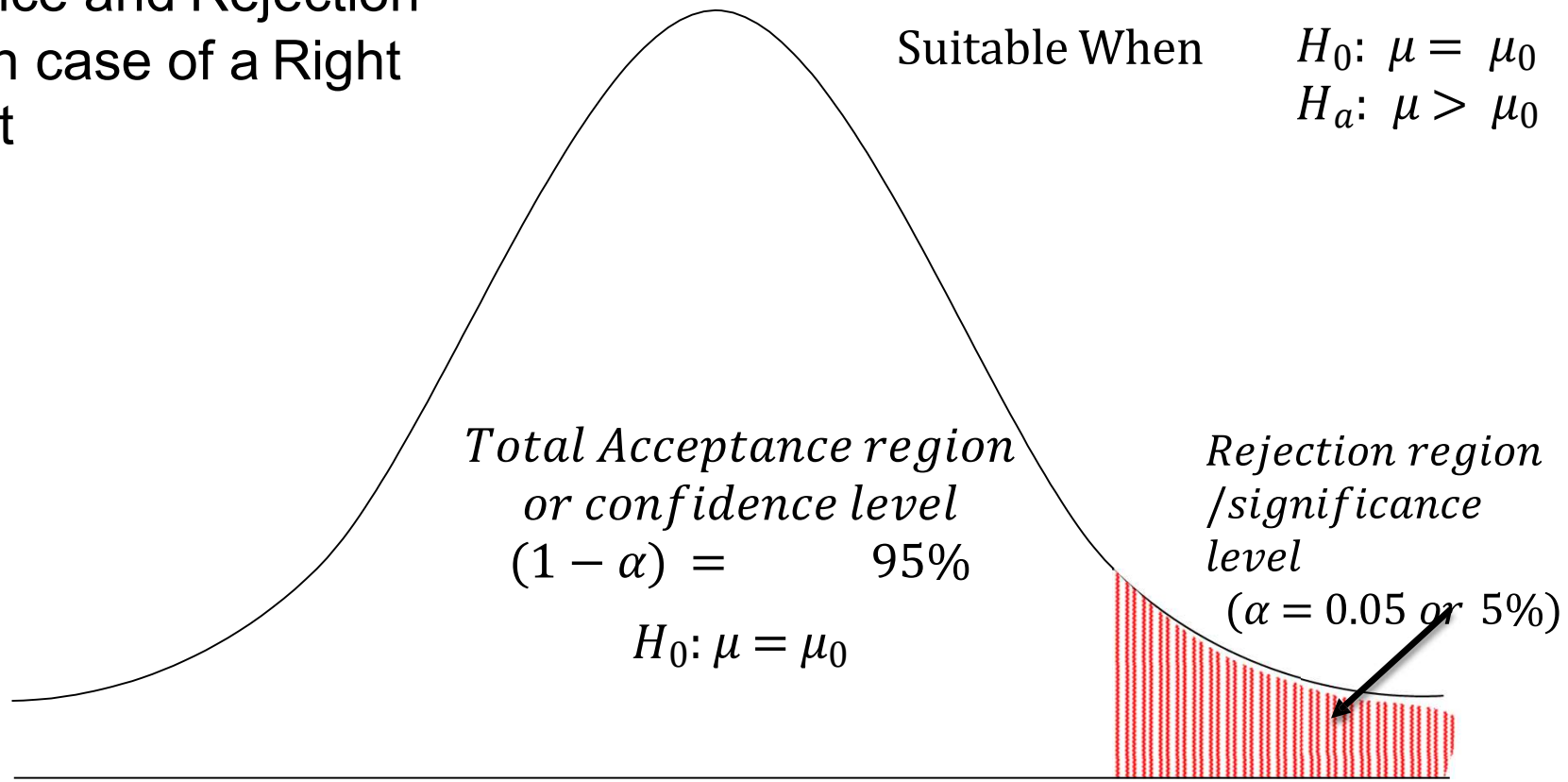
Acceptance and Rejection regions in case of a left tailed test

Suitable When $H_0: \mu = \mu_0$
$H_a: \mu < \mu_0$



Right tailed test @ 5% Significance level

Acceptance and Rejection regions in case of a Right tailed test



14

Z-TEST AND T-TEST

Z-Test for testing means

Test Condition

- Population normal and infinite
- Sample size large or small,
- Population variance is known
- Ha may be one-sided or two sided

Test Statistics

$$Z = \frac{X - \mu_{H_0}}{\sigma_p}$$

Z-Test for testing means

Test Condition

- Population normal and finite,
- Sample size large or small,
- Population variance is known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{X - \mu_{H_0}}{\frac{\sigma_p}{\sqrt{n}} \times \left[\sqrt{(N-n)N(-1\quad)}\right]}$$

Z-Test for testing means

Test Condition

- Population is infinite and may not be normal,
- Sample size is large,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$Z = \frac{X - \mu_{H_0}}{\frac{\sigma_s}{\sqrt{}}}$$

Z-Test for testing means

Test Condition

- Population is finite and may not be normal,
- Sample size is large,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{X - \mu_{H_0}}{\frac{\sigma_s}{\sqrt{n}} \times \left[\sqrt{(N-n)N(-1\quad)}\right]}$$

T-Test for testing means

Test Condition

- Population is infinite and normal,
- Sample size is small,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{X - \mu_{H_0}}{\frac{\sigma_s}{\sqrt{\ }}}$$

$$with\ d.f. = n - 1$$

$$\sigma_s = \sqrt{\frac{(X_i - X\ )^2}{(n-1)}}$$

T-Test for testing means

Test Condition

- Population is finite and normal,
- Sample size is small,
- Population variance is unknown
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{X - \mu_{H_0}}{\frac{\sigma_s}{\sqrt{n}} \times \left[\sqrt{(N-n)N(-1\quad)}\right]}$$

$$with\ d.f. = n - 1$$

$$\sigma_s = \sqrt{\frac{(X_i - X\ )^2}{(n-1)}}$$

Hypothesis testing for difference between means

Z-TEST, T-TEST

Z-Test for testing difference between means

Test Condition

- Populations are normal
- Samples happen to be large,
- Population variances are known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{X_1 - X_2}{\sqrt{\frac{\sigma_{p1}^2}{n_1} + \frac{\sigma_{p2}^2}{n_2}}}$$

Z-Test for testing difference between means

Test Condition

- Populations are normal
- Samples happen to be large,
- Presumed to have been drawn from the same population
- Population variances are known
- Ha may be one-sided or two sided

Test Statistics

$$z = \frac{X_1 - X}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

T-Test for testing difference between means

Test Condition

- Samples happen to be small,
- Presumed to have been drawn from the same population
- Population variances are unknown but assumed to be equal
- Ha may be one-sided or two sided

Test Statistics

$$t = \frac{X_1 - X_2}{\sqrt{\frac{(n_1 - 1)\sigma_{s1}^2 + (n_2 - 1)\sigma_{s2}^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$with\ d.f. = (n_1 + n_2 - 2)$$

Hypothesis Testing for comparing two related samples

PAIRED T-TEST

Paired T-Test for comparing two related samples

Test Condition

- Samples happens to be small
- Variances of the two populations need not be equal
- Populations are normal
- Ha may be one sided or two sided

Test Statistics

$$t = \frac{D - 0}{\frac{\sigma_{diff.}}{\sqrt{n}}}$$

$with\ (n-1)\ d.f.$

D = Mean of differences

$\sigma_{diff.}$ = Standard deviation of differences

$n = Number\ of\ matched\ pairs$

Hypothesis Testing of proportions

Z-TEST

Z-test for testing of proportions

Test Condition

- Use in case of qualitative data
- Sampling distribution may take the form of binomial probability distribution
- Ha may be one sided or two sided
- $Mean = n.p$
- $Standard\ deviation = \sqrt{n.p.q}$

Test statistics

$$z = \frac{p -}{p\sqrt{\frac{p.q}{n}}}$$

$p = proportion\ of\ \ sucess$

Hypothesis Testing for difference between proportions

Z-TEST

Z-test for testing difference between proportions

Test Condition

- Sample drawn from two different populations
- Test confirm, whether the difference between the proportion of success is significant
- Ha may be one sided or two sided

Test statistics

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

p_1 = proportion of success in sample one

p_2 = proportion of success in sample two

F-TEST

F-Test for testing equality of variances of two normal populations

Test conditions

- The populations are normal
- Samples have been drawn randomly
- Observations are independent; and
- There is no measurement error
- Ha may be one sided or two sided

Test statistics

$$F = \frac{\sigma_{s1}^2}{\sigma_{s2}^2}$$

$with\ (n_1 - 1)$ and $(n_2 - 1)$ d. f.

σ_{s1}^2 is the sample estimate for σ_{p1}^2

σ_{s2}^2 is the sample estimate for σ_{p2}^2

Limitations of the test of Hypothesis

- Testing of hypothesis is not decision making itself; but help for decision making
- Test does not explain the reasons as why the difference exist, it only indicate that the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell about the reason causing the difference.
- Tests are based on the probabilities and as such cannot be expressed with full certainty.
- Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypothesis.