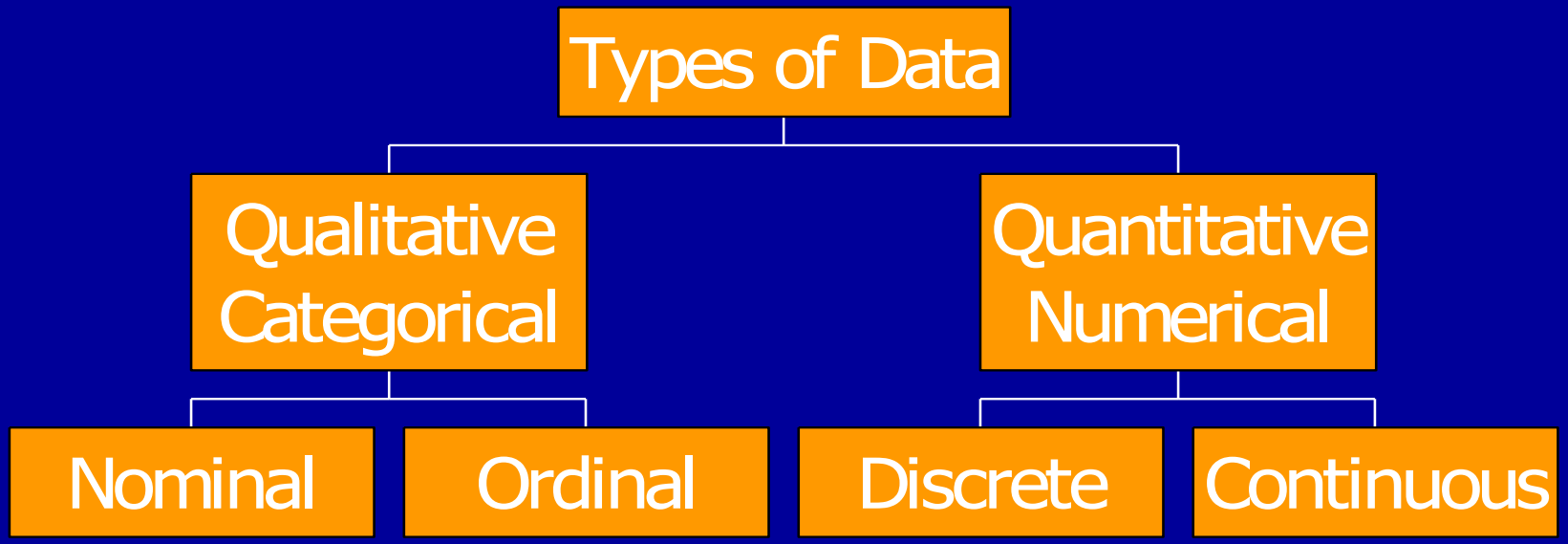
Applied Medical Statistics (Part I)

A Pathmeswaran
Department of Public Health
University of Kelaniya

1. Types of data, Measures of central tendency & dispersion, Normal/Skewed distribution, transformation
2. Graphical presentation of data
3. Sampling
4. Confidence intervals
5. P-value, Type I error & Type II error
6. Significance tests
    1. Z test
    2. Student's T test
    3. Chi-square test
    4. Non-parametric tests
7. Sample size calculations & Power

Types of variables or data

What are variables?


Types of Data
Qualitative Categorical
Quantitative Numerical
Nominal
Ordinal
Discrete
Continuous

Scales of measurement

- Interval scale
- Ratio scale

Collecting Data

- Questionnaires
- Observations
- Measurements

- Use of available data

Quality of Data

Measurements –
Instruments, Observers & Subjects

Minimize Variance and Bias

Calibration, Observer training and
Standardized procedure

		Variance	
		Low	High
Bias	Low	Precise and accurate	Imprecise AND THEREFORE INACCURATE, even though unbiased
	High	Precise but inaccurate	Imprecise and (doubly) inaccurate

Measurement Error

Results of measurement

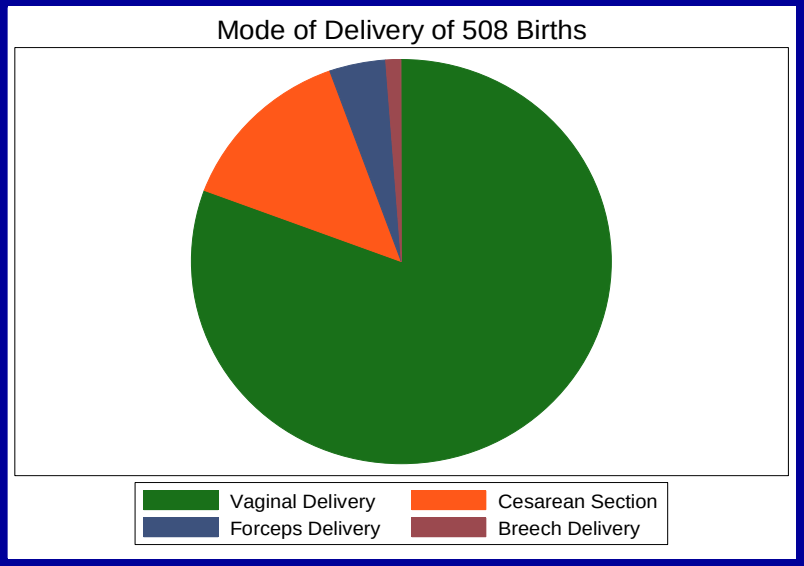
= True value + Measurement error

Summarizing Categorical Data

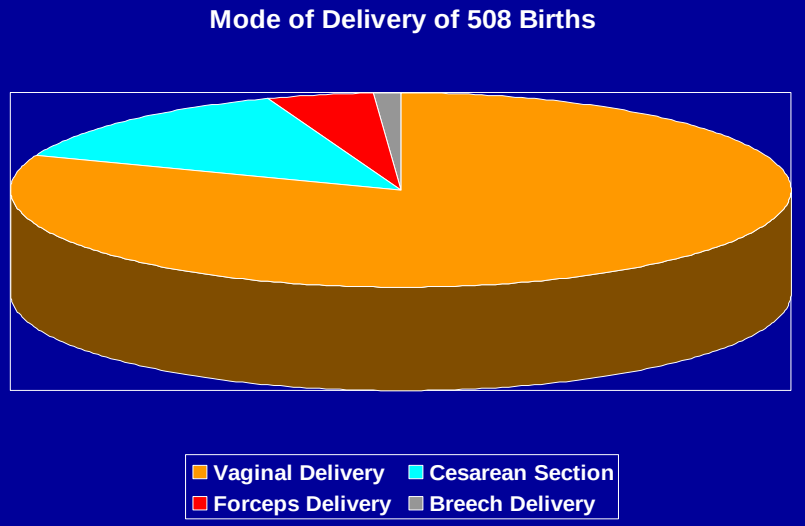
- Tables with percentages
- Graphs – Pie chart

Mode of Delivery of 508 Births

Mode of Delivery	Freq.
Vaginal Delivery	410
Cesarean Section	70
Forceps Delivery	22
Breech Delivery	6
Total	508

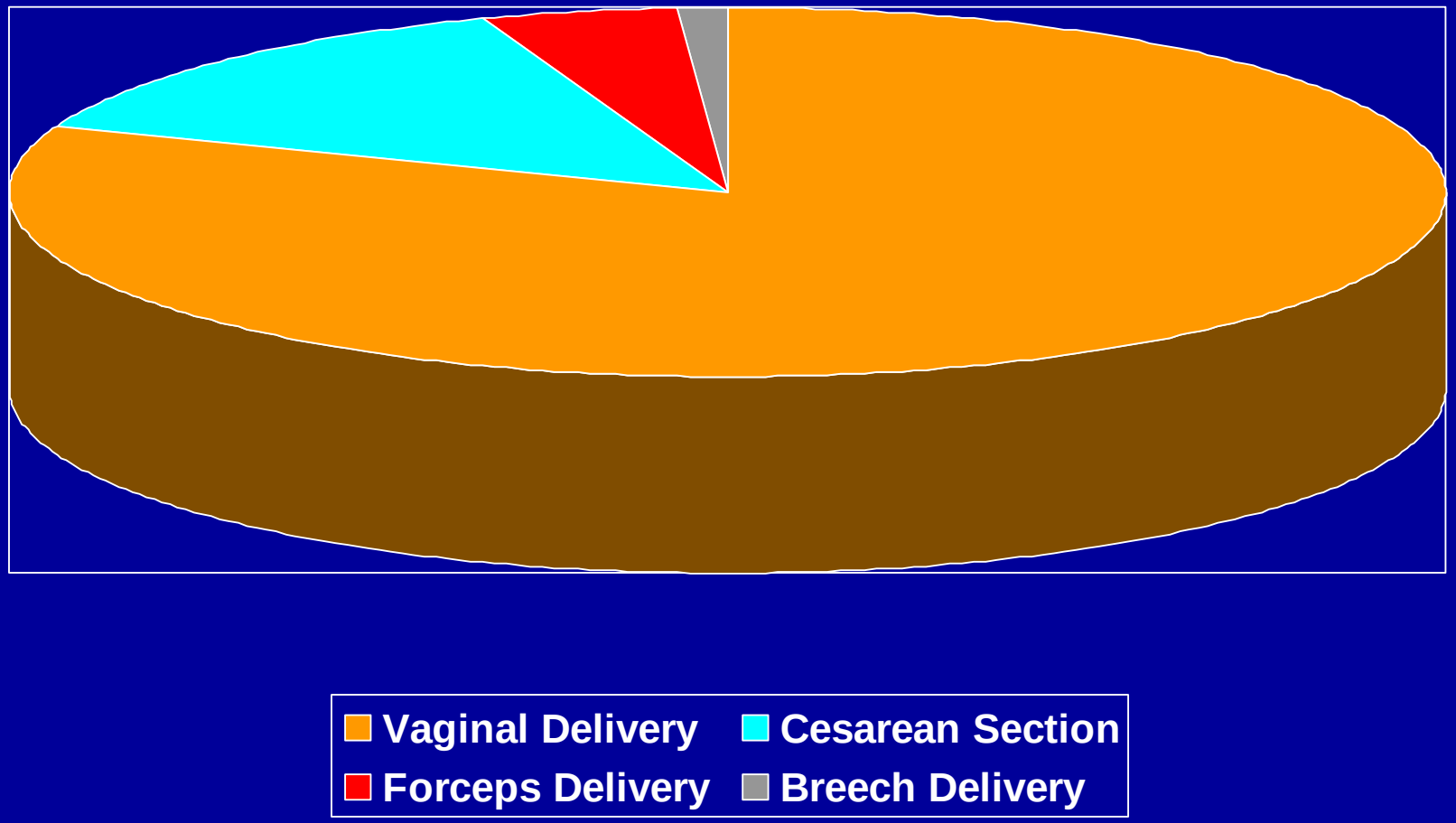

Mode of Delivery of 508 Births
Vaginal Delivery
Cesarean Section
Forceps Delivery
Breech Delivery

A

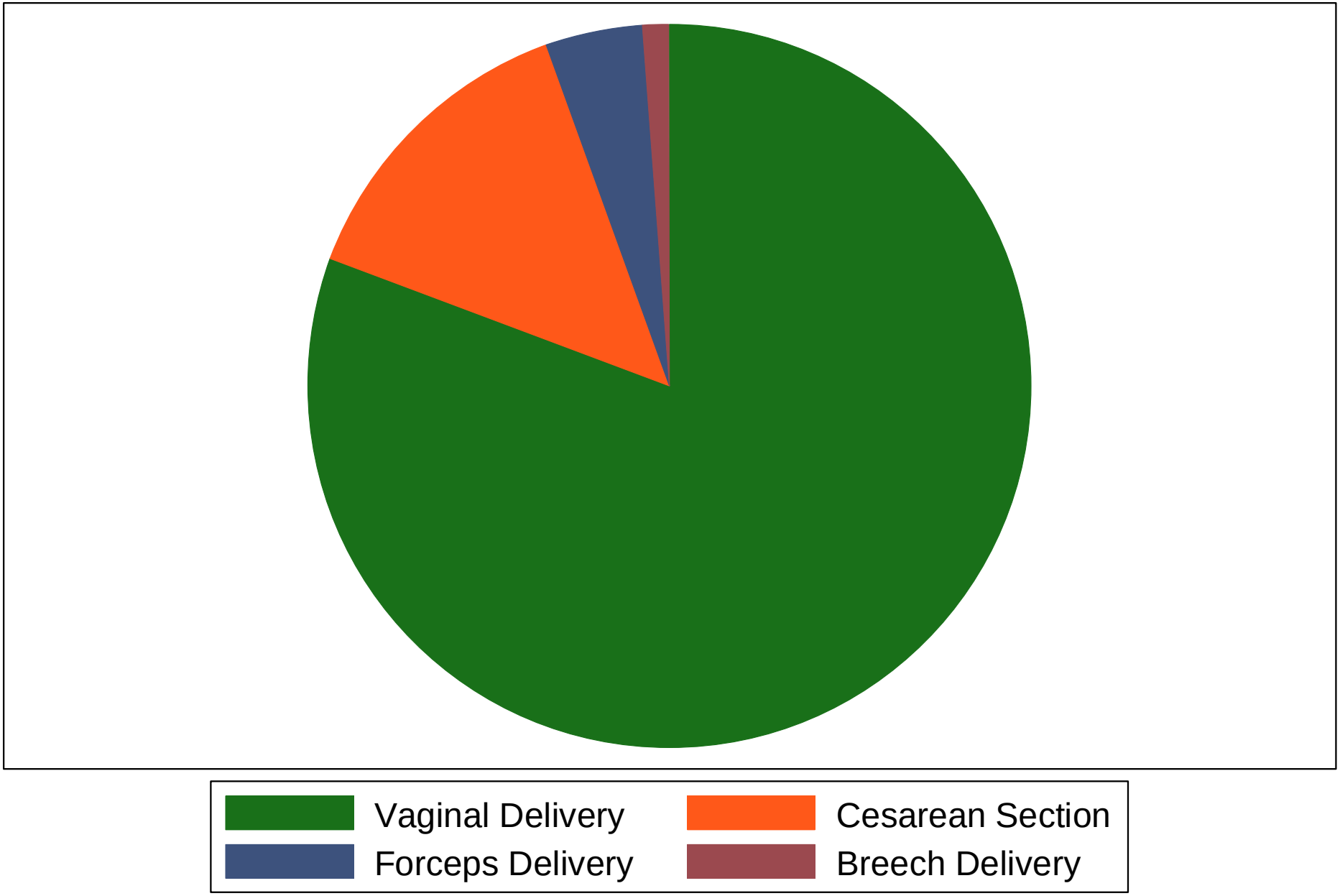

Mode of Delivery of 508 Births
Vaginal Delivery
Cesarean Section
Forceps Delivery
Breech Delivery

B

Mode of Delivery of 508 Births



Mode of Delivery of 508 Births



Mode of Delivery of 508 Births

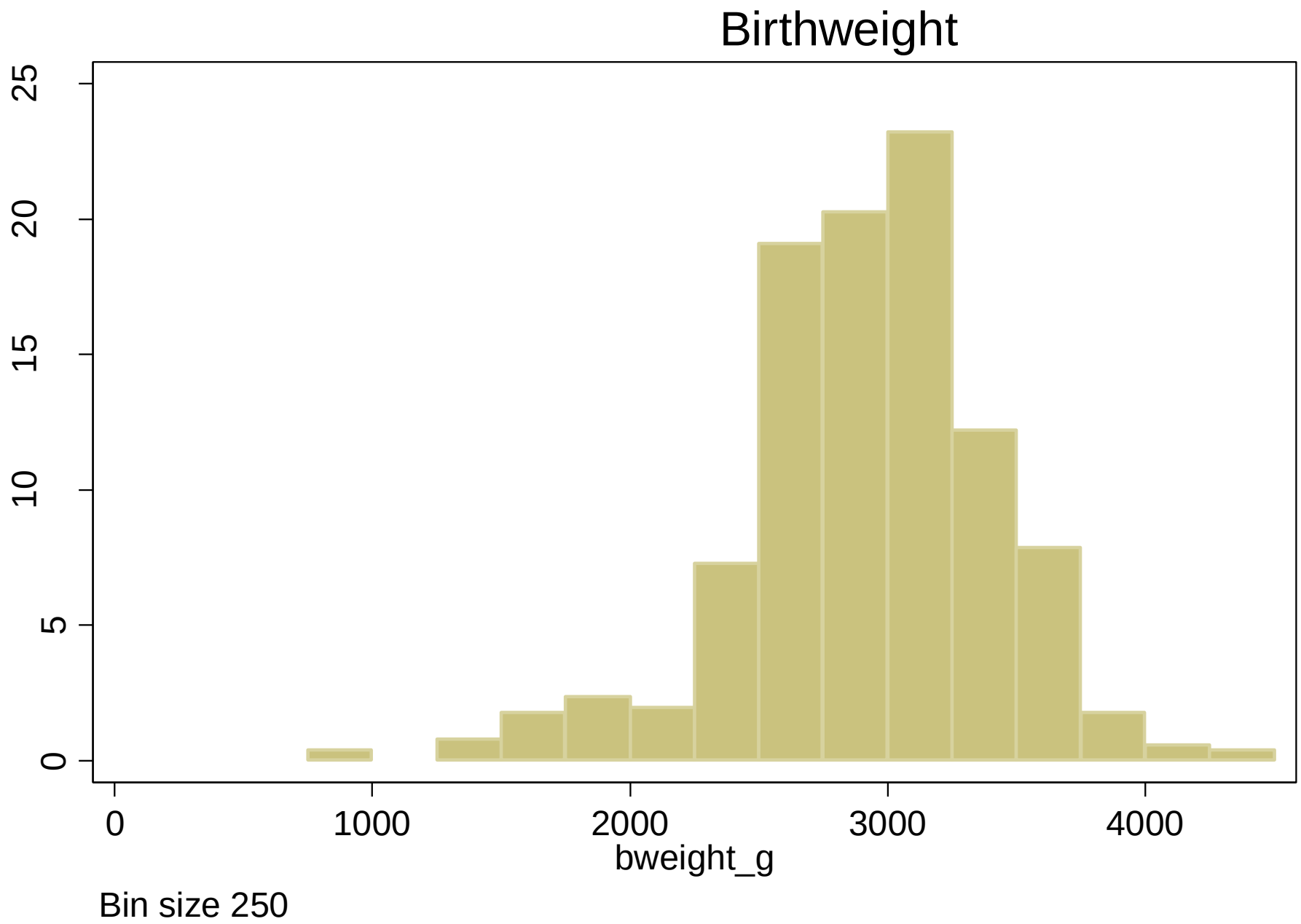
Mode of Delivery	Freq.	Percent	Cum.
Vaginal Delivery	410	80.7	80.7
Cesarean Section	70	13.8	94.5
Forceps Delivery	22	4.3	98.8
Breech Delivery	6	1.2	100.0
Total	508	100.0	

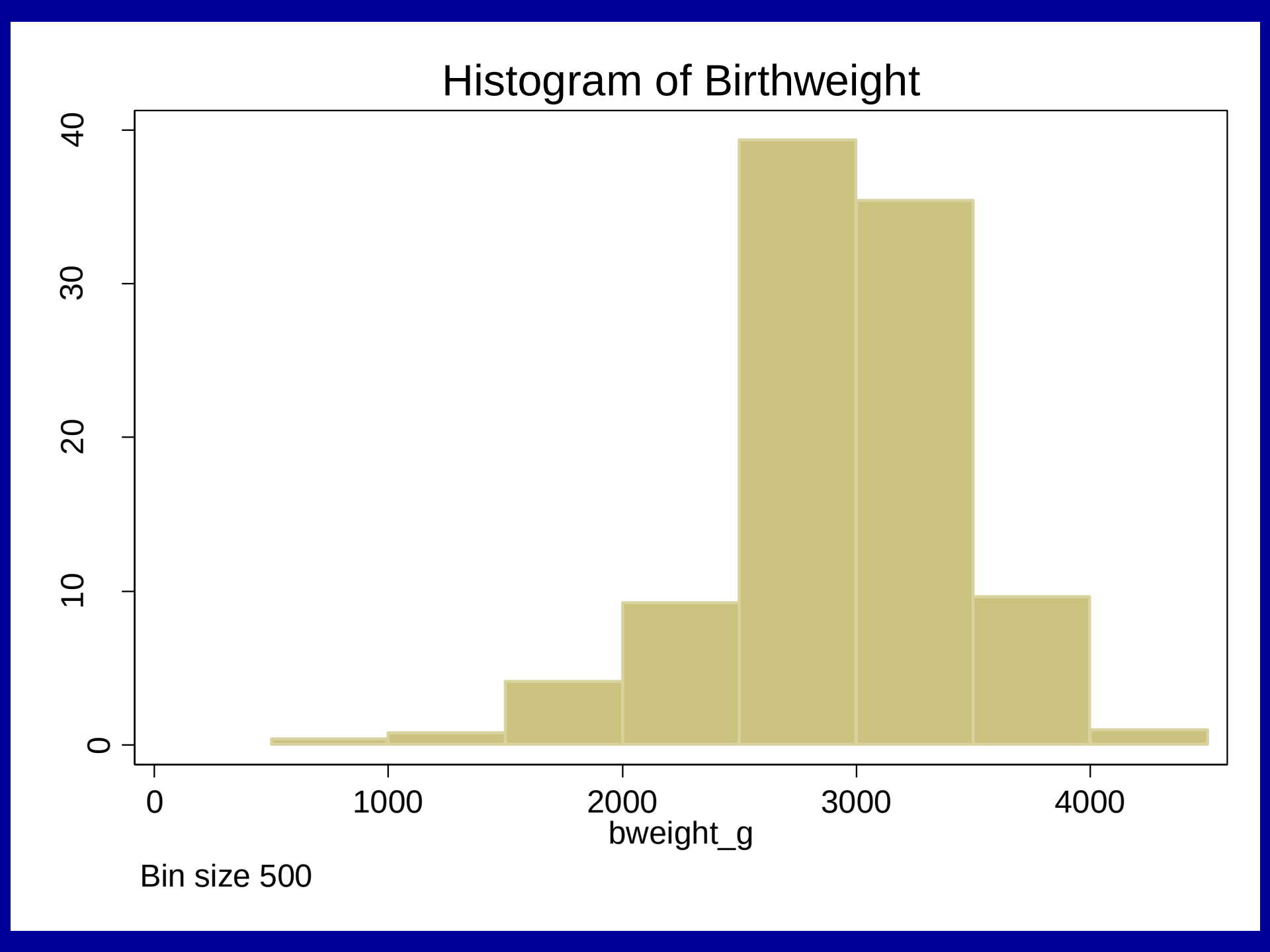
Stem and Leaf Plot of Pre Delivery Weight of 38 Pregnant Women

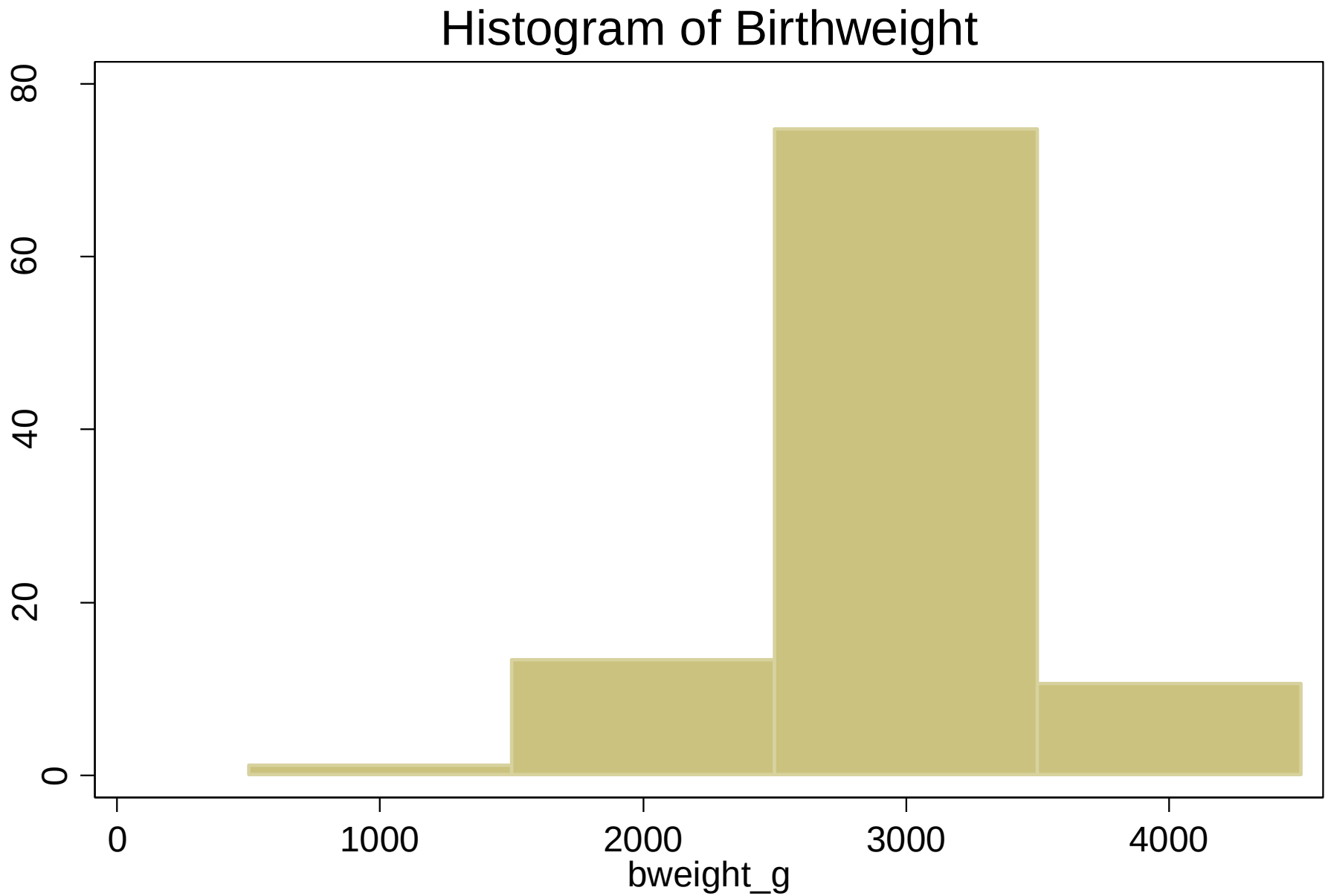
```
4*  |  4556699
5*  |  2333344455889
6*  |  0000111111144
7*  |  00005
```

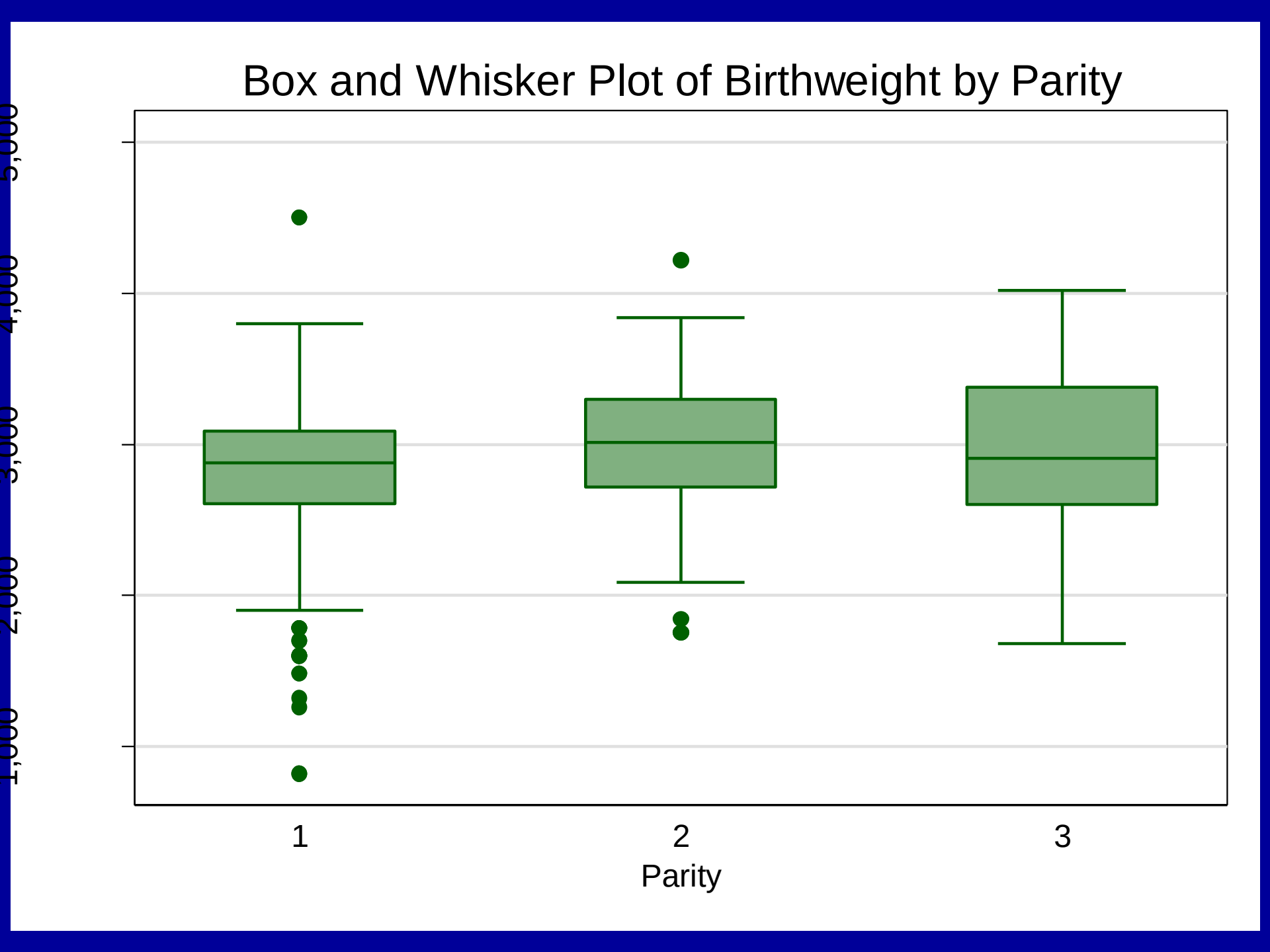
Pre Delivery Weight of 38 Pregnant Women

61	61	64	53	45	49	60	61
54	70	53	60	55	53	61	58
70	46	46	58	55	44	54	59
75	61	61	64	53	45	49	60
60	61	54	70	52	70		


Birthweight
25
20
15
10
5
0
0
1000
2000
3000
4000
bweight_g
Bin size 250


Histogram of Birthweight
0
10
20
30
40
0
1000
2000
3000
4000
bweight_g
Bin size 500


Histogram of Birthweight
0
20
40
60
80
0
1000
2000
3000
4000
bweight_g
Bin size 1000

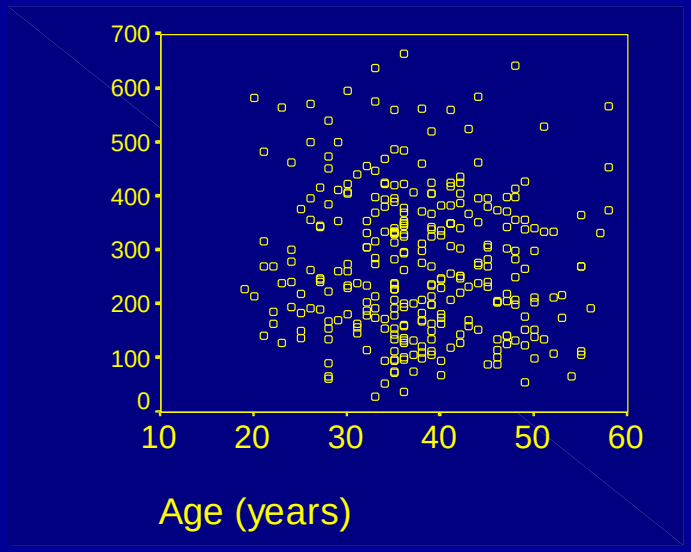

Box and Whisker Plot of Birthweight by Parity
5,000
4,000
3,000
2,000
1,000
1
2
3
Parity

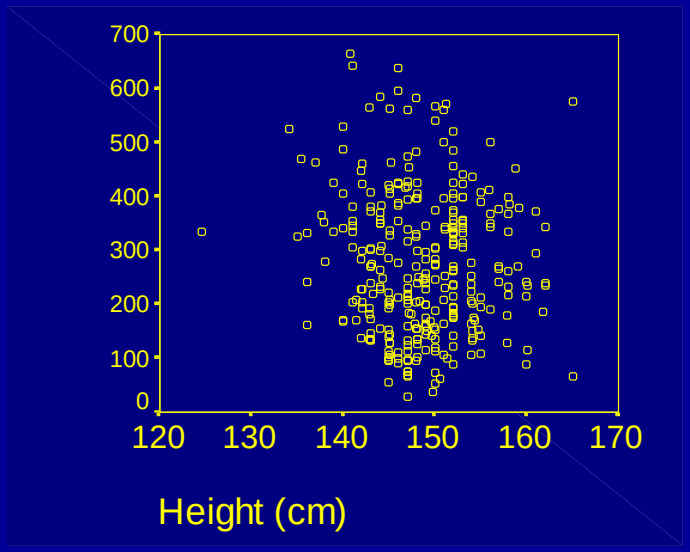
Results

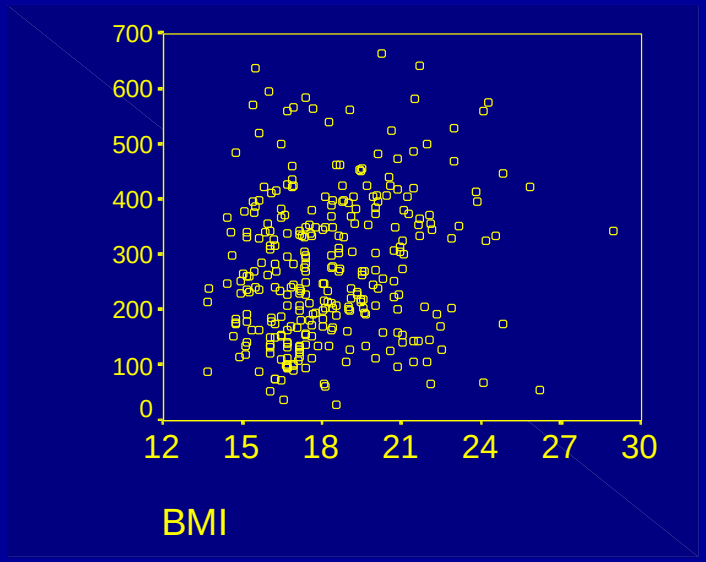
Charecteristics of study subjects (n = 304)

Age in years, mean (sd)	38 (8.4)
No. of Children, median (IQ range)	5 (3 – 6)
Height in cm, mean (sd)	148 (5.6)
BMI in kg/m2, mean (sd)	18.3 (2.5)
Haemoglobin in g/dl, mean (sd)	11.2 (1.5)
Leaves per month in kg (IQ range)	260 (170 – 270)

Productivity Vs Age, Height, BMI & Hb







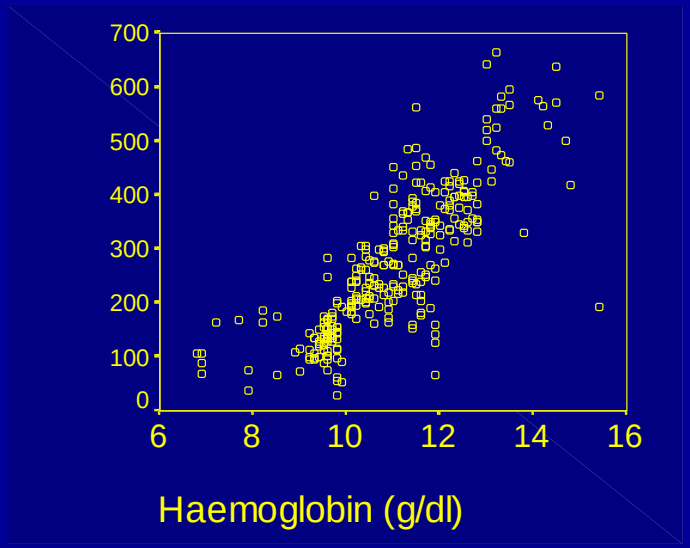
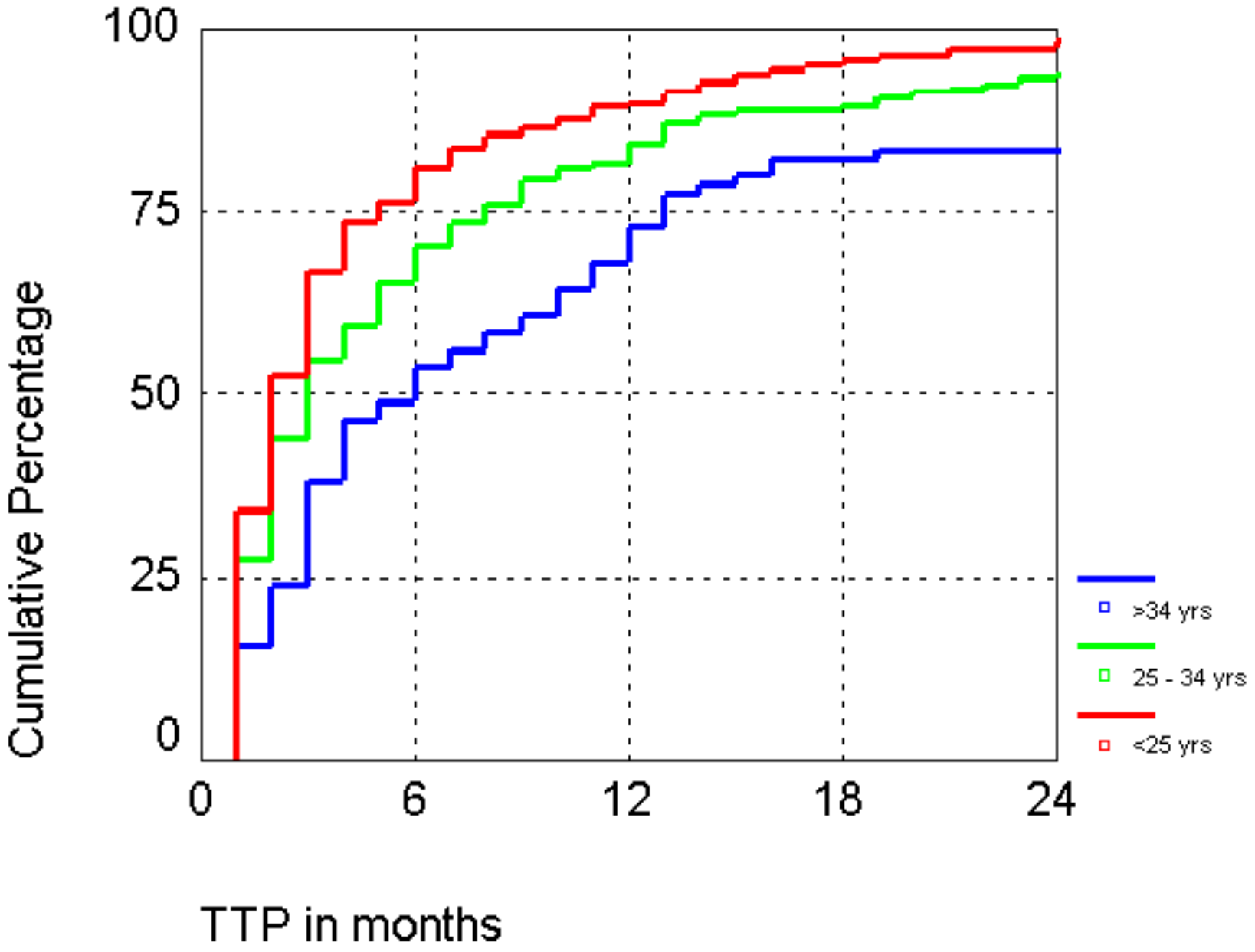
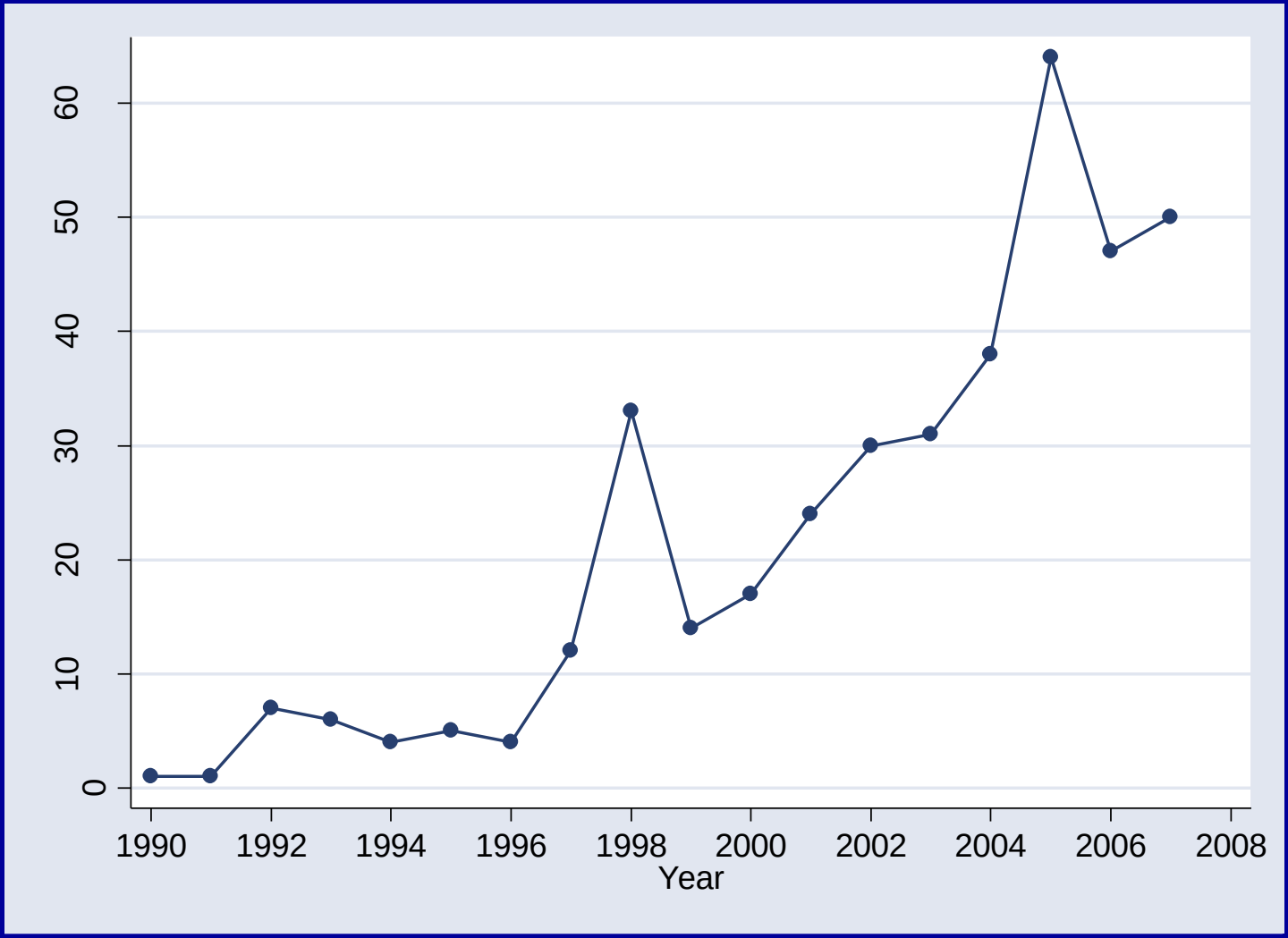


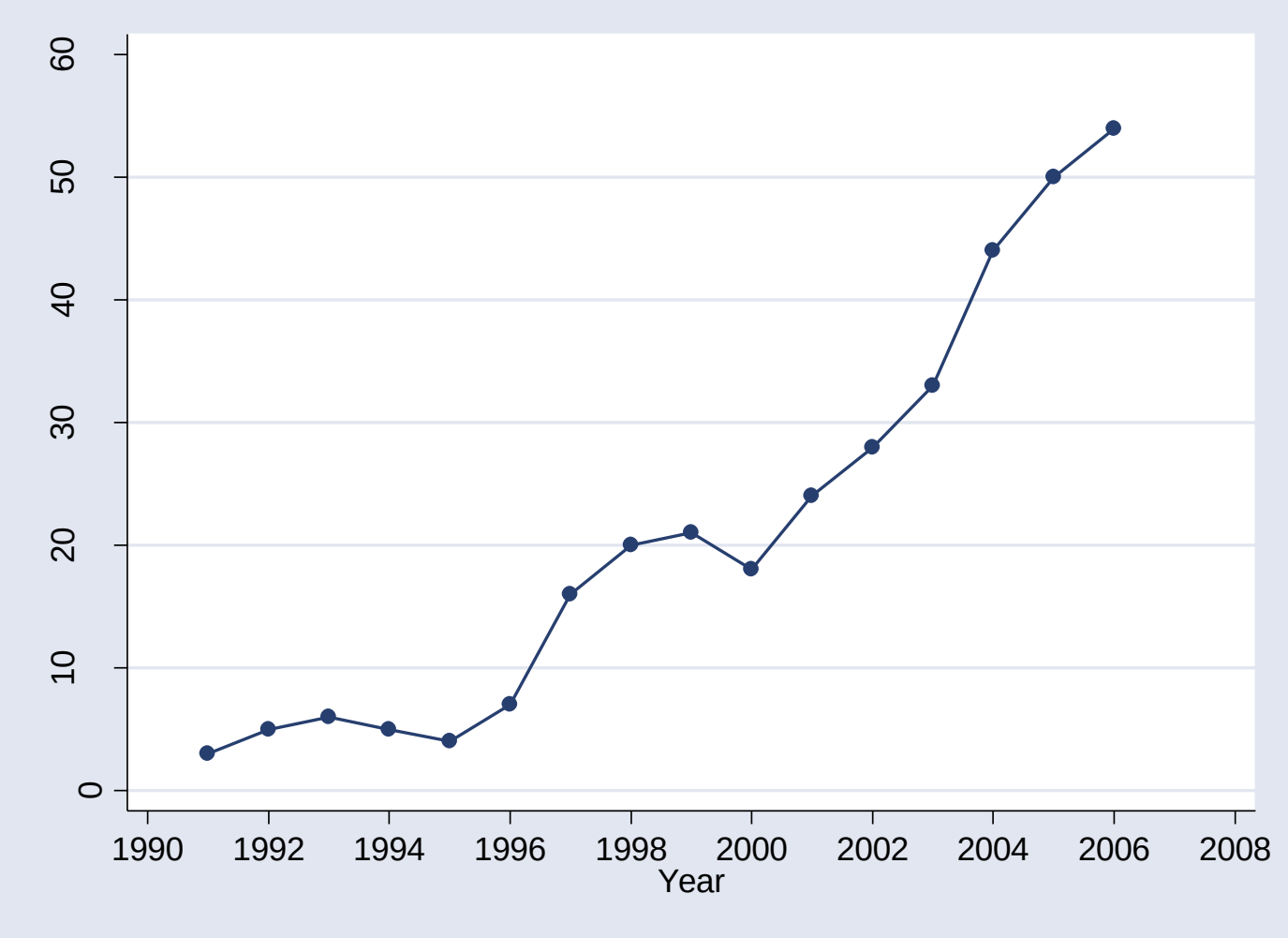
Figure: TTP by Age



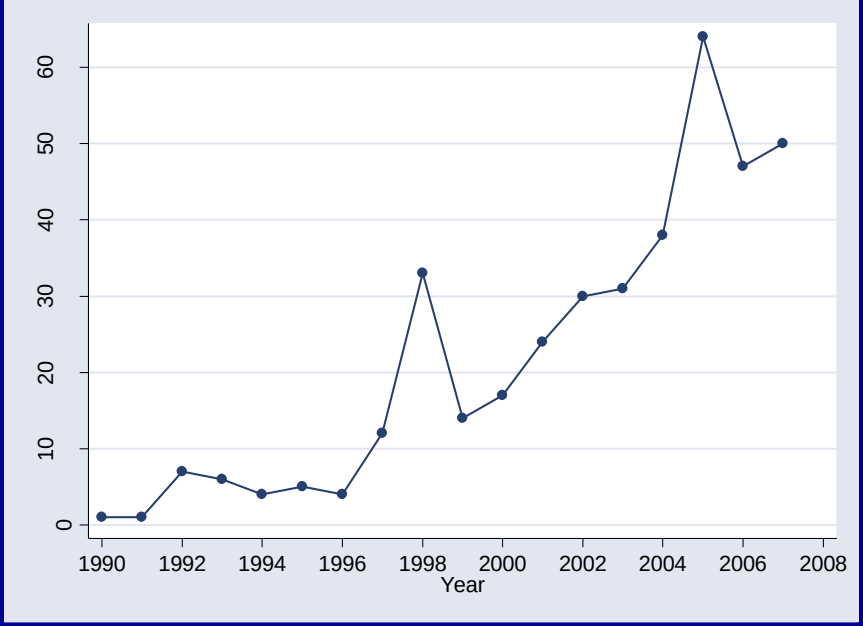
Reported number of HIV cases Sri Lanka 1990 - 2007

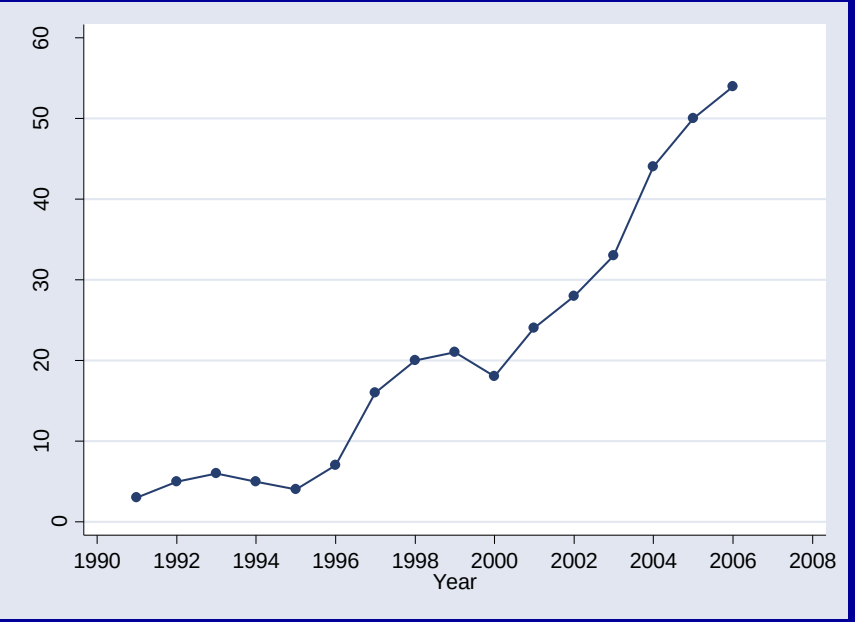


Reported number of HIV cases Sri Lanka 1990 - 2007



Reported number of HIV cases Sri Lanka 1990 - 2007





Rounding and significant figures

- Avoid misleading precision in summary statistics
- One decimal place more than original data
- Numbers with more than three significant figures rarely needed

Summarizing Numerical Data

- Central tendency/ Average - middle
- Dispersion/ Variation - spread

Measures of Central Tendency or Average

Mean

Median

Mode

Measures of dispersion or variation

Range – Difference between the highest and lowest values.

Quantiles – divisions of a distribution into ordered subgroups of equal size
- Quartiles
- Centiles

Standard Deviation

Quantiles

Set of values which divide a frequency distribution into equal groups, each containing the same fraction of the total population

Quartiles

Quintiles

Deciles

Centiles

Standard Deviation

- It indicates the spread of individual values around the middle (mean or sometimes the median)

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{(n-1)}}$$

Descriptive/Summary statistics

- Central tendency/average

  Mean, Median & Mode

- Dispersion/spread

  Range, Quartiles & Standard deviation

Which of the following is/ are true for the sample 3, 1, 2, 7, 2, 4, 2

1. The mean is 3
2. The median is 7
3. The mode is 2
4. The range is 1
5. The standard deviation is 2

Twenty HIV positive patients developed AIDS in January 2002. By December 2003, five of them were still alive. The survival times (in months) of the 15 who died were – 6, 7, 8, 8, 9, 10, 10, 11, 11, 12, 12, 15, 18, 18, & 20.

Determine the average survival time of these patients?

Platelet count of dengue patients

,000 per micro liter

Percentile	
5%	19
10%	25
25%	42
50%	68
75%	98
90%	126
95%	140

Obs	479
Mean	73
Std. Dev.	41

Normal Distribution

- The distribution of many biological variables in the population follow a pattern.

1. Data concentrated around the mean
2. Values on either side of the mean occur in equal frequency
3. Frequency of values fall as they get away from the mean

- Most values (68%) are within one standard deviation on either side of the mean
- Large majority (95%) are within two standard deviations on either side of the mean
- Almost all values (99.7%) are within three standard deviations on either side of the mean

Normal Range or 95% Reference Interval

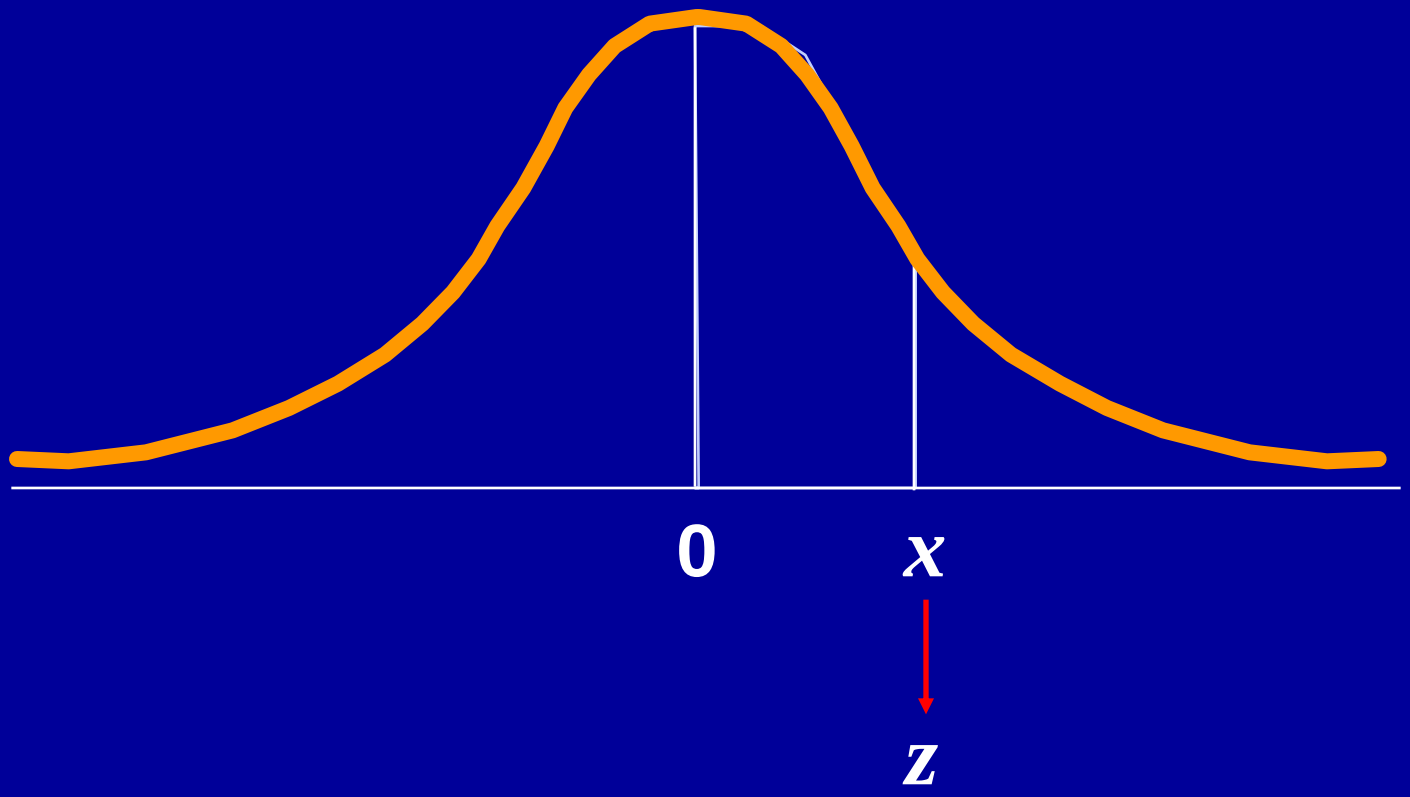
- Set of values within which 95% of values of healthy people will lie
- When the variable follows the Normal distribution this is
  Mean – 2 SD to Mean + 2 SD
- If not 2.5 centile to 97.5 centile

Z score

- The difference between a particular observation and the mean in terms of the standard deviation

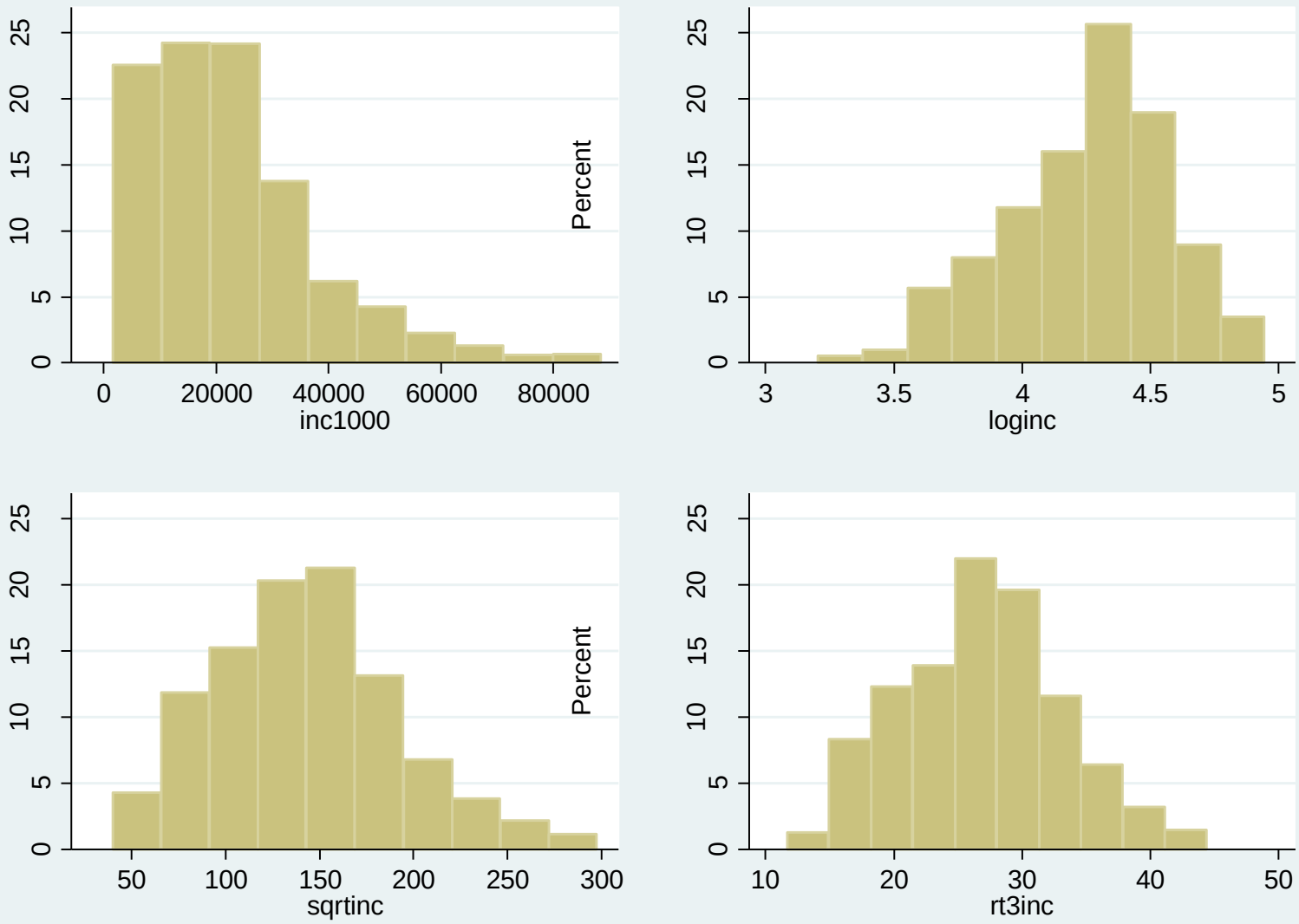
Standard Normal Distribution

$\mu = 0$ $\sigma = 1$



- Skew
- Transformation

Income distribution - raw & transformed data



Income distribution - raw & transformed data