

Statistics

Contents

1	1.1 – Key Words and Definitions	5
1.1	Key words	5
1.2	Types pf Data	5
1.3	Two Types of Data	5
1.3.1	Two types of Quantitative Data	6
1.4	4 Levels of Measurement	6
1.5	Design of Experiments/Observations	6
1.5.1	Observation vs. Experiment	6
1.5.2	Random	6
1.5.3	Common techniques to get a sample	6
1.6	Frequency Distribution	7
1.6.1	Touching Bar Chart	8

Chapter 1

1.1 – Key Words and Definitions

1.1 Key words

Data	Any observations that have been collected.
Statistics	Collect, analyze, summarize, interpret and draw conclusions from there.
Population	The complete set of elements being studied.
Samples	Some subset of the population.
Census	Collection from every member of a population.

Table 1.1: Statistics Vocabulary

→ If you take a sample, it must be collected **randomly**.

1.2 Types of Data

P-P Parameter	A characteristic of a population.
S-S Statistic	A characteristic of a sample.

Table 1.2: Statistics Vocabulary

1.3 Two Types of Data

Qualitative (Categorical)	Data that is non-numerical e.g. color, gender, race, zip-codes... Mathematical operations are meaningless .
Quantitative	Numerical e.g. height/weight, wages, temperature, time. Mathematical operations are meaningful .

Table 1.3: table

1.3.1 Two types of Quantitative Data

Discrete data	Countable or finite Numbers of eggs, dice...
Continuous Data:	Infinite number of possible values (not countable) Usually a measurement , e.g. temperature.

Table 1.4: Quantitative data

1.4 4 Levels of Measurement

Nominal	Categories not ordered. e.g. religion
Ordinal	Can be ordered, differences are meaningless Rank, color (spectrum)...
Interval	Ordered, differences are meaningful, no "Natural Zero" e.g. temperature
Ratio	Just like interval, but with a natural zero. e.g. amount of money

Table 1.5: Measurements

1.5 Design of Experiments/Observations

1.5.1 Observation vs. Experiment

An **observation** measures specific traits, but does **not** modify subjects.

An **experiment** applies a treatment and then measures the effect on the subjects.

1.5.2 Random

Each member of a population, has an equal chance of being selected in a sample.

Simple random sample

Each group of size 'n' has an equal chance of being selected.

1.5.3 Common techniques to get a sample

Table 1.6: 4 Common techniques to get a sample

Convenience sample	You use the results, which you easily get (not random)
Systematic sampling	Put a population in some order and select every " k^{th} " member.
Stratified Sample	Breaking population into sub-groups based on some characteristic, and then take a simple random sample out of each sub-groups.
Cluster sample	Divide population into "clusters" (regardless of characteristic), randomly select a certain number of clusters, and then collect data from the entire cluster.

1.6 Frequency Distribution

A frequency distribution is a list of values with corresponding frequencies.

Class width	Difference between two "lower class limits"
Lower class limit	Smallest value belonging to a class
Upper class limit	Highest value belonging to a class

Table 1.7: Frequency Distribution Terms

Steps:

1. Determine number of classes: 8

2. class width:

$$\frac{\text{Max Value} - \text{Min value}}{\text{number of classes}} \rightsquigarrow \frac{44 - 18}{8} \rightsquigarrow \frac{26}{8} \rightsquigarrow 3.25$$

Round **up**. $\rightsquigarrow 4$

3. Start with the minimum value: 18

4. Create classes with class width (4)

5. Find the class midpoint:

$$\frac{\text{upper class limit} + \text{lower class limit}}{2} \rightsquigarrow .$$

6. Class boundaries: used to separate classes without gaps.

class width: 4

Lower class limit: 18, 22, 26, ... 46

upper class limit: 21, 25 ... 49

class midpoint:

$$\frac{\text{upper class limit} + \text{lower class limit}}{2}$$

$\rightsquigarrow 19.5, 23.5, 27.5, 31.5, 35.5, 39.5, 43.5, 47.5$

class-width inbetween

class boundaries: Used to separate classes without gaps. 17.5, 21.5, 25.5, 29.5, 33.5, 37.5, 41.5, 49.5

Relative frequency distribution: Percentage

$$\frac{\text{class } f.}{\sum f.(n)}$$

Cumulative Frequency Distribution Adds sequential classes together.

Age	Freq.	Rel. Freq.	Cum. Freq.
18-21	25	58.1%	25
22-25	10	23.3%	35
26-29	4	9.3%	39
30-33	2	4.7%	41
34-37	1	2.3%	42
38-41	0	0%	42
42-43	1	2.3%	43
46-49	0	0%	43
n=43 $\sum f \uparrow$		100%	

Table 1.8: Frequency Distribution

1.6.1 Touching Bar Chart

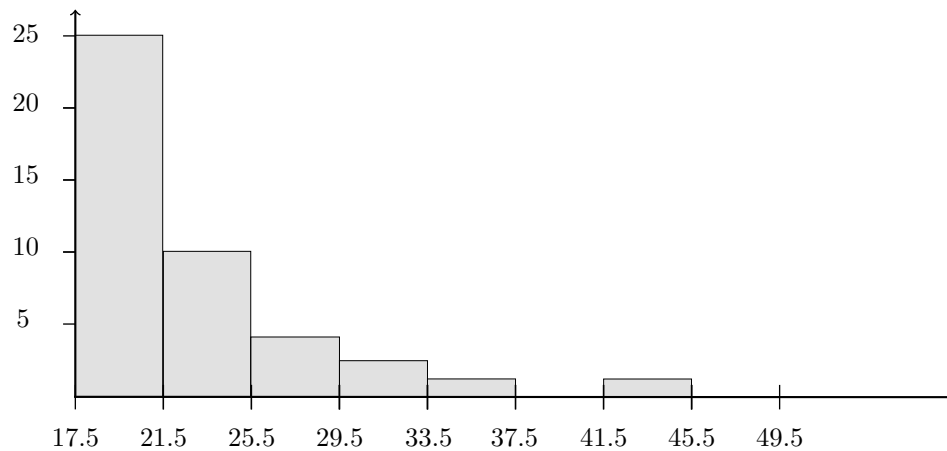


Figure 1.1: figures/stats-1

A cumulative chart would look exactly the same, but instead of having boundaries numbered it'd be in the middle of the bars with the cumulative frequency from class 1-8. And also the y-axis would be the percentage.

There is also a last one, where one takes the cumulative stuff, so that the graph columns are getting bigger and bigger...

Horizontal: Class midpoints or boundaries.

Vertical: Frequency.