

STAT 2000, Section 001
Introduction to Statistical Methods
Fall Semester 2002

Dr. Jürgen Symanzik

Utah State University

Department of Mathematics and Statistics

3900 Old Main Hill

Logan, UT 84322-3900

Tel.: (435) 797-0696

FAX: (435) 797-1822

e-mail: symanzik@sunfs.math.usu.edu

Contents

1	Introduction: What is Statistics?	1
2	Looking at Data — Distributions	1
2.1	Keywords	1
2.2	Displaying Distributions with Graphs	2
2.3	Describing Distributions with Numbers	8
2.4	The Normal Distributions	15
3	Looking at Data — Relationships	22
3.1	Scatterplots	22
3.2	Least-Squares Regression	28
3.3	Correlation	31
3.4	Residuals (or Errors)	32
3.5	Relations in Categorical Data	34
4	Producing Data	36
4.1	First Steps	36
4.2	Design of Experiments	37
4.3	Sampling Design	40
4.4	Toward Statistical Inference	41
5	Probability: The Study of Randomness	44
5.1	Randomness	44
5.2	Probability Models	44
5.3	Random Variables	47
5.4	Mean and Variance of a Random Variable	55
6	From Probability to Inference	59
6.1	Sampling Distributions for Counts and Proportions	59
6.2	The Sampling Distribution of a Sample Mean	67

7	Introduction to Inference	70
7.1	Estimating with Confidence	70
7.2	Tests of Significance	73
7.3	Use and Abuse of Tests	76
8	Inference for Distributions	78
8.1	Inference for the Mean of a Population	78
9	Useful Future Statistics Courses	81

1 Introduction: What is Statistics?

Three main topics discussed in this class:

(i) **(Simple) Data Analysis:**

methods to organize and describe data using graphs, summaries, and mathematical descriptions, e.g.,

(ii) **Data Production:**

methods to produce data, e.g., how to select samples and how to design experiments, e.g.,

(iii) **Statistical Inference:**

draw conclusions based on existing data and determine how reliable these conclusions are, e.g.,

2 Looking at Data — Distributions

2.1 Keywords

Individuals:

objects described by the data set, e.g.,

Variable:

characteristic of an individual, e.g.,

Types of Data:

these terms below determine our statistical analysis:

(i) **Qualitative or Categorical data (variable):**

data is non-numerical, e.g.,

(ii) **Quantitative data (variable):**

data is numerical

(a) **Discrete:**

(b) **Continuous:**

- (iii) (a) **Univariate data:**
one response is measured per individual
- (b) **Multivariate data:**
several responses are measured per individual
- (iv) **Distribution of a Variable:**
tells us which values a variable takes and how often it takes these values

2.2 Displaying Distributions with Graphs

Histograms:

most common graph of the distribution of a quantitative variable

Example: "Most Widely Held Stocks"

Class (Amount in \$) Tally Frequency (Number of Stocks)

Histogram:

Note:

- Vertical axis must start at 0
- No spaces between bars
- Each bar with same width

Example: "The Cost of Victories" – Wins

or better:

or best (in this example) after splitting stems:

Time Plots:

Plot of one variable versus time. Note that the time is **always** on the horizontal axis, increasing from left to right. Scale should be uniform, i.e., even if data is only available for a few year such as 1980, 1990, 1995, and 2000, the spacing between 1980 and 1990 **must** be twice as wide as the spacing between the other years.

Main components of a time plot are

- Trend: ...
- Seasonal Pattern: ...

Examples:

Bar Charts and Pie Charts: (for Categorical Data)

•Bar Chart:

Example: “Real Quiet the Real Thing?”

Note:

- Vertical axis must start at 0
- Spaces between bars

Other Examples:

•Pie Chart:

Example: “Real Quiet the Real Thing?”

Construction:

Sum up numbers, each slice is a fraction of the total size of the pie, i.e.,

Good design:

Start with the biggest “slice” at “noon” position. Move clockwise, using largest values first.

Avoid 3-dimensional slice charts – for those the area in the background seems to be smaller than the area in the foreground and thus the visual impression is misleading.