Economics 2370

Statistics and Probability

Sharon I. O'Donnell

Email: odonnell@bayou.uh.edu

Office: 246 M

Phone: (713)743-3818

Office hrs: 10:00 to 11:00 Tuesday and Thursday
by appointment or by chance

URL: http://www.uh.edu/~odonnell/econ2370

What this course will cover.

1. Presentations of data through graphs and charts,

2. Measures of central tendency and dispersion,

3. Probability and probability distributions,

4. Sampling plans and hypothesis testing of large
    and small samples.

Textbook, supplemental materials, etc.

Textbook Mendenhall et al: *Introduction to Probability and Statistics* 10th edition.

Material on the Web Lecture notes, syllabus, tutorial notebooks on the software, notebook examples from class, grades.

Software *Mathematica*

Methods of obtaining/accessing *Mathematica*

1. Purchase a cd of the software for $4.00

2. Use software in Social Work lab.

Grading

Note: Grades will be posted on the web page only.

1. The grade will be based on 2 computer projects (5% each) 2 midterm exams (25% each) and a final exam (40%).

2. Exams are closed notes, closed book.

3. I have included a list of optional problems. To motivate you to work on these problems, each test will include questions from this set.

Unannounced in-class quizzes

Math review for statistics

Introduction and Chapter 1

Motivation Why economists are interested in statistics and probability.
Why others should have a better understanding of statistics and probability.

Basics

Data → **Unit of analysis, experimental units, observations.**

Full set of data → **population.**

Random subset of population → **sample.**

Characteristics described by the data → **variable.**

Types of data sets (sample data sets or population data sets)

Data set can describe a large number of units at a moment in time → **Cross sectional data.** (e.q., 2000 US Census for Houston)

Data set can describe one type of unit over time → **time series.** (e.g., US GDP for the last 100 years.)

Data set can be a mixture of both → **panel data set.** (e.g., The GDPs of US, UK, and Germany for the last 100 years.)

Types of analyses

1. **Univariate analysis** - research analysis based on the examination of one variable.

2. **Bivariate** - analysis based on the examination of two variables. Often, the researcher is interested in understanding if there exists any association between the two variables.

3. **Multivariate** - three or more variables. Here, the researcher is also interested in the relationship among the variables.

Types of variables

1. Qualitative

   (a) Categorical-multiple categories

   (b) Ordinal-Values represent ranking or ordering

2. Quantitative (or interval) variable

   (a) Discrete - countable set of values.

   (b) Continuous

Other Data issues

Missing data Why does it exist? How can it affect the analysis?

Outliers

For each example below, what is the unit of analysis and variable type.

1. June, 1999 Unemployment rates for the counties in Texas.

2. Exit poll of the last Presidential election.

3. Responses from the question from a class course evaluation survey that rates the quality of the class.

Summation Operator

General concepts describing statistical tools

1. Descriptive Statistics

   (a) Concerned mainly with collecting, summarizing and interpreting data.

2. Inferential Statistics

   (a) Process of drawing conclusions about a population or making predictions using random samples.

Graphical representation of data

Graphical displays should

1. show the data.

2. induce the viewer to think about the substance
   rather than the methodology, graphic design,
   the technology of graphic production.

3. avoid distorting what the data have to say.

4. present many numbers in a small space.

5. make large data sets coherent.

6. encourage the eye to compare different pieces of data.

7. reveal the data at several levels of details, from a broad overview to a fine structure.

8. serve a reasonable purpose: description, exploration, tabulation, or decoration.

9. be closely integrated with the statistical and verbal descriptions of a data set.

Stem and leaf plot - graphical representation of interval and ordinal data using numbers.

1. Divide data into n main categories - **stem** portion of the graph.

2. Within each category, order values (the **leaves**) from lowest to highest.

3. List the stem in the first column, leaves in the remaining columns.

Example - Stem and leaf plot

Data

x1	27	x6	20	x11	47	x16	26
x2	25	x7	17	x12	34	x17	28
x3	31	x8	24	x13	26	x18	30
x4	19	x9	25	x14	31	x19	31
x5	40	x10	32	x15	22	x20	41

Plot

1 | 7 9
2 | 0 2 4 5 5 6 6 7 8
3 | 0 1 1 1 2 4
4 | 0 1 7

Part of Graphs

1. Body

2. X-axis (horizontal axis)

3. Y-Axis (vertical axis)

4. Z-Axis (in 3D graphs, the X & Y axes become the horizontal axes, the Z axis is the vertical)

5. Axis labels

6. Data Labels - often stored in a Legend.

Type of graphs - univariate analysis
Shows how the population or sample is distributed
by groups. Examines relative proportions.

1. Pie Chart

2. Bar Chart - **Histogram**

3. Box-plot

Type of graphs - bivariate analysis

Describes the association of two variables.

1. Scatterplot

Type of graphs - multivariate analysis

Describes the association and interaction of three variables.

1. 3D plots/Contours

Frequencies, Relative Frequencies and Cumulative Relative Frequencies.

1. Describes the distribution of univariate data.

2. Distribution $\rightarrow$ term I will use throughout the class.

Compute **Frequencies.**

1. For interval variables, group data into K mutually exclusive categories. For other variables, use existing categories.

2. Count within each category $\rightarrow$ **Frequency**

3. Compute Total number of data items $\rightarrow$ **n** (for sample), **N** (for population)

Compute **Relative Frequencies**

$$\frac{\text{Frequency of category } j}{\text{Total}}$$

1. Range of values from 0 to 1.00

Compute **Cumulative Relative Frequencies**

$$\frac{\sum_{i=1}^{j} \text{Frequency of category } i}{\text{Total}}$$

1. Range of values from 0 to 1.00

Compute **Relative Frequency as an angle** - used for PieCharts.

Relative Frequency of category j as an angle $=$

Relative Frequency of category $j * 360$

1. Range of values from 0 to 360

Example 1.3

Tools to analyze graphical information - univariate

1. Is the distribution symmetric or asymmetric?

2. Describing the peaks - unimodal or bimodal distribution.

3. Describing the shape - degree of skewness.