Statistics 101 Course Notes

Introduction to Quantitative Methods for Psychology and the Behavioral Sciences

Instructor: Alan Agresti

Course syllabus: At top of course home page, which is also at
www.stat.ufl.edu/~aa/harvard

Teaching fellows:
Joey Kelly, office hours in SC 600 Wed 1:30 to 2:30, Fri 11-12
Jon Hennessy, office hours in SC 608 Tues and Thurs 4-5

(or, by appointment for any of us)

1. Introduction

- Data - Information collected to gain knowledge about a field or to answer a question of interest.

- Data Sources include:
  - Surveys (Mail, Telephone, Internet)
  - Experiments

- *Statistics*- Set of methods for collecting/analyzing data (the art and science of learning from data)

Statistics provides methods for:

- *Design* - Planning/Implementing a study
  - Sample survey or experiment?
  - How to choose people (subjects) for the study, and how many?

- *Description* – Graphical and numerical methods for summarizing the data

- *Inference* – Methods for making predictions about a population (total set of subjects of interest), based on a sample (subset of the sample on which study collects data)

Examples

- How can we study whether a new therapy is better than a standard therapy for treating depression?
- How is crime rate for a county associated with factors such as % of residents with at least a high school degree, % unemployed, % in urban environment, … ?
- Is college GPA most correlated with IQ, average time studying per week, high school GPA, SAT score, number of hours spent on social activities, … ?

- *Parameter* – Numerical summary of the population
  - Population mean
  - Population proportion

- *Statistic* – Numerical summary of the sample

We use the sample statistic to make inferences about the population parameter.

Examples: parameters / statistics

Parameter	**Statistic**
% of all adult Americans who approve of Barack Obama's performance as President	% of 1000 adult Americans in a poll who approve of Obama's performance as President
Mean reaction time to red light in experiment when using (not using) cell phone while driving	Mean reaction time to red light for 100 students in experiment when using (not using) cell phone while driving

Note:

- Populations can be *actual* sets of people or *conceptual* (hypothetical)
- For good inferences, need sample to be representative of population
- *Statistical software* (such as SPSS) is used to analyze data

Software applies to *data files*

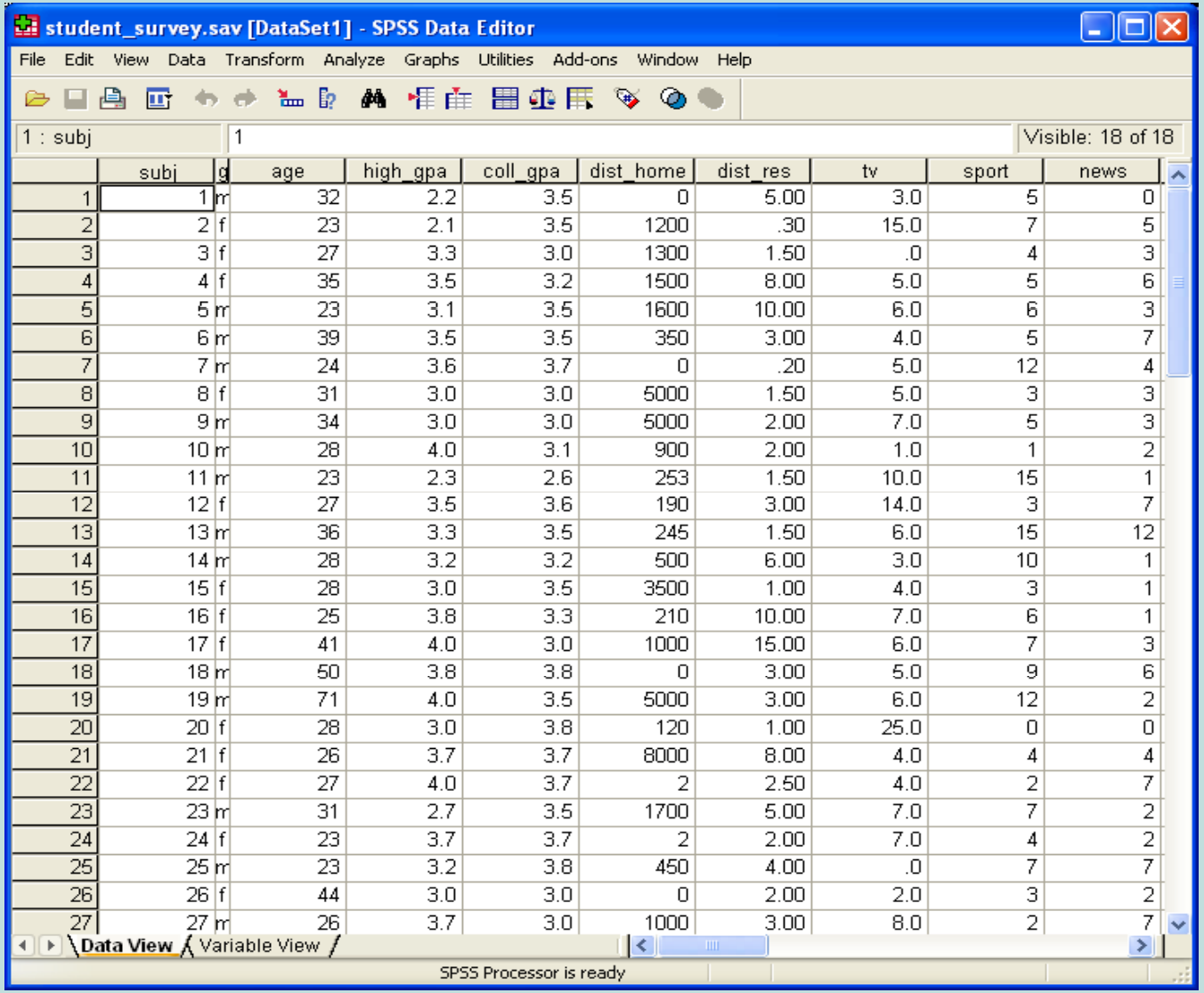
- Any one row contains observations for particular subject (person) in sample
- Any one column contains observations for a particular characteristic (“variable”) measured. The names of the characteristics are at top of file, in first row.

Examples: Go to
www.stat.ufl.edu/~aa/social/data.html

The first data file, from a survey of 60 students at Univ. of Florida, looks like:

subject	gen	age	high	coll	tv	veg	party	ideology	abor
1	m	32	2.2	3.5	3	n	r	6	n
2	f	23	2.1	3.5	15	y	d	2	y
3	f	27	3.3	3.0	0	y	d	2	y
4	f	35	3.5	3.2	5	n	i	4	y
5	M	23	3.1	3.5	6	n	i	1	y

When loaded by SPSS, looks like:



student_survey.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : subj | 1 | Visible: 18 of 18

	subj	g	age	high_gpa	coll_gpa	dist_home	dist_res	tv	sport	news
1	1	m	32	2.2	3.5	0	5.00	3.0	5	0
2	2	f	23	2.1	3.5	1200	.30	15.0	7	5
3	3	f	27	3.3	3.0	1300	1.50	.0	4	3
4	4	f	35	3.5	3.2	1500	8.00	5.0	5	6
5	5	m	23	3.1	3.5	1600	10.00	6.0	6	3
6	6	m	39	3.5	3.5	350	3.00	4.0	5	7
7	7	m	24	3.6	3.7	0	.20	5.0	12	4
8	8	f	31	3.0	3.0	5000	1.50	5.0	3	3
9	9	m	34	3.0	3.0	5000	2.00	7.0	5	3
10	10	m	28	4.0	3.1	900	2.00	1.0	1	2
11	11	m	23	2.3	2.6	253	1.50	10.0	15	1
12	12	f	27	3.5	3.6	190	3.00	14.0	3	7
13	13	m	36	3.3	3.5	245	1.50	6.0	15	12
14	14	m	28	3.2	3.2	500	6.00	3.0	10	1
15	15	f	28	3.0	3.5	3500	1.00	4.0	3	1
16	16	f	25	3.8	3.3	210	10.00	7.0	6	1
17	17	f	41	4.0	3.0	1000	15.00	6.0	7	3
18	18	m	50	3.8	3.8	0	3.00	5.0	9	6
19	19	m	71	4.0	3.5	5000	3.00	6.0	12	2
20	20	f	28	3.0	3.8	120	1.00	25.0	0	0
21	21	f	26	3.7	3.7	8000	8.00	4.0	4	4
22	22	f	27	4.0	3.7	2	2.50	4.0	2	7
23	23	m	31	2.7	3.5	1700	5.00	7.0	7	2
24	24	f	23	3.7	3.7	2	2.00	7.0	4	2
25	25	m	23	3.2	3.8	450	4.00	.0	7	7
26	26	f	44	3.0	3.0	0	2.00	2.0	3	2
27	27	m	26	3.7	3.0	1000	3.00	8.0	2	7

Data View | Variable View

SPSS Processor is ready