

Stata as a tool for teaching basic survey analysis skills to social science students

Joanna Sikora
School of Sociology
Australian National University

SOCY2038 Introduction to Quantitative Research Methods

- No prerequisites – assumed Year 10 mathematics
- Compulsory in BA Criminology
- Sociology majors
- Political science majors

Content

Week 1 Week 2 Week 3 Week 4 Week 5 Week 6 Week 7 Week 8 Week 9 Week 10 Week 11 Week 12

| | | | | | | | | | | | |
|--|---|-----------|--|-----------|-----------------------------|--|--|--|--|--|--|
| Structure of quantitative analysis paper | Designing the layout of crosstabulations, correlation matrices, OLS tables and writing about them | | | | | | | | | | |
| Stata: basic preparation of data and analysis | Descriptive statistics | Crosstabs | Correlations Bivariate and Multivariate OLS regression | | | | | | | | |
| Survey design: theories, deductive reasoning, question and questionnaire design, types of samples | | | | | | | | | | | |
| | Elementary statistical concepts | Sampling | Inference | Crosstabs | Correlations OLS regression | | | | | | |

Writing and reading skills
 Statistical software skills
 Conceptual skills
 Statistical knowledge

What (some) students say

“Some of the content is very difficult for people who aren't mathematically minded. They can't grasp some concepts that quickly.”

Did the lecturer stimulate your interest in the subject?

“An impossible feat! :P “

“Numerous students who I have spoken to said this was the HARDEST course in the whole criminology degree...”

How to get publication ready tables in few steps without user-written add-ons e.g. tabout; estab; estout?

```
tab fam_ssexk gender, row
```

```
+-----+
| Key   |
+-----+
| frequency |
| row percentage |
+-----+
```

```

A same-sex couple |           Gender
  with children | 1. Female   2. Male |           Total
-----+-----+-----+
  1. yes a family |           295     165 |           460
                |           64.13   35.87 |           100.00
-----+-----+-----+
  2. no not a family |           247     302 |           549
                |           44.99   55.01 |           100.00
-----+-----+-----+
                Total |           542     467 |           1,009

```

Table 1. In your opinion is a same-sex couple with children is a family?

| | Women | Men |
|------------------|------------|-------------|
| Yes, a family | 64% | 36% |
| No, not a family | 45% | 55% |
| Total | 54% | 46% |
| N | 542 | 467% |

How to get Stata to display 2 decimals in a correlation matrix?

```
correlate fnbzpayq fnprospq fnstudyq fnpaywkq fnineqq fe100svq felsavq feyrkidq
(obs=1,111)
```

| | fnbzpayq | fnprospq | fnstudyq | fnpaywkq | fnineqq | fe100svq | felsavq | feyrkidq |
|----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| fnbzpayq | 1.0000 | | | | | | | |
| fnprospq | 0.4906 | 1.0000 | | | | | | |
| fnstudyq | 0.2721 | 0.2984 | 1.0000 | | | | | |
| fnpaywkq | 0.3581 | 0.3181 | 0.3662 | 1.0000 | | | | |
| fnineqq | 0.3506 | 0.5089 | 0.2480 | 0.3319 | 1.0000 | | | |
| fe100svq | 0.0985 | 0.1258 | 0.0604 | 0.0614 | 0.0754 | 1.0000 | | |
| felsavq | 0.0898 | 0.1264 | 0.0466 | 0.0687 | 0.0862 | 0.8493 | 1.0000 | |
| feyrkidq | 0.0882 | 0.1206 | 0.0622 | 0.0651 | 0.1010 | 0.8435 | 0.7559 | 1.0000 |

How to get students from Stata output to this in as few steps as possible?

Table 2. Correlation coefficients, means and standard deviations for attitudes to functionality of inequality, age, education and earnings in ISSSA 2001

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. |
|-------------------------------------|--------|--------|--------|--------|---------|--------|--------|
| 1. Entrepreneurs should earn a lot | 1.00 | | | | | | |
| 2. Income differences are necessary | 0.49** | 1.00 | | | | | |
| 3. High salaries for prosperity | 0.37** | 0.33** | 1.00 | | | | |
| 4. Inequality for economic progress | 0.36** | 0.51** | 0.34** | 1.00 | | | |
| 5. Age in years | 0.11** | 0.08** | 0.14** | 0.15** | 1.00 | | |
| 6. Education in years | 0.05 | -0.05 | -0.06* | -0.02 | -0.31 | 1.00 | |
| 7. Annual earnings | 0.10** | 0.06* | 0.02 | 0.05 | -0.24** | 0.29** | 1.00 |
| Mean | 53.9 | 42.2 | 69.0 | 49.9 | 49.2 | 12.1 | 23,695 |
| Standard deviation | 24.6 | 23.5 | 20.2 | 23.9 | 15.4 | 2.9 | 31,744 |
| Min | 0 | 0 | 0 | 0 | 20 | 0 | 0 |
| Max | 100 | 100 | 100 | 100 | 89 | 18 | 288,39 |
| N | 53.9 | 42.2 | 69.0 | 49.9 | 49.2 | 12.1 | 23,695 |

Source: ISSSA 2001

**Coefficient significantly different from zero at $p = 0.01$

* Coefficient significantly different from zero at $p = 0.05$

How to get them from raw regression output to this?

Table 2. Support for government ownership of industry & commerce in Australia in 1995. Ordinary Least Squares Regression

| | Support for government-owned industry and commerce | | |
|-------------------------|--|--------------------------------|--|
| | Unstandardized coefficient b | Standard error SE | Standardized coefficient Beta |
| Age in years | -0.11** | 0.02 | -0.08 |
| Constant | 55.96** | 2.97 | |
| Adjusted R ² | 0.10 | | |
| (N of cases) | 2368 | | |

Source: ISEA 1995

Note *Coefficient significantly different from zero at $p = 0.05$
 **Coefficient significantly different from zero at $p = 0.01$