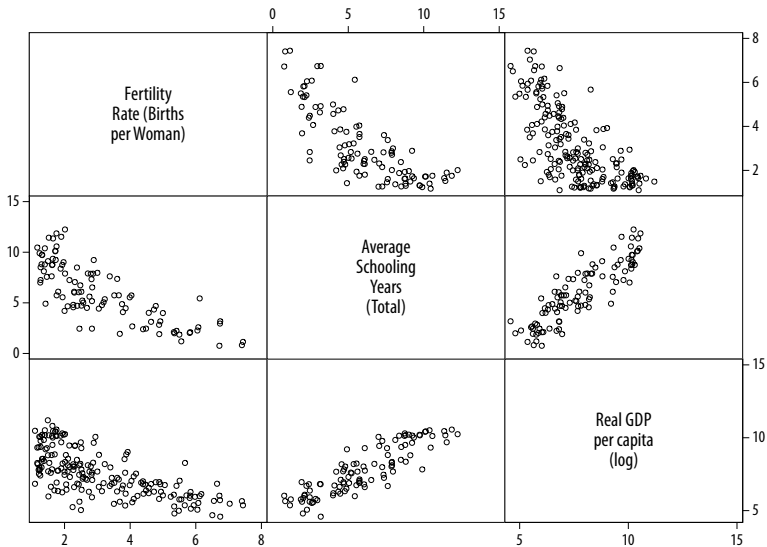


# Regression (II)

- 1 Multiple linear regression
- 2 Draft No. 2



# Multiple linear regression

```
. reg births schooling log_gdpc
```

Source	SS	df	MS
Model	150.301883	2	75.1509417
Residual	70.475313	83	.849100157
Total	220.777196	85	2.59737878

Number of obs = 86  
 F( 2, 83) = 88.51  
 Prob > F = 0.0000  
 R-squared = 0.6808  
 Adj R-squared = 0.6731  
 Root MSE = .92147

births	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
schooling	-.1976117	.0724595	-2.73	0.008	-.3417306	-.0534927
log_gdpc	-.4703416	.1324501	-3.55	0.001	-.7337796	-.2069036
_cons	7.950304	.6861182	11.59	0.000	6.585642	9.314965

# Multiple linear regression

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

## Partial derivatives

Each coefficient is calculated by **holding all others constant**.

## Least squares

The model is still optimized by minimizing the squared error terms.

## Sanity check

The model is still assuming *linear, additive* relationships.

```
reg births schooling log_gdpc, beta
```

Each variable can be normalized to fit  $\mathcal{D} \sim \mathcal{N}(0, 1)$ , so that their **standardized coefficients** have comparable standard deviation units:

births	Coef.	Std. Err.	t	P> t	Beta
schooling	-.1976117	.0724595	-2.73	0.008	-.3686479
log_gdpc	-.4703416	.1324501	-3.55	0.001	-.4800156
_cons	7.950304	.6861182	11.59	0.000	.

*(identical output for overall model fit omitted)*

## Sanity check

Interpret unstandardized coefficients; use standardization only for model comparisons.

## reg births schooling i.region

Categorical variables can be used as **dummies**, i.e. binary recodes of each category that are tested against a **reference category** to provide regression coefficients for the net effect of each category:

births	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
schooling	-.0415563	.0639718	-0.65	0.518	-.1688888	.0857763
log_gdpc	-.742187	.1380037	-5.38	0.000	-1.016876	-.4674975
region						
2	-.6523485	.5803126	-1.12	0.264	-1.807432	.5027349
3	.3682404	.254364	1.45	0.152	-.1380585	.8745393
4	1.411177	.2486027	5.68	0.000	.9163457	1.906008
5	1.167491	.337383	3.46	0.001	.4959471	1.839035
_cons	8.315004	.8006456	10.39	0.000	6.721359	9.908649

*(identical output for overall model fit omitted)*

# Where we are now

## Univariate statistics

- Introduction
- Dataset
- Variables

Assignment No. 1

*corrected* }  
*revised* }  
*appended* }

## Bivariate statistics

- Associations
- Correlations
- Simple OLS

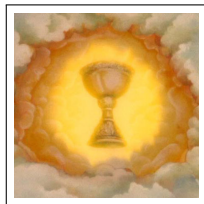
Assignment No. 2



## Statistical modelling

- Regressions
- Diagnostics
- Conclusion

Final paper



# Essential instructions

## Revise Draft No. 1

- go through corrections
- remove technical content
- **rewrite until concision**

Pay attention to paragraph limits and scientific style (esp. sources).

## Explore associations

- between DV and IVs (covariates, controls), or between two IVs
- with graphs and then with significance tests

Write up **substantive results** as sentences; cite significance tests and other statistics in brackets, e.g. ( $\rho = .7, p < .05$ ).



# Paper template, structure and style

LYNN WHITE University of Nebraska—Lincoln

Insert the summary statistics of your continuous variables and frequencies of your categorical variables in the appendix. Use the standard format shown at the end of this template, and mention the appendix in your text. All descriptive statistics should be in a single table with a level of precision of one decimal for the mean and standard deviation. Other figures like observations and relative frequencies should be rounded to integer values.

## Results

At each stage of your paper, write this section by following the specific instruction sets below. Keep the language simple and refer to significance tests only between brackets using *p*-values (e.g. Chi-squared test significant at  $p < .05$ , *t*-test significant at  $p < .01$ ).

### (First draft: Descriptive statistics)

Write up a maximum of four paragraphs that report on the descriptive statistics of your variables. Explain why you include each of them in your research design, and compare the mean with 95% confidence intervals of your dependent variable across categorical independent variables.<sup>4</sup> A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here.

### (Second draft: Association tests)

Write up a maximum of four paragraphs that report on the results of statistical tests of your hypotheses. Include graphs only for striking and significant relationships. If you feel confident enough with modeling, include preliminary findings from simple linear or logistic regression. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here. A wonderful paragraph goes here.

### (optional) Appendix Table & Figure, Correlations

If your research relies heavily on continuous variables, insert their scatterplot and correlation matrices in the appendix. Use the standard table format shown at the end of this template, and mention the appendix in your text.

### (optional) Appendix, Association

If your research relies heavily on categorical variables, express a visually striking relationship between two of your variables with a spineplot. Add a caption, a Chi-squared test result, and a citation as well as an interpretation in your text.

### (Final paper: Regression analysis)

Write up a maximum of four paragraphs that report on regression output to model the relationships between your variables. Take unstandardised and standardised "beta" coefficients

<sup>4</sup> In your code, you can include some graphs of the

## Writes of Passage: Writing an Empirical Journal Article

*This article provides advice about preparing research reports for submission to professional journals in general and Journal of Marriage and Family in particular. In addition to working through all the major parts of a research paper, I provide some general advice about writing, editing, and revising. The article is intended to help new professionals improve the quality of their journal submissions and the likelihood of successful publication.*

Writing research articles for professional journals is an art requiring good research skills, a clear sense of problem, and strong writing and editing skills. Assuming that years of graduate school have provided good research skills, I focus on the other requirements of writing a research article. My advice reflects the issues I most often raise when I review articles and 30 years of experience writing (and revising) research articles. I review guidelines for the major sections of the typical empirical research report and conclude with some suggestions about writing professionally. The emphasis is on writing for *Journal of Marriage and Family* (JMF), but the general principles apply across journals and substantive areas.

### WORKING THROUGH A RESEARCH PAPER

The format for a research paper is not set in stone. Each research problem is different, and

Department of Sociology, University of Nebraska—Lincoln, Lincoln, NE 68583-0324 (lwhite@unl.edu).

Key Words: research, theory, writing.

the organization of the paper will depend on whether it is exploratory research rather than theory testing. In addition, authors have some latitude in developing a personal style. Generally, however, each article needs an introduction, a literature review, a statement of the problem, description of method, results, and conclusion. The organization of the piece, the titles of various sections, and the relative weight of these sections vary from paper to paper and from journal to journal, but some general guidelines apply to reports of qualitative and quantitative research.

### Abstract

An abstract should summarize your study. In a few short sentences, it should state the research hypothesis, the sample, sample size, data used, and the findings. A starting sentence such as "Using data from a national sample of *n* women interviewed by telephone in 2002, we examine the relationship between *x* and *y*" will allow you to squeeze a lot of information into a few words. In a bare-bones fashion, without hyperbole or exaggeration, state the findings of the study. Examine prior issues of your target journal for abstract style and be sure to comply with the maximum length specified by the journal (120 words for JMF).

### Introduction

The introduction is critical to capturing the reader's attention and setting the tone for the paper. In approximately a single page, it should specify the research question, the data to be used, and the strengths of the design, and it

## The stab command

Syntax: `stab using Briatte_Petev_1, replace...`

- `su()` summarizes continuous variables
- `fre()` summarizes categorical variables
- `by()` creates multiple tables for comparison

Add the `corr` option to also export a correlation matrix.

`use datasets/nhis2009, clear`

```
stab using Briatte_Petev_1, replace ///  
    sum(age weight height) corr ///  
    fre(sex uninsured marstat health) ///  
    by(regionbr)
```

## Stata video tutorials

# STATa Tutorials: Using Do-files

# STATA Tutorials: Binary Logistic Regression

Methodology LSE Subscribe 53 videos

```

1. open c:\stata\bin\data\mccormick.dta
2. use "c:\stata\bin\data\mccormick.dta", clear
3. logit pctvoice gender age voice
4. logit
Iteration 0:   Log Likelihood = -307.22351
Iteration 1:   Log Likelihood = -95.09603
Iteration 2:   Log Likelihood = -94.89478
Iteration 3:   Log Likelihood = -94.87668
Iteration 4:   Log Likelihood = -94.87668

logistic regression

Log Likelihood = -94.87668

+-----+-----+
|                |      |      |      |      |
| pctvoice        |  Coef. | Std. Err. | z | P>|
|-----+-----+-----+-----+
| gender          | -0.097312 | 1.710557 | -0.23 | 0.819
| age             | 0.0288122 | 0.011035 | 2.15 | 0.033
| voice           | 1.196683 | 0.07486 | 16.12 | 0.000
| _cons           | -2.9336 | 1.61337 | -1.82 | 0.070
+-----+-----+-----+-----+

+-----+
| logit: |
+-----+

```

Source: LSE Methodology Institute, 2012.

# Thanks for your attention

## Project

- Name your paper and do-file like **Briatte\_Petev\_2**
- Make sure to print your paper to a slick **PDF**

## Readings

- *Stata Guide*, Sec. 11, 13 and 15

## Practice

- Replicate do-file
- **Use its structure for Draft No. 2**