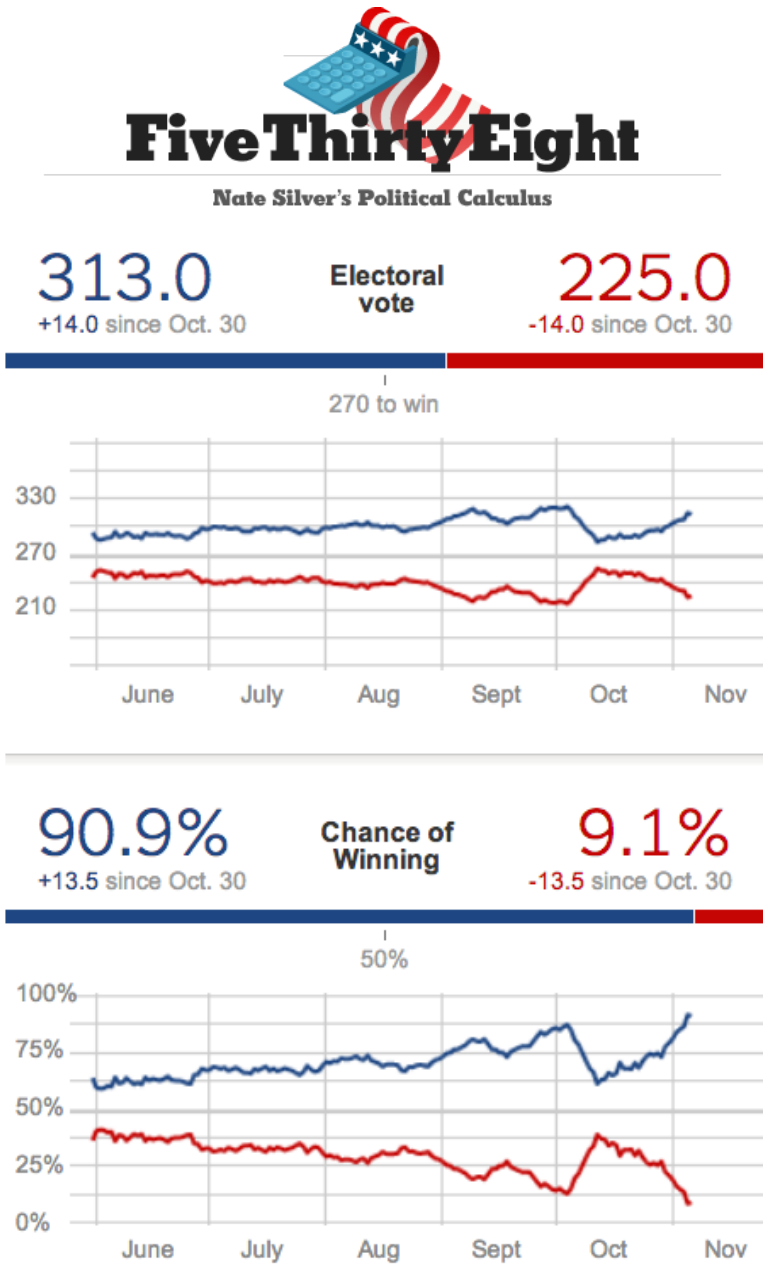
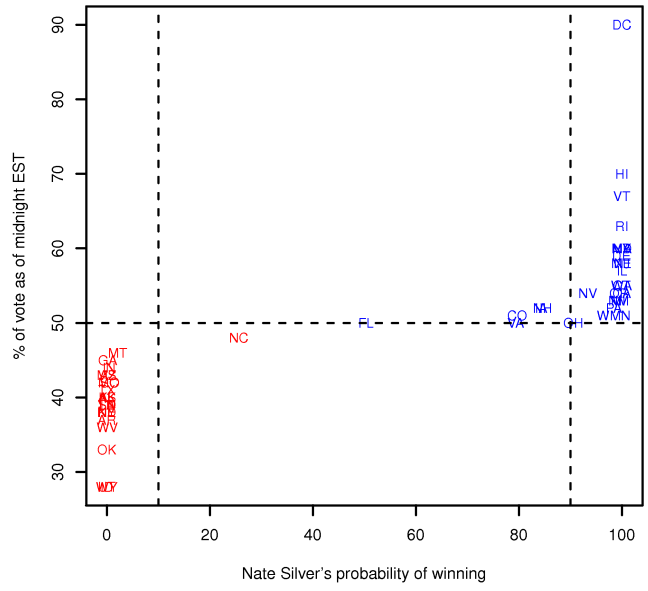
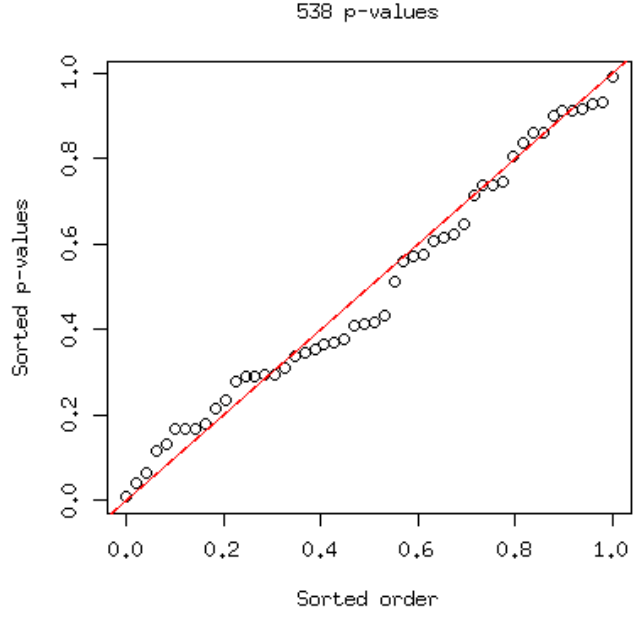
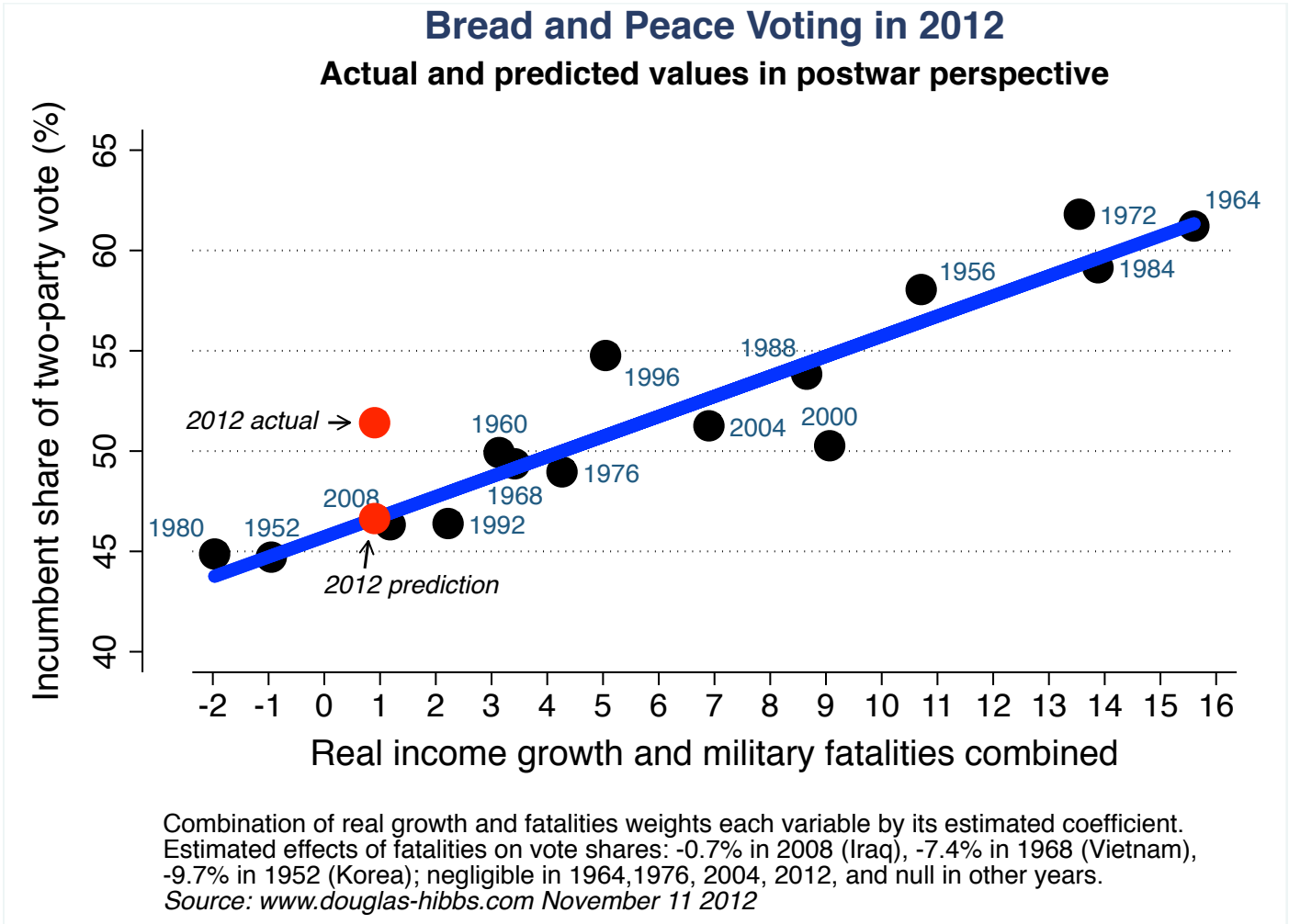
Simple regression

1. A simple linear model
2. Ordinary Least Squares (OLS)
3. Regression output
4. Draft No. 2


FiveThirtyEight
Nate Silver's Political Calculus
313.0
+14.0 since Oct. 30
Electoral vote
225.0
-14.0 since Oct. 30
270 to win
330
270
210
June
July
Aug
Sept
Oct
Nov
90.9%
+13.5 since Oct. 30
Chance of Winning
9.1%
-13.5 since Oct. 30
50%
100%
75%
50%
25%
0%
June
July
Aug
Sept
Oct
Nov


% of vote as of midnight EST
Nate Silver's probability of winning
DC
HI
VT
RI
NV
CO
NH
VA
OH
WI
MN
FL
NC
MT
OK
0
20
40
60
80
100
30
40
50
60
70
80
90


538 p-values
Sorted p-values
Sorted order
0.0
0.2
0.4
0.6
0.8
1.0



Combination of real growth and fatalities weights each variable by its estimated coefficient. Estimated effects of fatalities on vote shares: -0.7% in 2008 (Iraq), -7.4% in 1968 (Vietnam), -9.7% in 1952 (Korea); negligible in 1964,1976, 2004, 2012, and null in other years.
Source: www.douglas-hibbs.com November 11 2012

To what extent can trust in government be predicted from variations in economic growth?

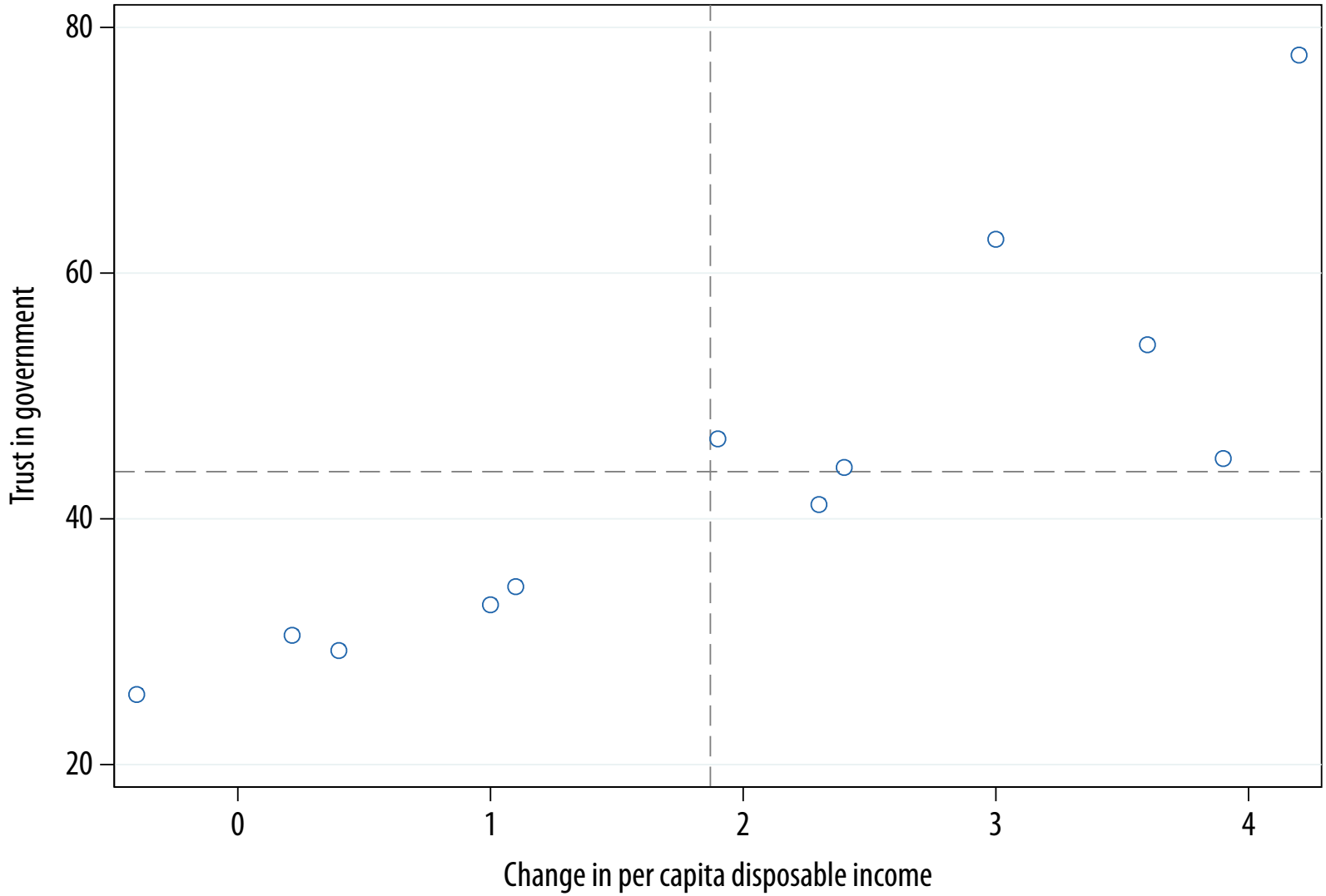
DV: Trust in government

"Just about always/Most of the time" (American National Election Studies)

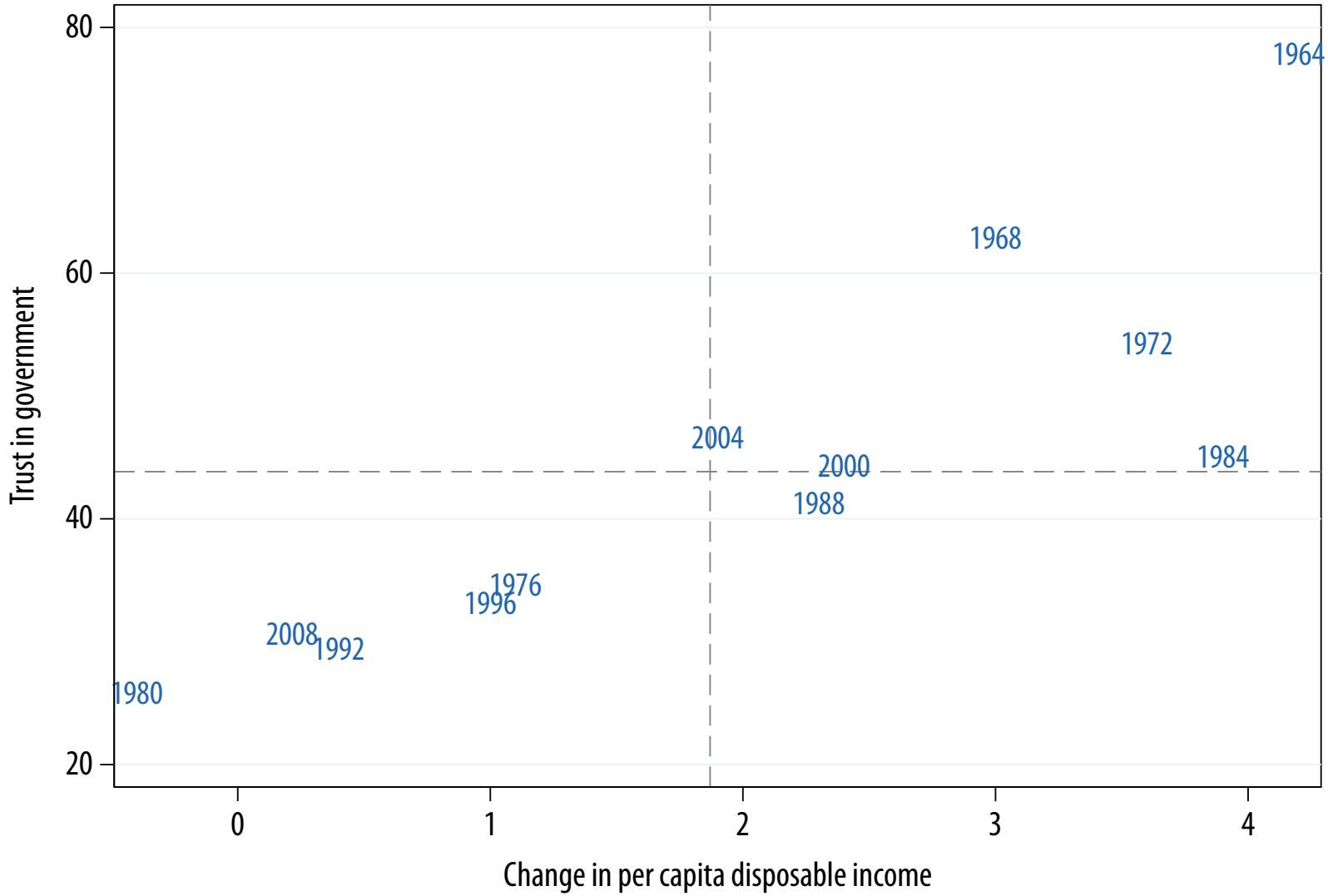
IV: Economic performance

Change in per capita disposable income (Bureau of Economic Analysis)

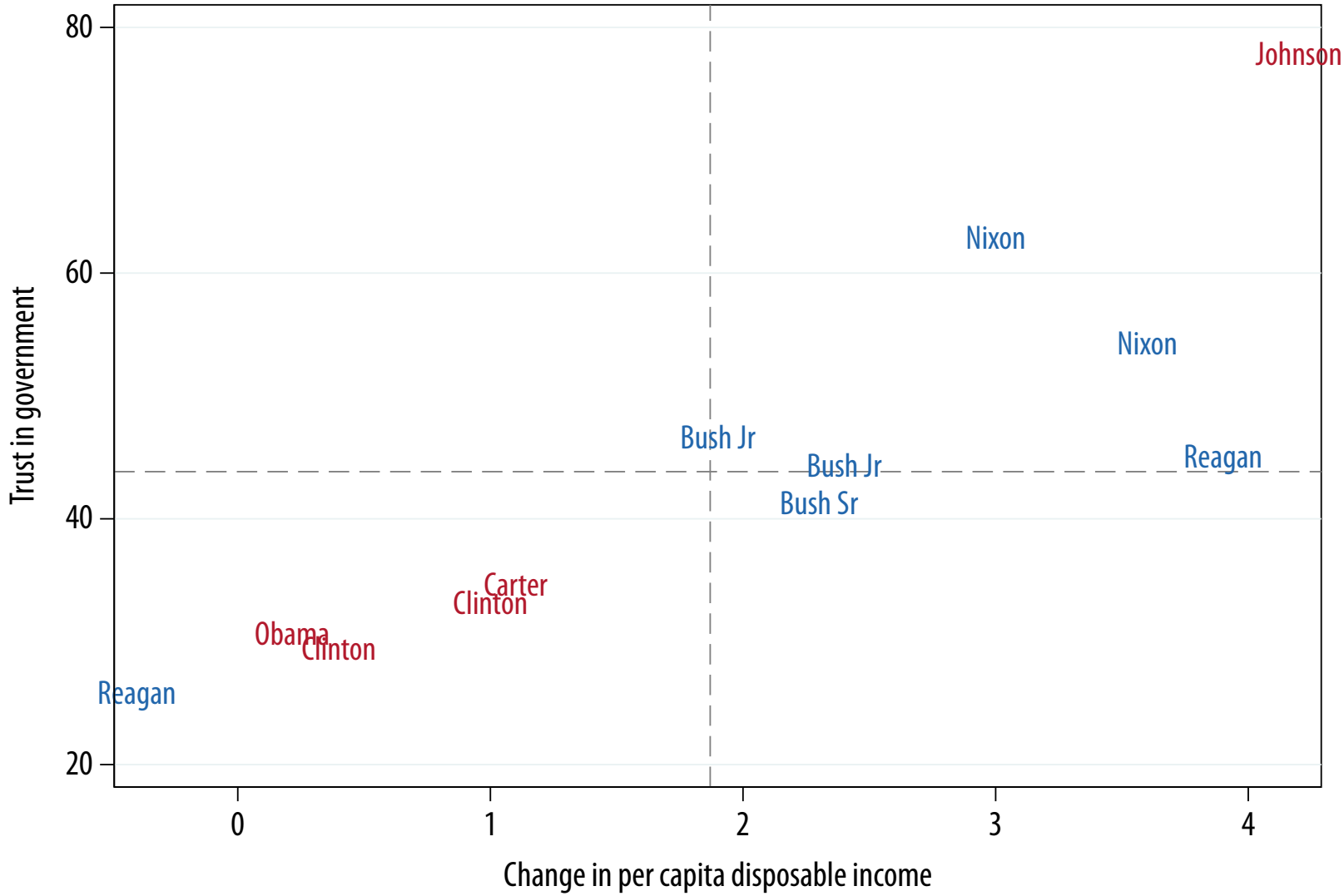
Example and data provided by John Sides.



Dashed lines at averages. Pearson correlation $\rho = .86$ significant at $p < .01$.



Dashed lines at averages. Pearson correlation $\rho = .86$ significant at $p < .01$.



Dashed lines at averages. Pearson correlation $\rho = .86$ significant at $p < .01$.

Simple linear regression

Equations

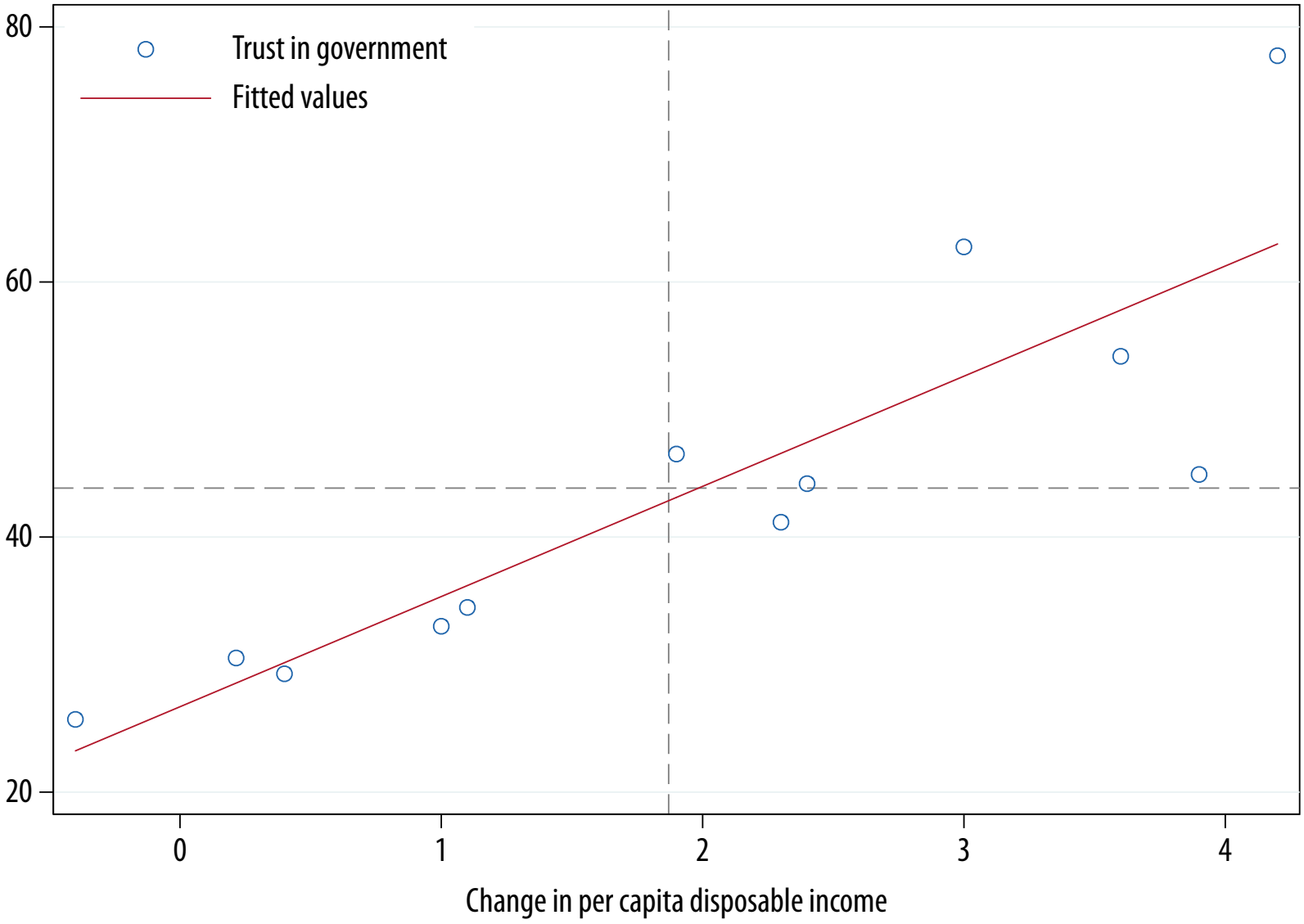
$$Y = \alpha + \beta X + \epsilon \quad \hat{Y} = \hat{\alpha} + \hat{\beta} X + \hat{\epsilon} \quad \epsilon = Y - \hat{Y}$$

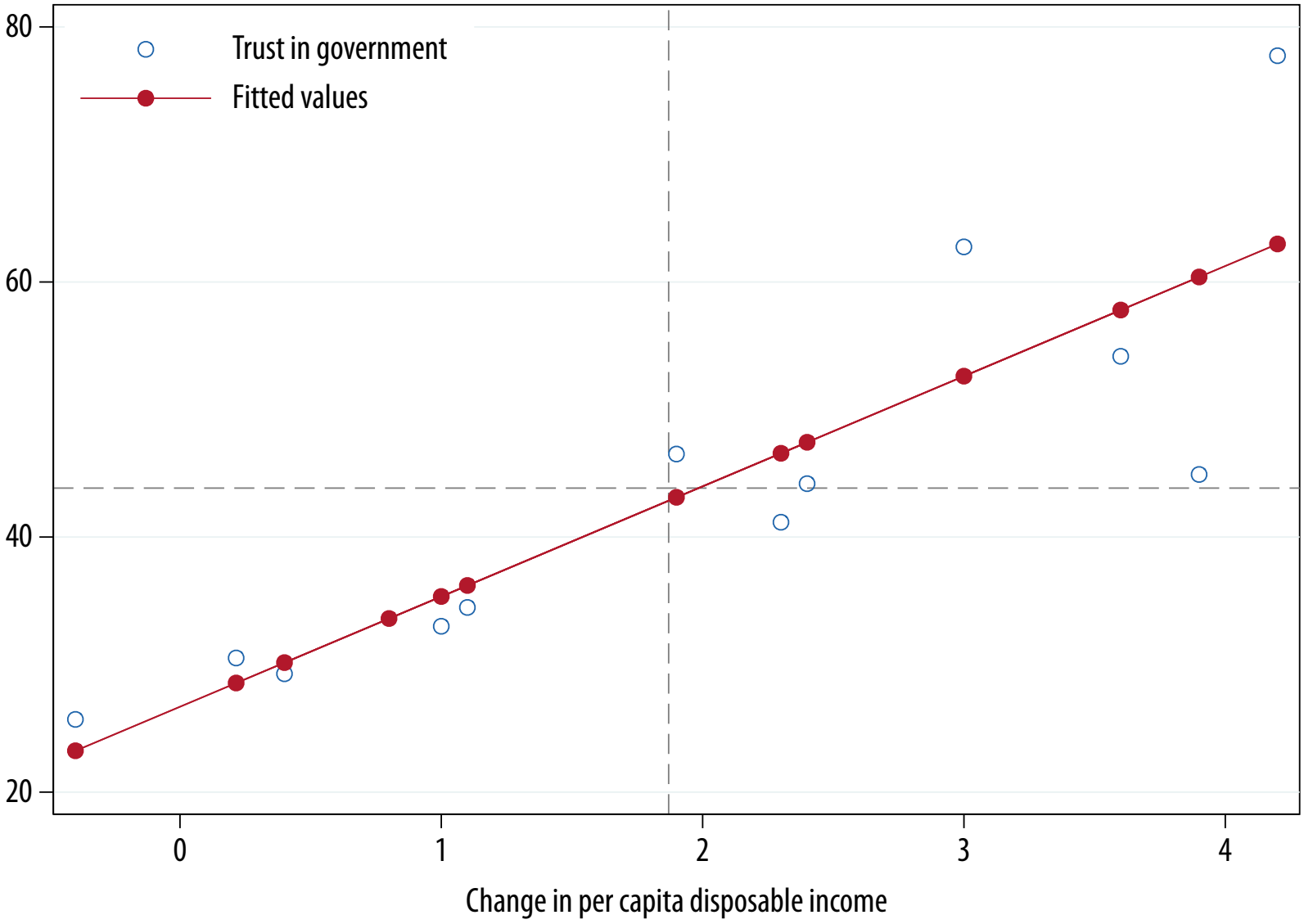
Parameters

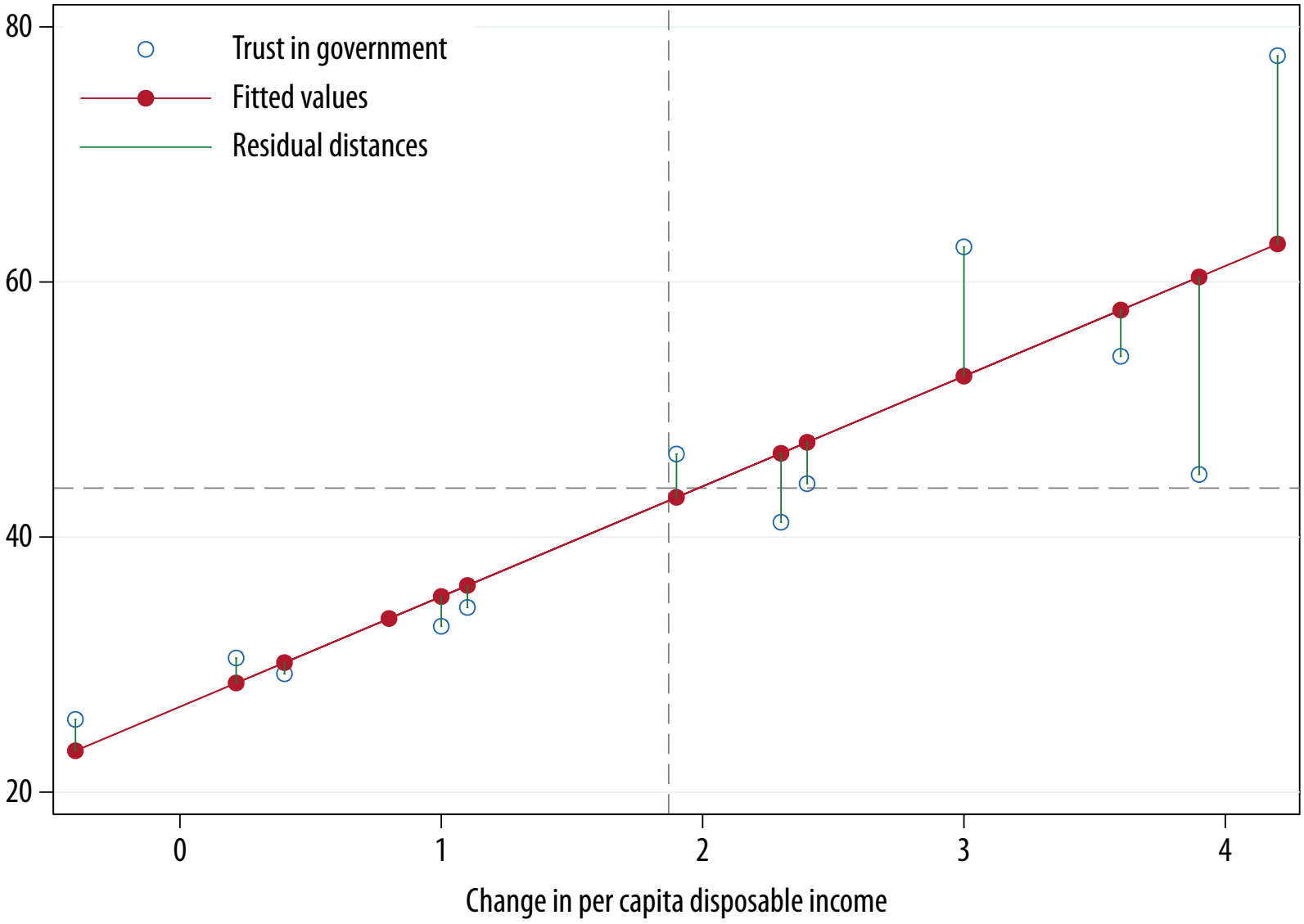
- Y is the dependent variable and $\hat{Y}$ its predicted value
- X is the independent variable used as a predictor of Y
- α is the constant (intercept)
- β is the regression coefficient (slope)
- ϵ is the error term (residuals)

Warning

The model assumes a *linear*, *additive* relationship.


Trust in government
Fitted values
80
60
40
20
0
1
2
3
4
Change in per capita disposable income


Trust in government
Fitted values
80
60
40
20
0
1
2
3
4
Change in per capita disposable income


Trust in government
Fitted values
Residual distances
80
60
40
20
0
1
2
3
4
Change in per capita disposable income

Ordinary Least Squares (OLS)

Error term

In a simple linear model $Y = \alpha + \beta X + \epsilon$, the regression coefficient β is calculated as to minimize the residual sum of squares

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \epsilon^2$$

where $Y_i - \hat{Y}_i$ is the residual (or error term) of each observation.

Parameter estimation

$$\beta = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}_X} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \qquad \alpha = \bar{Y} - \beta \bar{X}$$

reg y x

```
. regress trust income
```

Source	SS	df	MS
Model	1908.80221	1	1908.80221
Residual	643.906248	10	64.3906248
Total	2552.70846	11	232.064405

Number of obs	12
F(1, 10)	29.64
Prob > F	0.0003
R-squared	0.7478
Adj R-squared	0.7225
Root MSE	8.0244

trust	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	8.639373	1.586767	5.44	0.000	5.103836	12.17491
_cons	26.69501	3.888016	6.87	0.000	18.03197	35.35805

Top left: ANOVA table. Top right: model fit.
Bottom: regression coefficients.

Interpretation of fit

Number of observations N, significance test $H_0 : \beta = 0$, coefficient of determination R^2, root mean square error (RMSE).

```
. regress trust income

      Source |       SS       df       MS
-------------+------------------------------
       Model |  1908.80221     1  1908.80221
    Residual |  643.906248    10  64.3906248
-------------+------------------------------
       Total |  2552.70846    11  232.064405

                                      Number of obs =        12
                                      F(  1,    10) =     29.64
                                      Prob > F      =    0.0003
                                      R-squared     =    0.7478
                                      Adj R-squared =    0.7225
                                      Root MSE      =    8.0244

------------------------------------------------------------------------------
       trust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   8.639373   1.586767     5.44   0.000     5.103836    12.17491
       _cons |   26.69501   3.888016     6.87   0.000     18.03197    35.35805
------------------------------------------------------------------------------
```

Goodness of fit

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} = \frac{\text{residual sum of squares}}{\text{total sum of squares}}$$

As the fit improves, $RSS \to 0$ and $R^2 \to 1$.

```
Number of obs =        12
F(  1,    10) =     29.64
Prob > F      =    0.0003
R-squared     =    0.7478
Adj R-squared =    0.7225
Root MSE      =    8.0244
```

Sanity check

Focus on getting N and the RMSE right.

Interpretation of regression coefficients

A regression coefficient estimates the variation in Y predicted by a change in one unit of X (recall that $Y = \alpha + \beta X + \epsilon$)

```
. regress trust income

      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  1,    10) =   29.64
       Model |  1908.80221     1  1908.80221            Prob > F      =  0.0003
    Residual |  643.906248    10  64.3906248            R-squared     =  0.7478
-------------+------------------------------           Adj R-squared =  0.7225
       Total |  2552.70846    11  232.064405            Root MSE      =  8.0244

------------------------------------------------------------------------------
       trust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   8.639373   1.586767     5.44   0.000     5.103836    12.17491
       _cons |   26.69501   3.888016     6.87   0.000     18.03197    35.35805
------------------------------------------------------------------------------
```

trust	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	8.639373	1.586767	5.44	0.000	5.103836	12.17491
_cons	26.69501	3.888016	6.87	0.000	18.03197	35.35805

- The coefficient is the slope β of the regression line and the constant is its intercept, the coordinate of origin $\alpha = \hat{Y}_{X=0}$.
- The standard error, t-value and p-value test whether the coefficient is significantly different from 0.

Logarithmic coefficients: see UCLA mini-guide

Linear-linear relationships: $Y = \alpha + \beta X$

An increase of one unit of X is associated with an increase of β_1 units of Y.

Log-linear relationships: $\ln Y = \alpha + \beta X$

An increase of one unit of X is associated with a $100 \times \beta_1\%$ increase in Y (true effect: $Y \times \exp(\beta_1)$).

Linear-log relationships: $Y = \alpha + \beta \ln X$

A 1% increase in X is associated with a $0.01 \times \beta_1$ unit increase in Y (e.g. $\beta_1 \times \texttt{log(1.15)}$ for +15% in X).

Log-log relationships: $\ln Y = \alpha + \beta \ln X$

A 1% increase in X is associated with a $\beta_1\%$ increase in Y.

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. $(\rho = .7, p < .05)$.

Structure and style

Paper template

Group Names

Statistical Reasoning and Quantitative Methods, Fall 2012
Research Paper

Why Titles Matter: Evidence from Contemporary European Academia, 2012

Draft 1 · Draft 2 · Final Version · Date

Abstract

In one paragraph, write up a short summary of your work when you are done with the analysis. Make sure to mention the keywords and main results of your research. A few lines are enough. Also make sure to use scientific style throughout your paper. 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. A wonderful paragraph goes here. A wonderful paragraph goes here.

Feedback

Write a short numbered list of issues that you want to discuss in more detail about your research. Include references to your code by mentioning the line numbers of your do-file. All sorts of questions and comments are also very welcome! 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. A wonderful paragraph goes here. A wonderful paragraph goes here.

1

White (2005)

LYNN WHITE *University of Nebraska—Lincoln*

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 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

Department of Sociology, University of Nebraska—Lincoln, Lincoln, NE 68588-0324 (lwhite3@unl.edu).

Key Words: research, theory, writing.

Coursework

Project — *Stata Guide*, Sec. 13–15

- Correct and improve Draft No. 1
- Finalize association tests and interpretations
- Name your paper (PDF) and do-file like Briatte_Petev_2

Readings

- *Making History Count*, ch. 4 (simple regression)
- *Stata Guide*, Sec. 10 (association)
- *Stata Guide*, Sec. 11 (correlation and simple regression)

Enjoy your day and see you soon!