Simple regression

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
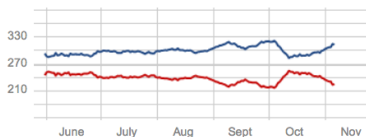
90.9%
+13.5 since Oct. 30

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

DC

HI
VT
RI

NV

NC

FL

VT

OK

WV

Nate Silver's probability of winning

0   20   40   60   80   100

538 p-values

Sorted p-values

Sorted order

0.0   0.2   0.4   0.6   0.8   1.0
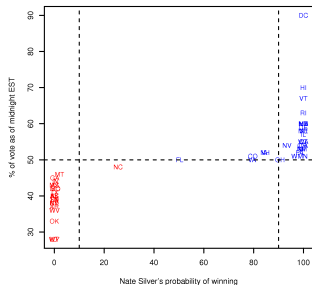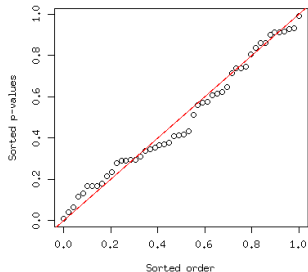
Bread and Peace Voting in 2012

Actual and predicted values in postwar perspective

Incumbent share of two-party vote (%) — vertical axis

Real income growth and military fatalities combined — horizontal axis

2012 actual →

2012 prediction

Data point labels: 1980, 1952, 2008, 1960, 1968, 1992, 1976, 1996, 2004, 1988, 2000, 1956, 1972, 1984, 1964
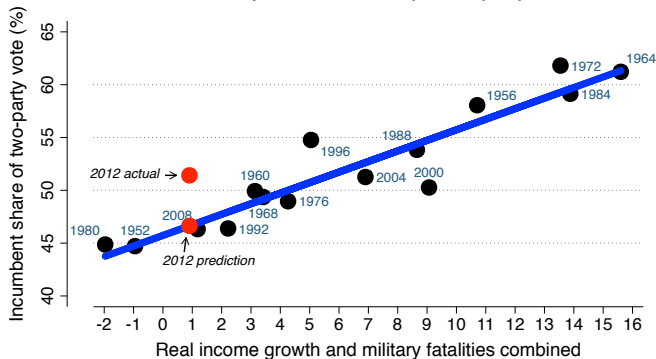
Combination of real growth and fatalities weights each variable by its estimated coefficient.
Estimated effects of fatalities on vote shares: -0.7% in 2008 (Iraq), -7.4% in 1968 (Vietnam),
-9.7% in 1952 (Korea); negligible in 1964,1976, 2004, 2012, and null in other years.
Source: www.douglas-hibbs.com November 11 2012

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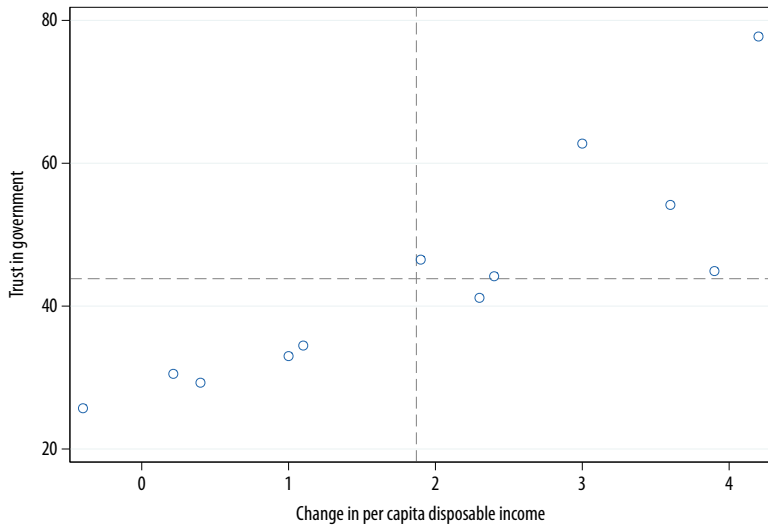
DV: Trust in government

"Just about always/Most of the time" (American National Election Studies)
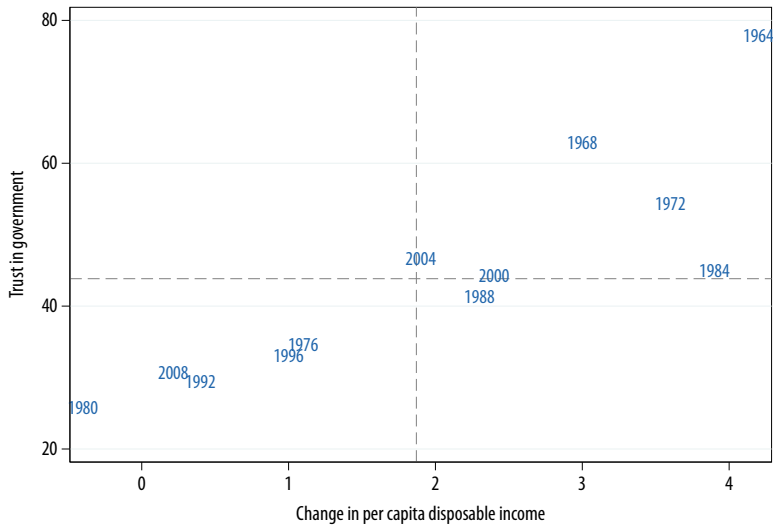
IV: Economic performance

Change in per capita disposable income (Bureau of Economic Analysis)
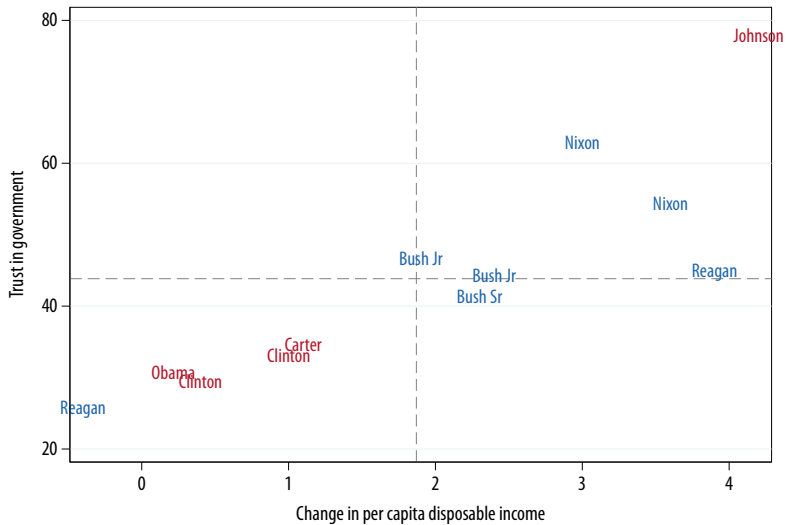
Example and data provided by John Sides.

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 $p < .01$.
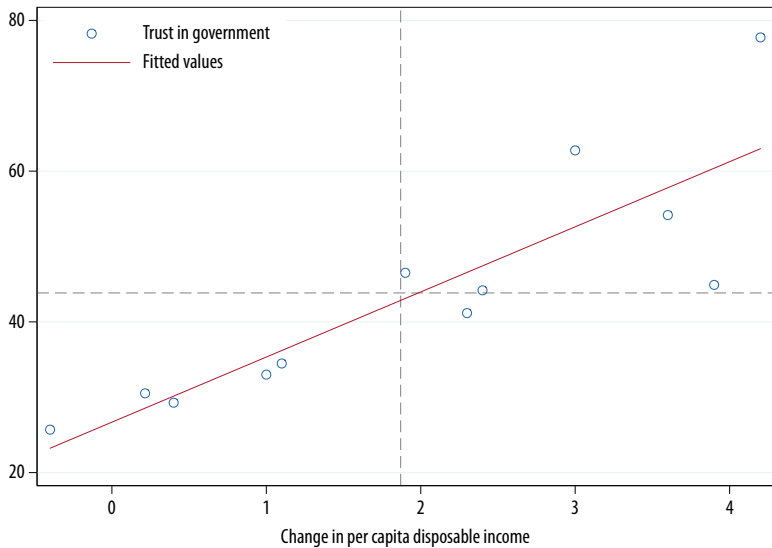
Simple linear regression

Equations

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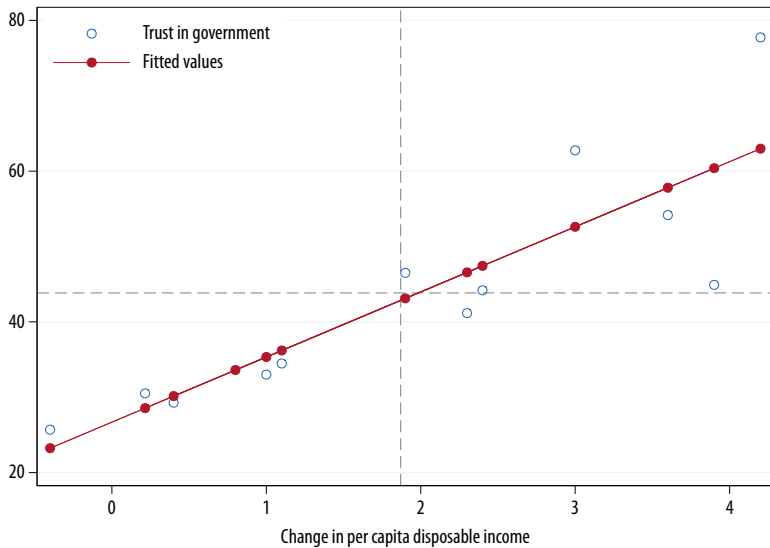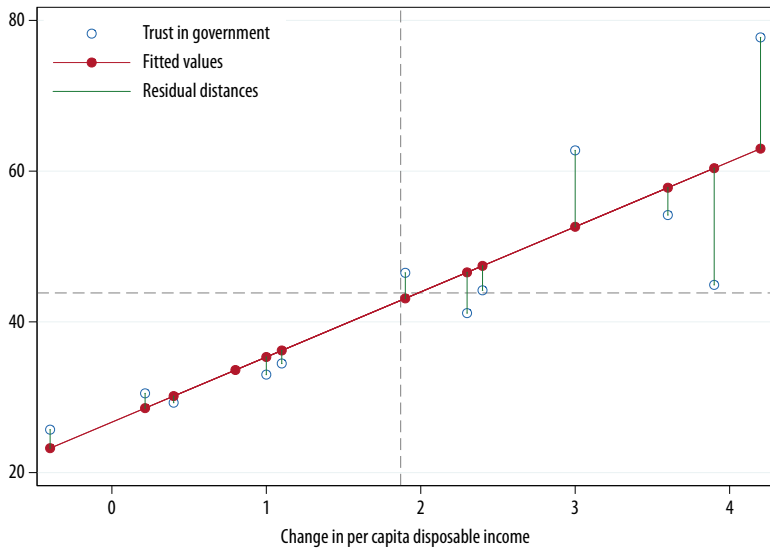
Parameters

- 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.

Change in per capita disposable income

Change in per capita disposable income

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

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

```
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).

Goodness of fit

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

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

Sanity check

Focus on getting N and the RMSE right.

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

Interpretation of regression coefficients

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

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.

Where we are now

Univariate statistics	**Bivariate statistics**	**Statistical modelling**

Univariate statistics

- Introduction
- Dataset
- Variables

Assignment No. 1

$$\left.\begin{array}{r} corrected \\ revised \\ appended \end{array}\right\}$$

Bivariate statistics

- Associations
- Correlations
- Simple OLS

Assignment No. 2

Statistical modelling

- Regressions
- Diagnostics
- Conclusion

Final paper
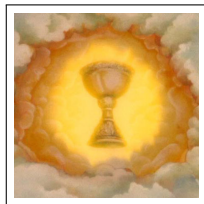
Essential instructions

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

White (2005)

Coursework

- 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!