Multiple Regression

Review from Lecture 3

Review questions from syllabus
- Questions about underlying theory
- Math and notation handout from first class

Review resampling in simple regression
- Regression diagnostics
- Smoothing methods (modern regr)

Methods of resampling – which to use?

Today

Multiple regression
- Leverage, influence, and diagnostic plots
- Collinearity

Resampling in multiple regression
- Ratios of coefficients
- Differences of two coefficients
- Robust regression

Yes! More t-shirts!

Locating a Maximum

Polynomial regression

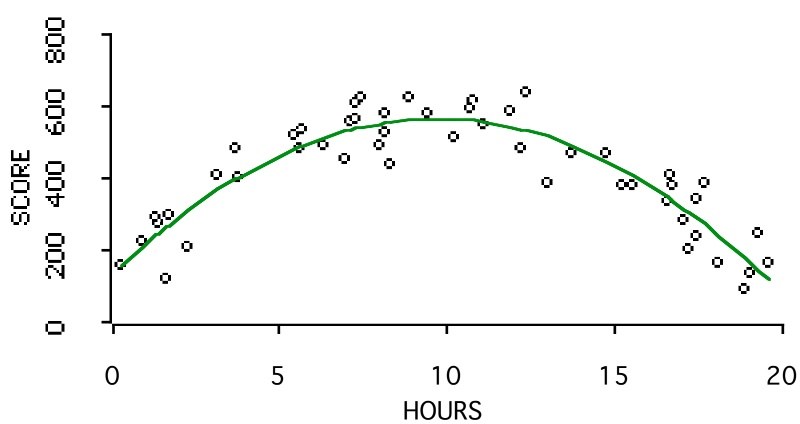
Why bootstrap in least squares regression?

Question
What amount of preparation (in hours) maximizes the average test score?

Quadmax.dat
Constructed data for this example

Results of fitting a quadratic

Scatterplot with added quadratic fit (or smooth)



Least squares regression of the model

$$Y = a + b\,X + c\,X^2 + \square$$

gives the estimates

$a = 136$ $b = 89.2$ $c = -4.60$

So, where's the maximum and what's its CI?

Inference for maximum

Position of the peak

The maximum occurs where the derivative of the fit is zero. Write the fit as

$$f(x) = a + b\,x + c\,x^2$$

and then take the derivative,

$$f'(x) = b + 2cx$$

Solving $f'(x^*) = 0$ for x^* gives

$$x^* = -\frac{b}{2c} \approx -89.2/(2)(-4.6) = 9.7$$

Questions

- What is the standard error of x^*?
- What is the 95% confidence interval?
- Is the estimate biased?

Bootstrap Results for Max Location

Standard error and confidence intervals (B=2000)

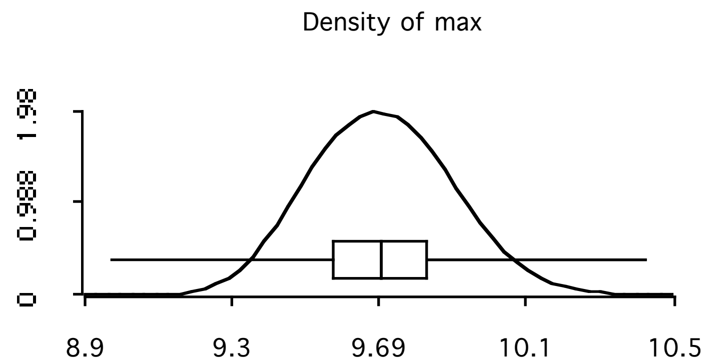
Standard error is slightly smaller with fixed resampling (no surprise)…

Observation resampling	SE* = 0.185
Fixed resampling	SE* = 0.174

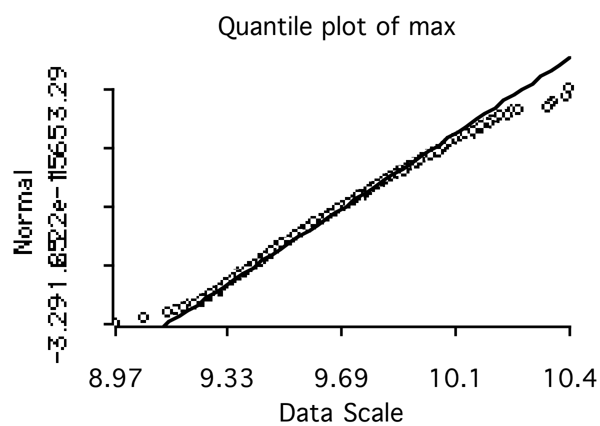
Both 95% intervals are [9.4, 10]

Bootstrap distribution

The kernel looks pretty normal.



Quantile plot reveals a lack of normality
- deviate from normality in the extremes
- “heavy tails” often occur with a ratio



Review - Resampling in Regression

Two approaches

Resample observations (X_i varies).

Resample residuals (X_i fixed).

Extended comparison

	Observations	Residuals
Equation-dependent	No	Yes
Assumption-dependent	Some	More
Preserves X values	No	Yes
Maintains (X,Y) assoc	Yes	No
Conditional inference	No	Yes
Agrees with usual SE	Maybe	Yes
Computing speed	Fast	Faster

Multiple Regression Model

Model

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \epsilon$$

where 1. Observations are independence
2. $E(\epsilon) = 0$, $Var(\epsilon)=\sigma^2$
3. Errors are normally distributed

plus *either* of the following
- X is "fixed"
- X is independent of ϵ, measured perfectly

Important questions

Which X_j belong in the model?
(a.k.a., variable selection)

Is the role of each predictor linear?

Do the predictors interact? Why additive?

Outliers?

Missing data?

Measurement error?

Easier question

Standard statistics issue:
What is a CI for coefficients if model holds?

Looking at a Multiple Regression

Example data duncan.dat

Duncan occupational status data, c.1957

Observations are 45 "occupations"

Y	Prestige = % rating from survey
X_1	Income = % above \$3500
X_2	Education = % HS

Common data set in Stine & Fox (1997)

Modeling question

Regression model

$$\text{Prestige} = \beta_0 + \beta_1 \text{Income} + \beta_2 \text{Educ} + \text{error}$$

Research question

Which effect is larger?

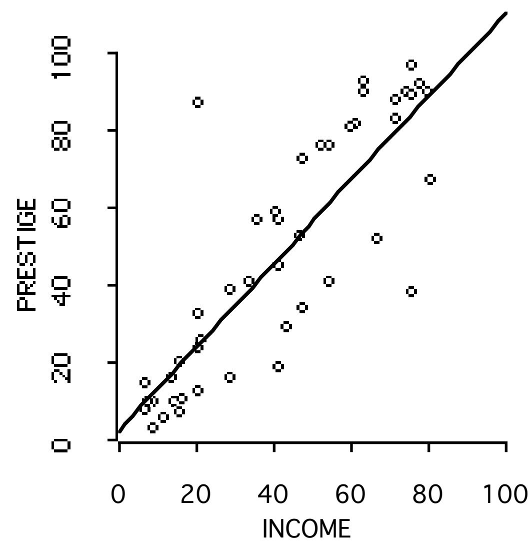
β_1 (income) or β_2 (education)?

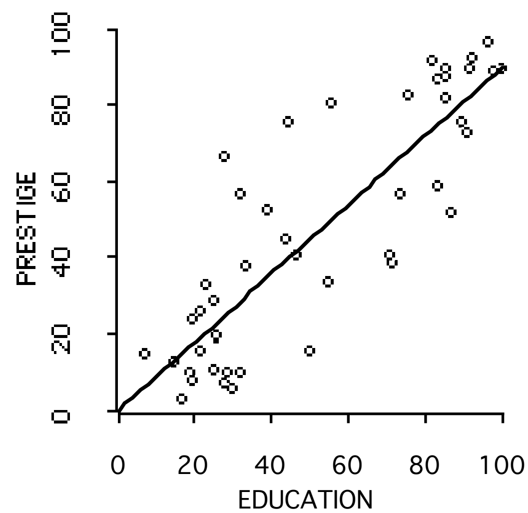
Which factor at the time of this data had a larger *partial* association with how the occupation was rated in this study? Note that the units of the predictors are comparable, so the question is sensible.

Exploratory Plots

Marginal scatterplots

Plot of prestige on each predictor reveals
Presence of outlying occupations (which?)





2.46 + 1.08 INCOME .284+ .902 EDUCATION

Comparison marginally

Income has steeper *marginal* slope.

Differences in regression between
marginal slope vs. partial slope

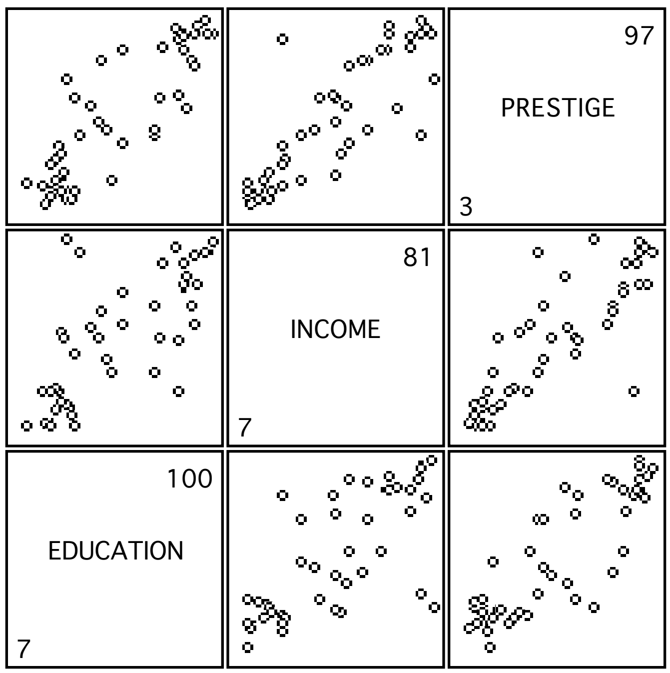
Partial slopes are likely to be smaller for both
– Draw the "graph" of the model
– Correlation between predictors
– Direct vs. indirect effects

Scatterplot Matrix

Connected scatterplots

Visual correlation matrix
- Marginal plots of Y on each predictor
- Includes plots between predictors

Linking and brushing are very powerful



Outlying occupations

Minister, railroad conductor and engineer show up as unusual ("special") occupations.

Are these leveraged? Influential?

OLS Fit for Duncan Model

Fitted model

Variable	Slope	SE	t	p-value
Constant	-6.0647	4.27	-1.42	0.16
INCOME	0.5987	0.12	5.00	0.00
EDUC	0.5458	0.10	5.55	0.00

$R^2 = 0.83$ $\quad s = 13.4$

Both slopes are significantly different from zero

Income has the larger partial slope.

*Are the slopes **significantly** different?*

Standard approach

Test whether a combination of coefficients is zero. Here, just the difference of slopes.

In general, you can test whether any weighted sum of the coefficients is different from zero

Difference is not significant ($t \approx 0.3$)

Graphical Analysis

"Model as lens"

Regression model focuses attention on certain features of the data.

Diagnostic plots

"Standard" residual plots
- Residuals on fit
- Residuals on each predictor, etc.

Leave-one-out diagnostic plots
- leverage on stud. residuals.
- influence plots (Cook's d)

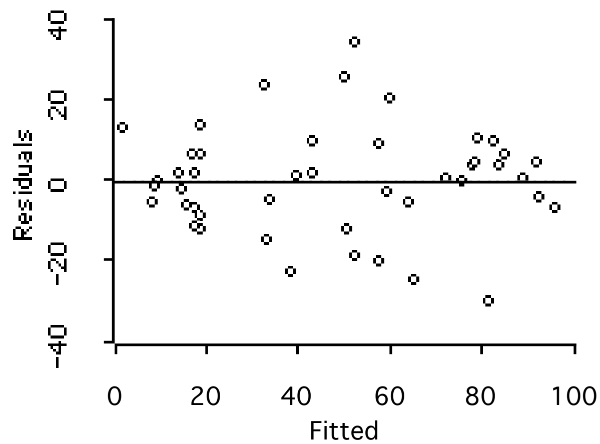
Diagnostic plots for each predictor
- Partial regression plot
- Partial residual plot

Familiarity with these?

Leverage and Outlier Plots

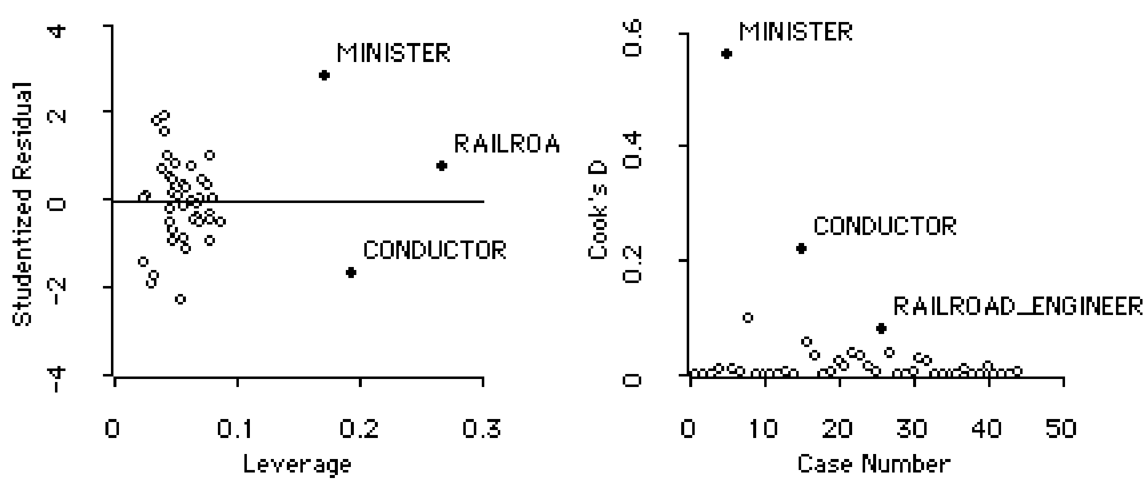
Old stand-by

Plot residuals on fitted values
- Plot suggests things are "OK"



Basic diagnostic plots

Several occupations are clear outliers



Regression Diagnostic Plots

Viewing the multiple regression

Idea

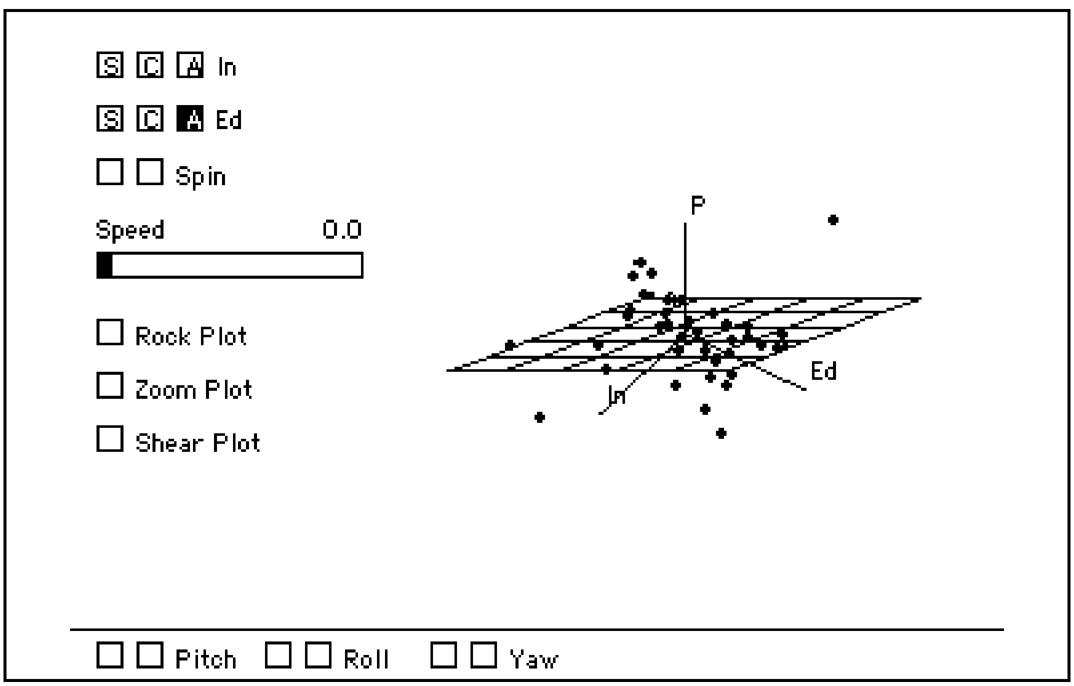
View multiple regression as a collection of "simple" regressions.

Each plot shows a simple regression view of a multiple regression coefficient.

Animated 3-d views

Look at fitted model in 3D.

Diagnostic plots are "edge-on" views



Diagnostic Plots

Two flavors

Partial regression plots

a.k.a. Added variable plots, leverage plots

Emphasize leveraged points, outliers

Interpretation of multiple regression slope
Construction of partial regression plot suggests what it means to “control for the other predictors” in a multiple regression model.

Partial residual plots

a.k.a. Component plots

Emphasize presence of nonlinear trends.
Partial regression plots typically do not show nonlinearity since they shows plots of residuals.

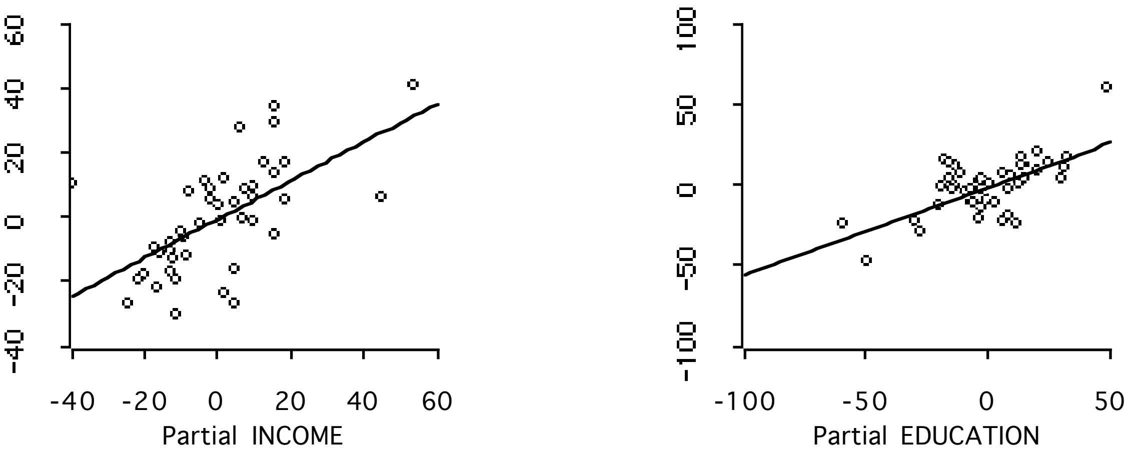
Component plots are similar to plots of residuals on original predictors.

Recent research in graphics (D. Cook)

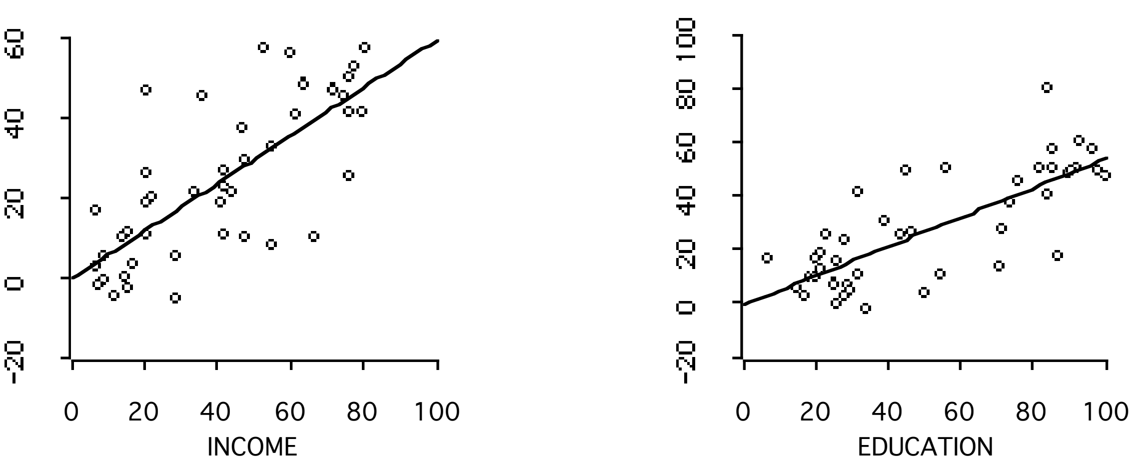
Diagnostic Plots for Duncan

Partial regression plots (added-variable plots)

Show effects of outliers in *multiple* regression
- Attenuate slope for income



Partial residual plots (component plots)



Diagnostics for the Difference

Impact of leverage points

Question

Do leveraged occupations affect difference?

Slick test: reformulate the model

$$\begin{aligned} \textit{Prestige} &= \beta_0 + \beta_1 \textit{Inc} + \beta_2\, \textit{Educ} + \epsilon \\ &= \beta_0 + (\beta_1 - \beta_2)\textit{Inc} + \beta_2(\textit{Educ} + \textit{Inc}) + \epsilon \end{aligned}$$

Coefficient of *Income* in reformulated model *is* the difference of the slopes in original.

Allows diagnostic plots for difference!

Results of reformulated model

Difference in coefficients is *not* significant

- Difference is 0.0529

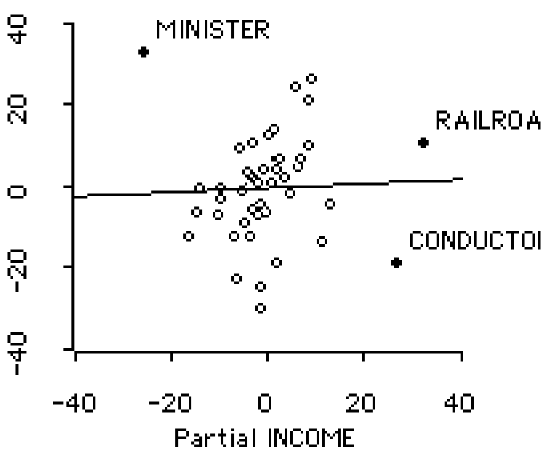
Variable	Slope	SE	t	p-value
Constant	-6.0647	4.27	-1.4	0.16
INCOME	**0.0529**	0.20	0.3	0.80
INC+ED	0.5458	0.10	5.6	0.00

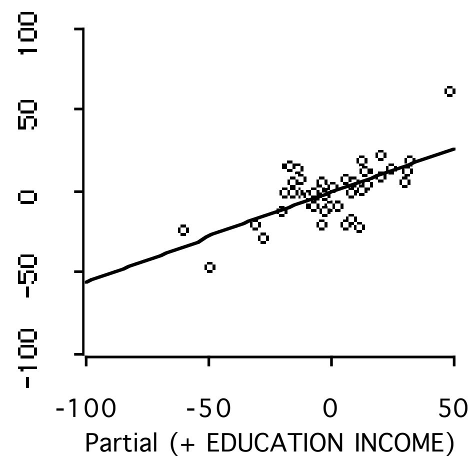
$R^2 = 0.83$ $s = 13.4$

Diagnostic Plots

Partial regression plots

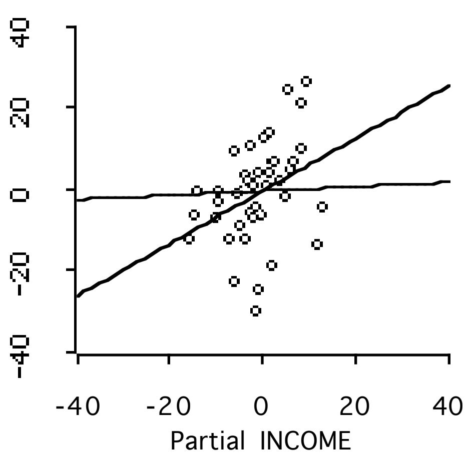
Three leveraged points attenuate difference.





What happens without outliers?

Revised fit suggests significant slope.



Further Analysis of Duncan Model

Leveraged / influential occupations

Ministers, railroad conductors, engineers

Fitted model without these

Use point state in plot to filter data.
= point-state 'normal

Fit with 42 remaining occupations...

Variable	Slope	SE	t	p-value
Constant	-6.32	3.7	-1.7	0.09
INCOME	0.93	0.15	6.1	0.00
EDUC	0.28	0.12	2.3	0.02

$R^2 = 0.876$ $s = 11.5$

Difference of the slopes now?

Use:test message as before, but with 42 observations gives...

t = 2.43 -> Significant

Is this honest?

Several subjective steps…

Robust Regression

Goals

Eliminate the subjectivity...
Which points ought we remove?

Fixes problems due to *few* large residuals.

Ought to control the leverage as well
Some do – they're called bounded influence.

Iterative calculation

Chicken-or-egg problem
To decide which points are outliers implies that you need to know a fitted line…

But to know which line to fit, you need to know which points are outliers.

Iteratively reweighted least squares

Now you want a faster computer!

Fitting a Robust Regression

Robust Fitting process

Iterations
- Start with OLS(not so good, really)
- Iteratively down-weight "outliers"

Plot weights on residuals
See which have been down-weighted. The down-weighting is done automatically.

Different choices for how to down-weight
- influence functions
- biweight, Huber... see references

Results for Duncan's data

Estimates using all 45 occupations (the three other occupations are back in for this fitting) via biweight show a difference...

Variable	Slope	SE	ratio
Constant	-7.4	3.0	-2.5
INCOME	0.76	0.084	9.0
EDUC	0.44	0.069	6.4

How to test? Do we believe the SE values offered by the robust regression?

Comparing the Robust Slopes

How to compare the slopes? Test the difference?

Bootstrap!

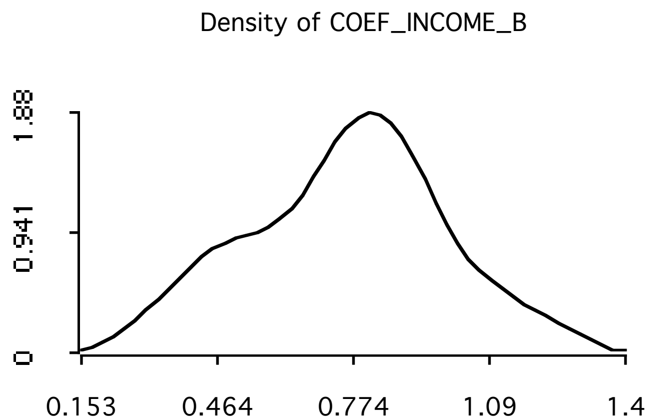
Are the results similar to the output? To OLS?

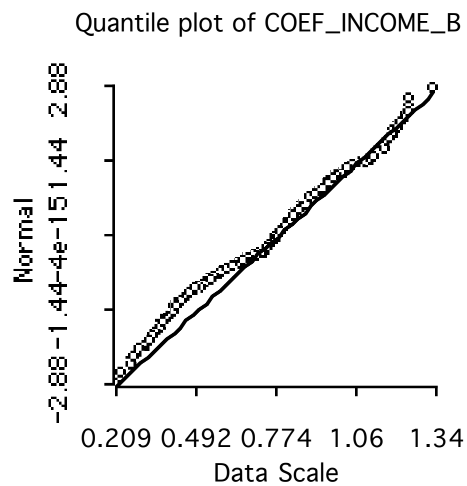
Bootstrap results

Big differences
- Bootstrap standard errors are much larger
- Output shows "asymptotic" formulas
- Asymptotics can be poor with small samples.

	Asymptotic	*Bootstrap*
SE(educ)	0.07	0.17
SE(inc)	0.08	0.22

Bootstrap distribution is *not* normal



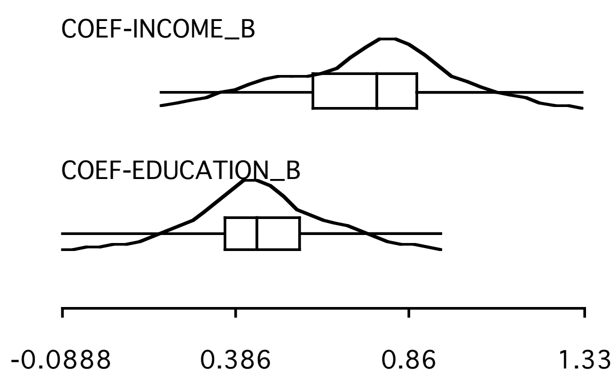


Comparison of Robust Slopes

Compare bootstrap distributions

AXIS compare command
– use the vertical option

Comparison boxes show □income > □education



Significant difference?

Form confidence interval for the differences
Subtract and look at the summary…

Mean = 0.293 SD = 0.381 n=500

2.5%	5%	25%	50%	75%	95%	97.5%
-0.502	-0.368	0.0027	0.332	0.527	0.931	1.05

Percentile interval implies difference is *not* significant.

Dilemma

Why was it significant with OLS without the three outlying occupations, but the robust regression does not think so?

One answer

Robust regression automatically down-weights points that are far from the fit… It down-weights *all* that are far from fit.

This much less focused than deleting three occupations.

A different answer

When resampling, the outliers may get replicated several times and "overwhelm" the robust regression.

Things to Take Away

Bootstrap resampling in multiple regression

Done in the same way as in simple regression.

Graphical regression analysis

Powerful tools for understanding a regression.

Robust regression guards against outliers.

Difference of robust slopes was large.

Similar to removing the outlier occupations.

"Objective" method allows SE calculation via bootstrap.

Other topics in regression (not discussed)

Estimating the accuracy of predictions

Judging the effects of variable selection

Review Questions

How does robust regression differ from LS regression?

Robust regression down-weights observations with extreme residuals, requiring for estimation an iterative procedure known as IRLS (iteratively reweighted least squares).

How are the two methods of resampling in regression affected by outliers?

Fixed X resampling scatters the outlier about the model, whereas random resampling keeps it located in the same spot. If the outlier is leveraged, it will produce bimodal distributions with random resampling, but not with residual resampling. It's often hard, as we have seen, to decide which is right.

Why are bootstrap intervals often similar to LS intervals? Different for robust estimators?

For residual resampling, the BS standard error agrees with the usual formula, up to a factor of n/(n-k) (where k = 1 + # fitted slopes)

$$SE^* = s^2\,(X'X)^{-1} \quad , \quad s^2 = RSS/n$$

Things will often differ for observation resampling since now the X's are random, not fixed, and the sampling variation is larger.

For robust estimators, the "standard formulas" are just rough approximations that are valid only for very large samples, and these approximations are typically inferior to what the BS gives.

How is a scatterplot matrix different from a correlation matrix? Similar?

It's like a graphical correlation matrix. Each plot shows more than a simple numerical value.

What's the difference between a component plot and an added variable plot? How are they similar?

Component plot (partial residual)
 - (Resids + Fit) on (Raw X)
Added variable plot (partial regression)
 - (Resids of Y on others) on

(Resids on X on others)

Both show a "simple regression view" of a multiple regression slope.

What is the problem with judging significance after removing outliers?

Removing data by hand effectively weights those observations zero, leaving the rest weighted one. How much variation does this manual weighting process introduce? The process is typically also a bit subjective. We often find what we want to find.

How does a simple robust regression determine which points have large/small weight?

The graph of the robust weights on the residuals reveals the underlying function that sets the weights, known as an influence function. Two popular influence (weight) functions are the Huber and biweight.

Why did the BS indicate that the difference in robust slopes was large, but not significant?

Evidently, the “smooth” downweighting of points done by the robust regression has more variation than our discrete elimination of the three leveraged occupations.

Why is the BS density of the robust slope estimator often bimodal?

Enough outliers can overwhelm the robust estimator. It will not eliminate them all, and in this case gives a value unlike that for a sample with fewer outliers.