

Robust Regression

The purpose of this exercise is to explore the use of robust methods in linear models for counteracting the effects of unusual data (outliers, influential cases). We will use the data from Duncan (1961), examining the relation between Prestige of occupational categories and the average Income and Education in those jobs. The first few cases are:

case	job	income	educ	prestige
1	Accountant	62	86	82
2	Pilot	72	76	83
3	Architect	75	92	90
4	Author	55	90	76
5	Chemist	64	86	90
6	Minister	21	84	87
7	Professor	64	93	93

1. The data is stored in N:\data\duncan.sas. Read it into SAS using

```
%include data(duncan);
```

2. Just to recall the basic analysis, fit a regression model predicting prestige from income and education. Recall that there were a few influential observations that affected the regression coefficients.

```
proc reg data=duncan;
  model prestige = Income Educ / influence;
  id job;
run;
```

3. Now, use the `robust` macro to fit a robust regression model, downweighting observations with large residuals. There are several robust methods available, but we'll use the (default) bisquare function. The plot step below shows the case weights on the final iteration.

```
%robust(data=duncan,
  response=prestige, model=income educ,
  id=job, proc=reg, function=bisquare,
  out=results);
proc plot data=results;
  plot _weight_ * case = job;
run;
```

4. The `robust` macro gives no details on the final model, but we can get these by running the same model using the weights in the final iteration.

```
proc reg data=resids;
  model prestige = income educ / influence;
  weight _weight_;
  id job; run;
```

Robust

- Now, try doing the same analysis with `proc robustreg`. There are many robust methods implemented, but here we'll just use the default method of M estimation. (In the SAS log, you will see a warning that methods MM and LTS are better with influential cases.) Note that `proc robustreg` can produce some useful plots using ODS Graphics.

```
ods rtf file='robdunc0.rtf' style=journal;
ods graphics on;
proc robustreg data=duncan
    plots=(ddplot(label=leverage) rdplot(label=leverage)) ;
    model prestige = income educ / diagnostics itprint ;
    id job;
    output out=resids r=residual weight=weight outlier=outlier;
run;
ods graphics off;
ods rtf close;
```

Robust Regression in R

Below is an R script to carry out much the same analysis in R, using the data Duncan in the car package and `rlm()` in the MASS package. This script is available in

N:\psyc6140\tutorials\robust.R.

```
library(car)
data(Duncan)
library(MASS)

# standard OLS
dunc.lm <- lm(prestige ~ income+education, data=Duncan)
summary(dunc.lm)

# M estimate
dunc.robust <- rlm(prestige ~ income+education, data=Duncan)
summary(dunc.robust)
# show small weights
wts <- dunc.robust$w
(cbind(Duncan,wts))[wts<0.5,]

# MM estimate
dunc.robust1 <- rlm(prestige ~ income+education, data=Duncan, method="MM")
summary(dunc.robust1)

wts <- dunc.robust$w
(cbind(Duncan,wts))[wts<0.5,]

# compare coefficients
rbind(coef(dunc.lm), coef(dunc.robust), coef(dunc.robust1))
```