multiple predictors ex.

Stat 209

[redacted]

File coleman.dat contains data from a random sample of 20 schools (from the East) from the 1966 Coleman Report.
The outcome measure C7 is the verbal mean test score for all sixth graders in the school. The predictor variables are: C2, staff salaries per pupil, C3, percent white collar fathers for the sixth graders; C4 is a SES composite measure (deviation) for the sixth graders, C5 Mean teacher's verbal test score, C6 6th grade mean mother's educational level (1 unit=2 school yrs)

outcome

	ssal	whcol	ses	tverb	momed	vach
1	3.83	28.87	7.20	26.60	6.19	37.01
2	2.89	20.10	-11.71	24.40	5.17	26.51
3	2.86	69.05	12.32	25.70	7.04	36.51
4	2.92	65.40	14.28	25.70	7.10	40.70
5	3.06	29.59	6.31	25.40	6.15	37.10
6	2.07	44.82	6.16	21.60	6.41	33.90
7	2.52	77.37	12.70	24.90	6.86	41.80
8	2.45	24.67	-0.17	25.01	5.78	33.40
9	3.13	65.01	9.85	26.60	6.51	41.01
10	2.44	9.99	-0.05	28.01	5.57	37.20
11	2.09	12.20	-12.86	23.51	5.62	23.30
12	2.52	22.55	0.92	23.60	5.34	35.20
13	2.22	14.30	4.77	24.51	5.80	34.9
14	2.67	31.79	-0.96	25.80	6.19	33.10
15	2.71	11.60	-16.04	25.20	5.62	22.70
16	3.14	68.47	10.62	25.01	6.94	39.70
17	3.54	42.64	2.66	25.01	6.33	31.80
18	2.52	16.70	-10.99	24.80	6.01	31.70
19	2.68	86.27	15.03	25.51	7.51	43.10
20	2.37	76.73	12.77	24.51	6.96	41.01

data are ascii

from week 1 math notes

I Two predictor model $Y \quad X_1 \quad X_2$

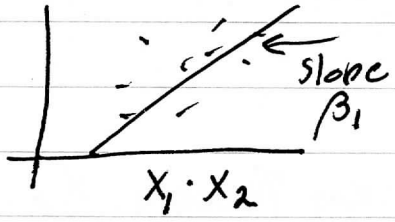
$$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

adjusted (partial) variable interpretation

$$\beta_1 = \beta_{YX_1 \cdot X_2} = \beta_{Y(X_1 \cdot X_2)}$$

adjusted var $X_1 \cdot X_2 = X_1 - \beta_{X_1 X_2} X_2$

same for β_2



fantasy: "holding constant"
reality: "reaming out"

extends to p predictors: $\beta_1 = \beta_{Y(X_1 \cdot \underset{\sim}{X}_{(1)})}$ $\underset{\sim}{X}_{(1)}$ all but X_1

(see MT, coleman ex, Berk, NWK)

school-level data, Coleman 1966, adjusted variables illustration

```
> ed = read.table(file="D:\\drr06\\stat209\\coleman.dat", header = T)
> attach(ed)
> cor(ed)
           ssal      whcol       ses      tverb     momed      vach
ssal  1.0000000 0.18113980 0.2296278 0.50266385 0.1967731 0.1922916
whcol 0.1811398 1.00000000 0.8271829 0.05105812 0.9271008 0.7534008
ses   0.2296278 0.82718291 1.0000000 0.18332924 0.8190633 0.9271611
tverb 0.5026638 0.05105812 0.1833292 1.00000000 0.1238087 0.3336495
momed 0.1967731 0.92710081 0.8190633 0.12380866 1.0000000 0.7329859
vach  0.1922916 0.75340081 0.9271611 0.33364951 0.7329859 1.0000000
> edreg = lm(vach ~ ssal+ whcol+ ses+ tverb+ momed)
> summary(edreg)
Call:
lm(formula = vach ~ ssal + whcol + ses + tverb + momed)
Residuals:
     Min       1Q   Median       3Q      Max
-3.94972 -0.61739  0.06235  0.73430  5.00176
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.94857   13.62755   1.464   0.1653
ssal        -1.79333    1.23340  -1.454   0.1680
whcol        0.04360    0.05326   0.819   0.4267
ses          0.55576    0.09296   5.979 3.38e-05 ***
tverb        1.11017    0.43377   2.559   0.0227 *
momed       -1.81092    2.02739  -0.893   0.3868
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.074 on 14 degrees of freedom
Multiple R-Squared: 0.9063,     Adjusted R-squared: 0.8728
F-statistic: 27.08 on 5 and 14 DF,  p-value: 9.927e-07

> momreg = lm(momed ~ ssal + whcol + ses + tverb)
     #this regression object is fit to momed using the other predictors
     #residuals(momreg) are the plain residuals from this fit

> momregadj = lm(vach ~ residuals(momreg))
> summary(momregadj)

Call:
lm(formula = vach ~ residuals(momreg))
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         35.082      1.333  26.322 8.03e-16 ***
residuals(momreg)   -1.811      5.826  -0.311     0.76
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5.96 on 18 degrees of freedom
Multiple R-Squared: 0.00534,    Adjusted R-squared: -0.04992
F-statistic: 0.09663 on 1 and 18 DF,  p-value: 0.7595
```

read in data

big pos corr

regression object

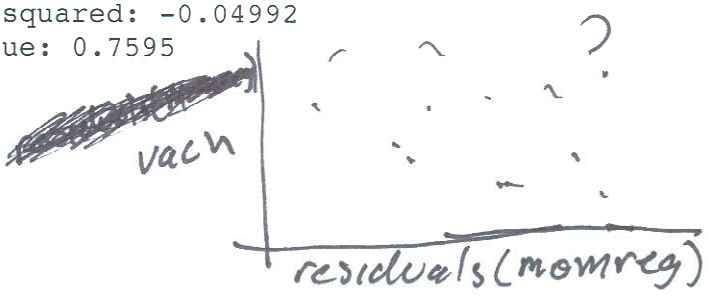
full regression MT result

only n=20 not signif. MT, not whole story

obtain multiple regr. coeff for momed via adjusted (partial) variable predictn

match coeff

exercise: should also plot. c.f. Berk, N/Wk tests



Stat 209
week 1

Coleman Data

```
R version 2.14.1 (2011-12-22)
> ed = read.table(file="http://www-stat.stanford.edu/~rag/stat209/coleman.dat
  > dim(ed)  [1] 20  6  > attach(ed)
> momreg = lm(momed ~ ssal + whcol + ses + tverb)
> momregadj = lm(vach ~ residuals(momreg))
> plot(residuals(momreg), vach, pch = 20)
>  abline(momregadj)
>  coef(momregadj)
(Intercept) residuals(momreg)
35.082500          -1.810922
```

plot adjusted variable regression

plotting char