

Lecture 12

Each year the *U.S. News and World Report* provides rankings of US colleges (and universities), based on data provided by the colleges themselves. The web page at <http://www.usnews.com/usnews/edu/college/corank.htm> documents (see *Rankings information*) the methods that US News uses.

The data for the 1994 rankings (downloaded from the web site for the Journal of Statistics Education: gopher://jse.stat.ncsu.edu:70/11/jse/) are linked to the *datasets* page at Stat100++ web site. The data consist of the following variables:

FICE = FICE (Federal ID number)
 name = College Name (abbreviated)
 State = State
 type = Public/private indicator (public=1, private=2)
 SATm = Average Math SAT score
 SATv = Average Verbal SAT score
 SAT = Average Combined SAT score
 ACT = Average ACT score
 SATm25 = First quartile - Math SAT
 SATm75 = Third quartile - Math SAT
 SATv25 = First quartile - Verbal SAT
 SATv75 = Third quartile - Verbal SAT
 ACT25 = First quartile - ACT
 ACT75 = Third quartile - ACT
 new.applied = Number of applications received
 new.accepted = Number of applicants accepted
 new.enrolled = Number of new students enrolled
 new.top10HS = Pct. new students from top 10% of H.S. class
 new.top25HS = Pct. new students from top 25% of H.S. class
 undergrads = Number of fulltime undergraduates
 parttime = Number of part-time undergraduates
 tuition.in = In-state tuition
 tuition = Out-of-state tuition
 room+board = Room and board costs
 room = Room costs
 board = Board costs
 cost.fees = Additional fees
 cost.books = Estimated book costs
 cost.pers = Estimated personal spending
 fac.phd = Pct. of faculty with Ph.D.'s
 fac.term = Pct. of faculty with terminal degree
 sf.ratio = Student/faculty ratio
 donate = Pct. alumni who donate
 exp.stud = Instructional expenditure per student
 graduate = Graduation rate
 College = College name (in full)

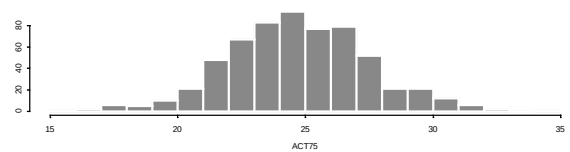
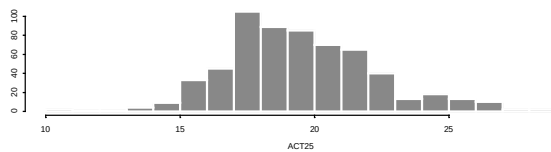
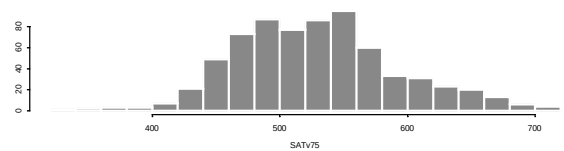
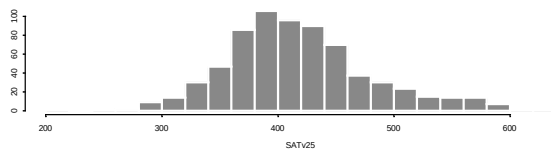
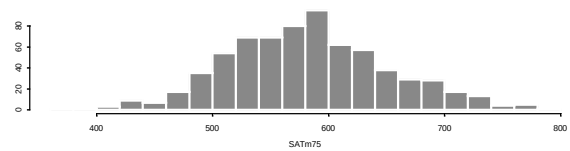
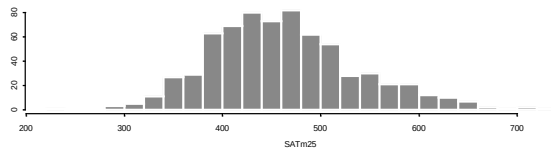
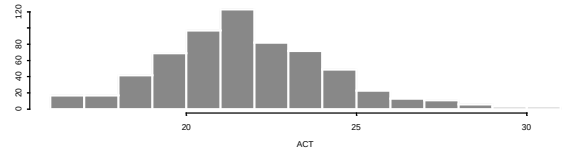
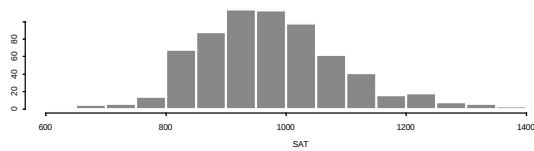
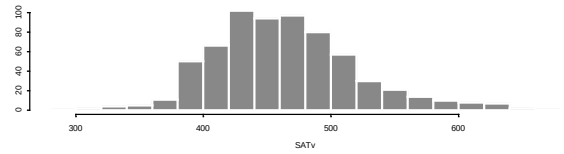
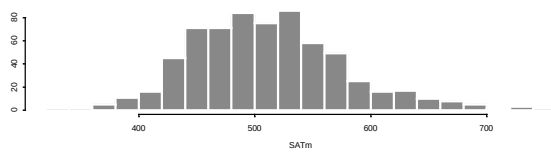
A star (*) in the original data indicates a missing value. The software that I use writes NA for missing values.

I extracted the subset of the data, consisting of the 1133 colleges for which there is AAUP salary data available.

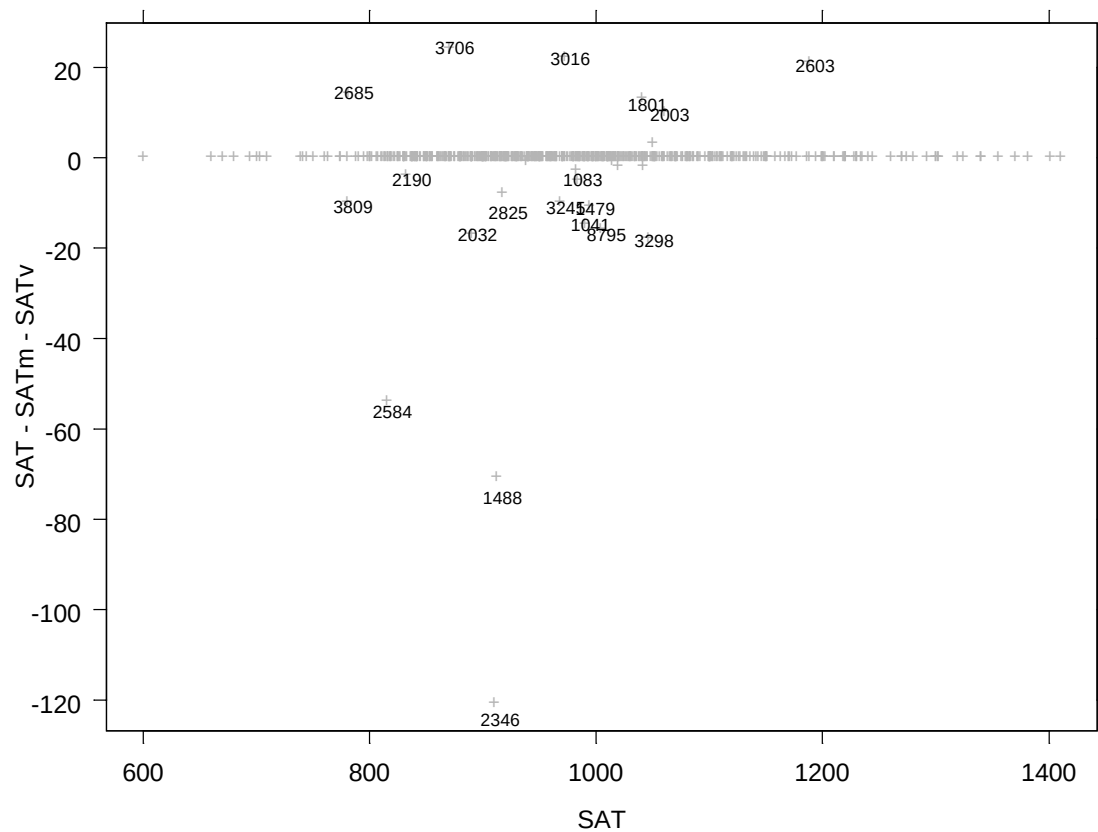
The *U.S. News* derives ranks by taking weighted combinations of scores derived from such data. For example, 'Student selectivity' contributes 15% to the ranking score, with average SAT/ACT scores contributing 40% towards the 'Student selectivity'.

Missing data complicate comparisons of SAT/ACT scores across colleges. For some colleges we have only first and third SAT quartiles; for others we have only average SAT scores; for others we have only ACT scores or quartiles. How should we combine these pieces of partial information into a single SAT measure?

	<i>SATm</i>	<i>SATv</i>	<i>SAT</i>	<i>ACT</i>	<i>SATm25</i>	<i>SATm75</i>	<i>SATv25</i>	<i>SATv75</i>	<i>ACT25</i>	<i>ACT75</i>
Min.	320	280	600	16	220	370	200	330	10	15
1st Qu.	467	426.2	896.5	21	420	540	380	490	18	23
Median	507	460	968	22	460	585	420	530	20	25
Mean	512.6	465.3	977.7	22.27	467.2	587.3	422.7	533.7	19.96	25.21
3rd Qu.	550	495.8	1043	24	510	630	450	570	22	27
Max.	750	665	1410	31	740	785	630	720	29	35
NA's	471	471	470	506	439	439	439	439	526	526



For all except 2 colleges with SAT present, both SATm and SATv are also present. For only 1 college with both SATm and SATv present, is SAT missing. In most cases $SAT = SATm + SATv$. No college has only two or three of the four SAT quartiles, or only one of the ACT quartiles.



Presence/absence of SAT/ACT scores

	*	SATm25
*	332	138
SAT	107	556

	*	ACT25
*	366	140
ACT	160	467

	*	sat (any form)
*	83	283
act (any form)	249	518

*

ACT

	*	SATm25
*	83	54
SAT	33	196

	*	SATm25
*	61	0
SAT	59	40

*

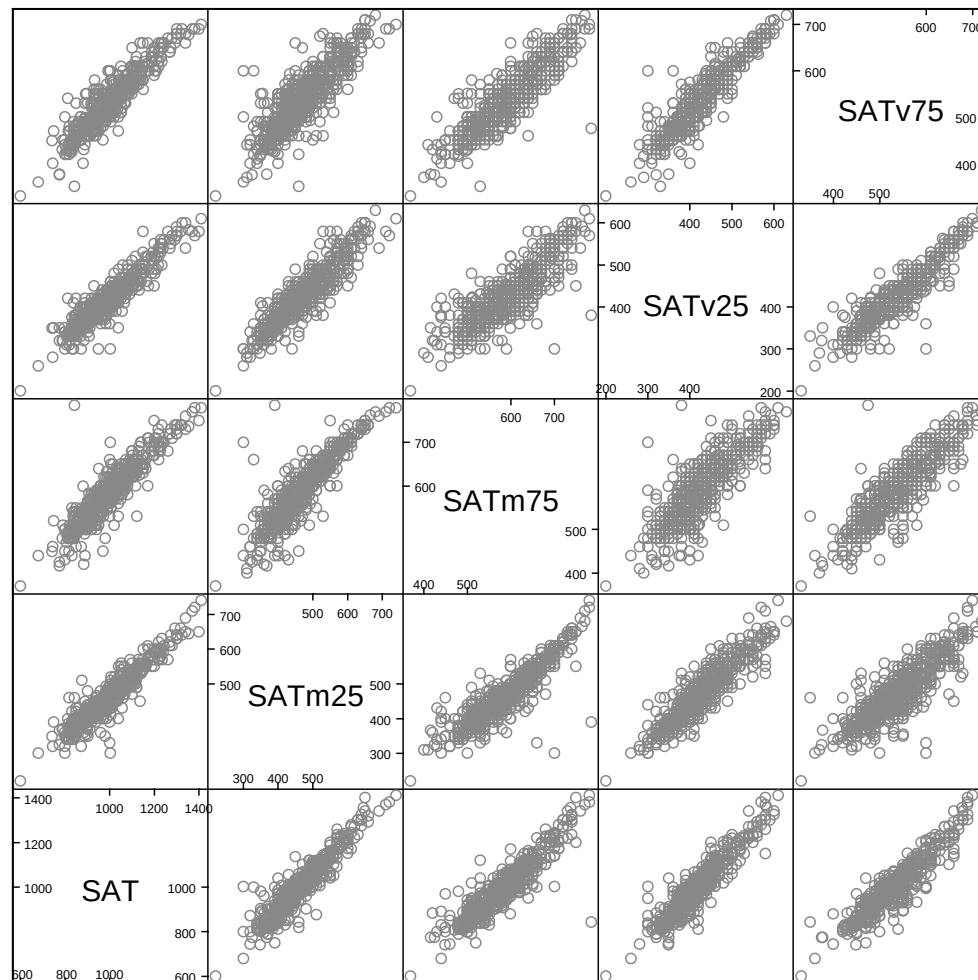
	*	SATm25
*	53	78
SAT	0	9

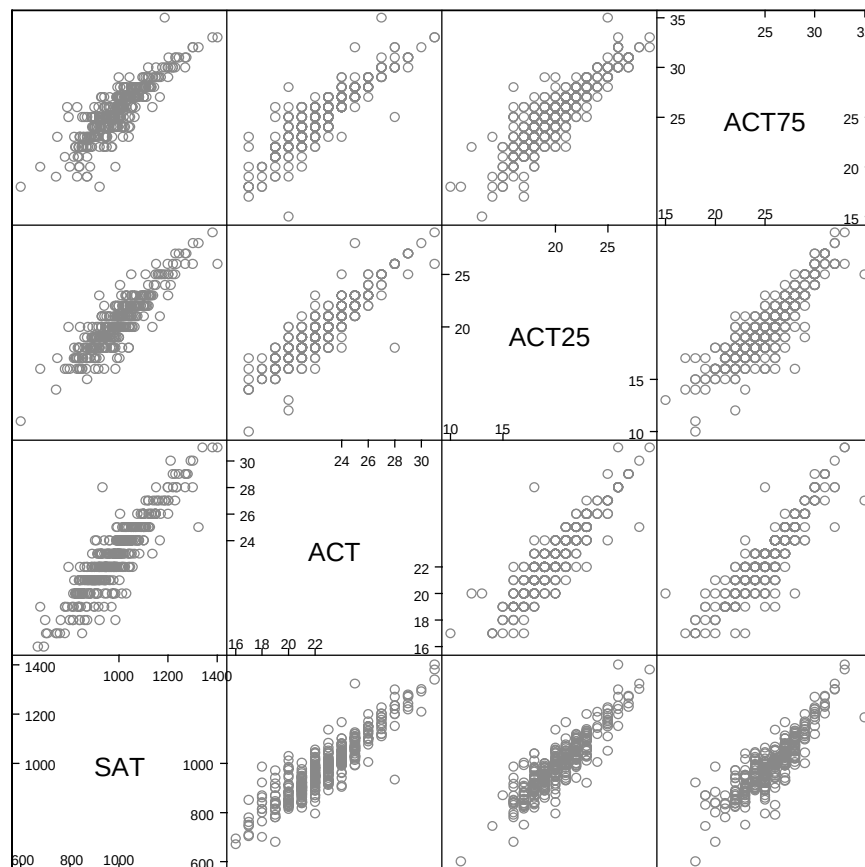
	*	SATm25
*	135	6
SAT	15	311

ACT25

How to predict SAT scores?

Which is better: use to the coarser ACT scores to predict SAT, vice versa?





Predict SAT from ACT

	Value	Std. Error	t value	Pr(> t)
(Intercept)	71.8210	21.1620	3.3939	0.0008
ACT	40.0024	0.9231	43.3366	0.0000

Residual standard error: 50.12 on 423 degrees of freedom
Multiple R-Squared: 0.8162

Predict SAT from ACT25 and ACT75

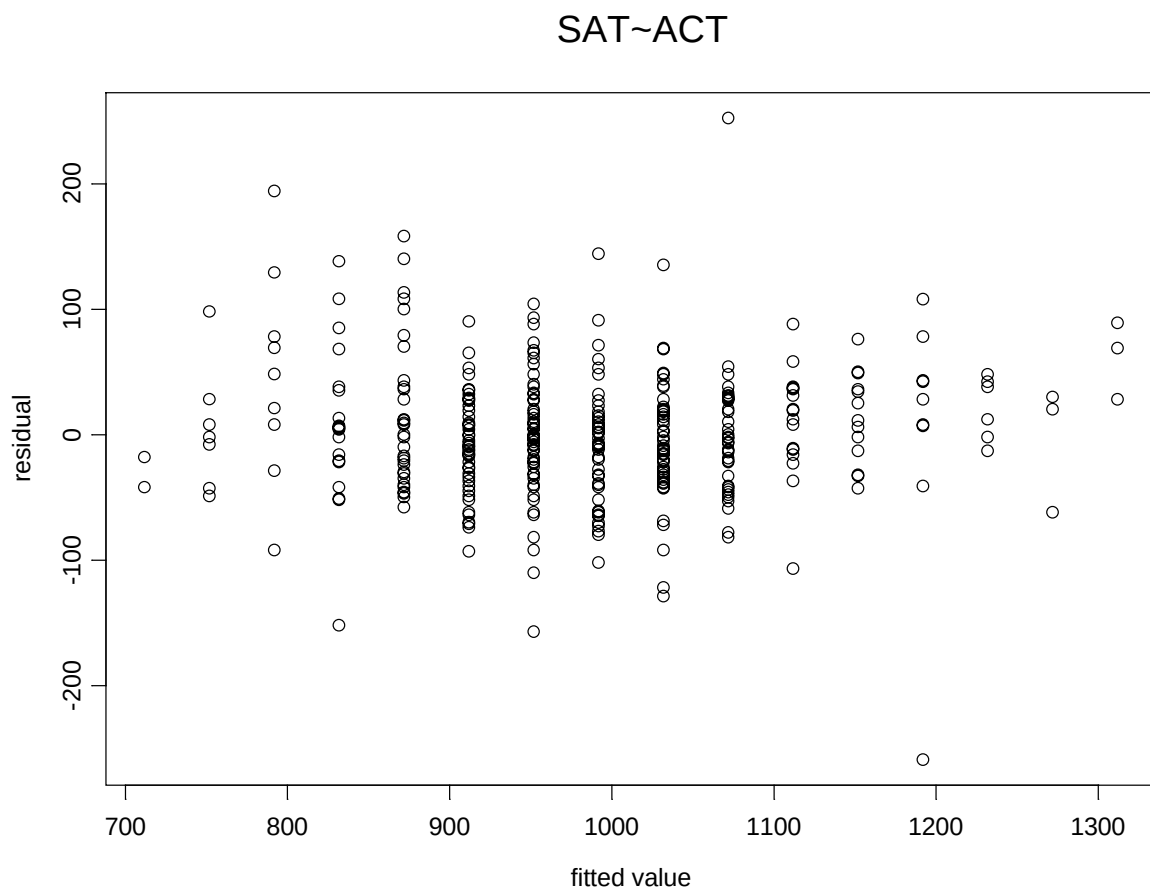
	Value	Std. Error	t value	Pr(> t)
(Intercept)	115.8312	24.9292	4.6464	0.0000
ACT25	23.7215	2.0022	11.8476	0.0000
ACT75	15.5143	1.9658	7.8921	0.0000

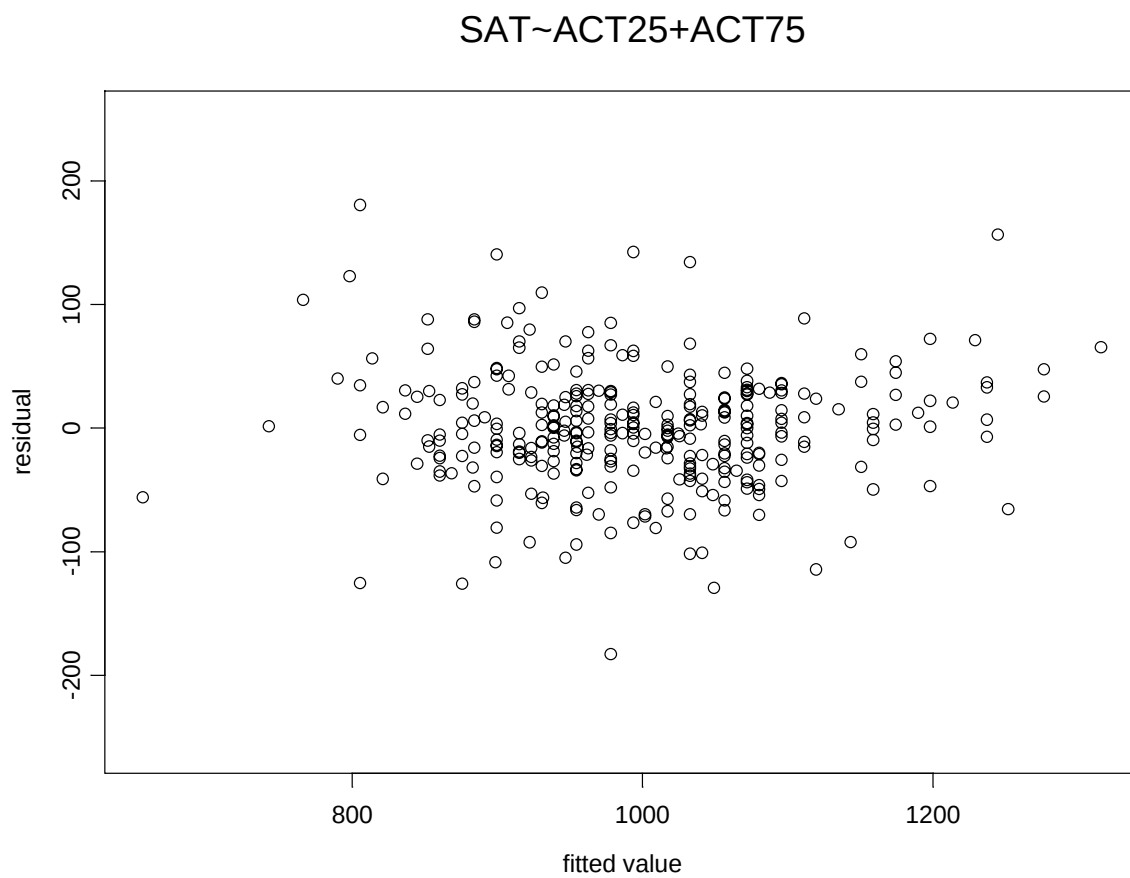
Residual standard error: 46.74 on 332 degrees of freedom
Multiple R-Squared: 0.8222

Predict SAT from SATm25 , SATm75 , SATv25 , SATv75

	Value	Std. Error	t value	Pr(> t)
(Intercept)	58.6716	10.8146	5.4252	0.0000
SATm25	0.6203	0.0502	12.3530	0.0000
SATm75	0.3112	0.0457	6.8154	0.0000
SATv25	0.3625	0.0658	5.5083	0.0000
SATv75	0.6018	0.0559	10.7662	0.0000

Residual standard error: 26.07 on 551 degrees of freedom
Multiple R-Squared: 0.9529





SAT~SATm25+SATm75+SATv25+SATv75

