

<div> <div>Chapter 12: Strategies for Variable Selection</div> <div>Class 19, 4/15/09 W</div> </div>	<div>Slide 1 Chapter 12: Strategies for Variable Selection</div> <div>NOTES:</div>
<div> <div>HW 12 due Friday 4/17/09</div> <div>Submit as Myname-HW12.doc (or *.rtf)</div> <ul style="list-style-type: none"> <li>HW 12 10.28: El Niño and Hurricanes <ul style="list-style-type: none"> <li>Due Friday 4/17/09 Noon</li> </ul> </li> <li>HW 13 Cammen's ingestion rate data. Note that this was a 2003 final exam problem <ul style="list-style-type: none"> <li>Read Cammen (1980) &amp; evaluate his regression model</li> <li>Due Friday 4/25/09 Noon This problem will count double!</li> </ul> </li> <li>Read Chapter 12: Selection of variables</li> <li>Run my overfitting syntax: overfitting.sps</li> <li>Read Campbell &amp; Kenney Chapters 4 &amp; 5 on the regression artefact and gender inequities <ul style="list-style-type: none"> <li>Run my Campbell &amp; Kenny syntax: RTMCK.sps</li> </ul> </li> <li>No Class Monday 4/20: Patriot's Day</li> </ul> </div>	<div>Slide 2 HW 12 due Friday 4/17/09</div> <div>NOTES:</div>
<div> <div>HW12: Cammen model</div> <p>Cammen (1980) compiled data from the literature on the ingestion rates of 22 deposit feeders. Deposit feeders are organisms that live in mud and sand and ingest mud and sand. Deposit feeders use the organic matter in the mud and sand for growth. Table 1 shows the species from the literature, their ingestion rates, the fraction organic matter in sediment, and the body weights of individual deposit feeders. Cammen (1980) used regression to estimate the ingestion rate of deposit feeders (<b>ING</b>) (mg dry weight/day) using the fraction organic matter in the sediment (<b>OM</b>) and body weight of the deposit feeder (<b>WT</b>). He regressed <math>\log_{10}</math> (<b>ING</b>) as the response variable with two explanatory variables <math>\log_{10}</math> (<b>WT</b>) and <math>\log_{10}</math> (<b>OM</b>). He deleted the three bivalves from his analyses because they appeared to be outliers, and based his regressions on the 19 non-bivalve species.</p> </div>	<div>Slide 3 HW12: Cammen model</div> <div>NOTES:</div>

Table 1. Data from **Cammen (1980)**. Loaded on **Prometheus** as **cammen.csv**, in case you wanted to examine the data (optional). The last 5 highlighted species are bivalve molluscs (indicated under **Taxon**). **WT** is the body weight of the deposit feeder (dry weight of the animal) in milligrams. **ING** is the ingestion rate (mg dry weight/day). Cammen scaled the ingestion rate to account for temperature effects (higher ingestion at higher temperatures). **CPI** is the organic matter content (% weight organic matter / % total sediment dry weight), expressed as %.

Species	Taxon	WT	ING	CPI
1 <i>Nereis virens</i>	Gastropod mollusc	0.2	0.57	16
2 <i>Nereis virens</i>	Gastropod mollusc	0.2	0.66	17
3 <i>Tubificoides</i>	Oligochaete (annelid)	0.27	0.46	29.7
4 <i>Nereis virens</i>	Crustacean	0.32	0.48	50
5 <i>Phoronopsis viridis</i>	Gastropod mollusc	0.46	2.7	14.4
6 <i>Nereis virens</i>	Gastropod mollusc	0.9	0.67	13
7 <i>Nereis virens</i>	Polychaete (annelid)	5.8	25.2	6.8
8 <i>Phoronopsis viridis</i>	Crustacean (annelid)	8.4	1.49	93
9 <i>Orchestoidea</i>	Crustacean	12.4	4.4	88
10 <i>Aricidea</i>	Polychaete (annelid)	20.4	24.0	2.2
11 <i>Tricoproctus</i>	Polychaete (annelid)	46	250	1
12 <i>Ampelisca</i>	Crustacean	53	300	4.2
13 <i>Uca</i>	Crustacean	63.9	19.9	51
14 <i>Scapharca</i>	Crustacean	65	59	23.6
15 <i>Phoronopsis</i>	Polychaete (annelid)	80	1007	0.7
16 <i>Aricidea</i>	Polychaete (annelid)	280	3400	1.2
17 <i>Aricidea</i>	Polychaete (annelid)	380	3400	0.4
18 <i>Aricidea</i>	Polychaete (annelid)	880	4750	0.64
19 <i>Microphallus</i>	Crustacean	2050	4680	2.1
20 <i>Macoma</i>	Bivalve mollusc	5.1	4.49	20
21 <i>Mytilus</i>	Bivalve mollusc	19.9	3.94	6.8
22 <i>Scapharca</i>	Bivalve mollusc	280	4.3	5.4

## Slide 4

NOTES:

- Was Cammen (1980) justified in dropping the three bivalve molluscs from his regression equation?
  - Consider both the case-wise diagnostic tests (residuals vs. predicted values, Cook's D, studentized residuals, and leverage values), and the results of fitting bivalves as a dummy variable.
  - Discuss the problems in using Cook's D, leverage, and studentized residuals in detecting outliers when more than one datum may be an outlier.
  - There is no strictly right or wrong answer to this question, but you must justify your choice with evidence from the regression analyses.
- There were 5 groups of animals in Cammen's data. Is there evidence that the ingestion rates as a function of weight and organic matter differ among these 5 groups?
- Based on your analyses, produce a graph showing the relationship between ingestion rate, body weight and organic matter.
- Write the regression equation expressing the relationship between ingestion rate, organic matter, and body weight. Pay attention to significant figures, and include an estimate of the standard error of the coefficients.
- If you found that the animal groups differed in ingestion rate, your final graphs and model should reflect this full model

## Slide 5

NOTES:

## Homework Presentations

- William Walker for HW 8,
- Steven Kichefski for HW 9 and
- Lisa Greber for HW10



## Slide 6 Homework Presentations

NOTES:

<div data-bbox="276 285 756 350" data-label="Section-Header"> <h2>Chapter 12: Strategies for variable selection</h2> </div>	<div data-bbox="815 132 1417 207" data-label="Section-Header"> <h3>Slide 7 Chapter 12: Strategies for variable selection</h3> </div> <div data-bbox="815 294 940 327" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="289 686 751 753" data-label="Section-Header"> <h2>Using multiple regression to test causal models</h2> </div> <div data-bbox="230 768 753 871" data-label="Text"> <p>Being in politics is like being a football coach. You have to be smart enough to understand the game and dumb enough to think it's important.-- Eugene McCarthy</p> </div> <div data-bbox="243 892 740 1022" data-label="Text"> <p>Application to Regression &amp; Chapter 12 To use multiple regression to test causal models, you have to know enough statistics to run the analysis, but you have to be dumb enough to think the approach is valid</p> </div>	<div data-bbox="815 657 1377 732" data-label="Section-Header"> <h3>Slide 8 Using multiple regression to test causal models</h3> </div> <div data-bbox="815 819 940 852" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="310 1220 721 1257" data-label="Section-Header"> <h2>Regression errors &amp; artifacts</h2> </div> <div data-bbox="230 1283 763 1575" data-label="List-Group"> <ul style="list-style-type: none"> <li>● A) Covariates are often necessary <ul style="list-style-type: none"> <li>▸ Fluoride &amp; cancer (Manly 1992)</li> <li>▸ Storks &amp; babies</li> </ul> </li> <li>● B) Multicollinearity: <ul style="list-style-type: none"> <li>▸ Interpreting Beta signs as effects when the magnitude and sign of Beta is a function of other variables in the equation</li> <li>▸ Handguns &amp; Crime rates (Lott &amp; Mustard vs. Ayers &amp; Donahue)</li> <li>▸ Peterson on school vouchers &amp; test scores</li> </ul> </li> <li>● C) The regression artifact and improper interpretation of the effects of covariates <ul style="list-style-type: none"> <li>▸ Math ability &amp; gender</li> <li>▸ The Bell Curve</li> </ul> </li> </ul> </div>	<div data-bbox="815 1182 1336 1222" data-label="Section-Header"> <h3>Slide 9 Regression errors &amp; artifacts</h3> </div> <div data-bbox="815 1306 940 1341" data-label="Text"> <p>NOTES:</p> </div>

## Does fluoride cause cancer?

Manly (1992) The design & analysis of research studies

- Yiamouyiannis & Burk 1977
  - Fluoridation began in 1952-1956
  - Fluoridated and non-fluoridated cities matched by population size
    - 10 largest non-fluoridated cities
    - Fluoridated cities of comparable size



Table 1.2. Cancer deaths per 100 000 population in fluoridated and non-fluoridated cities in the United States (Yiamouyiannis and Burk, 1977)

	Fluoridated cities	Non-fluoridated cities
1950	181	179 *
1970	217	197
Change	+36	+18

Why ?

## Slide 10 Does fluoride cause cancer?

NOTES:

## Cancer & Fluoride

Manly (1992) The design & analysis of research studies

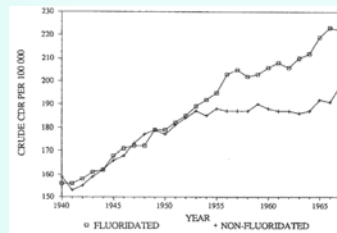


Figure 1.2. Crude cancer death rates per 100,000 of population for ten fluoridated and ten non-fluoridated cities in the United States, 1940-69. Fluoridation of cities took place over the period 1952-56.

Why ?

Yiamouyiannis & Burk 1977: 10 largest non-fl. Cities and matched fluoridated cities

Fluoridation took place from 1952-1956

Rebuttal: Oldham & Newell (1977) Applied Statistics

## Slide 11 Cancer & Fluoride

NOTES:

## Guidelines for predictive modeling

From Holmes' Causal modelling (Sage)

- Theorize before analyzing data or validate theory with additional data
- Formulate explicitly ordered hypotheses
- Measure covariation with an appropriate technique
- Examine measures of association to see if they are significant
- Reject competing models that are more complex or less based on theory
- Reject models that have "bad fit"

## Slide 12 Guidelines for predictive modeling

NOTES:

## Gallagher's addenda

From Harrell & Campbell & Kenney

- Don't use multiple regression to infer causation. When more than one variable is in the model, the sign and magnitude of the coefficients for an explanatory variable often depend on the value of other variables in the equation
- Don't use stepwise or other automated selection procedures
- Beware the regression artifact and control for it
  - Use repeated measures designs, structural equation models or corrections for the regression artifact.
  - Or, design a controlled experiment to properly assess the effect

## Slide 13 Gallagher's addenda

NOTES:

Display 12.1 p. 327

Average SAT scores by US State in 1982, and possible associated factors

State	SAT	Take	Income	Years	Public	Expend	Rank
1. Iowa	1089	2	526	16.79	87.8	25.60	85.7
2. North Dakota	1079	3	284	16.77	89.7	24.97	86.4
3. North Dakota	1068	3	117	16.17	88.3	26.62	89.9
4. Kansas	1067	5	338	16.30	91.9	23.14	88.1
5. Nebraska	1061	5	293	17.23	83.6	23.00	88.3
6. Minnesota	1053	8	263	17.91	92.7	26.48	89.4
7. Montana	1052	4	345	16.47	79.2	27.42	90.9
8. Utah	1023	4	233	16.75	79.2	27.42	90.9
9. Wisconsin	1011	10	394	16.87	73.3	27.69	84.2
10. Wisconsin	1011	10	394	16.87	73.3	27.69	84.2
11. Oklahoma	1000	5	156	17.93	83.2	26.07	89.5
12. Arkansas	999	4	293	14.45	88.9	13.71	89.3
13. Tennessee	999	6	116	17.72	83.7	14.88	83.4
14. New Mexico	993	7	285	16.14	92.1	17.80	83.9
15. Idaho	989	3	133	16.76	87.9	27.36	89.0
16. Kentucky	981	6	130	16.41	71.4	15.49	89.4
17. Colorado	981	6	133	16.83	89.2	26.36	89.3
18. Washington	981	19	309	16.23	87.5	26.13	87.1
19. Illinois	971	11	114	17.86	89.4	17.44	86.3
20. Illinois	971	11	114	17.86	89.4	17.44	86.3
21. Missouri	971	9	284	16.87	44.8	19.72	82.9
22. Michigan	971	10	322	16.42	67.7	26.79	86.9
23. Michigan	971	10	322	16.42	67.7	26.79	86.9
24. West Virginia	968	7	282	17.08	80.7	24.61	85.8
25. Alabama	964	6	131	16.37	89.6	13.84	92.9
26. Ohio	959	16	266	16.52	73.2	22.43	79.4
27. Alaska	923	14	489	15.12	96.5	50.00	76.6
28. Nevada	923	14	489	15.12	96.5	50.00	76.6
29. Oregon	906	40	263	14.48	92.2	30.49	79.3
30. Vermont	906	34	223	16.30	82.2	17.39	82.1
31. California	897	42	275	16.97	83.9	27.81	71.4
32. Connecticut	897	42	275	16.97	83.9	27.81	71.4
33. Maine	896	40	268	16.05	81.7	26.41	79.1
34. New York	896	19	136	16.86	80.4	13.18	79.3
35. Maryland	896	40	268	16.05	81.7	26.41	79.1
36. Florida	889	19	215	15.91	80.5	22.62	74.9
37. Maryland	889	19	215	15.91	80.5	22.62	74.9
38. Massachusetts	889	11	112	16.80	80.4	11.11	71.1
39. Delaware	888	43	246	16.79	80.7	13.14	69.9
40. Rhode Island	877	19	128	16.47	76.7	25.19	71.4
41. New Jersey	877	19	128	16.47	76.7	25.19	71.4
42. Texas	866	27	303	14.95	92.7	19.03	74.9
43. Indiana	866	48	236	14.46	81.4	23.49	69.9
44. Florida	877	47	234	15.31	92.8	19.92	79.3
45. Georgia	877	51	236	14.34	80.5	19.52	74.9
46. South Carolina	790	48	214	13.42	88.1	17.60	74.9

Case Study 12.1  
SAT Scores



## Slide 14

NOTES:

Display 12.2 p. 328

SAT's ADJUSTED FOR % TAKING EXAM AND THEIR MEDIAN CLASS RANK

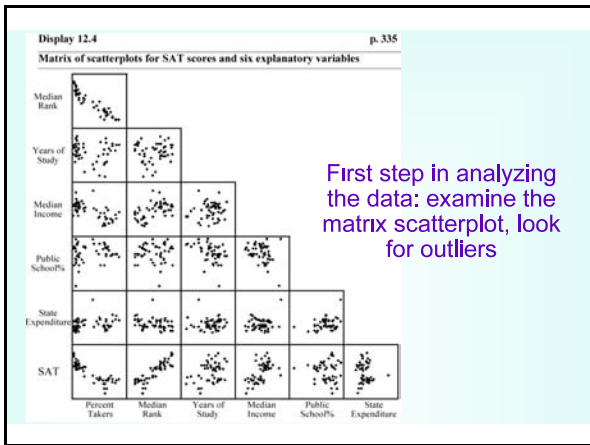
Rank	State	Adjusted SAT Score
1	New Hampshire	92
2	Massachusetts	88
3	Connecticut	88
4	California	88
5	Washington	88
6	Minnesota	88
7	Illinois	88
8	New York	88
9	Michigan	88
10	Wisconsin	88
11	North Dakota	88
12	South Dakota	88
13	Nebraska	88
14	Kansas	88
15	Idaho	88
16	Montana	88
17	Utah	88
18	Alaska	88
19	Alaska	88
20	Alaska	88
21	Alaska	88
22	Alaska	88
23	Alaska	88
24	Alaska	88
25	Alaska	88
26	Alaska	88
27	Alaska	88
28	Alaska	88
29	Alaska	88
30	Alaska	88
31	Alaska	88
32	Alaska	88
33	Alaska	88
34	Alaska	88
35	Alaska	88
36	Alaska	88
37	Alaska	88
38	Alaska	88
39	Alaska	88
40	Alaska	88
41	Alaska	88
42	Alaska	88
43	Alaska	88
44	Alaska	88
45	Alaska	88
46	Alaska	88
47	Alaska	88
48	Alaska	88
49	Alaska	88
50	Alaska	88

Case Study 12.1  
Final model  
SAT Scores =  $f(\%$   
Taking exam,  
median class rank)  
or  $(\%$  taking exam,  
rank and  
expenditure)  
Result: NH is #1,  
Massachusetts is  
11 or 32  
(expenditure)



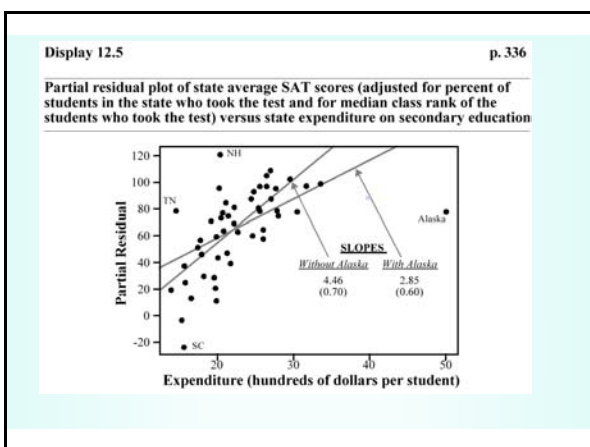
## Slide 15

NOTES:



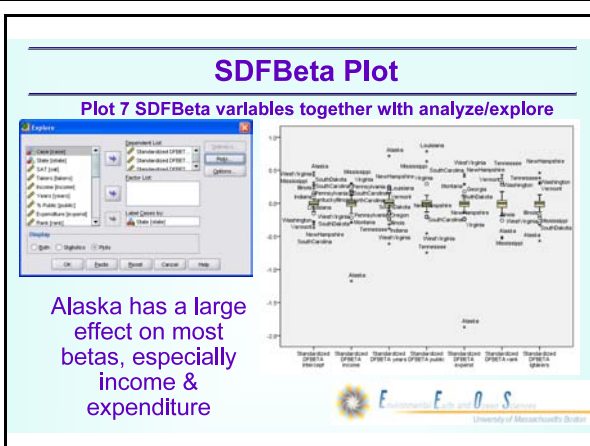
## Slide 16

NOTES:



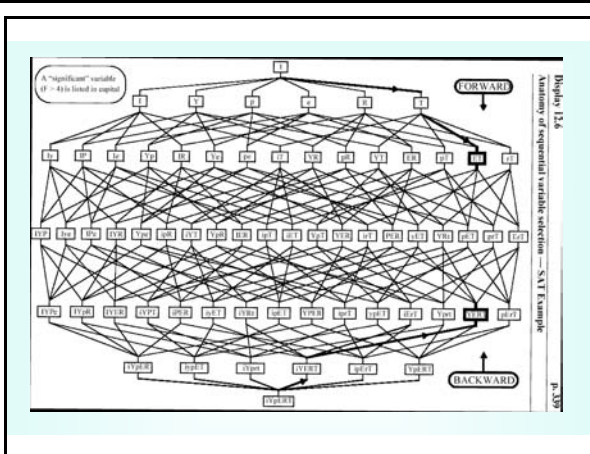
## Slide 17

NOTES:



## Slide 18 SDFBeta Plot

NOTES:

**Slide 19**

NOTES:

**All possible regression models**

$R^2$  and Adjusted  $R^2$  choose models with TOO many parameters

- Mallow's  $C_p$ 
  - $C_p = p + (n-p)(\sigma^2 - \sigma^2_{full}) / \sigma^2_{full}$
- Akaike information content
- Schwarz Bayesian Information Content (BIC)
  - $\ln(\log(\sigma^2) + p \log(n))$ .
  - Can be used to calculate posterior probabilities
- Neither available in SPSS without syntax
  - All are available in SPSS syntax
    - \Statistics SELECTION
  - The BIC in SPSS is different from Sleuth

**Slide 20 All possible regression models**

NOTES:

**SPSS selection criteria**

In syntax only: \Statistics SELECTION

Model Summary <sup>a</sup>									
Change Statistics									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	df1	df2	Sig. F Change	Selection Criteria
1	.388 <sup>a</sup>	.148	.086	1.32344	.148	1	610	.4	28.216
2	.728 <sup>b</sup>	.532	.492	1.00379	.384	14	341	.000	7.077
3	.894 <sup>c</sup>	.758	.737	.87174	.266	44	935	.000	-26.298

<sup>a</sup> Predictors: (Constant), In (Precipitation), In (Area), In (Runoff), In (Discharge), In (Deposition), In (NO<sub>3</sub> precipitation)  
<sup>b</sup> Predictors: (Constant), In (Precipitation), In (Area), In (Runoff), In (Discharge), In (Deposition), In (NO<sub>3</sub> precipitation), In (Density)  
<sup>c</sup> Predictors: (Constant), In (Precipitation), In (Area), In (Runoff), In (Discharge), In (Deposition), In (NO<sub>3</sub> precipitation), In (Density)  
<sup>d</sup> Dependent Variable: In (NO<sub>3</sub>)

**Slide 21 SPSS selection criteria**

NOTES:

## Bayes' Theorem

Larsen & Marx 2nd Edition (2001)

**Bayes' Theorem (Theorem 26.2 p. 65)**

Let  $\{A_i\}_{i=1}^n$  be a set of  $n$  events,  
each with positive probability,  
that partition  $S$  in such a way that

$$\bigcup_{i=1}^n A_i = S$$

and  $A_i \cap A_j = \emptyset$  for  $i \neq j$ .

For any event  $B$  (also defined on  $S$ ),  
where  $P(B) > 0$ ,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

for any  $i = 1, 2, \dots, n$ .

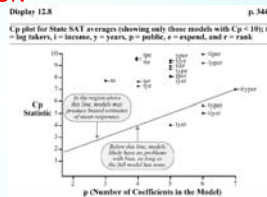
## Slide 22 Bayes' Theorem

NOTES:

## All possible regressions

All regression models in SAS, R & Matlab, not SPSS

- SAS procedure
- SPSS
  - /STATISTICS COEFF OUTS CI R ANOVA COLLIN TOL CHANGE SELECTION
- Matlab
  - Stixbox



## Slide 23 All possible regressions

NOTES:

## SPSS regression syntax

/STATISTICS ALL or /STATISTICS SELECTION

\* Case 1201- note the /STATISTICS=SELECTION.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/SELECT= istate NE 2

/MISSING LISTWISE

/STATISTICS ALL

/CRITERIA=PIN(.05) POUT(.10) CIN(95)

/NOORIGIN

/DEPENDENT sat

/METHOD=BACKWARD lgtakern income years public expend rank

/PARTIALPLOT ALL

/SCATTERPLOT=(\*ZRESID ,\*ZPRED )

/RESIDUALS ID( state )

/SAVE PRED COOK MCIN ICIN RESID .

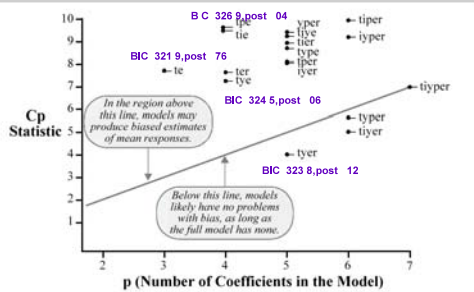
## Slide 24 SPSS regression syntax

NOTES:



Display 12.8

Cp plot for State SAT averages (showing only those models with  $C_p < 10$ ); t = log takers, i = income, y = years, p = public, e = expend, and r = rank

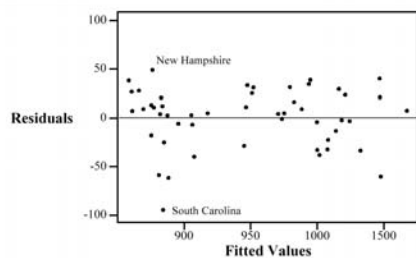


## Slide 25

NOTES:

Display 12.13

Scatterplot of residuals versus fitted values from the regression of state SAT average on percent takers and median class rank of takers



## Slide 26

NOTES:

## Trouble assessing significance

Display 12.13, page 363

Statistical measures for the contribution of expenditure to different models (Alaska removed)

Model	t-Statistic	2-sided p-value	% variation explained by expenditure	Previous Residual	Total Variation
t	0.27	.79	0.2	0.2	
i+e	0.40	.69	0.3	0.2	
y+e	0.86	.39	1.6	1.4	
p+e	0.18	.86	0.1	0.1	
ti+e	4.78	.00002	49.6	7.4	
ti	6.04	.000002	76.4	8.4	
ti+y	0.07	.95	0.0	0.0	
ti+p	0.11	.92	0.0	0.0	
ti+e	4.88	.000001	52.9	6.8	
ti+y+e	3.88	.000005	76.7	8.1	
ti+p+e	1.05	.30	2.5	2.1	
ti+y+e	4.20	.0001	78.2	4.3	
ti+y+e	5.11	.000001	82.8	6.3	
ti+p+e	6.07	.0000001	80.5	9.4	
ti+y+e	5.91	.0000004	77.7	8.2	
ti+y+e	6.17	.0000002	83.5	8.4	
ti+y+e	6.92	.0000001	82	8.9	
ti+y+e	4.32	.000008	43.1	4.2	
ti+y+e	5.20	.000005	60.4	6.2	
ti+y+e	5.27	.000001	70.1	8.0	
ti+y+e	5.59	.0000001	76.1	7.9	
ti+y+e	5.92	.0000004	78.7	7.8	
ti+y+e	4.80	.000002	52.3	5.3	
ti+y+e	4.78	.000002	51.9	5.3	
ti+y+e	5.23	.0000001	62.1	5.6	
ti+y+e	6.25	.00000001	84.5	8.8	
ti+y+e	4.27	.0001	42.4	4.1	
ti+y+e	4.23	.00009	41.7	4.1	
ti+y+e	5.00	.00001	58.2	5.1	
ti+y+e	5.94	.0000004	82.0	8.0	
ti+y+e	5.13	.0000005	61.3	5.4	
ti+y+e	4.64	.00003	51.2	4.5	

Expenditure is correlated with other explanatory variables, so the significance (and magnitude) depends on the other variables in the model. Oregon ranks 31st in average SAT but 46th based on money spent.

## Slide 27 Trouble assessing significance

NOTES:

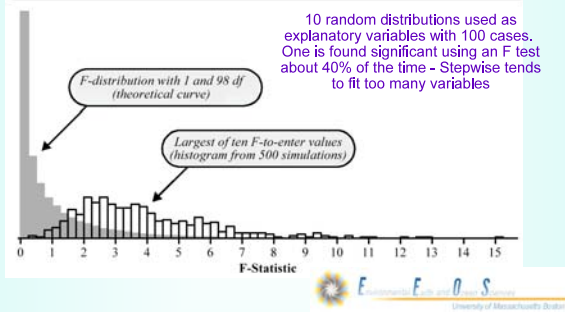
### Overfitting: why stepwise procedures should not be used to estimate p values.

### Slide 28 Overfitting: why stepwise procedures should not be used to estimate p values.

NOTES:

Display 12.7

Simulated distribution of the largest of ten F-statistics



### Slide 29

NOTES:

### Gallagher's overfitting.sps

```
* Overfitting simulation, inspired by
* Nontechnical Introduction to Overfitting in Regression-Type Models, Babyak (2004).
* Michael A Babyak. What You See May Not Be What You Get: A Brief, Nontechnical
* Introduction to Overfitting in Regression-Type Models.
* Psychosom Med 2004 66: 411-421.
* Written by E Gallagher, revised 4/12/05.
* Generate 100 cases, with 32 normally distributed variates.
new file.
input program.
loop #1 = 1 to 100.
  COMPUTE V1 = RV.normal (0,1) .
  COMPUTE V2 = RV.normal (0,1) .
  ..
  COMPUTE V32 = RV.normal (0,1) .
end case.
end loop.
end file.
end input program.
formats V1 to V32 (f4,2).
exe.
```

### Slide 30 Gallagher's overfitting.sps

NOTES:

## Results of Stepwise Selection

### 31 Random predictor variables

Coefficients <sup>a</sup>													
Step	Model	Unstandardized Coefficients				Standardized Coefficients				95% Confidence Interval for B			
		B	SE	Beta	t	B	SE	Beta	t	Lower Bound	Upper Bound	Lower Bound	Upper Bound
1	(Constant)	273	284			273	284						
	V1	273	284			273	284						
	V2	273	284			273	284						
	V3	273	284			273	284						
	V4	273	284			273	284						
2	(Constant)	298	282			298	282						
	V1	282	284			282	284						
	V2	282	284			282	284						
	V3	282	284			282	284						
	V4	282	284			282	284						
3	(Constant)	313	289			313	289						
	V1	411	283			425	4432			236	587		
	V2	257	283			253	2836			242	2802		
	V3	237	281			254	2566			251	2605		
	V4	192	287			237	2311			238	2324		
4	(Constant)	373	3088			373	3088						
	V1	432	282			434	4583			237	582		
	V2	282	285			282	2850			421	431		
	V3	226	280			243	2511			214	2647		
	V4	280	286			224	2369			225	2395		
5	(Constant)	382	3088			382	3088						
	V1	432	282			434	4583			237	582		
	V2	282	285			282	2850			421	431		
	V3	226	280			243	2511			214	2647		
	V4	280	286			224	2369			225	2395		

a. Dependent Variable: vt

25 (Constant)

V1

V2

V3

V4

V5

183

174

242

225

179

145

284

193

214

225

117

1166

279

284

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282

282</

Backward  
(added V23, V19) →

## Slide 31 Results of Stepwise Selection

NOTES:

## Harrell (2002, p. 56-57) on stepwise

Harrell's conclusion: Don't use stepwise!

- It yields  $R^2$  values that are biased high
- F and  $\chi^2$  distributions don't have their claimed distributions
- SE of regression coefficients are biased low and CI's and predicted values that are falsely narrow
- P-values too small
- Regression coefficients biased high in absolute value and need shrinkage.
- Rather than solving the problem of collinearity, variable selection is made arbitrary by collinearity
- It allows us not to think about the problem

## Slide 32 Harrell (2002, p. 56-57) on stepwise

NOTES:

## Multicollinearity, collinearity

- If the explanatory variables are strongly correlated
  - The regression coefficient estimates have a huge variance
  - They can change in sign and significance with a slight change in the data, bouncing betas
- Diagnostics (Variance Inflation Factors (VIF's) - see next page)
- Solutions for multicollinearity for OLS regression
  - Reduce the number of explanatory variables using theory & insight into the field
  - Ridge regression
  - Principal components regression



## Slide 33 Multicollinearity, collinearity

NOTES:

### Collinearity [multicollinearity]

- When one or more predictors can be predicted by other predictors, the standard error of the regression coefficients can be inflated and the corresponding tests have reduced power
- Assessed with Variance inflation factors (VIF) or tolerance
  - $VIF_i = 1 / (1 - R_i^2)$ , where  $R_i^2$  is the squared multiple correlation coefficient between explanatory variable 'i' and the other explanatory variables
  - Neter et al. (1996): **VIF's > 10** are cause for concern (but smaller VIF's can also be a problem)

### Slide 34 Collinearity [multicollinearity]

NOTES:

### Ways of detecting multicollinearity

Marayuma (1998, p. 64)

- When the variance (standard errors) of beta weights is large
- When signs on beta weights are inappropriate [e.g., larger classes  $\Rightarrow$  higher test scores]
- When regression weights and signs change radically upon the addition or removal of single variables
- When the Variance Inflation Factor is high (VIF > 6 or 7 as a very rough rule)
- When simple correlations are > 0.8-0.9
- When correlations among predictor variables >  $R^2$  for response with all predictor variables

### Slide 35 Ways of detecting multicollinearity

NOTES:

### Shooting Down the "More Guns, Less Crime" Hypothesis

Ian Ayres\* & John J. Donohue III

1250  
STANFORD LAW REVIEW  
[Vol. 57:1250]

The question merits investigation. Why do the results of the Lott model (Table 2) support the Lott thesis more than the results of the Zheng model (Table 2)? The answer seems to be to be somewhat surprising. As Table 2 documents, the two models include some different explanatory variables that one might think could have important implications. For example, Lott controls for population, density, and income per capita in the past, while Zheng controls for police, poverty, unemployment, and alcohol consumption. But these differences in the explanatory controls turn out to be largely unimportant. Interestingly, as we will discuss in the next section, what drives the entire difference between Tables 2 and 3 is that Lott includes a large number of potentially duplicative demographic variables. Indeed, the story is so extensive as to make multicollinearity a serious issue.<sup>62</sup> Specifically, while Zheng's model controls for the percent black and three age groupings, Lott's has thirty-six separate demographic percentages, breaking down each of these different race categories—black, white, and neither black nor white—and both sexes into six separate age categories from age ten up. The similarity of the results to the inclusion or exclusion of an array of highly collinear demographic variables serves as a cautionary tale for those who would rely upon panel data models of crime. Probably no one examining either Weingang's work or that of Lott and Donohue would suspect that conclusions reached from these models would be sensitive to these seemingly innocuous demographic controls.

crime from the Table 3 regressions, one finds utterly bizarre results. For example, an increase of one percentage point in the percentage of black males aged 30-39 would be expected to almost double the violent crime rate, while a similar increase in the percentage of black males aged 40-49 would lead to a drop in violent crime of 60%. Similarly, increasing the percentage of black males aged 50-64 would cause violent crime to jump by 145%, but increasing the percentage of black males over age 65 would lead to a 78% decline in violent crime. These nonsense results prevent us from understanding why the demographic controls can influence the estimates of shall-issue adoption so strongly.

Adding too many covariates can destroy a regression

### Slide 36

NOTES:

### Solutions to multicollinearity

- If the goal of the model is to produce predicted values for one analysis, then multicollinearity is **not** a problem. All variables can be included.
  - However, if the equation is to be used for new data, then the model will be badly overfitted, the predicted values will be biased
  - Significant coefficients could be spurious or nonsense
- Solutions
  - Variable selection procedures (cluster analysis of variables)
  - Principal components regression
    - Use principal component scores as explanatory variables
    - Principal component scores are orthogonal (uncorrelated)
  - Ridge regression
  - Structural equation modeling

### Slide 37 Solutions to multicollinearity

NOTES:

### Ridge regression

Available as a macro in SPSS, LISREL (not AMOS); increase variance for variables not covariance

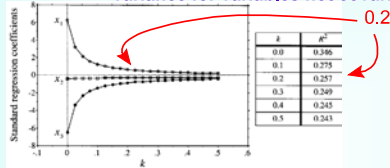


Figure 19.8 Ridge trace diagram showing the estimates of the standardized regression coefficients  $b_j$  explanatory variables  $x_1$  to  $x_3$  as a function of  $k$ . Table: decrease of  $R^2$  as a function of  $k$ .

A ridge regression parameter,  $k$ , is chosen using the ridge trace diagram ( $k=0.2$  in the above example [the base of the horn] from Draper & Smith) that 'shrinks' the regression coefficients, especially those coefficients (Beta's) that are strongly correlated. This offers a partial solution to the problem of collinearity.

### Slide 38 Ridge regression

NOTES:

### Case 11.2 Gender discrimination

### Slide 39 Case 11.2 Gender discrimination

NOTES:

Sex Discrimination Data

Beginning Salary	Year	Sex	Age	Education	Experience
10000	1970	0	40	12	10
10000	1970	0	41	13	11
10000	1970	0	42	14	12
10000	1970	0	43	15	13
10000	1970	0	44	16	14
10000	1970	0	45	17	15
10000	1970	0	46	18	16
10000	1970	0	47	19	17
10000	1970	0	48	20	18
10000	1970	0	49	21	19
10000	1970	0	50	22	20
10000	1970	0	51	23	21
10000	1970	0	52	24	22
10000	1970	0	53	25	23
10000	1970	0	54	26	24
10000	1970	0	55	27	25
10000	1970	0	56	28	26
10000	1970	0	57	29	27
10000	1970	0	58	30	28
10000	1970	0	59	31	29
10000	1970	0	60	32	30
10000	1970	0	61	33	31
10000	1970	0	62	34	32
10000	1970	0	63	35	33
10000	1970	0	64	36	34
10000	1970	0	65	37	35
10000	1970	0	66	38	36
10000	1970	0	67	39	37
10000	1970	0	68	40	38
10000	1970	0	69	41	39
10000	1970	0	70	42	40
10000	1970	0	71	43	41
10000	1970	0	72	44	42
10000	1970	0	73	45	43
10000	1970	0	74	46	44
10000	1970	0	75	47	45
10000	1970	0	76	48	46
10000	1970	0	77	49	47
10000	1970	0	78	50	48
10000	1970	0	79	51	49
10000	1970	0	80	52	50
10000	1970	0	81	53	51
10000	1970	0	82	54	52
10000	1970	0	83	55	53
10000	1970	0	84	56	54
10000	1970	0	85	57	55
10000	1970	0	86	58	56
10000	1970	0	87	59	57
10000	1970	0	88	60	58
10000	1970	0	89	61	59
10000	1970	0	90	62	60
10000	1970	0	91	63	61
10000	1970	0	92	64	62
10000	1970	0	93	65	63
10000	1970	0	94	66	64
10000	1970	0	95	67	65
10000	1970	0	96	68	66
10000	1970	0	97	69	67
10000	1970	0	98	70	68
10000	1970	0	99	71	69
10000	1970	0	100	72	70
10000	1970	0	101	73	71
10000	1970	0	102	74	72
10000	1970	0	103	75	73
10000	1970	0	104	76	74
10000	1970	0	105	77	75
10000	1970	0	106	78	76
10000	1970	0	107	79	77
10000	1970	0	108	80	78
10000	1970	0	109	81	79
10000	1970	0	110	82	80
10000	1970	0	111	83	81
10000	1970	0	112	84	82
10000	1970	0	113	85	83
10000	1970	0	114	86	84
10000	1970	0	115	87	85
10000	1970	0	116	88	86
10000	1970	0	117	89	87
10000	1970	0	118	90	88
10000	1970	0	119	91	89
10000	1970	0	120	92	90
10000	1970	0	121	93	91
10000	1970	0	122	94	92
10000	1970	0	123	95	93
10000	1970	0	124	96	94
10000	1970	0	125	97	95
10000	1970	0	126	98	96
10000	1970	0	127	99	97
10000	1970	0	128	100	98
10000	1970	0	129	101	99
10000	1970	0	130	102	100
10000	1970	0	131	103	101
10000	1970	0	132	104	102
10000	1970	0	133	105	103
10000	1970	0	134	106	104
10000	1970	0	135	107	105
10000	1970	0	136	108	106
10000	1970	0	137	109	107
10000	1970	0	138	110	108
10000	1970	0	139	111	109
10000	1970	0	140	112	110
10000	1970	0	141	113	111
10000	1970	0	142	114	112
10000	1970	0	143	115	113
10000	1970	0	144	116	114
10000	1970	0	145	117	115
10000	1970	0	146	118	116
10000	1970	0	147	119	117
10000	1970	0	148	120	118
10000	1970	0	149	121	119
10000	1970	0	150	122	120
10000	1970	0	151	123	121
10000	1970	0	152	124	122
10000	1970	0	153	125	123
10000	1970	0	154	126	124
10000	1970	0	155	127	125
10000	1970	0	156	128	126
10000	1970	0	157	129	127
10000	1970	0	158	130	128
10000	1970	0	159	131	129
10000	1970	0	160	132	130
10000	1970	0	161	133	131
10000	1970	0	162	134	132
10000	1970	0	163	135	133
10000	1970	0	164	136	134
10000	1970	0	165	137	135
10000	1970	0	166	138	136
10000	1970	0	167	139	137
10000	1970	0	168	140	138
10000	1970	0	169	141	139
10000	1970	0	170	142	140
10000	1970	0	171	143	141
10000	1970	0	172	144	142
10000	1970	0	173	145	143
10000	1970	0	174	146	144
10000	1970	0	175	147	145
10000	1970	0	176	148	146
10000	1970	0	177	149	147
10000	1970	0	178	150	148
10000	1970	0	179	151	149
10000	1970	0	180	152	150
10000	1970	0	181	153	151
10000	1970	0	182	154	152
10000	1970	0	183	155	153
10000	1970	0	184	156	154
10000	1970	0	185	157	155
10000	1970	0	186	158	156
10000	1970	0	187	159	157
10000	1970	0	188	160	158
10000	1970	0	189	161	159
10000	1970	0	190	162	160
10000	1970	0	191	163	161
10000	1970	0	192	164	162
10000	1970	0	193	165	163
10000	1970	0	194	166	164
10000	1970	0	195	167	165
10000	1970	0	196	168	166
10000	1970	0	197	169	167
10000	1970	0	198	170	168
10000	1970	0	199	171	169
10000	1970	0	200	172	170
10000	1970	0	201	173	171
10000	1970	0	202	174	172
10000	1970	0	203	175	173
10000	1970	0	204	176	174
10000	1970	0	205	177	175
10000	1970	0	206	178	176
10000	1970	0	207	179	177
10000	1970	0	208	180	178
10000	1970	0	209	181	179
10000	1970	0	210	182	180
10000	1970	0	211	183	181
10000	1970	0	212	184	182
10000	1970	0	213	185	183
10000	1970	0	214	186	184
10000	1970	0	215	187	185
10000	1970	0	216	188	186
10000	1970	0	217	189	187
10000	1970	0	218	190	188
10000	1970	0	219	191	189
10000	1970	0	220	192	190
10000	1970	0	221	193	191
10000	1970	0	222	194	192
10000	1970	0	223	195	193
10000	1970	0	224	196	194
10000	1970	0	225	197	195
10000	1970	0	226	198	196
10000	1970	0	227	199	197
10000	1970	0	228	200	198
10000	1970	0	229	201	199
10000	1970	0	230	202	200
10000	1970	0	231	203	201
10000	1970	0	232	204	202
10000	1970	0	233	205	203
10000	1970	0	234	206	204
10000	1970	0	235	207	205
10000	1970	0	236	208	206
10000	1970	0	237	209	207
10000	1970	0	238	210	208
10000	1970	0	239	211	209
10000	1970	0	240	212	210
10000	1970	0	241	213	211
10000	1970	0	242	214	212
10000	1970	0	243	215	213
10000	1970	0	244	216	214
10000	1970	0	245	217	215
10000	1970	0	246	218	216
10000	1970	0	247	219	217
10000	1970	0	248	220	218
10000	1970	0	249	221	219
10000	1970	0	250	222	220
10000	1970	0	251	223	221
10000	1970	0	252	224	222
10000	1970	0	253	225	223
10000	1970	0	254	226	224
10000	1970	0	255	227	225
10000	1970	0	256	228	226
10000	1970	0	257	229	227
10000	1970	0	258	230	228
10000	1970	0	259	231	229
10000	1970	0	260	232	230
10000	1970	0	261	233	231
10000	1970	0	262	234	232
10000	1970	0	263	235	233
10000	1970	0	264	236	234
10000	1970	0	265	237	235
10000	1970	0	266	238	236
10000	1970	0	267	239	237
10000	1970	0	268	240	238
10000	1970	0	269	241	239
10000	1970	0	270	242	240
100					

### SPSS output using forward, backward or stepwise

Model Summary<sup>a</sup>

Model	Selection Criteria			
	Akaike Information Criterion	Amemiya Prediction Criterion	Mallows' Prediction Criterion	Schwarz Bayesian Criterion
1	-395.813 <sup>a</sup>	858	38.600	-390.747
2	-407.042 <sup>b</sup>	761	23.681	-399.444
3	-410.713 <sup>c</sup>	731	19.134	-400.582
4	-415.357 <sup>d</sup>	691	13.330	-403.294
5	-419.552 <sup>e</sup>	665	9.705	-404.356
6	-421.539 <sup>f</sup>	651	7.718	-408.876
7	-427.248 <sup>g</sup>	612	2.501	-412.053

a. Predictors: (Constant), f (e<sup>2</sup>)  
b. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e)  
c. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e), v (s \* x)  
d. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e), v (s \* x), k (a\*x)  
e. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e), v (s \* x), k (a\*x), x (Experience)  
f. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e), k (a\*x), x (Experience)  
g. Predictors: (Constant), f (e<sup>2</sup>), n (s \* e), k (a\*x), x (Experience), q (e\*x)  
h. Dependent Variable: ln (Salary)

### Slide 43 SPSS output using forward, backward or stepwise

NOTES:

### Has gender equity really been rejected?

Campbell & Kenny: statistical equating often produces gender discrimination when there is none, and racial differences when there are none

### Slide 44 Has gender equity really been rejected?

NOTES:

### Statistical Equating & RTM

Campbell & Kenny: The regression artifact

- The sophomore jinx
- Spontaneous remission of depression
- Misclassification of individuals using standardized tests
- Perhaps:
  - Ashland cancer study
  - Washington D.C. vouchers
  - Sander's analysis of African-American failure on the bar exam
- Statistical equating
  - Regression to the mean leads to a bias in estimating gender differences using "equating"
  - Page 84: Ethnic differences in intellectual ability:
    - "We believe that the bias in statistical equating for ethnic differences in achievement and intelligence testing is underadjustment"

### Slide 45 Statistical Equating & RTM

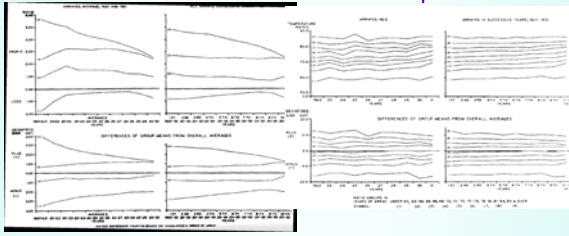
NOTES:

### Poor Horace Secrist (1933)

Identify companies that had lower than average profits and invest in them; he was aware of RTM  
Profits (left), temperature (right)

Stocks

Temperature



### Slide 46 Poor Horace Secrist (1933)

NOTES:

### Hotelling's (1933) JASA review

- Business varies, but average temperatures don't vary nearly as much
  - Secrist chose cities spread out throughout the country and looked at interannual variability
  - Small year-to-year variations compared to the big city-to-city variations
- Secrist rebuttal (1934)



### Slide 47 Hotelling's (1933) JASA review

NOTES:

### Hotelling's (1934) rejoinder

Quoted in Stigler's "Statistics on the Table"

"To 'prove' such a mathematical result [regression to the mean in annual reports] by a costly and prolonged numerical study of many kinds of business profit and expense ratios is analogous to proving the multiplication table by arranging elephants in rows and columns, and then doing the same for numerous other kinds of animals. The performance, though perhaps entertaining, and having a certain pedagogical value, is not an important contribution to either zoology or to mathematics."

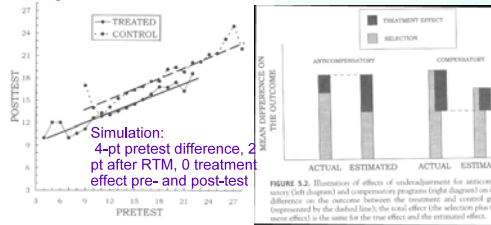
### Slide 48 Hotelling's (1934) rejoinder

NOTES:



## Statistical Equating

Effects on gender bias & racial differences  
**"Including a covariate, like socioeconomic status, can produce a racial or gender bias, when none really exists!"**



URE 5.1. Parallel regression lines for treatment and control groups ed by statistical equating (multiple regression).

## Slide 49 Statistical Equating

NOTES:

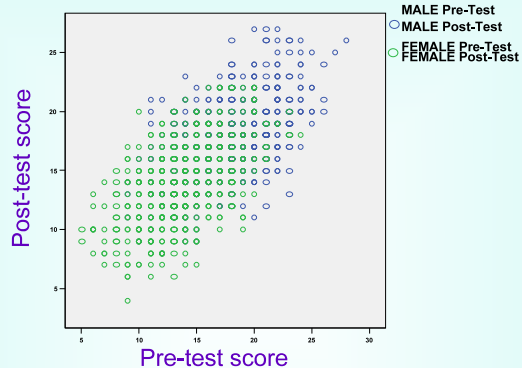
## A hypothetical test of gender effects

Read Campbell & Kenny Chapters 4 & 5

- Are women inferior in mathematics?
- Randomly select 500 women & 500 men for admission to a intense workshop on advanced mathematics.
- Give both groups a pretest of mathematical ability
  - In the simulation (rtm-ck.sps) generate test scores by 4 tosses of a die. Assign males 4 units higher score in both pre & post test
    - Males: sum of 4 dice + 4
    - Females: sum of 4 dice + 0.
- Assume that the workshop does NOTHING to improve ability for either group
- Retest each student, the post-test, which is modeled to have a correlation of 0.5 between pre- & post-test
  - 2 dice the same, 2 new dice throws for each student
- Test whether males did better than females in this advanced workshop, even after controlling for their previous math background

## Slide 50 A hypothetical test of gender effects

NOTES:



## Slide 51

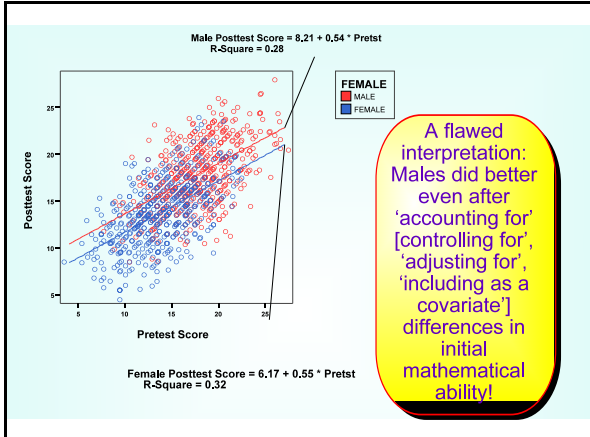
NOTES:

## Slide 52



NOTES:

## Slide 53



NOTES:

**Flawed interpretation: Females score 2 points less ( $1.9 \pm 0.4$ ) on the post-test, after ‘supposedly’ controlling for the effect of previous mathematical ability ( $p < 10^{-18}$ )**

		Coefficients <sup>a</sup>					
Model		Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	
1	(Constant)	8.148	.498		16.777	7.8E-036	7.194 9.099
	Pretest Score	.545	.028	.542	20.736	1.1E-079	.493 .596
	FEMALE	-1.927	.209	-.240	-9.198	2.1E-019	-2.338 -1.516

a. Dependent Variable: Posttest Score

But: the simulation is set so that the workshop didn't have any effect on either group!

**Slide 54 Flawed interpretation: Females score 2 points less ( $1.9 \pm 0.4$ ) on the post-test, after ‘supposedly’ controlling for the effect of previous mathematical ability ( $p < 10^{-18}$ )**

NOTES:

### Classic Analysis of covariance

Huge Male-female difference in post-workshop scores, after 'controlling' for pre-test ability

- \* Classic analysis of covariance (ANCOVA)
- \* to test for treatment effect
- \* with pretest as the covariate.

ANOVA postst BY treat(0,1) with pretst  
/STATISTICS=ALL.

		ANOVA <sup>a,b</sup>		Unique Method				
Posttest Score	Covariates	Pretest Score	Sum of Squares	df	Mean Square	F	Sig.	B
			3630.112	1	3630.112	429.995	1.08E-079	.545
	Main Effects	FEMALE	714.243	1	714.243	84.604	2.08E-019	
	Model		7654.148	2	3827.074	453.326	9.44E-141	
	Residual		8416.888	997	8.442			
	Total		16071.036	999	16.087			

a. Posttest Score by FEMALE with Pretest Score

b. All effects entered simultaneously

### Slide 55 Classic Analysis of covariance

NOTES:

### Repeated measures designs (Chapter 16) produce the correct solution: No effect of gender on post-test

There is no pre-test to post-test x gender interaction

Tests of Within-Subjects Effects

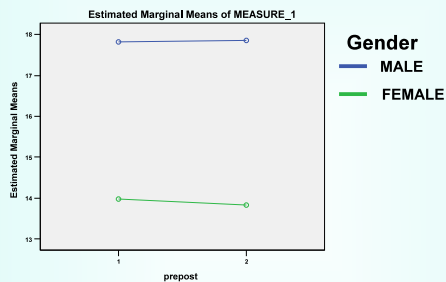
Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
prepost	1.458	1	1.458	.266	.606
	Greenhouse-Geisser	1.458	1.000	1.458	.266
	Huynh-Feldt	1.458	1.000	1.458	.266
	Lower-bound	1.458	1.000	1.458	.266
prepost * treat	4.232	1	4.232	.771	.380
	Greenhouse-Geisser	4.232	1.000	4.232	.771
	Huynh-Feldt	4.232	1.000	4.232	.771
	Lower-bound	4.232	1.000	4.232	.771
Error(prepost)	5476.310	998	5.487		
	Greenhouse-Geisser	5476.310	998.000	5.487	
	Huynh-Feldt	5476.310	998.000	5.487	
	Lower-bound	5476.310	998.000	5.487	

### Slide 56 Repeated measures designs (Chapter 16) produce the correct solution: No effect of gender on post-test

NOTES:

### Profiles from Repeated Measures ANOVA



### Slide 57 Profiles from Repeated Measures ANOVA

NOTES:

### Change score: Do paired $t$ tests on males & females separately

Paired Samples Test									
Paired Differences									
95% Confidence Interval of the Difference									
	Mean	Std. Deviation	Std. Error	Lower	Upper	t	df	Sig. (2-tailed)	
Pair 1 Pretest Score - Posttest Score	-.038	3.365	.150	-.334	.258	-.253	499	.801	

Paired Samples Test									
Paired Differences									
95% Confidence Interval of the Difference									
	Mean	Std. Deviation	Std. Error	Lower	Upper	t	df	Sig. (2-tailed)	
Pair 1 Pretest Score - Posttest Score	.146	3.260	.146	-.140	.432	1.002	499	.317	

### Slide 58 Change score: Do paired $t$ tests on males & females separately

NOTES:

### Why didn't regression & ANCOVA work?

See Cambell & Kenny (Ch 4-5) for full analysis

- Whenever there is less than perfect correlation between the covariate and the response, the effect of the covariate on the response is **not** removed by regression (=Analysis of covariance)
- This is due to regression to the mean
- Since the correlation between pre-test and post-test was set at  $r=0.5$ , only 50% of the pre-test effect can be 'explained' or accounted for by multiple regression
- Whenever the covariate is less than perfectly correlated with the response, multiple regression does not fully 'control for' or 'account for' the effects of the covariate.
  - Note that if the pre-test score had a correlation with the post-test score of 0.25, then only 1/4 of the pre-test difference would be accounted for by including pre-test as a covariate. There would be a 3-point advantage for males after including pre-test as a covariate

### Slide 59 Why didn't regression & ANCOVA work?

NOTES:

### Galton's regression to the mean

Son's height 1" taller than father's,  $r=0.5$ ,  $SD=2.5$ "

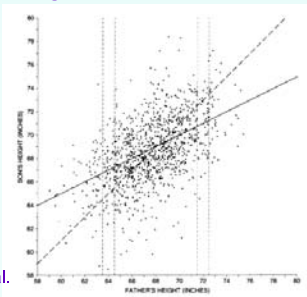


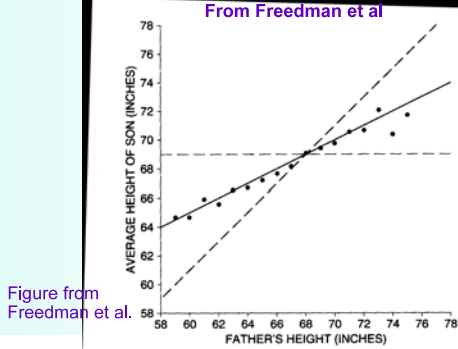
Figure from Freedman et al.

### Slide 60 Galton's regression to the mean

NOTES:

### RTM effect $\propto 1/r$

From Freedman et al



### Slide 61 RTM effect $\propto 1/r$

NOTES:

### Galton squeeze

If you naively use pretest as a covariate, you'll introduce an artifact in the analysis.

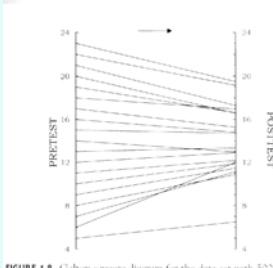


FIGURE 1.8. Galton squeeze diagram for the data set with 500 cases using pretest to predict posttest.

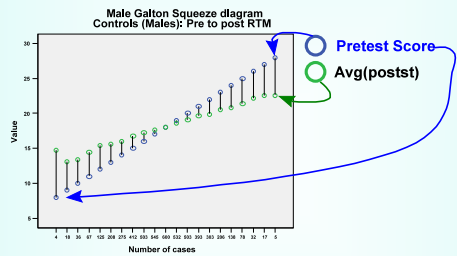
Using pre-test to predict post-test will be subject to 'regression to the mean.' If  $r$  between pre- and post is 0.5, only half of the pre-test effect will be accounted for.

### Slide 62 Galton squeeze

NOTES:

### Galton squeeze

Only about  $\frac{1}{2}$  the pretest effect is removed if the correlation is 0.5 between covariate and response. The other half appears as the male-female difference in the post-test scores

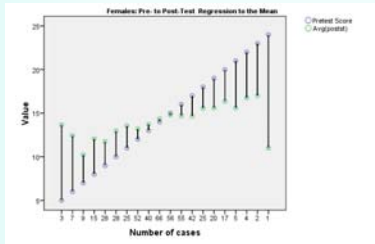


### Slide 63 Galton squeeze

NOTES:

### Galton squeeze, if $r=0.25$

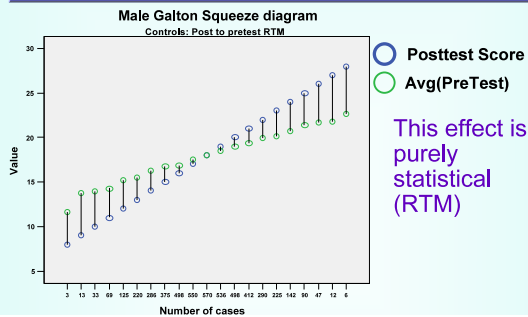
Only about  $\frac{1}{4}$  of the pretest effect is removed if the correlation is 0.25 between covariate and response. The other  $\frac{3}{4}$  appears as the male-female difference in the post-test scores



### Slide 64 Galton squeeze, if $r=0.25$

NOTES:

### Regression to the mean applies forward & backward



### Slide 65 Regression to the mean applies forward & backward

NOTES:

### The Regression Fallacy

Stigler (1999) Chapter 9 Regression toward the mean

- "I suspect that **the regression fallacy** is the most common fallacy in the statistical analysis of economic data." Milton Friedman (1992) [emphasis added]
- "The recurrence of **regression fallacies** is testimony to its subtlety, deceptive simplicity, and I speculate, to the wide use of the word regression to describe least squares fitting of curves, lines, and surfaces. Researchers may err because they believe they know about regression, yet in truth have never fully appreciated how Galton's concept works. History suggests that this will not change soon. Galton's achievement remains one of the most attractive triumphs in the history of statistics, but it is one that each generation must learn to appreciate anew, on that seemingly never loses its power to surprise."



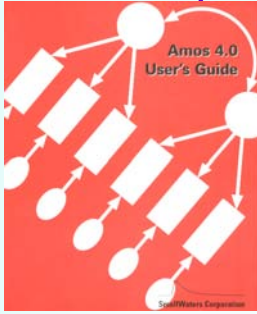
### Slide 66 The Regression Fallacy

NOTES:

<div data-bbox="298 170 738 207" data-label="Section-Header"> <h3>Statistical matching &amp; equating</h3> </div> <div data-bbox="318 214 711 239" data-label="Text"> <p>Creates 'bias' in assessing treatment effects</p> </div> <div data-bbox="230 245 773 489" data-label="List-Group"> <ul style="list-style-type: none"> <li>• <b>Matching:</b> If a covariate (e.g., pretest scores) is used to select groups, and there is less than perfect correlation between pre-and post-test assessments, then there will be regression to the mean. <ul style="list-style-type: none"> <li>▸ Each group will regress to its own mean</li> <li>▸ <b>The regression to the mean effect will produce a treatment difference due to the treatment when none may have existed.</b></li> <li>▸ Scaling College math performance vs. Gender based on categorical variables like (high school algebra I, Algebra I &amp; II, Algebra I, II &amp; Calculus) is still prone to the regression artifact</li> </ul> </li> <li>• <b>Equating:</b> If the covariate is weakly correlated with the presumed factor that it is controlling for (SES), &amp; the covariate is positively associated with the response, then differences among groups can be magnified by the addition of the covariate.</li> </ul> </div> <div data-bbox="532 499 786 541" data-label="Image"> </div>	<div data-bbox="815 132 1386 172" data-label="Section-Header"> <h3>Slide 67 Statistical matching &amp; equating</h3> </div> <div data-bbox="815 256 940 291" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="282 655 750 695" data-label="Section-Header"> <h3>Structural modeling vs. ANCOVA</h3> </div> <div data-bbox="277 699 763 726" data-label="Text"> <p>Cook &amp; Campbell 1979. Primer on Regression artifacts</p> </div> <div data-bbox="230 726 764 1001" data-label="List-Group"> <ul style="list-style-type: none"> <li>• "The usefulness of analysis of covariance is closely coupled to the assumption that each covariate be measured without error" <ul style="list-style-type: none"> <li>▸ Other assumptions too</li> <li>▸ Violation of this assumption could be disastrous</li> </ul> </li> <li>• Using unreliable covariates can produce treatment effects that do not exist and can mask strong treatment effects. <ul style="list-style-type: none"> <li>▸ Gender discrimination</li> <li>▸ Racial differences on standardized tests</li> </ul> </li> <li>• Really unreliable covariates can change the sign of a treatment effect</li> </ul> </div> <div data-bbox="532 989 786 1029" data-label="Image"> </div>	<div data-bbox="815 621 1273 693" data-label="Section-Header"> <h3>Slide 68 Structural modeling vs. ANCOVA</h3> </div> <div data-bbox="815 779 940 814" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="282 1176 756 1243" data-label="Section-Header"> <h3>Solutions to Equating &amp; matching problems</h3> </div> <div data-bbox="230 1262 758 1533" data-label="List-Group"> <ul style="list-style-type: none"> <li>• Need a procedure that can adjust for the effect of the covariate, to correct for the 'bias' due to the regression to the mean phenomenon</li> <li>• Equating &amp; ANCOVA, may be ok when <ul style="list-style-type: none"> <li>▸ Randomized assignment of subjects to cases <ul style="list-style-type: none"> <li>▪ Equating not needed at all for reliability, but only for increasing 'power'</li> </ul> </li> <li>▸ <b>If there is little correlation between the treatment groups and the covariate.</b></li> </ul> </li> <li>• Alternatives to multiple regression: Structural equation modeling, change-score analysis (Campbell &amp; Kenny 1999), Hierarchical linear models, James-Stein (empirical Bayes) estimators</li> </ul> </div>	<div data-bbox="815 1146 1425 1222" data-label="Section-Header"> <h3>Slide 69 Solutions to Equating &amp; matching problems</h3> </div> <div data-bbox="815 1306 940 1341" data-label="Text"> <p>NOTES:</p> </div>

## Structural equation modeling

AMOS: Analysis of moment structures



Covered in EEOS612:  
No time in EEOS611

## Slide 70 Structural equation modeling

NOTES:

## Path analysis and regression

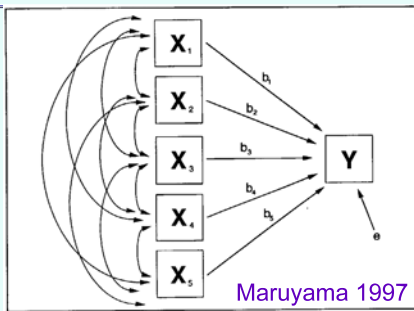


Figure 2.1. Regression Model With Five Predictor Variables

## Slide 71 Path analysis and regression

NOTES:

## Regression: a subset of structural equation modeling

The path diagram in Figure 1 shows a model for these data:

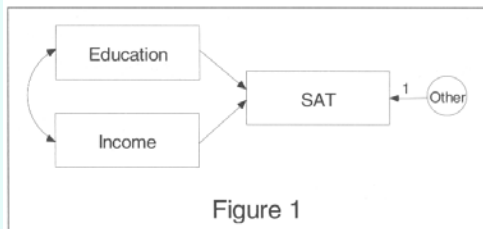


Figure 1

## Slide 72 Regression: a subset of structural equation modeling

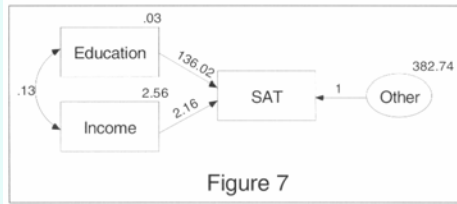
NOTES:



## AMOS graphical solutions

### Path coefficients (unstandardized or standardized)

Now to see the unstandardized estimates, simply click on **Unstandardized estimates** in the open dialog box next to your drawing area. Your path diagram should now look like **Figure 7**:

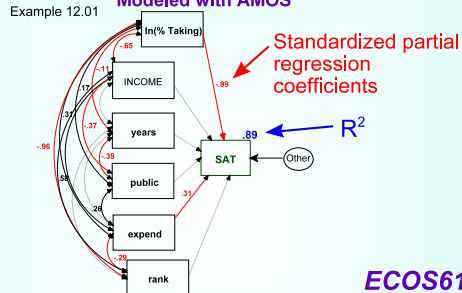


## Slide 73 AMOS graphical solutions

NOTES:

## Predicting SAT scores from states

Ramsey & Schafer (2001) "Statistical Sleuth" Ch. 12  
Modeled with AMOS



## Slide 74 Predicting SAT scores from states

NOTES:

## Results from a standard OLS regression

	Unstandardized Coefficients		Standardized Coefficients		t		Sig.		95% Confidence Interval for B	
	B	Std. Error	Beta		t				Lower Bound	Upper Bound
(Constant)	1008.646	167.708			6.020	.000			664.910	1352.382
INCOME	.46	.08	.37		6.000	.000			.31	.61
LOG-TEACHERS	-.052	.34	-.10		-.157	.874			-.739	.634

## Slide 75 Results from a standard OLS regression

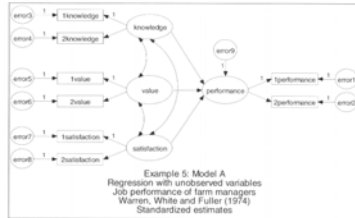
NOTES:

## From Path to Factor analysis

Latent variables (=unmeasured variables, Factors)

Model A

This path diagram presents a model for the eight subtests:



Four ellipses in the figure are labeled **knowledge**, **value**, **satisfaction** and **performance**. They represent the unobserved variables that are indirectly measured by the eight split-half tests.

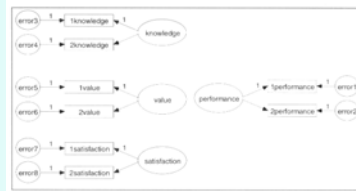
## Slide 76 From Path to Factor analysis

NOTES:

## Measurement & Structural submodels

Measurement model

The set of connections between the observed and unobserved variables is often called the measurement model. The current problem has four distinct measurement submodels:



Structural model

The model component connecting the latent variables is called the structural model:



## Slide 77 Measurement & Structural submodels

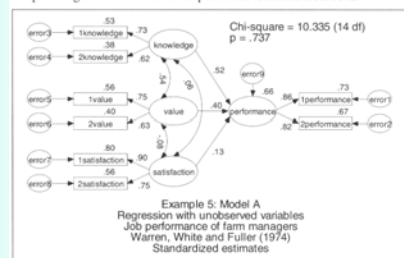
NOTES:

## AMOS Results

Chi-square, under  $H_0$ ,  $\approx$  d.f.

Amos Graphics output

The path diagram with standardized parameter estimates inserted is:



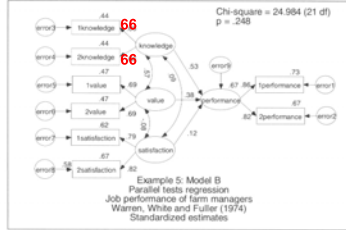
## Slide 78 AMOS Results

NOTES:

## Full vs Reduced models

### Test equal slopes model, fewer parameters

Now, here are the corresponding standardized estimates and squared multiple correlations in Amos Graphics output:



Testing Model B against Model A

## Slide 79 Full vs Reduced models

NOTES:

## How to handle covariates in RTM

### 213 11-year olds, pre- & post-test with training

#### The data

Olsson (1973) administered a battery of eight tests to 213 11-year-old students on two occasions. We will employ two of the eight tests, *Synonyms* and *Opposites*, in this example. Between the two administrations of the test battery, 108 of the students (the experimental group) received training that was intended to improve performance on the tests. The other 105 students (the control group) did not receive any special training. As a result of taking two tests on two occasions, each of the 213 students obtained four scores:

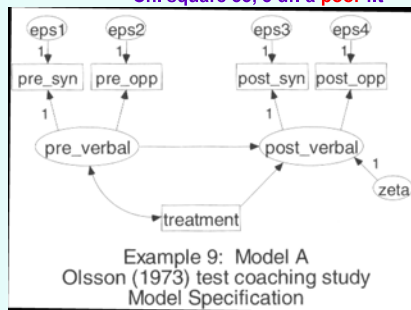
scores	explanation
pre_syn	Pretest scores on the Synonyms test
pre_opp	Pretest scores on the Opposites test
post_syn	Post-test scores on the Opposites test
post_opp	Post-test scores on the Synonyms test
treatment	A dichotomous variable taking on the value 1 for students who received the special training, and 0 for those who did not. This variable was created especially for the analyses in this example.

## Slide 80 How to handle covariates in RTM

NOTES:

## A reduced model

Chi square 33, 3 df: a poor fit

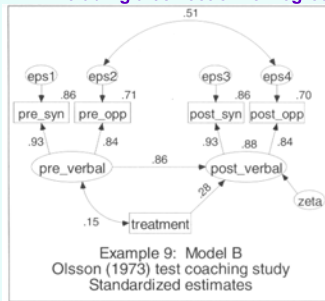


## Slide 81 A reduced model

NOTES:

## SEM, testing between groups

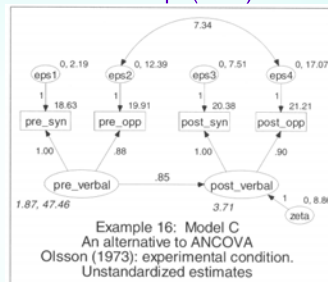
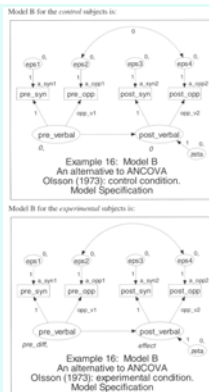
Including a correction for regression to the mean



## Slide 82 SEM, testing between groups

NOTES:

Testing treatment vs. Control with regression to mean; can estimate intercept (3.71)



## Slide 83

NOTES:

## MCAS Analyses and the thrip fallacy

## Slide 84 MCAS Analyses and the thrip fallacy

NOTES:

### Applications to SAT & MCAS

- SAT scores: can be analyzed using SEM
  - % Taking exams and expenditure per students are the most important variables
- How should socioeconomic factors be included in evaluating schools with MCAS
  - Strong collinearity among socio-economic variables
  - Gaudet & UMASS Donahue Institute
    - Socioeconomic variables are strongly correlated
    - Used principal component regression (didn't need to)
    - Could have used ridge regression
  - Tuerck, Beacon Hill Institute
    - Class size increases MCAS scores: probably an artifact, but need original data.
  - Chen & Ferguson (2002) simultaneous spatial autoregressive model (SAR)

### Slide 85 Applications to SAT & MCAS

NOTES:

### Gaudet's Ranking of MA Schools

1998 UMASS/Amherst Ph.D. and Donahue Institute Annual reports

- Gaudet's method for evaluating school quality
  - Socioeconomic variables from the 1990 census database, per student expenditure from MA DOE, MEAP results
  - 6 variables used in a "Major Axis" or principal components regression
    - average education level, average income, poverty rate, single-parent status, language spoken, and percentage of school-age population enrolled in private schools.
  - 86% of the variation in 1998 MCAS score is due to socioeconomic background of the students
  - Reduced to 85%, 83%, 81% and 81%MA
- Rerank 240 communities after controlling for 6 socioeconomic factors.

### Slide 86 Gaudet's Ranking of MA Schools

NOTES:

### The best 10th grade classes

Gaudet's ranking for President Bulger's office

District	ELA 10 Score	Overscore	District	Math 10 Score	Overscore
Berlin	255	10	Harvard	254	10
Boylston	251	8	Lenox	250	9
Lenox	251	8	Newburyport	251	8
Stoneham	250	8	Westborough	253	8
Northampton	248	8	Amesbury	246	8
Harvard	254	8	Northampton	245	7
Nauset	250	8	Gardner	240	7
Braintree	250	8	Nauset	247	7
Clinton	245	7	Shrewsbury	249	7
Wareham	244	7	Berlin	250	7
Shrewsbury	251	7	Boylston	250	7
Pentucket, Re	250	6	Braintree	247	6
Norwood	248	5	Nashoba	250	6
Westborough	251	5	Tyngsboro	245	6

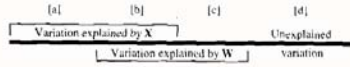
Similar to Case Study 12.1, the residual after fitting covariates (Socio-economic factors) is used to assess teaching Quality

### Slide 87 The best 10<sup>th</sup> grade classes

NOTES:

### The thrip/regression fallacy

Model in Response variable(s) Y can be partitioned



From Legendre & Legendre (1998)

Figure 10.10 Partition of the variation of a response variable  $y$  among two sets of explanatory variables  $X$  and  $W$ . The length of the horizontal line corresponds to 100% of the variation in  $y$ . Fraction (b) is the intersection of the linear effects of  $X$  and  $W$  on  $y$ . Adapted from Legendre (1993).

Andrewartha & Birch (1954) on 'weather' vs. Biological interactions controlling thrip abundance and Smith's critique

### Slide 88 The thrip/regression fallacy

NOTES:

### Chen & Ferguson (2002)

Evaluating school quality

$$Y_i = \beta_0 + \sum_{j=1}^4 \beta_j X_{ij} + \varepsilon_i \quad (A5.1)$$

where,  $Y_i, i = 1, 2, \dots, 226$  is the grand average of MCAS scores for years 1998, 1999, and 2000 for district  $i$ , and  $X_{ij}, j = 1, 2, 3, 4$  are the covariates of economic and demographic factors. They are AFRICAN-AMERICAN, PERCAP, TWOPHLD, and TAFDCPER. (LIM.ENG, which might quite reasonably be deemed a non-school related variable, is not used in this equation, since in combination with these variables alone it is not significant.) Once again, however, a Moran test indicates that the residuals of (A5.1) are spatially autocorrelated.



### Slide 89 Chen & Ferguson (2002)

NOTES:

Just as in the earlier equation we employ spatial models. Here the model is:

$$Y_i = \beta_0 + \sum_{j=1}^4 \beta_j X_{ij} + \delta_i + \varepsilon_i \quad (A5.2)$$

Again, as in Appendix 3, we estimate both a Conditional Spatial Autoregression (CAR) model using S-Plus and a Bayesian spatial approach estimated with WinBUGS. The estimated coefficients and p-values are listed in Table A5.3.

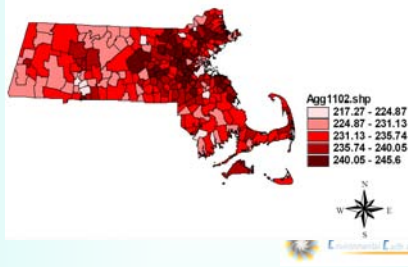
	S-PLUS	WinBUGS
INTERCEPT	221.54(.00)	224.20
AFRICAN	-0.160(.00)	-0.162
PERCAP	0.594(.00)	0.602
TWOPHLD	0.122(.00)	0.125
TAFDCPER	-2.124(.00)	-2.213

### Slide 90 Chen & Ferguson (2002)

NOTES:

### Spatially correlated residuals

MCAS Three Year Grand Average Scores 1998-2000



### Slide 91 Spatially correlated residuals

NOTES:

(See Text – Details of Economic/Demographic Equation Below)

RANK	SCHOOL	GRADV890	RES890	BAYRE890	3 YR TOT STUDS
1	AMHERST PELHAM	237.35	6.73	7.29	8709
2	LENOX	239.78	6.03	5.37	1701
3	HARVARD	245.60	5.57	6.10	2451
4	WESTBOROUGH	242.76	5.01	5.16	6251
5	BELMONT	243.95	4.61	4.02	7083
6	NAUSET	238.97	4.46	3.13	6837
7	NORTH READING	241.52	4.42	4.32	4468
8	NORTHAMPTON	234.23	4.20	4.21	6101
9	ACTON BOXBOROUGH	245.32	4.20	4.49	10307
10	HAMILTON WENHAM	241.37	3.89	3.62	5026
11	SANDWICH	240.02	3.79	3.04	7946
12	ARLINGTON	239.12	3.45	3.26	8153
13	NEWTON	243.60	3.34	2.59	22600
14	HADLEY	238.29	3.33	4.00	1274

### Slide 92 Chen & Ferguson (2002)

NOTES:

208	MARBLEHEAD	238.98	-2.94	-2.82	5612
209	BELLINGHAM	232.54	-3.00	-2.65	5359
210	SOUTH HADLEY	231.13	-3.01	-2.47	4797
211	SAUGUS	231.55	-3.04	-3.39	6638
212	WINCHENDON	227.46	-3.09	-3.43	3995
213	TAUNTON	226.04	-3.33	-3.20	15658
214	EASTHAMPTON	228.51	-3.49	-2.80	3731
215	MARLBOROUGH	231.44	-3.55	-3.57	8028
216	CAMBRIDGE	226.07	-3.67	-3.25	13788
217	LAWRENCE	217.50	-3.68	-2.76	21674
218	HAVERTHILL	226.92	-4.24	-3.86	16712
219	MAYNARD	231.77	-4.45	-4.00	2656
220	AVON	227.79	-4.45	-4.12	1660
221	LOWELL	222.05	-4.60	-4.56	29854
222	WESTPORT COMMUNITY	229.16	-5.02	-4.19	3871
223	NARRAGANSETT	227.87	-5.17	-5.27	2941
224	SOUTHERN BERKSHIRE	228.34	-6.08	-5.65	2157
225	DOVER SHERBORN	244.12	-6.11	-5.96	3760
226	WESTON	245.15	-6.40	-6.83	4087

### Slide 93 Chen & Ferguson (2002)

NOTES:

### What factors affect test scores?

Name of Variable	Definition	B-PLS	W-PLS
PERCAP	Per Capita Income	4707 (000)	0.4408
UNEMPLOYED	Unemployment Rate	1.048 (000)	0.1241
CRIME	Percentage of Pop. Under 18 in Crime	-1.388 (000)	-1.403
SALARY	Salary of Teachers	8489 (000)	0.04182
AFRICAN-AMERICAN	Percentage of African American	-0.2142 (000)	-0.2013
TEACHER	Teacher-Minimum Salary	0.1354 (000)	0.1297
INTERSTATE	Percentage of Interstate	0.0211 (000)	0.0184
LOTAL	Log of total population	3.3421 (001)	4.688
LDN ENG	Percentage of Urban	-0.1075 (014)	-0.1104
ATHLETIC	Percentage of Athletic	-0.0002 (015)	-0.007542
BOOKSQUEL	Percentage of Books	-0.0009 (015)	-0.00974
SENATE O	Senate Floor	0.0004 (016)	0.0762

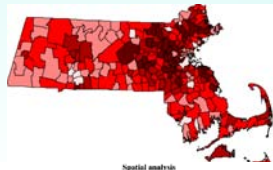


Figure 13.27 Partition of the variation of a response matrix  $Y$  between environmental covariates  $X_1$  and spatial covariates  $W$  (explanatory variables). The length of the horizontal line corresponds to 100% of the variation in  $Y$ . Compare to Fig. 13.10. (Adapted from Howard et al., 1995, and Eggen, 1995).

### Slide 94 What factors affect test scores?

NOTES:

### Beacon Hill Institute Study

Goal to rank schools & to evaluate educational policy

- Use 2000 MCAS scores as response variables
- Variables in Multiple regression:
  - Policy: % change in per pupil spending, percentage change in student-teacher ratios, number of students per computer
  - Socioeconomic: crime rates, % of workers that are professionals, % households headed by single females, Urban or non-urban
  - Choice variables: % students in charter schools, % students in METCO
  - Previous performance: 1994 MEAP scores

### Slide 95 Beacon Hill Institute Study

NOTES:

### Beacon Hill Results

Increase class sizes for "good schools"

- SES
  - School performance rises with % professionals or managers
  - School performance drops as the crime rate increases
  - School performance drops with higher % single parent households
  - Urbanized school districts have poorer performance
- Choice
  - Charter schools 'spur schools to do better'
  - METCO has no effect
  - % of students attending public schools positively associated with scores
- Policy implications
  - Spending doesn't improve performance
  - Increased class size for "good districts" improves performance
  - "Win-win situation" Increase class size in good districts by decreasing their funding and shift to poorer districts



### Slide 96 Beacon Hill Results

NOTES:



### The 15 best schools?

#### The 15 Best-Performing Massachusetts School Districts

DISTRICT (number of ratings for which district fell in the top 10)	Achieving Good Performance (G Rating)			Reducing Poor Performance (P Rating)		
	4 <sup>th</sup>	8 <sup>th</sup>	10 <sup>th</sup>	4 <sup>th</sup>	8 <sup>th</sup>	10 <sup>th</sup>
Hadley (5)	X	X	X		X	X
Clinton (3)	X	X		X		
Methuen (3)	X			X	X	
Stoneham (3)		X	X			X
Tyngsborough (3)	X		X			X
Nantucket (2)		X			X	
Chelsea (2)				X		X
Dighton-Rehoboth (2)		X			X	
Eastham (2)	X			X		
Everett (2)	X			X		
Hanover (2)		X			X	
Oxford (2)	X			X		
Provincetown (2)			X			X
Shrewsbury (2)			X			X
Sutton (2)	X			X		

### Slide 97 The 15 best schools?

NOTES:

### The 12 worst schools?

Beacon Hill Inst: Weighted average of 4th, 8th & 10th grades

#### The 12 Worst-Performing Massachusetts School Districts

DISTRICT (number of ratings for which district fell in the bottom 10)	Achieving Good Performance (G Rating)			Reducing Poor Performance (P Rating)		
	4 <sup>th</sup>	8 <sup>th</sup>	10 <sup>th</sup>	4 <sup>th</sup>	8 <sup>th</sup>	10 <sup>th</sup>
Narragansett (4)	X		X	X		X
Gateway (3)		X	X			X
Somerset (3)			X		X	X
Chesterfield-Goshen (2)	X			X		
Adams Cheshire (2)	X			X		
Hudson (2)		X			X	
Leicester (2)		X			X	
Millis (2)	X			X		X
Mount Greylock (2)		X			X	
Randolph (2)			X			X
Swampscott (2)			X			X
Watertown (2)		X	X			

### Slide 98 The 12 worst schools?

NOTES:

### The Worst 10th grade schools

#### Beacon Hill Institute

Foxborough	86	Taunton	210
Weston	22	Winchendon	192
Quabbin	128	Wareham	186
North Attleborough	171	Melrose	113
Berkshire Hills	133	Carver	187
Uxbridge	170	Leicester	142
Quabog Regional	168	Winthrop	188
Harvard	17	Westford	63
Peabody	193	Lunenburg	104
Longmeadow	46	Randolph	200
Southwick Tolland	199	Littleton	67
North Middlesex	88	Lincoln-Sudbury	36
Sutton	152	Watertown	132
Hopedale	135	Bellingham	174
Mount Greylock	60	Somerset	196
Douglas	172	Narragansett	191
Saugus	197	Swampscott	141
Taunton	210	Gateway	207

### Slide 99 The Worst 10<sup>th</sup> grade schools

NOTES:

<div data-bbox="292 170 742 205" data-label="Section-Header"> <h2>The Beacon Hill Institute Report</h2> </div> <div data-bbox="292 214 742 239" data-label="Section-Header"> <h3>Would increasing class size improve performance?</h3> </div> <div data-bbox="232 239 734 516" data-label="List-Group"> <ul style="list-style-type: none"> <li>• Beacon Hill study <ul style="list-style-type: none"> <li>▶ No attempt was made to assess colinearity among the many strongly correlated explanatory variables</li> <li>▶ Multicollinearity would invalidate many of their interpretations of betas, especially class size <ul style="list-style-type: none"> <li>▪ The authors should have calculated VIF's</li> <li>▪ Solutions <ul style="list-style-type: none"> <li>◦ Do ridge regression or principal components regression</li> <li>◦ Create a structural equation model for the hypotheses</li> </ul> </li> </ul> </li> <li>▶ A major conclusion from the study that increased class size improves MCAS performance runs counter to controlled experiments</li> </ul> </li> <li>• Experiments or quasi-experiments performed on class size indicate a negative correlation between class size and performance <ul style="list-style-type: none"> <li>▶ STAR</li> <li>▶ SAGE</li> </ul> </li> </ul> </div> <div data-bbox="532 499 786 541" data-label="Image"> </div>	<div data-bbox="815 132 1422 170" data-label="Section-Header"> <h2>Slide 100 The Beacon Hill Institute Report</h2> </div> <div data-bbox="815 256 938 291" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="326 657 696 690" data-label="Section-Header"> <h2>Class size and test scores</h2> </div> <div data-bbox="253 699 784 726" data-label="Section-Header"> <h3>Inference: reduced class size causes improved performance</h3> </div> <div data-bbox="232 726 771 966" data-label="List-Group"> <ul style="list-style-type: none"> <li>• The Tennessee Star Study <ul style="list-style-type: none"> <li>▶ A controlled experiment</li> <li>▶ Students randomly assigned to class sizes of 15 or 24</li> <li>▶ Long-lasting effects</li> </ul> </li> <li>• The Wisconsin SAGE study <ul style="list-style-type: none"> <li>▶ Students randomly assigned to small and large classes.</li> </ul> </li> <li>• Analysis of covariance (i.e., multiple regression) IS NOT a valid alternative to a randomized experiment</li> </ul> </div> <div data-bbox="532 987 786 1029" data-label="Image"> </div>	<div data-bbox="815 623 1304 657" data-label="Section-Header"> <h2>Slide 101 Class size and test scores</h2> </div> <div data-bbox="815 743 938 779" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="415 1146 599 1180" data-label="Section-Header"> <h2>Conclusions</h2> </div> <div data-bbox="232 1213 742 1474" data-label="List-Group"> <ul style="list-style-type: none"> <li>• Regression to the mean will be present whenever an explanatory variable (covariate) exhibits less than perfect correlation with the response variable. The higher the variability in the covariate, the more the regression to the mean effect</li> <li>• For pre-test vs. Post-test analyses, regressing with pretest score as an explanatory variable DOES NOT remove the effects of pre-test differences.</li> <li>• Better approaches: Repeated measures designs, hierarchical linear longitudinal models, or subtract pretest from posttest (called change score analysis)</li> </ul> </div> <div data-bbox="532 1474 786 1516" data-label="Image"> </div>	<div data-bbox="815 1110 1131 1146" data-label="Section-Header"> <h2>Slide 102 Conclusions</h2> </div> <div data-bbox="815 1232 938 1266" data-label="Text"> <p>NOTES:</p> </div>