Statistical Control and Mean Adjustment

1. Control

- Control is the process of eliminating threats to inference about which/how various independent variables (IV) contribute to variation on the dependent variable (DV). Control is about holding constant categories of variables.
- Confounding variables are often the threats that must be controlled. Confounders are variables that are related to both the model IVs and DVs but cause bias in regression and ANOVA if not included in the model or not controlled by design.
- Example: Research question – how does reading efficacy relate to reading achievement? Possible confounder: Sex of student. Control: Examine relation of reading efficacy and reading achievement separately by sex, or for only one sex.
- Control is built into experimental designs to provide internal validity so causal mechanisms leading to variation on DV can be associated with levels of IVs. This control by design helps to eliminate the effects of confounding variables.
- With experimental studies one attempts to obtain using various design methods such as
    - randomization of experimental units (e.g., people, plants, etc.),
    - including model covariates (e.g., variables that relate to or influence the DV),
    - including model factors or blocking variables (e.g. categorical variables that also influence the DV),
    - and holding constant levels of an IV to eliminate or manage the effects of that variable on the DV.

Examples of Control by Design

(a) Fertilizer for tomato yield

- IV: Four categories of fertilizer
    - Fish emulsion – nitrogen and other elements, applied at soil level periodically
    - Chicken manure - aged, added to soil during plantings, and again to soil periodically
    - Both
    - None
- DV: Yield, in pounds, of tomatoes
- Other experimentally designed controls of possible confounding variables:
    - 2 tomato plants per large pot,
    - same soil used,
    - same location within in a greenhouse to limit the effects of weather,
    - same amount of water applied at the same time daily,
    - and same amount of insecticide and fungicides applied to each pot.

(b) Classroom mathematics with background music

- IV: Three categories of music condition
    - Calming: Giazotto's Adagio: https://archive.org/details/albinoni-adagio-in-g-minor-12-sonatas-op.-6-c.rc.-claudio-scimone-i-solisti-veniti-piero-toso
    - Disruptive: John Contrane's The Father and the Son and the Holy Ghost: https://archive.org/details/cd_meditations_john-coltrane/90-0308/
    - None
- DV: Student performance on an algebra test
- Other experimentally designed controls of possible confounding variables:

- Six teachers involved, each teaches three sections of algebra, and each will teach one section with each background music playing (e.g., teacher A will each one section with calming, one with disruptive, and the third section with no music)
- Order of conditions will vary by teacher to control time of day effects (e.g. teacher A will teach Calming, Disruptive, and None; teacher B will teach Disruptive, None, and Calming; teacher C will teach None, Calming, then Disruptive; etc.)
- Teachers will use a common script each day when delivering the mathematics presentation
- Length of each lesson daily will be one hour
- Music will be played during the one-hour lesson and will be played on a common device with a common low volume (just above the threshold to hear in the background)

2. Statistical Control

- A means to partial, or hold constant, the effects of confounding variables (i.e., those that confuse interpretation of IV effects on the DV because the confounding variable correlates to both the IV and DV).
- The logic of statistical control is like control in experimental studies but done statistically. With correlational, or non-experimental data, statistical control does not offer causal interpretations like offered with data from experimental designs.
- Statistical control allows one to compare the IV effects upon the DV by holding constant the effects of the confounding variable. In essence, it is like examining the relation between the IV and DV for each level of the confounding variable.

Example of Statistical Control (Regression with two categorical variables)

Example data files so those interested can replicate the analyses:

SPSS [8g-Statistical-Control-and-Adjustment-Salary-Data.sav]
JASP [8g-Statistical-Control-and-Adjustment-Salary-Data.jasp]

Salary by Sex (controlling for Rank) – Fictional Data

- Is there a difference in faculty salary by sex?
- Salary reported in thousands of dollars
- Salary means
  - Females = $83.33
  - Males = $66.67

Descriptive Statistics

	Salary	
	Female	Male
Valid	30	30
Missing	0	0
Mean	83.3333	66.6667
Std. Deviation	19.0848	19.1155
Minimum	47.0000	45.0000
Maximum	105.0000	105.0000

- Regression results shows males earn $16.66 less than females (i.e., 66.67 – 83.33 = -16.66).

Coefficients

Model		Unstandardized	Standard Error	Standardized[a]	t	p
H_0	(Intercept)	75.0000	2.6747		28.0401	< .001
H_1	(Intercept)	83.3333	3.4872		23.8969	< .001
	Sex (Male)	−16.6667	4.9316		−3.3795	0.0013

[a] Standardized coefficients can only be computed for continuous predictors.

- Faculty rank (i.e., assistance, associate, professor) is a known factor in salary variation so it must be controlled to obtain a more realistic assessment of salary differences by sex.
- Below are salaries by rank and sex.

Descriptives

Descriptives - Salary

Rank	Sex	N	Mean	SD	SE	Coefficient of variation
Assistant	Female	5	50.0000	2.5495	1.1402	0.0510
	Male	15	50.0000	2.0702	0.5345	0.0414
Associate	Female	10	75.0000	2.4037	0.7601	0.0320
	Male	10	75.0000	2.4037	0.7601	0.0320
Professor	Female	15	100.0000	2.2039	0.5690	0.0220
	Male	5	100.0000	4.1231	1.8439	0.0412

- To statistically control for rank, add it to the regression of salary on sex. Results are provided below and now show that once rank is controlled, there is no salary difference between sexes as demonstrated in the descriptive table above. Thus, regression was able to offer control of rank, a potentially confounding variable, and clarify the salary by sex relation.

Coefficients

Model		Unstandardized	Standard Error	Standardized[a]	t	p
H_0	(Intercept)	75.000	2.675		28.040	< .001
H_1	(Intercept)	50.000	0.744		67.235	< .001
	Sex (Male)	2.149×10^{-14}	0.682		3.148×10^{-14}	1.000
	Rank (Associate)	25.000	0.782		31.976	< .001
	Rank (Professor)	50.000	0.836		59.823	< .001

[a] Standardized coefficients can only be computed for continuous predictors.

If salary is the same between the sexes for each rank, why do females have an overall higher mean salary when rank is ignored?

3. Statistical Adjustment

- When more than one IV is present in a regression model, the slope estimates are known as partial coefficients or partial slopes and represent the expected change in the DV for a one-unit change in the IV controlling for, or holding constant, the effects of other IVs in the regression equation.
- The **predicted means** from a regression equation take into account the partial effects of each IV, and these predicted means are known as **adjusted means** or **marginal means**.
- As previously noted, regression and ANOVA models are mathematically the same and both are part of the general linear model. Quantitative IVs in the ANOVA model are called covariates and the name for such models is ANCOVA (analysis of covariance). Like regression, group comparisons in ANCOVA are performed on marginal means (adjusted means) in which the DV means are adjusted for the contributions of each IV in the model.

Example of Statistical Adjustment (Regression with one categorical and one quantitative variable)

Example data files so those interested can replicate the analyses:

SPSS 8g-Statistical-Control-and-Adjustment-Pretest-Posttest-Data.sav
JASP 8g-Statistical-Control-and-Adjustment-Pretest-Posttest-Data.sav

The screenshot below shows fictional data used to illustrate statistical adjustment in regression and ANCOVA.

- Two group: Experimental and Control
- DV: Posttest scores
- Posttest means by group
  - Experimental m = 89.50
  - Control m = 83.50
  - Mean difference = 89.50 – 83.50 = 6.00
- A pretest is common for experimental studies because it provides a way to check for group equivalence, at least on the variable of interest, and provides a way to reduce error variance in the DV for regression and ANCOVA models which leads to additional precision for estimates and power for tests.
- To better illustrate adjustments, three pretest scores are provided in the data.
  - Pretest equal for both groups
    - Experimental m = 32.50
    - Control m = 32.50
  - Pretest control group lower: pretest is 5 points lower for the control group (5 points subtracted from initial pretest scores)
    - Experimental m = 32.50
    - Control m = 27.50
  - Pretest control group higher: pretest is 5 points higher for the control group (5 points added to initial pretest scores)
    - Experimental m = 32.50
    - Control m = 37.50
- Data for the example are shown below.

	Group	Posttest	Pretest-equal	Pretest-ControlLower	Pretest-ControlHigher
1	Control	73	30	25	35
2	Control	76	30	25	35
3	Control	79	25	20	30
4	Control	82	25	20 **-5**	30 **+5**
5	Control	85	40	35	45
6	Control	88	40	35	45
7	Control	91	35	30	40
8	Control	94	35	30	40
9	Experimental	79	30	30	30
10	Experimental	82	30	30	30
11	Experimental	85	25	25	25
12	Experimental	88	25	25	25
13	Experimental	91	40	40	40
14	Experimental	94	40	40	40
15	Experimental	97	35	35	35
16	Experimental	100	35	35	35

- In each scenario pretest scores correlate positively with posttest scores (the usual situation in experimental research)
- Descriptive statistics for each variable by group are shown below

Group Descriptives

	Group	N	Mean	SD	SE
Posttest	Control	8	83.500000000	7.348469228	2.598076211
	Experimental	8	89.500000000	7.348469228	2.598076211
Pretest-equal	Control	8	32.500000000	5.976143047	2.112885637
	Experimental	8	32.500000000	5.976143047	2.112885637
Pretest-ControlLower	Control	8	27.500000000	5.976143047	2.112885637
	Experimental	8	32.500000000	5.976143047	2.112885637
Pretest-ControlHigher	Control	8	37.500000000	5.976143047	2.112885637
	Experimental	8	32.500000000	5.976143047	2.112885637

Pretest Means Equal

If there are no differences in the pretest means for both groups, are any adjustments needed to account for pre-existing group differences? Use regression to calculate the adjusted mean for posttest scores.

Table A: No difference in pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	?
Control	32.50	83.50	?
Mean Difference =	0.00	6.00	?

Coefficients						
Model		Unstandardized	Standard Error	Standardized[a]	t	p
H_0	(Intercept)	86.500000000	1.936491673		44.668407926	< .001
H_1	(Intercept)	60.100000000	9.246995354		6.499408478	< .001
	Group (Experimental)	6.000000000	3.090929659		1.941163554	0.074233065
	Pretest-equal	0.720000000	0.276461153	0.536656315	2.604344197	0.021825324

[a] Standardized coefficients can only be computed for continuous predictors.

Prediction equation
Posttest' = 60.10 + 6.00 (Group) + 0.72 (pretest)

To obtain adjusted means for both groups, set the pretest score at the overall mean which is **32.50** and plug that value into the regression equation. Hence, the regression equation **holds constant** pretest scores at **32.50** and provides a prediction of the group means if both groups had the same mean pretest score.

Control group adjusted mean:
Posttest' = 60.10 + 6.00 (Group) + 0.72 (pretest)
Posttest' = 60.10 + 6.00 (0) + 0.72 (32.50)
Posttest' = 60.10 + 23.4
Posttest' = 83.50

Experimental group adjusted mean:
Posttest' = 60.10 + 6.00 (Group) + 0.72 (pretest)
Posttest' = 60.10 + 6.00 (1) + 0.72 (32.50)
Posttest' = 60.10 + 6.00 + 23.4
Posttest' = 89.50

Thus, the predicted means are the same as the observed means because there is no adjustment since both groups had the same mean on the covariate (pretest scores).

Table A: No difference in pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	89.50
Control	32.50	83.50	83.50
Mean Difference =	0.00	6.00	6.00

Pretest Mean for Control Group Lower

If the control group starts with a lower pretest mean score, are any adjustments needed to account for pre-existing group differences? Use regression to calculate the adjusted mean for posttest scores.

Table B: Control group starts study with lower pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	?
Control	27.50	83.50	?
Mean Difference =	0.00	6.00	?

Note that the control group starts the experiment with less knowledge and therefore a lower pretest score. This could partially explain why their posttest scores were lower. Therefore, pretest differences between the groups must be taken into account.

Coefficients

Model		Unstandardized	Standard Error	Standardized[a]	t	p
H_0	(Intercept)	86.500000000	1.936491673		44.668407926	< .001
H_1	(Intercept)	63.700000000	7.910606317		8.052480106	< .001
	Group (Experimental)	2.400000000	3.385943795		0.708812711	0.490957096
	Pretest-ControlLower	0.720000000	0.276461153	0.587877538	2.604344197	0.021825324

[a] Standardized coefficients can only be computed for continuous predictors.

Prediction equation
Posttest’ = 63.70 + 2.40 (Group) + 0.72 (pretest)

To obtain adjusted means for both groups, set the pretest score at the overall mean which is **30.00** and plug that value into the regression equation. Hence, the regression equation **holds constant** the pretest score at **30.00** and provides a prediction of the group means if both groups had the same mean pretest score.

Control group adjusted mean:
Posttest’ = 63.70 + 2.40 (Group) + 0.72 (pretest)
Posttest’ = 63.70 + 2.40 (0) + 0.72 (30.00)
Posttest’ = 63.70 + 21.6
Posttest’ = 85.30

Experimental group adjusted mean:
Posttest’ = 63.70 + 2.40 (Group) + 0.72 (pretest)
Posttest’ = 63.70 + 2.40 (1) + 0.72 (30.00)
Posttest’ = 63.70 + 2.40 + 21.6
Posttest’ = 87.70

The predicted means are now adjusted with the group starting with a lower pretest score adjusted upward and the group with the higher pretest score adjusted downward. Thus, regression has compensated the group means for their relative starting position on the covariate of pretest scores.

Table B: Control group starts study with lower pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	87.70
Control	27.50	83.50	85.30
Mean Difference =	0.00	6.00	2.40

Pretest Mean for Control Group Higher

If the control group starts with a higher pretest mean score, are any adjustments needed to account for pre-existing group differences? Use regression to calculate the adjusted mean for posttest scores.

Table C: Control group starts study with higher pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	?
Control	37.50	83.50	?
Mean Difference =	0.00	6.00	?

Coefficients

Model		Unstandardized	Standard Error	Standardized[a]	t	p
H_0	(Intercept)	86.500000000	1.936491673		44.668407926	< .001
H_1	(Intercept)	56.500000000	10.595173066		5.332616999	< .001
	Group (Experimental)	9.600000000	3.385943795		2.835250843	0.014048535
	Pretest-ControlHigher	0.720000000	0.276461153	0.587877538	2.604344197	0.021825324

[a] Standardized coefficients can only be computed for continuous predictors.

Prediction equation
Posttest' = 56.50 + 9.60 (Group) + 0.72 (pretest)

To obtain adjusted means for both groups, set the pretest score at the overall mean which is **35.00** and plug that value into the regression equation. Hence, the regression equation **holds constant** the pretest score at **35.00** and provides a prediction of the group means if both groups had the same mean pretest score.

Control group adjusted mean:
Posttest' = 56.50 + 9.60 (Group) + 0.72 (pretest)
Posttest' = 56.50 + 9.60 (0) + 0.72 (35.00)
Posttest' = 56.50 + 25.2
Posttest' = 81.7

Experimental group adjusted mean:
Posttest' = 56.50 + 9.60 (Group) + 0.72 (pretest)
Posttest' = 56.50 + 9.60 (1) + 0.72 (35.00)
Posttest' = 56.50 + 9.60 + 25.2
Posttest' = 91.30

The predicted means are now adjusted with the group starting with a lower pretest score adjusted upward and the group with the higher pretest score adjusted downward. Thus, regression has compensated the group means for their relative starting position on the covariate of pretest scores.

Table C: Control group starts study with higher pretest scores

Group	Pretest M	Observed Posttest M	Adjusted Posttest M
Experimental	32.50	89.50	91.30
Control	37.50	83.50	81.70
Mean Difference =	0.00	6.00	9.60

If these data were analyzed with the ANCOVA command in SPSS or JASP, the same results would be obtained. For example, below is JASP output from ANCOVA showing marginal means (adjusted means) using the pretest scores +5 for the control group (i.e., the results shown in Table C above). Note the calculated adjusted means above match the marginal means reported by JASP.

ANCOVA - Posttest ▼

Cases	Sum of Squares	df	Mean Square	F	p
Group	307.200000000	1	307.200000000	8.038647343	0.014048535
Pretest-ControlHigher	259.200000000	1	259.200000000	6.782608696	0.021825324
Residuals	496.800000000	13	38.215384615		

Note. Type III Sum of Squares

Marginal Means

Marginal Means - Group

Group	Marginal Mean	95% CI for Mean Difference		SE
		Lower	Upper	
Control	81.700000000	76.747798194	86.652201806	2.292294786
Experimental	91.300000000	86.347798194	96.252201806	2.292294786

4. Validity of Regression and Model Specification

- Gelman and Hill (2006, p 45) explain that validity of regression means that the
  - data analyzed must link to the research questions driving the study;
  - IV and DV measures should demonstrate score validity (i.e., scores represent the construct);
  - model should include related IVs and exclude unrelated IVs; and
  - the regression model should use data from a representative sample of the target population.
- Pedhazur (1997) notes that model specification refers to proper inclusion of relevant predictors, and ensuring the model is additive (e.g., interaction terms are addressed) and linear (e.g., non-linear components are addressed).
- The focus here is on inclusion of relevant predictors. Pedhazur explains that failure to include relevant predictors can lead to bias parameter estimates, i.e., regression coefficients can be wrong and misleading.

Example of Model Specification: Excluding a Relevant Predictor (Regression with multiple quantitative variables)

Example data files so those interested can replicate the analyses:

SPSS 8g-Statistical-Control-and-Adjustment-SAT-Data.sav
JASP 8g-Statistical-Control-and-Adjustment-SAT-Data.jasp

- Data: State level variables (n = 50) collected in 1998
  - DV: Mean Scholastic Aptitude Test (SAT) combined score for mathematics and verbal per state
  - IV: Average (mean) ratio of students to teachers (i.e., a proxy for class size) per state
  - IV: Mean teacher salary in thousands of dollars per state
- Predictions (hypotheses)
  - Positive relation between SAT and teacher salary – those states with higher salaries should see higher student performance on the SAT
  - Negative relation between SAT and class size – those states with a higher student to teacher ratio should see lower SAT scores
- Below is a screenshot of the data:

- sat_total = SAT combined score
- salary = teacher mean salary (1996?)
- ratio = ratio of students to teachers (1996?)

	state	ratio	salary	sat_percent	sat_total	region
1	Mississippi	17.5	26.818	4	1036	South
2	Utah	24.3	29.082	4	1076	West
3	South Dakota	14.4	25.994	5	1068	Midwest
4	Iowa	15.8	31.511	5	1099	Midwest
5	North Dakota	15.3	26.327	5	1107	Midwest
6	Arkansas	17.1	28.934	6	1005	South
7	Alabama	17.2	31.144	8	1029	South
8	Louisiana	16.8	26.461	9	1021	South
9	Oklahoma	15.5	28.172	9	1027	South
10	Missouri	15.5	31.189	9	1045	Midwest
11	Nebraska	14.5	30.922	9	1050	Midwest
12	Kansas	15.1	34.652	9	1060	Midwest
13	Wisconsin	15.9	37.746	9	1073	Midwest
14	Minnesota	17.5	35.948	9	1085	Midwest
15	Wyoming	14.9	31.285	10	1001	West
16	Kentucky	17	32.257	11	999	South
17	New Mexico	17.2	28.493	11	1015	West
18	Michigan	20.1	41.895	11	1033	Midwest

Regression Results

JASP regression results are shown below.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	965.920000000	10.581224805		91.286218546	< .001
H_1	(Intercept)	1113.877434677	93.002628595		11.976838198	< .001
	ratio	2.665842353	4.307096146	0.080749787	0.618941919	0.538943193
	salary	−5.538449227	1.642984598	−0.439790836	−3.370968441	0.001505615

SAT' = b0 + b1 ratio + b2 salary

SAT' = 1113.87 + 2.66 ratio + -5.53 salary

The hypotheses were not supported. The regression results show, instead, that:

- Class size (student to teacher ratio) is **positively** related to SAT scores – the larger the class size, the higher SAT scores, but the coefficient was not significant.
- Teacher salary was **negatively** related to SAT scores – the higher teacher salary, the lower SAT scores. The coefficient is statistically significant.

Data Oddness

Estimated regression coefficients seem odd and certainly contradict prior expectations. What else seems odd about the data?

Highest Scoring SAT	SAT Mean	Lowest Scoring SAT	SAT Mean
North Dakota	1107	South Carolina	844
Iowa	1099	Georgia	854
Minnesota	1085	North Carolina	865
Utah	1076	Pennsylvania	880
Wisconsin	1073	Indiana	882
South Dakota	1068	Rhode Island	888
Kansas	1060	Florida	889
Nebraska	1050	Hawaii	899
Illinois	1048	New York	892

Surprisingly there is a very large discrepancy in state mean scores as shown above. One should expect that mean scores would demonstrate less variability. The difference between the top and bottom scoring states is 1107-844 = 263 which is more than the combined test SD of 200.

What is the nature of students who take the SAT across states?

That question is not easily answered with the available data, but the College Board, writers of the SAT, do provide the percentage of high school graduates who take the SAT in each state, which is linked below and reported on many sites.

https://reports.collegeboard.org/sat-suite-program-results

The data table now includes the percentage of high school graduates, in the class of 1998, who took the SAT in each state.

Highest Scoring SAT	SAT Mean	SAT Participation Rate	Lowest Scoring SAT	SAT Mean	SAT Participation Rate
North Dakota	1107	5	South Carolina	844	58
Iowa	1099	5	Georgia	854	65
Minnesota	1085	9	North Carolina	865	60
Utah	1076	4	Pennsylvania	880	70
Wisconsin	1073	9	Indiana	882	58
South Dakota	1068	5	Rhode Island	888	70
Kansas	1060	9	Florida	889	57
Nebraska	1050	9	Hawaii	899	48
Illinois	1048	13	New York	892	74

Note the large discrepancy of graduates who completed the SAT between high scoring and low scoring SAT states. Clearly this is an important difference that must be considering in the analysis of SAT scores.

Why do some states have participation rates over 50% while others have participation rates under 10%?

The answer is competition, culture, and graduation requirement.

- The rival to the SAT is the ACT (developed by American College Testing). While most colleges and universities that require a standardized test for admission will accept either the SAT or ACT, the culture within a state seems to drive preferences.
- I remember, in the early 1980s, that everyone in my high school planning to attend college took the SAT; the ACT was not considered an option. Likely similar cultures of preference exist in various states for the ACT or the SAT, and this partially explains the variation in SAT participation rates across states. Local availability of proctoring

sites may be another reason. As a high school student, I recall the SAT being offered at a nearby high school, but don't recall the ACT being offered nearby.

- Some states require one or the other test as a high school graduation requirement. It serves as benchmark for assessment purposes.

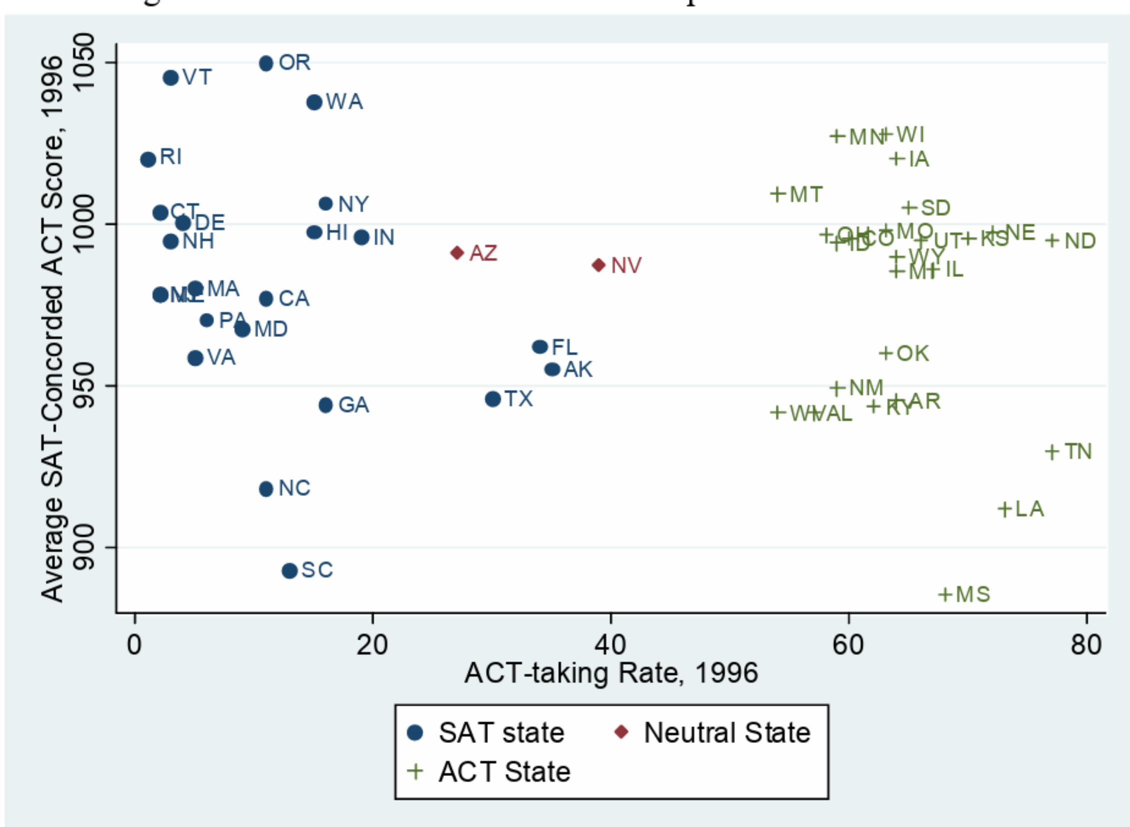
Why do those states with the lowest SAT participation rates have the highest means?

Likely the ACT is the most popular test in those states with lowest SAT participation rates (e.g., North Dakota, Iowa, Minnesota, etc.), so the only students taking the SAT are those who plan to attend a college in another state or country that prefers the SAT. Additionally, those students are also likely to be among the brightest and know that a high SAT score will be needed to gain admission to the out-of-state college or university. In short, many of the strongest students take the SAT in those lowest participation rate states while students with weaker performance tend to take the ACT.

Although I did not collect state participation rates of the ACT in 1998, I was able, in 2024, to find the document linked below which provides a graph showing SAT adjusted scores and ACT participation rates based upon 1996 results. The graph is shown below.

https://www.nber.org/system/files/working_papers/w14265/w14265.pdf

Figure 1B. Average 1996 ACT Score versus ACT Participation Rate



Using the above graph, I have provided approximate ACT participation rates for those states with the lowest SAT participation rate.

Highest Scoring SAT	ACT Participation Rate Approximate	SAT Participation Rate
North Dakota	78	5
Iowa	64	5
Minnesota	58	9
Utah	66	4
Wisconsin	63	9
South Dakota	65	5
Kansas	70	9
Nebraska	72	9
Illinois	68	13

As these results show, and as the document cited above demonstrates, there is selection bias in participation in the SAT across states. Given this, it is necessary to control for participation rate when modeling SAT scores since there is selection bias in who completes the SAT by state.

As this example demonstrates, it is important to consider model specification – predictor inclusion – to obtain proper coefficient estimates. Failure to include relevant, important predictors can provide misleading results as was shown by the first regression equation. This is an example of Simpson's paradox (effects change directions once a variable is controlled) and of variable suppression (effects become larger, weaker, or change direction after controlling for a variable).

Revised Regression Results

JASP regression results are shown below.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	965.920000000	10.581224805		91.286218546	< .001
H_1	(Intercept)	1057.898162003	44.328668530		23.864875646	< .001
	ratio	−4.639428046	2.121514211	−0.140530751	−2.186847499	0.033876517
	salary	2.552470111	1.004518263	0.202683624	2.540989253	0.014491262
	sat_percent	−2.913350207	0.228243593	−1.042070438	−12.764214640	< .001

SAT' = b0 + b1 ratio + b2 salary + b3 sat percent
SAT' = 1057.89 + -4.63 ratio + 2.55 salary + -2.91 sat percent

The hypotheses are now supported. The regression results show that:

- Class size (student to teacher ratio) is **negatively** related to SAT scores – the larger the class size, the lower SAT scores. The coefficient is statistically significant.
- Teacher salary was **positively** related to SAT scores – the higher teacher salary, the higher SAT scores. The coefficient is statistically significant.
- Additionally, the greater the percentage of students taking the SAT, the lower SAT mean scores. This coefficient is also significant.

Caution with Adjusted Means

Adjusted means use the regression line of the covariate for prediction, and it is possible that line is not consistent for subsets the population. This means it is possible an interaction exists between the covariate and groups compared. See Agresti text, section 13.5 for more information. If an interaction exists, then distinct regression slopes exist for each group, so adjust means must be considered across ranges of the covariate to provide a more appropriate understanding and interpretation of group differences. Read also Agresti textbook chapter 10 for more on control and adjustments.

Adjusted means are covered in more detail in the ANCOVA notes and presentations.

References

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Pedhazur, E. J., (1997). *Multiple Regression in Behavioral Research* (3rd ed.). Orlando, FL: Harcourt Brace.