More Regression

What else?

General Outline

- Sample size determination
- Dealing with categorical data
  - Dummy coding
  - Effects coding
  - Contrast coding
- Different approaches to MR
  - Standard
  - Sequential
  - Stepwise
- Model comparison
- Moderators and Mediators

Sample Size

- Programs such as Gpower or the MBESS package in R may be used to calculate the sample size needed for the desired power, alpha level, number of predictors and chosen effect size, assuming…
- You have perfectly reliable measures
- You meet all assumptions
- You have no outliers
- The effect size you enter is actually obtained

- In short, if a primary concern is rejecting a null hypothesis you better get big samples or hope for big effects

Categorical variables in regression

- As we've noted before, regression can actually handle different types of predictors, and in the social sciences we are often interested in differences between k groups

- Dummy coding
- 0s and 1s
- k-1 predictors will go into the regression equation leaving out one reference category (e.g. control)
  - Note that with only 2 categories the common 0-1 classification is sufficient to use directly in MR (2-1 = 1 dummy coded variable)
- Coefficients will be interpreted as change with respect to the reference variable (the one with all zeros)

Dummy coding

- In general, the coefficients are the *distances from the dummy values to the reference value*, controlling for other variables in the equation
- Perhaps the conceptually easier way to interpret it is not in terms of actual change in Y as we do in typical regression (though that's what we're doing), but how we do in Anova with regard to mean differences
  - The b coefficient is how far that group (which is coded with a 1) mean is from the reference group mean, whose mean can be seen in the output as the constant
  - The difference in coefficients is the difference between mean of the 1 group and the reference (left out) group
- If only two categories, the b represents the difference in their means

For 3 groups, 2 dummy coded variables
The 'all zero' group is the reference category

P1	P2	DV
1	0	3
1	0	5
1	0	7
1	0	2
0	1	3
0	1	6
0	1	7
0	1	7
0	0	8
0	0	8
0	0	9
0	0	9

Effects coding

- Makes comparisons in relation to the grand mean of all subgroups
- Reference category is coded as -1 for all predictor dummy variables, and your constant is now the grand mean
  - Still k-1 predictors for k groups
- Given this type of effect coding, a b of -1.5 for the dummy "Experimental Group 2" means that the expected value on the DV for the Experimental (i.e. its mean) is 1.5 less than the mean for all subgroups
- Dummy coding interprets b for the dummy category relative to the reference group (the left-out category), effects coding interprets it relative to the entire set of groups
- Perhaps more interpretable when no specific control group for reference

Example data and output

- A sales manager wishes to determine the best training program for new employees. He has performance scores for three groups: employees in training program 1, 2 or 3

ANOVA

Score on training exam

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2525.691	2	1262.846	12.048	.000
Within Groups	5974.724	57	104.820		
Total	8500.415	59			

Original grouping variable

Dummy Effects

12	1	44.92	1.00	.00	1.00	.00
13	1	67.04	1.00	.00	1.00	.00
14	1	62.99	1.00	.00	1.00	.00
15	1	66.63	1.00	.00	1.00	.00
16	1	65.53	1.00	.00	1.00	.00
17	1	59.58	1.00	.00	1.00	.00
18	1	85.65	1.00	.00	1.00	.00
19	1	64.55	1.00	.00	1.00	.00
20	1	83.74	1.00	.00	1.00	.00
21	2	72.85	.00	1.00	.00	1.00
22	2	88.17	.00	1.00	.00	1.00
23	2	80.82	.00	1.00	.00	1.00
24	2	71.27	.00	1.00	.00	1.00
25	2	81.50	.00	1.00	.00	1.00
26	2	47.56	.00	1.00	.00	1.00
27	2	81.04	.00	1.00	.00	1.00
28	2	81.38	.00	1.00	.00	1.00
29	2	82.96	.00	1.00	.00	1.00
30	2	75.98	.00	1.00	.00	1.00
31	2	77.35	.00	1.00	.00	1.00
32	2	69.31	.00	1.00	.00	1.00
33	2	61.69	.00	1.00	.00	1.00
34	2	64.87	.00	1.00	.00	1.00
35	2	75.43	.00	1.00	.00	1.00
36	2	59.83	.00	1.00	.00	1.00
37	2	89.65	.00	1.00	.00	1.00
38	2	59.10	.00	1.00	.00	1.00
39	2	76.14	.00	1.00	.00	1.00
40	2	74.46	.00	1.00	.00	1.00
41	3	82.33	.00	.00	-1.00	-1.00
42	3	89.69	.00	.00	-1.00	-1.00
43	3	81.01	.00	.00	-1.00	-1.00
44	3	85.09	.00	.00	-1.00	-1.00
45	3	74.14	.00	.00	-1.00	-1.00
46	3	75.93	.00	.00	-1.00	-1.00

Reference category

ANOVA

Score on training exam

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2525.691	2	1262.846	12.048	.000
Within Groups	5974.724	57	104.820		
Total	8500.415	59			

Categorical variables in regression: Equivalence of Regression and ANOVA

- Regular Anova

ANOVA

Score on training exam

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2525.691	2	1262.846	12.048	.000
Within Groups	5974.724	57	104.820		
Total	8500.415	59			

- Dummy

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2525.691	2	1262.846	12.048	.000[a]
	Residual	5974.724	57	104.820		
	Total	8500.415	59			

a. Predictors: (Constant), group2, group1

b. Dependent Variable: Score on training exam

- Effects

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2525.691	2	1262.846	12.048	.000[a]
	Residual	5974.724	57	104.820		
	Total	8500.415	59			

a. Predictors: (Constant), group2b, group1b

b. Dependent Variable: Score on training exam

Categorical variables in regression

- Relative to reference group
  - Both are doing less than method 3

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	79.279	2.289		34.630	.000
	group1	-15.699	3.238	-.622	-4.849	.000
	group2	-5.712	3.238	-.226	-1.764	.083

a. Dependent Variable: Score on training exam

- Relative to grand mean
  - Method 1 is noticeably less than grand mean, 2 training method is near it
  - Suggests 3 better than both (again)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	72.142	1.322		54.581	.000
	group1b	-8.562	1.869	-.587	-4.581	.000
	group2b	1.425	1.869	.098	.763	.449

a. Dependent Variable: Score on training exam

Contrast coding

- Test for specific group differences and trends
- Coding scheme the same as done in ANOVA
  - http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter5/sasreg5.htm
- We will talk more about contrast analyses with advanced ANOVA topics later, but it is enough to let you know you can still do it within regression

Methods of classical MR

- Confirmatory: Sequential/Hierarchical
- Exploratory: Stepwise

Sequential[1] (hierarchical) regression

- Researcher specifies some order of entry based on theoretical considerations
- They might start with the most important predictor(s) to see what they do on their own before those of lesser theoretical interest are added
- Otherwise, might start with those of lesser interest, and see what the important ones add to the equation (in terms of R^2)
- There isn't any real issue here, except that you can get the same sort of info by examining the squared semi-partial correlations with the 'all in' approach

- Comparison of Standard vs. Sequential
  - Standard at top, 3 different orderings follow
  - e b l
  - b l e
  - l b e
- Note that no matter how you start you end up the same once all the variables are in.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-1.467	.084		-17.390	.000			
	e	.937	.144	.254	6.502	.000	-.250	.364	.159
	b	-4.405	.179	-.647	-24.656	.000	-.799	-.829	-.605
	l	-.577	.037	-.631	-15.602	.000	-.603	-.685	-.383

a. Dependent Variable: duobs

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-3.728	.117		-31.861	.000			
	e	-.923	.214	-.250	-4.312	.000	-.250	-.250	-.250
2	(Constant)	-2.262	.092		-24.597	.000			
	e	-.814	.124	-.221	-6.575	.000	-.250	-.367	-.221
	b	-5.388	.229	-.791	-23.535	.000	-.799	-.816	-.790
3	(Constant)	-1.467	.084		-17.390	.000			
	e	.937	.144	.254	6.502	.000	-.250	.364	.159
	b	-4.405	.179	-.647	-24.656	.000	-.799	-.829	-.605
	l	-.577	.037	-.631	-15.602	.000	-.603	-.685	-.383

a. Dependent Variable: duobs

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-2.542	.088		-29.037	.000			
	b	-5.443	.246	-.799	-22.171	.000	-.799	-.799	-.799
2	(Constant)	-1.598	.088		-18.212	.000			
	b	-4.698	.185	-.690	-25.359	.000	-.799	-.836	-.667
	l	-.390	.025	-.426	-15.673	.000	-.603	-.686	-.412
3	(Constant)	-1.467	.084		-17.390	.000			
	b	-4.405	.179	-.647	-24.656	.000	-.799	-.829	-.605
	l	-.577	.037	-.631	-15.602	.000	-.603	-.685	-.383
	e	.937	.144	.254	6.502	.000	-.250	.364	.159

a. Dependent Variable: duobs

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-2.432	.148		-16.433	.000			
	l	-.552	.044	-.603	-12.615	.000	-.603	-.603	-.603
2	(Constant)	-1.598	.088		-18.212	.000			
	l	-.390	.025	-.426	-15.673	.000	-.603	-.686	-.412
	b	-4.698	.185	-.690	-25.359	.000	-.799	-.836	-.667
3	(Constant)	-1.467	.084		-17.390	.000			
	l	-.577	.037	-.631	-15.602	.000	-.603	-.685	-.383
	b	-4.405	.179	-.647	-24.656	.000	-.799	-.829	-.605
	e	.937	.144	.254	6.502	.000	-.250	.364	.159

a. Dependent Variable: duobs

- Take the squared semi-partials from the previous slide (appropriate model) and add them in to get your new R^2
- Example: from the first ordering, to get from E to E+B, take the 'part' correlation for B when it is second in the ordering (-.790), square it (which gives you ~.625), and add that to .063 (the R^2 with E only) to get my new R^2 of .688 with both of them in.
- Adding the squared semi-partial of L would provide the final model R^2

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)	-1.467	.084		-17.390	.000			
	e	.937	.144	.254	6.502	.000	-.250	.364	.159
	b	-4.405	.179	-.647	-24.656	.000	-.799	-.829	-.605
	l	-.577	.037	-.631	-15.602	.000	-.603	-.685	-.383

a. Dependent Variable: duobs

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.250[a]	.063	.059	1.46624	.063	18.592	1	278	.000
2	.829[b]	.688	.685	.84811	.625	553.907	1	277	.000
3	.913[c]	.834	.832	.61933	.146	243.437	1	276	.000

a. Predictors: (Constant), e

b. Predictors: (Constant), e, b

c. Predictors: (Constant), e, b, l

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.799[a]	.639	.637	.91025	.639	491.561	1	278	.000
2	.899[b]	.809	.807	.66387	.170	245.643	1	277	.000
3	.913[c]	.834	.832	.61933	.025	42.271	1	276	.000

a. Predictors: (Constant), b

b. Predictors: (Constant), b, l

c. Predictors: (Constant), b, l, e

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.603[a]	.364	.362	1.20774	.364	159.140	1	278	.000
2	.899[b]	.809	.807	.66387	.444	643.082	1	277	.000
3	.913[c]	.834	.832	.61933	.025	42.271	1	276	.000

a. Predictors: (Constant), l

b. Predictors: (Constant), l, b

c. Predictors: (Constant), l, b, e

Sequential regression

- The difference between standard and sequential regression is primarily one of description of results, not method and by default, the program you use is already doing some form of sequential regression
  - Type I Sums of Squares: straightforward sequential
    - The effect you see for a predictor depends on order of selection
    - Reduction in the error SS as each effect is added to the model
  - Type II SS: true hierarchical; SAS (proc reg), R, S-Plus, others default
    - The effect you see partials out other effects that do not include it
    - Reduction in error SS due to adding the term to the model after all other terms except those that contain it (i.e. interaction terms)
  - Type III: SPSS
    - The effect you see is as if it is entered last in the model[1]
    - Reduction in error SS due to adding the term after all other terms have been added to the model

- Sequential may provide a better description of your theory, and allow you to do sets at a time rather easily, but the info is technically there to do so in a standard regression

Stepwise (statistical) methods

- Statistical methods allow one to come up with a 'best' set of predictors from a given starting set
- These are used in exploratory endeavors as part of a theory generating process, and model search in general is a rapidly developing area of statistics
- Techniques include:
  - Backward
  - Forward
  - "Stepwise"
  - All possible subsets and Model Comparison/ Averaging
- In the following examples: consider predictors A, B, C and dependent variable Y

Stepwise

- Backward
  - All in
  - Remove least 'significant'[1] contributor
  - Rerun, do the same until all left are significantly contributing
- Forward
  - Start with predictor with largest correlation (validity) with DV
  - Add in the next predictor that results in e.g. greatest increase in R^2 (i.e. which has the largest semi-partial squared when it is second in the ordering)
    - The third will be added with the first two staying in
  - Continue until no significant change in model results from the addition of another variable

- Backward
  - 1st Model: ABC as predictors
  - 2nd Model: AB- both sig?
    - Yes, stop
    - No, continue
  - 3rd model: A (or B, whichever was sig)
- Forward
  - 1st Model: A
  - 2nd Model: AB
    - R^2 change significant with addition of B?
      - Yes, continue
        - Model 3: ABC
      - No, stop
        - Final Model: A
  - 3rd Model ABC?
    - R^2 increase significant with addition of C?
      - Yes, stop
        - Final model: ABC
      - No, stop
        - Final Model: AB

Stepwise

- "Stepwise"
  - Forward and backward are both stepwise procedures
  - "Stepwise" regression however specifically refers to the forward selection process, but if upon retest of a new model a variable previously significant is no longer so, it is removed
- All subsets
  - Starting with the full or unrestricted model of all predictors included, find the best group of variables based on some criterion (e.g. best 3 that produce the largest Adjusted R^2), or just choose from the results which one you like
  - More computer intensive but easily accomplished by desktops for typical academic research questions

- Stepwise
  - 1st Model: A
  - 2nd Model: AB
    - Is A still sig?
      - Yes, continue
        - Model 3: ABC
        - Reassess A and B
      - No, drop it
        - Model 3: BC
        - Reassess B
- All subsets
  - Full Model: ABC
  - All subset models:
    - A
    - B
    - C
    - AB
    - BC
    - AB

Problems with stepwise methods

- Capitalization on chance
  - Results are less generalizable and may only hold for your particular data
- The statistical program is dictating your ideas
  - "*Never* let a computer select predictors mechanically. The computer does not know your research questions nor the literature upon which they rest. It cannot distinguish predictors of direct substantive interest from those whose effects you want to control." (Singer & Willett)

Problems with stepwise methods

- Frank Harrell (noted stats guy)
  - It yields R-squared values that are badly biased high.
  - The F and chi-squared tests quoted next to each variable on the printout do not have the claimed distribution.
  - The method yields confidence intervals for effects and predicted values that are falsely narrow (See Altman and Anderson, Statistics in Medicine).
  - It yields P-values that do not have the proper meaning and the proper correction for them is a very difficult problem.
  - It gives biased regression coefficients that need shrinkage (the coefficients for remaining variables are too large; see Tibshirani, 1996).
  - It has severe problems in the presence of collinearity.
  - It is based on methods (e.g., F tests for nested models) that were intended to be used to test prespecified hypotheses.
  - Increasing the sample size doesn't help very much (see Derksen and Keselman).
  - *It allows us to not think about the problem.*

More

- Stepwise methods will not necessarily produce the best model if there are redundant predictors (common problem).
- Models identified by stepwise methods have an inflated risk of capitalizing on chance features of the data. They often fail when applied to new datasets. They are rarely tested in this way.
- Since the interpretation of coefficients in a model depends on the other terms included, "it seems unwise to let an automatic algorithm determine the questions we do and do not ask about our data".
- "It is our experience and strong belief that better models and a better understanding of one's data result from focused data analysis, guided by substantive theory."

- Judd and McClelland, *Data Analysis: A Model Comparison Approach*

The exploratory nature of science

- Stepwise regression is an exploratory technique, this is not debatable
  - Presenting stepwise results as standard, theory-generated results is to commit research fraud
- However, exploratory endeavors are an essential part of science and can be useful
- Exploratory regression has its place in that regard, but is almost completely useless as an inferential technique without validation either with a new data set or at the very least cross- or bootstrap-validation with the same data.
  - Without validation you might as well save yourself the trouble and simply provide the correlation matrix

Model comparison

- Science progresses not with a probabilistic refutation of a hypothesis no sees as viable to begin with, but rather pitting ideas against each other with real world evidence bearing out the details
- Is one set of predictors better at predicting a DV than another set?
- Multiple regression and other techniques allows us to pit one hypothesis/ theory vs. another, but how would we compare them?
- For outcome Y, any two models of predictors of Y can be seen as subsets as the full model with all predictors
  - Example
    - Theory 1: Y ~ A + B + C
    - Theory 2: Y ~ D + E + F
    - Unrestricted model Y ~ A + B + C + D + E + F
  - Alternative example
    - Theory 1: Y ~ A + B + C
    - Theory 2: Y ~ B + C + D
    - Unrestricted model Y ~ A + B + C + D
  - Alternative example:
    - Theory 1: Y ~ A + B
    - Theory 2 (unrestricted model): Y ~ A + B + AB (interaction)

Model comparison

- Compare R^2?
  - R^2 cannot be compared for models of different sizes
  - Also R^2s that aren't that far off could change to nonsig difference or even flip flop model orderings upon a new sample
    - Always remember that effect sizes, like p-values and other statistics are variable from sample to sample
- Other statistics do allow for such comparisons though would still require validation
  - Adjusted R^2 can be used as it takes into account the number of predictors
  - Fit statistics such as the Aikaike Information Criterion or Bayesian Information Criterion that you typically see in path analysis or SEM might be a better way to compare models[1]
    - When comparing the lower value is the better model
      - If both negative, the more negative value

Model comparison

- The problem with traditional stepwise approaches is that though a final model is selected, you don't get a good sense of where it fits in the grand scheme of other possibilities, and as parsimony is a goal of science, the final model selected for statistical significance reasons may still not be the best if simpler models would suffice
- The all subsets approach allows for this, as one could retain for example, the best model with the lowest BIC
- However the truth is that one study cannot provide definitive evidence that one model would be the best
- What one could do is out of the handful of most viable models, average coefficient values over them
- Even better, use a weighted average, such that coefficients from better fitting models get more weight

Model comparison: Newer approaches

- Bayesian Model Averaging:
  - Selects subsets of unrestricted models and penalizes complexity
  - Retains several viable models each with a p-value that equals the probability of the Model given the data, p(M|D)
    - The opposite of typical NHST p-values
  - Provides coefficients with interpretable p-values
    - Probability that the coefficient is not equal to zero given the final set of plausible models
  - When a final set of viable models are retained (note that while one model is best, several are still plausible good), averages predictor coefficients weighted by model performance (i.e. p(M|D)

Example regarding exploratory regression

- World data from class website
- 10 predictors of Aids rate for various countries[1]
- POPULATN: overall population
- DENSITY: density
- URBAN: urban population
- LIFEEXPF: life expectancy female (essentially redundant with male so arbitrarily chose fem)
- LITERACY
- POP_INCR: population increase in the past year
- GDP_CAP: gross domestic product
- CALORIES: daily caloric intake
- B_TO_D: birth to death ratio
- FERTILTY: fertility rate

All in and Forward

- Regular regression significant predictors
  - Life exp
  - Literacy
  - GDP

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	299.944	86.378		3.472	.001
	Population in thousands	-5.634E-6	.000	-.018	-.216	.830
	Number of people / sq. kilometer	-.004	.008	-.040	-.513	.610
	People living in cities (%)	-.025	.295	-.011	-.085	.932
	Average female life expectancy	-8.335	1.231	-1.721	-6.772	.000
	People who read (%)	2.857	.453	1.192	6.311	.000
	Population increase (% per year))	6.986	16.056	.143	.435	.665
	Gross domestic product / capita	.004	.001	.552	3.799	.000
	Daily calorie intake	.010	.014	.101	.702	.485
	Birth to death ratio	2.054	5.952	.078	.345	.731
	Fertility: average number of kids	5.082	7.161	.178	.710	.481

a. Dependent Variable: Number of aids cases / 100000 people

- Forward[1] selection final model
  - Life exp
  - Literacy
  - GDP
  - Population increase

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	191.139	34.897		5.477	.000
	Average female life expectancy	-2.320	.501	-.479	-4.630	.000
2	(Constant)	301.645	34.304		8.793	.000
	Average female life expectancy	-6.616	.836	-1.366	-7.917	.000
	People who read (%)	2.446	.414	1.020	5.913	.000
3	(Constant)	373.826	35.155		10.634	.000
	Average female life expectancy	-7.767	.797	-1.603	-9.740	.000
	People who read (%)	2.272	.374	.948	6.082	.000
	Gross domestic product / capita	.003	.001	.445	4.256	.000
4	(Constant)	335.828	36.457		9.212	.000
	Average female life expectancy	-8.339	.792	-1.721	-10.528	.000
	People who read (%)	2.830	.413	1.180	6.859	.000
	Gross domestic product / capita	.005	.001	.592	5.200	.000
	Population increase (% per year))	15.513	5.718	.318	2.713	.008

a. Dependent Variable: Number of aids cases / 100000 people

Backward

- Backward (last 3 models shown)
  - Life exp
  - Literacy
  - GDP
  - Population increase

- Stepwise
  - Life exp
  - Literacy
  - GDP
  - Population increase

5	(Constant)	278.118	68.139		4.082	.000
	Average female life expectancy	-8.153	1.055	-1.683	-7.731	.000
	People who read (%)	2.917	.422	1.216	6.920	.000
	Population increase (% per year))	11.086	8.260	.227	1.342	.184
	Gross domestic product / capita	.004	.001	.525	4.069	.000
	Daily calorie intake	.011	.013	.107	.788	.433
	Fertility: average number of kids	5.421	6.510	.190	.833	.408
6	(Constant)	285.574	67.291		4.244	.000
	Average female life expectancy	-7.838	.973	-1.618	-8.055	.000
	People who read (%)	2.893	.419	1.207	6.900	.000
	Population increase (% per year))	10.320	8.180	.211	1.262	.211
	Gross domestic product / capita	.004	.001	.566	4.808	.000
	Fertility: average number of kids	5.759	6.478	.202	.889	.377
7	(Constant)	335.828	36.457		9.212	.000
	Average female life expectancy	-8.339	.792	-1.721	-10.528	.000
	People who read (%)	2.830	.413	1.180	6.859	.000
	Population increase (% per year))	15.513	5.718	.318	2.713	.008
	Gross domestic product / capita	.005	.001	.592	5.200	.000

a. Dependent Variable: Number of aids cases / 100000 people

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	191.139	34.897		5.477	.000
	Average female life expectancy	-2.320	.501	-.479	-4.630	.000
2	(Constant)	301.645	34.304		8.793	.000
	Average female life expectancy	-6.616	.836	-1.366	-7.917	.000
	People who read (%)	2.446	.414	1.020	5.913	.000
3	(Constant)	373.826	35.155		10.634	.000
	Average female life expectancy	-7.767	.797	-1.603	-9.740	.000
	People who read (%)	2.272	.374	.948	6.082	.000
	Gross domestic product / capita	.003	.001	.445	4.256	.000
4	(Constant)	335.828	36.457		9.212	.000
	Average female life expectancy	-8.339	.792	-1.721	-10.528	.000
	People who read (%)	2.830	.413	1.180	6.859	.000
	Gross domestic product / capita	.005	.001	.592	5.200	.000
	Population increase (% per year))	15.513	5.718	.318	2.713	.008

a. Dependent Variable: Number of aids cases / 100000 people

Bayesian Model Averaging

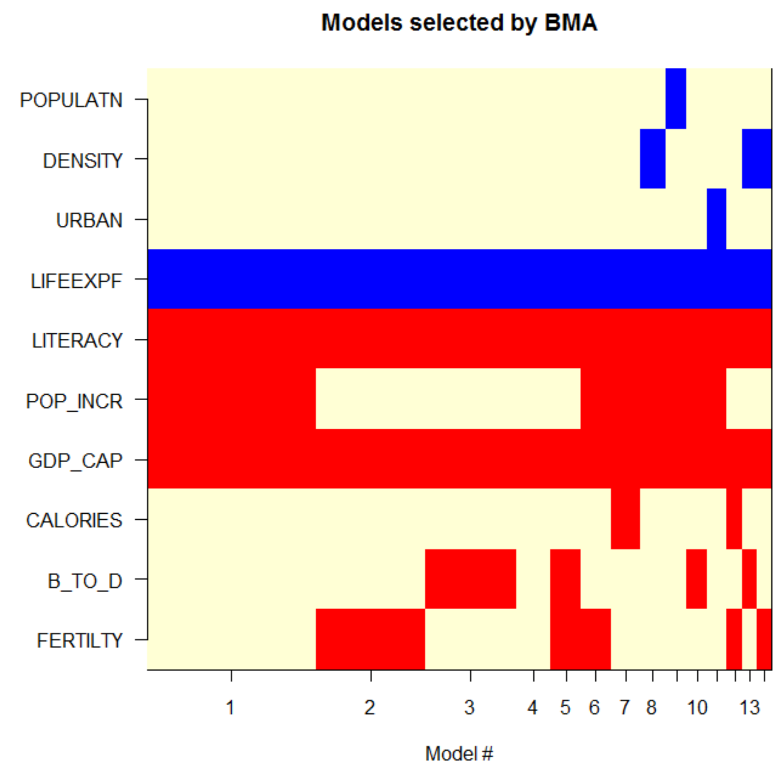
14 models were selected

Best 5 models (cumulative posterior probability = 0.6947):

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	3.244e+02	6.995e+01	335.828266	247.046940	398.932623	373.826389	305.132768
POPULATN	3.4	-3.027e-07	4.715e-06	.	.	.	.	.
DENSITY	8.8	-4.755e-04	2.839e-03	.	.	.	.	.
URBAN	3.1	-1.011e-04	4.845e-02	.	.	.	.	.
LIFEEXPF	100.0	-8.126e+00	1.048e+00	-8.338884	-7.143555	-8.945286	-7.766588	-8.080967
LITERACY	100.0	2.755e+00	4.306e-01	2.830309	2.774806	2.630557	2.272156	2.827419
POP_INCR	50.3	7.495e+00	8.889e+00	15.513096	.	.	.	.
GDP_CAP	100.0	4.333e-03	9.852e-04	0.004559	0.003786	0.004790	0.003427	0.004507
CALORIES	7.2	7.399e-04	4.483e-03	.	.	.	.	.
B_TO_D	25.0	1.292e+00	2.913e+00	.	.	6.346956	.	3.921060
FERTILTY	32.2	3.253e+00	5.671e+00	.	11.594738	.	.	7.700950
nVar				4	4	4	3	5
r2				0.629	0.625	0.623	0.590	0.634
BIC				-56.232515	-55.374972	-54.991118	-53.037193	-52.814231
post prob				0.270	0.176	0.145	0.055	0.049

Example regarding exploratory regression

- BMA
- Top 5 model (14 models viable in total) probabilities range from ~5% to 27%
- Best model
  - Life exp
  - Literacy
  - GDP
  - Population increase
- However across viable models
  - PopInc probability that it does not equal zero only 50%
  - 2nd best model includes Fertility
  - While the evidence for model 1 vs. model 5 is somewhat noticeable, there isn't much distinguishing the others.



Moderators and Mediators

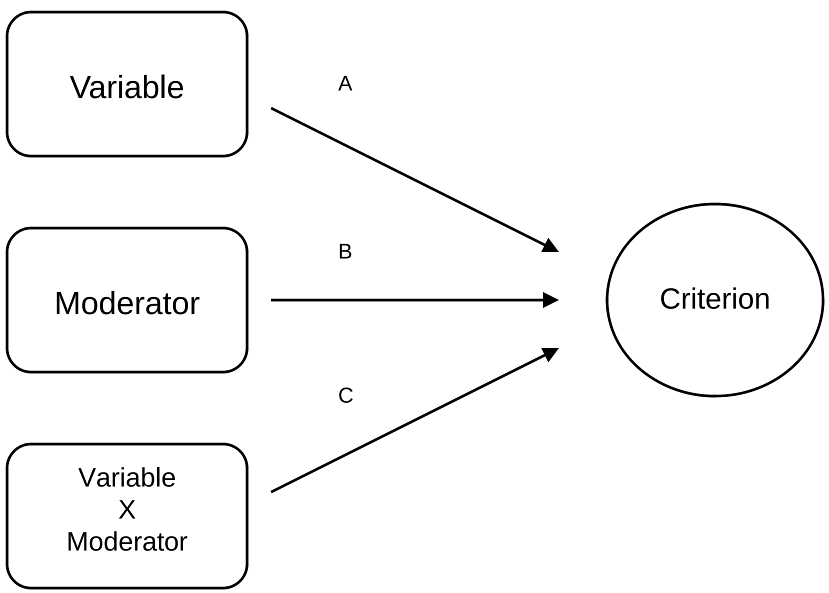
- Moderators and mediators are often a focus of investigation in MR
- First thing to do is keep them straight
- Moderators
  - Same thing as an interaction in ANOVA
- Mediators
  - With the mediator we have a different relationship between the predictor variables, such that one variable in effect accounts for the relationship between a predictor and dependent variable.
- As you can see, these are different models and will require different tests

Moderators

- In typical analysis of variance, interactions are part of the output by default.
- In multiple regression, interactions are not as often looked at (though probably should be more), and usually must be specified.
- As with anova (GLM), the interaction term must be added to the equation as the product of the predictors in question
- Continuous variables should be centered
  - Transform the variable to one in which the mean is subtracted from each response (alternatively one could use z-scores).
  - This will primarily deal with the collinearity issue we'd have with the other IVs correlating with the new interaction term
- In addition, categorical variables must be transformed using dummy or effects coding.
- To best understand the nature of the moderating relationship, one can look at the relationship between one IV at particular (fixed) levels of the other IV

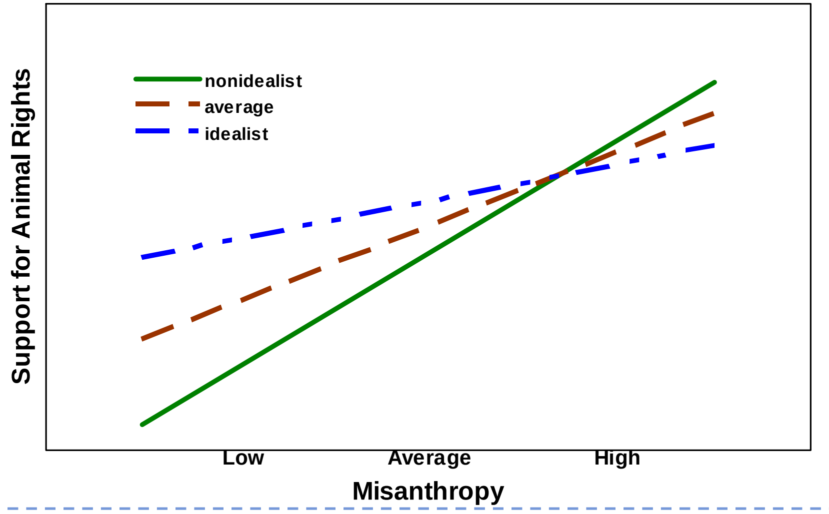
Moderators

- Also, we're not just limited to 2-way interactions
- One would follow the same procedure as before for e.g. a 3-way, and we'd have the same number and kind of interactions as we would in Anova



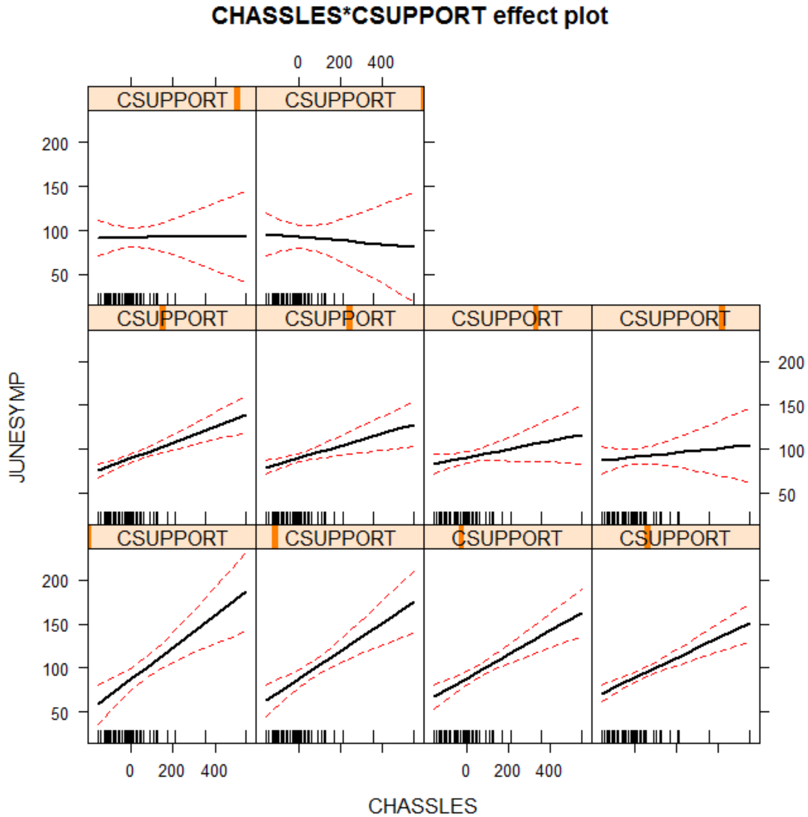
Moderators

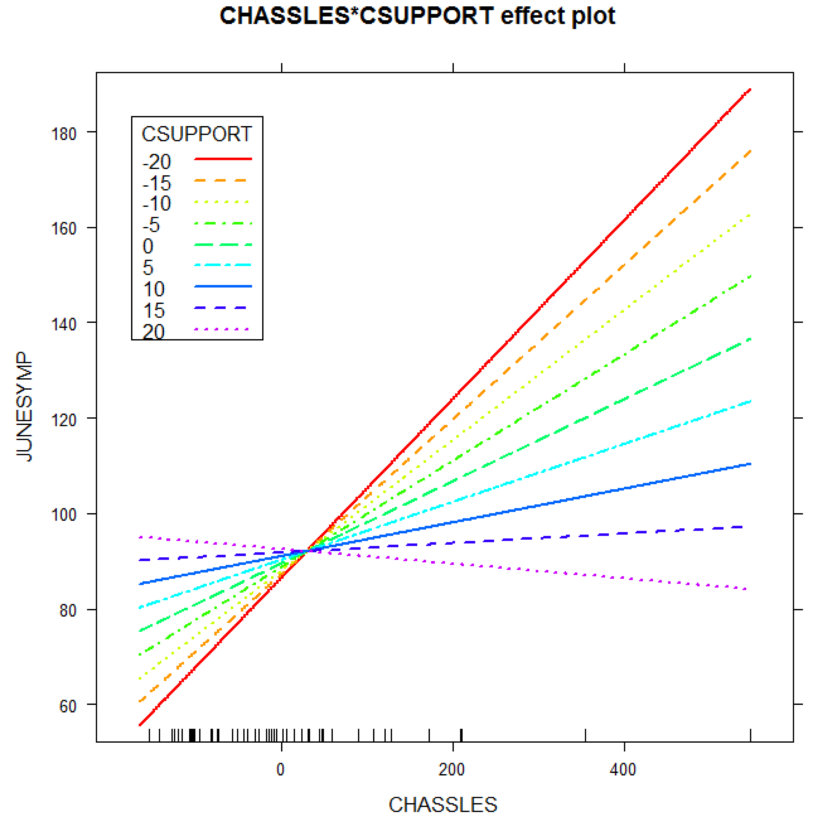
- To actually see what's interaction, we can fix values for one of the variables, computing a new regression equation for each fixed value
- Example[1]
- Support =.303∗ZMisanth + .067∗ZIdeal – .153∗ZInteract
  - For low, medium, and high values of the moderator, simply substitute low (-1), medium (0), and high (1) values of Ideal.
  - **Low Idealism**
  - AR = .303∗Misanth + .067∗(-1) + (-.153)∗(-1)∗Misanth
  - = .456∗Misanth - .067
  - **Mean Idealism**
  - AR = .303∗ Misanth + .067∗(0) + .153∗(0)∗ Misanth
  - = .303∗ Misanth
  - **High Idealism**
  - AR = .303∗ Misanth + .067∗(1) + (-.153)∗(1)∗ Misanth
  - = .15∗ Misanth + .067



Howell Text Example

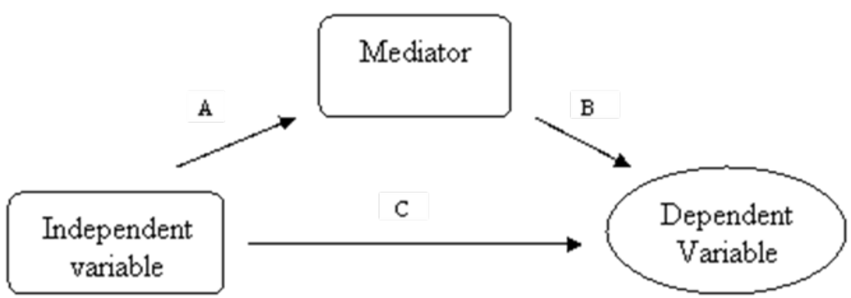
- Number of psychological symptoms predicted by number of daily hassles, social support and their interaction[1]





Mediators

- Mediator variables account for the relationship between a predictor and the dependent variable
  - Theoretically, one variable *causes* another variable which *causes* another
- In other words, by considering a mediator variable, a predictor no longer has a relationship with the DV (or more liberally, just noticeably decreases)



Mediators

- Testing for mediation
- To test for mediators, one can begin by estimating three regression equations
  - (1) The mediator predicted by the independent variable
  - (2) The dependent variable predicted by the independent variable
  - (3) The dependent variable predicted by the mediator and independent variable.
- To begin with, we must have significant relationships found for the equations (1) and (2).
- If the effect of the IV on the DV decreases dramatically when the mediator is present (e.g., its effect becomes nonsignificant), then the mediator may be accounting for the effects of the independent variable in question.
- Overall power for equation (3) is diminished due to the correlation between the independent variable and mediator, and so rigid adherence to the p-value may not tell the whole story.
- Also look at the size of the coefficients, namely, if the coefficient for the independent variable diminishes noticeably with the addition of the mediator to the equation.

Mediators (Sobel test)

- Sobel test for mediation
- Calculate *a*, which equals the unstandardized coefficient of the IV when predicting the DV by itself, and its standard error s_a. From equation (3) take the unstandardized coefficient *b* for the mediator and its standard error s_b.
- To obtain the statistic, input those calculations in the following variant of the Sobel's original formula:

$$z = \frac{a*b}{\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}}$$

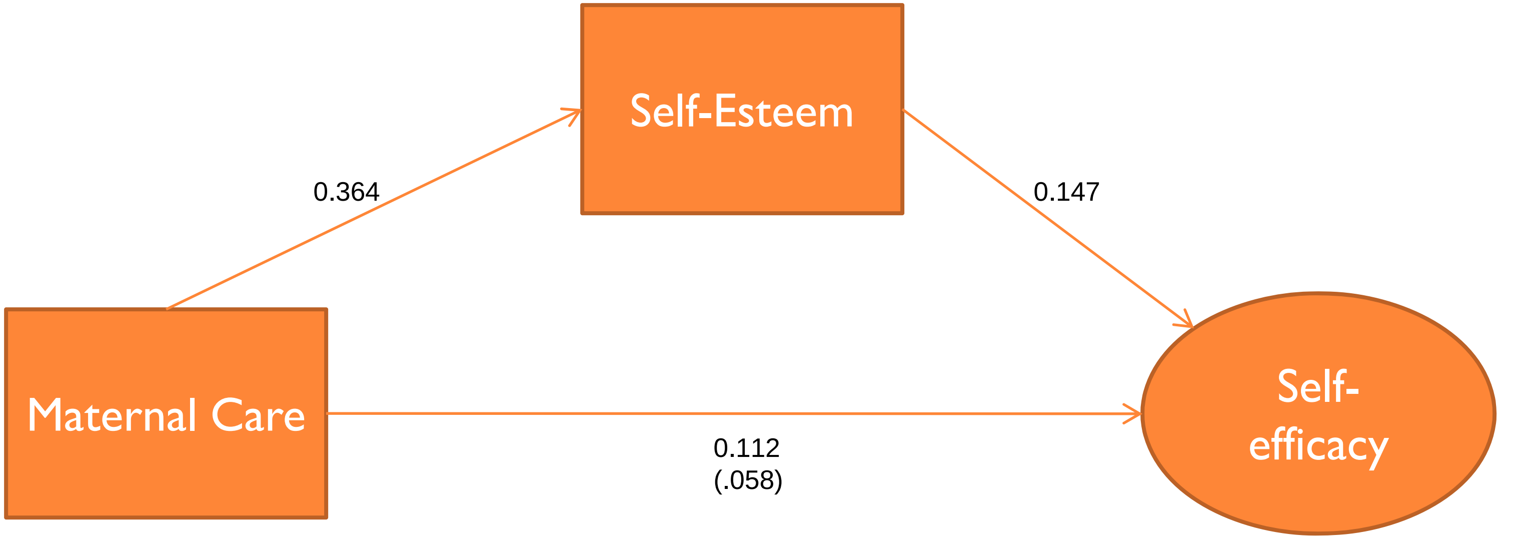
- Test of the null hypothesis that the mediated effect (i.e. a*b or c-c', where c' = the relationship after controlling for the mediator) equals zero in the population

Mediation (Howell example)

- Variables: Level of Maternal care received as a child, Self-esteem as a child, feelings of maternal self-efficacy as a mother with child of 5 months
- R Code
  library(multilevel)
  sobel(MCAREM,SE,MEQ5)
- Model 1
  - IV→ DV
- Model 2
  - Both → DV
- Model 3
  - IV → Mediator
- Indirect Effect:
  - Product of Model.3 coef and med coef from Model.2
- SE
  - Standard error for the Sobel test
- Z
  - Z statistic for the Sobel test.

```
$Model.1
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.260      0.141  23.199    0.000
pred           0.112      0.042   2.677    0.009

$Model.2
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.929      0.173  16.918    0.000
pred           0.058      0.044   1.334    0.185
med            0.147      0.048   3.041    0.003

$Model.3
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.257      0.294   7.687        0
pred           0.364      0.087   4.178        0

$Indirect.Effect
[1] 0.053

$SE
[1] 0.022

$z.value
[1] 2.459
```

Mediation graphic



Moderators and Mediators

- For mediation correlations among variables should not just be significant, but meaningful
  - Just b/c r = .2, p = .05 for your sample size doesn't make it a noticeable effect to determine mediation with
- Mediation should rarely be thought of outside the context of path analysis/sem, as two predictor models are overly simplistic
- If you find the word 'cause' untenable, mediation is not appropriate[2]
- Mediators and moderators can be used in the same analysis[1]
- More complex designs are often analyzed with path analysis or SEM