psyc3010 lecture 8

standard and hierarchical multiple regression

last week: correlation and regression
Next week: moderated regression

last week → this week

- last week we revised ***correlation & regression*** and took a look at some of the underlying principles of these methods [partitioning variance into SS regression ($\hat{Y}$ - $\bar{Y}$) and SS residual (Y - $\hat{Y}$).]
- We extended these ideas to the multiple predictor case (***multiple regression***) and touched upon indices of ***predictor importance***
- these week we go through two full examples of multiple regression
  - **standard regression**
  - **heirarchical regression**

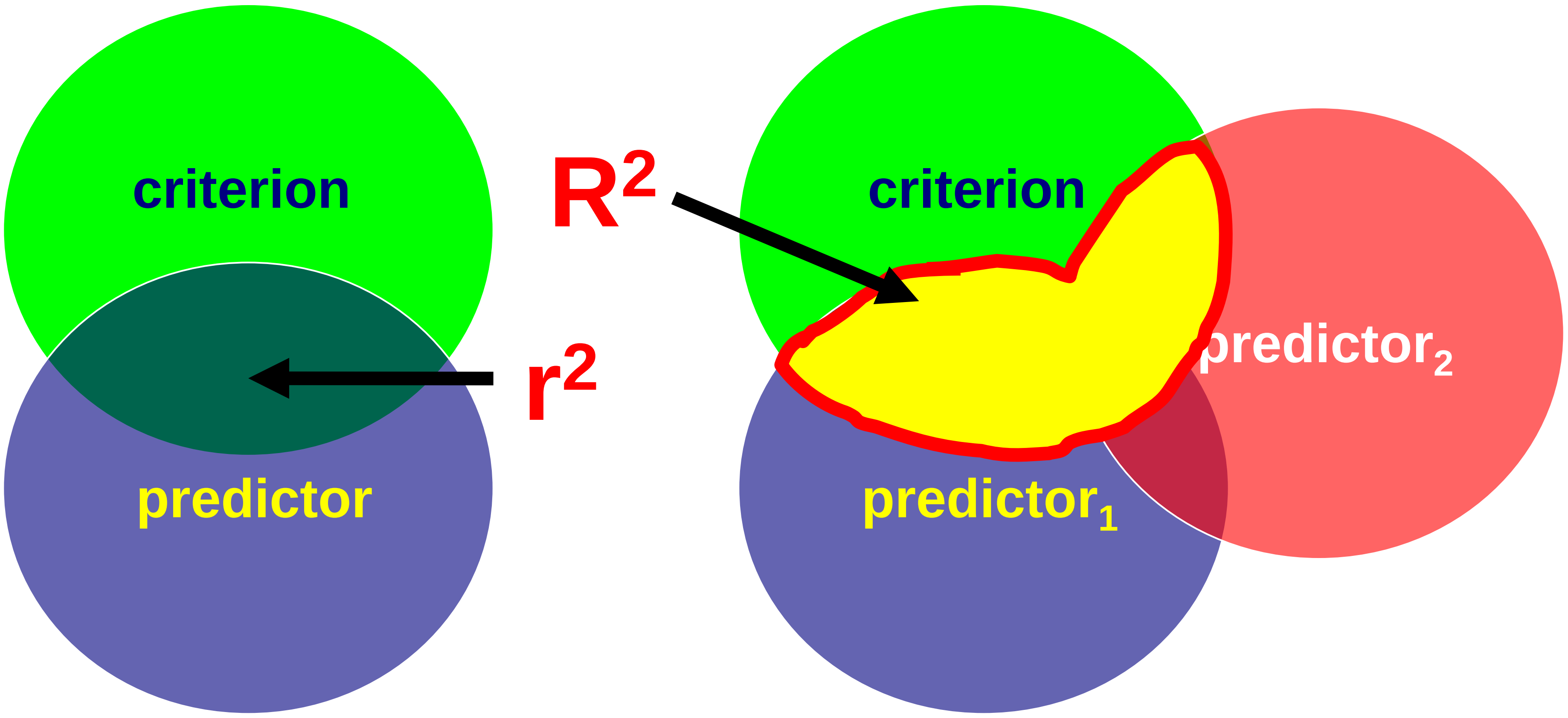
Indices of ***predictor importance:***

- r [Pearson or **zero-order correlation**] – a scale free measure of association – the standardised covariance between two factors
- r^2 [the **coefficient of determination**] – the proportion of variability in one factor (e.g., the DV) accounted for by another (e.g. an IV).
- b [unstandardised slope or **unstandardised regression coefficient**] – a scale dependent measure of association, the slope of the regression line – the change in units of Y expected with a 1 unit increase in X
- β [standardised slope or **standardised regression coefficient**] – a scale free measure of association, the slope of the regression line if all variables are standardised – the change in standard deviations in Y expected with a 1 standard deviation increase in X, controlling for all other predictors. $\beta = r$ in bivariate regression (when there is only one IV).
- pr^2 [**partial correlation squared**] – a scale free measure of association controlling for other IVs -- the proportion of residual variance in the DV (after other IVs are controlled for) uniquely accounted for by the IV.
- sr^2 [**semi-partial correlation squared**] – a scale free measure of association controlling for other IVs -- the proportion of total variance in the DV uniquely accounted for by the IV.

Comparing the different *r*s

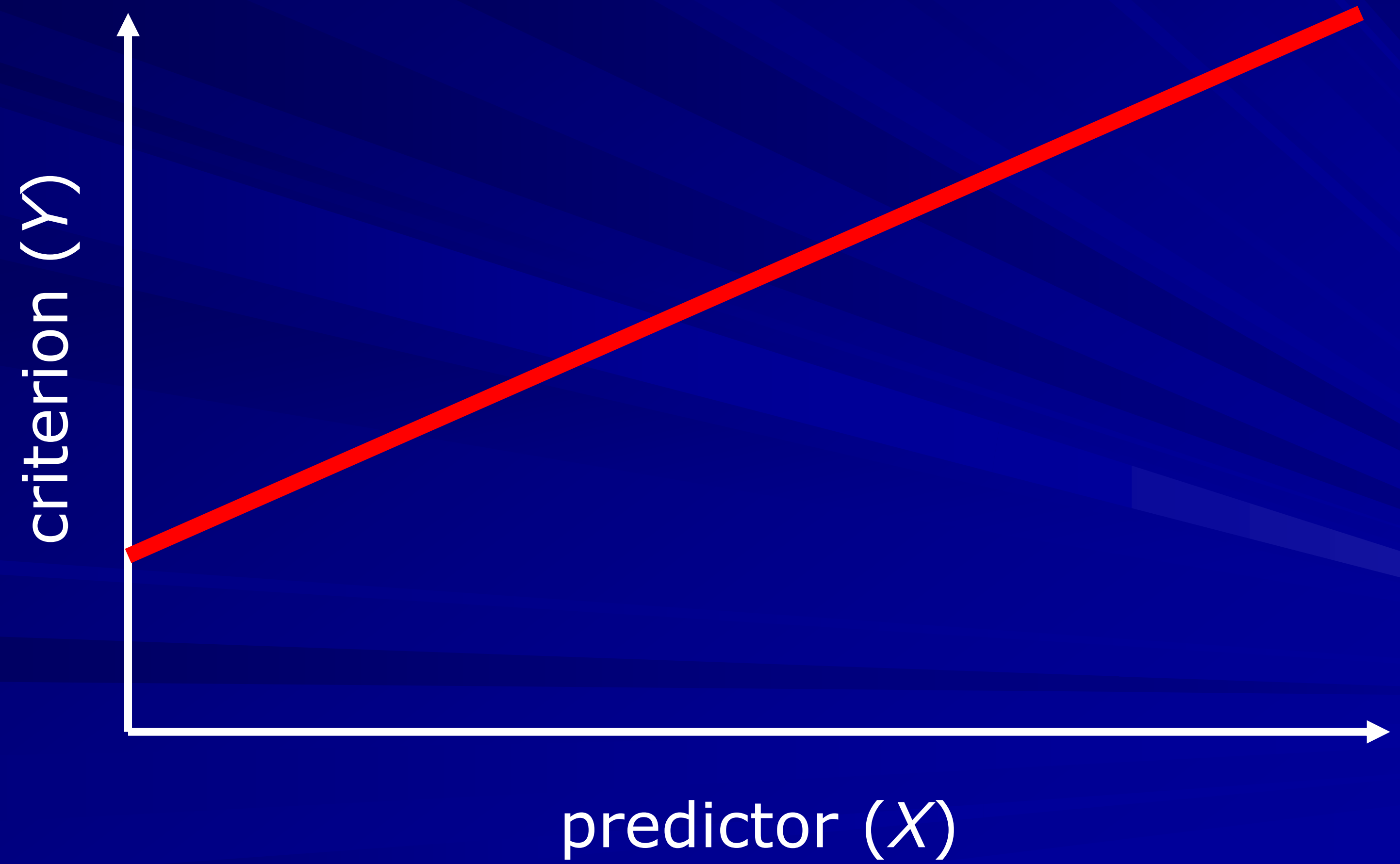
- The **zero-order (Pearson's) correlation** between IV and DV ignores extent to which IV is correlated with other IVs.
- The **semi-partial correlation** deals with unique effect of IV on total variance in DV – usually what we are interested in.
  - Conceptually similar to eta squared
  - Confusion alert: in SPSS the semi-partial r is called the part correlation. No one else does this though.
- The **partial correlation** deals with unique effect of the IV on residual variance in DV. More difficult to interpret – most useful when other IVs = control variables.
  - Conceptually similar to 'partial eta squared'
- Generally r > spr and pr > spr

bivariate vs multiple regression - model coefficient of determination



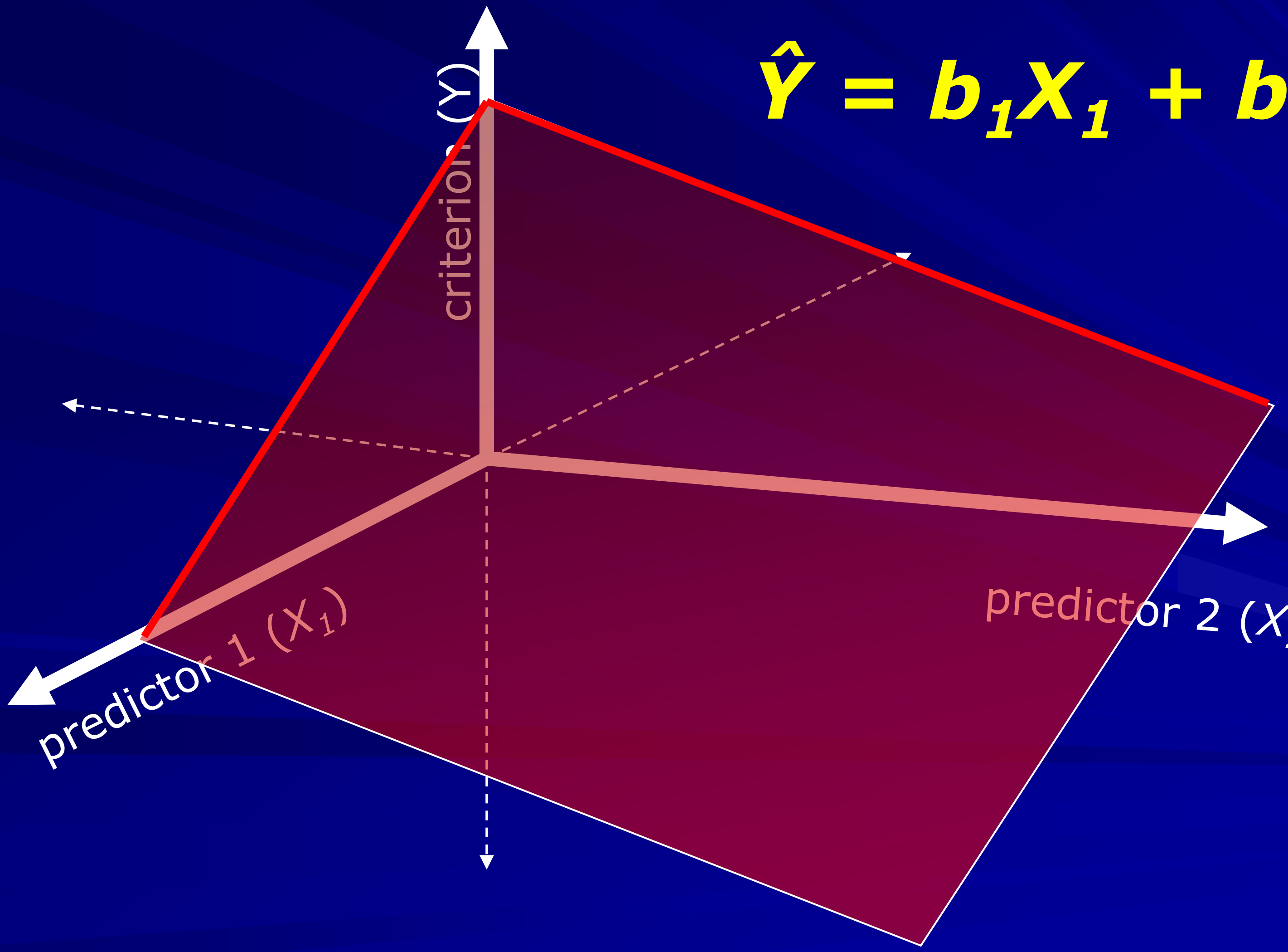
the linear model – one predictor
(2D space)

$$\hat{Y} = bX + a$$



the linear model – two predictors
(3D space)

$$\hat{Y} = b_1X_1 + b_2X_2 + a$$



the linear model – 2 predictors

- **criterion scores are predicted using the best *linear combination* of the predictors**
  - similar to the line-of-best-fit idea, but it becomes the *plane*-of-best-fit
  - equation derived according to the least-squares criterion – such that $\Sigma(Y-\hat{Y})^2$ is minimized
    - b_1 is the slope of the plane relative to the X_1 axis,
    - b_2 is the slope relative to the X_2 axis,
    - a is the point where the plane intersects the Y axis (when X_1 and X_2 are equal to zero)
- the idea extends to 3+ predictors but becomes tricky to represent graphically (i.e., hyperspace)

example

- new study...examine the amount of variance in academic achievement (GPA) accounted for by…
  - Minutes spent studying per week (questionnaire measure)
  - motivation (questionnaire measure)
  - anxiety (questionnaire measure)
- can use multiple regression to asses how much variance the predictors explain as a set (R^2)
- can also assess the relative importance of each predictor (r, b, β, pr^2, sr^2).

data table

subject	study (X_1)	motivation (X_2)	anxiety (X_3)	GPA (Y)
1	104	12	1	5.5
2	109	13	9	5.7
3	123	9	2	5.5
4	94	15	11	5.3
5	114	15	2	6.1
6	91	7	9	4.9
7	100	5	1	4.5
…				
29	107	10	6	5.9
30	119	8	2	6.0

preliminary statistics

Descriptive Statistics

	Mean	SD	N	alpha
study time	97.967	8.915	30	.88
motivation	14.533	4.392	30	.75
anxiety	4.233	1.455	30	.85
GPA	5.551	2.163	30	.82

Correlations

	ST	MOT	ANX	GPA
study time	1.00			
motivation	.313	1.00		
anxiety	.256	.536	1.00	
GPA	.637	.653	.505	1.00

preliminary statistics

Descriptive Statistics

	Mean	SD	N	alpha
study time	97.967	8.915	30	.88
motivation	14.533	4.392	30	.75
anxiety	4.233	1.455	30	.85
GPA	5.551	2.163	30	.82

means and standard deviations are used to obtain regression estimates, and are reported as preliminary stats when one conducts MR. They are needed to interpret coefficients, although descriptively they are not as critical in MR as they are for t-tests and anova

preliminary statistics

Descriptive Statistics

Cronbach's α is an index of internal consistency (reliability) for a continuous scale

best to use scales with high reliability *(α >.70)* if available – less error variance

alpha
.88
.75
.85
.82

Correlations

	ST	MOT	ANX	GPA
study time	1.00			
motivation	.313	1.00		
anxiety	.256	.536	1.00	
GPA	.637	.653	.505	1.00

preliminary statistics

Descriptive Statistics

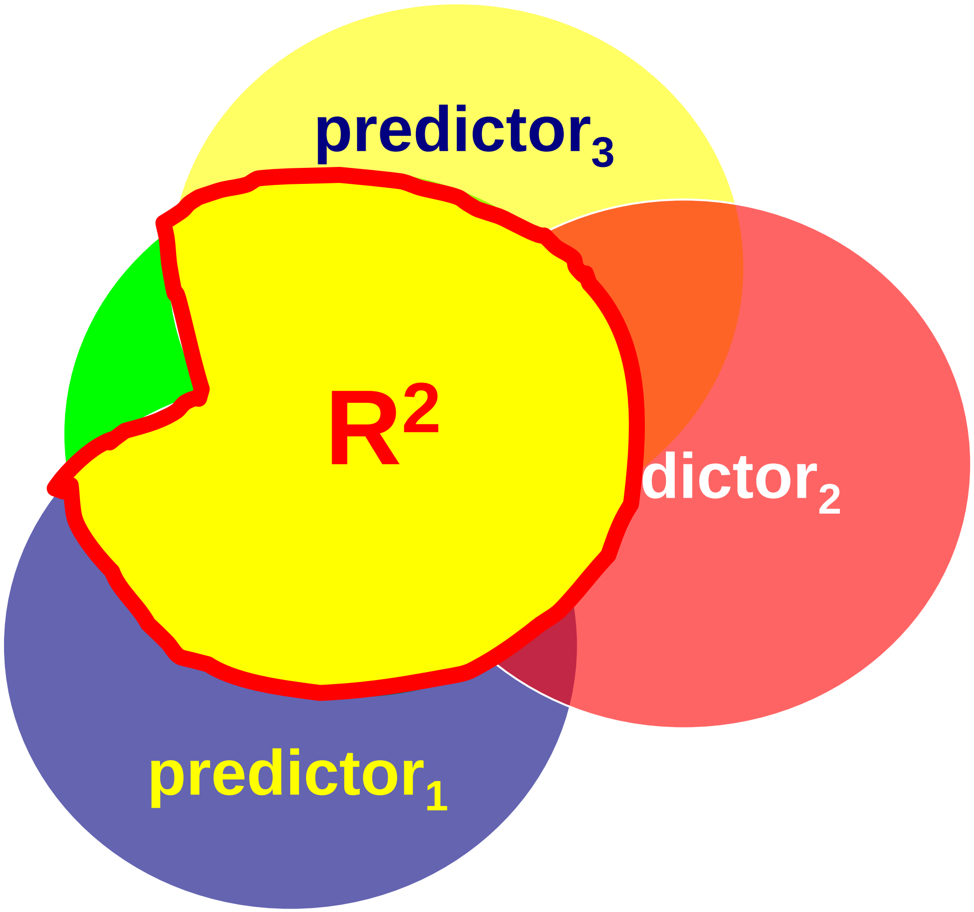
the correlation matrix, tells you the extent to which each predictor is related to the criterion (called **validities**), as well as intercorrelations among predictors (**collinearities**).

to maximise R^2 we want predictors that have high validities and low collinearity

Correlations

	IQ	MOT	ANX	GPA
IQ	1.00			
motivation	.313	1.00		
anxiety	.256	.536	1.00	
GPA	.637	.653	.505	1.00

principle of parsimony:



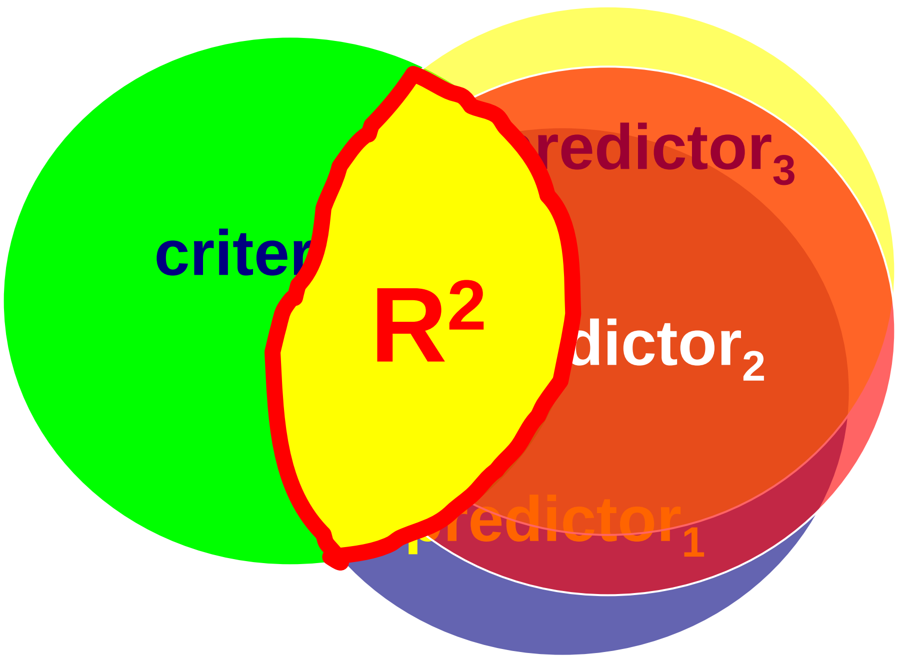
predictors are:

-highly correlated with criterion

-have low(er) correlations with one another

-Good

principle of parsimony:



predictors are:

-highly correlated with criterion

-highly correlated with one another

-Bad. Probably would delete some of the redundant IVs.

regression solution

$$\hat{Y} = b_1X_1 + b_2X_2 + a$$

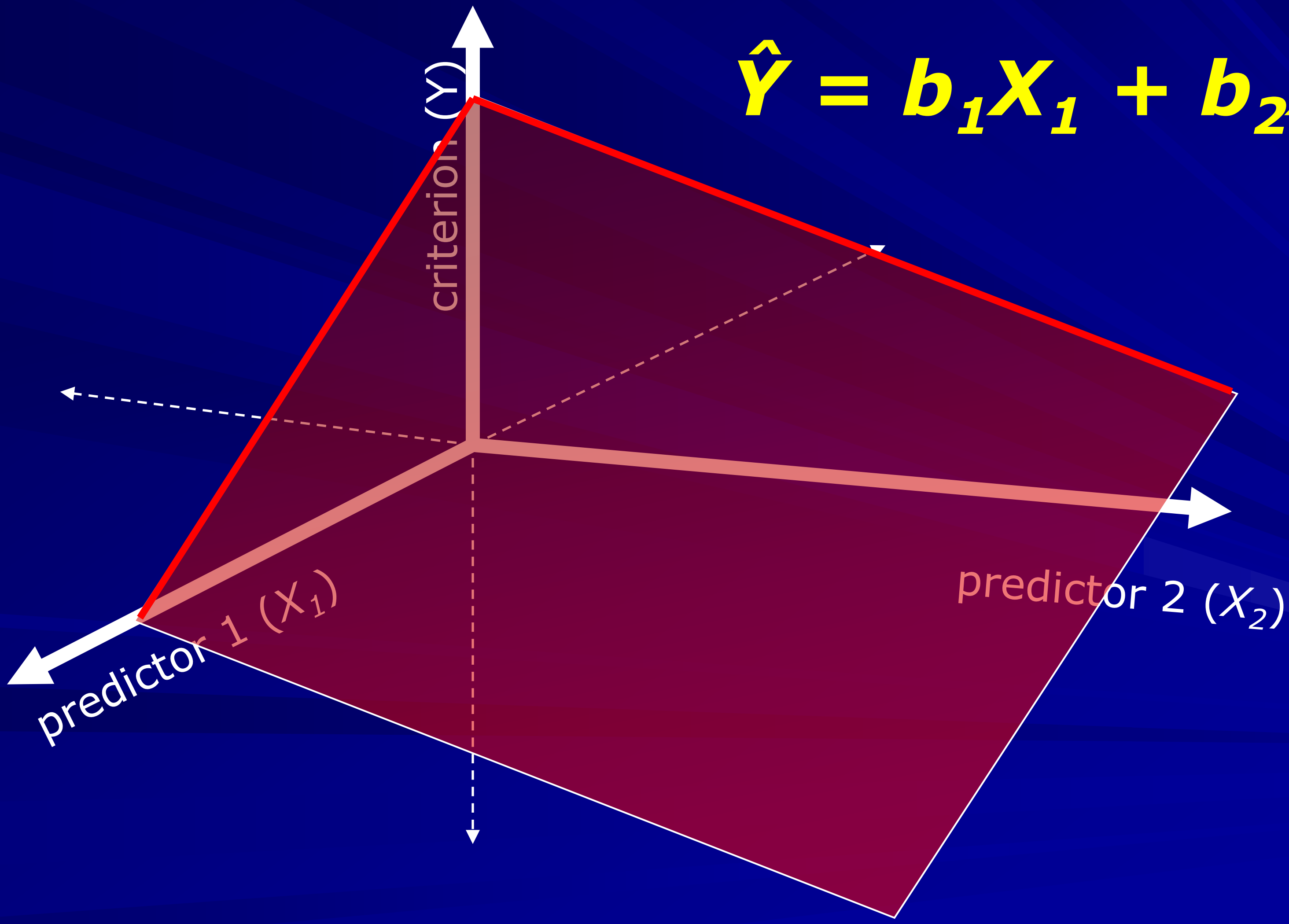
- calculation for multiple regression requires the solution of a set of parameters (one slope for each predictor – b values)
- E.g. with 2 IVs, the 2 slopes define the **plane-of-best-fit** that goes through the 3-dimensional space described by plotting the DV against each IVs
- Pick bs so that deviations of dots from the plane are minimized
  - these weights are derived through matrix algebra - beyond the scope of this course
- understand how with one variable, we model Y hat with a line described by 2 parameters (bX + a);
- with two, model Y hat as a plane described by 3 parameters (b1X1 + b2X2 + a)
- with p predictors, model Yhat as a p-dimensional hyperspace blob with p + 1 parameters (constant, and a slope for each IV).
- So Ŷ, the predicted value of Y, is modeled with a linear composite formed by multiplying each predictor by its regression weight / slope / coefficient (just like a linear contrast) and adding the constant:

$$\hat{Y} = .79\,\boldsymbol{ST} + 1.45\,\boldsymbol{MOT} + 1.68\,\boldsymbol{ANX} - 95.02$$

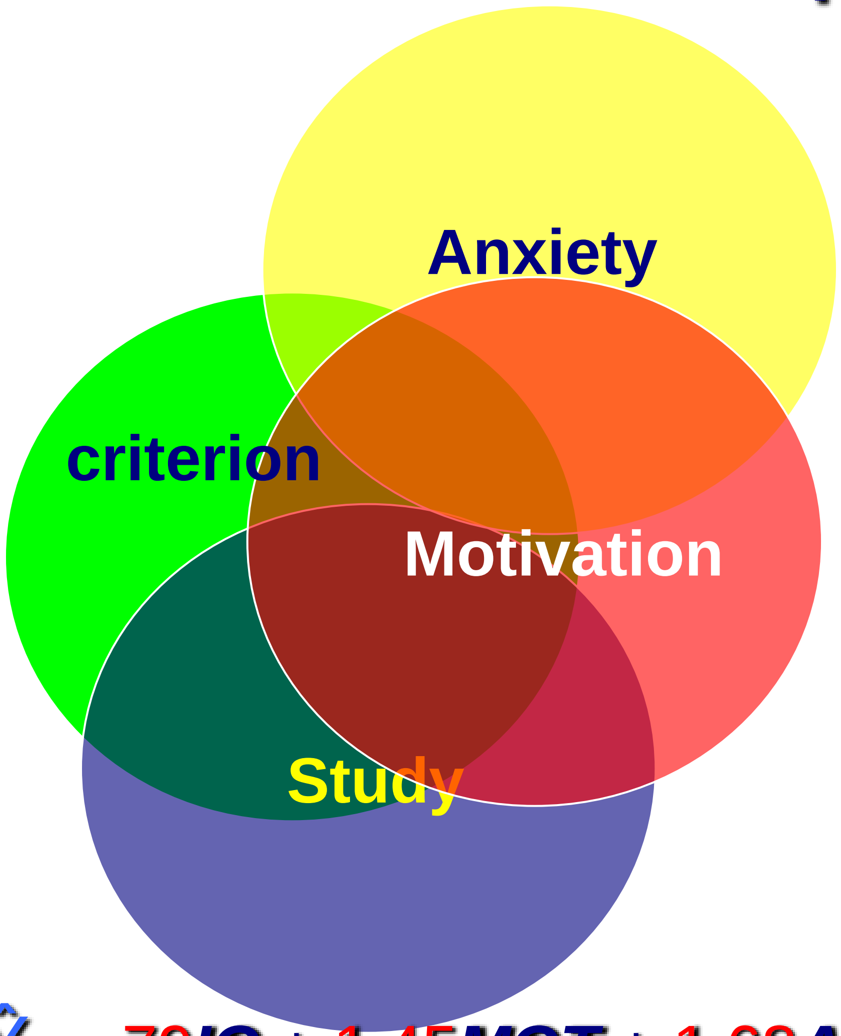
- the criterion (GPA) is regressed on this linear composite

the linear model – two predictors
(3D space)

$$\hat{Y} = b_1X_1 + b_2X_2 + a$$

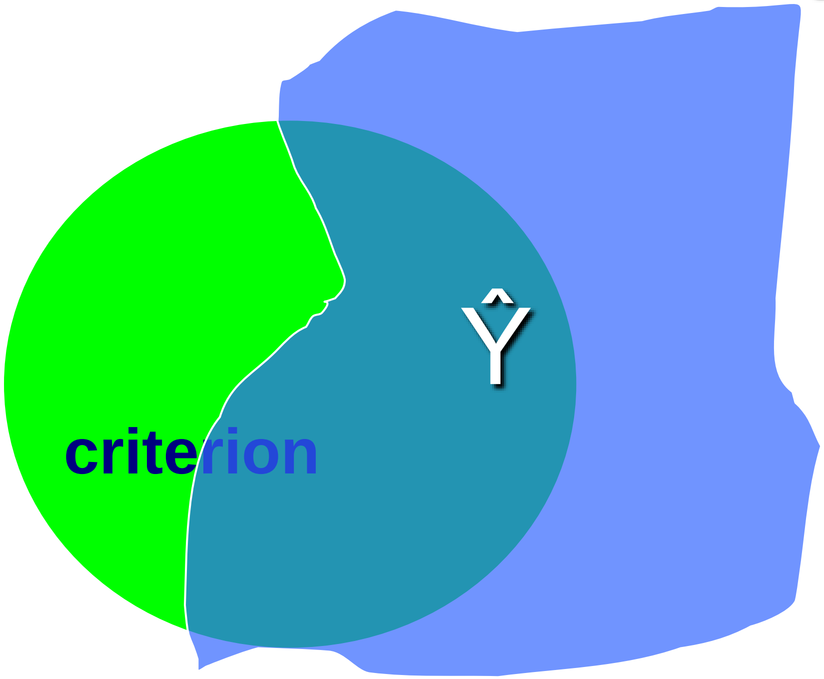


the linear composite



$\hat{Y} = .79\mathit{IQ} + 1.45\mathit{MOT} + 1.68\mathit{ANX} - 95.02$

the linear composite



…so we end up with two overlapping variables just like in bivariate regression *(only one is blue and weird and wibbly, graphically symbolising that underlying the linear relationship between the DV and Y hat, the linear composite, is a 4-dimensional space defined by the 3 IVs and the DV)*

the model: *R* and *R²*

- Despite the underlying complexity, the multiple correlation coefficient (R) is just a bivariate correlation between the criterion (GPA) and the best linear combination of the predictors ($\hat{Y}$)
- i.e., $R^2 = r^2_{Y\hat{Y}}$

  where $\hat{Y} = .79\boldsymbol{ST} + 1.45\boldsymbol{MOT} + 1.68\boldsymbol{ANX} - 95.02$
- accordingly, we can treat the model R exactly like r, ie:

i. calculate R adjusted:

$$\sqrt{1-\frac{(1-R^2)(N-1)}{N-2}}$$

ii. square R to obtain amount of variance accounted for in Y by our linear composite ($\hat{Y}$)

iii. test for statistical significance

the model: *R* and *R*²

- 1. In this example, R = .81
  - so R adj = $$\sqrt{1-\frac{(1-.65)(30-1)}{30-2}}$$

    = .798

- 2. R^2 = .65 (.638 adjusted)

"…therefore, 65% of the variance in participants' GPA was explained by the combination of their study time, motivation, and anxiety."

the model: ***R*** and ***R***^{2}

- 3. The overall model (R^2) is tested for significance –
  - H_0 – the relationship between the predictors (as a group) and the criterion is zero
  - H_1 – the relationship between the predictors (as a group) and the criterion is different from zero

$$F = \frac{(N - p - 1)R^2}{p(1 - R^2)}$$

$$= \frac{(30 - 3 - 1) \times .6518}{3 \times (1 - .6518)}$$

$$= 16.23$$

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Reminder of t-test for r

$$df = p, N - p - 1$$

Test of R^2 (analysis of regression)

$$F = \frac{(N-p-1)R^2}{p(1-R^2)} \qquad df = p, N-p-1$$

$$= \frac{R^2 / p}{(1-R^2)/(N-p-1)}$$

What we know (can account for) — R^2

What we don't know (can't account for) — $(1-R^2)$

= variance accounted for / df
———————————————
variance *not* accounted for (error) / df

$$= \frac{\text{MS REGRESSION}}{\text{MS RESIDUAL}}$$

the model: *R* and *R^2*

- Or perform same test via ***analysis of regression:***

$$SS_Y = SS_{Regression} + SS_{Residual}$$

$SS_Y = \sum (Y-\bar{Y})^2 = (5.5 - 5.551)^2 + (5.7 - 5.551)^2 \ldots$
$= 6667.46$

$SS_{Regression} = \sum (\hat{Y} - \bar{Y})^2 = (6.22 – 5.551)^2 + \ldots$
$= 4346.03$

$SS_{Residual} = \sum (Y - \hat{Y})^2$
$= SS_Y - SS_{Regression} = 6667.46 - 4346.03$
$= 2321.43$

the model: *R* and *R*^{2}

Summary Table for Analysis of Regression:

Model	Sums of Squares	df	Mean Square	F	sig
Regression	4346.03	3	1448.68	16.23	.000
Residual	2321.43	26	89.29		
Total	6667.46	29			

$$df = \mathrm{p}, N - \mathrm{p} - 1$$

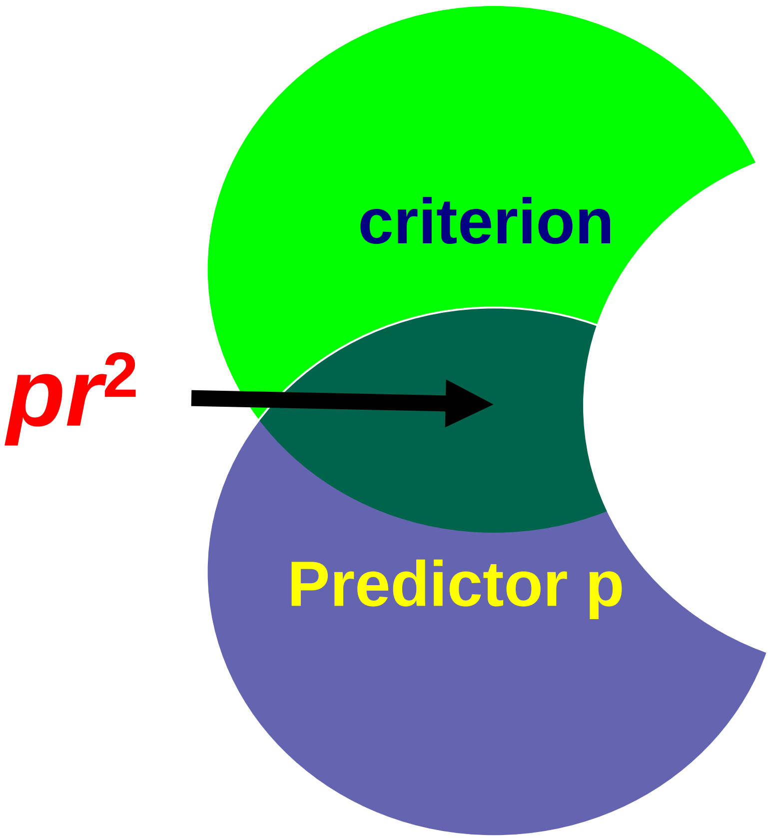
The model including study time, motivation, and anxiety accounted for significant variation in participants' GPA, F(3, 26) = 16.23, p < .001, R2 = .65.

individual predictors

- we already have our bivariate correlations (r) between each predictor and the criterion. In addition, SPSS gives us:
  – b – (unstandardised) partial regression coefficient
  – β - standardised partial regression coefficient
  – pr – partial correlation coefficient
  – sr – semi-partial correlation coefficient

  (as the calculations for these are all matrix algebra we will bypass that….)

pr – partial correlation coefficient



pr is the correlation between predictor p and the criterion, with the variance shared with the ***other predictors partialled out***

Can write $r_{01.2}$ [partial r between 0 and 1 excluding shared variance with 2]

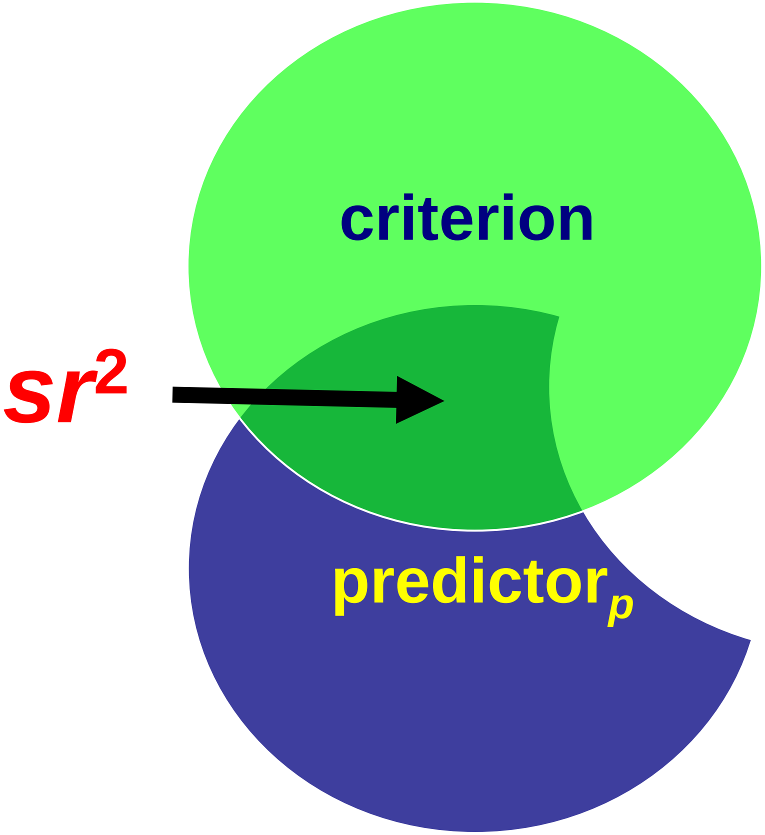
pr² indicates the proportion of residual variance in the criterion (DV variance left unexplained by the other predictors) that is explained by predictor p

pr_{ST} = .581; pr_{IQ}^2 = 33.7%

pr_{MOT} = .562 ; pr_{MOT}^2 = 31.5%

pr_{ANX} = .293; pr_{ANX}^2 = 8.5%.

sr – semi-partial correlation coefficient



sr is the correlation between predictor *p* and the criterion, with the variance shared with the **other predictors *partialled out*** of predictor *p*

Can write $r_{0(1.2)}$ [partial r between 0 and (1 excluding 2)]

sr² indicates the **unique** contribution to the total variance in the DV explained by predictor *p*

sr_{ST} = .469; sr_{IQ}^2 = 21.9%

sr_{MOT} = .411; sr_{MOT}^2 = 16.9%

sr_{ANX} = .224; sr_{ANX}^2 = 5%

shared variance ≈ 21% (R2 - ∑sr2, 65-44%)

$\hat{Y} = b_1X_1 + b_2X_2 + a$ → Zero-order coefficient – doesn't take other IVs into account

$\hat{Y} = b_{Y1.2}X_1 + b_{Y2.1}X_2 + a$

$b_{Y1.2}$ <u>first-order</u> coefficient

$b_{Y1.2} \neq b_{Y1}$ unless $r_{12} = 0$

→ First-order coefficient – takes 1 other IV into account

$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 + a$

$\hat{Y} = b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}X_3 + a$

$b_{Y1.23}$ <u>second-order</u> coefficient

→ Takes 2 other IVs into account

All reported coefficients (e.g. in SPSS) are highest order coefficients

tests of *b*s:

test importance of the predictor <u>in the context</u> of all the other predictors

divide *b* by its standard error. $df = N - p - 1$

$$t_{b1} = \frac{.789247}{.208497} = 3.785^* \quad \text{ST}$$

$$t_{b2} = \frac{1.453540}{.484508} = 3.000^* \quad \text{MOT}$$

$$t_{b3} = \frac{1.678871}{1.437221} = 1.168 \quad \text{ANX}$$

- ST contributes significantly to prediction of DV, after controlling for the other predictors, and so does MOT
- though a valid zero-order predictor of DV, anx does <u>not</u> contribute to the prediction, given ST and MOT

Importance of predictors

can't rely on *r*s (zero-order), because the predictors are interrelated

(predictor with a significant *r* may contribute nothing, once others are included; e.g., ANX)

partial regression coefficient (*b*s):

adjusted for correlation of the predictor with the other predictors

but

can't use relative magnitude of *b*s, because scale-bound

(importance of a given *b* depends on unit and variability of measure)

Standardized regression coefficients (*β* s):

rough estimate of relative contribution of predictors, because use same metric

can compare ***β*** s within a regression equation (but <u>not</u> necessarily across groups & settings – in that standard deviation of variables change)

Standardized regression coefficients:

$$\beta 1 = b1 \cdot \frac{s_1}{s_Y}$$

when IVs are not correlated:

$\beta = r$

when IVs are correlated:

β s (magnitudes, signs) are affected by pattern of correlations among the predictors

$$\hat{Z}_Y = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \ldots + \beta_p Z_p$$

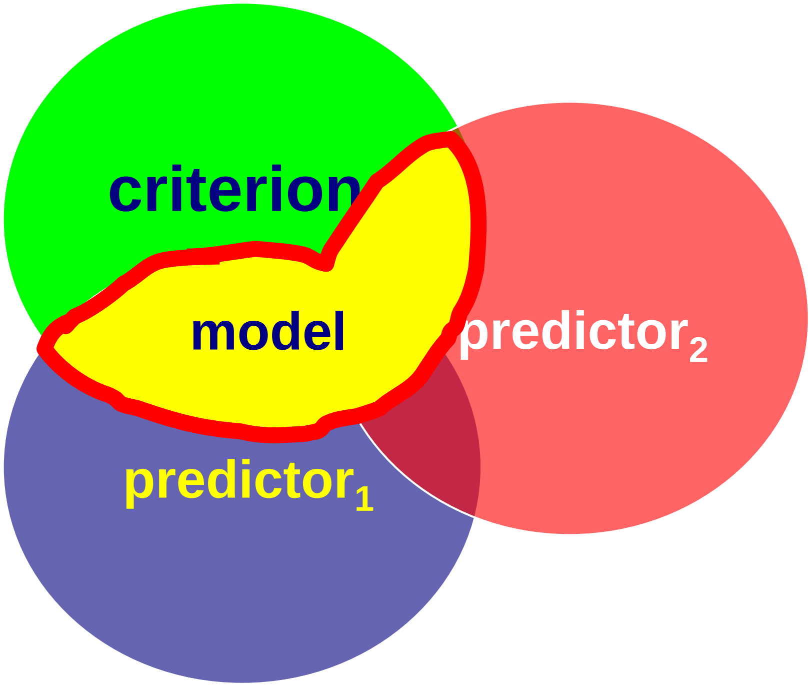
$$\hat{Z}_Y = .46\, Z_{ST} + .42\, Z_{MOT} + .16\, Z_{ANX}$$

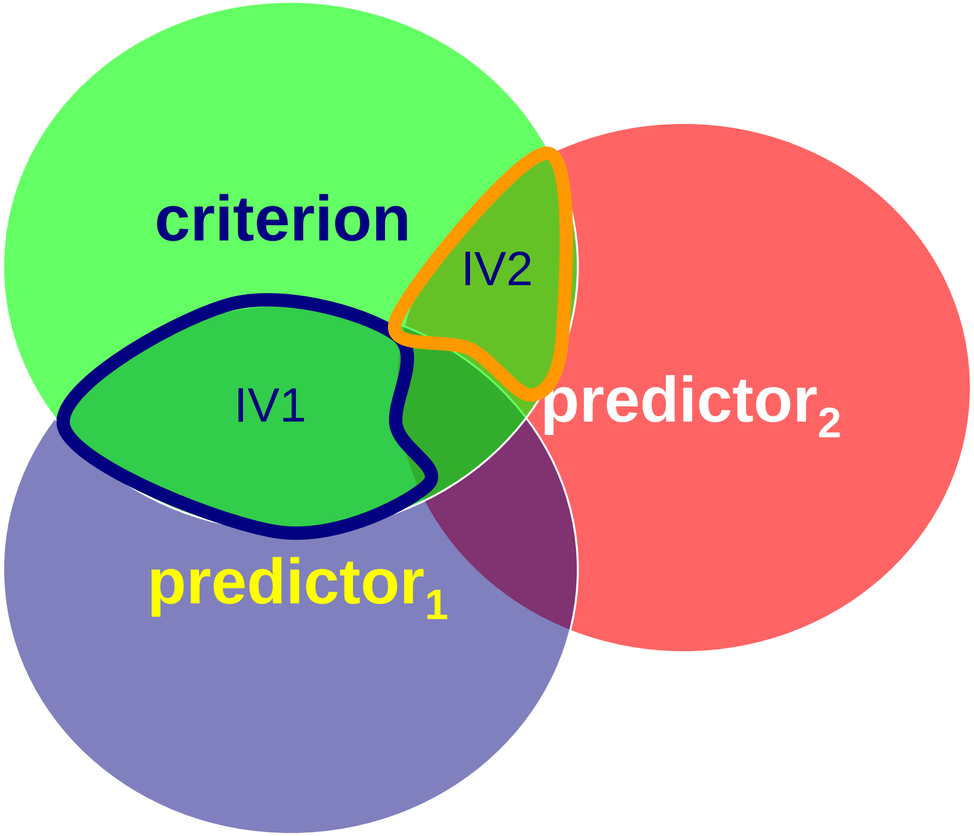
a one-SD increase in ST (<u>with all other variables held constant)</u> is associated with an increase of .46 SDs in DV

standard vs hierarchical regression

- **standard**
  - all predictors are entered ***simultaneously***
  - each predictor is evaluated in terms of what it adds to prediction beyond that afforded by all others
  - most appropriate when IVs are not intercorrelated
- **hierarchical**
  - predictors are entered ***sequentially*** in a pre-specified order
  - each predictor is evaluated in terms of what it adds to prediction *at its point of entry*
  - order of prediction based upon logic/theory

standard vs hierarchical multiple regression

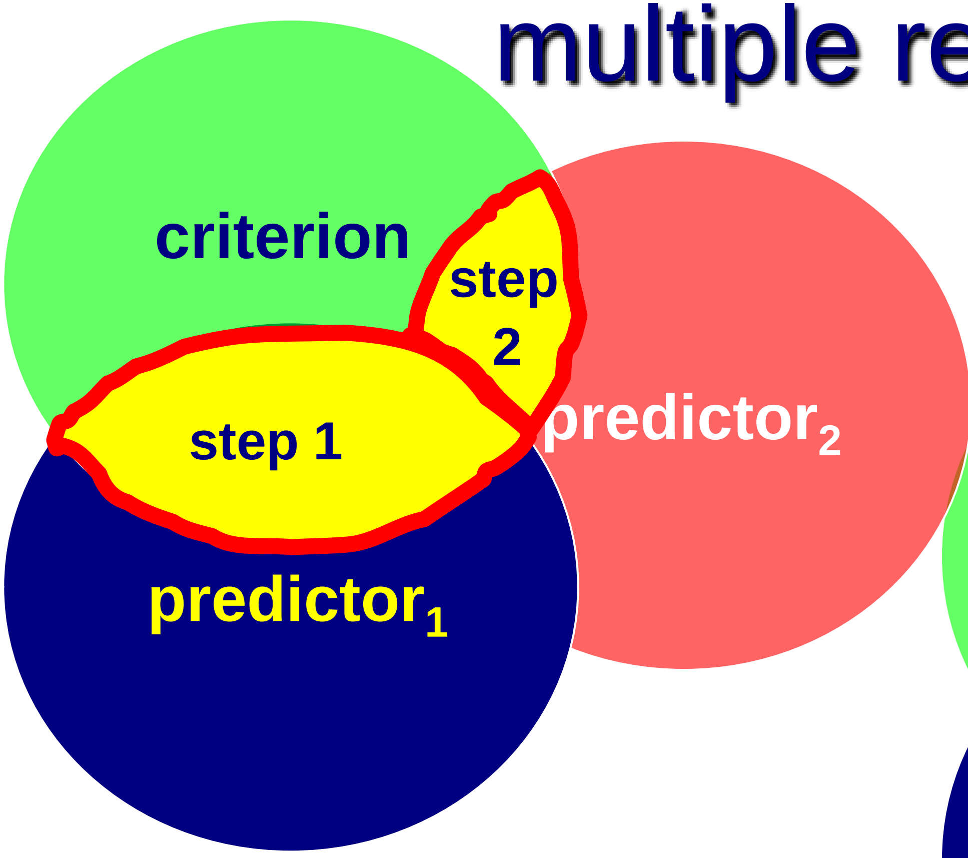


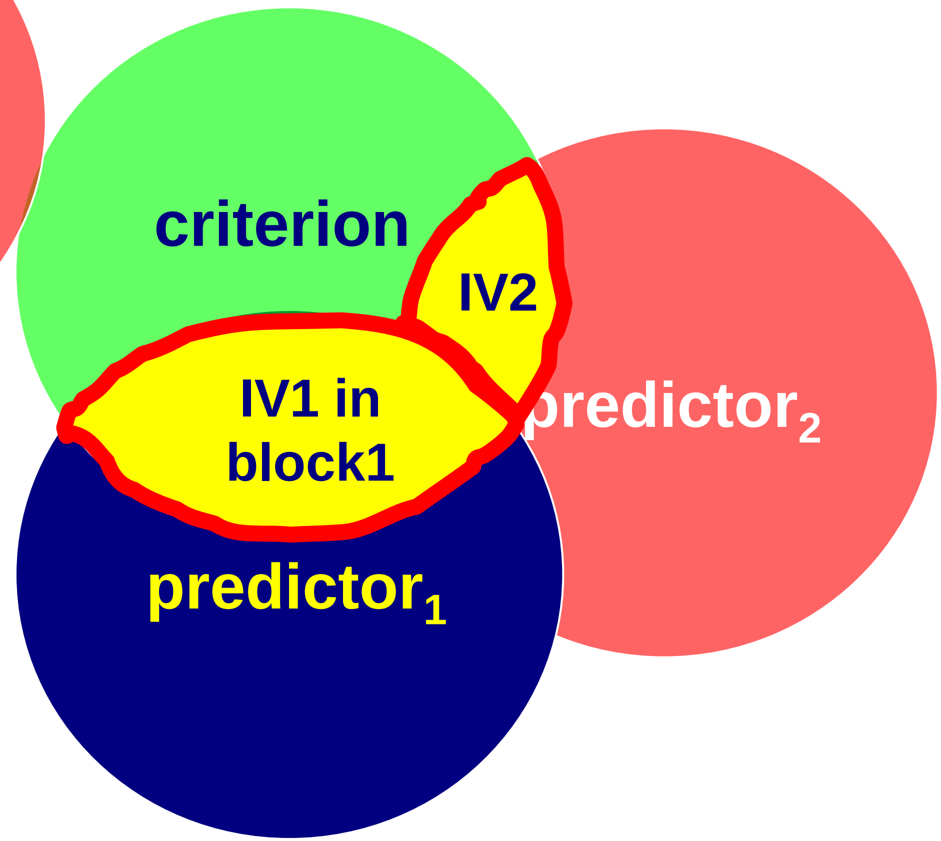


standard multiple regression:

- Model R2 assessed in 1 step
- b for each IV based on unique contribution only

standard vs hierarchical multiple regression





hierarchical multiple regression:

- Model R2 assessed in > 1 step
- Each step (“block”) add more IVs
- b for first IV based on total contribution; later IV on unique contribution

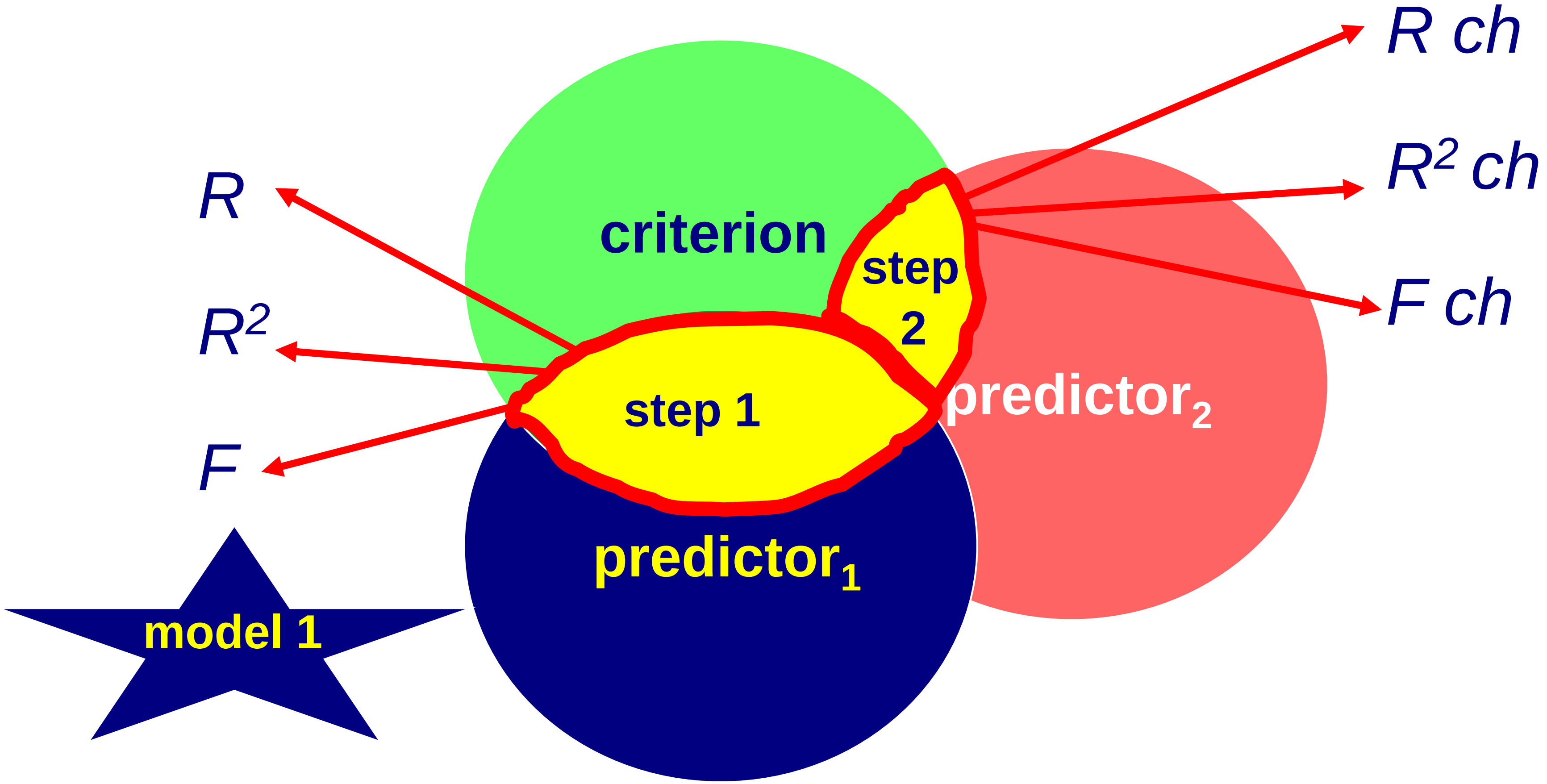
hierarchical regression

some rationales for order of entry:

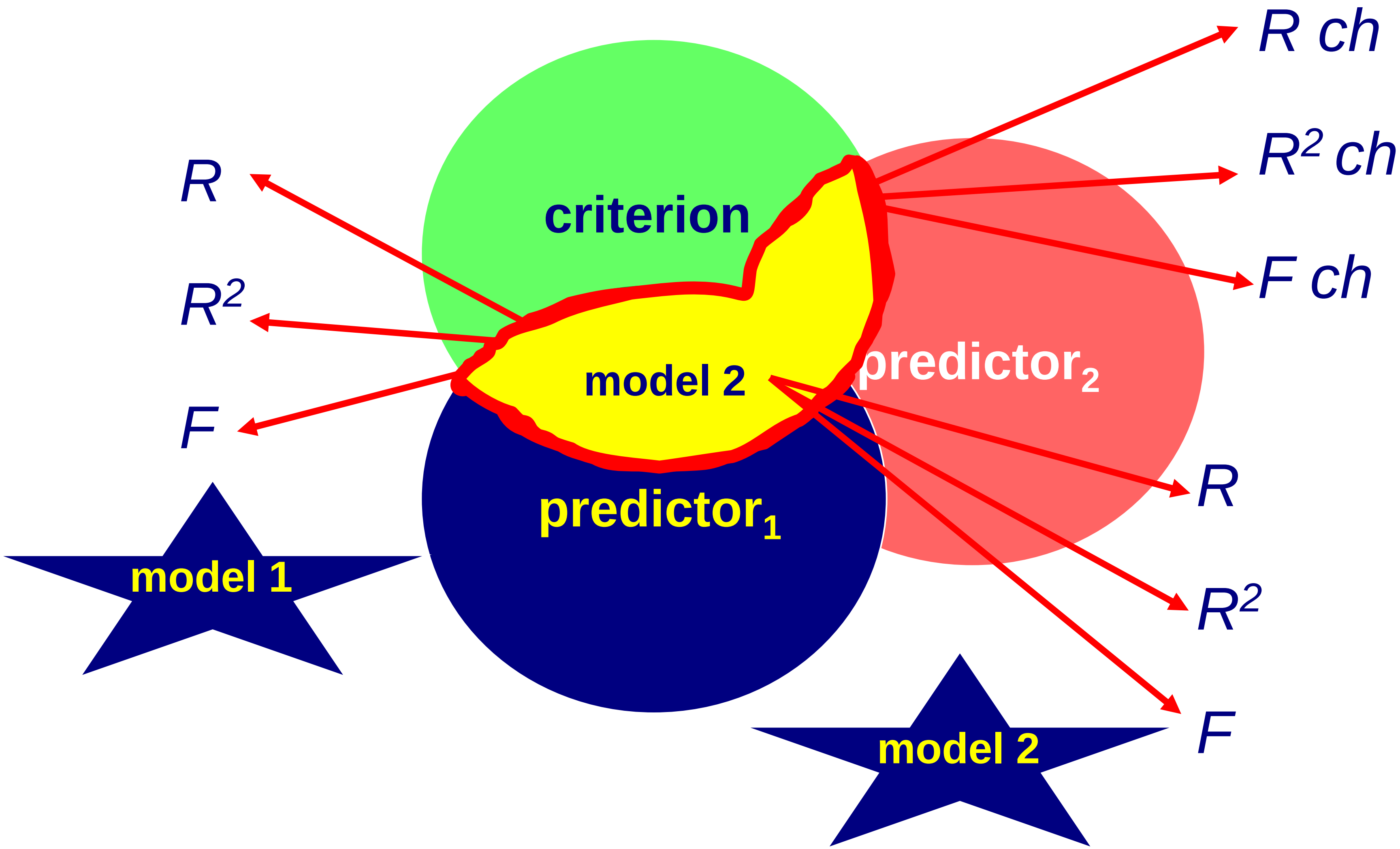
1. to partial out the effect of a control variable not of interest to the study
   - ***exactly the same idea as ancova –*** your 'covariate' in this case is the predictor entered at step 1
2. to build a sequential model according to some theory
   - e.g., broad measure of personality entered at step 1, more specific/narrow attitudinal measure entered at step 2

- order of entry is ***crucial*** to outcome and interpretation
- predictors can be entered singly or in blocks of >1
- now we will have an R, R^2, b, β, pr^2 sr^2 for EACH step to report
- also test *increment* in prediction at each block:
  - *R^2 change*
  - *F change*

hierarchical multiple regression



hierarchical multiple regression



testing hierarchical models

f = full(er) model [with more variables added]

r = reduced model

$$R^2 change = R^2 f - R^2 r$$

$$Fchange = \frac{(\mathrm{R}^2 f - \mathrm{R}^2 r)/(p_f - p_r)}{(1 - R^2 f)/N - p_f - 1)}$$

$$\mathrm{df} = p_f - p_r, \mathrm{N} - p_f - 1$$

an example:

suppose we wanted to repeat our GPA study using hierarchical regression..

- further suppose our real interest was motivation and study time, we just wanted to *control* for anxiety:
  - enter anxiety at step 1
  - enter motivation and study time at step 2
- preliminary statistics would be same as before
- model would be assessed sequentially
  - step 1 – prediction by anxiety
  - step 2 – prediction by motivation and study time ***above and beyond*** that explained by anxiety

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	14.836	2	26	.000

for model 1 – *R* and *R^2* are the same as bivariate r between GPA and Anxiety (as anxiety is the only variable in the model).

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	14.836	2	26	.000

here *R^2 ch* is just the same as *R^2* because it simply reflects the change from zero.

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	14.836	2	26	.000

for model 2 – *R* and R^2 are the same as our full standard multiple regression conducted earlier.

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	14.836	2	26	.000

R^2 ch tells us that by including study time and motivation we increase the amount of variance accounted for in GPA by 40%

(this is the critical bit!)

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	14.836	2	26	.000

alternatively, *R^2 ch* tells us that after controlling for anxiety, study time and motivation explain 40% of the variance in GPA

model summary

Model	R	R^2	R^2_{adj}	change statistics				
				R^2 ch	*F ch*	df1	df2	sig F *ch*
1	.505	.255	.228	.255	9.584	1	28	.004
2	.813	.652	.612	.397	**14.836**	2	26	.000

… and *F ch* tells us that this increment in the variance accounted is significant

(null hyp: R2 ch = 0)

anova

Summary Table for Analysis of Regression:

Model		Sums of Squares	df	Mean Square	F	sig
1	Regression	1702.901	1	1702.901	9.584	.004
	Residual	4964.567	28	177.306		
	Total	6667.46	29			
2	Regression	4346.03	3	1448.68	16.23	.000
	Residual	2321.43	26	89.29		
	Total	6667.46	29			

details for model 1 are just the same as those reported in the change statistics section on the previous page (as the change was relative to zero)

anova

Summary Table for Analysis of Regression:

Model		Sums of Squares	df	Mean Square	F	sig
1	Regression	1702.901	1	1702.901	9.584	.004
	Residual	4964.567	28	177.306		
	Total	6667.46	29			
2	Regression	4346.03	3	1448.68	16.23	.000
	Residual	2321.43	26	89.29		
	Total	6667.46	29			

details for model 2 test the overall significance of the model (and are therefore exactly the same as we would get if we had done a standard regression)

coefficients

	Model	B	SE	β	t	sig
1	constant	-80.233	7.595		7.009	.000
	ANX	5.268	1.700	.505	3.009	.004
2	constant	-95.02	3	1448.68	16.23	.000
	ANX	1.678	1.437	.16	1.168	.253
	ST	.789	.208	.42	3.785	.000
	MOT	1.453	.484	.46	3.000	.005

model 1 shows the coefficients for anxiety as the predictor of GPA (i.e., the variables included at step 1)

coefficients

Model		B	SE	β	t	sig
1	constant	-80.233	7.595		7.009	.000
	ANX	5.268	1.700	.505	3.009	.004
2	constant	-95.02	3		16.23	.000
	ANX	1.678	1.437	.16	1.168	.253
	ST	.789	.208	.42	3.785	.000
	MOT	1.453	.484	.46	3.000	.005

model 2 is identical to the coefficients table we would get in standard multiple regression if all predictors were entered simultaneously

summary of results

	step	R^2	F	R^2 ch	F ch
1	ANX	.255	9.604*	.255	9.584*
2	ST	.651	16.23*	.397	14.836*
	MOT				

some uses for hierarchical multiple regression (HMR)

- to control for nuisance variables
  - as we have done now
  - logic is same as for ancova
- to test mediation (briefly covered next week)
- to test moderated relationships (interactions)

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_1X_2 + c$$

Difference between structure of Standard and Hierarchical MR tests

Standard Multiple Regression:

1. Tests overall model R^2 automatically
2. Does not test subgroupings of variables (Blocks)
3. Tests unique effect of each IV (i.e., covariation of residual DV scores with IV once all other IVs' effects are controlled (partialled out))
4. Does not test for interactions automatically
5. Report Model R^2 with F test, plus each IVs' βs with t-tests, plus relevant follow-ups

Hierarchical Multiple Regression:

1. Tests overall model automatically
2. Tests each Block (subgrouping of variables) separately (2 sets of Fs)
3. Tests unique effect of each IV for variables in this block and earlier – but βs don't exclude overlapping variance with variables in later blocks
4. Does not test for interactions automatically – but use HMR to test manually (moderated MR next week)
5. Report each block R^2 change with F test, plus IVs' βs with t-tests from each block as entered, plus final model R^2 with F test, plus relevant follow-ups.
6. Depending on theory may or may not report betas for IVs from earlier blocks again if they change in later blocks
   - Usually not for if early block = control
   - Definitely yes if mediation test

some issues in SMR & HMR

- **multicollinearity and singularity**
  - this condition occurs when predictors are highly correlated (>.80 - 90)
  - diagnosed with high intercorrelations of IVs (collinearities) and a statistic called ***tolerance***
  - tolerance = ($\boldsymbol{1 - R^2_x}$)
  - $\boldsymbol{R^2_x}$ is the overlap between a particular predictor and all the other predictors
    - low tolerance = multicollinearity → singularity
    - high tolerance = relatively independent predictors
  - multicollinearity leads to unstable calculation of regression coefficients (***b***), even though $\boldsymbol{R^2}$ may be significant
- Some additional info about suppressor variables, handling missing data, and cross-validation is provided in the "Practice Materials" section of the web site

assumptions of multiple regression

- **distribution of residuals**
  - **normality**: conditional array of Y values are normally distributed around $\hat{Y}$ (assumption of normality in arrays)
  - **homoscedasticity**: variance of Y values are constant across different values of $\hat{Y}$ (assumption of "homogeneity of variance in arrays")
  - **linearity**: relationship between $\hat{Y}$ and errors of prediction
  - **independence** of errors
- **scales (predictor and criterion scores)**
  - **normality** (variables are normally distributed), **linearity** (there is a straight line relationship between predictors and criterion) predictors are **not singular** (extremely highly correlated)
  - measured using a **continuous** scale (interval or ratio)

In class next week:

- Moderated multiple regression
- Assignment 2

In the tutes:

- This week: Multiple regression, SPSS
- In 2 weeks: Moderated regression, SPSS

readings :

- Howell Ch 15
- Field Ch 5