# Variable Importance

# Variable Importance

- As the goal of much of social science research is explanation, determining variable importance is a key concern
- Unfortunately, the way it is typically done is crude, difficult to make sense of, and not done with much thought
- We will go over the standard fare then provide a ‘new’[1] technique that is highly interpretable and a means to test for differences among variable importance metrics

# “Controlling for”

- Let’s start with raw coefficients- What do they tell us?
- How much does Y change with a one unit change in X… *controlling for* the other variables in the model
- Now for some fun go ask random researchers what ‘controlling for’ actually means
- At a conceptual level it simply means we are considering this variable’s effects on the DV with respect to the other predictors
- What it actually means is that this is the average effect (slope) seen at each of the levels of the other variables
  - E.g. average effect of sex role identity on marital satisfaction for each value of age
- It does not mean experimental control, and it also doesn’t mean we’ve controlled for what might be potentially important variables that weren’t included in the model

# Variable importance: Statistical significance

- We can examine the output to determine which variables statistically significantly contribute to the model
- Standard error
  - measure of the variability that would be found among the different slopes estimated from other samples drawn from the same population
- This is typically not a good way to determine variable importance and should probably only be noted casually if at all
- However the standard error does allow us to get confidence intervals for the coefficients, which are desirable and should be a standard part of reporting

$$s^2_{y.12} = \frac{ss_{res}}{N-k-1}$$

$$s_{b1} = \sqrt{\frac{s^2_{y.12}}{\sum x_1^2\left(1-r_{12}^2\right)}}$$

$$t_{b1} = \frac{b_1}{s_{b1}}$$

# The standardized coefficient

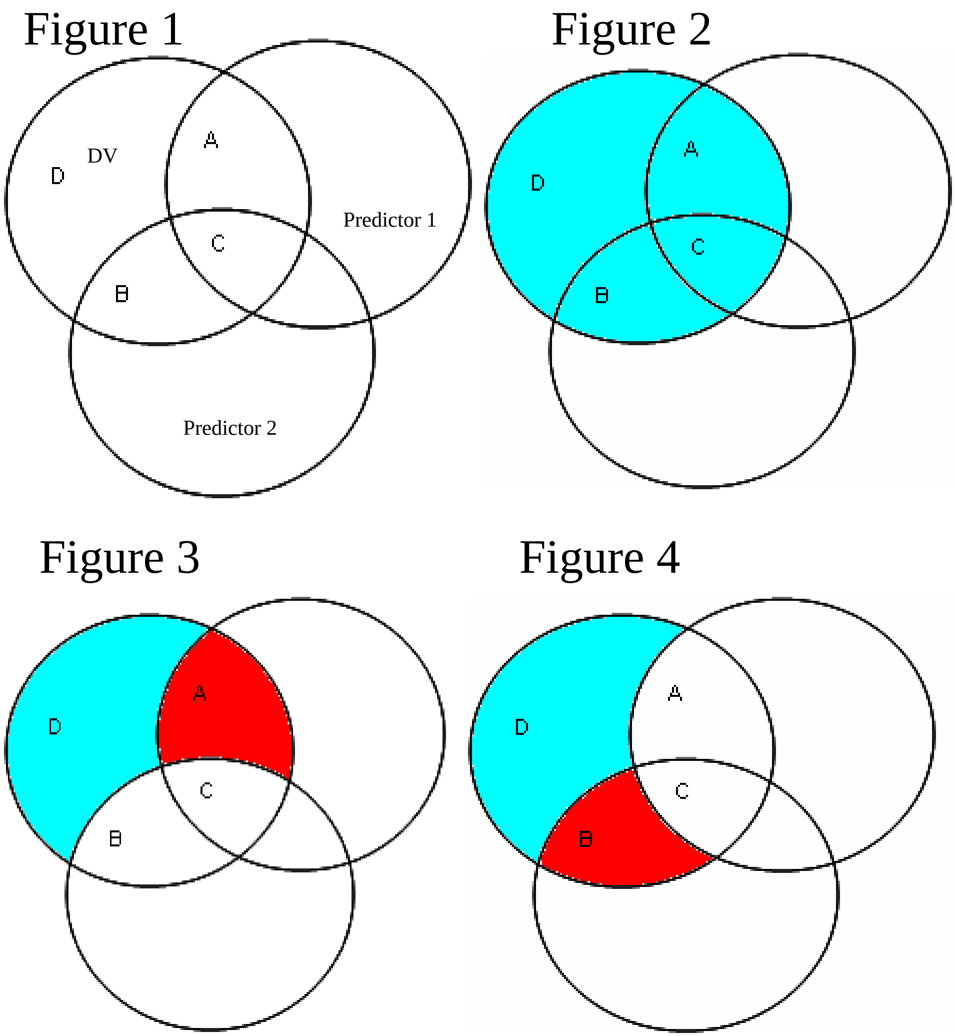
- If we standardized our variables before running the regression analysis (i.e. used the correlation matrices) we would have a standardized regression coefficient
  - How much Y would typically change given a one standard deviation change in X
  - In the simple setting it equals the Pearson r, in MR it is a type of partial correlation
- We like this because most of the measures used in psychology are on arbitrary scales (e.g. Likert)
- In this manner we can more easily compare one variable to another in absolute terms
- However, it is not a justifiable method of determining one variable is more important than another just because their coefficients differ
  - Did you really think they'd be the same going into the analysis?

# Other Standard Metrics

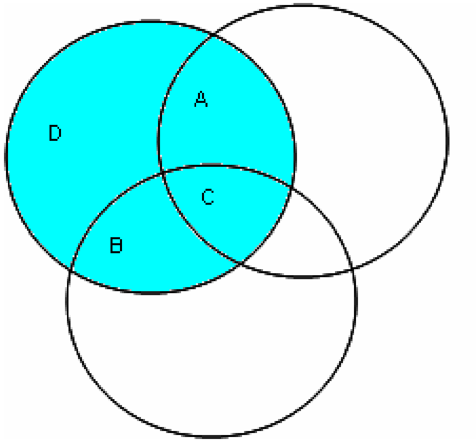
- Partial correlation
  - Predictor-DV correlation after partialling out the shared variance that both have with the other variables
  - Computationally:
    - $SS_{predictor} / (SS_{predictor} + SS_{residual})$
- Semi-partial Correlation
  - Predictor-DV correlation after partialling out the shared variance the DV has with other variables
  - When squared it represents the amount $R^2$ would increase if it was added last to the model
  - Computationally:
    - $SS_{predictor} / SS_{total}$
- Again though, the ordering one sees with these doesn't by itself tell you importance as the ordering would likely, if not dramatically change upon a new sample

# Variable importance: Partial correlation

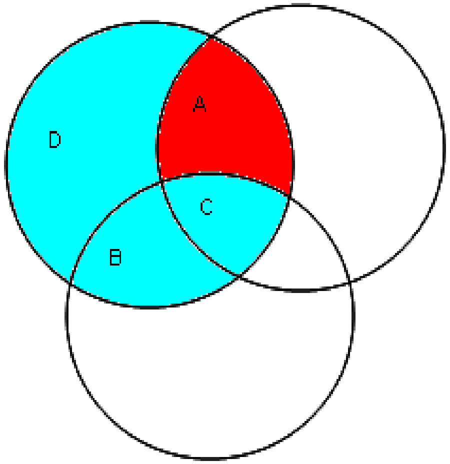
- Partial correlation is the contribution of a predictor after the contributions of the other predictors have been taken out of both that predictor and the DV
- A+B+C+D represents all the variability in the DV to be explained (fig. 1 and 2)
  - A+B+C = $R^2$ for the model
- The squared partial correlation is the amount a variable explains relative to the amount in the DV that is left to explain after the contributions of the other predictors have been removed from both the predictor and criterion
- It is A/(A+D) (fig. 3)
  - For Predictor 2 it would be B/(B+D) (fig. 4)

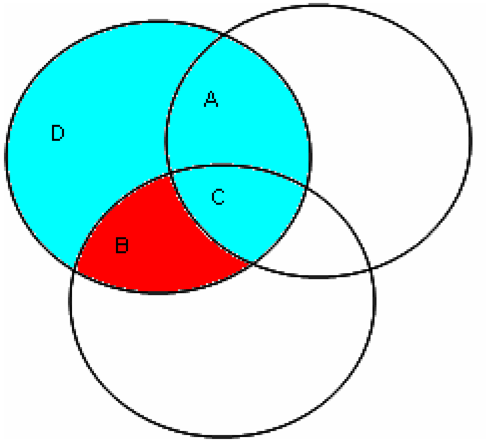


# Variable importance: Semipartial correlation




- The semipartial correlation (squared) is perhaps the more useful measure of contribution
- It refers to the unique contribution of A to the model, i.e. the relationship between the DV and predictor after the contributions of the other predictors have been removed *from the predictor in question*
- A/(A+B+C+D)
  - For predictor 2
    - B/(A+B+C+D)



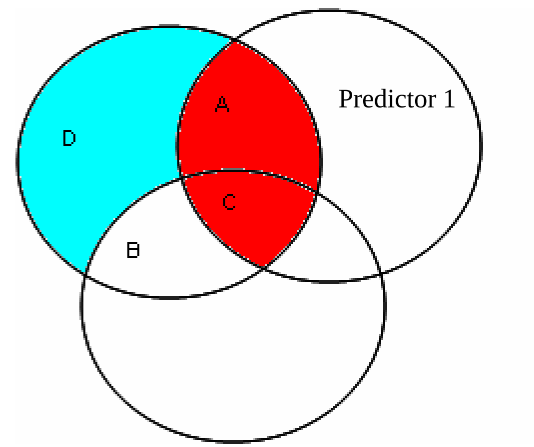



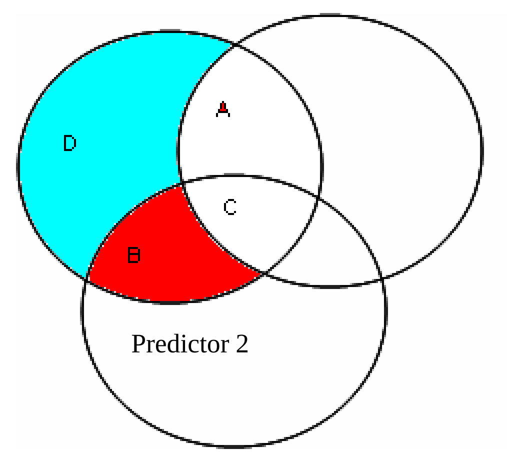

- Interpretation (of the squared value):
- Out of *all* the variance to be accounted for, how much does this variable explain that no other predictor does?
  - or
- How much would $R^2$ drop if the variable were removed?

$$R^2 = r_1^2 + sr_2^2 \ldots sr_p^2$$

# Variable importance

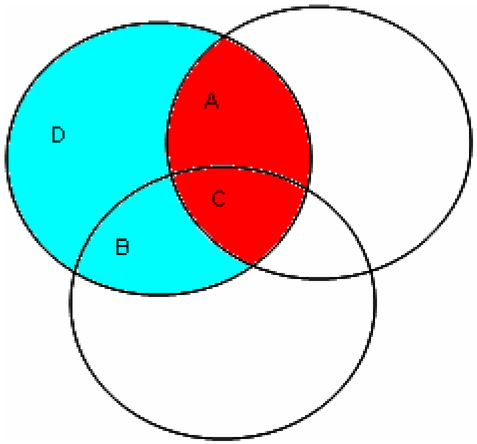
- Note that exactly how partial and semi-partial will be figured will depend on the *type* of multiple regression employed.
- The previous examples concerned a ‘standard’ multiple regression situation
- As a preview, for sequential (i.e. hierarchical) regression the partial correlation would be
  - $\text{Predictor}_1 = (A+C)/(A+C+D)$
  - $\text{Predictor}_2 = B/(B+D)$

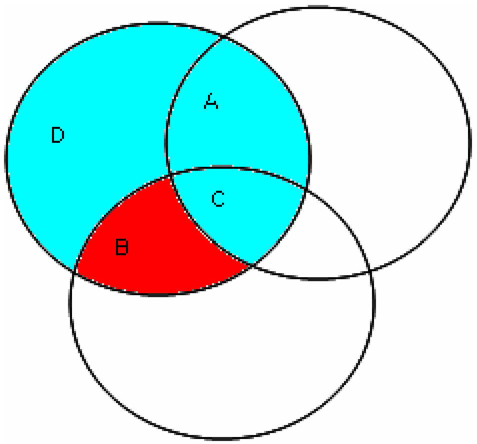


# Semipartial correlation

- For semi-partial correlation
  - $Predictor_1 = (A+C)/(A+B+C+D)$
  - $Predictor_2$ same as before
- The result for the addition of the second variable is the same as it would be in standard MR
- Thus if the goal is to see the unique contribution of a single variable after all others have been controlled for, the results of a sequential regression can be determined to an extent with the squared semipartial correlations from a regular MR
- In general terms, it is the unique contribution of the variable at the point it enters the equation (sequential or stepwise)



# Variable importance: Example data

- The semipartial correlation is labeled as ‘part’ correlation in SPSS[1]
- Here we can see that education level is really doing all the work in this model

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part |
| 1 | (Constant) | -27886.3 | 5529.479 | | -5.043 | .000 | | | |
| | Educational Level (years) | 4004.576 | 210.628 | .677 | 19.013 | .000 | .661 | .659 | .654 |
| | Months since Hire | 87.951 | 58.441 | .052 | 1.505 | .133 | .084 | .069 | .052 |
| | Previous Experience (months) | 11.936 | 5.803 | .073 | 2.057 | .040 | -.097 | .094 | .071 |

a. Dependent Variable: Current Salary

# Another example

- Mental health symptoms predicted by number of doctor visits, physical health symptoms, number of stressful life events

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .553[a] | .306 | .302 | 3.504 |

a. Predictors: (Constant), Stressful life events, Visits to health professionals, Physical health symptoms

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 2498.626 | 3 | 832.875 | 67.820 | .000[a] |
| | Residual | 5661.387 | 461 | 12.281 | | |
| | Total | 8160.013 | 464 | | | |

a. Predictors: (Constant), Stressful life events, Visits to health professionals, Physical health symptoms

b. Dependent Variable: Mental health symptoms

# Another example

- Here we see that physical health symptoms and stressful life events both significantly contribute to the model in a statistical sense
- Examining the unique contributions, physical health symptoms seem more ‘important’, but unless you run a procedure to provide a statistical test for their difference, you have *zero* evidence to suggest one is contributing significantly more than the other
  – In other words, that difference my just be due to sampling variability and unless you test that you cannot say that e.g. .381 and .223 are statistically different

**Coefficients**[a]

| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. | Correlations Zero-order | Correlations Partial | Correlations Part |
|---|---|---|---|---|---|---|---|---|---|
| 1 | (Constant) | .845 | .409 | | 2.067 | .039 | | | |
| | Visits to health professionals | -.001 | .017 | -.003 | -.075 | .940 | .256 | -.003 | -.003 |
| | Physical health symptoms | .761 | .078 | .434 | 9.818 | .000 | .505 | .416 | .381 |
| | Stressful life events | .007 | .001 | .238 | 5.754 | .000 | .370 | .259 | .223 |

a. Dependent Variable: Mental health symptoms

↑ ↑

# Variable Importance: Comparison

- Comparison of standardized coefficients, partial, and semi-partial correlation coefficients
- All of them are ‘partial’ correlations in the sense that they give an account of the relationship among a predictor and the DV after removing the effects of others in some fashion

$$\beta = \frac{r_{y1} - (r_{y2})(r_{12})}{1 - (r_{12})^2}$$

$$Partial = \frac{r_{y1} - (r_{y2})(r_{12})}{\sqrt{1 - r_{y2}^2}\sqrt{1 - (r_{12})^2}}$$

$$Semi - Partial = \frac{r_{y1} - (r_{y2})(r_{12})}{\sqrt{1 - (r_{12})^2}}$$

# Variable Importance: A more intuitive metric

- The problem with all of the previous measures of variable importance is that they don’t decompose $R^2$ into the relative contributions to it from the predictors
- There is one that provides an *average $R^2$ increase*, depending on the order a predictor enters into the model
  - 3 predictor example A **B** C; **B** A C, C A **B** etc.
- One way to think about it using what you’ve just learned is thinking of the squared semi-partial correlation whether a variable is first second third etc.
- This statistic is just an *average semi-partial*, and note that the average is for *all* possible permutations
  - E.g. the R-square contribution for B being first in the model includes B A C and B C A, both of which would of course be the same value

As Predictor 1: $R^2$ = .629
Note there are 2 models in which war would be first and have that value
(war, math, bush; war, bush, math)

# Outcome: Score on a Social Conservativism/Liberalism scale

**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 | .793[a] | .629 | .618 | 1.46685 | **.629** | 54.346 | 1 | 32 | .000 |
| 2 | .809[b] | .654 | .619 | 1.46367 | .025 | 1.070 | 2 | 30 | .356 |

a. Predictors: (Constant), war on terror

b. Predictors: (Constant), war on terror, mathmatical ability, grade for bush

As Predictor 2: $R^2$ change = .639 and .087
Here there are two models in which war would be second
(math, war, bush; bush, war, math)

**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 | .096[a] | .009 | -.022 | 2.39850 | .009 | .295 | 1 | 32 | .591 |
| 2 | .805[b] | .649 | .626 | 1.45116 | .639 | 56.417 | 1 | 31 | .000 |
| 3 | .809[c] | .654 | .619 | 1.46367 | .005 | .472 | 1 | 30 | .497 |

a. Predictors: (Constant), mathmatical ability

b. Predictors: (Constant), mathmatical ability, war on terror

c. Predictors: (Constant), mathmatical ability, war on terror, grade for bush

**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 | .742[a] | .551 | .537 | 1.61427 | .551 | 39.296 | 1 | 32 | .000 |
| 2 | .799[b] | .638 | .615 | 1.47194 | .087 | 7.488 | 1 | 31 | .010 |
| 3 | .809[c] | .654 | .619 | 1.46367 | .016 | 1.351 | 1 | 30 | .254 |

a. Predictors: (Constant), grade for bush

b. Predictors: (Constant), grade for bush, war on terror

c. Predictors: (Constant), grade for bush, war on terror, mathmatical ability

As Predictor 3: $R^2$ change = .098
There are 2 models in which war would be last and have that value.

**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 | .746[a] | .557 | .528 | 1.63025 | .557 | 19.453 | 2 | 31 | .000 |
| 2 | .809[b] | .654 | .619 | 1.46367 | .098 | 8.458 | 1 | 30 | .007 |

a. Predictors: (Constant), mathmatical ability, grade for bush

b. Predictors: (Constant), mathmatical ability, grade for bush, war on terror

# Interpretation

- The average of these* is the *average contribution to R square for a particular variable over all possible orderings*
  - In this case for war it is ~.36, i.e. on average, it increases R square 36% of variance accounted for
- Furthermore, if we add up the average R-squared contribution for all three…
  - .36+.28+.01 = .65
  - .65 is the $R^2$ for the model

# Example

- The output of interest is labeled as LMG below
  – LMG stands for Lindemann, Merenda and Gold, authors who introduced it in a textbook in the early 80s
- 'Last' is simply the squared semi-partial correlation
- 'First' is just the square of the simple bivariate correlation between predictor and DV
- 'Beta squared' is the square of the beta coefficient with all variables in
- 'Pratt' is the product of the standardized coefficient and the simple bivariate correlation
  – It too will add up to the model $R^2$ but is not recommended, one reason being that it can actually be negative

| | lmg | last | first | betasq | pratt |
|---|---|---|---|---|---|
| BUSH | 0.278 | 0.005 | 0.551 | 0.024 | 0.116 |
| MATH | 0.012 | 0.016 | 0.009 | 0.016 | 0.012 |
| WAR | 0.363 | 0.098 | 0.629 | 0.439 | 0.526 |

# Relative Importance Summary

- There are multiple ways to estimate a variable's contribution to the model, and some may be better than others
- A general approach:
- Check simple bivariate relationships
  - If you don't see worthwhile correlations with the DV there you shouldn't expect much from your results regarding the model*
    - Check for outliers and compare with robust measures also
  - You may detect that some variables are so highly correlated that one is redundant
- Statistical significance is not a useful means of assessing relative importance, nor is the raw coefficient typically
- Standardized coefficients and partial correlations are a first step
  - Compare standardized to simple correlations as a check on possible suppression
- Of typical output the semi-partial correlation is probably the more intuitive assessment
- The LMG is also intuitive, and is a natural decomposition of $R^2$, unlike the others