Linear Regression with Multiple Continuous Predictors

PSY 612

Regression Models with Multiple Predictors

- Psychological phenomena are often complex and explained by a confluence of *many* different variables
  - **Ex:** People's life satisfaction is explained by many different things, like their ability to meet their basic needs, the amount of social support they have from people around them, their feelings of success, etc.

- Regressions models that include multiple predictors simultaneously represent this complexity more than models that include only a single predictor

- Multiple regression allow researchers to test the significance of a specific predictor *when controlling for the other predictors in the model*

Example

Let's continue with our example predicting **goal achievement** from both **planfulness** and **grit.** As mentioned earlier, the researcher is particularly interested in whether planfulness predicts goal achievement **when controlling for grit.**

Goal achievement Y	**Planfulness X1**	**Grit X2**
48	5	4
21	4	2
15	3	3
6	1	1
10	2	2

Example

- Let's mean-center both of the predictor variables so the model's y-intercept will occur at a meaningful value (at the mean of each predictor).

Goal achievement Y	**Planfulness_C X1**	**Grit_C X2**
48	2	1.6
21	1	-0.4
15	0	0.6
6	-2	-1.4
10	-1	-0.4

Single Predictor Models

```r
model_planful <- lm(goal_ach ~ planfulness_c, data = data)
summary(model_planful)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     20.000      3.683   5.430   0.0123 *
planfulness_c    9.500      2.604   3.648   0.0356 *
---
```

```r
model_grit <- lm(goal_ach ~ grit_c, data = data)
summary(model_grit)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   20.000      4.423   4.521   0.0202 *
grit_c        12.500      4.338   2.882   0.0634 .
---
```

Interpreting the Parameter Estimates

- b_0 = 20, predicted goal achievement score for someone who scores equal to the mean on planfulness
- b_1 = 9.5, the predicted increase in goal achievement scores for every 1-unit increase in planfulness

- b_0 = 20, predicted goal achievement score for someone who scores equal to the mean on grit
- b_1 = 12.5, the predicted increase in goal achievement scores for every 1-unit increase in grit

Multiple Predictor Model

```r
model <- lm(goal_ach ~ planfulness_c + grit_c, data = data)
summary(model)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     20.000      4.044   4.946   0.0385 *
planfulness_c    6.500      5.154   1.261   0.3345
grit_c           5.000      7.148   0.699   0.5566
---
```

Question: Why do the slope coefficients for planfulness, b1, and grit, b2, not equal the slope coefficients for each from their single predictor models?

Interpreting the Parameter Estimates

- $b_0 = 20$ Predicted goal achievement score for someone who scores equal to the mean on planfulness *and* grit
- $b_1 = 6.5$ The predicted increase in goal achievement scores for every 1-unit increase in planfulness *controlling for grit*
- $b_2 = 5$ The predicted increase in goal achievement scores for every 1-unit increase in grit *controlling for planfulness*

How would you interpret the meaning of each of these parameter estimates?

What does "*controlling for grit*" mean?

- Remember that the coefficients in a regression model that includes multiple continuous predictors are conceptually akin to semipartial correlations
  - The relationship between Y and the part of X1 that is unrelated to X2
  - For our example, the regression coefficient for planfulness is the relationship between goal achievement and the part of planfulness that is unrelated to grit

- Q: How can we find the part of planfulness that is unrelated to grit?

Predict planfulness from grit and store the residuals:

```
model <- lm(planfulness_c ~ grit_c, data = data)

data$planfulness_residuals <- resid(model)
```

Predict goal achievement from the part of planfulness that is not related to grit:

```
model <- lm(goal_ach ~ planfulness_residuals, data = data)
summary(model)
```

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             20.000      8.066   2.479   0.0893 .
planfulness_residuals    6.500     10.283   0.632   0.5722
---
```

What does "*controlling for planfulness*" mean?

- For our example, the regression coefficient for grit is the relationship between goal achievement and the part of grit that is unrelated to planfulness

- Q: How can we find the part of grit that is unrelated to planfulness?

Predict grit from planfulness and store the residuals:

```
model <- lm(grit_c ~ planfulness_c, data = data)

data$grit_residuals <- resid(model)
```

Predict goal achievement from the part of grit that is not related to planfulness:

```
model <- lm(goal_ach ~ grit_residuals, data = data)
summary(model)
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)       20.00       8.43   2.372   0.0983 .
grit_residuals     5.00      14.90   0.336   0.7593
---
```

Model Visualization

- The full model equation was:

Goal_Ach' = 20 + 6.5*Planfulness_C + 5*Grit_C

- Remember that a model with two predictor variables is the geometric shape of a plane. We can visualize this plane using the scatter3d() function from the `car` package in R

```
scatter3d(formula = goal_ach ~ planfulness_c + grit_c, data = data)
```

Multicollinearity

- **Multicollinearity** occurs when the predictors in one's model are highly correlated with each other.

- Multicollinearity poses a problem because it:
  - Makes it difficult to detect the unique relationship between each predictor and the outcome
  - Makes it more difficult to detect significant predictors because confidence intervals around parameter estimates are wider (and standard errors are larger)

Ways of Detecting Multicollinearity

Tolerance: a measure of how much a predictor's variance is unique from the other predictors in the model

$$\text{Tolerance} = 1 - {R^2}_p$$

- where ${R^2}_p$ is the R-squared value resulting from a model in which a particular predictor, *p*, is predicted by all the other *p*-1 predictors in the model

$$\text{or VIF} = \frac{1}{Tolerance}$$

- A **low tolerance** (below 0.20) or a **high VIF** (above 5 or 10) indicates a problem with multicollinearity

The Model Comparison

Example

- Let's say the researcher is particularly interested in whether planfulness significantly predicts goal achievement *when controlling for grit*

Goal achievement Y	Planfulness_C X1	Grit_C X2
48	2	1.6
21	1	-0.4
15	0	0.6
6	-2	-1.4
10	-1	-0.4

Model Comparison

Model Comparison:

Model A: $Y_i = \beta_0 + \beta_1 Planfulness\_C_i + \beta_2 Grit\_C_i + \varepsilon_i$ PA = 3

Model C: $Y_i = \beta_0 + \beta_1 Grit\_C_i + \varepsilon_i$ PC = 2

Null & Alternative Hypotheses:

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

Estimate the parameters in Model C

Model C: $Y_i = \beta_0 + \beta_1 Grit\_C_i + \varepsilon_i$

- This is a model with a single continuous predictor, so we can use the same formulas we used last week to solve for b_1 and b_0

$$b_1 : \frac{\Sigma(x_i - M_x)(y_i - M_y)}{\Sigma(X - M_x)^2}$$

$$b_0 : M_Y - b_1 M_X$$

Estimate the parameters in Model C

X_i Grit_C	Y_i Goal Ach	$(X_i - M_X)$	$(Y_i - M_Y)$	$(X_i - M_X)*(Y_i - M_Y)$	$(X_i - M_X)^2$
1.6	48	1.6	28	1.6*28 = 44.8	2.56
-0.4	21	-0.4	1	-0.4	0.16
0.6	15	0.6	-5	-3	0.36
-1.4	6	-1.4	-14	19.6	1.96
-0.4	10	-0.4	-10	4	0.16
$M_X = 0$	$M_Y = 20$			SP = 65	$SS_X = 5.2$

$$b_1 = \frac{\Sigma(x_i - M_x)(y_i - M_y)}{\Sigma(X - M_x)^2} = \frac{SP}{SS_x}$$

$$b_1 = \frac{65}{5.2} = 12.5$$

$$b_0 = M_Y - b_1 M_X$$

$$b_0 = 20 - (12.5*0) = 20$$

Estimate of Model C:

$Y_i = 20 + 12.5*\text{Grit\_C}$

Evaluate the Fit of Model C

X_i Grit_C	Y_i Goal Ach	$Y_i' = 20 + 12.5*\text{Grit\_C}$	$(Y_i - Y'_i)$	$(Y_i - Y'_i)^2$
1.6	48	20 + (12.5*1.6) = 40	48-40 = 8	64
-0.4	21	20 + (12.5*-0.4) = 15	6	36
0.6	15	20 + (12.5*0.6) = 27.5	-12.5	156.25
-1.4	6	20 + (12.5*-1.4) = 2.5	3.5	12.25
-0.4	10	20 + (12.5*-0.4) = 15	-5	25

Estimate of Model C:

$Y_i = 20 + 12.5*\text{Grit\_C}$

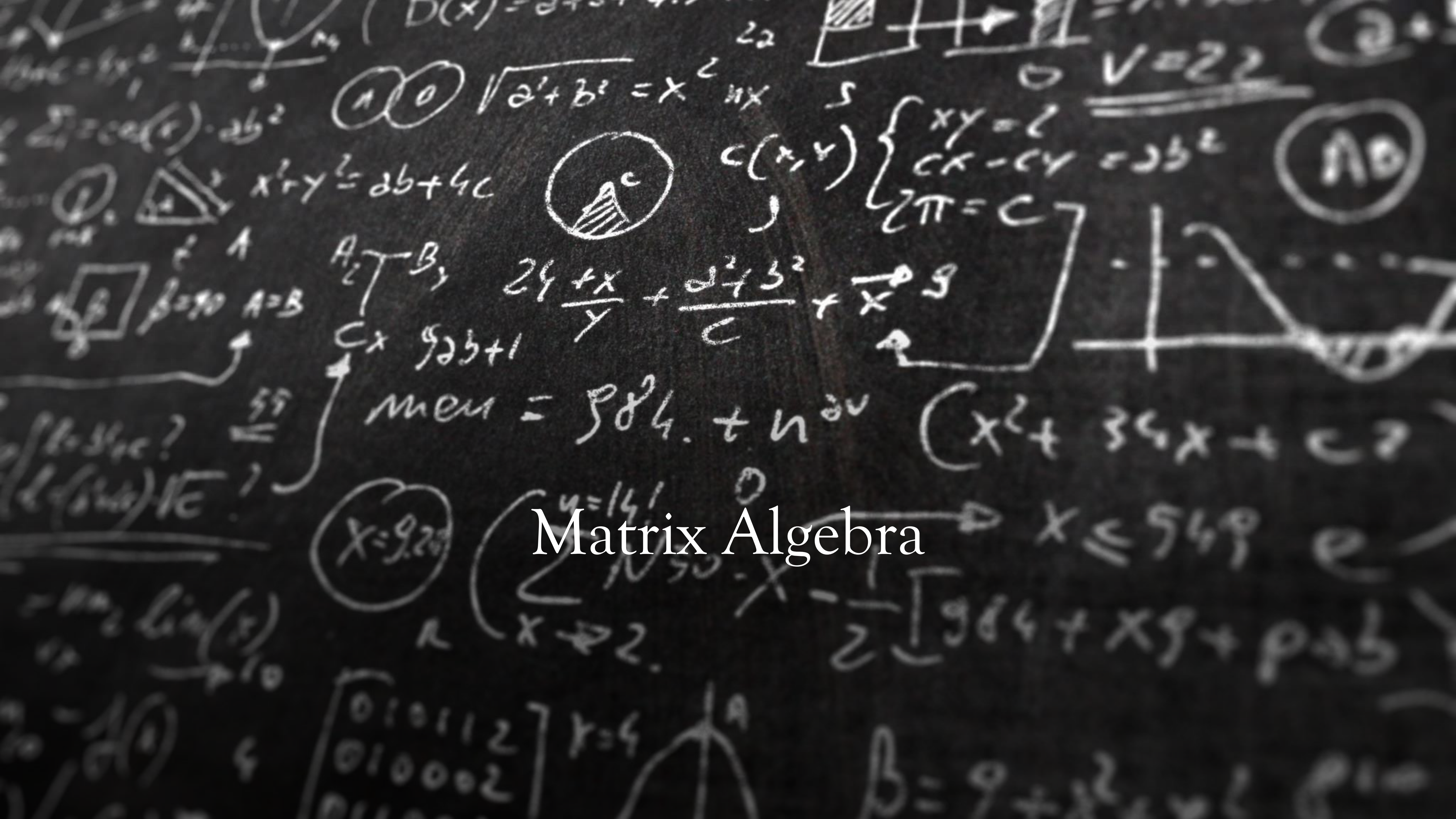
$SSE(C) = \Sigma(Y_i - Y'_i)^2$

$SSE(C) = 293.5$

Estimate the parameters in Model A

Model A: $Y_i = \beta_0 + \beta_1 Planfulness\_C_i + \beta_2 Grit\_C_i + \varepsilon_i$

- When there are multiple continuous predictors that are potentially correlated, we have to use matrix algebra to solve for the parameter estimates
  - A **matrix** is a rectangular array of numbers arranged in rows and columns



Matrix Algebra

Matrix Algebra Form

Representing the model in **matrix algebra form:**

$$Y = X\beta + \varepsilon$$

$$\underset{5\text{x}1}{\begin{bmatrix} 48 \\ 21 \\ 15 \\ 6 \\ 10 \end{bmatrix}} = \underset{5\text{x}3}{\begin{bmatrix} 1 & 2 & 1.6 \\ 1 & 1 & -0.4 \\ 1 & 0 & 0.6 \\ 1 & -2 & -1.4 \\ 1 & -1 & -0.4 \end{bmatrix}} * \underset{3\text{x}1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}} + \underset{5\text{x}1}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}}$$

- The dimensions of a matrix are given in the form [# of rows, # of columns]

Solving for β

We want to solve for the parameter estimates in matrix β. How can we solve for it in the matrix algebra equation?

$$Y = X\beta$$

$$\beta = X^{-1} * Y$$

- Unfortunately, the inverse of the X matrix, X^{-1}, is a little complicated to solve for:

$$X^{-1} = (X^T X)^{-1} X^T$$

- Thus, the parameter estimates in matrix algebra form are equal to:

$$\beta = (X^T X)^{-1} X^T * Y$$

Basic Matrix Algebra Operations

- Transpose of X (X^T)

$$\begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix} \longrightarrow \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

- Multiplying matrices

$$\underset{3x2}{\begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}} \times \underset{2x3}{\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}} = \underset{3x3}{\begin{bmatrix} (a*a + d*d) & (a*b + d*e) & (a*c + d*f) \\ (b*a + e*d) & (b*b + e*e) & (b*c + e*f) \\ (c*a + f*d) & (c*b + f*e) & (c*c + f*f) \end{bmatrix}}$$

  - The **inner dimensions** must match to multiply two matrices
  - The dimensions of the product matrix are equal to the outer dimensions of the two matrices that were multiplied

Basic Matrix Algebra Operations

- Inverse of X (X^{-1})
  - The inverse of a matrix, X, is the matrix that, when multiplied by X, produces the identity matrix

$$\begin{array}{ccc} X & X^{-1} & \text{Identity Matrix} \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix} & \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$$

- For a 2x2 matrix, the inverse is equal to:

$$X^{-1} = \frac{1}{ad - bc} \begin{bmatrix} a & -b \\ -c & d \end{bmatrix}$$

- Where "ad – bc" is called the **determinant** of matrix X

Basic Matrix Algebra Operations

For a 3x3 matrix:

$$\overset{X}{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}} \overset{X^{-1}}{\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}} = \overset{\text{Identity Matrix}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$$

- **To find X^{-1}:**
  1. Find the minor for each element in X by taking the part of the matrix remaining after excluding the row and the column containing that particular element
  2. Solve for the determinant of each 2x2 matrix
  3. Apply the following signs to each resulting determinant
  4. Divide each element by the determinant of X, the original 3x3 matrix

$$\begin{bmatrix} \begin{vmatrix} e & f \\ h & i \end{vmatrix} & \begin{vmatrix} d & f \\ g & i \end{vmatrix} & \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ \begin{vmatrix} a & c \\ g & i \end{vmatrix} & \begin{vmatrix} a & c \\ g & i \end{vmatrix} & \begin{vmatrix} a & b \\ g & h \end{vmatrix} \\ \begin{vmatrix} b & c \\ e & f \end{vmatrix} & \begin{vmatrix} a & c \\ d & f \end{vmatrix} & \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix}$$

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$det(X) = aei + bfg + cdh - afh - bdi - ceg$$

Estimate the parameters in Model A

Model A: $Y_i = \beta_0 + \beta_1 \text{Planfulness\_C}_i + \beta_2 \text{Grit\_C}_i + \varepsilon_i$

- Let's use those matrix algebra operations to solve for the parameter estimates for Model A.

$$\underset{Y}{\begin{bmatrix} 48 \\ 21 \\ 15 \\ 6 \\ 10 \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & 2 & 1.6 \\ 1 & 1 & -0.4 \\ 1 & 0 & 0.6 \\ 1 & -2 & -1.4 \\ 1 & -1 & -0.4 \end{bmatrix}} * \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}$$

- $\beta = (X^T X)^{-1} X^T * Y$

Evaluate the Fit of Model A

Estimate of Model A:

$Y_i = 20 + 6.5*Plan\_C + 5*Grit\_c$

X_1 Plan_C	X_2 Grit_C	Y Goal Ach	$Y_i' = 20 + 6.5*Plan\_C + 5*Grit\_c$	$(Y_i - Y'_i)$	$(Y_i - Y'_i)^2$
2	1.6	48	20 + (6.5*2) + (5*1.6) = 41	7	49
1	-0.4	21	20 + (6.5*1) + (5*-0.4) = 24.5	-3.5	12.25
0	0.6	15	20 + (6.5*0) + (5*0.6) = 23	-8	64
-2	-1.4	6	20 + (6.5*-2) + (5*-1.4) = 0	6	36
-1	-0.4	10	20 + (6.5*-1) + (5*-0.4) = 11.5	-1.5	2.25

$SSE(A) = \Sigma(Y_i - Y'_i)^2$

$SSE(A) = 163.5$

Comparing Fit of Model A vs Model C

SSR is how much additional error is explained by Model A compared to Model C:

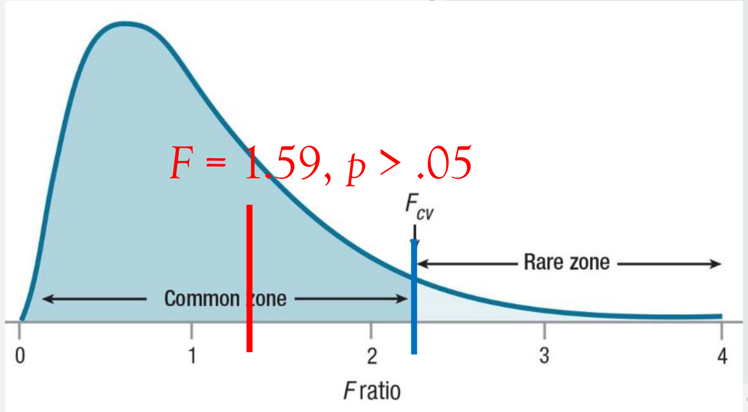
- SSR = SSE(C) – SSE(A)
- SSR = 293.5 – 163.5 = 130

PRE describes the reduction in error as a proportion:

- PRE = $\frac{SSR}{SSE(C)} = \frac{130}{293.5} = 0.44$

Model A explains 44% more of the variability in goal achievement scores compared to Model C.

Summary Table



$F_{CV}(1, 2) = 18.513$

Source	*SS*	*df*	*MS*	*F*	*p*
Planfulness_C	SSR = 130	PA-PC = 1	MSR = 130	F = 1.59	Use R to obtain
Grit_C					
Model A	SSE(A) = 163.5	n-PA = 2	MSE = 81.75		

To evaluate the significance of the *F*-statistic, we could either:

- Compare the *p*-value to an alpha of .05
- Compare the *F*-statistic to an *F*-critical value based on (PA-PC) & (n-PA) degrees of freedom

Performing the Analysis in R

Import (or set up) the data:

```
goal_ach <- c(48,21,15,6,10)
planfulness <- c(5,4,3,1,2)
grit <- c(4,2,3,1,2)

data <- cbind.data.frame(goal_ach,planfulness,grit)
```

Center the continuous predictors:

```
data$planfulness_c <- scale(data$planfulness, center = TRUE, scale = FALSE)
data$grit_c <- scale(data$grit, center = TRUE, scale = FALSE)
```

Fit the model:

```
model <- lm(goal_ach ~ planfulness_c + grit_c, data = data)
```

Summary Output

summary(model) output:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)     20.000      4.044   4.946   0.0385 *
planfulness_c    6.500      5.154   1.261   0.3345  
grit_c           5.000      7.148   0.699   0.5566  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.042 on 2 degrees of freedom
Multiple R-squared:  0.8522,	Adjusted R-squared:  0.7043 
F-statistic: 5.765 on 2 and 2 DF,  p-value: 0.1478
```

The significance of each parameter estimate is evaluated using a *t*-statistic that follows the general form:

- $t = \dfrac{parameter\ estimate}{standard\ error}$

To evaluate the significance of the *t*-statistic, we could either:

- Compare the *t*-statistic to *t*-critical values based on *df* = n-PA, or
- Compare the *p*-value to an alpha of .05

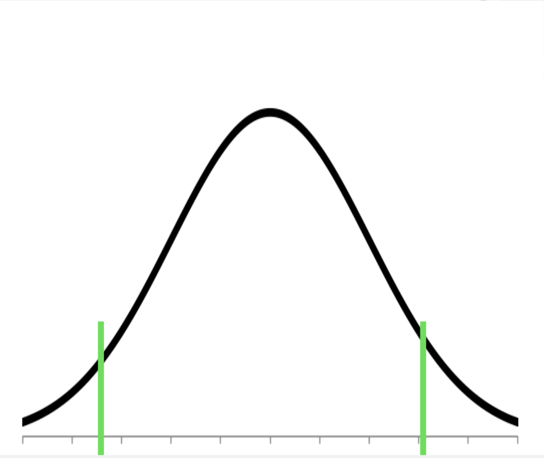
Example: Planfulness was a non-significant predictor of goal achievement, $t(2) = 1.26$, $p = .335$.

Parameter Estimates:

- $b_0 = 20$
- $b_1 = 6.5$
- $b_2 = 5$

Full model estimate:

Goal_Ach' = 20 + 6.5*Planfulness_C + 5*Grit_C



$t_{CV} = -4.303$ $t_{CV} = +4.303$

Anova output

Anova(model, type = 3) output:

```
Anova Table (Type III tests)

Response: goal_ach
             Sum Sq Df F value  Pr(>F)
(Intercept)  2000.0  1 24.4648 0.03853 *
planfulness_c 130.0  1  1.5902 0.33447
grit_c         40.0  1  0.4893 0.55665
Residuals     163.5  2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     20.000      4.044   4.946   0.0385 *
planfulness_c    6.500      5.154   1.261   0.3345
grit_c           5.000      7.148   0.699   0.5566
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.042 on 2 degrees of freedom
Multiple R-squared:  0.8522,    Adjusted R-squared:  0.7043
F-statistic: 5.765 on 2 and 2 DF,  p-value: 0.1478
```

Presents the resulting SSR and F-statistic for a Model A-Model C comparison including, and not including, (respectively) the predictor variable on each row.

- Redundant with the statistical tests in the summary( ) output (notice that each t = sqrt(F))
- Up to the researcher which form they'd like to present the significance test for a predictor variable in
- ANOVA table is helpful for checking our by-hand calculations

Example: Planfulness was a non-significant predictor of goal achievement, $F(1,2) = 1.59$, $p = .335$.

Confidence Intervals

Parameter Estimates:

- b_0 = 20 , 95%CI[2.60, 37.40]
- b_1 = 6.5 , 95%CI[-15.68, 28.68]
- b_2 = 5 , 95%CI[-25.76, 35.76]

Full model estimate:
Goal_Ach' = 20 + 6.5*Planfulness_C + 5*Grit_C

confint(model) output:

```
                    2.5 %   97.5 %
(Intercept)      2.602166 37.39783
planfulness_c -15.677973 28.67797
grit_c        -25.755315 35.75532
```

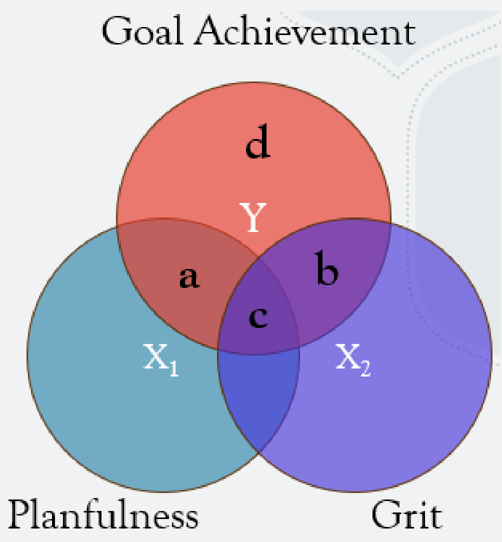
$$95\%CI = bp_{.123 \ldots p-1} \pm \sqrt{\frac{F_{CV} * MSE}{SS_x(1 - R^2{}_p)}}$$

- $b_{p.123 \ldots p-1}$ is the point estimate for predictor *p* controlling for the other *p*-1 predictors in the model
- F_{CV} is the *F*-critical value corresponding to (1, *n*-PA) degrees of freedom and α = .05
- MSE $= \frac{SSE(A)}{n - PA}$
- SS_X for the predictor variable of interest
- $1 - R^2{}_p$ = the tolerance of the predictor variable
    - Q: What happens to the 95%CI around the point estimate **as the tolerance of a predictor decreases**?

Effect Sizes

etaSquared(model, type = 3) output:

```
              eta.sq eta.sq.part
planfulness_c 0.11754069   0.4429302
grit_c        0.03616637   0.1965602
```



SS_{Total}

```
> var(data$goal_ach)*(nrow(data)-1)
[1] 1106
```

Eta-Squared: $\eta^2 = \frac{SSR}{SS_{Total}}$

For planfulness in our example:

- $\eta^2 = \frac{130}{1106} = 0.1175$
- Semi-partial correlation squared

Partial Eta-Squared: $\eta^2_{Partial} = \frac{SSR}{SSE(C)}$

For planfulness in our example:

- $\eta^2_{Partial} = \frac{130}{293.5} = 0.4429$
- Also called *PRE*
- Partial correlation squared

Standardized Regression Coefficients

Import (or set up) the data:

```
goal_ach <- c(48,21,15,6,10)
planfulness <- c(5,4,3,1,2)
grit <- c(4,2,3,1,2)

data <- cbind.data.frame(goal_ach,planfulness,grit)
```

Standardize <u>all</u> of the variables:

```
data$goal_ach_std <- scale(data$goal_ach, center = TRUE, scale = TRUE)
data$planfulness_std <- scale(data$planfulness, center = TRUE, scale = TRUE)
data$grit_std <- scale(data$grit, center = TRUE, scale = TRUE)
```

Fit the model:

```
model <- lm(goal_ach_std ~ planfulness_std + grit_std, data = data)
```

Standardized Regression Coefficients

summary(model) output:

```
Coefficients:
                              Estimate               Std. Error t value Pr(>|t|)
(Intercept)     -0.000000000000000003882  0.243170736794691366711   0.000    1.000
planfulness_std  0.618067337167283237243  0.490126289220616961906   1.261    0.334
grit_std         0.342842073207096698084  0.490126289220617128439   0.699    0.557

Residual standard error: 0.5437 on 2 degrees of freedom
Multiple R-squared:  0.8522,    Adjusted R-squared:  0.7043
F-statistic: 5.765 on 2 and 2 DF,  p-value: 0.1478
```

Goal_Ach_Std' = 0 + 0.62*Planfulness_Std + 0.34*Grit_Std

- b_0 = predicted goal achievement for someone who scores at the mean on planfulness and grit
- b_1 = the model predicts that goal achievement will increase by 0.62 standard deviations for every 1 standard deviation increase in planfulness
- b_2 = the model predicts that goal achievement will increase by 0.34 standard deviations for every 1 standard deviation increase in grit

Question: Can we say that planfulness is a stronger predictor of goal achievement than grit?

Unstandardized vs Standardized Regression Coefficients

- pg. 130-131

Relative Importance of Predictor Variables

Many researchers are tempted to make inferences about the relative importance of predictor variables in multiple regression models. It should be obvious that it is not reasonable to compare directly the relative magnitudes of b_j because they clearly depend on the unit of measurement. By making the unit of measurement for a given predictor variable smaller (e.g., using centimeters instead of inches to measure height), we can make b_j arbitrarily large without changing the fundamental impact of that variable in the model.

Unfortunately, the standardized regression coefficients do not solve the problem of making inferences about relative importance. First, they still depend on the particular range and distribution of the predictor in the particular dataset. If this distribution were to change for any reason—for example, if the researcher were to include a wider range of cases—then the standardized regression coefficients could change dramatically even if the underlying relationship between the predictor and the data remained the same (i.e., the partial regression coefficient is unchanged). Such dependence on the vagaries of a particular set of observations hardly provides a good basis for inferring relative importance. Second, relative importance is also clouded by redundancy. A regression coefficient (standardized or not) might be low not because its associated predictor variable has no relationship with Y, but because that predictor variable is at least partially redundant with the other predictors. Thus, standardized regression coefficients do not really allow us, as they were intended, to make relative comparisons of predictors. The fundamental problem is that we cannot frame a question about relative importance in terms of a Model A versus Model C comparison because both models would necessarily have the same predictors. Hence, our machinery for statistical inference cannot address the question of relative importance.

In sum, relative importance of predictor variables is a slippery concept. Any statement about relative importance must be accompanied by many caveats that it applies to this particular range of variables in this particular situation. As a result they are, in our opinion, seldom useful. Thus, although it is tempting to compare the importance of predictors in multiple regression, it is almost always best not to do so.