Multiple Regression

Last time

- Semi-partial and partial correlations

Today

- Introduction to multiple regression

Regression equation

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

- regression coefficients are "partial" regression coefficients
  - predicted change in Y for a 1 unit change in X, *holding all other predictors constant*
  - similar to semi-partial correlation -- represents part of each X

Interpretting multiple regression model

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

- Intercept is the value of Y when all predictors = 0
- Regression coefficients are the predicted change in Y for a 1 unit change in X, *holding all other predictors constant*

Interpretting multiple regression model

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

- Residual in simple regression can be thought of as a measure of Y that is left over after accounting for your DV

- Partial correlation can be created by:

  1. create a measure of X_1 that is independent of X_2
  2. create a measure of Y that is independent of X_2
  3. correlate the new measures

Example

```r
library(here)
stress.data = read.csv(here("data/stress.csv"))
library(psych)
describe(stress.data$Stress)
```

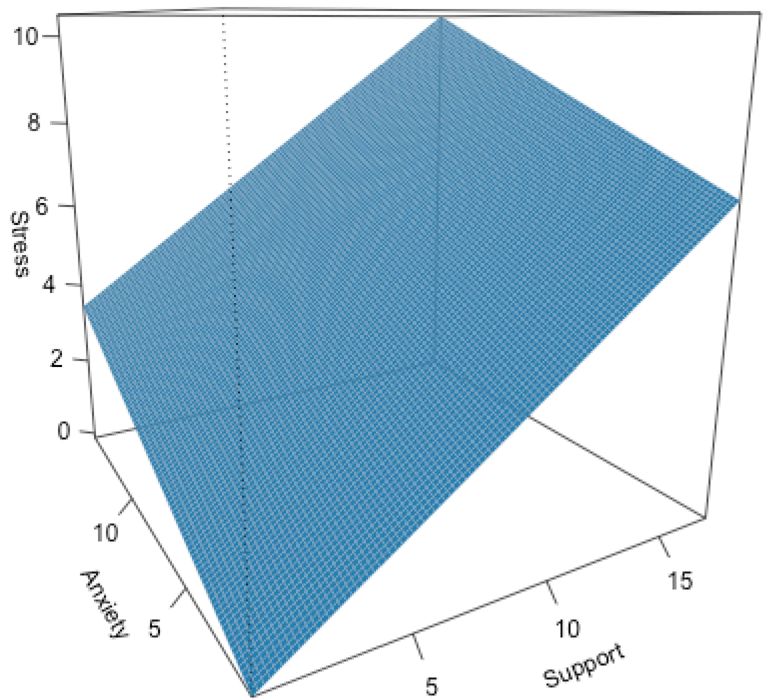
```
##    vars   n mean   sd median trimmed  mad  min   max range skew kurtos
## X1    1 118 5.18 1.88   5.27    5.17 1.65 0.62 10.32  9.71 0.08     0.
```

Example

```r
mr.model <- lm(Stress ~ Support + Anxiety, data = stress.data
summary(mr.model)
```

```
...
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31587    0.85596  -0.369 0.712792
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,    Adjusted R-squared:  0.3444
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
...
```

Visualizing multiple regression



Calculating coefficients

Just like with univariate regression, we calculate the OLS solution. As a reminder, this calculation will yield the estimate that reduces the sum of the squared deviations from the line:

Unstandardized

$$\hat{Y} = b_0 + b_1 X1 + b_2 X_2$$

$$\text{minimize} \sum (Y - \hat{Y})^2$$

Standardized

$$\hat{Z}_Y = b_1^* Z_{X1} + b_2^* Z_{X2}$$

$$\text{minimize} \sum (z_Y - \hat{z}_Y)^2$$

Calculating the standardized partial regression coefficient

$$b_1^* = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}$$

$$b_2^* = \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2}$$

Notice the similarity with semi-partial correlation

$$b_1^* = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}$$

$$sr = r_{y(1.2)} = \frac{r_{Y1} - r_{Y2}r_{Y12}}{\sqrt{1 - r_{12}^2}}$$

Relationships between partial, semi- and b*

All ways to represent the relationship between two variables while taking into account a third (or more!) variables.

- Each is a standardized effect, bounded by -1 and 1*. This means they can be compared.

Not equal calculations!

- If predictors are not correlated, r, sr $(r_{Y(1.2)})$ and b* are equal

*Standardized regression coefficients are not bounded by -1 and 1, but it's rare and usually a problem

Standardized mult regression coefficient b^*

$$\frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}$$

Semi-partial correlation $r_{y(1.2)}$

$$\frac{r_{Y1} - r_{Y2}r_{Y12}}{\sqrt{1 - r_{12}^2}}$$

Partial correlation $r_{y1.2}$

$$\frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1 - r_{Y2}^2}\sqrt{1 - r_{12}^2}}$$

```r
mod0 = lm(z_stress ~ z_anxiety + z_support,
          data = stress.data)

round(coef(mod0),3)
```

```
## (Intercept)   z_anxiety   z_support
##       0.000       0.339       0.710
```

```r
spcor.test(x = stress.data$A
           y = stress.data$S
           z = stress.data$S
```

```
## [1] 0.2797712
```

```r
pcor.test(x = stress.data$An
          y = stress.data$St
          z = stress.data$Su
```

```
## [1] 0.3339479
```

They're not the same, but they're close!

Review

Original Metric

$$b_1 = b_1^* \frac{s_Y}{s_{X1}}$$

$$b_1^* = b_1 \frac{s_{X1}}{s_Y}$$

Intercept

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

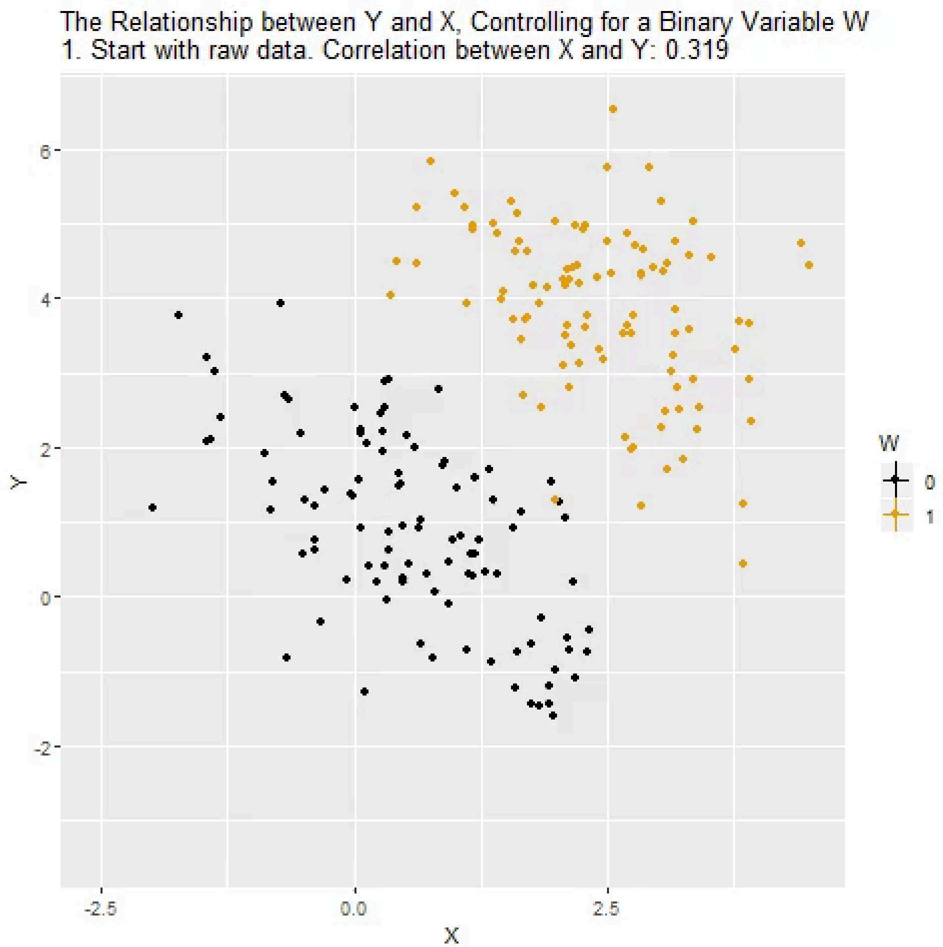
```
mr.model <- lm(Stress ~ Support + Anxiety, data = stress.data
summary(mr.model)
```

```
## 
## Call:
## lm(formula = Stress ~ Support + Anxiety, data = stress.data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.1958 -0.8994 -0.1370  0.9990  3.6995 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.31587    0.85596  -0.369 0.712792    
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,	Adjusted R-squared:  0.3444 
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
```

```r
mr.model <- lm(Stress ~ Support + Anxiety, data = stress.data
summary(mr.model)
```

```
##
## Call:
## lm(formula = Stress ~ Support + Anxiety, data = stress.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1958 -0.8994 -0.1370  0.9990  3.6995
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31587    0.85596  -0.369 0.712792
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,    Adjusted R-squared:  0.3444
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
```

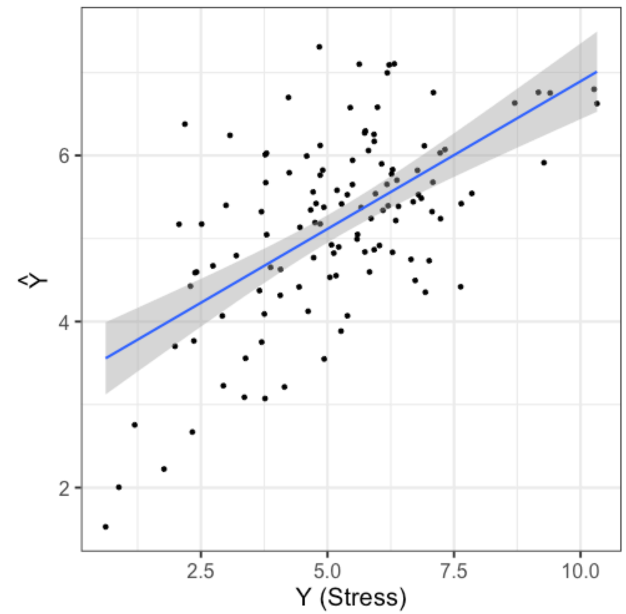
"Controlling for"



Estimating model fit

```
##
## Call:
## lm(formula = Stress ~ Support + Anxiety, data = stress.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1958 -0.8994 -0.1370  0.9990  3.6995
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31587    0.85596  -0.369 0.712792
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,    Adjusted R-squared:  0.3444
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
```

```
library(broom)
stress.data1 = augment(mr.mod
stress.data1 %>%
  ggplot(aes(x = Stress, y =
```



Multiple correlation, R

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

- $\hat{Y}$ is a linear combination of Xs
- $r_{Y\hat{Y}}$ = multiple correlation = R

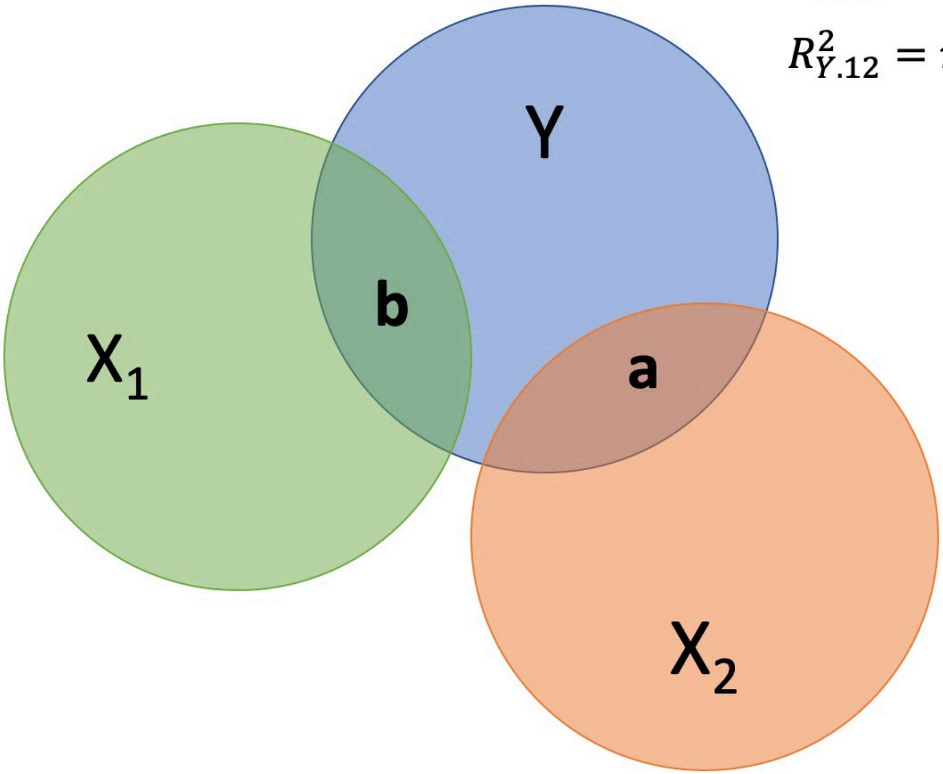
$$R = \sqrt{b_1^* r_{Y1} + b_2^* r_{Y2}}$$

$$R^2 = b_1^* r_{Y1} + b_2^* r_{Y2}$$

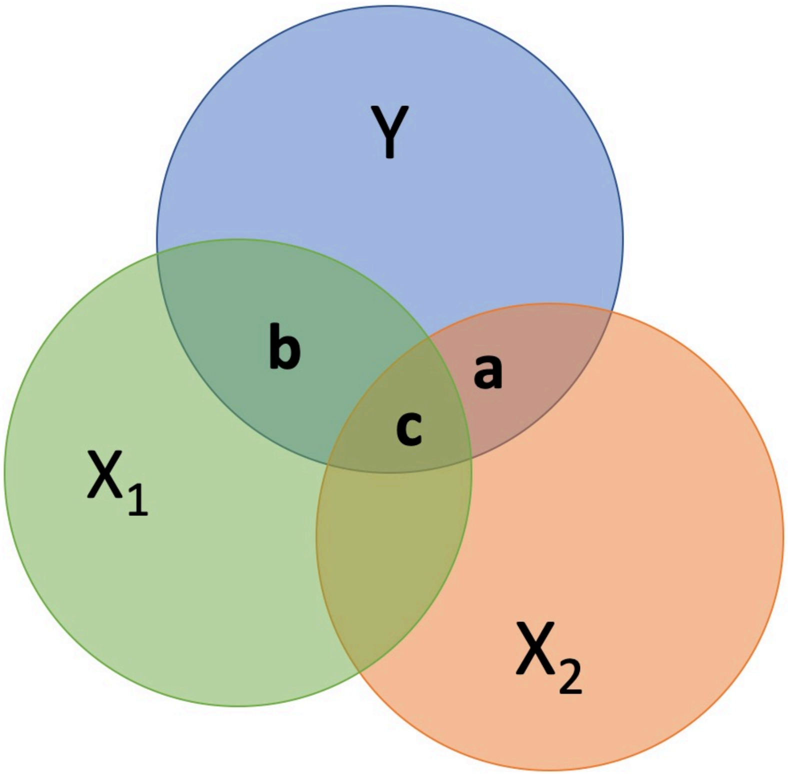
$$R^2_{Y.12} = b^*_{Y1.2} r_{Y1} + b^*_{Y2.1} r_{Y2}$$

$$R^2_{Y.12} = r_{Y1} r_{Y1} + r_{Y2} r_{Y2}$$

$$R^2_{Y.12} = r^2_{Y1} + r^2_{Y2}$$



$$R^2_{Y.12} = b^*_{Y1.2} r_{Y1} + b^*_{Y2.1} r_{Y2}$$



Decomposing sums of squares

We haven't changed our method of decomposing variance from the univariate model

$$\frac{SS_{regression}}{SS_Y} = R^2$$

$$SS_{regression} = R^2(SS_Y)$$

$$SS_{residual} = (1 - R^2)SS_Y$$

Significance tests

- R^2 (omnibus)
- Regression Coefficients
- Increments to R^2

R-squared, R^2

- Same interpretation as before
- Adding predictors into your model will increase R^2 – regardless of whether or not the predictor is significantly correlated with Y.
- Adjusted/Shrunken R^2 takes into account the number of predictors in your model

Adjusted R-squared, $\mathrm{Adj}R^2$

$$R_A^2 = 1 - \frac{Var_{res}}{Var_{total}}$$

$$R_A^2 = 1 - \frac{\frac{SS_{res}}{n-p-1}}{\frac{SS_{total}}{n-1}}$$

$$R_A^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Adjusted R-squared, $\mathrm{Adj}R^2$

$$R_A^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

- What happens if you add many IV's to your model that are uncorrelated with your DV?

- What happens as you add more covariates to your model that are highly correlated with your key predictor, X?

$$b_1^* = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}$$

ANOVA

```
summary(mr.model)
```

```
##
## Call:
## lm(formula = Stress ~ Support + Anxiety, data = stress.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1958 -0.8994 -0.1370  0.9990  3.6995
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31587    0.85596  -0.369 0.712792
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,    Adjusted R-squared:  0.3444
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
```

ANOVA

```
anova(mr.model)
```

```
## Analysis of Variance Table
##
## Response: Stress
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Support     1 113.151 113.151  49.028 1.807e-10 ***
## Anxiety     1  33.314  33.314  14.435 0.0002336 ***
## Residuals 115 265.407   2.308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(mr.model)
```

```
##
## Call:
## lm(formula = Stress ~ Support + Anxiety, data = stress.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1958 -0.8994 -0.1370  0.9990  3.6995
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31587    0.85596  -0.369 0.712792
## Support      0.40618    0.05115   7.941 1.49e-12 ***
## Anxiety      0.25609    0.06740   3.799 0.000234 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 115 degrees of freedom
## Multiple R-squared:  0.3556,    Adjusted R-squared:  0.3444
## F-statistic: 31.73 on 2 and 115 DF,  p-value: 1.062e-11
```

Test of individual regression coefficients

$$H_0 : \beta_X = 0$$
$$H_1 : \beta_X \neq 0$$

Test of individual regression coefficients

In the case of univariate regression:

$$se_b = \frac{s_Y}{s_X}\sqrt{\frac{1 - r^2_{xy}}{n - 2}}$$

In the case of multiple regression:

$$se_b = \frac{s_Y}{s_X}\sqrt{\frac{1 - R^2_{Y\hat{Y}}}{n - p - 1}}\sqrt{\frac{1}{1 - R^2_{i.jkl...p}}}$$

- As N increases...
- As variance explained increases...

Next time

More multiple regression

Can you...

- write out standardized and unstandardized regression equations?
- interpret the coefficients of a multiple regression?
- draw comparisons from ANOVA and regression?
- calculate R^2?