Housekeeping

- AP2 now due a week from Friday

Data and Files

- Files for Today:
  - albumsales2.dat
  - MLR.R
- Packages for Today:
  - lmtest
  - car
  - QuantPsyc

Data Sets

Appendix: R Examples for Common Data Manipulation Needs

newData = data.frame(oldData [, c(“V1", “V2" , “V3")]) ##Create new datafile with only variables of interest

newData = na.omit(oldData) ##Omit missing data from the data file

data$V1reverse = 6 - data$V1 ##Reverse code a question — number represents 1 more than the scale of the original question (i.e., this example is for a 5-point likert scale item)

data$V1 = (df$Q1 + df$Q2 + df$Q3)/3 ##Combine questions together as a factor

library(sjlabelled)
data$V1 <- as_label(df$V1)##Changes SPSS data that is designated as labelled to a factor with the given labels

library(dplyr)
newData = sample_n(oldData, 1000)##Use to randomly select 1000 cases from large data sets

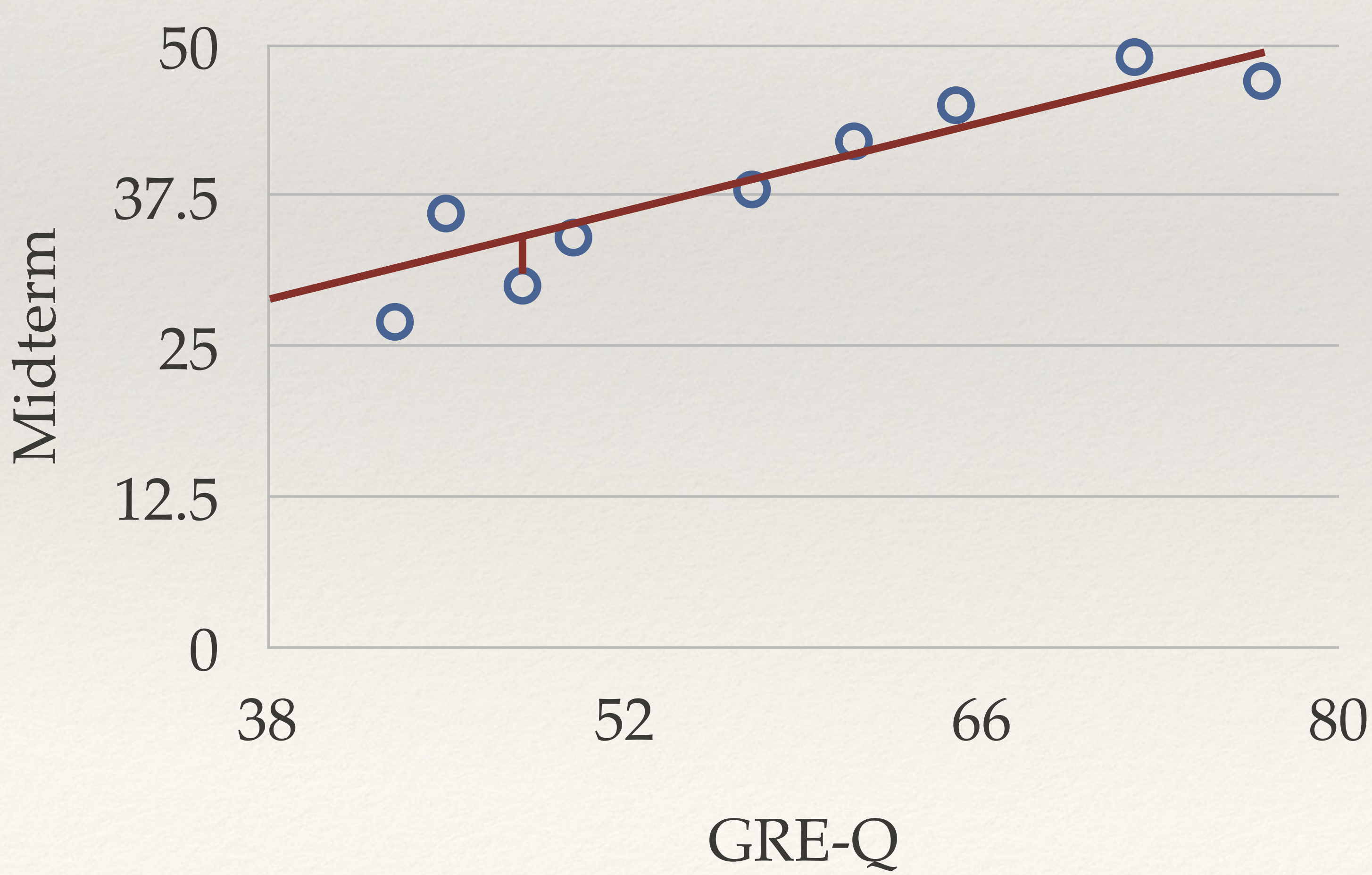
##For working with SPSS files (.sav files)
#import using the “from SPSS” option in R Studio

#run below code for any variable you want to be a factor with labels
library(sjlabelled)
data$V1 = as_factor(data$V1)

Prediction & Regression

$$Y_i' = a_{YX} + b_{YX}(X_i)$$

$$Y_i = a_{YX} + b_{YX}(X_i) + e_i$$



R Studio: The Linear Model Function

Assumption Tests

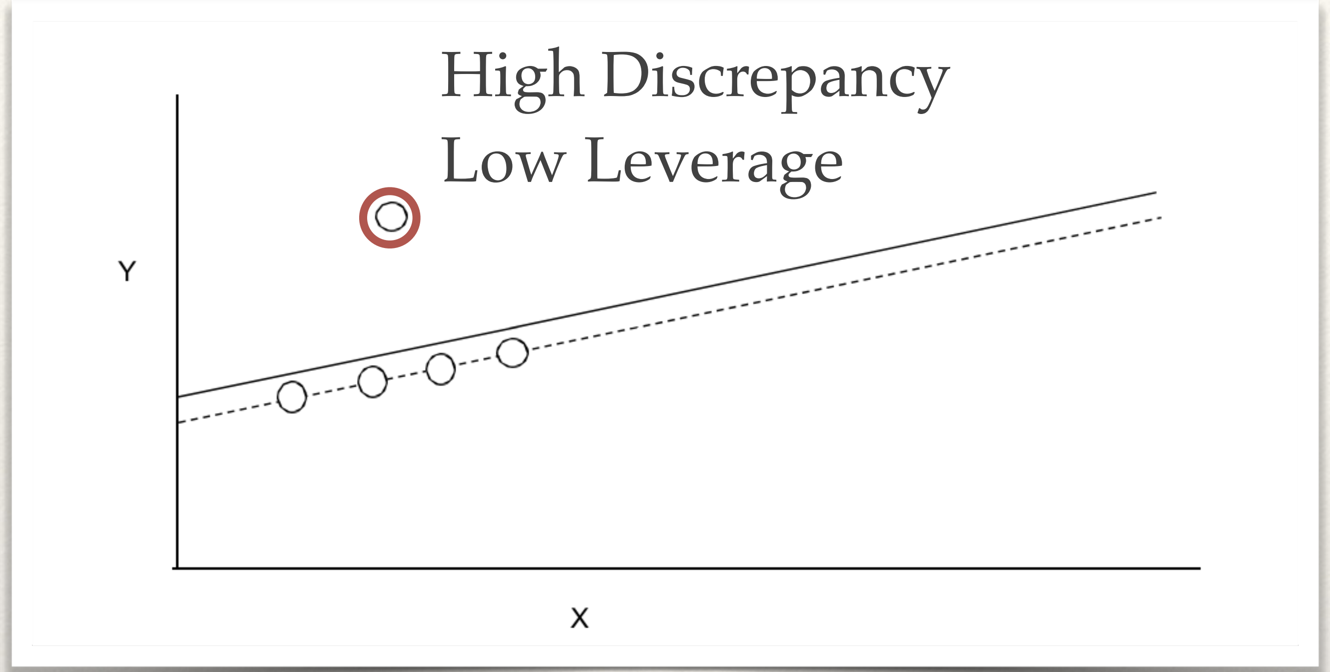
Generalizations

Fit

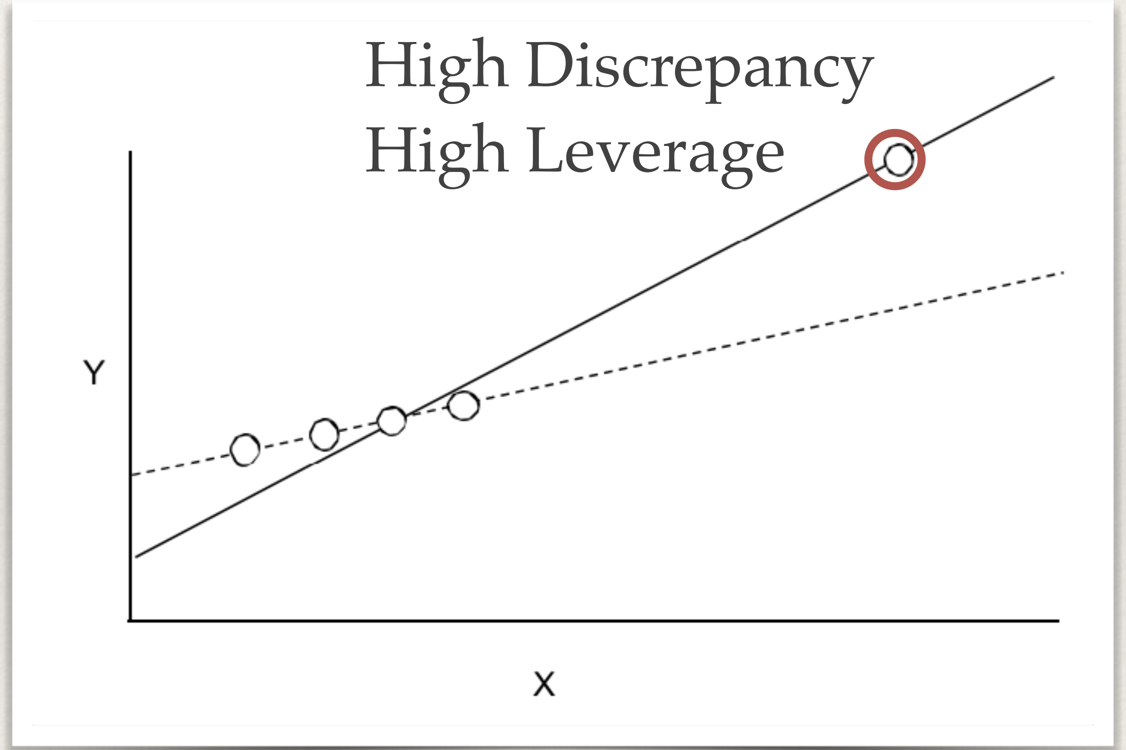
Diagnostics

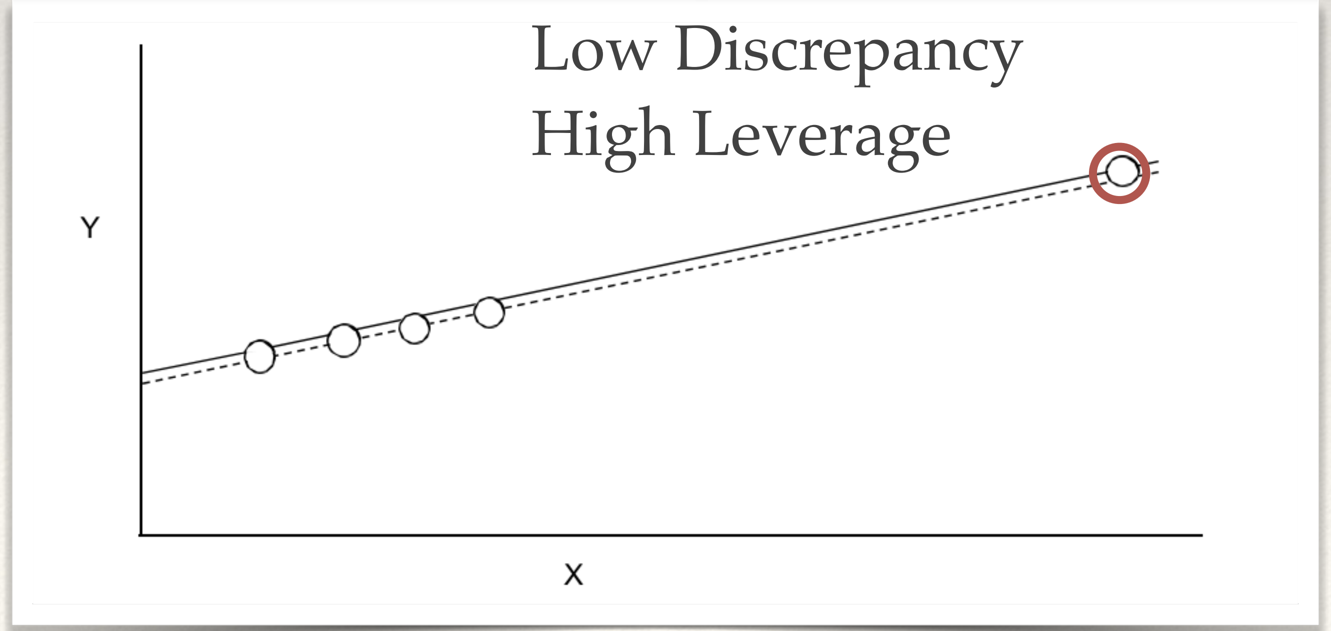
```
> summary(mod1)

Call:
lm(formula = sales ~ adverts, data = albumsales)

Residuals:
     Min       1Q   Median       3Q      Max
-152.949  -43.796   -0.393   37.040  211.866

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.341e+02  7.537e+00  17.799   <2e-16 ***
adverts     9.612e-02  9.632e-03   9.979   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 65.99 on 198 degrees of freedom
Multiple R-squared:  0.3346,	Adjusted R-squared:  0.3313
F-statistic: 99.59 on 1 and 198 DF,  p-value: < 2.2e-16
```

Outliers and Influential Cases



Diagnostics





R Studio:Outliers and Influential Cases

Investigate Thresholds

Use the which() command to see which cases, if any, are exceeding the give threshold

```
> ##Assess Outliers
> which(albumsales$std.residuals > 3 | albumsales$std.residuals < -3)
[1] 169
> which(albumsales$std.residuals > 3.29 | albumsales$std.residuals < -3.29)
integer(0)
> ##Assess Influential Cases
> which(albumsales$leverage > 3*mean(albumsales$leverage))
[1]  11  23  43  87  88 184
> which(albumsales$Cooks.distance > 4/(length(albumsales$adverts)-2))
[1]   1  10  42  55 125 169
> which(albumsales$Cooks.distance > 1)
integer(0)
```

Assumptions for Generalizations

Assumption Tests	
Normality of Error Term	Normality Tests on Residuals
Homoskedasticity	Breusch-Pagan
Independent Errors	Durbin-Watson

R Studio: Assumptions for Generalization

shapiro.test()	Command used to assess normality, in this case **normality of the error term**.
bptest()	Command for the Breusch-Pagan test used to assess **homoskedasticity**. Part of the "lmtest" package.
dwt()	Command for the Durbin-Watson test used to assess **Independence of the Error Term**. Part of the "car" package.

Options for Violations

- If there is a glaring error, fix it
- Several similar outliers may indicate a subpopulation
- Consider excluding the outlying/influential points
    - Parallel analyses, one with outliers, the other without
- Winsorizing - Replace outlying value with an arbitrary value just slightly larger than the largest non-outlying observation
- Use robust regression techniques rather than ordinary least squares estimation

APA Write Up

A simple linear regression model with Adverts predicting sales was significant, $F(1,198) = 99.59$, $p < .001$, and represented a large effect ($R^2 = .33$) as it accounted for 33% of the variance in Sales. Diagnostics for possible influential cases in this model did provide mixed results, with Cook's distances exceeding conservative thresholds, but no distance exceeding the more liberal threshold of 1. Assumption testing noted no concerns with normality and independence of the error terms, but the Breusch-Pagan test ($p < .05$) did note concerns with homoskedasticity. As such, significant findings, and generalizations from them, are likely not appropriate.

ESCP 8850

Multiple Linear Regression

February 28, 2022

Multiple Predictors

Grad GPA (Y)	GRE-Total (X_1)	Undergrad GPA (X_2)
4.0	145	3.2
3.9	120	3.7
3.8	125	3.6
3.7	130	2.9
3.6	110	3.5
3.5	100	3.3
3.4	95	3.0
3.3	115	2.7
3.2	105	3.1
3.1	90	2.8
3.0	105	2.4

$r =$	0.78
$r^2 =$	0.62

Prediction Equation

$$Y_i' = a + b_1(X_{1i}) + b_2(X_{2i}) \ldots b_m\left(X_{mi}\right)$$

Y′ = Y prime or the predicted Y value

a = the value of Y when all predictors are 0

b_m = change in Y for every one-unit change in X_m, with other X's held constant

X_m = Given score on variable X_m

Regression Equation

$Outcome_i = (model) + Error_i$

$$Y_i = a + b_1(X_{1i}) + b_2(X_{2i}) \ldots b_m\left(X_{mi}\right) + e_i$$

Y = Y value

a = the value of Y when all predictors are 0

b_m = change in Y for every one-unit change in X_m, with other X's held constant

X_m = Given score on variable X_m

e = error

Regression Equation

$$\text{GGPA}_i = a + b_1(\text{GRE}_i) + b_2(\text{UGPA}_i) + e_i$$

GGPA = Graduate GPA

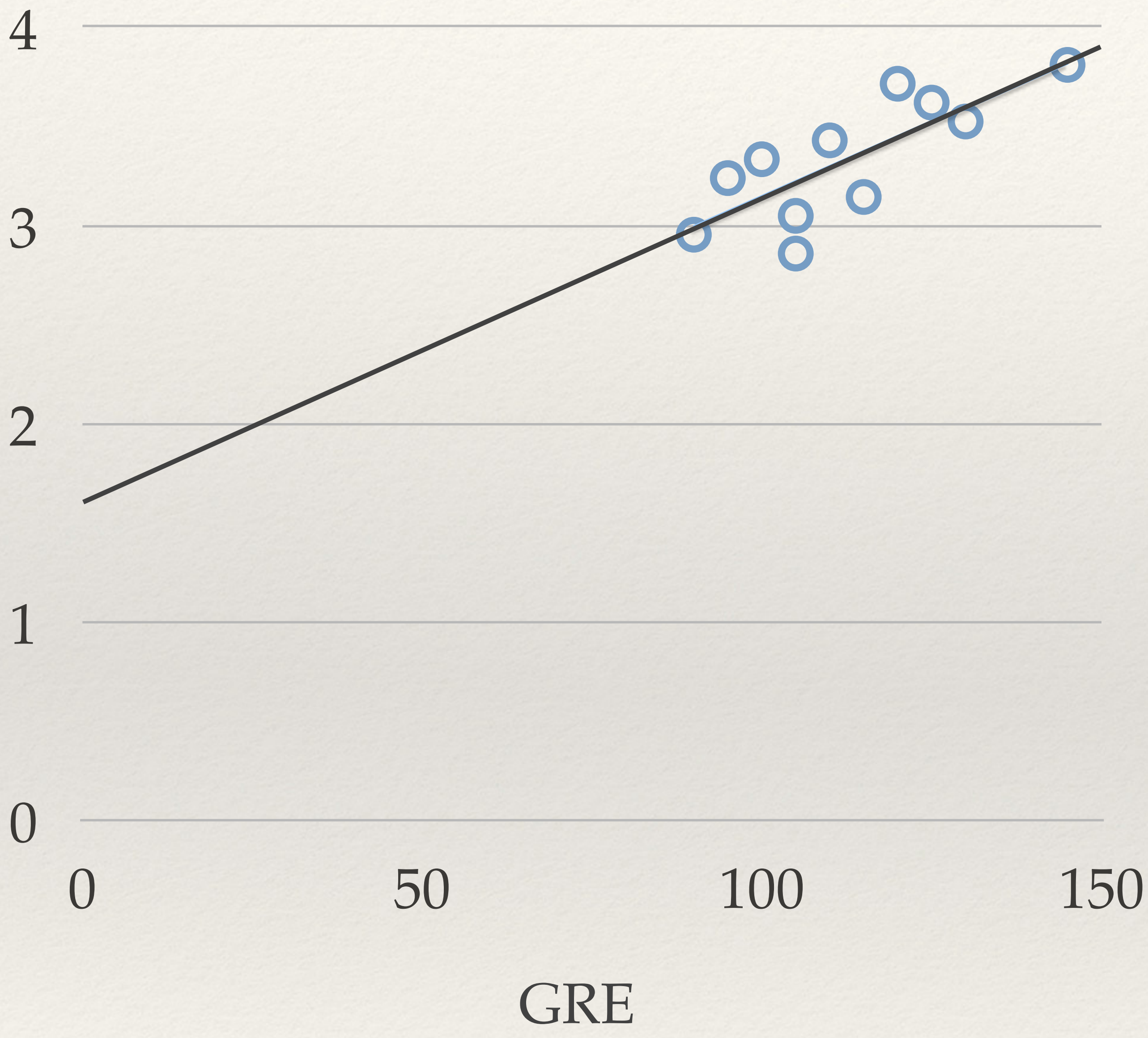
a = the value of Graduate GPA when GRE and UGPA are both 0

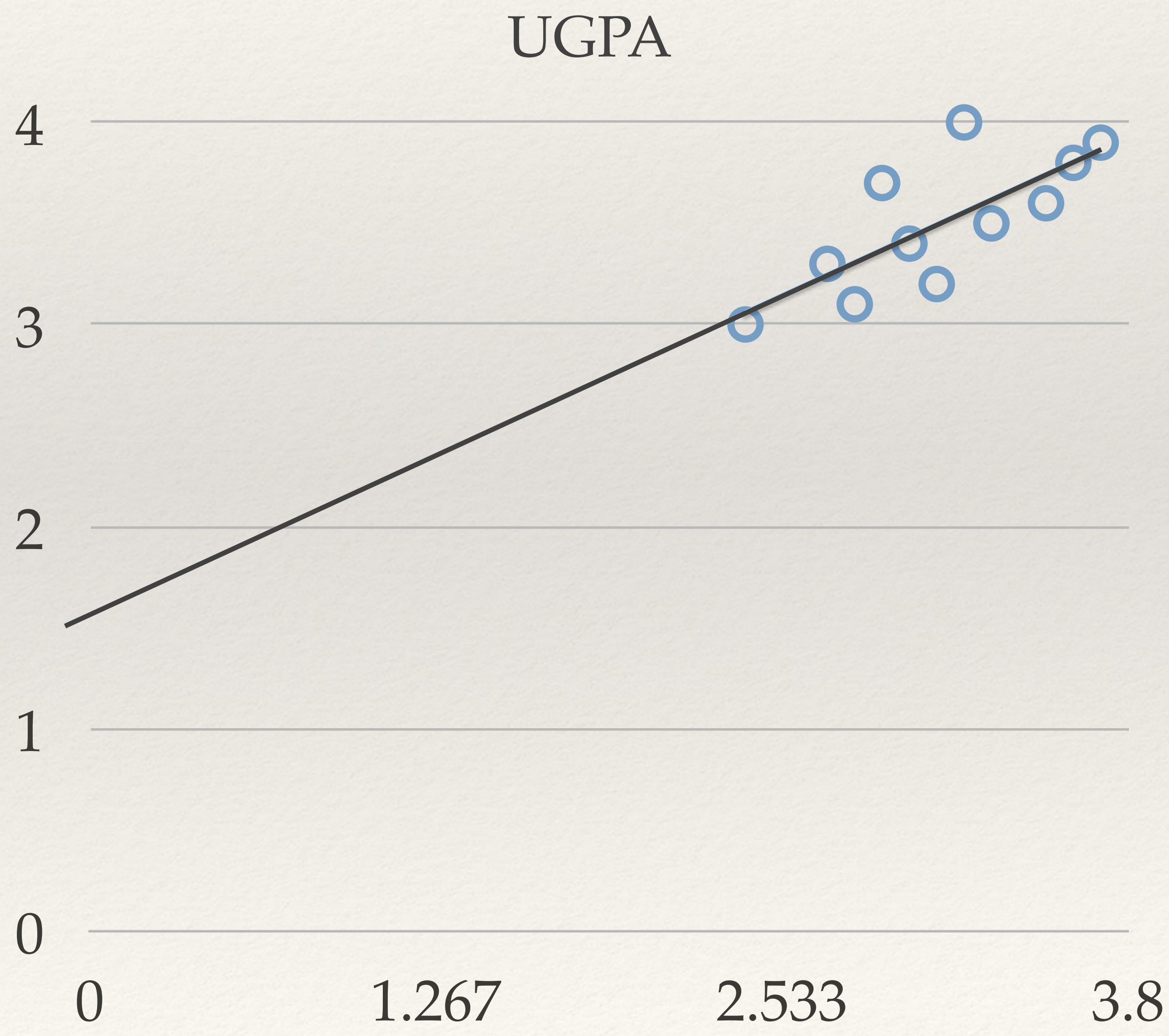
b_1 = change in Graduate GPA for every one-unit change in GRE Scores, with Undergraduate GPA held constant

b_2 = change in Graduate GPA for every one-unit change in Undergraduate GPA, with GRE held constant

e = error

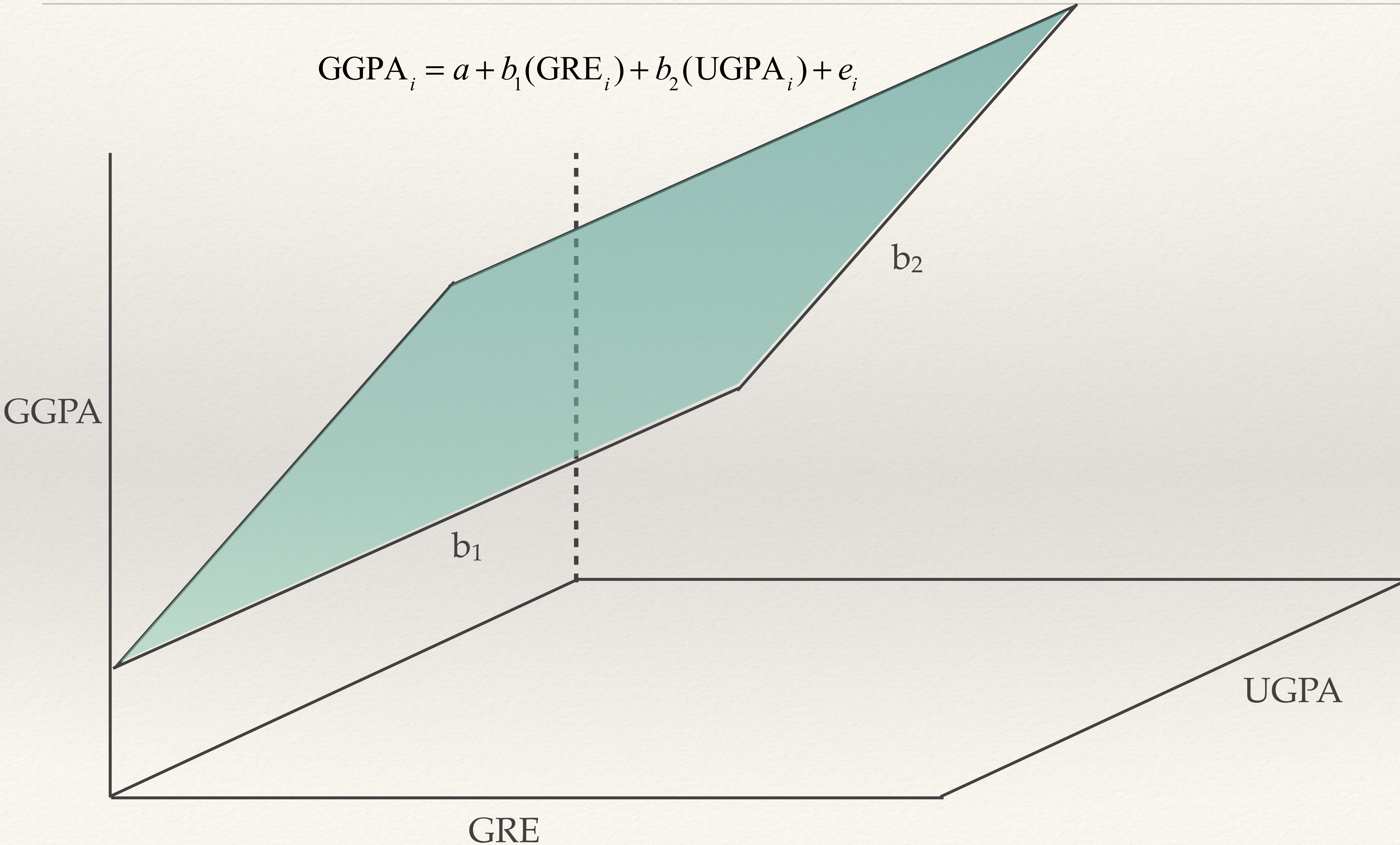
The Regression Plane



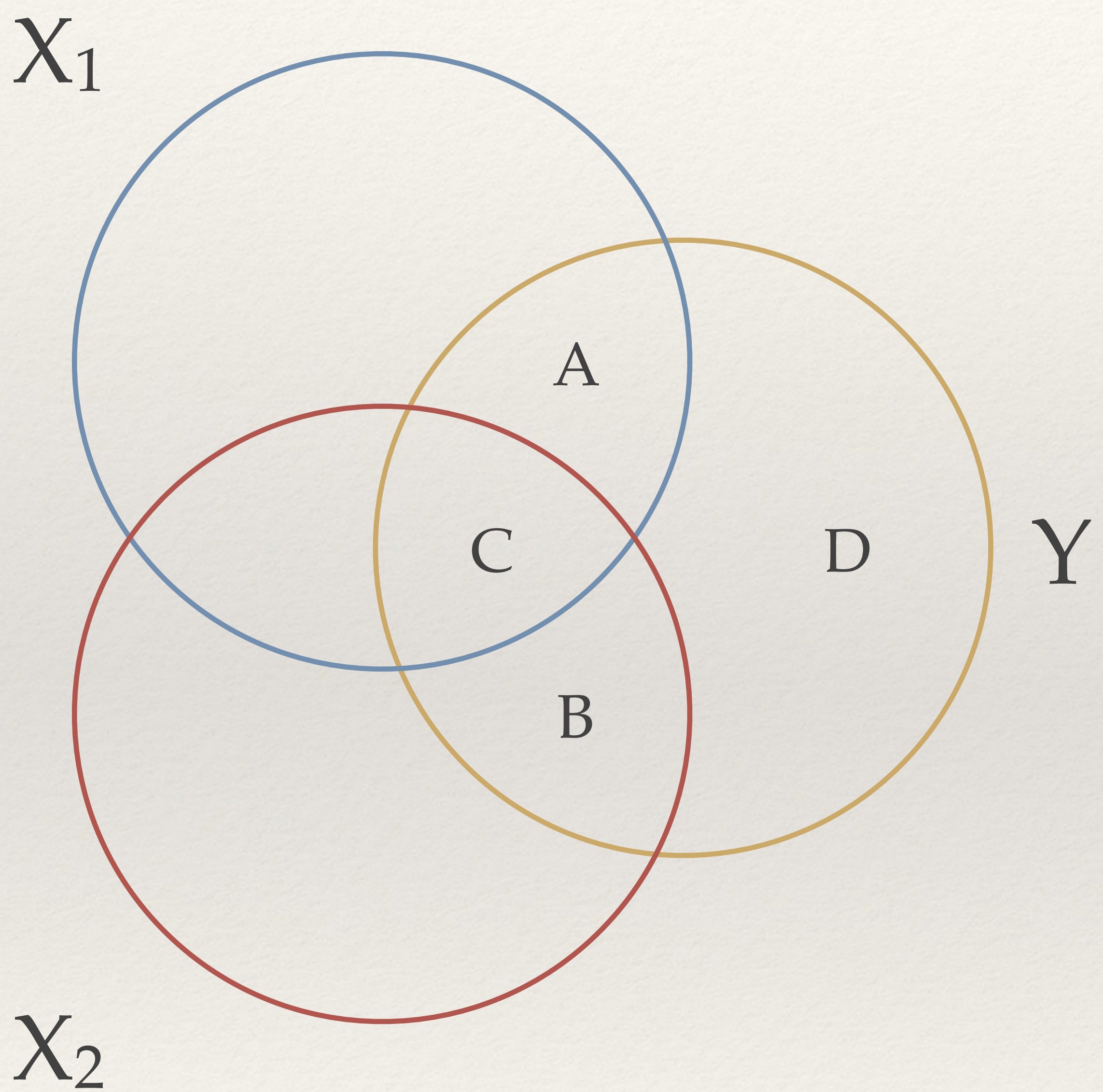


The Regression Plane

$$\mathrm{GGPA}_i = a + b_1(\mathrm{GRE}_i) + b_2(\mathrm{UGPA}_i) + e_i$$



Model Fit



Total Variance

$$SS_T = A + B + C + D$$

R-squared

$$r^2_{YX_1} = A + C$$

$$r^2_{YX_2} = B + C$$

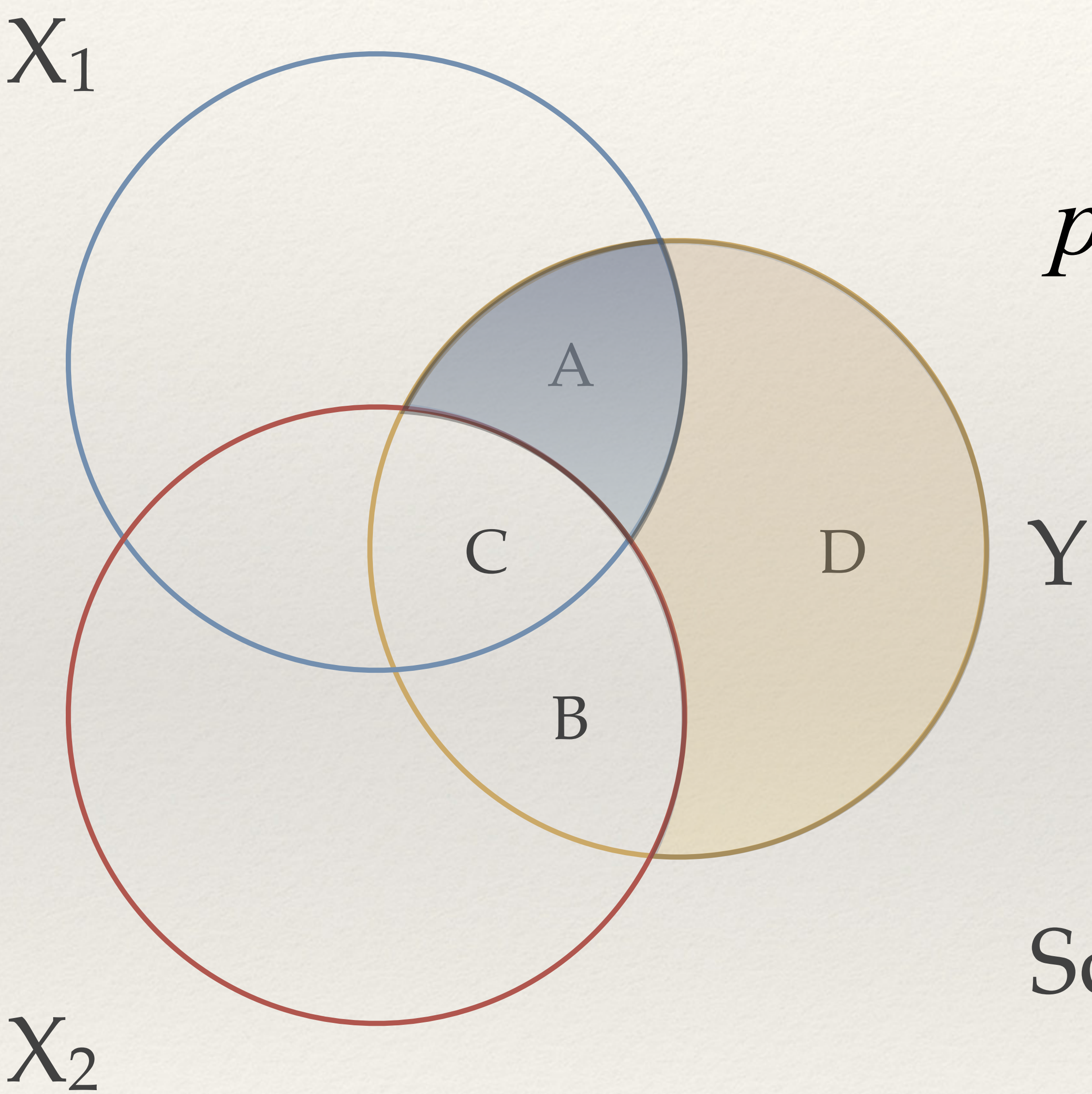
Multiple R-squared

$$R^2 = A + B + C$$

Unexplained Variance

$$1 - R^2 = D$$

Partial and Semi-Partial Squared Correlations



Squared Partial Correlations

$$pr^2_{YX_1.X_2} = \frac{\left[\left(A+B+C\right)-\left(B+C\right)\right]}{\left[1-\left(B+C\right)\right]}$$

Squared Semi-Partial Correlations

$$sr^2_{Y\left(X_1.X_2\right)} = \left(A+B+C\right)-\left(B+C\right)$$

Semi-Partial Correlations and Partial Slopes

Semi-Partial Correlations

$$r_{Y(X_1.X_2)} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r^2_{X_1X_2}}}$$

Partial Regression Slopes

$$b_1 = \frac{\left(r_{Y1} - r_{Y2} r_{12}\right) S_Y^2}{\left(1 - r^2_{12}\right) S_1^2}$$

Note, X's are typically removed from here for simplicity

Semi-Partial Correlations and Partial Slopes

Partial Regression Slopes

$$b_1 = \frac{\left(r_{Y1} - r_{Y2} r_{12}\right) S_Y^2}{\left(1 - r^2{}_{12}\right) S_1^2}$$

$$b_2 = \frac{\left(r_{Y2} - r_{Y1} r_{12}\right) S_Y^2}{\left(1 - r^2{}_{12}\right) S_2^2}$$

Intercept

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

Multiple Regression Example

Grad GPA (Y)	GRE-Total (X_1)	Undergrad GPA (X_2)
4.0	145	3.2
3.9	120	3.7
3.8	125	3.6
3.7	130	2.9
3.6	110	3.5
3.5	100	3.3
3.4	95	3.0
3.3	115	2.7
3.2	105	3.1
3.1	90	2.8
3.0	105	2.4

Variable	Mean	Variance
GGPA (Y)	3.5	0.11
GRE (X_1)	112.73	266.82
UGPA (X_2)	3.11	0.16

$r_{y1} = 0.78$

$r_{y2} = 0.75$

$r_{12} = 0.30$

Multiple Regression Example

$$b_1 = \frac{\left(r_{Y1} - r_{Y2} r_{12}\right) S_Y^2}{\left(1 - r^2{}_{12}\right) S_1^2} = .0125$$

A partial slope of .0125 means if GRE increases by one point, then GGPA increases by .0125 points, controlling for UGPA.

$$b_2 = \frac{\left(r_{Y2} - r_{Y1} r_{12}\right) S_Y^2}{\left(1 - r^2{}_{12}\right) S_2^2} = .4687$$

A partial slope of .4687 means if UGPA increases by one point, then GGPA increases by .4687 points, controlling for GRE.

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 = .6337$$

An intercept of .6337 means if scores on GRE and UGPA are both zero (although not probable), then GGPA is .6337.

Multiple Regression Example

Prediction Equation

$$Y'_i = .6337 + .0125(X_{1i}) + .4687(X_{2i})$$

If your GRE was 130 and your UGPA was 3.5, then your predicted GGPA is:

$$Y'_i = .6337 + .0125(130) + .4687(3.5) = 3.8992$$

Regression Equation

$$Y_i = .6337 + .0125(X_{1i}) + .4687(X_{2i}) + e_i$$

Multiple Regression Fit

$$R^2_{Y.1,\ldots,m} = \frac{SS_M}{SS_T}$$

SS_T = Sums of Squares Total

SS_R = Sums of Squares Residual

SS_M = Sums of Squares Model

$$R^2 = .9089$$

Cohen's Suggested Cutoffs for **Multiple Regression**:

.10 = Small

.30 = Medium

.50 = Large

Multiple Regression Fit

Adjusted R-Squared

$$R^2_{adj} = 1 - \left[\left(\frac{N-1}{N-k-1}\right)\left(\frac{N-2}{N-k-2}\right)\left(\frac{N+1}{2}\right)\right]\left(1-R^2\right)$$ *Stein's Formula*

$$R^2_{adj} = 1 - \left(1-R^2\right)\frac{(N-1)}{N-k}$$ *Formula used in R*

Akaike Information Criterion

$$AIC = N\ln\left(\frac{SS_M}{N}\right) + 2k$$

Diagnostics for Multiple Regression Fit

Discrepancy

How different is a point from the other points in a scatterplot? That is, does the data point follow the general trend?

Leverage

How different is a predictor value from the other predictor values? How much impact does any one case have on the predictor?

Influence

Do any values display both discrepancy and leverage? Is any one case altering the regression line?

Multiple Regression Generalizations

Model Significance

$$H_0 : R^2 = \rho^2$$

$$F = \frac{\left(\frac{.9089}{2}\right)}{\left(\frac{1-.9089}{11-2-1}\right)} = 39.9078$$

$$F = \frac{\left(\frac{R^2}{m}\right)}{\left(\frac{1-R^2}{N-m-1}\right)}$$

$$F_{crit} =_{.05} F_{2,8} = 4.46$$

Reject H_0

Multiple Regression Example

Predictor Significance

$$H_0 : b_{YX_1} = \beta_{YX_1}$$

$$H_0 : b_{YX_2} = \beta_{YX_2}$$

$$t = \frac{b_{YX_1}}{S_{b_{YX_1}}} = 5.447$$

Reject H_0

$$t = \frac{b_{YX_2}}{S_{b_{YX_2}}} = 5.030$$

Reject H_0

$$t_{crit} =_{.05} t_8 = 2.306$$

Assumptions for Generalizations

Variable Type

Non-zero Variance

Multicollinearity

External Variables

Residuals

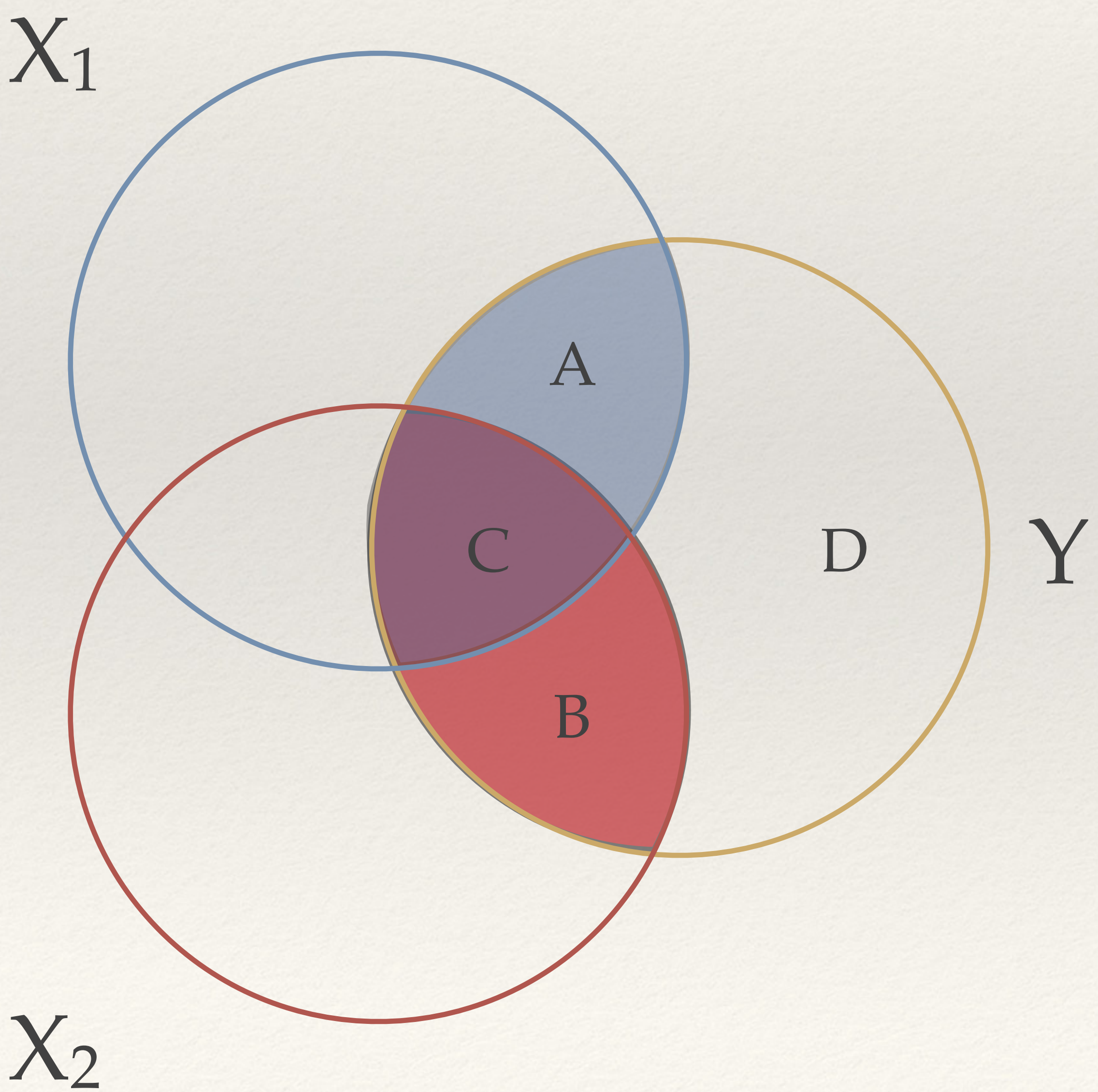
Independence

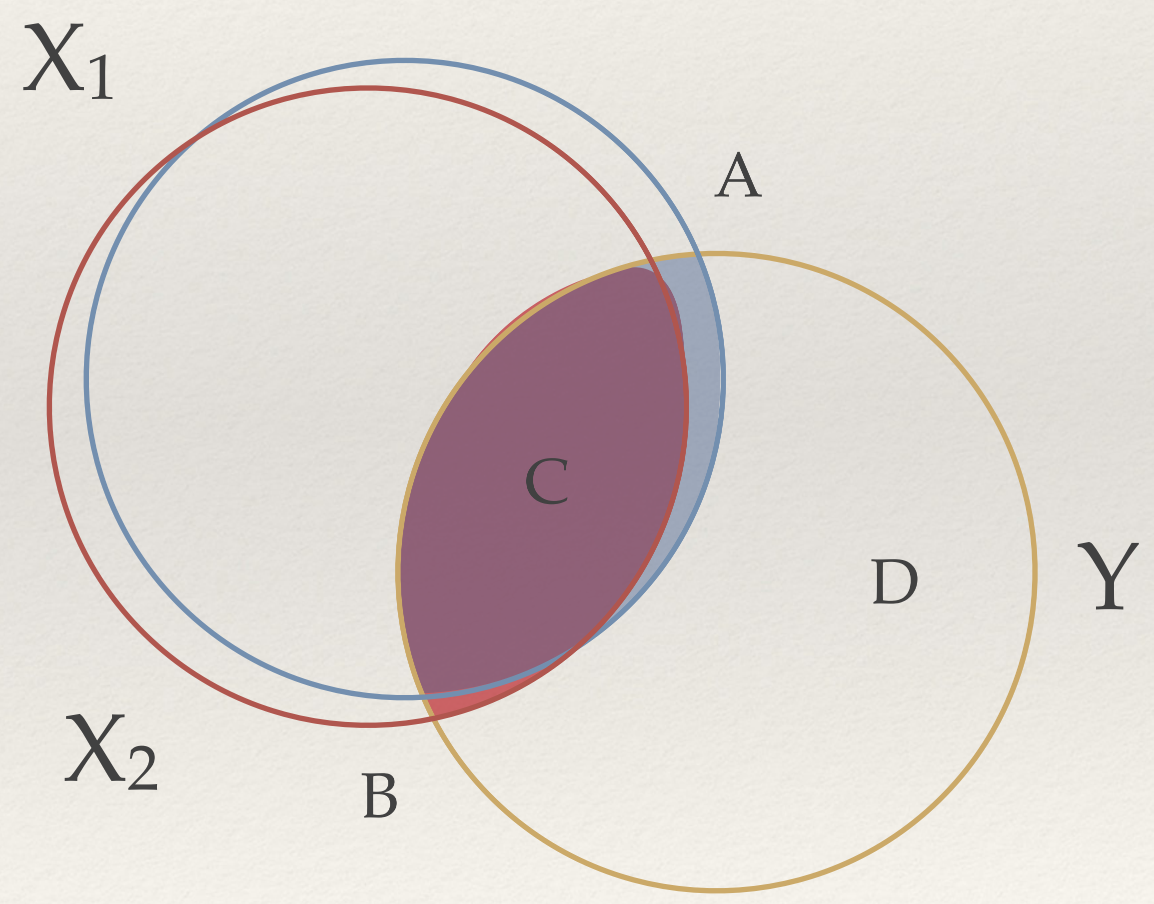
Linearity

Assumptions for Generalizations

Multicollinearity

Variance Inflation Factor (VIF)





The Standardized Prediction Model

$$z\left(Y_i^{'}\right) = b_1^{*} z_{1i} + b_2^{*} z_{2i} + \ldots + b_k^{*} z_{1i}$$

z(Y’) = The predicted Z score for Y

b^{*} = the standardized partial slope (beta weights)

z(X) = The Z score for X

*Note, these can be helpful in assessing the relative importance of given predictors to the model (i.e., units are in standard deviations). In essence, they are a **comparative** effect size measures for the predictors. They are also often reported as β.*

Methods of Entering Predictors

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

$$Y = X_1 + X_2$$

$$Y = X_1 + X_3 + X_5$$

Methods of Entering Predictors

Simultaneous

All variables are entered into the model at the same time

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

Sequential

Variables are systematically entered using a given technique in order to specify the model

Stepwise | **All Subsets** | **Hierarchical**

Methods of Entering Predictors

Stepwise

Variables are entered / removed one at a time with the impacts to the model being assessed with each step.

Backwards

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

Forwards

$$Y = X_1 + X_2 + X_3$$

Methods of Entering Predictors

All Subsets

Every possible model is assessed

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

All possible one-, two-, three-, four-, and five-predictor models would be analyzed.

Thus there will be 5 one-predictor models, 10 two-predictor models, 10 three-predictor models, 5 four-predictor models, and 1 five-predictor model.

Methods of Entering Predictors

Hierarchical

Variables are entered according to theoretical relevance and previous research

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

For any <u>nested</u> hierarchical models, comparisons can use an F-test of the R^2 change. The lowest AIC can also be used to make decisions about the best fitting models with the same outcome variable, regardless of nesting.

$$\Delta R^2 = R_L^2 - R_S^2$$

$$AIC = N \ln\left(\frac{SS_M}{N}\right) + 2k$$

Nested Models

$$Y = X_1 + X_2 + X_3 + X_4$$

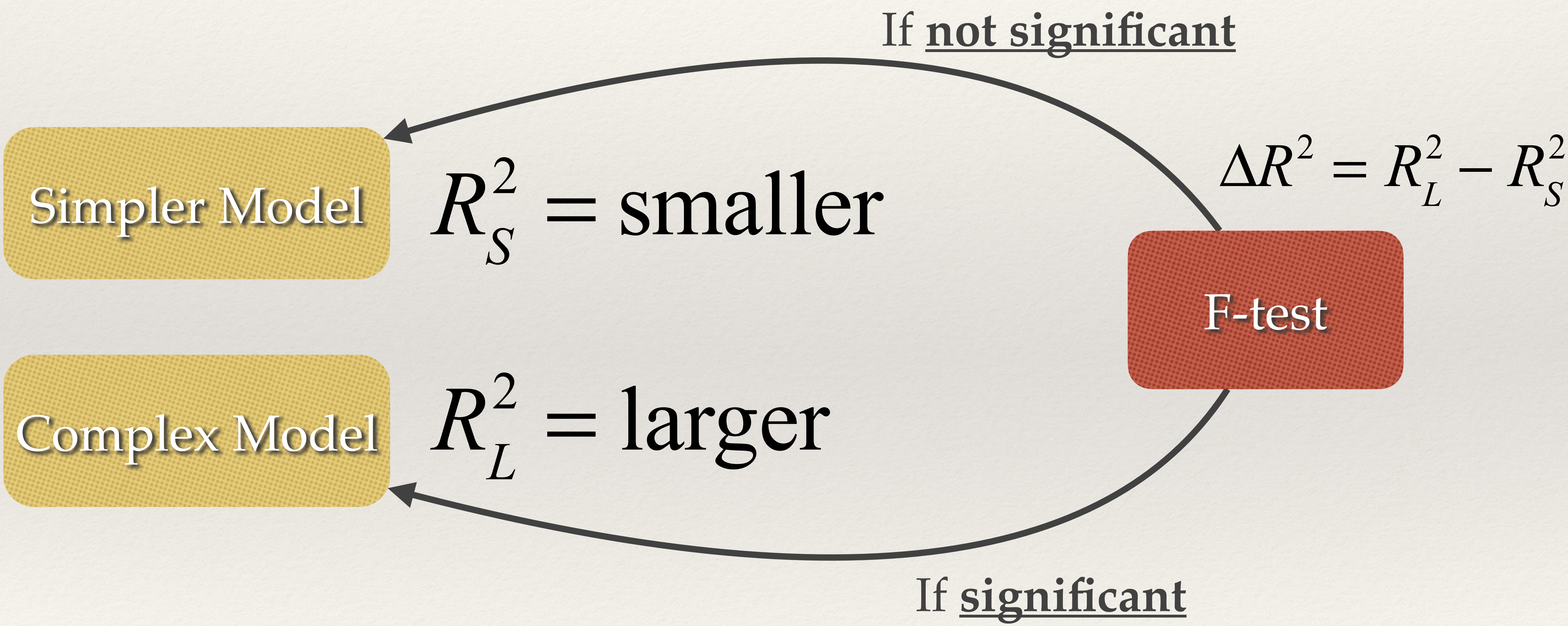
$$Y = X_1 + X_2 + X_4$$

$$Y = X_1 + X_2$$

$$Y = X_1$$

$$\Delta R^2 = R_L^2 - R_S^2$$

Decisions in Hierarchical Regression



Non-Nested Models

$$Y = X_1 + X_2 + X_3 + X_4$$

$$Y = X_1 + X_2 + X_5$$

Not Nested

$$Y = X_3 + X_5$$

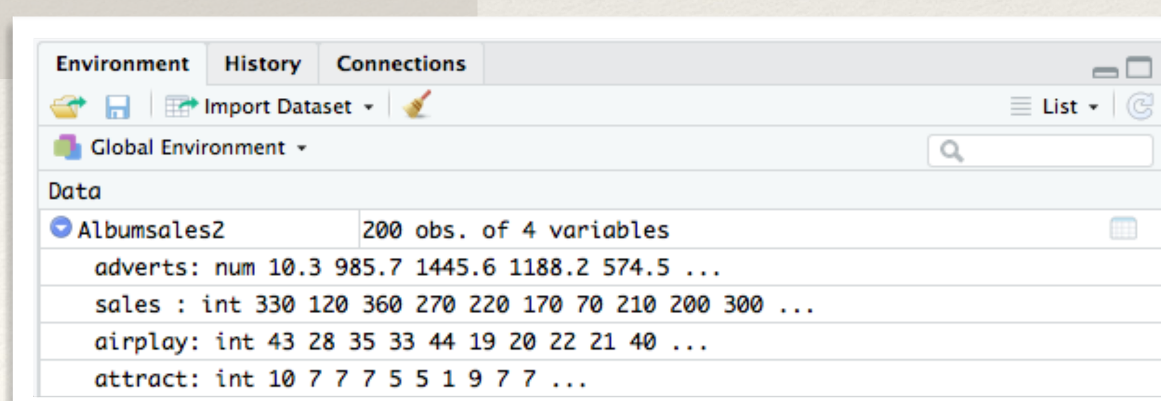
$$AIC = N\ln\left(\frac{SS_M}{N}\right) + 2k$$

R Studio: Data

Albumsales2

Variable	Explanation
adverts	Money spent on advertisement (in thousands of pounds)
Sales	Albums sold (in thousands)
Airplay	Number of times song was played on major radio station week prior to album release
Atrract	Mode Attractiveness of band on a scale of 0-10

Hierarchical



R Studio: Data

Hierarchical

$$\text{Sales}_i = a + e_i$$

Mod 0

$$\text{Sales}_i = a + b_1(\text{Adverts}_i) + e_i$$

Mod1 R²

Mod2 R²

$$\text{Sales}_i = a + b_1(\text{Adverts}_i) + b_2\left(\text{Air}_i\right) + b_2\left(\text{Attract}_i\right) + e_i$$

$$\Delta R^2 = R_L^2 - R_S^2$$

$\Delta R^2 =$ Mod2 R² − Mod1 R²

F-test

R Studio: The Linear Model Function

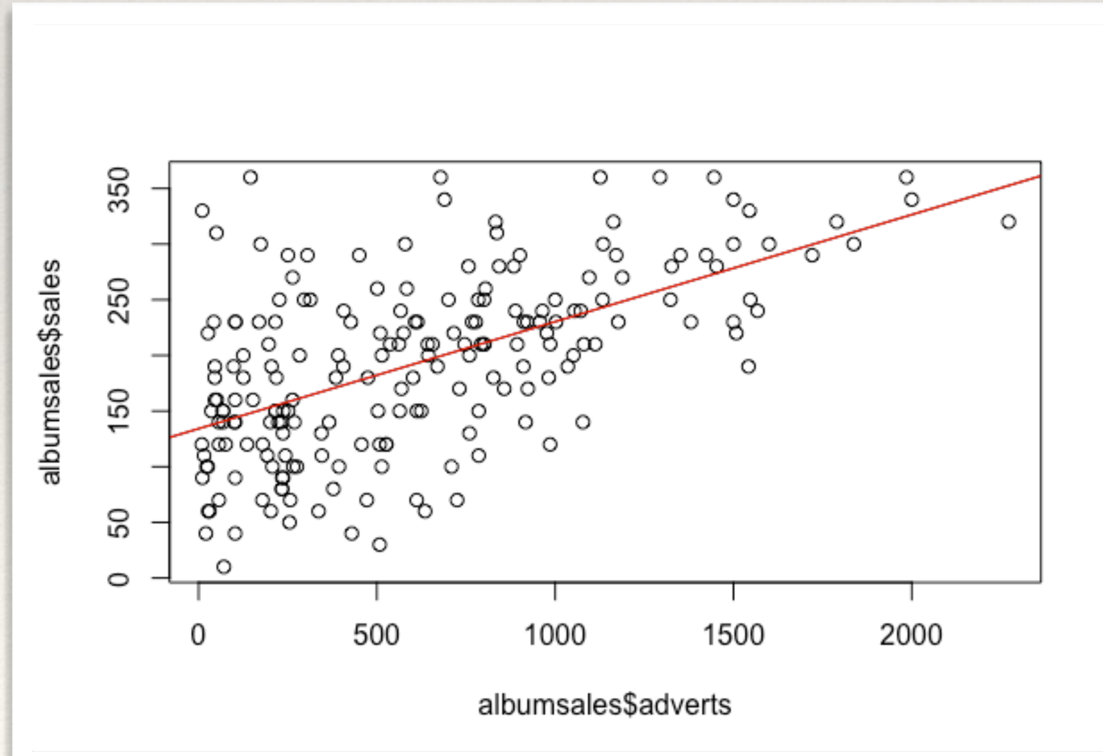
lm() General code for creating linear models in R.

mod1 = lm(sales ~ adverts, data = Albumsales2)

abline(a = 134.13994, b = .09612, col = 'red')

mod1

```
> mod1 = lm(sales ~ adverts, data = Albumsales2)
> mod1

Call:
lm(formula = sales ~ adverts, data = Albumsales2)

Coefficients:
(Intercept)      adverts
  134.13994      0.09612
```



R Studio: The Linear Model Function

lm() General code for creating linear models in R.

mod2 = lm(sales ~ adverts + airplay + attract, data = Albumsales2)

mod2

```
> mod2 = lm(sales ~ adverts + airplay + attract, data = Albumsales2)
> mod2

Call:
lm(formula = sales ~ adverts + airplay + attract, data = Albumsales2)

Coefficients:
(Intercept)      adverts      airplay      attract
  -26.61296      0.08488      3.36743     11.08634
```

R Studio: The Linear Model Function

summary()

General code for summarizing given object. If object is a linear model, a regression analysis is provided.

summary(mod1)

```
> summary(mod1)

Call:
lm(formula = sales ~ adverts, data = Albumsales2)

Residuals:
     Min       1Q   Median       3Q      Max
-152.949  -43.796   -0.393   37.040  211.866

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.341e+02  7.537e+00  17.799   <2e-16 ***
adverts     9.612e-02  9.632e-03   9.979   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 65.99 on 198 degrees of freedom
Multiple R-squared:  0.3346	Adjusted R-squared:  0.3313
F-statistic: 99.59 on 1 and 198 DF,  p-value: < 2.2e-16
```

summary(mod2)

```
> summary(mod2)

Call:
lm(formula = sales ~ adverts + airplay + attract, data = Albumsales2)

Residuals:
     Min       1Q   Median       3Q      Max
-121.324  -28.336   -0.451   28.967  144.132

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -26.612958  17.350001  -1.534    0.127
adverts       0.084885   0.006923  12.261  < 2e-16 ***
airplay       3.367425   0.277771  12.123  < 2e-16 ***
attract      11.086335   2.437849   4.548 9.49e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47.09 on 196 degrees of freedom
Multiple R-squared:  0.6647,	Adjusted R-squared:  0.6595
F-statistic: 129.5 on 3 and 196 DF,  p-value: < 2.2e-16
```

R Studio: The Linear Model Function

anova()

Will conduct F-test on given objects, including comparison of two models

anova(mod1, mod2)

```
> anova(mod1, mod2)
Analysis of Variance Table

Model 1: sales ~ adverts
Model 2: sales ~ adverts + airplay + attract
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    198 862264
2    196 434575  2    427690 96.447 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

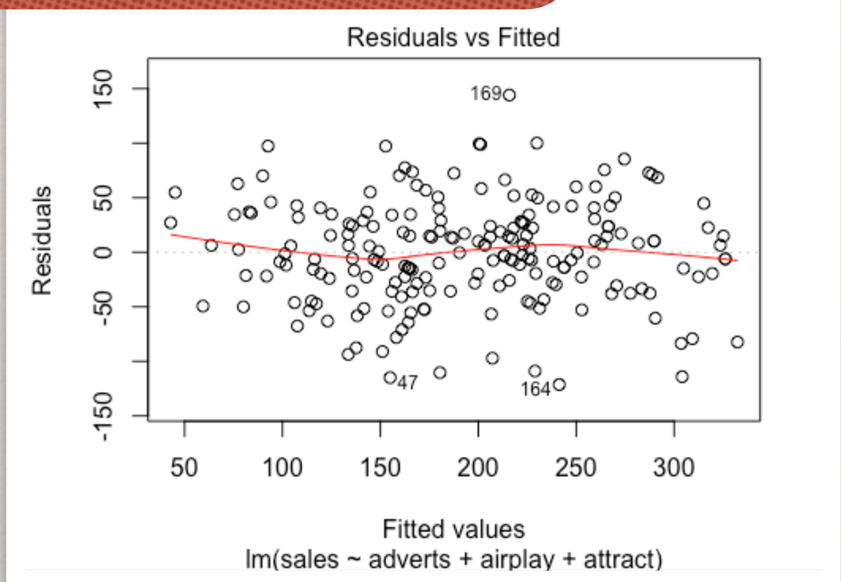
Mod2

$$\Delta R^2 = R_L^2 - R_S^2$$

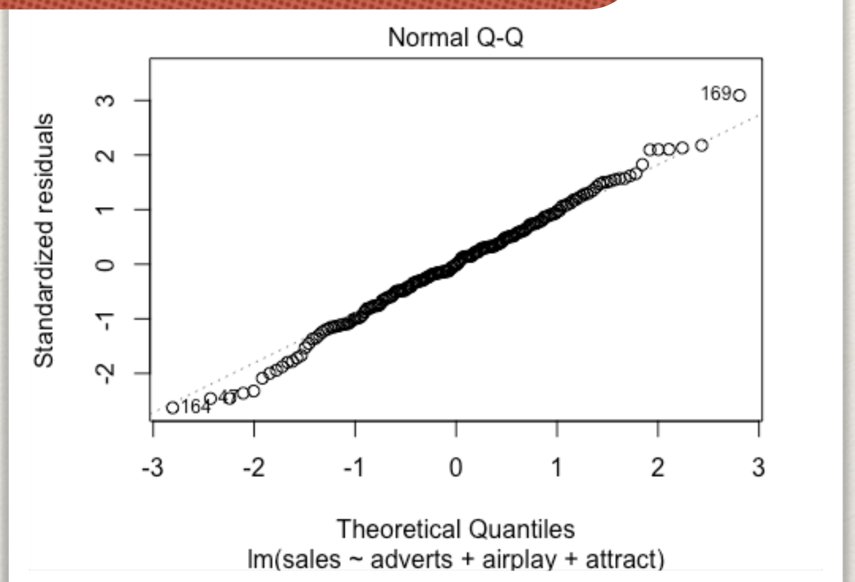
ChangeRmod1.2 = .6647 -.3346

R Studio: Visualizing Data

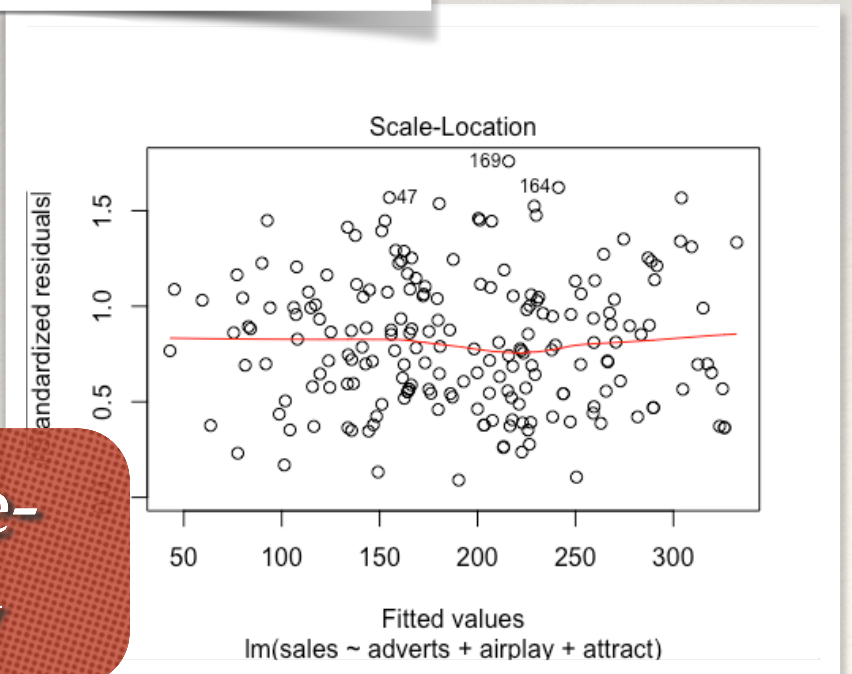
Linearity



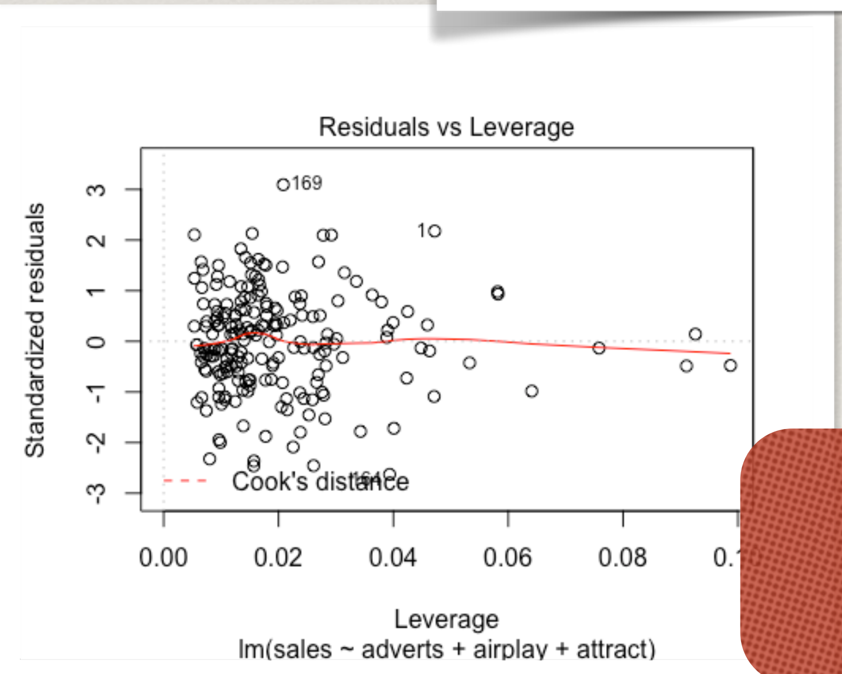
Normality of Error Term



plot(mod2)



Homoske-
dasticity



Influence

R Studio: Assessing Model

Diagnostics

Significance

```
Albumsales2$std.residuals = rstandard(mod2)

Albumsales2$leverage = hatvalues(mod2)

Albumsales2$Cooks.distance = cooks.distance(mod2)
```

R Studio:Outliers and Influential Cases

Diagnostics

which(Albumsales2$std.residuals > 3.29 | Albumsales2$std.residuals < -3.29)#Discrepancy

which(Albumsales2$leverage > 3*mean(Albumsales2$leverage))#Leverage

which(Albumsales2$Cooks.distance > 1)#Influence

```
> which(Albumsales2$std.residuals > 3.29 | Albumsales2$std.residuals < -3.29)#Discrepancy
integer(0)
> which(Albumsales2$leverage > 3*mean(Albumsales2$leverage))#Leverage
[1]   7  12 138 181 184
> which(Albumsales2$Cooks.distance > 1)#Influence
integer(0)
```

Assumptions for Generalizations

Assumptions

Albumsales2$residuals = resid(mod2)

shapiro.test(albumsales$residuals)

```
        Shapiro-Wilk normality test

data:  Albumsales2$residuals
W = 0.99483, p-value = 0.7253
```

library(lmtest)

bptest(mod2, studentize = FALSE)

```
        Breusch-Pagan test

data:  mod2
BP = 6.635, df = 3, p-value = 0.08449
```

library(car)

dwt(mod2)

```
lag Autocorrelation D-W Statistic p-value
  1       0.0026951      1.949819    0.72
Alternative hypothesis: rho != 0
```

Assumptions for Generalizations

vif() Part of "car" package, used to assess multicollinearity.

vif(mod2)#Nothing larger than 10

1/vif(mod2)#Below .2 maybe, below .1 is an issue

mean(vif(mod2))#Should not be substantially greater than 1

```
> ##Multicollinearity
> vif(mod2)#Nothing larger than 10
 adverts  airplay  attract 
1.014593 1.042504 1.038455 
> 1/vif(mod2)#Below .2 maybe, beolow .1 is an issue
  adverts   airplay   attract 
0.9856172 0.9592287 0.9629695 
> mean(vif(mod2))#Should not be substantially greater than 1
[1] 1.03185
```

R Studio: The Linear Model Function

summary()

General code for summarizing given object. If object is a linear model, a regression analysis is provided.

summary(mod2)

```
> summary(mod2)

Call:
lm(formula = sales ~ adverts + airplay + attract, data = Albumsales2)

Residuals:
     Min       1Q   Median       3Q      Max 
-121.324  -28.336   -0.451   28.967  144.132 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -26.612958  17.350001  -1.534    0.127    
adverts       0.084885   0.006923  12.261  < 2e-16 ***
airplay       3.367425   0.277771  12.123  < 2e-16 ***
attract      11.086335   2.437849   4.548 9.49e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47.09 on 196 degrees of freedom
Multiple R-squared:  0.6647,	Adjusted R-squared:  0.6595 
F-statistic: 129.5 on 3 and 196 DF,  p-value: < 2.2e-16
```

Step 2: Predictors

Step 1: Model

R Studio: The Linear Model Function

lm.beta()

Part of QuantPsyc package. Will provide standardized regression coefficients for given model

library(QuantPsyc)

lm.beta(mod2)

```
> lm.beta(mod2)
  adverts   airplay   attract
0.5108462 0.5119881 0.1916834
```

A standardized partial slope of .51 means if Adverts increases by one standard deviation, then Sales increases by .51 standard deviations, controlling for both airplay and attractiveness.

Results Table

	ΔR^2	b	SE_b	β	p
Step 1	0.33				< .001
Constant		134.14	7.54		< .001
Adverts		0.10	0.01	0.58	< .001
Step 2	0.33 Change				< .001 anova command
Constant		-26.61	17.35		> .05
Adverts		0.09	0.01	0.51	< .001
Airplay		3.37	0.28	0.51	< .001
Attract		11.09	2.44	0.19	< .001

APA Write Up

Hierarchical regression analysis (see Table 1) was used to determine that the more complex model with airplay and attractiveness accounted for significantly more variance ($\Delta R^2 = 0.33$, $p < .001$) than the simple linear model with only advertisements. Diagnostics for this more complex model noted no concerns with influential cases, and assumption testing found no concerns with multicollinearity, normality, homoskedasticity, and independence of the error. The more complex model was significant, $F(3, 196) = 129.50$, $p < .001$ and represented a large effect ($R^2_{adj} = 0.66$) as it accounted for 66% of the variance in sales. Advertisements significantly predicted sales while controlling for airplay and attractiveness, $t(196) = 12.61$, $p < .001$; airplay significantly predicted sales while controlling for advertisements and attractiveness, $t(196) = 12.12$, $p < .001$; and attractiveness significantly predicted sales while controlling for advertisements and airplay, $t(196) = 4.54$, $p < .001$. Advertisements ($\beta = 0.51$) and airplay ($\beta = 0.51$) had the strongest and near equal effects on sales, with an expected 0.51 standard deviation increase in sales for every one standard deviation increase in the respective variable, while controlling for the other variables in the model. Attractiveness ($\beta = 0.19$) had a smaller effect with a 0.19 standard deviation increase in sales for every one standard deviation increase while controlling for advertisements and attractiveness.

For Next Meeting

- Read: Posted Readings
- AP2: Check Canvas due date
- Final Project Data Plan