PSYC308C.DA2

Daniel Pinedo

**Prompt** You are employed by a local university to conduct research regarding the performance of incoming students during their first semester. The grades reported by the previous years has been markedly low, which is concerning for the reputation and quality of the school. The university is interested in student’s commitment to studying, amount of stress felt, and the average hours of sleep each student was getting and how these variables impact average test scores. You are put to the task of determining what factors impact this year’s incoming students exam scores and must report back to the university in order to develop an intervention to improve the grades of future first-years.

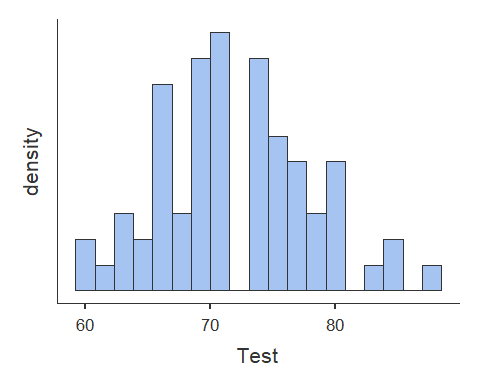
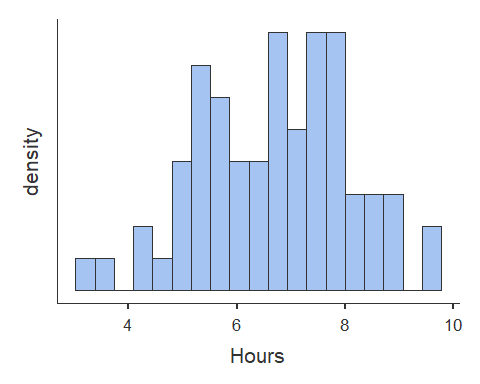
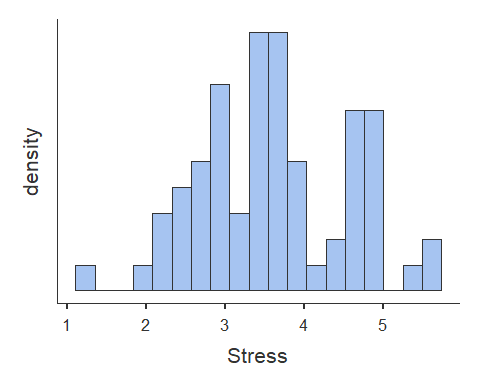
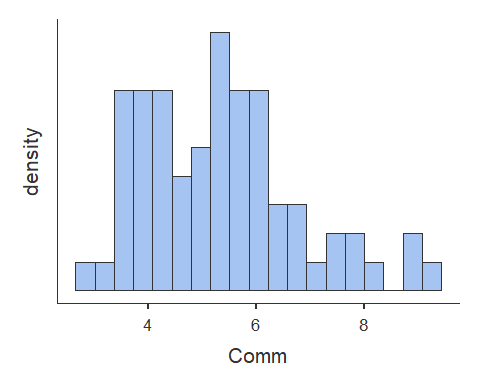
**Variables** Commitment (Comm) - average amount of commitment to studying in the past week, 1-10 scale. Stress - average amount of stress report by students in the past week, 1-10 scale. Hours of Sleep (Hours) - average amount of hours of sleep in the past week, 1-10 scale. Test Score (Test) (outcome) - average test score, 1-100 scale.

**Hypotheses** H0: no relationship between variables Ha: commitment + stress + hours predict test score N = 70

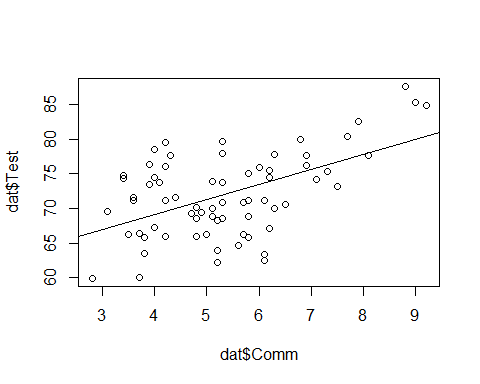
**Descriptive Statistics and Assumptions**

# Prerequisitites  
 # 1. Variables are measured on the continuous level  
  
# Assumptions  
 # 1. Normal Distribution for X and Y (Product) [i.e. histogram, skew +-3, kurtosis +-10]  
 # Histogram for Comm appears positively skewed and unimodal  
 # Histogram for Stress appears negatively skewed and unimodal  
 # Histogram for Hours appears symmetric and bimodal  
 # Histogram for Test appears symmetric and bimodal  
 # Skewness - Comm 0.65 Stress 0.14 Hours -0.21 Test 0.26  
 # Kurtosis - 0.11 -0.30 -0.28 -0.21   
   
 # 2. Linear Relationship beween X and Y  
 # Visual inspection of scatterplot and prediction model line indicate ...  
 # 3. Homoscedasticity  
 # a. Visual inspection of scatterplots indicate:   
 # higher variance at the lower end for Comm   
 # assymettric variance at the lower end for Stress  
 # higher variance at the upper end for Hours  
 # b. non-constant variance test - H0 = TRUE (PASS)  
   
 # 4. [Examine residuals (e = Y - Y~predicted~) to understand 2 and 3 mathematically]  
  
# Descriptives [Assumption 1]  
desc <- descriptives(data = dat,   
 vars = c('Comm', 'Stress', 'Hours', 'Test'),   
 hist = TRUE,   
 sd = TRUE,   
 range = TRUE,   
 skew = TRUE,   
 kurt = TRUE)  
desc

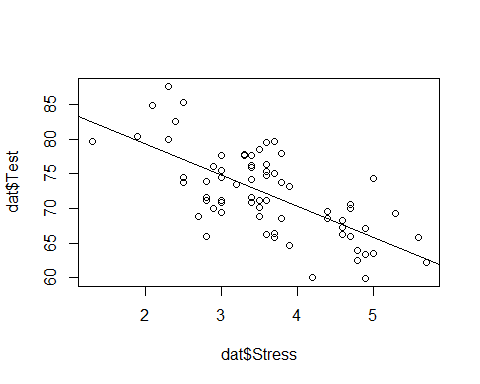
##   
## DESCRIPTIVES  
##   
## Descriptives   
## --------------------------------------------------------------   
## Comm Stress Hours Test   
## --------------------------------------------------------------   
## N 70 70 70 70   
## Missing 0 0 0 0   
## Mean 5.36 3.62 6.71 72.0   
## Median 5.25 3.55 6.85 71.2   
## Standard deviation 1.45 0.923 1.39 6.05   
## Range 6.40 4.40 6.40 27.7   
## Minimum 2.80 1.30 3.20 59.9   
## Maximum 9.20 5.70 9.60 87.6   
## Skewness 0.650 0.138 -0.214 0.256   
## Std. error skewness 0.287 0.287 0.287 0.287   
## Kurtosis 0.107 -0.304 -0.277 -0.206   
## Std. error kurtosis 0.566 0.566 0.566 0.566   
## --------------------------------------------------------------



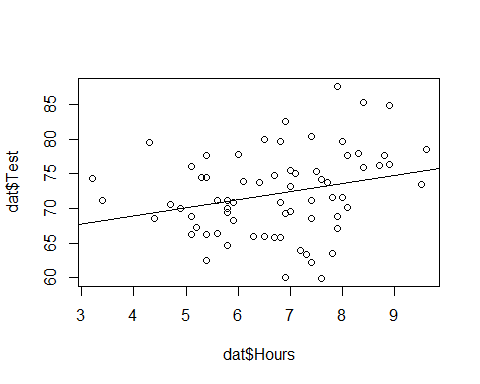
# Scatterplots [Assumption 2 and 3a]  
plot(dat$Comm, dat$Test, abline(lm(dat$Test ~ dat$Comm)))



plot(dat$Stress, dat$Test, abline(lm(dat$Test ~ dat$Stress)))



plot(dat$Hours, dat$Test, abline(lm(dat$Test ~ dat$Hours)))



# Homoscedasticity [Assumption 3b]  
  
#non-constant variance Chi-squared test [Chi-squared (df) = ##.##, p = .###]  
#H0 = homoscedastic - TRUE  
#Ha = heteroscedastic  
  
ncvTest(lm(Test ~ Comm + Stress + Hours, data = dat))

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 3.590164, Df = 1, p = 0.058123

**Correlations**

# Correlation  
cortable <- corrMatrix(data = dat,   
 vars = c('Comm', 'Stress', 'Hours', 'Test'),   
 flag = TRUE)  
cortable

##   
## CORRELATION MATRIX  
##   
## Correlation Matrix   
## ----------------------------------------------------------------   
## Comm Stress Hours Test   
## ----------------------------------------------------------------   
## Comm Pearson's r  -0.355 0.305 0.522   
## p-value  0.003 0.010 < .001   
##   
## Stress Pearson's r  -0.115 -0.684   
## p-value  0.343 < .001   
##   
## Hours Pearson's r  0.268   
## p-value  0.025   
##   
## Test Pearson's r    
## p-value    
## ----------------------------------------------------------------   
## Note. \* p < .05, \*\* p < .01, \*\*\* p < .001

**Simple Regression**

# Simple Regression Model 1  
# Start with the simpler model first - Stress is most correlated with outcome variable (Test)  
model1 <- linReg(data = dat,   
 dep = 'Test', #outcome  
 covs = c('Stress'), #predictors  
 blocks = list(c('Stress')), #order - doesn't matter for simple regression as there is only one variable  
 modelTest = TRUE, #significance test on model [H0: R squared = 0]  
 stdEst = TRUE) #standardized regression coefficient for individual variable [Stand. Estimate or Beta]  
model1 #print to screen

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## -----------------------------------------------------------   
## Model R R² F df1 df2 p   
## -----------------------------------------------------------   
## 1 0.684 0.467 59.7 1 68 < .001   
## -----------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate   
## ------------------------------------------------------------------------   
## Intercept 88.26 2.169 40.70 < .001   
## Stress -4.48 0.580 -7.72 < .001 -0.684   
## ------------------------------------------------------------------------

#This model is best fit for simple regression based on R squared and Beta Estimates  
  
# ALTERNATIVE  
model1.1<- lm(Test ~ Stress, data = dat)  
summary(model1.1)

##   
## Call:  
## lm(formula = Test ~ Stress, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.8125 -2.8887 -0.3681 3.0985 9.6469   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 88.2601 2.1686 40.699 < 2e-16 \*\*\*  
## Stress -4.4813 0.5801 -7.725 6.92e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.446 on 68 degrees of freedom  
## Multiple R-squared: 0.4674, Adjusted R-squared: 0.4596   
## F-statistic: 59.68 on 1 and 68 DF, p-value: 6.92e-11

# Simple Regression Model 2  
# Comm is second most correlated with outcome variable (Test)  
model2 <- linReg(data = dat,   
 dep = 'Test', #outcome  
 covs = c('Comm'), #predictors  
 blocks = list(c('Comm')), #order - doesn't matter for simple regression as there is only one variable  
 modelTest = TRUE, #significance test on model [H0: R squared = 0]  
 stdEst = TRUE) #standardized regression coefficient for individual variable  
model2 #print to screen

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## -----------------------------------------------------------   
## Model R R² F df1 df2 p   
## -----------------------------------------------------------   
## 1 0.522 0.272 25.4 1 68 < .001   
## -----------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate   
## ------------------------------------------------------------------------   
## Intercept 60.40 2.386 25.31 < .001   
## Comm 2.17 0.430 5.04 < .001 0.522   
## ------------------------------------------------------------------------

# model has predictive significance  
  
# ALTERNATIVE  
model2.1<- lm(Test ~ Comm, data = dat)  
summary(model2.1)

##   
## Call:  
## lm(formula = Test ~ Comm, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.1331 -3.4520 -0.8247 4.2373 9.9885   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 60.4005 2.3864 25.310 < 2e-16 \*\*\*  
## Comm 2.1693 0.4302 5.042 3.64e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.198 on 68 degrees of freedom  
## Multiple R-squared: 0.2721, Adjusted R-squared: 0.2614   
## F-statistic: 25.42 on 1 and 68 DF, p-value: 3.64e-06

# Simple Regression Model 3  
# Hours is third most correlated with outcome variable (Test)  
model3 <- linReg(data = dat,   
 dep = 'Test', #outcome  
 covs = c('Hours'), #predictors  
 blocks = list(c('Hours')), #order - doesn't matter for simple regression as there is only one variable  
 modelTest = TRUE, #significance test on model [H0: R squared = 0]  
 stdEst = TRUE) #standardized regression coefficient for individual variable  
model3 #print to screen

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## -----------------------------------------------------------   
## Model R R² F df1 df2 p   
## -----------------------------------------------------------   
## 1 0.268 0.0716 5.24 1 68 0.025   
## -----------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate   
## ------------------------------------------------------------------------   
## Intercept 64.20 3.488 18.41 < .001   
## Hours 1.17 0.509 2.29 0.025 0.268   
## ------------------------------------------------------------------------

# model has predictive significance  
  
# ALTERNATIVE  
model3.1<- lm(Test ~ Hours, data = dat)  
summary(model2.1)

##   
## Call:  
## lm(formula = Test ~ Comm, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.1331 -3.4520 -0.8247 4.2373 9.9885   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 60.4005 2.3864 25.310 < 2e-16 \*\*\*  
## Comm 2.1693 0.4302 5.042 3.64e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.198 on 68 degrees of freedom  
## Multiple R-squared: 0.2721, Adjusted R-squared: 0.2614   
## F-statistic: 25.42 on 1 and 68 DF, p-value: 3.64e-06

**Multiple Regression**

# Multiple regression test #A [Test ~ Stress + Comm] - best fit  
  
modelA <- linReg(data = dat,   
 dep = 'Test', #outcome  
 covs = c('Stress', 'Comm'), #predictors  
 blocks = list(c('Stress', 'Comm')), #order matters here if separate blocks of variables are provided  
 modelTest = TRUE,   
 stdEst = TRUE,   
 ciStdEst = TRUE,   
 r2Adj = TRUE)  
modelA

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.746 0.556 0.543 42.0 2 67 < .001   
## --------------------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## --------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## --------------------------------------------------------------------------------------------   
## Intercept 78.46 3.335 23.53 < .001   
## Stress -3.74 0.571 -6.55 < .001 -0.570 -0.744 -0.397   
## Comm 1.33 0.362 3.67 < .001 0.319 0.145 0.493   
## --------------------------------------------------------------------------------------------

# ALTERNATIVE  
modelA.1<- lm(Test ~ Stress + Comm, data = dat)  
summary(modelA.1)

##   
## Call:  
## lm(formula = Test ~ Stress + Comm, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.6690 -2.5243 -0.3167 2.5740 10.0199   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.4604 3.3346 23.529 < 2e-16 \*\*\*  
## Stress -3.7385 0.5705 -6.553 9.5e-09 \*\*\*  
## Comm 1.3271 0.3620 3.666 0.000488 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.088 on 67 degrees of freedom  
## Multiple R-squared: 0.5564, Adjusted R-squared: 0.5432   
## F-statistic: 42.02 on 2 and 67 DF, p-value: 1.493e-12

# Multiple regression test #A [Test ~ Stress + Comm + Hours]  
  
modelB <- linReg(data = dat,   
 dep = 'Test', #outcome  
 covs = c('Stress', 'Comm', 'Hours'), #predictors  
 blocks = list(c('Stress', 'Comm', 'Hours')), #order matters here if separate blocks of variables are provided  
 modelTest = TRUE,   
 stdEst = TRUE,   
 ciStdEst = TRUE,   
 r2Adj = TRUE)  
modelB

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.754 0.568 0.549 29.0 3 66 < .001   
## --------------------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ---------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## ---------------------------------------------------------------------------------------------   
## Intercept 75.848 3.833 19.79 < .001   
## Stress -3.733 0.567 -6.58 < .001 -0.569 -0.7422 -0.397   
## Comm 1.182 0.375 3.15 0.002 0.284 0.1041 0.464   
## Hours 0.502 0.370 1.36 0.179 0.115 -0.0543 0.285   
## ---------------------------------------------------------------------------------------------

# ALTERNATIVE  
modelB.1<- lm(Test ~ Stress + Comm, data = dat)  
summary(modelB.1)

##   
## Call:  
## lm(formula = Test ~ Stress + Comm, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.6690 -2.5243 -0.3167 2.5740 10.0199   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.4604 3.3346 23.529 < 2e-16 \*\*\*  
## Stress -3.7385 0.5705 -6.553 9.5e-09 \*\*\*  
## Comm 1.3271 0.3620 3.666 0.000488 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.088 on 67 degrees of freedom  
## Multiple R-squared: 0.5564, Adjusted R-squared: 0.5432   
## F-statistic: 42.02 on 2 and 67 DF, p-value: 1.493e-12

**Model Comparison**

# Hierarchical regression with model comparison (significance of R squared change)  
# 2 models plus comparison of them for final homework should be presented  
  
# Comparison Model 1  
 # Model B: Test ~ Stress + Comm + Hours  
 # Model A: Test ~ Stress + Comm [ best fit ]  
  
compare1 <- linReg(data = dat,   
 dep = 'Test',   
 covs = c('Stress', 'Comm', 'Hours'),  
 blocks = list(  
 list('Stress', 'Comm'), #Model A  
 list('Hours')), #Model B  
 modelTest = TRUE,  
 r2Adj = TRUE,  
 stdEst = TRUE,  
 ciStdEst = TRUE)  
compare1

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.746 0.556 0.543 42.0 2 67 < .001   
## 2 0.754 0.568 0.549 29.0 3 66 < .001   
## --------------------------------------------------------------------------   
##   
##   
## Model Comparisons   
## ----------------------------------------------------------------   
## Model Model <U+0394>R² F df1 df2 p   
## ----------------------------------------------------------------   
## 1 - 2 0.0120 1.84 1 66 0.179   
## ----------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## --------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## --------------------------------------------------------------------------------------------   
## Intercept 78.46 3.335 23.53 < .001   
## Stress -3.74 0.571 -6.55 < .001 -0.570 -0.744 -0.397   
## Comm 1.33 0.362 3.67 < .001 0.319 0.145 0.493   
## --------------------------------------------------------------------------------------------   
##   
##   
## MODEL 2  
##   
## Model Coefficients   
## ---------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## ---------------------------------------------------------------------------------------------   
## Intercept 75.848 3.833 19.79 < .001   
## Stress -3.733 0.567 -6.58 < .001 -0.569 -0.7422 -0.397   
## Comm 1.182 0.375 3.15 0.002 0.284 0.1041 0.464   
## Hours 0.502 0.370 1.36 0.179 0.115 -0.0543 0.285   
## ---------------------------------------------------------------------------------------------

# ALTERNATIVE  
stats::anova(modelB.1, modelA.1)

## Analysis of Variance Table  
##   
## Model 1: Test ~ Stress + Comm  
## Model 2: Test ~ Stress + Comm  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 67 1119.7   
## 2 67 1119.7 0 0

# Both statistical tests yield no significant difference between models B and A

# Comparison Model 2  
 # Model A: Test ~ Stress + Comm [ best fit ]  
 # Model 1: Test ~ Stress  
  
compare2 <- linReg(data = dat,   
 dep = 'Test',   
 covs = c('Stress', 'Comm'),  
 blocks = list(  
 list('Stress'), #Model A  
 list('Comm')), #Model B  
 modelTest = TRUE,  
 r2Adj = TRUE,  
 stdEst = TRUE,  
 ciStdEst = TRUE)  
compare2

##   
## LINEAR REGRESSION  
##   
## Model Fit Measures   
## --------------------------------------------------------------------------   
## Model R R² Adjusted R² F df1 df2 p   
## --------------------------------------------------------------------------   
## 1 0.684 0.467 0.460 59.7 1 68 < .001   
## 2 0.746 0.556 0.543 42.0 2 67 < .001   
## --------------------------------------------------------------------------   
##   
##   
## Model Comparisons   
## -----------------------------------------------------------------   
## Model Model <U+0394>R² F df1 df2 p   
## -----------------------------------------------------------------   
## 1 - 2 0.0890 13.4 1 67 < .001   
## -----------------------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## ------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## ------------------------------------------------------------------------------------------   
## Intercept 88.26 2.169 40.70 < .001   
## Stress -4.48 0.580 -7.72 < .001 -0.684   
## ------------------------------------------------------------------------------------------   
##   
##   
## MODEL 2  
##   
## Model Coefficients   
## --------------------------------------------------------------------------------------------   
## Predictor Estimate SE t p Stand. Estimate Lower Upper   
## --------------------------------------------------------------------------------------------   
## Intercept 78.46 3.335 23.53 < .001   
## Stress -3.74 0.571 -6.55 < .001 -0.570 -0.744 -0.397   
## Comm 1.33 0.362 3.67 < .001 0.319 0.145 0.493   
## --------------------------------------------------------------------------------------------

# ALTERNATIVE  
stats::anova(modelA.1, model1.1)

## Analysis of Variance Table  
##   
## Model 1: Test ~ Stress + Comm  
## Model 2: Test ~ Stress  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 67 1119.7   
## 2 68 1344.4 -1 -224.66 13.443 0.0004878 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Both statistical tests yield a significant difference between models A and 1

**Best Model with Centered Data**

#Predicted score on Y when all predictors are averaged vs   
 #{uncentered} predicted score on Y when all predictors are zero.  
  
#Stress  
dat$Stress.C <- dat$Stress - mean(dat$Stress)  
   
#Comm  
dat$Comm.C <- dat$Comm - mean(dat$Comm)  
  
modelA.2<- lm(Test ~ Stress.C + Comm.C, data = dat)  
summary(modelA.2)

##   
## Call:  
## lm(formula = Test ~ Stress.C + Comm.C, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.6690 -2.5243 -0.3167 2.5740 10.0199   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 72.0186 0.4886 147.393 < 2e-16 \*\*\*  
## Stress.C -3.7385 0.5705 -6.553 9.5e-09 \*\*\*  
## Comm.C 1.3271 0.3620 3.666 0.000488 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.088 on 67 degrees of freedom  
## Multiple R-squared: 0.5564, Adjusted R-squared: 0.5432   
## F-statistic: 42.02 on 2 and 67 DF, p-value: 1.493e-12

**Visualization with Centered Data**

# plotting a multiple regression model based on:   
 # Model A: Test ~ Stress + Comm (from lm command of model created 'modelA.1')  
  
# create predicted values from predictors and save in object  
model\_p <- ggpredict(modelA.2, terms = c('Stress.C', 'Comm.C'), full.data = TRUE, pretty = FALSE)  
  
# plot predicted line  
plot <- ggplot(model\_p, aes(x, predicted)) +  
 geom\_smooth(method = "lm", se = FALSE, fullrange=TRUE) + xlab("Score") + ggtitle("Plot of Model of Stress and Commitment Predicting Test Score") + ylab("Test Score") +  
 geom\_point() + theme\_minimal()  
  
plot

