Example5_1

Kevin Cummiskey

10/29/2019

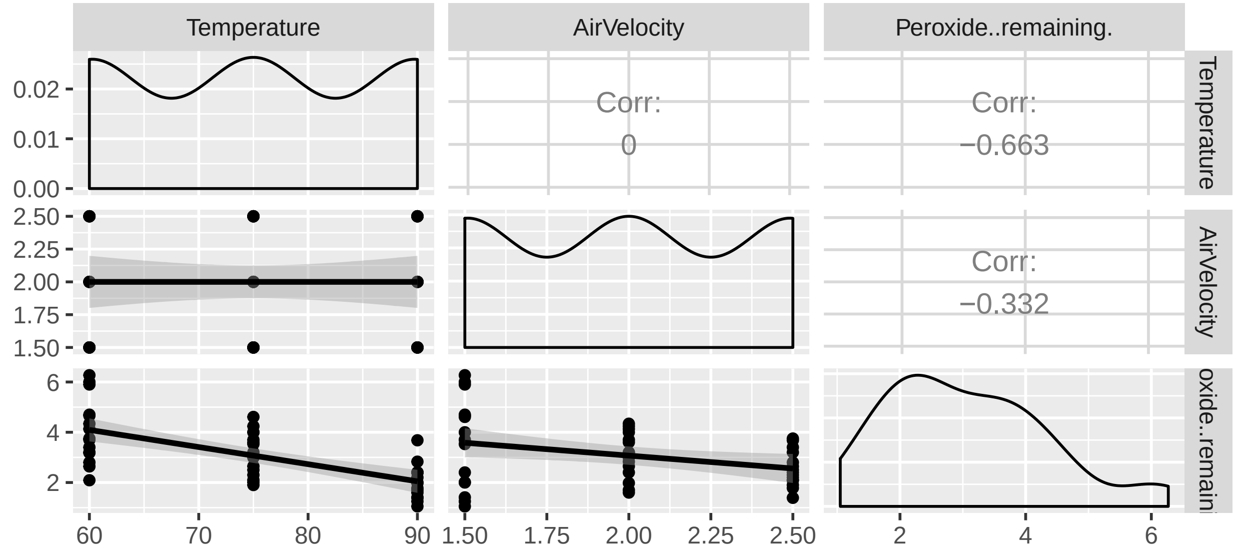
Review

You are interested in whether the linear association between body weight and IOCT time differs by cadet class. Given data, describe how you would perform an appropriate test. Please be specific on what models you would fit and the test statistic you would use.

Multiple Quantitative Explanatory Variables

Data Analysis

```r
nuts = read.table(file = "http://www.isi-stats.com/isi2/data/pistachioStudy.txt",
                  header = T)
library(GGally)
library(plotly)
nuts %>% select(Temperature, AirVelocity, Peroxide..remaining.) %>% ggpairs(lower = list(continuous = "
```



```r
#nuts %>% plot_ly(x = ~Temperature, y = ~AirVelocity, z = ~Peroxide..remaining.)
```

Main Effects Model

We fit the following model:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2)$$

where y_i is the peroxide remaining (%), $x_{1,i}$ is the Temperature (F), and $x_{2,i}$ is the AirVelocity (mph).

How do we interpret the coefficients in the model?

```
model_TempVel = lm(Peroxide..remaining. ~ Temperature + AirVelocity, data = nuts)
summary(model_TempVel)
```

```
##
## Call:
## lm(formula = Peroxide..remaining. ~ Temperature + AirVelocity,
##     data = nuts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57222 -0.45822 -0.05822  0.52978  2.14478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.23322    1.03126   9.923 1.41e-12 ***
## Temperature -0.06820    0.01065  -6.403 1.04e-07 ***
## AirVelocity -1.02400    0.31952  -3.205  0.00258 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8751 on 42 degrees of freedom
## Multiple R-squared:  0.5497, Adjusted R-squared:  0.5283
## F-statistic: 25.64 on 2 and 42 DF,  p-value: 5.289e-08
```

What do we conclude from these results?

How do the coefficients compare to the single variable models? Why?

```
model_Temp = lm(Peroxide..remaining. ~ Temperature, data = nuts)
summary(model_Temp)
```

```
##
## Call:
## lm(formula = Peroxide..remaining. ~ Temperature, data = nuts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00322 -0.69322 -0.06722  0.56678  2.17678
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.18522    0.89239   9.172 1.11e-11 ***
## Temperature -0.06820    0.01174  -5.808 6.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.9648 on 43 degrees of freedom
## Multiple R-squared:  0.4396, Adjusted R-squared:  0.4266
## F-statistic: 33.73 on 1 and 43 DF,  p-value: 6.957e-07
```

```
model_Vel = lm(Peroxide..remaining. ~ AirVelocity, data = nuts)
summary(model_Vel)
```

```
##
## Call:
## lm(formula = Peroxide..remaining. ~ AirVelocity, data = nuts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.53222 -0.67022 -0.05222  0.92978  2.68778
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.1182     0.9062   5.648 1.19e-06 ***
## AirVelocity  -1.0240     0.4439  -2.307    0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.216 on 43 degrees of freedom
## Multiple R-squared:  0.1101, Adjusted R-squared:  0.08942
## F-statistic: 5.321 on 1 and 43 DF,  p-value: 0.02596
```

Which variable is more important in explaining peroxide remaining?

How do we standardize a varible? Why do we standardize variables?

Model with standardized variables.

```
#In our data set, the variables are already standardized.
#If you needed to do it, you could use the scale function
nuts = nuts %>% mutate(temp.std = scale(Temperature))

model_std = lm(Peroxide..remaining. ~ std.temp + std.air, data = nuts)
summary(model_std)
```

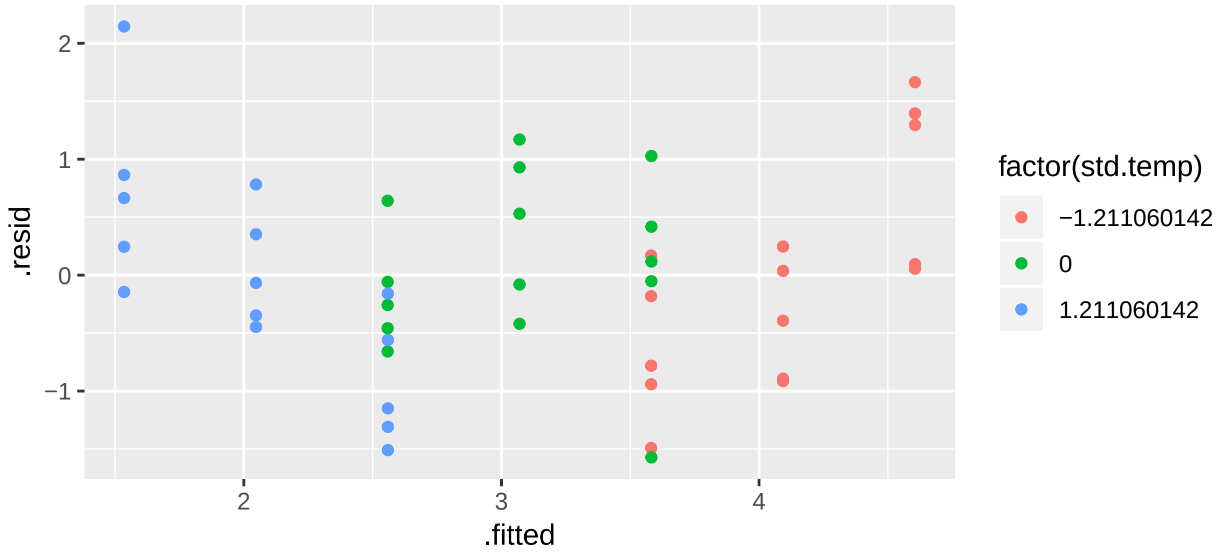
```
##
## Call:
## lm(formula = Peroxide..remaining. ~ std.temp + std.air, data = nuts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57222 -0.45822 -0.05822  0.52978  2.14478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0702     0.1304  23.537  < 2e-16 ***
## std.temp     -0.8447     0.1319  -6.403 1.04e-07 ***
## std.air      -0.4228     0.1319  -3.205  0.00258 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8751 on 42 degrees of freedom
## Multiple R-squared:  0.5497, Adjusted R-squared:  0.5283
## F-statistic: 25.64 on 2 and 42 DF,  p-value: 5.289e-08
```

How have the coefficients changed? p-values? How do we interpret the intercept?

Are the validity conditions met?

Residuals vs Predicted Values

```
model_std %>%
  fortify() %>%
  ggplot(aes(x = .fitted, y = .resid, color = factor(std.temp))) +
  geom_point()
```



Model with Interactions