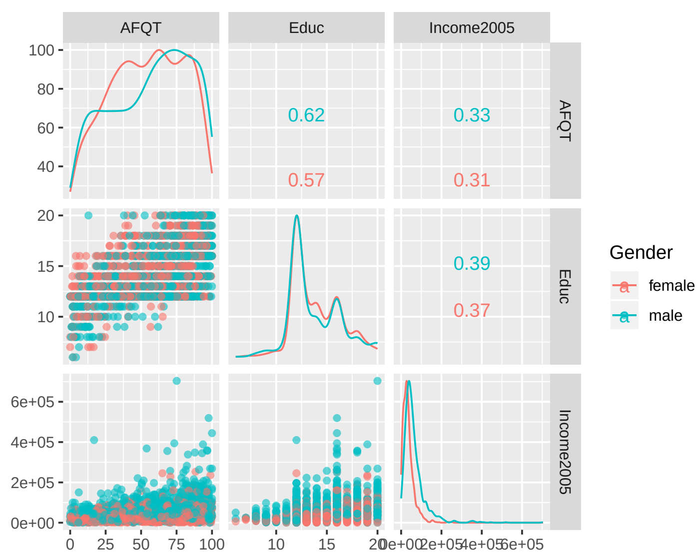
Homework #3

Dustin Leatherman

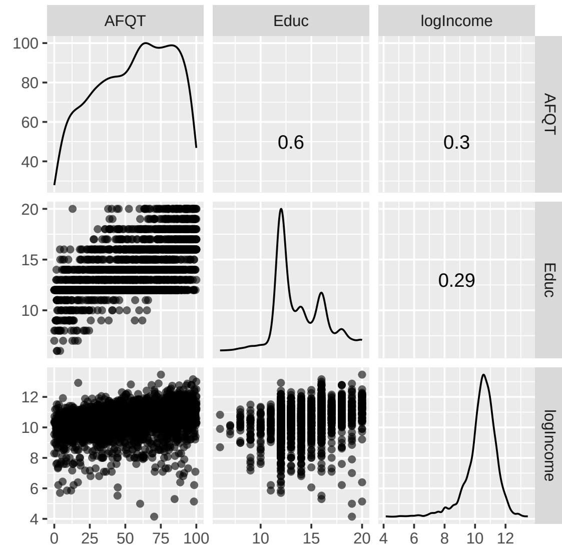
February 3, 2019

Income, Intelligence & Gender

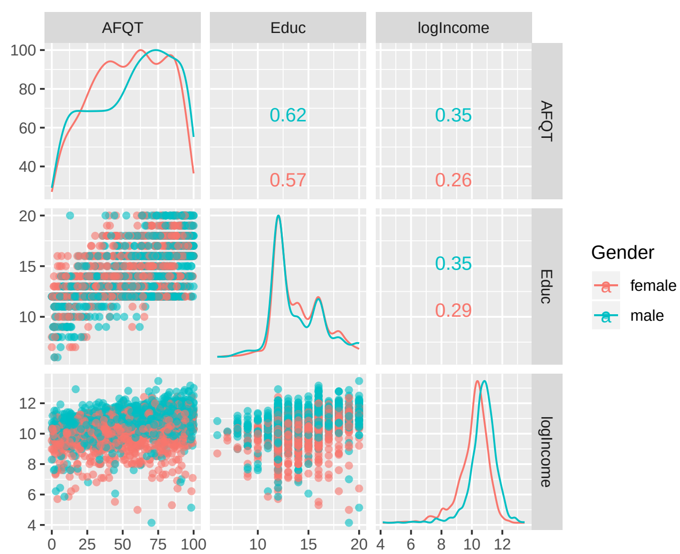
```r
# Common transformations on the dataset
data <-
  ex0923 %>%
  mutate(logIncome = log(Income2005))

# Scatterplot matrix for variables of interest grouped by Gender. No Transformations appplied
data %>%
  select("AFQT", "Educ", "Income2005", "Gender") %>%
  ggscatmat(color="Gender", alpha = 0.6)
```




```r
# Scatterplot matrix for variables of interest. I found that having gender separated made scatterplots 
data %>%
  select("AFQT", "Educ", "logIncome", "Gender") %>%
  ggscatmat(alpha = 0.6)
```




```r
# Scatterplot for variables of interest grouped by Gender. Transformations applied
data %>%
  select("AFQT", "Educ", "logIncome", "Gender") %>%
  ggscatmat(color="Gender", alpha = 0.6)
```



The scatterplot for Income has a funneling shape which indicates that it is skewed. This is confirmed by its density plot. Income tends to be right skewed so a log transformation has been applied to make it more normal. The density plots in the scatter matrix show the distribution of each variable categorized by Gender. The distribution of AFQT is left skewed for males while it is normalish looking for females. After applying a log transformation, logIncome appears to be left skewed but the distribution appears more normal with the peak logIncome for males being larger. There are two peaks in the Education Density graph, one at 12 years and 16 years. This indicates that a majority of the respondents had either a high school or bachelors degree. Besides Income, no other fields need to transformations applied.

Log-Income & Intelligence

Log-Income and AFQT Score have a positive correlation. The scatterplot indicates a linear increase in Log-Income as the AFQT score increases. There are many outliers as indicated by the by the dots that fall below the main cluster. It is interesting to note that a majority of the outliers that fall below the cluster appear to be female.

Log-Income & Education

Log-Income and Education have a positive correlation with Log-Income generally increasing for each year of education. It is interesting to note that the spread of Log-Income is largest at year 12 and decreases steadily after. There are some outliers for people with more years of education. For almost all years of education, females have generally have less Log-Income than males. The exception to this is year 19 where it appears fairly even between the two genders.

Intelligence & Education

AFQT Scores and Years of Education have a positive correlation. For every year of education, AFQT Scores generally increase. There is a small number of outliers who have higher AFQT scores with 10 or less years of education as well as a single outlier who has 20 years of education and a score of approximately 13.

Model

$\mu\{\text{logIncome}|\text{AFQT,Educ,Gender}\} = \beta_0 + \beta_1\text{MALE} + \beta_2\text{AFQT} + \beta_3\text{Educ}$

```
model <- lm(logIncome ~ Educ + AFQT + Gender, data=data, na.action = na.exclude)
summary(model)
```

```
##
## Call:
## lm(formula = logIncome ~ Educ + AFQT + Gender, data = data, na.action = na.exclude)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0906 -0.3301  0.1404  0.5091  2.5452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.7312115  0.1026287  85.076  < 2e-16 ***
## Educ        0.0769506  0.0084888   9.065  < 2e-16 ***
## AFQT        0.0059139  0.0007657   7.724  1.6e-14 ***
## Gendermale  0.6245093  0.0341748  18.274  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8661 on 2580 degrees of freedom
## Multiple R-squared:  0.2101, Adjusted R-squared:  0.2092
## F-statistic: 228.7 on 3 and 2580 DF,  p-value: < 2.2e-16
```

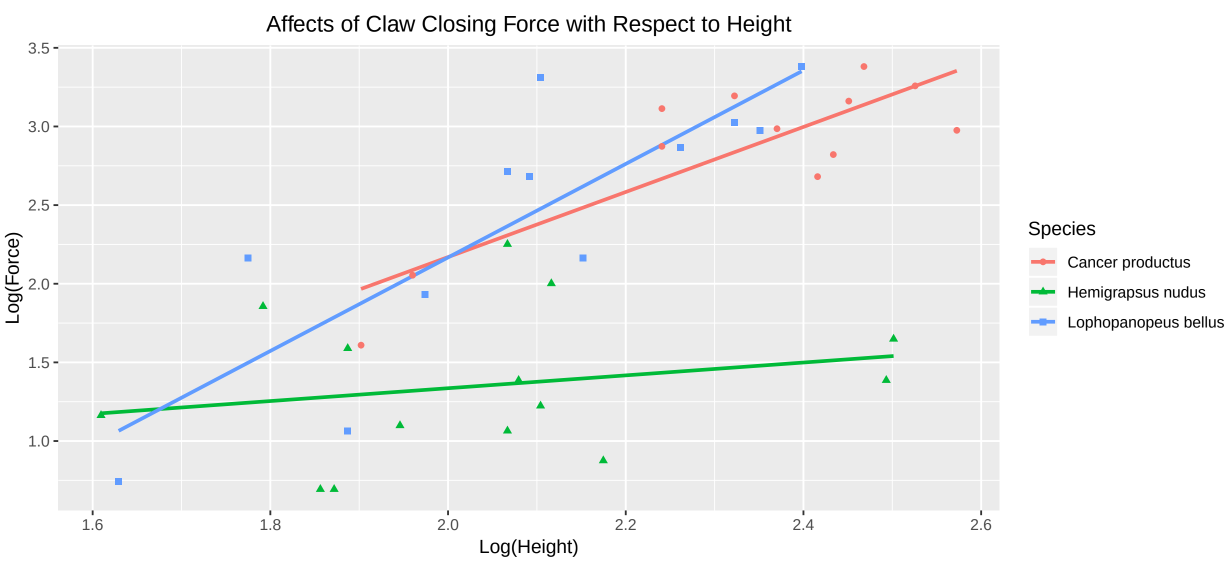
```
confint(model)
```

```
##                    2.5 %      97.5 %
## (Intercept) 8.529968655 8.932454398
## Educ        0.060305047 0.093596160
## AFQT        0.004412575 0.007415313
## Gendermale  0.557496441 0.691522169
```

There convincing evidence that being being male has a positive multiplicative effect on median income (Sum of Squares F-Test. p-value $< 2.2e\text{-}16$). The estimated median income of a male is $\exp(0.6245) = 1.8673$ times the median income of a female after accounting for Education and AFQT Score. With 95% confidence, the median income of a male is between $\exp(0.5575) = 1.7463$ and $\exp(0.6915) = 1.9967$ times the median income of a female after accounting for Education and AFQT Score.

Crab Force

```
ggplot(ex0722, aes(y = log(Force), x = log(Height), color = Species, shape = Species)) +
  geom_smooth(method = "lm",se = FALSE) +
  geom_point() +
    xlab("Log(Height)") +
    ylab("Log(Force)") +
    ggtitle("Affects of Claw Closing Force with Respect to Height") +
      theme(plot.title = element_text(hjust = 0.5))
```



$\mu\{\text{Log(Force)}|\text{Log(Height)}, \text{Species}\} = \beta_0 + \beta_1\text{log(Height)} + \beta_2\text{NUDUS} + \beta_3\text{BELLUS}+ \beta_4(\text{Log(Height)} \times \text{NUDUS}) + \beta_5(\text{Log(Height)} \times \text{BELLUS})$

```
ex0722$Species2 <- relevel(ex0722$Species, ref = "Lophopanopeus bellus")

# Interaction terms make a parallel lines model
model <- lm(log(Force) ~ log(Height) + Species + log(Height):Species, data = ex0722)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = log(Force) ~ log(Height) + Species + log(Height):Species,
##     data = ex0722)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76675 -0.28506 -0.02306  0.24249  0.88822
##
## Coefficients:
##                                         Estimate Std. Error t value
## (Intercept)                              -1.9673     1.4489  -1.358
## log(Height)                               2.0685     0.6208   3.332
## SpeciesHemigrapsus nudus                  2.4864     1.7606   1.412
## SpeciesLophopanopeus bellus              -1.8128     1.8533  -0.978
## log(Height):SpeciesHemigrapsus nudus     -1.6601     0.7889  -2.104
## log(Height):SpeciesLophopanopeus bellus   0.9052     0.8302   1.090
##                                         Pr(>|t|)
## (Intercept)                              0.18405
## log(Height)                              0.00219 **
## SpeciesHemigrapsus nudus                 0.16752
## SpeciesLophopanopeus bellus              0.33536
## log(Height):SpeciesHemigrapsus nudus     0.04330 *
## log(Height):SpeciesLophopanopeus bellus  0.28368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4329 on 32 degrees of freedom
## Multiple R-squared:  0.7945, Adjusted R-squared:  0.7624
## F-statistic: 24.75 on 5 and 32 DF,  p-value: 3.935e-10
```

```r
confint(model)
```

```
##                                              2.5 %      97.5 %
## (Intercept)                             -4.9186614  0.98411449
## log(Height)                              0.8039135  3.33298860
## SpeciesHemigrapsus nudus                -1.0997187  6.07255364
## SpeciesLophopanopeus bellus             -5.5878604  1.96235631
## log(Height):SpeciesHemigrapsus nudus    -3.2671495 -0.05312693
## log(Height):SpeciesLophopanopeus bellus -0.7857653  2.59616597
```

```r
# Change base comparison to use bellus instead for easier comparisons between Nudus and Bellus
model.rebase <- lm(log(Force) ~ log(Height) + Species2 + log(Height):Species2, data = ex0722)
summary(model.rebase)
```

```
##
## Call:
## lm(formula = log(Force) ~ log(Height) + Species2 + log(Height):Species2,
##     data = ex0722)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76675 -0.28506 -0.02306  0.24249  0.88822
##
```

```
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          -3.7800     1.1556  -3.271  0.00257
## log(Height)                           2.9737     0.5511   5.395 6.29e-06
## Species2Cancer productus              1.8128     1.8533   0.978  0.33536
## Species2Hemigrapsus nudus             4.2992     1.5283   2.813  0.00832
## log(Height):Species2Cancer productus  -0.9052     0.8302  -1.090  0.28368
## log(Height):Species2Hemigrapsus nudus -2.5653     0.7354  -3.488  0.00144
##
## (Intercept)                          **
## log(Height)                          ***
## Species2Cancer productus
## Species2Hemigrapsus nudus            **
## log(Height):Species2Cancer productus
## log(Height):Species2Hemigrapsus nudus **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4329 on 32 degrees of freedom
## Multiple R-squared:  0.7945, Adjusted R-squared:  0.7624
## F-statistic: 24.75 on 5 and 32 DF,  p-value: 3.935e-10
```

```r
confint(model.rebase)
```

```
##                                           2.5 %     97.5 %
## (Intercept)                           -6.133906 -1.4261452
## log(Height)                            1.851016  4.0962864
## Species2Cancer productus              -1.962356  5.5878604
## Species2Hemigrapsus nudus              1.186223  7.4121158
## log(Height):Species2Cancer productus  -2.596166  0.7857653
## log(Height):Species2Hemigrapsus nudus -4.063250 -1.0674268
```

Comparing NUDUS to BELLUS

β_0 + $\beta_1$log(Height) + β_2PRODUCTUS + β_3NUDUS + β_4(Log(Height) $\times$ PRODUCTUS) + β_5(Log(Height) $\times$ NUDUS)

$\rightarrow$ β_3 + β_5Log(Height)

There is convincing evidence that Hemigraspus Nudus has a weaker claw strength than Lophopanopeus Bellus after accounting for Log(Height) (p-value = 0.00144). It is estimated that Hemigraspus Nudus has an average weaker claw force than Lophopanopeus Bellus by 2.5653 Log-Force units. With 95% confidence, Hemigraspus Nudus mean claw strength is between 1.0674 Log-Force units and 4.0633 Log-Force units weaker than Lophopanopeus Bellus after accounting for Log(Height).

Comparing NUDUS to PRODUCTUS

β_0 + $\beta_1$log(Height) + β_2NUDUS + β_4Log(Height) - (β_0 + $\beta_1$log(Height))

$\rightarrow$ β_2 + β_4Log(Height)

There is moderate evidence that Hemigraspus Nudus has weaker claw strength than Cancer Productus after accounting for Log(Height) (p-value = 0.00144). It is estimated that Hemigraspus Nudus is an average of 1.6601 Log(Force) units weaker than Cancer Productus. With 95% confidence, Hemigraspus Nudus mean claw strength is between 0.0531 and 3.2671 Log(Force) units weaker than Cancer Productus after accounting for Log(Height).

Comparing BELLUS to PRODUCTUS

β_0 + $\beta_1$log(Height) + β_3BELLUS + β_5Log(Height) - (β_0 + $\beta_1$log(Height))

$\rightarrow$ β_3 + β_5Log(Height)

There is no evidence that Lophopanopeus Bellus has a difference in mean claw strength in Log(Force) units compared to Cancer Productus (p-value = 0.28368).