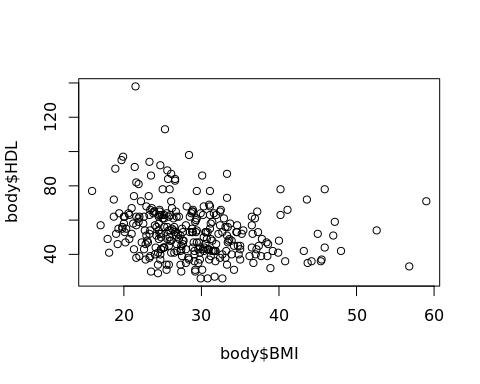
Lab05\_Massey

Matt Massey

6/19/2021

1. **Generate a scatterplot between the HDL and LDL variables.**

# Import your data set   
body <- read.table(file = "Lab5\_bodydata.txt", sep = "\t", header = TRUE)  
  
#You may need this later. This code converts numeric variables to categorial  
body$GENDER..1.M. = as.factor(body$GENDER..1.M.)  
  
# Scatterplot of BMI vs. HDL  
plot(x = body$BMI, y = body$HDL)



*\*\*\*Note: In the video, Dr. Pittard includes code for a regression line here, but the lab instructions do not.*

1. **Does there appear to be a linear association between BMI and HDL?**

Yes, but it is VERY weak and inconclusive with just this initial visual inspection.

1. **Does it appear that the relationship between BMI and HDL is linear? Weak or strong?**

There appears to be a VERY slight decrease in HDL as BMI increases. However, I would need more evidence to conclusively say this negative linear relationship is significant.

# Information from correlation and regression  
# Use Multiple R-squared for correlation  
summary(lm(HDL~BMI, data = body))

##   
## Call:  
## lm(formula = HDL ~ BMI, data = body)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -27.126 -10.214 -2.088 7.219 80.302   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 69.3992 3.8222 18.157 < 2e-16 \*\*\*  
## BMI -0.5443 0.1288 -4.227 3.15e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.02 on 298 degrees of freedom  
## Multiple R-squared: 0.05657, Adjusted R-squared: 0.05341   
## F-statistic: 17.87 on 1 and 298 DF, p-value: 3.15e-05

1. **What is the equation of the regression line?**

The general equation for a sample is – ŷ = b0 + b1x

For this specific sample, the equation is – HDL = 69.3992 – 0.05657(BMI)

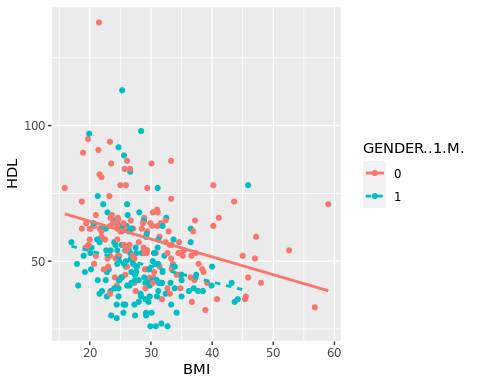
1. **Interpret the slope and intercept in the context of HDL and BMI.**

The y-intercept is 69.3992, so with a BMI of 0, the linear model predicts that HDL should be 69.3992. The slope of the regression line is -0.05657, which means that the linear regression model predicts that if BMI increases by one unit, the predicted HDL value will fall by 0.05657 units. Despite the weak negative linear relationship, we can see that the P-value for the linear regression is very low (3.15 e-5) and suggests there is sufficient evidence to reject the null hypothesis that the slope is 0. We can also take the square root of r2 to get the correlation coefficient, r, equal to -0.238, which is a very weak correlation between BMI and HDL.

1. **If the measurements are separated according to gender, do the scatterplots, correlation, or regression model change?**

# Plot separating male and female with corresponding regression lines. 0 = female, 1 = male:  
library(ggplot2)  
ggplot(body, aes(x=BMI, y=HDL, color = GENDER..1.M., lty=GENDER..1.M.)) + geom\_point() +   
 geom\_smooth(method="lm", se=FALSE)

## `geom\_smooth()` using formula 'y ~ x'



# Regression model using gender as a covariate  
# 0 = female, 1 = male  
# Using HDL as the response variable  
summary(lm(HDL~BMI + body$GENDER..1.M., data = body))

##   
## Call:  
## lm(formula = HDL ~ BMI + body$GENDER..1.M., data = body)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.406 -9.333 -2.384 5.403 74.459   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 77.0958 3.8105 20.233 < 2e-16 \*\*\*  
## BMI -0.6305 0.1221 -5.162 4.49e-07 \*\*\*  
## body$GENDER..1.M.1 -10.2039 1.6458 -6.200 1.89e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.16 on 297 degrees of freedom  
## Multiple R-squared: 0.1647, Adjusted R-squared: 0.1591   
## F-statistic: 29.28 on 2 and 297 DF, p-value: 2.482e-12

Above we see a scatterplot with the data separated by gender, where magenta is female (code=0) and cyan is male (code=1). The scatterplot also shows a multivariable linear regression model for male and female that results in two lines colored in the same manner. Visual inspection of the scatterplot and regression model show that females (magenta) have generally higher HDL compared to men, although both appear to have a similar negative relationship with BMI. The regression model also shows this by showing that men are significantly lower than women by almost 10 HDL units (-10.2039); the corresponding P-value for this difference is 1.89e-9 and suggests that this difference is significantly different.

1. **Write a brief conclusion on how gender affects or does not affect HDL.**

Females tend to have higher HDL cholesterol than males, although HDL decreases in both genders as BMI increases. It is not clear from this dataset what causes this difference in HDL between male and female, but it is significantly different.