Chelsey Kim

Analysis of the association between different variables in MIDUS and depression scores

Introduction: Depression is a common mood disorder that many adults encounter. Knowing what increases the risk of depression is important to study so measures can be taken to prevent this risk. The research questions being studied include the following. What is the association between depression and how healthy an individual perceives himself/herself to be? Health is an indicator of wellbeing; therefore, depressed individuals will have lower ratings of health. What is the association between depression and how often an individual laughs within a month? Depressed individuals tend to overlook humor; therefore, depressed individuals will laugh less often. What is the association between depression and household income? There will be no distinct relationship because money cannot buy happiness. What is the association between depression and how often an individual faces discrimination? Discrimination makes individuals feel inferior, making them more likely to become depressed.

Methods: The data set, MacArthur Study of Successful Midlife Development (MIDUS) in the US, surveys adults on factors related to health. I analyzed the Center for Epidemiological Studies Depression scores (CES-D), how often an individual laughed within the past month, self-rated health, household income in USD, and daily discrimination. The CES-D uses a screening test for depression, a score 16 or above indicating depression. The CES-D was recoded to remove missing values and categorized scores under 16 as "Not Depressed" and scores above as "Depressed". How often an individual spent laughing in the past month was recoded to remove missing values. Fishers test was used because the expected cell count was less than 5. Self-rated health was recoded so values of 1, 2, or 3 were classified as "Healthy" and values of 4 or 5 were classified as "Unhealthy". A two-sample t-test was used to compare the means of depression scores for the two categories. Household income was recoded to remove missing data. Income's of equal or less than $28,700, $28,700 to $86,100, and equal or more than $86,100 were classified as low, middle, and high incomes, respectively. ANOVA testing was used to compare the means of depression score for the three income groups. Tukey test was used to determine which specific income groups were significant. Daily discrimination is a score determined from the self-reported amount of discriminatory experiences experienced and was recoded to remove missing data. A linear regression model was used to see the correlation between the depression score and discrimination scores.

Results: Each test statistic was compared to an 0.05 alpha level of significance. Whether an individual is depressed and how often they laughed within the month is statistically significant ($p=<.001$, see table 2 in Appendix A). Healthy individuals have statistically lower depression scores than unhealthy individuals ($t=-6.141$, $p=<.001$). There is a statistical significance between household income and depression scores ($p=0.013$), specifically the high- and low-income comparison ($p=0.006$, see table 3 in Appendix A). The more discrimination an individual faces daily, the statistically higher the depression score. The linear regression equation is $y = 1.09 + 0.544x$ ($p=<.001$, see table 4 in Appendix A).

Discussion: The association between depression and the frequency of laughter supports my hypothesis. Negative thoughts consume depressed individuals, making it hard for them to laugh at small joys. The association between depression and health supports my hypothesis. If depressed, individuals typically are less likely to take care of themselves because they may have trouble sleeping, eating, or simply do not care enough to take care of themselves. The association between depression and income supports my hypothesis to an extent. There is no association between the Middle-Low income and Middle-High incomes because having more materials does not make individuals happier. However, income increases happiness when it is a matter of providing basic human needs. The association between depression and daily discrimination supports my hypothesis. Individuals that are discriminated against become unhappy with the things they cannot change about themselves, making them more likely to be unhappy with themselves as a whole. However, the correlation between the two is weak (0.303) and the residuals did not follow linearity, so the findings are not practically significant. There are several limitations to this study. Health, frequency of laughter, and daily discrimination are self-reported measures; therefore, different individuals may have different rating scales. Not all observations may be independent, because siblings were included in the study. Income classification was solely based on yearly income and did not consider any other variables that may affect class. In conclusion, knowing that these variables have an association with depression will allow individuals to be conscious of these indicators, so that they can seek help for themselves or others if needed.

Appendix A: Tables with descriptive statistics, tests of association, and p-values

Table 1. Descriptive statistics and tests of association between the Center for Epidemiological Studies Depression Inventory, Self-Rated Health, "Over the past month how often did you spend time laughing?", Household Income, and Daily Discrimination in the MacArthur study of Successful Midlife Development (MIDUS). N= 973

	Overall N= 973	p-value
Center for Epidemiological Studies Depression Inventory (CES-D)	8.020 (7.764)	
Laughter		
Never	8 (0.822%)	<.001
1-6 times	207 (21.274%)	
7+ times	758 (77.903%)	
Self-Rated Health		<.001
Healthy	886 (91.059%)	
Unhealthy	87 (8.941%)	
Household Income (USD)		0.013
Low	177 (18.191%)	
Middle	468 (48.099%)	
High	328 (33.710%)	
Daily Discrimination	12.727 (4.328)	<.001

Table 2. Descriptive statistics and Fisher's Test between the Center for Epidemiological Studies Depression Inventory ≥ 16 and the amount of laughter over the past month in the MacArthur study of Successful Midlife Development (MIDUS). N=973

Overall	**Center for Epidemiological Studies Depression Inventory (CES-D)**		**P-Value**
	Not Depressed (<16)	**Depressed (≥ 16)**	
N=973	911 (86.762%)	139 (13.238%)	
Laughter			<.001
Never	3 (0.286%)	5(0.476%)	
1-6 times	161 (15.333%)	64 (6.095%)	
7+ times	747(71.142%)	70 (6.666%)	

Table 3. Descriptive statistics and Tukey test between the Center for Epidemiological Studies Depression Inventory scores and Household Income in USD in the MacArthur study of Successful Midlife Development (MIDUS). N=973

Household Income Comparison	Difference	95% Confidence Interval		P-Value
		Lower	Upper	
Middle-Low	-1.290	-2.806	0.284	0.135
High-Low	-2.134	-3.803	-0.527	0.006
High-Middle	-0.844	-2.168	0.360	0.214

Table 4. Linear regression results of Center for Epidemiological Studies Depression Inventory scores regressed on Daily Discrimination in the MacArthur Study of Successful Midlife Development (MIDUS). N= 973

	Center for Epidemiological Studies Depression Inventory β	95% Confidence Interval		p-value
		Lower	Upper	
Daily Discrimination	0.544	0.436	0.652	<0.001
Constant	1.09	-0.350	2.544	0.137
Observations	973			
R^2	0.092			
F Statistic	98.32			<0.001
Correlation	0.303	0.245	0.359	<0.001

Appendix B: Bi-variate Plots of the Center for Epidemiological Studies Depression Inventory, Self-Rated Health, "Over the past month how often did you spend time laughing?", Household Income, and Daily Discrimination in the MacArthur study of Successful Midlife Development (MIDUS). N= 973.

Figure 1. Center for Epidemiological Studies Depression Inventory scores ≥ 16 and Center for Epidemiological Studies Depression Inventory scores < 16 are associated with laughter over the past month. N=973.

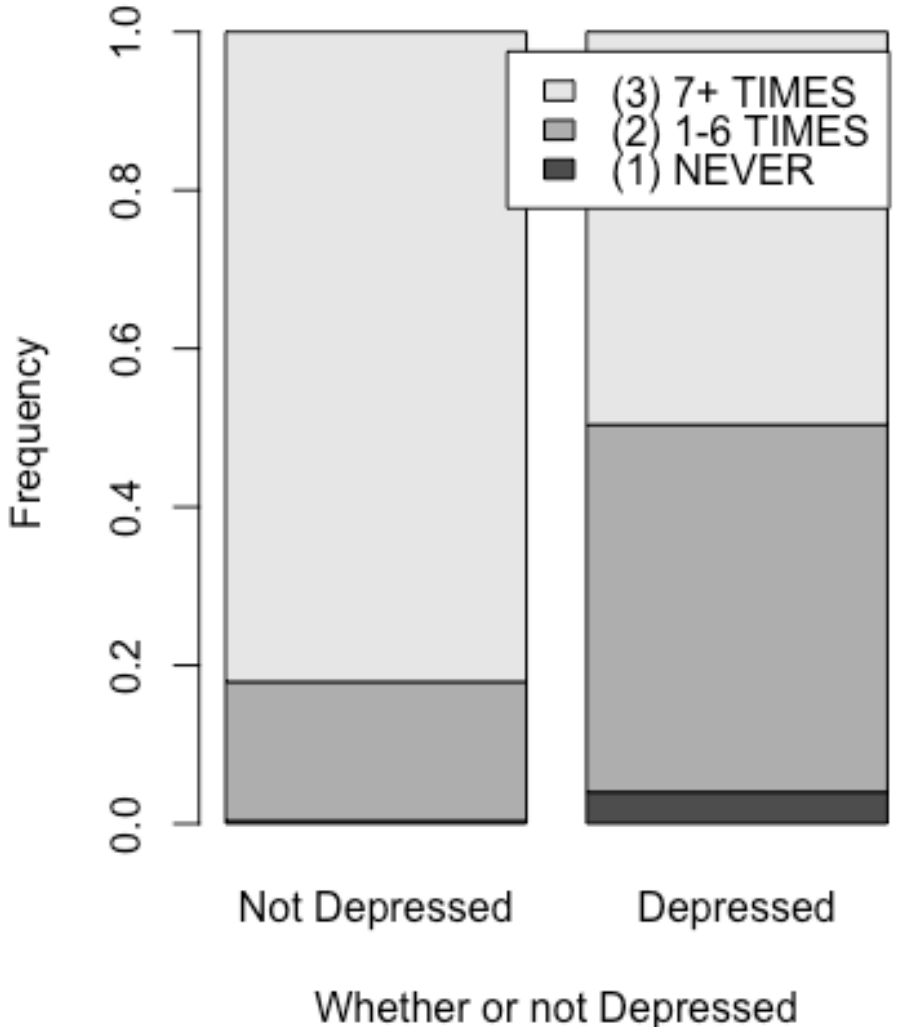


Figure 2. Center for Epidemiological Studies Depression Inventory scores are associated with Self-Rated Health . N=973.

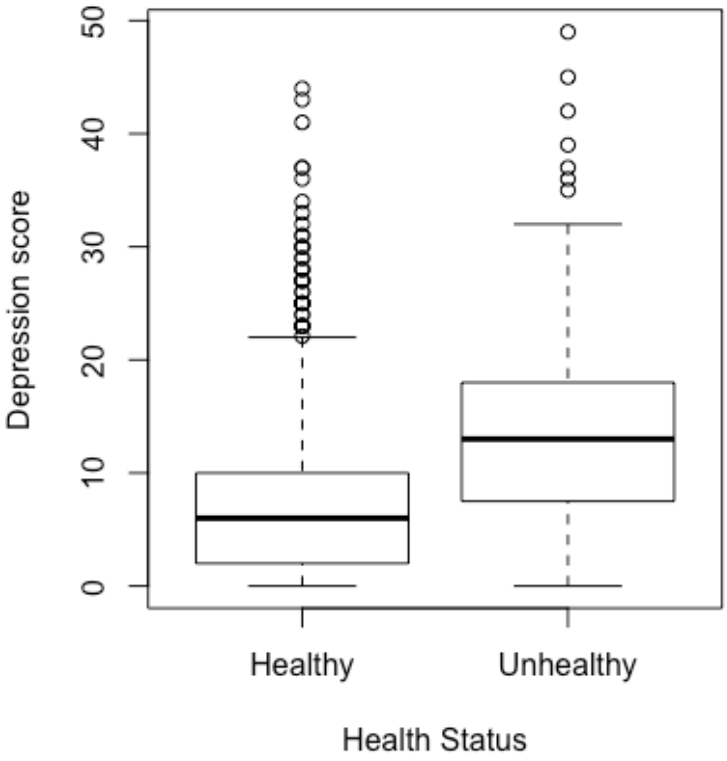


Figure 3. Center for Epidemiological Studies Depression Inventory scores are associated with Household Income in U.S. dollars. N=973.

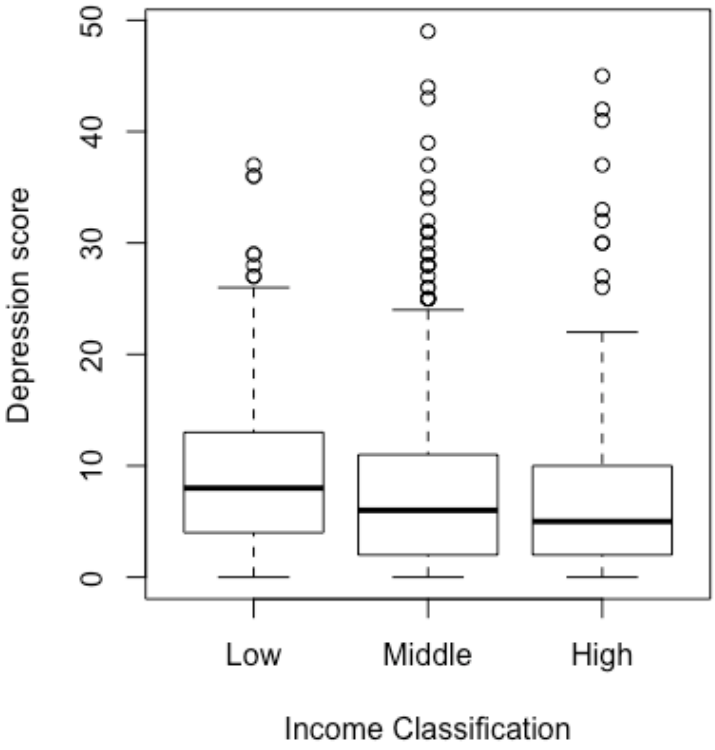


Figure 4. Center for Epidemiological Studies Depression Inventory scores and Household Income in U.S. dollars. N=973.

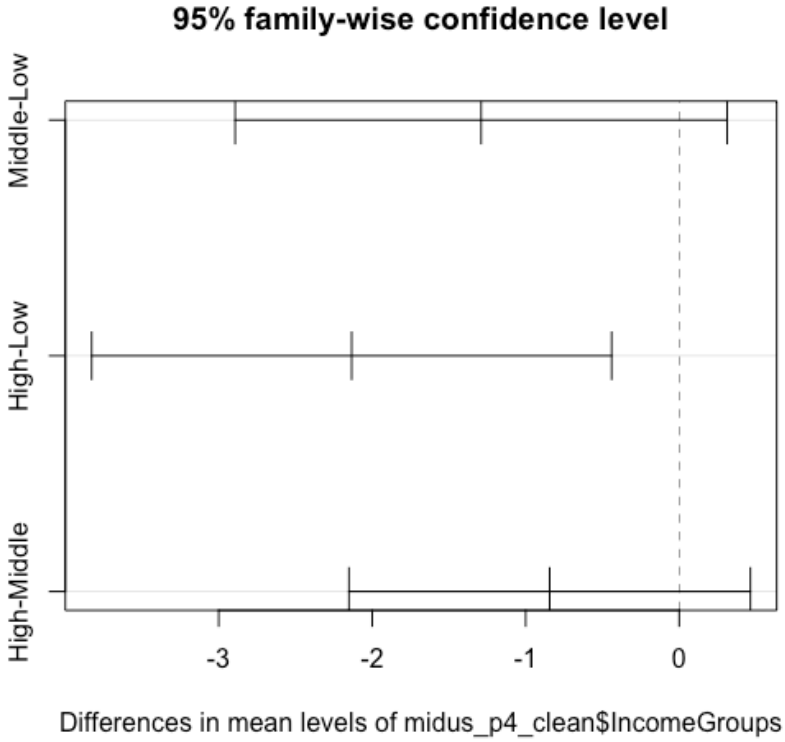
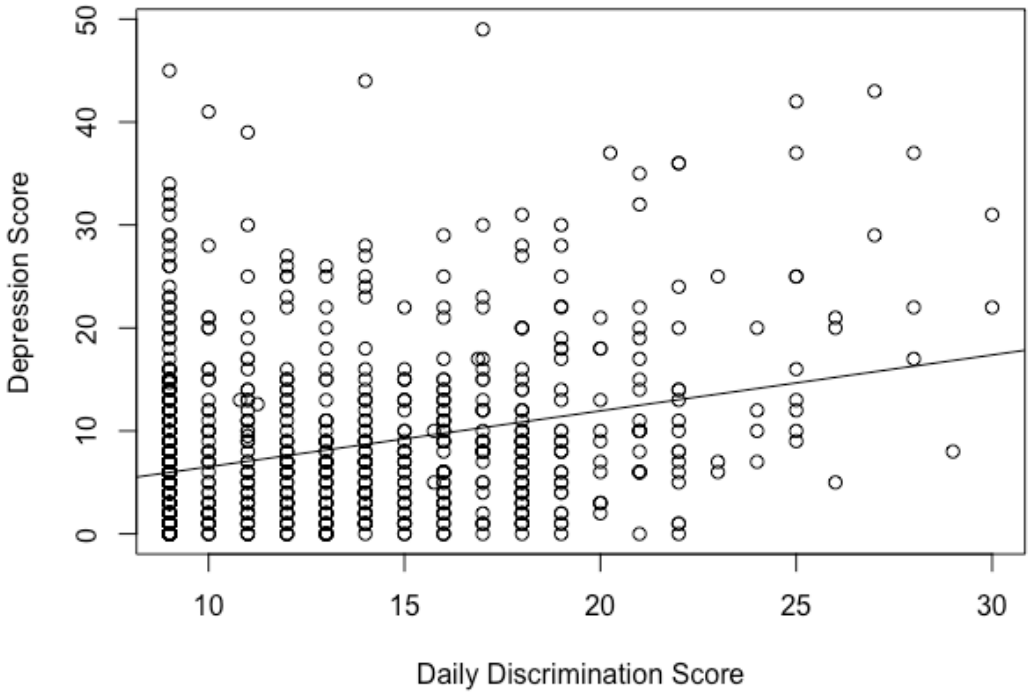


Figure 5. Center for Epidemiological Studies Depression Inventory scores and Daily Discrimination scores. N=973.



Appendix C: R Code

```
#Chelsey Kim
#December 9th, 2018
#Lab Section 4
#Final Project Submission
##############################

#Title: Analysis of the association between different variables in MIDUS and depression scores

#Setting working directory
setwd("/Users/chelseykim/Documents/QTM")
getwd()
#Getting the data set
midus_p4 <- read.csv('midus_p4.csv')

#Cleaning up the data on depression inventory
midus_p4$Depression<-midus_p4$B4QCESD
midus_p4[midus_p4$Depression=="NA",]
midus_p4$Depression[midus_p4$B4QCESD==NA] <-NA
summary(midus_p4$Depression)

#Cleaning up the data for the amount of times laughing
midus_p4$Laugh <- midus_p4$B4Q10I1
midus_p4[midus_p4$Laugh=="8",]
midus_p4$Laugh[midus_p4$B4Q10I1==8]<-NA
summary(midus_p4$Laugh)

#Cleaning up the data for household income
midus_p4$Income <- midus_p4$B1STINC1
midus_p4[midus_p4$Income=="9999998",]
midus_p4$Income[midus_p4$B1STINC1==9999998]<-NA
summary(midus_p4$Income)

#Cleaning up the data for the daily discrimination
midus_p4$DD <- midus_p4$B1SDAYDI
midus_p4[midus_p4$DD=="98",]
midus_p4$DD[midus_p4$B1SDAYDI==98]<-NA
summary(midus_p4$DD)

#Creating a new health variable
midus_p4$Health<-midus_p4$B1PA1
summary(midus_p4$Health)

#In order to have a consistent N, I am going to clean the data to omit all the NA
midus_p4_clean <- na.omit(midus_p4)

#Creating a dichotomous variable for Depression Inventory
midus_p4_clean$DepressionCategory <- factor(NA, levels=c("Not Depressed", "Depressed"))
midus_p4_clean$DepressionCategory[midus_p4_clean$Depression<16]<-"Not Depressed"
midus_p4_clean$DepressionCategory[midus_p4_clean$Depression>=16]<-"Depressed"

#Creating a dichotomous variable for Self-Rated Health
```

```
midus_p4_clean$HealthCategory <- factor(NA, levels=c("Healthy","Unhealthy"))
midus_p4_clean$HealthCategory [midus_p4_clean$Health=="1" | midus_p4_clean$Health=="2"|
midus_p4_clean$Health=="3"]<- "Healthy"
midus_p4_clean$HealthCategory [midus_p4_clean$Health=="4" | midus_p4_clean$Health=="5"]<-
"Unhealthy"

#Creating a factor variable with three levels for Household Income in USD
midus_p4_clean$IncomeGroups <- factor(NA, levels=c("Low","Middle", "High"))
midus_p4_clean$IncomeGroups[midus_p4_clean$Income<=28700]<-"Low"
midus_p4_clean$IncomeGroups[midus_p4_clean$Income>28700 & midus_p4_clean$Income<86100]<-
"Middle"
midus_p4_clean$IncomeGroups[midus_p4_clean$Income>=86100]<- "High"

#Descriptive statistics for Center for Epidemiological Studies Depression Inventory
mean(midus_p4_clean$Depression)
sd(midus_p4_clean$Depression)

#Descriptive statistics for the amount of laughter in the past month
LaughterTable <- table(midus_p4_clean$Laugh)
LaughterTable
prop.table(LaughterTable)*100

#Descriptive statistics for Self-Rated Health
HealthTable<- table(midus_p4_clean$HealthCategory)
HealthTable
prop.table(HealthTable)*100

#Descriptive statistics for Household Income
IncomeTable <- table(midus_p4_clean$IncomeGroups)
IncomeTable
prop.table(IncomeTable)*100

#Descriptive statistic for the amount of Daily Discrimination
mean(midus_p4_clean$DD)
sd(midus_p4_clean$DD)

#Checking expected cell counts for Chi-Squared test
Depression_Laugh <- table(midus_p4_clean$Laugh, midus_p4_clean$DepressionCategory)
Depression_Laugh
Depression_Laugh_Chi$expected
#Chi-Squared test would not work here because the expected cell count is less than 5

#Fisher's Test to see the relationship between whether or not an individual is depressed and the amount
they laugh within the past month
fisher.test(Depression_Laugh)


#Two Sample T-Test

#Checking assumptions
#Is the data normally distributed?
#Histogram for the relationship between healthy and depression scores
hist(midus_p4_clean$Depression[midus_p4_clean$HealthCategory=="Healthy"],
    main="Relationship Between Healthy individuals and Depression Scores", xlab="Depression Score")
```

```
#Histogram for the relationship between nonhealthy and depression scores
hist(midus_p4_clean$Depression[midus_p4_clean$HealthCategory=="Unhealthy"],
    main= "Relationship Between Unhealthy individuals and Depression Scores", xlab="Depression
Scores")
#Histograms are right skewed, but N>30 so test will go on

#Equal Variance because the standard deviations are relatively close to one another
sd(midus_p4_clean$Depression[midus_p4_clean$HealthCategory=="Healthy"])
sd(midus_p4_clean$Depression[midus_p4_clean$HealthCategory=="Unhealthy"])

#Two sample t-test to see the relationship between whether or not the individual is healthy and their
depression score
t.test(midus_p4_clean$Depression~midus_p4_clean$HealthCategory, var.equal=FALSE)


#ANOVA
boxplot(midus_p4_clean$Depression~midus_p4_clean$IncomeGroups)

#Variance for all the income groups are approximately equal
library(mosaic)
favstats(midus_p4_clean$Depression~midus_p4_clean$IncomeGroups)

#Checking normalization for each income group
par(mfrow=c(1,3), pty="s")
hist(midus_p4_clean$Depression[midus_p4_clean$IncomeGroups=="Low"])
hist(midus_p4_clean$Depression[midus_p4_clean$IncomeGroups=="Middle"])
hist(midus_p4_clean$Depression[midus_p4_clean$IncomeGroups=="High"])
#All of the histograms are right skewed, however N>30. Therefore, ANOVA test will still be run
dev.off()

#ANOVA Testing to see the relationship between individual's income classification and depression score
IncomeANOVA <- aov(midus_p4_clean$Depression~midus_p4_clean$IncomeGroups)
summary(IncomeANOVA)


#Linear Regression
#Checking the correlation between depression scores and daily discrimination scores
plot(midus_p4_clean$Depression~midus_p4_clean$DD, xlab="Daily Discrimination Score",
ylab="Depression Score")

#Estimating the correlation between the depression and daily discrimination
cor(midus_p4_clean$Depression, midus_p4_clean$DD)
#There is a weak, positive correlation between depression and daily discrimination

#Test to determine if the correlation is significantly different from zero
cor.test(midus_p4_clean$Depression, midus_p4_clean$DD)
#The true population correlation is significantly different from zero

#Linear Regression between depression score and daily discrimination score
Depression_DD <- lm(midus_p4_clean$Depression~midus_p4_clean$DD)
summary(Depression_DD)

#Adding a regression line estimated from the relationship between Depression and daily discrimination
abline(Depression_DD)
```

```
#Assessing model assumptions with residuals
rstandard(Depression_DD)
hist(rstandard(Depression_DD))
qqnorm(rstandard(Depression_DD))
qqline(rstandard(Depression_DD))
favstats(rstandard(Depression_DD))
#Does not follow linearity

#Assumptions regarding linear relationship and constant variance
plot(predict(Depression_DD), rstandard(Depression_DD))
abline(h=0, lty=2)
#There is constant variance

#Confidence Interval between depression score and daily discrimination score
confint(Depression_DD)


#Graphs

#Graph to see the relationship between whether or not an individual is depressed and the amount they
laugh within the past month
prop.table(Depression_Laugh,margin=2)
barplot(prop.table(Depression_Laugh,margin=2), xlab="Whether or not Depressed",
ylab="Frequency",legend=rownames(Depression_Laugh))

#Graph to see the relationship between whether or not an individual is healthy and their depression score
boxplot(midus_p4_clean$Depression~midus_p4_clean$HealthCategory, xlab="Health Status",
ylab="Depression score")

#Graph to see the relationship between depression scores and Household income category
boxplot(midus_p4_clean$Depression~midus_p4_clean$IncomeGroups, xlab="Income Classification",
ylab="Depression score")

#Graph to see the confidence interval comparing different income groups with their depression scores
TukeyHSD(IncomeANOVA)
plot(TukeyHSD(IncomeANOVA))
```