#Final Project #Author: Adarsh Sathyanarayanan #ID: 02188118 ###############################

#Importing the important libraries #################################

# Load necessary libraries

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(sandwich)  
library(readxl)  
library(ggplot2)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.2

## corrplot 0.95 loaded

library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

#load the data from the observations

Data\_input\_on\_WFH <- read\_excel("C:/Users/91886/Desktop/Data input on WFH.xlsx")  
data <- data.frame(Data\_input\_on\_WFH)

#Data Cleaning ##############

#Rename column for convenience

new\_column\_names <- c(  
 "Timestamp",  
 "Job\_Title",  
 "Industry",  
 "Current\_Work\_Arrangement",  
 "Pre\_Pandemic\_Arrangement",  
 "Age\_Range",  
 "Gender",  
 "Living\_Situation",  
 "Children\_Age",  
 "Elderly\_Care\_Level",  
 "Family\_Impact",  
 "Remote\_Work\_Policy",  
 "Ideal\_Work\_Arrangement",  
 "Work\_Life\_Balance\_Impact",  
 "Happiness\_Impact",  
 "Stress\_Experience",  
 "Comm\_Collab\_Before",  
 "Comm\_Collab\_After",  
 "AI\_Tool\_Usage",  
 "AI\_Tool\_Impact"  
)  
colnames(data) <- new\_column\_names  
  
# Create Happiness\_Score from Happiness\_Impact  
data$Happiness\_Score <- case\_when(  
 data$Happiness\_Impact == "Significantly improved" ~ 3,  
 data$Happiness\_Impact == "Slightly improved" ~ 2,  
 data$Happiness\_Impact == "No change" ~ 1,  
 data$Happiness\_Impact == "Slightly worsened" ~ -1,  
 data$Happiness\_Impact == "Significantly worsened" ~ -2,  
 TRUE ~ NA\_real\_ # Handle other cases like NA  
)  
  
# Create Stress\_Score (yes takes -1 because that indicates stress and no indicates no stress hence 1)  
data$Stress\_Score <- case\_when(  
 data$Stress\_Experience == "Yes" ~ -1,  
 data$Stress\_Experience == "No" ~ 1,  
 data$Stress\_Experience == "Unsure" ~ 0,  
 TRUE ~ NA\_real\_ # Handle other cases like NA  
)  
  
# Create Work Life Balance Score  
data$Work\_Life\_Balance\_Score <- case\_when(  
 data$Work\_Life\_Balance\_Impact == "Significantly improved" ~ 3,  
 data$Work\_Life\_Balance\_Impact == "Slightly improved" ~ 2,  
 data$Work\_Life\_Balance\_Impact == "No change" ~ 1,  
 data$Work\_Life\_Balance\_Impact == "Slightly worsened" ~ -1,  
 data$Work\_Life\_Balance\_Impact == "Significantly worsened" ~ -2,  
 TRUE ~ NA\_real\_ # Handle "Not applicable" and other cases as NA  
)  
  
# Create Family Impact score  
data$Family\_Impact\_Score <- case\_when(  
 data$Family\_Impact == "Very positively" ~ 2,  
 data$Family\_Impact == "Slightly positively" ~ 1,  
 data$Family\_Impact == "No impact" ~ 0,  
 data$Family\_Impact == "Slightly negative" ~ -1,  
 data$Family\_Impact == "Very negative" ~ -2,  
 TRUE ~ NA\_real\_ # Handle other cases or missing values as NA  
)  
  
# Create Gender score  
data$Gender\_Score <- case\_when(  
 data$Gender == "Male" ~ 0,  
 data$Gender == "Female" ~ 1,  
 data$Gender == "Other" ~ 2,  
 TRUE ~ NA\_real\_ # Handle other cases like NA  
)  
  
# Create children age score  
data$Children\_Age\_Score <- case\_when(  
 data$Children\_Age == "None" ~ 0,  
 data$Children\_Age == "1-2" ~ 1,  
 data$Children\_Age == "3+" ~ 2,  
 TRUE ~ NA\_real\_ # Handle other cases or missing values as NA  
)  
  
#Create age range score  
data$Age\_Range\_Score <- case\_when(  
 data$Age\_Range == "18-24" ~ 1,  
 data$Age\_Range == "25-34" ~ 2,  
 data$Age\_Range == "35-44" ~ 3,  
 data$Age\_Range == "45-54" ~ 4,  
 data$Age\_Range == "55-64" ~ 5,  
 TRUE ~ NA\_real\_ # Handle other cases like NA  
)  
  
#Create elderly care level score  
data$Elderly\_Care\_Level\_Score <- case\_when(  
 data$Elderly\_Care\_Level == "Not applicable" ~ 0,  
 data$Elderly\_Care\_Level == "Minimal assistance" ~ 1,  
 data$Elderly\_Care\_Level == "Moderate assistance (e.g., help with daily tasks)" ~ 2,  
 TRUE ~ NA\_real\_  
)  
  
#Create current work arrangement level score  
data$Current\_Work\_Arrangement\_Score <- case\_when(  
 data$Current\_Work\_Arrangement == "Full time in-office" ~ 1,  
 data$Current\_Work\_Arrangement == "Hybrid (some days remote, some in-office)" ~ 2,  
 data$Current\_Work\_Arrangement == "Remote" ~ 3,  
 TRUE ~ NA\_real\_  
)  
  
data$Living\_Situation\_Score <- case\_when(  
 data$Living\_Situation == "Alone" ~ 1,  
 data$Living\_Situation == "Roommates" ~ 2,  
 data$Living\_Situation == "Elderly family members (e.g., parents or grandparents)" ~ 3,  
 data$Living\_Situation == "Children" ~ 4,  
 data$Living\_Situation == "Alone, Roommates" ~ 5,  
 data$Living\_Situation == "Partner/spouse" ~ 6,  
 data$Living\_Situation == "Children, Partner/spouse" ~ 7,  
 data$Living\_Situation == "Children, Elderly family members (e.g., parents or grandparents), Partner/spouse" ~ 8,  
 data$Living\_Situation == "Elderly family members (e.g., parents or grandparents), Partner/spouse" ~ 9,  
 TRUE ~ NA\_real\_ # Handle other cases like NA  
)  
  
data$Children\_Age\_Score <- case\_when(  
 data$Children\_Age == "Not applicable" ~ 0,  
 data$Children\_Age == "Over 18 years old" ~ 1,  
 data$Children\_Age == "3-5 years old" ~ 2,  
 data$Children\_Age == "0-2 years old" ~ 3,  
 TRUE ~ NA\_real\_ # Handle other cases or missing values as NA  
)  
  
data$Comm\_Collab\_Before[is.na(data$Comm\_Collab\_Before)] <- 0  
data$Comm\_Collab\_After[is.na(data$Comm\_Collab\_After)] <- 0  
  
View(data)

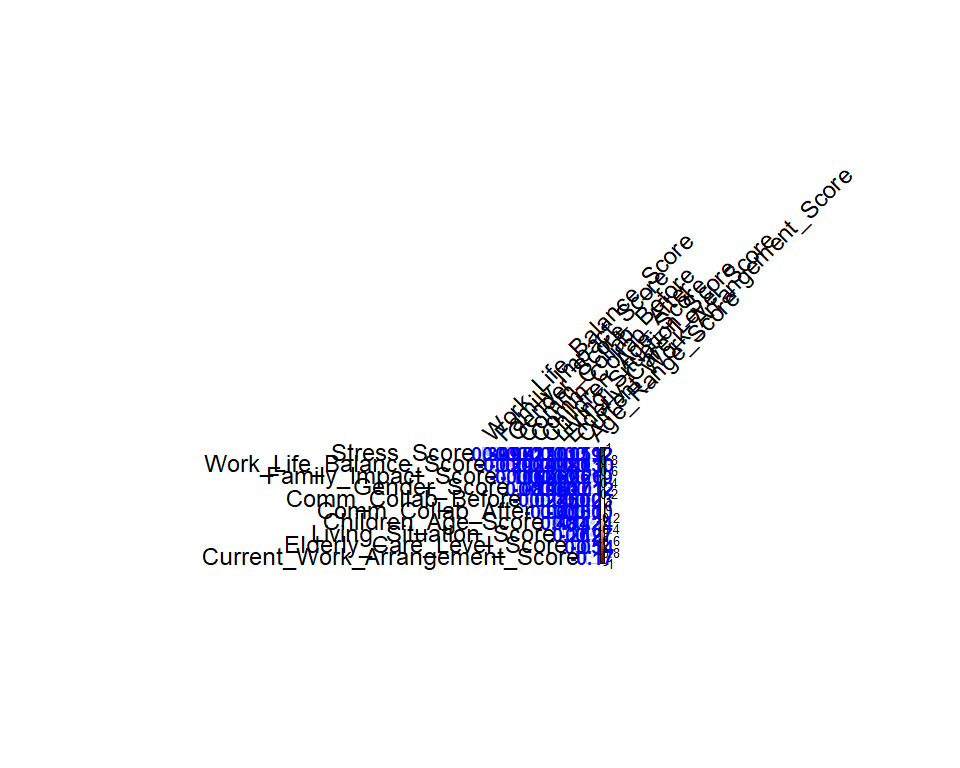
# Correlation Matrix

# Select relevant variables for correlation analysis  
cor\_data <- data[, c("Stress\_Score",   
 "Work\_Life\_Balance\_Score",   
 "Family\_Impact\_Score",   
 "Gender\_Score",   
 "Comm\_Collab\_Before",   
 "Comm\_Collab\_After",   
 "Children\_Age\_Score",   
 "Living\_Situation\_Score",   
 "Elderly\_Care\_Level\_Score",   
 "Current\_Work\_Arrangement\_Score",   
 "Age\_Range\_Score")]  
  
# Compute the correlation matrix using complete observations  
cor\_matrix <- cor(cor\_data, use = "complete.obs")  
  
# Visualize the correlation matrix with improved clarity  
corrplot(cor\_matrix,   
 method = "circle", # Use circle method for correlations  
 type = "upper", # Display only the upper triangle  
 tl.col = "black", # Text color for labels  
 tl.srt = 45, # Label angle  
 tl.cex = 1.5, # Increase label font size  
 addCoef.col = "blue", # Coefficients in blue  
 number.cex = 1.2, # Coefficient number size  
 mar = c(5, 5, 4, 4), # Adjust margins to avoid overlap  
 cl.lim = c(0, 1), # Color scale from 0 to 1  
 diag = FALSE, # Hide the diagonal  
 number.digits = 2) # Limit coefficients to 2 decimal places

## Warning in text.default(pos.xlabel[, 1], pos.xlabel[, 2], newcolnames, srt =  
## tl.srt, : "cl.lim" is not a graphical parameter

## Warning in text.default(pos.ylabel[, 1], pos.ylabel[, 2], newrownames, col =  
## tl.col, : "cl.lim" is not a graphical parameter

## Warning in title(title, ...): "cl.lim" is not a graphical parameter



#Null Hypothesis Testing ##########################

full\_model <- lm(Happiness\_Score ~ Stress\_Score + Work\_Life\_Balance\_Score + Family\_Impact\_Score + Gender\_Score +   
 Comm\_Collab\_Before + Comm\_Collab\_After + Children\_Age\_Score + Living\_Situation\_Score +   
 Elderly\_Care\_Level\_Score + Current\_Work\_Arrangement\_Score + Age\_Range\_Score, data = data)  
  
coeftest(full\_model, vcov = vcovHC(full\_model, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.456733 0.830326 -0.5501 0.5856723   
## Stress\_Score 0.452245 0.181704 2.4889 0.0175737 \*   
## Work\_Life\_Balance\_Score 0.427007 0.116573 3.6630 0.0007957 \*\*\*  
## Family\_Impact\_Score 0.026306 0.183838 0.1431 0.8870151   
## Gender\_Score -0.116211 0.316988 -0.3666 0.7160559   
## Comm\_Collab\_Before -0.062803 0.038600 -1.6270 0.1124601   
## Comm\_Collab\_After -0.011213 0.092279 -0.1215 0.9039632   
## Children\_Age\_Score 0.178329 0.162064 1.1004 0.2784776   
## Living\_Situation\_Score 0.195151 0.086344 2.2602 0.0299563 \*   
## Elderly\_Care\_Level\_Score 0.094884 0.156181 0.6075 0.5473161   
## Current\_Work\_Arrangement\_Score 0.702142 0.338338 2.0753 0.0451663 \*   
## Age\_Range\_Score -0.121187 0.161612 -0.7499 0.4582091   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Null Hypothesis (H₀):There is no significant change in happiness after the introduction of Work From Home (WFH) policies.

# Alternative Hypothesis (H₁):There is a significant change in happiness after the introduction of Work From Home (WFH) policies.

#Based on the above coefficient test we can safely say that the p value of Current\_Work\_Arrangement\_Score is 0.0452 (<0.05). #Hence we can safely reject Null hypothesis and state that Work from policies has impacted people’s happiness. #We consider Current\_Work\_Arrangement\_Score for disproving H0 because it reflects the change in work arrangement due to the introduction of WFH policies, which is the key factor in determining whether happiness has changed as a result.

#Understanding the factors that effect happiness ################################################

coeftest(full\_model, vcov = vcovHC(full\_model, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.456733 0.830326 -0.5501 0.5856723   
## Stress\_Score 0.452245 0.181704 2.4889 0.0175737 \*   
## Work\_Life\_Balance\_Score 0.427007 0.116573 3.6630 0.0007957 \*\*\*  
## Family\_Impact\_Score 0.026306 0.183838 0.1431 0.8870151   
## Gender\_Score -0.116211 0.316988 -0.3666 0.7160559   
## Comm\_Collab\_Before -0.062803 0.038600 -1.6270 0.1124601   
## Comm\_Collab\_After -0.011213 0.092279 -0.1215 0.9039632   
## Children\_Age\_Score 0.178329 0.162064 1.1004 0.2784776   
## Living\_Situation\_Score 0.195151 0.086344 2.2602 0.0299563 \*   
## Elderly\_Care\_Level\_Score 0.094884 0.156181 0.6075 0.5473161   
## Current\_Work\_Arrangement\_Score 0.702142 0.338338 2.0753 0.0451663 \*   
## Age\_Range\_Score -0.121187 0.161612 -0.7499 0.4582091   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Stress\_Score: Stress has a significant positive effect on happiness (p-value = 0.0176). #Work\_Life\_Balance\_Score: Work-life balance has a significant positive effect on happiness (p-value = 0.0008). #Living\_Situation\_Score: Living situation has a significant positive effect on happiness (p-value = 0.0300). #Current\_Work\_Arrangement\_Score: Current work arrangement (remote, hybrid, or in-office) has a significant positive effect on happiness (p-value = 0.0452).

#Refined model 1

refined\_model <- lm(Happiness\_Score ~ Stress\_Score + Work\_Life\_Balance\_Score +   
 Living\_Situation\_Score + Current\_Work\_Arrangement\_Score, data = data)  
coeftest(refined\_model, vcov = vcovHC(refined\_model, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.080598 0.556976 -1.9401 0.0589417 .   
## Stress\_Score 0.414173 0.194914 2.1249 0.0393817 \*   
## Work\_Life\_Balance\_Score 0.458051 0.116852 3.9199 0.0003132 \*\*\*  
## Living\_Situation\_Score 0.182881 0.061543 2.9716 0.0048355 \*\*   
## Current\_Work\_Arrangement\_Score 0.693472 0.251312 2.7594 0.0084736 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Refined model 2 #Stress may affect a person’s work life balance

refined\_model1 <- lm(Happiness\_Score ~ Stress\_Score \* Work\_Life\_Balance\_Score +   
 Living\_Situation\_Score + Current\_Work\_Arrangement\_Score, data = data)  
coeftest(refined\_model1, vcov = vcovHC(refined\_model1, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.076025 0.550450 -1.9548 0.0572830 .   
## Stress\_Score 0.396289 0.207463 1.9102 0.0629558 .   
## Work\_Life\_Balance\_Score 0.455341 0.124138 3.6680 0.0006824 \*\*\*  
## Living\_Situation\_Score 0.183313 0.063065 2.9067 0.0058090 \*\*   
## Current\_Work\_Arrangement\_Score 0.686392 0.246114 2.7889 0.0079142 \*\*   
## Stress\_Score:Work\_Life\_Balance\_Score 0.017940 0.124628 0.1440 0.8862269   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#The interaction between stress and work-life balance is not significant (p-value > 0.05). This means that stress and work-life balance do not interact in a way that significantly affects happiness.

#Refined model 3

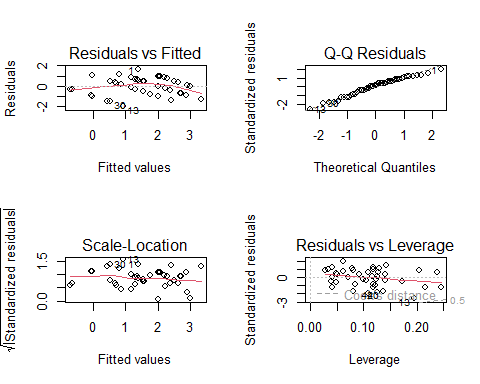
refined\_model2 <- lm(Happiness\_Score ~ Stress\_Score + Work\_Life\_Balance\_Score + Living\_Situation\_Score +   
 Current\_Work\_Arrangement\_Score\*Gender\_Score, data = data)  
coeftest(refined\_model2, vcov = vcovHC(refined\_model2, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value  
## (Intercept) -1.642048 0.791123 -2.0756  
## Stress\_Score 0.418560 0.180199 2.3228  
## Work\_Life\_Balance\_Score 0.396187 0.118658 3.3389  
## Living\_Situation\_Score 0.189564 0.063188 3.0000  
## Current\_Work\_Arrangement\_Score 1.077309 0.403550 2.6696  
## Gender\_Score 1.418573 0.844356 1.6801  
## Current\_Work\_Arrangement\_Score:Gender\_Score -0.906274 0.482767 -1.8773  
## Pr(>|t|)   
## (Intercept) 0.044245 \*   
## Stress\_Score 0.025233 \*   
## Work\_Life\_Balance\_Score 0.001799 \*\*  
## Living\_Situation\_Score 0.004577 \*\*  
## Current\_Work\_Arrangement\_Score 0.010835 \*   
## Gender\_Score 0.100550   
## Current\_Work\_Arrangement\_Score:Gender\_Score 0.067611 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#The interaction term between Current\_Work\_Arrangement\_Score and Gender\_Score is marginally significant with a p-value of 0.0676 (just above the 0.05 threshold). This indicates that the effect of work arrangement on happiness may differ slightly by gender.

#Choosing model1 ################

par(mfrow = c(2, 2)) # Arrange the plots in a 2x2 grid  
plot(refined\_model)



#From the first graph of Model 1, its evident there is behavior of non linear model from the curve

# Editing the stress data as it consists of -1 and 0 and performing log on top of it would result in null result

data$Stress\_Score\_adjusted <- data$Stress\_Score + 2  
  
#   
refined\_model3 <- lm(Happiness\_Score ~ log(Stress\_Score\_adjusted) + Work\_Life\_Balance\_Score +   
 Living\_Situation\_Score + Current\_Work\_Arrangement\_Score, data = data)  
coeftest(refined\_model3, vcov = vcovHC(refined\_model3, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.540734 0.626116 -2.4608 0.0179551 \*   
## log(Stress\_Score\_adjusted) 0.805017 0.367508 2.1905 0.0339596 \*   
## Work\_Life\_Balance\_Score 0.448127 0.116823 3.8359 0.0004042 \*\*\*  
## Living\_Situation\_Score 0.178606 0.061928 2.8841 0.0061103 \*\*   
## Current\_Work\_Arrangement\_Score 0.701267 0.253024 2.7715 0.0082109 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Visualizing the models ######################

refined\_models <- list(  
 full\_model, # Model with all terms  
 refined\_model1, # Model with interaction terms (Stress and Work-Life Balance)  
 refined\_model2, # Model with interaction terms (Current\_Work\_Arrangement and Gender)  
 refined\_model3 # Model with log term  
)  
  
stargazer(refined\_model3,  
 type = "text", # Use "text" for console output or "html" for R Markdown/HTML  
 title = "Regression Results for Final Model",  
 dep.var.labels = c("Happiness Score"),  
 covariate.labels = c("log(Stress Score Adjusted)",  
 "Work Life Balance Score",  
 "Living Situation Score",  
 "Current Work Arrangement Score"),  
 digits = 4,  
 out = "final\_model\_summary.txt")

##   
## Regression Results for Final Model  
## ==========================================================  
## Dependent variable:   
## ---------------------------  
## Happiness Score   
## ----------------------------------------------------------  
## log(Stress Score Adjusted) 0.8050\*\*   
## (0.3054)   
##   
## Work Life Balance Score 0.4481\*\*\*   
## (0.0986)   
##   
## Living Situation Score 0.1786\*\*\*   
## (0.0654)   
##   
## Current Work Arrangement Score 0.7013\*\*\*   
## (0.2248)   
##   
## Constant -1.5407\*\*\*   
## (0.5168)   
##   
## ----------------------------------------------------------  
## Observations 48   
## R2 0.6105   
## Adjusted R2 0.5743   
## Residual Std. Error 0.8503 (df = 43)   
## F Statistic 16.8493\*\*\* (df = 4; 43)   
## ==========================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

stargazer(refined\_models,   
 type = "text",   
 dep.var.labels = c("Happiness Score"), # Dependent variable label  
 covariate.labels = c("Stress Score", "Work Life Balance Score", "Living Situation Score",   
 "Current Work Arrangement Score", "Stress\*Work Life Balance",   
 "Current Work Arrangement\*Gender", "Log(Stress Score)"),  
 out = "project\_models\_comparison.txt")

##   
## =======================================================================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------------------------------  
## Happiness Score   
## (1) (2) (3) (4)   
## ---------------------------------------------------------------------------------------------------------------------------------------  
## Stress Score 0.452\*\* 0.396\* 0.419\*\*   
## (0.180) (0.197) (0.160)   
##   
## Work Life Balance Score 0.805\*\*   
## (0.305)   
##   
## Living Situation Score 0.427\*\*\* 0.455\*\*\* 0.396\*\*\* 0.448\*\*\*   
## (0.114) (0.101) (0.102) (0.099)   
##   
## Current Work Arrangement Score 0.026   
## (0.155)   
##   
## Stress\*Work Life Balance -0.116 1.419\*   
## (0.306) (0.825)   
##   
## Current Work Arrangement\*Gender -0.063   
## (0.039)   
##   
## Log(Stress Score) -0.011   
## (0.079)   
##   
## Children\_Age\_Score 0.178   
## (0.293)   
##   
## Current\_Work\_Arrangement\_Score:Gender\_Score -0.906\*\*   
## (0.441)   
##   
## Living\_Situation\_Score 0.195\*\* 0.183\*\*\* 0.190\*\*\* 0.179\*\*\*   
## (0.085) (0.067) (0.064) (0.065)   
##   
## Elderly\_Care\_Level\_Score 0.095   
## (0.209)   
##   
## Current\_Work\_Arrangement\_Score 0.702\*\* 0.686\*\*\* 1.077\*\*\* 0.701\*\*\*   
## (0.269) (0.233) (0.289) (0.225)   
##   
## Age\_Range\_Score -0.121   
## (0.200)   
##   
## Stress\_Score:Work\_Life\_Balance\_Score 0.018   
## (0.106)   
##   
## Constant -0.457 -1.076\*\* -1.642\*\*\* -1.541\*\*\*   
## (0.754) (0.523) (0.604) (0.517)   
##   
## ---------------------------------------------------------------------------------------------------------------------------------------  
## Observations 48 48 48 48   
## R2 0.651 0.606 0.646 0.610   
## Adjusted R2 0.545 0.559 0.594 0.574   
## Residual Std. Error 0.879 (df = 36) 0.866 (df = 42) 0.830 (df = 41) 0.850 (df = 43)   
## F Statistic 6.118\*\*\* (df = 11; 36) 12.908\*\*\* (df = 5; 42) 12.465\*\*\* (df = 6; 41) 16.849\*\*\* (df = 4; 43)  
## =======================================================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#The Final Linear model to support #Happiness\_Score=−1.5407+0.8050*log(Stress)+0.448*Work\_Life\_Balance\_Score+0.178*Living\_Situation\_Score+0.701*Current\_Work\_Arrangement\_Score