

# Thang\_Activity 4

2025-11-09

```
## Set global CRAN mirror
options(repos = c(CRAN = "https://cran.rstudio.com/"))

## Install the tableone package
install.packages("tableone")

## Installing package into 'C:/Users/Thang/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'tableone' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Thang\AppData\Local\Temp\RtmpATs4He\downloaded_packages

## Load the required packages

library(tableone)
library(readxl)
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(janitor)

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
library(knitr)

setwd("C:/Users/Thang/OneDrive/Documents/ActivityData_Assignment/Activity_4")
data <- read_excel("SuicideRisk_Data.xlsx")
```

## QUESTION 1.

```
# =====
# PART A: SABCS ~ Demographics
# =====

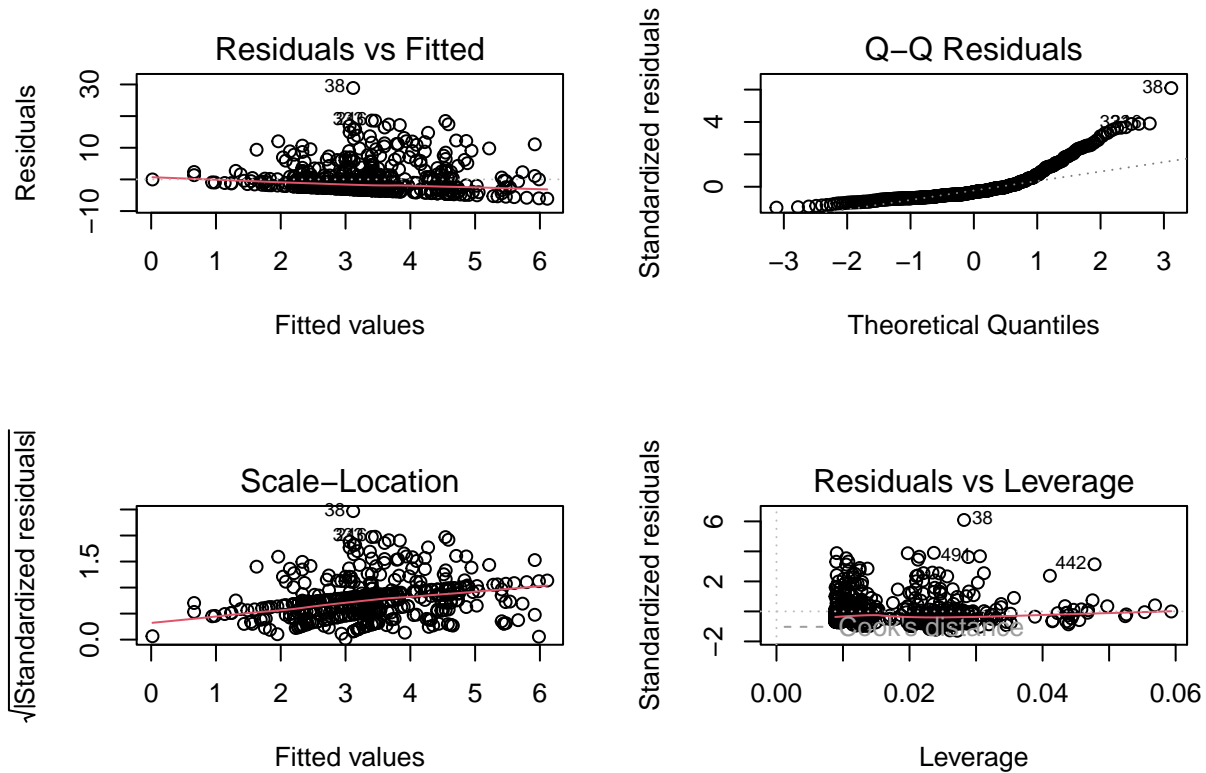
# Convert categorical variables to factors
data$GENDER <- factor(data$GENDER)
data$RACE <- factor(data$RACE)
data$ETHNICITY <- factor(data$ETHNICITY)
data$INCOME <- factor(data$INCOME)

model_a <- lm(SABCS_TOTAL_SUM ~ AGE + GENDER + RACE + ETHNICITY + INCOME, data = data)

summary(model_a)
```

```
##
## Call:
## lm(formula = SABCS_TOTAL_SUM ~ AGE + GENDER + RACE + ETHNICITY +
##     INCOME, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.1136 -2.9988 -1.5688  0.7734 28.8824
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.26744    1.20338   5.208 2.72e-07 ***
## AGE             -0.06428    0.03280  -1.960  0.0505 .
## GENDERMale       -0.28798    0.74409  -0.387  0.6989
## RACEOther        -0.10208    0.81829  -0.125  0.9008
## RACEWhite/Caucasian -0.35525    0.67499  -0.526  0.5989
## ETHNICITYNot Hispanic/Latino -1.18615    0.65198  -1.819  0.0694 .
## INCOME$51,000 - $75,000 -1.04222    0.67541  -1.543  0.1234
## INCOME$76,000 - $100,000 -0.31306    0.69360  -0.451  0.6519
## INCOME< $30,000    1.16960    0.66952   1.747  0.0812 .
## INCOME>$100,000    0.02945    0.67065   0.044  0.9650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.808 on 536 degrees of freedom
## Multiple R-squared:  0.03723,    Adjusted R-squared:  0.02107
## F-statistic: 2.303 on 9 and 536 DF,  p-value: 0.01522
```

```
# Diagnostics
par(mfrow=c(2,2))
plot(model_a)
```



```
# Reset plotting
par(mfrow=c(1,1))
```

The demographic model accounts for a mere 3.7% of the variation in suicidal risk, indicating that demographics alone do not serve as robust predictors of SABCS scores. Age has a small negative correlation, indicating a somewhat diminished suicide risk among older participants; nevertheless, the effect is minor. Ethnicity (Not Hispanic/Latino) and low income (<\$30,000) exhibit marginal correlations with SABCS results, however none serve as robust predictors. The majority of demographic factors, such as gender and ethnicity, lack statistical significance. In general, demographic traits don't give us much information on how likely someone in this group is to commit suicide.

```
# =====
# PART B: SABCS ~ Depression
# =====

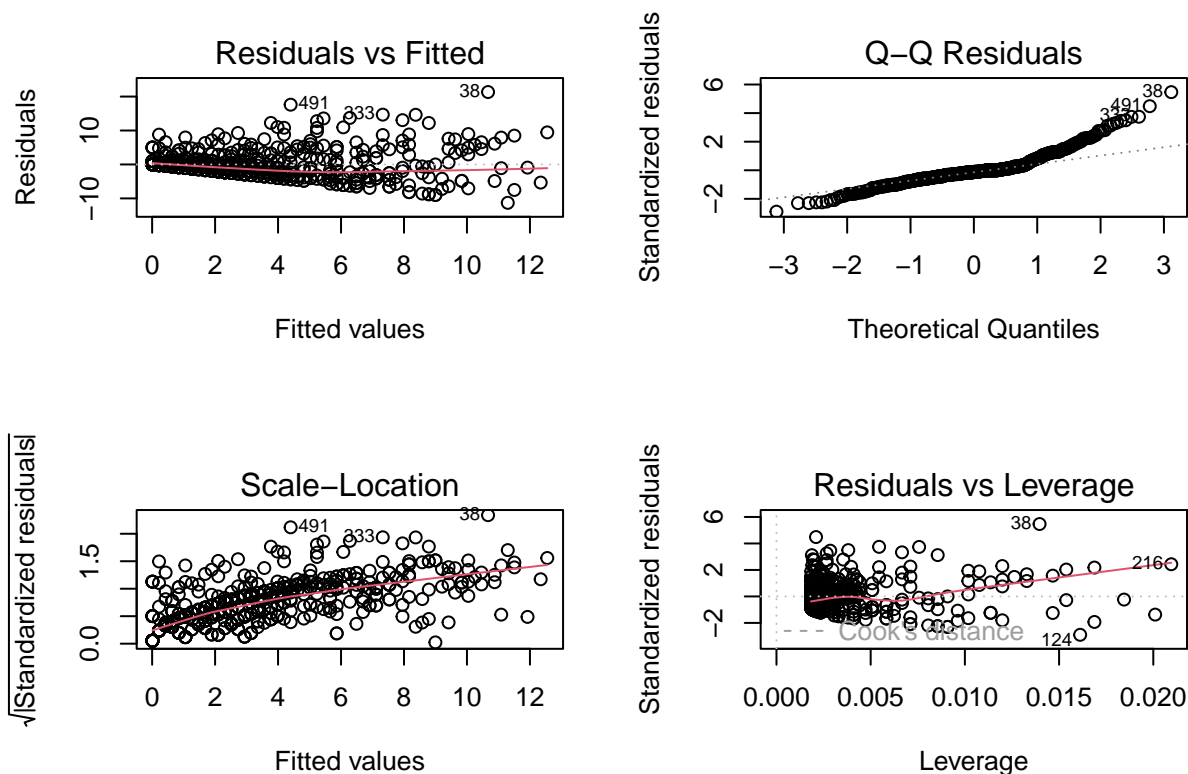
model_b <- lm(SABCS_TOTAL_SUM ~ CESDR_TOTAL_SUM, data = data)

summary(model_b)
```

```
##
```

```
## Call:
## lm(formula = SABCS_TOTAL_SUM ~ CESDR_TOTAL_SUM, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2968  -2.1009  -0.4739   0.9891  21.3302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.01088    0.25650   0.042   0.966
## CESDR_TOTAL_SUM 0.20900    0.01228  17.013 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.93 on 544 degrees of freedom
## Multiple R-squared:  0.3473, Adjusted R-squared:  0.3461
## F-statistic: 289.4 on 1 and 544 DF,  p-value: < 2.2e-16
```

```
# Diagnostics
par(mfrow=c(2,2))
plot(model_b)
```



```
par(mfrow=c(1,1))
```

There is a substantial and significant positive relationship between depressive symptoms (CESD-R scores) and suicide risk, which means that higher depression levels are linked to higher SABCS scores. The model

elucidates around 35% of the variance in suicide risk, far surpassing the demographic model. In this sample, depression is a strong indicator of the probability of suicide.

```
# =====
# PART C: SABCS ~ ACES 1-10
# =====

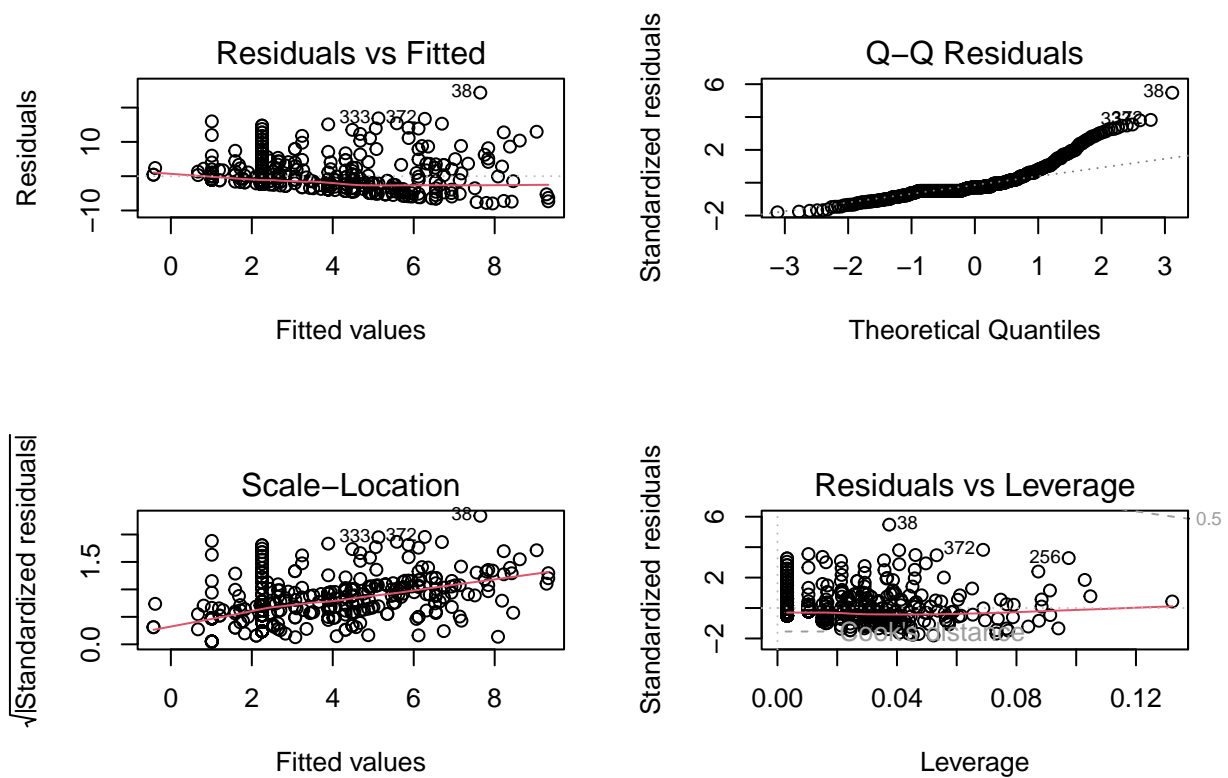
aces_vars <- paste0("ACES__", 1:10)

model_c <- lm(as.formula(
  paste("SABCS_TOTAL_SUM ~", paste(aces_vars, collapse = " + "))
), data = data)

summary(model_c)

##
## Call:
## lm(formula = as.formula(paste("SABCS_TOTAL_SUM ~", paste(aces_vars,
##   collapse = " + "))), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.9486 -2.2530 -1.2530  0.9881 24.3547
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.25301     0.25704   8.765  < 2e-16 ***
## ACES__1      0.19534     0.72244   0.270  0.786969
## ACES__2      2.06336     0.56740   3.637  0.000303 ***
## ACES__3      2.23049     0.56574   3.943  9.13e-05 ***
## ACES__4      0.09203     1.26894   0.073  0.942208
## ACES__5      0.81234     0.70763   1.148  0.251491
## ACES__6     -0.53680     0.65125  -0.824  0.410164
## ACES__7      0.57879     0.66378   0.872  0.383619
## ACES__8      1.63521     0.53899   3.034  0.002532 **
## ACES__9     -1.24110     0.47201  -2.629  0.008799 **
## ACES__10    -1.44145     0.99893  -1.443  0.149608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.533 on 535 degrees of freedom
## Multiple R-squared:  0.146, Adjusted R-squared:  0.13
## F-statistic: 9.147 on 10 and 535 DF, p-value: 5.074e-14

# Diagnostics
par(mfrow=c(2,2))
plot(model_c)
```



```
par(mfrow=c(1,1))

# Extract coefficients with p < 0.20
get_sig <- function(model, original_vars) {
  tbl <- summary(model)$coefficients
  sig_terms <- rownames(tbl)[tbl[,4] < 0.20] # significant rows (dummy names)
  sig_terms <- sig_terms[sig_terms != "(Intercept)"]

  matched <- c()
  for (term in sig_terms) {
    base <- original_vars[sapply(original_vars, function(v) startsWith(term, v))]
    matched <- c(matched, base)
  }

  unique(matched)
}

orig_a <- c("AGE", "GENDER", "RACE", "ETHNICITY", "INCOME")
orig_b <- c("CESDR_TOTAL_SUM")
orig_c <- paste0("ACES_--", 1:10)

sig_a <- get_sig(model_a, orig_a)
sig_b <- get_sig(model_b, orig_b)
sig_c <- get_sig(model_c, orig_c)

sig_a; sig_b; sig_c
```

```
## [1] "AGE" "ETHNICITY" "INCOME"
```

```
## [1] "CESDR_TOTAL_SUM"
```

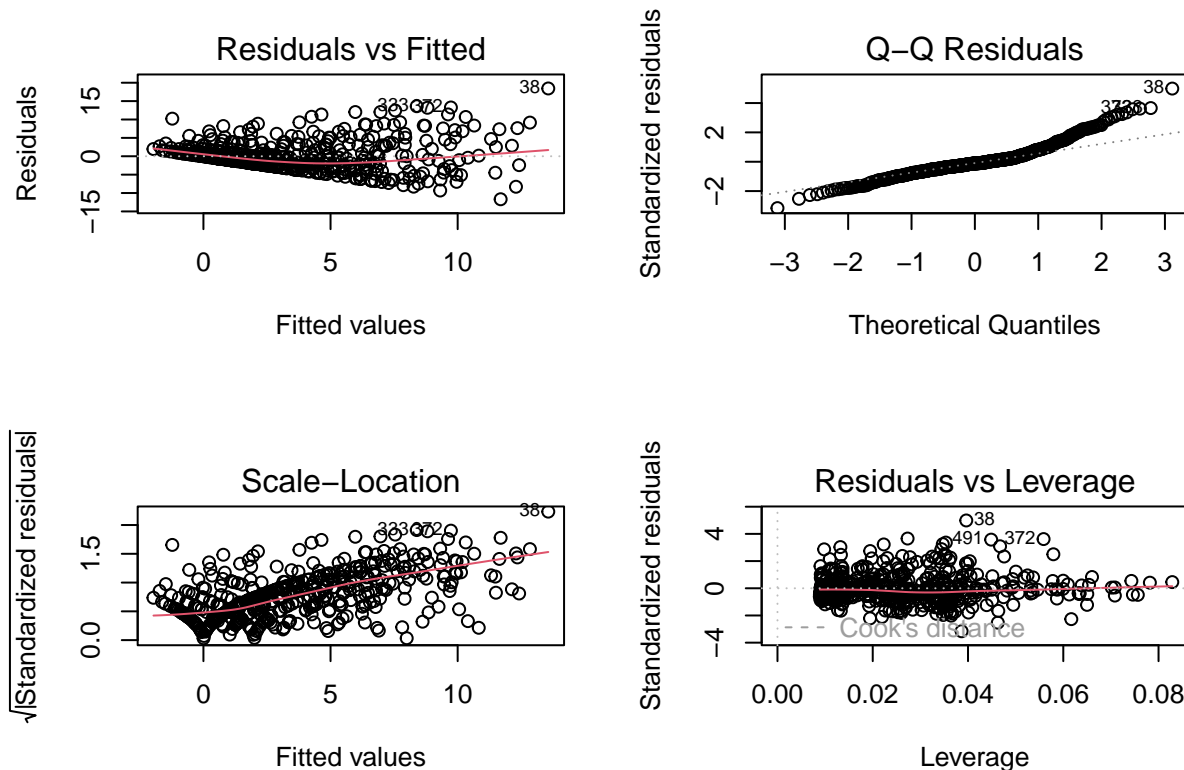
```
## [1] "ACES___2" "ACES___3" "ACES___8" "ACES___9" "ACES___1" "ACES___10"
```

Several ACE categories exhibit significant correlations with suicide risk, indicating that specific forms of childhood adversity are closely linked to elevated SABCS scores. In particular, ACEs 2, 3, and 8 are connected to a much increased risk of suicide, while ACE 9 shows a negative link. Most other ACE items do not have significant impacts, indicating that not all forms of childhood adversity equally influence suicidal thoughts or behaviors. The model accounts for around 14–15% of the variation in suicide risk, suggesting that childhood events are significant, though not the sole factor. Overall, ACEs help predict the risk of suicide, although the effects depend on the sort of adversity.

```
# =====  
# PART D  
# =====  
all_sig <- unique(c(sig_a, sig_b, sig_c))  
  
formula_d <- as.formula(  
  paste("SABCS_TOTAL_SUM ~", paste(all_sig, collapse = " + "))  
)  
  
model_d <- lm(formula_d, data = data)  
summary(model_d)  
  
##  
## Call:  
## lm(formula = formula_d, data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -11.6914  -1.9593  -0.4288   1.2998  18.4356   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)      1.89044    0.90389   2.091  0.03696 *      
## AGE              -0.05315    0.02637  -2.016  0.04432 *      
## ETHNICITYNot Hispanic/Latino -1.00572    0.50380  -1.996  0.04641 *      
## INCOME$51,000 - $75,000    -0.44277    0.52894  -0.837  0.40292        
## INCOME$76,000 - $100,000    0.20055    0.53890   0.372  0.70993        
## INCOME< $30,000           0.57864    0.52733   1.097  0.27301        
## INCOME>$100,000           0.78980    0.51999   1.519  0.12939        
## CESDR_TOTAL_SUM           0.18484    0.01269  14.560 < 2e-16 ***  
## ACES___2             1.36291    0.47280   2.883  0.00410 **      
## ACES___3             1.38402    0.45800   3.022  0.00263 **      
## ACES___8             0.83230    0.44182   1.884  0.06014 .        
## ACES___9            -1.27265    0.39121  -3.253  0.00121 **      
## ACES___1             0.11834    0.59304   0.200  0.84191        
## ACES___10            -0.21860    0.80080  -0.273  0.78498        
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##
```

```
## Residual standard error: 3.772 on 532 degrees of freedom
## Multiple R-squared:  0.4117, Adjusted R-squared:  0.3973
## F-statistic: 28.64 on 13 and 532 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2)); plot(model_d)
```



The combined model elucidates approximately 41% of the heterogeneity in suicide risk, representing a significant enhancement over the separate demographic and ACE models. Depression (CESDR scores) continues to be the most significant predictor, even after controlling for demographics and childhood adversity, demonstrating its critical role in suicidal risk. A number of ACE categories (ACES 2, 3, and 9) continue to exhibit substantial correlations, indicating that particular childhood events still have major implications beyond depression. Age and ethnicity exhibit minor yet statistically significant effects, whereas wealth demonstrates less influence. In general, integrating demographics, depression, and ACEs gives a better picture of the risk of suicide.

```
# =====
# PART E: sqrt(SABCS) model
# =====

# Create transformed variable
data$SABCS_sqrt <- sqrt(data$SABCS_TOTAL_SUM)

# Build formula using same predictors as Part D
formula_e <- as.formula(
  paste("SABCS_sqrt ~", paste(all_sig, collapse = " + "))
)
```



```

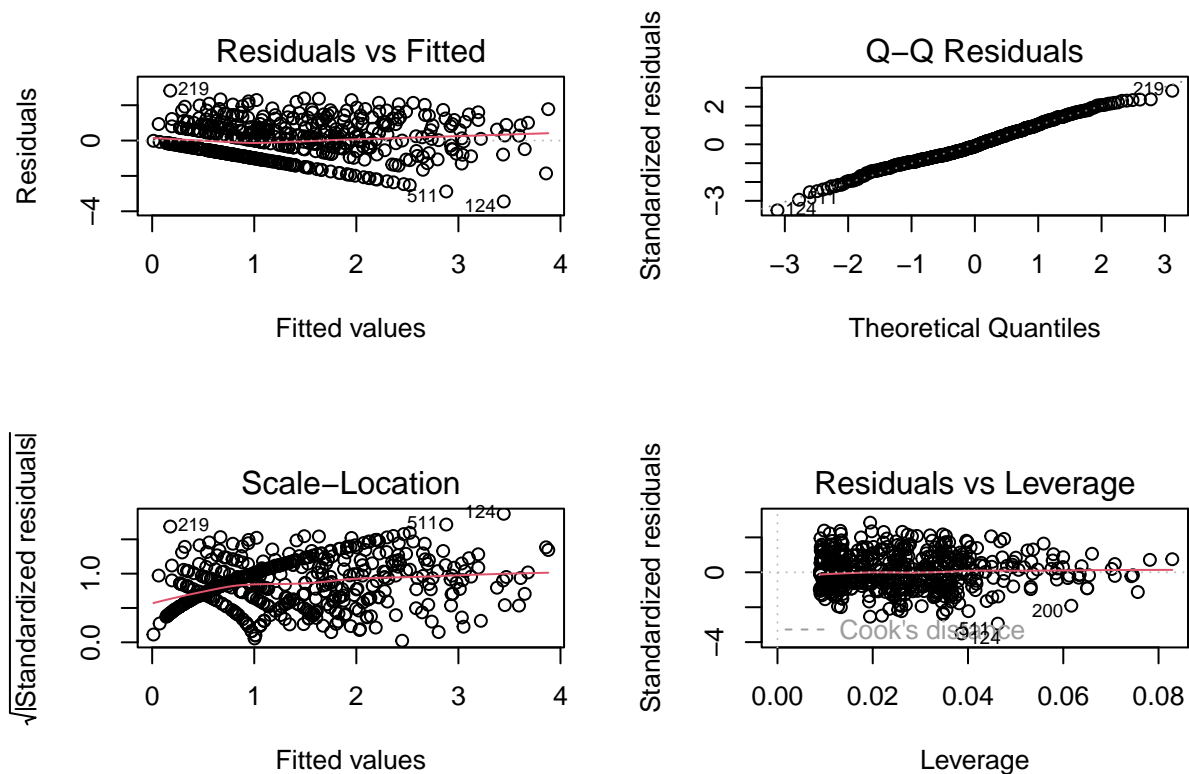
# Fit model
model_e <- lm(formula_e, data = data)

# Show results
summary(model_e)

##
## Call:
## lm(formula = formula_e, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4442 -0.6833 -0.0962  0.6821  2.8227
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.908047   0.240276   3.779 0.000175 ***
## AGE           -0.012019   0.007010  -1.715 0.086982 .
## ETHNICITYNot Hispanic/Latino -0.263033   0.133921  -1.964 0.050040 .
## INCOME$51,000 - $75,000 -0.218214   0.140606  -1.552 0.121268
## INCOME$76,000 - $100,000  0.041328   0.143254   0.288 0.773081
## INCOME< $30,000    0.131719   0.140176   0.940 0.347813
## INCOME>$100,000    0.147564   0.138227   1.068 0.286208
## CESDR_TOTAL_SUM    0.047502   0.003375  14.076 < 2e-16 ***
## ACES___2           0.354801   0.125681   2.823 0.004935 **
## ACES___3           0.331545   0.121747   2.723 0.006677 **
## ACES___8           0.231329   0.117447   1.970 0.049397 *
## ACES___9          -0.280136   0.103993  -2.694 0.007287 **
## ACES___1           0.059622   0.157644   0.378 0.705425
## ACES___10          0.120804   0.212873   0.567 0.570617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.003 on 532 degrees of freedom
## Multiple R-squared:  0.4, Adjusted R-squared:  0.3853
## F-statistic: 27.28 on 13 and 532 DF, p-value: < 2.2e-16

# Diagnostics
par(mfrow=c(2,2))
plot(model_e)

```



```
par(mfrow=c(1,1))
```

Following the square-root transformation, the model exhibits analogous patterns of important predictors as shown in Part D, with depression persisting as the most substantial factor linked to suicide risk. ACE items 2, 3, 8, and 9 still have significant effects, but the sizes of the effects are reduced because the transformation has made the outcome smaller. The residuals are more evenly distributed and have a smaller range, which means that the transformation makes the data more normal and stabilizes the variance. The converted model fits the data better overall, while still keeping the primary associations that were shown in the untransformed model.

```
# =====
# PART F: Compare Model D vs E
# =====

# Compare AIC & BIC
AIC(model_d, model_e)
```

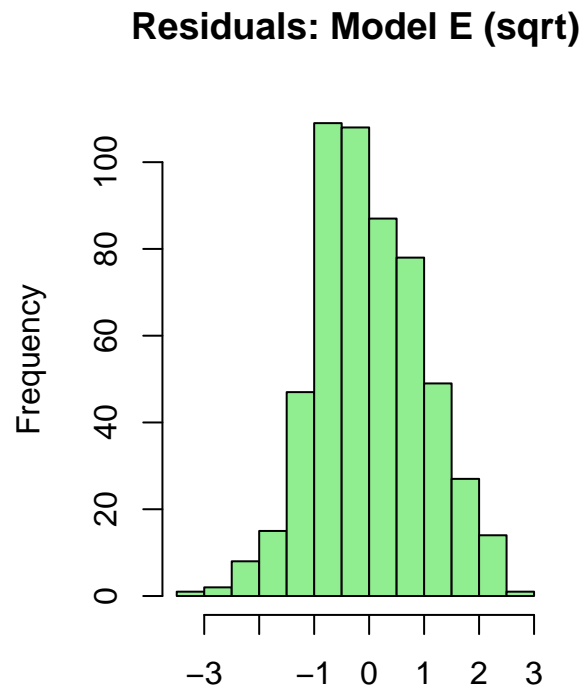
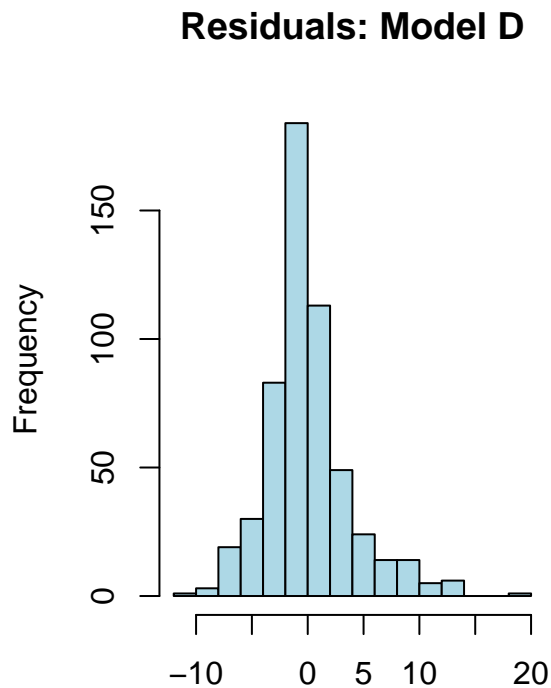
```
##          df      AIC
## model_d 15 3015.188
## model_e 15 1568.376
```

```
BIC(model_d, model_e)
```

```
##          df      BIC
```

```
## model_d 15 3079.727
## model_e 15 1632.915
```

```
# Residual histograms
par(mfrow=c(1,2))
hist(resid(model_d), main="Residuals: Model D", xlab="", col="lightblue")
hist(resid(model_e), main="Residuals: Model E (sqrt)", xlab="", col="lightgreen")
```



```
par(mfrow=c(1,1))

# Shapiro-Wilk normality test
shapiro.test(resid(model_d))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model_d)
## W = 0.93427, p-value = 9.056e-15
```

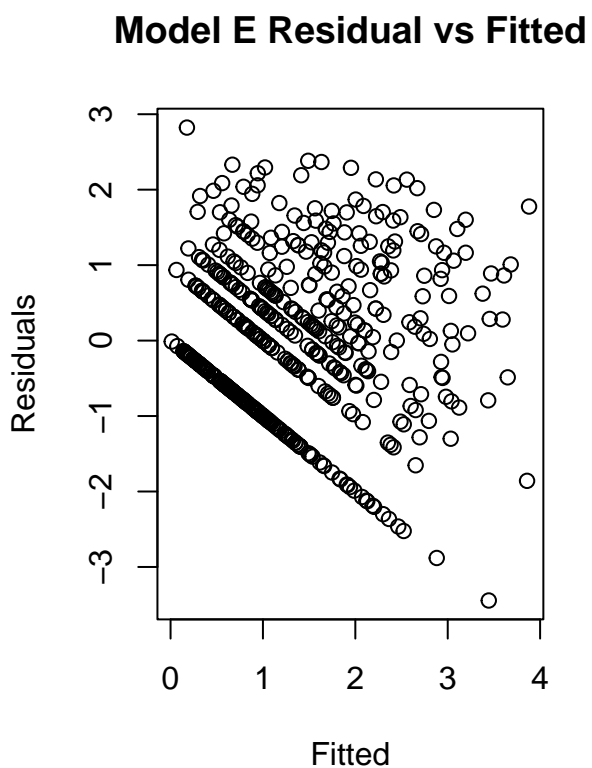
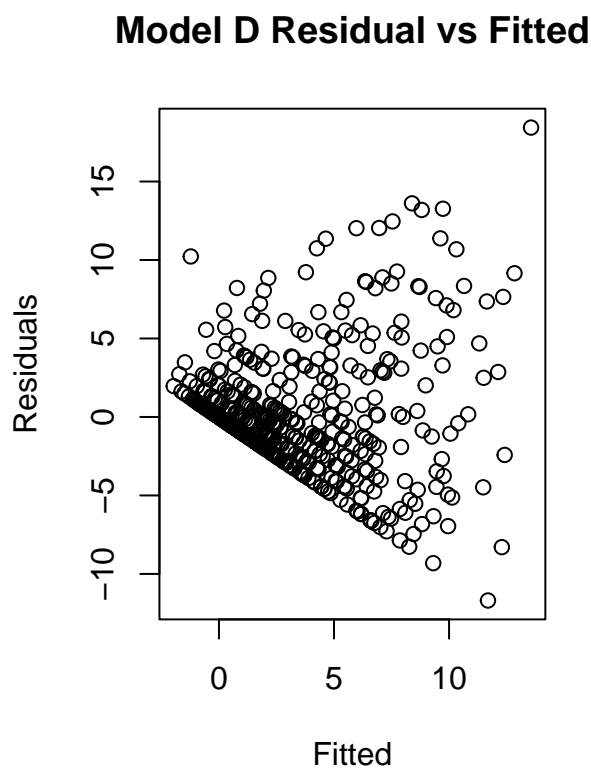
```
shapiro.test(resid(model_e))
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data: resid(model_e)
## W = 0.99344, p-value = 0.01771
```

```
# Residual vs Fitted comparison
par(mfrow=c(1,2))
plot(model_d$fitted.values, resid(model_d),
     main="Model D Residual vs Fitted",
     xlab="Fitted", ylab="Residuals")

plot(model_e$fitted.values, resid(model_e),
     main="Model E Residual vs Fitted",
     xlab="Fitted", ylab="Residuals")
```



```
par(mfrow=c(1,1))
```

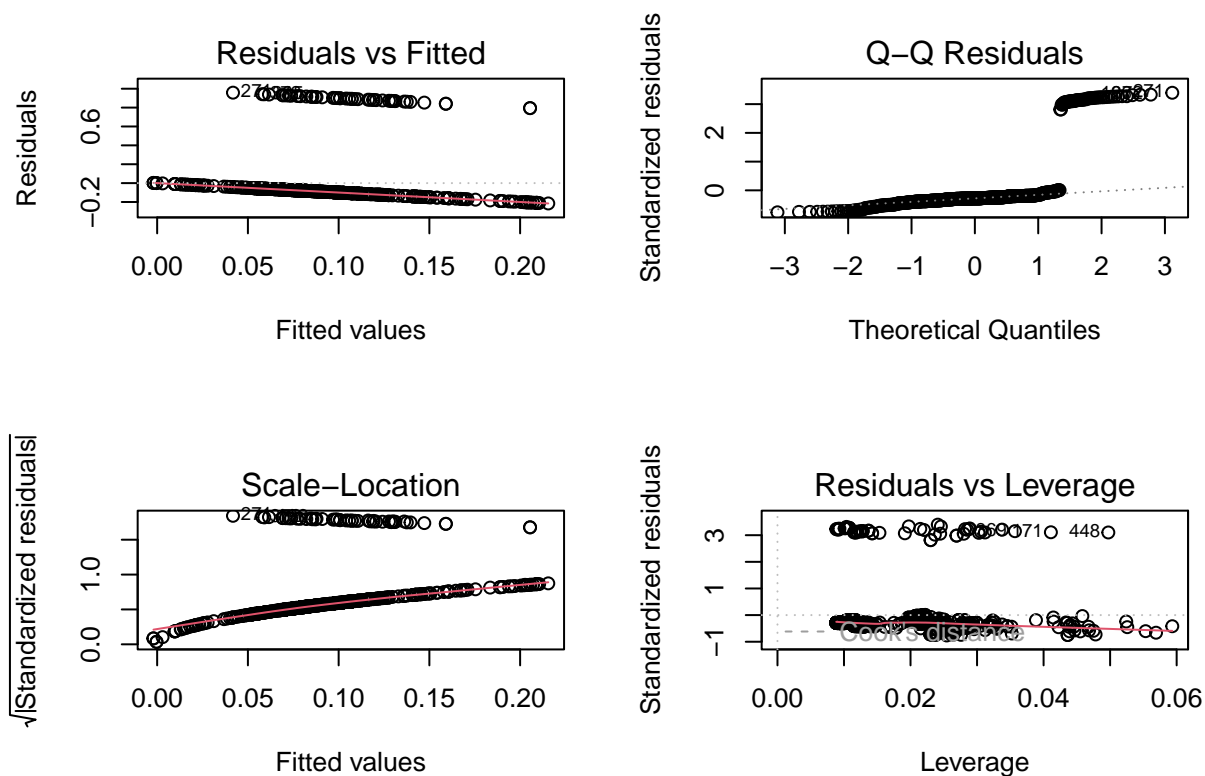
The square-root-transformed model (Model E) has much lower AIC and BIC values than the untransformed model (Model D). This means that it fits the data better. The Shapiro–Wilk test also shows that the residuals from Model E are substantially closer to a normal distribution ( $p = 0.018$ ) than those from Model D ( $p < 0.000000000000001$ ). This means that Model E is more normal. Even though the transformed model is harder to understand because the coefficients are based on the square root of suicide risk instead of the raw score, it fits the statistical assumptions better. Taking into account the trade-off between interpretability and statistical validity, Model E is a stronger choice because it fits better and is more trustworthy.

## QUESTION 2.

```
# =====  
# PART A: HX_SUICIDE ~ Demographics  
# =====  
  
model_a_hx <- lm(HX_SUICIDE ~ AGE + GENDER + RACE + ETHNICITY + INCOME, data = data)  
  
summary(model_a_hx)
```

```
##  
## Call:  
## lm(formula = HX_SUICIDE ~ AGE + GENDER + RACE + ETHNICITY + INCOME,  
##     data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.21551 -0.10245 -0.07885 -0.05568  0.95821   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    0.138268   0.071498   1.934   0.0537 .      
## AGE            0.001672   0.001949   0.858   0.3912      
## GENDERMale     0.012110   0.044210   0.274   0.7843      
## RACEOther     -0.038807   0.048618  -0.798   0.4251      
## RACEWhite/Caucasian  0.017234   0.040104   0.430   0.6676      
## ETHNICITYNot Hispanic/Latino -0.089846   0.038737  -2.319   0.0207 *     
## INCOME$51,000 - $75,000 -0.045096   0.040129  -1.124   0.2616      
## INCOME$76,000 - $100,000 -0.029880   0.041210  -0.725   0.4687      
## INCOME< $30,000  0.006487   0.039779   0.163   0.8705      
## INCOME>$100,000 -0.021926   0.039846  -0.550   0.5824      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.2857 on 536 degrees of freedom  
## Multiple R-squared:  0.01933,    Adjusted R-squared:  0.002864   
## F-statistic: 1.174 on 9 and 536 DF,  p-value: 0.3092
```

```
# Diagnostics  
par(mfrow=c(2,2))  
plot(model_a_hx)
```



```
par(mfrow=c(1,1))
```

The demographic model elucidates a mere 2% of the diversity in the history of suicide, as indicated by a  $R^2$  value of approximately 2%. Of all the demographic characteristics, only ethnicity (Not Hispanic/Latino) had a statistically significant link. People in this category had slightly lower suicide history scores. Other demographic factors, such as age, gender, ethnicity, and wealth, did not significantly forecast suicide history. In general, demographic traits by themselves are not good indicators of whether someone has ever thought about or attempted suicide.

```
# =====
# PART B: HX_SUICIDE ~ Depression
# =====

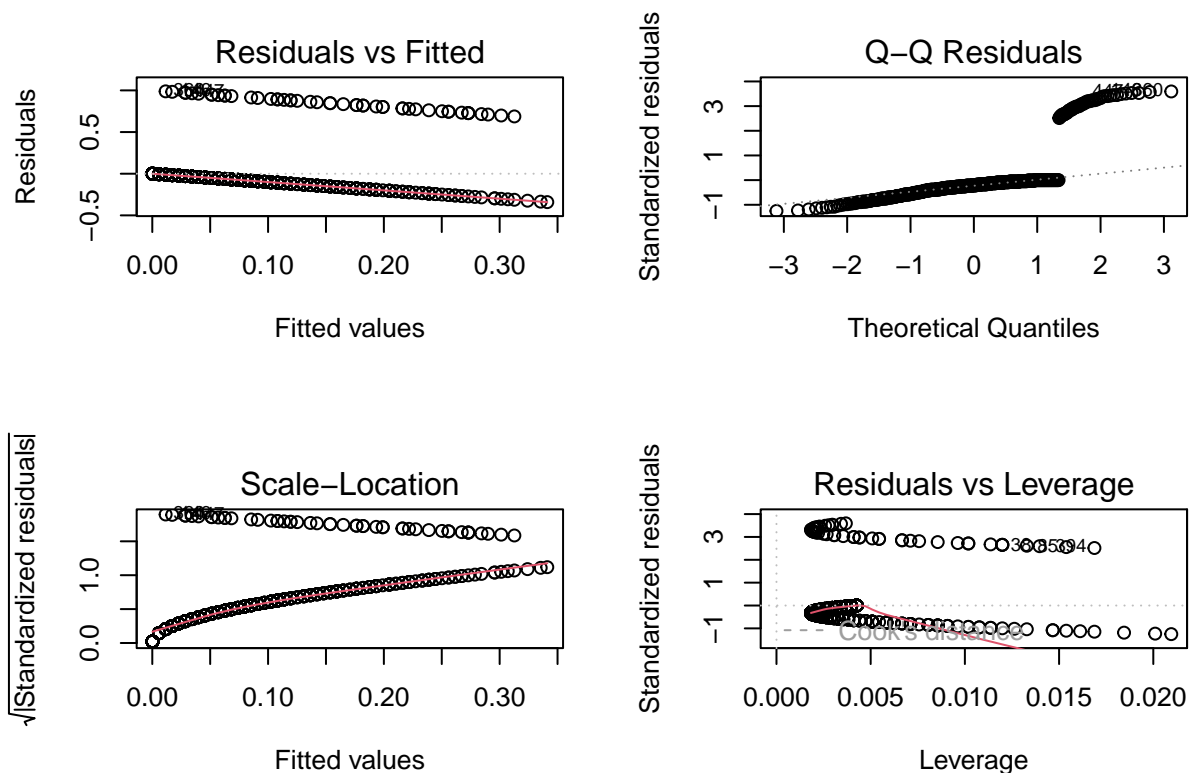
model_b_hx <- lm(HX_SUICIDE ~ CESDR_TOTAL_SUM, data = data)

summary(model_b_hx)

##
## Call:
## lm(formula = HX_SUICIDE ~ CESDR_TOTAL_SUM, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34112 -0.10812 -0.05698 -0.01720  0.98848
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0001492  0.0179847   0.008   0.993
## CESDR_TOTAL_SUM 0.0056829  0.0008614   6.598 9.92e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2755 on 544 degrees of freedom
## Multiple R-squared:  0.07409,    Adjusted R-squared:  0.07239
## F-statistic: 43.53 on 1 and 544 DF,  p-value: 9.92e-11
```

```
# Diagnostics
par(mfrow=c(2,2))
plot(model_b_hx)
```



```
par(mfrow=c(1,1))
```

Depression exhibits a substantial positive correlation with suicide history, suggesting that persons with elevated depressive symptoms are more inclined to disclose a previous suicide attempt or intention. The overall  $R^2$  is only about 7%, but this is still a big step up from the demographic-only model. These results indicate that depression is a significant psychological factor associated with suicide history, even when analyzed in isolation. The diagnostic plots show that the residual patterns are rather consistent, but the outcome distribution is still not quite normal because it is a binary-like variable.

```

# =====
# PART C: HX_SUICIDE ~ ACES 1-10
# =====

aces_vars <- paste0("ACES__", 1:10)

model_c_hx <- lm(as.formula(
  paste("HX_SUICIDE ~", paste(aces_vars, collapse = " + "))
), data = data)

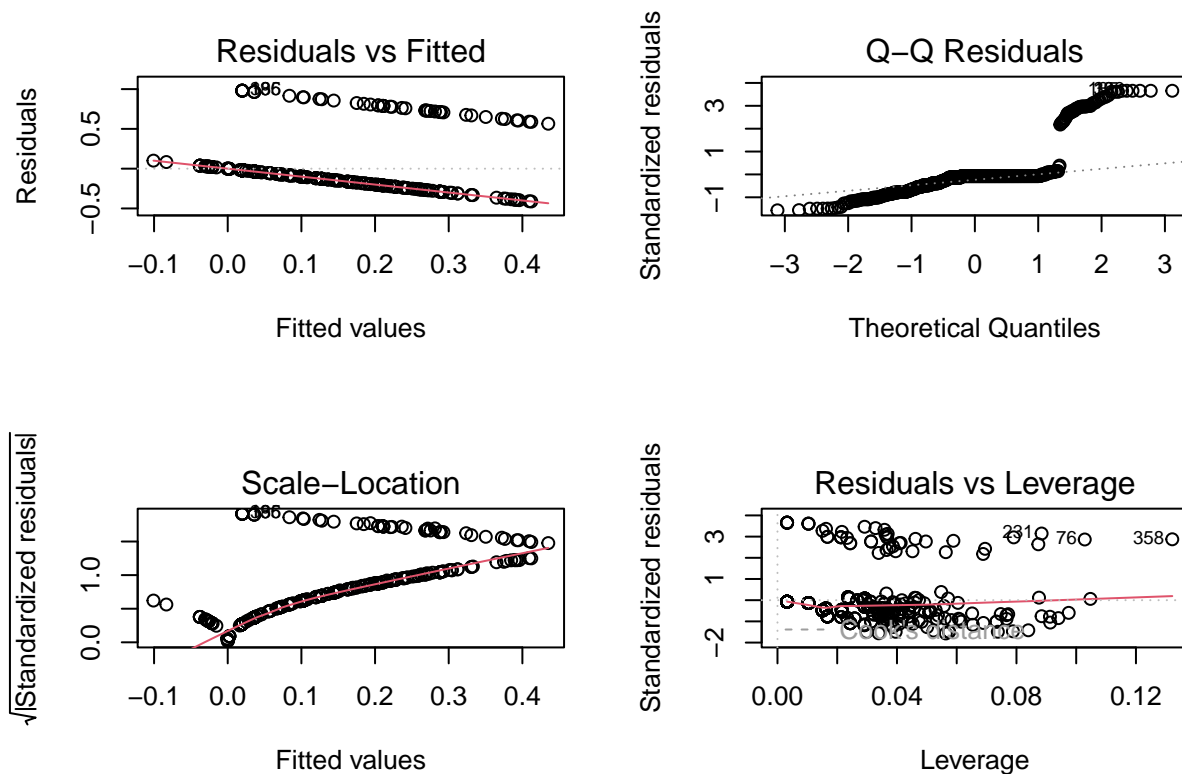
summary(model_c_hx)

##
## Call:
## lm(formula = as.formula(paste("HX_SUICIDE ~", paste(aces_vars,
##      collapse = " + "))), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41211 -0.10457 -0.01968 -0.01968  0.98032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.019678   0.015225   1.293 0.196738
## ACES__1      -0.019562   0.042791  -0.457 0.647742
## ACES__2       0.184569   0.033608   5.492 6.16e-08 ***
## ACES__3       0.083640   0.033510   2.496 0.012860 *
## ACES__4       0.006125   0.075162   0.081 0.935081
## ACES__5      -0.017496   0.041914  -0.417 0.676541
## ACES__6      -0.045473   0.038575  -1.179 0.238993
## ACES__7      -0.057519   0.039317  -1.463 0.144066
## ACES__8       0.107913   0.031925   3.380 0.000777 ***
## ACES__9       0.016308   0.027958   0.583 0.559940
## ACES__10      0.100126   0.059169   1.692 0.091189 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2685 on 535 degrees of freedom
## Multiple R-squared:  0.1355, Adjusted R-squared:  0.1193
## F-statistic: 8.383 on 10 and 535 DF, p-value: 1.011e-12

# Diagnostics
par(mfrow=c(2,2))
plot(model_c_hx)

```





```
par(mfrow=c(1,1))
```

Multiple ACE items were significantly associated with suicide history, including ACEs 2, 3, and 8. This indicates that certain types of childhood adversity—particularly emotional, physical, or household-related trauma—are linked to a higher likelihood of past suicidal behaviors. The model explained about 13% of the variation, suggesting that ACEs play a meaningful but not exclusive role in predicting suicide history. Several ACE items such as ACE 1, 4, and 5 were not significant, showing that not all childhood experiences contribute equally.

```
# =====
# PART D
# =====
orig_dem <- c("AGE", "GENDER", "RACE", "ETHNICITY", "INCOME")
orig_dep <- c("CESDR_TOTAL_SUM")
orig_aces <- paste0("ACES_...", 1:10)
get_sig <- function(model, original_vars) {
  tbl <- summary(model)$coefficients
  sig_terms <- rownames(tbl)[tbl[,4] < 0.20]
  sig_terms <- sig_terms[sig_terms != "(Intercept)"]

  matched <- c()
  for (term in sig_terms) {
    base <- original_vars[sapply(original_vars, function(v) startsWith(term, v))]
    matched <- c(matched, base)
  }
}
```

```

    unique(matched)
  }
  sig_dem <- get_sig(model_a_hx, orig_dem)
  sig_dep <- get_sig(model_b_hx, orig_dep)
  sig_aces <- get_sig(model_c_hx, orig_aces)

  sig_dem; sig_dep; sig_aces

## [1] "ETHNICITY"

## [1] "CESDR_TOTAL_SUM"

## [1] "ACES__2" "ACES__3" "ACES__7" "ACES__8" "ACES__1" "ACES__10"

# Combine all predictors at = 0.20
all_sig_hx <- unique(c(sig_dem, sig_dep, sig_aces))
all_sig_hx

## [1] "ETHNICITY"      "CESDR_TOTAL_SUM" "ACES__2"      "ACES__3"
## [5] "ACES__7"        "ACES__8"        "ACES__1"      "ACES__10"

formula_hx_final <- as.formula(
  paste("HX_SUICIDE ~", paste(all_sig_hx, collapse = " + "))
)

model_hx_final <- lm(formula_hx_final, data = data)

summary(model_hx_final)

##
## Call:
## lm(formula = formula_hx_final, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51035 -0.12176 -0.02784  0.02459  1.01653
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.0333775   0.0355485    0.939  0.34819
## ETHNICITYNot Hispanic/Latino -0.0700702   0.0343779   -2.038  0.04202 *
## CESDR_TOTAL_SUM    0.0040330   0.0008685    4.644 4.31e-06 ***
## ACES__2          0.1656330   0.0327465    5.058 5.82e-07 ***
## ACES__3          0.0453193   0.0313620    1.445  0.14903
## ACES__7         -0.0758090   0.0371180   -2.042  0.04160 *
## ACES__8          0.0899596   0.0312015    2.883  0.00409 **
## ACES__1         -0.0253719   0.0404691   -0.627  0.53096
## ACES__10         0.1254923   0.0565180    2.220  0.02681 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## Residual standard error: 0.2621 on 537 degrees of freedom
## Multiple R-squared:  0.1729, Adjusted R-squared:  0.1606
## F-statistic: 14.03 on 8 and 537 DF,  p-value: < 2.2e-16

library(broom)
library(dplyr)
library(knitr)
library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##      group_rows

# Create tidy table with confidence intervals
final_table_hx <- tidy(model_hx_final, conf.int = TRUE) %>%
  mutate(
    Estimate = round(estimate, 3),
    Std.Error = round(std.error, 3),
    t_value = round(statistic, 3),
    p_value = case_when(
      p.value < 0.001 ~ "<0.001",
      TRUE ~ sprintf("%.3f", p.value)
    ),
    `95% CI` = paste0("[", round(conf.low, 3), ", ", round(conf.high, 3), "]")
  ) %>%
  select(term, Estimate, Std.Error, t_value, p_value, `95% CI`)

# Clean variable labels
final_table_hx$term <- recode(final_table_hx$term,
  "(Intercept)" = "Intercept",
  "ETHNICITYNot Hispanic/Latino" = "Ethnicity: Not Hispanic/Latino",
  "CESDR_TOTAL_SUM" = "Depression (CESD-R)",
  "ACES___2" = "ACE 2",
  "ACES___3" = "ACE 3",
  "ACES___7" = "ACE 7",
  "ACES___8" = "ACE 8",
  "ACES___1" = "ACE 1",
  "ACES___10" = "ACE 10"
)

# Output decorated table
library(flextable)

##
## Attaching package: 'flextable'

## The following objects are masked from 'package:kableExtra':
##
##      as_image, footnote
```

```

ft <- flextable(final_table_hx)
ft <- autofit(ft)
ft <- add_header_lines(ft, "Table 2D. Final Multivariable Regression Model Predicting History of Suicide")

ft <- footnote(
  ft,
  i = 1, j = 1,
  value = as_paragraph(
    "Note: Coefficients represent regression estimates predicting HX_SUICIDE. Significance indicated by
  ),
  ref_symbols = "a"
)

ft

```

Table 2D. Final Multivariable Regression Model Predicting History of Suicide

term	Estimate	Std.Error	t_value	p_value	95% CI
Intercept <sup>a</sup>	0.033	0.036	0.939	0.348	[-0.036, 0.103]
Ethnicity: Not Hispanic/Latino	-0.070	0.034	-2.038	0.042	[-0.138, -0.003]
Depression (CESD-R)	0.004	0.001	4.644	<0.001	[0.002, 0.006]
ACE 2	0.166	0.033	5.058	<0.001	[0.101, 0.23]
ACE 3	0.045	0.031	1.445	0.149	[-0.016, 0.107]
ACE 7	-0.076	0.037	-2.042	0.042	[-0.149, -0.003]
ACE 8	0.090	0.031	2.883	0.004	[0.029, 0.151]
ACE 1	-0.025	0.040	-0.627	0.531	[-0.105, 0.054]
ACE 10	0.125	0.057	2.220	0.027	[0.014, 0.237]

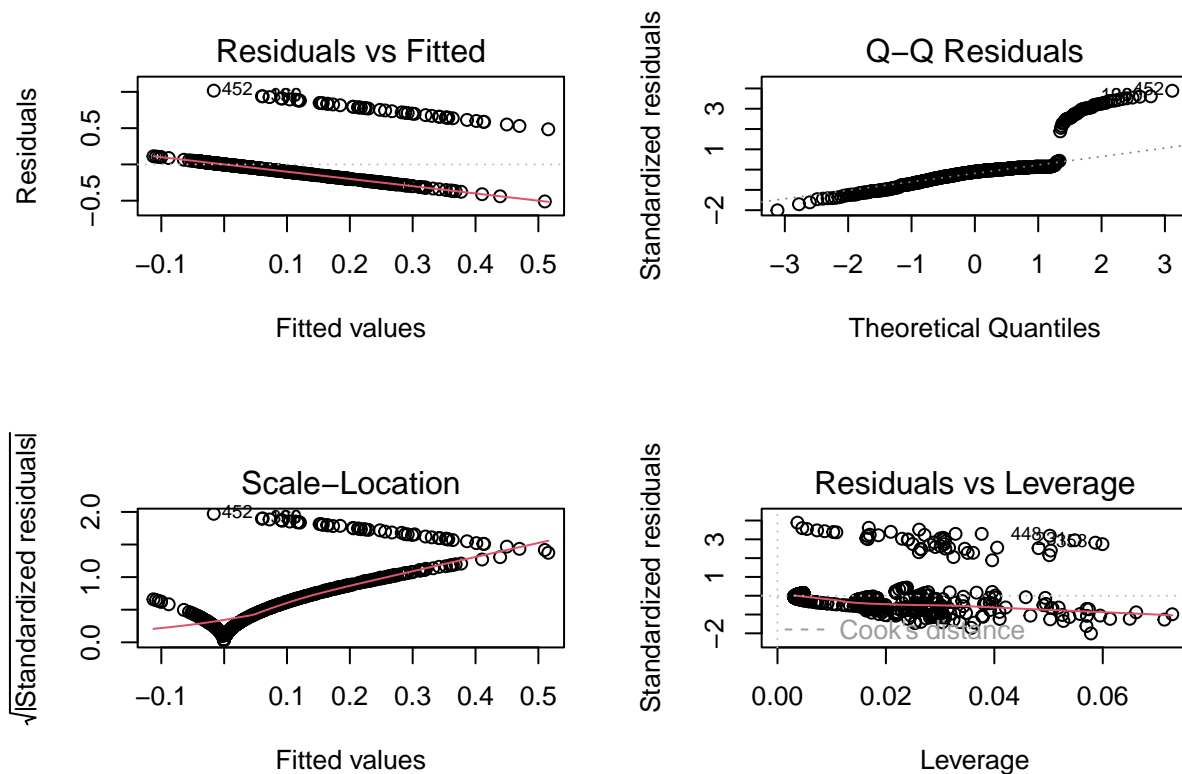
<sup>a</sup>Note: Coefficients represent regression estimates predicting HX\_SUICIDE. Significance indicated by p-values.

The final multivariable model, which includes all the factors chosen at  $\alpha = 0.20$ , explains around 17% of the differences in suicide history, which is better than any of the individual models. Depression continues to be a robust and substantial predictor, underscoring the importance of psychological distress in comprehending suicide risk. Certain ACE components, specifically ACEs 2, 7, 8, and 10, maintained their significance, indicating that particular childhood traumas persist in affecting suicide history, even after controlling for depression and demographic variables. Ethnicity continued to be a prominent demographic variable. These results indicate that a history of suicide is affected by current mental health and past life challenges.

```

# =====
# PART E
# =====
# Diagnostics
par(mfrow=c(2,2))
plot(model_hx_final)

```



```
par(mfrow=c(1,1))

# Normality test
shapiro.test(resid(model_hx_final))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model_hx_final)
## W = 0.71189, p-value < 2.2e-16
```

The residual diagnostics indicate that the model does not completely satisfy the assumption of normality, as validated by the Shapiro-Wilk test ( $p < .00000000000000022$ ). This is to be expected because HX\_SUICIDE is almost binary (0/1-like), which means that complete normal residuals are not statistically possible. Even yet, the model still gives useful information on what causes people to commit suicide. The residual plot does not show any strong signs of heteroscedasticity or extreme outliers, which means that the model is quite stable. Overall, the model is still useful for finding crucial correlates, even though the normalcy assumption isn't ideal.

### QUESTION 3.

From Model 1d (SABCS Final Model)

Significant Predictor: CESDR\_TOTAL\_SUM

Beta: 0.18484

Interpretation: Each additional point in depression score is associated with a 0.18-point increase in suicidal risk. This means depressive symptoms are a strong, meaningful predictor of higher SABCS scores.

Non-significant Predictor: ACES\_1

Beta: 0.11834

Interpretation: Although ACES\_1 has a small positive coefficient, it is not statistically significant, meaning it does not meaningfully predict suicidal risk. Any observed relationship may simply be due to chance.

## **From Model 2d (HX\_SUICIDE Final Model)**

Significant Predictor: CESDR\_TOTAL\_SUM

Beta: 0.004033

Interpretation: Each 1-point increase in depression score is associated with a small increase in suicide history score, meaning individuals with more depressive symptoms are more likely to report a prior suicide attempt or ideation.

Non-significant Predictor: ACES\_1

Beta: -0.02537

Interpretation: ACES\_1 is not statistically significant, indicating that this type of childhood adversity does not meaningfully predict past suicidal behavior when other variables are considered.

## **QUESTION 4.**

### **1. Study Design**

The structure of this dataset aligns most closely with a cross-sectional study, since all variables—life experiences, mental health measures, and suicide-related outcomes—were collected at a single point in time. It feels like a snapshot of each participant’s life at that moment, capturing their history and current emotional state simultaneously. Because nothing is measured across time, it does not function like a cohort or case-control design.

### **2. Limitation (1–2 sentences, humanized English)**

A major limitation of this “single-moment” approach is that it cannot reveal which experiences came first, making it hard to understand the true sequence of events. Without a sense of timing, any relationship we see may reflect correlation rather than a meaningful causal story.

### **3. How to Redesign the Study (3–4 sentences, humanized English)**

To address this issue, the study could be redesigned as a longitudinal cohort, where participants are followed over time and their emotional and environmental changes are recorded as they unfold. Tracking people across meaningful life stages would help establish whether depression, stress, or adversity occur before shifts in suicide risk. This design respects the fact that human experiences evolve rather than appear all at once. By observing these changes directly, we gain a clearer and more compassionate understanding of how struggles develop and how risks emerge over time.