

Effect of Medicaid Expansion on the Medicare Population Analysis

Section I: Model 1 - HCC ~ expansion status (Linear Regression)

The hierarchical condition category (HCC) score is a model developed by CMS to assign risk scores for patients based on health conditions. The risk score accounts for factors such as hospital admissions for acute and chronic issues, as well as various classes of diseases and conditions ranging in severity. The national average score is the reference point, and is equal to 1. Patients in poorer health than the average beneficiary would have a score greater than 1, whereas a patient with better health than the average beneficiary would have a score closer to 0 according to the CMS HCC model.

Although it is commonly used by insurance companies to estimate future patient costs, this metric can serve as a proxy to quantify the average individual health of a community. The average HCC score across U.S. counties is roughly normally distributed, with a median value of 0.96. The distribution is slightly right skewed, and a log transform might be considered if the model violates the LINE assumptions.

Exploratory Analysis

```
summary(county_dat$average_hcc_score)

##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.6000  0.9000  0.9600  0.9625  1.0200  1.5700

length(county_dat$average_hcc_score)

## [1] 3046

ggplot() +
  geom_histogram(aes(county_dat$average_hcc_score), col=1) +
  ggtitle('Normality of Outcome') +
  xlab('average hcc score')

## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

ggplot() +
  geom_histogram(aes(log10(county_dat$average_hcc_score)), fill=2, col =1) +
  ggtitle('Normality of Outcome (log)') +
  xlab('log(average hcc score)')

## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Normality of Outcome

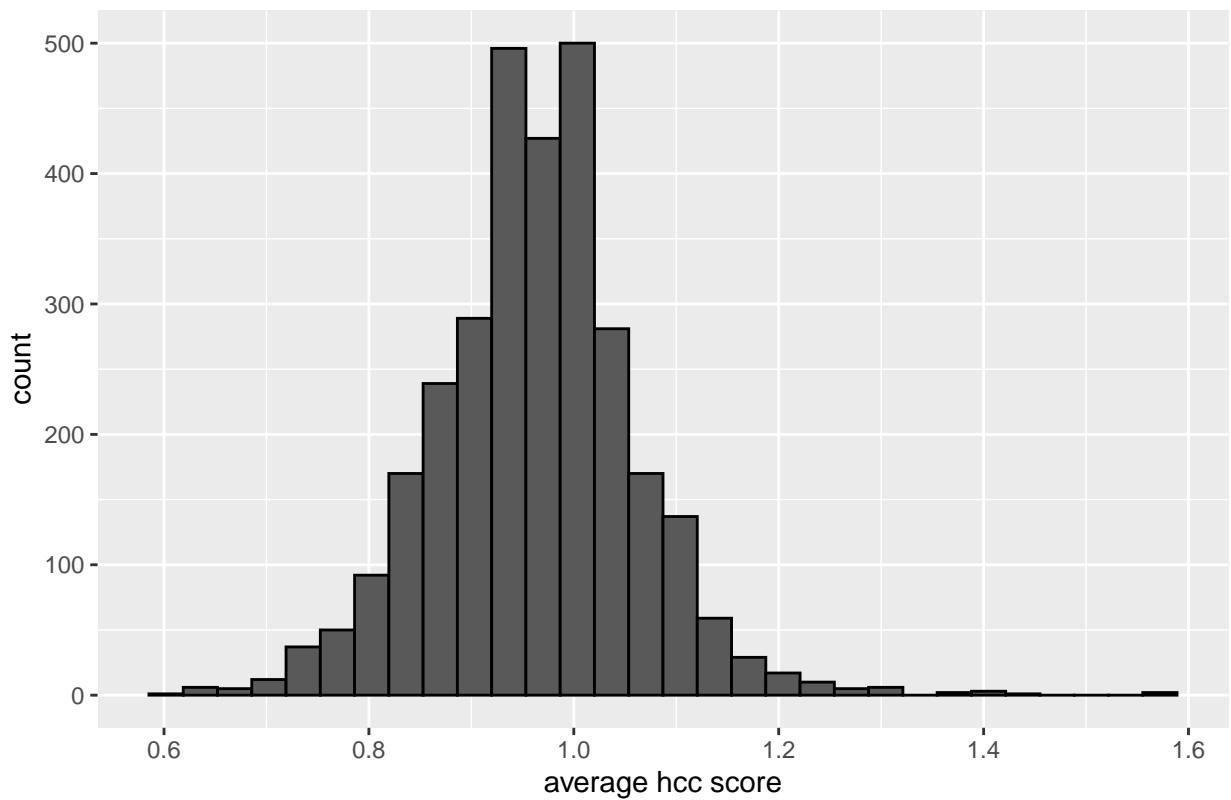


Figure 1: Figure 1

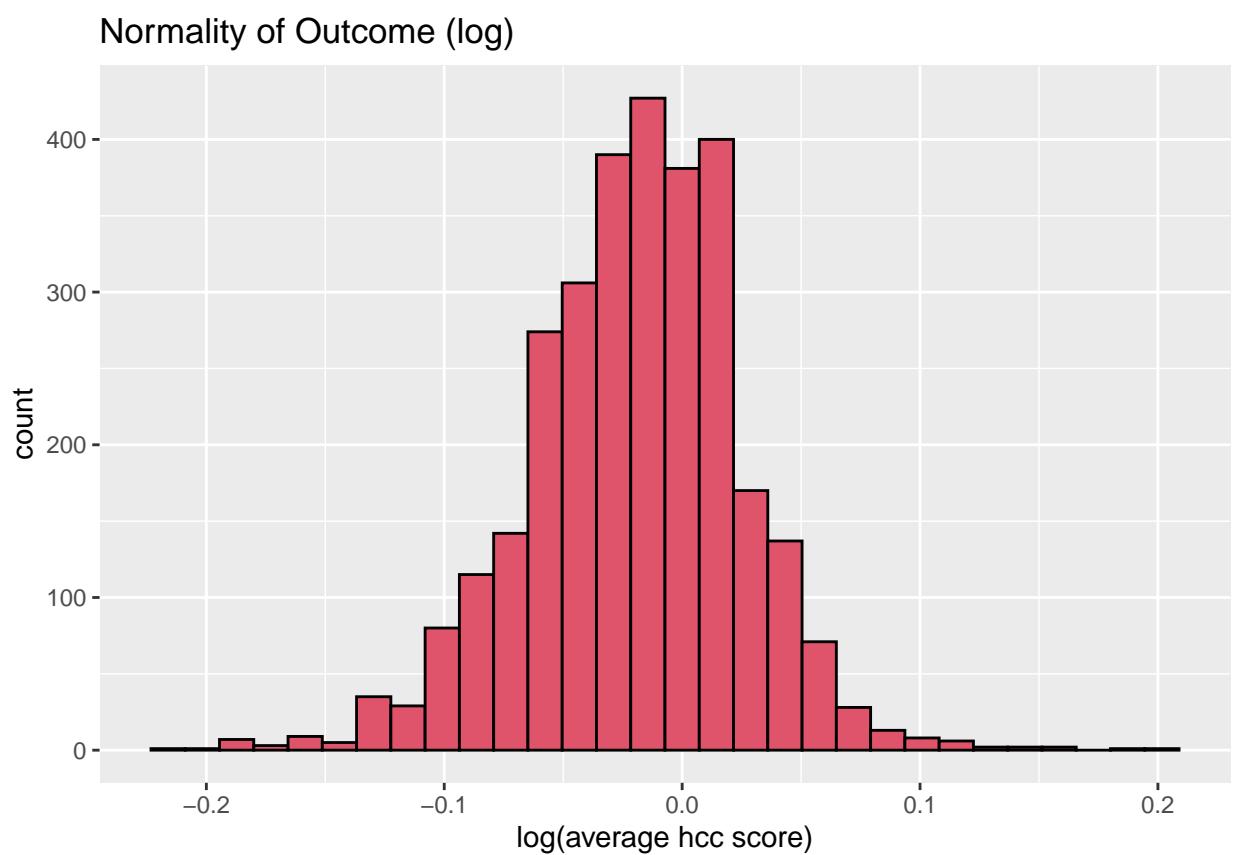
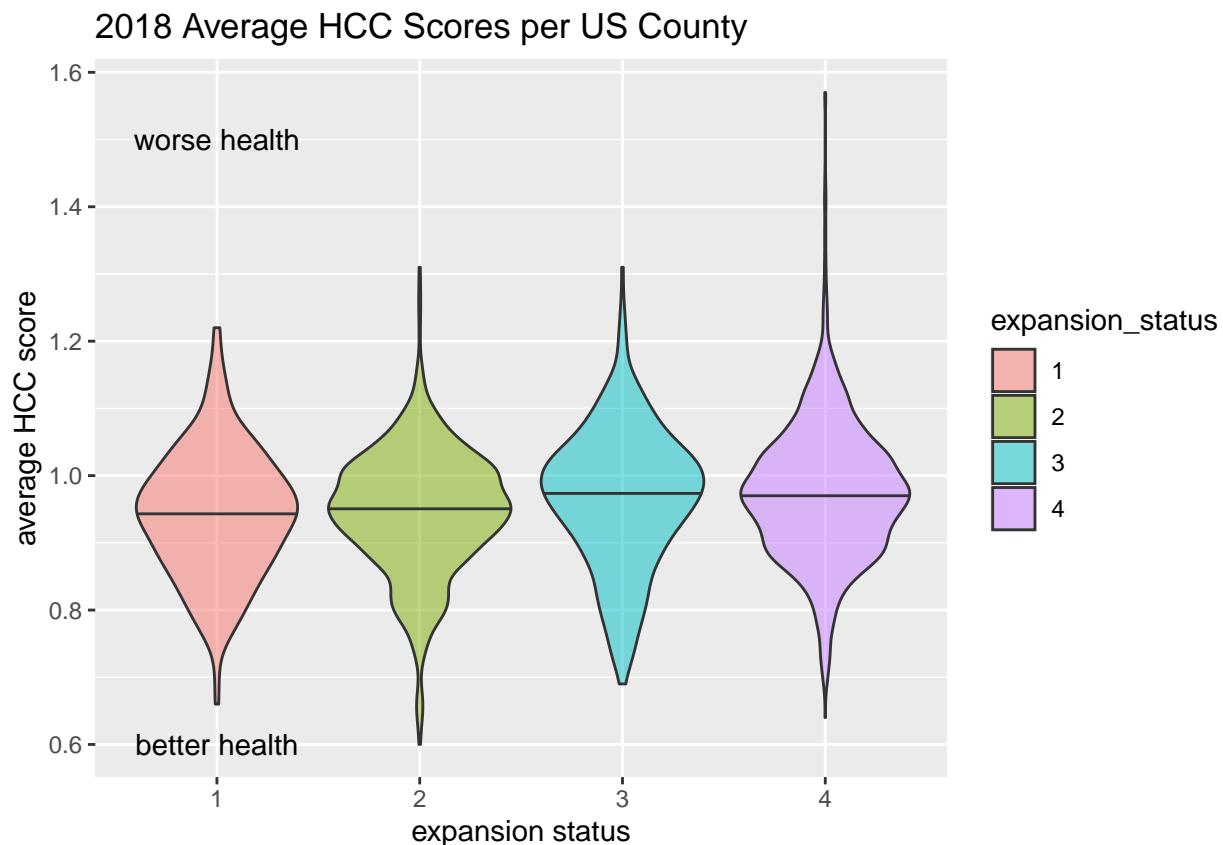


Figure 2: Figure 2

The primary covariate of interest is expansion status, and average age & percent eligible for medicaid are other important factors that should be adjusted for. There seems to be a slight difference in average HCC scores in the states that have not expanded & states that have expanded Medicaid, with larger variance in the counties that have not expanded.

```
county_dat %>% ggplot() +
  geom_violin(aes(expansion_status, average_hcc_score,
                  fill = expansion_status), alpha = 0.5,
              draw_quantiles = c(0.5)) +
  #scale_y_reverse() +
  ylab('average HCC score') +
  xlab('expansion status') +
  annotate("text", label = "better health", x = 1, y = 0.6) +
  annotate("text", label = "worse health", x = 1, y = 1.5) +
  ggttitle('2018 Average HCC Scores per US County')
```



Model Considerations

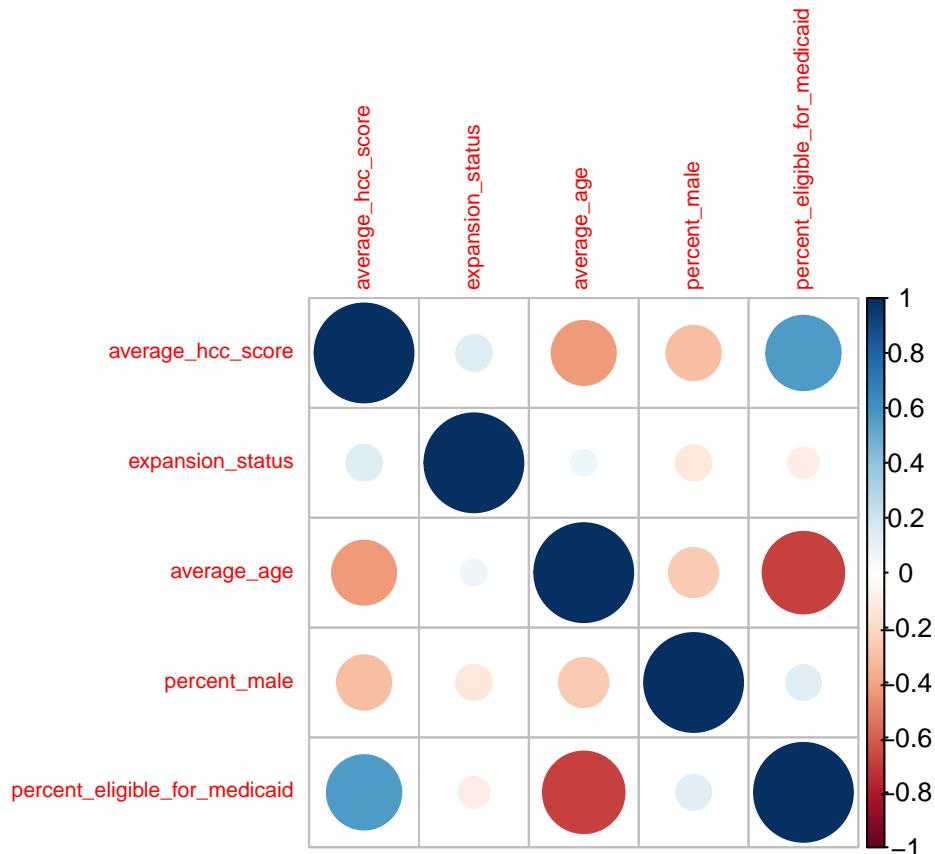
```
vars <- c("average_hcc_score",
        "expansion_status",
        "average_age",
        "percent_male",
        "percent_eligible_for_medicaid")
```

```
covars <- county_dat %>% select(vars) %>%
  mutate(expansion_status = as.integer(expansion_status)) %>% drop_na()
```

Assess Collinearity of Potential Covariates

```
## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(vars)' instead of 'vars' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```
corrplot(cor(covars), tl.cex = 0.7)
```



```
mlr_red <- lm(average_hcc_score ~ as.factor(expansion_status),
               data = county_dat)
summary(mlr_red)
```

```
##
## Call:
## lm(formula = average_hcc_score ~ as.factor(expansion_status),
##      data = county_dat)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -0.34534 -0.06376  0.00466  0.05722  0.59624
```

```

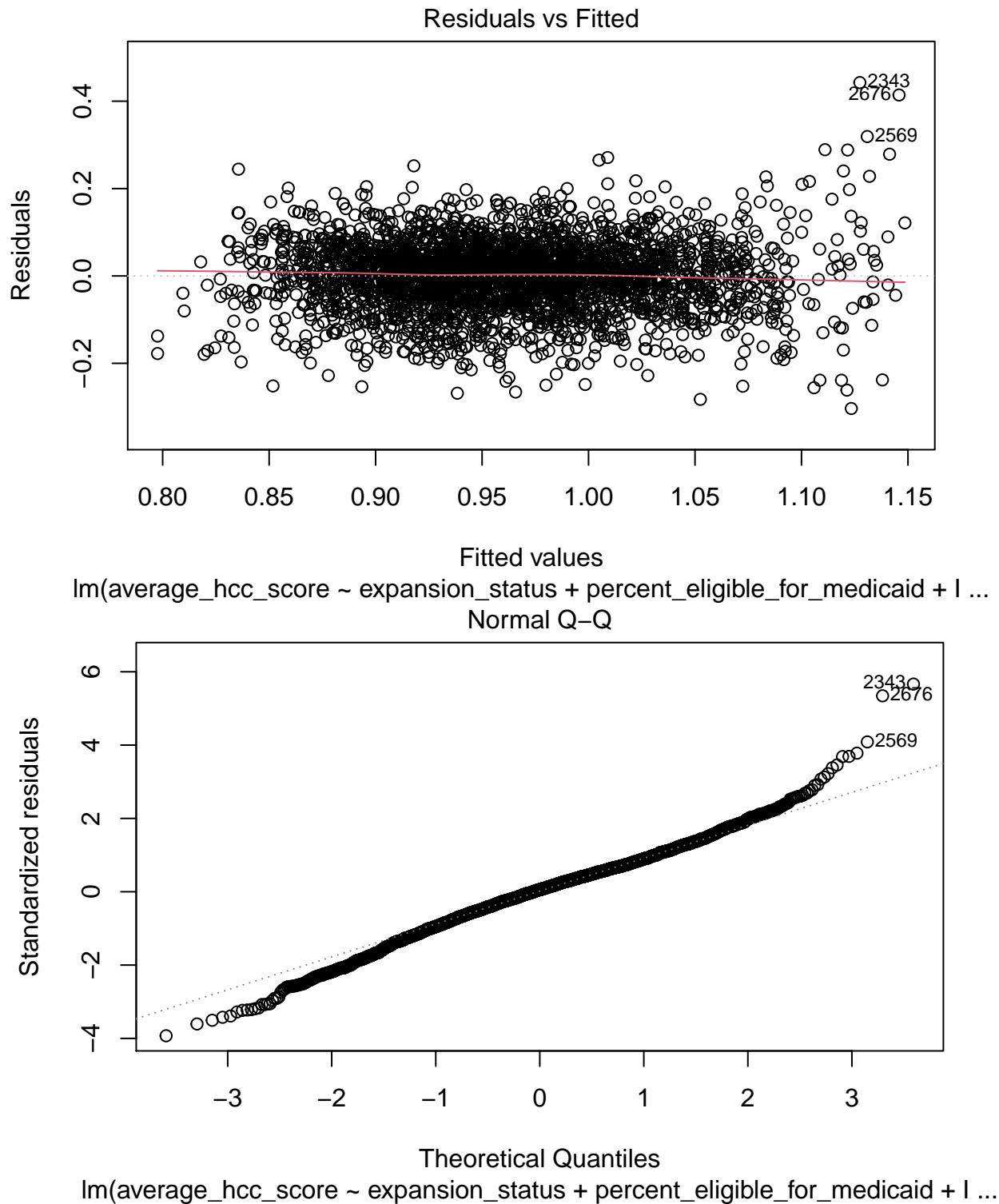
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|) 
## (Intercept)            0.942775  0.006765 139.351 < 2e-16 ***
## as.factor(expansion_status)2 0.002566  0.007539   0.340  0.73360
## as.factor(expansion_status)3 0.022435  0.008423   2.664  0.00777 ** 
## as.factor(expansion_status)4 0.030987  0.007196   4.306 1.71e-05 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.09781 on 3042 degrees of freedom
## Multiple R-squared:  0.01824,    Adjusted R-squared:  0.01727 
## F-statistic: 18.84 on 3 and 3042 DF,  p-value: 4.171e-12

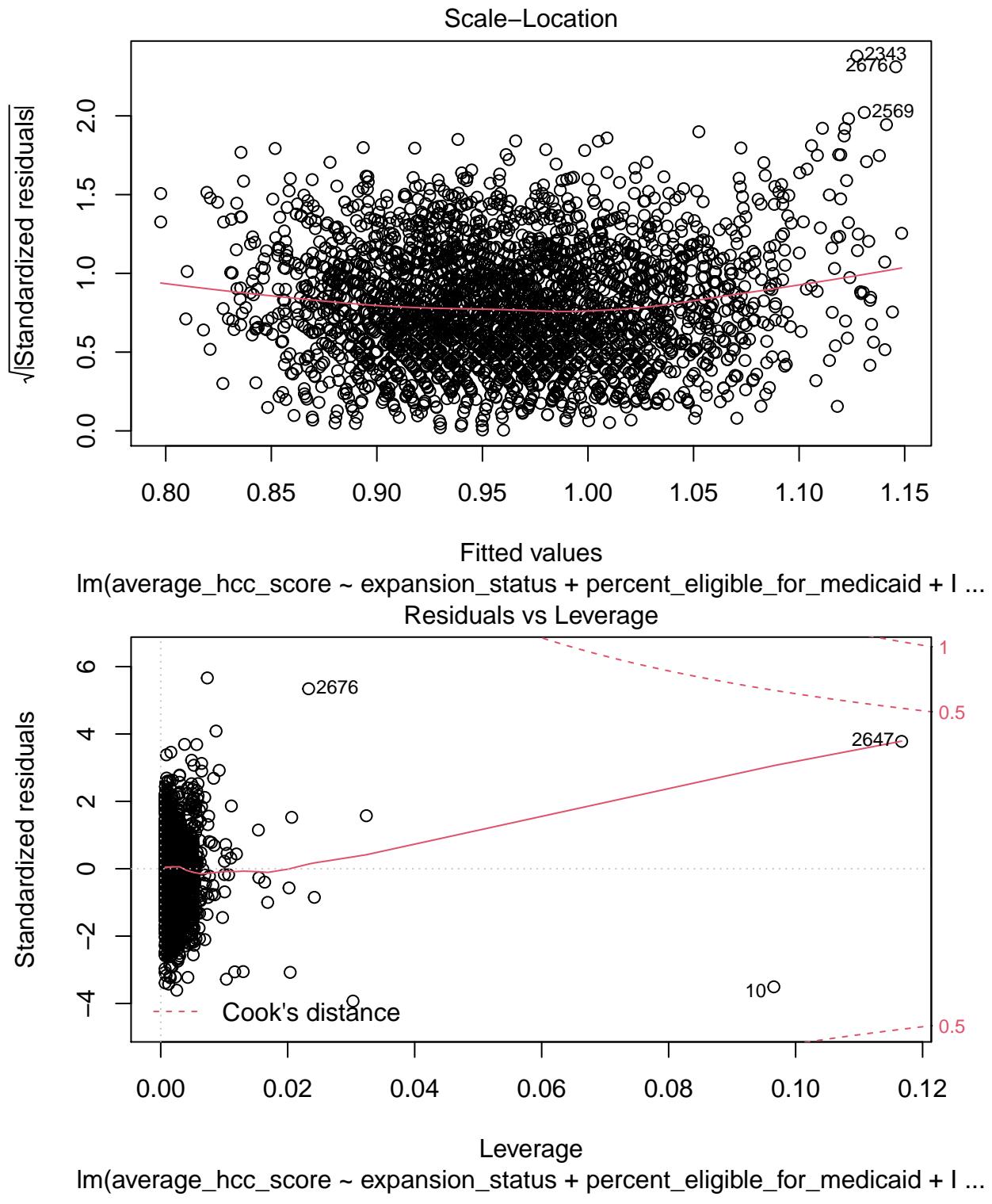
mlr <- lm(average_hcc_score ~ expansion_status +
           percent_eligible_for_medicaid +
           I(percent_eligible_for_medicaid**2),
           data = county_dat)
summary(mlr)

## 
## Call:
## lm(formula = average_hcc_score ~ expansion_status + percent_eligible_for_medicaid +
##     I(percent_eligible_for_medicaid^2), data = county_dat)
## 
## Residuals:
##      Min       1Q       Median       3Q       Max 
## -0.30341 -0.04604  0.00457  0.04884  0.44259 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|) 
## (Intercept)            0.744158  0.009241 80.527 < 2e-16 ***
## expansion_status2      0.009571  0.006050   1.582 0.113791
## expansion_status3      0.025738  0.006755   3.810 0.000141 *** 
## expansion_status4      0.049787  0.005794   8.592 < 2e-16 *** 
## percent_eligible_for_medicaid 1.116492  0.060283 18.521 < 2e-16 *** 
## I(percent_eligible_for_medicaid^2) -0.875561  0.113795 -7.694 1.91e-14 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.07843 on 3040 degrees of freedom
## Multiple R-squared:  0.3692, Adjusted R-squared:  0.3681 
## F-statistic: 355.8 on 5 and 3040 DF,  p-value: < 2.2e-16

plot(mlr)

```





The log transform does not remedy the normality violation of the model.

```
mlr_log <- lm(log(average_hcc_score) ~ expansion_status +
  # average_age + I(average_age**2) +
  percent_eligible_for_medicaid +
```

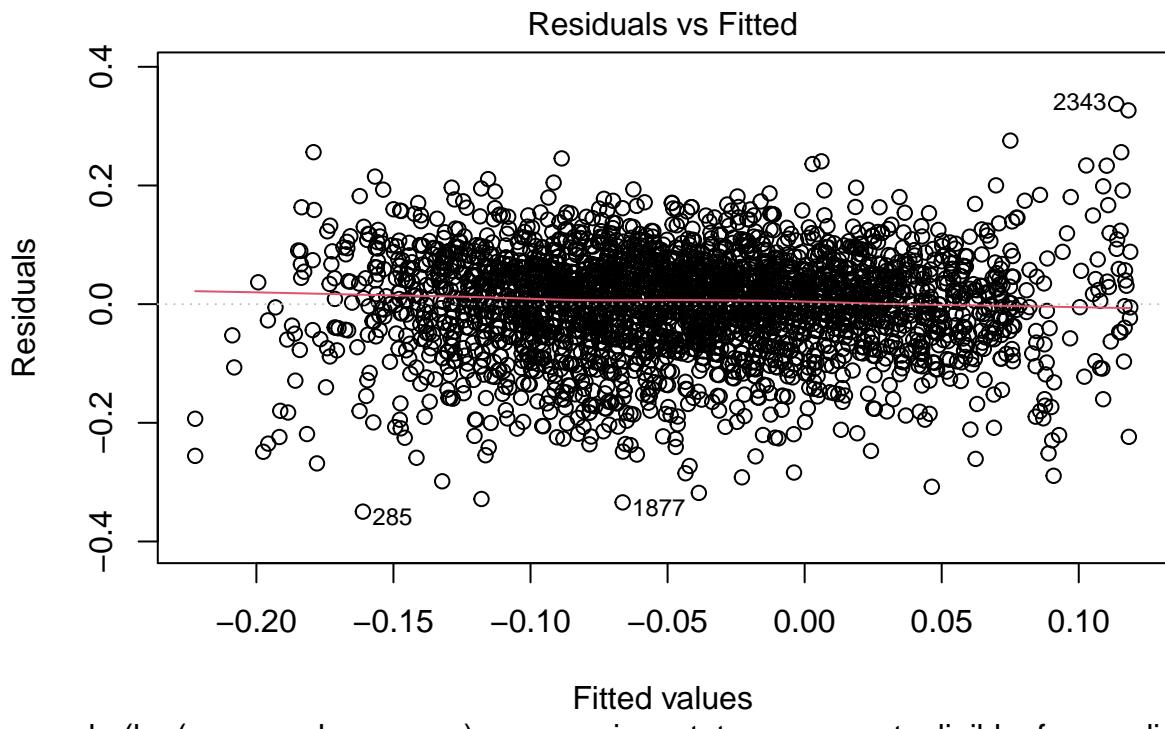
```

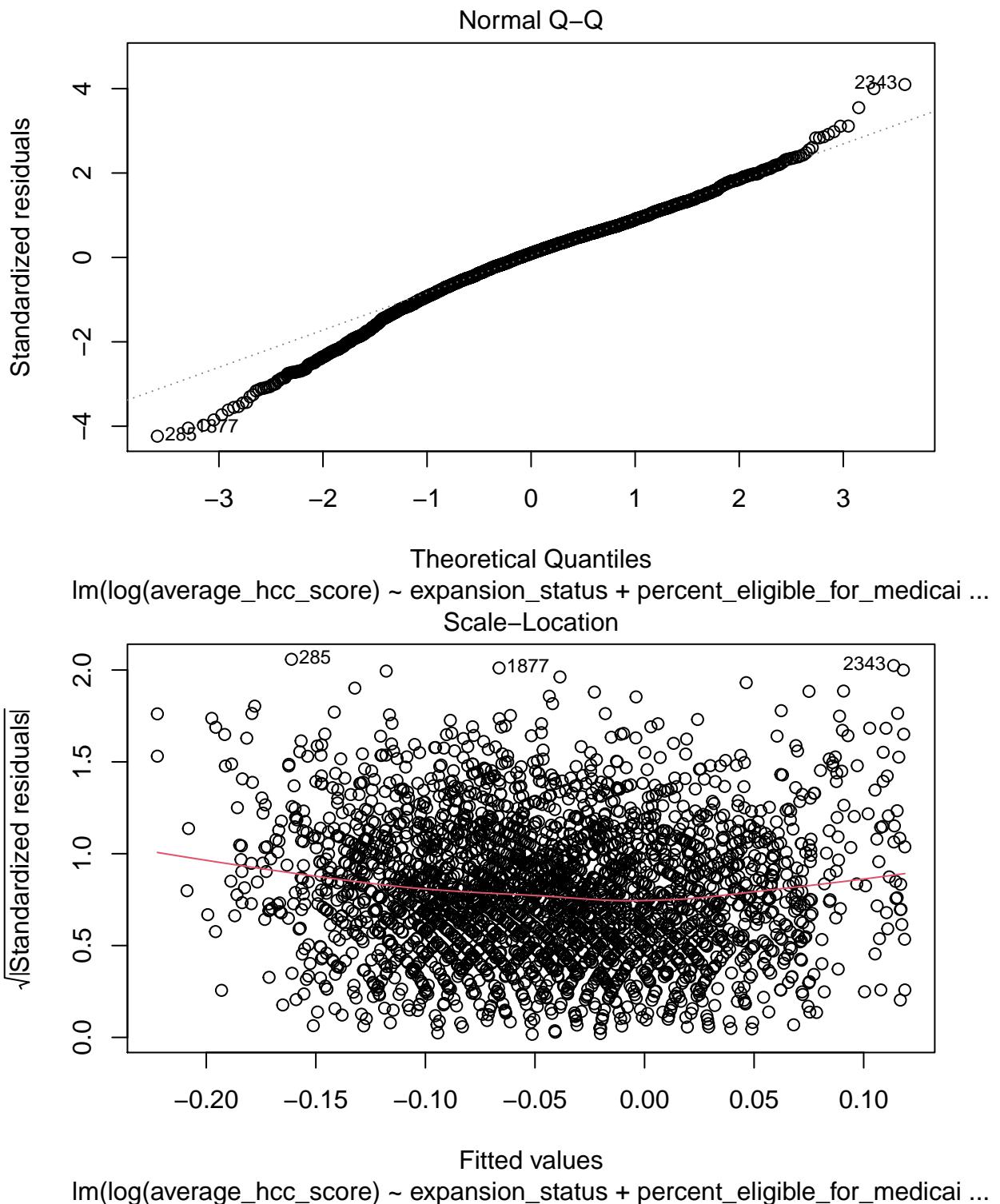
I(percent_eligible_for_medicaid**2),
  data = county_dat)
summary(mlr_log)

##
## Call:
## lm(formula = log(average_hcc_score) ~ expansion_status + percent_eligible_for_medicaid +
##     I(percent_eligible_for_medicaid^2), data = county_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.34968 -0.04565  0.00795  0.05264  0.33740 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)              -0.283270  0.009737 -29.092 < 2e-16 ***
## expansion_status2         0.010625  0.006375  1.667 0.095684 .  
## expansion_status3         0.026142  0.007117  3.673 0.000244 *** 
## expansion_status4         0.052189  0.006105  8.548 < 2e-16 *** 
## percent_eligible_for_medicaid 1.287246  0.063518 20.266 < 2e-16 *** 
## I(percent_eligible_for_medicaid^2) -1.183687  0.119902 -9.872 < 2e-16 *** 
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.08264 on 3040 degrees of freedom
## Multiple R-squared:  0.3603, Adjusted R-squared:  0.3593 
## F-statistic: 342.5 on 5 and 3040 DF,  p-value: < 2.2e-16

plot(mlr_log)

```







Consider average age & gender in model; both terms are significant & improve Adjusted R squared value. However, the linearity assumption seems to be violated based on the residuals vs fitted values plot. Also, the negative coefficients seem to be counterintuitive (HCC score would be expected to increase with increased age, not decrease as suggested by the negative coefficient).

```
mlr2 <- lm(average_hcc_score ~ expansion_status +
            average_age +
            percent_male +
            percent_eligible_for_medicaid +
            I(percent_eligible_for_medicaid**2),
            data = county_dat)
summary(mlr2)

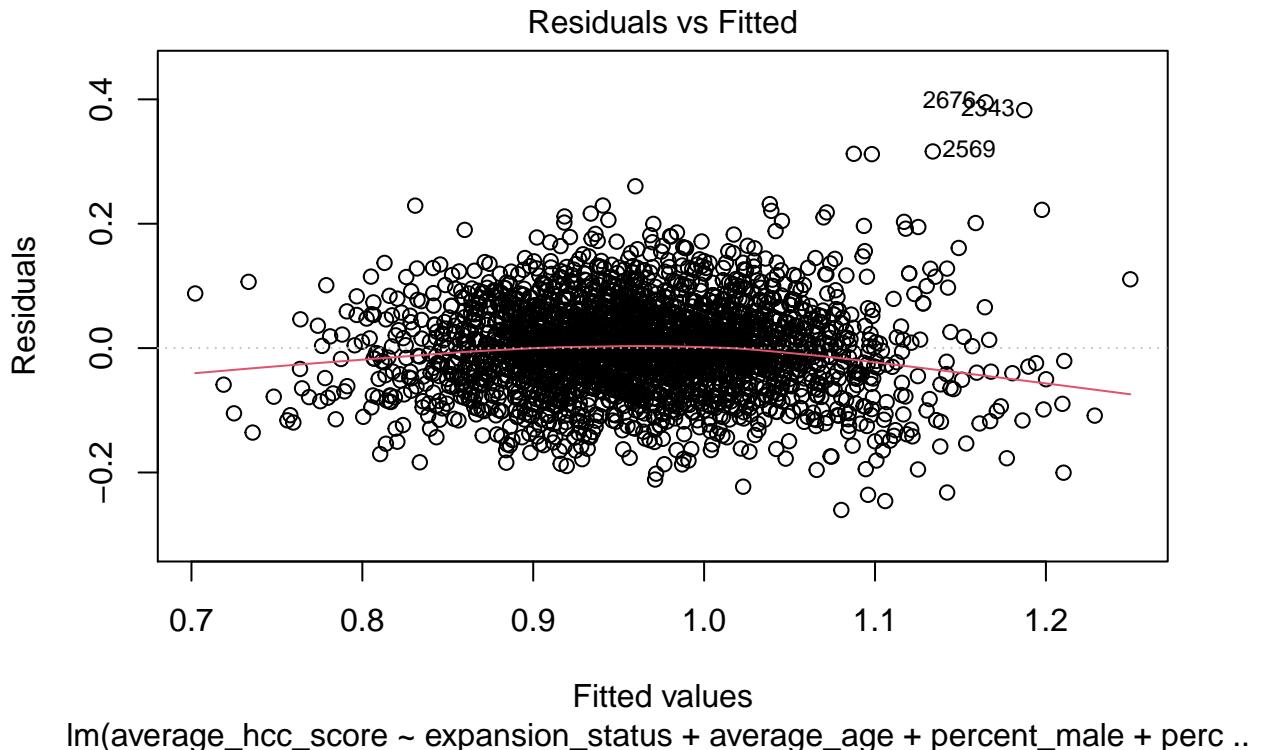
##
## Call:
## lm(formula = average_hcc_score ~ expansion_status + average_age +
##     percent_male + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##     data = county_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.26029 -0.04249  0.00016  0.04124  0.39525 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                2.3482968  0.0956672 24.547 < 2e-16 ***
## expansion_status2          0.0006056  0.0053668  0.113  0.9102    
## expansion_status3          0.0110246  0.0060051  1.836  0.0665 .  
## expansion_status4          0.0327772  0.0051644  6.347 2.53e-10 ***
```

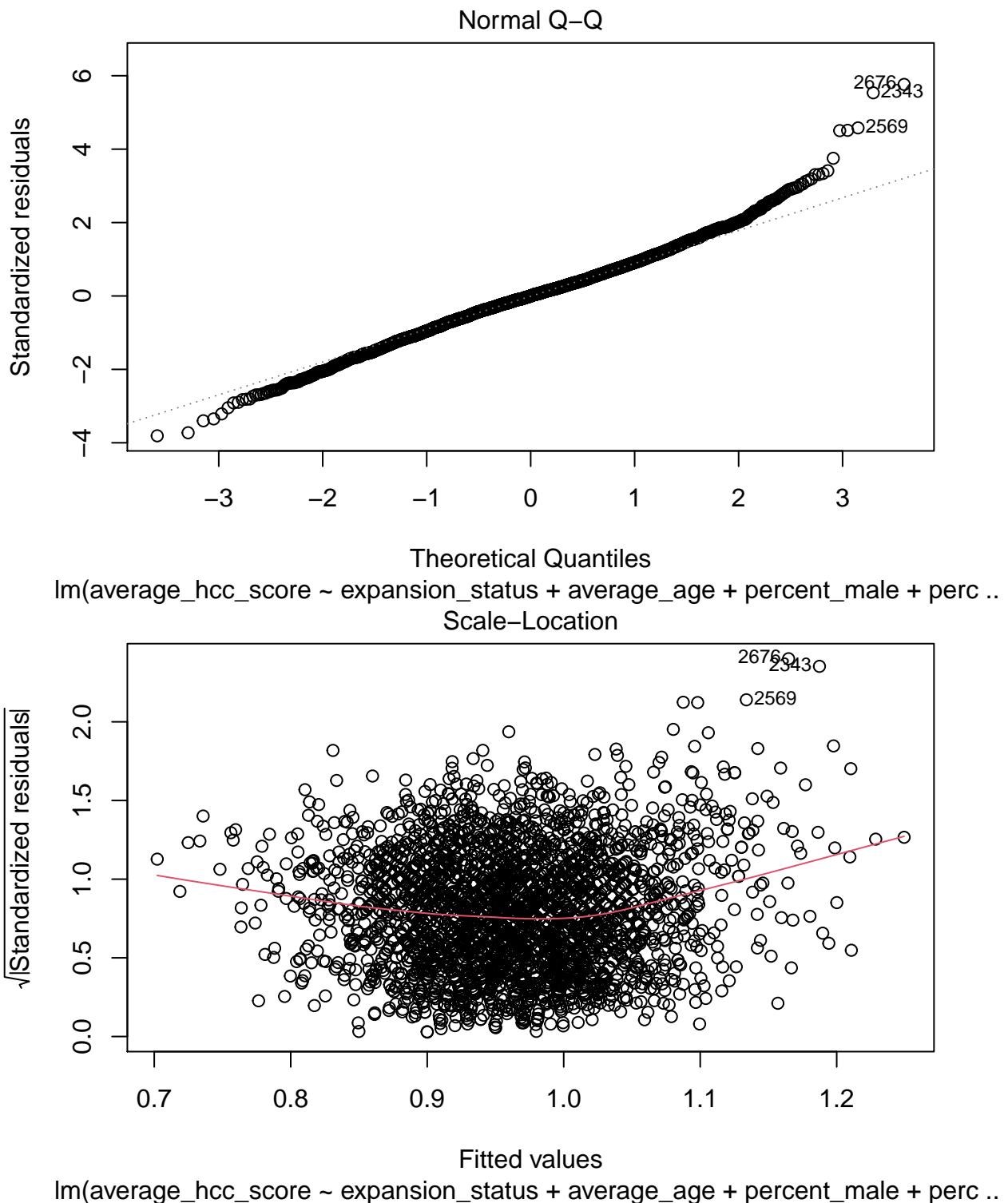
```

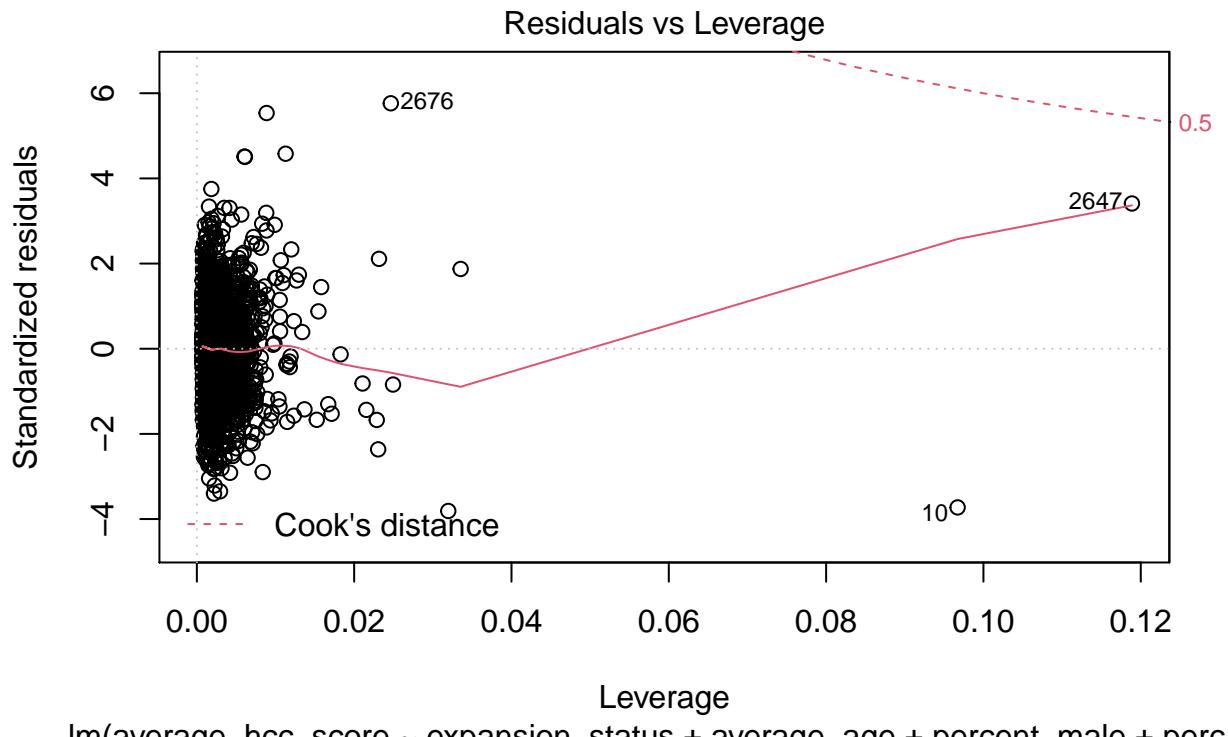
## average_age           -0.0099758  0.0011070  -9.012  < 2e-16 ***
## percent_male          -1.8028098  0.0623103 -28.933  < 2e-16 ***
## percent_eligible_for_medicaid 0.8197113  0.0603955 13.572  < 2e-16 ***
## I(percent_eligible_for_medicaid^2) -0.4482033  0.1047364  -4.279  1.93e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06944 on 3038 degrees of freedom
## Multiple R-squared:  0.5058, Adjusted R-squared:  0.5047
## F-statistic: 444.3 on 7 and 3038 DF,  p-value: < 2.2e-16

```

```
plot(mlr2)
```







The model with the additional linear & quadratic % eligible for medicaid terms is needed to explain the variation in the data. The reduced model is not sufficient (F-test $p < 0.05$).

Additionally, the model with the age & gender terms explain more variation in the data than the model considering only the linear & quadratic % eligible for medicaid terms.

```
anova(mlr_red, mlr)
```

```
## Analysis of Variance Table
##
## Model 1: average_hcc_score ~ as.factor(expansion_status)
## Model 2: average_hcc_score ~ expansion_status + percent_eligible_for_medicaid +
##           I(percent_eligible_for_medicaid^2)
##   Res.Df   RSS Df Sum of Sq    F    Pr(>F)
## 1   3042 29.101
## 2   3040 18.699  2     10.402 845.54 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(mlr, mlr2)
```

```
## Analysis of Variance Table
##
## Model 1: average_hcc_score ~ expansion_status + percent_eligible_for_medicaid +
##           I(percent_eligible_for_medicaid^2)
## Model 2: average_hcc_score ~ expansion_status + average_age + percent_male +
##           percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2)
##   Res.Df   RSS Df Sum of Sq    F    Pr(>F)
## 1   3040 18.699
```

```

## 2 3038 14.648 2 4.0512 420.12 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1

```

```
AIC(mlr2)
```

```
## [1] -7595.246
```

```
AIC(mlr)
```

```
## [1] -6855.463
```

Assess Effect Modification: The interaction terms are not statistically significant ($p>0.05$) when examining whether expansion status modifies the % eligible for medicaid in each county.

```

mlr_em <- lm(average_hcc_score ~ expansion_status +
               average_age +
               percent_male +
               percent_eligible_for_medicaid +
               I(percent_eligible_for_medicaid**2) +
               expansion_status*percent_eligible_for_medicaid,
               data = county_dat)
summary(mlr_em)

```

```

##
## Call:
## lm(formula = average_hcc_score ~ expansion_status + average_age +
##     percent_male + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2) +
##     expansion_status * percent_eligible_for_medicaid, data = county_dat)
##
## Residuals:
##      Min        1Q        Median       3Q        Max 
## -0.27287 -0.04207  0.00053  0.04186  0.38448 
##
## Coefficients:
##                                         Estimate Std. Error t value
## (Intercept)                         2.376045  0.096294 24.675
## expansion_status2                   0.008369  0.015373  0.544
## expansion_status3                  -0.003638  0.017408 -0.209
## expansion_status4                   0.019711  0.014726  1.339
## average_age                          -0.010376  0.001115 -9.309
## percent_male                         -1.784617  0.062531 -28.540
## percent_eligible_for_medicaid       0.783964  0.083827  9.352
## I(percent_eligible_for_medicaid^2)  -0.446826  0.105314 -4.243
## expansion_status2:percent_eligible_for_medicaid -0.037226  0.065485 -0.568
## expansion_status3:percent_eligible_for_medicaid  0.067180  0.074433  0.903
## expansion_status4:percent_eligible_for_medicaid  0.063294  0.063084  1.003
##                                         Pr(>|t|)    
## (Intercept)                         < 2e-16 ***
## expansion_status2                   0.586    
## expansion_status3                   0.834    

```

```

## expansion_status4          0.181
## average_age                < 2e-16 ***
## percent_male               < 2e-16 ***
## percent_eligible_for_medicaid < 2e-16 ***
## I(percent_eligible_for_medicaid^2) 2.27e-05 ***
## expansion_status2:percent_eligible_for_medicaid   0.570
## expansion_status3:percent_eligible_for_medicaid   0.367
## expansion_status4:percent_eligible_for_medicaid   0.316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06936 on 3035 degrees of freedom
## Multiple R-squared:  0.5074, Adjusted R-squared:  0.5058
## F-statistic: 312.6 on 10 and 3035 DF,  p-value: < 2.2e-16

```

The interaction term considering effect modification between average age & percent male is significant, and also results in more intuitive coefficients that align with expected directions for increasing males & older members of the county population.

The linearity of the model is also improved with the addition of the interaction term.

```

mlr_em_age_gender <- lm(average_hcc_score ~ expansion_status +
                        average_age + #I(average_age**2) +
                        percent_male + I(percent_male**2) +
                        percent_eligible_for_medicaid +
                        I(percent_eligible_for_medicaid**2) +
                        average_age*percent_male,
                        data = county_dat)
summary(mlr_em_age_gender)

```

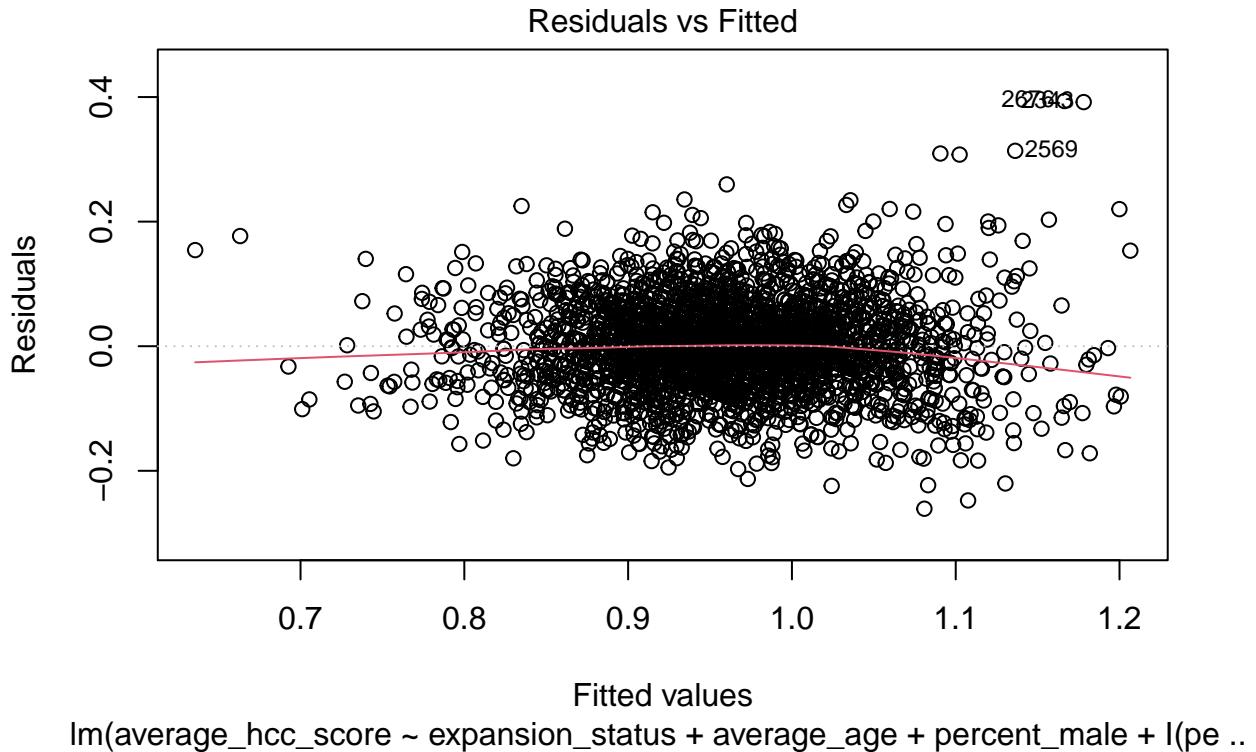
```

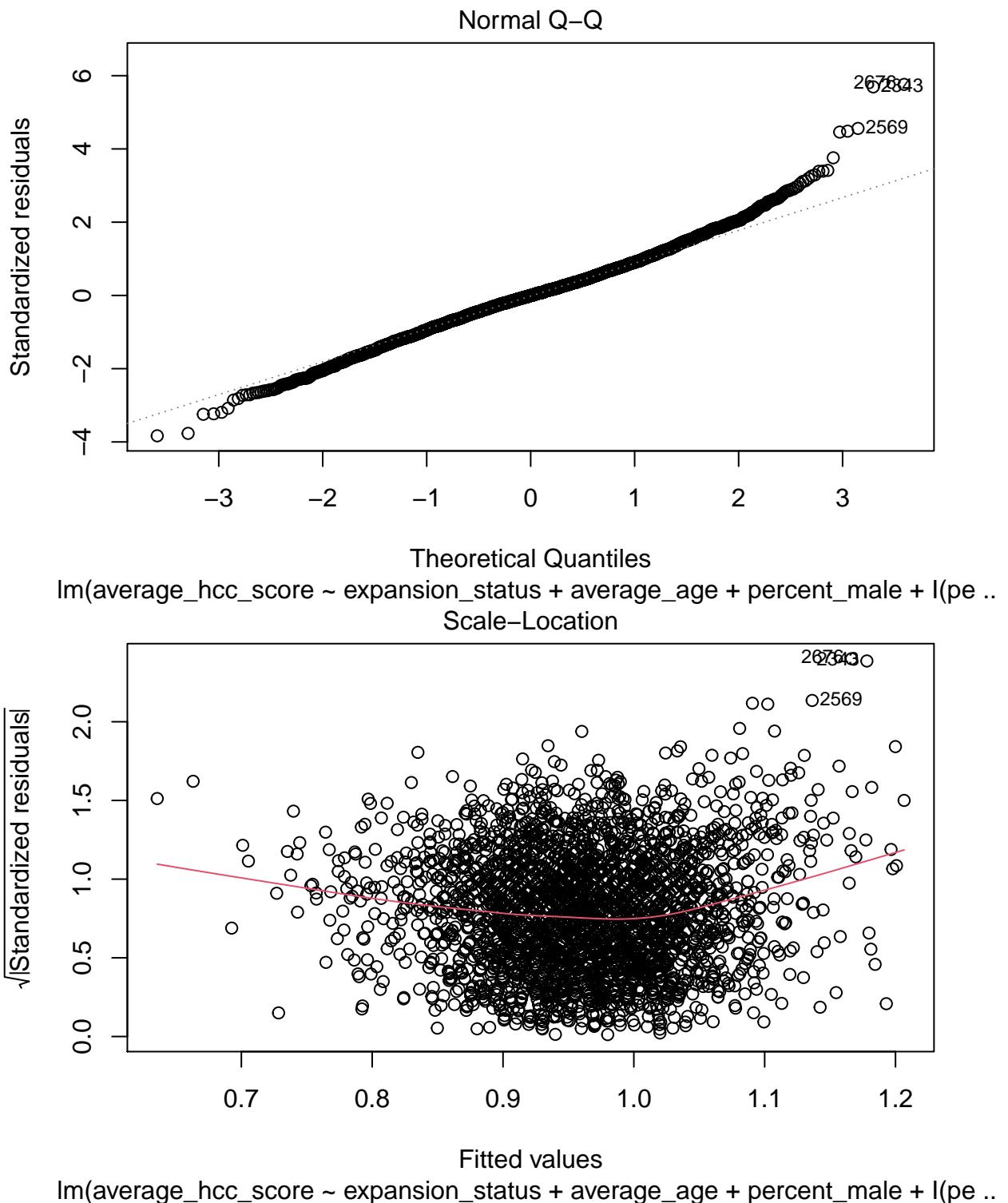
##
## Call:
## lm(formula = average_hcc_score ~ expansion_status + average_age +
##     percent_male + I(percent_male^2) + percent_eligible_for_medicaid +
##     I(percent_eligible_for_medicaid^2) + average_age * percent_male,
##     data = county_dat)
##
## Residuals:
##      Min        1Q        Median       3Q        Max 
## -0.26087 -0.04295 -0.00021  0.04058  0.39360 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 -3.543587  1.283061 -2.762  0.00578 **  
## expansion_status2            0.001065  0.005352  0.199  0.84232    
## expansion_status3            0.014587  0.006040  2.415  0.01579 *   
## expansion_status4            0.034507  0.005182  6.659 3.27e-11 ***  
## average_age                  0.049072  0.014906  3.292  0.00101 **  
## percent_male                 14.310687 3.357018  4.263 2.08e-05 ***  
## I(percent_male^2)           -7.576109  1.849526 -4.096 4.31e-05 ***  
## percent_eligible_for_medicaid 0.823500  0.060687 13.570 < 2e-16 ***  
## I(percent_eligible_for_medicaid^2) -0.452661  0.105487 -4.291 1.83e-05 ***  
## average_age:percent_male    -0.125855  0.031630 -3.979 7.08e-05 ***  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.06936 on 3035 degrees of freedom
## Multiple R-squared:  0.5074, Adjusted R-squared:  0.5058
## F-statistic: 312.6 on 10 and 3035 DF,  p-value: < 2.2e-16

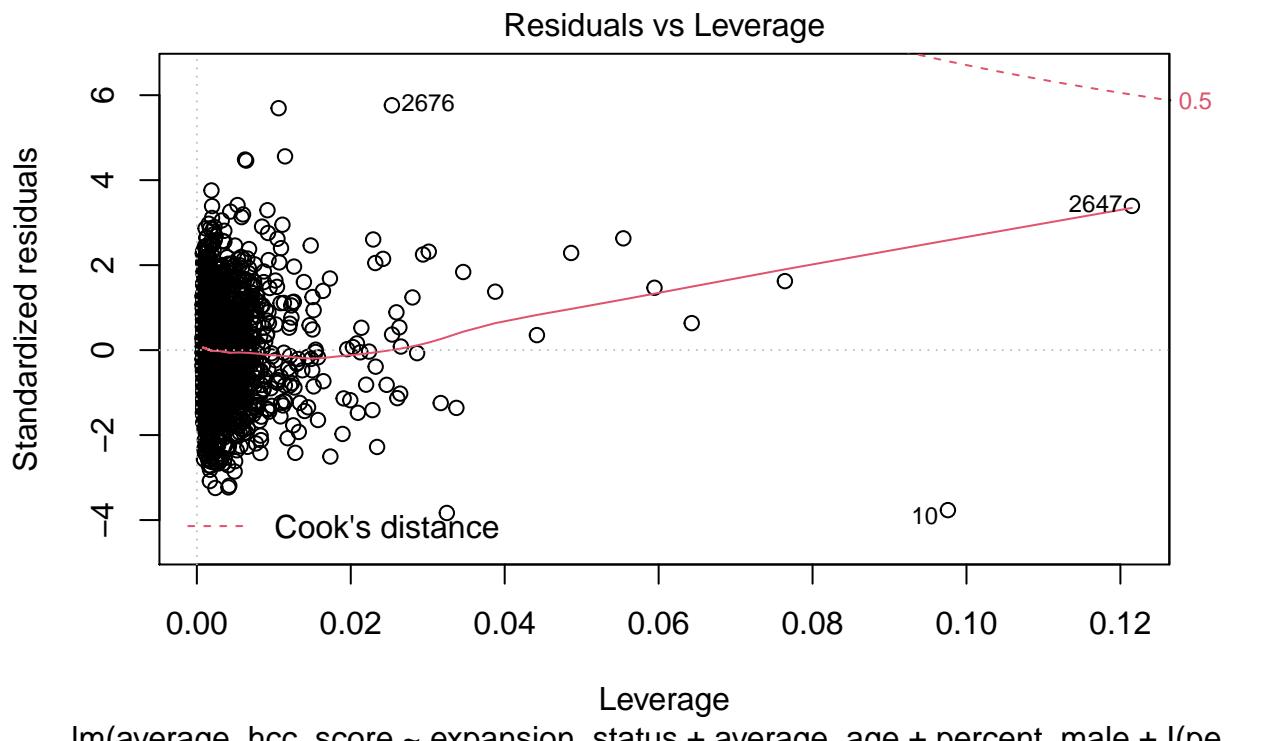
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.06919 on 3036 degrees of freedom  
## Multiple R-squared:  0.5096, Adjusted R-squared:  0.5081  
## F-statistic: 350.5 on 9 and 3036 DF,  p-value: < 2.2e-16
```

```
plot(mlr_em_age_gender)
```







lm(average_hcc_score ~ expansion_status + average_age + percent_male + I(pe ..

```

age <- 65
perc_male <- .5

# 1 unit change in % male
x <- 0.049072*age + 14.3*perc_male-7.57*perc_male**2 - 0.125855*age

y <- 0.049072*age + 14.3*(perc_male + 0.01)-7.57*((perc_male + 0.01)**2) - 0.125855*age

y - x

## [1] 0.066543

# 1 unit change in age
x <- 0.049072*age + 14.3*perc_male-7.57*perc_male**2 - 0.125855*age

y <- 0.049072*(age+1) + 14.3*perc_male-7.57*(perc_male**2) - 0.125855*(age+1)

y - x

## [1] -0.076783

m <- lm(average_hcc_score ~ percent_male + I(percent_male**2) +
           average_age +
           percent_male*average_age,
       data = county_dat)

plot(county_dat$percent_male, county_dat$average_hcc_score, col = 8,
      xlab = "percent male", ylab = "average hcc score")

```

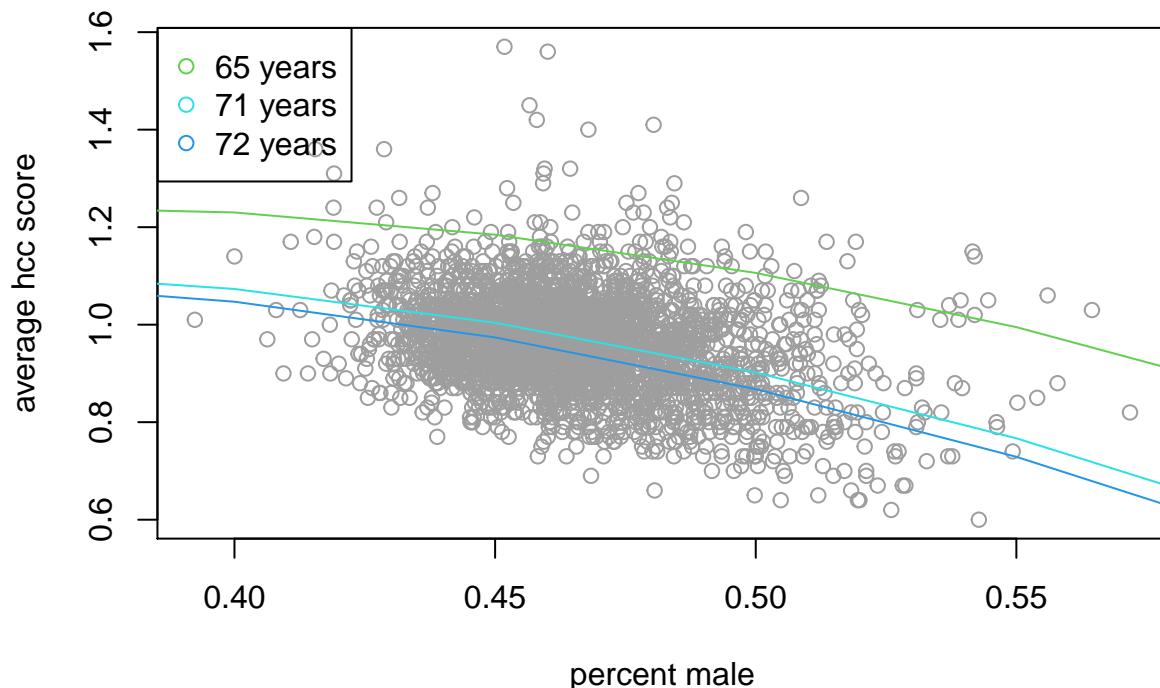
```

lines(seq(.3,.7,.05), predict(m, newdata = data.frame(average_age = 65,
                                                       percent_male = seq(.3,.7,.05))),
      col = 3
    )

lines(seq(.3,.7,.05), predict(m, newdata = data.frame(average_age = 71,
                                                       percent_male = seq(.3,.7,.05))),
      col = 5
    )

lines(seq(.3,.7,.05), predict(m, newdata = data.frame(average_age = 72,
                                                       percent_male = seq(.3,.7,.05))),
      col = 4
    )
legend("topleft", c("65 years", "71 years", "72 years"), pch=1, col=c(3,5,4))

```



```
AIC(mlr_em_age_gender)
```

```
## [1] -7614.531
```

```
anova( mlr2, mlr_em_age_gender)
```

```

## Analysis of Variance Table
##
## Model 1: average_hcc_score ~ expansion_status + average_age + percent_male +
##           percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2)
## Model 2: average_hcc_score ~ expansion_status + average_age + percent_male +
##           I(percent_male^2) + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2) +
##           average_age * percent_male

```

```

##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1    3038 14.648
## 2    3036 14.536  2    0.11155 11.649 9.127e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Robust Sandwich Errors for Coefficient Estimates Because the residual plot depicts a violation of homoscedasticity and the qq plot illustrates deviation from normality, robust sandwich estimation methods were used to calculate the standard errors for the coefficients in the final adjusted model. This method is more robust than the standard LSE method for estimating the standard errors, and can be used to derive more accurate confidence intervals.

```

library(lmtest)

## Loading required package: zoo

## 
## Attaching package: 'zoo'

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

library(sandwich)
coeftest(mlr_em_age_gender, vcov = vcovHC(mlr_em_age_gender, type="HC1"))

```

```

## 
## t test of coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           -3.5435873  1.6569047 -2.1387 0.032541 *  
## expansion_status2      0.0010648  0.0052007  0.2047 0.837782    
## expansion_status3      0.0145870  0.0057930  2.5181 0.011852 *  
## expansion_status4      0.0345069  0.0051904  6.6481 3.506e-11 *** 
## average_age            0.0490723  0.0189325  2.5920 0.009589 ** 
## percent_male           14.3106865 4.3593577  3.2828 0.001040 ** 
## I(percent_male^2)       -7.5761093 2.3504119 -3.2233 0.001281 ** 
## percent_eligible_for_medicaid 0.8234996 0.0969384  8.4951 < 2.2e-16 *** 
## I(percent_eligible_for_medicaid^2) -0.4526607 0.2091976 -2.1638 0.030558 *  
## average_age:percent_male -0.1258554  0.0402533 -3.1266 0.001785 ** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

# Confidence Intervals
(0.0010648 + c(-1,1)*1.96*0.0052007) #95% ci for expansion status 2

```

```

## [1] -0.009128572  0.011258172

```

```

(0.0145870 + c(-1,1)*1.96*0.0057930 ) #95% ci for expansion status 3

```

```

## [1] 0.00323272 0.02594128

```

```
(0.0345069 + c(-1,1)*1.96* 0.0051904) #95% ci for expansion status 4
```

```
## [1] 0.02433372 0.04468008
```

D. Model Summary

$$\begin{aligned} E(\text{average HCC score}) = & -3.54 \\ & 0.001 * \text{I(State Expanded Medicaid in 2014)} + \\ & 0.015 * \text{I(State Expanded Medicaid after 2014)} + \\ & 0.035 * \text{I(State has not Expanded Medicaid)} + \\ & 0.049 * (\text{average age}) + \\ & 14.31 * (\text{percent male}) \\ & - 7.58 * (\text{percent male})^2 + \\ & 0.82 * (\text{percent eligible for medicaid}) \\ & - 0.45 * (\text{percent eligible for medicaid})^2 \\ & - 0.13 * (\text{average age} * \text{percent male}) \end{aligned}$$

Section I: Model 2 - HCC 2018 vs 2010 ~ expansion status (Linear Regression)

Is there a difference in average HCC scores in 2018 vs 2010 for counties that expanded Medicaid early versus late?

Exploratory Analysis

Both the hcc ratio of 2018 vs 2010 values and the difference are normally distributed with long tails on both ends. Taking the log transform of the ratio does not substantially remedy the long tail issue.

```
summary(county_dat_combo$hcc_ratio)
```

```
##      Min. 1st Qu. Median    Mean 3rd Qu.    Max. 
## 0.7767  0.9881 1.0204 1.0229 1.0581 1.4848
```

```
ggplot() + geom_histogram(aes(county_dat_combo$hcc_ratio), col=1) +
  xlab('HCC ratio (2018 HCC/2010 HCC)') +
  ggtitle('Normality of Outcome')
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
summary(county_dat_combo$hcc_ratio)
```

```
##      Min. 1st Qu. Median    Mean 3rd Qu.    Max. 
## 0.7767  0.9881 1.0204 1.0229 1.0581 1.4848
```

Normality of Outcome

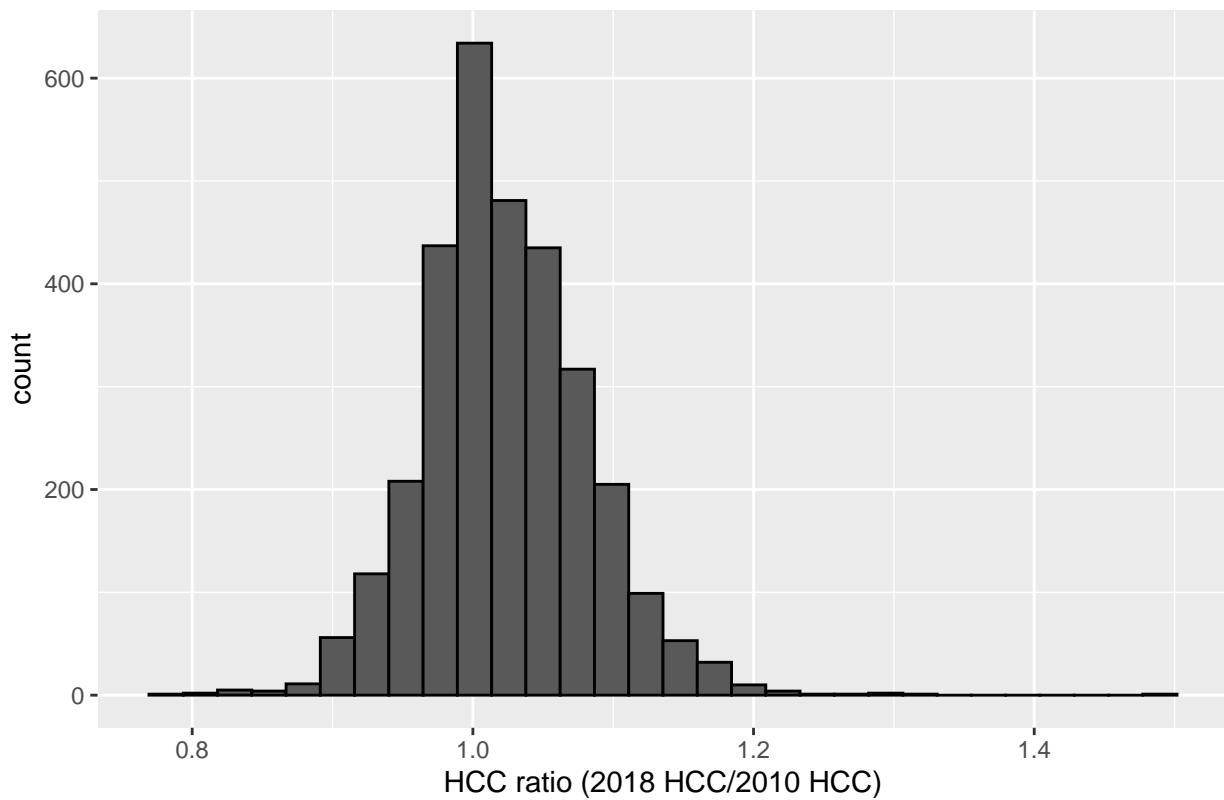


Figure 3: Figure 1a & 1b

```
ggplot() + geom_histogram(aes(log10(county_dat_combo$hcc_ratio)), col=1) +
  xlab('HCC ratio (log)') +
  ggtitle('Normality of Outcome (log)')
```

‘stat_bin()’ using ‘bins = 30’. Pick better value with ‘binwidth’.

Normality of Outcome (log)

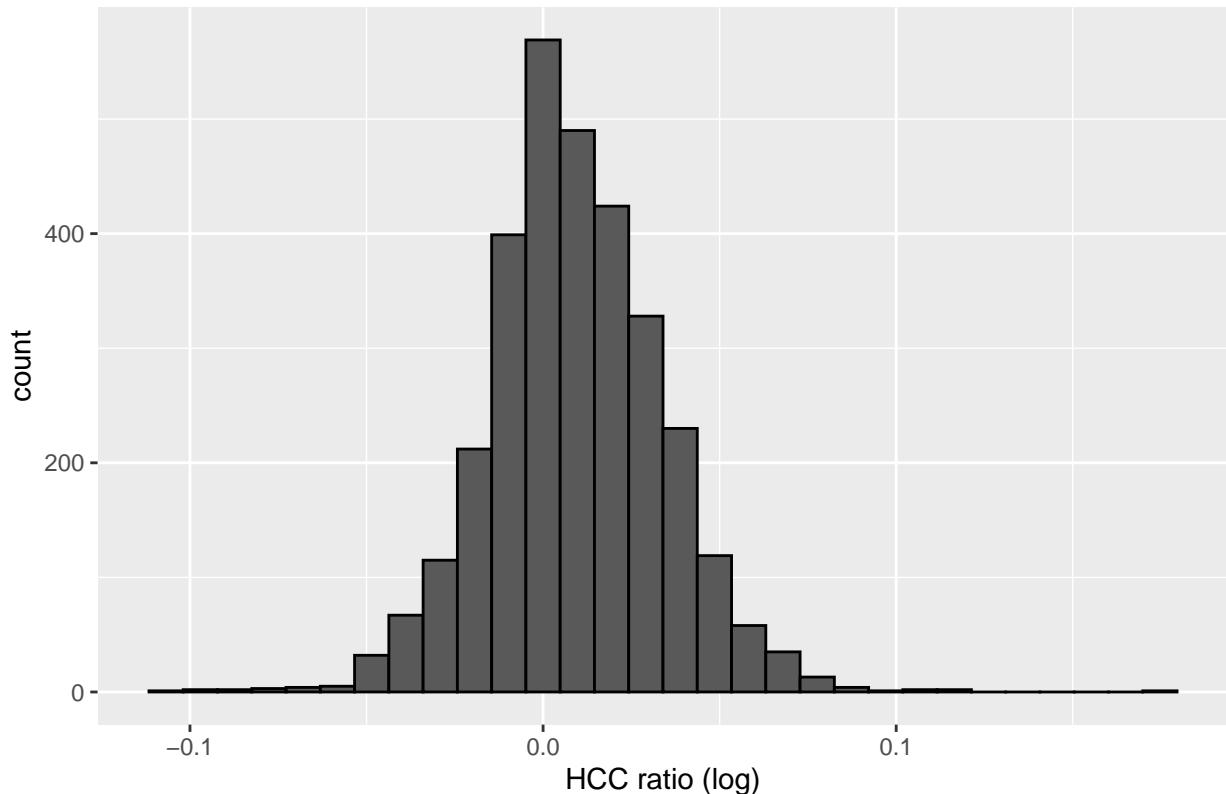


Figure 4: Figure 1a & 1b

```
summary(county_dat_combo$hcc_diff)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## -0.2300 -0.0100  0.0200  0.0208  0.0500  0.3200
```

```
ggplot() + geom_histogram(aes(county_dat_combo$hcc_diff), col=1) +
  xlab('HCC difference (2018 HCC-2010 HCC)') +
  ggtitle('Normality of Outcome')
```

‘stat_bin()’ using ‘bins = 30’. Pick better value with ‘binwidth’.

```
comp_red <- lm(hcc_diff ~ as.factor(expansion_status),
                 data = county_dat_combo)
summary(comp_red)
```

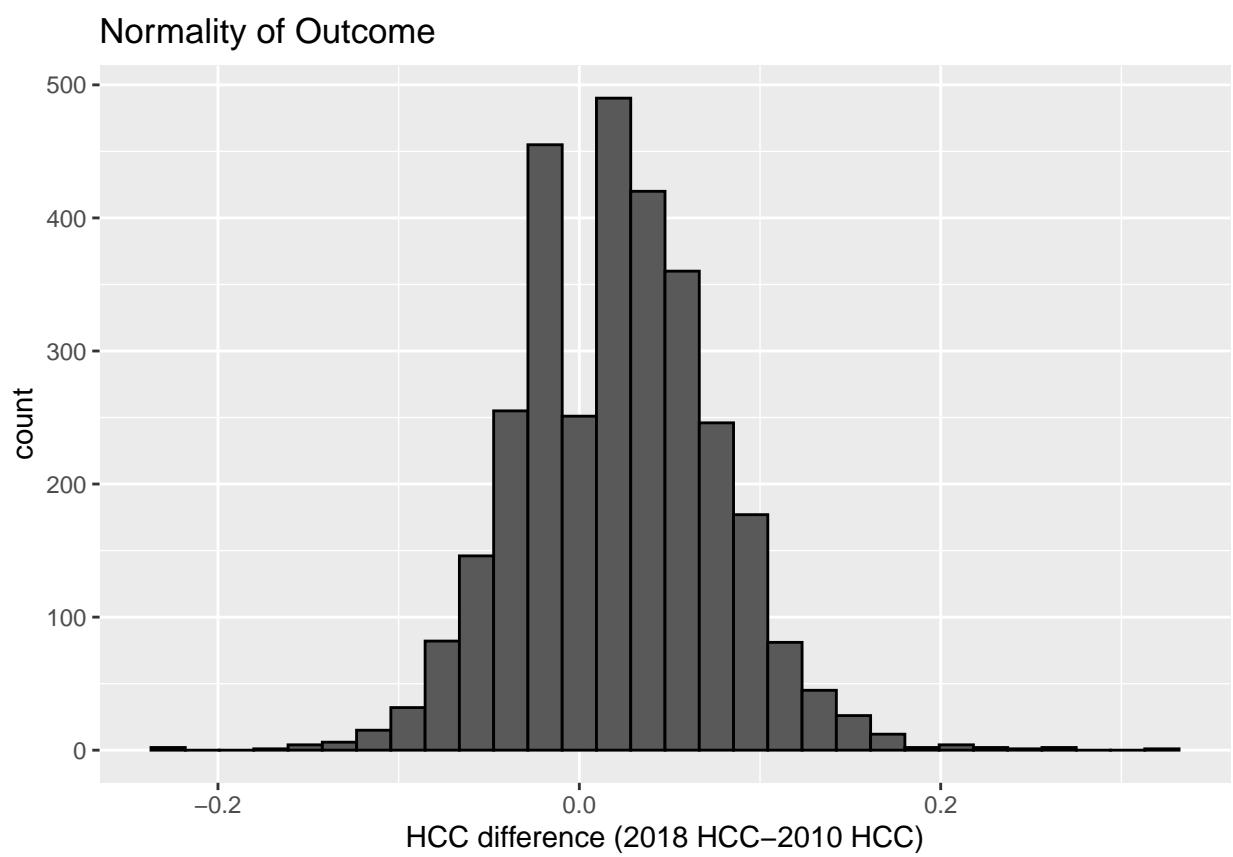


Figure 5: Figure 2

```

## 
## Call:
## lm(formula = hcc_diff ~ as.factor(expansion_status), data = county_dat_combo)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.26568 -0.03225 -0.00225  0.03432  0.30467 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             0.035681  0.003663  9.741 < 2e-16 ***
## as.factor(expansion_status)2 -0.018703  0.004082 -4.582 4.80e-06 ***
## as.factor(expansion_status)3 -0.020354  0.004539 -4.485 7.56e-06 *** 
## as.factor(expansion_status)4 -0.013431  0.003895 -3.448 0.000572 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.05346 on 3114 degrees of freedom
## Multiple R-squared:  0.008397, Adjusted R-squared:  0.007442 
## F-statistic:  8.79 on 3 and 3114 DF, p-value: 8.396e-06

```

```

comp1 <- lm(hcc_diff ~ as.factor(expansion_status) +
            delta_medicaid +
            delta_age +
            delta_male ,
            data = county_dat_combo)
summary(comp1)

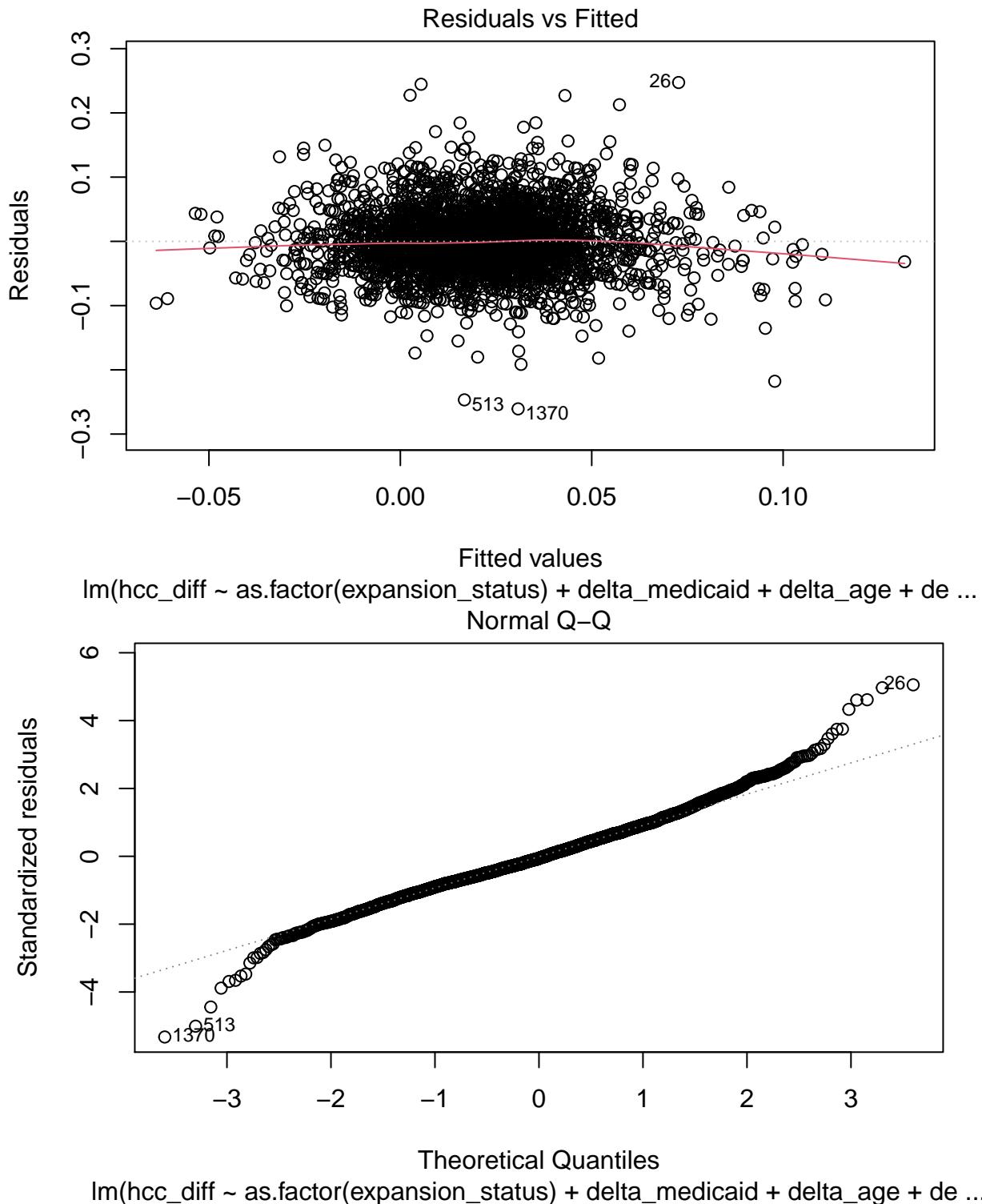
```

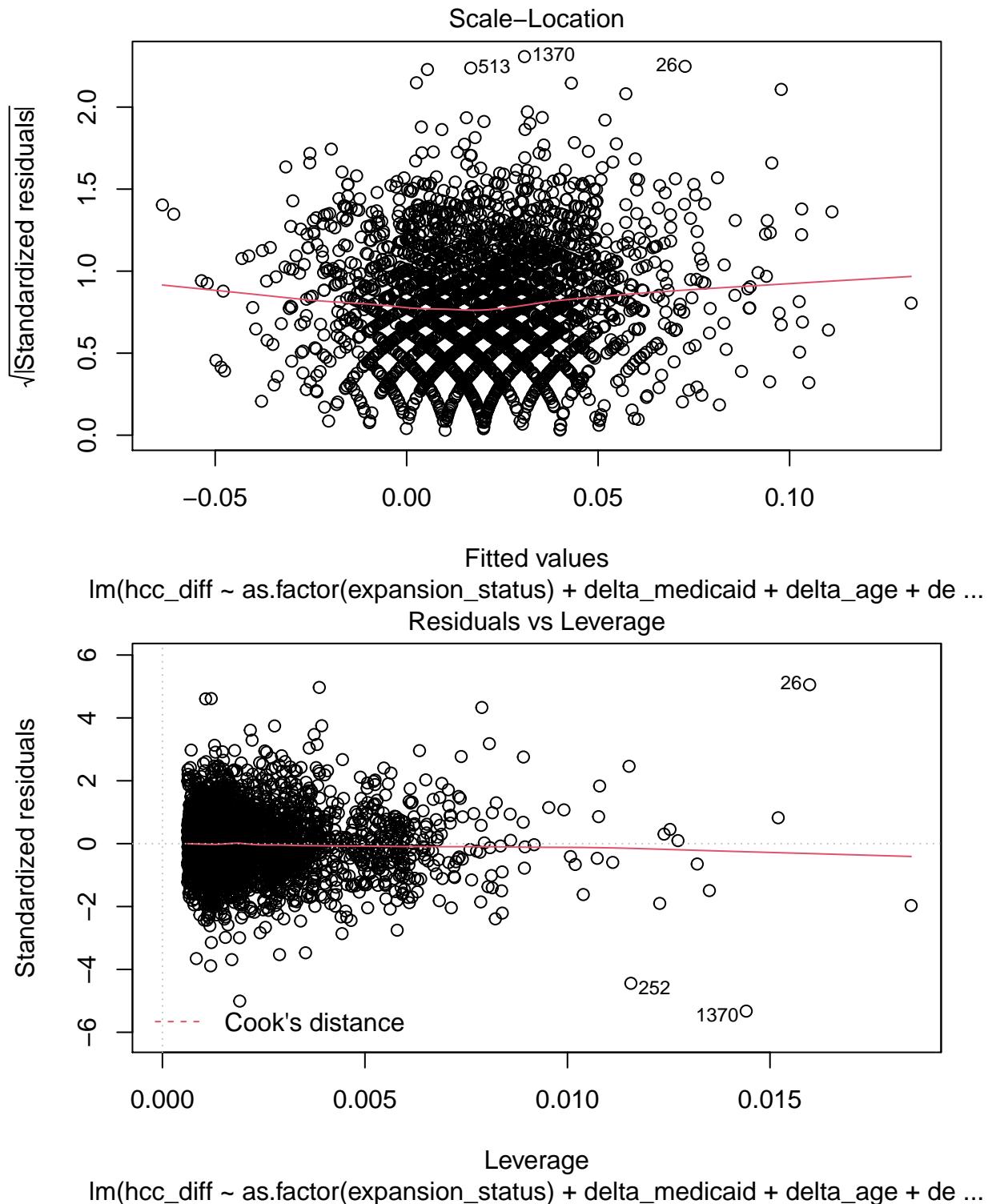
```

## 
## Call:
## lm(formula = hcc_diff ~ as.factor(expansion_status) + delta_medicaid +
##     delta_age + delta_male, data = county_dat_combo)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.260765 -0.031124 -0.002876  0.030114  0.247300 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             0.022123  0.003577  6.184 7.05e-10 *** 
## as.factor(expansion_status)2 -0.006265  0.003879 -1.615 0.106386 
## as.factor(expansion_status)3 -0.006567  0.004263 -1.540 0.123555 
## as.factor(expansion_status)4  0.013582  0.003938  3.449 0.000571 *** 
## delta_medicaid            0.132407  0.005777 22.921 < 2e-16 *** 
## delta_age                 1.044439  0.078688 13.273 < 2e-16 *** 
## delta_male                0.079834  0.030676  2.602 0.009299 ** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.04932 on 3111 degrees of freedom
## Multiple R-squared:  0.157, Adjusted R-squared:  0.1553 
## F-statistic: 96.54 on 6 and 3111 DF, p-value: < 2.2e-16

```

```
plot(comp1)
```





The additional quadratic delta_medicaid term improves the linearity of the final model, and is statistically significant.

```
comp2 <- lm(hcc_diff ~ as.factor(expansion_status) +
               delta_medicaid +
               delta_age +
```

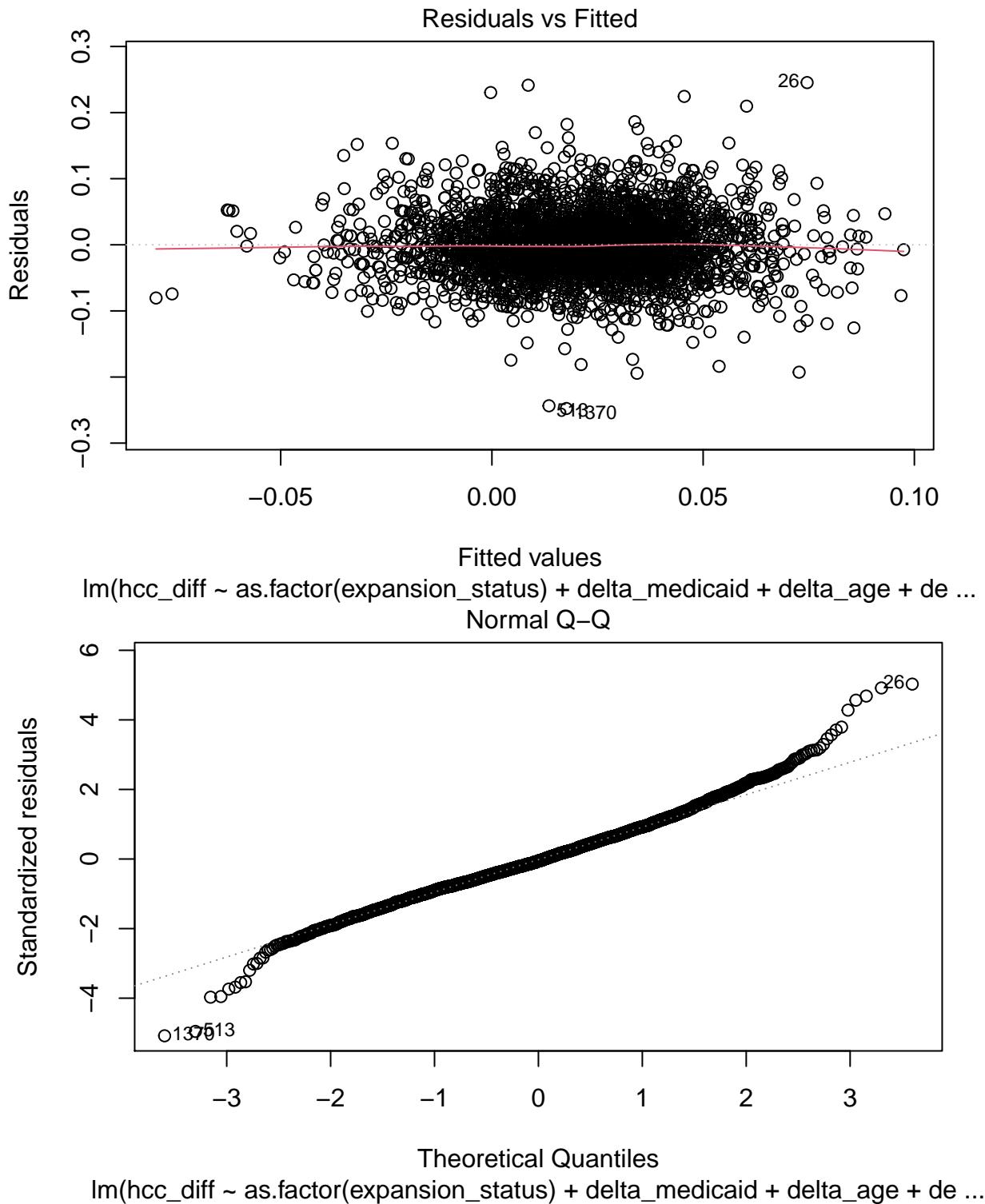
```

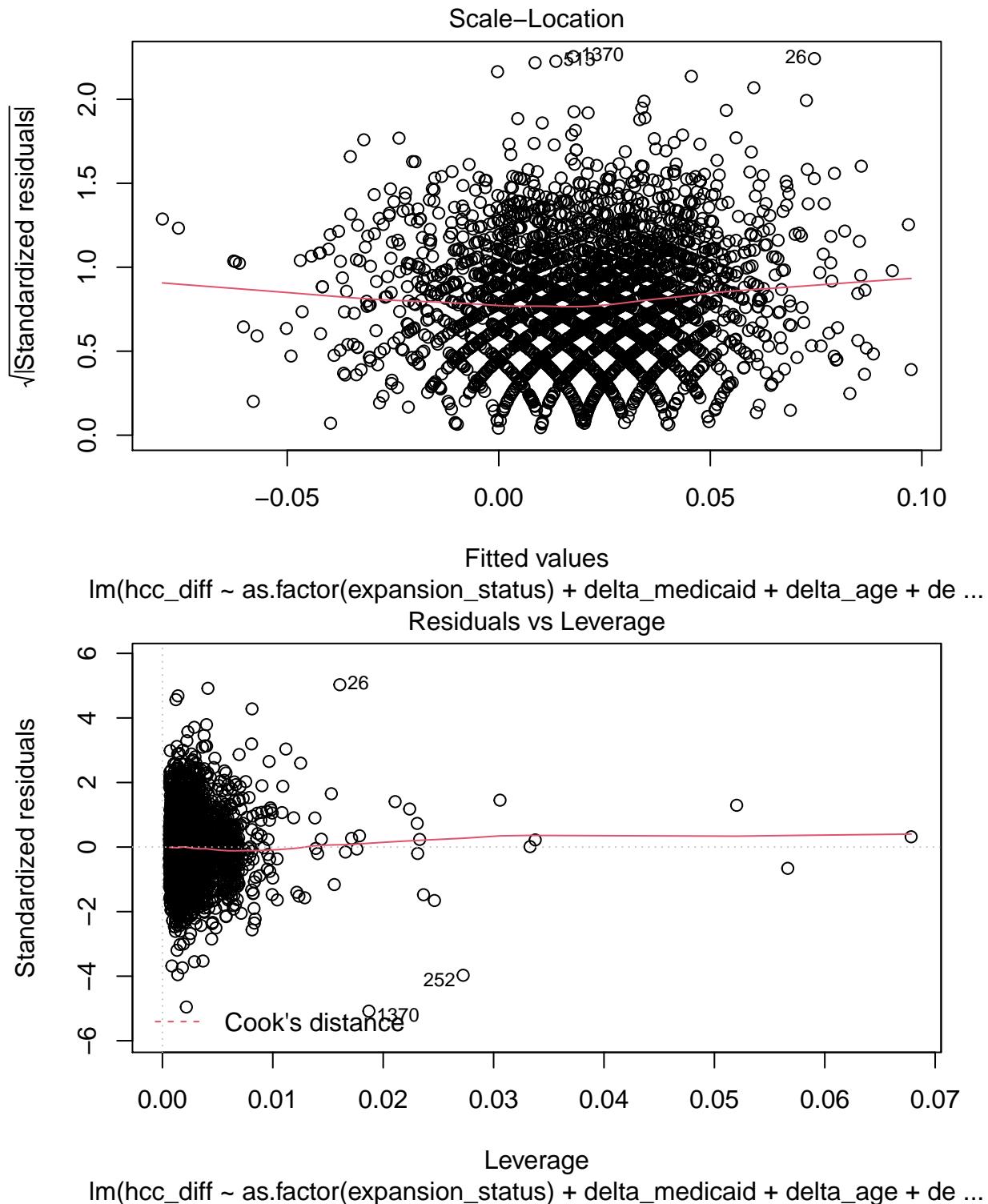
        delta_male +
        I(delta_medicaid**2) ,
  data = county_dat_combo)
summary(comp2)

##
## Call:
## lm(formula = hcc_diff ~ as.factor(expansion_status) + delta_medicaid +
##      delta_age + delta_male + I(delta_medicaid^2), data = county_dat_combo)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -0.247636 -0.031706 -0.002708  0.030112  0.245387
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             0.027361  0.003793   7.214 6.80e-13 ***
## as.factor(expansion_status)2 -0.009816  0.003966  -2.475  0.01338 *  
## as.factor(expansion_status)3 -0.010394  0.004355  -2.387  0.01706 *  
## as.factor(expansion_status)4  0.010692  0.003992   2.678  0.00744 ** 
## delta_medicaid            0.136209  0.005837  23.334 < 2e-16 ***
## delta_age                  1.036104  0.078518  13.196 < 2e-16 *** 
## delta_male                 0.086603  0.030644   2.826  0.00474 ** 
## I(delta_medicaid^2)       -0.054369  0.013345  -4.074  4.73e-05 ***
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04919 on 3110 degrees of freedom
## Multiple R-squared:  0.1614, Adjusted R-squared:  0.1596 
## F-statistic: 85.54 on 7 and 3110 DF,  p-value: < 2.2e-16

plot(comp2)

```





The full model with the additional delta age, percent male, and percent eligible for medicaid is needed to explain the variation in the data.

```
anova(comp_red, comp2)
```

```
## Analysis of Variance Table
```

```

## 
## Model 1: hcc_diff ~ as.factor(expansion_status)
## Model 2: hcc_diff ~ as.factor(expansion_status) + delta_medicaid + delta_age +
##           delta_male + I(delta_medicaid^2)
##   Res.Df   RSS Df Sum of Sq    F    Pr(>F)
## 1     3114 8.8997
## 2     3110 7.5261  4     1.3736 141.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
AIC(comp2)
```

```
## [1] -9924.365
```

Robust Sandwich Errors for Coefficient Estimates Because the residual plot depicts a violation of homoscedasticity and the qq plot illustrates deviation from normality, robust sandwich estimation methods were used to calculate the standard errors for the coefficients in the final adjusted model. This method is more robust than the standard LSE method for estimating the standard errors, and can be used to derive more accurate confidence intervals.

```

library(lmtest)
library(sandwich)
coeftest(comp2, vcov = vcovHC(comp2, type="HC1"))

```

```

## 
## t test of coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             0.0273608  0.0038189  7.1646 9.698e-13 ***
## as.factor(expansion_status)2 -0.0098163  0.0039447 -2.4884 0.0128823 *  
## as.factor(expansion_status)3 -0.0103938  0.0043592 -2.3843 0.0171705 *  
## as.factor(expansion_status)4  0.0106918  0.0039253  2.7238 0.0064891 ** 
## delta_medicaid          0.1362087  0.0064151 21.2326 < 2.2e-16 ***
## delta_age                1.0361040  0.0989823 10.4676 < 2.2e-16 *** 
## delta_male               0.0866028  0.0402591  2.1511 0.0315424 *  
## I(delta_medicaid^2)      -0.0543689  0.0149204 -3.6439 0.0002729 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

# Confidence Intervals
(-0.0098163 + c(-1,1)*1.96*0.0039447) #95% ci for 2 vs 1

```

```
## [1] -0.017547912 -0.002084688
```

```
( -0.0103938 + c(-1,1)*1.96* 0.0043592 ) #95% ci for 3 vs 1
```

```
## [1] -0.018937832 -0.001849768
```

```
(0.0106918 + c(-1,1)*1.96*0.0039253) #95% ci for 4 vs 1
```

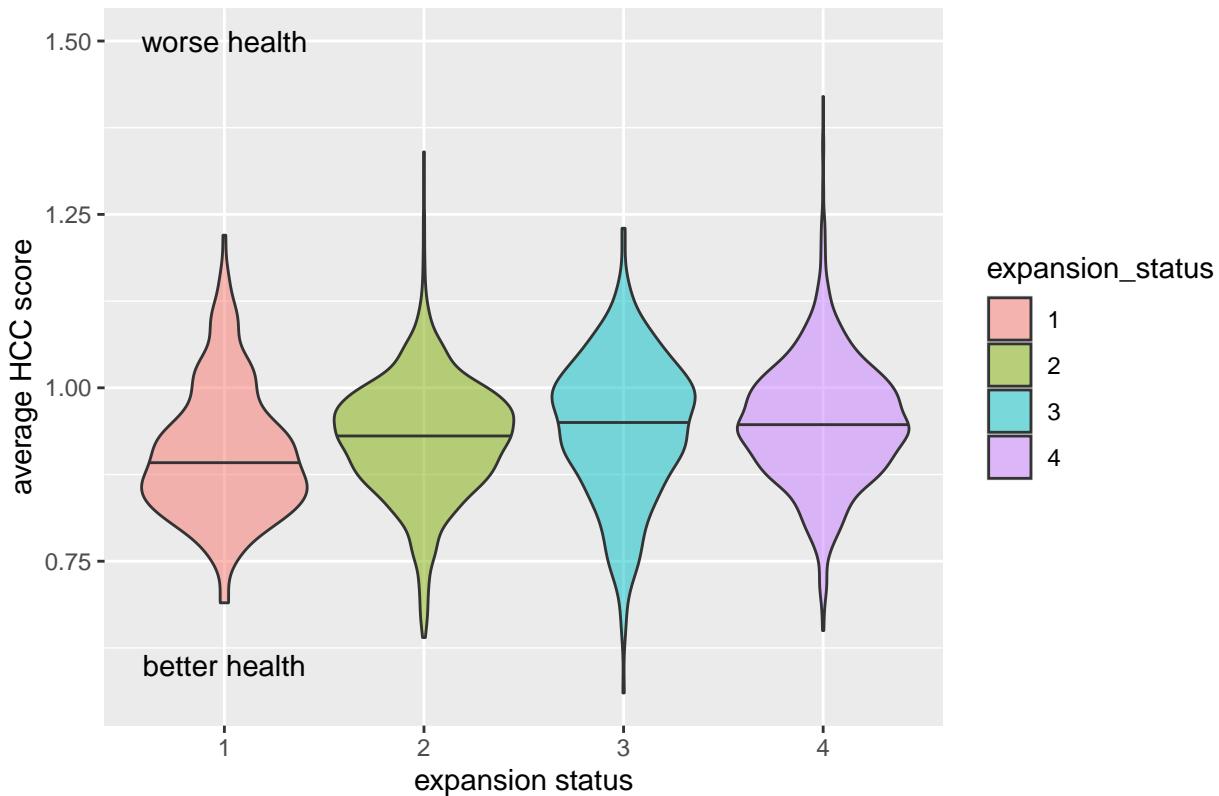
```
## [1] 0.002998212 0.018385388
```

Final Model Summary

$E(\text{average HCC score 2018} - \text{average HCC score 2010}) = 0.027$
– 0.0098 * I(State Expanded Medicaid in 2014)+
– 0.01 * I(State Expanded Medicaid after 2014)+
0.011 * I(State has not Expanded Medicaid)+
1.036 * (delta age)+
0.0866 * (delta male)+
0.136 * (delta percent eligible for medicaid)
– 0.0543 * (delta percent eligible for medicaid)²

```
county_dat_combo %>% ggplot() +
  geom_violin(aes(expansion_status, average_hcc_score_2010,
    fill = expansion_status), alpha = 0.5,
    draw_quantiles = c(0.5)) +
  # scale_y_reverse() +
  ylab('average HCC score') +
  xlab('expansion status') +
  annotate("text", label = "better health", x = 1, y = 0.6) +
  annotate("text", label = "worse health", x = 1, y = 1.5) +
  ggtitle('2010 Average HCC Scores per US County')
```

2010 Average HCC Scores per US County



Section III: Logistic Regression: Odds of High Emergency Dept Use/Hospital Readmission ~ Expansion Status

Create binary outcome categorizing hospital readmission rates & emergency department visits as high if the county rate exceeds the national average rate (hospital readmission - 18.06%, ED visits - 670 per 1000 beneficiaries)

Also, create multinomial class var with following labels:

- * 4 - hospital readmission rate & % beneficiaries with an ED visit above national average
- * 3 - beneficiaries w/ ED visit above national average (hospital readmission below)
- * 2 - hospital readmission rate above national average (ED below)
- * 1 - both hospital readmission & ED below

```
county_dat['high_hospital_readmission'] = ifelse(county_dat$hospital_readmission_rate >= .1806, 1, 0)

county_dat['high_ED'] = ifelse(county_dat$emergency_department_visits_per_1000_beneficiaries >= 670, 1, 0)

county_dat['hospital_ED_class'] = case_when(
  (county_dat$high_hospital_readmission == 1) &
    (county_dat$high_ED == 1) ~ 3,
  county_dat$high_ED == 1 ~ 2,
  county_dat$high_hospital_readmission == 1 ~ 2,
  TRUE ~ 1)
```

Emergency Department Visits

Expansion status of state is only predictive in the logistic model looking at high ED rates in counties; it is not a statistically significant predictor in the model with high hospitalization rates as the outcome (neither in binary - expanded vs non-expanded nor more verbose 4/3/2/1 classification)

```
mod.binary.ed = glm(high_ED ~
  as.factor(expansion_status) +
  percent_male +
  average_age +
  percent_eligible_for_medicaid +
  I(percent_eligible_for_medicaid**2),
  family = binomial(), data = county_dat)
summary(mod.binary.ed)

##
## Call:
## glm(formula = high_ED ~ as.factor(expansion_status) + percent_male +
##       average_age + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##       family = binomial(), data = county_dat)
##
## Deviance Residuals:
##    Min      1Q   Median      3Q      Max
## -2.8545 -0.7890  0.3651  0.7858  2.6277
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                29.23201   3.53935   8.259 < 2e-16 ***
## as.factor(expansion_status)2      0.48100   0.18347   2.622 0.008749 **
## as.factor(expansion_status)3      0.41179   0.20568   2.002 0.045273 *
## as.factor(expansion_status)4      0.62052   0.17606   3.524 0.000424 ***
## percent_male                  -24.36284   2.36449 -10.304 < 2e-16 ***
## average_age                   -0.32340   0.04105  -7.878 3.33e-15 ***
## percent_eligible_for_medicaid    34.78637   2.54669  13.659 < 2e-16 ***
## I(percent_eligible_for_medicaid^2) -44.50784   4.58289 -9.712 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 4188.5 on 3045 degrees of freedom
## Residual deviance: 3034.8 on 3038 degrees of freedom
## AIC: 3050.8
##
## Number of Fisher Scoring iterations: 4

mod.binary.ed.cont = glm(high_ED ~
  as.integer(expansion_status) +
  percent_male +
  average_age +
  percent_eligible_for_medicaid +
  I(percent_eligible_for_medicaid**2),
  family = binomial(), data = county_dat)
summary(mod.binary.ed.cont)
```

```

## 
## Call:
## glm(formula = high_ED ~ as.integer(expansion_status) + percent_male +
##       average_age + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##       family = binomial(), data = county_dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.8605 -0.7909  0.3659  0.7819  2.6677
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                29.40714   3.53049   8.329 < 2e-16 ***
## as.integer(expansion_status) 0.12982   0.04346   2.987  0.00281 **
## percent_male               -24.56220   2.35795 -10.417 < 2e-16 ***
## average_age                 -0.32284   0.04100  -7.874 3.44e-15 ***
## percent_eligible_for_medicaid 34.66931   2.53525 13.675 < 2e-16 ***
## I(percent_eligible_for_medicaid^2) -44.31853   4.55111  -9.738 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 4188.5 on 3045 degrees of freedom
## Residual deviance: 3039.2 on 3040 degrees of freedom
## AIC: 3051.2
##
## Number of Fisher Scoring iterations: 4

```

Reduced model treating expansion_status as a continuous covariate is sufficient; the categorical information is not needed.

```
anova(mod.binary.ed.cont, mod.binary.ed, test="Chisq")
```

```

## Analysis of Deviance Table
##
## Model 1: high_ED ~ as.integer(expansion_status) + percent_male + average_age +
##           percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2)
## Model 2: high_ED ~ as.factor(expansion_status) + percent_male + average_age +
##           percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2)
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      3040    3039.2
## 2      3038    3034.8  2    4.4276  0.1093

```

```
exp(0.48100 + c(-1,1)*0.18347*1.96) # expansion status 2 vs 1
```

```
## [1] 1.129075 2.317760
```

```
exp(0.41179 + c(-1,1)*0.20568*1.96) # expansion status 3 vs 1
```

```
## [1] 1.008695 2.259001
```

```

exp(0.62052 + c(-1,1)*0.17606*1.96) # expansion status 4 vs 1

## [1] 1.317113 2.626357

# continuous
exp(0.12982 + c(-1,1)*0.04346*1.96)

## [1] 1.045650 1.239864

# OR - category 4 vs 1
odds4 = exp(29.40714 + 4*0.12982)
odds1 = exp(29.40714 + 0.12982)
OR4v1 = odds4/odds1 # OR

```

```
library(ResourceSelection)
```

GOF - Hosmer Lemeshow Goodness of Fit

```

## ResourceSelection 0.3-5 2019-07-22

hoslem.test(county_dat$high_ED, predict(mod.binary.ed), g = 10)

##
```

```

## Hosmer and Lemeshow goodness of fit (GOF) test
##
```

```

## data: county_dat$high_ED, predict(mod.binary.ed)
## X-squared = -1948.5, df = 8, p-value = 1
```

```
hoslem.test(county_dat$high_ED, predict(mod.binary.ed.cont), g= 10)
```

```
##
```

```
## Hosmer and Lemeshow goodness of fit (GOF) test
```

```
##
```

```
## data: county_dat$high_ED, predict(mod.binary.ed.cont)
```

```
## X-squared = -1981, df = 8, p-value = 1
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
```

```
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## cov, smooth, var
```

```

roc_cont <- roc(county_dat$high_ED, predict(mod.binary.ed.cont))

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

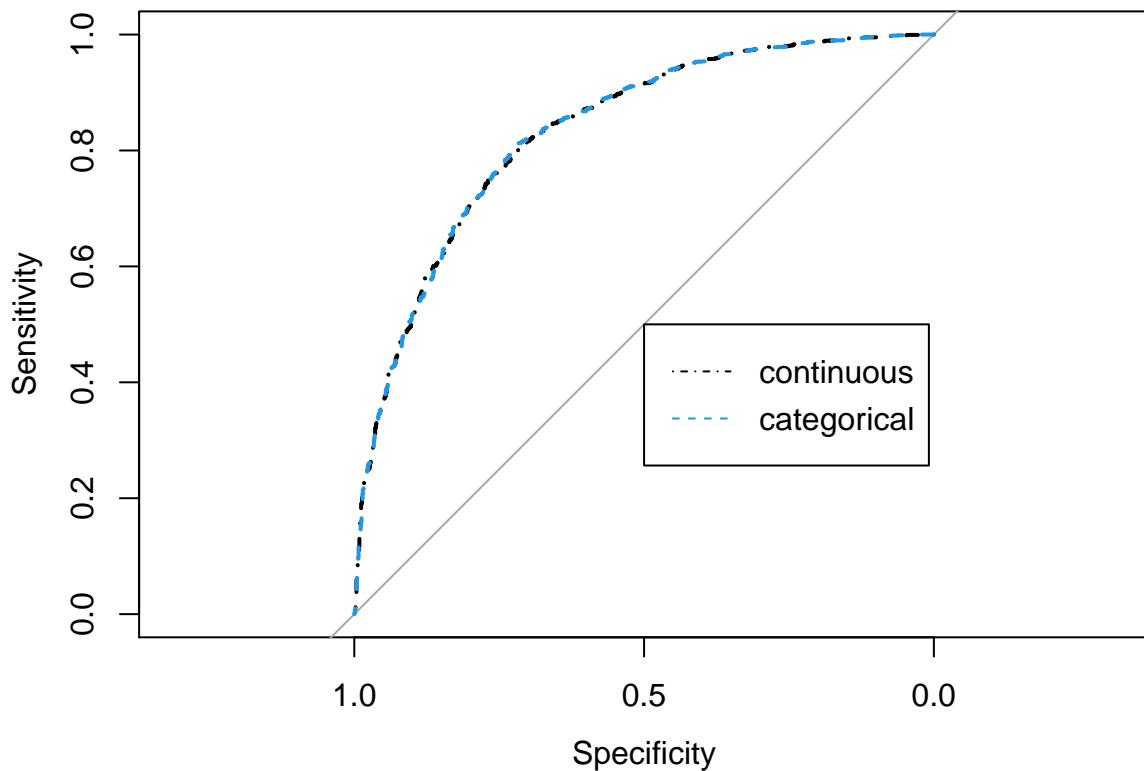
roc_cat <- roc(county_dat$high_ED, predict(mod.binary.ed))

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

plot(roc_cont, col = 1, lty = 4)
plot(roc_cat, add = TRUE, col = 4, lty = 2, alpha = .5)

legend(0.5, 0.5,
       legend = c("continuous", "categorical"),
       col = c(1, 4), lty = c(4, 2))

```



```

auc(roc_cont)

## Area under the curve: 0.8336

auc(roc_cat)

## Area under the curve: 0.8342

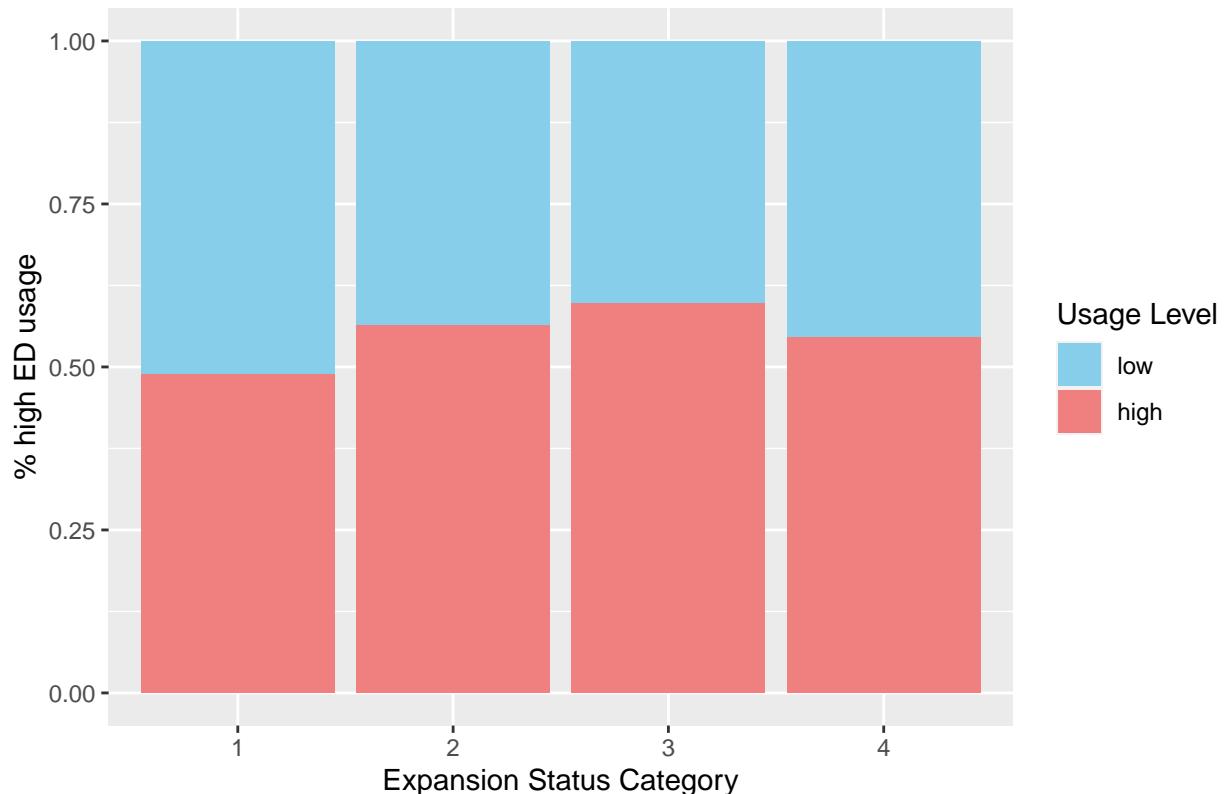
```

Final Model Summary

$$\logit(p(\text{high ED usage})) = 29.4 + 0.13 * (\text{expansion status}) - 24.56 * (\text{percent male}) - 0.322 * (\text{average age}) + 34.67 * (\text{percent eligible for medicaid}) - 44.32 * (\text{percent eligible for medicaid})^2$$

```
temp <- county_dat %>% group_by(expansion_status, high_ED) %>% count(high_ED)
temp %>% ggplot() +
  geom_bar(aes(x=expansion_status,
                y=n, fill = as.factor(high_ED)),
           position = "fill", stat='identity') +
  ylab('% high ED usage') + xlab('Expansion Status Category') +
  ggtitle('Emergency Department Usage Levels') +
  scale_fill_manual(name='Usage Level',
                    values = c('0' = "skyblue",
                               '1' = "lightcoral"),
                    labels = c('low', 'high'))
```

Emergency Department Usage Levels



Hospital Readmissions

```

mod.binary.hosp = glm(high_hospital_readmission ~
  as.factor(expansion_status) +
  percent_male +
  average_age +
  percent_eligible_for_medicaid,
  family = binomial(), data = county_dat)
summary(mod.binary.hosp)

##
## Call:
## glm(formula = high_hospital_readmission ~ as.factor(expansion_status) +
##       percent_male + average_age + percent_eligible_for_medicaid,
##       family = binomial(), data = county_dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.2497 -0.8481 -0.6165  1.1064  2.7787
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                26.57178   3.09240   8.593 < 2e-16 ***
## as.factor(expansion_status)2    0.27834   0.17757   1.567   0.117
## as.factor(expansion_status)3   -0.27100   0.20207  -1.341   0.180
## as.factor(expansion_status)4    0.19658   0.17157   1.146   0.252
## percent_male                 -17.57146   2.18202  -8.053 8.09e-16 ***
## average_age                  -0.28662   0.03611  -7.936 2.08e-15 ***
## percent_eligible_for_medicaid  5.36320   0.65799   8.151 3.61e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 3866.0 on 3045 degrees of freedom
## Residual deviance: 3443.4 on 3039 degrees of freedom
## AIC: 3457.4
##
## Number of Fisher Scoring iterations: 4

```

Multinomial Regression: Risk of High Emergency Dept Use and/or Hospital Readmission ~ Expansion Status

```

library(nnet)
mod.multi <- multinom(as.factor(hospital_ED_class) ~
  as.factor(expansion_status) +
  percent_male +
  average_age +
  percent_eligible_for_medicaid +
  I(percent_eligible_for_medicaid**2) +
  I(average_age**2),
  data = county_dat)

```

```

## # weights: 30 (18 variable)
## initial value 3346.373031
## iter 10 value 2877.441498
## iter 20 value 2687.437464
## iter 30 value 2663.789938
## iter 40 value 2663.730781
## iter 50 value 2663.721979
## iter 60 value 2663.606342
## iter 70 value 2663.091050
## final value 2663.058111
## converged

summary(mod.multi)

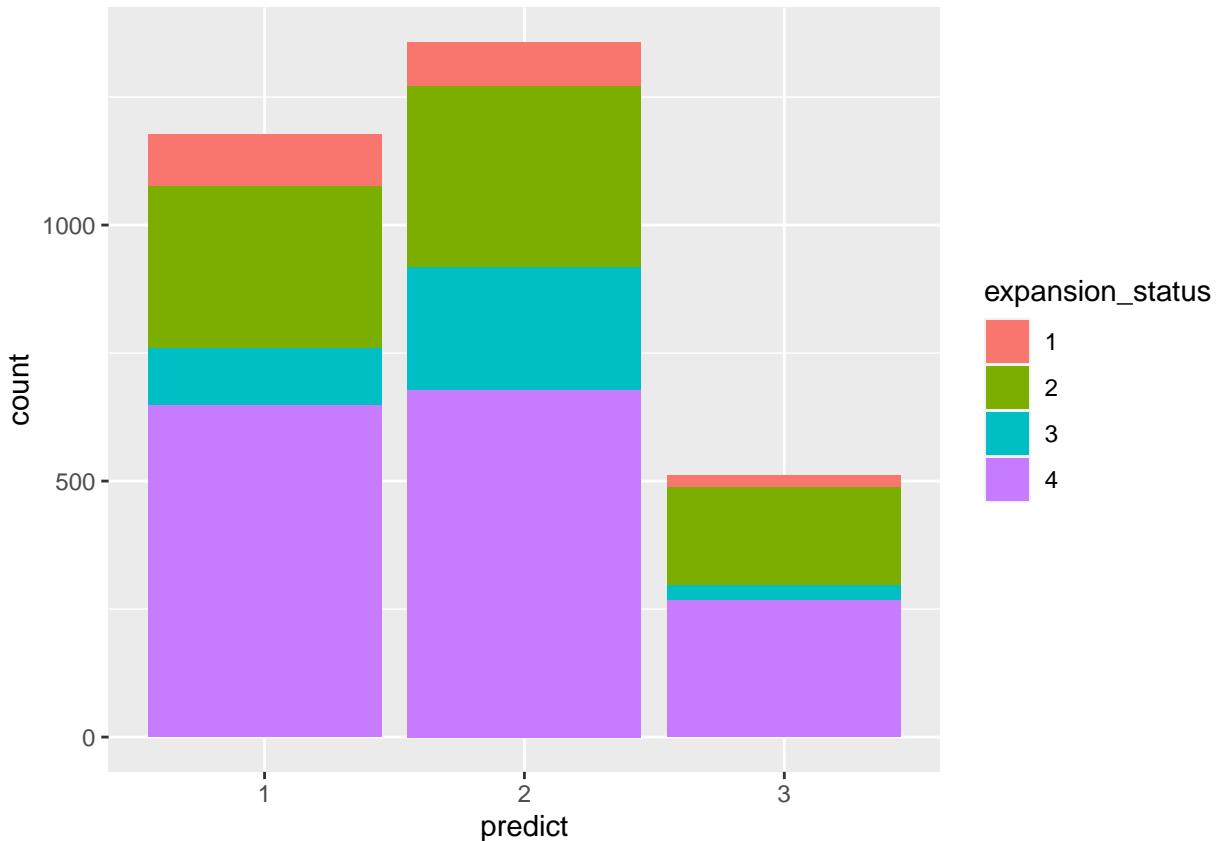
## Call:
## multinom(formula = as.factor(hospital_ED_class) ~ as.factor(expansion_status) +
##           percent_male + average_age + percent_eligible_for_medicaid +
##           I(percent_eligible_for_medicaid^2) + I(average_age^2), data = county_dat)
##
## Coefficients:
## (Intercept) as.factor(expansion_status)2 as.factor(expansion_status)3
## 2 -34.10016          0.4346568          0.37510679
## 3 53.44012          0.6711761          0.08902687
## as.factor(expansion_status)4 percent_male average_age
## 2          0.5546608     -20.56062    1.3464230
## 3          0.7291613     -36.45654   -0.7249963
## percent_eligible_for_medicaid I(percent_eligible_for_medicaid^2)
## 2          27.76756          -33.81324
## 3          38.75018          -45.32945
## I(average_age^2)
## 2      -0.011099119
## 3      0.001671282
##
## Std. Errors:
## (Intercept) as.factor(expansion_status)2 as.factor(expansion_status)3
## 2 0.0009394821          0.02741390          0.004877179
## 3 0.0011593232          0.02062059          0.002907146
## as.factor(expansion_status)4 percent_male average_age
## 2          0.03332274  0.0006059360  0.03389721
## 3          0.02444438  0.0007932748  0.04098273
## percent_eligible_for_medicaid I(percent_eligible_for_medicaid^2)
## 2          0.001034685          0.0003822978
## 3          0.001440342          0.0006571129
## I(average_age^2)
## 2      0.0004725000
## 3      0.0005751155
##
## Residual Deviance: 5326.116
## AIC: 5362.116

```

The model is not doing a good job classifying health outcome class based on expansion status, largely because the raw data does not have clear distinctions across the outcome categories.

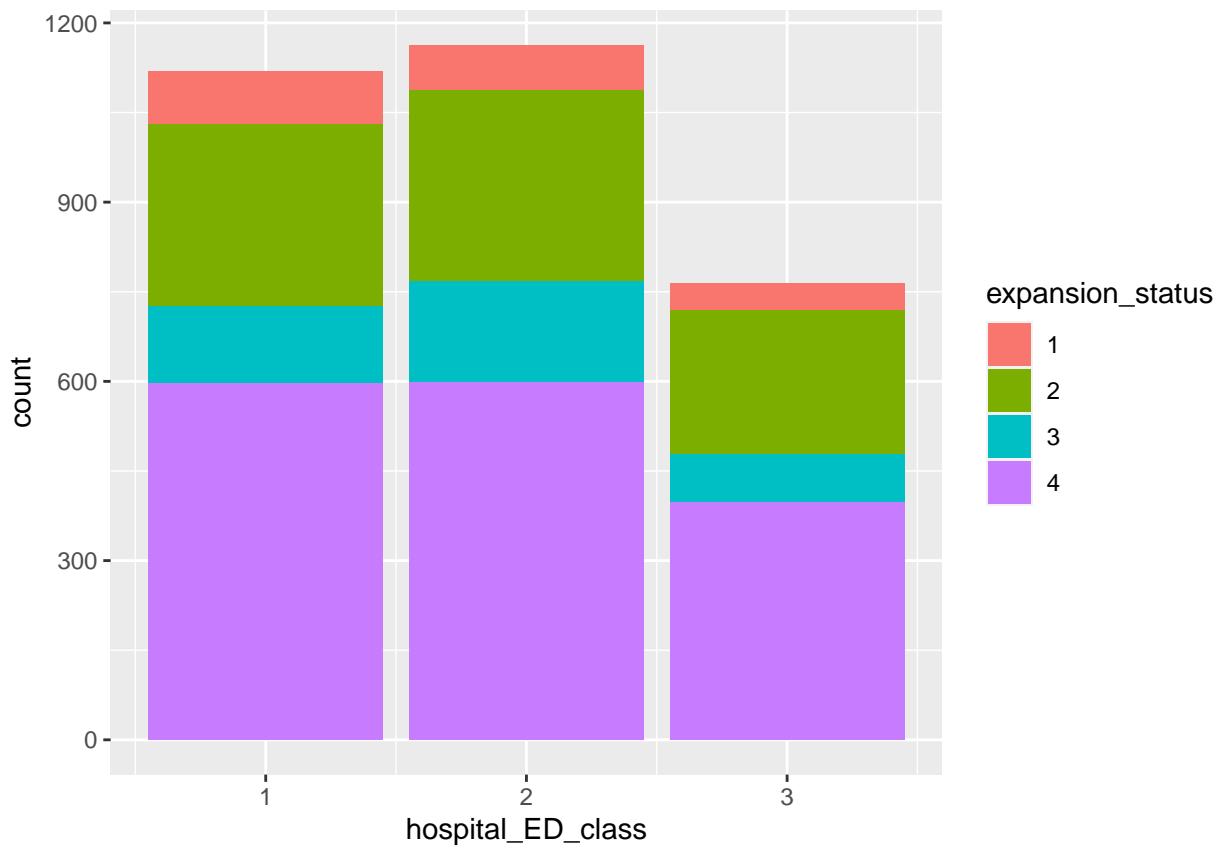
```
county_dat['predict'] <- predict(mod.multi)
county_dat %>% ggplot() + geom_histogram(aes(predict, fill = expansion_status), stat="count")
```

Warning: Ignoring unknown parameters: binwidth, bins, pad



```
county_dat %>% ggplot() + geom_histogram(aes(hospital_ED_class, fill = expansion_status), stat="count")
```

Warning: Ignoring unknown parameters: binwidth, bins, pad



Ordinal Regression: Odds of High Emergency Dept Use and/or Hospital Readmission ~ Expansion Status

```
library(VGAM)

## Loading required package: stats4

##
## Attaching package: 'VGAM'

## The following object is masked from 'package:lmtest':
## 
##     lrtest

## The following object is masked from 'package:gam':
## 
##     s

## The following object is masked from 'package:tidyverse':
## 
##     fill
```

```

ord.all <- vglm(hospital_ED_class ~
                 expansion_status +
                 percent_male +
                 average_age +
                 percent_eligible_for_medicaid +
                 I(percent_eligible_for_medicaid**2),
                 cumulative(parallel=TRUE, reverse=TRUE),
                 data=county_dat)

summary(ord.all)

##
## Call:
## vglm(formula = hospital_ED_class ~ expansion_status + percent_male +
##       average_age + percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##       family = cumulative(parallel = TRUE, reverse = TRUE), data = county_dat)
##
## Pearson residuals:
##             Min      1Q  Median      3Q     Max
## logitlink(P[Y>=2]) -6.645 -0.6259  0.2409  0.6042  9.577
## logitlink(P[Y>=3]) -2.781 -0.5718 -0.2235  0.4093  8.312
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1            29.13773   2.86319 10.177 < 2e-16 ***
## (Intercept):2            26.88181   2.85509  9.415 < 2e-16 ***
## expansion_status2         0.44145   0.15409  2.865  0.00417 **
## expansion_status3         0.06553   0.17235  0.380  0.70381
## expansion_status4         0.44713   0.14834  3.014  0.00258 **
## percent_male              -21.75771  1.90653 -11.412 < 2e-16 ***
## average_age                -0.31877  0.03295 -9.673 < 2e-16 ***
## percent_eligible_for_medicaid 28.58780  1.98938 14.370 < 2e-16 ***
## I(percent_eligible_for_medicaid^2) -36.10688  3.44578 -10.479 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y>=2]), logitlink(P[Y>=3])
##
## Residual deviance: 5365.752 on 6083 degrees of freedom
##
## Log-likelihood: -2682.876 on 6083 degrees of freedom
##
## Number of Fisher scoring iterations: 6
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
##           expansion_status2          expansion_status3
##           1.554954e+00          1.067720e+00
##           expansion_status4          percent_male
##           1.563822e+00          3.554223e-10
##           average_age          percent_eligible_for_medicaid

```

```

##                               7.270416e-01                               2.603295e+12
## I(percent_eligible_for_medicaid^2)                         2.084397e-16

Test proportional odds assumption - no overlap in beta estimates confidence interval. Proportional odds assumption is violated.

county_dat['ind3'] = ifelse(county_dat$hospital_ED_class == 3, 1, 0)
county_dat['ind32'] = ifelse(county_dat$hospital_ED_class == 1, 0, 1)

mod.3v12 <- glm(ind3 ~ expansion_status +
                  percent_male +
                  average_age +
                  percent_eligible_for_medicaid +
                  I(percent_eligible_for_medicaid**2),
                  family=binomial,
                  data=county_dat)

mod.32v1 <- glm(ind32 ~ expansion_status +
                  percent_male +
                  average_age +
                  percent_eligible_for_medicaid +
                  I(percent_eligible_for_medicaid**2),
                  family=binomial,
                  data=county_dat)

summary(mod.3v12)

## 
## Call:
## glm(formula = ind3 ~ expansion_status + percent_male + average_age +
##       percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##       family = binomial, data = county_dat)
## 
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max 
## -2.0996   -0.7184   -0.4320    0.5547    2.8688 
## 
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                29.52740   3.75773   7.858 3.91e-15 ***
## expansion_status2          0.37280   0.20374   1.830  0.0673 .  
## expansion_status3         -0.17955   0.23138  -0.776  0.4377    
## expansion_status4          0.33197   0.19819   1.675  0.0939 .  
## percent_male                 -21.77906  2.54873  -8.545 < 2e-16 ***
## average_age                  -0.34106  0.04262  -8.003 1.22e-15 ***
## percent_eligible_for_medicaid  22.20724  2.71752   8.172 3.04e-16 ***
## I(percent_eligible_for_medicaid^2) -26.63211  4.56332  -5.836 5.34e-09 ***
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```

## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 3431.2 on 3045 degrees of freedom
## Residual deviance: 2788.8 on 3038 degrees of freedom
## AIC: 2804.8
##
## Number of Fisher Scoring iterations: 5

summary(mod.32v1)

##
## Call:
## glm(formula = ind32 ~ expansion_status + percent_male + average_age +
##       percent_eligible_for_medicaid + I(percent_eligible_for_medicaid^2),
##       family = binomial, data = county_dat)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.9339 -0.7994  0.3925  0.7404  3.1234
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                30.18076   3.58293   8.423 < 2e-16 ***
## expansion_status2          0.51572   0.18499   2.788 0.005306 **
## expansion_status3          0.31301   0.20785   1.506 0.132069
## expansion_status4          0.61732   0.17673   3.493 0.000478 ***
## percent_male                 -24.93738  2.37775 -10.488 < 2e-16 ***
## average_age                  -0.31929  0.04173  -7.652 1.99e-14 ***
## percent_eligible_for_medicaid 30.46235  2.36156  12.899 < 2e-16 ***
## I(percent_eligible_for_medicaid^2) -36.09792  4.25296 -8.488 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 4005.7 on 3045 degrees of freedom
## Residual deviance: 2950.9 on 3038 degrees of freedom
## AIC: 2966.9
##
## Number of Fisher Scoring iterations: 5

# ord.all_potest <- vglm(hospital_ED_class ~
#                         as.factor(expansion_status) +
#                         percent_male +
#                         average_age +
#                         percent_eligible_for_medicaid +
#                         I(percent_eligible_for_medicaid**2),
#                         cumulative(parallel=FALSE, reverse=T),
#                         data=county_dat)
#
# summary(ord.all_potest)

```