Assignment 1

1) **Determine, via a single-predictor model, whether *# of Cases* is a significant predictor of *Risk-adjusted Rate.* Comment on the overall model fit and the amount of variance accounted for by the predictor. (2 points)**

```
Call:
lm(formula = Risk.adjusted.Rate ~ X..of.Cases, data = Performanc
e.Ratings.Cleaned)

Residuals:
   Min     1Q Median     3Q    Max
-5.342 -4.310 -2.461  2.626 97.678

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.3421386  0.1146578  46.592   <2e-16 ***
X..of.Cases  -0.0001609  0.0001029  -1.563    0.118
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.813 on 3628 degrees of freedom
Multiple R-squared:  0.0006732, Adjusted R-squared:  0.0003978
F-statistic: 2.444 on 1 and 3628 DF,  p-value: 0.1181
```

i) The null Hypothesis in this case is that there is no association between the number of cases and Risk-Adjusted Rate, and the alternate hypothesis is that there is an association between both the variables. In this case, the probability of getting a large t-value [t-value = -1.563] is more than 0.05 [p(t) = 0.118], leading to the rejection of null hypothesis. Hence, Number of cases is not a good predictor of Risk-Adjusted Rate, Considering the probability of t-statistic. ii) The F-statistic of the model is very tiny [F-Value = 2.444] and the probability of obtaining this value is more than 0.05 [p(F) = 0.1181]. Moreover, the amount of variance accounted for by the predictor Risk-Adjusted Rate is just 0.039% which is very tiny [Adjusted R- squared = 0.0003978]. Hence, the null hypothesis is accepted, and the model is not a good fit model.

2) **Next, determine whether in addition to *# of Cases*, *County* is a significant predictor. Interpret the coefficients and comment on which counties appear to differ significantly from the baseline county. (2 points).**
   i) The null Hypothesis in this case is that there is no association between the number of cases, counties and Risk-Adjusted Rate, and the alternate hypothesis is that there is an association between the variables. In this model, the number of cases is significant predictor of Risk adjusted rate as the probability of t-statistic is less than 0.05 [p(t) = 0.00353], implying that the probability of getting a t-value higher than -2.919 is very tiny.
   ii) The counties Alameda, Butte, Contra Costa, Fresno, Contra Costa, Humboldt, Kern, Los Angeles, Marin, Monterey, Napa, Orange, Riverside, Sacramento, San Bernardino, San Diego, San Francisco, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara,

Santa Cruz, Shasta, Sonoma, Stanislaus, Tulare, Ventura are significant predictors of the model as the probability of t-statistic of every county is less than 0.05 and they differ significantly from the baseline county.

```
Call:
lm(formula = Risk.adjusted.Rate ~ X..of.Cases + County, data = Performance.Ratings.Cleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-11.839  -4.216  -2.390   2.743  97.166

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.409e+01  3.496e+00   4.029 5.73e-05
X..of.Cases           -6.311e-04  2.162e-04  -2.919  0.00353
CountyAlameda         -9.323e+00  3.554e+00  -2.623  0.00874
CountyButte           -8.479e+00  3.687e+00  -2.300  0.02151
CountyContra  Costa   -1.067e+01  3.917e+00  -2.723  0.00649
CountyContra Costa    -8.350e+00  3.618e+00  -2.308  0.02107
CountyFresno          -9.038e+00  3.527e+00  -2.562  0.01044
CountyHumboldt        -1.073e+01  3.712e+00  -2.890  0.00388
CountyKern            -7.853e+00  3.553e+00  -2.210  0.02716
CountyLos Angeles     -8.230e+00  3.484e+00  -2.362  0.01823
CountyMarin           -9.680e+00  3.709e+00  -2.610  0.00910
CountyMonterey        -9.954e+00  3.588e+00  -2.774  0.00556
CountyNapa            -8.027e+00  3.595e+00  -2.233  0.02560
CountyOrange          -9.016e+00  3.497e+00  -2.578  0.00998
CountyRiverside       -9.213e+00  3.519e+00  -2.618  0.00888
CountySacramento      -9.365e+00  3.497e+00  -2.678  0.00744
CountySan Bernardino  -8.355e+00  3.500e+00  -2.387  0.01703
CountySan Diego       -8.717e+00  3.496e+00  -2.493  0.01271
CountySan Francisco   -8.251e+00  3.524e+00  -2.342  0.01925
CountySan Joaquin     -9.931e+00  3.582e+00  -2.773  0.00559
CountySan Luis Obispo -8.777e+00  3.698e+00  -2.373  0.01768
```

```
CountySan Mateo        -9.723e+00  3.560e+00  -2.731  0.00634
CountySanta Barbara    -1.032e+01  3.591e+00  -2.873  0.00409
CountySanta Clara      -9.519e+00  3.505e+00  -2.716  0.00664
CountySanta Cruz       -8.061e+00  3.702e+00  -2.178  0.02948
CountyShasta           -8.413e+00  3.589e+00  -2.344  0.01911
CountySolano           -2.244e+00  3.706e+00  -0.605  0.54495
CountySonoma           -1.042e+01  3.594e+00  -2.901  0.00375
CountyStanislaus       -8.916e+00  3.543e+00  -2.516  0.01190
CountyTulare           -8.750e+00  3.680e+00  -2.378  0.01747
CountyVentura          -8.504e+00  3.555e+00  -2.392  0.01682
CountyYuba             -5.855e+00  3.698e+00  -1.583  0.11344

(Intercept)           ***
X..of.Cases           **
CountyAlameda         **
CountyButte           *
CountyContra  Costa   **
CountyContra Costa    *
CountyFresno          *
CountyHumboldt        **
CountyKern            *
CountyLos Angeles     *
CountyMarin           **
CountyMonterey        **
CountyNapa            *
CountyOrange          **
CountyRiverside       **
CountySacramento      **
CountySan Bernardino  *
CountySan Diego       *
CountySan Francisco   *
```

```
CountyMonterey       **
CountyNapa           *
CountyOrange         **
CountyRiverside      **
CountySacramento     **
CountySan Bernardino *
CountySan Diego      *
CountySan Francisco  *
CountySan Joaquin    **
CountySan Luis Obispo *
CountySan Mateo      **
CountySanta Barbara  **
CountySanta Clara    **
CountySanta Cruz     *
CountyShasta         *
CountySolano
CountySonoma         **
CountyStanislaus     *
CountyTulare         *
CountyVentura        *
CountyYuba
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.775 on 3598 degrees of freedom
Multiple R-squared:  0.01998,    Adjusted R-squared:  0.01153
F-statistic: 2.366 on 31 and 3598 DF,  p-value: 3.159e-05
```

iii) The coefficients of the categorical variables are determined by the difference in the means of the baseline county and other counties, and their positive value implies that the mean of the county is higher than the baseline county and vice-versa. The coefficients of the significant counties are Alameda [-9.323e+00], Butte [-8.479e+00], Contra Costa [-1.067e+01], Fresno [-9.038e+00], Contra Costa [-8.350e+00], Humboldt [-1.073e+01], Kern [-7.853e+00], Los Angeles [-8.230e+00], Marin [-9.680e+00], Monterey [-9.954e+00], Napa [8.027e+00], Orange [-9.016e+00], Riverside [-9.213e+00], Sacramento [-9.365e+00], San Bernardino [-8.355e+00], San Diego [-8.717e+00], San Francisco [-8.251e+00], San Joaquin [9.931e+00], San Luis Obispo [-8.777e+00], San Mateo [-9.723e+00], Santa Barbara [1.032e+01], Santa Clara [-9.519e+00], Santa Cruz [-8.061e+00], Shasta [-8.413e+00], Sonoma [-1.042e+01], Stanislaus [-8.916e+00], Tulare [-8.750e+00], Ventura [-8.504e+00]. The coefficients of all the counties are negative implying the lower mean than the baseline county.

iv) The F-statistic of the model is very tiny [F-Value = 2.366] and the probability of obtaining this value is less than 0.05 [p(F) = 0.00003159]. Moreover, the amount of variance accounted for by the predictor Risk-Adjusted Rate is just 1.153% which is very tiny [Adjusted R- squared = 0.01153]. Hence, the null hypothesis is accepted, and the model is not a good fit model.

3) **To the previous model, add *Performance Measure* as a predictor and comment on the consequences of adding this a predictor on change in the overall model fit, in terms of statistical significance and practical significance. (3 points)**

   i) Upon adding the Performance Measure as a predictor, the county and the number of cases became insignificant predictors of risk-Adjusted Rate in this model, the Performance Measure is the only significant predictor of risk-Adjusted Rate. This means that there is some data redundancy between the county, Performance Measure, and the number of

cases. In this model, the categories CABG + Valve Operative Mortality [p(t) = < 2e-16], CABG Operative Mortality [p(t) = < 2e-16], Operative Mortality [p(t) = < 2e-16], and Post-Operative Stroke [p(t) = < 2e-16] are significant predictors of Risk adjusted rate as the probability of tstatistic is less than 0.05.

```
Call:
lm(formula = Risk.adjusted.Rate ~ X..of.Cases + County + Performance.Measure,
    data = Performance.Ratings.Cleaned)

Residuals:
    Min      1Q   Median      3Q      Max
-13.241  -1.979  -0.422   1.208   95.782

Coefficients:
                                          Estimate
(Intercept)                              1.349e+01
X..of.Cases                             -1.159e-04
CountyAlameda                           -1.856e+00
CountyButte                             -1.048e+00
CountyContra  Costa                     -2.904e+00
CountyContra Costa                      -1.003e+00
CountyFresno                            -1.644e+00
CountyHumboldt                          -3.223e+00
CountyKern                              -3.878e-01
CountyLos Angeles                       -7.598e-01
CountyMarin                             -2.185e+00
CountyMonterey                          -2.494e+00
CountyNapa                              -5.491e-01
CountyOrange                            -1.550e+00
CountyRiverside                         -1.704e+00
CountySacramento                        -2.037e+00
CountySan Bernardino                    -9.186e-01
CountySan Diego                         -1.264e+00
CountySan Francisco                     -8.158e-01
CountySan Joaquin                       -2.469e+00
```

```
CountySan Luis Obispo                                          -1.313e+00
CountySan Mateo                                                -2.239e+00
CountySanta Barbara                                            -2.850e+00
CountySanta Clara                                              -2.069e+00
CountySanta Cruz                                               -5.877e-01
CountyShasta                                                   -9.521e-01
CountySolano                                                    5.244e+00
CountySonoma                                                   -2.949e+00
CountyStanislaus                                               -1.490e+00
CountyTulare                                                   -1.339e+00
CountyVentura                                                  -1.032e+00
CountyYuba                                                      1.608e+00
Performance.Measure30-day Readmission                          1.114e+00
Performance.MeasureCABG + Valve Operative Mortality           -5.495e+00
Performance.MeasureCABG 30-Day Readmission                     6.144e-02
Performance.MeasureCABG Operative Mortality                   -9.209e+00
Performance.MeasureCABG Operative Mortality                   -9.868e+00
Performance.MeasureOperative Mortality                        -9.932e+00
Performance.MeasurePost-Operative Stroke                      -1.064e+01
Performance.MeasurePost-Operative Stroke                      -1.075e+01
                                                              Std. Error
(Intercept)                                                    2.883e+00
X..of.Cases                                                    1.756e-04
CountyAlameda                                                  2.873e+00
CountyButte                                                    2.977e+00
CountyContra  Costa                                            3.162e+00
CountyContra Costa                                             2.923e+00
CountyFresno                                                   2.852e+00
CountyHumboldt                                                 2.998e+00
CountyKern                                                     2.872e+00
```

```
CountyLos Angeles                                           2.818e+00
CountyMarin                                                 2.995e+00
CountyMonterey                                              2.900e+00
CountyNapa                                                  2.905e+00
CountyOrange                                                2.829e+00
CountyRiverside                                             2.846e+00
CountySacramento                                            2.827e+00
CountySan Bernardino                                        2.830e+00
CountySan Diego                                             2.828e+00
CountySan Francisco                                         2.849e+00
CountySan Joaquin                                           2.895e+00
CountySan Luis Obispo                                       2.987e+00
CountySan Mateo                                             2.878e+00
CountySanta Barbara                                         2.902e+00
CountySanta Clara                                           2.834e+00
CountySanta Cruz                                            2.989e+00
CountyShasta                                                2.900e+00
CountySolano                                                2.993e+00
CountySonoma                                                2.905e+00
CountyStanislaus                                            2.864e+00
CountyTulare                                                2.972e+00
CountyVentura                                               2.874e+00
CountyYuba                                                  2.986e+00
Performance.Measure30-day Readmission                       5.556e-01
Performance.MeasureCABG + Valve Operative Mortality         5.151e-01
Performance.MeasureCABG 30-Day Readmission                  5.893e-01
Performance.MeasureCABG Operative Mortality                 5.561e-01
Performance.MeasureCABG Operative Mortality                 5.374e-01
Performance.MeasureOperative Mortality                      6.802e-01
Performance.MeasurePost-Operative Stroke                    5.555e-01
```

```
                                          t value
(Intercept)                                 4.681
X..of.Cases                                -0.660
CountyAlameda                              -0.646
CountyButte                                -0.352
CountyContra  Costa                        -0.918
CountyContra Costa                         -0.343
CountyFresno                               -0.576
CountyHumboldt                             -1.075
CountyKern                                 -0.135
CountyLos Angeles                          -0.270
CountyMarin                                -0.729
CountyMonterey                             -0.860
CountyNapa                                 -0.189
CountyOrange                               -0.548
CountyRiverside                            -0.599
CountySacramento                           -0.721
CountySan Bernardino                       -0.325
CountySan Diego                            -0.447
CountySan Francisco                        -0.286
CountySan Joaquin                          -0.853
CountySan Luis Obispo                      -0.440
CountySan Mateo                            -0.778
CountySanta Barbara                        -0.982
CountySanta Clara                          -0.730
CountySanta Cruz                           -0.197
CountyShasta                               -0.328
CountySolano                                1.752
CountySonoma                               -1.015
CountyStanislaus                           -0.520
```

```
CountyTulare                                                -0.451
CountyVentura                                               -0.359
CountyYuba                                                   0.538
Performance.Measure30-day Readmission                        0.201
Performance.MeasureCABG + Valve Operative Mortality        -10.668
Performance.MeasureCABG 30-Day Readmission                   0.104
Performance.MeasureCABG Operative Mortality                -16.560
Performance.MeasureCABG Operative Mortality                -18.364
Performance.MeasureOperative Mortality                     -14.601
Performance.MeasurePost-Operative Stroke                   -19.156
Performance.MeasurePost-Operative Stroke                   -20.453
                                                           Pr(>|t|)
(Intercept)                                                2.96e-06
X..of.Cases                                                  0.5093
CountyAlameda                                                0.5182
CountyButte                                                  0.7248
CountyContra  Costa                                          0.3585
CountyContra Costa                                           0.7316
CountyFresno                                                 0.5644
CountyHumboldt                                               0.2823
CountyKern                                                   0.8926
CountyLos Angeles                                            0.7875
CountyMarin                                                  0.4658
CountyMonterey                                               0.3899
CountyNapa                                                   0.8501
CountyOrange                                                 0.5838
CountyRiverside                                              0.5494
CountySacramento                                             0.4711
CountySan Bernardino                                         0.7455
CountySan Diego                                              0.6549
```

```
CountySan Francisco                                          0.7746
CountySan Joaquin                                            0.3938
CountySan Luis Obispo                                        0.6601
CountySan Mateo                                              0.4367
CountySanta Barbara                                          0.3261
CountySanta Clara                                            0.4654
CountySanta Cruz                                             0.8442
CountyShasta                                                 0.7427
CountySolano                                                 0.0799
CountySonoma                                                 0.3101
CountyStanislaus                                             0.6030
CountyTulare                                                 0.6523
CountyVentura                                                0.7195
CountyYuba                                                   0.5903
Performance.Measure30-day Readmission                        0.8411
Performance.MeasureCABG + Valve Operative Mortality        < 2e-16
Performance.MeasureCABG 30-Day Readmission                   0.9170
Performance.MeasureCABG Operative Mortality                < 2e-16
Performance.MeasureCABG Operative Mortality                < 2e-16
Performance.MeasureOperative Mortality                     < 2e-16
Performance.MeasurePost-Operative Stroke                   < 2e-16
Performance.MeasurePost-Operative Stroke                   < 2e-16

(Intercept)                                                ***
X..of.Cases
CountyAlameda
CountyButte
CountyContra  Costa
CountyContra Costa
CountyFresno
```

```
CountyHumboldt
CountyKern
CountyLos Angeles
CountyMarin
CountyMonterey
CountyNapa
CountyOrange
CountyRiverside
CountySacramento
CountySan Bernardino
CountySan Diego
CountySan Francisco
CountySan Joaquin
CountySan Luis Obispo
CountySan Mateo
CountySanta Barbara
CountySanta Clara
CountySanta Cruz
CountyShasta
CountySolano                                          .
CountySonoma
CountyStanislaus
CountyTulare
CountyVentura
CountyYuba
Performance.Measure30-day Readmission
Performance.MeasureCABG + Valve Operative Mortality ***
Performance.MeasureCABG 30-Day Readmission
Performance.MeasureCABG Operative Mortality           ***
Performance.MeasureCABG Operative Mortality           ***
```

```
CountySan Joaquin
CountySan Luis Obispo
CountySan Mateo
CountySanta Barbara
CountySanta Clara
CountySanta Cruz
CountyShasta
CountySolano                                          .
CountySonoma
CountyStanislaus
CountyTulare
CountyVentura
CountyYuba
Performance.Measure30-day Readmission
Performance.MeasureCABG + Valve Operative Mortality ***
Performance.MeasureCABG 30-Day Readmission
Performance.MeasureCABG Operative Mortality           ***
Performance.MeasureCABG Operative Mortality           ***
Performance.MeasureOperative Mortality                ***
Performance.MeasurePost-Operative Stroke              ***
Performance.MeasurePost-Operative Stroke              ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.405 on 3590 degrees of freedom
Multiple R-squared:  0.3777,     Adjusted R-squared:  0.3709
F-statistic: 55.86 on 39 and 3590 DF,   p-value: < 2.2e-16
```

ii) The F-statistic of the model is better than the previous models [F-Value = 55.86] and the probability of obtaining this value is less than 0.05 [p(F) = < 2.2e-16]. Moreover, the amount of variance accounted for by the predictor Risk-Adjusted Rate is 37.09% which is

better than the previous models [Adjusted R- squared = 0.3709]. Hence, the alternate hypothesis is accepted, and the model is a good fit model.

4) **Next, add *# of Adverse Events* as a predictor and comment on the consequences of adding this a predictor on change in the overall model fit, in terms of statistical significance and practical significance. (3 points).**

   i) Upon adding the number of adverse events as a predictor, the county and the number of cases, and the number of adverse events became insignificant predictors of risk-Adjusted Rate in this model, the Performance Measure, and the county Solano are the only significant predictors of risk-Adjusted Rate. This means that there is some data redundancy between the county, Performance Measure, number of adverse events and the number of cases. In this model, the categories CABG + Valve Operative Mortality [p(t) = < 2e-16], CABG Operative Mortality [p(t) = < 2e-16], Operative Mortality [p(t) = < 2e-16], and Post-Operative Stroke [p(t) = < 2e-16] of performance measure are significant predictors of Risk adjusted rate as the probability of t-statistic is less than 0.05 and the category Solano [p(t) = 0.0485] is significant predictor in this model..

   ii) The F-statistic of the model is better than the previous models [F-Value = 54.49] and the probability of obtaining this value is less than 0.05 [p(F) = < 2.2e-16]. Moreover, the amount of variance accounted for by the predictor Risk-Adjusted Rate is 37.09% which is better than the previous models [Adjusted R- squared = 0.3709]. Hence, the alternate hypothesis is accepted, and the model is a good fit model, but it has similar fit as the previous model.

```
Call:
lm(formula = Risk.adjusted.Rate ~ X..of.Cases + County + Performance.Measure +
    X..of.Adverse.Events, data = Performance.Ratings.cleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-13.242  -1.983  -0.426   1.200  95.781

Coefficients:
                                            Estimate
(Intercept)                                1.243e+01
X..of.Cases                               -1.371e-04
CountyAlameda                             -8.585e-01
CountyButte                               -5.635e-02
CountyContra  Costa                       -1.906e+00
CountyContra Costa                        -5.649e-03
CountyFresno                              -6.559e-01
CountyHumboldt                            -2.220e+00
CountyKern                                 6.079e-01
CountyLos Angeles                          2.368e-01
CountyMarin                               -1.183e+00
CountyMonterey                            -1.497e+00
CountyNapa                                 4.490e-01
CountyOrange                              -5.519e-01
CountyRiverside                           -7.100e-01
CountySacramento                          -1.050e+00
CountySan Bernardino                       7.120e-02
CountySan Diego                           -2.690e-01
CountySan Francisco                        1.792e-01
CountySan Joaquin                         -1.477e+00
```

```
CountySan Luis Obispo                                     -3.161e-01
CountySan Mateo                                           -1.239e+00
CountySanta Barbara                                       -1.851e+00
CountySanta Clara                                         -1.073e+00
CountySanta Cruz                                           4.100e-01
CountyShasta                                               4.412e-02
CountySolano                                               6.242e+00
CountySonoma                                              -1.947e+00
CountyStanislaus                                          -5.008e-01
CountyTulare                                              -3.498e-01
CountyVentura                                             -3.622e-02
CountyYuba                                                 2.598e+00
Performance.Measure30-day Readmission                      1.491e-01
Performance.MeasureCABG + Valve Operative Mortality       -5.427e+00
Performance.MeasureCABG 30-Day Readmission                 9.137e-02
Performance.MeasureCABG Operative Mortality               -9.142e+00
Performance.MeasureCABG Operative Mortality               -9.797e+00
Performance.MeasureOperative Mortality                    -9.859e+00
Performance.MeasurePost-Operative Stroke                  -1.057e+01
Performance.MeasurePost-Operative Stroke                  -1.068e+01
X..of.Adverse.Events                                       2.162e-03
                                                          Std. Error
(Intercept)                                                3.082e+00
X..of.Cases                                                1.769e-04
CountyAlameda                                              3.049e+00
CountyButte                                                3.146e+00
CountyContra  Costa                                        3.323e+00
CountyContra Costa                                         3.097e+00
CountyFresno                                               3.026e+00
CountyHumboldt                                             3.169e+00
```

```
CountyKern                                                3.048e+00
CountyLos Angeles                                         2.998e+00
CountyMarin                                               3.166e+00
CountyMonterey                                            3.074e+00
CountyNapa                                                3.080e+00
CountyOrange                                              3.008e+00
CountyRiverside                                           3.023e+00
CountySacramento                                          3.002e+00
CountySan Bernardino                                      3.006e+00
CountySan Diego                                           3.006e+00
CountySan Francisco                                       3.026e+00
CountySan Joaquin                                         3.068e+00
CountySan Luis Obispo                                     3.157e+00
CountySan Mateo                                           3.055e+00
CountySanta Barbara                                       3.077e+00
CountySanta Clara                                         3.012e+00
CountySanta Cruz                                          3.159e+00
CountyShasta                                              3.075e+00
CountySolano                                              3.163e+00
CountySonoma                                              3.081e+00
CountyStanislaus                                          3.038e+00
CountyTulare                                              3.140e+00
CountyVentura                                             3.050e+00
CountyYuba                                                3.154e+00
Performance.Measure30-day Readmission                     5.569e-01
Performance.MeasureCABG + Valve Operative Mortality       5.198e-01
Performance.MeasureCABG 30-Day Readmission                5.901e-01
Performance.MeasureCABG Operative Mortality               5.603e-01
Performance.MeasureCABG Operative Mortality               5.422e-01
Performance.MeasureOperative Mortality                    6.843e-01
```

```
Performance.MeasureOperative Mortality                               6.843e-01
Performance.MeasurePost-Operative Stroke                             5.598e-01
Performance.MeasurePost-Operative Stroke                             5.312e-01
X..of.Adverse.Events                                                 2.215e-03
                                                          t value
(Intercept)                                                 4.032
X..of.Cases                                                -0.775
CountyAlameda                                              -0.282
CountyButte                                                -0.018
CountyContra  Costa                                        -0.574
CountyContra Costa                                         -0.002
CountyFresno                                               -0.217
CountyHumboldt                                             -0.701
CountyKern                                                  0.199
CountyLos Angeles                                           0.079
CountyMarin                                                -0.374
CountyMonterey                                             -0.487
CountyNapa                                                  0.146
CountyOrange                                               -0.184
CountyRiverside                                            -0.235
CountySacramento                                           -0.350
CountySan Bernardino                                        0.024
CountySan Diego                                            -0.089
CountySan Francisco                                         0.059
CountySan Joaquin                                          -0.481
CountySan Luis Obispo                                      -0.100
CountySan Mateo                                            -0.406
CountySanta Barbara                                        -0.601
CountySanta Clara                                          -0.356
CountySanta Cruz                                            0.130
```

```
CountyShasta                                                         0.014
CountySolano                                                         1.973
CountySonoma                                                        -0.632
CountyStanislaus                                                   -0.165
CountyTulare                                                       -0.111
CountyVentura                                                      -0.012
CountyYuba                                                          0.824
Performance.Measure30-day Readmission                              0.268
Performance.MeasureCABG + Valve Operative Mortality             -10.440
Performance.MeasureCABG 30-Day Readmission                        0.155
Performance.MeasureCABG Operative Mortality                     -16.317
Performance.MeasureCABG Operative Mortality                     -18.067
Performance.MeasureOperative Mortality                          -14.407
Performance.MeasurePost-Operative Stroke                        -18.888
Performance.MeasurePost-Operative Stroke                        -20.103
X..of.Adverse.Events                                              0.976
                                                          Pr(>|t|)
(Intercept)                                               5.65e-05
X..of.Cases                                                 0.4384
CountyAlameda                                               0.7783
CountyButte                                                 0.9857
CountyContra  Costa                                         0.5663
CountyContra Costa                                          0.9985
CountyFresno                                                0.8284
CountyHumboldt                                              0.4836
CountyKern                                                  0.8419
CountyLos Angeles                                           0.9370
CountyMarin                                                 0.7088
CountyMonterey                                              0.6263
CountyNapa                                                  0.8841
```

```
CountyOrange                                         0.8544
CountyRiverside                                      0.8143
CountySacramento                                     0.7265
CountySan Bernardino                                 0.9811
CountySan Diego                                      0.9287
CountySan Francisco                                  0.9528
CountySan Joaquin                                    0.6303
CountySan Luis Obispo                                0.9202
CountySan Mateo                                      0.6851
CountySanta Barbara                                  0.5476
CountySanta Clara                                    0.7216
CountySanta Cruz                                     0.8967
CountyShasta                                         0.9886
CountySolano                                         0.0485
CountySonoma                                         0.5274
CountyStanislaus                                     0.8691
CountyTulare                                         0.9113
CountyVentura                                        0.9905
CountyYuba                                           0.4101
Performance.Measure30-day Readmission                0.7889
Performance.MeasureCABG + Valve Operative Mortality  < 2e-16
Performance.MeasureCABG 30-Day Readmission           0.8770
Performance.MeasureCABG Operative Mortality          < 2e-16
Performance.MeasureCABG Operative Mortality          < 2e-16
Performance.MeasureOperative Mortality               < 2e-16
Performance.MeasurePost-Operative Stroke             < 2e-16
Performance.MeasurePost-Operative Stroke             < 2e-16
X..of.Adverse.Events                                 0.3290
```

```
(Intercept)                                          ***
X..of.Cases
CountyAlameda
CountyButte
CountyContra  Costa
CountyContra Costa
CountyFresno
CountyHumboldt
CountyKern
CountyLos Angeles
CountyMarin
CountyMonterey
CountyNapa
CountyOrange
CountyRiverside
CountySacramento
CountySan Bernardino
CountySan Diego
CountySan Francisco
CountySan Joaquin
CountySan Luis Obispo
CountySan Mateo
CountySanta Barbara
CountySanta Clara
CountySanta Cruz
CountyShasta
CountySolano                                         *
CountySonoma
CountyStanislaus
CountyTulare
CountyVentura
```

```
CountySan Joaquin
CountySan Luis Obispo
CountySan Mateo
CountySanta Barbara
CountySanta Clara
CountySanta Cruz
CountyShasta
CountySolano                                                       *
CountySonoma
CountyStanislaus
CountyTulare
CountyVentura
CountyYuba
Performance.Measure30-day Readmission
Performance.MeasureCABG + Valve Operative Mortality ***
Performance.MeasureCABG 30-Day Readmission
Performance.MeasureCABG Operative Mortality            ***
Performance.MeasureCABG Operative Mortality            ***
Performance.MeasureOperative Mortality                 ***
Performance.MeasurePost-Operative Stroke               ***
Performance.MeasurePost-Operative Stroke               ***
X..of.Adverse.Events
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.405 on 3589 degrees of freedom
Multiple R-squared:  0.3778,     Adjusted R-squared:  0.3709
F-statistic: 54.49 on 40 and 3589 DF,  p-value: < 2.2e-16
```

5) **Of all the models fit so far, identify based on relevant information, the best model in terms of statistical and practical significance. Comment on which predictors are significant and what these results mean. (3 points).**

```
> anova(lm.model.1, lm.model.2, lm.model.3, lm.model.4)
Analysis of Variance Table

Model 1: Risk.adjusted.Rate ~ X..of.Cases
Model 2: Risk.adjusted.Rate ~ X..of.Cases + County
Model 3: Risk.adjusted.Rate ~ X..of.Cases + County + Performance.Measure
Model 4: Risk.adjusted.Rate ~ X..of.Cases + County + Performance.Measure +
    X..of.Adverse.Events
  Res.Df    RSS Df Sum of Sq      F     Pr(>F)
1   3628 168394
2   3598 165142 30      3253   3.7116 4.459e-11 ***
3   3590 104869  8     60273 257.9158 < 2.2e-16 ***
4   3589 104841  1        28   0.9529    0.329
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

i)   Of all the models, model 2 and model 3 are statistically significant as the probability of F statistic is less than 0.05 and p(F) are respectively, 4.459e-11, and < 2.2e-16 respectively. And among model 2 and model 3, model 3 is more significant, considering the F-statistic and the degrees of freedom. The F-statistic of model 3 is 257.91, which is very large compared to model 2 and the degree of freedom is 3590, which is less than model 2.

ii) Keeping this significance in the mind, it can be said that in this model, the categories CABG + Valve Operative Mortality [p(t) = < 2e-16], CABG Operative Mortality [p(t) = < 2e-16], Operative Mortality [p(t) = < 2e-16], and Post-Operative Stroke [p(t) = < 2e-16] of

performance measure are significant predictors of Risk adjusted rate as the probability of tstatistic is less than 0.05.

6) **Next, determine whether the association between *# of Adverse Events* and *Risk-adjusted Rate* varies significantly across different levels of *Performance Measure*. Comment on the results. (3 points).**

    i) The coefficients of the categorical variables are determined by the difference in the means of the baseline county and other counties, and their positive value implies that the mean of the county is higher than the baseline county and vice-versa. The coefficients of the significant categories of performance measure are 30-day Readmission [1.470e-01], CABG + Valve Operative Mortality [-5.455e+00], CABG 30-Day Readmission [6.858e-02], CABG Operative Mortality [-9.188e+00], Operative Mortality [-9.883e+00], Post-Operative Stroke [-1.063e+01]. The coefficients of 30-day Readmission, and CABG 30-Day Readmission are positive implying the greater mean than the baseline performance measure, and all the other categories have lower mean than the baseline category.

```
Call:
lm(formula = Risk.adjusted.Rate ~ X..of.Adverse.Events + Performance.Measure,
    data = Performance.Ratings.Cleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-12.352  -1.968  -0.494   1.162  96.269

Coefficients:
                                                       Estimate
(Intercept)                                            1.220e+01
X..of.Adverse.Events                                   6.282e-04
Performance.Measure30-day Readmission                  1.470e-01
Performance.MeasureCABG + Valve Operative Mortality   -5.455e+00
Performance.MeasureCABG 30-Day Readmission             6.858e-02
Performance.MeasureCABG Operative Mortality           -9.188e+00
Performance.MeasureCABG Operative Mortality           -9.827e+00
Performance.MeasureOperative Mortality                -9.883e+00
Performance.MeasurePost-Operative Stroke              -1.063e+01
Performance.MeasurePost-Operative Stroke              -1.073e+01
```

    ii) In this model, the categories CABG + Valve Operative Mortality [p(t) = < 2e-16], CABG Operative Mortality [p(t) = < 2e-16], Operative Mortality [p(t) = < 2e-16], and Post-Operative Stroke [p(t) = < 2e-16] of performance measure are significant predictors of Risk adjusted rate as the probability of t-statistic is less than 0.05.

    iii) There is a large value of F-statistic in this model [F-value = 226.3], and the probability of the observations having this F-value is less than 0.05 [p(F) = < 2.2e-16]. Also, the amount of variance accounted for by the predictor Risk-Adjusted Rate is 35.85 [Adjusted R- squared = 0.3585]. Hence, the model has a good fit, and the alternate hypothesis is accepted.

```
                                                                         Pr(>|t|)
(Intercept)                                                               <2e-16
X..of.Adverse.Events                                                      0.689
Performance.Measure30-day Readmission                                     0.793
Performance.MeasureCABG + Valve Operative Mortality                       <2e-16
Performance.MeasureCABG 30-Day Readmission                                0.908
Performance.MeasureCABG Operative Mortality                               <2e-16
Performance.MeasureCABG Operative Mortality                               <2e-16
Performance.MeasureOperative Mortality                                    <2e-16
Performance.MeasurePost-Operative Stroke                                  <2e-16
Performance.MeasurePost-Operative Stroke                                  <2e-16

(Intercept)                                                               ***
X..of.Adverse.Events
Performance.Measure30-day Readmission
Performance.MeasureCABG + Valve Operative Mortality ***
Performance.MeasureCABG 30-Day Readmission
Performance.MeasureCABG Operative Mortality                               ***
Performance.MeasureCABG Operative Mortality                               ***
Performance.MeasureOperative Mortality                                    ***
Performance.MeasurePost-Operative Stroke                                  ***
Performance.MeasurePost-Operative Stroke                                  ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.458 on 3620 degrees of freedom
Multiple R-squared:   0.3601,      Adjusted R-squared:   0.3585
F-statistic: 226.3 on 9 and 3620 DF,   p-value: < 2.2e-16
```

7) Finally, comment on whether the model you built in (7) is better than all the preceding models on the basis of appropriate evidence. (3 points).

```
> anova(lm.model.3, lm.model.final)
Analysis of Variance Table

Model 1: Risk.adjusted.Rate ~ X..of.Cases + County + Performance.Measure
Model 2: Risk.adjusted.Rate ~ X..of.Adverse.Events + Performance.Measure
  Res.Df    RSS  Df Sum of Sq      F   Pr(>F)
1   3590 104869
2   3620 107835 -30   -2966.8 3.3855 1.523e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

i) Among all the models, model 5 is significant model as the probability of F-statistic is less than 0.05 and p(F) are respectively, 1.523e-09, considering the F-statistic and the degrees of freedom. The F-statistic of model 5 is 3.385, and the degree of freedom is 3620.

ii) The amount of variance accounted for by the predictor Risk-Adjusted Rate for model 5 is 35.85% [Adjusted R- squared = 0.3585].