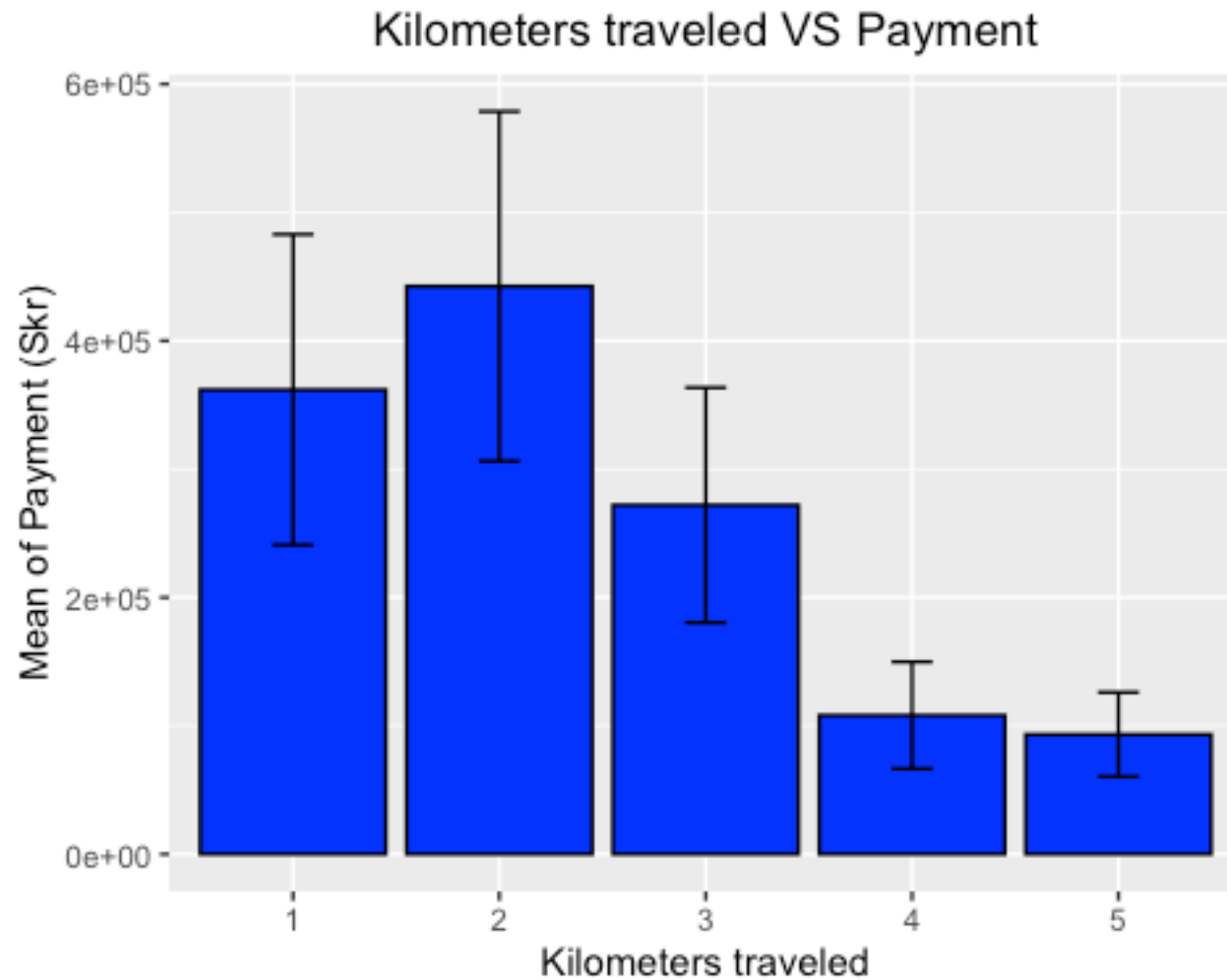


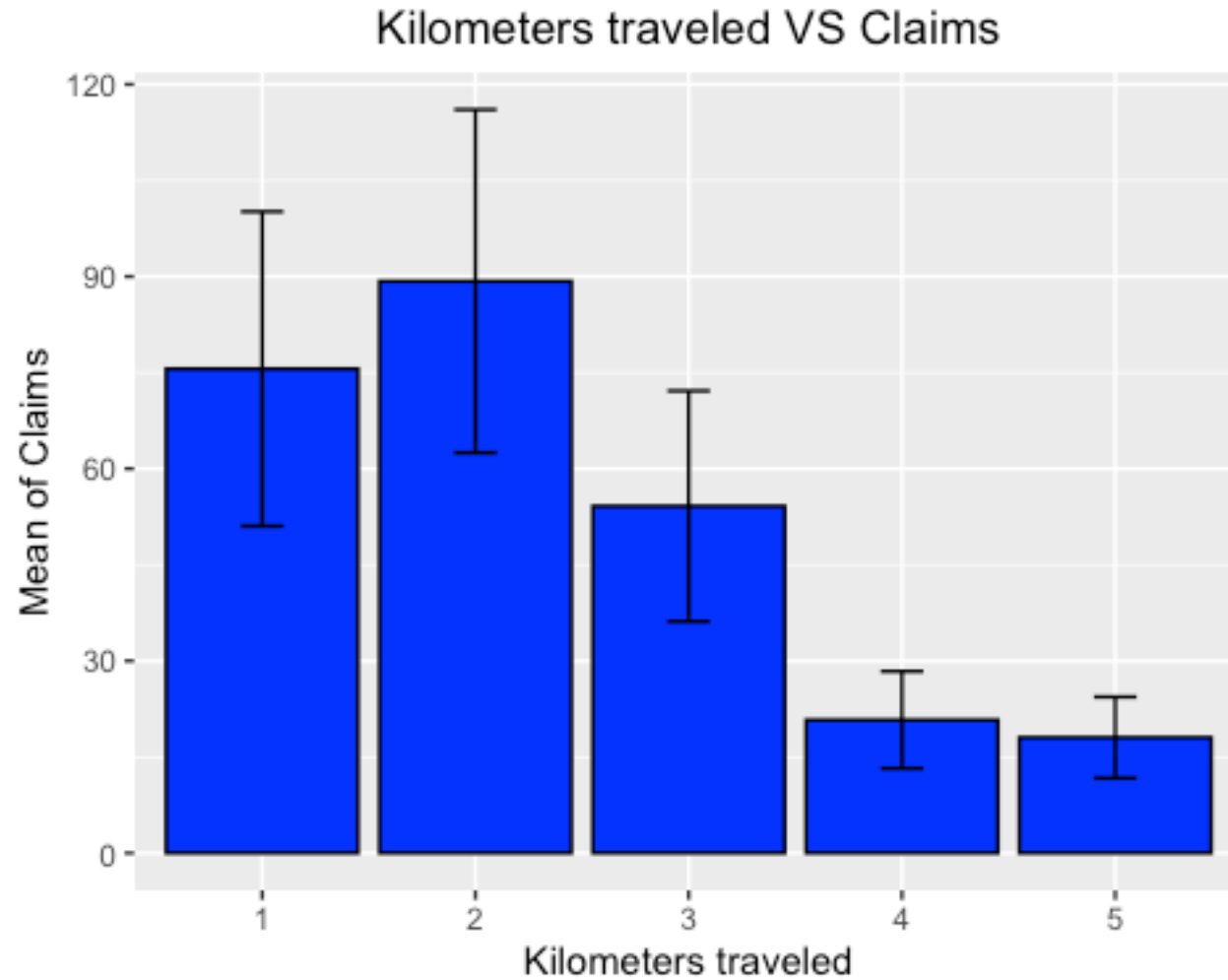
Assignment1

Ip Kim Wai (UP2067015)

Question A

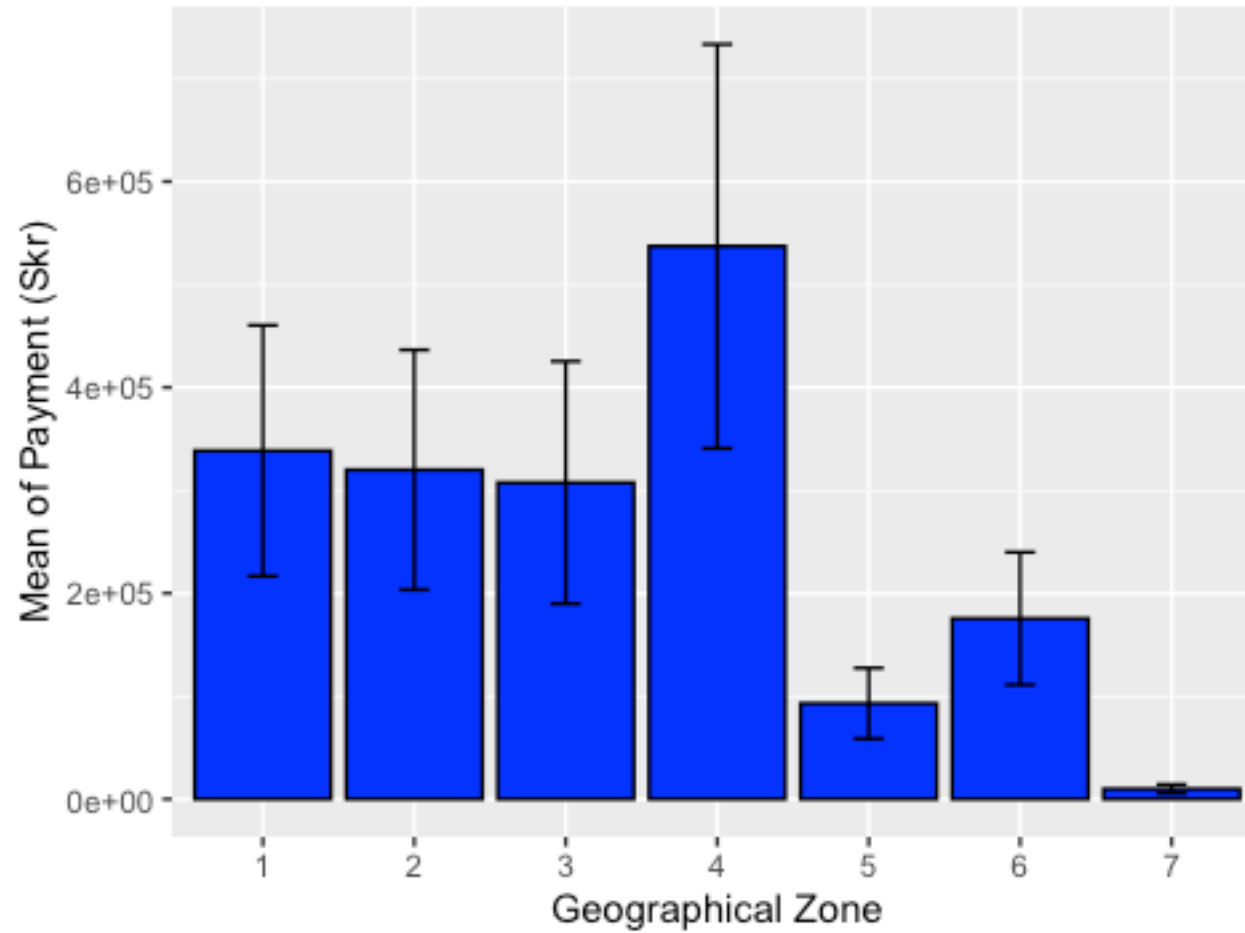


- Right-skewed
- Mean of payment of Kilometres 1 to 3 vary between 200000 to 450000 skr
- Mean of payment of Kilometres 4 and 5 are around 100000 skr
- Largest difference happens between Kilometres 2 and 5 (~350000 skr)



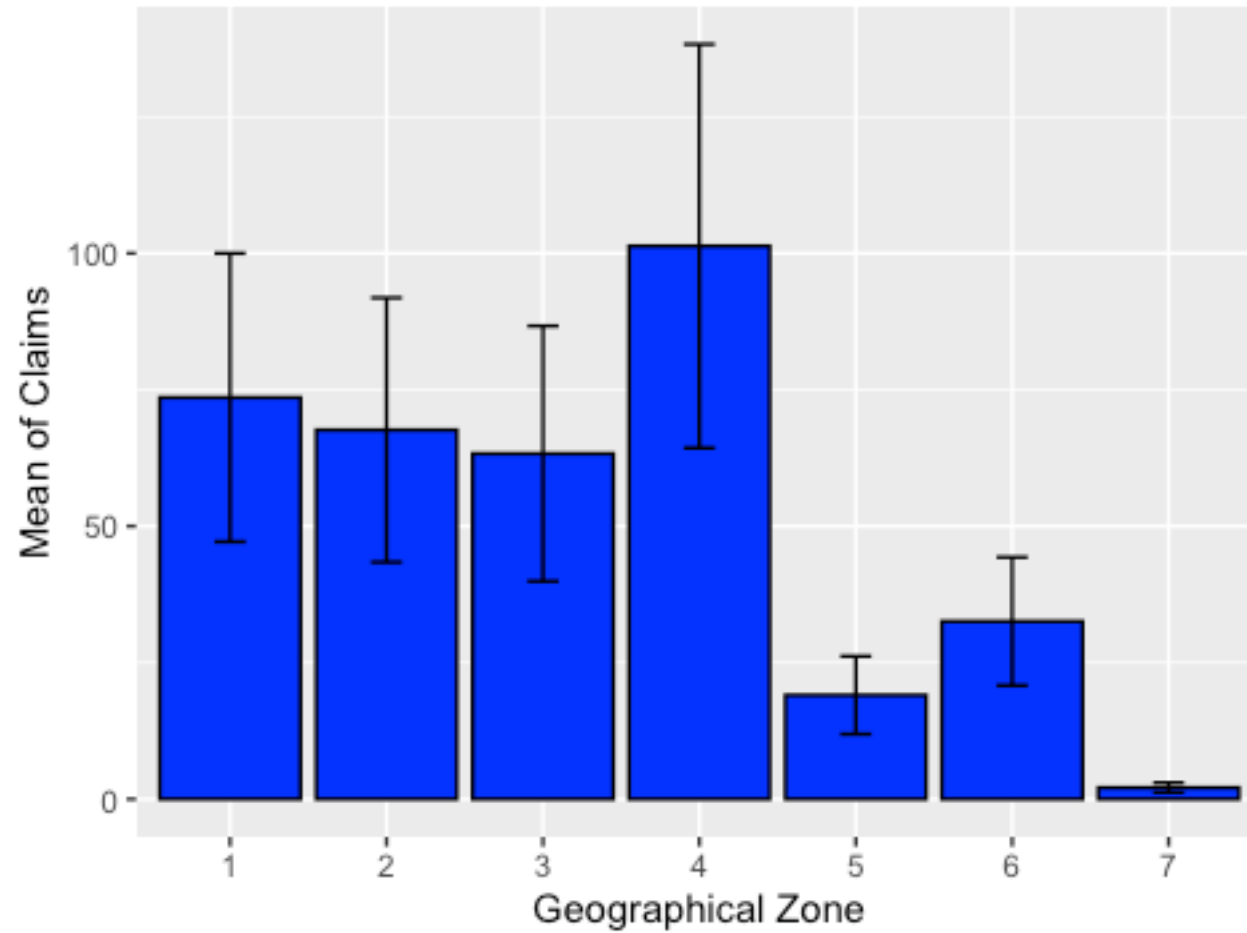
- Right-skewed
- Mean of claim amount of Kilometres 1 to 3 vary between 50 to 90 cases
- Mean of claim amount of kilometres 4 and 5 are around 20 cases
- Largest difference happens between Kilometres 2 and 5 (~70 cases)

Geographical Zone VS Payment



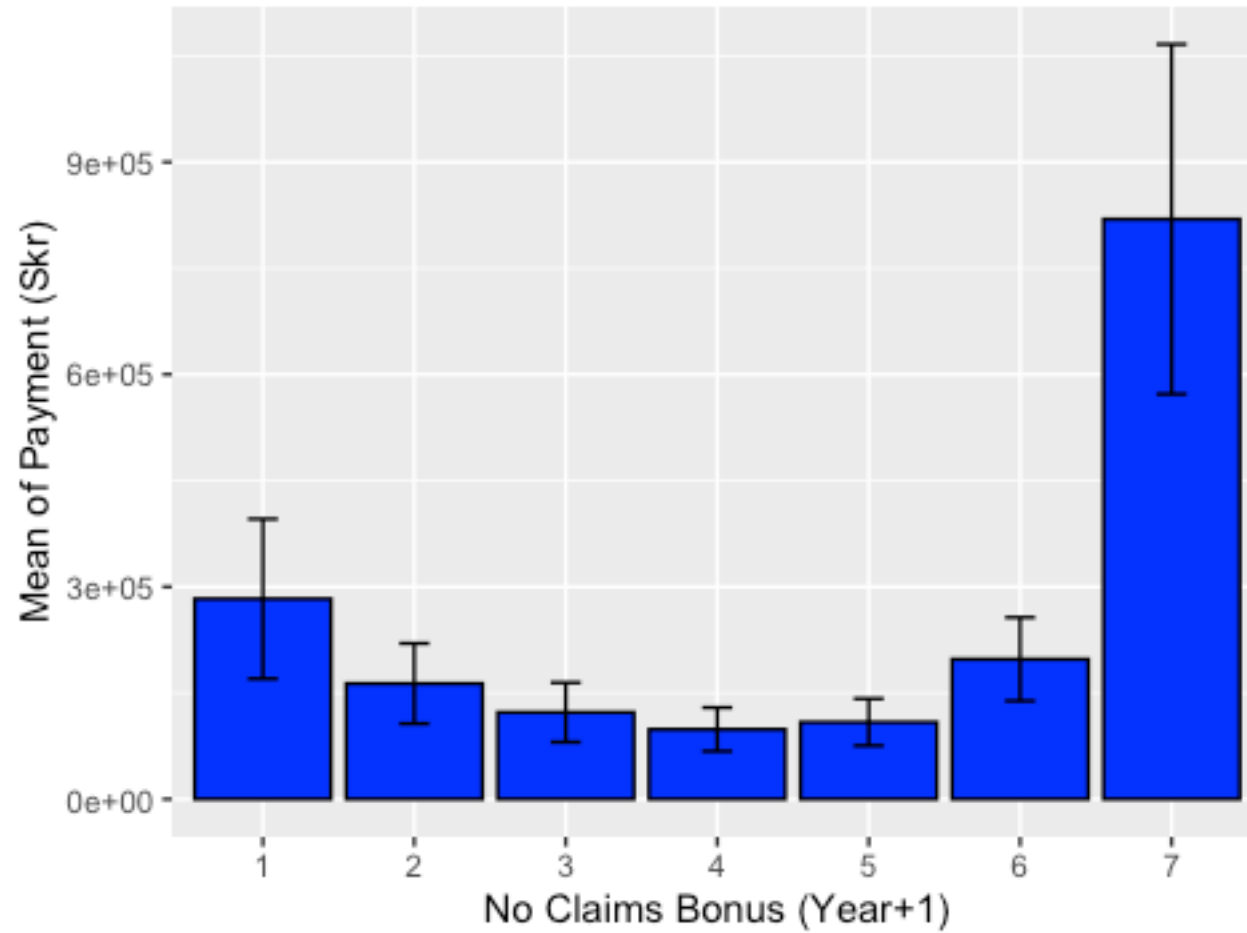
- Zone 1 to 4 have mean of claim amount vary between 300000 to 550000 skr
- Zone 5 to 7 have cases less than 200000 skr
- Difference between Zone 4 and Zone 7 is around 500000 skr

Geographical Zone VS Claims



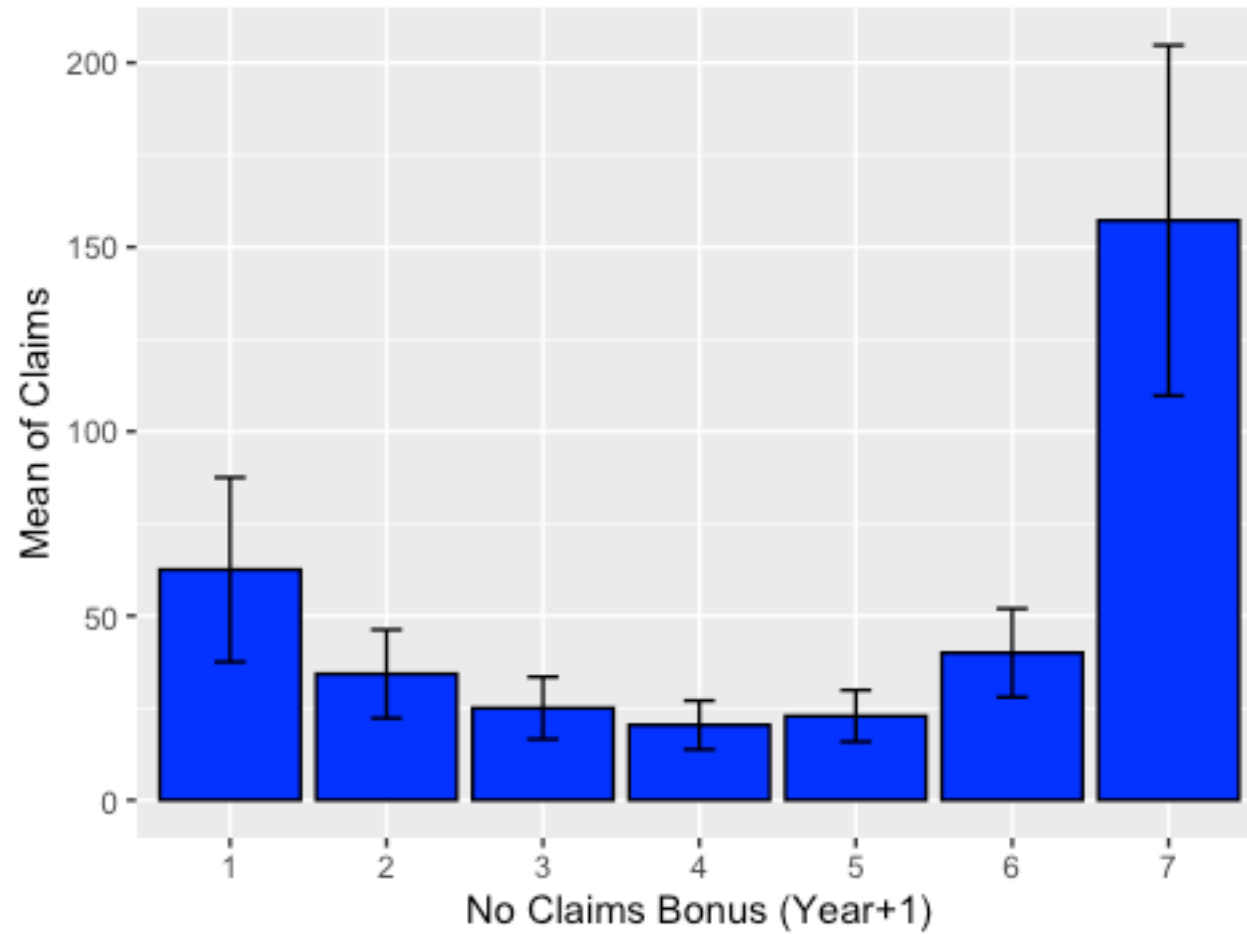
- Zone 1 to 4 have mean of claim amount vary between 60 to 105 cases
- Zone 5 to 7 have cases less than 40 cases
- Difference between Zone 4 and Zone 7 is around 100 cases

No Claims Bonus VS Payment

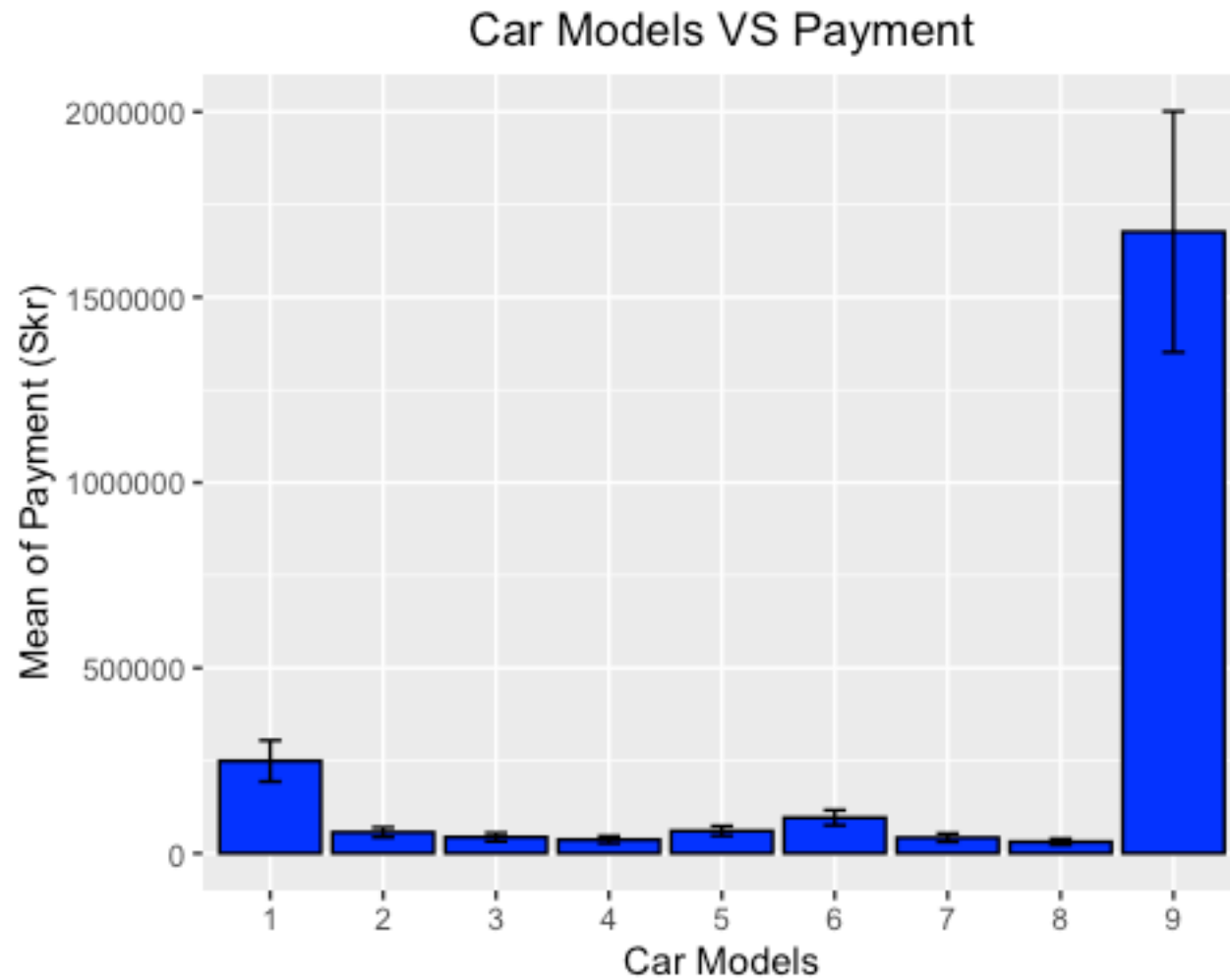


- U-shape distribution
- Mean claims of 6 years of bonus (7) particularly higher than the others (~800000 skr, nearly 500000 skr more than the next highest year)

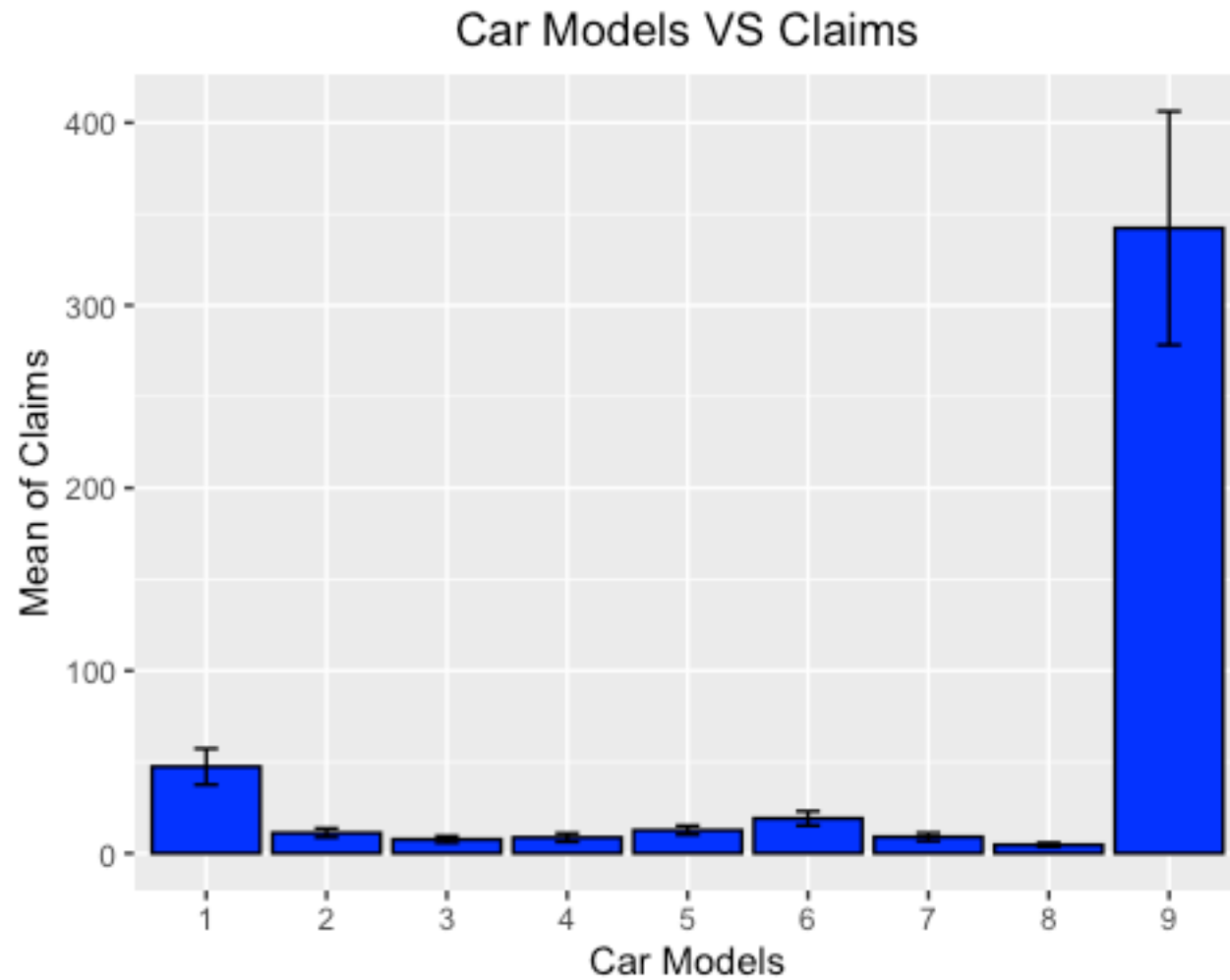
No Claims Bonus VS Claims



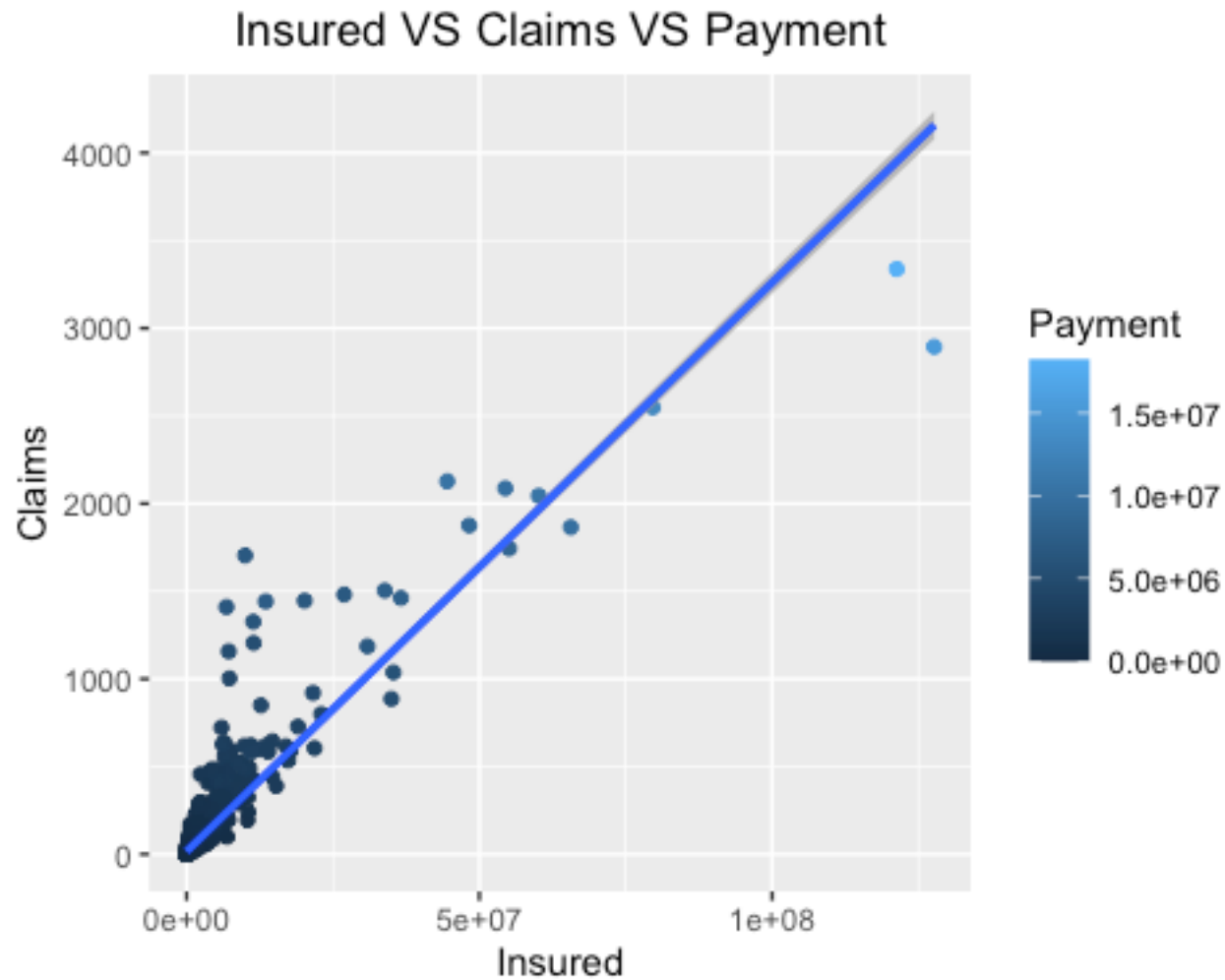
- U-shape distribution
- Mean claims of 6 years of bonus (7) particularly higher than the others (~150 cases, nearly 100 cases more than the next highest year)



- All classes are close in mean claims except for class9 (particularly high ~1700000 skr)



- All classes are close in mean claims except for class9 (particularly high ~350 cases)



Central tendency and dispersion measures

```
## Kilometres Zone Bonus Make Insured Claims
## 1:439 1:315 1:307 1 :245 Min. : 10 Min. : 0.00
## 2:441 2:315 2:312 2 :245 1st Qu.: 21610 1st Qu.: 1.00
## 3:441 3:315 3:310 9 :245 Median : 81525 Median : 5.00
## 4:434 4:315 4:310 5 :244 Mean : 1092195 Mean : 51.87
## 5:427 5:313 5:313 6 :244 3rd Qu.: 389782 3rd Qu.: 21.00
## 6:315 6:315 3 :242 Max. :127687270 Max. :3338.00
## 7:294 7:315 (Other):717
## Payment
## Min. : 0
## 1st Qu.: 2989
## Median : 27404
## Mean : 257008
## 3rd Qu.: 111954
## Max. :18245026
##
```

Insured amount:

- Range is $127687270 - 10 = 127687260$
- Interquartile range is $389782 - 21610 = 368172$

Claim amount:

- Range is $3338 - 0 = 3338$
- Interquartile range is $21 - 1 = 20$

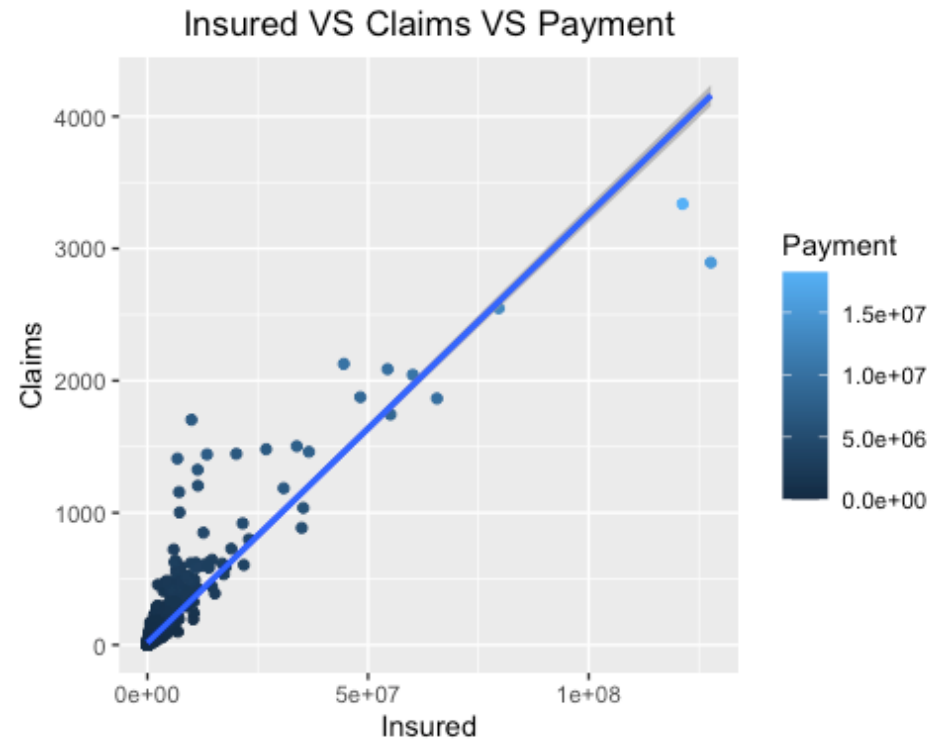
Payment:

- Range is $18245026 - 0 = 18245026$
- Interquartile range is $111954 - 2989 = 108965$

Question B

4 regular categorical variables + 3 continuous variables (cannot use regular categorical variables in correlation analysis)

Let's analyze again the last graph (scatter plot) in the answer to question A:



Outliers (No Pearson)

+ Our data set is not small (No Kendall)

+ Monotonic relationship (positive) among variables



Use Spearman method

```
##      Insured  Claims  Payment
## Insured 1.0000000 0.9333367 0.9030321
## Claims  0.9333367 1.0000000 0.9624433
## Payment 0.9030321 0.9624433 1.0000000
```

Large effect on “Payment”



Total payment is highly related to both the number of claims and the number of insured policy years

Question C

“Insured” and “Claims” have very strong correlation (as seen in the correlation analysis for question B) and that may bias our model (multicollinearity)

```
##      Insured  Claims  Payment
## Insured 1.0000000 0.9333367 0.9030321
## Claims 0.9333367 1.0000000 0.9624433
## Payment 0.9030321 0.9624433 1.0000000
```

Hierarchical method

We develop our first model with “Claims”. We assume “Claims” is highly important, so we use it as the first predictor:

```
##
## Call:
## lm(formula = Payment ~ Claims, data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1744858   -8545    2773   13386  1491369
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3362.29    2154.79  -1.56   0.119
## Claims      5020.08     10.35  485.11 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 97480 on 2180 degrees of freedom
## Multiple R-squared:  0.9908, Adjusted R-squared:  0.9908
## F-statistic: 2.353e+05 on 1 and 2180 DF, p-value: < 2.2e-16
```

Then we develop an advanced model by adding “Kilometres”, “Zone”, “Bonus”, and “Make” in one go:

```
##
## Call:
## lm(formula = Payment ~ Claims + Kilometres + Zone + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1689350 -21772   -190   22648 1355764
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -62206.38  10030.98  -6.201 6.69e-10 ***
## Claims       5059.74    12.41  407.870 < 2e-16 ***
## Kilometres2  11382.14    6282.18   1.812 0.070154 .
## Kilometres3  18546.92    6285.07   2.951 0.003202 **
## Kilometres4  23612.52    6343.76   3.722 0.000203 ***
## Kilometres5  22578.47    6376.93   3.541 0.000408 ***
## Zone2        11471.87    7422.07   1.546 0.122337
## Zone3        21010.68    7422.80   2.831 0.004690 **
## Zone4        58181.21    7429.68   7.831 7.53e-15 ***
## Zone5        30377.19    7465.22   4.069 4.89e-05 ***
## Zone6        44410.61    7439.10   5.970 2.77e-09 ***
## Zone7        33112.98    7618.77   4.346 1.45e-05 ***
## Bonus2       23223.57    7495.38   3.098 0.001971 **
## Bonus3       29502.51    7513.57   3.927 8.89e-05 ***
## Bonus4       28679.63    7517.12   3.815 0.000140 ***
## Bonus5       26319.68    7497.15   3.511 0.000456 ***
## Bonus6       28548.14    7475.35   3.819 0.000138 ***
## Bonus7       56743.88    7569.74   7.496 9.54e-14 ***
## Make2       -8928.58    8427.41  -1.059 0.289505
## Make3       -3955.68    8456.98  -0.468 0.640017
## Make4      -16004.86    8494.05  -1.884 0.059666 .
## Make5      -12543.28    8435.45  -1.487 0.137168
## Make6       -9520.15    8431.60  -1.129 0.258980
## Make7      -12209.93    8456.17  -1.444 0.148910
## Make8      -1369.76    8507.61  -0.161 0.872105
## Make9      -64246.26   9175.72  -7.002 3.36e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 93140 on 2156 degrees of freedom
## Multiple R-squared:  0.9917, Adjusted R-squared:  0.9916
## F-statistic: 1.032e+04 on 25 and 2156 DF, p-value: < 2.2e-16
```


We then use ANOVA table to compare both models:

```
## Analysis of Variance Table
##
## Model 1: Payment ~ Claims
## Model 2: Payment ~ Claims + Kilometres + Zone + Bonus + Make
##   Res.Df    RSS Df Sum of Sq   F    Pr(>F)
## 1   2180 2.0716e+13
## 2   2156 1.8704e+13 24 2.0118e+12 9.6622 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Both models are significantly better than the mean model (p-value < 0.05)
- The original model has a higher F-ratio with only 1 DF. However, the advanced model is more representative with 24 DF and a 0.0008 larger adjusted R-squared
- “Claims” has a particularly high t-value, which verifies the assumption we made in the beginning of Question C (“Claims” is highly important)
- The advanced model represents better of our data (p-value < 0.05)

Therefore, we will enter the testing section with our advanced model.

We found 45 *standardized residuals* with absolute value > 2.58 .

2.06% (more than 1% of our observation)  Remove poor residuals:

- Standardized residuals with absolute value > 3.29
- Cook's distance > 1.00
- Hat values of greater than twice the average hat value

We removed 1 poor residual.

DW test:

```
##  
## Durbin-Watson test  
##  
## data: PaymentModel2  
## DW = 1.9602, p-value = 0.109  
## alternative hypothesis: true autocorrelation is greater than 0
```

- Possible autocorrelation, though the effect could be very small

Variance Inflation Factor (VIF):

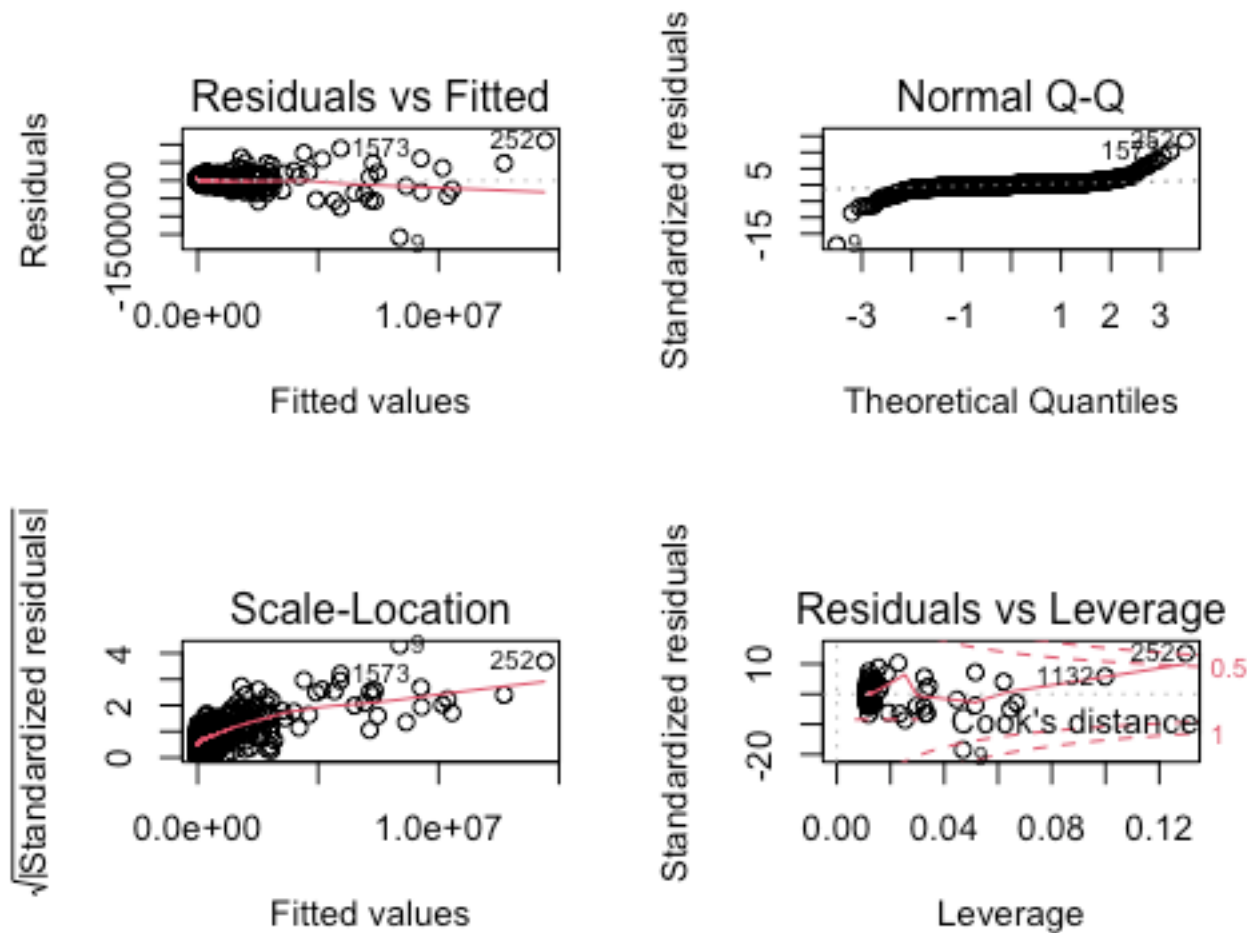
```
##          GVIF Df GVIF^(1/(2*Df))  
## Claims    1.608621 1    1.268314  
## Kilometres 1.036555 4    1.004498  
## Zone      1.046526 6    1.003797  
## Bonus     1.082664 6    1.006641  
## Make      1.457554 8    1.023827  
## [1] 2.435933
```

- No single predictor shows a strong linear relationship with other predictors (no VIF ≥ 10.00)
- May be one or more collinear explanatory variables (average VIF > 1.00)

Sample size:

- 2182 observation, far more than the recommended minimum ($50 + 5k$, where k is the number of predictors)
- Make our model more reliable

Linearity and homoscedasticity:



The top-left graph:

- Relationship between the fitted values and the standardized residuals
- Acceptable linear curve
- Data points are quite evenly dispersed around zero (homoscedasticity)

At last, we update the summary of our regression model:

```
##
## Call:
## lm(formula = Payment ~ Claims + Kilometres + Zone + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1581609 -21058      64   20886 1111063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -51782.40    9455.64  -5.476 4.85e-08 ***
## Claims       4980.51     12.59 395.695 < 2e-16 ***
## Kilometres2  8877.60     5910.96   1.502 0.133272
## Kilometres3 16878.54     5912.63   2.855 0.004350 **
## Kilometres4 19190.03     5972.81   3.213 0.001333 **
## Kilometres5 17839.41     6004.84   2.971 0.003003 **
## Zone2       11001.02     6981.32   1.576 0.115224
## Zone3       20196.75     6982.12   2.893 0.003859 **
## Zone4       55313.56     6990.51   7.913 3.99e-15 ***
## Zone5       26015.97     7026.66   3.702 0.000219 ***
## Zone6       41162.92     6999.96   5.880 4.73e-09 ***
## Zone7       27121.26     7175.17   3.780 0.000161 ***
## Bonus2      21099.83     7051.36   2.992 0.002800 **
## Bonus3      26593.48     7069.46   3.762 0.000173 ***
## Bonus4      25423.03     7073.33   3.594 0.000333 ***
## Bonus5      23326.20     7054.15   3.307 0.000959 ***
## Bonus6      26957.64     7032.02   3.834 0.000130 ***
## Bonus7      59379.76     7121.90   8.338 < 2e-16 ***
## Make2      -11798.66     7928.74  -1.488 0.136874
## Make3      -7153.68     7957.00  -0.899 0.368730
## Make4     -19238.69     7991.90  -2.407 0.016156 *
## Make5     -15327.06     7936.20  -1.931 0.053579 .
## Make6     -11780.78     7931.99  -1.485 0.137631
## Make7     -15326.80     7956.12  -1.926 0.054184 .
## Make8     -4942.63     8005.16  -0.617 0.537016
## Make9     -47401.86     8688.89  -5.455 5.45e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 87610 on 2155 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9913
## F-statistic: 9989 on 25 and 2155 DF, p-value: < 2.2e-16
```


Conclusion

In our survey of 2182 cases,
distance, location, bonus year, car model and claim amount all have **significant relationships** to insurance payment

Now we develop our second model with “Insured”.

We assume “Insured” is highly important, so we use it as the first predictor in our second model:

```
##
## Call:
## lm(formula = Payment ~ Insured, data = Insurance)
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -5946157 -75828  -70260  -30246  5343552
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.385e+04  7.971e+03   9.265  <2e-16 ***
## Insured      1.677e-01  1.383e-03 121.266  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 365600 on 2180 degrees of freedom
## Multiple R-squared:  0.8709, Adjusted R-squared:  0.8708
## F-statistic: 1.471e+04 on 1 and 2180 DF, p-value: < 2.2e-16
```

Then we develop an advanced model by adding “Kilometres”, “Zone”, “Bonus”, and “Make” in one go:

```
##
## Call:
## lm(formula = Payment ~ Insured + Kilometres + Zone + Bonus +
##   Make, data = Insurance)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -4705483 -76427  -4655   61437 4639327
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.923e+05  3.376e+04  8.657 < 2e-16 ***
## Insured      1.535e-01  1.373e-03 111.808 < 2e-16 ***
## Kilometres2  8.337e+04  2.129e+04  3.916 9.30e-05 ***
## Kilometres3  2.674e+04  2.132e+04  1.255 0.209764
## Kilometres4 -3.488e+04  2.148e+04 -1.624 0.104464
## Kilometres5 -3.463e+04  2.159e+04 -1.604 0.108809
## Zone2       -4.857e+04  2.517e+04 -1.930 0.053739 .
## Zone3       -8.112e+04  2.517e+04 -3.223 0.001288 **
## Zone4       -5.516e+04  2.527e+04 -2.183 0.029133 *
## Zone5       -1.467e+05  2.522e+04 -5.818 6.84e-09 ***
## Zone6       -1.272e+05  2.517e+04 -5.053 4.73e-07 ***
## Zone7       -1.864e+05  2.567e+04 -7.259 5.43e-13 ***
## Bonus2      -1.047e+05  2.539e+04 -4.125 3.85e-05 ***
## Bonus3      -1.386e+05  2.543e+04 -5.451 5.58e-08 ***
## Bonus4      -1.567e+05  2.543e+04 -6.163 8.51e-10 ***
## Bonus5      -1.563e+05  2.537e+04 -6.162 8.55e-10 ***
## Bonus6      -1.226e+05  2.534e+04 -4.840 1.39e-06 ***
## Bonus7      -8.646e+04  2.597e+04 -3.329 0.000886 ***
## Make2       -7.424e+04  2.855e+04 -2.600 0.009385 **
## Make3       -8.689e+04  2.865e+04 -3.033 0.002448 **
## Make4       -1.084e+05  2.877e+04 -3.769 0.000168 ***
## Make5       -7.293e+04  2.858e+04 -2.552 0.010790 *
## Make6       -8.356e+04  2.857e+04 -2.925 0.003483 **
## Make7       -8.926e+04  2.865e+04 -3.116 0.001857 **
## Make8       -8.780e+04  2.881e+04 -3.047 0.002337 **
## Make9       4.990e+05  2.972e+04 16.792 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 315800 on 2156 degrees of freedom
## Multiple R-squared:  0.9047, Adjusted R-squared:  0.9036
## F-statistic: 818.9 on 25 and 2156 DF, p-value: < 2.2e-16
```


We then use ANOVA table to compare both models:

```
## Analysis of Variance Table
##
## Model 1: Payment ~ Insured
## Model 2: Payment ~ Insured + Kilometres + Zone + Bonus + Make
##   Res.Df    RSS Df Sum of Sq   F    Pr(>F)
## 1   2180 2.9140e+14
## 2   2156 2.1504e+14 24 7.6353e+13 31.896 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Both models are significantly better than the mean model (p-value < 0.05)
- Original model has a higher F-ratio with only 1 DF. However, the advanced model is more representative with 24 DF and a 0.0328 larger adjusted R-squared
- “Insured” has a particularly high t-value, which verifies the assumption we made in the beginning of Question C (“Insured” is highly important)
- The advanced model represents better for our data (p-value < 0.05)

Therefore, we will enter the testing section with our advanced model.

We found 28 *standardized residuals* with absolute value > 2.58 .

1.28% (more than 1% of our observation)  Remove poor residuals:

- Standardized residuals with absolute value > 3.29
- Cook's distance > 1.00
- Hat values of greater than twice the average hat value

We removed 1 poor residual.

DW test:

```
##  
## Durbin-Watson test  
##  
## data: PaymentModel4  
## DW = 1.9655, p-value = 0.1338  
## alternative hypothesis: true autocorrelation is greater than 0
```

- Possible autocorrelation, though the effect could be very small.

Variance Inflation Factor (VIF):

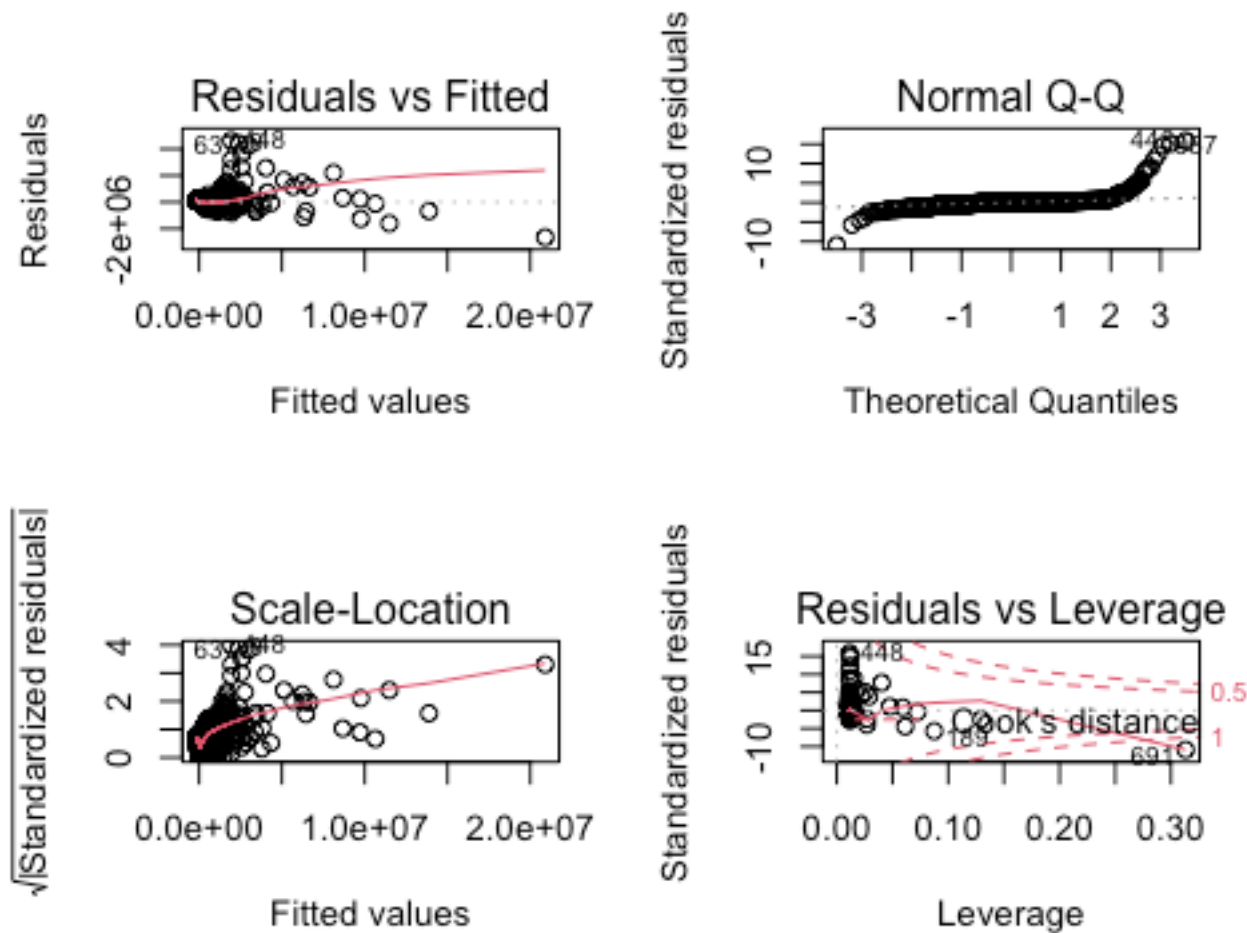
```
##          GVIF Df GVIF^(1/(2*Df))  
## Insured  1.359060 1    1.165787  
## Kilometres 1.023661 4    1.002927  
## Zone     1.029572 6    1.002432  
## Bonus    1.099418 6    1.007930  
## Make     1.218222 8    1.012413  
## [1] 2.394761
```

- No single predictor shows a strong linear relationship with other predictors (no VIF ≥ 10.00)
- May be one or more collinear explanatory variables (average VIF > 1.00)

Sample size:

- 2182 observation, far more than the recommended minimum ($50 + 5k$, where k is the number of predictors)
- Make our model more reliable

Linearity and homoscedasticity:



The top-left graph:

- Relationship between the fitted values and the standardized residuals
- Acceptable linear curve
- Data points are quite unevenly dispersed around zero (heteroscedasticity)

At last, we update the summary of our regression model:

```
##
## Call:
## lm(formula = Payment ~ Insured + Kilometres + Zone + Bonus +
##   Make, data = Insurance)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -2652372 -70012  -1935   60585  4627207
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.827e+05  3.134e+04  9.019  < 2e-16 ***
## Insured      1.672e-01  1.472e-03 113.597  < 2e-16 ***
## Kilometres2  6.895e+04  1.978e+04   3.486  0.000499 ***
## Kilometres3  2.254e+04  1.978e+04   1.139  0.254703
## Kilometres4 -2.922e+04  1.993e+04  -1.466  0.142848
## Kilometres5 -2.697e+04  2.004e+04  -1.346  0.178505
## Zone2       -5.126e+04  2.336e+04  -2.195  0.028288 *
## Zone3       -8.562e+04  2.336e+04  -3.665  0.000253 ***
## Zone4       -5.775e+04  2.345e+04  -2.463  0.013866 *
## Zone5       -1.376e+05  2.341e+04  -5.878  4.80e-09 ***
## Zone6       -1.239e+05  2.336e+04  -5.306  1.23e-07 ***
## Zone7       -1.713e+05  2.384e+04  -7.187  9.11e-13 ***
## Bonus2      -1.042e+05  2.356e+04  -4.420  1.04e-05 ***
## Bonus3      -1.371e+05  2.360e+04  -5.810  7.16e-09 ***
## Bonus4      -1.547e+05  2.360e+04  -6.556  6.88e-11 ***
## Bonus5      -1.557e+05  2.355e+04  -6.613  4.74e-11 ***
## Bonus6      -1.273e+05  2.352e+04  -5.411  6.95e-08 ***
## Bonus7      -1.234e+05  2.418e+04  -5.104  3.62e-07 ***
## Make2       -6.365e+04  2.651e+04  -2.402  0.016412 *
## Make3       -7.589e+04  2.659e+04  -2.854  0.004358 **
## Make4       -9.807e+04  2.670e+04  -3.673  0.000246 ***
## Make5       -6.238e+04  2.653e+04  -2.351  0.018807 *
## Make6       -7.724e+04  2.652e+04  -2.913  0.003621 **
## Make7       -7.826e+04  2.659e+04  -2.943  0.003283 **
## Make8       -7.487e+04  2.675e+04  -2.799  0.005168 **
## Make9        4.417e+05  2.775e+04  15.915  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 293100 on 2155 degrees of freedom
## Multiple R-squared:  0.9085, Adjusted R-squared:  0.9074
## F-statistic: 855.9 on 25 and 2155 DF, p-value: < 2.2e-16
```

Conclusion

In our survey of 2182 cases,
distance, location, bonus year, car model and insured amount all have **significant relationships** to insurance payment
However, compared to the first model, where adjusted R-squared is 0.9913, this model is less representative to the our data
Therefore we prefer the first model

Question D

This time, I will use *stepwise regression modeling* in *both directions*:

```
## Start: AIC=23160.03
## Claims ~ 1
##
##      Df Sum of Sq  RSS  AIC
## + Insured   1  73540770 15198022 19312
## + Make      8  23594134 65144658 22502
## + Bonus     6  4469115 84269677 23059
## + Zone      6  2220038 86518754 23117
## + Kilometres 4  1774202 86964590 23124
## <none>             88738792 23160
##
## Step: AIC=19311.82
## Claims ~ Insured
##
##      Df Sum of Sq  RSS  AIC
## + Make      8  3126865 12071157 18825
## + Zone      6   359554 14838468 19272
## + Bonus     6   335468 14862553 19275
## + Kilometres 4   143786 15054235 19299
## <none>             15198022 19312
## - Insured   1  73540770 88738792 23160
##
## Step: AIC=18825.2
## Claims ~ Insured + Make
##
##      Df Sum of Sq  RSS  AIC
## + Zone      6  424979 11646178 18759
## + Bonus     6   302253 11768904 18782
## + Kilometres 4   210824 11860333 18795
## <none>             12071157 18825
## - Make      8  3126865 15198022 19312
## - Insured   1  53073501 65144658 22502
##
## Step: AIC=18759
## Claims ~ Insured + Make + Zone
##
##      Df Sum of Sq  RSS  AIC
## + Bonus     6  297990 11348188 18714
## + Kilometres 4  224980 11421198 18724
## <none>             11646178 18759
## - Zone      6  424979 12071157 18825
## - Make      8  3192290 14838468 19272
## - Insured   1  51178355 62824533 22434
```

```
##
## Step: AIC=18714.44
## Claims ~ Insured + Make + Zone + Bonus
##
##      Df Sum of Sq  RSS  AIC
## + Kilometres 4   224352 11123836 18679
## <none>                11348188 18714
## - Bonus      6   297990 11646178 18759
## - Zone       6   420715 11768904 18782
## - Make       8   3162689 14510877 19235
## - Insured    1  46946083 58294272 22283
##
## Step: AIC=18678.87
## Claims ~ Insured + Make + Zone + Bonus + Kilometres
##
##      Df Sum of Sq  RSS  AIC
## <none>                11123836 18679
## - Kilometres 4   224352 11348188 18714
## - Bonus      6   297362 11421198 18724
## - Zone       6   435392 11559228 18751
## - Make       8   3241285 14365121 19221
## - Insured    1  45249641 56373477 22218
##
## Call:
## lm(formula = Claims ~ Insured + Make + Zone + Bonus + Kilometres,
##     data = Insurance)
##
## Coefficients:
## (Intercept)  Insured   Make2    Make3    Make4    Make5
##  7.130e+01  2.924e-05 -1.375e+01 -1.727e+01 -1.911e+01 -1.278e+01
##   Make6    Make7    Make8    Make9    Zone2    Zone3
## -1.514e+01 -1.611e+01 -1.813e+01  1.180e+02 -1.165e+01 -1.983e+01
##   Zone4    Zone5    Zone6    Zone7    Bonus2    Bonus3
## -2.059e+01 -3.574e+01 -3.416e+01 -4.461e+01 -2.533e+01 -3.334e+01
##   Bonus4    Bonus5    Bonus6    Bonus7 Kilometres2 Kilometres3
## -3.679e+01 -3.614e+01 -2.950e+01 -2.374e+01  1.423e+01  8.060e-01
## Kilometres4 Kilometres5
## -1.317e+01 -1.309e+01
```


The model suggests a **formula = Claims ~ Insured + Make + Zone + Bonus + Kilometres**

Then we take a look at the summary of our model:

```
##
## Call:
## lm(formula = Claims ~ Insured + Zone + Kilometres + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -983.95 -16.36   0.06  14.09 1222.44
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.130e+01  7.679e+00   9.284 < 2e-16 ***
## Insured      2.924e-05  3.122e-07  93.649 < 2e-16 ***
## Zone2       -1.165e+01  5.724e+00  -2.036 0.041887 *
## Zone3       -1.983e+01  5.724e+00  -3.464 0.000543 ***
## Zone4       -2.059e+01  5.747e+00  -3.583 0.000347 ***
## Zone5       -3.574e+01  5.737e+00  -6.230 5.60e-10 ***
## Zone6       -3.416e+01  5.724e+00  -5.969 2.79e-09 ***
## Zone7       -4.461e+01  5.839e+00  -7.641 3.23e-14 ***
## Kilometres2  1.423e+01  4.843e+00   2.938 0.003341 **
## Kilometres3  8.060e-01  4.848e+00   0.166 0.867982
## Kilometres4 -1.317e+01  4.884e+00  -2.697 0.007057 **
## Kilometres5 -1.309e+01  4.910e+00  -2.666 0.007737 **
## Bonus2      -2.533e+01  5.775e+00  -4.385 1.21e-05 ***
## Bonus3      -3.334e+01  5.784e+00  -5.765 9.35e-09 ***
## Bonus4      -3.679e+01  5.784e+00  -6.361 2.44e-10 ***
## Bonus5      -3.614e+01  5.771e+00  -6.263 4.55e-10 ***
## Bonus6      -2.950e+01  5.763e+00  -5.119 3.35e-07 ***
## Bonus7      -2.374e+01  5.907e+00  -4.019 6.03e-05 ***
## Make2       -1.375e+01  6.494e+00  -2.117 0.034346 *
## Make3       -1.727e+01  6.515e+00  -2.651 0.008088 **
## Make4       -1.911e+01  6.543e+00  -2.921 0.003523 **
## Make5       -1.278e+01  6.501e+00  -1.966 0.049478 *
## Make6       -1.514e+01  6.498e+00  -2.330 0.019899 *
## Make7       -1.611e+01  6.515e+00  -2.473 0.013469 *
## Make8       -1.813e+01  6.553e+00  -2.767 0.005712 **
## Make9        1.180e+02  6.759e+00  17.451 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 71.83 on 2156 degrees of freedom
## Multiple R-squared:  0.8746, Adjusted R-squared:  0.8732
## F-statistic: 601.7 on 25 and 2156 DF, p-value: < 2.2e-16
```

Now we will enter the testing section.

We found 29 *standardized residuals* with absolute value > 2.58 .

1.33% (more than 1% of our observation)  Remove poor residuals:

- Standardized residuals with absolute value > 3.29
- Cook's distance > 1.00
- Hat values of greater than twice the average hat value

We removed 1 poor residual.

DW test:

```
##  
## Durbin-Watson test  
##  
## data: ClaimsModel  
## DW = 1.9951, p-value = 0.3379  
## alternative hypothesis: true autocorrelation is greater than 0
```

- A DW value of 2.00 indicates no autocorrelation in our model

Variance Inflation Factor (VIF):

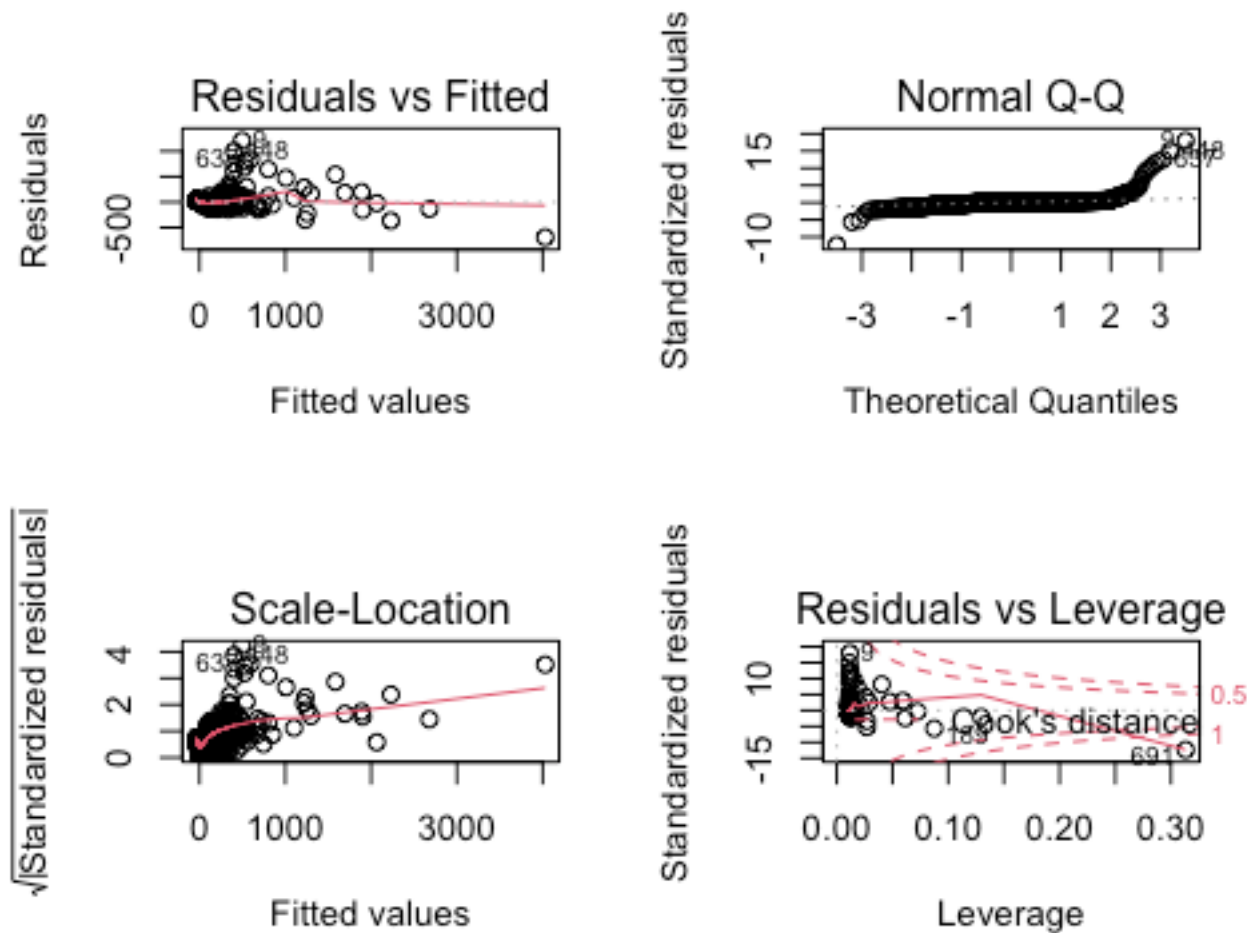
```
##          GVIF Df GVIF^(1/(2*Df))  
## Insured  1.359060 1    1.165787  
## Zone     1.029572 6    1.002432  
## Kilometres 1.023661 4    1.002927  
## Bonus    1.099418 6    1.007930  
## Make     1.218222 8    1.012413  
## [1] 2.394761
```

- No single predictor shows a strong linear relationship with other predictors (no VIF ≥ 10.00)
- May be one or more collinear explanatory variables (average VIF > 1.00)

Sample size:

- 2182 observation, far more than the recommended minimum ($50 + 5k$, where k is the number of predictors)
- Make our model more reliable

Linearity and homoscedasticity:



The top-left graph:

- Relationship between the fitted values and the standardized residuals
- Acceptable linear curve
- Data points are quite unevenly dispersed around zero (heteroscedasticity)

At last, we update the summary of our regression model again:

```
##
## Call:
## lm(formula = Claims ~ Insured + Zone + Kilometres + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -691.29  -15.69   0.90   13.75 1207.65
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.928e+01  7.216e+00  9.601 < 2e-16 ***
## Insured      3.212e-05  3.390e-07  94.736 < 2e-16 ***
## Zone2       -1.222e+01  5.378e+00  -2.271 0.023226 *
## Zone3       -2.077e+01  5.379e+00  -3.861 0.000116 ***
## Zone4       -2.113e+01  5.400e+00  -3.913 9.38e-05 ***
## Zone5       -3.383e+01  5.391e+00  -6.275 4.22e-10 ***
## Zone6       -3.349e+01  5.378e+00  -6.227 5.69e-10 ***
## Zone7       -4.147e+01  5.489e+00  -7.555 6.16e-14 ***
## Kilometres2  1.121e+01  4.554e+00  2.462 0.013898 *
## Kilometres3 -7.296e-02  4.556e+00  -0.016 0.987224
## Kilometres4 -1.199e+01  4.590e+00  -2.612 0.009071 **
## Kilometres5 -1.149e+01  4.614e+00  -2.489 0.012871 *
## Bonus2      -2.521e+01  5.426e+00  -4.645 3.60e-06 ***
## Bonus3      -3.303e+01  5.435e+00  -6.078 1.43e-09 ***
## Bonus4      -3.638e+01  5.435e+00  -6.694 2.76e-11 ***
## Bonus5      -3.601e+01  5.422e+00  -6.641 3.93e-11 ***
## Bonus6      -3.046e+01  5.415e+00  -5.626 2.09e-08 ***
## Bonus7      -3.147e+01  5.569e+00  -5.652 1.80e-08 ***
## Make2       -1.154e+01  6.103e+00  -1.890 0.058866 .
## Make3       -1.497e+01  6.123e+00  -2.445 0.014566 *
## Make4       -1.695e+01  6.149e+00  -2.756 0.005900 **
## Make5       -1.057e+01  6.110e+00  -1.730 0.083721 .
## Make6       -1.382e+01  6.106e+00  -2.263 0.023744 *
## Make7       -1.381e+01  6.123e+00  -2.256 0.024181 *
## Make8       -1.543e+01  6.159e+00  -2.505 0.012327 *
## Make9       1.060e+02  6.390e+00  16.582 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.49 on 2155 degrees of freedom
## Multiple R-squared:  0.8783, Adjusted R-squared:  0.8769
## F-statistic: 622.1 on 25 and 2155 DF, p-value: < 2.2e-16
```

Conclusion

In our survey of 2182 cases,
distance, location, bonus year, car model and insured amount all have **significant relationships** to claim amount

In respond to what extent the predictors affect claims number, by holding all other predictors (independent variables) constant, for every 1 unit increase in:

Insured, Claims increases by 69.28 cases
Zone2, Claims decreases by 12.22 cases
Zone3, Claims decreases by 20.77 cases
Zone4, Claims decreases by 21.13 cases
Zone5, Claims decreases by 33.83 cases
Zone6, Claims decreases by 33.49 cases
Zone7, Claims decreases by 41.47 cases
Kilometre2, Claims decreases by 11.21 cases
Kilometre4, Claims decreases by 11.99 cases
Kilometre5, Claims decreases by 11.49 cases
Bonus2, Claims decreases by 25.21 cases
Bonus3, Claims decreases by 33.03 cases
Bonus4, Claims decreases by 36.38 cases
Bonus5, Claims decreases by 36.01 cases
Bonus6, Claims decreases by 30.46 cases
Bonus7, Claims decreases by 31.47 cases
Make3, Claims decreases by 14.97 cases
Make4, Claims decreases by 16.95 cases
Make6, Claims decreases by 13.82 cases
Make7, Claims decreases by 13.81 cases
Make8, Claims decreases by 15.43 cases
Make9, Claims increases by 106.00 cases

Question E(i)

The number of claims predicted is 147 cases:

1

147.0687

The payment predicted is 735028 skr.

1

735028

Question E(ii)

The number of claims predicted is 471 cases.

1

470.7816

The payment predicted is 2274621 skr.

1

2274621

Question E(iii)

The number of claims predicted is 556 to 810 cases.

```
##      1
## 556.1285
##      1
## 810.3588
```

Number of payment predicted is 2764384 skr to 4030580 skr.

```
##      1
## 2764384
##      1
## 4030580
```

Question F

Recalling the above model developed in answer C, if all other predictors (independent variables) are held constant, each unit increase of the following variables has payment increases for (bolded = largest):

```
##
## Call:
## lm(formula = Payment ~ Claims + Kilometres + Zone + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1581609 -21058      64    20886 1111063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -51782.40   9455.64  -5.476 4.85e-08 ***
## Claims       4980.51    12.59 395.695 < 2e-16 ***
## Kilometres2  8877.60    5910.96   1.502 0.133272
## Kilometres3 16878.54    5912.63   2.855 0.004350 **
## Kilometres4 19190.03    5972.81   3.213 0.001333 **
## Kilometres5 17839.41    6004.84   2.971 0.003003 **
## Zone2        11001.02    6981.32   1.576 0.115224
## Zone3        20196.75    6982.12   2.893 0.003859 **
## Zone4        55313.56    6990.51   7.913 3.99e-15 ***
## Zone5        26015.97    7026.66   3.702 0.000219 ***
## Zone6        41162.92    6999.96   5.880 4.73e-09 ***
## Zone7        27121.26    7175.17   3.780 0.000161 ***
## Bonus2       21099.83    7051.36   2.992 0.002800 **
## Bonus3       26593.48    7069.46   3.762 0.000173 ***
## Bonus4       25423.03    7073.33   3.594 0.000333 ***
## Bonus5       23326.20    7054.15   3.307 0.000959 ***
## Bonus6       26957.64    7032.02   3.834 0.000130 ***
## Bonus7       59379.76    7121.90   8.338 < 2e-16 ***
## Make2       -11798.66    7928.74  -1.488 0.136874
## Make3       -7153.68    7957.00  -0.899 0.368730
## Make4       -19238.69    7991.90  -2.407 0.016156 *
## Make5       -15327.06    7936.20  -1.931 0.053579 .
## Make6       -11780.78    7931.99  -1.485 0.137631
## Make7       -15326.80    7956.12  -1.926 0.054184 .
## Make8       -4942.63    8005.16  -0.617 0.537016
## Make9       -47401.86    8688.89  -5.455 5.45e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 87610 on 2155 degrees of freedom
## Multiple R-squared:  0.9914, Adjusted R-squared:  0.9913
## F-statistic: 9989 on 25 and 2155 DF, p-value: < 2.2e-16
```

Recalling the above model developed in answer D, if all other predictors (independent variables) are held constant, each unit increase of the following variables has claim amount increases for (bolded = largest):

```
##
## Call:
## lm(formula = Claims ~ Insured + Zone + Kilometres + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -691.29  -15.69   0.90   13.75 1207.65
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.928e+01  7.216e+00  9.601 < 2e-16 ***
## Insured      3.212e-05  3.390e-07  94.736 < 2e-16 ***
## Zone2       -1.222e+01  5.378e+00  -2.271 0.023226 *
## Zone3       -2.077e+01  5.379e+00  -3.861 0.000116 ***
## Zone4       -2.113e+01  5.400e+00  -3.913 9.38e-05 ***
## Zone5       -3.383e+01  5.391e+00  -6.275 4.22e-10 ***
## Zone6       -3.349e+01  5.378e+00  -6.227 5.69e-10 ***
## Zone7       -4.147e+01  5.489e+00  -7.555 6.16e-14 ***
## Kilometres2  1.121e+01  4.554e+00  2.462 0.013898 *
## Kilometres3 -7.296e-02  4.556e+00  -0.016 0.987224
## Kilometres4 -1.199e+01  4.590e+00  -2.612 0.009071 **
## Kilometres5 -1.149e+01  4.614e+00  -2.489 0.012871 *
## Bonus2      -2.521e+01  5.426e+00  -4.645 3.60e-06 ***
## Bonus3      -3.303e+01  5.435e+00  -6.078 1.43e-09 ***
## Bonus4      -3.638e+01  5.435e+00  -6.694 2.76e-11 ***
## Bonus5      -3.601e+01  5.422e+00  -6.641 3.93e-11 ***
## Bonus6      -3.046e+01  5.415e+00  -5.626 2.09e-08 ***
## Bonus7      -3.147e+01  5.569e+00  -5.652 1.80e-08 ***
## Make2       -1.154e+01  6.103e+00  -1.890 0.058866 .
## Make3       -1.497e+01  6.123e+00  -2.445 0.014566 *
## Make4       -1.695e+01  6.149e+00  -2.756 0.005900 **
## Make5       -1.057e+01  6.110e+00  -1.730 0.083721 .
## Make6       -1.382e+01  6.106e+00  -2.263 0.023744 *
## Make7       -1.381e+01  6.123e+00  -2.256 0.024181 *
## Make8       -1.543e+01  6.159e+00  -2.505 0.012327 *
## Make9       1.060e+02  6.390e+00  16.582 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.49 on 2155 degrees of freedom
## Multiple R-squared:  0.8783, Adjusted R-squared:  0.8769
## F-statistic: 622.1 on 25 and 2155 DF, p-value: < 2.2e-16
```

For insured amount we do not have any developed model that targets it yet, so we will develop one below:

```
## Start: AIC=67857.44
## Insured ~ 1
##
##      Df Sum of Sq  RSS  AIC
## + Claims    1 5.7927e+16 1.1971e+16 64009
## + Make      8 9.8600e+15 6.0038e+16 67542
## + Bonus     6 4.6236e+15 6.5275e+16 67720
## + Zone      6 1.3270e+15 6.8571e+16 67828
## + Kilometres 4 9.6712e+14 6.8931e+16 67835
## <none>          6.9898e+16 67857
##
## Step: AIC=64009.23
## Insured ~ Claims
##
##      Df Sum of Sq  RSS  AIC
## + Make      8 8.4632e+14 1.1125e+16 63865
## + Bonus     6 4.5884e+14 1.1512e+16 63936
## + Zone      6 2.1090e+14 1.1760e+16 63982
## <none>          1.1971e+16 64009
## + Kilometres 4 3.8754e+13 1.1932e+16 64010
## - Claims    1 5.7927e+16 6.9898e+16 67857
##
## Step: AIC=63865.25
## Insured ~ Claims + Make
##
##      Df Sum of Sq  RSS  AIC
## + Bonus     6 3.7655e+14 1.0748e+16 63802
## + Zone      6 2.4903e+14 1.0876e+16 63828
## + Kilometres 4 6.4214e+13 1.1061e+16 63861
## <none>          1.1125e+16 63865
## - Make      8 8.4632e+14 1.1971e+16 64009
## - Claims    1 4.8913e+16 6.0038e+16 67542
##
## Step: AIC=63802.12
## Insured ~ Claims + Make + Bonus
##
##      Df Sum of Sq  RSS  AIC
## + Zone      6 2.3871e+14 1.0510e+16 63765
## + Kilometres 4 5.8343e+13 1.0690e+16 63798
## <none>          1.0748e+16 63802
## - Bonus     6 3.7655e+14 1.1125e+16 63865
## - Make      8 7.6402e+14 1.1512e+16 63936
## - Claims    1 4.4640e+16 5.5388e+16 67378
```

```
##
## Step: AIC=63765.11
## Insured ~ Claims + Make + Bonus + Zone
##
##      Df Sum of Sq    RSS   AIC
## + Kilometres 4 6.3284e+13 1.0446e+16 63760
## <none>                1.0510e+16 63765
## - Zone      6 2.3871e+14 1.0748e+16 63802
## - Bonus     6 3.6623e+14 1.0876e+16 63828
## - Make      8 7.9855e+14 1.1308e+16 63909
## - Claims    1 4.3477e+16 5.3987e+16 67334
##
## Step: AIC=63759.93
## Insured ~ Claims + Make + Bonus + Zone + Kilometres
##
##      Df Sum of Sq    RSS   AIC
## <none>                1.0446e+16 63760
## - Kilometres 4 6.3284e+13 1.0510e+16 63765
## - Zone      6 2.4365e+14 1.0690e+16 63798
## - Bonus     6 3.5976e+14 1.0806e+16 63822
## - Make      8 8.2530e+14 1.1272e+16 63910
## - Claims    1 4.2494e+16 5.2940e+16 67299
##
## Call:
## lm(formula = Insured ~ Claims + Make + Bonus + Zone + Kilometres,
##     data = Insurance)
##
## Coefficients:
## (Intercept)  Claims    Make2    Make3    Make4    Make5
## -1735451    27455    225840    315802    372456    199101
##   Make6    Make7    Make8    Make9    Bonus2    Bonus3
## 324465    283138    308991   -2044802    688625    894692
##   Bonus4    Bonus5    Bonus6    Bonus7    Zone2    Zone3
## 983131    985190    878785    1473402    358473    608836
##   Zone4    Zone5    Zone6    Zone7 Kilometres2 Kilometres3
## 891521    849612    891918    1001851   -390831   -168778
## Kilometres4 Kilometres5
## 71465    38389
```

The model suggests a **formula = Insured ~ Claims + Make + Bonus + Zone + Kilometres**

Then we take a look at the summary of our model:


```
##
## Call:
## lm(formula = Insured ~ Claims + Zone + Kilometres + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33005096  -369218   -37211   436161  49646999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1735450.8   237059.9  -7.321 3.46e-13 ***
## Claims       27455.3     293.2   93.649 < 2e-16 ***
## Zone2       358472.9   175404.2    2.044 0.041105 *
## Zone3       608835.6   175421.4    3.471 0.000529 ***
## Zone4       891520.6   175584.0    5.077 4.15e-07 ***
## Zone5       849611.5   176423.8    4.816 1.57e-06 ***
## Zone6       891918.1   175806.7    5.073 4.24e-07 ***
## Zone7      1001851.3   180052.7    5.564 2.96e-08 ***
## Kilometres2 -390830.6   148465.3   -2.632 0.008537 **
## Kilometres3 -168777.9   148533.7   -1.136 0.255960
## Kilometres4  71465.5   149920.6    0.477 0.633632
## Kilometres5  38389.1   150704.6    0.255 0.798955
## Bonus2      688624.7   177136.6    3.888 0.000104 ***
## Bonus3      894692.4   177566.6    5.039 5.08e-07 ***
## Bonus4      983130.8   177650.4    5.534 3.51e-08 ***
## Bonus5      985190.1   177178.6    5.560 3.02e-08 ***
## Bonus6      878785.0   176663.3    4.974 7.06e-07 ***
## Bonus7     1473402.4   178894.0    8.236 3.05e-16 ***
## Make2       225839.8   199163.2    1.134 0.256944
## Make3       315801.7   199862.0    1.580 0.114231
## Make4       372455.9   200738.0    1.855 0.063671 .
## Make5       199100.6   199353.2    0.999 0.318036
## Make6       324465.2   199262.2    1.628 0.103600
## Make7       283137.5   199842.8    1.417 0.156686
## Make8       308990.9   201058.4    1.537 0.124484
## Make9      -2044802.0   216847.9   -9.430 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2201000 on 2156 degrees of freedom
```



```
## Multiple R-squared:  0.8505, Adjusted R-squared:  0.8488
## F-statistic: 490.8 on 25 and 2156 DF,  p-value: < 2.2e-16
```

Now we will enter the testing section.

We found 27 *standardized residuals* with absolute value > 2.58 .

1.23% (more than 1% of our observation)  Remove poor residuals:

- Standardized residuals with absolute value > 3.29
- Cook's distance > 1.00
- Hat values of greater than twice the average hat value

We removed 2 poor residuals.

DW test:

```
##  
## Durbin-Watson test  
##  
## data: InsuredModel  
## DW = 1.9747, p-value = 0.1854  
## alternative hypothesis: true autocorrelation is greater than 0
```

- No autocorrelation in our model

Variance Inflation Factor (VIF):

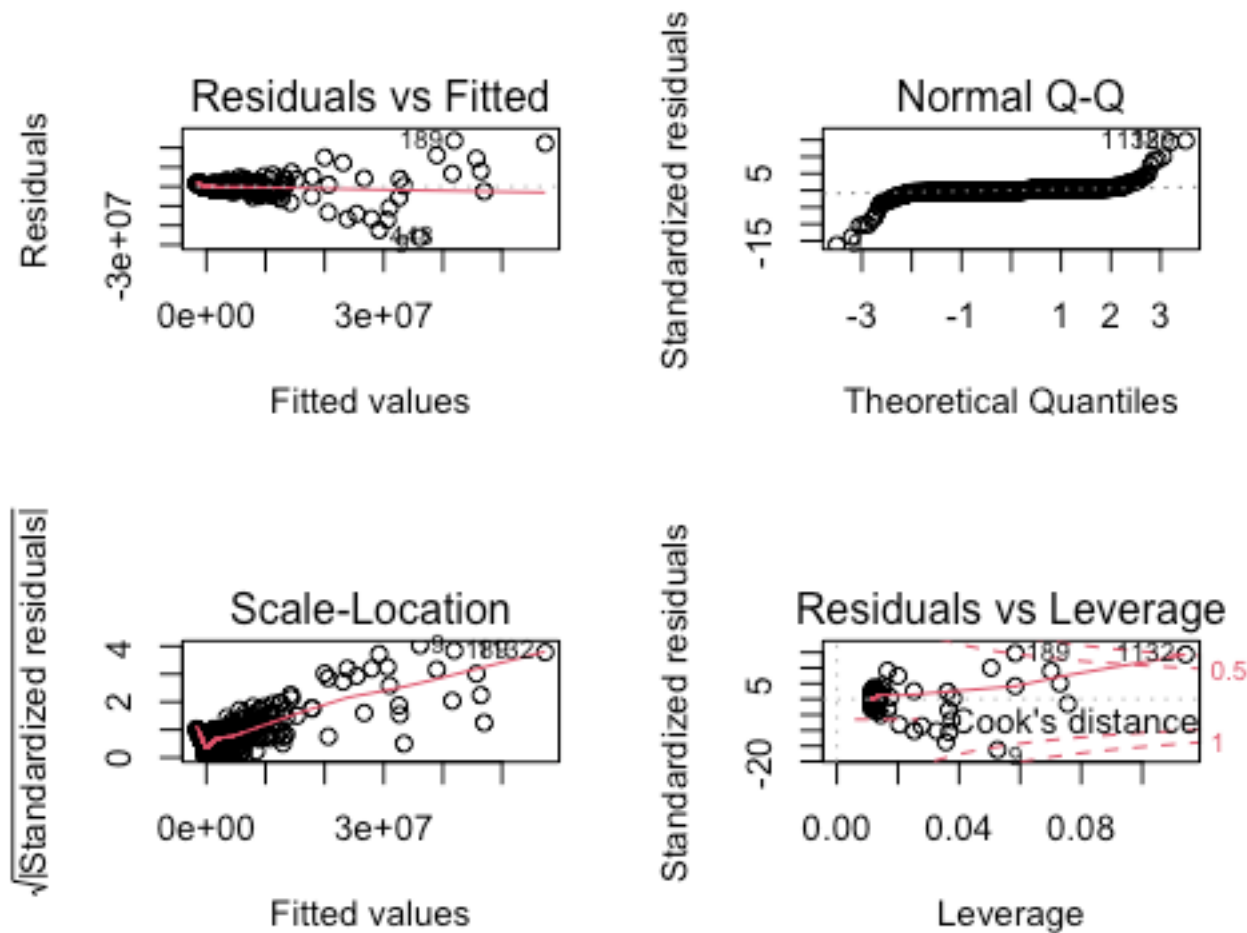
```
##          GVIF Df GVIF^(1/(2*Df))  
## Claims    1.637807 1    1.279768  
## Zone      1.047962 6    1.003912  
## Kilometres 1.038569 4    1.004742  
## Bonus     1.082684 6    1.006642  
## Make      1.486206 8    1.025073  
## [1] 2.440891
```

- No single predictor shows a strong linear relationship with other predictors (no VIF ≥ 10.00)
- May be one or more collinear explanatories (average VIF > 1.00)

Sample size:

- 2182 observation, far more than the recommended minimum ($50 + 5k$, where k is the number of predictors)
- Make our model more reliable

Lastly, we check for *linearity and homoscedasticity*:



The top-left graph:

- Relationship between the fitted values and the standardized residuals
- Acceptable linear curve
- Data points are quite evenly dispersed around zero (homoscedasticity)

At last, we update the summary of our regression model:

```
##
## Call:
## lm(formula = Insured ~ Claims + Zone + Kilometres + Bonus + Make,
##     data = Insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26139005 -239814  -33560   348589 23687512
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1216744.1   178623.6  -6.812 1.25e-11 ***
## Claims       22529.9     253.2   88.985 < 2e-16 ***
## Zone2        329202.3  131755.4    2.499  0.0125 *
## Zone3        558237.4  131772.5    4.236 2.37e-05 ***
## Zone4        682811.9  131989.5    5.173 2.51e-07 ***
## Zone5        578585.0  132697.5    4.360 1.36e-05 ***
## Zone6        690025.9  132154.9    5.221 1.95e-07 ***
## Zone7        629361.8  135574.7    4.642 3.66e-06 ***
## Kilometres2 -284463.5   111651.6   -2.548  0.0109 *
## Kilometres3 -130296.2   111607.0   -1.167  0.2432
## Kilometres4 -61048.5    112724.4   -0.542  0.5882
## Kilometres5 -113589.9   113332.0   -1.002  0.3163
## Bonus2       556848.1  133096.6    4.184 2.98e-05 ***
## Bonus3       714186.9  133456.2    5.351 9.65e-08 ***
## Bonus4       780557.6  133539.4    5.845 5.83e-09 ***
## Bonus5       799305.1  133169.7    6.002 2.28e-09 ***
## Bonus6       780028.6  132722.7    5.877 4.83e-09 ***
## Bonus7       1606935.4  134423.0   11.954 < 2e-16 ***
## Make2        47421.8   149668.1    0.317  0.7514
## Make3        116574.0   150209.6    0.776  0.4378
## Make4        171820.5   150868.5    1.139  0.2549
## Make5        26102.0   149806.7    0.174  0.8617
## Make6        183988.6   149716.5    1.229  0.2192
## Make7        89543.1   150190.5    0.596  0.5511
## Make8        86733.0   151128.4    0.574  0.5661
## Make9       -1036814.3   164913.3   -6.287 3.91e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1653000 on 2154 degrees of freedom
## Multiple R-squared:  0.8505, Adjusted R-squared:  0.8488
## F-statistic: 490.3 on 25 and 2154 DF, p-value: < 2.2e-16
```