LRM Practical End-semester Exam

Pranav Gopalkrishna, 1940223

2021-12-01

# DATA SET

This data set contains several fields of meteorological data from five different locations in Australia, measured across a span of around 8 years and 6 months, starting from 2008-12-01 and ending at 2017-06-25.

setwd("~/Documents/Study/computerScience/programming/r/data/")  
data = read.csv("weatherAustralia.csv")[c(-6:-8, -9:-11, -22, -23)]  
# (Categorical and redundant variables have been removed)  
head(data)

**Date Location MinTemp MaxTemp Rainfall WindSpeed9am WindSpeed3pm**  
01/12/08 Albury 13.4 22.9 0.6 20 24  
02/12/08 Albury 7.4 25.1 0.0 4 22  
03/12/08 Albury 12.9 25.7 0.0 19 26  
04/12/08 Albury 9.2 28.0 0.0 11 9  
05/12/08 Albury 17.5 32.3 1.0 7 20  
06/12/08 Albury 14.6 29.7 0.2 19 24

***(continuation...)***

**Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am**  
71 22 1007.7 1007.1 8 NA 16.9  
44 25 1010.6 1007.8 NA NA 17.2  
38 30 1007.6 1008.7 NA 2 21.0  
45 16 1017.6 1012.8 NA NA 18.1  
82 33 1010.8 1006.0 7 8 17.8  
55 23 1009.2 1005.4 NA NA 20.6

***(continuation...)***

**Temp3pm**  
21.8  
24.3  
23.2  
26.5  
29.7  
28.9

## Cleaning the data (removing null values and unused columns)

data = na.omit(data[c(-1, -2)])

## Data fields summary

# Target variable (response)...  
summary(data$Rainfall)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.000 0.000 2.975 0.600 371.000

We want to be able to predict rainfall. Hence, we want to grasp the relationship between the various variables and rainfall. For the best model, we want to eliminate the unreliable or insignificant predictors, and retain the ones that reduce variation error and increase R-square.

# Predictor variables (regressors)...  
summary(data[-3])

**MinTemp MaxTemp WindSpeed9am WindSpeed3pm Humidity9am**   
Min. :-3.3 Min. : 6.8 Min. : 0.00 Min. : 0.00 Min. : 5.00   
1st Qu.: 7.7 1st Qu.:19.2 1st Qu.: 9.00 1st Qu.:11.00 1st Qu.: 51.00   
Median :13.4 Median :24.3 Median :15.00 Median :17.00 Median : 64.00   
Mean :13.0 Mean :24.6 Mean :15.23 Mean :17.25 Mean : 64.25   
3rd Qu.:18.2 3rd Qu.:29.6 3rd Qu.:20.00 3rd Qu.:22.00 3rd Qu.: 80.00   
Max. :29.7 Max. :47.3 Max. :56.00 Max. :61.00 Max. :100.00

**Humidity3pm Pressure9am Pressure3pm Cloud9am**   
Min. : 1.00 Min. : 989.8 Min. : 982.9 Min. :0.000   
1st Qu.: 27.00 1st Qu.:1013.5 1st Qu.:1010.9 1st Qu.:1.000   
Median : 44.00 Median :1017.9 Median :1015.2 Median :4.000   
Mean : 46.07 Mean :1017.9 Mean :1015.2 Mean :3.947   
3rd Qu.: 64.00 3rd Qu.:1022.3 3rd Qu.:1019.4 3rd Qu.:7.000   
Max. :100.00 Max. :1036.8 Max. :1035.0 Max. :8.000

**Cloud3pm Temp9am Temp3pm**   
Min. :0.00 Min. : 0.30 Min. : 6.40   
1st Qu.:1.00 1st Qu.:13.40 1st Qu.:18.00   
Median :5.00 Median :18.60 Median :22.70   
Mean :4.32 Mean :18.09 Mean :23.18   
3rd Qu.:7.00 3rd Qu.:22.70 3rd Qu.:28.00   
Max. :8.00 Max. :37.70 Max. :46.70

#========================

# USING IN-BUILT STEPWISE MODEL SELECTION FUNCTION

This is to identify the best predictors among the available fields, by estimate the significance of their regression coefficients. For our purpose, we will use a combination of forward and backward selection, thereby eliminating as many irrelvant variables as possible while retaining as many relevant variables as possible.

## Full model (with all regressors)

fullModel = lm(data$Rainfall~., data)

This is the model with all the available regressors. This will be used as a source for the regressors, when we create the best fitting model.

summary(fullModel)

##   
## Call:  
## lm(formula = data$Rainfall ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.35 -3.98 -1.21 1.47 361.34   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 120.056261 24.071354 4.988 6.25e-07 \*\*\*  
## MinTemp -0.002039 0.065712 -0.031 0.975244   
## MaxTemp -0.842806 0.120480 -6.995 2.86e-12 \*\*\*  
## WindSpeed9am 0.195093 0.018529 10.529 < 2e-16 \*\*\*  
## WindSpeed3pm 0.022206 0.017245 1.288 0.197882   
## Humidity9am 0.119959 0.013523 8.871 < 2e-16 \*\*\*  
## Humidity3pm 0.079536 0.015623 5.091 3.65e-07 \*\*\*  
## Pressure9am -0.585801 0.089106 -6.574 5.21e-11 \*\*\*  
## Pressure3pm 0.454385 0.090321 5.031 4.99e-07 \*\*\*  
## Cloud9am 0.221519 0.060040 3.689 0.000226 \*\*\*  
## Cloud3pm -0.096283 0.060168 -1.600 0.109588   
## Temp9am 0.135117 0.095667 1.412 0.157884   
## Temp3pm 0.908171 0.135330 6.711 2.07e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.63 on 7791 degrees of freedom  
## Multiple R-squared: 0.1418, Adjusted R-squared: 0.1405   
## F-statistic: 107.3 on 12 and 7791 DF, p-value: < 2.2e-16

As can be seen, this is a relatively poor model, with an adjusted R-squared value of 0.1405. However, we see that the p-value for the F-statistic for this model is below 0.05, meaning that at least one regressor is significantly linearly related to the response, given a 0.05 significance level.

## Applying the step function

results = step(fullModel, direction = "both", scope = formula(fullModel))

## Start: AIC=36905.06  
## data$Rainfall ~ MinTemp + MaxTemp + WindSpeed9am + WindSpeed3pm +   
## Humidity9am + Humidity3pm + Pressure9am + Pressure3pm + Cloud9am +   
## Cloud3pm + Temp9am + Temp3pm  
##   
## Df Sum of Sq RSS AIC  
## - MinTemp 1 0.1 880330 36903  
## - WindSpeed3pm 1 187.4 880518 36905  
## - Temp9am 1 225.4 880556 36905  
## <none> 880330 36905  
## - Cloud3pm 1 289.3 880620 36906  
## - Cloud9am 1 1538.1 881868 36917  
## - Pressure3pm 1 2859.7 883190 36928  
## - Humidity3pm 1 2928.4 883259 36929  
## - Pressure9am 1 4883.5 885214 36946  
## - Temp3pm 1 5088.6 885419 36948  
## - MaxTemp 1 5529.4 885860 36952  
## - Humidity9am 1 8891.9 889222 36981  
## - WindSpeed9am 1 12526.1 892856 37013  
##   
## Step: AIC=36903.06  
## data$Rainfall ~ MaxTemp + WindSpeed9am + WindSpeed3pm + Humidity9am +   
## Humidity3pm + Pressure9am + Pressure3pm + Cloud9am + Cloud3pm +   
## Temp9am + Temp3pm  
##   
## Df Sum of Sq RSS AIC  
## - WindSpeed3pm 1 188.1 880519 36903  
## <none> 880330 36903  
## - Temp9am 1 289.8 880620 36904  
## - Cloud3pm 1 290.6 880621 36904  
## + MinTemp 1 0.1 880330 36905  
## - Cloud9am 1 1717.1 882048 36916  
## - Humidity3pm 1 2971.4 883302 36927  
## - Pressure3pm 1 3112.2 883443 36929  
## - Temp3pm 1 5134.2 885465 36946  
## - Pressure9am 1 5449.6 885780 36949  
## - MaxTemp 1 5553.4 885884 36950  
## - Humidity9am 1 9415.4 889746 36984  
## - WindSpeed9am 1 13630.4 893961 37021  
##   
## Step: AIC=36902.72  
## data$Rainfall ~ MaxTemp + WindSpeed9am + Humidity9am + Humidity3pm +   
## Pressure9am + Pressure3pm + Cloud9am + Cloud3pm + Temp9am +   
## Temp3pm  
##   
## Df Sum of Sq RSS AIC  
## <none> 880519 36903  
## + WindSpeed3pm 1 188.1 880330 36903  
## - Cloud3pm 1 361.7 880880 36904  
## - Temp9am 1 411.5 880930 36904  
## + MinTemp 1 0.8 880518 36905  
## - Cloud9am 1 1730.0 882249 36916  
## - Humidity3pm 1 2850.6 883369 36926  
## - Pressure3pm 1 3083.0 883602 36928  
## - Temp3pm 1 4952.7 885471 36944  
## - MaxTemp 1 5517.8 886036 36949  
## - Pressure9am 1 5533.1 886052 36950  
## - Humidity9am 1 9682.1 890201 36986  
## - WindSpeed9am 1 16266.6 896785 37044

### Obtaining the selected variables and their coefficients…

results$coefficients

## (Intercept) MaxTemp WindSpeed9am Humidity9am Humidity3pm Pressure9am   
## 126.85507103 -0.84018511 0.20213110 0.12117627 0.07743662 -0.58895064   
## Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm   
## 0.45134894 0.22172190 -0.10658198 0.15586088 0.88061073

From the above, we see that the final selected regressors contain all the possible regressors except MinTemp (minimum temperature for the day) and WindSpeed3pm (wind speed at 3PM).

### From the results of this function, we obtain the following model..

f = formula(Rainfall~.-MinTemp-WindSpeed3pm)  
finalModel = lm(f, data)  
# (All regressors except MinTemp and WindSpeed3pm)  
summary(finalModel)

##   
## Call:  
## lm(formula = f, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.37 -3.98 -1.21 1.46 361.45   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 126.85507 23.38773 5.424 6.00e-08 \*\*\*  
## MaxTemp -0.84019 0.12023 -6.988 3.01e-12 \*\*\*  
## WindSpeed9am 0.20213 0.01685 11.999 < 2e-16 \*\*\*  
## Humidity9am 0.12118 0.01309 9.257 < 2e-16 \*\*\*  
## Humidity3pm 0.07744 0.01542 5.023 5.20e-07 \*\*\*  
## Pressure9am -0.58895 0.08416 -6.998 2.81e-12 \*\*\*  
## Pressure3pm 0.45135 0.08641 5.224 1.80e-07 \*\*\*  
## Cloud9am 0.22172 0.05666 3.913 9.19e-05 \*\*\*  
## Cloud3pm -0.10658 0.05957 -1.789 0.0736 .   
## Temp9am 0.15586 0.08167 1.908 0.0564 .   
## Temp3pm 0.88061 0.13301 6.621 3.81e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.63 on 7793 degrees of freedom  
## Multiple R-squared: 0.1416, Adjusted R-squared: 0.1405   
## F-statistic: 128.6 on 10 and 7793 DF, p-value: < 2.2e-16

# ANALYSING THE FINAL MODEL

## Summary analysis

We see that the statistics of the final model are identical to the full model, which is not too surprising, since only two regressors were eliminated. However, this indicates that while we may have simplified the model, we have not improved its predictive power. As with the full model, the adjusted R-squared value is 0.1405, meaning that in the sample, only 14.05% of the variation in the response is explained by the model. Also, we see that the F-statistic for the model is significant for a 0.05 level of significance, meaning that there is significant linear relationship between the response and at least one regressor.

## Testing for autocorrelation

Autocorrelation is the presence of correlation between current and past values of the response variable. This can occur if past values have some impact on future values of the variable. Autocorrelation may also happen at a lag i.e. current values may be related to a past value before the adjacent past value. To check for autocorrelation, we wil use the Burbin-Watson test.

library(lmtest)  
dwtest(f, data = data)

##   
## Durbin-Watson test  
##   
## data: f  
## DW = 1.4907, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0

# (NOTE: f is the previously created formula for the final model)

p-value of the test statistic is much below 0.05, meaning it is significant for a 0.05 significance level (i.e. it is unlikely enough to be considered as significant indication for autocorrelation). Hence, we may conclude that there is autocorrelation within the response variable, and judging by the sign, this is a positive autocorrelation.

This means that we may conclude that past rainfall have a significant implact on future rainfall. Since our data set is a time-series data, and since rainfall tends to occur on a seasonal basis, and not randomly, this result makes sense, since a certain amount or range of amounts of rainfall is likely to be continually sustained over a period of many days, and likely to rise or fall consistently over a period of time.

## Testing for multicollinearity

Multicollinearity is the presence of moderate to strong linear correlation between two or more regressors within a linear regression model.

# Checking the correlation matrix of the regressors...  
library(ppcor)

## Loading required package: MASS

## Warning: package 'MASS' was built under R version 3.6.2

### Choosing columns of regressors

# (All rows from all except the 3rd column)  
X = data[-3]  
correlMatrix = pcor(X, method = "pearson")  
# Correlation coefficients...  
correlMatrix$estimate

## MinTemp MaxTemp WindSpeed9am WindSpeed3pm Humidity9am  
## MinTemp 1.00000000 0.061501477 0.287360163 -0.041028403 0.23913136  
## MaxTemp 0.06150148 1.000000000 0.001413421 0.020849195 -0.15248210  
## WindSpeed9am 0.28736016 0.001413421 1.000000000 0.312880631 -0.17160385  
## WindSpeed3pm -0.04102840 0.020849195 0.312880631 1.000000000 0.08526328  
## Humidity9am 0.23913136 -0.152482102 -0.171603850 0.085263279 1.00000000  
## Humidity3pm 0.12656099 0.090554589 -0.017431948 -0.095830896 0.70381012  
## Pressure9am -0.32655781 0.025034467 0.081144135 -0.048627207 -0.07411680  
## Pressure3pm 0.29058007 -0.052871806 -0.057961565 -0.007123758 0.07090027  
## Cloud9am 0.33043862 -0.084799063 -0.090626745 0.024121616 0.03993949  
## Cloud3pm 0.04880499 0.128229704 0.094365636 -0.129394513 -0.02217944  
## Temp9am 0.48869788 0.129212090 -0.128252087 0.199734586 -0.62411159  
## Temp3pm 0.09897482 0.798496305 -0.020550395 -0.151287078 0.40180129  
## Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm  
## MinTemp 0.12656099 -0.32655781 0.290580071 0.33043862 0.048804992  
## MaxTemp 0.09055459 0.02503447 -0.052871806 -0.08479906 0.128229704  
## WindSpeed9am -0.01743195 0.08114414 -0.057961565 -0.09062674 0.094365636  
## WindSpeed3pm -0.09583090 -0.04862721 -0.007123758 0.02412162 -0.129394513  
## Humidity9am 0.70381012 -0.07411680 0.070900266 0.03993949 -0.022179437  
## Humidity3pm 1.00000000 0.20197817 -0.190457830 0.06123252 0.109789204  
## Pressure9am 0.20197817 1.00000000 0.966030368 0.02530833 0.025231174  
## Pressure3pm -0.19045783 0.96603037 1.000000000 -0.03803721 -0.041267847  
## Cloud9am 0.06123252 0.02530833 -0.038037213 1.00000000 0.434162632  
## Cloud3pm 0.10978920 0.02523117 -0.041267847 0.43416263 1.000000000  
## Temp9am 0.59753255 -0.07868448 0.077509757 -0.20387758 -0.008395301  
## Temp3pm -0.55908710 0.19943063 -0.177403233 0.05400774 -0.130165762  
## Temp9am Temp3pm  
## MinTemp 0.488697876 0.09897482  
## MaxTemp 0.129212090 0.79849631  
## WindSpeed9am -0.128252087 -0.02055040  
## WindSpeed3pm 0.199734586 -0.15128708  
## Humidity9am -0.624111590 0.40180129  
## Humidity3pm 0.597532546 -0.55908710  
## Pressure9am -0.078684477 0.19943063  
## Pressure3pm 0.077509757 -0.17740323  
## Cloud9am -0.203877576 0.05400774  
## Cloud3pm -0.008395301 -0.13016576  
## Temp9am 1.000000000 0.29609802  
## Temp3pm 0.296098017 1.00000000

# p-values for the correlation coefficients...  
correlMatrix$p.value

## MinTemp MaxTemp WindSpeed9am WindSpeed3pm  
## MinTemp 0.000000e+00 5.515187e-08 4.681754e-148 2.911146e-04  
## MaxTemp 5.515187e-08 0.000000e+00 9.007120e-01 6.568790e-02  
## WindSpeed9am 4.681754e-148 9.007120e-01 0.000000e+00 1.390485e-176  
## WindSpeed3pm 2.911146e-04 6.568790e-02 1.390485e-176 0.000000e+00  
## Humidity9am 8.795699e-102 9.234336e-42 1.397152e-52 4.704133e-14  
## Humidity3pm 3.380155e-29 1.150207e-15 1.238468e-01 2.281923e-17  
## Pressure9am 4.374502e-193 2.709645e-02 7.263416e-13 1.748204e-05  
## Pressure3pm 1.698252e-151 3.008582e-06 3.046591e-07 5.294680e-01  
## Cloud9am 6.307329e-198 6.446478e-14 1.091785e-15 3.321191e-02  
## Cloud3pm 1.628419e-05 6.195419e-30 6.928841e-17 1.870213e-30  
## Temp9am 0.000000e+00 2.257739e-30 6.055079e-30 5.888778e-71  
## Temp3pm 1.988055e-18 0.000000e+00 6.965246e-02 3.958307e-41  
## Humidity9am Humidity3pm Pressure9am Pressure3pm  
## MinTemp 8.795699e-102 3.380155e-29 4.374502e-193 1.698252e-151  
## MaxTemp 9.234336e-42 1.150207e-15 2.709645e-02 3.008582e-06  
## WindSpeed9am 1.397152e-52 1.238468e-01 7.263416e-13 3.046591e-07  
## WindSpeed3pm 4.704133e-14 2.281923e-17 1.748204e-05 5.294680e-01  
## Humidity9am 0.000000e+00 0.000000e+00 5.705255e-11 3.699279e-10  
## Humidity3pm 0.000000e+00 0.000000e+00 1.500124e-72 1.437611e-64  
## Pressure9am 5.705255e-11 1.500124e-72 0.000000e+00 0.000000e+00  
## Pressure3pm 3.699279e-10 1.437611e-64 0.000000e+00 0.000000e+00  
## Cloud9am 4.205424e-04 6.301088e-08 2.546262e-02 7.829864e-04  
## Cloud3pm 5.022907e-02 2.467351e-22 2.591407e-02 2.681794e-04  
## Temp9am 0.000000e+00 0.000000e+00 3.495142e-12 7.279653e-12  
## Temp3pm 2.705105e-300 0.000000e+00 9.649100e-71 3.968809e-56  
## Cloud9am Cloud3pm Temp9am Temp3pm  
## MinTemp 6.307329e-198 1.628419e-05 0.000000e+00 1.988055e-18  
## MaxTemp 6.446478e-14 6.195419e-30 2.257739e-30 0.000000e+00  
## WindSpeed9am 1.091785e-15 6.928841e-17 6.055079e-30 6.965246e-02  
## WindSpeed3pm 3.321191e-02 1.870213e-30 5.888778e-71 3.958307e-41  
## Humidity9am 4.205424e-04 5.022907e-02 0.000000e+00 2.705105e-300  
## Humidity3pm 6.301088e-08 2.467351e-22 0.000000e+00 0.000000e+00  
## Pressure9am 2.546262e-02 2.591407e-02 3.495142e-12 9.649100e-71  
## Pressure3pm 7.829864e-04 2.681794e-04 7.279653e-12 3.968809e-56  
## Cloud9am 0.000000e+00 0.000000e+00 6.482116e-74 1.835719e-06  
## Cloud3pm 0.000000e+00 0.000000e+00 4.586560e-01 8.410638e-31  
## Temp9am 6.482116e-74 4.586560e-01 0.000000e+00 1.692871e-157  
## Temp3pm 1.835719e-06 8.410638e-31 1.692871e-157 0.000000e+00

### EXPLANATION

The p-value of a correlation coefficient for two variables (obtained from a sample) is the probability that you would have found the current result if the correlation coefficient were in fact zero (i.e. if you found correlation in the sample while there is none in the population i.e. wrongly rejected null hypothesis). If this probability is lower than the significance level, the correlation coefficient is said to be statistically significant. Significance level is the proportion of the lowest probability values of a statistic that, if the statistic actually takes such a value, you would consider it as significantly different from the population. Here, the population distribution contains probabilities of getting values assuming that the null hypothesis is true (in this case, null hypothesis is that there is zero correlation between the two variables).

### CONCLUSIONS

We see that most of the regressors have weak positive correlation. The only ones (that I can spot) with moderate correlation are Pressure3pm and WindSpeed3pm. The p-values for these coefficients are both significant and insignificant, meaning that some regressors are associated, maybe not strongly, but enough to have been less likely similar by chance.

## Searching for a potential cause of multicollinearity

We need the VIF (variance inflation factor i.e. the factor by which a variable’s regression coefficient has inflated from the constant variance of the error term) of each regressor. As a thumb rule, if the VIF of a regressor is greater than 10, it indicates that the particular regressor is the cause of the multicollinearity in the model.

library(car)  
vif(finalModel)

**MaxTemp WindSpeed9am Humidity9am Humidity3pm Pressure9am Pressure3pm**   
49.192340 1.153124 4.791248 8.639508 19.660955 19.890482

**Cloud9am Cloud3pm Temp9am Temp3pm**   
2.078465 1.904148 17.115005 56.452187

### CONCLUSIONS

We can see that MaxTemp, Pressure9am, Pressure3pm, Temp9am and Temp3pm have VIF values greater than 10. This indicates that all these regressors are contributing to the multicollinearity in the model. Significant correlation between regressors of the same thing in the same day (ex. Temp9am and Temp3pm) may be expected to be correlated, since they are taken in the same day merely 6 hourse apart, and the weather is not likely to change drastically in that time period. Also, some quantities are related physically to some degree, such as air pressure and wind speed. Given all this, it is not surprising that multicollinearity exists in the model.

## Testing for heteroscedasticity

Homoscedasticity in a model means that the error is constant i,e, error terms are equal for a given regressor value (hence, it must be measured across models taken from different samples of the same population). The best way for checking homoscedasticity is to make a scatterplot with the residuals against the regressor. However, due to a large number of variables in the model, we will not do this. Instead, to test the degree of heteroscedasticity in the response ‘Rainfall’, we will use the Breusch-Pagan test (note that the null hypothesis claims homoscedasticity).

library(lmtest)  
bptest(finalModel)

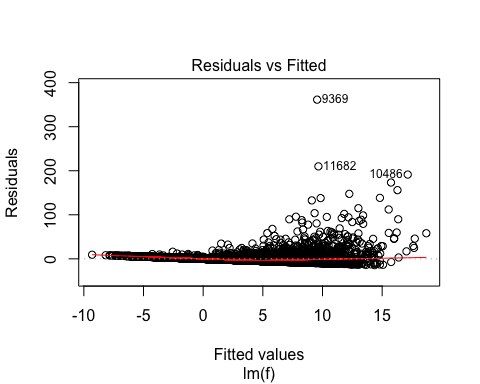
##   
## studentized Breusch-Pagan test  
##   
## data: finalModel  
## BP = 95.576, df = 10, p-value = 4.169e-16

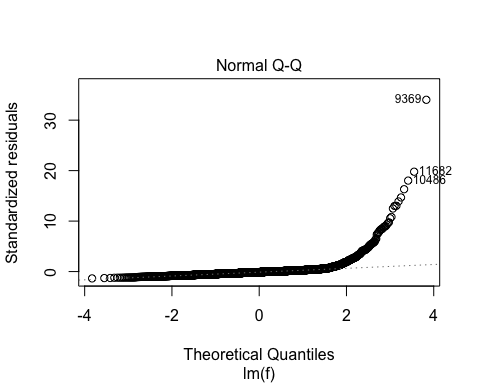
As we can see, the p-value of the test statistic BP is less than 0.05. This means we may reject the null hypothesis, hence accepting the claim that the error terms in the model are significantly unequal. This violates one of the assumptions of linear regression modelling, hence indicating that our model may not be a reliable predictor of the response i.e. Rainfall.

## Residual plots

Residual plots are designed to plot the key values of the model, after considering all the regressors. Jence, it is a way for us to analyse the model, and conclude on its reliablility.

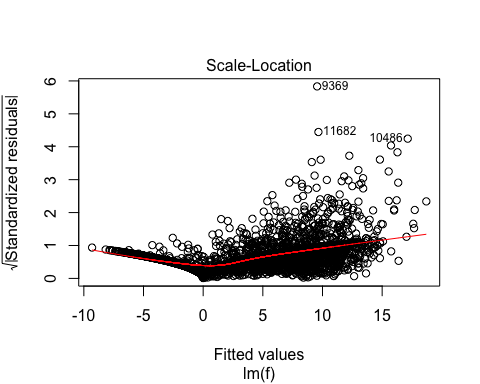
plot(finalModel)

 Residual vs. fitted plot shows a roughly expanding funnel, implying that the variance of the standardized residuals is not constant for all response values, and increases for larger values.

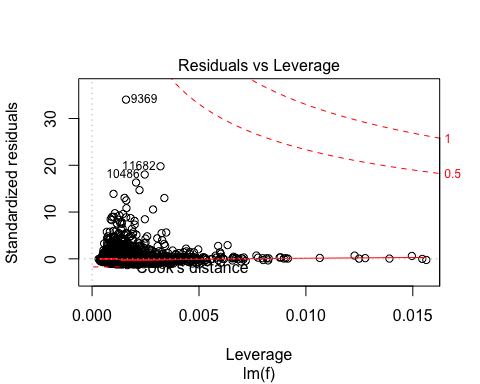


Q-Q plot shows ideal probability distribution until a certain point, but diverges greatly after the 2nd theoretical quartile. Nence we may conclude that

* The standardized residuals are not normally distributed
* All variables may not be linearly associated



Scale-location plot shows a roughly expanding closing funnel, implying that the variance of the standardized residuals is not constant for all predictor values (ages), and increases with fitted response values.

Residuals vs. leverage plot shows that all standardized residuals are completely outside Cook's distance, implying that there numerous outliers and extreme values (indicating some nature of Australian weather perhaps).