**STAT581 Assignment Report**

**Team Members**

Name ruid netid

Nishant Kumar 186003764 nak112

Kush Aswani 187001586 kaa189

Tejasva Tomar 189007957 tt425

**Background**

A tropical cyclone is a rapidly rotating storm system characterized by a low-pressure center, a closed low-level atmospheric circulation, strong winds, and a spiral arrangement of thunderstorms that produce heavy rain. Depending on its location and strength, a tropical cyclone is referred to by different names, including hurricane, tropical storm, cyclonic storm, tropical depression, and simply cyclone. A hurricane is a tropical cyclone that occurs in the Atlantic Ocean and northeastern Pacific Ocean. The Saffir–Simpson hurricane wind scale (SSHWS classifies hurricanes into five categories distinguished by the intensities of their sustained winds. To be classified as a hurricane, a tropical cyclone must have maximum sustained winds of at least 74 mph (33 m/s; 64 kn; 119 km/h) (Category 1). The highest classification in the scale, Category 5, consists of storms with sustained winds exceeding 156 mph (70 m/s; 136 kn; 251 km/h).

**Goal of Assignment**

We now try to analyze the information we get from List of Category 4 Hurricanes (<https://en.wikipedia.org/wiki/List_of_Category_4_Atlantic_hurricanes>) and List of Category 5 Hurricanes (<https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes>) and see what information we can get from them.

**Applying Monte-Carlo simulation**

We first create the variance to mean ratio from given data of category 4 hurricane. Then using rpois we generate a distribution for category 4 hurricane’s Lambda. Sample size will remain same as category 4 hurricane. We then calculate the variance to mean ratio of this sample. Then we generated 10000 samples like this and get the ratio for all samples. Then we compared all the ratios to the original ratio found that the ratios are same only 4 times out 10000. Thus, it means that this data is a over dispersed poisson distribution.

> library(readxl)

> FINAL\_TABLE\_CAT4 <- read\_excel("FINAL TABLE-CAT4.xlsx")

> View(FINAL\_TABLE\_CAT4)

> data=FINAL\_TABLE\_CAT4

> lambda=mean(data$OCCURANCES)

> variance=((var(data$OCCURANCES))\*17)/18

> variance=((var(data$OCCURANCES))\*16)/17

> rat=variance/lambda

> ratio=matrix(,nrow=10000,ncol=1)

> d1=matrix(,nrow=17,ncol=1)

> count=0

> for(x in 1:10000){

+ for(y in 1:17){

+ d1[y,1]=rpois(1,lambda)

+ }

+ ratio[x,1]=((var(d1[,1]))/mean(d1[,1]))

+ if(ratio[x,1]>=rat){

+ count=count+1

+ }

+ }

count

[1] 4

> sort(ratio,decreasing = TRUE)

[1] 3.267857 3.095652 2.987762 2.825203 2.781780 2.773026 2.703125 2.682540

[9] 2.678571 2.654948 2.636628 2.616803 2.587209 2.583000 2.436321 2.434375

[17] 2.431481 2.393478 2.382282 2.361650 2.344298 2.327756 2.321101 2.320697

[25] 2.320455 2.316116 2.315848 2.315848 2.309476 2.303279 2.299618 2.274336

[33] 2.270833 2.260000 2.259346 2.251025 2.250000 2.250000 2.243182 2.243164

[41] 2.232143 2.228306 2.226562 2.214286 2.214015 2.210630 2.210630 2.209961

[49] 2.200175 2.196429 2.195370 2.183036 2.175620 2.171739 2.168699 2.165138

[57] 2.163660 2.163321 2.157895 2.146516 2.145833 2.143885 2.142699 2.140496

[65] 2.136480 2.134783 2.129098 2.120327 2.113861 2.109127 2.107143 2.097917

[73] 2.095519 2.092822 2.079427 2.079348 2.079348 2.070248 2.069556 2.069318

[81] 2.067478 2.064693 2.060870 2.060547 2.056000 2.053704 2.052686 2.048673

[89] 2.045139 2.040888 2.031250 2.031250 2.030682 2.029867 2.027132 2.026575

[97] 2.024590 2.017562 2.011062 2.009375 2.007937 2.007937 2.005757 2.005435

[105] 2.005435 2.001008 2.000000 1.996429 1.996269 1.995000 1.993304 1.991667

[113] 1.990291 1.986957 1.986111 1.983871 1.981308 1.980410 1.977713 1.974330

[121] 1.974206 1.974099 1.973958 1.972727 1.968478 1.966435 1.964552 1.964440

[129] 1.962413 1.959402 1.958969 1.957341 1.955357 1.954918 1.954646 1.954000

[137] 1.950000 1.946121 1.944542 1.942913 1.938542 1.937000 1.936384 1.936384

[145] 1.934211 1.931193 1.928571 1.928295 1.927778 1.927381 1.927083 1.921875

[153] 1.920082 1.920000 1.917411 1.917373 1.916978 1.916667 1.916667 1.915404

[161] 1.913043 1.913043 1.912037 1.911822 1.907407 1.907407 1.906746 1.904915

[169] 1.903125 1.899364 1.898438 1.898230 1.898230 1.896930 1.895390 1.894628

[177] 1.894565 1.894565 1.893939 1.892857 1.892857 1.892276 1.892276 1.885417

[185] 1.885246 1.879425 1.877930 1.877404 1.876812 1.876812 1.876136 1.875000

[193] 1.875000 1.873016 1.873016 1.872706 1.867828 1.867828 1.867708 1.864815

[201] 1.863347 1.862500 1.862150 1.861641 1.860619 1.859848 1.859252 1.859252

[209] 1.857609 1.857609 1.857609 1.857609 1.856818 1.856818 1.854526 1.853211

[217] 1.850427 1.850000 1.848380 1.846154 1.845339 1.845339 1.843750 1.839130

[225] 1.836538 1.834906 1.832292 1.832292 1.829637 1.829637 1.829637 1.829464

[233] 1.828704 1.827331 1.825787 1.825188 1.822430 1.821828 1.821828 1.821429

[241] 1.820946 1.820946 1.820946 1.820652 1.817888 1.814583 1.814220 1.813462

[249] 1.813462 1.812500 1.812500 1.811224 1.809322 1.809055 1.807554 1.806569

[257] 1.805970 1.805970 1.805952 1.804612 1.803728 1.803571 1.802174 1.799107

[265] 1.798864 1.797115 1.797115 1.796875 1.796756 1.795940 1.795363 1.794811

[273] 1.793233 1.789352 1.789352 1.785714 1.785714 1.785714 1.785398 1.785088

[281] 1.783696 1.782658 1.781250 1.780769 1.780738 1.780534 1.780039 1.779167

[289] 1.777778 1.777228 1.775591 1.775591 1.775547 1.773352 1.773305 1.771694

[297] 1.771341 1.768750 1.767000 1.767000 1.767000 1.765625 1.765625 1.765476

[305] 1.765217 1.764423 1.763350 1.762931 1.762850 1.761458 1.759615 1.758858

[313] 1.758396 1.756188 1.755297 1.755297 1.754960 1.754717 1.754132 1.754132

[321] 1.754065 1.753226 1.750000 1.750000 1.750000 1.750000 1.748077 1.747788

[329] 1.746652 1.746652 1.746173 1.745902 1.744526 1.742991 1.742991 1.741453

[337] 1.741453 1.740909 1.738393 1.738095 1.737288 1.736789 1.736239 1.736239

[345] 1.734375 1.733221 1.733000 1.733000 1.732143 1.731870 1.731115 1.730324

[353] 1.729167 1.729015 1.728982 1.728516 1.728299 1.728261 1.727679 1.727679

[361] 1.727679 1.726815 1.726815 1.726293 1.726293 1.726293 1.725394 1.723684

[369] 1.722222 1.721591 1.719512 1.719136 1.719008 1.716000 1.715385 1.714623

[377] 1.714623 1.714286 1.714286 1.714286 1.713942 1.713346 1.711066 1.711066

[385] 1.711066 1.711066 1.711066 1.710648 1.710526 1.710177 1.709783 1.709677

[393] 1.708661 1.708333 1.708333 1.707974 1.706081 1.706081 1.705128 1.704762

[401] 1.701456 1.701446 1.701316 1.699038 1.699038 1.697674 1.697674 1.697674

[409] 1.697674 1.697500 1.695313 1.695313 1.694963 1.694575 1.693548 1.692540

[417] 1.691886 1.691372 1.691372 1.691304 1.689655 1.689212 1.689167 1.687500

[425] 1.687500 1.686966 1.683884 1.683884 1.683884 1.683884 1.683411 1.683263

[433] 1.683263 1.682692 1.682000 1.681202 1.681202 1.680070 1.678571 1.678571

[441] 1.678571 1.677752 1.677752 1.676250 1.676250 1.676230 1.676230 1.675403

[449] 1.675403 1.675197 1.674528 1.673246 1.672826 1.671875 1.671336 1.670968

[457] 1.670759 1.670635 1.670635 1.670635 1.668803 1.667793 1.666667 1.666346

[465] 1.666322 1.665414 1.662109 1.662109 1.662109 1.658854 1.658266 1.658266

[473] 1.655208 1.655000 1.654676 1.654348 1.654348 1.653761 1.653761 1.653761

[481] 1.653017 1.650568 1.650568 1.650407 1.650407 1.650407 1.650000 1.650000

[489] 1.649436 1.648760 1.648256 1.647246 1.647246 1.644318 1.644318 1.644048

[497] 1.644048 1.643692 1.642857 1.642857 1.641732 1.641393 1.641393 1.641393

[505] 1.641129 1.639563 1.639563 1.639563 1.637755 1.637755 1.637500 1.636905

[513] 1.635965 1.635965 1.635965 1.635965 1.635870 1.635870 1.634542 1.634434

[521] 1.634434 1.634211 1.633750 1.633750 1.633459 1.632653 1.632479 1.631000

[529] 1.631000 1.631000 1.629505 1.629505 1.629237 1.628289 1.625000 1.623992

[537] 1.623992 1.623975 1.623975 1.623975 1.620438 1.620438 1.619792 1.619792

[545] 1.618932 1.618932 1.618321 1.618321 1.617481 1.617325 1.616438 1.616379

[553] 1.616150 1.616150 1.616150 1.616150 1.615854 1.615854 1.615310 1.614583

[561] 1.614387 1.614387 1.614316 1.614316 1.614316 1.614000 1.613636 1.613636

[569] 1.612500 1.612500 1.611779 1.611229 1.611229 1.610360 1.610360 1.608268

[577] 1.608268 1.607143 1.607143 1.606855 1.606855 1.606855 1.606557 1.606557

[585] 1.605682 1.604167 1.603972 1.603175 1.603175 1.602083 1.599771 1.599638

[593] 1.599138 1.599138 1.599138 1.598837 1.598577 1.598301 1.598060 1.598060

[601] 1.598060 1.597345 1.597345 1.597222 1.596154 1.596154 1.594866 1.594866

[609] 1.594866 1.594866 1.594340 1.594340 1.594072 1.593750 1.593750 1.593220

[617] 1.592593 1.592593 1.591535 1.591346 1.591216 1.589718 1.589718 1.589416

[625] 1.589286 1.589139 1.587871 1.586364 1.586364 1.586364 1.586364 1.586364

[633] 1.586310 1.586174 1.585526 1.585387 1.584112 1.584112 1.584112 1.583333

[641] 1.581481 1.581481 1.581301 1.580435 1.580435 1.580435 1.580275 1.580044

[649] 1.580000 1.580000 1.579741 1.578540 1.578512 1.577991 1.577991 1.575893

[657] 1.575212 1.575212 1.575212 1.572917 1.572917 1.572917 1.572704 1.572581

[665] 1.572581 1.572165 1.572072 1.571429 1.571429 1.571429 1.570423 1.570000

[673] 1.569828 1.569656 1.569444 1.569444 1.567045 1.567045 1.567045 1.566832

[681] 1.566667 1.566667 1.566667 1.565891 1.565891 1.565741 1.564252 1.563095

[689] 1.562500 1.561957 1.561667 1.561422 1.561422 1.561404 1.561404 1.560950

[697] 1.560950 1.560780 1.560780 1.559829 1.559829 1.559735 1.558394 1.558394

[705] 1.558071 1.557203 1.557203 1.557203 1.557039 1.556920 1.555172 1.555172

[713] 1.554303 1.554303 1.554245 1.554245 1.553977 1.553977 1.553571 1.553571

[721] 1.553571 1.553571 1.552928 1.552579 1.552579 1.551215 1.549419 1.549419

[729] 1.548958 1.548958 1.548958 1.547727 1.547727 1.547727 1.546000 1.546000

[737] 1.546000 1.546000 1.544393 1.543478 1.543478 1.543388 1.543388 1.543388

[745] 1.543103 1.543103 1.543103 1.543103 1.543103 1.542763 1.542763 1.541667

[753] 1.541667 1.541339 1.541339 1.541284 1.540929 1.539195 1.539195 1.539195

[761] 1.538306 1.538306 1.538043 1.538043 1.537946 1.537879 1.537594 1.537594

[769] 1.537214 1.537214 1.536885 1.536885 1.536408 1.536408 1.536408 1.535714

[777] 1.535714 1.535577 1.533784 1.533784 1.533602 1.533565 1.532946 1.532946

[785] 1.532946 1.532946 1.532946 1.531250 1.531250 1.531250 1.531250 1.531250

[793] 1.531250 1.531250 1.530048 1.529472 1.529297 1.529000 1.528409 1.528409

[801] 1.528409 1.528351 1.527500 1.527500 1.525826 1.525826 1.525000 1.524784

[809] 1.524606 1.524533 1.524533 1.524123 1.524123 1.524123 1.524123 1.524123

[817] 1.523504 1.523504 1.522645 1.522124 1.521186 1.521186 1.521169 1.521169

[825] 1.520992 1.520833 1.519467 1.519467 1.519231 1.518973 1.518849 1.518849

[833] 1.517857 1.517086 1.515777 1.515625 1.513889 1.513889 1.513542 1.513542

[841] 1.512695 1.512695 1.512695 1.512000 1.512000 1.512000 1.512000 1.511861

[849] 1.508264 1.508264 1.508264 1.507576 1.507576 1.507576 1.507576 1.507246

[857] 1.506522 1.506466 1.506466 1.506466 1.506466 1.506250 1.505682 1.505556

[865] 1.505482 1.505342 1.504771 1.504771 1.504673 1.504673 1.504032 1.503713

[873] 1.503319 1.503178 1.503178 1.502778 1.502778 1.502717 1.502551 1.502294

[881] 1.502294 1.502049 1.500000 1.500000 1.500000 1.500000 1.500000 1.496350

[889] 1.496094 1.496094 1.496094 1.495495 1.495495 1.495495 1.495146 1.495146

[897] 1.495000 1.495000 1.494919 1.494919 1.494919 1.494213 1.491142 1.491142

[905] 1.490702 1.489773 1.489773 1.489662 1.489583 1.488806 1.488147 1.488043

[913] 1.487903 1.487179 1.486895 1.486895 1.486895 1.486842 1.486842 1.486842

[921] 1.486511 1.486111 1.485169 1.485119 1.485119 1.485119 1.485000 1.484813

[929] 1.484813 1.484631 1.484513 1.484513 1.484513 1.484375 1.483527 1.483527

[937] 1.483527 1.482798 1.482798 1.482798 1.482798 1.482143 1.482143 1.482143

[945] 1.482143 1.482143 1.479492 1.478125 1.478125 1.478000 1.477642 1.477642

[953] 1.477632 1.477431 1.476449 1.476351 1.474537 1.474537 1.474537 1.474515

[961] 1.474515 1.474515 1.474507 1.474409 1.474409 1.473684 1.473684 1.473684

[969] 1.473485 1.473140 1.473140 1.472948 1.472328 1.472328 1.472328 1.472328

[977] 1.472328 1.472028 1.470455 1.470455 1.470192 1.470192 1.470192 1.469758

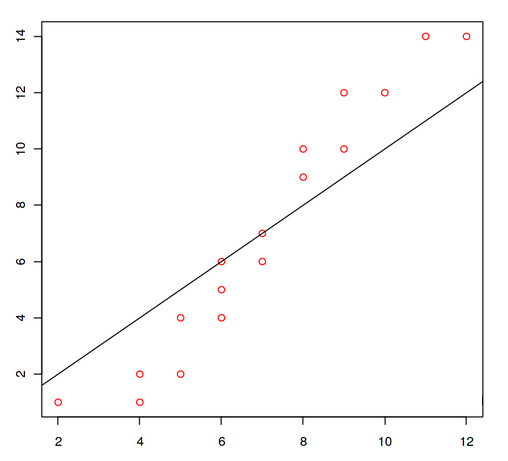
[985] 1.469758 1.469758 1.469565 1.469565 1.469017 1.469017 1.469017 1.469017

[993] 1.469017 1.468750 1.468750 1.468254 1.468202 1.468202 1.467213 1.467161

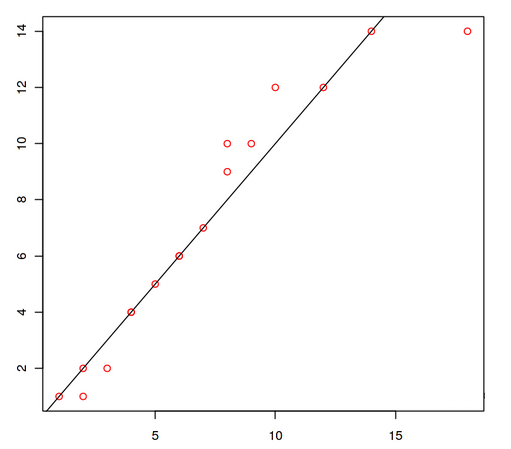
[ reached getOption("max.print") -- omitted 9000 entries ]

So -

We generate a QQplot comparison with Poisson distribution.



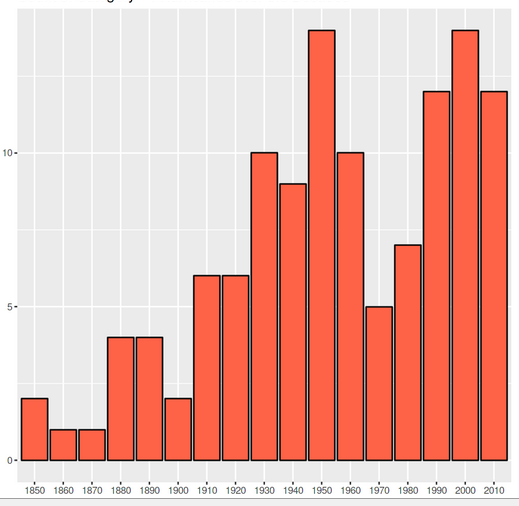
we compare it to a negative binomial distribution. Thus we get to know that this dataset is more like a negative binomial distribution.



**Analyzing Category 4 Hurricane data**

Category 4 hurricanes are tropical cyclones that reach Category 4 intensity on the Saffir–Simpson Hurricane Scale. Category 4 is the second-highest hurricane classification category on the Saffir–Simpson Hurricane Scale, and storms that are of this intensity maintain maximum sustained winds of 113–136 knots (130–156 mph, 209–251 km/h). Based on the Atlantic hurricane database, 94 hurricanes have attained Category 4 hurricane status since 1851, the start of modern meteorological record keeping

All data listed is provided by the NHC best track from decade 1851-2010.

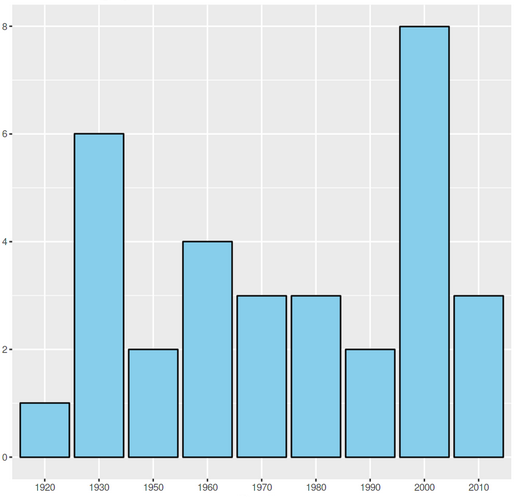


We see that the number of category-4 hurricanes occurring in a decade has increased with time gradually.

**Analyzing Category 5 Hurricane data**

A Category 5 Atlantic hurricane is one that is considered by the United States National Hurricane Center (NHC), to have had sustained wind speeds greater than 136 knots (157 mph; 252 km/h; 70 m/s) on the Saffir–Simpson scale. A total of 33 recorded tropical cyclones have reached Category 5 strength on the Saffir–Simpson hurricane wind scale in the Atlantic Ocean north of the equator, the Caribbean Sea, and the Gulf of Mexico. Hurricanes of such intensity occur once every three years in this region on average.

All data listed is provided by the NHC best track from decade 1920-2010.



We see that the number of category-5 hurricanes occurring in a decade has not increased with time gradually.

**Building Generalized linear model**

Linear regression is one of the most basic and commonly used type of predictive analysis.

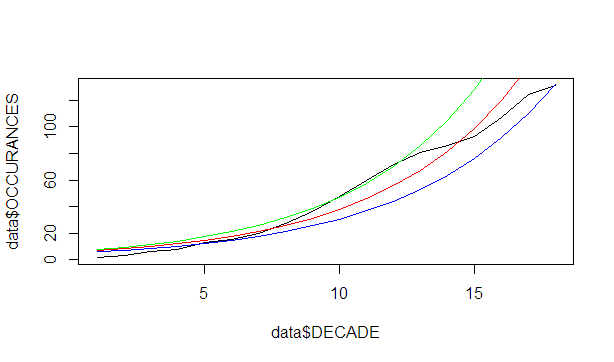
We try to build a regression model between the occurrences and the years of occurrence.

In the process we segregate data into the ratio of 80:20 of total timespan. We train the model based on the data we have from the first 80 percent of timespan. In this we use the Generalized Linear Progression Model. This trained model is then used to determine the values of the next 20 percent of timespan.

Now that we have two types of values for the last 20 percent, one achieved through prediction using the linear model and the other through the actual facts that we possess. We can clearly see that there is slight difference in the values obtained from prediction and the actual data that we have. It can also be noticed that few points are almost similar to the linear regression line that we have achieved from training our model. This way we can predict the values for next 10 years and so on and can also compare the results from Poisson distribution and the linear regression model.

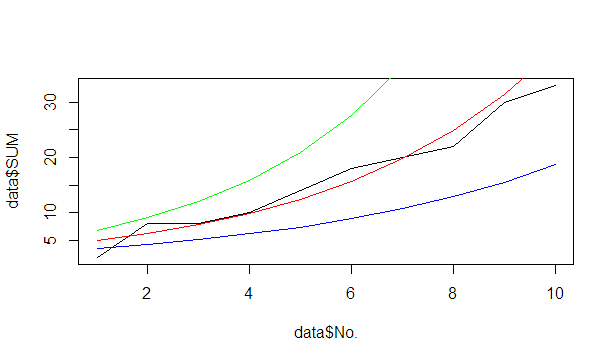
**GLM for category 4 hurricane data**

In this plot, we have used the GLM function in R Studio which also gives error rate. The green line indicates if we add error rate to the data. The red line is the linear model generated by the GLM function. The blue line indicates if subtract error rate from the data. Black line is the original data from <https://en.wikipedia.org/wiki/List_of_Category_4_Atlantic_hurricanes>



**GLM for category 5 hurricane data (decade-wise)**

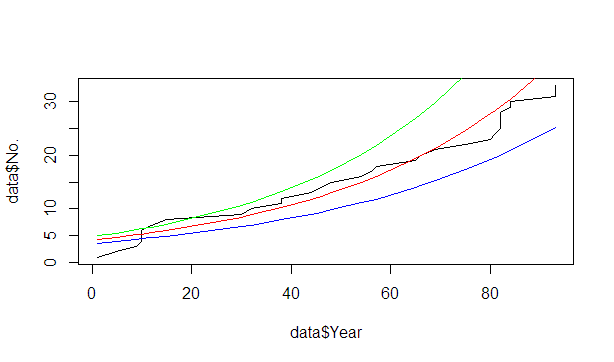
In this plot, we have used the GLM function in R Studio which also gives error rate. The green line indicates if we add error rate to the data. The red line is the linear model generated by the GLM function. The blue line indicates if subtract error rate from the data. Black line is the original data from <https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes>



As in category 5 there are only 10 decades so we get a very broad graph from it so to get more accurate representation we try to analyze year-wise occurrence of Category 5 hurricanes.

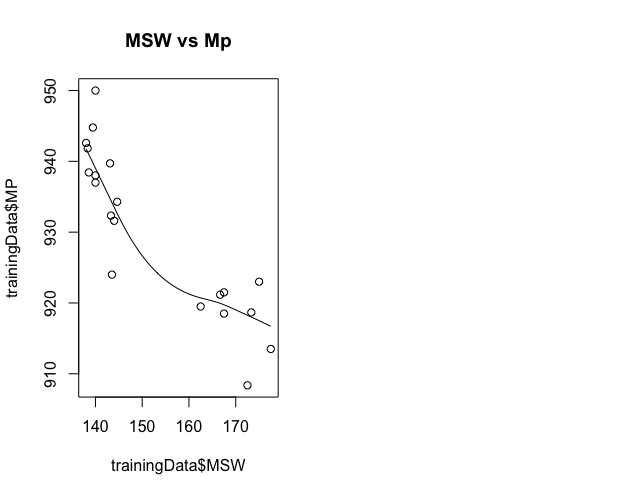
**GLM for category 5 hurricane data (year-wise)**

In this plot, we have used the GLM function in R Studio which also gives error rate. The green line indicates if we add error rate to the data. The red line is the linear model generated by the GLM function. The blue line indicates if subtract error rate from the data. Black line is the original data from <https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes>



**Analyzing Characteristics of hurricanes**

To use the data from other fields we try to plot few other graphs. One such graph is the relational graph between the Max Sustained Winds and the minimum pressure of the hurricanes. To do this we combined the data of both the category 4 and 5. We build a linear model between them. We use that data to predict the wind speeds of the next hurricane that might fall in this category. This can give us additional information regarding the hurricanes that might occur a specific number of times in the following decade which we can calculate using the Poisson distribution.



**Inference**

* We conclude from the Monte-Carlo simulation that the distribution is over dispersed Poisson distribution.
* In the third graph we can see that the observed data of Category 4 Hurricanes falls in the range generated by the GLM function.
* From Graph 4 and 5 we can observer there is not enough data to get pinpoint approximation information for future Category 5 hurricanes, but we can get a general approximation.
* Other data from data sets can be used to predict the further characteristics of future hurricanes

**Result**

From the data of category 4 and category 5 hurricanes we can say they follow an over dispersed Poisson distribution. So, we compare it to a negative binomial distribution. If we can gather more information about hurricanes we can now use Poisson Distribution to predict occurrences of hurricanes. To determine other characteristics of hurricane we can use other regression models to define those properties. As this is just a cursory analysis of the data, we show basic plots and we if need to dig deeper into the data we need to process larger datasets.