Untitled

Sean Park

2022 2 15

retail = read.csv("total\_data.csv", header = TRUE)  
names(retail) = c("Date","Temp","Rain","Wind","Humd","Snow","G\_temp", "A\_temp","Date1","Day","Total\_retail","Store\_retail","Deliv\_retail")  
retail = subset(retail,select = -Date1)  
summary(retail)

## Date Temp Rain Wind   
## Length:880 Min. :-7.958 Min. :0.000000 Min. :0.08333   
## Class :character 1st Qu.: 7.447 1st Qu.:0.000000 1st Qu.:0.95312   
## Mode :character Median :14.387 Median :0.000000 Median :1.28333   
## Mean :14.548 Mean :0.185928 Mean :1.43930   
## 3rd Qu.:22.240 3rd Qu.:0.004167 3rd Qu.:1.78333   
## Max. :30.629 Max. :7.058333 Max. :4.63750   
## Humd Snow G\_temp A\_temp   
## Min. : 23.08 Min. :0.000e+00 Min. :-3.792 Min. : 1.179   
## 1st Qu.: 53.73 1st Qu.:0.000e+00 1st Qu.: 6.616 1st Qu.: 7.829   
## Median : 67.98 Median :0.000e+00 Median :15.602 Median :16.163   
## Mean : 66.92 Mean :6.629e-05 Mean :15.287 Mean :16.001   
## 3rd Qu.: 80.04 3rd Qu.:0.000e+00 3rd Qu.:24.014 3rd Qu.:24.129   
## Max. :100.00 Max. :5.833e-02 Max. :34.779 Max. :29.829   
## Day Total\_retail Store\_retail Deliv\_retail   
## Length:880 Min. : 563400 Min. : 345580 Min. : 15100   
## Class :character 1st Qu.: 946500 1st Qu.: 566115 1st Qu.: 302750   
## Mode :character Median :1113750 Median : 667150 Median : 418150   
## Mean :1142001 Mean : 698353 Mean : 443648   
## 3rd Qu.:1316675 3rd Qu.: 804950 3rd Qu.: 562315   
## Max. :2771700 Max. :2533100 Max. :1068300

retail["Odds"] = (retail$Store\_retail/retail$Total\_retail)/(1-(retail$Store\_retail/retail$Total\_retail))  
summary(retail$Odds)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.4929 1.1657 1.6520 2.0403 2.2930 64.9603

retail["Deliv\_per"] = retail$Deliv\_retail/retail$Total\_retail  
summary(retail$Deliv\_per)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01516 0.30368 0.37708 0.38193 0.46174 0.66982

deliv\_per.glm = glm(Deliv\_per~Temp+Rain+Wind+Humd+Snow+G\_temp+A\_temp+Day, data=retail, family = binomial(link = "logit"))

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

summary(deliv\_per.glm)

##   
## Call:  
## glm(formula = Deliv\_per ~ Temp + Rain + Wind + Humd + Snow +   
## G\_temp + A\_temp + Day, family = binomial(link = "logit"),   
## data = retail)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9562 -0.1577 -0.0124 0.1568 0.6255   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.6093748 0.4881047 -1.248 0.212  
## Temp 0.0464188 0.0412928 1.124 0.261  
## Rain 0.0107234 0.1125126 0.095 0.924  
## Wind 0.0364122 0.1204100 0.302 0.762  
## Humd -0.0005073 0.0064585 -0.079 0.937  
## Snow 14.1926587 35.2445746 0.403 0.687  
## G\_temp -0.0618223 0.0486857 -1.270 0.204  
## A\_temp 0.0213227 0.0372003 0.573 0.567  
## Day목 0.0184586 0.2622497 0.070 0.944  
## Day수 0.0499716 0.2619014 0.191 0.849  
## Day월 0.0418481 0.2607894 0.160 0.873  
## Day일 0.1789103 0.2585730 0.692 0.489  
## Day토 -0.0335454 0.2623664 -0.128 0.898  
## Day화 -0.0236745 0.2620097 -0.090 0.928  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 45.029 on 879 degrees of freedom  
## Residual deviance: 41.806 on 866 degrees of freedom  
## AIC: 1004.8  
##   
## Number of Fisher Scoring iterations: 3

step(deliv\_per.glm, test="LRT")

## Start: AIC=1004.78  
## Deliv\_per ~ Temp + Rain + Wind + Humd + Snow + G\_temp + A\_temp +   
## Day

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - Day 6 42.724 993.7 0.91783 0.9885  
## - Humd 1 41.812 1002.8 0.00617 0.9374  
## - Rain 1 41.815 1002.8 0.00907 0.9241  
## - Wind 1 41.897 1002.9 0.09134 0.7625  
## - Snow 1 41.970 1003.0 0.16383 0.6857  
## - A\_temp 1 42.135 1003.1 0.32937 0.5660  
## - Temp 1 43.077 1004.0 1.27151 0.2595  
## - G\_temp 1 43.438 1004.4 1.63234 0.2014  
## <none> 41.806 1004.8

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=994.67  
## Deliv\_per ~ Temp + Rain + Wind + Humd + Snow + G\_temp + A\_temp

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - Humd 1 42.725 992.67 0.00102 0.9745  
## - Rain 1 42.732 992.68 0.00857 0.9262  
## - Wind 1 42.835 992.78 0.11126 0.7387  
## - Snow 1 42.880 992.83 0.15582 0.6930  
## - A\_temp 1 43.013 992.96 0.28935 0.5906  
## - Temp 1 43.940 993.89 1.21626 0.2701  
## - G\_temp 1 44.263 994.21 1.53930 0.2147  
## <none> 42.724 994.67

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=992.65  
## Deliv\_per ~ Temp + Rain + Wind + Snow + G\_temp + A\_temp

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - Rain 1 42.732 990.65 0.00756 0.9307  
## - Wind 1 42.867 990.79 0.14196 0.7063  
## - Snow 1 42.880 990.80 0.15489 0.6939  
## - A\_temp 1 43.016 990.94 0.29098 0.5896  
## - Temp 1 44.056 991.98 1.33116 0.2486  
## - G\_temp 1 44.325 992.25 1.60015 0.2059  
## <none> 42.725 992.65

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=990.69  
## Deliv\_per ~ Temp + Wind + Snow + G\_temp + A\_temp

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - Snow 1 42.887 988.85 0.15417 0.6946  
## - Wind 1 42.897 988.86 0.16423 0.6853  
## - A\_temp 1 43.045 989.01 0.31258 0.5761  
## - Temp 1 44.149 990.11 1.41689 0.2339  
## - G\_temp 1 44.446 990.41 1.71332 0.1906  
## <none> 42.732 990.69

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=989.81  
## Deliv\_per ~ Temp + Wind + G\_temp + A\_temp

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - Wind 1 43.088 988.01 0.20126 0.6537  
## - A\_temp 1 43.202 988.13 0.31585 0.5741  
## - Temp 1 44.238 989.16 1.35198 0.2449  
## - G\_temp 1 44.555 989.48 1.66886 0.1964  
## <none> 42.887 989.81

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=988.43  
## Deliv\_per ~ Temp + G\_temp + A\_temp

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## Df Deviance AIC LRT Pr(>Chi)  
## - A\_temp 1 43.453 986.79 0.36514 0.5457  
## - Temp 1 44.336 987.67 1.24789 0.2640  
## - G\_temp 1 44.766 988.10 1.67814 0.1952  
## <none> 43.088 988.43

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

##   
## Step: AIC=990.15  
## Deliv\_per ~ Temp + G\_temp

##   
## Call: glm(formula = Deliv\_per ~ Temp + G\_temp, family = binomial(link = "logit"),   
## data = retail)  
##   
## Coefficients:  
## (Intercept) Temp G\_temp   
## -0.46578 0.04169 -0.04075   
##   
## Degrees of Freedom: 879 Total (i.e. Null); 877 Residual  
## Null Deviance: 45.03   
## Residual Deviance: 43.45 AIC: 990.1

deliv\_per.glm.A = glm(Deliv\_per ~ Temp + G\_temp, family = binomial(link = "logit"),   
 data = retail)

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

summary(deliv\_per.glm.A)

##   
## Call:  
## glm(formula = Deliv\_per ~ Temp + G\_temp, family = binomial(link = "logit"),   
## data = retail)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.90976 -0.16226 -0.01274 0.15601 0.64900   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.46578 0.13621 -3.420 0.000627 \*\*\*  
## Temp 0.04169 0.03792 1.099 0.271579   
## G\_temp -0.04075 0.03394 -1.201 0.229912   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 45.029 on 879 degrees of freedom  
## Residual deviance: 43.453 on 877 degrees of freedom  
## AIC: 990.15  
##   
## Number of Fisher Scoring iterations: 3

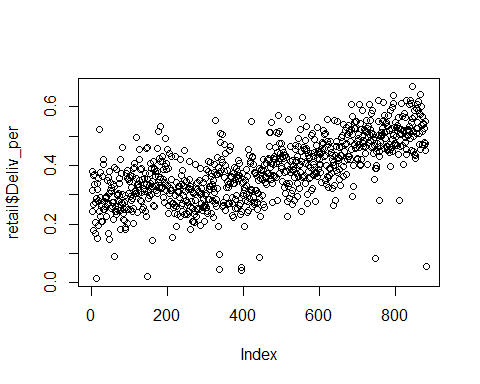
retail["isRain"] = retail["Rain"] != 0  
retail["isSnow"] = retail["Snow"] != 0  
deliv\_per.glm.B = glm(Deliv\_per~Temp+isRain+Wind+Humd+isSnow+G\_temp+A\_temp+Day, data=retail, family = binomial(link = "logit"))

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

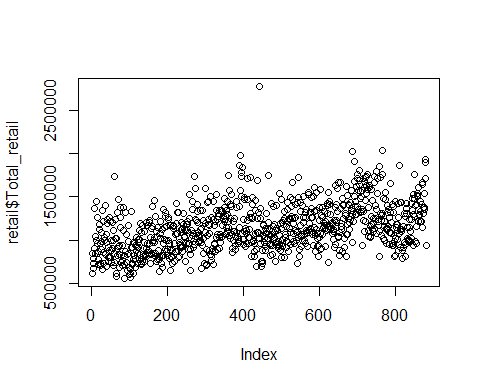
summary(deliv\_per.glm.B)

##   
## Call:  
## glm(formula = Deliv\_per ~ Temp + isRain + Wind + Humd + isSnow +   
## G\_temp + A\_temp + Day, family = binomial(link = "logit"),   
## data = retail)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.96118 -0.15463 -0.01265 0.15683 0.62822   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.615699 0.517246 -1.190 0.234  
## Temp 0.046739 0.041367 1.130 0.259  
## isRainTRUE 0.010689 0.213910 0.050 0.960  
## Wind 0.038477 0.118959 0.323 0.746  
## Humd -0.000507 0.007581 -0.067 0.947  
## isSnowTRUE 0.815361 2.058875 0.396 0.692  
## G\_temp -0.062294 0.048450 -1.286 0.199  
## A\_temp 0.021591 0.037082 0.582 0.560  
## Day목 0.018762 0.262238 0.072 0.943  
## Day수 0.051770 0.261263 0.198 0.843  
## Day월 0.043064 0.260656 0.165 0.869  
## Day일 0.179713 0.258924 0.694 0.488  
## Day토 -0.032795 0.262711 -0.125 0.901  
## Day화 -0.022538 0.261886 -0.086 0.931  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 45.029 on 879 degrees of freedom  
## Residual deviance: 41.813 on 866 degrees of freedom  
## AIC: 1004.9  
##   
## Number of Fisher Scoring iterations: 3

plot(retail$Deliv\_per)



plot(retail$Total\_retail)



retail["Index"] = seq(1,length(retail))

ind.lm = glm(Deliv\_per ~ Index, data = retail,family = binomial(link="logit"))

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

summary(ind.lm)

##   
## Call:  
## glm(formula = Deliv\_per ~ Index, family = binomial(link = "logit"),   
## data = retail)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.90179 -0.16481 -0.00842 0.16276 0.58502   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.459661 0.145372 -3.162 0.00157 \*\*  
## Index -0.002556 0.015052 -0.170 0.86513   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 45.029 on 879 degrees of freedom  
## Residual deviance: 45.000 on 878 degrees of freedom  
## AIC: 991.14  
##   
## Number of Fisher Scoring iterations: 3