SDS358_Project_RP4

Brent Bouslog

12/6/2020

Importing Data

```
mydata <- read.csv("nest_predation_dataset.csv")
```

Data Preprocessing

```
data <- mydata[c("Lat","Long","IncubationPeriod.days.", "NestlingPeriod.days.", "ClutchSize", "demeleva
data <- drop_na(data)
kable(head(data))
```

Lat	Long	IncubationPeriod.days.	NestlingPeriod.days.	ClutchSize	demelevation	temptrop2	Percent
40.7667520	-77.97765	10.5	11.0	3.5	380.32037	ntemp	
8.5666670	-67.58333	11.5	8.5	7.6	69.82164	trop	
-2.3108310	-80.83856	14.0	9.0	5.0	36.67687	trop	
-0.3591289	-80.35786	32.0	17.0	2.4	198.39531	trop	
9.3334480	-83.62788	18.0	22.5	2.0	706.53345	trop	
10.7167710	-61.30037	18.0	22.5	2.0	541.65424	trop	

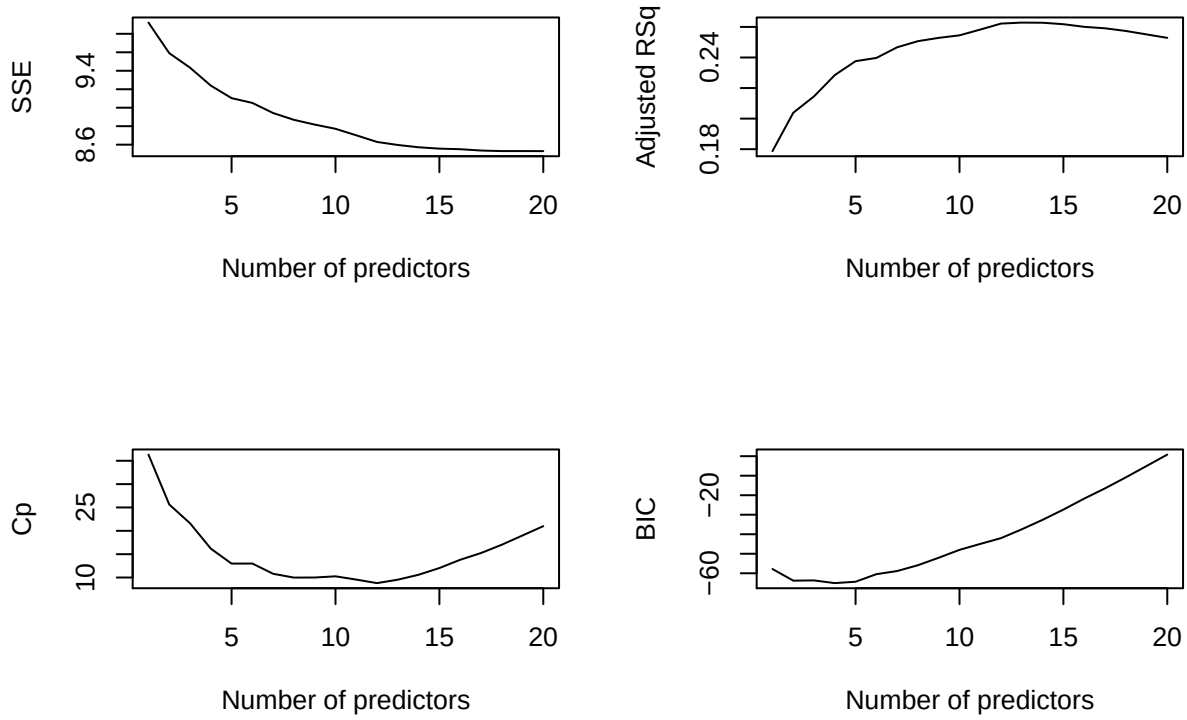
Best Subsets Regression: Multiple Linear Regression with Interaction Terms

```
data$ntemp <-ifelse(data$temptrop2 == 'ntemp', 1, 0)
data$stemp <-ifelse(data$temptrop2 == 'stemp', 1, 0)
```

```
# Calculate the squared predictor variables to include in the model and the interaction term:
data <-data %>% mutate(Lat.ntemp = Lat*ntemp, Lat.stemp = Lat*stemp, Long.ntemp = Long*ntemp, Long.stem
                       Incubate.ntemp = IncubationPeriod.days.*ntemp, Incubate.stemp = IncubationPeriod
                       Nestle.ntemp = NestlingPeriod.days.*ntemp, Nestle.stemp = NestlingPeriod.days.*s
                       ClutchSize.ntemp = ClutchSize*ntemp, ClutchSize.stemp = ClutchSize*stemp,
                       Elev.ntemp = demelevation*ntemp, Elev.stemp = demelevation*stemp)
```

```
# Find the best model for each number of predictors (with 8 predictors maximum)
models <-regsubsets(PercentageSuccessfulNests ~Lat +Lat.ntemp +Lat.stemp +Long +Long.ntemp +Long.stemp +
models.sum <-summary(models)

# Create four plots within a 2x2 frame to compare the different criteria
par(mfrow =c(2,2))
  # SSE
  plot(models.sum$rss, xlab ="Number of predictors", ylab ="SSE", type ="l")

  # R2
```

```r
plot(models.sum$adjr2, xlab ="Number of predictors", ylab ="Adjusted RSq", type ="l")

# Mallow's Cp
plot(models.sum$cp, xlab ="Number of predictors", ylab ="Cp", type ="l")

# BIC
plot(models.sum$bic, xlab ="Number of predictors", ylab ="BIC", type ="l")
```



It looks like 10 is the optimal number of predictors

```r
# Display the best model (selected predictors are indicated by *) for each number of predictors
models.sum$outmat
```

```
##           Lat Lat.ntemp Lat.stemp Long Long.ntemp Long.stemp
## 1  ( 1 )  " " " "       " "       " "  " "        " "
## 2  ( 1 )  " " " "       " "       " "  " "        " "
## 3  ( 1 )  " " "*"       " "       " "  " "        " "
## 4  ( 1 )  " " "*"       " "       " "  " "        " "
## 5  ( 1 )  " " "*"       " "       " "  " "        " "
## 6  ( 1 )  " " "*"       " "       " "  "*"        " "
## 7  ( 1 )  " " "*"       " "       " "  " "        "*"
## 8  ( 1 )  " " "*"       " "       " "  " "        " "
## 9  ( 1 )  " " "*"       " "       " "  "*"        " "
## 10  ( 1 ) " " "*"       " "       " "  " "        "*"
## 11  ( 1 ) " " "*"       " "       "*"  "*"        "*"
## 12  ( 1 ) " " "*"       " "       "*"  "*"        "*"
## 13  ( 1 ) "*" "*"       " "       "*"  "*"        "*"
```

```
## 14  ( 1 ) " " "*"       " "       "*"  "*"        "*"
## 15  ( 1 ) "*" "*"       " "       "*"  "*"        "*"
## 16  ( 1 ) "*" "*"       "*"       "*"  "*"        "*"
## 17  ( 1 ) "*" "*"       " "       "*"  "*"        "*"
## 18  ( 1 ) "*" "*"       "*"       "*"  "*"        "*"
## 19  ( 1 ) "*" "*"       "*"       "*"  "*"        "*"
## 20  ( 1 ) "*" "*"       "*"       "*"  "*"        "*"
##           IncubationPeriod.days. Incubate.ntemp Incubate.stemp
## 1  ( 1 )  " "                    " "            " "
## 2  ( 1 )  " "                    " "            " "
## 3  ( 1 )  " "                    " "            " "
## 4  ( 1 )  " "                    " "            " "
## 5  ( 1 )  "*"                    " "            " "
## 6  ( 1 )  "*"                    " "            " "
## 7  ( 1 )  "*"                    " "            " "
## 8  ( 1 )  "*"                    " "            " "
## 9  ( 1 )  "*"                    " "            " "
## 10  ( 1 ) "*"                    " "            " "
## 11  ( 1 ) "*"                    " "            " "
## 12  ( 1 ) "*"                    " "            " "
## 13  ( 1 ) "*"                    " "            " "
## 14  ( 1 ) "*"                    "*"            " "
## 15  ( 1 ) "*"                    "*"            " "
## 16  ( 1 ) "*"                    "*"            " "
## 17  ( 1 ) "*"                    "*"            " "
## 18  ( 1 ) "*"                    "*"            " "
## 19  ( 1 ) "*"                    "*"            "*"
## 20  ( 1 ) "*"                    "*"            "*"
##           NestlingPeriod.days. Nestle.ntemp Nestle.stemp ClutchSize
## 1  ( 1 )  " "                  "*"          " "          " "
## 2  ( 1 )  " "                  "*"          " "          "*"
## 3  ( 1 )  " "                  "*"          " "          " "
## 4  ( 1 )  " "                  "*"          " "          "*"
## 5  ( 1 )  " "                  "*"          " "          "*"
## 6  ( 1 )  " "                  "*"          " "          "*"
## 7  ( 1 )  " "                  "*"          "*"          "*"
## 8  ( 1 )  " "                  "*"          "*"          "*"
## 9  ( 1 )  " "                  "*"          "*"          "*"
## 10  ( 1 ) " "                  "*"          "*"          "*"
## 11  ( 1 ) " "                  "*"          "*"          "*"
## 12  ( 1 ) " "                  "*"          "*"          "*"
## 13  ( 1 ) " "                  "*"          "*"          "*"
## 14  ( 1 ) " "                  "*"          "*"          "*"
## 15  ( 1 ) " "                  "*"          "*"          "*"
## 16  ( 1 ) " "                  "*"          "*"          "*"
## 17  ( 1 ) " "                  "*"          "*"          "*"
## 18  ( 1 ) " "                  "*"          "*"          "*"
## 19  ( 1 ) " "                  "*"          "*"          "*"
## 20  ( 1 ) "*"                  "*"          "*"          "*"
##           ClutchSize.ntemp ClutchSize.stemp demelevation Elev.ntemp Elev.stemp
## 1  ( 1 )  " "              " "             " "          " "         " "
## 2  ( 1 )  " "              " "             " "          " "         " "
## 3  ( 1 )  " "              " "             " "          " "         " "
## 4  ( 1 )  " "              " "             " "          " "         " "
```

```
## 5  ( 1 )  " "              " "              " "          " "         " "
## 6  ( 1 )  " "              " "              " "          " "         " "
## 7  ( 1 )  " "              " "              " "          " "         " "
## 8  ( 1 )  " "              "*"              " "          " "         "*"
## 9  ( 1 )  " "              "*"              " "          " "         "*"
## 10  ( 1 ) " "              "*"              " "          " "         "*"
## 11  ( 1 ) " "              "*"              " "          " "         "*"
## 12  ( 1 ) "*"              "*"              " "          " "         "*"
## 13  ( 1 ) "*"              "*"              " "          " "         "*"
## 14  ( 1 ) "*"              "*"              " "          " "         "*"
## 15  ( 1 ) "*"              "*"              " "          " "         "*"
## 16  ( 1 ) "*"              "*"              " "          " "         "*"
## 17  ( 1 ) "*"              "*"              "*"          "*"         "*"
## 18  ( 1 ) "*"              "*"              "*"          "*"         "*"
## 19  ( 1 ) "*"              "*"              "*"          "*"         "*"
## 20  ( 1 ) "*"              "*"              "*"          "*"         "*"
##           ntemp stemp
## 1  ( 1 )  " "   " "
## 2  ( 1 )  " "   " "
## 3  ( 1 )  "*"   " "
## 4  ( 1 )  "*"   " "
## 5  ( 1 )  "*"   " "
## 6  ( 1 )  "*"   " "
## 7  ( 1 )  "*"   " "
## 8  ( 1 )  "*"   " "
## 9  ( 1 )  "*"   " "
## 10  ( 1 ) "*"   "*"
## 11  ( 1 ) " "   "*"
## 12  ( 1 ) " "   "*"
## 13  ( 1 ) " "   "*"
## 14  ( 1 ) "*"   "*"
## 15  ( 1 ) "*"   "*"
## 16  ( 1 ) "*"   "*"
## 17  ( 1 ) "*"   "*"
## 18  ( 1 ) "*"   "*"
## 19  ( 1 ) "*"   "*"
## 20  ( 1 ) "*"   "*"
```

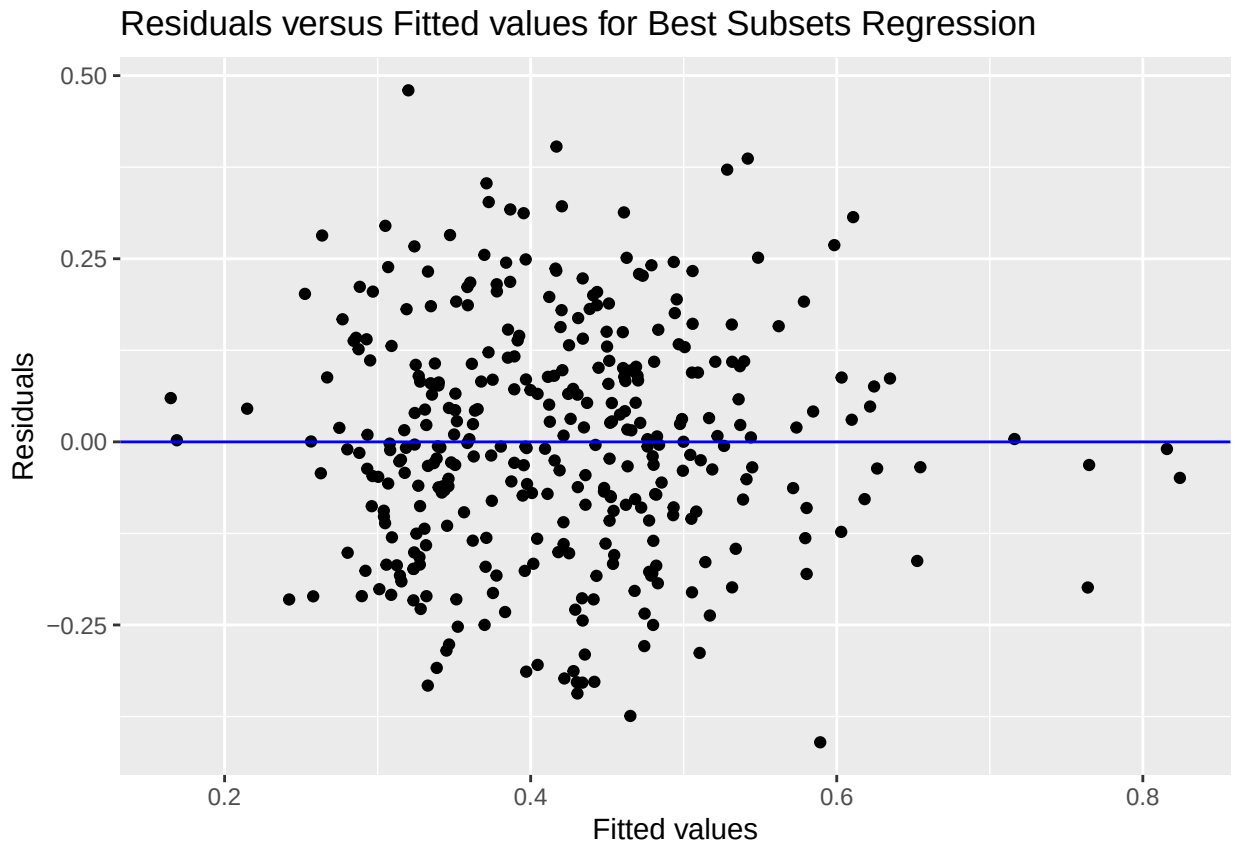
Including best 10 predictors, but maintaining hierarchy principle as well.

```
# Creating a model with the 6 predictors indicated as the best by the Best Subsets Regression
reg_sub <-lm(PercentageSuccessfulNests ~Lat +Lat.ntemp +Long +Long.stemp +IncubationPeriod.days. +Nestli

summary(reg_sub)
```

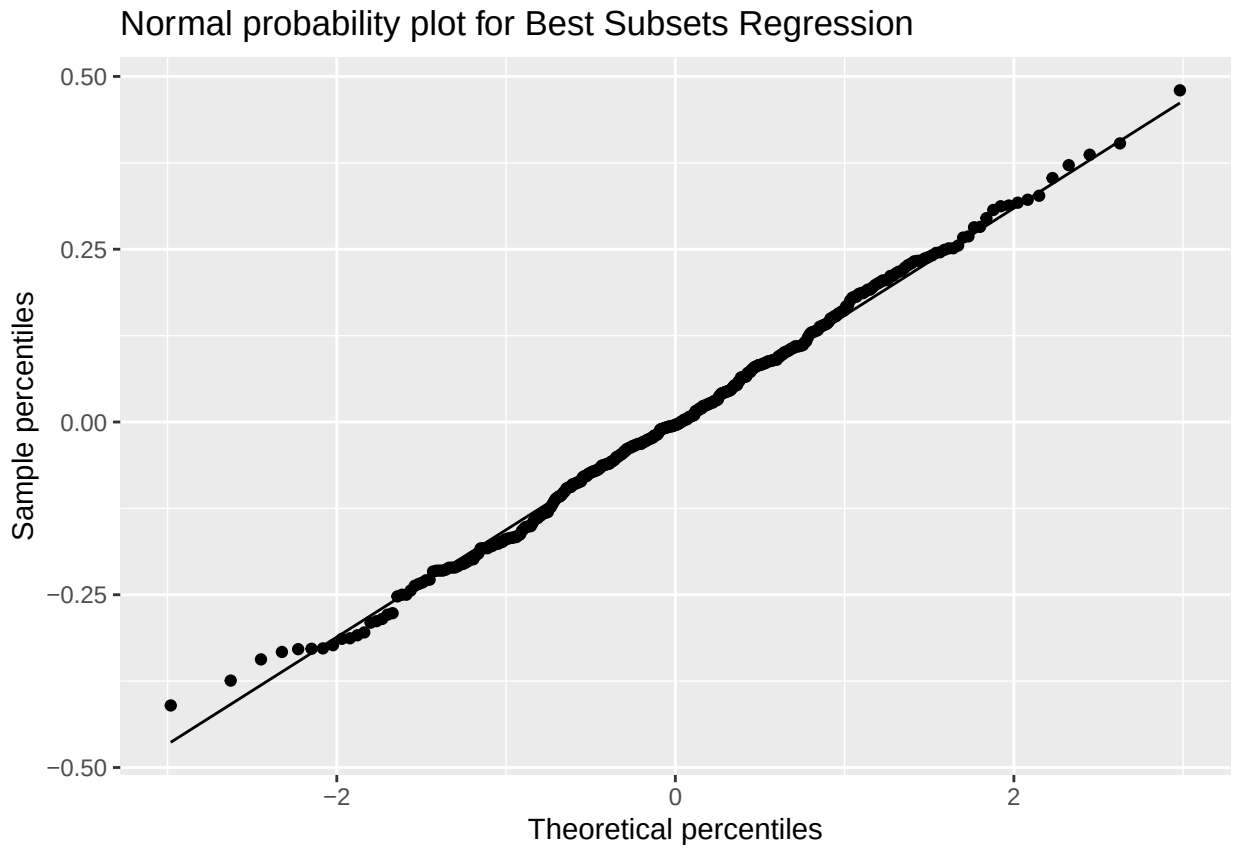
```
##
## Call:
## lm(formula = PercentageSuccessfulNests ~ Lat + Lat.ntemp + Long +
##     Long.stemp + IncubationPeriod.days. + NestlingPeriod.days. +
##     Nestle.ntemp + Nestle.stemp + ClutchSize + ClutchSize.stemp +
##     demelevation + Elev.stemp + ntemp + stemp, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41020 -0.10560 -0.00407  0.10373  0.47992
```

```
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              4.193e-01  9.006e-02   4.656 4.67e-06 ***
## Lat                      1.038e-03  1.456e-03   0.713 0.476327
## Lat.ntemp                5.043e-03  2.336e-03   2.159 0.031565 *
## Long                     1.638e-04  7.542e-04   0.217 0.828164
## Long.stemp               7.059e-03  4.838e-03   1.459 0.145536
## IncubationPeriod.days. -1.045e-02  4.235e-03  -2.468 0.014105 *
## NestlingPeriod.days.     1.619e-03  2.990e-03   0.542 0.588466
## Nestle.ntemp             1.569e-02  4.202e-03   3.734 0.000222 ***
## Nestle.stemp             2.001e-02  1.071e-02   1.869 0.062551 .
## ClutchSize               2.981e-02  1.087e-02   2.743 0.006413 **
## ClutchSize.stemp        -1.198e-01  5.553e-02  -2.158 0.031674 *
## demelevation            -8.282e-06  1.584e-05  -0.523 0.601349
## Elev.stemp              -2.708e-04  1.410e-04  -1.921 0.055526 .
## ntemp                   -3.588e-01  9.941e-02  -3.609 0.000354 ***
## stemp                    6.001e-01  3.197e-01   1.877 0.061430 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.162 on 333 degrees of freedom
## Multiple R-squared:  0.2783, Adjusted R-squared:  0.248
## F-statistic: 9.172 on 14 and 333 DF,  p-value: < 2.2e-16
```

```r
# Fit the model obtained from forward selection
data$resids <-residuals(reg_sub)
data$predicted <-predict(reg_sub)

ggplot(data, aes(x=predicted, y=resids)) +
  geom_point() +
  geom_hline(yintercept=0, color ="blue") +
  labs(title ="Residuals versus Fitted values for Best Subsets Regression", x ="Fitted values", y ="Resi
```



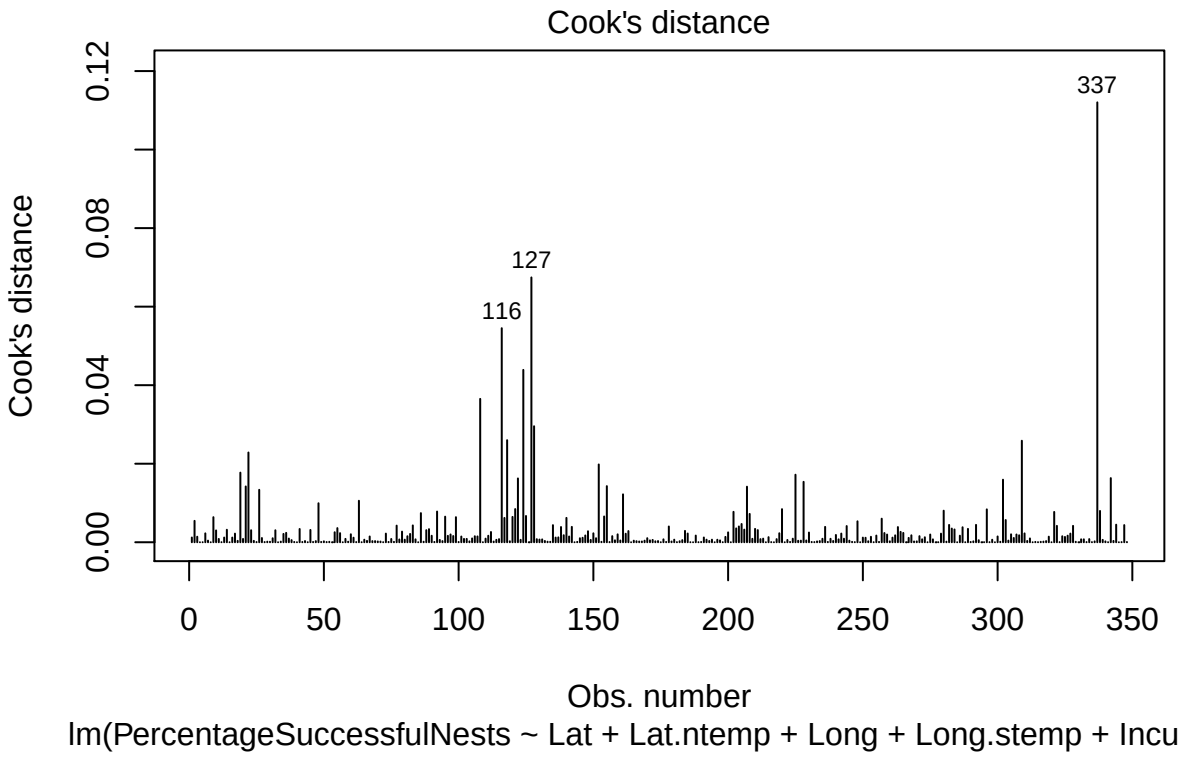

```
ggplot(data, aes(sample = resids)) +
  stat_qq() +
  stat_qq_line() +
  labs(title ="Normal probability plot for Best Subsets Regression", x ="Theoretical percentiles", y ="S
```




```
# ANOVA Table
anova(reg_sub)
```

```
## Analysis of Variance Table
##
## Response: PercentageSuccessfulNests
##                          Df Sum Sq Mean Sq F value    Pr(>F)
## Lat                       1 1.6357 1.63565 62.2947 4.311e-14 ***
## Lat.ntemp                 1 0.4266 0.42664 16.2488 6.889e-05 ***
## Long                      1 0.0117 0.01167  0.4445 0.5054124
## Long.stemp                1 0.0096 0.00961  0.3660 0.5456303
## IncubationPeriod.days.    1 0.0024 0.00244  0.0929 0.7607629
## NestlingPeriod.days.      1 0.3216 0.32162 12.2492 0.0005288 ***
## Nestle.ntemp              1 0.0797 0.07967  3.0344 0.0824403 .
## Nestle.stemp              1 0.1082 0.10822  4.1216 0.0431347 *
## ClutchSize                1 0.2117 0.21173  8.0638 0.0047935 **
## ClutchSize.stemp          1 0.0259 0.02592  0.9871 0.3211689
## demelevation              1 0.0377 0.03772  1.4367 0.2315290
## Elev.stemp                1 0.0363 0.03634  1.3842 0.2402296
## ntemp                     1 0.3720 0.37197 14.1666 0.0001977 ***
## stemp                     1 0.0925 0.09248  3.5221 0.0614300 .
## Residuals               333 8.7435 0.02626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Identifying outliers with Cook's distance
plot(reg_sub, which=4, cook.levels=cutoff)
```



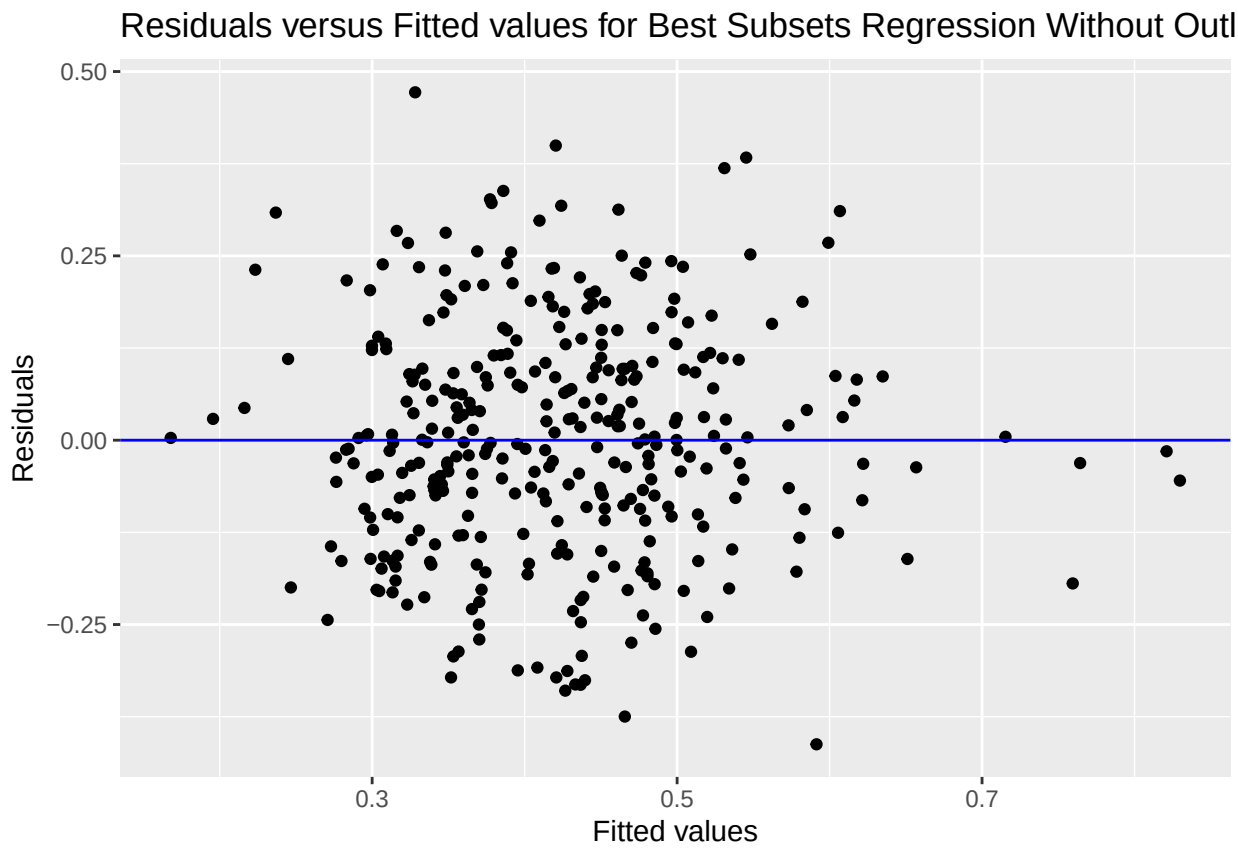

```
# If you identify an outlier, remove it by indexing the corresponding row
data_no_out <- data[-116,]
data_no_out <- data_no_out[-127,]
data_no_out <- data_no_out[-337,]
```

```
# Fit the regression model
reg_sub2 <-lm(PercentageSuccessfulNests ~Lat +Lat.ntemp +Long +Long.stemp +IncubationPeriod.days. +Nestl

# Display the summary table for the regression model
summary(reg_sub2)
```

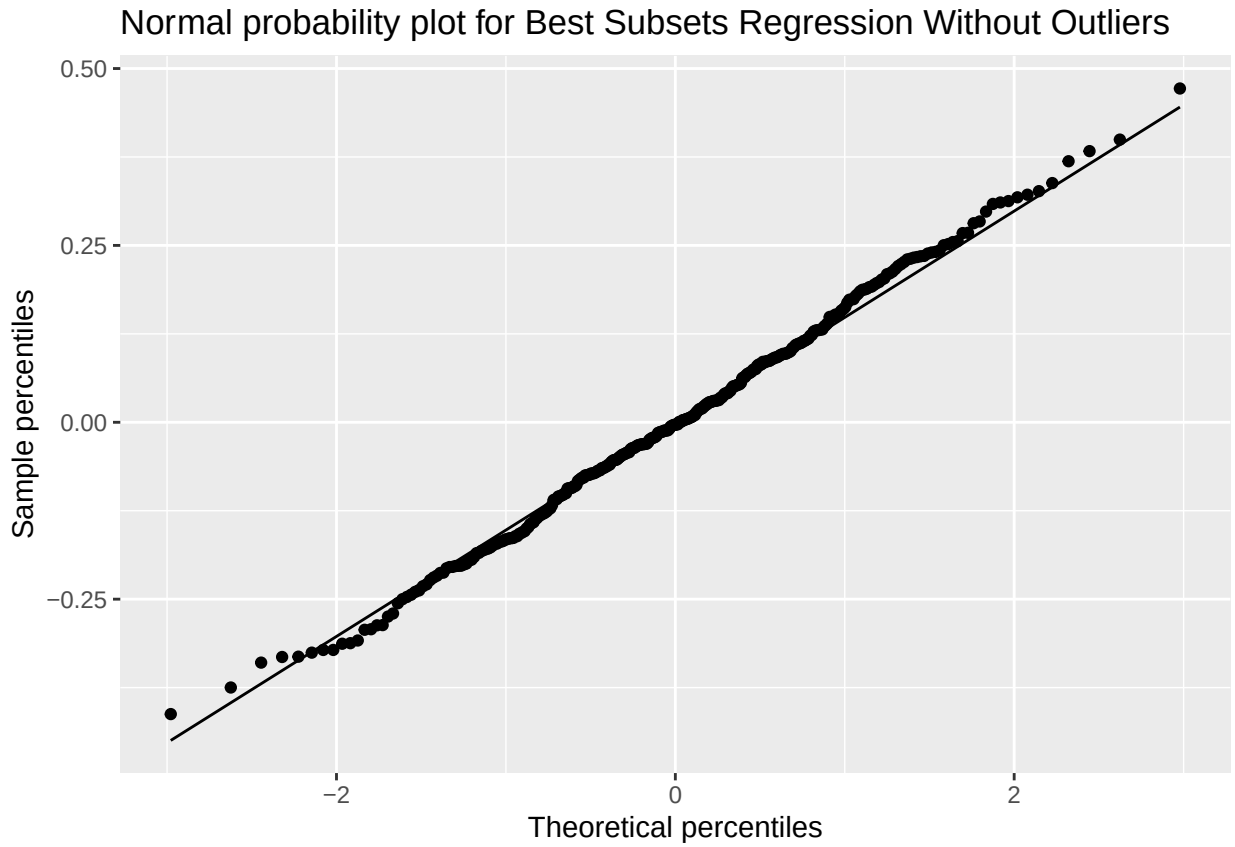
```
##
## Call:
## lm(formula = PercentageSuccessfulNests ~ Lat + Lat.ntemp + Long +
##     Long.stemp + IncubationPeriod.days. + NestlingPeriod.days. +
##     Nestle.ntemp + Nestle.stemp + ClutchSize + ClutchSize.stemp +
##     demelevation + Elev.stemp + ntemp + stemp, data = data_no_out)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41244 -0.10345 -0.00346  0.09930  0.47179
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.000e-01  8.970e-02   4.460 1.13e-05 ***
## Lat                     5.166e-04  1.461e-03   0.354  0.72389
## Lat.ntemp               5.465e-03  2.331e-03   2.345  0.01963 *
```

```
## Long                    1.086e-04  7.491e-04   0.145  0.88482
## Long.stemp              5.155e-03  5.019e-03   1.027  0.30520
## IncubationPeriod.days. -1.118e-02  4.231e-03  -2.641  0.00865 **
## NestlingPeriod.days.    3.966e-03  3.155e-03   1.257  0.20965
## Nestle.ntemp            1.358e-02  4.266e-03   3.183  0.00160 **
## Nestle.stemp            1.750e-02  1.068e-02   1.639  0.10217
## ClutchSize              2.898e-02  1.081e-02   2.681  0.00770 **
## ClutchSize.stemp       -1.371e-01  5.636e-02  -2.432  0.01553 *
## demelevation           -1.281e-05  1.583e-05  -0.809  0.41900
## Elev.stemp             -2.576e-04  1.400e-04  -1.840  0.06669 .
## ntemp                  -3.270e-01  9.958e-02  -3.284  0.00113 **
## stemp                   5.593e-01  3.205e-01   1.745  0.08186 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1609 on 330 degrees of freedom
## Multiple R-squared:  0.2769, Adjusted R-squared:  0.2462
## F-statistic: 9.024 on 14 and 330 DF,  p-value: < 2.2e-16
```

```r
# Fit the model obtained from forward selection
data_no_out$resids <-residuals(reg_sub2)
data_no_out$predicted <-predict(reg_sub2)

ggplot(data_no_out, aes(x=predicted, y=resids)) +
  geom_point() +
  geom_hline(yintercept=0, color ="blue") +
  labs(title ="Residuals versus Fitted values for Best Subsets Regression Without Outliers", x ="Fitted
```



```
ggplot(data_no_out, aes(sample = resids)) +
  stat_qq() +
  stat_qq_line() +
  labs(title ="Normal probability plot for Best Subsets Regression Without Outliers", x ="Theoretical p
```




```
# ANOVA Table
anova(reg_sub2)
```

```
## Analysis of Variance Table
##
## Response: PercentageSuccessfulNests
##                          Df Sum Sq Mean Sq F value    Pr(>F)
## Lat                       1 1.4908 1.49080 57.6027 3.312e-13 ***
## Lat.ntemp                 1 0.4888 0.48877 18.8855 1.851e-05 ***
## Long                      1 0.0131 0.01313  0.5073 0.4768288
## Long.stemp                1 0.0106 0.01059  0.4092 0.5228400
## IncubationPeriod.days.    1 0.0009 0.00088  0.0340 0.8538759
## NestlingPeriod.days.      1 0.4371 0.43713 16.8902 5.002e-05 ***
## Nestle.ntemp              1 0.0427 0.04272  1.6508 0.1997549
## Nestle.stemp              1 0.0719 0.07187  2.7768 0.0965867 .
## ClutchSize                1 0.1835 0.18348  7.0893 0.0081345 **
## ClutchSize.stemp          1 0.0559 0.05587  2.1587 0.1427139
## demelevation              1 0.0574 0.05736  2.2163 0.1375121
## Elev.stemp                1 0.0338 0.03376  1.3045 0.2542157
## ntemp                     1 0.3046 0.30455 11.7674 0.0006793 ***
## stemp                     1 0.0788 0.07884  3.0462 0.0818569 .
```

```
## Residuals              330 8.5407 0.02588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```