Croq Pain Restaurants

Day Group 4

Executive Summary RE: Croq'Pain Store Selection Recommendation
July 1995

Dear Michael Boutillon,

After conducting a thorough analysis of the 60 current Croq'Pain stores, we have refined your regression model in a manner that we believe will better serve you to predict the earnings of new potential store locations.

Our goal was two-fold:

1) Create a model that would use the optimal variables to make accurate predictions about the earnings of the existing store locations.
2) Make highly accurate predictions about the earnings of future store locations.

To guide our analysis, we made the following assumptions:

a) Earnings normalized with respect to the total population of the locality is the key factor.

b) A target performance ratio of 26% will be the threshold for our location recommendations.

c) The data obtained for the first 60 stores is representative of future locations.

Recommendations

In your case, there are many variables that could influence the earnings of your 60 stores. In order for our team to make an accurate prediction about the earnings of future locations, each of the variables provided in your model must be considered. A linear regression model is the most decisive tool to refine these variables and confidently establish which are the most significant in predicting earnings.

The regression analysis performed on the basis of the available data indicated that the size of the store, average income in locality, monthly rent per square meter in the same locale, number of 35-44 years old in the 3 km radius around the site, number of competitors in the 1 km radius, and the number of non-restaurant businesses in 1 km radius were dominant in predicting the projected earnings for the prospective location.

Based on this model, our recommendation of stores suitable to open are: 64 and 69 are as they are expected to perform beyond the targeted performance ratio of 26%.

Strengths of the Model

The model we adopted after a few iterations (See `Part A` below) accounts for factors that are correlated (such as Total and P15 - P55). As a result, we have removed such distracting variables. The R^2 value, 69.6%, serves to provide insights to the predicting power of our model. This means that 69.6% of the response variable variation can be explained by our linear model. P-values, which are measures of deciding whether the variables are statistically significant, are mostly less than 0.05, indicating that we are confident 95% of the time we will obtain test results at least as extreme as the results actually observed. These values suggest the regression model is relatively accurate in predicting unknown factors that are related to our data.

Potential Drawbacks of the Model

Potential weaknesses of using a linear regression that we discovered in our analysis were that given only 60 stores, the data may not be representative of the true significance of the predictors and may lead to biases. The factors that were not found to be significant in our regression model may prove to be significant once the number of restaurants increases. In addition, there may exist variables in the data that might not have been accounted for, but have a significant effect on earnings. However, our recommendation that you follow through with the adoption of our model still stands and we see it as the most viable option in predicting future earnings.

Please find below our detailed analysis and linear model iterations.

Kind regards,
Day Group 4

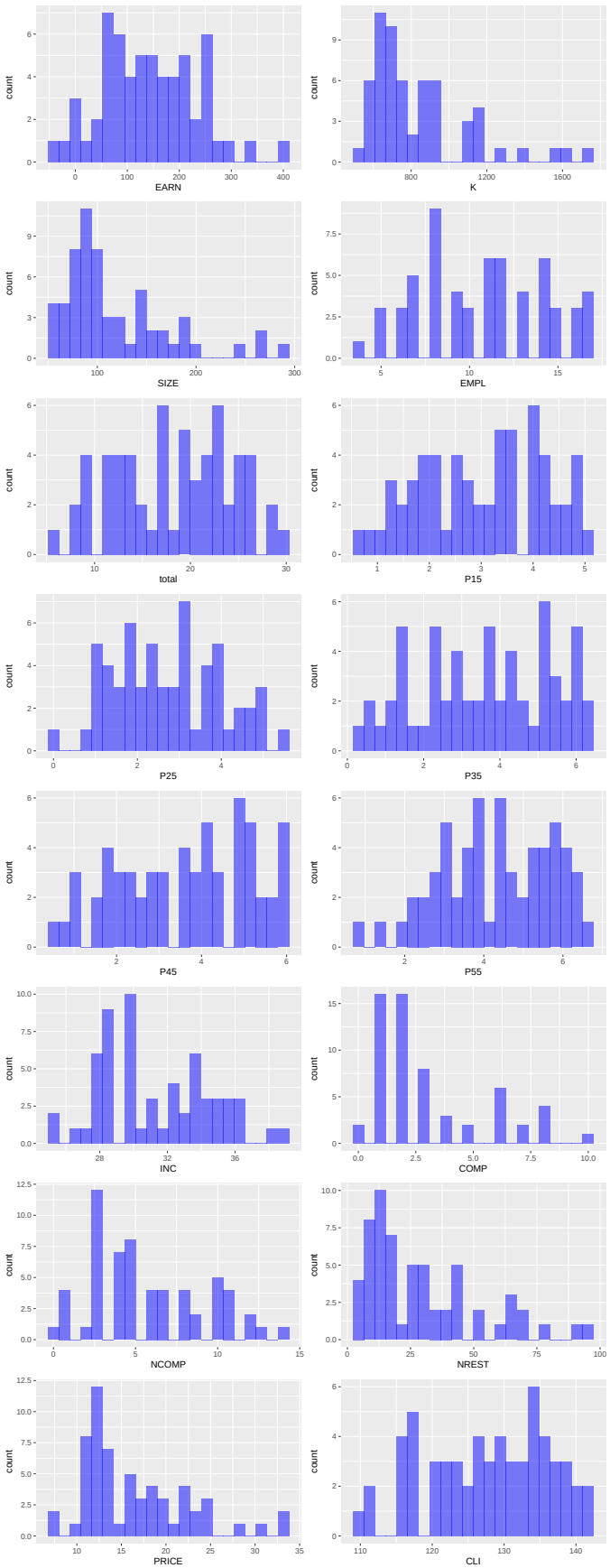
Before We Start Using the provided data, our goal is to find a model which will best help us predict the performance ratio and give Jean Gerard a list of stores which will likely succeed using our model.

Taking the look at the data, the first store has a seemingly anomalous data point. We will start out by changing the outlier to 28.339, citing user entry error. This base dataset we will be looking at will have the data for 60 stores.

Since the goal of the exercise is to make predictions of **EARN** for stores that have not yet been opened and that do not have **EARN** data, we will create a separate data set with the last 10 entries, which are missing **EARN** data.

Part A Visualize the data

Histograms for the dataset.

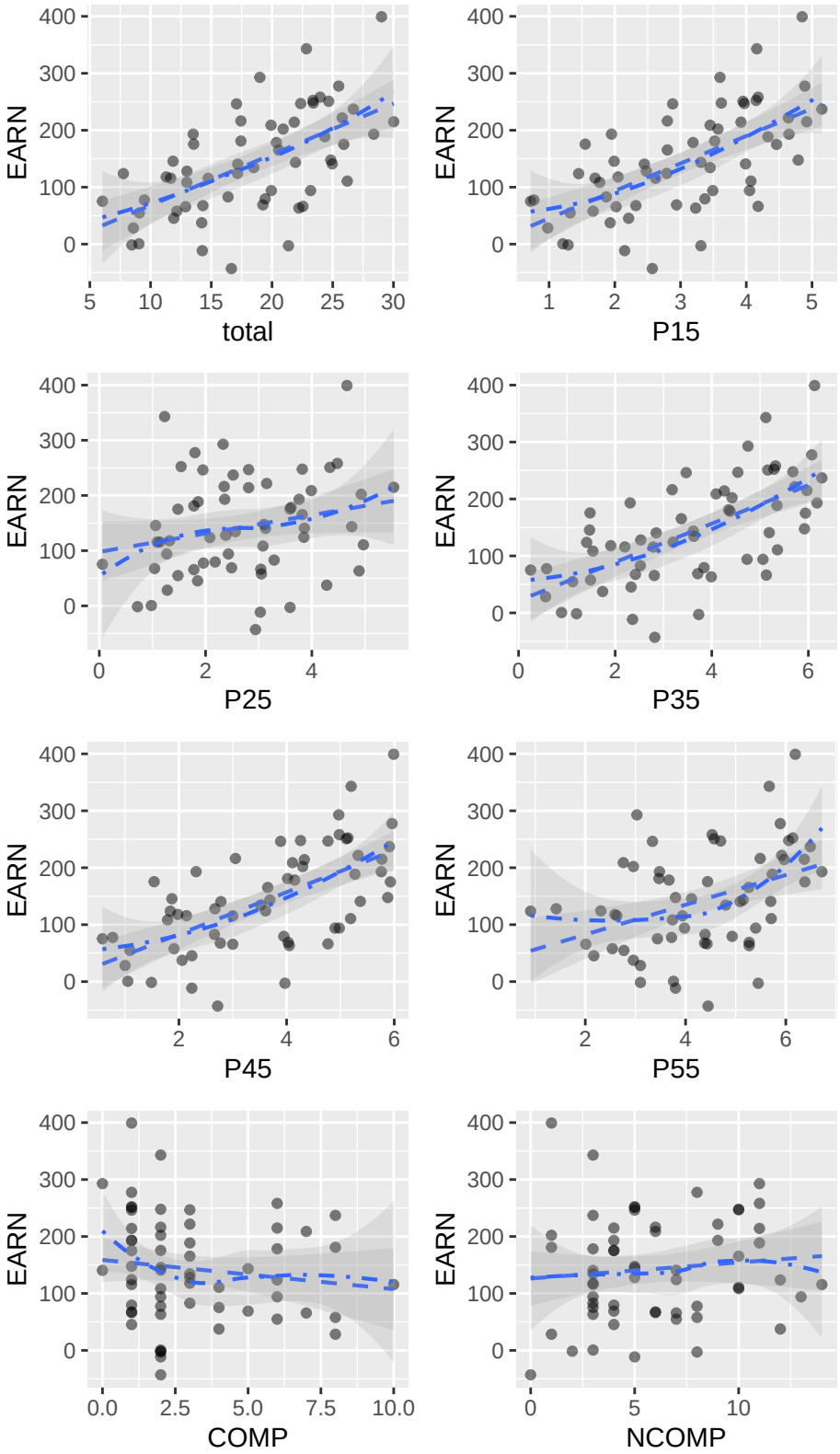

count
EARN
K
SIZE
EMPL
total
P15
P25
P35
P45
P55
INC
COMP
NCOMP
NREST
PRICE
CLI

The histograms above just give us a general idea of the distribution of the data points, but some plots are of interest.

1) `EARNINGS`: The distribution is fairly normal, indicating the usability of this without transformation, but we will continue to look at the scatter plots as well.
2) `SIZE` & `CLI` are skewed oppositely, indicating that while most of the stores are smaller is area, the `CLI` is fairly high across the data set.
3) `SIZE` & `PRICE` are skewed similarly, for the area of the stores and the rents for store spaces.
4) Population data does not follow a normal distribution, indicating we may have to make some changes before it's fully usable.

Overall, the distributions do not follow a normal trend because of the low number of data points we actually have available.

Scatterplots for P15 through P55 & competitions.



The scatterplots above show the populations against earnings, as well as the existence of stores that are competition on earnings. When looking at the population & total against earnings, we can observe a similar trend, which is understandable since the populations are a subset of the total population in a 3 km radius around the site.

Thus, these correlations make sense.

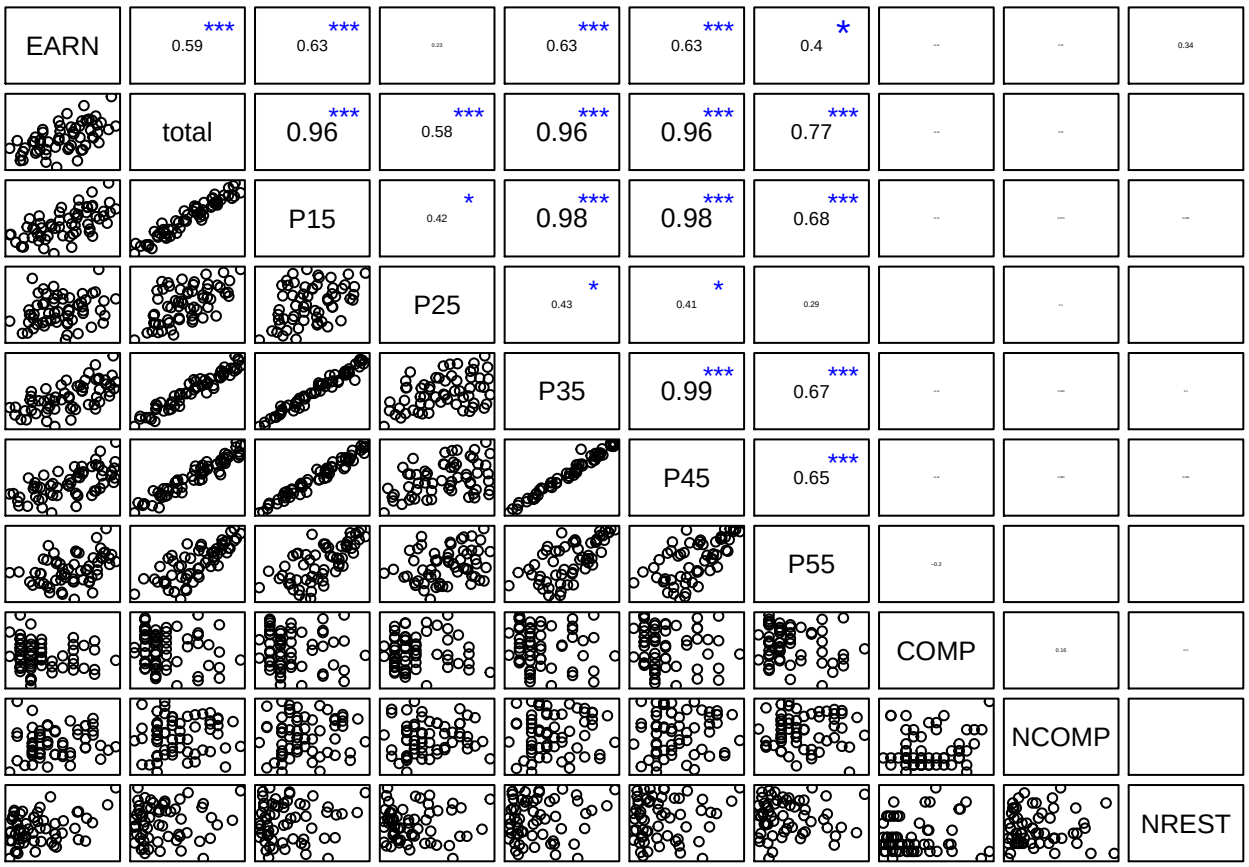
1) The correlations for `total` and `P15` through `P55` are directly correlated since the population of each age group has to be correlated with the total population in the 3 km radius around the site.
2) `P25` and `P55` look a little different from the rest (`15`, `35`, and `45`) of the scatterplots.

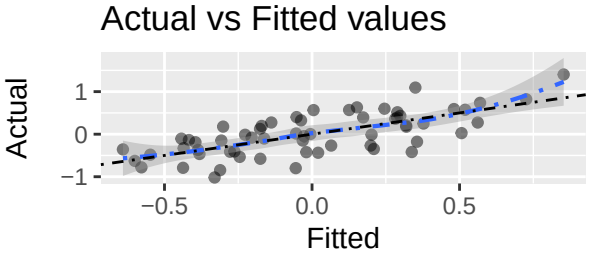
We will take a quantitative look at the correlations below.

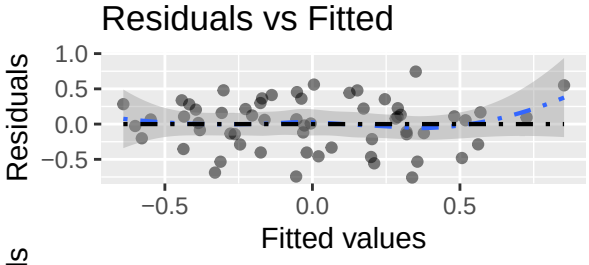
To account for the linearity we observed above, we will normalize the variables above by the total.

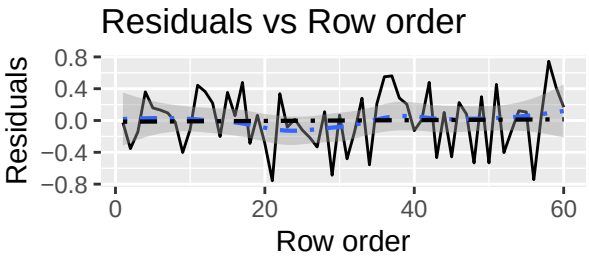
NON-TRANSFORMED data.

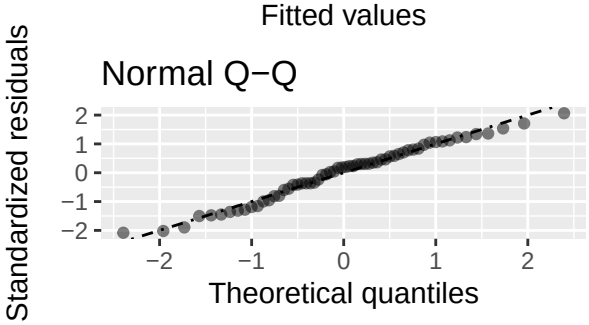
```
Linear regression (OLS)
Data     : CroqPainFix
Response variable    : EARN
Explanatory variables: total, P15, P25, P35, P45, P55, COMP, NCOMP, NREST
Null hyp.: the effect of x on EARN is zero
Alt. hyp.: the effect of x on EARN is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)      -0.000     0.049  -0.000   1.000
 total            -0.264     1.543  -0.171   0.865
 P15               0.581     0.637   0.911   0.366
 P25               0.014     0.338   0.041   0.968
 P35               0.295     0.752   0.392   0.697
 P45              -0.046     1.006  -0.046   0.964
 P55               0.017     0.390   0.043   0.966
 COMP             -0.119     0.105  -1.131   0.263
 NCOMP             0.081     0.104   0.780   0.439
 NREST             0.299     0.104   2.881   0.006 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.502,  Adjusted R-squared: 0.412
F-statistic: 5.596 df(9,50), p.value < .001
Nr obs: 60

Prediction error (RMSE):  0.35
Residual st.dev   (RSD):  0.383

Variance Inflation Factors
          VIF   Rsq
total 238.867 0.996
P45   101.481 0.990
P35    56.792 0.982
P15    40.738 0.975
P55    15.255 0.934
P25    11.485 0.913
COMP    1.108 0.097
NREST   1.082 0.076
NCOMP   1.077 0.072
```

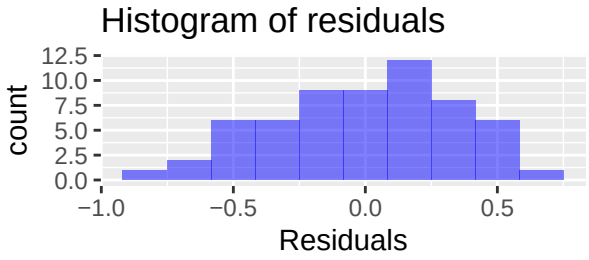


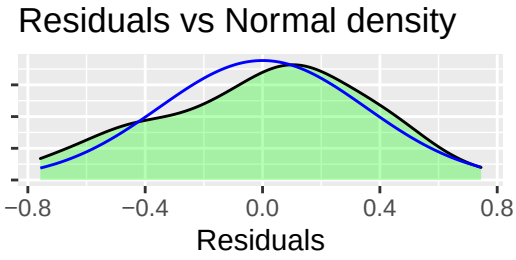






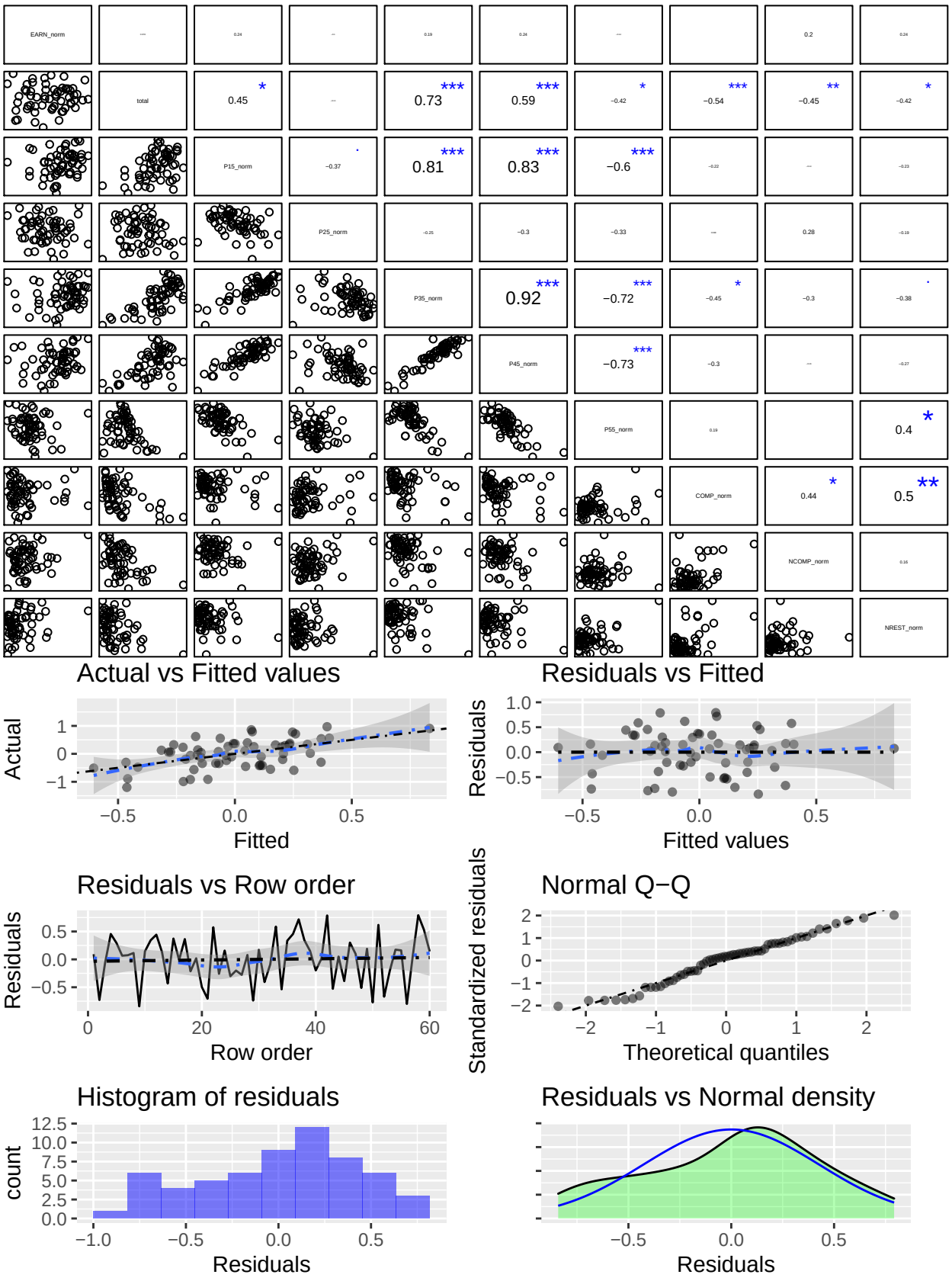






NORMALIZED DATA

```
Linear regression (OLS)
Data     : CroqPainFix
Response variable    : EARN_norm
Explanatory variables: total, P15_norm, P25_norm, P35_norm, P45_norm, P55_norm, COMP_norm, NCOMP_norm, N
Null hyp.: the effect of x on EARN_norm is zero
Alt. hyp.: the effect of x on EARN_norm is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)      -0.000     0.060  -0.000   1.000
 total             0.149     0.220   0.678   0.501
 P15_norm          0.173     0.246   0.700   0.487
 P25_norm         -0.074     0.286  -0.260   0.796
 P35_norm          0.082     0.450   0.181   0.857
 P45_norm         -0.111     0.435  -0.255   0.800
 P55_norm         -0.167     0.391  -0.426   0.672
 COMP_norm        -0.270     0.167  -1.619   0.112
 NCOMP_norm        0.364     0.155   2.351   0.023 *
 NREST_norm        0.476     0.156   3.060   0.004 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.272,  Adjusted R-squared: 0.141
F-statistic: 2.074 df(9,50), p.value 0.05
Nr obs: 60

Prediction error (RMSE):  0.423
Residual st.dev   (RSD):  0.463

Variance Inflation Factors
              VIF   Rsq
P35_norm   13.886 0.928
P45_norm   13.012 0.923
P55_norm   10.512 0.905
P25_norm    5.605 0.822
P15_norm    4.168 0.760
total       3.313 0.698
COMP_norm   1.916 0.478
NREST_norm  1.663 0.399
NCOMP_norm  1.644 0.392
```



The regressions above are at first glance, and take into account all the variables. Of course, this is not an

adequate model, and we will have to refine the model to better fit our data set.

But after initial observations, the preferred regression would be with the normalized data because:

1) Only 33% of the normalized variables have multicollenearity & high VIF, as opposed to 66% of the non-transformed data.
2) Higher VIF values make the variables effect on the model less trustworthy.
3) Looking at the collinearity graph, there is less quantititave and qualitative linearity amongst the normalized variables.

When looking at the model with the normalized data, there are only 2 suspicious variables, with VIF values slightly greater than 10 (`P45_norm` and `P35_norm`). We can attempt to fix this in our model and interpretation.

No evidence of heteroscedasticity for the non-transformed & normalized data looking at the Residuals vs. Fitted plot from the dashboard.

Regression Model: Part A

```
Linear regression (OLS)
Data     : CroqPainFix
Filter   : STOR <= 60
Response variable    : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P35_norm, COMP_norm, NREST_norm
Null hyp.: the effect of x on EARN_norm is zero
Alt. hyp.: the effect of x on EARN_norm is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)       0.000     0.038   0.000   1.000
 SIZE              0.514     0.079   6.538  < .001 ***
 INC               0.400     0.078   5.106  < .001 ***
 PRICE            -0.248     0.080  -3.114   0.003 **
 P35_norm          0.296     0.090   3.296   0.002 **
 COMP_norm        -0.095     0.096  -0.983   0.330
 NREST_norm        0.501     0.090   5.553  < .001 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.696,  Adjusted R-squared: 0.662
F-statistic: 20.262 df(6,53), p.value < .001
Nr obs: 60

Prediction error (RMSE):  0.273
Residual st.dev   (RSD):  0.291

Variance Inflation Factors
    COMP_norm NREST_norm P35_norm PRICE  SIZE   INC
VIF     1.623      1.423    1.413 1.111 1.078 1.074
Rsq     0.384      0.297    0.292 0.100 0.072 0.069
```

Since the majority of the population surveyed (55%) were between the ages of 25 and 35, the population variable being used is `P35_norm`. Since the goal is to also open more restaurants in business parks and in downtown, this is the target audience for whom we are attempting to build this model.

The variables we will be using to build our model are as follows: `SIZE`, `INC`, `PRICE`, `P35_norm`, `COMP_norm`, `NREST_norm`. The R^2 value is 0.696 for our model, and we have confirmed no collinearity between the

variables.

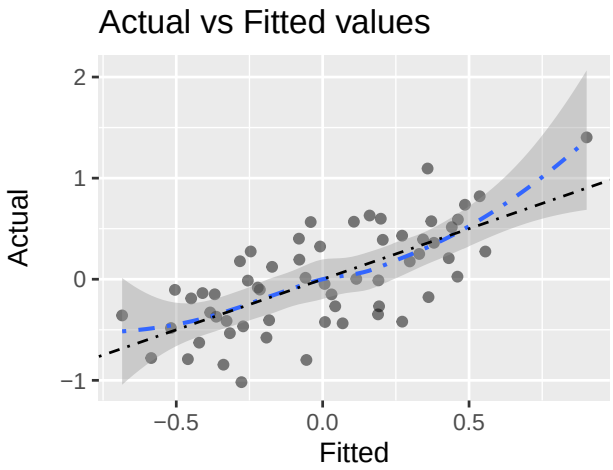
We will now use the **Step-wise** function to confirm the validity of our model.

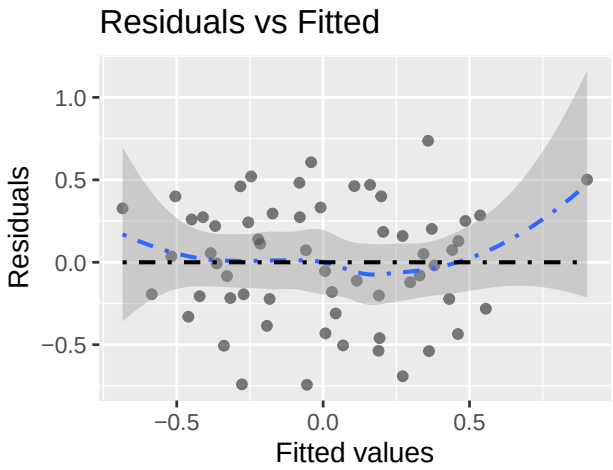
Step-wise for NON-TRANSFORMED data.

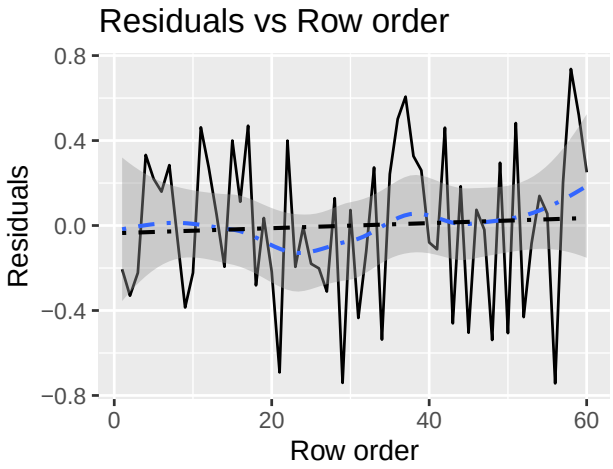
```
Start:  AIC=-105.99
EARN ~ total + P15 + P25 + P35 + P45 + P55 + COMP + NCOMP + NREST

        Df Sum of Sq    RSS      AIC
- P25    1   0.00024 7.3482 -107.993
- P55    1   0.00027 7.3483 -107.993
- P45    1   0.00030 7.3483 -107.992
- total  1   0.00431 7.3523 -107.960
- P35    1   0.02259 7.3706 -107.811
- NCOMP  1   0.08948 7.4375 -107.269
- P15    1   0.12207 7.4701 -107.006
- COMP   1   0.18814 7.5361 -106.478
<none>               7.3480 -105.995
- NREST  1   1.21962 8.5676  -98.781

Step:  AIC=-107.99
EARN ~ total + P15 + P35 + P45 + P55 + COMP + NCOMP + NREST

        Df Sum of Sq    RSS     AIC
- P55    1   0.00003 7.3483 -109.99
- P45    1   0.00155 7.3498 -109.98
- P35    1   0.02242 7.3707 -109.81
- total  1   0.02368 7.3719 -109.80
- NCOMP  1   0.09371 7.4420 -109.23
- P15    1   0.13386 7.4821 -108.91
- COMP   1   0.18814 7.5364 -108.48
<none>               7.3482 -107.99
- NREST  1   1.23015 8.5784 -100.71

Step:  AIC=-109.99
EARN ~ total + P15 + P35 + P45 + COMP + NCOMP + NREST

        Df Sum of Sq    RSS     AIC
- P45    1   0.00186 7.3501 -111.98
- P35    1   0.02239 7.3707 -111.81
- total  1   0.03820 7.3865 -111.68
- NCOMP  1   0.09368 7.4420 -111.23
- P15    1   0.13481 7.4831 -110.90
- COMP   1   0.19367 7.5419 -110.43
<none>               7.3483 -109.99
- NREST  1   1.23012 8.5784 -102.71

Step:  AIC=-111.98
EARN ~ total + P15 + P35 + COMP + NCOMP + NREST

        Df Sum of Sq    RSS     AIC
- P35    1   0.02452 7.3747 -113.78
- total  1   0.03949 7.3896 -113.66
- NCOMP  1   0.09199 7.4421 -113.23
```

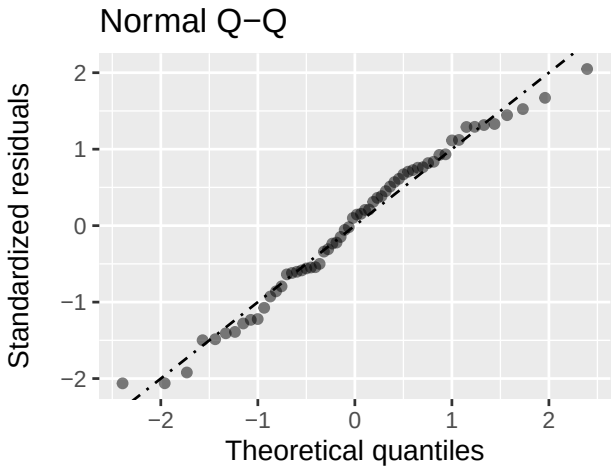
```
- P15    1   0.14000 7.4901 -112.84
- COMP   1   0.19208 7.5422 -112.43
<none>               7.3501 -111.98
- NREST  1   1.22826 8.5784 -104.71

Step:  AIC=-113.78
EARN ~ total + P15 + COMP + NCOMP + NREST

        Df Sum of Sq    RSS     AIC
- total  1   0.02305 7.3977 -115.59
- NCOMP  1   0.09136 7.4660 -115.04
- COMP   1   0.20284 7.5775 -114.15
<none>               7.3747 -113.78
- P15    1   0.59077 7.9654 -111.15
- NREST  1   1.37714 8.7518 -105.50

Step:  AIC=-115.59
EARN ~ P15 + COMP + NCOMP + NREST

        Df Sum of Sq     RSS      AIC
- NCOMP  1    0.0899  7.4876 -116.866
- COMP   1    0.1896  7.5873 -116.072
<none>                7.3977 -115.590
- NREST  1    1.3872  8.7849 -107.279
- P15    1    4.9716 12.3694  -86.747

Step:  AIC=-116.87
EARN ~ P15 + COMP + NREST

        Df Sum of Sq     RSS      AIC
- COMP   1    0.1521  7.6397 -117.659
<none>                7.4876 -116.866
- NREST  1    1.3744  8.8620 -108.754
- P15    1    5.1441 12.6317  -87.488

Step:  AIC=-117.66
EARN ~ P15 + NREST

        Df Sum of Sq     RSS      AIC
<none>                7.6397 -117.659
- NREST  1    1.2888  8.9286 -110.305
- P15    1    5.4293 13.0691  -87.446
-----------------------------------------------------
Backward stepwise selection of variables
-----------------------------------------------------
Linear regression (OLS)
Data     : CroqPainFix
Filter   : STOR <= 60
Response variable    : EARN
Explanatory variables: total, P15, P25, P35, P45, P55, COMP, NCOMP, NREST
Null hyp.: the effect of x on EARN is zero
Alt. hyp.: the effect of x on EARN is not zero
**Standardized coefficients shown (2 X SD)**
```

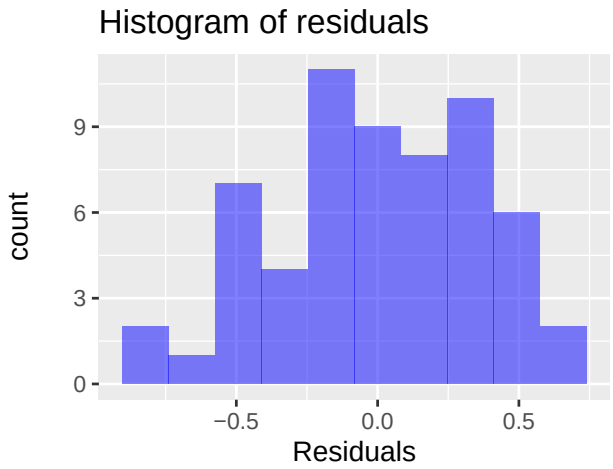
```
             coefficient std.error t.value p.value
 (Intercept)      -0.000     0.047  -0.000   1.000
 P15               0.608     0.096   6.365  < .001 ***
 NREST             0.296     0.096   3.101   0.003 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.482,  Adjusted R-squared: 0.464
F-statistic: 26.525 df(2,57), p.value < .001
Nr obs: 60

Prediction error (RMSE):  0.357
Residual st.dev   (RSD):  0.366

Variance Inflation Factors
      P15 NREST
VIF 1.005 1.005
Rsq 0.005 0.005
```

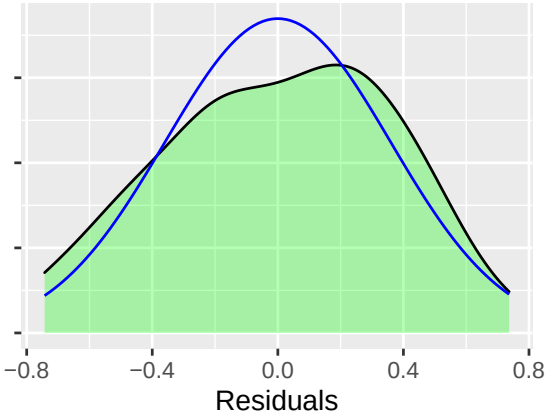










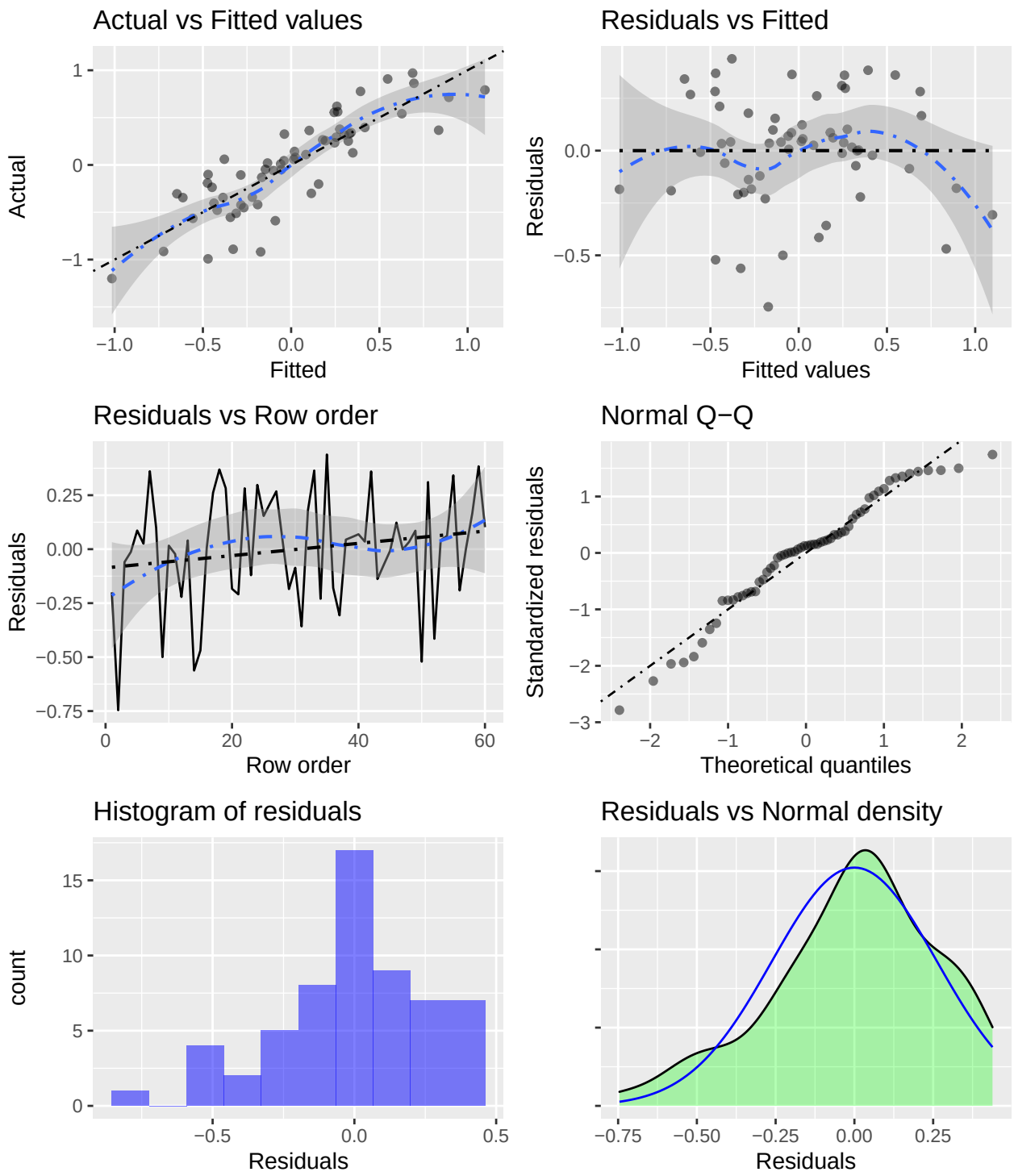


Step-wise for NORMALIZED DATA data.

```
Start:  AIC=-141.26
EARN_norm ~ SIZE + INC + PRICE + P15_norm + P25_norm + P35_norm +
    P45_norm + P55_norm + COMP_norm + NCOMP_norm + NREST_norm

             Df Sum of Sq    RSS     AIC
- P55_norm    1    0.0072 3.8263 -143.15
- P45_norm    1    0.0217 3.8409 -142.92
```

```
- P25_norm    1    0.0920 3.9111 -141.83
<none>                    3.8191 -141.26
- P35_norm    1    0.1354 3.9546 -141.17
- NCOMP_norm  1    0.1404 3.9595 -141.09
- COMP_norm   1    0.2330 4.0521 -139.71
- P15_norm    1    0.2987 4.1178 -138.74
- PRICE       1    0.5937 4.4128 -134.59
- INC         1    1.5728 5.3919 -122.57
- NREST_norm  1    2.8617 6.6808 -109.71
- SIZE        1    3.7604 7.5795 -102.13

Step:  AIC=-143.15
EARN_norm ~ SIZE + INC + PRICE + P15_norm + P25_norm + P35_norm +
    P45_norm + COMP_norm + NCOMP_norm + NREST_norm

             Df Sum of Sq    RSS     AIC
- P45_norm    1    0.0530 3.8793 -144.32
- P35_norm    1    0.1294 3.9558 -143.15
<none>                    3.8263 -143.15
- NCOMP_norm  1    0.1501 3.9764 -142.84
- P25_norm    1    0.1807 4.0071 -142.38
- COMP_norm   1    0.2606 4.0869 -141.19
- P15_norm    1    0.2937 4.1201 -140.71
- PRICE       1    0.5872 4.4136 -136.58
- INC         1    1.5659 5.3922 -124.56
- NREST_norm  1    2.9330 6.7593 -111.00
- SIZE        1    3.7628 7.5891 -104.06

Step:  AIC=-144.32
EARN_norm ~ SIZE + INC + PRICE + P15_norm + P25_norm + P35_norm +
    COMP_norm + NCOMP_norm + NREST_norm

             Df Sum of Sq    RSS     AIC
- P35_norm    1    0.0836 3.9629 -145.04
- NCOMP_norm  1    0.1188 3.9981 -144.51
<none>                    3.8793 -144.32
- P25_norm    1    0.1992 4.0785 -143.32
- P15_norm    1    0.2467 4.1260 -142.62
- COMP_norm   1    0.2940 4.1733 -141.94
- PRICE       1    0.6152 4.4944 -137.49
- INC         1    1.5772 5.4564 -125.85
- NREST_norm  1    2.8801 6.7594 -113.00
- SIZE        1    3.7100 7.5893 -106.06

Step:  AIC=-145.04
EARN_norm ~ SIZE + INC + PRICE + P15_norm + P25_norm + COMP_norm +
    NCOMP_norm + NREST_norm

             Df Sum of Sq    RSS     AIC
- NCOMP_norm  1    0.0875 4.0504 -145.73
<none>                    3.9629 -145.04
- P25_norm    1    0.2335 4.1964 -143.61
- COMP_norm   1    0.4379 4.4008 -140.75
- PRICE       1    0.6495 4.6124 -137.94
```

```
- P15_norm    1    1.3143 5.2772 -129.86
- INC         1    1.5329 5.4958 -127.42
- NREST_norm  1    2.8085 6.7714 -114.90
- SIZE        1    3.8516 7.8145 -106.30

Step:  AIC=-145.73
EARN_norm ~ SIZE + INC + PRICE + P15_norm + P25_norm + COMP_norm +
    NREST_norm

             Df Sum of Sq    RSS     AIC
<none>                    4.0504 -145.73
- P25_norm    1    0.3434 4.3939 -142.85
- COMP_norm   1    0.3528 4.4032 -142.72
- PRICE       1    0.7109 4.7614 -138.03
- P15_norm    1    1.3386 5.3890 -130.60
- INC         1    1.7157 5.7661 -126.54
- NREST_norm  1    2.8536 6.9040 -115.73
- SIZE        1    4.0818 8.1323 -105.91

-----------------------------------------------------
Backward stepwise selection of variables
-----------------------------------------------------
Linear regression (OLS)
Data     : CroqPainFix
Filter   : STOR <= 60
Response variable    : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P15_norm, P25_norm, P35_norm, P45_norm, P55_norm, COMP_norm, N
Null hyp.: the effect of x on EARN_norm is zero
Alt. hyp.: the effect of x on EARN_norm is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)      -0.000     0.036  -0.000   1.000
 SIZE              0.561     0.077   7.239  < .001 ***
 INC               0.360     0.077   4.693  < .001 ***
 PRICE            -0.233     0.077  -3.021   0.004 **
 P15_norm          0.374     0.090   4.146  < .001 ***
 P25_norm          0.178     0.085   2.100   0.041 *
 COMP_norm        -0.189     0.089  -2.128   0.038 *
 NREST_norm        0.560     0.092   6.053  < .001 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.725,  Adjusted R-squared: 0.688
F-statistic: 19.623 df(7,52), p.value < .001
Nr obs: 60

Prediction error (RMSE):  0.26
Residual st.dev   (RSD):  0.279

Variance Inflation Factors
    NREST_norm P15_norm COMP_norm P25_norm  SIZE PRICE   INC
VIF      1.619    1.543     1.499    1.364 1.136 1.122 1.116
Rsq      0.382    0.352     0.333    0.267 0.119 0.109 0.104
```



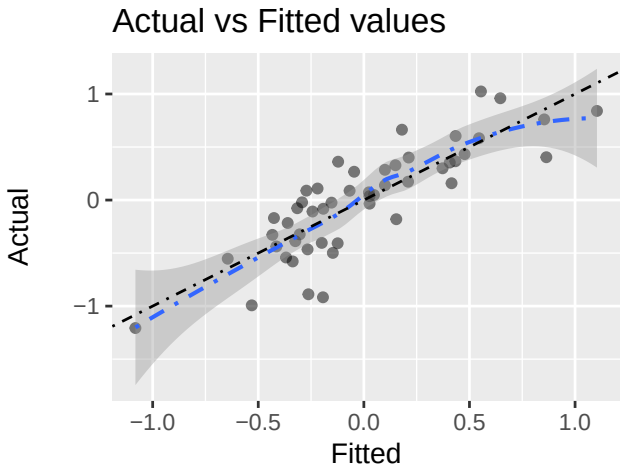
The step-wise functions above are purely statistical, thus do not take into consideration the VIF values, or the corresponding variables. When we make that determination based on the values, the non-transformed model still has VIF values that are too high. But the normalized model results in VIF values that are much lower, with a R^2 value of 0.725.

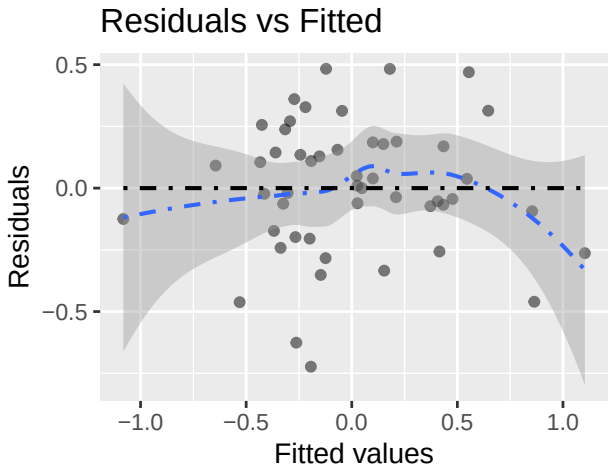
While the R^2 value above is higher than the intial 0.696 from our model, the observed difference is in the inclusion of the poplation variable. Although this generated model uses `P15_norm` and `P25_norm`, using `P35_norm` would be a better fit for the reccomendation we are making.

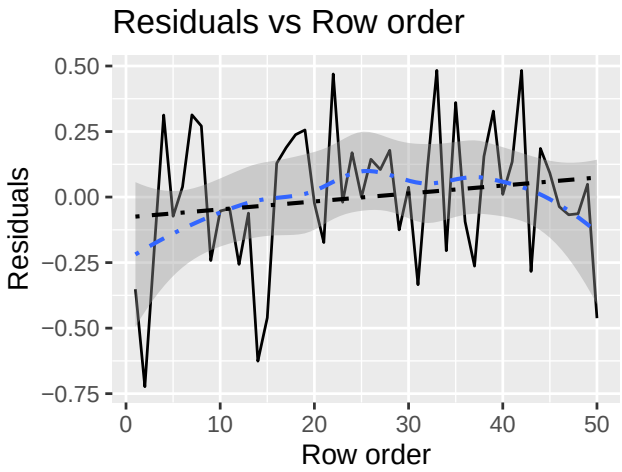
Re-modeled regression: Part B Using the data from stores that opened prior to 1994, we can rebuild the model. And this can be used to predict the normalized earnings for the last 10 stores.

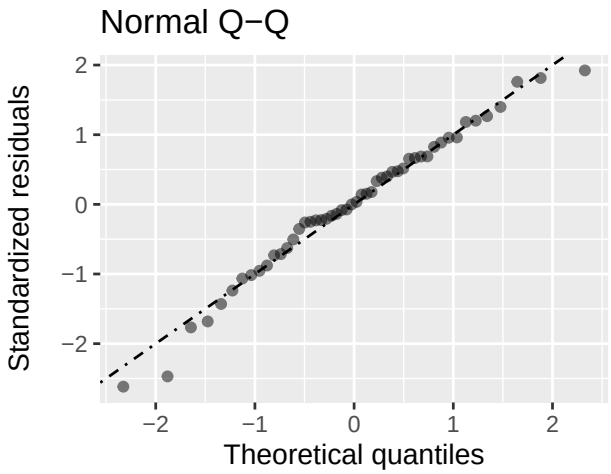
Using the same variables to build our model, we can observe the predicted values. We can also see the fit of this model with the data, when looking at the `Residuals vs Normal density` plot.

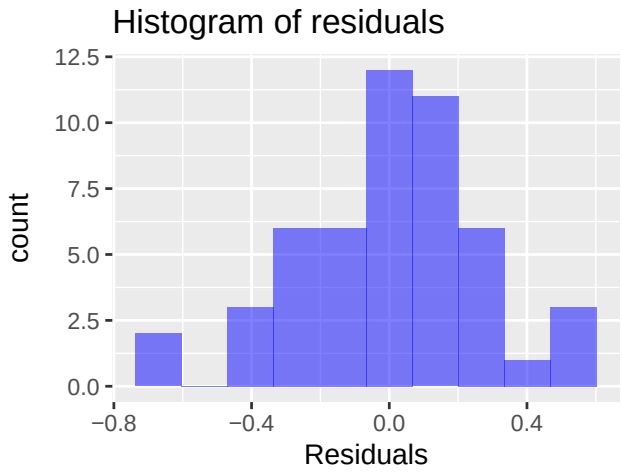
```
Linear regression (OLS)
Data     : CroqPainFix
Filter   : STOR <= 50
Response variable    : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P35_norm, COMP_norm, NREST_norm
Null hyp.: the effect of x on EARN_norm is zero
Alt. hyp.: the effect of x on EARN_norm is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)      -0.000     0.041  -0.000   1.000
 SIZE              0.410     0.090   4.536  < .001 ***
 INC               0.491     0.089   5.499  < .001 ***
 PRICE            -0.187     0.089  -2.099   0.042 *
 P35_norm          0.402     0.104   3.850  < .001 ***
 COMP_norm         0.002     0.107   0.022   0.983
 NREST_norm        0.581     0.101   5.750  < .001 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.708,  Adjusted R-squared: 0.668
F-statistic: 17.396 df(6,43), p.value < .001
Nr obs: 50

Prediction error (RMSE):  0.267
Residual st.dev   (RSD):  0.288

Variance Inflation Factors
    COMP_norm P35_norm NREST_norm  SIZE   INC PRICE
VIF     1.677    1.605      1.504 1.204 1.174 1.174
Rsq     0.404    0.377      0.335 0.170 0.149 0.148
```

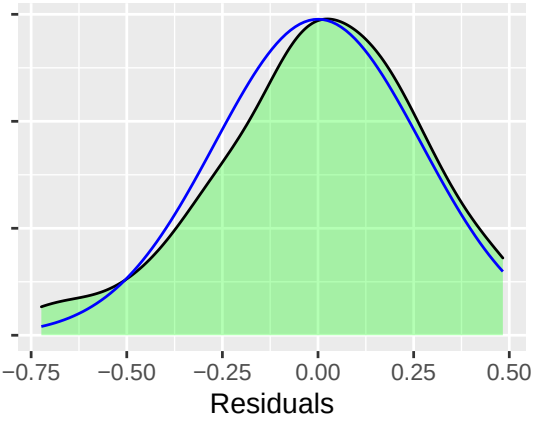













```
Linear regression (OLS)
Data                  : CroqPainFix
Filter                : STOR <= 50
Response variable     : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P35_norm, COMP_norm, NREST_norm
Prediction dataset    : CroqPainL

 SIZE    INC  PRICE P35_norm COMP_norm NREST_norm Prediction
```

```
 146 32.100 11.600    0.182     0.115      1.491      9.073
  62 32.700 18.000    0.218     0.543      5.435     12.216
  96 30.000 13.600    0.169     0.070      0.489      4.145
  86 34.400 16.500    0.195     0.231      1.309      8.048
  88 28.800 12.800    0.154     0.183      0.977      3.381
  72 28.700 24.300    0.175     0.094      0.468      1.399
 119 33.400 13.300    0.243     0.085      2.692     12.347
 285 27.600 18.300    0.224     0.088      1.752     11.895
 193 28.700 12.500    0.171     0.074      2.517      9.423
  92 36.000 14.100    0.238     0.039      1.216     10.840
```

COMP_norm is not statistically significant in this model, because p-value is still greater than 0.05, but we want to avoid ommitted variable bias, so included in the model to control for the model. In addition, when we are only looking at the data from 50 stores, competition will affect bottom-line earnings, and this a variable our model will take into account.

While these predicted values do not perfectly match the data provided for the last 10 stores, the tolerance level is within a comfortable range for us to accept this model.

To guage which stores should have been opened in 1994, we will base the decision on a threshold of 26% performance ratio.

To calulate performance ratio from our model, we will be doing

$\frac{Earnings * total}{K}$.

We only want to open stores which have a performance ratio greater than 26%. We observe this to be true in 3 stores: 51, 57 & 60. Thus, these are the only 3 stores which should have been opened in 1994.

Strategic Planning: Part C Using the model we built above, we will take a look at the 10 new stores to make a reccomendation on which ones should be opened.

```
Linear regression (OLS)
Data     : CroqPainFix
Response variable    : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P35_norm, COMP_norm, NREST_norm
Null hyp.: the effect of x on EARN_norm is zero
Alt. hyp.: the effect of x on EARN_norm is not zero
**Standardized coefficients shown (2 X SD)**

             coefficient std.error t.value p.value
 (Intercept)       0.000     0.038   0.000   1.000
 SIZE              0.514     0.079   6.538  < .001 ***
 INC               0.400     0.078   5.106  < .001 ***
 PRICE            -0.248     0.080  -3.114   0.003 **
 P35_norm          0.296     0.090   3.296   0.002 **
 COMP_norm        -0.095     0.096  -0.983   0.330
 NREST_norm        0.501     0.090   5.553  < .001 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.696,  Adjusted R-squared: 0.662
F-statistic: 20.262 df(6,53), p.value < .001
Nr obs: 60

Prediction error (RMSE):  0.273
Residual st.dev   (RSD):  0.291
```

```
Variance Inflation Factors
    COMP_norm NREST_norm P35_norm PRICE  SIZE   INC
VIF     1.623      1.423    1.413 1.111 1.078 1.074
Rsq     0.384      0.297    0.292 0.100 0.072 0.069

Linear regression (OLS)
Data                  : CroqPainFix
Response variable     : EARN_norm
Explanatory variables: SIZE, INC, PRICE, P35_norm, COMP_norm, NREST_norm
Prediction dataset    : CroqPain_holdout

 SIZE    INC  PRICE P35_norm COMP_norm NREST_norm Prediction
   54 38.400 21.700    0.064     0.596      2.683      4.862
  120 31.300 12.500    0.191     0.634      1.902      7.879
  135 29.000 21.700    0.196     0.084      1.092      5.866
  245 37.000 12.500    0.226     0.441      5.463     21.535
   96 30.000 18.000    0.143     3.429     10.857      9.258
  197 23.000 12.000    0.181     0.079      3.223      9.160
   93 25.200 33.400    0.169     0.120      0.300     -1.640
  169 30.100  9.000    0.094     0.425      1.169      6.779
  149 28.600 13.400    0.242     0.210      1.367      9.032
  150 34.900 15.400    0.079     0.237      4.269     11.069
```

Show 10 entries

Search:

STOR	P35_norm	COMP_norm	NREST_norm	pred_reg	PR
All	All	All	All	All	All
61	0.06	0.60	2.68	4.86	0.05
62	0.19	0.63	1.90	7.88	0.12
63	0.20	0.08	1.09	5.87	0.07
64	0.23	0.44	5.46	21.54	0.29
65	0.14	3.43	10.86	9.26	0.04
66	0.18	0.08	3.22	9.16	0.13
67	0.17	0.12	0.30	-1.64	-0.03
68	0.09	0.43	1.17	6.78	0.09
69	0.24	0.21	1.37	9.03	0.29
70	0.08	0.24	4.27	11.07	0.21

Showing 1 to 10 of 10 entries

Previous 1 Next

Stores with performance ratio greater than 26% are 64 & 69, and thus, should be opened.