**Matt Allen**

**DS 700**

**Assignment 3**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Exercise 1: Forecasting Chicken Wing Demand** | | | | |  |  |  |
| **a. Forecast the demand for week 7 using a five-period moving average** | | | | | | |  |
| Week | 1 | 2 | 3 | 4 | 5 | 6 | F7 |
| Demand | 650 | 521 | 563 | 735 | 514 | 596 | 585.8 |
|  |  |  |  |  |  |  |  |
| **b. Forecast the demand for week 7 using a three-period weighted moving average.** | | | | | | | |
| W1 | 0.5 |  |  |  |  |  | F7 |
| W2 | 0.3 |  |  |  |  |  | 599.2 |
| W3 | 0.2 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| **c. Forecast the demand for week 7 using exponential smoothing.** | | | | | | |  |
| alpha | 0.1 |  |  |  |  |  | F7 |
| F6 | 600 |  |  |  |  |  | 599.6 |
| A6 | 596 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| F7 = F6 + Alpha \* (A6 - F6) | | |  |  |  |  |  |

**d. What assumptions are made in each of the forecasts?**

In exercise a, I made the assumption that the average of the past five months is a good predictor of the chicken wing demand in the following month. In exercise b, I weighted the most current past periods the most with the assumption the future will be similar to near past. In exercise c, I assumed no trend or seasonality.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exercise 2: Forecasting tire demand** | | |  | |  | |  | |
|  | | | | | | | | |
| **Exercise 2: Forecasting tire demand** | | | |  | |  | |
| **a. Develop a spreadsheet using the first seven days of demand to determine the best exponential smoothing model for values of α = 0.2,**  **α = 0.3, and α = 0.4. Select the model with the smallest absolute deviation for seven periods.** | | | | | | | |
|  | | | | | | | |
|  | | | | | | | |
|  |  |  | |  | |  | |
| Total Months |  | Level smoothing parameter (alpha) | |  | | SumAbsDev | |
| 7 |  | 0.2 | |  | | 100.377088 | |
|  |  |  | |  | |  | |
| Day | Demand | Level Estimate | | Forecast | | AbsError | |
| 0 |  | 198 | |  | |  | |
| 1 | 200 | 198.4 | | 198 | | 2 | |
| 2 | 209 | 200.52 | | 198.4 | | 10.6 | |
| 3 | 215 | 203.416 | | 200.52 | | 14.48 | |
| 4 | 180 | 208.0992 | | 203.416 | | 23.416 | |
| 5 | 190 | 211.71904 | | 208.0992 | | 18.0992 | |
| 6 | 195 | 215.062848 | | 211.719 | | 16.71904 | |
| 7 | 200 | 218.0754176 | | 215.0628 | | 15.062848 | |

I chose alpha = 0.2. It had the smallest sum of absolute deviations. See spreadsheet.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Exercise 2: Forecasting tire demand** | | | |  | |  |  |
| **b. Develop another spreadsheet using the holdout sample for the second seven days to compare the best exponential smoothing model found in part a with a three-period moving average model. Compare the predictions on the basis of the total absolute deviation for the seven periods.** | | | | | | | |
|  | | | | | | | |
| Total Months |  |  | SumAbsDev | |
| 7 |  |  | 44 | |
|  |  |  |  | |
| Day | Demand | Forecast | AbsError | |
| 5 | 190 |  |  | |
| 6 | 195 |  |  | |
| 7 | 200 |  |  | |
| 8 | 208 | 195 | 13 | |
| 9 | 186 | 201 | 15 | |
| 10 | 193 | 198 | 5 | |
| 11 | 197 | 195.6666667 | 1.333333333 | |
| 12 | 188 | 192 | 4 | |
| 13 | 191 | 192.6666667 | 1.666666667 | |
| 14 | 196 | 192 | 4 | |

Based on the Sum of Absolute Deviations, the second model in (b) is a better predictor of tire demand

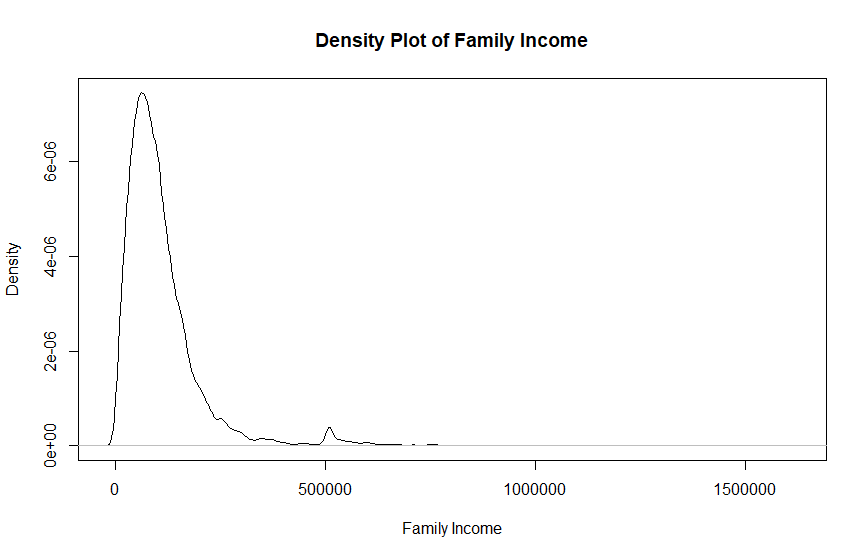
**c. What principles does this problem illustrate?**

This problem is an example of comparing models based on some statistic. Here the statistic is absolute deviation. I chose the model that minimized this statistic. The absolute deviation is the statistic I tried to minimize.

**Exercise 3: Predicting household income with logistic regression**

I created a new variable called FamilyIncomeGreaterThanOrEqualTo150000. This variable is set to TRUE if the existing variable FamilyIncome is greater than or equal to 150000 otherwise the value is FALSE.

Below is a density plot of Family Income. It appears to be right skewed.



For my first model, I am using the variables HouseCosts, and Insurance. I would expect a positive relationship between HouseCosts and Family income greater than or equal to 150000, and a positive relationship for Insurance as well.

The summary of the results is below.

Call:

glm(formula = FamilyIncomeGreaterThanOrEqualTo150000 ~ HouseCosts +

Insurance, family = "binomial", data = income\_data)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.8223 -0.6028 -0.4588 -0.3774 2.4395

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.936e+00 3.714e-02 -79.06 <2e-16 \*\*\*

HouseCosts 6.608e-04 1.765e-05 37.44 <2e-16 \*\*\*

Insurance 3.947e-04 1.935e-05 20.40 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 22808 on 22744 degrees of freedom

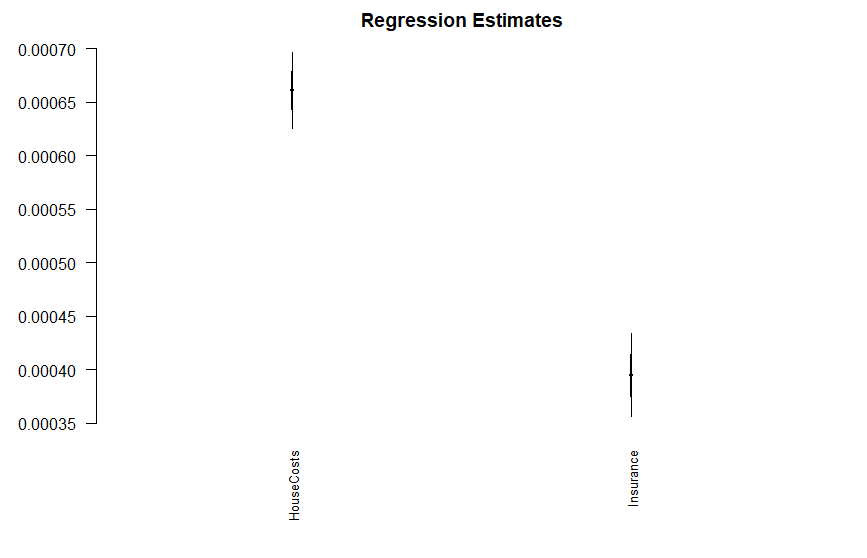
Residual deviance: 19238 on 22742 degrees of freedom

AIC: 19244

Number of Fisher Scoring iterations: 4

HouseCosts and Insurance do indeed have a positive relationship to income over $150,000. Null deviance is the measure of the effect of just the intercept. It is 22208. This seems very high. The residual deviance is a measure of goodness of fit of the entire model is 19238. This also seems high, but I tried four different models, and the AIC for this model was one of the lowest. From the summary the p-values of HouseCosts and Insurance are very low, which means they have a high degree of significance.

Below is a coefficient plot for HouseCosts and Insurance.



Of the four models, the one with the lowest AIC of 18908, and thus the better fit was with HouseCost, and NumWorkers. The summary is below.

Call:

glm(formula = FamilyIncomeGreaterThanOrEqualTo150000 ~ NumWorkers +

HouseCosts, family = "binomial", data = income\_data)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.5395 -0.6252 -0.4532 -0.2759 2.8351

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -4.009e+00 6.345e-02 -63.18 <2e-16 \*\*\*

NumWorkers 6.653e-01 2.501e-02 26.61 <2e-16 \*\*\*

HouseCosts 7.995e-04 1.656e-05 48.26 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 22808 on 22744 degrees of freedom

Residual deviance: 18902 on 22742 degrees of freedom

AIC: 18908