Statistical Analysis of the Nutritional Value Within Various Fast Food Companies

Lauryn Davis
STA 321 01

Fast Food Dataset Characteristics (Shape = 517, 17)

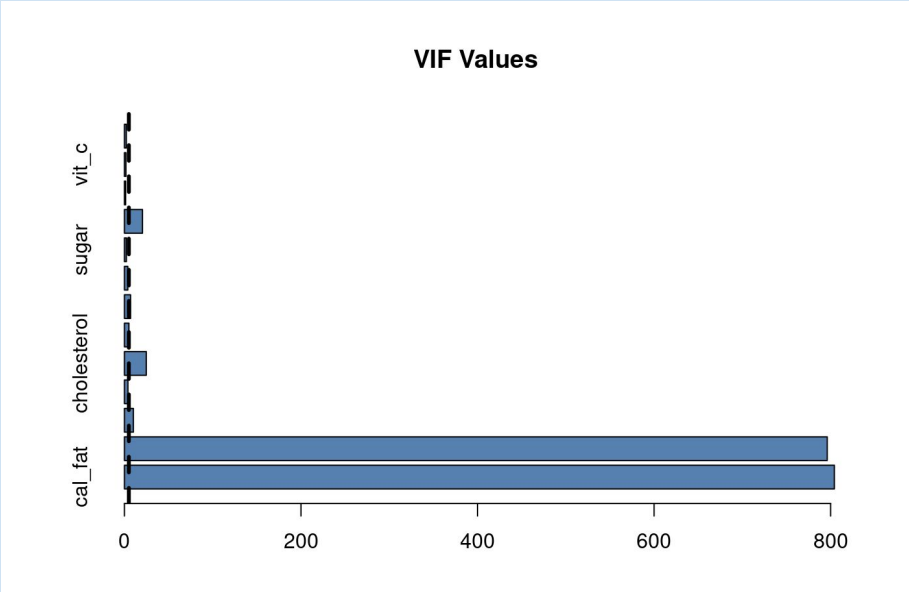
- Fast Food Companies (8): Mcdonalds, Chick Fil-A, Sonic, Arbys, Burger King, Dairy Queen, Subway, Taco Bell
- Food Item Variable
- Nutritional Quantitative Indicators: calories, total fat, saturated fat, trans fat, cholesterol, sodium, total carbohydrates, fiber, sugar, protein, vitamins A and C, calcium
- Miscellaneous indicator: If it is a salad item or not

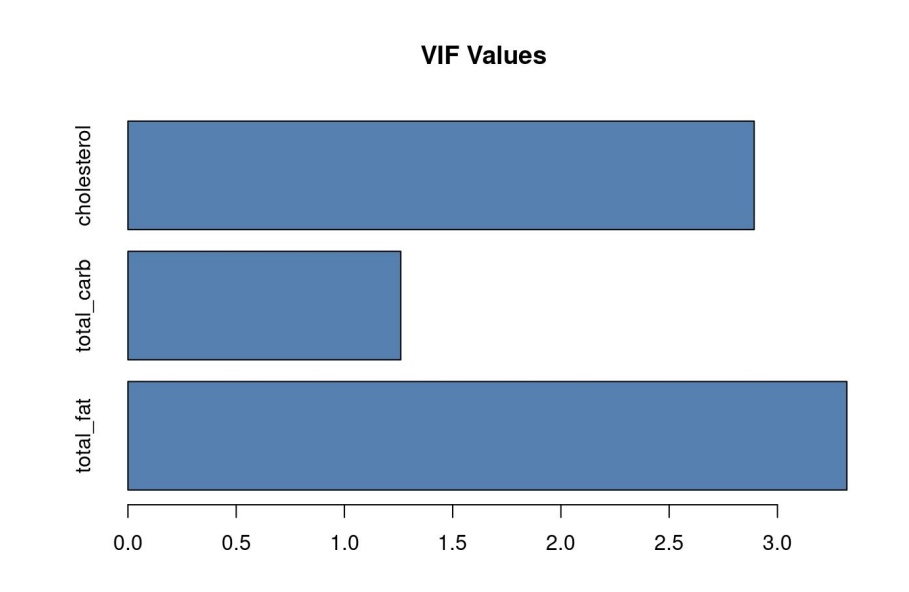
Research Questions

- Can we represent calories as a linear combination of various nutritional indicators?
- What restaurants have the largest weight within a model of “unhealthy” food items?

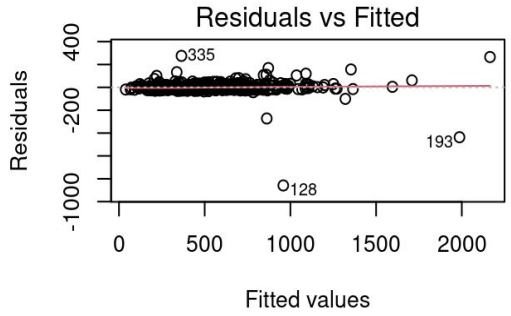
Multiple Regression Model for Calories:

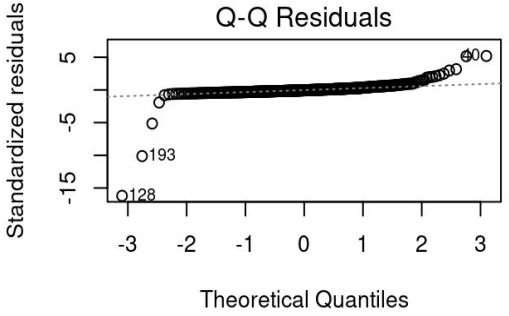
- Fixing Multicollinearity concerns:
- Final model <- lm(calories ~ **total_fat + total_ carb + cholesterol**). Adj R^2 = **96.41%**

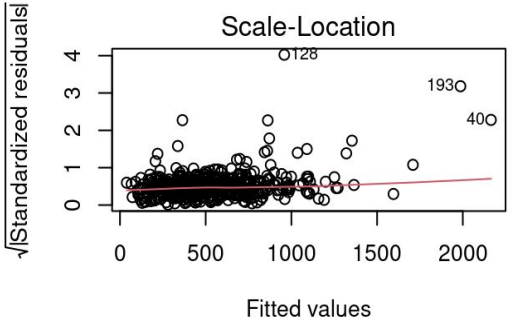


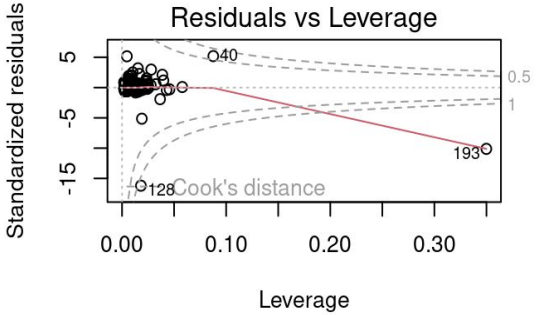


Why Checking Assumptions is Important:










```
##  Anderson-Darling normality test
##
## data:  rstandard(reduced_model)
## A = 73.99, p-value < 2.2e-16
```

lag	Autocorrelation	D-W Statistic	p-value
1	0.1218013	1.756271	0.014

Tried various transformations but nothing seemed to improve assumptions…

What's an appropriate amount of calories to eat in a day?

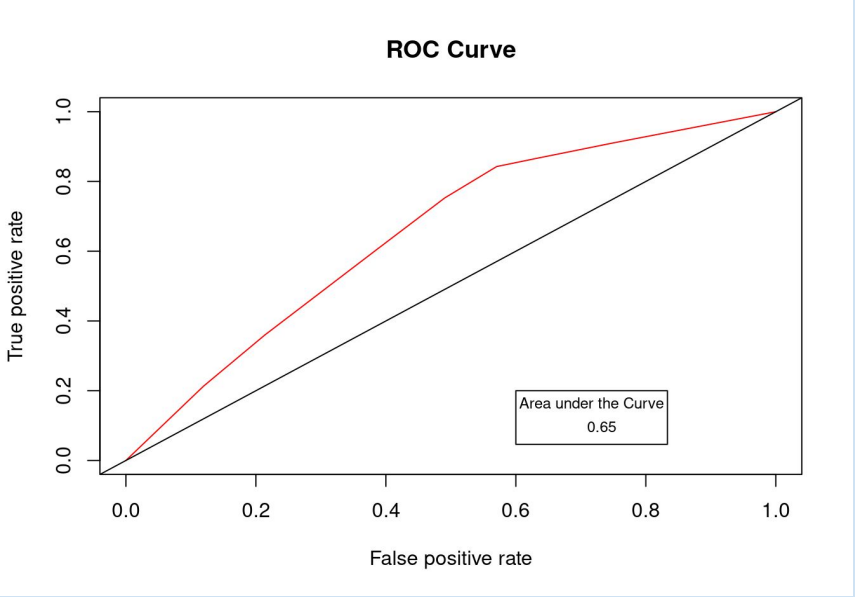
- "Generally, the recommended intake of calories in a day is 2,000 calories a day for women and 2,500 for men" (NHS).
- Median daily calorie intake for men and women = (2,000+ 2,500) / 2 = 2250 calories
- On average, in a very simplistic model, breakfast, lunch, and dinner should each be around 2250/3 = 750 calories per meal
- Create a binary indicator of a food item that is over this calories limit:

```
data$unhealthylevel <- ifelse(data$calories>750, 1, 0)
```

"What should my daily intake of calories be?", NHS, 28 Sept. 2023

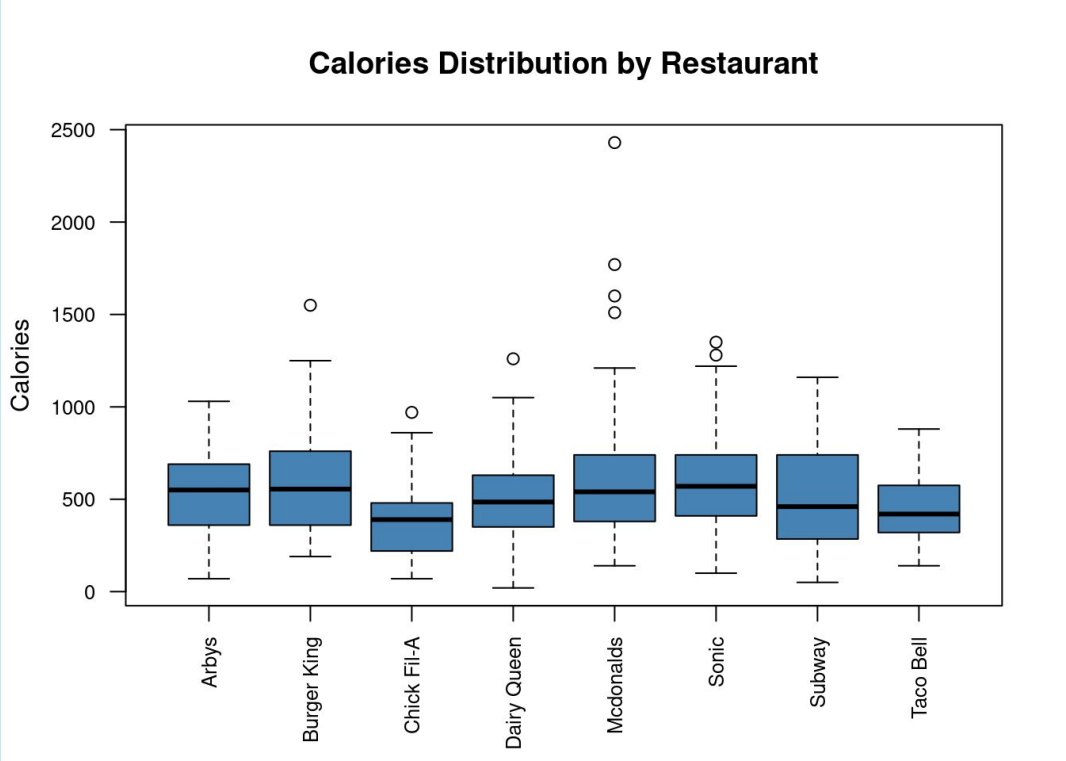
Logistic Regression Model

- Transform the restaurant variable to be a factor
- **Model<- glm(unhealthy_level ~ as.factor(restaurant))**
- Create a testing and training set to measure classification accuracy
- The model was able to predict if an item was unhealthy or not **58%** of the time!
- Pseudo R^2 = 5.6%



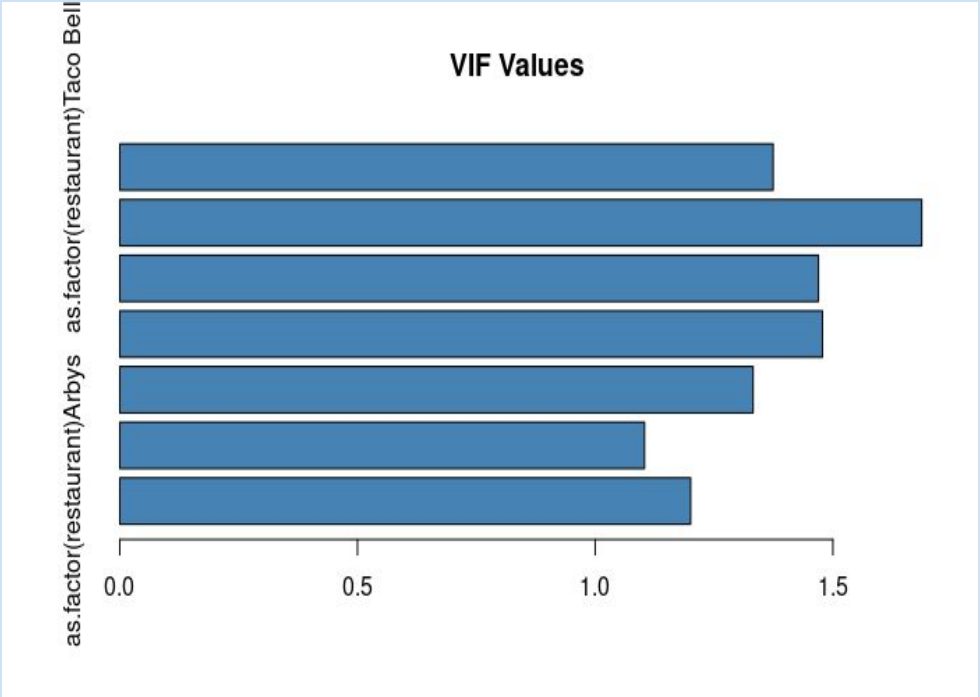

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: unhealthylevel

Terms added sequentially (first to last)


                     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                   514     474.13
as.factor(restaurant)  7   26.761       507     447.37 0.000368 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Calorie Distributions Per Restaurant



Checking Assumptions for Logistic Regression

- Multicollinearity satisfied
- No Influential Observations because all std. Residuals < 3.
- Hoslem goodness of fit not satisfied. P-value = <.0001
- Independence Durbin Watson test p-value = 0



Interpreting Results From Logistic Regression

	Mcdonalds	Taco Bell	Arby's	Chick Fil-A	Subway	Sonic	Burger King	DQ
Mcdonalds	1	.253** (-)	.265	.270	1.01	1.1	1.26	.796
Taco Bell	.395*** (+)	1	1.05** (+)	1.07	3.98	4.35*** (+)	4.98** (+)	3.14
Arby's	3.78	.953	1	1.02	3.79	4.14	4.75*** (+)	3
Chick Fil-A	3.69	.953	.98	1	3.72	4.06	4.66	2.941
Subway	.994	.251** (-)	.264	.269	1	1.09	1.253	.791
Sonic	.909	.230** (-)	.241	.246	.915	1	1.14	.724
Burger King	.793	.201*** (-)	.211*** (-)	.215*** (-)	.798	.872	1	.632
Dairy Queen	1.23	.318	.333	.3.4	1.26	1.38	1.58	1

Alpha Level:

*** : 0, ** : .001, * : .01.

+ : Column restaurant is unhealthier than reference

- : Column restaurant is healthier than reference

- The odds ratio of an item being unhealthy at Taco bell was .395 times **less** probable than at Mcdonald's.
- Restaurants within the Burger King model tend to be healthier

Conclusions

From Multiple Regression:

- Although assumptions were violated, it's interesting that total_fat, cholesterol, and total_carbs can predict calories so well
- Could look into predictors that are not as heavily associated with calories
- Always check assumptions!

From Logistic Regression:

- Although assumptions were violated, it's noteworthy that there was a statistically significant difference between the restaurant that was used in predicting an unhealthy item
- On average, Taco Bell tended to have the "healthiest" weights within the models
- On average, Burger king had the "unhealthiest" weight within the models
- Dairy Queen, Chick Fil-A, and Arby's had similar projections (however Arby's was significantly healthier than Burger King).