Jacqueline Vo

Regression Analysis Student Project

Spring 2011

Pizza Hut Slice Nutritional Value and Calorie Count

My project will study the relationship between the nutritional content in various types of Pizza Hut pizza and the calorie count per slice. I will regress the calorie count (Y) on 9 nutrients (explanatory variables: X1, X2,…, X9) that are listed on the website.

	Calories	Calories from Fat	Total Fat	Saturated Fat	Cholesterol	Sodium	Carbs	Fiber	Sugar	Protein
Cheese Only	240	90	10	4.5	0.025	0.53	27	1	2	11
Pepperoni	250	110	12	4.5	0.025	0.59	26	1	2	11
Supreme	290	130	14	5	0.03	0.65	27	2	2	12
Pepperoni & Mushroom	240	90	10	4	0.02	0.52	27	1	2	10
Italian Sausage & Red Onion	270	110	13	4.5	0.025	0.56	28	1	3	11
Ham & Pineapple	230	80	9	3.5	0.02	0.52	28	1	3	10
Veggie Lover's®	230	80	9	3.5	0.015	0.5	28	2	3	9
Meat Lover's®	330	160	18	7	0.04	0.83	27	1	2	14
Pepperoni Lover's®	290	130	14	6	0.035	0.73	27	1	2	13
Hawaiian Luau	260	100	12	4.5	0.025	0.61	28	1	3	11
Dan's Original	280	130	14	5	0.03	0.63	27	1	2	12
Triple Meat Italiano	290	130	15	5	0.03	0.7	27	1	2	13
Spicy Sicilian	270	120	13	5	0.025	0.7	27	2	2	11
Ultimate Cheese Lover's	270	120	13	5	0.025	0.58	26	1	2	12

Variables

- Y = Calories
- a = Intercept
- Bi = Coefficient
- X1 = Calories from fat
- X2 = Total Fat
- X3 = Saturated Fat
- X4 = Cholesterol
- X5 = Sodium
- X6 = Carbohydrates
- X7 = Dietary Fiber
- X8 = Sugar
- X9 = Protein

Regression Equation:

Y = a + B1X1 + B2X2 + B3X3 + B4X4 + B5X5 + B6X6 + B7X7 + B8X8 + B9X9

Hypothesis:

All coefficients Bi (i = 1,…,9) are zero

Regression Analysis: I will first regress calories on all 9 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-124.6816021	107.1750254	-1.163345673	0.309356213
X Variable 1	0.385426779	0.480344693	0.802396246	0.467288342
X Variable 2	5.30911047	3.91754494	1.355213674	0.246825985
X Variable 3	7.038750699	4.54785631	1.547707364	0.196605657
X Variable 4	-330.5387108	941.0172303	-0.35125681	0.743116738
X Variable 5	-42.1756872	35.45829983	-1.189444711	0.300042963
X Variable 6	8.074021777	3.998100946	2.019464212	0.113567049
X Variable 7	5.172110928	4.141893331	1.248731079	0.279857263
X Variable 8	-2.605848956	5.338257093	-0.48814602	0.650996647
X Variable 9	5.531377518	3.707386	1.491989	0.209978

Regression Stats	
Multiple R	0.997531
R Square	0.995068
Adjusted R Square	0.983971
Standard Error	3.561256
Observations	14

- The full regression produces this equation:

  Y = -124.68 + .39 (Calories from Fat) + 5.31 (Total Fat) + 7.04 (Saturated Fat) – 330.54 (Cholesterol) – 42.18 (Sodium) + 8.07 (Carbohydrates) + 5.17 (Dietary Fiber) – 2.61 (Sugar) + 5.53 (Protein)

- R Square = 99.75% so, overall, the listed nutrients are good predictors of calories since 99.5% of the variation in calories is explained by the 9 nutrients
- Looking closer, I see that some of the variables are not statistically significant. For example, variable X4, cholesterol, has a high p value of .74. I will remove this variable and regress calories on the remaining 8 nutrients

Regression Analysis 2: I will regress calories on the remaining 8 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-99.22313642	71.69750878	-1.383913306	0.224976115
X Variable 1	0.312030255	0.392774189	0.794426578	0.462969268
X Variable 2	5.865920331	3.253330632	1.803050779	0.131241124
X Variable 3	6.231873765	3.564432716	1.74834939	0.140821567
X Variable 5	-43.20523238	32.09008131	-1.346373415	0.235994907
X Variable 6	7.277903839	2.990982102	2.43328231	0.059140985
X Variable 7	5.762055233	3.438212411	1.675886927	0.154609098
X Variable 8	-2.173097796	4.716891026	-0.460705533	0.664354133
X Variable 9	4.786778617	2.762094328	1.733025034	0.143630343

Regression Stats	
Multiple R	0.997454644
R Square	0.994915767
Adjusted R Square	0.986780995
Standard Error	3.234036587
Observations	14

- R square stays approximately the same which means the 8 variable model is still overall a good model
- The F statistic went from 89.67 to 122.3
- I will remove variable X8, sugar, to determine if regression of calories on the remaining 7 variables is a better model

Regression Analysis 3: I will regress calories on the remaining 7 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-81.84821729	56.83342458	-1.440142274	0.199895305
X Variable 1	0.356648504	0.354780158	1.005266208	0.353573642
X Variable 2	5.33609468	2.836480002	1.881238252	0.108967791
X Variable 3	6.216059377	3.322058364	1.87114695	0.110502974
X Variable 5	-40.70842331	29.4797806	-1.380893022	0.216543875
X Variable 6	6.318547918	2.001129578	3.157490643	0.01962736
X Variable 7	5.946079233	3.182869917	1.868150251	0.110962989
X Variable 9	5.104612932	2.49280231	2.047740774	0.086512195

Regression Stats	
Multiple R	0.99734645
R Square	0.994699942
Adjusted R Square	0.988516541
Standard Error	3.014268295
Observations	14

- R square stays approximately the same which means the 7 variable model is still overall a good model
- The F statistic went from 122.3 to 160.9
- I will remove variable X1, calories from fat, to determine if regression of calories on the remaining 6 variables is a better model

Regression Analysis 4: I will regress calories on the remaining 6 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-49.87824842	47.13795477	-1.058133486	0.325136705
X Variable 2	7.958006264	1.11575534	7.132393615	0.000188237
X Variable 3	7.168740293	3.186416347	2.24978142	0.059217207
X Variable 5	-42.52843092	29.44631997	-1.444269809	0.191889566
X Variable 6	4.972194389	1.487994946	3.341539836	0.012393495
X Variable 7	7.754852315	2.627467991	2.951454535	0.02136269
X Variable 9	5.649908413	2.434903287	2.32038309	0.053364336

Regression Stats	
Multiple R	0.996898827
R Square	0.993807271
Adjusted R Square	0.988499218
Standard Error	3.016541089
Observations	14

- R square stays approximately the same at 99.7% which means the 6 variable model is still overall a good model
- The F statistic went from 160.9 to 187.2
- At this point there is only one more variable with a high p value, X5. I will remove variable X5, sodium, to determine if regression of calories on the remaining 5 variables is a better model

Regression Analysis 5: I will regress calories on the remaining 5 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-31.50205761	48.37081266	-0.651261699	0.533134509
X Variable 2	7.559156379	1.152074979	6.561340639	0.000176367
X Variable 3	4.727366255	2.878571936	1.642260941	0.139161721
X Variable 6	4.319701646	1.510914577	2.858997928	0.021184007
X Variable 7	6.100823045	2.520135176	2.420831669	0.041795821
X Variable 9	4.929269547	2.539834634	1.940783656	0.088239455

Regression Stats	
Multiple R	0.995972847
R Square	0.991961913
Adjusted R Square	0.986938108
Standard Error	3.214760278
Observations	14

- R square drops to 99.6% which is still high
- The F statistic went from 187.2 to 207.1
- I will remove variable X3, saturated fat, to determine if regression of calories on the remaining 4 variables is a better model

Regression Analysis 5: I will regress calories on the remaining 4 nutrients to determine which nutrients are the best predictors of calorie content.

	Coefficients	Standard Error	t Stat	P-value
Intercept	-35.59452524	52.66431671	-0.675875573	0.516109059
X Variable 2	8.115198175	1.200537692	6.759636311	8.27411E-05
X Variable 6	4.148845167	1.643304592	2.524696387	0.032517111
X Variable 7	6.820644425	2.705602696	2.520933482	0.032718498
X Variable 9	6.984602224	2.409448502	2.898838559	0.017628681

Regression Stats	
Multiple R	0.994611508
R Square	0.989252051
Adjusted R Square	0.984475185
Standard Error	3.504763868
Observations	14

- R square drops to 99.5% which is still high
- The F statistic went from to 207.1 to 207.0
- All X's are statistically significant

Conclusion:

I believe the best model is my 5th iteration (which included 4 of the original 9 nutrients) because all the variables are statistically significant. Although R and R square dropped slightly after each run, the last (fifth) regression still had R square above 99%. Furthermore, the F statistic increased with every regression. Since F was large to begin with (significantly greater than 1), I can reject the null hypothesis.

Through this analysis, I found that total fat, Carbohydrates, dietary fiber and protein are the greatest predictors of calorie content in Pizza Hut pizza slices.