

**DATE: 6/28/2021**  
**TO: Chocolate Bar Creators**  
**FROM: Charlie Evert**  
**RE: Predicting Candy Bar Calories Through Nutrients**

### **Introduction**

Consumers have many choices in terms of candy bars, although it is generally accepted that consumers aim to consume the lowest calorie option if such measures are marketed successfully. There are many factors that constitute the caloric make-up of candy bars, and each should be changed to produce palatable candy bars with the lowest possible calories. The following write-up aims to simplify how each factor contributes to calories, and can be used to target specific factors in the creation of candy bars to best capture the low-calorie market.

### **Data Summary**

The data is made up of macronutrients and micronutrients, and corresponding calories for each candy bar. There are many nutrients listed, with some correlating with other nutrients more than calories. This creates problems for data integrity, and must be sorted through. The data is best grouped prior to analysis. Data analysis for these groupings is located throughout the appendix, especially appendix L.

### **Findings**

The data has multicollinearity issues if left ungrouped. However, through the use of grouping into common factors, resultant models can be relied upon.

Through the use of this grouping, I have identified 6 factors, including: heart stopping factor, package density, hearty level, protein richness, carb richness and saltiness level. Each of these factors can, with a high degree of accuracy, predict calories.

The factors are not highly correlated with one another; thus, multicollinearity has been addressed.

### **Recommendations**

Candy bar calories are mainly influenced by how hearty, carb rich and heart stopping they are. Thus, increasing protein richness and reducing package density should lead to lower calorie chocolate bars. Additionally, candy bar makers should try to replace hearty, carb rich and heart stopping nutrients with increasing saltiness levels since this has a lesser effect on increasing calories than those factors. Candy makers should therefore target these factors in candy bar creation.

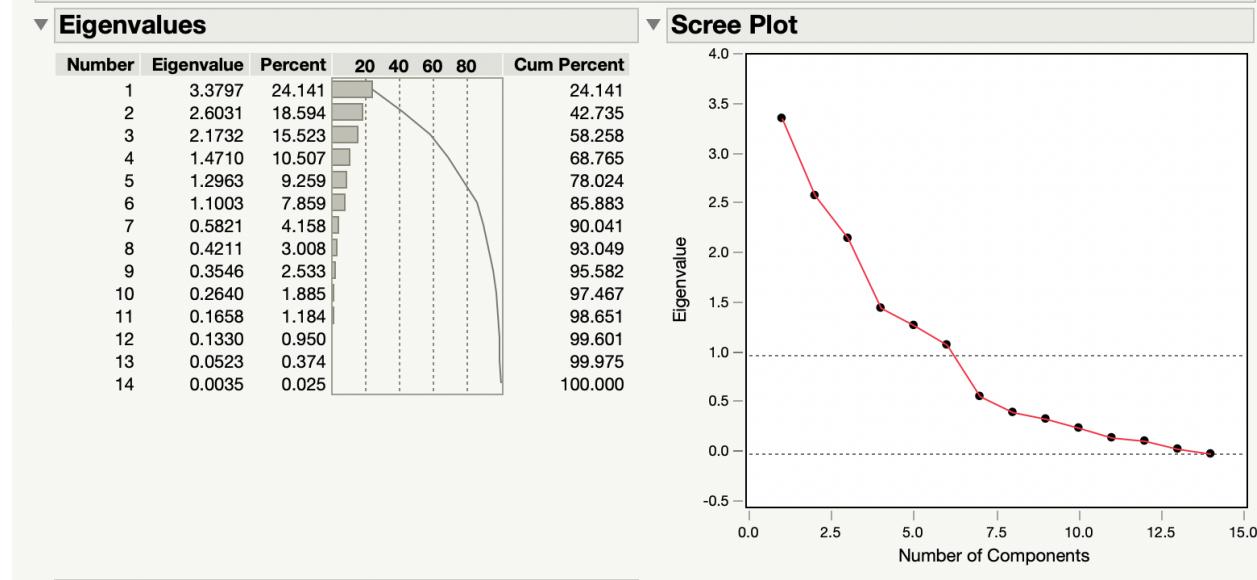
## Appendix A: Pairwise Correlations

Pairwise Correlations		Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-8	-6	-4	-2	0	.2	.4	.6	.8
Oz/pkg	Servings/pkg	0.9469	75	0.9170	0.9662	<.0001*											
Saturated fat g	Total fat g	0.7977	75	0.6970	0.8676	<.0001*											
Sugars g	Carbohydrate g	0.7003	75	0.6508	0.8000	<.0001*											
Protein g	Total fat g	0.6539	75	0.3915	0.7040	<.0001*											
Iron %RDI	Vitamin C %RDI	0.5880	75	0.3915	0.7040	<.0001*											
Iron %RDI	Vitamin A %RDI	0.5296	75	0.3407	0.6733	<.0001*											
Dietary fiber g	Total fat g	0.5091	75	0.3190	0.6598	<.0001*											
Calcium %RDI	Vitamin C %RDI	0.4963	75	0.3035	0.6500	<.0001*											
Protein g	Oz/pkg	0.4860	75	0.2911	0.6421	<.0001*											
Cholesterol g	Saturated fat g	0.4727	75	0.2753	0.6199	<.0001*											
Vitamin C %RDI	Vitamin A %RDI	0.4320	75	0.2373	0.6001	0.0001*											
Cholesterol g	Total fat g	0.4320	75	0.2373	0.6001	0.0001*											
Iron %RDI	Calcium %RDI	0.4303	75	0.2253	0.5988	0.0001*											
Calcium %RDI	Protein g	0.4202	75	0.2137	0.5909	0.0002*											
Iron %RDI	Iron %RDI	0.3777	75	0.1649	0.5569	0.0008*											
Dietary fiber g	Saturated fat g	0.3548	75	0.1391	0.5384	0.0018*											
Iron %RDI	Oz/pkg	0.3501	75	0.1350	0.5305	0.0025*											
Calcium %RDI	Vitamin A %RDI	0.3501	75	0.1340	0.5341	0.0021*											
Calcium %RDI	Total fat g	0.3492	75	0.1327	0.5338	0.0021*											
Calcium %RDI	Cholesterol g	0.3301	75	0.1115	0.5182	0.0038*											
Calcium %RDI	Saturated fat g	0.3249	75	0.1058	0.5140	0.0045*											
Protein g	Sodium mg	0.2799	75	0.0566	0.4766	0.150*											
Calcium %RDI	Oz/pkg	0.2427	75	0.0167	0.4452	0.359*											
Protein g	Saturated fat g	0.2420	75	0.0159	0.4445	0.359*											
Calcium %RDI	Oz/pkg	0.2416	75	0.0086	0.4396	0.559*											
Cholesterol g	Serving/pkg	0.2140	75	-0.0136	0.4206	0.6652											
Cholesterol g	Oz/pkg	0.2043	75	-0.0237	0.4122	0.7078											
Saturated fat g	Oz/pkg	0.1919	75	-0.0367	0.4013	0.0991											
Vitamin C %RDI	Protein g	0.1618	75	-0.0677	0.3750	0.1656											
Vitamin A %RDI	Sodium mg	0.1597	75	-0.0698	0.3731	0.1712											
Iron %RDI	Total fat g	0.1563	75	-0.0849	0.3565	0.2055											
Total fat g	Oz/pkg	0.1339	75	-0.0993	0.3509	0.2529											
Vitamin A %RDI	Protein g	0.1147	75	-0.1153	0.3330	0.3273											
Calcium %RDI	Dietary fiber g	0.1138	75	-0.1161	0.3322	0.3308											
Sugars g	Saturated fat g	0.1086	75	-0.1213	0.3275	0.3536											
Vitamin A %RDI	Cholesterol g	0.0947	75	-0.1352	0.3149	0.4191											
Vitamin C %RDI	Dietary fiber g	0.0929	75	-0.1370	0.3132	0.4280											
Saturated fat g	Oz/pkg	0.0785	75	-0.1521	0.2952	0.5181											
Carbohydrate g	Sodium mg	0.0727	75	-0.1569	0.2946	0.5355											
Sodium mg	Total fat g	0.0571	75	-0.1721	0.2804	0.6286											
Dietary fiber g	Oz/pkg	0.0545	75	-0.1746	0.2780	0.6425											
Sugars g	Cholesterol g	0.0450	75	-0.1838	0.2692	0.7014											
Iron %RDI	Saturated fat g	0.0280	75	-0.2002	0.2534	0.8114											
Protein g	Oz/pkg	0.0194	75	-0.2084	0.2453	0.8685											

The highest positive correlations are listed at the top left, and the highest negative correlations are listed at the bottom right. For Oz/pkg and servings per package there is the greatest positive correlation, and for Carbohydrates and servings/pkg there is the highest negative correlation (with much less of a correlation as compared to the greatest positive correlation). Sugars and oz/packaging are the least correlated of any variables. It is evident that there is an issue of multicollinearity with this data due to these high correlations between independent variables.

## Appendix B: Eigenvalues and Scree Plot

### Factor Analysis



Since Eigenvalues are greater than 1 for factors 1-6, these 6 factors will be analyzed. Although, slope did significantly change between factors 6 and 7 so one could alternatively use 7 factors.

**Appendix C: Significance Test****Significance Test**

Test	DF	ChiSquare	Prob>ChiSq
H0: no common factors.	91	855.807	<.0001*
HA: at least one common factor.			

Test	DF	Criterion	ChiSquare	Prob>ChiSq
H0: 6 factors are sufficient.	22	0.947	61.065	<.0001*
HA: more factors are needed.				

This test shows significance that 6 factors are sufficient for the model and that there are no common factors.

**Appendix D: Measures of Fit****▼ Measures of Fit**

Measures of Fit	Fit Index
Chi-Square without Bartlett's Correction	70.059
AIC	26.059
BIC	-24.925
Tucker and Lewis's Index	0.789
Root Mean Square Error of Approximation	0.172

Overall, the model fits the data quite well and therefore the model should be relied upon. This is evident by the high Tucker and Lewis' index.

## Appendix E: Factor Loading Tables & Defining Factors

### ▼ Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Saturated fat g	0.793913	0.060123	0.326717	-0.123808	0.090312	-0.098924
Total fat g	0.722522	0.002224	0.676514	-0.087603	-0.026033	0.090358
Cholesterol g	0.704619	0.127850	-0.123196	0.010023	-0.112724	0.004417
Oz/pkg	0.125168	0.987322	0.057856	-0.039152	0.043213	-0.030403
Serving/pkg	0.076723	0.960006	-0.102078	-0.007483	-0.215316	-0.085662
Protein g	0.145879	-0.021060	0.797687	0.301422	-0.144722	0.297097
Dietary fiber g	0.050955	0.000951	0.708063	-0.049592	-0.022480	-0.141418
Calcium %RDI	0.432822	0.204154	0.160628	0.838008	-0.007218	-0.111222
Vitamin C %RDI	-0.241272	-0.062178	0.013541	0.704721	-0.111160	-0.042690
Iron %RDI	-0.109599	-0.104289	0.391854	0.485345	-0.060642	-0.141431
Vitamin A %RDI	-0.034605	-0.041064	-0.090118	0.471105	-0.202373	0.168290
Carbohydrate g	-0.226888	-0.106814	-0.040786	-0.143626	0.953403	0.060259
Sugars g	0.120820	-0.052463	-0.102765	-0.161162	0.784749	-0.070796
Sodium mg	-0.044624	-0.094820	-0.002503	-0.031313	-0.014271	0.990629

The following labels each above factor based upon its commonalities along with a brief description regarding what each label means:

Factor 1: Heart stopping; how heart-stopping candy bars are in terms of fat & cholesterol.

Factor 2: Package Density; the make-up of how dense packages are.

Factor 3: Hearty Level; how “healthy” bars are in terms of low carb, high fat/protein diets.

Factor 4: Protein Richness; how much protein & accompanying nutrients bars have.

Factor 5: Carb Richness; how sugary and carb-heavy bars are.

Factor 6: Saltiness Level; how salty bars are.

## Appendix F: Analysis of Dataset using Factors

	Brand	Name	Heart Stopping	Package Density	Hearty Level	Protein Richness	Carb Richness	Saltiness Level
1	M&M/Mars	Snickers Peanut B...	0.93513385...	-0.282424532	0.87495109	-0.041059792	-0.208342148	1.5378254227
2	Hershey	Cookies 'n' Mint	0.48250918...	-0.555751158	-0.396635185	0.8874273462	-0.278213988	0.055449686
3	Hershey	Cadbury Dairy Milk	0.55703637...	2.438635694	-0.463786404	0.6551225615	-0.276303131	-0.391302397
4	M&M/Mars	Snickers	-0.552849716	1.4674333478	-0.582186577	-0.767428713	-1.27165078	0.2324347649
5	Charms	Sugar Daddy	-0.684152677	-0.261495728	-1.627395518	0.5547787162	1.1868017601	0.4516837932
6	M&M/Mars	Twix Peanut Butter	-0.004072532	-0.426613408	0.817687073	-0.747344677	-0.76393086	1.028136582
7	Hershey	Twizzler	-2.034227879	0.3271738806	-0.670680801	-0.109342786	0.6842289189	1.4617365914
8	Toblerone	0.74207132...	-0.871779896	-0.881196723	0.1741122175	-1.000220042	-1.149618298	
9	Nestle	Crunch	0.72061969	-0.5988952491	-0.605174748	0.7612693925	-0.112673547	-0.358217184
10	Hershey	Almond Joy	-0.291895751	0.9240168781	0.4291538603	-0.96446714	-0.792953003	0.195397862
11	Sherwood	Elana Mint	-0.261647633	-0.533177006	-0.100388636	-0.74556804	-0.369635499	-1.499014462
12	Hershey	Krackel	1.83042473	0.0496812482	0.4270050552	0.5923390962	2.1332835517	0.8312842927
13	M&M/Mars	M&Ms Peanut	-0.431424752	-0.427183705	0.8737641426	-0.241157535	-0.182353918	-1.159328454
14	Bit-O-Honey	Bit-O-Honey	-0.865628011	-0.268801162	-1.135911644	-0.01984899	0.8376404312	0.5229002924
15	Nestle	100 Grand	0.20376136	-0.554151379	-1.205455448	0.002167448	-0.165257928	-0.094865468
16	Hershey	Skor	1.427337079	-0.750919928	-1.405186359	-0.450237166	-0.827729623	0.6487944634
17	Hershey	Twix Caramel	0.520422831	-0.24159628	-0.123706245	-0.319567169	0.6944737118	0.7905156047
18	M&M/Mars	Milky Way Lite	-0.576013885	-0.373141117	-1.154916661	0.4224051517	0.048003024	0.4264762294
19	M&M/Mars	Mars	0.39794999	-0.424966197	-0.081912556	0.2308216356	0.1018996283	-0.161880485
20	Pearson	Peanut Nut Roll	-0.999662549	0.3560704405	1.8180638414	0.2146019518	0.7826058221	2.7884126153
21	Nestle	Raisinet	-0.519742837	-0.490762693	-0.257625371	0.009248592	-0.092281777	-1.372102162
22	Sherwood	Elana Mocca	0.76582737...	-0.606261202	-0.413061043	0.0459112941	-0.287181213	-0.906670137
23	Hershey	Bar None	0.40930323...	-0.550147965	0.3588698647	-0.465973778	-0.669372374	0.521457159
24	Brown & Haley	Almond Roca	2.17894618...	-0.709939802	-0.905745549	-1.384057548	-0.604100289	1.7043669925
25	Leaf	Payday	-1.03820366	-0.11654157	0.9181584226	0.3901999298	-0.619118934	1.8882736842
26	Just Born	Super Hot Tamales	-1.773855619	0.0738312256	-0.966256028	-0.459550791	2.3183563781	-1.152360035
27	Hershey	Rolo	0.9688721...	-0.31897223	-1.013496932	0.1680075889	0.6262432512	0.3998547357
28	Nestle	Butterfinger	-0.634654261	-0.052788655	0.2607622535	-1.000334965	0.8045984078	0.8371455064
29	Myerson	Big Cherry	-1.307606046	-0.305654275	0.992054007	-0.398546814	-0.297098054	-0.966013832
30	Hershey	Mr. Goodbar	0.95624087...	-0.009059816	2.9772879634	-0.24565067	0.7329874301	-0.824398463
31	Hershey	Golden Collection	1.70814937...	0.1075222831	2.8867194443	1.4473855912	1.5158532145	-0.497074653
32	Annabelle	U-No (Green)	1.10075675	-0.756885565	0.151907856	-0.470214058	-0.868794175	-0.721445539
33	Hershey	Reese's Peanut B...	-0.073237481	-0.453822904	0.3339230578	-0.559538016	-1.032846231	1.4413851295
34	M&M/Mars	Skittles	-1.27579784	0.0167357887	-0.843695314	-0.764963391	2.5243407854	-1.455362508
35	Hershey	Reese's Peanut B...	-0.138361449	-0.527372106	1.1498129232	-0.292402422	-1.223806171	-0.126320593
36	Adams & Brooks	Cup O Gold	-0.118332057	-0.784725182	-0.778207111	-0.637840253	-1.067171527	-1.524989436
37	Nestle	Baby Ruth	-0.4688644164	-0.040835944	0.3462997324	-0.480026236	0.5641864914	1.1750528769
38	Weider	Tiger Milk	-0.896561745	-0.5215721	-0.205631414	3.618462935	-0.990556847	0.2909291115
39	M&M/Mars	Dove	0.02753657	3.2919870757	0.2455211084	-0.620663622	-0.342729256	-0.691543585
40	Hershey	York Peppermint ...	-1.259846847	-0.469068661	-0.573550082	-0.726836221	-0.164347088	-1.433448287
41	M&M/Mars	3 Musketeers	-0.443304241	-0.015413839	-0.595987897	-0.377843422	1.5032980999	0.670344596
42	Annabelle	U-No (Blue)	1.10075675	-0.756885565	0.151907856	-0.470214058	-0.868794175	-0.721445539
43	Tootsie	Jr Mints	-0.921412091	-0.439325009	-0.915493368	-0.84882174	0.4686772648	-1.515490479
44	Hershey	Symphony (Blue)	1.00692310...	-0.684052925	-0.187694204	0.5471636823	-0.760151887	-0.368504405
45	Hershey	KitKat	1.79991913...	0.1505706489	0.8310272029	0.3852963579	2.5287566965	-0.101047658
46	Leaf	Whoppers	0.65632156...	-0.360135637	-1.1843855	1.0549675437	0.7459484459	1.1453776973

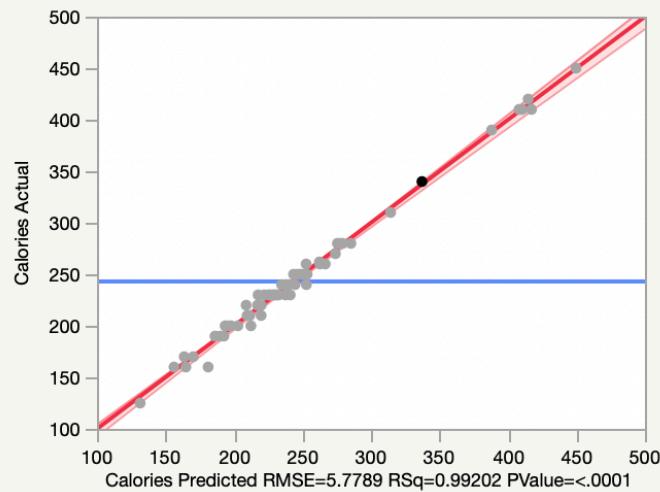
Using Snicker Peanut Butter by M&M/Mars, we can observe that the bar is very heart-stopping, not dense, very hearty, not at all protein-rich, not very carb-rich, and very salty.

Using Hershey's Almond Joy, we can observe that the bar is on the opposite end of the spectrum for most factors; it is not very heart-stopping, very densely packaged, mildly hearty, severely lacks protein richness, is very carb-rich and is not salty.

## Appendix G: Model & Effect Summary

### ▼ Whole Model

#### ▼ Actual by Predicted Plot



#### ▼ Effect Summary

Source	LogWorth	PValue
Factor3	59.603	0.00000
Factor5	54.057	0.00000
Factor1	49.463	0.00000
Factor6	18.702	0.00000
Factor4	13.085	0.00000
Factor2	3.423	0.00038

[Remove](#) [Add](#) [Edit](#)  FDR

Since all P Values are lower than .005, it is clear that the aforementioned 5 factors are all statistically significant for this model. This model illustrates a strong positive correlation between the factors, as demonstrated by the R squared value of .99202 listed underneath the graph of the multivariate function. Factors 6, 4 and 2 are less significant than 3, 5 and 1, but all 6 factors are clearly significant to predict the calories of a candy bar.

**Appendix H: Summary of Fit****▼ Summary of Fit**

RSquare	0.992016
RSquare Adj	0.991311
Root Mean Square Error	5.778879
Mean of Response	243.0267
Observations (or Sum Wgts)	75

With 75 observations, this model's R squared value of over .99 indicates a strong correlation between the factors as independent variables and calories as the dependent variable.

**Appendix I: Variance Analysis**

<b>Analysis of Variance</b>				
<b>Source</b>	<b>DF</b>	<b>Sum of Squares</b>	<b>Mean Square</b>	<b>F Ratio</b>
Model	6	282151.06	47025.2	1408.132
Error	68	2270.89	33.4	<b>Prob &gt; F</b>
C. Total	74	284421.95		<b>&lt;.0001*</b>

Not all of the means are equal, as indicated by a low P value of less than .005.

**Appendix J: Parameter Estimates**

<b>Parameter Estimates</b>				
<b>Term</b>	<b>Estimate</b>	<b>Std Error</b>	<b>t Ratio</b>	<b>Prob&gt; t </b>
Intercept	243.02667	0.667287	364.20	<.0001*
Factor1	29.313548	0.702344	41.74	<.0001*
Factor2	-2.516288	0.672425	-3.74	0.0004*
Factor3	41.972145	0.706675	59.39	<.0001*
Factor4	-6.367236	0.681672	-9.34	<.0001*
Factor5	32.952526	0.672337	49.01	<.0001*
Factor6	8.4804428	0.673793	12.59	<.0001*

Each parameter is statistically significant, as indicated by each variable possessing P values less than .005.

**Appendix K: Estimates****Prediction Expression**

243.02666667

+ 29.313547518 • Factor1

+ -2.516287503 • Factor2

+ 41.972144911 • Factor3

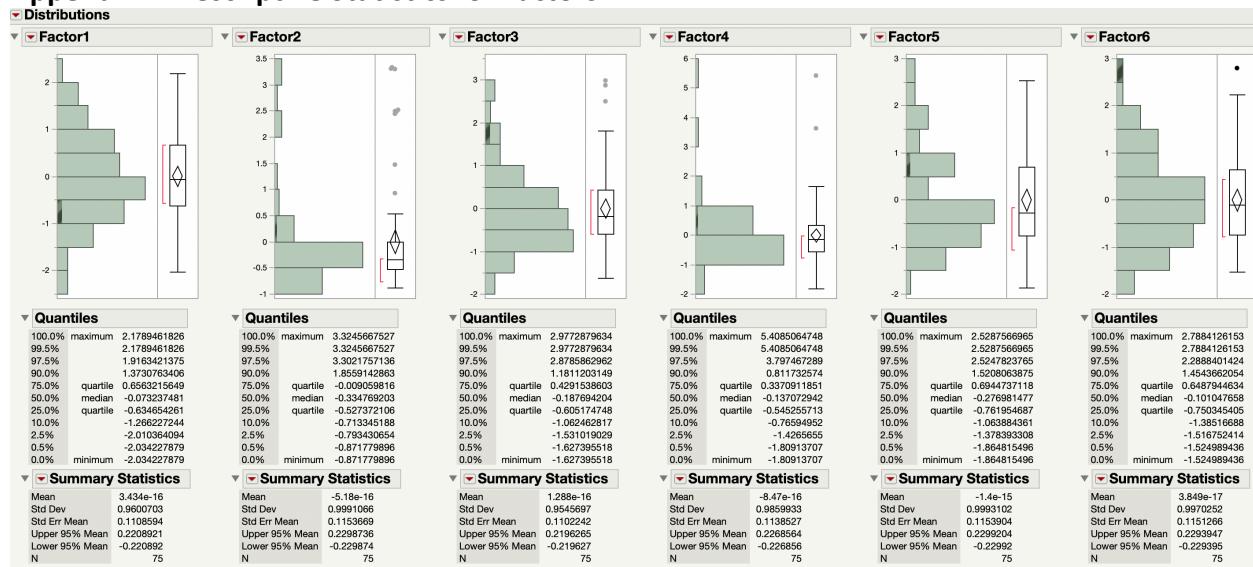
+ -6.367236328 • Factor4

+ 32.952526103 • Factor5

+ 8.4804427642 • Factor6

The multivariate model picture earlier in appendix G is made up of the above variable effects. Factors 1, 3, 5 and 6 increase the number of calories in the prediction, whereas factors 2 and 4 decrease the calories in the prediction.

## Appendix L: Descriptive Statistics for Factors



All factors follow a standard normal distribution, and the number of observations is consistent for each factor. Factor 1 has a leftward skew whereas the other 5 factors have rightward skews.

## Appendix M: Principal Components

