Min Luo

STA 5104

Are Expensive Healthy Cereals Worth It?

An Analysis on The Relationship between a Cereal's Price and Overall Rating

Introduction

Cereal is a staple food to many American households. In addition, its affordability and convenience make cereal the unsung hero in many college's students' diets. One could observe from a trip to the grocery store that the variety of cereal is overwhelming. How should a consumer, or more specifically, a college student with limited budget, know which one to purchase? Most consumers may fall for the Chivas Regal Effect where more expensive items are assumed to be better, though they do not know if costly cereals are necessarily better, until they've tried them all.

In this study, the relationship between the price and overall rating is analyzed. By knowing if there is a relationship between these two variables, consumers can make purchase decision based on the relationship, if one exists. For example, if there is a positive relationship between price and the overall rating, a consumer who wishes to buy the cereal with the highest quality could just look for the more expensive one.

Data

The data set used in this study is extracted from the Consumer Reports website. Consumer Reports is a non-for-profit organization dedicated to unbiased independent product testing. In the cereal product test, blind taste tests are conducted with trained testers. A sensory score and a nutrition score are given to each cereal, where the sensory score is based on sensory attributes like flavor and texture of the cereal, and the nutrition score is assigned based on the nutrition facts on the package. The scale of both scores is a 5-point scale, from 5 being excellent to 1 being poor. An overall score is computed based on the sensory and nutrition scores.

The cereals in this data set are cereals that are considered healthy, i.e. low in sugar and calories compared to other sugary cereals. Table 1 shows the variables and their description in the data set. This study focuses on the overall score and price per serving, where overall score is the dependent variable and price per serving is the independent variable.

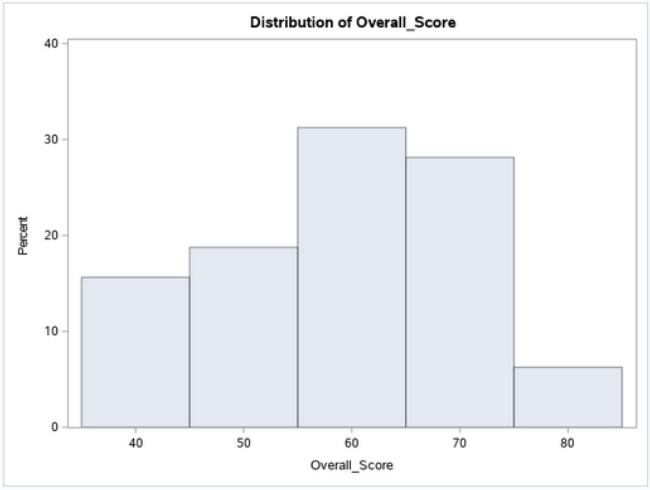
Table 1. Variables and description of Cereal data set

Variable	Description
Brand	Brand name and model of cereal
Price	Price of 1 whole unit of cereal
Overall Score	Score based on nutrient and sensory
Calories, Total Fat, Saturated Fat, Sodium, Fiber, Carbohydrates, Sugars, Protein	Information on nutrient facts label on package
Package Size	Weight of 1 unit of cereal in oz.
Price per Serving	Price by serving size

Analysis Result

The histogram of the shows that the Overall Score (Figure 1) is approximately normal; Result from Shapiro-Wilk test for normality (H_o: data is normally distributed, H_a: data is not normally distributed) shows a p-value of 0.7425, hence we fail to reject the null hypothesis, confirming that the data is normal. Price per Serving is skewed to the

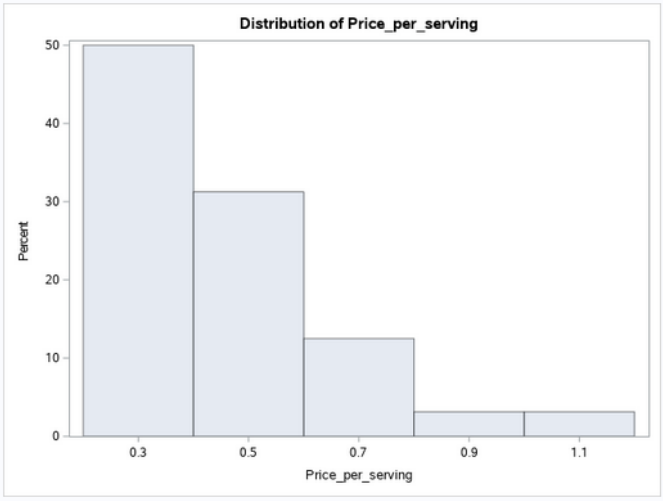
right, as shown in Figure 2, the result from Shapiro-Wilk test shows a p-value of 0.0020, indicating that the data is not normal.



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.978092	Pr < W	0.7425
Kolmogorov-Smirnov	D	0.082741	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.038063	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.248423	Pr > A-Sq	>0.2500

```
proc univariate data=cereal normal;
histogram Overall_score Price_per_Serving;
run;
```

Figure 1. Histogram of Overall Score (left) shows the variable is approximately normal



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.879912	Pr < W	0.0020
Kolmogorov-Smirnov	D	0.153882	Pr > D	0.0519
Cramer-von Mises	W-Sq	0.157435	Pr > W-Sq	0.0191
Anderson-Darling	A-Sq	1.023257	Pr > A-Sq	0.0094

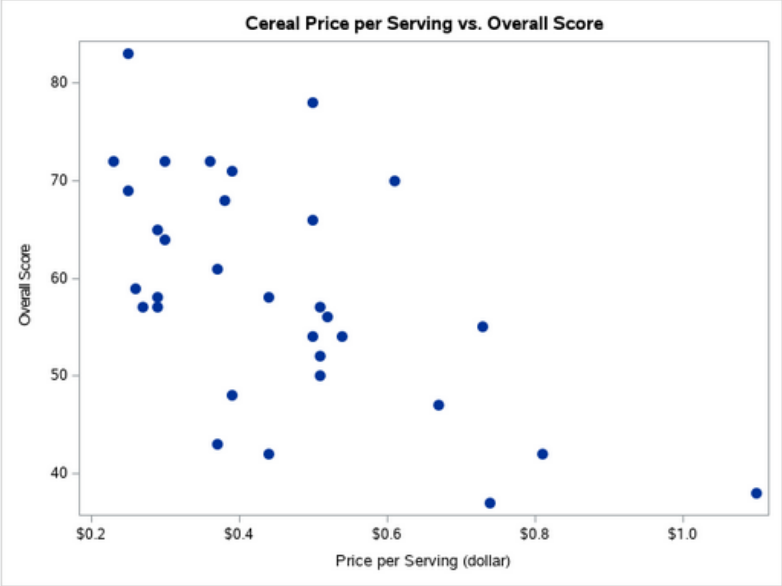
Figure 2. Histogram of Overall Score (left) shows the variable is right-skewed

A scatter plot is used to explore the relationship between Overall Score and Price per Serving as shown in Figure 3. The scatter plot shows that there is a negative curvilinear relationship between the two variables, i.e. as the price of the cereal increases, the overall score is expected to decrease. The Pearson and Spearman correlation coefficients are estimated to be -0.59315 and -0.59776. Result from both tests are significant with a p-value of 0.0003, we reject the null hypothesis that there is not a relationship between the two variables.

The negative association between Price per Serving and Overall Score suggests that expensive cereals tend to have low overall score, which is the opposite of the expectation. Now we shall consider the source of the low overall score. Recall that the overall score consists of two parts, sensory score and nutrition score. To study if there is a sensory difference between expensive and cheap cereals, the cereals are separated into two groups, based on their sensory score, where a sensory score of at least 4 is labelled high sensory, and a score below 4 is low sensory. A scatter plot of the Overall Score and Price per Serving is replotted with the data points colored based on their sensory score. As shown in Figure 5, there is not an obvious difference in price for the sensory scores. As shown

below in the test for normality, Overall Score is approximately normal, and Price per Serving is right-skewed. A log transformation is applied to Price per Serving. A normality test on the transformed Price per Serving shows that the transformed variable is normal. A t-test can now be applied, with the null hypothesis (H_o) being there is a difference between the mean price in high sensory cereals and low sensory cereals, and alternative hypothesis is there is not a mean price difference between the two groups.

The result of the t-test is shown in Figure 6. The t-test for the log transformed Price per Serving shows a p-value of 0.0230, hence we reject the null hypothesis, i.e. the result suggests that there is a difference in the mean price between high sensory cereals and low sensory cereals. In addition, the mean price is higher in the low sensory score group, as shown in the 95% confidence interval for the difference (Diff(1-2)) (-0.3173, -0.5877).




```
title "Cereal Price per Serving vs.
Overall Score";
proc sgplot data=cereal;
format Price_per_Serving dollar5.2;
scatter y=Overall_score
x=Price_per_Serving/
markerattrs=(size=10
symbol=Circlefilled);
run;
```

Figure 3. Scatter plot shows Overall Score and Price per Serving are negatively associated

The CORR Procedure

2 Variables:	Overall_Score Price_per_serving

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
Overall_Score	32	58.59375	11.69742	57.50000	37.00000	83.00000
Price_per_serving	32	0.45688	0.19430	0.41500	0.23000	1.10000

Pearson Correlation Coefficients, N = 32
Prob > |r| under H0: Rho=0

	Overall_Score	Price_per_serving
Overall_Score	1.00000	-0.59315 0.0003
Price_per_serving	-0.59315 0.0003	1.00000

Spearman Correlation Coefficients, N = 32
Prob > |r| under H0: Rho=0

	Overall_Score	Price_per_serving
Overall_Score	1.00000	-0.59776 0.0003
Price_per_serving	-0.59776 0.0003	1.00000

```
proc corr data=cereal pearson spearman;
var Overall_score Price_per_Serving;
run;
```

Figure 4. Pearson and Spearman correlation coefficients both show a non-zero correlation between Price per Serving and Overall Score

Figure 5. Scatter plot of Price per Serving and Overall Score with points colored based on their sensory score

SAS code for Figure 5

```
data cereal;
	if Sensory_Score >=4 then
		sen_grp=1;
	else
		sen_grp=2;
run;

title1 "Cereal Price per Serving vs. Overall Score";
title2 'Colored by Sensory';
proc sgplot data=cereal;
	format Price_per_Serving dollar5.2;
	scatter y=Overall_score x=Price_per_Serving/
		colorresponse=Sensory_Score
		markerattrs=(size=10 symbol=Circlefilled)
		filledoutlinedmarkers
		colormodel=TwoColorRamp;
		xaxis label='Price per Serving (dollar)';
		yaxis label='Overall Score';
run;
```

The TTEST Procedure

Variable: tPrice_per_Serving

sen_grp	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
High Sensory		12	-1.0583	0.3115	0.0899	-1.4697	-0.4943
Low Sensory		20	-0.7410	0.3891	0.0870	-1.3471	0.0953
Diff (1-2)	Pooled		-0.3173	0.3626	0.1324		
Diff (1-2)	Satterthwaite		-0.3173		0.1251		

sen_grp	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
High Sensory		-1.0583	-1.2562	-0.8604	0.3115	0.2206	0.5288
Low Sensory		-0.7410	-0.9231	-0.5589	0.3891	0.2959	0.5683
Diff (1-2)	Pooled	-0.3173	-0.5877	-0.0470	0.3626	0.2897	0.4846
Diff (1-2)	Satterthwaite	-0.3173	-0.5739	-0.0608			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	30	-2.40	0.0230
Satterthwaite	Unequal	27.358	-2.54	0.0172

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	11	1.56	0.4533

```
proc ttest data=cereal;
        format sen_grp sen_grp.;
        class sen_grp;
        var tPrice_per_Serving;
run;
```

Figure 6. T-test result showing the there is a difference in mean price between cereals with high sensory scores and the ones with low sensory scores

Conclusion

By analyzing the relationship between the price and overall score of different cereals, consumers can make purchase decision based on the relationship. The Pearson and Spearman correlation coefficients both show that there is a negative association between the two variables, indicating that more expensive cereals are expected to have a lower overall score on average. Furthermore, result from the t-test shows that expensive cereals tend to have a lower sensory rating. Hence, the overall results from this study suggests that, when it comes to healthy cereals, the cheaper ones are expected to have a higher quality.

Reference

"Cereals." *Healthy Cereal Ratings*, www.consumerreports.org/products/healthy-cereal/ratings-overview/.