Reckless or Responsible: A Multivariate Statistical Analysis of Consumer Spending

Emilola Abayomi *　　Erin Esp †　　Shannon Grant ‡

Abstract

As Americans spend more and save less, there is a need to evaluate variables which influence spending habits. First, we reduce the number of variables with Principal Components analysis and identify underlying factors by grouping correlated variables in Factor Analysis. Finally, we use Discriminant Analysis to develop a rule for classifying individual consumers as reckless or responsible.

Introduction

Money: is it really the path to destruction? For some, money can buy shoes, fast cars, and maybe happiness, but for others it can only buy more debt with no way out. For years, newspaper and magazine headlines have decried the spending habits of the American people. In 2005, for the first time since the Great Depression, consumers actually spent more than they made and the individual savings rate plunged to -0.5% [Bru06]. Why do people choose to live beyond their means?

In this study, we use multivariate statistical techniques to explore data from the 2004 Consumer Expenditure Survey by the Bureau of Labor Statistics. [Div06] Initially we were interested in twenty-one variables, including both consumer unit characteristics and expenditures expressed as a percentage of income. See Table 1 for a list of variables. By employing Principal Components Analysis, we reduced the dimensionality of the data set to five.

*Albany State University, Albany, GA; emilola.abayomi@asurams.edu
†Montana Tech, Butte, MT; jeesp@mtech.edu
‡University of Idaho, Moscow, ID; sgrant@uidaho.edu

We also used Factor Analysis to group correlated variables and draw conclusions about relationships between variables. Finally, we used Discriminant Analysis to classify individual families, known as consumer units, as responsible or reckless.

Before moving on, we must define our vector of observations $\underline{x}$. Let

$$\underline{x}^T = [x_1, x_2, \ldots, x_p] \sim MN(\underline{\mu}, \mathbf{\Sigma}),$$

where x_1, x_2, ..., x_p are the p variables recorded for each consumer unit.

Assessing Normality

Our data is in the form of a random sample of size n from a multivariate population. The multivariate statistical techniques we use are based on the assumption that the data comes from a multivariate normal distribution. To test this assumption, we examine the univariate normality of each of the p variables with a Quantile-Quantile (Q-Q) plot. To construct a Q-Q plot for the sample on a given variable x, order the n observations such that $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ and calculate the corresponding standardized quantiles $z_{(j)}$ for j=1,...,p. Plot the ordered pairs ($x_{(j)}$, $z_{(j)}$). Solve the standardizing equation for x to find the equation of the Q-Q plot line:

$$x_{(j)} = \mu + \sigma z_{(j)}$$

where $x_{(j)}$ is the sample data point and $z_{(j)}$ is the corresponding standardized z quantile. If the plot is linear, we accept the assumption that the variable is normally distributed. If the variable is not normally distributed but is essential to our analysis, we will apply a transformation to make our data more "normal-looking" [JW02]. A more rigorous assessment of linearity is a hypothesis test of the correlation coefficient between $x_{(j)}$ and $z_{(j)}$. The null hypothesis (H_0) is that the population correlation coefficient ρ=1, thus the population comes from a normal distribution. The alternate hypothesis (H_A) is that the population does not come from a normal distribution, or 0<ρ<1. The test statistic is the sample correlation coefficient, defined as

$$r_{(x_j),(z_j)} = \frac{\text{Cov}\,(x_j, z_j)}{\sqrt{\text{Var}(x_j)\text{Var}(z_j)}}.$$

H_0 is rejected at the significance level α if $r_{(x_j),(z_j)}$ is less than the critical value c, which is defined by

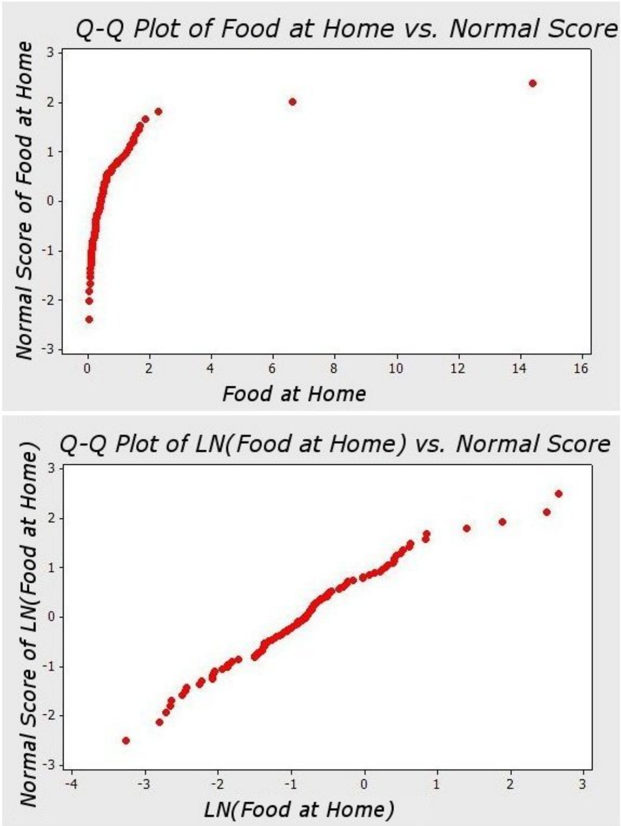
$$\text{P}(r < c|H_0) = \alpha.$$

Depending on the importance of the variable, we may include it in our analysis even if H_0 is rejected.

Table 1: Variables

Variable	Transformation	Description
Sex	none	Sex of reference person.
Region	none	Region of residence. Categorical: Northeast, South, Midwest, West.
Age	none	Age of reference person in years.
Tenure	none	Ownership status of housing: rented or owned.
Education	none	Education of reference person in years.
Race	none	Race of reference person. Categorical: White, Black, Native American, Asian, Pacific Islander, Multi-Race.
Income	Square root	Yearly income of consumer unit.
Food at Home	Natural log	Per person, percentage of income.
Transportation	Natural log	Percentage of income.
Food away from Home	Natural log	Per person, percentage of income.
Housing	Natural log	Percentage of income.
Cash contributions	Square root	Percentage of income.
Health Insurance	Square root	Percentage of income.
Tobacco	Square root	Percentage of income.
Health Care	Square root	Percentage of income.
Alcohol	Square root	Percentage of income.
Savings	Square root	Amount of money in savings account at end of last month, percentage of income.
Apparel	Square root	Per person, percentage of income.
Entertainment	Square root	Per person, percentage of income.

Figure 1: Top: Plot of Food At Home vs. corresponding normal score. Bottom: Plot of LN(Food At Home) vs. corresponding normal score.



Normality Results

For this data set we used a significance level α=0.01, sample size n=75, and we rejected H_0 if r<0.9771. Of the twenty-two original variables, five were normal. We rejected H_0 for Race; however, for the reasons stated above we neither transformed nor discarded the variable. We applied transformations to the other sixteen variables and failed to reject H_0 for five. Of the eleven remaining variables, we chose to discard Education and Trips as they were not crucial to our analysis. After discarding the two, we were left with nine transformed variables which were essential and approximately normal. See Appendix A for a list of variables and corresponding r values.

A visual example of the difference between a variable and a transformation of the variable can be displayed by Q-Q plots, as illustrated in Figure 1. Note the linearity of the second plot as compared to the nonlinearity of the first plot.

Principal Components Analysis

Principal Components Analysis (PCA) is a multivariate technique used to reduce the dimensionality of the data set by constructing linear combinations of x_1, x_2,...,x_p. Let the first Principal Component (PC) Y_1 be defined by

$$Y_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p = \underline{l_1}^T \underline{x},$$

where we choose $\underline{l_1}$ such that Y_1 accounts for the maximum variance out of the total variation. The second PC Y_2 is defined by

$$Y_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p = \underline{l_2}^T \underline{x},$$

where we choose $\underline{l_2}$ such that $\underline{l}_2^T\underline{l}_2 = 1$ and $\underline{l}_2^T\underline{l}_1 = 0$. In a similar way, we construct m components, where m is generally much smaller than p. We select m such that Y_1, Y_2,...,Y_m account for the majority of the variance. The variance contribution of the m PC's is

$$\frac{\sum_{i=1}^{m}\lambda_i}{\sum_{j=1}^{p}\lambda_j},$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ are the characteristic roots of the covariance matrix $\mathbf{\Sigma}$. A scree plot is a visual expression of the fraction of total variance in the data explained by each PC. To construct a scree plot, let the x-coordinates be i=1, 2, ..., p and let the y-coordinates be the corresponding λ_i's. Locating an elbow on the scree plot is an additional criterion for choosing m.

Principal Components Analysis Application

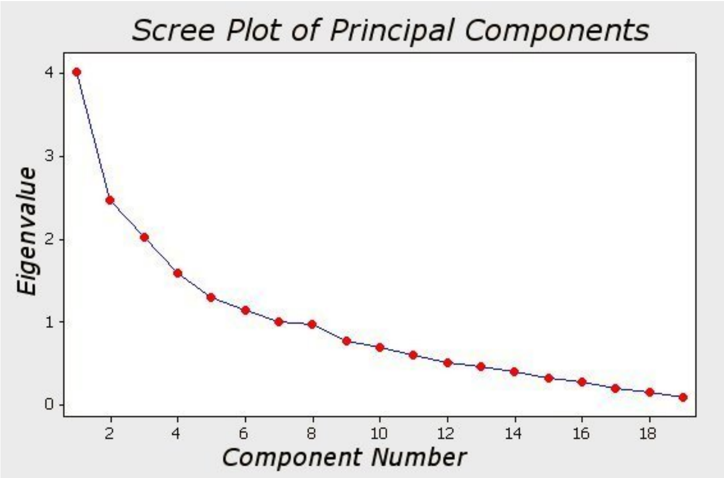
Using PCA, we were able to reduce the dimensionality of the data set to five components. Since there is no obvious elbow in the scree plot and hence no clear choice of m, and we chose to examine the contribution of each PC to the total variation. See Figure 2 for the scree plot.

Five PC's accounted for about 61% of the total variation. See Table 2 for the contribution of each PC to the total variation.

Table 2: PC Contributions

	PC 1	PC 2	PC 3	PC 4	PC 5
Eigenvalue	4.0105	2.5953	1.9544	1.5871	1.4021
Proportion	0.211	0.137	0.103	0.084	0.074
Cumulative	0.211	0.348	0.451	0.534	0.608

Figure 2: Scree plot for PCA.



As the first PC contrasted income with spending on food and housing, we chose to call it *Necessities.* The second PC emphasized age, health care, and health insurance, so we named it *Health and Aging.* The third PC compared age and health insurance with alcohol, apparel, and entertainment. We labeled this PC *Affluence.* PC 5 contrasted sex, apparel, health care, and health insurance with food at home, cash contributions, and savings, so we labeled it *Gender Attitudes.* The fourth PC did not suggest understandable relationships, so we left it unnamed. See Appendix B for the five PCs.

Factor Analysis

Factor Analysis (FA) is another multivariate technique commonly used to reduce the dimensionality of the data set. It involves writing each variable x_i as a linear combination of m common factors and a unique factor. Examination of these linear combinations reveals unobservable factors. We use the m-factor model, as defined by

$$x_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1$$

$$x_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2$$

$$\vdots$$

$$x_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p.$$

Here F_1, F_2, $\ldots$, F_m are unobservable common factors; ϵ_1, ϵ_2, $\ldots$, ϵ_p are unobservable unique factors; and l_{ij} is the factor loading of the j^{th} common factor on the i^{th} variable. The factor loading l_{ij} reflects the importance of the j^{th} common factor in the composition of the i^{th} variable. The m-factor matrix model $\underline{x}$ - $\underline{\mu} = \mathbf{L}\ \underline{F} + \underline{\epsilon}$ is based on the following assumptions:

1. The vector $\underline{F}$ of common factors is distributed multivariate normal with $E(\underline{F})=\underline{0}$ and $Cov(\underline{F})=\mathbf{I_m}$.

2. The vector $\underline{\epsilon}$ of uniqute factors is distributed multivariate normal with $E(\underline{\epsilon})=\underline{0}$ and $Cov(\underline{\epsilon})=\mathbf{\Psi}$, where $\mathbf{\Psi}$ is a diagonal matrix whose i^{th} diagonal element Ψ_i is the specificity (variance) of the i^{th} unique factor.

3. The vectors $\underline{F}$ and $\underline{\epsilon}$ are independent.

With these assumptions we can prove that $\mathbf{\Sigma}=\mathbf{L}\ \mathbf{L^T}+\mathbf{\Psi}$, where $\mathbf{\Sigma}$ is the covariance matrix. The proportion of total variance explained by the j^{th} factor is

$$\frac{\lambda_j}{\sum_{i=1}^{p}\lambda_i}$$

where λ_j is the j^{th} largest characteristic root of the covariance matrix $\mathbf{\Sigma}$. The goal of FA is to obtain $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$, the maximum likelihood estimates of $\mathbf{L}$ and $\mathbf{\Psi}$, and to interpret the factors.

We test the adequacy of the m-factor model using the null hypothesis H_0: $\mathbf{\Sigma} = \mathbf{L}\mathbf{L^T} + \mathbf{\Psi}$ and the alternative hypothesis H_A: $\mathbf{\Sigma}$ is any other positive definite matrix. Reject H_0 if $\chi^2 > \chi^2_{\alpha,\nu}$ where

$$\nu = \frac{1}{2}\left((p-m)^2 - p - m\right).$$

Our test statistic is

$$\chi^2 = \left[n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}m\right] \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathbf{T}} + \hat{\mathbf{\Psi}}|}{|\mathbf{S}|}.$$

Beginning with m=1, we test H_0. If H_0 is not rejected, we accept the 1-factor model. If H_0 is rejected, we increase m until H_0 is accepted.

Factor Analysis Application

We decided to use a six-factor model based on the proportion of variance explained by the factors, as the test of H_0 versus H_A was inconclusive for our m-factor model. Our first factor showed a high correlation among food and housing versus income, so we labeled this factor *Necessities*. The *Necessities* factor accounted for about 15% of the variance. The second factor did not

Table 3: Factor Contributions

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Variance	2.8333	1.9149	1.8856	1.4926	1.2589	0.9081
Proportion	0.149	0.101	0.099	0.079	0.066	0.048
Cumulative	0.149	0.25	0.349	0.428	0.494	0.542

suggest a meaningful relationship between variables so we left it unnamed. Factor 2 accounted for about 10% of the variance. We named factor 3 *Attitude Toward Health* as it contrasted health care and health insurance with tobacco; furthermore, factor 3 contributed 10% of the variance. Factor 4 contributed 8% of the variance and showed that alcohol, apparel and entertainment were highly correlated; thus we named this factor *Lifestyle*. We gave factor 5 the name *Age and Money Management* due to the high loadings on age, food away from home, cash contributions and savings. Factor 5 accounted for 6.6% of the variance. The high factor loading for education in factor 6 suggests the name *Education*; this factor contributed 5% of the total variance. See Appendix C for our six factors, which account for 54% of the total variance, and Table 3 for the contribution of each factor to the total variance.

Discriminant Analysis

Discriminant Analysis (DA) is used to classify an observed $\underline{x}$ into one of two populations, π_1 or π_2. We accomplish this goal by using the discriminant function, a linear function of $\underline{x}$ which maximizes the distance between the two populations. We assume the populations are multivariate normal; that is, $\pi_1 \equiv \text{MN}(\underline{\mu_1}, \mathbf{\Sigma_1})$ and $\pi_2 \equiv \text{MN}(\underline{\mu_2}, \mathbf{\Sigma_2})$ with $\mathbf{\Sigma_1} = \mathbf{\Sigma_2}$. Assume that we use samples of sizes n_1 and n_2 from π_1 and π_2, respectively. These are the training samples. The training samples are used to calculate the estimates of the mean and variance for each population, which are $\overline{\underline{x}}_1$, $\overline{\underline{x}}_2$, $\mathbf{S}_1$,and $\mathbf{S}_2$, .

Before using DA, we must determine if the covariance matrices are equal. Our null hypothesis H_0: $\mathbf{\Sigma_1} = \mathbf{\Sigma_2}$ will be tested against the alternate hypothesis H_A: $\mathbf{\Sigma_1} \neq \mathbf{\Sigma_2}$ using a likelihood ratio test. The test statistic is $\frac{M}{c}$, which is distributed χ^2_ν with ν degrees of freedom. Here, $\nu = \frac{1}{2}p(p+1)$ and p is the number of variables. The number M is defined by

$$M = \ln|\mathbf{S}| \sum_{i=1}^{2}(n_i - 1) - \sum_{i=1}^{2}\left[\ln|\mathbf{S_i}|(n_i - 1)\right],$$

where S_i is the sample covariance matrix, or estimate of Σ_i for i = 1,2 and S is the pooled sample covariance matrix as defined by

$$\mathbf{S} = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{2} (n_i - 1)\mathbf{S_i} \right).$$

Also, $\frac{1}{c}$ is defined by

$$\frac{1}{c} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)} \left[\left(\sum_{i=1}^{2} \frac{1}{n_i - 1} \right) - \frac{1}{n_1 + n_2 - 2} \right].$$

If $\frac{M}{c} < \chi^2_{\alpha,\nu}$, we fail to reject H_0. Therefore, we assume $\mathbf{\Sigma_1} = \mathbf{\Sigma_2}$ and we move on to classify observations in our training sample into π_1 or π_2 using the linear discriminant function.

The training sample contains observations which are already classified. By reclassifying them, we test the accuracy of our DA technique. The classification rule is as follows:

$$\begin{array}{l} \text{classify } \underline{x} \text{ into } \pi_1 \text{ if } \underline{\hat{a}}^T \ \underline{x} > \hat{h}, \\ \text{classify } \underline{x} \text{ into } \pi_2 \text{ if } \underline{\hat{a}}^T \ \underline{x} \le \hat{h}, \end{array}$$

where

$$\underline{\hat{a}} = S^{-1}(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)$$

and

$$\hat{h} = \frac{1}{2}(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)^T S^{-1}(\underline{\overline{x}}_1 + \underline{\overline{x}}_2).$$

The above classification rule is equivalent to classifying $\underline{x}$ into π_1 if the squared M-distance of $\underline{x}$ from π_1 is less than the squared M-distance of $\underline{x}$ from π_2. That is, we classify $\underline{x}$ into π_1 if

$$(\underline{x} - \underline{\overline{x}}_1)^T \mathbf{S}^{-1} (\underline{x} - \underline{\overline{x}}_1) < (\underline{x} - \underline{\overline{x}}_2)^T \mathbf{S}^{-1} (\underline{x} - \underline{\overline{x}}_2);$$

otherwise we classify $\underline{x}$ into π_2.

To evaluate the accuracy of the classification scheme, we use the training sample and calculate the Apparent Error Rate (APER), the ratio of misclassifications to the number of observations. Another technique for examining the error rate of the DA is calculating the Total Probability of Missclassification (TPM). First we determine the Mahalanobis distance, or M-distance (Δ_p), between the two populations, where

$$\hat{\Delta}_p^2 = (\underline{\overline{x}}_1 - \underline{\overline{x}}_2)^T \mathbf{S}^{-1} (\underline{\overline{x}}_1 - \underline{\overline{x}}_2).$$

Thus

$$TPM = \alpha = 2\Phi \left(-\frac{1}{2} \hat{\Delta}_p \right).$$

Table 4: Linear DA, Training Samples

	Classified: 1	*Classified: 2*	*Total*
True: 1	21	2	23
True: 2	3	21	24
Correct	21	21	42

Table 5: Linear DA, Test Sample

	Classified: 1	*Classified: 2*	*Total*
True: 1	12	1	13
True: 2	9	3	12
Correct	12	3	15

Discriminant Analysis Application

We divided our consumers into two populations based on the median of the square root of savings. Consumers with savings greater than the median were classified into π_1, the responsible group, while those with savings less than the median were classified into π_2, the reckless group. In our training sample, we set n_1=25 and n_2=25. Our testing sample was made up of the remaining 25 observations.

We tested H_0: $\mathbf{\Sigma_1} = \mathbf{\Sigma_2}$ against H_A: $\mathbf{\Sigma_1} \neq \Sigma_2$ to determine if the linear method was appropriate. Our test statistic $\frac{M}{c}$ was 126.371 with a p-value of 0.996, and our critical value χ^2 with 171 degrees of freedom and α=0.01 was 216.938. Since $\frac{M}{c} < \chi^2$, we accepted H_0 and concluded that the linear method could be used.

We ran linear DA on our training samples and our results are in Table 4. Forty-two of forty-seven observations were classified correctly, giving APER=0.106 and TPM=0.014. Note that three of the fifty observations from the training samples were not analyzed; this occurred due to missing values in these observations.

Linear DA was less successful in classifying our testing sample. Fifteen of twenty-five observations were classified correctly, giving APER=0.40. See Table 5 for these results.

Conclusions

We began this study with twenty-one variables and, after assessing normality and applying transformations, we were left with nineteen variables. Using PCA, we were able to reduce the number of variables to five PC's and account for 61% of the total variation. This model was satisfactory, as we were able to interpret and name four out of five components. Our other method for reducing dimensionality was FA, where we used a six-factor model which accounted for 54% of the variance. Though we could identify and name five of the six factors, we were not satisfied with this technique due to the inconclusiveness of the hypothesis test. Finally, we used linear DA to classify individual consumer units as reckless or responsible spenders. The training samples were classified successfully, with an APER of 11%. We were less happy with the DA results for the testing sample.

We were pleased to note the similarities in our components and factors; age, spending on necessities, and lifestyle choices were apparent influences in both models. Our most intriguing observation was how income and spending habits influenced savings. Surprisingly few consumers saved anything at all; however, age and savings were highly correlated, which we expected.

Acknowledgements

First we would like to thank Miami University, the NSF, and the NSA for sponsoring SUMSRI. Next we would like to thank the SUMSRI co-directors, Dr. Davenport and Dr. Waikar, for teaching and guiding us. We would also like to thank Shelly-Ann Meade, our graduate assistant, for her patient help. We would like to thank Dr. Farmer for helping us through the writing and revision process. We would especially like to thank Nathan Bialke of the University of Idaho for assistance in all matters technical. Finally, we would like to thank our fellow SUMSRI students for seven long weeks of memories.

References

[Bru06] Laura Bruce, *Negative personal savings rate: What does it mean?*, From Bankrate.com [Online], 2006, Available: http://www.bankrate.com/brm/news/sav/20060308a1.asp [Accessed 16 July 2006].

[Div06] Division of Consumer Expenditure Surveys, Branch of Information and Analysis, Bureau of Labor Statistics, *Consumer Expenditure Survey*, Data CD, 2006, Available from the BLS.

[JW02] Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, fifth ed., Prentice Hall, 2002.

Appendix A

Normality Assessment Results

Variable	r
Sex	1.00
Region	0.996
Age	0.992
Tenure	0.988
Education	0.983
Race	0.970
Natural log of Food Away from Home	0.994
Natural log of Transportation	0.994
Natural log of Food at Home	0.989
Natural log of Housing	0.988
Square root of Income	0.974
Square root of Cash Contributions	0.972
Square root of Health Insurance	0.971
Square root of Tobacco	0.955
Square root of Health Care	0.934
Square root of Alcohol	0.934
Square root of Savings	0.914
Square root of Apparel	0.911
Square root of Entertainment	0.908
Square root of Trips	0.918
Square root of Education	0.746

Appendix B

Five-Component PC Model

Variable	*PC 1*	*PC 2*	*PC 3*	*PC 4*	*PC 5*
Region	0.045	0.235	0.113	0.137	0.258
Tenure	0.275	0.157	-0.199	-0.078	0.035
Sex	-0.146	0.189	-0.200	-0.163	-0.327
Race	0.038	0.024	-0.254	-0.481	0.126
Age	-0.090	-0.372	-0.301	0.260	0.118
Education	-0.102	-0.136	0.178	-0.434	0.061
Sqrt Income	-0.360	-0.132	0.309	-0.119	-0.003
Ln Food at Home	0.448	0.036	-0.155	-0.002	-0.033
Ln Food Away from Home	0.340	-0.131	0.089	-0.126	0.290
Sqrt Alcohol	0.252	0.029	0.486	0.094	0.020
Ln Housing	0.386	0.127	-0.144	-0.077	0.032
Sqrt Apparel	0.203	-0.157	0.382	-0.056	-0.350
Ln Transportation	0.080	0.170	-0.042	-0.468	0.060
Sqrt Health Care	0.188	-0.419	-0.102	-0.079	-0.387
Sqrt Health Insurance	-0.001	-0.386	-0.285	0.035	-0.258
Sqrt Entertainment	0.332	-0.239	0.241	0.069	-0.037
Sqrt Cash Contributions	0.061	-0.307	-0.072	-0.212	0.272
Sqrt Tobacco	0.157	0.237	-0.177	0.341	-0.065
Sqrt Savings	-0.027	-0.298	-0.048	0.144	0.533

Appendix C

Six-Factor FA Model

Variable	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*
Region	0.079	0.025	0.292	-0.106	-0.083	-0.031
Tenure	0.416	0.287	0.091	0.095	-0.125	0.152
Sex	-0.194	-0.013	0.183	0.183	-0.180	0.204
Race	0.153	-0.076	-0.184	0.256	-0.113	0.314
Age	-0.290	0.185	-0.244	0.325	0.596	-0.215
Education	-0.219	-0.078	-0.166	-0.326	0.051	0.563
Sqrt Income	-0.720	-0.372	-0.071	-0.399	0.085	0.115
Ln Food at Home	0.786	0.514	-0.049	0.059	0.025	0.040
Ln Food Away	0.635	0.178	-0.133	-0.299	0.416	0.202
Sqrt Alcohol	0.374	0.105	0.021	-0.650	-0.037	-0.233
Ln Housing	0.661	0.362	0.064	0.052	-0.101	0.101
Sqrt Apparel	0.184	0.200	-0.372	-0.483	-0.223	-0.278
Ln Transportation	0.247	-0.105	0.040	0.013	-0.133	0.252
Sqrt Health Care	-0.000	0.580	-0.815	-0.000	-0.000	0.000
Sqrt Health Insurance	-0.272	0.398	-0.463	0.232	0.160	0.059
Sqrt Entertainment	0.416	0.329	-0.285	-0.431	0.157	-0.191
Sqrt Cash	0.043	0.028	-0.190	-0.075	0.454	0.229
Sqrt Tobacco	0.000	0.735	0.679	-0.000	-0.000	0.000
Sqrt Savings	-0.013	-0.126	-0.106	0.066	0.563	-0.166