Predictors of Hypertension and Prehypertension in Cal Poly Students

A Senior Project

Presented to

The Faculty of the Statistics Department

California Polytechnic State University, San Luis Obispo

In Partial Fulfillment

Of the Requirements for the Degree

Bachelor of Science

By

Toria Mock

TABLE OF CONTENTS

I. INTRODUCTION

I first became interested in applications of biostatistics when I participated in Boston University's Summer Institute for Training in Biostatistics in the summer of 2010. Analysis of FLASH data seemed a perfect outlet for me to put to use what I had learned through analyzing the Framingham Heart Study during the program.

My research project analyzes the contributing factors to hypertension and prehypertension in college students. This paper will describe the research I have done and all statistically significant factors that are found to be associated with hypertension/prehypertension in Cal Poly students. In Section II, the reader will find a description the data set and statistical methods I used, as well as some descriptive statistics of the sample. Section III discusses all significant analysis, and Section IV concludes the paper.

II. METHODS AND MATERIALS

Data

FLASH is a longitudinal study of the overall health of college students at Cal Poly. The physical assessment and online questionnaire capture physical indicators of overall health, such as blood pressure, BMI and resting heart rate, as well as mental indicators such as stress. The first cohort was surveyed in Fall 2009. The questionnaire is sent electronically to freshmen each fall and spring. Both the questionnaire and physical assessments are conducted on a voluntary basis. Each respondent is assigned a unique identification number by the data management team, which is used to correlate their physical assessment with their questionnaire, and differentiate them from other respondents. I have only considered those students who completed both a questionnaire and a physical assessment in my data set, regardless of the time it was completed. Therefore, the data I analyzed contains entries from Fall 2009, Spring 2010, and Fall 2010 and contains 716 respondents. Three of the 716 had no gender classified, and were therefore removed, leaving a data set of 713 observations. If a student completed both the questionnaire and assessment more than once during this time, only their first response was included in the data set.

The online questionnaire has six sections. These are: sociodemographic information, health perceptions and lifestyles, physical activity and exercise habits, stress, dietary habits and dining patterns, and sun exposure/sun block use. The first section includes questions pertaining to sex, race, parents' education and hometown. The second asks respondents to rank their overall health, and describe their alcohol and cigarette use. The next section asks respondents to rank their physical activity levels. The fourth section asks the student to rank their stress and mental health. The next section asks respondents to describe their eating habits, specifically how often they eat certain foods. Finally, the last section asks the student to rank the amount of time they spend outside and use of sun block.

After completing the online questionnaire, students are highly recommended to participate in the physical assessment. The physical assessment measures physical indicators of overall health. This includes blood pressure, BMI, and bodyfat percentage, among others. My data analysis primarily uses measures from the physical assessment. I also created a stress score from section four of the questionnaire. I analyzed the association between these variables and the probability of hypertension or prehypertension in college students.

The National Heart, Lung, and Blood Institute defines hypertension and prehypertension according to Table 1.

Table 1: Blood Pressure Class

Category	Systolic (top number)		Diastolic (bottom number)
Normal	Less than 120	*And*	Less than 80
Prehypertension	120–139	*Or*	80–89
Hypertension			
Stage 1	140–159	*Or*	90–99
Stage 2	160 or higher	*Or*	100 or higher

Blood pressure measures the force pushing outwards on arterial walls. Systolic blood pressure is defined as "the pressure when the heart beats while pumping blood". Diastolic blood pressure is "the pressure when the heart is at rest between beats". Hypertension and prehypertension are known to increase the risk of blood clots, plaque buildup, and tissue/organ damage from blocked arteries. For the first part of my research, students were classified into two blood pressure classes – normal and high. The high blood pressure class includes both prehypertension and hypertension stages one and two.

The stress score variable was created from responses to section four of the questionnaire. "A Global Measure of Perceived Stress," written by Sheldon Cohen, Tom Kamarck and Robin Mermelsteing and published in the Journal of Health and Social Behavior in 1983, defines the items and instructions for a perceived stress scale. The participant was asked to answer questions such as, "In the last month, how often have you been upset because of something that happened unexpectedly?" and asked to describe their response on a scale of zero (never) to four (very often). Descriptive statistics of the stress score in Cal Poly students can be found later in Section II.

The physical assessment score was created using the International Physical Activity Questionnaire (IPAQ) scoring criteria. The participants estimated the time they spent walking and exercising on a daily basis in the online questionnaire. These responses were used to classify respondents into high, moderate or low levels of physical activity. Descriptive statistics can be found in the last part of Section II.

Statistical Methods

I used logistic regression to model the probability of hypertension and prehypertension in Cal Poly students. A large part of my senior project was research on logistic regression and SAS, as I am not a statistics major.

Logistic regression is used for predicting the outcome of a binary dependent variable (in this case, hypertension/prehypertension vs. normal blood pressure). Logistic regression is similar to linear regression in that one or more independent variables, categorical or continuous, are used to predict an outcome. In contrast to linear regression, which models the expected value of a continuous outcome variable, logistic regression models the probability of an event. The regression model in my research modeled the probability of hypertension/prehypertension, compared to that of normal blood pressure.

Because logistic regression is modeling the probability of an outcome, the logistic function is used to ensure the model will only have expected values between zero and one.

The logistic function and the definition of variable z:

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$z = ln(\frac{P(hypertension/prehypertension)}{P(normal)}) = \alpha + \beta_1(gender) + \beta_2(BMI) + \dots$$

The variable z is such that α represents the background log odds, or the log odds of prehypertension/hypertension in a given respondent if there were no predictor variables. $\beta_1, \dots, \beta_k$ are the regression coefficients of the corresponding predictor variable, and represent the change in log odds that would result from a one-unit change in the corresponding predictor variable.

As shown in the appendix (Section V), SAS was used for logistic regression analysis and for most preliminary analysis. Specifically, the Logistic Procedure (PROC LOGISTIC, SAS 9.2 ®) was used for regression.

Preliminary Analyses and Descriptive Statistics

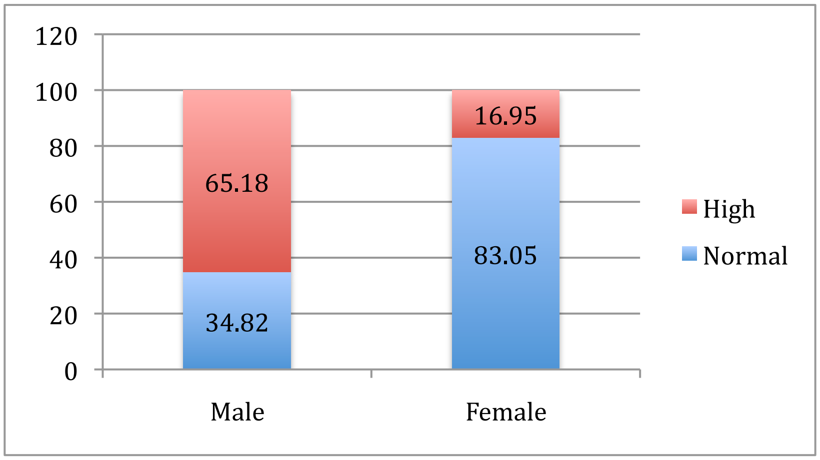
Descriptive statistics for this sample were generated in SAS. The data set has a total of 713 observations, including 359 males and 354 females. The data set was split into two classifications of blood pressure. Respondents classified as having high blood pressure were either hypertensive or prehypertensive. The remaining respondents had normal blood pressure levels. Of the 713 observations, 294 students were hypertensive or prehypertensive, and 419 had normal blood pressure levels.

Table 2 and Figure 1 show the distribution of blood pressure class by gender. A significantly higher proportion of males in this study are hypertensive/prehypertensive than females.

Table 2: Distribution of Blood Pressure Class by Gender

	Male		Female	
BP Class	*Frequency*	*Percent*	*Frequency*	*Percent*
Normal	125	34.82	294	83.05
Hypertension/Prehypertension	234	65.18	60	16.95

Figure 1: Distribution of Blood Pressure Class by Gender



Descriptive statistics for the explanatory variables of interest follow below.

Table 3: BMI Statistics

BMI (kg/m^2)	Male	Female	Overall
Sample Size	359	354	713
Mean	22.34	23.45	22.90
Std. Dev.	3.323	3.029	3.225

Figure 2: Distribution of BMI

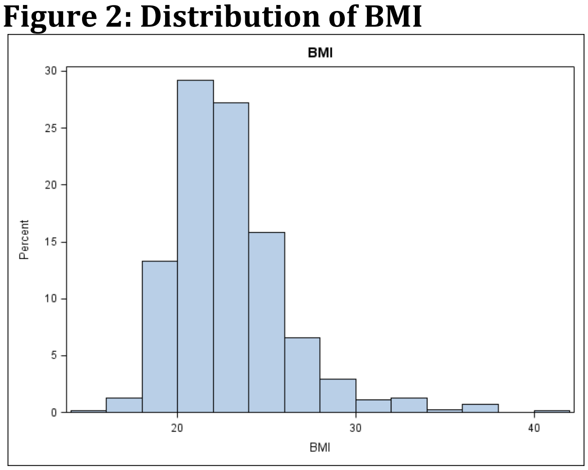


Figure 3: Distribution of BMI by Gender

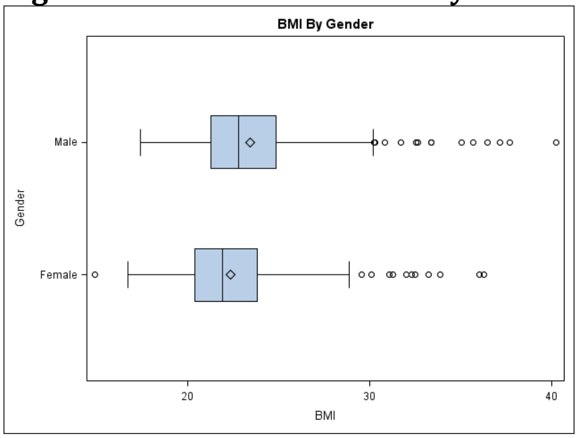


Table 4: Bodyfat Statistics

Bodyfat percentage	Male	Female	Overall
Sample Size	359	353	712
Mean	19.422	30.066	24.70
Std. Dev.	5.51	5.307	7.59

Figure 4: Distribution of Bodyfat

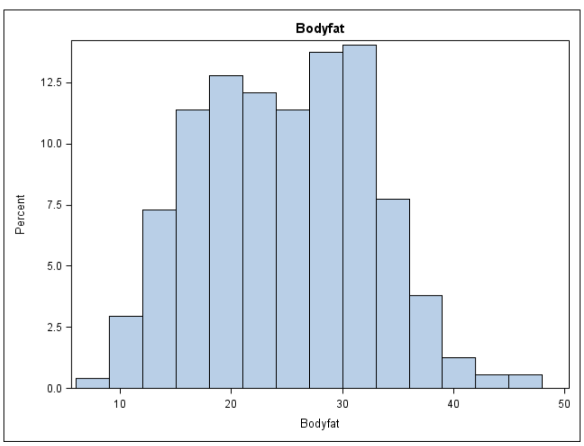


Figure 5: Distribution of Bodyfat by Gender

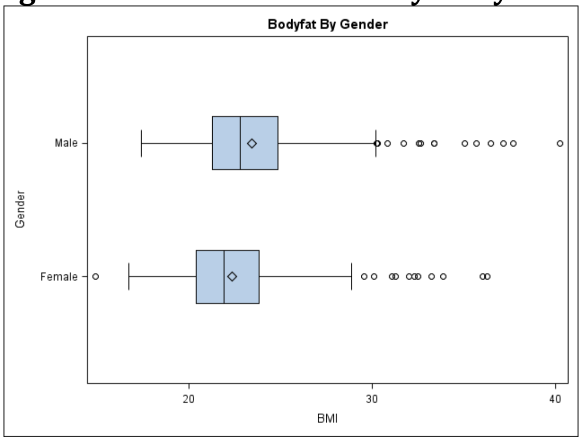


Figure 6: Distribution of Race

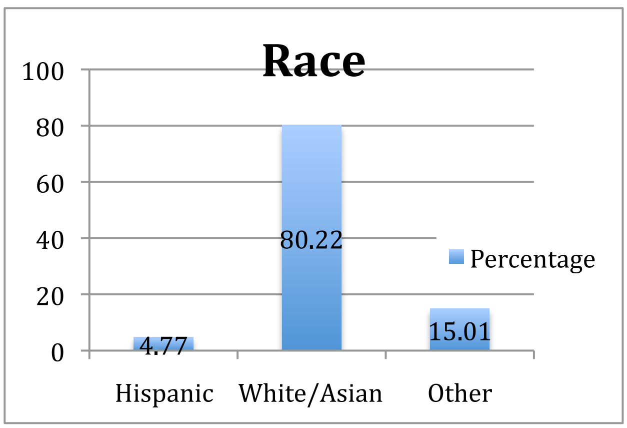


Table 5: Distribution of Race

Race	*Frequency*	*Percentage*
Hispanic	34	4.77
White/Asian	572	80.22
Other	107	15.01
Sample Size	713	100.00

Table 6 describes the distribution of stress score for the two college samples discussed in Cohen, Kamarck, and Mermelstein's article in the Journal of Health and Social Behavior. Table 7 and Figures 7, 8, and 9 show the distribution of stress score in the FLASH study.

Table 6: Stress Score in "A Global Measure of Perceived Stress", College Samples I & II

Stress Score	College Sample I	College Sample II
Sample size	114	332
Mean	23.67	23.18
Std. Dev.	7.79	7.31

Table 7: Stress Score Statistics

Stress Score	Male	Female	Overall
Sample Size	334	337	671
Mean	15.796	17.234	16.518
Std. Dev.	6.492	6.172	6.370

Figure 7: Stress Score Distribution

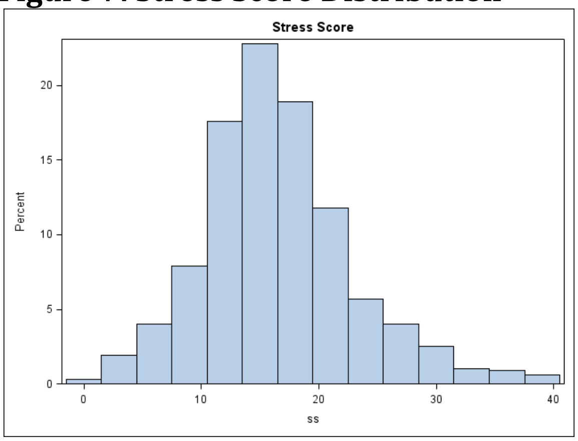


Figure 8: Stress Score Distribution - Women

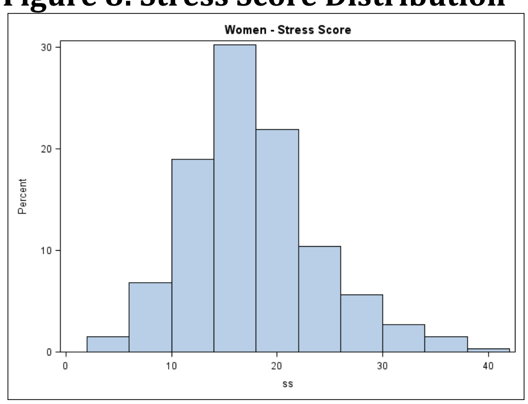
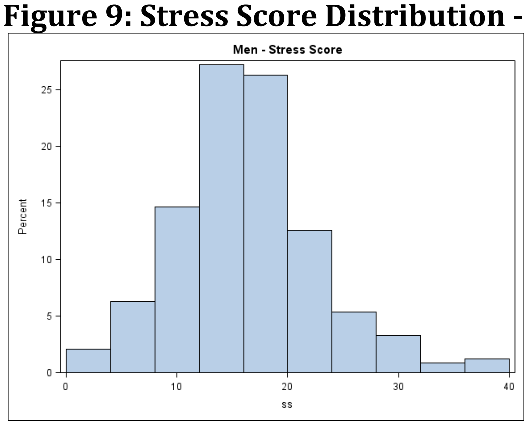


Figure 9: Stress Score Distribution - Men



IPAQ class was classified using the following protocol (from http://www.ipaq.ki.se):

1. **LOW**
   No activity is reported
   OR
   Some activity is reported but not enough to meet Categories 2 or 3.
2. **MODERATE**
   3 or more days of vigorous activity of at least 20 minutes per day
   OR
   5 or more days of moderate-intensity activity and/or walking of at least 30 minutes per day
   OR
   5 or more days of any combination of walking, moderate-intensity or vigorous-intensity activities achieving a minimum of at least 600 MET-minutes/week.
3. **HIGH**
   Vigorous-intensity activity on at least 3 days and accumulating at least 1500 MET-minutes/week
   OR
   7 or more days of any combination of walking, activities accumulating at least 3000 MET-minutes/week

Table 8: IPAQ Class Distribution

IPAQ Class	Male	Female	Overall
Sample Size	359	354	713
Low	19.78%	25.14%	22.44%
Moderate	24.51	32.77	28.61
High	55.71	42.09	48.95

Figure 10: IPAQ Class Distribution

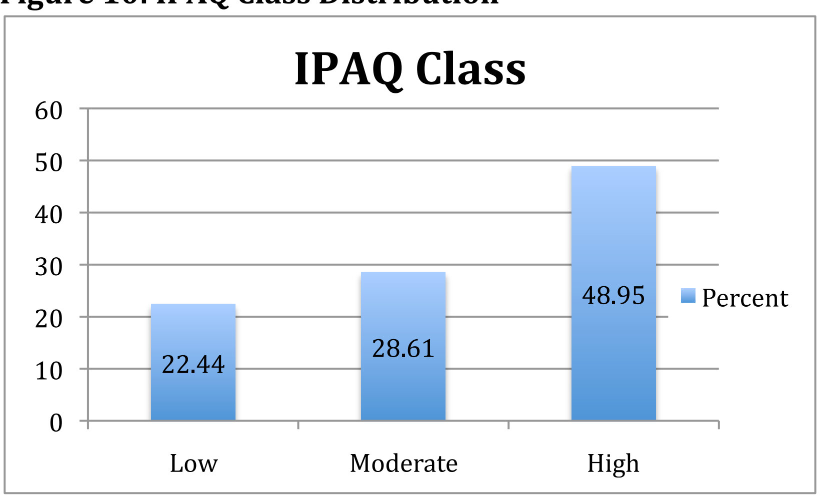


Figure 11: IPAQ Class Distribution by Gender

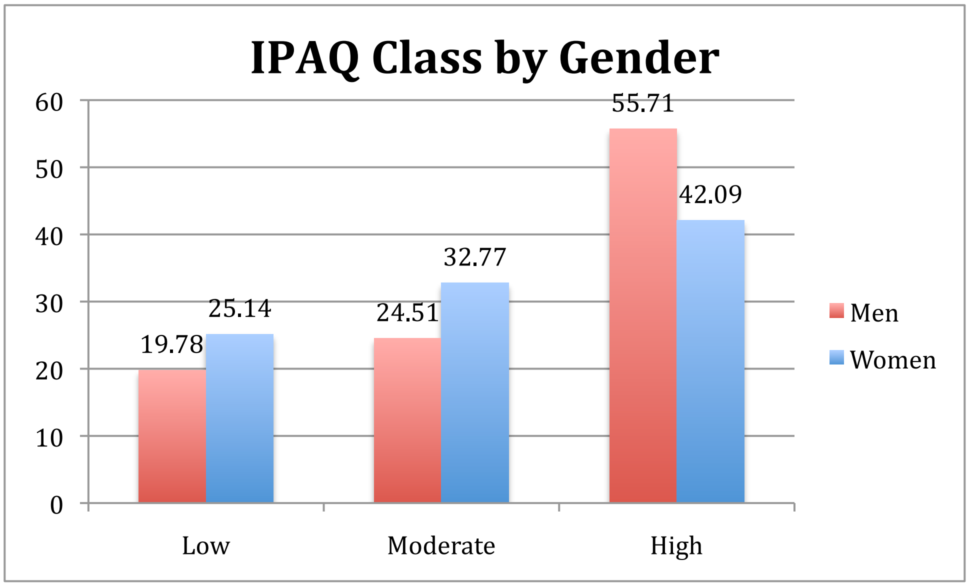


Table 9: Videogames per Weekday (Hours) Statistics

Videogames per Weekday (hrs)	Male	Female	Overall
Mean	1.798	0.828	1.31
Std. Dev.	2.258	0.945	1.79
Sample Size	334	338	672

Figure 13: Distribution of Videogames per Weekday

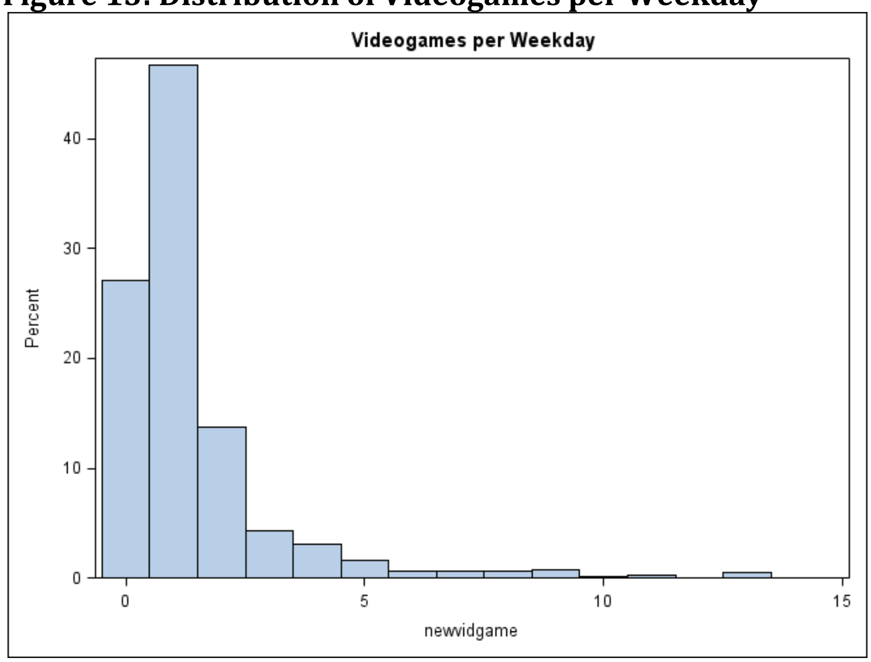
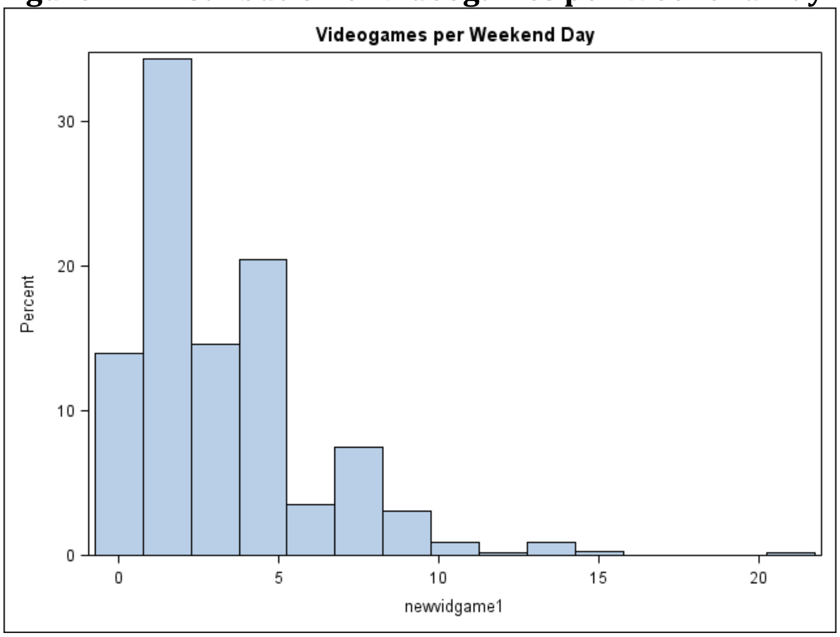


Table 10: Videogames per Weekend (Hours) Statistics

Videogames per Weekend Day (hrs)	Male	Female	Overall
Mean	3.119	3.440	3.28
Std. Dev.	2.723	2.736	2.73
Sample Size	337	341	678

Figure 12: Distribution of Videogames per Weekend Day



III. RESULTS

Binary Logistic Regression

I used binary logistic regression to analyze the association of hypertension/ prehypertension with the following variables: gender, race, BMI, bodyfat, stress score, IPAQ class, videogames per weekday (hours) and videogames per weekend day (hours). I also analyzed the interactions between videogames and gender (both for weekends and weekdays). I found the most significant factors to model blood pressure class were BMI, bodyfat, videogame hours per weekend day and the interaction between videogame hours per weekend day and gender. This model used 677 of the 713 observations. (Observations with missing values for the variables of interest were not included in the analysis.) Of the observations, 400 were classified as having normal blood pressure and 277 were hypertensive/prehypertensive. The significant output is shown in Table 11. The coefficient estimates in Table 11 were estimated using Fisher's Exact Method, and represent the change in log odds of prehypertension/hypertension associated with a one-unit increase in the explanatory variable. For example, a one-unit increase in BMI is associated with an increase of 0.2318 in the log odds of hypertension/prehypertension (when compared to normal blood pressure), all other variables remaining constant. The data analysis for this paper was generated using SAS software, Version 9.2 of the SAS System for Windows.[1]

Table 11: Significant Explanatory Variables – Binary Logistic Regression

Explanatory Variable	Coefficient Estimate	P-value	95% CI for Odds Ratio
Race (REF = Other)	-	0.6755	
Hispanic	-0.1098		
White/Asian	-0.0789		
Gender (REF=male)	1.0354	<0.0001	
BMI	0.2318	<0.0001	(1.163, 1.367)
Bodyfat	-0.0519	0.0208	(0.909, 0.992)
Videogames per Weekend Day	-0.0733	0.0408	
Videogames Weekend*Gender	-	0.0201	
Gender=Male	-0.0816		(0.780, 0.940)
Gender=Female			(0.910, 1.118)

We can conclude that BMI, bodyfat, and the interaction between hours of videogames played on the weekend and gender are significant predictors of blood pressure class. We also conclude from p-values that race does not have a significant association with hypertension/prehypertension in students. Neither stress score nor IPAQ class resulted in a significant association.

As stated above, the interaction between gender and time spent playing videogames on weekends is significant. This means that a change in amount of time spent playing videogames has a different association with chances of hypertension/prehypertension, depending on gender. Figure 14 shows the relationship between videogame hours per weekend day and log odds of prehypertension/hypertension by gender. It is easy to see how the relationships differ by gender. For males, an increase in videogames per weekend day is associated with a decrease in log odds of prehypertension/hypertension. This is not true for females. This is why the interaction between gender and time spent playing videogames was found to be significantly related to the probability of hypertension/prehypertension.

Figure 14: Interaction Plot

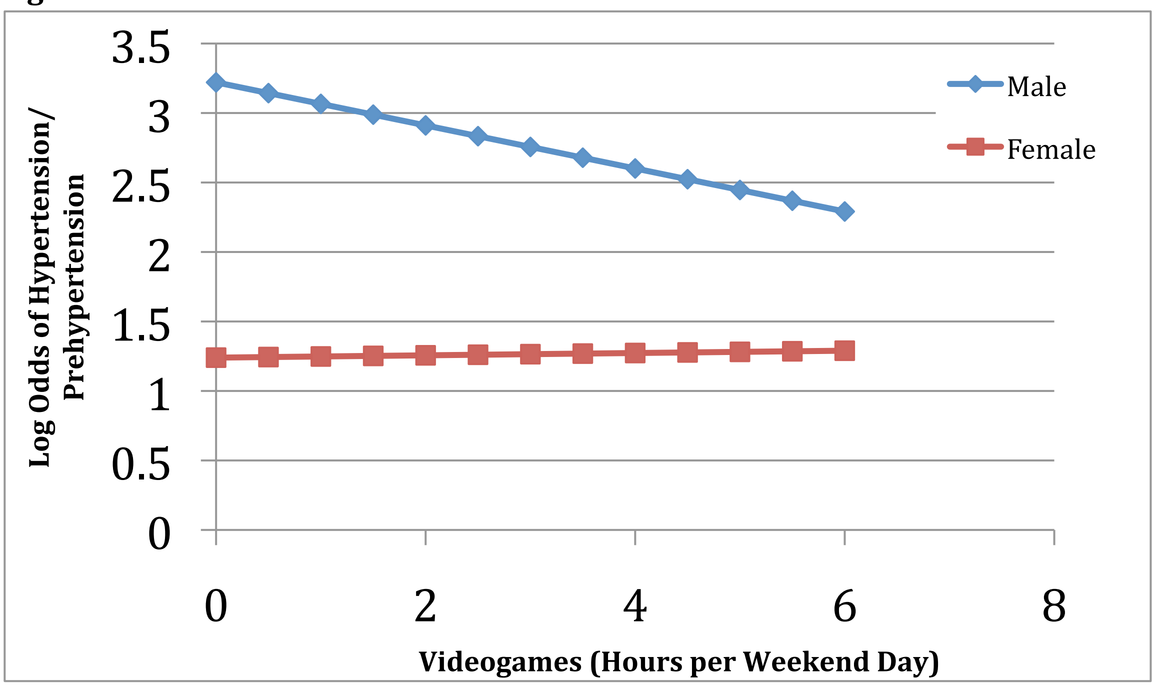


Figure 15 shows the odds ratios for hours spent playing videogames per weekend day by gender.

Figure 15: Odds Ratios

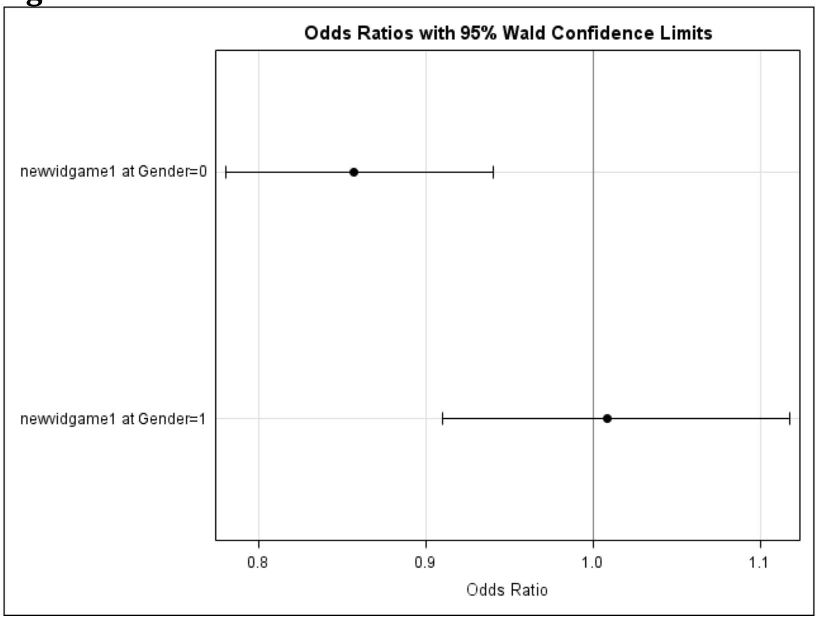


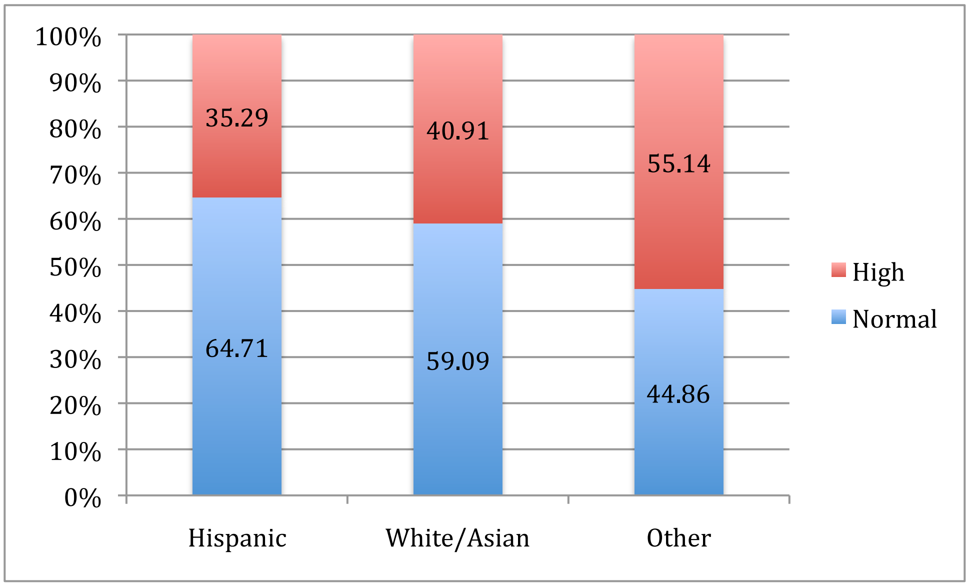
Figure 15 shows that time spent playing videogames on the weekend is significant for males, with regard to odds of pre-hypertension/hypertension (compared to normal blood pressure). Since the confidence interval for the odds ratio does not include one, it can be concluded that, for males in our sample, playing more videogames is associated with a lower probability of hypertension/prehypertension. From the odds ratios, we conclude that the odds of hypertension/prehypertension for males are multiplied by between 0.780 and 0.940 with a one-hour increase in time spent playing videogames on a weekend day. This means an increase in time spent playing videogames on the weekend is associated with a decrease in probability of hypertension/prehypertension in men. Since the odds ratio confidence interval for women includes one, we conclude that time spent playing videogames on the weekend does not significantly affect the odds of hypertension/prehypertension in women.

Table 12 and Figure 16 show the distribution of blood pressure class by race. Race was not found to be significantly associated with the probability of hypertension/prehypertension (p=0.675). This result could be influenced by the fact that over 80 percent of the sample falls in the White/Asian category.

Table 12: Distribution of Blood Pressure Class by Race

	Hispanic		White/Asian		Other	
BP Class	*Frequency*	*Percent*	*Frequency*	*Percent*	*Frequency*	*Percent*
Normal	22	64.71	338	59.09	48	44.86
Hypertension/ Prehypertension	12	35.29	234	40.91	59	55.14

Figure 16: Distribution of Blood Pressure Class by Race



Both BMI and bodyfat were significantly associated with blood pressure class. Since the confidence limits for the odds ratio are both greater than one, we can conclude that an increase in BMI is associated with a higher probability of hypertension/prehypertension. Specifically, the odds of hypertension/prehypertension are between 1.163 and 1.367 times higher for every one-unit increase in BMI.

Since the confidence limits for the odds ratio corresponding to the predictor variable bodyfat are both less than one, we can conclude that a one percentage point increase in bodyfat is associated with a lower probability of hypertension. Specifically, the odds of hypertension/prehypertension are between 0.909 and 0.992 times lower for every one percentage point increase in bodyfat. Although this seems counterintuitive, a possible explanation is that, on average, women have higher bodyfat than men, and, as shown in my descriptive statistics, a smaller proportion of women in our sample are hypertensive/prehypertensive than men.

Ordinal Logistic Regression

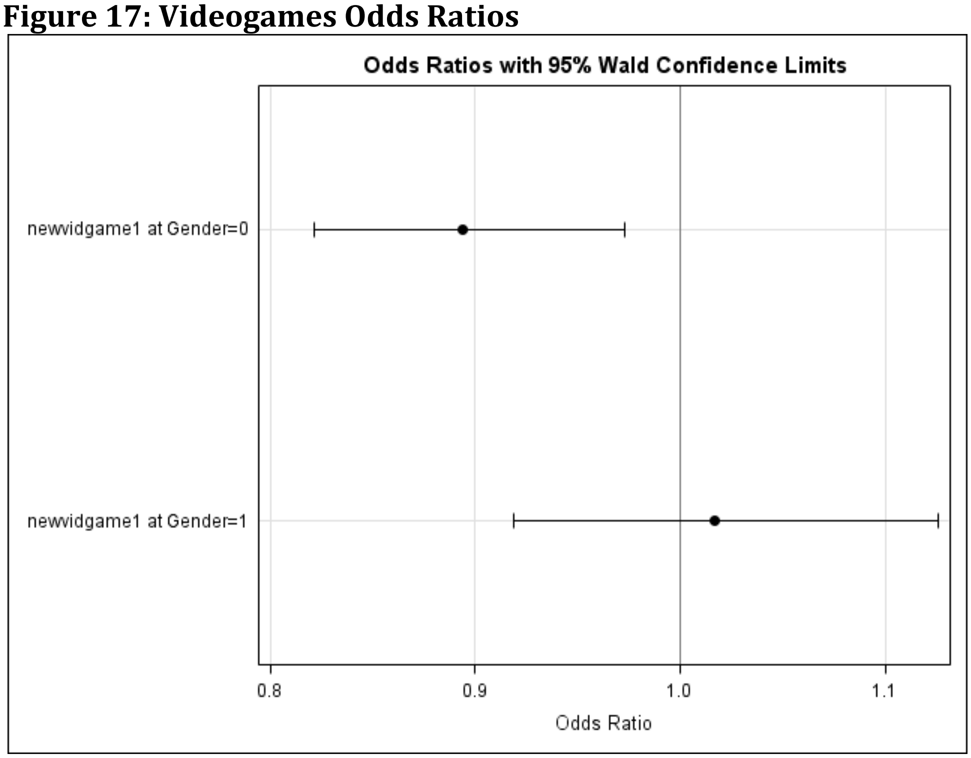
The probabilities of prehypertension and hypertension were also modeled using ordinal logistic regression. The respondents were classified into one of three blood pressure classes: normal, prehypertensive, and hypertensive, according to Table 1. Of the 713 observations, 37 (5.52%) were classified as hypertensive, 237 (35.37%) were prehypertensive, and 396 (59.10%) had normal blood pressure. I discovered that the most significant contributors on blood pressure class were BMI, bodyfat, and the interaction between gender and hours of videogames played per weekend day. The significant output is listed in Table 16. Coefficient estimates were calculated using Fisher's Exact method. The coefficients represent the change in the log odds of hypertension associated with a one-unit change in the explanatory variable. For example, a one-unit increase in BMI, all other variables remaining constant, is associated with an increase of the log odds of hypertension by 0.2075.

Table 13: Significant Explanatory Variables – Ordinal Logistic Regression

Explanatory Variable	Coefficient Estimate	P-value	95% Odds Ratio
Race (REF = Other)		0.8102	
Hispanic	-0.1834		
White/Asian	0.0483		
Gender (REF=male)	0.9827	<0.0001	
BMI	0.2075	<0.0001	(1.147, 1.132)
Bodyfat	-0.0499	0.0170	(0.913, 0.991)
Videogames per Weekend Day	-0.0476	0.1624	
Videogames Weekend*Gender		0.0551	
Gender=Male	-0.0644		(0.821, 0.973)
Gender=Female			(0.918, 1.126)

Overall, the ordinal logistic regression had the results we would expect given the binary model. Figure 17 shows the odds ratios for hours of videogames played per weekend day. Because the odds ratio confidence interval for males does not include one, we can conclude that an increase in videogames played (per weekend day) is associated with a decrease in odds of hypertension in men.

Figure 17: Videogames Odds Ratios



BMI, bodyfat, race and gender have similar p-values and odds ratios to the first regression model. The results for these variables can be interpreted in the same way as described in the Binary Logistic Regression section.

IV. CONCLUSION

The goal of my research was to determine what could be significant predictors of hypertension and prehypertension for college students at Cal Poly. I discovered that BMI, bodyfat, and the interaction between gender and videogames per weekend day are significantly associated with hypertension/prehypertension in Cal Poly students. For males, more time spent playing videogames was associated with a lower probability of hypertension and prehypertension. Also, higher BMI and lower bodyfat were associated with a higher probability of hypertension/prehypertension.

Amount of time playing videogames on weekends was also found to be significantly associated with hypertension/prehypertension. My research showed that more time playing videogames was associated with a lower probability of hypertension/prehypertension among males. Although we cannot state conclusively the cause for this association, a possible explanation would be that playing videogames relieves stress and therefore lowers blood pressure.

A remaining question from my research is why a decreased bodyfat is shown to be associated with an increased risk of hypertension/prehypertension. An intuitive response to this question is explained in the analysis section; namely, in our sample, women tend to have higher bodyfat and lower risk of hypertension than men. However, my research found no significant interaction between gender and bodyfat, meaning there is likely a different explanation that has not yet been considered.

I would have liked to further analyze the contributors to high blood pressure in college students using the data from the Heart Health section of the physical for the FLASH assessment. This includes measurements of glucose level, cholesterol, and triglyceride levels that likely would have been associated with hypertension.

REFERENCES

Cohen, Sheldon, Tom Kamarck, and Robin Mermelsteing. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24.4 (1983): 385-96. Print.

Conklin, Emily J., and Jongyoon Lee. "Dietary Patterns in Relation to Sleep and Stress in Cal Poly Freshman." Thesis. California Polytechnic State University, 2011.

"IPAQ Scoring Protocol." *Http://www.ipaq.ki.se/*. International Physical Activity Questionnaire, 4 Nov. 2010. Web. <https://sites.google.com/site/theipaq/scoring-protocol>.

Kleinbaum, David G. *Logistic Regression: A Self-learning Text*. New York: Springer, 1994. Print.

SAS Online Documentation. SAS. Web. <http://support.sas.com/documentation/>.

"What Is High Blood Pressure?" *Nhlbi.nhi.gov*. National Heart Lung and Blood Institute. Web. 22 Jan. 2012. <http://www.nhlbi.nih.gov/health/health-topics/topics/hbp/>.

ACKNOWLEDGEMENTS

- FLASH Study, Cal Poly, San Luis Obispo, CA 93407

V. APPENDIX

~~~~~~

```
                                     The SAS System

                                 The LOGISTIC Procedure

                                   Model Information

                     Data Set                      WORK.FLASH
                     Response Variable             bpstatus
                     Number of Response Levels     2
                     Model                         binary logit
                     Optimization Technique        Fisher's scoring


                         Number of Observations Read         713
                         Number of Observations Used         677


                                    Response Profile

                            Ordered                      Total
                              Value     bpstatus     Frequency

                                  1     high               277
                                  2     normal             400

                         Probability modeled is bpstatus='high'.

NOTE: 36 observations were deleted due to missing values for the response or explanatory
      variables.


                                 Class Level Information

                                                    Design
                            Class      Value       Variables

                            Gender     0             1
                                       1            -1

                            race       hispanic      1      0
                                       non-hisp      0      1
                                       other        -1     -1


                                Model Convergence Status

                     Convergence criterion (GCONV=1E-8) satisfied.
```

~~~~~~

```
                                     The SAS System

                                 The LOGISTIC Procedure

                                  Model Fit Statistics

                                                     Intercept
                                      Intercept            and
                         Criterion          Only     Covariates

                         AIC             918.050        712.862
                         SC              922.567        749.004
                         -2 Log L        916.050        696.862


                         Testing Global Null Hypothesis: BETA=0
```

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	219.1874	7	<.0001
Score	199.9173	7	<.0001
Wald	159.7575	7	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Gender	1	27.1422	<.0001
BMI	1	31.5514	<.0001
Bodyfat	1	5.3475	0.0208
race	2	0.7847	0.6755
newvidgame1	1	4.1863	0.0408
newvidgame1*Gender	1	5.4019	0.0201

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.2360	0.7698	30.2832	<.0001
Gender	0	1	1.0354	0.1987	27.1422	<.0001
BMI		1	0.2318	0.0413	31.5514	<.0001
Bodyfat		1	-0.0519	0.0225	5.3475	0.0208
race	hispanic	1	-0.1098	0.3001	0.1338	0.7146
race	non-hisp	1	-0.0789	0.1858	0.1805	0.6710
newvidgame1		1	-0.0733	0.0358	4.1863	0.0408
newvidgame1*Gender	0	1	-0.0816	0.0351	5.4019	0.0201

~~~~~~

The SAS System

The LOGISTIC Procedure

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
BMI		1.261	1.163	1.367
Bodyfat		0.949	0.909	0.992
race	hispanic vs other	0.742	0.273	2.018
race	non-hisp vs other	0.765	0.420	1.396

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.7	Somers' D	0.616
Percent Discordant	19.1	Gamma	0.618
Percent Tied	0.2	Tau-a	0.299
Pairs	110800	c	0.808

Wald Confidence Interval for Odds Ratios

Label	Estimate	95% Confidence Limits	
newvidgame1 at Gender=0	0.857	0.780	0.940
newvidgame1 at Gender=1	1.008	0.910	1.118
Gender 0 vs 1 at newvidgame1=3.2821	4.643	2.477	8.70
~~~~~~

~~~~~~

```
                              The SAS System

                          The LOGISTIC Procedure

                            Model Information

                 Data Set                      WORK.FLASH
                 Response Variable             bpstatus1
                 Number of Response Levels     3
                 Model                         cumulative logit
                 Optimization Technique        Fisher's scoring


                     Number of Observations Read         713
                     Number of Observations Used         670


                              Response Profile

                      Ordered                          Total
                        Value     bpstatus1        Frequency

                            1     hypertension            37
                            2     prehyp                 237
                            3     normal                 396

           Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 43 observations were deleted due to missing values for the response or explanatory
      variables.


                          Class Level Information

                                                 Design
                        Class      Value        Variables

                        Gender     0              1
                                   1             -1

                        race       hispanic       1      0
                                   non-hisp       0      1
                                   other         -1     -1


                         Model Convergence Status

                 Convergence criterion (GCONV=1E-8) satisfied.
```

~~~~~~

```
                              The SAS System

                          The LOGISTIC Procedure

                 Score Test for the Proportional Odds Assumption

                       Chi-Square       DF     Pr > ChiSq

                          16.8810        9         0.0506


                           Model Fit Statistics

                                                 Intercept
                                   Intercept           and
                     Criterion          Only    Covariates

                     AIC            1127.404       925.427
```

SC	1136.418	975.007
-2 Log L	1123.404	903.427

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	219.9768	9	<.0001
Score	201.8437	9	<.0001
Wald	169.9905	9	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Gender	1	24.5100	<.0001
BMI	1	32.3691	<.0001
Bodyfat	1	5.4395	0.0197
race	2	0.3336	0.8464
newvidgame1	1	1.2763	0.2586
newvidgame	1	0.3266	0.5677
newvidgame1*Gender	1	3.9744	0.0462
newvidgame*Gender	1	0.2623	0.6085

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	hypertension	1	-6.9065	0.7289	89.7825	<.0001
Intercept	prehyp	1	-3.8619	0.6747	32.7653	<.0001
Gender	0	1	0.9818	0.1983	24.5100	<.0001
BMI		1	0.2056	0.0361	32.3691	<.0001
Bodyfat		1	-0.0491	0.0210	5.4395	0.0197
race	hispanic	1	-0.1618	0.2959	0.2993	0.5844
race	non-hisp	1	0.0368	0.1797	0.0418	0.8379
newvidgame1		1	-0.0409	0.0362	1.2763	0.2586
newvidgame		1	-0.0559	0.0978	0.3266	0.5677
newvidgame1*Gender	0	1	-0.0717	0.0360	3.9744	0.0462
newvidgame*Gender	0	1	0.0503	0.0981	0.2623	0.6085

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
BMI	1.228	1.144	1.318
Bodyfat	0.952	0.914	0.992
race hispanic vs other	0.751	0.283	1.989
race non-hisp vs other	0.915	0.523	1.602

Association of Predicted Probabilities and Observed Responses

Percent Concordant	5.7	Somers' D	-.045
Percent Discordant	10.2	Gamma	-.285
Percent Tied	84.2	Tau-a	-.024
Pairs	117273	c	0.477

Wald Confidence Interval for Odds Ratios

Label	Estimate	95% Confidence Limits	
newvidgame1 at Gender=0	0.893	0.816	0.978
newvidgame1 at Gender=1	1.031	0.925	1.150
newvidgame at Gender=0	0.994	0.897	1.102
newvidgame at Gender=1	0.899	0.621	1.302

```
options formdlim="~" nodate nonumber
        rightmargin=.5 leftmargin=.5 topmargin=.5 bottommargin=.5;
libname sassy "F:/";

data work.flash;
set sassy.flash;
if gender=. then delete;

*computes bpstatus (high or normal);
if bps<120 and bpd<80 then bpstatus="normal";
else bpstatus="high";

*computes bpstatus1 (hyp, prehyp, or normal);
if bps>=140 or bpd>=90 then bpstatus1="hypertension";
else if bps>=120 or bpd>=80 then bpstatus1="prehyp";
else if bps>0 and bpd>0 then bpstatus1="normal";

*inverts necessary problems to compute stress score;
if dealtsucchassles=4 then dealtsucchassles1=0;
if copingwithchanges=4 then copingwithchanges1=0;
if confpersonalproblems=4 then confpersonalproblems1=0;
if goingyourway=4 then goingyourway1=0;
if ablecontrolirr=4 then ablecontrolirr1=0;
if ontopofthings=4 then ontopofthings1=0;
if ablecontroltime=4 then ablecontroltime1=0;

if dealtsucchassles=3 then dealtsucchassles1=1;
if copingwithchanges=3 then copingwithchanges1=1;
if confpersonalproblems=3 then confpersonalproblems1=1;
if goingyourway=3 then goingyourway1=1;
if ablecontrolirr=3 then ablecontrolirr1=1;
if ontopofthings=3 then ontopofthings1=1;
if ablecontroltime=3 then ablecontroltime1=1;

if dealtsucchassles=1 then dealtsucchassles1=3;
if copingwithchanges=1 then copingwithchanges1=3;
if confpersonalproblems=1 then confpersonalproblems1=3;
if goingyourway=1 then goingyourway1=3;
if ablecontrolirr=1 then ablecontrolirr1=3;
if ontopofthings=1 then ontopofthings1=3;
if ablecontroltime=1 then ablecontroltime1=3;

if dealtsucchassles=0 then dealtsucchassles1=4;
if copingwithchanges=0 then copingwithchanges1=4;
if confpersonalproblems=0 then confpersonalproblems1=4;
if goingyourway=0 then goingyourway1=4;
if ablecontrolirr=0 then ablecontrolirr1=4;
if ontopofthings=0 then ontopofthings1=4;
if ablecontroltime=0 then ablecontroltime1=4;

*computes stress score;
ss=sum(upsetunexpected,unablecontrol,nervousstressed,dealtsucchassle
s1,copingwithchanges1,
```

```
confpersonalproblems1,goingyourway1,couldnotcope,ablecontrolirr1,ont
opofthings1,angered,
   thinkingabout,ablecontroltime1,diffpilingup);

*computes ipaq class;
mpaweek=modphysactdays*(modphysacthrs*60+modphysactmins);
vpaweek=vigphysactdays*(vigphysacthrs*60+vigphysactmins);
walkweek=walkingdays*(walkinghrs*60+walkingmins);
walkMET=walkweek*3.3;
MET=sum(mpaweek*4,vpaweek*8,walkMET);

if vigphysactdays>=3 and MET>=1500 then ipaq="high";
else if vigphysactdays+modphysactdays+walkingdays>=7 and MET>=3000
    then ipaq="high";
else if vigphysactdays>=3 and vpaweek/vigphysactdays>=20
    then ipaq="moderate";
else if modphysactdays>=5 or (walkweek/7)>=30
    then ipaq="moderate";
else if vigphysactdays+modphysactdays+walkingdays>=5 and MET>=600
    then ipaq="moderate";
else ipaq="low";

if hispanic=2 then race="hispanic";
else if white=1 or asian=1 then race="non-hispanic white/asian";
else race="other";

if tabaccocigarettes=99 or tabaccocigarettes=0
    then cursmoke=0;
else cursmoke=1;
*current smokers have cursmoke=1;

if marijuana=99 or marijuana=0
    then curmj=0;
else curmj=1;
*current marijuana users have curmj=1;

if alcohol2=99 or alcohol2=0
    then curdrink=0;
else curdrink=1;
*current drinkers have curdrink=1;

*weekday tv variable;
length tv $ 20;
if avtvmovieweekday=0 or avtvmovieweekday=99 then tv="none";
else if 0<avtvmovieweekday<=2 or avtvmovieweekday=77 then tv="less
than 2 hrs";
else if avtvmovieweekday>2 then tv="greater than 2 hrs";

*create ORDINAL VIDEOGAMES variable (WKDAYS);
length videogames $ 20;
if avvideogamesweekday=0 or avvideogamesweekday=99  then
videogames="none";
else if 0<avvideogamesweekday<=2 or avvideogamesweekday=77 then
videogames="less than 2 hrs";
```

```
else if avvideogamesweekday>2 then videogames="greater than 2 hrs";

*videogames=77 means less than one hour a day - I defined this as
0.5 hours;
*create ORDINAL VIDEOGAMES1 variable (WEEKEND);
length videogames1 $ 20;
if avvideogamesweekend=0 or avvideogamesweekend=99  then
videogames1="none";
else if 0<avvideogamesweekend<=2 or avvideogamesweekend=77 then
videogames1="less than 2 hrs";
else if avvideogamesweekend>2 then videogames1="greater than 2 hrs";

*create CONTINUOUS variable NEWVIDGAME for analysis (WKDAY);
if avvideogamesweekday=0 or avvideogamesweekday=99  then
newvidgame=0;
else if avvideogamesweekday=77 then newvidgame=0.5;
else newvidgame=avvideogamesweekday;

*create CONTINUOUS variable NEWVIDGAME1 for analysis (WEEKEND);
if avvideogamesweekend=0 or avvideogamesweekend=99  then
newvidgame1=0;
else if avvideogamesweekend=77 then newvidgame1=0.5;
else newvidgame1=avvideogamesweekend;

*weekday internet variable;
length internet $ 20;
if avinternetweekday=0 or avinternetweekday=99 then internet="none";
else if 0<avinternetweekday<=2 or avinternetweekday=77 then
internet="less than 2 hrs";
else if avinternetweekday>2 then internet="greater than 2 hrs";

run;

*men and women only data sets;
data men women;
set flash;
if gender=0 then output men;
else if gender=1 then output women;
run;

*basic sample statistics;
proc sort data=flash; by gender; run;

*bodyfat by gender;
proc means data=flash;
var bodyfat;
by gender;
run;

proc means data=flash;
var ss;
run;

proc means data=flash;
var ss;
```

```
by gender;
run;

*bmi by gender;
proc means data=flash;
var bmi;
run;
proc means data=men;
var bmi;
run;
proc means data=women;
var bmi;
run;
proc freq data=flash;
tables race ss ipaq;
run;

proc freq data=flash;
tables ipaq*gender;
run;

proc means data=flash;
var newvidgame newvidgame1;
run;

proc freq data=flash;
tables bpstatus*gender bpstatus bpstatus*race;
run;

proc means data=flash;
var ss;
run;

*bpstatus by videogames and gender;
proc freq data=flash;
tables bpstatus*videogames;
by gender;
run;

*plots of bodyfat by gender;
ods graphics on;
proc sgplot data=flash;
title "Bodyfat";
histogram bodyfat;
run;
proc sgplot data=men;
title "Bodyfat of Men";
hbox bodyfat;
run;
ods graphics on;

*formats gender plots;
proc format;
 value sexfmt 0="Male"
              1="Female";
```

```
run;

*bmi by gender plot;
proc sgplot data=flash;
title "BMI By Gender";
hbox bmi / category=gender;
format gender sexfmt.;
run;
ods graphics off;

*histograms of videogames played per weekend day and weekday;
ods graphics on;
proc sgplot data=flash;
title "Videogames per Weekday";
histogram newvidgame;
run;
proc sgplot data=flash;
title "Videogames per Weekend Day";
histogram newvidgame1;
run;
ods graphics off;

*boxplots of videogames played by bpstatus and gender;
ods graphics on;
proc sgplot data=men;
title "Men - Weekend Videogames Played by BP Class";
hbox newvidgame / category=bpstatus;
format gender sexfmt.;
run;
proc sgplot data=women;
title "Women - Weekend Videogames Played by BP Class";
hbox newvidgame1 / category=bpstatus;
format gender sexfmt.;
run;
ods graphics off;

*stress score boxplot;
ods graphics on;
proc sgplot data=flash;
title "Stress Score";
hbox ss;
run;
ods graphics off;

*ss histogram (and by gender);
ods graphics on;
proc sgplot data=men;
title "Men - Stress Score";
histogram ss;
run;

proc sgplot data=women;
title "Women - Stress Score";
histogram ss;
run;
```

```
ods graphics off;

*analysis logistic regressions- BINARY MODEL;
ods graphics on;
proc logistic data=flash /*plots=all*/ ;
oddsratio newvidgame1;
oddsratio gender;
class gender race;
model bpstatus=gender bmi bodyfat race newvidgame1
newvidgame1*gender;
run;
ods graphics off;

*ordinal logistic regression;
ods graphics on;
proc logistic data=flash order=data desc /*plots=all */;
oddsratio newvidgame1;
oddsratio gender;
class gender race;
model bpstatus1=gender bmi bodyfat race newvidgame1
newvidgame1*gender;
run;
ods graphics off;

*other models;


proc logistic data=flash;
title "bpstatus=gender";
class gender;
model bpstatus= gender;
run;

proc logistic data=flash;
title "bpstatus=gender race";
class gender race;
model bpstatus= gender race;
run;

proc logistic data=flash order=data desc;
title "ordinal bpstatus=gender";
class gender;
model bpstatus1=gender;
run;

proc logistic data=flash order=data desc;
title "ordinal bpstatus=gender race";
class gender race;
model bpstatus1=gender race;
run;

proc logistic data=flash order=data desc;
title "full ordinal model for writeup";
class gender race;
```

```
model bpstatus1=gender bmi bodyfat race newvidgame1
newvidgame1*gender;
run;



/*
proc logistic data=women;
class ipaq tv videogames internet race;
model bpstatus= bmi   curdrink  ;
run;

proc logistic data=flash;
class gender;
model bpstatus=gender bmi bodyfat newvidgame;
run;

proc logistic data=flash;
class gender;
model bpstatus=gender bmi bodyfat newvidgame1;
run;

proc logistic data=flash;
class gender;
model bpstatus=gender bmi bodyfat newvidgame newvidgame*gender;
run;

proc logistic data=flash;
class gender;
model bpstatus=gender bmi bodyfat newvidgame1 newvidgame1*gender;
run;

ods graphics on;
proc logistic data=flash plots=all;
oddsratio newvidgame1;
class gender race;
model bpstatus=gender bmi bodyfat race newvidgame1
newvidgame1*gender;
run;
ods graphics off;

ods graphics on;
proc logistic data=flash plots=all;
oddsratio newvidgame1;
class gender race;
model bpstatus=gender bodyfat bmi  race newvidgame1
newvidgame1*gender;
run;
ods graphics off;

proc logistic data=flash;
class gender videogames1;
model bpstatus=gender bmi bodyfat videogames1*gender;
run;
*/
```