ACM 660 **PROJECT 1**

Harsha Kankanamge

A: The data set OBESITY.SAV represents data from a sample of 5857 subjects from the general population in the United States between 2009 and 2010. The goal is to identify risk factors that predict obesity. Using the variables WLKBIK (walk and use bicycle to get to and from places), SEDMIN (minutes of sedentary activity per week) and OBESE (body mass index status)

1. Conduct a descriptive analysis of the variables in the data set and summarize the results.

Frequency table of the BIM more than 35

bmi more than 35

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	4888	83.5	84.0	84.0
	yes	934	15.9	16.0	100.0
	Total	5822	99.4	100.0	
Missing	System	35	.6		
Total		5857	100.0		

In the sample 4888 people have BMI factor less than or equal 35 out of 5857 people. In general, 83.5% of the entire dataset represent that people who have BMI factor less than or equal 35, as well as 0.6% (35) of data missing in this column. 15.9% of people in the sample have “BMI more than 35”.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
minutes of sedentary activity per week	5786	0	840	314.63	185.081
Valid N (listwise)	5786				

Mean and Std. deviation of variable “minutes of sedentary activity per week” are respectively 314 minutes and 185.081. As well as it has maximum 840 minutes and minimum 0 minutes.

walk or bicycle

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	1524	26.0	26.0	26.0
	no	4333	74.0	74.0	100.0
	Total	5857	100.0	100.0	

In the sample 4333 people out of 5857 people do not “walk or bicycle”. It is 74% percent of the entire valid data. 1524 people do the “walk or bicycle” out of 5857. It is 26% of entire valid data.

bmi more than 35 * walk or bicycle

Crosstab

			walk or bicycle		
			yes	no	Total
bmi more than 35	no	Count	1335	3553	4888
		% within bmi more than 35	27.3%	72.7%	100.0%
	yes	Count	184	750	934
		% within bmi more than 35	19.7%	80.3%	100.0%
Total		Count	1519	4303	5822
		% within bmi more than 35	26.1%	73.9%	100.0%

72.7% of people who have BMI less than or equal to 35, do not “walk or bicycle” to get to and from places. It is 3553 out of 4888 people who have BMI less than or equal to 35. 80.3% of people who have BMI more than 35, do not “walk or bicycle” to get to and from places.

bmi more than 35 * moderate recreational activities

Crosstab

			moderate recreational activities		
			yes	no	Total
bmi more than 35	no	Count	1955	2932	4887
		% within bmi more than 35	40.0%	60.0%	100.0%
	yes	Count	277	657	934
		% within bmi more than 35	29.7%	70.3%	100.0%
Total		Count	2232	3589	5821
		% within bmi more than 35	38.3%	61.7%	100.0%

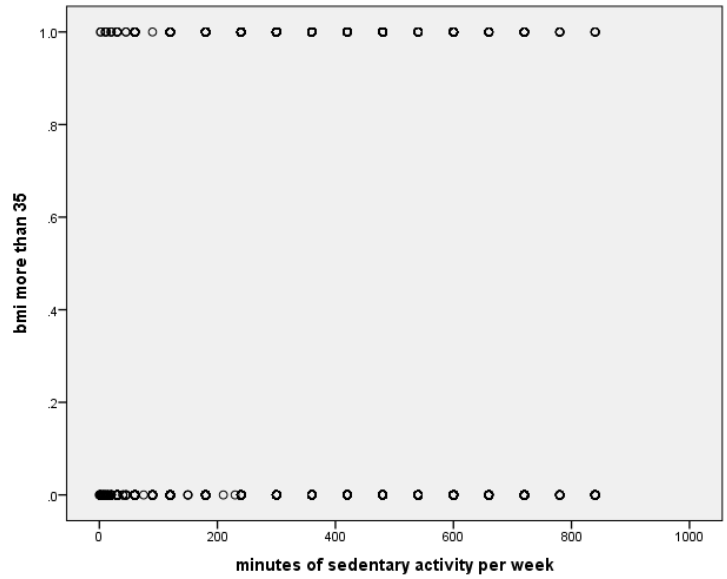
60.0% of people who have BMI less than or equal to 35, do not do moderate recreational activities. 70.3% of people how have BMI more than 35, do not do moderate recreational activities.

Descriptives

bmi more than 35				Statistic	Std. Error
minutes of sedentary activity per week	no	Mean		307.46	2.652
		95% Confidence Interval for Mean	Lower Bound	302.26	
			Upper Bound	312.66	
		5% Trimmed Mean		297.80	
		Median		300.00	
		Variance		34033.359	
		Std. Deviation		184.481	
		Minimum		0	
		Maximum		840	
		Range		840	
		Interquartile Range		240	
		Skewness		.706	.035
		Kurtosis		-.124	.070
	yes	Mean		349.40	6.090
		95% Confidence Interval for Mean	Lower Bound	337.45	
			Upper Bound	361.36	
		5% Trimmed Mean		344.38	
		Median		300.00	
		Variance		33898.771	
		Std. Deviation		184.116	
		Minimum		2	
		Maximum		840	
		Range		838	
		Interquartile Range		300	
		Skewness		.411	.081
		Kurtosis		-.575	.162

Mean value of SEDMIN (minutes of sedentary activity per week) is higher for people who have “BMI more than 35” (349.40) than BMI less than or equal to 35 (307.46). Std. deviation is almost same for both groups.

2. Form a scatter plot of OBESE versus SEDMIN and interpret the plot.



Most of the people have “minutes of sedentary activity per week” less than 400 either BMI more than 35 or less than or equal to 35. But when we look at the average, BMI more than 35 category has higher average for sedentary activity. Because category “bmi less than or equal to 35” most of data point stay very closer to the zero.

3. Write down the equation for the logistic function of OBESE on SEDMIN. Write down the equation for the logit transformation of this logistic regression model.

logistic regression model

$$p = \frac{1}{1+e^{-(b0+b1*SEDMIN)}} \quad or \quad p = \frac{e^{(b0+b1*SEDMIN)}}{1+e^{(b0+b1*SEDMIN)}}$$

logit transformation

$$\ln[\frac{p}{1-p}] = b0 + b1 * SEDMIN$$

Here $[\frac{p}{1-p}]$ is odds of success

4. Using a logistic regression model of OBESE on SEDMIN, estimate the maximum likelihood estimates of the parameters of the model. Using these estimates, write down the equation for the fitted values, i.e. the estimated logistic probabilities and also the estimated logit transformation.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
minutes of sedentary activity per week	Equal variances assumed	.230	.632	-6.306	5750	.000	-41.943	6.651	-54.983	-28.904
	Equal variances not assumed			-6.314	1283.469	.000	-41.943	6.643	-54.975	-28.912

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	SEDMIN	.001	.000	39.124	1	.000	1.001	1.001	1.002
	Constant	-2.053	.074	773.105	1	.000	.128		

a. Variable(s) entered on step 1: SEDMIN.

Estimated logit transformation

$$\ln[\frac{p^{\wedge}}{1-p^{\wedge}}] = -2.053+0.001*\text{SEDMIN}$$

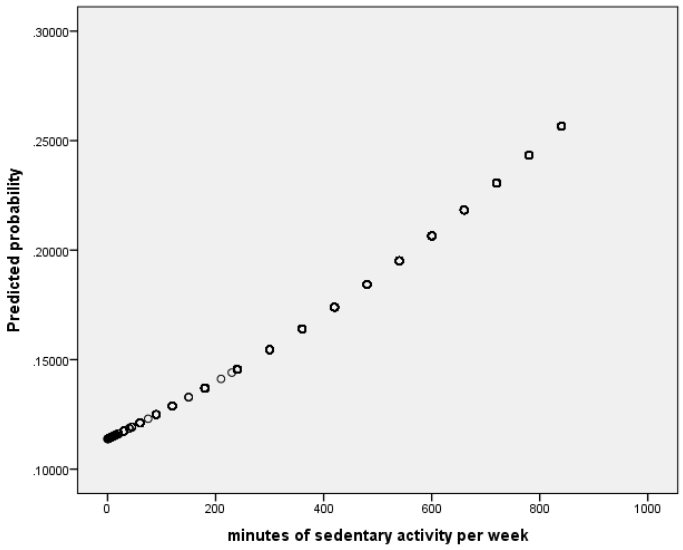
Estimated logistics probabilities

$$[\frac{p^{\wedge}}{1-p^{\wedge}}] = e^{-2.053+0.001*\text{SEDMIN}}$$

$$P^{\wedge} = \frac{e^{-2.053+0.001*\text{SEDMIN}}}{1+e^{-2.053+0.001*\text{SEDMIN}}}$$

If the person who have p^ less than or equal to 0.5 model will predict that person does not have bmi more than 35 if not model will predict bmi is more than 35.

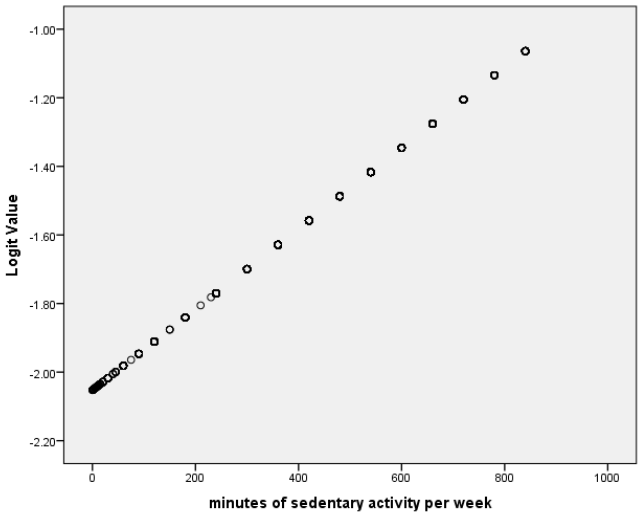
5. Form a scatter plot of the estimated logistic probabilities (fitted values or predicted values) versus SEDMIN and interpret the plot. Save 'predicted probabilities' and use it to form the scatter plot.



According to this graph, it is showing that nonlinear association between predicted probability for OBESE and SEDMIN. All most all the SEDMIN data points have predicted probability values very small for bmi.

6. Form a scatter plot of the logit values versus SEDMIN and interpret the plot.

Logit =ln(p^/1-p^)



According to this graph, it is showing that there is a linear relationship in between SEDMIN and ln(p^/1-p^). Therefore it is proving this linear equation $\ln[\frac{p^{\wedge}}{1-p^{\wedge}}]$= -2.053+0.001*SEDMIN

7. Determine and interpret the slope coefficient (log odds and odds) for every 210 minutes increase of sedentary activity per week (30 minutes every day for 7 days). Assess the significance of the slope coefficient for SEDMIN using the Likelihood ratio test, the Wald test and the Score test. Are the results of these tests consistent with one another? Use alpha = 0.05.

$$\ln[\frac{\hat{p}}{1-\hat{p}}] = -2.053+0.001*SEDMIN$$

The log odds for a change of 210 minutes in SEDMIN is 0.21 (0.001*210).

Coeff of SEDMIN (0.001) *210 implies that for each 210 minutes' increase in SEDMIN, the log odds of having BMI more than 35 increase by 0.21

Odds for an increase of 210 minutes per week in SEDMIN is exp (210*0.001) = exp (0.21) = 1.23

Indicates that for every 210 minutes' increase per week in SEDMIN, the risk of having BMI more than 35 increases by 1.23 times.

Block 0: Beginning Block

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	5111.987	-1.364
	2	5037.504	-1.640
	3	5036.951	-1.666
	4	5036.951	-1.666

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 5036.951

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.666	.036	2134.840	1	.000	.189

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	SEDMIN	39.505	1	.000
	Overall Statistics		39.505	1	.000

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients	
			Constant	SEDMIN
Step 1	1	5083.656	-1.570	.001
	2	4999.507	-1.991	.001
	3	4998.548	-2.051	.001
	4	4998.548	-2.053	.001
	5	4998.548	-2.053	.001

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 5036.951

d. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	38.403	1	.000
	Block	38.403	1	.000
	Model	38.403	1	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4998.548[a]	.007	.011

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			bmi more than 35		Percentage Correct
	Observed		no	yes	
Step 1	bmi more than 35	no	4838	0	100.0
		yes	914	0	.0
	Overall Percentage				84.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	SEDMIN	.001	.000	39.124	1	.000	1.001	1.001	1.002
	Constant	-2.053	.074	773.105	1	.000	.128		

a. Variable(s) entered on step 1: SEDMIN.

For testing

H0: β1=0

H1: β1≠0

Wald test statistics = 39.124 p-value = 0.000

p-value 0.00<0.05 Reject H0

Minutes of sedentary activity per week (SEDMIN) has significant contribution in the development of OBESE (bmi more than 35)

95% confidence limits for β1 (SEDMIN) are $\hat{\beta_1} \pm Z_{1-\frac{\alpha}{2}} \cdot S.E(\hat{\beta_1})$

0.001 ± 1.96 (0.000025) = (0.000951, 0.001049)

Confidence limits of $e^{\beta 1}$ are $e^{0.000951}$= 1.001 and $e^{0.001049}$= 1.002

We are 95% confidence that the mark of evidence of OBESE per one minute's increase in SEDMIN variance between 0.1% and 0.2%

When the -2LL value is small model is getting better

H0: β1=0

H1: β1≠0

-2LL with constant in the model 5036.951
-2LL with SEDMIN added to model is 4998.548 (smallest)
-2LL is smaller when the model with SEDMIN than model with only constant.

The likelihood ratio statistic is the change in -2LL

5036.951-4998.548 = 38.403 which is the Chi-square statistic and is significant (p-value = 0.00)

p-value<0.05 reject H0 Overall model is better model with SEDMIN.

H0: β1=0

H1: β1≠0

The score is of SEDMIN 39.505. p-value is 0.00

p-value< 0.05 Reject H0

There for SEDMIN is significant for prediction of OBESE (bmi more than 35).

Therefore, results of these tests consistent with one another.

8. Compute the logit and the estimated logistic probability of being obese with 420 minutes of sedentary activity per week (one hour daily).

Estimated logit transformation

$$\ln[\frac{p^\wedge}{1-p^\wedge}] = -2.053+0.001*\text{SEDMIN}$$

$$\ln[\frac{p^\wedge}{1-p^\wedge}] = -2.053+0.001*420 = -1.633$$

Estimated logistic probability

$$P^\wedge = \frac{e^{-2.053+0.001*\text{SEDMIN}}}{1+e^{-2.053+0.001*\text{SEDMIN}}} = 0.1634$$

9. Repeat exercises 4 and 5 above using the variable '' walk and use bicycle to get to and from places", WLKBIK as the only covariate

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	23.560[a]	1	.000		
Continuity Correction[b]	23.167	1	.000		
Likelihood Ratio	24.745	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.556	1	.000		
N of Valid Cases	5822				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 243.69.

b. Computed only for a 2x2 table

WLKBIK is significant

Categorical Variables Codings

		Frequency	Parameter coding (1)
walk or bicycle	yes	1519	1.000
	no	4303	.000

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	WLKBIK	.426	.088	23.300	1	.000	1.532	1.288	1.821
	Constant	-1.982	.079	635.096	1	.000	.138		

a. Variable(s) entered on step 1: WLKBIK.

Estimated logit transformation

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = -1.982 + 0.426 * \text{WLKBIK}$$

For WLKBIK =0 (yes)

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = -1.982$$

For WLKBIK =1 (no)

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = -1.982 + 0.426 = -1.556$$

Estimated logistics probabilities

$$\left[\frac{\hat{p}}{1-\hat{p}}\right] = e^{-1.982+0.426*\text{WLKBIK}}$$

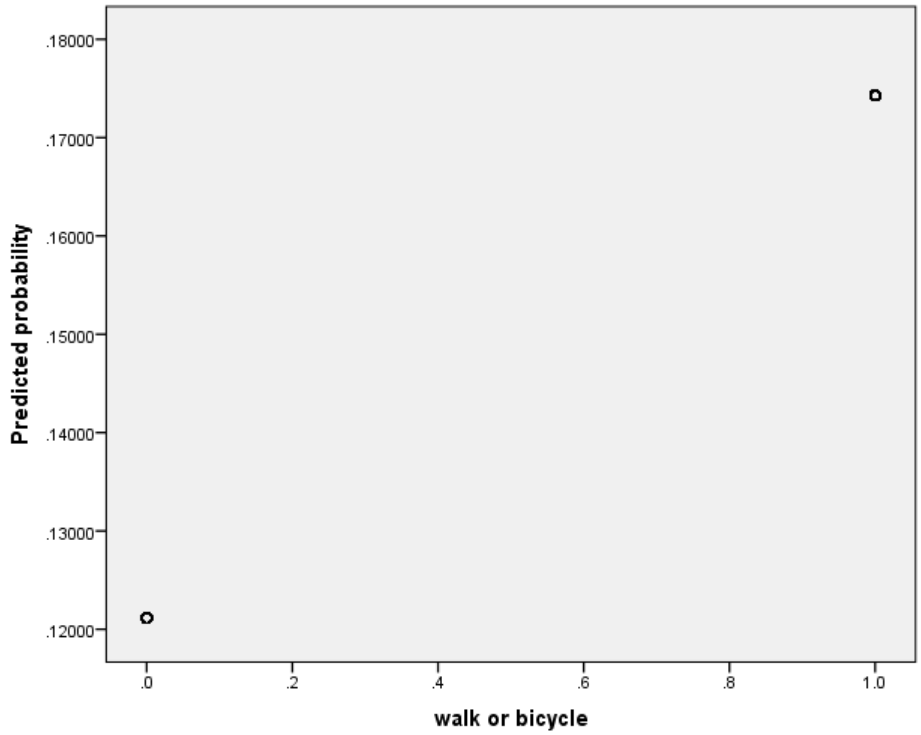
$$\hat{P} = \frac{e^{-1.982+0.426*\text{WLKBIK}}}{1+e^{-1.982+0.426*\text{WLKBIK}}}$$

For WLKBIK =0 (yes)

P^ = 0.121

For WLKBIK =1 (no)

P^ = 0.174



According to the graph predicted probability for those who do the walk or bicycle or not probability gather closer to two points 0.12 and 0.175. But both probabilities very closer to the zero. So predicted probability shows that regardless of "walk or bicycle" BMI less than or equal to 35.

10. Determine and interpret the slope coefficient (log odds and odds) for WLKBIK. Assess the significance of the slope coefficient using the Likelihood ratio test, the Wald test and the Score test. Are the results of these tests consistent with one another? Use alpha = 0.05.

$$\ln[\frac{p\hat{}}{1-p\hat{}}] = -1.982+0.426*\text{WLKBIK}$$

Log odds = b1= 0.426

Log odds increase by 0.426 when "walk or bicycle" status goes to not walk or bicycle.
(0.426 is the change in log odds comparing "do not walk or bicycle" to "walk or bicycle")

Odds $=e^{b1} = 1.532$

Odds of "BMI more than 35" among those people who do not "walk or bicycle" is 1.532 times as likely than those people who do "walk or bicycle"

Categorical Variables Codings

			Parameter coding
		Frequency	(1)
walk or bicycle	yes	1519	1.000
	no	4303	.000

Block 0: Beginning Block

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	5201.517	-1.358
	2	5128.256	-1.629
	3	5127.733	-1.655
	4	5127.733	-1.655

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 5127.733

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a,b]

Observed			Predicted		
			bmi more than 35		Percentage Correct
			no	yes	
Step 0	bmi more than 35	no	4888	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.655	.036	2148.001	1	.000	.191

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	WLKBIK	23.560	1	.000
	Overall Statistics		23.560	1	.000

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients	
			Constant	WLKBIK
Step 1	1	5184.226	-1.515	.213
	2	5103.927	-1.915	.378
	3	5102.988	-1.980	.425
	4	5102.988	-1.982	.426
	5	5102.988	-1.982	.426

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 5127.733

d. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24.745	1	.000
	Block	24.745	1	.000
	Model	24.745	1	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5102.988[a]	.004	.007

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted		
			bmi more than 35		Percentage Correct
			no	yes	
Step 1	bmi more than 35	no	4888	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	WLKBIK	.426	.088	23.300	1	.000	1.532	1.288	1.821
	Constant	-1.982	.079	635.096	1	.000	.138		

a. Variable(s) entered on step 1: WLKBIK.

For testing

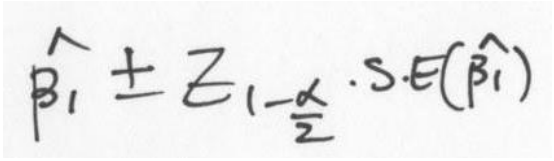
H0: β1=0

H1: β1≠0

Wald statistic = 23.30

p-value 0.00<0.05 Reject H0

WLKBIK has significant contribution in the development of OBESE (bmi more than 35)

95% confidence limits for β1 (WLKBIK) are 

0.426 ± 1.96 (0.088) = (0.2535, 0.5984)

Confidence limits of $e^{\beta 1}$ are $e^{0.2535}$= 1.288 and $e^{0.5984}$= 1.819

We are 95% confidence that variance of OBESE between 28.8% to 81.9% on status (yes or no) of WLKBIK.

When the -2LL value is small model is better

H0: β1=0

H1: β1≠0

-2LL with constant in the model 5127.733
-2LL with WLKBIK added to model is 5102.988 (smallest)
-2LL is smaller when the model with WLKBIK than model with only constant.

The likelihood ratio statistic is the change in -2LL

5127.733-5102.988 = 24.745 which is the Chi-square statistic

p-value (0.000) <0.05 Reject H0

WLKBIK is significant. Overall model with WLKBIK is the best model.

H0: β1=0

H1: β1≠0

The score is of WLKBIK 23.56. p-value is 0.00 <0.05 Reject H0
Therefore, WLKBIK is significant.

Therefore, results of these tests consistent with one another.

11. Compute and compare the logit and the estimated logistic probability of being obese for a person who does and does not walk or bike to places.

person who does not walk or bike to places (WLKBIK=1**)**
Estimated logit transformation

$$\ln[\frac{p^\wedge}{1-p^\wedge}] = -1.982+0.426*1$$

$$\ln[\frac{p^\wedge}{1-p^\wedge}] = -1.982+0.426 = -1.556$$

Logit for those who does not walk/bicycle is -1.556

Log odds for **person who does not walk or bike**=β0 + β1 = -1.982+0.426 = -1.556

Estimated logistic probability

$$P^\wedge = \frac{e^{-1.556}}{1+e^{-1.556}} = 0.174$$

Person who does not walk/bicycle have 0.174 probability of being obese.

person who does walk or bike to places (WLKBIK=0**)**
Estimated logit transformation

$$\ln[\frac{p^\wedge}{1-p^\wedge}] = -1.982$$

Logit for those who does walk/bicycle is -1.982

Log odds for **person who does walk or bike** =β0 = -1.982

Estimated logistic probability

$$P^\wedge = \frac{e^{-1.982}}{1+e^{-1.982}} = 0.121$$

Person who does walk/bicycle have 0.121 probability of being obese.

Comparing the log odds

0.426 is the change in log odds comparing “do not walk or bicycle” to “walk or bicycle”.

12. Interpret the Classification table and the Histogram of estimated probabilities.

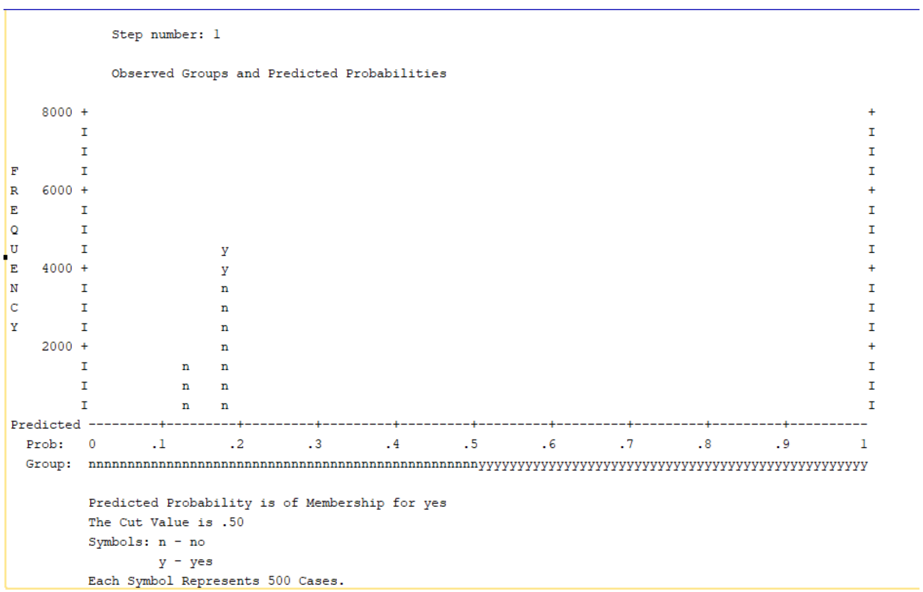
WLKBIK and BMI

Classification Table[a]

			Predicted		
			bmi more than 35		Percentage Correct
Observed			no	yes	
Step 1	bmi more than 35	no	4888	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. The cut value is .500

The model with variable WLKBIK can predict the “BMI more than 35” yes or no with 84.0% overall accuracy percentage. This model is predicting all 934 people who have “BMI more than 35” as “no” and all other people who have BMI less than or equal 35 were predicted 100% as it is.



In this histogram one symbol represent 500 cases. It shows as all most all the people do not have BMI more than 35. Because all the predicted probabilities very closer to the zero which means they do not have “bmi more than 35”. So this is conforming what we saw in previous classification table. So people who have BMI more than 35 also shows as they do not have BMI more than 35.

B: The data set OBESITY.SAV represents data from a sample of 5857 subjects from the general population in the United States between 2009 and 2010. The goal is to identify risk factors that predict obesity.

1. Using variables MODRECEXR (moderate recreational activity) and VIGRECEXR (vigorous recreational activity), create a new variable RECACT (recreational activity) with 4 groups based on the following variable values and identify the proportion of subjects in each group, after labelling the variable values.

1 INACTIVE (no moderate or vigorous recreational activity)
2 LOW (only moderate recreational activity)
3 MODERATE (only vigorous recreational activity)
4 HIGH (both moderate and recreational activity)

VIGWRK	MODWRK	WLKBIK	VIGRECEXR	MODRECEXR	SEDMIN	OBESE	Logit_sed	recact1
1	1	1	1	1	480	0	-1.49	1
1	1	1	1	1	240	1	-1.77	1
1	1	0	0	1	720	0	-1.20	3
1	1	1	1	1	240	0	-1.77	1
1	0	0	1	0	60	0	-1.98	2
1	0	1	1	1	540	0	-1.42	1
1	1	1	1	1	30	1	-2.02	1
1	1	1	1	1	.	0	.	1
1	0	1	0	0	720	0	-1.20	4
1	1	0	0	0	30	0	-2.02	4
1	1	1	1	1	600	0	-1.35	1
1	0	1	1	0	180	0	-1.84	2
1	1	1	1	1	60	0	-1.98	1
1	1	1	0	0	540	0	-1.42	4
0	1	1	1	0	180	0	-1.84	2
1	1	0	1	1	180	1	-1.84	1
1	1	0	1	1	120	0	-1.91	1
1	1	0	1	1	120	0	-1.91	1
1	0	1	1	1	240	0	-1.77	1
1	1	1	1	0	360	0	-1.63	2
1	1	0	0	1	120	0	-1.91	3
1	1	0	0	0	120	0	-1.91	4

recreational activity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	inactive	3237	55.3	55.3	55.3
	low	1542	26.3	26.3	81.6
	moderate	379	6.5	6.5	88.1
	high	698	11.9	11.9	100.0
	Total	5856	100.0	100.0	
Missing	System	1	.0		
Total		5857	100.0		

In the sample 55.3% represent inactive group, 26.3% represent low group, 6.5% represent moderate group and 11.9% represent high group. There is a one missing value.

recreational activity * bmi more than 35 Crosstabulation

			bmi more than 35		
			no	yes	Total
recreational activity	inactive	Count	2590	620	3210
		% within recreational activity	80.7%	19.3%	100.0%
		% within bmi more than 35	53.0%	66.4%	55.1%
	low	Count	1317	219	1536
		% within recreational activity	85.7%	14.3%	100.0%
		% within bmi more than 35	26.9%	23.4%	26.4%
	moderate	Count	342	37	379
		% within recreational activity	90.2%	9.8%	100.0%
		% within bmi more than 35	7.0%	4.0%	6.5%
	high	Count	638	58	696
		% within recreational activity	91.7%	8.3%	100.0%
		% within bmi more than 35	13.1%	6.2%	12.0%
Total		Count	4887	934	5821
		% within recreational activity	84.0%	16.0%	100.0%
		% within bmi more than 35	100.0%	100.0%	100.0%

People who are in the inactive group 19.3% have “bmi more than 35”.
People who are in the low group 14.3% have “bmi more than 35”.
People who are in the moderate group 9.8% have “bmi more than 35”.
People who are in the high group 8.3% have “bmi more than 35”.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	70.948[a]	3	.000
Likelihood Ratio	76.697	3	.000
Linear-by-Linear Association	68.762	1	.000
N of Valid Cases	5821		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 60.81.

Variable RECACT is significant.

2. In the logistic regression of OBESE on RECACT, use the first group and then the last group as the reference group, and interpret the odds ratios in each case.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
bmi more than 35 * recreational activity	5821	99.4%	36	0.6%	5857	100.0%

bmi more than 35 * recreational activity Crosstabulation

Count

		recreational activity				Total
		inactive	low	moderate	high	
bmi more than 35	no	2590	1317	342	638	4887
	yes	620	219	37	58	934
Total		3210	1536	379	696	5821

First group as reference group

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	5821	99.4
	Missing Cases	36	.6
	Total	5857	100.0
Unselected Cases		0	.0
Total		5857	100.0

a. If weight is in effect, see classification table for the total number of cases.

Categorical Variables Codings

		Frequency	Parameter coding		
			(1)	(2)	(3)
recreational activity	inactive	3210	.000	.000	.000
	low	1536	1.000	.000	.000
	moderate	379	.000	1.000	.000
	high	696	.000	.000	1.000

Block 0

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	5201.116	-1.358
	2	5127.906	-1.629
	3	5127.383	-1.655
	4	5127.383	-1.655

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 5127.383

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a,b]

Observed			Predicted		
			bmi more than 35		Percentage Correct
			no	yes	
Step 0	bmi more than 35	no	4887	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.655	.036	2147.399	1	.000	.191

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	recact	70.948	3	.000
		recact(1)	4.949	1	.026
		recact(2)	11.879	1	.001
		recact(3)	34.902	1	.000
	Overall Statistics		70.948	3	.000

Block 1

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients			
			Constant	recact(1)	recact(2)	recact(3)
Step 1	1	5148.908	-1.227	-.202	-.382	-.439
	2	5053.059	-1.418	-.336	-.688	-.813
	3	5050.693	-1.430	-.364	-.789	-.957
	4	5050.686	-1.430	-.364	-.794	-.968
	5	5050.686	-1.430	-.364	-.794	-.968

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 5127.383

d. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	76.697	3	.000
	Block	76.697	3	.000
	Model	76.697	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5050.686[a]	.013	.022

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			bmi more than 35		Percentage Correct
	Observed		no	yes	
Step 1	bmi more than 35	no	4887	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	recact			68.152	3	.000			
	recact(1)	-.364	.086	18.124	1	.000	.695	.587	.822
	recact(2)	-.794	.179	19.742	1	.000	.452	.318	.642
	recact(3)	-.968	.144	45.051	1	.000	.380	.286	.504
	Constant	-1.430	.045	1022.521	1	.000	.239		

a. Variable(s) entered on step 1: recact.

In the model overall ‘recact’ variable is significant

Wald statistic is 68.152 and p-value = 0.000<0.05

Effect of the “low”, “moderate” and “high” are statically significant when compare with the inactive group. All p-value <0.05

OR^(low)=0.695 The odds of Obese in the low group are 30.5% less likely than in the inactive group.
OR^(moderate)=0.452 The odds of Obese in the moderate group are 54.8% less likely than in the inactive group.
OR^(high)=0.380 The odds of Obese in the high group are 62% less likely than in the inactive group.

```
            Step number: 1

            Observed Groups and Predicted Probabilities

     4000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                   y                                                                                I
R    3000 +                   y                                                                                +
E         I                   y                                                                                I
Q         I                   n                                                                                I
U         I                   n                                                                                I
E    2000 +                   n                                                                                +
N         I                   n                                                                                I
C         I              y    n                                                                                I
Y         I              n    n                                                                                I
     1000 +              n    n                                                                                +
          I        n     n    n                                                                                I
          I        ny    n    n                                                                                I
          I        nn    n    n                                                                                I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy

          Predicted Probability is of Membership for yes
          The Cut Value is .50
          Symbols: n - no
                   y - yes
          Each Symbol Represents 250 Cases.
```

Last group as reference group

Categorical Variables Codings

			Parameter coding		
		Frequency	(1)	(2)	(3)
recreational activity	inactive	3210	1.000	.000	.000
	low	1536	.000	1.000	.000
	moderate	379	.000	.000	1.000
	high	696	.000	.000	.000

Block 0

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	5201.116	-1.358
	2	5127.906	-1.629
	3	5127.383	-1.655
	4	5127.383	-1.655

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 5127.383

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a,b]

			Predicted		
			bmi more than 35		Percentage Correct
	Observed		no	yes	
Step 0	bmi more than 35	no	4887	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.655	.036	2147.399	1	.000	.191

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	recact	70.948	3	.000
		recact(1)	56.782	1	.000
		recact(2)	4.949	1	.026
		recact(3)	11.879	1	.001
	Overall Statistics		70.948	3	.000

Block 1

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients			
			Constant	recact(1)	recact(2)	recact(3)
Step 1	1	5148.908	-1.667	.439	.237	.057
	2	5053.059	-2.232	.813	.478	.125
	3	5050.693	-2.387	.957	.594	.169
	4	5050.686	-2.398	.968	.604	.174
	5	5050.686	-2.398	.968	.604	.174

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 5127.383

d. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	76.697	3	.000
	Block	76.697	3	.000
	Model	76.697	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5050.686[a]	.013	.022

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted		
			bmi more than 35		Percentage Correct
			no	yes	
Step 1	bmi more than 35	no	4887	0	100.0
		yes	934	0	.0
	Overall Percentage				84.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	recact			68.152	3	.000			
	recact(1)	.968	.144	45.051	1	.000	2.633	1.985	3.494
	recact(2)	.604	.155	15.109	1	.000	1.829	1.349	2.480
	recact(3)	.174	.221	.621	1	.431	1.190	.772	1.835
	Constant	-2.398	.137	305.703	1	.000	.091		

a. Variable(s) entered on step 1: recact.

In the model overall 'recact' variable is significant

Wald statistic is 68.152 and p-value = 0.000<0.05

Effect of the" inactive" and "low" groups are statically significant when compare with the high group. p-value <0.05

But moderate group is not statically significant when compare with high group.
Wald statistic = 0.621 and p-value = 0.431 > 0.05 not significant

OR^ (inactive)=2.633 The odds of Obese in the inactive group are 2.633 times more likely to be occurred when compare with the high group.
OR^ (low)=1.829 The odds of Obese in the low group are 1.829 times as likely to be occurred when compare with the high group.
OR^ (high)=1.19 The odds of Obese in the moderate group are 1.19 (19%) times as likely to be occurred when compare with the high group.

3. Demonstrate that the value of the odds ratio obtained from the cross-classification of OBESE (columns) by RECACT (rows) is identical to the odds ratio from the logistic regression of OBESE on RECACT.

First group as reference group (Inactive group)

recreational activity * bmi more than 35 Crosstabulation

Count

		bmi more than 35		
		no	yes	Total
recreational activity	inactive	2590	620	3210
	low	1317	219	1536
	moderate	342	37	379
	high	638	58	696
Total		4887	934	5821

Odds ratio^ for low = ((219/1536)/(1317/1536))/ ((620/3210)/(2590/3210)) = 0.6946
Odd ratio^ for moderate = ((37/379)/(342/379))/ ((620/3210)/(2590/3210)) = 0.4519
Odd ratio^ for high = ((58/696)/(638/696))/ ((620/3210)/(2590/3210)) = 0.3797

The value of the odds ratio obtained from the cross-classification of OBESE (columns) by RECACT (rows) is identical to the odds ratio from the logistic regression of OBESE on RECACT.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	recact			68.152	3	.000			
	recact(1)	-.364	.086	18.124	1	.000	.695	.587	.822
	recact(2)	-.794	.179	19.742	1	.000	.452	.318	.642
	recact(3)	-.968	.144	45.051	1	.000	.380	.286	.504
	Constant	-1.430	.045	1022.521	1	.000	.239		

a. Variable(s) entered on step 1: recact.

Last group as reference group (high group)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	recact			68.152	3	.000			
	recact(1)	.968	.144	45.051	1	.000	2.633	1.985	3.494
	recact(2)	.604	.155	15.109	1	.000	1.829	1.349	2.480
	recact(3)	.174	.221	.621	1	.431	1.190	.772	1.835
	Constant	-2.398	.137	305.703	1	.000	.091		

a. Variable(s) entered on step 1: recact.

Odds ratio^ for inactive = ((620/3210)/(2590/3210))/ ((58/696)/(638/696))= 2.633
Odds ratio^ for low = ((219/1536)/(1317/1536))/ ((58/696)/(638/696))= 1.829
Odd ratio^ for moderate = ((37/379)/(342/379))/ ((58/696)/(638/696))= 1.19

The value of the odds ratio obtained from the cross-classification of OBESE (columns) by RECACT (rows) is identical to the odds ratio from the logistic regression of OBESE on RECACT.