



# ISDS 415

## Decision Support & Business Intelligence Systems

Logistic Regression HW

By

Archit Jajoo

Spring 2018

### Question Homework

A health researcher wants to be able to predict whether the incidence of heart disease can be predicted based on age, weight, gender and maximal aerobic capacity ( $VO_2\text{max}$ ) (an indicator of fitness and health). To this end, the researcher recruited 100 participants to perform a maximum  $VO_2\text{max}$  test as well as recording their age, weight and gender. The participants were also evaluated for the presence of heart disease. Run a logistic regression to determine whether the presence of heart disease could be predicted from their  $VO_2\text{max}$ , age, weight and gender. *Note:* this data is fictitious.

Give your answer showing the output from SPSS and your interpretation.

**Ans:**

We can see the variation from the model summary table by comparing the Cox & Snell R square and Nagelkerke R square. Looking at the values it can be determined that the variation in the dependent variable ranges from 31.6% to 43.7%. Also, looking at the classification table it can be determined that the cut-off value is 0.5 and anything greater than 0.5 suggests presence of heart disease under Yes column and vice-versa. (Refer Appendix for Model)

After performing the logistics regression with the dependent variable (heart disease) and independent variable (age, weight, gender and  $VO_2\text{max}$ ) it was found that the model is significant. The p value in the Sig. column was more than 0.5 for Weight which means that this factor did not contribute for predicting while the others did. (Refer Appendix for Model)

## Appendix

### Logistic Regression

[DataSet1] /Users/architjajoo/Desktop/d/CSUF classes/Semester 4/ISDS 415/HW 3/LR data file.sav

Case Processing Summary			
Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	97	100.0
	Missing Cases	0	.0
	Total	97	100.0
Unselected Cases		0	.0
Total		97	100.0

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

#### Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

#### Categorical Variables Codings

		Frequency	Parameter coding (1)
Gender	Female	35	1.000
	Male	62	.000

### Block 0: Beginning Block

#### Classification Table<sup>a,b</sup>

			Predicted		Percentage Correct
			Presence of Heart Disease		
Step 0	Observed		No	Yes	
	Presence of Heart Disease	No	64	0	100.0
		Yes	33	0	.0
	Overall Percentage				

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	-.662	.214	9.553	1	.002	.516

#### Variables not in the Equation

		Score	df	Sig.
Step 0	Variables	Age	16.445	.000
		Weight	8.822	.003
		Gender(1)	4.795	.029
		VO2max	5.905	.015
Overall Statistics		30.138	4	.000

## Block 1: Method = Enter

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	36.824	4	.000
	Block	36.824	4	.000
	Model	36.824	4	.000

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	87.564 <sup>a</sup>	.316	.437

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11.244	8	.188

### Contingency Table for Hosmer and Lemeshow Test

		Presence of Heart Disease = No		Presence of Heart Disease = Yes		Total
		Observed	Expected	Observed	Expected	
Step 1	1	10	9.806	0	.194	10
	2	10	9.483	0	.517	10
	3	10	9.010	0	.990	10
	4	9	8.276	1	1.724	10
	5	7	7.530	3	2.470	10
	6	7	6.782	3	3.218	10
	7	2	5.575	8	4.425	10
	8	4	4.203	6	5.797	10
	9	3	2.740	7	7.260	10
	10	2	.594	5	6.406	7

### Classification Table<sup>a</sup>

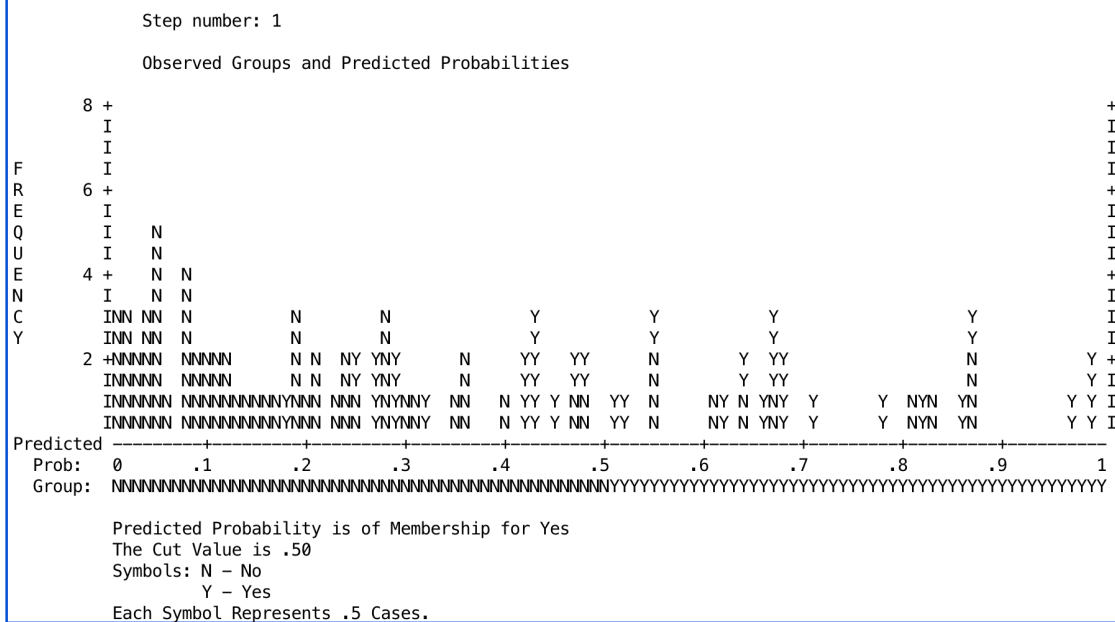
		Predicted		Percentage Correct
		Presence of Heart Disease No	Presence of Heart Disease Yes	
Step 1	Observed			
	Presence of Heart Disease	No	Yes	
	No	55	9	85.9
	Yes	15	18	54.5
	Overall Percentage			75.3

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 <sup>a</sup>	Age	.122	.036	11.652	1	.001	1.130	1.053	1.212
	Weight	-.013	.025	.269	1	.604	.987	.940	1.036
	Gender(1)	-2.859	1.002	8.146	1	.004	.057	.008	.408
	VO2max	-.173	.062	7.664	1	.006	.841	.745	.951
	Constant	3.602	4.382	.676	1	.411	36.671		

a. Variable(s) entered on step 1: Age, Weight, Gender, VO2max.



### Casewise List<sup>b</sup>

Selected Status <sup>a</sup>		Observed Presence of Heart Disease	Predicted	Predicted Group	Temporary Variable	
Case					Resid	ZResid
34	S	N**	.868	Y	-.868	-2.568
73	S	N**	.868	Y	-.868	-2.569

a. S = Selected, U = Unselected cases, and \*\* = Misclassified cases.

b. Cases with studentized residuals greater than 2.000 are listed.