

EXAM Take-Home 2: Part A.

Categorical Variables Codings

		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
ethnic	CAUCASIAN	26	.000	.000	.000	.000
	BLACK	4	1.000	.000	.000	.000
	HISPANIC	6	.000	1.000	.000	.000
	ASIAN	2	.000	.000	1.000	.000
	OTHER/DTS	2	.000	.000	.000	1.000
exercise	no	13	.000	.000		
	light	15	1.000	.000		
	regular	12	.000	1.000		

Block 0: Beginning Block**Classification Table^{a,b}**

		Predicted		
		2nd Heart Attack		Percentage
		0	1	Correct
Observed				
Step 0	2nd Heart Attack 0	0	20	.0
	1	0	20	100.0
Overall Percentage				50.0

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	.000	.316	.000	1	1.000	1.000

I ran a binary logistic regression analysis to determine which predictors are significant when predicting the likelihood of a second heart attack. The above information shows the predicted result without considering any predictors. As you can see in the classification table, the predicted percentage is about 50% accurate, which is not better than chance. There are also 2 categorical variables: exercise and ethnicity. I have included the categorical variable coding table for reference on how those were arranged. Further analysis is needed.

Block 1: Method = Forward Stepwise (Wald)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	35.370 ^a	.395	.526
2	27.108 ^b	.508	.677
3	18.318 ^c	.605	.806

- a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.
- b. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.
- c. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

The binary logistic regression analysis was then conducted using the Forward Stepwise (Wald) method. The -2 Log likelihood reflects that the best fit model includes total cholesterol level, smoking, and exercise information (Step 3, -2 Log likelihood = 18.318, Nagelkerke R^2 = .806). The only predictor not included is ethnicity, as there is not a significant impact on the risk of a second heart attack. Therefore, I will use Step 3 for the rest of the analysis.

Step	Chi-square	df	Sig.
1	3.772	8	.877
2	5.719	8	.679
3	1.487	8	.993

The Hosmer and Lemeshow test for Step 3 reflects a Chi-Square of 1.487, p = .993. Since α = .05, I can conclude that this test is not significant and therefore the model is a good fit for analysis.

Classification Table^a

	Observed	Predicted		Percentage Correct
		2nd Heart Attack 0	1	
Step 1	2nd Heart Attack 0	16	4	80.0
	1	5	15	75.0
	Overall Percentage			77.5
Step 2	2nd Heart Attack 0	18	2	90.0
	1	2	18	90.0
	Overall Percentage			90.0
Step 3	2nd Heart Attack 0	17	3	85.0
	1	2	18	90.0
	Overall Percentage			87.5

a. The cut value is .500

The above classification table for Step 3 shows a significant difference between the prediction accuracy without consideration of predictors (50%) and the prediction accuracy when 3 predictors (total cholesterol levels, smoking, and exercise) are considered (87.5%).

Part C.**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)		
							Lower	Upper	
Step 1 ^a	Total blood cholesterol level	.040	.012	11.169	1	.001	1.041	1.017	1.066
	Constant	-9.379	2.817	11.086	1	.001	.000		
Step 2 ^b	Total blood cholesterol level	.037	.013	7.749	1	.005	1.038	1.011	1.065
	smoking	2.706	1.047	6.672	1	.010	14.962	1.921	116.563
	Constant	-9.731	3.289	8.754	1	.003	.000		
Step 3 ^c	Total blood cholesterol level	.037	.016	5.413	1	.020	1.037	1.006	1.070
	exercise			4.982	2	.083			
	exercise(1)	-3.413	1.810	3.555	1	.059	.033	.001	1.145
	exercise(2)	-4.232	2.094	4.085	1	.043	.015	.000	.880
	smoking	3.655	1.641	4.959	1	.026	38.664	1.550	964.571
	Constant	-7.815	3.729	4.391	1	.036	.000		

a. Variable(s) entered on step 1: Total blood cholesterol level.

b. Variable(s) entered on step 2: smoking.

c. Variable(s) entered on step 3: exercise.

The above chart shows the key logistic regression information. When looking at Step 3, you can see that a higher blood cholesterol level will increase the risk of a second heart attack by 1.037 when compared to lower blood cholesterol levels. Light exercise will decrease the risk of a second heart attack by .033 in comparison to no exercise, making someone 30.3 times more likely to have a second heart attack with no exercise than if they lightly exercised. Normal exercise will decrease the risk of a second heart attack even further by .015, making someone 66.67 times more likely to have a second heart attack than if they exercised a normal amount. Smoking increases the risk of a second heart attack by 38.664. The most significant predictor of having a second heart attack is high blood cholesterol (alpha =.05, p=.020).

Part B.

The classification table for Step 3 in part A (also below) shows a significant difference between the prediction accuracy without consideration of predictors (50%) and the prediction accuracy when 3 predictors (total cholesterol levels, smoking, and exercise) are considered (87.5%), indicating this is an effective model to predict the second heart attack risk. The regression equation is:

$$\text{Logit}(p) = -7.815 + .037(\text{total cholesterol}) - 3.413(\text{light exercise}) - 4.232(\text{regular exercise}) + 3.655(\text{smoking})$$

Classification Table^a

		Observed	Predicted		Percentage Correct
			2nd Heart Attack 0	1	
Step 1	2nd Heart Attack	0	16	4	80.0
		1	5	15	75.0
	Overall Percentage				77.5
Step 2	2nd Heart Attack	0	18	2	90.0
		1	2	18	90.0
	Overall Percentage				90.0
Step 3	2nd Heart Attack	0	17	3	85.0
		1	2	18	90.0
	Overall Percentage				87.5

a. The cut value is .500