

## BSTA450 Assignment 5

- A. Use logistic regression to determine whether DIFF, PCTHIGH or PCTLOW are useful in predicting the winners of games in the NCAA basketball tournament. Use a 5% level of significance in any hypothesis test

Logistic Regression Results					
The LOGISTIC Procedure					
Model Information					
Data Set	WORK.SORTTEMPTABLESORTED				
Response Variable	WIN				
Number of Response Levels	2				
Model	binary logit				
Optimization Technique	Fisher's scoring				
Number of Observations Read		63			
Number of Observations Used		63			
Response Profile					
Ordered Value	WIN	Total Frequency			
1	0	21			
2	1	42			
Probability modeled is WIN='0'.					
Model Convergence Status					
Convergence criterion (GCONV=1E-8) satisfied.					
Model Fit Statistics					
Criterion	Intercept Only	Intercept and Covariates			
AIC	82.201	65.762			
SC	84.344	74.334			
-2 Log L	80.201	57.762			
Testing Global Null Hypothesis: BETA=0					
Test	Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio	22.4390	3	<.0001		
Score	18.0752	3	0.0004		
Wald	13.4570	3	0.0037		
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2041	4.0974	0.6115	0.4342
DIFF	1	0.4150	0.1137	13.3326	0.0003
PCTHIGH	1	3.2578	4.2213	0.5956	0.4403
PCTLOW	1	2.8826	4.2006	0.4709	0.4926
Odds Ratio Estimates					
Effect	Point Estimate		95% Wald Confidence Limits		
DIFF	1.514		1.212	1.892	
PCTHIGH	25.993		0.007	>999.999	
PCTLOW	17.861		0.005	>999.999	

DIFF chi-square = 13.3326

PCTHIGH chi-square = 0.5956

PCTLOW chi-square = 0.4709

### **DIFF :**

$$H_0 = B(\text{diff}) = 0$$

$$H_a = B(\text{diff}) \neq 0$$

Decision rule at  $\alpha = 5\%$  , degree of freedom = 1  $\chi^2 = 3.841$

-Reject  $H_0$  if  $\chi^2_{\text{model}} > 3.841$

-Do not reject  $H_0$  if  $\chi^2_{\text{model}} < 3.841$

$$\text{Diff chi-square} = 13.3326$$

Conclusion: As  $13.3326 > 3.841$ , we reject  $H_0$  and conclude that DIFF is useful for predicting the winner of the basketball games.

### **PCTHiGHT :**

$$H_0: B(\text{pctHigh}) = 0$$

$$H_a: B(\text{pctHigh}) \neq 0$$

Decision Rule at  $\alpha = 5\%$ , degrees of freedom = 1:

-Reject  $H_0$  if  $\chi^2_{\text{model}} > 3.841$

-Do not reject  $H_0$  if  $\chi^2_{\text{model}} < 3.841$

$$\text{Chi-Square pct high} = 0.5956$$

Conclusion: as  $0.5956 < 3.841$ , we do not reject  $H_0$  and conclude that PCTHIGH is not useful in predicting the winner of the games.

### **PCT LOW:**

$$H_0: B(\text{PctLow}) = 0$$

$$H_a: B(\text{PctLow}) \neq 0$$

Decision Rule at  $\alpha = 5\%$ , degrees of freedom 1,  $\chi^2 = 3.841$ :

-Reject  $H_0$  if  $\chi^2 > 3.841$

-Do not reject  $H_0$  if  $\chi^2 < 3.841$

Chi-Square PCT LOW= 0. 4709

Conclusion: as  $3.840 > 0.470$ , we do not reject  $H_0$  and conclude that PCTLow is not useful in predicting the winner of the games.

B. Add ROUND into your model. Use logistic regression to determine whether DIFF, PCTHIGH, PCTLOW or ROUND are useful in predicting the winners of games in the NCAA basketball tournament. Use a 5% level of significance in any hypothesis test.

#### The LOGISTIC Procedure

Model Information	
Data Set	WORK.SORTTEMPTABLESORTED
Response Variable	WIN
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	63
Number of Observations Used	63

Response Profile		
Ordered Value	WIN	Total Frequency
1	0	21
2	1	42

Probability modeled is WIN='0'.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	82.201	67.083
SC	84.344	77.799
-2 Log L	80.201	57.083

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.1173	4	0.0001
Score	18.2891	4	0.0011
Wald	13.4692	4	0.0092

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.6750	4.5419	1.0595	0.3033
DIFF	1	0.4705	0.1367	11.8472	0.0006
ROUND	1	-0.2727	0.3322	0.6738	0.4117
PCTHIGH	1	5.1695	4.8291	1.1460	0.2844
PCTLOW	1	3.9682	4.4766	0.7857	0.3754

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
DIFF	1.601	1.225	2.093
ROUND	0.761	0.397	1.460
PCTHIGH	175.835	0.014	>999.999
PCTLOW	52.887	0.008	>999.999

Diff-Chi Square: 11.8472

PCTHIGH Chi Square: 1.1460

PCTLOW Chi Square: 0.7857

ROUND-Chi square: 0.6738

### **DIFF:**

Ho:  $B(\text{diff}) = 0$

Ha:  $B(\text{diff}) \neq 0$

Decision rule at alpha 5%, degrees of freedom 1,  $X^2 = 3.841$

-Reject Ho if  $X^2 > 3.841$

-Do not reject Ho if  $X^2 < 3.841$

Chi-square diff = 11.8472

Conclusion: as  $11.8472 > 3.841$ , we reject Ho and conclude that DIFF is useful in predicting the winner of the games.

### **PCTHIGH:**

Ho:  $B(\text{pctHigh}) = 0$

Ha:  $B(\text{PctHigh}) \neq 0$

Decision Rule at alpha 5%, degrees of freedom 1,  $X^2 = 3.841$

-Reject Ho if  $X^2 > 3.841$

-Do not reject Ho if  $X^2 < 3.841$

PctHigh Chi square = 1.1460

Conclusion: as  $1.1460 < 3.841$ , we do not reject Ho and conclude that PCThigh is not useful in predicting the winner of the games.

### **PCTLOW:**

Ho:  $B(\text{pctlow}) = 0$

Ha:  $B(\text{pctlow}) \neq 0$

Decision Rule at alpha 5%, degrees of freedom 1,  $X^2 = 3.841$ :

-Reject  $H_0$  if  $X^2 > 3.841$

-Do not reject  $H_0$  if  $X^2 < 3.841$

pctLow Chi-Square = 0.7857

Conclusion: as  $0.7857 < 3.841$ , we do not reject  $H_0$  and conclude that pctLow is not useful in predicting the winner of the games

## ROUND:

$H_0: B(\text{round}) = 0$

$H_a: B(\text{ROUND}) \neq 0$

Decision rule at alpha 5%, degrees of freedom 1,  $X^2 = 3.841$ :

-Reject  $H_0$  if  $X^2 > 3.841$

-Do not reject  $H_0$  if  $X^2 < 3.841$

Round Chi square : 0.6738

Conclusion: as  $0.6738 < 3.841$ , we do not reject  $H_0$  and conclude that Round isn't useful in predicting the winner of the games.

C. Are there any other variables that might be useful in addition to ones given in this problem?  
Please explain.

- Number of scores and missed shot as positive and negative variables respectively can help predict which team has higher probability of winning (Free throw and three pointer rate to help determine which team has higher accuracy)
- Number of minutes of possession of the ball during a game can help determine which team is more dominant and more likely to win
- Offensive and defensive rebound and pass rate to help determine which team cooperates better between teammates and have a higher probability of winning
- Number of players as a first-round picks to determine which team has better players