

Regression modeling in practice

Week 4 assignment

Logistic regression model

I chose **addhealth** as my data set. The relationship between general health (H1GH1) and body mass index (BMI) was analyzed. Frequency of exercise (H1DA6), frequency of feeling happy (H1FS11), closeness with mother (HaWP9) and closeness with father (H1WP13) were analyzed as confounding variables.

Hypothesis: High BMI decreases the health situation.

General health has more than two levels, so I collapse it into 2 levels. 1 represents health. 0 represent not health.

The logistic regression model was processed: **H1GH1=BMI**.

The LOGISTIC Procedure		
Model Information		
Data Set	WORK.NEW2	
Response Variable	H1GH1	general health
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.2261	1	<.0001
Score	21.9912	1	<.0001
Wald	20.9922	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	8.4703	0.7397	131.1206	<.0001
BMI	1	-0.1236	0.0270	20.9922	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI	0.884	0.838	0.932

$$\text{General health} = -0.12 * \text{BMI} + 8.47$$

P-value is **less than 0.0001**, therefore at 95% confidence level general health and BMI are significantly associated.

Odds ratio is **0.884**, with 95% confidence interval **from 0.838 to 0.932**, which means people who are not healthy are more likely to have high BMI.

This result **corresponds to what I hypothesize**.

Some **potential confounding variables** were analyzed.

1. H1GH1=BMI H1DA6

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	7.4414	0.7721	92.8939	<.0001
BMI	1	-0.1241	0.0276	20.2160	<.0001
H1DA6	1	0.8919	0.2309	14.9241	0.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI	0.883	0.837	0.932
H1DA6	2.440	1.552	3.836

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	80.3	Somers' D	0.606
Percent Discordant	19.7	Gamma	0.606
Percent Tied	0.0	Tau-a	0.005
Pairs	162864	c	0.803

H1DA6 does not have significant evidence to be a confounding variable.

2. H1GH1=BMI H1FS11

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	7.2589	0.8304	76.4064	<.0001
BMI	1	-0.1265	0.0277	20.8223	<.0001
H1FS11	1	0.6774	0.2282	8.8121	0.0030

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI	0.881	0.835	0.930
H1FS11	1.969	1.259	3.079

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	77.7	Somers' D	0.555
Percent Discordant	22.2	Gamma	0.555
Percent Tied	0.0	Tau-a	0.005
Pairs	162864	c	0.777

H1FS11 does not show significant evidence to be confounding variable.

3. H1GH1=BMI H1WP9

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	8.3450	1.1825	49.8029	<.0001
BMI	1	-0.1239	0.0271	20.9299	<.0001
H1WP9	1	0.0283	0.2089	0.0184	0.8922

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI	0.883	0.838	0.932
H1WP9	1.029	0.683	1.549

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	68.6	Somers' D	0.373
Percent Discordant	31.3	Gamma	0.373
Percent Tied	0.1	Tau-a	0.003
Pairs	162864	c	0.687

P-value is 0.89, therefore at 95% confidence level, the relationship between mother is not correlated with the general health.

4. H1GH1= BMI H1WP13

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	8.3586	0.9593	75.9282	<.0001
BMI	1	-0.1241	0.0271	20.9384	<.0001
H1WP13	1	0.0243	0.1337	0.0330	0.8559

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI	0.883	0.838	0.931
H1WP13	1.025	0.788	1.332

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	68.6	Somers' D	0.373
Percent Discordant	31.3	Gamma	0.373
Percent Tied	0.0	Tau-a	0.003
Pairs	162864	c	0.687

P-value is 0.86, therefore at 95% confidence level, relationship with father is not correlated with the general health.

Code:

```

1 *load data;
2 LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
3 data new; set mydata.addhealth_pds;
4 *set aside missing values;
5 if H1GH1=6 then H1GH1=.; if H1GH1=8 then H1GH1=.;
6 if H1GH59A=96 then H1GH59A=.; if H1GH59A=98 then H1GH59A=.; if H1GH59A=99 then H1GH59A=.;
7 if H1GH59B=96 then H1GH59B=.; if H1GH59B=98 then H1GH59B=.; if H1GH59B=99 then H1GH59B=.;
8 if H1GH60=996 then H1GH60=.; if H1GH60=998 then H1GH60=.; if H1GH60=999 then H1GH60=.;
9 if H1DA6=6 then H1DA6=.; if H1DA6=8 then H1DA6=.;
10 *calculate body mass index;
11 H1GH59=H1GH59A * 12 + H1GH59B;
12 BMI=H1GH60 * 0.454 / (H1GH59 * 0.0254)**2;
13 *collapse response variable;
14 if H1GH1=2 then H1GH1=1;
15 if H1GH1=3 then H1GH1=1;
16 if H1GH1=4 then H1GH1=1;*healthy;
17 if H1GH1=5 then H1GH1=0;*not healthy;
18 *add labels;
19 label H1GH1="general health"
20       H1DA6="frequency of exercise"
21       H1FS11="frequency of feeling happy"
22       H1WP9="how close with mother"
23       H1WP13="how close with father";

24
25 *logistic regression model of general health and BMI;
26 proc logistic descending;
27 model H1GH1=BMI;
28 run;
29 *add frequency of exercise to this model;
30 proc logistic descending;
31 model H1GH1=BMI H1DA6;
32 run;
33 *add frequency of feeling happy to this model;
34 proc logistic descending; model H1GH1=BMI H1FS11;
35 run;
36 *add closeness with mother to this model;
37 proc logistic descending; model H1GH1=BMI H1WP9;
38 run;
39 *add closeness with father to this model;
40 proc logistic descending; model H1GH1=BMI H1WP13;
41 run;
42
43

```