

# Predict Passing the course with logistic regression

```
/* mathlogreg2.sas */
%include 'readmath.sas';
title2 'Predict Passing the course (Y-N) with Logistic Regression';

if sex = 'Female' then gender=1; else if sex = 'Male' then gender=0;
if tongue = 'English' then mtongue=1; else if tongue='Other' then mtongue=0;

options pagesize=100;

proc logistic descending order=internal;
  title3 'HS variables';
  model passed = hsgpa hscalcalc hsengl;

proc logistic descending order=internal;
  title3 'HS gpa and calc, course2 and diagnostic test';
  class course2 / param=ref;
  model passed = hsgpa hscalcalc course2 precalc calc;

/* Notice course2 is now a 2 df test. We seem to have lost everyone in
category zero (no data) by including the diagnostic test. See "Class Level
Information. Anyway, it's now useless. */

proc logistic descending order=internal;
  title3 'HS gpa and calc, precalc and total score';
  model passed = hsgpa hscalcalc precalc totscore;
  precalc_n_totscore: test precalc = totscore = 0;

/* Decision: Keep precalc rather than totscore */

proc logistic descending order=internal;
  title3 'Try gender, ethnic and mother tongue controlling for good stuff';
  class ethnic (param=ref ref='East Indian');
  /* Specifying a reference category that's not the last value */
  model passed = hsgpa hscalcalc precalc ethnic gender mtongue;
  contrast 'Demographics'
    ethnic 1 0 0 0 0,
    ethnic 0 1 0 0 0,
    ethnic 0 0 1 0 0,
    ethnic 0 0 0 1 0,
    ethnic 0 0 0 0 1,
    gender 1,
    mtongue 1 / e;

/* Decision: Forget about gender and ethnicity. */

proc logistic descending order=internal;
  title3 'My model: HS gpa, HS calculus mark, and Precalculus subtest';
  model passed = hsgpa hscalcalc precalc;

proc logistic descending order=internal;
  title3 'Stepwise Logistic Regression';
  class ethnic course / param=ref;
  model passed = gender mtongue ethnic
    hsgpa hscalcalc hsengl
    course precalc calc totscore
  / selection=stepwise slentry=0.05 slstay=0.05 ;
```

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 HS variables

1

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	435

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	257
2	No	178

Probability modeled is passed='Yes'.

NOTE: 144 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	590.611	465.610
SC	594.686	481.912
-2 Log L	588.611	457.610

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	131.0005	3	<.0001
Score	110.3360	3	<.0001
Wald	82.9074	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-15.9885	2.0503	60.8083	<.0001
hsgpa	1	0.1491	0.0331	20.3218	<.0001
hscal	1	0.0582	0.0126	21.3526	<.0001
hsengl	1	0.00437	0.0164	0.0711	0.7898

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
hsgpa	1.161	1.088	1.238
hscal	1.060	1.034	1.086
hsengl	1.004	0.973	1.037

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.9	Somers' D	0.619
Percent Discordant	19.0	Gamma	0.620
Percent Tied	0.2	Tau-a	0.300
Pairs	45746	c	0.810

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 HS gpa and calc, course2 and diagnostic test

2

The LOGISTIC Procedure

Model Information

Data Set WORK.MATH  
 Response Variable passed Passed the course  
 Number of Response Levels 2  
 Model binary logit  
 Optimization Technique Fisher's scoring

Number of Observations Read 579  
 Number of Observations Used 375

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	234
2	No	141

Probability modeled is passed='Yes'.

NOTE: 204 observations were deleted due to missing values for the response or explanatory variables.

#### Class Level Information

Class	Value	Design Variables	
course2	1	1	0
	2	0	1
	3	0	0

#### Model Convergence Status

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

#### Model Fit Statistics

Criterion	Intercept	Intercept
	Only	and Covariates
AIC	498.554	379.569
SC	502.481	407.057
-2 Log L	496.554	365.569

#### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	130.9852	6	<.0001
Score	108.3421	6	<.0001
Wald	79.9399	6	<.0001

#### Type 3 Analysis of Effects

Effect	DF	Wald	Pr > ChiSq
		Chi-Square	
hsgpa	1	14.6472	0.0001
hscal	1	18.1207	<.0001
course2	2	0.4383	0.8032
precalc	1	8.5427	0.0035
calc	1	1.6624	0.1973

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-14.6426	2.3386	39.2026	<.0001
hsgpa	1	0.1203	0.0314	14.6472	0.0001
hscal	1	0.0596	0.0140	18.1207	<.0001
course2 1	1	0.0616	0.7473	0.0068	0.9343
course2 2	1	-0.2244	0.5129	0.1915	0.6617
precalc	1	0.2617	0.0895	8.5427	0.0035
calc	1	0.0843	0.0654	1.6624	0.1973

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 HS gpa and calc, course2 and diagnostic test

3

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
hsgpa	1.128	1.060	1.199
hscal	1.061	1.033	1.091
course2 1 vs 3	1.064	0.246	4.601
course2 2 vs 3	0.799	0.292	2.183
precalc	1.299	1.090	1.548
calc	1.088	0.957	1.237

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.4	Somers' D	0.670
Percent Discordant	16.4	Gamma	0.671
Percent Tied	0.1	Tau-a	0.315
Pairs	32994	c	0.835

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 HS gpa and calc, precalc and total score

4

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	375

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	234
2	No	141

Probability modeled is passed='Yes'.

NOTE: 204 observations were deleted due to missing values for the response or explanatory variables.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	498.554	376.007
SC	502.481	395.642
-2 Log L	496.554	366.007

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	130.5468	4	<.0001
Score	108.2737	4	<.0001
Wald	79.7057	4	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-14.6351	2.2803	41.1914	<.0001
hsgpa	1	0.1181	0.0311	14.4227	0.0001
hscal	1	0.0592	0.0136	18.9109	<.0001
precalc	1	0.1812	0.1250	2.0999	0.1473
totscore	1	0.0821	0.0650	1.5969	0.2063

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
hsgpa	1.125	1.059	1.196
hscal	1.061	1.033	1.090
precalc	1.199	0.938	1.532
totscore	1.086	0.956	1.233

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.4	Somers' D	0.670
Percent Discordant	16.4	Gamma	0.671
Percent Tied	0.1	Tau-a	0.315
Pairs	32994	c	0.835

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
precalc_n_totscore	13.8587	2	0.0010

5

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 Try gender, ethnic and mother tongue controlling for good stuff

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	370

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	232
2	No	138

Probability modeled is passed='Yes'.

NOTE: 209 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables				
ethnic	Asian	1	0	0	0	0
	East Indian	0	0	0	0	0
	Eastern European	0	1	0	0	0
	European not Eastern	0	0	1	0	0
	Middle-Eastern and Pakistani	0	0	0	1	0
	Other and DK	0	0	0	0	1

Model Convergence Status

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



Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	490.784	379.703
SC	494.698	422.752
-2 Log L	488.784	357.703

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	131.0808	10	<.0001
Score	110.8226	10	<.0001
Wald	80.9993	10	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
hsgpa	1	10.3205	0.0013
hscal	1	24.3884	<.0001
precalc	1	13.5282	0.0002
ethnic	5	4.5364	0.4750
gender	1	0.8343	0.3610
mtongue	1	0.1917	0.6615

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square
Intercept	1	-14.2314	2.4312	34.2655
hsgpa	1	0.1043	0.0325	10.3205
hscal	1	0.0687	0.0139	24.3884
precalc	1	0.3205	0.0871	13.5282
ethnic Asian	1	-0.3522	0.4642	0.5758
ethnic Eastern European	1	-0.0314	0.5126	0.0038
ethnic European not Eastern	1	0.2889	0.4266	0.4587
ethnic Middle-Eastern and Pakistani	1	-0.4521	0.5419	0.6963
ethnic Other and DK	1	0.6539	0.8954	0.5334
gender	1	0.2465	0.2698	0.8343
mtongue	1	-0.1485	0.3391	0.1917

6

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 Try gender, ethnic and mother tongue controlling for good stuff

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
hsgpa	0.0013
hscal	<.0001
precalc	0.0002
ethnic Asian	0.4480
ethnic Eastern European	0.9512
ethnic European not Eastern	0.4982
ethnic Middle-Eastern and Pakistani	0.4040
ethnic Other and DK	0.4652
gender	0.3610
mtongue	0.6615

Odds Ratio Estimates

Effect	Point Estimate
hsgpa	1.110
hscal	1.071
precalc	1.378
ethnic Asian vs East Indian	0.703
ethnic Eastern European vs East Indian	0.969
ethnic European not Eastern vs East Indian	1.335
ethnic Middle-Eastern and Pakistani vs East Indian	0.636
ethnic Other and DK vs East Indian	1.923
gender	1.279
mtongue	0.862

Odds Ratio Estimates

95% Wald  
Confidence Limits

1.042	1.183
1.042	1.101
1.161	1.634
0.283	1.746
0.355	2.647
0.579	3.081
0.220	1.840
0.333	11.120
0.754	2.171
0.443	1.676

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.6	Somers' D	0.674
Percent Discordant	16.2	Gamma	0.675
Percent Tied	0.2	Tau-a	0.316
Pairs	32016	c	0.837

Coefficients of Contrast Demographics

Parameter	Row1	Row2	Row3
Intercept	0	0	0
hsgpa	0	0	0
hscal	0	0	0
precalc	0	0	0
ethnicAsian	1	0	0
ethnicEastern_European	0	1	0
ethnicEuropean_not_Eastern	0	0	1
ethnicMiddle_Eastern__and_Pakist	0	0	0
ethnicOther__and_DK	0	0	0
gender	0	0	0
mtongue	0	0	0

Coefficients of Contrast Demographics

Parameter	Row4	Row5	Row6
Intercept	0	0	0
hsgpa	0	0	0
hscal	0	0	0
precalc	0	0	0
ethnicAsian	0	0	0
ethnicEastern_European	0	0	0
ethnicEuropean_not_Eastern	0	0	0
ethnicMiddle_Eastern__and_Pakist	1	0	0
ethnicOther__and_DK	0	1	0
gender	0	0	1
mtongue	0	0	0

Coefficients of Contrast Demographics

Parameter	Row7
Intercept	0
hsgpa	0
hscal	0
precalc	0
ethnicAsian	0
ethnicEastern_European	0
ethnicEuropean_not_Eastern	0
ethnicMiddle_Eastern__and_Pakist	0
ethnicOther__and_DK	0
gender	0
mtongue	1

Contrast Test Results

Contrast	DF	Wald Chi-Square	Pr > ChiSq
Demographics	7	6.0125	0.5383

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 My model: HS gpa, HS calculus mark, and Precalculus subtest

8

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	375

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	234
2	No	141

Probability modeled is passed='Yes'.

NOTE: 204 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	498.554	375.618
SC	502.481	391.326
-2 Log L	496.554	367.618

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	128.9358	3	<.0001
Score	107.7971	3	<.0001
Wald	79.6583	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-14.7970	2.2683	42.5550	<.0001
hsgpa	1	0.1173	0.0310	14.3281	0.0002
hscal	1	0.0638	0.0132	23.3346	<.0001
precalc	1	0.2989	0.0844	12.5464	0.0004

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
hsgpa	1.124	1.058	1.195
hscal	1.066	1.039	1.094
precalc	1.348	1.143	1.591

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.1	Somers' D	0.663
Percent Discordant	16.7	Gamma	0.665
Percent Tied	0.2	Tau-a	0.312
Pairs	32994	c	0.832

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 Stepwise Logistic Regression

9

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	368

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	230
2	No	138

Probability modeled is passed='Yes'.

NOTE: 211 observations were deleted due to missing values for the response or explanatory variables.

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables				
ethnic	Asian	1	0	0	0	0
	East Indian	0	1	0	0	0
	Eastern European	0	0	1	0	0
	European not Eastern	0	0	0	1	0
	Middle-Eastern and Pakistani	0	0	0	0	1
	Other and DK	0	0	0	0	0
course	1	1	0			
	2	0	1			
	3	0	0			

Step 0. Intercept entered:

Model Convergence Status

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

-2 Log L = 486.911

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
111.4077	14	<.0001

Step 1. Effect hscalc entered:

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	488.911	398.544
SC	492.819	406.360
-2 Log L	486.911	394.544

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	92.3666	1	<.0001
Score	85.5385	1	<.0001
Wald	68.7679	1	<.0001

---

Gender, Ethnicity and Math performance  
Predict Passing the course (Y-N) with Logistic Regression  
Stepwise Logistic Regression

10

The LOGISTIC Procedure

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
38.7030	13	0.0002

NOTE: No effects for the model in Step 1 are removed.



Step 2. Effect hsgpa entered:

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	488.911	382.659
SC	492.819	394.384
-2 Log L	486.911	376.659

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	110.2512	2	<.0001
Score	94.7678	2	<.0001
Wald	71.1676	2	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
22.0384	12	0.0371

NOTE: No effects for the model in Step 2 are removed.

Step 3. Effect precalc entered:

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	488.911	370.567
SC	492.819	386.199
-2 Log L	486.911	362.567

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	124.3437	3	<.0001
Score	104.2190	3	<.0001
Wald	77.6336	3	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
8.7229	11	0.6475

NOTE: No effects for the model in Step 3 are removed.

NOTE: No (additional) effects met the 0.05 significance level for entry into the model.

Gender, Ethnicity and Math performance  
 Predict Passing the course (Y-N) with Logistic Regression  
 Stepwise Logistic Regression

11

The LOGISTIC Procedure

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	hscal		1	1	85.5385		<.0001
2	hsgpa		1	2	17.3610		<.0001
3	precalc		1	3	13.8105		0.0002

Summary of Stepwise Selection

Variable  
 Step Label

1 HS Calculus  
 2 High School GPA  
 3 Number precalculus correct

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
hsgpa	1	13.5684	0.0002
hscal	1	21.6572	<.0001
precalc	1	13.2413	0.0003

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-14.4886	2.2714	40.6869	<.0001
hsgpa	1	0.1147	0.0312	13.5684	0.0002
hscal	1	0.0616	0.0132	21.6572	<.0001
precalc	1	0.3098	0.0851	13.2413	0.0003

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
hsgpa	1.122	1.055	1.192
hscal	1.064	1.036	1.092
precalc	1.363	1.154	1.611

Association of Predicted Probabilities and Observed Responses

Percent Concordant	82.9	Somers' D	0.659
Percent Discordant	17.0	Gamma	0.659
Percent Tied	0.1	Tau-a	0.310
Pairs	31740	c	0.829