Chapter 6 Logistic Regression Example

Obs	ID	Split60	X4	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
1	1	0	1	8.5	3.9	2.5	5.9	4.8	4.9	6.0	6.8	4.7	4.3	5.0	5.1	3.7
2	2	0	0	8.2	2.7	5.1	7.2	3.4	7.9	3.1	5.3	5.5	4.0	3.9	4.3	4.9
3	3	0	1	9.2	3.4	5.6	5.6	5.4	7.4	5.8	4.5	6.2	4.6	5.4	4.0	4.5
4	4	0	1	6.4	3.3	7.0	3.7	4.7	4.7	4.5	8.8	7.0	3.6	4.3	4.1	3.0
5	5	0	0	9.0	3.4	5.2	4.6	2.2	6.0	4.5	6.8	6.1	4.5	4.5	3.5	3.5
6	6	0	1	6.5	2.8	3.1	4.1	4.0	4.3	3.7	8.5	5.1	9.5	3.6	4.7	3.3
7	7	0	1	6.9	3.7	5.0	2.6	2.1	2.3	5.4	8.9	4.8	2.5	2.1	4.2	2.0
8	8	0	1	6.2	3.3	3.9	4.8	4.6	3.6	5.1	6.9	5.4	4.8	4.3	6.3	3.7
9	10	0	1	6.4	4.5	5.1	6.1	4.7	5.7	5.7	8.4	5.4	5.3	4.1	5.8	4.4
10	11	0	0	8.7	3.2	4.6	4.8	2.7	6.8	4.6	6.8	5.8	7.5	3.8	3.7	4.0
11	12	0	1	6.1	4.9	6.3	3.9	4.4	3.9	6.4	8.2	5.8	5.9	3.0	4.9	3.2
12	14	0	0	9.2	3.9	5.7	5.5	2.4	8.4	4.8	7.1	6.7	3.0	4.5	2.6	4.2
13	15	0	1	6.3	4.5	4.7	6.9	4.5	6.8	5.9	8.8	6.0	5.4	4.8	6.2	5.2
14	16	0	0	8.7	3.2	4.0	6.8	3.2	7.8	3.8	4.9	6.1	5.0	4.3	3.9	4.5
15	17	0	1	5.7	4.0	6.7	6.0	3.3	5.5	5.1	6.2	6.7	5.4	4.2	6.2	4.5
16	20	0	1	9.1	4.5	3.6	6.4	5.3	5.3	7.1	8.4	5.8	6.7	4.5	6.1	4.4
17	24	0	1	9.3	2.4	2.6	7.2	2.2	7.2	4.5	6.2	6.4	4.2	6.7	4.4	4.5
18	27	0	0	8.5	3.0	7.2	5.8	4.1	7.6	3.7	4.8	6.9	6.7	5.3	3.8	4.4
19	29	0	0	8.5	3.0	5.7	6.0	2.3	7.6	3.7	4.8	5.8	6.0	5.7	3.8	4.4
20	30	0	1	7.6	3.6	3.0	4.0	5.1	4.2	4.6	7.7	4.9	7.2	4.7	5.5	3.5
21	31	0	0	6.9	3.4	8.5	4.3	4.5	6.4	4.7	5.2	7.7	3.3	3.7	2.7	3.3
22	32	0	1	8.1	2.5	7.2	4.5	2.3	5.1	3.8	6.6	6.8	6.1	3.0	3.5	3.0
23	33	0	1	6.7	3.7	6.5	5.3	5.3	5.1	4.9	9.2	5.7	4.2	3.5	4.5	3.4
24	35	0	1	6.7	4.0	5.2	3.9	3.0	5.4	6.8	8.4	6.2	6.0	2.5	4.3	3.5
25	36	0	0	8.7	3.2	6.1	4.3	3.5	6.1	2.9	5.6	6.1	6.5	3.1	2.9	2.5
26	37	0	0	9.0	3.4	5.9	4.6	3.9	6.0	4.5	6.8	6.4	4.3	3.9	3.5	3.5
27	38	0	1	9.6	4.1	6.2	7.3	2.9	7.7	5.5	7.7	6.1	4.4	5.2	4.6	4.9
28	43	0	0	9.3	5.1	4.6	6.8	5.8	6.6	6.3	7.4	5.1	4.1	4.6	4.6	4.3
29	44	0	1	5.1	5.1	6.6	6.9	4.4	5.4	7.8	5.9	7.2	5.2	4.9	6.3	4.5
30	45	0	0	8.0	2.5	4.7	7.1	3.6	7.7	3.0	5.2	5.1	3.9	4.3	4.2	4.7
31	46	0	1	5.9	4.1	5.7	5.9	5.8	6.4	5.5	8.4	6.4	5.1	5.2	5.8	4.8
32	47	0	0	10.0	4.3	7.1	6.3	2.9	5.4	4.5	3.8	6.7	3.7	5.0	4.0	3.5
33	48	0	1	5.7	3.8	6.8	7.5	5.7	5.7	6.0	8.2	6.6	4.8	6.5	7.3	5.2
34	49	0	1	9.9	3.7	3.7	6.1	4.2	7.0	6.7	6.8	5.9	7.2	4.5	3.4	3.9
35	50	0	0	7.9	3.9	4.3	5.8	4.4	6.9	5.8	4.7	5.2	3.6	4.1	4.2	4.3
36	52	0	0	8.2	2.7	3.7	7.4	2.7	7.9	3.1	5.3	5.3	5.0	4.5	4.3	4.9
37	53	0	1	9.4	2.5	4.8	6.1	3.2	7.3	4.6	6.3	6.3	9.2	4.7	4.6	4.6
38	54	0	0	6.9	3.4	5.7	4.4	3.3	6.4	4.7	5.2	6.4	4.4	3.2	2.7	3.3
39	56	0	0	9.3	3.8	7.3	5.7	3.7	6.4	5.5	7.4	6.6	5.9	4.1	3.2	3.4
40	58	0	0	7.6	3.6	5.2	5.8	5.6	6.6	5.4	4.4	6.7	6.4	4.6	3.9	4.0
41	60	0	1	9.9	2.8	7.2	6.9	2.6	5.8	3.5	5.4	6.2	7.0	5.6	4.9	4.0

42	61	0	0	8.7	3.2	8.4	6.1	2.8	7.8	3.8	4.9	7.2	4.5	5.4	3.9	4.5
43	63	0	0	8.8	3.9	3.8	5.1	4.3	4.7	4.8	5.8	5.0	7.2	4.4	3.7	2.9
44	64	0	1	7.7	2.2	6.3	4.5	2.4	4.7	3.4	6.2	6.0	4.7	3.3	3.1	2.6
45	65	0	1	6.6	3.6	5.8	4.1	4.9	4.7	4.8	7.2	6.5	3.9	3.5	3.6	2.8
46	67	0	1	5.7	4.0	7.9	6.4	2.7	5.5	5.1	6.2	7.5	6.4	5.0	6.2	4.5
47	68	0	1	5.5	3.7	4.7	5.4	4.3	5.3	4.9	6.0	5.6	2.5	4.5	5.9	4.3
48	69	0	1	7.5	3.5	3.8	3.5	2.9	4.1	4.5	7.6	5.1	5.2	4.0	5.4	3.4
49	72	0	0	6.7	3.2	3.0	3.7	4.8	6.3	4.5	5.0	5.2	2.5	2.9	2.6	3.1
50	79	0	0	9.3	3.5	6.3	7.6	5.5	7.5	5.9	4.6	6.6	3.1	5.2	4.1	4.6
51	80	0	1	7.1	3.4	4.9	4.1	4.0	5.0	5.9	7.8	6.1	3.5	2.6	3.1	2.7
52	81	0	0	9.9	3.0	7.4	4.8	4.0	5.9	4.8	4.9	5.9	6.9	3.2	4.3	3.8
53	86	0	1	7.5	3.5	4.1	4.5	3.5	4.1	4.5	7.6	4.9	2.8	3.4	5.4	3.4
54	87	0	1	5.0	3.6	1.3	3.0	3.5	4.2	4.9	8.2	4.3	7.6	2.4	4.8	3.1
55	88	0	0	7.7	2.6	8.0	6.7	3.5	7.2	4.3	5.9	6.9	7.7	5.1	3.9	4.3
56	92	0	1	7.1	4.2	4.1	2.6	2.1	3.3	4.5	9.9	5.5	3.5	2.0	4.0	2.4
57	94	0	1	9.3	3.5	5.4	7.8	4.6	7.5	5.9	4.6	6.4	4.9	4.8	4.1	4.6
58	95	0	0	9.3	3.8	4.0	4.6	4.7	6.4	5.5	7.4	5.3	4.8	3.6	3.2	3.4
59	98	0	0	8.7	3.2	3.3	3.2	3.1	6.1	2.9	5.6	5.0	4.3	3.1	2.9	2.5
60	100	0	1	7.9	3.0	4.4	5.1	5.9	4.2	4.8	9.7	5.7	5.8	3.4	5.4	3.5

Chapter 6 Logistic Regression Example

Obs	ID	Split60	X4	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
1	9	1	1	5.8	3.6	5.1	6.7	3.7	5.9	5.8	9.3	5.9	4.4	4.4	6.1	4.6
2	13	1	0	9.5	5.6	4.6	6.9	5.0	6.9	6.6	7.6	6.5	5.3	5.1	4.5	4.4
3	18	1	1	5.9	4.1	5.5	7.2	3.5	6.4	5.5	8.4	6.2	6.3	5.7	5.8	4.8
4	19	1	1	5.6	3.4	5.1	6.4	3.7	5.7	5.6	9.1	5.4	6.1	5.0	6.0	4.5
5	21	1	1	5.2	3.8	7.1	5.2	3.9	4.3	5.0	8.4	7.1	4.6	3.3	4.9	3.3
6	22	1	1	9.6	5.7	6.8	5.9	5.4	8.3	7.8	4.5	6.4	6.5	4.3	3.0	4.3
7	23	1	0	8.6	3.6	7.4	5.1	3.5	7.3	4.7	3.7	6.7	6.0	4.8	3.4	4.0
8	25	1	1	6.0	4.1	5.3	4.7	3.5	5.3	5.3	8.0	6.5	3.9	4.7	5.3	4.0
9	26	1	1	6.4	3.6	6.6	6.1	4.0	3.9	5.3	7.1	6.1	3.7	5.6	6.6	3.9
10	28	1	1	7.0	3.3	5.4	5.5	2.6	4.8	4.2	9.0	6.5	5.9	4.3	5.2	3.7
11	34	1	1	8.0	3.3	6.1	5.7	5.5	4.6	4.7	8.7	5.9	3.8	4.7	6.6	4.2
12	39	1	1	8.2	3.6	3.9	6.2	5.8	4.9	5.0	9.0	5.2	7.1	4.7	6.9	4.5
13	40	1	1	6.1	4.9	3.0	4.8	5.1	3.9	6.4	8.2	5.1	6.8	4.5	4.9	3.2
14	41	1	1	8.3	3.4	3.3	5.5	3.1	4.6	5.2	9.1	4.1	1.7	4.6	5.8	3.9
15	42	1	0	9.4	3.8	4.7	5.4	3.8	6.5	4.9	8.5	4.9	6.2	4.1	4.5	4.1
16	51	1	1	6.7	3.6	5.9	4.2	3.4	4.7	4.8	7.2	5.7	5.3	4.0	3.6	2.8
17	55	1	1	8.0	3.3	3.8	5.8	3.2	4.6	4.7	8.7	5.3	4.2	4.9	6.6	4.2
18	57	1	1	7.4	5.1	4.8	7.7	4.5	7.2	6.9	9.6	6.4	7.4	5.7	6.5	5.5
19	59	1	0	10.0	4.3	5.3	3.7	4.2	5.4	4.5	3.8	6.7	4.5	3.7	4.0	3.5
20	62	1	1	8.4	3.8	6.7	5.0	4.5	4.7	5.9	6.7	5.1	4.2	2.7	5.0	3.6
21	66	1	1	5.7	3.8	3.5	6.7	5.4	5.7	6.0	8.2	5.4	5.0	4.7	7.3	5.2

22	70	1	1	6.4	3.6	2.7	5.3	3.9	3.9	5.3	7.1	5.2	5.5	4.7	6.6	3.9
23	71	1	1	9.1	4.5	6.1	5.9	6.3	5.3	7.1	8.4	7.1	5.7	5.4	6.1	4.4
24	73	1	1	6.5	4.3	2.7	6.6	6.5	6.3	6.0	8.7	4.7	6.3	4.6	5.6	4.6
25	74	1	1	9.9	3.7	7.5	4.7	5.6	7.0	6.7	6.8	7.2	4.6	4.1	3.4	3.9
26	75	1	1	8.5	3.9	5.3	5.5	5.0	4.9	6.0	6.8	5.7	3.6	4.4	5.1	3.7
27	76	1	0	9.9	3.0	6.8	5.0	5.4	5.9	4.8	4.9	7.3	7.6	3.1	4.3	3.8
28	77	1	1	7.6	3.6	7.6	4.6	4.7	4.6	5.0	7.4	8.1	6.6	4.5	5.8	3.9
29	78	1	0	9.4	3.8	7.0	6.2	4.7	6.5	4.9	8.5	7.3	2.4	4.3	4.5	4.1
30	82	1	0	8.7	3.2	6.4	4.9	2.4	6.8	4.6	6.8	6.3	5.1	4.3	3.7	4.0
31	83	1	0	8.6	2.9	5.8	3.9	2.9	5.6	4.0	6.3	6.1	4.0	2.7	3.0	3.0
32	84	1	1	6.4	3.2	6.7	3.6	2.2	2.9	5.0	8.4	7.3	6.5	2.0	3.7	1.6
33	85	1	0	7.7	2.6	6.7	6.6	1.9	7.2	4.3	5.9	6.5	4.1	4.7	3.9	4.3
34	89	1	0	9.1	3.6	5.5	5.4	4.2	6.2	4.6	8.3	6.5	4.1	4.6	4.3	3.9
35	90	1	1	5.5	5.5	7.7	7.0	5.6	5.7	8.2	6.3	7.4	4.9	5.5	6.7	4.9
36	91	1	0	9.1	3.7	7.0	4.1	4.4	6.3	5.4	7.3	7.5	4.6	4.4	3.0	3.3
37	93	1	0	9.2	3.9	4.6	5.3	4.2	8.4	4.8	7.1	6.2	6.6	4.4	2.6	4.2
38	96	1	0	8.6	4.8	5.6	5.3	2.3	6.0	5.7	6.7	5.8	3.6	4.9	3.6	3.6
39	97	1	1	7.4	3.4	2.6	5.0	4.1	4.4	4.8	7.2	4.5	6.4	4.2	5.6	3.7
40	99	1	1	7.8	4.9	5.8	5.3	5.2	5.3	7.1	7.9	6.0	5.7	4.3	4.9	3.9

Chapter 6 Logistic Regression Example

The LOGISTIC Procedure

Model Information		
Data Set	WORK.HBAT60	
Response Variable	X4	X4 - Region
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Note: Most typical Regression Diagnostics, e.g., Proc REG and GLM are also availble in Proc Logistic.

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	X4	Total Frequency
1	0	26
2	1	34

Specifies the response value

Probability modeled is X4=0.

Stepwise Selection Procedure

Step 0. Intercept entered:

SAS Output

Logistic regression measures model estimation fit with the value of -2 times the log of the likelihood value, referred to as -2LL or -2 log likelihood. The minimum value for -2LL is 0, which corresponds to a perfect fit (likelihood = 1 and -2LL is then 0). Thus, the lower the -2LL value, the better the fit of the model. The -2LL value can be used to compare equations for the change in fit or to calculate measures comparable to the R2 measure in multiple regression.

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

-2 Log L	=	82.108

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.2683	0.2605	1.0603	0.3031

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
42.3497	13	<.0001

Wald statistic Test used in logistic regression for the significance of the logistic coefficient. Its interpretation is like the F or t values used for the significance testing of regression coefficients.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X6	1	11.9251	0.0006
X7	1	2.0517	0.1520
X8	1	1.6089	0.2046
X9	1	0.8656	0.3522
X10	1	0.7914	0.3737
X11	1	18.3231	<.0001
X12	1	8.6217	0.0033
X13	1	21.3297	<.0001
X14	1	0.4654	0.4951
X15	1	0.6138	0.4333
X16	1	0.0899	0.7644
X17	1	21.2035	<.0001
X18	1	0.1567	0.6922

Step 1. Effect X13 entered:

The “Model Fit Statistics” table contains the Akaike information criterion (AIC), the Schwarz Criterion (SC), and the negative of twice the log likelihood (–2 Log L) for the intercept only model and the fitted model. AIC and SC can be used to compare different models, and the ones with smaller values are preferred.

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

The Cox and Snell R2 measure operates in the same manner, with higher values indicating greater model fit. However, this measure is limited in that it cannot reach the maximum value of 1, so Nagelkerke proposed a modification that had the range of 0 to 1. Both of these additional measures are interpreted as reflecting the amount of variation accounted for by the logistic model, with 1.0 indicating perfect model fit.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	84.108	60.971
SC	86.202	65.160
-2 Log L	82.108	56.971

← Useful for comparing between models

R-Square	0.3423	Max-rescaled R-Square	0.4591

← Nagelkerke R-square

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	25.1363	1	<.0001

Test that all coefficients are equal to zero. Significant values indicate at least one coefficient not equal to zero. Comparable to overall model fit F-test in

Score	21.3297	1	<.0001
Wald	15.4710	1	<.0001

Original logistic coefficients centered around 0. Sign indicates direction of relationship

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	7.0082	1.8360	14.5698	0.0001
X13	1	-1.1287	0.2870	15.4710	<.0001

Exponentiated coefficients which are centered around 1.0

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
X13	0.323	0.184	0.568

Confidence interval of significant variable does not include 1.0

Concordant = case with actual value of 1.0 has higher predicted value than case with actual value of 0

Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.5	Somers' D	0.699
Percent Discordant	14.6	Gamma	0.705
Percent Tied	0.9	Tau-a	0.349
Pairs	884	c	0.850

Higher values indicate better fit

close to generalized R-square

Equal to area under ROC curve

All pairs of cases that do not match (1-0 or 0-1)

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
27.3151	12	0.0070

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
X13	1	15.4710	<.0001

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

Statistical Measures. The first statistical measure is the chi-square test for the change in the -2LL value from the base model, which is comparable to the overall F test in multiple regression. Smaller values of the -2LL measure indicate better model fit, and the statistical test is available for assessing the difference between both the base model and other proposed models (in a stepwise procedure, this test is always based on improvement from the prior step).

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
X6	1	4.8588	0.0275
X7	1	0.1321	0.7163
X8	1	0.0074	0.9315
X9	1	1.3794	0.2402
X10	1	0.1293	0.7192
X11	1	6.1543	0.0131
X12	1	2.7452	0.0975
X14	1	0.6395	0.4239
X15	1	0.3442	0.5574
X16	1	2.5284	0.1118
X17	1	13.7231	0.0002
X18	1	1.2061	0.2721

Step 2. Effect X17 entered:

Criterion - Underneath are various measurements used to assess the model fit. The first two, Akaike Information Criterion (AIC) and Schwarz Criterion (SC) are deviants of negative two times the Log-Likelihood (-2 Log L). AIC and SC penalize the log-likelihood by the number of predictors in the model.

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

Model converges and -2 Log L (analogous to Global F-test in PROC REG) shows that overall model is significant (see Residual Chi-square test p< .0001).

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	84.108	45.960
SC	86.202	52.243
-2 Log L	82.108	39.960

R-Square	0.5046	Max-rescaled R-Square	0.6769

AIC - This is the Akaike Information Criterion. It is calculated as AIC = -2 Log L + 2((k-1) + s), where k is the number of levels of the dependent variable and s is the number of predictors in the model. AIC is used for the comparison of nonnested models on the same sample. Ultimately, the model with the smallest AIC is considered the best, although the AIC value itself is not meaningful.

SC - This is the Schwarz Criterion. It is defined as -2 Log L + ((k-1) + s)*log(Σ fi), where fi's are the frequency values of the ith observation, and k and s were defined previously. Like AIC, SC penalizes for the number of predictors in the model and the smallest SC is most desirable and the value itself is not meaningful..

Test - These are three asymptotically equivalent Chi-Square tests. They test against the null hypothesis that at least one of the predictors' regression coefficient is not equal to zero in the model. The difference between them are where on the log-likelihood function they are evaluated.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1477	2	<.0001
Score	31.3228	2	<.0001
Wald	14.1772	2	0.0008

Likelihood Ratio - This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero in the model. The LR Chi-Square statistic can be calculated by -2 Log L(null model) - 2 Log L(fitted model) = , where L(null model) refers to the Intercept Only model and L(fitted model) refers to the Intercept and Covariates model.

Score - This is the Score Chi-Square Test that at least one of the predictors' regression coefficient is not equal to zero in the model.

Wald - This is the Wald Chi-Square Test that at least one of the predictors' regression coefficient is not equal to zero in the model.

-2 Log L - This is negative two times the log-likelihood. The -2 Log L is used in hypothesis tests for nested models and the value in itself is not meaningful.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	14.1917	3.7123	14.6143	0.0001
X13	1	-1.0791	0.3574	9.1148	0.0025
X17	1	-1.8439	0.6388	8.3314	0.0039

The **"Analysis of Maximum Likelihood Estimates"** table summarizes information regarding the independent variables including parameter estimates, variability, and significance.

Intercept Only - This column refers to the respective criterion statistics with no predictors in the model, i.e., just the response variable.

Intercept and Covariates - This column corresponds to the respective criterion statistics for the fitted model. A fitted model includes all independent variables and the intercept. We can compare the values in this column with the criteria corresponding Intercept Only value to assess model fit/significance.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X13	0.340	0.169	0.685
X17	0.158	0.045	0.553

The **"Odds Ratio Estimates"** table summarizes the significant independent variables and indicates their associated odds ratios and confidence limits.

Effect - the predictor variables that are interpreted in terms of odds ratios.

Point Estimate - the odds ratio corresponding to Effect. The odds ratio is obtained by exponentiating the Estimate, exp[Estimate]. The difference in the log of two odds is equal to the log of the ratio of these two odds. The log of the ratio of two odds is the log odds ratio. Hence, the interpretation of Estimate--the coefficient was interpreted as the difference in log-odds--could also be done in terms of log-odds ratio. When the Estimate is exponentiated, the log-odds ratio becomes the odds ratio. We can interpret the odds ratio as follows: for a one unit change in the predictor variable, the odds ratio for a positive outcome is expected to change by the respective coefficient, given the other variables in the model are held constant.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	92.1	Somers' D	0.843
Percent Discordant	7.8	Gamma	0.844
Percent Tied	0.1	Tau-a	0.421
Pairs	884	c	0.921

Percent Concordant - A pair of observations with different observed responses is said to be concordant if the observation with the lower ordered response value (= 0) has a lower predicted mean score than the observation with the higher ordered response value (= 1).

Percent Discordant - If the observation with the lower ordered response value has a higher predicted mean score than the observation with the higher ordered response value, then the pair is discordant.

Percent Tied - If a pair of observations with different responses is neither concordant nor discordant, it is a tie.

Pairs - This is the total number of distinct pairs in which one case has an observed outcome different from the other member of the pair.

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
20.2161	11	0.0425

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X13	1	9.1148	0.0025
X17	1	8.3314	0.0039

95% Wald Confidence Limits - This is the Wald Confidence Interval (CI) of an individual odds ratio, given the other predictors are in the model. For a given predictor variable with a level of 95% confidence, we'd say that we are 95% confident that upon repeated trials, 95% of the CI's would include the "true" population odds ratio. The CI is equivalent to the Chi-Square test statistic: if the CI includes one, we'd fail to reject the null hypothesis that a particular regression coefficient equals zero and the odds ratio equals one, given the other predictors are in the model. An advantage of a CI is that it is illustrative; it provides information on where the "true" parameter may lie and the precision of the point estimate for the odds ratio.

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

Somers' D - Somer's D is used to determine the strength and direction of relation between pairs of variables. Its values range from -1.0 (all pairs disagree) to 1.0 (all pairs agree). It is defined as (nc-nd)/t where nc is the number of pairs that are concordant, nd the number of pairs that are discordant, and t is the number of total number of pairs with different responses. In our example, it equals the difference between the percent concordant and the percent discordant divided by 100

Gamma - The Goodman-Kruskal Gamma method does not penalize for ties on either variable. Its values range from -1.0 (no association) to 1.0 (perfect association). Because it does not penalize for ties, its value will generally be greater than the values for Somer's D.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
X6	1	0.6561	0.4179
X7	1	3.5008	0.0613
X8	1	0.0063	0.9369
X9	1	0.6926	0.4053
X10	1	0.0914	0.7624
X11	1	3.4094	0.0648
X12	1	0.8492	0.3568
X14	1	2.3269	0.1272
X15	1	0.0257	0.8727
X16	1	0.0103	0.9192
X18	1	2.9074	0.0882

Tau-a - Kendall's Tau-a is a modification of Somer's D that takes into the account the difference between the number of possible paired observations and the number of paired observations with a different response. It is defined to be the ratio of the difference between the number of concordant pairs and the number of discordant pairs to the number of possible pairs (2(nc-nd)/(N(N-1)). Usually Tau-a is much smaller than Somer's D since there would be many paired observations with the same response.

c - is equivalent to the well known measure ROC. c ranges from 0.5 to 1, where 0.5 corresponds to the model randomly predicting the response, and a 1 corresponds to the model perfectly discriminating the response.

A Receiver Operating Characteristic Curve (ROC) is a standard technique for summarizing classifier performance over a range of trade-offs between true positive (TP) and false positive (FP) error rates.

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

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	X13		1	1	21.3297		<.0001	X13 - Competitive Pricing
2	X17		1	2	13.7231		0.0002	X17 - Price Flexibility

Sample ranked by predicted probability and the divided into 10 equally-sized groups(in this data that is 10 cases per group). Group 1 has lowest predicted probability, Group 10 the highest

Partition for the Hosmer and Lemeshow Test

Group	Total	X4 = 0 Observed	X4 = 0 Expected	X4 = 1 Observed	X4 = 1 Expected
1	6	0	0.01	6	5.99
2	7	0	0.11	7	6.89
3	6	0	0.16	6	5.84
4	6	1	0.39	5	5.61
5	6	1	2.41	5	3.59
6	6	6	3.61	0	2.39
7	6	4	4.12	2	1.88
8	6	5	5.03	1	0.97
9	6	4	5.34	2	0.66
10	5	5	4.81	0	0.19

Chi-Square-Based Measure. Hosmer and Lemeshow developed a classification test where the cases are first divided into approximately 10 equal classes. Then, the number of actual and predicted events is compared in each class with the chi-square statistic. This test provides a comprehensive measure of predictive accuracy that is based not on the likelihood value, but rather on the actual prediction of the dependent variable.

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
9.9225	8	0.2705

The "Hosmer and Lemeshow Goodness of Fit Test" indicates the quality of model fit. If the associated p-value is significant (p<0.05) this would be an indication that we need to rethink our analytic strategy.

SAS Output

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	26	0	34	0	43.3	100.0	0.0	56.7	.
0.100	25	24	10	1	81.7	96.2	70.6	28.6	4.0
0.200	25	24	10	1	81.7	96.2	70.6	28.6	4.0
0.300	25	25	9	1	83.3	96.2	73.5	26.5	3.8
0.400	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.500	24	28	6	2	86.7	92.3	82.4	20.0	6.7
0.600	20	28	6	6	80.0	76.9	82.4	23.1	17.6
0.700	16	31	3	10	78.3	61.5	91.2	15.8	24.4
0.800	13	31	3	13	73.3	50.0	91.2	18.8	29.5
0.900	6	33	1	20	65.0	23.1	97.1	14.3	37.7
1.000	0	34	0	26	56.7	0.0	100.0	.	43.3

1 - Positive Predictive Value

1 - Negative Predictive Value

Chapter 6 Logistic Regression Example

The LOGISTIC Procedure

Model Information		
Data Set	WORK.HBAT60	
Response Variable	X4	X4 - Region
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	60
Number of Observations Used	60

Response Profile		
Ordered Value	X4	Total Frequency
1	0	26
2	1	34

Probability modeled is X4=0.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	84.108	45.960
SC	86.202	52.243
-2 Log L	82.108	39.960

SAS Output

R-Square	0.5046	Max-rescaled R-Square	0.6769

Parameter - Underneath are the predictor variables in the model and the intercept.
DF - This column gives the degrees of freedom corresponding to the Parameter. Each Parameter estimated in the model requires one DF and defines the Chi-Square distribution to test whether the individual regression coefficient is zero, given the other variables are in the model.
Estimate - These are the binary logit regression estimates for the Parameters in the model. The logistic regression model models the log odds of a positive response (probability modeled is honcomp=1) as a linear combination the predictor variables. **We** can interpret the parameter estimates as follows: for a one unit change in the predictor variable, the difference in log-odds for a positive outcome is expected to change by the respective coefficient, given the other variables in the model are held constant.
Intercept - This is the logistic regression estimate when all variables in the model are evaluated at zero.

Standard Error - These are the standard errors of the individual regression coefficients. They are used in both the 95% Wald Confidence Limits and the Chi-Square test statistic.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1477	2	<.0001
Score	31.3228	2	<.0001
Wald	14.1772	2	0.0008

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	14.1917	3.7123	14.6143	0.0001
X13	1	-1.0791	0.3574	9.1148	0.0025
X17	1	-1.8439	0.6388	8.3314	0.0039

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
X13	0.340	0.169	0.685
X17	0.158	0.045	0.553

Association of Predicted Probabilities and Observed Responses

Percent Concordant	92.1	Somers' D	0.843
Percent Discordant	7.8	Gamma	0.844
Percent Tied	0.1	Tau-a	0.421
Pairs	884	c	0.921

Partition for the Hosmer and Lemeshow Test

Group	Total	X4 = 0		X4 = 1	
		Observed	Expected	Observed	Expected
1	6	0	0.01	6	5.99
2	7	0	0.11	7	6.89
3	6	0	0.16	6	5.84
4	6	1	0.39	5	5.61
5	6	1	2.41	5	3.59
6	6	6	3.61	0	2.39
7	6	4	4.12	2	1.88
8	6	5	5.03	1	0.97
9	6	4	5.34	2	0.66
10	5	5	4.81	0	0.19

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
9.9225	8	0.2705

Classification Table

	Correct		Incorrect		Percentages				

Chi-Square, DF and Pr > ChiSq - These are the Chi-Square test statistic, Degrees of Freedom (DF) and associated p-value (PR>ChiSq) corresponding to the specific test that all of the predictors are simultaneously equal to zero. We are testing the probability (PR>ChiSq) of observing a Chi-Square statistic as extreme as, or more so, than the observed one under the null hypothesis; the null hypothesis is that all of the regression coefficients in the model are equal to zero.

The **DF** defines the distribution of the Chi-Square test statistics and is defined by the number of predictors in the model.

Typically, **PR>ChiSq** is compared to a specified alpha level, our willingness to accept a type I error, which is often set at 0.05 or 0.01. The small p-value from the all three tests would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero.

Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.400	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.410	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.420	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.430	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.440	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.450	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.460	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.470	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.480	25	27	7	1	86.7	96.2	79.4	21.9	3.6
0.490	24	28	6	2	86.7	92.3	82.4	20.0	6.7
0.500	24	28	6	2	86.7	92.3	82.4	20.0	6.7
0.510	24	28	6	2	86.7	92.3	82.4	20.0	6.7
0.520	24	28	6	2	86.7	92.3	82.4	20.0	6.7
0.530	22	28	6	4	83.3	84.6	82.4	21.4	12.5
0.540	22	28	6	4	83.3	84.6	82.4	21.4	12.5
0.550	22	28	6	4	83.3	84.6	82.4	21.4	12.5
0.560	22	28	6	4	83.3	84.6	82.4	21.4	12.5
0.570	22	28	6	4	83.3	84.6	82.4	21.4	12.5
0.580	20	28	6	6	80.0	76.9	82.4	23.1	17.6
0.590	20	28	6	6	80.0	76.9	82.4	23.1	17.6
0.600	20	28	6	6	80.0	76.9	82.4	23.1	17.6

Chapter 6 Logistic Regression Example

Obs	X4	X13	X17	F_X4	I_X4	P_0	P_1
1	1	6.8	5.1	1	1	0.07241	0.92759
2	0	5.3	4.3	0	0	0.63264	0.36736
3	1	4.5	4.0	1	0	0.87653	0.12347
4	1	8.8	4.1	1	1	0.05393	0.94607
5	0	6.8	3.5	0	0	0.59870	0.40130
6	1	8.5	4.7	1	1	0.02540	0.97460
7	1	8.9	4.2	1	1	0.04082	0.95918
8	1	6.9	6.3	1	1	0.00761	0.99239
9	1	8.4	5.8	1	1	0.00381	0.99619
10	0	6.8	3.7	0	0	0.50781	0.49219
11	1	8.2	4.9	1	1	0.02431	0.97569
12	0	7.1	2.6	0	0	0.85016	0.14984
13	1	8.8	6.2	1	1	0.00119	0.99881
14	0	4.9	3.9	0	0	0.84719	0.15281
15	1	6.2	6.2	1	1	0.01924	0.98076
16	1	8.4	6.1	1	1	0.00219	0.99781
17	1	6.2	4.4	1	1	0.35159	0.64841

SAS Output

18	0	4.8	3.8	0	0	0.88133	0.11867
19	0	4.8	3.8	0	0	0.88133	0.11867
20	1	7.7	5.5	1	1	0.01394	0.98606
21	0	5.2	2.7	0	0	0.97345	0.02655
22	1	6.6	3.5	1	0	0.64928	0.35072
23	1	9.2	4.5	1	1	0.01740	0.98260
24	1	8.4	4.3	1	1	0.05723	0.94277
25	0	5.6	2.9	0	0	0.94275	0.05725
26	0	6.8	3.5	0	0	0.59870	0.40130
27	1	7.7	4.6	1	1	0.06917	0.93083
28	0	7.4	4.6	0	1	0.09315	0.90685
29	1	5.9	6.3	1	1	0.02206	0.97794
30	0	5.2	4.2	0	0	0.69759	0.30241
31	1	8.4	5.8	1	1	0.00381	0.99619
32	0	3.8	4.0	0	0	0.93793	0.06207
33	1	8.2	7.3	1	1	0.00030	0.99970
34	1	6.8	3.4	1	0	0.64209	0.35791
35	0	4.7	4.2	0	0	0.79825	0.20175
36	0	5.3	4.3	0	0	0.63264	0.36736
37	1	6.3	4.6	1	1	0.25186	0.74814
38	0	5.2	2.7	0	0	0.97345	0.02655
39	0	7.4	3.2	0	0	0.57585	0.42415
40	0	4.4	3.9	0	0	0.90485	0.09515
41	1	5.4	4.9	1	1	0.33833	0.66167
42	0	4.9	3.9	0	0	0.84719	0.15281
43	0	5.8	3.7	0	0	0.75219	0.24781
44	1	6.2	3.1	1	0	0.85632	0.14368
45	1	7.2	3.6	1	1	0.44621	0.55379
46	1	6.2	6.2	1	1	0.01924	0.98076
47	1	6.0	5.9	1	1	0.04062	0.95938
48	1	7.6	5.4	1	1	0.01858	0.98142
49	0	5.0	2.6	0	0	0.98205	0.01795
50	0	4.6	4.1	0	0	0.84127	0.15873
51	1	7.8	3.1	1	0	0.51462	0.48538
52	0	4.9	4.3	0	0	0.72615	0.27385
53	1	7.6	5.4	1	1	0.01858	0.98142
54	1	8.2	4.8	1	1	0.02909	0.97091
55	0	5.9	3.9	0	0	0.65332	0.34668
56	1	9.9	4.0	1	1	0.02049	0.97951
57	1	4.6	4.1	1	0	0.84127	0.15873
58	0	7.4	3.2	0	0	0.57585	0.42415
59	0	5.6	2.9	0	0	0.94275	0.05725
60	1	9.7	5.4	1	1	0.00196	0.99804

Chapter 6 Logistic Regression Example

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of F_X4 by I_X4		
	I_X4(Into: X4)		
F_X4(From: X4)	0	1	Total
0	25	1	26
	41.67	1.67	43.33
	96.15	3.85	
	80.65	3.45	
1	6	28	34
	10.00	46.67	56.67
	17.65	82.35	
	19.35	96.55	
Total	31	29	60
	51.67	48.33	100.00

Original Data Classifications Crosstabulations

Chapter 6 Logistic Regression Example

Obs	X4	X13	X17	F_X4	I_X4	P_0	P_1
1	1	9.3	6.1	1	1	0.00083	0.99917
2	0	7.6	4.5	0	1	0.09053	0.90947
3	1	8.4	5.8	1	1	0.00381	0.99619
4	1	9.1	6.0	1	1	0.00124	0.99876
5	1	8.4	4.9	1	1	0.01968	0.98032
6	1	4.5	3.0	1	0	0.97820	0.02180
7	0	3.7	3.4	0	0	0.98073	0.01927
8	1	8.0	5.3	1	1	0.01457	0.98543
9	1	7.1	6.6	1	1	0.00354	0.99646
10	1	9.0	5.2	1	1	0.00601	0.99399
11	1	8.7	6.6	1	1	0.00063	0.99937
12	1	9.0	6.9	1	1	0.00026	0.99974
13	1	8.2	4.9	1	1	0.02431	0.97569
14	1	9.1	5.8	1	1	0.00179	0.99821
15	0	8.5	4.5	0	1	0.03632	0.96368
16	1	7.2	3.6	1	1	0.44621	0.55379
17	1	8.7	6.6	1	1	0.00063	0.99937
18	1	9.6	6.5	1	1	0.00029	0.99971
19	0	3.8	4.0	0	0	0.93793	0.06207
20	1	6.7	5.0	1	1	0.09467	0.90533
21	1	8.2	7.3	1	1	0.00030	0.99970
22	1	7.1	6.6	1	1	0.00354	0.99646
23	1	8.4	6.1	1	1	0.00219	0.99781
24	1	8.7	5.6	1	1	0.00398	0.99602
25	1	6.8	3.4	1	0	0.64209	0.35791

26	1	6.8	5.1	1	1	0.07241	0.92759
27	0	4.9	4.3	0	0	0.72615	0.27385
28	1	7.4	5.8	1	1	0.01111	0.98889
29	0	8.5	4.5	0	1	0.03632	0.96368
30	0	6.8	3.7	0	0	0.50781	0.49219
31	0	6.3	3.0	0	0	0.86548	0.13452
32	1	8.4	3.7	1	1	0.15508	0.84492
33	0	5.9	3.9	0	0	0.65332	0.34668
34	0	8.3	4.3	0	1	0.06334	0.93666
35	1	6.3	6.7	1	1	0.00696	0.99304
36	0	7.3	3.0	0	0	0.68621	0.31379
37	0	7.1	2.6	0	0	0.85016	0.14984
38	0	6.7	3.6	0	0	0.58019	0.41981
39	1	7.2	5.6	1	1	0.01977	0.98023
40	1	7.9	4.9	1	1	0.03329	0.96671

Chapter 6 Logistic Regression Example

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of F_X4 by I_X4

F_X4(From: X4)	I_X4(Into: X4) 0	1	Total
0	9	4	13
	22.50	10.00	32.50
	69.23	30.77	
	81.82	13.79	
1	2	25	27
	5.00	62.50	67.50
	7.41	92.59	
	18.18	86.21	
Total	11	29	40
	27.50	72.50	100.00

Holdout Sample Classifications Crosstabulations