Logistic Regression

Logistic regression is a variation of the regression model. It is used when the dependent response variable is binary in nature. Logistic regression predicts the probability of the dependent response, rather than the value of the response (as in simple linear regression).

In this example, the dependent variable is frequency of sex (less than once per month versus more than once per month).

```
PROC LOGISTIC DESCENDING;
MODEL freqdum = age racenew happy church male married/EXPB;
RUN;
```

SAS chooses the smaller value to estimate its probability. If you include the “descending” option, then SAS will estimate the larger value. If you choose not to include the “descending” option, you will get the same results, except that each B will need to be multipled by negative 1 (-1) and the odds ratios inverted.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1438.921	1146.812
SC	1443.879	1176.563
-2 Log L**a**	1436.921	1134.812

a. **“-2 Log L”** is used in hypothesis testing for nested models. It is negative two times the log likelihood. Tables typically report the “intercept and covariates” value.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	302.1090	5	<.0001
Score	269.5576	5	<.0001
Wald	203.1289	5	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate**b c**	Standard Error	Chi-Square	Pr > ChiSq	Exp(Est)**d**
Intercept	1	6.7747	0.5080	177.8151	<.0001	875.397
AGE	1	-0.0609	0.00501	148.1741	<.0001	0.941
MARITAL	1	-1.7432	0.1661	110.0822	<.0001	0.175
RACENEW	1	-0.1217	0.1767	0.4745	0.4909	0.885
ATTEND	1	-0.0715	0.0282	6.4130	0.0113	0.931
HAPPY	1	-0.3346	0.1218	7.5451	0.0060	0.716

b. “Estimate” is the estimated coefficient, with the standard error.

c. For this example, the regression equation for the final model is:

$$SEXFREQ_{PREDICTED} = 6.575 - .061*age - 1.738*marital + .069*race - .070*attend - .334*happiness$$

d. Recall: When Exp(Est) is less than 1, increasing values of the variable correspond to decreasing odds of the event's occurrence. When Exp(Est) is greater than 1, increasing values of the variable correspond to increasing odds of the event's occurrence.

Intercept = Not interpretable in logistic regression.

Age = Increasing values of age correspond with decreasing odds of having sex more than once a month.

Marital = Increasing values of marital status (married to unmarried) correspond with decreasing odds of having sex more than once a month.

Race = Increasing values of race correspond with increasing odds of having sex more than once a month. Notice that this variable, however, is not significant.

Church Attendance = Increasing values of church attendance correspond with decreasing odds of having sex more than once a month.

Happiness = Increasing values of general happiness correspond with decreasing odds of having sex more than once a month. Recall that happiness is coded such that higher values indicate less happiness.

Odds Ratio Estimates**e**			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE	0.941**f**	0.932	0.950
MARITAL	0.175	0.126	0.242
RACENEW	0.885	0.626	1.252
ATTEND	0.931	0.881	0.984
HAPPY	0.716	0.564	0.909

e. “Exp(Est),” or the odds ratio, is the predicted change in odds for a unit increase in the predictor. When Exp(Est) is less than 1, increasing values of the variable correspond to decreasing odds of the event's occurrence. When Exp(B) is greater than 1, increasing values of the variable correspond to increasing odds of the event's occurrence.

f. If you subtract 1 from the odds ratio and multiply by 100, you get the percent change in odds of the dependent variable having a value of 1. For example, for age:

= 1 - (.941) = .051

= .051 * 100 = 5.1%

The odds ratio for age indicates that every unit increase in age is associated with a 5.1% decrease in the odds of having sex more than once a month.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.9	Somers' D	0.579
Percent Discordant	20.9	Gamma	0.580
Percent Tied	0.2	Tau-a	0.284
Pairs	271051	c	0.790

Logistic Regression (with non-linear variable)

```
PROC LOGISTIC DESCENDING;
MODEL freqdum = age marital racenew happy attend agesquar/EXPB;
RUN;
```

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	4.8068	0.7552	40.5113	<.0001	122.337
AGE	1	0.0289	0.0266	1.1772	0.2779	1.029
MARITAL	1	-1.6731	0.1705	96.3398	<.0001	0.188
RACENEW	1	-0.1216	0.1762	0.4768	0.4899	0.885
ATTEND	1	-0.0713	0.0284	6.2984	0.0121	0.931
HAPPY	1	-0.3529	0.1219	8.3801	0.0038	0.703
AGESQUAR	1	-0.00094	0.000280	11.4037	0.0007	0.999

The age squared variable is significant, indicating that age is non-linear.

Logistic Regression (with interaction term)

To test for two-way interactions (often thought of as a relationship between an independent variable (IV) and dependent variable (DV), moderated by a third variable), first run a regression analysis, including both independent variables (IV and moderator) and their interaction (product) term. It is highly recommended that the independent variable and moderator are standardized before calculation of the product term, although this is not essential. For this example, two dummy variables were created, for ease of interpretation. Sex was recoded such that 1=Male and 0=Female. Marital status was recoded such that 1=Currently married and 0=Not currently married. The interaction term is a product of these two dummy variables.

Regression Model (without interactions)

```
PROC LOGISTIC DESCENDING;
MODEL freqdum = age racenew happy church male married /EXPB;
RUN;
```

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	3.0471	0.3816	63.7725	<.0001	21.054
AGE	1	-0.0611	0.00506	145.7477	<.0001	0.941
RACENEW	1	-0.1490	0.1782	0.6989	0.4032	0.862
HAPPY	1	-0.3184	0.1228	6.7229	0.0095	0.727
ATTEND	1	-0.0595	0.0286	4.3146	0.0378	0.942
MALE	1	0.4436	0.1483	8.9514	0.0028	1.558
MARRIED	1	1.6983	0.1672	103.1272	<.0001	5.465

Regression Model (with interactions)

```
PROC LOGISTIC DESCENDING;
MODEL freqdum = age racenew happy church male married interact /EXPB;
RUN;
```

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	2.9288	0.3874	57.1643	<.0001	18.704
AGE	1	-0.0601	0.00509	139.1866	<.0001	0.942
RACENEW	1	-0.1729	0.1793	0.9297	0.3349	0.841
HAPPY	1	-0.3223	0.1233	6.8378	0.0089	0.724
ATTEND	1	-0.0563	0.0287	3.8375	0.0501	0.945
MALE	1	0.6488	0.1933	11.2615	0.0008	1.913
MARRIED	1	1.9360	0.2225	75.7215	<.0001	6.931
MALE*MARRIED	1	-0.5036	0.3024	2.7739	0.0958	0.604

The product term should be significant in the regression equation in order for the interaction to be interpretable. In this example, the interaction term is significant at the 0.1 level.

Regression Equation

$FREQDUM_{predicted}$ = 2.93 -.06*age - .17*White -.32*happy - .06*attend + 1.94*married + (.65 - .50*married) * male

Interpretation

Main Effects

The married coefficient represents the main effect for females (the 0 category). The effect for females is then 1.94, or the "marital" coefficient. The effect for males is 1.94 - .50, or 1.44.

The gender coefficient represents the main effect for unmarried persons (the 0 category). The effect for unmarried is then .65, or the "sex" coefficient. The effect for married is .65 - .50, or .15.

Interaction Effects

For a simple interpretation of the interaction term, plug values into the regression equation above.

Married Men =	$FREQDUM_{predicted}$ = 2.93 -.06*age - .17*White -.32*happy - .06*attend + 1.94**1* + (.65 - .50**1*) * *1*	= 2.43
Married Women =	$FREQDUM_{predicted}$ = 2.93 -.06*age - .17*White -.32*happy - .06*attend + 1.94**1* + (.65 - .50**1*) * *0*	= 2.28
Unmarried Men =	$FREQDUM_{predicted}$ = 2.93 -.06*age - .17*White -.32*happy - .06*attend + 1.94**0* + (.65 - .50**0*) * *1*	= 0.49
Unmarried Women =	$FREQDUM_{predicted}$ = 2.93 -.06*age - .17*White -.32*happy - .06*attend + 1.94**0* + (.65 - .50**0*) * *0*	= 0.34

In this example (age = 35 years; race = white-1; happiness = very happy-1; church attendance = never attends-0), we can see that (1) for both married and unmarried persons, males are reporting higher frequency of sex than females, and (2) married persons report higher frequency of sex than unmarried persons. The interaction tells us that the gender difference is greater for married persons than for unmarried persons.

Odds Ratios

Using "married" as the focus variable, we can say that the effect of being married on having sex more than once per month is greater for females.

Females: $e^{1.936} = 6.93$
Males: $e^{1.432} = 4.20$

Using "gender" as the focus variable, we can say that the effect of being male on having sex more than once per month is greater for marrieds.

Marrieds: $e^{0.15} = 1.16$
Unmarrieds: $e^{0.65} = 1.92$