Categorical Data Analysis Using SAS and Stata

Hsueh-Sheng Wu
Center for Family and Demographic Research

May 23, 2016

Outline

- Why do we need to learn categorical data analyses?
- A summary of different categorical data analyses
  - Analyses of contingency tables
  - Regression models
    - Logistic regression
    - Ordered logistic regression
    - Multinomial logistic regression
- Stata commands
- SAS commands
- Interpreting the results
- Predicted probability
- Conclusions

Why Do We Need to Learn Categorical Data Analysis?

- Four measurement levels
  - Nominal (e.g., gender, race)
  - Ordinal (e.g., attitude toward cohabitation)
  - Interval (e.g., temperature)
  - Ratio (e.g., income)
- Categorical variables are those measured at nominal and ordinal levels.
- Interval or ratio variables can be transformed into nominal or ordinal variables, but not the other way around.

BGSU
Center for Family and Demographic Research

What Is Special about Categorical Variable?

- The distribution of a categorical variable is described by its frequency and proportion rather than by its mean and variance.
- Statistical methods (i.e., t-test, correlation, OLS regression) designed for continuous dependent variables are not adequate for analyzing categorical dependent variables.
- The decision on how to analyze categorical variables is often based on:
  - The measurement level and number of categories in dependent variables
  - The measurement level and number of categories in independent variables
  - Sample size
  - Number of independent variables

BGSU
Center for Family and Demographic Research

Different Models for Categorical Dependent Variables

Categorical models address three types of questions:

- Examination of contingency tables
  - Proportions
  - Relative risks
  - Odds ratio
- How the characteristics of individuals affect the choice
  - Binary logistic regression
  - Ordered logistic regression
  - Multinomial logistic regression

Analyzing a Two-way Contingency Table

- Analyzing a 2x2 table

Table. Gender and Employment

	Employed	Unemployed
Male	200	200
Female	200	400

Table. Gender and Employment

	Employed	Unemployed
Male	ρ_1	$1-\rho_1$
Female	ρ_2	$1-\rho_2$

$$\text{Difference of Two Proportions} = \pi_1 - \pi_2 \approx \rho_1 - \rho_2$$

$$SE = \sqrt{\frac{\rho_1(1-\rho_1)}{n_1} + \frac{\rho_2(1-\rho_2)}{n_2}}$$

Analyzing a Two-way Contingency Table (Cont.)

$$\text{Relative Risk} = \frac{\pi_1}{\pi_2}$$

Odds Ratio

$$\text{Odds Ratio} = \frac{\text{Odds1}}{\text{Odds2}}$$

$$= \frac{\pi 1 / (1-\pi 1)}{\pi 2 / (1-\pi 2)} = \frac{\pi 11 / \lambda 12}{\pi 21 / \pi 22} = \frac{\pi 11 \cdot \pi 22}{\pi 12 \cdot \pi 21}$$

$$SE = \sqrt{\frac{1}{n11} + \frac{1}{n12} + \frac{1}{n21} + \frac{1}{n22}}$$

Example

- Data

  P1 = 200/400 = 0.5

  P2 = 200/600 = 0.33

  – Difference of two proportions

    P1 - P2 = 0.17

  – Relative risk

    P1/P2 = 1.51

  – Odds Ratio

    (200*400)/(200*200) = 2

BGSU
Center for Family and Demographic Research

Analyzing a Three-way Contingency Table

- A three-way contingency table can be viewed as multiple two-way contingency tables created at different levels of a third variable.

- Example:

Table. Relations among Country, Gender, and Employment

	County A		Country B	
	Employed	Unemploye	Employed	Unemployed
Male	180	120	20	80
Female	120	80	80	320

BGSU
Center for Family and Demographic Research

Example

– Difference of proportion

Country A: (180/300) – (120/200)=0

Country B: (20/100) –(80/320)=0

– Relative risk

Country A: (180/300)/(120/200)=0.6/0.6=1

Country B: (20/100) –(80/320)=0.2/0.2=1

– Odds Ratio

Country A: (180*80)/(120*120)=1

Country B: (20*320)*(80*80)=1

Models for Examining How Characteristics of Individuals Affect Choices

Logistic Regression

$$\log\left(\frac{p_1}{p_2}\right) = \log\left(\frac{p_1}{1-p_1}\right) = \alpha + \beta\chi$$

$$\pi(\chi) = \frac{\exp(\alpha + \beta\chi)}{1 + \exp(\alpha + \beta\chi)} = \frac{e^{\alpha+\beta\chi}}{1 + e^{\alpha+\beta\chi}}$$

Ordered Logistic Regression

$$p(Y \le j) = \pi_1 + \ldots + \pi_{j,}\ j = 1, \ldots, J$$

$$\operatorname{logit}[p(Y \le j)] = \log\left[\frac{p(Y \le j)}{1 - p(Y \le j)}\right] = \log\left[\frac{\pi_1 + \ldots + \pi_{j,}}{\pi_{j+1} + \ldots + \pi_{J,}}\right],\ j = 1, \ldots, J$$

BGSU Center for Family and Demographic Research

Models for Examining How Characteristics of Individuals Affect Choices (Cont.)

Multinomial Logistic Regression

$$\log(\frac{p_j}{p_J}) = \alpha_j + \beta_j \chi,\ j = 1, \ldots, J-1$$

$$\log(\frac{p_a}{p_b}) = \log(\frac{\frac{p_a}{p_J}}{\frac{p_b}{p_J}}) = \log(\frac{p_a}{p_J}) - \log(\frac{p_b}{p_J})$$

$$= (\alpha_{\mathrm{a}} + \beta_a \chi) - (\alpha_{\mathrm{b}} + \beta_b \chi)$$

$$= (\alpha_\alpha - \alpha_b) + (\beta_a - \beta_b)\chi$$

Relations among These Three Models

- Ordered logistic regression and multinomial logistic regression are an extension of logistic regression.

- Both ordered and multinomial logistic regression can be treated as models simultaneously estimating a series of logistic regression.

- Ordered logistic regression assumes different intercepts, but the same slope for different categories, while multinomial logistic regression assumes different intercept and slope parameters for different categories.

A List of Variables in the Data

variable name	variable label	Label Value	Label Label
aid	ID	57101310 - 99719978	
married	Marital Status	0	Not married
		1	Married
educ	Education	1	Less than High School
		2	High School
		3	Some college
		4	colleges or more
union	Union Status	0	single
		1	cohabiting
		2	married
female	Female	0	Male
		1	Female
age	Age		24-33
agesq	Age squared		576-1089
femaleage	Interaction term of female and age		0-33

BGSU

Center for Family and Demographic Research

Data for Logistic Regression, Ordered Logistic Regression, and Multinomial Logistic Regression

	aid	married	educ	union	female	age	agesq	femaleage
1	57101310	1	2	2	1	31	961	31
2	57103869	0	1	0	0	32	1024	0
3	57109625	0	1	0	0	27	729	0
4	57111071	0	3	0	0	27	729	0
5	57113943	0	3	1	0	29	841	0
6	57117542	0	1	0	0	28	784	0
7	57118381	1	3	2	1	25	625	25
8	57118943	1	4	2	1	29	841	29
9	57120005	0	4	0	0	26	676	0
10	57120046	1	3	2	0	31	961	0
11	57120371	1	2	2	1	31	961	31
12	57121404	1	4	2	1	28	784	28
13	57121476	0	2	0	1	27	729	27
14	57127241	0	3	1	1	26	676	26
15	57129567	0	3	1	0	27	729	0
16	57131432	1	3	2	0	29	841	0
17	57131909	0	3	1	0	26	676	0
18	57133772	0	3	0	1	26	676	26
19	57134457	0	3	0	1	28	784	28
20	57134967	0	1	1	1	26	676	26

Stata Commands

Logistic Regression

```
logit married female age femaleage
logit married female age femaleage, or
```

Ordered Logistic Regression

```
ologit educ female age femaleage
ologit educ female age femaleage, or
```

Multinomial Logistic Regression

```
mlogit union female age femaleage,
base(0)
```

BGSU

Center for Family and Demographic Research

SAS Commands

Logistic Regression

```
Proc Logistic data = in.annotated_3_2;
Format married marriedf. educ educf.;
Model married = educ female age femaleage;
run;
```

BGSU Center for Family and Demographic Research

SAS Commands

Ordered Logistic Regression

```
Proc Logistic data = in.annotated descending;
Format educ educf. female femalef.;
Model educ = female age femaleage;
run;

PROC QLIM data = in.annotated;
MODEL educ = female age
femaleage/DISCRETE (DIST=LOGISTIC);
RUN;
```

BGSU
Center for Family and Demographic Research

SAS and Stata Commands

Multinomial Logistic Regression

```
proc logistic data = in.annotated_3_2;
class union (ref = "0");
model union = female age femaleage/ link = glogit;
run;
```

Interpreting the Results

- The sample size
- The reference category
- The regression coefficients
- The odds ratio

Predicted Probability

- Predicted probability is useful to describe the results
- Odds = Exp(the sum of coefficients)
- Predicted Probability = Odds/(1+Odds)
- You can present predicted probability with graphs

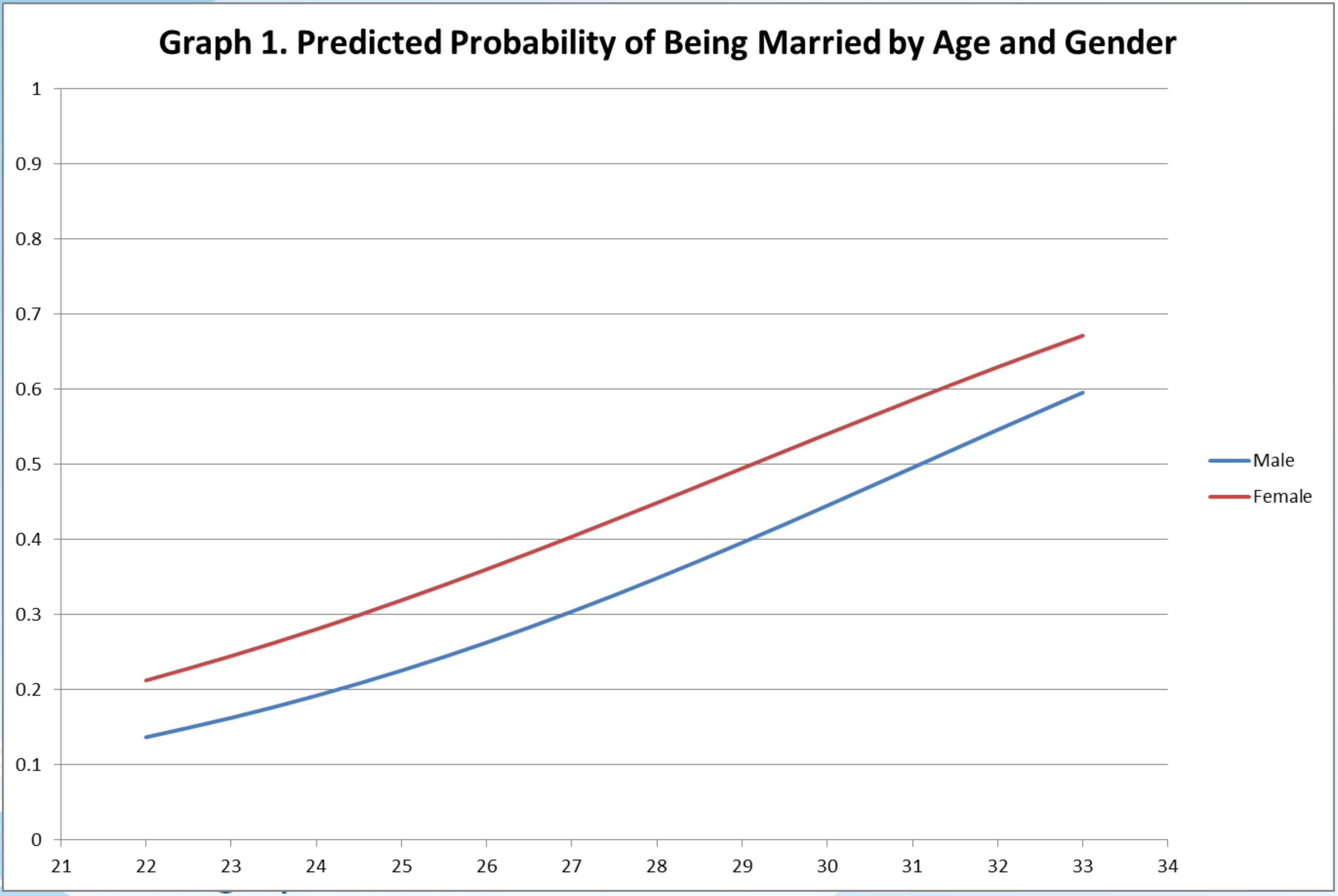
BGSU
Center for Family and Demographic Research

Predicted Probability (continued)

Table 3. Predicated Probability for Male and Female Respondents

Intercept	Female		Age		Age*Female		Sum of coefficents	Odds Ratio	Predicted Probability
coefficent	value	coefficent	value	coefficent	value	coefficent			
-6.295917	0	0.9380865	22	0.2025471	0	-0.0185092	-1.8398808	0.158836358	0.137065391
-6.295917	0	0.9380865	23	0.2025471	0	-0.0185092	-1.6373337	0.194497941	0.162828193
-6.295917	0	0.9380865	24	0.2025471	0	-0.0185092	-1.4347866	0.238166183	0.192353972
-6.295917	0	0.9380865	25	0.2025471	0	-0.0185092	-1.2322395	0.291638721	0.225789701
-6.295917	0	0.9380865	26	0.2025471	0	-0.0185092	-1.0296924	0.357116793	0.263143743
-6.295917	0	0.9380865	27	0.2025471	0	-0.0185092	-0.8271453	0.437295855	0.30424902
-6.295917	0	0.9380865	28	0.2025471	0	-0.0185092	-0.6245982	0.53547654	0.348736386
-6.295917	0	0.9380865	29	0.2025471	0	-0.0185092	-0.4220511	0.655700532	0.396026044
-6.295917	0	0.9380865	30	0.2025471	0	-0.0185092	-0.219504	0.802916946	0.44534328
-6.295917	0	0.9380865	31	0.2025471	0	-0.0185092	-0.0169569	0.983186059	0.495760877
-6.295917	0	0.9380865	32	0.2025471	0	-0.0185092	0.1855902	1.203928789	0.546264832
-6.295917	0	0.9380865	33	0.2025471	0	-0.0185092	0.3881373	1.474232182	0.595834212
-6.295917	1	0.9380865	22	0.2025471	22	-0.0185092	-1.3089967	0.270090903	0.212654781
-6.295917	1	0.9380865	23	0.2025471	23	-0.0185092	-1.1249588	0.324665843	0.245092636
-6.295917	1	0.9380865	24	0.2025471	24	-0.0185092	-0.9409209	0.390268272	0.280714363
-6.295917	1	0.9380865	25	0.2025471	25	-0.0185092	-0.756883	0.469126417	0.319323383
-6.295917	1	0.9380865	26	0.2025471	26	-0.0185092	-0.5728451	0.563918749	0.360580592
-6.295917	1	0.9380865	27	0.2025471	27	-0.0185092	-0.3888072	0.67786495	0.404004476
-6.295917	1	0.9380865	28	0.2025471	28	-0.0185092	-0.2047693	0.814835277	0.448985805
-6.295917	1	0.9380865	29	0.2025471	29	-0.0185092	-0.0207314	0.979482018	0.494817336
-6.295917	1	0.9380865	30	0.2025471	30	-0.0185092	0.1633065	1.177397507	0.540736133
-6.295917	1	0.9380865	31	0.2025471	31	-0.0185092	0.3473444	1.415304072	0.585973455
-6.295917	1	0.9380865	32	0.2025471	32	-0.0185092	0.5313823	1.701282367	0.629805454
-6.295917	1	0.9380865	33	0.2025471	33	-0.0185092	0.7154202	2.04504583	0.671597718

Predicted Probability (continued)



Conclusions

- If you have categorical dependent variables, you need to choose adequate methods to analyze them.
- You need to choose the regression models that fit your data and research questions.
- If you have event counts (e.g., the number of accidents), you need to use other models such as Poisson regression, Log-linear model, or Negative binomial regression for analyses.
- For additional help with categorical data analysis, feel free to contact me at wuh@bgsu.edu and 372-3119.

```
*************************
* Logistic Regression
*************************

. logit married female age femaleage

Iteration 0:   log likelihood = -3486.5653
Iteration 1:   log likelihood = -3394.8573
Iteration 2:   log likelihood = -3394.6262
Iteration 3:   log likelihood = -3394.6262

Logistic regression                               Number of obs   =       5114
                                                  LR chi2(3)      =     183.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -3394.6262                       Pseudo R2       =     0.0264

------------------------------------------------------------------------------
     married |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .9380865    .934408     1.00   0.315    -.8933196    2.769493
         age |   .2025471   .0242963     8.34   0.000     .1549273    .2501669
   femaleage |  -.0185092   .0326631    -0.57   0.571    -.0825278    .0455093
       _cons |  -6.295917   .6988726    -9.01   0.000    -7.665682   -4.926152
------------------------------------------------------------------------------

.
. logit married female age femaleage,or

Iteration 0:   log likelihood = -3486.5653
Iteration 1:   log likelihood = -3394.8573
Iteration 2:   log likelihood = -3394.6262
Iteration 3:   log likelihood = -3394.6262

Logistic regression                               Number of obs   =       5114
                                                  LR chi2(3)      =     183.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -3394.6262                       Pseudo R2       =     0.0264

------------------------------------------------------------------------------
     married | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   2.555088   2.387494     1.00   0.315     .4092948    15.95054
         age |   1.224518   .0297512     8.34   0.000     1.167573     1.28424
   femaleage |    .981661   .0320641    -0.57   0.571     .9207858    1.046561
       _cons |   .0018438   .0012886    -9.01   0.000     .0004686    .0072544
------------------------------------------------------------------------------

.
*************************
* Ordered Logistic Regression
*************************
.
. ologit educ female age femaleage

Iteration 0:   log likelihood = -6252.7947
Iteration 1:   log likelihood = -6206.1378
Iteration 2:   log likelihood = -6206.0689
Iteration 3:   log likelihood = -6206.0689

Ordered logistic regression                       Number of obs   =       5113
                                                  LR chi2(3)      =      93.45
                                                  Prob > chi2     =     0.0000
Log likelihood = -6206.0689                       Pseudo R2       =     0.0075

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.877126   .8172138     2.30   0.022      .275416    3.478835
         age |   -.010898   .0208886    -0.52   0.602    -.0518389    .0300429
   femaleage |  -.0494917   .0286535    -1.73   0.084    -.1056516    .0066682
-------------+----------------------------------------------------------------
       /cut1 |   -2.54953   .5995614                     -3.724648   -1.374411
       /cut2 |  -1.213573   .5982002                     -2.386024   -.0411226
       /cut3 |   .6857857   .5979509                     -.4861766    1.857748
```

```
------------------------------------------------------------------------------

.
. ologit educ female age femaleage,or

Iteration 0:   log likelihood = -6252.7947
Iteration 1:   log likelihood = -6206.1378
Iteration 2:   log likelihood = -6206.0689
Iteration 3:   log likelihood = -6206.0689

Ordered logistic regression                       Number of obs   =       5113
                                                  LR chi2(3)      =      93.45
                                                  Prob > chi2     =     0.0000
Log likelihood = -6206.0689                       Pseudo R2       =     0.0075

------------------------------------------------------------------------------
        educ | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   6.534694   5.340242     2.30   0.022     1.317079    32.42193
         age |   .9891612   .0206622    -0.52   0.602     .9494818    1.030499
   femaleage |   .9517131   .0272699    -1.73   0.084     .8997381     1.00669
-------------+----------------------------------------------------------------
       /cut1 |   -2.54953   .5995614                     -3.724648   -1.374411
       /cut2 |  -1.213573   .5982002                     -2.386024   -.0411226
       /cut3 |   .6857857   .5979509                     -.4861766    1.857748
------------------------------------------------------------------------------

.
.
**********************************
* Multinomial Logistic Regression
***********************************
.
. mlogit union female age femaleage, base(0)

Iteration 0:   log likelihood =  -5376.261
Iteration 1:   log likelihood = -5282.5618
Iteration 2:   log likelihood = -5282.3053
Iteration 3:   log likelihood = -5282.3053

Multinomial logistic regression                   Number of obs   =       5114
                                                  LR chi2(6)      =     187.91
                                                  Prob > chi2     =     0.0000
Log likelihood = -5282.3053                       Pseudo R2       =     0.0175

------------------------------------------------------------------------------
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
single       |  (base outcome)
-------------+----------------------------------------------------------------
cohabiting   |
      female |   .8175782   1.228829     0.67   0.506    -1.590882    3.226038
         age |   .0030635    .030982     0.10   0.921    -.0576601    .0637871
   femaleage |  -.0239002   .0435381    -0.55   0.583    -.1092333    .0614329
       _cons |  -.8159613   .8787378    -0.93   0.353    -2.538256    .9063331
-------------+----------------------------------------------------------------
married      |
      female |    1.22508   1.024113     1.20   0.232    -.7821444    3.232305
         age |   .2035444    .026309     7.74   0.000     .1519797    .2551092
   femaleage |   -.026943   .0358749    -0.75   0.453    -.0972566    .0433705
       _cons |  -5.930568   .7551726    -7.85   0.000    -7.410679   -4.450457
------------------------------------------------------------------------------
```

The SAS System

*** Logistic Regression***

The LOGISTIC Procedure

Model Information		
Data Set	IN.ANNOTATED_3_2	annotated_3_2 dataset written by Stat/Transfer Ver. 11.2.2106.0521
Response Variable	married	Current marital status
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	5114
Number of Observations Used	5114

Response Profile

Ordered Value	married	Total Frequency
1	married	2172
2	not married	2942

Probability modeled is married='married'.

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	6975.131	6797.252
SC	6981.670	6823.411
-2 Log L	6973.131	6789.252

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	183.8783	3	<.0001
Score	180.7636	3	<.0001
Wald	175.3151	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-6.2959	0.6989	81.1563	<.0001
female	1	0.9381	0.9344	1.0079	0.3154
age	1	0.2025	0.0243	69.4981	<.0001
femaleage	1	-0.0185	0.0327	0.3211	0.5709

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
female	2.555	0.409	15.951
age	1.225	1.168	1.284
femaleage	0.982	0.921	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	56.8	**Somers' D**	0.220
Percent Discordant	34.8	**Gamma**	0.240
Percent Tied	8.4	**Tau-a**	0.107
Pairs	6390024	**c**	0.610

The SAS System

*** Ordered Logistic Regression, Using “Proc Logistic” Procedure ***

The LOGISTIC Procedure

Model Information

Data Set	IN.ANNOTATED_3_2	annotated_3_2 dataset written by Stat/Transfer Ver. 11.2.2106.0521
Response Variable	educ	Education
Number of Response Levels	4	
Model	cumulative logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	5114
Number of Observations Used	5113

Response Profile

Ordered Value	educ	Total Frequency
1	colleges or more	1668
2	Some college	2211
3	Less than High School	399
4	High School	835

Probabilities modeled are cumulated over the lower Ordered Values.

Note: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

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

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
20.3374	6	0.0024

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	12511.589	12419.182
SC	12531.208	12458.419
-2 Log L	12505.589	12407.182

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	98.4077	3	<.0001
Score	97.9548	3	<.0001
Wald	97.1834	3	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	**colleges or more**	1	-0.5802	0.5974	0.9433	0.3314
Intercept	**Some college**	1	1.3187	0.5977	4.8672	0.0274
Intercept	**Less than High School**	1	1.8133	0.5981	9.1931	0.0024
female		1	2.0949	0.8184	6.5518	0.0105
age		1	-0.0146	0.0209	0.4922	0.4830
femaleage		1	-0.0572	0.0287	3.9675	0.0464

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
female	8.124	1.634	40.405
age	0.985	0.946	1.027
femaleage	0.944	0.893	0.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	52.3	**Somers' D**	0.120
Percent Discordant	40.3	**Gamma**	0.130

Association of Predicted Probabilities and Observed Responses			
Percent Tied	7.5	**Tau-a**	0.081
Pairs	8807799	**c**	0.560

The SAS System

*** Ordered Logistic Regression, Using “Proc QLIM” Procedure ***

The QLIM Procedure

Discrete Response Profile of educ		
Index	**Value**	**Total Frequency**
1	1	399
2	2	835
3	3	2211
4	4	1668

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	educ
Number of Observations	5113
Missing Values	1
Log Likelihood	-6206
Maximum Absolute Gradient	0.0002955
Number of Iterations	18
Optimization Method	Quasi-Newton
AIC	12424
Schwarz Criterion	12463

Goodness-of-Fit Measures		
Measure	**Value**	**Formula**
Likelihood Ratio (R)	93.451	2 * (LogL - LogL0)
Upper Bound of R (U)	12506	- 2 * LogL0
Aldrich-Nelson	0.0179	R / (R+N)
Cragg-Uhler 1	0.0181	1 - exp(-R/N)
Cragg-Uhler 2	0.0198	(1-exp(-R/N)) / (1-exp(-U/N))

Goodness-of-Fit Measures

Measure	Value	Formula
Estrella	0.0182	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.0159	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.0075	R / U
Veall-Zimmermann	0.0253	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.0596	
N = # of observations, K = # of regressors		

Algorithm converged.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	2.549530	0.599559	4.25	<.0001
female	1	1.877126	0.817215	2.30	0.0216
age	1	-0.010898	0.020889	-0.52	0.6019
femaleage	1	-0.049492	0.028654	-1.73	0.0841
_Limit2	1	1.335956	0.045318	29.48	<.0001
_Limit3	1	3.235315	0.055188	58.62	<.0001

The SAS System

*** Multinomial Logistic Regression ***

The LOGISTIC Procedure

Model Information		
Data Set	IN.ANNOTATED_3_2	annotated_3_2 dataset written by Stat/Transfer Ver. 11.2.2106.0521
Response Variable	union	Union Status
Number of Response Levels	3	
Model	generalized logit	
Optimization Technique	Newton-Raphson	

Number of Observations Read	5114
Number of Observations Used	5114

Response Profile

Ordered Value	union	Total Frequency
1	0	1936
2	1	1006
3	2	2172

Logits modeled use union=0 as the reference category.

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	10756.522	10580.611
SC	10769.601	10632.929
-2 Log L	10752.522	10564.611

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	187.9113	6	<.0001
Score	185.1487	6	<.0001
Wald	179.5235	6	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
female	2	1.4589	0.4822
age	2	69.4399	<.0001
femaleage	2	0.6271	0.7308

Analysis of Maximum Likelihood Estimates						
Parameter	**union**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	**1**	1	-0.8139	0.8787	0.8579	0.3543
Intercept	**2**	1	-5.9266	0.7551	61.5971	<.0001
female	**1**	1	0.8154	1.2288	0.4403	0.5070
female	**2**	1	1.2211	1.0241	1.4217	0.2331
age	**1**	1	0.00300	0.0310	0.0094	0.9229
age	**2**	1	0.2034	0.0263	59.7839	<.0001
femaleage	**1**	1	-0.0238	0.0435	0.2995	0.5842
femaleage	**2**	1	-0.0268	0.0359	0.5585	0.4549

Odds Ratio Estimates				
Effect	**union**	**Point Estimate**	**95% Wald Confidence Limits**	
female	**1**	2.260	0.203	25.124
female	**2**	3.391	0.456	25.236
age	**1**	1.003	0.944	1.066
age	**2**	1.226	1.164	1.290
femaleage	**1**	0.976	0.897	1.063
femaleage	**2**	0.974	0.907	1.044