Having Fun with Regressions Using SAS

March 16, 2009

BGSU
Center for Family and Demographic Research

What are the Different Proc for Regression in SAS?

- CALIS
- CATMOD
- GENMOD
- GLM
- LIFEREG
- LOESS
- LOGISTIC
- NLIN
- ORTHOREG
- PLS
- PROBIT
- REG
- RSREG
- TPSLINE
- TRANSREG

We are going to focus on:

- CATMOD-loglinear models and logistic
- GLM-multiple, polynomial, and weighted regression and ANOVA
- LOGISTIC-logistic models
- PROBIT-logistic and ordinal logistic
- REG-detailed linear regression

Some Advanced Models

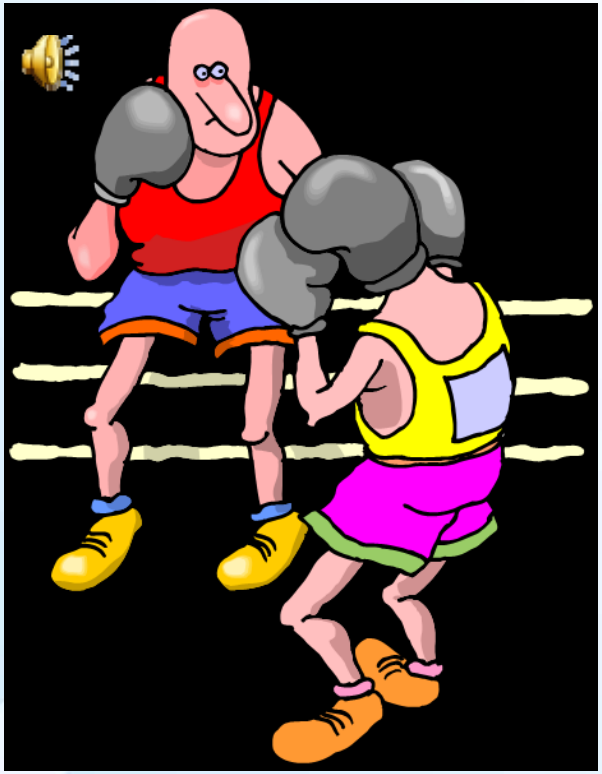
- Proc AUTOREG-time-series data with errors that are autocorrelated
- Proc MIXED-hieratical data

Side Note

- **data b.ws; set b.ws;**
- **white=(r1=1);**
- **black=(r1=2);**
- **Hisp=(r1=3);**
- **Other=(r1=4);run;**
- **proc freq; tables r1 white black hisp other ;run;**

Continuous Dependent Variable

- PROC REG VS PROC GLM



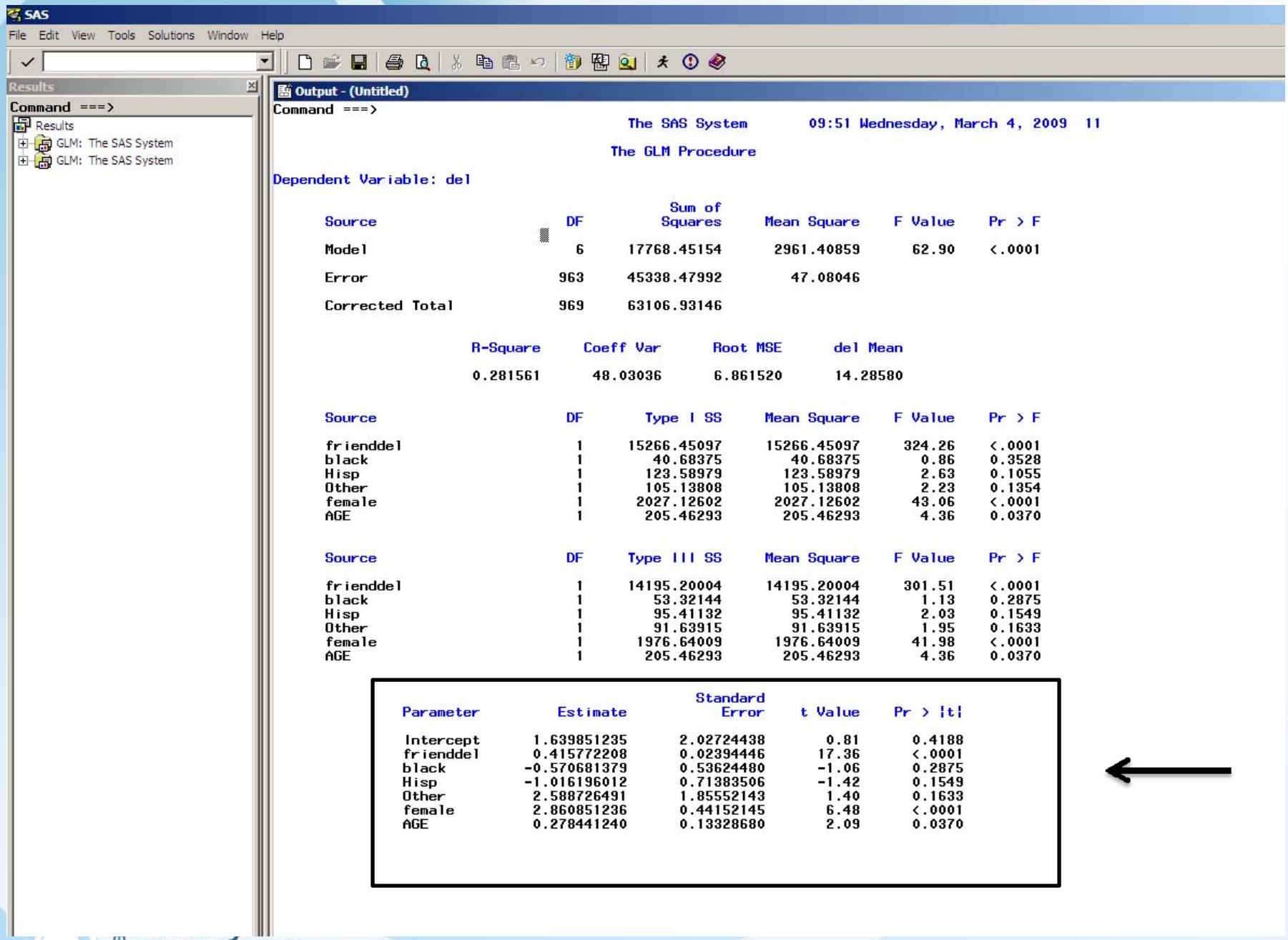
PROC GLM

- Can be used for a lot of different models including
  - Multiple Regression
  - ANOVA
  - Weighted Regression
- You DO NOT need a class statement in the code if you want regression coefficients.
- Will make interactions for you.

Proc GLM

```
proc glm;
model del= frienddel black hisp other female
   age;run;
```

BGSU
Center for Family and Demographic Research


SAS
File Edit View Tools Solutions Window Help
Results
Command ===>
Results
GLM: The SAS System
GLM: The SAS System
Output - (Untitled)
Command ===>
The SAS System 09:51 Wednesday, March 4, 2009 11
The GLM Procedure
Dependent Variable: del
Source DF Sum of Squares Mean Square F Value Pr > F
Model 6 17768.45154 2961.40859 62.90 <.0001
Error 963 45338.47992 47.08046
Corrected Total 969 63106.93146
R-Square Coeff Var Root MSE del Mean
0.281561 48.03036 6.861520 14.28580
Source DF Type I SS Mean Square F Value Pr > F
frienddel 1 15266.45097 15266.45097 324.26 <.0001
black 1 40.68375 40.68375 0.86 0.3528
Hisp 1 123.58979 123.58979 2.63 0.1055
Other 1 105.13808 105.13808 2.23 0.1354
female 1 2027.12602 2027.12602 43.06 <.0001
AGE 1 205.46293 205.46293 4.36 0.0370
Source DF Type III SS Mean Square F Value Pr > F
frienddel 1 14195.20004 14195.20004 301.51 <.0001
black 1 53.32144 53.32144 1.13 0.2875
Hisp 1 95.41132 95.41132 2.03 0.1549
Other 1 91.63915 91.63915 1.95 0.1633
female 1 1976.64009 1976.64009 41.98 <.0001
AGE 1 205.46293 205.46293 4.36 0.0370
Parameter Estimate Standard Error t Value Pr > |t|
Intercept 1.639851235 2.02724438 0.81 0.4188
frienddel 0.415772208 0.02394446 17.36 <.0001
black -0.570681379 0.53624480 -1.06 0.2875
Hisp -1.016196012 0.71383506 -1.42 0.1549
Other 2.588726491 1.85552143 1.40 0.1633
female 2.860851236 0.44152145 6.48 <.0001
AGE 0.278441240 0.13328680 2.09 0.0370

Proc GLM

proc glm;

model del= frienddel black hisp other female|age;**run;**

BGSU
Center for Family and Demographic Research

Output - (Untitled)

Command ===>

The SAS System 09:51 Wednesday, March 4, 2009 14

The GLM Procedure

Dependent Variable: del

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	18057.91260	2579.70180	55.09	<.0001
Error	962	45049.01886	46.82850		
Corrected Total	969	63106.93146			

R-Square	Coeff Var	Root MSE	del Mean
0.286148	47.90167	6.843135	14.28580

Source	DF	Type I SS	Mean Square	F Value	Pr > F
frienddel	1	15266.45097	15266.45097	326.01	<.0001
black	1	40.68375	40.68375	0.87	0.3515
Hisp	1	123.58979	123.58979	2.64	0.1046
Other	1	105.13808	105.13808	2.25	0.1344
female	1	2027.12602	2027.12602	43.29	<.0001
AGE	1	205.46293	205.46293	4.39	0.0365
female*AGE	1	289.46106	289.46106	6.18	0.0131

Source	DF	Type III SS	Mean Square	F Value	Pr > F
frienddel	1	14445.74745	14445.74745	308.48	<.0001
black	1	52.93458	52.93458	1.13	0.2880
Hisp	1	102.30558	102.30558	2.18	0.1397
Other	1	100.30662	100.30662	2.14	0.1436
female	1	145.66046	145.66046	3.11	0.0781
AGE	1	3.45036	3.45036	0.07	0.7861
female*AGE	1	289.46106	289.46106	6.18	0.0131

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	6.636082917	2.85062947	2.33	0.0201
frienddel	0.421129223	0.02397731	17.56	<.0001
black	-0.568608099	0.53480865	-1.06	0.2880
Hisp	-1.052491239	0.71207208	-1.48	0.1397
Other	2.709315323	1.85118530	1.46	0.1436
female	-7.128918048	4.04210802	-1.76	0.0781
AGE	-0.050943172	0.18767601	-0.27	0.7861
female*AGE	0.645156636	0.25949275	2.49	0.0131

Proc REG

- Linear regression
- Can have more than one model statement under the same proc
- Test for collinearity
- Plot

Proc Reg

```
proc reg;
model del= frienddel black hisp other female
   age;run;
```

BGSU
Center for Family and Demographic Research

Output - (Untitled)

Command ===>

The REG Procedure
Model: MODEL1
Dependent Variable: del

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	17768	2961.40859	62.90	<.0001
Error	963	45338	47.08046		
Corrected Total	969	63107			

Root MSE	6.86152	R-Square	0.2816
Dependent Mean	14.28580	Adj R-Sq	0.2771
Coeff Var	48.03036		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.63985	2.02724	0.81	0.4188
frienddel		1	0.41577	0.02394	17.36	<.0001
black		1	-0.57068	0.53624	-1.06	0.2875
Hisp		1	-1.01620	0.71384	-1.42	0.1549
Other		1	2.58873	1.85552	1.40	0.1633
female		1	2.86085	0.44152	6.48	<.0001
AGE	Respondents age	1	0.27844	0.13329	2.09	0.0370

Variance Inflation Factor

```
proc reg;
model del= frienddel black hisp other female
   age/VIF;run;
```

BGSU
Center for Family and Demographic Research

Output - (Untitled)

Command ===>

The SAS System 13:57 Wednesday, March 4, 2009 5

The REG Procedure
Model: MODEL1
Dependent Variable: del

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	17768	2961.40859	62.90	<.0001
Error	963	45338	47.08046		
Corrected Total	969	63107			

Root MSE	6.86152	R-Square	0.2816
Dependent Mean	14.28580	Adj R-Sq	0.2771
Coeff Var	48.03036		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	1.63985	2.02724	0.81	0.4188	0
frienddel		1	0.41577	0.02394	17.36	<.0001	1.07095
black		1	-0.57068	0.53624	-1.06	0.2875	1.04891
Hisp		1	-1.01620	0.71384	-1.42	0.1549	1.05554
Other		1	2.58873	1.85552	1.40	0.1633	1.00903
female		1	2.86085	0.44152	6.48	<.0001	1.00367
AGE	Respondents age	1	0.27844	0.13329	2.09	0.0370	1.05973

Standardized Coefficients

proc reg;

model del= frienddel black hisp other female age/stb;**run;**

Output - (Untitled)

Command ===>

The SAS System 13:57 Wednesday, March 4, 2009 20

The REG Procedure
Model: MODEL1
Dependent Variable: del

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	17768	2961.40859	62.90	<.0001
Error	963	45338	47.08046		
Corrected Total	969	63107			

Root MSE	6.86152	R-Square	0.2816
Dependent Mean	14.28580	Adj R-Sq	0.2771
Coeff Var	48.03036		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	Intercept	1	1.63985	2.02724	0.81	0.4188	0
frienddel		1	0.41577	0.02394	17.36	<.0001	0.49081
black		1	-0.57068	0.53624	-1.06	0.2875	-0.02977
Hisp		1	-1.01620	0.71384	-1.42	0.1549	-0.03995
Other		1	2.58873	1.85552	1.40	0.1633	0.03828
female		1	2.86085	0.44152	6.48	<.0001	0.17730
AGE	Respondents age	1	0.27844	0.13329	2.09	0.0370	0.05874

Model Selection Method

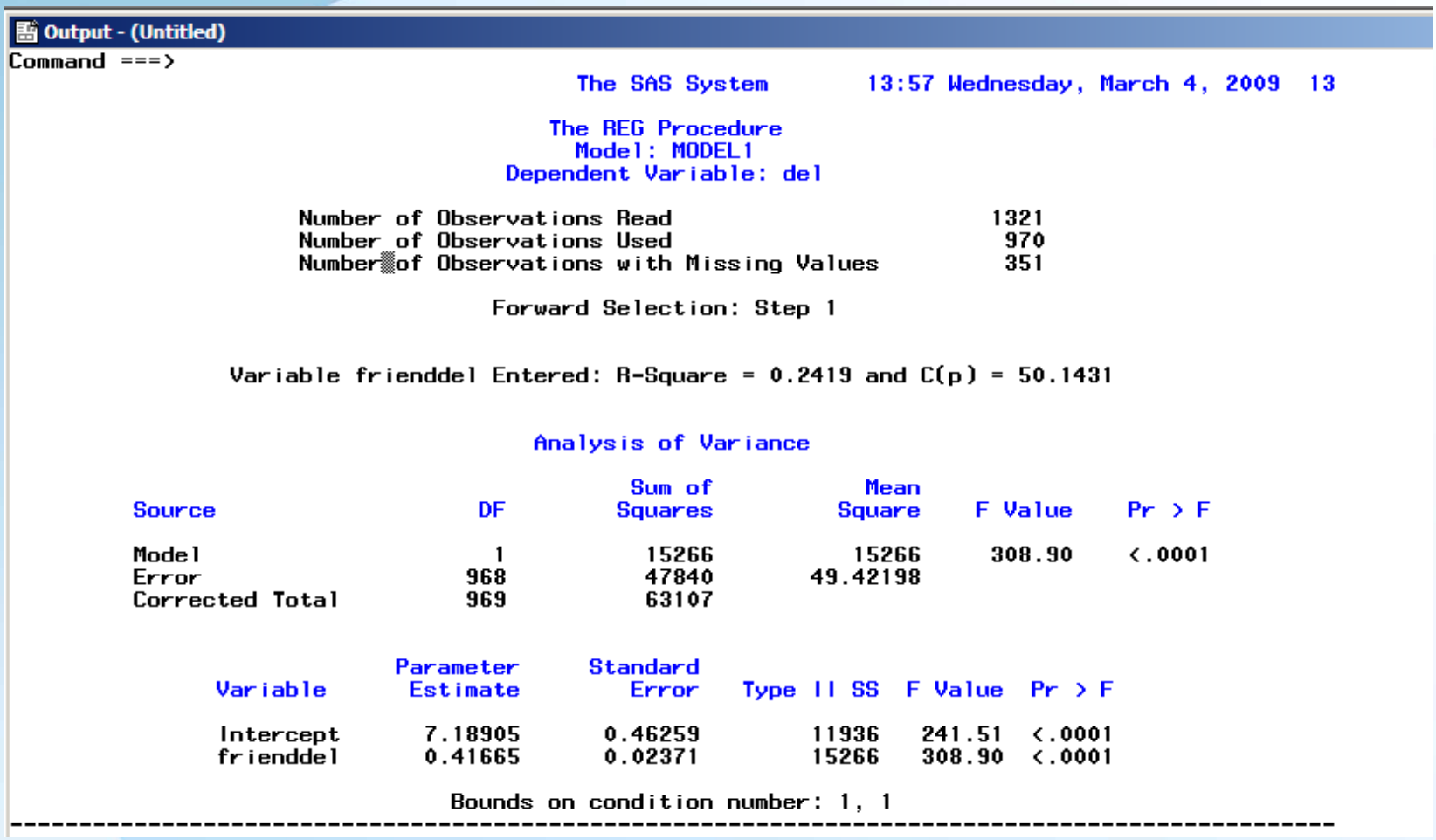
- Forward-variables are added one by one based on maxing the model fit
- Backward-eliminates based on the smallest contribution to the mode
- Stepwise-variables in the model can come out in a later step. Can put a p-value
- MaxR-list the best 1 variable model, 2 variable etc.
- MinR

Model Selection Method

proc reg;

model del= frienddel black hisp other female age/selection=forward;**run;**

BGSU
Center for Family and Demographic Research



Output - (Untitled)

Command ===>

The SAS System 13:57 Wednesday, March 4, 2009 13

The REG Procedure
Model: MODEL1
Dependent Variable: del

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Forward Selection: Step 1

Variable friendddel Entered: R-Square = 0.2419 and C(p) = 50.1431

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	15266	15266	308.90	<.0001
Error	968	47840	49.42198		
Corrected Total	969	63107			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	7.18905	0.46259	11936	241.51	<.0001
friendddel	0.41665	0.02371	15266	308.90	<.0001

Bounds on condition number: 1, 1

BGSU Center for Family and Demographic Research

Forward Selection: Step 2

Variable female Entered: R-Square = 0.2747 and C(p) = 8.2411

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	17333	8666.68778	183.09	<.0001
Error	967	45774	47.33563		
Corrected Total	969	63107			

The REG Procedure
Model: MODEL1
Dependent Variable: del

Forward Selection: Step 2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	5.60032	0.51261	5649.98729	119.36	<.0001
frienddel	0.42237	0.02322	15667	330.98	<.0001
female	2.92214	0.44221	2066.92460	43.67	<.0001

Bounds on condition number: 1.0014, 4.0056

BGSU
Center for Family and Demographic Research

proc reg;

model del= frienddel black hisp other female age/selection=rsquare;**run;**

Output - (Untitled)

Command ===>

The SAS System 13:57 Wednesday, March 4, 2009

The REG Procedure
Model: MODEL1
Dependent Variable: del

R-Square Selection Method

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Number in Model	R-Square	Variables in Model
1	0.2419	frienddel
1	0.0305	AGE
1	0.0264	female
1	0.0007	Other
1	0.0004	black
1	0.0002	Hisp
2	0.2747	frienddel female
2	0.2459	frienddel AGE
2	0.2439	frienddel Other
2	0.2434	frienddel Hisp
2	0.2426	frienddel black
2	0.0554	female AGE
2	0.0313	black AGE
2	0.0312	Other AGE
2	0.0306	Hisp AGE
2	0.0269	Other female
2	0.0267	black female
2	0.0267	Hisp female
2	0.0010	black Other
2	0.0009	Hisp Other
2	0.0005	black Hisp
3	0.2778	frienddel female AGE
3	0.2765	frienddel Other female
3	0.2759	frienddel Hisp female
3	0.2752	frienddel black female
3	0.2478	frienddel Other AGE
3	0.2473	frienddel Hisp AGE
3	0.2466	frienddel black AGE
3	0.2453	frienddel Hisp Other
3	0.2445	frienddel black Hisp
3	0.2444	frienddel black Other
3	0.0561	black female AGE
3	0.0560	Other female AGE
3	0.0556	Hisp female AGE
3	0.0319	black Other AGE
3	0.0313	black Hisp AGE

BGSU Center for Fam Der

Testing Coefficients

```
proc reg;
model del= frienddel black hisp other female
   age;
test1: test frienddel=0, age=0;
test2: test black-hisp=0;run;
```

test1: test frienddel=0, age=0;

- Tests if both frienddel and age are zero.
- If significant deleting frienddel and age would reduce the model fit

test2: test black-hisp=0;run;

- Tests if the coefficients of black and hisp are equal.
- If significant you CANNOT conclude the coefficients are equal.

Output - (Untitled)

Command ===>

The SAS System 09:06 Friday, March 6, 2009 1

The REG Procedure
Model: MODEL1
Dependent Variable: del

Number of Observations Read	1321
Number of Observations Used	970
Number of Observations with Missing Values	351

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	17768	2961.40859	62.90	<.0001
Error	963	45338	47.08046		
Corrected Total	969	63107			

Root MSE	6.86152	R-Square	0.2816
Dependent Mean	14.28580	Adj R-Sq	0.2771
Coeff Var	48.03036		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.63985	2.02724	0.81	0.4188
frienddel		1	0.41577	0.02394	17.36	<.0001
black		1	-0.57068	0.53624	-1.06	0.2875
Hisp		1	-1.01620	0.71384	-1.42	0.1549
Other		1	2.58873	1.85552	1.40	0.1633
female		1	2.86085	0.44152	6.48	<.0001
AGE	Respondents age	1	0.27844	0.13329	2.09	0.0370

Output - (Untitled)

Command ===>

The SAS System 09:06 Friday, March 6, 2009 2

The REG Procedure
Model: MODEL1

Test test1 Results for Dependent Variable del

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	8018.86781	170.32	<.0001
Denominator	963	47.08046		

We would lose in the model fit if we deleted friendddel and age

Output - (Untitled)

Command ===>

The SAS System 09:06 Friday, March 6, 2009 3

The REG Procedure
Model: MODEL1

Test test2 Results for Dependent Variable del

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	14.53004	0.31	0.5787
Denominator	963	47.08046		

Black and Hisp have the same coefficient

Binary Response Variable

```
proc logistic descending;
model highdel=frienddel black hisp other
   female age;run;
```

If your results seem reversed you probably forgot this option

Output - (Untitled)

Command ===>

The LOGISTIC Procedure

Model Information

Data Set	B.WS
Response Variable	highdel
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1321
Number of Observations Used	1300

Response Profile

Ordered Value	highdel	Total Frequency
1	1	47
2	0	1253

Probability modeled is highdel=1.

NOTE: 21 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	406.357	341.087
SC	411.527	377.278
-2 Log L	404.357	327.087

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	77.2702	6	<.0001
Score	128.7356	6	<.0001
Wald	68.7494	6	<.0001

BGSU
Center for Family and Demographic Research

Output - (Untitled)

Command ===>

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.2604	1.7758	27.1949	<.0001
frienddel	1	0.0823	0.0106	60.6830	<.0001
black	1	-0.1964	0.3987	0.2426	0.6223
Hisp	1	-0.3959	0.5270	0.5645	0.4525
Other	1	0.6958	1.0601	0.4308	0.5116
female	1	1.0061	0.3511	8.2119	0.0042
AGE	1	0.2353	0.1071	4.8283	0.0280

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
frienddel	1.086	1.064	1.108
black	0.822	0.376	1.795
Hisp	0.673	0.240	1.891
Other	2.005	0.251	16.014
female	2.735	1.374	5.443
AGE	1.265	1.026	1.561

Association of Predicted Probabilities and Observed Responses

Percent Concordant	82.8	Somers' D	0.670
Percent Discordant	15.8	Gamma	0.680
Percent Tied	1.5	Tau-a	0.047
Pairs	58891	c	0.835

Multinomial Logistic Regression

```
proc catmod;
direct frienddel black hisp other female age ;
model catdel=frienddel black hisp other
   female age /NOPROFILE;run;
```

***Have a direct statement**

***Put NOPROFILE**

Command ===>

The CATMOD Procedure

Data Summary

Response	catdel	Response Levels	3
Weight Variable	None	Populations	435
Data Set	WS	Total Frequency	966
Frequency Missing	355	Observations	966

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	2	90.55	<.0001
frienddel	2	84.47	<.0001
black	2	15.57	0.0004
Hisp	2	0.88	0.6441
Other	2	1.04	0.5950
female	2	40.39	<.0001
AGE	2	49.40	<.0001
Likelihood Ratio	856	763.62	0.9893

Analysis of Maximum Likelihood Estimates

Parameter	Function Number	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	10.7939	1.8805	32.95	<.0001
	2	4.7795	1.8390	6.75	0.0093
frienddel	1	-0.1455	0.0158	84.36	<.0001
	2	-0.0761	0.0118	41.87	<.0001
black	1	0.6349	0.4300	2.18	0.1398
	2	-0.0593	0.4179	0.02	0.8871
Hisp	1	-0.1076	0.5075	0.04	0.8322
	2	0.1174	0.4720	0.06	0.8036
Other	1	-0.6375	1.1388	0.31	0.5756
	2	-1.0201	1.1150	0.84	0.3603
female	1	-1.5022	0.3615	17.26	<.0001
	2	-0.6473	0.3479	3.46	0.0628
AGE	1	-0.3300	0.1135	8.45	0.0036
	2	-0.0198	0.1104	0.03	0.8579