```
* Probit and Logit Models in SAS;
* Copyright 2013 by Ani Katchova;

proc import out= work.data
datafile= "C:\Econometrics\Data\probit_insurance.csv"
dbms=csv replace; getnames=yes; datarow=2;
run;

proc means data=data;
var ins retire age hstatusg hhincome educyear married hisp;
run;

proc freq data=data;
tables ins;
run;

*OLS;
proc reg data=data;
model ins = retire age hstatusg hhincome educyear married hisp;
run;

*Logit;
proc logistic data=data descending;
model ins = retire age hstatusg hhincome educyear married hisp/ ctable pprob=0.5;
output out=lpred predicted=plogit;
run;

*Logit;
proc qlim data=data;
model ins = retire age hstatusg hhincome educyear married hisp / discrete
(dist=logit);
output out=mfx marginal;
run;

*Logit marginal effects;
proc means data=mfx mean std;
var Meff_P2_retire Meff_P2_age Meff_P2_hstatusg Meff_P2_hhincome Meff_P2_educyear
Meff_P2_married Meff_P2_hisp;
run;

*Logit predicted probabilities;
proc means data=lpred;
var ins plogit;
run;

*Probit;
proc logistic data=data descending;
model ins = retire age hstatusg hhincome educyear married hisp / link=probit ctable
pprob=0.5;
output out=ppred predicted=pprobit;
run;

*Probit;
proc qlim data=data;
```

```
model ins = retire age hstatusg hhincome educyear married hisp / discrete
(dist=normal);
output out=mfx marginal;
run;

*Probit marginal effects;
proc means data=mfx mean std;
var Meff_P2_retire Meff_P2_age Meff_P2_hstatusg Meff_P2_hhincome Meff_P2_educyear
Meff_P2_married Meff_P2_hisp;
run;

*Probit predicted probabilities;
proc means data=ppred;
var ins pprobit;
run;

*Probit predicting y=0;
proc probit data=data;
class ins;
model ins = retire age hstatusg hhincome educyear married hisp;
run;
```

The MEANS Procedure

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| ins | 3206 | 0.3870867 | 0.4871597 | 0 | 1.0000000 |
| retire | 3206 | 0.6247661 | 0.4842588 | 0 | 1.0000000 |
| age | 3206 | 66.9139114 | 3.6757942 | 52.0000000 | 86.0000000 |
| hstatusg | 3206 | 0.7046163 | 0.4562862 | 0 | 1.0000000 |
| hhincome | 3206 | 45.2639144 | 64.3393642 | 0 | 1312.12 |
| educyear | 3206 | 11.8986276 | 3.3046110 | 0 | 17.0000000 |
| married | 3206 | 0.7330006 | 0.4424610 | 0 | 1.0000000 |
| hisp | 3206 | 0.0726762 | 0.2596448 | 0 | 1.0000000 |



The FREQ Procedure

| ins | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| 0 | 1965 | 61.29 | 1965 | 61.29 |
| 1 | 1241 | 38.71 | 3206 | 100.00 |



The REG Procedure
Model: MODEL1
Dependent Variable: ins

| | |
|---|---|
| Number of Observations Read | 3206 |
| Number of Observations Used | 3206 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 7 | 62.84034 | 8.97719 | 41.14 | <.0001 |
| Error | 3198 | 697.78505 | 0.21819 | | |
| Corrected Total | 3205 | 760.62539 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.46711 | R-Square | 0.0826 |
| Dependent Mean | 0.38709 | Adj R-Sq | 0.0806 |
| Coeff Var | 120.67390 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 0.12709 | 0.16056 | 0.79 | 0.4287 |
| retire | 1 | 0.04085 | 0.01822 | 2.24 | 0.0250 |
| age | 1 | -0.00290 | 0.00242 | -1.20 | 0.2314 |
| hstatusg | 1 | 0.06556 | 0.01945 | 3.37 | 0.0008 |
| hhincome | 1 | 0.00049209 | 0.00013751 | 3.58 | 0.0004 |
| educyear | 1 | 0.02337 | 0.00287 | 8.15 | <.0001 |
| married | 1 | 0.12347 | 0.01936 | 6.38 | <.0001 |
| hisp | 1 | -0.12101 | 0.03367 | -3.59 | 0.0003 |



The LOGISTIC Procedure

| Model Information | |
|---|---|
| Data Set | WORK.DATA |
| Response Variable | ins |
| Number of Response Levels | 2 |
| Model | binary logit |
| Optimization Technique | Fisher's scoring |

| | |
|---|---|
| Number of Observations Read | 3206 |
| Number of Observations Used | 3206 |

| Response Profile | | |
|---|---|---|
| Ordered Value | ins | Total Frequency |
| 1 | 1 | 1241 |
| 2 | 0 | 1965 |

Probability modeled is ins='1'.

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

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| AIC | 4281.542 | 4005.757 |

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| SC | 4287.615 | 4054.339 |
| -2 Log L | 4279.542 | 3989.757 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 289.7856 | 7 | <.0001 |
| Score | 264.8691 | 7 | <.0001 |
| Wald | 238.0011 | 7 | <.0001 |

| Analysis of Maximum Likelihood Estimates | | | | | |
|---|---|---|---|---|---|
| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | 1 | -1.7155 | 0.7486 | 5.2516 | 0.0219 |
| retire | 1 | 0.1969 | 0.0842 | 5.4691 | 0.0194 |
| age | 1 | -0.0146 | 0.0113 | 1.6720 | 0.1960 |
| hstatusg | 1 | 0.3123 | 0.0917 | 11.6022 | 0.0007 |
| hhincome | 1 | 0.00230 | 0.000762 | 9.1386 | 0.0025 |
| educyear | 1 | 0.1143 | 0.0142 | 64.7335 | <.0001 |
| married | 1 | 0.5786 | 0.0933 | 38.4466 | <.0001 |
| hisp | 1 | -0.8096 | 0.1957 | 17.1137 | <.0001 |

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Point Estimate | 95% Wald Confidence Limits | |
| retire | 1.218 | 1.032 | 1.436 |
| age | 0.986 | 0.964 | 1.008 |
| hstatusg | 1.367 | 1.142 | 1.635 |
| hhincome | 1.002 | 1.001 | 1.004 |
| educyear | 1.121 | 1.090 | 1.153 |
| married | 1.784 | 1.485 | 2.142 |
| hisp | 0.445 | 0.303 | 0.653 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 66.5 | Somers' D | 0.335 |
| Percent Discordant | 33.1 | Gamma | 0.336 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Tied | 0.4 | Tau-a | 0.159 |
| Pairs | 2438565 | c | 0.667 |

Classification Table

| Prob Level | Correct | | Incorrect | | Percentages | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Event | Non-Event | Event | Non-Event | Correct | Sensi-tivity | Speci-ficity | False POS | False NEG |
| 0.500 | 337 | 1657 | 308 | 904 | 62.2 | 27.2 | 84.3 | 47.8 | 35.3 |



The QLIM Procedure

Discrete Response Profile of ins

| Index | Value | Total Frequency |
|---|---|---|
| 1 | 0 | 1965 |
| 2 | 1 | 1241 |

| Model Fit Summary | |
|---|---|
| Number of Endogenous Variables | 1 |
| Endogenous Variable | ins |
| Number of Observations | 3206 |
| Log Likelihood | -1995 |
| Maximum Absolute Gradient | 0.0001123 |
| Number of Iterations | 15 |
| Optimization Method | Quasi-Newton |
| AIC | 4006 |
| Schwarz Criterion | 4054 |

Goodness-of-Fit Measures

| Measure | Value | Formula |
|---|---|---|
| Likelihood Ratio (R) | 289.79 | 2 * (LogL - LogL0) |
| Upper Bound of R (U) | 4279.5 | - 2 * LogL0 |
| Aldrich-Nelson | 0.0829 | R / (R+N) |
| Cragg-Uhler 1 | 0.0864 | 1 - exp(-R/N) |
| Cragg-Uhler 2 | 0.1173 | (1-exp(-R/N)) / (1-exp(-U/N)) |
| Estrella | 0.0893 | 1 - (1-R/U)^(U/N) |
| Adjusted Estrella | 0.0845 | 1 - ((LogL-K)/LogL0)^(-2/N*LogL0) |

Goodness-of-Fit Measures

| Measure | Value | Formula |
|---|---|---|
| McFadden's LRI | 0.0677 | R / U |
| Veall-Zimmermann | 0.145 | (R * (U+N)) / (U * (R+N)) |
| McKelvey-Zavoina | 0.3302 | |

N = # of observations, K = # of regressors

Algorithm converged.

Parameter Estimates

| Parameter | DF | Estimate | Standard Error | t Value | Approx Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -1.715578 | 0.748643 | -2.29 | 0.0219 |
| retire | 1 | 0.196930 | 0.084205 | 2.34 | 0.0194 |
| age | 1 | -0.014596 | 0.011286 | -1.29 | 0.1959 |
| hstatusg | 1 | 0.312265 | 0.091672 | 3.41 | 0.0007 |
| hhincome | 1 | 0.002304 | 0.000762 | 3.02 | 0.0025 |
| educyear | 1 | 0.114263 | 0.014201 | 8.05 | <.0001 |
| married | 1 | 0.578636 | 0.093319 | 6.20 | <.0001 |
| hisp | 1 | -0.810306 | 0.195752 | -4.14 | <.0001 |



The MEANS Procedure

| Variable | Label | Mean | Std Dev |
|---|---|---|---|
| Meff_P2_retire | Marginal effect of retire on the probability of ins=2 | 0.0427616 | 0.0085144 |
| Meff_P2_age | Marginal effect of age on the probability of ins=2 | -0.0031693 | 0.000631046 |
| Meff_P2_hstatusg | Marginal effect of hstatusg on the probability of ins=2 | 0.0678058 | 0.0135010 |
| Meff_P2_hhincome | Marginal effect of hhincome on the probability of ins=2 | 0.000500208 | 0.000099598 |
| Meff_P2_educyear | Marginal effect of educyear on the probability of ins=2 | 0.0248111 | 0.0049402 |
| Meff_P2_married | Marginal effect of married on the probability of ins=2 | 0.1256459 | 0.0250176 |
| Meff_P2_hisp | Marginal effect of hisp on the probability of ins=2 | -0.1759510 | 0.0350340 |

The SAS System

The MEANS Procedure

| Variable | Label | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| ins | | 3206 | 0.3870867 | 0.4871597 | 0 | 1.0000000 |
| plogit | Estimated Probability | 3206 | 0.3870932 | 0.1418166 | 0.0340467 | 0.9649630 |



The LOGISTIC Procedure

| Model Information | |
|---|---|
| Data Set | WORK.DATA |
| Response Variable | ins |
| Number of Response Levels | 2 |
| Model | binary probit |
| Optimization Technique | Fisher's scoring |

| | |
|---|---|
| Number of Observations Read | 3206 |
| Number of Observations Used | 3206 |

Response Profile

| Ordered Value | ins | Total Frequency |
|---|---|---|
| 1 | 1 | 1241 |
| 2 | 0 | 1965 |

Probability modeled is ins='1'.

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

Model Fit Statistics

| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| AIC | 4281.542 | 4003.247 |
| SC | 4287.615 | 4051.830 |
| -2 Log L | 4279.542 | 3987.247 |

Testing Global Null Hypothesis: BETA=0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 292.2950 | 7 | <.0001 |
| Score | 264.8691 | 7 | <.0001 |
| Wald | 261.0477 | 7 | <.0001 |

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -1.0694 | 0.4553 | 5.5164 | 0.0188 |
| retire | 1 | 0.1183 | 0.0513 | 5.3139 | 0.0212 |
| age | 1 | -0.00887 | 0.00686 | 1.6699 | 0.1963 |
| hstatusg | 1 | 0.1977 | 0.0555 | 12.6989 | 0.0004 |
| hhincome | 1 | 0.00123 | 0.000437 | 7.9548 | 0.0048 |
| educyear | 1 | 0.0708 | 0.00849 | 69.4061 | <.0001 |
| married | 1 | 0.3623 | 0.0561 | 41.7896 | <.0001 |
| hisp | 1 | -0.4731 | 0.1102 | 18.4209 | <.0001 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 66.4 | Somers' D | 0.333 |
| Percent Discordant | 33.1 | Gamma | 0.334 |
| Percent Tied | 0.4 | Tau-a | 0.158 |
| Pairs | 2438565 | c | 0.666 |

Classification Table

| Prob Level | Correct | | Incorrect | | Percentages | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Event | Non-Event | Event | Non-Event | Correct | Sensi-tivity | Speci-ficity | False POS | False NEG |
| 0.500 | 334 | 1660 | 305 | 907 | 62.2 | 26.9 | 84.5 | 47.7 | 35.3 |



The QLIM Procedure

Discrete Response Profile of ins

| Index | Value | Total Frequency |
|---|---|---|
| 1 | 0 | 1965 |
| 2 | 1 | 1241 |

Model Fit Summary

| | |
|---|---|
| Number of Endogenous Variables | 1 |
| Endogenous Variable | ins |
| Number of Observations | 3206 |
| Log Likelihood | -1994 |
| Maximum Absolute Gradient | 3.49018E-6 |

| Model Fit Summary | |
|---|---|
| Number of Iterations | 15 |
| Optimization Method | Quasi-Newton |
| AIC | 4003 |
| Schwarz Criterion | 4052 |

| Goodness-of-Fit Measures | | |
|---|---|---|
| Measure | Value | Formula |
| Likelihood Ratio (R) | 292.3 | 2 * (LogL - LogL0) |
| Upper Bound of R (U) | 4279.5 | - 2 * LogL0 |
| Aldrich-Nelson | 0.0836 | R / (R+N) |
| Cragg-Uhler 1 | 0.0871 | 1 - exp(-R/N) |
| Cragg-Uhler 2 | 0.1183 | (1-exp(-R/N)) / (1-exp(-U/N)) |
| Estrella | 0.0901 | 1 - (1-R/U)^(U/N) |
| Adjusted Estrella | 0.0852 | 1 - ((LogL-K)/LogL0)^(-2/N*LogL0) |
| McFadden's LRI | 0.0683 | R / U |
| Veall-Zimmermann | 0.1461 | (R * (U+N)) / (U * (R+N)) |
| McKelvey-Zavoina | 0.1537 | |
| N = # of observations, K = # of regressors | | |

Algorithm converged.

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Parameter | DF | Estimate | Standard Error | t Value | Approx Pr > \|t\| |
| Intercept | 1 | -1.069319 | 0.458086 | -2.33 | 0.0196 |
| retire | 1 | 0.118357 | 0.051267 | 2.31 | 0.0210 |
| age | 1 | -0.008870 | 0.006899 | -1.29 | 0.1985 |
| hstatusg | 1 | 0.197736 | 0.055486 | 3.56 | 0.0004 |
| hhincome | 1 | 0.001233 | 0.000387 | 3.19 | 0.0014 |
| educyear | 1 | 0.070748 | 0.008478 | 8.34 | <.0001 |
| married | 1 | 0.362329 | 0.056002 | 6.47 | <.0001 |
| hisp | 1 | -0.473110 | 0.110439 | -4.28 | <.0001 |

The SAS System

The MEANS Procedure

| Variable | Label | Mean | Std Dev |
|---|---|---|---|
| Meff_P2_retire | Marginal effect of retire on the | 0.0419784 | 0.0074448 |

| Variable | Label | Mean | Std Dev |
|---|---|---|---|
| Meff_P2_age | probability of ins=2 | -0.0031459 | 0.000557915 |
| Meff_P2_hstatusg | Marginal effect of age on the probability of ins=2 | 0.0701324 | 0.0124379 |
| Meff_P2_hhincome | Marginal effect of hstatusg on the probability of ins=2 | 0.000437330 | 0.000077560 |
| Meff_P2_educyear | Marginal effect of hhincome on the probability of ins=2 | 0.0250926 | 0.0044502 |
| Meff_P2_married | Marginal effect of educyear on the probability of ins=2 | 0.1285099 | 0.0227912 |
| Meff_P2_hisp | Marginal effect of married on the probability of ins=2 | -0.1678014 | 0.0297595 |
| | Marginal effect of hisp on the probability of ins=2 | | |



The MEANS Procedure

| Variable | Label | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| ins | | 3206 | 0.3870867 | 0.4871597 | 0 | 1.0000000 |
| pprobit | Estimated Probability | 3206 | 0.3861134 | 0.1421375 | 0.0206439 | 0.9646857 |



The Probit Procedure

| Model Information | |
|---|---|
| Data Set | WORK.DATA |
| Dependent Variable | ins |
| Number of Observations | 3206 |
| Name of Distribution | Normal |
| Log Likelihood | -1993.623704 |

| | |
|---|---|
| Number of Observations Read | 3206 |
| Number of Observations Used | 3206 |

| Class Level Information | | |
|---|---|---|
| Name | Levels | Values |
| ins | 2 | 0 1 |

| Response Profile | | |
|---|---|---|
| Ordered Value | ins | Total Frequency |
| 1 | 0 | 1965 |

| Response Profile | | |
|---|---|---|
| Ordered Value | ins | Total Frequency |
| 2 | 1 | 1241 |

PROC PROBIT is modeling the probabilities of levels of ins having LOWER Ordered Values in the response profile table.

Algorithm converged.

| Type III Analysis of Effects | | | |
|---|---|---|---|
| Effect | DF | Wald Chi-Square | Pr > ChiSq |
| retire | 1 | 5.3296 | 0.0210 |
| age | 1 | 1.6529 | 0.1986 |
| hstatusg | 1 | 12.6996 | 0.0004 |
| hhincome | 1 | 10.1749 | 0.0014 |
| educyear | 1 | 69.6338 | <.0001 |
| married | 1 | 41.8584 | <.0001 |
| hisp | 1 | 18.3517 | <.0001 |

| Analysis of Maximum Likelihood Parameter Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | 1 | 1.0693 | 0.4581 | 0.1715 | 1.9671 | 5.45 | 0.0196 |
| retire | 1 | -0.1184 | 0.0513 | -0.2188 | -0.0179 | 5.33 | 0.0210 |
| age | 1 | 0.0089 | 0.0069 | -0.0047 | 0.0224 | 1.65 | 0.1986 |
| hstatusg | 1 | -0.1977 | 0.0555 | -0.3065 | -0.0890 | 12.70 | 0.0004 |
| hhincome | 1 | -0.0012 | 0.0004 | -0.0020 | -0.0005 | 10.17 | 0.0014 |
| educyear | 1 | -0.0707 | 0.0085 | -0.0874 | -0.0541 | 69.63 | <.0001 |
| married | 1 | -0.3623 | 0.0560 | -0.4721 | -0.2526 | 41.86 | <.0001 |
| hisp | 1 | 0.4731 | 0.1104 | 0.2567 | 0.6896 | 18.35 | <.0001 |