Ordinal Logistic and Probit Examples

Below is an example borrowed from Karen Seccombe's project[1] focusing on healthcare among welfare recipients in Oregon. The outcome for this model is a response to a question about how often the respondent cut meal sizes because of affordability, an indicator of food insecurity. Responses to two questions were coded into a single ordinal variable with three values, 0 = never or rarely, 1 = some months but not every month, and 2 = almost every month.

Case Processing Summary

		N	Marginal Percentage
cutmeal how often cut meal size	0 never or rarely	424	77.7%
	1 some months but not every month	56	10.3%
	2 almost every month	66	12.1%
Valid		546	100.0%
Missing		96	
Total		642	

Ordinal Logistic Model in SPSS

Regresson → ordinal → options (choose link: Logit)

```
plum cutmeal with mosmed depress1 educat marital
/link = logit
/print= parameter.
```

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	572.929			
Final	543.454	29.475	4	.000

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[cutmeal = 0]	1.884	.351	28.864	1	.000	1.197	2.571
	[cutmeal = 1]	2.659	.363	53.548	1	.000	1.947	3.371
Location	mosmed	.012	.024	.230	1	.631	-.036	.059
	depress1	.201	.039	26.158	1	.000	.124	.278
	educat	-.035	.115	.091	1	.762	-.260	.190
	marital	.463	.235	3.887	1	.049	.003	.922

Link function: Logit.

Odds ratios are not printed, but are easily computed by hand. For example, the odds ratio for `depress1` would be $e^{.201} = 1.22$. The odds ratios in this case represent the odds of moving from one category on Y to the next given an increment of X.

Ordered Logistic Model in R

Note: The `polr` function requires the outcome be a factor, and does not like categorical predictors. So, I converted predictors that were nonnumeric to numeric [I use lessR command below, but base R can be used too, e.g., `d$mosmed <- as.numeric(d$mosmed)`]. Missing data are also problematic with `polr`, so I used the following listwise deletion routine to remove cases with missing data on any of the variables in the model (using `lessR` code).

```
> library(lessR)
```

[1] Seccombe, K., Newsom, J.T., & Hoffman, K. (2006). Access to healthcare after welfare reform. *Inquiry, 43*, 167-179.

```
> #listwise deletion to match n from regression (needed to make sure nested test has same n)
> d = d[complete.cases(d[,c("cutmeal","mosmed","depress1","educat","marital")]),]
#always double check variable type changes and listwise deletion using str(d) and descriptive analysis

> #polr requires response to be a factor, so transform
> d$cutmeal <- factor(d$cutmeal)

> library(MASS)
> model <-polr(cutmeal ~ mosmed + depress1  + educat + marital,data=d,contrasts=NULL,method=c("logistic"))
> summary(model,digits = 3)

Re-fitting to get Hessian

Call:
polr(formula = cutmeal ~ mosmed + depress1 + educat + marital,
    data = d, contrasts = NULL, method = c("logistic"))

Coefficients:
           Value Std. Error t value
mosmed    0.0115     0.0239   0.483
depress1  0.2009     0.0391   5.137
educat   -0.0347     0.1153  -0.301
marital   0.4626     0.2365   1.956

Intercepts:
    Value  Std. Error t value
0|1  1.884  0.351      5.361
1|2  2.659  0.364      7.306

Residual Deviance: 718.9374
AIC: 730.9374

> #use AER coeftest and coefci for tests and confidence intervals
> library("AER")
> coeftest(model)

Re-fitting to get Hessian

z test of coefficients:

          Estimate Std. Error z value           Pr(>|z|)
mosmed    0.011520   0.023850  0.4830            0.62910
depress1  0.200902   0.039107  5.1372 0.0000002788310690
educat   -0.034684   0.115283 -0.3009            0.76352
marital   0.462583   0.236459  1.9563            0.05043
0|1       1.883932   0.351405  5.3611 0.0000000826971867
1|2       2.658861   0.363930  7.3060 0.0000000000002753

> coefci(model)

Re-fitting to get Hessian

                2.5 %     97.5 %
mosmed   -0.035330933 0.05837016
depress1  0.124081506 0.27772335
educat   -0.261142777 0.19177462
marital  -0.001909381 0.92707519
```

Ordered Probit Model in SPSS

Probit models in SPSS can be specified in several different ways. I use the PLUM procedure, but the user can use the *Ordinal* procedure (specifying probit link) or the *Probit* procedure through the menus. The *Probit* procedure requires specification of a variable with the count of total observed, so it is a less convenient approach. SPSS now has a *Generalized Linear Models* option through the menus in which ordinal logistic, probit models, Poisson, and negative binomial models can be tested.

Regresson → ordinal → options (choose link: Probit)

```
plum cutmeal with mosmed depress1 educat marital
/link = probit
/print= parameter summary.
```

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	572.929			
Final	541.558	31.371	4	.000

Link function: Probit.

Pseudo R-Square

Cox and Snell	.056
Nagelkerke	.075
McFadden	.042

Link function: Probit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[cutmeal = 0]	1.150	.198	33.700	1	.000	.761	1.538
	[cutmeal = 1]	1.584	.203	61.063	1	.000	1.187	1.982
Location	mosmed	.006	.014	.218	1	.640	-.020	.033
	depress1	.121	.023	27.759	1	.000	.076	.166
	educat	-.013	.065	.040	1	.841	-.141	.115
	marital	.260	.136	3.664	1	.056	-.006	.525

Link function: Probit.

As noted in the previous handout, standardized coefficients could be obtained in SPSS by prestandardizing the variables using the same N (e.g., using `DESCRIPTIVE VARS=mosmed(zmosmed)`) and ignoring the significance tests in the output.

Ordered Probit Model in R

(Note precautions above regarding missing data and outcome variable type also apply to probit models)

```
> library(MASS)
> model <-polr(cutmeal ~ mosmed + depress1  + educat + marital,data=d,contrasts=NULL,method=c("probit"))
> summary(model,digits = 3)

Call:
polr(formula = cutmeal ~ mosmed + depress1 + educat + marital,
    data = d, contrasts = NULL, method = c("probit"))

Coefficients:
            Value Std. Error t value
mosmed     0.00637     0.0137   0.466
depress1   0.12105     0.0230   5.267
educat    -0.01302     0.0649  -0.201
marital    0.25962     0.1359   1.910

Intercepts:
    Value  Std. Error t value
0|1  1.150  0.198      5.816
1|2  1.584  0.203      7.821

Residual Deviance: 717.0415
AIC: 729.0415

> #use AER coeftest and coefci for tests and confidence intervals
> coeftest(model1)

Re-fitting to get Hessian


z test of coefficients:

           Estimate Std. Error z value             Pr(>|z|)
mosmed     0.0063709  0.0136609  0.4664              0.64096
depress1   0.1210496  0.0229848  5.2665 0.000000139036058455
educat    -0.0130156  0.0648567 -0.2007              0.84095
marital    0.2596249  0.1359329  1.9099              0.05614
0|1        1.1495693  0.1976394  5.8165 0.000000006009265638
1|2        1.5843137  0.2025726  7.8210 0.000000000000005242

> coefci(model1)

Re-fitting to get Hessian

                2.5 %     97.5 %
mosmed    -0.020464160 0.03320602
depress1   0.075899142 0.16620014
educat    -0.140417971 0.11438673
marital   -0.007397164 0.52664703

#nested LR comparison assumes listwise deletion used so that N is the same for both nested models
> model0 <-polr(cutmeal ~ 1,data=d,contrasts=NULL,method=c("probit"))
> summary(model,digits = 3)
> model1 <-polr(cutmeal ~ mosmed + depress1  + educat + marital,data=d,contrasts=NULL,method=c("probit"))
> summary(model,digits = 3)
```

```
#requests likelihood ratio (G-squared) comparing the deviances from the two models
> anova(model0,model1,test="Chisq")

Likelihood ratio tests of ordinal regression models

Response: cutmeal
                                  Model Resid. df Resid. Dev   Test    Df LR stat.        Pr(Chi)
1                                     1       544   748.4121
2 mosmed + depress1 + educat + marital       540   717.0415 1 vs 2     4 31.37067 0.000002572061

> #use AER coeftest and coefci for tests and confidence intervals
> coeftest(model1)

Re-fitting to get Hessian


z test of coefficients:

           Estimate Std. Error z value              Pr(>|z|)
mosmed    0.0063709  0.0136609  0.4664               0.64096
depress1  0.1210496  0.0229848  5.2665 0.000000139036058455
educat   -0.0130156  0.0648567 -0.2007               0.84095
marital   0.2596249  0.1359329  1.9099               0.05614
0|1       1.1495693  0.1976394  5.8165 0.000000006009265638
1|2       1.5843137  0.2025726  7.8210 0.000000000000005242

> coefci(model1)

Re-fitting to get Hessian

                2.5 %     97.5 %
mosmed   -0.020464160 0.03320602
depress1  0.075899142 0.16620014
educat   -0.140417971 0.11438673
marital  -0.007397164 0.52664703

#obtaining the psuedo-R-sq values with modEvA package requires use of glm not polr
> model3=glm(cutmeal ~ mosmed + depress1  + educat + marital,data=d,family=binomial(link="probit"))
> summary(model3)

library(modEvA)
RsqGLM(model=model3)  #model on right side of equal sign is name of my model above

NOTE: Tjur R-squared applies only to binomial GLMs
$`CoxSnell`
[1] 0.04747709

$Nagelkerke
[1] 0.07255096

$McFadden
[1] 0.04578147

$Tjur
[1] NA

$sqPearson
[1] 0.05009599

#can get standardized coefficients with reghelper
> library(reghelper)
> beta(model3, x = TRUE, y = FALSE)
```

Sample Write-Up *(I report only on the ordinal logistic. The probit write-up would be the same except there is no OR and the standardized coefficients would hopefully be reported. I computed the OR by using* e^B*)*

An ordered logit model was estimated to investigate whether months on medical insurance, depression, education, and marital status predict how often meals were cut ("never," "some months," "almost every month"). Together, the predictors accounted for a significant amount of variance in the outcome, likelihood ratio $\chi^2(4) = 31.371$, $p < .001$. Only depression, *B* = .201, *SE* = .039, *OR* = 1.22, $p < .001$, and marital status, *B* = .463, *SE* = .235, *OR* = 1.59, $p = .049$, significantly independently predicted the frequency of cutting meals. Each point increase on the depression scale was associated with about 22% increase in the frequency of cutting meals compared to the lower frequency categories. Married individuals were approximately 50% more likely to have in increase in the frequency of cutting meals compared to the lower categories. Overall the model accounted for approximately 4% of the variance in the outcome, McFadden's pseudo-R^2 = .042.