Examples of Modeling Ordinal and Nominal Outcomes via SAS PROC LOGISTIC

The data for this example come from: http://www.ats.ucla.edu/stat/sas/dae/ologit.htm

SAS Syntax and Output for Data Manipulation:

```
* Import data and center/recode predictors;
DATA work.gradplan; SET lecture8.ologit;
      parentGD = pared;
      GPA3 = GPA-3;
      IF public=1 THEN private=0; ELSE private=1;
      LABEL apply=        "apply: 0=Not, 1=Eh, 2=Very"
            parentGD=     "parentGD: Parent Has Graduate Degree (0=N,1=Y)"
            private=      "private: Student Attends Private University (0=N,1=Y)"
            GPA3=         "GPA3: Student GPA (0=3)";
RUN;

TITLE1 "DESCRIPTIVES FOR STUDY VARIABLES";
PROC MEANS DATA=work.gradplan; VAR GPA; RUN;
PROC FREQ DATA=work.gradplan; TABLE Apply parentGD private; RUN; TITLE1;
```

apply: 0=Not, 1=Eh, 2=Very

APPLY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	220	55.00	220	55.00
1	140	35.00	360	90.00
2	40	10.00	400	100.00

Empty Model Predicting the CUMULATIVE Logit for Apply (treated as ORDINAL):

Submodel 1: $Logit(Apply_p > 0) = \beta_{01}$ → $Probability(Apply_p > 0) = \frac{exp(\beta_{01})}{1+exp(\beta_{01})}$

Submodel 2: $Logit(Apply_p > 1) = \beta_{02}$ → $Probability(Apply_p > 1) = \frac{exp(\beta_{02})}{1+exp(\beta_{02})}$

```
TITLE1 "EMPTY MODEL PREDICTING ORDINAL APPLICATION STATUS";
PROC LOGISTIC DATA=work.gradplan DESCENDING;
* Predicting 3-category apply UP via DESCENDING option;
  MODEL apply = / ITPRINT LINK=CLOGIT;
* Requesting logit intercept to be transformed into probability (add EXP for odds, too);
  ESTIMATE "Intercept" intercept 1 / ILINK CATEGORY=SEPARATE;
RUN;
```

Response Profile

Ordered Value	APPLY	Total Frequency
1	2	40
2	1	140
3	0	220

SAS is trying to help explain what it's doing… see, it's still predicting down, but it re-ordered your data so that up is now down, and down is now up…. That's not confusing at all! This is why we start with an empty model, to make sure we know what SAS is predicting.

Probabilities modeled are cumulated over the lower Ordered Values.

Maximum Likelihood Iteration History

Iter	Ridge	-2 Log L	Intercept_2	Intercept_1
0	0	741.205283	-2.197225	-0.200671

Last Evaluation of Gradient

Intercept_2	Intercept_1
-7.071E-15	6.883383E-15

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

-2 Log L = 741.205
```

Hooray! Our estimates are usable!

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 2	1	-2.1972	0.1667	173.8007	<.0001
Intercept 1	1	-0.2007	0.1005	3.9866	0.0459

Intercept 2 = logit of apply > 1 (so, 01 vs. 2 = submodel 2)

Intercept 1 = logit of apply > 0 (so, 0 vs.12 = submodel 1)

Estimate(s) : Table truncated to save space

Estimate = predicted logit

Mean = predicted probability

Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|	Mean	Standard Error of Mean
Intercept	2	-2.1972	0.1667	-13.18	<.0001	0.1000	0.01500
Intercept	1	-0.2007	0.1005	-2.00	0.0459	0.4500	0.02487

Probability of $Apply_p > 1 \;=\; exp(-2.1972)\,/\,[1 + exp(-2.1972)] \;=\; .100$
Probability of $Apply_p > 0 \;=\; exp(-0.2007)\,/\,[1 + exp(-0.2007)] \;=\; .450$

Probability of $Apply_p = 0$: $\quad 1 - Prob_{int1} = \; 1 - .45 = \; .55$
Probability of $Apply_p = 1$: $\quad Prob_{int1} - Prob_{int2} = .45 - .10 = \; .35$
Probability of $Apply_p = 2$: $\quad Prob_{int2} - 0 \quad = .10 - \; 0 \; = \; .10$

APPLY	Percent
0	55
1	35
2	10

Empty model should recreate these…

So now that we know what each submodel is predicting, we can add some predictors...

Submodel 1: $Logit(Apply_p > 0) = \beta_{01} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p)$

Submodel 2: $Logit(Apply_p > 1) = \beta_{02} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p)$

```
ODS GRAPHICS ON;
TITLE1 "CONDITIONAL MODEL PREDICTING ORDINAL APPLICATION STATUS";
TITLE2 "MAIN EFFECTS ONLY";
PROC LOGISTIC DATA=work.gradplan DESCENDING;
MODEL apply = gpa3 parentGD private / LINK=CLOGIT ITPRINT RSQUARE;
```

Rest of syntax will follow…

```
Probabilities modeled are cumulated over the lower Ordered Values.
```

Maximum Likelihood Iteration History

Iter	Ridge	-2 Log L	Intercept_2	Intercept_1	GPA3	parentGD	private
0	0	741.205283	-2.197225	-0.200671	0	0	0
1	0	717.464963	-2.408487	-0.411933	0.549246	1.052266	0.053785
2	0	717.025529	-2.504783	-0.412957	0.617084	1.043998	0.056955
3	0	717.024872	-2.510079	-0.414639	0.615596	1.047808	0.058544

```
Last Change in -2 Log L    0.0006568145
```

Last Evaluation of Gradient

Intercept_2	Intercept_1	GPA3	parentGD	private
-0.001054534	-0.002299624	0.0015795107	-0.002580072	-0.001586708

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

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
4.8446	3	0.1835

This is the global test of **whether proportional odds holds in these data**, with df = #slopes * *C*−2. Because the test statistic is nonsignificant, we can assume the submodels are the same except for their intercepts.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	745.205	727.025
SC	753.188	746.982
-2 Log L	741.205	717.025

SAS fits the empty model by default—that's what the "intercept only" column is for. But it's still useful to fit the empty model to see what the submodels predict.

R-Square 0.0587 Max-rescaled R-Square 0.0696

Pseudo-R^2 values (use with caution)

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.1804	3	<.0001
Score	23.4804	3	<.0001
Wald	24.3337	3	<.0001

SAS does some of your model comparisons for you, too! The likelihood ratio x^2 is the −2ΔLL (difference in −2LL) for df=3 between the empty and *current* models. Two other versions of the same idea are also provided.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 2	1	-2.5101	0.3103	65.4266	<.0001
Intercept 1	1	-0.4146	0.2732	2.3030	0.1291
GPA3	1	0.6156	0.2626	5.4963	0.0191
parentGD	1	1.0478	0.2684	15.2350	<.0001
private	1	0.0585	0.2886	0.0411	0.8393

Only one version of each predictor effect is provided because this ordinal model assumes proportional odds.

Interpret each effect…

Intercept 2:

Intercept 1:

GPA3:

parentGD:

private:

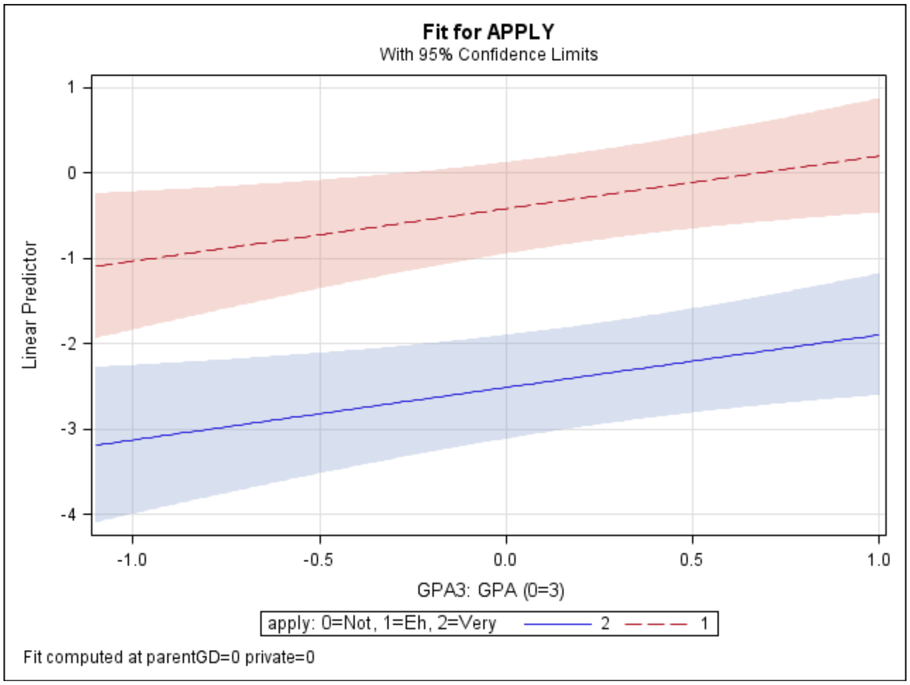
Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
GPA3	1.851	1.106	3.096
parentGD	2.851	1.685	4.826
private	1.060	0.602	1.867

Odds ratios whose confidence intervals do not include 1 are significant (so only the effect of private school is nonsignificant).

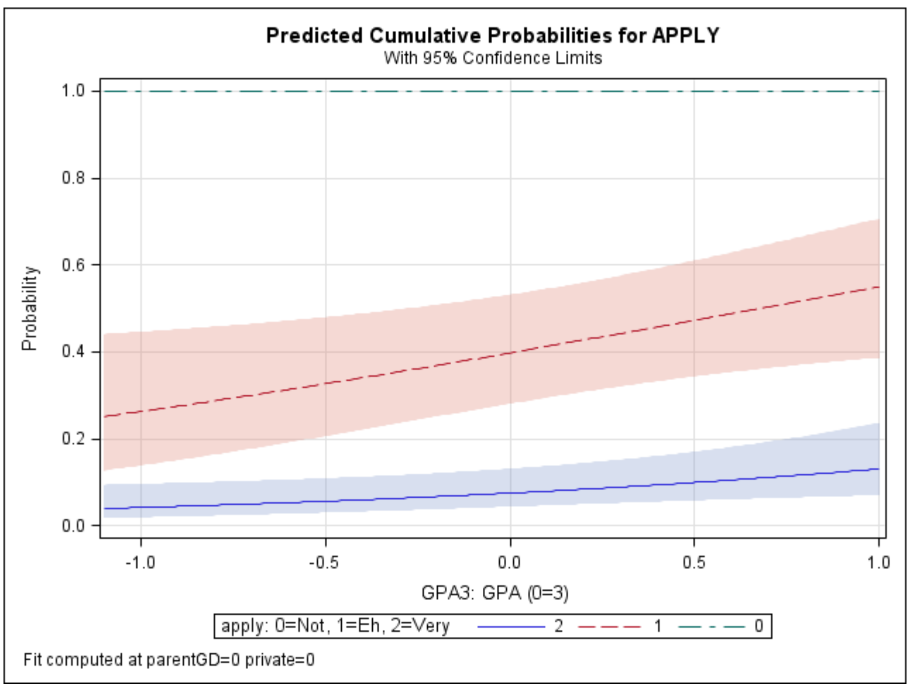
SAS made us some pictures from the EFFECTPLOT statement:
(in which the LINK option plots in logits or and ILINK plots in probability)

```
* Plot to demonstrate proportional odds (use LINK to get logit, ILINK to get probability);
* CLM=confidence limits, NOOBS=don't print people;
 EFFECTPLOT FIT(X=GPA3) / AT (parentGD=0 private=0) LINK CLM NOOBS;
```



The slope of the line for the GPA effect is the same across submodel—this is the “proportional odds” assumption. The lines differ by a constant amount, created by allowing a different intercept for each logit, no matter what GPA is.

Although these plots were created for parentGD =0 and private = 0, setting them at any other values would not change the GPA *slope* (because they don’t interact with GPA). It would change the *intercept*, however.



The effect of GPA on the probability is nonlinear. The lines “shut off” as they approach 0 (or 1, although that doesn’t happen here). That is the magic of the logit link—it prevents Y from going out of bounds!

There are many more EFFECTPLOT options as well (including to plot interactions).

Previously shown model estimates:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept 2	1	-2.5101	0.3103	65.4266	<.0001	beta02
Intercept 1	1	-0.4146	0.2732	2.3030	0.1291	beta01
GPA3	1	0.6156	0.2626	5.4963	0.0191	beta1
parentGD	1	1.0478	0.2684	15.2350	<.0001	beta2
private	1	0.0585	0.2886	0.0411	0.8393	beta3

Only one version of each predictor effect is provided because this ordinal model assumes proportional odds.

Now let's examine some additional requested estimates.

```
* Requesting conditional group means;
 ESTIMATE "Mean for No Grad Parent"   intercept 1 parentGD 0 / ILINK CATEGORY=SEPARATE;
 ESTIMATE "Mean for Yes Grad Parent"  intercept 1 parentGD 1 / ILINK CATEGORY=SEPARATE;

 ESTIMATE "Mean for Public School"    intercept 1 private  0 / ILINK CATEGORY=SEPARATE;
 ESTIMATE "Mean for Private School"   intercept 1 private  1 / ILINK CATEGORY=SEPARATE;

* Example of custom hypothesis test;
 ESTIMATE "Difference of Parent & School Effects" parentGD -1 private 1 /CATEGORY=SEPARATE;
RUN; ODS GRAPHICS OFF;
```

Estimate(s) : Table truncated to save space

Estimate = predicted logit

Mean = predicted probability

Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|	Mean	Error of Mean
Mean for No Grad Parent	2	-2.5101	0.3103	-8.09	<.0001	0.07515	0.02157
Mean for No Grad Parent	1	-0.4146	0.2732	-1.52	0.1291	0.3978	0.06545
Mean for Yes Grad Parent	2	-1.4623	0.3544	-4.13	<.0001	0.1881	0.05413
Mean for Yes Grad Parent	1	0.6332	0.3450	1.84	0.0665	0.6532	0.07815
Mean for Public School	2	-2.5101	0.3103	-8.09	<.0001	0.07515	0.02157
Mean for Public School	1	-0.4146	0.2732	-1.52	0.1291	0.3978	0.06545
Mean for Private School	2	-2.4515	0.1873	-13.09	<.0001	0.07933	0.01368
Mean for Private School	1	-0.3561	0.1177	-3.03	0.0025	0.4119	0.02850
Difference of Parent and School Effects	2	-0.9893	0.3860	-2.56	0.0104		
Difference of Parent and School Effects	1	-0.9893	0.3860	-2.56	0.0104		

Do the *p*-values for the intercepts (conditional means) provide useful information?

For what value of GPA are these means estimated? How would we select a different GPA?

Let's see how to add an interaction, such as between ParentGD and Private School...

Submodel 1: $Logit(Apply_p > 0) =$

$$\beta_{01} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$$

Submodel 2: $Logit(Apply_p > 1) =$

$$\beta_{02} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$$

```
ODS GRAPHICS ON;
TITLE2 "ADDING INTERACTION OF PARENT GRAD BY SCHOOL TYPE";
PROC LOGISTIC DATA=work.gradplan DESCENDING;
* Need to list group predictors on CLASS statement for plots;
  CLASS parentGD(REF='0') private (REF='0');
* Model now includes interactions via |;
  MODEL apply = gpa3 parentGD|private / LINK=CLOGIT ITPRINT RSQUARE;
```

| means "estimate all interactions among these predictors" (can also add @2 to say just all 2-way, @3 for 3-way, etc.

Rest of syntax will follow…

```
Probabilities modeled are cumulated over the lower Ordered Values.

    Class Level Information
                             Design
Class          Value      Variables
parentGD       0                 -1
               1                  1
private        0                 -1
               1                  1
```

By default, predictors on the CLASS statement get effect-coded. But in LOGISTIC you can add a REF option to choose the reference group (this is not the case in other PROCs, though).

```
                                    Maximum Likelihood Iteration History

Iter   Ridge      -2 Log L   Intercept_2   Intercept_1        GPA3   parentGD1     private1    parentGD1
                                                                                                private1
   0       0    741.205283     -2.197225     -0.200671           0           0            0           0
   1       0    716.344326     -1.921537      0.075017    0.555611    0.393527     0.150243    0.213109
   2       0    715.808783     -2.022062      0.075304    0.622824    0.395572     0.135206    0.191637
   3       0    715.807732     -2.023610      0.078070    0.620759    0.399014     0.134752    0.190783

Last Change in -2 Log L     0.0010512366

                                   Last Evaluation of Gradient
                                                                                           parentGD1
 Intercept_2       Intercept_1             GPA3         parentGD1          private1        1private1
-0.000560154      -0.001124494     0.0037301663      -0.001238268      -0.003273631       -0.0059295

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

Score Test for the Proportional Odds Assumption

Chi-Square        DF      Pr > ChiSq
    5.6363         4          0.2280
```

This is the global test of **whether proportional odds holds in these data**, with df = #slopes * *C*−2. Because the test statistic is nonsignificant, we can assume the submodels are the same except for their intercepts.

```
         Model Fit Statistics
                                  Intercept
                    Intercept           and
Criterion                Only    Covariates

AIC                   745.205       727.808
SC                    753.188       751.757
-2 Log L              741.205       715.808
```

SAS fits the empty model by default—that's what the "intercept only" column is for. But it's still useful to fit the empty model to see what the submodels predict.

```
R-Square     0.0615     Max-rescaled R-Square      0.0730
```

Pseudo-R^2 values (use with caution)

```
        Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio        25.3976        4         <.0001
Score                   25.0342        4         <.0001
Wald                    25.0651        4         <.0001
```

The likelihood ratio x^2 is the $-2\Delta LL$ (difference in $-2LL$) for df=4 between the empty and *current* models. Two other versions of the same idea are also provided.

```
             Type 3 Analysis of Effects
                                 Wald
Effect               DF     Chi-Square     Pr > ChiSq
GPA3                  1         5.5795         0.0182
parentGD              1         5.5464         0.0185
private               1         0.6255         0.4290
parentGD*private      1         1.3031         0.2536
```

By putting the binary predictors on the CLASS statement, we get their marginal main effects here (at the mean of other CLASS predictors). Below are their simple main effects at 0 instead.

```
                  Analysis of Maximum Likelihood Estimates
                                             Standard          Wald
Parameter              DF     Estimate        Error    Chi-Square     Pr > ChiSq
Intercept        2      1      -2.0236       0.2157       87.9770         <.0001
Intercept        1      1       0.0781       0.1767        0.1952         0.6586
GPA3                    1       0.6208       0.2628        5.5795         0.0182
parentGD         1      1       0.3990       0.1694        5.5464         0.0185
private          1      1       0.1348       0.1704        0.6255         0.4290
parentGD*private 1 1    1       0.1908       0.1671        1.3031         0.2536
```

Interpret each effect that is now different…

parentGD:

private:

parentGD*private:

Now let's revisit our model to compute predicted values and simple effects for this interaction:

Submodel 1: $Logit(Apply_p > 0) =$

$\beta_{01} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$

Submodel 2: $Logit(Apply_p > 1) =$

$\beta_{02} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$

Predicted values (intercepts or conditional means) come directly from the model equation:

No grad, public = intercept + $\beta_2(0) + \beta_3(0) + \beta_4(0)(0)$

No grad, private = intercept + $\beta_2(0) + \beta_3(1) + \beta_4(0)(1)$

Yes grad, public = intercept + $\beta_2(1) + \beta_3(0) + \beta_4(1)(0)$

Yes grad, private = intercept + $\beta_2(1) + \beta_3(1) + \beta_4(1)(1)$

```
* Requesting conditional group means;
ESTIMATE "Mean for No Grad,  Public School"
intercept 1 parentGD 0 private 0 parentGD*private 0 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for No Grad,  Private School"
intercept 1 parentGD 0 private 1 parentGD*private 0 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for Yes Grad, Public School"
intercept 1 parentGD 1 private 0 parentGD*private 0 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for Yes Grad, Private School"
intercept 1 parentGD 1 private 1 parentGD*private 1 / ILINK CATEGORY=SEPARATE;
```

Estimate = predicted logit

Mean = predicted probability

Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|	Mean	Standard Error of Mean
Mean for No Grad, Public School	2	-2.0236	0.2157	-9.38	<.0001	0.1167	0.02225
Mean for No Grad, Public School	1	0.07807	0.1767	0.44	0.6586	0.5195	0.04410
Mean for No Grad, Private School	2	-1.8889	0.1902	-9.93	<.0001	0.1314	0.02170
Mean for No Grad, Private School	1	0.2128	0.1525	1.40	0.1629	0.5530	0.03770
Mean for Yes Grad, Public School	2	-1.6246	0.3216	-5.05	<.0001	0.1646	0.04422
Mean for Yes Grad, Public School	1	0.4771	0.3064	1.56	0.1194	0.6171	0.07240
Mean for Yes Grad, Private School	2	-1.2991	0.2841	-4.57	<.0001	0.2143	0.04784
Mean for Yes Grad, Private School	1	0.8026	0.2764	2.90	0.0037	0.6905	0.05907

Predicted simple effects, conditional ONLY on interacting predictors:
Step 1: Find original model

Submodel 1: $Logit(Apply_p > 0) =$
$$\beta_{01} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$$
Submodel 2: $Logit(Apply_p > 1) =$
$$\beta_{02} + \beta_1(GPA_p - 3) + \beta_2(ParentGD_p) + \beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$$

Step 2: Find all terms related to <u>the effect of interest</u>, then combine like terms:

Parent effect = $\beta_2(ParentGD_p) + \beta_4(ParentGD_p)(Private_p)$
Parent effect = $[\beta_2 + \beta_4(Private_p)](ParentGD_p)$

School effect = $\beta_3(Private_p) + \beta_4(ParentGD_p)(Private_p)$
School effect = $[\beta_3 + \beta_4(ParentGD_p)](Private_p)$

Step 3: Build ESTIMATE statement out using terms in [ ], in which the variables listed represent the betas, and the numbers represent how many of each beta is needed given the interacting predictor values for which you want the effect

```
* Requesting simple effects per cell;
ESTIMATE "Parent Effect for Public"     parentGD 1 parentGD*private 0 / CATEGORY=SEPARATE;
ESTIMATE "Parent Effect for Private"    parentGD 1 parentGD*private 1 / CATEGORY=SEPARATE;
ESTIMATE "School Effect for No Grad"    private  1 parentGD*private 0 / CATEGORY=SEPARATE;
ESTIMATE "School Effect for No Grad"    private  1 parentGD*private 1 / CATEGORY=SEPARATE;
```

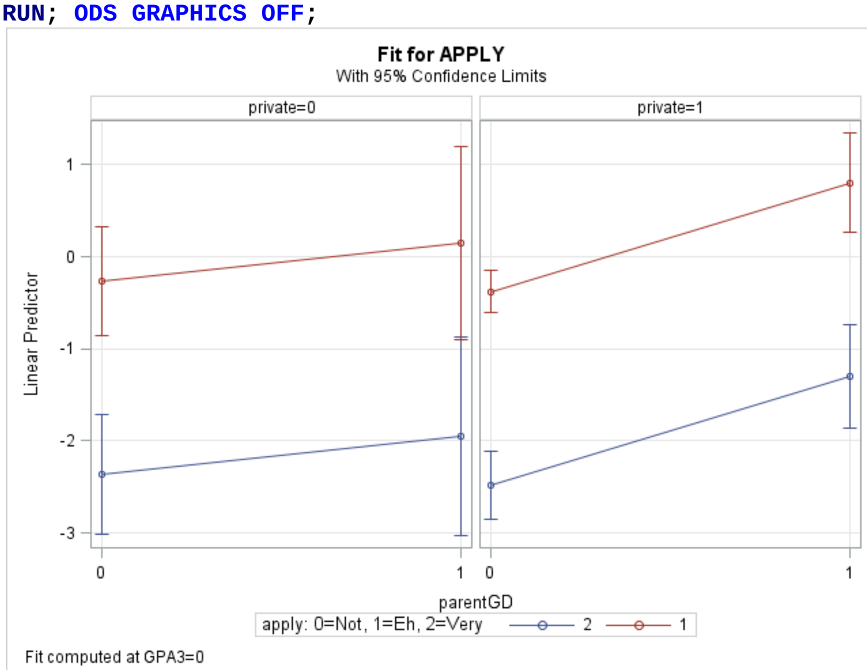
Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|
Parent Effect for Public	2	0.3990	0.1694	2.36	0.0185
Parent Effect for Public	1	0.3990	0.1694	2.36	0.0185
Parent Effect for Private	2	0.5898	0.1487	3.97	<.0001
Parent Effect for Private	1	0.5898	0.1487	3.97	<.0001
School Effect for No Grad	2	0.1348	0.1704	0.79	0.4290
School Effect for No Grad	1	0.1348	0.1704	0.79	0.4290
School Effect for No Grad	2	0.3255	0.2965	1.10	0.2723
School Effect for No Grad	1	0.3255	0.2965	1.10	0.2723

```
         Odds Ratio Estimates
                Point          95% Wald
Effect       Estimate      Confidence Limits
GPA3            1.860       1.111        3.114
```

By default SAS does not provide odds ratios for CLASS variables. But you can request them using an ODDSRATIO statement.

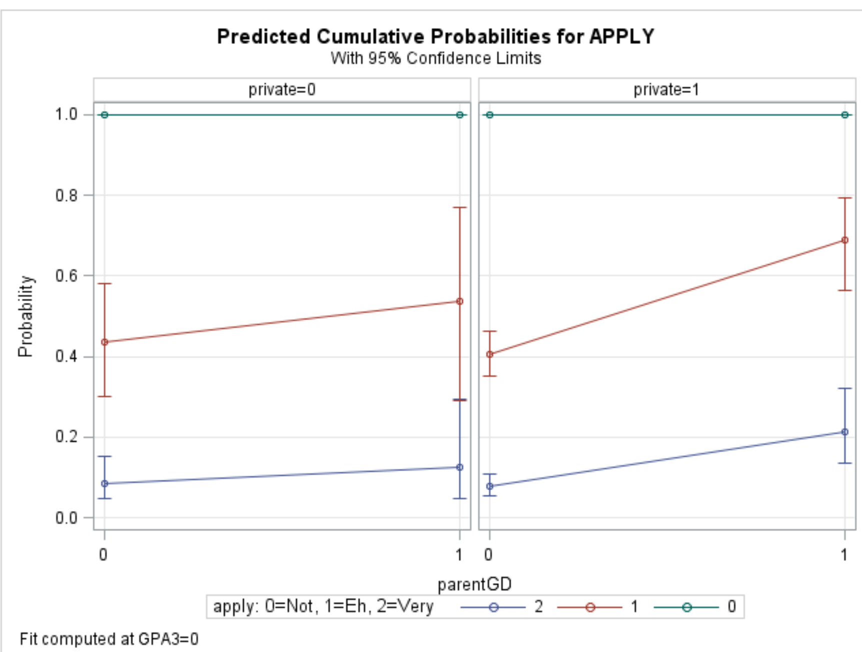
SAS made us some pictures from the EFFECTPLOT statement:
(in which the LINK option plots in logits or and ILINK plots in probability)

```
* Plot to demonstrate interactions (use LINK to get logit, ILINK to get probability);
* CLM=confidence limits, NOOBS=don't print people, AT is for covariate values;
EFFECTPLOT INTERACTION(X=parentGD PLOTBY=private) / AT (GPA3=0) LINK CLM NOOBS;
RUN; ODS GRAPHICS OFF;
```



The slope of the red and blue line for the ParentGD effect is the same across submodel—this is still the "proportional odds" assumption. The lines differ by a constant amount, created by allowing a different intercept for each logit submodel, no matter what ParentGD is.

The separate panels show the interaction, in which the effect of ParentGD is allowed to differ between public and private school students. The finding that the interaction is nonsignificant means the slopes are equivalent across panels, though.



This plot shows the effect of ParentGD on the probability for each submodel. It may be the case that the interaction looks stronger or weaker when plotted in terms of probabilities, but ultimately that is not relevant, as the interaction is evaluated on the logit scale, not the probability scale.

Given that the interaction is nonsignificant, we'll remove it from the model when demonstrating how to fit these same data to a generalized logit model for nominal responses, up next.

Empty Model Predicting the GENERALIZED Logit for Apply, <u>with 1 as the Reference Group</u>: (treated as NOMINAL now)

Submodel 0 if $Y_p = 0{,}1$: $Logit(Apply_p = 0) = \beta_{00}$ → $Probability(Apply_p = 0) = \frac{exp(\beta_{00})}{1+exp(\beta_{00})}$

Submodel 2 if $Y_p = 1{,}2$: $Logit(Apply_p = 2) = \beta_{02}$ → $Probability(Apply_p = 2) = \frac{exp(\beta_{02})}{1+exp(\beta_{02})}$

```
TITLE1 "EMPTY MODEL PREDICTING NOMINAL APPLICATION STATUS";
TITLE2 "1 AS REFERENCE GROUP, SO PREDICTING 1 VS 0, 1 VS 2";
PROC LOGISTIC DATA=work.gradplan;
  MODEL apply (REF='1')= / ITPRINT LINK=GLOGIT;
  * Requesting logit intercept to be transformed into probability (add EXP for odds, too);
  ESTIMATE "Intercept" intercept 1 / ILINK CATEGORY=SEPARATE; RUN;
```

Response Profile

Ordered Value	APPLY	Total Frequency
1	0	220
2	1	140
3	2	40

Logits modeled use APPLY=1 as the reference category.

Maximum Likelihood Iteration History

Iter	Ridge	-2 Log L	Intercept_0	Intercept_2
0	0	741.205283	0.451985	-1.252763

Last Evaluation of Gradient

Intercept_0	Intercept_2
4.996004E-14	-1.24345E-14

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

-2 Log L = 741.205

Analysis of Maximum Likelihood Estimates

Parameter	APPLY	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	0	1	0.4520	0.1081	17.4782	<.0001
Intercept	2	1	-1.2528	0.1793	48.8262	<.0001

Estimate(s) : Table truncated to save space

Estimate = predicted logit

Mean = predicted probability

Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|	Mean	Standard Error of Mean
Intercept	0	0.4520	0.1081	4.18	<.0001	0.5500	0.02487
Intercept	1	-1.2528	0.1793	-6.99	<.0001	0.1000	0.01500

Probability of $Yp = 0 = exp(-0.4520) / [1 + exp(-0.4520)] = .55$
Probability of $Yp = 2 = exp(-1.2528) / [1 + exp(-1.2528)] = .10$

Probability of $Y_p = 1$: $1 - Prob_{int1} - Prob_{int2} = .55 - .10 = .35$

APPLY	Percent
0	55
1	35
2	10

Note that SAS has changed its labeling of submodel 2 from “2” to “1” in the estimates table…

So now that we know what each submodel is predicting, we can add some predictors. Only this time, in the nominal model all predictor effects are estimated separately.

Submodel 0 if $Y_p = 0{,}1$:

$$Logit(Apply_p = 0) = \beta_{00} + \beta_{10}(GPA_p - 3) + \beta_{20}(ParentGD_p) + \beta_{30}(Private_p)$$

Submodel 2 if $Y_p = 1{,}2$:

$$Logit(Apply_p = 2) = \beta_{02} + \beta_{12}(GPA_p - 3) + \beta_{22}(ParentGD_p) + \beta_{32}(Private_p)$$

```
TITLE1 "CONDITIONAL MODEL PREDICTING NOMINAL APPLICATION STATUS";
TITLE2 "MAIN EFFECTS ONLY";
PROC LOGISTIC DATA=work.gradplan;
MODEL apply (REF='1') = gpa3 parentGD private / LINK=GLOGIT ITPRINT RSQUARE;
```

```
Logits modeled use APPLY=1 as the reference category.

                                     Maximum Likelihood Iteration History

Iter   Ridge       -2 Log L    Intercept_0    Intercept_2       GPA3_0       GPA3_2    parentGD_0   parentGD_2
  0        0     741.205283       0.451985      -1.252763            0            0             0            0
  1        0     715.419935       0.931890      -0.550967    -0.424655     0.438741     -0.874083     0.674152
  2        0     713.999132       0.950122      -0.751264    -0.449346     0.467087     -0.956242     0.436003
  3        0     713.993963       0.951534      -0.763990    -0.448751     0.475229     -0.951626     0.422600

 Iter      private_0        private_2
    0              0                0
    1      -0.399643        -0.941695
    2      -0.417219        -0.789161
    3      -0.418826        -0.778942

Last Change in -2 Log L     0.0051682727

                                        Last Evaluation of Gradient
Intercept_0   Intercept_2       GPA3_0       GPA3_2     parentGD_0     parentGD_2     private_0    private_2
0.0004101876 -0.001351514 -0.000075469 0.000039673  0.0001459101  -0.000961818 0.0001873177 0.000748594

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

         Model Fit Statistics
                                  Intercept
                 Intercept              and
Criterion             Only       Covariates
AIC                745.205          729.994
SC                 753.188          761.926
-2 Log L           741.205          713.994

R-Square     0.0658     Max-rescaled R-Square     0.0780

        Testing Global Null Hypothesis: BETA=0
Test                     Chi-Square       DF     Pr > ChiSq
Likelihood Ratio            27.2113        6         0.0001
Score                       28.2058        6         <.0001
Wald                        25.8399        6         0.0002
```

SAS fits the empty model by default—that's what the "intercept only" column is for. But it's still useful to fit the empty model to see what the submodels predict.

Pseudo-R^2 values (use with caution)

SAS does some of your model comparisons for you, too! The likelihood ratio x^2 is the $-2\Delta LL$ (difference in $-2LL$) for df=3 between the empty and *current* models. Two other versions of the same idea are also provided.

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
GPA3	2	4.9539	0.0840
parentGD	2	13.8522	0.0010
private	2	3.0371	0.2190

This is the omnibus test of whether each effect is significant *combining* across submodels. Effects per submodel are then given below.

Now every parameter is unique by submodel.

Analysis of Maximum Likelihood Estimates

Parameter	APPLY	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept	0	1	0.9515	0.3258	8.5285	0.0035	beta00
Intercept	2	1	-0.7640	0.4511	2.8684	0.0903	beta02
GPA3	0	1	-0.4488	0.2902	2.3911	0.1220	beta10
GPA3	2	1	0.4752	0.4871	0.9517	0.3293	beta12
parentGD	0	1	-0.9516	0.3171	9.0082	0.0027	beta20
parentGD	2	1	0.4226	0.4083	1.0715	0.3006	beta22
private	0	1	-0.4188	0.3433	1.4884	0.2225	beta30
private	2	1	-0.7789	0.4706	2.7398	0.0979	beta32

What are the estimates labeled APPLY 0 predicting?

What are the estimates labeled APPLY 2 predicting?

Odds Ratio Estimates

Effect	APPLY	Point Estimate	95% Wald Confidence Limits	
GPA3	0	0.638	0.361	1.128
GPA3	2	1.608	0.619	4.179
parentGD	0	0.386	0.207	0.719
parentGD	2	1.526	0.686	3.397
private	0	0.658	0.336	1.289
private	2	0.459	0.182	1.154

Odds ratios whose confidence intervals do not include 1 are significant.

Here's where our choice of 1 as the reference group comes in handy—we can now test proportional odds *for each slope*, which says that the slopes should be the same across ordinal submodels. In our case, submodel 0 is predicting 0 instead of 1 (down), whereas submodel 2 is predicting 2 instead of 1 (up). So if the assumption of proportional odds holds, then the slopes for a given predictor should be the same estimate across submodels, just opposite in sign. The TEST statements below refer to each submodel slope (using SAS's internal name) and ask if it is equal to −1* the other submodel's slope.

```
* Test proportional odds assumption per slope;
  GPA3:      TEST -GPA3_0     = GPA3_2;
  parentGD:  TEST -parentGD_0 = parentGD_2;
  private:   TEST -private_0  = private_2;
```

Linear Hypotheses Testing Results

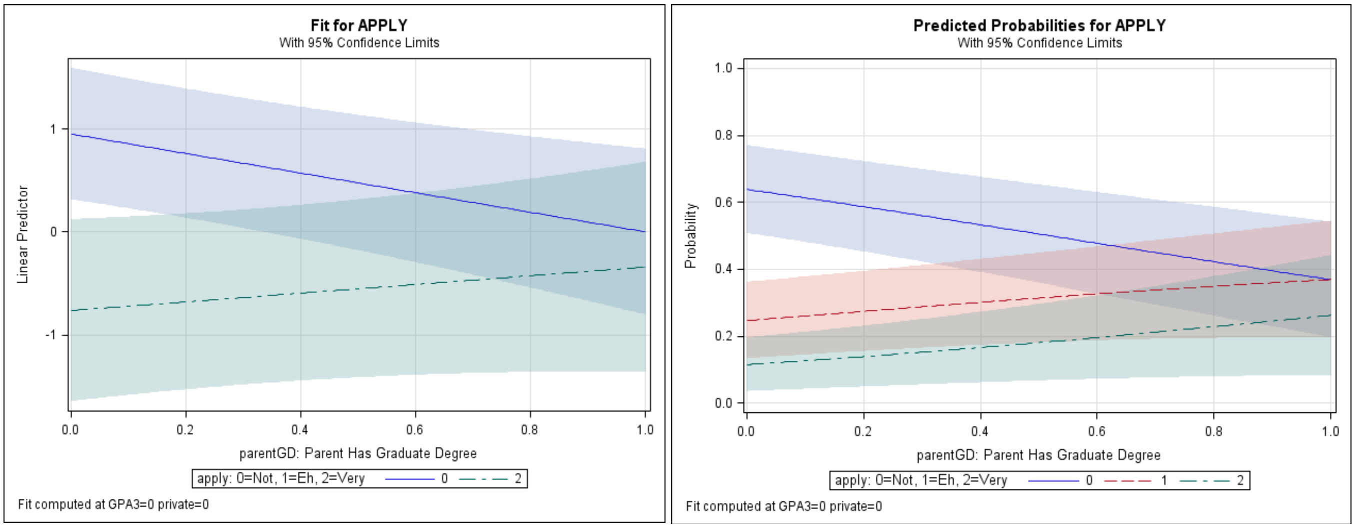
Label	Wald Chi-Square	DF	Pr > ChiSq
GPA3	0.0017	1	0.9673
parentGD	0.7857	1	0.3754
private	2.9767	1	0.0845

GPA3's slopes of 0.4488 predicting from 0 to 1 and 0.4752 predicting from 1 to 2 are not different.

ParentGD slopes of 0.9516 predicting from 0 to 1 and 0.4226 predicting from 1 to 2 are not different.

Private slopes of 0.4188 predicting from 0 to 1 and −0.7789 predicting from 1 to 2 are almost different.

```
* Plot to see non-proportional odds (use LINK to get logit, ILINK to get probability);
* CLM=confidence limits, NOOBS=don't print people;
  EFFECTPLOT FIT(X=parentGD) / AT (private=0 GPA3=0) LINK CLM NOOBS;
```



Even though the effect of parentGD in predicting "not likely" to apply to grad school rather than "somewhat likely" is significantly negative, it is not different from the nonsignificant positive effect of parent GD in predicting "very likely" rather than "somewhat likely" to apply.

```
* Requesting conditional group means;
ESTIMATE "Mean for No Grad Parent"   intercept 1 parentGD 0 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for Yes Grad Parent"  intercept 1 parentGD 1 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for Public School"    intercept 1 private  0 / ILINK CATEGORY=SEPARATE;
ESTIMATE "Mean for Private School"   intercept 1 private  1 / ILINK CATEGORY=SEPARATE;

* Example of custom hypothesis test;
ESTIMATE "Difference of Parent & School Effects" parentGD -1 private 1 /CATEGORY=SEPARATE;
RUN;
```

Estimate(s) : Table truncated to save space

Estimate = predicted logit

Mean = predicted probability

Label	apply: 0=Not, 1=Eh, 2=Very	Estimate	Standard Error	z Value	Pr > |z|	Mean	Standard Error of Mean
Mean for No Grad Parent	0	0.9515	0.3258	2.92	0.0035	0.6386	0.06687
Mean for No Grad Parent	1	-0.7640	0.4511	-1.69	0.0903	0.1149	0.04087
Mean for Yes Grad Parent	0	-0.00009	0.4111	-0.00	0.9998	0.3689	0.08788
Mean for Yes Grad Parent	1	-0.3414	0.5184	-0.66	0.5102	0.2622	0.09140
Mean for Public School	0	0.9515	0.3258	2.92	0.0035	0.6386	0.06687
Mean for Public School	1	-0.7640	0.4511	-1.69	0.0903	0.1149	0.04087
Mean for Private School	0	0.5327	0.1261	4.23	<.0001	0.5839	0.02887
Mean for Private School	1	-1.5429	0.2322	-6.64	<.0001	0.07327	0.01487
Difference of Parent and School Effects	0	0.5328	0.4568	1.17	0.2435		
Difference of Parent and School Effects	1	-1.2015	0.6105	-1.97	0.0491		