Logit, Probit & Ordinal Models

Lecture 1

Overview and Assumptions

Soc 512
Applied Multivariate Statistics

David Peters, Ph.D.
Department of Sociology
Iowa State University
OVERVIEW AND ASSUMPTIONS

OVERVIEW OF LOGISTIC REGRESSION

David J. Peters, Ph.D.
OVERVIEW

Logistic regression overview ...

- Technique that allows one to predict a categorical or ordinal DV from a set of predictors.
- Similar in purpose to OLS except DVs can be non-continuous.
- DV can be single or multiple discrete responses.
- IVs can be of any measurement scale.
- No assumption of normality of any variables.
- No assumption of homoscedasticity.
- No assumption of linearity between DV and IV, but assumes linearity between logits and IVs.
- Uses maximum likelihood (ML) rather than OLS estimation.
OVERVIEW

Types of research questions ...

Prediction of group membership or outcomes.

- Interested in which IVs drive a classification process or a given outcome.
- Mainly interested in effects of the IVs on sorting cases into groups or predicting the probability of an outcome for cases.

Analysis of a categorical or ordinal dependent variable.

- Allows analysis of discrete choices or categories that cannot be done using OLS.
- Very useful in social sciences with social groups (categories) or attitudes (ordinal).
- Employed extensively in economics and psychology, less so in sociology.
**Overview**

Types of research questions ...

Assess classification of cases into groups.

- Allows one to assess the accuracy of some classification scheme.
- Provides estimates of overall accuracy, and identifies poorly fitted cases and misclassifications.
- Permits one to improve the classification model by dropping IVs that do not contribute to classification, and assessing the effects of new predictors.
- Very useful in validating a cluster analysis solution.
OVERVIEW

Types of logistic regression ...

Binomial Logistic and Probit Regression

- Traditional logistic/probit regression with a 0-1 DV. Uses only two categories or one response (1) with a reference/null category (0).
- Very common in all social sciences. Typically referred to as just logit or probit regression.

Multinomial Logistic Regression

- Expands binomial to include more than two categories (1,2,3) or more than one response variable (0,1,2). However, a reference category for comparison MUST be specified.
- If no meaningful reference, then one must run multiple regressions with different reference categories.
In general ... 

Because logistic regression uses MLE rather than OLS, it avoids many of the typical assumptions tested in statistical analysis.

Does not assume normality of variables (both DV and IVs).
Does not assume linearity between DV and IVs.
Does not assume homoscedasticity.
Does not assume normal errors.

MLE allows more flexibility in the data and analysis because it has fewer restrictions.
ASSUMPTIONS

Meaningful DV categories.

- Must have a meaningful reference/base category, especially in multinomial models. If not, one must run all pair-wise contrasts.
- Categories must represent discrete units that are mutually exclusive and exhaustive.
- Must have logically ordered categories.
- ADVICE – always code reference category as zero.

Correct model specification.

- Like all techniques, assumes all relevant IVs and included and all irrelevant ones excluded.
- Based on theory and previous research.
ASSUMPTIONS

No outliers.

- Use typical techniques to identify and remove outliers.
- Use z-scores, histograms, and k-means clustering.
- Can also analyze residuals to identify outliers in the regression. Common practice is to examine standardized residuals to identify cases where the logit/probit model fit poorly.
- Create scatterplots of residual by DV; and residual by CaseID.

Types of Residuals ...

- Standardized Residual – typical regression residual.
- Pearson Chi-Square Residual – GFI residual (binomial only).
- Deviance Chi-Square Residual – GFI residual (binomial only).
**ASSUMPTIONS**

Adequate cell sample sizes.

- Adequate cell N is major concern in logit and probit regression.
- At MINIMUM have as many N and IVs in each cell of the DV; and have no missing data in the IVs.
- **Rule of 5** – Best to have at least 5 cases per IV in the model in each category of the DV.
- Check for large standard errors in the logit coefficients. If present one needs more N in certain cells.
- If model does not converge, small cell N may be the cause.
ASSUMPTIONS

No multicollinearity.

• Check zero-order correlation matrix for high values (ie \( r > 0.7 \)) and confirm with partial-order correl for offenders.

• If found, create interaction term (center the IVs first) or drop one of the IVs.

• Multicollinearity a problem if large standard errors are found in the logit coefficients.

• Relates to correct model specification assumption ... each IV should contribute something unique based on theory.
**ASSUMPTIONS**

**Linearity between logits/probits and IVs.**

- Although linearity between DV and IVs is not assumed, linearity between the logits/probits and the IVs are assumed.

**Methods to Assess ...**
- Plots of logits/probits by each IV.
- Log Cross-Products Test.

**Plots of Logits/Probits by IVs ...**

- Calculate the logit \((L_i)\) or probit \((I_i)\) for each predicted probability \((P_i)\) of the categories of the DV.
- Plot each logit/probit by each IV. Plots should show a linear scatter, indicating assumption is met.
- Traditional method to assess linearity.

\[
L_i = I_i = \log \left( \frac{P_i}{1 - P_i} \right)
\]
ASSUMPTIONS

Linearity between logits/probits and IVs.

Log Cross-Products Test ...

- For each IV ($X_i$), calculate the log cross-products of the variable by $X_{LPi} = (X_i \times \log(X_i))$.
- Run simple logit/probit regression that includes all IVs ($X_i$) and all log cross-products ($X_{LPi}$).
- Significant $X_{LP}$ indicates non-linearity between logits and that IV.
- However this is not a formal test, only an empirical rule.

Remedial measures ...

- Transform or dropping the IV is common, but no real guidance in literature ... use your judgment.
- Suggested that one only checks and reports violations, not transform or drop.
Independent errors.

- Like OLS, error terms are assumed uncorrelated.
- Typically occurs over time or space for secondary data. Can also occur with primary data that includes non-random measurement errors.
- No way to test for this in logit/probit regression. Must be assessed as part of research design and knowledge of the data. If you have data over time or spatial units, you likely have non-independent errors.
- If non-independent errors are suspected, this may indicate the need for multilevel models or other correction methods.
EXAMPLES AND SYNTAX

BASIC STEPS OF LOGIT/PROBIT REGRESSION
LOGIT/PROBIT REGRESSION - BASIC STEPS

Test ex ante assumptions …

- Meaning DV categories.
- Correct model specification.
- No outliers.
- Adequate cell size.
- No multicollinearity.
- Independent errors.

Chose appropriate CDF or link function …

Logit CDF – Used for binomial models.

Generalized Logit CDF – Used for multinomial models, but can also be used for binomial models as well.

Probit CDF – Used for binomial models under a normal CDF.
LOGIT/PROBIT REGRESSION – BASIC STEPS

Assess model fit …

• Interpret information criteria.
• Interpret Null Model Test.
• Interpret GFI tests.
• Interpret pseudo $R^2$.
• Calculate and interpret residuals.

Assess model parameters …

• Interpret logit/probit coefficients and their significance.
• Interpret odds-ratios and their significance (logit only).

Assess model classification …

• Calculate and inspect classification table.
• Interpret association coefficients.
Test ex post assumptions ...

• Linearity between logits/probits and IVs.

Additional information ...

• Correct reference category must be specified. If none exists, all comparisons must be made.
• Multinomial models must use the generalized logit CDF, although it can be used with binomial models as well.
• Multinomial probit models cannot be estimated.
• Multinomial models generate multiple predicted probabilities, multiple residuals, and multiple logits.
EXAMPLES AND SYNTAX

SOFTWARE AND SYNTAX
LOGIT/PROBIT REGRESSION - SOFTWARE

SAS or SPSS?

Both SAS and SPSS offer the same types of discreet choice models.

SAS advantages over SPSS…
• Probit models easier to estimate and better diagnostics.
• Automatically produces association measures.
• SAS has one procedure to do all discreet choice models, while SPSS has three separate ones (LOGISTIC, PROBIT, NOMREG).

SPSS advantages over SAS …
• Automatically produces classification tables.
• Automatically calculates residuals.
• Produces -2LL Ratio Tests to see if adding an IV improved model fit compared to a model excluding it.
• Easier to use interface and better charts and plots.
LOGISTIC REGRESSION VARIABLES pov01
   /METHOD=ENTER MINORTY NOHSD COLLEGE UNEMP
   /SAVE=PRED RESID LRESID ZRESID DEV
   /CLASSPLOT
   /PRINT=GOODFIT CI(95)
   /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

NOMREG pov01 (BASE=0 ORDER=ASCENDING) WITH MINORTY NOHSD COLLEGE UNEMP
   /CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
       LCONVERGE(0) PCONVERGE(0.000001) SINGULAR(0.00000001)
   /MODEL
   /STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE)
       ENTRYMETHOD(LR) REMOVALMETHOD(LR)
   /INTERCEPT=INCLUDE
   /PRINT=ASSOCIATION CLASSTABLE FIT PARAMETER SUMMARY LRT CPS STEP MFI
       IC
   /SAVE ESTPROB PREDCAT PCPROB ACPROB.
SPSS - Multinomial Models …

NOMREG pov123 (BASE=2 ORDER=ASCENDING) WITH MINORTY NOHSD COLLEGE
UNEMP
/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
   LCONVERGE(0) PCONVERGE(0.000001) SINGULAR(0.00000001)
/MODEL
/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE)
   ENTRYMETHOD(LR) REMOVALMETHOD(LR)
/INTERCEPT=INCLUDE
/PRINT=ASSOCIATION CLASSTABLE FIT PARAMETER SUMMARY LRT CPS STEP MFI
   IC
/SAVE ESTPROB PREDCAT PCPROB ACXPROB.
SAS – Binomial Models ...

proc logistic data=povcat;
class pov01 (param=ref ref='0');
model pov01 = minorty nohsd college unemp
   / link=logit aggregate scale=none rsquare expb stb lackfit
clodds=wald;
output out=pov01_logit predprobs=individual reschi=reschi
       resdev=resdev;

proc logistic data=povcat;
class pov01 (param=ref ref='0');
model pov01 = minorty nohsd college unemp
   / link=probit aggregate scale=none rsquare expb stb lackfit
clodds=wald;
output out=pov01_probit predprobs=individual reschi=reschi
       resdev=resdev;

proc freq data=pov01_logit;
tables _from_*_into_ / cumcol nocol;
SAS - Binomial Models ...

data plots; set pov01_logit;
   keep caseid pov01 _from_ _into_ ip_0 ip_1 reschi resdev cat0 cat1
       resid0 resid1 s_resid0 s_resid1;
   if (pov01 = 0) then cat0=1;
      else if (pov01 = 1) then cat0=0;
   if (pov01 = 1) then cat1=1;
      else if (pov01 = 0) then cat1=0;
   resid0 = (cat0 - ip_0);
   resid1 = (cat1 - ip_1);
   s_resid0 = (resid0 / (sqrt(ip_0*(1-ip_0))));
   s_resid1 = (resid1 / (sqrt(ip_1*(1-ip_1))));
SAS - Multinomial Models ...

```
proc logistic data=povcat;
class pov123 (param=ref ref='2');
model pov123 = minority nohsd college unemp
   / link=glogit aggregate scale=none rsquare expb stb lackfit
clodds=wald;
output out=pov123_logit predprobs=individual reschi=reschi
   resdev=resdev;

proc freq data=pov123_logit;
tables _from_*_into_ / cumcol nocol;
```
SAS - Multinominal Models ...

data plots; set pov123_logit;
    keep caseid pov123 _from_ _into_ ip_1 ip_2 ip_3 cat1 cat2 cat3 resid1 resid2 resid3 s_resid1 s_resid2 s_resid3;
    if (pov123 = 1) then cat1=1;
        else if (pov123 = 2 or 3) then cat1=0;
    if (pov123 = 2) then cat2=1;
        else if (pov123 = 1 or 3) then cat2=0;
    if (pov123 = 3) then cat3=1;
        else if (pov123 = 1 or 2) then cat3=0;
    resid1 = (cat1 - ip_1);
    resid2 = (cat2 - ip_2);
    resid3 = (cat3 - ip_3);
    s_resid1 = (resid1 / (sqrt(ip_1*(1-ip_1))));
    s_resid2 = (resid2 / (sqrt(ip_2*(1-ip_2))));
    s_resid3 = (resid3 / (sqrt(ip_3*(1-ip_3))));

Logit, Probit, Ordinal – Lecture 6
David J. Peters, Ph.D.
**LOGIT/PROBIT REGRESSION - SOFTWARE**

**SAS - Ordinal Models ...**

```sas
proc logistic data=povcat descending;
class pov123;
model pov123 = minority nohsd college unemp
   / link=cumlogit aggregate scale=none rsquare expb stb lackfit
clodds=wald;
output out=pov123_cumlogit predprobs=individual reschi=reschi
   resdev=resdev;

proc logistic data=povcat descending;
class pov123;
model pov123 = minority nohsd college unemp
   / link=cumprobit aggregate scale=none rsquare expb stb lackfit
clodds=wald;
output out=pov123_cumprobit predprobs=individual reschi=reschi
   resdev=resdev;

proc freq data=pov123_cumlogit;
tables _from_*_into_ / cumcol nocol;
```

**Logit, Probit, Ordinal – Lecture 6**

**Slide 14**

David J. Peters, Ph.D.
data plots; set pov123_cumlogit;
  keep caseid pov123 _from_ _into_ ip_1 ip_2 ip_3 cat1 cat2 cat3
       resid1 resid2 resid3 s_resid1 s_resid2 s_resid3;
  if (pov123 = 1) then cat1=1;
    else if (pov123 = 2 or 3) then cat1=0;
  if (pov123 = 2) then cat2=1;
    else if (pov123 = 1 or 3) then cat2=0;
  if (pov123 = 3) then cat3=1;
    else if (pov123 = 1 or 2) then cat3=0;
  resid1 = (cat1 - ip_1);
  resid2 = (cat2 - ip_2);
  resid3 = (cat3 - ip_3);
  s_resid1 = (resid1 / (sqrt(ip_1*(1-ip_1))));
  s_resid2 = (resid2 / (sqrt(ip_2*(1-ip_2))));
  s_resid3 = (resid3 / (sqrt(ip_3*(1-ip_3))));
EXAMPLE ... Persistently high and low poverty clusters.

GOAL: To determine which demographic factors influence whether a county is in the persistently high poverty or persistently low poverty cluster.

pov_3states.sas7bdat
- CASEID – county identifier.
- POV99 – percent 0-99 percent of poverty.
- MINORTY – percent non-white or Hispanic.
- NOHSD – percent without a high school education.
- COLLEGE – percent enrolled in 4-year colleges/universities.
- UNEMP – unemployment rate.
- POVO1 – binomial variable indicating high poverty (1) or not (0).
- POV123 – multinomial variable indicating low poverty (1), high poverty (3), and neither high or low poverty (2).
ASSESSING MODEL PARAMETERS
Assessing model parameters and predictors ...

For all logit and probit models ...

- Logit and Probit Coefficients
- Standardized Logit and Probit Coefficients

For logit models only ...

- Odds-Ratios or exp(b)
- $\chi^2$ Test of Predictors
Logits and Probits …

Logits and probits are the model parameters estimated by the MLE process. These parameters represent the implied COV generated by the model.

In other words, the parameters are what replicated the observed data/COV to produce $\Sigma(\theta)$, where the parameters represent $\theta$.

Logits and probits traditionally denoted as $b$ (beta), but technically they should be denoted as $\gamma$ (gamma) since they are ML estimators, not least squares ones.

However, in practice most disciplines and journals refer to them as $b$. 
**ASSESSING PARAMETERS**

**Logits and Probits ...**

Logits represent the change in the logistic CDF of the DV given a one unit change in the predictor / IV.

Recall that logits are defined as the log of the odds-ratio. Thus logits represent the change in the log odds-ratio of the DV.

\[
L_i = F^{-1}(p) \quad F(p) = \left( \frac{e^z}{1 + e^z} \right)
\]

Where …

\[ Z = b_0 + b_1X_{1i} + \ldots + b_kX_{ki} \]

\[ e = 2.71828 \quad dZ = \text{normalized values of } Z \]

Probits represent the change in the normal CDF of the DV given a one unit change in the predictor.

Probits are defined as the normal distribution of the predicted probabilities. Thus probits represent the change in the normalized probabilities of the DV.

\[
P_i = \Phi^{-1}(p) \quad \Phi(p) = (2\pi)^{-1/2} \int_{-\infty}^{Z} e\left(\frac{-Z^2}{2}\right)dZ
\]
ASSESSING PARAMETERS

Logits and Probits ...

Since logits and probits follow a general $\chi^2$ distribution, one can test whether the logits/probits are greater than zero.

However, logit and probit values themselves cannot directly be used as a test statistic because of scale differences.

Logits and probits are transformed and tested using the Wald’s $\chi^2$ statistic, which is the square of the logit/probit over the square of the parameter’s standard error.

$\chi^2_w$ is compared to a $\chi^2$ normal distribution with df equal to the rank of the linear hypotheses coefficient matrix ($L$). In terms of testing logits and probits, the rank or df is always equal to one.

$$\chi^2_{wi} = \left( \frac{b_i^2}{\sigma_i^2} \right) \quad df = r(L_i) = 1$$
ASSESSING PARAMETERS

Logits and Probits ...

Interpretation ...

• Larger values of the logits/probits indicates stronger effect on the CDF. Analogous to b in OLS regression.

• Walds $\chi^2$ tests the $H_0$ that the effect of the logits/probits on the CDF is zero. One wished to reject $H_0$ and find a significant $p$-value. Analogous to t-tests in OLS regression.

• Significant logits/probits indicates the predictor or IV contributes to explaining some outcome on the DV.

• However, it is difficult to substantively interpret logits/probits because the CDF scale is not a meaningful scale. This issue is solved with Odds-Ratios.
ASSESSING PARAMETERS

Standardized Logits and Probits …

To remove the effect of different scales of the IVs, one can standardize the logits/probits to the standard deviation of the predictor.

Allows easier interpretation of logits/probits. Most appropriate measure to assess the relative contribution of each predictor to explaining some outcome of the DV.

Interpretation ...

• Interpreted as the change in the logit or probit CDF given a one standard deviation unit change in the predictor/IV.
• Note differences in DV standardization for logit (sL) and probit (sP) models.
ASSESSING PARAMETERS

Odds-Ratios (Logistic Only) …

Interpreting changes in logits is difficult because log-odds are not in a meaningful scale.

Logits can be converted into odds-ratios to assess the relative “strength” of the IVs in affecting the probability of some outcome of the DV occurring.

In logistic regression, odds-ratios are the preferred measure for interpreting how predictors affect outcomes on the DV. Use odds-ratios for substantive interpretation of logistic models.

Odds-ratios are denoted as $\exp(b)$ or $\psi$ (psi). They are calculated by taking the logistic CDF constant to the power of the logit, controlling for units of change.
ASSESSING PARAMETERS

Odds-Ratios (Logistic Only) ...

\[ \psi_i = \left[ e^{b_i} \right]^c \]

Where …
\[ e = 2.71828 \quad b_i = \text{logit coefficient} \]
\[ c = \text{units of change, default is 1} \]

When \( c = 1 \), \( \psi \) is interpreted as the change in odds of the event occurring (DV moving from 0 to 1) given a one-unit change in the predictor.

When \( c = 10 \), \( \psi \) interpreted as the change in odds of the event occurring given a 10-unit change in the IV.

Changing units (c) is useful when IV scales are large.

For example …
- Effect of MHHI on odds more meaningful in $1000s than in $1s.
- Effect of UNEMP rate more meaningful in 0.10% than in 1%.
ASSESSING PARAMETERS

Odds-Ratios (Logistic Only) ...

Interpretation ...

Given a \( c \) unit change in the IV ...

- \( \psi = 1 \) indicates no change in odds of the event occurring.
- \( \psi > 1 \) indicates increase in odds of the event occurring.
- \( \psi < 1 \) indicates decrease in odds of the event occurring.

Significance can be assessed by examining the 95% confidence intervals of the odds-ratio estimate. Tests \( H_0 \) that there is no change in odds given a unit change in the IV ... or \( H_0: \psi = 1 \).

If the 95% CI includes 1, then fail to reject \( H_0 \) and conclude there is no significant change in odds.

Otherwise, reject \( H_0 \) and conclude there is a significant change in odds.
Odds-Ratios (Logistic Only) ...

Interpretation ...

Odds-ratios can be more easily interpreted by converting them into a percent change in odds.

Interpreted as the percent increase or decrease in the odds of an event occurring (DV moving from 0 to 1) given a one-unit (or c-unit) change in the IV.

\[ \psi_{\Delta i} = (\psi_i - 1) \times 100 \]

Limitations ...

- \( \Psi \) has an asymmetric scale, where the lowest limit is bounded by zero and the upper limit is \( +\infty \).
- Interpret excessively high values with caution. May indicate need to adjust \( \psi \) by units (c).
ASSESSING PARAMETERS

Odds-Ratios (Logistic Only) ...

Interpretation for binomial models...

• One unit change in the IV leads to a $\psi$ change in the odds of being in the outcome category ($y=1$) versus not being in the outcome category ($y=0$).

Interpretation for multinomial models...

• One unit change in the IV leads to a $\psi$ change in the odds of being in the current category ($y=j$) versus being in the reference category ($y=r$).
**ASSESSING PARAMETERS**

\( \chi^2 \) Test of Predictors (Multinomial Logistic Only) …

Tests whether significant mean differences exist across response level of the DV. MANOVA-type test based on T3SS.

In binomial logistic models this test is identical to the Walds \( \chi^2 \) test of parameters because there is only one response level (0-1).

In multinomial logistic models it tests for a difference in any pair of logits across all response levels of the DV.

Tests H0 of no difference in logits across \( j - 1 \) response levels. HA is at least one of the logits is different across levels. Compare to \( \chi^2 \) distribution with \( (j - 1) \) df with \( j \) being number of levels.

\[
\chi_{wi}^2 = \left( \frac{b_i^2}{\sigma_i^2} \right) \quad df = r(L) = (j - 1)
\]
ASSESSING PARAMETERS

χ² Test of Predictors (Multinomial Logistic Only) ...

Interpretation ...

- Not usually interpreted. Better to use b, b*, and ψ to interpret parameter significance.
- If interpreted, one wishes to find larger χ²W values and reject H0 of no differences across logits by finding a significant $p$-value. This indicates a significant difference in logits across levels of the DV.