Logistic Regression

Continued

Psy 524

Ainsworth

Equations

- Regression Equation

$$\overline{Y}_i = \frac{e^{A+B_1X_1+B_2X_2+B_3X_3}}{1+e^{A+B_1X_1+B_2X_2+B_3X_3}}$$

Equations

- The linear part of the logistic regression equation is used to find the probability of being in a category based on the combination of predictors
- Programs like SPSS and SAS separate discrete predictors with more than two levels into multiple dichotomous variables.

Equations

- Fall (0 = no, 1 = yes); Difficulty is continuous; season (1 = autumn, 2 = winter, 3 = spring)

Fall	Difficulty	Season
1	3	1
1	1	1
0	1	3
1	2	3
1	3	2
0	2	2
0	1	2
1	3	1
1	2	3
1	2	1
0	2	2
0	2	3
1	3	2
1	2	2
0	3	1

Equations

- Season is a discrete variable with three levels that would be turned into 2 separate variables season 1 and season 2.
- Season 1 is coded 1 for autumn and 0 otherwise; season 2 is coded 1 if winter and 0 otherwise; spring is coded when both are 0.

Fall	Difficulty	Season	Season1	Season2
1	3	1	1	0
1	1	1	1	0
0	1	3	0	0
1	2	3	0	0
1	3	2	0	1
0	2	2	0	1
0	1	2	0	1
1	3	1	1	0
1	2	3	0	0
1	2	1	1	0
0	2	2	0	1
0	2	3	0	0
1	3	2	0	1
1	2	2	0	1
0	3	1	1	0

Interpreting coefficients

- Good news – regression coefficients and their standard errors are found through advanced calculus methods of maximum likelihood (e.g. derivatives, etc.), so we're not getting into it.

Interpreting coefficientss

- Each coefficient is evaluated using a Wald test (really just a Z-test)

$$W_j = \frac{B_j}{SE_{B_j}}$$

Interpreting coefficients

Term	Coefficient	Standard Error	Wald Test (Z)
(Constant)	-1.776	1.89	-0.88
Difficulty	1.01	0.9	1.27
Season (1)	0.927	1.59	0.34
Season (2)	-0.418	1.39	-0.09

Interpreting coefficients

- The tests of the coefficients are approximate z-scores so they are tested as z-scores. None of the coefficients are significant in the sample data.
- The coefficients are placed into the model like in regular multiple regression in order to predict individual subjects' probabilities.

Goodness of fit

- Log-likelihood

$$\log - likelihood = \sum_{i=1}^{N} [Y_i \ln(\hat{Y}_i) + (1 - Y_i) \ln(1 - \hat{Y}_i)]$$

Goodness of fit

- Models are compared by taking 2 times the difference between the models log-likelihoods.

$\chi^2 = 2[(\text{log-likelihood for bigger model}) - (\text{log-likelihood for smaller model})]$

Note: models must be nested in order to be compared. Nested means that all components of the smaller model must be in the larger model.

Goodness of fit

- Often a model with intercept and predictors is compared to an intercept only model to test whether the predictors add over and above the intercept only. This is usually noted as $\chi 2=2[LL(B)-LL(0)]$

Loglikelihood for intercept only model (Yhat = $e^{.405}/1 + e^{.405}$)

Fall	Yhat	1-Yhat	Y*lnYhat	(1-Y)*(1-Yhat)	Σ[YlnYhat + (1 – Y)ln(1 – Yhat)
1	.60	.40	-.51	0	-.51
1	.60	.40	-.51	0	-.51
0	.60	.40	0	-.92	-.92
1	.60	.40	-.51	0	-.51
1	.60	.40	-.51	0	-.51
0	.60	.40	0	-.92	-.92
0	.60	.40	0	-.92	-.92
1	.60	.40	-.51	0	-.51
1	.60	.40	-.51	0	-.51
1	.60	.40	-.51	0	-.51
0	.60	.40	0	-.92	-.92
0	.60	.40	0	-.92	-.92
1	.60	.40	-.51	0	-.51
1	.60	.40	-.51	0	-.51
0	.60	.40	0	-.92	-.92
				Sum=	-10.11

Loglikelihood for intercept only model (Yhat = $e^{-1.776 + seas1(.927) + seas2(-.418)}$ / $1 + e^{-1.776 + seas1(.927) + seas2(-.418)}$)

Fall	Yhat	1-Yhat	Y*lnYhat	(1-Y)*(1-Yhat)	Σ[YlnYhat + (1 – Y)ln(1 – Yhat)
1	.899	0.101	-0.106	0	-0.106
1	.540	0.460	-0.616	0	-0.616
0	.317	0.683	0	-0.381	-0.381
1	.561	0.439	-0.578	0	-0.578
1	.698	0.302	-0.360	0	-0.360
0	.457	0.543	0	-0.611	-0.611
0	.234	0.766	0	-0.267	-0.267
1	.899	0.101	-0.106	0	-0.106
1	.561	0.439	-0.578	0	-0.578
1	.764	0.236	-0.269	0	-0.269
0	.457	0.543	0	-0.611	-0.611
0	.561	0.439	0	-0.823	-0.823
1	.698	0.302	-0.360	0	-0.360
1	.457	0.543	-0.783	0	-0.783
0	.899	0.101	0	-2.293	-2.293
				Sum=	-8.74

Goodness of Fit

- 2[-8.74 - (-10.11)] = 2.74
- the constant only model has one degree of freedom (for the constant) and the full model has 4 degrees of freedom (1 for the constant, and one for each predictor), the DF for the test is 4 – 1 = 3. The test of the chi-square is not significant at 3 DFs so the null is retained.
- Models with different numbers of predictors (nested) can also be compared in the same fashion.

Standardized Residuals

- Given a model you can calculate the standardized residual of each persons predicted probability (using the rather scary matrix formula on page 527)
- You can have SPSS save the standardized residuals and once this is done you can analyze them to see if any are above 3.3 and if they are the subject is an outlier according to the given model.

Types of Logistic Regression

- Direct or Simultaneous
- Sequential or User defined
- Stepwise or Statistical
- Probit vs. Logistic
  - Logistic assumes a categorical (qualitative) underlying distribution
  - Probit assumes a normal distribution and uses Z-scores to estimate the proportion under the curve.
  - Near .5 the analyses are similar they only differ at the extremes.

Inferential Tests

- Assessing goodness of fit for the model
  - There are many goodness of fit indices, so you need to keep in mind what is being compared to know whether a significant difference is good or not. Some tests significance means fit and others significance means lack of fit.

Inferential Tests

- Also consider sample sized when evaluating goodness of fit. Chi-square statistics are heavily influenced by sample size so that with a very large sample even minute differences will be significant.
  - If the sample size is large and the chi-square is significant this may not be important
  - Though if there is significance and the sample is relatively small than the effect is notable.

Inferential Tests

- Constant only vs. full model – here you want there to be a significant improvement to the prediction when all of the predictors are added to the model.
- Perfect model vs. proposed model – some programs test the proposed model against a perfect model (one that predicts perfectly) in this case you want the chi-square to be non-significant.

Inferential Tests

- Deciles of risk
  - Step 1: Subjects are ordered on there predicted probability
  - Step 2: Subjects are divided into 10 groups based on the probabilities (all subjects with .1or lower in lowest decile, .9 or higher in the highest decile, etc.)
  - Step 3: Divide subjects into groups according to their actual outcome (e.g. fall or no fall) creating a 2 X 10 matrix of observed frequencies for the example data.
  - Step 4: Expected frequencies are calculated and the observed frequencies are compared to the expected frequencies in a chi-square test. Fit is indicated by a non-significant chi-square.
  - In SPSS this is given by the Hosmer-Lemeshow test.

Test of individual predictors

- The Wald test is usually used to assess the significance of prediction of each predictor
- The Wald test is known to be overly conservative (increased type II error) and when a predictor is multinomial it does not give a test of the whole predictor but only the dummy coded versions of the predictor.

Number and type of outcomes

- Logistic regression with more than two outcome categories
  - If the response are ordered polytomous than k – 1 equations are made (k being the number of categories) which predicts the probability that a case is above a given category.
    - Defines thresholds – point in the data that separates category one form two, two from three, etc.
    - Calculates the probability that a person passes a given threshold
    - This is done for all categories except the last because the probability of being in a category above the highest is zero.

Number and type of outcomes

- If the responses are non-ordered multinomial than again k – 1 equations are created but the equations are predicting whether a person belongs to a category or not. An equation is made for all categories except the last.
  - SPSS ordinal (plum) is used for ordered polytomous and SPSS multinomial (nomreg) is used for un-ordered multinomial data.

Strength of association (pseudo R-square)

- There are several measures intended to mimic the R-squared analysis, but none of them are an R-squared. The interpretation is not the same, but they can be interpreted as an approximate variance in the outcome accounted for by the

Strength of association (pseudo R-square)

- McFadden's

$$\rho^2 = 1 - \frac{LL(B)}{LL(0)}$$

this value tends to be smaller than R-square and values of .2 to .4 are considered highly satisfactory.

Strength of association (pseudo R-square)

- Cox and Snell is also based on log-likelihood but it takes the sample size into account:

$$R^2_{CS} = 1 - \exp\left[-\frac{2}{n}[LL(B) - LL(0)]\right]$$

but it cannot reach a maximum of 1 like we would like so…

Strength of association (pseudo R-square)

- The Nagelkerke measure adjusts the C and S measure for the maximum value so that 1 can be achieved:

$$R_N^2 = \frac{R_{CS}^2}{R_{MAX}^2}, \text{where } R_{MAX}^2 = 1 - \exp[2(n^{-1})LL(0)]$$