Logistic Regression, Part III:
Hypothesis Testing, Comparisons to OLS

[This handout steals heavily from the SPSS Advanced Statistics User Guide. Also, Linear probability, logit, and probit models, by John Aldrich and Forrest Nelson, paper # 45 in the Sage series on Quantitative Applications in the Social Sciences; and Applied Logistic Regression Analysis Second Edition by Scott Menard, paper # 106 in that series.]

WARNING: As Menard more or less points out, notation is wildly inconsistent across authors and programs when it comes to Logistic regression. I'm trying to more or less follow Menard, but you'll have to learn to adapt to whatever the author or statistical program happens to use.

Overview. In this handout, we'll examine hypothesis testing in logistic regression and make comparisons between logistic regression and OLS. We'll use both SPSS and Stata in this handout. A separate handout provides more detail about using Stata. The optional appendix to this handout provides more detail on how some of the key calculations are done.

There are a number of logical analogs between OLS and Logistic regression, i.e. the math is different but the functions served are similar. I will summarize these first, and then explain each of them in more detail:

OLS Regression	Logical Analog in Logistic Regression
Total Sums of Squares	$-2LL_0$, DEV_0, D_0
Error/ Residual Sums of Squares	$-2LL_M$, DEV_M, D_M
Regression/Explained Sums of Squares	Model Chi Square, L^2, G_M
Global F	Model Chi Square, L^2, G_M
Incremental F Test	Chi-Square Contrast/ Incremental chi-square contrast
Incremental F Test and Wald test of the same hypotheses give identical results	Chi-square contrast between models and a Wald test of the same hypotheses generally do NOT give exactly identical results.

Using the same data as before, here is part of the output we get in SPSS when we do a logistic regression of Grade on Gpa, Tuce and Psi.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	15.404	3	.002
	Block	15.404	3	.002
	Model	15.404	3	.002

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	25.779	.382	.528

The more or less corresponding output from Stata is

```
. use http://www.nd.edu/~rwilliam/stats2/statafiles/logist.dta, clear
. logit  grade gpa tuce psi

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.496795
Iteration 2:   log likelihood = -12.929188
Iteration 3:   log likelihood = -12.889941
Iteration 4:   log likelihood = -12.889633
Iteration 5:   log likelihood = -12.889633

Logistic regression                               Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740
[Rest of output deleted]
```

Global tests of parameters. In OLS regression, if we wanted to test the hypothesis that all β's = 0 versus the alternative that at least one did not, we used a global F test. In logistic regression, we use a *likelihood ratio chi-square test* instead. SPSS refers to this as the Model chi-square while Stata calls it LR chi2. The value is 15.404. This is computed by contrasting a model which has no independent variables (i.e. has the constant only) with a model that does. Following is a general description of how it works.

The probability of the observed results given the parameter estimates is known as the *likelihood*. Since the likelihood is a small number less than 1, it is customary to use -2 times the log of the likelihood. -2LL (25.779 in the SPSS output; Stata reports LL which is -12.889633) is a measure of how well the estimated model fits the likelihood. A good model is one that results in a high likelihood of the observed results. This translates to a small number for -2LL (If a model fits perfectly, the likelihood is 1, and -2 times the log likelihood is 0).

-2LL is also called the Deviance, DEV, or simply D. Subscripts are often used to denote which model this particular deviance applies to. The smaller the deviance is, the better the model fits the data.

The "initial log likelihood function" is for a model in which only the constant is included. This is used as the baseline against which models with IVs are assessed. SPSS used to report it, but for some reason doesn't anymore (unless you request a lot of other boring optional output). It is easy to compute though; for any particular model, just add the Model chi-square + the -2 Log likelihood for the model, i.e. $-2LL_0 = 15.404 + 25.779 = 41.183$. (Stata reports LL_0, -20.59173, which is the log likelihood for iteration 0.)

$-2LL_0$, DEV_0, or simply D_0 are alternative ways of referring to the deviance for a model which has only the intercept. This is analogous to the Total Sums of Squares, SST, in OLS Regression.

When GPA, PSI, and TUCE are in the model, -2LL = 25.779. We can refer to this as DEV_M or simply D_M.

The -2LL for a model, or DEV_M, indicates the extent to which the model fails to perfectly predict the values of the DV, i.e. it tells how much improvement is needed before the predictors provide the best possible prediction of the dependent variable. DEV_M is analogous to the Error Sums of Squares, SSE, in OLS regression.

The addition of these 3 parameters reduces -2LL by 15.404, i.e.
$DEV_0 - DEV_M = 41.183 – 25.779 = 15.404$. This is reflected in the *Model Chi-square,* which Stata labels as LR chi2.

The Model Chi-Square, also called Model L^2 or G_M, is analogous to the Regression (explained) Sums of Squares, SSR, in OLS regression. It is also the direct counterpart to the Global F Test in regression analysis. A significant value tells you that one or more betas differ from zero, but it doesn't tell you which ones.

$$G_M = L^2 = DEV_0 - DEV_M$$

The significance level for the model chi-square indicates that this is a very large drop in chi-square, ergo we reject the null hypothesis. The effect of at least one of the IVs likely differs from zero.

You can think of the Deviance as telling you how bad the model still is, while the Model L^2 , aka G_M tells you how good it is.

Incremental Tests / Likelihood Ratio Chi-Square Tests. There is also an analog to the incremental F test. Just like with OLS, we can compare constrained and unconstrained models. We use an incremental chi-square square statistic instead of an incremental F statistic. (More commonly, you see phrases like chi-square contrasts.) The difference between the deviances of constrained and unconstrained models has a chi-square distribution with degrees of freedom equal to the number of constraints.

Incremental chi-square test/ chi-square contrast (analog to incremental F test)

L^2 = $DEV_{Constrained}$ - $DEV_{Unconstrained}$, d.f. = number of constraints

If the resulting chi-square value is significant, stick with the unconstrained model; if insignificant then the constraints can be justified. Alternatively, you'll get the same results using

L^2 = Model $L^2_{Unconstrained}$ – Model $L^2_{Constrained}$, d.f. = number of constraints

The notation L^2 is used to signify that this is a Likelihood Ratio Chi Square test (as opposed to, say, a Pearson Chi-Square test, which has less desirable properties). Again, notation is wildly inconsistent across authors. G^2 is another notation sometime used.

WARNING: In OLS, an incremental F test and a Wald test give you the same results. In logistic regression, a chi-square contrast between models and a Wald test generally do NOT give identical results. LR chi-square contrasts are considered better but in large samples it may not matter much.

Nested Models-SPSS [Read this on your own]. SPSS reports something called step, block and model chi-squares. These are a little confusing at first, but they are basically just different incremental and global LR chi-square tests. (As is its custom, SPSS is being internally inconsistent, and does not present results in the same format as it does for similar analyses with OLS regression.)

Note that, in the above, the Step, Block, and Model Chi-squares are all the same. This is because we entered all three variables at the same time. (This is similar to regression; if we enter all the variables at once, the F statistic and the F change statistic are one and the same, because we are moving from no variables to all the variables.) If, instead, we enter the variables sequentially, SPSS will do the incremental chi-square tests for us. The following command enters GPA, TUCE and PSI separately.

```
LOGISTIC REGRESSION VAR=grade
  /METHOD=ENTER gpa / enter tuce / enter psi
  /CRITERIA PIN(.05) POUT(.10) ITERATE(20) CUT(.5)
  /casewise pred resid sresid dfbeta outlier .
```

Here is the model information after GPA is entered.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	8.766	1	.003
	Block	8.766	1	.003
	Model	8.766	1	.003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	32.418	.240	.331

The inclusion of GPA reduces the deviance by 8.766. Since we have gone from 0 variables to 1 variable, step, block and model chi-squares are all the same, i.e. all three test H_0: $\beta_{GPA} = 0$. But, see what happens when we add TUCE:

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.435	1	.510
	Block	.435	1	.510
	Model	9.200	2	.010

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	31.983	.250	.345

The model chi-square, which once again is the equivalent of a global F, tells you that the effect of GPA or TUCE differs from 0. However, the Block chi-square tells you that the effects of variables entered in this block (in this case, TUCE only) do not significantly differ from 0. In other words, Step and Block both test H_0: $\beta_{TUCE} = 0$; but Model tests H_0: $\beta_{GPA} = \beta_{TUCE} = 0$

Hence, in SPSS, if you specify a hierarchy of models, the Block chi-square tells you whether any of the effects of the variables in that block significantly differ from 0. It is the logical equivalent of the F change statistic in regression.

Here is what we get when PSI is added in the final step.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6.204	1	.013
	Block	6.204	1	.013
	Model	15.404	3	.002

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	25.779	.382	.528

Here, Step and Block both test H_0: $\beta_{PSI} = 0$; but Model tests H_0: $\beta_{GPA} = \beta_{TUCE} = \beta_{PSI} = 0$. So, the block chi-square statistic of 6.204 tells us the effect of psi is statistically significant. Note too that the model chi-square jumps from 9.2 in block 2 to 15.404 in Block 3, a jump of 6.204. So, the block chi-square tells you how much the model chi-square would decline if Psi were dropped from the model. When doing model contrasts, you can use either the deviances or the model chi-squares; either way the incremental chi-square values will be the same.

Nested Models-Stata. In Stata, we can get incremental and global LR chi-square tests easily by using the `nestreg` command. We should include the `lr` option so we get likelihood ratio tests rather than Wald tests. The `quietly` option suppresses a lot of the intermediate information, but don't use it if you want to see those results.

```
. nestreg, lr quietly: logit  grade gpa tuce psi

Block  1: gpa
Block  2: tuce
Block  3: psi

  +----------------------------------------------------------------+
  | Block |        LL       LR     df  Pr > LR       AIC       BIC |
  |-------+--------------------------------------------------------|
  |     1 |  -16.2089     8.77      1   0.0031   36.4178  39.34928 |
  |     2 | -15.99148     0.43      1   0.5096  37.98296  42.38017 |
  |     3 | -12.88963     6.20      1   0.0127  33.77927  39.64221 |
  +----------------------------------------------------------------+
```

With Stata, you can also use the `lrtest` command to do likelihood ratio contrasts between models, e.g.

```
. quietly logit  grade gpa
. est store m1
. quietly logit  grade gpa tuce
. est store m2
. quietly logit  grade gpa tuce psi
. est store m3
. lrtest m1 m2

Likelihood-ratio test                                 LR chi2(1)  =      0.43
(Assumption: m1 nested in m2)                         Prob > chi2 =    0.5096

. lrtest m2 m3

Likelihood-ratio test                                 LR chi2(1)  =      6.20
(Assumption: m2 nested in m3)                         Prob > chi2 =    0.0127
```

Stepwise Logistic Regression-SPSS [Read this on your own]. Finally, note that the output also includes a step chi-square, which, in this case, is always the same as the block chi-square. If you were doing stepwise regression, however, the results would be different. Here, I tell SPSS to enter the variables using forward stepwise selection, dividing the variables into two blocks.

```
LOGISTIC REGRESSION VAR=grade
  /METHOD=fstep psi  / fstep gpa tuce
  /CRITERIA PIN(.50) POUT(.10) ITERATE(20) CUT(.5).
```

The default PIN value is .05, but I changed it to .5 so the insignificant TUCE would make it in. In the first block, psi alone gets entered.

Block 1: Method = Forward Stepwise (Conditional)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5.842	1	.016
	Block	5.842	1	.016
	Model	5.842	1	.016

Then, in the next block, the forward selection procedure causes gpa to get entered first, then tuce.

Block 2: Method = Forward Stepwise (Conditional)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	9.088	1	.003
	Block	9.088	1	.003
	Model	14.930	2	.001
Step 2	Step	.474	1	.491
	Block	9.562	2	.008
	Model	15.404	3	.002

So, looking at the final entries,

- The step chi-square, .474, tells you whether the effect of the variable that was entered in the final step, TUCE, significantly differs from zero. It is the equivalent of an incremental F test of the parameter, i.e. it tests H_0: $\beta_{TUCE} = 0$.
- The block chi-square, 9.562, tests whether either or both of the variables included in this block (GPA and TUCE) have effects that differ from zero. This is the equivalent of an incremental F test, i.e. it tests H_0: $\beta_{GPA} = \beta_{TUCE} = 0$.
- The model chi-square, 15.404, tells you whether any of the three IVs has significant effects. It is the equivalent of a global F test, i.e. it tests H_0: $\beta_{GPA} = \beta_{TUCE} = \beta_{PSI} = 0$.

Stepwise Logistic Regression-Stata. As with other Stata commands, you can use the `sw` prefix for stepwise regression. We can add the `lr` option so that likelihood-ratio, rather than Wald, tests are used when deciding the variables to enter next.

```
. sw, lr pe(.05) : logit  grade  gpa tuce psi

LR test               begin with empty model
p = 0.0031 <  0.0500  adding  gpa
p = 0.0130 <  0.0500  adding  psi

Logistic regression                               Number of obs   =         32
                                                  LR chi2(2)      =      14.93
                                                  Prob > chi2     =     0.0006
Log likelihood = -13.126573                       Pseudo R2       =     0.3625

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   3.063368    1.22285     2.51   0.012     .6666251     5.46011
         psi |   2.337776   1.040784     2.25   0.025     .2978755    4.377676
       _cons |  -11.60157   4.212904    -2.75   0.006    -19.85871   -3.344425
------------------------------------------------------------------------------
```

Tests of Individual Parameters. Testing whether any individual parameter equals zero proceeds pretty much the same way as in OLS regression. You can, if you want, do an incremental LR chi-square test. That, in fact, is the best way to do it, since the Wald test referred to next is biased under certain situations.

Or, SPSS reports the Wald statistic, which is $(b/s_b)^2$. When parameters are tested separately, we see that the effects of GPA and PSI are statistically significant, but the effect of TUCE is not. As we saw before,

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	GPA	2.826	1.263	5.007	1	.025	16.872
	TUCE	.095	.142	.452	1	.502	1.100
	PSI	2.378	1.064	4.992	1	.025	10.786
	Constant	-13.019	4.930	6.972	1	.008	.000

a. Variable(s) entered on step 1: PSI.

Note that the Wald statistic is less than the Block chi-square was (6.204) when we entered PSI last. The block chi-square should be considered the more accurate.

Stata reports basically the same information, except that it reports z values, which is b/s_b.

```
. logit  grade gpa tuce psi, nolog

Logistic regression                               Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   2.826113   1.262941     2.24   0.025     .3507938    5.301432
        tuce |   .0951577   .1415542     0.67   0.501    -.1822835    .3725988
         psi |   2.378688   1.064564     2.23   0.025       .29218    4.465195
       _cons |  -13.02135   4.931325    -2.64   0.008    -22.68657    -3.35613
------------------------------------------------------------------------------
```

With Stata, you can also continue to use the `test` command. The `test` command does Wald tests, which aren't as good as LR tests but which may be adequate in large samples, e.g.

```
. * Test whether effects of gpa and tuce are both 0
. test gpa tuce

 ( 1)  gpa = 0
 ( 2)  tuce = 0

           chi2(  2) =    6.35
         Prob > chi2 =    0.0418

. * Test whether effects of gpa and psi are equal
. test gpa = psi

 ( 1)  gpa - psi = 0

           chi2(  1) =    0.11
         Prob > chi2 =    0.7437
```

R^2 Analogs. As Menard points out in Applied Logistic Regression Analysis, Second Edition, several people have tried to come up with the equivalent of an R^2 measure for logistic regression. No one of these measures seems to have achieved widespread acceptance yet. One of the simplest and most popular formulas is

$$\text{Pseudo } R^2 = \text{Model } L^2/\ DEV_0 = 1 - DEV_M/DEV_0 = 1 - LL_M/LL_0$$

where, as you'll recall, DEV_0 (or $-2LL_0$) pertains to the baseline model with intercept only. (Menard refers to this as R^2_L; it is also called McFadden R^2 ; Stata just calls it Pseudo R^2. Be careful when reading, since the term Pseudo R^2 gets applied to a lot of different statistics.) This statistic will equal zero if all coefficients are zero. It will come close to 1 if the model is very good. In the present case, for the model with gpa, psi and tuce included,

$$\text{Pseudo } R^2 = \text{Model } L^2/\ DEV_0 = 15.404/41.183 = .374$$

For some weird reason, SPSS reports Pseudo R^2 in some of its related routines (NOMREG and PLUM) but not in Logistic Regression. Menard (p. 27) argues for the Pseudo R^2 statistic on the grounds that it is conceptually closest to OLS R^2 i.e. it reflects a proportionate reduction in the quantity actually being minimized, -2LL. However, as I explain in my categorical data class, you can make a logical case for most of the Pseudo R^2 measures.

The Cox and Snell statistic that SPSS reports seems to come fairly close to pseudo R^2 . The Cox-Snell R^2 uses a formula that is equivalent to

$$\text{Cox-Snell } R^2 = 1 - \exp(-\text{Model } L^2/N)$$

In this case, for the model with all 3 IVs included,

$$\text{Cox-Snell } R^2 = 1 - \exp(-\text{Model } L^2/N) = 1 - \exp(-15.404/32) = 1 - \exp(-.481375) = .382$$

The other R^2 reported by SPSS is an adjustment for Cox-Snell.

Other ways of assessing "Goodness of Fit." There are other ways to assess whether or not the model fits the data. For example, there is the *classification table*:

Classification Table[a]

			Predicted		
			GRADE		Percentage Correct
	Observed		.00	1.00	
Step 1	GRADE	.00	18	3	85.7
		1.00	3	8	72.7
	Overall Percentage				81.3

a. The cut value is .500

In the classification table, cases with probabilities ≥ .50 are predicted as having the event, other cases are predicted as not having the event. Ideally, you would like to see the two groups have very different estimated probabilities. In this case, of the 21 people who did not get A's, the model correctly predicted 18 would not but said that 3 would. Similarly, of the 11 who got A's, the model was right on 8 of them.

From the classification table, you can't tell how great the errors are. The 6 misclassified cases may have been within one or two percentage points of being classified correctly, or they may have been way off. For "rare" events, I am not sure how useful the table is. A 10% probability may be relatively high, but still not high enough to get the case classified as a 1 (e.g. there may be only 1 chance in a 1000 of the average 20 year old dying within the year; identifying those for whom the odds are 1 in 10 of dying may be quite useful.) In the latest versions of SPSS, you can make the cutoff point higher or lower than .50, which may make the classification table a little more useful. Menard goes on at some length about other possible classification/prediction strategies.

The equivalent command in Stata is `estat class` (you can also just use `lstat`)

```
. quietly logit  grade gpa tuce psi
. estat class

Logistic model for grade

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |         8             3  |         11
     -     |         3            18  |         21
-----------+--------------------------+-----------
   Total   |        11            21  |         32

Classified + if predicted Pr(D) >= .5
True D defined as grade != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   72.73%
Specificity                     Pr( -|~D)   85.71%
Positive predictive value       Pr( D| +)   72.73%
Negative predictive value       Pr(~D| -)   85.71%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   14.29%
False - rate for true D         Pr( -| D)   27.27%
False + rate for classified +   Pr(~D| +)   27.27%
False - rate for classified -   Pr( D| -)   14.29%
--------------------------------------------------
Correctly classified                        81.25%
--------------------------------------------------
```

Diagnostics. It can also be useful to run various diagnostics. These help to indicate areas or cases for which the model is not working well. For example, the command

```
LOGISTIC REGRESSION VAR=grade
  /METHOD=ENTER gpa psi tuce
  /CRITERIA PIN(.05) POUT(.10) ITERATE(20) CUT(.5)
  /casewise pred resid sresid dfbeta outlier .
```

will, for outlying cases, print (a) the predicted probability (b) the residual (observed value - predicted probability) (c) the studentized residual (d) the change in model parameters if the case is omitted. The OUTLIER parameter limits the printout to those cases with studentized residuals greater than or equal to 2 in magnitude. In a large sample you might want to be more selective, so you might change this to something like OUTLIER(3). If you want to get a listing for all the cases, just leave the OUTLIER parameter off. In this case, the output is

Casewise List[b]

Case	Selected Status[a]	Observed GRADE	Predicted	Temporary Variable Resid	SResid	DFB0	DFB1	DFB2	DFB3
2	S	1**	.111	.889	2.248	4.546	-1.058	.081	-.040
27	S	0**	.852	-.852	-2.070	2.958	-.471	-.633	-.057

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2.000 are listed.

Note that the greatest errors occur with cases 2 and 27. A closer examination of these outliers might give us ideas on how to improve the model.

Both Menard and the SPSS manual list other statistics for looking at residuals. Menard also briefly discusses some graphical techniques that can be useful. Also see Hamilton's Statistics with Stata for some ideas.

In Stata, you can again use the `predict` command to compute various outliers. As was the case with OLS, Stata tends to use different names than SPSS and does some computations differently.

```
. * Generate standardized residuals
. predict rstandard, rstandard
. extremes rstandard grade gpa tuce psi

  +--------------------------------------------+
  | obs:   rstandard   grade    gpa   tuce   psi |
  |--------------------------------------------|
  |  27.   -2.541286       0   3.51     26     1 |
  |  18.   -1.270176       0   3.12     23     1 |
  |  16.   -1.128117       0    3.1     21     1 |
  |  28.    -.817158       0   3.53     26     0 |
  |  24.   -.7397601       0   3.57     23     0 |
  +--------------------------------------------+

  +-------------------------------------------+
  |  19.   .8948758       1   3.39     17     1 |
  |  30.   1.060433       1      4     21     0 |
  |  15.   1.222325       1   2.83     27     1 |
  |  23.   2.154218       1   3.26     25     0 |
  |   2.   3.033444       1   2.39     19     1 |
  +-------------------------------------------+
```

Summary: Comparisons with OLS. There are many similarities between OLS and Logistic Regression, and some important differences. I'll try to highlight the most crucial points here.

OLS and its extensions	**Logistic Regression**
Estimated via least squares	Estimated via Maximum Likelihood.
Y is continuous, can take on any value	Y can only take on 2 values, typically 0 and 1
X's are continuous vars. Categorical variables are divided up into dummy variables	Same as OLS
X's are linearly related to Y; in the case of the LPM, X's are linearly related to P(Y=1)	X's are linearly related to log odds of event occurring. Log odds, in turn, are nonlinearly related to P(Y = 1).
Y's are statistically independent of each other, e.g., don't have serial correlation, don't include husbands and their wives as separate cases	Same as OLS

Robust standard errors can be used when error terms are not independent and identically distributed.	Same as OLS. Stata makes this easy (just add a `robust` parameter), SPSS does not.
There can be no perfect multicollinearity among the X's. High levels of multicollinearity can result in unstable sample estimates and large standard errors	Same as OLS. Techniques for detecting multicollinearity are also similar. In fact, as Menard points out, you could just run the corresponding OLS regression, and then look at the correlations of the IVs, the tolerances, variance inflation factors, etc. Or, use Stata's `collin` command.
Missing data can be dealt with via listwise deletion, pairwise deletion, mean substitution	Pairwise deletion isn't an option. Can't do "mean substitution" on the DV. Otherwise, can use techniques similar to those that we've described for OLS.
Global F test is used to test whether any IV effects differ from 0. d.f. = K, N-K-1	Model chi-square statistic (also known as Model L^2 or G^2 or G_M) is used for same purpose. D.F. = number of IVs in the model = K.
Incremental F test is used to test hypotheses concerning whether subset of coefficients = 0. If you specify variables in blocks, the F change statistic will give you the info you need.	LR Chi-square statistic is used. $DEV_{Constrained} - DEV_{Unconstrained}$. In SPSS, If you specify variables in blocks, the chi-square statistics for the step or block will give you the info you need.
T test or incremental F test is used to test whether an individual coefficient = 0	Can use a LR chi square test (preferable) or Wald statistic (probably usually ok, but not always).
Incremental F tests or T tests can be used to test equalities of coefficients within a model, equalities across populations, interaction effects.	Same basic procedures, substituting LR chi square tests for F tests.
Wald tests (as produced by the `test` command in stata) will produce the same results as incremental F tests. A nice thing about Wald tests is that they only require the estimation of the unconstrained model.	Wald tests can be performed, but they will generally NOT produce exactly the same results as LR tests. LR tests (which require the estimation of constrained and unconstrained models) are preferable, although in practice results will often be similar.
Can have interaction effects. Centering can sometimes make main effects easier to interpret. If you center the continuous vars, then the main effect of an IV like race is equal to the difference in the predicted values for an "average" black and white.	NOT quite the same as OLS. You can use interaction terms, but there are potential problems you should be aware of when interpreting results. See Allison (1999) or Williams (2007) for discussions. If you center, then the main effect of an IV like race is equal to the difference in the log odds for an "average" black and white.

Can do transformations of the IVs and DV to deal with nonlinear relationships, e.g. X^2, ln(X), ln(Y).	Same as OLS for the IVs, but you of course can't do transformations of the dichotomous DV.
Can plot Y against X, examine residuals, plot X against residuals, to identify possible problems with the model	Similar to OLS. Can examine residuals.
Can do mindless, atheoretical stepwise regression	Similar to OLS
R^2 tells how much of total variance is "explained".	Numerous Pseudo R^2 stats have been proposed. If you use one, make clear which one it is.
Can look at standardized betas.	There is actually a reasonable case for using standardized coefficients in logistic regression. Long & Freese's `spostado` routines include the `listcoef` command, which can do various types of standardization.
Can do path analysis. Can decompose association. Can estimate recursive and nonrecursive models. Programs like LISREL can deal with measurement error.	Most ideas of the "logic of causal order" still apply. But, many things, such as decomposition of effects, controlling for measurement error, estimating nonrecursive models, are much, much harder to do. There is work going on in this area, e.g. Lisrel, M-Plus, `gllamm` (an add-on routine to Stata).

Related Topics. Here is a super-quick look at other techniques for analyzing categorical data.

Probit. Probit models are an alternative to Logit models. They tend to produce almost identical results, and logit models are usually easier to work with. For some types of problems, there are more advanced probit techniques that can be useful.

Multinomial Logit. You can also have a dependent variable with more than two categories, e.g. the dependent variable might take the values Republican, Democrat, Other. The idea is that you talk about the probability of being in one group as opposed to another. In SPSS, use NOMREG, in Stata use `mlogit`.

Ordered Logit. Sometimes DVs are ordinal. Sometimes, it is ok to just treat them as interval-level and use OLS regression. But, other times an Ordered Logit routine is preferable. SPSS has PLUM. Stata has the built-in `ologit` and `oprobit`. Stata also has various user-written routines, including Williams's `oglm` and `gologit2`.

Appendix (Optional): Computing the log likelihood. This is adapted from J. Scott Long's Regression Models for Categorical and Limited Dependent Variables.

Define p_i as the probability of observing whatever value of y was actually observed for a given observation, i.e.

$$p_i = \begin{cases} \Pr(y_i = 1 \mid x_i) & \text{if } y_i = 1 \text{ is observed} \\ 1 - \Pr(y_i = 1 \mid x_i) & \text{if } y_i = 0 \text{ is observed} \end{cases}$$

If the observations are independent, the likelihood equation is

$$L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} p_i$$

The likelihood tends to be an incredibly small number, and it is generally easier to work with the log likelihood. Ergo, taking logs, we obtain the log likelihood equation:

$$\ln L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{N} \ln p_i$$

Before proceeding, let's see how this works in practice! Here is how you compute p_i and the log of p_i using Stata:

```
. quietly  logit grade gpa tuce psi
. * Compute probability that y = 1
. predict pi
(option p assumed; Pr(grade))
. * If y = 0, replace pi with probability y = 0
. replace pi = 1 - pi if grade == 0
(21 real changes made)
. * compute log of pi
. gen lnpi = ln(pi)

. list  grade pi lnpi, sep(8)

     +------------------------------+
     | grade         pi        lnpi |
     |------------------------------|
  1. |     0   .9386242   -.0633401 |
  2. |     1   .1110308   -2.197947 |
  3. |     0   .9755296   -.0247748 |
     |     --- Output deleted ---   |
 30. |     1    .569893   -.5623066 |
 31. |     1   .9453403   -.0562103 |
 32. |     1   .6935114   -.3659876 |
     +------------------------------+
```

So, this tells us that the predicted probability of the first case being 0 was .9386. The probability of the second case being a 1 was .111. The probability of the 3^{rd} case being a 0 was .9755; and so on. The likelihood is therefore

$$L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} p_i = .9386 * .1110 * .9755 * ... * .6935 = .000002524$$

which is a really small number; indeed so small that your computer or calculator may have trouble calculating it correctly (and this is only 32 cases; imagine the difficulty if you have hundreds of thousands). Much easier to calculate is the log likelihood, which is

$$\ln L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{N} \ln p_i = -.0633 + -2.198 + ... + -.366 = -12.88963$$

Stata's `total` command makes this calculation easy for us:

```
. total lnpi

Total estimation                    Number of obs    =      32

--------------------------------------------------------------
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
        lnpi |  -12.88963   3.127734     -19.26869   -6.510578
--------------------------------------------------------------
```

Note: The maximum likelihood estimates are those values of the parameters that make the observed data most likely. That is, the maximum likelihood estimates will be those values which produce the largest value for the likelihood equation (i.e. get it as close to 1 as possible; which is equivalent to getting the log likelihood equation as close to 0 as possible).