Logistic Regression, Part II:
The Logistic Regression Model (LRM) – Interpreting Parameters

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised January 26, 2022

This handout steals heavily from 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.

PROBABILITIES, ODDS AND LOG ODDS. The linear probability model (LPM) is

$$P(Yi = 1) = Pi = \alpha + \Sigma\beta X_{ik}$$

Among other things, a problem with this model is that the left hand side can only range from 0 to 1, but the right hand side can vary from negative infinity to positive infinity. One way of approaching this problem is to transform Pi to eliminate the 0 to 1 constraint. We can eliminate the upper bound (Pi = 1) by looking at the ratio Pi/(1 - Pi). Pi/(1 - Pi) is referred to as the *odds* of an event occurring. That is,

$$Odds_i = \frac{P_i}{1-P_i}$$

For example, if Pi = .90, the odds are 9 to 1 (or 9) that the event will happen. If Pi = .30, the odds are 3 to 7 (or .429 to 1, or simply .429) against the event occurring. If Pi = .50, the odds are 1 to 1 (even). The following table helps to illustrate how odds and probabilities are related to each other. It starts with odds of 10,000 to 1 against. Then, the odds are multiplied by 10 each row, until the odds become 10,000 to 1 in favor.

Odds	Probability	Change in Probability
0.0001	0.0100%	
0.001	0.0999%	0.0899%
0.01	0.9901%	0.8902%
0.1	9.0909%	8.1008%
1	50.0000%	40.9091%
10	90.9091%	40.9091%
100	99.0099%	8.1008%
1000	99.9001%	0.8902%
10000	99.9900%	0.0899%

Note the nonlinear relationship between odds and probability. At the low end, you'd prefer to have 100 to 1 odds against you rather than 10,000 to 1 odds against; but either way, you've still got less than a 1% chance of success. Conversely, it is better to have 10,000 to 1 odds in your favor rather than 100 to 1, but either way you've got better than a 99% chance of success. Hence, at the extremes, changes in the odds have little effect on the probability of success.

In the middle ranges, however, it is a very different story. As you go from 10 to 1 odds against to even odds of 1 to 1, the probability of success jumps from 9.9% to 50%, almost a 41% increase. And, as you go from even odds to 10 to 1 odds in your favor, the probability of success jumps from 50% to almost 91%.

The odds must be zero or positive, but there is no upper bound; as Pi approaches 1, Pi/(1 - Pi) goes toward infinity. But, there is still a lower bound of 0. We can eliminate the lower bound of 0 by taking the natural logarithm, ln[Pi/(1 - Pi)] the result of which can be any real number from negative to positive infinity. ln[Pi/(1 - Pi)] is referred to as the *log odds* of the event occurring. That is, the log odds are

$$LogOdds_i = \ln\left(\frac{P_i}{1-P_i}\right)$$

To show how the log odds, odds, and probability are related, we add the log odds to the previous table:

Log odds	Odds	Probability	Change in Probability
-9.2103	0.0001	0.0100%	
-6.9078	0.001	0.0999%	0.0899%
-4.6052	0.01	0.9901%	0.8902%
-2.3026	0.1	9.0909%	8.1008%
0.0000	1	50.0000%	40.9091%
2.3026	10	90.9091%	40.9091%
4.6052	100	99.0099%	8.1008%
6.9078	1000	99.9001%	0.8902%
9.2103	10000	99.9900%	0.0899%

Note that each 10-fold increase in the odds results in the log odds going up by 2.3026. This is because $e^{2.3026} = 10$. That is, what was a *multiplicative* relationship with the odds is now an *additive* relationship in the log odds. Similar to before, we see that, at the extremes, changes in the log odds of success have very little effect on the probability of success; but as you get away from the extremes, changes in log odds are associated with big changes in the probability of success.

This next table also shows the relationships. We start with very low log odds, and increase them by 1 with each row:

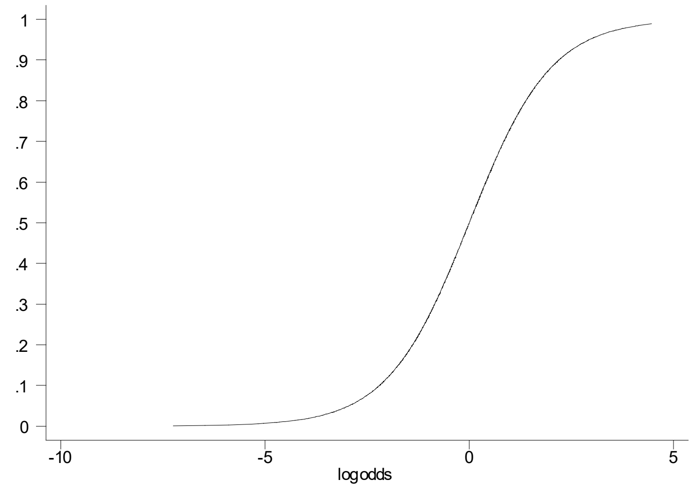
Log odds	Odds	Probability	Change in Probability
-9.0000	0.0001234	0.0123%	
-8.0000	0.0003355	0.0335%	0.0212%
-7.0000	0.0009119	0.0911%	0.0576%
-6.0000	0.0024788	0.2473%	0.1562%
-5.0000	0.0067379	0.6693%	0.4220%
-4.0000	0.0183156	1.7986%	1.1293%
-3.0000	0.0497871	4.7426%	2.9440%
-2.0000	0.1353353	11.9203%	7.1777%
-1.0000	0.3678794	26.8941%	14.9738%
0.0000	1.0000000	50.0000%	23.1059%
1.0000	2.7182818	73.1059%	23.1059%
2.0000	7.3890561	88.0797%	14.9738%
3.0000	20.0855369	95.2574%	7.1777%
4.0000	54.5981500	98.2014%	2.9440%
5.0000	148.4131591	99.3307%	1.1293%
6.0000	403.4287935	99.7527%	0.4220%
7.0000	1096.6331584	99.9089%	0.1562%
8.0000	2980.9579870	99.9665%	0.0576%
9.0000	8103.0839276	99.9877%	0.0212%

Note that the odds get multiplied by 2.7182818 with each row. This is the value of e. At the extremes, each 1 unit increase in the log odds has little effect on the probability of success, but in the middle ranges each 1 unit increase has fairly large effects.

Note also that:

- if the probability of success is less than 50%, the log odds are negative and the odds are less than 1;
- if the probability of success = 50%, the log odds are 0 and the odds = 1;
- if the probability of success is greater than 50%, the log odds are positive and the odds are greater than 1.

The following graph further helps to show how probabilities are related to the corresponding log odds:



Note that

- Although probabilities can range from 0 to 1, log odds can range from -∞ to +∞.

- Log odds follow an S-shaped curve. At the extremes, changes in the log odds produce very little change in the probabilities. In the middle of the S curve, changes in the log odds produce much larger changes in the probabilities.

- To put it another way, linear, additive increases in the log odds produce nonlinear changes in the probabilities.

- The relationship makes intuitive sense.

    - If things are very bad, they can't get that much worse, i.e. whether the odds are 1,000 to 1 against you or only 100 to 1, you still probably won't win, e.g. neither the F nor the D student will likely get an A.

    - Likewise, if things are very good, they can't get that much better. Regardless of whether the odds are 1,000 to 1 in your favor or only 100 to 1, you probably won't fail, e.g., the high A student and the very high A student both have great shots at an A.

    - It is in the middle ranges that changes are likely to make the most difference, e.g. both a B and a C student may have a shot at an A, but the B student's probability of success may be much higher.

- In short, in a regression analysis, log odds have many advantages over probabilities.

    - They have no upper or lower bounds.

    - Linear, additive increases in the log odds produce theoretically plausible nonlinear increases in probability.

- Hence, a method which predicts log odds has a great deal of appeal.

THE LOGISTIC REGRESSION MODEL (LRM). The *logistic regression model (LRM)* (also known as the logit model) can then be written as

$$\ln \frac{P(Yi=1)}{1-P(Yi=1)} = \ln(Odds_i) = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} = Zi$$

The above is referred to as the *log odds* and also as the *logit*. Zi is used as a convenient shorthand for $\alpha + \Sigma\beta_k X_{ik}$.

As Aldrich and Nelson and others note, there are several alternatives to the LRM which might be just as plausible or more plausible in a particular case. However,

- the LRM is comparatively easy from a computational standpoint
- there are many programs available which can estimate logistic regression models
- The LRM tends to work fairly well in practice

Note that, if we know either the odds or the log odds, it is easy to figure out the corresponding probability:

$$P_i = \frac{Odds_i}{1+Odds_i} = \frac{\exp(Z_i)}{1+\exp(Z_i)} = \frac{1}{1+\exp(-Z_i)}$$

So, for example (Confirm these calculations on your own)

if Odds = 1, P = .5; Odds = 3, P = .75; Odds = .5, P = .333.

If Z = 0, P = .5; Z = 1, P = .731; Z = -3, P = .0474.

ESTIMATION OF THE LRM. In linear regression we estimate the parameters of the model using the method of least squares. That is, we select regression coefficients that result in the smallest sums of squared distances between the observed and predicted values of the dependent variable.

In logistic regression, the parameters of the model are estimated using the method of *maximum likelihood.* That is, the coefficients that make our observed results most "likely" are selected. Since the logistic regression model is nonlinear, an iterative algorithm is necessary for parameter estimation. Because the data must be read multiple times, iterative procedures tend to be more time consuming.

The estimation procedure is too difficult to explain here; but fortunately, several programs, including R, SPSS, & Stata, include logistic regression routines.

LOGISTIC REGRESSION EXAMPLE. It will be easier to understand the LRM if we have an example in front of us. Here is how our earlier PSI example could be estimated using the Stata `logit` command.

```
. use https://www3.nd.edu/~rwilliam/statafiles/logist.dta, clear
. logit grade gpa tuce i.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
       1.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
------------------------------------------------------------------------------

. * Replay the results, this time getting the exponentiated coefficients
. logit, or

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

------------------------------------------------------------------------------
       grade | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   16.87972   21.31809     2.24   0.025     1.420194    200.6239
        tuce |   1.099832   .1556859     0.67   0.501     .8333651    1.451502
       1.psi |   10.79073   11.48743     2.23   0.025     1.339344    86.93802
       _cons |   2.21e-06   .0000109    -2.64   0.008     1.40e-10      .03487
------------------------------------------------------------------------------
```

INTERPRETATION OF PARAMETERS. Because the effect of the X's is nonlinear, interpretation of parameters is more difficult than in OLS regression. Recall the LRM can be written as

$$\ln \frac{P(Yi=1)}{1-P(Yi=1)} = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} = Zi$$

Recall that the left hand side stands for the log odds. Hence, a 1 unit increase in X_1 will result in a β_1 increase in the log odds. In the current example, if you had 2 people with identical GPA and TUCE scores, the log odds for the one who was in PSI would be 2.378 greater than the log odds for the person who wasn't.

Alas, most of us are not used to thinking in terms of log odds. Ergo, we may do slightly better if we express the model in terms of odds:

$$\frac{P(Yi=1)}{1-P(Yi=1)} = e^{\alpha+\sum_{k=1}^{K}\beta_k X_{ik}} = e^{a}e^{\beta_1 X_{i1}}e^{\beta_2 X_{i2}}\dots e^{\beta_K X_{iK}}$$

Hence, if X_1 increases by 1, the odds will increase by $\exp(\beta_1)$. So, for 2 otherwise identical students, the odds for the one in PSI would be exp(2.3783) = 10.79 times greater. This is shown in the Odds Ratio column of the printout.

Note that this does *not* mean that the one in PSI is 10.79 times more likely to get an A. This is best illustrated by plugging in some hypothetical numbers. Suppose, based on their GPA and TUCE scores, 5 students in a conventional class had a 1%, 10%, 50%, 90% and 99% chance of getting an A. The following table shows what their odds would be in the conventional class, what their odds would be in a PSI class, and their probability of getting an A in a PSI class:

Pi (Conventional)	Odds (Conv) = Pi/(1 - Pi)	Odds (PSI) = Odds (Conv) * 10.79	Pi (PSI) = Odds/(1 + Odds)	Change in Probability
1.00%	0.0101	0.109	9.83%	8.83%
10.00%	0.1111	1.199	54.51%	44.51%
50.00%	1.0000	10.787	91.52%	41.52%
90.00%	9.0000	97.079	98.98%	8.98%
99.00%	99.0000	1067.868	99.91%	0.91%

Note, for example, that a person with a 10% chance of an A in a regular class sees a huge increase in their chances of an A by getting into PSI. A person who already had a good chance of an A sees the same increase in their odds of getting an A, but a much smaller percentage increase.

By way of contrast, recall that these were the model parameters using OLS:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-1.498	.524		-2.859	.008
	GPA	.464	.162	.449	2.864	.008
	PSI	.379	.139	.395	2.720	.011
	TUCE	.010	.019	.085	.539	.594

a. Dependent Variable: GRADE

According to the OLS estimates, the person with a 1% chance in a conventional class would have a 39% chance in PSI, which is much greater than the 9.83% chance predicted in the logistic regression model. OLS overestimates the benefits of PSI for those with initially low and high probabilities of success, and underestimates the benefits of PSI for those in the middle.

In the above, we used hypothetical values for the Pi. An alternative, and perhaps more common, approach is to "plug in" reasonable values for the X's, and then see what effect changing one of the X's would have. For example:

Consider again the case of a student who has a GPA of 3.0, is taught by traditional methods, and has a score of 20 on TUCE. According to this model,

$$\ln \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} = Z_i$$

$$= -13.0187 + 3 * 2.8256 + 0 * 2.3783 + 20 * .0951$$

$$= -2.6399$$

NOTE: We can confirm this and subsequent calculations in Stata by using the `margins` command. The `predict(xb)` option gives the predicted log odds while the `predict(pr)` option (which is the default for logit and hence can be omitted) gives the predicted probability.

```
. use https://www3.nd.edu/~rwilliam/statafiles/logist.dta, clear
. quietly logit grade gpa tuce i.psi
. margins, at(gpa = 3 psi = 0 tuce = 20) predict(xb)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Linear prediction (log odds), predict(xb)
at           : gpa             =           3
               tuce            =          20
               psi             =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -2.639856   .9831621    -2.69   0.007    -4.566818   -.7128936
------------------------------------------------------------------------------
```

So, the log odds for this person is -2.6399, and the odds are exp(-2.6399) = .0714, i.e. about 1 to 14. Let's convert this into the corresponding probability:

$$P_i = \frac{1}{1 + \exp(-Z_i)} = \frac{1}{1 + \exp(2.6399)} = \frac{1}{15.012} = .0666$$

```
. margins, at(gpa = 3 psi = 0 tuce = 20)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
at           : gpa             =           3
               tuce            =          20
               psi             =           0
```

```
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .066617   .0611322     1.09   0.276    -.0531999    .1864339
------------------------------------------------------------------------------
```

So, according to the model, this person has less than a 7% chance of getting an A. Suppose, however, that this same person got placed in the PSI class. You would then get a logit of

$$\ln \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \alpha + \sum_{k=1}^{K} \beta_k X_{ik} = Z_i$$

$$= -13.0187 + 3 * 2.8256 + 1 * 2.3783 + 20 * .0951$$

$$= -.2616$$

```
. margins, at(gpa = 3 psi = 1 tuce = 20) predict(xb)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Linear prediction (log odds), predict(xb)
at           : gpa             =           3
               tuce            =          20
               psi             =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.2611683   .7374162    -0.35   0.723    -1.706477    1.184141
------------------------------------------------------------------------------
```

So, the log odds are -.2616, and the odds are exp(-.2616) = .7698, i.e. about 1 to 1.3. Again, note that the odds increased by exp(2.3783) = 10.79 times. The probability of getting an A would be

$$P_i = \frac{1}{1 + \exp(-Z_i)} = \frac{1}{1 + \exp(.2616)} = \frac{1}{2.299} = .43497$$

```
. margins, at(gpa = 3 psi = 1 tuce = 20)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
at           : gpa             =           3
               tuce            =          20
               psi             =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4350765   .1812458     2.40   0.016     .0798413    .7903118
------------------------------------------------------------------------------
```

Hence, getting into the PSI class would substantially increase the chances of getting an A — the person would have about a 37% better chance. (Which, incidentally, is about the same increase OLS regression predicted – but note that this person has about "average" GPAand TUCE scores. The means for GPA is 3.12, the mean for TUCE is 21.94)

Suppose, instead, that a student with a 4.0 average and a TUCE of 25 was in a traditional class. (You can confirm the calculations on your own if you want.) The log odds are .6612, and the odds are exp(.6612) = 1.937, i.e. the person is almost twice as likely to get an A as to not get an A. Pi is

```
. margins, at(gpa = 4 psi = 0 tuce = 25)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
at           : gpa             =           4
               tuce            =          25
               psi             =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .6597197   .2329773     2.83   0.005     .2030926    1.116347
------------------------------------------------------------------------------
```

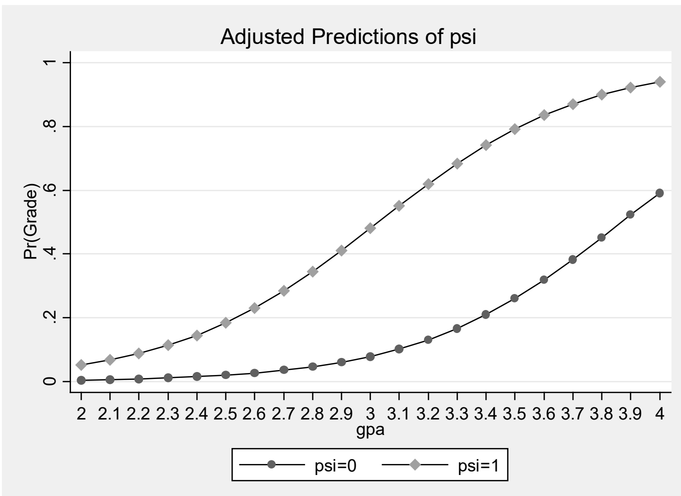
So, in a regular classroom, that student has about a 66% chance of an A. Put them in PSI, and the log odds are 3.0395, odds are 20.89 (again an increase of 10.79 times), and Pi is

```
. margins, at(gpa = 4 psi = 1 tuce = 25)

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
at           : gpa             =           4
               tuce            =          25
               psi             =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .9543808   .0560709    17.02   0.000     .8444837    1.064278
------------------------------------------------------------------------------
```

So, this person sees about a 29% improvement. Note that it would be impossible for this person to improve by 37%, like the first person did, because his or her probability of getting an A would then be greater than 1.

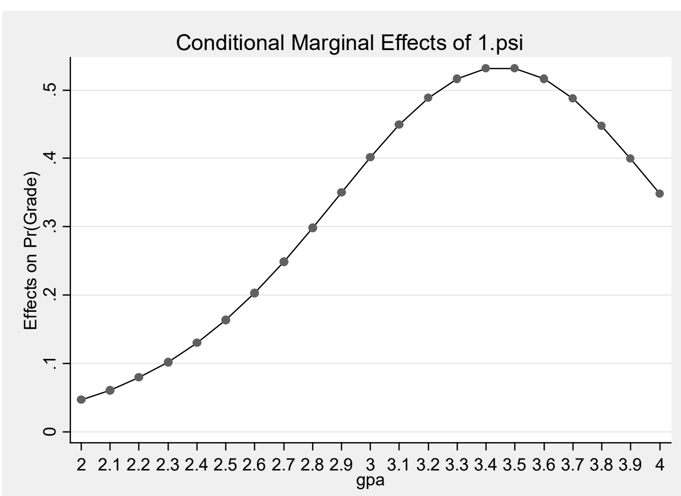
[OPTIONAL: USING THE MARGINS COMMAND EFFECTIVELY] Incidentally, the `margins` and `marginsplot` commands provide a nice way of generating several interesting comparisons at once. For example, the following shows the predicted scores for people who have average scores on TUCE and identical GPAs but who differ on their PSI status.

```
. quietly logit grade gpa tuce i.psi, nolog
. quietly margins psi, at(gpa = (2(.1)4) ) atmeans
. marginsplot, noci scheme(sj)

  Variables that uniquely identify margins: gpa psi
```



If we just want to see the predicted differences between those in psi and those not in psi,

```
. quietly margins, dydx(psi) at(gpa = (2(.1)4) ) atmeans
. marginsplot, noci scheme(sj)

  Variables that uniquely identify margins: gpa
```



With all the graphics, we see that the predicted differences in the probability of getting an A between those in psi and those not in psi is very small at low gpas, less than 10%. As GPA rises the gap gets much larger (more than 50%), up until around gpa = 3.4, and then the gap gets smaller as gpa continues to get bigger. In the last graph, when the confidence interval includes 0, the difference between those in psi and not in psi is not statistically significant, suggest that those with low B to low A GPAs are most likely to see a boost from being in psi.

Here are the actual data. In the last 3 columns, we compute what the probability of an A would have been in the student were not in PSI, then what the probability would be if the student were in Psi, and the gain in probability produced by being in psi. The data are sorted from lowest probability of success if not in Psi to the highest probability of success if not in Psi.

Gpa	Tuce	Psi	Grade	Logit if Not in Psi	Logit if in Psi	P if not in Psi	P if in PSI	Psi Gain
2.06	22	1	0	-5.10744	-2.72944	0.60%	6.13%	5.52%
2.39	19	1	1	-4.45986	-2.08186	1.14%	11.09%	9.94%
2.63	20	0	0	-3.68662	-1.30862	2.44%	21.27%	18.83%
2.92	12	0	0	-3.62708	-1.24908	2.59%	22.29%	19.70%
2.76	17	0	0	-3.60424	-1.22624	2.65%	22.68%	20.04%
2.66	20	0	0	-3.60184	-1.22384	2.65%	22.73%	20.07%
2.89	14	1	0	-3.52186	-1.14386	2.87%	24.16%	21.29%
2.74	19	0	0	-3.47076	-1.09276	3.02%	25.11%	22.09%
2.86	17	0	0	-3.32164	-0.94364	3.48%	28.02%	24.53%
2.83	19	0	0	-3.21642	-0.83842	3.86%	30.19%	26.33%
2.67	24	1	0	-3.19358	-0.81558	3.94%	30.67%	26.73%
2.87	21	0	0	-2.91338	-0.53538	5.15%	36.93%	31.78%
2.75	25	0	0	-2.8725	-0.4945	5.35%	37.88%	32.53%
2.89	22	0	0	-2.76186	-0.38386	5.94%	40.52%	34.58%
2.83	27	1	1	-2.45642	-0.07842	7.90%	48.04%	40.14%
3.1	21	1	0	-2.2634	0.1146	9.42%	52.86%	43.44%
3.03	25	0	0	-2.08122	0.29678	11.09%	57.37%	46.27%
3.12	23	1	0	-2.01688	0.36112	11.74%	58.93%	47.19%
3.39	17	1	1	-1.82386	0.55414	13.90%	63.51%	49.61%
3.16	25	1	1	-1.71384	0.66416	15.27%	66.02%	50.75%
3.28	24	0	0	-1.46972	0.90828	18.70%	71.26%	52.57%
3.32	23	0	0	-1.45168	0.92632	18.97%	71.63%	52.66%
3.26	25	0	1	-1.43124	0.94676	19.29%	72.05%	52.76%
3.57	23	0	0	-0.74518	1.63282	32.19%	83.66%	51.47%
3.54	24	1	1	-0.73496	1.64304	32.41%	83.79%	51.38%
3.65	21	1	1	-0.7091	1.6689	32.98%	84.14%	51.16%
3.51	26	1	0	-0.62974	1.74826	34.76%	85.17%	50.42%
3.53	26	0	0	-0.57322	1.80478	36.05%	85.87%	49.82%
3.62	28	1	1	-0.12888	2.24912	46.78%	90.46%	43.68%
4	21	0	1	0.28	2.658	56.95%	93.45%	36.50%
4	23	1	1	0.47	2.848	61.54%	94.52%	32.98%
3.92	29	0	1	0.81392	3.19192	69.29%	96.05%	26.76%

As you see, the gains from being in psi are initially small, then gradually get much bigger, and then start to drop again. Again, this is far different from OLS's across the board prediction of a 38% gain from being in Psi.

Of course, the above table focuses on predicted success. The actual benefits of Psi are made clear in the following table, which lists only the 11 students who got A's.

Gpa	Tuce	Psi	Grade	Logit if Not in Psi	Logit if in Psi	P if not in Psi	P if in PSI	Psi Gain
3.26	25	0	1	-1.43124	0.94676	19.29%	72.05%	52.76%
4	21	0	1	0.28	2.658	56.95%	93.45%	36.50%
3.92	29	0	1	0.81392	3.19192	69.29%	96.05%	26.76%
2.39	19	1	1	-4.45986	-2.08186	1.14%	11.09%	9.94%
2.83	27	1	1	-2.45642	-0.07842	7.90%	48.04%	40.14%
3.39	17	1	1	-1.82386	0.55414	13.90%	63.51%	49.61%
3.16	25	1	1	-1.71384	0.66416	15.27%	66.02%	50.75%
3.54	24	1	1	-0.73496	1.64304	32.41%	83.79%	51.38%
3.65	21	1	1	-0.7091	1.6689	32.98%	84.14%	51.16%
3.62	28	1	1	-0.12888	2.24912	46.78%	90.46%	43.68%
4	23	1	1	0.47	2.848	61.54%	94.52%	32.98%

Of the 14 students who were in Psi, 8 got A's, even though only 1 would have had better than a 50% chance at an A had they been in a conventional class. Conversely, only 3 of the 18 students in conventional classes got A's, and two of those had near-perfect GPAs coming into the class. Thus, it is reasonable to conclude that most of the students in Psi who got A's would not have done so had they been in a conventional class.

Assuming you don't want to present all the data, what sorts of values should you "plug in?"

- You might want to plug in the mean value for each continuous variable, to see how the "average" person does. Then, vary the value of one of the variables. For example, in this case you could plug in the means for GPA and TUCE, and then compute Pi when PSI = 0 and when PSI = 1. This would tell you how much better the "average" student would do in PSI.
- Better still, you might plug in values for a below-average student (someone with low GPA and low TUCE), an average student, and a very above-average student (with high GPA and high TUCE). This would indicate how much different types of students could be expected to benefit by PSI.

Another possible way of interpreting parameters is by looking at the relative effects of variables that are measured on the same scale.

- For example, suppose the model includes dummy variables for race and gender. If the effect of gender is larger than the effect of race, we could conclude that gender had the stronger effect.
- Or, suppose both years of education and years of job experience are DVs. You could look at the relative effects to see which was stronger. (Of course, you can also do these sorts of comparisons in a regular OLS regression.)

Summary. The assumptions of the logistic regression model are far more plausible than the assumptions of OLS. Unfortunately, because relationships are nonlinear rather than linear, parameters in logistic regression are not as easily interpretable as parameters in OLS regression. Some things you can do are

- Look at the T value (or Wald statistic) to see whether the effect is statistically significant. (We'll discuss hypothesis testing more in a later handout)

- Look at the sign of the effect, to see whether increases in the variable increase or decrease the probability of success

- Look at $\exp(\beta_k)$, to see how much a 1 unit increase in X_k changes the odds of success (keeping in mind that odds of success is not the same as probability of success)

- Plug in different values for the X variables, and see how changes in the value of an X variable affect the probability of success. Or, plug in different values for Pi, and see how changes in X change Pi. Values chosen should be reasonable ones.

- Look at the relative magnitudes of similarly measured variables, to determine which seem to have the greater impact

Coming up, we'll talk about significance tests, hypothesis testing, and diagnostics. We'll see that there are many parallels with OLS, although we do some things a little different.

Appendix: Some things to remember about logarithms [optional review if you need it]

$e = 2.71828$	e is an irrational number
$e^0 = 1$	indeed, anything to the 0 power (except 0) = 1
$\ln(e) = 1$	Ln = the natural log
$e^{\ln(a + b + c)} = a + b + c$	e.g. $e^{\ln(2 + 3 + 4)} = e^{2.1972} = 9$
$\ln (X^a) = a \ln(X)$	e.g. $\ln(7^2) = 2 \ln(7) = 3.8918$
$\ln (e^a) = a \ln(e) = a$	e.g. $\ln(e^2) = \ln(7.389) = 2$
$\ln (ab) = \ln(a) + \ln(b)$	e.g $\ln(2 * 8) = \ln(2) + \ln(8) = 2.7726$
$\ln(X^a Y^b) = a*\ln(X) + b*\ln(Y)$	e.g. $\ln(2^2 * 3^2) = 2*\ln(2) + 2*\ln(3) = 3.5835$
$e^{a+b+c} = e^a e^b e^c$	e.g. $e^{2+3+4} = e^2 e^3 e^4 = 8103.08$

Note also that

- e^X is often written as exp (X) for convenience. Exp(X) is called the *antilog* of X.

You can only take logarithms of positive, nonzero numbers (above rules involving logarithms assume positive numbers). Since we'll be working with probabilities, this won't be a problem for us.

Appendix: Constant Only Model

Here are the parameter estimates when only the constant is in the model:

```
. logit grade, nolog

Logistic regression                               Number of obs   =         32
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood =  -20.59173                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.6466272   .3721937    -1.74   0.082    -1.376113    .0828591
------------------------------------------------------------------------------
```

Since only the constant is entered, the log odds for every case are -.647. Hence, the predicted probability of success for every case is

$$P_i = \frac{1}{1+\exp(-Z_i)} = \frac{1}{1+\exp(.647)} = \frac{1}{1+1.9098} = .3437$$

`margins` whines when there are no independent variables, so to confirm we can do

```
. display 1 / (1 + exp(--.6466272))
.34374999
```

If you prefer, you can also work with the odds ratios directly:

```
. logit, or

Logistic regression                               Number of obs   =         32
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood =  -20.59173                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
       grade | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5238095   .1949586    -1.74   0.082     .2525582    1.086389
------------------------------------------------------------------------------

. display .5238095/(1 + .5238095)
.34374999
```

In the sample, 11 of 32 cases, or 34.37%, got A's. So, in a model with only the intercept, the intercept gives you the same information that a frequency distribution of the DV would give you (albeit in a rather convoluted way).