**LOGIT OUTPUT INTERPRETATION**

The odds of an event happening is defined as the probability that the event occurs divided by the probability that the event does not occur

* probability: the number of times the event occurs divided by the number of times the event could occur (possible values range from 0 to 1)
* odds: the probability that an event will occur divided by the probability that the event will not occur: probability(success) / probability(failure)
* odds ratio: the ratio of the odds of success for one group divided by the odds of success for the other group:  ( probability(success)A/probability(failure)A ) /  ( probability(success)B/probability(failure)B )
* log odds: the natural log of the odds

**Coefficient:** This means that the coefficients in a simple logistic regression are in terms of the log odds, that is, the coefficient 1.694596 implies that a one unit change in gender results in a 1.694596 unit change in the log of the odds.

Odd ratio: In our particular example, e1.694596 = 5.44 which implies that the odds of being admitted for males is 5.44 times that of females.

**Probit:**

A positive coefficient means that an increase in the predictor leads to an increase in the predicted probability.  A negative coefficient means that an increase in the predictor leads to a decrease in the predicted probability.

**gre** – The coefficient of **gre** is 0.0015244.  This means that an increase in GRE score increases the predicted probability of admission.

**gpa** – The coefficient of **gpa** is 0.4009853.  This means that an increase in GPA increases the predicted probability of admission.

**Marginal Effects:**

This means that for a 1% increase in price, the probability that a car is foreign increases by 0.005 on a [0,1] scale.

**MULTINOMIAL LOGIT: low ses relative to middle ses**

**science** – This is the multinomial logit estimate for a one unit increase in **science** test score for low **ses**relative to middle **ses** given the other variables in the model are held constant. If a subject were to increase his **science** test score by one point, the multinomial log-odds for low **ses** relative to middle **ses**would be expected to decrease by 0.024 unit while holding all other variables in the model constant.

**socst** – This is the multinomial logit estimate for a one unit increase in **socst** test score for low **ses**relative to middle **ses** given the other variables in the model are held constant. If a subject were to increase his **socst** test score by one point, the multinomial log-odds for low **ses** relative to middle **ses**would be expected to decrease by 0.039 unit while holding all other variables in the model constant.

**female** – This is the multinomial logit estimate comparing females to males for low **ses** relative to middle **ses** given the other variables in the model are held constant. The multinomial logit for females relative to males is 0.817 unit higher for being in low **ses** relative to middle **ses** given all other predictor variables in the model are held constant.

LOGISTIC REGRESSION ANALYSIS | STATA ANNOTATED OUTPUT

This page shows an example of logistic regression regression analysis with footnotes explaining the output.  These data were collected on 200 high schools students and are scores on various tests, including science, math, reading and social studies (**socst**).  The variable **female** is a dichotomous variable coded 1 if the student was female and 0 if male.

Because we do not have a suitable dichotomous variable to use as our dependent variable, we will create one (which we will call **honcomp**, for honors composition) based on the continuous variable **write**.  We do not advocate making dichotomous variables out of continuous variables; rather, we do this here only for purposes of this illustration.

**use https://stats.idre.ucla.edu/stat/data/hsb2, clear**

**generate honcomp = (write >=60)**

**logit honcomp female read science**

Iteration 0: log likelihood = -115.64441

Iteration 1: log likelihood = -84.558481

Iteration 2: log likelihood = -80.491449

Iteration 3: log likelihood = -80.123052

Iteration 4: log likelihood = -80.118181

Iteration 5: log likelihood = -80.11818

Logit estimates Number of obs = 200

LR chi2(3) = 71.05

Prob > chi2 = 0.0000

Log likelihood = -80.11818 Pseudo R2 = 0.3072

------------------------------------------------------------------------------

honcomp | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

female | 1.482498 .4473993 3.31 0.001 .6056111 2.359384

read | .1035361 .0257662 4.02 0.000 .0530354 .1540369

science | .0947902 .0304537 3.11 0.002 .035102 .1544784

\_cons | -12.7772 1.97586 -6.47 0.000 -16.64982 -8.904589

------------------------------------------------------------------------------

Iteration Log

Iteration 0: log likelihood = -115.64441

Iteration 1: log likelihood = -84.558481

Iteration 2: log likelihood = -80.491449

Iteration 3: log likelihood = -80.123052

Iteration 4: log likelihood = -80.118181

Iteration 5:a log likelihood = -80.11818

**a.**  This is a listing of the log likelihoods at each iteration.  (Remember that logistic regression uses maximum likelihood, which is an iterative procedure.)  The first iteration (called iteration 0) is the log likelihood of the "null" or "empty" model; that is, a model with no predictors.  At the next iteration, the predictor(s) are included in the model.  At each iteration, the log likelihood increases because the goal is to maximize the log likelihood.  When the difference between successive iterations is very small,  the model is said to have "converged", the iterating is stopped and the results are displayed.  For more information on this process, see [Regression Models for Categorical and Limited Dependent Variables](https://stats.idre.ucla.edu/stata/examples/long/) by J. Scott Long.

Model Summary

Logit estimates Number of obsc = 200

LR chi2(3)d = 71.05

Prob > chi2e = 0.0000

Log likelihood = -80.11818b Pseudo R2f = 0.3072

.  **Log likelihood** – This is the log likelihood of the final model.  The value -80.11818 has no meaning in and of itself; rather, this number can be used to help compare nested models.

c.  **Number of obs** – This is the number of observations that were used in the analysis.  This number may be smaller than the total number of observations in your data set if you have missing values for any of the variables used in the logistic regression.  Stata uses a listwise deletion by default, which means that if there is a missing value for any variable in the logistic regression, the entire case will be excluded from the analysis.

d.  **LR chi2(3)** – This is the likelihood ratio (LR) chi-square test.  The likelihood chi-square test statistic can be calculated by hand as 2\*(115.64441 – 80.11818) = 71.05.  This is minus two (i.e., -2) times the difference between the starting and ending log likelihood.  The number in the parenthesis indicates the number of degrees of freedom.  In this model, there are three predictors, so there are three degrees of freedom.

e.  **Prob > chi2** – This is the probability of obtaining the chi-square statistic given that the null hypothesis is true.  In other words, this is the probability of obtaining this chi-square statistic (71.05) if there is in fact no effect of the independent variables, taken together, on the dependent variable.  This is, of course, the p-value, which is compared to a critical value, perhaps .05 or .01 to determine if the overall model is statistically significant.  In this case, the model is statistically significant because the p-value is less than .000.

f.  **Pseudo R2** – This is the pseudo R-squared.  Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one.  There are a wide variety of pseudo-R-square statistics.  Because this statistic does not mean what R-square means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic with great caution.

Parameter Estimates

------------------------------------------------------------------------------

honcompg| Coef.h Std. Err.i zj P>|z|j [95% Conf. Interval]k

-------------+----------------------------------------------------------------

female | 1.482498 .4473993 3.31 0.001 .6056111 2.359384

read | .1035361 .0257662 4.02 0.000 .0530354 .1540369

science | .0947902 .0304537 3.11 0.002 .035102 .1544784

\_cons | -12.7772 1.97586 -6.47 0.000 -16.64982 -8.904589

------------------------------------------------------------------------------

g.  **honcomp** – This is the dependent variable in our logistic regression.  The variables listed below it are the independent variables.

h.  **Coef**. – These are the values for the logistic regression equation for predicting the dependent variable from the independent variable.  They are in log-odds units.  Similar to OLS regression, the prediction equation is

**log(p/1-p) = b0 + b1\*female + b2\*read + b3\*science**

where p is the probability of being in honors composition.  Expressed in terms of the variables used in this example, the logistic regression equation is

**log(p/1-p) = -12.7772 + 1.482498\*female + .1035361\*read + 0947902\*science**

These estimates tell you about the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale.  These estimates tell the amount of increase in the predicted log odds of honcomp = 1 that would be predicted by a 1 unit increase in the predictor, holding all other predictors constant.  Note: For the independent variables which are not significant, the coefficients are not significantly different from 0, which should be taken into account when interpreting the coefficients.  (See the columns with the z-values and p-values regarding testing whether the coefficients are statistically significant).  Because these coefficients are in log-odds units, they are often difficult to interpret, so they are often converted into odds ratios.  You can do this by hand by exponentiating the coefficient, or by using the **or** option with **logit** command, or by using the **logistic** command.

**female** – The coefficient (or parameter estimate) for the variable **female** is 1.482498.  This means that for a one-unit increase in **female** (in other words, going from male to female), we expect a 1.482498 increase in the log-odds of the dependent variable **honcomp**, holding all other independent variables constant.  
  **read** – For every one-unit increase in reading score (so, for every additional point on the reading test), we expect a .1035361 increase in the log-odds of **honcomp**, holding all other independent variables constant.  
  **science** – For every one-unit increase in science score, we expect a .0947902 increase in the log-odds of **honcomp**, holding all other independent variables constant.  
  **constant** – This is the expected value of the log-odds of **honcomp** when all of the predictor variables equal zero.  In most cases, this is not interesting.  Also, oftentimes zero is not a realistic value for a variable to take.

i.  **Std. Err.** – These are the standard errors associated with the coefficients.  The standard error is used for testing whether the parameter is significantly different from 0; by dividing the parameter estimate by the standard error you obtain a z-value (see the column with z-values and p-values).  The standard errors can also be used to form a confidence interval for the parameter, as shown in the last two columns of this table.

j.  **z** and **P>|z|** – These columns provide the z-value and 2-tailed p-value used in testing the null hypothesis that the coefficient (parameter) is 0.   If you use a 2-tailed test, then you would compare each p-value to your preselected value of alpha.  Coefficients having p-values less than alpha are statistically significant.  For example, if you chose alpha to be 0.05, coefficients having a p-value of 0.05 or less would be statistically significant (i.e., you can reject the null hypothesis and say that the coefficient is significantly different from 0).   If you use a 1-tailed test (i.e., you predict that the parameter will go in a particular direction), then you can divide the p-value by 2 before comparing it to your preselected alpha level.  With a 2-tailed test and alpha of 0.05, you may reject the null hypothesis that the coefficient for **female** is equal to 0.  The coefficient of 1.482498 is significantly greater than 0.     
  The coefficient for **read** is .1035361 significantly different from 0 using alpha of 0.05 because its p-value is 0.000, which is smaller than 0.05.  
  The coefficient for **science** is .0947902 significantly different from 0 using alpha of 0.05 because its p-value is 0.000, which is smaller than 0.05.

k. **[95% Conf. Interval]** – This shows a 95% confidence interval for the coefficient.  This is very useful as it helps you understand how high and how low the actual population value of the parameter might be.  The confidence intervals are related to the p-values such that the coefficient will not be statistically significant if the confidence interval includes 0.

Odds Ratios

In this next example, we will illustrate the interpretation of odds ratios.  We will use the **logistic** command so that we see the odds ratios instead of the coefficients.  In this example, we will simplify our model so that we have only one predictor, the binary variable **female**.  Before we run the logistic regression, we will use the **tab** command to obtain a crosstab of the two variables.

**tab female honcomp**

| honcomp

female | 0 1 | Total

-----------+----------------------+----------

male | 73 18 | 91

female | 74 35 | 109

-----------+----------------------+----------

Total | 147 53 | 200

If we divide the number of males who are in honors composition, 18, by the number of males who are not in honors composition, 73, we get the odds of being in honors composition for males, 18/73 = .24657534.  If we do the same thing for females, we get 35/74 = .47297297.  To get the odds ratio, which is the ratio of the two odds that we have just calculated, we get .47297297/.24657534 = 1.9181682.  As we can see in the output below, this is exactly the odds ratio we obtain from the **logistic** command.  The thing to remember here is that you want the group coded as 1 over the group coded as 0, so honcomp=1/honcomp=0 for both males and females, and then the odds for females/odds for males, because the females are coded as 1.

With regard to the 95% confidence interval, we do not want this to include the value of 1.  When we were considering the coefficients, we did not want the confidence interval to include 0.  If we exponentiate 0, we get 1 (exp(0) = 1).  Hence, this is two ways of saying the same thing.  As you can see, the 95% confidence interval includes 1; hence, the odds ratio is not statistically significant.  Because the lower bound of the 95% confidence interval is so close to 1, the p-value is very close to .05.

There are a few other things to note about the output below.  The first is that although we have only one predictor variable, the test for the odds ratio does not match with the overall test of the model.  This is because the z statistic is actually the result of a Wald chi-square test, while the test of the overall model is a likelihood ratio chi-square.  While these two types of chi-square tests are asymptotically equivalent, in small samples they can differ, as they do here.  Also, we have the unfortunate situation in which the results of the two tests give different conclusions.  This does not happen very often.  In a situation like this, it is difficult to know what to conclude.  One might consider the power, or one might decide if an odds ratio of this magnitude is important from a clinical or practical standpoint.

**logistic honcomp female**

Logistic regression Number of obs = 200

LR chi2(1) = 3.94

Prob > chi2 = 0.0473

Log likelihood = -113.6769 Pseudo R2 = 0.0170

------------------------------------------------------------------------------

honcomp | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

female | 1.918168 .6400451 1.95 0.051 .9973827 3.689024

------------------------------------------------------------------------------

For more information on interpreting odds ratios, please see [How do I interpret odds ratios in logistic](https://stats.idre.ucla.edu/stata/faq/how-do-i-interpret-odds-ratios-in-logistic-regression/)

**Coefficient:** This means that the coefficients in a simple logistic regression are in terms of the log odds, that is, the coefficient 1.694596 implies that a one unit change in gender results in a 1.694596 unit change in the log of the odds.

Odd ratio: In our particular example, e1.694596 = 5.44 which implies that the odds of being admitted for males is 5.44 times that of females.