**Descriptive Statistics**

```
. summarize age attend happy married male racenew sexfreq

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |      2751    46.28281    17.37049         18         89
      attend |      2743    3.662413     2.70204          0          8
       happy |      1369    1.821037    .6289519          1          3
     married |      2765    .4589512    .4984023          0          1
        male |      2765     .444123    .4969578          0          1
-------------+--------------------------------------------------------
     racenew |      2762    .7585083    .4280652          0          1
     sexfreq |      2151    2.825198    2.013244          0          6
```

**"Obs"** = The number of observations with valid responses for each variable. For general happiness there are 1369 valid responses.

**"Minimum"** = The minimum possible value for each variable. Recall that values for general happiness are coded (1) very happy, (2) pretty happy, and (3) not too happy.

**"Maximum"** = The maximum possible value for each variable.

Since we have coded race, gender, and marital status as a dummies, the mean can be interpreted as the percent of respondents reporting the focus category (1=white, 1=male, 1=married). For gender, 44% of the respondents are male. For race, 76% of the respondents are White. For marital status, 46% of the respondents are married.

**"Mean"** = The mean value for all 2751 cases. Respondents have a mean age of 46.28 years.

**"Std. Deviation"** = The standard deviation is a measure of dispersion around the mean. For example, the mean age for this sample is 46, with a standard deviation of 17. So, 95% of the cases in a normal distribution will be between 29 and 63.

**Frequencies**

In this example, we compare the "table" and "tabulate" commands in STATA. We recoded the missing values to "-9" so that it could be shown in the tables.

```
. table happy
```

```
--------------------------
general       |
happiness     |      Freq.
--------------+-----------
           -9 |      1,396
   very happy |        415
 pretty happy |        784
not too happy |        170
--------------------------
```

The "freq." column indicates simply the number of cases in each of the categories. In this example, 415 of the cases have a value of (1) very happy and 1396 of the cases have missing values (-9).

```
. tabulate happy
```

```
      general |
    happiness |      Freq.     Percent        Cum.
--------------+-----------------------------------
           -9 |      1,396       50.49       50.49
   very happy |        415       15.01       65.50
 pretty happy |        784       28.35       93.85
not too happy |        170        6.15      100.00
--------------+-----------------------------------
        Total |      2,765      100.00
```

The "tabulate" command provides "freq.," "percent," and the cumulative percent ("cum."). In this example, 415 of the cases (15.01%) have a value of (1) very happy and 1396 (50.49%) of the cases have missing values (-9).

## Correlations

The correlation tells you the magnitude and direction of the association between two variables.

```
. pwcorr happy sexfreq age , obs sig

             |    happy  sexfreq      age
-------------+---------------------------
       happy |   1.0000
             |
             |     1369
             |
     sexfreq |  -0.1053   1.0000
             |   0.0006
             |     1060     2151
             |
         age |  -0.0453  -0.4370   1.0000
             |   0.0944   0.0000
             |     1362     2143     2751
             |
```

Here is the correlation between age of respondent and frequency of sex. The correlation coefficient is -.437 (and is significant). This suggests a negative correlation with moderate magnitude. As age increases, the frequency of sex decreases. The correlation between age and frequency of sex is -.437. If we square this value, we get .190969, or 19.1 out of 100, or 19.1 percent. From this we can claim that 19.1% of the correlation in frequency of sex is attributed to respondent's age.

This cell represents the correlation (and significance and sample size) between age and general happiness. The top value (-.045) is the correlation coefficient. The middle value (.094) is the significance. The bottom value (1362) is the number of cases. In this example, the correlation is not significant, at the $p < .05$ level.

Note

The following general categories indicate a quick way of interpreting correlations.

| | |
|---|---|
| 0.0 – 0.2 | Very weak correlation |
| 0.2 – 0.4 | Weak correlation |
| 0.4 – 0.7 | Moderate correlation |
| 0.7 – 0.9 | Strong correlation |
| 0.9 – 1.0 | Very strong correlation |

**T-Test**

The Independent Sample T-Test compares the mean scores of two groups on a given variable. In this example, we compare “frequency of sex” for males versus females.

Null Hypothesis: The means of the two groups (males and females) are not significantly different.
Alternate Hypothesis: The means of the two groups (males and females) are significantly different.

```
. oneway sexfreq sex

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      134.322319      1   134.322319     33.64     0.0000
 Within groups      8579.95197   2149   3.99253233
------------------------------------------------------------------------
    Total           8714.27429   2150   4.05315083

Bartlett's test for equal variances:  chi2(1) =   6.3875  Prob>chi2 = 0.011
```

Bartlett’s test tests the hypothesis that our samples have equal variances. If the Chi-Square test is significant ($p < .05$), then we can reject the null hypothesis and assume that variances are unequal. In this case, Chi-Square is significant and we can reject the null hypothesis and assume that variances are unequal.

```
. ttest sexfreq , by(sex)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |     976    3.099385    .0612103    1.912271    2.979266    3.219504
  female |    1175    2.597447    .0602927     2.06673    2.479153     2.71574
---------+--------------------------------------------------------------------
combined |    2151    2.825198    .0434086    2.013244     2.74007    2.910325
---------+--------------------------------------------------------------------
    diff |            .5019384    .0865368                .3322339    .6716429
------------------------------------------------------------------------------
Degrees of freedom: 2149

                  Ho: mean(male) - mean(female) = diff = 0

     Ha: diff < 0               Ha: diff != 0              Ha: diff > 0
       t =   5.8003                t =   5.8003              t =   5.8003
   P < t =   1.0000          P > |t| =   0.0000          P > t =   0.0000
```

We can see here that males report higher frequencies of sex than females. However, we cannot tell from here whether or not this difference is significant.

Our “t” value is 5.842. There is a significant difference between the two groups. We can conclude, then, that males and females significantly differ in their reports of sexual frequency: males reported higher levels of sexual frequency than did females. The circled value indicates the difference in means (male score minus female score).

**Crosstab (with Chi-Square)**

A crosstabulation displays the number of cases in each category defined by two or more grouping variables.

The Chi Square Goodness of Fit test determines if the observed frequencies are different from what we would expect to find (we expect equal numbers in each group within a variable).

Null Hypothesis: There are approximately equal numbers of cases in each group.
Alternate Hypothesis: There are not equal numbers of cases in each group.

```
. tabulate happy freqdum , chi2

      general |     sexfreq dummy
    happiness | less than  more than |     Total
--------------+----------------------+----------
   very happy |       103        207 |       310
 pretty happy |       286        333 |       619
not too happy |        63         68 |       131
--------------+----------------------+----------
        Total |       452        608 |     1,060

          Pearson chi2(2) =  16.0387   Pr = 0.000
```

For example, we see that there are 207 cases reporting "very happy" for general happiness and "more than once per month" for frequency of sex.

The Chi-Square measure tests the hypothesis that the row and column variables in a crosstabulation are independent of one another. While the Chi-Square measure may indicate that there is a relationship between two variables, they do not indicate the strength or direction of the relationship.

**ANOVA**

The One-Way ANOVA compares the mean of one or more groups based on one independent variables (or factor). We assume that the dependent variable is normally distributed and that groups have approximately equal variance on the dependent variable.

Null Hypothesis: There are no significant differences between groups' mean scores.
Alternate Hypothesis: There is a significant difference between groups' mean scores.

In this example, we compare "frequency of sex" by church attendance, recoded from 9 groups to 3 groups (0=not often, 1=sometimes, 2=often).

```
. anova sexfreq church

                        Number of obs =    2142     R-squared     =  0.0132
                        Root MSE      = 2.00297     Adj R-squared =  0.0122

           Source |  Partial SS    df       MS           F     Prob > F
       -----------+----------------------------------------------------
            Model |  114.423493     2  57.2117464      14.26     0.0000
                  |
           church |  114.423493     2  57.2117464      14.26     0.0000
                  |
         Residual |  8581.45139  2139  4.01189873
       -----------+----------------------------------------------------
            Total |  8695.87488  2141  4.06159499
```

$$F = \frac{\text{variance between groups}}{\text{variance expected due to chance (error)}} = \frac{57.212}{4.012} = 14.26$$

If the sample means are clustered closely together (i.e., small differences), the variance will be small; if the means are spread out (i.e., large differences), the variances will be larger.

Our F value is 14.261. Our significance level is .000. We can conclude that there is a significant difference between the three groups. To determine which groups are different from one another, we use the "multiple comparisons" results below.

**Simple Linear Regression**

Simple Linear Regression tells you the amount of variance accounted for by one variable in predicting another variable.

```
. reg sexfreq age marital racenew happy attend

      Source |       SSc       df       MS              Number of obs =    1052
-------------+------------------------------           F(  5,  1046) =  101.49d
       Modela|  1382.66217     5  276.532435           Prob > F      =  0.0000
    Residualb|  2850.06406  1046  2.72472664           R-squared     =  0.3267e
-------------+------------------------------           Adj R-squared =  0.3234
       Total |  4232.72624  1051  4.02733229           Root MSE      =  1.6507f

------------------------------------------------------------------------------
     sexfreq |      Coef.   Std. Err.      tg    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0554982   .0030401   -18.26   0.000    -.0614636   -.0495327
     marital |  -1.372326   .1071273   -12.81   0.000    -1.582535   -1.162117
     racenew |  -.2153774   .1252702    -1.72   0.086    -.4611869    .0304321
       happy |  -.2618401    .084642    -3.09   0.002    -.4279275   -.0957527
      attend |  -.0674279   .0196243    -3.44   0.001    -.1059354   -.0289204
       _cons |      8.298   .2966264    27.97   0.000     7.715949     8.88005
------------------------------------------------------------------------------
```

**a.** The output for Model displays information about the variation accounted for by the model.

**b.** The output for Residual displays information about the variation that is not accounted for by your model. And the output for Total is the sum of the information for Regression and Residual.

**c.** A model with a large model sum of squares in comparison to the residual sum of squares indicates that the model accounts for most of variation in the dependent variable. Very high residual sum of squares indicate that the model fails to explain a lot of the variation in the dependent variable, and you may want to look for additional factors that help account for a higher proportion of the variation in the dependent variable. In this example, we see that 32.5% of the total sum of squares is made up from the regression sum of squares. You may notice that the $R^2$ for this model is also .325 (this is not a coincidence!).

**d.** If the significance value of the F statistic is small (smaller than say 0.05) then the independent variables do a good job explaining the variation in the dependent variable. If the significance value of F is larger than say 0.05 then the independent variables do not explain the variation in the dependent variable. For this example, the model does a good job explaining the variation in the dependent variable.

**e.** The "R Square" tells us how much of the variance of the dependent variable can be explained by the independent variable(s). In this example, 32.7% of the variance in frequency of sex is explained by differences in all included variables.

**f.** The standard error of the estimate (aka, the root mean square error), is the standard deviation of the error term, and is the square root of the Mean Square Residual (or Error).

**g.** The t statistics can help you determine the relative importance of each variable in the model. The t-statistic is calculated by divided the variable's unstandardized coefficient by its standard error. As a guide regarding useful predictors, look for t values well below -1.96 or above +1.96. As you can see, the race variable is the only t-value not within this range (note also that this is the only variable that is not significant).

Regression Equation

$$SEXFREQ_{predicted} = 8.298 - .055*age - 1.372*marital - .215*race - .262*happy - .067*attend$$

If we plug in values in for the independent variables (age = 35 years; marital = currently married-1; race = white-1; happiness = very happy-1; church attendance = never attends-0), we can predict a value for frequency of sex:

$$SEXFREQ_{predicted} = 8.298 - .055*35 - 1.372*1 - .215*1 - .262*1 - .067*0$$
$$= 4.524$$

As this variable is coded, a 35-year old, White, married person with high levels of happiness and who never attends church would be expected to report their frequency of sex between values 4 (weekly) and 5 (2-3 times per week).

If we plug in 70 years, instead, we find that frequency of sex is predicted at 2.599, or approximately 1-2 times per month.

Model Interpretation

*Constant* = The predicted value of "frequency of sex", when all other variables are 0. In this example, a value of 8.298 is not interpretable, since the valid responses for frequency of sex range from 0-6. Important to note, values of 0 for all variables is not interpretable either (i.e., age cannot equal 0).

*Age* = For every unit increase in age (in this case, year), frequency of sex will decrease by .055 units.

*Marital Status* = For every unit increase in marital status, frequency of sex will decrease by 1.372 units. Since marital status has only two categories, we can conclude that currently married persons have more sex than currently unmarried persons.

*Race* = For every unit increase in race, frequency of sex will decrease by .215 units. For example, the difference between non-White (0) to White (1) would be .215 units.

*Happiness* = For every unit increase in happiness, frequency of sex will decrease by .262 units. Recall that happiness is coded such that higher scores indicate less happiness. For this example, then, higher levels of happiness predict higher frequency of sex.

*Church Attendance* = For every unit increase in church attendance, frequency of sex decreases by .067 units.

**Simple Linear Regression (with Age-Squared Variable)**

It is known that some variables are often non-linear, or curvilinear. Such variables may be age or income. In this example, we include the original age variable and an age squared variable.

```
. reg sexfreq age marital racenew happy attend agesquar

      Source |       SS       df       MS              Number of obs =    1052
-------------+------------------------------           F(  6,  1045) =   85.55
       Model |  1394.24116     6  232.373526           Prob > F      =  0.0000
    Residual |  2838.48508  1045  2.71625366           R-squared     =  0.3294
-------------+------------------------------           Adj R-squared =  0.3255
       Total |  4232.72624  1051  4.02733229           Root MSE      =  1.6481

------------------------------------------------------------------------------
     sexfreq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0204952   .0172229    -1.19   0.234    -.0542906    .0133002
     marital |   -1.31425   .1105973   -11.88   0.000    -1.531268   -1.097232
     racenew |  -.2183136   .1250833    -1.75   0.081    -.4637567    .0271295
       happy |  -.2713968   .0846369    -3.21   0.001    -.4374745   -.1053191
      attend |  -.0670664   .0195946    -3.42   0.001    -.1055156   -.0286172
    agesquar |  -.0003523   .0001706    -2.06   0.039    -.0006872   -.0000175
       _cons |   7.466803   .4997854    14.94   0.000     6.486106      8.4475
------------------------------------------------------------------------------
```

The age squared variable is significant, indicating that age is non-linear.

**Simple Linear Regression (with interaction term)**

In a linear model, the effect of each independent variable is always the same. However, it could be that the effect of one variable depends on another. In this example, we might expect that the effect of age is dependent on sex. In the following example, we include an interaction term, age*sex.

To test for two-way interactions (often thought of as a relationship between an independent variable (IV) and dependent variable (DV), moderated by a third variable), first run a regression analysis, including both independent variables (IV and moderator) and their interaction (product) term. It is highly recommended that the independent variable and moderator are standardized before calculation of the product term, although this is not essential. The product term should be significant in the regression equation in order for the interaction to be interpretable.

For this example, two dummy variables were created, for ease of interpretation. Sex was recoded such that 1=Male and 0=Female. Marital status was recoded such that 1=Currently married and 0=Not currently married. The interaction term is a cross-product of these two dummy variables.

Regression Model (*without* interactions)

```
. reg sexfreq age racenew happy attend male married

      Source |       SS       df       MS              Number of obs =    1052
-------------+------------------------------           F(  6,  1045) =   87.90
       Model |  1419.69041     6  236.615068           Prob > F      =  0.0000
    Residual |  2813.03583  1045  2.69190031           R-squared     =  0.3354
-------------+------------------------------           Adj R-squared =  0.3316
       Total |  4232.72624  1051  4.02733229           Root MSE      =  1.6407

------------------------------------------------------------------------------
     sexfreq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0547746    .003028   -18.09   0.000    -.0607164   -.0488329
     racenew |  -.2307659   .1245824    -1.85   0.064    -.4752261    .0136942
       happy |  -.2421688   .0842976    -2.87   0.004    -.4075806    -.076757
      attend |  -.0562567   .0197369    -2.85   0.004    -.0949852   -.0175281
        male |   .3852511    .103874     3.71   0.000     .1814257    .5890764
     married |   1.317949   .1074847    12.26   0.000     1.107038    1.528859
       _cons |   5.295466   .2582801    20.50   0.000     4.788659    5.802273
------------------------------------------------------------------------------
```

Regression Model (*with* interactions)

```
. reg sexfreq age racenew happy attend male married interact

      Source |       SS       df       MS              Number of obs =    1052
-------------+------------------------------           F(  7,  1044) =   77.31
       Model |  1444.99234     7  206.427478           Prob > F      =  0.0000
    Residual |  2787.73389  1044  2.67024319           R-squared     =  0.3414
-------------+------------------------------           Adj R-squared =  0.3370
       Total |  4232.72624  1051  4.02733229           Root MSE      =  1.6341

------------------------------------------------------------------------------
     sexfreq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0532626   .0030556   -17.43   0.000    -.0592584   -.0472668
     racenew |  -.2625718   .1245097    -2.11   0.035    -.5068896   -.0182541
       happy |  -.2508243   .0840049    -2.99   0.003    -.4156619   -.0859866
      attend |  -.0525717   .0196938    -2.67   0.008    -.0912157   -.0139278
        male |   .6725887   .1393423     4.83   0.000     .3991658    .9460117
     married |   1.630128    .147462    11.05   0.000     1.340772    1.919484
    interact |  -.6434146   .2090208    -3.08   0.002    -1.053563   -.2332659
       _cons |   5.135983   .2624047    19.57   0.000     4.621082    5.650883
------------------------------------------------------------------------------
```

Regression Equation

$SEXFREQ_{predicted}$ = 5.136 -.053*age - .263*race -.251*happy - .053*attend + 1.630*married + (.673 - .643*married) * male

Interpretation

*Main Effects*

The married coefficient represents the main effect for females (the 0 category). The effect for females is then 1.63, or the “marital” coefficient. The effect for males is 1.63 - .643, or .987.

The gender coefficient represents the main effect for unmarried persons (the 0 category). The effect for unmarried is then .673, or the “sex” coefficient. The effect for married is .673 - .643, or .03.

*Interaction Effects*

For a simple interpretation of the interaction term, plug values into the regression equation above.

| | | |
|---|---|---|
| Married Men = | $SEXFREQ_{predicted}$ = 5.136 -.053*35 - .263*1 -.251*1 - .053*0 + 1.630**1* + (.673 - . 643**1*) * *1* | = 5.455 |
| Married Women = | $SEXFREQ_{predicted}$ = 5.136 -.053*35 - .263*1 -.251*1 - .053*0 + 1.630**1* + (.673 - . 643**1*) * *0* | = 5.425 |
| Unmarried Men = | $SEXFREQ_{predicted}$ = 5.136 -.053*35 - .263*1 -.251*1 - .053*0 + 1.630**0* + (.673 - . 643**0*) * *1* | = 3.825 |
| Unmarried Women = | $SEXFREQ_{predicted}$ = 5.136 -.053*35 - .263*1 -.251*1 - .053*0 + 1.630**0* + (.673 - . 643**0*) * *0* | = 3.795 |

In this example (age = 35 years; race = white-1; happiness = very happy-1; church attendance = never attends-0), we can see that (1) for both married and unmarried persons, males are reporting higher frequency of sex than females, and (2) married persons report higher frequency of sex than unmarried persons. The interaction tells us that the gender difference is greater for married persons than for unmarried persons.

## Logistic Regression

Logistic regression is a variation of the regression model. It is used when the dependent response variable is binary in nature. Logistic regression predicts the probability of the dependent response, rather than the value of the response (as in simple linear regression).

In this example, the dependent variable is frequency of sex (less than once per month versus more than once per month).

```
. logistic freqdum age marital racenew attend happy

Logistic regression                               Number of obs   =       1052
                                                  LR chi2(5)      =     302.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -567.40591                       Pseudo R2       =     0.2102b

------------------------------------------------------------------------------
     freqdum | Odds Ratioa  Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9408818   .0047102   -12.17   0.000     .9316952     .950159
     marital |     .17496   .0290688   -10.49   0.000     .1263327    .2423048
     racenew |   .8854209   .1564181    -0.69   0.491      .626292    1.251765
      attend |   .9309624   .0262984    -2.53   0.011     .8808194    .9839599
       happy |   .7156043    .087177    -2.75   0.006     .5636079    .9085918
------------------------------------------------------------------------------
```

**a.** The “Odds Ratio” is the predicted change in odds for a unit increase in the predictor. When the Odds Ratio is less than 1, increasing values of the variable correspond to decreasing odds of the event's occurrence. When the Odds Ratio is greater than 1, increasing values of the variable correspond to increasing odds of the event's occurrence.

If you subtract 1 from the odds ratio and multiply by 100, you get the percent change in odds of the dependent variable having a value of 1. For example, for age:

= 1 - (.941) = .059

= .059 * 100 = 5.9%

The odds ratio for age indicates that every unit increase in age is associated with a 5.9% decrease in the odds of having sex more than once a month.

**b.** The R-Square statistic cannot be exactly computed for logistic regression models, so these approximations are computed instead. Larger pseudo r-square statistics indicate that more of the variation is explained by the model, to a maximum of 1.

Interpretation

Recall: When Exp(B) is less than 1, increasing values of the variable correspond to decreasing odds of the event's occurrence. When Exp(B) is greater than 1, increasing values of the variable correspond to increasing odds of the event's occurrence.

*Constant* = Not interpretable in logistic regression.

*Age* = Increasing values of age correspond with decreasing odds of having sex more than once a month.

*Marital* = Increasing values of marital status (married to unmarried) correspond with decreasing odds of having sex more than once a month.

*Race* = Increasing values of race correspond with increasing odds of having sex more than once a month. Notice that this variable, however, is not significant.

*Church Attendance* = Increasing values of church attendance correspond with decreasing odds of having sex more than once a month.

*Happiness* = Increasing values of general happiness correspond with decreasing odds of having sex more than once a month. Recall that happiness is coded such that higher values indicate less happiness.

**Logistic Regression (with non-linear variables)**

It is known that some variables are often non-linear, or curvilinear. Such variables may be age or income. In this example, we include the original age variable and an age squared variable.

```
. logistic freqdum age marital racenew attend happy agesquar

Logistic regression                                 Number of obs   =       1052
                                                    LR chi2(6)      =     314.11
                                                    Prob > chi2     =     0.0000
Log likelihood = -561.40465                         Pseudo R2       =     0.2186

------------------------------------------------------------------------------
     freqdum | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.029322   .0274165     1.09   0.278     .9769654    1.084485
     marital |   .1876667   .0319891    -9.82   0.000     .1343674    .2621081
     racenew |   .8854594   .1559968    -0.69   0.490     .6269129    1.250633
      attend |   .9311651   .0264615    -2.51   0.012     .8807195    .9845002
       happy |   .7026621    .085654    -2.89   0.004     .5533319    .8922928
    agesquar |   .9990557   .0002795    -3.38   0.001      .998508    .9996037
------------------------------------------------------------------------------
```

The age squared variable is significant, indicating that age is non-linear.

**Logistic Regression (with interaction term)**

To test for two-way interactions (often thought of as a relationship between an independent variable (IV) and dependent variable (DV), moderated by a third variable), first run a regression analysis, including both independent variables (IV and moderator) and their interaction (product) term. It is highly recommended that the independent variable and moderator are standardized before calculation of the product term, although this is not essential. For this example, two dummy variables were created, for ease of interpretation. Sex was recoded such that 1=Male and 0=Female. Marital status was recoded such that 1=Currently married and 0=Not currently married. The interaction term is a product of these two dummy variables.

Regression Model (without interactions)

```
. logistic freqdum age racenew happy attend male married

Logistic regression                               Number of obs   =       1052
                                                  LR chi2(6)      =     311.10
                                                  Prob > chi2     =     0.0000
Log likelihood = -562.91057                       Pseudo R2       =     0.2165

------------------------------------------------------------------------------
     freqdum | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9407224   .0047616   -12.07   0.000     .9314359    .9501014
     racenew |   .8615719   .1535544    -0.84   0.403     .6075547    1.221793
       happy |   .7273111   .0893135    -2.59   0.010     .5717327    .9252252
      attend |   .9422746   .0269726    -2.08   0.038     .8908649     .996651
        male |   1.558383   .2310822     2.99   0.003     1.165347    2.083978
     married |   5.464756   .9139101    10.16   0.000     3.937479    7.584435
------------------------------------------------------------------------------
```

Regression Model (with interactions)

```
. logistic freqdum age racenew happy attend male married interact

Logistic regression                               Number of obs   =       1052
                                                  LR chi2(7)      =     313.87
                                                  Prob > chi2     =     0.0000
Log likelihood =  -561.5232                       Pseudo R2       =     0.2184

------------------------------------------------------------------------------
     freqdum | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9417151   .0047935   -11.80   0.000     .9323667    .9511572
     racenew |   .8412302   .1508375    -0.96   0.335      .591956    1.195474
       happy |   .7244743   .0892968    -2.61   0.009     .5689921    .9224436
      attend |   .9452654   .0271617    -1.96   0.050     .8935009    1.000029
        male |   1.913216   .3698852     3.36   0.001     1.309784    2.794655
     married |    6.93103   1.542039     8.70   0.000     4.481461    10.71953
    interact |   .6043344   .1827447    -1.67   0.096     .3341046    1.093131
------------------------------------------------------------------------------
```

The product term should be significant in the regression equation in order for the interaction to be interpretable. In this example, the interaction term is significant at the 0.1 level.

Interpretation

*Main Effects*

The married coefficient represents the main effect for females (the 0 category). The effect for females is then 1.94, or the "marital" coefficient. The effect for males is 1.94 - .50, or 1.44.

The gender coefficient represents the main effect for unmarried persons (the 0 category). The effect for unmarried is then .65, or the "sex" coefficient. The effect for married is .65 - .50, or .15.

*Odds Ratios*

Using "married" as the focus variable, we can say that the effect of being married on having sex more than once per month is greater for females.

Females: $e^{1.936} = 6.93$

Males: $e^{1.432} = 4.20$

Using "gender" as the focus variable, we can say that the effect of being male on having sex more than once per month is greater for marrieds.

Marrieds: $e^{0.15} = 1.16$

Unmarrieds: $e^{0.65} = 1.92$