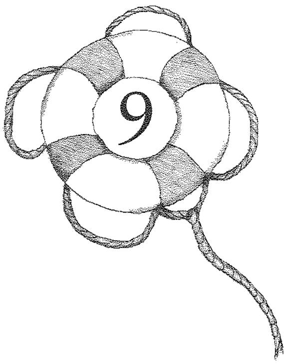


Presenting your Results

In this chapter, we will take you through some techniques for presenting your results to a scientific or policy audience. This part of research is given far less attention than it should be. It is one thing to produce a bunch of statistical findings – it is quite another to be able to communicate them clearly to your peers and to a non-technical audience. It takes quite a bit of practice to be able to do this effectively.

We will use a worked example throughout this chapter to illustrate how we go about starting with a research question, developing hypotheses around the research question, presenting statistical results, and creating discussion about the results of our statistical tests. Throughout the worked example, we are using a different data set than in previous chapters but still from the first (1991) wave of the British Household Panel Survey, which can be obtained from the UK Data Archive (www.data-archive.ac.uk).

DECIDING ON A RESEARCH QUESTION

The first thing that you need to do when you are undertaking research is to decide on a feasible research question. The feasibility of a research question depends on many things, including your interests, the time and funding that you have, and your skill set. You may have to narrow a very wide topic to something more specific if you have a limited amount of time in which to produce results. Or you may have to modify a research question if you don't have the analytic skills to answer your original question (provided you don't have the time or desire to learn the new skills whilst answering your research question!).

Suppose you are interested in attitudes towards gender roles. A very broad research question would be, 'What determines attitudes towards gender roles?' But in reality, it is likely that the answer to this question would require very detailed information about individuals' environments when they were growing up, as well as detailed information about their parents' beliefs and behaviours. It may be the case that you don't have such detailed information. You may want to narrow your research question to something more specific, such as, 'How do adults' characteristics influence their attitudes about gender roles?'

REVIEWING THE LITERATURE

Before undertaking any study, it important to first review the existing literature on your general topic. Conducting a literature review is beyond the scope of this book, but one of the authors has written about this task elsewhere (see Neuman and Robson 2008). A quick search would show you that Burt and Scott (2002), Fortin (2005) and McDaniel (2008) have all published studies that examine gender role attitudes. Careful review of these articles and others would help you become familiar with theories in this area of study and the findings of these studies would assist you in developing testable hypotheses.

DEVELOPING HYPOTHESES

Hypotheses come from three general places: theory, previous research, and exploration. Theory and previous research can obviously guide your expectations about what you might find in your data. In many cases, however, researchers working in new areas may not have previous research or a suitable theory to draw from and therefore might undertake exploratory analysis to uncover patterns in the data. Sometimes scientists use hypotheses that are derived from a combination of theory and research, and also have additional hypotheses that are exploratory.

From the literature, we would be able to make the following hypotheses:

H1: Women will have more liberal gender role attitudes than men.
H2: Younger people will have more liberal gender role attitudes than older people.
H3: Married people will have less liberal gender role attitudes than other marital statuses.

H4: Religious people will have less liberal gender role attitudes than non-religious people.
H5: Education will be positively associated with liberal gender role attitudes.
H6: Ethnic minorities, particularly Asians, will be less liberal than White respondents.
H7: There will be an interaction between sex and income such that there is a stronger positive association between income and liberal gender role attitudes for women.

We will also include an exploratory hypothesis to test in our analyses:

H8: There will be an interaction between sex and marital status on gender role attitudes.

EXPLORING THE DATA AND SELECTING MEASURES

Our analyses here are going to be based on the same survey that we have been using for the earlier chapters of the book. In 'real life', you may have a choice of data sets from which you could select the data on which you want to test your hypotheses. Or you may have collected your own data for the express purpose of answering a set of research questions.

From the above research questions, we will need measures of: gender role attitudes, sex, age, ethnicity, marital status, religiosity, education, and income.

Recall from Chapter 3 that we constructed a scale that assessed attitudes towards gender roles. After some analysis, it was determined that the following items would be kept in the scale:

opfama: A pre-school child is likely to suffer if his or her mother works.
opfamb: All in all, family life suffers when the woman has a full-time job.
opfamc: A woman and her family would all be happier if she goes out to work.
opfamd: Both the husband and wife should both contribute to the household income.

opfame: Having a full-time job is the best way for a woman to be an independent person.
opfamf: A husband's job is to earn money; a wife's job is to look after the home and family.
opfamh: Employers should make special arrangements to help mothers combine jobs and childcare.
opfami: A single parent can bring up children as well as a couple.

The response categories for all items were: 1, strongly agree; 2, agree; 3, neither agree nor disagree; 4, disagree; and 5, strongly disagree. We reverse coded items *opfamc, opfamd, opfame, opfamh* and *opfami* and then added all the items together to give a scale with a minimum of 8 and a maximum of 40. People who score 8 express very conservative attitudes, while those around the 40 mark would be very liberal. We could have used command **alpha**, but, as shown in Chapter 3, it rescales the variables and their values become less intuitive (although it does produce mathematically the same scale). In this chapter, we will call this variable *genderroles*.

We know from previous chapters that we have a dummy variable that measures sex called *female*, a variable measuring age called *age*, and a marital status variable *mastat*, and a variable that measures monthly income called *fimn*. We also have a seven-category variable that measures education, called *qfachi*, as well as a dummy variable that indicates if a respondent was active in a religious group, which is called *activerel*.

As this example involves adults' characteristics we restrict the sample to those 18 and over.

```
keep if age>17
```

UNIVARIATE ANALYSIS

Before undertaking a detailed analysis, it is important that we get our hands dirty with the data and really get familiar with the variables of interest. All analyses should begin at the univariate (i.e. one-variable) level. We cannot overemphasize that it is important to get to know your variables before you throw them into more complex analyses. In real life, you should also be aware of any sampling issues that are present in your data (i.e. do you need to include any weights to adjust for sampling?). Don't forget to specify your missing values!

We can check our variables by running a **summarize** command on our dichotomous and interval variables:

su genderroles female age fimn activerel

```
. su genderroles female age fimn activerel

    Variable |      Obs       Mean    Std. Dev.     Min       Max
-------------+----------------------------------------------------
 genderroles |     9188   25.50424    4.768196        8        40
      female |     9920   .5333669    .4989106        0         1
         age |     9920   45.49526    18.02041       18        97
        fimn |     9582   758.1702    742.0371        0     11297
   activerel |     9572   .1019641    .3026169        0         1
```

We can see that the mean of *genderroles* is 25.50 with a standard deviation of 4.77. The sample is about 53% female, and the average age is 45.50 years. As well, the average monthly income (*fimn*) is £758.17 and about 10% of the respondents are actively involved in religious groups (*activerel*).

We will now tabulate the categorical variables in our data set.

```
. ta mastat

-> tabulation of mastat

  marital status |      Freq.     Percent        Cum.
-----------------+-----------------------------------
         married |      6,009       60.57       60.57
living as couple |        670        6.75       67.33
         widowed |        866        8.73       76.06
        divorced |        434        4.38       80.43
       separated |        189        1.91       82.34
   never married |      1,752       17.66      100.00
-----------------+-----------------------------------
           Total |      9,920      100.00
```

We can see that the majority of sample members are married (just over 60%), with the next largest category being never married (about 18%).

```
. ta qfachi

highest academic |
   qualification |      Freq.     Percent        Cum.
-----------------+-----------------------------------
  higher degree  |        122        1.28        1.28
      1st degree |        598        6.25        7.53
hnd,hnc,teaching |        496        5.18       12.71
         a level |      1,349       14.10       26.81
         o level |      2,320       24.25       51.06
             cse |        469        4.90       55.96
   none of these |      4,213       44.04      100.00
-----------------+-----------------------------------
           Total |      9,567      100.00
```

The largest category in the variable measuring highest academic qualification (*qfachi*) is 'none of these', which can be interpreted as having only compulsory schooling or less. Just over 7% of the sample had a university degree or higher.

```
. ta race

    ethnic group |
      membership |      Freq.     Percent        Cum.
-----------------+-----------------------------------
           white |      9,196       96.15       96.15
     black-carib |         65        0.68       96.83
   black-african |         42        0.44       97.27
     black-other |         26        0.27       97.54
          indian |         99        1.04       98.58
       pakistani |         42        0.44       99.02
     bangladeshi |          6        0.06       99.08
         chinese |          9        0.09       99.17
other ethnic grp |         79        0.83      100.00
-----------------+-----------------------------------
           Total |      9,564      100.00
```

In terms of ethnic group membership (*race*), we can see that over 96% of the sample is White. Some of the categories, like 'Black other', 'Bangladeshi', and 'Chinese', are also very small: 26, 6, and 9, respectively. We need to think if there are ways of collapsing the categories so that we do not have problems with this variable later. If we try to make a number of dummy variables out of this variable the way it is currently coded, we will run into problems with the smaller groups – they will be associated with a lot of 'error' (indicated by large standard errors), or the estimation techniques will simply kick them out of the estimation procedure due to collinearity problems.

There are always debates around how to 'best' collapse ethnic group categories, and there is no one best way. Here, we are going to group all the 'Black' categories together, create a single group for Indian, Pakistani, and Bangladeshi called 'Asian', and group 'Chinese' with 'other'. Of course, the Asian group masks the differences between Muslim and non-Muslim Asians and creating the single category 'Black' also loses the major cultural differences between Caribbean Blacks and African Blacks. Also putting Chinese with 'Other' simply loses the uniqueness of the Chinese in a very heterogeneous and basically undefined group. But in real-life research, such decisions must be made.

```
gen race2=race
recode race2 3=2 4=2 5=3 6=3 7=3 8=4 9=4
lab def race2 1 "white" 2 "black" 3 "asian" ///
   4 "other"
lab val race2 race2
```

As a final step, we check our new variable to make sure the recoding was done properly.

```
tab race2
```

```
. tab race2

      race2 |      Freq.     Percent        Cum.
------------+-----------------------------------
      white |      9,196       96.15       96.15
      black |        133        1.39       97.54
      asian |        147        1.54       99.08
      other |         88        0.92      100.00
------------+-----------------------------------
      Total |      9,564      100.00
```

When reading academic articles and reports, the first table that you often see is a table of descriptive statistics. It is a good idea to make such a table to give the reader some indication of the characteristics of your sample. However, notice that in the previous tables, the N differs quite a bit. For *age* and *female*, there are 10,264 observations, but for *genderroles* there are 9515 and for *activerel* there are 9902.

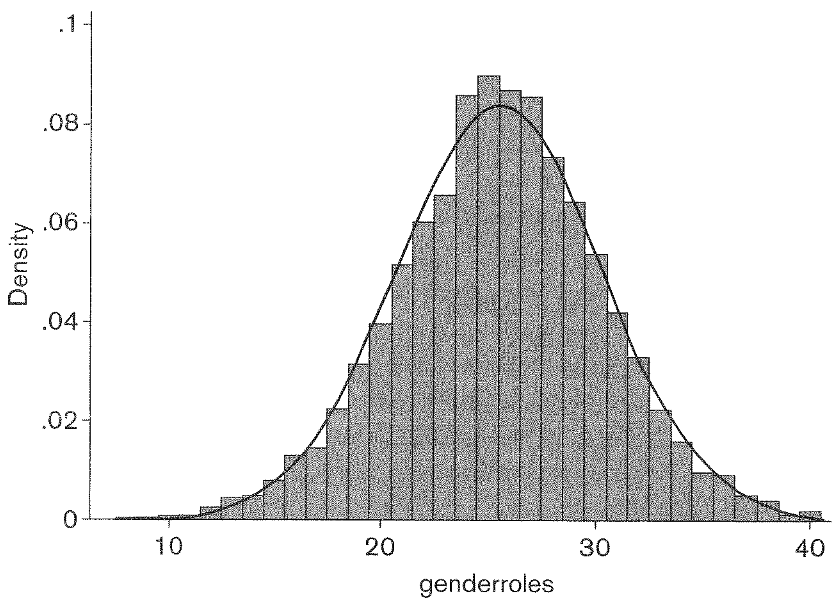
Why are the numbers of observations different for the variables? This is due to people not answering survey items. As the variable *genderroles* is a composite score of eight items, a person had to have answered every one of the eight items to be included in the scale – at least that is how we constructed it here (see Chapter 3 for alternative techniques that allow for individuals to be missing on one or more of the items). And some people simply don't like to answer certain types of questions, such as those concerning income or religious beliefs.

Because of these differing *N* sizes, it is better to wait to produce our final table of descriptive statistics (i.e. the one to include in our report) until after we have done our multivariate estimations. This is because after we do regressions, we can get the descriptive statistics for our 'estimation sample' – that is, the subsample of cases that have data on all our variables of interest.

We also need to check the distribution of our dependent variable so that we know its properties for when we want to conduct multivariate analyses. We can check this visually with the **histogram** command.

```
histogram genderroles, discrete
```

```
. histogram genderroles, discrete
(start=8, width=1)
```



From the histogram we can see that our dependent variable of interest is reasonably normally distributed. It is surprising to see how 'normal' it is, as it is remarkable how few interval variables (in our experience at least) display such tidy distributive characteristics.

BIVARIATE TESTS

Before we directly test our hypotheses, we should undertake some bivariate tests. One of the assumptions of many multivariate techniques is that the independent variables are not highly correlated with one another. We can check this assumption with the **corr** command.

We have two categorical variables in our analysis – *qfachi* and *mastat*. We cannot simply correlate these variables with the others because the numbers associated with their categories are nominal. We need to convert these variables to sets of dummy variables. We can do this with the **xi:** command.

```
xi: su i.qfachi i.mastat i.race2
```

```
. xi: su i.qfachi i.mastat i.race2
i.qfachi   _Iqfachi_1-7   (naturally coded; _Iqfachi_1 omitted)
i.mastat   _Imastat_1-6   (naturally coded; _Imastat_1 omitted)
i.race2    _Irace2_1-4    (naturally coded; _Irace2_1 omitted)

  Variable |     Obs        Mean    Std. Dev.      Min     Max
-----------+--------------------------------------------------
_Iqfachi_2 |    9567    .0625065    .2420859         0       1
_Iqfachi_3 |    9567    .0518449    .2217253         0       1
_Iqfachi_4 |    9567    .1410055    .3480455         0       1
_Iqfachi_5 |    9567    .2425003    .4286176         0       1
_Iqfachi_6 |    9567    .0490227    .2159267         0       1
-----------+--------------------------------------------------
_Iqfachi_7 |    9567    .4403679    .4964572         0       1
_Imastat_2 |    9920    .0675403    .2509681         0       1
_Imastat_3 |    9920    .0872984     .282286         0       1
_Imastat_4 |    9920      .04375    .2045487         0       1
_Imastat_5 |    9920    .0190524    .1367162         0       1
-----------+--------------------------------------------------
_Imastat_6 |    9920    .1766129      .38136         0       1
 _Irace2_2 |    9564    .0139063    .1171083         0       1
 _Irace2_3 |    9564    .0153701    .1230263         0       1
 _Irace2_4 |    9564    .0092012    .0954854         0       1
```

You can see that this has created a set of dummies for *qfachi* and *mastat*. This process by default drops the lowest coded variable as the reference category. However, for the **corr** command, we will need all categories for *qfachi*, *mastat*, and *race2* to be dummy coded. We can make the ones corresponding to category 1 for each variable manually. In the case of *mastat*, category 1 corresponds to being married; for *qfachi*, respondents in category 1 have a higher degree; for *race2*, category 1 corresponds to being White.

```
gen married=mastat==1
gen higherdeg=qfachi==1
gen white=race2==1
```

Now we can create a correlation matrix (partial table shown) for all the variables:

```
corr genderroles age female married _Im* ///
  higherdeg _Iq* fimn activerel _Irace2* white
```

```
. corr genderroles age female married _Im* ///
    higherdeg _Iq* fimn activerel _Irace2* white
(obs=9163)

             | gender~s      age   female married _Imast~2 _Imast~3  _Imast~4
-------------+-----------------------------------------------------------------
 genderroles |   1.0000
         age |  -0.3263   1.0000
      female |   0.1472   0.0426   1.0000
     married |  -0.1487   0.1280  -0.0549   1.0000
   _Imastat_2 |   0.1181  -0.1952  -0.0160  -0.3419   1.0000
   _Imastat_3 |  -0.0950   0.4483   0.1626  -0.3749  -0.0815   1.0000
   _Imastat_4 |   0.0410   0.0232   0.0505  -0.2710  -0.0589  -0.0646   1.0000
   _Imastat_5 |   0.0424  -0.0256   0.0472  -0.1760  -0.0383  -0.0420  -0.0303
```

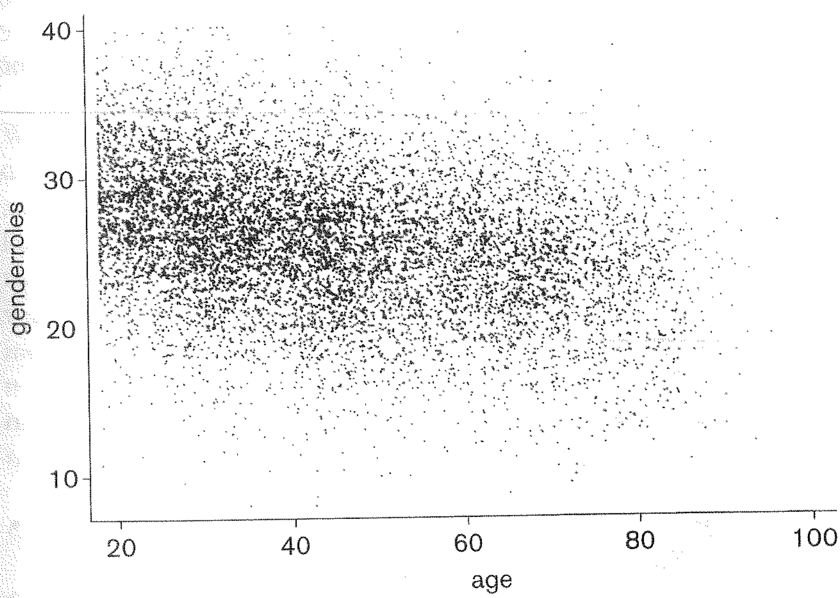
In the correlation matrix, we look for correlations that are higher than about 0.60. We want variables to be correlated with the dependent variable. Because we put *genderroles* first in our list of variables, the correlates with it will be in the first column. What we are trying to spot is if the correlations between our independent variables are of concern. In the full matrix (not presented) we observe that having low education (*_Iqfac~7*) is quite strongly correlated with *age* (0.4877). Of course, the categories of a variable converted to dummies will be correlated with each other, often quite highly. In this example being *White* and being Asian (*_Irace~3*) are correlated at −0.6211 (not shown). Apart from these unavoidable correlations between the dummies, there is nothing that raises alarm in this correlation matrix.

If there were a large correlation of, say, 0.70 between *age* and *fimn*, for example, we would have to make a decision about dropping one of these variables as multivariate estimation techniques would not be able to properly capture the individual effects of variables that are so highly correlated. Substantively, they are obviously different, but if they are correlated so highly that they are not 'mathematically' different enough for Stata (or any other software program, or even hand calculation for that matter!) to tell them apart.

Note that it was quite 'in fashion' to publish correlation matrices up until about 10 years ago. Now it is very rarely done. If you are writing a technical report, you may want to include such a matrix in an Appendix, but nowadays it is rarely a main part of a scholarly social sciences paper.

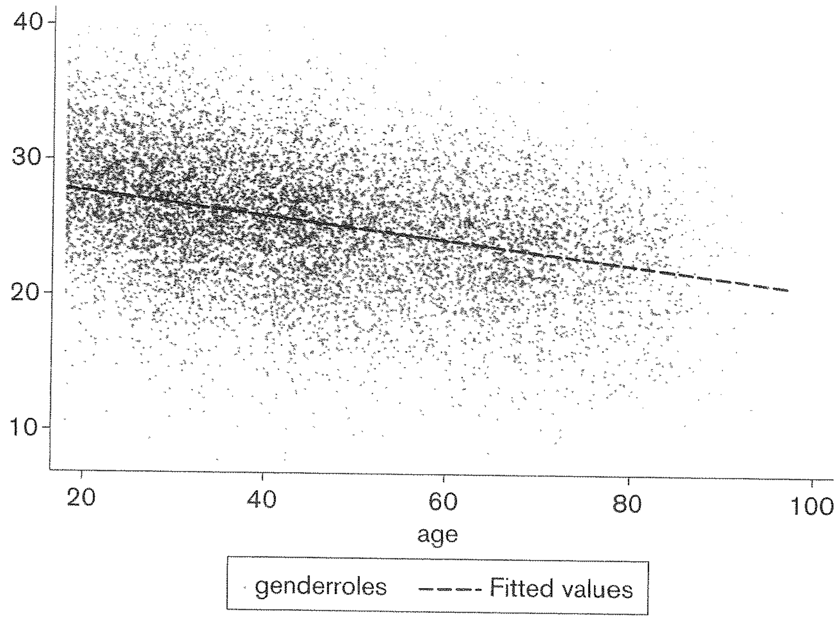
We conclude our bivariate tests with some scatterplots.

```
scatter genderroles age, msymbol(point) jitter(3)
```



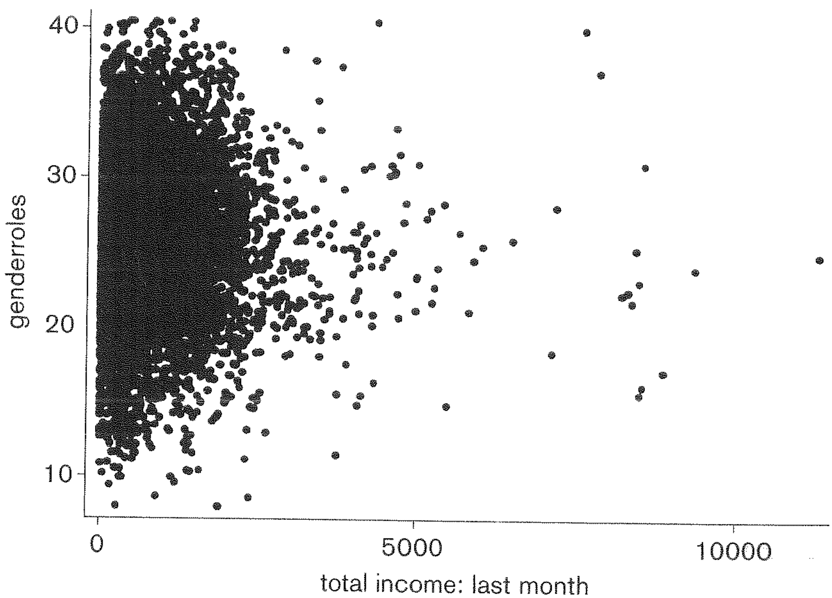
The scatterplot between *genderroles* and *age* reveals that there is evidence of a downward negative linear association. We can also add a linear fit line to display this association:

```
scatter genderroles age, msymbol(point) ///
   jitter(3) || lfit genderroles age
```



Now we look at the relationship between *genderroles* and total income (*fimn*):

```
scatter genderroles fimn
```



Immediately you can see that the association between these two variables does not look strongly linear. And there are a number of outliers – one in particular in excess of £10,000. We need to examine this income variable more closely.

su fimn, detail

```
. su fimn, detail

            total income: last month
-------------------------------------------------
   Percentiles   Smallest
 1%        0          0
 5%       56          0
10%   134.23          0    Obs              9582
25%   281.45          0    Sum of Wgt.      9582
50%      550               Mean         758.1702
                  Largest  Std. Dev.    742.0371
75% 1023.333   8628.875
90% 1602.562   8716.667   Variance       550619
95%  2006.54   9455.773   Skewness     3.412333
99% 3447.667      11297   Kurtosis     27.19502
```

We know that income variables are often highly skewed – and the details provided from the **su** output reveal this. The mean (758) and median (550) are very different, and the skewness (3.41) and kurtosis (27.19) are also very large. In the previous chapter, we took the natural logarithm of income to help normalize it, as it was our dependent variable. We could do this, but transforming income as an independent variable is not often done because it is not as important when it is an dependent variable. The robustness of the multivariate tests that are commonly used can cope with non-normally distributed independent variables.

What we should do, however, is examine what happens if we eliminate one of the bigger outliers:

corr genderroles fimn

```
. corr genderroles fimn
(obs=9188)

             |  gender~s     fimn
-------------+------------------
 genderroles |    1.0000
        fimn |    0.0278   1.0000
```

corr genderroles fimn if fimn<10000

```
. corr genderroles fimn if fimn<10000
(obs=9187)

             | gender~s     fimn
-------------+------------------
 genderroles |   1.0000
        fimn |   0.0282   1.0000
```

The correlations reveal that removal of the outlier only improves the correlation coefficient by 0.004 – not very much. What if we limit the sample to those who reported income which was at or below the 75th percentile?

corr genderroles fimn if fimn<1023

```
. corr genderroles fimn if fimn<1023
(obs=6839)

             | gender~s     fimn
-------------+------------------
 genderroles |   1.0000
        fimn |   0.1085   1.0000
```

The correlation on a subsample of those with incomes at or below the 75th percentile improves the strength of the association significantly. So, let's see if we take those below the 95th percentile.

corr genderroles fimn if fimn<2006

```
. corr genderroles fimn if fimn<2006
(obs=8716)

             | gender~s     fimn
-------------+------------------
 genderroles |   1.0000
        fimn |   0.0926   1.0000
```

The correlation drops, but only slightly. Therefore, it appears that those 472 or so cases with incomes above £2006 per month are suppressing the association for the majority of the sample. There are some advanced methods that allow you to model these changes in slopes such as splines, but they are beyond the scope of this book. For this example we will further restrict our sample to those with incomes below the 95th percentile. But before doing this it is worth seeing how many cases will be dropped. We know from the correlations that we will lose about 470 cases who have non-missing values on *genderroles* and *fimn*, but if we use a **keep if fimn<2006** command then we will also drop those missing (.) on *fimn*. Remember that Stata stores missing values (.) as a very large number.

ta fimn if missing(fimn), miss

```
. ta fimn if missing(fimn), miss

  total income: |
     last month |      Freq.     Percent        Cum.
----------------+-----------------------------------
              . |        338      100.00      100.00
----------------+-----------------------------------
          Total |        338      100.00
```

Or: **count if missing (fimn)**

```
. count if missing(fimn)
338
```

keep if fimn<2006

```
. keep if fimn<2006
(819 observations deleted)
```

This shows us that we had 338 cases missing on *fimn* so when we use the **keep** command we can see that Stata has deleted those plus those under £2006 per month income for a total of 819 cases.

MULTIVARIATE TESTS

As discussed in Chapter 8, there are a multitude of multivariate tests to choose from. You need to pick the one that fits with your hypotheses and the nature of your data. We are testing causal relationships (i.e. that a variety of characteristics influence attitudes about gender roles). Our dependent variable is normally distributed. Therefore ordinary least squares regression is a suitable tool for testing our hypotheses, and we include two interaction terms (Jaccard and Turrisi 2003); see Box 9.1.

Box 9.1: The treatment of dummy variables in interaction terms

We should mention that we have to put **i.** in front of **female** so that Stata knows that it is a categorical variable. However, because dummy variables are a special case of categorical variables (see Chapter 3), if we were interacting two dummies (e.g. *female* and *activerel*), we could just type **i.female*activrel** instead of **i.female*i.activerel**. By way of example, we will show you both ways of doing it.

xi: regress genderroles i.female*i.activerel

```
. xi: regress genderroles i.female*i.activerel
i.female          _Ifemale_0-1       (naturally coded; _Ifemale_0 omitted)
i.activerel       _Iactiverel_0-1    (naturally coded; _Iactiverel_0 omitted)
i.fem~e*i.act~l   _IfemXact_#_#      (coded as above)

      Source |       SS       df       MS              Number of obs =    8712
-------------+------------------------------           F(  3,  8708) =   93.68
       Model |  6160.01762     3  2053.33921           Prob > F      =  0.0000
    Residual |  190871.233  8708  21.9190667           R-squared     =  0.0313
-------------+------------------------------           Adj R-squared =  0.0309
       Total |   197031.25  8711  22.6186718           Root MSE      =  4.6818

------------------------------------------------------------------------------
 genderroles |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ifemale_1 |   1.377705   .1061882    12.97   0.000     1.169551    1.585858
 _Iactivere~1 | -2.025562   .2970706    -6.82   0.000     -2.60789   -1.443233
 _IfemXact_~1 |  .3858559   .3601664     1.07   0.284    -.3201554    1.091867
       _cons |   24.94691   .0784668   317.93   0.000      24.7931    25.10072
------------------------------------------------------------------------------
```

xi: regress genderroles i.female*activerel

```
. xi: regress genderroles i.female*activerel
i.female          _Ifemale_0-1       (naturally coded; _Ifemale_0 omitted)
i.female*acti~l   _IfemXactiv_#      (coded as above)

      Source |       SS       df       MS              Number of obs =    8712
-------------+------------------------------           F(  3,  8708) =   93.68
       Model |  6160.01762     3  2053.33921           Prob > F      =  0.0000
    Residual |  190871.233  8708  21.9190667           R-squared     =  0.0313
-------------+------------------------------           Adj R-squared =  0.0309
       Total |   197031.25  8711  22.6186718           Root MSE      =  4.6818

------------------------------------------------------------------------------
 genderroles |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ifemale_1 |   1.377705   .1061882    12.97   0.000     1.169551    1.585858
   activerel |  -2.025562   .2970706    -6.82   0.000     -2.60789   -1.443233
 _IfemXacti~1 |  .3858559   .3601664     1.07   0.284    -.3201554    1.091867
       _cons |   24.94691   .0784668   317.93   0.000      24.7931    25.10072
------------------------------------------------------------------------------
```

You can see that the results are identical.

One advantage of using dummy variables as 'interval' variables (as one is essentially doing when dropping the **i.**) is that when you have several interactions with one dummy variable, say *female*, you are not given redundant results.

Suppose we were to interact *female* with marital status, age, and education in our estimation of their impact on *genderroles*.

xi: regress genderroles i.female*age ///
i.female*i.mastat i.female*i.qfachi

```
. xi: regress genderroles i.female*age ///
>     i.female*i.mastat i.female*i.qfachi
i.female          _Ifemale_0-1         (naturally coded; _Ifemale_0 omitted)
i.female*age      _IfemXage_#          (coded as above)
i.mastat          _Imastat_1-6         (naturally coded; _Imastat_1 omitted)
i.fem~e*i.mas~t   _IfemXmas_#_#        (coded as above)
i.qfachi          _Iqfachi_1-7         (naturally coded; _Iqfachi_7 omitted)
i.fem~e*i.qfa~i   _IfemXqfa_#_#        (coded as above)

      Source |       SS       df       MS              Number of obs =    8703
-------------+------------------------------           F( 25,  8677) =   62.78
       Model |  30147.0345    25  1205.88138           Prob > F      =  0.0000
    Residual |  166667.822  8677  19.2080007           R-squared     =  0.1532
-------------+------------------------------           Adj R-squared =  0.1507
       Total |  196814.856  8702  22.6171979           Root MSE      =  4.3827

------------------------------------------------------------------------------
 genderroles |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Ifemale_1 |   1.475481   .4216284     3.50   0.000     .6489888    2.301973
         age |  -.0840634   .0053219   -15.80   0.000    -.0944955   -.0736313
 _IfemXage_1 |  -.0069617    .007236    -0.96   0.336    -.0211461    .0072226
  _Ifemale_1 |  (dropped)
  _Imastat_2 |   1.736397   .2829492     6.14   0.000     1.181749    2.291045
  _Imastat_3 |   .0771449   .4010154     0.19   0.847    -.7089406    .8632303
  _Imastat_4 |   .9312998   .3949254     2.36   0.018     .1571523    1.705447
  _Imastat_5 |   .3175018   .6461857     0.49   0.623    -.9491755    1.584179
  _Imastat_6 |    .581275   .1986193     2.93   0.003      .191934     .970616
_IfemXmas_~2 |  -.7865847   .3891229    -2.02   0.043    -1.549358   -.0238114
_IfemXmas_~3 |   .9940145   .4604245     2.16   0.031     .0914732    1.896556
_IfemXmas_~4 |   .5622266   .4858379     1.16   0.247     -.390131    1.514584
_IfemXmas_~5 |   1.352576   .7614352     1.78   0.076    -.1400176     2.84517
_IfemXmas_~6 |   .5004244   .2780129     1.80   0.072    -.0445469    1.045396
  _Ifemale_1 |  (dropped)
  _Iqfachi_1 |   1.649272   .6709021     2.46   0.014     .3341446    2.964399
  _Iqfachi_2 |   1.199016   .3157355     3.80   0.000     .5800992    1.817932
  _Iqfachi_3 |   -.307903   .3245798    -0.95   0.343    -.9441565    .3283504
  _Iqfachi_4 |  -.2654178   .2197034    -1.21   0.227    -.6960887    .1652531
  _Iqfachi_5 |  -.3738898   .1966644    -1.90   0.057    -.7593987     .011619
  _Iqfachi_6 |   -.858516   .3636598    -2.36   0.018    -1.571376   -.1456564
_IfemXqfa_~1 |   .1745363    1.05091     0.17   0.868    -1.885497     2.23457
_IfemXqfa_~2 |  -.1261066   .4392844    -0.29   0.774    -.9872082     .734995
_IfemXqfa_~3 |   .8151526   .4492686     1.81   0.070    -.0655206    1.695826
_IfemXqfa_~4 |   .2182148   .3177161     0.69   0.492    -.4045842    .8410138
_IfemXqfa_~5 |   .2559116   .2588053     0.99   0.323    -.2514082    .7632313
_IfemXqfa_~6 |   .0685341   .4747771     0.14   0.885    -.8621418      .99921
       _cons |    28.3311   .3177356    89.17   0.000     27.70826    28.95394
------------------------------------------------------------------------------
```

You can see from the output that there are two spaces in the output that say 'dropped'. That is because the main effect of *female* had already been added to the estimation. As there is not a **xi:** option that allows us to specify that the main effects of a categorical variable interacting with a categorical variable (**i.var1*i.var2**) are not reported, we must think of something else.

As dummy variables have some of the 'properties' of interval variables we can use the **xi:** formats **i.var1*var2** as well as **i.var1|var2**. In both of these formats, the second variable is expected to be interval. In the first format, **i.var1*var2**, all interactions as well as the main effects of both variables are reported. In the second format, **i.var1|var2**, all interactions between *var1* (categorical) and *var2* (interval) are created, with the main effects of *var1* not being reported. We can use *female* as our interval variable because it is a dummy and use the **|** option so that Stata doesn't try to enter the main effects of *female* numerous times.

```
xi: regress genderroles i.female|age ///
  i.mastat*female i.qfachi*female
```

```
. xi: regress genderroles i.female|age ///
>     i.mastat*female i.qfachi*female
i.female          _Ifemale_0-1        (naturally coded; _Ifemale_0 omitted)
i.female|age      _IfemXage_#         (coded as above)
i.mastat          _Imastat_1-6        (naturally coded; _Imastat_1 omitted)
i.mastat*female   _ImasXfemal_#       (coded as above)
i.qfachi          _Iqfachi_1-7        (naturally coded; _Iqfachi_7 omitted)
i.qfachi*female   _IqfaXfemal_#  (coded as above)

      Source |       SS       df       MS              Number of obs =    8703
-------------+------------------------------           F( 25,  8677) =   62.78
       Model |  30147.0345    25  1205.88138           Prob > F      =  0.0000
    Residual |  166667.822  8677  19.2080007           R-squared     =  0.1532
-------------+------------------------------           Adj R-squared =  0.1507
       Total |  196814.856  8702  22.6171979           Root MSE      =  4.3827

------------------------------------------------------------------------------
 genderroles |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0840634   .0053219   -15.80   0.000    -.0944955   -.0736313
  _IfemXage_1 |  -.0069617    .007236    -0.96   0.336    -.0211461    .0072226
  _Imastat_2 |   1.736397   .2829492     6.14   0.000     1.181749    2.291045
  _Imastat_3 |   .0771449   .4010154     0.19   0.847    -.7089406    .8632303
  _Imastat_4 |   .9312998   .3949254     2.36   0.018     .1571523    1.705447
  _Imastat_5 |   .3175018   .6461857     0.49   0.623    -.9491755    1.584179
  _Imastat_6 |    .581275   .1986193     2.93   0.003      .191934     .970616
      female |   1.475481   .4216284     3.50   0.000     .6489888    2.301973
_ImasXfema~2 |  -.7865847   .3891229    -2.02   0.043    -1.549358   -.0238114
_ImasXfema~3 |   .9940145   .4604245     2.16   0.031     .0914732    1.896556
```

```
_ImasXfema~4 |   .5622266   .4858379    1.16  0.247    -.390131   1.514584
_ImasXfema~5 |   1.352576   .7614352    1.78  0.076   -.1400176    2.84517
_ImasXfema~6 |   .5004244   .2780129    1.80  0.072   -.0445469   1.045396
  _Iqfachi_1 |   1.649272   .6709021    2.46  0.014    .3341446   2.964399
  _Iqfachi_2 |   1.199016   .3157355    3.80  0.000    .5800992   1.817932
  _Iqfachi_3 |   -.307903   .3245798   -0.95  0.343   -.9441565   .3283504
  _Iqfachi_4 |  -.2654178   .2197034   -1.21  0.227   -.6960887   .1652531
  _Iqfachi_5 |  -.3738898   .1966644   -1.90  0.057   -.7593987    .011619
  _Iqfachi_6 |   -.858516   .3636598   -2.36  0.018   -1.571376  -.1456564
_IqfaXfema~1 |   .1745363    1.05091    0.17  0.868   -1.885497    2.23457
_IqfaXfema~2 |  -.1261066   .4392844   -0.29  0.774   -.9872082    .734995
_IqfaXfema~3 |   .8151526   .4492686    1.81  0.070   -.0655206   1.695826
_IqfaXfema~4 |   .2182148   .3177161    0.69  0.492   -.4045842   .8410138
_IqfaXfema~5 |   .2559116   .2588053    0.99  0.323   -.2514082   .7632313
_IqfaXfema~6 |   .0685341   .4747771    0.14  0.885   -.8621418     .99921
       _cons |    28.3311   .3177356   89.17  0.000    27.70826   28.95394
------------------------------------------------------------------------------
```

You can see here that the results are much tidier, with no 'dropped' messages. This solves the problem of how to tidy up your output with an interaction of a dummy variable with numerous predictors. If the variable you want to interact with numerous other predictors is a categorical variable with numerous categories, however, there is no 'quick fix' to get Stata to stop inserting the main effects several times. There is no 'error' in the results – you will just have to remove the 'dropped' comments manually when you are creating your tables.

To test all of our hypotheses in one model, we can use the command:

**xi: regress genderroles age i.female*i.mastat ///
i.qfachi i.female|fimn activerel i.race2**

```
. xi: regress genderroles age i.female*i.mastat ///
>   i.qfachi i.female|fimn activerel i.race2
i.female          _Ifemale_0-1     (naturally coded; _Ifemale_0 omitted)
i.mastat          _Imastat_1-6     (naturally coded; _Imastat_1 omitted)
i.fem~e*i.mas~t   _IfemXmas_#_#    (coded as above)
i.qfachi          _Iqfachi_1-7     (naturally coded; _Iqfachi_1 omitted)
i.female|fimn     _IfemXfimn_#     (coded as above)
i.race2           _Irace2_1-4      (naturally coded; _Irace2_1 omitted)

      Source |       SS     df       MS         Number of obs =    8692
-------------+--------------------------        F( 24,  8667) =   77.31
       Model | 34656.0981    24 1444.00409      Prob > F      =  0.0000
    Residual | 161881.671 8667  18.677936       R-squared     =  0.1763
-------------+--------------------------        Adj R-squared =  0.1741
       Total | 196537.769 8691 22.6139419       Root MSE      =  4.3218
```

```
-------------------------------------------------------------------------------
 genderroles |      Coef.  Std. Err.       t   P>|t|   [95% Conf. Interval]
-------------+-----------------------------------------------------------------
         age |  -.0827785   .0036312  -22.80   0.000   -.0898965  -.0756605
  _Ifemale_1 |   .5136454   .1953023    2.63   0.009    .1308066   .8964842
  _Imastat_2 |   1.616538   .2740168    5.90   0.000      1.0794   2.153676
  _Imastat_3 |   .0765492   .3888046    0.20   0.844   -.6856002   .8386987
  _Imastat_4 |   .9034001   .3914352    2.31   0.021    .1360941   1.670706
  _Imastat_5 |   .2874111   .6373313    0.45   0.652   -.9619098   1.536732
  _Imastat_6 |   .5797909   .1911909    3.03   0.002    .2050114   .9545705
_IfemXmas_~2 |  -1.063059   .3728356   -2.85   0.004   -1.793905  -.3322125
_IfemXmas_~3 |   .8186293    .428668    1.91   0.056   -.0216619    1.65892
_IfemXmas_~4 |   .0240032   .4828787    0.05   0.960   -.9225538   .9705602
_IfemXmas_~5 |    .893311   .7512081    1.19   0.234   -.5792355   2.365857
_IfemXmas_~6 |   .2944977   .2572233    1.14   0.252    -.209721   .7987164
  _Iqfachi_2 |  -.6707918   .5413836   -1.24   0.215   -1.732032   .3904489
  _Iqfachi_3 |  -1.627406   .5445569   -2.99   0.003   -2.694867  -.5599454
  _Iqfachi_4 |  -1.673665   .5217301   -3.21   0.001    -2.69638  -.6509502
  _Iqfachi_5 |  -1.764963   .5151843   -3.43   0.001   -2.774847   .7550791
  _Iqfachi_6 |  -2.160175   .5498084   -3.93   0.000    -3.23793   1.082419
  _Iqfachi_7 |  -1.433279   .5141618   -2.79   0.005   -2.441158   .4253994
        fimn |   .0000712   .0001491    0.48   0.633    -.000221   .0003634
_IfemXfimn_1 |   .0019329    .000216    8.95   0.000    .0015094   .0023563
  activerel  |  -1.305594   .1581752   -8.25   0.000   -1.615655   .9955332
   _Irace2_2 |   1.495979   .4201054    3.56   0.000    .6724727   2.319486
   _Irace2_3 |  -1.820171   .3985591   -4.57   0.000   -2.601442   1.038901
   _Irace2_4 |  -.6684578   .4874699   -1.37   0.170   -1.624015   .2870991
       _cons |    29.7689   .5729856   51.95   0.000    28.64571   30.89209
-------------------------------------------------------------------------------
```

The information at the top of the output tells us that category 1 in *mastat* has been omitted, as have the categories 1 for *qfachi* and 1 for *race2*. These correspond to being married, having a higher degree, and being White. If the categories that are omitted seem reasonable for testing our hypotheses, you can just leave them. But if it seems more logical to change the reference category, use the **char** command. One of our hypotheses is that education will be positively associated with liberal gender role attitudes, so it probably makes more sense to have the lowest education coded as the reference category, because if support for our hypothesis is found with the default reference category, our coefficients will all be negative – which isn't 'wrong', but less intuitive.

```
char qfachi [omit] 7
xi: regress genderroles age i.female*i.mastat ///
   i.qfachi i.female|fimn activerel i.race2
```

```
. char qfachi [omit] 7
. xi: regress genderroles age i.female*i.mastat ///
>   i.qfachi i.female|fimn activerel i.race2
i.female          _Ifemale_0-1      (naturally coded; _Ifemale_0 omitted)
i.mastat          _Imastat_1-6      (naturally coded; _Imastat_1 omitted)
i.fem~e*i.mas~t   _IfemXmas_#_#     (coded as above)
i.qfachi          _Iqfachi_1-7      (naturally coded; _Iqfachi_7 omitted)
i.female|fimn     _IfemXfimn_#      (coded as above)
i.race2           _Irace2_1-4       (naturally coded; _Irace2_1 omitted)

      Source |       SS       df       MS              Number of obs =    8692
-------------+------------------------------           F( 24,  8667) =   77.31
       Model |  34656.0981    24  1444.00409           Prob > F      =  0.0000
    Residual |  161881.671  8667   18.677936           R-squared     =  0.1763
-------------+------------------------------           Adj R-squared =  0.1741
       Total |  196537.769  8691  22.6139419           Root MSE      =  4.3218

------------------------------------------------------------------------------
 genderroles |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0827785   .0036312   -22.80   0.000    -.0898965   -.0756605
  _Ifemale_1 |   .5136454   .1953023     2.63   0.009     .1308066    .8964842
  _Imastat_2 |   1.616538   .2740168     5.90   0.000       1.0794    2.153676
  _Imastat_3 |   .0765492   .3888046     0.20   0.844    -.6856002    .8386987
  _Imastat_4 |   .9034001   .3914352     2.31   0.021     .1360941    1.670706
  _Imastat_5 |   .2874111   .6373313     0.45   0.652    -.9619098    1.536732
  _Imastat_6 |   .5797909   .1911909     3.03   0.002     .2050114    .9545705
_IfemXmas_~2 |  -1.063059   .3728356    -2.85   0.004    -1.793905   -.3322125
_IfemXmas_~3 |   .8186293    .428668     1.91   0.056    -.0216619     1.65892
_IfemXmas_~4 |   .0240032   .4828787     0.05   0.960    -.9225538    .9705602
_IfemXmas_~5 |    .893311   .7512081     1.19   0.234    -.5792355    2.365857
_IfemXmas_~6 |   .2944977   .2572233     1.14   0.252     -.209721    .7987164
 _Iqfachi_1  |   1.433279   .5141618     2.79   0.005     .4253994    2.441158
 _Iqfachi_2  |   .7624871   .2247183     3.39   0.001     .3219858    1.202988
 _Iqfachi_3  |  -.1941275   .2278045    -0.85   0.394    -.6406785    .2524234
 _Iqfachi_4  |  -.2403863   .1581545    -1.52   0.129    -.5504066    .0696341
 _Iqfachi_5  |   -.331684   .1274664    -2.60   0.009    -.5815485   -.0818196
 _Iqfachi_6  |  -.7268959   .2308039    -3.15   0.002    -1.179326   -.2744653
        fimn |   .0000712   .0001491     0.48   0.633     -.000221    .0003634
_IfemXfimn_1 |   .0019329    .000216     8.95   0.000     .0015094    .0023563
   activerel |  -1.305594   .1581752    -8.25   0.000    -1.615655   -.9955332
   _Irace2_2 |   1.495979   .4201054     3.56   0.000     .6724727    2.319486
   _Irace2_3 |  -1.820171   .3985591    -4.57   0.000    -2.601442   -1.038901
   _Irace2_4 |  -.6684578   .4874699    -1.37   0.170    -1.624015    .2870991
       _cons |   28.33562   .2710218   104.55   0.000     27.80435    28.86689
------------------------------------------------------------------------------
```

Let's go through the results. We can see from our adjusted R-squared that our variables explain just over 17% of the variance in gender roles. It isn't brilliant, but it isn't bad either.

We will use $p < 0.05$ to determine statistically significant effects. In terms of *age*, the effect is statistically significant. Each additional year of age reduces a person's score on *genderroles* by −0.083, independent of the effects of the other variables. Compared to males, being female is associated with a 0.514 increase on *genderroles*.

For marital status all the coefficients are relative to the omitted category of married. We need to remember that the Stata-generated dummies for this variable have suffixes that correspond to the original value labels: *_Imastat_2* corresponds to 'living as a couple', *_Imastat_3* to 'widowed', *_Imastat_4* to 'divorced', *_Imastat_5* to 'separated', and *_Imastat_6* to 'never married.' Compared to married people, living as a couple was associated with a 1.617 increase in *genderroles*. Being divorced, compared to being married, was associated with a 0.903 increase in *genderroles*. Being single, compared to being married, was also associated with a 0.580 increase in *genderroles*. The other categories were not significantly different than married at the 0.05 level.

The next lines in our output correspond to the exploratory test of the interaction between sex and marital status. We can see that three of the interactions are statistically significant. The omitted category here is married males (those who are omitted on both *female* and *mastat*), so all the results are relative to this group. Significant interactions tell that that the slopes are significantly different for the groups under consideration and cannot be interpreted literally. So the significant interaction between *female* and living as a couple (*_IfemXmas_~2*) means that there is a statistically significant difference in the effect of living as a couple on attitudes towards gender roles between men and women. We can't be entirely certain of what that difference is without additional analyses, which we will cover later in this chapter.

In terms of educational attainment, there are four statistically significant coefficients, which are relevant to the 'none of these' category on *qfachi*, which we interpret as being largely comprised of those with only compulsory schooling. We can see that having a higher degree (*_Iqfachi_1*) or a university degree (*_Iqfachi_2*), compared to having only compulsory schooling, are associated with a 1.433 and 0.762 increase in the *genderroles* measure, respectively. The other two statistically significant results are for having O levels (compulsory school leaving age qualifications) and CSE (Certificate of Secondary Education) relative to having only compulsory schooling. Both, however, have negative coefficients, suggesting that those who have these marginal qualifications are less liberal compared to those with only compulsory schooling (or less). This may have to do with older people being more likely to be in these categories (O levels, for example, have been replaced by an alternative qualification). This may be something that a researcher would want to explore in future research. You could create a three way interaction between *age*, *female*, and *qfachi*, add it to the model and see what happens.

The overall effect of income (*fimn*) was positive and non-significant, with each unit (£1) increase in income being associated with an increase in *genderroles* of 0.00007. This is a very small coefficient, but this is due to the way the variable *fimn* is measured – in pounds.

The interaction between *female* and *fimn* (*_IfemXfimn_1*) was significant, though. This suggests that as females earn more, they tend to have higher scores on *genderroles*, and that this is significantly different from the effect that *fimn* has on males' gender role attitudes. We will explore this interaction in more depth later in this chapter.

The relationship between being active in a religious group (*activerel*) and *genderroles* is in the expected direction. Compared to respondents who were not active in a religious group, being active in a religious group was associated with a 1.306 decrease in *genderroles*. Finally, in terms of ethnic group membership, we find that relative to Whites, being Black (*_Irace2_2*) is associated with a 1.496 increase in *genderroles*, being Asian (*_Irace2_3*) with a 1.820 decrease in *genderroles* and no significant difference for 'other' (*_Irace2_4*).

Let's review our hypotheses and see what we can determine so far:

H1: Women will have more liberal gender role attitudes than men: supported.
H2: Younger people will have more liberal gender role attitudes than men: supported.
H3: Married people will have less liberal gender role attitudes than other marital statuses: supported.
H4: Religious people will have less liberal gender role attitudes than non-religious people: supported.
H5: Education will be positively associated with liberal gender role attitudes: somewhat supported.
H6: Ethnic minorities, particularly Asians, will be less liberal than White respondents: somewhat supported.
H7: There will be an interaction between sex and income such that there is a stronger positive association between income and liberal gender role attitudes for women: supported.
H8: There will be an interaction between sex and marital status on gender role attitudes: supported.

We have found at least some support for all of our hypotheses. In real life, it is very rare to find support for all your hypotheses. Even when you fail to find support for your hypotheses, such a 'non-finding' can be a finding in itself (but again, in 'real life' it is actually quite difficult to publish such findings, unfortunately).

If we are satisfied with our models and don't want to make any further adjustments, we can now start thinking about making tables and graphs for our paper.

Making tables

Recall that, earlier in the chapter, we urged you to not make tables of descriptive statistics until you have run your final model and have an 'estimation sample'. We can get descriptive statistics now using the **`if e(sample)`** option or by keeping only the cases in the final regression by using the command **`keep if e(sample)`**. It is very important to note that the following command must be used immediately after the final regression you are using, because it is only the last estimates that are stored in memory.

```
gen noqual=qfachi==7
xi: su genderroles age female i.mastat married ///
   i.qfachi noqual fimn activerel i.race2 white if e(sample)
```

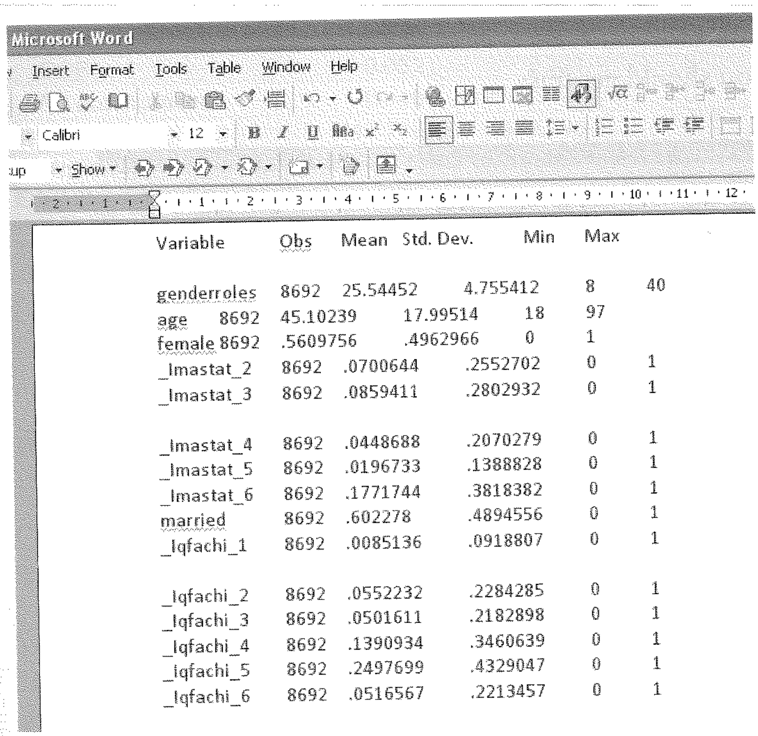
```
. gen noqual=qfachi==7
. xi: su genderroles age female i.mastat married ///
>    i.qfachi noqual fimn activerel i.race2 white if e(sample)
i.mastat     _Imastat_1-6    (naturally coded; _Imastat_1 omitted)
i.qfachi     _Iqfachi_1-7    (naturally coded; _Iqfachi_7 omitted)
i.race2      _Irace2_1-4     (naturally coded; _Irace2_1 omitted)

    Variable |    Obs       Mean    Std. Dev.    Min    Max
-------------+---------------------------------------------
 genderroles |   8692   25.54452    4.755412      8     40
         age |   8692   45.10239    17.99514     18     97
      female |   8692   .5609756    .4962966      0      1
  _Imastat_2 |   8692   .0700644    .2552702      0      1
  _Imastat_3 |   8692   .0859411    .2802932      0      1
-------------+---------------------------------------------
  _Imastat_4 |   8692   .0448688    .2070279      0      1
  _Imastat_5 |   8692   .0196733    .1388828      0      1
  _Imastat_6 |   8692   .1771744    .3818382      0      1
     married |   8692    .602278    .4894556      0      1
  _Iqfachi_1 |   8692   .0085136    .0918807      0      1
```

```
-------------+--------------------------------------------------------
  _Iqfachi_2 |      8692    .0552232    .2284285          0          1
  _Iqfachi_3 |      8692    .0501611    .2182898          0          1
  _Iqfachi_4 |      8692    .1390934    .3460639          0          1
  _Iqfachi_5 |      8692    .2497699    .4329047          0          1
  _Iqfachi_6 |      8692    .0516567    .2213457          0          1
-------------+--------------------------------------------------------
      noqual |      8692    .4455821    .4970585          0          1
        fimn |      8692    649.7487    479.7897          0       2005
  activerel  |      8692    .0998619    .2998331          0          1
  _Irace2_2  |      8692    .0125403    .1112854          0          1
  _Irace2_3  |      8692    .0139208    .1171693          0          1
-------------+--------------------------------------------------------
  _Irace2_4  |      8692    .0092039    .0954998          0          1
      white  |      8692     .964335    .1854641          0          1
```

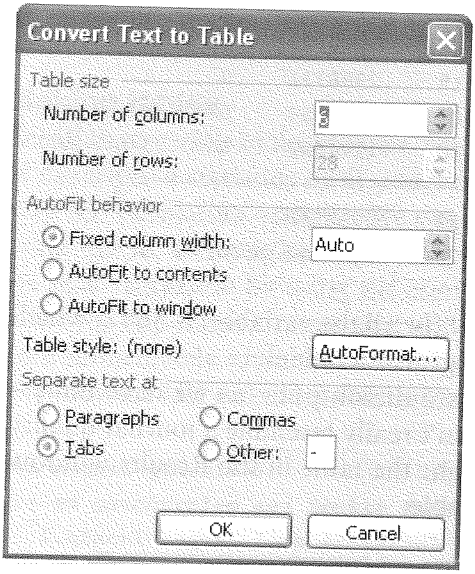
You will notice now that the *N* for all the variables is 8692, which is the same as the *N* for our regression model.

While you can cut and paste the output – as we have done – into a Word document, it doesn't really resemble a journal-quality table. You will have to highlight the table in the Results window and then go to **Edit → Copy table**.

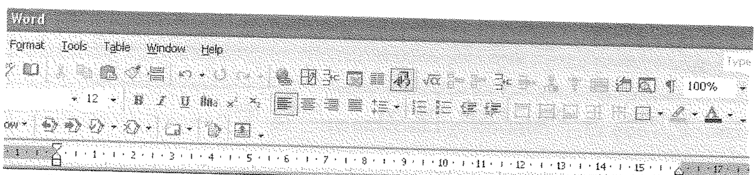


Variable	Obs	Mean	Std. Dev.	Min	Max
genderroles	8692	25.54452	4.755412	8	40
age	8692	45.10239	17.99514	18	97
female	8692	.5609756	.4962966	0	1
_Imastat_2	8692	.0700644	.2552702	0	1
_Imastat_3	8692	.0859411	.2802932	0	1
_Imastat_4	8692	.0448688	.2070279	0	1
_Imastat_5	8692	.0196733	.1388828	0	1
_Imastat_6	8692	.1771744	.3818382	0	1
married	8692	.602278	.4894556	0	1
_Iqfachi_1	8692	.0085136	.0918807	0	1
_Iqfachi_2	8692	.0552232	.2284285	0	1
_Iqfachi_3	8692	.0501611	.2182898	0	1
_Iqfachi_4	8692	.1390934	.3460639	0	1
_Iqfachi_5	8692	.2497699	.4329047	0	1
_Iqfachi_6	8692	.0516567	.2213457	0	1

Now open a blank Word document and press **Paste**. The result will be ugly. Now, in Word, select the contents of the table then go to **Table** → **Convert Text to Table** and the following dialogue box will appear. More often than not, it is 'smart' enough to guess the number of columns and rows that you want. Check that these are correct then click **OK**.



Then you should get this:



Variable	Obs	Mean	Std. Dev.	Min	Max
genderroles	8692	25.54452	4.755412	8	40
age	8692	45.10239	17.99514	18	97
female	8692	.5609756	.4962966	0	1
_Imastat_2	8692	.0700644	.2552702	0	1
_Imastat_3	8692	.0859411	.2802932	0	1
_Imastat_4	8692	.0448688	.2070279	0	1
_Imastat_5	8692	.0196733	.1388828	0	1
_Imastat_6	8692	.1771744	.3818382	0	1
married	8692	.602278	.4894556	0	1
_Iqfachi_1	8692	.0085136	.0918807	0	1
_Iqfachi_2	8692	.0552232	.2284285	0	1
_Iqfachi_3	8692	.0501611	.2182898	0	1
_Iqfachi_4	8692	.1390934	.3460639	0	1
_Iqfachi_5	8692	.2497699	.4329047	0	1
_Iqfachi_6	8692	.0516567	.2213457	0	1

Some may prefer to copy the table into Excel to do the initial formatting rather than Word. With a bit of tidying up, we have the following table:

Table of descriptive statistics ($N = 8692$)

Variable	Mean	Std. Dev.	Min	Max
Dependent variable				
Gender roles	25.545	4.755	8	40
Independent variables				
Age	45.102	17.995	18	97
Female	0.561	0.496	0	1
Marital status				
Living as a couple	0.070	0.255	0	1
Widowed	0.086	0.280	0	1
Divorced	0.045	0.207	0	1
Separated	0.020	0.139	0	1
Single	0.177	0.382	0	1
Married	0.602	0.489	0	1
Educational attainment				
Higher degree	0.009	0.092	0	1
University degree	0.055	0.228	0	1
HND, HNC, teaching	0.050	0.218	0	1
A levels	0.139	0.346	0	1
O levels	0.250	0.433	0	1
CSE	0.052	0.221	0	1
No qualifications	0.446	0.497	0	1
Income	649.749	479.790	0	2005
Active in religious group	0.100	0.300	0	1
Ethnicity				
Black	0.013	0.111	0	1
Asian	0.014	0.117	0	1
Other	0.009	0.095	0	1
White	0.964	0.185	0	1

The next table you will want to produce is a table of your regression findings. Let us run the regression again (without displaying the output) to make sure it is the most recent thing in Stata's memory.

```
xi: regress genderroles age i.female*i.mastat ///
    i.qfachi i.female|fimn activerel i.race2
```

We will now use the command **esttab** to make a regression table. You might have to install it first. If so, type **findit esttab** and follow the instructions. There is also a useful online tutorial by the author of **esttab** at http://repec.org/bocode/e/estout/index.html (Jann 2005, 2007).

If you just type **esttab** after the regression, you will get the following, partial, output in your Results window:

```
_Iqfachi_3          -0.194
                   (-0.85)

_Iqfachi_4          -0.240
                   (-1.52)

_Iqfachi_5          -0.332**
                   (-2.60)

_Iqfachi_6          -0.727**
                   (-3.15)

fimn             0.0000712
                    (0.48)

_Ifemxfimn_1       0.00193***
                    (8.95)

activerel           -1.306***
                   (-8.25)

_Irace2_2            1.496***
                    (3.56)

_Irace2_3           -1.820***
                   (-4.57)

_Irace2_4           -0.668
                   (-1.37)

_cons                28.34***
                  (104.55)
------------------------------
N                     8692
------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

As you can see, the results have the unstandardized coefficient and the *t* statistic (in parentheses). At the bottom of the output, there is a note about the *t* statistics being in parentheses and that the stars correspond to * $p<0.05$, ** $p<0.01$, *** $p<0.001$. If you copy this to an Excel spreadsheet, you can edit it from there.

We can tell Stata to make some adjustments to what is displayed. Note that there are far too many options with the **esttab** command to discuss here in any detail. You really need to look at the help menu for this and related commands to truly customize your tables to your liking and to the requirements of specific disciplines and journals.

esttab, label se ar2, using regression1.rtf

Here, we have requested that variable labels (**label**) be printed instead of the value labels, that the standard errors (**se**) be reported instead of *t* statistics, and that the adjusted R^2 (**ar2**) is reported. When you use the option **using**, the resulting table is saved as a file in your active directory. Also, when you run the command, a message is returned:

```
. esttab, label se ar2, using regression1.rtf
(output written to regression1.rtf)
```

The file `regression1.rtf` is an active link (usually shown in blue) – clicking on it automatically opens up the .rtf (rich text format) document in Word. You should note that you can save the file in many different formats – we are just using .rtf documents to keep things simple. When you click on the active link, your document will open. Again, you will have to tidy it up and add the proper variable names where you used Stata-generated dummy variables. After some tidying up, we have a table that looks like this:

Regression of attitudes towards gender roles on various individual characteristics

	b
Age	−0.0828***
	(0.00363)
Female	0.514**
	(0.195)
Marital status (ref: Married)	
Living as a couple	1.617***
	(0.274)
Widowed	0.0765
	(0.389)
Divorced	0.903*
	(0.391)
Separated	0.287
	(0.637)
Never married, single	0.580**
	(0.191)
Female*Living as a couple	−1.063**
	(0.373)

Female*Widowed	0.819
	(0.429)
Female*Divorced	0.0240
	(0.483)
Female*Separated	0.893
	(0.751)
Female*Never married	0.294
	(0.257)
Educational attainment (ref: No qualifications)	
Higher degree	1.433**
	(0.514)
University degree	0.762***
	(0.225)
HND, HNC, teaching	−0.194
	(0.228)
A levels	−0.240
	(0.158)
O levels	−0.332**
	(0.127)
CSE	−0.727**
	(0.231)
Income	0.0000712
	(0.000149)
Female*Income	0.00193***
	(0.000216)
Active in religious group	−1.306***
	(0.158)
Ethnicity (ref: White)	
Black	1.496***
	(0.420)
Asian	−1.820***
	(0.399)
Other	−0.668
	(0.487)
Constant	28.34***
	(0.271)
Observations	8692
Adjusted R^2	0.174

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Understanding interactions

We have two statistically significant interactions: between *female* and a category of *mastat* and between *female* and income (*fimn*). It is often useful, particularly if your audience is non-technical, to give more information about what your interaction means.

Because both of our interactions are with *female*, what the regression coefficients are telling us is that the slopes for males and females on the categorical variables are significantly different from one another. One useful way of getting to the bottom of the interaction is to run the model separately for the variables in the interaction term. So, for example, we can run the models separately for men and women. Instead of just pasting the output for the separate estimations below, we are going to save the results and make a table with both of them using the **estimates store** and **esttab** command.

First, we run a regression for only females (remembering to take out the interactions)

```
xi: regress genderroles age i.mastat ///
   i.qfachi fimn activerel i.race2 if female==1
```

We then get Stata to store these results as a model called 'female'.

```
estimates store female
```

We then run the same model on males:

```
xi: regress genderroles age i.mastat ///
   i.qfachi fimn activerel i.race2 if female==0
```

We store the results as a model called 'male'.

```
estimates store male
```

We then use **esttab** to create a table with results by requesting that models 'female' and 'male' be displayed, with variable labels (**label**), standard errors (**se**), adjusted R^2 (**ar2**), and with only the set of marital status variables (***mastat***) and income (**fimn**) displayed using the **keep** option (as these are the coefficients we are interested in comparing). We write **mtitles** so that each model is given the name we stated above (i.e. 'female' and 'male'). If we don't specify it, the dependent variable would appear instead. We are using the option **replace** in case we want to rerun the models for whatever reason. This option overwrites any existing files with the same name. If we wanted to fix any mistakes and we hadn't written **replace**, Stata would return the following message:

```
file interaction.rtf already exists
r(602);
```

```
esttab female male, label se ar2 ///
   keep(*mastat* fimn) mtitles, ///
   using interaction.rtf, replace
```

```
. esttab female male, label se ar2 ///
   keep(*mastat* fimn) mtitles, ///
   (using interaction.rtf, replace
(output written to interaction.rtf)
```

After clicking on the active link, we obtain the following table:

	(1) female	(2) male
mastat==2	0.574* (0.268)	1.566*** (0.277)
mastat==3	0.858*** (0.227)	0.0882 (0.390)
mastat==4	0.890** (0.287)	0.896* (0.385)
mastat==5	1.182** (0.404)	0.262 (0.627)
mastat==6	0.878*** (0.196)	0.537** (0.203)
total income: last month	0.00211*** (0.000177)	0.000000752 (0.000152)
Observations	4876	3816
Adjusted R^2	0.165	0.150

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Graphing interactions

Interactions presented in a table of regression results are difficult to interpret and not very intuitive. So it is useful to visually display what they are telling you. Understanding and graphing interaction terms between a categorical variable and an interval variable is considerably easier than getting to grips with an interaction between two categorical variables. So, we'll start with the interaction between sex and income.

First run the estimation:

```
xi: regress genderroles age i.female*i.mastat ///
   i.qfachi i.female|fimn activerel i.race2
```

Then request predicted/fitted values of *genderroles*:

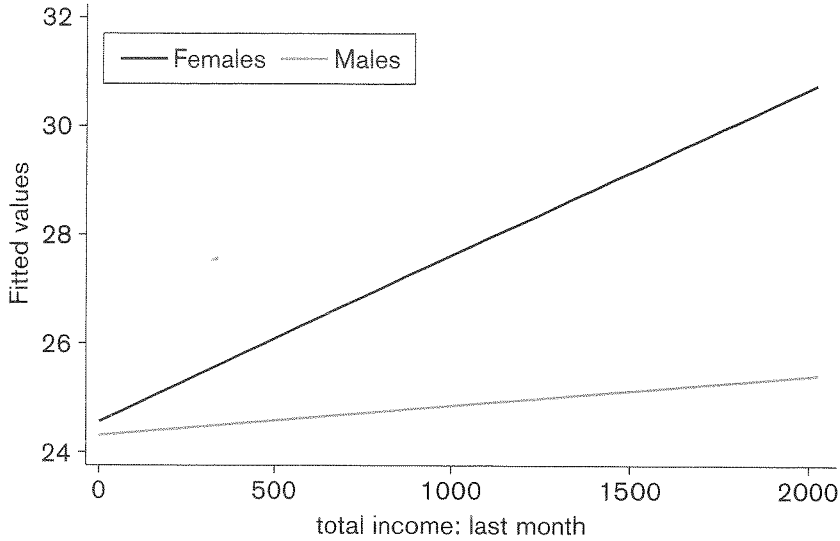
```
predict xb
```

Now we show you two different ways to graph the interaction. The graphs are slightly different, but both show substantively that, for women, as income increases so too do their liberal gender role attitudes, whereas for men there is little, if any, effect, as shown by the flat line.

First, we graph the predicted values (*xb*) against income (*fimn*) separately for males and females, using a linear fit (**lfit**) graph. We use a linear fit so that that a single line is presented. Not using this option and simply requesting a line graph would result in a crazy looking graph resembling a large scribble. The **legend** option tells Stata to label the lines as 'Females' and 'Males'.

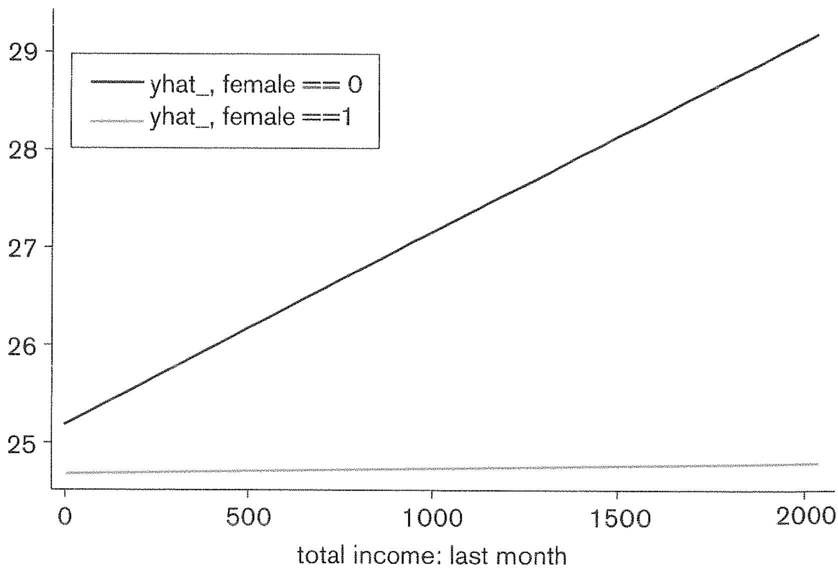
```
twoway (lfit xb fimn if female==1) ///
   (lfit xb fimn if female==0), ///
   legend(order (1 "Females" 2 "Males"))
```

Alternatively, we could use the **xi3** and **postgr3** combination of commands introduced in Chapter 8. One issue with this method is that to get accurate predictions of the regression model a variable can only be in one interaction term. In our previous model *female* was in two interaction terms so, for this example, we remove the **i.female*i.mastat** term and run the



regression model. The **postgr3** command produces the graph below which tells us pretty much the same as the one above.

```
xi3: regress genderroles age i.mastat ///
   i.qfachi i.female*fimn activerel i.race2
postgr3 fimn,by(female)
```



Box 9.2: Doing commands 'quietly'

If you want to rerun a regression just to make sure you have the right estimates in the memory but don't want to see the results you can prefix the command with **qui:** which is short for **quietly**. The **quietly** prefix can be used with commands other than **regress**. You can see from the example command below that it is possible to use the **qui:** prefix before the **xi3:** prefix.

```
qui: xi3: regress genderroles age ///
   i.female*i.mastat i.qfachi fimn activerel ///
   i.race2
```

Recall, from earlier in the chapter, that it was marital status categories 2 (living as a couple), 3 (widowed), and 6 (never married) that were significantly different from 1 (married). We can see from the table on p. 350 that the effect of living as a couple on *genderroles* for males is over twice the size of the effect for females. Being widowed, on the other hand, has a very large effect for women (0.858), but not statistically significant effect for men. Similarly, being separated has a larger effect on women (1.182) than men (0.262).

While the coefficients for these marital statuses look rather different for males and females, the lack of their statistical significance in an interaction suggests that their slopes are not significantly different from one another. This is likely to be due to the smaller number of cases when you break down the separated and divorced categories by sex:

```
ta mastat sex if e(sample)
```

```
. ta mastat sex if e(sample)

                  |          sex
   marital status |      male   female |     Total
------------------+--------------------+----------
          married |     2,370    2,865 |     5,235
living as couple  |       292      317 |       609
          widowed |       139      608 |       747
         divorced |       129      261 |       390
        separated |        47      124 |       171
    never married |       839      701 |     1,540
------------------+--------------------+----------
            Total |     3,816    4,876 |     8,692
```

Interactions between dummy coded categorical variables are quite tricky to understand and even more tricky to present in a meaningful way. This is mainly because the regression coefficients are not really 'slopes' but differences between groups. An interaction between a categorical variable and an interval variable, as above, clearly shows a difference in the slopes for the effect of income on gender roles for men and women. Rather than talk about differences in differences, we shall call the coefficients for the marital status categories 'slopes', and so we are still looking for differences in slopes but with the added complication that all the dummy variable coefficients are relative to the same category, in this case those who are married. Let's look at the relevant part of the regression results again:

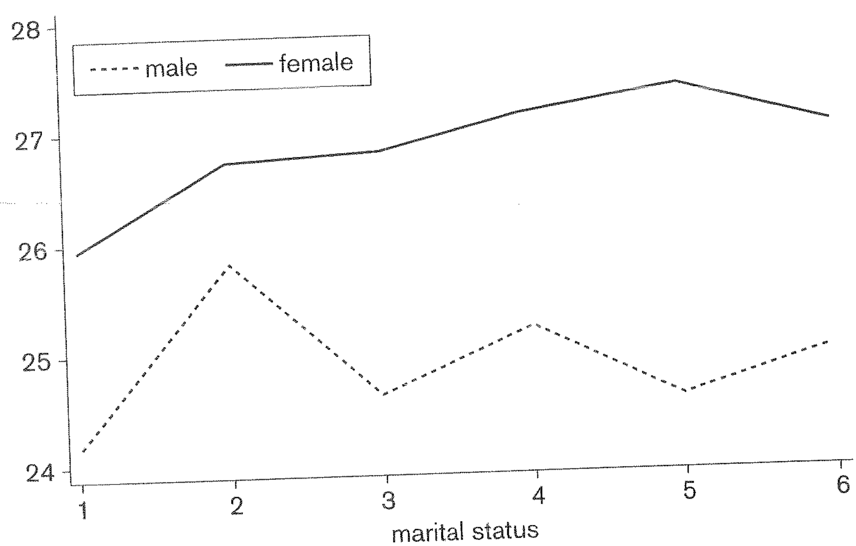
```
  _Ifemale_1 |  .5136454  .1953023   2.63  0.009   .1308066   .8964842
 _Imastat_2 |  1.616538  .2740168   5.90  0.000     1.0794   2.153676
 _Imastat_3 |  .0765492  .3888046   0.20  0.844  -.6856002   .8386987
 _Imastat_4 |  .9034001  .3914352   2.31  0.021   .1360941   1.670706
 _Imastat_5 |  .2874111  .6373313   0.45  0.652  -.9619098   1.536732
 _Imastat_6 |  .5797909  .1911909   3.03  0.002   .2050114   .9545705
_IfemXmas_~2 | -1.063059  .3728356  -2.85  0.004  -1.793905  -.3322125
_IfemXmas_~3 |  .8186293   .428668   1.91  0.056  -.0216619    1.65892
_IfemXmas_~4 |  .0240032  .4828787   0.05  0.960  -.9225538   .9705602
_IfemXmas_~5 |   .893311  .7512081   1.19  0.234  -.5792355   2.365857
_IfemXmas_~6 |  .2944977  .2572233   1.14  0.252   -.209721   .7987164
```

We can see that there is a main effect for the *female* variable where women, on average, report more liberal attitudes to gender roles than men. Then three of the marital status categories have significant main effects in that those living together, those separated, and those who have never been married all have, on average, more liberal attitudes to gender roles. The one significant interaction term between sex and marital status is `_IfemXmas_~2` which is for the 'living together' category. As marital status categories are dummy coded with married as the reference category, this interaction tells us that the 'slope' between 'married' and 'living together' categories is different for men and women. None of the other interaction terms are significant, which tells us that the 'slope' *between those categories and being married* is the same for men and women. These results do not tell us if those who are divorced are different from those who are separated. This is a drawback with using dummy coding, and some prefer to use effect coding to get round this issue of choosing a reference category. See Chapter 8 for an example of effect coding.

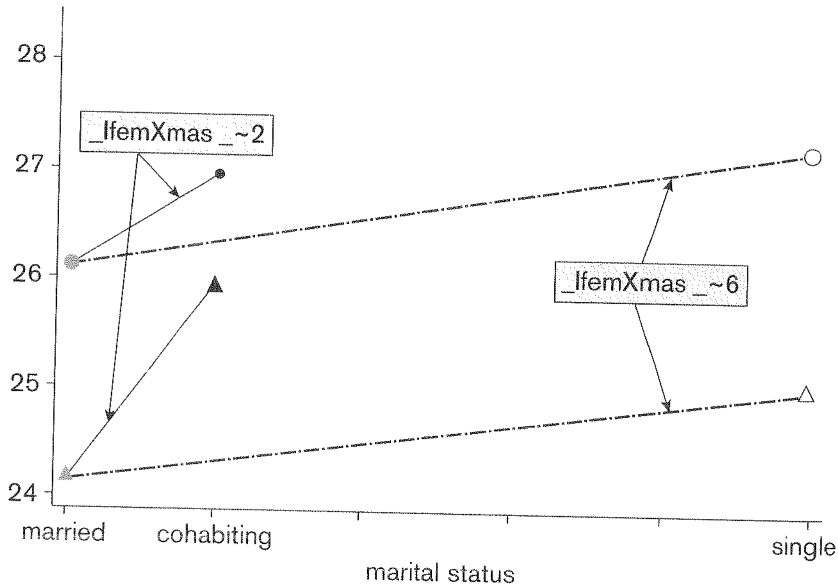
At the risk of being redundant, let's have a look at this using some graphs. We have done these graphs using the **postgr3** command and then using the Graph Editor to show you what is possible in the Editor.

```
postgr3 mastat,by(female)
```

The basic **postgr3** command for the sex and marital status interaction model produces the following line chart. A line chart is not technically correct for this, but it gives enough information. In this graph you can see that the solid line is for women and the dashed line is for men. The average difference between these lines represents the main effect of the female variable, but what the interaction is looking at is the difference in the 'slopes' between each of the categories and the married categories.



Below we have used the Graph Editor and taken out the lines and replaced the category points with markers: circles for women and triangles for men. We have also taken out the information for the widowed, separated, and divorced categories. To help make our point three categories are enough. We have plotted the 'slopes' between the 'married' category and the 'living together' (cohabiting) category with solid lines and between the married category and the never married (single) category with dashed lines. Hopefully, this makes it clear that the `_IfemXmas_~2` interaction term is the difference between the two solid line 'slopes'. In other words, the difference (slope) between those who are married and those who are living together is significantly different for men and women. The difference (slope) is greater for men, which is also shown by the negative sign on the interaction term's coefficient in the regression results as women are represented by *female*=1. Now compare the solid line 'slopes' with the dashed line 'slopes'. The dashed lines are almost parallel which indicates that there is no gender difference in the differences (slopes) *between those being married and those who have never been married*. This is shown in the regression results by the `_IfemXmas_~6` interaction term not being significant. Again, it is worth noting that from these results we cannot say anything about differences in 'slopes' between other pairs of categories such as between divorced and widowed. If you wish you can draw in the other three 'slopes' between married and widowed, married and separated, and married and divorced for both men and women in the first line graph and see how they are reasonably parallel, which is reflected in the non-significant interaction terms `_IfemXmas_~3`, `_IfemXmas_~4` and `_IfemXmas_~5` respectively.



The usefulness of interactions between categorical variables is open to debate. Take this example and ask what this difference of differences (slopes) actually means. The way we have worded the example, where all categories are relative to those who are married, implies what happens when someone changes from that category to another, but is that change logical or the norm? It might make a bit more sense to compare those who are living together and those who are married with those who have never been married, as a common social process is from single to cohabiting to married. Not all who are married moved from the cohabiting or single categories as there will be people who were in the divorced or widowed categories who then married. However, it makes little sense to compare those who have never been married with those who are separated, divorced or widowed as it is not possible to move from being single to being separated, divorced or widowed without first being married.

WRITING UP YOUR FINDINGS

A 'typical' research article in the social and behavioural sciences is organized in the following way:

1. Introduction
2. Literature review and theory
3. Rationale for current study (highlighting any gaps, shortcomings, and/or contradictions in the existing literature) and hypotheses
4. Description of data, variables, and analytic approach
5. Results
6. Discussion
7. Conclusion

The best way to learn how to do these steps is to read lots of articles in your discipline and organize your papers in a similar way. We've discussed here how to create hypotheses, test them, understand your output, and make tables and graphs to display your results. In our opinion, the graphical display of results is something that is truly underrated in the teaching of social statistics – and it is a skill that is much appreciated by novices, policy-makers, and non-technical people who are trying to make sense of quantitative reports and articles. You should always try to make complex statistical output as simple to understand as possible. While you may very much like large tables of numbers (we sympathize completely), they can be daunting and far from user-friendly to your proposed readership.

Discussing your results and tying them back in with the literature review is a skill that you can only develop over time. Your first attempts are likely to sound like the Results section regurgitated, but it is important to link the findings with previous research and theory. It is an art, if you don't mind our saying so. This is likely to be the section that you will have to rewrite several times. It should also include any shortcomings in your analysis. If you don't acknowledge shortcomings, people reviewing your work will be certain to remind you of them. In the example analysis undertaken in this chapter, we would be sure to talk about how the results for education were interesting and unexpected, and why this might be so (i.e. that older people might be in some of the classifications). We would also highlight the shortcomings of how we measured ethnicity and how the results might be masking important differences between people in the groups. The discussion section is also a good place to talk about recommendations for future research.