Stata Pointers

The following example illustrates the analysis of data entered as frequencies, rather than one line per case. The [weight=freq] part of each command tells STATA to regard each line as representing the number of individuals given by the variable I have arbitrarily called freq. The same analysis could be achieved on a data set of one line per case by omitting [weight=freq]. The required data set would omit the freq variable and have 31 lines identical to the first line in the listing below, 11 lines identical to the second , etc.

```
. insheet using transf.csv , comma
(4 vars, 12 obs)
. list
          hosp       type      freq       trans
  1.    Hosp-A       Hips        31  Transf-Yes
  2.    Hosp-B       Hips        11  Transf-Yes
  3.    Hosp-C       Hips        91  Transf-Yes
  4.    Hosp-A      Knees        16  Transf-Yes
  5.    Hosp-B      Knees         7  Transf-Yes
  6.    Hosp-C      Knees         9  Transf-Yes
  7.    Hosp-A       Hips        88   Transf-No
  8.    Hosp-B       Hips        34   Transf-No
  9.    Hosp-C       Hips       251   Transf-No
 10.    Hosp-A      Knees       240   Transf-No
 11.    Hosp-B      Knees        80   Transf-No
 12.    Hosp-C      Knees        99   Transf-No

. describe

Contains data
  obs:            12
 vars:             4
 size:           480 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
hosp            str6   %9s
type            str5   %9s
freq            int    %8.0g
trans           str10  %10s
-------------------------------------------------------------------------------
```

The output above shows that except for freq variables have been entered as string variables (see storage type column). I The following creates numerical variables with labels. These tend to work better for many types of analysis. I also drop Hospital A to reproduce the analysis from class.

```
. encode hosp, gen(hospit)
. encode type, gen(proc)
. encode trans , gen(transf)
. drop if hospit == 1
(4 observations deleted)
```

Now I'll try some simple tabulations.

```
. tab hospit

     hospit |      Freq.     Percent        Cum.
------------+-----------------------------------
     Hosp-B |          4       50.00       50.00
     Hosp-C |          4       50.00      100.00
------------+-----------------------------------
      Total |          8      100.00
```

To get useful results, I invoke the weight option.

```
. tab hospit [weight = freq]
(frequency weights assumed)

     hospit |      Freq.     Percent        Cum.
------------+-----------------------------------
     Hosp-B |        132       22.68       22.68
     Hosp-C |        450       77.32      100.00
------------+-----------------------------------
      Total |        582      100.00
```

Next I'll get all the two-way crosstabulations using the tab2 command.

```
. tab2 hospit proc transf [weight=freq] , chi2
(frequency weights assumed)

-> tabulation of hospit by proc

           |         proc
    hospit |      Hips      Knees |     Total
-----------+----------------------+----------
    Hosp-B |        45         87 |       132
    Hosp-C |       342        108 |       450
-----------+----------------------+----------
     Total |       387        195 |       582

          Pearson chi2(1) =  80.4601   Pr = 0.000

-> tabulation of hospit by transf

           |        transf
    hospit | Transf-No  Transf-Ye |     Total
-----------+----------------------+----------
    Hosp-B |       114         18 |       132
    Hosp-C |       350        100 |       450
-----------+----------------------+----------
     Total |       464        118 |       582

          Pearson chi2(1) =   4.6545   Pr = 0.031

-> tabulation of proc by transf

           |        transf
      proc | Transf-No  Transf-Ye |     Total
-----------+----------------------+----------
      Hips |       285        102 |       387
     Knees |       179         16 |       195
-----------+----------------------+----------
     Total |       464        118 |       582

          Pearson chi2(1) =  26.4297   Pr = 0.000
```

Now my first attempt at getting the MH estimate and test.

```
. mhodds hospit transf proc [weight = freq]
(frequency weights assumed)
Response hospit not coded 0/1
```

Rats - most of the epitab commands are very fussy and require that the varables are 0/1. This can be done as below by subtraction. The codes for Hospital B and C are 2 and 3 (1's corresponded to the dropped Hospital A data). Now everything works.

```
. gen hospdum = hospit - 2
. gen trandum = transf - 1
. gen procdum = proc - 1
. mhodds hospdum trandum procdum [ weight = freq]
(frequency weights assumed)

Mantel-Haenszel estimate of the odds ratio
Comparing trandum==1 vs. trandum==0, controlling for procdum

------------------------------------------------------------------
 Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
------------------------------------------------------------------
   1.093461       0.09        0.7672         0.605073   1.976053
------------------------------------------------------------------
```

Missing Data Values

To fix up data where some values are meant to designate missing use the mvencode command - see the on-line help!