```
#### Biost 517: Applied Biostatistics I
#### Emerson, Fall 2007

#### Homework #1 Key
#### Annotated Stata Log File
#### October 13, 2007

#### NOTE: I most definitely did not want you to hand in such
#### output as this. I do this to aid you in understanding
#### how I got the answers for the Key.

#### Comments edited into the log file produced by Stata are
#### on the lines that start with the four '#' signs and are
#### printed in italics.

#### The Stata commands are put in bold face.

#### Stata output is displayed in regular typeface in blue.

#### Reading in the data from the textfile
. infile case ptid sex age alb alkphos alt ast bili chol using pbcscreen.txt

'case' cannot be read as a number for case[1]
'ptid' cannot be read as a number for ptid[1]
'sex' cannot be read as a number for sex[1]
'age' cannot be read as a number for age[1]
'alb' cannot be read as a number for alb[1]
'alkphos' cannot be read as a number for alkphos[1]
'alt' cannot be read as a number for alt[1]
'ast' cannot be read as a number for ast[1]
'bili' cannot be read as a number for bili[1]
'chol' cannot be read as a number for chol[1]
'NA' cannot be read as a number for chol[9]
'NA' cannot be read as a number for chol[14]
'NA' cannot be read as a number for chol[16]
'NA' cannot be read as a number for alt[19]
'NA' cannot be read as a number for chol[19]
'NA' cannot be read as a number for chol[20]
'NA' cannot be read as a number for alt[21]
'NA' cannot be read as a number for chol[21]
'NA' cannot be read as a number for chol[26]
'NA' cannot be read as a number for chol[31]
```

```
'NA' cannot be read as a number for alb[53]
'NA' cannot be read as a number for bili[53]
'NA' cannot be read as a number for chol[53]
'NA' cannot be read as a number for chol[68]
'NA' cannot be read as a number for chol[69]
'NA' cannot be read as a number for chol[71]
'NA' cannot be read as a number for chol[73]
'NA' cannot be read as a number for chol[74]
'NA' cannot be read as a number for alb[77]
'NA' cannot be read as a number for chol[77]
'NA' cannot be read as a number for chol[78]
'NA' cannot be read as a number for chol[80]
'NA' cannot be read as a number for chol[81]
'NA' cannot be read as a number for chol[82]
'NA' cannot be read as a number for chol[83]
'NA' cannot be read as a number for chol[84]
'NA' cannot be read as a number for alb[87]
'NA' cannot be read as a number for bili[87]
'NA' cannot be read as a number for chol[87]
'NA' cannot be read as a number for alkphos[92]
'NA' cannot be read as a number for alt[92]
'NA' cannot be read as a number for ast[92]
'NA' cannot be read as a number for chol[93]
'NA' cannot be read as a number for chol[97]
'NA' cannot be read as a number for alb[98]
'NA' cannot be read as a number for bili[98]
'NA' cannot be read as a number for chol[98]
'NA' cannot be read as a number for chol[99]
'NA' cannot be read as a number for chol[100]
'NA' cannot be read as a number for chol[103]
'NA' cannot be read as a number for chol[107]
'NA' cannot be read as a number for alb[110]
'NA' cannot be read as a number for chol[110]
'NA' cannot be read as a number for chol[113]
'NA' cannot be read as a number for chol[115]
'NA' cannot be read as a number for chol[118]
'NA' cannot be read as a number for chol[119]
'NA' cannot be read as a number for chol[120]
'NA' cannot be read as a number for chol[121]
'NA' cannot be read as a number for chol[123]
'NA' cannot be read as a number for chol[125]
```

```
'NA' cannot be read as a number for chol[131]
'NA' cannot be read as a number for alb[132]
'NA' cannot be read as a number for chol[132]
'NA' cannot be read as a number for alb[133]
'NA' cannot be read as a number for chol[133]
'NA' cannot be read as a number for chol[134]
'NA' cannot be read as a number for chol[135]
'NA' cannot be read as a number for alb[137]
'NA' cannot be read as a number for chol[137]
'NA' cannot be read as a number for chol[138]
'NA' cannot be read as a number for chol[141]
'NA' cannot be read as a number for chol[142]
'NA' cannot be read as a number for chol[145]
'NA' cannot be read as a number for chol[146]
'NA' cannot be read as a number for chol[148]
'NA' cannot be read as a number for chol[150]
'NA' cannot be read as a number for chol[151]
'NA' cannot be read as a number for chol[152]
'NA' cannot be read as a number for chol[153]
'NA' cannot be read as a number for chol[154]
'NA' cannot be read as a number for chol[155]
'NA' cannot be read as a number for chol[160]
'NA' cannot be read as a number for chol[171]
'NA' cannot be read as a number for alb[172]
'NA' cannot be read as a number for bili[172]
'NA' cannot be read as a number for chol[172]
'NA' cannot be read as a number for chol[174]
'NA' cannot be read as a number for chol[213]
'NA' cannot be read as a number for alb[215]
'NA' cannot be read as a number for bili[215]
'NA' cannot be read as a number for chol[215]
'NA' cannot be read as a number for chol[221]
'NA' cannot be read as a number for alb[233]
'NA' cannot be read as a number for bili[233]
'NA' cannot be read as a number for chol[233]
'NA' cannot be read as a number for alb[234]
'NA' cannot be read as a number for bili[234]
'NA' cannot be read as a number for chol[234]
'NA' cannot be read as a number for bili[270]
'NA' cannot be read as a number for chol[270]
'NA' cannot be read as a number for chol[271]
```

```
'NA' cannot be read as a number for alb[275]
'NA' cannot be read as a number for bili[275]
'NA' cannot be read as a number for chol[275]
'NA' cannot be read as a number for chol[276]
'NA' cannot be read as a number for chol[279]
'NA' cannot be read as a number for chol[281]
'NA' cannot be read as a number for chol[282]
'NA' cannot be read as a number for chol[285]
'NA' cannot be read as a number for chol[288]
'NA' cannot be read as a number for chol[300]
'NA' cannot be read as a number for alb[311]
'NA' cannot be read as a number for bili[311]
'NA' cannot be read as a number for chol[311]
'NA' cannot be read as a number for alb[312]
'NA' cannot be read as a number for bili[312]
'NA' cannot be read as a number for chol[312]
'NA' cannot be read as a number for alb[320]
'NA' cannot be read as a number for bili[320]
'NA' cannot be read as a number for chol[320]
'NA' cannot be read as a number for alb[341]
'NA' cannot be read as a number for chol[341]
'NA' cannot be read as a number for alb[348]
'NA' cannot be read as a number for bili[348]
'NA' cannot be read as a number for chol[348]
'NA' cannot be read as a number for alb[349]
'NA' cannot be read as a number for bili[349]
'NA' cannot be read as a number for chol[349]
'NA' cannot be read as a number for alb[350]
'NA' cannot be read as a number for bili[350]
'NA' cannot be read as a number for chol[350]
'NA' cannot be read as a number for alb[351]
'NA' cannot be read as a number for bili[351]
'NA' cannot be read as a number for chol[351]
'NA' cannot be read as a number for alb[356]
'NA' cannot be read as a number for bili[356]
'NA' cannot be read as a number for chol[356]
'NA' cannot be read as a number for alb[357]
'NA' cannot be read as a number for bili[357]
'NA' cannot be read as a number for chol[357]
'NA' cannot be read as a number for chol[368]
'NA' cannot be read as a number for chol[386]
```

```
'NA' cannot be read as a number for alkphos[389]
'NA' cannot be read as a number for alt[389]
'NA' cannot be read as a number for ast[389]
'NA' cannot be read as a number for chol[389]
'NA' cannot be read as a number for chol[391]
'NA' cannot be read as a number for alb[395]
'NA' cannot be read as a number for bili[395]
'NA' cannot be read as a number for chol[395]
'NA' cannot be read as a number for alb[396]
'NA' cannot be read as a number for bili[396]
'NA' cannot be read as a number for chol[396]
'NA' cannot be read as a number for chol[416]
'NA' cannot be read as a number for chol[418]
'NA' cannot be read as a number for chol[420]
'NA' cannot be read as a number for chol[422]
'NA' cannot be read as a number for chol[424]
'NA' cannot be read as a number for chol[425]
'NA' cannot be read as a number for chol[426]
'NA' cannot be read as a number for chol[427]
'NA' cannot be read as a number for chol[428]
'NA' cannot be read as a number for chol[429]
'NA' cannot be read as a number for alb[430]
'NA' cannot be read as a number for chol[430]
'NA' cannot be read as a number for alb[439]
'NA' cannot be read as a number for bili[439]
'NA' cannot be read as a number for chol[439]
'NA' cannot be read as a number for alb[443]
'NA' cannot be read as a number for bili[443]
'NA' cannot be read as a number for chol[443]
'NA' cannot be read as a number for alb[449]
'NA' cannot be read as a number for bili[449]
'NA' cannot be read as a number for chol[449]
'NA' cannot be read as a number for chol[467]
'NA' cannot be read as a number for chol[469]
'NA' cannot be read as a number for alb[502]
'NA' cannot be read as a number for bili[502]
'NA' cannot be read as a number for chol[502]
'NA' cannot be read as a number for chol[508]
'NA' cannot be read as a number for chol[516]
'NA' cannot be read as a number for alkphos[519]
'NA' cannot be read as a number for alkphos[528]
```

```
'NA' cannot be read as a number for chol[530]
'NA' cannot be read as a number for chol[535]
'NA' cannot be read as a number for chol[536]
(536 observations read)

 #### Drop the first case, because it was just the column headings
. drop in 1
(1 observation deleted)

#### Declare the format to provide approximately 3 significant digits in print out
. format age bili %9.1f
. format alb %9.2f
. format alkphos alt ast chol %9.0f

#### Save the data file so I don't have to do all of the above again
. save pbcscreen
file adultfev.dta saved

#### Checking to see if all subject ID numbers are unique.
#### I do this using the "egen" command with the by()
#### option which will generate a new variable containing a
#### constant equal to the count of nonmissing data. Then
#### when I do a table of that new constant, I find that
#### there are 535 cases with the value 1. Had there been
#### a duplicate subject ID number, I might have found, say,
#### 533 cases with a value of 1 and 2 cases with a value of 2.
. egen idcnt= count(ptid), by(ptid)
. table idcnt

----------------------
    idcnt |      Freq.
----------+-----------
        1 |        535
----------------------

#### Descriptive statistics for the entire sample in the format I like.
#### I omit case and ptid, because those are just labels. I omit sex
#### because it is a binary variable: All I would care about is the
#### frequency of each sex.
#### Note the fact that I specified the statistics that I wanted, I
#### specified that the statistics were to be in columns, and I specified
#### that I wanted Stata to use the formats that I had pre-specified for
#### the variables.
```

```
. tabstat age alb alkphos alt ast bili chol, stat(n mean sd min p25 p50 p75 max) col(stat) format

    variable |         N      mean        sd       min       p25       p50       p75       max
-------------+----------------------------------------------------------------------------------
         age |     535.0      51.9       9.6       1.0      46.0      51.0      59.0      80.0
         alb |    505.00      3.97      0.44      1.80      3.80      4.00      4.30      5.20
     alkphos |       531       372       321        60       174       282       458      3741
         alt |       531       111       353         8        42        64       111      5550
         ast |       533        95       245        12        46        69       102      5550
        bili |     511.0       1.1       2.1       0.1       0.5       0.7       1.1      35.2
        chol |       422       250        72        79       205       236       284       716
-------------------------------------------------------------------------------------------------

. tabulate sex, row column
        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         34        6.36        6.36
          2 |        501       93.64      100.00
------------+-----------------------------------
      Total |        535      100.00
```

```
#### Now doing the same within groups defined by sex. Note
#### that I either had to sort the data first or use
#### the command "bysort" instead of "by".
. by sex: tabstat age alb alkphos alt ast bili chol, stat(n mean sd min p25 p50 p75 max) col(stat) format
-----------------------------------------------------------------------------------------------------------
-> sex = 1

    variable |         N      mean        sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------------------
         age |      34.0      53.0       8.1      37.0      46.0      52.0      60.0      66.0
         alb |     31.00      4.01      0.37      3.00      3.80      4.10      4.30      4.60
     alkphos |        34       334       179       107       203       303       439       744
         alt |        34       117       106        18        50        87       147       550
         ast |        34       100        98        23        50        70       100       550
        bili |      31.0       1.4       2.0       0.2       0.6       0.8       1.1      10.2
        chol |        27       263        95       138       194       237       301       513
-------------+--------------------------------------------------------------------------------

-----------------------------------------------------------------------------------------------------------
-> sex = 2

    variable |         N      mean        sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------------------
         age |     501.0      51.8       9.7       1.0      46.0      51.0      59.0      80.0
         alb |    474.00      3.96      0.45      1.80      3.80      4.00      4.30      5.20
     alkphos |       497       375       329        60       174       281       458      3741
         alt |       497       110       364         8        42        62       108      5550
         ast |       499        95       252        12        45        69       103      5550
        bili |     480.0       1.1       2.1       0.1       0.5       0.7       1.1      35.2
        chol |       395       249        70        79       205       235       283       716
-------------+--------------------------------------------------------------------------------
```