Homework1 Questions – (Answers at the end)

Please answer the following questions using Stata and submit your output. You may find the **log** command useful for saving your output.
The charts can be either copied and pasted into MS Word (right click on the chart for the options), or saved on disk (.png format is recommended) and inserted into MS Word manually, through Insert->Picture menu. The latter method is sometimes more reliable, while the former one is faster.

1. Compare the distributions of the book/market ratio, market capitalization, dividend/price ratio, price/earnings between 1998 and 1990:
    a. Compute summary statistics for these variables across all companies. Compare with the results for 1990.
    b. Create 10 level groups, based on market capitalization. Compute average book/market ratio, dividend/price ratio, price/earnings ratio for each group. Create bar charts (see tutorial for an example).
    c. Comment on differences between 1990 and 1998.
2. Consider a strategy of buying stock of companies with high book/market ratio and selling stock of those with low such ratio. Compute the profits such a strategy would have produced during 1999-2003.
3. List all of the most closely related companies in 1998 (`hsicig=739`). Display the company name, ticker, main line of business, book value, market capitalization and past year return.
4. Repeat the analysis in p.1 for the entire business services industry alone in 1990 and 1998 (`hsicmg=73`). What can you tell about these companies? How has the industry changed since 1990, compared to the rest of the economy? Hint: to avoid tedious repeating of the **if** conditions, you may check the Stata help on **keep** and **drop** commands and utilize one of them.
5. Make a scatter plot of the earnings/price ratio and market value of the companies with `hsicig==739`. Put ticker marks on the chart. Do any companies stand out? What can you tell about HGC?

Stata tutorial

Contact Information

With all questions concerning Stata, this tutorial and exercise questions, or other related issues, please contact Andrei Jirnyi, a-jirnyi@northwestern.edu, phone (847) 491-4270(o), (847)858-2537(c).

Why Stata?

There are many ways to do numerical analysis. Most of you are probably familiar with Microsoft Excel. Why learn another tool, and what advantages does it have?

1. Fast and easy data transformation
2. Fast and easy data analysis
3. Ability to handle large datasets
4. Easy automation of repeated operations
5. A large number of built-in statistical procedures

How to get Stata

Kellogg has established a Stata "GradPlan" -- this is an agreement with StataCorp that allows Kellogg faculty, students and staff to buy Stata at the lowest possible cost. The details can be found at http://www.kellogg.northwestern.edu/rc/sec/manuals/stata/stata-gradplan.htm. Intercooled Stata is the recommended version, its' price is $145.

The software is also available on the PCs located in the computing labs on the Evanston and Chicago campuses. Intercooled Stata version 8.2 is installed in the "special software" workstations at the Jacobs Center (ten workstations) and Chicago campus (six workstations) computer laboratories.

Some preliminary steps

Conventions in this document

In this handout, I'll put a dot in front of Stata commands in this text, and display them in **`bold courier,`** like this:

```
.help
```

Note that you do not need to type the dot!
The parameters of the commands are sometimes given in text in ***`italic,`*** required parameters in ***`<angle brackets>,`*** and optional ones in **`[`*`square brackets`*`]`**. The actual Stata output is given in `plain courier.`

Stata windows

Once you start Stata for Windows, you'll see four main windows:

1. **Stata Results**: (on Windows typically with black background): this is where you can see the estimation results and all other output that you ask Stata to provide.
2. **Stata Command**: this is where you enter all commands
3. **Review**: this window keeps a complete history of the commands you've entered
4. **Variables**: once the dataset is loaded, all the available variables are listed here.

The commands you enter are repeated in the results window in white color, while Stata output is typically in green and yellow.

The working directory

Whenever you ask Stata to read or write something on disk, it will try to do it in your *working directory* (unless otherwise specified). To see the current working directory, you may type the following command:

```
.pwd
```

You probably want to change this directory to the one you keep the data in, to do so use a **`cd`** command:

```
.cd "c:\aj\work\value-investing"
```

both forward (/) and reverse (\) slashes work, and the quotation marks are required only when you have spaces in the path name.

To see the contents of the directory, use **`ls`** or **`dir`**:

```
.ls
```

```
  <dir>   3/14/06 17:27  .
  <dir>   3/14/06 17:27  ..
4040.7k   3/13/06 12:54  piotroski00.pdf
   5.8k   3/10/06 12:59  spore_stata.csv
  28.0k   3/14/06 17:27  statatut.doc
```

To get help on a command, type **`.man <command>`**

```
.man ls
```

Some of you might have recognized these commands from DOS or UNIX. In fact, many other DOS and UNIX commands work in Stata, such as **`copy, erase, mkdir, cat, type, shell.`**

Keeping track of your work

One of the advantages Stata and other similar software has over the menu-driven systems, such as spreadsheets, is the ability to keep track of all your work and of all the results that you've obtained, and to quickly and easily repeat it all on a different dataset or with some other modifications.
There are two main commands that I recommend that you do at the very beginning of your session: **`log`** and **`cmdlog`**:

```
.log using session.log
.cmdlog using commands.txt
```

The first will create a file named `session.log` in your home directory[1]. This file is a simple text file, that could be opened in any text editor, such as Microsoft Word or Notepad.
The second command will create a file `commands.txt`, containing all the commands that you have entered. This file can later be easily edited into a Stata script for automatic execution.
Whenever you do something that you would like to make a particular note of in the script or the output file, you may just enter a comment in the command line by typing an asterisk (*) first:

```
.* This is a comment
```

This can greatly simplify future parsing of the command file in particular.

Once your work is done, you should close the log files by typing

```
.log close
.cmdlog close
```

Loading the data into Stata

The first step in any analysis is to read the dataset. There are many data formats, such as Excel spreadsheets, text files, or SAS datasets.
This time we'll start with an existing Stata dataset, `nyse90.dta`. It contains data on accounting information and average future 5-year returns on publicly traded stocks with a closing price of $5 or above on December 31, 1990. Please download the file from the class website and save it into your working directory. After that, start Stata and load the dataset by typing the following command:

```
.*Loading the data...
.clear all
.set mem 100m
.use stocks90
```

[1] It is essential to explicitly specify the extension `.log` for the output log – otherwise Stata will create a file `session.smcl` in its' own format, that can only be viewed from within Stata by entering a command **`view session.smcl`**.

The first command clears any possibly open datasets from memory, so that you can load a new one. This step is particularly recommended to include in any script files that you may write.

Once the data is loaded, inspect it by chosing **Data->Data Editor** in the main Stata menu. You'll see a spreadsheet, much like an Excel one, of all the data in memory. Each row represents the data for one company in 1990. From now on I'll refer to the rows of data as "observations". It is a fairly large data set including many variables that we will not use in this tutorial.

If you scroll around the spreadsheet a bit, you might notice that some apparently numerical variables have a value of "." in the cell. This means that this particular variable is missing for this company. For example, we do not have data on 1990 earnings for M.L. Macadamia Orchards, LP.

Summarizing the data

First, we would like to list the variables:

```
. d

Contains data from stocks90.dta
  obs:         1,811
 vars:           352                          29 Mar 2006 07:45
 size:     2,571,620 (97.5% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
name            str32  %32s                   company name
ticker          str5   %5s                    exchange ticker symbol
permno          long   %12.0g                 crsp permanent number
gvkey           long   %12.0g                 standard and poor's identifier
hsiccd          int    %8.0g                  header sic code
hsicig          int    %8.0g                  header sic industry group
ret90           float  %9.0g                  annual return on common stock
                                                in 1990
retstd          float  %9.0g                  holding period return
ret5            float  %9.0g                  average annual return on common
                                                stock for 1991-1995
vol             double %12.0g                 share volume
shrout          double %12.0g                 shares outstanding
prc             float  %9.0g                  closing price or bid/ask average
linkdt          int    %dD_m_Y                first effective date of link
linkenddt       int    %dD_m_Y                last effective date of link
fyr             byte   %8.0g                  fiscal yearend month of data
data1           float  %9.0g                  cash and short-term investments
                                                (mm$)
data2           double %10.0g                 receivables - total (mm$)
data3           float  %9.0g                  inventories - total (mm$)
assets_cur      float  %9.0g                  current assets - total (mm$)

.............

be              float  %9.0g                  book equity
me              float  %9.0g                  market equity
mv              float  %9.0g
z50             float  %9.0g
ebi             float  %9.0g
z19             float  %9.0g
ecomm           float  %9.0g
bme             float  %9.0g                  book/market equity ratio
hsicmg          float  %9.0g                  header sic major group
-------------------------------------------------------------------------------
Sorted by:
```

At the beginning we can see some summary information on this data file. We know that it is named `stocks90.dta`, that it has 1811 observations and 352 variables (or, in other words, that it contains 352 data items, describing 1811 different companies), and that it occupies about 2.5M of memory.

Below we can see, among other things, the *variable names* – short names for different data items, that we'll use to refer to them in Stata commands, and *variable labels* – short descriptions of each item. We'll soon learn how to create both.

We do not need to go to the spreadsheet data browser to get a quick look at the data. Just use the **`.list`** command (usually abbreviated as **.l**):

```
. l name  ticker assets_tot sales_net in 1/5
```

```
     +-------------------------------------------------------------+
     |                        name   ticker   assets~t   sales_~t |
     |-------------------------------------------------------------|
  1. |  CHIQUITA BRANDS INTL INC      CQB   2174.437    4272.66 |
  2. |      DEKALB GENETICS CORP    SEEDB      238.1        274 |
  3. |             DOLE FOOD INC      DOL    2423.27   3003.213 |
  4. | M L MACADAMIA ORCHARDS LP      NUT     74.567     14.186 |
  5. |  PIONEER HI BRED INTL INC      PHB   1005.992    964.453 |
     +-------------------------------------------------------------+
```

(your output may look a little differently, depending on the size of your results window). This command lists several columns

A few things to note here:

- **`"in 1/3"`** at the end instructed Stata to only use the observations 1 through 3
- Some variable names have been shortened to fit into the table (`sales_~t` instead of `sales_net`). When you are entering Stata commands, it is sufficient to enter only as many characters as to uniquely specify the variable:

`. l name ticker sales in 1/10`

```
    +-----------------------------------------------+
     |                       name   ticker   sales_~t |
     |-----------------------------------------------|
  1. |  CHIQUITA BRANDS INTL INC      CQB    4272.66 |
  2. |      DEKALB GENETICS CORP    SEEDB        274 |
  3. |             DOLE FOOD INC      DOL   3003.213 |
  4. | M L MACADAMIA ORCHARDS LP      NUT     14.186 |
  5. |  PIONEER HI BRED INTL INC      PHB    964.453 |
     |-----------------------------------------------|
  6. |          SERVICEMASTER CO      SVM    1825.75 |
  7. |       I P TIMBERLANDS LTD      IPT    195.461 |
  8. |  RAYONIER TIMBERLANDS L P      LOG    107.562 |
  9. |                 B H P LTD      BHP      11853 |
 10. |      CLEVELAND CLIFFS INC      CLF      309.9 |
     +-----------------------------------------------+
```

Let us also see some statistics:

```
. su

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        name |         0
      ticker |         0
      permno |      1811     48897.9     22724.1      10064      93105
       gvkey |      1811    10164.17    11280.82       1004     160785
      hsiccd |      1811    4751.088    1892.273          0       9511
-------------+--------------------------------------------------------
      hsicig |      1811    474.7929      189.21          0        951
       ret90 |      1811   -.0621237    .5000388  -.8105354    14.3767
      retstd |      1804    .0999556    .0467829   .0071527   .4288884
        ret5 |      1809    .1655348     .222925  -.7883611   4.006406
         vol |      1811    2.156981    3.939127      .0009    47.0501
-------------+--------------------------------------------------------
      shrout |      1811    48.89816    94.83598       .073   1246.185
         prc |      1811     26.7208     157.644      5.125       6675

.............

          be |      1584    1372.901    3349.589   -2587.88      46693
          me |      1753    1794.967    4814.561   8.870875   64567.18
-------------+--------------------------------------------------------
          mv |      1566    6104.499    16204.48     12.584   211465.5
         z50 |      1804   -.7642056    75.11449    -2424.8        671
         ebi |      1361     271.226     771.631  -1390.251      11600
         z19 |      1810    3.975649    19.92937      -.526        559
       ecomm |      1576    137.8852    439.4368    -4448.7       6433
-------------+--------------------------------------------------------
         bme |      1566    .9123388    .6511382  -5.262994   6.456691
      hsicmg |      1811    47.13418    18.93703          0         95
```

`.su` is a shortcut for **`.summarize`** here. The first column of the output gives the number of *non-missing numerical observations* (note that `ticker` and `name` have 0 in it since they are not numerical), followed by sample statistics.

We can do some very quick analysis already now. For example, one may be interested in what was the average relationship between past 1-year return and future 5-year returns in 1990. For a brief look, let us display a correlation table, using **`.cor:`**

```
. cor ret90 ret5
(obs=1809)

             |    ret90     ret5
-------------+------------------
       ret90 |   1.0000
        ret5 |  -0.0548   1.0000
```

We can see that the correlation is negative (-5%) here – that is, companies, whose stock did poorly in 1990, on average did better for the following 5 years.

Generating new variables

While this file contains a few useful data items, we would like to look at other numbers as well – in particular, at some useful financial ratios. Just in Excel we would create new columns, in Stata to do that we need to create new variables. This is done with a command **.gen**erate.

```
. gen eps=earn/shrout
(269 missing values generated)
. gen epr=eps/prc
(269 missing values generated)
. gen dp=div/prc
(43 missing values generated)
```

We have just created five new variables: eps[2], epr, and dp. Note that in all cases some missing values were created. This happened because we do not have data on all items for all companies. For example, because we do not have total earnings for NUT, we were not able to compute earnings per share either.

```
. l ticker earn shrout eps in 1/5
```

	ticker	earn	shrout	eps
1.	CQB	348.996	45.219	7.717906
2.	SEEDB	20.7	5.581	3.709013
3.	DOL	448.802	59.27	7.572161
4.	NUT	.	7.5	.
5.	PHB	646.341	30.924	20.90095

Now our dataset contains 355 variables:

```
. d
Contains data from stocks90.dta
  obs:         1,811
 vars:           355                          29 Mar 2006 07:45
 size:     2,593,352 (97.5% of memory free)
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
name            str32  %32s                   company name
ticker          str5   %5s                    exchange ticker symbol
permno          long   %12.0g                 crsp permanent number
.............
eps             float  %9.0g
epr             float  %9.0g
dp              float  %9.0g
------------------------------------------------------------------------------
Sorted by:
     Note:  dataset has changed since last saved
```

[2] One thing that we need to watch out for here is measurement units: in our case both earnings are recorded in millions of dollars, and the number of shares outstanding in millions of shares, so it is OK to divide them. If, say, number of shares were recorded in thousands, we'd have to multiply it by 1000.

The five new variables are at the end. Note that unlike other variables, they do not yet have labels. It's a good idea to add them right now, so that we won't forget what they are. This is done with **`.label var`** command:

```
. label var eps "Earnings per share"
. label var epr "Earnings to price ratio"
. label var dp "Dividend to price ratio"

.d

Contains data from stocks90.dta
  obs:         1,811
 vars:           355                          29 Mar 2006 07:45
 size:     2,593,352 (97.5% of memory free)
------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
name            str32  %32s                   company name
ticker          str5   %5s                    exchange ticker symbol
permno          long   %12.0g                 crsp permanent number

.......................

eps             float  %9.0g                  Earnings per share
epr             float  %9.0g                  Earnings to price ratio
dp              float  %9.0g                  Dividend to price ratio
------------------------------------------------------------------------------
Sorted by:
     Note:  dataset has changed since last saved
```

The note in the last line reminds us that we have modified the dataset. It might be a good idea to **`.save`** it now, so that the work we've just done wouldn't be lost in case something unexpected happens:

```
.save, replace
```

The ***`[,replace]`*** option specifies that if a file with the current name already exists, it would be overwritten. To save it under a different name (e.g. `nyse90_new`) instead, we'd have to specify it like this:

```
.save stocks90_new, replace
```

A few more statistics

Now, once we created our new variables, let us take a look at them. Here are some useful variables, including those we just computed:

```
. d bme me dp epr eps

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
bme             float  %9.0g                  book/market equity ratio
me              float  %9.0g                  market equity
dp              float  %9.0g                  Dividend to price ratio
epr             float  %9.0g                  Earnings to price ratio
eps             float  %9.0g                  Earnings per share
```

First, compute some summary statistics:

```
. su bme me dp epr eps

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         bme |      1566    .9123388    .6511382  -5.262994   6.456691
          me |      1753    1794.967    4814.561   8.870875   64567.18
          dp |      1768     .048625    .1420018          0   4.351515
         epr |      1542    2.000054    24.07292  -27.17606   765.8135
         eps |      1542    55.94497    632.4668  -268.3636   19911.15
```

Now let us check how they are related to each other and to returns. First, compute a correlation matrix:

```
.  cor ret5 ret90 epr me bme dp
(obs=1525)

             |     ret5    ret90      epr       me      bme       dp
-------------+------------------------------------------------------
        ret5 |   1.0000
       ret90 |  -0.0769   1.0000
         epr |  -0.0113  -0.0137   1.0000
          me |  -0.0397   0.0919   0.0624   1.0000
         bme |   0.1374  -0.3386   0.0385  -0.1394   1.0000
          dp |   0.0026  -0.0376   0.0035  -0.0117   0.0630   1.0000
```

First, one thing you probably noticed that the correlation between `ret5` and `ret90` has changed – it is now -7.7%. This happened because all companies that had missing values in *any* of the variables were excluded from computations here.

Other than that, we can see that there is a small negative correlation between future returns and E/P ratio and the market value. On the other hand, companies with a higher D/P and especially B/M ratios did better than others.

One problem with these simple statistics is that they do not tell us much about how significant the relationships are, how large are the deviations from the average, etc.

We can see that there are a few large outliers that confuse the picture a little. To make it a little more transparent, we can break the companies into several groups, based on their B/M ratio, market capitalization and other variables:

Regression analysis

To evaluate the significance of the potential relationships, let us run a linear regression of future returns on potential predictor variables:

```
. regress ret5 ret90 epr me bme dp

      Source |       SS       df       MS              Number of obs =    1525
-------------+------------------------------           F(  5,  1519) =    6.38
       Model |  1.69165501     5  .338331003           Prob > F      =  0.0000
    Residual |   80.568012  1519  .053040166           R-squared     =  0.0206
-------------+------------------------------           Adj R-squared =  0.0173
       Total |   82.259667  1524  .053976159           Root MSE      =   .2303

------------------------------------------------------------------------------
        ret5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ret90 |  -.0200095   .0161318    -1.24   0.215    -.0516525    .0116336
         epr |  -.0001481   .0002445    -0.61   0.545    -.0006276    .0003314
          me |  -8.32e-07   1.17e-06    -0.71   0.476    -3.12e-06    1.46e-06
         bme |   .0444557    .009725     4.57   0.000     .0253798    .0635315
          dp |  -.0102525   .0393889    -0.26   0.795    -.0875149      .06701
       _cons |   .1304556   .0109711    11.89   0.000     .1089356    .1519757
------------------------------------------------------------------------------
```

The strongest and most significant predictor of future returns was clearly the B/M ratio. But, while other variables are not very significant in such a regression, this does not necessarily mean that a strategy, based on them, would not be profitable. One thing that a regression like this one fails to account for is the potential non-linear relationships between returns and predicting variables.

Portfolio construction

To further explore the potential relationships, let us construct simulation portfolios, based on our potential variables. To do so, we shall rank the stocks in accordance with the value of each such variable and compute the returns to a strategy of buying potential winners and selling short potential losers. Let us start with the book/market ratio, as the most potentially significant predictor.

First, let us create a *categorical variable* that forms 10 ranked groups of stocks, based on the book/market ratio. So far we've seen only one way of constructing new variables, that by using **`gen.`** Another such command is **`xtile`**:

```
. xtile me10=me, nq(10)
```

This command creates a new variable `me10`, based on values of `me`, grouped into ten (**`"q(10)"`**) groups. This last statement is an *option* for a command – an additional instruction on how exactly to do the

computations. The 10 groups are created to be approximately equal in size, to check that, let us tabulate the values of this new variable:

```
. tab me10
```

```
         10 |
  quantiles |
      of me |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        176       10.04       10.04
          2 |        175        9.98       20.02
          3 |        175        9.98       30.01
          4 |        176       10.04       40.05
          5 |        175        9.98       50.03
          6 |        175        9.98       60.01
          7 |        176       10.04       70.05
          8 |        175        9.98       80.03
          9 |        175        9.98       90.02
         10 |        175        9.98      100.00
------------+-----------------------------------
      Total |      1,753      100.00
```

The new variable takes values from 1 to 10, indicating that the corresponding stock belongs to that particular ranked group. Let us also check how the average values of different variables differ across these 10 groups:

```
. tabstat eps epr mktval bme dp ret5, s(mean) by(me10)
```

```
Summary statistics: mean
  by categories of: me10 (10 quantiles of me)

    me10 |       eps       epr  mktval~s       bme        dp      ret5
---------+--------------------------------------------------------------
       1 |  6.791798  .6965072  54.43373  1.387439  .0722566   .219612
       2 |  6.202881  .4946279  77.86297  1.052264  .0488254  .1886405
       3 |  7.762573   .551816  122.7272  1.041249  .0533842  .1833389
       4 |  9.226702  .5470097  150.6228  .8518151  .0402081  .1514274
       5 |  7.394273  .3534503  205.8057  .7774228  .0558928  .1510664
       6 |   11.4845  .5100749  386.4259  .9004726  .0537308  .1481498
       7 |   9.05238  .3185214  589.5316   .886191  .0423823  .1346627
       8 |  30.48633  1.324385  1031.318  .8770901  .0432942  .1924418
       9 |  73.11799  3.636735   1936.98  .8267867  .0419831  .1585319
      10 |  341.6869  9.889199  6106.672  .6531582  .0371551  .1376024
---------+--------------------------------------------------------------
   Total |  56.45648   2.01975  1169.715  .9123388  .0489159  .1665508
------------------------------------------------------------------------
```

The **tabstat** command computes the statistic, specified in the option "**s()**", for each value of a categorical variable, specified in "**by()**".

We can see that the average future return on common stock for the smallest companies is 22%, much higher than that of the largest companies (13.7%). Note that these average returns can be interpreted as returns on an equally-weighted portfolio of all stocks in that group.

Similarly, we can create groups based, for example, on the book/market ratio, and compute similar statistics there:

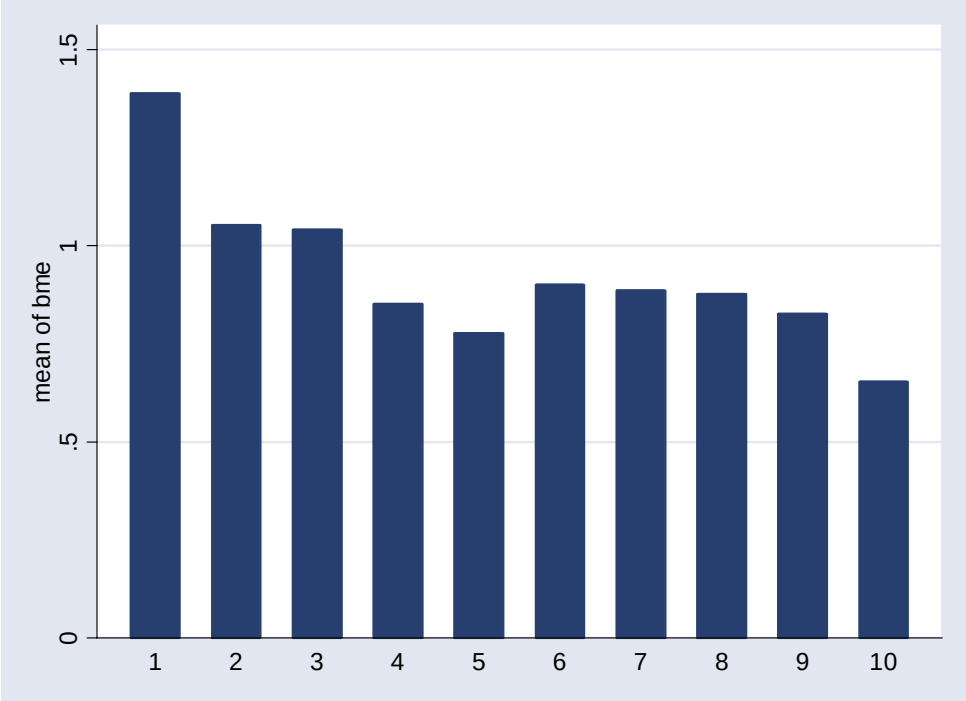
```
. tabstat eps epr mktval bme dp ret5, s(mean) by(bme10)

Summary statistics: mean
  by categories of: bme10 (10 quantiles of bme)

   bme10 |       eps       epr  mktval~s       bme        dp      ret5
---------+--------------------------------------------------------------
       1 |  10.54549  .1328178  610.1211  .0863635  .0429207  .1347002
       2 |  13.84976  .2927945  684.6179  .3563609  .0183622  .1117098
       3 |  71.83744  2.400372  1109.809   .481466  .0255522  .1372841
       4 |  57.49137  1.728079   1147.17   .612406  .0249824  .1524704
       5 |  129.1861  4.570877  1930.534  .7276851  .0442441  .2038519
       6 |  69.98866  1.585766  1449.554  .8738779  .0661087  .1737286
       7 |  23.81782  1.114036   908.648   1.02227  .0567815  .1698623
       8 |  11.39654  .4380677  1050.622   1.22347  .0481891  .1733748
       9 |  21.31581  1.215803  1559.935  1.487134  .0547627  .2073983
      10 |  154.0027  6.683219  1250.361  2.258993  .0706334   .235436
---------+--------------------------------------------------------------
   Total |  56.45648   2.01975  1169.715  .9123388  .0452488  .1699369
------------------------------------------------------------------------
```

Here we can see that the average future return on common stock for the companies with the highest B/M ratio is 23.5%, much higher than that of the lowest group (13.5%).

A summary bar plot

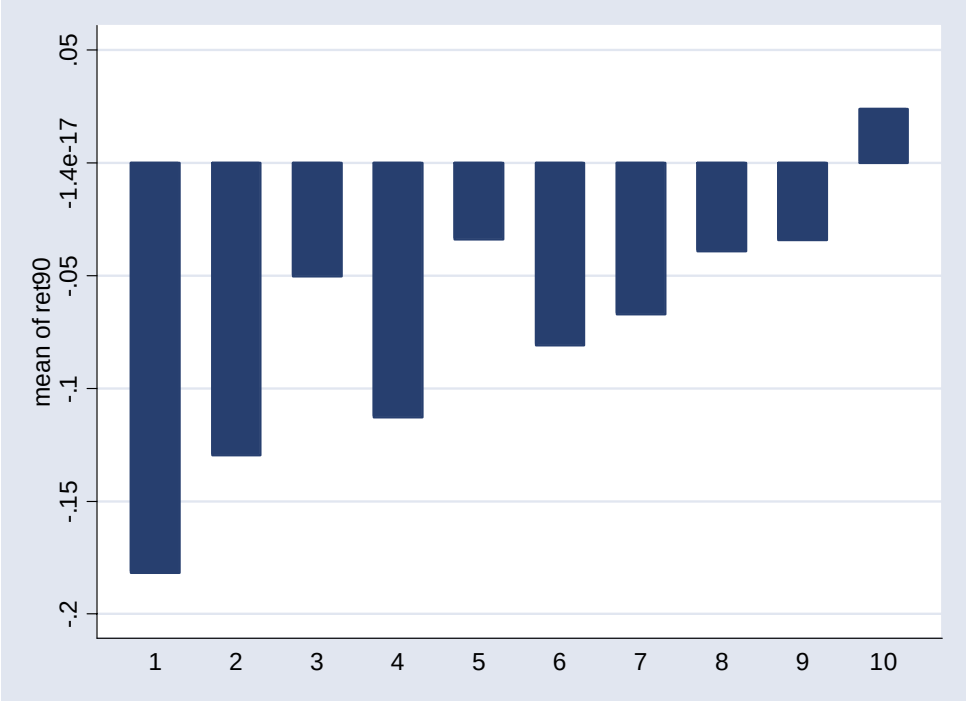
Of course, the book/market ratio and the market cap are correlated, as one can easily see from the summary tables we produced earlier. We can also look at their relationship on a plot:

```
. gr bar (mean) bme, over(me10)
```



The **gr bar (mean)** command automatically computed the mean of `bme` and plotted it over all the categories of `me10`. As expected, we see a negative relationship, especially pronounced at the lowest market cap levels: these are the stocks that have recently experienced poor returns.

```
. gr bar (mean) ret90, over(me10)
```



Hudson General analysis

Now let us look at some statistics for Hudson General.
First, load the data file for 1998:

```
.use all98.dta
```

Generate some additional statistics:

```
. gen eps=earn/shrout
. gen epr=eps/prc
. gen dp=div/prc
```

Take a look at Hudson General (HGC):

```
. l name ticker iname hsiccd if ticker=="HGC"
```

```
     +-------------------------------------------------------------------+
     |                name   ticker                            iname   hsiccd |
     |-------------------------------------------------------------------|
 19. | HUDSON GENERAL CORP      HGC   AIRPORTS & TERMINAL SERVICES     7394 |
     +-------------------------------------------------------------------+
```

Here, the **if ticker==”HGC”** statement has restricted the **list** command to the observations that satisfy the condition (note that to specify an equality, the “=” sign has to be repeated). `hsiccd` is the SIC industry code of the company, in this case 7394. The first three digits of this code (739) specify an industry group, and the first two (73) specify a “major industry group”, in this case Business Services. Let us see what other companies are there in the same industry group:

```
. l name ticker iname  if hsicig==739
```

```
     +-------------------------------------------------------------------+
     |                        name   ticker                          iname |
     |-------------------------------------------------------------------|
  1. |                   I B P INC      IBP            MEAT PACKING PLANTS |
 19. |         HUDSON GENERAL CORP      HGC   AIRPORTS & TERMINAL SERVICES |
304. |                   LASON INC     LSON   CMP PROGRAMMING,DATA PROCESS |
372. | GEOTEL COMMUNICATIONS CORP     GEOC           PREPACKAGED SOFTWARE |
565. |                  ARCADIS NV    ARCAF           ENGINEERING SERVICES |
     +-------------------------------------------------------------------+
```

Now let’s see some statistics on those companies:

```
. l ticker eps epr mktval bme dp ret98 ret5 if hsicig==739
```

```
     +-------------------------------------------------------------------------------------+
     | ticker        eps        epr     mktval        bme         dp       ret98        ret5 |
     |-------------------------------------------------------------------------------------|
  1. |    IBP   11.32418   .3888131    2703.79   .5023689   .0034335    .3974356   -.0685974 |
 19. |    HGC   16.75358   .2659299    109.935   .7996095    .015873    .3359825     2.08204 |
304. |   LSON   2.082292   .0357859   902.9536    .335761          0    1.185446    -.854272 |
372. |   GEOC   .3261969    .008757    1009.14   .0556112          0       3.768    1.603755 |
565. |  ARCAF   3.313642     .42078   148.6406   .6761166   .0392889   -.2409649     .127984 |
     +-------------------------------------------------------------------------------------+
```

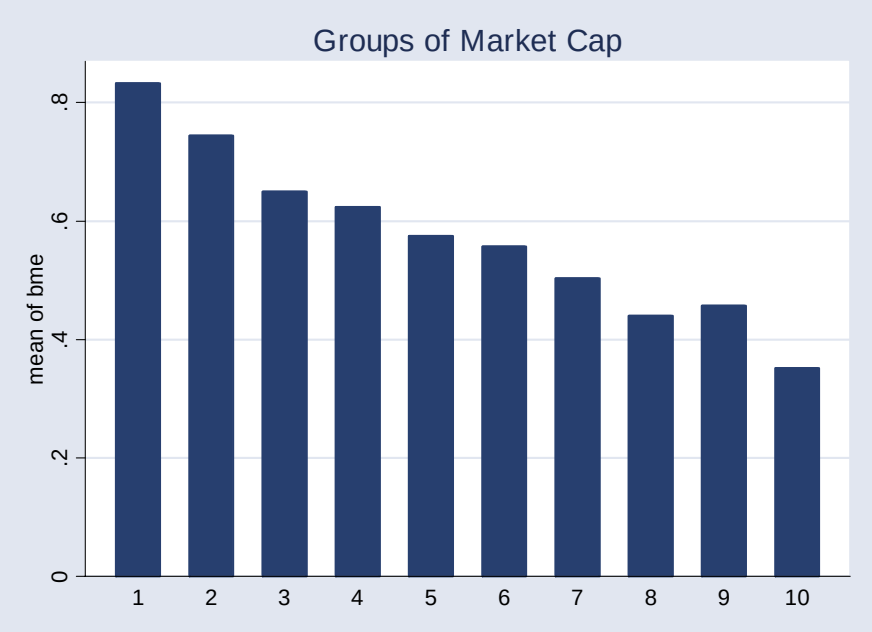
We can also display data on a scatter plot, for example like this:

```
. scatter bme me if hsicig==739, mlabel(ticker)
```

The last parameter **"mlabel(ticker)"** instructed Stata to place near each data point a label – in this case, the company's ticker symbol.

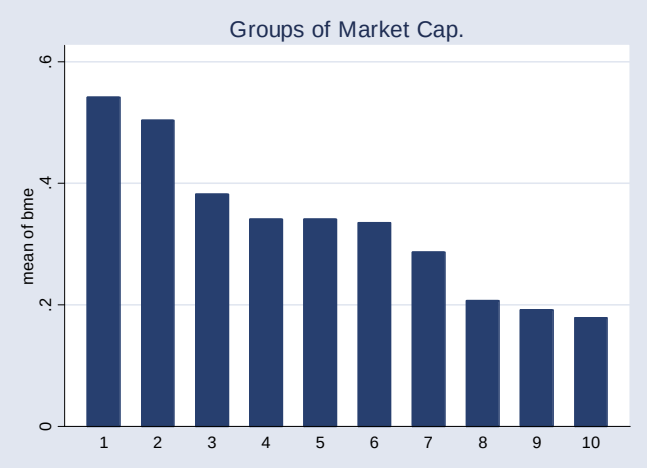
Homework1 Answers

```
. /* Part 1a */
. su bme me dp epr

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         bme |      5240    .5686128    .4472267  -2.075402   5.267616
          me |      5731    2907.297    13112.76    4.20525   333672.2
          dp |      5930    .0185453    .0484703          0   2.271852
         epr |      5274    1.230452    16.39802  -55.38419   668.0734

. /* Part 1b */
. xtile me10=me, nq(10)

. tabstat eps  epr mktval bme dp, s(mean) by(me10)

Summary statistics: mean
  by categories of: me10 (10 quantiles of me)

    me10 |       eps       epr    mktval       bme        dp
---------+--------------------------------------------------
       1 |   3.08183  .2147178  32.36216  .8327547  .0215133
       2 |  2.176366  .1434744  65.96726  .7439956  .0178496
       3 |  2.215206  .1106453  102.8396  .6500231  .0208976
       4 |  3.090467  .1480313  154.3239  .6234538  .0188834
       5 |  4.386719  .2248164  238.7836  .5744724  .0177955
       6 |  5.530824  .2404611  364.5323  .5568566  .0185632
       7 |  5.938522  .2599116  580.7353  .5036616  .0181217
       8 |  17.85195  .8504423  1042.505  .4400238  .0170886
       9 |  123.1223  4.060059  2276.379  .4573416  .0156977
      10 |  286.6091  5.458061  18484.25  .3514821  .0170179
---------+--------------------------------------------------
   Total |  48.78835  1.248927  2333.867  .5686128  .0183414
------------------------------------------------------------

.gr bar (mean) bme, over(me10) title("Groups of Market Cap.")
```




```
. /* Part 2 */
. xtile bme10=bme, nq(10)

. tabstat ret5 eps epr mktval bme dp, s(mean) by(bme10)

Summary statistics: mean
  by categories of: bme10 (10 quantiles of bme)

   bme10 |      ret5       eps       epr    mktval       bme        dp
---------+--------------------------------------------------------------
       1 |  .1114811  -9.46297 -.4852172  7784.574 -.0065247  .0059963
       2 |  .0213247  5.851604  .0932791  5261.255  .1719405  .0065261
       3 |  .0553223  34.56831   .768163   2896.32  .2665535  .0064367
       4 |  .0162675  40.98703  .8380865  3362.296  .3572687  .0096505
       5 |  .1016196  26.10176  .5733465  1616.691  .4517523  .0109517
       6 |   .172112   102.555  3.134946  1051.511  .5468151  .0142634
       7 |  2.736393  76.46099  .9267213   1045.78  .6534455  .0165595
       8 |  .2607621  44.77619  1.512045  719.8845  .7879832  .0203103
       9 |  .1332738   115.362  3.086587   661.424  .9677376  .0263889
      10 |   .149084  37.04397  1.737113   383.291  1.489156   .027886
---------+--------------------------------------------------------------
   Total |  .3756401  47.51699  1.220517  2478.303  .5686128  .0145021
------------------------------------------------------------------------

. /* Part 3 */
. l name ticker iname be me eps if hsicig==739

     +-------------------------------------------------------------------------------------------------+
     |                     name   ticker                          iname         be         me        eps |
     |-------------------------------------------------------------------------------------------------|
 171. |                I B P INC      IBP            MEAT PACKING PLANTS   1350.692   2688.645   11.32418 |
1532. |        HUDSON GENERAL CORP      HGC   AIRPORTS & TERMINAL SERVICES     70.638    88.34062   16.75358 |
5199. |                LASON INC     LSON   CMP PROGRAMMING,DATA PROCESS    287.039    854.8907   2.082292 |
5280. | GEOTEL COMMUNICATIONS CORP     GEOC           PREPACKAGED SOFTWARE     55.757   1002.621   .3261969 |
6025. |                ARCADIS NV    ARCAF           ENGINEERING SERVICES    102.527    151.641   3.313642 |
     +-------------------------------------------------------------------------------------------------+
```

```
. /* Part 4 */
.
. save nyse98a, replace
file nyse98a.dta saved

. keep if hsicmg==73
(5463 observations deleted)
.
. /* Part 4.1a */
. su bme me dp epr

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         bme |       536    .3317828    .3033717  -.3734393   2.601382
          me |       539    1850.572    12264.12       9.25   267686.3
          dp |       558     .001977    .0110414          0   .1877807
         epr |       546    .4061718    9.494692  -55.38419   202.2024
.
. /* Part 4.1b */
. xtile me10=me, nq(10)
. tabstat eps  epr mktval bme dp, s(mean) by(me10)

Summary statistics: mean
  by categories of: me10 (10 quantiles of me)

    me10 |       eps       epr    mktval       bme        dp
---------+--------------------------------------------------
       1 |  .1309977 -.0085133  34.65505  .5413646  .0008217
       2 |  .4968812 -.0254595  75.62775  .5034658   .003104
       3 |  -.026774 -.0303351   125.114  .3820832  .0007111
       4 | -.7313988 -.0809841   186.313  .3404816  .0008445
       5 |  2.345342  .1149984  275.2439  .3406725  .0002394
       6 |  4.770126  .0891176  376.9049  .3350131  .0010716
       7 |  .2362984 -.0891747  557.8234  .2871116  .0031214
       8 |  .8485455  .0244454  1010.179  .2074182  .0012348
       9 |  .6089149  .0107427  1914.736  .1919285  .0010746
      10 |  304.5731  5.233272  15175.14  .1789721  .0061833
---------+--------------------------------------------------
   Total |  30.87461  .5160783  1948.681  .3317828  .0018326
------------------------------------------------------------
.gr bar (mean) bme, over(me10) title("Groups of Market Cap.")
```



```
. /* Part 5 */
. scatter epr me if hsicig==739, mlabel(ticker)
```