TO START THIS MATHEMATICA NOTEBOOK YOU CLICK ITS FILENAME.
You will have to use a computer in a university lab (e.g. Wells Hall B-Wing)

This *Mathematica* notebook contains a number of useful functions described in the handout and briefly indicated below. The first time you attempt to use one of these functions a panel will pop up asking "Do you want to evaluate all the initialization cells?" to which you must answer yes.

To enter a given command line you click on the screen whereupon a horizontal line should appear at the cursor. When right brackets are in view on the *Mathematica* panel you want to click at a place where a horizontal line will extend between two such brackets if you desire a new line. If you attempt to type multiple commands into a single bracketed location *Mathematica* will become confused.

Type the command you wish to execute then PRESS THE ENTER KEY ON THE NUMERIC KEYPAD. This is required because *Mathematica* wants to use the return or other enter key to move to the next line. You do nor want to move to a new line. You want to enter a command. That is why you must use the ENTER key on the numeric keypad.

To save your work select save from the pull down file menu, which saves it as a *Mathematica* .nb (notebook) file. If you wish to print your work at home select print then the option of saving as a PDF. You will be unable to work with the .nb *Mathematica* file itself unless you have *Mathematica* installed (unlikely) but you can transport and print the .pdf file virtually anywhere.

Click the line below and press ENTER on the numeric keypad.

```
In[44]:= size[{4.5, 7.1, 7.8, 9.1}]

Out[44]= 4
```

Just above, I clicked to open a new line then typed

size[{4.5, 7.1, 7.8, 9.1}]

followed by a press of the numeric keypad ENTER key. Notice that off to the right of the entry there are nested brackets joining the command line and its output 4 = the number of data items in {4.5, 7.1, 7.8, 9.1}.

▪ A complete list of the commands in this notebook and what they do.

size[{4.5, 7.1, 7.8, 9.1}] returns 4
mean[{4.5, 7.1, 7.8, 9.1}] returns the mean 7.125
median[{4.5, 7.1, 7.8, 9.1}] returns the median of the list {4.5, 7.1, 7.8, 9.1}
s[{4.5, 7.1, 7.8, 9.1}] returns the sample standard deviation s=1.93628
sd[{4.5, 7.1, 7.8, 9.1}] returns the n-divisor version of standard deviation s=1.67686
r[x, y] returns the sample correlation $r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\,\sqrt{\overline{y^2}-\bar{y}^2}}$ for paired data.
sample[{4.5, 7.1, 7.8, 9.1}, 10] returns 10 samples from`{4.5, 7.1, 7.8, 9.1}`
ci[{4.5, 7.1, 7.8, 9.1}, 1.96] returns a 1.96 coefficient CI for the mean from given data
bootci[mean, {4.5, 7.1, 7.8, 9.1}, 10000, 0.95] returns 0.95 bootstrap ci for pop mean
smooth[{4.5, 7.1, 7.8, 9.1}, 0.2] returns the density for data at bandwidth 0.2
smooth2[{4.5, 7.1, 7.8, 9.1}, 0.2] returns the density for data at bandwidth 0.2
overlaid with normal densities having sd = 0.2 around each data value
smoothdistribution[{{1, 700},{4 ,300}}, 0.2] returns the density at bandwidth 0.2
for a list consisting of 700 ones and 300 fours.
popSALES is a file of 4000 sales amounts used for examples
entering **popSALES** will spill 4000 numbers onto the screen. To prevent
that enter **popSALES;** instead (the appended semi-colon suppresses output).
betahat[matrix x, data y] returns the least squares coefficients $\hat{\beta}$ for a fit of the model $y = x\beta + \epsilon$.
resid[matrix x, data y] returns the estimated errors $\hat{\epsilon} = y - x\hat{\beta}$ (see **betahat** above).
R[matrix x, data y] returns the **multiple correlation** between the fitted values $x\hat{\beta}$ and data y.
xquad[matrix x] returns the full quadratic extension of a design matrix with constant term
xcross[matrix x] returns the extension of x to include all products of differing columns.
betahatCOV[x matrix, data y] returns the estimated covariance matrix of the vector betahat β.
normalprobabilityplot[data, dotsize] returns a normal probability plot for data (e.g. with dotsize .01).
t[df, conf] returns the t-score used in lieu pf z-score in a CI for confidence conf (t[Infinity, .95] ~ 1.96).
Tprob[t, df] returns P(|T| < t) for df (e.g. T[1.96, Infinity] ~ 0.95).

HW6 involves executing the code below. Your results will vary from those shown.

Submit a printout of your work below. Your results will be compared with those of other students.

Sign your work.

```
In[32]:= mean[popSALES]

Out[32]= 15.2302

In[33]:= sd[popSALES]

Out[33]= 9.41429
```

In *Mathematica* the percent character % refers to the output of the very last command execution.

The next line finds a sample of 40 from popSALES.

```
In[34]:= mysample = sample[popSALES, 40];
```

The next line finds the sample mean. The line following that finds a 68% z - CI for the population mean using the formula you have learned for MOE of sample mean.

```
In[35]:= mean[mysample]

Out[35]= 15.7382

In[36]:= mean[mysample] + 1.00 {-1, 1} s[mysample] / Sqrt[40]

Out[36]= {14.1424, 17.3341}
```

The next line finds the above confidence interval via a bundled routine. It outputs
{mean, n, s, z (or t), CI}
and is specific to the purpose of using the MOE approach to determine the (same) CI for the mean.

```
In[37]:= ci[mysample, 1.00]

Out[37]= {15.7382, 40., 10.0932, 1., {14.1424, 17.3341}}
```

Has your regular ci covered μ?

Yes, 15.2302 is within the range {14.1424, 17.3341}

Around what percentage of regular 68% ci for μ cover μ?

68%

The next line employs 10000 replications of

a. Selecting a with - repl eq - pr sample of 40 from mysample (a bootstrap-sample).
b. Calculating the bootstrap-sample mean.

The routine then increasingly orders the 10000 values $|\overline{X}^* - \overline{X}|$.
Finally, it determines the 68th percentile ♯*of the above 10000 values.
The 68% bootstrap ci for μ is then reported as $\overline{X}$+ {-1,1} ♯*.

By hand, write the brief probability expression setting forth the performance claim made for the regular 68% ci for μ.

P(|xbar - mu| < $\frac{10.0932}{\sqrt{40}}$) = .68

By hand, write the brief probability expression seting forth the performance claim made for the bootstrap 68% ci for μ.

P(|xbar - mu| < ♯*) = .68

In[45]:= **bootci[mean, mysample, 10 000, 0.68]**

Out[45]//MatrixForm=

Confidence Level	0.68
Estimator	mean
Estimate	15.7382
Sample Size	40
number of replications	10 000
bootstrap ci Half Width	1.55775
bootstrap confidence interval	{14.1805, 17.296}

Has your bootstrap 68% ci covered μ?

yes it has, 15.2302 is within the range {14.1753, 17.3012}

Is your 68% bootstrap ci in fairly close agreement with your regular 68% ci?

{14.1424, 17.3341} is definitely in agreement with {14.1753, 17.3012}

Try the bootstrap again. There will be a slight difference due to the randomness of the 10000 replications upon which the bootstrap is based in these examples.

still in agreement, new interval {14.1805, 17.296}

Next, we turn our attention to estimating the population median.

In[39]:= **median[popSALES]**

Out[39]= 13.21

In[40]:= **median[mysample]**

Out[40]= 13.51

We will not present a formula for a "regular" 68 % ci for the population median, for which we have not offered any formula for a "regular" 68% ci.

The bootstrap offers a solution.

In[41]:= **bootci[median, mysample, 10 000, 0.68]**

Out[41]//MatrixForm=

Confidence Level	0.68
Estimator	median
Estimate	13.51
Sample Size	40
number of replications	10 000
bootstrap ci Half Width	2.69
bootstrap confidence interval	{10.82, 16.2}

Has the bootstrap 68 % interval covered the population median?

Yes it has.

Around what percentage of your classmates should answer yes to the above question.

68%

Additional exercises.

1. Women are scored $x = 1$. They are 3/27 of the population. On score y they average \$4 .67 with variance 9.34.

   Men are scored $x = 0$. They are 24/27 of the population. On score y they average 5.22 with variance 7.33.

Determine:

Within component of variance Y = E Var (Y | X).

$$\text{E Var}(Y \mid X) = 9.34(3/27) + 7.33(24/27) = 7.5533$$

Between component of variance Y = Var (E (Y | X)).

$$Y = \text{Var}(\text{E}(Y \mid X)) = 4.67^2(3/27) + 5.22^2(24/27) - (4.67(3/27) + 5.22\,(24/27))^2 = 0.02988$$

$$\text{Var } Y = \text{E Var}(Y \mid X) + \text{Var}(\text{E}(Y \mid X))$$

$$\text{Var } Y = \text{E Var}(Y \mid X) + \text{Var}(\text{E}(Y \mid X)) = 7.58318$$

2. For the joint x, y probability distribution

x \ y	4	6
1	.1	.3
3	.2	.4

Determine:

a. Conditional distribution P(Y = 4 | X = 1), P(Y = 6 | X = 1).

$P(Y = 4 \mid X = 1) = \frac{.1}{(.1+.3)} = .25$

$P(Y = 6 \mid X = 1) = \frac{.3}{(.1+.3)} = .75$

b. E(Y | X = 1) = 4 P(Y = 4 | X = 1) + 6 P(Y = 6 | X = 1).

$E(Y \mid X = 1) = 4\,P(Y = 4 \mid X = 1) + 6\,P(Y = 6 \mid X = 1) = 4(.25) + 6(.75) = 5.5$

c. $\mathrm{Var}(Y \mid X = 1) = (4^2\,P(Y = 4 \mid X = 1) + 6^2 P(Y = 6 \mid X = 1)) - (E(Y \mid X = 1))^2$.

$= 4^2(.25) + 6^2(.75) - 5.5^2 = .75$

d. Var Y (direct calculation from the table).

$\mathrm{Var}\,Y = E(Y^2) - (E(Y))^2$

$E(Y^2) = 4^2(.1 + .2) + 6^2(.3 + .4) = 30$

$(E(Y))^2 = (4(.1 + .2) + 6(.3 + .4))^2 = 29.16$

$\mathrm{Var}\,Y = E(Y^2) - (E(Y))^2 = .84$

e. within component E Var(Y | X) = .4 Var(Y | X = 1) + .6 Var(Y | X = 3).

$$E(Y \mid X=3) = 4\left(\frac{.2}{.2+.4}\right) + 6\left(\frac{.4}{.2+.4}\right) = 5.33333$$
$$Var(Y \mid X=3) = 4^2\left(\frac{.2}{.2+.4}\right) + 6^2\left(\frac{.4}{.2+.4}\right) - 5.33333^2 = .889$$

$$E\,Var(Y \mid X) = .4(.75) + .6(.889) = 0.8334$$

f. between component Var E(Y | X) = variance of the distribution
E(Y | X = 1) with probability .4
E(Y | X = 3) with probability .6

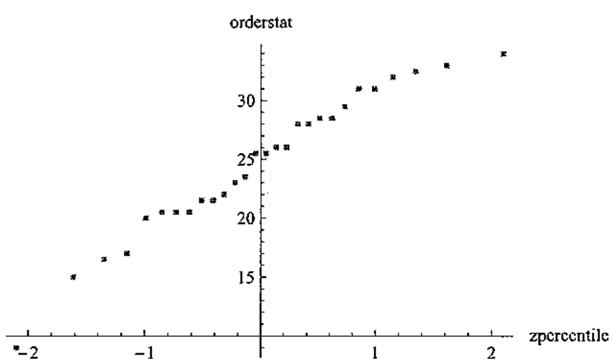
$$5.5^2\,(.4) + 5.33333^2\,(.6) - (\,5.5(.4) + 5.33333\,(.6))^2 = .006667$$

g. verify that Var Y = within + between

$0.8334 + .006667 = .840067 \approx .84$

```
In[42]:= examscores = {9, 21.5, 16.5, 28.5, 28, 28.5, 22, 26, 29.5, 21.5, 20.5,
   32.5, 20.5, 32, 34, 25.5, 17, 33, 28, 31, 20.5, 20, 15, 31, 23, 25.5, 23.5, 26}
Out[42]= {9, 21.5, 16.5, 28.5, 28, 28.5, 22, 26, 29.5, 21.5, 20.5, 32.5,
   20.5, 32, 34, 25.5, 17, 33, 28, 31, 20.5, 20, 15, 31, 23, 25.5, 23.5, 26}
In[48]:= normalprobabilityplot[examscores, 0.01]
```

In[47]:=




```
In[49]:= ci[examscores, 1.96]
Out[49]= {24.625, 28., 6.0898, 1.96, {22.3693, 26.8807}}
```

By inspection more than 5% of that data was not within our confidence interval so I would conclude that this data is not normally distributed.

However the plot does appear to be fairly linear which provides evidence that exam scores may in fact be somewhat normally distributed