stat200

A number of statistical routines are programmed into this Mathematica notebook file. To run them you must boot the notebook from a university lab by

a. navigating to www.stt.msu.edu/~lepage
b. clicking on STT200
c. clicking on stat200 (stat200 will launch)
d. clicking on the 1+1 line below
e. performing SHIFT+ENTER.

Begin by clicking the line below then performing SHIFT +ENTER. Respond YES when you are asked whether to evaluate cells.

```
1 + 1
```

```
2
```

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 simple (i.e. equally weighted) mean 7.125
avg[{4.5, 7.1, 7.8, 9.1}] also returns the simple (i.e. equally weighted) mean 7.125
mean[{4.5, 7.1}, {2, 5}] returns the {2/7, 5/7} weighted mean 6.357 of {4.5, 7.1}
avg[{4.5, 7.1}, {2, 5}] also returns the {2/7, 5/7} weighted mean 6.357 of {4.5, 7.1}
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 = $\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2}-\overline{x}^2}\,\sqrt{\overline{y^2}-\overline{y}^2}}$ for paired data.
regrtable[x,y] returns a table illustrating calculations of $\overline{x}$, $\overline{y}$, $\overline{x^2}$, $\overline{y^2}$, $\overline{xy}$.
regrstats[x, y] returns $\overline{x}$, $\overline{y}$, s_x, s_y, r, and the slope of the regression line = r $\frac{s_y}{s_x}$.
regrplot[x,y] returns the plot of (x, y) pairs overlaid with the regression line.
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;** (append semi-colon to suppress output).
betahat[matrix x, data y] returns the least squares coefficients $\hat{\beta}$ for a fit of the model y = x β + ϵ.
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 betâhat β.

By clicking on any of the examples below you can execute it afresh by performing SHIFT+ENTER. Or click anywhere between lines, or at the end of the file, to make a fresh line and type your own examples.

Median of a list (i.e. Q_2).

```
med[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}]

5.
```

25 % quartile Q_1.

```
qtile[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}, .25]
```

```
2.
```

75 % quartile Q_3.

```
qtile[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}, .75]
```

```
7.
```

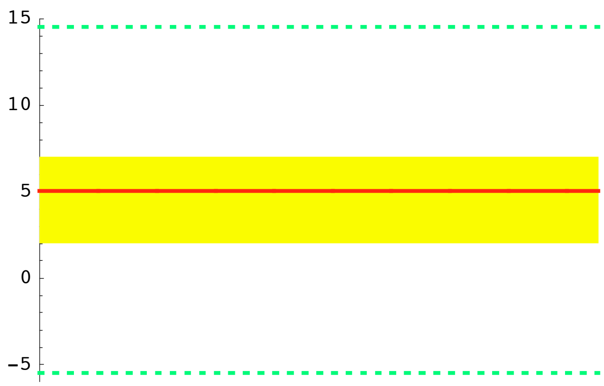
Interquartile range $Q_3 - Q_1$.

```
iqr[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}]
```

```
5.
```

Boxplot with median in red and interval $[Q_1, Q_3]$ illustrated as a yellow band.
Green lines at $Q_1 - 1.5$ **iqr** and $Q_3 + 1.5$ **iqr** are not usually shown in the plot.
The largest and smallest data values in the green zone are plotted with whiskers.
Do this by hand for the example below.

```
boxplot[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}]
```




```
- Graphics -
```

The computer can illustrate the calculations for a boxplot.

```
specimendata = {1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}
```

```
{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}
```

Here is the calculation of the upper green line of the boxplot just above.

```
qtile[specimendata, 0.75] + 1.5 iqr[specimendata]
```

```
14.5
```

Simple average (equal weight average) $\frac{\sum x}{n}$.

```
avg[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}]
```

```
9.09091
```

Sample standard deviation s = $\sqrt{\frac{\sum (x-\overline{x})^2}{n-1}}$.

```
s[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}]
```

```
15.7382
```

Simple average.

```
avg[{1, 2, 3, 4}]
```

```
2.5
```

Simple average is also denoted "mean."

```
mean[{1, 2, 3, 4}]
```

```
2.5
```

Simple average is a weighted average by equal weights 1.

```
avg[{1, 2, 3, 4}, {1, 1, 1, 1}]
```

```
2.5
```

Only the relative sizes of the weights matters. We may as well use weights of 2.

```
avg[{1, 2, 3, 4}, {2, 2, 2, 2}]
```

```
2.5
```

Average of a larger list with duplicate values.

```
avg[{1, 1, 1, 1, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4}]
```

```
2.64286
```

The above average is a weighted average of 4 ones, 3 twos, 1 three, and 6 fours.

```
? avg
```

```
Global`avg

avg[list_] := Mean[list] 1.

avg[list_, wts_] := Mean[list wts] 1. / Mean[wts]
```

```
avg[{1, 2, 3, 4}, {4, 3, 1, 6}]
```

```
2.64286
```

```
avg[{4.5, 7.1}, {2, 5}]
```

```
6.35714
```

Sample standard deviation.

```
s[{1, 2, 3, 4}]
```

```
1.29099
```

The sample standard deviation above is the same as a weighted sample standard deviation.

```
s[{1, 2, 3, 4}, {1, 1, 1, 1}]
```

```
1.29099
```

The sample standard deviation of an unequally distributed list.

```
s[{1, 1, 1, 1, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4}]
```

```
1.33631
```

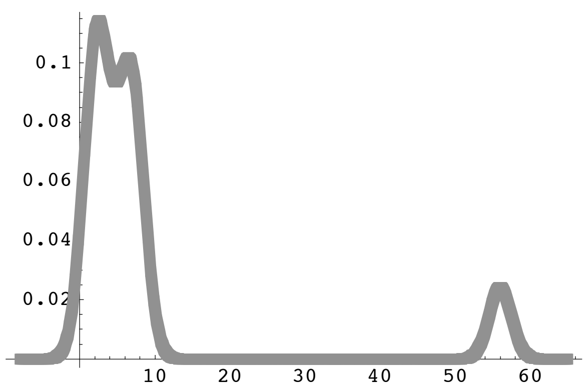
The same s above, calculated as a weighted sample standard deviation. Since "s" uses n-1 divisor, the weights must be actual frequency counts (as opposed to weights in "avg" whose relative sizes are all that matter).

```
s[{1, 2, 3, 4}, {4, 3, 1, 6}]
```

```
1.33631
```

Smoothing a list (an alternative to the histogram). First, with bandwidth 1.5

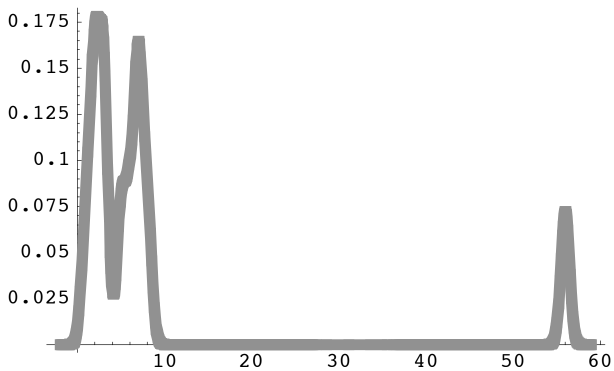
```
smooth[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}, 1.5]
```




```
- Graphics -
```

Smoothing a list (an alternative to the histogram). This time with a smaller bandwidth 0.5 which reveals more detail but may be too responsive to small chance variations in data.

```
smooth[{1, 2, 2, 6, 5, 3, 3, 8, 56, 7, 7}, 0.5]
```




```
- Graphics -
```

Ch. 5 Exercise 15. ne/mw and w/se states are entered as mydat**a** and mydat**b**.

```
mydata = {1, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 10, 10, 11, 12}

{1, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 9, 9, 9, 10, 10, 11, 12}

Length[mydata]

19

med[mydata]

8.

qtile[mydata, 0.25]

5.

qtile[mydata, 0.75]

9.

iqr[mydata]

4.

mydatb = {1, 6, 9, 9, 9, 10, 10, 11, 11, 13, 14, 14,
  14, 14, 15, 17, 18, 20, 20, 21, 21, 23, 24, 26, 30, 30, 31, 40, 66}

{1, 6, 9, 9, 9, 10, 10, 11, 11, 13, 14, 14, 14,
 14, 15, 17, 18, 20, 20, 21, 21, 23, 24, 26, 30, 30, 31, 40, 66}

Length[mydatb]

29

med[mydatb]

15.
```

```
qtile[mydatb, 0.25]
```

10.5

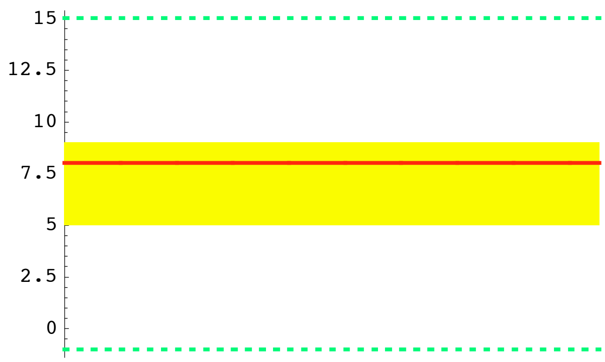
```
qtile[mydatb, 0.75]
```

23.5

```
iqr[mydatb]
```

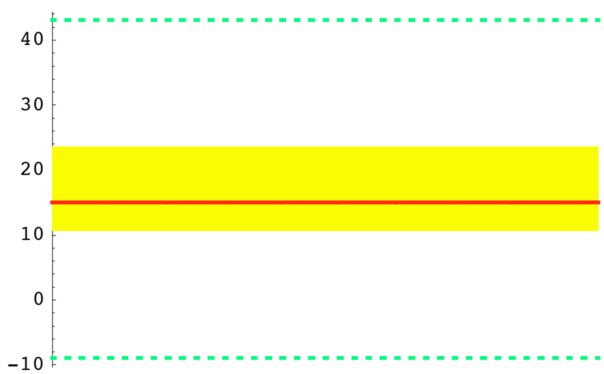
13.

```
boxplot[mydata]
```



- Graphics -

```
boxplot[mydatb]
```



- Graphics -

▪ Exercise. 7-17

```
hp = {170, 205, 190, 125, 310, 285, 127, 140, 215, 210, 170, 140, 194, 115}
```

{170, 205, 190, 125, 310, 285, 127, 140, 215, 210, 170, 140, 194, 115}

```
mpg = {22, 20, 15, 31, 10, 13, 29, 25, 21, 23, 18, 23, 21, 29}
```

{22, 20, 15, 31, 10, 13, 29, 25, 21, 23, 18, 23, 21, 29}

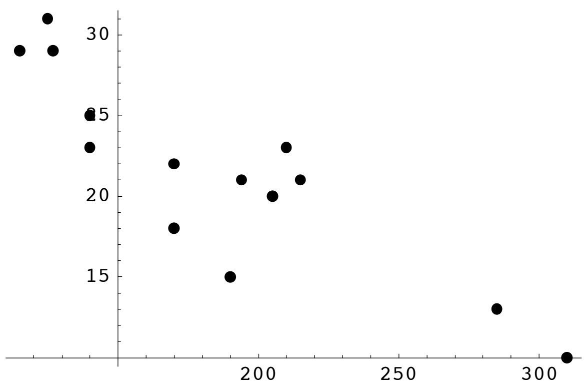
regrtable[hp, mpg]

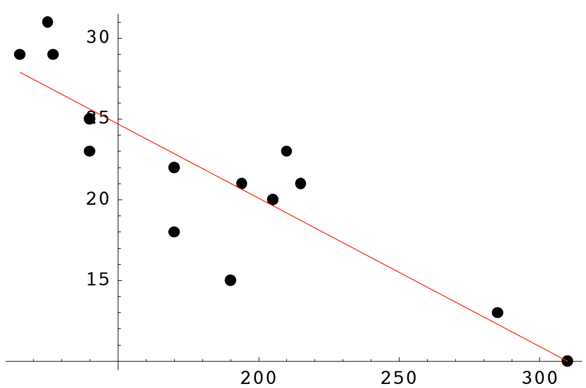
x	y	x^2	y^2	xy
170	22	28900	484	3740
205	20	42025	400	4100
190	15	36100	225	2850
125	31	15625	961	3875
310	10	96100	100	3100
285	13	81225	169	3705
127	29	16129	841	3683
140	25	19600	625	3500
215	21	46225	441	4515
210	23	44100	529	4830
170	18	28900	324	3060
140	23	19600	529	3220
194	21	37636	441	4074
115	29	13225	841	3335
_	_	_	_	_
185.429	21.4286	37527.9	493.571	3684.79

regrstats[hp, mpg]

{185.429, 21.4286, 58.189, 6.08547, -0.877954, -0.0918175}

regrplot[hp, mpg]





- Graphics -

■ Exercise. 7-18

```
mar = {22, 17, 40, 5, 37, 19, 23, 6, 7, 53, 34}
```

{22, 17, 40, 5, 37, 19, 23, 6, 7, 53, 34}

```
oth = {4, 3, 21, 1, 16, 8, 14, 3, 3, 31, 24}
```

{4, 3, 21, 1, 16, 8, 14, 3, 3, 31, 24}

```
?regrstats
```

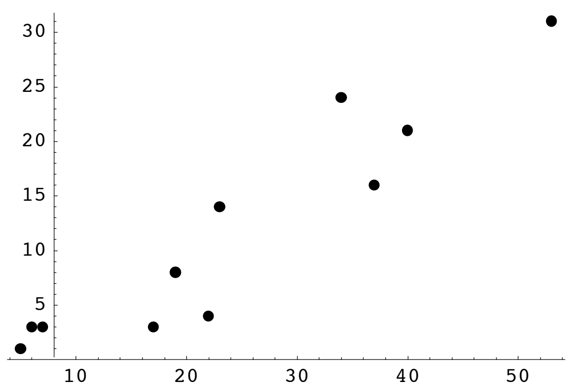
Global`regrstats

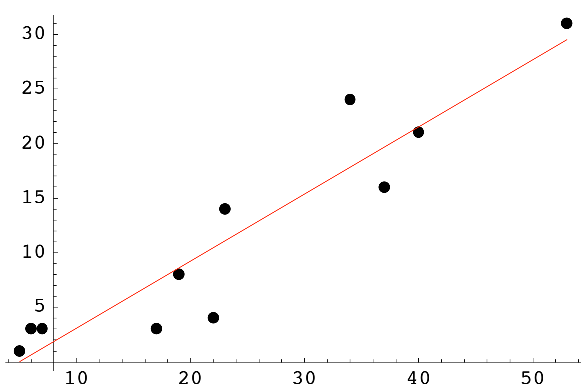
regrstats[x_, y_] := $\left\{\text{mean}[x], \text{mean}[y], s[x], s[y], r[x, y], \frac{r[x,y]\, s[y]}{s[x]}\right\}$ 1.

```
regrstats[mar, oth]
```

{23.9091, 11.6364, 15.5528, 10.2399, 0.9341, 0.615003}

```
regrplot[mar, oth]
```





- Graphics -

■ Exercise. 7-28

```
fat = {19, 31, 34, 35, 39, 39, 43}
```

{19, 31, 34, 35, 39, 39, 43}

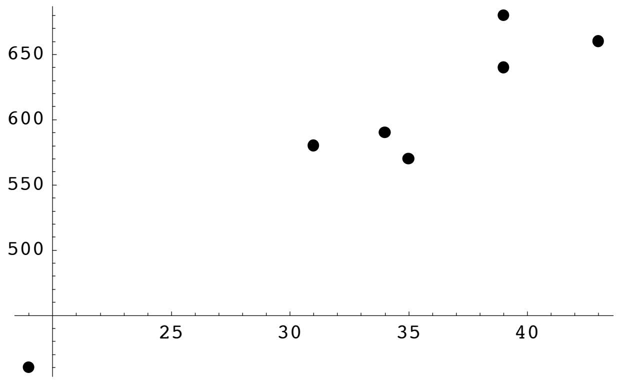
```
sod = {410, 580, 590, 570, 640, 680, 660}
```

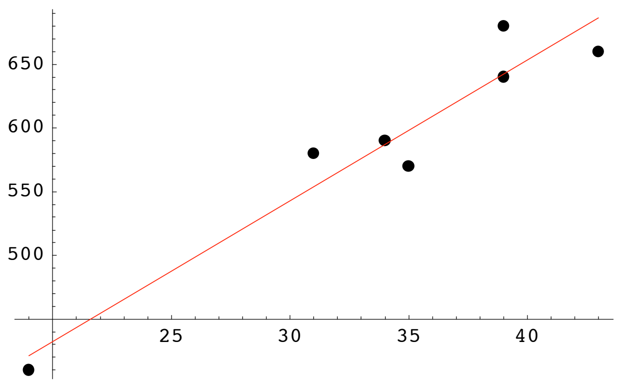
{410, 580, 590, 570, 640, 680, 660}

```
regrstats[fat, sod]
```

{34.2857, 590., 7.80415, 89.8146, 0.960633, 11.0555}

```
regrplot[fat, sod]
```





- Graphics -

■ Exercise. 7-17

```
fat = {19, 31, 34, 35, 39, 39, 43}
```

{19, 31, 34, 35, 39, 39, 43}

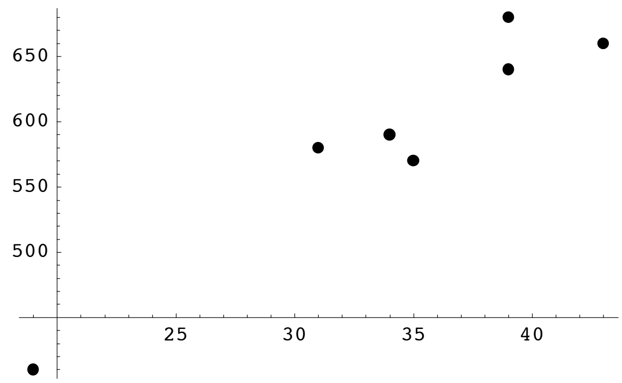
```
sod = {410, 580, 590, 570, 640, 680, 660}
```

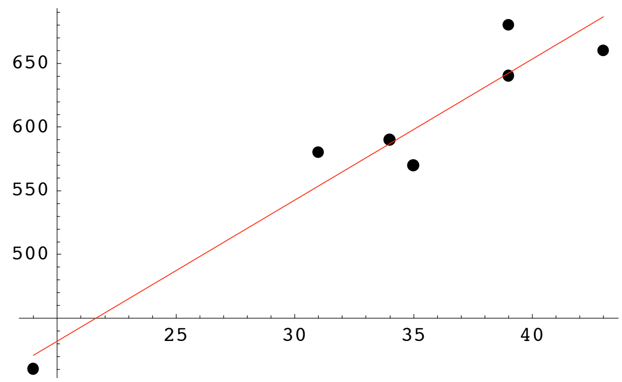
{410, 580, 590, 570, 640, 680, 660}

```
regrstats[fat, sod]
```

{34.2857, 590., 7.80415, 89.8146, 0.960633, 11.0555}

```
regrplot[fat, sod]
```





- Graphics -

■ Exercise. 7-18

```
mar = {22, 17, 40, 5, 37, 19, 23, 6, 7, 53, 34}
```

{22, 17, 40, 5, 37, 19, 23, 6, 7, 53, 34}

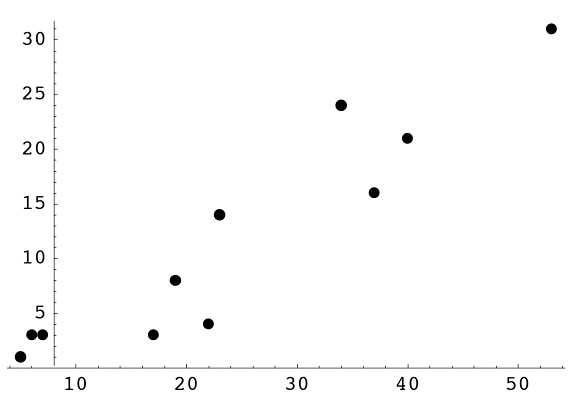
```
oth = {4, 3, 21, 1, 16, 8, 14, 3, 3, 31, 24}
```

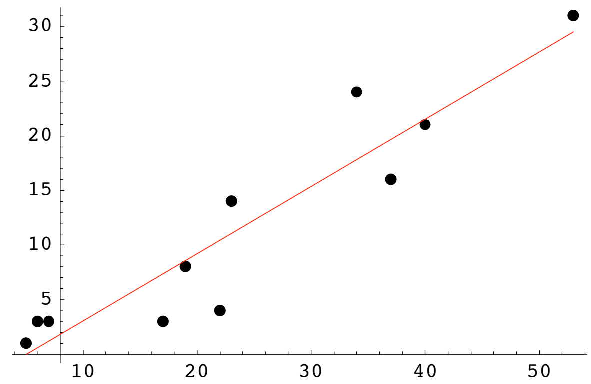
{4, 3, 21, 1, 16, 8, 14, 3, 3, 31, 24}

```
regrstats[mar, oth]
```

{23.9091, 11.6364, 15.5528, 10.2399, 0.9341, 0.615003}

```
regrplot[mar, oth]
```





- Graphics -

■ Exercise. 7-28

```
fat = {19, 31, 34, 35, 39, 39, 43}
```

```
{19, 31, 34, 35, 39, 39, 43}
```

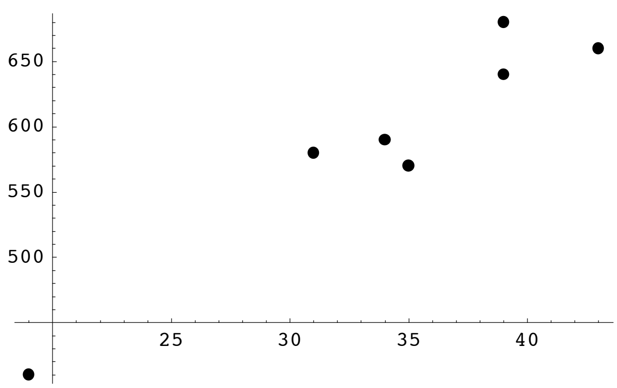
```
sod = {410, 580, 590, 570, 640, 680, 660}
```

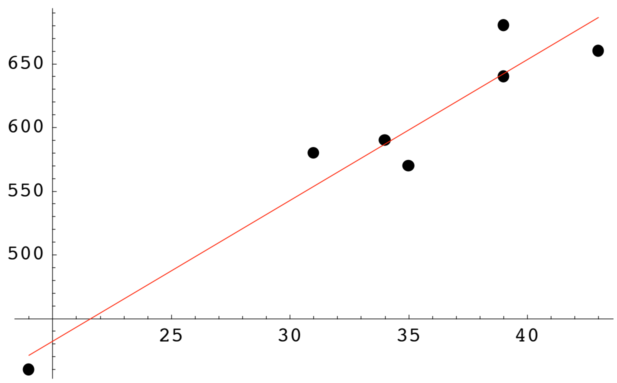
```
{410, 580, 590, 570, 640, 680, 660}
```

```
regrstats[fat, sod]
```

```
{34.2857, 590., 7.80415, 89.8146, 0.960633, 11.0555}
```

```
regrplot[fat, sod]
```






```
- Graphics -
```

■ Exercise. 8-43

```
wt2 = {175, 181, 200, 159, 196, 192, 205, 173,
  187, 188, 188, 240, 175, 168, 246, 160, 215, 159, 146, 219}
```

```
{175, 181, 200, 159, 196, 192, 205, 173, 187,
 188, 188, 240, 175, 168, 246, 160, 215, 159, 146, 219}
```

```
bf2 = {6, 21, 15, 6, 22, 31, 32, 21, 25, 30, 10, 20, 22, 9, 38, 10, 27, 12, 10, 28}
```

```
{6, 21, 15, 6, 22, 31, 32, 21, 25, 30, 10, 20, 22, 9, 38, 10, 27, 12, 10, 28}
```

```
?regrstats
```

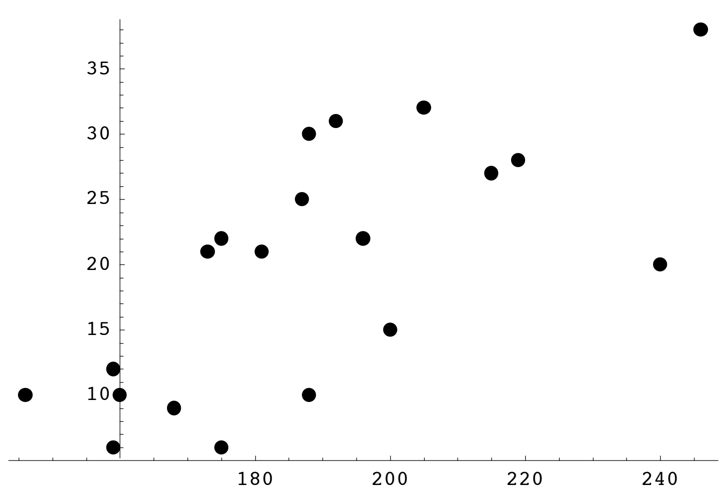
```
Global`regrstats
```

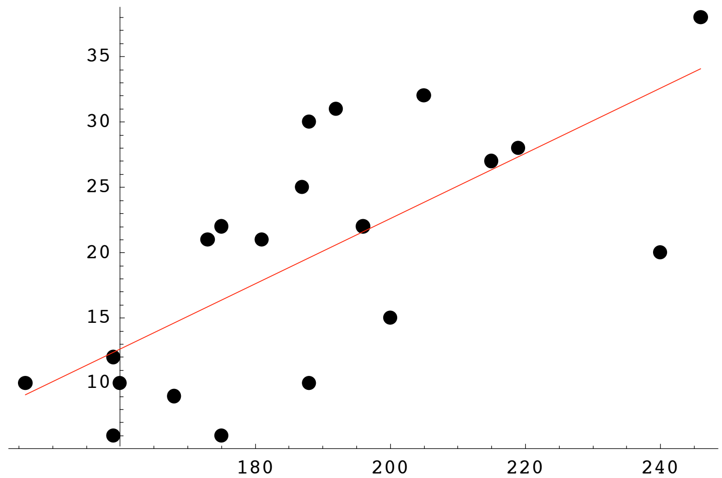
regrstats[x_, y_] := $\left\{\text{mean}[x], \text{mean}[y], s[x], s[y], r[x, y], \frac{r[x,y]\, s[y]}{s[x]}\right\}$ 1.

```
regrstats[wt2, bf2]
```

```
{188.6, 19.75, 26.6624, 9.5635, 0.696633, 0.249874}
```

```
regrplot[wt2, bf2]
```






```
- Graphics -
```

```
?mean
```

```
Global`mean

mean[list_] := Mean[list] 1.
```

■ Exercise. 8-45

```
hj = {1.6, 1.75, 1.63, 1.72, 1.81, 1.78, 1.69, 1.66, 1.78, 1.81, 1.72, 1.75, 1.72,
  1.75, 1.66, 1.78, 1.84, 1.72, 1.81, 1.66, 1.84, 1.84, 1.78, 1.72, 1.72, 1.78}
```

```
{1.6, 1.75, 1.63, 1.72, 1.81, 1.78, 1.69, 1.66, 1.78, 1.81, 1.72, 1.75, 1.72,
 1.75, 1.66, 1.78, 1.84, 1.72, 1.81, 1.66, 1.84, 1.84, 1.78, 1.72, 1.72, 1.78}
```

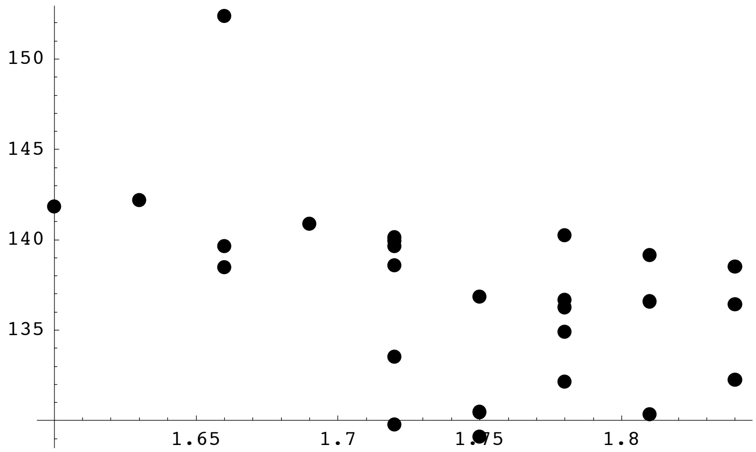
```
m800 = {141.82, 129.08, 142.17, 133.52, 139.14, 136.25, 140.86, 139.64,
  134.9, 136.57, 140.11, 130.45, 129.77, 136.83, 152.36, 136.67, 138.49,
  139.94, 130.32, 138.47, 132.24, 136.41, 140.25, 138.56, 139.65, 132.15}
```

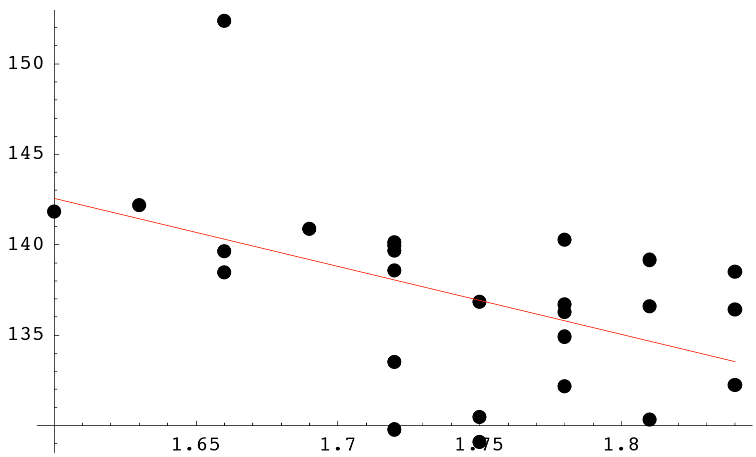
```
{141.82, 129.08, 142.17, 133.52, 139.14, 136.25, 140.86, 139.64,
 134.9, 136.57, 140.11, 130.45, 129.77, 136.83, 152.36, 136.67, 138.49,
 139.94, 130.32, 138.47, 132.24, 136.41, 140.25, 138.56, 139.65, 132.15}
```

```
regrstats[hj, m800]
```

```
{1.74308, 137.178, 0.065895, 4.98274, -0.49593, -37.5004}
```

```
regrplot[hj, m800]
```





- Graphics -