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

- navigating to www.stt.msu.edu/~lepage
- clicking on the (folder) STT200
- clicking on the (program) stat200 2-10-09 (stat200 2-10-09 will launch)
- clicking on the 1+1 line just below
- performing SHIFT+ENTER.
- responding YES to the pop-up (evaluates cells).

1 + 1
A partial list of the commands in this notebook and what they do.

- **median**\([\{4.5, 7.1, 7.8, 9.1\}]\) returns the median of \(\{4.5, 7.1, 7.8, 9.1\}\)
- **qtile**\([\{4.5, 7.1, 7.8, 9.1\}, 0.7]\) returns the 70\(^{th}\) percentile of \(\{4.5, 7.1, 7.8, 9.1\}\)
- **iqr**\([\{4.5, 7.1, 7.8, 9.1\}]\) returns the inter-quartile range of \(\{4.5, 7.1, 7.8, 9.1\}\)
- **boxplot**\([\{4.5, 7.1, 7.8, 9.1\}]\) returns a boxplot of \(\{4.5, 7.1, 7.8, 9.1\}\) except without whiskers but with fences.
- **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**\([\{4.5, 7.1, 7.8, 9.1\}, \{2, 5\}]\) 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 \(\hat{\sigma} = 1.67686\)

\[ r[x, y] \] returns the sample correlation \(r = \frac{\bar{x}\bar{y} - \bar{xy}}{\sqrt{s_x^2 - \bar{x}^2} \sqrt{s_y^2 - \bar{y}^2}} \) for paired data.

\[ \text{regrtable}[x,y] \] returns a table illustrating calculations of \(\bar{x}, \bar{y}, s_x, s_y, r, \) and the slope of the regression line \(r = \frac{s_y}{s_x} = r = \frac{\hat{c}}{\hat{a}}\).

\[ \text{regrstats}[x,y] \] returns \(\bar{x}, \bar{y}, s_x, s_y, r, \) and the slope of the regression line \(r = \frac{s_y}{s_x} = r = \frac{\hat{c}}{\hat{a}}\).

\[ \text{regrplot}[x,y] \] returns the plot of \((x, y)\) pairs overlaid with the regression line.

\[ \text{sample}[\{4.5, 7.1, 7.8, 9.1\}, 10] \] returns 10 samples from \(\{4.5, 7.1, 7.8, 9.1\}\)

\[ \text{ci}[\{4.5, 7.1, 7.8, 9.1\}, 1.96] \] returns a 1.96 coefficient CI for the mean from given data

\[ \text{bootci}[\text{mean}, \{4.5, 7.1, 7.8, 9.1\}, 10000, 0.95] \] returns 0.95 bootstrap ci for pop mean

\[ \text{smooth}[\{4.5, 7.1, 7.8, 9.1\}, 0.2] \] returns the density for data at bandwidth 0.2

\[ \text{smooth2}[\{4.5, 7.1, 7.8, 9.1\}, 0.2] \] returns the density for data at bandwidth 0.2

\[ \text{smoothdistribution}[\{\{1, 700\}, \{4, 300\}\}, 0.2] \] returns the density at bandwidth 0.2

\[ \text{popSALES} \] is a file of 4000 sales-amounts (used for examples)

Entering \[\text{popSALES} \] will spill 4000 numbers onto the screen.

To prevent that enter \[\text{popSALES;} \] (append semi-colon to suppress output).

\[ \text{betahat0}[\text{list x, list y}] \] returns the least squares intercept and slope for a straight line fit of the model \(y = b_0 + b_1 x + \epsilon\).

\[ \text{betahat}[\text{matrix x, list y}] \] returns the least squares coefficients \(\hat{\beta}\) for a fit of the matrix model \(y = x\beta + \epsilon\).

\[ \text{resid0}[\text{list x, list y}] \] returns the estimated errors \(\hat{\epsilon} = y - x\hat{\beta}\) (see \text{betahat0} above).

\[ \text{resid}[\text{matrix x, list y}] \] returns the estimated errors \(\hat{\epsilon} = y - x\hat{\beta}\) (see \text{betahat} above).

\[ \text{R}[\text{matrix x, list y}] \] returns the multiple correlation \(R\) between the fitted values \(x\hat{\beta}\) and data \(y\) in the matrix model setup.

\[ \text{xquad}[\text{matrix x}] \] returns the full quadratic extension of a design matrix with constant term

\[ \text{xcross}[\text{matrix x}] \] returns the extension of \(x\) to include all products of differing columns.

\[ \text{betahatCOV}[\text{x matrix, list y}] \] returns the estimated covariance matrix of the vector \(\hat{\beta}\) in the matrix model setup.

\[ \text{normalprobabilitplot}[\text{list, 0.02}] \] returns the normal probability plot of the list using plotting-dot size 0.02.

By clicking on any of the examples below you can execute it afresh by performing \text{SHIFT+ENTER}. Or click anywhere between lines, or at the end of the file, to make a fresh line and type your own examples.
Defining data to Mathematica, Exercise 33, page 190.

\[
\]

\[
\text{Length}[\text{calories}] = 20
\]

\[
time = \{21.4, 30.8, 37.7, 33.5, 32.8, 39.5, 22.8, 34.1, 33.9, 43.8, 42.4, 43.1, 29.2, 31.3, 28.6, 32.9, 30.6, 35.1, 33.0, 43.7\}
\]

\[
\text{Length}[\text{time}] = 20
\]

* Plotting \((x, y)\) data and the regression line for data already defined to Mathematica.

\[
\text{regrplot[time, calories]} \quad \text{(_plot)}
\]

* Calculating regression statistics \(\bar{x}, \bar{y}, S_x, S_y, r, r \frac{s_y}{s_x}\).

\[
\text{regrstats[time, calories]} \quad \text{(_regression statistics)}
\]

\[
\{34.01, 456., 6.31647, 29.9403, -0.649167, -3.07707\}
\]

* Calculating the intercept \(\bar{y} - \bar{x} r \frac{s_y}{s_x}\) of the regression line (it is off the plot).

\[
456 - 34.01 (-3.07707) = 560.651
\]
Calculating the fraction of $\sigma_y^2$ explained by regression of $y = \text{calories}$ on $x = \text{time}$.

```
r[time, calories]^2
0.421417
```

Normal probability plot of residuals with plotting-dot size 0.02.

```
normalprobabilityplot[resid0[time, calories], 0.02]
```
Table illustrating calculations of $x, \ y, \ x^2, \ y^2, \ xy$.

In[658]= regtable[time, calories]

Out[658]//MatrixForm=

\[
\begin{array}{cccc}
 x & y & x^2 & y^2 & xy \\
 21.4 & 472 & 57.96 & 227784 & 10100.8 \\
 30.8 & 498 & 948.64 & 248004 & 15338.4 \\
 37.7 & 465 & 1421.29 & 216225 & 17530.5 \\
 33.5 & 456 & 1122.25 & 207936 & 15276. \\
 32.8 & 423 & 1075.84 & 178929 & 13874.4 \\
 39.5 & 437 & 1560.25 & 190969 & 17261.5 \\
 22.8 & 508 & 519.84 & 258064 & 11582.4 \\
 34.1 & 431 & 1162.81 & 185761 & 14697.1 \\
 33.9 & 479 & 1149.21 & 229441 & 16238.1 \\
 43.8 & 454 & 1918.44 & 206116 & 19885.2 \\
 42.4 & 450 & 1797.76 & 202500 & 19080. \\
 43.1 & 410 & 1857.61 & 168100 & 17671. \\
 29.2 & 504 & 852.64 & 254016 & 14716.8 \\
 31.3 & 437 & 979.69 & 190969 & 13678.1 \\
 28.6 & 489 & 817.96 & 239121 & 13985.4 \\
 32.9 & 436 & 1082.41 & 190096 & 14344.4 \\
 30.6 & 480 & 936.36 & 230400 & 14688. \\
 35.1 & 439 & 1232.01 & 192721 & 15408.9 \\
 33. & 444 & 1089. & 197136 & 14652. \\
 43.7 & 408 & 1909.69 & 166464 & 17829.6 \\
 34.01 & 456. & 1194.58 & 208788. & 15391.9 \\
\end{array}
\]

Computing correlation \( r = \frac{\bar{xy} - \bar{x} \bar{y}}{\sqrt{\frac{x^2 - \bar{x}^2}{x^2 - \bar{x}^2}} \sqrt{\frac{y^2 - \bar{y}^2}{y^2 - \bar{y}^2}}} \) (slight errors due to rounding in the above).

In[659]=

\[ \frac{15391.9 - 34.01 \times 456}{\sqrt{1194.58 - 34.01^2} \sqrt{208788 - 456^2}} \]

Out[659]= -0.649207