HW 6 Due in recitation 2-24-09

Go to a computer lab before recitation. Launch stat200 2-10-09 found on www.stt.msu.edu/~lepage (be sure to launch the 2-10-09 edition near the end of the file list). *Mathematica* will launch. Follow the instructions on Lecture Outline 2-20-09 and do the following:

1. Enter the following matrix to *Mathematica*:
myx= {{1, 2.3, 3.6}, {1, 2.4, 3.5}, {1, 2.0, 3.1}, {1, 2.4, 3.7}, {1, 2.5, 3.6}}

```
In[51]:= myx = {{1, 2.3, 3.6},
    {1, 2.4, 3.5}, {1, 2.0, 3.1},
    {1, 2.4, 3.7}, {1, 2.5, 3.6}}

Out[51]= {{1, 2.3, 3.6},
   {1, 2.4, 3.5}, {1, 2., 3.1},
   {1, 2.4, 3.7}, {1, 2.5, 3.6}}
```

2. Enter y = the last five digits of your student number. For example, if your student number ends in 47680 you enter:
myy = {4, 7, 6, 8, 0}

```
In[52]:= myy = {4, 7, 6, 8, 0}

Out[52]= {4, 7, 6, 8, 0}
```

3. Compute the coefficients $\hat{b}_0, \hat{b}_1, \hat{b}_2$ of a least squares fit of the model y = $b_0, + b_1\, x_1 + b_2\, x_2$ for the n = 5 data values.

In[53]:=
```
mybetahats = betahat[myx, myy]
```
Out[53]=
```
{7.73563, -16.092, 9.88506}
```

4. Compute the multiple correlation R.

In[54]:=
```
R[myx, myy]
```
Out[54]=
```
0.472217
```

5. Determine the fraction of s_y^2 explained by regression on the columns of myx.

In[55]:=
```
0.472217^2
```
Out[55]=
```
0.222989
```

6. Determine a 95% CI for b_1 that would apply if n were large (here it is only 5) and specified assumptions on the "errors in regression" were satisfied.

In[57]:=
```
MatrixForm[betahatCOV[myx, myy]]
```
Out[57]//MatrixForm=

$$\begin{pmatrix} 893.385 & 110.746 & -327.774 \\ 110.746 & 491.214 & -357.247 \\ -327.774 & -357.247 & 330.453 \end{pmatrix}$$

In[58]:=
```
-16.092 + {-1, 1} 1.96 Sqrt[491.214]
```
Out[58]=
```
{-59.5322, 27.3482}
```

The role of n = 5 is concealed in the above calculation of CI. Had n been large we'd have seen a narrower (more informative) CI.

7. Calculate the predicted value $\hat{y}$ for independent variable values

$$\{1, 2.4, 3.0\}.$$

It is the value 1 $\hat{b}_0$ + 2.4 $\hat{b}_1$+ 3.0 $\hat{b}_2$ and is simply calculated using the "dot product" below.

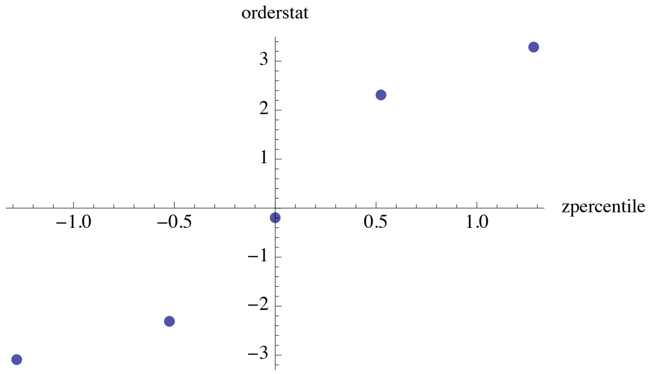
In[60]:=
```
{1, 2.4, 3.0}.mybetahats
```

Out[60]=
```
-1.22989
```

8. The residuals are the vertical discrepancies between the points of the plot and the regression surface, written y - $\hat{y}$. A normal probability plot of these residuals will give us an idea as to whether the y-scores appear to have been tossed from the underlying model by means of independent normal random errors. Look to see if the plot is roughly a straight line (of course this is only a toy example; n is woefully small).

In[111]:= `normalprobabilityplot[resid[myx, myy], .02]`

Out[111]=



Submit a one page printout of your results (as above) in recitation.

Here are some **additional exercises** for you to work through. You will be quizzed on these ideas in recitation but **don't hand them in**. I can respond to questions in lecture.

1-4. A model for the strength of a concrete mixture is

strength = b_0 + b_1 agg + b_2 add + b_3 temp + b_4 cure

where

agg is a measure of aggregate in the mix
add is the amount of an additive to the mix
temp is a measure of the temperature during curing

cure is the time allowed to cure before strength testing

1. What is the dependent variable? List the independent variables (including constant term).

2. The coefficients obtained from least squares (i.e. regression) are $\hat{b}_0 = 28.2$, $\hat{b}_1 = 1.22$, $\hat{b}_2 = 2.31$, $\hat{b}_3 = 0.26$, $\hat{b}_4 = 0.36$. Determine the estimated strength for a mix

agg = .3 add = 6.3 temp = 47 cure = 12.

3. For R = 0.8 give the fraction of s_y^2 explained by regression on the independent variables.

4. Suppose $s_y = 34$ and R = 0.8. For an elliptical plot, give the distribution of the y values in the vertical cylinder (not strip) for

agg = .3 add = 6.3 temp = 47 cure = 12.

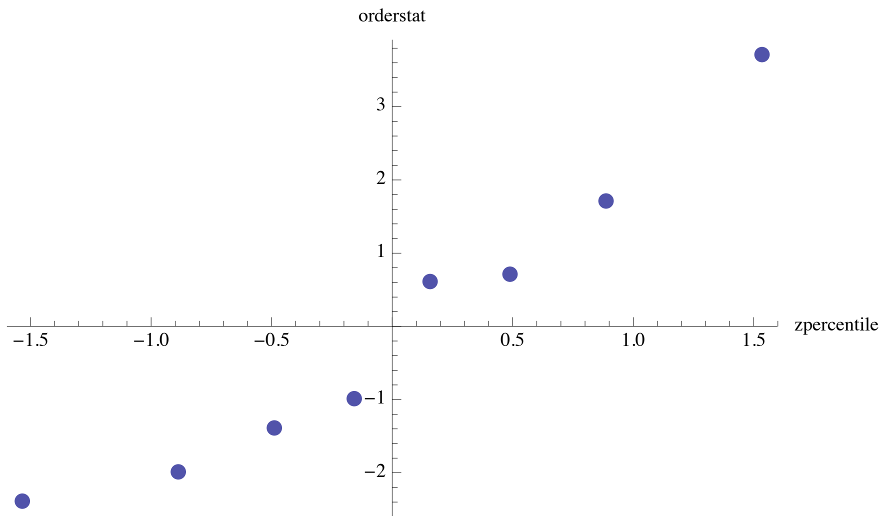
Give the mean, sd, and form of the distribution.

5. Suppose the residuals are

{3.7125, 1.7125, 0.7125, -1.3875, -2.3875, -0.9875, -1.9875, 0.6125}.

Here is a normal probability plot of these residuals (required computer).

```
normalprobabilityplot[{3.7125, 1.7125, 0.7125,
  -1.3875, -2.3875, -0.9875, -1.9875, 0.6125}, .02]
```



The lowest value seems to be pulled back a bit towards the center (short left tail) but it could easily be spurious since only a single residual is doing it.

6. Suppose the diagonal entry of betahatCOV for "cure" is 78.79. For large n, if the residuals plot looks like a straight line, we might offer a 95% CI for the coefficient of cure. Give this CI.