Sample Questions for Quiz I
ST370 - Probability and Statistics for Engineers
The ACTUAL quiz will consists of less number of problems.

1. A small data set consists of the numbers 34, 27, 26, 42, 38, 40, 31, 30, 35, 38. Compute the a) the mean, b) the median and c) IQR.

2. A statistics instructor informed his class that the mean, median and s.d. of exam score were 74, 82 and 9.45 respectively.
   a) Sketch a picture of what the histogram (or frequency polygon) of scores might look like (based on the above summary values).
   b) How would you interpret your histogram? Choose one of the following:
      (i) Majority of students in the class studied for the exam, but few students had not.
      (ii) Majority of students in the class didn’t study for the exam, but for few students.
      (iii) None of the above.
   c) Suppose the class consisted of 25 students and the instructor forgot to grade one paper (e.g. above mean is actually based on 24 graded papers). How high would the score on the last paper have to be to raise the class average by one point? By two points?

3. Given are the percent NaOH used as a pretreatment chemical, $X_1$, the pretreatment time in minutes, $X_2$, and the resulting value of a specific surface area variable, $Y$ (with units of cm$^3$/h), for 9 batches of pulp produced from a mixture of hardwoods at a treatment temperature of 75° Celsius in mechanical pulping.

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>3.0</th>
<th>3.0</th>
<th>3.0</th>
<th>9.0</th>
<th>9.0</th>
<th>9.0</th>
<th>15.0</th>
<th>15.0</th>
<th>15.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$</td>
<td>30</td>
<td>60</td>
<td>90</td>
<td>30</td>
<td>60</td>
<td>90</td>
<td>30</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>$Y$</td>
<td>5.95</td>
<td>5.60</td>
<td>5.44</td>
<td>6.22</td>
<td>5.85</td>
<td>5.61</td>
<td>8.36</td>
<td>7.30</td>
<td>6.43</td>
</tr>
</tbody>
</table>

   a) The primary response variable is :
   b) Name TWO experimental factors and give their levels :
   c) What could be used as a blocking variable in this experiment?
   d) What is an obvious weakness in above data collection?
   e) The fitted equation is given by $\hat{Y} = 4.98 + 0.14X_1 - 0.02X_2$ ($R^2 = 0.8065$). How would you interpret the regression coefficients of $X_1$ and $X_2$ in the above equation?
4. $Y = \text{Drying time (in hours)}$ and $X = \text{solvent content (in \%)}$ of sample batches of an experimental coating substance are plotted in above figure. The summary statistics for the data are as follows:

$\bar{X} = 3.41$, $\bar{Y} = 1.92$, $s_X = 0.6505$, $s_Y = 0.2741$ and $r = -0.9734$.

a) Compute the slope and intercept of the predicted line $\hat{Y} = a + bX$ along with coefficient of determination, $R^2$.

b) Interpret the values of $b$ and $R^2$ obtained in a).

c) Interpret the above residual plots. Discuss what they suggest about the appropriateness of the fitted line.

d) Based on the fitted line, what drying time would you predict for an additional solvent content of 3.8\%? Can you predict the drying time for an additional solvent content of 5.7\%?

5. Give an example in each case to distinguish the following:

(a) Population and sample.

(b) Quantitative and qualitative data.
6. Suppose that a complete factorial treatment structure is to be used to study how a variable $Y$ is affected by variables $X_1$, $X_2$ and $X_3$. Given that $X_1$ has 3 levels, $X_2$ has 2 levels and $X_3$ has 2 levels.

a) $Y$ is called a __________ variable. (fill in)

b) How many total treatment combinations are there?

c) If the answer to b) is $k$, and the experimenter decides to use a randomized complete block with 4 blocks, how many total experimental runs will need to be made? (in terms of $k$ or use the number from b))

d) Why would anyone want to use blocking?

7. Suggested problems from the text:

1-5, 1-14, 2-19, 2-29, 2-31, 2-40, 4-1, 4-11 and 4-21.

*Solutions to above problems will NOT be posted on the web. However, I'll solve the problems in class.*