Exam 1 is Wed. Sept. 18. **No notes. You are allowed a TI-30 calculator.** The exam covers ch. 1,2,3,4 and part of 5. The pull out section of the text, colored boxes and bold face text have important ideas. Also see summary ch. 7 on p. 179-181. Some of the types of problems that appear on Exam 1 are listed below.

The following problem is very important for both the midterm and the final.

1) Given a list of n numbers, find the mean $\overline{x}$, median, and standard deviation SD $= S$. Recall that $\overline{x} = \sum x/n$, $S = \sqrt{\sum(x-\overline{x})^2/(n-1)}$ and the median is found by finding the middle number(s) of the data ordered from smallest to largest. See Q1 2, Q2 3,4, Q3 and HW2 A).

2) You should know how to make a bar graph for categorical data. Bar graph bars should have bases that have the **same length**.

Do not make breaks in the vertical axis of bar graphs and histograms because the area of the bars is proportional to the percentage of cases in each class.

3) Given a list of numbers, make a distribution table and a histogram. Given a histogram and a rule for the endpoints (eg bar includes right endpoint but not the left endpoint), you should be able to make a distribution table. See Q1 3 and HW1 C).

4) Given a list of numbers, make a stemplot. **Include the stem units and leaf units on the plot.** For example the number 205 will have stem 20 and leaf 5 with stem units = tens and leaf units = ones. Note that $20(10) + 1(5) = 205$. See Q1 2 and HW1 B).

5) Sometimes a list of numbers is presented as a stemplot, then you are asked to find the mean, median, and SD of the list. The stem and leaf units are used to determine what the list of numbers is.

6) Given a list of numbers or Minitab output, find the 5 number summary: min, Q1, median, Q3, and max. Recall that the data is **sorted from smallest to largest.** The median is the "middle number", Q1 is the median of the sorted numbers to the left of the median, and Q3 is the median of the numbers to the right of the median. Use the 5 number summary to make a boxplot. See Q1 1, Q2 1, and HW1 D),E).

7) Given a boxplot, bar graph, stemplot, histogram or timeplot, be able to give a short summary of what the plot tells you. For example are outliers present, is the histogram symmetric, right skewed or left skewed? how does the proportion of one category compare to the proportion of another category? See HW1 B), C).

8) From a story problem, you should be able to determine the individuals and the variables. You should know whether the variable is categorical or discrete. See HW1 A).

The following problem is very important for both the midterm and the final.

9) Know how to do a **forwards calculation using table A.** In the story problem you will be told that X is approximately normal with some mean and SD. You will be

given one or two X^* values and asked **to find a proportion** or probability or chance. Draw a line and mark down the mean and the X^* values. Standardize each X value by taking the z-score $Z^* = (X^* - \mu)/\sigma$. If you want the chance that X is **less than** X^*, then table A gives the correct value. If you want the chance that X is **greater than** X^* take 1 - table A value. If you want the chance that X **is between** X_1^* and X_2^* subtract the smaller value from the larger value of table A. Given a z-score, to use table A you use the leftmost column and top row of table A. Intersect this row and column to get a 4 digit decimal which is equal to the area to the left of Z^*. See Q2 5, Q3, and HW2 C), E).

The following problem is very important for both the midterm and the final.

10) Know how to do a **backwards calculation using table A.** Here you are **given a proportion and asked to find** one or two X^* values. Table A gives areas to the **left** of Z^*. So if you are asked to find the top 5%, that is the same as finding the bottom 95%. If you are asked to find the bottom 25%, table A gives the correct value. If you are asked to find the two values containing the middle 95%, then 5% of the area is outside of the middle. Hence .025 area is to the left of $X^*(lo)$ and $.025 + .95 = .975$ area is to the left of $X^*(hi)$. Once you know the area to the left of X^*, find the largest 4 digit number smaller than the desired area and the smallest 4 digit number larger than the desired area. These two numbers will be found in the middle of table A. Take the number closest to the desired area, and to find the corresponding Z^*, examine the row and column containing the number. Go along the row to the entry in the leftmost column of table A and go along the column to the top row of table A. For example, if your 4 digit number is .9750, $Z^* = 1.96$. To get the corresponding X^*, use $X^* = \mu + \sigma Z^*$. See Q3, HW2 D), F).

11) Given a density that is box shaped with base from a to b, know that the height of the density is $1/(b-a)$ and that the chance that X is between c and d where $a \leq c < d \leq b$ is given by (base)(height) $= (d-c)/(b-a)$. See HW2 B).

12) Given the table used to compute the correlation with a few missing entries, be able to compute the missing entries and then get $r = \sum z_x z_y/(n-1)$. See Q3 1 and HW3 A). Here $z_x = (x - \overline{x})/s_x$ and $z_y = (y - \overline{y})/s_y$ where x is the x value, $\overline{x}$ is the sample mean of the $x's$ and s_x is the sample standard deviation of the $x's$.

The following three problems are very important for both the midterm and the final.

13) Be able to find the least squares line $\hat{y} = a + bx$ from Minitab output (a is under *Coef* and to the right of *Constant* while b is below a). See HW3 H).

14) Be able to find the least squares line $\hat{y} = a + bx$ given 2 means, 2 SD's and the correlation r. Recall that $b = rs_y/s_x$ and $a = \bar{y} - b\bar{x}$. Remember that **the response** y is what you want to predict. The explanatory variable is x. See HW3 E).

15) Given the line $\hat{y} = a + bx$, be able to predict y for a given value of x. See HW3 E)b, H).