1. Phone surveys are sometimes used to rate TV shows. Such a survey records several variables listed below. Which ones of them are categorical and which are quantitative?
- the number of people watching the show – **quant.**
- the ages of people watching the show – **quant.**
– the name of the show being watched – **categorical.**
– whether they watch the show on cable, or on satellite TV or neither – **categorical.**

2. Below are the team-by-team statistics on home runs in early part of 1994 season. They are given as stemplots separately for the American and the National League teams. American league: National League:

a) Find the five point summary for the home runs of the National League teams.

29 46 50.5 55 67

b) What is the mean number of home runs by the American League teams?

56.93

c) What is the standard deviation of home runs by the American League teams?

12.69

d) For each of the following statements, decide whether it is correct or not (yes/no), by looking at the shape of the stemplots:
- The American League plot is reasonably symmetric. **- yes**
- The National league plot is skewed to the right **- no**
- The median number of home runs hit by American league teams was higher than by National League teams **- yes**
- The median number of home runs for the National League is higher than the mean number of home runs **- yes**
e) Make side-by-side boxplots for the two leagues. Compare the boxplots.

3. The distribution of actual weight of 8.0 oz. chocolate bars produced by a certain machine is Normal with a mean of 8.1 oz. and a standard deviation of 0.1 oz.

a) Find the proportion of chocolate bars weighing under 8.0 oz.

variable X=weight of chocolate bar
X is N(8.1, 0.1)
proportion of bars under 8 Oz is the same as P(X< 8) = (standardize) P(z < (8-8.1)/0.1)=
P(z < -1) = 0.1587

b) What is the proportion of chocolate bars weighing between 8.2 and 8.3 oz.

It is equal to P(8.2 <X < 8.3) =P((8.2-8.1)/0.1 < z < (8.3-8.1)/0.1) = P (1 < z < 2)=
0.9772-0.8413 = 0.1359

c) What weight should be put on the chocolate bar wrappers so that only 1% of bars are
   underweight?

Need to find L such that P(X < L)=0.01. First find z such that P(z < L) =0.01 by using
table A backward: z= -2.33. Unstandardize: X=z*0.1 + 8.1 = 7.867 Oz

4. A student wonders if people of similar heights tend to date each other. Here are the
data that she collected (heights in inches):
Women 66 64 66 65 70 65
Men   72 68 70 68 74 69
a) Find correlation between women's and men's heights

0.91

b) Indicate whether the following statements are correct (yes/no):
–   There is a positive association between the heights of men and women **- yes**
- There is a positive correlation between the heights of men and women, because women
are always smaller than the men they date **- No**
- A positive correlation between the heights of men and women means that there would
be no outliers in a scatterplot of these data **- No**
–   A positive correlation between the heights of men and women means that a scatterplot
   of these data would lie along a horizontal line – **No**

c) Find the equation of the least-squares regression line for these data, using women's height as an explanatory variable.

d) Make a scatterplot of the data

e) Add the regression line to the scatterplot

f) According to your analysis, if a woman's height is 64 inches, what would you predict the height of the man she dates to be?

g) Make a plot of the residuals for this data. Does it indicate that there are outliers in the data? Why? Does it indicate that a straight line is a good summary for the data? Why?

5. All human blood can be of one of four types, called O, A, B, or AB. The distribution of types varies a bit with race. Choose a black American at random. Here are the approximate probabilities that the person you choose will have blood type O, B, or AB.

Blood type O A B AB
Probability .50 ? .20 .05

A) What is the probability that the chosen person has blood type A?

0.25

B) What is the probability that the chosen person has a blood type other than O?

0.5

C) What is the probability that the chosen person has blood type that is either B or AB?

0.25

6. A car insurer has found that repair claims have a mean of $920 and a standard deviation of $870. Suppose that the next 100 claims can be regarded as a random sample from the long-run claims process.

a) What is the distribution of the average of the next 100 claims?

N(920,87)

b) What is the probability that the average of the 100 claims is larger than $1000?

P(xbar > 1000) = P (z > (1000-920)/87) = P(z>0.92) = 1-0.8212=0.1788

c) If the average of the next 100 claims turns out to be $1020, is this a strong evidence that the actual mean of repair claims is higher than $920?

Ho: $\mu=920$
Ha: $\mu>920$
$z=(1020-920)/87=1.15$
$P=1-0.8749 = 0.1251$
P-large, i.e. cannot reject Ho. No it is not strong evidence that the actual mean is higher thn 920.

7. In a certain town 60% of the households own mutual funds, 40% own individual stocks, and 20% own both mutual funds and individual stocks. (You may want to use a Venn diagram to help you with this problem, but it is not required)
a) Find the proportion of households that own mutual funds, but not individual stocks.

40%

b) What is the proportion of households that own neither mutual funds nor the individual stocks?

20%

c) Find the proportion of households that own either mutual funds but no individual stocks or individual stocks but no mutual funds.

60%

d) A person from this town is selected at random. What is the probability that that person owns both mutual funds and individual stocks?

.2

8. The water diet requires you to drink two cups of water every half-hour, but otherwise allows you to eat anything. Four adult volunteers agree to test the diet. They are weighed before they begin the diet and after six weeks on the diet. The weights (in pounds) are

Person	1	2	3	4
Weight before the diet	180	125	240	150
Weight after six weeks	170	130	215	152

For the population of all adults assume that the weight loss after six weeks on the diet (weight before minus weight after) is Normally distributed with mean .

a) Is this a one-sample, matched-pairs or two-sample design?

matched pairs

b) Does these data indicate that the diet leads to weight loss (use 99% level of significance)?

before - after =10   -5   25   -2
xbar=7, s=13.64
Ho: $\mu=0$
Ha: $\mu>0$
z=(7-0)/(13.64/2)=1.03
P=1-.8485=.1515
P-large (P>0.01), cannot reject Ho. No, there is no evidence of significant weight loss.

c) Give a 98% confidence interval for μ
ci)
7±2.326*(13.64/2)
The average weight difference (before minus after) is between -8.86 and 22.86 pounds, with 98% confidence