**FRAMINGHAM DIDACTIC DATA SET**

**Resource 01 - Solutions   
*8-3-2015***

1. Briefly describe the study, including the type (observational, experimental), outcome variable(s), exposure(s) or treatment(s) of interest and any obvious advantages/disadvantages of the design.

**This is an observational study because we are only observing cardiovascular risk factors, not randomizing subjects to risk factors (quite unethical). The outcome variable of interest is whether or not the patient develops CHD. Here there is not one primary exposure of interest, but rather this is exploratory and they are looking to identify any risk factors. An advantage of this design is the large sample size, the sense of tradition (since the participant’s parents and grandparents were involved in the study) and the developing trust with these researchers. A disadvantage of this design is the large number of potential confounders and the data is only collected every two years.**

2. Briefly describe the characteristics of the dataset, including sample size and the type of variables (e.g. categorical, nominal, ordinal, numeric, discrete, and continuous).

**This dataset consists of 4699 patients who were free of coronary heart disease at their baseline exam. The variables collected are:**

**gender (categorical - nominal)**

**systolic blood pressure (numeric- continuous)**

**diastolic blood pressure (numeric - continuous)**

**serum cholesterol (numeric - continuous)**

**age (numeric - discrete)**

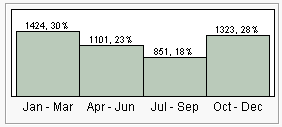
**body mass index (numeric - continuous)**

**month exam occurred (categorical - nominal)**

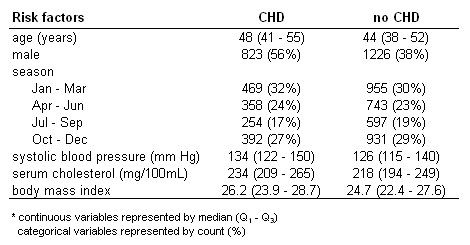
**cardiovascular heart disease fate (categorical - nominal)**

**The important thing is that they get categorical vs. numeric correct and that they get nominal vs. ordinal correct. Do not fuss over discrete vs. continuous since these are often not well defined.**

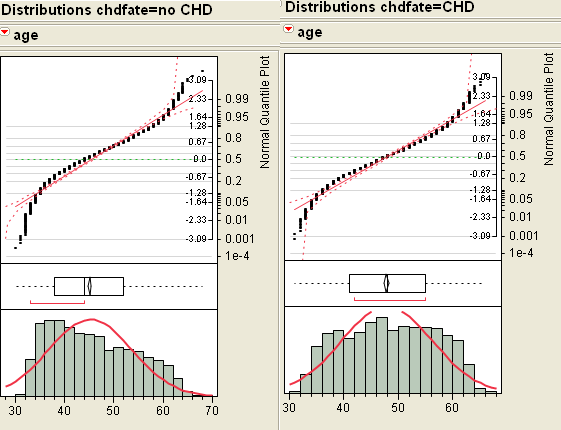
3. To capture the seasonal effect, create a new variable season where 1 = baseline exam occurred in January – March, 2 = April – June, 3 = July – September, 4 = October – December. Be prepared to show how you did this (screenshot of formula is fine) and the resulting distribution of your new variable in your presentation.

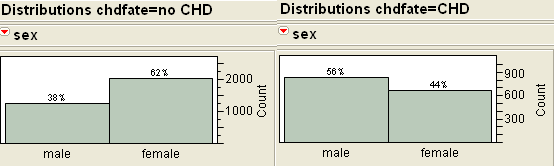


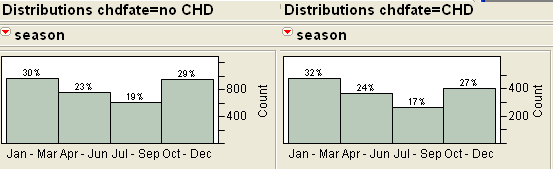
4. The researchers were interested in comparing those who develop CHD with those who do not with respect to potential risk factors. Create a “Table 1” demonstrating this comparison with respect to age, gender, season, systolic blood pressure, serum cholesterol and body mass index. Be sure to use appropriate descriptive statistics and be ready to defend your selections.

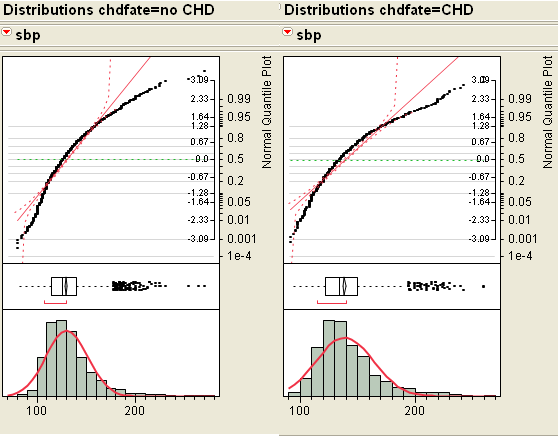


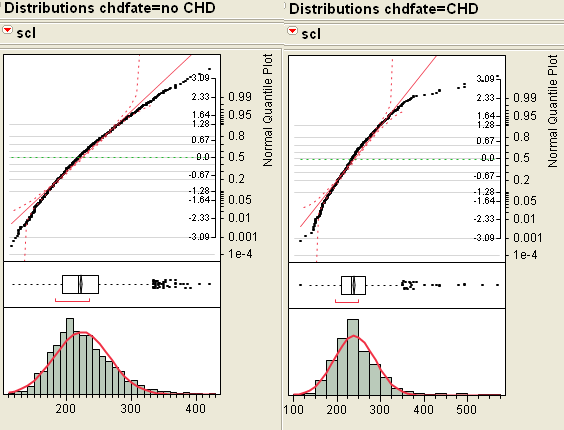
**Evidence of distributions:**

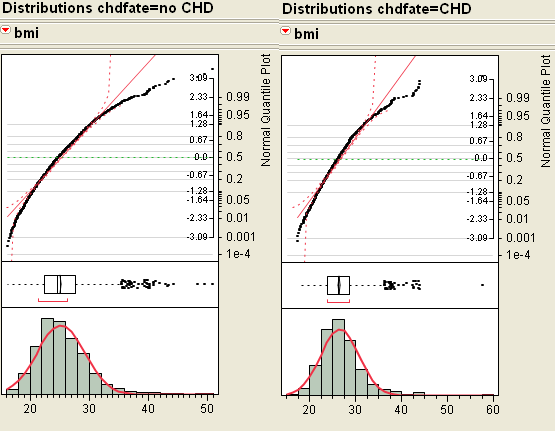




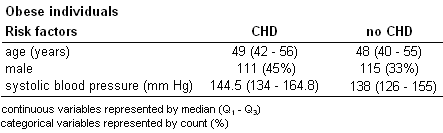




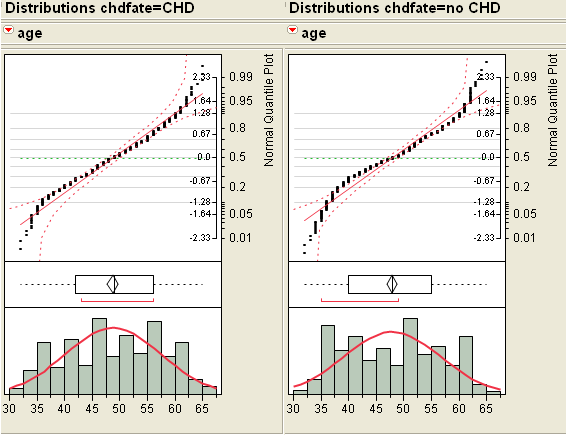


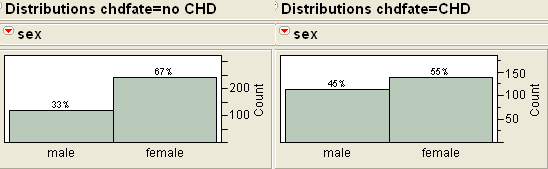


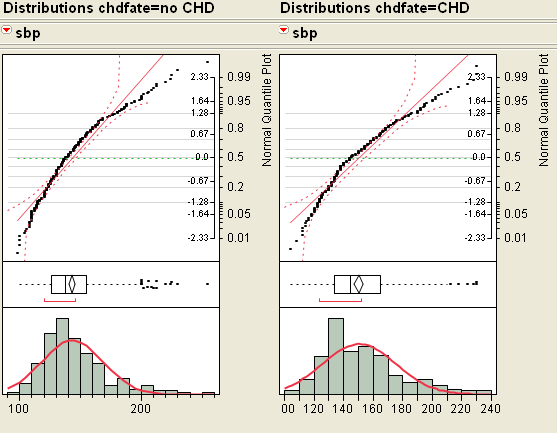
5. There is specific interest in congestive heart disease (CHD) and the obese. An adult who has a BMI of 30 or higher is considered obese. For the subset of obese participants, create a “Table 2” comparing those who develop CHD with those who do not with respect to age, gender and sbp.



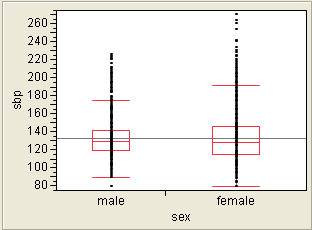
**Evidence of distributions:**



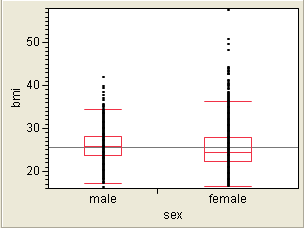




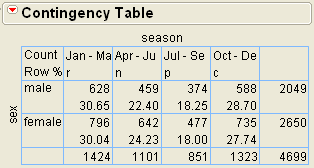
6. Use graphs to explore the relationship between gender and the following: systolic blood pressure, body mass index and season. Interpret the findings.



**It appears that the sbp distribution for males and females are about the same with a slightly larger variance in the female distribution.**

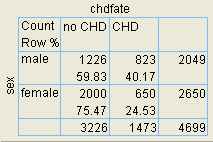
****

**It appears that on average males have a higher bmi, but females values are more spread out.**



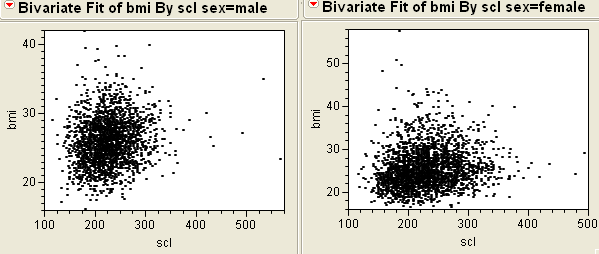
**Season does not appear to be associated with gender.**

7. Claims have been made that CHD is more common in males than females. Use graphs to investigate this claim. What do you conclude?



**This data seems to support that claim that chd is associated with gender (males being more likely).**

8. Use graphs to explore the relationship between serum cholesterol and body mass index by gender. Does the association look the same for males and females? Interpret the findings.



**It appears that as serum cholesterol increases, body mass index also increases, slightly more for males.**

**Ask students if these scatterplots are very informative. Why?**

**What could be done to enhance these plots and make them more informative? (correlation ellipse or regression line… we’ll talk about each later in the course).**

9. Describe the population to which these results can be generalized.

**These results can be generalized to men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts (some could argue that Framingham is representative of a larger geographic region).**