

Assignment #1

Due: October 8th at 12pm in the box marked stat 333 inside room MS 315.

Assignments will not be accepted if they are passed in after this time.

Assignments must be understandable (logically correct as well as legible).

Marks will be deducted if the assignment does not have a cover page with your name on it and if it is not stapled.!

- A sample of 51 individuals are selected for participation in a study of cardiovascular risk factors. The following data represent the ages of enrolled individuals measured in years. Here age is measured in the usual way with a person being recorded as 65, for example, until the day he/she turns 66. The data are as follows:

60 66 71 76 77 72 66 62 87 75 70 66 65 70 75 85 66 76 71 77 72
66 63 64 67 73 77 85 75 70 65 77 73 67 64 83 75 68 65 65 68 73
82 79 77 73 73 68 67 65 65

 - Construct a stem –and –leaf diagram of the data. You may want to do increments of 5 instead of 10.
 - Does the graphic in (A) support the claim that the data came from a reasonably bell shaped distribution (normal distribution)?
 - Use the stem-and-leaf display of part (A) to determine the median, the quartiles, and the interquartile range. Calculate by hand. Find the mean and standard deviation. Do the values calculated aid in the decision of determining if the data are normal or not? Explain.
 - Construct a boxplot of the data by hand. Indicate fences and if there are any outliers.
 - Organize the subjects into age classes starting at 60 with a class width of 5. Plot the frequency histogram.
- Natural killer (NK) cell activity was measured for three groups of subjects: those who had low, medium, and high scores on the Social Readjustment Rating Scale. The original observations, sample sizes, are provided by Irwin and his colleagues (1987) are given in the table below.

Natural killer cell activity (lytic units^a)

Low score ^b	moderate score ^b	High score ^b
22.2	15.1	10.2
97.8	23.2	11.3
29.1	10.5	11.4
37.0	13.9	5.3
35.8	9.7	14.5
44.2	19.0	11.0
82.0	19.8	13.6
56.0	9.1	33.4
9.3	30.1	25.0
19.9	15.5	27.0
39.5	10.3	36.3
12.8	11.0	17.7
37.4		

^a One lytic unit is defined as the number of effector cells killing 20% of the target cells.

^b The grouping is based on scores from the Social Readjustment Rating Scale.

You can use the computer for the values and graphs for these questions.

- Give two different estimates for the central measure of NK cell activity for each group. Which is the best measure for each group and why? Make sure to include any graphs if needed.
- How would you describe variability of NK cell activity for each group and why? Find the value.
- What is the range of NK cell activity for each group?
- What percentage of your sample had low scores on the Social Readjustment Rating Scale? Show calculations.
- You want to examine to see if NK cell activity is related to a subjects score on the Social Readjustment Rating Scale (low, moderate, High).

- (i) How would you look at this graphically?
- (ii) What conclusions can you draw from your graph and descriptive statistics about the NK cell activity and the three groups?
3. Consider a screening tool comprised of a series of questions designed to assess whether patients exhibit symptoms of depression. Each participant completes the series of questions and his/her scores are summed producing a total score. Patients are then classified as at low, moderate or high risk for depression based on their total scores. In the present study each patient completes the screening tool and, in addition, undergoes an extensive psychiatric examination. Based on the psychiatric examination, patients are classified as clinically depressed or not.

Clinical Assessment	Risk for Depression (Screening tool)			Total
	Low	Moderate	High	
Not Depressed	522	127	39	688
Depressed	28	18	11	57
Total	550	145	50	745

- (a) What is the probability that a patient who is at low risk for depression is clinically depressed?
- (b) What is the probability that a patient who is at moderate risk for depression is clinically depressed?
- (c) What is the probability that a patient who is at high risk for depression is clinically depressed?
- (d) Is the clinical assessment of depression independent of the patient's risk for depression? Explain.
- (e) What is the sensitivity (probability that a clinically depressed patient scores moderate/high) and specificity (probability that a non-clinically depressed patient tests scores low) for this study? What do these probabilities say about the screening tool?
- (f) Suppose that 60% of the participants in the study are male. In addition, suppose that 70% of the participants who are classified as clinically depressed are male.
- Are the events male and clinically depressed mutually exclusive? Explain using actual probabilities.
 - Are the events male and clinically depressed independent? Explain using actual probabilities.
 - Who is more likely to be clinically depressed- a male or a female in this study? Explain.