University of California, Santa Cruz
Department of Statistics
Baskin School of Engineering
Statistical Methods for the Biological,
Environmental, and Health Sciences
STAT 007

Answers to the Midterm

1. In an observational study subjects with and without heart disease were selected. Then researchers looked back in time and collect information regarding their identification, age, gender, and aspirin consumption, from hospital records.

   (a) Identify whether the above observational study is cross-sectional, retrospective, or prospective. Justify your answer.

   *The above observation study is an retrospective (or case-study) study **[2pts.]**. One group has heart disease the other does not and measurements for those subjects is collected from previous records, here hospital records **[2pts.]**.*

   (b) Identify whether the gender of the subjects in the study is of nominal, ordinal, interval or ratio level of measurement. Justify your answer.

   *The gender of the subjects is of nominal level of measurement **[2pts.]**. There is no order in that measurement **[2pts.]**.*

2. A professor wants to compare men and women performances in her classes. To do this, she conducts a survey by randomly selecting three different classes and surveying all of the students as they left those classes. One of the questions in the survey is how much time they spend studying for her class.

   (a) Identify which of the following types of sampling is used by the professor: random, systematic, convenience, stratified, or cluster. Justify your answer.

   *The sample collection method is cluster sampling **[2pts.]**. All the individuals from the same heterogeneous group are sampled. The groups are randomly sampled. **[2pts.]***

   (b) Which of the following types of graphs should she use to compare the amount of time that men and women spend studying for her class: histograms, dot graphs, stem-and-leaf graphs, frequency polygons, or time-series graphs. Justify your answer.

   *Frequency polygons should be used to compare the amount of time that men and women spend studying for the class. **[4pts.]***

3. The white blood cells counts (measured in 1000 cells/mL) and red blood cell counts (measured in million cells/mL) were measured to each of 12 subjects. The following statistics were computed for both cell counts:

   white blood cells counts: mean=6.73, median=7, midrange=7.4, standard deviation=1.46, variance=2.1316, range=4.6
   red blood cells counts: mean=4.57, median=6.3, midrange=4.75, standard deviation=0.42, variance=0.1764, range=1.5

   (a) Which blood cells count shows more variation, white blood cells counts or red blood cells counts? Justify your answer.

   To compare the variation on both groups we use the coefficient of variation. The measurements are in different units and the means and variances of both cell counts are very different ***[2pts.]****. For white blood cells counts* $CV = (1.46/6.73)100\% = 21.69\%(0.2169)$ ***[0.5pts.]****. For red blood cells counts* $CV = (0.42/4.57)100\% = 9.19\%(0.0919)$ ***[0.5pts.]****. So white blood cell counts seem to have more variation.* ***[1pts.]***

   (b) The highest value observed for white blood cells counts is 8.7, while the highest value observed for red blood cells counts is 5.5. Which of these two values is more extreme relative to the group from which they came?

   To determine which of these two values is more extreme we compute their z scores ***[2pts.]****. For white blood cells:* $zScore = (8.7 - 6.73)/1.46 = 1.3493$ ***[0.5pts.]****. For red blood cells:* $zScore = (5.5 - 4.57)/0.42 = 2.2142$ ***[0.5pts.]****. So, the red blood cells counts 5.5 seems to be more extreme than the white blood cells counts of 8.7****[1pts.]***

4. Consider a test that has 8 multiple choice questions. Assume that a student makes random guesses for each of the 8 questions and that the probability of correctly answering each question at random is 0.2.

   Let x be the random variable that describes the number of correct answers that the student has in the test.

   (a) Find the probability that no question is correctly answered by the student. Justify your answer.

   We assume that x is a random variable that follows a Binomial distribution with parameters $n = 8$ *and* $p = 0.2$ ***[2pts.]****. We need to compute* $P(0) = 0.1677$*, from the table at the back of the book* ***[2pts.]****.*

   (b) It is observed that the student answered 6 questions correctly. Would you say that this is a lucky or an unlucky student? Justify your answer.

We will use the rule of thumb to determine whether the value 6 is significantly high or not. If it is significantly high, the student was very lucky because observing such a large value under the assumption that $p = 0.2$ *would be rare* ***[2pts.]****. Note that* $\mu+2\sigma = n*p+2*\sqrt{n*p*(1-p)} = 8*0.2+2*\sqrt{8*0.2*0.8} = 3.8627$*. Therefore, the student got lucky* ***[2pts.]****.*
Just saying that the student is lucky because observing 6 out of 8 correct answers under the assumption that $p = 0.2$ *receives* ***[2pts.]****.*

5. Female pulse rates vary on a continuous scale and their mean parameter is 74 beats per minutes and their standard deviation parameter is 12.5 beats per minutes.

   Let x be the random variable that describes the pulse rate of a female.

   (a) Normal pulse rates for females vary between 60 and 100 beats per minutes. Find the probability that a randomly selected female has normal pulse rate.

   We assume that x follows a normal distirbution with mean $\mu = 74$ *and standard deviation* $\sigma = 12.5$*.* ***[1pts]****. We need to compute* $P(60 < x < 100)$***[1pts.]*** $= P((60-74)/12.5 < z < (100-74)/12.5) = P(z < 2.4) - P(z < -1.12)$***[1pts.]*** $= 0.9918 - 0.1314 = 0.8604$***[1pts.]***

   (b) Find the pulse rate of a female that separates the lowest 30% from the highest 70%.

   From the table of zScores we get that $z = -0.5244$ ***[2pts]*** *and so, the pulse rate of a female that separates the lowest 30% from the highest 70% is* $x = -0.5244*12.5+74 = 67.445$ ***[2pts]***