BMI 713: Computational Statistics for Biomedical Sciences

Assignment 2

September 16, 2010 (due Sept 23)

1 Random variables and distributions

1. Assume that a die is fair, i.e. if the die is rolled once, the probability of getting each of the six numbers is 1/6. Calculate the probability of the following events.

   - Rolling the die once, what is the probability of getting a number less than 3?
   - (Optional) Rolling the die twice, what is the probability that the sum of two rolling numbers is less than 3?

2. Let p be the probability of obtaining a head when flipping a coin. Suppose that Bob flipped the coin n ($n \geq 1$) times. Let X be the total number of head he obtained.

   - What distribution does the random variable X follow? Is X a discrete or continuous random variable?
   - What is the probability of $X = k$, i.e. what is $P(X = k)$ ($0 \leq k \leq n$)? (Write down the mathematical formula for calculating this probability.)
   - What is the probability of $X \geq k$, i.e. what is $Pr(X \geq k)$?
   - Suppose $p = 0.4$ and $n = 10$. Calculate the probabilities $Pr(X = 3)$ and $Pr(X \geq 3)$. (You may need the functions `dbinom` and `pbinom` in R to calculate these two probabilities. Use `?dbinom` and `?pbinom` to get help information of these two functions).

3. In a population of certain type of fish, the lengths of the individual fish follow a normal distribution. The mean length of the fish is 54.00 mm and the standard deviation is 4.50 mm. Answer the following questions.

   - What percentage of the fish are less than 63 mm?
   - What percentage of the fish are more than 50 mm?
   - Suppose that you randomly selected 10 fish from the population. What is the probability that the average length of the 10 fish is between 52 mm to 56 mm?

2 Hypothesis testing

1. In each of the following situations, state an appropriate null hypothesis H_0 and alternative hypothesis H_1. Be sure to identify the parameters that you use to state the hypotheses.

   (a) An experiment on learning in animals measures how long it takes a mouse to find its way through a maze. The mean time is 18 seconds for one particular maze. A researcher thinks that a loud noise will cause the mice to complete the maze slower. She measures how long each of 10 mice takes with a noise as stimulus.

   (b) A pharmaceutical company developed a new drug for certain type of cancer and the company believed that the drug can significantly increase patient's survival time after surgery. The mean survival time after surgery is 16 months. The company selected 20 volunteers who had the surgery, treated them with the drug and records their survival time.

2. Determine if the following statements are true.

   (a) P-value is the probability that the null hypothesis is true.

   (b) P-value is the probability, computed assuming that H_0 is true, that the test statistics will take a value at least as extreme as that actually observed.

   (c) Standard normal distribution has fatter tail than student t-distribution.

   (d) When the sample size is small, t-test is more accurate than z-test.

3. The mean level of calcium in the blood in healthy young adults is about 9.5 milligrams per deciliter. A clinic in Boston measures the blood calcium level of 10 healthy pregnant women as follows

   $$9.09,\ 9.82,\ 9.58,\ 9.03,\ 10.48,\ 9.35,\ 9.85,\ 9.36,\ 9.64,\ 9.43.$$

   Is this an indication that the mean calcium level in the population from which these women come differs from 9.5?

   - State the null hypothesis H_0 and and the alternative hypothesis H_1.
   - Calculate the mean and standard deviation of the blood calcium level of the 10 women.
   - Calculate the z-score and perform the z-test based on the z-score. What is the P-value?
   - Calculate the t-statistic and perform the t-test. What is the P-value? (Optional) What is the degree of freedom of this t-test?