BIOSTAT 701 Homework 2

Note: Only use R (or other software) if it is indicated in the question. Full credit will not be given to answers without work shown.

1. (5 pts) Let Y be a binomial random variable with n = 10 and p = 0.4. Find the following values without using R:

a. $P(Y \leqslant 2)$

b. $P(Y > 2)$ you should be able to get this pretty easily, without adding up lots of numbers.

2. (10 pts) In an inspection of automobiles in a city, 30% of all automobiles had emissions that did not meet EPA regulation. We take a random sample of 9 automobiles. Let *X* denote the number of automobiles in the sample that fail the inspection. Construct a table that shows the probabilities and cumulative probabilities *[Note that you may use R to find the probabilities, etc and then type those in to create the table .]* Use the table to answer:

a. What is the probability that exactly 2 of the 9 failed the inspection?
b. What is the probability that 6 and more of the 9 failed the inspection?
c. What is the probability that at least 2 of the 9 failed the inspection?

3. (10 pts) Over a long period of time in a large multinational corporation, 10% of all sales trainees are rated as outstanding, 75% are rated as excellent/good, 10% are rated as satisfactory, and 5% are considered unsatisfactory.

a. Compute the probability that three out of the 10 trainees are rated as satisfactory or excellent/good.
b. For the 10 trainees selected from part a, compute the probability that 1 is rated as satisfactory.

4. (10 pts) The SAT are scored so as to yield a mean of 500 and a standard deviation of 100. These scores are close to being normally distributed. An exclusive club wishes to invite those scoring in the top 5% on the College Boards to join.

c. Find percentages of scores that is between 480 and 590.
d. What score is required to be invited to join the club?
e. What score separates the top 60% of the population from the bottom 40%? What do we call this value? *(hint: the score that separates the top 20% of the population from the bottom 80% is called the 80th percentile)*

5. (5 pts) In a large construction company, 9 workers are randomly selected, only 2 are female. Is there evidence to doubt that the proportion of female workers is 0.3 or more? (A probability less than a default value of 0.05 is considered doubtful. If the probability is more than 0.05, one fails to doubt.) Note that here it is not appropriate to use the empirical rule approach since the distribution is not close enough to normal distribution.

6. (10 pts) The College Boards, which are administered each year to many thousands of high school students, are scored so as to yield a mean of 500 and a standard deviation of 100. These scores are close to being normally distributed. What percentage of the scores can be expected to satisfy each condition?

a. Greater than 750

b. Greater than 600

c. Less than 450

d. Between 450 and 600.

e. A class of 16 students is from the reference population. What is the probability that the average SAT score of the class is between 450 and 600?

7. (10 pts) Simulate a data set with 100 individuals, each drawn from a standard normal distribution. Square these values, then add them. Is this sum near 100? What distribution have you simulated? Do you obtain similar

results using an R function which samples from this distribution directly?

8. (10 pts) R has a function which you can use to directly simulate multinomial random variables. But, as an exercise, suppose that the only tool available to you is a function for simulating Bernoulli random variables. How could you repeatedly apply this function to simulate a multinomial random variable? (Describing the algorithm is sufficient, but it is so much better if you can program this in R.)

9. (10 pts) BIOSTAT701 will have a quiz consisting of 15 questions with three true-or-false questions, and the rest are multiple-choice questions. Find the probability that a student will answer exactly three questions correctly. Specify any assumptions that you make (if any).

10. (5 pts) Suppose you perform n independent trials, where each trial results in exactly one of k possible outcomes, with probabilities $p_1, p_2, \dots, p_k$ (where $p_1 + p_2 + \dots + p_k = 1$). Let X_i be the number of times outcome i occurs, derive the PMF for $(X_1, \dots, X_k)$.

10. (15 pts) Verify the following properties of indicator functions.

(a) $1_A(x) = 1 - 1_{\bar{A}}(x)$, where $\bar{A}$ denotes the complement of A.

(b) $1_{A_1 \cap A_2 \cap \cdots \cap A_n}(x) = 1_{A_1}(x)\, 1_{A_2}(x) \cdots 1_{A_n}(x)$.

(c) $1_{A_1 \cup A_2 \cup \cdots \cup A_n}(x) = \max\{\, 1_{A_1}(x),\, 1_{A_2}(x), \ldots,\, 1_{A_n}(x)\}$.

(d) $[\, 1_A(x)]^2 = 1_A(x)$.