1. (30%) True or False. Justify your answer.

a. If E and F are two independent events, then the sum of P(E) and P(F) cannot exceed 1.

b. There is a one in six chance of rolling a pair if two dice are rolled.

c. If two events are mutually exclusive, then the sum of their probabilities is 1.

d. If two events are independent, then they are automatically mutually exclusive.

e. If $P(E \cup F) = P(E) + P(F)$, then E and F must be disjoint.

f. If a coin is tossed five times, then the probability of heads coming up twice is the same as the probability of heads coming up three times.

g. If heads has come up six times in a row, then tails is more likely to come up on the seventh toss.

h. If you toss a coin seven times, is more likely that heads will come up six times and tails once than that heads will come up seven times in a row.

i. A random variable must assign a different number to each possible outcome.

j. The histogram of X will be highest at the expected value of X.

k. The expected value of X is half-way between the largest and smallest possible values of X.

l. If m is the median of X, then it must be the case that $P(X \leq m) = 1/2$ and $P(X \geq m) = 1/2$.

m. If E(X) is the expected value of X, and $x_1, x_2, \ldots, x_n$ are X-scores obtained in an experiment, then $(x_1 + x_2 + \ldots + x_n) / n = E(X)$.

n. We should expect the actual values of X obtained in experiments to be within one standard deviation away from the mean.

o. For any X, at least 3/4 of its values obtained in experiments will lie within 2 standard deviations of its mean.

2.(10%) An Urn contains *n* white and *m* black balls. One ball at a time is randomly withdrawn until the first white ball is drawn. Find the expected number of black balls that are withdrawn.

3. (10%) At a clinic, on the average, only 70% of potential new patients can talk immediately with a social worker when they call. The other 30% are asked to leave their phone numbers. About 60% of the time a social worker can return the call on the same day, and the other 40% of the time the caller is contacted on the following day. Experience at the clinic indicates that the probability a caller will actually visit the clinic for consultation is 0.8 if the caller was immediately able to speak to a social worker, whereas it is 0.6 and 0.4, respectively, if the patient's call was returned the same day or the following day.

a. What percentage of callers who visit the clinic for consultation?

b. What percentage of patients that visit the clinic didn't have to have their calls returned?

4. (10%) Consider the following questions regarding sampling.
   a. (4%) The process of acquiring data incurs costs and such costs are often proportionate to the sample size. Yet, we desire to have a larger sample size, if cost of sampling is manageable. Why? Please give explanations from the viewpoint of statistics.
   b. (2%) What is *sampling without replacement*? Please explain.
   c. (4%) The *sampling error* is inevitable in sampling. But, assessing the sampling error is not trivial because the population parameter considered is unknown. What would be your assessment of the sampling error? Please elaborate it.
5. (20%) Consider the fundamental yet puzzling issue of adopting textbooks written in English on the collegiate level. It is believed that adoption of textbooks written in English has a long-term influence on preparing students for graduate studies. But, college students often have a relatively short-term goal—passing each test. Thus, students are apt to use the Chinese versions of textbooks, after considering the limited preparation time they have for each test, even if instructors adopt the English versions of textbooks. To justify students' decisions, you are asked to test whether college students who study the English versions of textbooks and those who study the Chinese versions of textbooks have different test performances in courses in which textbooks are written in English and tests are prepared in English.
   a. (4%) Describe and justify the data acquisition process, and prepare the acquired data in a tabular form, where each sample size is set to be 30. (You may randomly assign data values in the table.)
   b. (8%) If you consider conducting a parametric test, how does the testing procedure work? What would the corresponding assumption(s) be? (The table prepared in item a might be helpful in answering this question.)
   c. (8%) If you consider conducting a nonparametric test, how does the testing procedure work? What would the corresponding assumption(s) be? (Again, the table prepared in item a might be helpful in answering this question.)
6. (20%) Consider the following questions regarding simple linear regression.
   a. (4%) What is the simple linear regression model? What are the model assumptions?
   b. (4%) Construct the general form of the ANOVA table for simple linear regression.
   c. (4%) Show that the *total sum of squares* is equal to the sum of two terms in the ANOVA table considered in item b.
   d. (4%) Note that a particular testing procedure is used in the ANOVA table considered in item b. Please explain the logic underlying the use of this test.
   e. (4%) Do the least squares method give good estimators for the parameters of the simple linear regression model? If yes, please justify them.