Statistics and Data Analysis
Midterm Exam

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

Note. This exam is in-class and open everything (including all kinds of electronic devices). However, an exam taker is not allowed to communicate with any person during the exam. Cheating will result in severe penalty. You do not need to return the problem sheet.

1. (10 points) Consider a population of the following 20 numbers:

$$178, 172, 175, 184, 192, 175, 165, 178, 177, 175, 180, 182, 187, 183, 160, 178, 179, 162, 170, 171.$$

   (a) (4 points) Draw a histogram with four classes $[155, 165)$, $[165, 175)$, ..., and $[185, 195)$. Which class has the highest frequency? What is that frequency?

   (b) (3 points) Draw a pie chart that illustrates the proportions of odd and even numbers.

   (c) (3 points) Calculate the mean, median, and mode of these 20 numbers.

2. (10 points) Load the data set "SDA-Fa14_data_wholesale.txt". Consider the sales data collected from channel 2 and region 1.

   (a) (2 points) Calculate the means, medians, and sample variances for milk sales.

   (b) (2 points) Calculate the means, medians, and sample variances for fresh food sales.

   (c) (3 points) Find the sample coefficients of variation for milk and fresh food sales. Which variable has higher variability?

   (d) (3 points) Calculate the correlation coefficient for milk and fresh food sales. Are them positively or negatively correlated?

3. (10 points) You have decided that you will buy a lottery ticket if and only if your expected earning is larger than the ticket price. The ticket costs \$5. With probability 0.001, you win \$1000; with probability 0.03, you win \$100; with probability 0.1, you win \$5. Should you buy a ticket for this lottery? Why?

4. (10 points) Let X be the number of heads after 3 flips of an unfair coin with

$$\Pr(X = \text{Head}) = 0.2 = 1 - \Pr(X = \text{Tail}).$$

   (a) (2 points) List all the possible values of X.

   (b) (5 points) Find the probability distribution of X.

   (c) (3 points) Is the distribution symmetric, positively skewed, or negatively skewed? Why?

5. (10 points; 5 points each) Consider a random variable X with the following probability distribution:

x	5	6	7	10
$\Pr(X = x)$	0.5	0.25	0.125	0.125

   (a) Draw a bar chart to illustrate the probability distribution.

   (b) What are the expected value, variance, and standard deviation of X?

6. (10 points) Calculate the following probabilities:

   (a) (4 points) If $X \sim \text{ND}(-10, 4)$, find $\Pr(X \leq -7)$.

(b) (4 points) If $X \sim \text{ND}(4, 10)$, find $\Pr(0 \leq X \leq 13)$.

(c) (2 points) If $X \sim \text{ND}(4, 10)$, find $\Pr(X = 0)$.

7. (10 points) The daily demand of a product $X \sim \text{ND}(70, 8)$. At the end of each day, you place an order to order q units from your supplier. The products will be ready at your store the next morning. Unsold products will become valueless.

   (a) (3 points) If $q = 80$, what is the probability of having a shortage?

   (b) (3 points) Suppose you want to have a service level of 90%, i.e., with probability 90% you will fulfill all daily demands. Find the minimum q (must be an integer) that achieves this service level.

   (c) (4 points) For $q = 50$, 60, 70, 80, and 90, find the corresponding service level. Then draw a scatter plot for the 5 pairs of order quantities and service levels.

8. (10 points) A laptop manufacturer produces a type of laptop starting from 2013. At that time, the average battery life is 6 hours. Last month, a new technology was introduced, and the average battery life was expected to increase. The manufacturer wants to know whether the new technology really brought a significant improvement and make the current average battery life $\mu > 7$. For simplicity, let's assume that the current standard deviation is known to be 0.6 hour.

   (a) (3 points) Suppose that we will randomly draws 36 laptops and obtain a sample mean $\overline{X}$. What is the sampling distribution of $\overline{X}$?

   (b) (3 points) If μ is still 6, how likely will we get a sample mean that is below 5.8?

   (c) (4 points) If μ is still 6, find a such that $\Pr(\overline{X} > a) = 0.05$.

9. (20 points; 5 points each) Qualitatively answer the following true-and-false questions regarding interval estimation for the population mean. For each problem (no matter your answer is true or false), briefly explain the reason.

   (a) Suppose the population variance is unknown. For a given confidence level, enlarging the confidence interval requires us to increase the sample size.

   (b) Suppose the population variance is known. For a given sample size, increasing the confidence level will enlarge the confidence interval.

   (c) Suppose the population follows a normal distribution, the sample size is large, and the population variance is known. In this case, we can use the z distribution to do estimation.

   (d) Suppose the population follows a non-normal distribution, the sample size is small, and the population variance is unknown. In this case, we can use the t distribution to do estimation.