Statistics and Data Analysis
Distributions and Sampling

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

1. In the "Bike" sheet, the daily casual, registered, and total bike rentals are recorded for two years. Sometimes the daily casual rentals account for more than $\frac{1}{3}$ of the total rentals. We are interested in whether that will happen for a given future day.
   (a) According to the two-year data, what is the probability for that to happen on a future day?
   (b) According to the two-year data, what is the probability for that to happen on a future non-working day?
   (c) According to only the data in 2012, what is the probability for that to happen on a future day?

2. Consider the MS Excel worksheet "Bike." We want to use the 731 observed "registered" values to infer the probability distribution of daily registered bike rentals in *any* day. The given histogram suggests that daily registered bike rental are normally distributed.

   (a) Calculate the mean $\bar{x}$ and sample standard deviation s for the 731 observed "registered" values.

   (b) Suppose that $X \sim \mathrm{ND}(\bar{x}, s)$, find the probabilities for X to be within each class.

   (c) Find the theoretical frequencies of X for all classes.

   (d) Compare the two sequences of observed and theoretical frequencies. Are they close to each other?

3. Consider the MS Excel worksheet “Bike” and the distribution fitting task we just did. Let k be the number of classes, O_i be the observed frequency of class i, and T_i be the theoretical frequency of class i. We define

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - T_i)^2}{T_i}$$

as a measurement of how the theoretical distribution is close to the observed distribution.

(a) Calculate χ^2 for $\mathrm{ND}(\bar{x}, s)$.

(b) Calculate χ^2 for $\mathrm{ND}(3500, 1500)$.

4. Weights of bags of candies produced in a factory follow ND$(2, \sigma)$. The quality control officer randomly draws n bags and calculate the sample mean $\bar{x}$. If $\bar{x} > 2.2$ or $\bar{x} < 1.8$, we will get punished. What is the probability for us to get punished under the following conditions?

   (a) $\sigma = 0.5$ and $n = 5$?

   (b) $\sigma = 0.1$ and $n = 5$?

   (c) $\sigma = 0.5$ and $n = 3$?

   (d) $\sigma = 0.1$ and $n = 3$?

5. Let X be the battery life (in hours) of laptops manufactured in our factory. Suppose that $X \sim \text{ND}(5.5, 0.6)$. Assume that the battery lives of two laptops are independent.

(a) Let X_1 and X_2 be the battery lives of the first and second drawn laptops. Find the probability distribution of $\frac{X_1+X_2}{2}$.

(b) Suppose now n laptops are sampled. Let X_i be the battery life of the ith drawn laptop and $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$ be the sample mean. Find the probability distribution of $\bar{x}$.

(c) An inspection rule is adopted: If the sample mean of 2 laptops is less than 5.2, the whole batch will be rejected. Find the probability of rejection.

6. Let X be the battery life (in hours) of laptops manufactured in our factory. Suppose that $\mu_X = 5.5$, $\sigma_X = 0.6$, and we have no more information about X's distribution. Assume that the battery lives of two laptops are independent. n laptops are sampled. Let X_i be the battery life of the ith drawn laptop and $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$ be the sample mean.

   (a) An inspection rule is adopted: If the sample mean of $n = 36$ laptops is less than 5.2, the whole batch will be rejected. Find the probability of rejection.

   (b) What if $n = 16$?