

# Statistics and Data Analysis, Fall 2015

## Homework 2: Probability, Distributions, and Sampling

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

This homework is due **6:35 pm, November 2, 2015**. Each student should submit her/his own **hard copy** to the instructor at the beginning of the class. All the data for this homework are contained in the file “SDA-Fa15\_hw02\_data.xlsx”. Discuss with your classmates but NEVER copy one’s work.

1. (25 points; 5 points each) Let  $X$  be a discrete random variable whose sample space is  $\{0, 5, 10, 11, 12, 15\}$ . Its probability distribution is:

$x$	0	5	10	11	12	15
$\Pr(X = x)$	0.2	0.1	0.3	0.2	0.1	?

- (a) Find  $\Pr(X = 15)$ .
- (b) Find  $\mathbb{E}[X]$ , the expected value of  $X$ .
- (c) Find  $\text{Var}(X)$ , the variance of  $X$ .
- (d) Let  $Y = X + 5$  be another random variable. Find the probability distribution of  $Y$ .
- (e) Find  $\mathbb{E}[Y]$  and  $\text{Var}(Y)$ .
2. (25 points; 5 points each) A lottery ticket costs \$100. With probability 0.1%, you win \$10000. With probability 5%, you win \$1000. Otherwise you win nothing.
- (a) Find the expected value of your profit (earning minus cost) of buying this ticket.
- (b) Find the standard deviation of your profit of buying this ticket.
- (c) Suppose that you may pay \$50 to buy an insurance. Once you buy the insurance, you get paid \$50 if you do not win anything in the lottery. Find the expected value of your profit of buying both the ticket and insurance.
- (d) Find the standard deviation of your profit of buying both the ticket and insurance.
- (e) Is it a good idea to buy the insurance? Make your own comments.
- Note.** This question has no standard solution. Your answer will be graded based on the logic.
3. (25 points; 5 points each) Let  $X$  be the battery life (in hours) of laptops manufactured in our factory.<sup>1</sup> Once a batch of 1000 laptops are manufactured, a laptop will be randomly drawn from the batch for a test. If the battery life is less than  $t$  hours, a second laptop will be randomly drawn for the same test. If both laptops fail the test, the whole batch will be rejected; otherwise, the batch will be accepted. Suppose that  $X \sim \text{ND}(\mu, 0.5)$ . Assume that the battery lives of two laptops are independent.
- (a) If  $\mu = 5.5$  and  $t = 5.2$ , find the probability for the first laptop to fail the test.
- (b) If  $\mu = 5.5$  and  $t = 5.2$ , find the probability for the batch to be rejected.
- (c) If  $\mu = 5.5$ , find a value  $t$  so that the probability for the first laptop to fail the test is 5%.
- (d) If  $\mu = 5.5$ , find a value  $t$  so that the probability for the batch to be rejected is 1%.
- (e) If  $t = 5.2$ , find a value  $\mu$  so that the probability for the batch to be rejected is 1%.

---

<sup>1</sup>The battery life of a laptop refers to how long this laptop can work after a single charge.

4. (25 points; 5 points each) A bank let its officers make telemarketing phone calls to sell long-term deposits. 4119 records of past calls are recorded in the sheet “Bank.” The meanings of the columns are listed below:

- age: The age of the customer.
- job: The job of the customer. Possible values include admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, and unknown.
- marital: The marital status of the customer. Possible values include divorced (which means divorced or widowed), married, single, and unknown.
- education: The education level of the customer. Possible values include basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, and unknown.
- default: Whether the customer has any default record. Possible values are no, yes, and unknown.
- housing: Whether the customer has housing loan. Possible values are no, yes, and unknown.
- loan: Whether the customer has non-housing loan. Possible values are no, yes, and unknown.
- duration: The length of the last call (in seconds).
- buy: Whether the customer decides to buy a long-term deposit in that call. Possible values are no and yes.

Answer the following questions based on the data set.

- (a) Do you think “education” and “buy” are independent? Why or why not?
- (b) Estimate the probability distribution of the job of the next customer.
- (c) Draw a histogram for “age” by using  $[0, 10)$ ,  $[10, 20)$ , ..., and  $[80, 90)$  as classes.
- (d) Fit a normal distribution to “age” based on the histogram obtained above. Use sample mean and sample standard deviation to be the mean and standard deviation of the normal distribution. For each class, find the observed and theoretical frequencies. Find the total difference (i.e., the sum of the differences in all the classes).
- (e) Comment on your task in Part (d). Do you think it is good to estimate “age” as a normal random variable?

**Note.** This question has no standard solution. Your answer will be graded based on the logic.

5. (0 point; do not submit your answers for this problem) Continue from Problem 3. Let  $\mu = 5.5$ .

- (a) Let  $X_1$  and  $X_2$  be the battery lives of the first and second drawn laptops. Find the probability distribution, mean, and standard deviation of  $\frac{X_1+X_2}{2}$ .
- (b) Suppose now  $n$  laptops may be sampled. Let  $X_i$  be the battery life of the  $i$ th drawn laptop and  $\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$  be the sample mean. Find the probability distribution, mean, and standard deviation of  $\bar{x}$ .
- (c) A new inspection rule is adopted: If the sample mean of  $n$  laptops is less than  $t$ , the whole batch will be rejected. Let  $t = 5.2$  and  $n = 2$ , find the probability of rejection. Compare your answer with that in Problem 3b. Explain why they are the same or different.