Statistics and Data Analysis
Homework 3

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

1. Consider a random variable X with the following probability distribution:

x	1	2	3	4
$\Pr(X = x)$	0.5	0.25	0.125	0.125

   (a) Draw a bar chart to illustrate the probability distribution.
   (b) What is the expected value of X, $\mathbb{E}[X]$?
   (c) What is the variance of X, $\text{Var}(X)$?
   (d) What is the standard deviation of X?

2. Consider a random variable X that is uniformly distributed between 0 and 10.
   Note. We will write $X \sim \text{Uni}(a, b)$ if X is uniformly distributed between a and b.

   (a) Find $\Pr(X = 3)$.
   (b) Find $\Pr(X \geq 20)$.
   (c) Find $\Pr(X \geq 6)$.
   (d) Find $\Pr(2 \leq X \leq 7)$.
   (e) Find the pdf of X.

3. By using R, one may easily generate random numbers following a given distribution, which is equivalent to sampling from a population that follows that distribution. In particular, we may use the function `sample(x, size)` to randomly draw a sample of size `size` from the population `x` *without replacement*.

   (a) Try `x <- seq(1, 10, 1)` and then `sample(x, 10)` for a few times. What do you get?
   (b) Now try `sample(x, 10, replace = TRUE)` for a few times. Is there anything different?
   (c) How to randomly draw one student in our class to win a prize? How to randomly draw three?

4. Let X be the outcome of rolling a fair dice. We know it follows a discrete uniform distribution between 1 and 6. Let's roll this dice for 10000 times and count the frequencies:

```
trial <- 10000
x <- rep(0, trial)
for(i in 1:trial)
{
  x[i] <- sample(1:6, 1)
}
table(x)
```

   Does the outcome look uniform?

5. Let X_1, X_2, and X_3 be the outcome of rolling three fair dices. What is the probability distribution of $Y = X_1 + X_2 + X_3$?

   (a) What are the possible values of Y?
   (b) Let's roll these dices for 10000 times and count the frequencies:

```
trial <- 10000
x <- rep(0, trial)
for(i in 1:trial)
{
  x[i] <- sum(sample(1:6, 3, replace = TRUE))
}
oc <- as.data.frame(table(x))
barplot(oc$Freq / trial, names.arg = oc$x)
```

6. We have used the R function `pnorm(q, mean, sd)` to calculate the *left-tail probability* of normal distributions. Let's do some more practices.

   (a) If $X \sim \text{ND}(10, 2)$, find $\Pr(X \leq 7)$.

   (b) If $X \sim \text{ND}(10, 2)$, find $\Pr(X \geq 13)$.

   (c) If $X \sim \text{ND}(200, 50)$, find $\Pr(120 \leq X \leq 170)$.

   (d) If $Z \sim \text{ND}(0, 1)$, find $\Pr\left(\frac{120 - 200}{50} \leq Z \leq \frac{170 - 200}{50}\right)$.

7. Besides `pnorm()`, there are other functions related to the normal distribution:

   (a) The function `rnorm(n, mean, sd)` generates `n` random samples from a normal population with mean `mean` and standard deviation `sd`. To see this, try `hist(rnorm(100, 10, 2))`. How to make the output look more like a normal distribution?

   (b) The function `qnorm(p, mean, sd)` returns a number `x` such that `p` equals `pnorm(x, mean, sd)`. To see how this works, try `q <- qnorm(0.05, 10, 2)` and `p <- pnorm(q, 10, 2)`.

   (c) If $X \sim \text{ND}(80, 10)$, find x such that $\Pr(X \leq x) = 0.8$.

8. The daily demand of a product $X \sim \text{ND}(80, 10)$. 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) If $q = 95$, what is the probability of having a shortage?

   (b) 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) For each integer between 60 and 100, find the corresponding service level. Then draw a scatter plot for the 41 pairs of order quantities and service levels.

   (d) Write R codes without `qnorm()` to solve Part (b).
   Hint. The service levels found in Part (c) are useful.