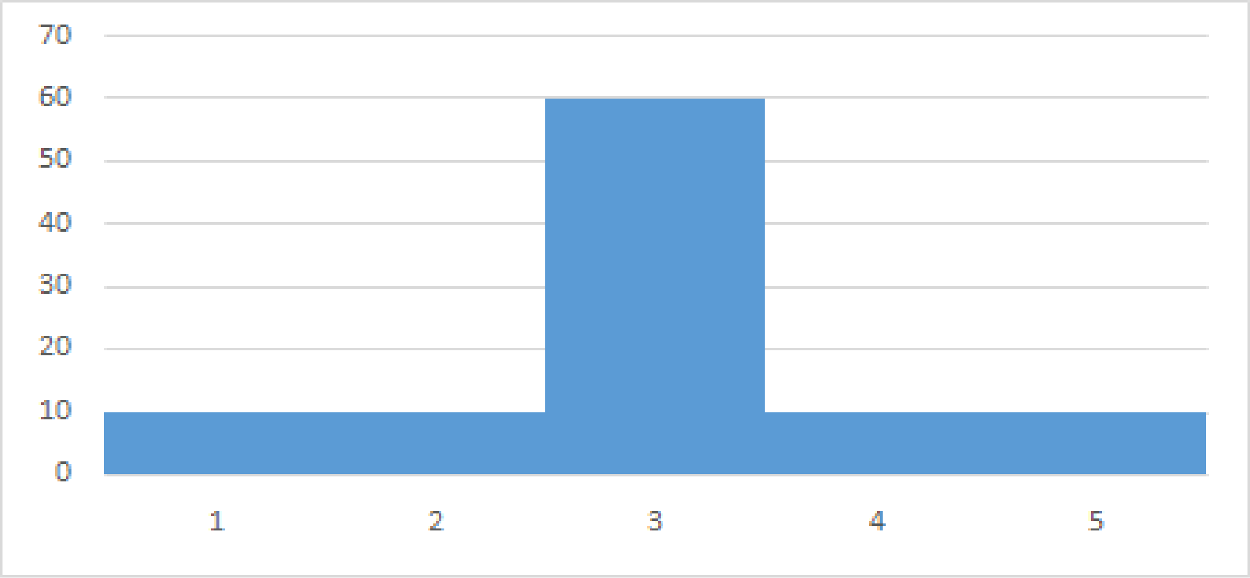
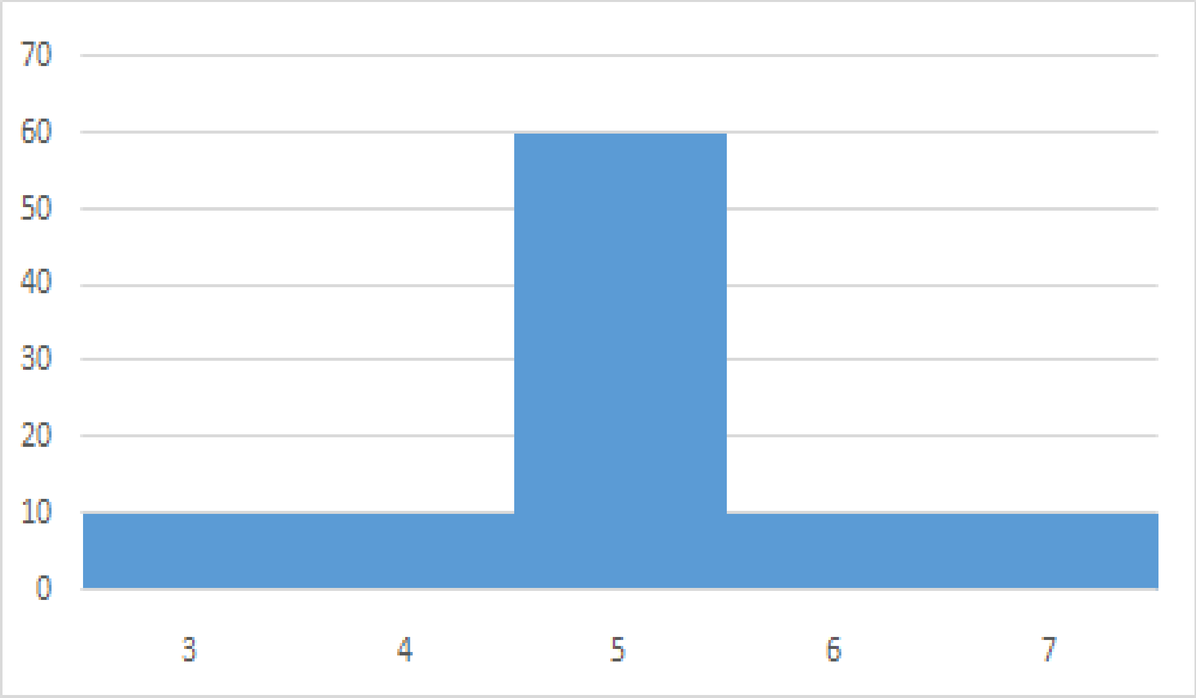
Linear Functions of Random Variables

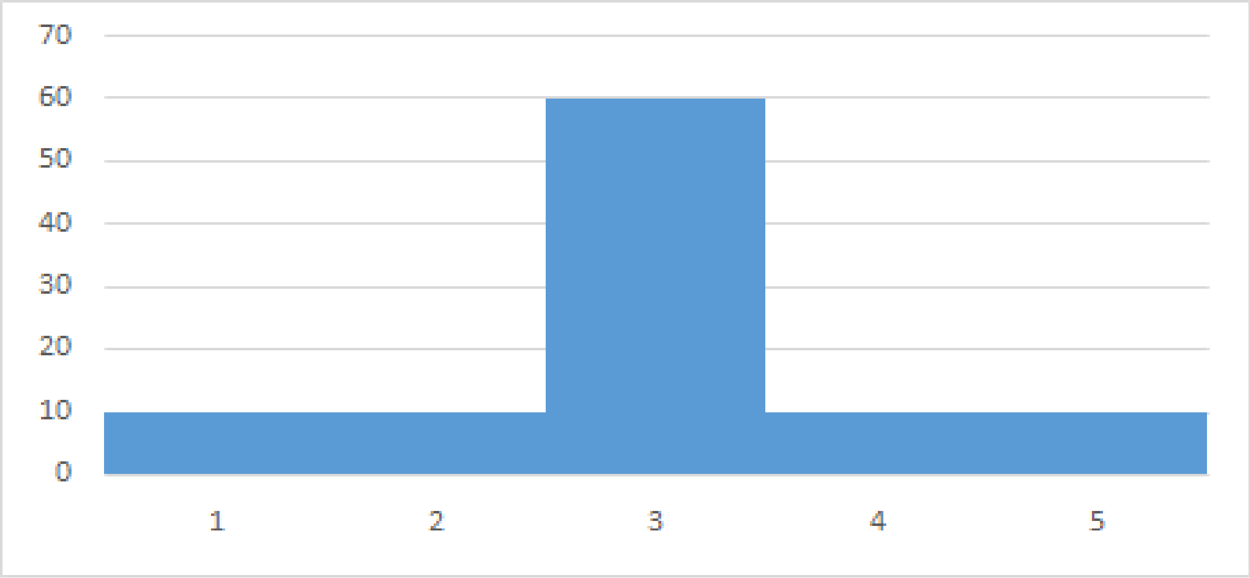
Dr.Mamatha.H.R
Professor
Department of Computer Science and Engineering
PES University
Bangalore

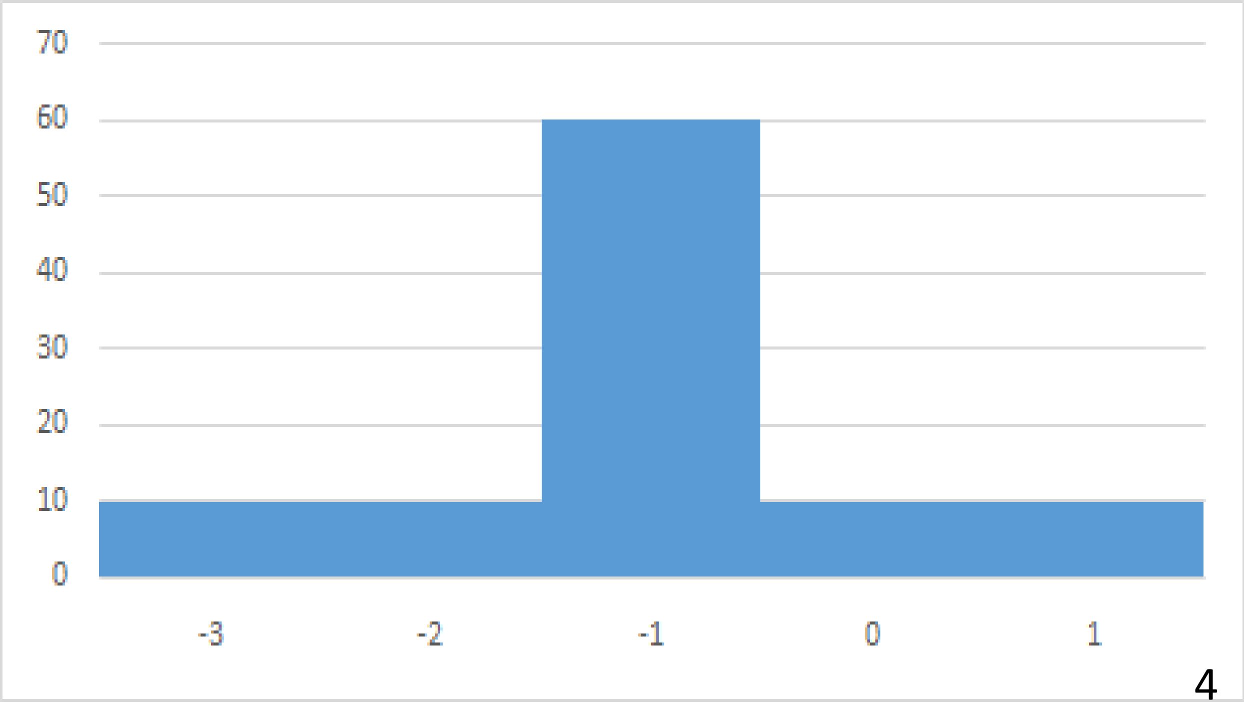
Course material created using text book

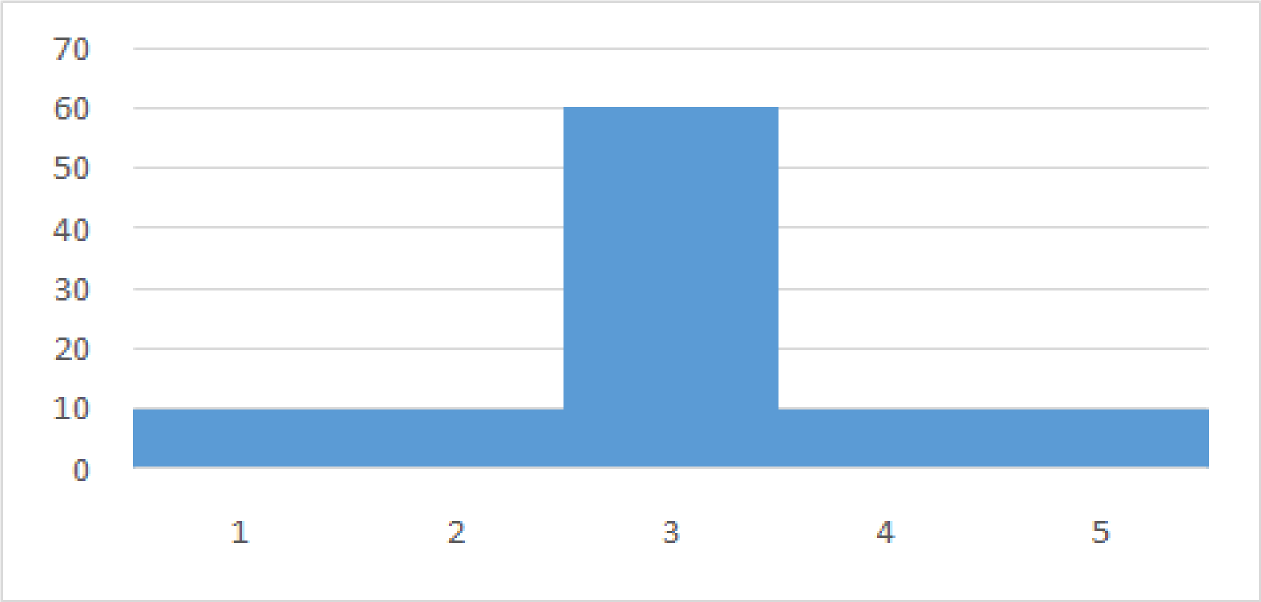
- we often construct new random variables by performing arithmetic operations on other random variables.
- For example, we might add a constant to a random variable, multiply a random variable by a constant, or add two or more random variables together.


70
60
50
40
30
20
10
0
1
2
3
4
5

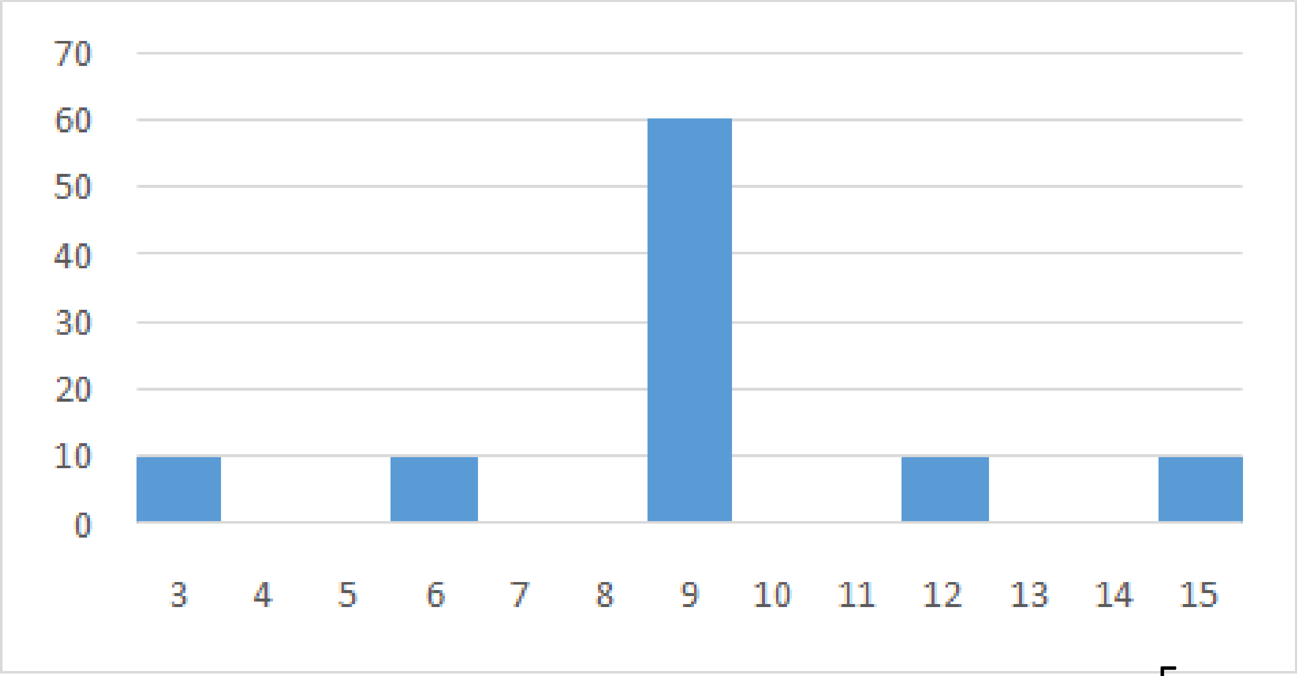

70
60
50
40
30
20
10
0
3
4
5
6
7


70
60
50
40
30
20
10
0
1
2
3
4
5


70
60
50
40
30
20
10
0
-3
-2
-1
0
1



70
60
50
40
30
20
10
0
1
2
3
4
5



70
60
50
40
30
20
10
0
3
4
5
6
7
8
9
10
11
12
13
14
15

Adding a Constant

- When a constant is added to a random variable, the mean is increased by the value of the constant, but the variance and standard deviation are unchanged.

If X is a random variable and b is a constant, then

$$\mu_{X+b} = \mu_X + b$$

$$\sigma^2_{X+b} = \sigma^2_X$$

Multiplying by a Constant

- Often we need to multiply a random variable by a constant
- For example, to convert to a more convenient set of units.
- multiplication by a constant affects the mean, variance, and standard deviation of a random variable

- In general, when a random variable is multiplied by a constant, its mean is multiplied by the same constant

If X is a random variable and a is a constant, then

$$\mu_{aX} = a\mu_X$$

- In general, when a random variable is multiplied by a constant, its variance is multiplied by the *square of the constant.*

If X is a random variable and a is a constant, then

$$\sigma_{aX}^2 = a^2\sigma_X^2$$

$$\sigma_{aX} = |a|\sigma_X$$

- If a random variable is multiplied by a constant and then added to another constant,the effect on the mean and variance

If X is a random variable, and a and b are constants, then

$$\mu_{aX+b} = a\mu_X + b$$

$$\sigma^2_{aX+b} = a^2\sigma^2_X$$

$$\sigma_{aX+b} = |a|\sigma_X$$

Marie has a part-time job walking dogs to earn money on weekends. The following probability distribution represents the probability of having a particular number of clients on any given day. If she earns $2.75 per client, how much could she expect to earn each day, on average, and what is the standard deviation of her expected earnings?

# clients	20	25	30	35	40
probability	.15	.35	.30	.15	.05

Start by finding the mean:

$\mu X = 20 \times .15 + 25 \times .35 + 30 \times .3 + 35 \times .15 + 40 \times .05 = 28$

Use the mean to find the variance:

$\sigma^2 X = (20-28)2 \times .15 + (25-28)2 \times .35 + (30-28)2 \times .30 + (35-28)2 \times .15 + (40-28)2 \times .05 = 28.5$

Use the variance to find the standard deviation: $\sigma X = \sqrt{\ } 28.5 = 5.3$

Now we can find her average income by multiplying the mean, 28 by Marie’s rate, $2.75, to get her **average daily income of $77**.

Finally, we can multiply the calculated standard deviation, 5.3, by the rate, \$2.75, to get the standard deviation of her income: $5.3 \times \$2.75 = \$14.58$

What all this means is that Marie can expect to average \$77 per day, on average, give or take about \$14.58.

Means of Linear Combinations of Random Variables

If $X_1, X_2, \ldots, X_n$ are random variables, then the mean of the sum $X_1 + X_2 + \cdots + X_n$ is given by

$$\mu_{X_1+X_2+\cdots+X_n} = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n} \tag{2.47}$$

If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then the random variable

$$c_1 X_1 + \cdots + c_n X_n$$

is called a **linear combination** of $X_1, \ldots, X_n$.

If X and Y are random variables, and a and b are constants, then

$$\mu_{aX+bY} = \mu_{aX} + \mu_{bY} = a\mu_X + b\mu_Y \tag{2.48}$$

More generally, if $X_1, X_2, \ldots, X_n$ are random variables and $c_1, c_2, \ldots, c_n$ are constants, then the mean of the linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is given by

$$\mu_{c_1X_1+c_2X_2+\cdots+c_nX_n} = c_1\mu_{X_1} + c_2\mu_{X_2} + \cdots + c_n\mu_{X_n} \tag{2.49}$$

Independent Random Variables

- The notion of independence for random variables is very much like the notion of independence for events.
- Two random variables are independent if knowledge concerning one of them does not affect the probabilities of the other.
- When two events are independent,the probability that both occur is found by multiplying the probabilities for each event.
- Let *X be* a random variable and let *S be a set of numbers. The notation "X $\in$ S" means that the*
- value of the random variable *X is in the set S.*

If X and Y are **independent** random variables, and S and T are sets of numbers, then

$$P(X \in S \text{ and } Y \in T) = P(X \in S)P(Y \in T) \tag{2.50}$$

More generally, if $X_1, \ldots, X_n$ are independent random variables, and $S_1, \ldots, S_n$ are sets, then

$$P(X_1 \in S_1 \text{ and } X_2 \in S_2 \text{ and} \cdots \text{ and } X_n \in S_n) = \\ P(X_1 \in S_1)P(X_2 \in S_2) \cdots P(X_n \in S_n) \tag{2.51}$$

Rectangular plastic covers for a compact disc (CD) tray have specifications regarding length and width. Let X be the length and Y be the width, each measured to the nearest millimeter, of a randomly sampled cover. The probability mass function of X is given by $P(X = 129) = 0.2$, $P(X = 130) = 0.7$, and $P(X = 131) = 0.1$. The probability mass function of Y is given by $P(Y = 15) = 0.6$ and $P(Y = 16) = 0.4$. The area of a cover is given by $A = XY$. Assume X and Y are independent. Find the probability that the area is 1935 mm^2.

Solution

The area will be equal to 1935 if $X = 129$ and $Y = 15$. Therefore

$$\begin{aligned} P(A = 1935) &= P(X = 129 \text{ and } Y = 15) \\ &= P(X = 129)P(Y = 15) \qquad \text{since } X \text{ and } Y \text{ are independent} \\ &= (0.2)(0.6) \\ &= 0.12 \end{aligned}$$

Variances of Linear Combinations of Independent Random Variables

If $X_1, X_2, \ldots, X_n$ are *independent* random variables, then the variance of the sum $X_1 + X_2 + \cdots + X_n$ is given by

$$\sigma^2_{X_1+X_2+\cdots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n} \qquad (2.52)$$

If $X_1, X_2, \ldots, X_n$ are *independent* random variables and $c_1, c_2, \ldots, c_n$ are constants, then the variance of the linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is given by

$$\sigma^2_{c_1X_1+c_2X_2+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + c_2^2\sigma^2_{X_2} + \cdots + c_n^2\sigma^2_{X_n} \qquad (2.53)$$

If X and Y are *independent* random variables with variances σ_X^2 and σ_Y^2, then the variance of the sum $X + Y$ is

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \tag{2.54}$$

The variance of the difference $X - Y$ is

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 \tag{2.55}$$

The fact that the variance of the difference is the *sum* of the variances may seem counterintuitive. However, it follows from Equation (2.53) by setting $c_1 = 1$ and $c_2 = -1$.

A piston is placed inside a cylinder. The clearance is the distance between the edge of the piston and the wall of the cylinder and is equal to one-half the difference between the cylinder diameter and the piston diameter. Assume the piston diameter has a mean of 80.85 cm with a standard deviation of 0.02 cm. Assume the cylinder diameter has a mean of 80.95 cm with a standard deviation of 0.03 cm. Find the mean clearance. Assuming that the piston and cylinder are chosen independently, find the standard deviation of the clearance.

Solution

Let X_1 represent the diameter of the cylinder and let X_2 the diameter of the piston. The clearance is given by $C = 0.5X_1 - 0.5X_2$. Using Equation (2.49), the mean clearance is

$$\begin{aligned}\mu_C &= \mu_{0.5X_1-0.5X_2}\\ &= 0.5\mu_{X_1} - 0.5\mu_{X_2}\\ &= 0.5(80.95) - 0.5(80.85)\\ &= 0.050\end{aligned}$$

Since X_1 and X_2 are independent, we can use Equation (2.53) to find the standard deviation σ_C:

$$\begin{aligned}\sigma_C &= \sqrt{\sigma^2_{0.5X_1-0.5X_2}}\\ &= \sqrt{(0.5)^2\sigma^2_{X_1} + (-0.5)^2\sigma^2_{X_2}}\\ &= \sqrt{0.25(0.02)^2 + 0.25(0.03)^2}\\ &= 0.018\end{aligned}$$

Independence and Simple Random Samples

- When a simple random sample of numerical values is drawn from a population, each
  item in the sample can be thought of as a random variable.
- The items in a simple random sample may be treated as independent, except when the sample is a large proportion (more than 5%) of a finite population

If $X_1, X_2, \ldots, X_n$ is a simple random sample, then $X_1, X_2, \ldots, X_n$ may be treated as independent random variables, all with the same distribution.

- When *X1, . . . , Xn are independent random variables, all with the same distribution, it is*
- sometimes said that *X1, . . . , Xn are* ***independent and identically distributed (i.i.d.).***

Sampling Distributions

In inferential statistics, we want to use characteristics of the sample (i.e. a **statistic**) to estimate the characteristics of the population (i.e. a **parameter**).

If we obtain a random sample and calculate a sample statistic from that sample, the sample statistic is a random variable .

The population parameters, however, are fixed. If the statistic is a random variable, can we find the distribution? The mean? The standard deviation?

The answer is yes! This is why we need to study the sampling distribution of statistics. So what is a sampling distribution?

The sampling distribution of a statistic is a probability distribution based on a large number of samples of size n from a given population.

Sampling Distribution of the Sample Mean

In this example, the population is the weight of six pumpkins (in pounds) displayed in a carnival "guess the weight" game booth. You are asked to guess the average weight of the six pumpkins by taking a random sample without replacement from the population.

Pumpkin	A	B	C	D	E	F
Weight (in pounds)	19	14	15	9	10	17

Since we know the weights from the population, we can find the population mean.

$\mu=(19+14+15+9+10+17)/6=14$ pounds

To demonstrate the sampling distribution,

let's start with obtaining all of the possible samples of size n=2 from the populations, sampling without replacement.

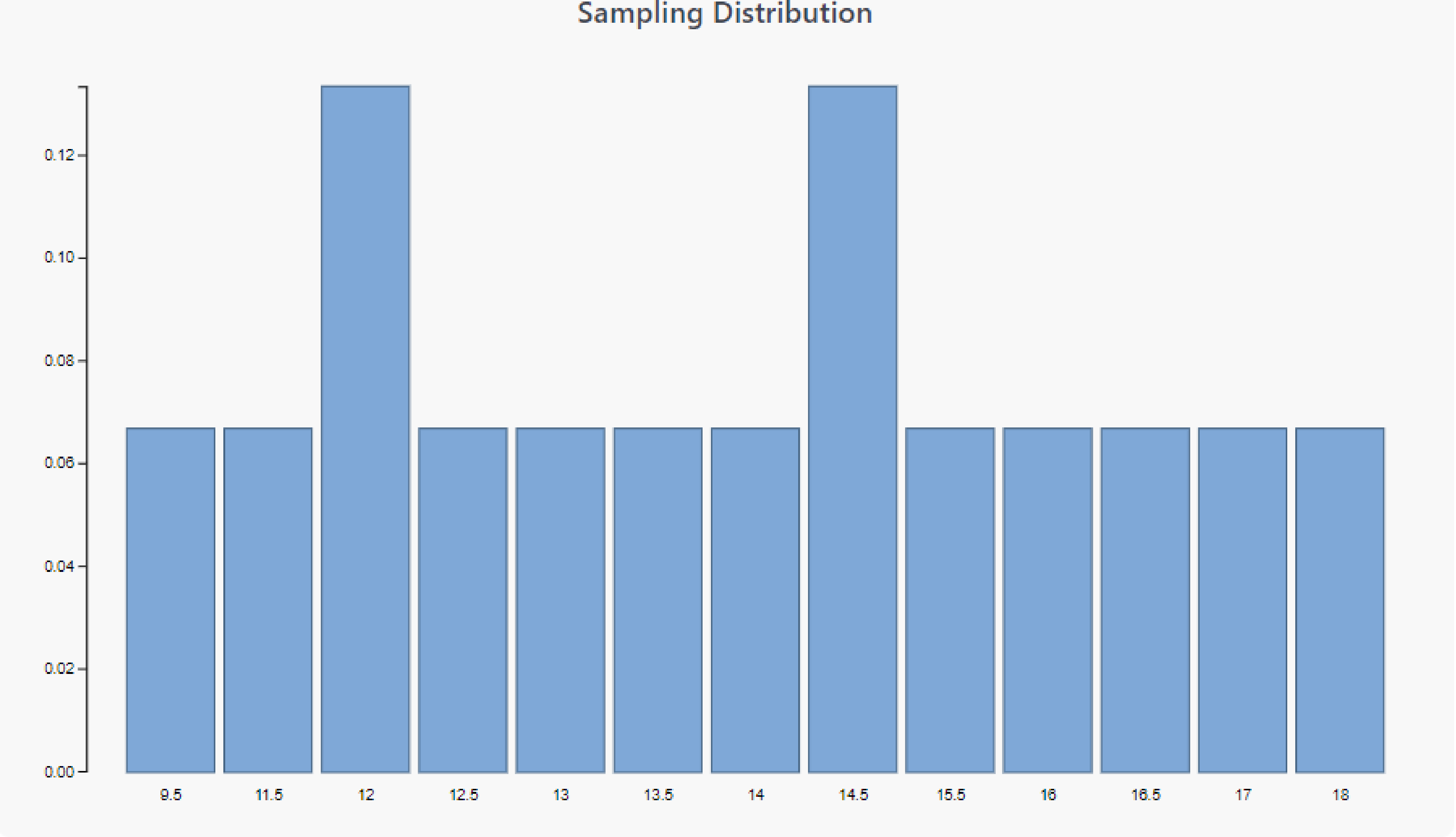
The table show all the possible samples, the weights for the chosen pumpkins, the sample mean and the probability of obtaining each sample. Since we are drawing at random, each sample will have the same probability of being chosen.

Sample	Weight	$\bar{x}$	Probability
A, B	19, 14	16.5	1/15
A, C	19, 15	17.0	1/15
A, D	19, 9	14.0	1/15
A, E	19, 10	14.5	1/15
A, F	19, 17	18.0	1/15
B, C	14, 15	14.5	1/15
B, D	14, 9	11.5	1/15
B, E	14, 10	12.0	1/15
B, F	14, 17	15.5	1/15
C, D	15, 9	12.0	1/15
C, E	15, 10	12.5	1/15
C, F	15, 17	16.0	1/15
D, E	9, 10	9.5	1/15
D, F	9, 17	13.0	1/15
E, F	10, 17	13.5	1/15

We can combine all of the values and create a table of the possible values and their respective probabilities.

X	9.5	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.5	16.0	16.5	17.0	18.0
Probability	1/15	1/15	2/15	1/15	1/15	1/15	1/15	2/15	1/15	1/15	1/15	1/15	1/15

The table is the probability table for the sample mean and it is the sampling distribution of the sample mean weights of the pumpkins when the sample size is 2.



the chance that the sample mean is exactly the population mean is only 1 in 15, very small.

Now that we have the sampling distribution of the sample mean, we can calculate the mean of all the sample means. In other words, we can find the mean (or expected value) of all the possible $\bar{x}$'s.

The mean of the sample means is

$$\mu_{\bar{x}}=\sum \bar{x}_i f(\bar{x}_i)=9.5(1/15)+11.5(1/15)+12(2/15)+12.5(1/15)+13(1/15)+13.5(1/15)+14(1/15)+14.5(2/15)+15.5(1/15)+16(1/15)+16.5(1/15)+17(1/15)+18(1/15)=14$$

let's do the same thing as above but with sample size n=5

Sample	Weights	$\bar{x}$	Probability
A, B, C, D, E	19, 14, 15, 9, 10	13.4	1/6
A, B, C, D, F	19, 14, 15, 9, 17	14.8	1/6
A, B, C, E, F	19, 14, 15, 10, 17	15.0	1/6
A, B, D, E, F	19, 14, 9, 10, 17	13.8	1/6
A, C, D, E, F	19, 15, 9, 10, 17	14.0	1/6
B, C, D, E, F	14, 15, 9, 10, 17	13.0	1/6

The sampling distribution is:

X	13.0	13.4	13.8	14.0	14.8	15.0
Probability	1/6	1/6	1/6	1/6	1/6	1/6

The mean of the sample means is...

μ=(1/6)(13+13.4+13.8+14.0+14.8+15.0)

=14 pounds

The Mean and Variance of a Sample Mean

- The most frequently encountered linear combination is the sample mean.
- Specifically,if *$X1, \ldots, Xn$ is a simple random sample from a population with mean μ and variance*
- *σ^2, then the sample mean X is the linear combination*

$$\overline{X} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n$$

From this fact we can compute the mean and variance of $\overline{X}$.

$$\begin{aligned}
\mu_{\overline{X}} &= \mu_{\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n} \\
&= \frac{1}{n}\mu_{X_1} + \cdots + \frac{1}{n}\mu_{X_n} && \text{(using Equation 2.49)} \\
&= \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu \\
&= (n)\left(\frac{1}{n}\right)\mu \\
&= \mu
\end{aligned}$$

X_i are identically distributed, which means they have the same mean μ

- the items in a simple random sample may be treated as independent random variables. Therefore

$$
\begin{aligned}
\sigma_{\overline{X}}^2 &= \sigma^2_{\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n} \\
&= \frac{1}{n^2}\sigma^2_{X_1} + \cdots + \frac{1}{n^2}\sigma^2_{X_n} \qquad \text{(using Equation 2.53)} \\
&= \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 \\
&= (n)\left(\frac{1}{n^2}\right)\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the sample mean $\overline{X}$ is a random variable with

$$\mu_{\overline{X}} = \mu \tag{2.56}$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \tag{2.57}$$

The standard deviation of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{2.58}$$

A process that fills plastic bottles with a beverage has a mean fill volume of 2.013 L and a standard deviation of 0.005 L. A case contains 24 bottles. Assuming that the bottles in a case are a simple random sample of bottles filled by this method, find the mean and standard deviation of the average volume per bottle in a case.

Solution

Let $V_1, \ldots, V_{24}$ represent the volumes in 24 bottles in a case. This is a simple random sample from a population with mean $\mu = 2.013$ and standard deviation $\sigma = 0.005$. The average volume is $\overline{V} = (V_1 + \cdots + V_{24})/24$. Using Equation (2.56),

$$\mu_{\overline{V}} = \mu = 2.013$$

Using Equation (2.58),

$$\sigma_{\overline{V}} = \frac{\sigma}{\sqrt{24}} = 0.001$$