

5.2 – Random Variables:

Objectives:

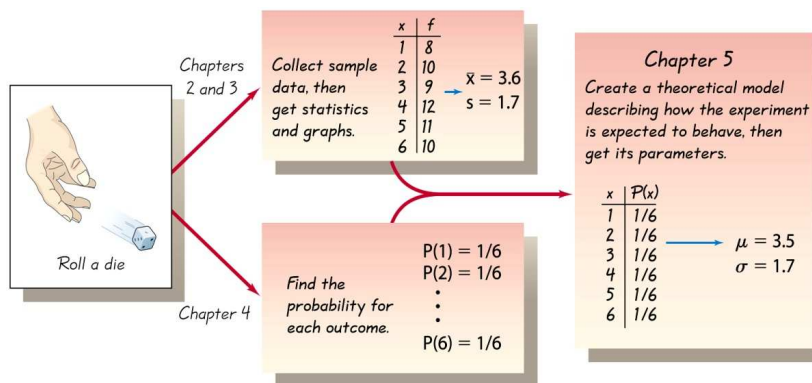
1. Distinguish between discrete and continuous random variables
2. Determine whether a probability distribution is given
3. Find the mean and standard deviation of a probability distribution
4. Identify unusual results with probabilities
5. Determine the expected value.

This chapter combines the methods of **descriptive statistics** presented in Chapter 2 and 3 and those of **probability** presented in Chapter 4 to describe and analyze **probability distributions**.

Probability Distributions describe what will probably happen instead of what actually did happen, and they are often given in the format of a graph, table, or formula.

Combining Descriptive Methods and Probabilities:

In this chapter we will construct probability distributions by presenting possible outcomes along with the relative frequencies we expect.



In order to fully understand probability distributions, we must first understand the concept of a random variable, and be able to distinguish between discrete and continuous random variables.

Random Variables:

A random variable is a variable (typically represented by x) that has a single numerical value, determined by chance, for each outcome of a procedure

Probability Distribution:

A probability distribution is a description that gives the probability for each value of the random variable; often expressed in the format of a graph, table, or formula

x	$P(x)$
0	0.22
1	0.14
2	0.26
3	0.09
4	0.29

Types of Random Variables:

In chapter 1 we learned about discrete and continuous data. Random variables may represent either discrete or continuous data.

Discrete Random Variable

either a finite number of values or countable number of values, where “countable” refers to the fact that there might be infinitely many values, but they result from a counting process

Continuous Random Variable

infinitely many values and those values can be associated with measurements on a continuous scale without gaps or interruptions

Example: Identify the given random variable as being discrete or continuous.

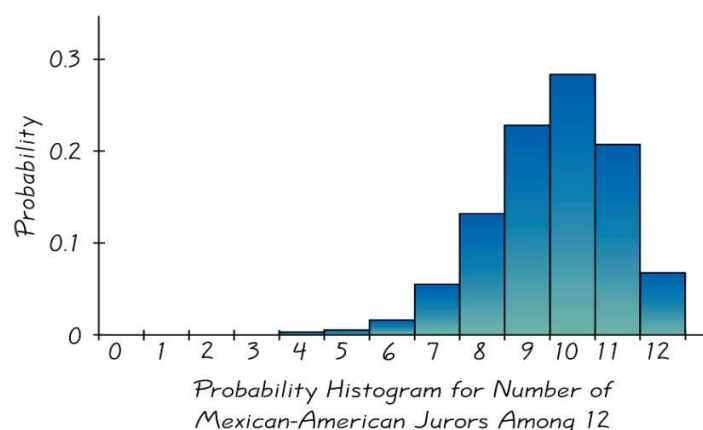
1. The cost of a randomly selected orange
2. The pH level in a shampoo
3. The number of phone calls between New York and California on Thanksgiving day
4. The height of a randomly selected student
5. The braking time of a car
6. The number of field goals kicked in a football game

Solutions:

1. Discrete
2. Continuous
3. Discrete
4. Continuous
5. Continuous
6. Discrete

Probability Histograms:

The probability histogram is very similar to a relative frequency histogram, but the vertical scale shows probabilities. The total area of all the rectangles of a probability histogram is 1. This correspondence between area and probability is very useful in statistics and will be further explored in future lessons.



Requirements for Probability Distributions:

Every probability distribution must satisfy each of the following requirements.

1. $\sum P(x) = 1$ where x assumes all possible values.
2. $0 \leq P(x) \leq 1$ for every individual value of x .

The first event comes from the simple fact that the random variable x represents all possible events in the sample space, so we can be certain that one of the events will occur.

The second requirement is an extension of the rule that all probabilities must be between 0 and 1.

Example: Is the following a probability distribution. If not, identify the requirement that is not satisfied.

x	$P(x)$
0	0.142
1	0.182
2	0.214
3	0.175
4	0.229
5	0.171

Solution: It is NOT a probability distribution because the sum of the probabilities is greater than 1.

Example: Is the following a probability distribution. If not, identify the requirement that is not satisfied.

x	$P(x)$
0	0.290
1	0.278
2	-0.047
3	0.063
4	0.292
5	0.124

Solution: It is NOT a probability distribution because one of the probabilities is less than 0.

Example: If a person is randomly selected from a certain town, the probability distribution for the number, x , of siblings is as described in the accompanying table.

x	$P(x)$
0	0.24
1	0.32
2	0.21
3	0.14
4	0.05
5	0.03

Solution: It is NOT a probability distribution because the sum of the probabilities is 0.99, not 1.

Mean, Variance and Standard Deviation of a Probability Distribution:

The mean, variance, and standard deviation for a probability distribution can be found using the following formulas.

$$\mu = \sum [x \cdot P(x)] \quad \text{Mean for a probability distribution}$$

$$s^2 = \sum [(x - \mu)^2 \cdot P(x)] \quad \text{Variance for a probability distribution}$$

$$s^2 = \sum [x^2 \cdot P(x)] - \mu^2 \quad \text{Variance (shortcut) for a probability distribution}$$

$$s = \sqrt{\sum [x^2 \cdot P(x)] - \mu^2} \quad \text{Standard Deviation for a probability distribution}$$

Example: The random variable x is the number of houses sold by a realtor in a single month at the Sellmore Real Estate office. Its probability distribution is as follows. Find the mean and the standard deviation for the probability distribution.

Houses Sold (x)	Probability $P(x)$
0	0.24
1	0.01
2	0.12
3	0.16
4	0.01
5	0.14
6	0.11
7	0.21

Solution: The best way to find the mean and standard deviation is to use the formulas in an excel spreadsheet by making a table as follows:

x	$P(x)$	$xP(x)$	$[(x - \mu)^2 \cdot P(x)]$
0	0.24	0	3.110
1	0.01	0.01	0.068
2	0.12	0.24	0.307
3	0.16	0.48	0.058
4	0.01	0.04	0.002
5	0.14	0.7	0.274
6	0.11	0.66	0.634
7	0.21	1.47	2.428
		3.6	2.623

Therefore, the mean is 3.6 and the standard deviation is 2.623

Example: In a certain town, 40% of adults have a college degree. The accompanying table describes the probability distribution for the number of adults (among 4 randomly selected adults) who have a college degree. Find the mean and standard deviation for the probability distribution.

x	$P(x)$
0	0.1296
1	0.3456
2	0.3456
3	0.1536
4	0.0256

Solution: The best way to find the mean and standard deviation is to use the formulas in an excel spreadsheet by making a table as follows:

x	$P(x)$	$xP(x)$	$[(x - \mu)^2 \cdot P(x)]$
0	0.1296	0	0.332
1	0.3456	0.3456	0.124
2	0.3456	0.6912	0.055
3	0.1536	0.4608	0.301
4	0.0256	0.1024	0.147
		1.6	0.980

Therefore, the mean is 1.6 and the standard deviation is 0.980

Identifying *Unusual* Results with the Range Rule of Thumb

A simple tool for understanding standard deviation is the range rule of thumb. This rule is based on the principle that for many data sets 95% of the data values lie within 2 standard deviations of the mean. This rule may be expressed as:

$$\text{Minimum Value} = \text{Mean} - 2 \text{ times Standard Deviation}$$

$$\text{Maximum Value} = \text{Mean} + 2 \text{ times Standard Deviation}$$

Rare Event Rule for Inferential Statistics

If the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

- ❖ Unusually high: x successes among n trials is an unusually high number of successes if $P(x \text{ or more}) \leq 0.05$.
- ❖ Unusually low: x successes among n trials is an unusually low number of successes if $P(x \text{ or fewer}) \leq 0.05$.

Example: Suppose that weight of adolescents is being studied by a health organization and that the accompanying table describes the probability distribution for three randomly selected adolescents, where x is the number who are considered morbidly obese. Is it unusual to have no obese subjects among three randomly selected adolescents?

x	$P(x)$
0	0.111
1	0.215
2	0.450
3	0.224

Solution: The mean of the probability distribution is 1.787 and the standard deviation is 0.915.

x	$P(x)$	$xP(x)$	$[(x - \mu)^2 \cdot P(x)]$
0	0.111	0	0.354
1	0.215	0.215	0.133
2	0.45	0.9	0.020
3	0.224	0.672	0.330
		1.787	0.915

- ❖ Using the range rule of thumb we have $1.787 - 2(0.915) = -0.043$. Because $x = 0$ is within this range, it is NOT unusual to have no obese subjects among three randomly selected adolescents.
- ❖ Using the rare event rule, $P(event) = 0.111 \geq 0.05$ and is therefore not a rare or unusual event.

Example: Suppose that voting in municipal elections is being studied and that the accompanying table describes the probability distribution for four randomly selected people, where x is the number that voted in the last election. Is it unusual to find four voters among four randomly selected people?

x	$P(x)$
0	0.23
1	0.32
2	0.26
3	0.15
4	0.04

Solution: The mean of the probability distribution is 1.45 and the standard deviation is 1.117.

x	$P(x)$	$xP(x)$	$[(x - \mu)^2 \cdot P(x)]$
0	0.23	0	0.484
1	0.32	0.32	0.065
2	0.26	0.52	0.079
3	0.15	0.45	0.360
4	0.04	0.16	0.260
		1.45	1.117

- ❖ Using the range rule of thumb we have $1.45 - 2(1.117) = 3.684$. Because $x = 4$ is not within this range, it is considered an unusual value.
- ❖ Using the rare event rule, $P(event) = 0.04 \leq 0.05$ and is therefore a rare or unusual event.

Expected Value:

The expected value of a discrete random variable is denoted by E , and it represents the mean value of the outcomes. It is obtained by using the following formula:

$$E = \sum [x \cdot P(x)]$$

Note that this is the same formula we use for finding the mean of a probability distribution.

Example: Suppose you buy 1 ticket for \$1 out of a lottery of 1,000 tickets where the prize for the one winning ticket is to be \$500. What is your expected value?

Solution: Make a table using excel.

Event	x	P(x)	xP(x)
Win	499	0.001	0.499
Lose	-1	0.999	-0.999
Expected Value			-0.5

The expected value is -\$0.50.

Example: A 28-year-old man pays \$ 165 for a one-year life insurance policy with coverage of \$140,000. If the probability that he will live through the year is 0.9994, what is the expected value for the insurance policy?

Solution: Make a table using excel.

Event	x	P(x)	xP(x)
live	-165	0.9994	-164.901
die	139,835	0.0006	83.901
Expected Value			-81

The expected value is -\$81.00

Example: Suppose you pay \$ 3.00 to roll a fair die with the understanding that you will get back \$ 5.00 for rolling a 1 or a 6, nothing otherwise. What is your expected value?

Solution: Make a table using excel.

Event	x	P(x)	xP(x)
win	2	0.334	0.668
lose	-3	0.666	-1.998
Expected Value			-1.33

The expected value is -\$1.33.

Example: Professor Strong has \$10,000 to invest, and his financial analyst recommends two types of junk bonds. The A bonds have a 6% annual yield with a default rate of 1%. The B bonds have an 8% annual yield with a default rate of 5%. If the bond defaults the \$10,000 is lost. Which of the two bonds is better? Why? Should he select either bond? Why or why not?

Solution: Make a table using excel. From a comparison of the tables below it can be seen that Bond A has the greatest expected value and is therefore the better investment. However, either bond would be a good investment since both expected values are positive.

Bond A	x	P(x)	xP(x)
Yield	600	0.99	594
Default	-10,000	0.01	-100
Expected Value			494.00

Bond B	x	P(x)	xP(x)
Yield	800	0.95	760
Default	-10,000	0.05	-500
Expected Value			260.00