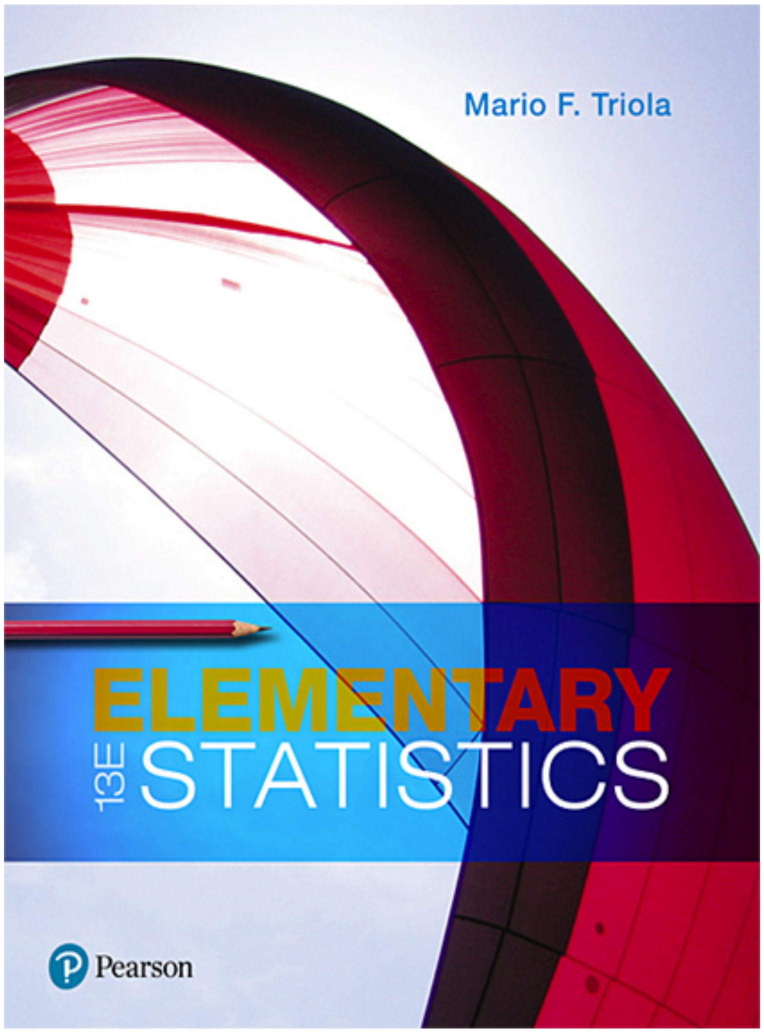
Elementary Statistics

Thirteenth Edition



Chapter 5

Probability Distributions

Probability Distributions

5-1 Probability Distributions

5-2 Binomial Probability Distributions

5-3 Poisson Probability Distributions

Key Concept

The focus of this section is the **binomial probability distribution** and methods for finding probabilities.

Easy methods for finding the mean and standard deviation of a binomial distribution are also presented.

As in other sections, we stress the importance of **interpreting** probability values to determine whether events are **significantly low** or **significantly high.**

Binomial Probability Distribution (1 of 2)

- Binomial Probability Distribution
  - A **binomial probability distribution** results from a procedure that meets these four requirements:
    1. The procedure has a **fixed number of trials.** (A trial is a single observation.)
    2. The trials must be **independent,** meaning that the outcome of any individual trial doesn’t affect the probabilities in the other trials.

Binomial Probability Distribution (2 of 2)

- Binomial Probability Distribution
  - A **binomial probability distribution** results from a procedure that meets these four requirements:
    3. Each trial must have all outcomes classified into exactly **two categories,** commonly referred to as **success** and **failure.**
    4. The probability of a success remains the same in all trials.

Notation for Binomial Probability Distributions (1 of 3)

S and *F* (success and failure) denote the two possible categories of all outcomes.

$P(S) = p(p$ = probability of a success)

$P(F) = 1 - p = q(q$ = probability of a failure)

n the fixed number of trials

Notation for Binomial Probability Distributions (2 of 3)

x - a specific number of successes in *n* trials, so *x* can be any whole number between 0 and *n*, inclusive

p - probability of **success** in **one** of the *n* trials

q - probability of **failure** in **one** of the *n* trials

P(*x*) - probability of getting exactly *x* successes among the *n* trials

Notation for Binomial Probability Distributions (3 of 3)

The word **success** as used here is arbitrary and does not necessarily represent something good. Either of the two possible categories may be called the success *S* as long as its probability is identified as *p*.

CAUTION When using a binomial probability distribution, always be sure that *x* and *p* are **consistent** in the sense that they both refer to the **same** category being called a success.

Example: Twitter (1 of 8)

When an adult is randomly selected (with replacement), there is a 0.85 probability that this person knows what Twitter is (based on results from a Pew Research Center survey). Suppose that we want to find the probability that exactly three of five randomly selected adults know what Twitter is.

a. Does this procedure result in a binomial distribution?

b. If this procedure does result in a binomial distribution, identify the values of n, x, p, and q.

Example: Twitter (2 of 8)

Solution

a. This procedure does satisfy the requirements for a binomial distribution, as shown below.

1. The number of trials (5) is fixed.
2. The 5 trials are independent because the probability of any adult knowing Twitter is not affected by results from other selected adults.

Example: Twitter (3 of 8)

Solution

3. Each of the 5 trials has two categories of outcomes: The selected person knows what Twitter is or that person does not know what Twitter is.
4. For each randomly selected adult, there is a 0.85 probability that this person knows what Twitter is, and that probability remains the same for each of the five selected people.

Example: Twitter (4 of 8)

Solution

b. Having concluded that the given procedure does result in a binomial distribution, we now proceed to identify the values of *n, x, p,* and *q*.

1. With five randomly selected adults, we have $n = 5$.
2. We want the probability of exactly three who know what Twitter is, so $x = 3$.

Example: Twitter (5 of 8)

Solution

3. The probability of success (getting a person who knows what Twitter is) for one selection is 0.85, so $p = 0.85$.
4. The probability of failure (not getting someone who knows what Twitter is) is 0.15, so $q = 0.15$.

Example: Twitter (6 of 8)

Solution

Again, it is very important to be sure that *x* and *p* both refer to the same concept of "success." In this example, we use *x* to count the number of people who know what Twitter is, so *p* must be the probability that the selected person knows what Twitter is. Therefore, *x* and *p* do use the same concept of success: knowing what Twitter is.

Treating Dependent Events as Independent

5% Guideline for Cumbersome Calculations

When sampling without replacement and the sample size is no more than 5% of the size of the population, treat the selections as being **independent** (even though they are actually dependent).

Methods for Finding Binomial Probabilities (1 of 4)

Method 1: Binomial Probability Formula

$$P(x) = \frac{n!}{(x-n)!\,x!} \cdot p^x \cdot q^{n-x} \quad \text{for } x = 0, 1, 2, 3, \ldots, n$$

where

n = number of trials

x = number of successes among n trials

p = probability of success in any one trial

q = probability of failure in any one trial ($q = 1 - p$)

Example: Twitter (7 of 8)

Given that there is a 0.85 probability that a randomly selected adult knows what Twitter is, use the binomial probability formula to find the probability that when five adults are randomly selected, exactly three of them know what Twitter is. That is, apply the previous formula to find $P(3)$ given that $n = 5$, $x = 3$, $p = 0.85$, and $q = 0.15$.

Example: Twitter (8 of 8)

Solution

Using the given values of *n, x, p,* and *q* in the binomial probability formula, we get

$$P(3) = \frac{5!}{(5-3)!3!} \cdot 0.85^3 \cdot 0.15^{5-3}$$

$$= \frac{5!}{2!3!} \cdot 0.614125 \cdot 0.0225$$

$$= (10)(0.614125)(0.0225)$$

$$= 0.138178$$

$$= 0.138 \text{ (round to three significant digits)}$$

The probability of getting exactly three adults who know Twitter among five randomly selected adults is 0.138.

Methods for Finding Binomial Probabilities (2 of 4)

Method 2: Using Technology

Technologies can be used to find binomial probabilities. The screen displays on the next slide list binomial probabilities for n = 5 and p = 0.85, as in the previous example. Notice that in each display, the probability distribution is given as a table.

Methods for Finding Binomial Probabilities (3 of 4)

Method 2: Using Technology

Statdisk

Binomial Probability

Number of Trials, n: 5

Success Prob, p: 0.85

Evaluate

Results for all values of x are provided unless you enter a specific value for x here:

Mean: 4.2500
Standard Deviation: 0.7984
Variance: 0.6375

x	P(x)	P(x or fewer)	P(x or greater)
0	0.0000759	0.0000759	1.0000000
1	0.0021516	0.0022275	0.9999241
2	0.0243844	0.0266119	0.9977725
3	0.1381781	0.1647900	0.9733881
4	0.3915047	0.5562947	0.8352100
5	0.4437053	1.0000000	0.4437053

Clear Copy

TI-83/84 Plus CE

NORMAL FLOAT AUTO REAL RADIAN MP

L1	L2	L3	L4	L5
0	7.6E-5	------	------	------
1	.00215			
2	.02438			
3	.13818			
4	.3915			
5	.44371			
------	------			

Excel

	A	B
1	x	P(x)
2	0	7.594E-05
3	1	0.0021516
4	2	0.0243844
5	3	0.1381781
6	4	0.3915047
7	5	0.4437053

Minitab

x	P(X = x)
0	0.000076
1	0.002152
2	0.024384
3	0.138178
4	0.391505
5	0.443705

Example: Overtime Rule in Football (1 of 4)

We previously noted that between 1974 and 2011, there were 460 NFL football games decided in overtime, and 252 of them were won by the team that won the overtime coin toss. Is the result of 252 wins in the 460 games equivalent to random chance, or is 252 wins **significantly high?** We can answer that question by finding the probability of 252 wins or more in 460 games, assuming that wins and losses are equally likely.

Example: Overtime Rule in Football (2 of 4)

Solution

Using the notation for binomial probabilities, we have $n = 460$, $p = 0.5$, $q = 0.5$, and we want to find the sum of all probabilities for each value of x from 252 through 460. The formula is not practical here, because we would need to apply it 209 times—we don’t want to go there. Table A-1 (Binomial Probabilities) doesn’t apply because $n = 460$, which is way beyond the scope of that table. Instead, we wisely choose to use technology.

Example: Overtime Rule in Football (3 of 4)

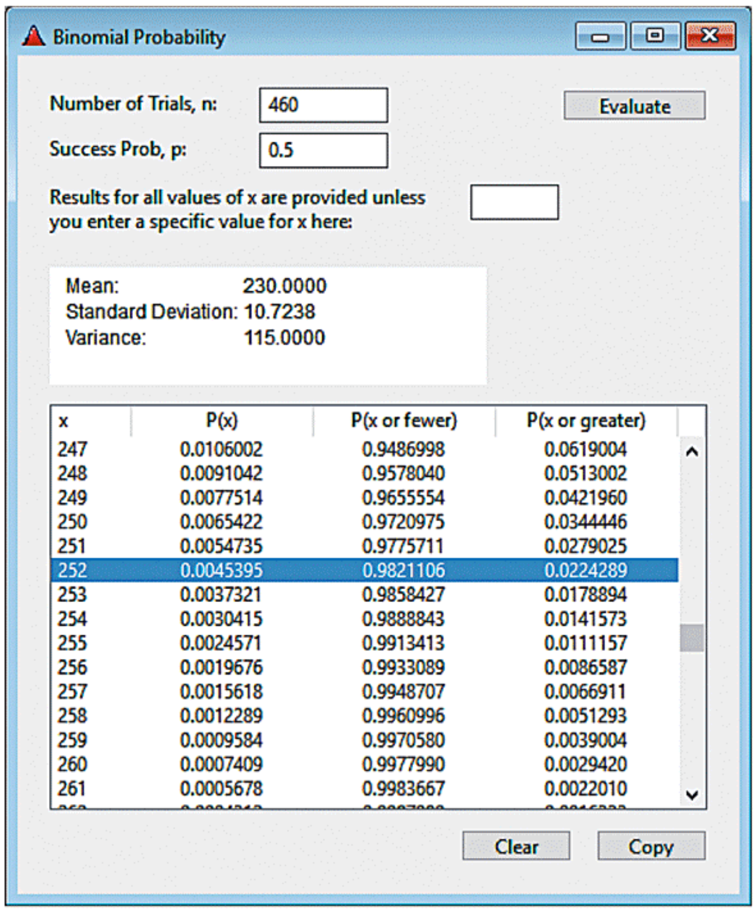
Solution

The Statdisk display on the next page shows that the probability of 252 or more wins in 460 overtime games is 0.0224 (rounded), which is low (such as less than 0.05). This shows that it is unlikely that we would get 252 or more wins by chance. If we effectively rule out chance, we are left with the more reasonable explanation that the team winning the overtime coin toss has a better chance of winning the game.

Example: Overtime Rule in Football (4 of 4)

Solution

Statdisk



x	P(x)	P(x or fewer)	P(x or greater)
247	0.0106002	0.9486998	0.0619004
248	0.0091042	0.9578040	0.0513002
249	0.0077514	0.9655554	0.0421960
250	0.0065422	0.9720975	0.0344446
251	0.0054735	0.9775711	0.0279025
252	0.0045395	0.9821106	0.0224289
253	0.0037321	0.9858427	0.0178894
254	0.0030415	0.9888843	0.0141573
255	0.0024571	0.9913413	0.0111157
256	0.0019676	0.9933089	0.0086587
257	0.0015618	0.9948707	0.0066911
258	0.0012289	0.9960996	0.0051293
259	0.0009584	0.9970580	0.0039004
260	0.0007409	0.9977990	0.0029420
261	0.0005678	0.9983667	0.0022010

Methods for Finding Binomial Probabilities (4 of 4)

Method 3: Using Table A-1 in Appendix A

This method can be skipped if technology is available. Table A-1 in Appendix A lists binomial probabilities for select values of n and p. It cannot be used if $n > 8$ or if the probability p is not one of the 13 values included in the table.

To use the table of binomial probabilities, we must first locate n and the desired corresponding value of x. At this stage, one row of numbers should be isolated. Now align that row with the desired probability of p by using the column across the top. The isolated number represents the desired probability. A very small probability, such as 0.000064, is indicated by 0+.

Example: Devil of a Problem (1 of 3)

Based on a Harris poll, 60% of adults believe in the devil. Assuming that we randomly select five adults, use Table A-1 to find the following:

a. The probability that exactly three of the five adults believe in the devil

b. The probability that the number of adults who believe in the devil is at least two

Example: Devil of a Problem (2 of 3)

Solution

a. The following excerpt from the table shows that when $n = 5$ and $p = 0.6$, the probability for $x = 3$ is given by $P(3) = 0.346$.

TABLE A-1

Binomial Probabilities

n	*x*	.01
5	0	.951
	1	.048
	2	.001
	3	0+
	4	0+
	5	0+

	p	
.50	.60	.70
.031	.010	.002
.156	.077	.028
.313	.230	.132
.313	.346	.309
.156	.259	.360
.031	.078	.168

x	*P*(*x*)
0	.010
1	.077
2	.230
3	.346
4	.259
5	.078

Example: Devil of a Problem (3 of 3)

Solution

b. The phrase “at least two” successes means that the number of successes is 2 or 3 or 4 or 5.

$$
\begin{aligned}
P(\text{at least 2 believe in the devil}) &= P(2 \text{ or } 3 \text{ or } 4 \text{ or } 5) \\
&= P(2) + P(3) + P(4) + P(5) \\
&= 0.230 + 0.346 + 0.259 + 0.078 \\
&= 0.913
\end{aligned}
$$

Using Mean and Standard Deviation for Critical Thinking

For Binomial Distributions

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard Deviation: $\sigma = \sqrt{npq}$

Range Rule of Thumb

Significantly low values $\leq (\mu - 2\sigma)$

Significantly high values $\geq (\mu + 2\sigma)$

Values not significant: Between $(\mu - 2\sigma)$ **and** $(\mu + 2\sigma)$

Example: Using Parameters to Determine Significance (1 of 4)

A previous example involved n = 460 overtime wins in NFL football games. We get p = 0.5 and q = 0.5 by assuming that winning the overtime coin toss does not provide an advantage, so both teams have the same 0.5 chance of winning the game in overtime.

a. Find the mean and standard deviation for the number of wins in groups of 460 games.

b. Use the range rule of thumb to find the values separating the numbers of wins that are significantly low or significantly high.

c. Is the result of 252 overtime wins in 460 games significantly high?

Example: Using Parameters to Determine Significance (2 of 4)

Solution

a. With $n = 460$, $p = 0.5$, and $q = 0.5$, previous formulas can be applied as follows:

$$\mu = np = (460)(0.5) = 230.0 \text{ games}$$

$$\sigma = \sqrt{npq} = \sqrt{(460)(0.5)(0.5)} = 10.7 \text{ games}$$

For random groups of 460 overtime games, the mean number of wins is 230.0 games, and the standard deviation is 10.7 games.

Example: Using Parameters to Determine Significance (3 of 4)

Solution

b. The values separating numbers of wins that are significantly low or significantly high are the values that are two standard deviations away from the mean. With $\mu = 230.0$ games and $\sigma = 10.7$ games, we get

$$(\mu - 2\sigma) = 230.0 - 2(10.7) = 208.6 \text{ games}$$

$$(\mu + 2\sigma) = 230.0 + 2(10.7) = 251.4 \text{ games}$$

Example: Using Parameters to Determine Significance (4 of 4)

Solution

Significantly low numbers of wins are 208.6 games or fewer, significantly high numbers of wins are 251.4 games or greater, and values not significant are between 208.6 games and 251.4 games.

c. The result of 252 wins is significantly high because it is greater than the value of 251.4 games found in part (b).

Using Probabilities to Determine When Results Are Significantly High or Low

Significantly high number of successes:

x successes among n trials is **significantly high** if the probability of x or more successes is 0.05 or less. That is, x is a significantly high number of successes if $P(x \text{ or more}) \leq 0.05$.

Significantly low number of successes:

x successes among n trials is **significantly low** if the probability of x or fewer successes is 0.05 or less. That is, x is a significantly low number of successes if $P(x \text{ or fewer}) \leq 0.05$.

Rationale for the Binomial Probability Formula

The number of outcomes with exactly *x* successes among *n* trials

The probability of *x* successes among *n* trials for any one particular order

$$P(x) = \frac{n!}{(x-n)!x!} \cdot p^{x} \cdot q^{n-x}$$