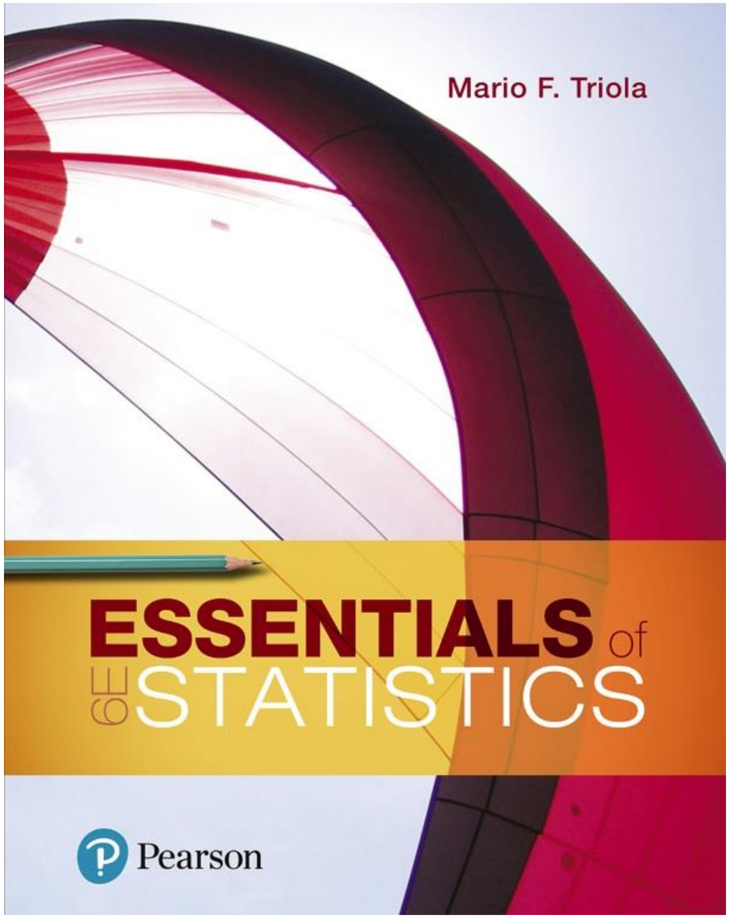
Elementary Statistics

Sixth Edition



Chapter 5

Probability Distributions

Basic Concepts of Probability Distribution

- Probability Distribution
  - A **probability distribution** is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

Basic Concepts of Probability Distribution

- Discrete Random Variable
  - A **discrete random variable** has a collection of values that is finite or countable. (If there are infinitely many values, the number of values is countable if it is possible to count them individually, such as the number of tosses of a coin before getting heads.)

Probability Distribution Requirements

Every probability distribution must satisfy each of the following three requirements.

1. There is a **numerical** (not categorical) random variable *x*, and its number values are associated with corresponding probabilities.
2. $\sum P(x) = 1$ where *x* assumes all possible values. (The sum of all probabilities must be 1, but sums such as 0.999 or 1.001 are acceptable because they result from rounding errors.)
3. $0 \le P(x) \le 1$ for every individual value of the random variable *x*. (That is, each probability value must be between 0 and 1 inclusive.)

Parameters of a Probability Distribution

- **Mean, *μ*,** for a probability distribution

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

- **Variance**, ***σ*²**, for a probability distribution

$$\sigma^2 = \sum \left[(x-\mu)^2 \cdot P(x) \right]$$ (This format is easier to understand.)

- **Standard deviation**, ***σ***, for a probability distribution

$$\sigma = \sqrt{\sum \left[x^2 \cdot P(x) \right] - \mu^2}$$

Expected Value

- Expected Value
  - The **expected value** of a discrete random variable x is denoted by E, and it is the mean value of the outcomes, so $E = \mu$ and E can also be found by evaluating $\sum[x \cdot P(x)]$.

Example: Finding the Mean, Variance, and Standard Deviation (1 of 5)

The table describes the probability distribution for the number of heads when two coins are tossed. Find the mean, variance, and standard deviation for the probability distribution described.

***x*: Number of Heads When Two Coins Are Tossed**	***P*(*x*)**
0	0.25
1	0.50
2	0.25

Example: Finding the Mean, Variance, and Standard Deviation (2 of 5)

x	$P(x)$	$x \cdot P(x)$	$(x - \mu)^2 \cdot P(x)$
0	0.25	$0 \cdot 0.25 = 0.00$	$(0 - 1.0)^2 \cdot 0.25 = 0.25$
1	0.50	$1 \cdot 0.50 = 0.50$	$(1 - 1.0)^2 \cdot 0.50 = 0.00$
2	0.25	$2 \cdot 0.25 = 0.50$	$(2 - 1.0)^2 \cdot 0.25 = 0.25$
Total		1.00 ↑ $\mu = \sum [x \cdot P(x)]$	0.50 ↑ $\sigma^2 = \sum [(x - \mu)^2 \cdot P(x)]$

Solution

The two columns at the left describe the probability distribution. The two columns at the right are for the purposes of the calculations required.

Example: Finding the Mean, Variance, and Standard Deviation (3 of 5)

x	$P(x)$	$x \cdot P(x)$	$(x - \mu)^2 \cdot P(x)$
0	0.25	$0 \cdot 0.25 = 0.00$	$(0 - 1.0)^2 \cdot 0.25 = 0.25$
1	0.50	$1 \cdot 0.50 = 0.50$	$(1 - 1.0)^2 \cdot 0.50 = 0.00$
2	0.25	$2 \cdot 0.25 = 0.50$	$(2 - 1.0)^2 \cdot 0.25 = 0.25$
Total		1.00 ↑ $\mu = \sum [x \cdot P(x)]$	0.50 ↑ $\sigma^2 = \sum [(x - \mu)^2 \cdot P(x)]$

Solution

Mean: $\sum [x \cdot P(x)] = 1.0$

Variance: $\sigma^2 = \sum \left[(x - \mu)^2 \cdot P(x) \right] = 0.5$

Example: Finding the Mean, Variance, and Standard Deviation (4 of 5)

x	$P(x)$	$x \cdot P(x)$	$(x - \mu)^2 \cdot P(x)$
0	0.25	$0 \cdot 0.25 = 0.00$	$(0 - 1.0)^2 \cdot 0.25 = 0.25$
1	0.50	$1 \cdot 0.50 = 0.50$	$(1 - 1.0)^2 \cdot 0.50 = 0.00$
2	0.25	$2 \cdot 0.25 = 0.50$	$(2 - 1.0)^2 \cdot 0.25 = 0.25$
Total		1.00 ↑ $\mu = \sum [x \cdot P(x)]$	0.50 ↑ $\sigma^2 = \sum [(x - \mu)^2 \cdot P(x)]$

Solution

The standard deviation is the square root of the variance, so

$$\text{Standard deviation: } \sigma = \sqrt{0.5}$$
$$= 0.707107 \approx 0.7$$

Example: Finding the Mean, Variance, and Standard Deviation (5 of 5)

Interpretation

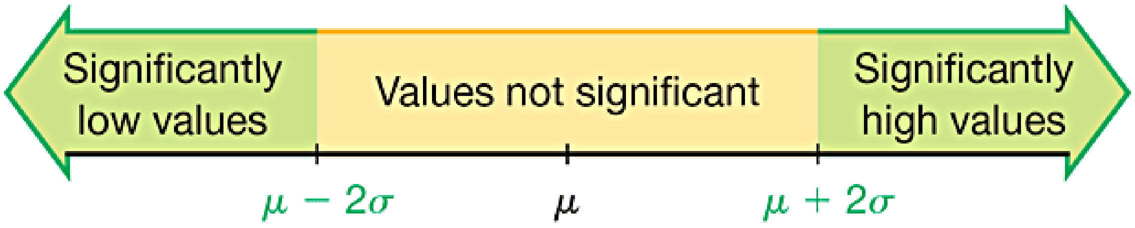
When tossing two coins, the mean number of heads is 1.0 head, the variance is 0.50 $heads^2$, and the standard deviation is 0.7 head.

Also, the expected value for the number of heads when two coins are tossed is 1.0 head, which is the same value as the mean. If we were to collect data on a large number of trials with two coins tossed in each trial, we expect to get a mean of 1.0 head.

Identifying Significant Results with the Range Rule of Thumb

Range Rule of Thumb for Identifying Significant Values

- **Significantly low** values are ($\mu - 2\sigma$) or lower.
- **Significantly high** values are ($\mu + 2\sigma$) or higher.
- **Values not significant**: Between ($\mu - 2\sigma$) and ($\mu + 2\sigma$).



Example: Identifying Significant Results with the Range Rule of Thumb (1 of 3)

We found that when tossing two coins, the mean number of heads is $\mu = 1.0$ head and the standard deviation is $\sigma = 0.7$ head. Use those results and the range rule of thumb to determine whether 2 heads is a significantly high number of heads.

Example: Identifying Significant Results with the Range Rule of Thumb (2 of 3)

Solution

Using the range rule of thumb, the outcome of 2 heads is significantly high if it is greater than or equal to $(\mu + 2\sigma)$. With $\mu = 1.0$ head $\sigma = 0.7$ head, we get

$$(\mu + 2\sigma) = 1 + 2(0.7) = 2.4 \text{ heads}$$

Significantly high numbers of heads are 2.4 and above.

Example: Identifying Significant Results with the Range Rule of Thumb (3 of 3)

Interpretation

Based on these results, we conclude that 2 heads is not a significantly high number of heads (because 2 is not greater than or equal to 2.4).

Identifying Significant Results with Probabilities: (1 of 2)

- Significantly high number of successes:
  - x successes among n trials is a **significantly high** number of successes 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 **or more**) ≤ 0.05.

The value 0.05 is not absolutely rigid. Other values, such as 0.01, could be used to distinguish between results that are significant and those that are not significant.

Identifying Significant Results with Probabilities: (2 of 2)

- Significantly low number of successes:
  - x successes among n trials is a **significantly low** number of successes 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 **or fewer**) ≤ 0.05.

The value 0.05 is not absolutely rigid. Other values, such as 0.01, could be used to distinguish between results that are significant and those that are not significant.

The Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular outcome is very small and the outcome occurs **significantly less than or significantly greater than** what we expect with that assumption, we conclude that the assumption is probably not correct.

Binomial Probability Distribution

- 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.
    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

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

Mean and Standard Deviation

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** $(\mu + 2\sigma)$