

# **Advanced Statistics**

**Dr. Syed Faisal Bukhari**

**Associate Professor**

**Department of Data Science**

**Faculty of Computing and Information Technology**

**University of the Punjab**

# Textbooks

- ❑ **Probability & Statistics for Engineers & Scientists,**  
Ninth Edition, Ronald E. Walpole, Raymond H.  
Myer

# References

Readings for these lecture notes:

- ❑ **Probability & Statistics for Engineers & Scientists**, Ninth edition, Ronald E. Walpole, Raymond H. Myer

These notes contain material from the above book.

# Mathematical Expectation

If **two coins** are tossed **16 times** and  **$X$**  is the number of **heads** that occur per toss, then the values of  **$X$**  are **0, 1, and 2**.

Suppose that the experiment yields **no heads, one head,** and **two heads** a total of **4, 7, and 5** times, respectively.

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06$$

- ❑ This is an **average value** of the data and yet it is not a possible outcome of  **$\{0, 1, 2\}$** .
- ❑ Hence, an average is **not necessarily a possible** outcome for the experiment
- ❑ For instance, a **salesman's average monthly** income is **not likely** to be equal to any of his **monthly paychecks**.

x	f	fx
0	4	0
1	7	7
2	5	10
	$\sum f = 16$	$\sum fx = 17$
$\bar{x} = \frac{17}{16}$ =1.06		

x	P(X)	xP(X)
0	$\frac{4}{16}$	0
1	$\frac{7}{16}$	$\frac{7}{16}$
2	$\frac{5}{16}$	$\frac{10}{16}$
	$\sum P(X) = \frac{16}{16} = 1$	$\sum xP(X) = \frac{17}{16} = 1.0625$

- ❑ The numbers  **$4/16$ ,  $7/16$ , and  $5/16$**  are the fractions of the total tosses resulting in **0, 1, and 2 heads**, respectively. These **fractions** are also the **relative frequencies** for the different values of  $X$  in our experiment.
- ❑ In fact, then, we can **calculate the mean**, or **average**, of a set of data by knowing the distinct values that occur and their **relative frequencies**, **without any knowledge of the total number of observations in our set of data**.

□ Therefore, if **4/16, or 1/4**, of the tosses result in **no heads**, **7/16** of the tosses result in **one head**, and **5/16** of the tosses **result in two heads**, the **mean number of heads per toss** would be **1.06** no matter whether the total number of **tosses were 16, 1000, or even 10,000**.



- ❑ This method of **relative frequencies** is used to calculate the **average number of heads** per toss of two coins that we might **expect in the long run**.
- ❑ We shall refer to this average value as the **mean of the random variable  $X$**  or the **mean of the probability distribution of  $X$**  and write it as  $\mu_x$  or simply as  $\mu$  when it is clear to which random variable we refer.

- It is also common among statisticians to refer to this **mean** as the **mathematical expectation**, or **the expected value** of the **random variable  $X$** , and denote it as  **$E(X)$** .

# Mean or Expected value of X

**Theorem:** Let  $X$  be a random variable with probability distribution  $f(x)$ . The **mean**, or **expected value**, of  $X$  is

$$\mu = E(X) = \sum_x xf(x)$$

if  **$X$  is discrete**, and

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

if  **$X$  is continuous**

**Example :** A lot containing **7 components** is sampled by a quality inspector; the lot contains **4 good components** and **3 defective components**. A **sample of 3** is taken by the inspector. Find the **expected value** of the number of **good components** in this sample.

$$N = 7$$

$$n = 3$$

$$k = 4$$

$$P(X = x) = h(x; N, n, k) = \binom{k}{x} \binom{N-k}{n-x} / \binom{N}{n}, \max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}$$

Let  $X$  represent the number of good components in the sample.

$$\max\{0, n-(N-k)\} = \max\{0, 3-(7-4)\} = \max(0, 0) = 0$$

$$\min\{n, k\} = \min(3, 4) = 3$$

$x$	$P(X = x)$	$x P(X)$
0	$\frac{1}{35}$	0
1	$\frac{12}{35}$	$\frac{12}{35}$
2	$\frac{18}{35}$	$\frac{36}{35}$
3	$\frac{4}{35}$	$\frac{12}{35}$
	$\sum P(X) = \frac{16}{16} = 1$	$\sum xP(X) = \frac{60}{35} = 1.7143$

## Cont.

- Thus, if a sample of **size 3 is selected** at random over and over again from a lot of **4 good components** and **3 defective** components, it will contain, on average, **1.7 good components**.

**Example:** A salesperson for a medical device company has **two appointments** on a given day.

At the first appointment, he believes that he has a **70% chance** to make the deal, from which he can **earn \$1000** commission if successful. On the other hand, he thinks he only has a **40% chance** to make the deal at the second appointment, from which, if successful, he can **make \$1500**. What is his **expected commission** based on his own probability belief? Assume that the appointment results are **independent** of each other.



## First appointment:

Probability of **commission** = **0.70**, **Commission** = **1000**

Probability of **no commission** =  $1 - 0.70 = 0.30$ , **Commission** = **0**

## Second appointment:

Probability of **commission** = **0.40**, **Commission** = **1500**

Probability of **no commission** =  $1 - 0.40 = 0.60$ , **Commission** = **0**

**Since** appointment results are **independent**.

Let X denotes total commission from appointment 1 and appointment 2

X	$P(X = x)$	$xP(X)$
$0 + 0 = 0$	$(0.30)(0.60) = 0.18$	0
$1000 + 0 = 1000$	$(0.70)(0.60) = 0.42$	420
$0 + 1500 = 1500$	$(0.30)(0.40) = 0.12$	180
$1000 + 1500 = 2500$	$(0.70)(0.40) = 0.28$	700
	$\sum P(X) = 1$	$\sum xP(X)$ $= \$1300$

**Example :** Let  $X$  be the random variable that denotes the life in hours of a certain electronic device. The **probability density function** is

$$f(x) = \begin{cases} \frac{20,000}{x^3}, & x > 100, \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected life of this type of device.

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

if **X** is continuous

$$\mu = E(X) = \int_{100}^{\infty} \mathbf{x} \times \frac{20,000}{x^3} dx$$

$$= \int_{100}^{\infty} \frac{20,000}{\mathbf{x^2}} dx$$

$$= 20,000 \int_{100}^{\infty} x^{-2} dx$$

$$= -20,000 [x^{-1}]_{100}^{\infty} = -20,000\left(\frac{1}{\infty} - \frac{1}{100}\right)$$

$$= 200$$