Probability and Statistics - 3130006

Unit-3: Basic Statistics

July 23, 2020

Contents

1 Introduction

discrete random variable is described by probability function or probability mass function. Similarly, a continuous random variable is described by its probability density function. Instead of a function, a more compact description can be made by a few parameters, known as statistical measures, that are representative of the distribution. In descriptive statistics, statistical measures are used to summarize a set of observations in order to communicate the information as simply as possible. The observations are described in
(i) a measure of location or central tendency, such as arithmetic mean
(ii) a measure of statistical dispersion like standard deviation
(iii) a measure of the shape of the distribution like skewness or kurtosis
(iv) if more than one variable is measured, a measure of statistical dependence such as correlation coefficient.

2 Measures of Central Tendency

. In statistics, a central tendency or measure of central tendency is a central or typical value of a probability distribution. It is also called a center or location of the distribution. Measures of central tendency are often called averages. An average is a single value which can be taken as a representative of the whole distribution. There are five types of measures of central tendency or averages which are commonly used.
(i) Arithmetic mean or mean or expectation
(ii) Median
(iii) Mode
(iv) Geometric mean
(v) Harmonic mean

Mean

Mean The mean or average value (μ) of the probability distribution of a discrete random variable X is called as expectation and is denoted by $E(X)$.
$\mu = E(X) = \sum_{i=1}^{\infty} x_i p(x_i) = \sum xp(x)$ where $p(x)$ is the probability mass function of the discrete random variable X.

Median

Median The median is the point which divides the entire distribution into two equal parts.

If X is a random variable, the value of $X = x$ for which the cumulative distribution function $F(x) = \frac{1}{2}$ is called the median of X. For a discrete random variable X, if there exits no x such that $F(x) = \frac{1}{2}$ then the median M of probability distribution is given by $M = \frac{1}{2}(x_k + x_{k+1})$
where $F(x_k) < \frac{1}{2}$ and $F(x_{k+1}) > \frac{1}{2}$ and x_k and x_{k+1} are two consecutive values of X.

Geometric Mean

The geometric mean G of a random variable X is defined by $logG = E(log(x))$. The geometric mean of the probability distribution of a discrete random variable X is given by
$logG = \sum_{i=1}^{\infty}(logx_i)p(x_i) = \sum(logx)p(x)$
where $p(x)$ is the probability mass function of the discrete random variable X.

Harmonic Mean

The harmonic mean of a random variable X is defined by
$\frac{1}{H} = E(\frac{1}{X})$. The harmonic mean of the probability distribution of a discrete random H variable X is given by
$\frac{1}{H} = \sum_{i=1}^{\infty}(\frac{1}{x_i})p(x_i) = \sum(\frac{1}{x})p(x)$
where $p(x)$ is the probability mass function of the discrete random variable X.

3 Measures of Dispersion

A measure of central tendency is a representative value of the random variable. But it is important to know how the values are clustered around or scattered away from the measure of central tendency. The property of the random variable or its distribution by which its values are clustered around or scattering away from the central value is called dispersion. There are three types of measures of dispersion which are commonly used.
(i) Quartile Deviation
(ii) Mean Deviation
(iii) Standard Deviation

Quartile Deviation

Quartile deviation or semi-inter quartile range of the prob- ability distribution of a discrete random variable X is given by $Q = \frac{1}{2}(Q_3 - Q_1)$
where Q_1 and Q_3 are the first and third quartiles of the distribution respectively.

Mean Deviation

Mean deviation of the probability distribution of a discrete random variable X is given by

$MD = \sum |x - \mu| p(x)$

where $p(x)$ is the probability mass function of the discrete random variable X. **Standard Deviation**

Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by the Greek letter σ.

$SD = \sigma = \sqrt{E(X^2) - [E(X)]^2} = \sqrt{E(X^2) - \mu^2}$

Variance Deviation

Variance Variance characterizes the variability in the distributions since two distributions with same mean can still have different dispersion of data about their means. Variance of the probability distribution of a discrete random variable X is given by

$Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$

Example 1 *A random variable X has the following distribution:*

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Find (i) mean (ii) variance, and (iii) $P(1 < x < 6)$.

Solution:

(i)

$$\begin{aligned} Mean = \mu = &\sum xp(x) \\ =&1\left(\frac{1}{36}\right) + 2\left(\frac{3}{36}\right) + 3\left(\frac{5}{36}\right) + 4\left(\frac{7}{36}\right) + 5\left(\frac{9}{36}\right) + 6\left(\frac{11}{36}\right) \\ =&\frac{161}{36} \\ =&4.47 \end{aligned}$$

(ii)

$$
\begin{aligned}
Variance = \sigma^2 = & \sum x^2 p(x) - \mu^2 \\
= & 1\left(\frac{1}{36}\right) + 4\left(\frac{3}{36}\right) + 9\left(\frac{5}{36}\right) + 16\left(\frac{7}{36}\right) + 25\left(\frac{9}{36}\right) + 36\left(\frac{11}{36}\right) - (4.47)^2 \\
= & 1.99
\end{aligned}
$$

(iii)

$$
\begin{aligned}
P(1 < X < 6) = & P(X=2) + P(X=3) + P(X=4) + P(X=5) \\
= & \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36} \\
= & \frac{24}{36} \\
= & 0.67
\end{aligned}
$$

Example 2 *The probability distribution of a random variable X is given bellow. Find (i)$E(X)$ (ii) Var(X) (iii) $E(2X-3)$, and (iv) Var$(2X-3)$*

X	-2	-1	0	1	2
$P(X=x)$	0.2	0.1	0.3	0.3	0.1

Solution:

(i)

$$
\begin{aligned}
E(X) = & \sum xp(x) \\
= & -2(0.2) - 1(0.1) + 0(0.3) + 1(0.3) + 2(0.1) \\
= & 0
\end{aligned}
$$

(ii)

$$
\begin{aligned}
Var(X) = & \sum x^2 p(x) - \mu^2 \\
= & 4(0.2) + 1(0.1) + 0(0.3) + 1(0.3) + 4(0.1) - 0 \\
= & 1.6
\end{aligned}
$$

(iii)

$$E(2X-3) = 2E(X) - 3$$
$$= 0 - 3$$
$$= -3$$

(iv)

$$Var(2X-3) = (2)^2 Var(X) \quad (\because Var(kX) = k^2 Var(X) \text{ and } Var(X+k) = Var(X))$$
$$= 4(1.6)$$
$$= 6.4$$

Example 3 *A machine produces an average of 500 items during the first week of the month and on average of 400 items during the last week of the month, the probability for these being 0.68 and 0.32 respectively. Determine the expected value of the production.*

Solution:

Let X be the random variable which denotes the items produced by the machine. The probability distribution is

X	500	400
$P(X=x)$	0.68	0.32

Expected value of the production

$$E(X) = \sum xp(x)$$
$$= 500(0.68) + 400(0.32)$$
$$= 468$$

Example 4 *A discrete random variable has the probability mass function given below:*

X	-2	-1	0	1	2	3
$P(X=x)$	0.2	k	0.1	2k	0.1	2k

Find k, mean, and variance.

Solution:

Since $P(X = x)$ is a probability mass function,

$$\begin{aligned}\sum P(X = x) &= 1\\ 0,2 + k + 0.1 + 2k + 0.1 + 2k &= 1\\ 5k + 0.4 &= 1\\ 5k &= 0.6\\ k &= \frac{0.6}{5}\\ k &= \frac{3}{25}\end{aligned}$$

Hence, the probability distribution is

X	-2	-1	0	1	2	3
$P(X = x)$	$\frac{2}{10}$	$\frac{3}{25}$	$\frac{1}{10}$	$\frac{6}{25}$	$\frac{1}{10}$	$\frac{6}{25}$

$$\begin{aligned}Mean = E(X) = \sum xp(x)\\ &= (-2)\left(\frac{2}{10}\right) + (-1)\left(\frac{3}{25}\right) + (0)\left(\frac{1}{10}\right) + (1)\left(\frac{6}{25}\right) + (2)\left(\frac{1}{10}\right) + (3)\left(\frac{6}{25}\right)\\ &= \frac{6}{25}\end{aligned}$$

$$\begin{aligned}Variance = Var(X) = \sum x^2p(x) - [E(X)]^2\\ &= (4)\left(\frac{2}{10}\right) + (1)\left(\frac{3}{25}\right) + 0 + (1)\left(\frac{6}{25}\right) + (4)\left(\frac{1}{10}\right) + (9)\left(\frac{6}{25}\right) - \left(\frac{6}{25}\right)^2\\ &= \frac{73}{250} - \frac{36}{625}\\ &= \frac{293}{625}\end{aligned}$$