# Introduction to Statistics (MAT 283)

Dipti Dubey

Department of Mathematics
Shiv Nadar University

# Table of Contents

EXAMPLE: A lot of 8 TV sets includes 3 that are defective. If 4 of the sets are chosen at random for shipment to a hotel, how many defective sets can they expect?

Variance of Random Variables:

Let $X$ be a random variable with mean $\mu_X$ and probability density function $f$. The variance, $Var(X)$ of the random variable $X$ is defined as

$$Var(X) = E((X - \mu_X)^2).$$

It is also denoted as $\sigma_x^2$.

The positive square root of the variance is called the standard deviation of the random variable X. Like variance, the standard deviation also measures the spread.

**Theorem**: If $X$ is a random variable with mean $E(X)$ and variance $Var(X)$, then

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

**Proof.** Let $\mu_X = E(X)$, we have

$$\begin{aligned} Var(X) &= E[(X - \mu_X)^2)] \\ &= E[X^2 - 2\mu_X X + X^2] \\ &= E(X^2) - 2\mu_X E(X) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned}$$

**Theorem**: If $Var(X)$ exists and $Y = a + bX$, then

$$Var(Y) = b^2 Var(X).$$

**Proof**: We have

$$\begin{aligned} Var(a+bX) &= E[(a+bX-E(a+bX))^2] \\ &= E[(a+bX-a-b(E(X)))^2] \\ &= E[b^2(X-(E(X))^2] \\ &= b^2 Var(X). \end{aligned}$$

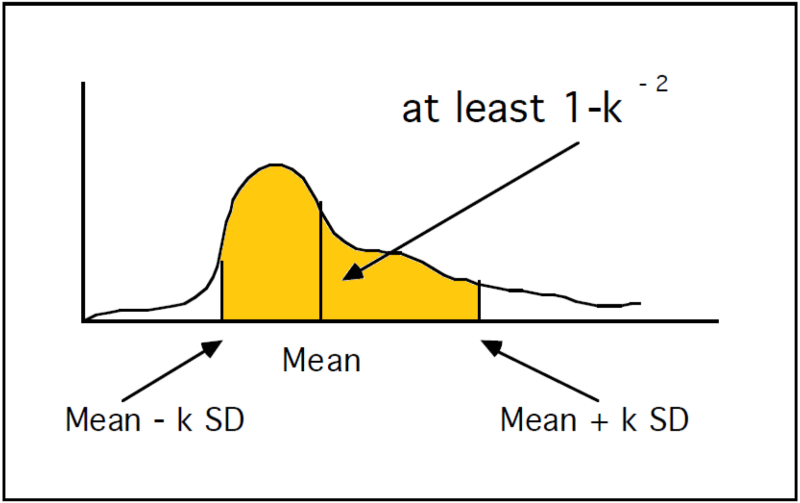
# Table of Contents

## Chebyshev's Inequality:

Let $X$ be a random variable with probability density function $f$. If $\mu$ and $\sigma > 0$ are the mean and standard deviation of $X$, then

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

for any nonzero real positive constant $k$.


at least $1-k^{-2}$
Mean
Mean - k SD
Mean + k SD

**Proof**.We assume that the random variable $X$ is continuous. If $X$ is not continuous we replace the integral by summation in the following proof. From the definition of variance, we have the following:

$$\begin{aligned}
\sigma^2 &= E[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma}(x-\mu)^2 f(x)dx \\
&\quad + \int_{\mu+k\sigma}^{\infty}(x-\mu)^2 f(x)dx
\end{aligned}$$

Since the integrand $(x-\mu)^2 f(x)$ is nonnegative, we get

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx.$$

If $x \in (-\infty, \mu - k\sigma)$,

$$\begin{aligned} x &\leq \mu - k\sigma \\ \implies k\sigma &\leq \mu - x \\ \implies k^2\sigma^2 &\leq (\mu - x)^2. \end{aligned} \tag{1}$$

If $x \in (\mu + k\sigma, \infty)$,

$$\begin{aligned} x &\geq \mu + k\sigma \\ \implies -k\sigma &\leq \mu - x \\ \implies k^2\sigma^2 &\leq (\mu - x)^2. \end{aligned} \tag{2}$$

Thus if $x \notin (\mu - k\sigma, \mu + k\sigma)$, by (1) and (2) we get $k^2\sigma^2 \leq (\mu - x)^2$. Therefore,

$$\sigma^2 \geq k^2\sigma^2 \int_{-\infty}^{\mu - k\sigma} f(x)dx + k^2\sigma^2 \int_{\mu + k\sigma}^{\infty} f(x)dx$$

$$
\begin{aligned}
&\frac{1}{k^2} \geq P(X \leq \mu - k\sigma) + P(X \geq \mu + k\sigma) \\
\implies &\frac{1}{k^2} \geq P(|X - \mu| \geq k\sigma) \\
\implies &\frac{1}{k^2} \geq 1 - P(|X - \mu| < k\sigma)
\end{aligned}
$$

which is

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

This completes the proof.

EXAMPLE: Use Chebyshev's inequality to find what percent of the values will fall between 123 and 179 for a data set with mean of 151 and standard deviation of 14.

Note that $k = 2$ here and $1 - \frac{1}{k^2} = \frac{3}{4}$. This implies that 75% of the data values are between 123 and 179.

# Table of Contents

**Moment Generating Function:** Let $X$ be a random variable with probability density function $f$. A real valued function $M : \mathbb{R} \to \mathbb{R}$ defined by

$$M(t) = E(e^{tX})$$

is called the moment generating function of $X$ if this expected value exists for all $t$ in the interval $-h < t < h$ for some $h > 0$.

**Moment Generating Function:** Let $X$ be a random variable with probability density function $f$. A real valued function $M : \mathbb{R} \to \mathbb{R}$ defined by

$$M(t) = E(e^{tX})$$

is called the moment generating function of $X$ if this expected value exists for all $t$ in the interval $-h < t < h$ for some $h > 0$.

Using the definition of expected value of a random variable, we obtain

$$M(t) = \begin{cases} \sum_{x \in R_X} e^{tx} f(x) & \text{if X is discrete} \\ \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if X is continuous.} \end{cases}$$

Example: Let X have the PDF

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{if} \quad x > 0 \\ \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} e^{-x/2} dx \\ &= \frac{1}{2} \int_0^\infty e^{(t-\frac{1}{2})x} dx \\ &= \frac{1}{1-2t}, \qquad t < \frac{1}{2}. \end{aligned}$$

$$[Use : \textstyle\int_0^\infty e^{-ax} dx = \frac{1}{a}, \quad a > 0.]$$

To explain why we refer to this function as a "moment-generating" function, let us substitute for $e^{tx}$ its Maclaurins series expansion,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots + \frac{t^rx^r}{r!} + \ldots$$

For discrete case, thus we get

$$M(t) = \sum_x [1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots + \frac{t^rx^r}{r!} + \ldots] f(x)$$

$$= \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \ldots \frac{t^r}{r!}\sum_x x^r f(x) + \ldots$$

$$= 1 + E(X)t + E(X^2)\frac{t^2}{2!} + \ldots + E(X^r)\frac{t^r}{r!} + \ldots$$

Theorem: $\frac{d^r M(t)}{dt^r}\Big|_{t=0} = E(X^r)$.

Example: Let X have the PDF

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Recall

$$M(t) = \frac{1}{1-2t}, \qquad t < \frac{1}{2}.$$

Then

$$M'(t) = \frac{2}{(1-2t)^2}, \qquad M''(t) = \frac{8}{(1-2t)^3}, \qquad t < \frac{1}{2}.$$

and hence

$$E(X) = 2, \; E(X^2) = 8, \text{ and } \; Var(X) = 4.$$