# IX. Joint Probability Distributions

Instructor: **Yanlin Qi**

Institute of Transportation Studies
Department of Statistics
University of California, Davis

# Today



- Review of Joint Probability Distributions
    - Examples
- Linear Functions
- Numerical Summaries

# Joint Probability Distributions - Review

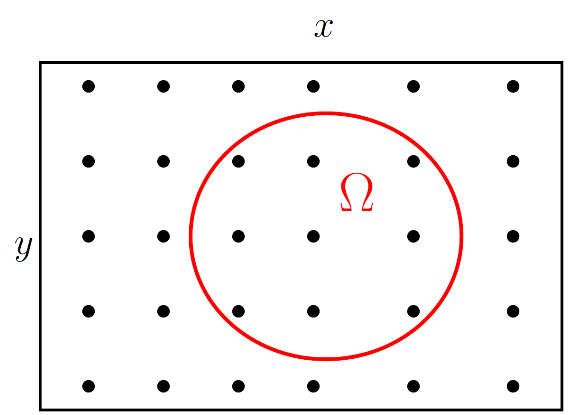
# Joint PMFs for Discrete RVs

- Let X and Y be two discrete random variables on the same sample space. We define their joint pmf as a function $f : \mathrm{R}^2 \to \mathrm{R}$ with

$$f(x,y) = \begin{cases} P(X = x, Y = y), & \text{for all feasible pairs } (x,y) \\ 0, & \text{otherwise} \end{cases}$$

$$P((X,Y) \in \Omega) = \sum_{(x,y)\in\Omega} f(x,y)$$

Then for any region $\Omega \subset \mathrm{R}^2$



# Properties of Joint Probability Functions

- Any joint PMF $f(x, y) : \mathrm{R}^2 \rightarrow \mathrm{R}$, must satisfy (and vice versa)

1. $f(x, y) \geq \mathbf{0}$ for all $x, y \in \mathbf{R}$

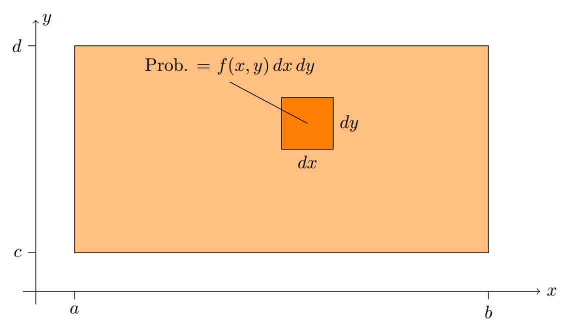
2. $$\sum_x \sum_y f_{XY}(x, y) = 1$$

3. $f_{XY}(x, y) = P(X = x, Y = y)$

# joint PDFs for Continuous RVs



- **Definition** The joint PDF $f(x, y)$ describes the likelihood of the continuous random variables X and Y taking on specific values x and y.
  - X takes values in [a, b]
  - Y takes values in [c, d]
  - (X, Y) takes values in [a, b] × [c, d]
  - Joint PDF: f(x, y)
  - $f(x, y)dxdy$ is the probability of being in the small square.



# Joint Probability of Continuous RVS

- The **joint probability density function** of the continuous random variables X and Y, is denoted as $f_{XY}(x, y)$ and satisfies

1. $f_{XY}(x, y) \geq 0 \quad \text{for all } x, y$

2. $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{XY}(x, y)\,dx\,dy = 1$

3. For any region *R* of two-dimensional space

$$P([X, Y] \in R) = \iint\limits_{R} f_{XY}(x, y)\,dx\,dy$$

# Marginal PMFs



- **From joint to marginal**
  - **Marginal PMFs:** Individual PMFs $f_X(x), f_Y(y)$.
- *Proposition.* Let $f(x, y)$ be the joint PMF for X, Y . Then

$$f_X(x) = \sum_y f(x,y), \quad \text{and} \quad f_Y(y) = \sum_x f(x,y).$$

# Marginal PDFs



- The **marginal PDF** of a random variable is obtained by integrating the joint PDF over the other variable(s).
- Marginal PDF of $X$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\,dy = \int_{R_y} f_{XY}(x, y)\,dy$$

# Conditional Distribution Function

## Definition

- The conditional distribution function of $Y$ given $X = x$ (with $f_X(x) \neq 0$)) is defined as

$$f(\underbrace{y}_{\text{variable}} \mid \underbrace{x}_{\text{fixed}}) = \frac{f(x, y)}{f_X(x)}, \quad \text{for all feasible } y$$

## Independence

- 1. Two discrete RVs $X$, $Y$ are independent if

$$f(x, y) = f_X(x) f_Y(y), \quad \text{for all } x, y$$

- $X$, $Y$ are independent if all conditional distributions of $Y$ are identical to its marginal distribution:

$$f(y \mid x) = f_Y(y), \quad \text{for all } x, y$$

# Independence

- 1. Two discrete RVs $X$, $Y$ are independent if

$$f(x,y) = f_X(x) f_Y(y), \quad \text{for all } x, y$$

- $X$, $Y$ are independent if all conditional distributions of $Y$ are identical to its marginal distribution:

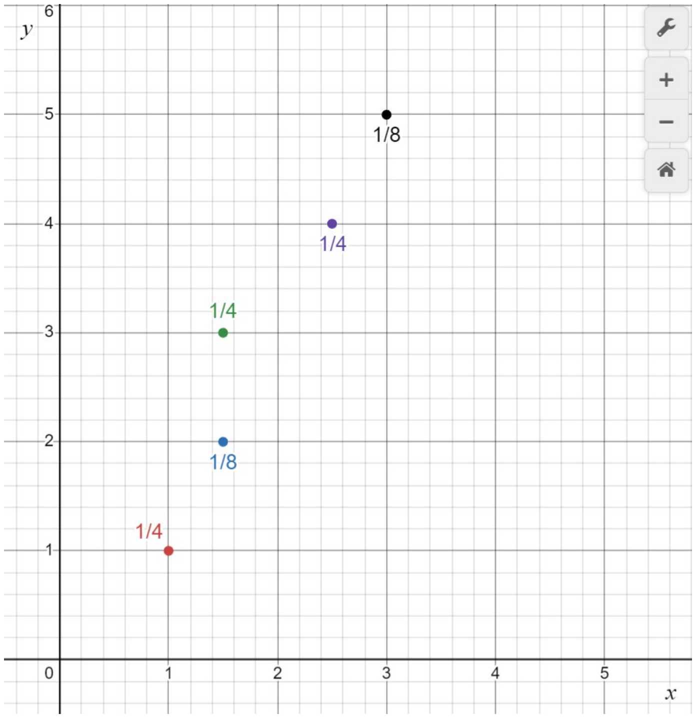
$$f(y \mid x) = f_Y(y), \quad \text{for all } x, y$$

# Example 1



- Show that the following function satisfies the properties of a joint probability mass function

| $x$ | $y$ | $f_{XY}(x, y)$ |
|---|---|---|
| 1 | 1 | 1/4 |
| 1.5 | 2 | 1/8 |
| 1.5 | 3 | 1/4 |
| 2.5 | 4 | 1/4 |
| 3 | 5 | 1/8 |




Determine the following:

(a) $P(X < 2.5, Y < 3)$

(b) $P(X < 2.5)$

(c) $P(Y < 3)$

(d) $P(X > 1.8, Y > 4.7)$

(e) $E(X)$, $E(Y)$, $V(X)$, and $V(Y)$.

(f) Marginal probability distribution of the random variable $X$

(g) Conditional probability distribution of $Y$ given that $X = 1.5$

# Example 1



- First, f(x,y) >= 0. Le R denote the range of (X,Y)
- Then, $\sum_R f(x,y) = \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} = 1$
- a) P(X < 2.5, Y < 3) =

  f(1.5,2)+f(1,1) = 1/8+1/4=3/8
- b) P(X < 2.5) =

  f (1.5, 2) + f (1.5, 3) +f(1,1)= 1/8 + 1/4+1/4 = 5/8
- c) P(Y < 3) =

  f (1.5, 2)+f(1,1) = 1/8+1/4=3/8
- d) P(X > 1.8, Y > 4.7) =

  f ( 3, 5) = 1/8

# Example 1



- e) $E(X) =$

  $1(1/4)+ 1.5(3/8) + 2.5(1/4) + 3(1/8) = 1.8125$

  $E(Y) =$

  $1(1/4)+2(1/8) + 3(1/4) + 4(1/4) + 5(1/8) = 2.875$

  $V(X) = E(X^2)-[E(X)]^2 =$

  $[1^2(1/4)+1.5^2(3/8)+2.5^2(1/4)+3^2(1/8)]-1.8125^2 =0.4961$

  $V(Y) = E(Y^2)-[E(Y)]^2 =$

  $[1^2(1/4)+2^2(1/8)+3^2(1/4)+4^2(1/4)+5^2(1/8)]-2.875^2 = 1.8594$

# Example 1

- f) marginal distribution of x

| x | f(x) |
|---|---|
| 1 | ¼ |
| 1.5 | 3/8 |
| 2.5 | ¼ |
| 3 | 1/8 |

- g) Conditional

| y | $f_{Y\|1.5}(y)$ |
|---|---|
| 2 | (1/8)/(3/8)=1/3 |
| 3 | (1/4)/(3/8)=2/3 |

# Example 2



- Determine the value of $c$ that makes the function $f(x,y) = c(x+y)$ a joint probability density function over the range $0 < x < 3$ and $x < y < x+2$
  - a) $P(X < 1, Y < 2)$
  - b) $P(1 < X < 2)$
  - c) Expression for $P(1 < Y)$
  - d) Expression for $P(X < 2, Y < 2)$
  - e) Expression for $E(X)$
  - f) Marginal probability distribution of X

# Example 2

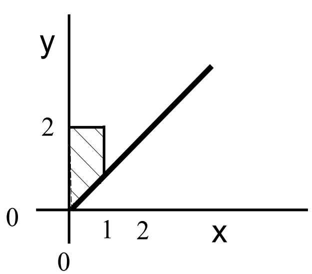
- Finding c

$$c\int_0^3\int_x^{x+2}(x+y)\,dy\,dx = \int_0^3 xy + \frac{y^2}{2}\Big|_x^{x+2}dx$$

$$= \int_0^3\left[x(x+2) + \frac{(x+2)^2}{2} - x^2 - \frac{x^2}{2}\right]dx$$

$$= c\int_0^3(4x+2)\,dx = \left[2x^2 + 2x\right]_0^3 = 24c$$

- c = 1/24
- a) P(X < 1, Y < 2) equals the integral of $f_{XY}(x,y)$



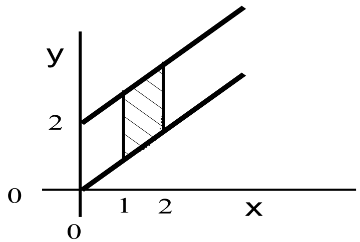

$$P(X<1,Y<2) = \frac{1}{24}\int_0^1\int_x^2(x+y)\,dy\,dx = \frac{1}{24}\int_0^1 xy + \frac{y^2}{2}\Big|_x^2 dx = \frac{1}{24}\int_0^1 2x + 2 - \frac{3x^2}{2}dx =$$

$$\frac{1}{24}\left[x^2 + 2x - \frac{x^3}{2}\Big|_0^1\right] = 0.10417$$

# Example 2



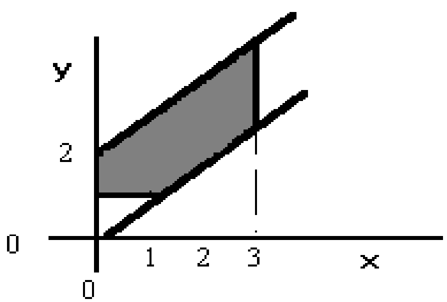
- b) P(1 < X < 2) equals the integral of $f_{XY}(x,y)$




$$P(1<X<2)=\frac{1}{24}\int_1^2\int_x^{x+2}(x+y)\,dy\,dx=\frac{1}{24}\int_1^2 xy+\left.\tfrac{y^2}{2}\right|_x^{x+2}dx$$

$$=\frac{1}{24}\int_1^2(4x+2)\,dx=\frac{1}{24}\left[\left.2x^2+2x\right|_1^2\right]=\tfrac{1}{3}.$$

- c) P (Y>1)




$$P(Y>1)=1-P(Y\le 1)=1-\tfrac{1}{24}\int_0^1\int_x^1(x+y)\,dy\,dx=1-\frac{1}{24}\int_0^1\left.(xy+\frac{y^2}{2})\right|_x^1$$

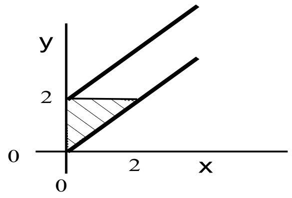
$$=1-\frac{1}{24}\int_0^1 x+\frac{1}{2}-\frac{3}{2}x^2\,dx=1-\frac{1}{24}\left.\left(\frac{x^2}{2}+\frac{1}{2}-\frac{1}{2}x^3\right)\right|_0^1$$

$$=1-0.02083=0.9792$$

# Example 2



- d) P (X<2, Y<2)




$$= \frac{1}{24}\int_0^2\int_x^2 (x+y)dydx = \frac{1}{24}\int_0^2 xy + \tfrac{y^2}{2}\Big|_x^2 dx$$

$$= \frac{1}{24}\int_0^2 (2x+2-1.5x^2)dx = \frac{1}{24}\left[x^2+2x-\frac{x^3}{2}\Big|_0^2\right] = \tfrac{1}{6}.$$

- e) E(X)

$$E(X) = \frac{1}{24}\int_0^3\int_x^{x+2} x(x+y)dydx = \frac{1}{24}\int_0^3 x^2y + \tfrac{xy^2}{2}\Big|_x^{x+2} dx$$

$$= \frac{1}{24}\int_0^3 (4x^2+2x)dx = \frac{1}{24}\left[\frac{4x^3}{3}+x^2\Big|_0^3\right] = \frac{15}{8}$$

- fx(x)

$$f_X(x) = \frac{1}{24}\int_x^{x+2}(x+y)dy = \frac{1}{24}\left[xy+\tfrac{y^2}{2}\Big|_x^{x+2}\right] = \frac{x}{6}+\frac{1}{12}$$

# Linear Functions

# Linear Functions of RVs



- **Content:**
- A linear function of a single random variable (RV) $X$ is an expression of the form:

$$Y = aX + b$$

- More generally, given RVs $X_1, X_2, \dots, X_p$ and constants $c_1, c_2, \dots, c_p$

$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$$

is a linear combination of $X_1, X_2, \dots, X_p$

# Linear Functions of RVs



- **Motivation**: A RV is sometimes defined as a function of one or more RVs
- Example 3.
  - The RVs $X_1$ and $X_2$ denote length and width of a manufactured part,

$Y = 2X_1 + 2X_2$ is a RV representing the perimeter of the part

# Mean and Variance of a Linear Function



- The mean of a Linear Function $Y = aX + b$:

$$E[Y] = E[aX + b] = aE[X] + b$$

Expectation operator is linear

- The variance of $Y = aX + b$:

$$Var(Y) = Var(aX + b) = a^2 Var(X)$$

# Mean and Variance of a Linear Function



- The expected value (mean) of a linear function

$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$$

- is

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p)$$

- The variance of a linear function is

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j)$$

- If $X_1, X_2, \ldots, X_p$ are independent

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p)$$

# Example 4



- The RVs $X_1$ and $X_2$ denote length and width of a manufactured part
  - E($X_1$) = 2 centimeters with standard deviation 0.1
  - E($X_2$) = 5 centimeters with standard deviation 0.2
  - Cov($X_1$, $X_2$) = -0.005

- $Y = 2X_1 + 2X_2$ is a RV representing the perimeter of the part

  E(Y)= 2E($X_1$)+2E($X_1$)= 2(2)+2(5)= 14 centimeters

  - And

$$
\begin{aligned}
V(Y) &= 2^2(0.1^2) + 2^2(0.2^2) + 2 \times 2 \times 2(-0.005) \\
&= 0.04 + 0.16 - 0.04 = 0.16 \text{ centimeters squared}
\end{aligned}
$$

# Covariance



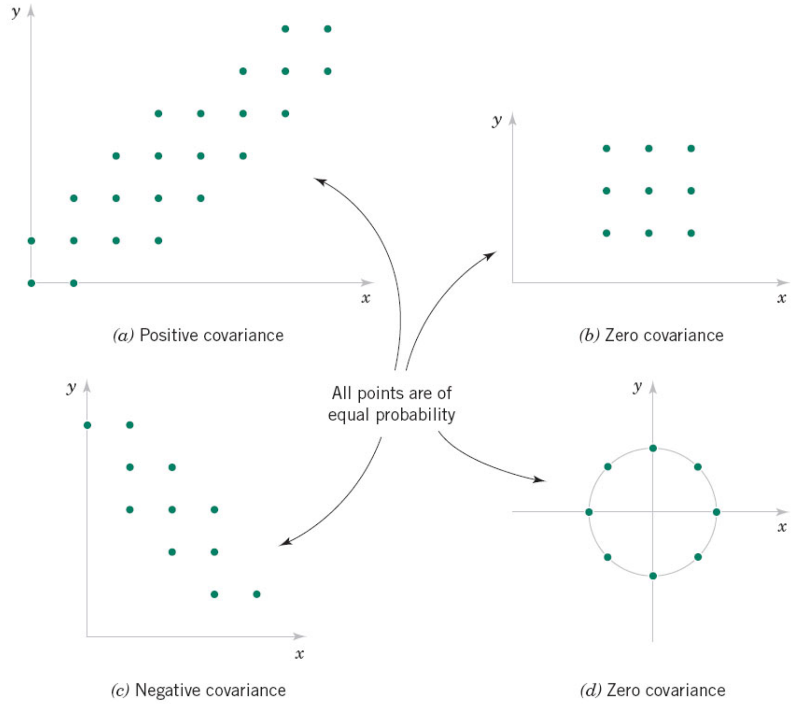
- Covariance between two RVs X and Y is defined as:

$$\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$$

- Covariance is a measure of linear relationship

Positive Covariance

Negative Covariance




Zero Covariance

no linear relationship

# Covariance and Statistical Independence



- If two random variables are statistically independent, the covariance between them is 0
  - However, the converse is not necessarily true

# Mean and Variance of an Average

- The particular linear function that performs the average of $p$ RVs is used quite often
- If $X_1, X_2, \ldots, X_p$ are independent RVs with

$$E(X_i) = \mu \qquad V(X_i) = \sigma^2$$

- Then their average

$$\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$$

has mean and variance equal to

$$E(\overline{X}) = \mu \qquad V(\overline{X}) = \frac{\sigma^2}{p}$$

# Covariance



- Let $X$ be a random variable with mean $\mu_X$, and let $Y$ be a random variable with mean, $\mu_Y$. The expected value of $(X-\mu_X)(Y-\mu_Y)$ is called:
  - The covariance between X and Y, denoted *Cov(X, Y)*
- For discrete random variables

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y)$$

- An equivalent expression is

$$Cov(X,Y) = E(XY) - \mu_x\mu_y = \sum_x \sum_y xyP(x,y) - \mu_x\mu_y$$

# Correlation



- Another measure of the linear relationship between two RVs that is often easier to interpret than the covariance

$$\rho_{X_i X_j} = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{V(X_1)V(X_2)}} = \frac{\sigma_{X_i X_j}}{\sigma_{X_i}\sigma_{X_j}}$$

- Note that, for any two RVs:

$$-1 \leq \rho_{X_i X_j} \leq +1$$

- In case the RVs are independent, then

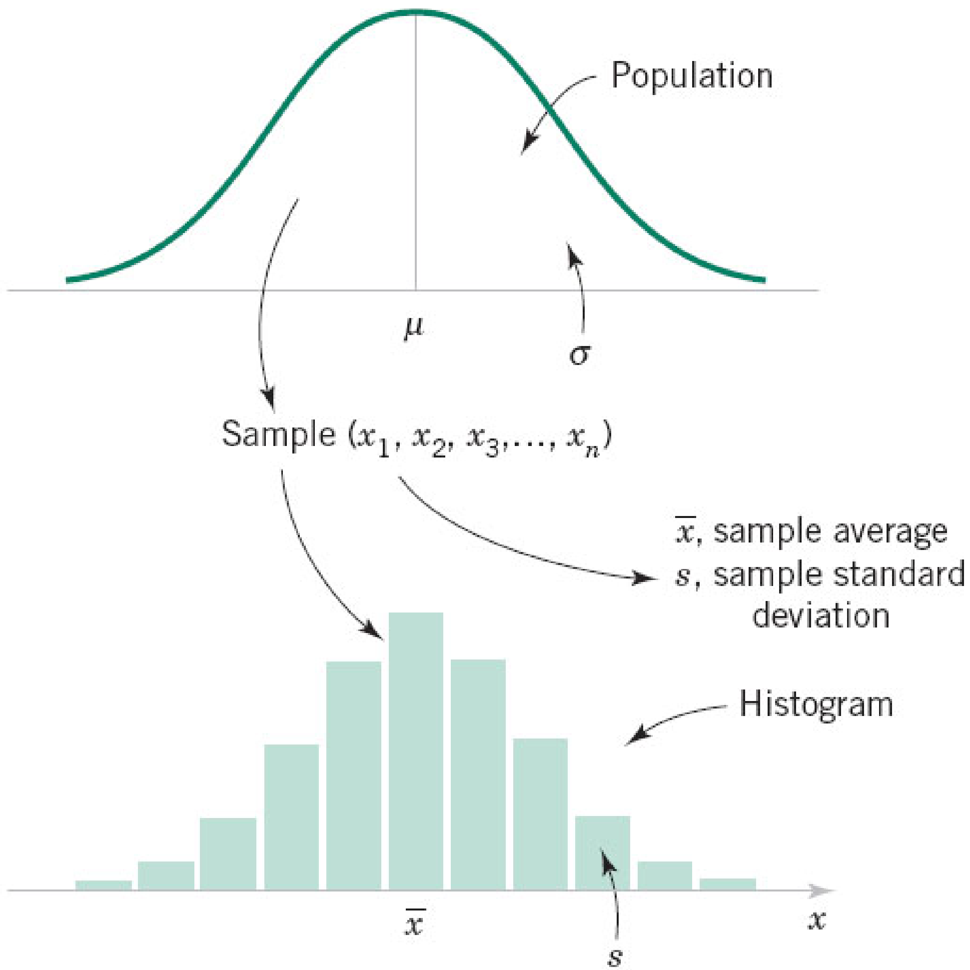
$$\sigma_{X_i X_j} = \rho_{X_i X_j} = 0$$

# Numerical Summaries

# Numerical Summaries



- We usually work with data obtained from observations in samples (size $n$) extracted from larger populations (size $N$)



# Numerical Summaries

- Numerical summaries are statistical measures that provide concise descriptions of a dataset.

# Sample Mean



- We can characterize the central tendency in the data by the sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where $x_1, x_2, \ldots, x_n$ are $n$ observations in a sample selected from some larger population

- The sample mean $\bar{x}$ is a reasonable estimate of the population mean $\mu$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

# Sample Variance



- We can characterize the variability or scatter in the data by the sample variance or sample standard deviation

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s = \sqrt{s^2}$$

where $x_1, x_2, \ldots, x_n$ are $n$ observations in a sample selected from some larger population

- The sample variance $s^2$ is a reasonable estimate of the population variance $\sigma^2$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad \sigma = \sqrt{\sigma^2}$$

# Alternative Equation for the Variance

- Computing the differences $(x_i - \bar{x})$ may be tedious
- An alternative formulation may be used

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

- If we have computed the sample mean, eventually simplifies to

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

# Example 6

- Consider a set of 8 measurements of a pull-off force (in N) of a certain fastening device

| $i$ | $x_i$ | $(x_i - x)^2$ |
|---|---|---|
| 1 | 12.6 | 0.16 |
| 2 | 12.9 | 0.01 |
| 3 | 13.4 | 0.16 |
| 4 | 12.3 | 0.49 |
| 5 | 13.6 | 0.36 |
| 6 | 13.5 | 0.25 |
| 7 | 12.6 | 0.16 |
| 8 | 13.1 | 0.01 |
| | **104.0** | **1.60** |

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{104}{8} = 13.0 \text{ N}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1.60}{7} = 0.2286 \text{ N}^2$$

$$s = \sqrt{0.2286} = 0.48 \text{ N}$$

$$v = \frac{s}{\bar{x}} = \frac{0.48}{13.0} = 0.037 = 3.7\%$$

**Coefficient of variation**

# Example 7

- The concentration of a solution is measured using the same instrument, as follows:
  - 63.2, 67.1, 65.8, 64.0, 65.1, 65.3 g/l
- Calculate the sample mean
  - Ans.: $\overline{x} = (63.2+67.1+65.8+64.0+65.1+65.3)/6 = 65.083$ g/l
- Calculate the sample variance and standard deviation
  - Ans.: $s^2 = 1.86869$ $\quad s = 1.367$
- The coefficient of variation is low: $v = s/\overline{x} = 0.021 = 2.1\%$