2. Means of Linear Combinations of R. V.

develop **some useful properties** that will simplify the calculations of means

2. Means of Linear Combinations of R. V.

develop **some useful properties** that will simplify the calculations of means

Theorem 4.5 If a and b are constant, then

$$E(aX + b) = aE(X) + b.$$

2. Means of Linear Combinations of R. V.

develop **some useful properties** that will simplify the calculations of means

Theorem 4.5 If a and b are constant, then

$$E(aX + b) = aE(X) + b.$$

PROOF By the definition of an expected value,

$E(aX + b) = \int_{-\infty}^{\infty}(ax + b)f(x)dx$

2. Means of Linear Combinations of R. V.

develop **some useful properties** that will simplify the calculations of means

Theorem 4.5 If a and b are constant, then

$$E(aX + b) = aE(X) + b.$$

PROOF By the definition of an expected value,

$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x) dx$
$= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx = aE(X) + b$

Linear Combinations of Random Variables

Remark: All the results are valid for both discrete and continuous random variables. Proofs are given only for the continuous case.

Linear Combinations of Random Variables

Remark: All the results are valid for both discrete and continuous random variables. Proofs are given only for the continuous case.

Two Corollaries:

Linear Combinations of Random Variables

Remark: All the results are valid for both discrete and continuous random variables. Proofs are given only for the continuous case.

Two Corollaries:

Corollary 1 Setting $a = 0$, we see that $E(b) = b$.

Linear Combinations of Random Variables

Remark: All the results are valid for both discrete and continuous random variables. Proofs are given only for the continuous case.

Two Corollaries:

Corollary 1 Setting $a = 0$, we see that $E(b) = b$.

Corollary 2 Setting $b = 0$, we see that $E(aX) = aE(X)$.

Linear Combinations of Random Variables

Theorem 4.6 The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

Linear Combinations of Random Variables

Example 4.17 Let X be a random variable with probability distribution as follows

0	1	2	3
$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X-1)^2$.

Solution......

Linear Combinations of Random Variables

Example 4.17 Let X be a random variable with probability distribution as follows

0	1	2	3
$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X-1)^2$.

Solution......

$E(X-1)^2 = 1.$

Linear Combinations of Random Variables

Example 4.18 The weekly demand for a certain drink, in thousand of liters, at a chain of convenience stores is a continuous random variable $g(X) = X^2 + X - 2$, where X has the density function

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value for the weekly demand of the drink.

Solution......

Linear Combinations of Random Variables

Example 4.18 The weekly demand for a certain drink, in thousand of liters, at a chain of convenience stores is a continuous random variable $g(X) = X^2 + X - 2$, where X has the density function

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value for the weekly demand of the drink.

Solution……

$E(X^2 + X - 2) = 5/2.$

Linear Combinations of Random Variables

the case of two random variables

Linear Combinations of Random Variables

the case of two random variables

Theorem 4.7 The expected value of the sum or difference of two or more functions of random variables X and Y is the sum or difference of the expected values of the functions. That is

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)].$$

Linear Combinations of Random Variables

Theorem 4.8 Let X and Y be two independent random variables. Then

$$E(XY) = E(X)E(Y).$$

Linear Combinations of Random Variables

Theorem 4.8 Let X and Y be two independent random variables. Then

$$E(XY) = E(X)E(Y).$$

PROOF……

3. Variance and Covariance

The mean or expected value of a random variable describes where the probability distribution is **centered**.

3. Variance and Covariance

The mean or expected value of a random variable describes where the probability distribution is **centered**.

It is important, but not an adequate description of the shape of the distribution.

Variance

Example

two different probability distributions have the same mean

1	2	3		
0.3	0.4	0.3		

and

0	1	2	3	4
0.2	0.1	0.3	0.3	0.1

Variance

Example

two different probability distributions have the same mean

1	2	3
0.3	0.4	0.3

and

0	1	2	3	4
0.2	0.1	0.3	0.3	0.1

We need to characterize the **variability** in the distribution.

Variance

The most important measure of variability of a r.v. X is obtained by letting $g(X) = (X - \mu_X)^2$ in Theorem 4.1.

Variance

The most important measure of variability of a r.v. X is obtained by letting $g(X) = (X - \mu_X)^2$ in Theorem 4.1.

We call it the **variance of the random variable** X, denoted by $Var(X)$ or σ_X^2.

Variance

The most important measure of variability of a r.v. X is obtained by letting $g(X) = (X - \mu_X)^2$ in Theorem 4.1.

We call it the **variance of the random variable** X, denoted by $Var(X)$ or σ_X^2.

Definition 4.3

Let X be a discrete random variable with probability distribution $p(x)$ and mean μ_X. The variance of X is

$$\sigma_X^2 = E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 p(x).$$

Variance

Definition 4.3'

Let X be a continuous random variable with density function $f(x)$ and mean μ_X. The variance of X is

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)dx.$$

Variance

Definition 4.3'

Let X be a continuous random variable with density function $f(x)$ and mean μ_X. The variance of X is

$$\sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx.$$

- The positive square root of the variance, σ_X, is called the **standard deviation** of X.

Variance

Remark:

σ_X^2 will be much smaller for a set of x values that are close to μ_X than it would be for a set of values that vary considerably from μ_X.

Variance

Example 4.8 Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A is

1	2	3
0.3	0.4	0.3

and for company B is

0	1	2	3	4
0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that of company A.

Variance

an alternative and preferred formula for finding σ^2,

Variance

an alternative and preferred formula for finding σ^2,

Theorem 4.2 The variance of a random variable X is

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

Variance

an alternative and preferred formula for finding σ^2,

Theorem 4.2 The variance of a random variable X is

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

PROOF ......

Variance

an alternative and preferred formula for finding σ^2,

Theorem 4.2 The variance of a random variable X is

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

PROOF ......

- Theorem 4.2 often simplifies the calculations

Variance

Example 4.9 Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X

0	1	2	3
0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ_X^2

Solution......

Variance

Example 4.9 Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X

0	1	2	3
0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ_X^2

Solution......

$\sigma_X^2 = 0.4979$

Variance

Example 4.10 The weekly demand for Pepsi, in thousands of liters, from a local chain of efficiency stores, is a continuous random variable X having the probability density

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2 \\ 0, & \text{elsewhere}, \end{cases}$$

Find the mean and variance of X.

Solution......

Variance

Example 4.10 The weekly demand for Pepsi, in thousands of liters, from a local chain of efficiency stores, is a continuous random variable X having the probability density

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2 \\ 0, & \text{elsewhere,} \end{cases}$$

Find the mean and variance of X.

Solution......

$\mu_X = 5/3, \qquad \sigma_X^2 = 1/18.$

Variance

The variance or standard deviation only has meaning when we compare two or more distributions that have **the same units of measurement**.

Variance

The variance or standard deviation only has meaning when we compare two or more distributions that have **the same units of measurement**.

We could compare the variances of the distributions of contents, measured in liters, for two companies bottling orange juice.

Variance

The variance or standard deviation only has meaning when we compare two or more distributions that have **the same units of measurement**.

We could compare the variances of the distributions of contents, measured in liters, for two companies bottling orange juice. And the large value would indicate the company whose product is more variable or less uniform.

Variance

The variance or standard deviation only has meaning when we compare two or more distributions that have **the same units of measurement**.

We could compare the variances of the distributions of contents, measured in liters, for two companies bottling orange juice. And the large value would indicate the company whose product is more variable or less uniform.

It would not be meaningful to compare the variance of a distribution of heights to the variance of a distribution of aptitude scores.

Variance

consider the variance of $g(X)$, denoted by $\sigma^2_{g(X)}$

Variance

consider the variance of $g(X)$, denoted by $\sigma^2_{g(X)}$

Theorem 4.3 Let X be a discrete random variable with probability distribution $p(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 p(x).$$

Variance

Theorem 4.3' Let X be a continuous random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x) dx.$$

Variance

Example 4.11 Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

0	1	2	3
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Solution......

Variance

Example 4.11 Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

0	1	2	3
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Solution......

$$\sigma^2_{2X+3} = 4$$

Variance

Example 4.12 Let X be a random variable having the density function given

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere,} \end{cases}$$

Find the variance of the random variable $g(X) = 4X + 3$.

Solution......

Variance

Example 4.12 Let X be a random variable having the density function given

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere}, \end{cases}$$

Find the variance of the random variable $g(X) = 4X + 3$.

Solution......

$\sigma^2_{4X+3} = 51/5$

Covariance

If $g(X, Y) = (X - \mu_X)(Y - \mu_Y)$, then $E[g(X, Y)]$ is called the covariance of X and Y, and denoted by σ_{XY} or $Cov(X, Y)$.

Covariance

If $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$, then $E[g(X,Y)]$ is called the covariance of X and Y, and denoted by σ_{XY} or $Cov(X,Y)$.

Definition 4.4 Let X and Y be discrete random variables with joint probability distribution $p(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y).$$

Covariance

Definition 4.4' Let X and Y be continuous random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx d$$

Covariance

covariance: a measurement of the nature of the association between two random variables.

Covariance

covariance: a measurement of the nature of the association between two random variables.

If large values of X often result in large values of Y, or small values of X result in small values of Y. Thus the product $(X - \mu_X)(Y - \mu_Y)$ will tend to be positive.

Covariance

covariance: a measurement of the nature of the association between two random variables.

If large values of X often result in large values of Y, or small values of X result in small values of Y. Thus the product $(X - \mu_X)(Y - \mu_Y)$ will tend to be positive.

On the other hand, if large X values often result in small Y values, the product $(X - \mu_X)(Y - \mu_Y)$ will tend to be negative.

Covariance

the alternative and preferred formula for $Cov(X, Y)$:

Covariance

the alternative and preferred formula for $Cov(X, Y)$:

Theorem 4.4 The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y).$$

Covariance

the alternative and preferred formula for $Cov(X, Y)$:

Theorem 4.4 The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y).$$

PROOF......

Covariance

When X and Y are statistically independent, the covariance is zero.

Covariance

When X and Y are statistically independent, the covariance is zero.

Two variables may have zero covariance and still not be statistically independent.

Covariance

Example 4.13 The number of blue refills X and the number of red refills Y, when 2 refills for a ballpoint pen are selected at random from a certain box, was described in Example 3.8. Find the covariance of X and Y.

Covariance

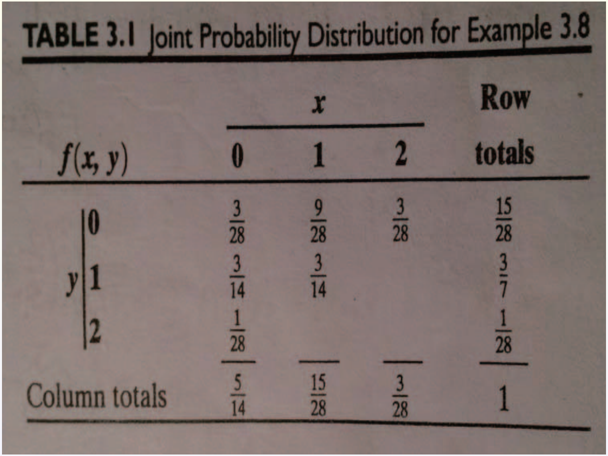

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x = 0$	$x = 1$	$x = 2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Covariance

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x=0$	$x=1$	$x=2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

$$\sigma_{XY} = -9/56$$

Covariance

Example 4.14 The fraction X of male runners and the fraction Y of female runners who compete in marathon races is described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \leq x \leq 1, 0 \leq y \leq x \\ 0, & \text{elsewhere,} \end{cases}$$

Find the covariance of X and Y.

Solution......

Covariance

Example 4.14 The fraction X of male runners and the fraction Y of female runners who compete in marathon races is described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \leq x \leq 1, 0 \leq y \leq x \\ 0, & \text{elsewhere,} \end{cases}$$

Find the covariance of X and Y.

Solution......

$\sigma_{XY} = 4/225$

Covariance

Although the covariance between two random variables does provide information regarding the nature of the relationship, the magnitude of σ_{XY} does not indicate anything regarding the strength of the relationship.

Covariance

Although the covariance between two random variables does provide information regarding the nature of the relationship, the magnitude of σ_{XY} does not indicate anything regarding the strength of the relationship.

Because σ_{XY} is not scale free. Its magnitude will depend on the units measured for both X and Y.

Correlation coefficient

the **correlation coefficient** is a scale-free

Correlation coefficient

the **correlation coefficient** is a scale-free

Definition 4.5 Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Correlation coefficient

the **correlation coefficient** is a scale-free

Definition 4.5 Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- ρ_{XY} is free of the units of X and Y
- $-1 \leq \rho_{XY} \leq 1$
- $\rho_{XY} = 0$ when $\sigma_{XY} = 0$.

Correlation coefficient

Suppose $Y = a + bX$,

Correlation coefficient

Suppose $Y = a + bX$,

$\rho_{XY} = 1$, if $b > 0$, and

Correlation coefficient

Suppose $Y = a + bX$,

$\rho_{XY} = 1$, if $b > 0$, and
$\rho_{XY} = -1$, if $b < 0$.

Correlation coefficient

Suppose $Y = a + bX$,

$\rho_{XY} = 1$, if $b > 0$, and
$\rho_{XY} = -1$, if $b < 0$.

How to prove?

Linear Combinations of Random Variables

some **useful properties** that will simplify the caculations of variance of random variables

Linear Combinations of Random Variables

some **useful properties** that will simplify the caculations of variance of random variables

Theorem 4.9 If a and b are constants, then

$$\sigma^2_{aX+b} = a^2\sigma^2_X.$$

PROOF......

Linear Combinations of Random Variables

some **useful properties** that will simplify the caculations of variance of random variables

Theorem 4.9 If a and b are constants, then

$$\sigma^2_{aX+b} = a^2\sigma^2_X.$$

PROOF......

Linear Combinations of Random Variables

Corollary 1 Setting $a = 1$, we see that

$$\sigma^2_{X+b} = \sigma^2_X.$$

Linear Combinations of Random Variables

Corollary 1 Setting $a = 1$, we see that

$$\sigma^2_{X+b} = \sigma^2_X.$$

Corollary 2 Setting $b = 0$, we see that

$$\sigma^2_{aX} = a^2\sigma^2_X.$$

Linear Combinations of Random Variables

Theorem 4.10 X and Y are random variables, a and b are constants, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

PROOF......

Linear Combinations of Random Variables

Corollary 1 If X and Y are independent random variables, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Linear Combinations of Random Variables

Corollary 1 If X and Y are independent random variables, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Corollary 2 If X and Y are independent random variables, then

$$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Linear Combinations of Random Variables

Generalizing to a linear combination of n independent r.v.

Linear Combinations of Random Variables

Generalizing to a linear combination of n independent r.v.

Corollary 3 If $X_1, X_2, \ldots X_n$ are independent random variables, then

$$\sigma^2_{a_1 X_1 + a_2 X_2 + \cdots + a_n X_n} = a_1^2 \sigma^2_{X_1} + a_2^2 \sigma^2_{X_2} + \cdots + a_n^2 \sigma^2_{X_n}.$$

Linear Combinations of Random Variables

Example 4.20 If X and Y are random variables with variance $\sigma_X^2 = 2, \sigma_Y^2 = 4$, and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

Solution......

Linear Combinations of Random Variables

Example 4.20 If X and Y are random variables with variance $\sigma_X^2 = 2, \sigma_Y^2 = 4$, and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

Solution......

$\sigma_Z^2 = 130$

Linear Combinations of Random Variables

Example 4.21 Let X and Y denote the amount of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variance $\sigma_X^2 = 2$ and $\sigma_Y^2 = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

Solution......

Linear Combinations of Random Variables

Example 4.21 Let X and Y denote the amount of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variance $\sigma_X^2 = 2$ and $\sigma_Y^2 = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

Solution......

$\sigma_Z^2 = 30$