

# Review of Statistical Theory

## Part 1

Dragos Ailoae  
dailoae@gradcenter.cuny.edu

Advanced Economics and Business Statistics  
ECON-4400w - Spring 2022

Brooklyn College  
Feb 7, 2022

# Review of Statistical Theory

1. **The probability framework for statistical inference**
2. Estimation
3. Testing
4. Confidence Intervals

## The probability framework for statistical inference

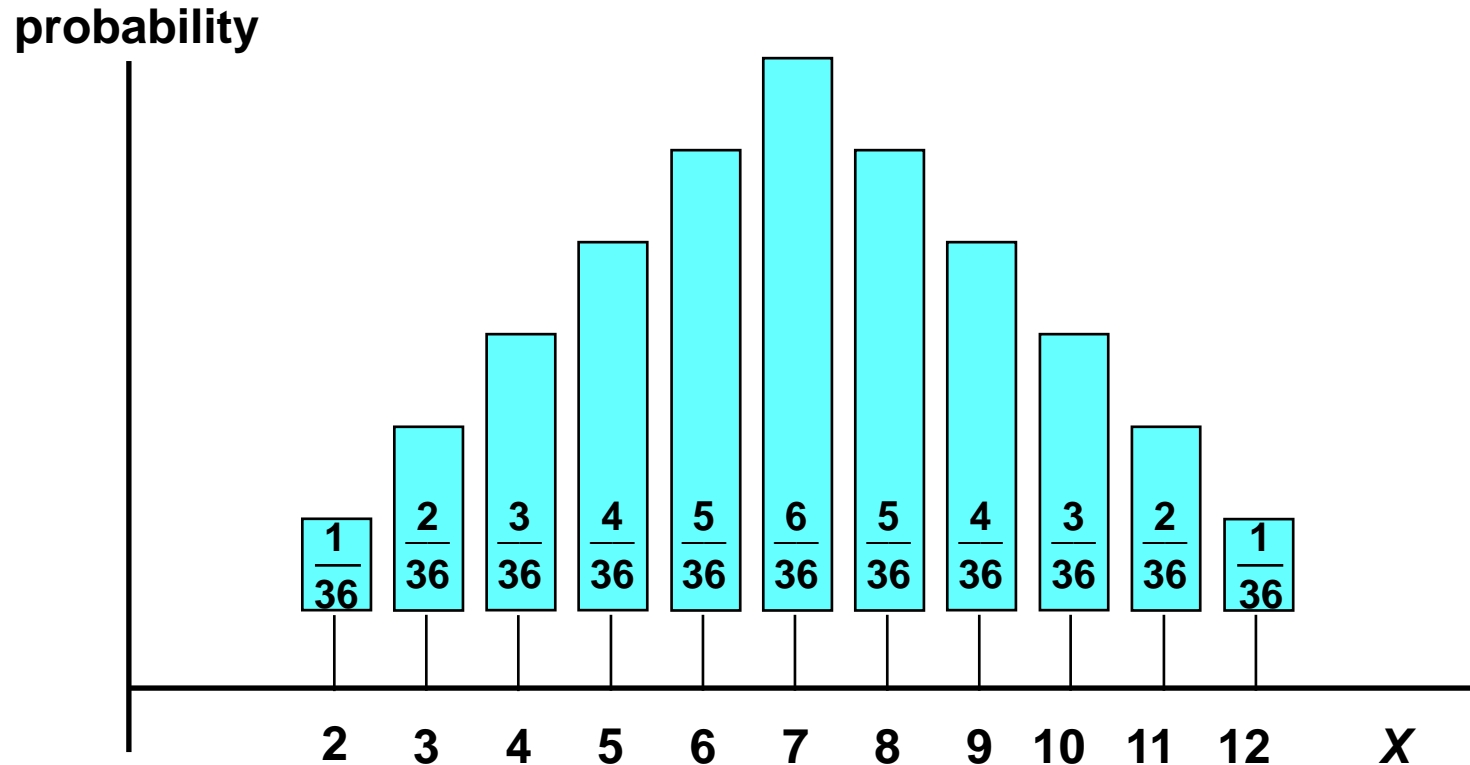
- a) **Random variable, distribution**
- b) **Moments of a distribution (mean, variance, standard deviation, covariance, correlation)**
- c) **Conditional distributions and conditional means**
- d) **Distribution of a sample of data drawn randomly from a population:  $Y_1, \dots, Y_n$**

Experiment: Toss a Pair of “Fair” Dice and Calculate the Sum the Dots on the Two Upward Facing Sides					
Face Values - Outcomes	Events Sum of Dots	Number of Outcomes that produce the Event			Probability
(1,1)	2		1		1/36 = 0.0278
(1,2) (2,1)	3		2		2/36 = 0.0556
(1,3) (2,2) (3,1)	4		3		3/36 = 0.0833
(1,4) (2,3) (3,2) (4,1)	5		4		4/36 = 0.1111
(1,5) (2,4) (3,3) (4,2) (5,1)	6		5		5/36 = 0.1389
(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	7		6		6/36 = 0.1667
(2,6) (3,5) (4,4) (5,3) (6,2)	8		5		5/36 = 0.1389
(3,6) (4,5) (5,4) (6,3)	9		4		4/36 = 0.1111
(4,6) (5,5) (6,4)	10		3		3/36 = 0.0833
(5,6) (6,5)	11		2		2/36 = 0.0556
(6,6)	12		1		1/36 = 0.0278
Total # of outcomes	1+2+3+...+2+1 =		36		

# The probability framework: nomenclature

- The mutually exclusive potential results of a random process are called the **outcomes**.
- The **probability** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The set of all possible outcomes is called the **sample space**. An **event** is a subset of the sample space, that is, an event is a set of one or more outcomes.
- A **random variable** is a numerical summary of a random outcome. The number of times your computer crashes while you are writing a term paper is random and takes on a numerical value, so it is a random variable.
- Some random variables are **discrete** and some are **continuous**. As their names suggest, a discrete random variable takes on only a discrete set of values, like 0, 1, 2, c, whereas a continuous random variable takes on a continuum of possible values.

# Probability distribution for dice example

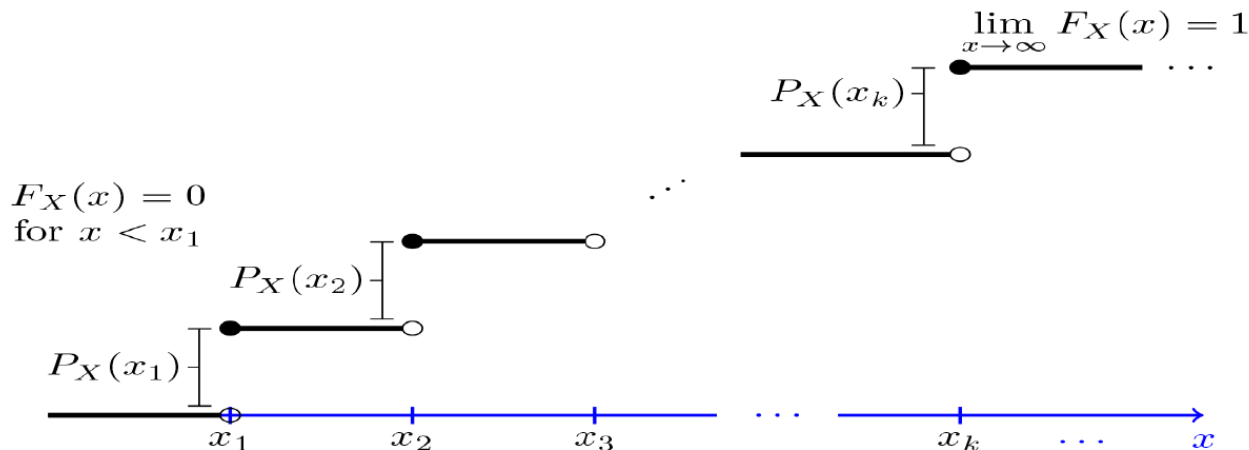


**EXAMPLE:  $X$  IS THE SUM OF TWO DICE**

# Probability Distribution of a Discrete Random Variable

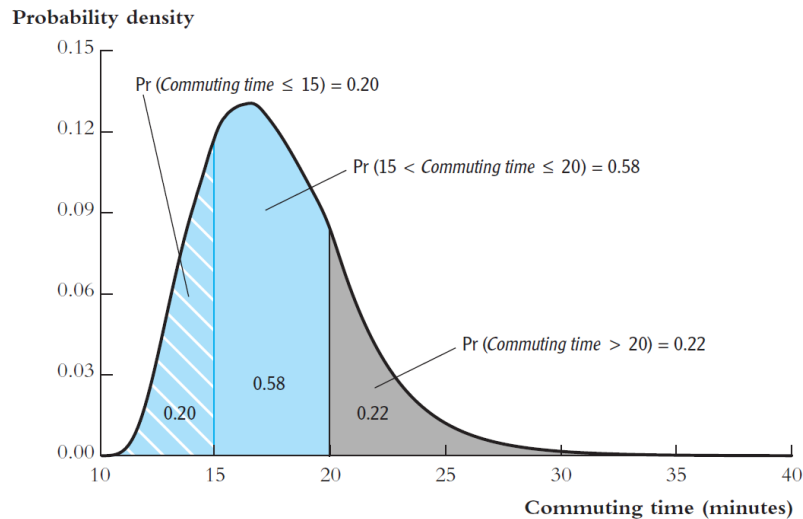
- The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur. These probabilities sum to 1.
- The **cumulative probability distribution** is the probability that the random variable is less than or equal to a particular value. A cumulative probability distribution is also referred to as a cumulative distribution function, a C.D.F., or a cumulative distribution.

## *CDF of discrete random variables.*



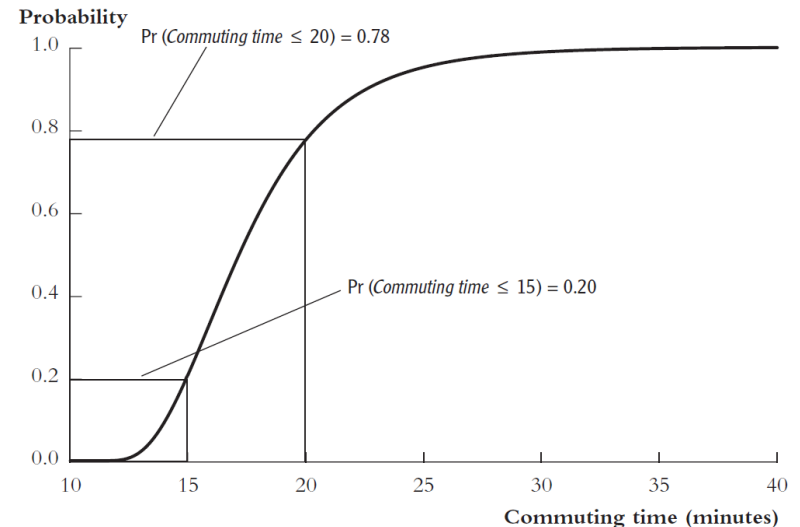
# Probability Distribution of a Continuous Random Variable

## Probability Density Function (PDF)



(b) Probability density function of commuting time

## Cumulative Distribution Function (CDF)



(a) Cumulative distribution function of commuting time

## (b) Moments of a population distribution: mean, variance, standard deviation, covariance, correlation

***mean*** = expected value (expectation) of  $Y$  *a.k.a.* **first moment**

$$= E(Y)$$

$$= \mu_Y$$

= long-run average value of  $Y$  over repeated realizations of  $Y$

$$\mathbf{variance} = E(Y - \mu_Y)^2$$

$$= \sigma_Y^2$$

= measure of the squared spread of the distribution

= second central moment

$$\text{standard deviation} = \sqrt{\text{variance}} = \sigma_Y$$



## (b) Moments of a population distribution: mean, variance, standard deviation, covariance, correlation (cont'd)

- The **expected value** of a random variable  $Y$ , denoted  $E(Y)$ , is the long-run average value of the random variable over many repeated trials or occurrences.
- The expected value of a discrete random variable is computed as a weighted average of the possible outcomes of that random variable, where the weights are the probabilities of that outcome.

### KEY CONCEPT

### Expected Value and the Mean

## 2.1

Suppose the random variable  $Y$  takes on  $k$  possible values,  $y_1, \dots, y_k$ , where  $y_1$  denotes the first value,  $y_2$  denotes the second value, and so forth, and that the probability that  $Y$  takes on  $y_1$  is  $p_1$ , the probability that  $Y$  takes on  $y_2$  is  $p_2$ , and so forth. The expected value of  $Y$ , denoted  $E(Y)$ , is

$$E(Y) = y_1 p_1 + y_2 p_2 + \cdots + y_k p_k = \sum_{i=1}^k y_i p_i \quad (2.3)$$

where the notation  $\sum_{i=1}^k y_i p_i$  means “the sum of  $y_i p_i$  for  $i$  running from 1 to  $k$ .” The expected value of  $Y$  is also called the mean of  $Y$  or the expectation of  $Y$  and is denoted  $\mu_Y$ .

# Expected value / weighted average in Excel

## Using SUM()

	A	B	C	D	E
	Y (Sum of Dots)	Number of Outcomes that produce the Event	Probability (colB / 36)	Probability weighted Y (colA*colC)	
1					
2	2	1	0.0278	0.06	
3	3	2	0.0556	0.17	
4	4	3	0.0833	0.33	
5	5	4	0.1111	0.56	
6	6	5	0.1389	0.83	
7	7	6	0.1667	1.17	
8	8	5	0.1389	1.11	
9	9	4	0.1111	1.00	
10	10	3	0.0833	0.83	
11	11	2	0.0556	0.61	
12	12	1	0.0278	0.33	
13	E(Y) =SUM(D2:D12)				7.00

## Using SUMPRODUCT()

	A	B	C	D
	Y (Sum of Dots)	Number of Outcomes that produce the Event	Probability (colB / 36)	Probability weighted Y (colA*colC)
1				
2	2	1	0.0278	0.06
3	3	2	0.0556	0.17
4	4	3	0.0833	0.33
5	5	4	0.1111	0.56
6	6	5	0.1389	0.83
7	7	6	0.1667	1.17
8	8	5	0.1389	1.11
9	9	4	0.1111	1.00
10	10	3	0.0833	0.83
11	11	2	0.0556	0.61
12	12	1	0.0278	0.33
13	E(Y)			7.00
14	=SUMPRODUCT(A2:A12,C2:C12)			
15				

## (b) Moments of a population distribution: mean, variance, standard deviation, covariance, correlation (cont'd)

The **variance** and **standard deviation** measure the dispersion or the “spread” of a probability distribution.

### Variance and Standard Deviation

KEY CONCEPT

2.2

The variance of the discrete random variable  $Y$ , denoted  $\sigma_Y^2$ , is

$$\sigma_Y^2 = \text{var}(Y) = E[(Y - \mu_Y)^2] = \sum_{i=1}^k (y_i - \mu_Y)^2 p_i. \quad (2.5)$$

The standard deviation of  $Y$  is  $\sigma_Y$ , the square root of the variance. The units of the standard deviation are the same as the units of  $Y$ .

# Variance / Standard Deviation in Excel

## Variance (Y)

## Standard Deviation (Y)

	A	B	C	D	E	F	G	H
	Y (Sum of Dots)	Number of Outcomes that produce the Event	Probability (colB / 36)	Probability weighted Y (colA*colC)	Difference From the Mean (colA - E(Y) or (colA - 7))	Difference From the Mean Squared (colE ^ 2)		
1								
2	2	1	0.0278	0.06	-5	25		
3	3	2	0.0556	0.17	-4	16		
4	4	3	0.0833	0.33	-3	9		
5	5	4	0.1111	0.56	-2	4		
6	6	5	0.1389	0.83	-1	1		
7	7	6	0.1667	1.17	0	0		
8	8	5	0.1389	1.11	1	1		
9	9	4	0.1111	1.00	2	4		
10	10	3	0.0833	0.83	3	9		
11	11	2	0.0556	0.61	4	16		
12	12	1	0.0278	0.33	5	25		
13			E(Y)	7.00		Variance(Y)	5.83	
14			E(Y)	7.00		=SUMPRODUCT(F2:F12,C2:C12)		

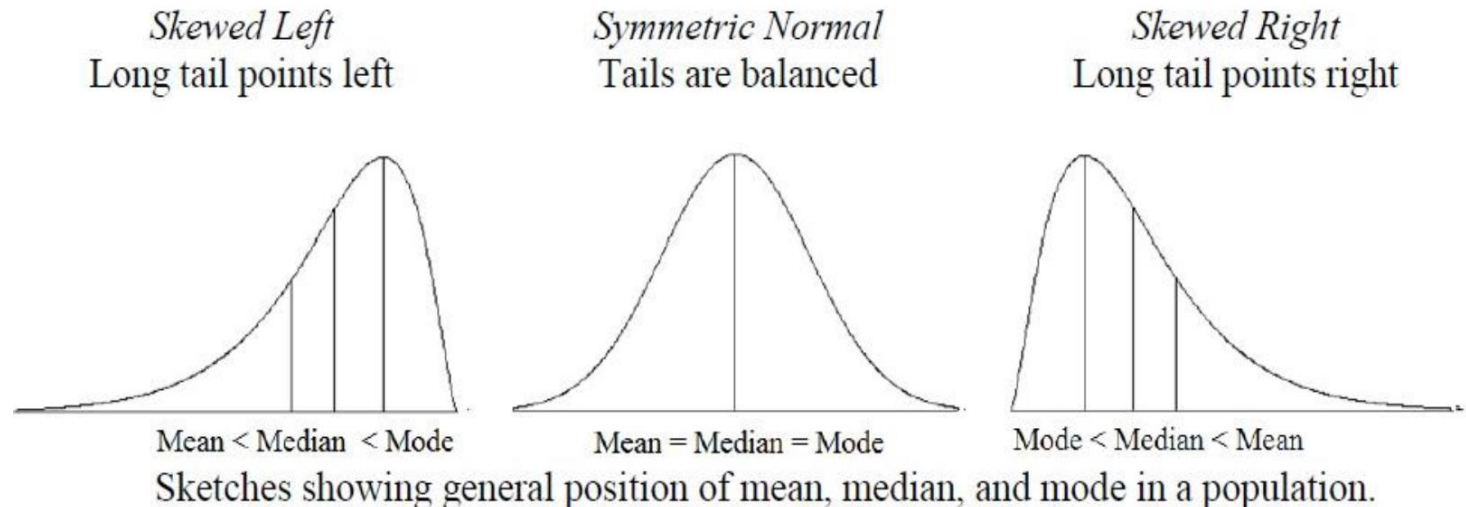
	A	B	C	D	E	F
	Y (Sum of Dots)	Number of Outcomes that produce the Event	Probability (colB / 36)	Probability weighted Y (colA*colC)	Difference From the Mean (colA - E(Y) or (colA - 7))	Difference From the Mean Squared (colE ^ 2)
1						
2	2	1	0.0278	0.06	-5	25
3	3	2	0.0556	0.17	-4	16
4	4	3	0.0833	0.33	-3	9
5	5	4	0.1111	0.56	-2	4
6	6	5	0.1389	0.83	-1	1
7	7	6	0.1667	1.17	0	0
8	8	5	0.1389	1.11	1	1
9	9	4	0.1111	1.00	2	4
10	10	3	0.0833	0.83	3	9
11	11	2	0.0556	0.61	4	16
12	12	1	0.0278	0.33	5	25
13			E(Y)	7.00		Variance(Y)
14			E(Y)	7.00		5.83
15						
16						
17						2.42
18						=SQRT(F14)
						StdDev(Y)

## (b) Moments of a population distribution: mean, variance, standard deviation, covariance, correlation (cont'd)

$$\text{skewness} = \frac{E[(Y - \mu_Y)^3]}{\sigma_Y^3}$$

= third central moment; measure of asymmetry of a distribution

- *skewness* = 0: distribution is symmetric
- *skewness* > 0: distribution has long right tail (if <0 then long left tail)



## (b) Moments of a population distribution: mean, variance, standard deviation, covariance, correlation (cont'd)

$$\text{kurtosis} = \frac{E[(Y - \mu_Y)^4]}{\sigma_Y^4}$$

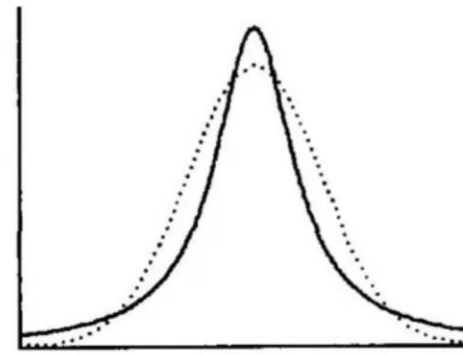
= fourth central moment; measure of mass in tails

= measure of probability of large values

- *kurtosis* = 3: normal distribution
- *kurtosis* > 3: heavy tails (“*leptokurtotic*”)



Negative kurtosis  
(platykurtic)



Positive kurtosis  
(leptokurtic)

# 2 random variables: joint distributions and covariance

- Random variables  $X$  and  $Y$  have a *joint distribution*
- The joint probability distribution can be written as the function

$$\Pr(X = x, Y = y) = f(x, y)$$

- The *covariance* between  $X$  and  $Y$  is

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

$$\text{cov}(X, Y) = E[(XY)] - E(X)E(Y) \equiv E[(XY)] - \mu_X \mu_Y = \sigma_{XY}$$

- The covariance is a measure of the linear association between  $X$  and  $Y$ ; its units are units of  $X \times$  units of  $Y$
- $\text{cov}(X, Y) > 0$  means a positive relation between  $X$  and  $Y$
- If  $X$  and  $Y$  are independently distributed, then  $\text{cov}(X, Y) = 0$  (but not vice versa!!)
- The covariance of a r.v. with itself is its variance:

$$\text{cov}(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = \sigma_X^2$$

# Means, Variances, and Covariances of Sums of Random Variables

Let  $X$ ,  $Y$ , and  $V$  be random variables, let  $\mu_X$  and  $\sigma_X^2$  be the mean and variance of  $X$ , let  $\sigma_{XY}$  be the covariance between  $X$  and  $Y$  (and so forth for the other variables), and let  $a$ ,  $b$ , and  $c$  be constants. Equations (2.29) through (2.35) follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \quad (2.29)$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \quad (2.30)$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \quad (2.31)$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \quad (2.32)$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \quad (2.33)$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y, \quad (2.34)$$

$$|\text{corr}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality)}. \quad (2.35)$$



# Discrete Bivariate Probability Distribution

X \ Y	Y				Marginal Distr. of X	$\sum_y f(x,y)$
		-4	1	3		
	1	0.15	0.20	0.05	0.40	
	2	0.10	0.25	0.25	0.60	
Marginal Distr. of Y $\sum_x f(x,y)$		0.25	0.45	0.30	1.00	$\sum_x \sum_y f(x,y) = 1$

$$\text{cov}(X,Y) = E[(XY)] - E(X)E(Y)$$

$$\begin{aligned} E(X, Y) = X*Y*f(x,y) &= 1(-4)(0.15) + 1(1)(.2) + 1(3)(.05) + \\ &\quad + 2(-4)(.1) + 2(1)(.25) + 2(3)(.25) \\ &= -0.6 + 0.2 + 0.05 + -0.8 + 0.5 + 1.5 \\ &= -0.25 + 1.2 = 0.95 \end{aligned}$$

$$E(X) = (1)(0.4) + 2(0.6) = 1.6$$

$$E(Y) = (-4)(0.25) + (1)(0.45) + (3)(0.3) = 0.35$$

$$E(X)E(Y) = 1.6 * 0.35 = 0.56$$

$$\text{cov}(X,Y) = 0.95 - 0.56 = 0.39$$

The *correlation coefficient* is defined in terms of the covariance:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = r_{XY}$$

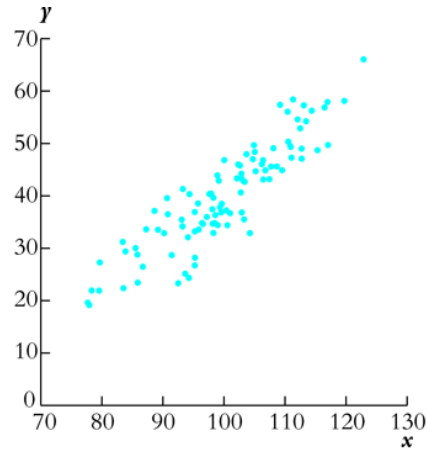
- $-1 \leq \text{corr}(X, Y) \leq 1$
- $\text{corr}(X, Y) = 1$  mean perfect positive linear association
- $\text{corr}(X, Y) = -1$  means perfect negative linear association
- $\text{corr}(X, Y) = 0$  means no linear association

Example:

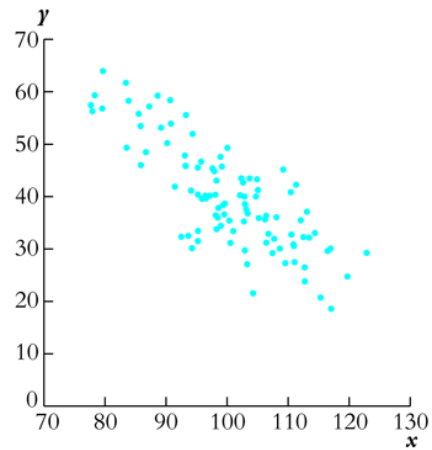
$$\sigma_X = 2.05, \sigma_Y = 1.50, \text{cov}(X, Y) = 2.24$$

$$r_{XY} = 2.24 / (2.05)(1.5) = 0.73$$

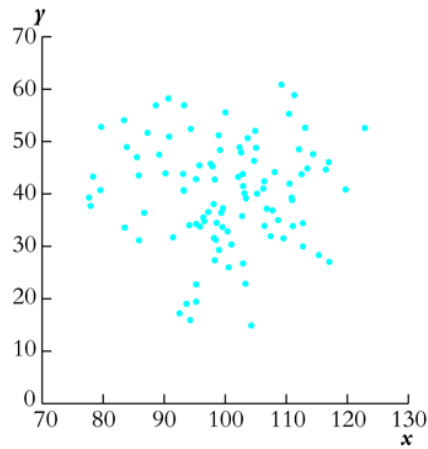
# *The correlation coefficient measures linear association*



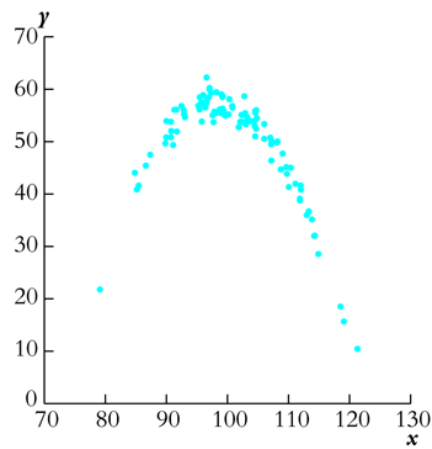
**(a)** Correlation = +0.9



**(b)** Correlation = -0.8



**(c)** Correlation = 0.0



**(d)** Correlation = 0.0 (quadratic)

## (c) Conditional distributions and conditional means

### *Conditional distributions*

- The distribution of  $Y$ , given value(s) of some other random variable,  $X$
- Ex: the distribution of test scores, given that  $STR < 20$

### *Conditional expectations and conditional moments*

- *conditional mean* = mean of conditional distribution  
 $= E(Y|X = x)$  (*important concept and notation*)
- *conditional variance* = variance of conditional distribution
- *Example*:  $E(\text{Test score}|STR < 20)$  = the mean of test scores among districts with small class sizes

*The difference in means is the difference between the means of two conditional distributions:*

## (c) Conditional distributions and conditional means (cont'd)

$$\Delta = E(\text{Test score} | STR < 20) - E(\text{Test score} | STR \geq 20)$$

Other examples of conditional means:

- Wages of all female workers ( $Y = \text{wages}$ ,  $X = \text{sex}$ )
- Mortality rate of those given an experimental treatment ( $Y = \text{live/die}$ ;  $X = \text{treated/not treated}$ )
- If  $E(X|Z) = \text{const}$ , then  $\text{corr}(X,Z) = 0$  (not necessarily vice versa however)

*The conditional mean is a (possibly new) term for the familiar idea of the group mean*

## (c) Conditional distributions and conditional means (cont'd)

The conditional mean plays a key role in prediction:

- Suppose you want to predict a value of  $Y$ , and you are given the value of a random variable  $X$  that is related to  $Y$ . That is, you want to predict  $Y$  given the value of  $X$ .
  - For example, you want to predict someone's income, given their years of education.
- A common measure of the quality of a prediction  $m$  of  $Y$  is the mean squared prediction error (MSPE), given  $X$ ,  $E[(Y - m)^2|X]$
- Of all possible predictions  $m$  that depend on  $X$ , the conditional mean  $E(Y|X)$  has the smallest mean squared prediction error (*optional proof is in Appendix 2.2*).

## (c) Conditional distributions and conditional means (cont'd)

**Example.** Suppose that  $(X, Y)$  is a bivariate discrete random variable such that the point  $(1, 2)$  occurs with probability  $1/8$ ,  $(1, 3)$  with probability  $3/8$ ,  $(2, 3)$  with probability  $1/4$ , and  $(3, 1)$  with probability  $1/4$ . Then  $(X, Y)$  assumes as values only one of these four points.

	$Y = 1$	$Y = 2$	$Y = 3$	marginal of $X$
$X = 1$	0	$1/8$	$3/8$	$1/2$
$X = 2$	0	0	$1/4$	$1/4$
$X = 3$	$1/4$	0	0	$1/4$
marginal of $Y$	$1/4$	$1/8$	$5/8$	1

## (c) Conditional distributions and conditional means (cont'd)

We compute the **conditional probability** function of  $Y$  given  $X = 1$ . Note that  $P[Y = y \mid X = 1] = 0$  except for  $y = 2, 3$ . Thus,

$$P[Y = 2 \mid X = 1] = \frac{P[X = 1, Y = 2]}{P[X = 1]} = \frac{1/8}{1/2} = 1/4;$$
$$P[Y = 3 \mid X = 1] = \frac{P[X = 1, Y = 3]}{P[X = 1]} = \frac{3/8}{1/2} = 3/4.$$

Note that once again  $\sum_y P[Y = y \mid X = 1] = 1$ .

We compute the **conditional mean** of  $Y$  given that  $X = 1$  :

$$E[Y \mid X = 1] = 2 \cdot 1/4 + 3 \cdot 3/4 = 11/4.$$

Also, we compute the conditional mean of  $X$  given that  $Y = 3$ . The conditional distribution of  $X$  given  $Y = 3$  :

$$p_{X|Y=3}(1) = 3/5; p_{X|Y=3}(2) = 2/5; p_{X|Y=3}(3) = 0.$$

Thus  $E[X \mid Y = 3] = 1 \cdot 3/5 + 2 \cdot 2/5 = 7/5$ .