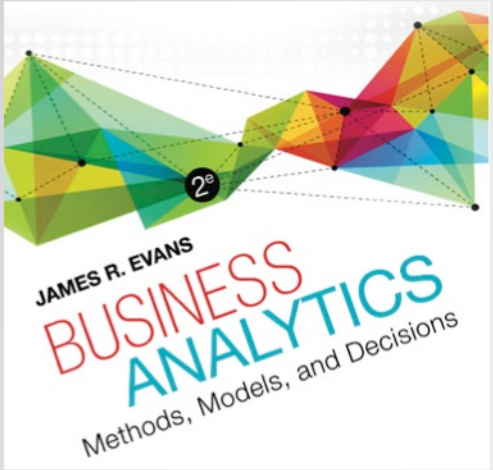


Chapter 5 Probability Distributions and Data Modeling

Basic Concepts of Probability

- **Probability** is the likelihood that an outcome occurs. Probabilities are expressed as values between 0 and 1.
- An **experiment** is the process that results in an outcome.
- The **outcome** of an experiment is a result that we observe.
- The **sample space** is the collection of all possible outcomes of an experiment.

Definitions of Probability

Probabilities may be defined from one of three perspectives:

- Classical definition: probabilities can be deduced from theoretical arguments
- Relative frequency definition: probabilities are based on empirical data
- Subjective definition: probabilities are based on judgment and experience

Example 5.1 Classical Definition of Probability

Roll 2 dice

- 36 possible rolls (1,1), (1,2),…(6,5), (6,6)
- Probability = number of ways of rolling a number divided by 35; e.g., probability of a 3 is 2/36

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

Suppose two consumers try a new product.

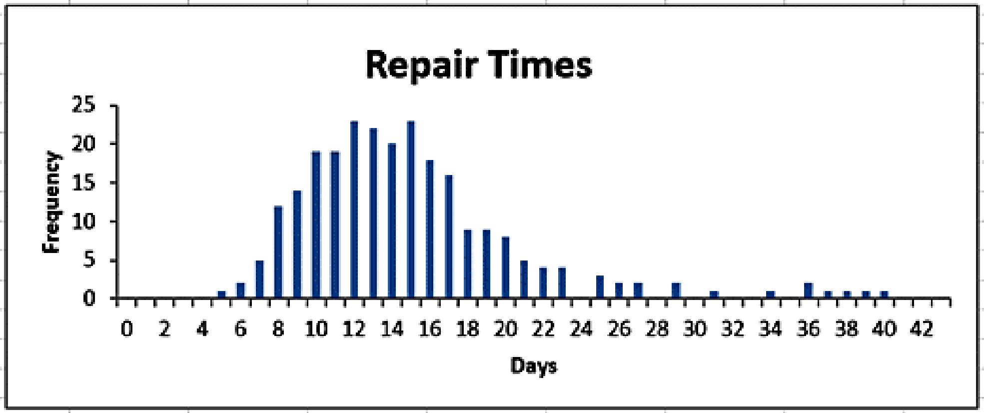
- Four outcomes:
  1. like, like
  2. like, dislike
  3. dislike, like
  4. dislike, dislike
- Probability at least one dislikes product = 3/4

Example 5.2: Relative Frequency Definition of Probability

- Use relative frequencies as probabilities
- Probability a computer is repaired in 10 days = 0.076

Computer Repair Times

Days	Frequency	Relative Frequency	Cumulative Percentage
0	0	0.000	0.0%
1	0	0.000	0.0%
2	0	0.000	0.0%
3	0	0.000	0.0%
4	0	0.000	0.0%
5	1	0.004	0.4%
6	2	0.008	1.2%
7	5	0.020	3.2%
8	12	0.048	8.0%
9	14	0.056	13.6%
10	19	0.076	21.2%
11	19	0.076	28.8%
12	23	0.092	38.0%
13	22	0.088	46.8%



Probability Rules and Formulas

- Label the n outcomes in a sample space as $O_1, O_2, \ldots, O_n$, where O_i represents the i[th] outcome in the sample space. Let $P(O_i)$ be the probability associated with the outcome O_i.
- The probability associated with any outcome must be between 0 and 1.

$$0 \leq P(O_i) \leq 1 \text{ for each outcome } O_i \qquad (5.1)$$

- The sum of the probabilities over all possible outcomes must be equal to 1.

$$P(O_1) + P(O_2) + \ldots + P(O_n) = 1 \qquad (5.2)$$

Probabilities Associated with Events

- An **event** is a collection of one or more outcomes from a sample space.
- <u>Rule 1</u>. The probability of any event is the sum of the probabilities of the outcomes that comprise that event.

Example 5.3: Computing the Probability of an Event

Consider the events:

- Rolling 7 or 11 on two dice

  Probability = 6/36 + 2/36 = 8/36.

- Repair a computer in 7 days or less

  Probability =

  $= O_1 + O_2 + O_3 + O_4 + O_5 + O_6 + O_7$

  = 0 + 0 + 0 + 0 + .004 + .008 + .020

  = 0.032

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

Computer Repair Times

Days	Frequency	Relative Frequency
0	0	0.000
1	0	0.000
2	0	0.000
3	0	0.000
4	0	0.000
5	1	0.004
6	2	0.008
7	5	0.020
8	12	0.048

Complement of an Event

- If A is any event, the complement of *A*, denoted A^c, consists of all outcomes in the sample space not in *A*.
- Rule 2. The probability of the complement of any event *A* is $P(A^c) = 1 - P(A)$.

Example 5.4: Computing the Probability of the Complement of an Event

Dice example:

- $A = \{7, 11\}$

  $P(A) = 8/36$

- $A^c = \{2, 3, 4, 5, 6, 8, 9, 10, 12\}$
- Using Rule 2:

$P(A^c) = 1 - 8/36 = 28/36$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

Union of Events

- The union of two events contains all outcomes that belong to either of the two events.
  - If A and B are two events, the probability that some outcome in either A or B (that is, the union of A and B) occurs is denoted as $P(A \text{ or } B)$.
- Two events are mutually exclusive if they have no outcomes in common.
- Rule 3. If events A and B are mutually exclusive, then $P(A \text{ or } B) = P(A) + P(B)$.

Example 5.5: Computing the Probability of Mutually Exclusive Events

Dice Example:

- $A = \{7, 11\}$: $P(A) = 8/36$
- $B = \{2, 3, 12\}$: $P(B) = 4/36$
- $P(A \text{ or } B)$ = Union of events A and B

$$= P(A) + P(B)$$

$$= 8/36 + 4/36 = 12/36$$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

Non-Mutually Exclusive Events

- The notation *(A and B)* represents the intersection of events *A* and *B* – that is, all outcomes belonging to both A and B .
- Rule 4. If two events *A* and *B* are not mutually exclusive, then $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

Example 5.6: Computing the Probability of Non-Mutually Exclusive Events

Dice Example:

- $A = \{2, 3, 12\}$: $P(A) = 4/36$
- B = {even number} : $P(B) = 18/36$
- $(A \text{ and } B) = \{2, 12\}$: $P(A \text{ and } B) = 2/36$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

  $= 4/36 + 18/36 - 2/36$

  $= 20/36$

Die Sum	Frequency
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1
Sum	36

Joint and Marginal Probability

- The probability of the intersection of two events is called a **joint probability.**
- The probability of an event, irrespective of the outcome of the other joint event, is called a **marginal probability.**

Application of Joint and Marginal Probability

- A sample of 100 individuals were asked to evaluate their preference for three new proposed energy drinks in a blind taste test.
- The sample space consists of two types of outcomes corresponding to each individual: gender (F = female or M = male) and brand preference (B_1, B_2, or B_3).
- Define a new sample space consisting of the outcomes that reflect the different combinations of outcomes from these two sample spaces.
  - O_1 = the respondent is female and prefers brand 1
  - O_2 = the respondent is female and prefers brand 2
  - O_3 = the respondent is female and prefers brand 3
  - O_4 = the respondent is male and prefers brand 1
  - O_5 = the respondent is male and prefers brand 2
  - O_6 = the respondent is male and prefers brand 3
- The probability of each of these events is the intersection of the gender and brand preference event. For example, $P(O_1) = P(F \text{ and } B_1)$

Example 5.7: Applying Probability Rules to Joint Events

- *Energy Drink Survey*
- The joint probabilities of gender and brand preference are calculated by dividing the number of respondents corresponding to each of the six outcomes listed above by the total number of respondents, 100.
  - E.g., $P(F \text{ and } B_1) = P(O_1) = 9/100 = 0.09$

	A	B	C	D	E	F	G	H	I
1	**Energy Drink Survey**								
2									
3	**Respondent**	**Gender**	**Brand Preference**						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2		Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
12	9	Male	Brand 1		Female	0.09	0.06	0.22	0.37
13	10	Female	Brand 3		Male	0.25	0.17	0.21	0.63
14	11	Male	Brand 3		Grand Total	0.34	0.23	0.43	1
15	12	Male	Brand 2						
16	13	Female	Brand 3						

Joint probabilities

Example 5.7: Continued

- The marginal probabilities for gender and brand preference are calculated by adding the joint probabilities across the rows and columns
  - E.g., the event F, (respondent is female) is comprised of the outcomes O_1, O_2, and O_3, and therefore $P(F) = P(F \text{ and } B_1) + P(F \text{ and } B_2) + P(F \text{ and } B_3) = 0.37$

	A	B	C	D	E	F	G	H	I
1	Energy Drink Survey								
2									
3	Respondent	Gender	Brand Preference						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2		Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
12	9	Male	Brand 1		Female	0.09	0.06	0.22	0.37
13	10	Female	Brand 3		Male	0.25	0.17	0.21	0.63
14	11	Male	Brand 3		Grand Total	0.34	0.23	0.43	1
15	12	Male	Brand 2						
16	13	Female	Brand 3						

Marginal probabilities

Joint/Marginal Probability Rule

- Calculations of marginal probabilities leads to the following probability rule:

- Rule 5. If event A is comprised of the outcomes $\{A_1, A_2, \ldots, A_n\}$ and event B is comprised of the outcomes $\{B_1, B_2, \ldots, B_n\}$, then

$$P(A_i) = P(A_i \text{ and } B_1) + P(A_i \text{ and } B_2) + \ldots + P(A_i \text{ and } B_n)$$

Example 5.7 Continued

- Events F and M are mutually exclusive, as are events B_1, B_2, and B_3 since a respondent may be only male or female and prefer exactly one of the three brands. We can use Rule 3 to find, for example, $P(B_1 \text{ or } B_2) = 0.34 + 0.23 = 0.57$.
- Events F and B_1, however, are not mutually exclusive because a respondent can be both female and prefer brand 1. Therefore, using Rule 4, we have $P(F \text{ or } B_1) = P(F) + P(B_1) – P(F \text{ and } B_1) = 0.37 + 0.34 – 0.09 = 0.62$.

	A	B	C	D	E	F	G	H	I
1	Energy Drink Survey								
2									
3	Respondent	Gender	Brand Preference						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2		Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
12	9	Male	Brand 1		Female	0.09	0.06	0.22	0.37
13	10	Female	Brand 3		Male	0.25	0.17	0.21	0.63
14	11	Male	Brand 3		Grand Total	0.34	0.23	0.43	1
15	12	Male	Brand 2						
16	13	Female	Brand 3						

Conditional Probability

- Conditional probability is the probability of occurrence of one event *A*, given that another event *B* is known to be true or has already occurred.

Example 5.8 Computing a Conditional Probability in a Cross-Tabulation

- Suppose we know a respondent is male. What is the probability that he prefers Brand 1?
- <u>Using cross-tabulation</u>: Of 63 males, 25 prefer Brand 1, so the probability of preferring Brand 1 given that a respondent is male = 25/63
- <u>Using joint probability table</u>: divide the joint probability 0.25 (the probability that the respondent is male and prefers brand 1) by the marginal probability 0.63 (the probability that the respondent is male).

	A	B	C	D	E	F	G	H	I
1	Energy Drink Survey								
2									
3	Respondent	Gender	Brand Preference						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2		Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
12	9	Male	Brand 1		Female	0.09	0.06	0.22	0.37
13	10	Female	Brand 3		Male	0.25	0.17	0.21	0.63
14	11	Male	Brand 3		Grand Total	0.34	0.23	0.43	1
15	12	Male	Brand 2						
16	13	Female	Brand 3						

Example 5.9: Conditional Probability in Marketing

- *Apple Purchase History*
- The PivotTable shows the count of the type of second purchase given that each product was purchased first.
- Probability of purchasing an iPad given that a customer already purchased an iMac = 2/13

	A	B
1	Apple Products Purchase History	
2		
3	First Purchase	Second Purchase
4	iPod	iMac
5	iPhone	MacBook
6	iMac	iPhone
7	iPhone	iPod
8	iPod	iPhone
9	MacBook	iPod
10	iPhone	MacBook
11	MacBook	iPhone
12	iPod	MacBook

	A	B	C	D	E	F	G
1							
2							
3	Count of Second Purchase	Column Labels					
4	Row Labels	iMac	iPad	iPhone	iPod	MacBook	Grand Total
5	iMac		2	3	2	6	13
6	iPad	1		1	2	10	14
7	iPhone	3	4		14	21	42
8	iPod	3	12	12		30	57
9	MacBook	8	16	26	24		74
10	Grand Total	15	34	42	42	67	200

Conditional Probability Formula

- The conditional probability of an event A given that event B is known to have occurred is

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \qquad (5.3)$$

- We read the notation *P(A|B)* as "the probability of *A* given *B*."

Example 5.10: Using the Conditional Probability Formula

Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
Female	0.09	0.06	0.22	0.37
Male	0.25	0.17	0.21	0.63
Grand Total	**0.34**	**0.23**	**0.43**	**1**

- $P(B_1|M) = P(B_1 \text{ and } M)/\ P(M) = (0.25)/(0.63) = 0.397$
- $P(B_1|F) = P(B_1 \text{ and } F)/\ P(F) = (0.09)/(0.37) = 0.243$
- Summary of conditional probabilities:

P(Brand|Gender)	Brand 1	Brand 2	Brand 3
Male	0.397	0.270	0.333
Female	0.243	0.162	0.595

- Applications in marketing and advertising.

Variations of the Conditional Probability Formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \tag{5.3}$$

- *P(A and B) = P(A | B) P(B)*
- *P(B and A) = P(B | A) P(A)*
  - Note: *P(A and B) = P(B and A)*

- Multiplication law of probability:

$$P(A \text{ and } B) = P(A|B)\,P(B) = P(B|A)\,P(A) \tag{5.4}$$

Extension of the Multiplication Law

- Suppose $B_1, B_2, \ldots, B_n$ are mutually exclusive events whose union comprises the entire sample space. Then

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n) \qquad \textbf{(5.5)}$$

Example 5.11: Using the Multiplication Law of Probability

- "Texas Hold 'Em" Poker
- Probability of pocket aces (two aces in hand)
- A_1 = Ace on first card; A_2 = Ace on second card
- $P(A_1 \text{ and } A_2) = P(A_2|A_1)\, P(A_1)$

$$= (3/51)\ (4/52)$$

$$= 0.004525$$

Independent Events

- Two events *A* and *B* are **independent** if
  $P(A \mid B) = P(A)$.
- *Energy Drink Survey* example: the probability of preferring a brand depends on gender.
- Thus, we may say that brand preference and gender are not independent.

Example 5.12: Determining if Two Events are Independent

- Are Gender and Brand Preference Independent?
- $P(B_1) = 0.34$

Joint Probability Table	Brand 1	Brand 2	Brand 3	Grand Total
Female	0.09	0.06	0.22	0.37
Male	0.25	0.17	0.21	0.63
Grand Total	**0.34**	**0.23**	**0.43**	**1**

- $P(B_1|M) = 0.397$

P(Brand|Gender)	Brand 1	Brand 2	Brand 3
Male	0.397	0.270	0.333
Female	0.243	0.162	0.595

- Because $0.397 \neq 0.34$, Gender and Brand Preference are not independent.

Multiplication Law for Independent Events

- If two events are independent, then we can simplify the multiplication law of probability in equation (5.4)

$$P(A \text{ and } B) = P(A \mid B)\, P(B) = P(B \mid A)\, P(A) \qquad \textbf{(5.4)}$$

- by substituting *P(A)* for *P(A | B)*:

$$P(A \text{ and } B) = P(B)\, P(A) = P(A)P(B) \qquad \textbf{(5.6)}$$

Example 5.13: Using the Multiplication Law for Independent Events

Dice Rolls:

- Rolling pairs of dice are independent events since they do not depend on the previous rolls.
- A = {roll a sum of 6 on first roll}
- B = {roll a sum of 2, 3, or 12 on second roll}
- Using formula (5.5): $P(A \text{ and } B) = P(A)\,P(B)$

$$= (5/36)\,(4/36) = 0.0154$$

Random Variables

- A **random variable** is a numerical description of the outcome of an experiment.
- A **discrete random variable** is one for which the number of possible outcomes can be counted.
- A **continuous random variable** has outcomes over one or more continuous intervals of real numbers.

Example 5.14: Discrete and Continuous Random Variables

Examples of <u>discrete</u> random variables:

- outcomes of dice rolls
- whether a customer likes or dislikes a product
- number of hits on a Web site link today

Examples of <u>continuous</u> random variables:

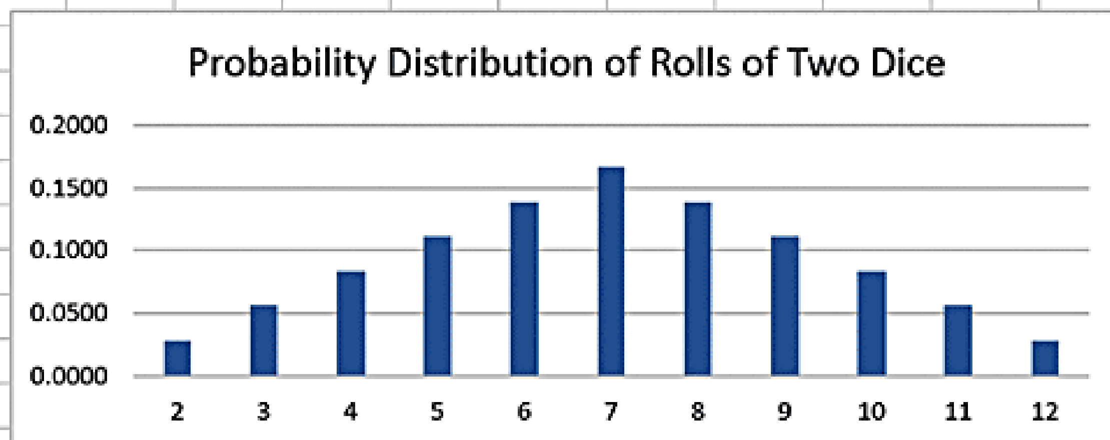
- weekly change in DJIA
- daily temperature
- time between machine failures

Probability Distributions

- A **probability distribution** is a characterization of the possible values that a random variable may assume along with the probability of assuming these values.
- We may develop a probability distribution using any one of the three perspectives of probability: classical, relative frequency, and subjective.

Example 5.14 Probability Distribution of Dice Rolls

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Dice Roll Probabilities**												
2													
3	**Outcome**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
4	**Number of Ways**	1	2	3	4	5	6	5	4	3	2	1	36
5	**Probability**	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278	1



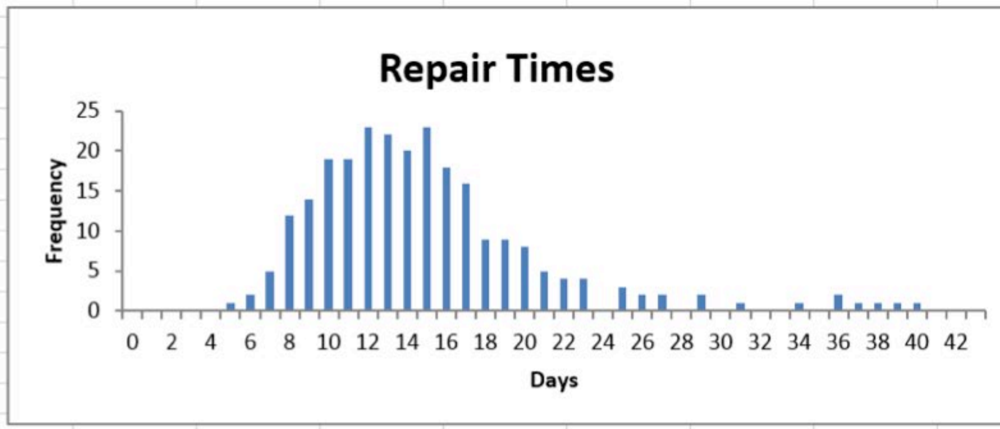
Empirical Probability Distributions

- We can calculate the relative frequencies from a sample of empirical data to develop a probability distribution. Because this is based on sample data, we usually call this an **empirical probability distribution**.
- An empirical probability distribution is an approximation of the probability distribution of the associated random variable, whereas the probability distribution of a random variable, such as one derived from counting arguments, is a theoretical model of the random variable.

Empirical Probability Distribution Example

Computer Repair Times

Days	Frequency	Relative Frequency	Cumulative Percentage
0	0	0.000	0.0%
1	0	0.000	0.0%
2	0	0.000	0.0%
3	0	0.000	0.0%
4	0	0.000	0.0%
5	1	0.004	0.4%
6	2	0.008	1.2%
7	5	0.020	3.2%
8	12	0.048	8.0%
9	14	0.056	13.6%
10	19	0.076	21.2%
11	19	0.076	28.8%
12	23	0.092	38.0%
13	22	0.088	46.8%



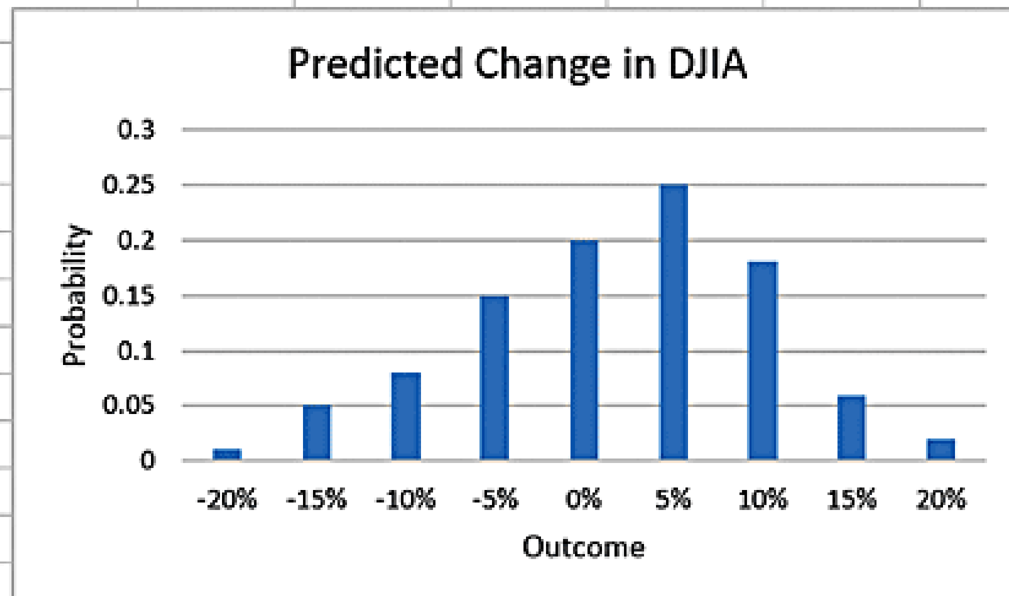
Subjective Probability Distributions

- We could simply specify a probability distribution using subjective values and expert judgment.
- This is often done in creating decision models for phenomena for which we have no historical data.

Example 5.16: A Subjective Probability Distribution

- Distribution of an expert's assessment of how the DJIA might change next year.

	A	B
1	Change in DJIA	Subjective Probability
2	-20%	0.01
3	-15%	0.05
4	-10%	0.08
5	-5%	0.15
6	0%	0.2
7	5%	0.25
8	10%	0.18
9	15%	0.06
10	20%	0.02



Discrete Probability Distributions

- For a discrete random variable *X*, the probability distribution of the discrete outcomes is called a **probability mass function** and is denoted by a mathematical function, *f(x)*.
  - The symbol x_i represents the i^{th} value of the random variable *X* and $f(x_i)$ is the probability.
- Properties:
  - the probability of each outcome must be between 0 and 1
  - the sum of all probabilities must add to 1

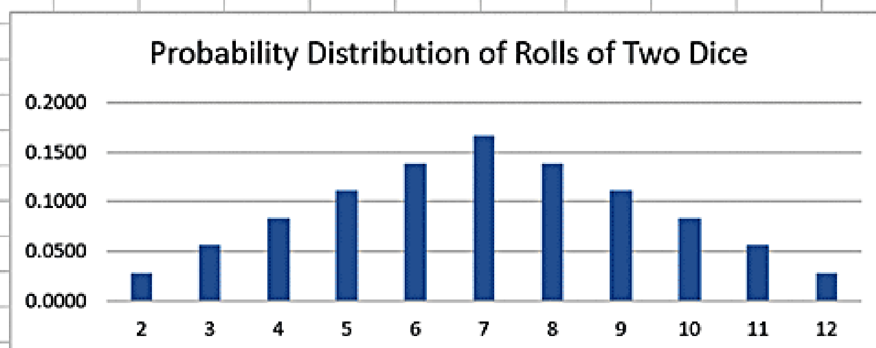
$$0 \leq f(x_i) \leq 1 \quad \text{for all } i \tag{5.7}$$

$$\sum_i f(x_i) = 1 \tag{5.8}$$

Example 5.17: Probability Mass Function for Rolling Two Dice

- x_i = values of the random variable X, which represents sum of the rolls of two dice
  - $x_1 = 2$, $x_2 = 3$, …, $x_{10} = 11$, $x_{11} = 12$
- $f(x_1) = 1/36 = 0.0278$; $f(x_2) = 2/36 = 0.0556$, etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Dice Roll Probabilities**												
2													
3	**Outcome**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
4	**Number of Ways**	1	2	3	4	5	6	5	4	3	2	1	36
5	**Probability**	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278	1



Cumulative Distribution Function

- A cumulative distribution function, $F(x)$, specifies the probability that the random variable X assumes a value less than or equal to a specified value, x; that is,

$$F(x) = P(X \leq x)$$

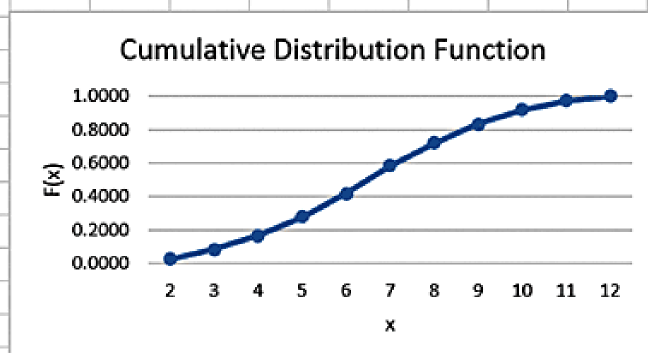
Example 5.18: Using the Cumulative Distribution Function

- Probability of rolling a 6 or less = $F(6) = 0.1667$
- Probability of rolling between 4 and 8:

$$= P(4 \le X \le 8) = P(3 < X \le 8) = P(X \le 8) - P(X \le 3)$$

$$= 0.7222 - 0.0833 = 0.6389$$

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Dice Roll Probabilities**												
2													
3	**Outcome**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
4	**Number of Ways**	**1**	**2**	**3**	**4**	**5**	**6**	**5**	**4**	**3**	**2**	**1**	**36**
5	**Probability**	**0.0278**	**0.0556**	**0.0833**	**0.1111**	**0.1389**	**0.1667**	**0.1389**	**0.1111**	**0.0833**	**0.0556**	**0.0278**	**1**
6	**Cumulative Probability**	**0.0278**	**0.0833**	**0.1667**	**0.2778**	**0.4167**	**0.5833**	**0.7222**	**0.8333**	**0.9167**	**0.9722**	**1.0000**	



Expected Value of a Discrete Random Variable

- The expected value of a random variable corresponds to the notion of the mean, or average, for a sample.
- For a discrete random variable *X*, the expected value, denoted *E[X]*, is the weighted average of all possible outcomes, where the weights are the probabilities:

$$E[X] = \sum_{i=1}^{\infty} x_i f(x_i) \tag{5.9}$$

Example 5.19: Computing the Expected Value

- Rolling two dice
  - $E[X] = 2(0.0278) + 3(0.0556) + 4(0.0833) + 5(0.1111) + 6(0.1389) + 7(0.1667) + 8(0.1389) + 9(0.1111) + 10(0.0833) + 11(0.0556) + 12(0.0278) = 7$

	A	B	C
1	**Expected Value Calculations**		
2			
3	**Outcome, x**	**Probability, f(x)**	**x*f(x)**
4	2	0.0278	0.0556
5	3	0.0556	0.1667
6	4	0.0833	0.3333
7	5	0.1111	0.5556
8	6	0.1389	0.8333
9	7	0.1667	1.1667
10	8	0.1389	1.1111
11	9	0.1111	1.0000
12	10	0.0833	0.8333
13	11	0.0556	0.6111
14	12	0.0278	0.3333
15		Expected value	7.0000

Example 5.20: Expected Value on Television

The Apprentice

- Teams were required to select an artist (mainstream or avant-garde) and sell their art for the most money possible. A back-of-the-envelope expected value calculation would have easily predicted the winner.

Deal or No Deal

- Contestant had 5 briefcases left with $100, $400, $1000, $50,000 or $300,000 in them.
- Expected value of briefcases is $70,300.
- Banker offered contestant $80,000 to quit, which was higher than the expected value. The probability of choosing the $300,000 briefcase was only 0.2, so the decision should have been easy to make.

Expected Value and Decision Making

- The expected value is a “long-run average” and is appropriate for decisions that occur on a repeated basis.
- For one-time decisions, however, you need to consider the downside risk and the upside potential of the decision.

Example 5.21: Expected Value of a Charitable Raffle

x	$f(x)$
$-\$50$	0.999
$\$24,950$	0.001

- Cost of raffle ticket is $50
- 1000 raffle tickets are sold.
- Winning prize is $25,000
- $E[X] = -\$25$
- If you played this game repeatedly over the long run, you would lose an average of $25.00 each time you play.
- For any one game, you would either lose $50 or win $24,950.
  - Is the risk of losing $50 worth the potential of winning $24,950?

Example 5.22: Airline Revenue Management

- Full and discount airfares are available for a flight.
- Full-fare ticket costs \$560
- Discount ticket costs \$400
- X = ticket price paid
- $p = 0.75$ (the probability of selling a full-fare ticket)
- $E[X] = 0.75(\$560) + 0.25(0) = \$420$

- The airline should not discount full-fare tickets because the expected value of a full-fare ticket is greater than the cost of a discount ticket.
- Break-even point: $\$400 = p(\$560)$ or $p = 0.714$

Variance of a Discrete Random Variable

- The variance, *Var[X]*, of a discrete random variable *X* is a weighted average of the squared deviations from the expected value:

$$\mathrm{Var}[X] = \sum_{j=1}^{\infty} (x_j - E[X])^2 f(x_j) \tag{5.10}$$

Example 5.23: Computing the Variance of a Random Variable

- Rolling two dice

	A	B	C	D	E	F
1	Variance Calculations					
2						
3	Outcome, x	Probability, f(x)	x*f(x)	(x - E[X])	(x - E[X])^2	(x - E[X])^2*f(x)
4	2	0.0278	0.0556	-5.0000	25.0000	0.6944
5	3	0.0556	0.1667	-4.0000	16.0000	0.8889
6	4	0.0833	0.3333	-3.0000	9.0000	0.7500
7	5	0.1111	0.5556	-2.0000	4.0000	0.4444
8	6	0.1389	0.8333	-1.0000	1.0000	0.1389
9	7	0.1667	1.1667	0.0000	0.0000	0.0000
10	8	0.1389	1.1111	1.0000	1.0000	0.1389
11	9	0.1111	1.0000	2.0000	4.0000	0.4444
12	10	0.0833	0.8333	3.0000	9.0000	0.7500
13	11	0.0556	0.6111	4.0000	16.0000	0.8889
14	12	0.0278	0.3333	5.0000	25.0000	0.6944
15		Expected value	7.0000		Variance	5.8333

Bernoulli Distribution

- Two possible outcomes, "success" and "failure," each with a constant probability of occurrence; p is the probability of a success and $1 - p$ is the probability of a failure
- Typically, $x = 1$ represents "success" and $x = 0$ represents "failure"
- Probability mass function:

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} \quad \textbf{(5.11)}$$

- $E[X] = p$
- $\mathrm{Var}[X] = p(1 - p)$

Example 24: Using the Bernoulli Distribution

- The Bernoulli distribution can be used to model whether an individual responds positively (x = 1), or negatively (x = 0) to a telemarketing promotion.
- For example, if you estimate that 20% of customers contacted will make a purchase, the probability distribution that describes whether or not a particular individual makes a purchase is Bernoulli with p = 0.2

Binomial Distribution

- Models *n* independent replications of a Bernoulli experiment, each with a probability *p* of success.
  - *X* represents the number of successes in these *n* experiments
- Probability mass function:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{for } x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{(5.12)}$$

- The number of ways of choosing *x* distinct items from a group of *n* items and is

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} \qquad \textbf{(5.13)}$$

where *n*! (*n* factorial) = *n*(*n* - 1)(*n* - 2) . . . (2)(1), and 0! is defined as 1.

- Expected value = *np*; variance = *np(1 – p)*

Example 5.25: Computing Binomial Probabilities

- Suppose 10 individuals receive a telemarking promotion. Each individual has a 0.2 probability of making a purchase. Find the probability that exactly 3 of the 10 individuals make a purchase.
- The probability distribution that x individuals out of 10 calls will make a purchase is:

$$f(x) = \begin{cases} \binom{10}{x}(0.2)^x(0.8)^{10-x}, & \text{for } x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

- Excel function:

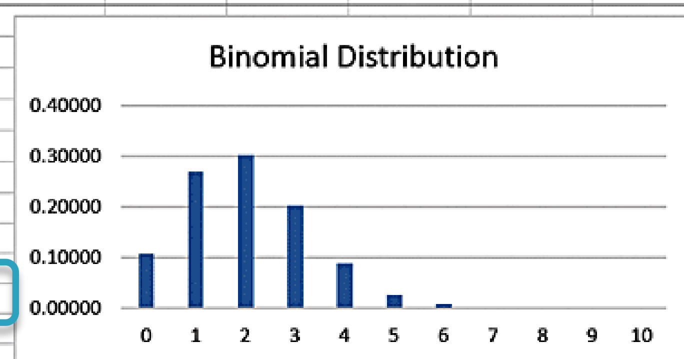
  =BINOM.DIST(*number_s, trials, probability_s, cumulative*)

- If *cumulative* is set to TRUE, then this function will provide cumulative probabilities; otherwise the default is FALSE, and it provides values of the probability mass function, *f(x)*.

Example 5.26: Using Excel's Binomial Distribution Function

- The probability that exactly 3 of 10 individuals will make a purchase is $P(x = 3)$: =BINOM.DIST(3,10,0.2,TRUE) = 0.20133
- The probability that 3 or fewer of 10 individuals will make a purchase is $P(x \leq 3)$: =BINOM.DIST(3,10,0.2,FALSE) = 0.87913

	A	B	C
1	**Binomial Probabilities**		
2			
3	**n**	10	
4	**p**	0.2	
5			
6	**x**	**f(x)**	**F(x)**
7	0	0.10737	0.10737
8	1	0.26844	0.37581
9	2	0.30199	0.67780
10	3	0.20133	0.87913
11	4	0.08808	0.96721
12	5	0.02642	0.99363
13	6	0.00551	0.99914
14	7	0.00079	0.99992
15	8	0.00007	1.00000
16	9	0.00000	1.00000
17	10	0.00000	1.00000



Shapes and Skewness of the Binomial Distribution

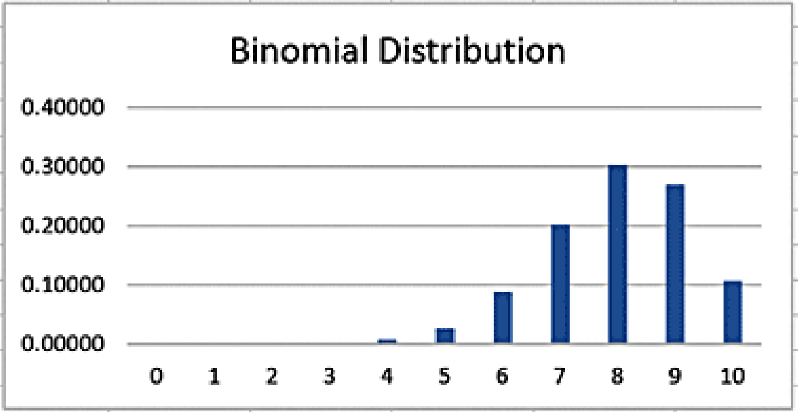
- The binomial distribution is symmetric when $p = 0.5$; positively skewed when $p < 0.5$, and negatively skewed when $p > 0.5$.

Example of negatively-skewed distribution

Binomial Probabilities

n	10
p	0.8

x	f(x)	F(x)
0	0.00000	0.00000
1	0.00000	0.00000
2	0.00007	0.00008
3	0.00079	0.00086
4	0.00551	0.00637
5	0.02642	0.03279
6	0.08808	0.12087
7	0.20133	0.32220
8	0.30199	0.62419
9	0.26844	0.89263
10	0.10737	1.00000



Poisson Distribution

- Models the number of occurrences in some unit of measure (often time or distance).
- There is no limit on the number of occurrences.
- The average number of occurrence per unit is a constant denoted as λ.
- Probability mass function:

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & \text{for } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

- Expected value = λ; variance = λ

Example 5.27: Computing Poisson Probabilities

- Suppose the average number of customers arriving at a Subway restaurant during lunch hour is λ =12 per hour.
- The probability that exactly *x* customers arrive during the hour is given by the Poisson distribution with a mean of 12.

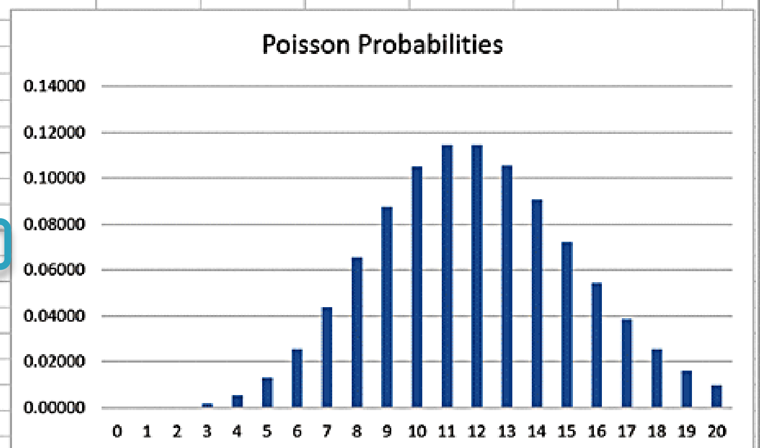
$$f(x) = \begin{cases} \dfrac{e^{-12}12^x}{x!}, & \text{for } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- Excel function: =POISSON.DIST(*x, mean, cumulative*)

Example 5.28: Using Excel's Poisson Distribution Function

- With $\lambda = 12$, the probability that $X = 1$ is
  =POISSON.DIST(A7,B3,FALSE) = 0.00007
- The probability that $X \leq 4$ is
  =POISSON.DIST(A10,B3,TRUE) = 0.00760

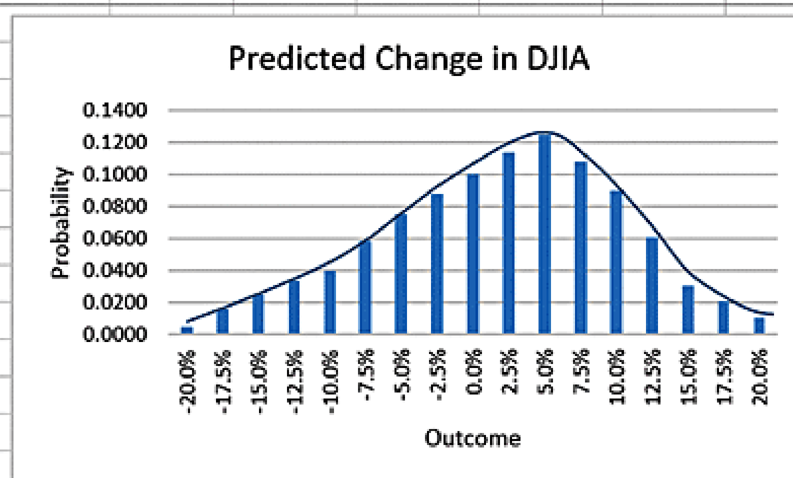
	A	B	C
1	**Poisson Probabilities**		
2			
3	**Mean**	12	
4			
5	**x**	**f(x)**	**F(x)**
6	0	0.00001	0.00001
7	1	0.00007	0.00008
8	2	0.00044	0.00052
9	3	0.00177	0.00229
10	4	0.00531	0.00760
11	5	0.01274	0.02034
12	6	0.02548	0.04582
13	7	0.04368	0.08950
14	8	0.06552	0.15503
15	9	0.08736	0.24239
16	10	0.10484	0.34723
17	11	0.11437	0.46160
18	12	0.11437	0.57597
19	13	0.10557	0.68154
20	14	0.09049	0.77202
21	15	0.07239	0.84442
22	16	0.05429	0.89871
23	17	0.03832	0.93703
24	18	0.02555	0.96258
25	19	0.01614	0.97872
26	20	0.00968	0.98840



Continuous Probability Distributions

- A **probability density function** is a mathematical function that characterizes a continuous random variable

	A	B
1	Change in DJIA	Subjective Probability
2	-20.0%	0.0050
3	-17.5%	0.0159
4	-15.0%	0.0250
5	-12.5%	0.0334
6	-10.0%	0.0403
7	-7.5%	0.0584
8	-5.0%	0.0750
9	-2.5%	0.0884
10	0.0%	0.1000
11	2.5%	0.1134
12	5.0%	0.1250
13	7.5%	0.1084
14	10.0%	0.0900
15	12.5%	0.0609
16	15.0%	0.0300
17	17.5%	0.0209
18	20.0%	0.0100



Continuous Probability Distributions

- Properties
  - $f(x) \geq 0$ for all values of x
  - Total area under the density function equals 1.
  - $P(X = x) = 0$
  - Probabilities are only defined over intervals.
  - $P(a \leq X \leq b)$ is the area under the density function between a and b.

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \tag{5.15}$$

Uniform Distribution

- The **uniform distribution** characterizes a continuous random variable for which all outcomes between a minimum (*a*) and a maximum (*b*) are equally likely.
- Density function:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \le x \le b \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{(5.16)}$$

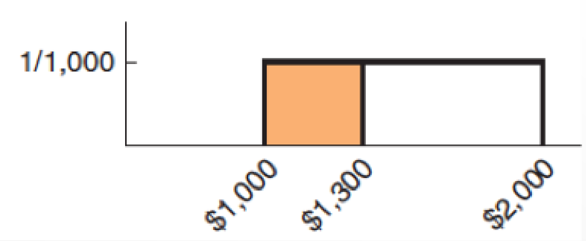
- Cumulative distribution function:

$$F(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \le x \le b \\ 1, & \text{if } b < x \end{cases} \qquad \textbf{(5.17)}$$

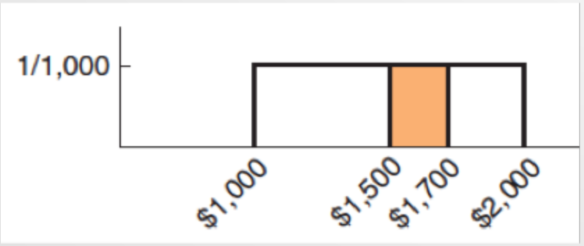
- Expected value = $(a + b)/2$; variance = $(b - a)^2/12$

Example 5.29: Computing Uniform Probabilities

- Sales revenue for a product varies uniformly each week between \$1000 and \$2000.
- Probability that sales revenue will be less than x = \$1,300.
  - $F(1{,}300) = (1{,}300 - 1{,}000) / (2{,}000 - 1{,}000) = 0.30$



- Probability that revenue will be between \$1,500 and \$1,700.
  - $P(1{,}500 \le X \le 1{,}700) = P(X \le 1{,}700) - P(X \le 1{,}500) = F(1{,}700) - F(1{,}500)$ $= F(1{,}700) - F(1{,}500) = 0.7 – 0.5 = 0.2$



Discrete Uniform Distribution

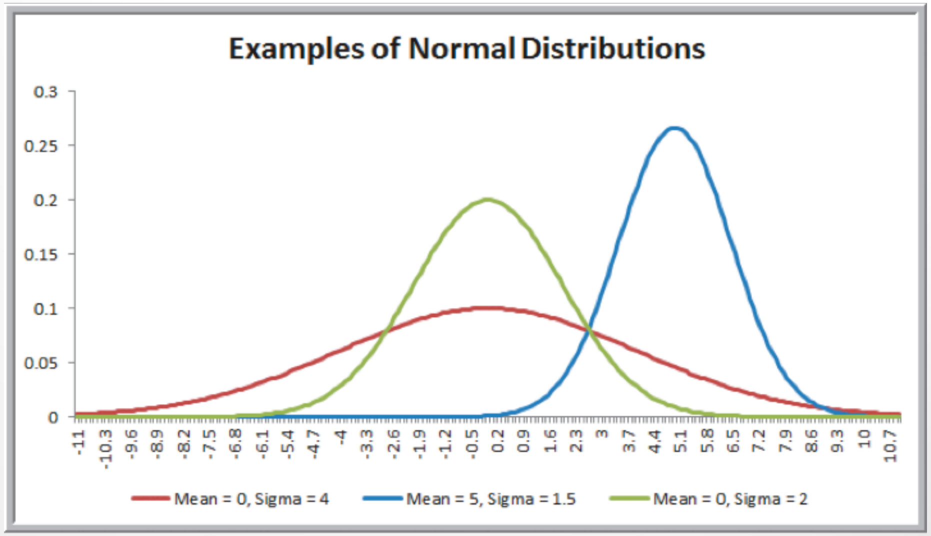
- A variation of the uniform distribution is one for which the random variable is restricted to integer values between *a* and *b* (also integers); this is called a **discrete uniform distribution**.
    - Example: roll of a single die. Each of the numbers 1 through 6 have a 1/6 probability of occurrence.

Normal Distribution

- $f(x)$ is a bell-shaped curve
- Characterized by 2 parameters:

  μ (mean)

  σ (standard deviation)
- Properties
  1. Symmetric
  2. Mean = Median = Mode
  3. Range of X is unbounded
  4. Empirical rules apply



Computing Normal Probabilities

- Excel function:

=NORM.DIST(*x, mean, standard_deviation, cumulative*).

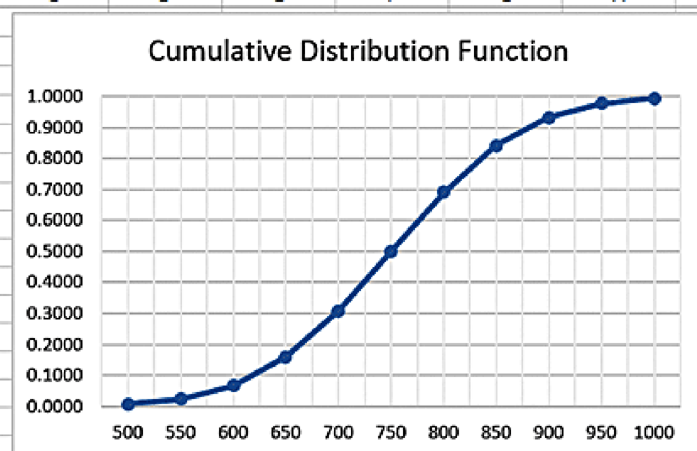
  - NORM.DIST(*x, mean, standard_deviation, TRUE*) calculates the cumulative probability
  - If cumulative is set to FALSE, the function simply calculates the value of the density function f(x), which has little practical application.

Example 5.30 Using the NORM.DIST Function to Compute Normal Probabilities

- The distribution for customer demand (units per month) is normal with mean = 750 and standard deviation = 100
- Find the probability that demand will be:

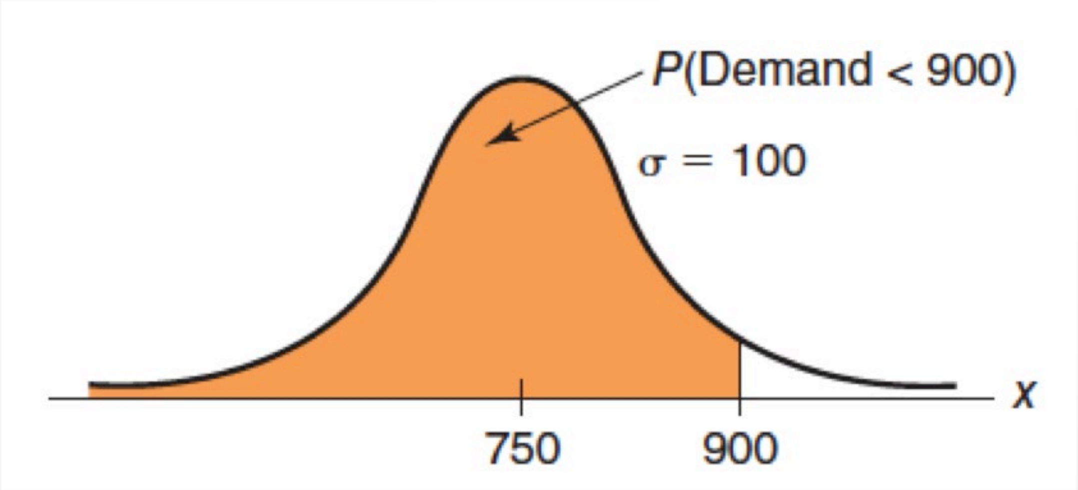
1. at most 900 units/month
2. exceed 700 units/month
3. be between 700 and 900 units/month

	A	B
1	Normal Probabilities	
2		
3	Mean	750
4	Standard Deviation	100
5		
6	x	F(x)
7	500	0.0062
8	550	0.0228
9	600	0.0668
10	650	0.1587
11	700	0.3085
12	750	0.5000
13	800	0.6915
14	850	0.8413
15	900	0.9332
16	950	0.9772
17	1000	0.9938



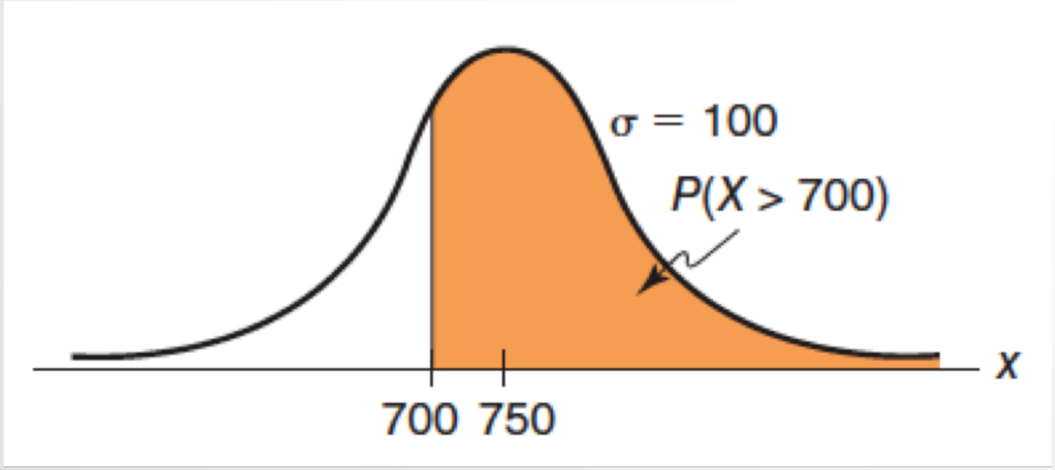
Example 5.30: Question 1

- Probability that demand will be at most 900 units, or $P(X \leq 900)$:
  - =NORM.DIST(900,750,100,TRUE) = 0.9332.



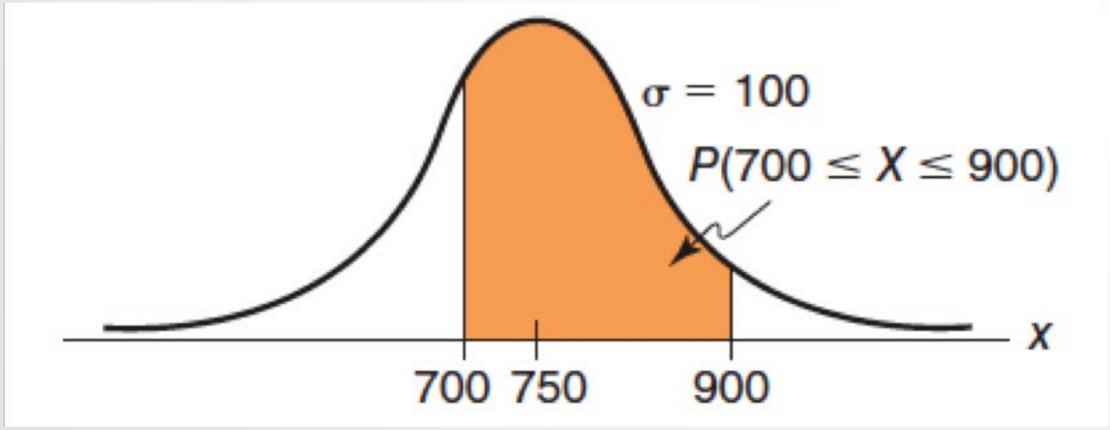
Example 5.30: Question 2

- Probability that demand will exceed 700 units, or P(X > 700).
  - =1-NORM.DIST(700,750,100,TRUE) = 1 - 0.3085 = 0.6915



Example 5.30: Question 3

- Probability that demand will be between 700 and 900, or $P(700 < X < 900)$:
  - =NORM.DIST(900,750,100,TRUE) - NORM.DIST(700,750,100,TRUE) =0.9332 - 0.3085 = 0.6247



The NORM.INV Function

- Normal Inverse function:

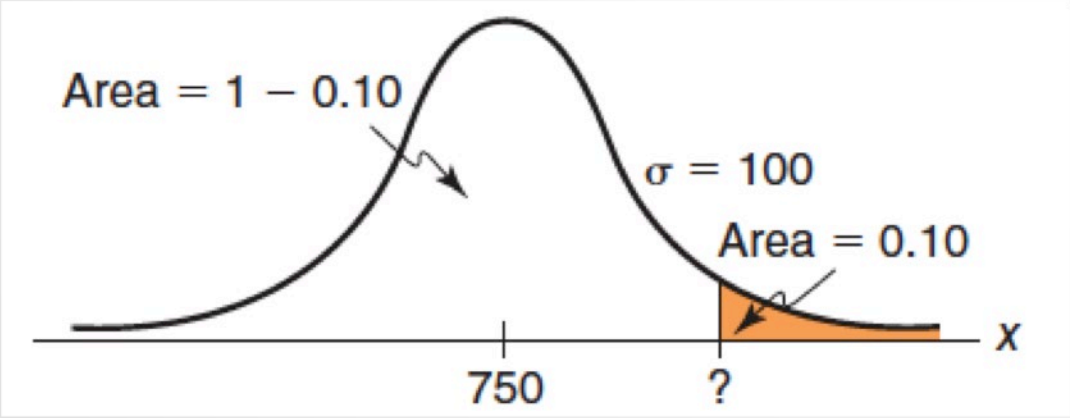
=NORM.INV(*probability, mean, stdev*)

provides the *x* value with *F(x)* = *probability*

Example 5.31: Using the NORM.INV Function

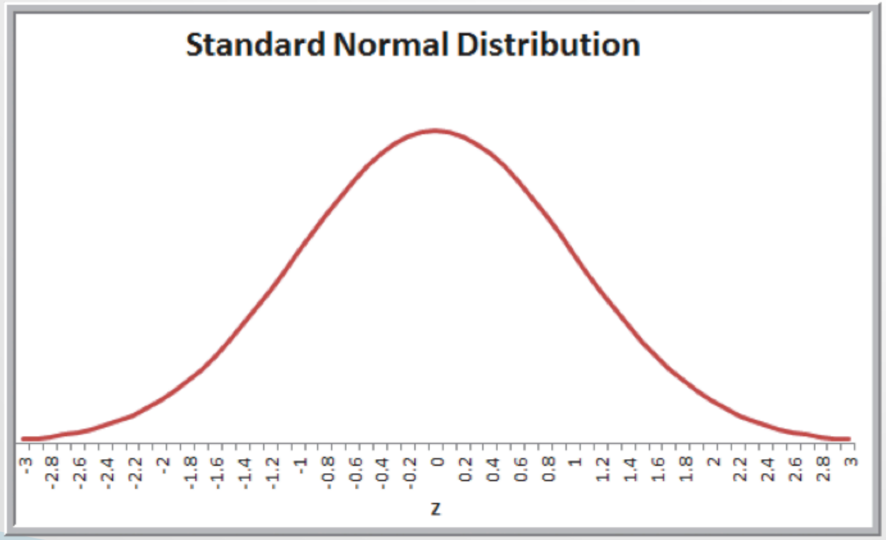
- What level of demand would be exceeded at most 10% of the time?
- Find x such that $F(x) = 0.90$:

  = NORM.INV(0.90, 750, 100) = 878.155



Standard Normal Distribution

- A standard normal distribution is a normal distribution with a mean of 0 and standard deviation of 1.
  - A standard normal random variable is denoted by *Z*.
  - The scale along the z-axis represents the number of standard deviations from the mean of zero.
  - The Excel function =NORM.S.DIST(z) finds probabilities for the standard normal distribution.



Example 5.32: Computing Probabilities with the Standard Normal Distribution

- Verify the empirical rules using Excel.
- Example: The probability within one standard deviation of the mean is $P(-1 < Z < 1)$

  = NORMS.DIST(1) – NORMS.DIST(-1)

  = 0.84134 – 0.15866

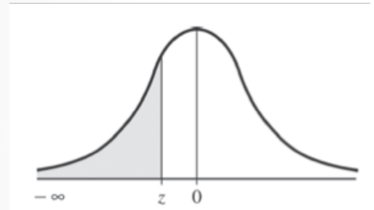
  = 0.6827

  ~ 68%

	A	B	C	D	E	F	G	H
1	**Standard Normal Probabilities**							
2								
3	**z**	**F(z)**		**a**	**b**	**F(a)**	**F(b)**	**F(b) - F(a)**
4	-3	0.00135		-1	1	0.15866	0.84134	0.6827
5	-2	0.02275		-2	2	0.02275	0.97725	0.9545
6	-1	0.15866		-3	3	0.00135	0.99865	0.9973
7	0	0.50000						
8	1	0.84134						
9	2	0.97725						
10	3	0.99865						

Using Standard Normal Distribution Tables

- Table 1 of Appendix A



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
−3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
−3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
−3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
−3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
−3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
−3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035

- We may compute probabilities for any normal random variable *X* having a mean μ and standard deviation σ by converting it to a standard normal random variable *Z*:

$$z = \frac{(x - \mu)}{\sigma} \tag{5.20}$$

Example 5.33: Computing Probabilities with Standard Normal Tables

- From Example 5.30, what is the probability that demand will be at least 900 units/month?
- $z = (900 - 750)/100 = 1.50$
- Using Table 1 in Appendix A, we find:
  - $P(X < 900) = P(Z < 1.50) = 0.93319$

z	.00
0.0	.5000
0.1	.5398
0.2	.5793
0.3	.6179
0.4	.6554
0.5	.6915
0.6	.7257
0.7	.7580
0.8	.7881
0.9	.8159
1.0	.8413
1.1	.8643
1.2	.8849
1.3	.9032
1.4	.9192
1.5	.9332

Exponential Distribution

- Models the time between randomly occurring events
- Density function:

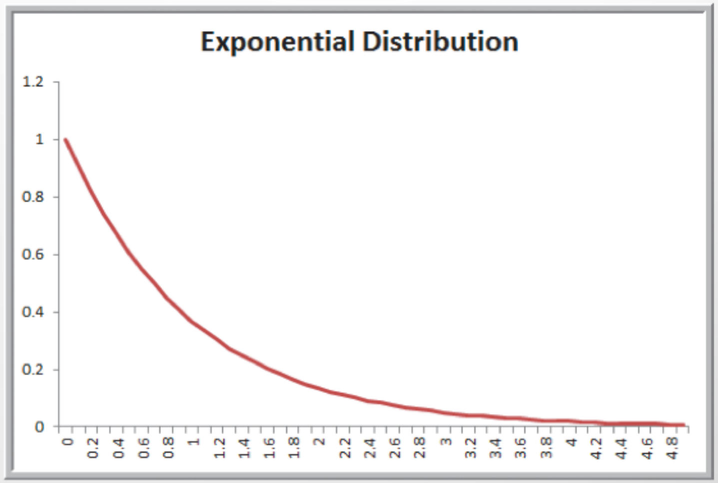
$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0 \qquad \textbf{(5.21)}$$

- Cumulative distribution function:

$$F(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0 \qquad \textbf{(5.22)}$$

- Mean = μ = 1/λ
- Excel function:
  - =EXPON.DIST(*x, lambda, cumulative*)

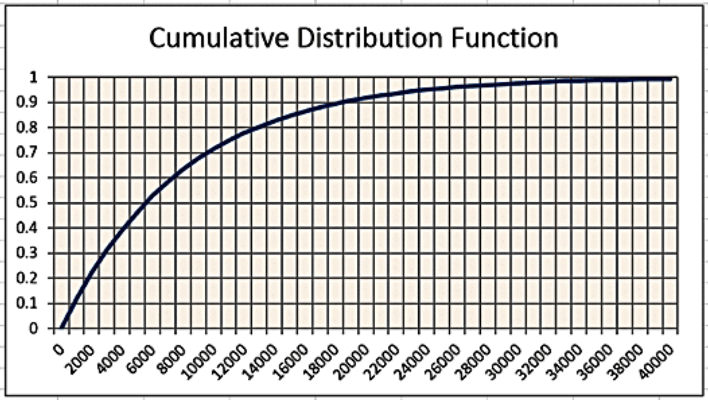
If the number of events occurring during an interval of time has a Poisson distribution, then the time between events is exponentially distributed.



Example 5.34: Using the Exponential Distribution

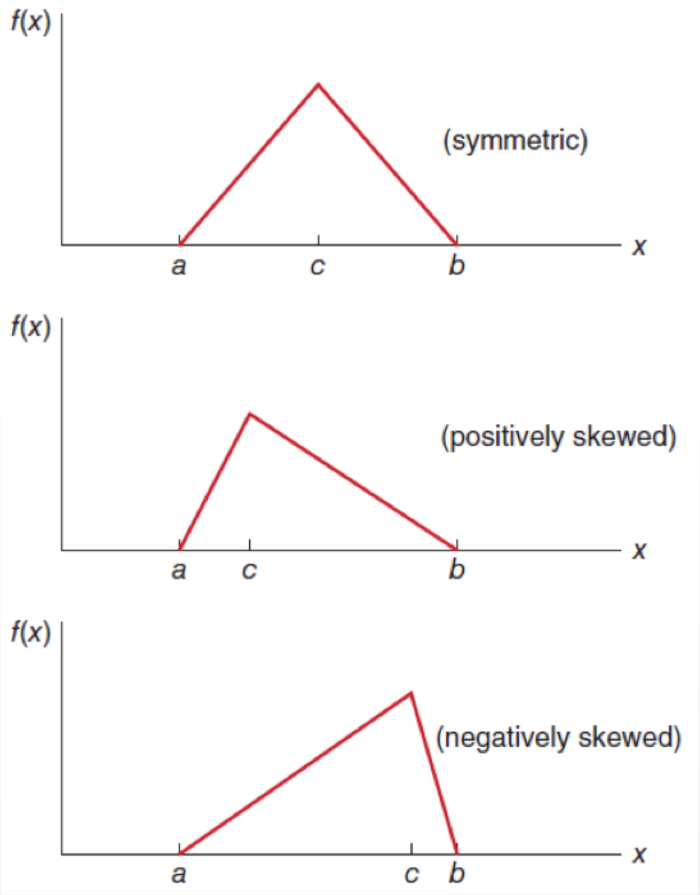
- The mean time to failure of a critical engine component is μ = 8,000 hours. What is the probability of failing before 5000 hours?
- $P(X < x)$ =EXPON.DIST(*x, lambda, cumulative*)
- $\lambda = 1/8000$
- $P(X < 5000)$ =EXPON.DIST(5000, 1/8000, TRUE)
  = 0.4647

	A	B
1	**Exponential Probabilities**	
2		
3	**Mean**	8000
4	**Lambda**	0.000125
5		
6	**x**	**F(x)**
7	0	0
8	1000	0.117503
9	2000	0.221199
10	3000	0.312711
11	4000	0.393469
12	5000	0.464739
13	6000	0.527633
14	7000	0.583138
15	8000	0.632121
16	9000	0.675348
17	10000	0.713495
18	11000	0.74716
19	12000	0.77687
20	13000	0.803088
21	14000	0.826226
22	15000	0.846645
23	16000	0.864665



Other Useful Distributions

- Triangular Distribution
- Lognormal Distribution
- Beta Distribution



Random Sampling from Probability Distributions

- A random number is one that is uniformly distributed between 0 to 1.
  - Excel function: =RAND( )

	A	B
1	Random Numbers	
2		
3	Sample	Random Number
4	1	0.326510048
5	2	0.743390121
6	3	0.801687688
7	4	0.804777187
8	5	0.848401291
9	6	0.614517898
10	7	0.452136913
11	8	0.600374163
12	9	0.533963502
13	10	0.638112424

Example 5.35: Sampling from the Distribution of Dice Outcomes

Probability distribution

x	$f(x)$	$F(x)$
2	0.028	0.028
3	0.056	0.083
4	0.083	0.167
5	0.111	0.278
6	0.139	0.417
7	0.167	0.583
8	0.139	0.722
9	0.111	0.833
10	0.083	0.917
11	0.056	0.972
12	0.028	1.000

Intervals for random sampling

Interval	Outcome
0 to 0.028	2
0.028 to 0.083	3
0.083 to 0.167	4
0.167 to 0.278	5
0.278 to 0.417	6
0.417 to 0.583	7
0.583 to 0.722	8
0.722 to 0.833	9
0.833 to 0.917	10
0.917 to 0.972	11
0.972 to 1	12

1. Generate a random number
2. Find the interval in which it falls
3. Use the associated outcome as the sample

Example 5.36: Using the VLOOKUP Function

- Sample from the probability distribution of predicted change in the Dow Jones Industrial Average index
- Compute *F(x)* and assign intervals to outcomes
- Generate random numbers using the Excel function =RAND( )
    - E.g. Cell J2: =VLOOKUP(I2,$E2:$G$10,3)

	A	B	C	D	E	F	G	H	I	J
1	Change in DJIA	f(x)	F(x)		Interval		Change in DJIA		Random Number	Outcome
2	-20%	0.01	0.01		0	0.01	-20%		0.530612386	5%
3	-15%	0.05	0.06		0.01	0.06	-15%		0.232776591	-5%
4	-10%	0.08	0.14		0.06	0.14	-10%		0.780924503	10%
5	-5%	0.15	0.29		0.14	0.29	-5%		0.363267546	0%
6	0%	0.2	0.49		0.29	0.49	0%		0.489479718	0%
7	5%	0.25	0.74		0.49	0.74	5%		0.062832805	-10%
8	10%	0.18	0.92		0.74	0.92	10%		0.53878251	5%
9	15%	0.06	0.98		0.92	0.98	15%		0.52525315	5%
10	20%	0.02	1		0.98	1	20%		0.99381738	20%
11									0.840872917	10%

Sampling from Common Probability Distributions

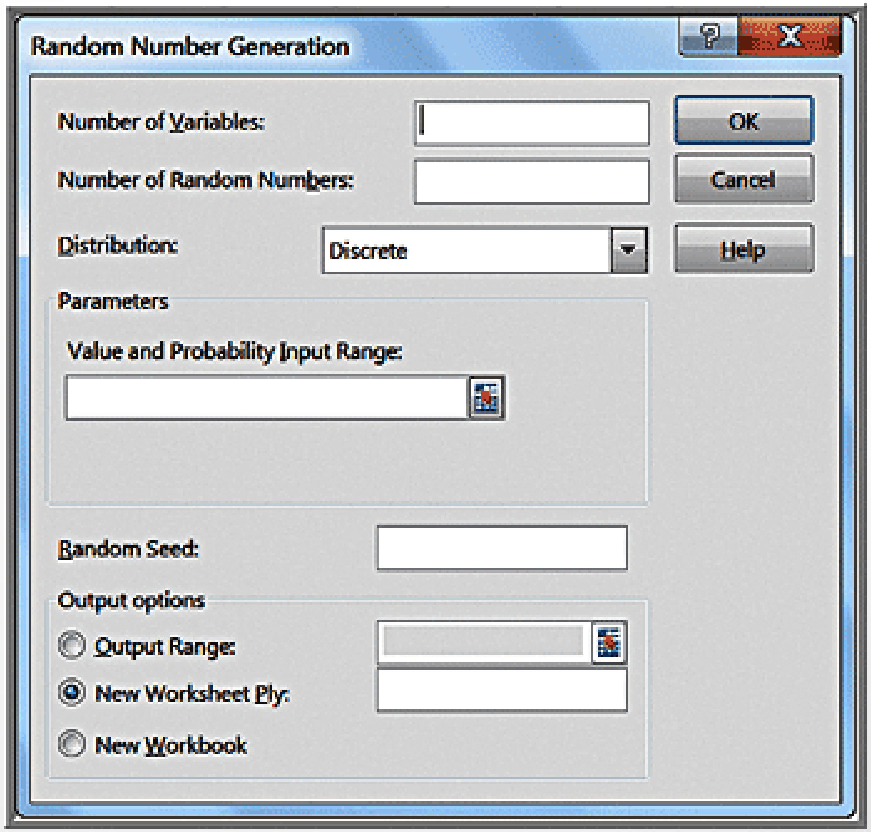
- A value randomly generated from a specified probability distribution is called a **random variate**.
  - Example: Uniform distribution

$$U = a + (b - a)*\mathrm{RAND}() \qquad (5.23)$$

- *Analysis Toolpak Random Number Generation* Tool
  - Can sample from uniform, normal, Bernoulli, binomial, Poisson, patterned, and discrete distributions.
  - Can also specify a **random number seed** – a value from which a stream of random numbers is generated. By specifying the same seed, you can produce the same random numbers at a later time.

Example 5.37: Using Excel's *Random Number Generation* Tool

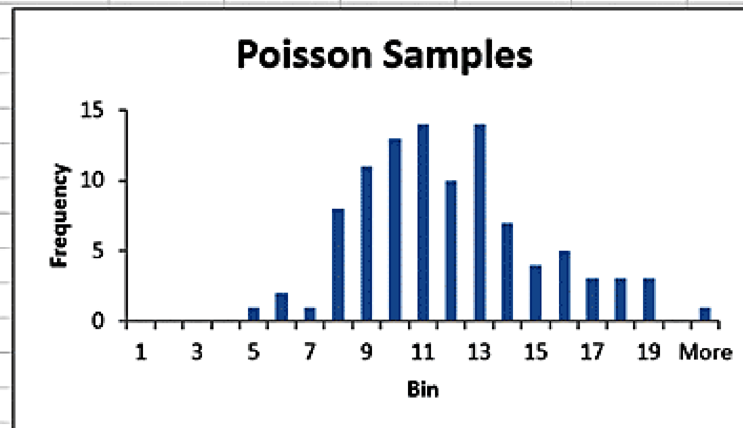
- Generate 100 outcomes from a Poisson distribution with a mean of 12
  - *Number of Variables* = 1.
  - *Number of Random Numbers* = 100
  - *Distribution* = Poisson
  - Dialog changes and prompts you to enter *Lambda* (mean of Poisson) = 12



Example 5.37 Results

Poisson Samples

Sample	Value	Bin	Frequency
1	10	1	0
2	11	2	0
3	11	3	0
4	16	4	0
5	18	5	1
6	11	6	2
7	6	7	1
8	19	8	8
9	12	9	11
10	13	10	13
11	14	11	14
12	16	12	10
13	14	13	14
14	15	14	7
15	15	15	4
16	12	16	5
17	15	17	3
18	8	18	3
19	10	19	3
20	9	20	0
21	13	More	1



(Histogram created manually)

Using Excel Functions to Generate Random Variates

- Normal: =NORM.INV*(RAND( ), mean, stdev)*
- Standard normal: =NORM.S.INV*(RAND( ))*

Example 5.38: A Sampling Experiment for Evaluating Capital Budgeting Projects

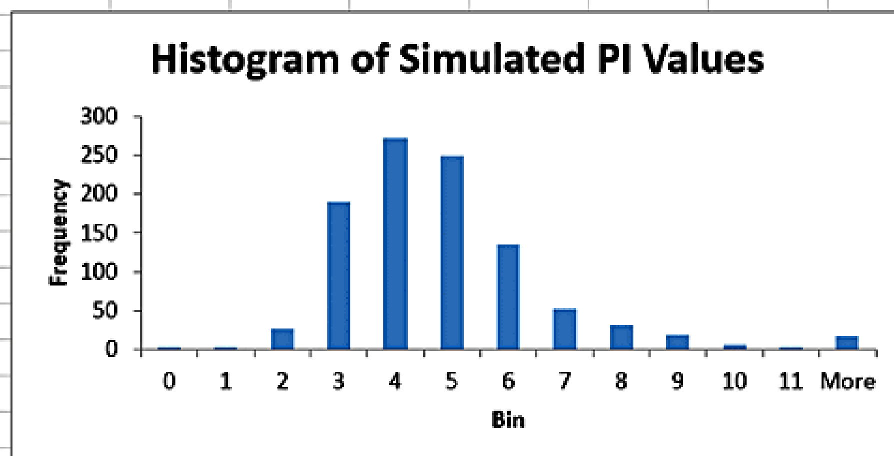
- In finance, one way of evaluating capital budgeting projects is to compute a profitability index: *PI = PV / I,*
  - *PV* is the present value of future cash flows
  - *I* is the initial investment
- What is the probability distribution of *PI* when PV is estimated to be normally distributed with a mean of $12 million and a standard deviation of $2.5 million, and the initial investment is also estimated to be normal with a mean of $3.0 million and standard deviation of $0.8 million.?

Example 5.38 Continued

- Column F:
  =NORM.INV(RAND( ), 12, 2.5)
- Column G:
  =NORM.INV(RAND( ), 3, 0.8)

	A	B	C	D	E	F	G	H
1	Profitability Index Analysis				Experiment	PV	I	PI
2					1	11.79045	2.116217	5.571475
3		Mean	Standard Deviation		2	10.62588	2.839064	3.742741
4	PV	12	2.5		3	12.22324	1.049416	11.64765
5	I	3	0.8		4	11.25269	3.947846	2.850337
6					5	11.3254	3.995613	2.83446
7	Mean PI for 1000 Experiments			4.365203	6	15.02659	3.324238	4.52031
8					7	12.79318	3.255405	3.929827
9					8	13.19409	3.000283	4.397616
10					9	12.7466	3.532532	3.608346
11					10	12.5399	3.675463	3.411789

	A	B
1	*Bin*	*Frequency*
2	0	1
3	1	1
4	2	26
5	3	189
6	4	272
7	5	249
8	6	135
9	7	52
10	8	31
11	9	18
12	10	6
13	11	3
14	More	17



Analytic Solver Platform Distribution Functions

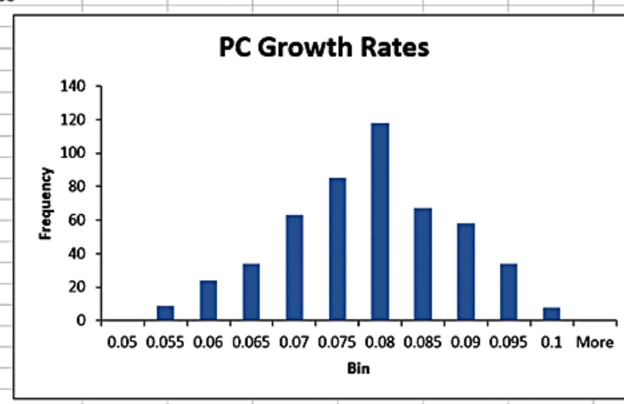
- *Analytic Solver Platform* provides Excel functions to generate random variates for many distributions

Distribution	*Analytic Solver Platform* Function
Bernoulli	PsiBernoulli(*probability*)
Binomial	PsiBinomial(*trials*, *probability*)
Poisson	PsiPoisson(*mean*)
Uniform	PsiUniform(*lower*, *upper*)
Normal	PsiNormal(*mean*, *standard deviation*)
Exponential	PsiExponential(*mean*)
Discrete Uniform	PsiDisUniform(*values*)
Geometric	PsiGeometric(*probability*)
Negative Binomial	PsiNegBinomial(*successes*, *probability*)
Hypergeometric	PsiHyperGeo(*trials*, *success*, *population size*)
Triangular	PsiTriangular(*minimum*, *most likely*, *maximum*)
Lognormal	PsiLognormal(*mean*, *standard deviation*)
Beta	PsiBeta(*alpha*, *beta*)

Example 5.39: Using *Analytic Solver Platform* Distribution Functions

- An energy company was considering offering a new product and needed to estimate the growth in PC ownership.
- Using the best data and information available, they determined that the minimum growth rate was 5.0%, the most likely value was 7.7%, and the maximum value was 10.0% (a triangular distribution).
  - A portion of 500 samples that were generated using the function PsiTriangular(5%, 7.7%, 10%):

	A	B
1	PC Ownership Growth Rates	
2		
3	Minimum	5%
4	Most Likely	7.70%
5	Maximum	10%
6		
7	Sample	Growth Rate
8	1	5.49%
9	2	6.16%
10	3	9.23%
11	4	8.49%
12	5	6.68%
13	6	7.63%
14	7	8.08%
15	8	8.54%
16	9	7.56%
17	10	9.26%
18	11	7.98%
19	12	7.46%
20	13	8.00%
21	14	6.54%
504	497	5.65%
505	498	7.39%
506	499	5.60%
507	500	6.50%



Data Modeling and Distribution Fitting

- Using sample data may limit our ability to predict uncertain events that may occur because potential values outside the range of the sample data are not included.
- A better approach is to identify the underlying probability distribution from which sample data come by “fitting” a theoretical distribution to the data and verifying the goodness of fit statistically.
  - Examine a histogram for clues about the distribution’s shape
  - Look at summary statistics such as the mean, median, standard deviation, coefficient of variation, and skewness

Example 5.40: Analyzing Airline Passenger Data

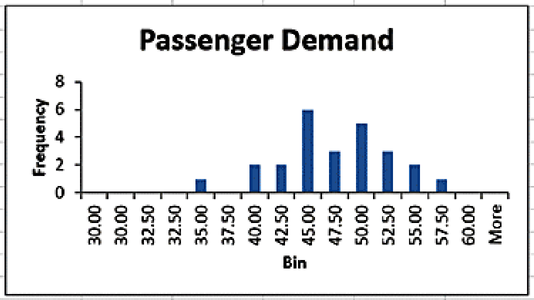
- Sample data on passenger demand for 25 flights

Airline Passengers

Sample Data
36
55
47
45
48
43
42
56
40
47
44
46
53
45
44
45
45
41
47
46
40
46
42
41
58

Bin	*Frequency*
30.00	0
30.00	0
32.50	0
32.50	0
35.00	1
37.50	0
40.00	2
42.50	2
45.00	6
47.50	3
50.00	5
52.50	3
55.00	2
57.50	1
60.00	0
More	0

Passengers	
Mean	45.68
Standard Error	1.043584
Median	45
Mode	45
Standard Deviation	5.217918
Sample Variance	27.22667
Kurtosis	0.707219
Skewness	0.823163
Range	22
Minimum	36
Maximum	58
Sum	1142
Count	25

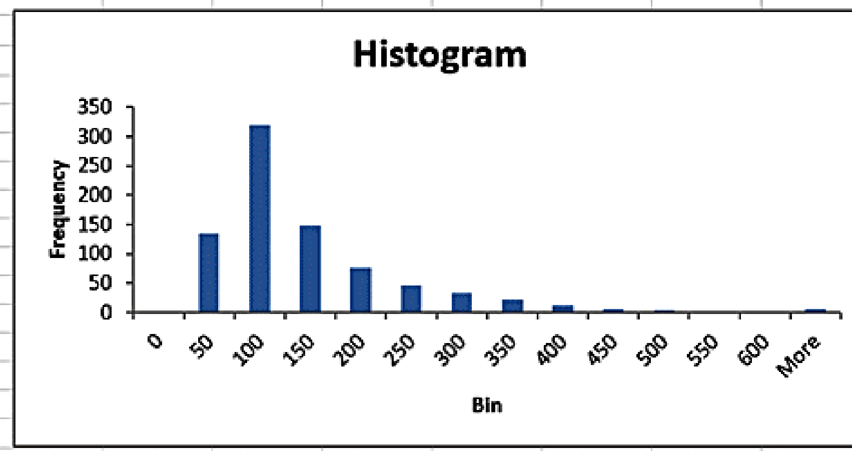


- The histogram shows a relatively symmetric distribution. The mean, median, and mode are all similar, although there is moderate skewness. A normal distribution is not unreasonable.

Example 5.41: Analyzing Airport Service Times

- Sample data on service times for 812 passengers at an airport's ticketing counter

	A	B	C	D
1	Service Times at an Airport Ticketing Counter			
2				
3	Times (sec.)		*Service Times*	
4	26			
5	358		Mean	126.2783
6	218		Standard Error	3.691221
7	71		Median	88
8	190		Mode	83
9	36		Standard Deviation	105.1836
10	45		Sample Variance	11063.59
11	68		Kurtosis	8.707526
12	45		Skewness	2.413577
13	66		Range	867
14	132		Minimum	9
15	372		Maximum	876
16	125		Sum	102538
17	16		Count	812
18	37			



  - It is not clear what the distribution might be. It does not appear to be exponential, but it might be lognormal or another distribution.

Goodness of Fit

- A better approach that simply visually examining a histogram and summary statistics is to analytically fit the data to the best type of probability distribution.
- Three statistics measure goodness of fit:
  - Chi-square (need at least 50 data points)
  - Kolmogorov-Smirnov (works well for small samples)
  - Anderson-Darling (puts more weight on the differences between the tails of the distributions)
- *Analytic Solver Platform* has the capability of fitting a probability distribution to data.

Example 5.42: Fitting a Distribution to Airport Service Times

1. Highlight the data

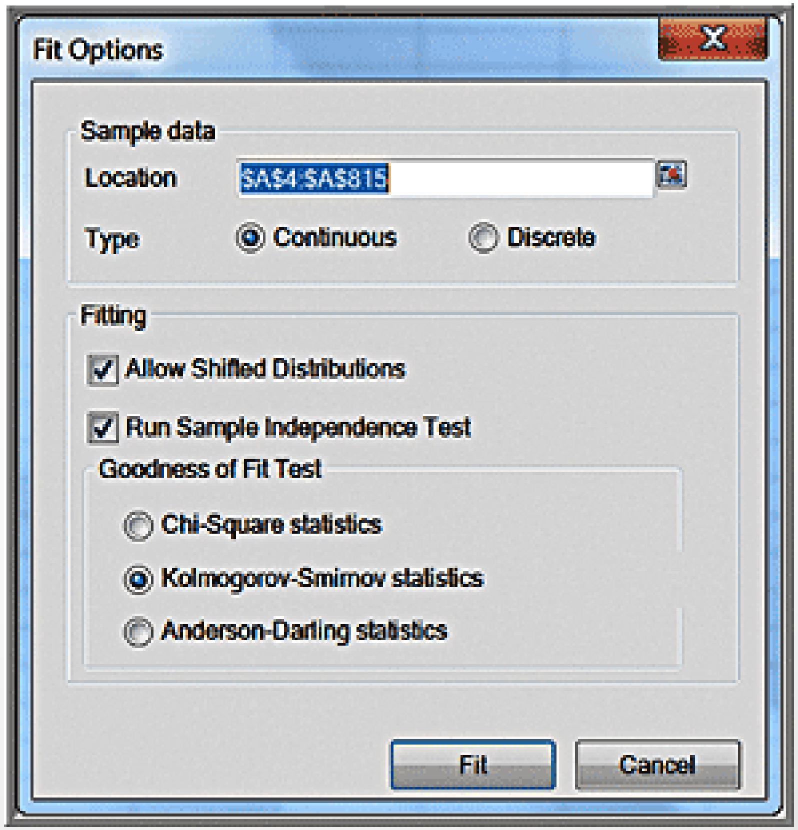
   Analytic Solver Platform > Tools > Fit

2. *Fit Options* dialog

   Type: Continuous

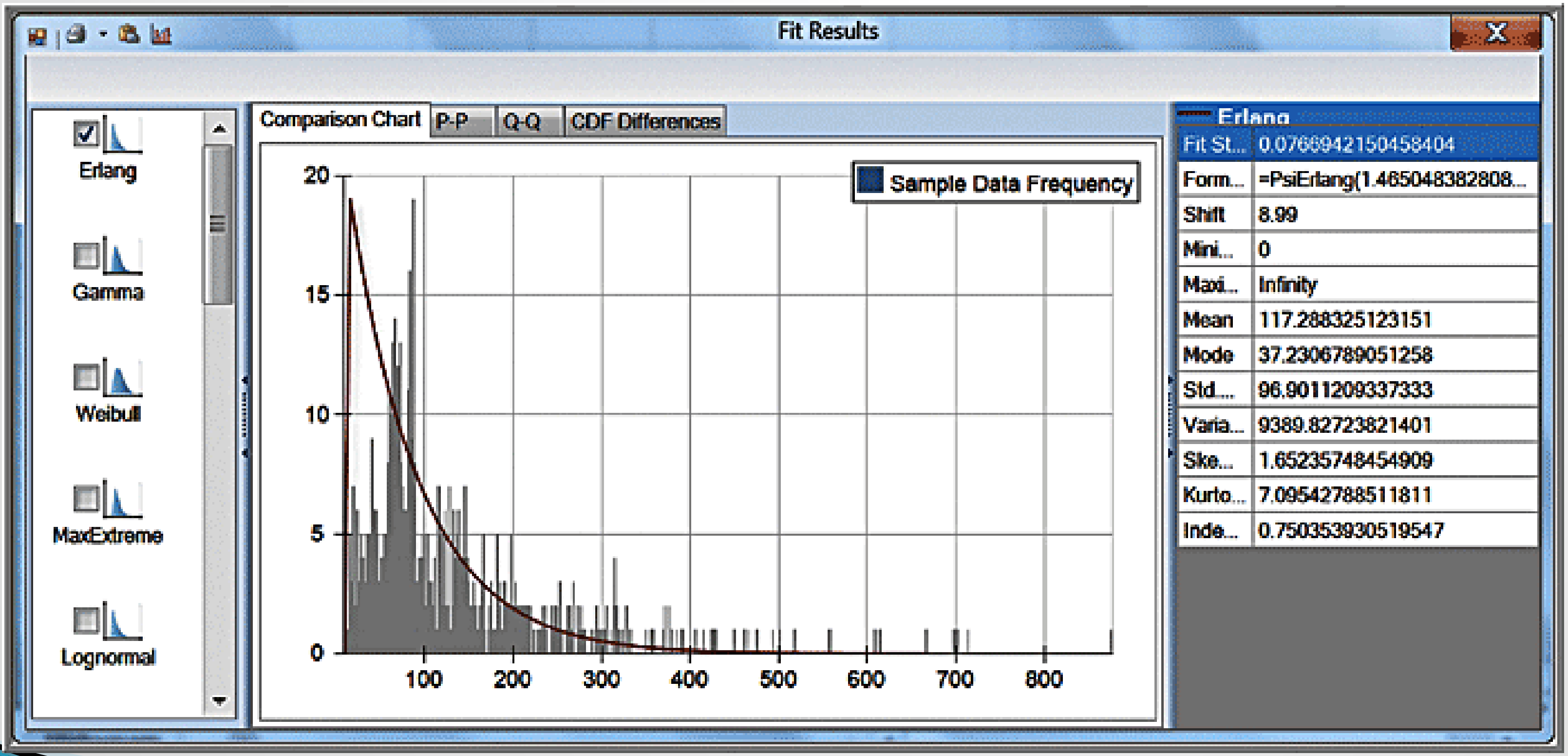
   Test: Kolmorgov-Smirnov

   Click *Fit* button



Example 5.42 Continued

- The best-fitting distribution is called an Erlang distribution.



Google Analytics Project 1 Email me (albert.kalim@asbury.edu) your answers by Sunday, 6/5,11:59 p.m. ET (10 points total)

Log in to your Google Analytics dashboard here and click Sign in to Analytics.

After you logged in, focus on the Audience tools on the left hand side. **Tell me as much as you can any information about the audience (visitors to this website) for the month of April 2022**. You will be graded based on your familiarity with the tools and how you read/interpret the data.