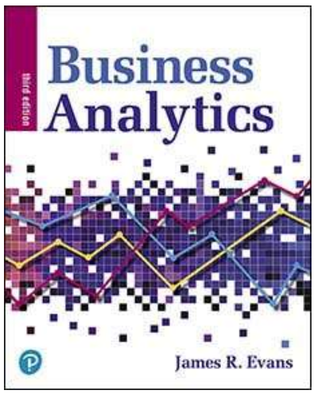
Business Analytics: Methods, Models, and Decisions
Third Edition



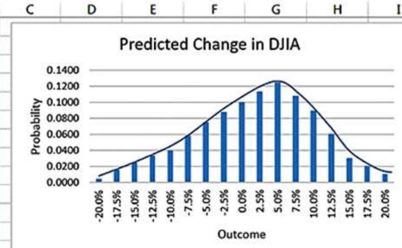
Chapter 5
Probability Distributions and Data Modeling

Part 2

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



Properties of 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*.

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \qquad \textbf{(5.17)}$$

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:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{(5.18)}$$

- Cumulative distribution function:

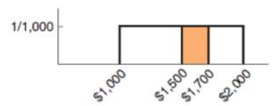
$$F(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1, & \text{if } b < x \end{cases} \qquad \textbf{(5.19)}$$

- Expected value $= \dfrac{(a+b)}{2}$; variance $= \dfrac{(b-a)^2}{12}$

Example 5.33: 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.
  - $F(1{,}300) = \dfrac{(1{,}300 - 1{,}000)}{(2{,}000 - 1{,}000)} = 0.30$

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

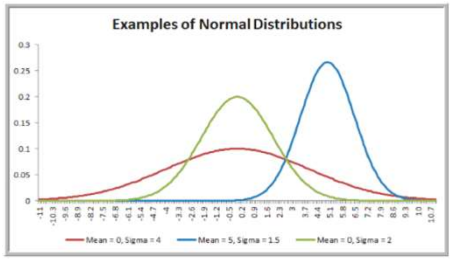


Discrete Uniform Distribution

- 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 $\frac{1}{6}$ 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



Pearson Slide - 7

Excel Normal Distribution Function

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

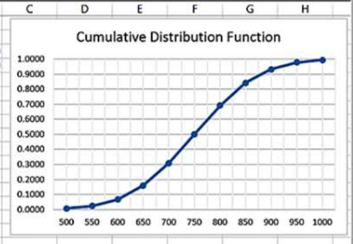
- NORM.DIST(*x*, *mean*, *standard_deviation*, *TRUE*) calculates the cumulative probability.
- If cumulative is set to FALSE, the function simply calculates the value of the density function $f(x)$, which has little practical application.

Pearson Slide - 8

Example 5.34 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

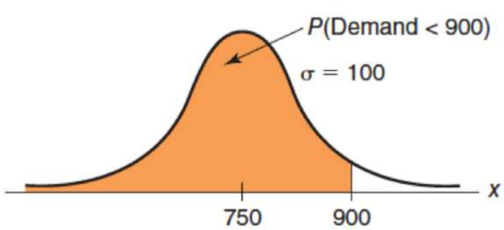
	A	B
1	**Normal Probabilities**	
2		
3	**Mean**	750
4	**Standard Deviation**	100
5		
6	x	F(x)
7	500	0.0062
8	550	0.0228
9	600	0.0668
10	650	0.1587
11	700	0.3085
12	750	0.5000
13	800	0.6915
14	850	0.8413
15	900	0.9332
16	950	0.9772
17	1000	0.9938



 Slide - 9

Example 5.34: Question 1

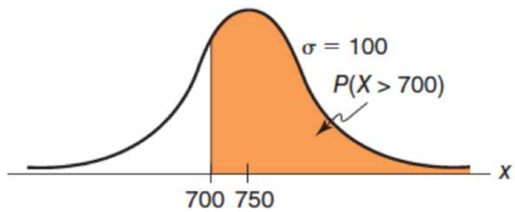
- Probability that demand will be at most 900 units, or $P(X \le 900)$:
  - $=$NORM.DIST(900,750,100,TRUE)$=$0.9332.



 Slide - 10

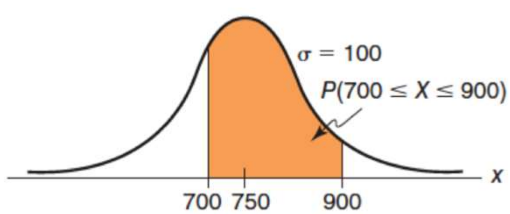
Example 5.34: Question 2

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



Example 5.34: 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

Suppose that we know the cumulative probability but don't know the value of x. The Excel function

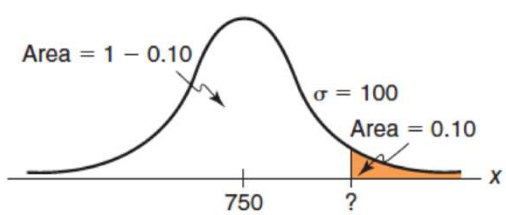
$$=\text{NORM.INV}(\textit{probability, mean, stdev})$$

provides the *x* value for a given cumulative probability.

 Slide - 13

Example 5.35: Using the NORM.INV Function

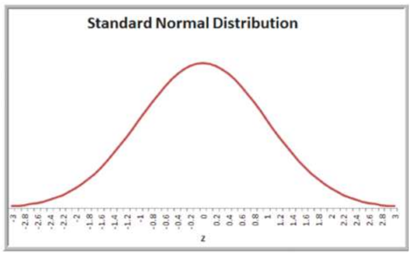
- What level of demand would be exceeded at most 10% of the time?
- Find *x* such that $F(x) = 0.90$:
  $=\text{NORM.INV}(0.90,750,100) = 878.155$



 Slide - 14

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.



 Slide - 15

Example 5.36: Computing Probabilities with the Standard Normal Distribution

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

=NORM.S.DIST(1) – NORM.S.DIST(-1)

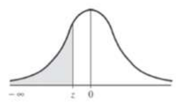
$$= 0.84134 - 0.15866$$
$$= 0.6827$$
$$\sim 68\%$$

	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						

 Slide - 16

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:

$$z = \frac{x - \mu}{\sigma} \qquad \textbf{(5.22)}$$

Example 5.37: Computing Probabilities with Standard Normal Tables

- From Example 5.34, what is the probability that demand will be at least 900 units/month?
- $$z = \frac{(900 - 750)}{100} = 1.50$$
- 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:

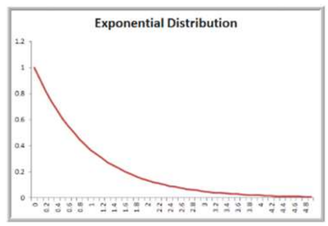
$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0 \qquad \textbf{(5.23)}$$

- Cumulative distribution function:

$$F(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0 \qquad \textbf{(5.24)}$$

- Mean $= \mu = \dfrac{1}{\lambda}$

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



Pearson Slide - 19

Excel Exponential Distribution Function

=EXPON.DIST(*x, lambda, cumulative*)

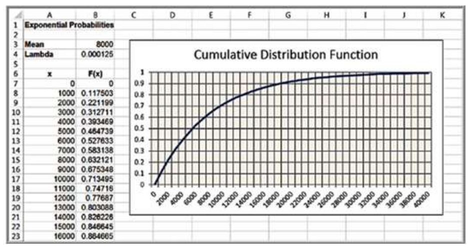
As with other Excel probability distribution functions, cumulative is either TRUE or FALSE, with TRUE providing the cumulative distribution function.

Pearson Slide - 20

Example 5.38: Using the Exponential Distribution

- The mean time to failure of a critical engine component is $\mu = 8{,}000$ hours. What is the probability of failing before 5000 hours?
  - $P(X < x) = \text{EXPON.DIST}(x,\ lambda,\ cumulative)$
  - $\lambda = \dfrac{1}{8000}$
  - $P(X < 5000) = \text{EXPON.DIST}\left(5000,\ \dfrac{1}{8000}, \text{TRUE}\right)$

$$= 0.4647$$



	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

Data Modeling and Distribution Fitting

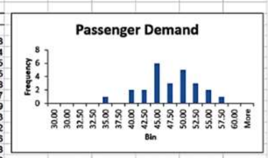
- Using sample data may limit our ability to predict uncertain events that may occur because potential values outside the range of the sample data are not included.
- A better approach is to identify the underlying probability distribution from which sample data come by "fitting" a theoretical distribution 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 5.39: Analyzing Airline Passenger Data

- Sample data on passenger demand for 25 flights

Airline Passengers

Sample Data	Bin	Frequency	Passengers	
36	30.00	0		
55	30.00	0		
47	32.50	0	Mean	45.68
45	32.50	0	Standard Error	1.043584
48	35.00	1	Median	45
43	37.50	0	Mode	45
42	40.00	2	Standard Deviation	5.217918
56	42.50	2	Sample Variance	27.22667
40	45.00	6	Kurtosis	0.707219
47	47.50	3	Skewness	0.823163
44	50.00	5	Range	22
46	52.50	3	Minimum	36
53	55.00	2	Maximum	58
45	57.50	1	Sum	1142
44	60.00	0	Count	25
45	More	0		
45				
41				
47				
46				
40				
46				
42				
41				
58				



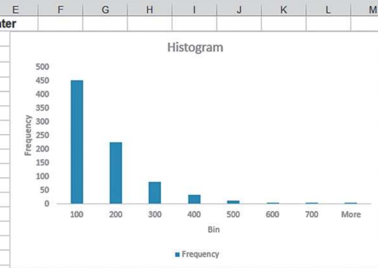
- The histogram shows a relatively symmetric distribution. The mean, median, and mode are all similar, although there is moderate skewness. A normal distribution is not unreasonable.

Example 5.40: Analyzing Airport Service Times

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

Service Times at an Airport Ticketing Counter

Times (sec.)	Service Times		Bin
26			100
358	Mean	126.28	200
218	Standard Error	3.6912	300
71	Median	88	400
190	Mode	83	500
36	Standard Deviatic	105.18	600
45	Sample Variance	11064	700
68	Kurtosis	8.7075	
45	Skewness	2.4136	
66	Range	867	
132	Minimum	9	
372	Maximum	876	
125	Sum	102538	
16	Count	812	
37			



- It appears to be exponential, but there is a difference between the mean and standard deviation, suggesting that it might be some other distribution.

