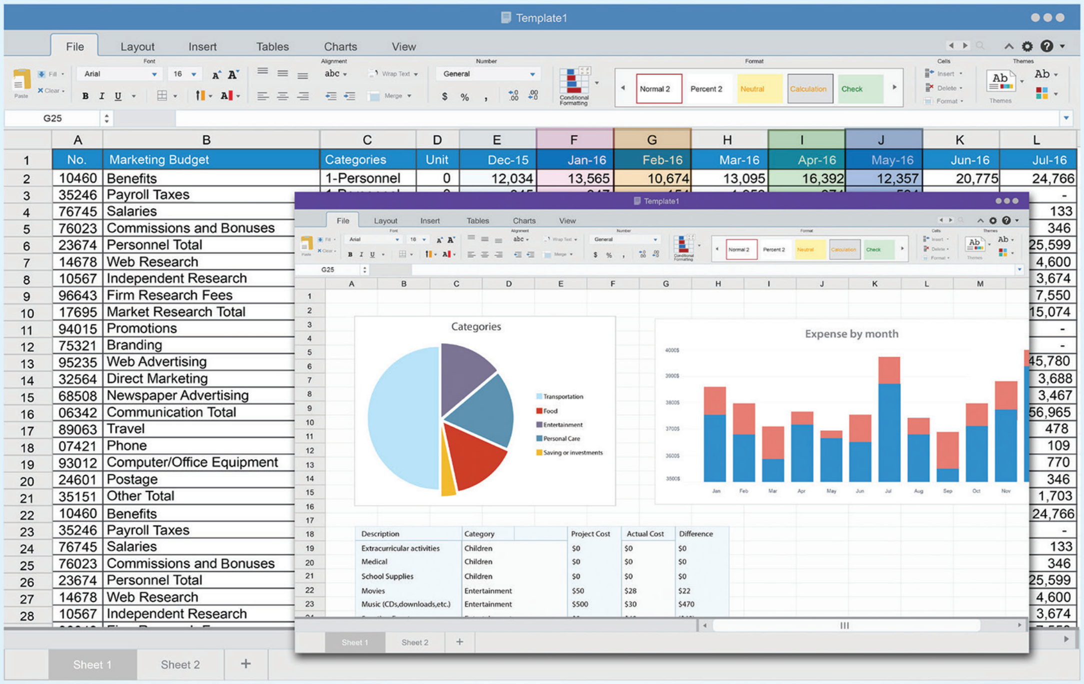
GLOBAL EDITION

Statistics for Managers

Using Microsoft® Excel®

NINTH EDITION

David M. Levine • David F. Stephan • Kathryn A. Szabat



	A	B	C	D	E	F	G	H	I	J	K	L
1	No.	Marketing Budget	Categories	Unit	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16
2	10460	Benefits	1-Personnel	0	12,034	13,565	10,674	13,095	16,392	12,357	20,775	24,766
3	35246	Payroll Taxes										-
4	76745	Salaries										133
5	76023	Commissions and Bonuses										346
6	23674	Personnel Total										25,599
7	14678	Web Research										4,600
8	10567	Independent Research										3,674
9	96643	Firm Research Fees										7,550
10	17695	Market Research Total										15,074
11	94015	Promotions										-
12	75321	Branding										-
13	95235	Web Advertising										45,780
14	32564	Direct Marketing										3,688
15	68508	Newspaper Advertising										3,467
16	06342	Communication Total										56,965
17	89063	Travel										478
18	07421	Phone										109
19	93012	Computer/Office Equipment										770
20	24601	Postage										346
21	35151	Other Total										1,703
22	10460	Benefits										24,766
23	35246	Payroll Taxes										-
24	76745	Salaries										133
25	76023	Commissions and Bonuses										346
26	23674	Personnel Total										25,599
27	14678	Web Research										4,600
28	10567	Independent Research										3,674

Description	Category	Project Cost	Actual Cost	Difference
Extracurricular activities	Children	$0	$0	$0
Medical	Children	$0	$0	$0
School Supplies	Children	$0	$0	$0
Movies	Entertainment	$50	$28	$22
Music (CDs,downloads,etc.)	Entertainment	$500	$30	$470

A Roadmap for Selecting a Statistical Method

Data Analysis Task	For Numerical Variables	For Categorical Variables
Describing a group or several groups	Ordered array, stem-and-leaf display, frequency distribution, relative frequency distribution, percentage distribution, cumulative percentage distribution, histogram, polygon, cumulative percentage polygon **(Sections 2.2, 2.4)** Mean, median, mode, geometric mean, quartiles, range, interquartile range, standard deviation, variance, coefficient of variation, skewness, kurtosis, boxplot, normal probability plot **(Sections 3.1, 3.2, 3.3, 6.3)** Index numbers **(online Section 16.7)** Dashboards **(Section 17.2)**	Summary table, bar chart, pie chart, doughnut chart, Pareto chart **(Sections 2.1 and 2.3)**
Inference about one group	Confidence interval estimate of the mean **(Sections 8.1 and 8.2)** *t* test for the mean **(Section 9.2)** Chi-square test for a variance or standard deviation **(online Section 12.7)**	Confidence interval estimate of the proportion **(Section 8.3)** *Z* test for the proportion **(Section 9.4)**
Comparing two groups	Tests for the difference in the means of two independent populations **(Section 10.1)** Wilcoxon rank sum test **(Section 12.4)** Paired *t* test **(Section 10.2)** *F* test for the difference between two variances **(Section 10.4)** Wilcoxon signed ranks test **(online Section 12.8)**	*Z* test for the difference between two proportions **(Section 10.3)** Chi-square test for the difference between two proportions **(Section 12.1)** McNemar test for two related samples **(online Section 12.6)**
Comparing more than two groups	One-way analysis of variance for comparing several means **(Section 11.1)** Kruskal-Wallis test **(Section 12.5)** Randomized block design **(online Section 11.3)** Two-way analysis of variance **(Section 11.2)**	Chi-square test for differences among more than two proportions **(Section 12.2)**
Analyzing the relationship between two variables	Scatter plot, time series plot **(Section 2.5)** Covariance, coefficient of correlation **(Section 3.5)** Simple linear regression **(Chapter 13)** *t* test of correlation **(Section 13.7)** Time-series forecasting **(Chapter 16)** Sparklines **(Section 2.7)**	Contingency table, side-by-side bar chart, PivotTables **(Sections 2.1, 2.3, 2.6)** Chi-square test of independence **(Section 12.3)**
Analyzing the relationship between two or more variables	Colored scatter plots, bubble chart, treemap **(Section 2.7)** Multiple regression **(Chapters 14 and 15)** Dynamic bubble charts **(Section 17.2)** Regression trees **(Section 17.3)** Cluster analysis **(Section 17.4)**	Multidimensional contingency tables **(Section 2.6)** Drilldown and slicers **(Section 2.7)** Logistic regression **(Section 14.7)** Classification trees **(Section 17.3)** Multiple correspondence analysis **(Section 17.5)**

CASES

Managing Ashland MultiComm Services

The Ashland MultiComm Services (AMS) marketing department wants to increase subscriptions for its *3-For-All* telephone, cable, and Internet combined service. AMS marketing has been conducting an aggressive direct-marketing campaign that includes postal and electronic mailings and telephone solicitations. Feedback from these efforts indicates that including premium channels in this combined service is a very important factor for both current and prospective subscribers. After several brainstorming sessions, the marketing department has decided to add premium cable channels as a no-cost benefit of subscribing to the *3-For-All* service.

The research director, Mona Fields, is planning to conduct a survey among prospective customers to determine how many premium channels need to be added to the *3-For-All* service in order to generate a subscription to the service. Based on past campaigns and on industry-wide data, she estimates the following:

Number of Free Premium Channels	Probability of Subscriptions
0	0.02
1	0.04
2	0.06
3	0.07
4	0.08
5	0.085

1. If a sample of 50 prospective customers is selected and no free premium channels are included in the *3-For-All* service offer, given past results, what is the probability that
   - **a.** fewer than 3 customers will subscribe to the *3-For-All* service offer?
   - **b.** 0 customers or 1 customer will subscribe to the *3-For-All* service offer?
   - **c.** more than 4 customers will subscribe to the *3-For-All* service offer?
   - **d.** Suppose that in the actual survey of 50 prospective customers, 4 customers subscribe to the *3-For-All* service offer. What does this tell you about the previous estimate of the proportion of customers who would subscribe to the *3-For-All* service offer?
2. Instead of offering no premium free channels as in Problem 1, suppose that two free premium channels are included in the *3-For-All* service offer. Given past results, what is the probability that
   - **a.** fewer than 3 customers will subscribe to the *3-For-All* service offer?
   - **b.** 0 customers or 1 customer will subscribe to the *3-For-All* service offer?
   - **c.** more than 4 customers will subscribe to the *3-For-All* service offer?
   - **d.** Compare the results of (a) through (c) to those of Problem 1.
   - **e.** Suppose that in the actual survey of 50 prospective customers, 6 customers subscribe to the *3-For-All* service offer. What does this tell you about the previous estimate of the proportion of customers who would subscribe to the *3-For-All* service offer?
   - **f.** What do the results in (e) tell you about the effect of offering free premium channels on the likelihood of obtaining subscriptions to the *3-For-All* service?
3. Suppose that additional surveys of 50 prospective customers were conducted in which the number of free premium channels was varied. The results were as follows:

Number of Free Premium Channels	Number of Subscriptions
1	5
3	6
4	6
5	7

How many free premium channels should the research director recommend for inclusion in the *3-For-All* service? Explain.

Digital Case

Apply your knowledge about expected value in this continuing Digital Case from Chapters 3 and 4.

Open **BullsAndBears.pdf**, a marketing brochure from EndRun Financial Services. Read the claims and examine the supporting data. Then answer the following:

1. Are there any "catches" about the claims the brochure makes for the rate of return of Happy Bull and Worried Bear funds?
2. What subjective data influence the rate-of-return analyses of these funds? Could EndRun be accused of making false and misleading statements? Why or why not?
3. The expected-return analysis seems to show that the Worried Bear fund has a greater expected return than the Happy Bull fund. Should a rational investor never invest in the Happy Bull fund? Why or why not?

CHAPTER 5

EXCEL GUIDE

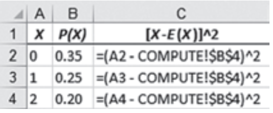
EG5.1 The PROBABILITY DISTRIBUTION for a DISCRETE VARIABLE

Key Technique Use **SUMPRODUCT(*X cell range, P*(*X*) *cell range*)** to compute the expected value. Use **SUMPRODUCT(*squared differences cell range, P*(*X*) *cell range*)** to compute the variance.

Example Compute the expected value, variance, and standard deviation for the number of interruptions per day data of Table 5.1 on page 207.

Workbook Use the **Discrete Variable workbook** as a model.

For the example, open to the **DATA worksheet** of the **Discrete Variable workbook**. The worksheet contains the column A and B entries needed to compute the expected value, variance, and standard deviation for the example. Unusual for a DATA worksheet in this book, column C contains formulas. These formulas use the expected value that cell B4 in the COMPUTE worksheet of the same workbook computes (first three rows shown below) and are equivalent to the fourth column calculations in Table 5.3.



	A	B	C
1	*X*	*P(X)*	[*X*-*E*(*X*)]^2
2	0	0.35	=(A2 - COMPUTE!$B$4)^2
3	1	0.25	=(A3 - COMPUTE!$B$4)^2
4	2	0.20	=(A4 - COMPUTE!$B$4)^2

For other problems, modify the DATA worksheet. Enter the probability distribution data into columns A and B and, if necessary, extend column C, by first selecting cell C7 and then copying that cell down as many rows as necessary. If the probability distribution has fewer than six outcomes, select the rows that contain the extra, unwanted outcomes, right-click, and then click Delete in the shortcut menu.

Appendix F further explains the SUMPRODUCT function that the COMPUTE worksheet uses to compute the expected value and variance.

EG5.2 BINOMIAL DISTRIBUTION

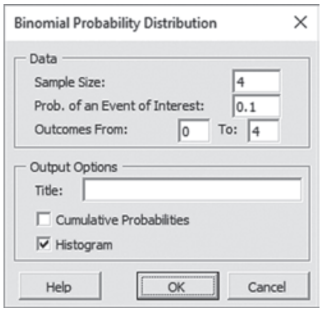
Key Technique Use the **BINOM.DIST(*number of events of interest, sample size, probability of an event of interest,* FALSE)** function.

Example Compute the binomial probabilities for $n = 4$ and $\pi = 0.1$, and construct a histogram of that probability distribution, similar to Figures 5.2 and 5.3 on page 214.

PHStat Use **Binomial**.

For the example, select **PHStat → Probability & Prob. Distributions → Binomial**. In the procedure's dialog box (shown below):

1. Enter **4** as the **Sample Size**.
2. Enter **0.1** as the **Prob. of an Event of Interest**.
3. Enter **0** as the **Outcomes From** value and enter **4** as the (Outcomes) **To** value.
4. Enter a **Title**, check **Histogram**, and click **OK**.



Binomial Probability Distribution

Data

Sample Size: 4

Prob. of an Event of Interest: 0.1

Outcomes From: 0 To: 4

Output Options

Title:

☐ Cumulative Probabilities

☑ Histogram

Check **Cumulative Probabilities** before clicking **OK** in step 4 to have the procedure include columns for $P(\leq X)$, $P(<X)$, $P(>X)$, and $P(\geq X)$ in the binomial probabilities table.

Workbook Use the **Binomial workbook** as a template and model.

For the example, open to the **COMPUTE worksheet** of the **Binomial workbook**, shown in Figure 5.2 on page 214. The worksheet already contains the entries needed for the example. For other problems, change the sample size in cell B4 and the probability of an event of interest in cell B5. If necessary, extend the binomial probabilities table by first selecting cell range A18:B18 and then copying that cell range down as many rows as necessary. To construct a histogram of the probability distribution, use the Appendix Section B.6 instructions.

For problems that require cumulative probabilities, use the CUMULATIVE worksheet in the Binomial workbook. The SHORT TAKES for Chapter 5 explains and documents this worksheet.

EG5.3 POISSON DISTRIBUTION

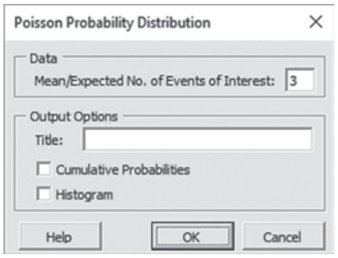
Key Technique Use the **POISSON.DIST(*number of events of interest, the average or expected number of events of interest,* FALSE)** function.

Example Compute the Poisson probabilities for the Figure 5.7 customer arrival problem on page 219.

PHStat Use **Poisson**.

For the example, select **PHStat ➔ Probability & Prob. Distributions ➔ Poisson**. In this procedure's dialog box (shown below):

1. Enter **3** as the **Mean/Expected No. of Events of Interest**.
2. Enter a **Title** and click **OK**.



Poisson Probability Distribution

Data

Mean/Expected No. of Events of Interest: 3

Output Options

Title:

□ Cumulative Probabilities

□ Histogram

Check **Cumulative Probabilities** before clicking **OK** in step 2 to have the procedure include columns for $P(\leq X)$, $P(<X)$, $P(>X)$, and $P(\geq X)$ in the Poisson probabilities table. Check **Histogram** to construct a histogram of the Poisson probability distribution.

Workbook Use the **Poisson workbook** as a template.

For the example, open to the **COMPUTE worksheet** of the **Poisson workbook**, shown in Figure 5.7 on page 219. The worksheet already contains the entries for the example. For other problems, change the mean or expected number of events of interest in cell E4. To construct a histogram of the probability distribution, use the Appendix Section B.6 instructions.

For problems that require cumulative probabilities, use the CUMULATIVE worksheet in the Binomial workbook. The Short Takes for Chapter 5 explains and documents this worksheet.

6 The Normal Distribution and Other Continuous Distributions

CONTENTS

OBJECTIVES

- Compute probabilities from the normal distribution
- Use the normal distribution to solve business problems
- Use the normal probability plot to determine whether a set of data is approximately normally distributed
- Compute probabilities from the uniform distribution.

MyTVLab watch learn share subscribe

Watch for Free This Month's Exclusive Content

The Adventures of Dirk "Sunny" Lande

A detuned guitar, runaway acoustic feedback, and a blown amp causes a tear in the fabric of the metaverse and propels musician Dirk ("but everyone calls me Sonny") Lande READ MORE

Whitney Meets Taylor

Having completed her adventure meeting Waldo, Whitney encounters a teddy bear named Taylor in this delightful children's story READ MORE

MyTVLab Mobile Streaming Video Series Winner

Two Bananas: Kitchen Counter

The sequel to *Two Bananas: Kitchen Table* finds the crime-solving bananas on a counter, looking for missing kitchen contractor READ MORE

MyTVLab Documentary of the Month

Uncovering The Orange Roofs

One man's quest to revisit the golden age of automobile travel leads him to uncover a famous roadside symbol of the past READ MORE

Presented by VLABGo

MyTVLab Highlight on Travel

Contact your Internet Provider for special MyTVLab Packages

Privacy Policy Terms of Service

© MyTVLab, LLC. All rights reserved.

▼ USING STATISTICS
Normal Load Times at MyTVLab

You are the vice president in charge of sales and marketing for MyTVLab, a web-based business that has evolved into a full-fledged, subscription-based streaming video service. To differentiate MyTVLab from the other companies that sell similar services, you decide to create a "Why Choose Us" web page to help educate new and prospective subscribers about all that MyTVLab offers.

As part of that page, you have produced a new video that samples the content MyTVLab streams as well as demonstrates the relative ease of setting up MyTVLab on many types of devices. You want this video to download with the page so that a visitor can jump to different segments immediately or view the video later, when offline.

You know from research (Kishnan and Sitaraman) and past observations, Internet visitors will not tolerate waiting too long for a web page to load. One wait time measure is load time, the time in seconds that passes from first pointing a browser to a web page until the web page is fully loaded and content such as video is ready to be viewed. You have set a goal that the load time for the new sales page should rarely exceed 10 seconds (too long for visitors to wait) and, ideally, should rarely be less than 1 second (a waste of company Internet resources).

To measure this time, you point a web browser at the MyTVLab corporate test center to the new sales web page and record the load time. In your first test, you record a time of 6.67 seconds. You repeat the test and record a time of 7.52 seconds. Though consistent to your goal, you realize that two load times do not constitute strong proof of anything, especially as your assistant has performed his own test and recorded a load time of 8.83 seconds.

Could you use a method based on probability theory to ensure that most load times will be within the range you seek? MyTVLab has recorded past load times of a similar page with a similar video and determined the mean load time of that page is 7 seconds, the standard deviation of those times is 2 seconds, that approximately two-thirds of the load times are between 5 and 9 seconds, and about 95% of the load times are between 3 and 11 seconds.

Could you use these facts to assure yourself that the load time goal you have set for the new sales page is likely to be met?

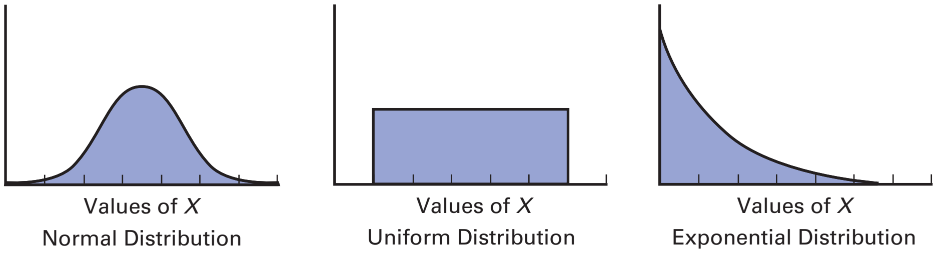
Chapter 5 discusses how to use probability distributions for a *discrete* numerical variable. In the MyTVLab scenario, you are examining the load time, a *continuous* numerical variable. You are no longer considering a table of discrete (specific) values, but a continuous range of values. For example, the phrase "load times are between 5 and 9 seconds" includes *any* value between 5 and 9 and not just the values 5, 6, 7, 8, and 9. If you plotted the phrase on a graph, you would draw a *continuous* line from 5 to 9 and not just plot five discrete points.

When you add information about the shape of the range of values, such as two-thirds of the load times are between 5 and 9 seconds or about 95% of the load times are between 3 and 11 seconds, you can visualize the plot of all values as an area under a curve. If that area under the curve follows the well-known pattern of certain continuous distributions, you can use the continuous probability distribution for that pattern to estimate the likelihood that a load time is within a range of values. In the MyTVLab scenario, the past load times of a similar page describes a pattern that conforms to the pattern associated with the normal distribution, the subject of Section 6.2. That would allow you, as the vice president for sales and marketing, to use the normal distribution with the statistics given to determine if your load time goal is likely to be met.

6.1 Continuous Probability Distributions

Continuous probability distributions vary by the shape of the area under the curve. Figure 6.1 visualizes the normal, uniform, and exponential probability distributions.

FIGURE 6.1
Three continuous probability distributions



Some distributions, including the normal and uniform distributions in Figure 6.1, show a symmetrical shape. Distributions such as the right-skewed exponential distribution do not. In symmetrical distributions the mean equals the median, whereas in a right-skewed distribution the mean is greater than the median. Each of the three distributions also has unique properties.

The **normal distribution** is not only symmetrical, but bell-shaped, a shape that (loosely) suggests the profile of a bell. Being bell-shaped means that most values of the continuous variable will cluster around the mean. Although the values in a normal distribution can range from negative infinity to positive infinity, the shape of the normal distribution makes it very unlikely that extremely large or extremely small values will occur.

Section 6.4 further discusses the uniform distribution.

The online Section 6.5 further discusses the exponential distribution.

The **uniform distribution**, also known as the *rectangular distribution*, contains values that are equally distributed in the range between the smallest value and the largest value. In a uniform distribution, every value is equally likely.

The **exponential distribution** contains values from zero to positive infinity and is right-skewed, making the mean greater than the median. Its shape makes it unlikely that extremely large values will occur.

Besides visualizations such as those in Figure 6.1, a continuous probability distribution can be expressed mathematically as a *probability density function*. A **probability density function** for a specific continuous probability distribution, represented by the symbol $f(X)$, defines the distribution of the values for a continuous variable and can be used as the basis for calculations that determine the likelihood or probability that a value will be within a certain range.

6.2 The Normal Distribution

The most commonly used continuous probability distribution, the normal distribution, plays an important role in statistics and business. Because of its relationship to the Central Limit Theorem (see Section 7.2), the distribution provides the basis for classical statistical inference and can be used to approximate various discrete probability distributions. For business, many continuous variables used in decision making have distributions that closely resemble the normal distribution. The normal distribution can be used to estimate the probability that values occur within a specific range or interval. This probability corresponds to an area under a curve that the normal distribution defines. Because a single point on a curve, representing a specific value, cannot define an area, the area under any single point/specific value will be 0. Therefore, when using the normal distribution to estimate values of a continuous variable, the probability that the variable will be exactly a specified value is always zero.

By the rule the previous paragraph states, the probability that the load time is exactly 7, or any other specific value, is zero.

For the MyTVLab scenario, the load time for the new sales page would be an example of a continuous variable whose distribution approximates the normal distribution. This approximation enables one to estimate probabilities such as the probability that the load time would be between 7 and 10 seconds, the probability that the load time would be between 8 and 9 seconds, or the probability that the load time would be between 7.99 and 8.01 seconds.

Exhibit 6.1 presents four important theoretical properties of the normal distribution. The distributions of many business decision-making continuous variables share the first three properties, sufficient to allow the use of the normal distribution to *estimate* the probability for specific ranges or intervals of values.

EXHIBIT 6.1

Normal Distribution Important Theoretical Properties

Symmetrical distribution. Its mean and median are equal.

Bell-shaped. Values cluster around the mean.

Interquartile range is roughly 1.33 standard deviations. Therefore, the middle 50% of the values are contained within an interval that is approximately two-thirds of a standard deviation below and two-thirds of a standard deviation above the mean.

The distribution has an infinite range ($-\infty < X < \infty$). Six standard deviations approximate this range (see page 235).

Table 6.1 presents the fill amounts, the volume of liquid placed inside a bottle, for a production run of 10,000 one-liter water bottles. Due to minor irregularities in the machinery and the water pressure, the fill amounts will vary slightly from the desired target amount, which is a bit more than 1.0 liters to prevent underfilling of bottles and the subsequent consumer unhappiness that such underfilling would cause.

TABLE 6.1
Fill amounts for 10,000 one-liter water bottles

Fill Amount (liters)	Relative Frequency
< 1.025	48/10,000 = 0.0048
1.025 < 1.030	122/10,000 = 0.0122
1.030 < 1.035	325/10,000 = 0.0325
1.035 < 1.040	695/10,000 = 0.0695
1.040 < 1.045	1,198/10,000 = 0.1198
1.045 < 1.050	1,664/10,000 = 0.1664
1.050 < 1.055	1,896/10,000 = 0.1896
1.055 < 1.060	1,664/10,000 = 0.1664
1.060 < 1.065	1,198/10,000 = 0.1198
1.065 < 1.070	695/10,000 = 0.0695
1.070 < 1.075	325/10,000 = 0.0325
1.075 < 1.080	122/10,000 = 0.0122
1.080 or above	48/10,000 = 0.0048
Total	1.0000