Online Tutorial 1

Statistical Tools for Managers

Tutorial Outline

Statistical applications permeate the subject of operations management because so much of decision making depends on probabilities that are based on limited or uncertain information. This tutorial provides a review of several important statistical tools that are useful in many chapters of the text. An understanding of the concepts of probability distributions, expected values, and variances is needed in the study of decision trees, quality control, forecasting, queuing models, work measurement, learning curves, inventory, simulation, project management, and maintenance.

DISCRETE PROBABILITY DISTRIBUTIONS

In this section, we explore the properties of **discrete probability distributions**, that is, distributions in which outcomes are not continuous. When we deal with discrete variables, there is a probability value assigned to each event. These values must be between 0 and 1, and they must sum to 1. Example T1 relates to a sampling of student grades.

Example T1

The dean at East Florida University, Nancy Beals, is concerned about the undergraduate statistics training of new MBA students. In a sampling of 100 applicants for next year's MBA class, she asked each student to supply his or her final grade in the course in statistics taken as a sophomore or junior. To translate from letter grades to a numeric score, the dean used the following system:

5. A **4.** B **3.** C **2.** D **1.** F

The responses to this query of the 100 potential students are summarized in the table below. Also shown is the probability for each possible grade outcome. This discrete probability distribution is computed using the relative frequency approach. Probability values are also often shown in graph form as in Figure T1.1.

PROBABILITY DISTRIBUTION FOR GRADES

GRADE LETTER OUTCOME	SCORE VARIABLE (x)	NUMBER OF STUDENTS RESPONDING	PROBABILITY, $P(x)$
A	5	10	0.1 = 10/100
B	4	20	0.2 = 20/100
C	3	30	0.3 = 30/100
D	2	30	0.3 = 30/100
F	1	10	0.1 = 10/100
		Total = 100	1.0 = 100/100

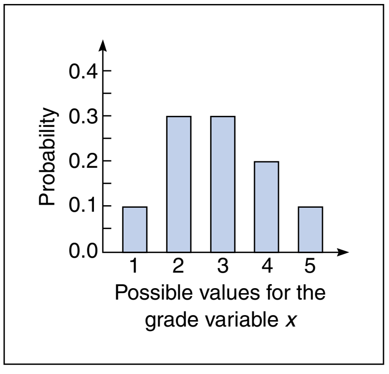
This distribution follows the three rules required of all probability distributions:

1. The events are mutually exclusive and collectively exhaustive.
2. The individual probability values are between 0 and 1 inclusive.
3. The total of the probability values sum to 1.

The graph of the probability distribution in Example T1 gives us a picture of its shape. It helps us identify the central tendency of the distribution (called the expected value) and the amount of variability or spread of the distribution (called the variance). Expected value and variance are discussed next.

FIGURE T1.1 ■

Probability Function for Grades



Expected Value of a Discrete Probability Distribution

Once we have established a distribution, the first characteristic we are usually interested in is the "central tendency" or average of the distribution.[1] We computed the **expected value**, a measure of central tendency, as a weighted average of the values of the variable:

$$E(x) = \sum_{i=1}^{n} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + \cdots + x_n P(x_n) \quad \text{(T1-1)}$$

where x_i = variable's possible values
$P(x_i)$ = probability of each of the variable's possible values

The expected value of any discrete probability distribution can be computed by: (1) multiplying each possible value of the variable x_i by the probability $P(x_i)$ that outcome will occur, and (2) summing the results, indicated by the summation sign, Σ. Example T2 shows such a calculation.

Example T2

Here is how the expected grade value can be computed for the question in Example T1.

$$\begin{aligned} E(x) &= \sum_{i=1}^{5} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3) + x_4 P(x_4) + x_5 P(x_5) \\ &= (5)(.1) + (4)(.2) + (3)(.3) + (2)(.3) + (1)(.1) \\ &= 2.9 \end{aligned}$$

The expected grade of 2.9 implies that the mean statistics grade is between D (2) and C (3), and that the average response is closer to a C, which is 3. Looking at Figure T1.1, we see that this is consistent with the shape of the probability function.

Variance of a Discrete Probability Distribution

In addition to the central tendency of a probability distribution, most decision makers are interested in the variability or the spread of the distribution. The **variance** of a probability distribution is a number that reveals the overall spread or dispersion of the distribution.[2] For a discrete probability distribution, it can be computed using the following equation:

$$\text{Variance} = \sum_{i=1}^{n} (x_i - E(x))^2 P(x_i) \quad \text{(T1-2)}$$

where x_i = variable's possible values
$E(x)$ = expected value of the variable
$P(x_i)$ = probability of each possible value of the variable

[1]If the data we are dealing with have not been grouped into a probability distribution, the measure of central tendency is called the arithmetic mean, or simply, the average. Here is the mean of the following seven numbers: 10, 12, 18, 6, 4, 5, 15.

$$\text{Arithmetic mean, } \overline{X} = \frac{\Sigma X}{n} = \frac{10 + 12 + 18 + 6 + 4 + 5 + 15}{7} = 10$$

[2]Just as the variance of a probability distribution shows the dispersion of the data, so does the variance of ungrouped data, that is, data not formed into a probability distribution. The formula is: Variance = $\Sigma(X - \overline{X})^2/n$. Using the numbers 10, 12, 18, 6, 4, 5, and 15, we find that $\overline{X}$ = 10. Here are the variance computations:

$$\begin{aligned} \text{Variance} &= \frac{(10-10)^2 + (12-10)^2 + (18-10)^2 + (6-10)^2 + (4-10)^2 + (5-10)^2 + (15-10)^2}{7} \\ &= \frac{0 + 4 + 64 + 16 + 36 + 25 + 25}{7} \\ &= \frac{170}{7} = 24.28 \end{aligned}$$

We should also note that when the data we are looking at represent a *sample* of a whole set of data, we use the term $n - 1$ in the denominator, instead of n, in the variance formula.

To compute the preceding variance, the expected value is subtracted from each value of the variable squared, and multiplied by the probability of occurrence of that value. The results are then summed to obtain the variance.

A related measure of dispersion or spread is the **standard deviation**. This quantity is also used in many computations involved with probability distributions. The standard deviation, σ, is just the square root of the variance:

$$\sigma = \sqrt{\text{variance}} \tag{T1-3}$$

Example T3 shows a variance and standard deviation calculation.

Example T3

Here is how this procedure is done for the statistics grade survey question:

$$\begin{aligned}\text{Variance} &= \sum_{i=1}^{5} (x_i - E(x))^2 P(x_i) \\ &= (5 - 2.9)^2(.1) + (4 - 2.9)^2(.2) + (3 - 2.9)^2(.3) + (2 - 2.9)^2(.3) + (1 - 2.9)^2(.1) \\ &= (2.1)^2(.1) + (1.1)^2(.2) + (.1)^2(.3) + (-.9)^2(.3) + (-1.9)^2(.1) \\ &= .441 + .242 + .003 + .243 + .361 \\ &= 1.29\end{aligned}$$

The standard deviation for the grade question is

$$\begin{aligned}\sigma &= \sqrt{\text{variance}} \\ &= \sqrt{1.29} = 1.14\end{aligned}$$

CONTINUOUS PROBABILITY DISTRIBUTIONS

There are many examples of continuous variables. The time it takes to finish a project, the number of ounces in a barrel of butter, the high temperature during a given day, the exact length of a given type of lumber, and the weight of a railroad car of coal are all examples of continuous variables. Variables can take on an infinite number of values, so the fundamental probability rules must be modified for continuous variables.

As with discrete probability distributions, the sum of the probability values must equal 1. Because there are an infinite number of values of the variables, however, the probability of *each value* of the variable *must be 0*. If the probability values for the variable values were greater than 0, then the sum would be infinitely large.

The Normal Distribution

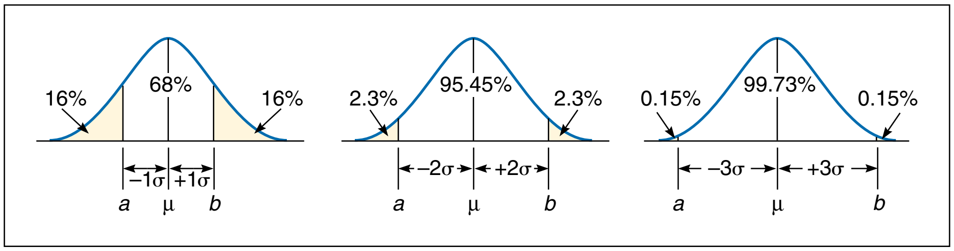
One of the most popular and useful continuous probability distributions is the **normal distribution**, which is characterized by a bell-shaped curve. The normal distribution is completely specified when values for the mean, μ, and the standard deviation, σ, are known.

The Area Under the Normal Curve Because the normal distribution is symmetrical, its midpoint (and highest point) is at the mean. Values of the *x*-axis are then measured in terms of how many standard deviations they are from the mean.

The area under the curve (in a continuous distribution) describes the probability that a variable has a value in the specified interval. The normal distribution requires complex mathematical calculations, but tables that provide areas or probabilities are readily available. For example, Figure T1.2 illustrates three commonly used relationships that have been derived from standard normal tables (a procedure we discuss in a moment). The area from point *a* to point *b* in the first drawing represents the probability, 68%, that the variable will be within 1 standard deviation of the mean. In the middle graph, we see that about 95.45% of the area lies within plus or minus 2 standard deviations of the mean. The third figure shows that 99.73% lies between $\pm 3\sigma$.

FIGURE T1.2 ■

Three Common Areas under Normal Curves



Translated into an application, Figure T1.2 implies that if the expected lifetime of an experimental computer chip is $\mu = 100$ days, and if the standard deviation is $\sigma = 15$ days, then we can make the following statements:

1. 68% of the population of computer chips (technically, 68.26%) studied have lives between 85 and 115 days (namely, $\pm 1\sigma$).
2. 95.45% of the chips have lives between 70 and 130 days ($\pm 2\sigma$).
3. 99.73% of the computer chips have lives in the range from 55 to 145 days ($\pm 3\sigma$).
4. Only 16% of the chips have lives greater than 115 days (from first graph, the area to the right of $+1\sigma$).

Using the Standard Normal Table To use a table to find normal probability values, we follow two steps.

Step 1: Convert the normal distribution to what we call a *standard normal distribution*. A standard normal distribution is one that has a mean of 0 and a standard deviation of 1. All normal tables are designed to handle variables with $\mu = 0$ and $\sigma = 1$. Without a standard normal distribution, a different table would be needed for each pair of μ and σ values. We call the new standard variable z. The value of z for any normal distribution is computed from the equation:

$$z = \frac{x - \mu}{\sigma} \quad \text{(T1-4)}$$

where x = value of the variable we want to measure
μ = mean of the distribution
σ = standard deviation of the distribution
z = number of standard deviations from the mean, μ to x

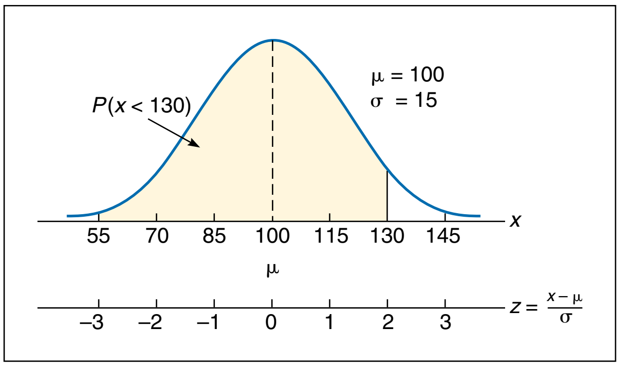
For example, if $\mu = 100$, $\sigma = 15$, and we are interested in finding the probability that the variable x is less than 130, then we want $P(x < 130)$.

$$z = \frac{x - \mu}{\sigma} = \frac{130 - 100}{15} = \frac{30}{15} = 2 \text{ standard deviations}$$

This means that the point x is 2.0 standard deviations to the right of the mean. This is shown in Figure T1.3.

FIGURE T1.3 ■

Normal Distribution Showing the Relationship between *z* Values and *x* Values



Step 2: Look up the probability from a table of normal curve areas. Appendix I in the textbook is such a table of areas for the standard normal distribution. One of the ways it is set up is to provide the area under the curve to the left of any specified value of z.

Let us see how Appendix I can be used. The column on the left lists values of z, with the second decimal place of z appearing in the top row. For example, for a value of $z = 2.00$ as just computed, find 2.0 in the left-hand column and .00 in the top row. In the body of the table, we find that the area sought is .97725, or 97.7%. Thus:

$$P(x < 130) = P(z < 2.00) = 97.7\%$$

This suggests that if the mean lifetime of a computer chip is 100 days with a standard deviation of 15 days, the probability that the life of a randomly selected chip is less than 130 days is 97.7%.

By referring to Figure T1.2, we see that this probability could also have been derived from the middle graph. Note that $1.0 - .977 = .023 = 2.3\%$, which is the area in the right-hand tail of the curve.

Example T4 illustrates the use of the normal distribution further.

Example T4

Holden Construction Co. builds primarily three- and four-unit apartment buildings (called triplexes and quadraplexes) for investors, and it is believed that the total construction time in days follows a normal distribution. The mean time to construct a triplex is 100 days, and the standard deviation is 20 days. If the firm finishes this triplex in 75 days or less, it will be awarded a bonus payment of $5,000. What is the probability that Holden will receive the bonus?

FIGURE T1.4 ■

Probability Holden Will Receive the Bonus by Finishing in 75 Days

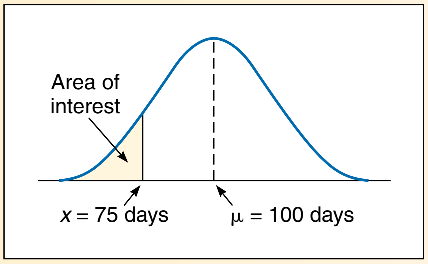


Figure T1.4 illustrates the probability we are looking for in the shaded area. The first step is to compute the z value:

$$z = \frac{x - \mu}{\sigma} = \frac{75 - 100}{20} = \frac{-25}{20} = -1.25$$

This z value indicates that 75 days is -1.25 standard deviations to the left of the mean. But the standard normal table is structured to handle only positive z values. To solve this problem, we observe that the curve is symmetric. The probability Holden will finish in *less than 75 days is equivalent* to the probability it will finish in *more than 125 days*. We first find the probability Holden will finish in less than 125 days. That value was .89435. So the probability it will take more than 125 days is

$$P(x < 125) = .89435$$

$$\text{Thus, } P(x > 125) = 1.0 - P(x < 125) = 1.0 - .89435 = .10565$$

Thus, the probability of completing the triplex in 75 days is .10565, or about 10%.

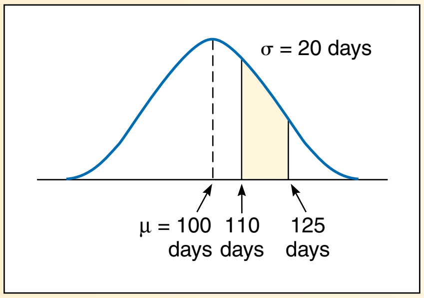
A second example: What is the probability the triplex will take between 110 and 125 days? We see in Figure T1.5 that

$$P(110 < x < 125) = P(x < 125) - P(x < 110)$$

That is, the shaded area in the graph can be computed by finding the probability of completing the building in 125 days or less *minus* the probability of completing it in 110 days or less.

FIGURE T1.5 ■

Probability of Holden Completion between 110 and 125 Days



Recall that $P(x < 125 \text{ days})$ is equal to .89435. To find $P(x < 110 \text{ days})$, we follow the two steps developed earlier.

1. $z = \frac{x - \mu}{\sigma} = \frac{110 - 100}{20} = \frac{10}{20} = .50$ standard deviation
2. From Appendix I, we see that the area for $z = .50$ is .69146. So the probability the triplex can be completed in less than 110 days is .69146. Finally,

$$P(110 < x < 125) = .89435 - .69146 = .20289$$

The probability that it will take between 110 and 125 days is about 20%.

SUMMARY

The purpose of this tutorial is to assist readers in tackling decision-making issues that involve probabilistic (uncertain) information. A background in statistical tools is quite useful in studying operations management. We examined two types of probability distributions, discrete and continuous. Discrete distributions assign a probability to each specific event. Continuous distributions, such as the normal, describe variables that can take on an infinite number of values. The normal, or bell-shaped, distribution is very widely used in business decision analysis and is referred to throughout this book.

KEY TERMS

Discrete probability distributions *(p. T1-2)*
Expected value *(p. T1-3)*
Variance *(p. T1-3)*
Standard deviation *(p. T1-4)*
Normal distribution *(p. T1-4)*

DISCUSSION QUESTIONS

1. What is the difference between a discrete probability distribution and a continuous probability distribution? Give your own example of each.
2. What is the expected value and what does it measure? How is it computed for a discrete probability distribution?
3. What is the variance and what does it measure? How is it computed for a discrete probability distribution?
4. Name three business processes that can be described by the normal distribution.

PROBLEMS

· **T1.1** Sami Abbasi Health Food stocks five loaves of Vita-Bread. The probability distribution for the sales of Vita-Bread is listed in the following table. How many loaves will Sami sell on the average?

NUMBER OF LOAVES SOLD	PROBABILITY
0	.05
1	.15
2	.20
3	.25
4	.20
5	.15

T1.2 What are the expected value and variance of the following probability distribution?

Variable, *x*	Probability
1	.05
2	.05
3	.10
4	.10
5	.15
6	.15
7	.25
8	.15

T1.3 Sales for Hobi-cat, a 17-foot catamaran sailboat, have averaged 250 boats per month over the last 5 years with a standard deviation of 25 boats. Assuming that the demand is about the same as past years and follows a normal curve, what is the probability that sales will be less than 280 boats next month?

T1.4 Refer to Problem T1.3. What is the probability that sales will be more than 265 boats during the next month? What is the probability that sales will be under 250 boats next month?

T1.5 Precision Parts is a job shop that specializes in producing electric motor shafts. The average shaft size for the E300 electric motor is .55 inch, with a standard deviation of .10 inch. It is normally distributed. What is the probability that a shaft selected at random will be between .55 and .65 inch?

T1.6 Refer to Problem T1.5. What is the probability that a shaft size will be greater than .65 inch? What is the probability that a shaft size will be between .53 and .59 inch? What is the probability that a shaft size will be under .45 inch?

T1.7 An industrial oven used to cure sand cores for a factory manufacturing engine blocks for small cars is able to maintain fairly constant temperatures. The temperature range of the oven follows a normal distribution, with a mean of 450°F and a standard deviation of 25°F. Kamvar Farahbod, president of the factory, is concerned about the large number of defective cores that have been produced in the last several months. If the oven gets hotter than 475°F, the core is defective. What is the probability that the oven will cause a core to be defective? What is the probability that the temperature of the oven will range from 460° to 470°F?

T1.8 Bill Hardgrave, production foreman for the Virginia Fruit Company, estimates that the average sales of oranges is 4,700 and the standard deviation is 500 oranges. Sales follow a normal distribution.

a) What is the probability that sales will be greater than 5,500 oranges?
b) What is the probability that sales will be greater than 4,500 oranges?
c) What is the probability that sales will be less than 4,900 oranges?
d) What is the probability that sales will be less than 4,300 oranges?

T1.9 Lori Becher has been the production manager of Medical Suppliers, Inc. (MSI), for the last 17 years. MSI is a producer of bandages and arm slings. During the last 5 years, the demand for the No-Stick bandage has been fairly constant. On the average, sales have been about 87,000 packages of No-Stick. Lori has reason to believe that the distribution of No-Stick follows a normal curve, with a standard deviation of 4,000 packages. What is the probability sales will be less than 81,000 packages?

BIBLIOGRAPHY

Berenson, Mark L., and David M. Levine. *Business Statistics.* 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2001.

Canavos, G. C., and D. M. Miller. *An Introduction to Modern Business Statistics.* Belmont, CA: Duxbury Press, 1993.

Hamburg, Morris, and Peg Young. *Statistical Analysis for Decision Making*, 6th ed. Fort Worth: Dryden Press, 1994.

Levin, R. I., D. S. Rubin, J. P. Stinson, and E. S. Gardner. *Quantitative Approaches to Management*, 11th ed. New York: McGraw-Hill, 2001.

Render, B., R. M. Stair, and M. Hanna. *Quantitative Analysis for Management*, 8th ed. Upper Saddle River, NJ: Prentice Hall, 2003.