The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Basic Statistics

Topic 16

Exam Focus

This topic addresses the concepts of expected value, variance, standard deviation, covariance, correlation, skewness, and kurtosis. The characteristics and calculations of these measures will be discussed. For the exam, be able to calculate the mean and variance of a random variable, and the covariance and correlation between two random variables. Also, be able to identify and interpret the first four moments of a statistical distribution.

The word *statistics* is used to refer to data (e.g., the average return on XYZ stock was 8% over the last ten years) and the methods we use to analyze data. Statistical methods fall into one of two categories, descriptive statistics or inferential statistics.

Descriptive statistics are used to summarize the important characteristics of large data sets. The focus of this topic is on the use of descriptive statistics to consolidate a mass of numerical data into useful information.

Inferential statistics, which will be discussed in subsequent topics, pertain to the procedures used to make forecasts, estimates, or judgments about a large set of data on the basis of the statistical characteristics of a smaller set (a sample).

A *population* is defined as the set of all possible members of a stated group. A cross-section of the returns of all of the stocks traded on the New York Stock Exchange (NYSE) is an example of a population.

It is frequently too costly or time consuming to obtain measurements for every member of a population, if it is even possible. In this case, a sample may be used. A sample is defined as a subset of the population of interest. Once a population has been defined, a sample can be drawn from the population, and the sample's characteristics can be used to describe the population as a whole. For example, a sample of 30 stocks may be selected from all of the stocks listed on the NYSE to represent the population of all NYSE-traded stocks.

Measures of Central Tendency

LO 16.1: Interpret and apply the mean, standard deviation, and variance of a random variable.

LO 16.2: Calculate the mean, standard deviation, and variance of a discrete random variable.

Measures of central tendency identify the center, or average, of a data set. This central point can then be used to represent the typical, or expected, value in the data set.

To compute the **population mean**, all the observed values in the population are summed (ΣX) and divided by the number of observations in the population, *N*. Note that the population mean is unique in that a given population only has one mean. The population mean is expressed as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

The **sample mean** is the sum of all the values in a sample of a population, ΣX, divided by the number of observations in the sample, *n*. It is used to make *inferences* about the population mean. The sample mean is expressed as:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Note the use of *n*, the sample size, versus *N*, the population size.

Example: Population mean and sample mean

Assume you and your research assistant are evaluating the stock of AXZ Corporation. You have calculated the stock returns for AXZ over the last 12 years to develop the following data set. Your research assistant has decided to conduct his analysis using only the returns for the five most recent years, which are displayed as the bold numbers in the data set. Given this information, **calculate** the population mean and the sample mean.

Data set: 12%, **25%**, **34%**, 15%, **19%**, 44%, **54%**, 33%, 22%, 28%, **17%**, 24%

Answer:

$$\mu = \text{population mean} = \frac{12+25+34+15+19+44+54+33+22+28+17+24}{12}$$
$$= 27.25\%$$

$$\overline{X} = \text{sample mean} = \frac{25+34+19+54+17}{5} = 29.8\%$$

The population mean and sample mean are both examples of arithmetic means. The arithmetic mean is the sum of the observation values divided by the number of observations. It is the most widely used measure of central tendency and has the following properties:

- All interval and ratio data sets have an arithmetic mean.
- All data values are considered and included in the arithmetic mean computation.
- A data set has only one arithmetic mean (i.e., the arithmetic mean is unique).
- The sum of the deviations of each observation in the data set from the mean is always zero.

The arithmetic mean is the only measure of central tendency for which the sum of the deviations from the mean is zero. Mathematically, this property can be expressed as follows:

$$\text{sum of mean deviations} = \sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

Example: Arithmetic mean and deviations from the mean

Compute the arithmetic mean for a data set described as:

Data set: [5, 9, 4, 10]

Answer:

The arithmetic mean of these numbers is:

$$\overline{X} = \frac{5+9+4+10}{4} = 7$$

The sum of the deviations from the mean (of 7) is:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = (5-7)+(9-7)+(4-7)+(10-7) = -2 + 2 - 3 + 3 = 0$$

Unusually large or small values can have a disproportionate effect on the computed value for the arithmetic mean. The mean of 1, 2, 3, and 50 is 14 and is not a good indication of what the individual data values really are. On the positive side, the arithmetic mean uses all the information available about the observations. The arithmetic mean of a sample from a population is the best estimate of both the true mean of the sample and the value of the next observation.

The **median** is the midpoint of a data set when the data is arranged in ascending or descending order. Half the observations lie above the median and half are below. To determine the median, arrange the data from the highest to the lowest value, or lowest to highest value, and find the middle observation.

The median is important because the arithmetic mean can be affected by extremely large or small values (outliers). When this occurs, the median is a better measure of central tendency than the mean because it is not affected by extreme values that may actually be the result of errors in the data.

Example: The median using an odd number of observations

What is the median return for five portfolio managers with 10-year annualized total returns of: 30%, 15%, 25%, 21%, and 23%?

Answer:

First, arrange the returns in descending order.

30%, 25%, 23%, 21%, 15%

Then, select the observation that has an equal number of observations above and below it—the one in the middle. For the given data set, the third observation, 23%, is the median value.

Example: The median using an even number of observations

Suppose we add a sixth manager to the previous example with a return of 28%. What is the median return?

Answer:

Arranging the returns in descending order gives us:

30%, 28%, 25%, 23%, 21%, 15%

With an even number of observations, there is no single middle value. The median value in this case is the arithmetic mean of the two middle observations, 25% and 23%. Thus, the median return for the six managers is 24.0% = 0.5(25 + 23).

Consider that while we calculated the mean of 1, 2, 3, and 50 as 14, the median is 2.5. If the data were 1, 2, 3, and 4 instead, the arithmetic mean and median would both be 2.5.

The **mode** is the value that occurs most frequently in a data set. A data set may have more than one mode or even no mode. When a distribution has one value that appears most frequently, it is said to be unimodal. When a set of data has two or three values that occur most frequently, it is said to be bimodal or trimodal, respectively.

Example: The mode

What is the mode of the following data set?

Data set: [30%, 28%, 25%, 23%, 28%, 15%, 5%]

Answer:

The mode is 28% because it is the value appearing most frequently.

The **geometric mean** is often used when calculating investment returns over multiple periods or when measuring compound growth rates. The general formula for the geometric mean, *G*, is as follows:

$$G = \sqrt[n]{X_1 \times X_2 \times \ldots \times X_n} = (X_1 \times X_2 \times \ldots \times X_n)^{1/n}$$

Note that this equation has a solution only if the product under the radical sign is non-negative.

When calculating the geometric mean for a returns data set, it is necessary to add 1 to each value under the radical and then subtract 1 from the result. The geometric mean return (R_G) can be computed using the following equation:

$$1 + R_G = \sqrt[n]{(1+R_1) \times (1+R_2) \times \ldots \times (1+R_n)}$$

where:
R_t = the return for period t

Example: Geometric mean return

For the last three years, the returns for Acme Corporation common stock have been –9.34%, 23.45%, and 8.92%. **Compute** the compound annual rate of return over the 3-year period.

Answer:

$$1 + R_G = \sqrt[3]{(-0.0934+1) \times (0.2345+1) \times (0.0892+1)}$$

$$1 + R_G = \sqrt[3]{0.9066 \times 1.2345 \times 1.0892} = \sqrt[3]{1.21903} = (1.21903)^{1/3} = 1.06825$$

$$R_G = 1.06825 - 1 = 6.825\%$$

Solve this type of problem with your calculator as follows:

- On the TI, enter 1.21903 [y^x] 0.33333 [=], or 1.21903 [y^x] 3 [1/x] [=]
- On the HP, enter 1.21903 [ENTER] 0.33333 [y^x], or 1.21903 [ENTER] 3 [1/x] [y^x]

Note that the 0.33333 represents the one-third power.

Professor's Note: The geometric mean is always less than or equal to the arithmetic mean, and the difference increases as the dispersion of the observations increases. The only time the arithmetic and geometric means are equal is when there is no variability in the observations (i.e., all observations are equal).

Expectations

LO 16.3: Interpret and calculate the expected value of a discrete random variable.

LO 16.5: Calculate the mean and variance of sums of variables.

The **expected value** is the weighted average of the possible outcomes of a random variable, where the weights are the probabilities that the outcomes will occur. The mathematical representation for the expected value of random variable X is:

$$E(X) = \Sigma P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$

Here, E is referred to as the expectations operator and is used to indicate the computation of a probability-weighted average. The symbol x_1 represents the first observed value (observation) for random variable X; x_2 is the second observation, and so on through the nth observation. The concept of expected value may be demonstrated using probabilities associated with a coin toss. On the flip of one coin, the occurrence of the event "heads" may be used to assign the value of one to a random variable. Alternatively, the event "tails" means the random variable equals zero. Statistically, we would formally write:

if heads, then X = 1

if tails, then X = 0

For a fair coin, P(heads) = P(X = 1) = 0.5, and P(tails) = P(X = 0) = 0.5. The expected value can be computed as follows:

$$E(X) = \Sigma P(x_i)x_i = P(X = 0)(0) + P(X = 1)(1) = (0.5)(0) + (0.5)(1) = 0.5$$

In any individual flip of a coin, X cannot assume a value of 0.5. Over the long term, however, the average of all the outcomes is expected to be 0.5. Similarly, the expected value of the roll of a fair die, where X = number that faces up on the die, is determined to be:

$$E(X) = \Sigma P(x_i)x_i = (1/6)(1) + (1/6)(2) + (1/6)(3) + (1/6)(4) + (1/6)(5) + (1/6)(6)$$

$$E(X) = 3.5$$

We can never roll a 3.5 on a die, but over the long term, 3.5 should be the average value of all outcomes.

The expected value is, statistically speaking, our "best guess" of the outcome of a random variable. While a 3.5 will never appear when a die is rolled, the average amount by which our guess differs from the actual outcomes is minimized when we use the expected value calculated this way.

Professor's Note: When we had historical data earlier, we calculated the mean or simple arithmetic average. The calculations given here for the expected value (or weighted mean) are based on probability models, whereas our earlier calculations were based on samples or populations of outcomes. Note that when the probabilities are equal, the simple mean is the expected value. For the roll of a die, all six outcomes are equally likely, so $\frac{1+2+3+4+5+6}{6} = 3.5$ *gives us the same expected value as the probability model. However, with a probability model, the probabilities of the possible outcomes need not be equal, and the simple mean is not necessarily the expected outcome, as the following example illustrates.*

Example: Expected earnings per share

The probability distribution of EPS for Ron's Stores is given in the figure below. **Calculate** the expected earnings per share.

EPS Probability Distribution

Probability	*Earnings Per Share*
10%	£1.80
20%	£1.60
40%	£1.20
30%	£1.00
100%	

Answer:

The expected EPS is simply a weighted average of each possible EPS, where the weights are the probabilities of each possible outcome.

$$E(EPS) = 0.10(1.80) + 0.20(1.60) + 0.40(1.20) + 0.30(1.00) = £1.28$$

Properties of expectation include:

1. If *c* is any constant, then:

   $$E(cX) = cE(X)$$

2. If *X* and *Y* are any random variables, then:

   $$E(X + Y) = E(X) + E(Y)$$

Professor's Note: This property displays the mean of the sum of random variables. It is simply the sum of the individual random variable means.

3. If *c* and *a* are constants, then:

   $$E(cX + a) = cE(X) + a$$

4. If *X* and *Y* are independent random variables, then:

   $$E(XY) = E(X) \times E(Y)$$

5. If *X* and *Y* are NOT independent, then:

   $$E(XY) \neq E(X) \times E(Y)$$

6. If *X* is a random variable, then:

   $$E(X^2) \neq [E(X)]^2$$

Variance and Standard Deviation

The mean and variance of a distribution are defined as the first and second moments of the distribution, respectively. **Variance** is defined as:

$$\text{Var}(X) = E[(X - \mu)^2]$$

The square root of the variance is called the **standard deviation**. The variance and standard deviation provide a measure of the extent of the dispersion in the values of the random variable around the mean.

Properties of variance include:

1. $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$

   where $\mu = E(X)$

2. If *c* is any constant, then:

   $\text{Var}(c) = 0$

3. If *c* is any constant, then:

   $\text{Var}(cX) = c^2 \times \text{Var}(X)$

4. If *c* is any constant, then:

   $\text{Var}(X + c) = \text{Var}(X)$

5. If *a* and *c* are constants, then:

   $\text{Var}(aX + c) = a^2 \times \text{Var}(X)$

6. If *X* and *Y* are independent random variables, then:

   $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

   $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

7. If *X* and *Y* are independent and *a* and *c* are constants, then:

   $\text{Var}(aX + cY) = a^2 \times \text{Var}(X) + c^2 \times \text{Var}(Y)$

Example: Computing variance and standard deviation

What is the variance and standard deviation of the sum of points in tossing a single coin if heads = 2 points and tails = 10 points?

Answer:

$$\mu = (2 + 10) / 2 = 6$$

$$\text{Var}(X) = (2 - 6)^2 \times 0.5 + (10 - 6)^2 \times 0.5$$

$$\text{Var}(X) = 8 + 8 = 16$$

$$\text{standard deviation}(X) = \sqrt{16} = 4$$

Covariance and Correlation

LO 16.4: Calculate and interpret the covariance and correlation between two random variables.

The variance and standard deviation measure the dispersion, or volatility, of only one variable. In many finance situations, however, we are interested in how two random variables move in relation to each other. For investment applications, one of the most frequently analyzed pairs of random variables is the returns of two assets. Investors and managers frequently ask questions such as, "What is the relationship between the return for Stock A and Stock B?" or "What is the relationship between the performance of the S&P 500 and that of the automotive industry?" As you will soon see, the covariance provides useful information about how two random variables, such as asset returns, are related.

Covariance is the expected value of the product of the deviations of the two random variables from their respective expected values. A common symbol for the covariance between random variables X and Y is Cov(X,Y). Since we will be mostly concerned with the covariance of asset returns, the following formula has been written in terms of the covariance of the return of asset i, R_i, and the return of asset j, R_j:

$$\text{Cov}(R_i,R_j) = E\{[R_i - E(R_i)][R_j - E(R_j)]\}$$

This equation simplifies to:

$$\text{Cov}(R_i,R_j) = E(R_i,R_j) - E(R_i) \times E(R_j)$$

Properties of covariance include:

1. If X and Y are independent random variables, then:

   $$\text{Cov}(X,Y) = 0$$

2. The covariance of random variable X with itself is the variance of X.

   $$\text{Cov}(X,X) = \text{Var}(X)$$

3. If a, b, c, and d are constants, then:

   $$\text{Cov}(a + bX, c + dY) = b \times d \times \text{Cov}(X,Y)$$

4. If X and Y are NOT independent, then:

   $$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \times \text{Cov}(X,Y)$$

   $$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \times \text{Cov}(X,Y)$$

Professor's Note: When discussing the properties of variance, we showed the variance of the sum of independent random variable variances. The covariance term was not present in this earlier expression because the variables did not influence each other. However, when random variables are not independent, two times the covariance of the random variables must be included as demonstrated in the above property.

To aid in the interpretation of covariance, consider the returns of a stock and of a put option on the stock. These two returns will have a negative covariance because they move in opposite directions. The returns of two automotive stocks would likely have a positive covariance, and the returns of a stock and a riskless asset would have a zero covariance because the riskless asset's returns never move, regardless of movements in the stock's return.

Example: Covariance

Assume that the economy can be in three possible states (S) next year: boom, normal, or slow economic growth. An expert source has calculated that P(boom) = 0.30, P(normal) = 0.50, and P(slow) = 0.20. The returns for Stock A, R_A, and Stock B, R_B, under each of the economic states are provided in the table below. What is the covariance of the returns for Stock A and Stock B?

Answer:

First, the expected returns for each of the stocks must be determined.

$$E(R_A) = (0.3)(0.20) + (0.5)(0.12) + (0.2)(0.05) = 0.13$$

$$E(R_B) = (0.3)(0.30) + (0.5)(0.10) + (0.2)(0.00) = 0.14$$

The covariance can now be computed using the procedure described in the following table:

Covariance Computation

Event	*P(S)*	R_A	R_B	$P(S) \times [R_A - E(R_A)] \times [R_B - E(R_B)]$
Boom	0.3	0.20	0.30	(0.3)(0.2 – 0.13)(0.3 – 0.14) = 0.00336
Normal	0.5	0.12	0.10	(0.5)(0.12 – 0.13)(0.1 – 0.14) = 0.00020
Slow	0.2	0.05	0.00	(0.2)(0.05 – 0.13)(0 – 0.14) = 0.00224
				$Cov(R_A, R_B) = \Sigma P(S) \times [R_A - E(R_A)] \times [R_B - E(R_B)] = 0.00580$

In practice, the covariance is difficult to interpret. This is mostly because it can take on extremely large values, ranging from negative to positive infinity, and, like the variance, these values are expressed in terms of squared units.

To make the covariance of two random variables easier to interpret, it may be divided by the product of the random variables' standard deviations. The resulting value is called the correlation coefficient, or simply, **correlation**. The relationship between covariances, standard deviations, and correlations can be seen in the following expression for the correlation of the returns for asset *i* and *j*:

$$\text{Corr}(R_i, R_j) = \frac{\text{Cov}(R_i, R_j)}{\sigma(R_i)\sigma(R_j)}, \text{ which implies } \text{Cov}(R_i, R_j) = \text{Corr}(R_i, R_j)\sigma(R_i)\sigma(R_j)$$

The correlation between two random return variables may also be expressed as $\rho(R_i, R_j)$, or $\rho_{i,j}$.

Properties of correlation of two random variables R_i and R_j are summarized here:

- Correlation measures the strength of the linear relationship between two random variables.
- Correlation has no units.
- The correlation ranges from –1 to +1. That is, $-1 \leq \text{Corr}(R_i, R_j) \leq +1$.
- If $\text{Corr}(R_i, R_j) = 1.0$, the random variables have perfect positive correlation. This means that a movement in one random variable results in a proportional positive movement in the other relative to its mean.

- If $\text{Corr}(R_i, R_j) = -1.0$, the random variables have perfect negative correlation. This means that a movement in one random variable results in an exact opposite proportional movement in the other relative to its mean.
- If $\text{Corr}(R_i, R_j) = 0$, there is no linear relationship between the variables, indicating that prediction of R_i cannot be made on the basis of R_j using linear methods.

Example: Correlation

Using our previous example, **compute** and **interpret** the correlation of the returns for stocks A and B, given that $\sigma^2(R_A) = 0.0028$ and $\sigma^2(R_B) = 0.0124$ and recalling that $\text{Cov}(R_A,R_B) = 0.0058$.

Answer:

First, it is necessary to convert the variances to standard deviations.

$$\sigma(R_A) = (0.0028)^{½} = 0.0529$$

$$\sigma(R_B) = (0.0124)^{½} = 0.1114$$

Now, the correlation between the returns of Stock A and Stock B can be computed as follows:

$$\text{Corr}(R_A, R_B) = \frac{0.0058}{(0.0529)(0.1114)} = 0.9842$$

The interpretation of the possible correlation values is summarized in Figure 1.

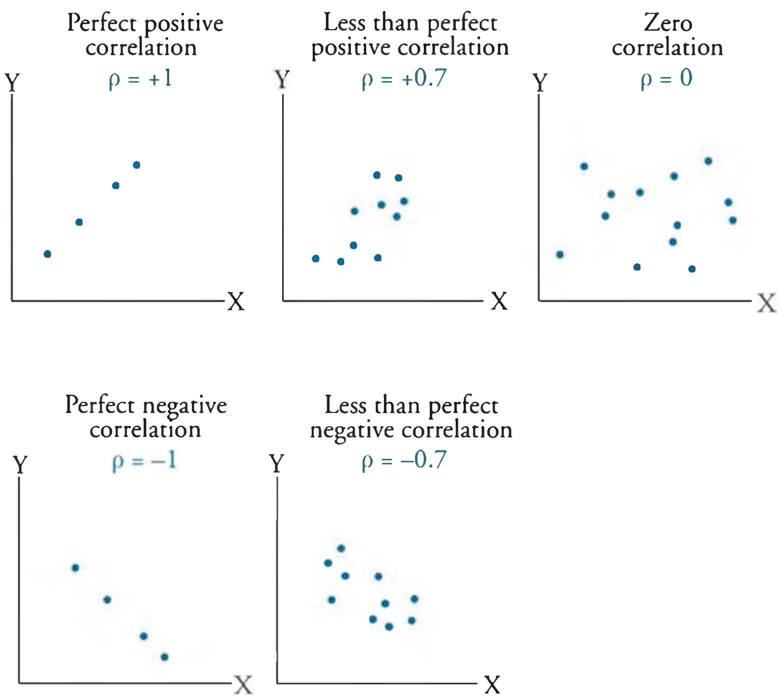
Figure 1: Interpretation of Correlation Coefficients

Correlation Coefficient (ρ)	*Interpretation*
$\rho = +1$	perfect positive correlation
$0 < \rho < +1$	a positive linear relationship
$\rho = 0$	no linear relationship
$-1 < \rho < 0$	a negative linear relationship
$\rho = -1$	perfect negative correlation

Interpreting a Scatter Plot

A scatter plot is a collection of points on a graph where each point represents the values of two variables (i.e., an X/Y pair). Figure 2 shows several scatter plots for the two random variables *X* and *Y* and the corresponding interpretation of correlation. As shown, an upward-sweeping scatter plot indicates a positive correlation between the two variables, while a downward-sweeping plot implies a negative correlation. Also illustrated in Figure 2 is that as we move from left to right in the rows of scatter plots, the extent of the linear relationship between the two variables deteriorates, and the correlation gets closer to zero. Note that for $\rho = 1$ and $\rho = -1$, the data points lie exactly on a line, but the slope of that line is not necessarily +1 or −1.

Figure 2: Interpretations of Correlation



Moments and Central Moments

LO 16.6: Describe the four central moments of a statistical variable or distribution: mean, variance, skewness and kurtosis.

The shape of a probability distribution can be described by the "moments" of the distribution. Raw moments are measured relative to an expected value raised to the appropriate power. The first raw moment is the **mean** of the distribution, which is the expected value of returns:

$$E(R) = \mu = \sum_{i=1}^{n} p_i R_i^1$$

where:
p_i = probability of event i
R_i = return associated with event i

Generalizing, the *k*th raw moment is the expected value of R^k:

$$E(R^k) = \sum_{i=1}^{n} p_i R_i^k$$

Raw moments for k > 1 are not very useful for our purposes, however, central moments for k > 1 are important.

Central moments are measured relative to the mean (i.e., central around the mean). The *k*th central moment is defined as:

$$E(R-\mu)^k = \sum_{i=1}^{n} p_i (R_i - \mu)^k$$

Professor's Note: Since central moments are measured relative to the mean, the first central moment equals zero and is, therefore, not typically used.

The second central moment is the **variance** of the distribution, which measures the dispersion of data.

$$\text{variance} = \sigma^2 = E\left[(R-\mu)^2\right]$$

Professor's Note: Since moments higher than the second central moment can be difficult to interpret, they are typically standardized by dividing the central moment by σ^k.

The third central moment measures the departure from symmetry in the distribution. This moment will equal zero for a symmetric distribution (such as the normal distribution).

$$\text{third central moment} = E\left[(R-\mu)^3\right]$$

The **skewness** statistic is the standardized third central moment. Skewness (sometimes called *relative skewness*) refers to the extent to which the distribution of data is not symmetric around its mean. It is calculated as:

$$\text{skewness} = \frac{E\left[(R-\mu)^3\right]}{\sigma^3}$$

The fourth central moment measures the degree of clustering in the distribution.

$$\text{fourth central moment} = E\left[(R-\mu)^4\right]$$

The **kurtosis** statistic is the standardized fourth central moment of the distribution. Kurtosis refers to the degree of peakedness or clustering in the data distribution and is calculated as:

$$\text{kurtosis} = \frac{E\left[(R-\mu)^4\right]}{\sigma^4}$$

Kurtosis for the normal distribution equals 3. Therefore, the **excess kurtosis** for any distribution equals:

$$\text{excess kurtosis} = \text{kurtosis} - 3$$

Although additional central moments can be calculated, risk management is not often concerned with anything beyond the fourth central moment.

Skewness and Kurtosis

LO 16.7: Interpret the skewness and kurtosis of a statistical distribution, and interpret the concepts of coskewness and cokurtosis.

A distribution is symmetrical if it is shaped identically on both sides of its mean. Distributional symmetry implies that intervals of losses and gains will exhibit the same frequency. For example, a symmetrical distribution with a mean return of zero will have losses in the –6% to –4% interval as frequently as it will have gains in the +4% to +6% interval. The extent to which a returns distribution is symmetrical is important because the degree of symmetry tells analysts if deviations from the mean are more likely to be positive or negative.

Skewness, or skew, refers to the extent to which a distribution is not symmetrical. Nonsymmetrical distributions may be either positively or negatively skewed and result from the occurrence of outliers in the data set. Outliers are observations with extraordinarily large values, either positive or negative.

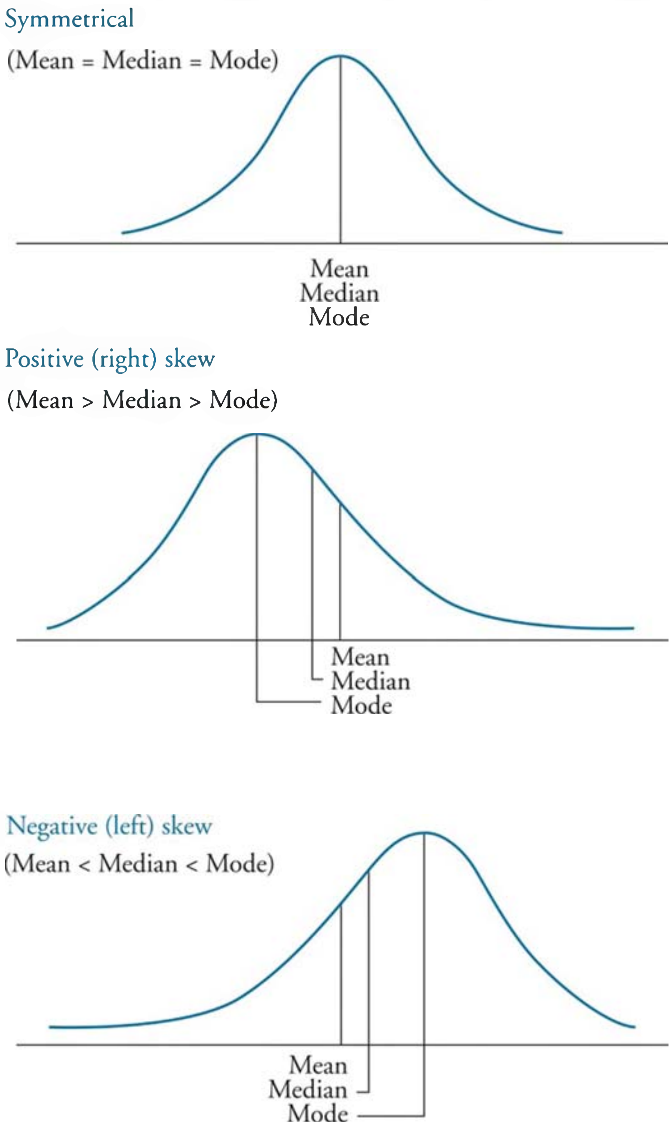
- A *positively skewed* distribution is characterized by many outliers in the upper region, or right tail. A positively skewed distribution is said to be skewed right because of its relatively long upper (right) tail.
- A *negatively skewed* distribution has a disproportionately large amount of outliers that fall within its lower (left) tail. A negatively skewed distribution is said to be skewed left because of its long lower tail.

Skewness affects the location of the mean, median, and mode of a distribution.

- For a symmetrical distribution, the mean, median, and mode are equal.
- For a positively skewed, unimodal distribution, the mode is less than the median, which is less than the mean. The mean is affected by outliers; in a positively skewed distribution, there are large, positive outliers which will tend to "pull" the mean upward, or more positive. An example of a positively skewed distribution is that of housing prices. Suppose you live in a neighborhood with 100 homes; 99 of them sell for $100,000, and one sells for $1,000,000. The median and the mode will be $100,000, but the mean will be $109,000. Hence, the mean has been "pulled" upward (to the right) by the existence of one home (outlier) in the neighborhood.
- For a negatively skewed, unimodal distribution, the mean is less than the median, which is less than the mode. In this case, there are large, negative outliers that tend to "pull" the mean downward (to the left).

Professor's Note: The key to remembering how measures of central tendency are affected by skewed data is to recognize that skew affects the mean more than the median and mode, and the mean is "pulled" in the direction of the skew. The relative location of the mean, median, and mode for different distribution shapes is shown in Figure 3. Note the median is between the other two measures for positively or negatively skewed distributions.

Figure 3: Effect of Skewness on Mean, Median, and Mode

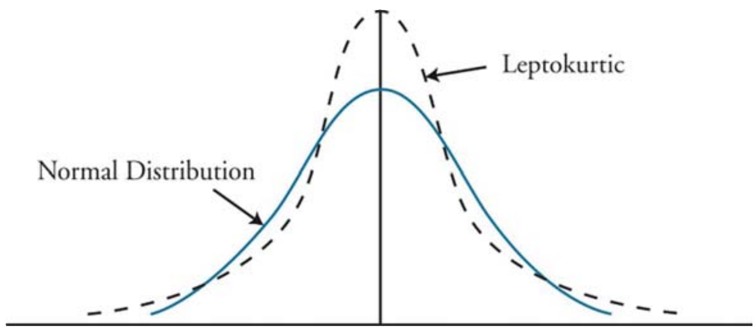


Kurtosis is a measure of the degree to which a distribution is more or less "peaked" than a normal distribution. **Leptokurtic** describes a distribution that is more peaked than a normal distribution, whereas **platykurtic** refers to a distribution that is less peaked (or flatter) than a normal distribution. A distribution is **mesokurtic** if it has the same kurtosis as a normal distribution.

As indicated in Figure 4, a leptokurtic return distribution will have more returns clustered around the mean and more returns with large deviations from the mean (fatter tails).

Relative to a normal distribution, a leptokurtic distribution will have a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean. This means there is a relatively greater probability of an observed value being either close to the mean or far from the mean. With regard to an investment returns distribution, a greater likelihood of a large deviation from the mean return is often perceived as an increase in risk.

Figure 4: Kurtosis



A distribution is said to exhibit **excess kurtosis** if it has either more or less kurtosis than the normal distribution. The computed kurtosis for all normal distributions is three. Statisticians, however, sometimes report excess kurtosis, which is defined as kurtosis minus three. Thus, a normal distribution has excess kurtosis equal to zero, a leptokurtic distribution has excess kurtosis greater than zero, and platykurtic distributions will have excess kurtosis less than zero.

Kurtosis is critical in a risk management setting. Most research about the distribution of securities returns has shown that returns are not normally distributed. Actual securities returns tend to exhibit both skewness and kurtosis. Skewness and kurtosis are critical concepts for risk management because when securities returns are modeled using an assumed normal distribution, the predictions from the models will not take into account the potential for extremely large, negative outcomes. In fact, most risk managers put very little emphasis on the mean and standard deviation of a distribution and focus more on the distribution of returns in the tails of the distribution—that is where the risk is. In general, greater positive kurtosis and more negative skew in returns distributions indicates increased risk.

Coskewness and Cokurtosis

Previously, we identified moments and central moments for mean and variance. In a similar fashion, we can identify cross central moments for the concept of covariance. The third cross central moment is known as **coskewness** and the fourth cross central moment is known as **cokurtosis**.

To illustrate the importance of these concepts in risk management, suppose we are analyzing the returns data from four different stocks over a 7-year time period (shown in Figure 5). Although returns vary over time, the mean, variance, skewness, and kurtosis of all stock returns are the same under this scenario. In addition, the covariance between returns for Stock 1 and Stock 2 is equal to the covariance between returns for Stock 3 and Stock 4.

Figure 5: Stock Returns

	Stocks			
Time	1	2	3	4
1	0.0%	−2.4%	−12.6%	−12.6%
2	−2.4%	−12.6%	−5.3%	−5.3%
3	−12.6%	2.4%	0.0%	−2.4%
4	−5.3%	−5.3%	−2.4%	12.6%
5	2.4%	0.0%	2.4%	0.0%
6	5.3%	5.3%	5.3%	5.3%
7	12.6%	12.6%	12.6%	2.4%

By combining Stock 1 and Stock 2 into Portfolio A, and Stock 3 and Stock 4 into Portfolio B (shown in Figure 6), we find that the returns for Portfolio A and Portfolio B have the same mean and variance. However, these combined return sets do not have the same skewness (i.e., the coskewness between stocks in the portfolios is different). The reason for this difference is that the ranking of returns over time (e.g., from best to worst) is different for each stock, and when combined in a portfolio, these differences skew the portfolio returns distribution. For example, the worst return for Stock 1 occurred during time period 3, but in Portfolio A, the worst return occurred during time period 2. Similarly, the best return for Stock 4 occurred during time period 4, but in Portfolio B, the best return occurred during time period 7.

Figure 6: Portfolio Returns

	Portfolio	
Time	A	B
1	−1.2%	−12.6%
2	−7.5%	−5.3%
3	−5.1%	−1.2%
4	−5.3%	5.1%
5	1.2%	1.2%
6	5.3%	5.3%
7	12.6%	7.5%

From a risk management standpoint, it is helpful to know that the worst outcome in Portfolio B is 1.7 times greater than the worst outcome in Portfolio A. So, although the mean and variance of these portfolios are equal, shortfall risk expectations can differ depending on time period. This is important information to know, however, most risk models choose to ignore the effects of coskewness and cokurtosis. The reason being is that as the number of variables increase, the number of coskewness and cokurtosis terms will increase rapidly, making the data much more difficult to analyze. Practitioners instead opt to use more tractable risk models, such as GARCH (see Topic 28), which capture the essence of coskewness and cokurtosis by incorporating time-varying volatility and/or time-varying correlation.

The Best Linear Unbiased Estimator

LO 16.8: Describe and interpret the best linear unbiased estimator.

In upcoming topics, we will continue to discuss statistics and explore how sample parameters can be used to draw conclusions about population parameters. **Point estimates** are single (sample) values used to estimate population parameters, and the formula used to compute a point estimate is known as an **estimator**.

There are certain statistical properties that make some estimates more desirable than others. These desirable properties of an estimator are unbiasedness, efficiency, consistency, and linearity.

- An *unbiased* estimator is one for which the expected value of the estimator is equal to the parameter you are trying to estimate. For example, because the expected value of the sample mean is equal to the population mean $\left[E(\bar{x}) = \mu\right]$, the sample mean is an unbiased estimator of the population mean.
- An unbiased estimator is also *efficient* if the variance of its sampling distribution is smaller than all the other unbiased estimators of the parameter you are trying to estimate. The sample mean, for example, is an unbiased and efficient estimator of the population mean.
- A *consistent* estimator is one for which the accuracy of the parameter estimate increases as the sample size increases. As the sample size increases, the sampling distribution bunches more closely around the population mean.
- A point estimate is a *linear* estimator when it can be used as a linear function of sample data.

If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the **best linear unbiased estimator** (BLUE).

Key Concepts

LO 16.1

To compute the population mean, all the observed values in the population are summed and divided by the number of observations in the population.

Variance and standard deviation provide a measure of the extent of the dispersion in the values of the random variable around the mean.

LO 16.2

The mean of a population is expressed as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Variance of a random variable is defined as:

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$
where $\mu = E(X)$

The square root of the variance is called the standard deviation.

LO 16.3

Expected value is the weighted average of the possible outcomes of a random variable, where the weights are the probabilities that the outcomes will occur. The expectation of a random variable X having possible values $x_1, \ldots, x_n$ is defined as:

$$E(X) = x_1 P(X = x_1) + \ldots + x_n P(X = x_n)$$

LO 16.4

Covariance measures the extent to which two random variables tend to be above and below their respective means for each joint realization. It can be calculated as:

$$\mathrm{Cov}(A,B) = \sum_{i=1}^{N} P_i\left(A_i - \bar{A}\right)\left(B_i - \bar{B}\right)$$

Correlation is a standardized measure of association between two random variables; it ranges in value from –1 to +1 and is equal to:

$$\frac{\mathrm{Cov}(A,B)}{\sigma_A \sigma_B}$$

LO 16.5

If *X* and *Y* are any random variables, then:

$$E(X + Y) = E(X) + E(Y)$$

If *X* and *Y* are independent random variables, then:

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

If *X* and *Y* are NOT independent, then:

$$Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X,Y)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2 \times Cov(X,Y)$$

LO 16.6

The shape of a probability distribution is characterized by its raw moments and central moments. The first raw moment is the mean of the distribution. The second central moment is the variance. The third central moment divided by the cube of the standard deviation measures the skewness of the distribution, and the fourth central moment divided by the fourth power of the standard deviation measures the kurtosis of the distribution.

LO 16.7

Skewness describes the degree to which a distribution is nonsymmetric about its mean.

- A right-skewed distribution has positive skewness and a mean that is higher than the median that is higher than the mode.
- A left-skewed distribution has negative skewness and a mean that is lower than the median that is lower than the mode.

Kurtosis measures the peakedness of a distribution and the probability of extreme outcomes.

- Excess kurtosis is measured relative to a normal distribution, which has a kurtosis of three.
- Positive values of excess kurtosis indicate a distribution that is leptokurtic (fat tails, more peaked).
- Negative values of excess kurtosis indicate a platykurtic distribution (thin tails, less peaked).

Like mean and variance, we can generalize covariance to cross central moments. The third cross central moment is coskewness and the fourth cross central moment is cokurtosis.

LO 16.8

Desirable statistical properties of an estimator include unbiasedness (sign of estimation error is random), efficiency (lower sampling error than any other unbiased estimator), consistency (variance of sampling error decreases with sample size), and linearity (used as a linear function of sample data).

Concept Checkers

1. A distribution of returns that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean:
   A. is positively skewed.
   B. is a symmetric distribution.
   C. has positive excess kurtosis.
   D. has negative excess kurtosis.

2. The correlation of returns between Stocks A and B is 0.50. The covariance between these two securities is 0.0043, and the standard deviation of the return of Stock B is 26%. The variance of returns for Stock A is:
   A. 0.0331.
   B. 0.0011.
   C. 0.2656.
   D. 0.0112.

Use the following data to answer Questions 3 and 4.

Probability Matrix

Returns	R_B = 50%	R_B = 20%	R_B = –30%
R_A = –10%	40%	0%	0%
R_A = 10%	0%	30%	0%
R_A = 30%	0%	0%	30%

3. Given the probability matrix above, the standard deviation of Stock B is closest to:
   A. 0.11.
   B. 0.22.
   C. 0.33.
   D. 0.15.

4. Given the probability matrix above, the covariance between Stock A and B is closest to:
   A. –0.160.
   B. –0.055.
   C. 0.004.
   D. 0.020.

5. A discrete uniform distribution (each event has an equal probability of occurrence) has the following possible outcomes for X: [1, 2, 3, 4]. The variance of this distribution is closest to:
   A. 1.00.
   B. 1.25.
   C. 1.50.
   D. 2.00.

Concept Checker Answers

1. C A distribution that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean will be leptokurtic and will exhibit excess kurtosis (positive). The distribution will be taller and have fatter tails than a normal distribution.

2. B $\text{Corr}(R_A, R_B) = \dfrac{\text{Cov}(R_A, R_B)}{[\sigma(R_A)][\sigma(R_B)]}$

$$\sigma^2(R_A) = \left[\frac{\text{Cov}(R_A, R_B)}{\sigma(R_B)\text{Corr}(R_A, R_B)}\right]^2 = \left[\frac{0.0043}{(0.26)(0.5)}\right]^2 = 0.0331^2 = 0.0011$$

3. C Expected return of Stock B = (0.4)(0.5) + (0.3)(0.2) + (0.3)(–0.3) = 0.17

$\text{Var}(R_B) = 0.4(0.5 - 0.17)^2 + 0.3(0.2 - 0.17)^2 + 0.3(-0.3 - 0.17)^2 = 0.1101$

Standard deviation = $\sqrt{0.1101} = 0.3318$

4. B $\text{Cov}(R_A, R_B) = 0.4(-0.1 - 0.08)(0.5 - 0.17) + 0.3(0.1 - 0.08)(0.2 - 0.17) + 0.3(0.3 - 0.08)(-0.3 - 0.17) = -0.0546$

5. B Expected value = (1/4)(1 + 2 + 3 + 4) = 2.5

Variance = $(1/4)[(1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2] = 1.25$

Note that since each observation is equally likely, each has 25% (1/4) chance of occurrence.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Distributions

Topic 17

Exam Focus

This topic explores common probability distributions: uniform, Bernoulli, binomial, Poisson, normal, lognormal, chi-squared, Student's t, and F. You will learn the properties, parameters, and common occurrences of these distributions. Also discussed is the central limit theorem, which allows us to use sampling statistics to construct confidence intervals for point estimates of population means. For the exam, focus most of your attention on the binomial, normal, and Student's t distributions. Also, know how to standardize a normally distributed random variable, how to use a z-table, and how to construct confidence intervals.

Parametric and Nonparametric Distributions

Probability distributions are classified into two categories: parametric and nonparametric. **Parametric distributions**, such as a normal distribution, can be described by using a mathematical function. These types of distributions make it easier to draw conclusions about the data; however, they also make restrictive assumptions, which are not necessarily supported by real-world patterns. **Nonparametric distributions**, such as a historical distribution, cannot be described by using a mathematical function. Instead of making restrictive assumptions, these types of distributions fit the data perfectly; however, without generalizing the data, it can be difficult for a researcher to draw any conclusions.

LO 17.1: Distinguish the key properties among the following distributions: uniform distribution, Bernoulli distribution, Binomial distribution, Poisson distribution, normal distribution, lognormal distribution, Chi-squared distribution, Student's t, and F-distributions, and identify common occurrences of each distribution.

The Uniform Distribution

The **continuous uniform distribution** is defined over a range that spans between some lower limit, *a*, and some upper limit, *b*, which serve as the parameters of the distribution. Outcomes can only occur between *a* and *b*, and since we are dealing with a continuous distribution, even if $a < x < b$, $P(X = x) = 0$. Formally, the properties of a continuous uniform distribution may be described as follows:

- For all $a \leq x_1 < x_2 \leq b$ (i.e., for all x_1 and x_2 between the boundaries *a* and *b*).
- $P(X < a \text{ or } X > b) = 0$ (i.e., the probability of *X* outside the boundaries is zero).
- $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$. This defines the probability of outcomes between x_1 and x_2.

Don't miss how simple this is just because the notation is so mathematical. For a continuous uniform distribution, the probability of outcomes in a range that is one-half the whole range is 50%. The probability of outcomes in a range that is one-quarter as large as the whole possible range is 25%.

Example: Continuous uniform distribution

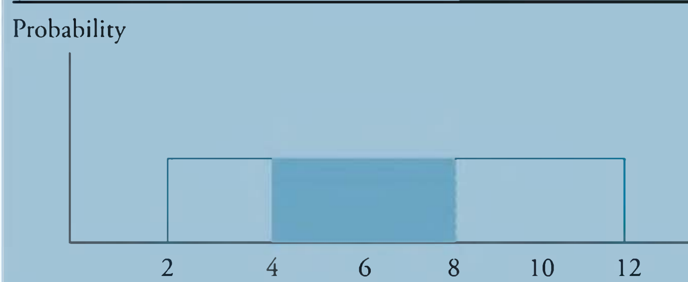
X is uniformly distributed between 2 and 12. **Calculate** the probability that *X* will be between 4 and 8.

Answer:

$$\frac{8-4}{12-2}=\frac{4}{10}=40\%$$

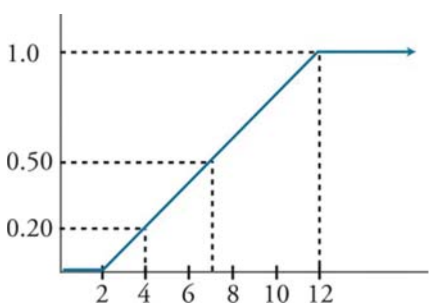
The figure below illustrates this continuous uniform distribution. Note that the area bounded by 4 and 8 is 40% of the total probability between 2 and 12 (which is 100%).

Continuous Uniform Distribution



Since outcomes are equal over equal-size possible intervals, the cumulative distribution function (cdf) is linear over the variable's range. The cdf for the distribution in the above example, Prob (X < x), is shown in Figure 1.

Figure 1: CDF for a Continuous Uniform Variable



The probability function for a continuous random variable is called the probability density function (pdf) and is denoted f(x). Symbolically, the probability density function for a continuous uniform distribution is expressed as:

$$f(x)=\frac{1}{b-a}\text{ for } a \le x \le b\text{, else } f(x)=0$$

The mean and variance, respectively, of a uniform distribution are:

$$E(x) = \frac{a+b}{2}$$

$$Var(x) = \frac{(b-a)^2}{12}$$

The Bernoulli Distribution

A Bernoulli distributed random variable only has two possible outcomes. The outcomes can be defined as either a "success" or a "failure." The probability of success, p, may be denoted with the value "1" and the probability of failure, $1-p$, may be denoted with the value "0." Bernoulli distributed random variables are commonly used for assessing whether or not a company defaults during a specified time period. In the default example, the random variable equals "1" in the event of default and "0" in the event of survival.

The Binomial Distribution

A **binomial random variable** may be defined as the number of "successes" in a given number of trials, whereby the outcome can be either "success" or "failure." The probability of success, p, is constant for each trial and the trials are independent. A binomial random variable for which the number of trials is 1 is called a Bernoulli random variable. Think of a trial as a mini-experiment (or Bernoulli trial). The final outcome is the number of successes in a series of n trials. Under these conditions, the binomial probability function defines the probability of x successes in n trials. It can be expressed using the following formula:

$$p(x) = P(X = x) = (\text{number of ways to choose } x \text{ from } n)p^x(1-p)^{n-x}$$

where:

$$(\text{number of ways to choose x from n}) = \frac{n!}{(n-x)!x!}$$

p = the probability of "success" on each trial [don't confuse it with p(x)]

So the probability of exactly x successes in n trials is:

$$p(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

Example: Binomial probability

Assuming a binomial distribution, **compute** the probability of drawing three black beans from a bowl of black and white beans if the probability of selecting a black bean in any given attempt is 0.6. You will draw five beans from the bowl.

Answer:

$$P(X=3)=p(3)=\frac{5!}{2!3!}(0.6)^3(0.4)^2=(120\ /\ 12)(0.216)(0.160)=0.3456$$

Some intuition about these results may help you remember the calculations. Consider that a (very large) bowl of black and white beans has 60% black beans and that each time you select a bean, you replace it in the bowl before drawing again. We want to know the probability of selecting exactly three black beans in five draws, as in the previous example.

One way this might happen is BBBWW. Since the draws are independent, the probability of this is easy to calculate. The probability of drawing a black bean is 60%, and the probability of drawing a white bean is 1 – 60% = 40%. Therefore, the probability of selecting BBBWW, in order, is 0.6 × 0.6 × 0.6 × 0.4 × 0.4 = 3.456%. This is the $p^3(1-p)^2$ from the formula and *p* is 60%, the probability of selecting a black bean on any single draw from the bowl. BBBWW is not, however, the only way to choose exactly three black beans in five trials. Another possibility is BBWWB, and a third is BWWBB. Each of these will have exactly the same probability of occurring as our initial outcome, BBBWW. That's why we need to answer the question of how many ways (different orders) there are for us to choose three black beans in five draws. Using the formula, there are $\frac{5!}{(5-3)!3!}=10$ ways; 10 × 3.456% = 34.56%, the answer we computed above.

Expected Value and Variance of a Binomial Random Variable

For a given series of *n* trials, the expected number of successes, or E(X), is given by the following formula:

expected value of X = E(X) = np

The intuition is straightforward; if we perform *n* trials and the probability of success on each trial is *p*, we expect *np* successes.

The variance of a binomial random variable is given by:

variance of X = np(1 – p) = npq

Professor's Note: q = 1 – p is the probability that the event will fail to occur in a single trial (i.e., the probability of failure).

Example: Expected value of a binomial random variable

Based on empirical data, the probability that the Dow Jones Industrial Average (DJIA) will increase on any given day has been determined to equal 0.67. Assuming the only other outcome is that it decreases, we can state p(UP) = 0.67 and p(DOWN) = 0.33. Further, assume that movements in the DJIA are independent (i.e., an increase in one day is independent of what happened on another day).

Using the information provided, **compute** the expected value of the number of up days in a 5-day period.

Answer:

Using binomial terminology, we define success as UP, so p = 0.67. Note that the definition of success is critical to any binomial problem.

$$E(X \mid n = 5, p = 0.67) = (5)(0.67) = 3.35$$

Recall that the "|" symbol means *given*. Hence, the preceding statement is read as: the expected value of X given that n = 5, and the probability of success = 67% is 3.35.

Using the equation for the variance of a binomial distribution, we find the variance of *X* to be:

$$\text{Var}(X) = np(1 - p) = 5(0.67)(0.33) = 1.106$$

We should note that since the binomial distribution is a discrete distribution, the result X = 3.35 is not possible. However, if we were to record the results of many 5-day periods, the average number of up days (successes) would converge to 3.35.

Binomial distributions are used extensively in the investment world where outcomes are typically seen as successes or failures. In general, if the price of a security goes up, it is viewed as a success. If the price of a security goes down, it is a failure. In this context, binomial distributions are often used to create models to aid in the process of asset valuation.

Professor's Note: We will examine binomial trees for stock option valuation in Book 4.

The Poisson Distribution

The Poisson distribution is another discrete probability distribution with a number of real-world applications. For example, the number of defects per batch in a production process or the number of calls per hour arriving at the 911 emergency switchboard are discrete random variables that follow a Poisson distribution.

While the Poisson random variable *X* refers to the *number of successes per unit*, the parameter lambda (λ) refers to the average or *expected number of successes per unit*. The mathematical expression for the Poisson distribution for obtaining *X* successes, given that λ successes are expected, is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

An interesting feature of the Poisson distribution is that both its mean and variance are equal to the parameter, λ.

Example: Using the Poisson distribution (1)

On average, the 911 emergency switchboards receive 0.1 incoming calls per second. What is the probability that in a given minute exactly 5.0 phone calls will be received, assuming the arrival of calls follows a Poisson distribution?

Answer:

We first need to convert the seconds into minutes. Note that λ, the expected number of calls per minute, is (0.1)(60) = 6.0. Hence:

$$P(X = 5) = \frac{6^5 e^{-6}}{5!} = 0.1606 = 16.06\%$$

This means that, given the average of 0.1 incoming calls per second, there is a 16.06% chance there will be five incoming phone calls in a minute.

Example: Using the Poisson distribution (2)

Assume there is a 0.01 probability of a patient experiencing severe weight loss as a side effect from taking a recently approved drug used to treat heart disease. What is the probability that out of 200 such procedures conducted on different patients, five patients will develop this complication? Assume that the number of patients developing the complication from the procedure is Poisson-distributed.

Answer:

Let X = expected number of patients developing the complication from the procedure
= np = (200)(0.01) = 2

$$P(X = 5) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^5 e^{-2}}{5!} = 0.036 = 3.6\%$$

This means that given a complication rate of 0.01, there is a 3.6% probability that 5 out of every 200 patients will experience severe weight loss from taking the drug.

The Normal Distribution

The normal distribution is important for many reasons. Many of the random variables that are relevant to finance and other professional disciplines follow a normal distribution. In the area of investment and portfolio management, the normal distribution plays a central role in portfolio theory.

The probability density function for the normal distribution is:

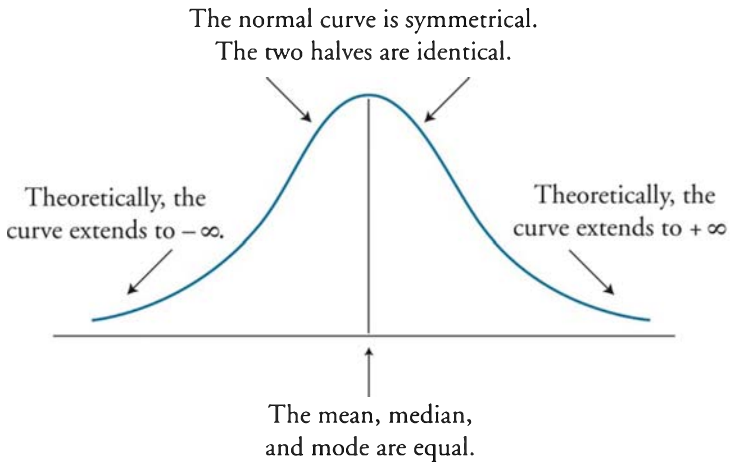
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The normal distribution has the following key properties:

- It is completely described by its mean, μ, and variance, σ^2, stated as $X \sim N(\mu, \sigma^2)$. In words, this says that "*X* is normally distributed with mean μ and variance σ^2."
- Skewness = 0, meaning the normal distribution is symmetric about its mean, so that $P(X \leq \mu) = P(\mu \leq X) = 0.5$, and mean = median = mode.
- Kurtosis = 3; this is a measure of how flat the distribution is. Recall that excess kurtosis is measured relative to 3, the kurtosis of the normal distribution.
- A linear combination of normally distributed independent random variables is also normally distributed.
- The probabilities of outcomes further above and below the mean get smaller and smaller but do not go to zero (the tails get very thin but extend infinitely).

Many of these properties are evident from examining the graph of a normal distribution's probability density function as illustrated in Figure 2.

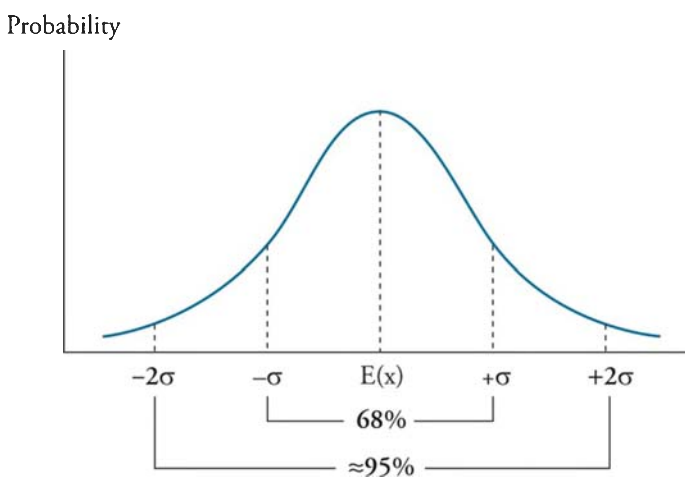
Figure 2: Normal Distribution Probability Density Function



A **confidence interval** is a range of values around the expected outcome within which we expect the actual outcome to be some specified percentage of the time. A 95% confidence interval is a range that we expect the random variable to be in 95% of the time. For a normal distribution, this interval is based on the expected value (sometimes called a point estimate) of the random variable and on its variability, which we measure with standard deviation.

Confidence intervals for a normal distribution are illustrated in Figure 3. For any normally distributed random variable, 68% of the outcomes are within one standard deviation of the expected value (mean), and approximately 95% of the outcomes are within two standard deviations of the expected value.

Figure 3: Confidence Intervals for a Normal Distribution



In practice, we will not know the actual values for the mean and standard deviation of the distribution, but will have estimated them as $\bar{X}$ and s. The three confidence intervals of most interest are given by:

- The 90% confidence interval for X is $\bar{X} - 1.65s$ to $\bar{X} + 1.65s$.
- The 95% confidence interval for X is $\bar{X} - 1.96s$ to $\bar{X} + 1.96s$.
- The 99% confidence interval for X is $\bar{X} - 2.58s$ to $\bar{X} + 2.58s$.

Example: Confidence intervals

The average return of a mutual fund is 10.5% per year and the standard deviation of annual returns is 18%. If returns are approximately normal, what is the 95% confidence interval for the mutual fund return next year?

Answer:

Here μ and σ are 10.5% and 18%, respectively. Thus, the 95% confidence interval for the return, *R*, is:

$$10.5 \pm 1.96(18) = -24.78\% \text{ to } 45.78\%$$

Symbolically, this result can be expressed as:

$$P(-24.78 < R < 45.78) = 0.95 \text{ or } 95\%$$

The interpretation is that the annual return is expected to be within this interval 95% of the time, or 95 out of 100 years.

The Standard Normal Distribution

A standard normal distribution (i.e., *z*-distribution) is a normal distribution that has been standardized so it has a mean of zero and a standard deviation of 1 [i.e., N~(0,1)]. To standardize an observation from a given normal distribution, the *z-value* of the observation must be calculated. The *z*-value represents the number of standard deviations a given observation is from the population mean. *Standardization* is the process of converting an observed value for a random variable to its *z*-value. The following formula is used to standardize a random variable:

$$z = \frac{\text{observation} - \text{population mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

Professor's Note: The term z-value will be used for a standardized observation in this topic. The terms z-score and z-statistic are also commonly used.

Example: Standardizing a random variable (calculating *z*-values)

Assume the annual earnings per share (EPS) for a population of firms are normally distributed with a mean of $6 and a standard deviation of $2.

What are the *z*-values for EPS of $2 and $8?

Answer:

If EPS = x = \$8, then $z = (x - \mu) / \sigma = (\$8 - \$6) / \$2 = +1$

If EPS = x = \$2, then $z = (x - \mu) / \sigma = (\$2 - \$6) / \$2 = -2$

Here, z = +1 indicates that an EPS of $8 is one standard deviation above the mean, and z = –2 means that an EPS of $2 is two standard deviations below the mean.

Calculating Probabilities Using *z*-Values

Now we will show how to use standardized values (*z*-values) and a table of probabilities for *Z* to determine probabilities. A portion of a table of the cumulative distribution function for a standard normal distribution is shown in Figure 4. We will refer to this table as the *z*-table, as it contains values generated using the cumulative density function for a standard normal distribution, denoted by F(Z). Thus, the values in the *z*-table are the probabilities of observing a *z*-value that is less than a given value, *z* [i.e., $P(Z < z)$]. The numbers in the first column are *z*-values that have only one decimal place. The columns to the right supply probabilities for *z*-values with two decimal places.

Note that the *z*-table in Figure 4 only provides probabilities for positive *z*-values. This is not a problem because we know from the symmetry of the standard normal distribution that $F(-Z) = 1 - F(Z)$. The tables in the back of many texts actually provide probabilities for negative *z*-values, but we will work with only the positive portion of the table because this may be all you get on the exam. In Figure 4, we can find the probability that a standard normal random variable will be less than 1.66, for example. The table value is 95.15%. The probability that the random variable will be less than –1.66 is simply 1 – 0.9515 = 0.0485 = 4.85%, which is also the probability that the variable will be greater than +1.66.

Figure 4: Cumulative Probabilities for a Standard Normal Distribution

CDF Values for the Standard Normal Distribution: The z-Table										
z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.5	.6915	Please note that several of the rows have been deleted to save space.*								
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

*A complete cumulative standard normal table is included in the Appendix.

Professor's Note: When you use the standard normal probabilities, you have formulated the problem in terms of standard deviations from the mean. Consider a security with returns that are approximately normal, an expected return of 10%, and standard deviation of returns of 12%. The probability of returns greater than 30% is calculated based on the number of standard deviations that 30% is above the expected return of 10%. 30% is 20% above the expected return of 10%, which is 20 / 12 = 1.67 standard deviations above the mean. We look up the probability of returns less than 1.67 standard deviations above the mean (0.9525 or 95.25% from Figure 4) and calculate the probability of returns more than 1.67 standard deviations above the mean as 1 – 0.9525 = 4.75%.

Example: Using the z-table (1)

Considering again EPS distributed with μ = \$6 and σ = \$2, what is the probability that EPS will be \$9.70 or more?

Answer:

Here we want to know P(EPS > \$9.70), which is the area under the curve to the right of the *z*-value corresponding to EPS = \$9.70 (see the distribution below).

The *z*-value for EPS = \$9.70 is:

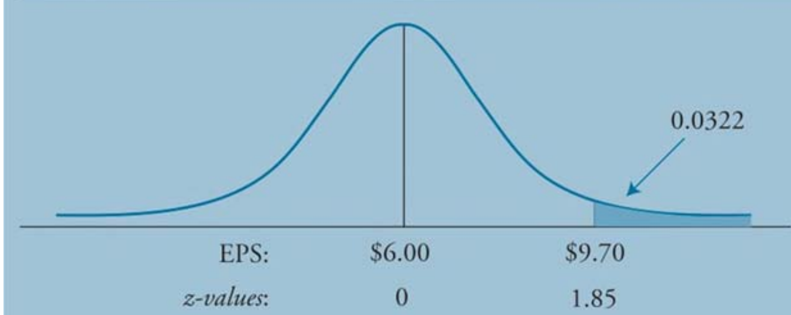
$$z = \frac{(x-\mu)}{\sigma} = \frac{(9.70-6)}{2} = 1.85$$

That is, \$9.70 is 1.85 standard deviations above the mean EPS value of \$6.

From the *z*-table we have F(1.85) = 0.9678, but this is P(EPS ≤ 9.70). We want P(EPS > 9.70), which is 1 – P(EPS ≤ 9.70).

P(EPS > 9.70) = 1 – 0.9678 = 0.0322, or 3.2%

P(EPS>$9.70)



Example: Using the *z*-table (2)

Using the distribution of EPS with μ = $6 and σ = $2 again, what percent of the observed EPS values are likely to be less than $4.10?

Answer:

As shown graphically in the distribution below, we want to know P(EPS < $4.10). This requires a 2-step approach like the one taken in the preceding example.

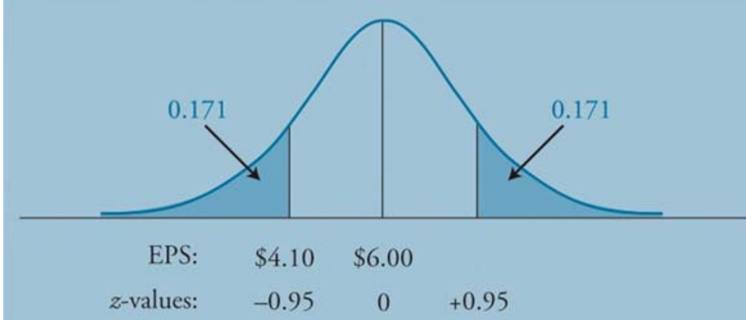
First, the corresponding *z*-value must be determined as follows:

$$z = \frac{(\$4.10 - \$6)}{2} = -0.95,$$

So $4.10 is 0.95 standard deviations below the mean of $6.00.

Now, from the *z*-table for negative values in the back of this book, we find that F(–0.95) = 0.1711, or 17.11%.

Finding a Left-Tail Probability



The *z*-table gives us the probability that the outcome will be more than 0.95 standard deviations below the mean.

The Lognormal Distribution

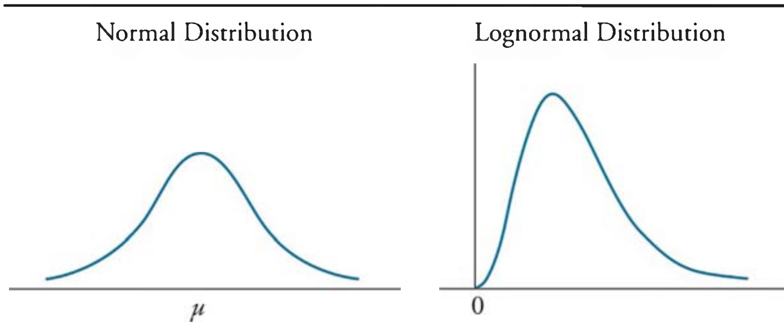
The lognormal distribution is generated by the function e^x, where x is normally distributed. Since the natural logarithm, ln, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

The probability density function for the lognormal distribution is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}$$

Figure 5 illustrates the differences between a normal distribution and a lognormal distribution.

Figure 5: Normal vs. Lognormal Distributions



In Figure 5, we can see that:

- The lognormal distribution is skewed to the right.
- The lognormal distribution is bounded from below by zero so that it is useful for modeling asset prices which never take negative values.

If we used a normal distribution of returns to model asset prices over time, we would admit the possibility of returns less than –100%, which would admit the possibility of asset prices less than zero. Using a lognormal distribution to model *price relatives* avoids this problem. A price relative is just the end-of-period price of the asset divided by the beginning price (S_1/S_0) and is equal to (1 + the holding period return). To get the end-of-period asset price, we can simply multiply the price relative times the beginning-of-period asset price. Since a lognormal distribution takes a minimum value of zero, end-of-period asset prices cannot be less than zero. A price relative of zero corresponds to a holding period return of –100% (i.e., the asset price has gone to zero).

The Central Limit Theorem

LO 17.2: Describe the central limit theorem and the implications it has when combining independent and identically distributed (i.i.d.) random variables.

LO 17.3: Describe i.i.d. random variables and the implications of the i.i.d. assumption when combining random variables.

The central limit theorem states that for simple random samples of size *n* from a *population* with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean $\bar{x}$ approaches a normal probability distribution with mean μ and variance equal to $\frac{\sigma^2}{n}$ as the sample size becomes large. This is possible because, when the sample size is large, the sums of independent and identically distributed (i.i.d.) random variables (the individual items drawn for the sample) will be normally distributed.

The central limit theorem is extremely useful because the normal distribution is relatively easy to apply to hypothesis testing and to the construction of confidence intervals. Specific inferences about the population mean can be made from the sample mean, *regardless of the population's distribution*, as long as the sample size is "sufficiently large," which usually means $n \geq 30$.

Important properties of the central limit theorem include the following:

- If the sample size *n* is sufficiently large ($n \geq 30$), the sampling distribution of the sample means will be approximately normal. Remember what's going on here: random samples of size *n* are repeatedly being taken from an overall larger population. Each of these random samples has its own mean, which is itself a random variable, and this set of sample means has a distribution that is approximately normal.
- The mean of the population, μ, and the mean of the distribution of all possible sample means are equal.
- The variance of the distribution of sample means is $\frac{\sigma^2}{n}$, the population variance divided by the sample size.

Student's *t*-Distribution

Student's *t*-distribution, or simply the *t*-distribution, is a bell-shaped probability distribution that is symmetrical about its mean. It is the appropriate distribution to use when constructing confidence intervals based on *small samples* (n < 30) from populations with *unknown variance* and a normal, or approximately normal, distribution. It may also be appropriate to use the *t*-distribution when the population variance is unknown and the sample size is large enough that the central limit theorem will assure that the sampling distribution is approximately normal.

Student's *t*-distribution has the following properties:

- It is symmetrical.
- It is defined by a single parameter, the degrees of freedom (df), where the degrees of freedom are equal to the number of sample observations minus 1, n – 1, for sample means.
- It has more probability in the tails (fatter tails) than the normal distribution.
- As the degrees of freedom (the sample size) gets larger, the shape of the t-distribution more closely approaches a standard normal distribution.

When *compared to the normal distribution*, the *t*-distribution is flatter with more area under the tails (i.e., it has fatter tails). As the degrees of freedom for the *t*-distribution increase, however, its shape approaches that of the normal distribution.

The degrees of freedom for tests based on sample means are n – 1 because, given the mean, only n – 1 observations can be unique.

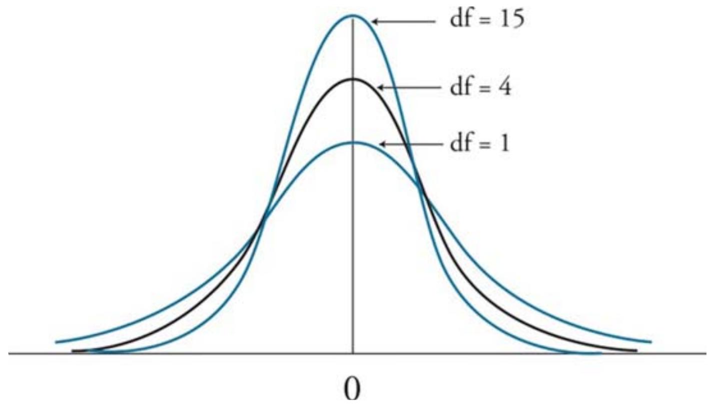
The table in Figure 6 contains one-tailed critical values for the t-distribution at the 0.05 and 0.025 levels of significance with various degrees of freedom (df). Note that, unlike the *z*-table, the *t*-values are contained within the table and the probabilities are located at the column headings. Also note that the level of significance of a *t*-test corresponds to the *one-tailed probabilities*, *p*, that head the columns in the *t*-table.

Figure 6: Table of Critical *t*-Values

	One-Tailed Probabilities, p	
df	*p = 0.05*	*p = 0.025*
5	2.015	2.571
10	1.812	2.228
15	1.753	2.131
20	1.725	2.086
25	1.708	2.060
30	1.697	2.042
40	1.684	2.021
50	1.676	2.009
60	1.671	2.000
70	1.667	1.994
80	1.664	1.990
90	1.662	1.987
100	1.660	1.984
120	1.658	1.980
∞	1.645	1.960

Figure 7 illustrates the different shapes of the *t*-distribution associated with different degrees of freedom. The tendency is for the *t*-distribution to look more and more like the normal distribution as the degrees of freedom increase. Practically speaking, the greater the degrees of freedom, the greater the percentage of observations near the center of the distribution and the lower the percentage of observations in the tails, which are thinner as degrees of freedom increase. This means that confidence intervals for a random variable that follows a *t*-distribution must be wider (narrower) when the degrees of freedom are less (more) for a given significance level.

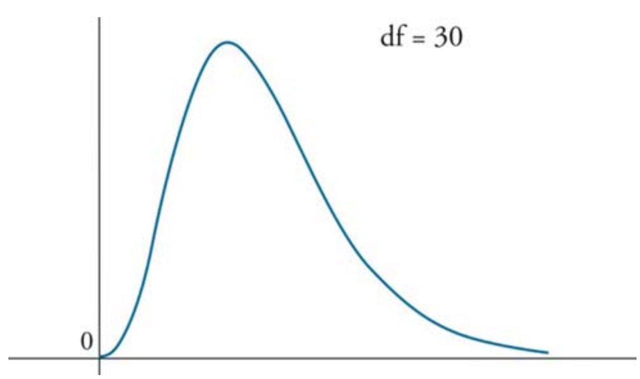
Figure 7: *t*-Distributions for Different Degrees of Freedom (df)



The Chi-Squared Distribution

As you will see in Topic 19, hypothesis testing of the population variance requires the use of a chi-squared distributed test statistic, denoted χ^2. The chi-square distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.

Figure 8: Chi-Squared Distribution



The chi-squared test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:
n = sample size
s^2 = sample variance
σ_0^2 = hypothesized value for the population variance

The chi-squared test compares the test statistic to a critical chi-squared value at a given level of significance to determine whether to reject or fail to reject a null hypothesis. Note that since the chi-squared distribution is bounded below by zero, chi-squared values cannot be negative.

The *F*-Distribution

As you will also see in Topic 19, the hypotheses concerned with the equality of the variances of two populations are tested with an *F*-distributed test statistic. Hypothesis testing using a test statistic that follows an *F*-distribution is referred to as the *F*-test. The *F*-test is used under the assumption that the populations from which samples are drawn are normally distributed and that the samples are independent.

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as:

$$F = \frac{s_1^2}{s_2^2}$$

where:
s_1^2 = variance of the sample of n_1 observations drawn from Population 1
s_2^2 = variance of the sample of n_2 observations drawn from Population 2

An *F*-distribution is presented in Figure 9. As indicated, the *F*-distribution is right-skewed and is truncated at zero on the left-hand side. The shape of the *F*-distribution is determined by *two separate degrees of freedom*, the numerator degrees of freedom, df_1, and the denominator degrees of freedom, df_2.

Note that $n_1 - 1$ and $n_2 - 1$ are the degrees of freedom used to identify the appropriate critical value from the *F*-table (provided in the Appendix).

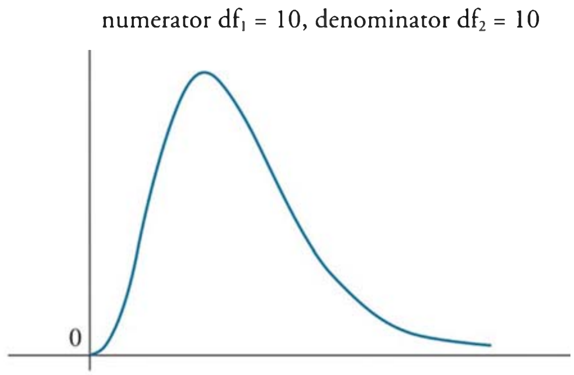
Some additional properties of the *F*-distribution include the following:

- The *F*-distribution approaches the normal distribution as the number of observations increases (just as with the *t*-distribution and chi-squared distribution).
- A random variable's *t*-value squared (t^2) with n – 1 degrees of freedom is *F*-distributed with 1 degree of freedom in the numerator and n – 1 degrees of freedom in the denominator.
- There exists a relationship between the *F*- and chi-squared distributions such that:

$$F = \frac{\chi^2}{\text{\# of observations in numerator}}$$

as the # of observations in denominator $\rightarrow \infty$

Figure 9: *F*-Distribution



Mixture Distributions

LO 17.4: Describe a mixture distribution and explain the creation and characteristics of mixture distributions.

The distributions discussed in this topic, as well as others, can be combined to create unique probability density functions. It may be helpful to create a new distribution if the underlying data you are working with does not currently fit a predetermined distribution. In this case, a newly created distribution may assist with explaining the relevant data.

To illustrate a mixture distribution, suppose that the returns of a stock follow a normal distribution with low volatility 75% of the time and high volatility 25% of the time. Here we have two normal distributions with the same mean, but different risk levels. To create a mixture distribution from these scenarios, we randomly choose either the low or high volatility distribution, placing a 75% probability on selecting the low volatility distribution. We then generate a random return from the selected distribution. By repeating this process several times, we will create a probability distribution that reflects both levels of volatility.

Mixture distributions contain elements of both parametric and nonparametric distributions. The distributions used as inputs (i.e., the component distributions) are parametric, while the weights of each distribution within the mixture are nonparametric. The more component distributions used as inputs, the more closely the mixture distribution will follow the actual data. However, more component distributions will make it difficult to draw conclusions given that the newly created distribution will be very specific to the data.

By mixing distributions, it is easy to see how we can alter skewness and kurtosis of the component distributions. Skewness can be changed by combining distributions with different means, and kurtosis can be changed by combining distributions with different variances. Also, by combining distributions that have significantly different means, we can create a mixture distribution with multiple modes (e.g., a bimodal distribution).

Creating a more robust distribution is clearly beneficial to risk managers. Different levels of skew and/or kurtosis can reveal extreme events that were previously difficult to identify. By creating these mixture distributions, we can improve risk models by incorporating the potential for low-frequency, high-severity events.