Key Concepts

LO 18.1

Bayes' theorem is defined for two random variables *A* and *B* as follows:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

LO 18.2

The primary difference between the Bayesian and frequentist approaches is that the Bayesian approach is based on a prior belief regarding the probability of an event occurring, while the frequentist approach is based on a number or frequency of events occurring during the most recent sample.

LO 18.3

Bayes' theorem can be extended to include more than two possible outcomes. Given the numerous calculations involved when incorporating multiple states, it is helpful to solve these types of problems using spreadsheet software.

Concept Checkers

Use the following information to answer Questions 1 through 3

Suppose a manager for a fund of funds uses historical data to categorize managers as excellent or average. Based on historical performance, the probabilities of excellent and average managers outperforming the market are 80% and 50%, respectively. Assume that the probabilities for managers outperforming the market is independent of their performance in prior years. In addition, the fund of funds manager believes that only 15% of total fund managers are excellent managers. Assume that a new manager started three years ago and beat the market in each of the past three years.

1. Using the Bayesian approach, what is the approximate probability that the new manager is an excellent manager today?
   A. 18.3%.
   B. 27.5%.
   C. 32.1%.
   D. 42.0%.

2. What is the approximate probability that the new manager will outperform the market next year using the Bayesian approach?
   A. 31.9%.
   B. 51.2%.
   C. 62.6%.
   D. 80.0%.

3. What is the probability that the new manager will outperform the market next year using the frequentist approach?
   A. 41.9%.
   B. 51.2%.
   C. 80.0%.
   D. 100.0%.

Use the following information to answer Questions 4 and 5

Suppose a pension fund gathers information on portfolio managers to rank their abilities as excellent, average, or below average. The analyst for the pension fund forms prior beliefs regarding the probabilities of a manager outperforming the market based on historical performances of all managers. There is a 10% probability that a manager is excellent, a 60% probability that a manager is average, and a 30% probability that a manager is below average. In addition, the probabilities of a manager outperforming the market are 75% for an excellent manager, 50% for an average manager, and 25% for a below average manager. Assume the probability of the manager outperforming the market is independent of the prior year performance.

4. What is the probability of a new manager outperforming the market two years in a row?
   A. 18.50%.
   B. 22.50%.
   C. 37.25%.
   D. 56.25%.

5. Suppose a new manager just outperformed the market two years in a row. Using Bayesian analysis, what is the updated belief or probability that the new manager is excellent?
   A. 20.0%.
   B. 22.5%.
   C. 25.0%.
   D. 27.5%.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

Concept Checker Answers

1. **D** Excellent managers are expected to outperform the market 80% of the time. The probability of an excellent manager outperforming three years in a row is 0.8^3 or 51.2%. Similarly, the probability of an average manager outperforming the market three years in a row is determined by taking the 50% probability to the third power: $0.5^3 = 0.125$.

   The probability that the new manager is excellent after beating the market three years in a row is determined by the following Bayesian approach:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)}$$

   The denominator is the unconditional probability of outperforming the market for three years in a row. This is computed as a weighted average of the probabilities of outperforming three years in a row for each type of manager.

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.512 \times 0.15) + (0.125 \times 0.85) \\ &= 0.0768 + 0.10625 \\ &= 0.18305 \end{aligned}$$

   With this information, we can now apply the Bayesian approach as follows:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.512 \times 0.15}{0.18305} = 41.956\%$$

2. **C** The probability of the new manager outperforming the market next year is the unconditional probability of outperforming the market based on the new probability that the new manager is an excellent manager after outperforming the market three years in a row. From Question 1, we determined the probability that the new manager is excellent after beating the market three years in a row as:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.512 \times 0.15}{0.18305} = 41.956\%$$

   The probability that the new manager is average after beating the market three years in a row is determined as:

$$P(A \mid O) = \frac{P(O \mid A) \times P(A)}{P(O)} = \frac{0.125 \times 0.85}{0.18305} = 58.044\%$$

   Next, these new probabilities are now used to determine the unconditional probability of outperforming the market next year.

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.8 \times 0.41956) + (0.5 \times 0.58044) \\ &= 0.3356 + 0.2902 \\ &= 0.6258 \text{ or } 62.58\% \end{aligned}$$

3. **D** The frequentist approach determines the probability based on the outperformance for the most recent sample size. In this example, there are only three years of data and the new manager outperformed the market every year. Thus, there is a 100% probability under this approach (3 out of 3) that the new manager will outperform the market next year.

4. **B** To answer this question, you need to determine the unconditional probability of outperforming the market two years in a row. The first step is to calculate the probability of each type of manager outperforming the market two years in a row.

   The probability that an excellent manager outperforms the market two years in a row is:

   $P(O \mid p = 0.75) = 0.75^2 = 0.5625$

   The probability that an average manager outperforms the market two years in a row is:

   $P(O \mid p = 0.5) = 0.5^2 = 0.25$

   The probability that a below average manager outperforms the market two years in a row is:

   $P(O \mid p = 0.25) = 0.25^2 = 0.0625$

   Next, calculate the unconditional probability that a new manager outperforms the market two years in a row based on prior expectations or beliefs:

   $P(O) = (10\% \times 56.25\%) + (60\% \times 25\%) + (30\% \times 6.25\%) = 0.05625 + 0.15 + 0.01875 = 0.225$ or 22.5%

5. **C** From Question 4, we know the unconditional probability that a new manager outperforms the market two years in a row based on prior expectations or beliefs is:

   $P(O) = (10\% \times 56.25\%) + (60\% \times 25\%) + (30\% \times 6.25\%) = 0.05625 + 0.15 + 0.01875 = 0.225$ or 22.5%

   With this information, we can apply Bayes' theorem to update our beliefs that the manager is excellent:

$$P(p = 0.75 \mid O) = \frac{P(O \mid p = 0.75) \times P(p = 0.75)}{P(O)} = \frac{0.5625 \times 0.1}{0.225} = 25\%$$

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:

Hypothesis Testing and Confidence Intervals

Topic 19

Exam Focus

This topic provides insight into how risk managers make portfolio decisions on the basis of statistical analysis of samples of investment returns or other random economic and financial variables. We first focus on the estimation of sample statistics and the construction of confidence intervals for population parameters based on sample statistics. We then discuss hypothesis testing procedures used to conduct tests concerned with population means and population variances. Specific tests reviewed include the *z*-test and the *t*-test. For the exam, you should be able to construct and interpret a confidence interval and know when and how to apply each of the test statistics discussed when conducting hypothesis testing.

Applied Statistics

In many real-world statistics applications, it is impractical (or impossible) to study an entire population. When this is the case, a subgroup of the population, called a sample, can be evaluated. Based upon this sample, the parameters of the underlying population can be estimated.

For example, rather than attempting to measure the performance of the U.S. stock market by observing the performance of all 10,000 or so stocks trading in the United States at any one time, the performance of the subgroup of 500 stocks in the S&P 500 can be measured. The results of the statistical analysis of this sample can then be used to draw conclusions about the entire population of U.S. stocks.

Simple random sampling is a method of selecting a sample in such a way that each item or person in the population being studied has the same likelihood of being included in the sample. As an example of simple random sampling, assume you want to draw a sample of five items out of a group of 50 items. This can be accomplished by numbering each of the 50 items, placing them in a hat, and shaking the hat. Next, one number can be drawn randomly from the hat. Repeating this process (experiment) four more times results in a set of five numbers. The five drawn numbers (items) comprise a simple random sample from the population. In applications like this one, a random-number table or a computer random-number generator is often used to create the sample. Another way to form an approximately random sample is systematic sampling, selecting every *n*th member from a population.

Sampling error is the difference between a sample statistic (the mean, variance, or standard deviation of the sample) and its corresponding population parameter (the true mean, variance, or standard deviation of the population). For example, the sampling error for the mean is as follows:

$$\text{sampling error of the mean} = \text{sample mean} - \text{population mean} = \bar{x} - \mu$$

Mean and Variance of the Sample Average

It is important to recognize that the sample statistic itself is a random variable and, therefore, has a probability distribution. The **sampling distribution** of the sample statistic is a probability distribution of all possible sample statistics computed from a set of equal-size samples that were randomly drawn from the same population. Think of it as the probability distribution of a statistic from many samples.

For example, suppose a random sample of 100 bonds is selected from a population of a major municipal bond index consisting of 1,000 bonds, and then the mean return of the 100-bond sample is calculated. Repeating this process many times will result in many different estimates of the population mean return (i.e., one for each sample). The distribution of these estimates of the mean is the *sampling distribution of the mean*. It is important to note that this sampling distribution is distinct from the distribution of the actual prices of the 1,000 bonds in the underlying population and will have different parameters.

To illustrate the mean of the sample average, suppose we have selected two independent and identically distributed (i.i.d.) variables at random, X_1 and X_2, from a population. Since these two variables are i.i.d., the mean and variance for both observations will be the same, respectively.

Recall from Topic 16, the mean of the sum of two random variables is equal to:

$$E(X_1 + X_2) = \mu_X + \mu_X = 2\mu_X$$

Thus, the mean of the sample average, $E(\overline{X})$, will be equal to:

$$E\left(\frac{X_1 + X_2}{2}\right) = \frac{2\mu_X}{2} = \mu_X$$

More generally, we can say that for *n* observations:

$$E(\overline{X}) = \mu_X$$

By applying the properties of variance for the sums of independent random variables, we can also calculate the variance of the sample average. Recall, that for independent variables, the covariance term in the variance equation will equal zero. For two observations, the variance of the sum of two random variables will equal:

$$\text{Var}(X_1 + X_2) = 2\sigma_x^2$$

Thus, when applying the following variance property:

$$\text{Var}(aX_1 + cX_2) = a^2 \times \text{Var}(X_1) + c^2 \times \text{Var}(X_2)$$

and assuming *a* and *c* are equal to 0.5, the variance of the sample average, $\text{Var}(\overline{X})$, will be equal to $\frac{\sigma_X^2}{2}$. In more general terms, $\text{Var}(\overline{X}) = \frac{\sigma_X^2}{n}$ for *n* observations, and the standard deviation of the sample average is equal to $\frac{\sigma_X}{\sqrt{n}}$. This standard deviation measure is known as the **standard error**.

These properties help define the distributional characteristics of the sample distribution of the mean and allow us to make assumptions about the distribution when the sample size is large.

LO 19.1: Calculate and interpret the sample mean and sample variance.

Population and Sample Mean

Recall from Topic 16, that in order 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*.

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

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.

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

Population and Sample Variance

Dispersion is defined as the *variability around the central tendency*. The common theme in finance and investments is the tradeoff between reward and variability, where the central tendency is the measure of the reward and dispersion is a measure of risk.

The **population variance** is defined as the average of the squared deviations from the mean. The population variance (σ^2) uses the values for all members of a population and is calculated using the following formula:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Example: Population variance, σ^2

Assume the following 5-year annualized total returns represent all of the managers at a small investment firm (30%, 12%, 25%, 20%, 23%). What is the population variance of these returns?

Answer:

$$\mu = \frac{[30+12+25+20+23]}{5} = 22\%$$

$$\sigma^2 = \frac{\left[(30-22)^2 + (12-22)^2 + (25-22)^2 + (20-22)^2 + (23-22)^2\right]}{5} = 35.60\left(\%^2\right)$$

Interpreting this result, we can say that the average variation from the mean return is 35.60% squared. Had we done the calculation using decimals instead of whole percents, the variance would be 0.00356.

A major problem with using the variance is the difficulty of interpreting it. The computed variance, unlike the mean, is in terms of squared units of measurement. How does one interpret squared percents, squared dollars, or squared yen? This problem is mitigated through the use of the *standard deviation*. The population standard deviation, σ, is the square root of the population variance and is calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X - \mu)^2}{N}}$$

Example: Population standard deviation, σ

Using the data from the preceding example, **compute** the population standard deviation.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

Answer:

$$\sigma = \sqrt{\frac{(30-22)^2 + (12-22)^2 + (25-22)^2 + (20-22)^2 + (23-22)^2}{5}}$$

$$= \sqrt{35.60} = 5.97\%$$

Calculated with decimals instead of whole percents, we would get:

$$\sigma^2 = 0.00356 \text{ and } \sigma = \sqrt{0.00356} = 0.05966 = 5.97\%$$

Since the population standard deviation and population mean are both expressed in the same units (percent), these values are easy to relate. The outcome of this example indicates that the mean return is 22% and the standard deviation about the mean is 5.97%.

The **sample variance**, s^2, is the measure of dispersion that applies when we are evaluating a sample of n observations from a population. The sample variance is calculated using the following formula:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

The most noteworthy difference from the formula for population variance is that the denominator for s^2 is $n - 1$, one less than the sample size n, where σ^2 uses the entire population size N. Another difference is the use of the sample mean, $\overline{X}$, instead of the population mean, μ. Based on the mathematical theory behind statistical procedures, the use of the entire number of sample observations, n, instead of $n - 1$ as the divisor in the computation of s^2, will systematically underestimate the population parameter, σ^2, particularly for small sample sizes. This systematic underestimation causes the sample variance to be what is referred to as a biased estimator of the population variance. Using $n - 1$ instead of n in the denominator, however, improves the statistical properties of s^2 as an estimator of σ^2. Thus, s^2, as expressed in the equation above, is considered to be an unbiased estimator of σ^2.

Example: Sample variance

Assume that the 5-year annualized total returns for the five investment managers used in the preceding examples represent only a sample of the managers at a large investment firm. What is the sample variance of these returns?

Answer:

$$\overline{X} = \frac{[30+12+25+20+23]}{5} = 22\%$$

$$s^2 = \frac{\left[(30-22)^2+(12-22)^2+(25-22)^2+(20-22)^2+(23-22)^2\right]}{5-1} = 44.5\left(\%^2\right)$$

Thus, the sample variance of $44.5(\%^2)$ can be interpreted to be an unbiased estimator of the population variance. Note that 44.5 "percent squared" is 0.00445 and you will get this value if you put the percent returns in decimal form [e.g., $(0.30 - 0.22)^2$, and so forth.].

As with the population standard deviation, the sample standard deviation can be calculated by taking the square root of the sample variance. The sample standard deviation, *s*, is defined as:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Example: Sample standard deviation

Compute the sample standard deviation based on the result of the preceding example.

Answer:

Since the sample variance for the preceding example was computed to be $44.5(\%^2)$, the sample standard deviation is:

$$s = [44.5(\%^2)]^{1/2} = 6.67\% \text{ or } \sqrt{0.00445} = 0.0667$$

The results shown here mean that the sample standard deviation, $s = 6.67\%$, can be interpreted as an unbiased estimator of the population standard deviation, σ.

The **standard error** of the sample mean is the standard deviation of the distribution of the sample means.

When the standard deviation of the population, σ, is *known*, the standard error of the sample mean is calculated as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where:
$\sigma_{\bar{x}}$ = standard error of the sample mean
σ = standard deviation of the population
n = size of the sample

Example: Standard error of sample mean (known population variance)

The mean hourly wage for Iowa farm workers is $13.50 with a *population standard deviation* of $2.90. **Calculate** and **interpret** the standard error of the sample mean for a sample size of 30.

Answer:

Because the population standard deviation, σ, *is known*, the standard error of the sample mean is expressed as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$2.90}{\sqrt{30}} = \$0.53$$

Professor's Note: On the TI BAII Plus, the use of the square root key is obvious. On the HP 12C, the square root of 30 is computed as: [30] [g] [$\sqrt{x}$].

This means that if we were to take all possible samples of size 30 from the Iowa farm worker population and prepare a sampling distribution of the sample means, we would get a distribution with a mean of $13.50 and standard error of $0.53.

Practically speaking, the *population's standard deviation is almost never known*. Instead, the standard error of the sample mean must be estimated by dividing the standard deviation of the *sample* mean by $\sqrt{n}$:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Example: Standard error of sample mean (unknown population variance)

Suppose a sample contains the past 30 monthly returns for McCreary, Inc. The mean return is 2% and the *sample* standard deviation is 20%. **Calculate** and **interpret** the standard error of the sample mean.

Answer:

Since σ is unknown, the standard error of the sample mean is:

$$s_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{30}} = 3.6\%$$

This implies that if we took all possible samples of size 30 from McCreary's monthly returns and prepared a sampling distribution of the sample means, the mean would be 2% with a standard error of 3.6%.

Example: Standard error of sample mean (unknown population variance)

Continuing with our example, suppose that instead of a sample size of 30, we take a sample of the past 200 monthly returns for McCreary, Inc. In order to highlight the effect of sample size on the sample standard error, let's assume that the mean return and standard deviation of this larger sample remain at 2% and 20%, respectively. Now, **calculate** the standard error of the sample mean for the 200-return sample.

Answer:

The standard error of the sample mean is computed as:

$$s_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{200}} = 1.4\%$$

The result of the preceding two examples illustrates an important property of sampling distributions. Notice that the value of the standard error of the sample mean decreased from 3.6% to 1.4% as the sample size increased from 30 to 200. This is because as the sample size increases, the sample mean gets closer, on average, to the true mean of the population. In other words, the distribution of the sample means about the population mean gets smaller and smaller, so the standard error of the sample mean decreases.

Population and Sample Covariance

The covariance between two random variables is a statistical measure of the degree to which the two variables move together. The covariance captures the linear relationship between one variable and another. A positive covariance indicates that the variables tend to move together; a negative covariance indicates that the variables tend to move in opposite directions.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

The population and sample covariances are calculated as:

$$\text{population cov}_{XY} = \frac{\sum_{i=1}^{N}(X_i - \mu_X)(Y_i - \mu_Y)}{N}$$

$$\text{sample cov}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

The actual value of the covariance is not very meaningful because its measurement is extremely sensitive to the scale of the two variables. Also, the covariance may range from negative to positive infinity and it is presented in terms of squared units (e.g., percent squared). For these reasons, we take the additional step of calculating the correlation coefficient (see Topic 16), which converts the covariance into a measure that is easier to interpret.

Confidence Intervals

LO 19.2: Construct and interpret a confidence interval.

Confidence interval estimates result in a range of values within which the actual value of a parameter will lie, given the probability of 1 – α. Here, alpha, α, is called the *level of significance* for the confidence interval, and the probability 1 – α is referred to as the *degree of confidence*. For example, we might estimate that the population mean of random variables will range from 15 to 25 with a 95% degree of confidence, or at the 5% level of significance.

Confidence intervals are usually constructed by adding or subtracting an appropriate value from the point estimate. In general, confidence intervals take on the following form:

point estimate ± (reliability factor × standard error)

where:
point estimate = value of a sample statistic of the population parameter
reliability factor = number that depends on the sampling distribution of the point estimate and the probability that the point estimate falls in the confidence interval, (1 – α)
standard error = standard error of the point estimate

If the population has a *normal distribution with a known variance*, a confidence interval for the population mean can be calculated as:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where:

$\bar{x}$ = point estimate of the population mean (sample mean)

$z_{\alpha/2}$ = reliability factor, a standard normal random variable for which the probability in the right-hand tail of the distribution is $\alpha/2$. In other words, this is the *z*-score that leaves $\alpha/2$ of probability in the upper tail.

$\frac{\sigma}{\sqrt{n}}$ = the standard error of the sample mean where σ is the known standard deviation of the population, and *n* is the sample size

The most commonly used standard normal distribution reliability factors are:

$z_{\alpha/2}$ = 1.65 for 90% confidence intervals (the significance level is 10%, 5% in each tail).

$z_{\alpha/2}$ = 1.96 for 95% confidence intervals (the significance level is 5%, 2.5% in each tail).

$z_{\alpha/2}$ = 2.58 for 99% confidence intervals (the significance level is 1%, 0.5% in each tail).

Do these numbers look familiar? They should! In Topic 17, we found the probability under the standard normal curve between z = –1.96 and z = +1.96 to be 0.95, or 95%. Owing to symmetry, this leaves a probability of 0.025 under each tail of the curve beyond z = –1.96 or z = +1.96, for a total of 0.05, or 5%—just what we need for a significance level of 0.05, or 5%.

Example: Confidence interval

Consider a practice exam that was administered to 36 FRM Part I candidates. The mean score on this practice exam was 80. Assuming a population standard deviation equal to 15, **construct** and **interpret** a 99% confidence interval for the mean score on the practice exam for 36 candidates. *Note that in this example the population standard deviation is known, so we don't have to estimate it.*

Answer:

At a confidence level of 99%, $z_{\alpha/2} = z_{0.005} = 2.58$. So, the 99% confidence interval is calculated as follows:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 80 \pm 2.58 \frac{15}{\sqrt{36}} = 80 \pm 6.45$$

Thus, the 99% confidence interval ranges from 73.55 to 86.45.

Confidence intervals can be interpreted from a probabilistic perspective or a practical perspective. With regard to the outcome of the practice exam example, these two perspectives can be described as follows:

- *Probabilistic interpretation*. After repeatedly taking samples of exam candidates, administering the practice exam, and constructing confidence intervals for each sample's mean, 99% of the resulting confidence intervals will, in the long run, include the population mean.
- *Practical interpretation*. We are 99% confident that the population mean score is between 73.55 and 86.45 for candidates from this population.

Confidence Intervals for a Population Mean: Normal With Unknown Variance

If the distribution of the *population is normal with unknown variance*, we can use the *t*-distribution to construct a confidence interval:

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where:

$\bar{x}$ = the point estimate of the population mean

$t_{\alpha/2}$ = the *t*-reliability factor (i.e., *t*-statistic or critical *t*-value) corresponding to a *t*-distributed random variable with n – 1 degrees of freedom, where *n* is the sample size. The area under the tail of the *t*-distribution to the right of $t_{\alpha/2}$ is $\alpha/2$.

$\frac{s}{\sqrt{n}}$ = standard error of the sample mean

s = sample standard deviation

Unlike the standard normal distribution, the reliability factors for the *t*-distribution depend on the sample size, so we can't rely on a commonly used set of reliability factors. Instead, reliability factors for the *t*-distribution have to be looked up in a table of Student's *t*-distribution, like the one at the back of this book.

Owing to the relatively fatter tails of the *t*-distribution, confidence intervals constructed using *t*-reliability factors ($t_{\alpha/2}$) will be more conservative (wider) than those constructed using *z*-reliability factors ($z_{\alpha/2}$).

Example: Confidence intervals

Let's return to the McCreary, Inc. example. Recall that we took a sample of the past 30 monthly stock returns for McCreary, Inc. and determined that the mean return was 2% and the sample standard deviation was 20%. Since the population variance is unknown, the standard error of the sample was estimated to be:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{30}} = 3.6\%$$

Now, let's construct a 95% confidence interval for the mean monthly return.

Answer:

Here, we will use the *t*-reliability factor because the population variance is unknown. Since there are 30 observations, the degrees of freedom are 29 = 30 – 1. Remember, because this is a two-tailed test at the 95% confidence level, the probability under each tail must be $\alpha/2$ = 2.5%, for a total of 5%. So, referencing the one-tailed probabilities for Student's *t*-distribution at the back of this book, we find the critical *t*-value (reliability factor) for $\alpha/2$ = 0.025 and df = 29 to be $t_{29,\,2.5}$ = 2.045. Thus, the 95% confidence interval for the population mean is:

$$2\% \pm 2.045\left(\frac{20\%}{\sqrt{30}}\right) = 2\% \pm 2.045(3.6\%) = 2\% \pm 7.4\%$$

Thus, the 95% confidence has a lower limit of –5.4% and an upper limit of +9.4%.

We can interpret this confidence interval by saying we are 95% confident that the population mean monthly return for McCreary stock is between –5.4% and +9.4%.

Professor's Note: You should practice looking up reliability factors (i.e., critical t-values or t-statistics) in a t-table. The first step is always to compute the degrees of freedom, which is n – 1. The second step is to find the appropriate level of alpha or significance. This depends on whether the test you're concerned with is one-tailed (use α) or two-tailed (use α/2). To look up $t_{29,\,2.5}$, *find the 29 df row and match it with the 0.025 column; t = 2.045 is the result. We'll do more of this in our study of hypothesis testing.*

Confidence Interval for a Population Mean: Nonnormal With Unknown Variance

We now know that the *z*-statistic should be used to construct confidence intervals when the population distribution is normal and the variance is known, and the *t*-statistic should be used when the distribution is normal but the variance is unknown. But what do we do when the distribution is *nonnormal*?

As it turns out, the size of the sample influences whether or not we can construct the appropriate confidence interval for the sample mean.

- If the *distribution is nonnormal* but the *population variance is known*, the *z*-statistic can be used as long as the sample size is large ($n \geq 30$). We can do this because the central limit theorem assures us that the distribution of the sample mean is approximately normal when the sample is large.
- If the *distribution is nonnormal* and the *population variance is unknown*, the *t*-statistic can be used as long as the sample size is large ($n \geq 30$). It is also acceptable to use the *z*-statistic, although use of the *t*-statistic is more conservative.

This means that if we are sampling from a nonnormal distribution (which is sometimes the case in finance), *we cannot create a confidence interval if the sample size is less than 30*. So, all else equal, make sure you have a sample of at least 30, and the larger, the better.

Figure 1: Criteria for Selecting the Appropriate Test Statistic

When sampling from a:	*Test Statistic*	
	Small Sample *(n < 30)*	*Large Sample* *(n ≥ 30)*
Normal distribution with *known* variance	*z*-statistic	*z*-statistic
Normal distribution with *unknown* variance	*t*-statistic	*t*-statistic*
Nonnormal distribution with *known* variance	not available	*z*-statistic
Nonnormal distribution with *unknown* variance	not available	*t*-statistic*

* The *z*-statistic is theoretically acceptable here, but use of the *t*-statistic is more conservative.

All of the preceding analysis depends on the sample we draw from the population being random. If the sample isn't random, the central limit theorem doesn't apply, our estimates won't have the desirable properties, and we can't form unbiased confidence intervals. Surprisingly, creating a *random sample* is not as easy as one might believe. There are a number of potential mistakes in sampling methods that can bias the results. These biases are particularly problematic in financial research, where available historical data are plentiful, but the creation of new sample data by experimentation is restricted.

Hypothesis Testing

LO 19.3: Construct an appropriate null and alternative hypothesis, and calculate an appropriate test statistic.

Hypothesis testing is the statistical assessment of a statement or idea regarding a population. For instance, a statement could be, "The mean return for the U.S. equity market is greater than zero." Given the relevant returns data, hypothesis testing procedures can be employed to test the validity of this statement at a given significance level.

A **hypothesis** is a statement about the value of a population parameter developed for the purpose of testing a theory or belief. Hypotheses are stated in terms of the population parameter to be tested, like the population mean, μ. For example, a researcher may be interested in the mean daily return on stock options. Hence, the hypothesis may be that the mean daily return on a portfolio of stock options is positive.

Hypothesis testing procedures, based on sample statistics and probability theory, are used to determine whether a hypothesis is a reasonable statement and should not be rejected or if it is an unreasonable statement and should be rejected. The process of hypothesis testing consists of a series of steps shown in Figure 2.

Figure 2: Hypothesis Testing Procedure*

State the hypothesis
↓
Select the appropriate test statistic
↓
Specify the level of significance
↓
State the decision rule regarding the hypothesis
↓
Collect the sample and calculate the sample statistics
↓
Make a decision regarding the hypothesis
↓
Make a decision based on the results of the test

* (*Source*: Wayne W. Daniel and James C. Terrell, *Business Statistics, Basic Concepts and Methodology*, Houghton Mifflin, Boston, 1997.)

The Null Hypothesis and Alternative Hypothesis

The **null hypothesis**, designated H_0, is the hypothesis the researcher wants to reject. It is the hypothesis that is actually tested and is the basis for the selection of the test statistics. The null is generally a simple statement about a population parameter. Typical statements of the null hypothesis for the population mean include H_0: $\mu = \mu_0$, H_0: $\mu \leq \mu_0$, and H_0: $\mu \geq \mu_0$, where μ is the population mean and μ_0 is the hypothesized value of the population mean.

Professor's Note: The null hypothesis always includes the "equal to" condition.

The **alternative hypothesis**, designated H_A, is what is concluded if there is sufficient evidence to reject the null hypothesis. It is usually the alternative hypothesis the researcher is really trying to assess. Why? Since you can never really prove anything with statistics, when the null hypothesis is discredited, the implication is that the alternative hypothesis is valid.

The Choice of the Null and Alternative Hypotheses

The most common null hypothesis will be an "equal to" hypothesis. The alternative is often the hoped-for hypothesis. When the null is that a coefficient is equal to zero, we hope to reject it and show the significance of the relationship.

When the null is less than or equal to, the (mutually exclusive) alternative is framed as greater than. If we are trying to demonstrate that a return is greater than the risk-free rate, this would be the correct formulation. We will have set up the null and alternative hypothesis so rejection of the null will lead to acceptance of the alternative, our goal in performing the test.

Hypothesis testing involves two statistics: the *test statistic* calculated from the sample data and the *critical value* of the test statistic. The value of the computed test statistic relative to the critical value is a key step in assessing the validity of a hypothesis.

A test statistic is calculated by comparing the point estimate of the population parameter with the hypothesized value of the parameter (i.e., the value specified in the null hypothesis). With reference to our option return example, this means we are concerned with the difference between the mean return of the sample and the hypothesized mean return. As indicated in the following expression, the test statistic is the difference between the sample statistic and the hypothesized value, scaled by the standard error of the sample statistic.

$$\text{test statistic} = \frac{\text{sample statistic} - \text{hypothesized value}}{\text{standard error of the sample statistic}}$$

The standard error of the sample statistic is the adjusted standard deviation of the sample. When the sample statistic is the sample mean, $\bar{x}$, the standard error of the sample statistic for sample size *n*, is calculated as:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

when the population standard deviation, σ, *is known*, or

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

when the population standard deviation, σ, *is not known*. In this case, it is estimated using the standard deviation of the sample, *s*.

Professor's Note: Don't be confused by the notation here. A lot of the literature you will encounter in your studies simply uses the term $\sigma_{\bar{x}}$ *for the standard error of the test statistic, regardless of whether the population standard deviation or sample standard deviation was used in its computation.*

As you will soon see, a test statistic is a random variable that may follow one of several distributions, depending on the characteristics of the sample and the population. We will look at four distributions for test statistics: the *t*-distribution, the *z*-distribution (standard normal distribution), the chi-squared distribution, and the *F*-distribution. The critical value for the appropriate test statistic—the value against which the computed test statistic is compared—depends on its distribution.

One-Tailed and Two-Tailed Tests of Hypotheses

LO 19.4: Differentiate between a one-tailed and a two-tailed test and identify when to use each test.

The alternative hypothesis can be one-sided or two-sided. A one-sided test is referred to as a **one-tailed test**, and a two-sided test is referred to as a **two-tailed test**. Whether the test is one- or two-sided depends on the proposition being tested. If a researcher wants to test whether the return on stock options is greater than zero, a one-tailed test should be used. However, a two-tailed test should be used if the research question is whether the return on options is simply different from zero. Two-sided tests allow for deviation on both sides of the hypothesized value (zero). In practice, most hypothesis tests are constructed as two-tailed tests.

A two-tailed test for the population mean may be structured as:

$$H_0: \mu = \mu_0 \text{ versus } H_A: \mu \neq \mu_0$$

Since the alternative hypothesis allows for values above and below the hypothesized parameter, a two-tailed test uses two critical values (or rejection points).

The *general decision rule for a two-tailed test* is:

Reject H_0 if: test statistic > upper critical value or
test statistic < lower critical value

Let's look at the development of the decision rule for a two-tailed test using a *z*-distributed test statistic (a *z*-test) at a 5% level of significance, $\alpha = 0.05$.

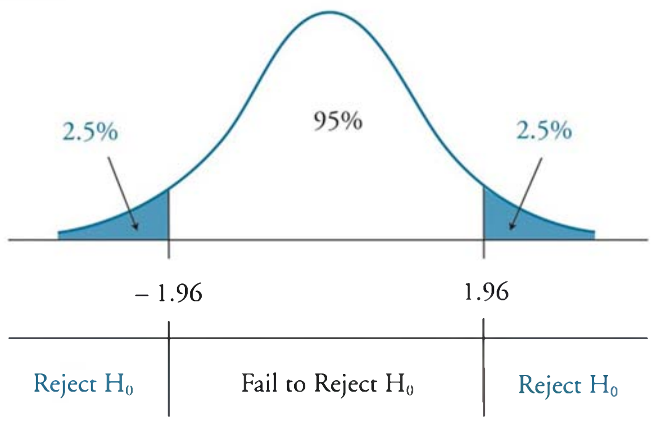
- At $\alpha = 0.05$, the computed test statistic is compared with the critical z-values of ±1.96. The values of ±1.96 correspond to $\pm z_{\alpha/2} = \pm z_{0.025}$, which is the range of *z*-values within which 95% of the probability lies. These values are obtained from the cumulative probability table for the standard normal distribution (*z*-table), which is included at the back of this book.
- If the computed test statistic falls outside the range of critical *z*-values (i.e., test statistic > 1.96, or test statistic < –1.96), we reject the null and conclude that the sample statistic is sufficiently different from the hypothesized value.
- If the computed test statistic falls within the range ±1.96, we conclude that the sample statistic is not sufficiently different from the hypothesized value ($\mu = \mu_0$ in this case), and we fail to reject the null hypothesis.

The *decision rule* (rejection rule) *for a two-tailed z-test* at $\alpha = 0.05$ can be stated as:

Reject H_0 if: test statistic < –1.96 or
test statistic > 1.96

Figure 3 shows the standard normal distribution for a two-tailed hypothesis test using the *z*-distribution. Notice that the significance level of 0.05 means that there is 0.05 / 2 = 0.025 probability (area) under each tail of the distribution beyond ±1.96.

Figure 3: Two-Tailed Hypothesis Test Using the Standard Normal (*z*) Distribution



Example: Two-tailed test

A researcher has gathered data on the daily returns on a portfolio of call options over a recent 250-day period. The mean daily return has been 0.1%, and the sample standard deviation of daily portfolio returns is 0.25%. The researcher believes the mean daily portfolio return is not equal to zero. **Construct** a hypothesis test of the researcher's belief.

Answer:

First, we need to specify the null and alternative hypotheses. The null hypothesis is the one the researcher expects to reject.

H_0: $\mu_0 = 0$ versus H_A: $\mu_0 \neq 0$

Since the null hypothesis is an equality, this is a two-tailed test. At a 5% level of significance, the critical *z*-values for a two-tailed test are ±1.96, so the decision rule can be stated as:

Reject H_0 if: test statistic < –1.96 or test statistic > +1.96

The *standard error* of the sample mean is the adjusted standard deviation of the sample. When the sample statistic is the sample mean, *x*, the standard error of the sample statistic for sample size *n* is calculated as:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Since our sample statistic here is a sample mean, the standard error of the sample mean for a sample size of 250 is $\frac{0.0025}{\sqrt{250}}$ and our test statistic is:

$$\frac{0.001}{\left(\frac{0.0025}{\sqrt{250}}\right)} = \frac{0.001}{0.000158} = 6.33$$

Since 6.33 > 1.96, we reject the null hypothesis that the mean daily option return is equal to zero. Note that when we reject the null, we conclude that the sample value is significantly different from the hypothesized value. We are saying that the two values are different from one another *after considering the variation in the sample*. That is, the mean daily return of 0.001 is statistically different from zero given the sample's standard deviation and size.

For a one-tailed hypothesis test of the population mean, the null and alternative hypotheses are either:

Upper tail: H_0: $\mu \leq \mu_0$ versus H_A: $\mu > \mu_0$, or
Lower tail: H_0: $\mu \geq \mu_0$ versus H_A: $\mu < \mu_0$

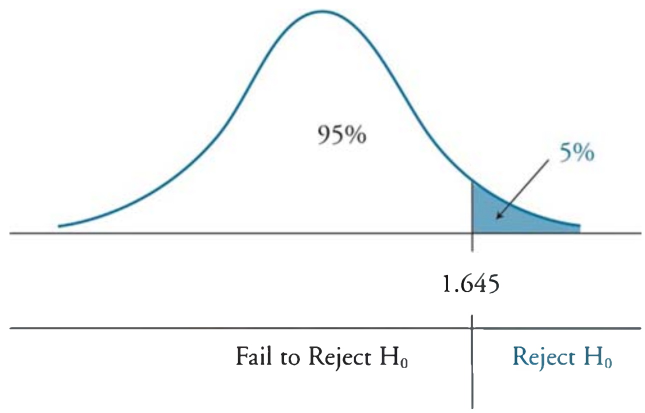
The appropriate set of hypotheses depends on whether we believe the population mean, μ, to be greater than (upper tail) or less than (lower tail) the hypothesized value, μ_0. Using a *z*-test at the 5% level of significance, the computed test statistic is compared with the critical values of 1.645 for the upper tail tests (i.e., H_A: $\mu > \mu_0$) or –1.645 for lower tail tests (i.e., H_A: $\mu < \mu_0$). These critical values are obtained from a *z*-table, where $-z_{0.05} = -1.645$ corresponds to a cumulative probability equal to 5%, and the $z_{0.05} = 1.645$ corresponds to a cumulative probability of 95% (1 – 0.05).

Let's use the upper tail test structure where H_0: $\mu \leq \mu_0$ and H_A: $\mu > \mu_0$.

- If the calculated test statistic is greater than 1.645, we conclude that the sample statistic is sufficiently greater than the hypothesized value. In other words, we reject the null hypothesis.
- If the calculated test statistic is less than 1.645, we conclude that the sample statistic is not sufficiently different from the hypothesized value, and we fail to reject the null hypothesis.

Figure 4 shows the standard normal distribution and the rejection region for a one-tailed test (upper tail) at the 5% level of significance.

Figure 4: One-Tailed Hypothesis Test Using the Standard Normal (*z*) Distribution



Example: One-tailed test

Perform a *z*-test using the option portfolio data from the previous example to test the belief that option returns are positive.

Answer:

In this case, we use a one-tailed test with the following structure:

H_0: $\mu \leq 0$ versus H_A: $\mu > 0$

The appropriate decision rule for this one-tailed *z*-test at a significance level of 5% is:

Reject H_0 if: test statistic > 1.645

The test statistic is computed the same way, regardless of whether we are using a one-tailed or two-tailed test. From the previous example, we know the test statistic for the option return sample is 6.33. Since 6.33 > 1.645, we reject the null hypothesis and conclude that mean returns are statistically greater than zero at a 5% level of significance.

Type I and Type II Errors

Keep in mind that hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population. We must be aware that there is some probability that the sample, in some way, does not represent the population and any conclusion based on the sample about the population may be made in error.

When drawing inferences from a hypothesis test, there are two types of errors:

- Type I error: the rejection of the null hypothesis when it is actually true.
- Type II error: the failure to reject the null hypothesis when it is actually false.

The significance level is the probability of making a Type I error (rejecting the null when it is true) and is designated by the Greek letter alpha (α). For instance, a significance level of 5% ($\alpha = 0.05$) means there is a 5% chance of rejecting a true null hypothesis. When conducting hypothesis tests, a significance level must be specified in order to identify the critical values needed to evaluate the test statistic.

The decision for a hypothesis test is to either reject the null hypothesis or fail to reject the null hypothesis. Note that it is statistically incorrect to say "accept" the null hypothesis; it can only be supported or rejected. The decision rule for rejecting or failing to reject the null hypothesis is based on the distribution of the test statistic. For example, if the test statistic follows a normal distribution, the decision rule is based on critical values determined from the standard normal distribution (*z*-distribution). Regardless of the appropriate distribution, it must be determined if a one-tailed or two-tailed hypothesis test is appropriate before a decision rule (rejection rule) can be determined.

A decision rule is specific and quantitative. Once we have determined whether a one- or two-tailed test is appropriate, the significance level we require, and the distribution of the test statistic, we can calculate the exact critical value for the test statistic. Then we have a decision rule of the following form: if the test statistic is (greater, less than) the value X, reject the null.

The Power of a Test

While the significance level of a test is the probability of rejecting the null hypothesis when it is true, the power of a test is the probability of correctly rejecting the null hypothesis when it is false. The power of a test is actually one minus the probability of making a Type II error, or 1 – P(Type II error). In other words, the probability of rejecting the null when it is false (power of the test) equals one minus the probability of *not* rejecting the null when it is false (Type II error). When more than one test statistic may be used, the power of the test for the competing test statistics may be useful in deciding which test statistic to use. Ordinarily, we wish to use the test statistic that provides the most powerful test among all possible tests.

Figure 5 shows the relationship between the level of significance, the power of a test, and the two types of errors.

Figure 5: Type I and Type II Errors in Hypothesis Testing

	True Condition	
Decision	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision Type II error
Reject H_0	Incorrect decision Type I error Significance level, α, = P(Type I error)	Correct decision Power of the test = 1 – P(Type II error)

Sample size and the choice of significance level (Type I error probability) will together determine the probability of a Type II error. The relation is not simple, however, and calculating the probability of a Type II error in practice is quite difficult. Decreasing the significance level (probability of a Type I error) from 5% to 1%, for example, will increase the probability of failing to reject a false null (Type II error) and, therefore, reduce the power of the test. Conversely, for a given sample size, we can increase the power of a test only with the cost that the probability of rejecting a true null (Type I error) increases. For a given significance level, we can decrease the probability of a Type II error and increase the power of a test, only by increasing the sample size.

The Relation Between Confidence Intervals and Hypothesis Tests

A confidence interval is a range of values within which the researcher believes the true population parameter may lie.

A confidence interval is determined as:

$$\left\{\left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} - \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right] \leq \begin{matrix}\text{population}\\\text{parameter}\end{matrix} \leq \left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} + \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right]\right\}$$

The interpretation of a confidence interval is that for a level of confidence of 95%, for example, there is a 95% probability that the true population parameter is contained in the interval.

From the previous expression, we see that a confidence interval and a hypothesis test are linked by the critical value. For example, a 95% confidence interval uses a critical value associated with a given distribution at the 5% level of significance. Similarly, a hypothesis test would compare a test statistic to a critical value at the 5% level of significance. To see this relationship more clearly, the expression for the confidence interval can be manipulated and restated as:

$$-\text{critical value} \leq \text{test statistic} \leq +\text{critical value}$$

This is the range within which we fail to reject the null for a two-tailed hypothesis test at a given level of significance.

Example: Confidence interval

Using option portfolio data from the previous examples, **construct** a 95% confidence interval for the population mean daily return over the 250-day sample period. Use a *z*-distribution. **Decide** if the hypothesis $\mu = 0$ should be rejected.

Answer:

Given a sample size of 250 with a standard deviation of 0.25%, the standard error can be computed as $s_{\bar{x}} = {}^{s}\!/\!_{\sqrt{n}} = 0.25/\sqrt{250} = 0.0158\%$.

At the 5% level of significance, the critical *z*-values for the confidence interval are $z_{0.025}$ = 1.96 and $-z_{0.025}$ = −1.96. Thus, given a sample mean equal to 0.1%, the 95% confidence interval for the population mean is:

$0.1 - 1.96(0.0158) \le \mu \le 0.1 + 1.96(0.0158)$, or
$0.069\% \le \mu \le 0.1310\%$

Since there is a 95% probability that the true mean is within this confidence interval, we can reject the hypothesis $\mu = 0$ because 0 is not within the confidence interval.

Notice the similarity of this analysis with our test of whether $\mu = 0$. We rejected the hypothesis $\mu = 0$ because the sample mean of 0.1% is more than 1.96 standard errors from zero. Based on the 95% confidence interval, we reject $\mu = 0$ because zero is more than 1.96 standard errors from the sample mean of 0.1%.

Statistical Significance vs. Economic Significance

Statistical significance does not necessarily imply economic significance. For example, we may have tested a null hypothesis that a strategy of going long all the stocks that satisfy some criteria and shorting all the stocks that do not satisfy the criteria resulted in returns that were less than or equal to zero over a 20-year period. Assume we have rejected the null in favor of the alternative hypothesis that the returns to the strategy are greater than zero (positive). This does not necessarily mean that investing in that strategy will result in economically meaningful positive returns. Several factors must be considered.

One important consideration is transactions costs. Once we consider the costs of buying and selling the securities, we may find that the mean positive returns to the strategy are not enough to generate positive returns. Taxes are another factor that may make a seemingly attractive strategy a poor one in practice. A third reason that statistically significant results may not be economically significant is risk. In the above strategy, we have additional risk from short sales (they may have to be closed out earlier than in the test strategy). Since the statistically significant results were for a period of 20 years, it may be the case that there is significant variation from year to year in the returns from the strategy, even though the mean strategy return is greater than zero. This variation in returns from period to period is an additional risk to the strategy that is not accounted for in our test of statistical significance.

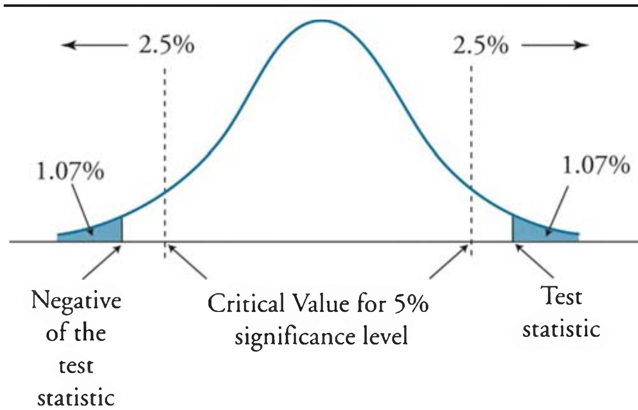
Any of these factors could make committing funds to a strategy unattractive, even though the statistical evidence of positive returns is highly significant. By the nature of statistical tests, a very large sample size can result in highly (statistically) significant results that are quite small in absolute terms.

The *p*-Value

The *p*-value is the probability of obtaining a test statistic that would lead to a rejection of the null hypothesis, assuming the null hypothesis is true. It is the smallest level of significance for which the null hypothesis can be rejected. For one-tailed tests, the *p*-value is the probability that lies above the computed test statistic for upper tail tests or below the computed test statistic for lower tail tests. For two-tailed tests, the *p*-value is the probability that lies above the positive value of the computed test statistic *plus* the probability that lies below the negative value of the computed test statistic.

Consider a two-tailed hypothesis test about the mean value of a random variable at the 95% significance level where the test statistic is 2.3, greater than the upper critical value of 1.96. If we consult the *z*-table, we find the probability of getting a value greater than 2.3 is (1 – 0.9893) = 1.07%. Since it's a two-tailed test, our *p*-value is 2 × 1.07 = 2.14%, as illustrated in Figure 6 At a 3%, 4%, or 5% significance level, we would reject the null hypothesis, but at a 2% or 1% significance level, we would not. Many researchers report *p*-values without selecting a significance level and allow the reader to judge how strong the evidence for rejection is.

Figure 6: Two-Tailed Hypothesis Test with *p*-Value = 2.14%



The *t*-Test

When hypothesis testing, the choice between using a critical value based on the *t*-distribution or the *z*-distribution depends on sample size, the distribution of the population, and whether the variance of the population is known.

The *t*-test is a widely used hypothesis test that employs a test statistic that is distributed according to a *t*-distribution. Following are the rules for when it is appropriate to use the *t*-test for hypothesis tests of the population mean.

Use the t*-test if the population variance is unknown* and either of the following conditions exist:

- The sample is large ($n \geq 30$).
- The sample is small ($n < 30$), but the distribution of the population is normal or approximately normal.

If the sample is small and the distribution is non-normal, we have no reliable statistical test.

The computed value for the test statistic based on the *t*-distribution is referred to as the *t*-statistic. For hypothesis tests of a population mean, a *t*-statistic with $n - 1$ degrees of freedom is computed as:

$$t_{n-1} = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

where:
$\overline{x}$ = sample mean
μ_0 = hypothesized population mean (i.e., the null)
s = standard deviation of the sample
n = sample size

Professor's Note: This computation is not new. It is the same test statistic computation that we have been performing all along. Note the use of the sample standard deviation, s*, in the standard error term in the denominator.*

To conduct a *t*-test, the *t*-statistic is compared to a critical *t*-value at the desired level of significance with the appropriate degrees of freedom.

In the real world, the underlying variance of the population is rarely known, so the *t*-test enjoys widespread application.

The *z*-Test

The *z*-test is the appropriate hypothesis test of the population mean when the *population is normally distributed with known variance.* The computed test statistic used with the *z*-test is referred to as the *z*-statistic. The *z*-statistic for a hypothesis test for a population mean is computed as follows:

$$\text{z-statistic} = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

where:
$\overline{x}$ = sample mean
μ_0 = hypothesized population mean
σ = standard deviation of the population
n = sample size

To test a hypothesis, the *z*-statistic is compared to the critical *z*-value corresponding to the significance of the test. Critical *z*-values for the most common levels of significance are displayed in Figure 7. You should memorize these critical values for the exam.

Figure 7: Critical *z*-Values

Level of Significance	*Two-Tailed Test*	*One-Tailed Test*
0.10 = 10%	±1.65	+1.28 or –1.28
0.05 = 5%	±1.96	+1.65 or –1.65
0.01 = 1%	±2.58	+2.33 or –2.33

When the *sample size is large* and the *population variance is unknown*, the *z*-statistic is:

$$\text{z-statistic} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

where:
$\bar{x}$ = sample mean
μ_0 = hypothesized population mean
s = standard deviation of the sample
n = sample size

Note the use of the sample standard deviation, *s*, versus the population standard deviation, σ. Remember, this is acceptable if the sample size is large, although the *t*-statistic is the more conservative measure when the population variance is unknown.

Example: *z*-test or *t*-test?

Referring to our previous option portfolio mean return problem once more, **determine** which test statistic (*z* or *t*) should be used and the difference in the likelihood of rejecting a true null with each distribution.

Answer:

The population variance for our sample of returns is unknown. Hence, the *t*-distribution is appropriate. With 250 observations, however, the sample is considered to be large, so the *z*-distribution would also be acceptable. This is a trick question—either distribution, *t* or *z*, is appropriate. With regard to the difference in the likelihood of rejecting a true null, since our sample is so large, the critical values for the *t* and *z* are almost identical. Hence, there is almost no difference in the likelihood of rejecting a true null.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

LO 19.5: Interpret the results of hypothesis tests with a specific level of confidence.

Example: The *z*-test

When your company's gizmo machine is working properly, the mean length of gizmos is 2.5 inches. However, from time to time the machine gets out of alignment and produces gizmos that are either too long or too short. When this happens, production is stopped and the machine is adjusted. To check the machine, the quality control department takes a gizmo sample each day. Today, a random sample of 49 gizmos showed a mean length of 2.49 inches. The population standard deviation is known to be 0.021 inches. Using a 5% significance level, **determine** if the machine should be shut down and adjusted.

Answer:

Let μ be the mean length of all gizmos made by this machine, and let $\bar{x}$ be the corresponding mean for the sample.

Let's follow the hypothesis testing procedure presented earlier in Figure 2. Again, you should know this process.

Statement of hypothesis. For the information provided, the null and alternative hypotheses are appropriately structured as:

H_0: $\mu = 2.5$ (The machine does not need an adjustment.)
H_A: $\mu \neq 2.5$ (The machine needs an adjustment.)

Note that since this is a two-tailed test, H_A allows for values above and below 2.5.

Select the appropriate test statistic. Since the population variance is known and the sample size is > 30, the *z*-statistic is the appropriate test statistic. The *z*-statistic is computed as:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Specify the level of significance. The level of significance is given at 5%, implying that we are willing to accept a 5% probability of rejecting a true null hypothesis.

State the decision rule regarding the hypothesis. The ≠ sign in the alternative hypothesis indicates that the test is two-tailed with two rejection regions, one in each tail of the standard normal distribution curve. Because the total area of both rejection regions combined is 0.05 (the significance level), the area of the rejection region in each tail is 0.025. You should know that the critical *z*-values for $\pm z_{0.025}$ are ±1.96. This means that the null hypothesis should not be rejected if the computed z-statistic lies between –1.96 and +1.96 and should be rejected if it lies outside of these critical values. The decision rule can be stated as:

Reject H_0 if: z-statistic < $-z_{0.025}$ or z-statistic > $z_{0.025}$, or equivalently,
Reject H_0 if: z-statistic < –1.96 or z-statistic > + 1.96

Collect the sample and calculate the test statistic. The value of $\bar{x}$ from the sample is 2.49. Since σ is given as 0.021, we calculate the *z*-statistic using σ as follows:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{2.49 - 2.5}{0.021 / \sqrt{49}} = \frac{-0.01}{0.003} = -3.33$$

Make a decision regarding the hypothesis. The calculated value of the *z*-statistic is –3.33. Since this value is less than the critical value, $-z_{0.025} = -1.96$, it falls in the rejection region in the left tail of the *z*-distribution. Hence, there is sufficient evidence to reject H_0.

Make a decision based on the results of the test. Based on the sample information and the results of the test, it is concluded that the machine is out of adjustment and should be shut down for repair.

The Chi-Squared Test

The *chi-squared test* is used for hypothesis tests concerning the variance of a normally distributed population. Letting σ^2 represent the true population variance and σ_0^2 represent the hypothesized variance, the hypotheses for a two-tailed test of a single population variance are structured as:

H_0: $\sigma^2 = \sigma_0^2$ versus H_A: $\sigma^2 \neq \sigma_0^2$

The hypotheses for one-tailed tests are structured as:

H_0: $\sigma^2 \leq \sigma_0^2$ versus H_A: $\sigma^2 > \sigma_0^2$, or
H_0: $\sigma^2 \geq \sigma_0^2$ versus H_A: $\sigma^2 < \sigma_0^2$

Hypothesis testing of the population variance requires the use of a chi-squared distributed test statistic, denoted χ^2. The chi-squared distribution is asymmetrical and approaches the normal distribution in shape as the degrees of freedom increase.

To illustrate the chi-squared distribution, consider a two-tailed test with a 5% level of significance and 30 degrees of freedom. As displayed in Figure 8, the critical chi-squared values are 16.791 and 46.979 for the lower and upper bounds, respectively. These values are obtained from a chi-squared table, which is used in the same manner as a *t*-table. A portion of a chi-squared table is presented in Figure 9.

Note that the chi-squared values in Figure 9 correspond to the probabilities in the right tail of the distribution. As such, the 16.791 in Figure 8 is from the column headed *0.975* because 95% + 2.5% of the probability is to the right of it. The 46.979 is from the column headed *0.025* because only 2.5% probability is to the right of it. Similarly, at a 5% level of significance with 10 degrees of freedom, Figure 9 shows that the critical chi-squared values for a two-tailed test are 3.247 and 20.483.

Figure 8: Decision Rule for a Two-Tailed Chi-Squared Test

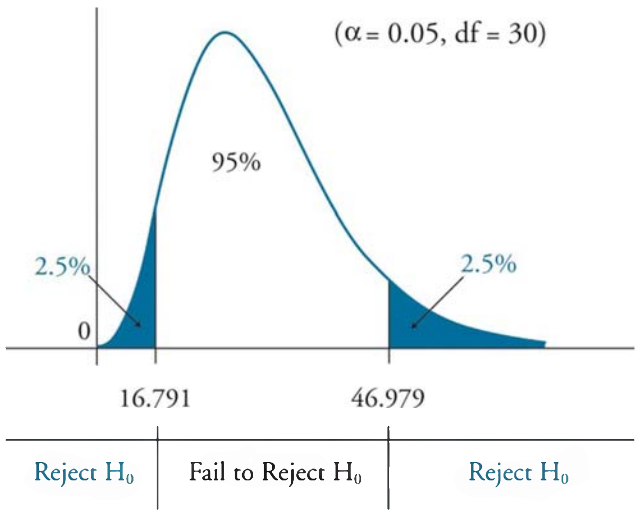


Figure 9: Chi-Squared Table

Degrees of Freedom	*Probability in Right Tail*					
	0.975	*0.95*	*0.90*	*0.1*	*0.05*	*0.025*
9	2.700	3.325	4.168	14.684	16.919	19.023
10	3.247	3.940	4.865	15.987	8.307	20.483
11	3.816	4.575	5.578	17.275	19.675	21.920
30	**16.791**	18.493	20.599	40.256	43.773	**46.979**

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