Question #1 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373

The average salary for a sample of 61 CFA charterholders with 10 years experience is $200,000, and the sample standard deviation is $80,000. Assume the population is normally distributed. Which of the following is a 99% confidence interval for the population mean salary of CFA charterholders with 10 years of experience?

A) $160,000 to $240,000.

B) $172,754 to $227,246.

C) $172,514 to $227,486.

Explanation

If the distribution of the population is *normal*, but we *don't know* the population variance, we can use the Student's t-distribution to construct a confidence interval. Because there are 61 observations, the degrees of freedom are 60. From the student's t table, we can determine that the reliability factor for $t_{\alpha/2}$, or $t_{0.005}$, is 2.660. Then the 99% confidence interval is $200,000 ± 2.660($80,000 / √61) or $200,000 ± 2.660 × $10,243, or $200,000 ± $27,246.

(Study Session 3, Module 11.2, LOS 11.j)

Question #2 of 89

A traffic engineer is trying to measure the effects of carpool-only lanes on the expressway. Based on a sample of 1,000 cars at rush hour, he finds that the mean number of occupants per car is 2.5, with a standard deviation of 0.4. Assuming that the population is normally distributed, what is the confidence interval at the 5% significance level for the number of occupants per car?

A) 2.288 to 2.712.

B) 2.455 to 2.555.

C) 2.475 to 2.525.

Explanation

The Z-score corresponding with a 5% significance level (95% confidence level) is 1.96. The confidence interval is equal to: 2.5 ± 1.96(0.4 / √1,000) = 2.475 to 2.525. (We can use Z-scores because the size of the sample is so large.)

(Study Session 3, Module 11.2, LOS 11.j)

Question #3 of 89

Studies of performance of a sample of mutual fund managers *most likely* suffer from:

A) sample-selection bias.

B) look-ahead bias.

C) survivorship bias.

Explanation

Studies of the performance of mutual fund managers often suffer from survivorship bias as poorly performing funds are closed down and are not included in the sample.

(Study Session 3, Module 11.2, LOS 11.k)

Question #4 of 89

A sample of 100 individual investors has a mean portfolio value of $28,000 with a standard deviation of $4,250. The 95% confidence interval for the population mean is *closest* to:

A) $27,575 to $28,425.

B) $19,500 to $28,333.

C) $27,159 to $28,842.

Explanation

Confidence interval = mean ± t_c\{S / √n\}

= 28,000 ± (1.98) (4,250 / √100) or 27,159 to 28,842

If you use a *z*-statistic because of the large sample size, you get 28,000 ± (1.96) (4,250 / √100) = 27,167 to 28,833, which is closest to the correct answer.

(Study Session 3, Module 11.2, LOS 11.j)

Question #5 of 89

A study reports that from 2002 to 2004 the average return on growth stocks was twice as large as that of value stocks. These results *most likely* reflect:

A) survivorship bias.

B) look-ahead bias.

C) time-period bias.

Explanation

Time-period bias can result if the time period over which the data is gathered is either too short because the results may reflect phenomenon specific to that time period, or if a change occurred during the time frame that would result in two different return distributions. In this case the time period sampled is probably not large enough to draw any conclusions about the long-term relative performance of value and growth stocks, even if the sample size within that time period is large.

Look-ahead bias occurs when the analyst uses historical data that was not publicly available at the time being studied. Survivorship bias is a form of sample selection bias in which the observations in the sample are biased because the elements of the sample that *survived* until the sample was taken are different than the elements that dropped out of the population.

(Study Session 3, Module 11.2, LOS 11.k)

Question #6 of 89

From the entire population of McDonald's franchises, an analyst constructs a sample of the monthly sales volume for 20 randomly selected franchises. She calculates the mean sales volume for those 20 franchises to be $400,000. The sampling distribution of the mean is the probability distribution of the:

A) mean monthly sales volume estimates from all possible samples of 20 observations. ✓

B) monthly sales volume for all McDonald's franchises. ✗

C) mean monthly sales volume estimates from all possible samples. ✗

Explanation

The sampling distribution of a sample statistic is a probability distribution made up of all possible *sample statistics* computed from samples *of the same size* randomly drawn from the same population, along with their associated probabilities.

(Study Session 3, Module 11.1, LOS 11.a)

Question #7 of 89

Which of the following is *least likely* a step in stratified random sampling?

A) The population is divided into strata based on some classification scheme. ✗

B) The sub-samples are pooled to create the complete sample. ✗

C) The size of each sub-sample is selected to be the same across strata. ✓

Explanation

In stratified random sampling we first divide the population into subgroups, called strata, based on some classification scheme. Then we randomly select a sample from each stratum and pool the results. *The size of the samples from each strata is based on the relative size of the strata relative to the population and are not necessarily the same across strata.*

(Study Session 3, Module 11.1, LOS 11.c)

Question #8 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690

A random sample of 25 Indiana farms had a mean number of cattle per farm of 27 with a sample standard deviation of five. Assuming the population is normally distributed, what would be the 95% confidence interval for the number of cattle per farm?

A) 25 to 29.

B) 22 to 32.

C) 23 to 31.

Explanation

The standard error of the sample mean = 5 / √25 = 1

Degrees of freedom = 25 − 1 = 24

From Student's t-table, $t_{5/2}$ = 2.064

The confidence interval is: 27 ± 2.064(1) = 24.94 to 29.06 or 25 to 29.

(Study Session 3, Module 11.2, LOS 11.j)

Question #9 of 89

What is the 95% confidence interval for a population mean with a known population variance of 9, based on a sample of 400 observations with mean of 96?

A) 95.118 to 96.882.

B) 95.706 to 96.294.

C) 95.613 to 96.387.

Explanation

Because we can compute the population standard deviation, we use the z-statistic. A 95% confidence level is constructed by taking the population mean and adding and subtracting the product of the z-statistic reliability ($z_{\alpha/2}$) factor times the known standard deviation of the population divided by the square root of the sample size (note that the population variance is given and its positive square root is the standard deviation of the population): $x \pm z_{\alpha/2} \times (\sigma / n^{1/2}) = 96 \pm 1.96 \times (9^{1/2} / 400^{1/2}) = 96 \pm 1.96 \times (0.15) = 96 \pm 0.294$ = 95.706 to 96.294.

(Study Session 3, Module 11.2, LOS 11.j)

Question #10 of 89

Which of the following statements about sampling errors is *least* accurate?

A) Sampling error is the error made in estimating the population mean based on a sample mean.

B) Sampling error is the difference between a sample statistic and its corresponding population parameter.

C) Sampling errors are errors due to the wrong sample being selected from the population.

Explanation

Sampling error is the difference between a sample statistic (the mean, variance, or standard deviation of the sample) and its corresponding population parameter (the mean, variance, or standard deviation of the population).

(Study Session 3, Module 11.1, LOS 11.b)

Question #11 of 89

An analyst is asked to calculate standard deviation using monthly returns over the last five years. These data are *best* described as:

A) cross-sectional data.

B) time series data.

C) systematic sampling data.

Explanation

Time series data are taken at equally spaced intervals, such as monthly, quarterly, or annual. Cross sectional data are taken at a single point in time. An example of cross-sectional data is dividend yields on 500 stocks as of the end of a year.

(Study Session 3, Module 11.1, LOS 11.d)

Question #12 of 89

A population has a mean of 20,000 and a standard deviation of 1,000. Samples of size n = 2,500 are taken from this population. What is the standard error of the sample mean?

A) 20.00.

B) 400.00.

C) 0.04.

Explanation

The standard error of the sample mean is estimated by dividing the standard deviation of the sample by the square root of the sample size: $s_x = s / n^{1/2} = 1000 / (2500)^{1/2} = 1000 / 50 = 20$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #13 of 89

Which of the following statements about sample statistics is *least* accurate?

A) The *z*-statistic is used to test normally distributed data with a known variance, whether testing a large or a small sample.

B) There is no sample statistic for non-normal distributions with unknown variance for either small or large samples.

C) The *z*-statistic is used for nonnormal distributions with known variance, but only for large samples.

Explanation

There is no sample statistic for non-normal distributions with unknown variance for small samples, but the t-statistic is used when the sample size is large.

(Study Session 3, Module 11.2, LOS 11.k)

Question #14 of 89

A research paper that reports finding a profitable trading strategy without providing any discussion of an economic theory that makes predictions consistent with the empirical results is *most likely* evidence of:

A) a sample that is not large enough.

B) data mining.

C) a non-normal population distribution.

Explanation

Data mining occurs when the analyst continually uses the same database to search for patterns or trading rules until he finds one that *works*. If you are reading research that suggests a profitable trading strategy, make sure you heed the following warning signs of data mining:

Evidence that the author used many variables (most unreported) until he found ones that were significant.

The lack of any economic theory that is consistent with the empirical results.

(Study Session 3, Module 11.2, LOS 11.k)

Question #15 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373

Based on Student's t-distribution, the 95% confidence interval for the population mean based on a sample of 40 interest rates with a sample mean of 4% and a sample standard deviation of 15% is *closest to:*

A) 1.261% to 6.739%.

B) -0.794% to 8.794%.

C) -0.851% to 8.851%.

Explanation

The standard error for the mean = s/(n)$^{0.5}$ = 15%/(40)$^{0.5}$ = 2.372%. The critical value from the t-table should be based on 40 – 1 = 39 df. Since the standard tables do not provide the critical value for 39 df the closest available value is for 40 df. This leaves us with an approximate confidence interval. Based on 95% confidence and df = 40, the critical t-value is 2.021. Therefore the 95% confidence interval is approximately: 4% ± 2.021(2.372) or 4% ± 4.794% or -0.794% to 8.794%.

(Study Session 3, Module 11.2, LOS 11.j)

Question #16 of 89

When sampling from a population, the *most* appropriate sample size:

A) is at least 30.

B) minimizes the sampling error and the standard deviation of the sample statistic around its population value.

C) involves a trade-off between the cost of increasing the sample size and the value of increasing the precision of the estimates.

Explanation

A larger sample reduces the sampling error and the standard deviation of the sample statistic around its population value. However, this does not imply that the sample should be as large as possible, or that the sampling error must be as small as can be achieved. Larger samples might contain observations that come from a different population, in which case they would not necessarily improve the estimates of the population parameters. Cost also increases with the sample size. When the cost of increasing the sample size is greater than the value of the extra precision gained, increasing the sample size is not appropriate.

(Study Session 3, Module 11.2, LOS 11.k)

Question #17 of 89

Which of the following statements about sampling and estimation is *most* accurate?

A) Time-series data are observations over individual units at a point in time.

B) A confidence interval estimate consists of a range of values that bracket the parameter with a specified level of probability, $1 - \beta$.

C) A point estimate is a single estimate of an unknown population parameter calculated as a sample mean.

Explanation

Time-series data are observations taken at specific and equally-spaced points.

A confidence interval estimate consists of a range of values that bracket the parameter with a specified level of probability, $1 - \alpha$.

(Study Session 3, Module 11.2, LOS 11.h)

Question #18 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850

A traffic engineer is trying to measure the effects of carpool-only lanes on the expressway. Based on a sample of 20 cars at rush hour, he finds that the mean number of occupants per car is 2.5, with a standard deviation of 0.4. If the population is normally distributed, what is the confidence interval at the 5% significance level for the number of occupants per car?

A) 2.387 to 2.613.

B) 2.313 to 2.687.

C) 2.410 to 2.589.

Explanation

The reliability factor corresponding with a 5% significance level (95% confidence level) for the Student's t-distribution with (20 − 1) degrees of freedom is 2.093. The confidence interval is equal to: $2.5 \pm 2.093(0.4 / \sqrt{20}) = 2.313$ to 2.687. (We must use the Student's t-distribution and reliability factors because of the small sample size.)

(Study Session 3, Module 11.2, LOS 11.j)

Question #19 of 89

A sample of size n = 25 is selected from a normal population. This sample has a mean of 15 and a sample variance of 4. What is the standard error of the sample mean?

A) 2.0.

B) 0.8.

C) 0.4.

Explanation

The standard error of the sample mean is estimated by dividing the standard deviation of the sample by the square root of the sample size. The standard deviation of the sample is calculated by taking the positive square root of the sample variance $4^{1/2} = 2$. Applying the formula: $s_x = s / n^{1/2} = 2 / (25)^{1/2} = 2 / 5 = 0.4$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #20 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373

The approximate 99% confidence interval for the population mean based on a sample of 60 returns with a mean of 7% and a sample standard deviation of 25% is *closest* to:

A) 0.546% to 13.454%.

B) 1.584% to 14.584%.

C) -1.584% to 15.584%.

Explanation

The standard error for the mean = s / $(n)^{0.5}$ = 25% / $(60)^{0.5}$ = 3.227%. The critical value from the t-table should be based on 60 − 1 = 59 df. Since the standard tables do not provide the critical value for 59 df the closest available value is for 60 df. This leaves us with an approximate confidence interval. Based on 99% confidence and df = 60, the critical t-value is 2.660. Therefore the 99% confidence interval is approximately: 7% ± 2.660(3.227) or 7% ± 8.584% or -1.584% to 15.584%.

If you use a z-statistic, the confidence interval is 7% ± 2.58(3.227) = -1.326% to 15.326%, which is closest to the correct choice.

(Study Session 3, Module 11.2, LOS 11.j)

Question #21 of 89

From a population of 5,000 observations, a sample of n = 100 is selected. Calculate the standard error of the sample mean if the population standard deviation is 50.

A) 4.48.

B) 50.00.

C) 5.00.

Explanation

The standard error of the sample mean equals the standard deviation of the population divided by the square root of the sample size: $50 / 100^{1/2} = 5$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #22 of 89

Which of the following statements about sampling and estimation is *most* accurate?

A) The probability that a parameter lies within a range of estimated values is given by α.

B) The standard error of the sample means when the standard deviation of the population is known equals $\sigma / \sqrt{n}$, where σ = sample standard deviation adjusted by n − 1.

C) The standard error of the sample means when the standard deviation of the population is unknown equals $s / \sqrt{n}$, where s = sample standard deviation.

Explanation

The probability that a parameter lies within a range of estimated values is given by 1 − α. The standard error of the sample means when the standard deviation of the population is known equals $\sigma / \sqrt{n}$, where σ = *population* standard deviation.

(Study Session 3, Module 11.2, LOS 11.h)

Question #23 of 89

Frank Grinder is trying to introduce sampling into the quality control program of an old-line manufacturer. Grinder samples 38 items and finds that the standard deviation in size is 0.019 centimeters. What is the standard error of the sample mean?

A) 0.00204.

B) 0.00615.

C) 0.00308.

Explanation

If we do not know the standard deviation of the population (in this case we do not), then we estimate the standard error of the sample mean = the standard deviation of the sample / the square root of the sample size = $0.019 / \sqrt{38} = 0.00308$ centimeters.

(Study Session 3, Module 11.1, LOS 11.f)

Question #24 of 89

Monthly Gross Domestic Product (GDP) figures from 1990-2000 are an example of:

A) cross-sectional data.

B) systematic data.

C) time-series data.

Explanation

A time-series is a group of observations taken at specific and equally spaced points in time. Cross-sectional data are observations taken at a single point in time.

(Study Session 3, Module 11.1, LOS 11.d)

Question #25 of 89

Cumulative Z-Table

z	0.05	0.06	0.07	0.08	0.09
2.4	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9970	0.9971	0.9972	0.9973	0.9974

The average return on the Russell 2000 index for 121 monthly observations was 1.5%. The population standard deviation is assumed to be 8.0%. What is a 99% confidence interval for the mean monthly return on the Russell 2000 index?

A) -0.4% to 3.4%.

B) -6.5% to 9.5%.

C) 0.1% to 2.9%.

Explanation

Because we know the population standard deviation, we use the z-statistic. The z-statistic reliability factor for a 99% confidence interval is 2.575. The confidence interval is 1.5% ± 2.575[(8.0%)/√121] or 1.5% ± 1.9%.

(Study Session 3, Module 11.2, LOS 11.j)

Question #26 of 89

Sampling error can be defined as:

A) rejecting the null hypothesis when it is true.

B) the standard deviation of a sampling distribution of the sample means.

C) the difference between a sample statistic and its corresponding population parameter.

Explanation

This is the definition.

(Study Session 3, Module 11.1, LOS 11.b)

Question #27 of 89

An analyst has reviewed market data for returns from 1980–1990 extensively, searching for patterns in the returns. She has found that when the end of the month falls on a Saturday, there are usually positive returns on the following Thursday. She has engaged in:

A) data snooping.

B) biased selection.

C) data mining.

Explanation

Data mining refers to the extensive review of the same database searching for patterns.

(Study Session 3, Module 11.2, LOS 11.k)

Question #28 of 89

According to the Central Limit Theorem, the distribution of the sample means is approximately *normal* if:

A) the standard deviation of the population is known.

B) the sample size $n > 30$.

C) the underlying population is normal.

Explanation

The Central Limit Theorem states that if the sample size is sufficiently large (i.e. greater than 30) the sampling distribution of the sample means will be approximately normal.

(Study Session 3, Module 11.1, LOS 11.e)

Question #29 of 89

Construct a 90% confidence interval for the mean starting salaries of the CFA charterholders if a sample of 100 recent CFA charterholders gives a mean of 50. Assume that the population variance is 900. All measurements are in $1,000.

A) $50 \pm 1.645(30)$.

B) $50 \pm 1.645(900)$.

C) $50 \pm 1.645(3)$.

Explanation

Because we can compute the population standard deviation, we use the z-statistic. A 90% confidence level is constructed by taking the population mean and adding and subtracting the product of the z-statistic reliability ($z_{\alpha/2}$) factor times the known standard deviation of the population divided by the square root of the sample size (note that the population variance is given and its positive square root is the standard deviation of the population): $x \pm z_{\alpha/2} \times (\sigma / n^{1/2}) = 50 \pm 1.645 \times (900^{1/2} / 100^{1/2}) = 50 \pm 1.645 \times (30 / 10) = 50 \pm 1.645 \times (3)$. This is interpreted to mean that we are 90% confident that the above interval contains the true mean starting salaries of CFA charterholders.

(Study Session 3, Module 11.2, LOS 11.j)

Question #30 of 89

A nursery sells trees of different types and heights. Suppose that 75 trees chosen at random are sold for planting at City Hall. These 75 trees average 60 inches in height with a standard deviation of 16 inches.

Using this information, construct a 95% confidence interval for the mean height of all trees in the nursery.

A) 0.8 ± 1.96(16).

B) 60 ± 1.96(1.85).

C) 60 ± 1.96(16).

Explanation

Because the sample size is sufficiently large, we can use the *z*-statistic. A 95% confidence level is constructed by taking the sample mean and adding and subtracting the product of the *z*-statistic reliability factor ($z_{\alpha/2}$) times the standard error of the sample mean: $x \pm z_{\alpha/2} \times (s / n^{1/2}) = 60 \pm (1.96) \times (16 / 75^{1/2}) = 60 \pm (1.96) \times (16 / 8.6603) = 60 \pm (1.96) \times (1.85)$.

(Study Session 3, Module 11.2, LOS 11.j)

Question #31 of 89

A sample of five numbers drawn from a population is (5, 2, 4, 5, 4). Which of the following statements concerning this sample is *most* accurate?

A) The mean of the sample is $\sum X / (n - 1) = 5$.

B) The sampling error of the sample is equal to the standard error of the sample.

C) The variance of the sample is: $\sum(x_1 - \text{mean of the sample})^2 / (n - 1) = 1.5$.

Explanation

The mean of the sample is $\sum X / n = 20 / 5 = 4$. The sampling error of the sample is the difference between a sample statistic and its corresponding population parameter.

(Study Session 3, Module 11.1, LOS 11.b)

Question #32 of 89

The table below is for five samples drawn from five separate populations. The far left columns give information on the population distribution, population variance, and sample size. The right-hand columns give three choices for the appropriate tests: Z = *z*-statistic, and t = *t*-statistic. "None" means that a test statistic is not available.

Sampling From			*Test Statistic Choices*		
Distribution	*Variance*	*n*	***One***	***Two***	***Three***
Normal	5.60	75	Z	Z	Z
Non-normal	n/a	45	Z	*t*	*t*
Normal	n/a	1000	Z	*t*	*t*
Non-normal	14.3	15	*t*	none	*t*
Normal	0.056	10	Z	Z	*t*

Which set of test statistic choices (One, Two, or Three) matches the correct test statistic to the sample for all five samples?

A) One.

B) Three.

C) Two.

Explanation

For the exam: COMMIT THE FOLLOWING TABLE TO MEMORY!

When you are sampling from a:	and the sample size is **small**, use a:	and the sample size is **large**, use a:
Normal distribution with a *known* variance	*Z*-statistic	*Z*-statistic
Normal distribution with an *unknown* variance	*t*-statistic	*t*-statistic*
Nonnormal distribution with a *known* variance	not available	*Z*-statistic
Nonnormal distribution with an *unknown* variance	not available	*t*-statistic*

(Study Session 3, Module 11.2, LOS 11.j)

Question #33 of 89

Which of the following statements about confidence intervals is *least* accurate? A confidence interval:

A) has a significance level that is equal to one minus the degree of confidence.

B) expands as the probability that a point estimate falls within the interval decreases.

C) is constructed by adding and subtracting a given amount from a point estimate.

Explanation

A confidence interval contracts as the probability that a point estimate falls within the interval decreases.

(Study Session 3, Module 11.2, LOS 11.h)

Question #34 of 89

A sample of 25 junior financial analysts gives a mean salary (in thousands) of 60. Assume the population variance is known to be 100. A 90% confidence interval for the mean starting salary of junior financial analysts is *most* accurately constructed as:

A) 60 ± 1.645(2).

B) 60 ± 1.645(4).

C) 60 ± 1.645(10).

Explanation

Because we can compute the population standard deviation, we use the z-statistic. A 90% confidence level is constructed by taking the population mean and adding and subtracting the product of the z-statistic reliability ($z_{\alpha/2}$) factor times the known standard deviation of the population divided by the square root of the sample size (note that the population variance is given and its positive square root is the standard deviation of the population): $x \pm z_{\alpha/2} \times (\sigma / n^{1/2}) = 60 +/- 1.645 \times (100^{1/2} / 25^{1/2}) = 60 +/- 1.645 \times (10 / 5) = 60 +/- 1.645 \times 2$.

(Study Session 3, Module 11.2, LOS 11.j)

Question #35 of 89

If the true mean of a population is 16.62, according to the central limit theorem, the mean of the distribution of sample means, for all possible sample sizes n will be:

A) 16.62.

B) indeterminate for sample with n < 30.

C) 16.62 / √n.

Explanation

According to the central limit theorem, the mean of the distribution of sample means will be equal to the population mean. n > 30 is only required for distributions of sample means to approach normal distribution.

(Study Session 3, Module 11.1, LOS 11.e)

Question #36 of 89

When is the t-distribution the appropriate distribution to use? The t-distribution is the appropriate distribution to use when constructing confidence intervals based on:

A) large samples from populations with known variance that are nonnormal.

B) small samples from populations with known variance that are at least approximately normal.

C) small samples from populations with unknown variance that are at least approximately normal.

Explanation

The t-distribution is the appropriate distribution to use when constructing confidence intervals based on small samples from populations with unknown variance that are either normal or approximately normal.

(Study Session 3, Module 11.2, LOS 11.i)

Question #37 of 89

The sample mean is an unbiased estimator of the population mean because the:

A) sampling distribution of the sample mean has the smallest variance of any other unbiased estimators of the population mean.

B) expected value of the sample mean is equal to the population mean.

C) sample mean provides a more accurate estimate of the population mean as the sample size increases.

Explanation

An unbiased estimator is one for which the expected value of the estimator is equal to the parameter you are trying to estimate.

(Study Session 3, Module 11.1, LOS 11.g)

Question #38 of 89

The sample mean is a consistent estimator of the population mean because the:

A) sample mean provides a more accurate estimate of the population mean as the sample size increases.

B) sampling distribution of the sample mean has the smallest variance of any other unbiased estimators of the population mean.

C) expected value of the sample mean is equal to the population mean.

Explanation

A consistent estimator provides a more accurate estimate of the parameter as the sample size increases.

(Study Session 3, Module 11.1, LOS 11.g)

Question #39 of 89

If the number of offspring for females of a certain mammalian species has a mean of 16.4 and a standard deviation of 3.2, what will be the standard error of the sample mean for a survey of 25 females of the species?

A) 1.28.

B) 3.20.

C) 0.64.

Explanation

The standard error of the sample mean when the standard deviation of the population is known is equal to the standard deviation of the population divided by the square root of the sample size. In this case, 3.2 / √25 = 0.64.

(Study Session 3, Module 11.1, LOS 11.f)

Question #40 of 89

Which of the following would result in a wider confidence interval? A:

A) greater level of significance.

B) higher degree of confidence.

C) higher alpha level.

Explanation

A higher degree of confidence (e.g. 99% instead of 95%) would require a higher reliability factor (2.575 instead of 1.96 assuming a normal distribution). A wider confidence interval corresponds to a lower alpha significance level and the point estimate does not affect the width of the confidence interval.

(Study Session 3, Module 11.2, LOS 11.j)

Question #41 of 89

The average mutual fund return calculated from a sample of funds with significant survivorship bias would *most likely* be:

A) an unbiased estimate of the mean return of the population of all mutual funds if the sample size was large enough.

B) larger than the mean return of the population of all mutual funds.

C) smaller than the mean return of the population of all mutual funds.

Explanation

If we try to draw any conclusions from an analysis of a mutual fund database with survivorship bias, we overestimate the average mutual fund return, because we don't include the poorer-performing funds that dropped out. A larger sample size from a database with survivorship bias will still result in a biased estimate.

(Study Session 3, Module 11.2, LOS 11.k)

Question #42 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373

From a sample of 41 monthly observations of the S&P Mid-Cap index, the mean monthly return is 1% and the sample variance is 36. For which of the following intervals can one be *closest* to 95% confident that the population mean is contained in that interval?

A) 1.0% ± 1.9%.

B) 1.0% ± 6.0%.

C) 1.0% ± 1.6%.

Explanation

If the distribution of the population is *nonnormal*, but we *don't know* the population variance, we can use the Student's t-distribution to construct a confidence interval. The sample standard deviation is the square root of the variance, or 6%. Because there are 41 observations, the degrees of freedom are 40. From the Student's *t* distribution, we can determine that the reliability factor for $t_{0.025}$, is 2.021. Then the 95% confidence interval is 1.0% ± 2.021(6 / √41) or 1.0% ± 1.9%.

(Study Session 3, Module 11.2, LOS 11.j)

Question #43 of 89

The central limit theorem concerns the sampling distribution of the:

A) sample standard deviation.

B) population mean.

C) sample mean.

Explanation

The central limit theorem tells us that for a population with a mean m and a finite variance σ^2, the sampling distribution of the *sample means* of all possible samples of size n will approach a normal distribution with a mean equal to μ and a variance equal to σ^2 / n as *n* gets large.

(Study Session 3, Module 11.1, LOS 11.e)

Question #44 of 89

An analyst wants to generate a simple random sample of 500 stocks from all 10,000 stocks traded on the New York Stock Exchange, the American Stock Exchange, and NASDAQ. Which of the following methods is *least likely* to generate a random sample?

A) Using the 500 stocks in the S&P 500.

B) Assigning each stock a unique number and generating a number using a random number generator. Then selecting the stock with that number for the sample and repeating until there

C) Listing all the stocks traded on all three exchanges in alphabetical order and selecting every 20th stock.

Explanation

The S&P 500 is not a random sample of all stocks traded in the U.S. because it represents the 500 largest stocks. The other two choices are legitimate methods of selecting a simple random sample.

(Study Session 3, Module 11.1, LOS 11.a)

Question #45 of 89

A traffic engineer is trying to measure the effects of carpool-only lanes on the expressway. Based on a sample of 100 cars at rush hour, he finds that the mean number of occupants per car is 2.5, and the sample standard deviation is 0.4. What is the standard error of the sample mean?

A) 5.68.

B) 1.00.

C) 0.04.

Explanation

The standard error of the sample mean when the standard deviation of the population is not known is estimated by the standard deviation of the sample divided by the square root of the sample size. In this case, $0.4 / \sqrt{100} = 0.04$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #46 of 89

When sampling from a nonnormal distribution with an known variance, which statistic should be used if the sample size is *large* and if the respective sample size is *small*?

A) *z*-statistic; *z*-statistic.

B) *t*-statistic; *t*-statistic.

C) *z*-statistic; not available.

Explanation

When you are sampling from a:	and the sample size is small, use a:	and the sample size is large, use a:
Normal distribution with a *known* variance	*z*-statistic	*z*-statistic
Normal distribution with an *unknown* variance	*t*-statistic	*t*-statistic*
Nonnormal distribution with a *known* variance	not available	*z*-statistic
Nonnormal distribution with an *unknown* variance	not available	*t*-statistic*

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

(Study Session 3, Module 11.2, LOS 11.k)

Question #47 of 89

Suppose the mean debt/equity ratio of the population of all banks in the United States is 20 and the population variance is 25. A banking industry analyst uses a computer program to select a random sample of 50 banks from this population and compute the sample mean. The program repeats this exercise 1000 times and computes the sample mean each time. According to the central limit theorem, the sampling distribution of the 1000 sample means will be approximately normal if the population of bank debt/equity ratios has:

A) a Student's *t*-distribution, because the sample size is greater than 30.

B) a normal distribution, because the sample is random.

C) any probability distribution.

Explanation

The central limit theorem tells us that for a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample means of all possible samples of size *n* will be approximately normally distributed with a mean equal to μ and a variance equal to σ^2/n, *no matter the distribution of the population*, assuming a large sample size.

(Study Session 3, Module 11.1, LOS 11.e)

Question #48 of 89

Which of the following statements regarding the central limit theorem (CLT) is *least* accurate? The CLT:

A) holds for any population distribution, assuming a large sample size.

B) gives the variance of the distribution of sample means as σ^2 / n, where σ^2 is the population variance and n is the sample size.

C) states that for a population with mean μ and variance σ^2, the sampling distribution of the sample means for any sample of size n will be approximately normally distributed.

Explanation

This question is asking you to select the inaccurate statement. The CLT states that for a population with mean μ and a finite variance σ^2, the sampling distribution of the sample means becomes approximately normally distributed *as the sample size becomes large*. The other statements are accurate.

(Study Session 3, Module 11.1, LOS 11.e)

Question #49 of 89

Which of the following is NOT a prediction of the central limit theorem?

A) The variance of the sampling distribution of sample means will approach the population variance divided by the sample size.

B) The standard error of the sample mean will increase as the sample size increases.

C) The mean of the sampling distribution of the sample means will be equal to the population mean.

Explanation

The standard error of the sample mean is equal to the sample standard deviation divided by the square root of the sample size. As the sample size increases, this ratio decreases. The other two choices are predictions of the central limit theorem.

(Study Session 3, Module 11.1, LOS 11.e)

Question #50 of 89

A sample size of 25 is selected from a normal population. This sample has a mean of 15 and the population variance is 4.

Using this information, construct a 95% confidence interval for the population mean, m.

A) 15 ± 1.96(2).

B) 15 ± 1.96(0.4).

C) 15 ± 1.96(0.8).

Explanation

Because we can compute the population standard deviation, we use the z-statistic. A 95% confidence level is constructed by taking the population mean and adding and subtracting the product of the z-statistic reliability ($z_{\alpha/2}$) factor times the known standard deviation of the population divided by the square root of the sample size (note that the population variance is given and its positive square root is the standard deviation of the population): $x \pm z_{\alpha/2} \times (\sigma / n^{1/2}) = 15 \pm 1.96 \times (4^{1/2} / 25^{1/2}) = 15 \pm 1.96 \times (0.4)$.

(Study Session 3, Module 11.2, LOS 11.j)

Question #51 of 89

Sampling error is the:

A) estimation error created by using a non-random sample.

B) difference between the point estimate of the mean and the mean of the sampling distribution.

C) difference between a sample statistic and its corresponding population parameter.

Explanation

Sampling error is the difference between any sample statistic (the mean, variance, or standard deviation of the sample) and its corresponding population parameter (the mean, variance or standard deviation of the population). For example, the sampling error for the mean is equal to the sample mean minus the population mean.

(Study Session 3, Module 11.1, LOS 11.b)

Question #52 of 89

The sample of per square foot sales for 100 U.S. retailers in December 2004 is an example of:

A) cross-sectional data.

B) time-series data.

C) unbiased data.

Explanation

Cross-sectional data are a sample of observations taken at a single point in time. A time-series is a sample of observations taken at specific and equally spaced points in time.

(Study Session 3, Module 11.1, LOS 11.d)

Question #53 of 89

The sample mean return of Bartlett Co. is 3% and the standard deviation is 6% based on 30 monthly returns. What is the confidence interval of a two tailed z-test of the population mean with a 5% level of significance?

A) 2.61 to 3.39.

B) 0.85 to 5.15.

C) 1.90 to 4.10.

Explanation

The standard error of the sample is the standard deviation divided by the square root of n, the sample size. 6% / $30^{1/2}$ = 1.0954%.

The confidence interval = point estimate +/- (reliability factor × standard error)

confidence interval = 3 +/- (1.96 × 1.0954) = 0.85 to 5.15

(Study Session 3, Module 11.1, LOS 11.f)

Question #54 of 89

If the variance of the sampling distribution of an estimator is smaller than all other unbiased estimators of the parameter of interest, the estimator is:

A) consistent.

B) efficient.

C) unbiased.

Explanation

An estimator is efficient if the variance of its sampling distribution is smaller than that of all other unbiased estimators of the parameter.

(Study Session 3, Module 11.1, LOS 11.g)

Question #55 of 89

Thomas Merton, a car industry analyst, wants to investigate a relationship between the types of ads used in advertising campaigns and sales to customers in certain age groups. In order to make sure he includes manufacturers of all sizes, Merton divides the industry into four size groups and draws random samples from each group. What sampling method is Merton using?

A) Stratified random sampling.

B) Cross-sectional sampling.

C) Simple random sampling.

Explanation

In stratified random sampling, we first divide the population into subgroups based on some relevant characteristic(s) and then make random draws from each group.

(Study Session 3, Module 11.1, LOS 11.c)

Question #56 of 89

Melissa Cyprus, CFA, is conducting an analysis of inventory management practices in the retail industry. She assumes the population cross-sectional standard deviation of inventory turnover ratios is 20. How large a random sample should she gather in order to ensure a standard error of the sample mean of 4?

A) 25.

B) 20.

C) 80.

Explanation

Given the population standard deviation and the standard error of the sample mean, you can solve for the sample size. Because the standard error of the sample mean equals the standard deviation of the population divided by the square root of the sample size, $4 = 20 / n^{1/2}$, so $n^{1/2} = 5$, so $n = 25$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #57 of 89

The range of possible values in which an actual population parameter may be observed at a given level of probability is known as a:

A) significance level.

B) confidence interval.

C) degree of confidence.

Explanation

A confidence interval is a range of values within which the actual value of a parameter will lie, given a specified probability level. A point estimate is a single value used to estimate a population parameter. An example of a point estimate is a sample mean. The degree of confidence is the confidence level associated with a confidence interval and is computed as $1 - \alpha$.

(Study Session 3, Module 11.2, LOS 11.h)

Question #58 of 89

Sunil Hameed is a reporter with the weekly periodical The *Fun Finance Times*. Today, he is scheduled to interview a researcher who claims to have developed a successful technical trading strategy based on trading on the CEO's birthday (sample was taken from the Fortune 500). After the interview, Hameed summarizes his notes (partial transcript as follows). The researcher:

- was defensive about the lack of economic theory consistent with his results.
- used the same database of data for all his tests and has not tested the trading rule on out-of-sample data.
- excluded stocks for which he could not determine the CEO's birthday.
- used a sample cut-off date of the month before the latest market correction.

Select the choice that *best* completes the following: Hameed concludes that the research is flawed because the data and process are biased by:

A) data mining, time-period bias, and look-ahead bias.

B) data mining, sample selection bias, and time-period bias.

C) sample selection bias and time-period bias.

Explanation

Evidence that the researcher used *data mining* is that he was defensive about the lack of economic theory consistent with his results and that he used the same database of data for all his tests. One way to *avoid* data mining is to test the trading rule on out-of-sample data. *Sample selection bias* occurs when some data is systematically excluded from the analysis, usually because it is not available. Here, the researcher excluded stocks for which he could not determine the CEO's birthday. *Time-period bias* can result if the time period is too short or too long. Here, it is likely that the period was too short since the researcher used a cut-off date of the month before the latest market correction. *Note:* this could be an additional example of data mining.

We are not given enough information to determine if the researcher is guilty of look-ahead bias (which occurs when the analyst uses historical data that was not publicly available at the time being studied).

(Study Session 3, Module 11.2, LOS 11.k)

Question #59 of 89

Which of the following statements about a confidence interval for a population mean is *most* accurate?

A) When a *z*-statistic is acceptable, a 95% confidence interval for a population mean is the sample mean plus-or-minus 1.96 times the sample standard deviation.

B) If the population variance is unknown, a large sample size is required in order to estimate a confidence interval for the population mean.

C) For a sample size of 30, using a *t*-statistic will result in a wider confidence interval for a population mean than using a *z*-statistic.

Explanation

Although the *t*-distribution begins to approach the shape of a normal distribution for large sample sizes, at a sample size of 30 a *t*-statistic produces a wider confidence interval than a *z*-statistic. A confidence interval for the population mean is the sample mean plus-or-minus the appropriate critical value times the *standard error*, which is the standard deviation divided by the square root of the sample size. If a population is normally distributed, we can use a *t*-statistic to construct a confidence interval for the population mean from a small sample, even if the population variance is unknown.

(Study Session 3, Module 11.2, LOS 11.j)

Question #60 of 89

With 60 observations, what is the appropriate number of degrees of freedom to use when carrying out a statistical test on the mean of a population?

A) 59.

B) 61.

C) 60.

Explanation

When performing a statistical test on the mean of a population based on a sample of size n, the number of degrees of freedom is n – 1 since once the mean is estimated from a sample there are only n – 1 observations that are free to vary. In this case the appropriate number of degrees of freedom to use is 60 – 1 = 59.

(Study Session 3, Module 11.2, LOS 11.i)

Question #61 of 89

A range of estimated values within which the actual value of a population parameter will lie with a given probability of 1 − α is a(n):

A) α percent point estimate.

B) (1 − α) percent confidence interval.

C) α percent confidence interval.

Explanation

A 95% confidence interval for the population mean (α = 5%), for example, is a range of estimates within which the actual value of the population mean will lie with a probability of 95%. Point estimates, on the other hand, are *single* (sample) values used to estimate population parameters. There is no such thing as a α percent *point estimate* or a (1 − α) percent *cross-sectional point estimate.*

(Study Session 3, Module 11.2, LOS 11.h)

Question #62 of 89

Which of the following characterizes the typical construction of a confidence interval *most* accurately?

A) Standard error +/- (Point estimate / Reliability factor).

B) Point estimate +/- (Reliability factor × Standard error).

C) Point estimate +/- (Standard error / Reliability factor).

Explanation

We can construct a confidence interval by adding and subtracting some amount from the point estimate. In general, confidence intervals have the following form:

Point estimate +/- Reliability factor x Standard error

Point estimate = the value of a sample statistic of the population parameter

Reliability factor = a number that depends on the sampling distribution of the point estimate and the probability the point estimate falls in the confidence interval (1 – α)

Standard error = the standard error of the point estimate

(Study Session 3, Module 11.2, LOS 11.h)

Question #63 of 89

Segment of the table of critical values for Student's t-distribution:

Level of Significance for a One-Tailed Test		
df	**0.050**	**0.025**
Level of Significance for a Two-Tailed Test		
df	**0.10**	**0.05**
28	1.701	2.048
29	1.699	2.045
30	1.697	2.042
40	1.684	2.021

For a *t*-distributed test statistic with 30 degrees of freedom, a one-tailed test specifying the parameter greater than some value and a 95% confidence level, the critical value is:

A) 1.697.

B) 2.042.

C) 1.684.

Explanation

This is the critical value for a one-tailed probability of 5% and 30 degrees of freedom.

(Study Session 3, Module 11.2, LOS 11.i)

Question #64 of 89

The average return on small stocks over the period 1926-1997 was 17.7%, and the standard deviation of the sample was 33.9%. Assuming returns are normally distributed, the 95% confidence interval for the return on small stocks next year is:

A) –16.2% to 51.6%.

B) –48.7% to 84.1%.

C) 16.8% to 18.6%.

Explanation

A 95% confidence interval is ± 1.96 standard deviations from the mean, so 0.177 ± 1.96(0.339) = (–48.7%, 84.1%).

(Study Session 3, Module 11.2, LOS 11.j)

Question #65 of 89

Which one of the following statements about the t-distribution is *most* accurate?

A) The t-distribution approaches the standard normal distribution as the number of degrees of freedom becomes large. ✓

B) Compared to the normal distribution, the t-distribution has less probability in the tails. ✗

C) The t-distribution is the only appropriate distribution to use when constructing confidence intervals based on large samples. ✗

Explanation

As the number of degrees of freedom grows, the t-distribution approaches the shape of the standard normal distribution. Compared to the normal distribution, the t-distribution has fatter tails. When choosing a distribution, three factors must be considered: sample size, whether population variance is known, and if the distribution is normal.

(Study Session 3, Module 11.2, LOS 11.i)

Question #66 of 89

From a population with a known standard deviation of 15, a sample of 25 observations is taken. Calculate the standard error of the sample mean.

A) 1.67. ✗

B) 0.60. ✗

C) 3.00. ✓

Explanation

The standard error of the sample mean equals the standard deviation of the population divided by the square root of the sample size: $s_x = s / n^{1/2} = 15 / 25^{1/2} = 3$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #67 of 89

The average U.S. dollar/Euro exchange rate from a sample of 36 monthly observations is $1.00/Euro. The population variance is 0.49. What is the 95% confidence interval for the mean U.S. dollar/Euro exchange rate?

A) $0.7713 to $1.2287. ✓

B) $0.8075 to $1.1925. ✗

C) \$0.5100 to \$1.4900.

Explanation

The population *standard deviation* is the square root of the variance ($\sqrt{0.49} = 0.7$). Because we know the population standard deviation, we use the z-statistic. The z-statistic reliability factor for a 95% confidence interval is 1.960. The confidence interval is \$1.00 ± 1.960(\$0.7 / $\sqrt{36}$) or \$1.00 ± \$0.2287.

(Study Session 3, Module 11.2, LOS 11.j)

Question #68 of 89

An article in a trade journal suggests that a strategy of buying the seven stocks in the S&P 500 with the highest earnings-to-price ratio at the end of the calendar year and holding them until March 20 of the following year produces significant trading profits. Upon reading further, you discover that the study is based on data from 1993 to 1997, and the earnings-to-price ratio is calculated using the stock price on December 31 of each year and the annual reported earnings per share for that year. Which of the following biases is *least likely* to influence the reported results?

A) Survivorship bias.

B) Time-period bias.

C) Look-ahead bias.

Explanation

Survivorship bias is not likely to significantly influence the results of this study because the authors looked at the stocks in the S&P 500 at the beginning of the year and measured performance over the following three months. Look-ahead bias could be a problem because earnings-price ratios are calculated and the trading strategy implemented at a time before earnings are actually reported. Finally, the study is conducted over a relatively short time period during the long bull market of the 1990s. This suggests the results may be time-specific and the result of time-period bias.

(Study Session 3, Module 11.2, LOS 11.k)

Question #69 of 89

The following data are available on a sample of advertising budgets of 81 U.S. manufacturing companies: The mean budget is \$10 million. The sample variance is 36 million. The standard error of the sample mean is:

A) \$1,111.

B) \$667.

C) \$400.

Explanation

The sample standard deviation is the square root of the variance: $(36{,}000{,}000)^{1/2}$ = \$6,000. The standard error of the sample mean is estimated by dividing the standard deviation of the sample by the square root of the sample size: $\sigma_{mean} = s / (n)^{1/2} = 6{,}000 / (81)^{1/2}$ = \$667.

(Study Session 3, Module 11.1, LOS 11.f)

Question #70 of 89

Joseph Lu calculated the average return on equity for a sample of 64 companies. The sample average is 0.14 and the sample standard deviation is 0.16. The standard error of the mean is *closest* to:

A) 0.0025.

B) 0.0200.

C) 0.1600.

Explanation

The standard error of the mean = $\sigma/\sqrt{n} = 0.16/\sqrt{64} = 0.02$.

(Study Session 3, Module 11.1, LOS 11.f)

Question #71 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373

From a sample of 41 orders for an on-line bookseller, the average order size is $75, and the sample standard deviation is $18. Assume the distribution of orders is normal. For which interval can one be exactly 90% confident that the population mean is contained in that interval?

A) $70.27 to $79.73.

B) $71.29 to 78.71.

C) $74.24 to $75.76.

Explanation

If the distribution of the population is *normal*, but we *don't know* the population variance, we can use the Student's t-distribution to construct a confidence interval. Because there are 41 observations, the degrees of freedom are 40. From Student's *t* table, we can determine that the reliability factor for $t_{\alpha/2}$, or $t_{0.05}$, is 1.684. Then the 90% confidence interval is $75.00 ± 1.684($18.00 / √41), or $75.00 ± 1.684 × $2.81 or $75.00 ± $4.73

(Study Session 3, Module 11.2, LOS 11.j)

Question #72 of 89

Which of the following is the *best* method to avoid data mining bias when testing a profitable trading strategy?

A) Test the strategy on a different data set than the one used to develop the rules.

B) Increase the sample size to at least 30 observations per year.

C) Use a sample free of survivorship bias.

Explanation

The *best* way to avoid data mining is to test a potentially profitable trading rule on a data set different than the one you used to develop the rule (out-of-sample data). A larger sample size won't prevent data mining, and you can still data mine a database free of survivorship bias.

(Study Session 3, Module 11.2, LOS 11.k)

Question #73 of 89

A simple random sample is a sample constructed so that:

A) the sample size is random.

B) each element of the population is also an element of the sample.

C) each element of the population has the same probability of being selected as part of the sample.

Explanation

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 (non-zero) likelihood of being included in the sample.

(Study Session 3, Module 11.1, LOS 11.a)

Question #74 of 89

In which one of the following cases is the t-statistic the appropriate one to use in the construction of a confidence interval for the population mean?

A) The distribution is nonnormal, the population variance is known, and the sample size is at least 30.

B) The distribution is nonnormal, the population variance is unknown, and the sample size is at least 30.

C) The distribution is normal, the population variance is known, and the sample size is less than 30.

Explanation

The t-distribution is the theoretically correct distribution to use when constructing a confidence interval for the mean when the distribution is nonnormal and the population variance is unknown but the sample size is at least 30.

(Study Session 3, Module 11.2, LOS 11.j)

Question #75 of 89

A scientist working for a pharmaceutical company tries many models using the same data before reporting the one that shows that the given drug has no serious side effects. The scientist is guilty of:

A) data mining.

B) look-ahead bias.

C) sample selection bias.

Explanation

Data mining is the process where the same data is used with different methods until the desired results are obtained.

(Study Session 3, Module 11.2, LOS 11.k)

Question #76 of 89

Which statement *best* describes the properties of Student's t-distribution? The t-distribution is:

A) symmetrical, and defined by two parameters.

B) symmetrical, and defined by a single parameter.

C) skewed, and defined by a single parameter.

Explanation

The t-distribution is symmetrical like the normal distribution but unlike the normal distribution is defined by a single parameter known as the degrees of freedom.

(Study Session 3, Module 11.2, LOS 11.i)

Question #77 of 89

The confidence interval for a parameter is calculated as:

A) Point Estimate ± Reliability Factor × Standard Error.

B) Point Estimate ± Standard Error.

C) Point Estimate × Reliability Factor ± Standard Error.

Explanation

The confidence interval for a parameter is calculated as Point Estimate ± Reliability Factor × Standard Error. The reliability factor is based on the assumed distribution of the point estimate and the degree of confidence (1 − a) for the confidence interval.

(Study Session 3, Module 11.2, LOS 11.j)

Question #78 of 89

An analyst has compiled stock returns for the first 10 days of the year for a sample of firms and estimated the correlation between these returns and changes in book value for these firms over the just ended year. What objection could be raised to such a correlation being used as a trading strategy?

A) Use of year-end values causes a sample selection bias.

B) The study suffers from look-ahead bias.

C) Use of year-end values causes a time-period bias.

Explanation

The study suffers from look-ahead bias because traders at the beginning of the year would not be able to know the book value changes. Financial statements usually take 60 to 90 days to be completed and released.

(Study Session 3, Module 11.2, LOS 11.k)

Question #79 of 89

Which of the following statements regarding confidence intervals is *most* accurate?

A) The lower the significance level, the wider the confidence interval.

B) The higher the significance level, the wider the confidence interval.

C) The lower the degree of confidence, the wider the confidence interval.

Explanation

A higher degree of confidence requires a wider confidence interval. The degree of confidence is equal to one minus the significance level, and so the wider the confidence interval, the higher the degree of confidence and the lower the significance level.

(Study Session 3, Module 11.2, LOS 11.j)

Question #80 of 89

A local high school basketball team had 18 home games this season and averaged 58 points per game. If we assume that the number of points made in home games is normally distributed, which of the following is *most likely* the range of points for a confidence interval of 90%?

A) 26 to 80.

B) 34 to 82.

C) 24 to 78.

Explanation

This question has a bit of a trick. To answer this question, remember that the mean is at the midpoint of the confidence interval. The correct confidence interval will have a midpoint of 58. (34 + 82) / 2 = 58.

(Study Session 3, Module 11.2, LOS 11.j)

Question #81 of 89

The practice of repeatedly using the same database to search for patterns until one is found is called:

A) sample selection bias.

B) data mining.

C) data snooping.

Explanation

The practice of data mining involves analyzing the same data so as to detect a pattern, which may not replicate in other data sets, also known as torturing the data until it confesses.

(Study Session 3, Module 11.2, LOS 11.k)

Question #82 of 89

Student's *t*-Distribution

Level of Significance for One-Tailed Test						
df	**0.100**	**0.050**	**0.025**	**0.01**	**0.005**	**0.0005**
Level of Significance for Two-Tailed Test						
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690

Books Fast, Inc., prides itself on shipping customer orders quickly. Books Fast sampled 27 of its customers within a 200-mile radius and found a mean delivery time of 76 hours, with a sample standard deviation of 6 hours. Based on this sample and assuming a normal distribution of delivery times, what is the confidence interval for the mean delivery time at 5% significance?

A) 65.75 to 86.25 hours.

B) 68.50 to 83.50 hours.

C) 73.63 to 78.37 hours.

Explanation

The confidence interval is equal to 76 + or − (2.056)(6 / $\sqrt{27}$) = 73.63 to 78.37 hours.

Because the sample size is small, we use the t-distribution with (27 − 1) degrees of freedom.

(Study Session 3, Module 11.2, LOS 11.j)

Question #83 of 89

Which one of the following distributions is described entirely by the degrees of freedom?

A) Student’s t-distribution.

B) Normal distribution.

C) Lognormal distribution.

Explanation

Student's t-distribution is defined by a single parameter known as the degrees of freedom.

(Study Session 3, Module 11.2, LOS 11.i)

Question #84 of 89

A statistical estimator is unbiased if:

A) the variance of its sampling distribution is smaller than that of all other estimators.

B) an increase in sample size decreases the standard error.

C) the expected value of the estimator is equal to the population parameter.

Explanation

Desirable properties of an estimator are unbiasedness, efficiency, and consistency. An estimator is unbiased if its expected value is equal to the population parameter it is estimating. An estimator is efficient if the variance of its sampling distribution is smaller than that of all other unbiased estimators. An estimator is consistent if an increase in sample size decreases the standard error.

(Study Session 3, Module 11.1, LOS 11.g)

Question #85 of 89

An analyst divides the population of U.S. stocks into 10 equally sized sub-samples based on market value of equity. Then he takes a random sample of 50 from each of the 10 sub-samples and pools the data to create a sample of 500. This is an example of:

A) systematic cross-sectional sampling.

B) stratified random sampling.

C) simple random sampling.

Explanation

In stratified random sampling we first divide the population into subgroups, called strata, based on some classification scheme. Then we randomly select a sample from each stratum and pool the results. The size of the samples from each strata is based on the relative size of the strata relative to the population. 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 (non-zero) likelihood of being included in the sample.

(Study Session 3, Module 11.1, LOS 11.c)

Question #86 of 89

Frank Grinder is trying to introduce sampling into the quality control program of an old-line manufacturer. Currently, each item is individually inspected to make sure it meets size tolerances. For all items manufactured during August, the standard deviation of size was 0.02 centimeters. If Grinder takes a sample of 30 items and finds a standard deviation of size of 0.019 centimeters, what is the standard error of the sample mean?

A) 0.00365.

B) 0.00200.

C) 0.00600.

Explanation

If we know the standard deviation of the population (in this case we do), then the standard error of the sample mean = the standard deviation of the population / the square root of the sample size = 0.02 / √30 = 0.00365 centimeters.

(Study Session 3, Module 11.1, LOS 11.f)

Question #87 of 89

An equity analyst needs to select a representative sample of manufacturing stocks. Starting with the population of all publicly traded manufacturing stocks, she classifies each stock into one of the 20 industry groups that form the Index of Industrial Production for the manufacturing industry. She then selects four stocks from each industry. The sampling method the analyst is using is *best* characterized as:

A) stratified random sampling.

B) nonrandom sampling.

C) random sampling.

Explanation

In stratified random sampling, a researcher classifies a population into smaller groups based on one or more characteristics, takes a simple random sample from each subgroup, and pools the results.

(Study Session 3, Module 11.1, LOS 11.c)

Question #88 of 89

The sampling distribution of a statistic is:

A) always a standard normal distribution.

B) the same as the probability distribution of the underlying population.

C) the probability distribution consisting of all possible sample statistics computed from samples of the same size drawn from the same population.

Explanation

A sample statistic itself is a random variable, so it also has a probability distribution. For example, suppose we start with a sample of the prices of 200 stocks, and we calculate the sample mean of a random sample of 40 of those stocks. If we repeat this many times, we will have many different estimates of the sample mean. The distribution of these estimates of the mean is the sampling distribution of the mean. A statistic's sampling distribution is not necessarily normal or the same as that of the population.

(Study Session 3, Module 11.1, LOS 11.a)

Question #89 of 89

The central limit theorem states that, for any distribution, as *n* gets larger, the sampling distribution:

A) becomes larger.

B) approaches the mean.

C) approaches a normal distribution.

Explanation

As *n* gets larger, the variance of the distribution of sample means is reduced, and the distribution of sample means approximates a normal distribution.

(Study Session 3, Module 11.1, LOS 11.e)