

LO.a: Define simple random sampling and a sampling distribution.

1. A descriptive measure of a population is *best* described as:
 - A. an estimator.
 - B. a parameter.
 - C. a sample statistic.

LO.b: Explain sampling error.

2. The difference between the observed value of a statistic and the quantity intended to estimate is known as:
 - A. a parameter.
 - B. a sampling error.
 - C. a systematic error.

LO.c: Distinguish between simple random and stratified random sampling.

3. Paul uses the stratified random sampling technique of categorization for his research project. He first divides the country into 15 regions; then for each region, he categorizes the population into three age groups; young, adults, and elderly. Finally, he divides these categories into males and females. The number of cells or strata the sampling plan entails is *closest* to:
 - A. 20.
 - B. 45.
 - C. 90.
4. Which of the following methods produces precise parameter estimates?
 - A. Simple random sampling.
 - B. Stratified random sampling.
 - C. Indexing.

LO.d: Distinguish between time-series and cross-sectional data.

5. Which of the following is *most likely* to be an example of cross-sectional data?
 - A. Financial ratios of Siemens for the years 2007, 2008, and 2009.
 - B. Financial ratios of Engro, Nestle, and Unilever for the year 2009.
 - C. Financial ratios of Engro, Nestle and Unilever for the years 2007, 2008, and 2009.
6. An analyst compiles EPS data for a sample of companies randomly drawn from the KSE-100 index. He gathers the data from the companies' fiscal year 2013 annual reports. This data is *best* characterized as:
 - A. time-series data.
 - B. longitudinal data.
 - C. cross-sectional data.

7. Julian Spencer is conducting a research for his project. He makes use of the household budget data of the Scandinavian countries for FY2012. This is *most likely* an example of:
- cross sectional data.
 - time series data.
 - panel data.

LO.e: Explain the central limit theorem and its importance.

8. The population distribution of real estate investment returns in a country is found to be negatively skewed in a particular year. Given that the sample size is large, the sampling distribution of the sample mean returns would be characterized by:

	Distribution	Mean	Variance
A	Negative skewed	Less than population mean	Equal to population variance
B	Normal distribution	Equal to population mean	Less than population variance
C	Normal distribution	Less than population mean	Equal to population variance

9. According to the central limit theorem, the sampling distribution of the sample mean will be approximately normal for a large-size sample if the:
- population distribution is normal.
 - population distribution is symmetric.
 - population is described by any probability distribution.
10. An analyst draws samples from a population with distribution that is negatively skewed. Given the sample size is large, the sampling distribution of the sample mean is *most likely*:
- approximately normally distributed.
 - to have a variance equal to that of the entire population.
 - to have a mean smaller than the mean of the entire population.
11. Which of the following statements about central limit theorem are *most likely* correct?
- Statement I: The value of n should exceed 30 for the sample to be considered large.
- Statement II: The central limit theorem states that the variance of the distribution of the sample mean is $\frac{\sigma^2}{(n-1)}$.
- Statement III: The sample mean follows an approximately normal distribution for large sample sizes.
- Statements I and II.
 - Statements I and III.
 - Statements II and III.

LO.f: Calculate and interpret the standard error of the sample mean.

12. From the following information, the standard error of the sample mean is *closest* to:

Mean	33
Number of observations	82
Variance	53

- A. 0.63.
B. 0.75.
C. 0.80.
13. A sample of 40 observations has a mean of 5 and a standard deviation of 22. The standard error of the sample mean is *closest* to:
A. 0.55.
B. 0.80.
C. 3.48.
14. Following sample of 10 items is drawn from the population:
2, 5, -5, 12, 9, 11, 0, 1, 20, -7
If the population variance is unknown, the standard error of the sample mean is *closest* to:
A. 2.5.
B. 2.6.
C. 2.8.
15. Ali manages a portfolio of 250 stocks. He draws a sample of 50 stocks, which has a standard deviation of returns of 15% and a mean annual return of 20%. Standard error of the sample mean is *closest* to:
A. 0.9500%.
B. 0.0212%.
C. 2.1200%.
16. From a large population you draw a sample of 64. The sample mean is 12 and the standard deviation is 3. What is the standard error of the sample mean?
A. 0.375.
B. 0.378.
C. 0.667.

LO.g: Identify and describe desirable properties of an estimator.

17. Which of the following is *least likely* a desirable property of an estimator?
A. Consistency.
B. Efficiency.
C. Reliability.
18. An analyst considers a large sample size to estimate a population parameter. A larger sample size is *most likely* to improve the estimator's:
A. consistency.

- B. efficiency.
- C. unbiasedness.

19. As the sample size increases, the probability of an estimate being close to the value of the population parameter increases. This desirable property of estimators is known as:
- A. Consistency.
 - B. Efficiency.
 - C. Unbiasedness.

LO.h: Distinguish between a point estimate and a confidence interval estimate of a population parameter.

20. All else held constant, the width of a confidence interval is *most likely* to be smaller if:
- A. sample size is larger and degree of confidence is higher.
 - B. sample size is larger and degree of confidence is lower.
 - C. sample size is smaller and degree of confidence is lower.
21. A researcher is unsure of the underlying population distribution and wants to increase the reliability of the parameter estimates. Which of the following is *least likely* to achieve her objective?
- A. Increase in the sample size.
 - B. Use of point estimates rather than confidence intervals.
 - C. Use of the t-distribution rather than the normal distribution to establish confidence intervals.
22. Which of the following *most likely* represents the degree of confidence?
- A. α .
 - B. $1-\alpha$.
 - C. $\alpha - 1$.
23. The average dividend yield for a sample of 50 companies is 4.5% and its standard deviation is 3.75%. The 95% confidence interval for this sample is *most likely*:
- A. [3.46, 5.54].
 - B. [3.13, 5.87].
 - C. [3.62, 5.38].

LO.i: Describe properties of Student's t-distribution and calculate and interpret its degrees of freedom.

24. Which of the following is *least accurate* about student's t-distribution in comparison with the normal distribution? Student's t-distribution:
- A. has fatter tails.
 - B. is more peaked.
 - C. is symmetrical.

25. Which of the following would result in the t-distribution to approach a standard normal distribution?
- A. An increase in degrees of freedom.
 - B. An increase in standard deviation.
 - C. Greater probability in the tails.

LO.j: Calculate and interpret a confidence interval for a population mean, given a normal distribution with 1) a known population variance, 2) an unknown population variance, or 3) an unknown variance and a large sample size.

26. A researcher wants to set the confidence intervals for the population mean of an approximately normal distribution. Assuming the sample size is small, the student's t-distribution is the preferred method when the variance is:
- A. unknown.
 - B. too large.
 - C. negative.
27. An analyst wants to predict the return on the KSE Index for the current year using the following data and assumptions:
- Sample size = 60 stocks
 - Mean return for those stocks in the sample for the previous year = 15%
 - Sample variance = 5%
 - Confidence interval = 99%
 - Population variance is unknown
- If she assumes that the KSE return this year will be the same as it was last year, which of the following is the *best* estimate of the 99% confidence interval for this year's KSE return?
- A. 0.16665 to 0.13335.
 - B. 0.09345 to 0.20656.
 - C. 0.07454 to 0.22452.
28. A researcher wants to set the confidence interval for the population mean. The distribution is normal, the sample size is small and the variance is unknown. It is *most* appropriate for the researcher to use:
- A. the t-distribution.
 - B. the z-distribution.
 - C. either distribution.
29. Use the following values from Student's t-distribution to establish a 95% confidence interval for the population mean given a sample size of 11, a sample mean of 8.0, and a sample standard deviation of 15. Assume that the population from which the sample is drawn is normally distributed and the population variance is not known. Note that the probabilities given below are one-tailed probabilities.

Degrees of freedom	$p = 0.10$	$p = 0.05$	$p = 0.025$	$p = 0.01$
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718

- A. A lower bound of 4.96 and an upper bound of 11.04.
- B. A lower bound of -1.95 and an upper bound of 17.95.
- C. A lower bound of -2.08 and an upper bound of 18.08.

30. A sample of 125 observations drawn from a normally distributed population has a sample mean of 8 and a sample standard deviation of 12. Using the extract from the z -distribution given below, find the 95% confidence interval for the population mean:

Cumulative Probabilities for a Standard Normal Distribution $P(Z = x) = N(x)$ for $x = 0$ or $P(Z = z) = N(z)$ for $z = 0$			
x or z	0.04	0.05	0.06
1.6	0.9495	0.9505	0.9515
1.7	0.9591	0.9599	0.9608
1.8	0.9671	0.9678	0.9686
1.9	0.9738	0.9744	0.9750

The 95% confidence interval is *closest* to:

- A. 5.90 and 10.10.
- B. 6.23 and 9.77.
- C. 7.80 and 8.20.

31. A sample of 74 observations drawn from a normally distributed population has a sample mean of 11 and a sample standard deviation of 19. The extract from the z -distribution is given below.

Cumulative Probabilities for a Standard Normal Distribution $P(Z = x) = N(x)$ for $x = 0$ or $P(Z = z) = N(z)$ for $z = 0$			
x or z	0.04	0.05	0.06
1.6	0.9495	0.9505	0.9515
1.7	0.9591	0.9599	0.9608
1.8	0.9671	0.9678	0.9686
1.9	0.9738	0.9744	0.9750

The 95% confidence interval is *closest* to:

- A. 5.90 and 16.10.
- B. 6.67 and 15.33.
- C. 7.37 and 14.63.

32. Assume a population with a non-normal distribution and known variance, and a large sample size. The statistic *most likely* used to estimate the mean of this population is:

- A. F.
- B. t.
- C. z.

33. In the confidence interval equation given below, $z_{\alpha/2}$ is known as the:

$$\text{Confidence Interval} = \mu \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

- A. Point estimate.
- B. Reliability factor.
- C. Standard error.

LO.k: Describe the issues regarding selection of the appropriate sample size, data-mining bias, sample selection bias, survivorship bias, look-ahead bias, and time-period bias.

34. Extensive searching through a dataset for statistically significant patterns is known as:

- A. data mining.
- B. data diddling.
- C. hypothesis testing.

35. Which of the following biases is most prevalent when evaluating mutual funds?

- A. Look ahead bias.
- B. Survivorship bias.
- C. Time period bias.

Solutions

1. B is correct. A descriptive measure of a population is best described as a parameter.
2. B is correct. The difference between the observed value of a statistic and the quantity intended to estimate is the sampling error.
3. C is correct. This is an application of the multiplication rule of counting. The total number of cells is the product of 15, 3, and 2, which is 90.
4. B is correct. Stratified random sampling produces more precise parameter estimates than simple random sampling. Indexing is an investment strategy.
5. B is correct. Option B is an example of cross-sectional data which can be defined as data on some characteristic of individuals or groups at a single point in time. Option A is an example of time-series data. Option C is an example of longitudinal data.
6. C is correct. Data on some characteristics of companies at a single point in time are cross-sectional data.
7. A is correct. Cross sectional data are observations that represent individuals, groups, geographical regions, or companies at a single point in time. So, this is an example of cross sectional data.
8. B is correct. According to the central limit theorem, the sampling distribution of sample means is approximately a normal distribution with its mean is equal to the population mean. Its variance is equal to population variance divided by the sample size and hence would be less than population variance. The Central Limit Theorem is independent of the probability distribution of the given population when the sample size is large and thus the skewness of population does not matter.
9. C is correct. Based on the central limit theorem, when the sample size is large, we can compute confidence intervals for the population mean based on the normal distribution regardless of the distribution of the underlying population.
10. A is correct. The central limit theorem establishes that the sampling distribution of sample means will be approximately normal, will have a mean equal to the population mean, and will have a variance equal to the population variance divided by the sample size.
11. B is correct. Statement II is incorrect. This is because the central limit theorem states that the variance of the distribution of the sample mean is $\frac{\sigma^2}{n}$.

12. C is correct. The standard error of the sample mean is the sample standard deviation (or the population standard deviation if known) divided by the square root of the sample size. In this case, the standard error of the sample mean = $\sqrt{53}/\sqrt{82} = 0.80$.
13. C is correct. The standard error of the sample mean, when we know the sample standard deviation, is sample standard deviation divided by square root of sample size:
 $22/\sqrt{40} = 3.478505$.
14. B is correct. The sample mean is 4.8. The sample variance is 68.8 and sample standard deviation is 8.3. If the population variance is unknown the standard error of sample mean is calculated as sample standard deviation divided by square root of sample size. $8.3/\sqrt{10} = 2.6$
15. C is correct. The standard error of the mean is $s = \frac{\sigma}{\sqrt{n}}$, hence $s = 0.15/ (50)^{0.5} = 0.0212 = 2.12\%$.
16. A is correct. Standard Error = $\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{64}} = 0.375$.
17. C is correct. The desirable properties of an estimator are unbiasedness, consistency and efficiency.
18. A is correct. Unbiasedness and efficiency are properties of an estimator's sampling distribution that hold for any size sample. A consistent estimator is one for which the probability of estimates close to the value of the population parameter increases as sample size increases.
19. A is correct. The property defined by the statement in the question is that of consistency.
20. B is correct. As we increase the degree of confidence, the confidence interval becomes wider. A larger sample size decreases the width of a confidence interval.
21. B is correct. Point estimates are not as reliable as confidence interval estimates. A conservative approach to confidence intervals relies on the t -distribution rather than the normal distribution, and use of the t -distribution will increase the reliability of the confidence interval. Increasing the sample size increases the reliability of the confidence intervals.
22. B is correct. The degree of confidence is $1 - \alpha$ for a confidence interval.
23. A is correct. Confidence Interval = $\mu \pm z \left(\frac{\sigma}{\sqrt{n}} \right) = 4.5 \pm 1.96 \left(\frac{3.75}{\sqrt{50}} \right) = [3.46, 5.54]$
24. B is correct. The Student's t -distribution has fatter tails and is less peaked compared to the normal distribution.

25. A is correct. As the degrees of freedom increases, the tails are no longer fat, and the t-distribution approaches a standard normal distribution.
26. A is correct. The Student's t-distribution is the preferred option when the sample size is small and the variance is unknown.
27. C is correct. Confidence interval = $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$. Here, $s = \sqrt{0.05} = 0.2236$, $n = 60$ and $\bar{X} = 0.15$. Since n is large we can use the z-table. For a 99% confidence level the z-value is 2.58. The estimated interval is $0.15 \pm 2.58 * (0.2236 / \sqrt{60})$ or 0.07454 to 0.22452.
28. A is correct. When the sample size is small, the Student's t-distribution is preferred if the variance is unknown.
29. C is correct. With a sample size of 11, there are 10 degrees of freedom. The confidence interval concept is based on a two-tailed approach. For a 95% confidence interval, 2.5% of the distribution will be in each tail. Thus, the correct t-statistic to use is 2.228.
 $8.0 \pm 2.228 \times 15 / \sqrt{11}$
 $8.0 \pm 10.08 = -2.08$ to 18.08 .
30. A is correct. The 95% confidence interval uses $z_{0.025}$ as the reliability factor. The cumulative probability value closest to 0.975 provides the appropriate value of $z_{0.025}$ which is $z_{0.025} = 1.96$. The confidence interval is:
 $8 \pm 1.96 \times (12 / \sqrt{125}) = 5.90$ and 10.10 .
31. B is correct. The 95% confidence interval uses $z_{0.025}$ as the reliability factor. The cumulative probability value closest to 0.975 provides the appropriate value of $z_{0.025}$ which is $z_{0.025} = 1.96$. The confidence interval is:
 $11 \pm 1.96 \times (19 / \sqrt{74}) = 6.67$ and 15.33 .
32. C is correct. z-test statistic is most likely used for a large sample size with non-normal distribution, and known variance.
33. B is correct. Confidence Interval = Point estimate \pm Reliability factor (Standard error)
Hence, it is the reliability factor.
34. A is correct. Extensive searching through a dataset for statistically significant patterns is known as data mining.
35. B is correct. The mutual funds that do not perform well or are no longer in business are often not found in the dataset. As a result, a sample often only consists of surviving funds leading to survivorship bias.