Learning Module 7: Estimation and Inference

Q.422 Which statistic is *most likely* to be used for a limited normal sample size (less than 30) with an unknown variance?

A. t-test.

B. z-test.

C. The sample size is too small to make any kind of judgment.

The correct answer is **A.**

A t-test is used if you do not know the population's standard deviation and have a limited sample (less than 30). For a sample size larger than 30, you can use the z-test.
B is incorrect. The z-test is used when the population is normally distributed. The z test can also test when the population is substantial, thanks to the central limit theorem.

C is incorrect. For a sample less than 30 with unknown variance, a t-test is appropriate.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean

Q.424 A distribution has a mean of 11% and a standard deviation of 20%. The interval of this distribution using a 95% confidence interval is *closest to*;

      A. -9% to 31%.

      B. -28.2% to 50.2%.

      C. -21.9% to 43.9%.

The correct answer is **B.**

Based on the information given in the question the confidence interval is given by:

$$11\% \pm 20(1.96) = 11\% \pm 39.20 = -28.2\% \text{ to } 50.2\%$$

The value of 1.96 is based on the fact that 95% of the area of a normal distribution is within 1.96 standard deviations of the mean.

Interpretation of the result: 95% of all observations lie between -28.2% and 50.2%.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.993 The *most likely* outcome of an increase in sample size is:

    A. the sample standard deviation increases.

    B. the sample mean gets closer to the population mean.

    C. the sample mean varies from the population mean to a large extent.

The correct answer is **B.**

The standard error of the sample mean is used by analysts to estimate the precision of the sample mean to the population mean. A lower value of the standard more precise estimate.

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

To answer this question, we will use a constant σ (3); we will then vary n (10 to represent a smaller sample and 30 a larger sample) $\sigma_x = \frac{3}{\sqrt{10}} = 0.948 \, \sigma_x = \frac{3}{\sqrt{30}} = 0.548$. As seen above, a larger n (30) size gives a lower SE (0.548) A smaller n (10) gives a higher SE (0.948).

The larger the sample, the more accurate the sample mean, as we consider a larger number of outcomes.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.996 As the degrees of freedom increases and the t-statistic approaches towards the z-statistic, the tails of the distribution *most likely* become:

      A. thicker.

      B. thinner.

      C. unchanged.

The correct answer is **B.**

The degrees of freedom (number of observations-1) of a t distribution determine its peakedness. The tails of the t-distribution are usually thicker than the normal distribution, so when the number of observations increases and the degrees of freedom increase, the t-distribution becomes more spiked/peaked with thinner tails.

A is incorrect. The tails of the t-distribution is usually thicker than the normal distribution.

C is incorrect. There is an expected change on the tails of the t-distribution with an increase in the degrees of freedom.

Q.1877 The *most accurate* definition of a sample error is the:

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

      B. difference between a sample statistic and its corresponding population parameter.

      C. difference between the point estimate of the mean and the mean of the sampling distribution.

The correct answer is **B.**

Sampling error is the difference between any sample statistic and its corresponding population parameter. An example of sampling error would be the difference between the sample mean $\bar{X}$ and the population mean, μ.

A is incorrect. It is the definition of random sampling and not a sampling error.

C is incorrect. It is the definition of standard error, and not a sampling error.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.1879 The Central Limit Theorem is *most appropriately* concerned with the sampling distribution of the:

      A. sample mean.

      B. population mean.

      C. sample standard deviation.

The correct answer is **B.**

The central limit theorem states that the sample mean X of random samples of sizes n drawn from a population with mean μ and variance σ^2 will have an approximately normal distribution with mean μ and var $\frac{\sigma^2}{n}$ as the sample size n becomes large.

A is incorrect. The sampling distribution is majorly concerned with the sample mean and not the Central Limit Theorem.

C is incorrect. The sample standard deviation is concerned with the value of the sample variance.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.1884 The following information is available on a sample of advertising budgets taken from 81 U.S companies:

- The mean advertising budget is 10 million.
- The sample variance is 36 million.

The standard error of the sample mean is *closest to*:

      A. 0.667.

      B. 1,667.

      C. 11,384

The correct answer is **A.**

The standard error of the sample mean ($S\bar{X}$) can be calculated using the formula:

$$S_{\bar{X}} = \frac{s}{\sqrt{n}}$$

Where:

s is the sample standard deviation.

n is the sample size.

The sample standard deviation is the square root of the variance $= \sqrt{36 \text{ million}} = 6$ million.
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 $= \frac{6}{\sqrt{81}} = 0.667$.

B is incorrect. Results from the mean advertising budget divided by the sample standard deviation of the mean square root.

C is incorrect. The result of dividing the sample variance by the sample standard deviation is the square root of the variance.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.2752 Which of the following is the *most appropriate* example of a simple random sample?

    A. An analyst randomly selects AAA-rated corporate bonds as a sample to estimate the default risk of the U.S. Bond market.

    B. An analyst randomly selects large-cap index stocks from the Mumbai Stock Exchange to measure the holding period return of Indian stocks.

    C. An analyst randomly selects the GDP growth rate of five random countries from the European Union to measure the GDP growth of European Union countries.

The correct answer is **C.**

The five European countries are selected randomly (without any specific criteria) to measure the GDP growth rate of European Union countries.

A is incorrect. It is also an example of stratified random sampling. It applies the criteria of selecting only AAA-rated corporate bonds.

B is incorrect. It is an example of stratified random sampling. This is because the analyst only takes Large-cap index stock (Criteria or Strata is applied).

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.2753 A survey team in Srilanka conducted a health survey in a village by dividing the village population into three different age ranges. The first range consisted of the population aged between 0 to 20 years, the second range was of the population above 20 years and below 45 years, and the last range was comprised of all the villagers above 45 years of age. A sample was drawn from each range according to the proportion of that range. The results showed that the healthiest people were between the ages of 0 and 20 years. The *most likely* type of sampling technique used by the survey team is;

      A. simple random sampling.

      B. stratified random sampling.

      C. systematic random sampling.

The correct answer is **B.**

In Stratified Random Sampling, the population is divided into subpopulations based on some criteria or strata (in our example, the criteria were different age ranges). The simple random sample is then drawn from the subpopulation or stratum as per the proportion of that stratum in the total population.

A is incorrect. Simple random sampling does not involve the division of a population into strata. In simple random sampling, all the population elements have an equal chance of being randomly selected.

C is incorrect. In systematic random sampling, the sample is collected by using a pre-determined regular interval, obtained by dividing the population size, N, by the sample size, n.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.2754 Bogotara is a hypothetical country whose bond market is composed of 2,000 corporate and government bonds. An analyst interested in investing in Bagotara's bond market calculated the mean return of 11.09% based on the sample of 200 bonds, while the population mean return of Bagotara's bond market is only 9.89%. Which of the following errors is *most likely* related to the analyst's estimation?

      A. Mean error.

      B. Sampling error.

      C. Standard deviation.

The correct answer is **B.**

In statistics, sampling error is incurred when the statistical characteristics of a population are estimated from a subset, or sample, of that population. Since the sample does not include all members of the population, statistics on the sample, such as means and quantiles, generally differ from the characteristics of the entire population, which are known as parameters.
A is incorrect. The mean error refers to uncertainty in measurement or the difference between the measured and true/correct values.

C is incorrect. The following steps guide the calculation of the standard deviation error;
Step 1: Calculate the mean (Total of all samples divided by the number of samples).
Step 2: Calculate each measurement's deviation from the mean (Mean minus the individual measurement).
Step 3: Square each deviation from the mean. Squared negatives become positive.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.2755 An analyst is given the financial statements of the top five firms in the logistics sector to conduct a cross-sectional analysis. Which of the following methods is *most appropriate* for cross-sectional data analysis?

   A. The analyst should compare the profit margin of each of the five firms within the sector.

   B. The analyst should compare the returns of each of the five firms with their historical returns.

   C. The analyst should use the EPS of the last 5 years of any firm to estimate the growth in the sector.

The correct answer is **A.**

Option A uses the cross-sectional data or the data of peer firms to conduct the analysis, while options B) and C) provide a time series data analysis.
Cross-sectional data is a type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at the same time or without regard to differences in time. On the other hand, time-series data refers to observations made over a period of time at regular intervals.

B and C are incorrect. It also provides for time-series analysis.

Q.2757 Which of the following is *least likely* a property of the central limit theorem?

A. The variance of distributions of the sample mean is calculated as $\frac{\sigma^2}{\sqrt{n}}$.

B. The mean of the population and the mean of all of the distributions of the sample's means are equal.

C. If the sample size is equal to or larger than 30, we can assume that the sample mean is normally distributed.

The correct answer is **A.**

The variance of the central limit theorem is calculated as $\frac{\sigma^2}{n}$ and not $\frac{\sigma^2}{\sqrt{n}}$.

B is incorrect. The mean of the sample means is equal to the population mean.

C is incorrect. It is an appropriate property of the central limit theorem. If the sample size is equal to or larger than 30, we can assume that the sample mean is normally distributed.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.2758 Which of the following is the *most appropriate* definition of the standard error of the sample mean?

    A. The variance of the distribution of the sample mean.

    B. The standard deviation of the distribution of the sample means.

    C. Difference between the population mean and the sample mean.

The correct answer is **B.**

The standard error of the sample mean is the standard deviation of the distribution of the sample means. It is a measure of how spread out the means of different samples are likely to be from the population mean.

It is calculated as the standard deviation of the population divided by the square root of the sample size.

This is why it's called the standard "error" – it tells us where the mean of our sample is likely to fall in relation to the true population mean

A is incorrect. The standard error of the sample mean is the standard deviation, not the variance, of the sample, mean distribution.

C is incorrect. The difference between the population mean and the sample mean is the sampling error, not the standard error.

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.2759 The mean return on the stocks of automotive companies is $26.5, while the sample standard deviation of 36 automotive companies is $3.1. The standard error of the sample mean is *closest to*:

    A. $0.52

    B. $0.60

    C. $0.74

The correct answer is **A.**

The standard error of the sample mean ($\sigma_{\bar{x}}$) can be calculated using the formula: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$\text{Standard Error} = \frac{\text{Standard deviation of the sample mean}}{\sqrt{\text{Sample size}}} = \frac{\$3.1}{\sqrt{36}} = \frac{3.1}{6} = \$0.5166667 \approx \$0.52$$

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.2761 The mean of a population of 1,000 observations is 61. If the mean and the variance of a sample of 225 observations are 49 and 25 respectively, then the standard error of the sample mean is *closest to*:

    A. 0.33.

    B. 0.46

    C. 0.71

The correct answer is **A.**

If the variance of the sample is 25 ($s^2 = 25$), then the standard deviation s = 5.

$$\text{The standard error of the sample mean of 225 observation} = \frac{\text{Standard Deviation of the sample mean}}{\sqrt{\text{Sample size}}} =$$

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.2767 Which of the following is *least likely* an example of out-of-sample testing?

A. An analyst developed a model to measure the risk of small-cap equity stocks. The model is then applied to large-cap stocks to measure their significance level.

B. An analyst developed a return model based on the S&P 500 daily index from January 1st, 2015 to March 30th, 2015. He then applies the model on data collected between April 1st, 2015, and May 15th, 2015 to test its significance.

C. An analyst developed a model to forecast bond yield variations based on weekly yield variations in 10-year U.S. bonds for the first quarter of 2009 to the fourth quarter of 2011. To test its significance, the model was then applied to yield variations from the third quarter of 2010 to the second quarter of 2011.

The correct answer is **C.**

Out-of-sample testing is defined as using the developed model based on in-the-sample data and tested on out-of-sample data to test its statistical significance. It tests the model on in-the-sample data. The sample data is data that is available at the time of building the model. The sample data is unknown when building the model, and the model should forecast that.

A is incorrect. It's out-of-sample testing since the analyst applies the model on large-cap stocks, yet it was developed to measure the risk of small-cap equity..

B is incorrect. It's also out-of-sample testing since the range of data collection to be used on the model data differs from that of the daily designed index..

Q.3468 A researcher wants to conduct a study to determine the level of literacy in his country. He randomly selects a few universities and conducts a survey among the school students. With respect to his study, the sampling can be *best described* as:

  A. systematic sampling.

  B. simple random sampling.

  C. stratified random sampling.

The correct answer is **C.**

The study aims to determine the level of literacy in the country. However, the survey is conducted among university students. Thus, the sampling can be best defined as stratified random sampling.

A is incorrect. In Systematic random sampling, a sample is selected using a pre-determined regular interval known as k. K is obtained by dividing the population size N by the sample size n.

B is incorrect. In simple random sampling, a sample is obtained by randomly picking out elements from a population without first forming groups of elements with similar characteristics from within the population. Each element in simple random sampling has an equal chance of being selected.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: Compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.3469 Which of the following is the *most accurate* statement regarding the central limit theorem? From a population with mean and a finite variance:

 A. the sample mean approaches a normal distribution for systematic random samples of size *n* for *n* < 30.

 B. the sample mean approaches a normal distribution for systematic random samples of size *n* as *n* becomes larger.

 C. the sample mean approaches a normal distribution for simple random samples of size *n* as *n* becomes larger.

The correct answer is **C.**

The central limit theorem states that, for simple random samples of size n from a population with mean μ and finite variance σ^2, the sample mean approximately has a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$ as n becomes large. The theorem holds for n greater than or equal to 30.
CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.3470 Large random samples of size *n* are repeatedly taken from a large population. Thereafter, the mean of the random samples is taken to create a distribution. If the population mean is μ , then the mean of the distribution created by computing the mean of the random sample will *most likely* be:

    A. nμ

    B. $\frac{\mu^2}{n}$

    C. μ

The correct answer is **C.**

According to the central limit theorem, the population for a simple random sample of size n from a population with mean and finite variances, the sample mean approaches a normal distribution as n becomes larger. The mean of the random samples taken is equal to the mean of the population.

We can also tackle this question from the perspective of the qualities of a good point estimator. A good point estimator (in this case, the sample mean) must be unbiased. An estimator is unbiased if its expected value is equal to its true value (in this case, if the true value is μ then the expected value must also be μ)

A good point estimator must also be consistent. If an estimator is consistent, as the sample size increases, the probability that the estimator "closes in" on the actual value of the parameter approaches 1. Since the sample is already large in this particular question, as per the consistent quality of a good estimator, the sample mean must be equal to the population mean (which has been given as x).

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.3475 An equity research analyst wants to create an index that contains an equal representation of each segment of the market. The *most appropriate* sampling method to create the index is:

    A. stratified sampling.

    B. systematic sampling.

    C. simple random sampling.

The correct answer is **A.**

Stratified sampling divides stocks with similar properties in a group. Therefore, an index created through stratified sampling will have an equal representation of each segment.

B is incorrect. In Systematic random sampling, a sample is selected using a pre-determined regular interval known as k. K is obtained by dividing the population size N by the sample size n.

C is incorrect. In simple random sampling, a sample is obtained by randomly picking out elements from a population without first forming groups of elements with similar characteristics from within the population. Each element in simple random sampling has an equal chance of being selected.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: Compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Q.3737 The mean hourly wage for coal workers in the U.S. is $15.5 with a population standard deviation of $3.2. If the sample size is 30, the standard error of the sample mean is *closest to*:

    A. 0.206

    B. 0.584

    C. 0.813

The correct answer is **B.**

Since the standard deviation for the population is known,

$$\text{Standard error of the mean} = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{30}} = 0.5842$$

Interpretation: If we were to take a number of samples of size 30 from the US coal workers population and proceed to prepare a sampling distribution of the sample means, the means of the various samples would vary from each other by 0.5842. The standard error gives analysts an idea of how precisely the sample mean estimates the population mean. The smaller the sample mean, the more precise the estimate. The standard error in this question is small; this implies that this particular sample has almost accurately estimated the population mean. The mean hourly wage for the population of coal workers from which this sample was obtained is, therefore, likely to be around $15.5

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.3744 A population has a known mean of 100. Suppose 36 samples are randomly drawn from this population with replacement. The observed mean is 97.8 and the standard deviation is 10. The standard error of the sample mean is *closest to*:

    A. 0.360

    B. 1.011

    C. 1.667

The correct answer is **C.**

The standard error of the sample mean is;

$$\text{Standard error of the sample mean} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{36}} = 1.667$$

Interpretation: The means of the 36 samples randomly drawn from the population vary from each other by 1.6667. The standard error gives analysts an idea of how precisely the sample mean estimates the population mean. The smaller the standard error, the better. In this case, the sample mean is not so small. As a result, the sample might not be a very good estimator of the population mean.

A is incorrect. It assumes the following calculation:

$$\text{Standard error of the sample mean} = \frac{36}{10} = 0.36$$

B is incorrect. It assumes the following calculation;

$$\text{Standard error of the sample mean} = \frac{\sigma}{\sqrt{\text{observed} - \text{mean}}} = \frac{10}{\sqrt{97.8}} = 1.011$$

CFA Level I, Volume 1, Topic 3 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

Q.4024 XYZ Associates, a qualified audit firm, wants to audit the books of accounts for ABC Ltd for the financial year 2021. During their fieldwork, XYZ discovered that ABC Ltd had a significant number of financial transactions in their books. Due to time constraints, XYZ couldn't audit all the transactions. To manage this, XYZ Audit Firm employed a sampling technique that allowed them to make a practical and informed selection of transactions to audit, acknowledging that the technique might introduce some bias but still provide valuable insights.

      A. Cluster sampling.

      B. Judgemental sampling.

      C. Convenience sampling.

The correct answer is **B.**

Judgemental sampling refers to the process of handpicking a few elements from a population built on professional knowledge. Thus XYZ Associates will employ the judgmental sampling technique in sampling the financial transactions of ABC Limited because of time constraints. The downside of using this sampling technique is that XYZ Associates may be biased hence yielding skewed results.

A is incorrect. Cluster sampling involves classifying the population into subgroups known as clusters, and then certain clusters are selected as a whole through a simple random sampling technique. Thus, XYZ Associates cannot use this sampling technique in their audit process.

C is incorrect. A population element is selected during convenience sampling based on how easily a researcher can access the element. Note that samples are selected conveniently in this method, so they may not necessarily represent the whole population, thus compromising the sampling accuracy.

CFA Level I, Volume 1, Topic 2 - Quantitative Methods, Learning Module 7, Estimation and Inference, LOS 7a: Compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.