ICS Part-II Statistics — Punjab Board

CHAPTER 11

Sampling Techniques and Sampling Distributions

Statistics • 2nd Year • ICS • Punjab Board • Free PDF Notes

Chapter 11: Sampling Techniques and Sampling Distributions

Almost nothing in real-world statistics is based on studying an entire population -- it is far too slow, expensive, or simply impossible. Instead, statisticians study a carefully chosen sample and use it to draw conclusions about the whole population. This chapter builds the foundation for that entire process: how populations and samples are defined, how a representative sample is actually selected (probability vs non-probability methods, simple random sampling, stratified sampling), and what kinds of errors can creep in along the way.

The second and more theoretical half of the chapter introduces sampling distributions -- the probability distribution of a statistic (like the sample mean or sample proportion) across all possible samples of a given size. This concept, and especially the Central Limit Theorem introduced here, is the single most important theoretical bridge in the entire course: it is what allows the normal distribution from Chapter 10 to be used for inference about sample means and proportions in every chapter that follows (Estimation, Hypothesis Testing).

Learning Objectives

- Define population, sample, sampling, parameter, and statistic, and distinguish clearly between each pair
- Distinguish probability sampling from non-probability sampling and describe common methods of each
- Explain and apply simple random sampling with and without replacement, including counting the number of possible samples
- Describe stratified sampling and explain when it is preferred over simple random sampling
- Distinguish sampling error from non-sampling error, and define accuracy, precision, and bias
- Explain what a sampling distribution of a statistic is and define its standard error

- State and apply the properties of the sampling distribution of the sample mean, with and without the finite population correction
- State the Central Limit Theorem and explain its importance in statistical inference
- Find the mean and standard error of the sampling distributions of the difference of two means, sample proportion, difference of two proportions, and sample variance

Key Concepts

11.1-11.2 Population and Sample

A population is the totality of all observations (or all sampling units) possessing some common, specified characteristic that a researcher wants to study -- it may be finite (a limited, countable number of units, such as all registered voters in a city) or infinite (an unlimited number of units, such as all possible outcomes of tossing a coin indefinitely, or all items that could ever be produced by an ongoing production process).

A sample is a part of the population selected in such a way that it is expected to represent the characteristics of the whole population reasonably well. Sampling is the procedure used to select such a representative sample. A sample survey studies only a part of the population, whereas a census (complete enumeration) studies every single unit. Samples are generally preferred over a census because they save time, cost far less, are often more feasible for large or inaccessible populations, and -- perhaps counter-intuitively -- can sometimes give MORE accurate results than a census, since a smaller, well-managed data collection effort allows more careful measurement and closer supervision, reducing non-sampling errors. The limitations of sampling are that it can never give perfectly exact results (only estimates) and requires careful, expert planning to avoid a biased or unrepresentative sample.

11.3 Sampling Design

A sampling design is the complete plan for selecting a sample from a given population. Its building blocks are the sampling unit (the individual element, or group of elements, that will actually be selected, e.g., a household, a student, a

factory) and the sampling frame -- a complete list of all sampling units in the population, from which the sample is actually drawn (e.g., a voter list, a list of registered students). A good sampling frame should be accurate, complete, and up to date; a poor frame is one of the most common sources of non-sampling error.

Sampling methods fall into two broad families: non-probability sampling, in which the chance of any particular element being included in the sample cannot be determined in advance (examples include convenience sampling, judgment/ purposive sampling, and quota sampling) -- these are quick and cheap but their results cannot be generalized to the population with a known level of confidence; and probability sampling (also called random sampling), in which every element of the population has a known, non-zero probability of being selected. Probability sampling is the standard for scientific inference because it allows the sampling error to be objectively measured and controlled. Selection may be done with replacement (a selected unit is returned to the population and can be chosen again) or without replacement (a selected unit is removed and cannot be chosen again).

11.6 Simple Random Sampling

Simple random sampling (SRS) is the most basic probability sampling method: a sample of size n is drawn from a population of size N such that every possible sample of that size has an equal chance of being selected. It can be carried out physically using the lottery method (writing all N units on identical slips, mixing thoroughly, and drawing n of them) or, more commonly and more rigorously, using a table of random numbers (or a random-number generator), where units of the population are numbered and n numbers are read off the table according to a fixed, unbiased rule.

Theorem 11.1 gives the counting rule that underlies every sampling-distribution calculation in this chapter: if sampling is done WITH replacement, the number of possible ordered samples of size n from a population of size N is N raised to the power n (N^n) -- because each of the n draws independently has N possible outcomes. If sampling is done WITHOUT replacement, the number of possible ordered samples of size n is N(N-1)(N-2)...(N-n+1), written compactly as NPn (N permute n) -- because the first draw has N possible outcomes, the second has

only N-1 remaining, the third only N-2, and so on, since a selected unit cannot be chosen again. These two counting formulas (N^n and NPn) are the starting point of virtually every worked example on sampling distributions in this chapter -- they tell you exactly how many rows the sampling distribution table will have before you even start building it.

11.7 Stratified Sampling

Stratified sampling is used when the population is heterogeneous but can be divided into non-overlapping, internally homogeneous subgroups called strata (for example, dividing a student population into strata by grade level, or a national population into strata by province). A simple random sample is then drawn independently from each stratum, and the stratum samples are combined into the final overall sample. Stratified sampling is preferred over plain SRS whenever the strata differ meaningfully from each other but units within the same stratum are fairly similar, because it guarantees representation from every subgroup and typically produces estimates with a smaller standard error than an SRS of the same total size drawn from the unstratified population.

11.8 Errors in Sampling

No sample-based estimate is ever perfectly exact; the discrepancy between an estimate and the true population value can arise from two very different sources. Sampling error is the difference E = T - theta between a sample statistic T (such as the sample mean) and the true population parameter theta (such as the population mean) that arises purely because only a sample, and not the entire population, was observed -- it exists even if every single measurement in the sample was made perfectly, and it can generally be reduced by increasing the sample size. Non-sampling errors, by contrast, can occur even in a complete census, and arise from sources such as faulty questionnaire design, non-response, data-entry mistakes, or interviewer bias -- these do NOT necessarily shrink as the sample size grows, and in fact tend to grow more troublesome in very large surveys where quality control becomes harder to maintain.

Two further concepts describe the quality of an estimate. Accuracy refers to how close an estimate is to the true population value -- a highly accurate estimate has very little sampling error. Precision refers to how closely repeated estimates

(from repeated samples of the same size and design) cluster around each other, regardless of whether they are close to the true value -- it is measured by the standard error of the estimator; a small standard error means high precision. Bias is defined as the difference between the expected value of an estimator T and the true parameter, Bias = E(T) - theta; an unbiased estimator has Bias = 0, meaning it is correct 'on average' across all possible samples, even though any single sample estimate may still differ from the true value due to ordinary sampling error.

11.9-11.10 Parameters, Statistics, and the Idea of a Sampling Distribution

A parameter is any descriptive numerical measure calculated from the ENTIRE population -- since the population is fixed, a parameter is a constant with one single true (though often unknown) value. Key parameters include the population total tau = sum of all x-j, the population mean mu = tau/N, the population variance sigma-squared = [sum of (x-j - mu)^2]/N, and the population proportion pi = k/N (where k is the number of elements possessing some attribute).

A statistic is any descriptive numerical measure calculated from a SAMPLE -- because the particular sample drawn will differ from one instance to the next, a statistic is a random variable, taking a different value for each possible sample; it therefore has its own probability distribution. Key statistics include the sample total (sum of x-i), the sample mean x-bar = (sum of x-i)/n, the sample proportion p = x/n, and two versions of sample variance: the divide-by-n version s-squared = [sum of (x-i - x-bar)^2]/n and the divide-by-(n-1) version s-hat-squared = [sum of (x-i - x-bar)^2]/(n-1) -- the two are related by n.s^2 = (n-1).s-hat^2.

The sampling distribution of a statistic is the probability distribution formed by the values of that statistic computed from EVERY possible sample of a given size that could be drawn from the population. The standard error of a statistic is defined as the standard deviation of its sampling distribution -- it measures how much the statistic is expected to vary from one random sample to the next, and it is the single most important number needed for statistical inference (confidence intervals and hypothesis tests, covered in the next two chapters) because it quantifies the precision of an estimate.

11.11-11.12 Sampling Distribution of the Sample Mean and the Central Limit Theorem

The sampling distribution of the sample mean X-bar is the probability distribution of the means of all possible simple random samples of size n drawn from a population with mean mu and variance sigma-squared. Its properties are given by a sequence of theorems. Theorem 11.2: the mean of the sampling distribution of X-bar always equals the population mean, mu-Xbar = E(X-bar) = mu, regardless of sample size or replacement method -- so X-bar is an unbiased estimator of mu. Theorem 11.3: when sampling is done with replacement from an infinite population (or a finite one with replacement), the variance of X-bar is the population variance divided by the sample size, sigma-Xbar-squared = sigma-squared/n, so the standard error is sigma-Xbar = sigma/sqrt(n) -- notice the standard error shrinks as the sample size n grows, which is exactly why larger samples give more precise estimates.

Theorem 11.4 extends this to sampling WITHOUT replacement from a finite population of size N: the variance picks up an extra multiplier called the finite population correction (fpc), sigma-Xbar-squared = (sigma-squared/n).[(N-n)/(N-1)]. The fpc is always less than or equal to 1, so sampling without replacement always gives an equal or SMALLER standard error than sampling with replacement -- and when the sample size n is small relative to the population size N (a small 'sampling fraction' n/N), the fpc is so close to 1 that it can often be safely ignored in practice.

Theorem 11.5 states that if the ORIGINAL population itself is normally distributed, then X-bar is exactly normally distributed with mean mu and variance sigma-squared/n, for ANY sample size (even n=1). But most real populations are not normal -- which is where Theorem 11.6, the Central Limit Theorem (CLT), becomes indispensable: for a LARGE sample size, the sampling distribution of X-bar is approximately normal with mean mu and variance sigma-squared/n, REGARDLESS of the shape of the original population's distribution. The larger the sample size, the better this normal approximation becomes. This is the single most powerful and far-reaching result in this chapter: it means the standardized variable Z = (X-bar - mu)/(sigma/sqrt(n)) can be treated as approximately standard normal even when nothing at all is known about the shape of the

underlying population -- and it is this fact alone that makes the confidence intervals and hypothesis tests of Chapters 12 and 13 possible for virtually any kind of data.

11.13 Sampling Distribution of the Difference Between Two Sample Means

When comparing two populations (for example, comparing the average lifetime of light bulbs from two different manufacturers), we work with the difference X-bar-1 minus X-bar-2 between two INDEPENDENT sample means. Theorem 11.7: the mean of this difference always equals the difference of the population means, mu = mu-1 - mu-2. Theorem 11.8: for independent samples with replacement (or from infinite populations), the variance of the difference is the SUM of the two individual sampling variances, sigma-squared = sigma-1-squared/n-1 + sigma-2-squared/n-2 -- variances of independent random variables add even though we are taking a DIFFERENCE, because Var(A-B) = Var(A) + Var(B) when A and B are independent. Theorem 11.9 adds the finite population correction to each term for sampling without replacement from finite populations, and Theorem 11.10 confirms that if both original populations are normal, the difference X-bar-1 - X-bar-2 is exactly normally distributed (and by the Central Limit Theorem, approximately normal for large samples even if the populations are not normal).

11.14-11.15 Sampling Distribution of Sample Proportion and Its Difference

When the characteristic of interest is qualitative with two outcomes (success/failure, yes/no), the population proportion is pi = k/N and the sample proportion is P = X/n, where X counts the number of 'successes' in the sample. Theorem 11.11: for sampling with replacement (or from an infinite/Bernoulli population), P has mean mu-P = pi and variance sigma-P-squared = pi(1-pi)/n -- in fact the exact sampling distribution of P in this case is directly linked to the binomial distribution from Chapter 9. The shape of the sampling distribution of P is right-skewed when pi < 0.5, left-skewed when pi > 0.5, and symmetric when pi = 0.5, but as n grows large it approaches a normal shape by the Central Limit Theorem. Theorem 11.12 adds the finite population correction for sampling without

replacement, linking the sampling distribution of P in that case to the hypergeometric distribution, also from Chapter 9.

Just as with means, we often compare two proportions from two independent Bernoulli populations. Theorem 11.13: the mean of the difference P1 - P2 equals pi-1 - pi-2. Theorem 11.14: for independent samples with replacement, the variance of the difference is again the sum of the two individual variances, sigma-squared = pi-1(1-pi-1)/n-1 + pi-2(1-pi-2)/n-2. Theorem 11.15 adds the finite population correction to each term for sampling without replacement from finite populations.

11.16-11.17 Other Sampling Distributions: Sample Variance

Besides the mean and the proportion, every other sample statistic -- the sample median, the sample variance, the sample standard deviation -- has its own distinct sampling distribution, and a full specification of any sampling distribution must always state three things together: the population being sampled, the particular statistic being computed, and the sample size being used, since changing any one of these three produces a genuinely different sampling distribution. For the sample variance S-squared = [sum of (X-i - X-bar)^2]/n (computed with the divide-by-n convention), the sampling distribution has the property mu-S-squared = E(S-squared) = [(n-1)/n].sigma-squared -- notice this is slightly LESS than the true population variance sigma-squared, which is exactly why the alternative divide-by-(n-1) version of sample variance is preferred as an unbiased estimator in later chapters on estimation.

Important Definitions

What is a population?: The totality of all observations (sampling units) possessing some common, specified characteristic that a study wants to examine; may be finite or infinite.

What is a sample?: A part of the population selected in such a way that it is expected to represent the characteristics of the whole population.

What is the difference between a parameter and a statistic?: A parameter is a descriptive measure computed from the entire population and is therefore a

fixed constant; a statistic is a descriptive measure computed from a sample and is therefore a random variable that varies from sample to sample.

What is a sampling frame?: A complete list of all the sampling units in a population, from which a sample is actually drawn.

What is the difference between probability and non-probability sampling?: In probability sampling every element of the population has a known, non-zero chance of selection, so results can be generalized with a known confidence level; in non-probability sampling the selection chances cannot be determined, so results cannot be reliably generalized to the population.

What is sampling error?: The difference E = T - theta between a sample statistic and the true population parameter, arising purely because only a sample rather than the whole population was observed; it generally decreases as sample size increases.

What is the difference between accuracy and precision?: Accuracy is how close an estimate is to the true population value; precision is how closely repeated estimates cluster around each other (measured by the standard error), regardless of whether they are close to the true value.

What is a sampling distribution?: The probability distribution formed by the values of a statistic computed from every possible sample of a given size that could be drawn from a population.

What is the standard error of a statistic?: The standard deviation of the sampling distribution of that statistic; it measures how much the statistic is expected to vary from sample to sample.

What is the finite population correction (fpc)?: The factor (N-n)/(N-1) applied to the variance of a sampling distribution when sampling is done without replacement from a finite population; it is always <=1 and can often be ignored when the sample is a small fraction of the population.

State the Central Limit Theorem.: For a large sample size, the sampling distribution of the sample mean X-bar is approximately normal with mean mu and variance sigma-squared/n, regardless of the shape of the original population's distribution; the approximation improves as n increases.

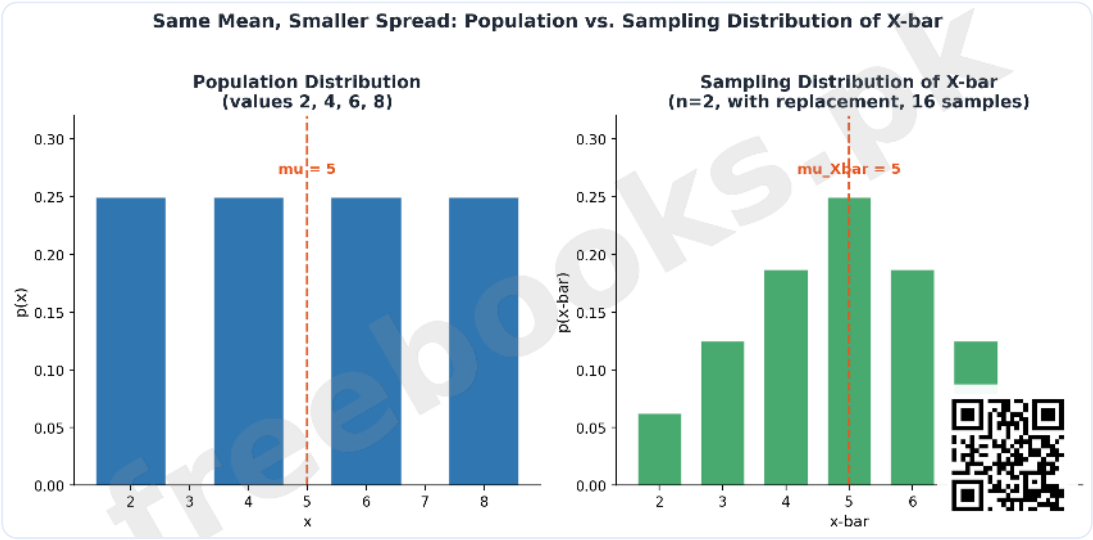
Key Facts and Relations

Topic	Key Fact / Relation
Population parameters	mu = (sum x_j)/N; sigma^2 = (sum x_j^2)/N - mu^2; pi = k/N
Sample statistics	x-bar = (sum x_i)/n; s^2 = [sum(x_i - x-bar)^2]/n; p = x/n
Number of samples (Theorem 11.1)	With replacement: N^n | Without replacement: NPn = N(N-1)...(N-n+1)
Mean of X-bar (Thm 11.2)	mu_Xbar = E(X-bar) = mu
Variance of X-bar, with replacement (Thm 11.3)	sigma_Xbar^2 = sigma^2/n; SE = sigma/sqrt(n)
Variance of X-bar, without replacement (Thm 11.4)	sigma_Xbar^2 = (sigma^2/n).[(N-n)/(N-1)]
Central Limit Theorem (Thm 11.6)	For large n: X-bar ~approx N(mu, sigma^2/n), regardless of population shape
Z for sample mean	Z = (X-bar - mu)/(sigma/sqrt(n))
Difference of means: mean & variance (Thm 11.7-11.8)	mu = mu1-mu2; sigma^2 = sigma1^2/n1 + sigma2^2/n2
Sample proportion: mean & variance (Thm 11.11-11.12)	mu_P = pi; sigma_P^2 = pi(1-pi)/n [x fpc if without replacement]
Difference of proportions: mean & variance (Thm 11.13-11.14)	mu = pi1-pi2; sigma^2 = pi1(1-pi1)/n1 + pi2(1-pi2)/n2
Mean of sample variance (S^2, divide by n)	mu_(S^2) = E(S^2) = [(n-1)/n].sigma^2

Diagrams

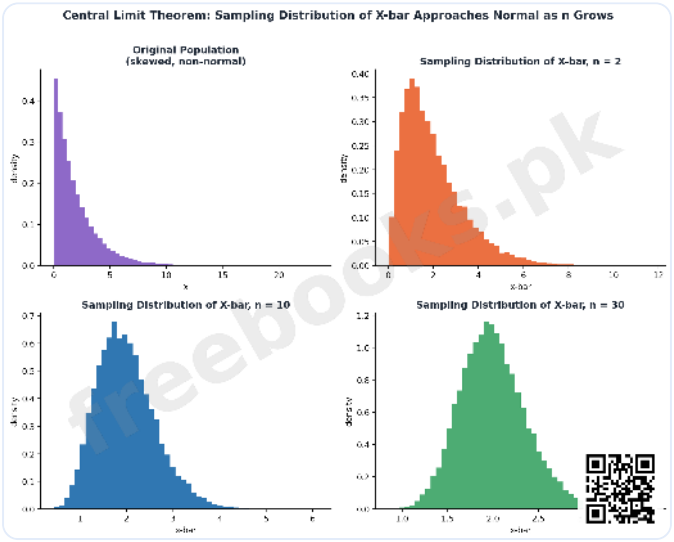
Population Distribution vs. Sampling Distribution of the Mean: A side-by-side comparison of a flat (uniform) population distribution for the values 2,4,6,8 against the mound-shaped sampling distribution of the sample mean for all 16

possible samples of size 2 drawn with replacement -- both centred at the same mean of 5, but the sampling distribution of X-bar is visibly narrower, illustrating Theorem 11.2 (equal means) and Theorem 11.3 (reduced variance)



Population Distribution vs. Sampling Distribution of the Mean

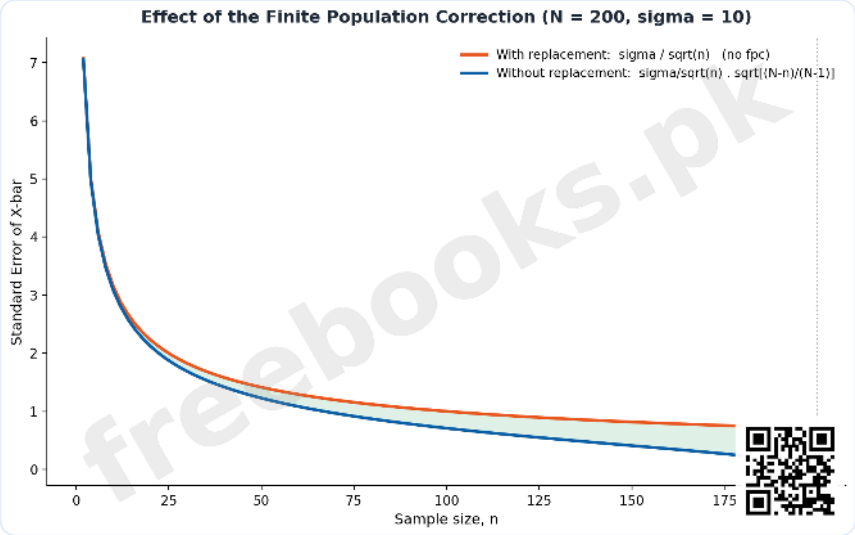
The Central Limit Theorem in Action: Three panels showing a skewed (non-normal) population distribution alongside simulated sampling distributions of the sample mean for increasing sample sizes n=2, n=10, and n=30, illustrating how the sampling distribution of X-bar becomes progressively more bell-shaped and narrower as n grows, regardless of the population's original skewed shape



The Central Limit Theorem in Action

Effect of the Finite Population Correction Factor: A line chart showing the standard error of the sample mean against sample size n for a fixed population

size N, comparing the with-replacement formula sigma/sqrt(n) against the without-replacement formula that includes the finite population correction, showing how the fpc pulls the standard error down further as the sample size approaches the population size



Effect of the Finite Population Correction Factor

Short Questions & Answers

Distinguish between a parameter and a statistic with one example each.

A parameter is a fixed numerical measure from the whole population, e.g., the population mean mu; a statistic is a numerical measure from a sample, and varies from sample to sample, e.g., the sample mean x-bar.

Why is sampling generally preferred over a complete census?

Sampling saves time and cost, is more feasible for large or infinite populations, and -- because a smaller data collection effort allows closer supervision -- can sometimes even be more accurate than a census by reducing non-sampling errors.

What is the key difference between sampling error and non-sampling error?

Sampling error arises only because a sample rather than the whole population was studied and generally shrinks as sample size grows; non-sampling error can occur even in a full census (e.g., from bad questionnaires or data-entry mistakes) and does not necessarily shrink with sample size.

Using Theorem 11.1, how many distinct samples of size n=3 can be drawn without replacement from a population of N=6 units?

NPn = N(N-1)(N-2) = 6(5)(4) = 120 possible samples.

Why does the finite population correction factor make the standard error smaller, not larger?

Because the fpc, (N-n)/(N-1), is always less than or equal to 1 when n < N, so multiplying the with-replacement variance by this factor can only reduce (or leave unchanged) the resulting standard error -- sampling without replacement gives more information per unit sampled.

Why is the Central Limit Theorem so important for statistical inference?

It guarantees that the sampling distribution of the sample mean is approximately normal for large samples regardless of the shape of the original population, which allows normal-distribution-based methods (confidence intervals, hypothesis tests) to be applied to sample means from virtually any kind of population.

Long Questions & Answers

Explain the concept of a sampling design, distinguishing probability from non-probability sampling, and describe simple random sampling and stratified sampling as two concrete probability sampling methods.

Before any statistical analysis of sample data can be trusted, careful thought has to be given to precisely how that sample was actually chosen in the first place, since the method of selection determines whether the resulting sample can be reasonably expected to represent the population it was drawn from. A sampling design is the complete, deliberate plan governing exactly how a sample will be selected from a given population, and it rests on two foundational building blocks. The first is the sampling unit, which is simply the individual element or defined group of elements that will actually be selected during the sampling process -- this might be a single household in a survey of family income, an individual student in a survey of academic performance, or an entire factory in a survey of industrial output, depending entirely on what the study is designed to measure. The second building block is the sampling frame, which is a complete,

itemized list of every single sampling unit that exists within the population under study, and it is from this list that the actual sample will eventually be drawn -- common real-world examples include an electoral register listing every registered voter, or an official enrollment list containing every currently registered student at an institution. The quality of this sampling frame matters enormously in practice, because a frame that is inaccurate, incomplete, or badly out of date is one of the single most common and consequential sources of non-sampling error encountered in real survey work, since units that are missing from the frame have no chance whatsoever of ever being selected into the sample, regardless of how careful the subsequent random selection procedure itself might be. Once a sampling frame has been properly established, the actual sampling methods available for selecting units from it fall broadly into two fundamentally different families. The first family is non-probability sampling, in which the precise probability of any particular individual element ending up included in the final sample simply cannot be calculated or determined in advance -- common examples include convenience sampling, where units are selected purely because they happen to be easy to reach, judgment or purposive sampling, where a researcher's own expert opinion decides which units to include, and quota sampling, where predetermined numbers from various subgroups are filled without genuine randomization within each subgroup. Such non-probability methods are certainly quick to carry out and inexpensive, but because the underlying probability of selection is unknown, the results they produce cannot be legitimately or objectively generalized back to the wider population with any statistically defensible, quantifiable level of confidence. The second, generally far more scientifically rigorous family is probability sampling, also frequently referred to as random sampling, in which every single element within the population is guaranteed to have some known, non-zero probability of ultimately being selected into the sample. This particular property is precisely what allows the resulting sampling error to be objectively measured, quantified, and rigorously controlled through the mathematical machinery of sampling distributions covered later in this very chapter, which is exactly why probability sampling methods form the accepted foundation of essentially all serious scientific and statistical inference. Simple random sampling, commonly abbreviated as SRS, represents the most basic and conceptually fundamental probability sampling method of all: a sample of some particular size n is drawn

from a population of size N in such a careful manner that every conceivable possible sample of that exact size has been given an precisely equal chance of ultimately being the one selected. In practical terms, SRS can be physically carried out through the lottery method, which involves writing every single population unit onto identical, indistinguishable slips of paper, thoroughly mixing the entire collection of slips together, and then drawing out exactly n of them at random; alternatively, and considerably more commonly in modern practice, SRS is instead carried out using a table of random numbers or an appropriate computerized random-number generator, whereby every unit in the population first gets assigned its own unique identifying number, after which n of these numbers are systematically read off according to some fixed, entirely unbiased selection rule. A crucial theoretical result underlying essentially every single sampling-distribution calculation that appears throughout the remainder of this chapter is the precise counting rule for determining exactly how many distinct possible samples actually exist: when sampling is carried out with replacement, the total number of distinct ordered samples of size n is N raised to the power n, conventionally written N^n, since each and every one of the n individual draws independently offers exactly N possible outcomes; but when sampling is instead carried out without replacement, the total count of distinct ordered samples of size n instead becomes N multiplied by (N-1) multiplied by (N-2), continuing on down through (N-n+1), a quantity conventionally denoted as NPn, precisely because each successive draw after the very first one necessarily has one fewer remaining possible outcome available, since previously selected units simply cannot be chosen again a second time. Stratified sampling represents an important and frequently more efficient refinement upon plain, unstructured simple random sampling, and it becomes especially valuable specifically in situations where the overall population being studied is genuinely heterogeneous as a whole, yet can nevertheless be meaningfully divided into a number of distinct, non-overlapping subgroups, formally termed strata, within each of which the individual member units are comparatively far more similar and homogeneous to one another -- a good illustrative example would be dividing an entire national population up into separate strata corresponding to its various different provinces, or alternatively dividing an entire student population up into separate strata according to each student's particular grade level. Under this particular sampling design, an entirely independent simple random sample is

then drawn separately from within each individual stratum, after which all of these individual stratum-level samples are subsequently combined together to form the single final overall sample used for the study. Stratified sampling proves genuinely preferable to an unstratified simple random sample of the exact same total overall size specifically whenever the various strata differ substantially and meaningfully from each other in their underlying characteristics, precisely because this particular sampling design guarantees that every single relevant subgroup of the population will be adequately represented within the final sample, and because it also typically manages to produce final estimates carrying a meaningfully smaller standard error compared to what an equivalently-sized unstratified simple random sample would have produced from that same overall heterogeneous population.

State the Central Limit Theorem in full, and explain its relationship to Theorems 11.2 through 11.5 on the sampling distribution of the sample mean, including why it is considered the single most important theoretical result of this chapter.

The theoretical heart of this entire chapter builds up in a careful, deliberate sequence of theorems, each one adding a further layer of generality on top of the last, and understanding precisely how each theorem in this sequence relates to the ones immediately before and after it is absolutely essential for genuinely grasping why the Central Limit Theorem occupies the uniquely central position that it does in modern statistical theory. Theorem 11.2 establishes the very first and most fundamental building block of the entire sequence: no matter which particular sample size n happens to be used, and entirely regardless of whether the actual sampling itself is conducted with or without replacement, the mean of the sampling distribution of the sample mean X-bar always turns out to equal precisely the true population mean mu itself, formally expressed as mu-Xbar equals E(X-bar) equals mu. This particular property carries an enormously important practical implication: it guarantees that the sample mean X-bar constitutes what statisticians formally term an unbiased estimator of the unknown population mean mu, meaning that even though any single individual sample's calculated mean will typically differ somewhat from the true population mean purely due to ordinary sampling error, the AVERAGE value of X-bar computed across every single conceivable sample that could possibly be drawn

will nevertheless land exactly precisely on the true population mean, with absolutely no systematic bias whatsoever pulling estimates consistently either too high or too low. Theorem 11.3 then proceeds to add crucial additional information specifically regarding the VARIABILITY, rather than merely the central tendency, of this same sampling distribution: specifically for the particular case of sampling with replacement, the variance of the sampling distribution of X-bar works out to equal the original population variance sigma-squared divided by the sample size n, yielding the well known standard error formula sigma-Xbar equals sigma divided by the square root of n. This particular formula carries a critically important practical takeaway of its own: as the sample size n is progressively increased, this standard error correspondingly and continuously shrinks smaller and smaller, which is exactly the precise underlying mathematical reason why larger samples reliably yield noticeably more precise, tightly clustered estimates of the unknown population mean than smaller samples do. Theorem 11.4 then extends this same basic variance result specifically to the alternative case of sampling WITHOUT replacement from a finite population, introducing in the process the finite population correction factor (N-n) divided by (N-1), a quantity that is always mathematically guaranteed to be less than or equal to 1 whenever the sample size n happens to be smaller than the population size N, and this multiplicative correction factor consequently always shrinks the resulting standard error down somewhat further still whenever sampling without replacement is actually employed, compared against what sampling with replacement of that identical sample size would otherwise have produced. Theorem 11.5, meanwhile, takes an important but comparatively more restrictive and narrowly limited step further: specifically IF the original underlying population itself happens to already be normally distributed, then the resulting sampling distribution of X-bar is not merely approximately normal but is instead EXACTLY normally distributed with mean mu and variance sigma-squared over n, and this particular exact result additionally holds true regardless of whatever specific sample size n happens to be used, even including the extreme case of a sample size as small as just a single solitary observation, n equals 1. The single crucial practical limitation of Theorem 11.5, however, is that its validity depends entirely and completely upon the rather strong underlying assumption that the original population being sampled from is already normally distributed in the first place -- and this particular assumption is one that, in the overwhelming majority

of genuinely real-world practical situations encountered by working statisticians, either cannot be verified with any real confidence at all, or else can be confidently verified as being straightforwardly false. This is precisely the exact point at which Theorem 11.6, the Central Limit Theorem itself, becomes so extraordinarily indispensable and powerful within the whole of statistical theory. The Central Limit Theorem states, in its complete and proper full form, that for a sufficiently large sample size, the sampling distribution of the sample mean X-bar is approximately normally distributed with mean mu and variance sigma-squared divided by n, and this remarkable approximate normality result holds true entirely regardless of whatever particular shape or form the original underlying population's own probability distribution actually happens to take -- the original population could be heavily and noticeably skewed, could be flat and uniform in shape, could be distinctly bimodal with two separate peaks, or could take on virtually any other conceivable non-normal shape whatsoever, and yet the resulting sampling distribution of its sample mean will nevertheless still end up looking approximately bell-shaped and normal, provided only that the sample size being used is sufficiently large. Moreover, the overall quality and accuracy of this particular normal approximation itself continues to improve steadily and progressively for larger and larger sample sizes, meaning that even population distributions that are quite dramatically and severely non-normal in their original underlying shape will nevertheless still yield a sampling distribution of X-bar that already looks remarkably close to normal once the sample size employed becomes reasonably large, generally taken in most practical contexts to be somewhere in the vicinity of thirty or more individual observations. The overwhelming practical importance of this single specific result, considered across the whole of statistics as a discipline, is genuinely difficult to overstate. Because the Central Limit Theorem removes entirely the otherwise crucially restrictive assumption of Theorem 11.5 that the original population itself must already be normally distributed, it consequently allows the standardized variable Z, defined as the quantity (X-bar minus mu) divided by (sigma divided by the square root of n), to be legitimately treated as approximately following a standard normal distribution for large samples drawn from virtually ANY kind of population distribution whatsoever, however irregular or unusual its original shape might happen to be. It is precisely this single specific fact, considered in isolation, that ultimately makes it genuinely possible to construct valid

confidence intervals for an unknown population mean, and equally to conduct valid hypothesis tests concerning that same unknown population mean, using the exact same standard normal-distribution-based methodology explored in full technical detail back in Chapter 10 -- and this crucial theoretical connection is exactly the specific reason why the Central Limit Theorem is so very widely and consistently regarded, by working statisticians and instructors alike, as constituting the single most important and far-reaching theoretical bridge to be found anywhere within the whole of this particular statistics course, directly connecting the probability theory developed earlier in the book to the practical methods of statistical inference on estimation and hypothesis testing that immediately follow in the two chapters yet to come.

Multiple Choice Questions (MCQs)

A parameter differs from a statistic in that a parameter is: (A) A random variable (B) A constant computed from the entire population (C) Always unknown (D) Computed only from a sample

Correct answer: (B) A constant computed from the entire population. A parameter is a fixed, constant value computed from the whole population; a statistic is a random variable computed from a sample.

Using Theorem 11.1, the number of possible samples of size n=2 drawn WITH replacement from a population of N=5 is: (A) 10 (B) 20 (C) 25 (D) 5

Correct answer: (C) 25. With replacement: N^n = 5^2 = 25.

Using Theorem 11.1, the number of possible samples of size n=2 drawn WITHOUT replacement from a population of N=5 is: (A) 25 (B) 20 (C) 10 (D) 5

Correct answer: (B) 20. Without replacement: NPn = N(N-1) = 5(4) = 20.

The finite population correction factor (N-n)/(N-1) is: (A) Always greater than 1 (B) Always equal to 1 (C) Always less than or equal to 1 (when n<N) (D) Always negative

Correct answer: (C) Always less than or equal to 1 (when n<N). Since n < N in practical sampling, (N-n)/(N-1) is always <= 1, which is why sampling without replacement reduces the standard error.

Sampling error, unlike non-sampling error, arises: (A) Only in a complete census (B) Purely because a sample rather than the whole population was studied (C) From data-entry mistakes (D) From poorly worded questions

Correct answer: (B) Purely because a sample rather than the whole population was studied. Sampling error exists purely because only a sample (not the full population) is observed, and generally decreases as sample size increases.

Precision of an estimator is measured by its: (A) Bias (B) Mean (C) Standard error (D) Population size

Correct answer: (C) Standard error. Precision refers to how tightly repeated estimates cluster together, measured by the standard error of the estimator.

By Theorem 11.2, the mean of the sampling distribution of X-bar equals: (A) sigma^2/n (B) mu (C) 0 (D) sigma

Correct answer: (B) mu. Theorem 11.2 states mu_Xbar = E(X-bar) = mu, regardless of sample size or replacement method.

The Central Limit Theorem states that for a large sample size, the sampling distribution of X-bar is approximately normal: (A) Only if the population is already normal (B) Regardless of the shape of the population distribution (C) Only for infinite populations (D) Only when sampling without replacement

Correct answer: (B) Regardless of the shape of the population distribution. The power of the CLT is precisely that it holds for large n regardless of the population's original shape.

For sampling with replacement, the variance of the sampling distribution of the difference between two independent sample means X1-bar - X2-bar is: (A) sigma1^2/n1 - sigma2^2/n2 (B) sigma1^2/n1 + sigma2^2/n2 (C) (sigma1^2+sigma2^2)/(n1+n2) (D) sigma1.sigma2/(n1.n2)

Correct answer: (B) sigma1^2/n1 + sigma2^2/n2. Variances of independent random variables ADD even when taking a difference: Var(X1-bar - X2-bar) = sigma1^2/n1 + sigma2^2/n2 (Theorem 11.8).

The mean of the sampling distribution of the sample proportion P equals: (A) n (B) pi(1-pi) (C) pi (D) N/n

Correct answer: (C) pi. Theorem 11.11 states mu_P = E(P) = pi, the true population proportion.

Quick Revision Summary

- Population = whole group (parameter, constant); Sample = part studied (statistic, random variable)
- Sampling frame = list of all sampling units; probability sampling gives every unit a known chance of selection
- Theorem 11.1: with replacement N^n samples; without replacement NPn = N(N-1)...(N-n+1) samples
- Sampling error: T-theta, shrinks with larger n | Non-sampling error: from bad frames/questions/entry, doesn't shrink with n
- Accuracy = closeness to true value | Precision = closeness of repeated estimates (measured by SE) | Bias = E(T)-theta
- Thm 11.2-11.3: mu_Xbar=mu; sigma_Xbar^2=sigma^2/n (with replacement)
- Thm 11.4: without replacement, multiply variance by fpc = (N-n)/(N-1), always <=1
- Thm 11.6 (CLT): for large n, X-bar ~approx N(mu, sigma^2/n) regardless of population shape
- Difference of means: mu=mu1-mu2, sigma^2=sigma1^2/n1+sigma2^2/n2 (variances ADD for independent samples)
- Sample proportion: mu_P=pi, sigma_P^2=pi(1-pi)/n [x fpc if without replacement]

Exam Tips

- Always identify FIRST whether sampling is with or without replacement -- it changes both the counting formula (N^n vs NPn) and the variance formula (fpc or not)
- When a question says 'infinite population' or doesn't mention N, treat it as sampling with replacement -- no fpc needed
- For difference-of-means or difference-of-proportions problems, remember variances ADD even though you're computing a difference

- Build the sampling distribution table systematically: list all samples, compute the statistic for each, then tally into a frequency/probability table before computing mean and variance
- The CLT applies to means and proportions for LARGE n -- don't assume normality for small samples unless the population itself is stated to be normal
- Double-check verification questions (e.g., 'verify mu_Xbar = mu') by computing both sides independently and confirming they match numerically