MODULE 3

Sampling and Sampling Distributions

BSM301 · Mathematics III

...drawing a sample to learn about the whole population.

What We'll Cover

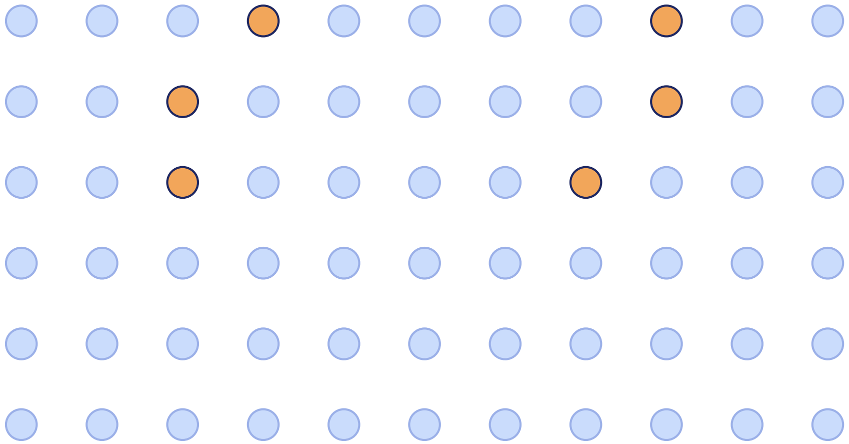
1. **Population & Sample**
   Parameters vs. statistics, SRSWR vs. SRSWOR
2. **Sampling Distributions**
   Built by hand from a tiny population
3. **Standard & Probable Error**
   Quantifying how much a statistic varies
4. **Sampling Distribution of the Mean**
   Standard error, with and without replacement
5. **The Central Limit Theorem**
   Why sample means turn out approximately Normal
6. **Sampling Distribution of the Proportion**
   The categorical cousin of the sample mean
7. **Sampling Distribution of Variance**
   Degrees of freedom and the chi-square distribution
8. **Putting It Together**
   A comparison table and real-world applications

Population vs. Sample

Every statistical inference starts with this distinction

POPULATION (N)

Every unit that could possibly be studied



Amber dots = a SAMPLE (n) drawn from the population

Population

The complete set of all items or individuals under study, of size N.

Sample

A subset of size n, drawn from the population, used to learn about it.

Why it matters: we almost never observe the whole population — we use sample statistics to estimate unknown population parameters, and quantify how much uncertainty that introduces.

Why Sample Instead of Taking a Census?

Studying the entire population is often impossible or unwise

1

Cost

Surveying every device, user, or transaction is far more expensive than a representative subset

2

Time

A full census can take so long the population has already changed by the time it finishes

3

Destructive testing

Testing a battery's lifetime to failure destroys it — you can't do that to every unit

4

Infeasibility

Some populations are unbounded or unknown in size, like all future network packets

Random Sampling: With vs. Without Replacement

Two ways to draw a random sample, and why the difference matters

SRSWR

Simple Random Sampling With Replacement

- Each selected unit is returned to the population before the next draw
- The same unit can be selected more than once
- Every draw has exactly the same probability, 1/N

Example: randomly sampling IP addresses from a rotating pool where an address can appear again

SRSWOR

Simple Random Sampling Without Replacement

- A selected unit is set aside — it cannot be drawn again
- Every unit appears at most once in the sample
- Draw probabilities change slightly after each draw
- Needs a finite population correction factor in formulas

Example: auditing 20 unique blockchain wallets out of a fixed set of 500 for compliance review

Population Parameters vs. Sample Statistics

Parameters describe the population; statistics estimate them from a sample

Quantity	Population Parameter	Sample Statistic
Mean	μ	$\bar{x}$ (x-bar)
Variance	σ^2	s^2
Standard Deviation	σ	s
Proportion	P	$\hat{p}$ (p-hat)
Size	N	n

Rule of thumb: Greek letters (μ, σ, σ^2) always describe the fixed but usually unknown population; Roman letters ($\bar{x}$, s, s^2) describe what we actually calculate from the data in hand. A statistic is a random variable — it changes from sample to sample — while a parameter is a fixed constant.

The Idea of a Sampling Distribution

What happens if we could draw every possible sample?

1. Pick a sample size n from a population of size N
2. List every possible sample of that size (or imagine drawing one repeatedly)
3. Compute the statistic (e.g. the mean) for each possible sample
4. The probability distribution of all those statistic values is the sampling distribution

Key idea: a sampling distribution is a distribution of a statistic, built from all possible samples — not a distribution of raw data.

Worked Example — Building a Sampling Distribution

EXAMPLE

A tiny population, small enough to enumerate every possible sample

Setup: Population = {2, 4, 6, 8} ($N = 4$). Draw samples of size $n = 2$, with replacement (SRSWR).

1. Population mean: $\mu = (2+4+6+8)/4 = 5$
2. Population variance: $\sigma^2 = \Sigma(xi-\mu)^2/N = 5$ ($\sigma = 2.236$)

Next: with replacement, every pair (x_1, x_2) from {2,4,6,8}×{2,4,6,8} is equally likely — that's 4 × 4 = 16 equally likely ordered samples. The next slide lists all 16 and their sample means.

Worked Example — All 16 Possible Samples

EXAMPLE

Each ordered pair (x_1, x_2) is equally likely, probability 1/16

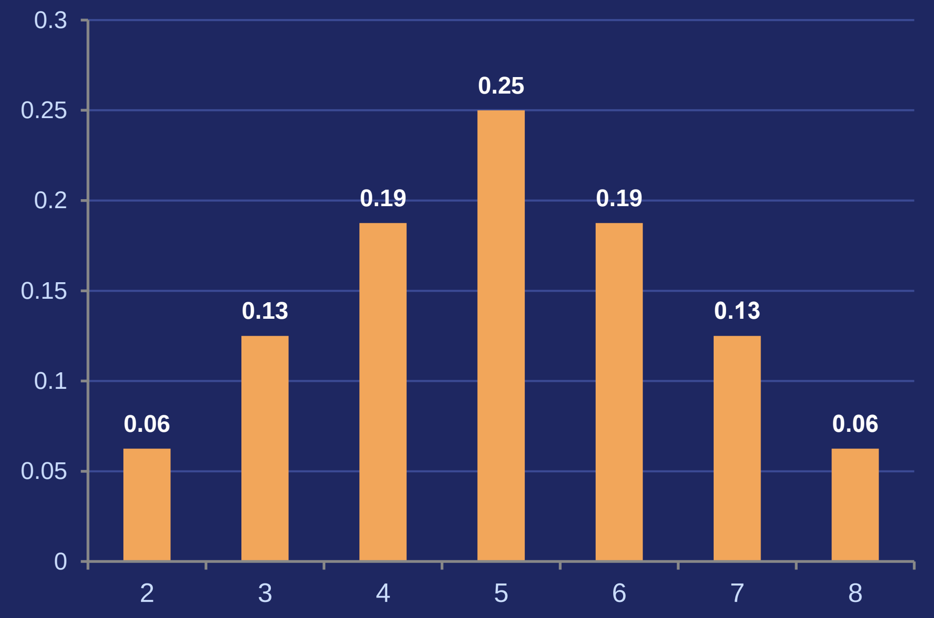
Sample mean $\bar{x}$ for every (x_1, x_2) combination:

$x_1 \setminus x_2$	2	4	6	8
2	2.0	3.0	4.0	5.0
4	3.0	4.0	5.0	6.0
6	4.0	5.0	6.0	7.0
8	5.0	6.0	7.0	8.0

Read it as: row = first draw x_1, column = second draw x_2, cell = the sample mean for that ordered pair. Drawing 2 then 6 gives $\bar{x}$ = 4.0, same as drawing 6 then 2.

Frequency of Each Sample Mean

$\bar{x}$	2	3	4	5	6	7	8
freq	1	2	3	4	3	2	1



Worked Example — Two Remarkable Facts

EXAMPLE

The sampling distribution's own mean and variance follow exact rules

Fact 1 — The sampling distribution is centred on μ

1. $E(\bar{x}) =$ (mean×freq)/16 = 80/16 = 5 = μ ✓

On average, across all possible samples, the sample mean lands exactly on the population mean — the sample mean is an unbiased estimator of μ.

Fact 2 — The sampling distribution is tighter than the population

1. $\mathrm{Var}(\bar{x}) = \Sigma(\text{mean}-\mu)^2 \times \text{freq}/16 = 40/16 = 2.5 \quad = \quad \sigma^2/n = 5/2 = 2.5$ ✓
2. $\mathrm{SE}(\bar{x}) = \sqrt{2.5} \approx 1.581$ — smaller than $\sigma = 2.236$

Standard Error

The standard deviation of a sampling distribution

SE(statistic) = **standard deviation of the statistic's sampling distribution**

Measures

How much a statistic (e.g. $\bar{x}$) varies from sample to sample

Shrinks with n

Larger samples give more precise, lower-SE estimates

Not the same as σ

σ measures spread within one sample; SE measures spread across many samples' statistics

Drives confidence intervals

Margins of error and hypothesis tests are built directly on SE

Probable Error

A older, related measure of a statistic's variability

$$PE = 0.6745 \times SE$$

Where 0.6745 comes from: for a Normal distribution, exactly 50% of values lie within ±0.6745 standard deviations of the mean — so PE marks the boundary of the middle 50% of a normally distributed statistic.

Worked Example

From our tiny-population example, $SE(\bar{x}) = 1.581$.

$PE = 0.6745 \times 1.581 \approx 1.066$

In practice: modern statistics almost always reports SE (or a confidence interval) rather than PE — but PE still appears in some older textbooks and exam syllabi.

Standard Error of the Mean — With Replacement

The formula that powers most of applied statistics

$$SE(\bar{x}) = \sigma / \sqrt{n}$$

σ

the population standard deviation

n

the sample size

Effect of n↑

SE shrinks proportional to $1/\sqrt{n}$ — quadrupling n halves the SE

Assumes

SRSWR, or an infinite / very large population

Intuition: averaging more observations smooths out individual randomness — but with diminishing returns, since halving SE requires quadrupling n, not just doubling it.

Standard Error of the Mean — Without Replacement

A correction factor accounts for sampling a finite population

$$SE(\bar{x}) = \sigma/\sqrt{n} \times \sqrt{[(N-n)/(N-1)]}$$

The finite population correction (FPC): $\sqrt{[(N-n)/(N-1)]}$ is always ≤ 1, so SRSWOR always gives a smaller (or equal) standard error than SRSWR for the same n — sampling without replacement is slightly more informative.

When n is small vs. N

FPC ≈ 1, so the correction barely matters (common rule: skip it if $n/N < 5\%$)

When n is a large fraction of N

FPC shrinks noticeably — the correction cannot be ignored

Worked Example — Comparing SRSWR and SRSWOR

EXAMPLE

Same population, same sample size — different standard errors

Setup: A population of $N = 100$ devices has $\sigma = 15$. A sample of $n = 25$ is drawn.

SRSWR

1. $SE = \sigma/\sqrt{n}$
2. $= 15/\sqrt{25} = 15/5$

$SE = 3.00$

SRSWOR

1. $SE = (\sigma/\sqrt{n}) \times \sqrt{[(N-n)/(N-1)]}$
2. $= 3 \times \sqrt{(75/99)} = 3 \times 0.8704$

$SE = 2.611$

Notice: $2.611 < 3.00$ — SRSWOR gives the smaller standard error, as expected.

The Central Limit Theorem

Perhaps the single most important result in all of statistics

For a sufficiently large sample size n, the sampling distribution of the mean $\bar{x}$ is approximately

Normal(μ, σ^2/n)

— no matter what shape the original population's distribution has.

"Sufficiently large"

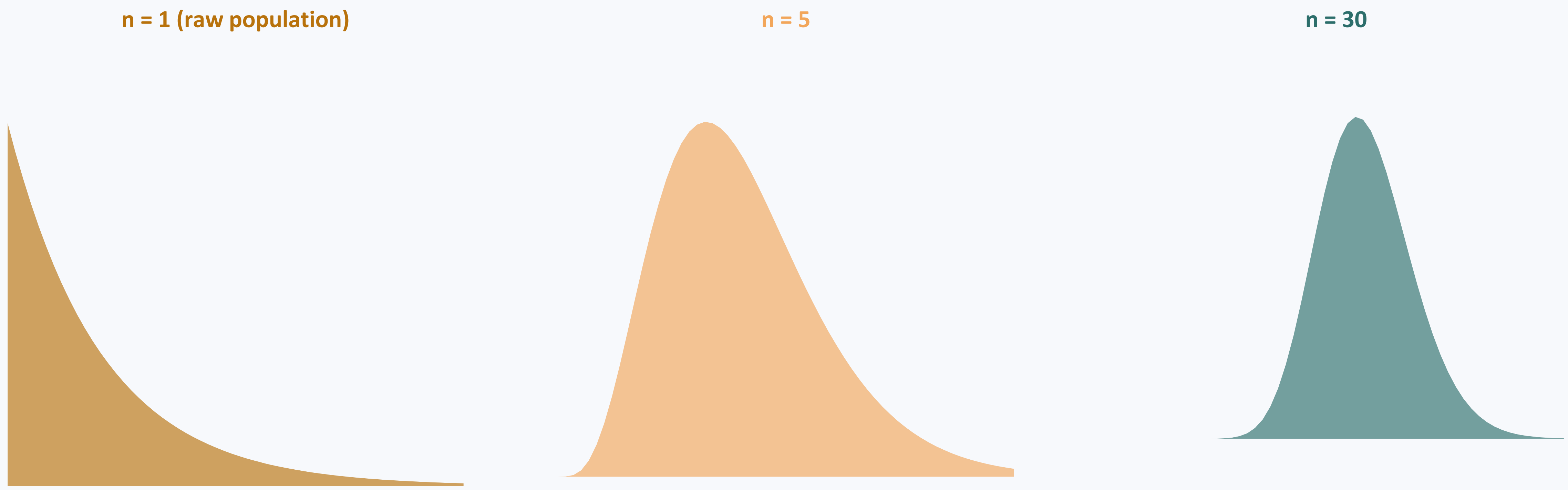
commonly taken as $n \geq 30$ as a rule of thumb (smaller n works fine if the population is already close to Normal)

Why it matters

it lets us use Normal-distribution tools (z-scores, confidence intervals) for the sample mean, even from skewed or unknown populations

Watching the Central Limit Theorem in Action

A skewed population's sampling distribution of the mean, as n grows



What's happening: the population here is right-skewed (like an Exponential distribution). As n grows from 1 to 30, the sampling distribution of $\bar{x}$ becomes tighter and more symmetric — converging toward the bell shape the CLT predicts.

EXAMPLE

Worked Example — Applying the CLT

Using the Normal approximation for a sample mean

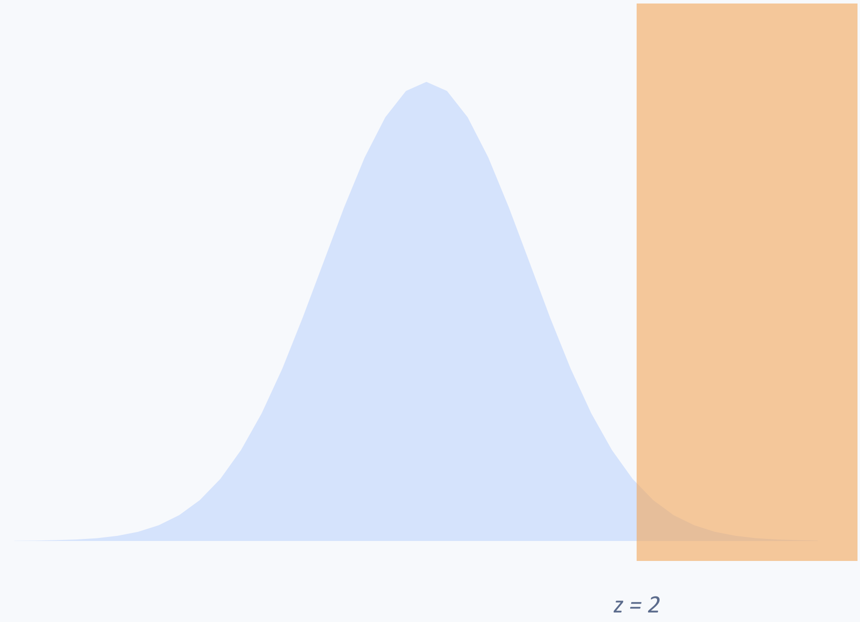
Question: IoT sensor readings have population mean $\mu = 100$ and $\sigma = 20$. For a sample of $n = 64$ readings, find $P(\bar{x} > 105)$.

Step-by-Step

1. Standard error: $SE = 20/\sqrt{64} = 2.50$
2. Standardize: $z = (105-100)/2.5 = 2.00$
3. Look up table: $P(Z > 2.00) = 0.0228$

$P(\bar{x} > 105) \approx 2.28\%$

Sampling distribution of $\bar{x}$ (standardized)



Sampling Distribution of the Proportion

The categorical cousin of the sample mean

$$\hat{p} = X / n$$

X = number of "successes" in the sample, n = sample size

E($\hat{p}$)

= P (unbiased: the sample proportion averages out to the true population proportion)

Var($\hat{p}$)

= PQ/n, where Q = 1 − P

Shape

approximately Normal when n is large (a version of the CLT for proportions)

Examples

% of malicious packets, % of failed devices, % of confirmed transactions

Rule of thumb for "large n": the Normal approximation for $\hat{p}$ works well when both $n \cdot P \geq 5$ and $n \cdot Q \geq 5$ — i.e. the sample is expected to contain a reasonable number of both successes and failures.

Standard Error of the Sample Proportion

Same logic as the mean, adapted for a 0/1 outcome

$$SE(\hat{p}) = \sqrt{(PQ / n)}$$

P, Q

the (usually unknown) true population proportion and its complement, $Q = 1 - P$

When P is unknown

substitute the sample proportion $\hat{p}$ for P — gives an estimated SE

SRSWOR version

multiply by the same finite population correction $\sqrt{[(N-n)/(N-1)]}$ used for the mean

Largest SE

occurs at $P = 0.5$, where PQ is maximized — proportions near 0 or 1 have smaller SE

Worked Example — Sample Proportion

EXAMPLE

A network security screening scenario

Question: 5% of all packets on a network are malicious ($P = 0.05$). In a sample of $n = 500$ packets, find $P(\hat{p} > 0.07)$.

Step-by-Step

1. **Standard error:** $SE = \sqrt{(0.05 \times 0.95/500)} = 0.0097$
2. **Standardize:** $z = (0.07 - 0.05)/0.0097 = 2.05$
3. **Look up table:** $P(Z > 2.05) = 0.0201$

$P(\hat{p} > 0.07) \approx 2.01\%$

Interpretation: even though the true malicious-packet rate is only 5%, a sample of 500 packets shows more than 7% malicious about 2% of the time just by chance — a security dashboard alerting at 7% would false-alarm roughly 1 time in 50, purely from sampling variability.

Sampling Distribution of the Variance

Just like $\bar{x}$ and $\hat{p}$, the sample variance s^2 is itself a random variable

Definition: $s^2 = \Sigma(x_i - \bar{x})^2 / (n-1)$ — every different sample of size n from the same population gives a slightly different s^2. The distribution of all those possible s^2 values is the sampling distribution of the variance.

$E(s^2)$

$= \sigma^2$ — s^2 (with divisor $n-1$) is an unbiased estimator of the population variance

Why $n-1$, not n?

dividing by $n-1$ exactly corrects a small downward bias that dividing by n would introduce

Population variance known

we can standardize s^2 directly against the known σ^2

Population variance unknown

the more common real-world case — covered next

Case: Population Variance Is Unknown

The realistic scenario — and why it needs a new distribution

1. In practice, σ^2 is almost never known — we only have the sample and its s^2
2. We'd like to standardize s^2 the way we standardize $\bar{x}$, but σ^2 itself is missing
3. Using $\bar{x}$ (estimated from the same sample) instead of μ "uses up" one degree of freedom
4. The resulting standardized quantity no longer follows a Normal shape — it follows the chi-square distribution

Coming up: we formalize "degrees of freedom", then meet the chi-square distribution that s^2 actually follows.

Degrees of Freedom

How many values are actually free to vary?

Definition: the number of independent pieces of information available to estimate a quantity — usually (sample size) minus (number of parameters already estimated from the same data).

A Concrete Illustration

Suppose 5 numbers must have a mean of 6, so they must sum to 30.

Pick the first 4 freely: 4, 6, 8, 7 (sum so far = 25).

The 5th number is now forced: 30 − 25 = 5. It has no freedom left.

In s^2: each deviation $(x_i - \bar{x})$ is used, but the n deviations always sum to exactly zero — so once $n-1$ of them are known, the last is determined. That's why s^2 has $n-1$ degrees of freedom.

The Chi-Square (χ^2) Distribution

Built from squared, standardized Normal variates

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

where each Zi ~ Normal(0, 1), independently

Parameter

k = degrees of freedom = the number of squared Normal terms summed

Support

$\chi^2 \geq 0$ — it's a sum of squares, so it can never be negative

Shape

right-skewed for small k, approaching a Normal shape as k grows large

Where it's used

testing variances, goodness-of-fit tests, and tests of independence

Chi-Square and the Sample Variance

The key relationship that makes s^2 usable in practice

$$(n - 1)s^2 / \sigma^2 \sim \chi^2_{n-1}$$

What this means: take the sample variance s^2, scale it by $(n-1)/\sigma^2$, and the result follows a chi-square distribution with $n-1$ degrees of freedom — this is what lets us build confidence intervals and hypothesis tests for a population variance.

Why this matters in practice: network engineers use this to test whether latency variance has changed after an infrastructure update; quality teams use it to test whether a sensor's measurement variance stays within spec — both are literally testing hypotheses about σ^2 using this chi-square relationship.

Mean and Variance of the Chi-Square Distribution

Two simple formulas, both depending only on k

Mean

$$E(\chi^2) = k$$

Variance

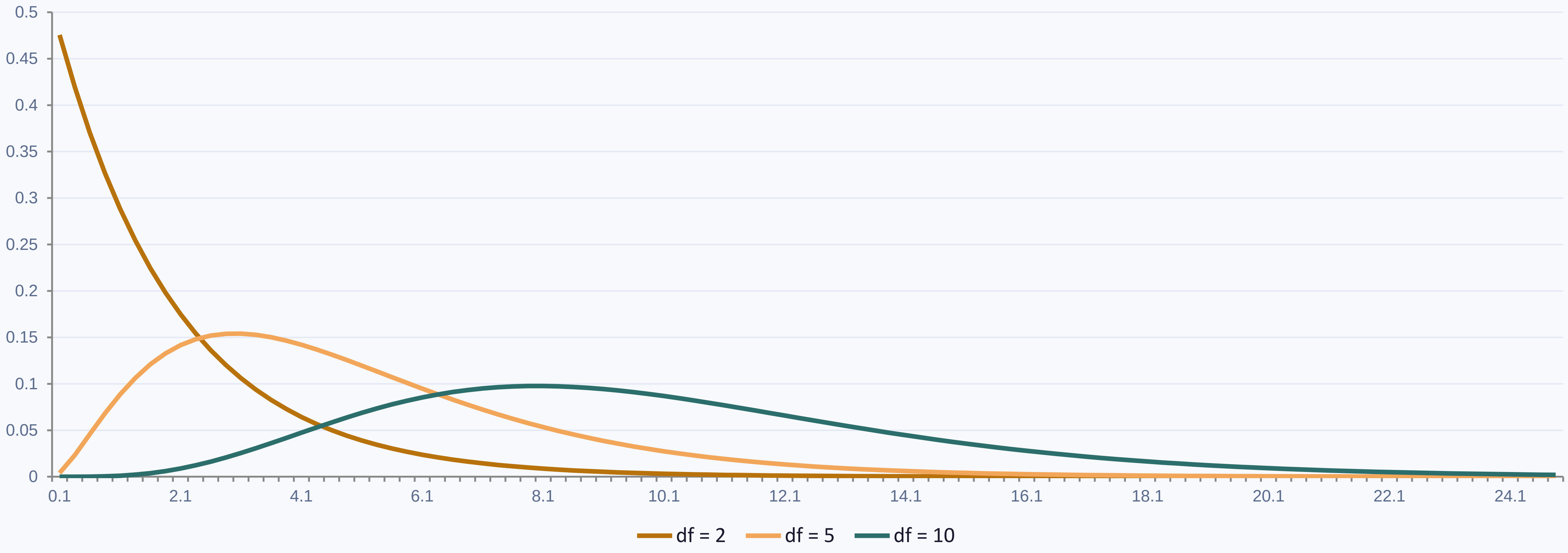
$$\mathrm{Var}(\chi^2) = 2k$$

Sanity check with df = 1: a single squared standard Normal, Z^2, has $E(Z^2) = \mathrm{Var}(Z) = 1 = k$, matching the formula — and its variance works out to $2 = 2k$ as well.

Intuition: as k grows, both the centre (k) and the spread ($\sqrt{2k}$) grow, but the spread grows more slowly than the centre — which is exactly why the distribution looks progressively more symmetric and Normal-like for large k.

Chi-Square Shapes for Different Degrees of Freedom

Right-skewed for small k, increasingly symmetric for large k



Notice: **the peak shifts right and the curve flattens as df increases — each curve's peak sits near df−2, consistent with mean = df.**

Worked Example — A Chi-Square Calculation

EXAMPLE

Testing whether a sample's variance is consistent with a claimed σ^2

Question: A sample of $n = 10$ sensor readings gives $s^2 = 25$. The manufacturer claims the population variance is $\sigma^2 = 20$. Compute the chi-square statistic.

Step-by-Step

1. **Degrees of freedom:** $df = n-1 = 9$
2. **Chi-square statistic:** $\chi^2 = (n-1)s^2/\sigma^2 = 9 \times 25/20$
3. $= 225/20 = 11.25$

$\chi^2 = 11.25$ with $df = 9$

Interpreting the result: the critical value $\chi^2_{0.05,9} \approx 16.92$. Since $11.25 < 16.92$, this sample's variance is well within the range expected under the manufacturer's claim — there's no strong evidence the true variance differs from 20.

Formula Cheat-Sheet

Every formula from this module, in one place

Quantity	Formula
SE of the mean (SRSWR)	$\sigma / \sqrt{n}$
SE of the mean (SRSWOR)	$(\sigma/\sqrt{n}) \times \sqrt{[(N-n)/(N-1)]}$
Probable Error	$PE = 0.6745 \times SE$
SE of a proportion	$\sqrt{(PQ/n)}$
Chi-square from sample variance	$\chi^2 = (n-1)s^2/\sigma^2$, $df = n-1$
Mean, variance of χ^2	$E(\chi^2) = k$, $Var(\chi^2) = 2k$

All Sampling Distributions at a Glance

One table to revise the whole module

Statistic	Sampling Distribution	Mean	Standard Error	Needs σ^2 known?
Sample mean $\bar{x}$	≈ Normal (by CLT)	μ	$\sigma/\sqrt{n}$	Yes (or use s)
Sample proportion $\hat{p}$	≈ Normal (large n)	P	$\sqrt{PQ/n}$	No — uses P, Q
$\mathbf{(n-1)s^2/\sigma^2}$	Chi-square, df = n−1	n−1	$\sqrt{2(n-1)}$	Yes

The common thread: every sampling distribution answers the same question — "how much would this statistic vary if we repeated the sampling process over and over?" — just for a different statistic each time.

Looking ahead: Module 4 (Estimation) uses these exact standard errors to build confidence intervals; Module 5 (Hypothesis Testing) uses these exact sampling distributions to decide whether an observed difference is statistically significant.

Sampling Theory in Practice

The same ideas, applied across your specialization tracks

Internet of Things

Sample mean

Estimating average battery life from a sample of deployed sensors

Sample proportion

Estimating the fraction of a sensor fleet reporting corrupted readings

Chi-square

Testing whether a new firmware's variance in measurement error has changed

Cyber Security

Sample mean

Estimating average response time of a detection system from sampled logs

Sample proportion

Estimating the true rate of malicious traffic from a sampled packet capture

Chi-square

Testing whether login-attempt variance has shifted after a policy change

Blockchain

Sample mean

Estimating average transaction fee from a sample of recent blocks

Sample proportion

Estimating the share of failed transactions from a sample of the mempool

Chi-square

Testing whether block-time variance is consistent with the protocol's target

Common Pitfalls to Avoid

Mistakes that show up again and again in exams and in practice

Confusing σ with SE

σ describes spread within one sample; SE describes spread of the statistic across many samples — they are related by √n, not equal

Forgetting the FPC

skipping the finite population correction when n is a large fraction of N overstates the standard error

Using n instead of n−1

for the sample variance, dividing by n instead of n−1 introduces a small but systematic downward bias

Misreading "large n" for CLT

$n \geq 30$ is a rule of thumb, not a guarantee — a heavily skewed population may need a much larger n before the Normal approximation is trustworthy

Applying Normal tools to χ^2

the chi-square distribution is not symmetric — don't use z-tables for it; use chi-square tables with the correct df

Ignoring what df represents

always double-check which quantity's degrees of freedom you're using — n−1 for a single variance is not the same df used elsewhere

Key Takeaways

1. A sampling distribution describes how a statistic (mean, proportion, variance) varies across all possible samples — it is not the distribution of raw data.
2. Standard error is the standard deviation of that sampling distribution, and it shrinks as sample size grows, roughly as $1/\sqrt{n}$.
3. The Central Limit Theorem lets the sample mean be treated as approximately Normal for large n, regardless of the population's shape.
4. The sample proportion follows the same logic as the mean, with $SE = \sqrt{PQ/n}$.
5. The chi-square distribution governs $(n-1)s^2/\sigma^2$, with degrees of freedom $n-1$, mean k, and variance 2k.
6. Next: Module 4 — Estimation of Parameters, turning these standard errors into confidence intervals.