A LEVEL MATHS 9709 · PAPER 6 PROBABILITY & STATISTICS 2 · 6.4 · ~4 MIN READ

Sampling and estimation

Populations and random samples, the distribution of the sample mean, the Central Limit Theorem, unbiased estimates of the mean and variance, and confidence intervals for a population mean and a population proportion.

Populations & samples | Random sampling | Distribution of $\bar{X}$ | Central Limit Theorem | Unbiased estimates | s^2 with n − 1 | Confidence intervals for μ | Confidence intervals for p

RS by Riyaz Sir
mentorriyaz.in

Paper tag = **where it is examined** FORMULA LIST (MF19) given in the exam M1 A1 B1 how marks are awarded highlight = comes up again and again Test yourself at the end

IN 30 SECONDS

- **$\bar{X}$ has mean μ and variance σ^2/n.** Larger samples give means that cluster more tightly around μ.
- **Central Limit Theorem:** for large n (n ≥ 30 or so), $\bar{X}$ is approximately normal whatever the population's distribution.
- **Unbiased estimate of σ^2:** $s^2 = [\Sigma x^2 - (\Sigma x)^2/n]/(n - 1)$. Dividing by n − 1 corrects for using $\bar{x}$ in place of μ.
- **Confidence interval for μ:** $\bar{x} \pm z \times \sigma/\sqrt{n}$. A 95% interval uses z = 1.96; 90% uses 1.645; 99% uses 2.576.

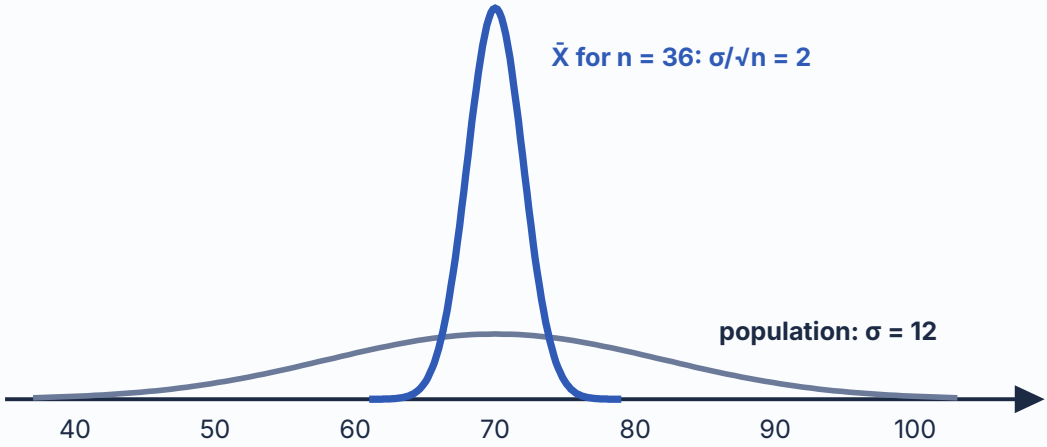
BY THE END YOU CAN

- Explain populations, samples and random sampling, and why a sample might not be random.
- Use the distribution of the sample mean, including the Central Limit Theorem.
- Calculate unbiased estimates of the population mean and variance.
- Find confidence intervals for a population mean and for a population proportion, and interpret them.

1 Samples and the sample mean

$\bar{X} \sim N(\mu, \sigma^2/n)$

- **Random sample:** every possible sample of size n is equally likely to be chosen. Choosing the first 30 customers through the door isn't random.
- **Population normal:** $\bar{X} \sim N(\mu, \sigma^2/n)$ exactly, for any n. **Population not normal:** $\bar{X} \approx N(\mu, \sigma^2/n)$ for large n (CLT).



Sample means cluster much more tightly around μ, and are approximately normal for large n.

Sample means are much less spread out than single values, and approximately normal for large n.

WORKED EXAMPLE

Masses are N(80, 10^2). Find the probability that the mean of a random sample of 25 exceeds 83.

1. $\bar{X}$ ~ N(80, 100/25 = 4) M1 A1
2. **z = 3/2 = 1.5 ⇒ 1 − 0.9332 = 0.0668** A1

WORKED EXAMPLE

A population has mean 5 and variance 4 (not normal). Find the approximate probability that the mean of 40 values is less than 4.6.

1. by the CLT, $\bar{X}$ ≈ N(5, 0.1) B1
2. z = −0.4/√0.1 = −1.265 M1
3. **0.103** A1

COMMON MISTAKE

Quoting the CLT for a normal population, or for a small sample from a non-normal one. The CLT is needed only when the population isn't normal, and it needs n large.

Riyaz Sir

2 Estimation

Unbiased estimates

$$\bar{x} = \Sigma x/n \; ; \; s^2 = [1/(n-1)][\Sigma x^2 - (\Sigma x)^2/n]$$ FORMULA LIST (MF19)

s^2 is the unbiased estimate of σ^2

WORKED EXAMPLE

A sample of 50 has $\Sigma x = 1240$ and $\Sigma x^2 = 31\,200$. Find unbiased estimates of μ and σ^2.

1. $\bar{x} =$ **24.8** B1
2. $s^2 = (31\,200 - 1240^2/50)/49 = 448/49$ M1
3. **$= 9.14$** A1

Confidence intervals

$$\mu: \bar{x} \pm z\,\sigma/\sqrt{n} \; ; \; p: \hat{p} \pm z\sqrt{[\hat{p}(1-\hat{p})/n]}$$

use s for σ when n is large and σ is unknown

WORKED EXAMPLE

Use the sample above to find a 95% confidence interval for μ.

1. $24.8 \pm 1.96\sqrt{(9.143/50)}$ M1
2. 24.8 ± 0.838 A1
3. **$(23.96, 25.64)$** A1

WORKED EXAMPLE

In a random sample of 200 students, 64 use the library weekly. Find a 90% confidence interval for the proportion p.

1. $\hat{p} = 0.32$ B1
2. $0.32 \pm 1.645\sqrt{(0.32 \times 0.68/200)} = 0.32 \pm 0.0543$ M1 A1
3. **$(0.266, 0.374)$** A1

- **Interpretation:** if many such intervals were made, about 95% would contain μ. It doesn't mean "μ is in this interval with probability 0.95".

COMMON MISTAKE

Using n instead of n – 1 for s^2, or σ instead of $\sigma/\sqrt{n}$ in the interval. Both appear every year in the Paper 6 reports.

LINKING QUESTION

Linear combinations explain why $\mathrm{Var}(\bar{X}) = \sigma^2/n$: see Linear combinations; the normal is in The normal distribution; and tests on μ are in Hypothesis tests.

Test yourself — cover the answers first

QUICK CHECK

1. X ~ N(10, 9). What is the distribution of the mean of 16 values?
2. What z-value gives a 99% confidence interval?
3. Why divide by n − 1 for s^2?
4. When do you need the Central Limit Theorem?
5. Is "the first 20 names on a list" a random sample?

EXAM-STYLE

6. A sample of 10: $\Sigma x = 52$, $\Sigma x^2 = 290$. Find unbiased estimates of μ and σ^2.
7. How large a sample is needed for the 95% interval for μ to have width at most 2, if $\sigma = 6$?
8. A 95% confidence interval for p is (0.22, 0.34). Find $\hat{p}$ and n.

Show answers

1. $N(10, 9/16)$
2. 2.576
3. to give an unbiased estimate of σ^2
4. when the population isn't normal and n is large
5. no
6. 5.2; $(290 - 270.4)/9 = 2.18$
7. $2 \times 1.96 \times 6/\sqrt{n} \leq 2 \Rightarrow \sqrt{n} \geq 11.76 \Rightarrow n = 139$
8. $\hat{p} = 0.28$; $1.96\sqrt{(0.28 \times 0.72/n)} = 0.06 \Rightarrow n = 215$

Stuck on a question?

Send it to me on WhatsApp. I read every message.

Share with a friend

Ask Riyaz Sir

Want regular help? Send this to your parent →