

PROBABILITY & STATISTICS · JNTUH R18

Solved Previous Year Questions

12 papers (Nov-2020 to Apr/May-2026), grouped unit-wise. Every numerical answer was checked by computation.

Apr/May-2026

Oct/Nov-2025

May-2025

Nov-2024

Mar/Apr-2024

Oct-2023

Apr/May-2023

Nov/Dec-2022

Mar/Apr-2022

Oct/Nov-2021

Jul-2021

Nov-2020

HOW TO READ THIS DOCUMENT

Each question carries an orange tag with the paper(s) and marks. When a question appeared in more than one paper, all papers are listed in one tag group. Theory questions are answered first in each unit (short model answers). Numerical questions follow, fully solved. The "Unit X, §" references point to the full explanation in your unit notes.

Note: the Apr/May-2026 scan has no page for Q4–Q9, so those questions are not available.

Unit 1 Probability

A. Theory questions — short model answers

PYQ. State Bayes' theorem and explain its significance. Oct-2025 · 3M

If $E_1, \dots, E_n$ are mutually exclusive and exhaustive with $P(E_i) > 0$, and A is any event with $P(A) > 0$, then $P(E_k/A) = \frac{P(E_k)P(A/E_k)}{\sum_i P(E_i)P(A/E_i)}$. **Significance:** it finds the probability of a cause after the effect is seen (prior $\rightarrow$ posterior). Full proof: Unit 1, §1.9.

PYQ. State the multiplication theorem of probability. May-2025 · 2M

For events A, B with $P(A) \neq 0$: $P(A \cap B) = P(A)P(B/A)$. For independent events, $P(A \cap B) = P(A)P(B)$.

PYQ. State and prove the addition theorem. Nov/Dec-2022 · 3M

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Proof: $A \cup B = A \cup (\bar{A} \cap B)$ (disjoint) and $B = (\bar{A} \cap B) \cup (A \cap B)$ (disjoint); apply the axiom of union to both. Full proof: Unit 1, §1.5.

PYQ. Define probability (classical). Mar/Apr-2024 · 2M

If a trial has n mutually exclusive, equally likely, exhaustive outcomes and m favour E , then $P(E) = \frac{m}{n}$.

B. Numerical and derivation questions — fully solved

PYQ. Define sample space and give an example. Oct-2025 · 2M Nov-2022 · 2M

Sample space: The set of all possible outcomes of a random experiment is the sample space S .

Example: Throw one die. $S = \{1, 2, 3, 4, 5, 6\}$. The event "even number" is $E = \{2, 4, 6\}$.

PYQ. A five-figure number is formed by the digits 0, 1, 2, 3, 4 (without repetition). Find the probability that the number is divisible by 4. Nov-2024 · 6M

Step 1 Find $n(S)$. The first place cannot be 0. So $n(S) = 4 \times 4 \times 3 \times 2 \times 1 = 96$.

Step 2 A number is divisible by 4 when its **last two digits** are divisible by 4. Possible endings: 04, 12, 20, 24, 32, 40.

Step 3 Endings with 0 (04, 20, 40): the other 3 digits fill 3 places freely: $3! = 6$ each. Total = 18.

Step 4 Endings without 0 (12, 24, 32): 0 is left. First place cannot be 0: $2 \times 2 \times 1 = 4$ each. Total = 12.

Step 5 $n(E) = 18 + 12 = 30$. $P(E) = \frac{30}{96} = \frac{5}{16}$. 5/16

PYQ. What is the probability of getting 9 cards of the same suit if 12 cards are drawn at random?

Mar/Apr-2024 · 3M

$n(S) = {}^{52}C_{12}$. Select the suit in 4 ways. Select 9 cards from that suit: ${}^{13}C_9$. Select 3 other cards from the other 39 cards: ${}^{39}C_3$.

$$P = \frac{4 \cdot {}^{13}C_9 \cdot {}^{39}C_3}{{}^{52}C_{12}} \approx 0.000127$$

PYQ. A die is rolled. If the number is odd, what is the probability that it is prime? Nov-2024 · 3M

The number is odd, so the new sample space is $\{1, 3, 5\}$. $n(S) = 3$. Odd primes are $\{3, 5\}$. $n(E) = 2$.

P = 2/3

PYQ. In a sample of 446 cars stopped at a road block, only 67 drivers had their seat belts fastened. Estimate the probability that a driver will have the seat belt fastened. Apr/May-2026 · 3M

$n(S) = 446$, $n(E) = 67$. $P(E) = \frac{67}{446} \approx 0.150$. ≈ 0.15

PYQ. (a) Ten coins are tossed. Find the probability of at least eight heads. (b) A coin is tossed 12 times. Find the probability of at least one head. Apr/May-2023 · 2M Nov-2022 · 2M

(a) $n(S) = 2^{10} = 1024$. $n(E) = {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10} = 45 + 10 + 1 = 56$. $P = \frac{56}{1024} = \frac{7}{128}$.

(b) $P = 1 - P(\text{no head}) = 1 - \frac{1}{2^{12}} = \frac{4095}{4096}$.

PYQ. State and prove any two axioms of probability. Oct-2025 · 5M Apr/May-2026 · 2M (state)

Axiom 2 (Certainty): $P(S) = 1$.

Proof: $n(S)$ outcomes are all favourable to S . So $P(S) = \frac{n(S)}{n(S)} = 1$.

Axiom 3 (Union): If $E_1 \cap E_2 = \phi$, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Proof: $P(E_1 \cup E_2) = \frac{n(E_1 \cup E_2)}{n(S)}$.

By the principle of inclusion and exclusion, $n(E_1 \cup E_2) = n(E_1) + n(E_2) - n(E_1 \cap E_2) = n(E_1) + n(E_2) - 0$.

So $P(E_1 \cup E_2) = \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} = P(E_1) + P(E_2)$. ■

Axiom 1 (Positivity): $P(E) \geq 0$. Proof: $n(E) \geq 0$ and $n(S) > 0$. So $P(E) = n(E)/n(S) \geq 0$.

PYQ. A card is drawn from a well-shuffled pack of cards. Find the probability that it is either a spade or an ace. May-2025 · 3M Oct-2023 · 3M

$n(S) = 52$. $E_1 = \text{spade: } P(E_1) = \frac{13}{52}$. $E_2 = \text{ace: } P(E_2) = \frac{4}{52}$. $E_1 \cap E_2 = \text{ace of spades: } P = \frac{1}{52}$.
 $P(E_1 \cup E_2) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$. 4/13

PYQ. 60% of people like tea, 50% like coffee and 30% like both. Find the probability that a person likes at least one of the two. Oct-2025 · 5M

$P(T \cup C) = P(T) + P(C) - P(T \cap C) = 0.6 + 0.5 - 0.3 = 0.8$. 0.8

PYQ. Given $P(A) = 0.30$, $P(B) = 0.62$, $P(A \cap B) = 0.12$. Find (i) $P(A \cup B)$ (ii) $P(\bar{A} \cap B)$ (iii) $P(A \cap \bar{B})$ (iv) $P(\bar{A} \cup \bar{B})$. Nov/Dec-2022 · 5M

- (i) $P(A \cup B) = 0.30 + 0.62 - 0.12 = 0.80$
- (ii) $P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.62 - 0.12 = 0.50$
- (iii) $P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.30 - 0.12 = 0.18$
- (iv) $P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 1 - 0.12 = 0.88$

PYQ. If $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{3}$, $P(A \cap B) = \frac{1}{6}$, find (i) $P(A \cup B)$ (ii) $P(A \cap B^c)$ (iii) $P(A \cap B)^c$.

Jul-2021 · 7M

- (i) $\frac{1}{4} + \frac{2}{3} - \frac{1}{6} = \frac{3+8-2}{12} = \frac{3}{4}$ (ii) $\frac{1}{4} - \frac{1}{6} = \frac{1}{12}$ (iii) $1 - \frac{1}{6} = \frac{5}{6}$

PYQ. Show that for any two events A and B, $P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$.

Apr/May-2023 · 5M

Step 1 $A \cap B \subseteq A$. By theorem 3, $P(A \cap B) \leq P(A)$.

Step 2 $A \subseteq A \cup B$. By theorem 3, $P(A) \leq P(A \cup B)$.

Step 3 By the addition theorem, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Because $P(A \cap B) \geq 0$ (axiom 1), $P(A \cup B) \leq P(A) + P(B)$.

Step 4 Join the three results: $P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$. ■

PYQ. Define conditional probability. Nov-2024 · 2M May-2025 · 3M

Let A and B be two events with $P(A) \neq 0$. The probability of B , given that A has occurred, is the conditional probability of B given A : $P(B/A) = \frac{P(A \cap B)}{P(A)}$.

PYQ. Find the probability of drawing two red balls in succession from a bag of 4 red and 5 black balls, (i) with replacement (ii) without replacement. Apr/May-2026 · 4M Class notes

Total balls = 9. E_1 = first ball red, E_2 = second ball red.

(i) With replacement: $P(E_1) = \frac{4}{9}$, $P(E_2/E_1) = \frac{4}{9}$. $P(E_1 \cap E_2) = \frac{4}{9} \cdot \frac{4}{9} = \frac{16}{81}$.

(ii) Without replacement: $P(E_1) = \frac{4}{9}$. After one red ball goes out, 3 red of 8 remain: $P(E_2/E_1) = \frac{3}{8}$.

By the multiplication theorem, $P(E_1 \cap E_2) = \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{6}$. (i) 16/81 (ii) 1/6

PYQ. A box has 4 red and 6 blue balls. Two balls are drawn without replacement. Find the probability that both are red. Oct-2025 · 5M

$$P = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15} \quad (\text{or } \frac{{}^4C_2}{{}^{10}C_2} = \frac{6}{45} = \frac{2}{15}). \quad \text{2/15}$$

PYQ. Two marbles are drawn in succession from a box of 10 red, 15 white, 20 blue and 30 orange marbles, without replacement. Find the probability that (i) both are orange (ii) neither is white. Mar/Apr-2024 · 5M

Total = 75.

$$(i) P = \frac{30}{75} \cdot \frac{29}{74} = \frac{29}{185}. \quad (ii) \text{ Non-white} = 60: P = \frac{60}{75} \cdot \frac{59}{74} = \frac{118}{185}.$$

PYQ. Two aeroplanes bomb a target in succession. The probability that each scores a hit is 0.3 and 0.2. The second plane bombs only if the first misses. Find the probability that (i) the target is hit (ii) the target is hit by the second plane (iii) both fail. May-2025 · 7M Mar/Apr-2022 · 8M

$$P(A) = 0.3, P(\bar{A}) = 0.7, P(B) = 0.2, P(\bar{B}) = 0.8.$$

Step 1 Target is hit = "A hits" or "A misses and then B hits": $P[A \cup (\bar{A} \cap B)] = P(A) + P(\bar{A})P(B) = 0.3 + 0.7(0.2) = 0.44$.

Step 2 Hit by the second plane = $P(\bar{A} \cap B) = 0.7 \times 0.2 = 0.14$.

Step 3 Both fail = $P(\bar{A} \cap \bar{B}) = 0.7 \times 0.8 = 0.56$. (Check: $0.44 + 0.56 = 1$.)

(i) 0.44 (ii) 0.14 (iii) 0.56

PYQ. A, B and C aim to shoot a balloon. A succeeds 4 times in 5, B 3 times in 4, C 2 times in 3. If all three aim together, find the probability that at least two of them hit. Nov/Dec-2022 · 5M Class notes

$$P(A) = \frac{4}{5}, P(\bar{A}) = \frac{1}{5}; P(B) = \frac{3}{4}, P(\bar{B}) = \frac{1}{4}; P(C) = \frac{2}{3}, P(\bar{C}) = \frac{1}{3}.$$

At least two = exactly two + all three:

$$P = P(ABC) + P(AB\bar{C}) + P(A\bar{B}C) + P(\bar{A}BC)$$

$$= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} + \frac{4}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{2}{5} + \frac{1}{5} + \frac{2}{15} + \frac{1}{10} = \frac{5}{6}. \quad \text{5/6}$$

PYQ. Of 10 girls in a class, 3 have blue eyes. 2 girls are chosen at random. Find the probability that (i) both have blue eyes (ii) at least one has blue eyes. Oct-2023 · 3M

$$n(S) = {}^{10}C_2 = 45. \text{ (i) } {}^3C_2/45 = 3/45 = \frac{1}{15}. \text{ (ii) } 1 - {}^7C_2/45 = 1 - \frac{21}{45} = \frac{8}{15}.$$

PYQ. Compute the probability that at least two 7's appear (as a sum) in four tosses of a pair of fair dice. Mar/Apr-2024 · 5M

From the two-dice table (Unit 1, Fig 1.4), $P(\text{sum } 7) = \frac{6}{36} = \frac{1}{6} = p, q = \frac{5}{6}$. Each toss is independent.

$$P(\text{at least two}) = 1 - P(0) - P(1) = 1 - \left(\frac{5}{6}\right)^4 - 4 \cdot \frac{1}{6} \left(\frac{5}{6}\right)^3 = 1 - \frac{625}{1296} - \frac{500}{1296} = \frac{171}{1296} = \frac{19}{144}.$$

PYQ. A and B toss a fair coin alternately. The first to get a head wins. If A starts, find his chance of winning. Apr/May-2023 · 5M Jul-2021 · 8M

$$p_1 = p_2 = \frac{1}{2}, q_1 q_2 = \frac{1}{4}. P(A) = \frac{1/2}{1 - 1/4} = \frac{2}{3}. \text{ Then } P(B) = 1 - \frac{2}{3} = \frac{1}{3}. \quad \text{2/3}$$

PYQ. A man alternately tosses a coin and throws a die, beginning with the coin. Find the probability that he gets a head before he gets a 3 or 4 on the die. Apr/May-2026 · 6M Nov-2024 · 4M (5 or 6)

$$\text{Coin: } p_1 = P(H) = \frac{1}{2}, q_1 = \frac{1}{2}. \text{ Die: } p_2 = P(3 \text{ or } 4) = \frac{2}{6} = \frac{1}{3}, q_2 = \frac{2}{3}.$$

He wins when the head comes first: on turn 1, 3, 5, ...

$$P = \frac{1}{2} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{2} + \left(\frac{1}{2} \cdot \frac{2}{3}\right)^2 \frac{1}{2} + \dots = \frac{1/2}{1 - \frac{1}{2} \cdot \frac{2}{3}} = \frac{1/2}{2/3} = \frac{3}{4}.$$

For "5 or 6" (Nov-2024), $p_2 = \frac{1}{3}$ again, so the answer is also $\frac{3}{4}$. 3/4

PYQ. A and B throw alternately with a pair of dice. A wins if he throws 6 before B throws 7. B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is 30/61. Oct-2023 · 7M

Class notes

From the two-dice table (Unit 1, Fig 1.4): sum 6 = {(1,5), (5,1), (2,4), (4,2), (3,3)}: $p_1 = \frac{5}{36}, q_1 = \frac{31}{36}$.

Sum 7 has 6 ways: $p_2 = \frac{1}{6}, q_2 = \frac{5}{6}$.

$$P(A) = \frac{p_1}{1 - q_1 q_2} = \frac{5/36}{1 - \frac{31}{36} \cdot \frac{5}{6}} = \frac{5/36}{1 - \frac{155}{216}} = \frac{5}{36} \cdot \frac{216}{61} = \frac{30}{61}. \blacksquare$$

PYQ. Coloured balls are in three boxes: Box 1 (4 white, 1 red, 3 blue), Box 2 (2 white, 3 red, 5 blue), Box 3 (3 white, 4 red, 3 blue). A box is selected at random and a ball is drawn. Find the probability that the ball is (a) red (b) blue. Apr/May-2026 · 10M Class notes

	Box X	Box Y	Box Z
White	4	2	3
Red	1	3	4
Blue	3	5	3
Total	8	10	10

Step 1 Each box has the same chance: $P(X) = P(Y) = P(Z) = \frac{1}{3}$.

Step 2 Let R = red ball. $P(R/X) = \frac{1}{8}$, $P(R/Y) = \frac{3}{10}$, $P(R/Z) = \frac{4}{10}$.

Step 3 By total probability: $P(R) = \frac{1}{3} \cdot \frac{1}{8} + \frac{1}{3} \cdot \frac{3}{10} + \frac{1}{3} \cdot \frac{4}{10} = \frac{1}{3} \left[\frac{5+12+16}{40} \right] = \frac{33}{120} = \frac{11}{40}$.

Step 4 Let B = blue. $P(B/X) = \frac{3}{8}$, $P(B/Y) = \frac{5}{10}$, $P(B/Z) = \frac{3}{10}$.
 $P(B) = \frac{1}{3} \left[\frac{3}{8} + \frac{5}{10} + \frac{3}{10} \right] = \frac{1}{3} \left[\frac{15+20+12}{40} \right] = \frac{47}{120}$.

(a) 11/40 (b) 47/120

PYQ. Box I: 2 red, 3 white, 5 blue. Box II: 4 red, 1 white, 3 blue. Box III: 3 red, 4 white, 3 blue. A box is selected at random and a ball is drawn. Find the probability that the ball is (a) red (b) blue.

May-2025 · 10M

Totals: 10, 8, 10. $P(R) = \frac{1}{3} \left[\frac{2}{10} + \frac{4}{8} + \frac{3}{10} \right] = \frac{1}{3} [0.2 + 0.5 + 0.3] = \frac{1}{3}$.

$P(B) = \frac{1}{3} \left[\frac{5}{10} + \frac{3}{8} + \frac{3}{10} \right] = \frac{1}{3} [0.5 + 0.375 + 0.3] = \frac{47}{120}$. **(a) 1/3 (b) 47/120**

PYQ. In a bolt factory, machines A, B, C make 20%, 30%, 50% of the output. Of their output 6%, 3%, 2% are defective. A bolt is drawn and found defective. Find the probability that it is from A, B, C.

Class notes

Oct/Nov-2021 (similar)

Machine	$P(E_i)$	$P(D/E_i)$	Product	$P(E_i/D)$
A	0.2	0.06	0.012	$0.012/0.031 = 12/31$
B	0.3	0.03	0.009	$9/31$
C	0.5	0.02	0.010	$10/31$
Total $P(D)$			0.031	1 ✓

PYQ. A businessman goes to hotels X, Y, Z 20%, 50%, 30% of the time. 5%, 4%, 8% of the rooms in X, Y, Z have faulty plumbing. Find the probability that a room with faulty plumbing is in hotel Z.

Nov-2024 · 10M

Mar/Apr-2022 (hotel Y)

Class notes

Let F = faulty plumbing.

Hotel	$P(\cdot)$	$P(F/\cdot)$	Product	Posterior
X	0.20	0.05	0.010	$10/54 = 5/27$
Y	0.50	0.04	0.020	$20/54 = 10/27$
Z	0.30	0.08	0.024	$24/54 = 4/9$
$P(F)$			0.054	1 ✓

$$P(Z/F) = \frac{0.3 \times 0.08}{0.054} = \frac{4}{9} \approx 0.444. \quad \text{P(Z/F) = 4/9} \quad \text{P(Y/F) = 10/27}$$

PYQ. Of three men, the chances that a politician, a businessman or an academician will be appointed vice-chancellor are 0.5, 0.3, 0.2. The probabilities that research is promoted by them are 0.3, 0.7, 0.8. (a) Find the probability that research is promoted. (b) If research is promoted, find the probability that the VC is an academician. Mar/Apr-2024 · 10M Nov-2020 Class notes

Let R = research is promoted.

Person	Prior	$P(R/\cdot)$	Product	Posterior
Politician (A)	0.5	0.3	0.15	$15/52 \approx 0.288$
Businessman (B)	0.3	0.7	0.21	$21/52 \approx 0.404$
Academician (C)	0.2	0.8	0.16	$16/52 = 4/13 \approx 0.308$
$P(R)$			0.52	1 ✓

$$(a) P(R) = 0.52 = 13/25 \quad (b) P(C/R) = 4/13$$

PYQ. Three companies X, Y, Z produce TVs. X produces twice as many as Y. Y and Z produce the same number. 2% of X, 2% of Y and 4% of Z are defective. A TV chosen at random is defective. Find the probability that it is produced by company Z, and by company X. Apr/May-2023 · 10M Class notes

$$P(X) = 2P(Y), P(Y) = P(Z). \text{ So } 2P(Y) + P(Y) + P(Y) = 1 \Rightarrow P(Y) = \frac{1}{4}, P(X) = \frac{1}{2}, P(Z) = \frac{1}{4}.$$

$$P(D) = \frac{1}{2}(0.02) + \frac{1}{4}(0.02) + \frac{1}{4}(0.04) = 0.01 + 0.005 + 0.01 = 0.025.$$

$$P(Z/D) = \frac{0.01}{0.025} = \frac{2}{5}; P(X/D) = \frac{0.01}{0.025} = \frac{2}{5}. \quad \text{P(Z/D) = P(X/D) = 2/5}$$

CAUTION

One TV comes from one company only. In your class notes the two answers are multiplied ($\frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25}$). In the exam, give both answers $\frac{2}{5}$ separately. If the examiner means "Z or X", add them: $\frac{2}{5} + \frac{2}{5} = \frac{4}{5}$.

PYQ. A slip is given to X, who marks it with a plus sign with probability $1/3$, else a minus sign. X passes it to Y, Y to Z, Z to W, W to a referee. Each of Y, Z, W changes the sign with probability $2/3$. The referee sees a plus. Find the probability that X originally wrote a plus. Oct-2023 · 10M

Step 1 The sign stays the same if the number of changes is even (0 or 2). Let $c = \frac{2}{3}$ (change), $k = \frac{1}{3}$ (keep).

$$P(\text{even}) = k^3 + {}^3C_2 c^2 k = \frac{1}{27} + 3 \cdot \frac{4}{9} \cdot \frac{1}{3} = \frac{1}{27} + \frac{12}{27} = \frac{13}{27}. P(\text{odd}) = \frac{14}{27}.$$

Step 2 Let E_1 = X wrote plus ($\frac{1}{3}$), E_2 = X wrote minus ($\frac{2}{3}$), A = referee sees plus.

$$P(A/E_1) = \frac{13}{27} \text{ (even changes)}, P(A/E_2) = \frac{14}{27} \text{ (odd changes)}.$$

Step 3

$$\text{Bayes: } P(E_1/A) = \frac{\frac{1}{3} \cdot \frac{13}{27}}{\frac{1}{3} \cdot \frac{13}{27} + \frac{2}{3} \cdot \frac{14}{27}} = \frac{13}{13 + 28} = \frac{13}{41}.$$

13/41

PYQ. If there are 4 defective items among 9, and 3 items are selected, find the probability that (i) 2 are defective (ii) one is defective. **Oct/Nov-2021 · 6M**

$$n(S) = {}^9C_3 = 84. \text{ (i) } \frac{{}^4C_2 \cdot {}^5C_1}{84} = \frac{30}{84} = \frac{5}{14}. \text{ (ii) } \frac{{}^4C_1 \cdot {}^5C_2}{84} = \frac{40}{84} = \frac{10}{21}.$$

PYQ. Machines A, B, C produce 20%, 30%, 50% of an item. The probabilities of defective items are 0.02, 0.03, 0.04. An item is found defective. Find the probability that it is from (i) A (ii) B (iii) C.

Oct/Nov-2021 · 9M

Machine	Prior	P(D/·)	Product	Posterior
A	0.2	0.02	0.004	4/33 = 0.121
B	0.3	0.03	0.009	9/33 = 0.273
C	0.5	0.04	0.020	20/33 = 0.606
P(D)			0.033	1 ✓

Unit 2 Random Variables & Distributions

A. Theory questions — short model answers

PYQ. State the PMF and PDF of random variables. Oct-2025 · 5M

PMF (discrete): $p(x_i) = P(X = x_i)$ with $p(x_i) \geq 0$, $\sum p(x_i) = 1$. **PDF** (continuous): $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$, $P(a \leq X \leq b) = \int_a^b f(x)dx$.

PYQ. Compare Binomial, Poisson and Normal distributions with regard to mean and variance.

Oct-2025 · 5M

Binomial: mean np , variance npq (variance $<$ mean). Poisson: mean = variance = λ . Normal: mean μ , variance σ^2 (independent). Full table: Unit 2, §2.8.

PYQ. State the properties of the Normal distribution. Oct-2025 · 3M

Bell-shaped, symmetric about μ ; mean = median = mode; total area 1; asymptotic to the x-axis; inflection points at $\mu \pm \sigma$; 68.26 / 95.44 / 99.73 % within 1, 2, 3 σ .

PYQ. Explain the terms population and sample. Mar/Apr-2024 · 2M

Population: the full set of observations under study (size N). Sample: a part of the population examined to learn about it (size n).

B. Numerical and derivation questions — fully solved

PYQ. A pair of fair dice is rolled. Let X = maximum of the two numbers. Find the probability distribution of X , its mean and variance. Oct-2023 · 3M Class notes

$n(S) = 36$. $X = k$ when the larger number is k : there are $2k - 1$ such pairs. So $P(X = k) = \frac{2k - 1}{36}$.

x	1	2	3	4	5	6
$P(x)$	1/36	3/36	5/36	7/36	9/36	11/36

Mean $\mu = \frac{1+6+15+28+45+66}{36} = \frac{161}{36} = 4.472$. $\sum x^2 p = \frac{791}{36} = 21.972$.
 $\sigma^2 = 21.972 - (4.472)^2 = 1.971$; $\sigma = 1.404$.

PYQ. A die is rolled twice. X = minimum of the two numbers. Find the distribution of X , its mean and variance. Nov/Dec-2022 · 10M Class notes

$P(X = k) = \frac{13 - 2k}{36}$:

x	1	2	3	4	5	6
$P(x)$	11/36	9/36	7/36	5/36	3/36	1/36

$$\mu = \frac{11+18+21+20+15+6}{36} = \frac{91}{36} = 2.528. \quad \sum x^2 p = \frac{11+36+63+80+75+36}{36} = \frac{301}{36}.$$

$$\sigma^2 = \frac{301}{36} - \left(\frac{91}{36}\right)^2 = 1.971; \quad \sigma = 1.404. \quad \mu = 2.528, \sigma^2 = 1.971$$

PYQ. A person wins Rs. 80 if 3 heads occur, Rs. 30 if 2 heads, Rs. 10 if only 1 head, in a single toss of 3 fair coins. If the game is fair, how much must he lose if no heads occur? **Mar/Apr-2024 · 3M**

Let he lose Rs. a . A game is **fair** when the expected gain $E(X) = 0$.

No. of heads	3	2	1	0
Gain x	80	30	10	$-a$
$P(x)$	1/8	3/8	3/8	1/8

$$E(X) = \frac{80+90+30-a}{8} = 0 \Rightarrow a = 200. \quad \text{He must lose Rs. 200}$$

PYQ. A random variable X takes the values $-2, -1, 0, 1, 2$ such that $P(X = 0) = P(X < 0) = P(X > 0)$ (equal probabilities on each side). Find the probability distribution and the distribution function of X . **Jul-2021 · 7M**

$P(X < 0) + P(X = 0) + P(X > 0) = 1$, so each is $\frac{1}{3}$. Split each side equally: $P(-2) = P(-1) = \frac{1}{6}$, $P(1) = P(2) = \frac{1}{6}$.

x	-2	-1	0	1	2
$P(x)$	1/6	1/6	1/3	1/6	1/6
$F(x)$	1/6	2/6	4/6	5/6	1

PYQ. A continuous RV X has PDF $f(x) = k(1 - x^2)$ for $0 < x < 1$, and 0 elsewhere. (a) Find k and $P(X > 0.5)$. (b) Using the distribution function, find $P(X < 0.3)$ and $P(0.4 < X < 0.6)$. (c) Find the mean. **May-2025 · 10M** [Class notes](#)

Step 1 $\int_0^1 k(1 - x^2)dx = 1 \Rightarrow k \left[x - \frac{x^3}{3} \right]_0^1 = k \cdot \frac{2}{3} = 1 \Rightarrow k = \frac{3}{2}.$

Step 2 $P(X > 0.5) = \frac{3}{2} \left[x - \frac{x^3}{3} \right]_{0.5}^1 = \frac{3}{2} \left[\frac{2}{3} - \frac{11}{24} \right] = \frac{3}{2} \cdot \frac{5}{24} = \frac{5}{16} = 0.3125.$

Step 3 Distribution function: $F(x) = \int_0^x \frac{3}{2}(1 - t^2)dt = \frac{3}{2} \left(x - \frac{x^3}{3} \right), 0 \leq x \leq 1.$

Step 4 $P(X < 0.3) = F(0.3) = \frac{3}{2}(0.3 - 0.009) = 0.4365.$
 $P(0.4 < X < 0.6) = F(0.6) - F(0.4) = \frac{3}{2}(0.528) - \frac{3}{2}(0.37867) = 0.224.$

Step 5 Mean = $\int_0^1 x \cdot \frac{3}{2}(1 - x^2)dx = \frac{3}{2} \left[\frac{x^2}{2} - \frac{x^4}{4} \right]_0^1 = \frac{3}{2} \cdot \frac{1}{4} = \frac{3}{8}.$

$k = 3/2 \cdot P(X > 0.5) = 0.3125 \cdot 0.4365 \cdot 0.224 \cdot \mu = 3/8$

Class notes also find $P(0.1 < X < 0.2) = \frac{3}{2} \left[0.1 - \frac{0.007}{3} \right] = 0.1465.$

PYQ. If $f(x) = y_0 e^{-|x|}, -\infty < x < \infty$, is a PDF, find y_0 . (Class notes also: find mean, variance and $P(0 < X < 4)$.) **Apr/May-2026 · 3M** **Nov-2024 · 3M** [Class notes](#)

$e^{-|x|}$ is an even function. So $\int_{-\infty}^{\infty} y_0 e^{-|x|} dx = 2y_0 \int_0^{\infty} e^{-x} dx = 2y_0 [-e^{-x}]_0^{\infty} = 2y_0(0 + 1) = 1 \Rightarrow y_0 = \frac{1}{2}.$

Mean: $\int x \cdot \frac{1}{2} e^{-|x|} dx = 0$ (odd function). **Variance:** $\sigma^2 = 2 \cdot \frac{1}{2} \int_0^{\infty} x^2 e^{-x} dx - 0 = \frac{2!}{1^3} = 2.$

$$P(0 < X < 4) = \frac{1}{2} \int_0^4 e^{-x} dx = \frac{1}{2}(1 - e^{-4}) = 0.4908. \quad \mathbf{y_0 = 1/2, \mu = 0, \sigma^2 = 2}$$

PYQ. Find the mean and variance of $f(x) = e^{-x}$, $x > 0$. Apr/May-2023 · 3M

$$\mu = \int_0^\infty x e^{-x} dx = 1! = 1. \quad E(X^2) = \int_0^\infty x^2 e^{-x} dx = 2! = 2. \quad \sigma^2 = 2 - 1 = 1. \quad \mathbf{\mu = 1, \sigma^2 = 1}$$

PYQ. The PDF of X is $f(x) = \frac{1}{2} \sin x$ for $0 \leq x \leq \pi$, and 0 elsewhere. Find the mean, mode and median. Also find $P(0 < X < \pi/2)$. Mar/Apr-2022 · 9M Class notes

$$\text{Mean: } \mu = \frac{1}{2} \int_0^\pi x \sin x dx = \frac{1}{2} \left[-x \cos x \right]_0^\pi + \frac{1}{2} \int_0^\pi \cos x dx = \frac{1}{2}(\pi) + \frac{1}{2}[\sin x]_0^\pi = \frac{\pi}{2}.$$

$$\text{Mode: } f'(x) = \frac{1}{2} \cos x = 0 \Rightarrow x = \frac{\pi}{2}. \quad f''\left(\frac{\pi}{2}\right) = -\frac{1}{2} \sin \frac{\pi}{2} = -\frac{1}{2} < 0. \quad \text{So mode} = \frac{\pi}{2}.$$

$$\text{Median: } \int_0^M \frac{1}{2} \sin x dx = \frac{1}{2} \Rightarrow \frac{1}{2}[1 - \cos M] = \frac{1}{2} \Rightarrow \cos M = 0 \Rightarrow M = \frac{\pi}{2}.$$

$$P(0 < X < \frac{\pi}{2}) = \frac{1}{2}[-\cos x]_0^{\pi/2} = \frac{1}{2}. \quad \mathbf{Mean = Mode = Median = \pi/2}$$

PYQ. A continuous RV has PDF $f(x) = kxe^{-\lambda x}$ for $x \geq 0$, $\lambda > 0$, and 0 otherwise. Find (i) k (ii) mean (iii) variance. Nov-2020 · 7M Class notes

$$\text{Use } \int_0^\infty x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}}.$$

$$(i) k \cdot \frac{1!}{\lambda^2} = 1 \Rightarrow k = \lambda^2. \quad (ii) \mu = \lambda^2 \int_0^\infty x^2 e^{-\lambda x} dx = \lambda^2 \cdot \frac{2}{\lambda^3} = \frac{2}{\lambda}.$$

$$(iii) E(X^2) = \lambda^2 \cdot \frac{3!}{\lambda^4} = \frac{6}{\lambda^2}; \quad \sigma^2 = \frac{6}{\lambda^2} - \frac{4}{\lambda^2} = \frac{2}{\lambda^2}.$$

PYQ. If X is a continuous RV and $Y = aX + b$, prove that $E(Y) = aE(X) + b$ and $Var(Y) = a^2 Var(X)$.

Nov/Dec-2022 · 3M

$$E(aX + b) = \int (ax + b)f(x)dx = a \int xf(x)dx + b \int f(x)dx = aE(X) + b \cdot 1.$$

$$Var(Y) = E[(Y - E(Y))^2] = E[(aX + b - aE(X) - b)^2] = E[a^2(X - E(X))^2] = a^2 E[(X - \mu)^2] = a^2 Var(X). \blacksquare$$

PYQ. $X \sim B(n = 10, p = 0.5)$. Find $P(X = 6)$. Oct-2025 · 5M

$$P(6) = {}^{10}C_6 (0.5)^{10} = \frac{210}{1024} = 0.2051.$$

PYQ. The mean of a binomial distribution is 3 and the variance is 9/4. Find (i) n (ii) $P(1 \leq X < 6)$.

Nov-2024 · 5M

$$np = 3, npq = \frac{9}{4} \Rightarrow q = \frac{9/4}{3} = \frac{3}{4}, p = \frac{1}{4}, n = \frac{3}{1/4} = 12.$$

$$P(1 \leq X < 6) = \sum_{r=1}^5 {}^{12}C_r \left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{12-r} = 0.9139. \quad \mathbf{n = 12, P = 0.9139}$$

PYQ. In a binomial distribution of 5 independent trials, the probabilities of 1 and 2 successes are 0.4096 and 0.2048. Find the parameter p , mean and variance. Mar/Apr-2022 · 6M Class notes

$$\frac{P(2)}{P(1)} = \frac{{}^5C_2 p^2 q^3}{{}^5C_1 p q^4} = \frac{2p}{q} = \frac{0.2048}{0.4096} = \frac{1}{2} \Rightarrow 4p = q = 1 - p \Rightarrow p = \frac{1}{5}.$$

$$\text{Mean} = np = 1; \text{variance} = npq = 5 \cdot \frac{1}{5} \cdot \frac{4}{5} = 0.8.$$

PYQ. The probability of a man hitting a target is $1/3$. (a) If he fires 5 times, find the probability of at least 2 hits. (b) How many times must he fire so that the probability of at least one hit is more than 90%? Mar/Apr-2024 · 10M Class notes

$$(a) P(X \geq 2) = 1 - P(0) - P(1) = 1 - \left(\frac{2}{3}\right)^5 - 5 \cdot \frac{1}{3} \left(\frac{2}{3}\right)^4 = 1 - \frac{32}{243} - \frac{80}{243} = \frac{131}{243} = 0.5391.$$

$$(b) 1 - \left(\frac{2}{3}\right)^n > 0.9 \Rightarrow \left(\frac{2}{3}\right)^n < 0.1. \text{ Try: } n = 5: 0.132; n = 6: 0.088 \checkmark. \quad \mathbf{n = 6}$$

PYQ. Out of 800 families with 5 children each, how many would you expect to have (i) 3 boys (ii) 5 girls (iii) 2 or 3 boys (iv) at least one boy? Assume equal probabilities for boys and girls. Nov-2020 · 8M Class notes

$$n = 5, p = q = \frac{1}{2}, P(r) = \frac{{}^5C_r}{32}. \text{ Expected no.} = N \cdot P \text{ with } N = 800.$$

$$(i) 800 \cdot \frac{10}{32} = 250 \quad (ii) 800 \cdot \frac{1}{32} = 25 \quad (iii) 800 \cdot \frac{20}{32} = 500 \quad (iv) 800 \cdot \frac{31}{32} = 775$$

PYQ. Estimate the probability that out of 2000 persons exactly 5 will suffer a bad reaction, if the probability of a bad reaction is 0.001. Mar/Apr-2024 · 5M Class notes (exactly 3, more than 2, none, more than 1)

$$\lambda = np = 2000 \times 0.001 = 2. \quad P(5) = \frac{e^{-2} 2^5}{5!} = \frac{32}{120} e^{-2} = 0.0361.$$

$$\text{Class version: } P(3) = \frac{4}{3} e^{-2} = 0.1804; P(X > 2) = 1 - e^{-2}(1 + 2 + 2) = 0.3233; P(0) = e^{-2} = 0.1353; P(X > 1) = 1 - 3e^{-2} = 0.5940.$$

(The class notes write 0.1604 for $P(3)$. The correct value is 0.1804.)

PYQ. A car hire firm has two cars. The number of demands per day is Poisson with mean 1.5. Find the proportion of days on which (i) neither car is used (ii) some demand is refused. Oct-2023 · 5M

$$\lambda = 1.5. (i) P(0) = e^{-1.5} = 0.2231.$$

$$(ii) \text{ A demand is refused when demand } > 2: P(X > 2) = 1 - e^{-1.5} \left[1 + 1.5 + \frac{1.5^2}{2} \right] = 1 -$$

$$0.2231(3.625) = 0.1912. \quad \mathbf{\approx 19\% \text{ of days}}$$

PYQ. A Poisson variate has $P(X = 1) = P(X = 2)$. Find (i) the mean (ii) $P(X \geq 1)$ (iii) $P(1 \leq X \leq 4)$. Mar/Apr-2022 · 7M Class notes

$$e^{-\lambda} \lambda = \frac{e^{-\lambda} \lambda^2}{2} \Rightarrow \lambda = 2 \ (\lambda \neq 0). \text{ Mean} = 2.$$

$$P(X \geq 1) = 1 - e^{-2} = 0.8647. \quad P(1 \leq X \leq 4) = e^{-2} \left[2 + 2 + \frac{4}{3} + \frac{2}{3} \right] = 6e^{-2} = 0.8120.$$

$$\text{Class notes also: } P(4) = \frac{2}{3} e^{-2} = 0.0902; P(1 < X < 4) = P(2) + P(3) = 0.4511.$$

PYQ. X is Poisson with mean 2. Find (i) $P(X = 2)$ (ii) $P(1 \leq X < 3)$ (iii) $P(X = 0)$. Jul-2021 · 8M

$$(i) 2e^{-2} = 0.2707 \quad (ii) P(1) + P(2) = 4e^{-2} = 0.5413 \quad (iii) e^{-2} = 0.1353$$

PYQ. 10% of bean seeds do not germinate. Seeds are sold in packets of 200 with a guarantee of 90% germination. Find the probability that a packet violates the guarantee. Apr/May-2023 · 5M

Class notes (5%)

The guarantee fails if more than $200 \times 0.10 = 20$ seeds do not germinate. $\lambda = np = 200 \times 0.10 = 20$.

$$P(X > 20) = 1 - \sum_{r=0}^{20} \frac{e^{-20} 20^r}{r!} = 0.4409.$$

Class notes version (5% fail): $\lambda = 10$, $P(X > 20) = 1 - \sum_{r=0}^{20} \frac{e^{-10} 10^r}{r!} = 0.0016$.

PYQ. Find the area under the standard normal curve (i) to the left of $z = -1.78$ (ii) between $z = -0.80$ and $z = 1.53$. May-2025 · 3M

(i) $P(Z < -1.78) = 0.5 - A(1.78) = 0.5 - 0.4625 = 0.0375$.

(ii) $A(0.80) + A(1.53) = 0.2881 + 0.4370 = 0.7251$.

PYQ. X is normal with mean 30 and s.d. 5. Find (i) $P(26 \leq X \leq 40)$ (ii) $P(|X - 30| > 5)$.

Nov-2024 · 5M

Class notes

(i) $z_1 = \frac{26-30}{5} = -0.8$, $z_2 = \frac{40-30}{5} = 2$. Opposite sides: $A(0.8) + A(2) = 0.2881 + 0.4772 = 0.7653$.

(ii) $|X - 30| \leq 5 \iff 25 \leq X \leq 35 \iff -1 \leq Z \leq 1$: probability $= 2A(1) = 0.6826$.

$P(|X - 30| > 5) = 1 - 0.6826 = 0.3174$. (i) 0.7653 (ii) 0.3174

Class notes also: $P(X \geq 45) = P(Z \geq 3) = 0.5 - 0.4987 = 0.0013$.

PYQ. For a normal variate with mean 1 and s.d. 3, find (i) $P(3.45 \leq X \leq 6.19)$ (ii) $P(-1.43 \leq X \leq 6.19)$.

Nov-2020 · 7M

Class notes

(i) $z_1 = \frac{3.45-1}{3} = 0.82$, $z_2 = \frac{6.19-1}{3} = 1.73$. Same side: $A(1.73) - A(0.82) = 0.4582 - 0.2939 = 0.1643$.

(ii) $z_1 = \frac{-1.43-1}{3} = -0.81$: $A(0.81) + A(1.73) = 0.2910 + 0.4582 = 0.7492$.

PYQ. The masses of 300 students are normally distributed with mean 68 kg and s.d. 3 kg. How many students have masses (a) greater than 72 kg (b) less than or equal to 64 kg (c) between 65 and 71 kg inclusive? Nov/Dec-2022 · 10M Class notes

(a) $z = \frac{72-68}{3} = 1.33$. $P = 0.5 - A(1.33) = 0.5 - 0.4082 = 0.0918$. No. = $300 \times 0.0918 = 27.5 \approx 28$.

(b) $z = \frac{64-68}{3} = -1.33$. $P = 0.5 - 0.4082 = 0.0918$. No. ≈ 28 .

(c) $z_1 = -1$, $z_2 = 1$. $P = 2(0.3413) = 0.6826$. No. = $300 \times 0.6826 = 204.8 \approx 205$.

PYQ. The marks of 1000 students are normally distributed with mean 78% and s.d. 11%. (a) How many students got marks above 90%? (b) What was the highest mark of the lowest 10%? (c) Within what limits did the middle 90% lie? Apr/May-2023 · 10M Class notes

(a) $z = \frac{90-78}{11} = 1.09$. $P = 0.5 - A(1.09) = 0.5 - 0.3621 = 0.1379$. No. = $1000 \times 0.1379 \approx 138$.

(b) Lowest 10%: area from z_1 to 0 is 0.4. Table: $z_1 = -1.28$. $x = 78 - 1.28(11) = 63.9 \approx 64\%$.

(c) Middle 90%: 0.05 in each tail, so $A(z) = 0.45 \Rightarrow z = 1.645$. Limits = $78 \pm 1.645(11) = 78 \pm 18.1$.

≈ 60% to 96%

PYQ. The weights of 800 male students are normally distributed with mean 140 lb and s.d. 10 lb. Find the number of students whose weights are (i) between 138 and 148 lb (ii) more than 152 lb.

Mar/Apr-2022 · 8M

Class notes

(i) $z_1 = -0.2$, $z_2 = 0.8$: $A(0.2) + A(0.8) = 0.0793 + 0.2881 = 0.3674$. No. = $800 \times 0.3674 \approx 294$.

(ii) $z = 1.2$: $0.5 - 0.3849 = 0.1151$. No. = $800 \times 0.1151 \approx 92$.

The 2022 paper prints $\mu = 100$. That is a misprint: with $\mu = 100$ the answer is almost 0. The standard problem (Schaum) uses $\mu = 140$.

PYQ. Find the mean and s.d. of a normal distribution in which 7% of items are under 35 and 89% are under 63. Mar/Apr-2024 · 5M Class notes

$$A(z_1) = 0.5 - 0.07 = 0.43 \Rightarrow z_1 = -1.48 \text{ (left)}. A(z_2) = 0.89 - 0.5 = 0.39 \Rightarrow z_2 = 1.23.$$

$$\frac{35-\mu}{\sigma} = -1.48 \dots (1); \frac{63-\mu}{\sigma} = 1.23 \dots (2). \text{ Subtract: } \frac{28}{\sigma} = 2.71 \Rightarrow \sigma = 10.33; \mu = 35 + 1.48(10.33) = 50.29.$$

$$\mu \approx 50.3, \sigma \approx 10.3$$

PYQ. In a normal distribution, 10.03% of items are under 25 kg and 89.97% are under 70 kg. Find the mean and s.d. Oct-2023 · 5M

$$A(z_1) = 0.5 - 0.1003 = 0.3997 \Rightarrow z_1 = -1.28. A(z_2) = 0.8997 - 0.5 = 0.3997 \Rightarrow z_2 = 1.28.$$

The two points are symmetric, so $\mu = \frac{25+70}{2} = 47.5. \frac{70-47.5}{\sigma} = 1.28 \Rightarrow \sigma = 17.58.$

$$\mu = 47.5 \text{ kg}, \sigma = 17.58 \text{ kg}$$

PYQ. A mean of a normal population equals the standard error of the mean of samples of size 64. Find the probability that the mean of a sample of size 64 will be negative. Apr/May-2023 · 5M

$$\mu = \frac{\sigma}{\sqrt{64}} = \frac{\sigma}{8}. Z = \frac{\bar{x}-\mu}{\sigma/8}. \text{ For } \bar{x} = 0: Z = \frac{0-\sigma/8}{\sigma/8} = -1.$$

$$P(\bar{x} < 0) = P(Z < -1) = 0.5 - 0.3413 = 0.1587.$$

PYQ. A population consists of the five numbers 2, 3, 6, 8, 11. Consider all possible samples of size 2 drawn **with replacement**. Find (a) the mean of the population (b) the s.d. of the population (c) the mean of the sampling distribution of means (d) the s.d. of the sampling distribution of means.

Class notes
Nov-2024 (size 3)

$$(a) \mu = \frac{2+3+6+8+11}{5} = 6. (b) \sigma^2 = \frac{16+9+0+4+25}{5} = 10.8, \sigma = 3.286.$$

$$(c) \text{ No. of samples} = N^n = 5^2 = 25. \text{ The 25 sample means add to 150. } \mu_{\bar{x}} = \frac{150}{25} = 6.$$

	2	3	6	8	11
2	2	2.5	4	5	6.5
3	2.5	3	4.5	5.5	7
6	4	4.5	6	7	8.5
8	5	5.5	7	8	9.5
11	6.5	7	8.5	9.5	11

$$(d) \sum (\bar{x} - 6)^2 = 135. \sigma_{\bar{x}}^2 = \frac{135}{25} = 5.4, \sigma_{\bar{x}} = 2.32.$$

$$\text{Check by formula: } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.286}{\sqrt{2}} = 2.32 \checkmark.$$

$$\text{Nov-2024 version (size 3, with replacement): } 5^3 = 125 \text{ samples; } \mu_{\bar{x}} = 6; \sigma_{\bar{x}} = \frac{3.286}{\sqrt{3}} = 1.897.$$

$$6, 3.286, 6, 1.897$$

PYQ. A population consists of 2, 3, 5, 6, 8. Consider all samples of size 3 drawn **without replacement**. Find (a) μ (b) σ (c) mean of the sampling distribution of means (d) s.d. of the sampling distribution of means. May-2025 · 10M

$$(a) \mu = \frac{24}{5} = 4.8. (b) \sigma^2 = \frac{\sum x^2}{5} - \mu^2 = \frac{138}{5} - 23.04 = 4.56, \sigma = 2.135.$$

$$(c) {}^5C_3 = 10 \text{ samples. Means: } 3.33, 3.67, 4.33, 4.33, 4.67, 5, 5.33, 5.33, 5.67, 6.33. \text{ Their sum is 48, so}$$

their mean = $4.8 = \mu$.

(d) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{2.135}{\sqrt{3}} \sqrt{\frac{2}{4}} = 0.872$. **4.8, 2.135, 4.8, 0.872**

PYQ. A population consists of 3, 6, 9, 15, 27. Consider all samples of size 2 drawn without replacement. Find (a)–(d) as above. **Oct-2023 · 10M**

$\mu = 12$; $\sigma^2 = \frac{81+36+9+9+225}{5} = 72$, $\sigma = 8.485$; ${}^5C_2 = 10$ samples; $\mu_{\bar{x}} = 12$;

$\sigma_{\bar{x}} = \frac{8.485}{\sqrt{2}} \sqrt{\frac{3}{4}} = 5.196$. **12, 8.485, 12, 5.196**

PYQ. Populations 5, 10, 14, 15, 18, 24 (Oct-2021) and 4, 8, 12, 16, 20, 24 (Nov-2020): all samples of size 2 without replacement. **Oct/Nov-2021 · 15M** **Nov-2020 · 8M**

Population	μ	σ^2 (σ)	No. of samples	$\mu_{\bar{x}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{2}} \sqrt{\frac{4}{5}}$
5,10,14,15,18,24	14.33	35.56 (5.963)	15	14.33	3.771
4,8,12,16,20,24	14	46.67 (6.831)	15	14	4.320 (variance 18.67)

Class notes use 5, 10, 14, 18, 13, 24: $\mu = 14$, $\sigma = 5.97$, $\sigma_{\bar{x}} = 3.78$.

PYQ. If the mean and s.d. of a sampling distribution are 100 and 15, find the standard error for a sample of size 25. **Oct-2025 · 5M** **Class notes**

S.E. = $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$.

Unit 3 Testing of Hypothesis I (Large Samples)

A. Theory questions — short model answers

PYQ. Write all the steps involved in testing a hypothesis.

Apr/May-2026 · 3M

Nov/Dec-2022

Oct/Nov-2021

(1) State H_0 (2) State H_1 (3) Select α (4) Calculate the test statistic $Z = \frac{t - E(t)}{S.E.(t)}$ (5) Read the critical value (6) If $|Z| < z_\alpha$ accept H_0 , else reject. Unit 3, §3.7.

PYQ. Define critical region and level of significance.

Apr/May-2026 · 2M

Nov-2024 · 2M

LOS α : the maximum probability of a Type I error we accept (5% or 1%). **Critical region:** the region of the sampling distribution where we reject H_0 ; its area is α .

PYQ. Define Type I and Type II errors.

Oct-2025 · 2M

Apr/May-2023

Jul-2021

Type I: reject H_0 when it is true (probability α). Type II: accept H_0 when it is false (probability β).

PYQ. Null and alternative hypothesis.

Mar/Apr-2024 · 3M

Oct-2023

Nov/Dec-2022

H_0 : the hypothesis of no difference, tested for rejection (e.g. $\mu = 60$). H_1 : the complement accepted when H_0 is rejected ($\mu \neq 60$, $\mu > 60$, $\mu < 60$).

PYQ. Importance of estimation; types of estimators; unbiased estimator; properties of a good estimator.

Apr/May-2026

May-2025

Apr/May-2023

Estimation finds an unknown parameter from a sample. Types: point and interval. Unbiased: $E(\hat{\theta}) = \theta$. Good estimator: unbiased, consistent, efficient, sufficient. Unit 3, §3.1–3.2.

PYQ. Explain the test for equality of two means based on large samples.

Nov-2020 · 6M

$H_0 : \mu_1 = \mu_2$; $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; compare with 1.96 (5%). Unit 3, §3.9.

B. Numerical and derivation questions — fully solved

PYQ. Define point and interval estimation. Explain with examples.

Oct-2025 · 5M

Use the definitions in 3.2 and 3.3. Example: 50 students, $\bar{x} = 62$ kg, $\sigma = 5$ kg.

Point estimate of μ : 62 kg. 95% interval estimate: $62 \pm 1.96 \cdot \frac{5}{\sqrt{50}} = 62 \pm 1.386 = (60.61, 63.39)$ kg.

PYQ. To estimate the average time visitors take to move from one building to another, a sample of size n is used. Given $\sigma = 1.40$ minutes, how large should the sample be to be 99% confident that the error E is at most 0.25? **May-2025 · 3M** **Nov-2024 · 3M**

$$n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 = \left(\frac{2.58 \times 1.40}{0.25} \right)^2 = (14.448)^2 = 208.7. \text{ Round up. } \mathbf{n = 209}$$

PYQ. A random sample of 400 items has mean 82 and s.d. 18. Find the maximum error of estimate at 95% confidence. **Nov/Dec-2022 · 3M**

$$E = 1.96 \times \frac{18}{\sqrt{400}} = 1.96 \times 0.9 = 1.764. \mathbf{E = 1.764}$$

PYQ. The efficiency expert of a computer company tested 40 engineers. The mean time to assemble a component is 12.73 min and s.d. 2.06 min. (a) If $\bar{x} = 12.73$ is used as a point estimate, find the maximum error with 99% confidence. (b) Construct a 98% confidence interval. (c) With what confidence can we assert that the sample mean does not differ from the true mean by more than 30 seconds? **Mar/Apr-2024 · 10M** **Oct-2023 · 3M (part c)**

$$\text{S.E.} = \frac{s}{\sqrt{n}} = \frac{2.06}{\sqrt{40}} = 0.3257.$$

Step 1 (a) $E = 2.58 \times 0.3257 = 0.840$ min.

Step 2 (b) $12.73 \pm 2.33 \times 0.3257 = 12.73 \pm 0.759 = (11.97, 13.49)$ min.

Step 3 (c) 30 seconds = 0.5 min. $z = \frac{E}{\text{S.E.}} = \frac{0.5}{0.3257} = 1.535$. Confidence = $2A(1.535) = 2(0.4376) = 0.875$.

(a) 0.84 min (b) (11.97, 13.49) (c) $\approx 87.5\%$

PYQ. A random sample of 100 is taken from a population with $\sigma = 5.1$. The sample mean is 21.6. Construct a 95% confidence interval for μ . **Mar/Apr-2022 · 6M**

$$21.6 \pm 1.96 \times \frac{5.1}{10} = 21.6 \pm 1.0 = (20.6, 22.6).$$

Same method: $n = 400$, s.d. 10, mean 60 $\rightarrow (59.02, 60.98)$ **Oct/Nov-2021**; $n = 200$, mean 50, s.d. 9 $\rightarrow (48.75, 51.25)$

Mar/Apr-2022.

PYQ. Construct a 95% confidence interval for the true proportion of computer literates if 47 out of 150 persons from rural areas are computer literates. **Apr/May-2023 · 5M** **Jul-2021 · 7M**

$$p = \frac{47}{150} = 0.3133, q = 0.6867. \text{ S.E.} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.3133 \times 0.6867}{150}} = 0.0379.$$

$$\text{CI} = 0.3133 \pm 1.96(0.0379) = 0.3133 \pm 0.0742 = (0.239, 0.388). \mathbf{23.9\% \text{ to } 38.8\%}$$

PYQ. Explain the difference between one-tailed and two-tailed tests. Oct-2025 · 3M

One-tailed test	Two-tailed test
H_1 has $>$ or $<$.	H_1 has $\neq$.
Critical region is in one tail; area α .	Critical region is in both tails; area $\alpha/2$ each.
At 5%: $z_\alpha = 1.645$.	At 5%: $z_\alpha = 1.96$.
Example: new bulbs are better ($\mu > 1000$).	Example: mean life is 1000 ($\mu \neq 1000$).

PYQ. A sample of 50 students has a mean weight of 62 kg. The population s.d. is 5 kg. Formulate H_0 and H_1 to test if the population mean is 60 kg at 5% level. Oct-2025 · 5M

Step 1 $H_0 : \mu = 60$ kg.

Step 2 $H_1 : \mu \neq 60$ kg (two-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $Z = \frac{62 - 60}{5/\sqrt{50}} = \frac{2}{0.7071} = 2.83$.

Step 5 Tabulated $z_{0.05} = 1.96$.

Step 6 $|Z| = 2.83 > 1.96$. Reject H_0 . The population mean is not 60 kg.

PYQ. A sample of 400 items is taken from a population whose s.d. is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate the 95% confidence interval for the population mean. May-2025 · 10M

Step 1 $H_0 : \mu = 38$.

Step 2 $H_1 : \mu \neq 38$ (two-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $Z = \frac{40 - 38}{10/\sqrt{400}} = \frac{2}{0.5} = 4$.

Step 5 $z_{0.05} = 1.96$.

Step 6 $4 > 1.96$. Reject H_0 . The sample is not from a population with mean 38.

95% CI: $40 \pm 1.96(0.5) = 40 \pm 0.98 = (39.02, 40.98)$. Z = 4, reject; CI (39.02, 40.98)

PYQ. An oceanographer wants to check whether the depth of the ocean in a region is 57.4 fathoms, as previously recorded. At $\alpha = 0.05$, what can he conclude if soundings at 40 random locations gave a mean of 59.1 fathoms with s.d. 5.2? Also calculate the 95% confidence interval. Mar/Apr-2024 · 10M

Step 1 $H_0 : \mu = 57.4$.

Step 2 $H_1 : \mu \neq 57.4$.

Step 3 $\alpha = 0.05$.

Step 4 $S.E. = \frac{5.2}{\sqrt{40}} = 0.822. Z = \frac{59.1 - 57.4}{0.822} = 2.07.$

Step 5 $z_{0.05} = 1.96.$

Step 6 $2.07 > 1.96.$ Reject H_0 . The depth is not 57.4 fathoms.

95% CI: $59.1 \pm 1.96(0.822) = (57.49, 60.71)$ fathoms.

PYQ. The average depth of ocean in a region is recorded as 67.4 fathoms. Is there reason to believe this at the 0.01 level, if readings at 40 locations showed a mean of 69.3 with s.d. 5.4? **Oct-2023 · 10M**

$$H_0 : \mu = 67.4; H_1 : \mu \neq 67.4; \alpha = 0.01. Z = \frac{69.3 - 67.4}{5.4/\sqrt{40}} = \frac{1.9}{0.8538} = 2.23.$$

$z_{0.01} = 2.58. 2.23 < 2.58.$ Accept H_0 . The recorded depth 67.4 is acceptable.

PYQ. It is claimed that a random sample of 49 tires has a mean life of 15200 km and is drawn from a population with mean 15150 km and s.d. 1200 km. Test the claim at (i) 0.01 (ii) 0.05. **Nov/Dec-2022 · 5M**

$$H_0 : \mu = 15150; H_1 : \mu \neq 15150. Z = \frac{15200 - 15150}{1200/7} = \frac{50}{171.43} = 0.29.$$

$0.29 < 2.58$ and $0.29 < 1.96.$ Accept H_0 at both levels. The claim is valid.

PYQ. A trucking firm suspects the claim that the average life of certain tires is at least 28,000 miles. 40 tires gave a mean life of 27,463 miles with s.d. 1,348 miles. What can it conclude if the probability of a Type I error is at most 0.01? **Mar/Apr-2022 · 8M**

Step 1 $H_0 : \mu = 28000.$

Step 2 $H_1 : \mu < 28000$ (left-tailed; the claim "at least" is doubted).

Step 3 $\alpha = 0.01.$

Step 4 $Z = \frac{27463 - 28000}{1348/\sqrt{40}} = \frac{-537}{213.1} = -2.52.$

Step 5 Left-tailed $z_{0.01} = -2.33.$

Step 6 $-2.52 < -2.33:$ Z is in the critical region. Reject H_0 . The claim is not true.

PYQ. The means of two large samples of sizes 1000 and 2000 are 67.5 and 68.0 inches. Can the samples be regarded as drawn from the same population of s.d. 2.5 inches? **May-2025 · 10M**

Mar/Apr-2022 · 9M

Step 1 $H_0 : \mu_1 = \mu_2$ (same population).

Step 2 $H_1 : \mu_1 \neq \mu_2.$

Step 3 $\alpha = 0.05.$

Step 4 $Z = \frac{67.5 - 68.0}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = \frac{-0.5}{2.5 \times 0.03873} = -5.16.$

Step 5 $z_{0.05} = 1.96.$

Step 6 $|Z| = 5.16 > 1.96$. Reject H_0 . The samples are not from the same population.

PYQ. Test the significance of the difference between two sample means: $n_1 = 40$, $\bar{x}_1 = 78$, $s_1^2 = 40$; $n_2 = 35$, $\bar{x}_2 = 72$, $s_2^2 = 35$ at 5% level. **Oct-2025 · 5M**

$$H_0 : \mu_1 = \mu_2; H_1 : \mu_1 \neq \mu_2. Z = \frac{78 - 72}{\sqrt{\frac{40}{40} + \frac{35}{35}}} = \frac{6}{\sqrt{2}} = 4.24 > 1.96. \text{ Reject } H_0. \text{ The difference is significant.}$$

PYQ. A company claims that its bulbs are superior to those of its main competitor. A sample of 40 bulbs has mean life 647 h with s.d. 27 h. A sample of 40 bulbs of the competitor has mean life 638 h with s.d. 31 h. Does this substantiate the claim at the 0.05 level? **Nov-2024 · 10M**

Step 1 $H_0 : \mu_1 = \mu_2$.

Step 2 $H_1 : \mu_1 > \mu_2$ ("superior" $\rightarrow$ right-tailed).

Step 3 $\alpha = 0.05$.

Step 4
$$Z = \frac{647 - 638}{\sqrt{\frac{27^2}{40} + \frac{31^2}{40}}} = \frac{9}{\sqrt{18.225 + 24.025}} = \frac{9}{6.5} = 1.38.$$

Step 5 Right-tailed $z_{0.05} = 1.645$.

Step 6 $1.38 < 1.645$. Accept H_0 . The data do not support the claim of superiority.

PYQ. A sample of 100 bulbs by manufacturer A has mean life 1190 h, s.d. 90 h. A sample of 75 bulbs by B has mean life 1230 h, s.d. 120 h. Is there a difference between the mean lives at the 0.05 level? **Apr/May-2023 · 5M**

$$H_0 : \mu_1 = \mu_2; H_1 : \mu_1 \neq \mu_2. Z = \frac{1190 - 1230}{\sqrt{\frac{8100}{100} + \frac{14400}{75}}} = \frac{-40}{\sqrt{81 + 192}} = \frac{-40}{16.52} = -2.42.$$

$|Z| = 2.42 > 1.96$. Reject H_0 . The mean lives are different.

PYQ. 6400 men have mean weight 67.85 kg, s.d. 2.56 kg. 1600 women have mean 68.55 kg, s.d. 2.52 kg. Do the data indicate that men are, on average, heavier than women? **Nov/Dec-2022 · 5M**

$$H_0 : \mu_1 = \mu_2; H_1 : \mu_1 > \mu_2 \text{ (right-tailed)}. Z = \frac{67.85 - 68.55}{\sqrt{\frac{2.56^2}{6400} + \frac{2.52^2}{1600}}} = \frac{-0.70}{0.0707} = -9.91.$$

Right-tailed $z_{0.05} = 1.645$. Z is negative, so it is not in the right-tail critical region. Accept H_0 : the data do **not** show that men are heavier. (In fact $|Z|$ is very large, so women are significantly heavier in this sample.)

PYQ. In a random sample of 1200 cars making a right turn at a junction, 314 drove into the wrong lane. Test whether actually 30% of all drivers make this mistake. Use (a) 0.05 (b) 0.01 level. Comment on the results. **Nov-2024 · 10M**

Step 1 $H_0 : P = 0.30$.

Step 2 $H_1 : P \neq 0.30$.

Step 3 (a) $\alpha = 0.05$, (b) $\alpha = 0.01$.

Step 4 $p = \frac{314}{1200} = 0.2617$. $Z = \frac{0.2617 - 0.30}{\sqrt{0.3 \times 0.7/1200}} = \frac{-0.0383}{0.01323} = -2.90$.

Step 5 (a) 1.96 (b) 2.58.

Step 6 $|Z| = 2.90$ is greater than both. Reject H_0 at both levels. Comment: the true proportion is not 30%; it is less than 30%. The result is strong, because it is significant even at 1%.

PYQ. A manufacturer claims that only 4% of his products are defective. A random sample of 500 had 100 defective. Test the hypothesis at the 0.05 level. Mar/Apr-2022 · 6M

$H_0 : P = 0.04; H_1 : P > 0.04$ (right-tailed). $p = \frac{100}{500} = 0.2$.

$Z = \frac{0.2 - 0.04}{\sqrt{0.04 \times 0.96/500}} = \frac{0.16}{0.00876} = 18.26 > 1.645$. Reject H_0 . The claim is false.

PYQ. If 57 out of 150 patients with a disease are cured by allopathy and 33 out of 100 by homeopathy, is there reason to believe that allopathy is better than homeopathy at the 0.05 level? Oct-2023 · 10M Oct/Nov-2021 · 8M

Step 1 $H_0 : P_1 = P_2$.

Step 2 $H_1 : P_1 > P_2$ ("better" $\rightarrow$ right-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $p_1 = \frac{57}{150} = 0.38$, $p_2 = \frac{33}{100} = 0.33$. $P = \frac{57+33}{250} = 0.36$, $Q = 0.64$.
S.E. = $\sqrt{0.36 \times 0.64 \left(\frac{1}{150} + \frac{1}{100}\right)} = \sqrt{0.2304 \times 0.01667} = 0.0620$. $Z = \frac{0.05}{0.062} = 0.81$.

Step 5 Right-tailed $z_{0.05} = 1.645$.

Step 6 $0.81 < 1.645$. Accept H_0 . There is no reason to believe allopathy is better.

PYQ. The owner of a machine shop tests two snack vending machines 250 times each. The first fails 13 times and the second fails 7 times. Test at 0.05 whether the difference between the sample proportions is significant. Apr/May-2023 · 5M Jul-2021 · 8M

$p_1 = \frac{13}{250} = 0.052$, $p_2 = \frac{7}{250} = 0.028$. $P = \frac{20}{500} = 0.04$, $Q = 0.96$.

$Z = \frac{0.052 - 0.028}{\sqrt{0.04 \times 0.96 \left(\frac{2}{250}\right)}} = \frac{0.024}{0.01753} = 1.37 < 1.96$. Accept H_0 . The difference is not significant.

PYQ. In a city, 125 men in a sample of 500 were smokers. In another city, 375 in a sample of 1000 were smokers. Does this indicate a greater proportion of smokers in the second city? Apr/May-2023 · 5M

$H_0 : P_1 = P_2; H_1 : P_1 < P_2$ (left-tailed). $p_1 = 0.25$, $p_2 = 0.375$. $P = \frac{500}{1500} = \frac{1}{3}$, $Q = \frac{2}{3}$.

$Z = \frac{0.25 - 0.375}{\sqrt{\frac{1}{3} \cdot \frac{2}{3} \left(\frac{1}{500} + \frac{1}{1000}\right)}} = \frac{-0.125}{0.0258} = -4.84 < -1.645$. Reject H_0 . Yes, the second city has a greater proportion of smokers.

PYQ. In city A, 20% of a random sample of 900 school boys had a slight physical defect. In city B, 18.5% of 1600 boys had the same defect. Is the difference between the populations significant?

Nov/Dec-2022 · 5M

$$P = \frac{900(0.2) + 1600(0.185)}{2500} = \frac{180 + 296}{2500} = 0.1904, Q = 0.8096.$$

$$Z = \frac{0.2 - 0.185}{\sqrt{0.1904 \times 0.8096 \left(\frac{1}{900} + \frac{1}{1600}\right)}} = \frac{0.015}{0.01636} = 0.92 < 1.96. \text{ Accept } H_0. \text{ The difference is not significant.}$$

PYQ. A company surveys a new plan. In one group of 500 workers, 62% favoured the plan. In another group of 400 workers, 41% were against it. Is there a significant difference between the two groups at the 2% level? Oct/Nov-2021 · 8M

Group 2 favour: $100\% - 41\% = 59\%$. $p_1 = 0.62, p_2 = 0.59. P = \frac{310+236}{900} = 0.6067, Q = 0.3933.$

$$Z = \frac{0.03}{\sqrt{0.6067 \times 0.3933 \left(\frac{1}{500} + \frac{1}{400}\right)}} = \frac{0.03}{0.0328} = 0.92. \text{ At 2\% (two-tailed) } z = 2.33. \text{ Accept } H_0. \text{ No significant difference.}$$

PYQ. 400 men and 600 women were asked about a flyover. 200 men and 325 women were in favour. Test whether the proportions of men and women in favour are the same at 5%. Nov-2020 · 9M

$$p_1 = 0.5, p_2 = 0.5417. P = \frac{525}{1000} = 0.525. Z = \frac{-0.0417}{\sqrt{0.525 \times 0.475 \left(\frac{1}{400} + \frac{1}{600}\right)}} = \frac{-0.0417}{0.0322} = -1.29. |Z| < 1.96$$

. Accept H_0 .

Unit 4 Testing of Hypothesis II (Small Samples)

A. Theory questions — short model answers

PYQ. State the t-distribution and its applications. May-2025 · 2M Oct-2025 · 2M

$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ with $\nu = n - 1$; PDF $y_0(1 + t^2/\nu)^{-(\nu+1)/2}$. Applications: single mean, two means, paired data, CI for μ . Unit 4, §4.2.

PYQ. State the χ^2 -distribution; properties; conditions for validity. Apr/May-2026 Nov-2024

Mar/Apr-2024 Oct-2023 Apr/May-2023

$\chi^2 = \sum \frac{(O-E)^2}{E}$. Mean ν , variance 2ν , positive, right-skewed. Conditions: $N \geq 50$, each $E \geq 5$, independent observations, linear constraints.

PYQ. Explain the χ^2 -test for goodness of fit / independence of attributes. Oct-2025 Apr/May-2023

Goodness of fit: E from the assumed distribution, $\nu = n - 1 - (\text{parameters estimated})$. Independence: $E = \frac{RT \times CT}{N}$, $\nu = (r - 1)(c - 1)$. Unit 4, §4.7.

PYQ. Properties and applications of the F-distribution; test for equality of variances. Oct-2025

Nov/Dec-2022

$F = S_1^2/S_2^2$ (larger on top), $\nu = (n_1 - 1, n_2 - 1)$; positive, right-skewed, depends only on ν_1, ν_2 . Unit 4, §4.6.

PYQ. Write the test statistic for the small-sample t-test. Nov/Dec-2022 · 2M

$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$, $S^2 = \frac{\sum(x - \bar{x})^2}{n-1}$, $\nu = n - 1$.

B. Numerical and derivation questions — fully solved

PYQ. A random sample of 10 boys had IQs 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. (i) Do these data support the assumption of a population mean IQ of 100? (ii) Find a reasonable range in which most of the mean IQ values of samples of 10 boys lie. Nov-2020 · 8M Class notes

x	70	120	110	101	88	83	95	98	107	100	$\Sigma = 972$
$(x - \bar{x})^2$	739.84	519.84	163.84	14.44	84.64	201.64	4.84	0.64	96.04	7.84	1833.6

$\bar{x} = 97.2$, $S^2 = \frac{1833.6}{9} = 203.73$, $S = 14.27$.

Step 1 $H_0 : \mu = 100$.

Step 2 $H_1 : \mu \neq 100$.

Step 3 $\alpha = 0.05$.

Step 4 $t = \frac{97.2 - 100}{14.27/\sqrt{10}} = -0.62$, $|t| = 0.62$, $\nu = 9$.

Step 5 $t_{0.025}(9) = 2.262.$

Step 6 $0.62 < 2.262.$ Accept H_0 . The mean IQ is 100.

(ii) 95% range: $97.2 \pm 2.262 \times \frac{14.27}{\sqrt{10}} = (86.99, 107.41).$

PYQ. In 1950 in India the mean life expectancy was 50 years. A random sample of 11 persons gave life expectancies 58.2, 56.6, 54.2, 50.4, 44.2, 61.9, 57.5, 53.4, 49.7, 55.4, 57.0. Does it confirm the expected view? Apr/May-2023 · 6M Jul-2021 · 8M

$$\sum x = 598.5, \bar{x} = 54.41, \sum (x - \bar{x})^2 = 236.07, S^2 = \frac{236.07}{10} = 23.61, S = 4.859.$$

Step 1 $H_0 : \mu = 50.$

Step 2 $H_1 : \mu \neq 50.$

Step 3 $\alpha = 0.05.$

Step 4 $t = \frac{54.41 - 50}{4.859/\sqrt{11}} = \frac{4.41}{1.465} = 3.01, \nu = 10.$

Step 5 $t_{0.025}(10) = 2.228.$

Step 6 $3.01 > 2.228.$ Reject H_0 . The data do not confirm the view; the mean life expectancy is more than 50 years.

PYQ. Two horses A and B were tested for the time (in seconds) to run a track. A: 28, 30, 32, 33, 33, 29, 34. B: 29, 30, 30, 24, 27, 29. Test whether the two horses have the same running capacity. May-2025 · 10M Oct-2023 · 10M Class notes

x	28	30	32	33	33	29	34	$\Sigma x = 219$
$(x - \bar{x})^2$	10.80	1.65	0.51	2.94	2.94	5.22	7.37	31.43
y	29	30	30	24	27	29		$\Sigma y = 169$
$(y - \bar{y})^2$	0.69	3.36	3.36	17.36	1.36	0.69		26.83

$$\bar{x} = 31.29, \bar{y} = 28.17. S^2 = \frac{31.43+26.83}{7+6-2} = 5.297, S = 2.301.$$

Step 1 $H_0 : \mu_1 = \mu_2.$

Step 2 $H_1 : \mu_1 \neq \mu_2.$

Step 3 $\alpha = 0.05.$

Step 4 $t = \frac{31.29 - 28.17}{2.301\sqrt{\frac{1}{7} + \frac{1}{6}}} = \frac{3.12}{1.280} = 2.44, \nu = 11.$

Step 5 $t_{0.025}(11) = 2.201.$

Step 6 $2.44 > 2.201.$ Reject H_0 . The two horses do **not** have the same running capacity.

CORRECTION TO CLASS NOTES

The class notes reject H_0 but then write "the horses have the same capacity". When H_0 is rejected, the conclusion is the **opposite**: they do not have the same capacity.

PYQ. Two samples: size 10, mean 15, sum of squares of deviations 90; size 12, mean 14, sum of squares 108. Test whether the samples came from the same normal population. Oct/Nov-2021 · 15M

Nov-2020 · 7M

Class notes

Same normal population means equal variances **and** equal means. So do the F-test first, then the t-test.

F-test: $S_1^2 = \frac{90}{9} = 10$, $S_2^2 = \frac{108}{11} = 9.818$. $F = \frac{10}{9.818} = 1.02$; $F_{0.05}(9, 11) = 2.90$. $1.02 < 2.90$: accept $\sigma_1^2 = \sigma_2^2$.

t-test: $S^2 = \frac{90+108}{20} = 9.9$, $S = 3.146$. $t = \frac{15 - 14}{3.146\sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.74$; $t_{0.025}(20) = 2.086$. Accept $\mu_1 = \mu_2$.

Both are accepted. The samples come from the same normal population.

PYQ. To test the effect of a fertilizer on rice, 24 plots were used, half treated and half untreated. Untreated: mean 4.8 quintals, s.d. 0.4. Treated: mean 5.1 quintals, s.d. 0.36. Can we conclude that the fertilizer gives a significant improvement? Apr/May-2023 · 10M

$$n_1 = n_2 = 12. S^2 = \frac{12(0.36)^2 + 12(0.4)^2}{22} = \frac{1.5552 + 1.92}{22} = 0.158, S = 0.3974.$$

Step 1 $H_0 : \mu_1 = \mu_2$.

Step 2 $H_1 : \mu_1 > \mu_2$ (treated is better; right-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $t = \frac{5.1 - 4.8}{0.3974\sqrt{\frac{1}{12} + \frac{1}{12}}} = \frac{0.3}{0.1622} = 1.85, \nu = 22.$

Step 5 $t_{0.05}(22) = 1.717$.

Step 6 $1.85 > 1.717$. Reject H_0 . The fertilizer gives a significant improvement.

PYQ. Gain in weight (lbs) of pigs fed on two diets. A: 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25. B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22. Do the two diets differ significantly?

Nov-2020 · 7M

Class notes

$$\bar{x} = 28, \bar{y} = 30, \sum(x - \bar{x})^2 = 380, \sum(y - \bar{y})^2 = 1410. S^2 = \frac{1790}{25} = 71.6.$$

$$t = \frac{28 - 30}{\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}} = -0.61, \nu = 25; t_{0.025}(25) = 2.060. \text{ Accept } H_0. \text{ The diets do not differ significantly.}$$

PYQ. Average weekly losses of worker-hours due to accidents in 10 plants before and after a safety programme: Before 45, 73, 46, 124, 33, 57, 83, 34, 26, 17. After 36, 60, 44, 119, 35, 51, 77, 29, 24, 11. Is the programme effective at the 0.05 level? Nov-2024 · 10M

$d = x - y$	9	13	2	5	-2	6	6	5	2	6	$\Sigma d = 52$
$(d - \bar{d})^2$	14.44	60.84	10.24	0.04	51.84	0.64	0.64	0.04	10.24	0.64	149.6

$$\bar{d} = 5.2, S_d^2 = \frac{149.6}{9} = 16.62, S_d = 4.077.$$

Step 1 $H_0 : \mu_d = 0$ (no effect).

Step 2 $H_1 : \mu_d > 0$ (losses decrease; right-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $t = \frac{5.2}{4.077/\sqrt{10}} = 4.03, \nu = 9$.

Step 5 $t_{0.05}(9) = 1.833$.

Step 6 $4.03 > 1.833$. Reject H_0 . The safety programme is effective.

PYQ. An investigator states that husbands are more intelligent than wives. IQs of 10 couples: Husbands 117, 105, 97, 105, 123, 109, 86, 78, 103, 107. Wives 106, 98, 87, 104, 116, 95, 90, 69, 108, 85. Test with a reasonable test at 0.05. Nov/Dec-2022 · 10M Class notes

Each husband is paired with his wife. So the **reasonable test is the paired t-test**.

d : 11, 7, 10, 1, 7, 14, -4, 9, -5, 22. $\bar{d} = 7.2, \sum(d - \bar{d})^2 = 603.6, S_d = 8.19$.

Step 1 $H_0 : \mu_d = 0$.

Step 2 $H_1 : \mu_d > 0$ (right-tailed).

Step 3 $\alpha = 0.05$.

Step 4 $t = \frac{7.2}{8.19/\sqrt{10}} = 2.78, \nu = 9$.

Step 5 $t_{0.05}(9) = 1.833$.

Step 6 $2.78 > 1.833$. Reject H_0 . Husbands are more intelligent than wives.

Class notes treat the samples as independent: $\bar{x} = 103, \bar{y} = 95.8, S = 13.51, t = 1.19 < 1.734$, accept H_0 . Write the paired test in the exam, and mention the reason (data are in pairs).

PYQ. Find (i) $F_{0.05}$ for $\nu_1 = 15, \nu_2 = 7$ (ii) $F_{0.95}$ for $\nu_1 = 12, \nu_2 = 15$. May-2025 · 3M Oct-2023 · 3M

(i) From the table, $F_{0.05}(15, 7) = 3.51$. (ii) $F_{0.95}(12, 15) = \frac{1}{F_{0.05}(15, 12)} = \frac{1}{2.62} = 0.382$.

PYQ. If S_1^2 and S_2^2 are the variances of independent random samples of sizes n_1 and n_2 from normal populations with equal variances, find $P\left(\frac{S_1^2}{S_2^2} < 4.03\right)$. Nov-2024 · 3M Mar/Apr-2024 · 3M

The source problem (Miller & Freund) has $n_1 = 10, n_2 = 15$. So $F = \frac{S_1^2}{S_2^2}$ has $(\nu_1, \nu_2) = (9, 14)$. From the F-table, $F_{0.01}(9, 14) = 4.03$. So $P(F > 4.03) = 0.01$ and $P\left(\frac{S_1^2}{S_2^2} < 4.03\right) = 1 - 0.01 = 0.99$.

0.99

In the exam, write the assumed sample sizes $n_1 = 10, n_2 = 15$ in the first line.

PYQ. Pumpkins: two samples of 11 and 9 have sample s.d.s 0.8 and 0.5. Test that the true variances are equal against the alternative that they are not, at the 10% level. Nov-2024 · 10M Class notes

$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11(0.64)}{10} = 0.704, S_2^2 = \frac{9(0.25)}{8} = 0.281$.

Step 1 $H_0 : \sigma_1^2 = \sigma_2^2$.

Step 2 $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Step 3 $\alpha = 0.10$ (two-tailed $\rightarrow$ use $F_{0.05}$ in the upper tail).

Step 4 $F = \frac{0.704}{0.281} = 2.51, \nu = (10, 8)$.

Step 5 $F_{0.05}(10, 8) = 3.35$.

Step 6 $2.51 < 3.35$. Accept H_0 . The variances are equal.

PYQ. In one sample of 8 observations, the sum of squares of deviations from the mean is 84.4. In another sample of 10 it is 102.6. Test at 5% whether the populations have the same variance.

Jul-2021 · 15M

[Class notes](#)

$S_1^2 = \frac{84.4}{7} = 12.06, S_2^2 = \frac{102.6}{9} = 11.40. F = \frac{12.06}{11.40} = 1.06, \nu = (7, 9). F_{0.05}(7, 9) = 3.29. 1.06 < 3.29$. Accept H_0 : same variance.

PYQ. Household net expenditure on health care (% of income). South: 15, 8.0, 3.8, 6.4, 27.4, 19.0, 35.3, 13.6. North: 18.8, 23.1, 10.3, 8.0, 18.0, 10.2, 15.2, 19.0, 20.2. Test the equality of variances.

Mar/Apr-2024 · 10M

[Class notes](#)

South: $n_1 = 8, \bar{x} = 16.06, \sum(x - \bar{x})^2 = 823.18, S_1^2 = \frac{823.18}{7} = 117.60$.

North: $n_2 = 9, \bar{y} = 15.87, \sum(y - \bar{y})^2 = 219.50, S_2^2 = \frac{219.50}{8} = 27.44$.

Step 1 $H_0 : \sigma_1^2 = \sigma_2^2$.

Step 2 $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Step 3 $\alpha = 0.05$.

Step 4 $F = \frac{117.60}{27.44} = 4.29, \nu = (7, 8)$.

Step 5 $F_{0.05}(7, 8) = 3.50$.

Step 6 $4.29 > 3.50$. Reject H_0 . The variances are not equal.

PYQ. Time taken by workers using two methods. A: 20, 16, 26, 27, 23, 22. B: 27, 33, 42, 35, 32, 34, 38. Do the variances differ significantly? Mar/Apr-2024 · 10M

A: $\bar{x} = 22.33, \sum(x - \bar{x})^2 = 81.33, S_1^2 = \frac{81.33}{5} = 16.27$. B: $\bar{y} = 34.43, \sum(y - \bar{y})^2 = 133.71, S_2^2 = \frac{133.71}{6} = 22.29$.

Larger on top: $F = \frac{22.29}{16.27} = 1.37, \nu = (6, 5). F_{0.05}(6, 5) = 4.95$. Accept H_0 . The variances do not differ significantly.

PYQ. Sample A: 10, 12, 14, 16, 18. Sample B: 8, 10, 12, 14, 16. Test for equality of variances.

Oct-2025 · 5M

Both samples have the same spread: $\sum(x - \bar{x})^2 = 40$ for each, so $S_1^2 = S_2^2 = 10. F = 1.00 < F_{0.05}(4, 4) = 6.39$. Accept H_0 .

PYQ. A pair of dice is thrown 360 times. Frequencies of sums 2–12: 8, 24, 35, 37, 44, 65, 51, 42, 26, 14, 14. Are the dice fair at the 0.05 level? May-2025 · 10M Oct-2023 · 10M Nov-2020 · 8M [Class notes](#)

For fair dice, $P(\text{sum} = k) = \frac{1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1}{36}$. $E = 360 \cdot P$.

Sum	2	3	4	5	6	7	8	9	10	11	12
O	8	24	35	37	44	65	51	42	26	14	14
E	10	20	30	40	50	60	50	40	30	20	10
$\frac{(O-E)^2}{E}$	0.4	0.8	0.833	0.225	0.72	0.417	0.02	0.1	0.533	1.8	1.6

Step 1 H_0 : the dice are fair.

Step 2 H_1 : the dice are not fair.

Step 3 $\alpha = 0.05$.

Step 4 $\chi^2 = 7.45$, $\nu = 11 - 1 = 10$.

Step 5 $\chi_{0.05}^2(10) = 18.307$.

Step 6 $7.45 < 18.307$. Accept H_0 . The dice are fair.

PYQ. A die is thrown 120 times. (a) Faces 1–6: 20, 21, 19, 18, 22, 20. (b) Faces 1–6: 10, 15, 25, 25, 18, 27. Test whether the die is fair at 5%. Oct-2025 · 5M Jul-2021 · 15M

$E = 20$ for each face, $\nu = 5$, $\chi_{0.05}^2(5) = 11.07$.

(a) $\chi^2 = \frac{0+1+1+4+4+0}{20} = 0.5 < 11.07$. Accept: fair.

(b) $\chi^2 = \frac{100+25+25+25+4+49}{20} = \frac{228}{20} = 11.4 > 11.07$. Reject: the die is not fair.

PYQ. The numbers of automobile accidents per week in a community: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these figures consistent with the belief that accident conditions were the same during this 10-week period? Mar/Apr-2022 · 9M

Total = 100, so $E = 10$ for each week. $\chi^2 = \frac{4+4+100+64+16+0+25+16+1+36}{10} = 26.6$, $\nu = 9$, $\chi_{0.05}^2(9) = 16.92$. $26.6 > 16.92$. Reject H_0 . The conditions were not the same.

PYQ. Fit a Poisson distribution to the data and test the goodness of fit at 0.05: x : 0, 1, 2, 3, 4; f : 419, 352, 154, 56, 19. Nov/Dec-2022 · 10M

Step 1 $N = 1000$, $\lambda = \bar{x} = \frac{0+352+308+168+76}{1000} = 0.904$.

Step 2 $P(0) = e^{-0.904} = 0.4049$. Use $P(r+1) = \frac{\lambda}{r+1}P(r)$: expected $E = 404.9, 366.1, 165.5, 49.9$, and the last class (4 or more) = $1000 - 986.3 = 13.7$.

Step 3 $\chi^2 = \frac{(419-404.9)^2}{404.9} + \frac{(352-366.1)^2}{366.1} + \frac{(154-165.5)^2}{165.5} + \frac{(56-49.9)^2}{49.9} + \frac{(19-13.7)^2}{13.7} = 4.67$.

Step 4 $\nu = n - 2 = 5 - 2 = 3$ (one parameter λ estimated). $\chi_{0.05}^2(3) = 7.815$.

Step 5 $4.67 < 7.815$. Accept H_0 . The Poisson distribution is a good fit.

PYQ. Fish from lake A (10): 11.5, 10.8, 11.6, 9.4, 12.4, 11.4, 12.2, 11, 10.6, 10.8 ppm. Lake B (8): 11.8, 12.6, 12.2, 12.5, 11.7, 12.1, 10.4, 12.6. The method variances are known: 0.09 (A) and 0.16 (B). Can we reject (at 5%) the claim that the lakes are equally contaminated? **Oct/Nov-2021 · 8M**

The population variances are **known**, so use Z (not t). $\bar{x}_A = 11.17$, $\bar{x}_B = 11.99$.

$$Z = \frac{11.17 - 11.99}{\sqrt{\frac{0.09}{10} + \frac{0.16}{8}}} = \frac{-0.8175}{0.1703} = -4.80. |Z| > 1.96. \text{ Reject } H_0: \text{ the lakes are not equally contaminated.}$$

Unit 5 Curve Fitting

A. Theory questions — short model answers

PYQ. State the principle of least squares (and its advantage). all 8 papers

The best-fit curve makes $S = \sum [y_i - f(x_i)]^2$ minimum. Advantage: one unique best-fit curve; errors do not cancel; simple linear normal equations. Unit 5, §5.1.

PYQ. Normal equations for $y = ax^2 + bx + c$. Apr/May-2023 · 2M

$$\sum y = a \sum x^2 + b \sum x + nc; \sum xy = a \sum x^3 + b \sum x^2 + c \sum x; \sum x^2 y = a \sum x^4 + b \sum x^3 + c \sum x^2.$$

B. Numerical and derivation questions — fully solved

PYQ. Fit a straight line to the data: x: 1, 2, 3, 4, 5; y: 2, 4, 5, 4, 5. Oct-2025 · 5M Class notes

x	y	x^2	xy
1	2	1	2
2	4	4	8
3	5	9	15
4	4	16	16
5	5	25	25
15	20	55	66

Normal equations ($n = 5$): $5a + 15b = 20$... (i); $15a + 55b = 66$... (ii).

(i) $\times 3$: $15a + 45b = 60$. Subtract from (ii): $10b = 6 \Rightarrow b = 0.6$. Then $a = \frac{20-9}{5} = 2.2$. **$y = 2.2 + 0.6x$**

PYQ. By the method of least squares, find the straight line that best fits: x: 1, 2, 3, 4, 5; y: 14, 27, 40, 55, 68. Nov-2020 · 8M

$$\sum x = 15, \sum y = 204, \sum x^2 = 55, \sum xy = 14 + 54 + 120 + 220 + 340 = 748.$$

$$5a + 15b = 204; 15a + 55b = 748. \text{ Solve: } 10b = 748 - 612 = 136 \Rightarrow b = 13.6; a = \frac{204-204}{5} = 0.$$

$$\mathbf{y = 13.6x}$$

PYQ. Fit a straight line: x: 1, 2, 3, 4, 5, 6; y: 3.7, 6.2, 8.7, 11.2, 13.7, 16.2. Jul-2021 · 7M

$$\sum y = 59.7, \sum xy = 252.7. 6a + 21b = 59.7; 21a + 91b = 252.7. \text{ Solve: } b = 2.5, a = 1.2. \mathbf{y = 1.2 + 2.5x}$$

(The points lie exactly on this line.)

PYQ. The table shows how many weeks six persons worked at an automobile inspection station and the number of cars each inspected between noon and 2 PM. Fit $y = a + bx + cx^2$. x : 2, 7, 9, 1, 5, 12; y : 13, 21, 23, 14, 15, 21. Apr/May-2026 · 5M Class notes

x	y	x^2	x^3	x^4	xy	x^2y
2	13	4	8	16	26	52
7	21	49	343	2401	147	1029
9	23	81	729	6561	207	1863
1	14	1	1	1	14	14
5	15	25	125	625	75	375
12	21	144	1728	20736	252	3024
36	107	304	2934	30340	721	6357

Normal equations ($n = 6$):

$$6a + 36b + 304c = 107 \dots (i)$$

$$36a + 304b + 2934c = 721 \dots (ii)$$

$$304a + 2934b + 30340c = 6357 \dots (iii)$$

Solve (eliminate a , then b): $a = 10.822$, $b = 1.717$, $c = -0.0650$. $y = 10.822 + 1.717x - 0.065x^2$

CORRECTION TO CLASS NOTES

The class notes write equation (i) as $6a + 36b + 304c = 721$. The right side must be $\sum y = 107$. That error gives the wrong curve $y = -15.584 + 21.704x + 0.1089x^2$. Check: at $x = 12$, it gives $y \approx 260$, but the data value is 21.

PYQ. Find a least squares quadratic curve: x : 1, 2, 3, 4, 5, 6; y : 6, 4, 3, 5, 4, 2. Mar/Apr-2024 · 8M

Oct-2023 · 8M Class notes

Sums: $\sum x = 21$, $\sum x^2 = 91$, $\sum x^3 = 441$, $\sum x^4 = 2275$, $\sum y = 24$, $\sum xy = 75$, $\sum x^2y = 301$.

$6a + 21b + 91c = 24 \dots (i)$; $21a + 91b + 441c = 75 \dots (ii)$; $91a + 441b + 2275c = 301 \dots (iii)$.

Solve: $a = 5.8$, $b = -0.514$, $c = 0$. $y = 5.8 - 0.514x (+ 0 \cdot x^2)$

For this data the best parabola has $c = 0$, so it is the same as the straight line of Example 1. This is a correct and complete answer; state it in one line.

PYQ. Fit a parabola $y = a + bx + cx^2$: x : 1, 2, 3, 4, 5, 6, 7; y : 2.3, 5.2, 9.7, 16.5, 29.4, 35.5, 54.4.

Nov/Dec-2022 · 10M

$n = 7$, $\sum x = 28$, $\sum x^2 = 140$, $\sum x^3 = 784$, $\sum x^4 = 4676$, $\sum y = 153$, $\sum xy = 848.6$, $\sum x^2y = 5053$.

$7a + 28b + 140c = 153$; $28a + 140b + 784c = 848.6$; $140a + 784b + 4676c = 5053$.

Solve: $a = 2.371$, $b = -1.093$, $c = 1.193$. $y = 2.371 - 1.093x + 1.193x^2$

SHORT-CUT

When x values are equally spaced, put $X = x - 4$ (middle value). Then $\sum X = \sum X^3 = 0$ and the equations become very easy.

PYQ. Fit a least squares parabola: X: 0, 0.2, 0.4, 0.7, 0.9, 1; Y: 1.016, 0.768, 0.648, 0.401, 0.272, 0.193.

Oct/Nov-2021 · 15M

$$\sum x = 3.2, \sum x^2 = 2.5, \sum x^3 = 2.144, \sum x^4 = 1.9234, \sum y = 3.298, \sum xy = 1.1313, \sum x^2y = 0.7442.$$

$$6a + 3.2b + 2.5c = 3.298; 3.2a + 2.5b + 2.144c = 1.1313; 2.5a + 2.144b + 1.9234c = 0.7442.$$

$$y = 0.999 - 1.009x + 0.213x^2$$

PYQ. For the "cars inspected" data (x: 2, 7, 9, 1, 5, 12; y: 13, 21, 23, 14, 15, 21), fit $y = ae^{bx}$.

May-2025 · 8M

Class notes

x	y	$Y = \log_{10} y$	x^2	xY
2	13	1.1139	4	2.2278
7	21	1.3222	49	9.2554
9	23	1.3617	81	12.2553
1	14	1.1461	1	1.1461
5	15	1.1761	25	5.8805
12	21	1.3222	144	15.8664
36		7.4422	304	46.6315

$$6A + 36B = 7.4422 \dots (i); 36A + 304B = 46.6315 \dots (ii). (i) \times 6: 36A + 216B = 44.6532. \text{ Subtract: } 88B = 1.9783 \Rightarrow B = 0.02248. A = \frac{7.4422 - 0.8093}{6} = 1.1055.$$

$$a = 10^{1.1055} = 12.75, b = \frac{0.02248}{0.4343} = 0.0518. \quad y = 12.75 e^{0.0518x}$$

PYQ. Determine a and b by least squares such that $y = ae^{bx}$: x: 2, 4, 6, 8, 10; y: 4.077, 11.084, 30.128, 81.897, 222.62. Jul-2021 · 8M

$$Y = \log_{10} y: 0.6103, 1.0447, 1.4790, 1.9133, 2.3476. \sum Y = 7.3948, \sum xY = 53.055, \sum x = 30, \sum x^2 = 220.$$

$$5A + 30B = 7.3948; 30A + 220B = 53.055. \text{ Solve: } B = 0.21715, A = 0.17606.$$

$$a = 10^{0.17606} = 1.50, b = \frac{0.21715}{0.4343} = 0.50. \quad y = 1.5 e^{0.5x}$$

PYQ. Fit $Y = Ae^{BX}$: X: 2, 4, 6, 8, 10, 12; Y: 1.8, 1.5, 1.4, 1.1, 1.1, 0.9. Oct/Nov-2021 · 7M

$$\log_{10} Y: 0.2553, 0.1761, 0.1461, 0.0414, 0.0414, -0.0458. \text{ Sum} = 0.6145; \sum X \log Y = 2.2877; \sum X = 42; \sum X^2 = 364.$$

$$6a_0 + 42b_0 = 0.6145; 42a_0 + 364b_0 = 2.2877 \rightarrow b_0 = -0.02877, a_0 = 0.3038. A = 10^{0.3038} = 2.013, B =$$

$$\frac{-0.02877}{0.4343} = -0.0662. \quad Y = 2.013 e^{-0.0662X}$$

PYQ. For the cars data, fit $y = ax^b$. Nov-2024 · 10M Class notes

x	y	$X = \log x$	$Y = \log y$	X^2	XY
2	13	0.3010	1.1139	0.0906	0.3353
7	21	0.8451	1.3222	0.7142	1.1174
9	23	0.9542	1.3617	0.9106	1.2994
1	14	0	1.1461	0	0
5	15	0.6990	1.1761	0.4886	0.8221
12	21	1.0792	1.3222	1.1647	1.4269
		3.8785	7.4422	3.3686	5.0011

$6A + 3.8785b = 7.4422$; $3.8785A + 3.3686b = 5.0011$. Solve: $b = 0.2209$, $A = 1.0976$. $a = 10^{1.0976} = 12.52$.

$$y = 12.52 x^{0.2209}$$

PYQ. Fit $y = ax^b$: X: 50, 100, 250, 500, 1000; Y: 108, 53, 24, 9, 5. Mar/Apr-2022 · 7M

$\log x$: 1.6990, 2.0000, 2.3979, 2.6990, 3.0000 (sum 11.7959; $\sum X^2 = 28.9211$).

$\log y$: 2.0334, 1.7243, 1.3802, 0.9542, 0.6990 (sum 6.7911); $\sum XY = 14.8853$.

$5A + 11.7959b = 6.7911$; $11.7959A + 28.9211b = 14.8853$. Solve: $b = -1.040$, $A = 3.8116$, $a = 10^{3.8116} = 6481$. $y = 6481 x^{-1.04}$

PYQ. Estimate y at $x = 5$ by fitting a least squares curve $y = \frac{b}{x(x-a)}$: x: 3.6, 4.8, 6.0, 7.2, 8.4, 9.6, 10.8; y: 0.83, 0.31, 0.17, 0.10, 0.07, 0.05, 0.04. Apr/May-2023 · 10M

Step 1 Make it linear: $x(x-a) = \frac{b}{y} \Rightarrow \frac{1}{xy} = \frac{x-a}{b} = -\frac{a}{b} + \frac{1}{b}x$. Put $Y = \frac{1}{xy}$, $A = -\frac{a}{b}$, $B = \frac{1}{b}$: $Y = A + Bx$.

Step 2 Y: 0.3347, 0.6720, 0.9804, 1.3889, 1.7007, 2.0833, 2.3148. $\sum Y = 9.4748$, $\sum xY = 79.599$, $\sum x = 50.4$, $\sum x^2 = 403.2$.

Step 3 $7A + 50.4B = 9.4748$; $50.4A + 403.2B = 79.599$. Solve: $B = 0.2822$, $A = -0.6786$.

Step 4 $b = \frac{1}{B} = 3.543$, $a = -Ab = 2.404$. So $y = \frac{3.543}{x(x-2.404)}$.

Step 5 At $x = 5$: $y = \frac{3.543}{5(2.596)} = 0.273$.

$$y(5) \approx 0.27$$

Unit 6 Correlation & Regression

A. Theory questions — short model answers

PYQ. Explain types of correlation. Oct-2023 · 3M

Positive / negative; simple / multiple; partial / total; linear / non-linear. Unit 6, §6.1.

PYQ. Discuss the correlation coefficient. Apr/May-2023 · 3M

$r = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$; $-1 \leq r \leq 1$; unit-free; independent of change of origin and scale.

PYQ. Define the regression coefficients and give their properties. Nov/Dec-2022 · 3M

$b_{yx} = r \frac{\sigma_y}{\sigma_x}$, $b_{xy} = r \frac{\sigma_x}{\sigma_y}$; $r^2 = b_{xy} b_{yx}$; same sign; both cannot exceed 1; AM $\geq r$; independent of origin, not scale. Unit 6, §6.4.

B. Numerical and derivation questions — fully solved

PYQ. Calculate Karl Pearson's correlation coefficient: X: 10, 20, 30, 40, 50; Y: 12, 21, 32, 43, 54.

Oct-2025 · 5M

[Class notes](#)

x	$X = x - 30$	X^2	y	$Y = y - 32.4$	Y^2	XY
10	-20	400	12	-20.4	416.16	408
20	-10	100	21	-11.4	129.96	114
30	0	0	32	-0.4	0.16	0
40	10	100	43	10.6	112.36	106
50	20	400	54	21.6	466.56	432
150		1000	162		1125.2	1060

$\bar{x} = 30$, $\bar{y} = 32.4$. $r = \frac{1060}{\sqrt{1000 \times 1125.2}} = 0.9993$. Very strong positive correlation. **$r = 0.9993$**

PYQ. Find Karl Pearson's coefficient: Wages 100, 101, 102, 102, 100, 99, 97, 98, 96, 95; Cost of living 98, 99, 99, 97, 95, 92, 95, 94, 90, 91. Nov-2020 · 7M [Class notes](#)

$\bar{x} = 99$, $\bar{y} = 95$. $\sum X^2 = 54$, $\sum Y^2 = 96$, $\sum XY = 61$. $r = \frac{61}{\sqrt{54 \times 96}} = \frac{61}{72} = 0.847$.

PYQ. Find a suitable coefficient of correlation: Fertilizer (tons) 15, 18, 20, 24, 30, 35, 40, 50; Productivity (tons) 85, 93, 95, 105, 120, 130, 150, 160. Nov-2020 · 8M

Both are measured numbers, so Karl Pearson's r is suitable. $\bar{x} = 29$, $\bar{y} = 117.25$. $\sum X^2 = 1022$, $\sum Y^2 = 5343.5$, $\sum XY = 2317$.

$r = \frac{2317}{\sqrt{1022 \times 5343.5}} = 0.9915$. (Ranks of x and y agree completely, so Spearman's $\rho = 1$.)

PYQ. Distinguish between Karl Pearson's and Spearman's rank correlation coefficients.

Oct-2025 · 3M

Karl Pearson's r	Spearman's ρ
Uses the actual values of x and y.	Uses only the ranks.
For quantitative (measured) data.	For qualitative data (beauty, honesty) or ranked data.
$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$	$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$
Affected by extreme values.	Less affected by extreme values.
Measures linear relationship only.	Measures any monotonic relationship.

PYQ. Marks of 10 students in C++ (x): 6, 5, 8, 8, 7, 6, 10, 4, 9, 7 and Statistics (y): 8, 7, 7, 10, 5, 8, 10, 6, 8, 6. Compute Spearman's rank correlation coefficient. Apr/May-2026 · 10M [Class notes](#)

x	Rank x	y	Rank y	D	D ²
6	7.5	8	4	3.5	12.25
5	9	7	6.5	2.5	6.25
8	3.5	7	6.5	-3	9
8	3.5	10	1.5	2	4
7	5.5	5	10	-4.5	20.25
6	7.5	8	4	3.5	12.25
10	1	10	1.5	-0.5	0.25
4	10	6	8.5	1.5	2.25
9	2	8	4	-2	4
7	5.5	6	8.5	-3	9
$\sum D^2$					79.5

Step 1 Ties in x: 8 (×2), 7 (×2), 6 (×2) → three groups with $m = 2$. Ties in y: 10 (×2), 8 (×3), 7 (×2), 6 (×2) → $m = 2, 3, 2, 2$.

Step 2 Correction = $\frac{1}{12} [6(2^3 - 2) + (3^3 - 3)] = \frac{1}{12} [36 + 24] = 5$.

Step 3 $\rho = 1 - \frac{6(79.5 + 5)}{10(100 - 1)} = 1 - \frac{507}{990} = 0.488$.

ρ = 0.488

PYQ. Heights of 12 fathers (X): 65, 63, 67, 64, 68, 62, 70, 66, 68, 67, 69, 71 and their eldest sons (Y): 68, 66, 68, 65, 69, 66, 68, 65, 71, 67, 68, 70. Compute Spearman's rank correlation coefficient.

May-2025 · 10M

[Class notes](#)

Ranks X: 9, 11, 6.5, 10, 4.5, 12, 2, 8, 4.5, 6.5, 3, 1. Ranks Y: 5.5, 9.5, 5.5, 11.5, 3, 9.5, 5.5, 11.5, 1, 8, 5.5, 2. $\sum D^2 = 72.5$. Ties: X: 67 (×2), 68 (×2). Y: 68 (×4), 66 (×2), 65 (×2).

$$\text{Correction} = \frac{1}{12}[(8-2) + (8-2) + (64-4) + (8-2) + (8-2)] = \frac{84}{12} = 7.$$

$$\rho = 1 - \frac{6(72.5 + 7)}{12(143)} = 1 - \frac{477}{1716} = 0.722. \quad \rho = 0.722$$

PYQ. Marks of 8 students in Economics (x): 78, 56, 36, 66, 25, 75, 82, 62 and Statistics (y): 84, 44, 57, 58, 60, 68, 62, 58. Compute Spearman's rank correlation. Nov-2024 · 10M

Ranks x: 2, 6, 7, 4, 8, 3, 1, 5. Ranks y: 1, 8, 7, 5.5, 4, 2, 3, 5.5 (58 occurs twice → ranks 5, 6 → 5.5).

$D: 1, -2, 0, -1.5, 4, 1, -2, -0.5$. $\sum D^2 = 1 + 4 + 0 + 2.25 + 16 + 1 + 4 + 0.25 = 28.5$. Correction = $\frac{1}{12}(8-2) = 0.5$.

$$\rho = 1 - \frac{6(28.5 + 0.5)}{8(63)} = 1 - \frac{174}{504} = 0.655. \quad \rho = 0.655$$

PYQ. If $x = 2y + 3$ and $y = \frac{x}{8} + 6$ are the regression lines of x on y and y on x respectively, find the correlation coefficient. Apr/May-2026 · 3M Nov-2024 · 3M Mar/Apr-2024 · 3M Class notes

$$b_{xy} = 2, b_{yx} = \frac{1}{8}. r^2 = 2 \times \frac{1}{8} = \frac{1}{4}, r = +\frac{1}{2} \text{ (both coefficients are positive).}$$

$$\text{Means: } x = 2y + 3 \text{ and } x = 8y - 48 \Rightarrow 2y + 3 = 8y - 48 \Rightarrow y = 8.5, x = 20. \quad r = 0.5, \bar{x} = 20, \bar{y} = 8.5$$

PYQ. For army personnel, the regression of weight of kidneys (y) on weight of heart (x) is $y = 0.399x + 6.394$, and the regression of x on y is $x = 1.212y - 2.461$. Find r. May-2025 · 3M

$$b_{yx} = 0.399, b_{xy} = 1.212. r = \sqrt{0.399 \times 1.212} = \sqrt{0.4836} = 0.695.$$

PYQ. Marks of 10 students. Internal: 25, 28, 30, 32, 35, 36, 38, 39, 42, 45. University: 20, 26, 29, 30, 25, 18, 26, 35, 35, 46. Obtain the two regression equations and find (a) the most likely internal marks for university marks of 25 (b) the most likely university marks for internal marks of 30.

Mar/Apr-2024 · 10M

Class notes

x	X	X^2	y	Y	Y^2	XY
25	-10	100	20	-9	81	90
28	-7	49	26	-3	9	21
30	-5	25	29	0	0	0
32	-3	9	30	1	1	-3
35	0	0	25	-4	16	0
36	1	1	18	-11	121	-11
38	3	9	26	-3	9	-9
39	4	16	35	6	36	24
42	7	49	35	6	36	42
45	10	100	46	17	289	170
350		358	290		598	324

$$\bar{x} = 35, \bar{y} = 29. b_{yx} = \frac{324}{358} = 0.905, b_{xy} = \frac{324}{598} = 0.542.$$

Step 1 Y on X: $y - 29 = 0.905(x - 35)$, i.e. $y = 0.905x - 2.68$.

Step 2 X on Y: $x - 35 = 0.542(y - 29)$, i.e. $x = 0.542y + 19.29$.

Step 3 (a) $y = 25$: $x = 35 + 0.542(25 - 29) = 32.83 \approx 33$.

Step 4 (b) $x = 30$: $y = 29 + 0.905(30 - 35) = 24.47 \approx 24$.

$$r = \sqrt{0.905 \times 0.542} = 0.700.$$

The class notes use 32 for the 5th university mark (the PYQ has 25). With 32: $\bar{y} = 29.7$, $b_{xy} = 0.553$, answers 33 and 25.

PYQ. Find the correlation coefficient and the lines of regression, and predict Y if X = 67.5. X: 65, 66, 67, 67, 68, 69, 71, 73; Y: 67, 68, 64, 68, 72, 70, 69, 70. Jul-2021 · 15M Class notes

$$\bar{x} = 68.25, \bar{y} = 68.5, \sum X^2 = 49.5, \sum Y^2 = 40, \sum XY = 21.$$

$$r = \frac{21}{\sqrt{49.5 \times 40}} = 0.472. \quad b_{yx} = \frac{21}{49.5} = 0.4242, \quad b_{xy} = \frac{21}{40} = 0.525.$$

$$\text{Y on X: } y - 68.5 = 0.4242(x - 68.25). \quad \text{X on Y: } x - 68.25 = 0.525(y - 68.5).$$

$$\text{At } x = 67.5: y = 68.5 + 0.4242(-0.75) = 68.18. \quad \mathbf{Y \approx 68.18}$$

PYQ. Find the coefficient of correlation and the regression lines: x: 6, 2, 10, 4, 8; y: 9, 11, 5, 8, 7.

Nov/Dec-2022 · 10M

$$\bar{x} = 6, \bar{y} = 8. \quad X: 0, -4, 4, -2, 2; \quad Y: 1, 3, -3, 0, -1. \quad \sum X^2 = 40, \sum Y^2 = 20, \sum XY = 0 - 12 - 12 + 0 - 2 = -26.$$

$$r = \frac{-26}{\sqrt{800}} = -0.919. \quad b_{yx} = -0.65, \quad b_{xy} = -1.3.$$

$$\text{Y on X: } y - 8 = -0.65(x - 6) \Rightarrow y = -0.65x + 11.9. \quad \text{X on Y: } x - 6 = -1.3(y - 8) \Rightarrow x = -1.3y + 16.4.$$

PYQ. Find the regression line of Y on X and the correlation coefficient: x: 1, 5, 3, 2, 1, 1, 7, 3; y: 6, 1, 0, 0, 1, 2, 1, 5. Oct/Nov-2021 · 15M Class notes

$$\bar{x} = 2.875, \bar{y} = 2. \quad \sum X^2 = 32.875, \sum Y^2 = 36, \sum XY = -10.$$

$$b_{yx} = \frac{-10}{32.875} = -0.304. \quad \text{Y on X: } y - 2 = -0.304(x - 2.875) \Rightarrow y = -0.304x + 2.875.$$

$$r = \frac{-10}{\sqrt{32.875 \times 36}} = -0.291.$$

Correction: the class notes get $\sum XY = -5.75$ because of two sign slips (row 2: $2.125 \times (-1) = -2.125$; row 5: $(-1.875)(-1) = +1.875$). The correct total is -10.

PYQ. For 25 pairs: $\sum x = 127$, $\sum y = 100$, $\sum x^2 = 760$, $\sum y^2 = 449$, $\sum xy = 500$. Later, two pairs were found copied as (8, 14) and (8, 6) instead of (8, 12) and (6, 8). Find the correct r and the correct regression lines. Oct-2023 · 10M Apr/May-2023 · 10M Class notes

Step 1 $\sum x = 127 - 8 - 8 + 8 + 6 = 125$; $\sum y = 100 - 14 - 6 + 12 + 8 = 100$.

Step 2 $\sum x^2 = 760 - 64 - 64 + 64 + 36 = 732$; $\sum y^2 = 449 - 196 - 36 + 144 + 64 = 425$.

Step 3 $\sum xy = 500 - 112 - 48 + 96 + 48 = 484$.

Step 4 $\bar{x} = 5, \bar{y} = 4. \quad S_{xx} = 732 - \frac{125^2}{25} = 107; \quad S_{yy} = 425 - 400 = 25; \quad S_{xy} = 484 - \frac{125 \times 100}{25} = -16.$

Step 5 $r = \frac{-16}{\sqrt{107 \times 25}} = -0.309.$

Step 6 $b_{yx} = \frac{-16}{107} = -0.1495, \quad b_{xy} = \frac{-16}{25} = -0.64.$
Y on X: $y - 4 = -0.1495(x - 5) \Rightarrow y = -0.1495x + 4.748.$

$$X \text{ on } Y: x - 5 = -0.64(y - 4) \Rightarrow x = -0.64y + 7.56.$$

$$r = -0.309$$

CORRECTION TO CLASS NOTES

The class notes put the raw sums into $r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = \frac{484}{\sqrt{732 \times 425}} = 0.868$. That formula needs deviations from the mean. Subtract $\frac{\sum x \sum y}{n}$ etc. first, as in step 4.

PYQ. For 10 pairs: $\sum x = 66$, $\sum y = 69$, $\sum x^2 = 476$, $\sum y^2 = 521$, $\sum xy = 485$. Two pairs were copied as (2, 3), (7, 5) instead of (4, 6), (9, 8). Find the correct r . [Mar/Apr-2022 · 15M](#) [Class notes](#)

Correct: $\sum x = 70$, $\sum y = 75$, $\sum x^2 = 520$, $\sum y^2 = 587$, $\sum xy = 540$.

$S_{xx} = 520 - 490 = 30$, $S_{yy} = 587 - 562.5 = 24.5$, $S_{xy} = 540 - 525 = 15$.

$r = \frac{15}{\sqrt{30 \times 24.5}} = 0.553$. $b_{yx} = 0.5$, $b_{xy} = 0.612$. Y on X: $y - 7.5 = 0.5(x - 7)$; X on Y: $x - 7 = 0.612(y - 7.5)$.