Sure! Here are some important exam questions for **Probability and Statistics** for **BCA 3rd Semester**:

Important Probability Questions

1. **Define Probability. Explain the different approaches to probability.**
2. **State and prove the addition theorem of probability.**
3. **What is conditional probability? Explain with an example.**
4. **State and prove Bayes' theorem.**
5. **Explain the concept of random variables and their types.**
6. **Find the probability distribution of a binomial random variable.**
7. **Explain the difference between discrete and continuous probability distributions.**
8. **Find the mean and variance of a binomial distribution.**
9. **Define the Poisson distribution and state its properties.**
10. **Solve problems on probability using permutations and combinations.**

Important Statistics Questions

1. **Define statistics. Explain the types of statistics.**
2. **Explain different types of data with examples.**
3. **What are measures of central tendency? Explain mean, median, and mode with formulas.**
4. **Calculate the mean, median, and mode for a given data set.**
5. **Explain the measures of dispersion. Calculate variance and standard deviation.**
6. **What is the difference between population and sample?**
7. **Explain correlation and regression with formulas.**
8. **Find the correlation coefficient for given data.**
9. **Explain the concept of skewness and kurtosis.**
10. **Discuss the importance and method of sampling in statistics.**

Long Code/Problem-Based Questions (Important for Board Exam)

- **Find mean, median, mode, variance, and standard deviation for a given data set.**
- **Solve problems on conditional probability and Bayes' theorem.**
- **Find the probability of events using binomial or Poisson distribution.**
- **Calculate the correlation coefficient and regression equations for bivariate data.**

Absolutely! Here are some important **equation-based solved questions** for Probability and Statistics for BCA 3rd Semester, along with step-by-step solutions.

---

1. Find the Mean, Median, Mode, Variance, and Standard Deviation for the following data:

Class Interval	Frequency (f)
0 - 10	5
10 - 20	8
20 - 30	12
30 - 40	7
40 - 50	3

Solution:

Step 1: Calculate the class marks (x):

Class Interval	Frequency (f)	Mid-point (x)
0 - 10	5	5
10 - 20	8	15
20 - 30	12	25
30 - 40	7	35
40 - 50	3	45

Step 2: Calculate f×xf \times x:

Frequency (f)	Mid-point (x)	f×xf \times x
5	5	25
8	15	120

Frequency (f)	Mid-point (x)	f×xf \times x
12	25	300
7	35	245
3	45	135

Sum of frequencies, Σf=5+8+12+7+3=35\Sigma f = 5 + 8 + 12 + 7 + 3 = 35

Sum of f×xf \times x, Σfx=25+120+300+245+135=825\Sigma fx = 25 + 120 + 300 + 245 + 135 = 825

Mean (x¯\bar{x}):

x¯=ΣfxΣf=82535=23.57\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{825}{35} = 23.57

Median:

- Find cumulative frequency (CF):

Class Interval	Frequency (f)	CF
0 - 10	5	5
10 - 20	8	13
20 - 30	12	25
30 - 40	7	32
40 - 50	3	35

- Median class: N2=352=17.5\frac{N}{2} = \frac{35}{2} = 17.5, lies in the class where CF ≥ 17.5 → **20-30**

Median formula:

Median=l+(N2−Ffm)×h\text{Median} = l + \left(\frac{\frac{N}{2} - F}{f_m} \right) \times h

Where:

- l=20l = 20 (lower boundary of median class)
- N=35N = 35
- F=13F = 13 (CF before median class)
- fm=12f_m = 12 (frequency of median class)
- h=10h = 10 (class width)

Calculate:

Median=20+(17.5−1312)×10=20+(4.512)×10=20+3.75=23.75\text{Median} = 20 + \left(\frac{17.5 - 13}{12} \right) \times 10 = 20 + \left(\frac{4.5}{12} \right) \times 10 = 20 + 3.75 = 23.75

Mode:

Mode class = class with highest frequency = 20 - 30 (frequency = 12)

Mode formula:

Mode=l+(f1−f02f1−f0−f2)×h\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h

Where:

- l=20l = 20 (lower boundary of mode class)
- f1=12f_1 = 12 (frequency of mode class)
- f0=8f_0 = 8 (frequency of previous class)
- f2=7f_2 = 7 (frequency of next class)
- h=10h = 10 (class width)

Calculate:

Mode=20+(12−82×12−8−7)×10=20+(424−15)×10=20+(49)×10=20+4.44=24.44\text{Mode} = 20 + \left(\frac{12 - 8}{2 \times 12 - 8 - 7} \right) \times 10 = 20 + \left(\frac{4}{24 - 15} \right) \times 10 = 20 + \left(\frac{4}{9} \right) \times 10 = 20 + 4.44 = 24.44

Variance (σ2\sigma^2) and Standard Deviation (σ\sigma):

Calculate f×x2f \times x^2:

Frequency (f)	Mid-point (x)	x2x^2	f×x2f \times x^2
5	5	25	125
8	15	225	1800

Frequency (f)	Mid-point (x)	x2x^2	f×x2f \times x^2
12	25	625	7500
7	35	1225	8575
3	45	2025	6075

Sum of fx2f x^2, Σfx2=125+1800+7500+8575+6075=23,075\Sigma f x^2 = 125 + 1800 + 7500 + 8575 + 6075 = 23,075

Variance formula:

σ2=Σfx2Σf−(ΣfxΣf)2=2307535−(23.57)2=659.29−555.36=103.93\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f} \right)^2 = \frac{23075}{35} - (23.57)^2 = 659.29 - 555.36 = 103.93

Standard deviation:

σ=103.93=10.20\sigma = \sqrt{103.93} = 10.20

Final answers:

- **Mean** = 23.57
- **Median** = 23.75
- **Mode** = 24.44
- **Variance** = 103.93
- **Standard Deviation** = 10.20

---

2. Solve a problem on Conditional Probability:

Given:

- Probability that a student passes Mathematics = 0.7
- Probability that a student passes English = 0.8
- Probability that a student passes both subjects = 0.6

Find:

- Probability that a student passes Mathematics if he passes English.

Solution:
P(M)=0.7,P(E)=0.8,P(M∩E)=0.6P(M) = 0.7, \quad P(E) = 0.8, \quad P(M \cap E) = 0.6

Conditional probability formula:

P(M|E)=P(M∩E)P(E)=0.60.8=0.75P(M | E) = \frac{P(M \cap E)}{P(E)} = \frac{0.6}{0.8} = 0.75

Answer: Probability that a student passes Mathematics given he passes English is **0.75**.

---

3. Solve a Bayes' Theorem problem:

There are two boxes:

- Box 1 contains 3 red and 2 green balls.
- Box 2 contains 4 red and 4 green balls.

A box is selected at random and one ball is drawn. If the ball drawn is red, find the probability that it is drawn from Box 1.

Solution:

Define events:

- B1B_1: Box 1 selected
- B2B_2: Box 2 selected
- RR: Red ball drawn

Given:

P(B1)=P(B2)=12P(B_1) = P(B_2) = \frac{1}{2} P(R|B1)=35,P(R|B2)=48=12P(R | B_1) = \frac{3}{5}, \quad P(R | B_2) = \frac{4}{8} = \frac{1}{2}

Using Bayes' theorem:

P(B1|R)=P(R|B1)P(B1)P(R|B1)P(B1)+P(R|B2)P(B2)=35×1235×12+12×12=310310+14=310310+2.510=3/105.5/10=35.5=611≈0.545P(B_1 | R) = \frac{P(R | B_1) P(B_1)}{P(R | B_1) P(B_1) + P(R | B_2) P(B_2)} = \frac{\frac{3}{5} \times \frac{1}{2}}{\frac{3}{5} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}} = \frac{\frac{3}{10}}{\frac{3}{10} + \frac{1}{4}} = \frac{\frac{3}{10}}{\frac{3}{10} + \frac{2.5}{10}} = \frac{3/10}{5.5/10} = \frac{3}{5.5} = \frac{6}{11} \approx 0.545

Answer: Probability that the ball is from Box 1 given it is red = **6/11 ≈ 0.545**

Got it! Here's a **long, detailed question with a complete solution** for **Probability and Statistics** suitable for BCA 3rd semester exams:

Question (Long):

The following table shows the marks obtained by 35 students in a statistics exam:

Marks Interval	Number of Students (f)
0 - 10	5
10 - 20	8
20 - 30	12
30 - 40	7
40 - 50	3

1. Calculate the **mean**, **median**, and **mode** of the marks.
2. Calculate the **variance** and **standard deviation** of the marks.
3. Find the **probability** that a randomly selected student scored:

   a) More than 30 marks

   b) Between 10 and 30 marks

Solution:

Step 1: Calculate Class Midpoints (x)

Marks Interval	f	Mid-point (x)
0 - 10	5	5
10 - 20	8	15

Marks Interval	f	Mid-point (x)
20 - 30	12	25
30 - 40	7	35
40 - 50	3	45

Step 2: Calculate Mean

Calculate f×xf \times x:

f	x	f×xf \times x
5	5	25
8	15	120
12	25	300
7	35	245
3	45	135

Sum of frequencies: N=5+8+12+7+3=35N = 5 + 8 + 12 + 7 + 3 = 35

Sum of f×xf \times x: Σfx=25+120+300+245+135=825\Sigma fx = 25 + 120 + 300 + 245 + 135 = 825

Mean (x¯\bar{x}):

x¯=ΣfxN=82535=23.57\bar{x} = \frac{\Sigma fx}{N} = \frac{825}{35} = 23.57

Step 3: Calculate Median

Calculate cumulative frequencies (CF):

Marks Interval	f	CF
0 - 10	5	5
10 - 20	8	13

Marks Interval	f	CF
20 - 30	12	25
30 - 40	7	32
40 - 50	3	35

Median class is where CF ≥ N2=17.5\frac{N}{2} = 17.5 → 20 - 30

Median formula:

Median=l+(N2−Ffm)×h\text{Median} = l + \left(\frac{\frac{N}{2} - F}{f_m} \right) \times h

Where:

- l=20l = 20 (lower boundary of median class)
- F=13F = 13 (CF before median class)
- fm=12f_m = 12 (frequency of median class)
- h=10h = 10 (class width)

Calculate:

Median=20+(17.5−1312)×10=20+4.512×10=20+3.75=23.75\text{Median} = 20 + \left(\frac{17.5 - 13}{12} \right) \times 10 = 20 + \frac{4.5}{12} \times 10 = 20 + 3.75 = 23.75

Step 4: Calculate Mode

Mode class is the class with the highest frequency → 20 - 30 (frequency 12)

Mode formula:

Mode=l+(f1−f02f1−f0−f2)×h\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h

Where:

- l=20l = 20 (lower boundary of mode class)
- f1=12f_1 = 12 (frequency of mode class)
- f0=8f_0 = 8 (frequency of previous class)
- f2=7f_2 = 7 (frequency of next class)
- h=10h = 10 (class width)

Calculate:

Mode=20+(12−82×12−8−7)×10=20+424−15×10=20+49×10=20+4.44=24.44\text{Mode} = 20 + \left(\frac{12 - 8}{2 \times 12 - 8 - 7} \right) \times 10 = 20 + \frac{4}{24 - 15} \times 10 = 20 + \frac{4}{9} \times 10 = 20 + 4.44 = 24.44

Step 5: Calculate Variance and Standard Deviation

Calculate f×x2f \times x^2:

f	x	x2x^2	f×x2f \times x^2
5	5	25	125
8	15	225	1800
12	25	625	7500
7	35	1225	8575
3	45	2025	6075

Sum Σfx2=125+1800+7500+8575+6075=23075\Sigma f x^2 = 125 + 1800 + 7500 + 8575 + 6075 = 23075

Variance:

σ2=Σfx2N−(ΣfxN)2=2307535−(23.57)2=659.29−555.36=103.93\sigma^2 = \frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N} \right)^2 = \frac{23075}{35} - (23.57)^2 = 659.29 - 555.36 = 103.93

Standard Deviation:

σ=103.93=10.20\sigma = \sqrt{103.93} = 10.20

Step 6: Find Required Probabilities

Total students, N=35N = 35

a) Probability that a student scores **more than 30**:

Number of students scoring more than 30 = Frequency of classes 30-40 and 40-50 = 7+3=107 + 3 = 10

P(marks>30)=1035=27≈0.286P(\text{marks} > 30) = \frac{10}{35} = \frac{2}{7} \approx 0.286

b) Probability that a student scores **between 10 and 30**:

Number of students scoring between 10 and 30 = Frequency of classes 10-20 and 20-30 = 8+12=208 + 12 = 20

P(10<marks<30)=2035=47≈0.571P(10 < \text{marks} < 30) = \frac{20}{35} = \frac{4}{7} \approx 0.571

Final Summary:

Measure	Value
Mean	23.57
Median	23.75
Mode	24.44
Variance	103.93
Standard Deviation	10.20
Probability (Marks > 30)	0.286
Probability (10 < Marks < 30)	0.571

Sure! Here's a **long question with detailed solution** from another important chapter of Probability & Statistics — **Correlation and Regression** — which is commonly asked in BCA 3rd sem exams.

Chapter: Correlation and Regression

Question (Long):

The following table shows the marks obtained by 6 students in Mathematics (X) and Statistics (Y):

Student	Mathematics (X)	Statistics (Y)
1	45	50

Student	Mathematics (X)	Statistics (Y)
2	50	55
3	55	52
4	60	60
5	65	65
6	70	70

1. Calculate the **correlation coefficient** between Mathematics and Statistics marks.
2. Find the **regression equations** of Y on X and X on Y.
3. Interpret the results.

Solution:

Step 1: Prepare the table with necessary values:

Student	X	Y	X2X^2	Y2Y^2	XYXY
1	45	50	2025	2500	2250
2	50	55	2500	3025	2750
3	55	52	3025	2704	2860
4	60	60	3600	3600	3600
5	65	65	4225	4225	4225
6	70	70	4900	4900	4900

Calculate sums:

ΣX=45+50+55+60+65+70=345\Sigma X = 45 + 50 + 55 + 60 + 65 + 70 = 345
ΣY=50+55+52+60+65+70=352\Sigma Y = 50 + 55 + 52 + 60 + 65 + 70 = 352
ΣX2=2025+2500+3025+3600+4225+4900=20275\Sigma X^2 = 2025 + 2500 + 3025 + 3600 + 4225 + 4900 = 20275 ΣY2=2500+3025+2704+3600+4225+4900=20954\Sigma Y^2 = 2500 + 3025 + 2704 + 3600 + 4225 + 4900 = 20954 ΣXY=2250+2750+2860+3600+4225+4900=20585\Sigma XY = 2250 + 2750 + 2860 + 3600 + 4225 + 4900 = 20585

Step 2: Calculate the correlation coefficient rr:

Formula:

r=nΣXY−ΣXΣY(nΣX2−(ΣX)2)(nΣY2−(ΣY)2)r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{ \left(n\Sigma X^2 - (\Sigma X)^2 \right) \left(n\Sigma Y^2 - (\Sigma Y)^2 \right)}}

Where n=6n = 6.

Calculate numerator:

6×20585−345×352=123510−121440=20706 \times 20585 - 345 \times 352 = 123510 - 121440 = 2070

Calculate denominator:

(6×20275−3452)(6×20954−3522)=(121650−119025)(125724−123904)=2625×1820=4777500=2185.79\sqrt{(6 \times 20275 - 345^2)(6 \times 20954 - 352^2)} = \sqrt{(121650 - 119025)(125724 - 123904)} = \sqrt{2625 \times 1820} = \sqrt{4777500} = 2185.79

Therefore:

r=20702185.79=0.947r = \frac{2070}{2185.79} = 0.947

Step 3: Calculate Regression Equations

Regression of Y on X (predict Y from X):

bY|X=nΣXY−ΣXΣYnΣX2−(ΣX)2=2070121650−119025=20702625=0.7886b_{Y|X} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} = \frac{2070}{121650 - 119025} = \frac{2070}{2625} = 0.7886

Mean values:

X¯=3456=57.5,Y¯=3526=58.67\bar{X} = \frac{345}{6} = 57.5, \quad \bar{Y} = \frac{352}{6} = 58.67

Equation of regression line:

Y−Y¯=bY|X(X−X¯)Y - \bar{Y} = b_{Y|X} (X - \bar{X}) Y−58.67=0.7886(X−57.5)Y - 58.67 = 0.7886 (X - 57.5) Y=58.67+0.7886(X−57.5)Y = 58.67 + 0.7886 (X - 57.5)

Regression of X on Y (predict X from Y):

$$b_{X|Y} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma Y^2 - (\Sigma Y)^2} = \frac{2070}{125724 - 123904} = \frac{2070}{1820} = 1.137$$

Equation of regression line:

$$X - \bar{X} = b_{X|Y} (Y - \bar{Y}) \quad X - 57.5 = 1.137 (Y - 58.67)$$
$$X = 57.5 + 1.137 (Y - 58.67)$$

Step 4: Interpretation

- The **correlation coefficient** $r = 0.947$ shows a **strong positive linear relationship** between Mathematics and Statistics marks.
- The regression equations allow us to **predict** one variable given the other. For example, if a student scores 60 in Mathematics:

$$Y = 58.67 + 0.7886 (60 - 57.5) = 58.67 + 0.7886 \times 2.5 = 58.67 + 1.97 = 60.64$$

So, predicted Statistics marks ≈ 60.64.

Great! I'll prepare **long, solved questions** with detailed answers for these important chapters in Probability & Statistics for your BCA 3rd sem exam:

---

1. Probability Distributions (Binomial & Poisson)

Question:

A factory produces light bulbs. The probability of producing a defective bulb is 0.02. If bulbs are produced in batches of 100, find:
a) The probability that exactly 3 bulbs are defective.
b) The mean and variance of defective bulbs in a batch.

Solution:

Given:
n=100n = 100, p=0.02p = 0.02, q=1−p=0.98q = 1-p = 0.98

a) Using **Binomial distribution formula**:

P(X=x)=(nx)pxqn−xP(X = x) = \binom{n}{x} p^x q^{n-x}

Find P(X=3)P(X=3):

P(3)=(1003)(0.02)3(0.98)97P(3) = \binom{100}{3} (0.02)^3 (0.98)^{97}

Calculate:

(1003)=100×99×983×2×1=161700\binom{100}{3} = \frac{100 \times 99 \times 98}{3 \times 2 \times 1} = 161700 P(3)=161700×(0.000008)×(0.133)≈161700×1.064×10−6=0.172P(3) = 161700 \times (0.000008) \times (0.133) \approx 161700 \times 1.064 \times 10^{-6} = 0.172

Probability exactly 3 defective bulbs = 0.172

---

b) Mean and variance of binomial distribution:

Mean=np=100×0.02=2\text{Mean} = np = 100 \times 0.02 = 2
Variance=npq=100×0.02×0.98=1.96\text{Variance} = npq = 100 \times 0.02 \times 0.98 = 1.96

---

2. Sampling

Question:

Explain the difference between **Population** and **Sample**. Describe the advantages of sampling over census.

Answer:

- **Population:** The entire group or set of items/people under study. Example: All students of a university.
- **Sample:** A subset of the population selected for analysis. Example: 200 students selected randomly from the university.

Advantages of Sampling over Census:

1. **Cost-effective:** Sampling requires fewer resources than studying the entire population.
2. **Time-saving:** Sampling is quicker to conduct than a census.
3. **Feasibility:** Sometimes census is impractical or impossible; sampling provides a viable alternative.
4. **Data accuracy:** Sampling can reduce errors due to better control and handling.

---

3. Hypothesis Testing

Question:

A manufacturer claims that the mean lifetime of a bulb is 1200 hours. A sample of 50 bulbs gave a mean lifetime of 1180 hours with a standard deviation of 100 hours. Test the claim at 5% level of significance.

Solution:

Step 1: State hypotheses

H0:μ=1200(claim)H_0: \mu = 1200 \quad (\text{claim}) Ha:μ≠1200(alternative)H_a: \mu \neq 1200 \quad (\text{alternative})

Step 2: Given data

- Sample size n=50n = 50
- Sample mean x¯=1180\bar{x} = 1180
- Standard deviation s=100s = 100
- Significance level α=0.05\alpha = 0.05

Step 3: Test statistic (Z-test since n>30n > 30)

Z=x¯−μs/n=1180−1200100/50=−20100/7.07=−2014.14=−1.414Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1180 - 1200}{100/\sqrt{50}} = \frac{-20}{100/7.07} = \frac{-20}{14.14} = -1.414

Step 4: Critical value

At 5% significance (two-tailed), critical Z values = ±1.96

Step 5: Decision

- Since −1.96<−1.414<1.96-1.96 < -1.414 < 1.96, we **fail to reject H0H_0**.
- There is not enough evidence to reject the manufacturer's claim.

Absolutely! Here's a **long question with detailed solution** on **ANOVA (Analysis of Variance)** suitable for your BCA 3rd semester exam:

Question (ANOVA):

A company wants to test whether three different training programs produce different average scores in employee performance. The scores of employees after training are as follows:

Program A	Program B	Program C
85	78	90
88	74	85
90	80	88
86	76	84

Using **one-way ANOVA** at 5% significance level, test whether there is a significant difference between the mean scores of the three programs.

Solution:

Step 1: Organize the data

Program A	Program B	Program C
85	78	90
88	74	85
90	80	88
86	76	84

Step 2: Calculate group means and overall mean

- n1=n2=n3=4n_1 = n_2 = n_3 = 4 (number of observations per group)

Calculate sums and means:

Program	Sum	Mean
A	85+88+90+86 = 349	X¯1=3494=87.25\bar{X}_1 = \frac{349}{4} = 87.25
B	78+74+80+76 = 308	X¯2=3084=77.00\bar{X}_2 = \frac{308}{4} = 77.00
C	90+85+88+84 = 347	X¯3=3474=86.75\bar{X}_3 = \frac{347}{4} = 86.75

Overall mean (X¯\bar{X}):

X¯=349+308+34712=100412=83.67\bar{X} = \frac{349 + 308 + 347}{12} = \frac{1004}{12} = 83.67

Step 3: Calculate Sum of Squares Between Groups (SSB)

SSB=∑ni(X¯i−X¯)2SSB = \sum n_i (\bar{X}_i - \bar{X})^2
SSB=4(87.25−83.67)2+4(77.00−83.67)2+4(86.75−83.67)2SSB = 4 (87.25 - 83.67)^2 + 4 (77.00 - 83.67)^2 + 4 (86.75 - 83.67)^2

Calculate each term:

- (87.25−83.67)2=(3.58)2=12.8164(87.25 - 83.67)^2 = (3.58)^2 = 12.8164
- (77.00−83.67)2=(−6.67)2=44.4889(77.00 - 83.67)^2 = (-6.67)^2 = 44.4889
- (86.75−83.67)2=(3.08)2=9.4864(86.75 - 83.67)^2 = (3.08)^2 = 9.4864

So,

SSB=4×12.8164+4×44.4889+4×9.4864=51.2656+177.9556+37.9456=267.17SSB = 4 \times 12.8164 + 4 \times 44.4889 + 4 \times 9.4864 = 51.2656 + 177.9556 + 37.9456 = 267.17

Step 4: Calculate Sum of Squares Within Groups (SSW)

Calculate variance within each group:

For Program A:

SSW1=∑(X1i−X¯1)2=(85−87.25)2+(88−87.25)2+(90−87.25)2+(86−87.25)2SSW_1 = \sum (X_{1i} - \bar{X}_1)^2 = (85 - 87.25)^2 + (88 - 87.25)^2 + (90 - 87.25)^2 + (86 - 87.25)^2 =(−2.25)2+(0.75)2+(2.75)2+(−1.25)2=5.0625+0.5625+7.5625+1.5625=14.75= (-2.25)^2 + (0.75)^2 + (2.75)^2 + (-1.25)^2 = 5.0625 + 0.5625 + 7.5625 + 1.5625 = 14.75

For Program B:

SSW2=(78−77)2+(74−77)2+(80−77)2+(76−77)2=1+9+9+1=20SSW_2 = (78 - 77)^2 + (74 - 77)^2 + (80 - 77)^2 + (76 - 77)^2 = 1 + 9 + 9 + 1 = 20

For Program C:

SSW3=(90−86.75)2+(85−86.75)2+(88−86.75)2+(84−86.75)2SSW_3 = (90 - 86.75)^2 + (85 - 86.75)^2 + (88 - 86.75)^2 + (84 - 86.75)^2 =10.5625+3.0625+1.5625+7.5625=22.75= 10.5625 + 3.0625 + 1.5625 + 7.5625 = 22.75

Sum of squares within groups:

SSW=SSW1+SSW2+SSW3=14.75+20+22.75=57.5SSW = SSW_1 + SSW_2 + SSW_3 = 14.75 + 20 + 22.75 = 57.5

Step 5: Calculate degrees of freedom

- Between groups: dfbetween=k−1=3−1=2df_{between} = k - 1 = 3 - 1 = 2
- Within groups: dfwithin=N−k=12−3=9df_{within} = N - k = 12 - 3 = 9
- Total degrees of freedom: dftotal=N−1=11df_{total} = N - 1 = 11

Step 6: Calculate Mean Squares

MSB=SSBdfbetween=267.172=133.59MSB = \frac{SSB}{df_{between}} = \frac{267.17}{2} = 133.59
MSW=SSWdfwithin=57.59=6.39MSW = \frac{SSW}{df_{within}} = \frac{57.5}{9} = 6.39

Step 7: Calculate F-statistic

F=MSBMSW=133.596.39=20.9F = \frac{MSB}{MSW} = \frac{133.59}{6.39} = 20.9

Step 8: Find critical value and conclusion

At 5% significance level, from F-distribution table for df1=2df_1 = 2 and df2=9df_2 = 9, critical F0.05≈4.26F_{0.05} \approx 4.26

Since 20.9>4.2620.9 > 4.26, **reject H0H_0**.

Conclusion:

There is a significant difference in the mean scores of the three training programs.

Absolutely! Here's a **long question with detailed solution** on **ANOVA (Analysis of Variance)** suitable for your BCA 3rd semester exam:

Question (ANOVA):

A company wants to test whether three different training programs produce different average scores in employee performance. The scores of employees after training are as follows:

Program A	Program B	Program C
85	78	90
88	74	85
90	80	88
86	76	84

Using **one-way ANOVA** at 5% significance level, test whether there is a significant difference between the mean scores of the three programs.

Solution:

Step 1: Organize the data

Program A	Program B	Program C
85	78	90
88	74	85
90	80	88
86	76	84

Step 2: Calculate group means and overall mean

- $n_1 = n_2 = n_3 = 4$ (number of observations per group)

Calculate sums and means:

Program	Sum	Mean
A	85+88+90+86 = 349	$\bar{X}_1 = \frac{349}{4} = 87.25$
B	78+74+80+76 = 308	$\bar{X}_2 = \frac{308}{4} = 77.00$
C	90+85+88+84 = 347	$\bar{X}_3 = \frac{347}{4} = 86.75$

Overall mean ($\bar{X}$):

$$\bar{X} = \frac{349 + 308 + 347}{12} = \frac{1004}{12} = 83.67$$

Step 3: Calculate Sum of Squares Between Groups (SSB)

$$SSB = \sum n_i (\bar{X}_i - \bar{X})^2$$

$$SSB = 4 (87.25 - 83.67)^2 + 4 (77.00 - 83.67)^2 + 4 (86.75 - 83.67)^2$$

Calculate each term:

- $(87.25 - 83.67)^2 = (3.58)^2 = 12.8164$
- $(77.00 - 83.67)^2 = (-6.67)^2 = 44.4889$
- $(86.75 - 83.67)^2 = (3.08)^2 = 9.4864$

So,

$$SSB = 4 \times 12.8164 + 4 \times 44.4889 + 4 \times 9.4864 = 51.2656 + 177.9556 + 37.9456 = 267.17$$

Step 4: Calculate Sum of Squares Within Groups (SSW)

Calculate variance within each group:

For Program A:

$$SSW_1 = \sum (X_{1i} - \bar{X}_1)^2 = (85 - 87.25)^2 + (88 - 87.25)^2 + (90 - 87.25)^2 + (86 - 87.25)^2$$

$$= (-2.25)^2 + (0.75)^2 + (2.75)^2 + (-1.25)^2 = 5.0625 + 0.5625 + 7.5625 + 1.5625 = 14.75$$

For Program B:

$$SSW_2 = (78 - 77)^2 + (74 - 77)^2 + (80 - 77)^2 + (76 - 77)^2 = 1 + 9 + 9 + 1 = 20$$

For Program C:

$$SSW_3 = (90 - 86.75)^2 + (85 - 86.75)^2 + (88 - 86.75)^2 + (84 - 86.75)^2$$

$$= 10.5625 + 3.0625 + 1.5625 + 7.5625 = 22.75$$

Sum of squares within groups:

$$SSW = SSW_1 + SSW_2 + SSW_3 = 14.75 + 20 + 22.75 = 57.5$$

Step 5: Calculate degrees of freedom

- Between groups: $df_{between} = k - 1 = 3 - 1 = 2$
- Within groups: $df_{within} = N - k = 12 - 3 = 9$
- Total degrees of freedom: $df_{total} = N - 1 = 11$

Step 6: Calculate Mean Squares

$$MSB = \frac{SSB}{df_{between}} = \frac{267.17}{2} = 133.59$$

$$MSW = \frac{SSW}{df_{within}} = \frac{57.5}{9} = 6.39$$

Step 7: Calculate F-statistic

$$F = \frac{MSB}{MSW} = \frac{133.59}{6.39} = 20.9$$

Step 8: Find critical value and conclusion

At 5% significance level, from F-distribution table for $df_1 = 2$ and $df_2 = 9$, critical $F_{0.05} \approx 4.26$

Since $20.9 > 4.26$, **reject H_0**.

Conclusion:

There is a significant difference in the mean scores of the three training programs.

Great! Here's a detailed solved question and answer for **Regression Analysis**, another important topic in Probability & Statistics for your BCA 3rd semester.

Chapter: Regression Analysis

Question (Long):

The marks obtained by 5 students in Mathematics (X) and Computer Science (Y) are as follows:

Student	Mathematics (X)	Computer Science (Y)
1	40	42
2	50	48
3	60	55
4	70	65
5	80	70

1. Calculate the regression equations of Y on X and X on Y.
2. Predict the Computer Science marks for a student who scored 65 in Mathematics.

Solution:

Step 1: Calculate necessary sums and means

Student	X	Y	X2X^2	Y2Y^2	XYXY
1	40	42	1600	1764	1680
2	50	48	2500	2304	2400
3	60	55	3600	3025	3300
4	70	65	4900	4225	4550
5	80	70	6400	4900	5600

Calculate sums:

ΣX=40+50+60+70+80=300\Sigma X = 40 + 50 + 60 + 70 + 80 = 300 ΣY=42+48+55+65+70=280\Sigma Y = 42 + 48 + 55 + 65 + 70 = 280 ΣX2=1600+2500+3600+4900+6400=19000\Sigma X^2 = 1600 + 2500 + 3600 + 4900 + 6400 = 19000 ΣY2=1764+2304+3025+4225+4900=16218\Sigma Y^2 = 1764 + 2304 + 3025 + 4225 + 4900 = 16218 ΣXY=1680+2400+3300+4550+5600=17,530\Sigma XY = 1680 + 2400 + 3300 + 4550 + 5600 = 17,530

Number of observations, n=5n = 5.

Step 2: Calculate means

X¯=ΣXn=3005=60\bar{X} = \frac{\Sigma X}{n} = \frac{300}{5} = 60 Y¯=ΣYn=2805=56\bar{Y} = \frac{\Sigma Y}{n} = \frac{280}{5} = 56

Step 3: Calculate regression coefficients

Regression of Y on X:

bY|X=nΣXY−ΣXΣYnΣX2−(ΣX)2=5×17530−300×2805×19000−3002b_{Y|X} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} = \frac{5 \times 17530 - 300 \times 280}{5 \times 19000 - 300^2}

Calculate numerator:

5×17530=87650,300×280=840005 \times 17530 = 87650, \quad 300 \times 280 = 84000 87650−84000=365087650 - 84000 = 3650

Calculate denominator:

5×19000=95000,3002=900005 \times 19000 = 95000, \quad 300^2 = 90000 95000−90000=500095000 - 90000 = 5000

So,

bY|X=36505000=0.73b_{Y|X} = \frac{3650}{5000} = 0.73

Regression equation of Y on X:

Y−Y¯=bY|X(X−X¯)Y - \bar{Y} = b_{Y|X} (X - \bar{X}) Y−56=0.73(X−60)Y - 56 = 0.73 (X - 60) Y=56+0.73(X−60)Y = 56 + 0.73 (X - 60)

Regression of X on Y:

bX|Y=nΣXY−ΣXΣYnΣY2−(ΣY)2=36505×16218−2802b_{X|Y} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma Y^2 - (\Sigma Y)^2} = \frac{3650}{5 \times 16218 - 280^2}

Calculate denominator:

$$5 \times 16218 = 81090, \quad 280^2 = 78400$$
$$81090 - 78400 = 2690$$

So,

$$b_{X|Y} = \frac{3650}{2690} = 1.356$$

Regression equation of X on Y:

$$X - \bar{X} = b_{X|Y} (Y - \bar{Y})$$
$$X - 60 = 1.356 (Y - 56)$$
$$X = 60 + 1.356 (Y - 56)$$

Step 4: Predict Computer Science marks for a student scoring 65 in Mathematics

Use regression equation of Y on X:

$$Y = 56 + 0.73 (65 - 60) = 56 + 0.73 \times 5 = 56 + 3.65 = 59.65$$

Final answers:

- Regression equation of Y on X:

  $$Y = 56 + 0.73 (X - 60)$$

- Regression equation of X on Y:

  $$X = 60 + 1.356 (Y - 56)$$

- Predicted Computer Science marks for $X=65$ is approximately **59.65**.

. Probability and Statistics

Group A: Very Short Answer Questions *(10 × 2 = 20)*

1. Define probability.
2. What is a sample space?
3. Define mean and median.
4. What is standard deviation?
5. Define random variable.
6. What is a discrete probability distribution?
7. What is permutation?
8. What is ANOVA?
9. Define null hypothesis.
10. Write the formula for combination.

Group B: Short Answer Questions *(5 × 6 = 30)*

11. Find the mean and variance of: 2, 4, 6, 8, 10.
12. A coin is tossed 3 times. Find the probability of getting exactly 2 heads.
13. Differentiate between population and sample.
14. What are the types of sampling methods?
15. Solve: A committee of 3 is to be formed from 5 boys and 4 girls. In how many ways can it be done?
16. Explain types of correlation.

Group C: Long Answer Questions *(2 × 15 = 30)*

17. Solve: Fit a binomial distribution for the following data:

x:	0	1	2	3	4
f:	5	10	20	30	35

18. Explain and perform one-way ANOVA for a given dataset.
19. Derive the formula for standard deviation and explain its significance.

✓ Probability & Statistics Revision Notes

---

1. Basic Concepts of Probability

- **Experiment:** Process with uncertain outcomes (e.g., tossing a coin).
- **Sample Space (S):** Set of all possible outcomes.
- **Event (E):** Subset of sample space.

---

2. Probability

- Probability of an event EE is:

P(E)=Number of favorable outcomesTotal number of outcomeswhere 0≤P(E)≤1P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} \quad \text{where } 0 \leq P(E) \leq 1

- **Certain event:** P(E)=1P(E) = 1
- **Impossible event:** P(E)=0P(E) = 0

---

3. Types of Events

- **Independent Events:** Occurrence of one does not affect the other.
- **Mutually Exclusive Events:** Cannot occur simultaneously.
- **Complementary Events:** P(E′)=1−P(E)P(E') = 1 - P(E).

---

4. Laws of Probability

- P(S)=1P(S) = 1
- Addition Rule:

P(A∪B)=P(A)+P(B)−P(A∩B)P(A \cup B) = P(A) + P(B) - P(A \cap B)

- Multiplication Rule (for independent events):

P(A∩B)=P(A)×P(B)P(A \cap B) = P(A) \times P(B)

---

5. Random Variable

- Variable taking numerical values assigned to outcomes.
- **Discrete:** Finite or countable values.
- **Continuous:** Infinite values in interval.

---

6. Mean, Median, Mode

- **Mean (Average):**

x¯=∑xin\bar{x} = \frac{\sum x_i}{n}

- **Median:** Middle value in ordered data.
- **Mode:** Most frequent value.

---

7. Variance and Standard Deviation

- **Variance:**

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

- **Standard Deviation:**

$$\sigma = \sqrt{\sigma^2}$$

---

8. Probability Distributions

- **Binomial Distribution:**

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- **Poisson Distribution:** For rare events.

---

9. Example Problem:

Q: A die is rolled once. Find the probability of getting an even number.

Solution:
Sample space $S = \{1, 2, 3, 4, 5, 6\}$
Favorable outcomes $E = \{2, 4, 6\}$

$$P(E) = \frac{3}{6} = \frac{1}{2}$$

Solved Probability Questions

---

1. Question:

A bag contains 5 red, 3 green, and 2 blue balls. One ball is drawn at random. Find the probability that the ball drawn is:

a) Red
b) Not green
c) Blue or red

Solution:

- Total balls = 5 + 3 + 2 = 10

a) Probability (Red) = Number of red balls / Total balls

P(R)=510=12P(R) = \frac{5}{10} = \frac{1}{2}

b) Probability (Not green) = 1 – Probability (Green)

P(not green)=1−310=710P(\text{not green}) = 1 - \frac{3}{10} = \frac{7}{10}

c) Probability (Blue or Red) = P(Blue) + P(Red) (since mutually exclusive)

P(B∪R)=210+510=710P(B \cup R) = \frac{2}{10} + \frac{5}{10} = \frac{7}{10}

---

2. Question:

Two coins are tossed together. Find the probability of getting:
a) Two heads
b) At least one tail
c) No head

Solution:

- Sample space: HH, HT, TH, TT (4 outcomes)

a) Two heads (HH)

P=14P = \frac{1}{4}

b) At least one tail = outcomes with tail (HT, TH, TT)

P=34P = \frac{3}{4}

c) No head means both tails (TT)

P=14P = \frac{1}{4}

---

3. Question:

A die is rolled twice. Find the probability that:
a) Sum of the two numbers is 7
b) Both numbers are even
c) At least one number is 4

Solution:

- Total outcomes = $6 \times 6 = 36$

a) Sum = 7 → pairs: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) → 6 outcomes

P=636=16P = \frac{6}{36} = \frac{1}{6}

b) Both even → even numbers: 2,4,6
Pairs: (2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6) → 9 outcomes

P=936=14P = \frac{9}{36} = \frac{1}{4}

c) At least one 4 → Total outcomes with no 4 in either die = $5 \times 5 = 25$
So, outcomes with at least one 4 = $36 - 25 = 11$

P=1136P = \frac{11}{36}

---

4. Question:

If two cards are drawn successively without replacement from a well-shuffled deck of 52 cards, find the probability that:
a) Both cards are aces
b) First card is king and second is queen

Solution:

a) Both aces:

- Number of aces = 4
- Probability 1st ace = 4/52
- Probability 2nd ace = 3/51 (since no replacement)

P=452×351=122652=1221P = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}

b) First king, second queen:

- Number of kings = 4, queens = 4
- Probability 1st king = 4/52
- Probability 2nd queen = 4/51

P=452×451=162652=4663P = \frac{4}{52} \times \frac{4}{51} = \frac{16}{2652} = \frac{4}{663}