Correlation & Regression

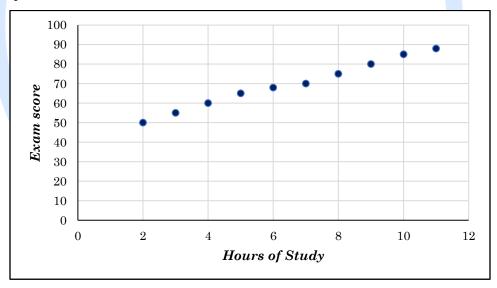
1. A researcher is studying the relationship between hours of study per week and exam scores for a group of students. The data collected from 10 students is as follows:

Student ID	1	2	3	4	5	6	7	8	9	10
Hours of Study	5	7	10	3	8	6	9	4	2	11
Exam Score	65	70	85	55	75	68	80	60	50	88

- a) Determine the direction of association using appropriate graphical method.
- b) Are "Hours of study" and "Exam score" correlated? If yes, comment about the strength of their relationship.

Solution:

a) Scatter plot:



Comment:

b) From graph it is clear that "Hours of study" and "Exam score" are correlated. For comment about the strength of their relationship, we have to calculate the Karl Pearson's Correlation Coefficient.

Here, X = Hours of Study, Y = Exam score

$$\sum X = 65, \sum Y = 696, \sum X^2 = 505, \sum Y^2 = 49868, \sum XY = 4866$$

$$\therefore r = 0.997$$

Comment:

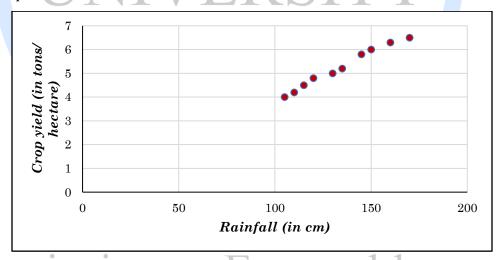
2. A researcher collects data on two variables: annual rainfall (in cm) and crop yield (in tons per hectare) over a 10-year period from a specific region.

Year	1	2	3	4	5	6	7	8	9	10
Rainfall	120	135	150	115	160	145	110	170	105	130
Crop yield	4.8	5.2	6.0	4.5	6.3	5.8	4.2	6.5	4.0	5.0

- a) Apply appropriate correlation measures method to determine the strength of association between Rainfall and Crop yield. Is the relationship between rainfall and crop yield strong or weak?
- b) Create a scatter plot for the given data. Does the scatter plot suggest the same relationship direction as the you got in (a)?
- c) If the correlation is strong, does it imply that an increase in rainfall will always lead to an increase in crop yield? Can we assume causality from correlation in this case?

Solution:

- a) Hints: Karl Pearson's correlation coefficient, r = 0.99
- b) Scatter plot:



Comment:

c) Even if the correlation coefficient is strong, it does not imply causality. While more rainfall generally corresponds to higher crop yield, other factors like soil quality, farming techniques, and pest control could affect the crop yield. Correlation only measures the linear relationship, not cause and effect.

3. A scientist is analyzing the relationship between temperature (in degrees Celsius) and electricity consumption (in kilowatt-hours) in a residential area over 12 days. The aim is to determine how changes in temperature affect electricity usage. To achieve this aim, he did bellow calculation:

$$\sum Temperature = 272; \sum Electricity\ Consumption = 4630$$

$$\sum (Temperature)^2 = 6876; \sum (Electricity\ Consumption)^2 = 1835100$$

$$\sum (Temperature \times E.\ Consumption) = 110790$$

- a) Calculate correlation coefficient using above information.
- b) Suppose 12th value of Temperature is 33 and Electricity consumption is 470. After calculating correlation coefficient, the scientist notices that the 12th value of electricity consumption was misreported and should be 490. Calculate the corrected correlation coefficient.
- c) Using above information, show that correlation coefficient is a symmetric measure.

Solution:

- a) Try yourself.
- b) Corrected correlation coefficient:

Here, X = Temperature, Y = Electricity consumption

	Before correction	After correction		
-	$\sum X = 272$	$\sum X = 272$		
	$\Sigma Y = 4630$	$\Sigma Y = (4630 - 470) + 490 = ???$		
	$\sum X^2 = 6876$	$\sum X^2 = 6876$		
т	$\sum Y^2 = 1835100 \bullet$	$\sum Y^2 = (1835100 - 470^2) + 490^2 = ???$		
1 n	$\sum XY = 110790$	$\sum XY = (110790 - (33 \times 470)) + (33 \times 490) = ???$	C	e

[Using these corrected values, calculate the corrected correlation coefficient]

c) Hints: Show that, $r_{xy} = r_{yx}$

4. A water resources engineer is investigating how annual rainfall (in mm) affects the water level (in meters) of a major river in Bangladesh over 10 years. The engineer collects data from different hydrological stations located near the river to understand the impact of varying rainfall on the river's water level. The data collected is as follows:

Year	1	2	3	4	5	6	7	8	9	10
Water level	8.5	9.0	9.5	8.7	9.2	10.0	8.0	10.5	9.8	11.0
Rainfall	1200	1300	1400	1250	1350	1500	1100	1600	1450	1700

- a) Determine how strongly and in which direction annual rainfall (in mm) is related to the water level (in meters). [Ans: r = 0.99]
- b) Fit an appropriate regression model using variable "Annual Rainfall (in mm)" and "Water Level (in meters)". [Ans: intercept = 2.44, Slope = 0.005]
- c) Provie interpretation about regression parameters.
- d) Predict the water level value when the annual rainfall is 800 mm.
- e) Does your model predict well? Provide a mathematical explanation. [Ans: $R^2 = 0.99$]
- 5. A transportation planner is analyzing how traffic volume (in thousands of vehicles) affects the average travel time (in minutes) on a major highway over 10 days. The goal is to understand how an increase in traffic volume impacts travel time and to develop a predictive model to estimate travel times based on traffic data. The data collected is as follows:

Day	1	2	3	4	5	6	7	8	9	10
Traffic Volume	40	45	50	42	48	55	38	60	53	65
Travel Time	30	35	38	32	36	42	28	46	40	50

- a) Determine how strongly and in which direction traffic volume (in thousands of vehicles) is related to average travel time (in minutes). [Ans: 0.998]
- b) Fit an appropriate regression model using the variables "Traffic Volume (in 000s)" and "Average Travel Time (in minutes)". [Ans: Intercept = 2.33, Slope = 1.25]
- c) Provide interpretation of regression parameters.
- d) Predict the travel time when the traffic volume is 30,000 vehicles.
- e) Does your model predict well? Provide a mathematical explanation based on model fit metrics. [Ans: $R^2 = 0.99$]

6. A company maintains a database that tracks employee performance based on two variables: **Hours of training (Training Hours)**: The number of hours an employee spends on training each month. **Monthly Sales (Sales)**: The sales (in dollars) generated by each employee in a month. A random sample of 10 employees provides the following data:

Employee ID	101	102	103	104	105	106	107	108	109	110
Training hours (X)	5	7	6	8	10	4	9	7	6	5
Sales (\$) (Y)	3000	3500	3200	4000	4500	2800	4200	3800	3400	3100

- a) Calculate the correlation coefficient between the number of training hours and the monthly sales. Interpret the result.
- b) Perform a simple linear regression analysis to model the relationship between **Training Hours (X)** and **Sales (Y)**. Use the data provided to find the regression equation.
- c) Based on the regression equation, predict the monthly sales for an employee who trains for **8 hours**.
- d) Discuss whether the correlation and regression analyses imply a causal relationship between training hours and sales performance.

Solution:

a) **Correlation Coefficient:**

To calculate the correlation coefficient (r), we use the formula:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right] \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right]}}$$

Where: n is the number of data points (10 in this case), X is the number of training hours, Y is the monthly sales. After applying the data, suppose the calculated correlation coefficient is r = 0.91. This suggests a strong positive correlation between training hours and monthly sales, indicating that employees who undergo more training tend to generate higher sales.

b) Simple Linear Regression:

The formula for the linear regression equation is:

$$Y = b_0 + b_1 X$$

Where:
$$b_1 = \frac{\sum XY - \frac{\sum X\Sigma Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$
 (slope), $b_0 = \bar{Y} - b_1 \bar{X}$ (intercept).

After calculation, the regression equation may look like:

Sales =
$$2000 + 250 \times Training_Hours$$

This means that for every additional hour of training, the sales increase by \$250, starting from a base sales amount of \$2000.

- c) **Prediction:** To predict the sales for an employee who trains for 8 hours, plug in X = 8 into the regression equation: Sales = $2000 + 250 \times 8 = 2000 + 2000 = 4000$ Thus, an employee who trains for 8 hours is predicted to generate \$4000 in sales.
- 7. Imagine a small electronics company in Dhaka that has been tracking its daily sales and profits over a period of 6 consecutive days. The company sells various electronic items like mobile phones, laptops, and accessories. The following are the sales ('000 BDT) and profits ('000 BDT) for the last 6 days:

Day	1	2	3	4	5	6
Sales	6	7	_8/	11	12	10
Profit	1	1	3	5	6	4

- a) Calculate the Pearson's correlation coefficient between sales and profit, and interpret your findings. [Ans: r = 0.98]
- b) Use an appropriate graphical method to find the relationship between sales and profit.
- c) Fit a least-squares regression line of profit on sales and interpret the regression parameters. [Intercept = -4.38, Slope = 0.85]
- d) If the sale is 15,000 BDT, estimate the profit.
- e) Evaluate how well the regression line fits the data using the coefficient of determination. $[R^2 = 0.96]$

8. A department of transportation's study on driving speed and mileage for midsize automobile resulted in the following table:

Driving speed	30	40	50	55	25
Mileage	27	25	30	35	22

- a) Is there any relationship between Driving speed and Mileage? Verify your answer. [Ans: 0.89]
- b) Find the regression equation of driving speed on mileage. [Intercept = -23.31, Slope = 2.28]
- c) What will be mileage when speed is 45?
- d) Test the fitness of your regression model with explanation. $[R^2 = 0.79]$



Basic Concepts of Probability

1. In the Engineering department, the academic paths of 348 students, across different years and specializations, are summarized in the contingency table below:

		Sub	jects		
	Mechanical	Electrical	Civil	CS	Total
1st Year	6 <i>x</i>	5 <i>x</i>	4 <i>x</i>	3 <i>x</i>	90
2nd Year	35	A	15	В	90
3rd Year	10y	15 <i>y</i>	12 <i>y</i>	7 <i>y</i>	88
4th Year	15	15	20	С	
Total		90			

- a) Complete the table.
- b) Determine the probability (with interpretation) that a randomly selected student is either in the "4th Year" or choosing "Computer Science (CS)".

Solution:

a) Calculate the value x,

$$6x + 5x + 4x + 3x = 90$$

$$\therefore x = 5$$

Calculate the value y,

$$10y + 15y + 12y + 7y = 88$$

$$\therefore y = 2$$

Now, we reconstruct the table,

		Sub	jects		
	Mechanical	Electrical	Civil	CS	Total
1st Year	30	25	20	15	90
2nd Year	35	20	15	20	90
3rd Year	20	30	24	14	88
4th Year	15	15	20	30	80
Total	100	90	79	79	348
	7 1 1 1	II g	LAU		

b) Let,

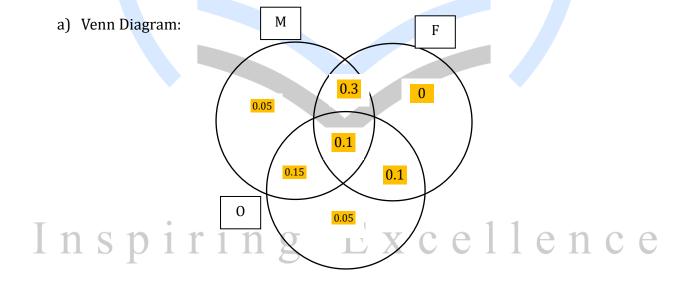
A = Event of 4th year student

B = Event of student choosing CS

Here,
$$P(A) = \frac{80}{348}$$
; $P(B) = \frac{79}{348}$; $P(A \& B) = \frac{30}{348}$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.37 (Approx.)$$

- 2. In a survey of business students, respondents were asked about their preferences for studying three major subjects: Marketing (M), Finance (F), and Operations (O). Each student could choose to study one or more subjects. The probabilities for each subject are as follows: The probability of a student choosing Marketing is 60%. The probability of a student choosing Finance is 50%. The probability of a student choosing Operations is 40%. The probability of a student choosing both Marketing and Finance is 30%. The probability of a student choosing both Finance and Operations is 25%. The probability of a student choosing all three subjects (Marketing, Finance, and Operations) is 10%.
 - a) Construct a Venn diagram to represent the probabilities.
 - b) If a business student is selected randomly, what is the probability that he/she is studying:
 - i. exactly one subjects?
 - ii. At least one subject?
 - iii. At most one subject?
 - iv. No subject?



b)
$$P(Exactly \ one \ subject) = P(Only \ M) + P(Only \ F) + P(Only \ O) =???$$

 $P(At \ least \ one) = P(M \cup F \cup O) =???$
 $P(At \ most \ one) = P(Only \ M) + P(Only \ F) + P(Only \ O) + P(None) =???$

- 3. In a certain residential suburb, 60% of all households get Internet service from the local cable company, 80% get television service from that company, and 50% get both services from that company. If a household is randomly selected, what is the probability that it gets at least one of these two services from the company, and what is the probability that it gets exactly one of these services from the company? **[Try yourself]**
- 4. A and B are two weak students in Statistics. A can answer correctly 15% of the questions and B can answer correctly 10% of the questions. A and B both can answer 2% of the questions. A question is selected at random. Find the probability that (a) at least one of them can answer correctly, (b) No one can answer correctly, (c) only one can answer correctly.

Let us denote the events,

A: A can answer the question correctly

B: *B* can answer the question correctly

Here, P(A) = 0.15; P(B) = 0.10; $P(A \cap B) = 0.02$

- a) $P(at \ least \ one) = P(A \cup B) = P(A) + P(B) = P(A \cap B) = ???$
- b) $P(No \ one) = P(\overline{A \cup B}) = 1 P(A \cup B) = ???$
- c) $P(Only\ one) = P(A \cap \bar{B}) + P(B \cap \bar{A}) = ???$
- 5. In a large corporation, two senior managers, Ms. Smith and Mr. Johnson, are frequently required to travel abroad for business meetings. Ms. Smith travels abroad 70% of the time during a given year, while Mr. Johnson travels abroad 45% of the time during the same period.
 - a) Calculate the probability that on a randomly selected day in the year, both Ms. Smith and Mr. Johnson will be abroad. [Ans: 0.315 or 31.5%]
 - b) Determine the probability that either Ms. Smith or Mr. Johnson will be abroad on a given day. [Ans: 0.835 or 83.5%]
 - c) What is the probability that neither Ms. Smith nor Mr. Johnson will be abroad on a randomly selected day? $[P(\bar{A} \cap \bar{B}) = 1 P(A \cup B)]$

- 6. Tickets are numbered from 1 to 100. They are well shuffled and a ticket is drawn at random. What is the probability that the drawn ticket has
 - a) An old number
 - b) A number 4 or multiple of 4
 - c) A number which is greater than 70
 - d) A number which is a square?

Here are 100 tickets, the total number of exhaustive, mutually exclusive and equally likely cases is 100.

a) Let A denote the event that the ticket drawn has an odd number. Since there 50 odd numbered tickets, the number of cases favorable to the event A is 50,

$$\therefore P(A) = \frac{50}{100} = 0.5$$

Comment:

b) Let B denote the event that the drawn ticket has a number 4 or multiple of 4. The numbers favorable to event "B" are 4, 8, 12, 16, 20 92, 96, 100. The total number of cases will be 25.

$$\therefore P(B) = \frac{25}{100} = \frac{1}{4}$$

Comment:

c) Let C be the event that the drawn ticket has a no greater than 70. Since the No. greater than 70 are 71, 72 100. Therefore 30 cases are favorable to the event C.

$$\therefore P(C) = \frac{30}{100}$$

Comment:

d) Let D be the event that the drawn ticket has a number which is a square. Since the squares between 1 to 100 are 1, 4, 9, 16, 25, 36, 49, 64, 81, 100.

$$\therefore P(D) = \frac{10}{100}$$

Comment:

- 7. Afsana feels that the probability that she will pass mathematics is 2/3 and statistics is 5/6. If the probability that she will pass both the course is 3/5, what is the probability that she will pass at least one of the courses? [Ans: 9/10]
- 8. Suppose A and B are two mutually exclusive events with P(A) = 3.5 and P(B) = 0.15. Find
 - a) $P(A \cup B)$ [Ans: 0.5]
 - b) P(A') [Ans: 0.65]
 - c) $P(A \cap B)$ [Ans: 0]
 - d) $P(A' \cup B')$ [Ans: 1]
 - e) $P(A' \cap B')$ [Ans: 0.5]



Conditional Probability

- 1. Suppose a balanced die is rolled once.
 - a. Find the probability that a number divisible by 3 is rolled given that the die comes up even.
 - b. Find the probability that the die comes up even given that a number divisible by 3 is rolled. [Ans: 0.5]
 - c. Find the probability that a number divisible by 3 is rolled given that die comes up at most 4. [Ans: 0.25]
 - d. Find the probability that the die comes up at most 4 given that a number divisible by 3 is rolled. [Ans: 0.5]

Solution:

BRAC

a) Let,

 $D3 = Event of number division by 3 = {3,6}$

 $E = Event of even number = \{2,4,6\}$

Here, the sample space, $S = \{1,2,3,4,5,6\}$

$$P(D3) = \frac{2}{6}; P(E) = \frac{3}{6}; P(D3 \cap E) = \frac{1}{6}$$

$$\therefore P(D3|E) = \frac{P(D3 \cap E)}{P(E)} = 0.333$$

Comment: Given that die comes up even, there is a 33.3% chance of rolling a number divisible by 3.

- b) Try yourself
- c) Try yourself
- d) Try yourself

- 2. In a certain coastal city, during the hurricane season, it is known that hurricanes occur on 60% of the days. When a hurricane occurs, there is an 85% chance that it will also lead to heavy rainfall.
 - a) Calculate the probability that on a given day during the hurricane season, both a hurricane and heavy rainfall will occur.
 - b) Determine the probability that heavy rainfall will occur given that a hurricane has occurred on a particular day.

a) Let,

H = Event of hurricanes occur

R = Event of heavy rainfall

$$P(H) = 0.6; P(R|H) = 0.85$$

$$\therefore P(R \cap H) = P(R|H) \times P(H) = ???$$

- b) Try yourself.
- 3. In a game of chance, you roll two six-sided dice. One die is red and the other is blue.
 - a) What is the probability that the sum of the numbers rolled is 7 given that the number on the red die is 4? [Ans: 1/6]
 - b) What is the probability that the number on the blue die is 5 given that the sum of the numbers rolled is 9? [Ans: 1/4]
 - c) What is the probability of rolling a sum less than or equal to 6 given that the number on the red die is even? [Ans: 1/3]
 - d) What is the probability that the number on the red die is odd given that the sum of the numbers rolled is 8? [Ans: 2/5]

- 4. A software company employs 80 developers. Out of these, 50 developers are proficient in JavaScript (JS), 40 are proficient in Python, and 20 are proficient in both JavaScript and Python.
 - a) What is the probability that a randomly selected developer is proficient in either JavaScript or Python?
 - If a developer is known to be proficient in Python, what is the probability that they b) are also proficient in JavaScript (i.e., conditional probability)?
 - **Are the events** "Proficient in JavaScript" and "Proficient in Python" independent? Explain your reasoning using probability concepts.

a) Probability of a developer being proficient in JavaScript or Python:

We can use the formula for the union of two events:

$$P(JS \cup Python) = P(JS) + P(Python) - P(JS \cap Python)$$

Where:

$$- P(JS) = \frac{50}{80} = 0.625$$

-
$$P(JS) = \frac{50}{80} = 0.625$$
,
- $P(Python) = \frac{40}{80} = 0.5$,

-
$$P(JS \cap Python) = \frac{20}{80} = 0.25.$$

So,
$$P(JS \cup Python) = 0.625 + 0.5 - 0.25 = 0.875$$

The probability that a developer is proficient in either JavaScript or Python is 0.875.

b) Conditional Probability: Given that the developer is proficient in Python, what is the probability that they are also proficient in JavaScript?

The conditional probability formula is:

$$P(JS|Python) = \frac{P(JS \cap Python)}{P(Python)}$$
Substituting the values:

$$P(JS|Python) = \frac{0.25}{0.5} = 0.5$$

So, the probability that a developer is proficient in JavaScript given that they are proficient in Python is **0.5**.

c) Independence Check:

Two events *A* and *B* are independent if:

$$P(A \cap B) = P(A) \times P(B)$$

Here:

$$P(JS) \times P(Python) = 0.625 \times 0.5 = 0.3125$$

Since $P(JS \cap Python) = 0.25$, which is not equal to 0.3125, the events "Proficient in JavaScript" and "Proficient in Python" are **not independent**.

5. In rolling two balanced dice, if the sum of the two volumes is 8 what is the probability that one of the values is 3?

Solution: Let A be the event that one of the values is 3. Let B be the event that the sum is 8. If we consider the sample space of the experiment consisting of 2 dice, the event AB consists of the sample points (3, 5) and (5, 3). The event B consists of the sample points (2, 6), (3,5), (4,4), (5, 3) and (6, 2). Now,

$$P(AB) = \frac{2}{36}; P(B) = \frac{5}{36}$$

 $\therefore P(A|B) = \frac{P(AB)}{P(B)} = ???$

- 6. Two balanced dice, one black and one red are thrown and the number of dots on their upper faces are noted, let b be the outcomes of the black die and r be the outcomes of the red die. Answer the following:
 - a) List a sample space of the experiment.
 - b) What is the probability that r > 4 and $b \le 5$? [Ans: 0.27]
 - c) What is the probability that the difference of the two dice is less than three? [Ans: 0.66]

- 7. The probability that a person picked at random from a population will exhibit the symptom of certain disease is 0.2, and the probability that a person picked at random has the disease is 0.23. the probability that a person who has the symptom also has the disease is 0.18. A person selected at random from the population does not have the symptom. What is the probability that the person has the disease? [Hints: P(D|S'), Ans: 0.0625]
- 8. A particular medicine was given to a group of people for a specific disease.

	Cured	Not cured
Male	20	15
Female	25	10

One person is selected. What is the probability that,

- a) The person is cured? [Ans: 45/70]
- b) The person is male? [Ans: 35/70]
- c) The person is male and cured? [Ans: 20/70]
- d) The person is female and cured? [Ans: 55/70]
- e) The person is cured given that the person is male? [Ans: 20/35]
- f) The person is cured or not cured? [Ans: 1]
- 9. Two students A and B are asked to develop a computer programme. Previous knowledge tells that A becomes successful in 60% cases and B becomes successful in 70% cases. If they work independently. What is the probability that
 - a) The programme will be developed?
 - b) None will be successful?
 - c) A will be successful under the condition that B fails?

Solution: P(A) = 0.6, P(B) = 0.7. As they work independently, $P(A \cap B) = 0.6 \times 0.7 = 0.42$

- a) $P(A \cup B) = 0.88$
- b) $P(A' \cap B') = 1 P(A \cup B) = 0.12$
- c) P(A|B') = 0.6

With replacement and Without replacement

- 1. A bag contains 5 red balls and 3 green balls. You are asked to draw 2 balls from the bag one after another, with replacement (meaning after drawing a ball, you put it back in the bag).
 - a) What is the probability of drawing two red balls consecutively?
 - b) What is the probability of drawing one red ball and one green ball in two draws, in any order?

Solution:

a) Probability of drawing a red ball on the first draw is 5/8.

Sine the ball is replaced, the probability of drawing a red ball in the second draw is still 5/8. Now,

$$P(2 \ red \ balls) = \frac{5}{8} \times \frac{5}{8} = \frac{25}{64}$$

So, the probability of drawing two red balls consecutively with replace is 25/64.

b) There are two possible favorable outcomes:

Red on the first draw, green on the second draw.

Green on the first draw, red on the second draw.

$$\therefore P(Red\ first, Green\ Second) = \frac{5}{8} \times \frac{3}{8} = \frac{15}{64}$$

And,

$$P(Green\ first, Red\ second) = \frac{3}{8} \times \frac{5}{8} = \frac{15}{64}$$

Now,

$$P(1 \ red, 1 \ green \ in \ any \ order) = \frac{15}{64} + \frac{15}{64} = \frac{15}{32}$$

So, the probability of drawing one red ball and one green ball in two draws is 15/32.

2. A jar contains 5 blue marbles, 4 green marbles, and 3 red marbles. You are asked to draw 3 marbles one after another, without replacement. What is the probability that all three marbles are blue?

Solution:

Probability of drawing a blue marble on the first draw is 5/12 After drawing one blue marble, there are 4 blue marbles left and 11 marbles in total. The probability of drawing a second blue marble is 4/11.

After drawing two blue marbles, there are 3 blue marbles left and 10 marbles in total. The probability of drawing a third blue marble is 3/10.

Alternative:

Selecting 3 marbles from 12 in 12C3 ways.

Selecting 3 blue marbles from 5 blue in 5C3 ways.

$$\therefore P(3 blue marbles) = \frac{5C3}{12C3} = \frac{1}{22}$$

Now, the probability that all three marbles are blue,

$$P(3 \ blue \ marbles) = \frac{5}{12} \times \frac{4}{11} \times \frac{3}{10} = \frac{1}{22}$$

- 3. A box contains seven balls- 2 red, 3 blue, and 2 yellow. Consider an experiment that consists of drawing a ball from the box.
 - a) What is the probability that the first ball drawn is yellow?
 - b) What is the probability that the same-colored ball is drawn twice without replacement?
 - c) What is the probability that the same-colored ball is drawn twice with replacement? [Ans:0.35]

Solution:

- a) $P(Yellow \ on \ the \ first \ draw) = \frac{Number \ of \ yellow \ balls}{Total \ number \ of \ balls} = \frac{2}{7}$
- b) If the first ball is red, $P(Red\ on\ both\ draws) = \frac{2}{7} \times \frac{1}{6}$ If the first ball is blue, $P(Blue\ on\ both\ draws) = \frac{3}{7} \times \frac{2}{6}$

If the first ball is yellow, $P(Yellow \ on \ both \ draws) = \frac{2}{7} \times \frac{1}{6}$

$$\therefore P(\textit{Same colored ball twice with out rep.}) = \left(\frac{2}{7} \times \frac{1}{6}\right) + \left(\frac{3}{7} \times \frac{2}{6}\right) + \left(\frac{2}{7} \times \frac{1}{6}\right)$$

c) Try yourself.

- 4. A box contains 20 bulbs, of which 5 are defective. If 3 of the bulbs are selected at random without replacement, what is the probability that all three bulbs are defective? [Ans: 0.0088]
- 5. A shelf contains 8 fiction books and 12 non-fiction books. You randomly pick 2 books, one after the other, with replacement. What is the probability of picking a fiction book first and a non-fiction book second? [Ans: 6/25]

