PR. 2 Expectation Decider

Q1 What is probability ?

⇒ Probability is a measure of how likely an event is to occur expressed as a number between 0 and 1 (or 0% to 100%)

Formula:

$$P(\text{event}) = \frac{\text{number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

Q2. key probability Terminology.

⇒ 1) experiment :- A process or action whose outcome in uncertain (e.g. selecting a student)

~~2) outcome~~ :-

2) sample space :- the set of all possible outcomes of an experiment (e.g the set of all 200 student)

3) event :- A subset of the sample space (e.g. the event that a student passed the exam)

4) mutually Exclusive event :- events that cannot occur at the same time

5) indipendent Events :- Events where the occurrence of one does not affect the probability of the other

Q3. Give at last three probability event examples from the dataset?

=> 1) Probability of passing the exam

$$\text{Probability} = \frac{104}{200} = 0.52$$

2) probability of High Attendance:

$$\text{Probability} = \frac{85}{200} = 0.425$$

Attendance Abouv 80%

3) probability of high study Hours:

$$\text{probability} = \frac{\text{study Hours}}{\text{sample space}}$$

study Hours more than 20 hours :- 71

$$\text{probability} = \frac{71}{200} = 0.355$$

Q4 Types of probability

1) Empirical probability :- Empirical probability is the calculated based on actual observations or experiments, not theory.

Formula

$$\text{Empirical probability} = \frac{\text{All pass student}}{\text{sample space}}$$

All pass student = 104
Sample space = 200

empirical probability = $\frac{104}{200}$ = 0.52

empirical probability is 0.52

2) Theoretical probability :- Theoretical probability is based on reasoning about the possible outcomes of an event assuming they are equally likely.

~~theori~~ theoretical probability of getting a hands on a single coin toss.

Theoretical probability = 0.5

Q 5) Random variable and probability distribution

i) Define a random variable for the event Number of students passing the final exam out of 3 randomly selected student.

* A list of student is created with their pass / fail status.
* From that list, we first filter only those students who passed the exam
* Than we use random.sample(5) to randomly select 5 students from the pass group.

* Finally, we print the name of the 5 randomly selected passed students.

ii) mean and variance.

n = total number of ~~selected~~ randomly selected student.

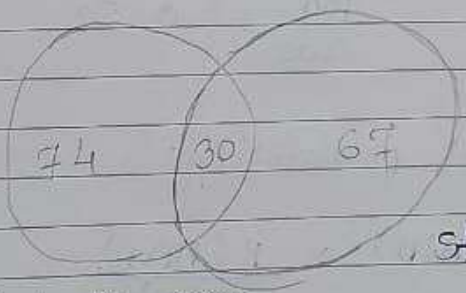
p = probability of randomly selected student of passing.

mean = $n * p$.

variance = $n * p * (1 - p)$

Q6 Venn diagram in probability.

Student pass the exam = 104
Student participates in group discussions = 97



Student pass the exam

Student participates in group discussion.

Q7 probability calculations:

1) Joint probability of participates in group discussion and pass exam.

Joint probability = P(A and B)

Joint probability = (48.5 and 52)

$\frac{97}{200} = 0.485$ $\frac{104}{200} = 0.52$

2) marginal probability of "pass exam"

P(Pass exam) = 52

$\frac{104}{200} = 0.52.$

3) conditional probability of pass exam given participation in group discussion.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

number of students with high study Hours = 71

number of students ~~with~~ who passed and had High study Hours = 68

$$P(A|B) = \frac{\text{[illegible]}}{\text{[illegible]}} = \frac{68}{71} = 0.9577$$

conclusion

The analysis of the synthetic student dataset provides valuable insight for the "Expectation Decider" model. key finding include:

* The overall passing rate in the sample is 52%.

* Study hours show the strongest ~~core~~ correlation with passing, with a conditional probability of 95.77% for students with high study hours.

* Group discussion participation is dependent event with passing suggesting it is a benefical factor.

* A venn diagram is a useful tool to understand the relationship between students who passed the exam and those who participated in the group discussion
1) student who only pass = 74
2) student who only participated = 67
3) Students who did both = 30

* This insights confirm that study hours attendance and group discussion participation are all statically significant factors that should be hevily weighted in the final predictive model

2025/12/12 18:49