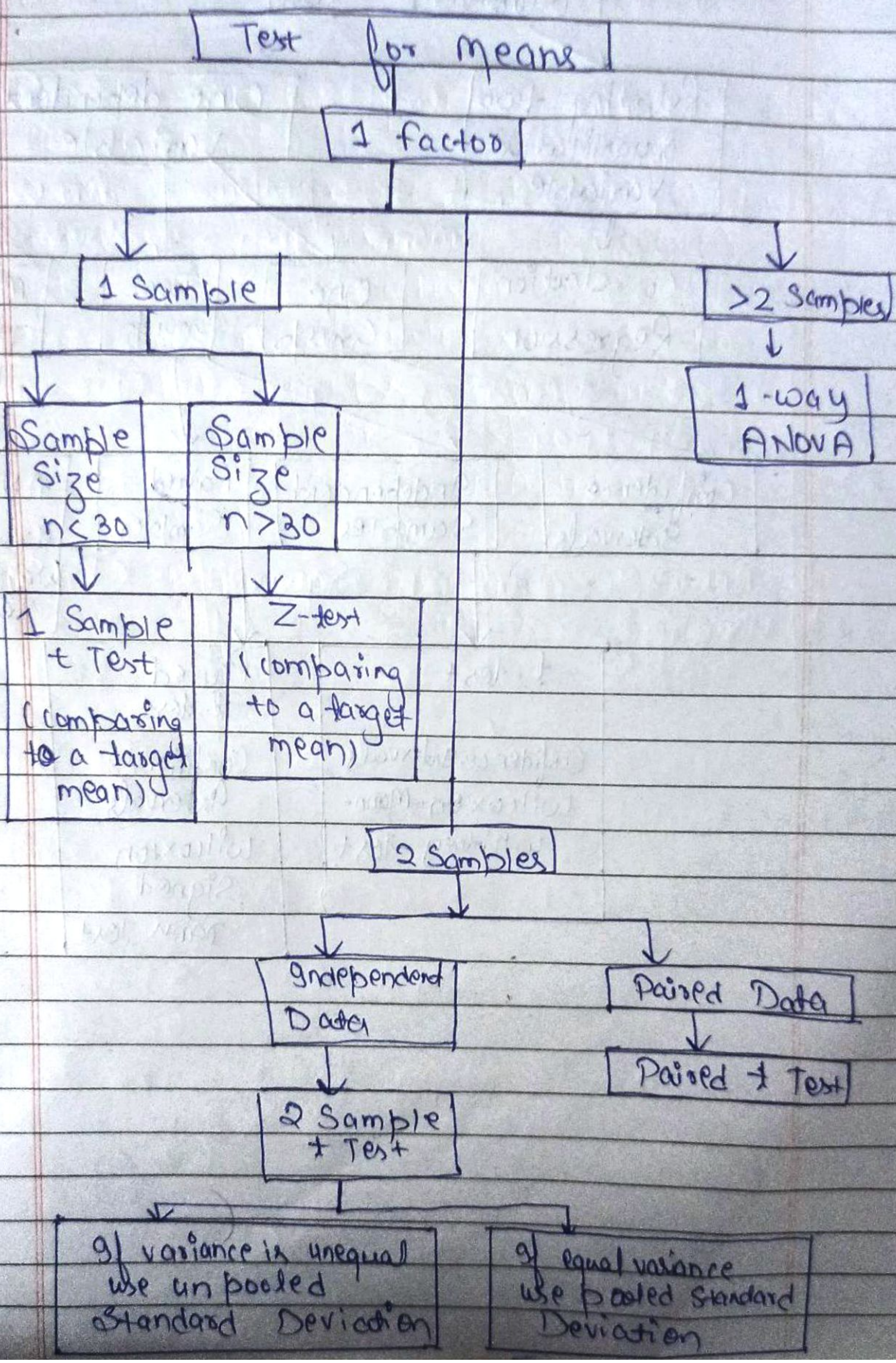
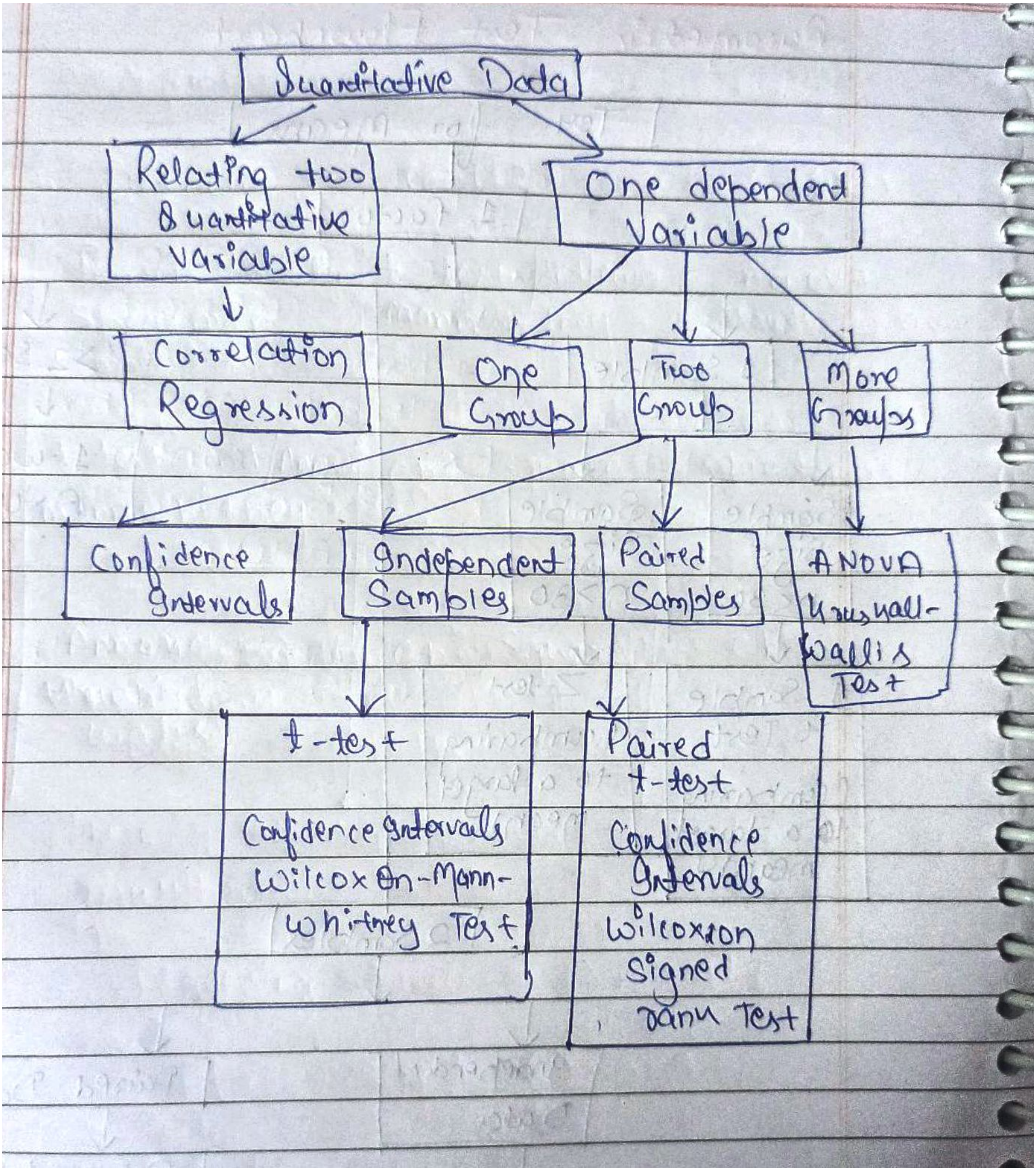
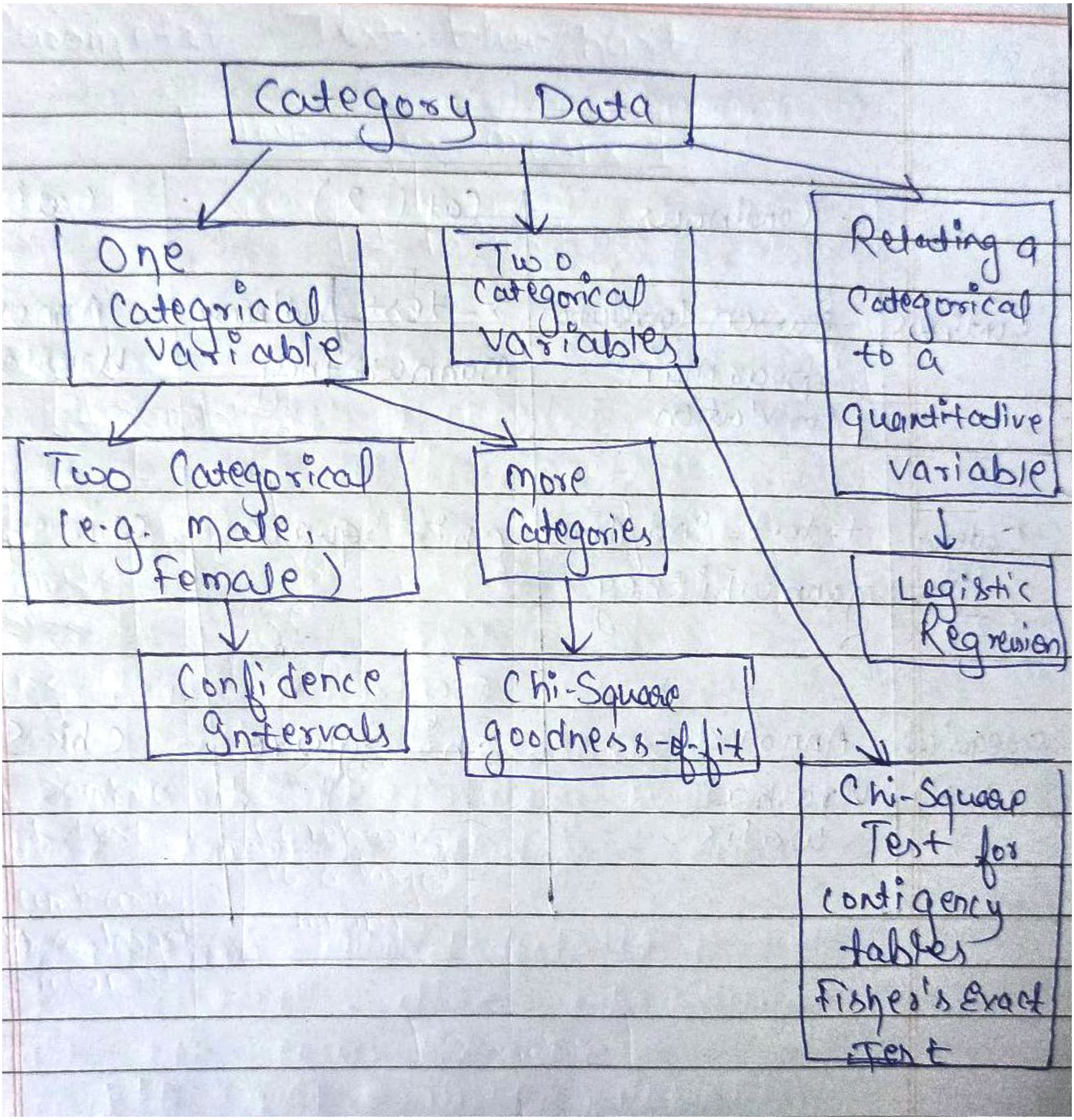
	Continous	Cat (=2)	Cat (>2)
Continous	Pearson Correlation Spearman Correlation	t-test_ind/Annova mann whitney	Annova Kruskal walis
Categorical (2)	t-test_ind/Annova mann whitney	Chi-Square	Chi-Square
categorical (>2)	Annova Kruskal walis	Chi Square	Chi Square

Parametric Test Flowchart




Quantitative Data
Relating two Quantitative variable
One dependend Variable
Correlation Regression
One Group
Two Groups
More Groups
Confidence Intervals
Independent Samples
Paired Samples
ANOVA Kruskall-Wallis Test
t-test
Confidence Intervals
Wilcoxon-Mann-whitney Test
Paired t-test
Confidence Intervals
Wilcoxon Signed rank Test


Category Data
One categrical variable
Two categorical variables
Relating a categorical to a quantitative variable
Two Categorical (e.g. male, female)
more Categories
Logistic Regression
Confidence Intervals
Chi-Square goodness-of-fit
Chi-Square Test for contigency tables
Fisher's Exact Test

Type - I and Type - II Error

Null Hypothesis	True	False
Rejected	Type - I Error False Positive Probability = α	Correct Decision True Positive Probability = $1-\beta$
Not Rejected	Correct Decision True Negative Probability = $1-\alpha$	Type - II Error False Negative Probability = β

Bias :-

In statistics, bias is a term which ~~divides~~ defines the tendency of the measurement process.

It means that it evaluates the over or underestimation of the value of the population parameter.

It is essentially when a model ~~or~~ statistic in unrepresentative of the population, and there are several sources of Bias that cause this.

Types of Statistical Bias :-

1. Selection Bias

Selection Bias is the phenomenon of selecting individuals, groups or data for analysis in such a way that proper randomization is not achieved, ultimately resulting in a sample that is not representative of the population.

Within selection bias there are several types of selection bias:-

- Sampling Bias :- Refers to a Biased sample caused by non-random sampling.

- Time Interval Bias :- Bias caused by intentionally specifying a certain range of time to support the desired conclusion.

2. Susceptibility Bias :- includes clinical susceptibility bias, protopathic bias, and identication bias, which all relate to the idea of potentially mixing up cause

and effect and correlation.

- Confirmatio Bias:- The tendency to favour information that confirms one's beliefs.

- Survivorship Bias:- The phenomen where only those that "survived" a long process are included or excluded in an analysis, thus creating a biased sample.

- Omitted variable Bias:-
This is bias that stems from the absence of relevant variables in a model. In machine learning, removing relevant and/or too many variables result in an underfit model.

- Recall Bias:
Recall Bias is a type of information bias where participants do not 'recall' previous events, memories, or details.

This is also related to recency bias, where we tend to remember things better that have happened more recently.

Observer Bias:-

This is the Bias that stems from the subjective viewpoint of observers and how they asses subjective criteria or record subjective information.

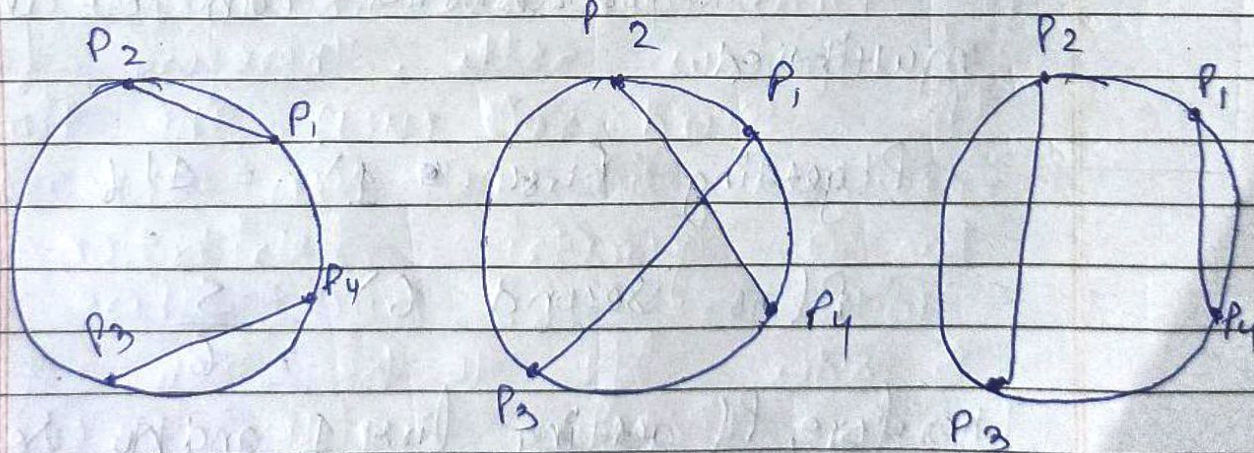
Funding Bias

Also known as sponsorship bias, it is the tendency to skew a study or the results of a study to support financial sponsor.

Probability Interview Problems

(Bloomberg - Easy)
Say you draw a circle and choose two chords at random. What is the probability that these chords intersect.

The two probability that two random chords intersect can be derieved by using a simple counting argument. Suppose that you pick four points at random on thec circle. Label the points according to their polar angles as P1, P2, P3 and P4.



Possible configuration of Two Chords

Consequently, the probability that two random chords intersect is 1/3 because the chords intersect in only one of the three possible arrangements.

Tesla Easy

A fair six-Sided die is rolled twice. What is the probability of getting 1 on the first roll and not getting 6 on the second roll?

The two events mentioned are independents.
The first roll of the die is independent of the second roll.

There, the probabilities can be directly multiplied

P(getting first 1) = 1/6

P(no second 6) = $\frac{5}{6}$

Therefore, P(getting first 1 and no second 6) = $\frac{1}{6} \times \frac{5}{6}$

$= \frac{5}{36}$

Uber - Easy

Say you roll three dice, one by one. What is the probability that you obtain 3 numbers in a strictly ~~incra~~ increasing order?

S = Sample Space for the experiment of throwing ~~dice~~ three dice one by one.

$n(S) = 6 \times 6 \times 6 = 216$

A = event of getting the numbers strictly in increasing order.

$A = \{(1,2,3), (1,2,4), (1,2,5), (1,2,6), (1,3,4), (1,3,5), (1,3,6), (1,4,5), (1,4,6), (1,5,6), (2,3,4), (2,3,5), (2,3,6), (2,4,5), (2,4,6), (2,5,6), (3,4,5), (3,4,6), (3,5,6), (4,5,6)\}$

$n(A) = 20$

$P(A) = \frac{n(A)}{n(S)} = \frac{20}{216} = \frac{5}{54}$

(Citadel : Medium)

What is expected Number of rolls needed to see all 6 sides of a fair die?

Let u denotes the number of sides seen from rolls already. The first roll will always result in a new side being seen. If you have seen u sides, then the probability of rolling an unseen value will be $(6-u)/6$, since there are $6-u$ values you have not seen, and 6 possible outcomes each roll.

Note that each roll is independent of previous rolls. Therefore, for the second roll ($u=1$), the time until a side not seen appears has a geometric distribution with $p = 5/6$, since there are 5 sides left to be seen of the 6. Likewise, after two sides ($u=2$), the time taken is a geometric distribution with $p = 4/6$. This continues for all of the sides until all are seen.

Number of rolls needed, and Thus we have

$$E[X] = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 6 \sum_{p=1}^{6} \frac{1}{p}$$

$$= \frac{60 + 72 + 90 + 120 + 180 + 60}{60}$$

$$= 1 + 1.2 + 1.5 + 2 + 3 + 6$$

$$= 14.7 \text{ rolls}$$