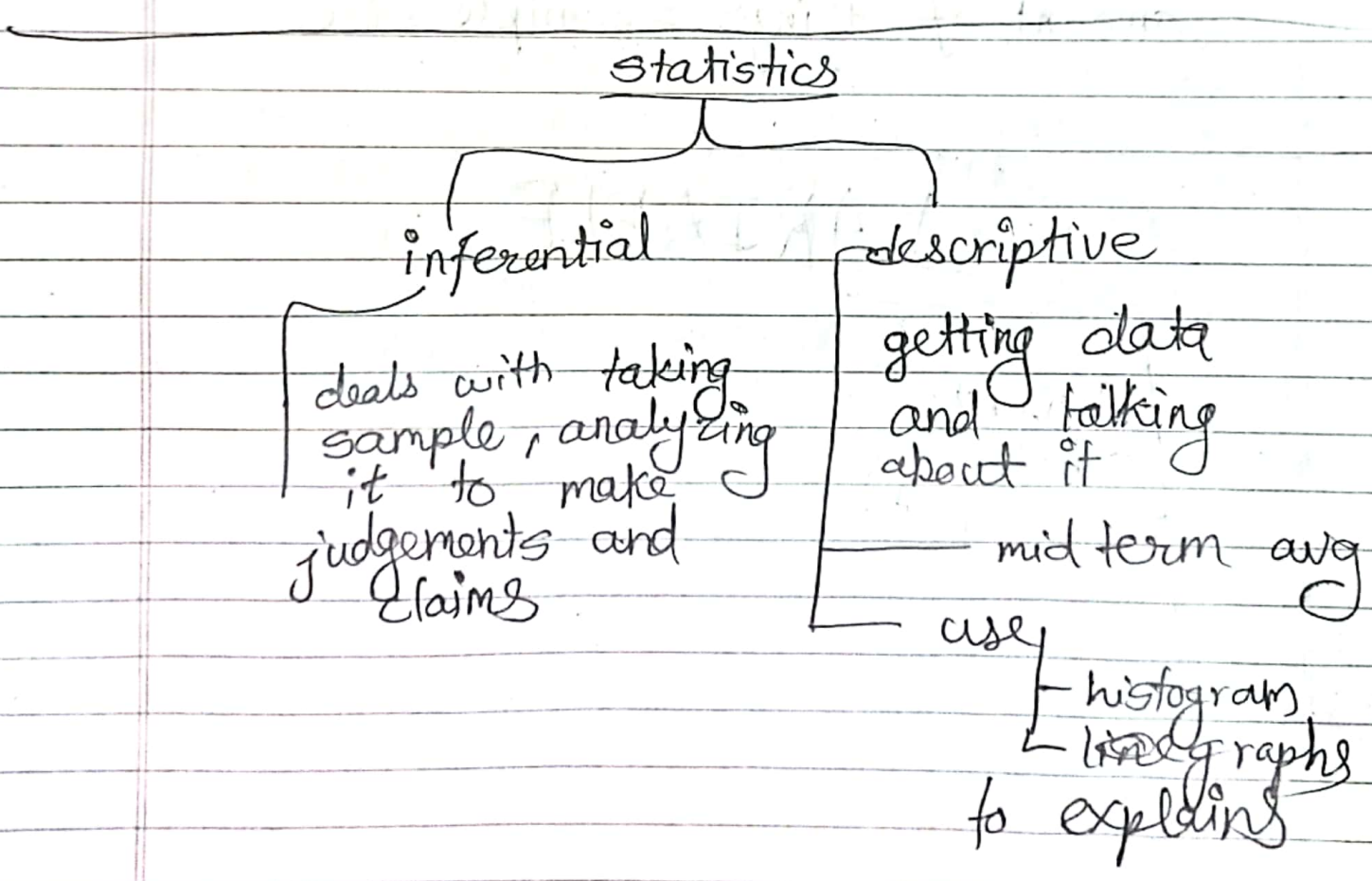
Statistics

The collectn and interpretin of data

(I) measure } variability
(II) analyze }

(I) height
(II) weight
(III) hair color
(IV) food preferences



Date
Page

Definition

Populatn

refers to total amount of things

Sample

refers to a small part of the populatn that is used for study

amount of things = sample size

VARIABLE

- height
- hair color
- weight

Measure variable

Quantitative data

data that is measured in numbers
OR
makes sense to perform arithmetic calculaths with

Quantitative variables

- height
- weight
- midterm score

Categorical data

refers to the value that place things into different groups or categories

Categorical variables

- hair color
- Type of cat
- Letter grade

Date
Page

Categorical variable

Categorical and Ordinal

if there is logical ordering to the values of a categorical variable

sorting garna sakinxa categorical value

Categorical and Nominal

no logical ordering to the values of a categorical variable.

example: skin color.

sorting garna but no logic in it just like alphabetical order

Quantitative variable

discrete and

refer to variables that can only be measured in certain numbers

example : no. of children, people, pets, animals

2.7 2, 3,

Continous

refers to variable that can take on any numerical value

~~example~~

weight

even 170.683

Variable

- what we measure from each people

Date
Page

- Bar charts } categorical data
- Pie charts
- histograms } Quantitative data
- timeplots
- way of displaying data

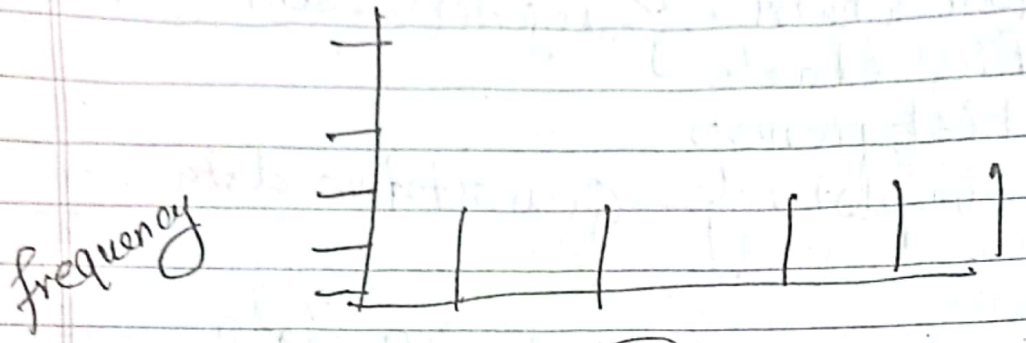
pie charts : relative size of each value in relation to whole

bar chart :- displays frequency and values of categorical variable.

Qual

Quant

- Stemplots
- histograms
- timeplots



histogram is in simple terms

- frequency distribution
  - It can be written into table format

By convention of statistics

each interval does not include the right end point

just like python for loop numbers

Date
Page

110 - 119
120 - 129 — no continuity

Instead,

110 - 120
120 - 130 — continuity

Reative frequency

Describing distribut^n using numbers

- mode
- median } measures of center
- mean
- range
- std deviat^n } measures of spread

mode: measures of centre
: refers to the value that is most frequently observed

median: middle of an ordered dataset

mean: measures of centre
: arithmetic average

median = physical middle point

range = max - min

standard deviation = how close to the value in a data set are to the mean

high std = high variability

Variance : related to std (square ka difference nuncha Std sanga)

The effects of transforming data on spread and centre

Outlier : a data value that is numerically distant from a data set

It can be smallest or largest

Atypical & surprising

The mean of is affected by the presence of outliers

Per capital income = total income / total populatn

Michael jordan - american basketball player

1984 north caroling

1980's Geography graduate highest avg salary starting