Pit-Stop Revision Handwritten Notes

Feel free to pass this around!

Stats + Mech 1

① Data Collection

Population = Set of items of interest

Census = Measures / observes every member of a population

Sample = Observations about a small group to find info about population

Types of Sampling

1. Simple Random Sampling:

- Every sample of size "n" has an equal chance of being selected.

2. Systematic Sampling:

- Required items are chosen at regular intervals from a list
  Sample size of 20 taken from 100, take every fifth member.

3. Stratisfied Sampling

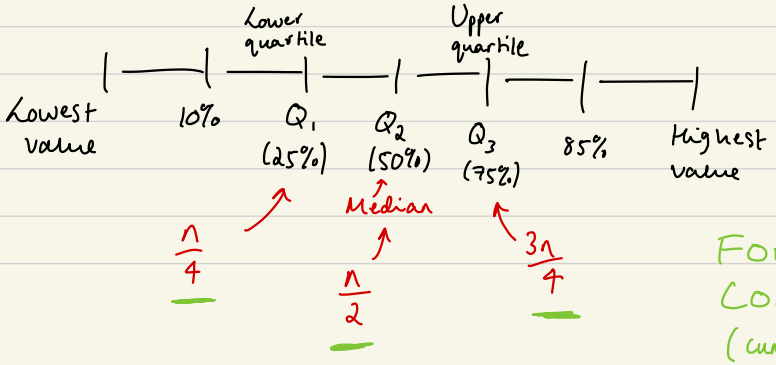
- Population is divided into mutually exclusive categories (Male + female), random sample taken from each.

(2) Measures of Location and Spread

Mode = Most common Median = Middle

Mean / Average: $\bar{x} = \frac{\Sigma x}{n}$ — $\bar{x}$: Mean; Σx: Sum of data values; n: no. of data values

Mean in frequency table: $\bar{x} = \frac{\Sigma fx}{\Sigma f}$ — Σfx: Sum of frequency x data value; Σf: Sum of frequency



* If $\frac{n}{4}$ or $\frac{3n}{4}$ is a whole number, go half up. If not whole, round up.

FOR GROUPED CONTINUOS Data! (cumulative frequency table)

Interpolation:

Length of time spent on internet (minutes)	30–31	32–33	34–36	37–39
Frequency	2	25	30	13

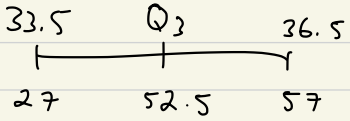
33.5 ↓ 36.5 ↓

70

Cumulative frequency | 2 | 27 | 57 | 70

↑ 52.5

For Q_3: $\frac{3 \times 70}{4} = 52.5$



$$\frac{Q_3 - 33.5}{36.5 - 33.5} = \frac{52.5 - 27}{57 - 27}$$

↓ Solve for Q_3

You can write the 10^{th} percentile as P_{10}

↳ (10% of 70)

For discrete data (distinct values) use $\frac{n}{4}$, $\frac{3n}{4}$ rules.

Range = Largest value − Lowest value.

Interquartile Range (IQR) = $Q_3 - Q_1$

Interpercentile Range = $90^{th} - 10^{th}$ percentile

↑ OR whichever percentile you're working with.

Variance (measure of spread) σ^2 = $\frac{\Sigma x^2}{n} - \frac{\Sigma x}{n} = \frac{\Sigma f x^2}{\Sigma f} - \frac{\Sigma f x}{\Sigma f}$

Standard Deviation σ = $\sqrt{\text{Variance}}$

$= \sqrt{\frac{\Sigma x^2}{n} - \frac{\Sigma x}{n}}$

$= \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \frac{\Sigma f x}{\Sigma f}}$

If working with grouped data for x, use the mid points.

E.g.

$2 \leq h \leq 3 \quad x = 2.5$

Coding : $y = \frac{x - a}{b}$

Mean of coded : $\bar{y} = \frac{\bar{x} - a}{b}$

Standard Deviation of coded : $\sigma_y = \frac{\sigma_x}{b}$

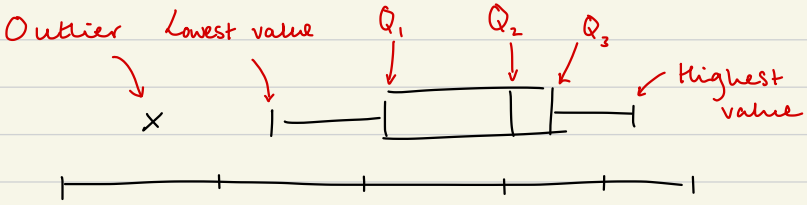
(3) Representations of Data

Anomaly = Error Outlier = Extreme result

Outliers : * $Q_3 + k(Q_3 - Q_1)$

* $Q_1 - k(Q_3 - Q_1)$

Removing anomalies = cleaning data



Cumulative Frequency = Add successive frequency values

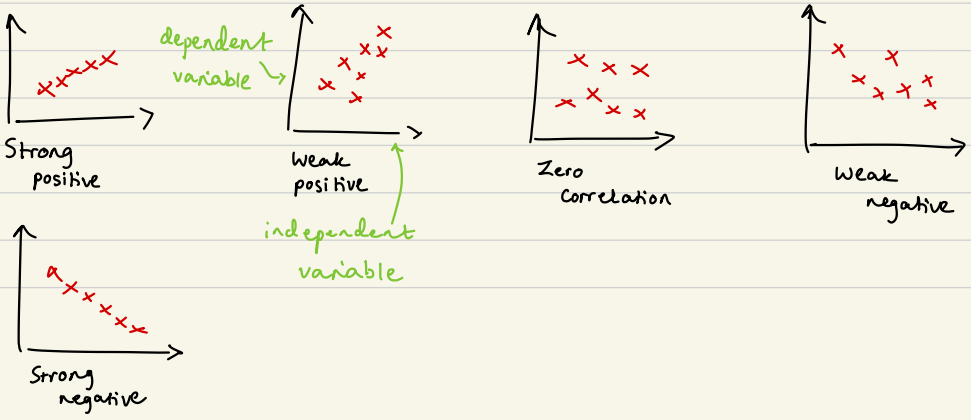
Frequency	Cumulative Frequency	
3	3	
4	7	(3+4)
2	9	(7+2)
9	18	(9+9)

Histograms: Frequency density = $\frac{\text{Frequency}}{\text{Class width}}$

Area of bar = Frequency

* Join midpoints of each bar to draw a frequency polygon.

④ Correlation



Regression Line : $y = a + bx$

(basically $y = mx + c$)

* The stronger the correlation, the more accurate the regression line.

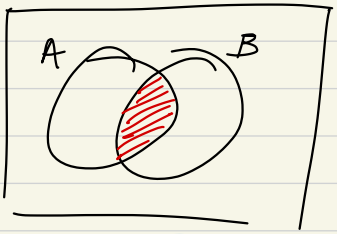
* Making a prediction outside the range of data is extrapolation

* If you know the value of independent variable, you can use a regression line to predict the dependant variable.

(5) Probability

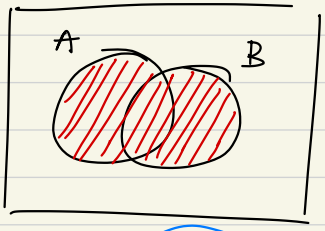
Sample Space: A set of all possible values.

Spinner 2 \ Spinner 1	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8



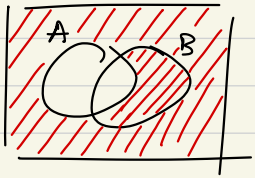
$P(A \cap B)$ Intersection

'A and B'



$P(A \cup B)$ Union

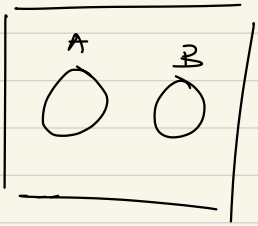
'A or B'



$P(A') = 1 - P(A)$

'Not A'

Mutually Exclusive:



$P(A \text{ or } B) = P(A) + P(B)$

Independent:
(One event has no effect on the other)

$P(A \text{ and } B) = P(A) \times P(B)$

(6) Statistical Distributions

Random Variable: Outcome depends on random event

Probability Distribution: Outlines probability of all outcomes in a sample space.

Discrete uniform distribution = All outcomes have the same probability.

Sum of probabilities add up to 1.

- $\sum P(X = x) = 1$

Random variable

Binomial Distribution

Model X as $X \sim B(n, p)$

n = no. of trials

p = probability of sucess

Probability of Binomial Distribution: $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$

n = index

p = parameter

$\frac{n!}{r!(n-r)!}$ or ${}^nC_r$

Binomial Cumulative Distribution (Binomial CD)

$P(X \leq x)$ → Use table in the data sheet OR Binomial C.D in calculator

Phrases

"Greater than 3" = $X > 3$ = $1 - P(X \leq 3)$

"No more than 3" = $X \leq 3$ = $P(X \leq 3)$

"At least 3" = $X \geq 3$ = $1 - P(X \leq 2)$

"Fewer than 3" = $X < 3$ = $P(X \leq 2)$

"At most 3" = $X \leq 3$ = $P(X \leq 3)$

(7) Hypothesis Testing

You can test a hypothesis about a population by carrying out an experiment or taking a sample of the population.

* Statistic calculated / result = Test statistic

You need 2 hypotheses:

* H_0 = Null hypothesis. Assume TRUE
  - H_0 : $p = \ldots$
* H_1 = Alternate hypothesis.
  - $H_1 : p < \ldots$ and $H_1 : p > \ldots$ (one-tailed test)
  - $H_1 : p \neq \ldots$ (two-tailed test)

Critical region: A region of the probability distribution, if the test statistic falls within it, you reject H_0.

Critical value = First value to fall within critical region

Significance Level: Threshold before you can reject H_0

Actual Significance Level: Probability of incorrectly rejecting the null hypothesis

One tailed test = One critical region on one end.

Two tailed test = Two critical regions on either end.

* One tailed test: Find these two

1. Express the distribution. $B(n, p)$

2. Identify H_0 and H_1. $H_0: p = \ldots$
$H_1: p < \ldots$ and $p > \ldots$

3. Calculate probability of test statistic: $P(X \leq \square) = \square$

4. Compare result to significance level (% as decimal)

$\square >$ Sig. lev. (5% / 0.05) $\square <$ sig. lev. (0.05)

5. Conclusion. Reject or Accept H_0.

* Two tailed test:

1. Express H_0 and H_1. $H_0: p = \ldots$
$H_1: p \neq \ldots$

2. Assume H_0 is true, write binomial distribution $B(n, p)$ (Think of it as does or doesn't occur)

3. What is your expected value? $= np$

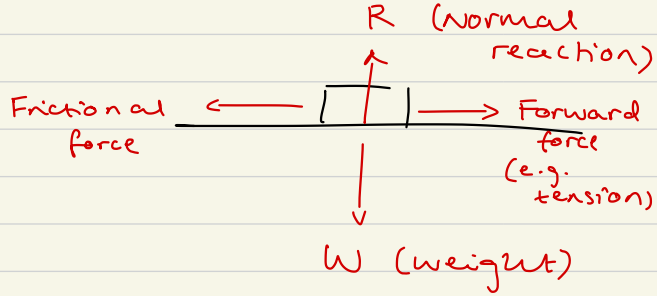
4. Probability of observed value?

5. Compare calculated probability to HALF the significance level.

6. Accept or Reject H_0.

(8) Modelling in Mechanics

Mass = kg
Distance = m
Time = s
Speed / Velocity = ms^{-1}
Acceleration = ms^{-2}
Force = N

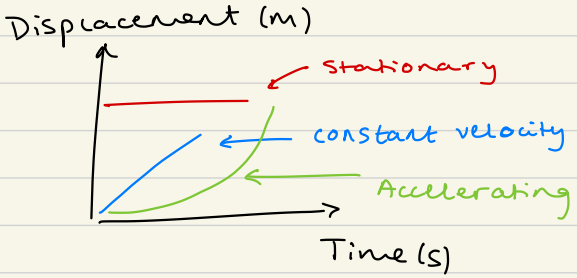


(9) Constant Acceleration

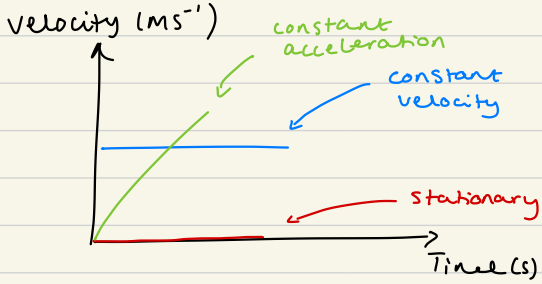
$v = u + at$ $\quad s = ut + \frac{1}{2}at^2$ $\quad s = vt - \frac{1}{2}at^2$

$s = \left(\frac{u+v}{2}\right)t$ $\quad v^2 = u^2 + 2as$

v = Final velocity (ms^{-1})
u = Initial velocity (ms^{-1})
a = acceleration (ms^{-2})
t = time (s)
s = displacement (m)



gradient = speed



gradient = acceleration

Area = displacement

* $g = ms^{-2}$

(10) Forces and Motion

Resultant Force = Sum of all forces.

$F = ma$

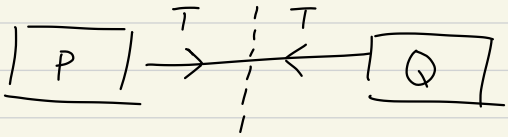
- F: Force (N)
- m: mass (kg)
- a: Acceleration (ms^{-2})

$W = mg$

- W: Weight (N)
- m: Mass (kg)
- g: gravity (ms^{-2})

$\longrightarrow\!\!>$ = Acceleration arrow

T = Tension, acts towards each other.



(11) Variable Acceleration

- $v = \frac{ds}{dt}$ —> Velocity is the rate of change of acceleration.

- $a = \frac{dv}{dt} = \frac{d^2s}{dt^2}$ —> Acceleration is the rate of change of velocity

- $s = \int v\,dt$ - $v = \int a\,dt$