Representation of Data

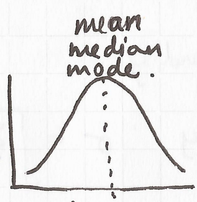
Types of Data
- Qualitative (descriptive)
- Quantitative (numeric)
  - Discrete (age)
  - Continuous (weight)

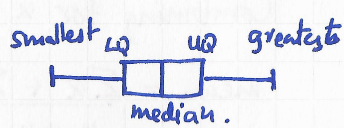
Histogram

- Continuous data
- Frequency density = $\dfrac{\text{Frequency}}{\text{Class width}}$
- Area = frequency
- Height = frequency density

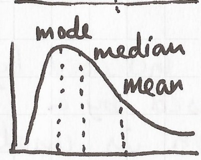
Skewness

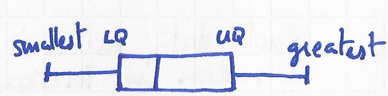
Symmetrical



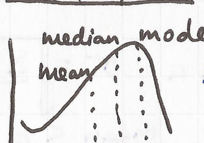


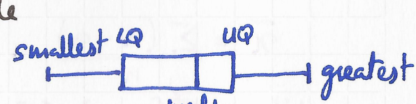
Positively Skewed





Negatively Skewed





Median

If n is odd, median $= \frac{n+1}{2}^{th}$ value

If n is even, median $= \frac{n}{2}^{th}$ value.

Mean

$$\bar{x} = \frac{\sum x}{n} \quad \text{or} \quad \bar{x} = \frac{\sum xf}{\sum f}$$

Standard Deviation

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \qquad \sigma = \sqrt{\frac{\sum (x-\bar{x})^2 f}{\sum f}}$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \qquad \sigma = \sqrt{\frac{\sum x^2 f}{\sum f} - \bar{x}^2}$$

Variance $= \sigma^2$

Combining for x and y

$$\text{mean} = \frac{\sum x + \sum y}{n_1 + n_2} \qquad \sigma = \sqrt{\frac{\sum x^2 + \sum y^2}{n_1 + n_2} - \text{mean}^2}$$

Each data value is increased by constant a

- mean is increased by a
- standard deviation is unaltered.

$$\bar{x} = \frac{\sum (x-a)}{n} + a \qquad \sigma = \sqrt{\frac{\sum (x-a)^2}{n} - \left(\frac{\sum (x-a)}{n}\right)^2}$$

When, $\sum (x-a) < 0$, mean $< a$

$\sum (x-a) = 0$ mean $= a$

$\sum (x-a) \geqslant 0$ mean $> a$

Permutations and Combinations

- No. of ways of arranging n unlike objects in a line is $n!$.
- Total arrangements for a word with repeated letters:

$$\frac{(\text{number of letters})!}{(\text{repeated letter})!}$$

- When 2 are to be together: consider them as one
- When 2 aren't to be together: total - impossible.

Permutations - order matters

n: items
r: spaces.

$$nP_r = \frac{n}{(n-r)!}$$

Combinations - order does not matter.

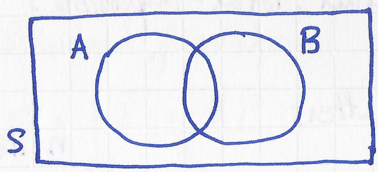
$$nC_r = \binom{n}{r} = \frac{n}{r!\,(n-r)!}$$

Probability

- $P(A) \Rightarrow$ Probability of event A
- $P(A') = 1 - P(A) \Rightarrow$ Probability of not A

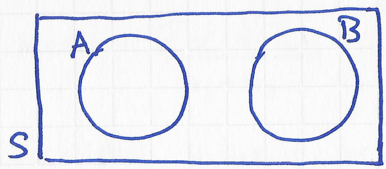
- $P(A) = 0 \Rightarrow$ impossible
- $P(A) = 1 \Rightarrow$ certain.
- $P(A) + P(A') = 1$

Combined Events



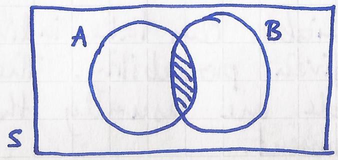
$A \cap B \Rightarrow$ A and B
$A \cup B \Rightarrow$ A or B or both
$P(A \cap B) \Rightarrow P(A)$ and $P(B) = P(A) \times P(B)$
$P(A \cup B) \Rightarrow P(A)$ or $P(B)$ or $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Mutually Exclusive Events



$P(A \cap B) = 0$
$P(A \cup B) = P(A) + P(B)$

Conditional Probability



$A \mid B \Rightarrow P(A)$ given that B has occured

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

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

If A and B are independant events,

$$P(A \mid B) = P(A)$$
$$P(B \mid A) = P(B)$$

So,

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

Discrete Random Variables

- A discrete random variable can take individual values each with a given probability. The values of the variable are usually the outcome of an experiment.

- Given that,

x	x_1	x_2	x_3	...	x_n
$P(X=x)$	p_1	p_2	p_3	...	p_n

$$P(X=x_1) + P(X=x_2) + \dots + P(X=x_n) = 1$$

$$\Rightarrow \sum_{\text{all } x} P(X=x) = 1$$

The Expectation of X (Mean)

$$E(X) = \mu$$

The expected value of a random variable or its mean:

$$E(X) = \sum xp \qquad E(X^2) = \sum x^2 p$$

x : value of X
p : probability of x

$$E(X \pm a) = E(X) \pm a$$
$$E(aX) = aE(X)$$

The Variance of X

It is a measure of the spread of X about the expected mean μ

$$Var(X) = \Sigma x^2 p - (\mu)^2$$

$$Var(X) = \Sigma (x-\mu)^2 p$$

Binomial Distribution

$X \sim B(n, p)$

$P(X=r) = \binom{n}{r}(p)^r (q)^{1-r}$

n : no. of trials
p : probability of success
q : probability of failure (1-p)

$E(X) = np$
(mean)

$Var(X) = npq$

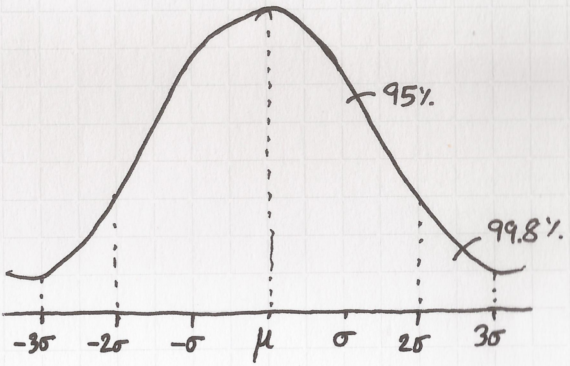
$\sigma = \sqrt{npq}$

Properties :
- Fixed no. of trials (n)
- 2 outcomes only
- outcomes of each trial is independant
- p is constant for each trial

Normal Distribution

X = a continuous random variable.

$X \sim N(\mu, \sigma^2)$



Standardising the normal variable X

$Z = \dfrac{X-\mu}{\sigma}$ $Z \sim N(0,1)$