Pop vs Sample

Pop	Sample
• N	n
• all items params	subset of pop statistics
• hard to gather data	• cheap & fast • needs random sample

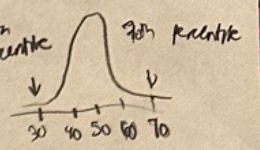
Types of Data

- Categorical
- Numerical
  - continuous (can be infinitely divided into smaller measurements)
  - discrete

bar plots vs hists to graphs

- categorical
- • quantitative
  • intervals as categories
  • frequency table in intervals

percentile



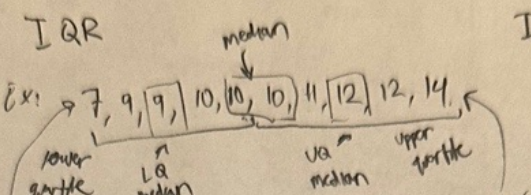
trim: top means <10% & >90% outliers removed

Mean - $\frac{\text{Sum of all values}}{\#\text{ of values}}$

Median - The middle number

Mode - most common number

IQR

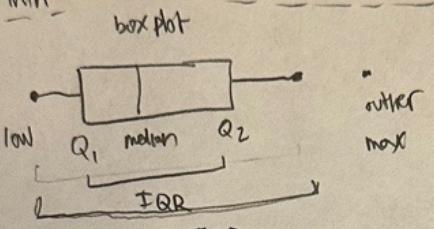


IQR = (UQ median − LQ median)

or 12 − 9 = 3

IQR = $Q_2 - Q_1$

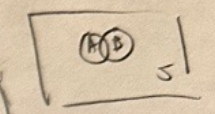
box plot



mild outlier: 1.5 · IQR

Extreme outlier: 3 · IQR

venn Diagram



Experiment - process w/ randomness

outcome - results of Experiment

ex: weather,

S - Sample space in {} braces

Event - Collection of outcomes

↳ A = {1, 2, 5, 1, 2}

Null set - no outcomes (∅)

3 operations:

complement (Aᶜ)	opposite
Intersection (A∩B)	A & B
Union (A∪B)	A or B

3 probability Axioms

1) non-negativity: for all events A, $P(A) \geq 0$
2) unitarity: $P(S) = 1$
3) sigma additivity: for disjoint events $A_1, A_2, A_3, \ldots$

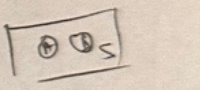
$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\sum_{i=1}^{\infty} A_i\right)$

$\bigcup_{i=1}^{n}$, $\bigcap_{i=1}^{n}$, $\sum_{i=1}^{n}$,

IF A & B disjoint, then they are mutually exclusive

$A \cap B = \emptyset$



1) $P(\emptyset) = 0$
2) $P(A^c) = 1 - P(A)$
3) for any event A, $P(A) \leq 1$
4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

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

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

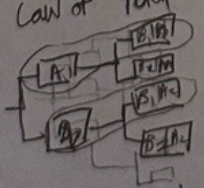
Conditional Probability

$P(A|B)$ ← probability A will happen given B

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

$P(A \cap B) = P(A)P(B|A)$ multiplication rule

Law of total probability



make a tree diagram, then add all probabilities that fulfill ex condition

$P(B) = P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)$

Independent Events

All 3 must be true

1) $P(A \cap B) = P(A)P(B)$
2) $P(A|B) = P(A)$
3) $P(B|A) = P(B)$

= A & B are independent iff $P(A \cap B) = P(A)P(B)$

Bayes Theorem

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

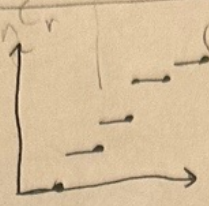
$$\Rightarrow P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)} \quad i = 1, \ldots, k$$

	order doesn't matter (combination)	order does matter (permutation)
No replacement	$\binom{n}{r} = \frac{n!}{r!(n-r)!} = {}_nC_r$ (we choose r things from n values)	$\frac{n!}{(n-k)!} = {}_nP_k$ (k = # of spots remaining)

Addition rule
- Two separate tasks $A \nmid B$
  A = m ways, B = n ways
  then $A \nmid B$ can be done $m+n$ ways

Product rule
- subsequent tasks $A \nmid B$
  A = m ways, B = n ways
  A then B can be done mn ways

Pmf - probability of a certain value (discrete only)
Cdf - Describes distribution of RV's (discrete & continuous)



$P(a \le X \le b) = F(b) - F(a-)$ ← largest X less than a
$P(a \le X \le b) = F(b) - F(a-1)$
$P(X \ge a) = 1 - P(X \le a-1)$
$p(x) = P(X = x)$ ⟹ pmf
$F(x) = P(X \le x) = \sum_{y: y \le x} p(y)$
↑ cdf

x	3	5	6	8	…
$F(x)$	0.1	0.2	0.1	0.5	…
$P(X=x)$	0.1	0.3	0.4	0.9	…

add $F(x)$ and prev value

x	4	7	9	11
$P(X=x)$	0.3	0.4	0.2	0.1
$F(x)$	0.3	0.7	0.9	1

use to find pmf

Linearity properties
$E(aX+b) = aE(X) + b$
$V(aX+b) = \sigma^2_{aX+b} = a^2 \cdot \sigma_X^2$ and $\sigma_{aX+b} = |a| \cdot \sigma_X$
$\sigma_{aX} = |a| \cdot \sigma_X$, $\sigma_{X+b} = \sigma_X$
$SD(X+b) = SD(X)$

Expected value & Variance of RV
$E(X) = \mu = \sum_{\text{all } x} x p(x)$, $E[g(x)] = \sum_{\text{all } x} g(x) p(x)$

Variance
$E[(X-\mu)^2] = \sum_{\text{all } x} (x-\mu)^2 p(x) = \sigma^2$
$E[(X-\mu)^2] = E(X^2) - [E(X)]^2$
$= V(X) = \sigma^2$

Binomial RV
$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$

In R: pbinom(q, size, prob)

$X \sim Bin(n,p) \Rightarrow B(x; n, p) = P(X \le x) = \sum_{y=0}^{x} b(y; n, p)$ $x = 0, 1, \dots, n$
$E(X) = np$, $V(X) = np(1-p)$
- Fixed trials, X = # of successes in n trials

Negative Binomial RV
$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$ $x = 0, 1, \dots$
$E(x) = \frac{r(1-p)}{p}$ $V(x) = \frac{r(1-p)}{p^2}$
- x = # of failures before r successes
- In R: pnbinom(q, size, prob)

Poisson RV
$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!}$ $x = 0, 1, \dots$ $\mu > 0$
$E(x) = V(x) = \mu$ ← avg # of events in a time frame (μ)
- No fixed # trials
- X = # of events in a given time