Multiplication r different experiments performed sequentially (e.g. expt 1, 2 has n_1, n_2 outcomes, … total = $\mathbf{n_1 n_2 \ldots n_r}$).
• Total number of ways to form 3-digit numbers from a few digits = Product of number of ways to pick each digit
• Number of ways to arrange people: If restriction, consider deterministic ordering. Else, add all possible orderings. E.g. MWMWMWM, WMWMWM = **2 × 3! × 3!** ***This AND That***

Addition Experiment performed by k different procedures. (eg procedure 1, 2 in n_1, n_2 ways…total = $\mathbf{n_1+n_2\ldots+n_r}$). Cases **do not overlap**. (#ways to pick even or odd numbers)

PERMUTATION

- Ordered Selection and arrangement of r objects out of n, given by $P_r^n = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-(r-1))$
- Note, if $r = n$, then $P_r^n = n!$, thus $n!$ is the number of ways to arrange n objects in order
- **Circular Permutation**: $(n-1)!$ if anticlockwise and clockwise order are distinct, and $\frac{(n-1)!}{2!}$ if both orders are the same
- Can be thought of: selecting r objects first without order, then permuting the r objects (P_r^r)
- **Permutation with sets of indistinguishable objects**: $(\frac{n}{n_1})(\frac{n-n_1}{n_2})(\frac{n-n_1-n_2}{n_3})\ldots = \frac{n!}{n_1!n_2!n_3!\ldots}$, where $n_1, n_2\ldots$ represents the different sets of indistinguishable objects (e.g. 4 red box, 3 red box, how many ways to permute if boxes of same colour indistinguishable)

• Find the **complement**: If 2 ppl must not be tgt, find #ways to arrange everyone − #ways they are grp tgt
• #4-letter words, all letters different = P_4^{26} n = 26 , r = 4

COMBINATION / CHOOSE

Unordered selection of r objects out of n, given by $C_r^n or (^n_r) = \frac{n!}{r!(n-r)!} = (^n_{n-r})$ for each combination, permute r objects

- $(^n_r) \times P_r^r = P_r^n$, so $(^n_r) = \frac{P_r^n}{P_r^r}$
- Use combination iff order is not important (no. of ways to pick 2 men and 1 woman from 4 men and 5 women)
- Note: **no. of choices do not always depend on the entire selection** (e.g. split set of 7 elements into 2 sets of 3 and 4, thus no. of choices = $(\frac{7}{3})$)
- e.g. Given box with 4 blue ball, 4 green ball and 2 red ball, P(2 blue, 1 green, 1 red) = $\frac{(^4_2)(^4_1)(^2_1)}{(^{10}_4)}$
- **r-Combinations with Repetitions allowed** (no. of ways r objects can be selected from n categories of objects with repetition): $(^{r+n-1}_{r})$

• #ways to choose sth (incl nothing): Line up in a row, & at every object, 2 choices (**pick/don't pick**) ⇝ total no. of choices = 2^k

	Order Matters	Order don't matter
No Repetition	n^k	$(^{k+n-1}_{k})$
Repetition	P_k^n	(^n_k)

PROBABILITY (chance / how likely an event will occur)

0≤P(A)≤1. P(A)=1 if A occurs every time. Sample space P(S)=1. P(A∪B)=P(A)+P(B) if mutually exclusive. P(∅)=0.
At least one of the two/**OR = Union** **AND = Intersection**
P(**only one** of the two) = P(A) + P(B) − **2 P(A∩B)**

- If S is a sample space, and A is an event, and $\exists x \in S$ and $x \notin A$, then $P(A) \not< 1$, as $P(x)$ can be 0
- For any 2 events A, B, $P(A) + P(B) \geq P(A \cup B)$
- For any 2 events A, B, $P(A) > P(A \cap B)$ is **NOT** always true
- For any 2 events A, B, $P(A) \leq P(B) \not\Rightarrow A \subset B$ (counterexample: $P(A) = P(B)$ but $A \neq B$, hence $A \not\subset B$
- Suppose $P(F) = P(G) = 0.4$, then $P(F \cup G) \leq 0.8$
- If $P(A)$ and $P(A \cap B)$ given, cannot deduce $P(B)$
- If $A \neq B$, then $P(A) \neq P(B)$ is **NOT** always true
- If A, B are independent, then $P(A \cup B) = 1 - (1 - P(A))(1 - P(B))$
- If $P(A) = 1 - P(B')$, then $P(A) = P(B) = 1 - P(B')$
- If $(A \cap B') \cup (A' \cap B) = \varnothing$, then $A = B$ (exists at the intersection only)
- If A, B are mut. excl, and $P(A) > 0$, $P(B) > 0$, then A, B are NOT indep
- **If CDF represents a cont. rv, then** $F(X = x) = 0, \forall x \in X$
- If CDF is step function, then the jumps in the function correspond to the actual $F(x)$ value at the x jump, then $E(X) = \sum$ of the product of the x jump and the $F(x)$ value

Finite Sample Space w/ Equally Likely Outcomes

P(1 girl out of 3 children) = $3 \times (\frac{1}{2})^3$ (Consider birth order) GBB BGB BBG BBB GGB GBG BGG GGG

- Assume that all outcomes in the sample space are equally as likely ($P(a_1) = P(a_2)\ldots$)
- For any event $A \subset S$, $P(A)$ = no. of sample pt in A / no. of sample pt in S > check for indep of people first (adjust max according and what exactly is the unique factor (bday mth, day,)
- Claim that in a room of ≥ 50 people, there is a chance that at least 2 ppl have the same birthday: consider event A = ≥ 2 ppl same bday, then A' = everyone same, $P(A') = \frac{365 * 364 * \ldots}{365^n}$ and $P(A) = 1 - P(A')$; as $n \to \infty, P(A) \to 1$
- Consider the inverse (how large of a grp to be selected such that P(some one bday = my bday) = 0.5: P(all different bdays) = $(\frac{364}{365})^n$, so we need $1 - (\frac{364}{365})^n \geq 0.5$, and we can get $n \geq 252.7$
- Why the difference? first: any day can be the shared birthday, second: only one day can be the shared birthday

CONDITIONAL PROBABILITY

- For any 2 event A, B, $P(A) > 0$, conditional probability of B given A occurred: $P(B|A) = \frac{P(A \cap B)}{P(A)}$ (if A never occurs, then cond prob is 0)
- **Reduced Sample Space**: under the knowledge that A occurs, A is our new reduced sample space
- **Multiplication Rule**: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$, if $P(A) \neq 0$ and $P(B) \neq 0$ respectively
- **Inverse Probability Formula**: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$
- $P(A'|B) = 1 - P(A|B)$ $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$

P(X) = P(X1) P(X2|X1) P(X3|X2∩X1) P(X4|X1∩X2∩X3)

P(only 1 red) $P((A_1 \cap A_2') \cup (A_1' \cap A_2)) = P(A_1 \cap A_2') + P(A_1' \cap A_2)$

The equality above holds since $(A_1 \cap A_2') \cap (A_1' \cap A_2) = \varnothing$.

$P(A_1 \cap A_2') = P(A_1)P(A_2'|A_1)$ $P(A_1' \cap A_2) = P(A_1')P(A_2|A_1')$

MUTUALLY EXCLUSIVE / DISJOINT: A ∩ B = ∅

A ∩ B = ∅, no element in common. **P (A∩B) = ∅**. If so, it implies that if A and B are **independent**, then **P(A) = 0** or **P(B) = 0**.
•P(A)>0, P(B)>0, **P(A∩B)=∅** ≠ P(A)P(B) ⇝NOT INDEPENDENT

INDEPENDENCE: P(A ∩ B) = P(A) P(B) A ⊥ B

•If **P(A)=1**, then A and B are **independent**.
• If **P(A) ≠ 0**, **A ⊥ B** iff **P(B|A)=P(A)**, knowing A does not change prob of B. If so, implies that if **P(A), P(B) > 0**, then A and B are **not mutually exclusive**, and A ⊥ B ⇝ P(A∩B)=P(A)P(B) >0 **≠ 0**.
• If A ⊥ B ⇝ Any combination of A, B, A', B' is independent, A' ⊥ B.
• S and ∅ are independent of any other event:
P(A∩S) = P(A) = P(A)P(S) **P(A∩∅) = P(∅) = 0 = P(A)P(∅)**

INDEPENDENCE ≠ MUTUALLY EXCLUSIVE

Probability, results may be same Set theory, **results different**

LAW OF TOTAL PROBABILITY

- **Partition**: $A_1, A_2 \ldots A_n$ mutually exclusive, and $\bigcup_{i=1}^n A_i = S$, then $A_1, A_2 \ldots A_n$ forms a partition of S
- **Law of Total Probability**: Suppose $A_1, A_2 \ldots A_n$ is a partition of S, then for any event B, $P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$
  - Can think of it as the sum of the intersections between each partition and the event, within the sample space (which is partitioned by $A_1, A_2 \ldots$ P(A∩B) P(A'∩B)
- **Special case**: $P(B) = P(A)P(B|A) + P(A')P(B|A')$
- Monty Hall: P(win) = P(car is behind door)P(win | car is behind door) + P(car not behind door)P(win | car not behind door), and $P(W) = \frac{1}{3}P(A)P(W|A) + \frac{2}{3}P(A')P(W|A')$, by switching, we set $P(W|A) = 0$ and $P(W|A') = 1$ and we get $P(W) = \frac{2}{3} > \frac{1}{3}$ as $P(W|A) = 1$ and $P(W|A') = 0$ original C' | C | C' switch opened

Bayes' Theorem

- $A_1, A_2 \ldots A_n$ forms a partition of S, then for any event B, $k \in [1, n]$, $P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(B \cap A_i)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_k)}$ 1. Identify event of interest 2. Identify partition
- Special case: $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$
- NOTE: **The partition to consider is always the first term in the cond. prob (e.g. consider A's partition when finding A | B**

PROBABILITY DISTRIBUTION / FUNCTIONS

- **Discrete**: no. of values in R_X is finite or countable
  - **Probability Mass function (PMF) / Probability Function**

$$f(x) = \begin{cases} P(X = x), x \in R_x \\ 0, x \notin R_x (improbable) \end{cases}$$

x	0	1	2
$f(x)$	1/2	1/6	1/3

  - Collection of pairs $(x_i, f(x_i)), i = 1, 2, 3\ldots$ is the **probability distribution of X**

Properties of PMF but not necessarily drv eg $f(x) = [\frac{1}{b-a}, 0]$

- Not necessarily increasing
- $f(x_i) \geq 0, \forall x_i \in R_X \Rightarrow$ No. of possible values of f(x) is countable/finite
- $f(x) = 0, \forall x_i \notin R_X$
- $\sum_{i=1}^n f(x_i) = \sum_{x_i \in R_x} = 1$, i.e. sum of all pmf values must be 1
- $\forall B \subset \mathbb{R}, P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i)$, i.e. for any event B, prob of drv X ∈ B (x is some outcome in B) is the sum of the PMF of values of $x \in B \cap R_X$ (P(X ∈ B) = $f(x_1) + f(x_2) + \ldots$
- **Continuous**: no. of values in R_X is an interval or collection of intervals
  - **Probability Function / Probability Density Function (PDF)**

Properties of PDF

- Not necessarily increasing ★ f(x) is a pdf if (1) and (2) are true. Check these conditions!
- ★ $f(x) \geq 0, \forall x \in R_x$ (1) ✓
- ★ $f(x) = 0, \forall x \notin R_x$ (1) ✓
- $f(x)$ **may take on values more than 1**
- ★ $\int_{R_x} f(x)dx = 1$, equivalent to $\int_{-\infty}^{\infty} f(x)dx = 1$, since $f(x) = 0, \forall x \notin R_X$ (2) ✓ ★ Integrate along the bounds
- $\forall a, b$ such that $a \leq b$, $P(a \leq X \leq b) = \int_a^b f(x)dx$
- $\forall x_0$, $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$, hence **P(A) = 0** but **A ≠ ∅** No area at one single point
- $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$, graphically all are represented by the area under the graph $f(x)$ Since point mass at endpoints a and b are 0

-- CUMULATIVE DISTRIBUTION FUNCTION (defined ∀ℝ) --

- For any rv X, $F(x) = P(X \leq x)$ (applies to drv and crv)
- $P(X = x) = P(X \leq x) - P(X < x)$ $\lim_{x \to a^+} F(x) = F(a)$: approaching from the left, the limit should be finite (i.e. $F(x) = P(X \leq x)$)
- Always **right-continuous**: approaching from the left, the limit should be finite (i.e. $F(x) = P(X \leq x)$)
- **Should be always non-decreasing**, $x_1 < x_2, F(x_1) \leq F(x_2)$
- cdf ⇏ pdf and pmf ⇏ pdf (one-to-one correspondence, bijective)
- Ranges for both $F(x)$ and $f(x)$
  - $0 \leq F(x) \leq 1$ FOR BOTH
  - discrete: $0 \leq f(x) \leq 1$ just need to make sure integration result is 1
  - continuous: $f(x) \geq 0$ but not always the case that $f(x) \leq 1$ (e.g. $f(x) = 2, 0 < x \leq 0.5$)

Convergence to 0 and 1: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

- Given $F(x)$ for RV X, $F(X) > 1 - P(X \geq x)$, $F(x) = 1 - P(X \geq x)$ and $F(x) = P(X = x)$ are POSSIBLE, but $F(x) < 1 - P(X \geq x)$, since $F(x) = P(X \leq x)$ and $1 - P(X \geq x) = P(X < x)$
- $F(x) \geq 1 - P(X \geq x)$

Discrete

P(a<X≤b) = P(X≤b) − P(X≤a) = F(b)−F(a)
P(a<X<b) = P(X≤b) − P(X≤a) = F(b−)−F(a)

- $F(x) \geq f(x)$
- $F(x) = \sum_{t \in R_X, t \leq x} f(t) = \sum_{t \in R_X, t \leq x} P(X = t)$ (meaning F(x) = P(X ≤ x) = f(a) + f(b) + … + f(x − 1) ★ Derive cdf
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a-)$ (largest value of R_X smaller than a, $F(a-) = lim_{x \to a-}$
- e.g. $P(2 \leq x \leq 5) = P(X \leq 5) - P(X < 2) = F(5) - F(2-) = F(5) - F(1)$, since the values are discrete
- e.g. $F(1.5) = F(1)$ if there are no values between the numbers such that $x \in R_X$, otherwise that equality is wrong
- For the CDF graph, the gaps between the different intervals represent the $F(x)$ value at the x value x
- Let $F(x)$ be cdf of drv, then $\exists x_1, x_2 \in \mathbb{R}$ such that $F(x_1) = F(x_2) > 0$ Graph looks like a step function

Continuous ⇝ Integration

- $F(x)$ **must be differentiable at all points**
- $F(x) = \int_{-\infty}^{x} f(t)dt$ (area under PDF) cdf ⇝ pdf no need ≠ now value is fine P(X=a)=0 P(a≤X≤b)=F(a)−F(b)
- $f(x) = \frac{d}{dx}F(x)$ (gradient of $F(x)$ is pdf at x) pdf
- $P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$

EXPECTATION

Mean/Expectation of DRV: Average value of rv X, defined by $\mu_X = E(X) = \sum_{x_i \in R_X} x_i f(x_i)$ (e.g. $E(X) = X_1 P(X = X_1) + \ldots$)

Mean/Expectation of CRV: Average value of rv X, defined by $\mu_X = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{x \in R_X} xf(x)dx$

(INTEGRATE ALONG THE BOUNDS ALWAYS)

$E(X)$ may not always be possible value of rv X

Properties of Expectation

- $E(E(X)) = E(X)$ In general, $E(\frac{1}{g(x)}) \neq \frac{1}{E(g(x))}$
- $E(aX) = aE(X)$, $E(aX + b) = aE(X) + b$, for discrete ($\sum_{x \in R_X} (ax + b)f(x)$) and discrete ($\int_{-\infty}^{\infty} (ax + b)f(x)dx$)
- $E(a_1X_1 + \ldots + a_kX_k) = a_1E(X_1) + \ldots + a_kE(X_k)$
- $E(X(X - 1)) = E(X^2 - X) = E(X^2) - E(X)$
- $E(X + Y) = E(X) + E(Y)$
- If $g(\cdot)$ is a function, then $E(e^{k}+b) = \sum (e^{x}+b)P(X=x)$
  - X is discrete with pmf $f(x)$: $E[g(X)] = \sum_{x \in R_X} g(x)f(x)$
  - X is continuous with pdf $f(x)$: $E[g(X)] = \int_{R_X} g(x)f(x)dx$
  - good if LHS is hard to compute (maybe g() not known)
  - Product extends to **all** $\mathbf{x} \in R_X$
- There **ARE** cases where $E(X)$ does not exist (think integration of harmonic series $\frac{1}{x}$, e.g. $\int x \times \frac{1}{x}$)
- For a **continuous** rv, $P(X = x) = 0 \forall x \in \mathbb{R}$, so doesnt mean that $X \subset \mathbb{R}$ and thus $P(X = x) = 0 \forall x \in \mathbb{R}$ is impossible
- $E(E(X)) = E(X)$, $E(c) = c$
- Let X be a discrete rv, then $|E(X)| < \infty \Rightarrow$ FALSE, let $P(X = n) = \frac{6}{\pi^2}\frac{1}{n^2}, n \in \mathbb{Z}^+$, then $E(X) = \sum_{n=1}^{\infty} \frac{6}{\pi^2}\frac{1}{n^2}n = \sum_{n=1}^{\infty} \frac{6}{\pi^2}\frac{1}{n} = \infty$ (harmonic series, note that if $\frac{1}{n^2}$, then convergent)
- **If R_X is finite, then $|E(X)|$ is finite**: If X is discrete, then $x \in R_X$, $|R_X| < \infty$ and $|E(X)| < \infty$. Now, $|E(X)| < \sum_x xf(x) = \sum |max(\{R_X\})| \cdot f(x) \leq \sum |max(\{R_X\})| = |R_X| \cdot max(\{R_X\})$ which is finite
- $Y = X^2 + 2$, then $E(Y) = E(X^2) + 2$ and $V(Y) = E(Y^2) - [E(Y)]^2$

If cannot compute E(X), compute ***E(X + a -a)*** *[both equiv]*

- Let X be a rv, then $V(X) \geq |E(X)| \Rightarrow$ FALSE, V and E has no relationship; ↓ V, the more narrow the distribution
- Let X be a cont. rv, then $\forall x \in R_X, |x| \leq 100 \to |E(X)| < \infty \Rightarrow$ TRUE, $E(X) = |\int_{x \in R_X} xf(x)dx| \leq \int_{x \in R_X} |x|f(x)dx \leq \int_{x \in R_X} |100|f(x) \leq |100|$, as $\int_{x \in R_X} f(x) = 1$
- Let X be a discrete rv, then $\forall x \in R_X, |x| \leq 100 \to |E(X)| < \infty \Rightarrow$ TRUE, $E(X) = |\sum_{x \in R_X} xf(x)dx| \leq \sum_{x \in R_X} |x|f(x)dx \leq \sum_{x \in R_X} |100|f(x) \leq |100|$, as $\sum_{x \in R_X} f(x) = 1$

VARIANCE σ^2

Variance • **Standard Deviation** of X: $\sigma_X = \sqrt{V(X)}$

- If $g(x) = (x - \mu_X)^2$, then $E[g(X)]$ is the variance
- Definition: $\sigma_X^2 = V(X) = E(X - \mu_X)^2$, applicable to both discrete or cont.
- **Discrete with pmf** $f(x)$: $V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$
- **Continuous with pdf** $f(x)$: $V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)$
- $V(X) \geq 0$ for any X, and is only equal if X is a constant ($P(X = E(X)) = 1$) $V(X^2+2) = E[(x^2+2)^2] - [E(X^2+2)]^2$
- $V(aX + b) = a^2V(X)$
- $V(X) = E(X^2) - [E(X)]^2 \geq 0$, $E(X^2) \geq [E(X)]^2$

DISCRETE

DISCRETE UNIFORM DISTRIBUTION

- If X is a DRV with values $x_1, x_2, \ldots$, each with **equal** probability, Discrete uniform distribution

$$f_X(x) = \begin{cases} \frac{1}{k}, x = x_1, x_2, \ldots \\ 0, \text{otherwise} \end{cases} \quad (3)$$

Probability mass function

- **Expectation** $\mu_X = E(X) = \sum_{i=1}^{k} x_i f_X(x_i) = \frac{1}{k}\sum_{i=1}^{k} x_i$
- **Variance** $\sigma_X^2 = V(X) = E(X^2) - E(X)^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu_X^2$

BERNOULLI RV X ~ Bernoulli(p)

If X is no. of successes in **Bernoulli trial** (random experiment), then X only has **2 possible outcomes (success / failure)**:

$$f_X(x) = P(X = x) = \begin{cases} p, x = 1 \\ 1 - p, x = 0 \end{cases} = p^x(1-p)^{1-x} \text{ for } x = 0, 1 \quad (4)$$

success failure

Define $q = 1 - p$ failure success Bernoulli Process

$f_X(0) = q$, $f_X(1) = p$ sequence of repeatedly performed independent and identical (idd) Bernoulli trials, Bernoulli process generates sequence of idd Bernoulli RV $(X_1, X_2 \ldots)$

- **Expectation** $\mu_X = E(X) = p$
- **Variance** $\sigma_X^2 = V(X) = p(1-p) = pq$

BINOMIAL RV X ~ Bin(n, p)

- Counts no. of successes in n trials of a Bernoulli process, where
  - Probability of success for each trial is p
  - Trials are **independent** X~Bin(n=1,p=p) = X~Bernoulli(p)
- PDF is: success failures

$$P(X = x) = (^n_x)p^x(1-p)^{n-x}, \text{for } x = 0, 1, 2 \ldots, n \quad (5)$$

- **Expectation** $E(X) = np = n \cdot E(X_{bernoulli})$
- **Variance** $V(X) = np(1-p) = npq = n \cdot V(X_{bernoulli})$
- Given a $X \sim Bin(n, p)$, can be expressed as $X = X_1 + X_2 + \ldots + X_n$, X can be expressed as: $X = X_1, X_2, \ldots, X_n$ (X_i are **idd** Bernoulli(p) RV) [this is bidirectional r/s]
- Follows that $X + Y$ is independent iff X, Y are independent (p must be the same)

For $i = 1, \ldots, 4$, let $B_i \sim$ Bin(100, 0.9) be the number of copies of the ith bit amongst the 100 copies sent that are not flipped. Then for a correctable message to be received, we need $B_1 > 50$, $B_2 > 50$, $B_3 > 50$ and $B_4 > 50$. $P(B_1 > 50, B_2 > 50, B_3 > 50, B_4 > 50)$

Independence $= P(B_1 > 50)P(B_2 > 50)P(B_3 > 50)P(B_4 > 50)$

For each i, $P(B_i) > 50$ $\left[\sum_{k=51}^{100} (^{100}_{k}) 0.9^k 0.1^{100-k}\right]^4 > 0.999$.

- If $X \sim Bin(n, p)$, $n > 1$, then $\forall n_1, n_2 \in \mathbb{Z} : n_1 + n_2 = n$, then $\exists X_1 \sim Bin(n_1, p), X_2 \sim Bin(n_2, p)$ such that $X = X_1 + X_2$
- $X \sim Bin(n, p)$, $Y \sim Bin(m, p)$ then $X + Y \sim Bin(n + m, p)$ Only true if X_1, X_2 / X, Y are independent. Check VARIANCE

NEGATIVE BINOMIAL RV X ~ NB(k, p)

No. of trials needed for k successes **to occur**, X is no. of idd Bernoulli(p) trials needed for k successes to occur P(X > x) = P(A ∩ B) = P(A) P(B)

$$f_X(x) = P(X = x) = (^{x-1}_{k-1})p^k(1-p)^{x-k}, \text{for } x = k, k+1, k+2 \ldots \quad (6)$$

success failure

Expectation $E(X) = \frac{k}{p}$

Variance $V(X) = \frac{(1-p)k}{p^2}$ $Y \sim Bin(x-1, p)$: $(^{x-1}_{k-1})p^k(1-p)^{x-k}$

Event $\{X = x\}$ = {x trials needed to get k success} = {k-1 successes in first x-1 trials ∩ {success on x^{th} trial} = $A \cap B$

- $P(A)$ given by binom prob of k-1 successes in first x-1 trials
- $P(B)$ given by berm prob of success on k^{th} trial = p
- Since both events are independent, multiplying both together

GEOMETRIC RV X ~ Geom(p) = X ~ NB(k=1, p=p)

Special case of NB, represents number of idd Bernoulli(p) trials needed **until first success occurs** $X \sim 1$

$$f_X(x) = P(X = x) = (1-p)^{x-1}p \quad (7)$$

failures success

Expectation $E(X) = \frac{1}{p}$

Variance $V(X) = \frac{(1-p)}{p^2}$ $P(X \geq k) = (1-p)^{k-1}$

$1 + r + r^2 + \ldots = \frac{1-r^n}{1-r}$ eg $\sum pq^{n-1}$, q=1-p

GP formula may be useful when computing CDF of GRV (e.g $\sum pq^{n-1}$, $q = 1 - p$)

POISSON RV X ~ Poisson(λ)

X represents the number of events occurring in a fixed period of time or fixed region (temporal, units of measurement), $\lambda > 0$

$$f_X(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k > 0, \in \mathbb{Z} \quad (8)$$

expected number of occurrences during given period/region

k=0,1,… = number of occurrence

Expectation $E(X) = \lambda$

Variance $V(X) = \lambda$

$P(X \leq k) = \sum_0^{k_0} \frac{e^{-\lambda}\lambda^k}{k!}$ Scale interval ⇒ $X_1 \sim Poisson(bk)$ for interval bT ↑ by b>0

If $X_1 \sim Poisson(k)$ for some interval T, when scaling interval

POISSON PROCESS X ~ Poisson(αT)

Continuous time process, counting no. of occurrences within some interval • No simultaneous occurrences

α: rate parameter • No. of occurrences in disjoint intervals are independent

Poisson dist. can be added iff both dist. are independent with no overlaps

In general, $X, Y \sim Poisson(\lambda_i) \to X + Y \not\sim Poisson(\lambda_x + \lambda_y)$

$X \sim Poisson(\lambda)$, then $\forall \lambda_1, \lambda_2 \in \mathbb{R}^+ : \lambda = \lambda_1 + \lambda_2$, then $\exists$ a RV $X_i \sim Poisson(\lambda_i), X = X_1 + X_2$

- True
- Given an interval, we can always partition it into 2 independent, non-overlapping intervals (refer to pt 1)

Given independent $X_1 \sim Poisson(2)$, $X_2 \sim Poisson(3)$, then $E(X_1|X_1 + X_2 = 10) = \frac{P(X_1 = x, X_1 + X_2 = 10)}{P(X_1 + X_2 = 10)} = \frac{P(X_1 = x)P(X_2 = 10 - x)}{P(X_1 + X_2 = 10)}$

Poisson Approximation to Binomial

- $X \sim Bin(n, p)$, $n \to \inf, p \to 0$, such that $\lambda = np$ is constant, then $X \sim Poisson(np)$

$\lim_{n \to \inf, p \to 0} P(X = x) = \frac{e^{-np}(np)^x}{x!}$

- n must be sufficiently large
- p is close to 0
- n >>> p
- np < n(1-p)

Use when n is too large and binomial

n≥20, p≤0.05 or n≥100, np≤10

- **When approximating and asked for E(X) or V(X), always use the non-approximated distribution values**

CONTINUOUS

CONTINUOUS UNIFORM DISTRIBUTION X ~ U(a, b)

Uniform prob across interval (a, b), **PDF** given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, a \leq x \leq b \\ 0, \text{otherwise} \end{cases} \quad (10)$$

eg time

$P(X \geq a) = \int_a^b f_X(x)\,dx$ P(X≥10|X≥5) ≠ P(X≥10)

Expectation $E(X) = \frac{a+b}{2} = \int_a^b x \cdot \frac{1}{b-a}dx$

Variance $V(X) = \frac{(b-a)^2}{12} = E(X^2) - E(X)^2 = \int_a^b x^2 \cdot \frac{1}{b-a}dx - (\int_a^b x \cdot \frac{1}{b-a}dx)^2$

CDF is given by

$$F_X(x) = \begin{cases} 0, x < a; \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, x > b \end{cases} = \int_{-\infty}^{x} f_X(t)dt = \int_a^x \frac{1}{b-a}dt \quad (11)$$

EXPONENTIAL DISTRIBUTION X ~ Exp(λ) = Exp(1/μ)

Continuous counterpart to Geom, model waiting time to **first success in continuous time**

PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases} = \begin{cases} \frac{1}{\mu}e^{-\frac{x}{\mu}}, x \geq 0 \\ 0, x < 0 \end{cases} \quad (12)$$

$\mu = \frac{1}{\lambda}$ ↑ mean ↑ "expected" ⇒ change μ to ↑

Expectation $E(X) = \int_0^\infty x\lambda e^{-\lambda x}dx = \frac{1}{\lambda} = \mu$

Variance $V(X) = \int_0^\infty x^2\lambda e^{-\lambda x}dx = \frac{1}{\lambda^2} = \mu^2$

CDF is given by

CDF

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases} = \int_0^x \lambda e^{-\lambda t}dt = 1 - e^{-\frac{x}{\mu}} \quad (13)$$

P(X≤x)

Noting that $P(a \leq X \leq b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$ $a, b > 0$

$P(X > x) = e^{-\lambda x}, x > 0$ ★ P(X≤x) = 1−P(X>x)

MEMORYLESS property:

- If $X \sim Exp(\lambda)$, $\lambda > 0$, then $\forall s, t \in \mathbb{R}+$, **P(X > s+t | X > s) = P(X > t)**
- prob of time t from time 0 → s equals prob from time 0 → t
- PDF is strictly decreasing
- λ→∞, left side of fn concentrates, right side flatter & →0 faster
- X~Exp(2), X~Exp(4), then **Z=X+Y ~ Exp(6)** as **E(X+Y) = E(Z)**
- If X1 … Xn are idd Exp(λ) variables, then if X=min{X1…Xn}, X~Exp(**n**λ) (multiply Exp() distributions tgt, then consider pdf)

GC SHORTCUTS

- $P(X = x)$: binompdf() BINOMIAL
- $P(X \leq x)$: binomcdf() [its ≤ not <]
- $P(X = x)$: geometpdf() GEOMETRIC
- $P(X \leq x)$: geometcdf() [its ≤ not <]
- $P(X = x)$: poissonpdf() POISSON
- $P(X \leq x)$: poissoncdf() [its ≤ not <] P(X>x)=1−P(X≤x)
- $P(X = x)$: normalpdf() NORMAL and STANDARD NORMAL
- $P(X \leq x)$: normalcdf() [its ≤ not <]

NORMAL DISTRIBUTION X ~ N(μ, σ²)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{(2\sigma^2)}}$$ Range: $-\infty < x < \infty$

Expectation E(X) = μ **Variance** V(X) = σ²

Area under curve always 1 ($f_X(x)$ is pdf). If **same σ, then same shape (with μ shifted)**. If σ **increases, then curve flattens.**

X~N(1,σ₁²), Y~N(0,σ₂²), σ₁²>σ₂²>0 ⇒ P(X<1)=0.5 < **P(Y<1)>0.5**

Probabilities of N distr **only dependent on mean μ, NOT σ.**

- Let X~N(0, 8²). As 2k+1 is always odd, the #times X is multiplied by itself is odd for X^{2k+1}. Also **P(−X ≤ x) = P(X ≥ −x) = P(X ≤ x)**. Thus, by symmetry, −X~N(0, 8²) too. Then **E(−X^{2k+1})= −E(X^{2k+1}) = E(X^{2k+1})** which is only possible if E(X)=0
- X ~ N(μ_1, σ_1^2) , Y ~ N(μ_2, σ_2^2) ⇝ **X + Y ~ N($\mu_1 + \mu_2$, $\sigma_1^2 + \sigma_2^2$)**
- $X_1 \ldots X_{100}$ are iid normal variables with E(X_i) = 5, V(X_i) = 4 ⇒ **$X = X_1 + \ldots X_{100} \sim$N(100 E($X_i$) , 100 V($X_i$))** by CLT

STANDARD NORMAL DISTRIBUTION X ~ N(0, 1)

PDF $\phi = f_Z(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ $Z = \frac{X-\mu}{\sigma}$ **E(X)=0 V(X) =1**

- $P(X_1 < X < X_2) \Leftrightarrow P(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}) = \Phi(\frac{x_2-\mu}{\sigma}) -$ $\Phi(\frac{x_2-\mu}{\sigma})$ where **CDF** $\Phi(\cdot) = \int_{-\infty}^{z} \phi(t)dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{t^2}{2}}dt$
- Symmetrical around 0, **P(Z ≥ 0) = P(Z ≤ 0) = Φ(0) = 0.5**
- $\Phi(z) = P(Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z)$!!!
- $P(-a \leq Z \leq a) = \Phi(a) - P(z \leq -a) = 2\,\Phi(a) - 1$!!!
- **−Z~N(0, 1)** $\sigma Z + \mu \sim N(\mu, \sigma^2)$ (linear transformation)

QUANTILE α^{th} (upper) quantile, $0 \leq \alpha \leq 1$, is the number x_α satisfying **$P(X \geq x_\alpha) = \alpha$**. **$P(Z \geq z_\alpha) = P(Z \leq -z_\alpha) = \alpha$**

$P(|Z| \leq z) = 1 - P(|Z| > z) = 1 - P(Z > z) - P(Z < z) = 1 - 2P(Z > z)$

NORMAL APPROXIMATION TO BINOMIAL

• When n→∞, p remains constant and near 0. Use when **np > 5** & **n(1-p) > 5**. If X~*Bin*(n, p), $Z = \frac{X-E(X)}{\sqrt{V(X)}} = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$

• Do **CONTINUITY CORRECTION BEFORE NORMALISATION**:

LHS:	$\leq \Rightarrow a - 0.5$	$< \Rightarrow a + 0.5$
RHS:	$< \Rightarrow a - 0.5$	$\leq \Rightarrow a + 0.5$

$P(X = k) \approx P(k - 0.5 < X < k + 0.5)$

$P(a < X < b) \approx P(a + 0.5 < X < b - 0.5)$

$P(X \leq c) = P(0 \leq X \leq c) \approx P(-0.5 < X < c + 0.5)$

$P(X > c) = P(c < X \leq n) \approx P(c + 0.5 < X < n + 0.5)$

SIMPLE RANDOM SAMPLE Every subset of n observations (members) of the population has same prob of being selected.
• #possible sample sizes of n from N population = *N choose n*.

INFINITE POPULATION Equals to sampling from finite population with replacement. Prob within a distribution varies, but does not vary across samples X1…Xn of same distribution.
• **Joint Probability Function**: $f_{X1\ldots Xn}(x_1\ldots x_n) = f_{X1}\bullet(x_1)\ldots f_{Xn}\bullet(x_n)$

STATISTICS

Statistic [**RANDOM VARIABLE**] = function of $(X_1\ldots X_n)$ [random sample of n observations]. Replace with observed ⇒ Realisation.

SAMPLE MEAN $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **[random variable]**

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad E(\bar{X}) = E(X)$

SAMPLE VARIANCE $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ **[random var]**

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad E(S^2) = \sigma^2$

MEAN AND VARIANCE OF $\bar{X}$

Sampling distribution of the sample mean $\bar{X}$ from a population with mean μ_X and variance σ_X^2

Note, $\sigma_{\bar{X}}$ is the variance for the sampling distribution for the sample mean, not of the population

$\mu_{\bar{X}} = E(\bar{X}) = \mu_X$ **[valid estimator for population mean]**

$\sigma_{\bar{X}}^2 = V(\bar{X}) = \frac{\sigma_X^2}{n}$

For an inf population, $n \to \infty$, variance becomes smaller and smaller, accuracy of $\bar{X}$ as an estimator improves

STANDARD ERROR Describes spread of sampling distribution. Is $\sigma_{\bar{X}}$, the standard deviation of sampling distribution of $\bar{X}$.

Standard error of $\bar{X}$ shows how much $\bar{x}$ differs among samples.

CENTRAL LIMIT THEOREM

Law of large numbers If $X_1\ldots X_n$ are iid RV with same μ and σ^2, ⇒ As $n \to \infty$, $P(|\bar{X} - \mu| > \epsilon \to 0, \epsilon \in \mathbb{R}$ i.e. $\bar{X} \to \mu$ i

• When $P(|\bar{X} - \mu| > \frac{\sigma}{\sqrt{n}}) > c$ then $P(Z > 1) > c$

CLT: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \to Z \sim N(0,1)$ (Approximate for large **n ≥ 30**, non-normal sample. Exact for normal sample)

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\lim_{n\to\infty} P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq x) = \Phi(x)$

$P(a \leq \bar{X} \leq b) = P(\frac{a-\mu}{\sigma/\sqrt{n}} \leq \bar{X} \leq \frac{b-\mu}{\sigma/\sqrt{n}})$

χ^2(n) distribution, *n* degrees of freedom

Z^2 is a χ^2 distribution with 1 deg of freedom. If $Z_1\ldots Z_n$ are iid standard normal RV, then $Z_1^2+\ldots+Z_n^2 \sim \chi^2(n)$. All χ^2 density functions have long right tail, determined by degree of freedom.

Expectation: *n* **Variance: *2n*** ***n large*** ⇒ $\chi^2 \approx N(n, 2n)$

• Both $Y1\sim\chi^2(n)$, $Y2\sim\chi^2(n)$ independent ⇒ **$Y1+Y2 \sim \chi^2(n+m)$**

• $\chi^2(\alpha, n)$ s.t. $Y \sim \chi^2(n)$, $P(Y > \chi^2(\alpha, n)) = \alpha$

If $X_i \sim N(\mu, \sigma^2)\ \forall i$, use $\frac{(n-1)S^2}{\sigma^2} \sim \chi2(n-1)$ for *n* idd RV with

E(X) = μ, Var(X) = σ^2, sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

• If S^2 is variance of a random sample size n taken from a normal population w variance σ^2: $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi2(n-1)$

t distribution, t(*n*), with *n* degrees of freedom

Independent Z~N(0,1) and **U~ χ^2(n)** ⇒ $T = \frac{Z}{\sqrt{U/n}} \sim t(n)$

• $\frac{U}{n} = \frac{Z_1^2+\ldots+Z_n^2}{n}$ and $E(Z_1^2)=1$ when n→∞ ⇒ **T~N(0,1) if n ≥ 30**

• **Expectation E(T)=0** **Variance V(T) =** $\frac{n}{n-2}$, $n > 2$

• Right tail probability: $P(T > t_{n;\alpha}) = \alpha$ [same shape as N]

• $X_1\ldots X_n$ idd **normal** RV with mean *μ*, *var* σ^2 ⇒ $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim$ **t(n-1)**

F distribution, F(*m, n*), with (*m, n*) degrees of freedom

Independent U~ $\chi^2(m)$ and **V~ $\chi^2(n)$** ⇒ $F = \frac{U/m}{V/n} \sim F(m,n)$

• $E(X) = \frac{n}{n-2}$, $n > 2$ $\quad Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $n > 4$

• $F = \frac{U/(n_1-1)}{V/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

• $F \sim (m,n) \Rightarrow \frac{1}{F} \sim (n,m)$ $\quad F(m,n;1-\alpha) = \frac{1}{F(n,m;\alpha)}$

• Upper tail probability of $F(m,n;\alpha)$ *s.t.* $P(F > F(m,n;\alpha)) = \alpha$

UNBIASED ESTIMATOR

$E(\hat{\theta}) = \theta$ where $\hat{\theta}$ is an estimator of parameter θ, is a RV based on sample, and has mean value equal to true value of parameter.

Confidence Intervals / Test Statistics: Population Mean: • 100(1 – α)% confidence interval formulas for population mean μ, • test statistics for the (null) hypothesis: $H_0: \mu = \mu_0$

!! Remember to SQUARE ROOT VARIANCE σ^2 or s^2 TO GET σ or s !!

	Population	σ	n	Test Statistic	E	n for desired E , α	Confidence Interval
I	**Normal**	Known	Any	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathbf{N(0,1)}$	$z_{\alpha/2}\cdot(\sigma/\sqrt{n})$	$\left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$	$\bar{x} \pm z_{\alpha/2}\cdot(\sigma/\sqrt{n})$
II	**Any**	Known	Large				
III	**Normal**	Unknown	Small	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim \mathbf{t_{n-1}}$ ***n-1** degrees of freedom*	$t_{n-1;\,\alpha/2}\cdot(\frac{s}{\sqrt{n}})$	$\left(\frac{t_{n-1;\,\alpha/2}\cdot s}{E_0}\right)^2$	$\bar{x} \pm t_{n-1;\,\alpha/2}\cdot(s/\sqrt{n})$ ***n-1** degrees of freedom*
IV	**Any**	Unknown	Large n≥30	$Z = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim \mathbf{N(0,1)}$	$z_{\alpha/2}\cdot(s/\sqrt{n})$	$\left(\frac{z_{\alpha/2}\cdot s}{E_0}\right)^2$	$\bar{x} \pm z_{\alpha/2}\cdot(s/\sqrt{n})$

HYPOTHESIS TESTING & REJECTION OF H_0

Can use Z-Test in GC **Can use T-Test in GC**

Given $H_0: \mu = \mu_0$		**p-value** (note: *z* = computed test statistic)	**Rejection region (Reject H_0) (NOTE α, or α/2)**	
Two-sided	$H_1: \mu \neq \mu_0$	$P(\|Z\| > \|z\|) = 2P(Z > \|z\|) = 2P(Z < -\|z\|)$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$t < -t_{n-1,\,\alpha/2}$ or $t > t_{n-1,\,\alpha/2}$
One-sided	$H_1: \mu < \mu_0$	$P(Z < -\|z\|)$ (only left tail)	$z < -z_{\alpha}$ **(α, not α/2 !)**	$t < -t_{n-1,\,\alpha}$ **(α, not α/2 !)** ***n-1** degrees of freedom*
	$H_1: \mu > \mu_0$	$P(Z > \|z\|)$ (only right tail)	$z > z_{\alpha}$ **(α, not α/2 !)**	$t < -t_{n-1,\,\alpha}$ **(α, not α/2 !)**

p-value & **significance level α**	**or CONFIDENCE INTERVAL**	**⇒ Reject H_0 ?**
p-value < α	**Confidence interval does NOT contain μ_0**	**⇒ Reject H_0**
p-value ≥ α	**Confidence interval contains μ_0**	**⇒ DO NOT reject H_0**

INDEPENDENT 2-samples Test

(Samples and individuals within each sample are independent. Complete randomisation.)

Population	Variances σ_1^2, σ_2^2	n_1, n_2	Test Statistic	100(1-α)% Confidence Interval
Any	**Known** (use σ) , **Unequal**	$n_1 \geq 30$, $n_2 \geq 30$	$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
Normal	$V(\bar{X}) = \sigma_1^2/n_1$ $V(\bar{Y}) = \sigma_2^2/n_2$	Any	$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$	
Any	**Unknown** (use *s*) , **Unequal** $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2$ $S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$	$n_1 \geq 30$ $n_2 \geq 30$	$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$ $Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0,1)$	$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
Normal	**Unknown , Equal** Check **equal variance assumption:** $\frac{1}{2} \leq \frac{S_1}{S_2} \leq 2$ $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ **!! Square root S_p^2 to get S_p !!**	$n_1 < 30$ $n_2 < 30$	$T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}} \sim t_{n_1+n_2-2}$ $T = \frac{(\bar{X}-\bar{Y})}{S_p\sqrt{1/n_1+1/n_2}} \sim t_{n_1+n_2-2}$	$(\bar{x}-\bar{y}) \pm t_{n_1+n_2-2;\,\alpha/2}\, s_p\sqrt{1/n_1+1/n_2}$ $(\bar{x}-\bar{y})$ is the point estimate for $\mu_1 - \mu_2$
Any		$n_1 \geq 30$ $n_2 \geq 30$	$Z = \frac{(\bar{X}-\bar{Y})}{S_p\sqrt{1/n_1+1/n_2}} \sim N(0,1)$	$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\, s_p\sqrt{1/n_1+1/n_2}$

• **Tests COMPARING MEANS : INDEPENDENT samples - Testing $H_0: \mu_1 - \mu_2 = \delta_0$** **Can use 2-SampZTest in GC**

Population	Variances $\sigma_1^2\sigma_2^2$	n_1, n_2	TEST STATISTIC
Normal	**KNOWN**	**≥ 30**	$Z = \frac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$
	UNKNOWN	**≥ 30**	$Z = \frac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{S_1^2/n_1+S_2^2/n_2}} \sim N(0,1)$
Normal	**UNKNOWN & EQUAL**	**< 30**	$Z = \frac{(\bar{X}-\bar{Y})-\delta_0}{S_p\sqrt{1/n_1+1/n_2}} \sim t_{n_1+n_2-2}$

*NOTE: S_1^2 → replaced by **SAMPLE** variances, NOT population variance*

Rejection regions (NOTE α, or α/2) and p-values:

H_1	Rejection region	p-value
$\mu_1 - \mu_2 > \delta_0$	$z > z_{\alpha}$	$P(Z > \|z\|)$
$\mu_1 - \mu_2 < \delta_0$	$z < z_{\alpha}$	$P(Z < -\|z\|)$
$\mu_1 - \mu_2 \neq \delta_0$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$2\,P(Z > \|z\|)$

• Average is different → $H_1: \mu_1 - \mu_2 \neq 0$

• Mean is higher for A than B → $H_1: \mu_A - \mu_B > 0$

MATCHED PAIRS Samples ($n_1 = n_2 = n$) $\mu_D = \mu_1 - \mu_2$

(observations are dependent WITHIN the pair, independent (randomized) between pairs.) *Eg before and after*

Population	Mean and Variance	n	Test Statistic		100(1-α)% Confidence Interval for μ_D
Normal	$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$ $S_D^2 = \frac{\sum_{i=1}^{n}(D_i-\bar{D})^2}{n-1}$	n < 30	$W = \frac{\bar{D}-\mu_D}{S_D/\sqrt{n}}$	**W ~ t_{n-1}**	$\bar{d} \pm t_{n-1;\,\alpha/2}\cdot(s_D/\sqrt{n})$
		n ≥30		**W ~ N(0, 1)**	$\bar{d} \pm z_{\alpha/2}\cdot(s_D/\sqrt{n})$

*** Remember to SQUARE ROOT S_D^2 to get S_D**

• **Tests COMPARING MEANS : PAIRED data - Testing $H_0: \mu_D = \mu_{D_0}$** → Use **Test statistic** $W = \frac{\bar{D}-\mu_{D_0}}{S_D/\sqrt{n}}$

MAXIMUM ERROR OF ESTIMATE

$E = z_{\alpha/2}\cdot(\sigma/\sqrt{n})$

Minimum sample size so that with probability 1 – α, the error is at most E_0 : $n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$

(1 – α) CI ⇒ $z_{\alpha/2}$: distr → invNorm(1 – α/2, …)

α/2 or α

90% CI, α = 0.10: $z_{\alpha/2} = z_{0.05} = 1.645$ *invNorm(0.95…)*

95% CI, α = 0.05: $z_{\alpha/2} = z_{0.025} = 1.96$ *invNorm(0.975…)*

96% CI, α = 0.04: $z_{\alpha/2} = z_{0.02} = 2.05$ *invNorm(0.98…)*

98% CI, α = 0.02: $z_{\alpha/2} = z_{0.01} = 2.326$ *invNorm(0.99…)*

99% CI, α = 0.01: $z_{\alpha/2} = z_{0.005} = 2.576$ *invNorm(0.995…)*

0.995 quantile of standard normal = $z_{0.005} = 2.576$

(1 – α) CI ⇒ $t_{n-1;\,\alpha/2}$: **α/2** distr → invT(1 – α/2, degree of freedom)

90% CI: invT(0.95, n) for $t_{n-1;\,0.05}$ ** 0.95=(1- 0.10/2)

95% CI, α = 0.05 ⇒ invT(0.975, n-1) , $t_{10;\,0.025} = 2.228$

98% CI, α = 0.02 ⇒ invT(0.99, n-1) , $t_{9;\,0.01} = 2.821$

99% CI, α = 0.01 ⇒ invT(0.995, n-1) , $t_{4;\,0.005} = 4.604$

$t_{n-1;\,\alpha}$: distr → invT(1 – α, degree of freedom) α

95% CI, α = 0.05 ⇒ invT(0.95, n-1) , $t_{7;\,0.05} = 1.895$

CONFIDENCE INTERVAL

• Interval estimator: Rule for calculating, from the sample, an interval (a,b) where you are fairly certain the parameter lies in.

• **Degree of confidence / Confidence Level** = 1 – α

• $P(a < \mu < b) = 1-\alpha$ $\quad (a, b)$ is the (1-α) confidence interval

a, b are RV, until the interval is constructed.

• CI at the 1–α level of confidence is the (1–α) confidence that a population parameter is found between the interval. Also the % of samples that will have population parameter within interval.

• **$(1 - \alpha)$ CI = $\bar{X} \pm E$** where E = maximum error of estimate

⇒ Since confidence interval $(a, b) = \bar{X} \pm E$,

Reject $H_0 \Leftrightarrow \mu_0 \notin (a, b)$ **DO NOT reject $H_0 \Leftrightarrow \mu_0 \in (a, b)$**

Confidence intervals can be used to perform **two-sided tests.**

• $\bar{X} \pm E$ has probability (1 – α) of containing μ

POOLED ESTIMATOR S_p^2

• For **small** samples with **equal variance**

• **Assume equal variance** when $\frac{1}{2} \leq \frac{S_1}{S_2} \leq 2$

• If σ is unknown ⇒ Estimate it. Under the equal variance assumption, S_1^2 and S_2^2 are unbiased estimators of σ^2.

• POOLED ESTIMATOR, $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

PAIRED DATA (X_i, Y_i)

• Matched pairs (X_i, Y_i), where $X_1 \ldots X_n$ is a random sample from population 1 and $Y_1 \ldots Y_n$ is a random sample from population 2.

• X_i and Y_i are dependent. (X_i, Y_i) and (X_j, Y_j) are independent for any i ≠ j.

• $D = X_i - Y_i$ and $\mu_D = \mu_1 - \mu_2$. Treat $D_1 \ldots D_n$ as a random sample from a population with mean μ_D and variance σ_D^2.

HYPOTHESIS TESTING

1. Set null and alternative hypothesis. Assume H_0 unless there is overwhelming evidence against it. Alternative hypothesis H_1 is the hypothesis we want to prove.

2. Set significance level α = P(Type I error) = P(Reject H_0 | H_0 is true)

Given	Do not reject H_0	Reject H_0
H_0 is true	Correct	**Type I error** [*serious*]
H_0 is false	**Type II error**	Correct

• β = P(Type II error) = P(Do not reject **H_0** | **H_0** is false)

• Power of the test = 1 – β = P(Reject **H_0** | **H_0** is false)

• To decrease both errors *simultaneously*, reduce sample size.

3. Identify test statistics (⇒ How unlikely it is to observe the sample, assuming H_0 is true), **distribution, and rejection criteria.**

• α ⇒ Divide possible values of test statistics into **REJECTION REGION (CRITICAL REGION)** and acceptance region.

• E.g. $N\sim(\mu, 2^2)$, $X_1\ldots X_n$ RVs from that distribution, if $H_0: \mu = 0$ vs $H_1: \mu \neq 100$, and reject when $|z_{obs}| > 2$ where $Z = \bar{X}/(\sigma/\sqrt{n})$. Suppose we actually reject H_0, what is the probability that we fail to reject H_0 given μ=2 ⇒ **Type II error**.

⇒ P(|Z| ≤ 2 | μ=2) = P(– 2 ≤ $\frac{\bar{X}}{\sigma/\sqrt{n}}$ ≤2 | μ=2). Standardise by subtracting 2/(2-5) from all terms (represents μ=2 part)

⇒ P(– 7 ≤ $\frac{\bar{X}-2}{\sigma/\sqrt{n}}$ ≤ -3 | μ=2) = P(– 7 ≤ $\hat{Z}$ ≤ -3 | μ=2)

4. Compute observed test statistic. (Refer to first table)

5. Reject or do not reject the null hypothesis.

P-VALUE / OBSERVED LEVEL OF SIGNIFICANCE

Probability of obtaining a test statistic at least as extreme as (≤ or ≥) than the observed sample value, **given H_0 is true**.

• **p-value < α ⇒ reject H_0** (test statistic in rejection region)

• **p-value ≥ α ⇒ DO NOT reject H_0**

• If asked for $\Phi(c)$ values, and if **RIGHT** tail ($\mu > \mu_0$) was tested, **negate the z or t value** from the test (NOT the α value)