Topic 3
Discrete & Continuous Probability Distributions

Tutorial agenda:

1. tutorial questions + go through key concepts
2. how to use calculator
3. try excel

Key concept

T4 L3

Recap Variables

Random Variable

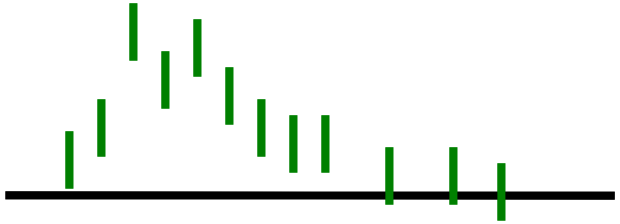
outcomes of an experiment with probabilistic occurrence

Topic 3 T4

Discrete Random Variable

a **counting process**

(e.g. number of courses you are taking in this semester)

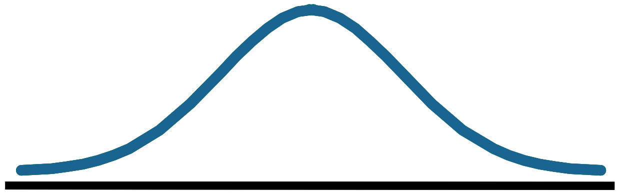


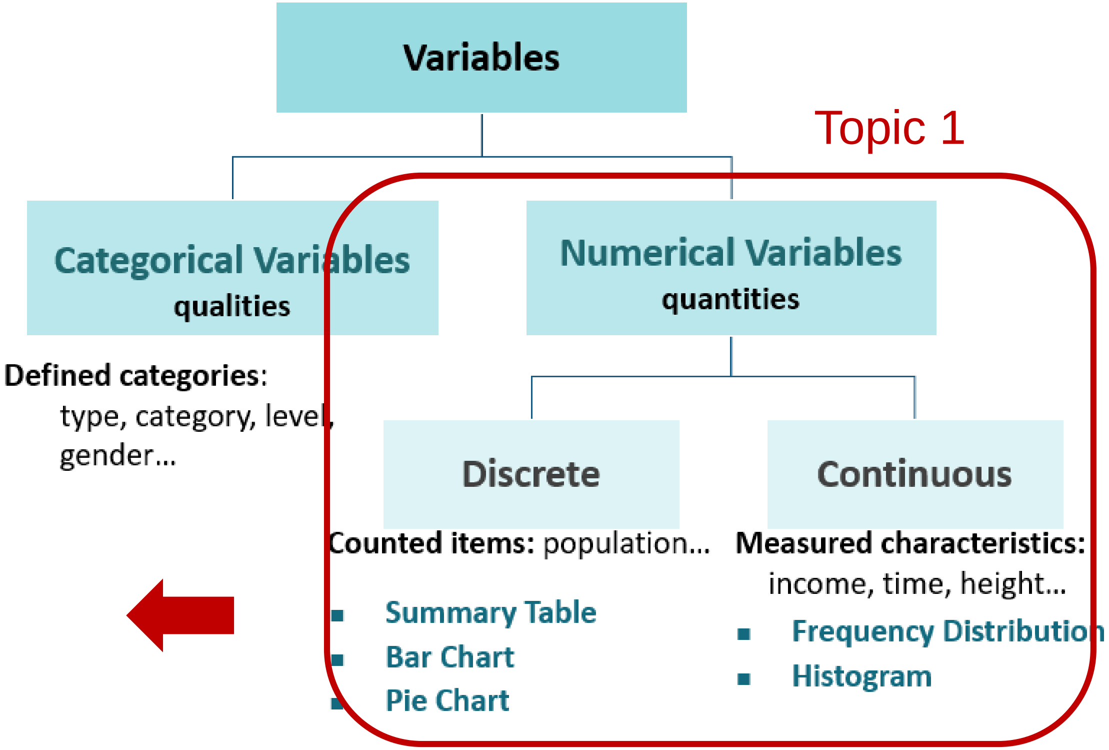
Topic 3 T5

Continuous Random Variable

a **measurement**

(e.g. your annual salary, or your weight)





Variables

Topic 1

Categorical Variables
qualities

Defined categories:
type, category, level, gender…

Numerical Variables
quantities

Discrete

Counted items: population…

- Summary Table
- Bar Chart
- Pie Chart

Continuous

Measured characteristics:
income, time, height…

- Frequency Distributio
- Histogram

Probability Distribution

- Discrete Probability Distribution
- Binomial Distribution

- Continuous Probability Distribution
- Normal Distribution

Discrete Probability Distributions

- Probability distribution
  - Mutually exclusive
    - = nothing in common = independent
  - Collectively exhaustive
    - = nothing left out = sum is 1 = all possible outcomes
  - Source: Priori knowledge / empirical approach

- Sumamry measures
  - Center - Expected value/mean

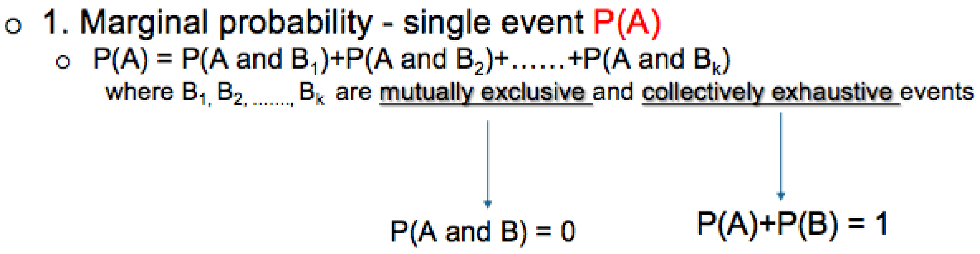
$$\mu = E(X) = \sum_{i=1}^{N} x_i P(X = x_i)$$

  - Variation
    - Variance
    - Standard deviation (s.d.)

$$\sigma^2 = \sum_{i=1}^{N} [x_i - E(X)]^2 P(X = x_i)$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [x_i - E(X)]^2 P(P = x_i)}$$

Topic 2



- 1. Marginal probability - single event P(A)
  - P(A) = P(A and B$_1$)+P(A and B$_2$)+……+P(A and B$_k$)
    where B$_1$, B$_2$, ……, B$_k$ are mutually exclusive and collectively exhaustive events

P(A and B) = 0

P(A)+P(B) = 1

Topic 1

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left(X_i - \mu\right)^2}{N} \qquad S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}$$

Binomial Distribution

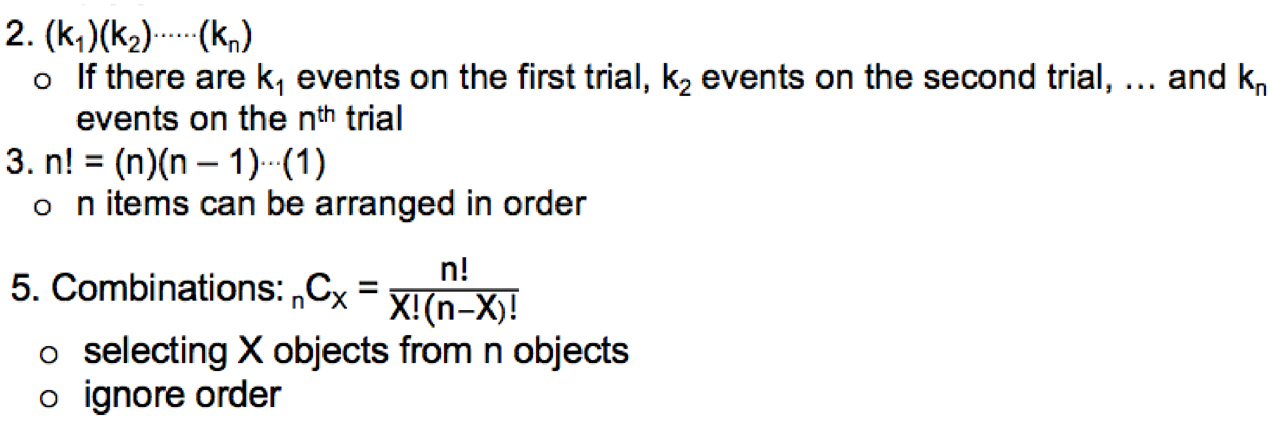
- Bi = 2
- Binomial = 2 outcomes
  - a special case of Discrete Distribution
  - Mutually exclusive
  - Collectively exhaustive

- What is the probability that x out of n obs meet the criterion?

$$p(X=x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

  - X - event of interest/meet the criterion
  - x - no. of event of interest
  - n - no. of obs
  - p - probability of an event of interest

Topic 2



2. $(k_1)(k_2)\cdots\cdots(k_n)$
   - If there are k_1 events on the first trial, k_2 events on the second trial, … and k_n events on the n^{th} trial
3. $n! = (n)(n-1)\cdots(1)$
   - n items can be arranged in order

5. Combinations: ${}_nC_X = \frac{n!}{X!(n-X)!}$
   - selecting X objects from n objects
   - ignore order

Binomial Distribution

- Summary measures
  - Center - Expected value/mean

$$\mu = np$$

  - Variation
    - Variance
    - Standard deviation (s.d.)

$$\sigma^2 = np(1-p)$$

$$\sigma = \sqrt{np(1-p)}$$

  - Shape
    - p

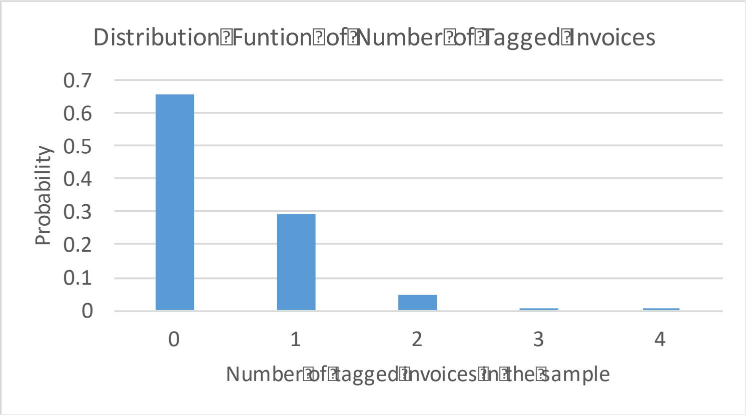
$p = 0.5$ => symmetrical

$p < 0.5$ => right-skewed

$p > 0.5$ => left-skewed

- p <=> concentration

n = 4, p = 0.1



Distribution and Summary measures

- Steps of calculating measures (center, variation)
  - 1. calculate the probability of each x out of n
    - $P(X=x_i)$
  - 2. construct a distribution table
    - x_i and $P(X=x_i)$
  - 3. calculate measures

$$p(X=x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

x_i	$P(X=x_i)$
0	0.59049
1	0.32805
2	0.0729
3	0.0081
4	0.00045
5	0.00001

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(X = x_i)$$

$$\sigma^2 = \sum_{i=1}^{N} [x_i - E(X)]^2 P(X = x_i)$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [x_i - E(X)]^2 P(P = x_i)}$$

Prove and Understand

- 1. Probability mean/variation V.S. Normal mean/variation

- 2. Binominal distribution probability

What is the probability that, among 3 students, any 2 of them getting a pass in the test, with the probability of passing the test equals 0.7?

Student	A	B	C	Probability
Case 1	P	P	F	0.7 × 0.7 × 0.3 = 0.147
Case 2	P	F	P	0.7 × 0.3 × 0.7 = 0.147
Case 3	F	P	P	0.3 × 0.7 × 0.7 = 0.147
Total				0.147+0.147+0.147=0.441

$$p(X=x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

- 3. Binomial mean/variation

Key concept

T6 L3

Probability Distribution

- Discrete Probability Distribution
- Binomial Distribution

- Continuous Probability Distribution
- Normal Distribution

Continuous Probability Distribution

- Continuous – uncountable number of values
  - In practice, a discrete variable with large range of values is also considered as continuous variable

- Probability distribution

Amount of Fill (liters)	Frequency	Relative Frequency
< 1.025	48	0.0048
1.025 < 1.030	122	0.0122
1.030 < 1.035	325	0.0325
1.035 < 1.040	695	0.0695
1.040 < 1.045	1198	0.1198
1.045 < 1.050	1664	0.1664
1.050 < 1.055	1896	0.1896
1.055 < 1.060	1664	0.1664
1.060 < 1.065	1198	0.1198
1.065 < 1.070	695	0.0695
1.070 < 1.075	325	0.0325
1.075 < 1.080	122	0.0122
1.080 or above	48	0.0048
		1.0000

Topic 1

- Summary table
  - Frequency distribution
  - Cumulative percentage distribution

Amount Spent ($)	Frequency	Relative Frequency
0 - < 100	40	0.40
100 - < 200	22	0.22
200 - < 300	15	0.15
300 - < 400	7	0.07
400 - < 500	5	0.05
500 - < 600	7	0.07
600 - < 700	0	0.00
700 - < 800	2	0.02
800 - < 900	2	0.02
900 - < 1000	0	0.00
Total	100	1.00

Continuous Probability Distribution

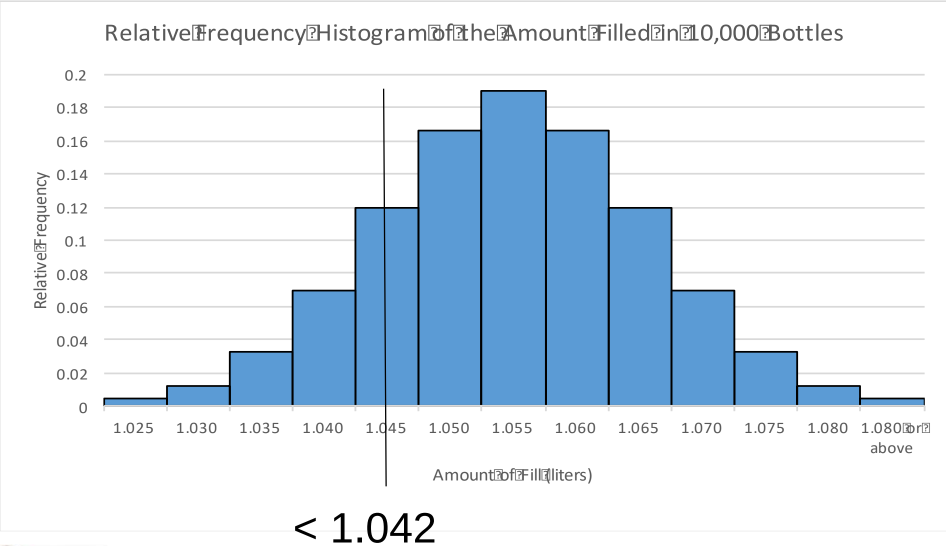
- Relative frequency histogram
  - relative frequency = percentage = Area
  - Area under a single point, $P(X=\boldsymbol{x})=0$
  - area ≈ Percentage polygon area

- Probability Density Function
  - Likelihood for X to be a certain value x_i
  - $f(x) \geq 0$ for all x_i of X
  - Total area below $f(x) = 1$
  - most important form - Normal Density Function

Connect mid-points Fit a math curve

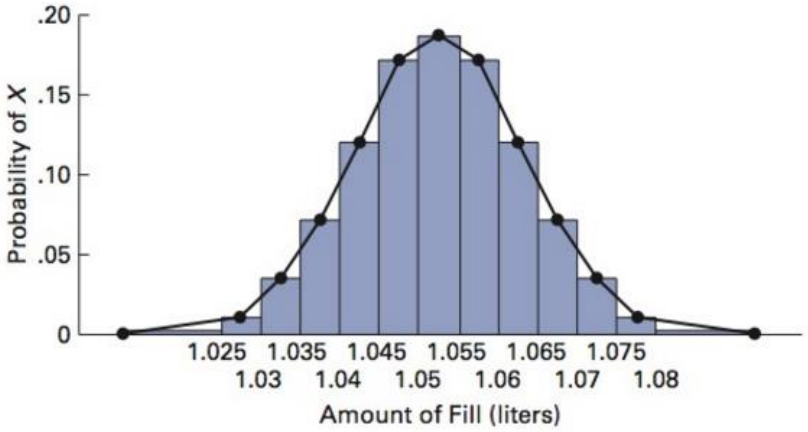
- Histogram ----------> Polygon ----------> Mathematical Curve

(multi-rectangles) (a smooth line)



< 1.042

Know how to calculate: $P(X < 1.042)$



Normal Distribution

- continuous random variable X follows Normal Distribution $\mathrm{X} \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(\frac{1}{2})[\frac{x-\mu}{\sigma}]^2}$$

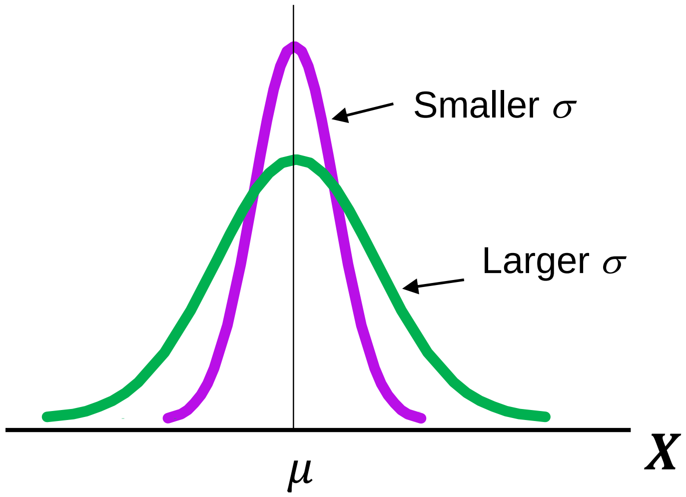
where $x =$ any value that the continuous random variable X can take in the range of -∞ to + ∞

$\mu =$ mean of the population

$\sigma =$ standard deviation of the population

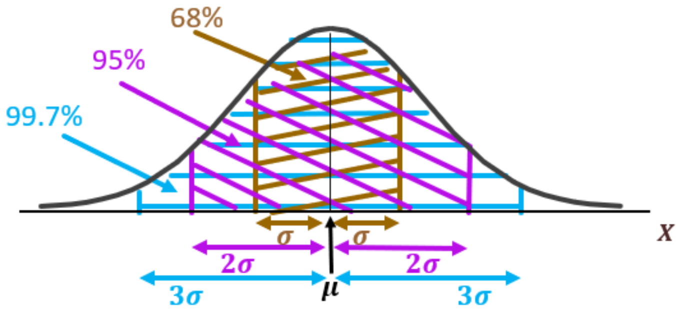
$e =$ constant, 2.71828…

$\pi =$ constant, 3.14159…



- Bell shaped, symmetrical about $\mathrm{X} = \mu$
- mean=median=mode
- spread is determined by σ
  - For smaller σ - clustered more closely around μ
  - For larger σ - more spread out and away from μ
- Has an infinite theoretical range, i.e. $-\infty$ to $+\infty$
- Follows the Empirical Rule

- Area within $\mu \pm \sigma$ equals 68% approximately
- Area within $\mu \pm 2\sigma$ equals 95% approximately
- Area within $\mu \pm 3\sigma$ equals 99.7% approximately

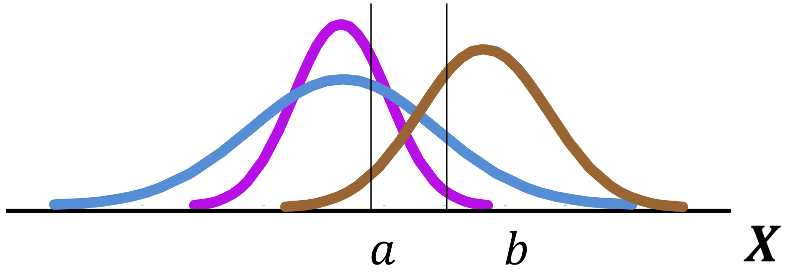


Normal Distribution

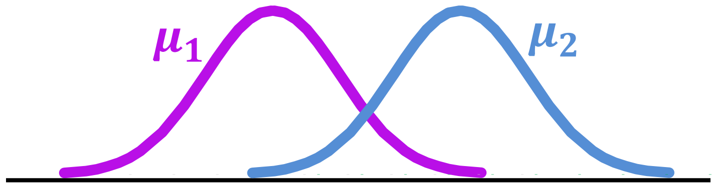
- Some rules
  - total area under the curve = 1
  - Probability = Area under the curve
  - Probability of any individual value is zero by definition, i.e. $P(X=a)=0$

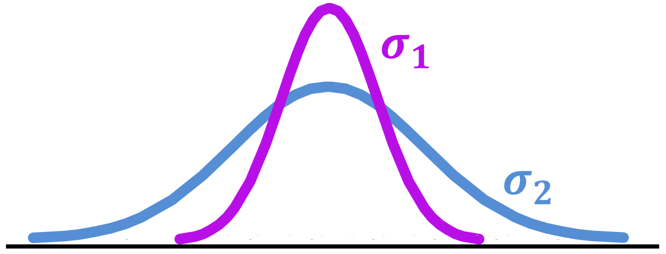
$$P(a \leq X \leq b) = P(a < X < b)$$

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi\sigma}} \int_a^b e^{-(\frac{1}{2})[\frac{x-\mu}{\sigma}]^2} \, dx$$



- How do the distribution change when
  - change μ
    - Increase? Decrease?
  - change σ
    - Increase? Decrease?
  - (important for understanding why we can standardize any normal distribution)

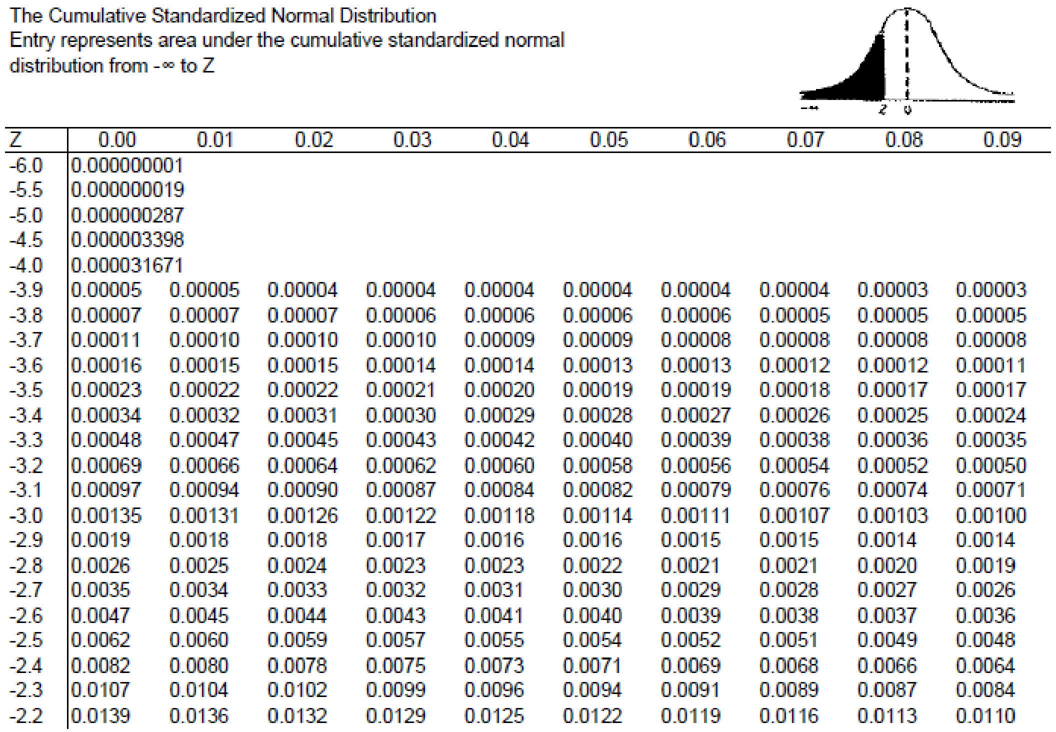




The Standardized Normal Distribution

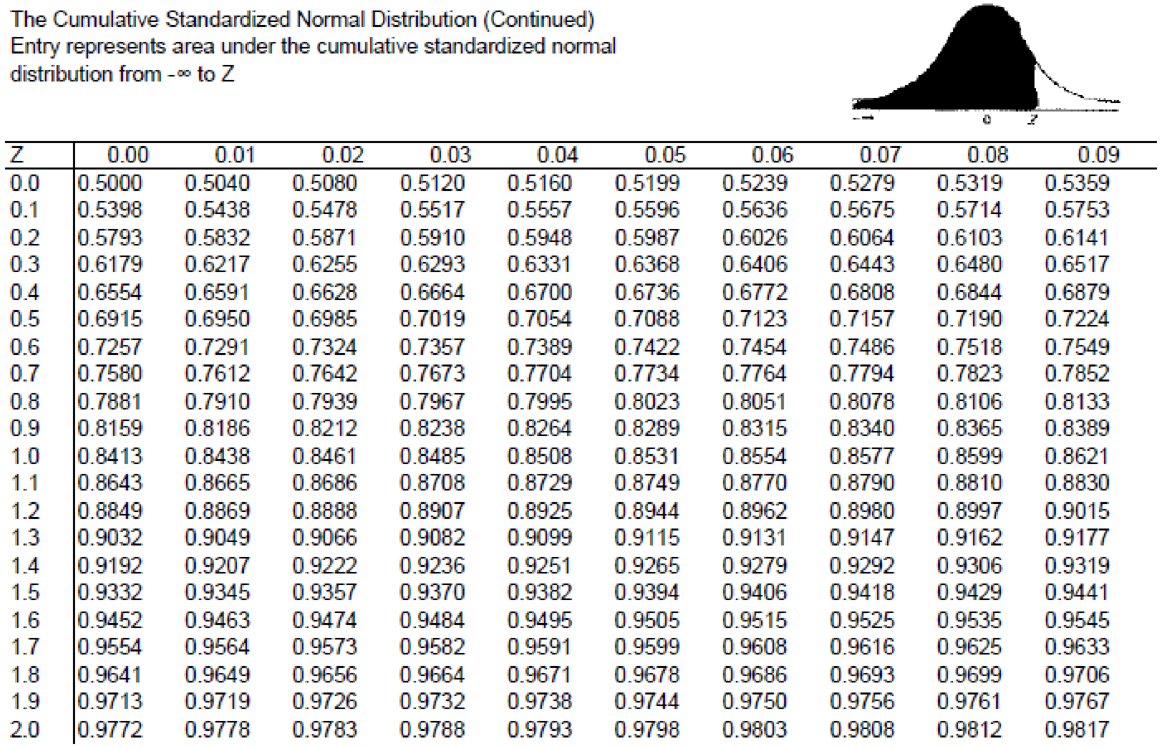
- Standard Normal Distribution $Z \sim N(0,1^2)$
  - μ=0, σ=1
  - Cumulative standardized normal tables = Z table
    - Find the probability of P(Z<a)

Know how to use the table



The Cumulative Standardized Normal Distribution
Entry represents area under the cumulative standardized normal
distribution from -∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-6.0	0.000000001									
-5.5	0.000000019									
-5.0	0.000000287									
-4.5	0.000003398									
-4.0	0.000031671									
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110



The Cumulative Standardized Normal Distribution (Continued)
Entry represents area under the cumulative standardized normal
distribution from -∞ to Z

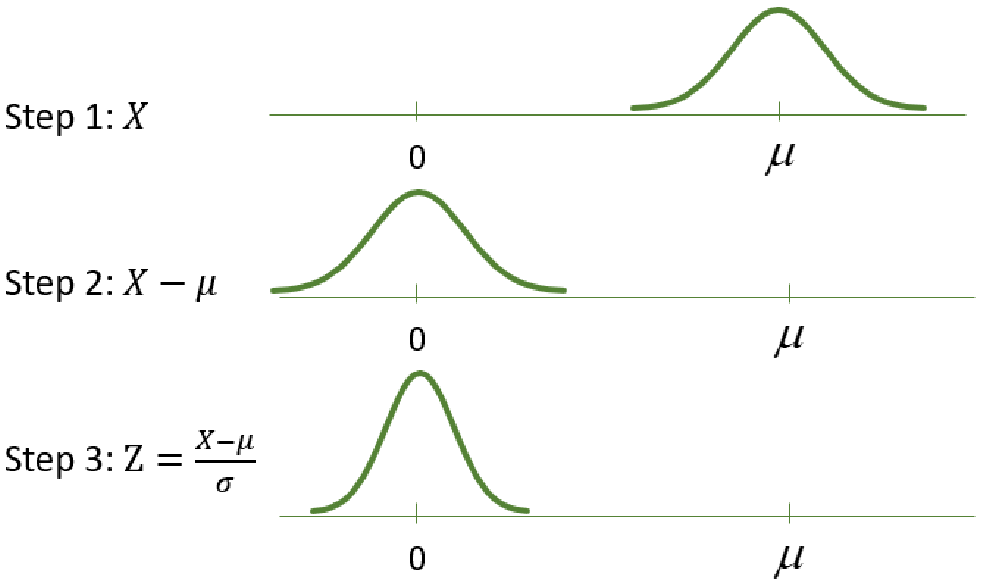
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

The Standardized Normal Distribution

- For any $X \sim N(\mu, \sigma^2)$, it can be standardized to $Z \sim N(0, 1^2)$ with the following formula

$$\boldsymbol{Z = \frac{X - \mu}{\sigma}}$$

why we can standardize any normal distribution???



The Standardized Normal Distribution

Know how to calculate probabilities

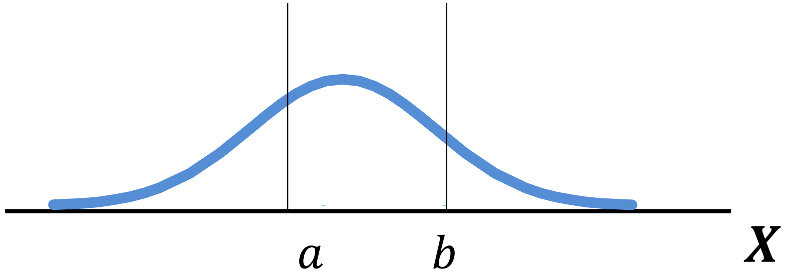
$$P(X = a) = 0$$

$$P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) \\ = P\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

$$P(X \geq a) = 1 - P(X \leq a) \\ = 1 - P\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) \\ = P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

$$P(X \leq a) + P(X \geq b) = P(X \leq a) + [1 - P(X \leq b)] \\ = P\left(Z \leq \frac{a-\mu}{\sigma}\right) + \left[1 - P\left(Z \leq \frac{b-\mu}{\sigma}\right)\right]$$



$\boldsymbol{X \sim N(\mu, \sigma^2)}$ ➔ $\boldsymbol{Z \sim N(0, 1^2)}$ ➔ **Probability**

Step 1: always change to $\leq$ first
because Z table by definition is P(Z<a)

Step 2: use formula to standardize

$$Z = \frac{X-\mu}{\sigma}$$

Step 3: calculate Z

Step 4: find probability in Z table

The Standardized Normal Distribution

$\boldsymbol{X \sim N(\mu, \sigma^2) \leftarrow Z \sim N(0, 1^2) \leftarrow}$ **Probability**

Step 1: always change to $\leq$ first
because Z table by definition is P(Z<a)

Step 2: find Z with given probability in Z table
If the prob is exactly in the middle of two numbers, take an average

Step 3: use formula to calculate X

$$X = Z\sigma + \mu$$

Know how to recover X

In class exercise:
1. in the top 5%
2. the middle 50%