R：Statistical Programming Methods
R：程式、機率與統計

Sampling Distribution

Exploring the data distribution

- A grouping of data into mutually exclusive classes showing the number of observations
  - Frequency Table
  - Percentiles and Box plot
  - Histogram
  - Density Plot
    - As opposed to the histogram, the density plot can smooth out the distribution of values and reduce the noise

Probability Distribution 機率分布

- A listing of all outcomes of an experiment and the probability associated with each outcome.

- The probability of a particular outcome is between 0 and 1 inclusive.
- The outcomes are mutually exclusive events.
- The list is exhaustive. The sum of the probabilities of the various events is equal to 1.

Random Variables 隨機變數

- A quantity resulting from an experiment that, by chance, can assume different values.

  - Discrete Random Variables 離散隨機變數
    - The numbers of heads appearing when a coin is tossed 3 times
    - The number of students earning an A in the class

  - Continuous Random Variables 連續隨機變數
    - The times of commercial flight between Taiwan and Hong Kong
    - The length of each song on an album

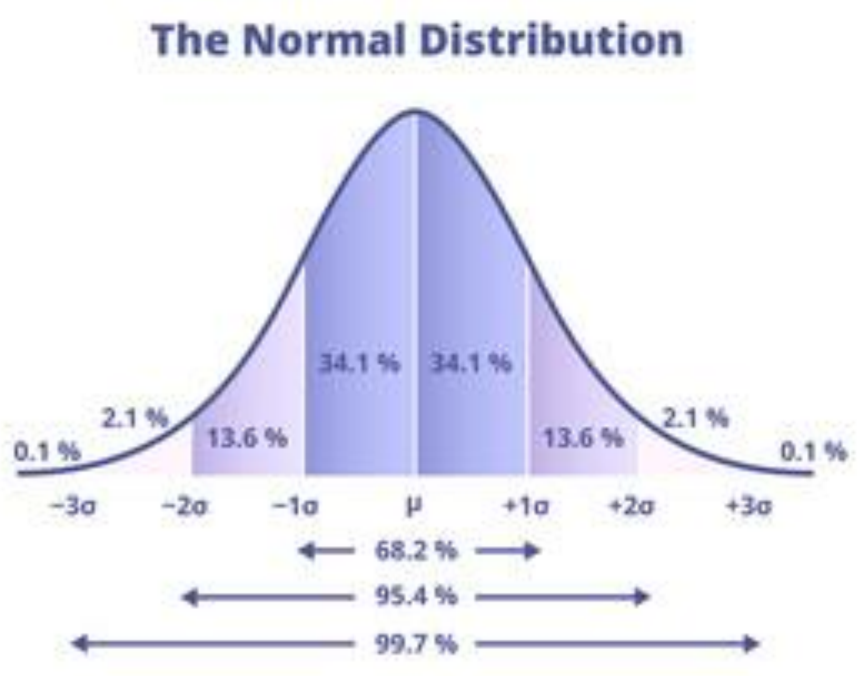
The Normal Distribution 常態分布

- Belongs to the family of continuous probability distributions
- The normal equation for

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

  - Where x is the normal random variable, σ is the standard deviation, and μ is the mean
- The normal equation is the probability density function for the normal distribution

Normal Distribution



Also:

- Mean = median = mode
- 1σ away ~ 68%
- 2σ away ~ 95%
- 3σ away ~ 99.7%
- Symmetry (50% of values to the left, 50% to the right)

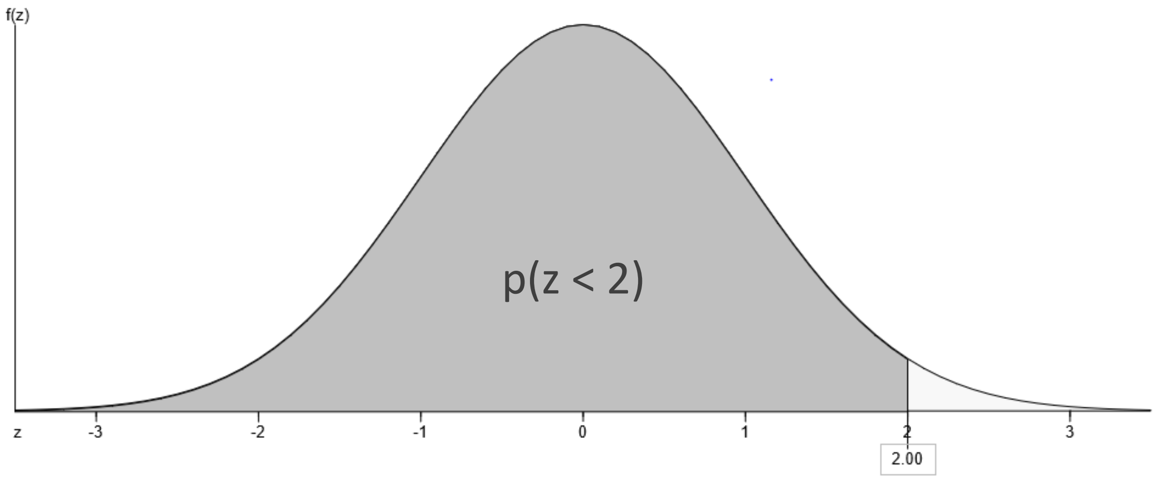
Importance of Normal Distribution

- Many things tend to <u>**approximately**</u> follow the normal distribution.
  - Central Limit Theorem 中央極限定理

- Normally distributed populations and samples allow us to use parametric tests (有母數分析方法)which are:
  - Simple to use and to interpret
  - Statistically powerful

Standard Normal Distribution

- Special case of the normal distribution, where μ = 0, and σ = 1



- The shaded area represents the probability of an event occurring under the standard normal population.
- For example:
  - If we picked a random number from a population with mean of 0 and standard deviation of 1, what is the probability of that number being LESS than 2?
    - Calculus
    - Z-tables
    - Software

Using Z-tables

- The probability of X between a and b

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \longrightarrow$$

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	0.7517	0.7549
0.7	0.7580	0.7611	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	0.9505	0.9515	0.9525	0.9535	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9554	0.9554	0.9554	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9641	0.9641	0.9641	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9713	0.9713	0.9713	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9772	0.9772	0.9772	0.9808	0.9812	0.9817

The probability of z less than 2

Z-tables

- Only works if mean is 0 and standard deviation is 1
  - So only for the standard normal distribution

- To work for any normal distribution we need to use a conversion formula:
  - $Z = \frac{X - \mu}{\sigma}$