STATISTICAL ANALYSIS

DESCRIPTIVE AND INFERENTIAL STATISTIC

PROBABILITY

There are two approach to the probability:

- **Classical approach: the number of successful outcomes is divided by the total number of equally likely outcomes.**

- **Frequency approach: is the frequency of an event occurring in large number of trials. For example, you flip a coin 1000 times and the number of occurrences of head up is 520. The probability of head up is 520/1000=0.52.**

Probability—Examples

Let us try to understand a probability with regard to frequency approach. The probability of an occurrence for an event labelled A is defined as the ratio of the number of events where event A occurs to the total number of possible events that could occur.

If you toss a coin, only two events can occur, either a head up or a tail.

***P*(H) denotes probability of event head is up. You can calculate the probability of head coming up using the formula :**

$$P(\mathrm{H}) = \frac{Number\ of\ time\ head\ is\ up}{(Number\ of\ time\ head\ is\ up + Number\ of\ time\ tail\ is\ up)}$$

VARIABLES

Variable: is the fundamental element of statistical analysis. And it represents the shape of our measurements.

Broadly, variable can be classified into two types:

- **Categorical variables (qualitative): which has two sub groups; nominal and ordinal**
- **Quantitative variable: which has two sub groups as well; discrete and continuous variables**

CATEGORICAL VARIABLES

- **Ordinal variable: observations can take a value that can be logically ordered or ranked. Examples of ordinal categorical variables include academic grades (i.e. A, B, C), clothing size (i.e. small, medium, large, extra large) and attitudes (i.e. strongly agree, agree, disagree, strongly disagree).**
- **Nominal variable: observations can take a value that is not able to be organised in a logical sequence. Examples of nominal categorical variables include sex, business type, eye colour, religion and brand.**

QUANTITATIVE VARIABLE

- **Continuous variable: is a numeric variable. Observations can take any value between a certain set of real numbers. The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows. Examples of continuous variables include height, time, age, and temperature.**
- **Discrete variable: is a numeric variable. Observations can take a value based on a count from a set of distinct whole values. A discrete variable cannot take the value of a fraction between one value and the next closest value. Examples of discrete variables include the number of registered cars, number of business locations, and number of children in a family, all of which measured as whole units (i.e. 1, 2, 3 cars)**

MEAN, MOD& MEDIAN

Mean: sum of all individual observations divided by the sum of the observations.

There are

several types of means, such as arithmetic mean, geometric mean and harmonic mean.

arithmetic mean: The following numbers are the rats' weights:

132, 139, 134, 141, 145, 141, 140, 166, 186, 183, find out the arithmetic mean?

$$= \frac{\sum X}{N} = \frac{1507}{10} = 150.7 \text{ g}$$

MEAN, MOD& MEDIAN

geometric mean: we can calculate the this kind of mean by convers the measuring numbers to the log. By using the same example:

Rats' body weights (g): Calculation of geometric mean of body weight of rats

Body weight (g)

Linear scale 132, 139, 134, 141, 145, 141, 140, 166, 186, 183

Log scale 2.12, 2.14, 2.13, 2.15, 2.16, 2.15, 2.15, 2.22,

2.27, 2.26

$=(\sum X)/N$ = 21.7/10= 2.17 you can take the anti-log to find out the mean in linear scale= 147.9g

MEAN, MOD& MEDIAN

Harmonic mean: is calculated by finding the mean of the reciprocals of the values and then finding the reciprocal of the mean.

Linear scale 132, 139, 134, 141, 145, 141, 140, 166, 186, 183

Reciprocal= 1/value= 1/132= 0.0076

Reciprocal 0.0076, 0.0072, 0.0075, 0.0071, 0.0069, 0.0071,

0.0071, 0.0060, 0.0054, 0.0055

Harmonic mean= 0.0673/10= 0.0067

MEAN, MOD& MEDIAN

MOD: can be defined as the most frequent value among the data set. Example:

130, 140, 140, 150, 140, 160, 140, 110, 120

The mod of this data set is 140(appears 4 times).

MEAN, MOD& MEDIAN

Median: is type of central tendency measurements and can be calculated by ranking the values from the lowest to the highest value or vice versa.

130, 140, 150, 160, 110, 120

- **rank the values:**

110, 120, 130, 140, 150, 160, 170

- **140 is the median of the values.**
- **Exercise find out the median of the following data set:**

  13, 16, 12, 11, 30, 28, 33, 27.

VARIANCE

Variance: measures how far a set of numbers is spread out

Rat No.	Weight (g)	(X-*mean* **)**	$(X\text{-mean})^2$
1	245	-7.6	57.76
2	254	+1.4	1.96
3	239	-13.6	184.96
4	266	+13.4	179.56
5	259	+6.4	40.96
No.	5		
sum	1263	0	465.2
Mean	252.6		

VARIANCE

Variance = 465.2/(5–1) = 116.3

STANDARD DEVIATION& STANDARD ERROR

Standard deviation is the square root of variation:

SD =$\sqrt{\text{Variance}}$ = $\sqrt{116.3}$ = ±10.78

Standard deviation is a useful measure to explain the distribution of the

sample observations around the mean

Standard Error (SE) : is the SD of the mean. It is considered as a measure of the precision of the sample mean

SE = SD/ $\sqrt{n}$ = 10.78/ $\sqrt{5}$ = ± 4.82

SE measures the variation in the means of the samples

DISTRIBUTION

It is important to know how the data are distributed for selecting a statistical tool for data analysis.

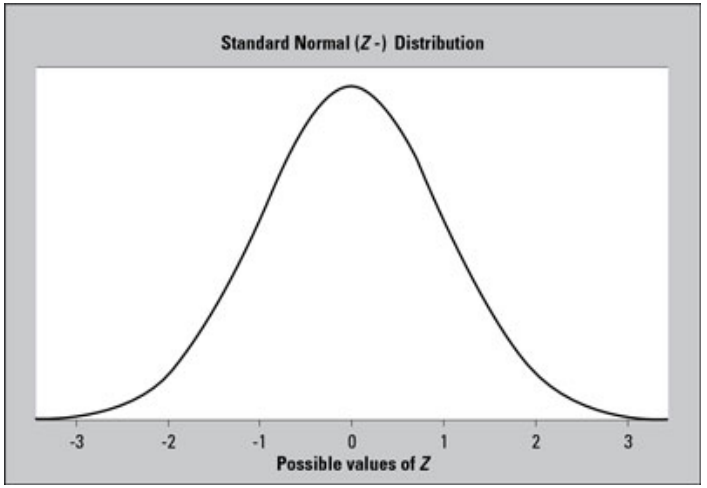
Type of data distribution:

- Normal distribution (Gaussian):
- Binomial distribution:
- Poisson distribution:
- Exponential distribution:
- Skewness and Kurtosis

NORMAL DISTRIBUTION

Data should be tested for normality before starting with any statistical analysis

Normal distribution



TEST FOR ANALYSING DATA DISTRIBUTION AND HOMOGENEITY

Data must be tested firstly to determine the type of their distribution. In general, we have to make sure that our data following the normal distribution.

Mainly, the data show either normal or non-normal distribution.

What I should do to test the data distribution?

There are three main tests can be used:

- **Kolmogorov-Smirnov test**
- **Shapiro-Wilk W test**
- **Levene's homogeneity test**