U3A Maths

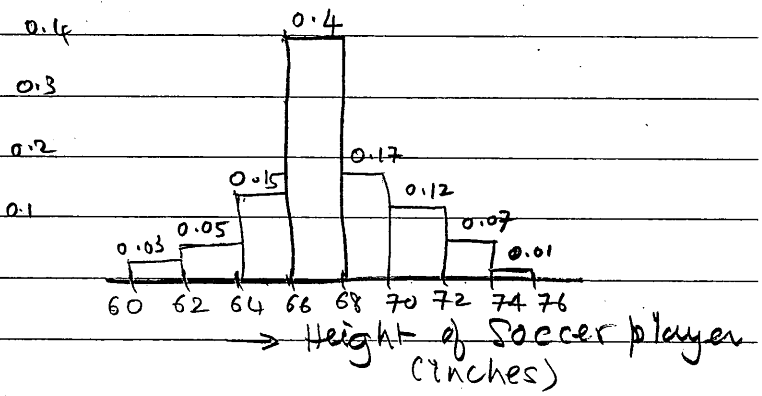
13-May-2022 Term 2 / Week 3

Normal Distribution

- For a large data set, it is convenient to express the data as a "Grouped Frequency Table"
- Expressing the frequency value as a "ratio" of the total number of data values makes the table very powerful
- Such a ratio is called "Relative Frequency"

-2-

- The relative frequency essentially represents the probability of a data value occuring in that group (range).
- For our soccer player height example, a bar chart of relative frequency is as shown below:

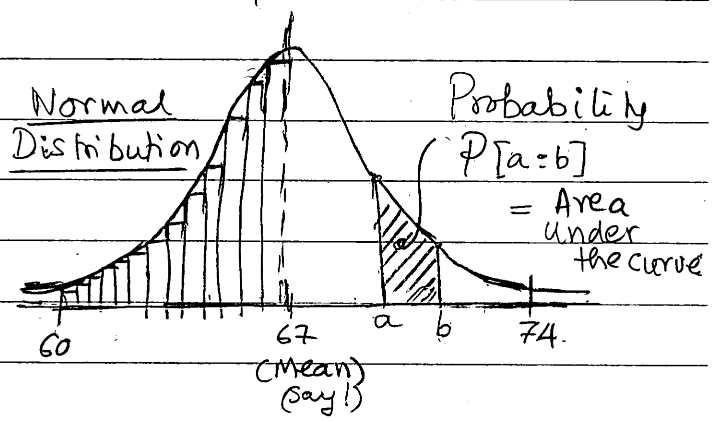


-3-

- Expressing (defining!) the relative frequency as the area of the bar, instead of the height of the bar makes the bar graph even more powerful!
- This will enable the interpolation of the probability value - even within each bar!

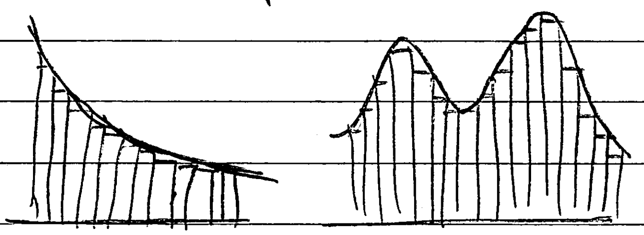
  Ex: $P[66:67] = \frac{1}{2} \times 0.4 = 0.2$
- If we had a large set of data and a smaller group size, the bar chart can be considered to be (almost!) continuous & smooth curve

-4-



- The above shape of the curve is not a coincidence, it occurs frequently for many randomly occuring data values.
- Hence it is called "Normal (probability) distribution Curve"

- Having said that, other probability distributions can also occur in practice, Some examples are as below.



Exponential   Multimodal.

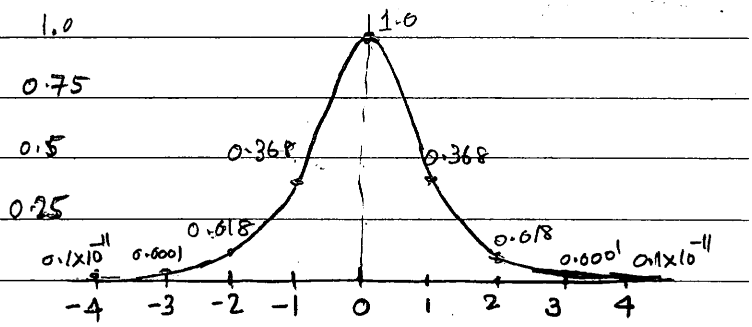
etc

- For the present, let us explore Normal Distribution, and in future we will explore other types.
- Finally, note that the above distributions assume "continuous data"

- For mathematical analysis, first we need find a mathematical function to represent the "idealised" normal distribution curve.

- Let us see how the following function plot looks like:

$f(x) = e^{-x^2}$ for $x = -4$ to 4



- The "idealised" normal distribution curve is based on the above function

- Note that we are interested in the "area under the curve", which can be obtained by integration!

- The area under normal distribution curve for given values of mean (μ) and Std. dev (σ) was first defined by Carl.F. Gauss

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

Area under the curve from $x = a$ to b

- The above equation can be simplified using Z-score, we have:

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

Also, $\frac{dZ}{dx} = \frac{1}{\sigma}$ $\therefore$ $dx = \frac{dZ}{\sigma}$

- Rewriting, we have:

$$P[a;b] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}Z^2} dZ$$

The above equation is called "Standard" normal distribution. The results of the equation is given as a table in Statistics books!

H.W. why do we need $\left(\frac{1}{\sqrt{2\pi}}\right)$ & $\left(\frac{1}{2}\right)$?