U3A Maths

6-May-2022 Term 2/Week 2

Review of Term 1

Let us consider an example.

- The heights of 200 male soccer players (in inches) are as given below:

  61, 73, 65, 63, 73, 66, 65, 68, 75, 67, 68, 64, 71, 69, 70, 62, 67, 66, 68, 69, 65, 71, 67, 65, 73, 74, 62, etc.

- What is the "typical" height of the soccer player
  - Average, Mean, Median, Mode etc.

-2-

- For statistical analysis we use the "Mean" value

$$\bar{x} = \frac{\Sigma x}{N} \quad \text{(sample) (size = N)}$$

$$\mu = \frac{\Sigma x}{n} \quad \text{(Popln) (size = n)}$$

- We are also interested "Mean of the deviations" from the mean value

Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma (x-\mu)^2}{n}} \quad \text{(Population)}$$

$$S_N = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}} \quad \text{(Sample)}$$

-3-

- Corrected Std. Dev. for Samples

$$S_{N-1} = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N} \cdot \left(\frac{N}{N-1}\right)}$$

Bessel Correction Factor

This is useful when estimating population std. dev. (σ) based on std. dev. of samples.

- Z Score

expresses the data values as standardised values, which are useful to data visualisation.

$$Z = \left(\frac{x-\mu}{\sigma}\right)$$

-4-

Ex: For the above data set (population), it is given that

Mean (μ) = 67.6 inches

Std. Dev. (σ) = 2.8 inches.

Calculate Z-Scores for a soccer player with a heigt of (a) 65 (b) 75 (c) 62

(a) $Z = \left(\frac{x-\mu}{\sigma}\right) = \frac{65-67.6}{2.8} = -0.92$

(b) $Z = (75-67.6)/2.8 = +2.64$

(c) $Z = (62-67.6)/2.8 = -2.0$

For "Normal" distribution:

$Z \Rightarrow -1$ to $+1 \Rightarrow$ Mean $\pm$ 1 Std dev $\Rightarrow$ 68%

$Z \Rightarrow -2$ to $+2 \Rightarrow$ Mean $\pm$ 2 Std dev $\Rightarrow$ 95%

$Z \Rightarrow -3$ to $+3 \Rightarrow$ Mean $\pm$ 3 Std dev $\Rightarrow$ 99.7%

-5-

- For large number of data **Grouped Frequency Table** provides an efficient way to present the data values (Note: Rel. Freq = $f/\Sigma f$)
- For our soccer players data.

Group (class)	Freq (f)	Relative freq
60 - 62	6	0.03
>62 - 64	10	0.05
>64 - 66	30	0.15
>66 - 68	80	0.40
>68 - 70	34	0.17
>70 - 72	24	0.12
>72 - 74	14	0.07
>74 - 76	2	0.01
	$\Sigma f = 200$	1.00

-6-

- Relative Frequency is more useful in practice, as it corresponds to "probability" of a data value occuring in that group.

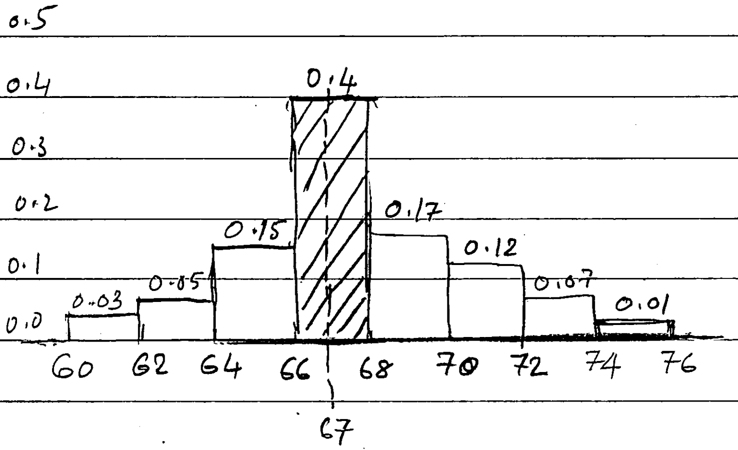
- For Ex, probability of a soccer player's height between >66 to 68 inches is 0.4 or 40%.

$$P[66:68] = 0.4$$

- In fact a "barchart" of Relative Freq values provides for even better analysis.

-7-

Relative Freq Bar Chart



- Let us "define" that Rel. Freq or "probability" of the group corresponds to the area of the bar (rather than the height).
- The above definition provides for more "powerful" analysis!

-8-

- For Ex, the probability of a soccer player's height between >66 to 70 inches is:

$$P[66:70] = P[66:68] + P[68:70]$$
$$= 0.4 + 0.17 = 0.57$$
or 57%

- Also, the probability of a given player's height between >66 to 67 inches is approx. half the area of the bar!

$$P[66:67] \simeq 0.4/2$$
$$= 0.2 \text{ or } 20\%$$

- The above analysis assumes uniform distribution between 66 to 68, which may not be true in practice!