Measures of Dispersion (Spread)

1. **Range—difference between largest and smallest values**

$$R = x_L - x_S$$

2. **Percentiles or quantiles**

p^{th} **percentile is the value V_p such that p percent of the sample points are $\leq V_p$.**

P_{50} **= median**

Def: After sorting observations from smallest to largest, p^{th} percentile =

a. $(k+1)^{th}$ largest sample point if $np/100$ is not an integer (k = largest integer less than $np/100$)

b. average of the $(np/100)^{th}$ and $(np/100+1)^{th}$ largest observations if $np/100$ is an integer

Example: For ER data, find the 25th percentile.

$n = 30$, $p = 25 \Rightarrow np/100 = 30*25/100 = 7.5$
(not an integer)

Therefore, 25th percentile

= (7 + 1)th largest value = 34

Ages of 30 patients seen in the ER of a hospital on a Friday night

35	32	21	43	39	60
36	12	54	45	37	53
45	23	64	10	34	22
36	45	55	44	55	46
22	38	35	56	45	57

Ordered Array

10	23	36	43	45	55
12	32	36	44	46	*56*
21	*34*	*37*	*45*	*53*	57
22	35	38	45	54	60
22	35	39	45	55	64

Find the 50th percentile.

n = 30, p = 50 ⇒ np/100 = 30*50/100 = 15 (an integer)

Therefore, 50th percentile

= average of (30*50/100)th and (30*50/100 +1)th largest values

= average of 15th and 16th largest values

= average of 39, 43 = 41

Find P_{40}.

Some Measures of Spread

(To eliminate the extremities)

$P_{90} - P_{10}$

$P_{75} - P_{25}$ **– Called Inter-quartile Range**

3. Variance

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}$$

Example: Fasting blood glucose levels of six children

74
68
60
65
71
82

$$\overline{x} = 70$$

x_i	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$
74	**74 – 70 = 4**	**16**
68	**-2**	**4**
60	**-10**	**100**
65	**-5**	**25**
71	**1**	**1**
82	**12**	**144**
Total	**0**	**290 =** $\sum \left(x_i - \overline{x}\right)^2$

s^2 **= 290 / 5 = 58**

Use an alternative formula if the data set is large or if the mean has a long decimal.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

	x_i	x_i^2
1.	74	5476
2.	68	4624
3.	60	3600
4.	65	4225
5.	71	5041
6.	82	6724
	Total = 420	29,690

$$s^2 = \frac{29{,}690 - \frac{420^2}{6}}{5}$$

$$= \frac{29{,}690 - 29{,}400}{5}$$

$$= \frac{290}{5} = 58$$

4. Standard Deviation

$$s = \sqrt{s^2}$$

$$= \sqrt{\mathbf{58}}$$

$$= \mathbf{7.62}$$

Translation:

$x_1, x_2, \ldots, x_n$ **– original sample.**

$x_1+c, x_2+c, \ldots, x_n+c$ **– translated sample**

Def: $y_i = x_i + c,\ i = 1, \ldots, n$

Therefore, $s_y^2 = s_x^2$

Scaling:

$x_1, x_2, \ldots, x_n$ **– original sample.**

$cx_1, cx_2, \ldots, cx_n$ **– rescaled sample**

Def: $y_i = cx_i,\ i = 1, \ldots, n$

$$s_y^2 = c^2 s_x^2$$

$$s_y = c s_x$$

Original sample	Scaled sample
0.013	13
0.024	24
0.010	10
0.009	9
0.014	14
$\Sigma = 0.070$	$\Sigma = 70$
$\bar{x} = 0.014$	$\bar{y} = 14 = 1000*0.014 = c*\bar{x}$

$$s_x^2 = \frac{(.013-.014)^2 + \ldots + (.014-.014)^2}{4}$$

$$= \frac{.000142}{4}$$

$$= .0000355$$

$$s_y^2 = \frac{(13-14)^2 + \ldots + (14-14)^2}{4}$$

$$= \frac{142}{4}$$

$$= 35.5 = .0000355*1{,}000{,}000 = s_x^2 * c^2$$

5. Coefficient of Variation

The series of data for which the coefficient of variation is large indicates that the group is more variable and it is less stable or less uniform

$$CV = \frac{s}{\bar{x}} * 100\%$$

Example:

Blood Glucose in Children:

$$\bar{x} = 70, \quad s = 7.62$$

Blood Glucose in Adults:

$$\bar{x} = 110, \quad s = 18.65$$

$$CV_C = \frac{7.62}{70} * 100\% = 10.9\%$$

$$CV_A = \frac{18.65}{110} * 100\% = 17.0\%$$