Coding

Rules of coding

Suppose our original variable (e.g. heights in cm) was x. Then y would represent the heights with 10cm added on to each value.

Coding	Effect on $\bar{x}$	Effect on σ
$y = x + 10$		
$y = 3x$		
$y = 2x - 5$		

You might get any **linear** coding (i.e. using $\times + \div -$). We might think that any operation on the values has the same effect on the mean. But note for example that **squaring** the values would not square the mean; we already know that $\Sigma x^2 \neq (\Sigma x)^2$ in general.

Quick-fire Questions

Old mean $\bar{x}$	Old σ_x	Coding	New mean $\bar{y}$	New σ_y
36	4	$y = x - 20$		
		$y = 2x$	72	16
35	4	$y = 3x - 20$		
		$y = \frac{x}{2}$	20	$\frac{3}{2}$
11	27	$y = \frac{x + 10}{3}$		
		$y = \frac{x - 100}{5}$	40	5

Example Exam Question

The following table summarises the times, *t* minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) *t*	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(*a*) Estimate the mean and standard deviation of these data. **(5)**
(*b*) Use linear interpolation to estimate the value of the median. **(2)**
(*c*) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**
(*d*) Estimate the interquartile range of this distribution. **(2)**
(*e*) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data. **(1)**

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) *t*	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(*f*) Without further calculations, explain the effect this would have on each of the estimates found in parts (*a*), (*b*), (*c*) and (*d*). **(3)**