Height Investigation Solutions

TASK A - Shape of Distributions

a) Children tend to round their heights to the nearest 5 cm.

b) The distribution for girls' heights is slightly positively skewed whereas the distribution for boys' heights may be considered symmetrical.

TASK B - Plotting a Histogram

a)

Height (cm)	130-(140)	140-(150)	150-(160)	160-(165)	165-(170)	170-(180)	180-190
Frequency	2	9	51	45	16	7	1
Class width	10	10	10	5	5	10	10
Frequency density: $\frac{frequency}{class\ width}$	0.2	0.9	5.1	9	3.2	0.7	0.1

Table 1: Grouped data for a random sample of heights of 13 year old UK girls taken from *CensusAtSchool* 2010/2011. Sample size 131.

b) This is a better representation of the data than in figure 1 as there are no columns with zero frequency. This shows the shape of the distribution more clearly.

c) This data may be normally distributed. Especially as height is a measurement.

TASK C - Calculating a mean and standard deviation from grouped data

a) For girls the values for the mean and median are not the same as the mean is distorted by a few extreme values. For boys the value of the mean is very close to the median as the distribution is symmetrical.

b) It is advised to use the mean and standard deviation as measures of location and spread for these data as there are very few extreme values to distort these statistics.
Using the data in table 3 calculate the mean and standard deviation.

c)

Height (cm)	130 -	140 -	150 -	160 -	165 -	170 -	180 -190	Total
Frequency (f)	2	9	51	45	16	7	1	131
Midpoint (x)	135	145	155	162.5	167.5	175	185	
Midpoint x Frequency (fx)	270	1,305	7,905	7,312.5	2,680	1,225	185	20,882.5
Frequency × Midpoint (fx^2)	36,450	189,225	1,225,275	1,188,281	448,900	214,375	34,225	3,336,731

Table 3: Grouped data for a random sample of heights of 13 year old UK children taken from *CensusAtSchool* 2010/2011.

Height Investigation Solutions

d) Mean = $\sum\frac{fx}{n}$ = $\frac{38522.5}{131}$ = 159.41 = 159.4 cm

e) Sample variance = $\frac{1}{n-1}(\sum fx^2 - n\bar{x}^2)$

= $\frac{1}{130}(3336731 - 131 \times 159.41^2)$

= 60.14

Therefore, sample standard deviation = $\sqrt{60.14}$ = 7.75 = 7.8 cm.

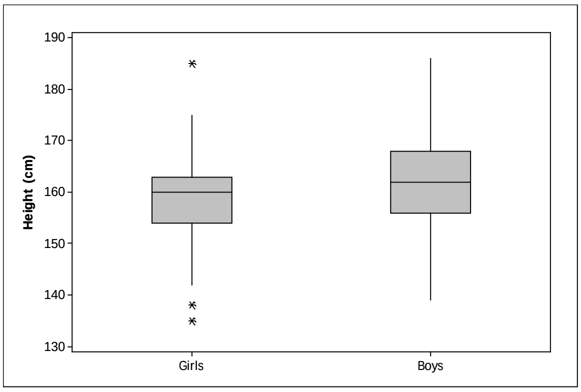
f) The answers are approximate as the values in each group have been estimated as the midpoints.

g) Nowadays computers are used to calculate statistics for large datasets so approximations are not needed. However this may be useful when using secondary data and only have access to grouped data.

TASK D - Plotting and interpreting boxplots

a) To check for outlier for 13 year old UK girls.
Lower limit = 154 - 1.5×(160 – 154) = 145 cm
Since the minimum value of 135 cm < 145 cm
The height of 135 cm is considered an outlier.
Upper limit = 163 + 1.5×(160 – 154) = 172 cm
Since the maximum value of 185 cm > 172 cm
The height of 185 cm is considered an outlier.
Similarly for 13 year old UK boys
Lower limit = 156 - 1.5×(168 - 156) = 138 cm
Since the minimum value of 156 cm > 138 cm
The height of 138 cm is not considered an outlier.
Upper limit = 168 + 1.5 × (168 – 156) = 186 cm
Since the maximum value of 186 cm is equal to the upper limit, 186 cm is not considered an outlier.

b) The boxplot suggests 13 year old boys and girls heights are not very different. There is more spread in the heights of the boys in this sample.



Heights of 13 year old UK boys and girls taken from *CensusAtSchool* 2010/2011.