Height Investigation S1 Worksheet

Are 13 year old UK boys taller than 13 year old UK girls?

To investigate this question we have taken a random sample of size 240 of 13 year old children from the 14,000 who completed the online *CensusAtSchool* questionnaire in 2010/2011.

Firstly, we must look at the shape of the distributions of 13 year old boys' and girls' heights to decide the statistics we need to calculate, for example do we calculate the mean or the median.

Figures 1 and 2 show the distributions of the heights for this random sample of 13 year old UK children, split by gender.

TASK A - Shape of Distributions

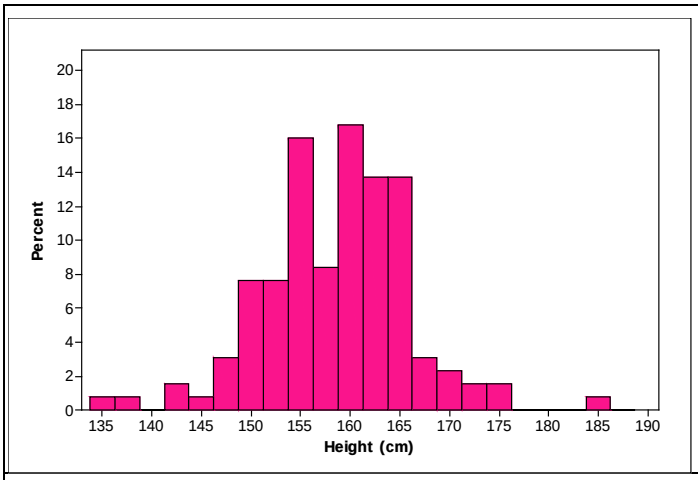

Figure 1: Histogram for a random sample of heights of 13 year old UK girls taken from *CensusAtSchool* 2010/2011. Sample size 131.

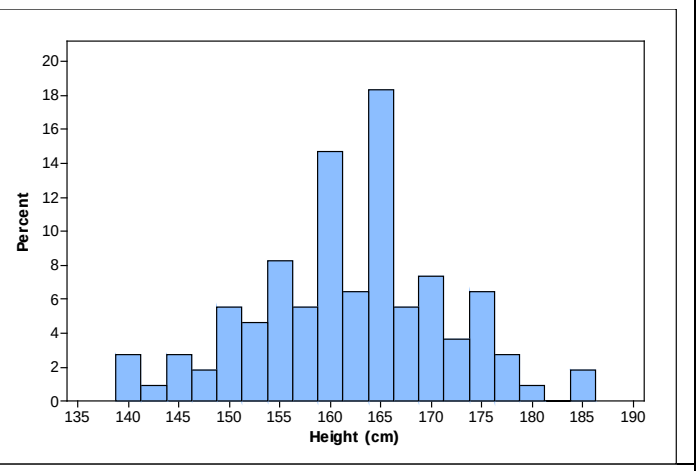

Figure 2: Histogram for a random sample of heights of 13 year old UK boys taken from *CensusAtSchool* 2010/2011. Sample size 109.

Interpret the shape of the distributions in figures 1 and 2.

a) Why do you think there are peaks at 150, 155, 160, 165, 170 and 175 cm?

b) What else do you notice about the shapes of the distributions in figures 1 and 2?

TASK B - Plotting a Histogram

It may be better to plot the same data using different column widths. Table 1 shows the data for the 131 UK girls from TASK A split into different height groups.

Note: The area of the columns in a histogram is proportional to the frequency for the group the column is representing. Figures 1 and 2 are using frequency for the vertical axis because the groups are equal. (Using frequency density on the vertical axis would give the same distribution shape, in this case, as the groups are of equal widths.)

Height Investigation S1 Worksheet

In table 1 130–(140) means from 130 up to and not including 140. The group widths are not all the same, e.g. the group 150–(160) has a width of 10 cm, whereas the group 165–(170) has a width of 5 cm. So the heights of some of the columns have to be adjusted to ensure the area of each column remains proportional the frequency of the group it is representing.

One way of doing this is to calculate the frequency density for each group and plotting frequency density on the vertical axis.

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							
Frequency density: $\frac{frequency}{class\ width}$							

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

a) Complete table 1 and plot a histogram for these data with the class sizes shown.

b) Do you think this a better representation of the data than in figure 1? Explain your answer.

c) Do you consider these data to be normally distributed?

TASK C (Calculating a mean and standard deviation from grouped data)

Below, in table 2, are the descriptive statistics for the random sample of the heights of 13 year old UK girls and boys represented in figures 1 and 2.

Variable	N Sample size	Mean	Standard Deviation	Minimum	Lower Quartile Q1	Median Q2	Upper Quartile Q3	Maximum
UK Girls' Height(cm)	131	158.6	7.4	135.0	154.0	160.0	163.0	185.0
UK Boys' Height(cm)	109	162.0	9.5	139.0	156.0	162.0	168.0	186.0

Table 2: Descriptive statistics for random samples of heights of 13 year old UK girls and of 13 year old UK boys taken from *CensusAtSchool* 2010/2011. (Table generated by Minitab.)

a) By inspection of the statistics in table 2, explain why there is a slight difference in the mean and median values for the heights of 13 year old UK girls whereas the mean and median values for the heights of 13 year old UK boys are the same.

b) In this analysis, is it better to use the mean and standard deviation, as measures of location and spread, rather than the median and the interquartile range? Explain your answer.

Height Investigation S1 Worksheet

Check that the mean and standard deviation values for the random sample of 13 year old UK girls taken from *CensusAtSchool* 2010/2011 in table 2 are approximately correct using the information in table 3.

c) Complete the table 3 and use this to calculate the statistics for d) and e).

Height (cm)	130-(140)	140-(150)	150-(160)	160-(165)	165-(170)	170-(180)	180-190	Total
Frequency (f)	2	9	51	45	16	7	1	
Midpoint (x)								
Midpoint x Frequency (fx)								
Frequency × Midpoint (fx^2)								

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

d) Mean = $\sum \frac{fx}{n}$

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

f) Explain why your answers are only approximately equal to those in table 2.

g) Explain why it is not likely that you will need to approximate these statistics nowadays.

TASK D (Plotting and interpreting boxplots)

On average are 13 year old UK boys taller than 13 year old UK girls?

As previously stated, table 2 shows the descriptive statistics for a random sample of 13 year old UK girls and a random sample of 13 year old UK boys taken from *CensusAtSchool* 2010/2011. Use the statistics in table 2 to investigate if 13 year old UK boys are taller than 13 year old UK girls.

a) Are the minimum and maximum heights for 13 year old boys and girls in table 2 considered outliers?
Use: Lower limit = Lower quartile – 1.5 * interquartile range
Upper limit = Upper quartile + 1.5 * interquartile range

b) Plot two boxplots, one for UK boys and one for UK girls, presented in a manner to allow comparisons between these two distributions.

c) By inspection of the boxplots, do you think there is a difference between boys' and girls' heights for the UK children in this sample?