

Importance of Data Processing



ALICE LEE 1858-1939 (STATISTICIAN)
WIKIPEDIA/ PUBLIC DOMAIN.



KARL PEARSON 1857-1936- EQUALLY DISTINGUISHED AS MATHEMATICIAN, LECTURER, WRITER, AND ORGANIZER OF STATISTICAL RESEARCH"
WIKI PEDIA/ PUBLIC DOMAIN.

Vasilliki found some interesting information about Alison Lee. She was born in England at the end of the nineteenth century and she was one of the first women that took a degree in mathematics. She demonstrated using statistics that there is no correlation between skull size and intelligence, and therefore there is no mental superiority of men to women because of the tendency of men to have a larger head than that of women. She worked together with Karl Pearson , who was an English Mathematician and Biostatician.

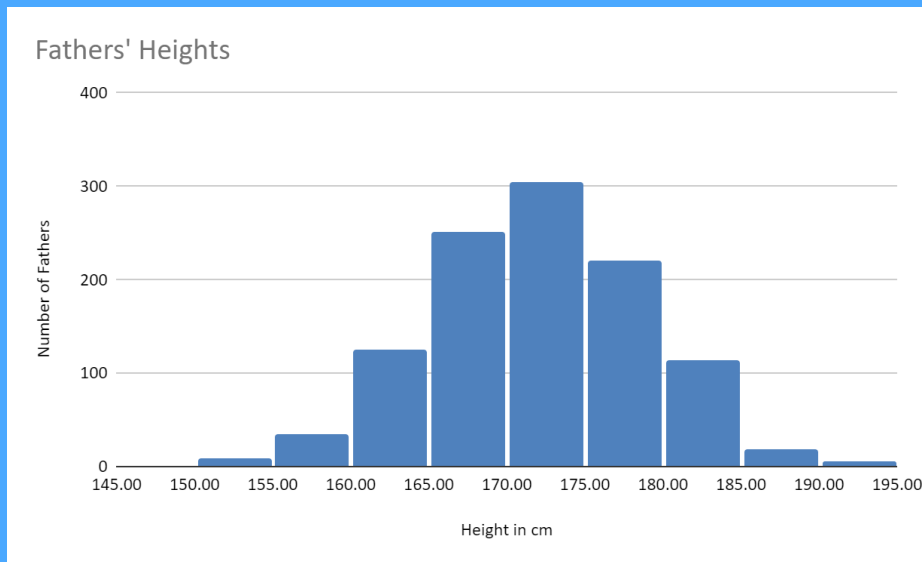
They conducted a survey about the assumption that Sons grow taller than their Fathers . For that reason they gathered data in the following manner: They asked more than 1000 families to measure the heights of the Father and the son of the family who was older than 18 years old .

USING HISTOGRAMS FOR DATA PROCESSING

When John saw the tables with the data he said that it is impossible to compare them because they are too many. Vassiliki proposed they use the diagrams which they have already learned in the previous lessons. Which of the diagrams (Stem and Leaf plot , dot plot or histogram) would you choose in order to represent the data? Explain your opinion.

Niki thought it's not convenient to use dot plot because the data are too many to draw 1068 dots. Also the stem and leaf plot is not suitable because the values are not natural numbers and it's not so easy to read the heights of every father . So she thought that it's more appropriate to use Histogram .She decided to use Excel in order to construct the Histogram.

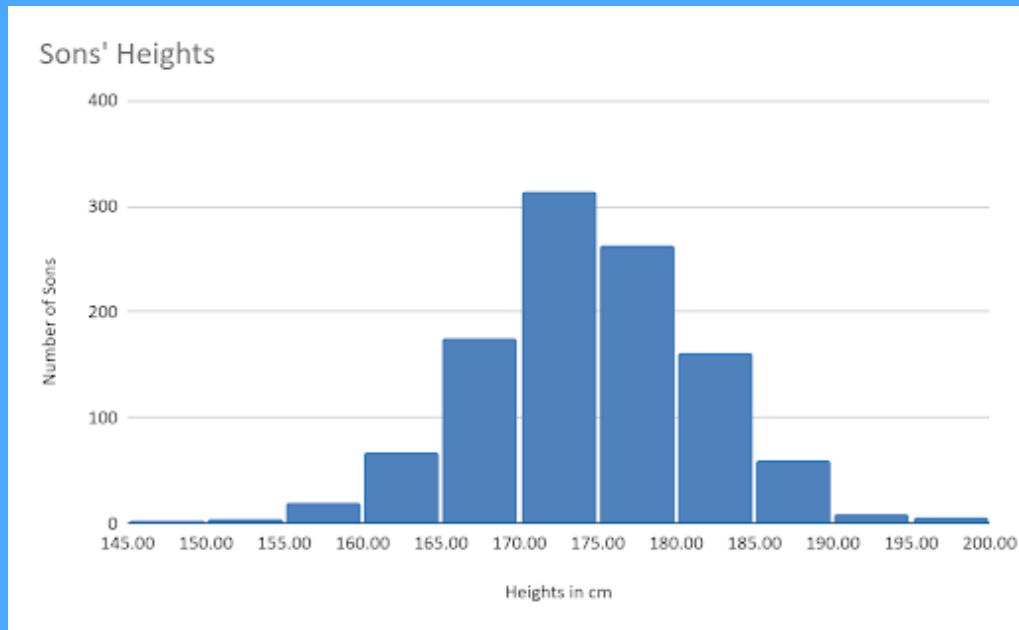
She chose to distribute the fathers' heights in groups with the bucket to be 5cm. She made a graph like this.



Observe the Histogram above and try to draw as many conclusions as you can about the distribution of Fathers' height. Try to answer the next questions:

- a) Which column has the bigger height ? What does that mean ?
-
-
- b) What is the range of the heights of most of the Fathers ? Explain your opinion.
-
-
- 3) Does the Histogram appear to have some kind of symmetry? What does that mean?
-
-

Try to construct on google docs the Histogram of the heights of the sons with bucket 5 cm. Is your histogram the same as the diagram below?



Compare the two Histograms and try to answer to the following questions :

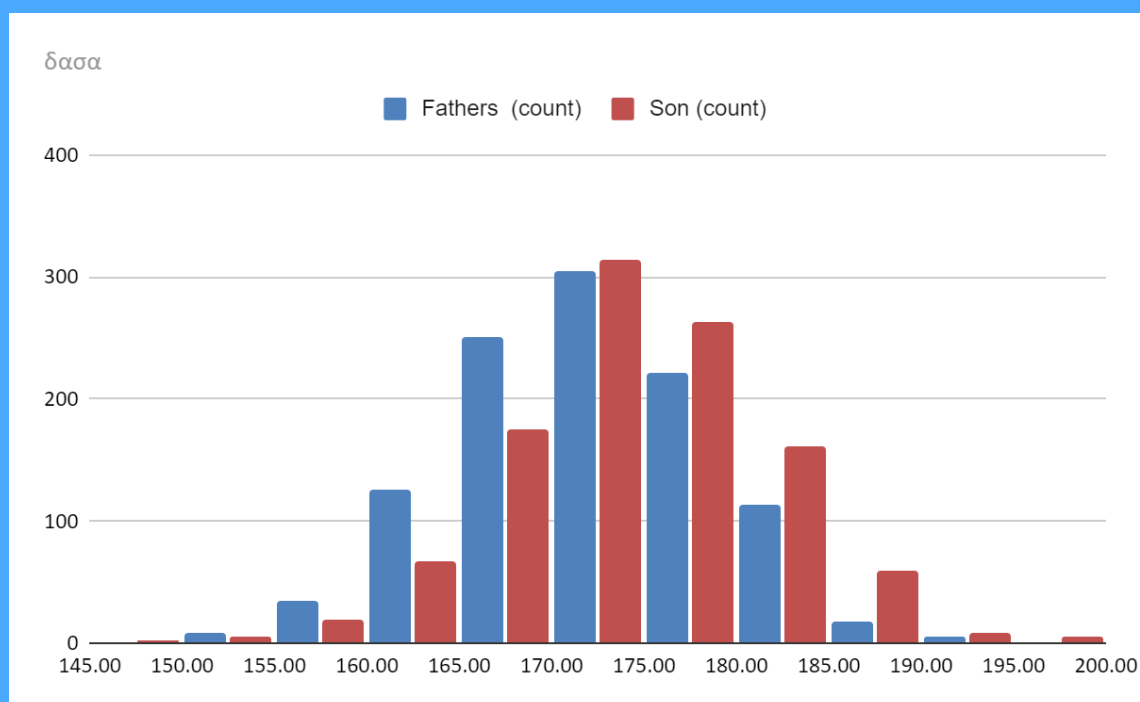
What is the height of most of the fathers and what of the sons ?

What is the height of the tallest Fathers and what of the tallest Son ?

Can you decide if the sons are growing taller than their fathers?

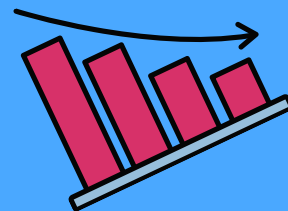
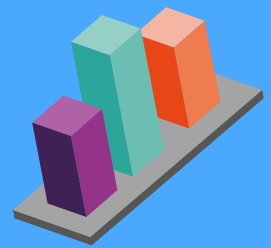
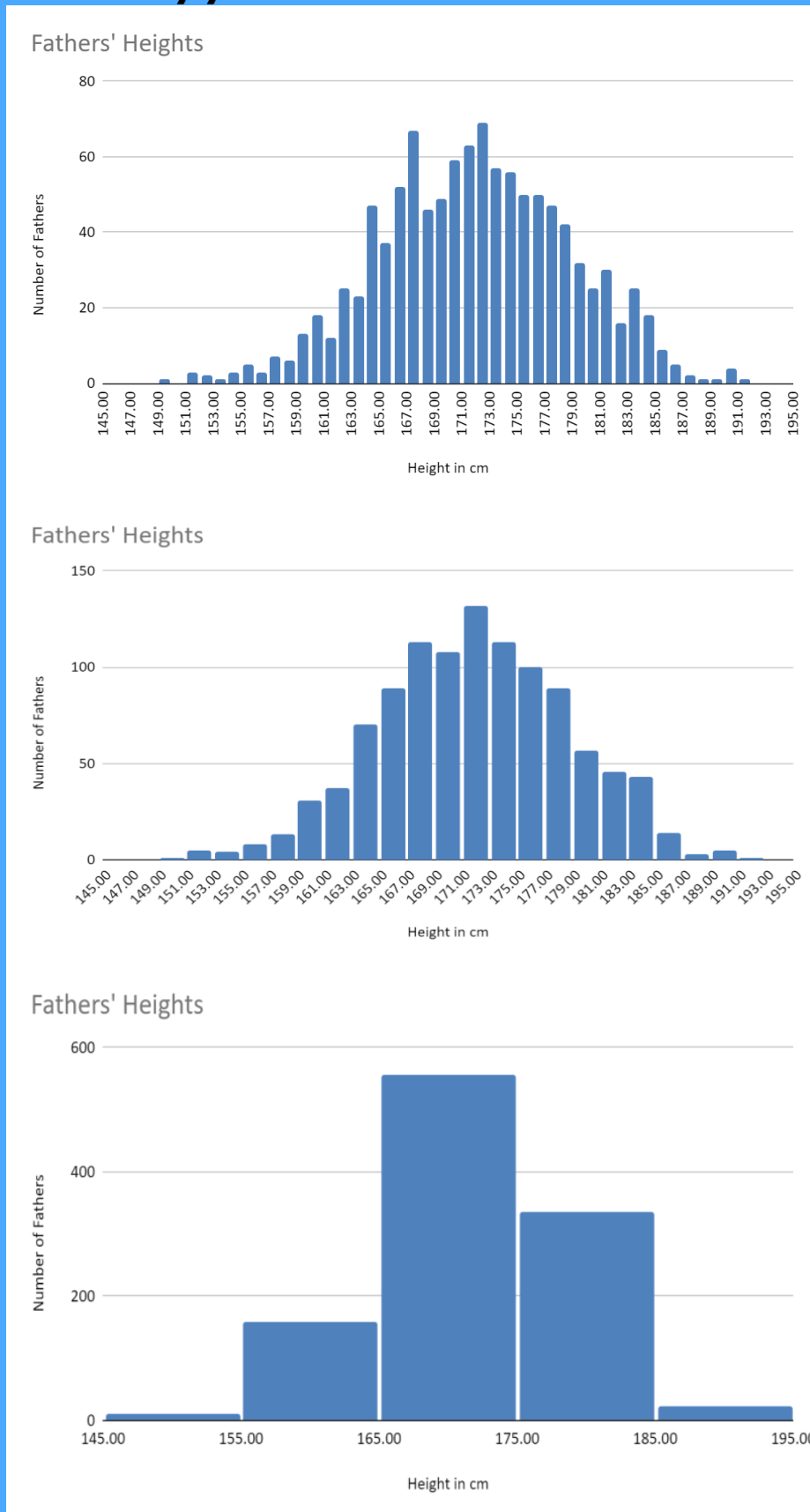
John couldn't make his mind about whether or not the Sons are in general taller than their fathers by comparing the two Histograms. Therefore he decided to put together the two Histograms in one diagram, like the one below.

Is it easier now to make the comparison? Can you draw conclusions? Explain your reason.



Comparing Histograms

1. Construct the histograms of the heights of Fathers with bucket 1, 2 and 10 .
2. John constructed the histograms of Fathers with bucket 1, 2 and 10 . Compare them. Which one do you think that is describing the distribution of Fathers' heights better ? Justify your answer.



Homework

Use the Appendix 2 in order to construct the histograms of the heights of Sons with bucket 1, 2 and 10 .