University of York

Department of Health Sciences

Applied Biostatistics

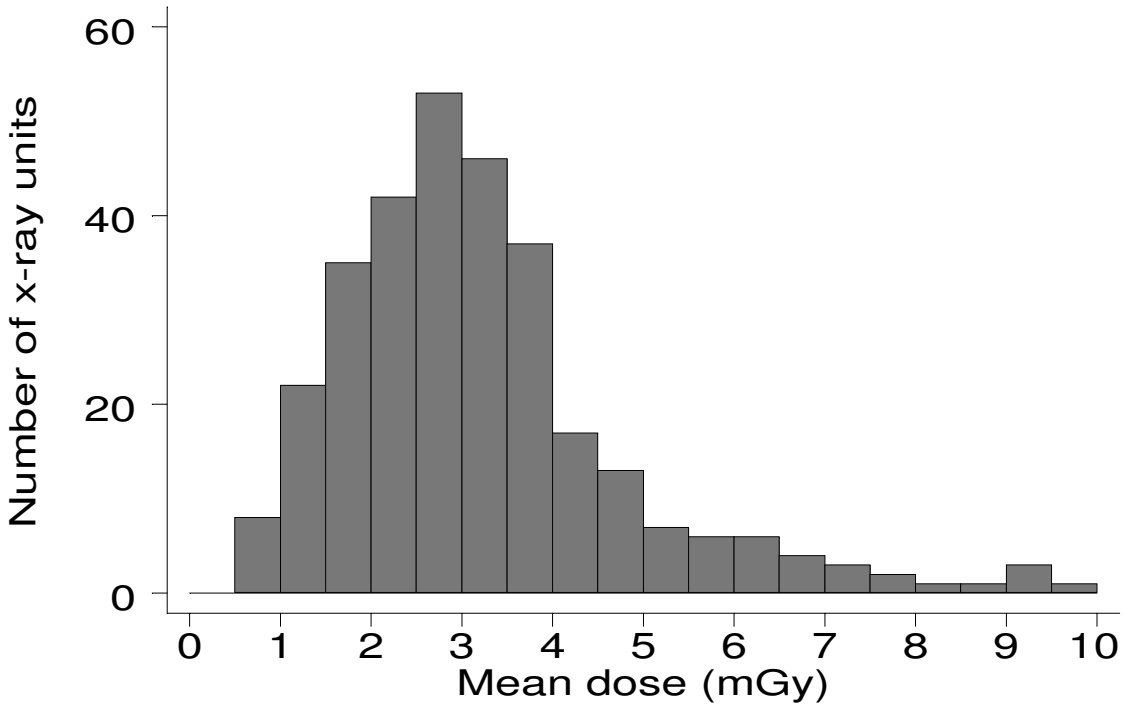
Suggested answers to Exercise: Summarising data

Question 1

(a) *The authors report that the does had 'median of 2.43 mGy, and the 75th percentile was 3.37 mGy.' What does this statement mean? What is the relationship between the third quartile and the 75th percentile?* The median is a number so that half the doses are less than it and half greater. If there is an odd number of observations it will be the middle observation if we arrange them in ascending order, if odd it will be the average of the middle pair. Thus half the doses were less than 550 μGy. The third quartile is the value so that three quarters of the observations are below it, hence 3/4 of the observations are below 660 μGy. The third quartile and the 75th percentile are two names for the same thing.

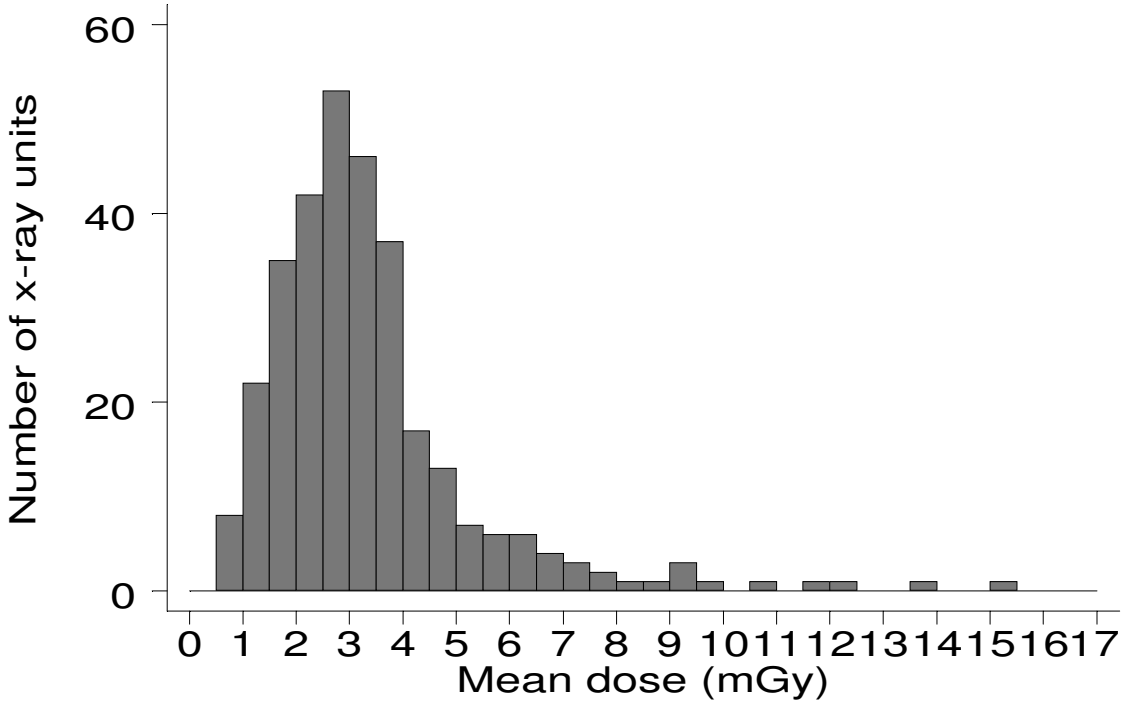
(b) *Figure 2 is described as a 'Frequency plot'. What other name is usually given to such a diagram? How could the presentation of this graph be improved?* This 'frequency plot' is usually called a histogram. The presentation could be improved by joining the bars together. As it is, the graph implies that there can be no observations between 2 and 2.5, etc. But this is not true; these are mean doses and so could take any value within the range. The three-dimensional effect contributes nothing and makes the graph more difficult to read.

The graph could be shown like this:



(c) *What term would be used to describe the shape of this frequency distribution?* The distribution is positively skew (skew to the right), because the long tail is on the right.

(d) *Five points have been omitted. What would be the effect on the distribution of including them?* The missing points are somewhere to the right of this figure. They might look something like this:



The distribution would appear more skew, because we have lengthened the long tail.

Question 2

(a) *What is wrong with this statement?* The three quartiles are the values which divide the distribution into four equal parts. They are points, not groups of people. They therefore do not have averages. The word is sometimes incorrectly used to mean one of the four groups into which the three quartiles divide the population, more correctly called `fourths'. In this sense there is no `middle quartile' because there are four of them and each fourth contains 25% of the population. (The newspaper in question is *The Guardian*, notorious for misprints, so this may not be quite what the author wrote.)

(b) *What is the more usual name for the `middle quartile'?* The middle quartile is the median.

Question 3

(a) *Which Honourable Member is correct, if any, and why?* Mr Lloyd is correct that the mean is not the same as the median but his definition of the median is wrong. It is the middle number when the numbers are arranged in ascending order and not the difference between the minimum and maximum values. So using his first example (2, 2, 5, 6, 7), the median here would be 5 and not 3.5 as asserted. He is correct that altering the extreme values will not change the median but his argument is confused.

Mr Carrington is correct that the median is not the mid-point between the first number and the last but is wrong in his definition. His definition appears to be

referring to the mode of a distribution, i.e. the category which has the highest frequency, although his explanation is not very clear. His definition of the average is wrong (the sample multiplied by the number of items). He is wrong in stating that the median will change if the number at the bottom of the scale changes. This will not change the median but will alter the mean.

(b) *What would be the effect on the skewness of the earnings distribution if the minimum wage were made a fixed proportion of the median, assuming that this figure was then higher than the current wage of some members of the population?* The distribution would become more positively skew. The smallest wages would be increased and the short, left-hand tail of the distribution would be made even shorter.