STAB22 TUT21

Chong Chen
University of Toronto, Scarborough
Department of Computer and Mathematical Sciences

February 1, 2019

1 Recall

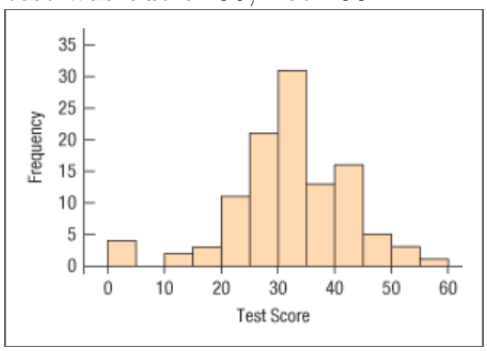
Quantitative variable: a variable that takes numerical values for which arithmetic operations such as adding and averaging make sense.

2 Histogram

Histogram: slice up all the possible values of the variable into bins and represent the counts that fall into the bins as bars.

2.1 Example

Below is a histogram of 110 students' test scores on a one-hour test in STA220 at the University of Toronto during summer 2007. The test was out of 50, not 100.



(a). Approximately what percentage of students got A grades? An A was 40 or higher, since the test was out of 50 (some scored above 50 due to bonus marks).

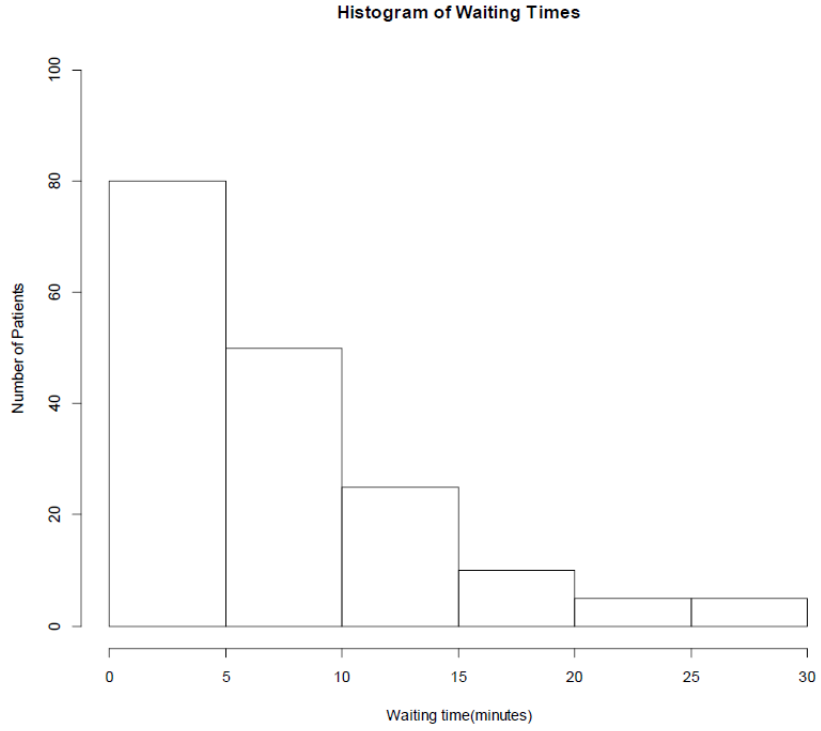
(b). What percent got C or D grades? (i.e., between 25 and 35 in the graph since any scores of 35 went into the higher bin)

(c). Write a brief description of this distribution (shape, centre, spread, and unusual features). Can you account for any of the features you see here?

2.2 Exercise

2.2.1

The histogram below displays the frequencies of waiting times, in minutes, for 175 patients in a doctor's office. To simplify, you may assume that no waiting times are exactly equal to one of the bin boundaries.



Use this information for this question and the next one.

Based on this information, which of the following statements is/are correct?
(I) Approximately 80% of the patients in this doctor's office waited less then 5 minutes.
(II) The mean waiting time is smaller than the median waiting time.
(III) The median waiting time is between 5 and 10 minutes.

(a) Only statement (I) is correct.
(b) Only statement (II) is correct.
(c) Only statement (III) is correct.
(d) Only statements (II) and (III) are correct.
(e) None of the three statements are correct.

2.2.2

Use the information in the previous question to answer this question.
The first quartile of waiting times is
(a) smaller than 5 minutes.
(b) between 5 and 10 minutes.
(c) between 10 and 15 minutes.
(d) between 15 and 20 minutes.
(e) larger than 20 minutes.

3 Five-number summary

The five-number summary of a set of observations consists of the smallest observation, the first quartile, the median, the third quartile, and the largest observation, written in order from smallest to largest. In symbols, the five-number summary is:

Minimum, First quartile, Median, Third quartile, Maximum

These five numbers give a reasonably complete description of both the center and the spread of the distribution.

3.1 Example

Here are costs of 10 electric smoothtop ranges rated very good or excellent by Consumer Reports on their website www.consumerreports.org.
850, 900, 1400, 1200, 1050, 1000, 750, 1250, 1050, 565

Find these statistics:

(a). Mean

(b). Median and quartiles

(c). Range and IQR

3.2 Boxplot

3.2.1 Exercise

A random sample generated from StatCrunch produced the following summary statistics:

N	Min.	Q1	Median	Mean	Q3	Max.
200	21.02	34.08	38.56	39.49	44.46	66.58

Based on these statistics and the 1.5*IQR rule for outliers, what can be stated about the number of outliers in this data set?

(a) There are no outliers in this data set.
(b) There is exactly one outlier in this data set.
(c) There is at least one outlier (i.e. one or more outliers) in this data set.
(d) There are no more than two outliers in this data set.
(e) There are no more than three outliers in this data set.