STAB22 TUT21

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

February 8, 2019

1 Last time

1.1 The Five-Number Summary

The minimum, The first quartile, Median, The third quartile, The maximum.

1.2 Boxplot

A boxplot displays the information in the five-number summary with the following elements:

1. Scale on left from the minimum to the maximum
2. Draw box from Q_1 (bottom) to Q_3 (top)
3. Draw horizontal line inside box at median
4. Upper fence: $Q_3 + 1.5 * IQR$ Lower fence: $Q_1 - 1.5 * IQR$
5. Draw lines connecting box to most extreme value within fences
6. Plot values outside fences individually. These are suspected outliers and deserve to be investigated.

1.3 Example

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.

2 Mean

Suppose you are given data $x_1, x_2, x_3, ..., x_n$, the mean is:

$$\bar{x} = \sum_{i=1}^{n} x_i$$

3 Standard deviation

Suppose you are given data $x_1, x_2, x_3, ..., x_n$, with the mean is $\bar{x}$, and the standard deviation is:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

4 Normal distribution

4.1 Example

Assume the cholesterol levels of adult women can be described by a Normal model with a mean of 188 mg/dL and a standard deviation of 24.

a. Draw and label the Normal model.

b. What percent of adult women do you expect to have cholesterol levels over 200 mg/dL?

c. What percent of adult women do you expect to have cholesterol levels between 150 and 170 mg/dL?

d. Estimate the interquartile range of the cholesterol levels.

e. Above what value are the highest 15% of women's cholesterol levels?

4.2 Exercises

4.2.1

A smelt is a type of food fish. Smelt lengths are normally distributed with mean 15 cm and standard deviation 1 cm. How long are the longest 10 percent of

smelts?

(a) bigger than 10.14 cm
(b) bigger than 16.28 cm
(c) 10.14 cm
(d) less than 16.28 cm
(e) 16.28 cm

4.2.2

Assume that the distribution of IQ scores of a group of students is normally distributed. The summary statistics of the IQ scores of this group of students are given below:

Column	n	Mean	Std. Dev.	Min	Q1	Median	Q3	Maximum
IQ scores	78	110.00	13.00	72.00	103.00	110.00	118.00	136.00

Use this information for this question and the following two questions. What percent of students in this group have an IQ score of 97 or greater?

(a) 2.5
(b) 16
(c) 68
(d) 84
(e) 95

4.2.3

Same information as above.
How high must an IQ score be in order to place in the top 1% of scores?

(a) 80 and higher
(b) 93 and higher
(c) 126 and higher
(d) 136 and higher
(e) 140 and higher

4.2.4

We transform the IQ scores by multiplying each student's IQ score by 0.9 and then adding 5. (Example: if a student has 80 before transformation, his score after the transformation is $(0.9 * 80) + 5$). Calculate the value of the IQR (interquartile range) of the IQ scores after this transformation.

(a) 13.5

(b) 15
(c) 18.5
(d) 20
(e) 64

4.2.5

The blood cholesterol levels for young women (aged 20 to 34) follow a normal distribution with mean 185 milligrams per deciliter (mg/dl) and standard deviation 39 mg/dl.
Complete the following sentence: 95% of young women have blood cholesterol level between

(a) 68 and 302
(b) 107 and 263
(c) 117 and 253
(d) 146 and 224
(e) 165 and 205