

Descriptive Statistics

Name: _____

Date: _____

Concepts

- Types of Variables and Frequency Tables (theory)
- Histograms and the Shape of a Distribution (theory)
- Mean, Median and Mode
- Variance and Standard Deviation
- Quartiles, Boxplots and Outliers

Types of Variables and Frequency Tables

1. A beekeeper records five variables for each hive in an apiary: the bee subspecies; the colony's strength, rated *weak*, *moderate* or *strong*; the number of frames holding brood; the mass of honey harvested (kg); and the hive's inventory number, painted on the lid.

- a) Classify each variable as categorical nominal, categorical ordinal, quantitative discrete or quantitative continuous.
- b) For 20 hives the numbers of brood frames were:

Brood frames	3	4	5	6	7
Number of hives	2	3	6	5	4

Add a relative-frequency row and a cumulative relative-frequency row to the table.

- c) What proportion of the hives have at most 5 brood frames? At least 6?

Histograms and the Shape of a Distribution

2. A regional airline records the percentage of seats filled on 50 flights. Each class includes its left endpoint; the last class also includes 100.

Seats filled (%)	[50, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100]
Flights	2	4	8	15	21

- a) Give the relative frequency and the cumulative relative frequency of each class.
- b) Describe the shape of the histogram of these data: its modal class, its symmetry or skew, and the direction of any tail.
- c) In which class does the median lie? What proportion of flights were at least 80% full?

Mean, Median and Mode

3. A customer-support agent logs how many emails she answers in each of 10 one-hour slots: 14, 9, 17, 12, 9, 20, 11, 9, 15, 14.

- a) Find the mean, the median and the mode.
- b) In an eleventh slot she answers 35 emails. Find the new mean and the new median, and say which of the two moved more and why.

4. A quality inspector counts the defective bolts in each of 40 boxes.

Defective bolts	0	1	2	3	4
Boxes	12	14	8	4	2

Find the mean, the median and the mode of the number of defective bolts per box.

Variance and Standard Deviation

5. A library counts the books returned through its outdoor drop box on 8 days: 12, 15, 9, 14, 10, 13, 11, 12. Treating these days as a sample, find the sample variance $s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$ and the sample standard deviation s .

6. For 25 calls to a software help line, the lengths x (in minutes) give $\sum x = 500$ and $\sum x^2 = 10\,384$.

- a) Find $\bar{x}$, and find s using $s^2 = \frac{\sum x^2 - (\sum x)^2/n}{n - 1}$.
- b) Each call is billed \$4.00 plus \$1.50 per minute. Find the mean and the standard deviation of the billed amounts.

Quartiles, Boxplots and Outliers

7. Use this convention: order the data and split them into a lower and an upper half, leaving out the median itself when n is odd; Q_1 is the median of the lower half and Q_3 the median of the upper half. A value is an outlier when it lies more than $1.5 \times \text{IQR}$ below Q_1 or above Q_3 .

The masses (kg) of 12 checked suitcases on one flight were 14, 21, 18, 23, 8, 19, 22, 17, 20, 16, 31, 18.

- a) Give the five-number summary and the IQR.
- b) Find the fences and identify any outliers.
- c) Where do the whiskers of the modified boxplot end?

8. Use the convention that a value is an outlier when it lies *more than* $1.5 \times \text{IQR}$ below Q_1 or above Q_3 . The lengths of the 60 episodes of a podcast have five-number summary (in minutes) minimum 2, $Q_1 = 14$, median 18, $Q_3 = 22$, maximum 41. The five shortest episodes last 2, 9, 10, 11, 12 minutes and the five longest 30, 31, 33, 35, 41 minutes.

- a) Find the fences and list every outlier.
- b) Describe the modified boxplot: where the box and the whiskers end.
- c) About what fraction of the episodes last between 14 and 22 minutes?

 Synthesis — drawing on several topics in this unit

9. A traffic survey records the number of occupants in 20 cars crossing a bridge:

Occupants	1	2	3	4	5
Cars	8	6	3	2	1

- Classify the variable.
- Find the mean and the median number of occupants.
- Find the sample standard deviation s , using $s^2 = \frac{\sum fx^2 - (\sum fx)^2/n}{n-1}$.
- Describe the shape of the distribution and check that the relation between the mean and the median agrees with it.

Want the harder problems and the worked solutions?

The *Challenge Problems* set and the *Answer Keys* for Descriptive Statistics — with the reasoning written out in full — are available at tutorinmontreal.ca.