# Lesson 19: Comparing populations with friends

## Goals

- Apply reasoning about centre and spread to determine whether two populations are likely to be meaningfully different, and explain (orally and in writing) the reasoning.

- Coordinate (orally) visual displays of data with descriptions of shape, measures of centre, and measures of spread.

- Determine what information is needed to solve problems about using samples to compare populations. Ask questions to elicit that information.

## Learning Targets

- I can decide what information I need to know to be able to compare two populations based on a sample from each.

## Lesson Narrative

Students continue to practise comparing populations by using samples from each population. Outside of the classroom, people who wish to compare groups will not usually have all of the useful statistics presented to them in a nice package, so they will need to determine what information to gather and then work through the comparison process. In this lesson, students are paired so that one student is presented with a situation and question while the other student has information to help solve the question. They must work together to answer the question by asking their own questions and explaining how each piece of information will be useful. An optional activity is also included in which students are asked to compare data from a sample of one population to statistics from a sample of a second population.

### Building On

- Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

### Addressing

- Use measures of centre and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. For example, decide whether the words in a chapter of a year 8 science book are generally longer than the words in a chapter of a year 5 science book.

### Instructional Routines

- Information Gap Cards

- Discussion Supports

### Required Materials

**Pre-printed slips, cut from copies of the blackline master**

| | |
|---|---|
| Info Gap: Comparing populations<br>**Problem Card 1**<br><br>A chewing gum company records the amount of time, in minutes, it takes for a person to find that the gum has lost its flavour for two different flavours of gum. A sample of 20 people is used for mint and another sample of 20 people is used for cinnamon gum. Is there a meaningful difference in the amount of time for the different flavours? Explain your reasoning. | Info Gap: Comparing populations<br>**Data Card 1**<br><br>• The distributions are not symmetrical<br>• Mean for mint: 74.5 minutes<br>• Median for mint: 65 minutes<br>• IQR for mint: 20 minutes<br>• Range for mint: 68 minutes<br>• Mean for cinnamon: 108.75 minutes<br>• Median for cinnamon: 112.5 minutes<br>• IQR for cinnamon: 22.5 minutes<br>• Range for cinnamon: 71 minutes |
| Info Gap: Comparing populations<br>**Problem Card 2**<br><br>A year 9 English teacher is interested in grammar scores for students coming from the two year 8 English teachers. Is there a meaningful difference in the means for student scores on a pretest coming from each year 8 teacher based on the samples?<br>Explain your reasoning. | Info Gap: Comparing populations<br>**Data Card 2**<br><br>Scores from Teacher A<br><br>| 70 | 75 | 80 | 80 | 85 | 85 | 85 | 90 | 95 | 100 |<br><br>Scores from Teacher B<br><br>| 70 | 70 | 70 | 75 | 75 | 80 | 80 | 80 | 90 | 100 | |

### Required Preparation

One copy of the blackline master from Comparing Populations cut into cards for every 2 students.

### Student Learning Goals

Let's ask important questions to compare groups.

## 19.1 Features of Graphic Representations

**Warm Up: 5 minutes**

In this warm-up, students review the useful information that can be gained from different graphical representations of data in preparation for comparing groups based on samples from each.

### Launch

For classes that may need help remembering the different representations, consider displaying an example of each type of graphical representation mentioned.

**Student Task Statement**

Dot plots, histograms, and box plots are different ways to represent a data set graphically.

Which of those displays would be the easiest to use to find each feature of the data?

1.   the mean

2.   the median

3.   the range

4.   the interquartile range

5.   the symmetry

**Student Response**

1.   Dot plot

2.   Box plot

3.   Dot plot

4.   Box plot

5.   Dot plot or histogram

**Activity Synthesis**

Poll the class for their answers to each of the problems. Select at least one student to share their reasoning for each question.

# 19.2 Info Gap: Comparing Populations

**30 minutes**

In this info gap activity, students work together to compare two populations from information about samples from each of the populations. Students must pay attention to the information they need in order to solve the problem and the types of question they could ask to get to the answer.

The info gap structure requires students to make sense of problems by determining what information is necessary, and then to ask for information they need to solve it. This may take several rounds of discussion if their first requests do not yield the information they need. It also allows them to refine the language they use and ask increasingly more precise questions until they get the information they need.

**Instructional Routines**
•   Information Gap Cards

## Launch

Arrange students in groups of 2. In each group, distribute a problem card to one student and a data card to the other student. After you review their work on the first problem, give them the cards for a second problem and instruct them to switch roles.

Tell students they will continue to work with comparing measures of centre for populations. Explain the Info Gap structure and consider demonstrating the protocol if students are unfamiliar with it. There are step-by-step instructions in the student task statement.

*Engagement: Develop Effort and Persistence.* Display or provide students with a physical copy of the written directions. Check for understanding by inviting students to rephrase directions in their own words. Keep the display of directions visible throughout the activity.
*Supports accessibility for: Memory; Organisation Conversing*: This activity uses *Information Gap* to give students a purpose for discussing information necessary to use data from samples to compare two populations. Display questions or question starters for students who need a starting point such as: "Can you tell me . . . (specific piece of information)", and "Why do you need to know . . . (that piece of information)?"
*Design Principle(s): Cultivate Conversation*

## Student Task Statement

Your teacher will give you either a *problem card* or a *data card*. Do not show or read your card to your partner.

If your teacher gives you the *problem card*:

1.  Silently read your card and think about what information you need to be able to answer the question.

2.  Ask your partner for the specific information that you need.

3.  Explain how you are using the information to solve the problem.

    Continue to ask questions until you have enough information to solve the problem.

4.  Share the *problem card* and solve the problem independently.

5.  Read the *data card* and discuss your reasoning.

If your teacher gives you the *data card*:

1.  Silently read your card.

2.  Ask your partner *"What specific information do you need?"* and wait for them to *ask* for information.

If your partner asks for information that is not on the card, do not do the calculations for them. Tell them you don't have that information.

3. Before sharing the information, ask "*Why do you need that information?*" Listen to your partner's reasoning and ask clarifying questions.

4. Read the *problem card* and solve the problem independently.

5. Share the *data card* and discuss your reasoning.

Pause here so your teacher can review your work. Ask your teacher for a new set of cards and repeat the activity, trading roles with your partner.

**Student Response**
1. There is a meaningful difference. Since the data is not symmetrical, it makes sense to use the median and IQR to compare the data. The medians are more than 2 IQRs apart ($112.5 - 65 > 2 \times 22.5$), so there is a meaningful difference between the flavours.

2. No. The mean for teacher A is 83 and range is 25. The mean for teacher B is 79 and the range is 30. So, the difference in means is small in comparison with the ranges.

**Are You Ready for More?**

Is there a meaningful difference between top sports performance in two different decades? Choose a variable from your favourite sport (for example, home runs in baseball, kills in volleyball, aces in tennis, saves in football, etc.) and compare the leaders for each year of two different decades. Is the performance in one decade meaningfully different from the other?

**Student Response**

Answers vary.

**Activity Synthesis**

The purpose of the discussion is to help students understand the types of questions they need to answer in order to compare groups.

Select several groups to share their answers and reasoning for each of the problems. In particular help students understand why the information "The distributions are not symmetrical" was important for solving the first problem.

Consider asking these discussion questions:

• "What was the most important question you asked for the first problem? For the second problem?"

• "What are some other ways the information could have been given to solve the problems?" (Instead of the characteristics for the first question, a box plot could have

been presented. The second question could have had a dot plot or characteristics like the first problem.)

- "If the distributions for the first problem had been symmetrical, would the answer have been the same?" (Yes. The difference in means is 34.25 which is small in comparison with the ranges, so there would not have been enough information to say that the two population means are meaningfully different.)

## 19.3 Comparing to Known Characteristics

**Optional: 15 minutes**

In this optional activity, students compare two populations using samples again, but this time only one sample is given. For the other sample, the characteristics have either been calculated already or are the focus of the question. This type of analysis is useful when comparing two similar populations as in this activity or when comparing a group against a standard.

### Instructional Routines
- Discussion Supports

### Launch

Keep students in groups of 2.

Tell students that sometimes it is useful to compare one group to a standard or another group where the important characteristics have already been computed. In these problems, a random sample from one group is given and characteristics of the second group is either given or sought.

Allow students 10 minutes partner work time followed by a whole-class discussion.

### Student Task Statement
1. A university graduate is considering two different companies to apply to for a job. Acme Corp lists this sample of salaries on their website:

    £45 000

    £55 000

    £140 000

    £70 000

    £60 000

    £50 000

    What typical salary would Summit Systems need to have to be meaningfully different from Acme Corp? Explain your reasoning.

2. A factory manager is wondering whether they should upgrade their equipment. The manager keeps track of how many faulty products are created each day for a week.

6      7      8      6      7      5      7

The new equipment guarantees an average of 4 or fewer faulty products per day. Is there a meaningful difference between the new and old equipment? Explain your reasoning.

## Student Response

1. A median salary greater than £97 500 (or less than £17 500). Since the salaries for Acme Corp have a large value far from the other values, the measure of centre chosen should be the median. The median salary for Acme Corp is £57 500 and the IQR is £20 000, so the median salary for Summit Systems must be greater than 2 IQRs above or less than 2 IQRs below Acme's median ($57\,500 + 2 \times 20\,000 = 97\,500$ or $57\,500 - 2 \times 20\,000 = 17\,500$).

2. Yes. The mean for the sample for the current machine is 6.57 faulty products per day and the range is 3 faulty products per day. The difference in means is 2.57 faulty products per day (since $6.57 - 4 = 2.57$) which is large compared with the range, so there is a meaningful difference in the mean number of faulty products per day.

## Activity Synthesis

The purpose of the discussion is to help students understand how to compare groups when one set of characteristics are known and the other group is represented by sample data.

Select some groups to share their answers and reasoning for the two problems.

Consider asking these discussion questions:

- "How did you determine what to use as a typical value for the first problem?" (Since there was one value much greater than the others, the distribution would not be symmetrical, so median is a more appropriate measure of centre.)

- "How did you determine what measure of centre to use for the second problem?" (Since the data were all close, either value could be used, but the new equipment reported the "average" or mean, so man should be used for the sample as well.)

- "The manufacturer for the new equipment guarantees 4 flaws *or fewer* per day with the new equipment. If the new equipment produces only 3 flaws per day does that change the answer for the second problem?" (No. There is an even greater difference between the current and new equipment, so it is even more meaningful.)

- "What other factors would the university graduate want to consider other than the meaningful difference in median salary between the two companies?" (In addition to the other factors for a job such a benefits, relationship with co-workers, type of work

being done at each company, etc., the graduate should consider the salary for the type of job he will get at the company. For example, if his degree is in computer science, he may be looking at a job with computers rather than sales or some other department within the company, so he might be able to get a better comparison of salaries that way.)

• "What other factors would the factory manager want to consider other than the meaningful difference in flaws for the equipment?" (The cost of the frequent flaws as well as the cost of the new equipment will probably factor into her decision to buy new equipment. The age of the current equipment and maintenance for older equipment compared to new equipment may also be important.)

*Representation: Internalise Comprehension.* Use colour and annotations to illustrate connections between representations. As students share their diagrams and reasoning, use colour and annotations to scribe their thinking on a display of each problem so that it is visible for all students.
*Supports accessibility for: Visual-spatial processing; Conceptual processing Speaking: Discussion Supports.* Use this routine to support whole-class discussion. For each response or observation that is shared, ask students to restate and/or revoice what they heard using precise mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker whether their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped clarify the original statement. This will provide more students with an opportunity to produce language as they interpret the reasoning of others.
*Design Principle(s): Support sense-making*

## Lesson Synthesis

Ask students what information is important to collect when attempting to compare large groups and why each of these pieces of information is useful. Ask students if they can think of other situations in which it might be helpful to compare two large groups by generating a sample and collecting information.

# 19.4 A Different Box Plot

**Cool Down: 5 minutes**

This cool-down assesses whether students understand how different groups can be compared based on data from a box plot. In particular, students need to be able to read the median and IQR from the box plot and use that information to construct an additional box plot that would be meaningfully different.

### Student Task Statement

Use the box plot to answer the questions.
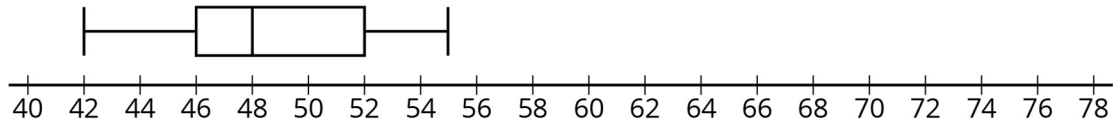
1.  What measure of centre is shown in the box plot? What measure of variability? What are the values for each of these characteristics?

2.  Draw another box plot with the same measure of variability that is meaningfully different from the one shown.

**Student Response**

1.  Median = 48. IQR = 6

2.  Answers vary. Correct responses show a box plot with an IQR of 6 and median greater than or equal to 60 (or less than or equal to 36).

## Student Lesson Summary

When using samples to comparing two populations, there are a lot of factors to consider.

*   Are the samples representative of their populations? If the sample is biased, then it may not have the same centre and variability as the population.

*   Which characteristic of the populations makes sense to compare—the mean, the median, or a proportion?

*   How variable is the data? If the data is very spread out, it can be more difficult to make conclusions with certainty.

Knowing the correct questions to ask when trying to compare groups is important to correctly interpret the results.

## Lesson 19 Practice Problems

### Problem 1 Statement

An agent at an advertising agency asks a random sample of people how many episodes of a TV show they watch each day. The results are shown in the dot plot.
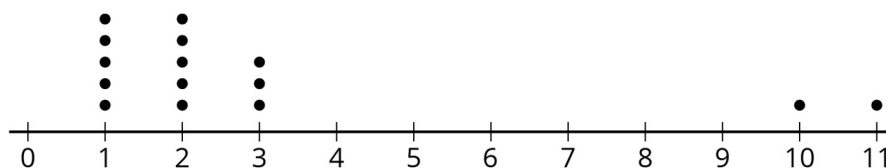


The agency currently advertises on a different show, but wants to change to this one as long as the typical number of episodes is not meaningfully less.

a.  What measure of centre and measure of variation would the agent need to find for their current show to determine if there is a meaningful difference? Explain your reasoning.

b.  What are the values for these same characteristics for the data in the dot plot?

c.  What numbers for these characteristics would be meaningfully different if the measure of variability for the current show is similar? Explain your reasoning.

**Solution**

a.  Median and IQR, since the dot plot shows a distribution that is not symmetrical, so it would make sense to compare medians.

b.  Median: 2 episodes; IQR: 2 episodes

c.  The other show would need to have a median of at least 6 episodes, since the medians would need to be at least 2 IQRs apart and $2 + 2 \times 2 = 6$.

**Problem 2 Statement**

Jada wants to know if there is a meaningful difference in the mean number of friends on social media for teens and adults. She looks at the friend count for the 10 most popular of her friends and the friend count for 10 of her parents' friends. She then computes the mean and range of each sample and determines there is a meaningful difference.

Jada's dad later tells her he thinks she has not come to the right conclusion. Jada checks her calculations and everything is right. Do you agree with her dad? Explain your reasoning.

**Solution**

Yes. She did not select her samples randomly, so they may not be representative of teens and adults.

**Problem 3 Statement**

The mean weight for a sample of a certain kind of ring made from platinum is 8.21 grams. The mean weight for a sample of a certain kind of ring made from gold is 8.61 grams. Is there a meaningful difference in the weights of the two types of rings? Explain your reasoning.

**Solution**

The answer is unknown with this information.

**Problem 4 Statement**

The lengths in feet of a random sample of 20 male and 20 female humpback whales were measured and used to create the box plot.

length in feet
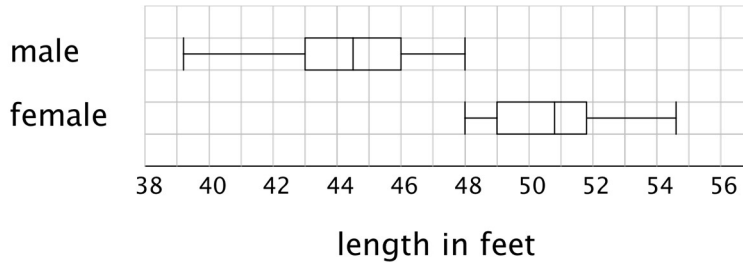
Estimate the median lengths of male and female humpback whales based on these samples.

**Solution**

Males: 44.6 feet. Females: 50.9 feet.