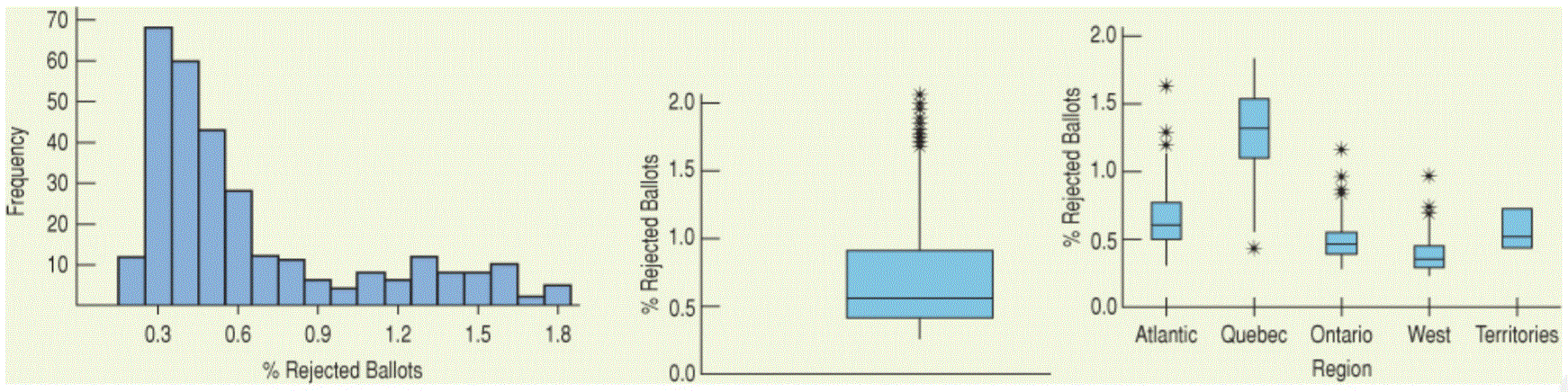
1. Spoiled ballots are a real threat to democracy. Below are displays of data from Elections Canada showing percentage of spoiled ballots (number of spoiled ballots divided by total number of ballots) for all 308 electoral ridings in the 2011 Canadian federal election, and another display comparing results by region.



1. Describe what the histogram says about spoiled ballots.

There is a large amount of ridings that between 0.3% and 0.6% of the ballots are spoiled. The histogram is unimodal, and very skewed to the right. The Histogram doesn’t appear to have any outliers.

1. What does the histogram suggest that you can’t see in the boxplot?

The histogram shows how many ballots are rejected per riding, the boxplot only shows percentages, and not by how many ridings.

1. Are there many outliers present in the nationwide data? Explain.

There are 9 outliers in the nationwide data. There is so much data clustered around the beginning of the dataset that the 9 largest percentages of rejected ballots are beyond the IQR.

1. Describe the regional variation in spoiled ballots.

The regions vary greatly in terms of amount of rejected ballots. Quebec on average has the most rejected ballots per riding, while the western provinces have on average a much lower percentage, however they still have a riding that has more spoiled ballots than most of the ridings in the territories.

1. Why is one of the regional boxplots missing whiskers?

Because the dataset for those ridings is so small that the IQR happens to include the min and the max

1. Which region had the worst result? The best result?

Quebec had the worst result, the most spoiled ballots. The western provinces had the best outcome with on average less than .5% of ballots being spoiled.

1. How does the last display above clarify what is seen in the histogram?

Even though it doesn’t show numbers, each of the boxes represents a number of different ridings, and if you ignore everything except the boxes, these ridings follow a similar shape as the histogram in the first image.

1. Your statistics teacher has announced that the lower of your two tests will be dropped. You got a 90 on test 1 and an 80 on test 2. You’re all set to drop the 80 until she announces that she grades “on a curve.” She standardized the scores in order to decide which is the lower one. If the mean on the first test was 88 with a standard deviation of 4 and the mean on the second was 75 with a standard deviation of 5
2. Which one will be dropped?

The first test will be the one that is dropped. If you scored 90% on the first test, then you scored only half a standard deviation about the mean of the scores. However, if you scored an 80% on the second test, then you scored a full standard deviation about the class average, which grading on a curve, would be higher.

1. Does this seem “fair”?

It depends on a lot of factors. This would increase an R score by putting you up above the rest of the class, but it would lower a final grade. This is assuming that the tests are weighted evenly, and that all other tests would simply absorb the grade of the lowest test.

1. SAT test scores are required of applicants to many U.S. universities. In 1995, the Educational Testing Services (ETS) adjusted the scores of SAT tests. Before ETS re-centered the SAT Verbal test, the mean of all test scores was 450.
2. How would adding 50 points to each score affect the mean?

Adding 50 points to the scores of the SATs would increase the mean by 50 as well. When shifting the data, the mean will increase by the amount that you’ve shifted by.

1. The standard deviation was 100 points. What would the standard deviation be after adding 50 points?

If you added 50 points to every score for the SATs, then the standard deviation would not change. Adding 50 points to each score would create a shifting of the data, which does not change standard deviation.

1. Suppose we drew boxplots of test takers’ scores a year before and a year after the ‘+50’ re-centering. How would the boxplots of the two years differ?

Assuming that the classes before and after the ‘+50’ point shifting of the data were about the same strength of class, if you were to draw two boxplots representing both sets of scores, then you would likely see two very similarly shaped plots. However, all data points after the shift, would also be shifted accordingly. So each number in your 5 number summary would also increase by 50 points.

1. A company manufacture wheels for in-line skating. The diameter of the wheels has a mean of 3 inches and a standard deviation of 0.1 inches. Because so many of its customers use the metric system, the company decides to report their production statistics in millimeters (1 inch = 25.4 mm). They report that the standard deviation is now 2.54mm. A corporate executive is worried about this increase in variation. Should he be concerned? Explain.

No, the corporate executive should not be worried about this increase in the standard deviation of the diameter of the wheels. Going from the diameter of the wheel in inches to the diameter in millimeters is a scaling of the dataset since you’re multiplying the diameters by 2.54. Because of this, if the standard deviation is .1 inch, then with a scaling of data, the standard deviation will also be affected by this scaling, unlike with the shifting of data.

1. As a group, the Dutch are among the tallest people in the world. The average Dutch man is 184 cm tall (and the average Dutch woman is 170.8 cm tall). If a Normal model is appropriate and the standard deviation for men is about 8 cm, what percentage of all Dutch men will be over 2 meters tall?

Approximately 2.5% of men in the Netherlands will be over 2m tall.

1. Suppose it take you 20 minutes, on average, to drive to school, with a standard deviation of 2 minutes. Suppose a Normal model is appropriate for the distribution of driving times.
2. How often will you arrive at school in less than 22 minutes?

You will arrive at school in less than 22 minutes 84% of the time.

1. How often will it take you more than 24 minutes? Do you think the distribution of your driving times is unimodal and symmetric?

It will take you more than 24 minutes to arrive to school about 2.5% of the time. If you’re using assuming that the normal model is appropriate for the distribution, then the distribution will be unimodal and symmetric.

1. What does this say about the accuracy of your predictions? Explain.

Because the distribution can be assumed as a normal model, it means that predictions you make will be fairly accurate. The model is unimodal and symmetric, and therefore your data can be predicted due to fewer outliers in the set.