MA 151: Using Minitab to Visualize and Explore Data
The Low Fat vs. Low Carb Debate

September 5, 2018

1 Introduction to the Data

We will be working with a synthetic data set meant to resemble the weight loss outcomes reported in [1], the study referenced in the Vox article "Why do dieters succeed or fail? The answers have little to do with food." In this study, 609 overweight adults took part in a dietary weight loss intervention called the Diet Intervention Examining The Factors Interacting with Treatment Success (DIETFITS). The participants were randomized into one of two groups: a healthy low fat (HLF) diet or a healthy low carb (HLC) diet. As part of the intervention, the participants participated in 22 small group sessions that covered ways to incorporate the healthy diet into their lifestyle. Their weights were recorded before the intervention and at a 12-month follow up. We will be considering the change in their weights from baseline to followup,

$$\text{Weight Change} = (\text{Weight at Followup}) - (\text{Weight at Baseline}) \tag{1}$$

The weight changes are reported in pounds.

2 Loading Data

Download the data from the course website by clicking the Weight Loss Data link under Lecture 2. Move the file somewhere you can easily access it, like your desktop.

The data is stored in an MPJ file, which stands for Minitab Project file. This is a proprietary format of Minitab, so data stored in this format can only be opened by Minitab.

The Minitab Project file stores both the data itself, and all of the analyses that you perform on the data. The data are stored in three worksheets, titled `Weight Change Data Low Fat`, `Weight Change Data Low Carb`, and `Weight Change Data with Groups`. The analyses will be listed in the `Session` window as you perform them.

3 Sorting Data

Pull up the project manager, and select the `Weight Change Data Low Fat` worksheet from the left panel. The data are stored in a spreadsheet, much like you might have seen in Excel. The values are currently listed as they were recorded: as a participant completed the study, the change in their weight was entered into the spreadsheet. To get a handle on the data, a good first step is to sort the data set. To do this, right-click anywhere in the `C1` column, and select

> `Sort Columns` → `Entire Worksheet` → `Smallest to Largest`

The weight changes should now be sorted from smallest (most negative) to largest (most positive).

Hands On 1

What was the greatest weight loss in the low fat group? The greatest weight gain?

Sort the data for the low carb group. What was the greatest weight loss in the low carb group? The greatest weight gain?

4 Creating a Dotplot

It's hard to get a feeling for how the weight differences are distributed just by looking at the table of numbers. Bring up the `Weight Change Data Low Fat` worksheet. Generate a dotplot of the weight changes by selecting

> `Graph` → `Dotplot...`

from the menu bar. Choose `Simple` from the `One Y` category. Choose 'Weight Change Low Fat (lb)' as the variable to construct a dotplot from by double-clicking on 'Weight Change Low Fat (lb)' in the left-most panel. Then click `OK`. You should see a dotplot that looks like the following figure:

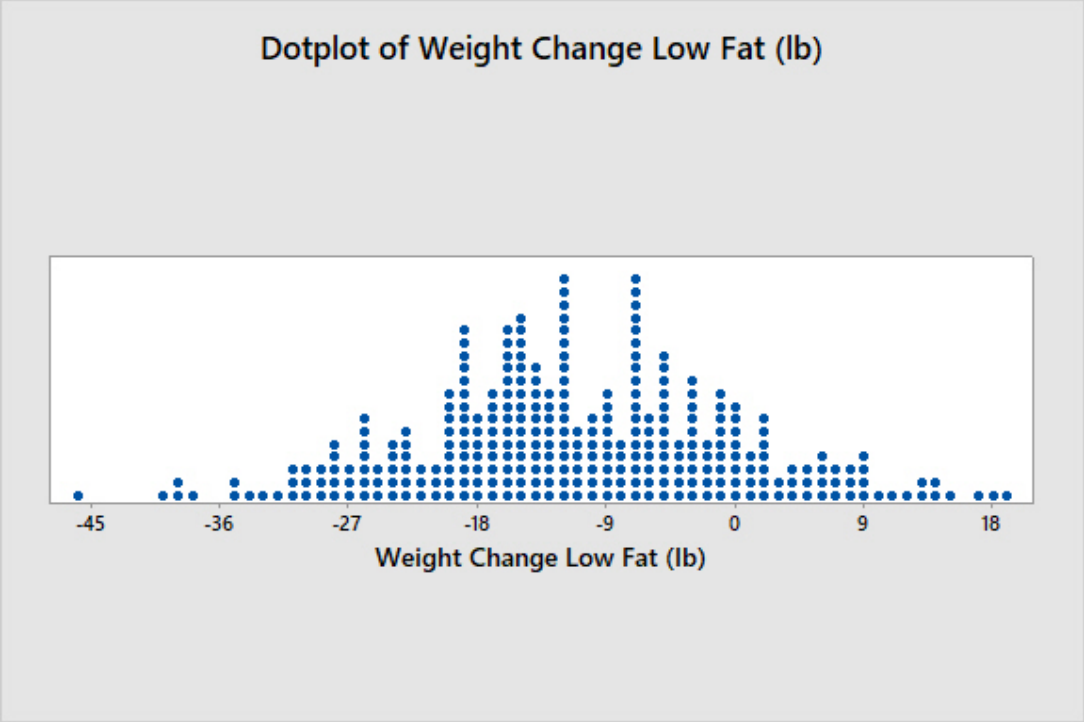

Figure 1:

To save the figure, select

`File` → `Save Graph As...`

from the menu bar. From the `Save as type:` dropdown select `JPEG 24 bit color (*.jpg)`. Then save the figure to somewhere you will be able to access it later.

Hands On 2

Each dot in the dotplot corresponds to a single data point. Hover over the left-most dot. What row in the table does this data point correspond to? Do the same for the right-most dot. What row in the table does this data point correspond to?

Remember: every dot tells a story.

Follow the same procedure to create a dotplot for the Low Carb group.

Hands On 3

How do the two dotplots compare? Are the horizontal axes lined up? The vertical axes? What impact does this have on your ability to compare the two data sets?

To compare the weight change in the two groups, we need to include an additional column indicating the group that the individual was assigned to. This has been done in the `Weight Change Data with Groups` Worksheet. Select this worksheet using either the `Project Manager` or by maximizing the worksheet from the bottom row of minimized windows. To plot each group together, select

`Graph` → `Dotplot...`

from the menu bar as before. Now, choose `With Groups` from the `One Y` category. Choose 'Weight Change (lb)' as the `Graph variables:` and Group as the Categorical variable[1]. You should see the dotplot below:

[1] 'Categorical variable' is another name for a qualitative / nominal variable, since the data come from a set of categories.

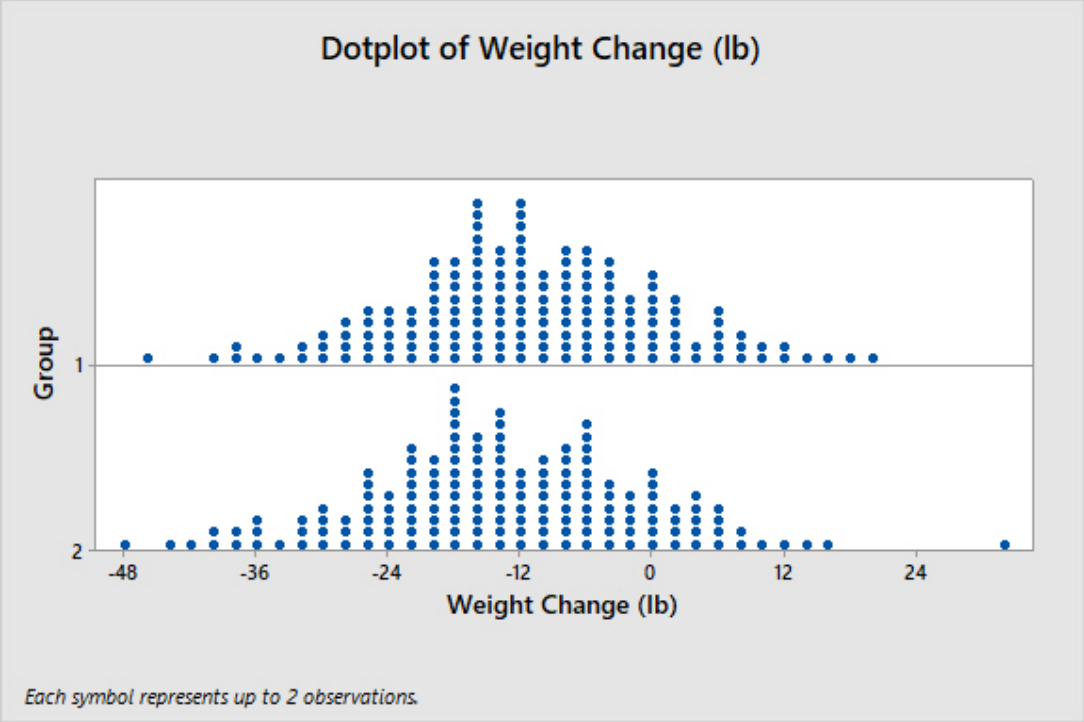

Figure 2:

Hands On 4

In the combined dotplot, which group corresponds to the low fat group? The low carb group? How can you tell?

Which group has more negative weight changes? **Reminder:** a negative weight change corresponds to weight *loss*.

Figure 3: Lorem ipsum.

5 Creating a Histogram

Dotplots are just one way to represent the frequency of measurements. They are largely used as a teaching tools (every dot is a data point), and you are not likely to see them in practice. A related representation that is more common is a histogram of the data. In the histogram, instead of representing each data point individually, we represent all data points within a given bin in terms of the height of the bar centered over that bin.

To create histograms of the data in the two groups, select

`Graph` → `Histogram...`

from the menu bar, and select the `With Groups` option. Select the `Graph variables:` and `Categorical variables` as you did with the grouped dotplots. You should see the following plot:

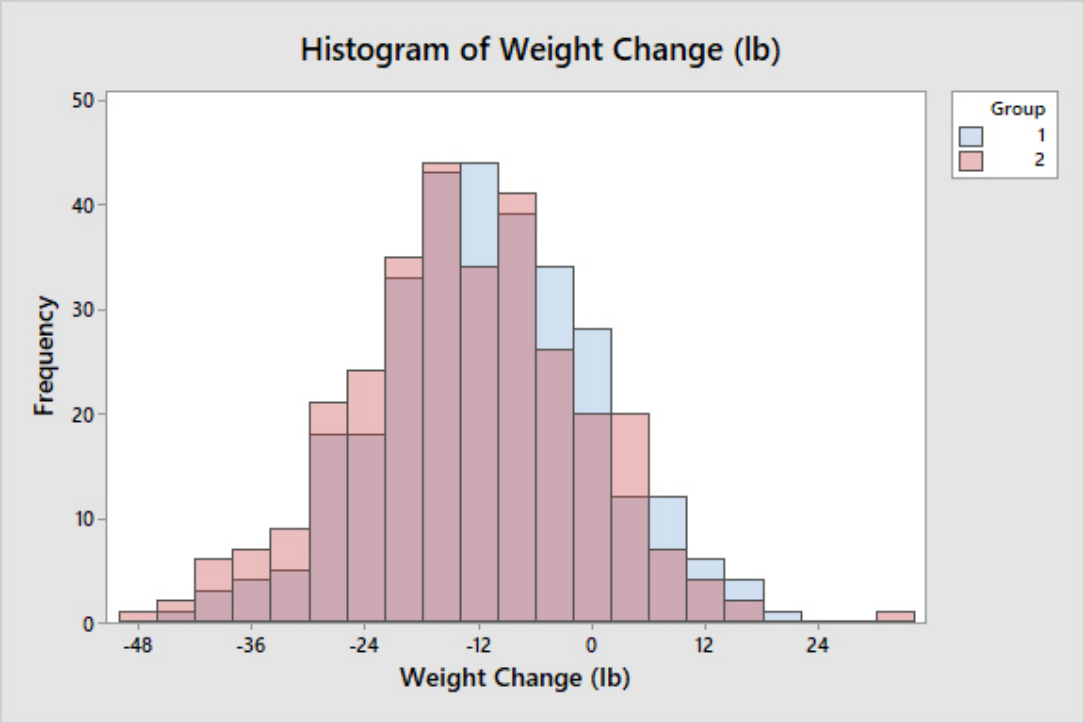

Figure 4:

Hands On 5

How do the grouped histograms compare to the grouped dotplots? Do they convey the same information?

You can modify the placement and number of bins by double clicking on one of the two histograms. In the `Edit Bars` dialog box, select the `Binning` tab. The options for the Interval Type are Midpoint and Cutpoint. The Midpoint option specifies the centers of the bins, and the width of the bins are determined by the number of bins. The Cutpoint option specifies the left and right endpoints for a bin.

Hands On 6

Switch from the Midpoint option to the Cutpoint option. What happens to the bins?

Change the `Interval Definition` to `Number of Intervals`, and try 2, 3, and 10 for the Number of intervals. What happens as you decrease the number of bins? Increase the number of bins?

You may have noticed that the histograms for both groups are relatively 'bell-shaped': there are many weight changes near the center of the histogram (indicated by the increasing height of the bars towards the center), and fewer and fewer weight changes at either extreme (indicated by the decreasing height of the bars). This type of data shape is commonly captured by a normal[2] curve. A normal curve is specified by two numbers, a mean and a standard deviation, that can

[2] 'Normal' in the sense of a standard of reference.

be estimated from the data. We can overlay the normal curves fit to each data set by selecting

`Graph` → `Histogram...`

from the menu bar, and selecting the `With Fits and Groups` option.

You should see the following plot:

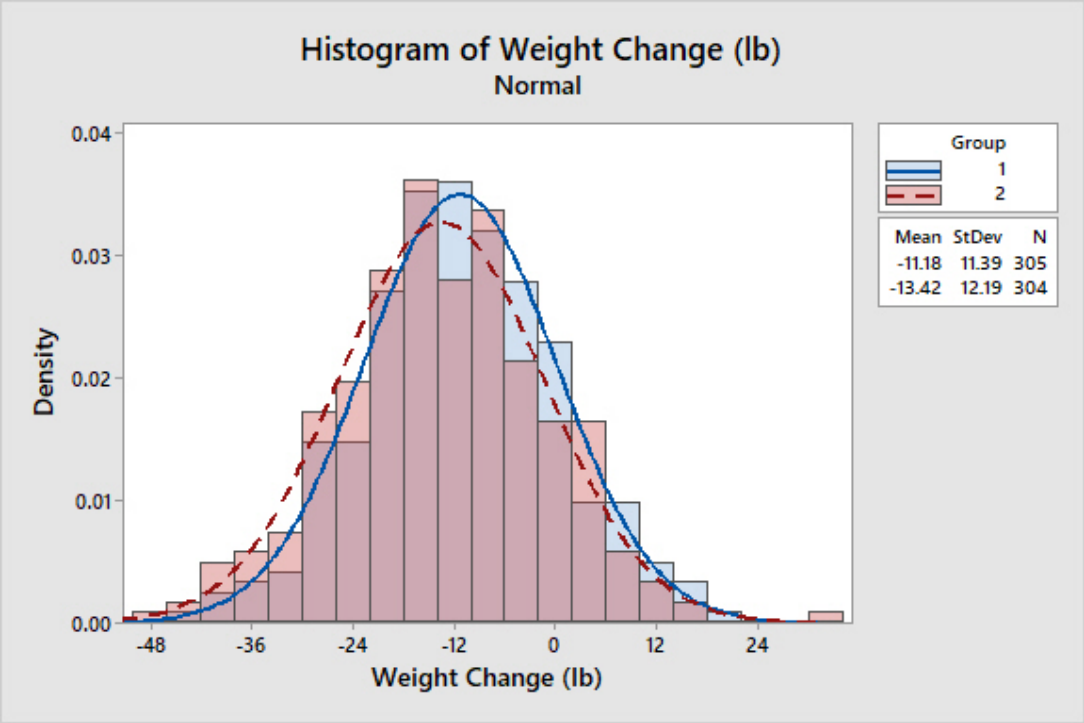

Figure 5:

We will return to the normal curve later in the semester, as it is a good model for many (but not all!) phenomena in the biological, medical, and social sciences.

6 Creating a Boxplot

As the final summary of the data, create boxplots for the two groups. **Hint:** Use the same instructions as for dotplots and histograms, but substitute the word boxplot into the instructions. You should see the following boxplots

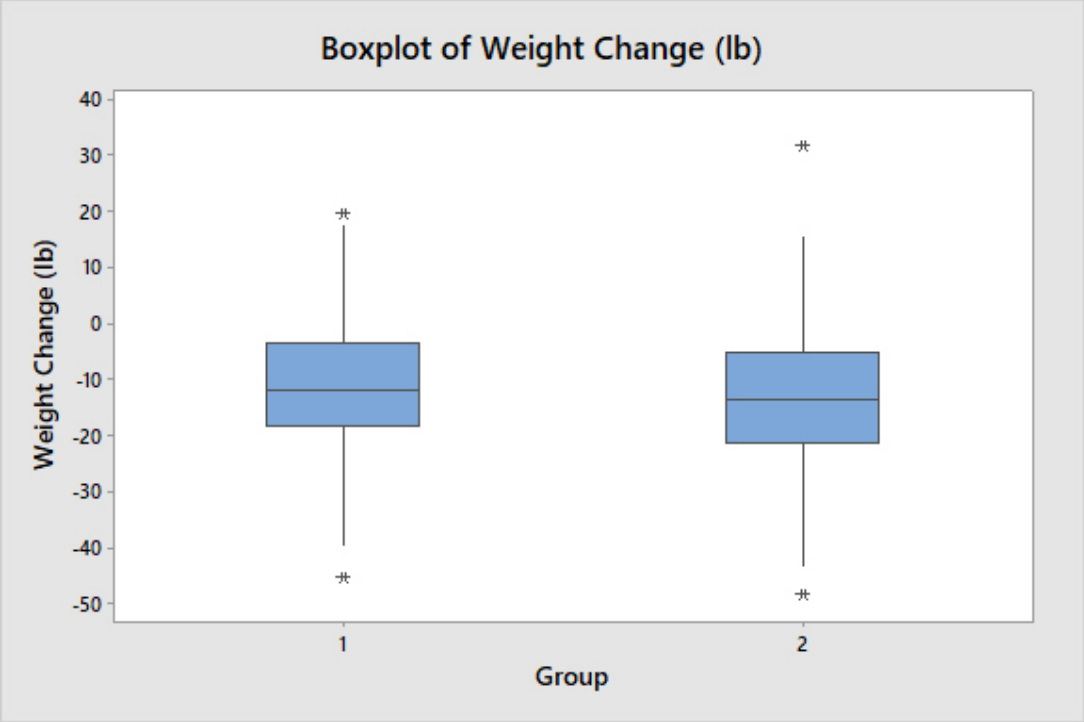

Figure 6:

Hands On 7

How do the boxplots compare to the dotplots? To the histograms? Which convey much information? Which are more succinct?

You can obtain the Five Number Summary of each data set by hovering the mouse pointer over each group.

Hands On 8

What is the 5 Number Summary for each group? Based on these summaries, which group had the better weight loss outcome: the low fat group or the low carb group?

7 Computing a Suite of Descriptive Statistics

Distributional summaries, like dotplots, histograms, and boxplots, give us a lot of information about a data set. But sometimes, a statistical question rests on a single characteristic of the data set. These single characteristics are known as descriptive statistics, since they (a) describe a feature of the data set and (b) are a statistic (feature) of the data set alone, and not the larger population. To compute a suite of descriptive statistics about each of the groups in the diet study, select

`Stat` → `Basic Statistics` → `Display Descriptive Statistics`

from the menu bar, and enter the appropriate variable and group indicator. The descriptive statistics will be displayed in the `Session` window.

Hands On 9

Using the Descriptive Statistics summary in the `Session` window, answer the following questions.

1. How many subjects were in the low fat group? The low carb group?
2. Where is the 5 Number Summary, and how does it compare to the 5 Number summary from the boxplots?
3. Which group, on average, had the largest change in weight?

We cannot reach any conclusions as to which dietary intervention would perform best for overweight individuals in the US just by considering the descriptive statistics. To reach a conclusion will require additional assumptions about how the particular sample of individuals relate to the larger population, and going from a statistic of the sample (like the difference of the mean weight changes in the sample) to a parameter (the difference in the expected weight change in the overweight US population). We will return to this question when we learn about inferential statistics later in the course.

References

[1] Christopher D Gardner, John F Trepanowski, Liana C Del Gobbo, Michelle E Hauser, Joseph Rigdon, John PA Ioannidis, Manisha Desai, and Abby C King. Effect of low-fat vs low-carbohydrate diet on 12-month weight loss in overweight adults and the association with genotype pattern or insulin secretion: the dietfits randomized clinical trial. *The Journal of the American Medical Association*, 319(7):667–679, 2018.