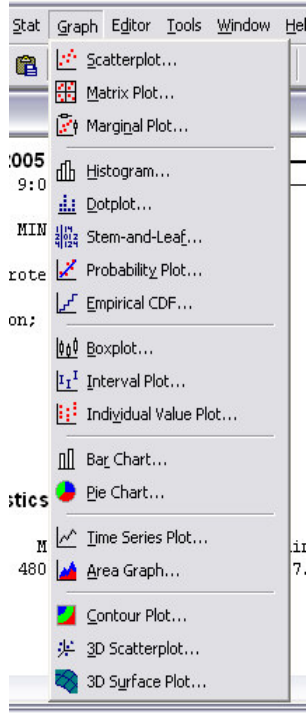


Lab 2

More graphical summaries

1. Plotting data under the Graph menu

a. There are many types of plots Minitab can produce for you

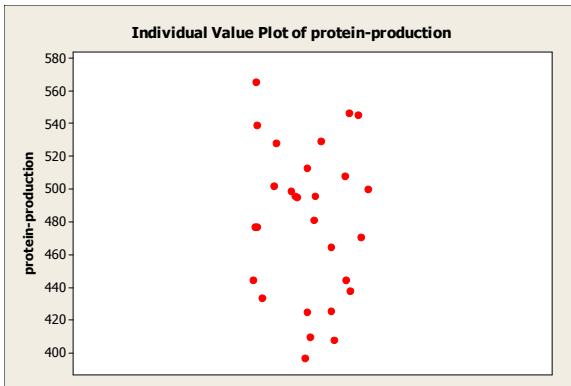


Consider the Ch. 2 Holstein.txt data.

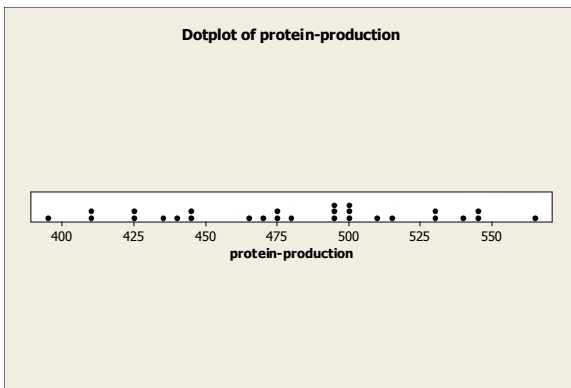
protein-production	496	445
425	502	546
481	529	471
477	500	495
434	465	445
410	539	565
397	408	499
438	513	508
545	496	426
528	477	

Let's look at a number of plots that are available to us. The plots I discuss in detail here are all shown below.

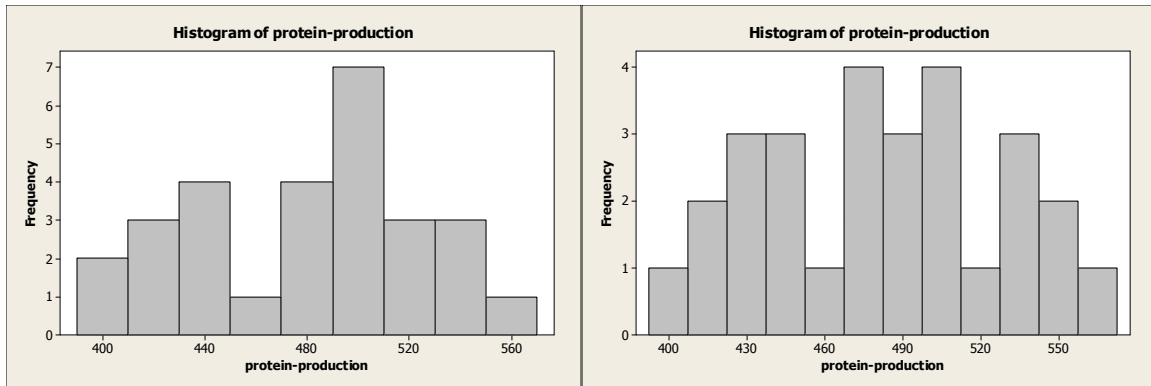
- b. Individual value plot. The horizontal axis is meaningless in this plot. It randomly displaces the values horizontally so that they do not overlap. Look at the values vertically for a sense of how spread the data are, and if they tend to cluster around particular values.



- c. Dotplot. The dotplot is similar to the individual values plot in that every observation is represented by a point. But this plot is more organized in histogram-type style.



- d. Histogram. The bars help smooth out the data to give a general idea of the shape of the data. The first histogram below is plotted with the default settings. If you right-click on the bars, and select “edit bars” and choose the Binning tab, you can specify the number of intervals you want. More bins will give you more detail and less smoothing, while less will do just the opposite. The second histogram has 12 intervals instead of 9. Notice that the plot appears trimodal (or even quadmodal) instead of the unimodal or bimodal shape of the first histogram. Therefore, the number of bins may change some of your interpretations of the shape of the data. My suggestion is to smooth the histogram out in your mind so as to not see characteristics of the data that aren’t really there. For example, by looking at either of these histograms I would conclude that the data are roughly symmetric, unimodal, with no outliers. The first histogram appears slightly left-skewed, but it is very mild.



- e. Stem-and-leaf plot. This is a plot you can quickly do by hand for small datasets. Below there are three columns. The first column is not part of the plot, and is simply counting how many observations appear in each tail, with the location of median indicated by parentheses. The second column is the stem, this is the numeric part of each data value except the last digit. The third column are the leaves, these are the last digit of each data value. In this case the data have first been rounded to the 10s place before being plotted. For example, the first few values plotted on this plot is 39 (first row), 40 and 41 (second row), and so on.

Stem-and-Leaf Display: protein-production

Stem-and-leaf of protein-production N = 28
Leaf Unit = 10

```

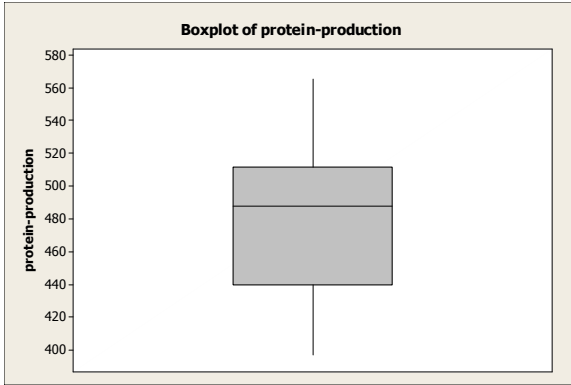
1   3   9
3   4  01
7   4  2233
9   4  44
13  4  6777
(5) 4  89999
10  5  0001
6   5  223
3   5  44
1   5  6

```

- f. Boxplot (or box and whiskers plot). The boxplot is simply a plot of the five number summary. The tips of the whiskers are the minimum and maximum values in the data. The box in the middle includes the median in the center, and the first and third quartiles (Q1 and Q3) at the edges of the box.

Descriptive Statistics: protein-production

Variable	Minimum	Q1	Median	Q3	Maximum
protein-producti	397.00	439.75	488.00	511.75	565.00



2. A nice first view of the data is given by the Stat/Basic Statistics/Graphical Summary. This plot provides a histogram with a normal distribution curve overlaid to help you decide whether the data appear normal (which is a necessary condition for hypothesis testing later). Below the histogram is a box plot, and below there are confidence intervals for both the mean and median. To the right is a hypothesis test for whether the data appear normal (if the p-value is small (eg. less than .05) we might conclude the data are not normal), also sample statistics such as mean and standard deviation are provided. Below that is the five-number summary which are the values the boxplot are plotted from. And finally the interval endpoints for three confidence intervals are given. In one quick step almost everything you might want to know about a set of data of one variable are provided in this plot. I love you, Minitab.

