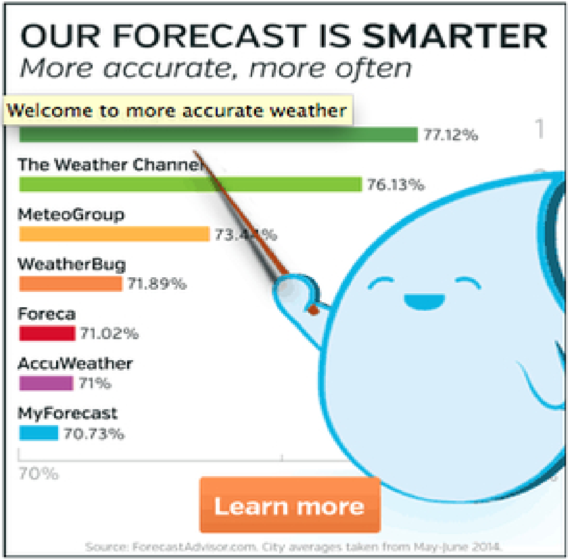
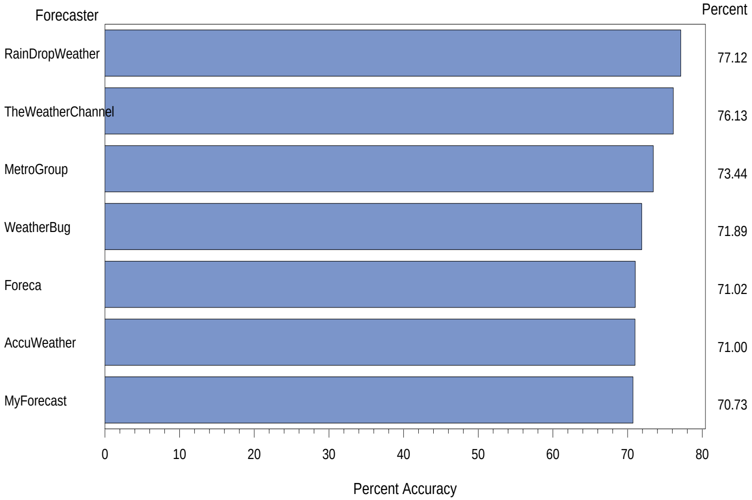

OUR FORECAST IS SMARTER
More accurate, more often
Welcome to more accurate weather
77.12%
1
The Weather Channe
76.13%
MeteoGroup
WeatherBug
71.89%
Foreca
71.02%
AccuWeather
71%
MyForecast
70.73%
70%
Learn more
Source: ForecastAdvisor.com. City averages taken from May-June 2014.

Our Forecast Is Smarter



Side-by-Side Boxplots

Boxplots can also be useful in comparing different groups of data. The next plot shows side-by-side boxplots comparing the weights of bread from different ovens (SAS code here). Compare this plot with the stratified plot used for the same purpose in the Chapter 1 lecture notes.

$\approx$

Side-by-Side Boxplots

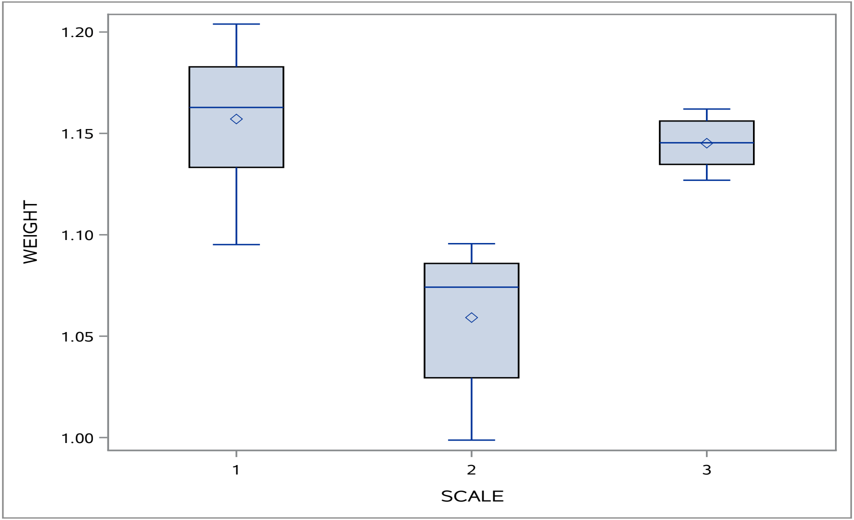


Figure: 16: Side-by-Side Boxplots for the bread data.

What Tukey Meant

Recall the quote from John Tukey which began this chapter:

“Numerical quantities focus on expected values, graphical summaries on unexpected values.”

What Tukey Meant

Recall the quote from John Tukey which began this chapter:

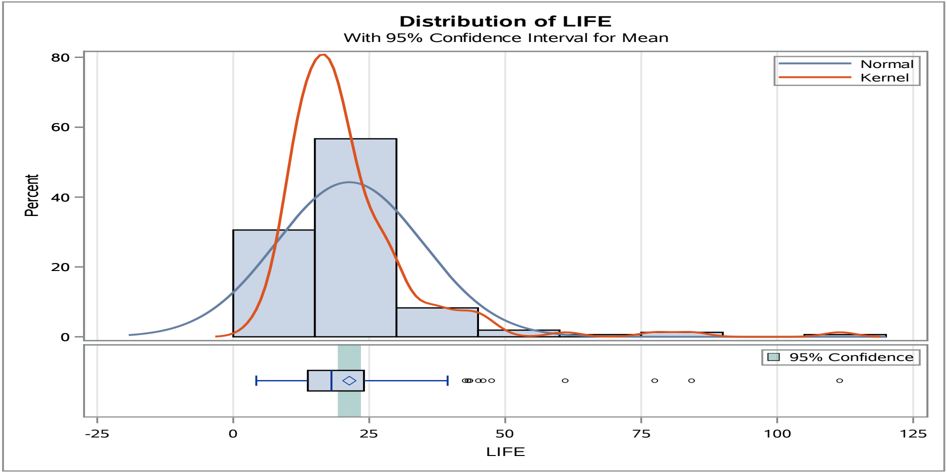
"Numerical quantities focus on expected values, graphical summaries on unexpected values."

With what you have learned, you should be able to appreciate it more, as the following graph will show.

$\approx$

What Tukey Meant

The numerical summaries tell us about most of the data; for the transformer data the median is around 18. On the graphs, however, our eyes are drawn to the exceptions, here the outliers.



Resistant Summary Measures

Resistant Summary Measures

- Summary measures are **resistant** if they are not seriously affected by outliers.

Resistant Summary Measures

- Summary measures are **resistant** if they are not seriously affected by outliers.
- The median and IQR are resistant measures of location and spread.

Resistant Summary Measures

- Summary measures are **resistant** if they are not seriously affected by outliers.
- The median and IQR are resistant measures of location and spread.
- The mean and standard deviation are not resistant.

$\approx$

Resistant Summary Measures

Though the mean is not resistant, for many data sets without outliers it is a better measure of location than the median. Two measures which attempt to add some resistance to the mean, while retaining its good properties when there are no outliers, are the **trimmed mean** and the **Winsorized mean**.

$\approx$

Resistant Summary Measures

- The k-times trimmed mean omits the k largest and k smallest data values and takes the mean of the remaining ones.
- To compute the k-times Winsorized mean, first create a new data set by replacing the k smallest data values with the value of the $k+1$st smallest, and the k largest data values with the value of the $k+1$st largest, while leaving the other data values untouched. The k-times Winsorized mean is the mean of all values in this new data set.

$\approx$

Example: Trimmed and Winsorized Means

As an example, consider again the last data set used to illustrate computation of quartiles. In ascending order the values are: 121, 133, 143, 144, 167, 172, 240, 250, 260.

To compute the 2-times trimmed mean, discard the two largest values (250, 260) and the two smallest values (121, 133), and take the average of the remaining values. The value of the two times trimmed mean is therefore

$(143 + 144 + 167 + 172 + 240)/5 = 173.2.$

$\approx$

To compute the 2-times Winsorized mean for these same data (121, 133, 143, 144, 167, 172, 240, 250, 260), set the two largest values to the value of the third largest value, 240, and set the two smallest values to the value of the third smallest value, 143, giving a modified data set: 143, 143, 143, 144, 167, 172, 240, 240, 240. Then take the average of these values. The answer is $181.\overline{3}$.

$\approx$

Recap, Chapter 2: Summarizing Data:

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same
  - Causes of common patterns

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same
  - Causes of common patterns
- Summary measures for stationary data distributions and when each is appropriate

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same
  - Causes of common patterns
- Summary measures for stationary data distributions and when each is appropriate
- Boxplots and outliers

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same
  - Causes of common patterns
- Summary measures for stationary data distributions and when each is appropriate
- Boxplots and outliers
- Resistant summary measures

Recap, Chapter 2: Summarizing Data:

- Displaying stationary data distributions
  - Bar charts, frequency histograms
  - Analysis of same
  - Causes of common patterns
- Summary measures for stationary data distributions and when each is appropriate
- Boxplots and outliers
- Resistant summary measures

$\approx$