STAT 252

Week 8

May 27, 2019 – June 5, 2019

**Time Series**

Motivation: Measuring a quantity (quantitative variable) over time and to predict future values

Examples:

* Daily log returns on stock.
* Monthly Consumer Price Index (*CPI*) values are equally spaced by month, not by day
* Daily recording of Weather
* Weekly observations of weight
* Yearly birth rate

Question 1: What do all of these examples have in common?

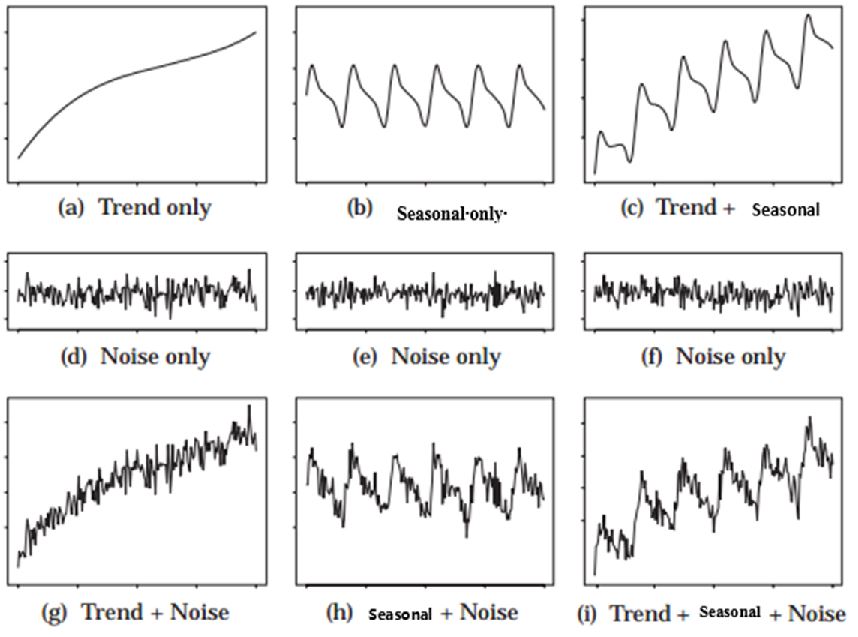
Question 2: Think of 2 examples that could be measured through time.

**Time Series Process:**

1. Recognize that the question at hand, can be solved with Time Series
2. Check assumptions
3. View Data
4. Determine the Type of Pattern
   * Determine the Time Interval
   * Determine the variable that is being measured/counted
   * Look Trend/Seasonality/Cycle
5. Determine and implement the appropriate time series model based on specified pattern
6. Assess model metrics
7. Forecast values

**Assumptions:**

* Consecutive observations are equally spaced
* Apply a time observation index
* This may only hold approximately

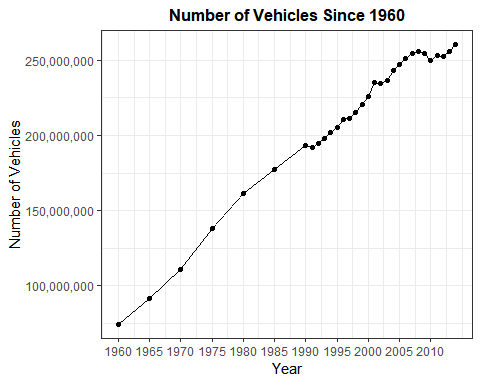


Question 3: What is occurring in (a)?

Question 4: What is occurring in (b)?

Question 5: What is occurring in (c)?

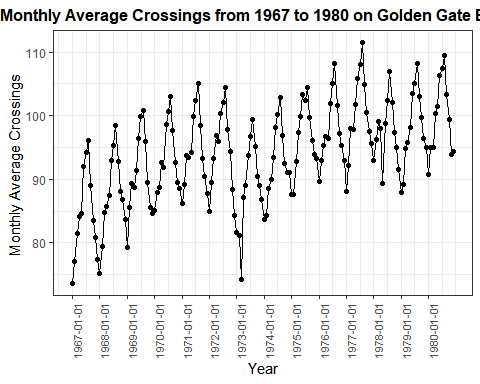
**Linear Trends (Positive and Negative)**



Question 6: What is the variable that is being measured or counted in this problem?

Question 7: What is the time interval in this question? Is this a problem?

Question 8: Describe the pattern in the graph.



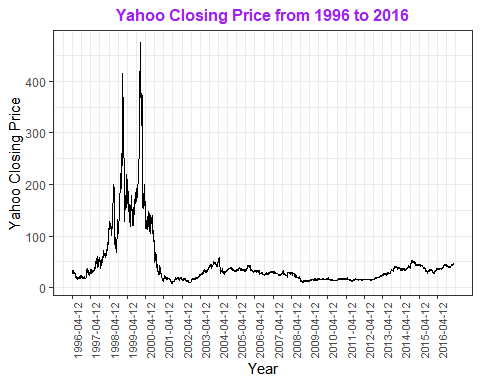
Question 9: What is the time interval in this problem? How else could the x-axis be represented?

Question 10: What is the variable that is being measured or counted in this problem?

Question 11: Describe the pattern in this graph.

Question 12: Does the plot of bridge crossings display a trend as well as seasonality? If so, describe the trend and explain how it shows up.

Question 13: Why do you think the average crossings per month is slightly increasing over the years in this graph?

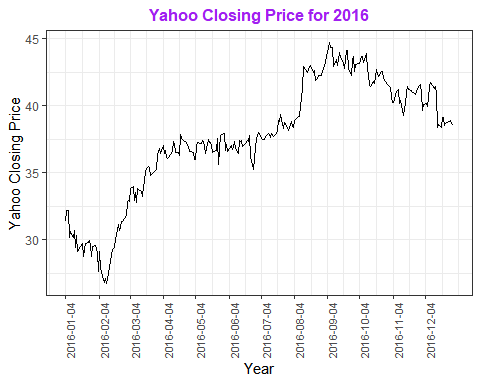


Question 14: What is the time interval in this problem?

Question 15: What is the variable that is being measured or counted in this problem?

Question 16: Identify a feature evident in this plot that did not appear in earlier examples.

Question 17: What is the difference between cyclic and seasonal?



Question 18: What are the benefits of zooming in?

Question 19: Break the series into intervals based on the line increasing, decreasing, or remaining constant.

**Smoothing**

Smoothing is simple a technique used to modify a time series so that predictions can be made in which the pattern is removed.

In other words, the ability to be predict within the context of time series, a smooth line will be created in which the patterns (variability) seen in the time series are accounted for.

Notation:

- current time period

- next time period

- previous time period

- actual quantity at time

- smoothed value at time

- forecast for time period

**Smoothing Technique 1: Moving Average ()**

where is the number past data points

* Large values for put more weight on past data points (Smoother)
* Small values for put more weight on recent data points

Where MA is Moving Average. Definition - A moving average (MA) of length L simply takes the average of the current value and the previous values.

To Predict/Forecast when using Moving Average future values are defined as:

Therefore, one period average Forecast at provides the following equation:

the above equation implies that the average of the past values is the predicted value for

Question 20: Find the moving average of the following numbers where and .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 2 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 3 |  |  |  |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |
| 2 |  |  |  |  |
|  |  |  |  |  |

For the data point, , the error is as defined as:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | Error |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Smoothing Technique 2: Exponential Smoothing ()**

where is the value between 0 and 1.

* Small values for put more weight on past data points
* Large values for put more weight on recent data points

Remember that:

* is the current value
* is the previous smoothed value is SES

Typically we let Where the forecasted value is defined as:

Question 21: Find the exponential smoothing of the following numbers where and .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 2 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 3 |  |  |  |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |
| 2 |  |  |  |  |

For the data point, , the error is as defined as:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | Error |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Model Metrics**

* **Mean absolute deviation (MAD)** Expresses accuracy in the same units as the data, which helps conceptualize the amount of error. Outliers have less of an effect on MAD than on MSE. is the sample size.
* **Mean squared error (MSE)** A commonly-used measure of accuracy of fitted time series values. Outliers have more influence on MSE than MAD. is the sample size.
* **Mean absolute percentage error (MAPE)** Expresses accuracy as a percentage of the error. Because this number is a percentage, it may be easier to understand than the other statistics. For example, if the MAPE is 5, on average the forecast is off by 5%. is the sample size.

Question 22: Find the MAD, MSE, and MAPE for Question 20.

Question 23: Find the MAD, MSE, and MAPE for Question 21.

**Autocorrelation**

Time series data will have errors that are not independent, we say that this data is **autocorrelated**.

If residuals vs. time were plotted and autocorrelation is present, there would be a pattern.

The Durbin-Watson statistic can detect **first-order** autocorrelation from the residuals of a regression analysis.

The null hypothesis:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ vs. The alternative hypothesis:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Quality Control**

Quality control methods in general and control charts in particular, have been very effective tools for management and industry.

They have played an important role in improving the quality of goods and service produced by manufacturers and service providers.

We will examine three of the more widely used types of control charts:

* -Charts for Process Location
* -Charts fot Process Variation
* -Charts for Process Attributes

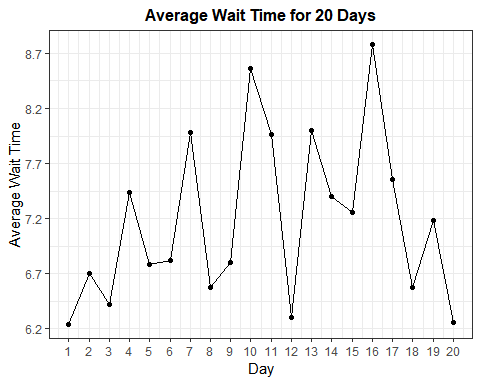
Motivation: On each of 20 different days, the waiting time (in minutes) was recorded for five randomly selected customers who entered an airline check-in line at Reno Airport. The airline gathered these data in order to investigate whether their waiting times are consistent from day to day.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| days | wait\_1 | wait\_2 | wait\_3 | wait\_4 | wait\_5 | mean | range |
| 1 | 5.9 | 4.9 | 5.6 | 7.8 | 7 |  |  |
| 2 | 6.5 | 6.6 | 6.6 | 6 | 7.8 |  |  |
| 3 | 10.1 | 4.9 | 4.5 | 6.3 | 6.3 |  |  |
| 4 | 7.1 | 5.5 | 11.3 | 5 | 8.3 |  |  |
| 5 | 7.3 | 5.7 | 9.4 | 4.9 | 6.6 |  |  |
| 6 | 10.4 | 3.6 | 4.8 | 7.6 | 7.7 |  |  |

Question 24: Record the first 6 means in the table above.

Question 25: Record the first 6 range values in the table above.

# -Chart



Question 26: The overall mean is , draw a horizontal line at this value. Does it appear that this process is “in control” with respect to location?

To assess whether the waiting time process is “under control,” we ask whether any of the sample means are farther from the overall mean than would be expected by normal variation.

Most control charts are based on the idea of considering a process to be in control if all of the sample means are within 3 standard deviations of the overall mean.

You know from the Central Limit Theorem that the standard deviation of a sample mean X is given by which is approximated by

But this is typically not done, because the sample standard deviation is difficult to compute by hand.

Instead we base the control limits on the average of the sample ranges, denoted by .

In the above example .

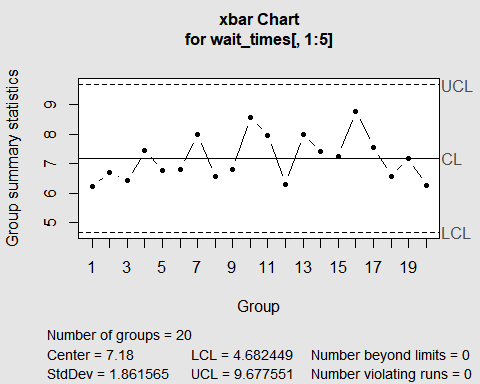
The control limits for an -Chart (of Process Location) are:   
  
and

where is the appropriate constant, dependent on the sample size n and tabulated in Table Q. (Table Q is on last page)  
  
Essentially is an unbiased estimator of .

Question 27: Use Table Q to determine the appropriate value of for these data.

Question 28: Calculate the upper and lower control limits for the -Chart

Your Calculations can be seen in the plot below



Question 29: According to these control limits, are any of the 20 sample mean waiting times out of control?

Question 30: Suppose that every single waiting time in all 20 samples was 30 minutes longer than it actually was. How would this change (if at all) the values of ? Justify your answers.

Question 31: Referring to question 30, would any of the sample means now be “out of control”?

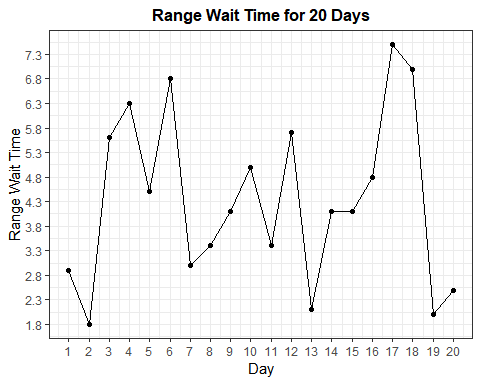
New Question: what if Kenny recorded the time in seconds and not minutes, how would this change (if at all) the values of ? Justify your answers.

What could be said about linear transformations in terms of quality control?

Notice that “in control” just means that none of the sample means is inconsistent with the pattern of variation shown in the others.

Being in control does not necessarily mean that the values are acceptable from a service or manufacturing perspective.

# R-Chart



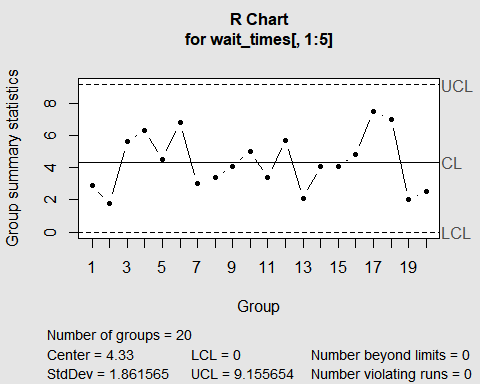
Question 32: Draw a horizontal line for . Does it appear that this process is “in control” with respect to variation?

The control limits for an -Chart (of Process Variation) are: and where and are the appropriate constants, again dependent on the sample size n and tabulated in Table Q.

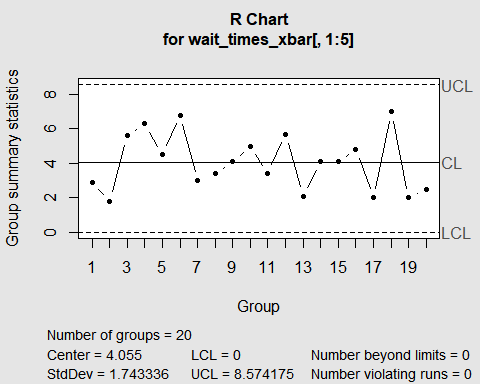
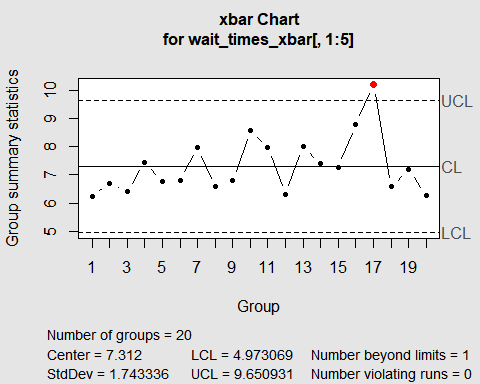
Question 33: Use Table Q to determine the appropriate values of and for these data.

Question 34: Calculate the upper and lower control limits for the R-Chart.

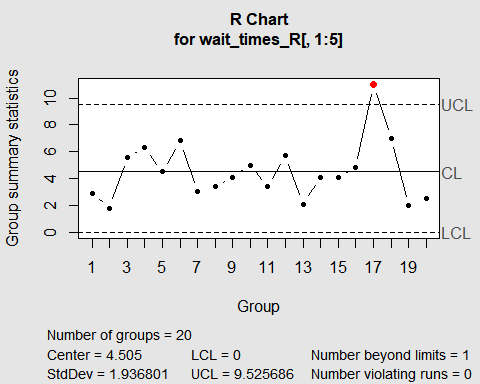
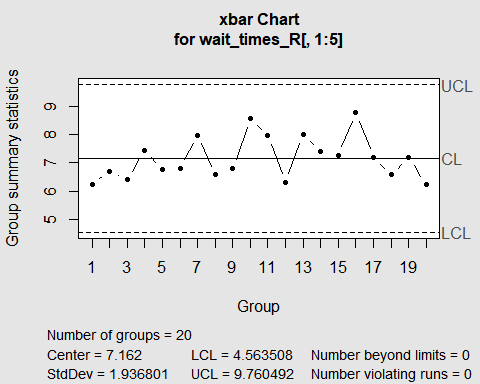
Question 35: According to plot below, are any of the 20 sample ranges (of waiting times) out of control?



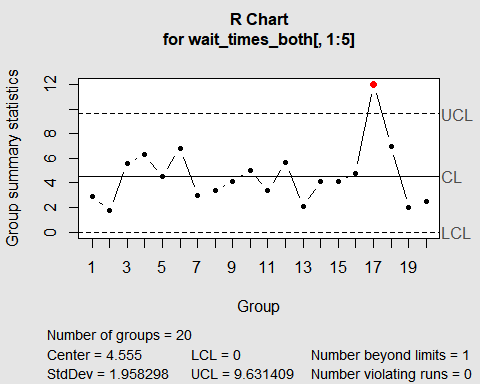
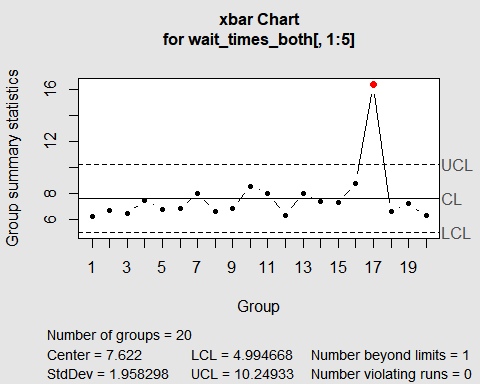
Question 36: How would one make the process out of control with respect to location but remain in control with respect to variation? Below is an example.



Question 37: How would one make the process out of control with respect to variation but remain in control with respect to location? Below is an example.



Question 38: How would one make the process out of control with respect to both location and variation? Below is an example.



# -Chart

Now we turn to control charts for attributes, in other words for a (binary) qualitative variables.

Motivation: Suppose a group of basketball players shoot 25 foul shots each at the end of each day’s practice. The coach keeps track of how many shots they miss each day for 10 days, wanting to assess which players have a foul shooting process that is “in control” and which do not.

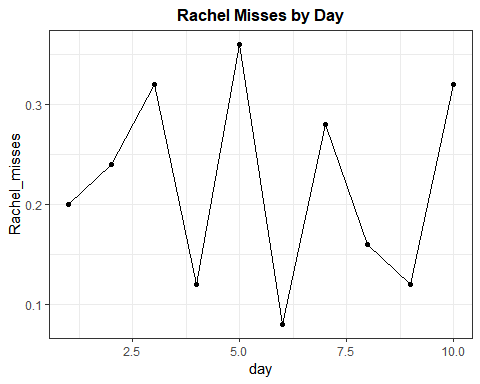
Question 39: What are the observational units and variable here?

Suppose the data on number of misses per day turn out to be:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| day | Rachel misses | Sam misses | Tierra misses | Uma misses | Veronica misses |
| 1 | 5 | 15 | 6 | 13 | 0 |
| 2 | 6 | 14 | 7 | 10 | 0 |
| 3 | 8 | 17 | 9 | 5 | 1 |
| 4 | 3 | 16 | 15 | 3 | 3 |
| 5 | 9 | 19 | 7 | 2 | 2 |
| 6 | 2 | 20 | 5 | 0 | 0 |
| 7 | 7 | 15 | 6 | 4 | 1 |
| 8 | 4 | 14 | 5 | 1 | 5 |
| 9 | 3 | 17 | 7 | 6 | 2 |
| 10 | 8 | 18 | 0 | 3 | 0 |
| Mean Proportion |  |  |  |  |  |

Question 40: Intuitively, which players seem to have their foul shooting process under control and which do not?

Question 41: Below is a chart for Rachel’s missed foul shots proportions. Does this process appear to be under control for Rachel? What symbol is used for proportions?



Question 42: Calculate the average of Rachel’s sample proportion values. (We use the symbol for this.)

Question 43: Is this value equal to the overall proportion of shots that she missed, for the 10 days combined? Explain why or why not.

The mean subgroup proportion is the center line for the p-chart.

To determine the lower and upper control limits, we need to determine the standard error of .

Question 48: Recall from your study of the binomial distribution, the standard deviation of a sample proportion .

Question 44: What is the standard error of ?

Question 45: In the notation of quality control, and using a pooled estimate for the unknown population proportion, what is the standard error of ?

Question 46: Calculate this standard error for Rachel’s sample proportion of misses.

The control limits for a - chart are:

Where is the mean subgroup proportion and n is the common sample size of the subgroups.

Question 47: Determine the lower and upper control limits for Rachel’s basketball shooting.

Question 48: Below is Rachel’s basketball shooting process in a -chart, is this in control? Explain.

