Using Control Charts for Communicating Variation

Dr Austin R Brown

Kennesaw State University

Introduction

- Over the past couple of class sessions, we have learned how to use various visual tools for communicating how some quantitative variable changes over time (in the case of a time series chart) or between groups (as we saw with boxplots).

- Another tool we have at our disposal that is incredibly common in many organizations is something called a *control chart*.

- A control chart is essentially a time series chart, but one in which we attempt to differentiate between typically occurring values and unusual values by incorporating statistical information.
  - For example, if I manage a plant that manufactures blue jeans, I need to make sure that if we are manufacturing jeans with a 34 inch inseam, that we are, on average, consistently hitting that target.

Introduction

- Instead of measuring every single pair of jeans, suppose I randomly sample every 10th pair of jeans manufactured and collect a total of 10 pairs per hour. I do this over the course of 30 hours.
  - Below represents the data structure:

Hour	Observation 1	Observation 2	Observation 3	Observation 4
1	34.05	34.11	33.92	33.89
2	33.22	33.54	33.25	34.48
3	33.86	33.55	34.46	34.23
4	34.66	33.73	34.06	33.82
5	33.27	33.79	33.86	33.56

Introduction

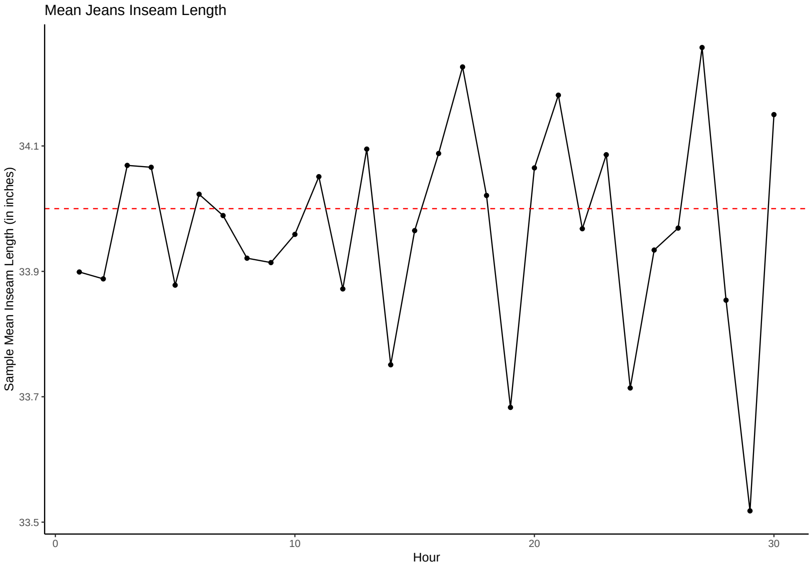
- Logically, we could plot these data in a time series chart much like we did with the pitching data.
- Let's try this by taking the mean at each time point and plotting it in a time series chart.

Introduction

```r
jeans2 <- bind_cols(
  jeans |>
    select(Hour),
  apply(jeans[,-1],1,mean)
)
names(jeans2)[2] <- "X-Bar"

jeans2 |>
  ggplot(aes(x=Hour,y=`X-Bar`)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 34,color='red',linetype='dashed') +
  labs(x = "Hour",
       y = "Sample Mean Inseam Length (in inches)",
       title = "Mean Jeans Inseam Length") +
  theme_classic()
```

Introduction



Introduction

- We see that at the beginning of the time series, the means deviate somewhat from the target value of 34 inches but not too much.

- But as time goes on, we see that there is much more deviation from target with some points being more than a full inch away from target.
  - Not ideal!!

- However, based on this, how can we know if the deviations we're observing, big or small, are meaningfully different from target, from a statistical perspective?

- There's a test for that! We could use a one-sample t-test in order to make that determination.

Introduction to Control Charts

- But what's better is that we can visually perform this one-sample t-test without having to manually calculate 30 different p-values.

- This is the logic behind a *control chart*.

- All control charts have the same key elements:
  1. A plotting statistic (like a test statistic in hypothesis testing)
  2. Control limit(s) (like critical values in hypothesis testing)
  3. A target value that the process ideally adheres to (like a null value in hypothesis testing)

Introduction to Control Charts

- Traditionally, if we are monitoring the mean of some univariate process, the classical Shewhart $\bar{X}$-chart is what we use.
- The plotting statistic for the Shewhart $\bar{X}$-chart is the sample mean for each time point.
- The target value can be estimated from historical data or specified explicitly.
- So then how do we calculate control limits?

Introduction to Control Charts

- Remember in prior courses when we learned about critical values? We specified those using some value of α, say $\alpha = 0.05$.
- We can do that in control charting, too, to obtain control limits (these are called α limits).
- But what is much more common is to use what are called σ limits, limits that are calculated as functions of the plotting statistic's standard deviation.

Introduction to Control Charts

- Recall, the standard error of the sample mean is:

$$SE[\bar{X}] = \frac{s}{\sqrt{n}}$$

- where s represents the sample standard deviation and n represents the sample size.

- In process control, we do it a little differently and I can briefly explain why:

Introduction to Control Charts

- Remember that we use the below formula to calculate an unbiased estimate of variance:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- And, we generally use the square root of this estimator as the estimate of standard deviation:

$$s = \sqrt{s^2}$$

Introduction to Control Charts

- However, s is not an unbiased estimator of σ. For small sample sizes ($n < 100$), it will systematically underestimate the true value.

- This can be shown mathematically, but that's not the point of this class.

- The constant by which s biases σ is called c_4. So we need to divide s by c_4 to get an unbiased estimate for the process standard deviation.

$$E[s] = c_4 \sigma$$

Introduction to Control Charts

- Now, at each time point, we can calculate a sample standard deviation s.
- Then, we average them across all sampled time points to get a quantity $\bar{s}$.
- We can then add and subtract this quantity from our target value to obtain control limits:
  - Note, we can estimate the target value using the mean of the sample means if we don't have a target value known.

$$UCL = \mu_0 + L\frac{\bar{s}}{c_4\sqrt{n}}$$

$$UCL = \mu_0 - L\frac{\bar{s}}{c_4\sqrt{n}}$$

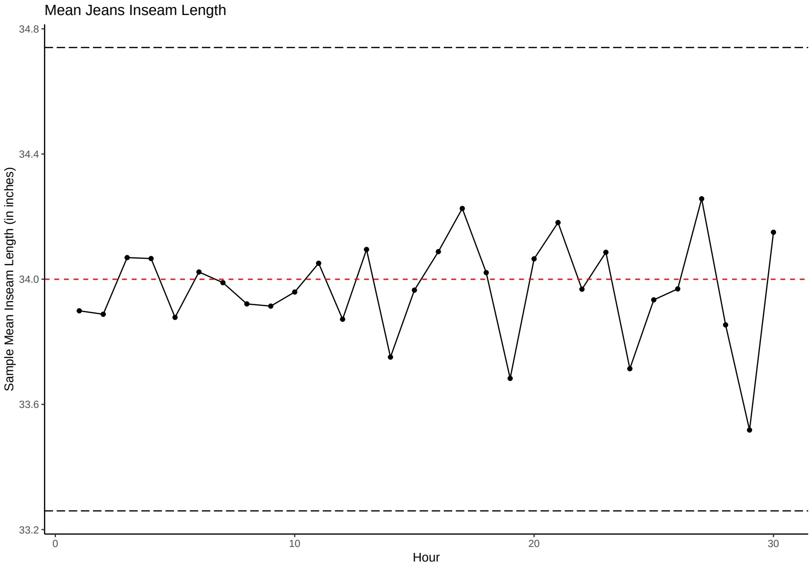
Introduction to Control Charts

- We generally choose $L = 3$ and can obtain c_4 from this table: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://web.mit.edu/2.810/www/files/readings/ControlChartConstantsAndFormulae.pdf
- Since our subgroup size is $n = 10$, we can use $c_4 = 0.9727$.
- Now we can explicitly calculate our control limits using R and plot them on our time series graph!

Introduction to Control Charts

```
## Okay, if we want to obtain control limits, we first have
## to calculate our mean standard deviation: ##
sbar <- mean(apply(jeans[,-1],1,sd))
## Since n = 10, c4 = 0.9727 ##
c4 <- 0.9727
## If L = 3, then we can calculate the Upper and Lower control limits:
L <- 3
UCL <- 34 + L*sbar/(c4*sqrt(10))
LCL <- 34 - L*sbar/(c4*sqrt(10))
## Now adding this info to our plot: ##
jeans2 |>
  ggplot(aes(x=Hour,y=`X-Bar`)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 34,color='red',linetype='dashed') +
  geom_hline(yintercept = UCL,color='black',linetype='longdash') +
  geom_hline(yintercept = LCL,color='black',linetype='longdash') +
  labs(x = "Hour",
       y = "Sample Mean Inseam Length (in inches)",
       title = "Mean Jeans Inseam Length") +
  theme_classic()
```

Introduction to Control Charts



Introduction to Control Charts

- Based on what we've observed here, we would conclude that our control chart exhibits statistical control.
- This means, we don't have evidence to suggest that the mean length of inseam differs substantially from 34 inches.
- Let's see how we can get these graphs using some built in packages: