MA22004

Seminar 10

Dr Eric Hall

10/12/2020

Announcements

Reminders

- It is week 10! This is our **last seminar before** the Class Test 3.
- You should have read §8 Quality Control on **Perusall**.
- Please look over the feedback from Class Test 2.

Upcoming

- Lab 7 due **Thursday 10 Dec** at **17:00** (late accepted until Sunday w/o penalty).
- Worksheet #10 Review is already posted to Blackboard.
- Class Test 3 is **Wednesday 16 Dec** from **14:00–17:00**.

What will Class Test 3 cover?

Class Test 3 will be longer than the first two tests and will be cumulative in nature, but will focus on the following topics.

- Categorical data (χ^2 tests for goodness of fit and independence)
- Linear regression (linear models, least squares estimates, inference for variances and least squares parameters, correlation)
- ANOVA (single factor tests and confidence intervals)

> Revise inferences (hypothesis tests and confidence intervals) for single samples and two samples.

Monday’s workshop will include a review of important topics.

Quality Control (QC)

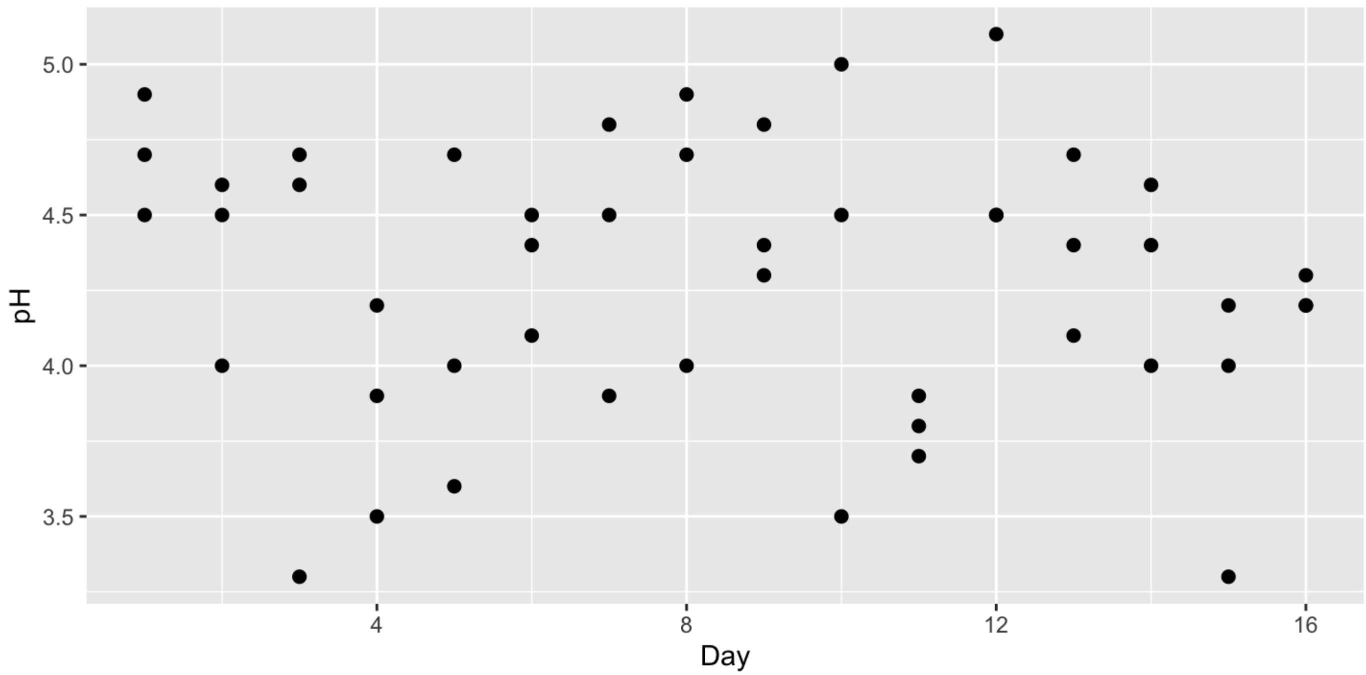
- Identify the sources of random variations in output processes that might have *assignable causes*.
- **Control charts** help us to recognize when industrial processes are no longer controlled.
- Control charts use time series data to calculate the running value of a quality statistic.
- If the quality statistic exceed upper/lower control limits, process is deemed to be out of control (quality negatively impacted).

> Workflow:
>
> time series $\longrightarrow$ control chart $\longrightarrow$ identify assignable causes

Beer Data: IPA

A brewer produces 3 batches of IPA per day and measures the pH value of each batch, which influences saccharification.

```
ggplot(ipa, aes(x = Day, y = pH)) + geom_point(size = 2)
```



3σ Control Chart

We consider the typical 3σ control chart for the mean pH $\overline{X}$.

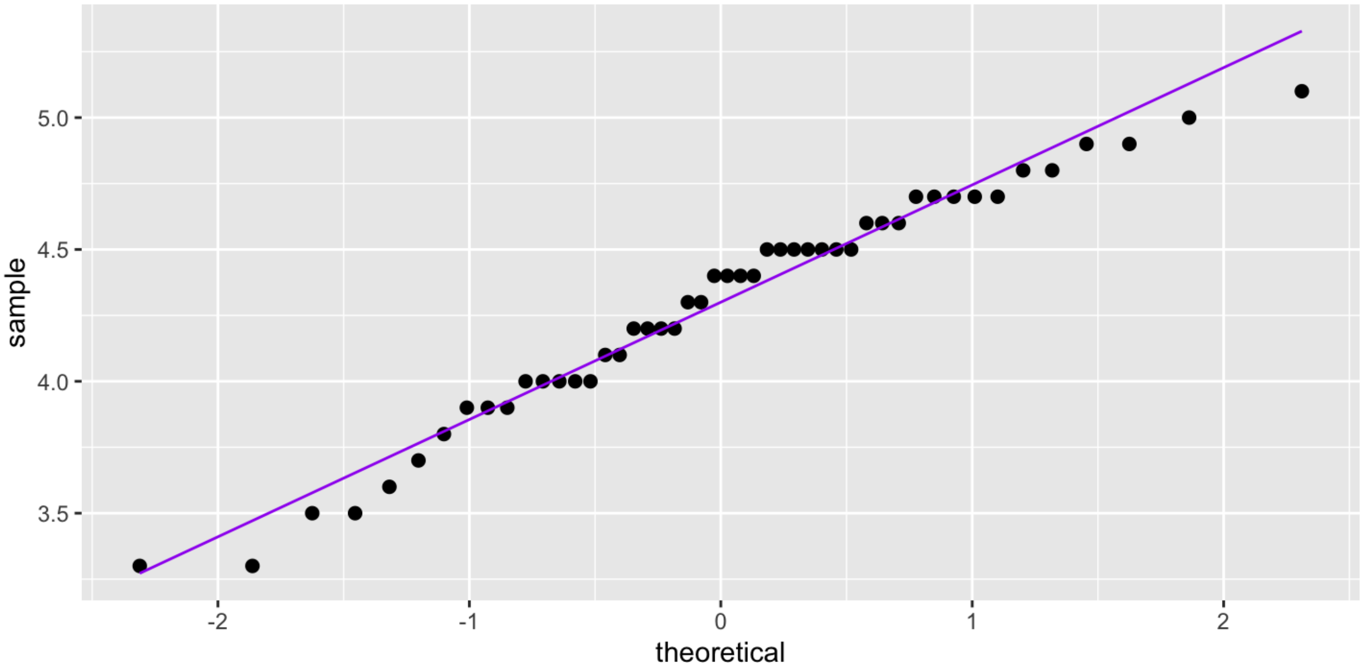
Assume generating process X is normally distributed with unknown μ and σ.

Our control region is specified to by three standard deviations: the process remains in control if it stays three deviations within a baseline value.

Why consider three standard deviations from the mean?

Checking for normality: `qqplot`

```
ggplot(ipa, aes(sample = pH)) + stat_qq(size = 2) + stat_qq_line(color = "purple")
```



Checking for normality: `shapiro.test`

Shapiro–Wilk tests the null hypothesis that a given sample came from a normally distributed population. Thus if P-value small, reject null and do no proceed.

```
shapiro.test(ipa$pH)
```

```

	Shapiro-Wilk normality test

data:  ipa$pH
W = 0.96781, p-value = 0.2079
```

> ⚠ Failing to reject null does *not* tell us the population is normally distributed…

Checking for normality ?

> If the sample size is *large*, which should we use (`qqplot` vs `shapiro.test`)?

Constructing a control chart

Day	pH Observations	$\bar{x}$	s	Range
1	4.7, 4.5, 4.9	4.700	0.2000	0.4
2	4.0, 4.6, 4.5	4.367	0.3215	0.6
3	4.7, 3.3, 4.6	4.200	0.7810	1.4
4	3.9, 3.5, 4.2	3.867	0.3512	0.7
5	4.0, 4.7, 3.6	4.100	0.5568	1.1
6	4.4, 4.5, 4.1	4.333	0.2082	0.4

$\hat{\mu}$

$\hat{\sigma}$

Estimating σ

Recall that the sample standard deviation is a biased estimator for the standard deviation.

For $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$,

$$\mathbf{E}(S) = a_n \cdot \sigma$$

where the *bias correction* can be computed exactly:

$$a_n = \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}.$$

n	3	4	5	6	7	8
a_n	0.886	0.921	0.94	0.952	0.959	0.965

Unbiased estimator σ

Let $\overline{S} = \frac{1}{k} \sum_{i=1}^{k} S_i$, where S_i are the sample sd for k days.

Control limits based on sample sds

$$\text{LCL} = \hat{\mu} - 3\frac{\bar{s}}{a_n\sqrt{n}}$$

$$\text{UCL} = \hat{\mu} + 3\frac{\bar{s}}{a_n\sqrt{n}}$$

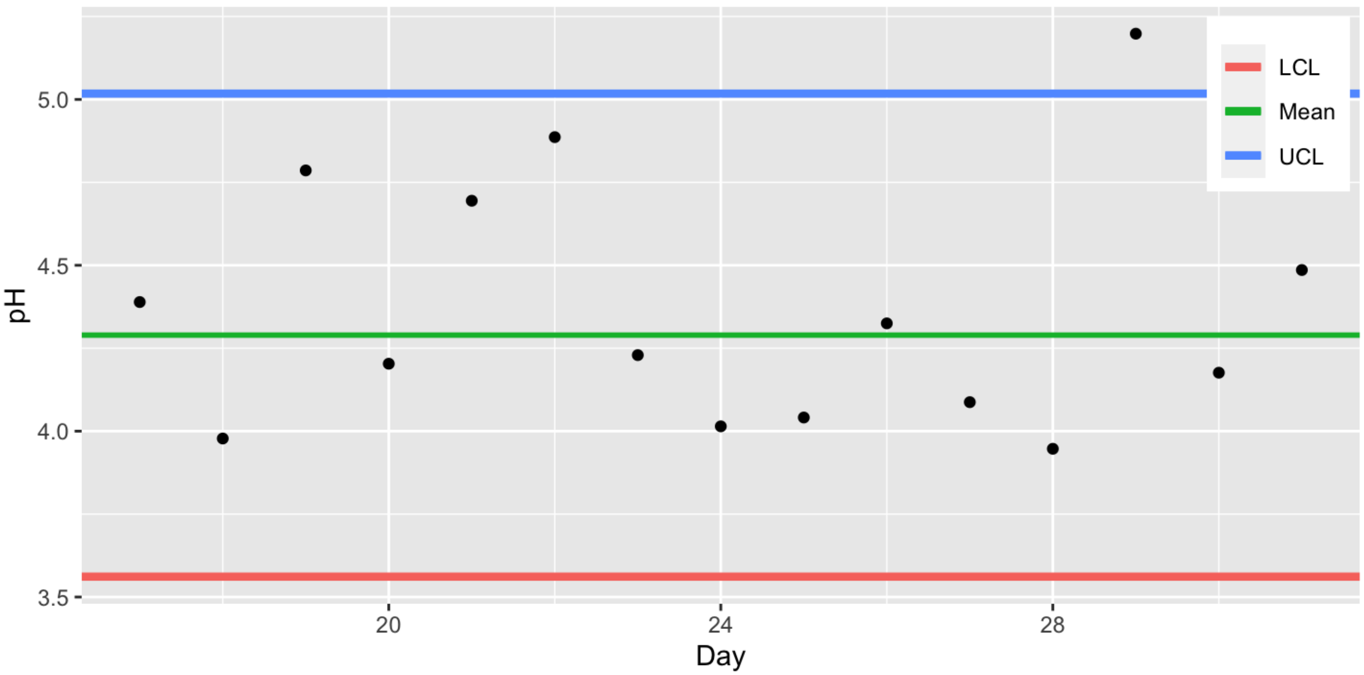
where

$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k}\bar{x}_i$

$\bar{s} = \frac{1}{k}\sum_{i=1}^{k}s_i$

Control Chart

- Assume process is in-control over the 16 days and compute the limits.
- Violation would require that we seek to identify an assignable cause.



Summary

Today we discussed control charts for Quality Control.

We learned about bias correction for estimating the population standard deviation.