Review topics today:

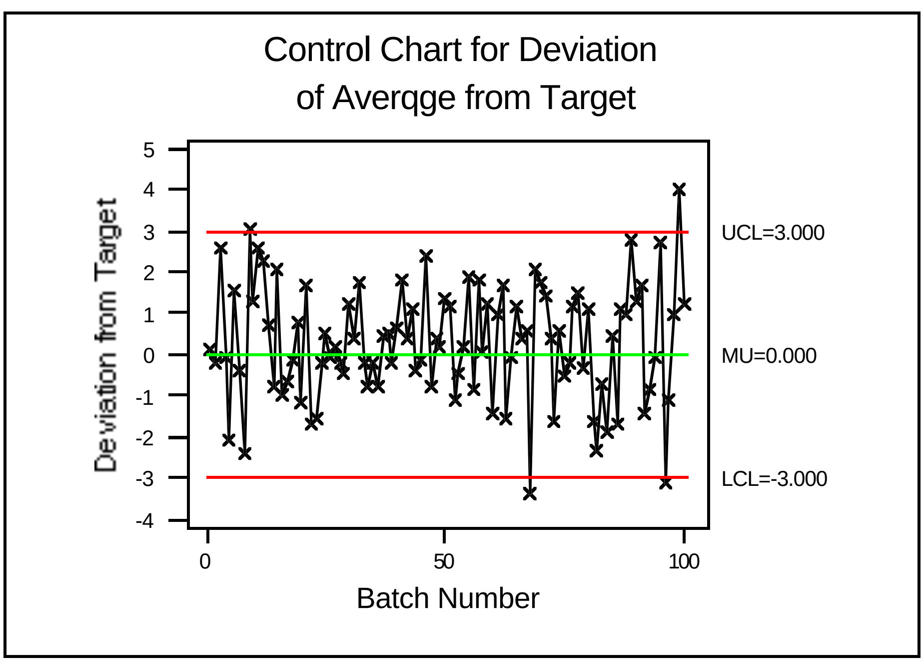
7. Quality Control
8. Hypothesis Testing
9. Optimization and Spatial Data
10. Survival Analysis

and course evaluations

I will hold a tutorial in Monday's class (Dec 2). I will not present anything so come with questions. I will hand out the Statistics Workshop Evaluation Forms at that time (Dec 2).

7. Quality Control

Main ideas: regular monitoring of quality (hourly, daily, ...)
use averaging to get normality (daily sample of output in manufacturing)
management by exception – use distribution extremes (±3SDs from mean)
reduction of variation by elimination of identifiable causes
increased profitability from reduction of variation
e.g. sort lumber - charge quality prices



8. Hypothesis Testing

There were two methods for comparing two population means:

1. **Estimate** the difference (population mean 1 – population mean 2). This is the confidence interval approach . See if the CI for the difference includes 0 as a credible value.
2. **Test** whether the difference could be 0. Ask, what is the probability of a sample difference as large as I observed IF the population difference in means is 0?

The hypothesis test we did was concerned with a comparison of pulse rates for males with pulse rates for females. Our data was from the students in the class at the time, and this is not really a random sample from any population. But we suppose (for the purpose of the demonstration only) that the data we collected were two random samples from two populations, male students and female students.

We had mean ± SD for each sample:

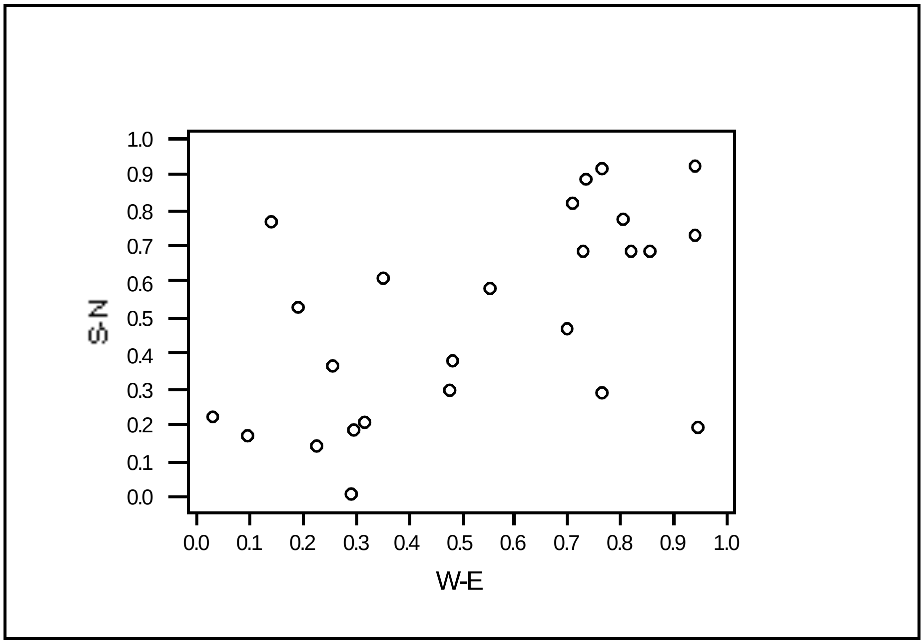
```
37.7 ± .75 pulses for females
35.8 ± .83 pulses for males
```

The difference between these two means is 1.9 and the SD of the difference was computed to be 1.1. ($\sqrt{.75^2 + .86^2} = 1.1$)

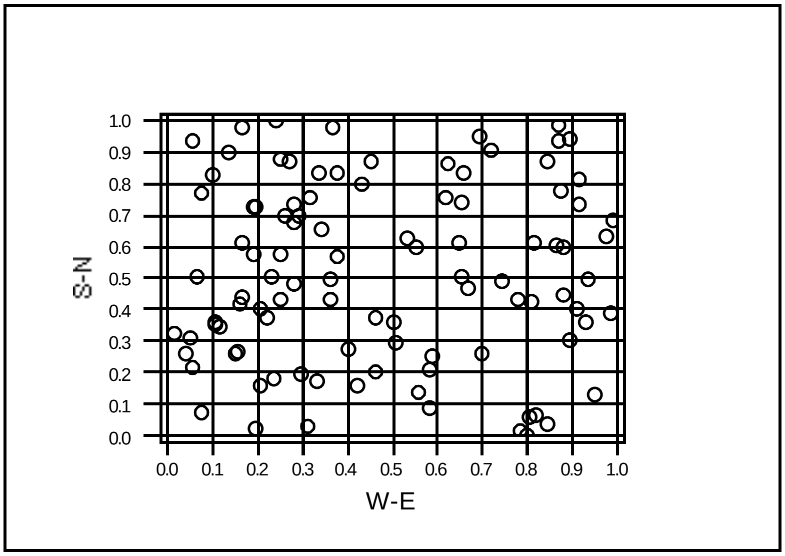
If the population differences were equal, then the average difference on the sample means would be 0. So the question is, how likely is it that a sample difference of 1.9 would occur when the the mean difference is 0 and the SD of the sample difference is 1.1? The answer is that it is between 1 and 2 SDs above the mean so the probability is between 16% and 2.5% that this would happen. This gives us weak evidence that the population means are different. (The exact percentage turns out to be 4%).

"When something happens that is rare under ordinary circumstances, this is evidence that the circumstances are not ordinary. "

9. Optimization and Spatial Data



The above spatial distribution of 25 points was generated in a way that makes every point have the same chance of being selected. This is called a **uniform** distribution. If a data scatter like this were observed for crime locations in a city, we might be tempted to consider what it was about the clusters that made crime more likely there. However, while there appears to be some clumping, if we are tempted to conclude there is some reason for it, we would be wrong since this data is actually showing the amount of clumping that occurs from a sample of a uniform distribution. The point is that the clumping has to be stronger than this to justify looking for a reason.



This graph was intended to show how "clumping-ness" could be described by a distribution:

Row	noincell	nocells
1	0	34
2	1	39
3	2	20
4	3	7
5	4 or more	0

The only thing to notice is that, although the population from which the points were sampled was uniform, the sample does not look uniform. We did not have 100 points such that each cell had one point.

10. Survival

The data we collected was

1. duration of observation = current date – license date
2. accident yet?

We did not have a single AFST – accident free survival time. And yet we could estimate the distribution of the ASFT. The key was sorting the data by 1. and observing 2.