Using Statistics in Optimization and Analysis of Bioprocessing

-- 20 Tips for Success --

George Bernstein, Ph.D.
Double Dragon Consulting Inc.

Harsha Ingle, Ph.D.
GlaxoSmithKline

<u>Topics</u>

- Biases
- Organizational
- Data
  - Collecting it
  - Looking at it
  - Analyzing it
- Statistical Analysis

Biases

1. Many people don’t like or trust statistics
   - Didn’t like it in college
   - Don’t understand the limitations/power of statistics
   - Have been burned in previous projects

- To implement a successful statistical project, you must:
  - Follow good project management principles
  - Establish clear objectives
  - Reasonable timelines
  - Establish a multidisciplinary team,
  - Set clear realistic goals
  - Be able to communicate progress/results in plain English

2. Statistics don’t hate you so don’t hate them

- Part of being a good scientist requires knowledge of statistical principles
- Statistics is a language that takes time and patience to understand
- Budget some time to talk with a statistician before you begin any large project
- Make a point to learn the strategic use of the tools

3. Don’t ask, Don’t tell?

- Statistics makes you accountable:
  - If you just spent $100K on a study, and your management team likes it, do you really want to find out if your results are statistically significant?
  - If you can't afford a negative answer, don't ask the question.

Organizational

4. Bring your statistician in early

- Unlike dropping your receipts on your tax accountant’s desk, you can’t throw data at your statistician after the fact and expect him/her to make sense of it

5. Get everyone involved

- Studies take time and are expensive
  - Better to start well than to re-do the study
  - You may not know all the factors that can affect the study.
- Get as many people that might influence the study involved in the planning:
  - Brainstorming
  - Buy-in

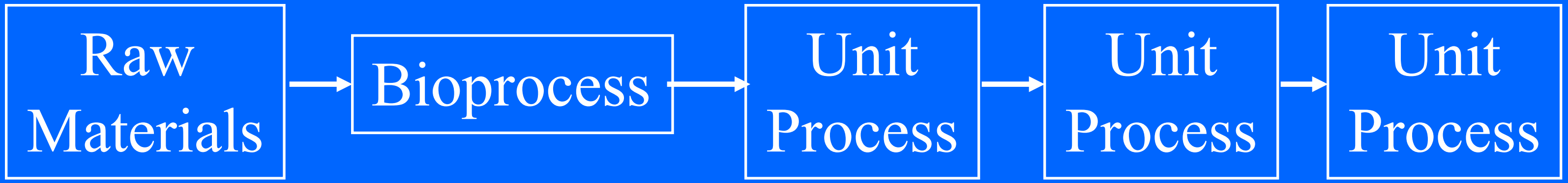
Data

6. The second most valuable resource a company has is its production data

(the first is the people)

- We love being data pack rats because it makes us feel secure that our processes are in control
- But more important than this reflex is to determine the right data and use it.
- To make use of your production data
  – You must know what data are important
  – Remember 21CFR Part 11

Different processes, different data



7. Do you have a good sampling plan ?

- A large number of 486s are due to bad sampling plans. For example:
  - Due to long QC testing schedules, samples are often taken at the beginning or start of the process
  - Samples that are hard to obtain are ignored
- The result?
  - Samples that are not representative of your process
  - Wrong conclusions

8. Know your errors

- Different error types
  - Type I error (concluding that things are different when they aren’t)
  - Type II error (concluding that things aren’t different when they are)
  - Data entry errors (always double check!).
  - Computational errors.
- Have methods to validate your data before you begin your analysis.

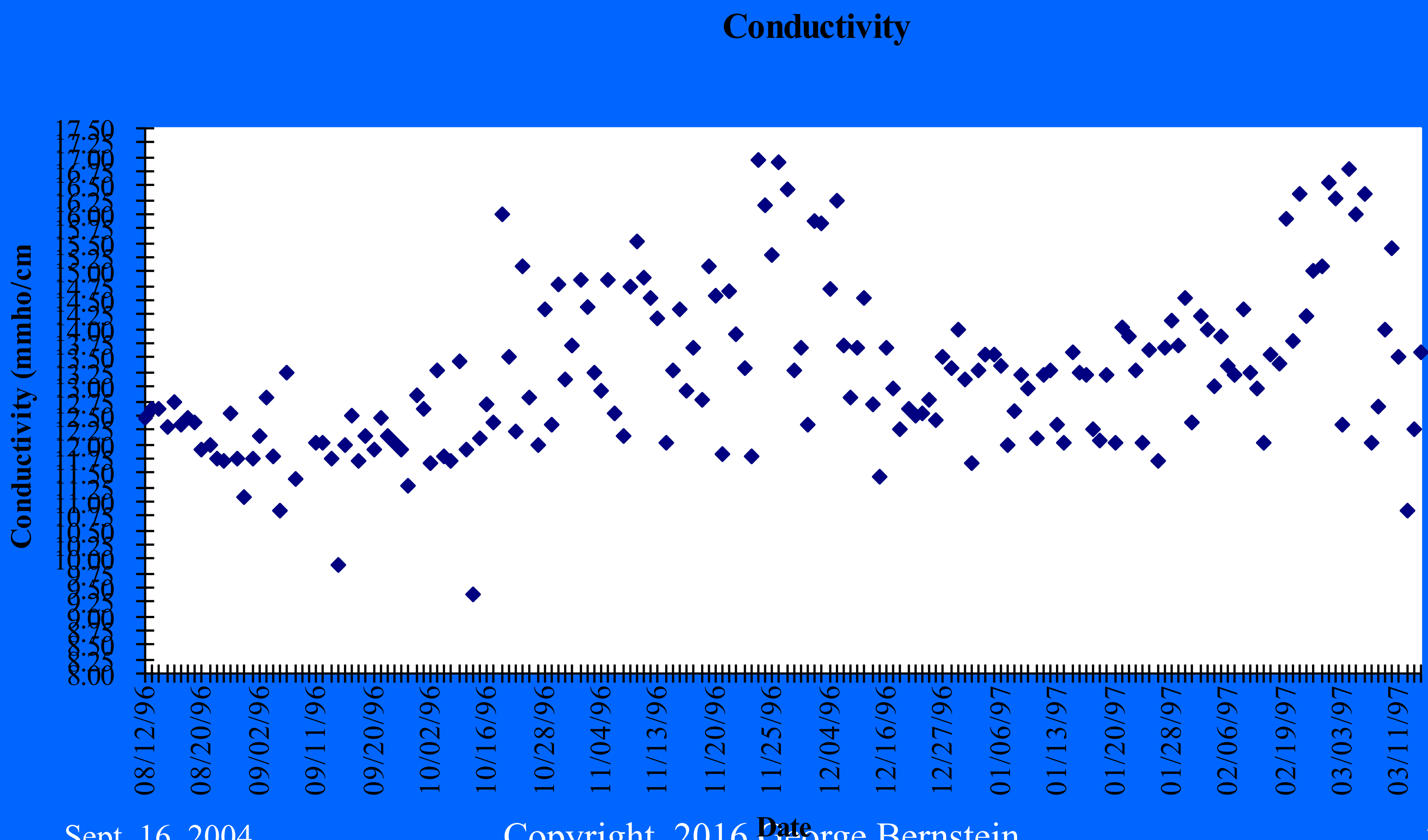
9. Collecting data is not even half the job

10. Expect 90% of your efforts in data cleansing

- Once you collect it, you have to:
  - organize it
  - analyze it correctly
  - act on it
- Technology has made it easy to collect volumes of data.
- Probably 95% of all data that is collected is archived without being used. This is expensive.

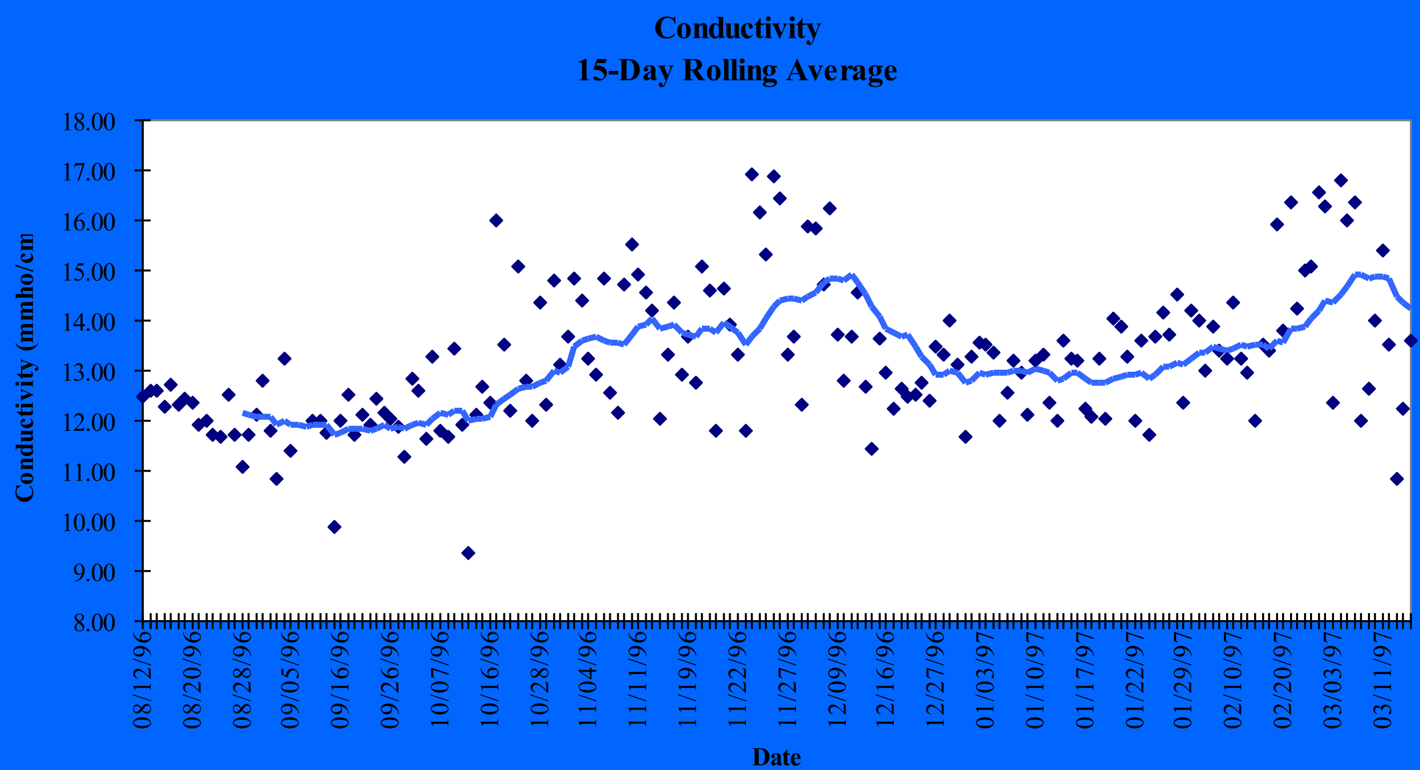
11. Use graphical techniques to

- Analyze your process
- Look for patterns, trends


Conductivity
Conductivity (mmho/cm
Date
08/12/96
08/20/96
09/02/96
09/11/96
09/20/96
10/02/96
10/16/96
10/28/96
11/04/96
11/13/96
11/20/96
11/25/96
12/04/96
12/16/96
12/27/96
01/06/97
01/13/97
01/20/97
01/28/97
02/06/97
02/19/97
03/03/97
03/11/97

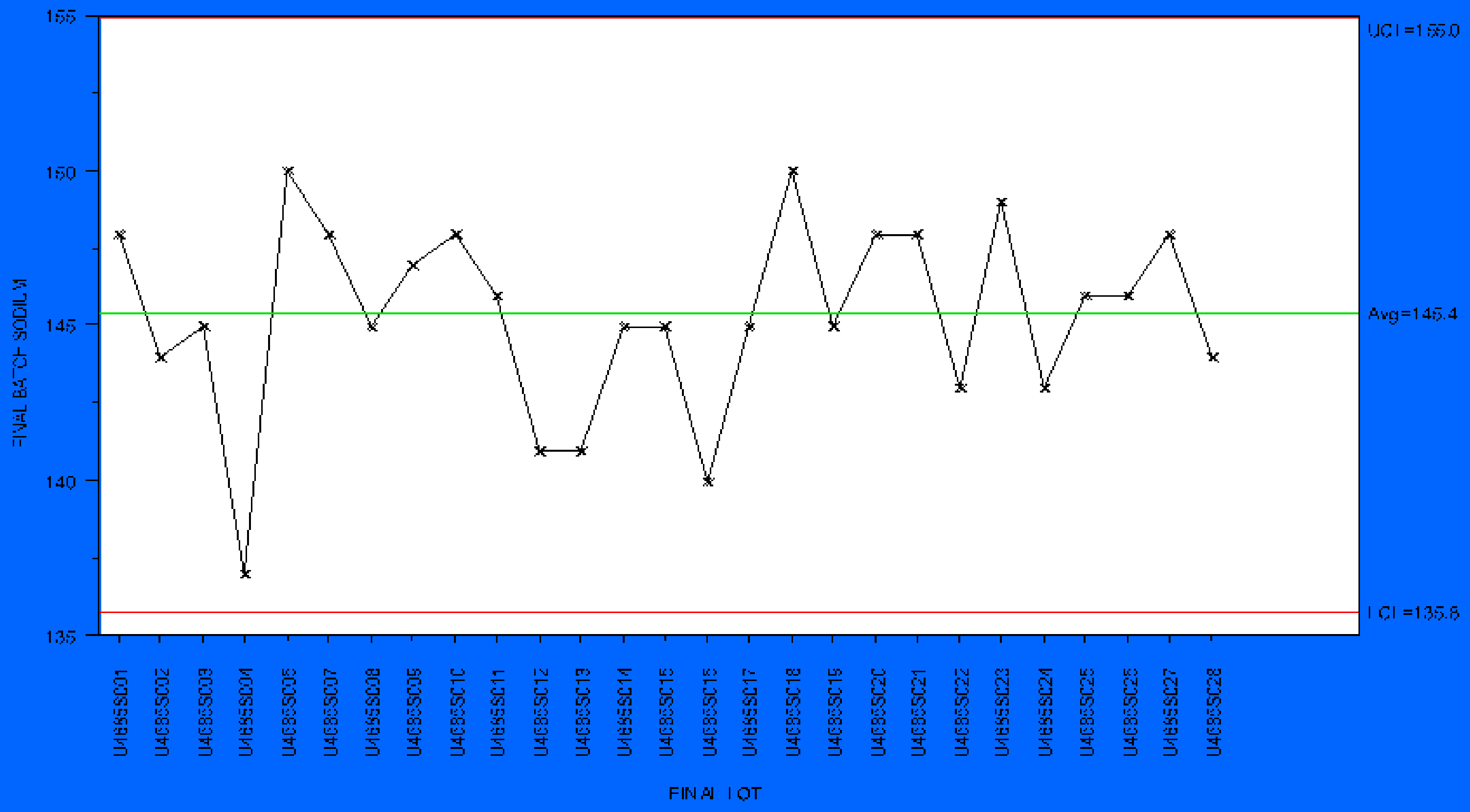
Raw material

- Seasonal variation?
- Effect on process?



12. Run charts are an invaluable way to track/analyze your process

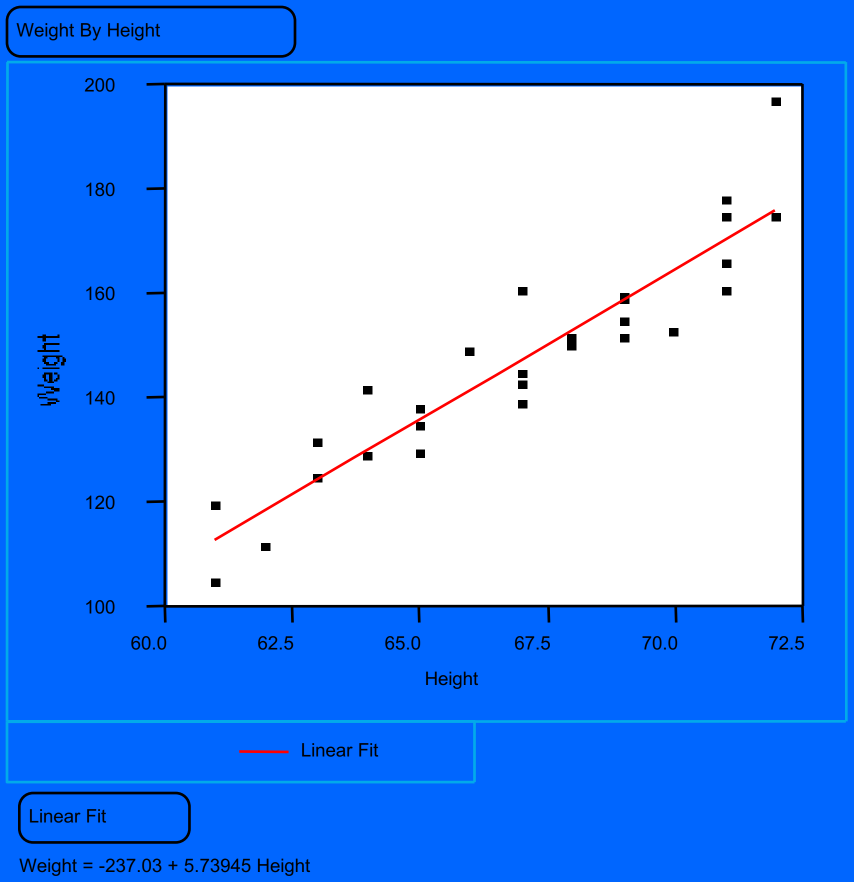
- Control charts are a way to:
  - Look at a process over time
  - Separate random variations from a trend
  - Identify a causative reason for a process shift
  - Track/trend your process so you can take action **before** a problem occurs/recurs.


155
150
145
140
135
UCL=155.0
Avg=145.4
LCL=135.8
FINAL LOT

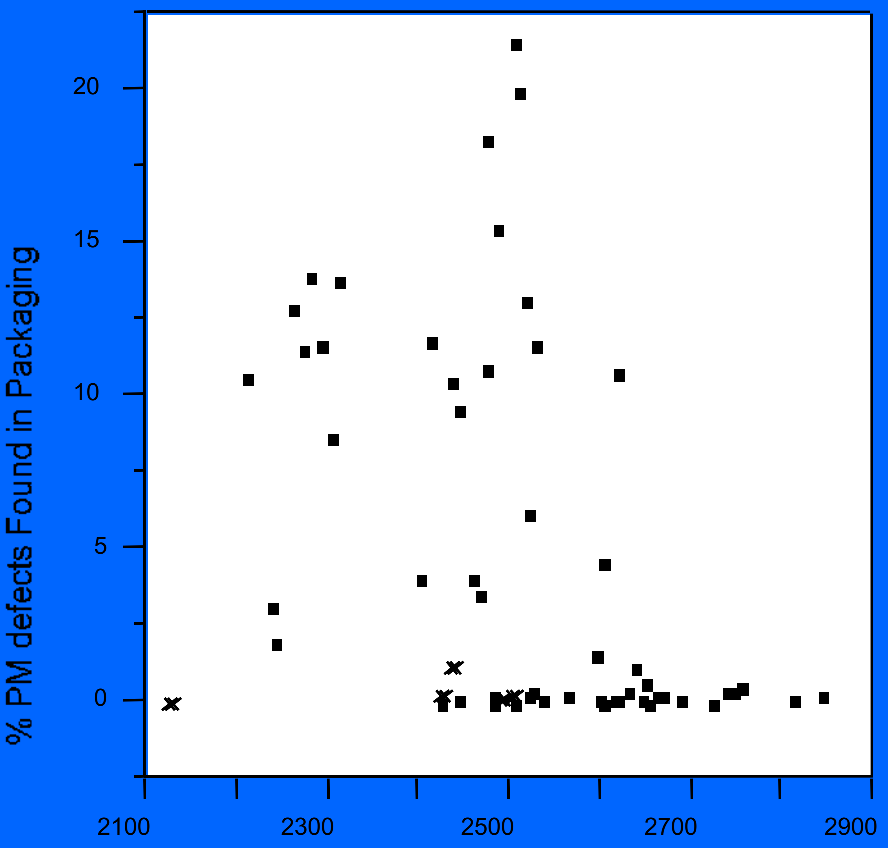
13. No matter what, statistics will not disprove the laws of nature

- Be able to explain what the statistics say
  - This is the purpose of a multifunctional team
  - Beware of data artifacts
  - Test your conclusions in the lab

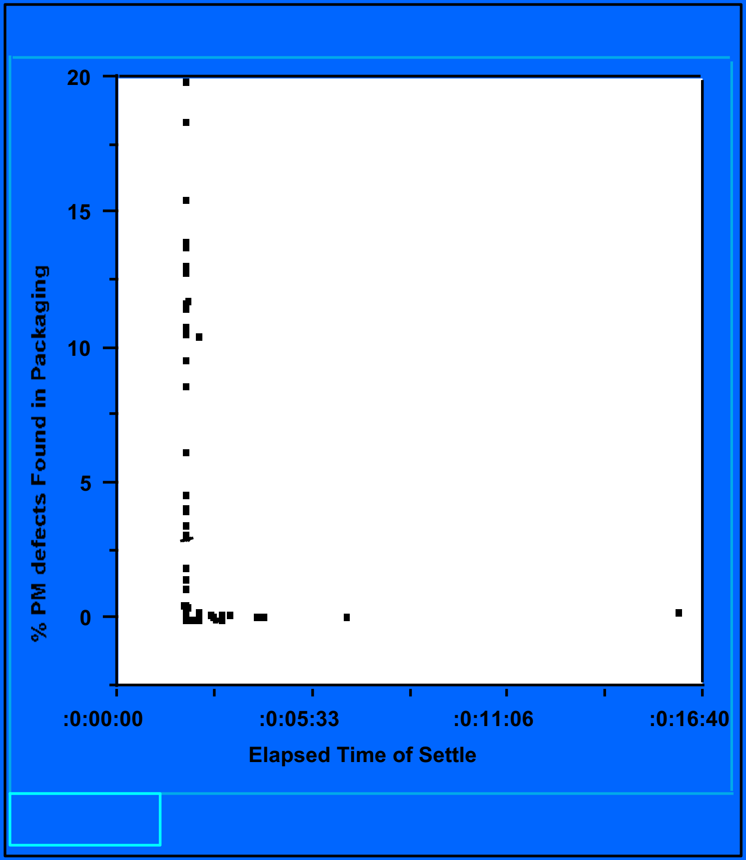
We wish that most data
Would look like this

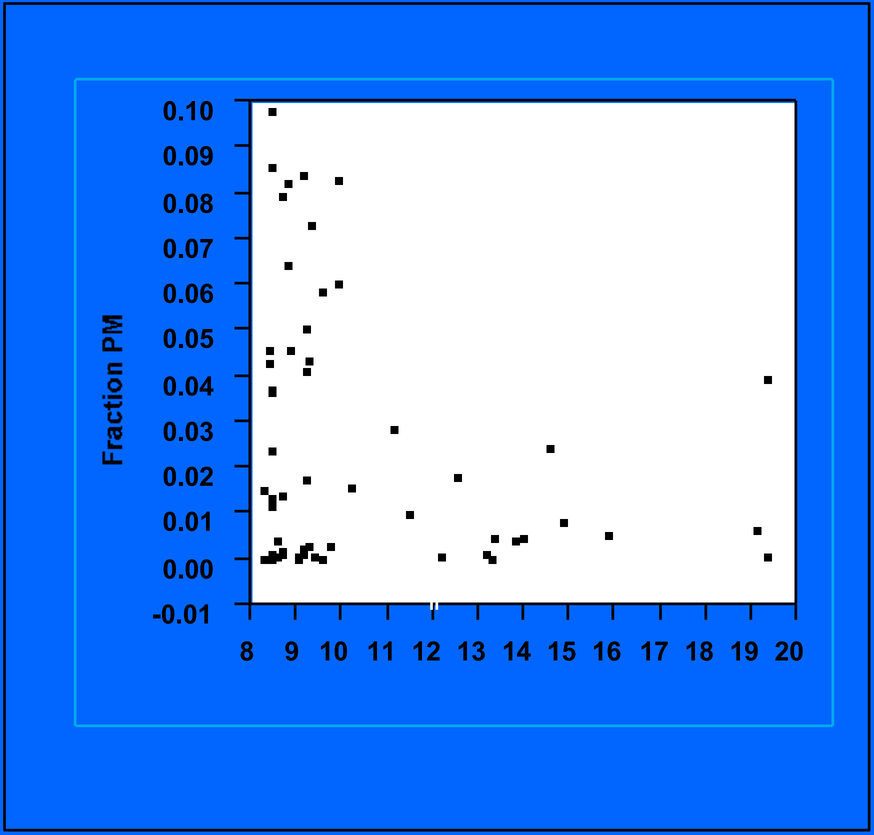


Unfortunately, is often looks like this



What conclusions can you draw?





Statistical Analysis

14. Purchasing a statistics package does not guarantee that you will be doing good statistics no matter how good the software

- Lots of statistical packages available:
  - Excel (Analyze-it)
  - SAS (ECO, JMP,
  - Systat
  - SPSS
  - S-Plus
  - Minitab

15. Know your statistics package

- Do you know what assumptions are being made by the software?
- Just because you got an answer doesn't mean it's the right one.
- Does the test expect the data to be normally distributed?
- If not, is using non-parametric methods more suitable?

16. One statistics tool doesn't fit all studies

Example:

- A manager is familiar with t-tests and wants to use it for comparing means of two groups for all studies.
  - That may not always be appropriate.
  - If there are more than 2 groups, an ANOVA may be more appropriate or ANCOVA, if there is an influencing variable that can be used as a covariate.

17. Data Modeling – keep it simple

- Given enough variables, you can fit anything
  - –Keep to the critical process parameters
  - –Consider SOME interaction terms
  - –Occam’s Razor
- Make sure variables chosen are controllable

18. Optimization involves experimentation

- Use Design of Experiments (DOE) to
  - Maximize the information
  - Minimize the number of experiments

19. You cannot predict a response outside of your solution space

- You may have found a local maximum but it may not
  - Be the optimal solution
  - Scale up
  - Represent stable conditions
- Which means more experiments

20. Document all work including data, study results, and conclusions to ensure compliance with GMPs