Minitab 15 Statistical Software

Common Practitioner Mistakes in Data Analysis

Jennifer Atlas, *Minitab Inc.*
jatlas@,minitab.com

Outline

1. Define Practitioner
2. Six Sigma Training
3. General Statistics
4. Basic Statistics
5. Regression
6. Quality Tools
7. Design of Experiments
8. Closing Remarks

Minitab 15 Statistical Software

1. Practitioners

- Who are they?
  - Anyone who needs to analyze data
  - Not sophisticated
  - Very little formal training in statistics
  - For many it is more about getting the answer

Minitab 15 Statistical Software

1. Practitioners

- Who are they?
  - Software is very important
  - Want a handful of tools they can use for any situation
  - Want very clear “rules of thumb”

Minitab 15 Statistical Software

2. Six Sigma Training

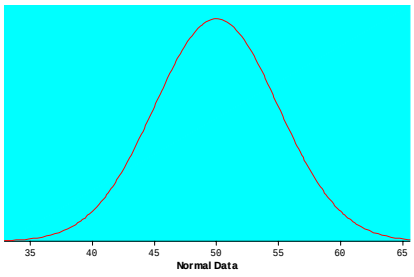
- Anywhere from 2 to 5 weeks of training
- Approximately 50% of time on statistics
- Many times there is no time to practice
- Many companies do not have a Statistician or MBB

Minitab 15 Statistical Software

3. General Statistics

Normality

- Data must be normal to do any analysis



Minitab 15 Statistical Software

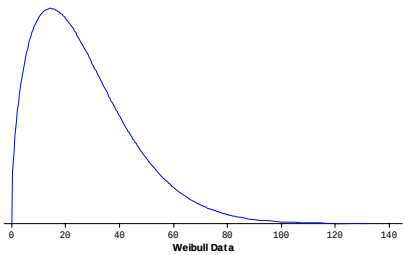
3. General Statistics

- Consultant: “If your data is not normal, find out why it is not normal and fix it.”
- Just a month ago: “If the data is not normal, you cannot do any statistics on it.”
- Most are not aware of other techniques: transformations, nonparametrics

Minitab 15 Statistical Software

3. General Statistics

- Ignore very skewed distributions with little attention paid to normality
- EX: Capability Analysis



Minitab 15 Statistical Software

3. General Statistics

One-sided Specification Limits

- EX: Ceiling tiles fire resistant to at least 20 feet
- Most are around 20 but some are much higher
- Confusing since this causes problems in the analysis---increase variation and forces distribution to be nonnormal

Minitab 15 Statistical Software

4. Basic Statistics

Confusing a CI and a PI

- EX. Weights of cereal boxes
- Want to talk about some sort of confidence on individual boxes
- Really thinking about capability

Minitab 15 Statistical Software

4. Basic Statistics

Overlapping CIs

- For a Two-Sample t-test or One-Way ANOVA, just do individual CIs on the means and see if they overlap
- The CIs can overlap by as much as 18% and still be significantly different

Minitab 15 Statistical Software

4. Basic Statistics

Continuous Versus Categorical

- Take continuous data and make it categorical---many times binary to use simpler tools such as proportions test or chi-square test
- Loss of information and power

Minitab 15 Statistical Software

4. Basic Statistics

Continuous Versus Categorical

- Treat categorical data as continuous
- EX: make defect/nondefect into 0/1
  - now it is continuous---right?????

Minitab 15 Statistical Software

5. Regression

R^2

- Too much emphasis on R^2
- 1. Should always be above 90%---agricultural, taste testing
- 2. Small R^2 but important predictors
- 3. Predictive Models --- with a large R^2 value but a low R^2 predicted

Minitab 15 Statistical Software

5. Regression

Assumptions

- Many people check the data for assumptions instead of the residuals from the model
- Most of the focus is on normality and not constant variance

Minitab 15 Statistical Software

6. Quality Tools

Gage R&R

- Using a sample of parts that is not representative of the population of parts
- 1. Pull off the process at relatively the same time-under estimate process variation (gage looks bad)
- 2. Ask JOE for parts (usually scrap laying around)---over estimate process variation (gage looks good)

Minitab 15 Statistical Software

6. Quality Tools

Control Charts

- Confusing Specification Limits and Control Limits
- Not taking a large enough base period to get good estimates of the control limits
- Not Fixing the Control Limits once they have a stable process

Minitab 15 Statistical Software

7. Design of Experiments

Variation

- Difference between replicates and repeats
- Are they really being taught that they must reset between runs?
- Error based on repeats is small

Minitab 15 Statistical Software

7. Design of Experiments

Screening

- Factorial designs are only for screening
- You cannot find the best settings from a factorial design
- You cannot build a model from a factorial design

Minitab 15 Statistical Software

7. Design of Experiments

Reducing Models

- Many are taught to reduce the model one term at a time

  --- biasing the error term

- Orthogonal designs

Minitab 15 Statistical Software

7. Design of Experiments

Center Points

- Many times they put all of the center points together at the end of the design---repeats
- Copying a design of a popular textbook such as Montgomery
- “Why did I run these center points?”

Minitab 15 Statistical Software

7. Design of Experiments

Advanced Tools

- Handling of Covariates in the Analysis
- Binary Response Data---logistic regression
- Hard-to-Change Factors
- Mixture Experiments (with process variables)

Minitab 15 Statistical Software

8. Closing Remarks

- Movements such as Six Sigma have brought statistics to the mainstream
- Majority of people using statistics are not Statisticians (as it should be!)
- Very little formal training
- Doing some “wrong” things with statistics versus not even looking at data

Minitab 15 Statistical Software

8. Closing Remarks

- Our customers are a pretty good sample of quality professionals, engineers, scientists, financial analysts, Belts of all sorts
- I have shared some of the things I have seen in the last 8 years

THANK YOU!

Minitab 15 Statistical Software