



Beyond the Formula III

**What Makes Statistics Different Than
Mathematics?**

By Patti Cyr

Eastman Kodak Company

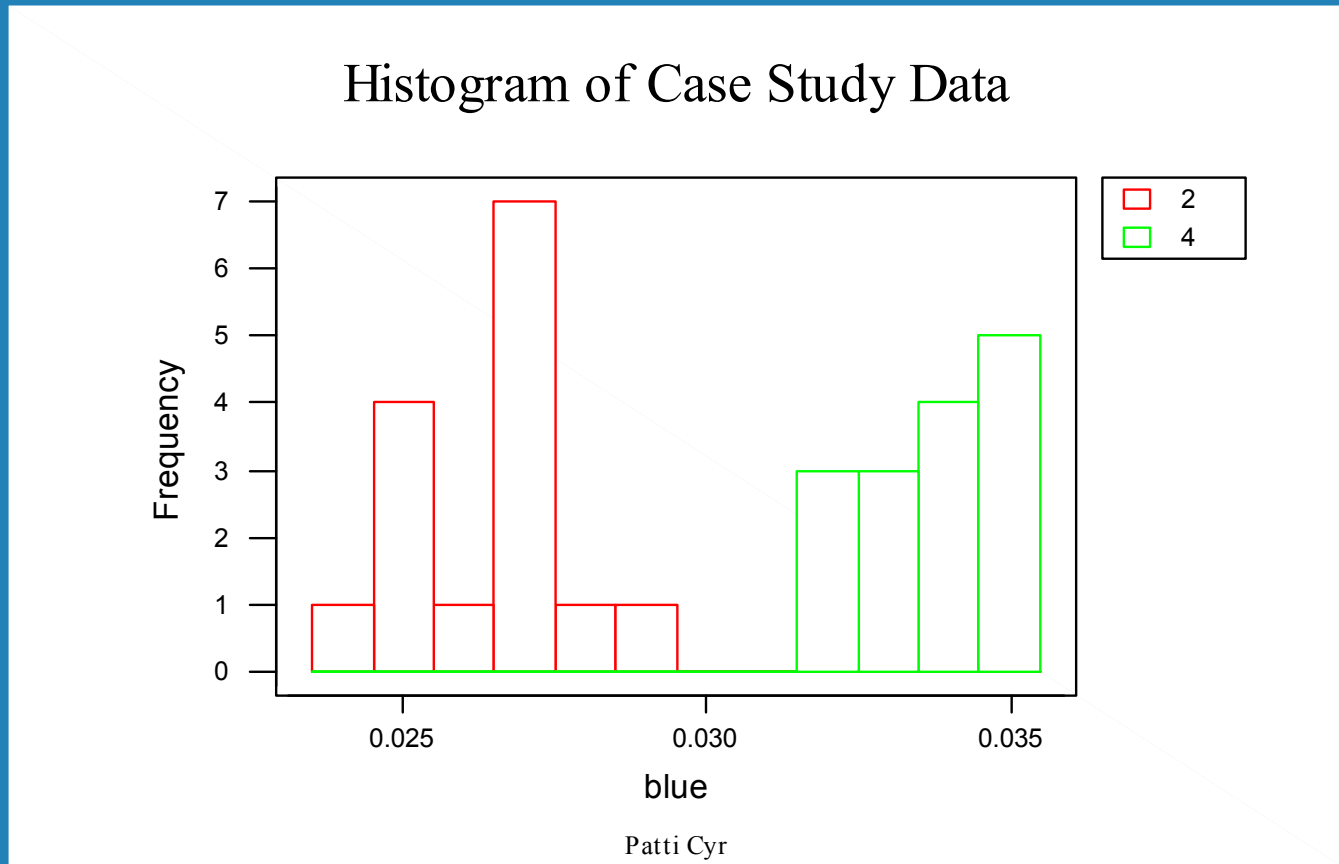
Case Study Data

❧ **Are these two groups of data significantly different?**

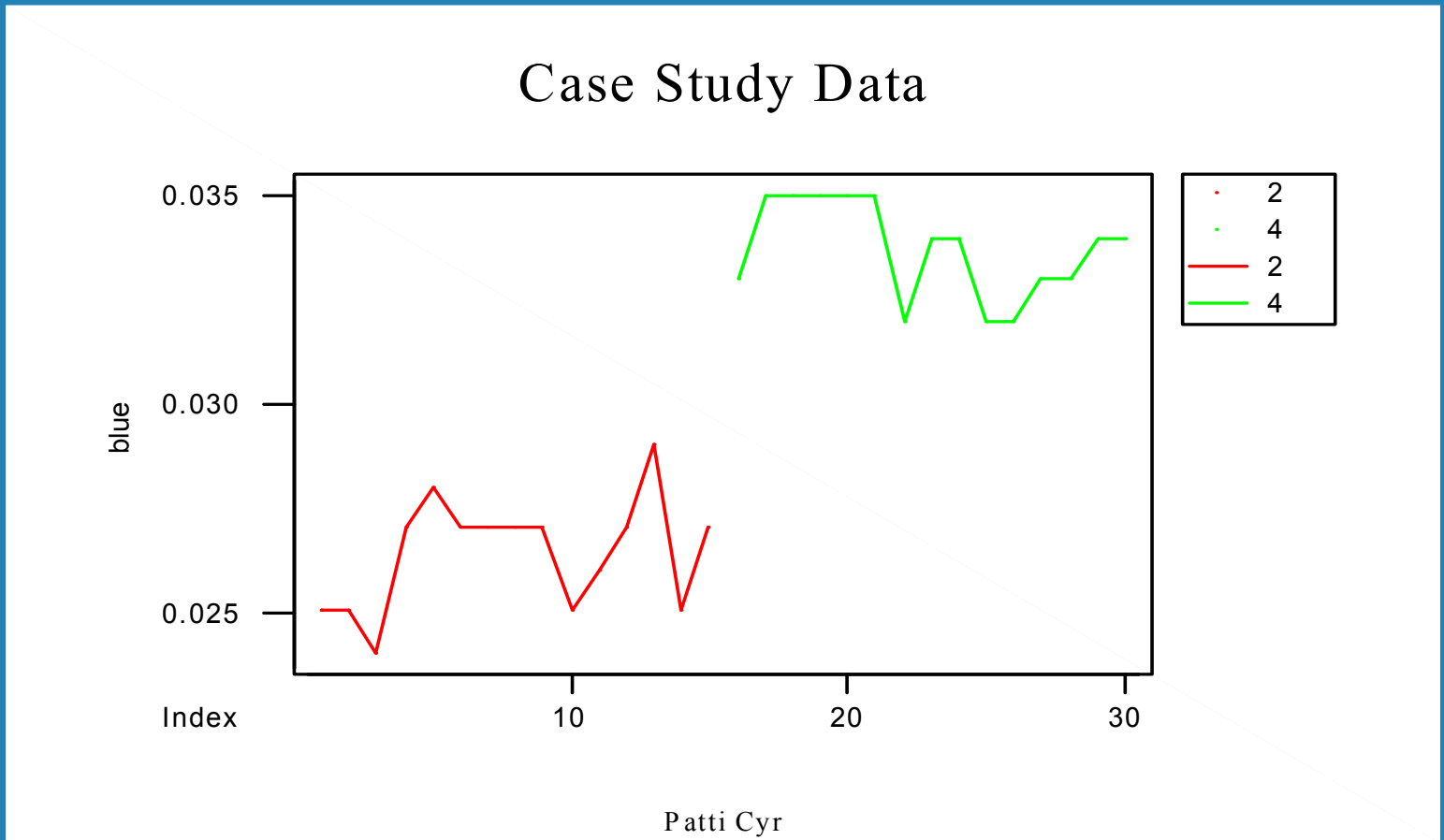
❧ **Start the analysis by plotting the data**

group 1	group 2
0.029	0.035
0.028	0.035
0.027	0.035
0.027	0.035
0.027	0.035
0.027	0.034
0.027	0.034
0.027	0.034
0.027	0.034
0.027	0.034
0.026	0.033
0.025	0.033
0.025	0.033
0.025	0.032
0.025	0.032
0.024	0.032

Case Study



Case Study Data



Case Study Data

- ∞ **Examining the plots, the two groups look different**
- ∞ **A Student's t-test can be used to test for statistical significance**



Application of the Formula

- ∞ **Application of the formula indicates a significant difference between the two groups**
- ∞ **This is the mathematical side of the analysis**

The Statistical Aspect

- ⌚ **What conclusion should be drawn from the statistical significance?**
- ⌚ **It depends on what question is being answered, and that depends on how the data is structured**

Facts About the Case Study

- ⌚ **The two groups represent measurements off two different instruments**
- ⌚ **The two groups of data have been generated measuring the same sample over and over**
- ⌚ **The sample was measured three times per day on five different days on each instrument**



Facts About the Case Study Data

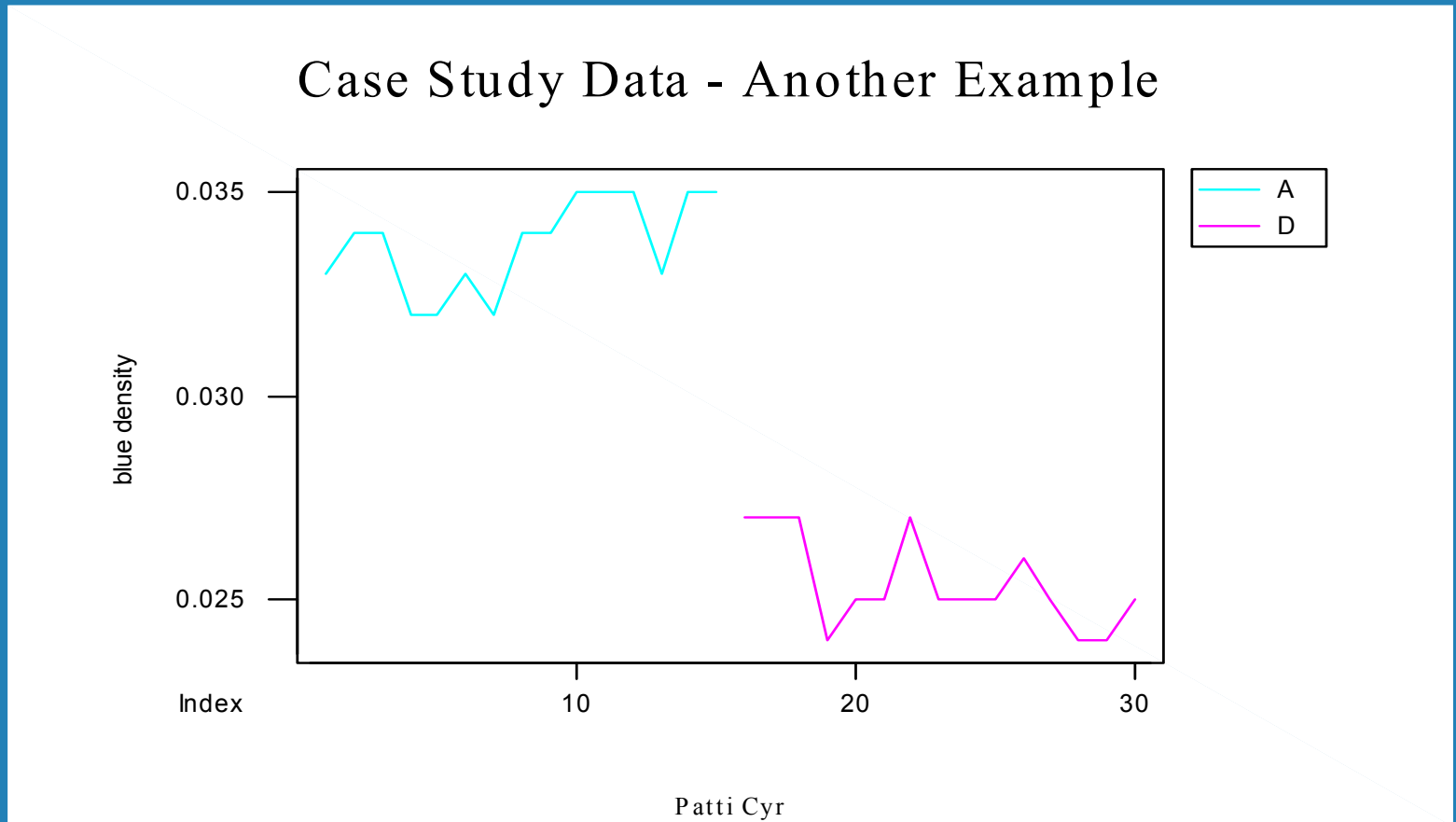
- ⌚ **The difference detected by the formula is actually a difference in measurement devices**



The Essence of the Problem

- ⌚ **What would happen if two different samples were measured, one on each instrument?**
- ⌚ **Could you say whether differences seen were due to the sample or the instrument?**

For Example



What Makes Statistics Different Than Mathematics?

- ⌚ Before drawing conclusions based on the results from the formula, careful consideration must be paid to the source of the data and the source of the measurements

What Makes Statistics Different Than Mathematics?

- ⌚ Questions to address who, what, where, when, why, and how are needed in order to know which question is actually being answered

The Data of the Case Study

- ◊ **Who decided the samples were needed?**
 - **Engineers who were trying to understand differences in measured values between four different instruments**
- ◊ **What were they interested to know?**
 - **They wanted to know if differences exist**
- ◊ **Where were the samples obtained?**
 - **Stock samples of typical materials were used**

The Data of the Case Study

∞ **When were the samples taken?**

- **All samples were collected at the same time**

∞ **Why were these particular samples taken?**

- **These four samples spanned the 'normal' range of density**

∞ **How were the samples stored?**

- **The samples were mounted in slide covers for protection**

The Measurements of the Case Study Data

⌚ Who made the measurements?

- Separate people tested each group in this example

⌚ What was measured?

- Optical blue density of 'clear' samples

⌚ Where were the samples measured?

- The measurements were made in the center of the mounted slide

The Measurements of the Case Study Data

- ❏ **When were the measurements done?**
 - Each set of measurements was made three reading per day over five days
- ❏ **Why were these particular tests selected?**
 - Internal customers questioned differences in blue density
- ❏ **How were the measurements done?**
 - Using standard procedure

The Data Provided

- ∩ **The entire dataset has been provided for the conference**
- ∩ **Four samples are included**
- ∩ **Four instruments are included**
- ∩ **Measurements of visual, red, green, and blue densities are included**
- ∩ **Measurements for instruments 1, 2, and 4 all done by the same person**

The Data Provided

- ⌚ **The column labeled 'test' gives the order of measurements each day**
- ⌚ **The column labeled 'day' gives the order of data collection by time**
- ⌚ **I hope you can use this data in many ways to illustrate the concepts discussed in this presentation**



The Last Thought

- ⌚ **An example in Box, Hunter, and Hunter discusses the evaluation of sole material for shoes for children**
- ⌚ **How should you use ten children to compare two different sole materials?**

What are the Options?

Ω Option 1 -

- **Have five children wear shoes made with material A**
- **Have five children wear shoes made with material B**
- **Compare the tread wear differences between the two groups after a predetermined amount of time**

What are the Options?

Ω Option 2 -

- **Have each child wear a pair of shoes made with one foot material A and the other foot material B**
- **Make sure to randomize which feet get which material**
- **Compare the tread wear differences between the feet of each child after a predetermined amount of time**

Why is Option 2 a Better Choice?

- ⌚ **The first option includes the differences in activity level of the children with the tread wear differences**
- ⌚ **The second option ensures that both materials see the same activity so differences in feet are differences in sole material wear**

Summary

- ⌚ **The formulas used in statistics are valid**
- ⌚ **They yield an evaluation of significance when applied**
- ⌚ **The conclusions you draw based on the results must be tempered with the information content of the data**
- ⌚ **How was the data collected?**