Statistics for Science Fair – Some Basics

Statistics is Your Friend!

- Begin with the end in mind. Consider statistics when you are:
  - Developing your hypothesis
  - Designing your experiment
  - Analyzing your data
  - Framing your conclusions
  - Presenting your work

Case Study

- Student wanted to study how the chemical composition of rocks was changed by meteorite impacts.
- For one part of the experiment, the student used two different techniques to compare the chemical composition of rocks before and after impact.

3 Principles of Design

- **Replication**
  - Aim for at least 30 replicates in each group
- **Control**
- **Randomization**
  - Let chance choose for you (simple random sample)
  - Consider using a coin, a table of random numbers, a dice, etc.

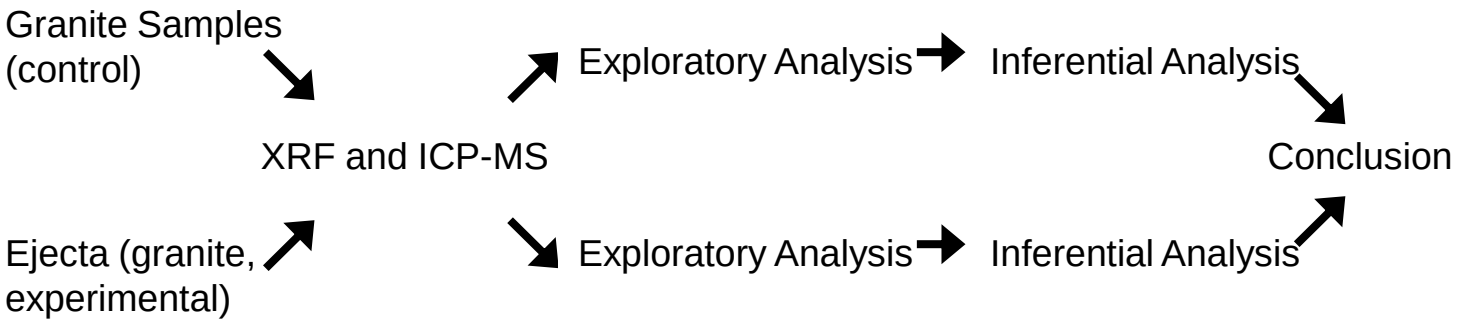
Design: An Example

- 100 granite samples were collected and need to be divided into two groups.
- Why randomize?
- How to randomize?
  - Number from 000 to 099. First 50 random numbers go to control group.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	8	0	9	4	2	5	2	5	8	2	4	7	1	3	4	7	7	4	3	3
2	3	5	6	3	2	1	9	8	8	2	1	1	9	0	4	5	2	6	1	8
3	1	3	3	0	6	3	3	1	3	7	5	3	9	6	9	3	8	7	3	8
4	3	5	6	5	0	0	1	6	2	2	4	3	6	4	3	2	4	7	9	6
5	7	8	5	0	5	9	2	5	5	5	8	8	7	3	1	1	2	1	9	2

Showing Your Design

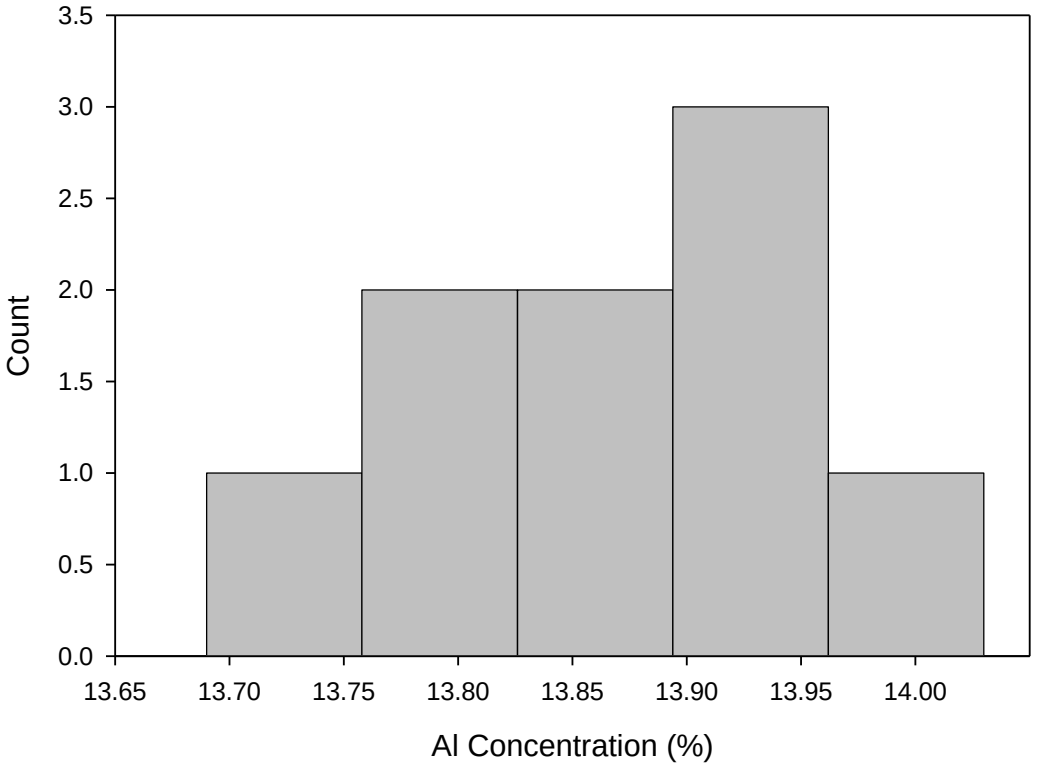
- Using a flow chart can help effectively convey information about your experimental design



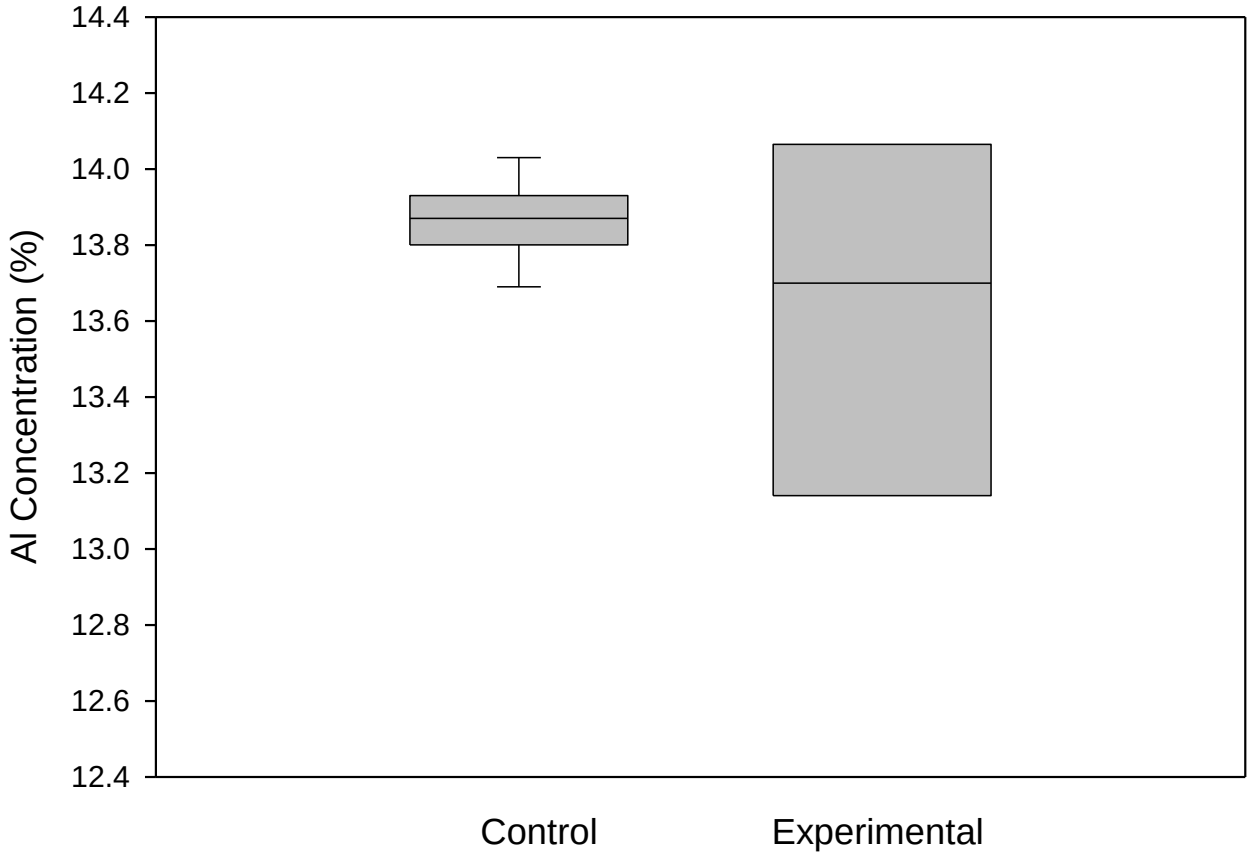
Exploratory Data Analysis

- The first step is analyzing data is called exploratory data analysis (EDA)
- First, use graphs to explore the data
- Second, use numerical measures to describe the data
- Never skip EDA—it allows you to decide how to approach the inferential analysis

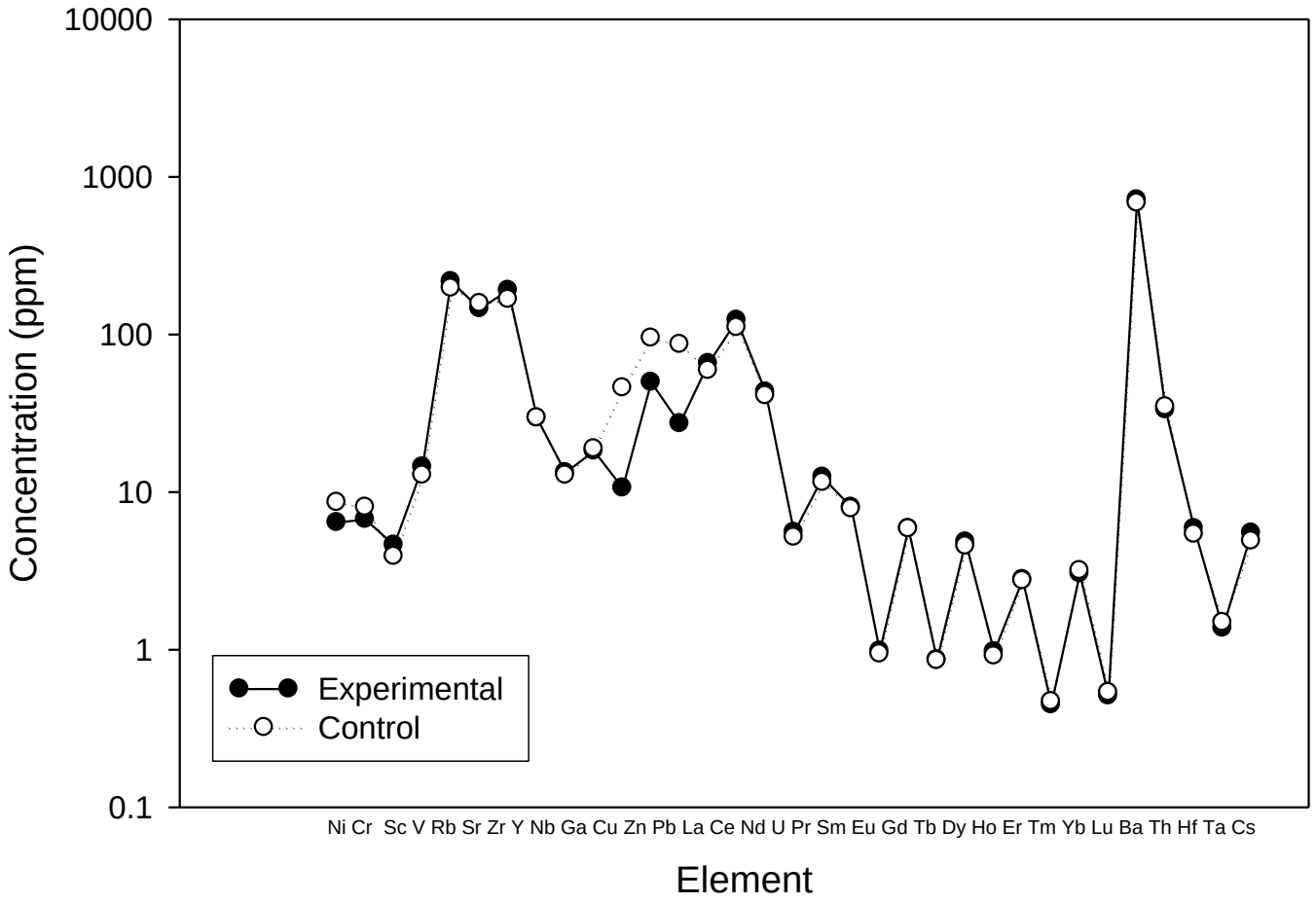
Histogram


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
13.65
13.70
13.75
13.80
13.85
13.90
13.95
14.00
Al Concentration (%)

Side-by-Side Boxplot



Spidergram: Elemental Concentration in Pre- and Post-Impact Materials



Moving Toward Inference

- Inferential analysis uses the language of probability
- How we approach inference is based on exploratory data analysis
- We say something is statistically significant if the probability that the observed result is due to chance is less than some specified probability

Hypothesis Testing

- Hypothesis testing is one of the huge ideas of inference
- We state a null hypothesis, H_0, and an alternative hypothesis, H_a
- **Granite Example:**
  - H_0: $\mu_{control} = \mu_{experimental}$
  - H_a: $\mu_{control} \neq \mu_{experimental}$

Types of Hypothesis Tests

- ***t*-tests** allow you to compare two means
- ***Z*-tests** allow you to compare two proportions
- **Chi-squared tests** allow you to determine the goodness of fit of data to a given distribution

The P-value

- The output of a hypothesis test is a P-value, which is the probability that the observed difference between the two groups is due to chance variation.
- Define a cutoff of significance. $P \leq 0.05$ is often used.
- If the p-value of the hypothesis test is less than the specified value, the difference is statistically significant

Back to Our Case Study

- EDA showed that Zn appeared to be different between the control and experimental studies. Is that difference significant at the 0.05 level?
- P = 0.001
- The result is statistically significant at the 0.05 level. The difference between the two groups is probably not due to chance variation.

Technology

- TI 83 Plus, 84 Plus, and 89 calculators
- Microsoft Excel
- Minitab
- R (free)
- S-PLUS
- MYSTAT (free)
- www.whfreeman.com/tps3e

Statistics is Your Friend!

- Begin with the end in mind. Consider statistics when you are:
  - Developing your hypothesis
  - Designing your experiment
  - Analyzing your data
  - Framing your conclusions
  - Presenting your work
- Weave statistics into your design, your analysis, your written work, and presentation