Practical Stats
Statistics, down to earth

What's New in Statistics?

1. Permutation Tests

Dennis R. Helsel
PracticalStats.com

1

Practical Stats
Statistics, down to earth

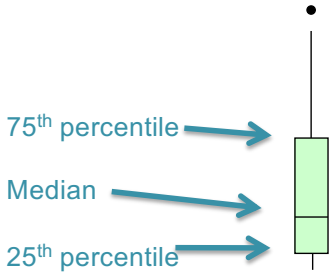
What's New that My Employees Should Know About?

1. Permutation Tests. Never worry about a normal distribution again.
2. Free software for environmental statistics
3. Better methods for handling nondetect data
4. New and better methods for finding the best regression line.

2

Practical Stats
Statistics, down to earth

Boxplots quickly show how close data are to a normal distribution



Skewness
(Top portion is larger than bottom portion - -not a normal distribution)

Symmetric data
(Top portion similar size to bottom portion – may approximate a normal distribution)

3

Practical Stats
Statistics, down to earth

1. Permutation Tests. Never worry about a normal distribution again.

Field data are usually skewed (not a normal distribution)

Parametric tests assume data look like a normal distribution. When this is not true they often do not find differences that are there

Nonparametric tests don't have a problem ("more power") with skewed data, but test differences in percentiles (medians)

How to test differences in means with skewed data?

4

Practical Stats
Statistics, down to earth

Bottom line: the p-value

p-value is the 'universal translator' for a statistical test

Answers the question "how believable is 'no difference between groups' – no signal?"

p = 0.03 means that there is only a 3 in 100 chance of seeing the signal strength of my data when there really is no difference

Tradition is that when the probability is less than 5% (0.05), reject 'no signal' and find that there is a difference, there is a correlation…etc.

5

Practical Stats
Statistics, down to earth

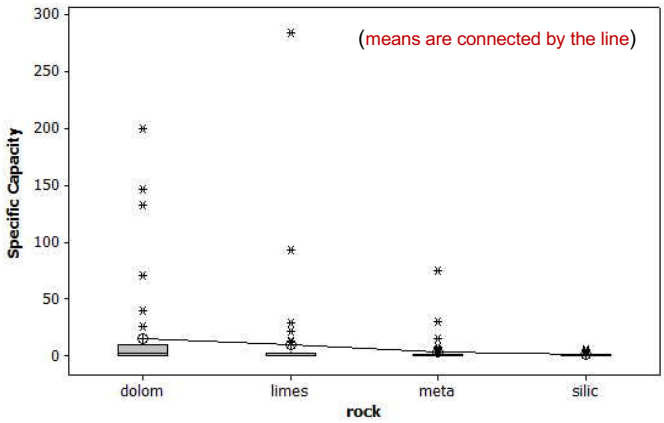
Testing Means: Old Method #1

Hope the parametric method works

Is specific capacity the same in all 4 rock types? Data are non-normal – should worry that p-values are artificially high.

50 obs per group

t-test: p=0.07
Conclude "not evidence that groups means differ"



6

Practical Stats
Statistics, down to earth

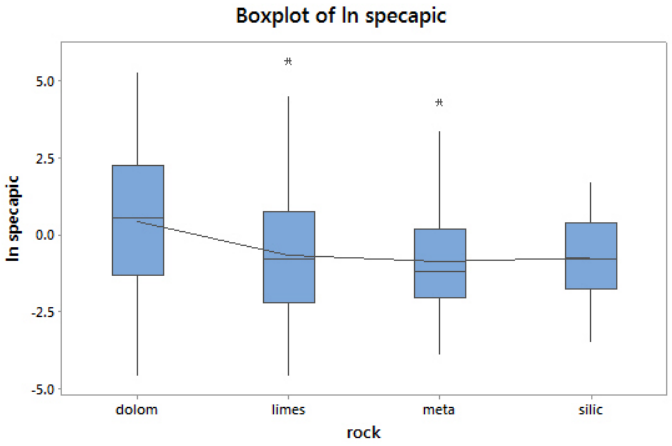
Old Method #2: test using logarithms

Now data look much more like a normal distribution, so I can use ANOVA (on logs).

(means are connected by the line)

p=0.007
Mean logs are significantly different

This tests whether the geometric means or MEDIANS differ, not means



7

Practical Stats
Statistics, down to earth

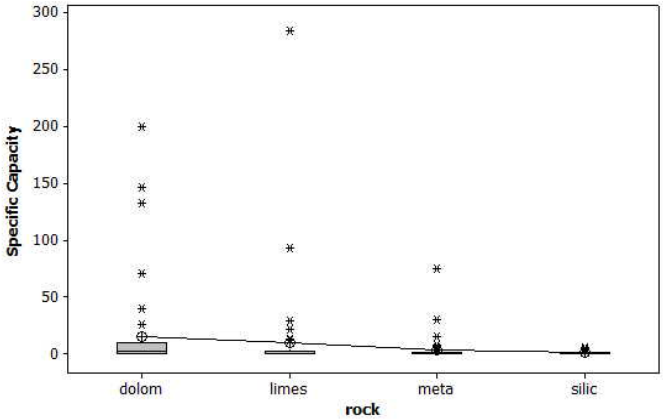
Old Method #3

Use a nonparametric test

Great test to determine if MEDIANS differ. Doesn't test means.

(means are connected by the line)

KW test:
p=0.009
Medians are significantly different



8

Practical Stats
Statistics, down to earth

When should the mean be of interest?

Mass, volume, totals

Cumulative long-term exposure

When regulations specify a mean

Estimating a mean and testing for differences between means (background vs. possible contamination) for skewed data has been one of the hardest things to do well in statistics

9

Practical Stats
Statistics, down to earth

If the mean is the objective, use a permutation test

Makes no distributional assumptions about the population sampled. (Does not require assumption of normality)

Does not rely on the Central Limit Theorem

Uses only the observed data and all possible rearrangements or permutations of the data

10

Practical Stats
Statistics, down to earth

If these are the original data

SP CAPACITY	ROCK
6	Dolomite
5	Dolomite
10	Dolomite
16	Limestone
8	Limestone
22	Limestone
18	Limestone

11

Practical Stats
Statistics, down to earth

Shuffle the group names (1 of 1000s possibilities)

Do this thousands of times

SP CAPACITY	ROCK
6	Limestone
5	Dolomite
10	Limestone
16	Dolomite
8	Limestone
22	Dolomite
18	Limestone

12

Practical Stats
Statistics, down to earth

Permutation Test for comparing group means

If there is no difference between group means, the data could be randomly reassigned to any group.

"Shuffle" the SITE names many times.

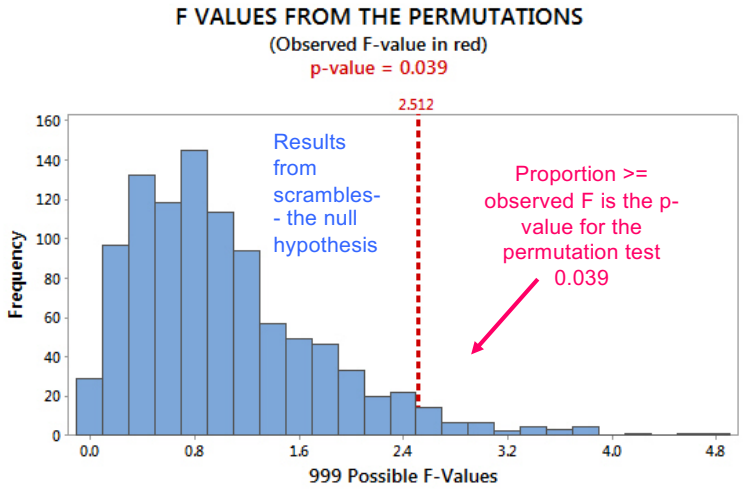
Compute a test result for each shuffle.

p-value equals percent of shuffled reps with a test stat >= the one result for the original data

13

Practical Stats
Statistics, down to earth

Permutation Test for Difference between Mean Specific Capacity



14

Practical Stats
Statistics, down to earth

Summary: Permutation Tests

Can be used instead of t-test, ANOVA, or other parametric tests, avoiding assumption of normal distribution

Are in several statistical software packages today

More power to see differences between means for skewed data than parametric tests

Finally a method that can see differences between means for the skewed data common to environmental sciences

15

Practical Stats
Statistics, down to earth

Resources – more info on these topics

Free Newsletters
http://www.PracticalStats.com/news/

Webinars and Training Courses
https://PracticalStats.teachable.com

Textbooks
http://practicalstats.com/books/

1. Statistical Methods in Water Resources 2nd Edition (2020) link to download on the Practical Stats Books page.
2. Statistics for Censored Environmental Data using Minitab and R Wiley (2012)

16

Practical Stats
Statistics, down to earth

Questions?

For answers to questions from the live broadcast --
see the Q&A pdf file underneath this video

17

Practical Stats
Statistics, down to earth

More Details Are On Our OnlineTraining Site

https://practicalstats.teachable.com

where you'll fine more videos and courses.

Dennis Helsel
PracticalStats.com

18