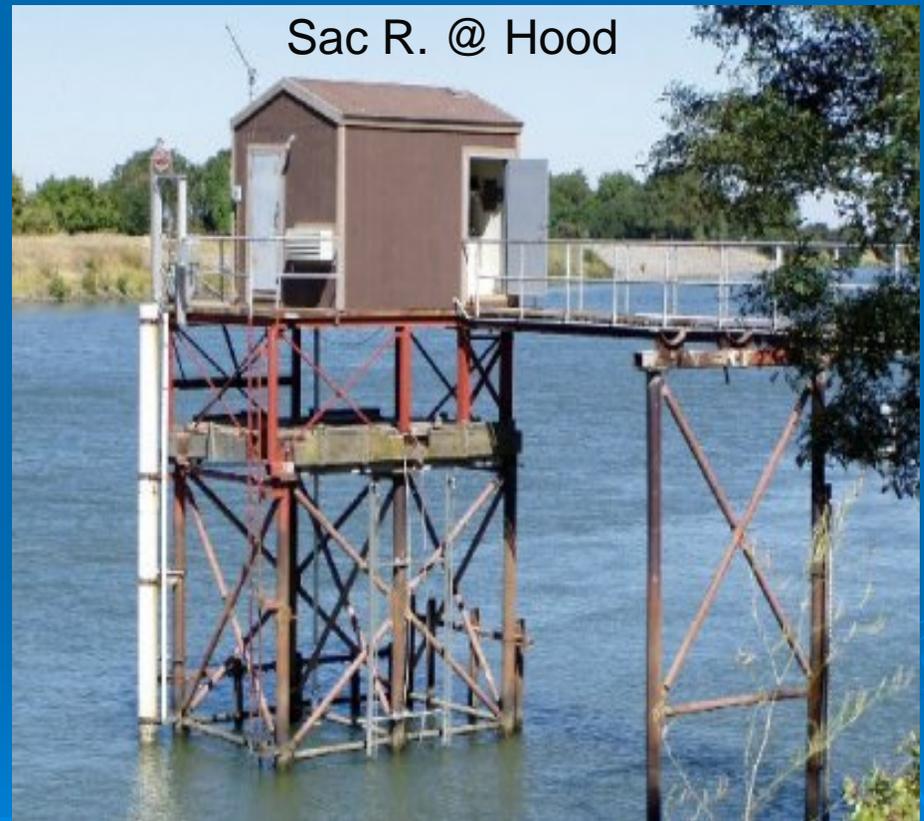
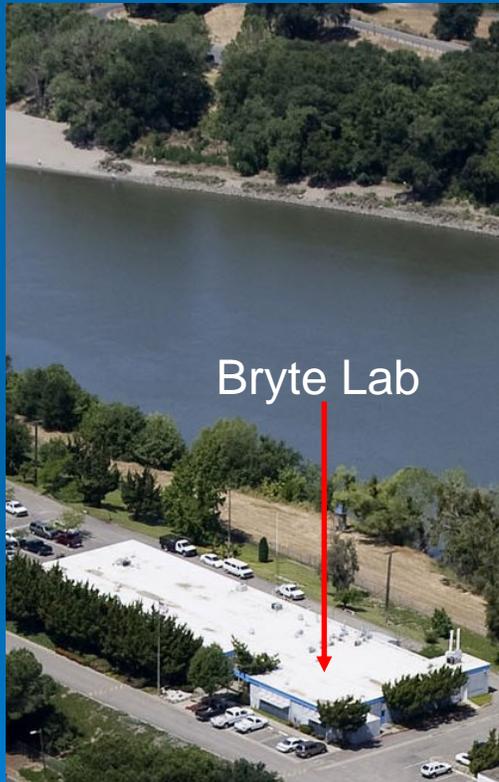


# Comparisons of Field & Bryte Lab Carbon Data



Murage Ngatia: April 21, 2009

# Field vs Field: Shimadzu vs Sievers

Shimadzu  
4100 (*field*)

Sievers  
800 (*field*)



# Field vs Lab: OI 1010 vs Field

OI 1010  
*Lab Instrument*



# Why Care About Organic Carbon?

- In drinking water treatment, OC reacts with disinfectants to form DBPs
- OC (organic matter) is one of the most important components in the transfer & transport of pesticides in soil & water
- OC plays a role in climate change: Carbon Capture Farming anyone?

# Field Data Summary: Apr '02- Apr '07

<b>Instrument</b>	<b>Method</b>	<b>*# of Analyses</b>	<b>Daily data capture</b>	<b>Common Days</b>
Sievers 800/900	UV/ pers Ox	110, 564	94%	1665
Shimadzu 4100	Combustion	62, 583	95%	1665

Common Days Sievers, Shimadzu & Bryte Lab = 169

# Inter Comparisons of 3 Instruments' Data

---

- Are Sievers & Shima data intercomparable?
- Are Sievers/Shima & lab data comparable?

(Essentially accuracy & precisions questions)

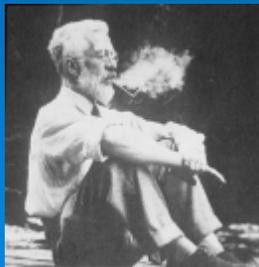
---

- I will only discuss Reproducibility using:  
Classical vs Equivalence  
Statistical Tests

# Classical Statistics

- Also called “frequentist” statistics or “Null Hypothesis Statistical Testing” (NHST)
- **Implicitly** involve a null hypothesis,  $H_0: A-B = 0$ 
  - t-test  $H_0$  : No difference between 2 instruments
  - Regression analysis,  $H_0$ : Slope is zero
  - Trend analysis  $H_0$ : No trend (trend slope is zero)
- Hybrid of opposing ideas:

R.  
Fisher  
( $p$  values &  $H_0$ )



J.  
Neyman



Pearson



(Level of significance,  $\alpha$ , &  $H_a$ )

# Classical Statistics

- NHST are **overwhelmingly** utilized but are not universally accepted

## Criticisms:

- In the real world, all null hypotheses are false!  
“All we know about the world teaches us that the effects of A and B are always different—in some decimal place—for any A and B. Thus asking ‘Are the effects different?’ is foolish” (Tukey 1991).
- $p$  values are arbitrary: sample size dependent
- Statistical significance does not mean practical significance i.e stat. signif. results may be trivial!

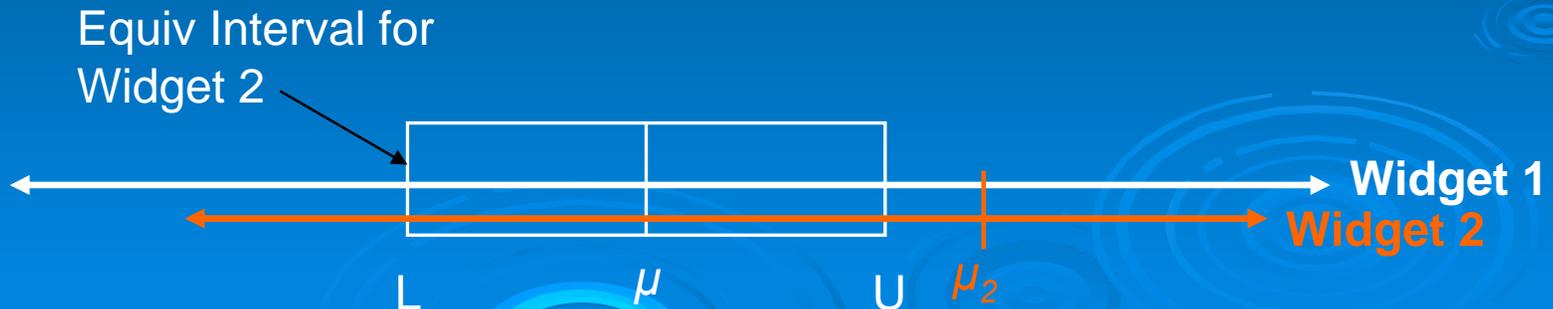
# Equivalence testing

Admit that some differences are trivial!

## Procedure:

- Determine a baseline
- Pre-define a difference of no importance/interest (equivalence interval, 'goal posts')
- Conduct the test

Example: Compare 2 widgets: 1 & 2



# Very Brief Description of the Mechanics of Equivalence Tests

- Note: There are more than one ways to test equivalence
- Example presented utilizes two one sided  $t$ -tests (TOST), the preferred procedure
- It has been proven that the 2-one sided  $t$ -tests are same as testing whether the commonly calculated 90% CI is within the equivalence interval

# Equivalence

## Procedure

- Sievers = baseline
- Equiv. Interval: 20%
- Interval based on lab precision

## Equivalence Tests

- 1)  $H_0$ : Shimadzu OC are equivalent to Sievers
2.  $H_0$ : Shimadzu, Sievers & Bryte OC are equivalent

# NHST

## Procedure

- 1) Paired t-test ( $n=1665$ )  
( $H_0$ : Sievers OC = Shima OC;
- 2) Kruskal-Wallis ANOVA:  
 $H_0$ : Sievers, Shimadzu, Bryte OC data are equal:  
i.e.  $H_0$ : Sievers OC = Shimadzu OC = Bryte Lab OC

# Results of NHST

## ➤ Paired *t*-test:

Shimadzu daily averages are statistically significantly different from Sievers

( $p < 0.01$ ,  $n = 1665$ )

## ➤ K-W ANOVA

Shimadzu data were significantly different from both Sievers & Bryte data

( $n = 169$ )

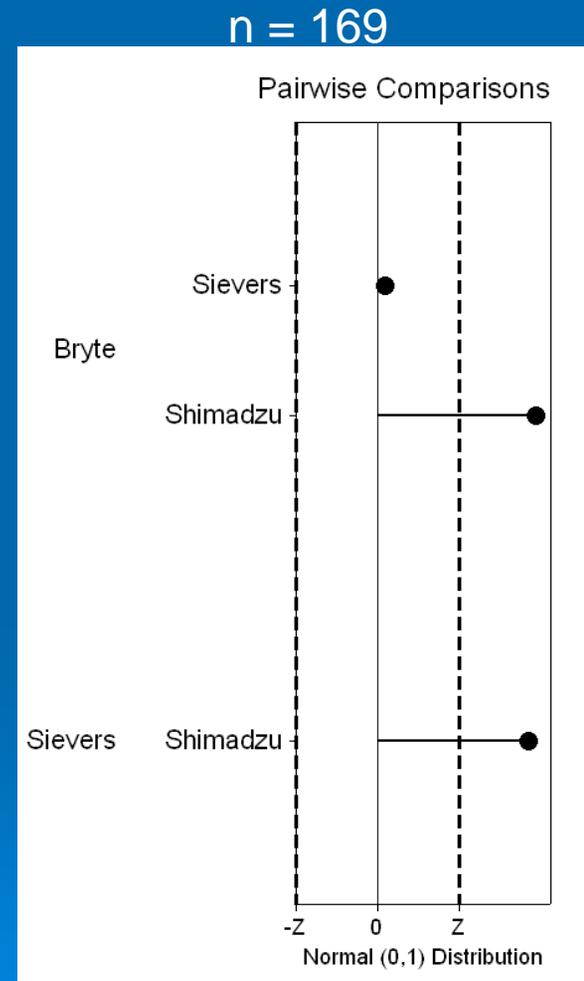
# Paired t-test

	<u>n</u>	<u>Mean</u>	<u>SE Mean</u>
Shima	1665	2.45916	0.02559
Sievers	1665	2.27784	0.02437
Diff	1665	0.181321	0.01277

95% CI: (0.156260, 0.206383)

t-test of mean diff,  $p = 0.000$

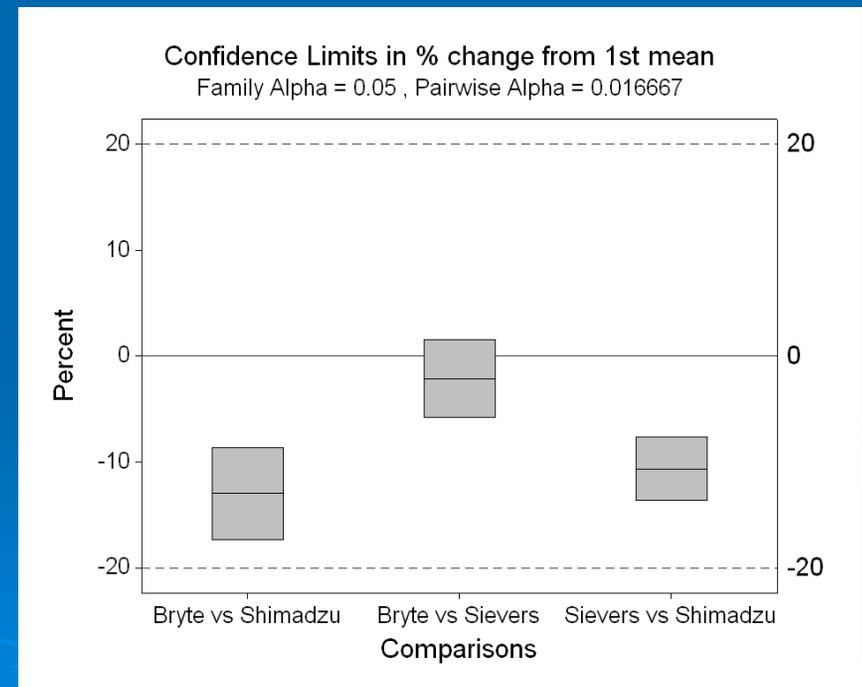
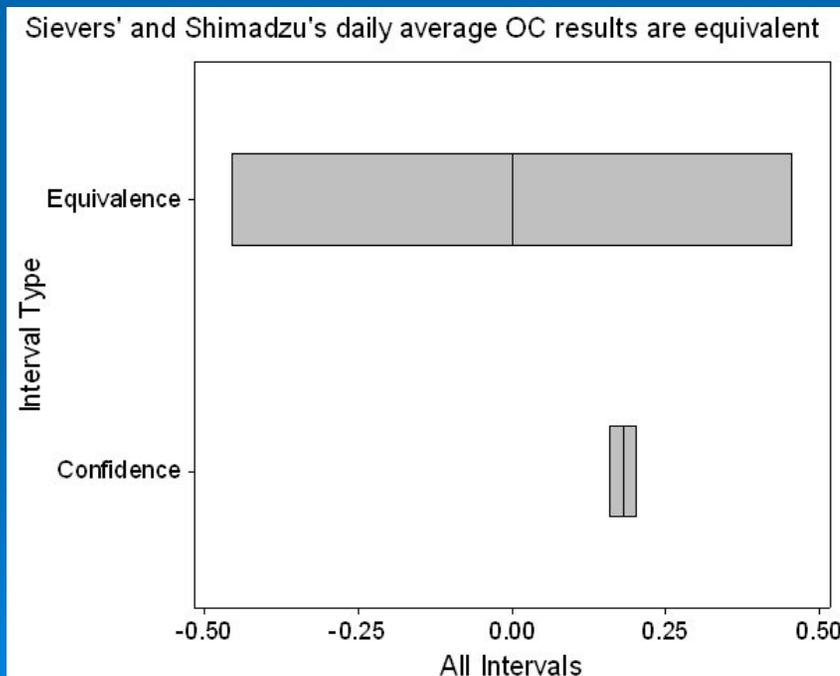
# KW-ANOVA



# Equivalence Tests Results

Sievers & Shimadzu  
are Equivalent

Sievers, Shimadzu &  
Bryte data are Equivalent



# Are equivalence statistical tests fringe science?

---

- They have been used in pharmaceuticals by FDA (The *Yellow Book*) since 1984; the EU has since followed.
- A generic drug must be bioequivalent to the original drug
  - **Note:** The test is not that the effectiveness of the generic drug is equal to that of the original drug -- i.e., there is no NHST here!
  - Rather the test is whether the effectiveness of the generic drug is within 20% of the approved drug.
  - There are no  $p$  values to report
  - In other words: ARE THE 2 DRUGS EQUIVALENT?

# Closing thoughts

- Equivalence tests are good alternatives to NHST for the analyses of the large datasets generated by DWR's online instruments
- Equivalence intervals are not easy to set.
- Equivalence Interval must be set a priori.
- Futile to try to use classical tests and equivalence tests on the same dataset i.e. to compare each stat method to the other.