Statistics for and with Physical Scientists

R Pruim

eCOTS 2014

Context

I teach at a 4-year liberal arts college in which

- roughly half of the students graduate with a professional degree (engineering, nursing, education, speach pathology, etc.)
- the sciences are strongly represented ($\sim 25\%$ and growing)
- statistics required for biology and engineering (dedicated courses for each), and for computer science, but not for physics and chemistry

Some History

0. Biology asks for feedback on math/stat in lab manual
   - not a physical science, but got the ball rolling
   - also teach *Intro Biostats*
1. Physics asks for feedback on their data analysis manual
   - has led to periodic conversation bursts
2. Creation of new Stats course for Engineers
   - 2-hour required course, spring sophomore year
   - taught for 2 years by colleague, 2 years by me
3. Creation of home-brewed materials to teach engineers
   - used this spring for second time
4. Interaction with 2 engineers about topics, data, etc.

The odd place of statistics in the sciences

- the more a science uses data and the more mathematics their students take, the less likely they are to take a college statistics course
- regardless of the amount of statistics a program wants its students to know, the maximal dose is 1 course
    - less than a course is preferable
- everyone (a) complains about their lack of statistical background, and (b) thinks they can teach statistics
- statistical idiosynchrasies (a.k.a. culture) grow up within disciplines (including Stat Ed)

Intro Stats is part of the problem, not part of the solution

An ~~obvious~~ wrong solution: have the scientists take Intro Stats

- for all the attention statisticians have placed on Intro Stats, most physical scientists would be sorely disappointed to have their students spend an entire semester to learn what is in those courses.

We need to talk

We need to talk – and listen

The customer is not always right

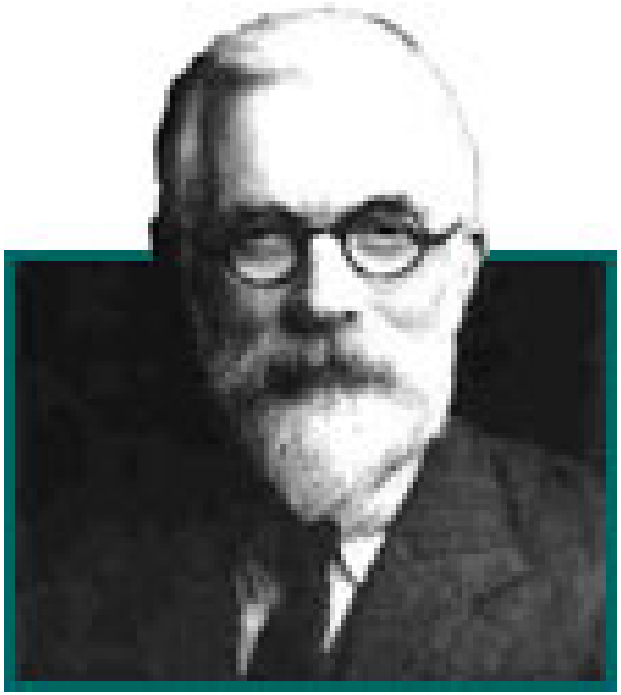
- scientific communities may perpetuate statistical errors or misconceptions
- scientists are often unaware of statistical options
  - *a* way becomes *the* way
- (most) scientists have no real experience teaching statistics
  - don't know what concepts are challenging
  - don't always appreciate prerequisite understanding
  - don't know current trends in Stat Ed
  - may not appreciate difference between doing and understanding

The customer is not always wrong

I believe sanity and realism can be restored to the teaching of Mathematical Statistics most easily and directly by entrusting such teaching largely to men and women who have had personal experience of research in the Natural Sciences.

The customer is not always wrong

I believe sanity and realism can be restored to the teaching of Mathematical Statistics most easily and directly by entrusting such teaching largely to men and women who have had personal experience of research in the Natural Sciences.



R. A. Fisher

Fisher isn't the only one

JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

A Statistics Curriculum for the Undergraduate Chemistry Major

Nicholas E. Schlotter*

Department of Chemistry, Hamline University, St. Paul, Minnesota 55104, United States

Supporting Information

ABSTRACT: Our ability to statistically analyze data has grown significantly with the maturing of computer hardware and software. However, the evolution of our statistics capabilities has taken place without a corresponding evolution in the curriculum for the undergraduate chemistry major. Most faculty understands the need for a statistical educational component, but there is little consensus as to the exact nature of what is to be taught and who should teach it. Because of the large number of courses required for the undergraduate chemistry major, it seems unlikely that requiring a course on statistics will be practical at most institutions. Additionally, it is unlikely that the typical high school education will address the needed statistics or the software training to prepare students for the chemistry courses. Therefore, the chemistry faculty must teach the statistics needed by the majors. The faculty needs to focus on statistics useful to the chemist and this is distinctly different than what is often encountered in biology, medicine, psychology, and business. A starting point is suggested for a discussion on a statistics curriculum that addresses the needs of the chemistry majors.

KEYWORDS: *First-Year Undergraduate, Second-Year Undergraduate, Upper-Division Undergraduate, Curriculum, Interdisciplinary/Multidisciplinary, Computer-Based Learning, Mathematics/Symbolic Mathematics, Statistical Mechanics*

Fisher isn't the only one

A Statistics Curriculum for the Undergraduate Chemistry Major

Schlotter, J. Chem. Educ., 2013, 90 (1), pp 51–55

- the **chemistry community needs to** have a discussion about statistics and to **decide what statistics should be taught**
- statistics curriculum needs to be **embedded in chemistry courses**
- to require a course in probability and statistics reaches the breaking point for required courses
- even when students take a statistics course in the mathematics department, it **does not address many of the needs** of chemistry majors
- focusing on how to **test data relative to a model**
- using **Excel** to do calculations and plot results

The customer is not always wrong

What the scientist knows

- science – context matters

The customer is not always wrong

What the scientist knows

- science – context matters
- how statistics is (or could be) used in science courses/labs/research

The customer is not always wrong

What the scientist knows

- science – context matters
- how statistics is (or could be) used in science courses/labs/research
- a (subset of) important statistical applications

The customer is not always wrong

What the scientist knows

- science – context matters
- how statistics is (or could be) used in science courses/labs/research
- a (subset of) important statistical applications
- what things we do that are of no use to them

So what do the engineers want?

I asked, and they said this:

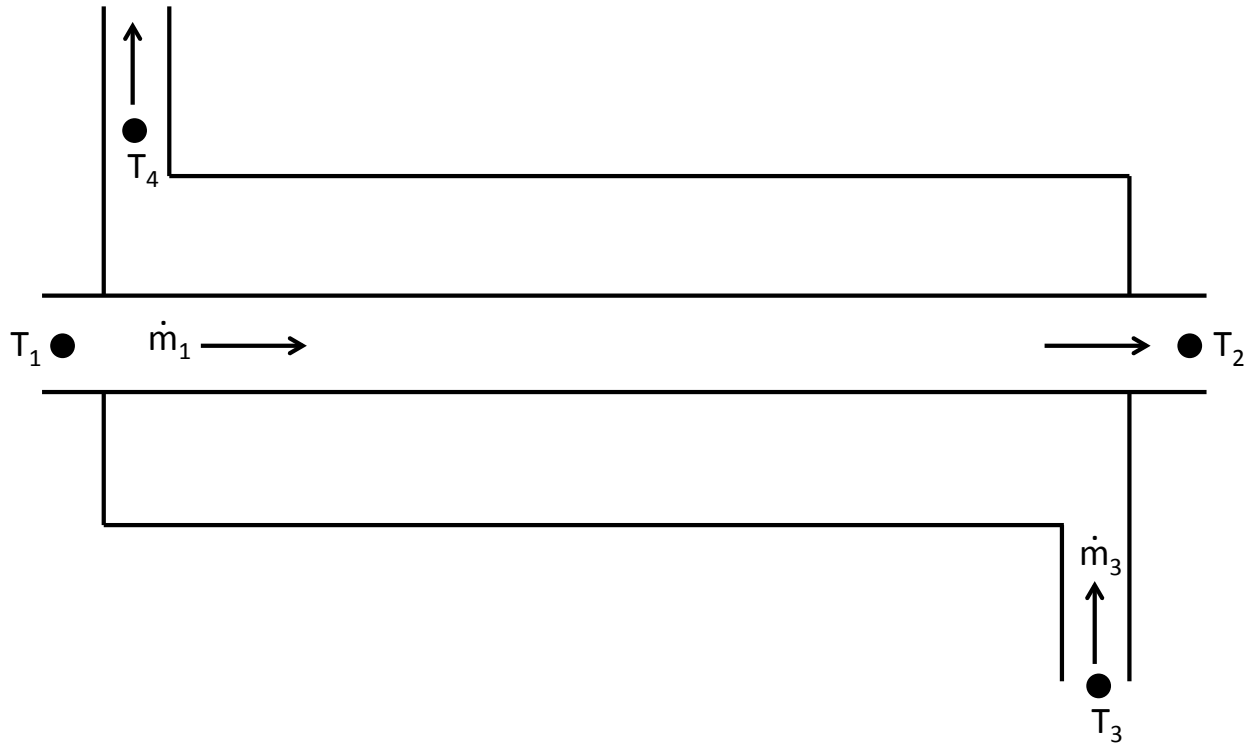
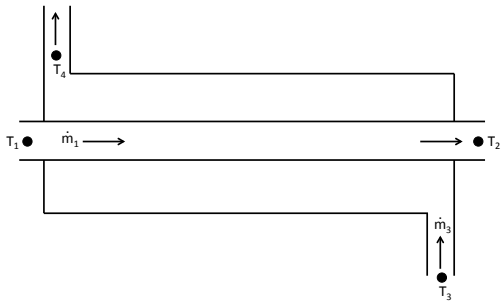


Figure : Heat exchanger with statepoints.

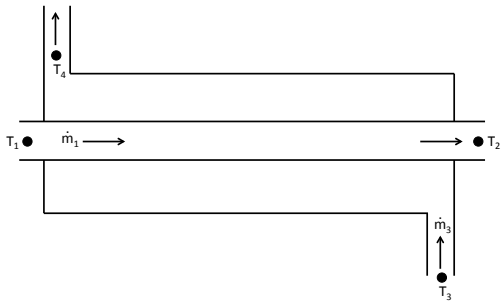
So what do the engineers want?

$$
\begin{aligned}
U &= \frac{\dot{Q}}{A\Delta T_{lm}} \\
\dot{Q} &= \dot{m}_1 C_p (T_2 - T_1) \\
\Delta T_{lm} &= \frac{(T_1 - T_4) - (T_2 - T_3)}{\log\left(\frac{T_1 - T_4}{T_2 - T_3}\right)}
\end{aligned}
$$



So what do the engineers want?

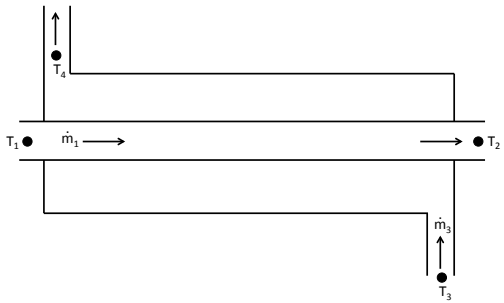
$$
\begin{aligned}
U &= \frac{\dot{Q}}{A\Delta T_{lm}} \\
\dot{Q} &= \dot{m}_1 C_p (T_2 - T_1) \\
\Delta T_{lm} &= \frac{(T_1 - T_4) - (T_2 - T_3)}{\log\left(\frac{T_1 - T_4}{T_2 - T_3}\right)}
\end{aligned}
$$



Estimate U with uncertainty

So what do the engineers want?

$$
\begin{aligned}
U &= \frac{\dot{Q}}{A\Delta T_{lm}} \\
\dot{Q} &= \dot{m}_1 C_p (T_2 - T_1) \\
\Delta T_{lm} &= \frac{(T_1 - T_4) - (T_2 - T_3)}{\log\left(\frac{T_1 - T_4}{T_2 - T_3}\right)}
\end{aligned}
$$



Estimate U with uncertainty

Oh, and $n = 1$.

A simpler problem

Given a bag of dimes, how can we estmate the number of dimes without counting them all? How accurate is such an estimate?

- What would an Intro Stats student do with this?

A simpler problem

Given a bag of dimes, how can we estmate the number of dimes without counting them all? How accurate is such an estimate?

- What would an Intro Stats student do with this?
  - collect a sample of bags of dimes
  - weigh each bag, and count dimes in each bag
  - fit a linear regression

Where do numbers come from?

Given a bag of dimes, how can we estmate the number of dimes without counting them all? How accurate is such an estimate?

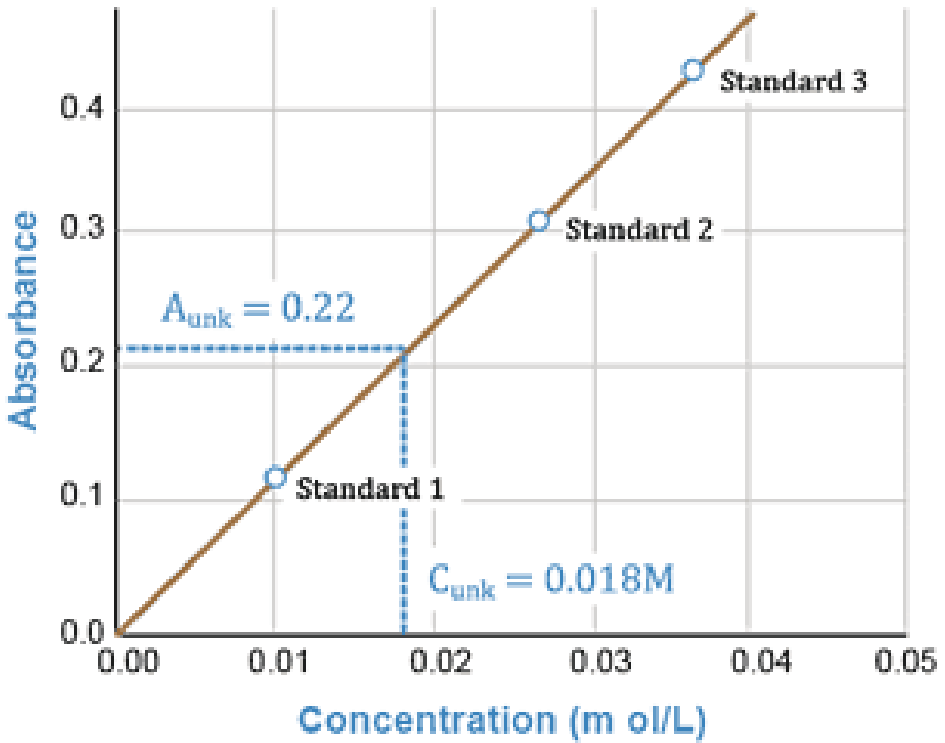
$$N = \frac{T}{\overline{d}}$$

- uncertainty in $\overline{d}$: $\frac{s}{\sqrt{n}}$
- uncertainty in T depends on measuring device and protocol
- uncertainty in N depends on
  - two uncertainties above
  - division

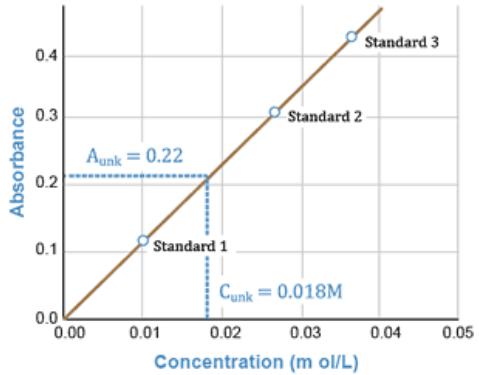
Another Example: Spectrophotometry

Spectrophotometry

1. prepare several dilutions of a standard solution (`conc`)
2. measure absorbance of each with a spectrophotometer (`abs`)
3. should be a linear relation between the two

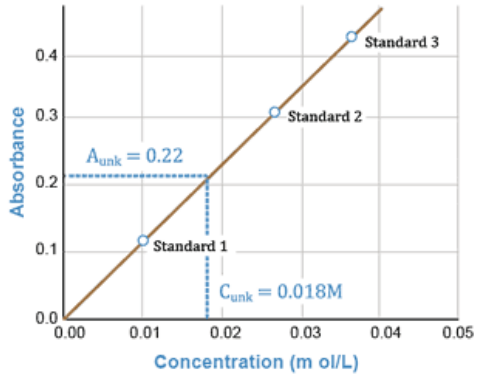


Spectrophotometry



for a substance with unknown `conc`, measure `abs` and use data to estimate `conc`

Spectrophotometry



for a substance with unknown `conc`, measure `abs` and use data to estimate `conc`

- Which variable is the "response variable" here?
- Does "independent variable" = "expanatory variable"?
- How do we estimate concentration with uncertainty?

So what do we (statisticians) do?

1. Find conversation partners
   - be prepared for a long, slow conversation
   - make sure you are really communicating (vocab!)

So what do we (statisticians) do?

1. Find conversation partners
   - be prepared for a long, slow conversation
   - make sure you are really communicating (vocab!)
2. Attack the labs
   - that's where data analysis happens
   - often home brewed, so change can happen more quickly

So what do we (statisticians) do?

1. Find conversation partners
   - be prepared for a long, slow conversation
   - make sure you are really communicating (vocab!)
2. Attack the labs
   - that's where data analysis happens
   - often home brewed, so change can happen more quickly
3. Help in downstream courses
   - We haven't acheived much until what happens in statistics makes a difference in the science courses

So what do we (statisticians) do?

1. Find conversation partners
   - be prepared for a long, slow conversation
   - make sure you are really communicating (vocab!)
2. Attack the labs
   - that's where data analysis happens
   - often home brewed, so change can happen more quickly
3. Help in downstream courses
   - We haven't acheived much until what happens in statistics makes a difference in the science courses
4. Focus on estimation with uncertainty
   - Let scientists help pick applications

But Wait, There's More

11 ET today:

Cassandra Wolos Pattanayak from Wellesley College

Reaching Students Outside the Statistics Classroom:

Integrating Statistics into the Coursework and Research Experiences of Non-Statistics Majors