Teacher Notes about "A Year's Worth of Inquiry on Day One"

William Thill, Harvard Westlake: wthill@hw.com

Rationale: Why a Day One Survey?

- Developing students' Statistical literacy, statistical reasoning, and statistical thinking must be done in context of real data.
- Students ask better questions, make better judgments, and do better work when they collect the data themselves.
- I want students ***to be curious, ask questions, make conjectures, and utilize their experiences*** **from the very first day.**
- I want students to investigate contexts authentically. Textbook problems prevent students from getting answers to their questions about context/ design, etc…
- I want students to practice picking the right statistical techniques for their research question.
- I want to respond to student questions **on the fly** with useful examples **immediately, effectively, and easily.**

Logistics: Brief explanation -

- On the first day of AP Statistics, our teachers (and the ones at Loomis Chaffee School) hand out the survey (hard copy), provide materials for measuring /collecting data, and facilitate student data collection (cat herding).
- As students work, teachers interview students: *"What are you wondering about? Do any research questions come to mind?"* Teachers record those questions for the class in a public record.
- Students are given a URL for the survey, username + password and enter their response to questions.
- After all(?) students respond to the survey, I look at the raw data in Fathom, do some "sanitizing" of responses, and provide the data file to students on our course website.
- Students and teachers have access to the data base every day, all year.

Uses for the Day One survey database:

1.Questions & AP Statistics topics that emerge:

- **Basic measurement / numeracy skills:**
  - Using measurement devices correctly (tape measures, rulers, scales – you'd be surprised!)
  - Using appropriate units, converting them correctly (adding 100 to measurements for heights, mm←→ cm), a scale/ ruler, knowing units, diameter, millimeters, dollars, etc.
  - Choosing appropriate ways to measure (pulse rates)
- **Describing/summarizing categorical and quantitative data:**
  - Analyzing quantitative data distributions:
    - Unimodal /symmetric (diameter of a tennis ball with a ruler, sometimes height/armspans
    - Skewed distributions (# Facebook Friends, Haircut costs)
    - Multi-modal distributions (how old is your teacher?)
    - Outliers (haircut costs, pulse rates, # Facebook friends)
  - Distributions of Categorical Data:
    - Conjecturing associations between gender / TV show, school/SAT attitudes, student / parent political affiliation

        - Comparative bar charts / ribbon charts (Do boys' TV habits doffer form girls'?)
    - Associations (or apparent independence) between quantitative variables
        - Height/arm span – strongly linear, using the line "height=armspan" vs. the least squares linear model to gain information
        - Guessed / actual BP weight: heteroschedasticity
        - Actual BP weight vs. "guess – actual" heteroschedasticity , transforming variables to achieve homoschedasticity
- **Design of studies / experiments**
    - The AP Stats kids: are they representative of a population of interest? Why (not)? What leaps of faith are you making?
    - Guessed & actual BP weights, sit & stand pulse rates: Paired data or independent groups?
    - Discussion of randomizing orders of treatments: Sit/ Stand variables.
    - Association / causation: knowing the difference: Is this a controlled experiment?
- **Understanding randomness / probability**
    - empirical results in short term vs. long term expectations. "Toss three coins":
    - Tree Diagrams → discrete RV: "toss three coins,"
    - Roll the die 10 times, record # threes: binomial models
    - Sampling distributions+ CLT: Large data set can be viewed as a population of students, sampling from this population w/ Fathom, create a measurement, and re-sample
- **Inference: making conclusions about populations from samples of data**
    - Discussing whether conditions for inference are satisfied – any question about populations will involve this.
    - Inference for proportions 1 & 2 sample:
        - What proportion of seniors at independent schools favor abolishing the SAT?
        - Which students are more likely to favor abolishing the SAT? LC students or HW students?
    - Inference for means (tests + CI) :
        - Do HW students, on average, tend to overload their backpacks (>20 lbs, according to one research study)?
        - Do standing PR's tend to be higher than sitting PR"s? if so by about how much?
        - Does Mr. Thill appear younger than his real age?
    - Inference for associations
        - Are "#facebook friends" and "amount spent on lunch associatied?" (Short answer: for the kids surveyed this part year, **(yes)**
        - For HS seniors, Is handedness associated with being a varsity athlete?
        - Benfords Law: Does # FB follow the predicted distribution from Benford's Law)?

2. How the database gets used:

In class:

Examples: Demonstrations / Discussions, responses to student questions

Quick 5-10 minute investigations

Assignments / Assessement:

"Turn-in" assignments based on answering research questions from data base, to demonstrate competency in AP Stats standards

Test questions for in-class tests

Case Studies for critical thinking / writing assessment

Benefits I observe:

- Student raise real questions of interest that can be explored and investigated on the first day of class.
- New / Different examples beyond the textbook are more easily accessible, dynamic.
- **Students can investigate the context - not possible with examples from other resources.**
- It's an equalizer - students without research question experience get their first chance to do this with this data base (catching up to science aficionados)
- Database nurtures curiosity and inquiry in my students
- Database nurtures curiosity and inquiry with fellow teachers of AP Stats
- Examples for tests / etc. involve context students understand, because they collected the data.

Things to keep in mind:

- Provide data to students in multiple formats : Fathom, tab-delimited text files, Excel,
- Triple check your data base for errors that will block students' access to the learning goals of your activity. Too many errors, outliers, failure to meet conditions will confuse, not support your students.
- Establish clear expectations for student success and make sure your students understand those goals. This is tougher to do with open ended tasks.
- Smaller, focused tasks tend to work better than larger open ended tasks for me.
- Use rubrics and examples of student work as benchmarks for your students.
- If using data base for assessment, work through what you want kids to achieve. Be sure the intended outcome comes to pass.
- Give students class time to access data, think of good questions, and refine the precision of their thinking / communication.
- Monitor student work as you go. This reduces student confusions and cheating.