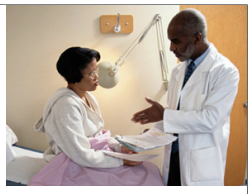


Adult Statistical Literacy: Describing and Assessing

Jennifer J. Kaplan

ICOTS-8

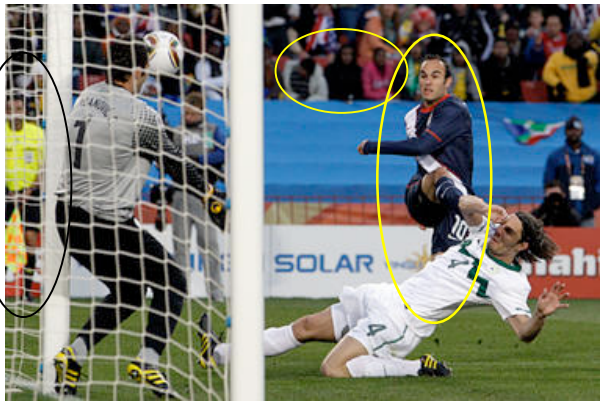
12 July 2010

There are two parts to this talk.

The first part, describing, is theoretical, but based in the literature

The second part, assessing, is experimental, using data to extend an existing framework fro school children to adults.

The pictures represent people involved in statistical literacy activities: the baby is researchir and buying stocks, the doctor and patient are going over medical results and making care decisions, the journalist is writing a report about a student and the people reading the newspaper are reading the report and incorporating the information into their world view.



Theoretical Foundation - Inspired by Roxy Peck

From Roxy's USCOTS 2005 plenary talk:

Spectators: some are interested and appreciative; some are bored and disinterested. In statistics, goal is someone who is interested and excited and enjoys "watching." Courses sh train to appreciate beauty and usefulness of statistics.

Referees: may not be expert players, but they know the rules and recognize bad behavior a call a penalty when they see it. In statistics, they need to know the rules and be able to critic evaluate the work of others, but don't necessarily need the skills of a player.

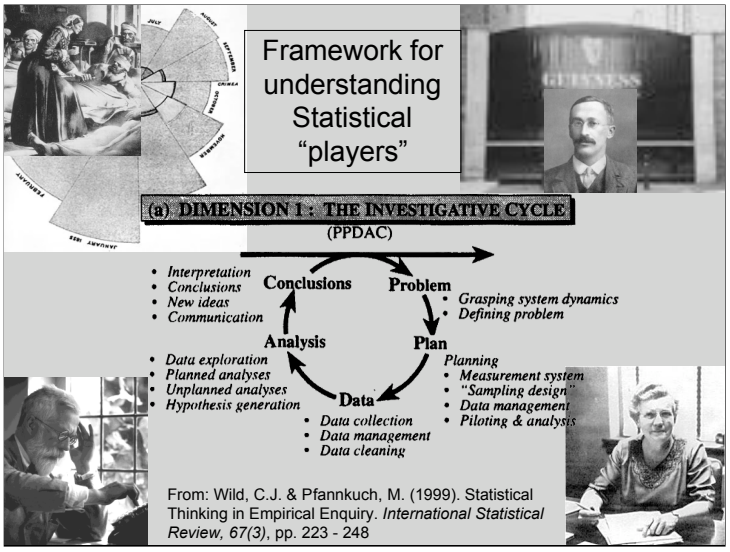
Players: have varying skill levels from recreational to professional, but have to know the rule and handle the ball. In statistics need to know the rules to recognize good and bad behavior need to be able to produce good behavior, and will have varying skill levels.

Players: (on slide) people involved in quantitative research. Statisticians and researchers in other fiel who use quantitative methods or mixed methods.

Referees: (on slide) people who work with the quantitative researchers, co-authors -- geneticists, biologists, climate scientists, education researchers, geographers, medical professionals running clini trials

Consumers: (**not on slide)** doctors, nurses, patients, journalists, investors, qualitative researchers in f that have quantitative researchers, regular people making decisions every day.

Make the point that there are **levels** of all of these things including Fans -- there are the face painters the ones yawning at two ends of the spectrum.



Cycle of Statistical Enquiry

We have a robust notion of what statistical players do as they enact statistical enquiry. The of this research program is to have as robust a notion of what expert statistical consumershi entails.

Existing Definitions of Statistical Literacy

Include all the definitions listed on the IASE - ISLP website - http://www.stat.auckland.ac.nz/~iase/islp/def

Key Words to notice:

Interpret, evaluate, discuss, communicate, appreciate, in context

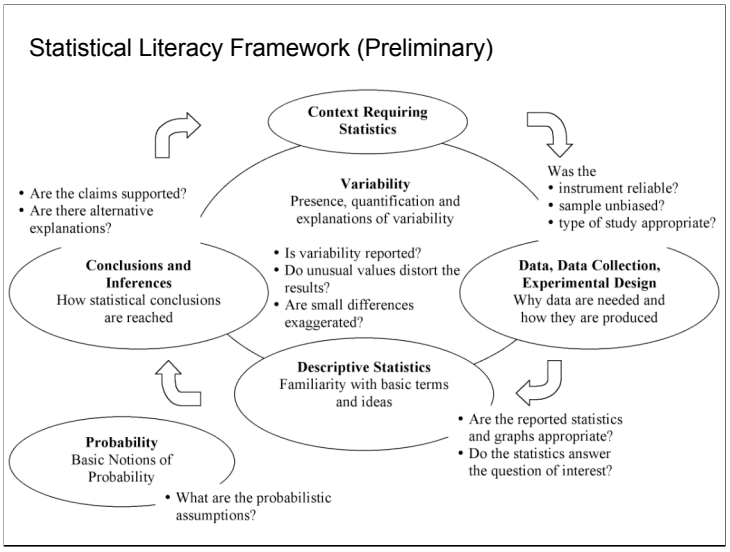
Elements of Statistical Literacy

The five authors provide lists of elements that make up statistical literacy.

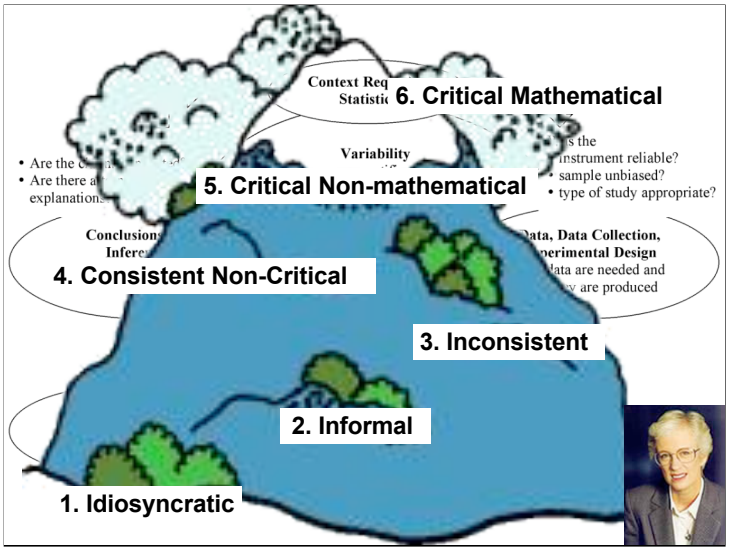
The elements were reorganized according to a scheme that more closely follows the traditional topics taught in school level statistics and probability because:

The overarching goal of the research program suggested by this pilot work is to provide a description expert adult statistical literacy to use as a target for instruction which, following the suggestion of Wallman (1993) utilizes the expertise of statisticians as a basis and that of Gal (2002) can be implem via large educational systems and academic institutions.

In addition, Iddo Gal's "worry questions" were used to supplement the competencies given by the otl authors.



The proceedings paper contains this information in tabular form and is more comprehensive



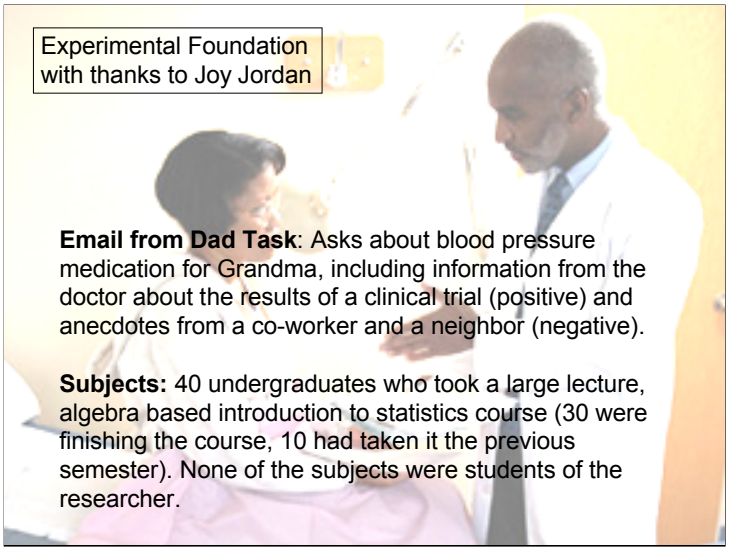
Watson and Callingham (2003) Framework

Originally I thought about using a ladder as a graphic here, but then I realized that with the cycle on t previous slide, a circular stair case might be more appropriate. I decided that a circular stair case too rigid a model and that learning is probably more like climbing a mountain, with multiple path to the top.

The frameworl was developed from large quantitative study: 3852 students in Tasmania. Grades 3, 5 with 27% from 3 and 9, 23% grade 6, 11% grade 5 and the rest roughly split between 7 and 8. Da were collected in 1993, 1995, 1997, 2000 on 80 items.

The 4 items with the highest number of responses were used to anchor the difficult measurements of the items. Difficulty levels and fit statistics of all steps on all items were calculated (using Quest Software)

Task Step = level of response denoted by coding



The Study

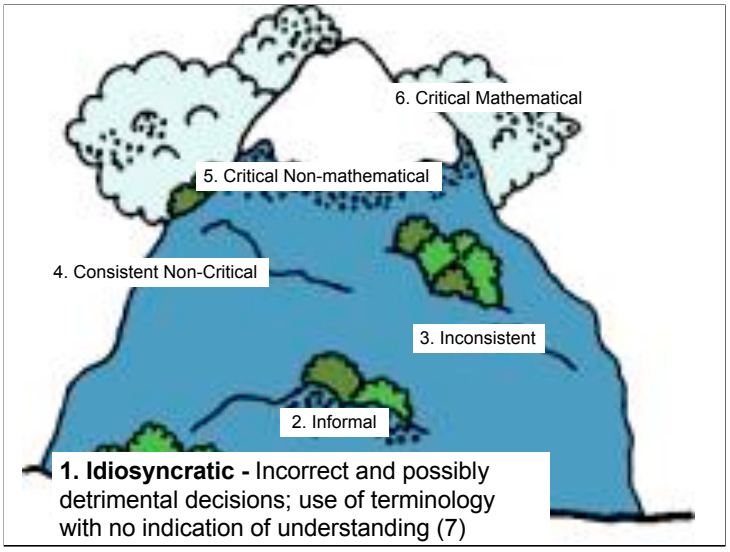
Thirty-two of the subjects (75%) were women.

Twenty-four (60%) reported having a G.P.A. of 3.5 or higher.

Thirty-three subjects (82.5%) reported earning an A or B in statistics.

The four concentrations with the highest number of participants were pre-pharmacy, 10 (25%), nursi
(20%), advertising, 6 (15%) and psychology, 4 (10%).

The number of words used in the responses to the Email from Dad task was roughly normally distrib
with a mean of 116 words and standard deviation 56 words.



Idiosyncratic - Incorrect and possibly detrimental decisions; use of terminology with no indication o understanding

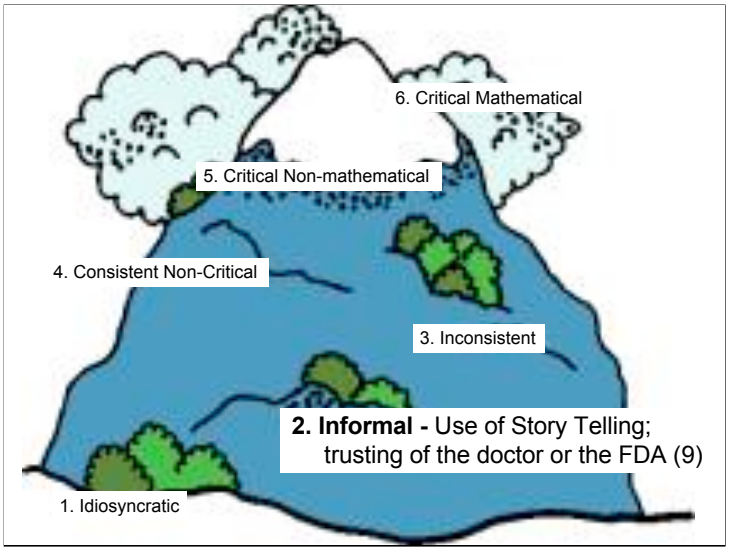
Examples:

The p-value indicates that there is only a 0.1% chance that the medication works. Keep Grandma on t medication.

The p-value indicates that there is only a 0.1% chance that the medication does NOT work. Take Gra off the medication.

The p-value means that the experimenters found a very rare situation so the medication is unlikely to work. Sally and Larry are proof of this. Take Grandma off the medication.

Take Grandma off the medication. A double blind experiment means that both the experimenters and subjects did not know what was given or taken. So the experiments did not know whether they gave t subjects the Makemewell or placebo, and for the subjects did not know if they were taken the placeb Makemewell. Even worst it could have been something else

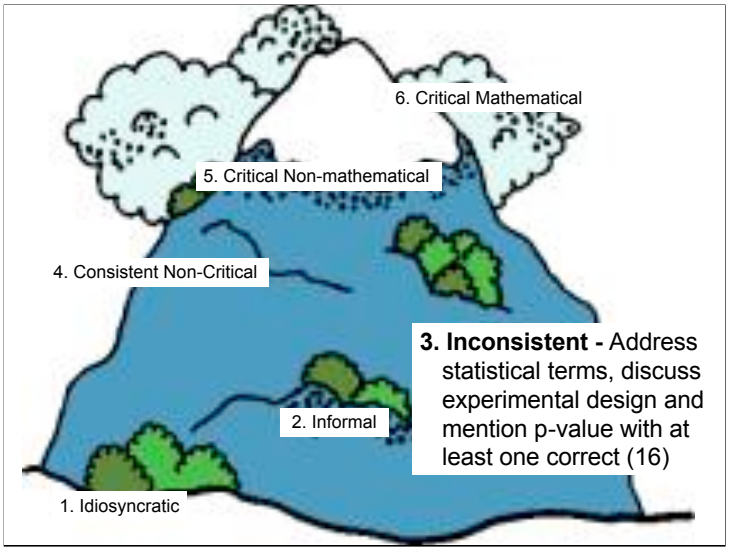


Informal - Use of Story Telling; trusting of the doctor or the FDA

No quantitative mention of the p-value; trust in the doctor/experimenters; directions to go back and a doctor follow up questions.

The results indicate that there is a high chance that the drug lowers blood pressure, but go back and a doctor about side effects and whether the drug has been tested against other drugs.

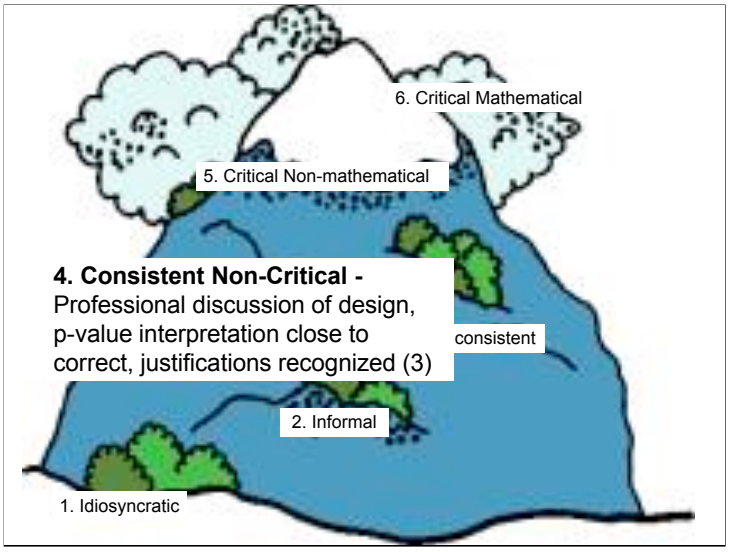
The experiment sounds good and both the doctor and the FDA approved the drug and I trust them. As Grandma how she is feeling. As long as she feels fine, keep her on the medication.



Inconsistent - Address statistical terms, discuss experimental design and mention p-value with at lea correct

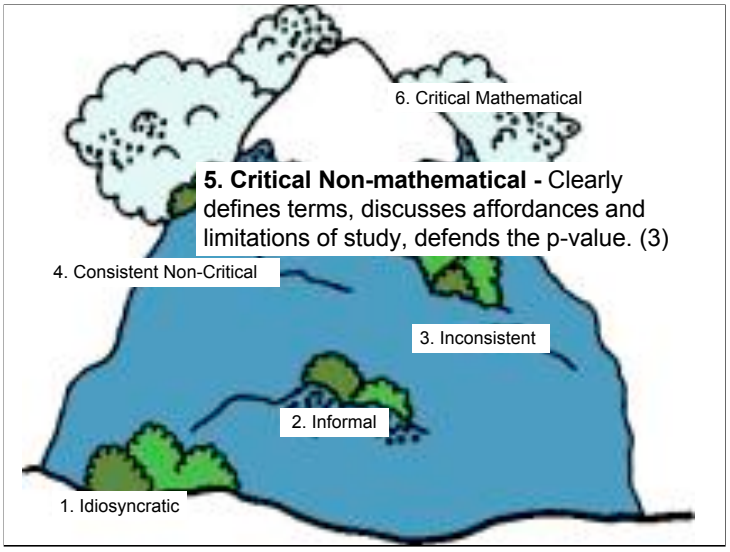
I trust the double-blind experiment's results more than the anecdotes from the neighbors. The neighb could be exceptions or outliers and their results might be very rare. In a double-blind experiment neit the doctor nor the patients knew which treatment or pill they were getting until afterwards so bias fro doctors and patients on how they think the reactions to the drugs would be are avoided. I think Grand should stay on the medication because it is significantly better at lowering blood pressure than nothin all and the chance of findings being based solely on chance is very small.

(HIGH 3) The test being conducted has a small sample size, which could make this study less accura There is a 0.1% chance that the doctor got these results by chance which makes the study seem plaus because it is low. The study does not account for side effects and the doctor did not mention if any of patients had an increase in blood pressure. If Grandma's blood pressure goes up, she should probably taking the medication. Some people always have different side effects when taking medications beca people's bodies do not act the same as others.



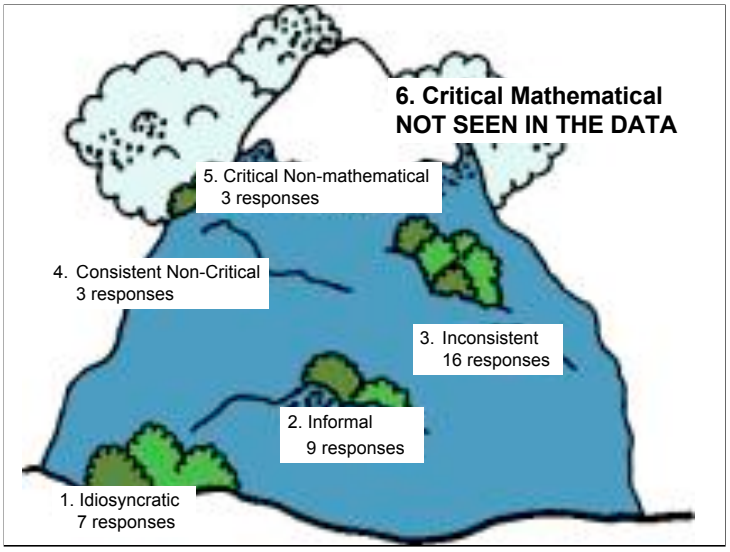
Consistent non-critical - Professional discussion of design, p-value interpretation close to correct, justifications recognized

The study conducted was a randomized, double-blind experiment. This means that the subjects were chosen randomly and were imposed with treatments. Double-blind means that neither the conductor experiment or the subject knew what the treatment was. This method allows for smaller error in that subjects will not give false results. For example, if a subject does not know they are on placebo, it is likely that they will falsely give lower response. The amount of subjects in this experiment may be to small. It also may have been a goo idea to also test the Makemewell drug against other treatments be placebo. The significance level of $p < 0.0001$ is extremely small. The usual level is 0.05. This level te you when the data is significant statistically and allows you to analyze whether or not the drug provic good result. A very small significance level fives very strong evidence that the Makemewell drug wo better than placebo. Of course, with experiments there are chances that it is not effective, which may the case for Sally. Also, this experiment does not test for side effects. I do not think that Grandma sh be taken off of this medication because the evidence for this drug is very strong due to the small significance level.

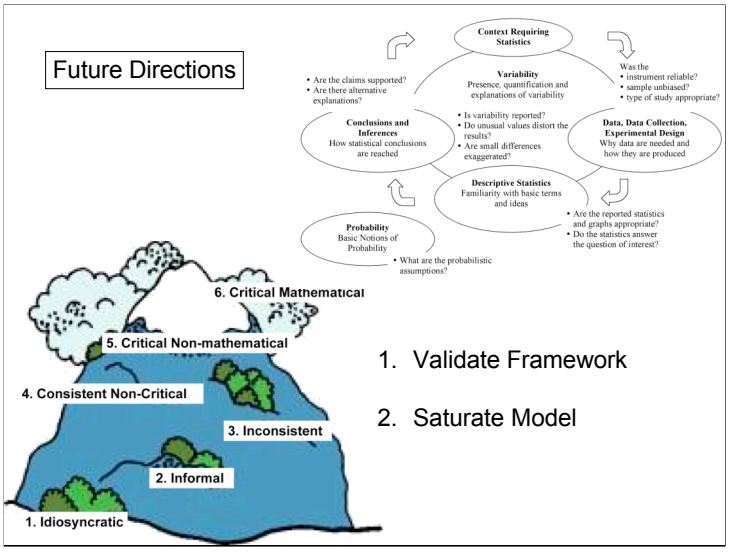


Critical, non-mathematical - Clearly defines terms, discusses affordances and limitations of study, defends the p-value.

I'm terribly sorry to hear about grandma. Basically, the experiment that was done for the drug had re good test results based on what you've told me. You said it was double-blind, which simply means n the doctors performing the experiments or the patients knew who was receiving an actual medication (Makemewell) and who was getting a placebo (essentially a sugar-pill that would have no real effects except psychological in some cases). The doctor would've compared the results of those who were a were not on the medication at the end. The way it works is they look to see what the chances are of g the result of lower blood pressure (as the experiment did show) if in actuality the medication didn't w This is the weird p-value (< .001) you sent me. If the Makemewell drug didn't lower blood pressure result that the experimental group would happen .1% of the time (very rare indeed!). As such I would that grandma should probably stay on the medication. The two cases you told me about could be the of so many other factors (like paranoid old people). They may not have been properly screened for p existing conditions that may effect them while taking the drug. And, of course, all drugs do have sid effects. I would say if grandma has been on it a while then it's safe to say that she can stay on it. Tha for the cash. You rule.



From the writing samples, we were not convinced that any of the subjects could transfer a correct numerical interpretation of the p-value to other contexts. This may be a limitation of t study/data rather than the subjects.



Future Directions

Doing two separate data collection:

1. Interviewing Ph.D. statisticians who teach an algebra based introduction to statisti course about their ideas about statistical literacy and to comment on the preliminar framework and to ask about the email from dad task and to get ideas for other task
2. Designing more tasks and then implementing them in a cross sectional study using undergraduate statistics majors and minors, and masters' and Ph.D. students in statistics. A second task under development is based on some of the graphs that we published in the IPCC report for policy makers on global climate change. The subj will first be asked to make conclusions based on the graphs and then to respond to conclusions made in the report based on the information in the graphs.

Questions?

Thank You!