


# Math is Music – Stats is Literature

*Or why are there no six year old novelists?*

Dick De Veaux - Williams College  
With thanks to Paul Velleman,  
Cornell University


August 4, 2005      Beyond the Formula      1



# Prodigies

- **Math, music, chess**
  - Gauss, Pascal
  - Mozart, Schubert, Mendelssohn
  - Bobby Fischer
- **Why these three areas?**
- **Each creates its own world with its own set of rules**
  - There is no “experience” required
  - Once you know the rules, you are free to create anything


August 4, 2005      Beyond the Formula      2



# Prodigies in Literature?

- **Mary Wollstonecraft Shelley**
  - Age 19
  - Created Frankenstein, an imaginary creature
- **Others?**
- **Why?**
  - Literature is about the world, not about rules. It deals with life’s experience and the wisdom we develop over time.


August 4, 2005      Beyond the Formula      3



# Statistics – What do students find hard?

- “Understood the material in class, but found it hard to do the homework”
- “Should be more like a math course, with everything laid out beforehand”
- “More problems in class should be like the HW and tests”


August 4, 2005      Beyond the Formula      4



# What is “easy”?

- **The math part – well, not “easy”, but...**
  - Math is axiomatic – logical – laid out beforehand
  - Give one example, we can change the numbers and it still makes sense
- **The hard part**
  - Putting it all together
    - Real world
    - Experience
    - Methods

August 4, 2005      Beyond the Formula      5



# Teaching Calculus

- **Of course, it’s not easy**
- **But (1<sup>st</sup> semester) Calculus has fewer concepts to get across**
  - Functions and Graphs (review)
  - Limits (review?)
  - Continuity (some review)
  - Derivative – max and min
  - Implicit differentiation
  - Antiderivative – area
  - Fundamental Theorem

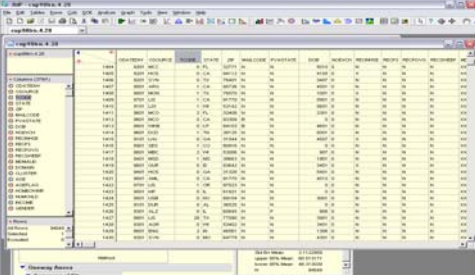
August 4, 2005      Beyond the Formula      6

## What about Statistics?

- **Scientific Method**
  - How to draw conclusions about the world given data?
  - Population and Samples
- **Center and Spread**
  - Variation is everywhere
- **Randomness (is our friend)**
- **Models**
  - Models for data
  - Models for relationships
  - Models for sampling
- **Probability**
  - Counting
  - Elementary manipulation
- **Inference**
  - Sampling Distribution
  - Confidence Interval
  - Hypothesis Testing

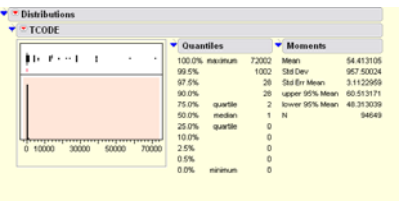
August 4, 2005 Beyond the Formula 7

## What's "Hard"? -- Example



August 4, 2005 Beyond the Formula 8

## T-Code



August 4, 2005 Beyond the Formula 9

## What does it mean?

T-Code	Title				
0					
1	MR.	17	JUDGE	50	BLDR
1001	MR. & MRS.	17002	JUDGE & MRS.	56	MAJOR
1002	MR. & MRS.	18	MAJOR	59002	LIEUTENANT & MRS.
2	MRS.	18002	MAJOR & MRS.	62	LORD
2002	MESDAIRES	19	SENATOR	63	CARDINAL
3	MRS.	20	GOVERNOR	64	FRIEND
3003	MESSES	21002	SERGEANT & MRS.	65	FRIENDS
4	DR.	22002	COLONEL & MRS.	68	ARCHDEACON
4002	DR. & MRS.	24	LIEUTENANT	69	CANCY
4004	DOCTORS	26	MONSIGNOR	70	BISHOP
5	MADAME	27	FRIEND	72002	REVEREND & MRS.
6	SERGEANT	28	MR.	73	PASTOR
9	RABBI	28028	MRS.	75	ARCHBISHOP
10	PROFESSOR	29	BISHOP	80	SPECIALIST
10002	PROFESSOR & MRS.	31	AMBASSADOR	87	PRIVATE
10010	PROFESSORS	31002	AMBASSADOR & MRS	89	SEAMAN
11	ADMIRAL	33	CANTOR	90	ARMAN
11002	ADMIRAL & MRS.	36	BROTHER	91	JUSTICE
12	GENERAL	37	BAR	92	MR. JUSTICE
12002	GENERAL & MRS.	38	COMMODORE	100	M.
13	COLONEL	40	FATHER	103	WALLS
13002	COLONEL & MRS.	42	SISTER	104	CHANCELLOR
14	CAPTAIN	43	PRESIDENT	106	REPRESENTATIVE
14002	CAPTAIN & MRS.	44	MASTER	107	SECRETARY
15	COMMANDER	46	MOTHER	108	LT. GOVERNOR
15002	COMMANDER & MRS.	47	CHAPLAIN		

August 4, 2005 Beyond the Formula 10

## What's Hard? Five Unnatural Acts


- Think *Critically*
- Be *Skeptical*
- Focus not on what we know, but on what we *don't know*
- Think first about *Variation*
- Think clearly about *Conditioning* and *Rare* events

August 4, 2005 Beyond the Formula 11

## 1. Think Critically

- **Challenge the data's credentials.**
- **Challenge how they were collected**
  - Look for bias
  - Have they ever done this in Calculus?
- **Know what we want to know.**
  - What's the QUESTION?
- **Plot the data**
  - And ask about lurking variables
- **Check Assumptions and Conditions.**


August 4, 2005 Beyond the Formula 12



## 2. Be Skeptical

- **Being skeptical is part of critical thinking**
  - Be cautious about making claims based on data.
- **“Trust every analysis, but plot the residuals.”**
  - Skeptical statisticians expect the unexpected, so we go looking for it.
- **Question the analysis –**
  - Not just is the answer correct, but
  - Is it appropriate?
  - Did it answer the question
  - Statistical vs. Practical significance


August 4, 2005      Beyond the Formula      13



## 3. Focus on What We Don't Know

- **In most science and math courses, we focus on what we know**
- **Statisticians are a bit weird**


August 4, 2005      Beyond the Formula      14



## Confidence Intervals

- **We don't say “The mean is 31.2”.**
- **We don't say “The mean is probably 31.2”**
- **We don't say “The mean is close to 31.2”.**
- **All we can manage is**
  - “*The mean is close to 31.2.... Probably*
  - *(and, in fact, I'm willing to admit I may be wrong and to spend the effort to give you a whole interval of plausible values and then to spend extra effort to estimate how likely it is that even that interval is wrong.)*”


August 4, 2005      Beyond the Formula      15



## 4. Thinking about Variation

- **Students find it easier to think about values rather than variation, but**  
*Statistics is about Variation*


August 4, 2005      Beyond the Formula      16



## Example

- **A town has two hospitals**
  - Large hospital about 100 babies a day
  - Smaller hospitals about 15 babies a day
- **Over the course of the year, which hospital (if either) would probably have more days in which more than 60% of the babies born are male?**


August 4, 2005      Beyond the Formula      17



## 5. Probability

- **Conditional, joint, rare events; randomness**
  - This is just plain hard.
- **It is easy to show that we don't naturally think clearly about conditional probabilities.**
  - But we need to in order to make rational decisions in the world

August 4, 2005      Beyond the Formula      18




## Linda

(Tversky & Kahneman)

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she participated in antinuclear demonstrations.


August 4, 2005 Beyond the Formula 19



## Order these in order of Likelihood

- a) Linda is a teacher in an elementary school
- b) Linda works in a bookstore and takes yoga classes.
- c) Linda is active in the feminist movement.
- d) Linda is a psychiatric social worker
- e) Linda is a member of the League of Women Voters.
- f) Linda is a bank teller.
- g) Linda is an insurance salesperson.
- h) Linda is a bank teller who is active in the feminist movement.


August 4, 2005 Beyond the Formula 20



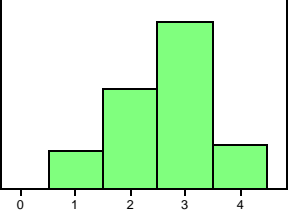
## Pick a number at Random

# 1 2 3 4


August 4, 2005 Beyond the Formula 21



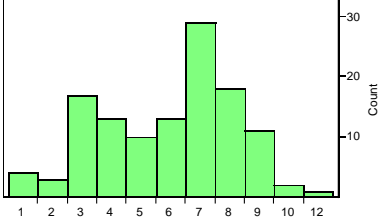
## Random?




August 4, 2005 Beyond the Formula 22



## Random II



August 4, 2005 Beyond the Formula 23



## So, Is Statistical Thinking Unnatural?

- We haven't evolved to be Statisticians.
- Our students who think Statistics is an unnatural subject are right. This isn't how humans think naturally.
- But it is how humans think rationally. And it is how scientists think. This is the way we must think if we are to make progress in understanding how the world works and, for that matter, how we ourselves work.

August 4, 2005 Beyond the Formula 24

## How can we teachers help?

- **Make the course relevant**
  - Examples that speak to students
  - Talk to them about problems – not parameters
  - We have a lot of ground to cover, make it important to them!
- **Give them an outline for putting the real world into a framework**
  - What's the problem? (Plan)
    - The W's
    - The model
    - The method
  - What are the mechanics? (Do)
  - What have we learned? (Report)
  - What next? (Act)

August 4, 2005 Beyond the Formula 25

## Many paths through these woods..

- **Order of topics less important than introducing one concept at a time**
  - Don't say samples, populations, randomness, inference, bias on day 1.
- **Deal upfront with vague concepts**
  - Skewed vs. symmetric
  - Center, spread
  - Unimodal or not?

August 4, 2005 Beyond the Formula 26

## Clearing the woods

- **Trim out unnecessary topics**
  - How to choose bin widths
  - Formulas for grouped data
  - Shortcut formulas
  - Testing mean, sigma known
  - Combinatorics
  - Probability?


August 4, 2005 Beyond the Formula 27

## Start with the Problem

- **Unless you know what the problem is, don't start the analysis**
  - Don't even collect the data
    - CEO at First USA
    - Xerox
- **Identify the data you have**
  - Know the source
  - Identify the W's

August 4, 2005 Beyond the Formula 28

## The W's



Year	Winner	Country	Time	Speed	Stages	Distance	Starters	Finishers
1903	Maurice Garin	Fra	94.33.00	25.3	6	2428	60	21
1904	Henri Cornet	Fra	96.05.00	24.3	6	2388	88	23
1905	Louis Trousselier	Fra	112.18.09	27.3	11	2975	60	24
1906	Rene Pottier	Fra	185.47.26	24.5	13	4637	82	14
1907	Lucien Petit-Breton	Fra	156.22.30	28.5	14	4488	93	33
1908	Lucien Petit-Breton	Fra	156.09.31	28.7	14	4488	114	36
...								
1999	Lance Armstrong	USA	91.32.16	40.3	20	3687	180	141
2000	Lance Armstrong	USA	92.33.08	39.56	21	3662	180	128
2001	Lance Armstrong	USA	86.17.28	40.02	20	3453	199	144
2002	Lance Armstrong	USA	82.05.12	39.93	20	3278	189	153
2003	Lance Armstrong	USA	83.41.12	40.94	20	3427	189	147
2004	Lance Armstrong	USA	83.36.02	40.553	20	3391.1	188	147
2005	Lance Armstrong	USA						

August 4, 2005 Beyond the Formula 29

## Honesty is the best policy


- **Be honest about models**
  - Tell them Statistics is really about models
  - A model is a simplification of reality.
- **We know the model's not perfect**
  - So be sure to check if it's appropriate!

August 4, 2005 Beyond the Formula 30



## All Models are Wrong...

George Box:  
 “All models are wrong... but some are useful”  
 “Statisticians, like artists, have the bad habit of falling in love with their models”



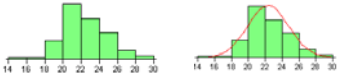
But, statisticians love models--  
*because they are wrong.*

What do we focus on?  
**residuals!**  
 what the model fails to account for

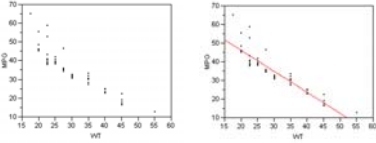
August 4, 2005 Beyond the Formula 31

## Common Models

- Probability models



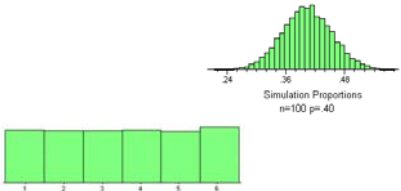
- Regression model



August 4, 2005 Beyond the Formula 32

## Common Models

- Simulation – by “hand” or computer



Simulation Proportions  
 $n=100$   $p=.40$

August 4, 2005 Beyond the Formula 33

## “Pay Dirt” Models

- Sampling distribution models
  - By now students know that models are idealized
  - They’ve seen probability models and simulations: CLT follows naturally
- Null hypothesis models
  - Wrong (we hope) but useful

August 4, 2005 Beyond the Formula 34

## Models...

Require assumptions  
 Because they are idealized, they are only really true under idealized assumptions

Are described by parameters  
 Parameters refer to models of populations, not to the populations themselves

August 4, 2005 Beyond the Formula 35

## Assumptions and Conditions

- Some assumptions we must just assume. (Pretend)
- Many can be checked for plausibility with appropriate conditions
  - Often the conditions are graphical (Remember the 3 rules)
- Few are really true

August 4, 2005 Beyond the Formula 36

## Conditions to Check

- **Summary statistics**
  - Quantitative data condition.

Variable -- TCODE	
Mean	54.41
Std Dev	957.50
Std Err Mean	3.11
upper 95% Mean	60.51
lower 95% Mean	48.31

- **T-test**
  - Assumption is that data are Normal
    - Rule of thumb? 30? 50? 100?
    - Nearly normal condition--make a picture

August 4, 2005      Beyond the Formula      37

## Technology frees us

- **Use Technology**
  - Calculation is for calculators and statistics packages.
  - Let the technology do the work, so students can think about statistical thinking.
  - Let them do it so we can “play Statistics”

August 4, 2005      Beyond the Formula      38

## Play Stats

August 4, 2005      39

## More Help – Reality Checks

- **Emphasize the concepts over the formulas.**
  - The answer is wrong if it makes no sense -- even if you pushed the buttons you meant to push or gave the command you intended
- **Check that the results are plausible**

9. **Professors.** A friend tells you about a recent study dealing with the number of years of teaching experience among current college professors. He remembers the mean but can't recall whether the standard deviation was 6 months, 6 years, or 16 years. Tell him which one it must have been, and why.

August 4, 2005      Beyond the Formula      40

## Report -- Act

- **Plot the data, but then say what you see.**
  - Give guidance for how to “see”
- **Reject the null hypothesis, but then provide a CI to assess effect size.**
  - Emphasize interplay between tests and CI
- **Think about costs and consequences.**
  - Don't be satisfied with “I rejected Ho”


August 4, 2005      Beyond the Formula      41

## Be Honest Part II

- **Acknowledge common misapplications and misinterpretations of statistics.**

What can go wrong?


August 4, 2005      Beyond the Formula      42



## Take Home Messages

- **Stats is about the real world**
  - It's messy – literature not music
  - Motivate by rooting the course in examples and real data that's relevant to students
  - Tell the story of Statistics so students take home a complete picture, not a set of tools
  - Technology frees the student to *think* about the world
  - We need to give the student a structure for a chaotic world (Deming)
    - Make them better problem solvers
  - Help them with unnatural thinking

August 4, 2005      Beyond the Formula      43



## Thank you !!

August 4, 2005      Beyond the Formula      44