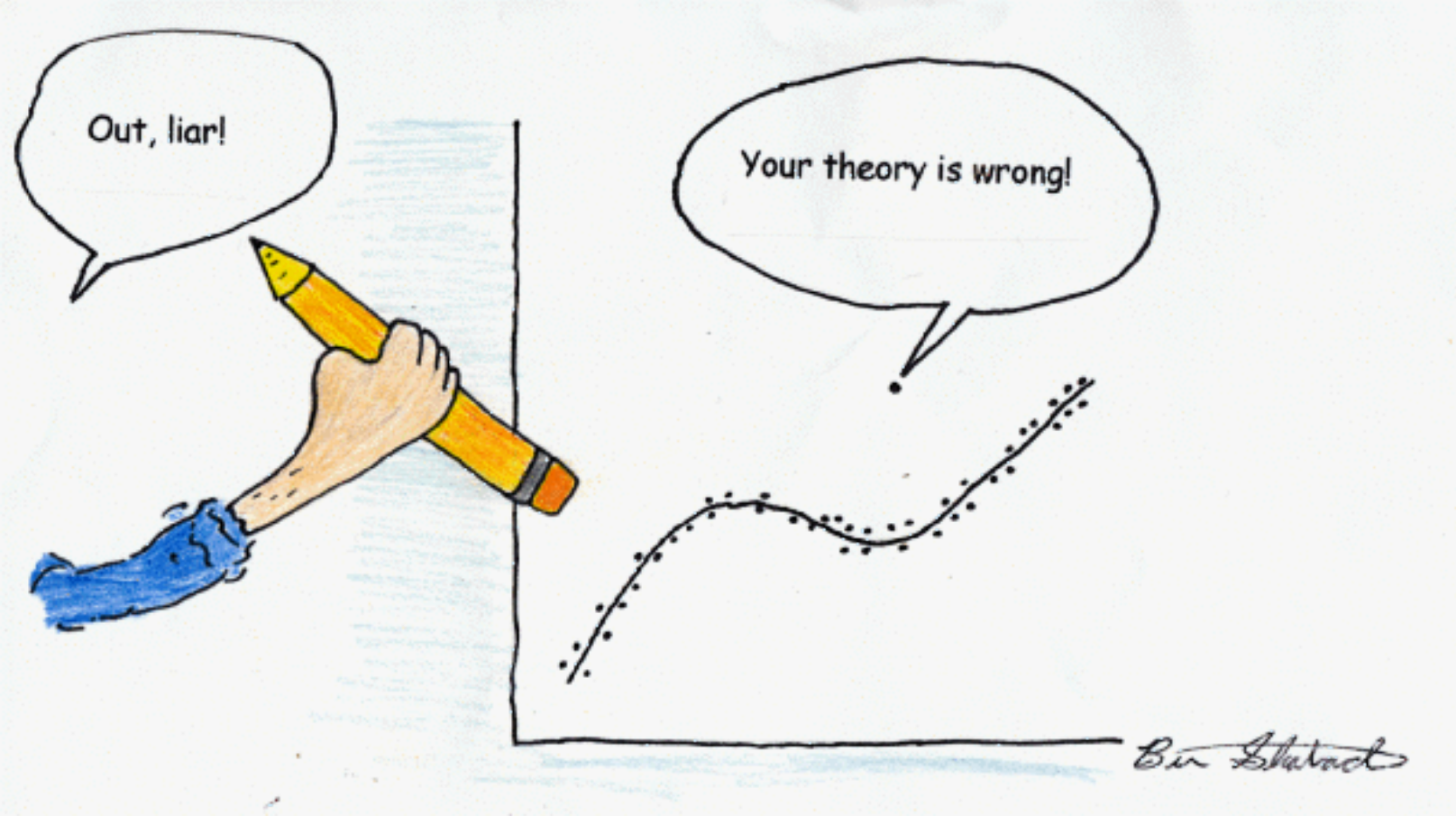
Biology 4605 / 7220
Model Based Statistics in Biology



Biology 4605 / 7220
Model Based Statistics in Biology

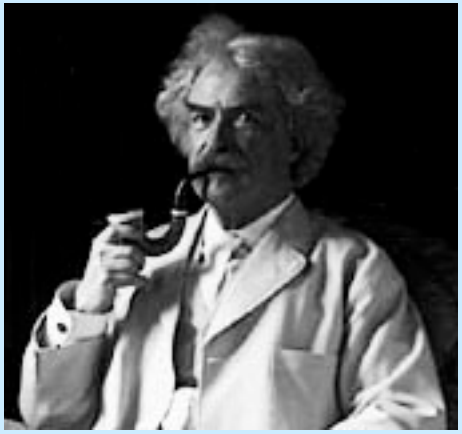
Topics:

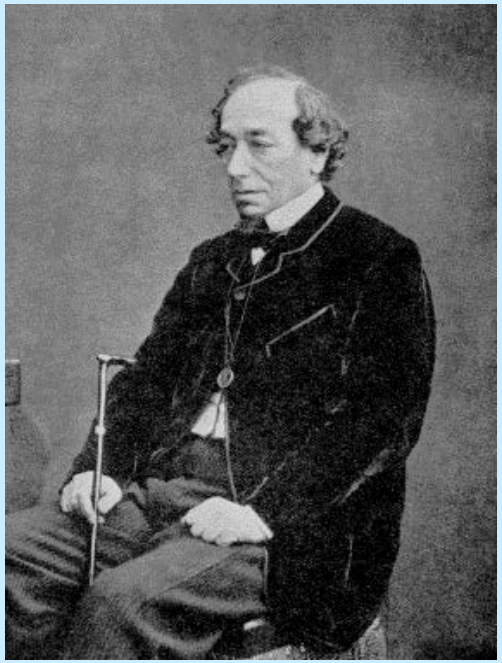
- **Lies, Damn Lies, and Statistics**
- **Statistics are Balderdash - Get rid of them!**
- **Statistics are like taxes - inevitable**
- **Hypothesis testing is statistical flotsam**
- **Model Based Statistics**
- **Discarding more flotsam**
- **Learning model-based statistics**

There are three kinds of lies; lies, damned lies and statistics.

5100 hits with Google in Sept 2002.

Who said it ?

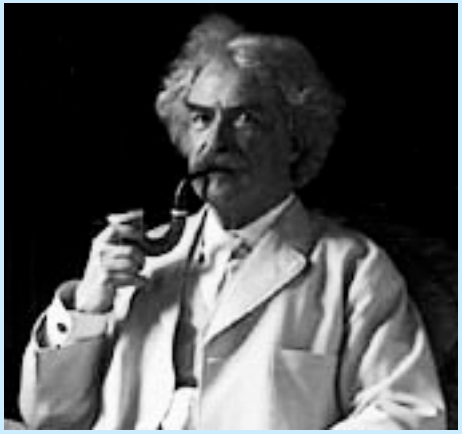




There are three kinds of lies; lies, damned lies and statistics.

5100 hits with Google in Sept 2002.

Who said it ?



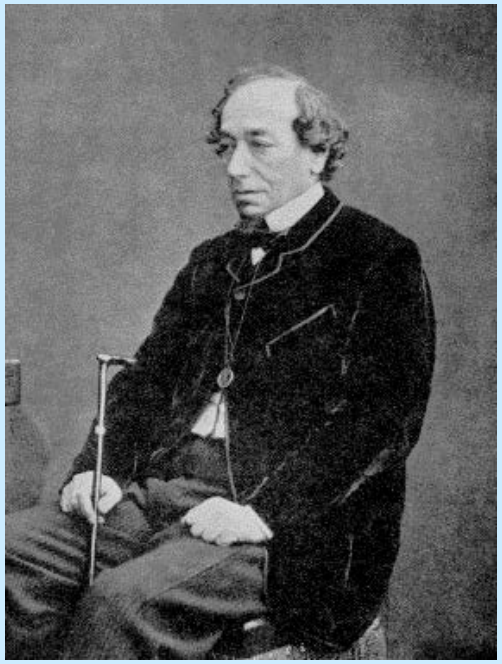
In his Autobiography (1924 vol 1 p 246) Mark Twain attributed the quote to Benjamin Disraeli

Did Disraeli say it ?

There are three kinds of lies; lies, damned lies and statistics.

Did Disraeli say it ?

No



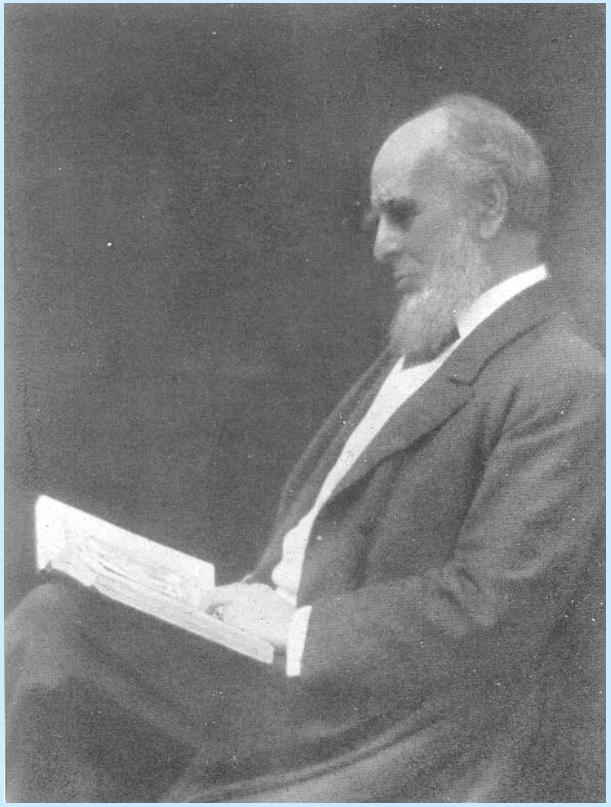
There are three kinds of lies; lies, damned lies and statistics.

If Disraeli didn't say it

then who did?

Leonard Henry Courtney

British economist and politician (1832-1918)



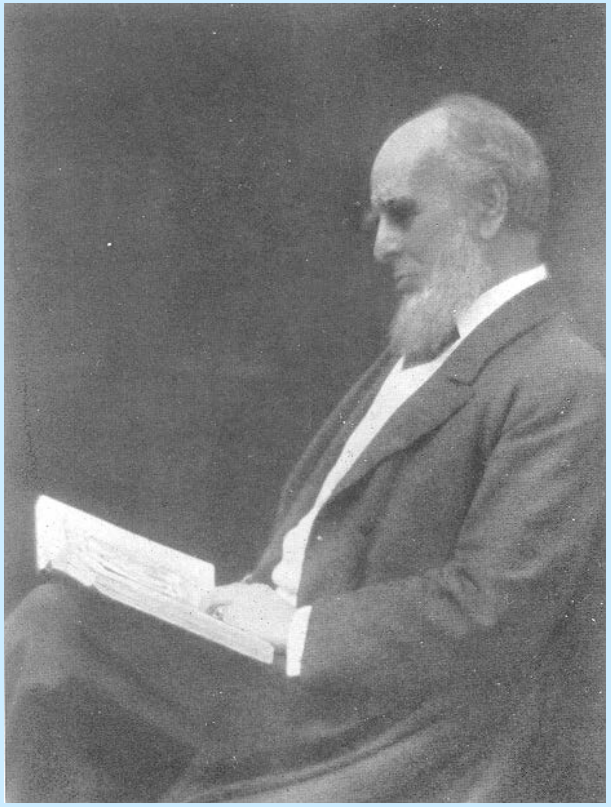
There are three kinds of lies; lies, damned lies and statistics.

The quote can be found in a statistics journal !!!!

Journal of the Royal Statistical Society, No. 59 (1896)

Leonard Henry Courtney

British economist and politician (1832-1918)



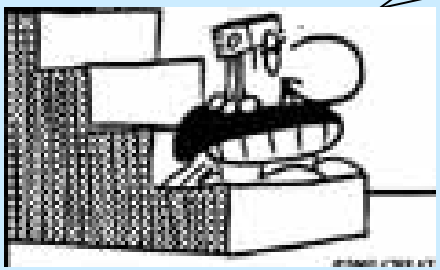
There are three kinds of lies; lies, damned lies and statistics.

If you think lawyers lie, try statistics.

Numbers are to paint what statisticians are to artists.

Paint is to art as numbers are to statisticians.

Numbers are to statisticians what paint is to artists.



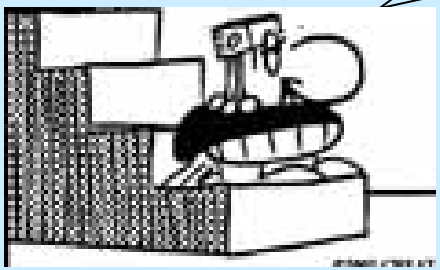
There are three kinds of lies; lies, damned lies and statistics.

When in doubt , use statistics.

Statistics are numbers looking for an argument.

When all else fails, use statistical reasoning.

If you want to really lie use statistics.

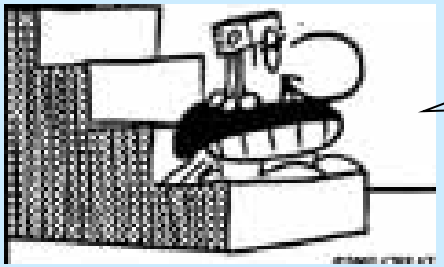


Like other occult techniques of divination, the statistical method has a private jargon deliberately contrived to obscure its methods from non-practitioners.

--G. O. Ashley.

Confusing students is its only function.
It frustrates and mystifies, in conjunction.

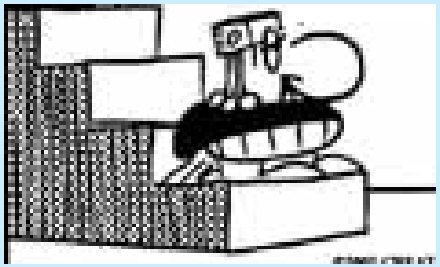
Man without statistics is like a fish without a bicycle.



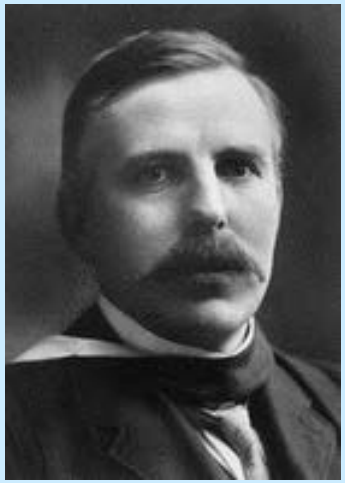
Like other occult techniques of divination, the statistical method has a private jargon deliberately contrived to obscure its methods from non-practitioners.

--G. O. Ashley.

I'm going to abolish statistics from my kingdom.



You can't do that.

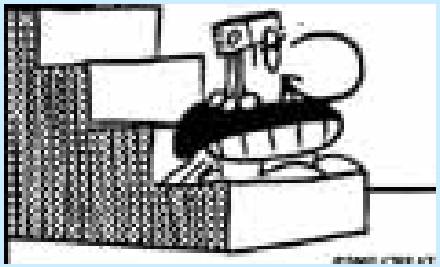


If your experiment needs statistics, you ought to have done a better experiment.

--Ernest Rutherford (1871-1937)

Of course I can.
Lord Rutherford says so.

That's an unsourced quote.

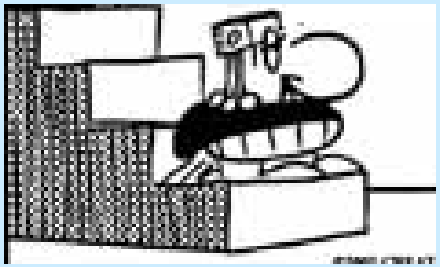


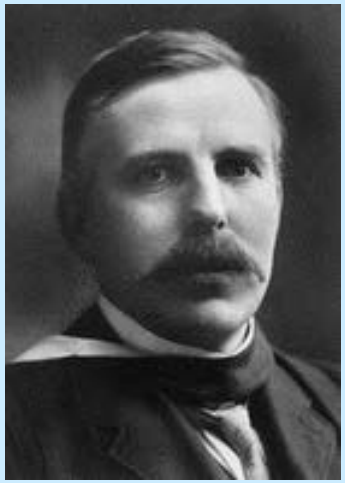
“If your experiment needs statistics, you ought to have done a better experiment.” Attributed to Rutherford by N.T.J. Bailey.

A latin square design uses statistics (two blocking factors in an ANOVA).

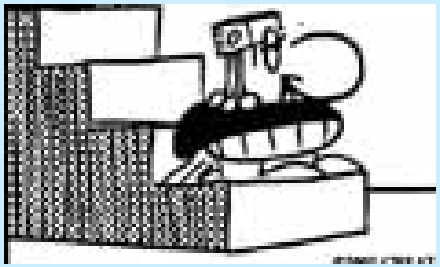
Now according to that quote, if you use a latin square design, you should have done a better experiment.

Once wrong doesn’t mean always wrong.

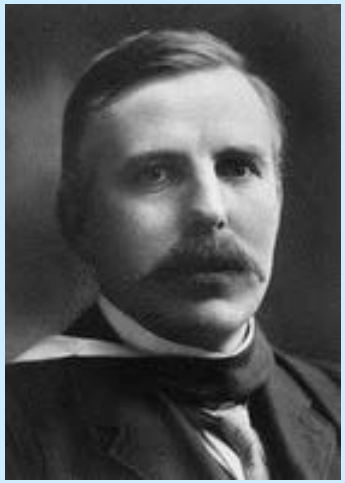




I say again.
Statistics are balderdash.
Get rid of them.



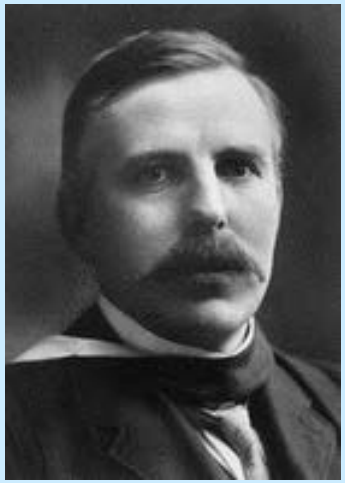
Rutherford did experiments.
Here's a quote.



“If your experiment needs statistics, you ought to have done a better experiment.” Attributed to Rutherford by N.T.J. Bailey.

Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis

--R.A. Fisher 1935



Google hits in July 2018

3689 hits on

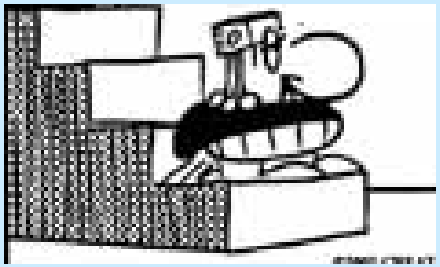
“If you need statistics”

4110 hits on Fisher quote

Fisher won the argument.
5 million hits on “null hypothesis”

That doesn’t prove Fisher was right.

I have it on good authority he was wrong.

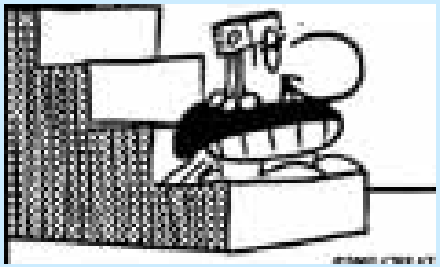


Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis

--R.A. Fisher 1935

It is "extraordinarily difficult to find a statistician who argues explicitly in favor of the retention of significance tests...."

--Oakes 1986

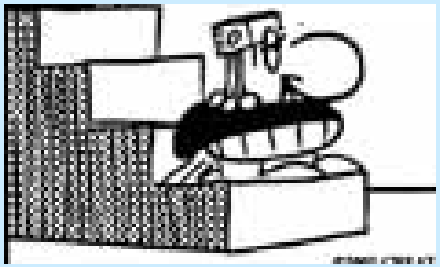


Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis

--R.A. Fisher 1935

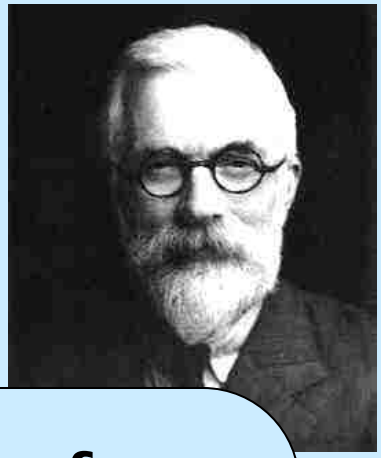
M.R. Nester (1996) collected over 125 quotes against use of hypothesis tests.

D.F. Parkhurst collected 80 quotes against continued use of significance tests (1996)



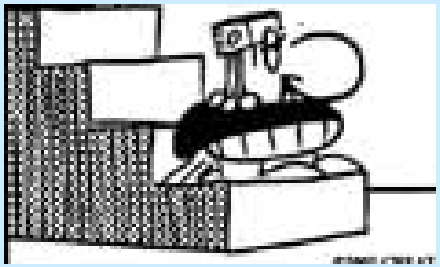
Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis

--R.A. Fisher 1935



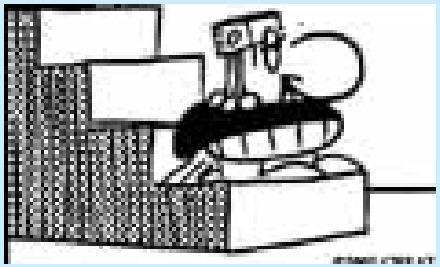
Everyone will have his own pet assortment of flotsam; mine include most of the theory of significance testing, including multiple comparison tests, and non parametric statistics.

John Nelder, 1971. Fisher's successor as Director of the Statistics Department at the Rothamsted Station.



Hypothesis testing is malarky
Get rid of it.

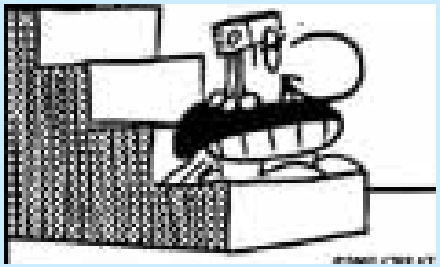
OK, Ok. Let's not get carried away.
Let's get another opinion.



You are always better off thinking about why a model could generate your data and then testing that model.

--L. Wilkinson *et al*. 1992

What's all this stuff about models?



I'm glad you asked.

Learn to write the model instead of memorizing tests.

What test do I use?

Identify a test by name.

Check its assumptions.

Use automated routines provided in a package.

Sort through the output for a p-value.

Report whether p was less than 5%.

Problem solving

What is the response variable?

What are the explanatory variables?

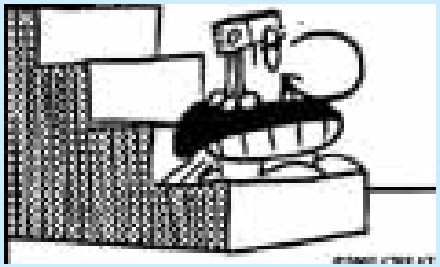
Write the model.

Check the residuals. Model appropriate? Error structure correct?

Take corrective action.

Report the model, parameter values, and standard errors.

What? Me write a model?

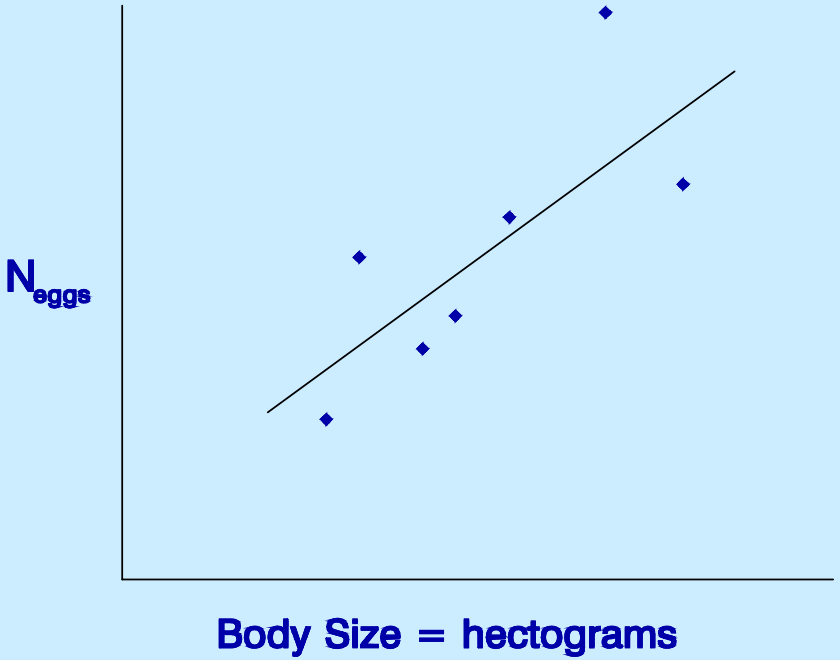


No Problem. Can you write the equation for a straight line?

Regression

Data = Model + Residual

Y = mX + b + Residual



See. No Problem

It's easy.

Write the model.
Discard the search for tests.

The General Linear Model.

Data =	Model	+ Normal Residual	
Data =	[Two means]	+ Normal residual	t-test
Data =	[Several means]	+ Normal residual	Oneway Anova
Data =	[Two factors]	+ Normal residual	Twoway Anova
Data =	[Line]	+ Normal residual	Regression
Data =	[Line + factors}	+ Normal residual	AnCova

While we are at it, let's discard some flotsam.

1. **Statistical tests of assumptions**
2. **Checking assumptions before analysis.**
3. **Non-parametric (rank based) tests.**
4. ***A posteriori* tests.**

Discarding the Flotsam

Statistical tests of assumptions are like going out in a rowboat to see if it is too windy to set sail in the Queen Mary.

--Roger Green

Statistical tests have little capacity to detect problems when it matters (*n* small).

Statistical tests detect small violations when even large violations do not matter (*n* large).

Statistical tests of assumptions reliably lead to the wrong conclusion.

Discarding the Flotsam

Checking statistical assumptions before analysis is like taking rabies shots before before being bitten. --D. Schneider

The assumptions for computing a p-value from F, t, or X^2 distributions:

1. **Residuals sum to zero**
2. **Residuals are homogeneous**
3. **Residuals are independent**
4. **Residuals are normally distributed**

Discarding the Flotsam

Non-parametric test: An assumption-free procedure for investigating a hypothesis of no practical interest.

--S.J. Penn

As biologists, we are more interested in the magnitude of the difference between groups, than we are in the ranking of the groups.

Data = Model + Any Residual

Dept. of Biology

Dept. of Statistics

Discarding the Flotsam

***A posteriori* tests.**

-turns lying over to the computer.

-almost always less effective than planned comparisons.

***A priori* (planned) tests.**

-an important mode of biological reasoning.

-almost always more effective than *a posteriori* comparisons.

Learning
Model-based Statistics

Elementary statistics courses for biologists tend to lead to the use of a stereotyped set of tests:

1 without critical attention to the underlying model involved;

2 without due regard to the precise distribution of sampling errors;

3 with little concern for the scale of measurement;

4 careless of dimensional homogeneity;

5 without considering the ideal transformation;

6 without any attempt at model simplification;

7 with too much emphasis on hypothesis testing and too little emphasis on parameter estimation.

--M.J. Crawley 1993

Gary Ramseyer's 1997 Advice on Teaching Statistics

1. **Utilize group activities**

2. **Use open-book examinations!**

3. **Use power not timed examinations. …another great anxiety-reducing tool.**

4. **I firmly believe that a comprehensive basic course that covers the waterfront of statistical techniques is WORTHLESS.**

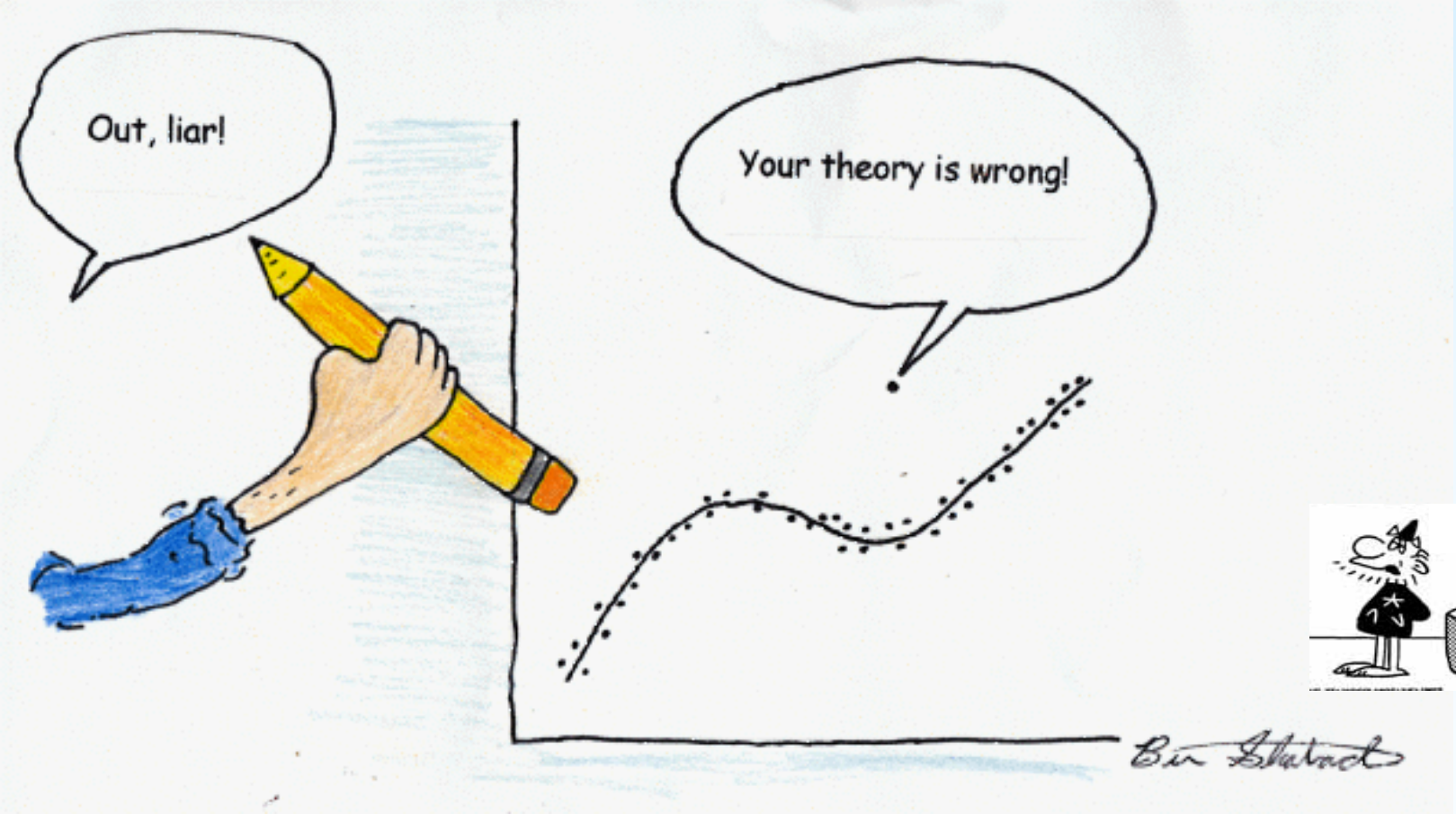
5. **Emphasize the handful of reoccurring themes in basic statistics, such as variability or dispersion of the data.**

6. **Give graded assignments on a weekly basis consisting of one or two problems.**

7. **And . . .**

Gary Ramseyer's 1997 Advice on Teaching Statistics

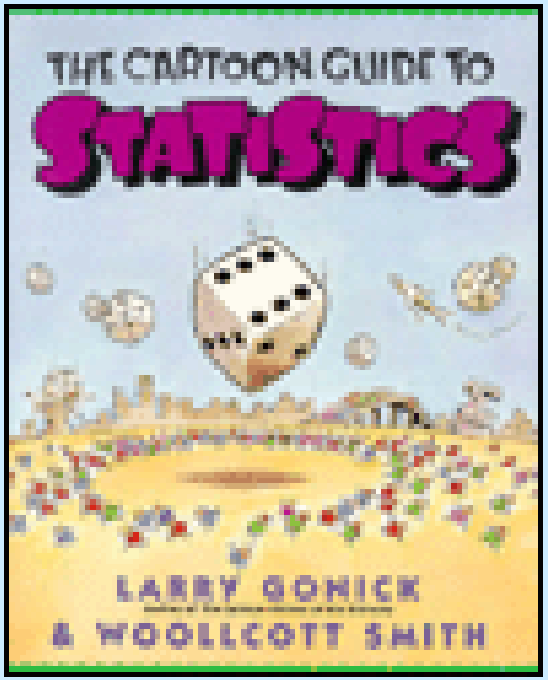
7. Show students how statistical methodology can have relevance in their lives. Above everything else, maintain a sense of humor and don't take yourself so seriously.



Gary Ramseyer's Gallery of Statistical Humor

 # 28 Did you hear about the statistician who went out on a limb to obtain a nested design?

#45 One day there was a fire in a wastebasket in the Dean's office and in rushed a physicist, a chemist, and a statistician . . .



Model based Statistics in Biology

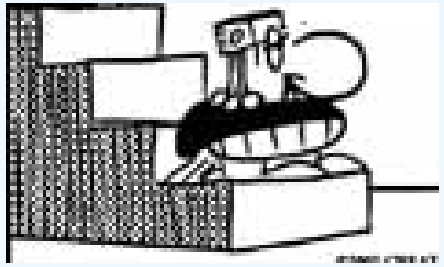
That's all very nice, but . . .

$$\frac{Q(M)}{Q_{ref}(M)} = \left(\frac{M}{M_{ref}}\right)^{\beta}$$

$$Q(M) = (Q(M_{ref}) \cdot M_{ref}{}^{-\beta}) \cdot M^{\beta}$$

$$Q(M) = k_{ref} \cdot M^{\beta}$$

I don't have time and it's Greek to me . . .



Models in Biology. And the good news is . . .

	Spreadsheet	Statistical packages				
	Excel	**MTab**	**SAS**	**SPSS**	**Splus**	**Systat**
Spreadsheet visible	T	T	No	T	T	T
Pull down menus	T	T	No	T	T	T
Basic statistical functions	T	T	T	T	T	T
Return a column of data in random order.	No	T	T	T	?	?
General Linear Model	??	T	T	T	T	T
Residual analysis	T	T	T	T	T	T
Logistic regression	No	T	T	T	T	T
Generalized Linear Model	No	No	T	T	T	T

Model Based Statistics in Biology

They are easy in one of these packages.

They can be taught at the 2nd year level in University.

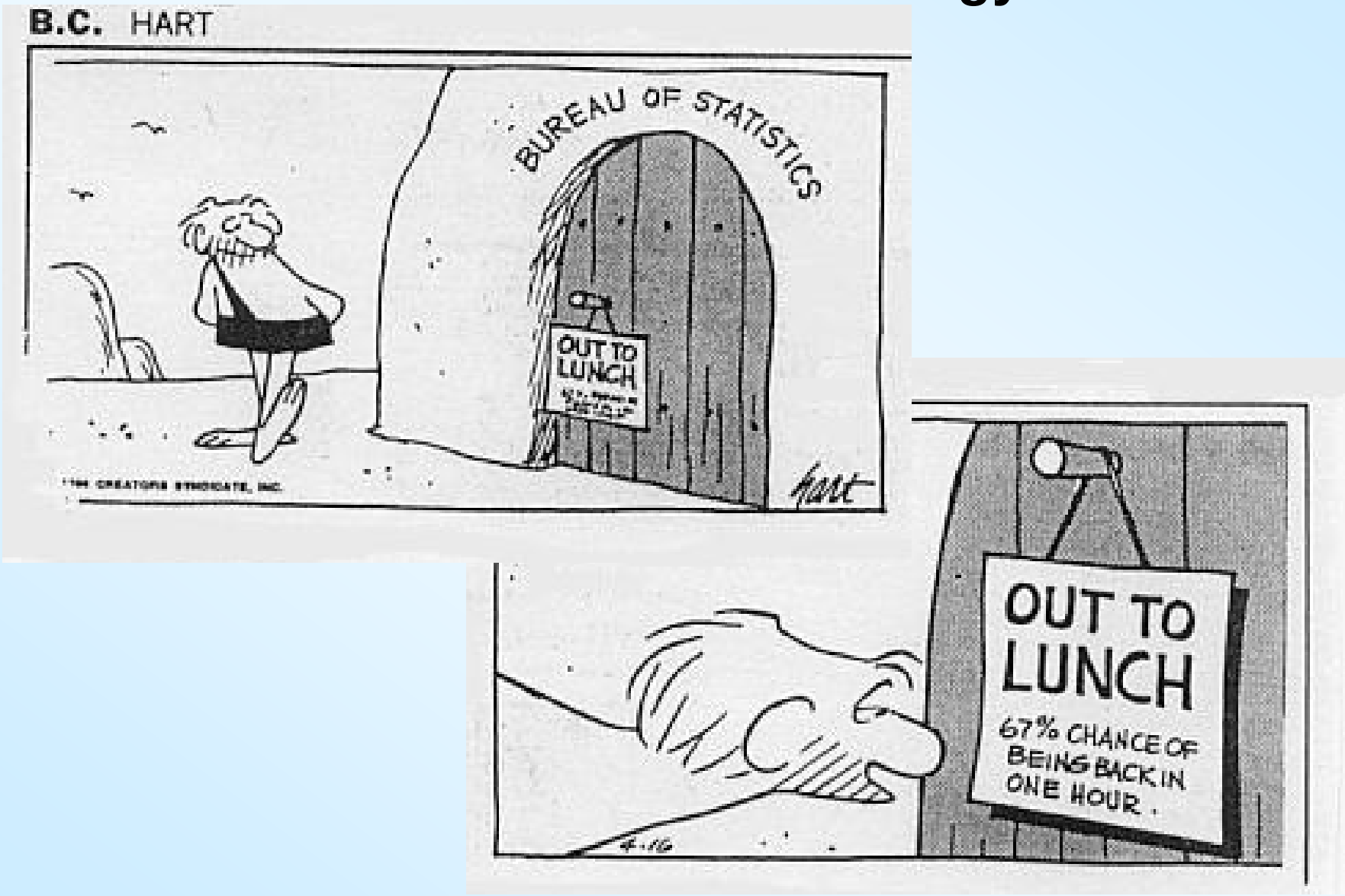
It's better practice than looking up some test.

It's quicker than looking up some test.

Then I decree that all must take this course . . .

Model Based Statistics in Biology

Then I decree that all must take this course . . .

Let's not get carried away….

Statistical Models in Biology