

Assessment in Statistics: A two-edged sword



Introduction

- A two edged-sword - can damage as well as do good work
- Does the way we assess maintain and improve standards or does it lower them
- Promoting Statistical Illiteracy as well as literacy

Two Examples

- Response of a Secretary of State to their TIMSS results
- *We must teach to the test better*
- An example from testing the English National Curriculum
- Objective 5/4c. Use the mean and the range of a set of data



The Two Buses



- I can either catch bus 1 or bus 2 on my way home from school. This is how long I had to wait for bus 1 on my last 5 journeys:
10 min 8 min 5 min 9 min 8 min

This is how long I had to wait for bus 2 on my last 5 journeys:
16 min 1 min 2 min 15 min 1 min

Continued



- The mean waiting time for bus 1 is 8 min and for bus 2 is 7 min. The range for bus 1 is 5 min and for bus 2 is 15 min
- Use the mean and the range to choose which is the best bus to catch.

Assessment

- Drives learning and teaching
- WYTIWYG
- We live in a society unhealthily preoccupied with assessment



Types of Assessment

- Illustrated by why the assessment is being done: e.g.
- The student to know what (s)he does & does not know
- The teacher to know what the student knows
- The teacher to be able to judge effectiveness of teaching
- The institution to give a grade
- Society to assess the effectiveness of the teacher or institution. This gets confounded with raising and maintaining standards, accountability, targets

Purposes of assessment (i)

- Many purposes; formative, summative etc. but fundamentally
- To ensure that students have learned and can apply
- But 'learned' at what level
- And the nature of 'apply'



Purposes of assessment (ii)

- Formative: ... appropriate next steps
- Diagnostic: ... remedial help
- Summative ... overall achievement
- Evaluative ... assess work of school/institution



Summary of purposes

- Formative and diagnostic is to improve student learning - collaborative
- Summative is to give a grade or a mark - judgmental
- Evaluative is to grade an institution - a distorting pressure



Good Formative Assessment

- Is embedded in the teaching & learning
- Shares learning goals with students
- Helps students know & recognise standards
- Provides feedback to aid improvement
- Is committed to every student improving
- Involves students in self-assessment
- Involves students & staff in continual review of progress

Marking & feedback

- Decrease the use of extrinsic awards
- Provide more oral feedback rather than relying on marking/written feedback
- Help students develop skills in reviewing their own work & that of peers
- Be clear whether to get pupils to correct their own work based on purpose served

• From: Good Assessment Practice in Mathematics. HMI 1477. Office for Standards in Education

Bloom's Taxonomy



Form of learning	Verbs
Knowledge	Define. State. List. Recognise Show. Label. Name. Identify
Comprehension	Explain. Clarify. Discuss. Review. Describe. Recognise Interpret. Relate.
Application	Demonstrate. Use. Calculate Estimate. Fit. Implement. Apply
Analysis	Investigate. Solve. Interpret Explore. Analyse. Explain. Compare. Contrast.
Synthesis	Design. Formulate. Model. Improve. Adapt. Develop. Construct. Devise. Combine.
Evaluation	Assess. Criticise. Contrast. Review. Distinguish. Decide.

Methods of assessment

Form of learning	Verbs	Methods
Knowledge	Define. State. List. Recognise Show. Label. Name. Identify	quiz part exam question
Comprehension	Explain. Clarify. Discuss. Review. Describe. Recognise	Multi choice, assignment exam
Application	Demonstrate. Use. Calculate Estimate. Fit. Implement.	Coursework. Apply specific techniques to particular problem
Analysis	Investigate. Solve. Interpret Explore. Analyse. Explain. Compare. Contrast.	Analyse a data set Oral presentation. Case studies Analyse computer output
Synthesis	Design. Formulate. Model. Improve. Adapt. Develop. Construct. Devise. Combine.	Design and carry out an investigation. Prepare a report & Present. Essay. Open ended questions.
Evaluation	Assess. Criticise. Contrast. Review. Distinguish.	Critically assess others' projects Defend own presentation

LTSNAssessment with a purpose

Summative & Evaluative Assessment Distort



- An example from UK Health Service
- The false logic. If X knows this they will be able to do Y; they can do Y therefore they know X
- Summative and evaluative assessment encourage defence mechanisms and breed fear
- Formative assessment requires trust, no judgement and openness

How psychology students see assessment (Newstead 2003)

- It is to grade people
- Developing an argument is low priority; amount & accuracy of information covered is important
- Assessment neither motivates students nor promotes learning.
- Even highly motivated students resort to a shallow approach - this is adaptive if the student is to take a variety of different assessments in a short time period
- No correlation between learning approach & final degree

A Quote



- *We are in danger of preparing our students for a life of tests rather than the tests of life*



Improving Quality - apply Deming Approach

- Developed in Industry: see how it applies to education and assessment
- Quality improvement NOT quality control
- Quality is delighting the customer, not just meeting his/her requirements
- Work with the customer to decide on what customer considers quality and how it might be measured to show improvement (but the aim is quality, not the measure)
- Empower the person who is at the point of decision making

The 14 Points (see handout)

1. Constant purpose to improve;
2. adopt new philosophy of leadership;
3. stop using inspection to achieve quality, build in quality;
4. single trusted supplier;
5. improve constantly;
6. on job training;
7. leadership to help people do a better job;

The 14 Points (contd)

8. drive out fear;
9. break down departmental barriers
10. eliminate slogans & targets;
11. eliminate quotas & management by numbers - substitute leadership;
12. remove barriers that stop pride in workmanship - incl. annual or merit rating;
13. institute programme of education & self-improvement;
14. transformation to quality is everybody's job.

Apply to Education (1)

- Constantly and continually seek for improvement (1,5)
- Applies to students, teachers, administrators, governments
- In the context of quality as previously defined
- Consider who is the customer and who the supplier



Apply to Education (2)

- Leadership is to help people do a better job (7) take responsibility that leads to change (2)
- Again applies to all



Apply to Education (3)

- Eliminate quotas, targets, numerical goals (11) and inspection for quality (3) build quality in from the beginning
- Develop pride in work & personal responsibility (12a) remove barriers such as inappropriate rewards (12b)



Apply to Education (4)

- Training on the job (6) education and self improvement (13)
- Note the delegation of responsibility to the receiver
- Drive out fear (8) slogans & exhortations (10) encourage cooperation (9) with all involved (14)



Two Principles

- The *CONTENT* principle: assessment should reflect the content that is most important for students to learn
- The *PROCESS* principle: assessment should enhance the sound development of concepts, insights and deeper levels of thinking



An Example (from many)

- 5. The managing director of a company is keen to assess the impact that retail prices may have on sales. He therefore obtains, in a summarised form, the following information from the Monthly Digest of Statistics.

Item Group	Single Item Index 1995	Single Item Index 1998	Weight
Food	250	291	208
Alcoholic Drink	305	310	77
Tobacco	358	364	38
Housing	308	320	149
Fuel and Light	391	402	67

Question (contd)

- Show that the total weighted index of retail prices for 1995 is 299 (3 marks)
- Work out the equivalent weighted index of retail prices for 1998. (3 marks)
- Hence calculate the all item (aggregate) weighted index of retail prices for 1998 using 1995 as the base year. (2 mks)
- Using your earlier results, if the food group were excluded from the calculations in part (c) what effect would this have had on the resultant index? (2 marks)

What is NOT asked?

Purposes of assessment (ii)

- the student to know
- the teacher to know what the student knows
- the teacher to know effectiveness of teaching
- the award of a grade
- Society to assess effectiveness

The End

What makes a first course in statistics?

Peter Holmes
RSS Centre for Statistical Education

Statistical Literacy

- *Growing literature*
- *Many terms: numeracy, statistical numeracy, statistical literacy etc*
- *ability to use numbers in practice - particularly in context of statistics*
- *reflect how statistics is used in practice*

An early definition from UK

- Statistical numeracy requires a feel for numbers, an appreciation of levels of accuracy, the making of sensible estimates, a common sense approach to data in supporting an argument, the awareness of the variety of interpretation of figures and a judicious understanding of widely used concepts such as mean and percentages. All these are part of everyday living.

What is Statistics?

- What statisticians do!!
- Statistics changes numbers into information.
- Statistics is the art & science of deciding what are the appropriate data to collect, deciding how to collect them efficiently and then using them to give information, answer questions, draw inferences and make decisions.

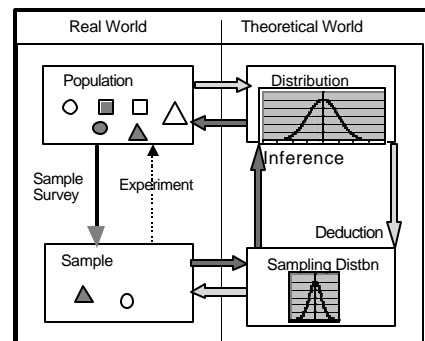
What is Statistics?

- Statistics uses the language & ideas of probability to describe inferences and risk.
- Statistics uses samples to get insight into different populations.
- Statistics is making decisions when there is uncertainty

On to a first course

- After a basic statistical literacy there are different needs for different types of people (the businessman differs from the experimental scientist)
- The strength of statistics based on:
 - Well-designed experiments - chance
 - Small observational studies - bias
 - Large observational studies - confounding

The Statistical Process



What does the Sample tell us about the Population?

- Well designed experiments : avoid confounding and bias, use chance
- Small surveys try to match the 'gold standard' to eliminate bias, typically estimate proportions and probability
- Large surveys or observational studies - essentially we know the population

Large observational studies

- Are growing in number- often connected with legislation or business practice
- May be both sample and population e.g patients attending hospital are a (poor) sample of the neighbouring population; the total population of those who attend and a sample of people who attend hospitals so data are used for league tables.

The High C's of Curriculum Development

- Content
 - Technique
 - Process
- Context
- Concept
 - Level of understanding required
 - For what purpose

The big ideas in Statistics

- bias (in sample, questionnaire, response)
- causality, census, chance,
- conditional probability, confidence interval,
- confounding, correlation,
- cross-section study, dependent
- distribution (population, sample, probability)

The big ideas in Statistics

- error, estimate, experiment,
- explanatory variable, fit, forecast.
- hypothesis test, independent, index number,
- inference, interaction, longitudinal study,
- mean, measuring scale, model (modelling)
- moving average, multivariate ,
- observational data, outlier, percentage

The big ideas in Statistics

- population, power of test, predict
- probability, random, rate, representative
- response variable, risk, sample,
- seasonality, significance, spatial statistics
- standard deviation, standard error
- standardise, survey, time series, trend
- type 1 error, type 2 error, utility,
- variability, variable, variance

Implications for the first course

- Too much in previous slides for one course
- What do they know already?
 - From general background
 - With an AP Statistics background
- What do they need to know?
 - For general education purposes
 - For their particular subjects

Implications for the first course

- The balance between *use* and *do* statistics
- The place of general examples such as those in *Statistics a Guide to the Unknown*
- Some specific things for specific majors
 - Economics: time series, index numbers
 - Industry: statistics process control
 - Business: reading tables; understand confounding, surveys
 - Psychology: experimental design, significance
 - Government: spatial and time series

An example:

Relationships & causes

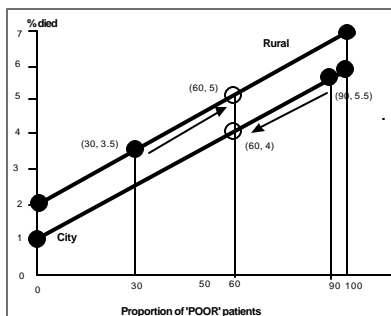
- *Correlation is not causation* is essentially a negative message - we can go beyond that
- Just to identify possible confounding factors is again essentially negative for decision making
- To be able to allow for the confounding effect means that it is possible to make stronger claims about cause and effect
- Simpson's paradox is just an extreme case of the effect of a confounding variable

The Hospitals

Hospital	Patient Conditions	Numbers	# died	% died
City	Good	100	1	1.0%
	Poor	900	54	6.0%
	Total	1000	55	5.5%
Rural	Good	700	14	2.0%
	Poor	300	21	7.0%
	Total	1000	35	3.5%

- Note that *City* hospital has better record for *good* and for *poor*, but overall *Rural* hospital has better record.
- Confounded by proportion of *poor* patients

The Hospitals



The Hospitals

- At every proportion of *poor* patients the Rural Hospital had a 1% higher death rate than the City Hospital
- Raw totals compare 90% *poor* in *City* with 30% in *Rural*
- Overall there were 60% *poor* so standardize and compare there.

The End

What makes a first course in statistics?

Peter Holmes
RSS Centre for Statistical Education

Statistical Literacy

- *Growing literature*
- *Many terms: numeracy, statistical numeracy, statistical literacy etc*
- *ability to use numbers in practice - particularly in context of statistics*
- *reflect how statistics is used in practice*

An early definition from UK

- *Statistical numeracy requires a feel for numbers, an appreciation of levels of accuracy, the making of sensible estimates, a common sense approach to data in supporting an argument, the awareness of the variety of interpretation of figures and a judicious understanding of widely used concepts such as mean and percentages. All these are part of everyday living.*

What is Statistics?

- *What statisticians do!!*
- *Statistics changes numbers into information.*
- *Statistics is the art & science of deciding what are the appropriate data to collect, deciding how to collect them efficiently and then using them to give information, answer questions, draw inferences and make decisions.*

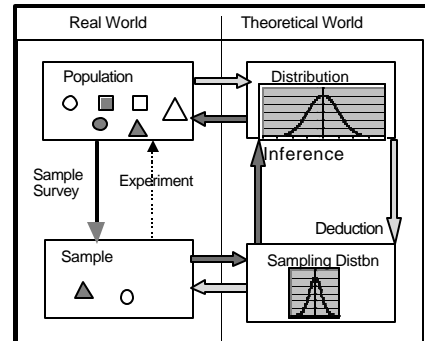
What is Statistics?

- *Statistics uses the language & ideas of probability to describe inferences and risk.*
- *Statistics uses samples to get insight into different populations.*
- *Statistics is making decisions when there is uncertainty*

On to a first course

- After a basic statistical literacy there are different needs for different types of people (the businessman differs from the experimental scientist)
- The strength of statistics based on:
 - Well-designed experiments - chance
 - Small observational studies - bias
 - Large observational studies - confounding

The Statistical Process



What does the Sample tell us about the Population?

- Well designed experiments : avoid confounding and bias, use chance
- Small surveys try to match the 'gold standard' to eliminate bias, typically estimate proportions and probability
- Large surveys or observational studies - essentially we know the population

Large observational studies

- Are growing in number- often connected with legislation or business practice
- May be both sample and population e.g patients attending hospital are a (poor) sample of the neighbouring population; the total population of those who attend and a sample of people who attend hospitals so data are used for league tables.

The High C's of Curriculum Development

- Content
 - Technique
 - Process
- Context
- Concept
 - Level of understanding required
 - For what purpose

The big ideas in Statistics

- bias (in sample, questionnaire, response)
- causality, census, chance,
- conditional probability, confidence interval,
- confounding, correlation,
- cross-section study, dependent
- distribution (population, sample, probability)

The big ideas in Statistics

- error, estimate, experiment,
- explanatory variable, fit, forecast.
- hypothesis test, independent, index number,
- inference, interaction, longitudinal study,
- mean, measuring scale, model (modelling)
- moving average, multivariate ,
- observational data, outlier, percentage

The big ideas in Statistics

- population, power of test, predict
- probability, random, rate, representative
- response variable, risk, sample,
- seasonality, significance, spatial statistics
- standard deviation, standard error
- standardise, survey, time series, trend
- type 1 error, type 2 error, utility,
- variability, variable, variance

Implications for the first course

- Too much in previous slides for one course
- What do they know already?
 - From general background
 - With an AP Statistics background
- What do they need to know?
 - For general education purposes
 - For their particular subjects

Implications for the first course

- The balance between *use* and *do* statistics
- The place of general examples such as those in *Statistics a Guide to the Unknown*
- Some specific things for specific majors
 - Economics: time series, index numbers
 - Industry: statistics process control
 - Business: reading tables; understand confounding, surveys
 - Psychology: experimental design, significance
 - Government: spatial and time series

An example:

Relationships & causes

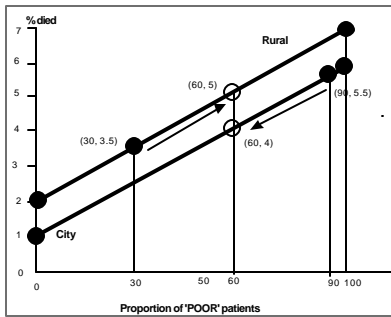
- *Correlation is not causation* is essentially a negative message - we can go beyond that
- Just to identify possible confounding factors is again essentially negative for decision making
- To be able to allow for the confounding effect means that it is possible to make stronger claims about cause and effect
- Simpson's paradox is just an extreme case of the effect of a confounding variable

The Hospitals

Hospital	Patient Category	Numbers	# died	% died
City	Good	100	1	1.0%
	Poor	900	54	6.0%
	Total	1000	55	5.5%
Rural	Good	700	14	2.0%
	Poor	300	21	7.0%
	Total	1000	35	3.5%

- Note that *City* hospital has better record for *good* and for *poor*, but overall *Rural* hospital has better record.
- Confounded by proportion of *poor* patients

The Hospitals



The Hospitals

- At every proportion of *poor* patients the Rural Hospital had a 1% higher death rate than the City Hospital
- Raw totals compare 90% *poor* in *City* with 30% in *Rural*
- Overall there were 60% *poor* so standardize and compare there.

The End