THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Research Methods comp4809/7809 Week 9

Evaluation and Analysis of Research

Prof Jane Hunter
jane@itee.uq.edu.au

Overview

- Week 1: Getting started
- Week 2: Find literature and methods
- Week 3: Understand the issues, build methods skills
- Weeks 4-5: Start writing, make progress
- Week 6: Draft report, refine your understanding
- Week 7: Invited Lecture, Full report

Overview

Weeks 8-11 - Communication

- Week 8: Presentation skills;
  - Elevator pitch
- Week 9: Evaluation and analysis of research;
  - Presentation of seminar title and take home message
- Week 10: Invited lecture (Andrew Bradley);
  - How to do a poster and demo
- Week 11: Invited lecture (Penny Sanderson)
- Week 12: Invited lecture (Jane Hunter);
  - Seminar Debrief
- Week13: Planning the rest of the project;
  The project timeline revisited

D6 Elevator pitch

- An elevator pitch starts with a big picture problem, and shows how it can be addressed

  Security is a major issue for file transfer. The problem is not just the security of files themselves, but also a lack of understanding of the problem in the community. I'm developing software for doing XYZ so that the problem can be taught better. In particular …

- After listening to a pitch, ask yourself
  - Why is this work important? (80% of the msg)
  - What is the person going to do? (20% of the msg)

- A pitch should first engage the listener (the hook), and then sell them the solution.
  - If a pitch starts with "My project is to do X" it misses the engagement
  - If a pitch doesn't say what you will do, it misses the solution

D7 Draft of seminar presentation

- Title of your seminar
- Take home message of your seminar (20-50 words)
- Topic for each slide. Either
  - Power Point draft; or
  - List the contents of the slides in text form
- Write five questions that you would like the audience to ask you at the end of your seminar.

Lecture 9
Evaluation and analysis of research

- Next Semester: What will you write in the second half of your project report?
  - Results
  - Discussion
  - Conclusions
- Plan your analyses now

Plan your analyses now

- The time to think about methodology, collation of results, and analyses or evaluation is when studies are first designed.
- This information is important both for the planning stages of research, and also for the presenting your work in the seminar.

Why plan now?

- The most common complaint of statistics advisors is that the data is inherently flawed because information is missing or the wrong information was collected
- The *right* time to design analyses is when a project is designed.

Evaluation has two parts

- How will you know whether your project works?
  - Record results (observation)
  - Analyse the results
  - Interpret the analyses
  - Make conclusions
- How will you convince your readers of your conclusions?
  - Clearly present the evidence
  - Clearly show how your conclusions arise from your evidence

What do the following mean?

- Methodology
- Methods
- Results
- Analysis
- Conclusions

- What's the difference between
  - a research program
  - a research project

Definitions

Methodology covers all the procedures used in a field to discover new information.

- In some fields it might involve mathematical proofs, in another it might be neural network simulations, extreme programming, ethnographic studies, design and development of new software
- “Methodology” covers the general case – the framework for your area

Methods are the specific application of a methodology to your project.

Research Methods

- Surveys
  - Literature (journals, magazines, conferences, books, Web)
  - Research/commercial, hardware/software, systems, standards
- Theoretical research (improve understanding)
  - Theories, Models, frameworks
- Applied research
  - Implementations/applications – something new or a new combination of existing tools
  - Solve problem or generate improvements – service, efficiency, cost, flexibility, scalability
- Comparative Analyses/Evaluation
  - qualitative and quantitative
- Usability Testing and user feedback

Definitions

Results are the indisputable numbers or the facts that are observed.

Analysis or evaluation is how to turn the observed facts into meaningful knowledge.

Conclusions are the claims that are made about what you've discovered or produced.

Example:
Comparing two algorithms

Methodology covers all the procedures used in a field to discover new information.
Methods are the specific application of a methodology to your project.
Results are the indisputable numbers or the facts that are observed.
Analysis or evaluation is how to turn the observed facts into meaningful knowledge.
Conclusions are the claims that are made about what you've discovered or produced.

- Methodology could be computational simulations of the two algorithms on a set of benchmark problems
- Methods are the details of the algorithms, including parameters, and the specific set of problems.
- Results to be collected could be their performance on a set of tasks. The results are likely to be collated into tables with standard deviations and presented as graphs with error bars.
- Analyses could include a statistical test to determine if the performances were statistically different. They could include comparing the resources used by the algorithms, and comparison to other algorithms (eg to best linear tests known).
- Conclusions could be that one algorithm performed better than the other (supported by the statistical test) on a particular type of task, but the other algorithm was superior on tasks with different requirements.

D8 Progress to date: Description of preliminary results or pilot studies

- Describe the progress made on the substance of the project to date, including description of pilot studies, code written etc. (300-500 words)

- **Rationale.** Efforts at this stage of semester are often directed towards the seminar, but you also need to be making progress on the substance of the project.
- **Tip.** Good time management is useful for keeping a project on track.

D9 Description of Research Methodology and Analysis or evaluation techniques

- Clearly describe the planned methodology for your project. Use diagrams where relevant. This should be in a form that may be directly useful for your final project report (minimum 1 page)

- Describe the results you intend to collect and the form in which they will be collated. Use tables or graphs to show the relevant variables or axes. (minimum 1 page)

- Clearly describe the analysis or evaluation techniques that are relevant to your project. Refer to particular analyses in the literature, and where relevant include specific diagrams that you will be using as a model for your own project. (minimum 200 words)

Scenario:
HCI Study of immersion in games

- Situation
  - A local company is interested in making highly profitable games and employs a researcher to study factors that affect enjoyment in games.
  - Initial studies indicate that immersion (the ability to become totally absorbed) is one important factor.
  - The company wants further research to study what factors break immersion in games.

- Research project
  - The researcher sets up a server and invites gamers to play, and observes players
  - The methodology is called ‘cyber ethnography’
  - The methods are how cyber ethnography is used in this case

Moving from methodology to a specific set of methods

- There are an infinite number of things to observe but the observer has finite attention
  - How will you decide on the factors?

- What will you report to the sponsors?
  - What to observe
  - How to record it
  - How to collate the findings into a report
  - How to interpret the findings
  - What to tell the sponsors

Don’t confuse these concepts!

- Hunch
- Speculation
- Hypothesis
- Proof

The nature of evidence: Evaluating classification results

- A classification system is any system where events or objects are sorted into classes. Examples include
  - A medical test (positive or negative)
  - Comparing a new machine learning algorithm with an existing one

- The accuracy of a classification system can be reported in several ways.
  - The simplest (and often most misleading) is percentage correct.
    - Its advantage is ease of understanding and comparison with other results.
    - However, it is a single figure, and is influenced by many different factors, the main one being threshold cutoff for classification.

- The initial distribution in the training samples has a major impact on performance of a classifier. Consider a fictitious medical scenario…

Case 1: Medical test results

- A man goes to his doctor for a routine checkup, including a test for <disease> required for insurance purposes. He has no lifestyle or health factors that indicate a high risk for <disease>, but the test comes back positive.

- You are the doctor and you know:
  - The diagnostic test is 99.99 % accurate^.
  - <disease> occurs in his community in about 1 in 10,000 people.
  - What is the chance that the person actually has <disease>?
    a. 99.99%
    b. highly likely (much more than 50%)
    c. 50%
    d. highly unlikely (<< 50%)
    e. 0.01%

^The accuracy is actually much lower in reality.

Case 1: cont

- Accuracy by itself can be misleading:
  - The test could be 99.99% accurate just by predicting that no one had the disease.

- If the test were 99.99% accurate on both positive and negative cases, then the information can be rephrased:
  - In 10,000 cases, the test will be right 9,999 times & wrong once
  - In 10,000 cases, 9,999 people will not have the disease
  - So if 10,000 people are tested, how many people will test positive?

Four possible outcomes in a diagnostic test

- negative cases will test negative (true negatives)
- negative cases will test positive (false positive)
- positive cases will test positive (true positive)
- positive cases will test negative (false negative)

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease			
Patient healthy			
Total			

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease			**1**
Patient healthy			**9999**
Total			**10,000**

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease	.9999	.0001	**1**
Patient healthy			**9999**
Total			**10,000**

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease	.9999	.0001	**1**
Patient healthy	.9999	9998.0001	**9999**
Total			**10,000**

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease	.9999	.0001	**1**
Patient healthy	.9999	9998.0001	**9999**
Total	**1.9998**	**9998.0002**	**10,000**

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease	.9999	.0001	**1**
Patient healthy	.9999	9998.0001	**9999**
Total	**1.9998**	**9998.0002**	**10,000**

Restating the odds provides a clear way to describe the results

- 10,000 cases are likely to produce one true positive and one false positive.
  - Thus, the man's chances are ~50%.

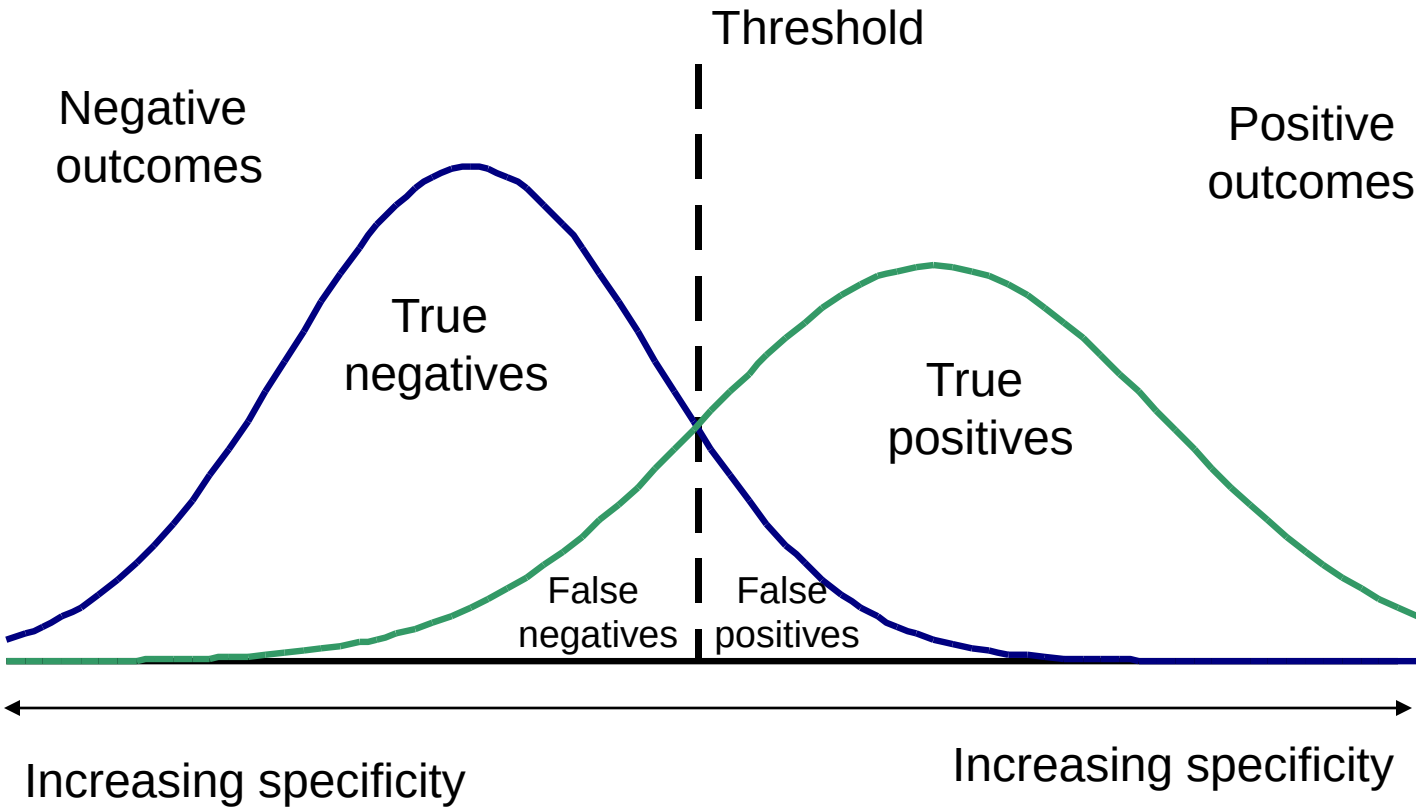
- If <disease> was much rarer in his community, say 1 in 100,000 cases, would his chances be different?

- If the test was less accurate, say 70%, would his chances be different?

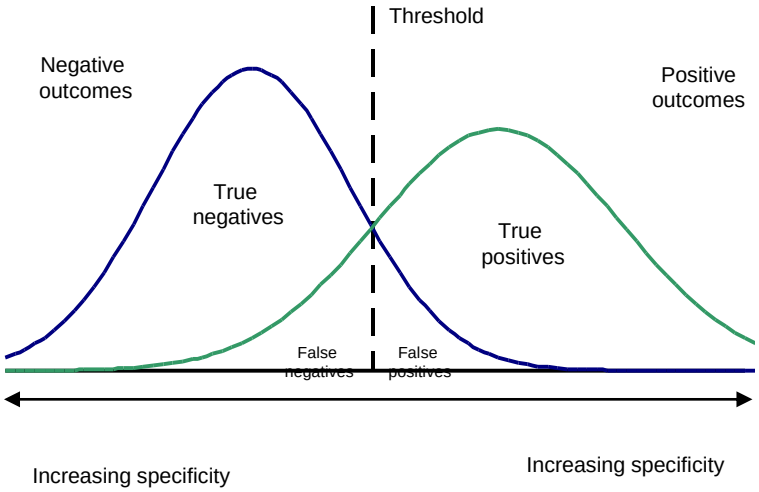
Tests are tuned to the costs of different types of errors

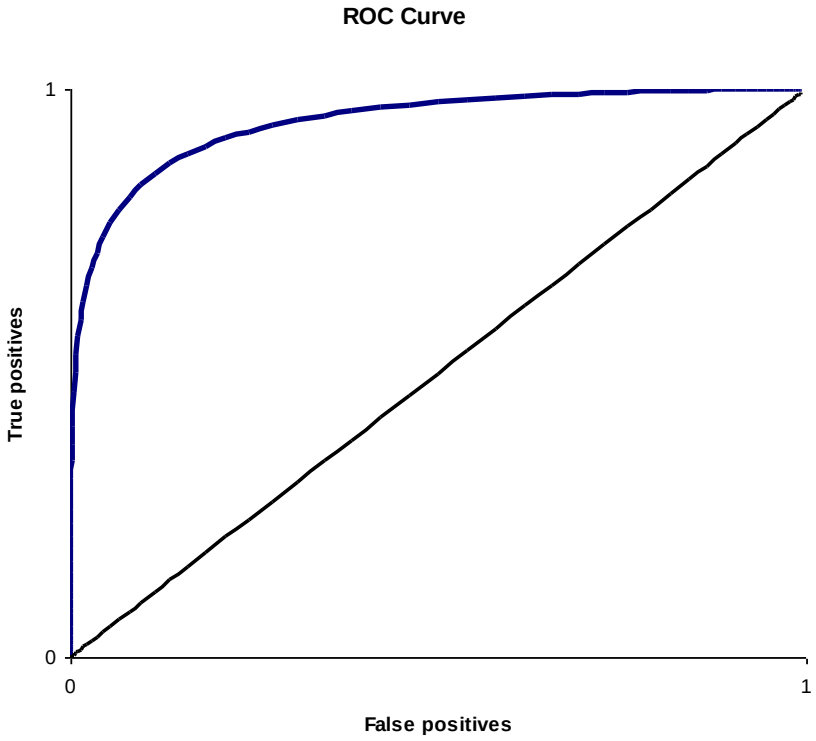
- Scenario: you're drafting legislation to ensure that insurance companies don't discriminate based on inaccurate tests
- Would you rather
  - 95% of cases with the disease denied insurance, even though a hundred times as many people who didn't have the disease were also denied insurance
  - Nearly everyone who didn't have the disease was allowed to get insurance, even through it meant that many people who did have the disease were also allowed to get insurance?
- What would your advice be if you were
  - An advocate for the insurance company?
  - An advocate for individuals being tested?

Discriminant Analysis



ROC curve
(receiver operating curve)





Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**
Patient has the disease	True positive (TP)	False negative (FN)	**Positive population (PP)**
Patient healthy	False positive (FP)	True negative (TN)	**Negative population (NP)**
Total	**Classed as positive (CP)**	**Classed as negative (CN)**	**Total population (N)**

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**	
Patient has the disease	True positive (TP)	False negative (FN)	**Positive population (PP)**	Sensitivity = TP/PP
Patient healthy	False positive (FP)	True negative (TN)	**Negative population (NP)**	Specificity = TN/NP
Total	**Classed as positive (CP)**	**Classed as negative (CN)**	**Total population (N)**	
	Positive predictive value PPV = TP/CP =Prob (disease present |pos test)	Negative predictive value NPV = FP/CN =Prob (disease absent |neg test)	Prevalence of the disease = PP/N	Accuracy =(TP+TN)/N

Case 2. Cancer

- Cervical cancer is the leading case of cancer in women, accounting for 6% of cancers.
- Current tests have a sensitivity of 80% and specificity of 70%.
- The incidence rate of new cases is 18.7 per 100,000.
  The prevalence is difficult to estimate, but is probably about 10 times as high: for this exercise, assume a prevalence of 200 per 100,000 (i.e., 0.2%).

- If a test result is diagnosed as abnormal, what is the chance that the person actually has the disease?
  a. 99.99%
  b. highly likely (much more than 50%)
  c. 50%
  d. highly unlikely (<< 50%)
  e. 0.01%

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**	
Patient has the disease				Sensitivity =
Patient healthy				Specificity =
Total				
	PPV =			Accuracy =

Diagnostic outcomes

	Test result is positive	**Test result is negative**	**Total**	
Patient has the disease	160	40	**200**	Sensitivity =160/200 = 0.8
Patient healthy	29,940	69,840	**99,800**	Specificity = 69860/99800= 0.7
Total	**30,100**	**69,900**	**100,000**	
	PPV = 160/30100 0.0053 (or 0.53%)			Accuracy = (160+69860)/100000 = 0.702 (or 70%)

D8. Description of preliminary results or pilot studies

- Describe the progress made on the substance of the project to date, including description of pilot studies, code written etc.

- If your project includes writing software, state the backup system you are using and other software engineering tools.

- Efforts at this stage of semester are often directed towards the seminar, but you also need to be making progress on the substance of the project.

In class exercise: Seminar preparation

- State
  - your name
  - title of your seminar
  - topic of research
  - take-home message

* The take home message is the one thing that you want your audience to remember