An MS Degree in CS Research

- 3 semesters in technical writing
- 2 semesters in statistics
- Experimental design
- Oral communications
- Teaching
- Web design
- Professionalism
- Grant writing

(But we only have this course!)

Getting Started

The syllabus.

The three topic areas:
- Acculturation/Professionalism: How to thrive as a researcher
- Technical presentation
- Statistics/Experimental design & analysis

In addition, we will often encounter the following theme: How do you evaluate quality?

CS5014 Course Goals

- Communicate the “culture” and “psychology” of the research enterprise
- Improve written communication skills
- Experimental Methods
  - Experimental design
  - Analysis techniques (statistics)
- Training in the “research game”
  - Proposal writing (& reviewing)
  - Paper writing (& reviewing)
  - Surviving (& and becoming successful) as a grad student
  - Surviving (& and becoming successful) as a researcher

Problem: The students in the class are very diverse!

Evaluate:

- papers
- proposals
- people
- programs

And briefly, other presentation skills including oral and web design. But we don’t have time to test oral presentation skills.
Discussion Question

How do you decide if a piece of research work is good or not?

Career Options for Research

Job titles:
- grad student
- post doc (not so common in CS)
- faculty (beware issues of tiers, expectations, fairness)
- industrial or government research job (what is “research?”)

Fact: Most of you will not end up with a career in “research” (at least not in a narrow or elitist sense).
- But you probably will use a lot of the information covered here, even after you graduate.

Assignment

Look at these web documents:
- “How to do Research At the MIT AI Lab” edited by David Chapman http://www.cs.indiana.edu/mit.research.how.to.html
- “How to have a Bad Career in Research/Academia” by David A. Patterson http://www.cs.berkeley.edu/~patterson/talks/BadCareer.pdf

Why are YOU here?

Certification
Knowledge/expertise in content areas
Other work skills
Get what you need for certain types of jobs
[Delay employment?]
Do it in a timely manner
How is graduate education different from undergraduate education?
Why do Research?

Learn a set of work skills that you can’t get from classes
• Significant writing task
• Independent/unstructured work task
• Do something “real”

Become a true expert in something

It’s the only way to get a PhD

A key concern with doing a research degree:
• When will it end?

Faculty Effects on Timely Graduation

Faculty do not want to keep you here as a perpetual slaves.

Faculty do have many conflicting demands on their time that might make them inattentive to your progress.

Faculty do not necessarily have good management skills.

Students should be proactive in fostering good management practice by their supervising faculty.
• Frequent meetings
• Make progress through a series of short-term milestones
• Clear schedules, clear responsibilities and goals

Graduate Research

Steps to a degree:
• Take a bunch of classes
• Find an advisor
• Pick a topic
• Do the work
• Write it up

WRONG! WRONG! WRONG!

This is how to delay (or kill) getting a degree.

Grad Research: How to Succeed (1)

Initially:
• Don’t just take classes – interact with faculty their students (research groups)
• Attend talks, read papers

Get PhD qualifier (or equivalent) out of the way early.

Identify an advisor early, and get involved in research
• GRA, Independent Study, or volunteer
• Research topic often comes from the work, not work from the topic.

Set deadlines or milestones... And keep them!

What do CS PhDs do? Roughly:
• 50% Industry
• 10% Government
• 40% Academia

Need to balance classes with advancing the research work.
Intersperse writing with working
- Documenting literature review is an ongoing process
- Constantly write progress reports
- Publish if possible

Learn how to write!

Picking an Advisor

How not to be successful:
- Pick advisor on one criteria: support, topic, personality.

How to succeed: Pick best compromise (for YOU!)
- Research area
- Support opportunity
- Physical environment for getting work done
- Intellectual environment for getting work done
- Peer support system (research group)
- Personality: Interaction at a personal level
- Personality: Management style (hands on vs. hands off)
- Level of attention
- Track record on timely graduation
- Professional advancement

Working with an Advisor

The most important thing is frequent interaction
- Meet often
- Document your progress (in writing)
- Have short and long-term milestones

The Role of the Advisor

- Serves as your mentor
- Serves as a source of technical assistance
- Helps you do the best work you are capable of, given constraints of time, ability, knowledge, etc.
- Helps you find resources you need (financial, equipment, psychological support)
- Helps you prepare for preliminary exam and defense
- “Keeps the Department and other hostile entities off your back.”
- Introduces you and promotes your work to important people in the field
- Gives you advice and direction on your work
- Gives you advice on career directions
- Helps you find a job

Given that all students graduating from CS@VT have at least a reasonable level of technical proficiency, the primary distinguishing determiner for success is communications ability.
What is Research Work?

- Ideally you are a member of a research group.
- Ideally, you are being supported to “do research.”
- Ideally, what you are supposed to do has a direct relationship to completing your thesis/dissertation.
- Even in an ideal world, you will be asked to contribute to the group
  - As opposed to your own work
  - System administrator, cleaning the lab, etc.
  - Such off-task work can generate useful work skills
  - Key: Balancing your needs vs. the group’s (or advisor’s) needs

The Role of Research Topic in Deciding Your Future Career

Independent Study: None.

Masters Thesis: Minimal.

PhD: As much or as little as you want.

Moral:
- The topic has to motivate you to finish your degree.
- It doesn’t need to dominate the rest of your life.

Corollary:
- Don’t obsess. Do work.

Does the Research Experience Meet Your Goals?

As a part of the degree process
- Make sure your writing improves.
- Make sure you learn how to schedule and pace.
- Make sure you get experience at independent/unstructured problem solving.
- Publish, become known.
- Become an expert in something you “own.”

CS Department Research (1)

Human-Computer Interaction/Graphics/Cyber Arts
- Bowman, Yong Cao, Ehrich, Gracanin, Harrison, McCrickard, North, Perez, Tatar, Quek

Computational Biology and Bioinformatics
- Barrett, Yang Cao, Choi, Heath, Marathe, Murali, Onufriev, Ramakrishnan, Sebutal, Shaffer, Vulikanti, Watson, Zhang

Systems, Networking, Grids, Security
- Back, Butt, Cameron, Feng, Kafura, Ribbens, Varadarajan, Tilevich, Yao

HPC/Computational Science and Engineering
- Yang Cao, Ribbens, Sandu, Shaffer, Sotelino, Watson
Software Engineering
- Arthur, Balci, Edwards, Ryder

Data/Digital Libraries
- Fox

Digital Education
- Edwards, Fox, Harrison, McCrickard, North, Perez, Ryder, Shaffer, Tatar, Tilevich

CS Faculty (1)

Sean Arthur: Software engineering (IV&V), translators
Godmar Back: Operating Systems
Osman Balci: Software engineering (IV&V), modeling and simulation
Chris Barrett: Bioinformatics
Doug Bowman: HCI, Virtual Environments
Vicky Choi: Algorithms, quantum computing
Steve Edwards: Software engineering, programming languages
Roger Ehrich: Image processing, HCI, computers in K12 education
Ed Fox: Multimedia, digital libraries, digital education
Denis Gracanin: Virtual reality, modeling and simulation
Steve Harrison: HCI, CSCW
Lenny Heath: Algorithms, graph theory, bioinformatics

CS Faculty (2)

Dennis Kafura: Software engineering (OOP), distributed systems, grids, security, health informatics
Madhav Marathe: Bioinformatics
Scott McCrickard: HCI, peripheral/mobile interfaces
T.M. Murali: Bioinformatics, gene expression and microarray processing, algorithms
Chris North: HCI, information visualization
Alexey Onufriev: Bioinformatics, protein-folding models, physics-based models, HPC
Manuel Perez: HCI, user interface software, digital education
Francis Quek: HCI, computer vision, machine learning
Naren Ramakrishnan: Recommender systems, PSEs, data mining
Cal Ribbens: HPC, grid computing, numerical methods
Barbara Ryder: Software Engineering

CS Faculty (3)

Adrian Sandu: HPC, computational science, air quality models
Joao Setubal: Bioinformatics, algorithms
Cliff Shaffer: Data structures, PSEs, computational biology, digital education
Deborah Tatar: HCI, cognition, digital education, CSCW
Eli Tilevich: Software Engineering, Web Technologies, digital education
Srinidhi Varadarajan: Parallel computation (cluster), networking
Layne Watson: Numerical analysis (optimization), PSEs, computational biology, scientific computing
Daphne Yao: Security
Liqing Zhang: Bioinformatics
Writing as an Endeavor

- Speaking and writing skills make or break any professional career (technical expertise a bonus)
- Writing has a lot of similarities to programming
- Tools of the trade (programming):
  - Compiler manuals, language manuals, programming guides
  - Compilers
  - CASE tools
  - Organizational skills, discipline
  - User interface
  - Debugging
- Programming design principles:
  - Who will use the program?
  - What does the program need to accomplish?

Writing Tools of the Trade

- Document processor/text editor
- Dictionary
- Thesaurus
- Spelling/grammar checkers
- Style and usage guides
- Technical writing guides

Writing design principles:
- Who will use the document?
- What does the document need to accomplish?

Fundamental Philosophy

The purpose of writing is to convey something TO A READER.
- Your goal is to have an (appropriate) impact
- Tone matters. Consider an email to a faculty member about a class.

All other things being equal, the better writer has a competitive advantage
- Speak to the reader
- Reduce Cognative Load

General Advice

How to learn how to write
- Write. A lot.
- Have a good writer edit your work.
- Study writing. Read manuals. Pay attention to good/bad writing in what you read.

Keep your writing simple. Aim for economy of words.

How to begin: Just type. Don’t worry about quality to start.

Syntactic consistency is vitally important: Carelessness kills

Mechanics:
- Composing at a keyboard vs. composing on paper
- Editing at a keyboard vs. editing on paper
Reducing Cognitive Load

Your job as a writer is to make yourself understood. Try to reduce cognitive load on the reader.
- Put a definition close to where it is used
- Keep notation simple and self-evident
- Take advantage of layout to emphasize important things (equations, lists, definitions, etc.)
- Remind readers of things as appropriate
- Be consistent in all respects (syntax, meaning, notation)
- Respect convention
- Use parallelism: If one thought is similar to another, the wording should reflect that
- All other things being equal, shorter is better
- Emphasize the important points in some way. There can only be so many important points!

Non-native English Speakers

Will you remain in this country (or another English-speaking country) after graduation?

After graduation, will you be publishing in English? Giving presentations in English?

Unfortunately, the only solution to becoming proficient in a language is to use it constantly.

Page Layout

How do you go about designing a page layout?

Question: Are you an expert in page layout?
- The correct answer is almost certainly “no.”

Most people don’t design/manufacture their own pharmaceuticals, aircraft, numerical software, security software, or other critical systems their lives depend on.

Leave designing page layouts to an expert.

Ideally, your software gives you a suitable page layout.
- For most professional writing, you must follow a dictated standard.

Software Support

Writing a major document (dissertation, book) can be made much easier (or possible) by appropriate automated tool support:
- Page layout support
- Typesetting support (tables, figures, mathematics)
- Rich ability to incorporate figure formats
- Tables of contents, figures, etc
- Index
- Cross referencing
- Spell checker
- Style checker

Contrast MS Word to \LaTeX\ for these tasks.

I familiar with \LaTeX, more than Word. \LaTeX\ features:
- An extensive system for defining styles for the entire document “look and feel”.
- Support for typesetting tables (a little cumbersome compared to MS Word, but more flexible)
- Supports wide set of figure formats (more cumbersome than Word)
- Extensive mathematics typesetting support (more flexible, easier to use than Word, but has a learning curve on the commands).
- Support for auto-generating tables of contents, figures, etc.
- Support for auot-generating an index (you must tag the terms)
- Extensive, “natural” support for cross-referencing, to get chapters, sections, tables, figures, etc. numbered right, and to reference them within the text
- As an ASCII format, one uses a 3rd party editor including spell-checker (often clunkier than MS Word) or style checker.

Convention reduces the level of effort for the reader.
How to Improve Your Writing

Ear: “The ability to hear whether a given word order, sentence, or term is correct.”

“The simplest way to improve your expository writing substantially is to learn to avoid a limited set of extremely common errors.”

“Fix 5 things.”

Some Writing Issues

- When editing (not composing!!) you should consciously strive to whittle away at the number of words needed to express yourself
- Don’t tell the reader how to think
- Avoid banality.
- Paragraphs should be logical units
- Break up long stretches of text
- Sentence openers should be strong (keep reader’s interest)
- Use punctuation to clarify (help reader parse)

Some English Hotspots (1)

Than and that typos

Articles (the, a):
- Do not use “the” (with plural or uncountable nouns) to talk about things in general.
- Do not use singular countable nouns without articles.

Hyphenated phrases are used to denote binding. Note that “-ly” replaces hyphenation. Otherwise, use hyphenation sparingly.
- object-oriented programming
- artificially induced intelligence
- The well-known theorem is a theorem that is well known.

Some English Hotspots (2)

Avoid using “very” and other waste words

“In order to” can usually drop “in order”

“Use” not “usage” or “utilize”.

Avoid passive voice (use “x did y” in preference to “y was done by x”).

Don’t “make decisions.” Decide!
Don’t “make commitments.” Commit!
Some English Hotspots (3)

Avoid “it is” and “there are”.

Be careful of ambiguous pronouns.

Proper names are capitalized. “In Section 3 there is a reference to the fourth section.”

Use present tense when referring to other places within the current document. “This is proved in Section 3” not “This will be proved in Section 3” or “This was proved in Section 3.”

Always use a spell checker! Multiple times during the process.

Some English Hotspots (4)

“i.e.,” and “e.g.,” ALWAYS go inside parentheses; “that is” and “for example” go outside parentheses. “etc.” should also be inside parentheses, use “and so on” outside.

- For example (e.g., you want to give an example), this is how you should do it.
- All these latin abbreviations (“i.e.,” “e.g.,” etc.) go inside. Outside you should put a, b, and so on.

“et al.” should not appear in the bibliography (except possibly for EXTREMELY long author lists).

Research Papers: Audience

Who is going to read your paper?

- Why should anyone read your paper?
- Perhaps this question should be answered explicitly, and as early as possible.

Any document is written for an audience. The audience can have a huge impact on the tone and level of the document, even after deciding the content.

- Students? Industry professionals? Researchers?
- Readers in the field? Out of the field?

Raising Audience Share

The more accessible your paper, the greater your potential audience share.

- Explain terms
- Avoid Jargon
- Use good style
- Even an expert won’t be offended by a little bit of extra explanation (within reason)
- Always pitch a little low
**Structure and Organization**

- Top-down paper design
- First step: Explicitly write down a **small** number of contributions, issues, points, etc. that the paper is “about”. This allows you to:
  - Write the abstract
  - Budget space (do this early!)
- If you have a length limit, begin by writing over the limit and then pruning down.
- Even if no length limit, keep the document as short as possible
  - Need to balance “supporting” the reader vs. redundancy
- Support the skim reader
  - Important results should have a visual tag

Some people like to write the abstract last. I think the best way is to write it first (to help you define the paper topic), then write the paper, then revise the abstract after you found out what the paper was actually about!

Papers seem to always be improved by pruning stuff down to meet a page limit.

---

**What is in a Paper?**

- Title
- Author/address list
- Date (or version)
- Abstract
- Keyword list
- Body of paper (intro, exposition, conclusions, future work)
- Figures and captions
- Statistics and other presentations of quantitative data
- Lists
- Theorems/proofs
- References
- Acknowledgments

**Title**

500 people read the title for every one that reads the paper

Goals of the title
- Help someone decide whether to read abstract, or paper (when scanning ToC or reference list)
- Give description of content
- Catch reader’s interest

Title should be specific to the paper
- Test: Could many of your papers use that same title?
- Test: Could that title refer to a paper on a different topic?

REALLY, REALLY avoid passive voice!

Ideally, the title would be an abstract of the paper

**Who is an Author?**

Most papers have multiple authors.

Who gets listed as an author? There are many types of contributions. Not all make an author.
- Gives key idea behind the work
- Discusses ideas with other contributors
- Writes the paper
- Codes, does labwork
- Develops proofs
- Collects data
- Analysis (statistics, etc)
- Edits paper
- Lends credibility

Note that most of what I say is not absolute. Most of the suggestions are desirable traits/goals that might be impractical in a specific situation.
Ordering Authors

What is the order of listing? Lots of possibilities

- Who did the most work? (measured how?)
- Who is most senior?
- Who can “sell” the paper best?
- Alphabetically or randomly?
- Sometimes lab director is last
- Sometimes students are first, or last

Life is unfair. Even if the paper authorship listing is fair, citation and public perception is not.

Warning: Decide on your professional name, and its rendition, and stick with it!
Be sensitive to your co-author’s name rendition. It’s not your right to decide.

Versioning

The key point is not to become confused about which version you are editing.

- Coordinating with co-authors
- “Passing the pen”
- SVN/CVS can help
- “Alternate” drafts to experiment with something can let you confuse yourself
- One strategy is to put a date stamp or version number into the document

The Abstract

Purpose: To summarize the contents of the paper. Explicitly, to enable the reader to decide whether to read the whole paper.

- Indicate key conclusions. Mention findings, not just say “analysis was conducted”
- (In a research paper), claim some new result

Warning: Abstract is often read separately from the paper (e.g., abstract service).

- The rest of paper might not be available
- Abstract must stand on its own
- No ability to reference anything in document (figures, citations)
- Avoid equations or heavy typesetting (might not transfer)
- Make easy to understand by non-native speakers

The Abstract (cont)

- Bad: Build abstract from sentences in intro.
- Good: Write a one paragraph mini-paper, designed for that purpose
- When to write abstract?
  - Strategy 1: Write abstract after the paper is essentially complete
  - Strategy 2: Write abstract first as part of planning (then revise at end)
- Journal paper-style abstracts are typically 200-300 words
- Avoid starting with “In this paper…”
- An “extended abstract” is a different animal. That is really a short paper (say 2-3 pages).
Keywords

The purpose of a keyword list is to support computer search of documents.
- You need to predict what potential readers of your paper will search for.
- Typically, don’t need to repeat words found in the title (they often get picked up automatically).

The Introduction (1)

The first sentence of the paper is crucial
- This is another opportunity to hook the reader... or to lose them
- Definitions are boring

Often, the initial opening sentence you come up with is so generic that it can be dropped from the paper with no loss. If that is true, then do so!!

Try to keep introduction section short

The Introduction (2)

Goals of the introduction:
- Define problem (what is this about?)
- Motivate the problem (why do I care?)
- Outline content of paper (what is plan of attack?)
- Usually summarize results (what is the solution?)

Typically give an outline of the paper at end of introduction, with a sentence for each section.

Reporting Results

- In general, you need to provide enough information that the knowledgeable reader could reproduce the work
- Enough data need to be provided so the reader can have confidence in the analysis and conclusion
- Clearly state assumptions
- Clearly state limitations
  - Tell what you didn’t do as well as what you did
  - Might be an opportunity for future work
- If appropriate, use a standard analysis or a standard testbed. That means less uncertainty to the reader, so more confidence in the result.
- When possible and appropriate, present in terms of the traditional scientific approach of hypothesis-experiment-conclusion
Displays and Captions

Displays are things like tables, figures, algorithms, etc., that are set apart from the text, and have a caption.

In general, the caption should summarize (briefly) the display.

Captions need to be self contained. Skim readers might well only read the displays and their captions, not the paper.

Within the paper text, you need to clearly refer the reader to the display at appropriate times (don’t count on them noticing the display on their own).

Conclusions

The Conclusions section should not simply be a repeat of prior material in the paper.

If it’s there, it should provide conclusions.

If there is nothing new to say, leave this section out.

This section often includes suggestions for future work (and might be titled “Conclusions and Future Work”)

Citations (1)

Citation here means the place where an item in the reference list is referred to.

Approach 1: Give a number
- Usually [1,2] or cite\(^1\),\(^2\)
- No info to the reader about the reference, but it’s easy to find in reference list
- Compact (especially the superscript form)
- Often good style to mention the author(s) when using this form, for the most significant references
- The order might be alphabetical (better) or by cite (worse)

Citations (2)

Approach 2: Give a tag, usually derived from names and dates
- Such as [Shaf98a] or maybe [CLR90]
- Potentially gives a clue to the reference content, but not much
- Not necessarily easy to find
- Reasonably compact

Goals:
- Convey info about cite to reader
- Be space efficient
- Make easy to find cite in reference list
- Make reference list well organized (similar references should be together)

Which loses the compactness!

Better or worse depends on which way the inference is going. Alphabetical is good when paper reader wants to guess the cite. Cite order is good when the reference list reader wants to guess/look up where it is cited in the paper.

First example is one name, the next is initials for 3 authors.
Citations (3)

Approach 3: Give information such as name and date
- Such as the book by Shaffer (2001a) or maybe some other good books (Cormen, et al. 1990, Preperata and Shamos 1985)
- This might tell the reader a lot about the reference.
- Takes a lot of space
- Not at all easy to find in reference list, in fact could be ambiguous
- Some references don’t have obvious authors, how to construct citation?

Occasionally, the full reference appears in the text, not in a list at the end (so the citation is the reference info)

Reference Lists

The reference list says a lot about a paper.
- What is the paper about?
- Is the author adequately familiar with prior work?
- Where can I go for more background, or to learn the field?

There are many variations on reference style
- It’s usually specified for you what style to use... so use it!
- When done manually, reference lists are one of the biggest sources of bugs
- Never do it manually. Always use something like BibTeX or Endnote. If your document processing system doesn’t give you equivalent support for formatting reference lists, switch to a real document processing system.

Reference Lists (cont)

When using reference software, put as much data into the database as possible, even if not all of it will actually appear in that citation style. You might need it later.

A good reference list style will
- Give the reader enough information about the reference to recover it
- Give the reader enough information to easily recover it/understand what it is
- Not take up more space than necessary

Examples of Reference List Styles


Names: Typically initials, sometimes last name first, sometimes last name last, sometimes mixed.

Journal title might contain abbreviations

Issue number might or might not be given, volume number is always given

The year might be in the cite identifier (if there is one), after the author, or near the end of the reference

Many like to reverse name order for first author, so that last name (ordering key) is obvious.
Examples of Reference List Styles (cont)

I believe in giving the reader more information, over saving space in the reference list (if you need more space, drop some of the less important citations).

Some information commonly given is a holdover from olden times, and is now useless for real people. Prime example: City of publication for books.

Of course, you often have no choice in style to use.

Reference List Hotspots

Views on citing URLs are in flux. Stay tuned.

“Personal Communication” is a valid citation (though a weak authority because the reader cannot recover the source).

There is nearly always some typesetting involved:

- Some part of the title is always italic
- Some items are often bold (volume number)
- Some styles use a small-caps font for the authors.

Acknowledgments

Acknowledgments can come at the beginning of the work, at the end of the work, or in a footnote in an appropriate place.

If the work was supported by a grant, **always** say that somewhere. Give a grant number (sponsor’s nomenclature) if there is one.

Also you will likely want to acknowledge those who helped proof the draft, gave ideas, or otherwise helped.

“I would like to thank...” You **are** thanking them so there’s no “would like to” about it!

Revising a Paper

“Simply go through what you have written and try to curb the length of sentences, question every passive verb and if possible make it active, prune redundant words, and look for nouns used instead of verbs.”

“Every single word that I publish I write at least six times.”

Typical process I follow:

- Get it all typed in (at terminal)
- Major reorganization into rough form, with initial scan for style/correctness (at terminal)
- First copy-edit round with significant alteration/modification/reorganization
- 1-4 more copy-edit rounds.
Tips for Revision

A person can look at something “fresh” only so many times.
- Time can help... let it sit a couple days if schedule permits
- Don’t give it to coauthor/advisor/helper prematurely. If you do, they will merely tell you what you would have discovered yourself on the next proofreading.

For research paper writeups, the writing should not all wait until the end!

Does it Make Sense?

- Does the ordering of the pieces (typically, sections) make sense?
  - What are the prerequisites for each section, in terms of what you need to know from other sections?
- Do the paragraphs follow one from another in logical progression?
  - Is each paragraph a logical subunit? Should you merge or split paragraphs?
  - Does each paragraph have enough “support” in the sense that the reader has the requisite knowledge to understand it? Requisite knowledge might be internal (where it came in the paper) or external (what a reader needs to know to read the paper).
- Does each sentence make sense? Does each sentence follow logically from the prior sentence?

Checklist

- Delete any word, phrase, sentence whose loss does not change the force or meaning
- Replace unnecessary long words with shorter (utilize \(\rightarrow\) use)
- Refactor sentences and paragraphs to put similar parts together
- Look for ambiguous phrases, unnecessary repetition, passive voice
- Are all claims supported?
- Check the math, then do it again.
- Is the notation as simple as possible?
- Did you cite enough? Too much?

When do You Publish a Paper?

There are some tradeoffs.
- More papers is better to a bean counter (least publishable unit)
- More content per paper is better for readers and society
- A paper can only be so long
- A paper has to be “ready” and “enough”

Paper lifecycle
- Possible interim (internal) reports, meetings minutes, emails
- Technical report
- Conference paper (possibly more than one)
- Journal paper
Where Should you Publish?

- There is definitely a pecking order
  - Journals have a quality hierarchy
  - Conferences have a quality hierarchy
  - Generally, journals outrank conferences
  - Other considerations: Books (and book chapters), invited papers
- Citation rates is one (semi-objective) measure of quality
- Generally, it only “counts” if it is refereed
- Your goal is to get the most “credit” possible
  - Don’t pitch too high, don’t pitch too low
  - Be aware of the audience

What affects probability of success?

- Appropriateness for the audience
- Acceptance rate (10-50% typically)

Mechanics (1)

Some venues take longer than others

- Conferences, special issues, etc. have submission date, and usually publication date
- Some journals need an outrageous amount of time to review and publish

Each venue normally has specs to meet

- Where is it sent?
- Email vs. web form submission
- Layout? These specs might be vague, or highly constrained

Mechanics (2)

Cover letter: what venue, what conditions, who is point of contact, potential changes of address, relationship to prior publication

Send email or fill out web form

Getting reviews, revising, negotiate with editor

Copy editing

- Set to house style
- Style edit (magazines vs. journals)

Proofs – they always contain errors

Resubmission

- When resubmitting a paper for re-review, include a letter.
- The letter should list in detail each of the points raised by the editor and reviewers.
- Hopefully the points are “addressed” in the rewrite. On occasion, the author might instead justify why the change should not/was not made.
- Make it clear what parts of the paper have been changed, and how.
- Be sure to make clear on the revised copy that it is the revised version.
- Depending on the situation, the editor might or might not have the revised paper re-reviewed.
Mechanics of Paper Handling (1)

- Manuscript sent by author to journal editor
- Number assigned to manuscript by publication staff
- Manuscript sent to editor-in-chief
- Manuscript assigned/sent to member of editorial board (editor)
- Editor sends paper to 2-4 reviewers, along with a proposed deadline and reviewing guidelines/forms
- Editor reminds reviewers that reviews are past due
- Editor receives reviews from reviewers

Mechanics of Paper Handling (2)

- Editor makes decision (or sends to editor-in-chief for decision)
  - Accept
  - Accept with minor revision
  - Require rewrite with re-review
  - Reject
- Six months is typical for this process (one iteration).

Paper Reviewing: The Players (1)

What is a journal editor? and what do they do?

- A person
- Typically, a member of academia who is essentially a volunteer
- Assign papers to reviewers
- Insure that the reviews get done
- Make decisions on which papers get accepted
- Might arrange for special issues, etc.

Paper Reviewing: The Players (2)

What is a reviewer?

- A person
- Typically, a member of academia, almost always a volunteer
- Could be well established, could be a student
- Its not all they do in life
- Their job is to write a review, communicate information to the editor and the authors, and do it in a timely fashion.
Typical Review Criteria (1)

1. Is the paper appropriate for the journal? (Topic, level, etc.)
2. Is the work original, and correct? (content quality) For a review paper, will it appeal to the journal audience?
3. Is the presentation clear and well organized?
4. Is the notation well conceived and consistent?
5. Does the paper appropriately cite prior work, and place itself appropriately in relation to the field?
6. Is the title appropriate?
7. Is the abstract appropriate?
8. Is the introduction appropriate?

Review Criteria (2)

If the paper fails 1 or 2, it will be rejected. Possibly the editor will suggest an alternate venue.

If the paper is weak on 2 in some way, it will probably require a complete-review of the revision (suggestions for revision will hopefully be included in information returned to author)

If the paper fails 3 badly enough it will probably be rejected or require major rewrite and re-review

If the paper fails 4 or 5, it will probably require appropriate rewrite, with good chance it will require some re-review

If the paper fails 6 through 8, it will probably need revision without re-review

How to be a Reviewer (1)

Task: Help an editor decide whether a paper is suitable (or will be suitable after revision) for publication. (Expert Witness)

Assuming the paper had no errors, would it be worthy of publication?
- Is the paper “interesting”?
- Originality (New?)
- How much contribution
- Appropriateness for this audience

Is it correct? (True?)
- You have to read thoroughly enough
- You have to know enough

How to be a Reviewer (2)

Is the presentation satisfactory?
- References appropriate
- English satisfactory, style satisfactory
- Sufficiently complete

In general:
- Can it be improved?
- Should it be required to improve?
Reviewing Mechanics

- If you can’t be on time, either return manuscript immediately or renegotiate the deadline immediately.
- Use a two-pass approach: Decide if its “reasonable” before going into details.
- You need to communicate to the editor your reasons for a decision.
- The more feedback you can provide to the authors on how to improve things, the better.
  - Note that this is very different than saying: The more obligation to improve, the better
- Your review doesn’t need to include a repeat of material in the paper (such as a summary). However, you might provide a “summary” that is a new interpretation of the paper.

Considerate Reviewing

- Don’t be unreasonable in your expectations
- If you recommend rejection, it will be for big reasons, not little reasons. Be sure not to mix reasons for rejection with picky details for possible improvement.
- Try to also comment on positive points, even when the paper is bad.
- Always be polite and not over-critical.
- Remember the material is a manuscript, and should be viewed for what it could be just as much as what it is (assuming opportunity for revision)
- The information is given to you in trust, so don’t abuse that trust.

Reviewing and Your Reputation

Reviewing takes time. Is it worth it?
- You see things earlier
- You become better tuned to your peer community
- You can influence the field
- Possible editorship or involvement in planning boards

You will establish a reputation (for good or bad) through your reviewing
- Conscientious? Reliable?
- Nasty or nice?
- A list vs. B list

PhD Qualifier

- Every department at all universities have them
  - All systems are bad, some are just less bad than others
  - Our current system was initiated in 2002
- At VT it is more of a process than an exam.
- Rules: Gain 6 points among 3 areas (3 points max for each)
  - GPA
  - Research accomplishment
  - Qualifier exam
- Rationale: Get people started on research early, less emphasis on preparing for some exam
- Risk: Not a clear termination point
- Questions and comments?
PhD Preliminary Exam

- An official VT university exam
- Structure defined by the department
- We expect a significant document
  - a proposal
  - literature review
  - “contract” for what is an acceptable dissertation
- We tend not to focus on knowledge
- Try to answer two fundamental questions:
  - Is it worth a PhD?
  - Can you do it?

Writing a Thesis (1)

What’s the big deal?
- It’s long: Psychological factors
- It’s long: Document management issues
- It’s long: Project management issues
- It’s unstructured: Decisions must be made

Structure
- Introduction: Lay out the problem
- Literature Review: How do you fit in?
- Content Sections
- Analysis (if appropriate)
- Conclusions and Future Work
- Reference List

Writing a Thesis (2)

Begin earlier rather than later
- Might do initial literature review while gearing up for doing work
- If the work naturally falls into project parts, can write parts as they are completed
- Ideally, you will already have submitted papers for publication well before completing the thesis. Depends on the scope of the work.

Key to success: Make it look like it was done right!

Research Defense

- Not an official VT exam
  - This event belongs to the student
  - The goal is to determine if the committee thinks that the student is on track to successfully complete.
  - Avoid nasty surprises at the final defense.
Thesis Defense

In our department, it's primarily a presentation of the thesis work.
- Like a (long) talk at a (small) conference
- But there are likely to be lots of questions

Some possible goals of the defense are to:
- Make sure you did it
- Make sure you understand what you did
- Make sure you understand the significance/context of what you did
- Probe your general understanding of the field of the thesis
- Make sure the committee understands what you did
- Test your ability to present

Mechanics

Generally about 2 hours long (shorter for MS, longer for PhD)

Presentation is typically preceded by a private discussion among the committee on student's record, ground rules, and problems

Presentation section is officially public, though it might or might not have been advertised

Private discussion among committee follows presentation, during which a decision is (usually) reached

Last step is communicating the decision to the student, including any requirements for modification to the thesis

How to Pick a Research Topic? (1)

Ideally, you want to have “impact.”

Dimensions:
- Easy vs. Hard
- Important vs. not important
- Obvious vs. not obvious

Where to look?
- Obvious, easy, important things have all been done
- Easy, unimportant things don’t help you (or society) much
- The easiest way to have impact is to find (relatively) easy and important things that are not obvious

Example: Dijkstra’s algorithm is only “not obvious” when it was first looked at.

How to Pick a Research Topic? (2)

What makes things obvious?
- The more some topic is clearly within one discipline, the more “obvious” (well plowed) it is.
- A lot of success can be gained by looking at the cracks between disciplines
- Problem: You need expertise in multiple disciplines

One good strategy is to get a “client.” That way you know that you are working on a “real” problem that someone cares about.
Three step process

- Name your topic
- Add an Indirect Question
- Answer So What?
- I am working on...
- Because I want to find out...
- So that the reader can understand...

Preliminary Proposal

- What is wrong with the world that you want to address with your research?
  - The Problem.
- Hypothesis: A claim about what will fix the problem
- Explanation for why this will fix the problem
  - Could have sub-hypotheses or sub-goals
- Work plan for what you will actually do
  - Probably with a time line

What Does a Researcher Do? (1)

Products:
- Publish papers (conferences, journals, “invited” papers in conferences or journals, workshops, seminars, books)
- Submit proposals (and other requests for support)
- Write reviews (public vs. private)
- Write evaluations (letters of recommendation, employee evaluation)

What Does a Researcher Do? (2)

Activities:
- Study the literature (learn), keep up with things
- Analyze: Statistics, hypotheses (experimental design)
- Serve on editorial boards, review panels, etc (service)
- Hunt for jobs
- Hunt for money (proposals)
- Collaborate: Peers, superiors, employees
- Teach
- Mentor
PostDocs

- Soft money positions
- Supply and Demand
- Biology
  - Oversupply of researchers
  - Lack of alternatives
  - Competition for hard money positions
  - Requirement for postdoc experience to be competitive
  - “Underclass” of “permanent temporary” positions

- Computer Science
  - Periods of undersupply of faculty
  - Industry absorbs most with CS degree, at all levels
  - Less culture of postdocs
  - Can be a worthwhile training experience

Getting a Faculty Position

Supply and Demand plays a big role.

- Historically, there has been a deficit of qualified applicants for faculty
- There are temporary periods of relative oversupply

Research vs. Teaching positions

A key feature of the job is diversity

- Warning: If you are not good at frequent context switches, then don’t do this!

Tenure

- An explicit, proactive point in time at which a decision is made to keep a faculty member in place or fire him/her.
- Usually done during the sixth year
- Rule of thumb: Is the replacement likely to be better or worse?
- Very different from Industry or Government
  - Firing only for cause
  - Once tenured, in theory its a permanent position
  - Still plenty of controls on performance
  - Original motivation was “academic freedom.”
  - My philosophy: Fundamental tradeoff between a high threshold of (initial and projected) performance requirement vs. security.

Some Issues of Ethics and Philosophy

Feynman Quote (see ftp://ftp.ncsu.edu/pub/eos/pub/jwilson/see-final.pdf)

- Academic dishonesty issues:
  - Unauthorized collaboration
  - Plagiarism
  - Fabrication of results

Issues: Correctness (reproducability); Allocation of credit.

Unauthorized collaboration is an issue for schoolwork. Plagiarism is an issue both for school and research work.
IRB Process

- Virginia Tech IRB infrastructure has a lot of experience
  - They know how to handle "standard" cases with little fuss
  - They have a well-defined process for doing most things
- You need IRB approval if you want to publish anything that includes data derived from individuals (a big motivator!)
- Multi-institution work can be complex to approve

Who Funds Grant Proposals? (1)

The Government:
- "Classic" technical research proposal
- Equipment or programmatic support
- Consulting or procurement contract for industry
- Charitable/community organization support

Industry:
- "Classic" technical research proposal
- Equipment donation
- Charitable/community organization sponsorship

Who Funds Grant Proposals? (2)

Foundations:
- Highly targeted research
- Programmatic support
- Charitable/community organization support

Internal Proposals:
- University, industry, government
- Technical research

How Proposals are Reviewed (1)

Who does reviewing depends on the funder
- Foundation: Board of directors
- Industry: Employees
- Some federal agencies: Agency employees or professional "handlers"
- "Classic" proposal solicitation with lots of external "peer" reviewers

Many of the same issues in place as with paper reviews
- Too big a pile, too short a time
- Obey format requirements!!!!!
- Make life easy for reviewers, don’t become a target
- They might not be technical experts
How Proposals are Reviewed (2)

Typical process:
- Reviewers categorize into bins (excellent, poor)
- Reviewer “panel” might synthesize individual reviews
- Administrator makes final decision based in large measure on bin “scores” of reviewer aggregate

Warning: Evaluation (of proposals, papers, applications, exams) is never an exact science. There’s a lot of luck and randomness in the process.
- For any binary decision from a pile of objects, there’s always three categories: Define yes, grey area, define no.

Writing the Proposal

Balancing act between your goals and the funder’s goals

Two fundamental approaches:
- Write to the solicitation
- Find a solicitation that matches your funding goal

Q: Why should anyone give you money?
A: Because you fulfill their goal better than the competition.

You must focus heavily on the funder’s goals

Strategies for Success (1)

[Taken from VT Research Division pages]

Know the funder! Contact before and during the proposal writing process can increase your chances of success by as much as 300%. Try to find out about general trends and any new areas of interest.

Preview successful applications from grant seekers with projects similar to your own. You’ll get good ideas and an understanding of the competition.

Be clear about your goals before you begin the application process. Draw up a plan that outlines your project goals for at least the next five years.

Strategies for Success (2)

Research your potential funders thoroughly and apply what you learn. “Fitting” your proposal into their program by ignoring their guidelines just won’t work.

Your original contact letter to a foundation or corporation is important. Make it as strong and to-the-point as possible.

If possible have program officials review a 3-5 page summary of your proposal. You’ll learn if you’re on the right track, and may save some time in the long run.

Read and follow the instructions. Especially with federal agencies, not following the rules exactly can ruin your chances, even if your proposal is a brilliant one.
Strategies for Success (3)

Prove the existence of the problem you propose to solve with data, statistics, case studies.
Keep it simple! A successful proposal is clear, factual, supportable, and professional.
Get straight to the point—don’t waste time or words

Don’t rush writing the proposal—take your time and get it right. Give yourself enough time to prepare the application properly and still meet the deadlines. If you don’t have time to do it right, don’t apply at all.

Strategies for Success (4)

“Marketing” is important. Look professional, involve key community figures when you can, and be sure your organization will appeal to the funder.

Management skills and experience will show the funder that you have the potential for success.

Have clear performance standards. Outline measurement indicators and determine result areas. A strong proposal proves that it is likely to achieve its goals.

Many funders like applications involving more than one organization. More likely to be convincing if the group has a standing relationship prior to writing the proposal.

Strategies for Success (5)

Know your budget! It should: be presented separately from the application, be realistic, reflect your needs, and all the figures should be correct. Keeping a record of how your costs were determined is also important.
- Don’t make yourself a liar
- Samauri Sword metaphor
- Be careful of commitments in budget re-negotiations

Keep all standard information (resumes, community statistics) in files that can be constantly updated. This saves time and allows you to concentrate on the “real” content of your proposal.

Strategies for Success (6)

Some writing hints:
- Follow the funder’s preferred format, if there is one
- Avoid jargon
- Open every section with a strong, clear sentence
- Make your proposal interesting to read
Strategies for Success

If you don’t succeed with your proposal, try to learn from the experience.

Never argue with the officials or grant reviewers. Instead, ask them where you may have gone wrong, and whether they think you should try submitting again.

Review your proposal carefully and try to identify areas that could be improved.

If it’s a good idea, improve the proposal and try again. Many (eventually) successful proposals do not get funded the first time.

Research Funding Trends

Interdisciplinary is “in”

- Probably healthy
- The best problems are real ones
- The best work is usually at the interface of different fields

Big is “in”

- Not necessarily healthy
- Easier to administer than individual grants

Targeted research is “in”

- Only as healthy as the judgment of those doing the targeting

Performance Evaluation

Examples:

- Evaluate design alternatives
- Compare two or more computers, programs, algorithms
  - Speed, memory, usability
- Determine optimum value of a parameter (tuning, optimization)
- Locate bottlenecks
- Characterize load
- Prediction of performance on future loads
- Determine number and size of components required

Examples

- Which is the best sorting algorithm?
- What factors effect data structure visualizations?
- Code-tune a program
- Which interface design is better?
- What are the best parameter values for a biological model?
Evaluation Issues

- System
  - Hardware, software, network: Need clear bounding for the "system" under study
- Technique
  - Measurement, simulation, analytical modeling
- Metrics
  - Response time, transactions per second
- Workload: The requests a user gives to the system
- Statistical techniques
- Experimental design
  - Maximize information, minimize number of experiments

A Common Fallacy

For propositions we want to believe, we ask only that the evidence not force us to believe otherwise — a rather easy standard to meet, given the equivocal nature of much information. For propositions we want to resist, however, we ask whether the evidence compels such a distasteful conclusion — a much more difficult standard to achieve. For desired conclusions, in other words, it is as if we ask ourselves “Can I believe this?”; but for unpalatable conclusions we ask “Must I believe this?” The evidence required for affirmative answers to these two questions are enormously different. By framing the question in such ways, however, we can often believe what we prefer to believe, and satisfy ourselves that we have an objective basis for doing so.

— Thomas Gilovich, *How We Know What Isn’t So*

The Error of the One-Sided Hypothesis

Consider the hypothesis “X performs better than Y”.

The danger is the following chain of reasoning:
- Could this hypothesis be true?
- I have evidence that the hypothesis might be true.
- Therefore it is true.
- What got ignored is any evidence that the hypothesis might not be (or is not) true.

A fair hypothesis (null hypothesis):
- “X and Y perform (effectively) the same.”

Statistics: The Basic Idea

Most CS research work involves samples from some “population.”
- HCI subjects
- Performance experiments on workloads for programs, systems, networks

We want to make inferences about a population by studying relatively small samples chosen from it.

The parameter value (population value) is fixed.

The statistic (value) is a random variable.

The values for the statistic come from some sampling distribution.
Statistics Questions (1)

Given a sample (chosen how?) and a statistic, what can we say about how close our estimate is to the true value?

- What is the typical difference between an estimate and the true value? (standard deviation)
- What is range that it is almost certain to be in? (confidence interval)
- What is the probability that the true value is at least \( x \)? (hypothesis test)

Statistics Questions (2)

Other issues:
- How well are two factors correlated?
- How can we predict an outcome given values for various factors? (regression modeling)
- Which factors matter? How do we optimize with respect to many factors? (experimental design)
- Are two populations the same? (ANOVA)

Variable Types (1)

Variables can be classified into types, with increasing ability to perform mathematical calculations.

Categorical variables
- Nominal values, no rank ordering possible
  - ex: Sex (male or female)

Ordinal measurement
- Rank is important, but cannot compare the values any further
  - ex: Position in race (first, second, third), we have no idea whether second is close or far from first

Variable Types (2)

Interval measurement
- Equal distance between units, but no absolute zero value so cannot take ratios
  - ex: Temperature in Fahrenheit, 60 degrees is 30 units more than 30 degrees, but its not twice as hot

Ratio Scales
- There is a fixed zero, so ratios make sense
  - ex: One salary value can be twice another salary value
Key Parameters (1)

Mean (expected value):
Population: \( \mu = E(x) = \sum_{i=1}^{n} p_i x_i \)
Sample: \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \)

Variance:
Population: \( \sigma^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \)
Sample: \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \)

\( \sigma / \mu \)

Coefficient of Variation:

Key Parameters (2)

Variance:
We divide by \( n - 1 \) because the mean variance over all samples is too low otherwise (biased vs. unbiased estimator).

Standard Deviation: Square root of variance.
* standard deviation is in units of the mean

Standard Error (of the mean): \( s / \sqrt{n} \)

Coefficient of Variation: \( \sigma / \mu \)

Mean, Median, Mode

Question: How do we characterize a population with one number?
- Indices of Central Tendency
- Favorites are Mean, Median, Mode
- The term “average” has no meaning

Cumulative Distribution Function: The probability of a variable taking a value less than or equal to \( a \):
\( F_x(a) = P(x \leq a) \)

Quantile: Denoted \( x_\alpha \), the \( x \) value at which the CDF takes a value \( \alpha \) is called \( \alpha \)-quantile.

Mean or Expected Value: \( \mu, E(x) \)

Median: The 50-percentile (or 0.5-quantile)

Mode: The most likely or most common value

Choosing Mean, Median, Mode

Each has its appropriate use (and inappropriate uses)

You can’t take the median or mean of categorical data. Use Mode.

Does the total have value? (Ex: Total CPU time for five database queries vs. number of windows open for each query.) Probably use mean for the times, median for windows.

Are the data skewed? Use Median
- GTA’s mistake
- Otherwise, can use mean or median

Class grades (at least in US) are nearly always skewed. Generally, median is the preferred value. But it is harder to compute than mean. They are the same for a normal distribution (and so mean is used in most statistics since it is easy to calculate).
“Then there is the man who drowned crossing a stream with an average depth of six inches.”

Mean, median, mode attempt to provide a single “characteristic” value for the population.
- But a single value might not be meaningful
- Some populations have similar values throughout, others vary a lot
- People generally prefer systems whose response time is more consistent

Indices of Dispersion (2)

There are different ways to measure variability

Range (max - min)
- Poor, tends to be unbounded, unstable over a range of observations, susceptible to outliers
- Easy to calculate

Variance ($\sigma^2$), standard deviation ($\sigma$)

10- and 90-precentiles

Outliers: Measurement error? Key instance?

Stem-and-Leaf Plots; Dotplots

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>259</td>
</tr>
<tr>
<td>5</td>
<td>011135556678</td>
</tr>
<tr>
<td>6</td>
<td>067789</td>
</tr>
<tr>
<td>7</td>
<td>012334555666699</td>
</tr>
<tr>
<td>8</td>
<td>000012223344456668</td>
</tr>
<tr>
<td>9</td>
<td>013</td>
</tr>
</tbody>
</table>

Note that the stem-and-leaf plot suffers from binning issues. You can’t tell from the shape of the plot where the gaps are in the data.

Histograms (1)

- Need to compute min and max of data
- Need to determine a cell size
- Lots of controversy
- Small cell size = few observations/cell, large variation in observations/cell
  - Rule of thumb: Do not permit less than 5 observations in a cell
- Large cell size loses distribution information
- Some like variable cell sizes
  - Vary bar width, not just bin values
- Use density, not frequency
- Beware binning problems
Histograms (2)

Boxplots

Scatterplots

Probability: Basic Concepts (1)

Sample Space: The set of possible outcomes.
Event: A subset of a sample space.
Random Variable: Assigns a numerical value to each outcome in a sample space. Each such value has a probability.
Permutations: \( n! \) ways to arrange \( n \) (distinct) objects.
Combinations: How many ways to choose \( k \) items from \( n \) (distinct) objects?

\[
\binom{n}{k} = \frac{n!}{(n-k)!k!}
\]
Probability: Basic Concepts (2)

Conditional Probability: A probability that is based on a part of a sample space.

Independent Events: Knowing that one event has occurred does not in any way change our estimate of the probability of the other event.

Gambler’s fallacy

Independent Variables: Knowledge of one does not affect the probabilities of the other.

\( p(x, y) = p_x(x)p_y(y) \)

Covariance (1)

Covariance: Given random variables \( x \) and \( y \) with means \( \mu_x \) and \( \mu_y \):

\[
\text{Cov}(x, y) = \sigma_{xy}^2 = \mu(x-\mu_x)(y-\mu_y) = \mu_{xy} - \mu_x\mu_y
\]

This is a measure of the degree to which two variables are related.

Covariance (2): Example 2.68

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>( p_x(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
<td>0.10</td>
<td>0.20</td>
<td>0.35</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.15</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.10</td>
<td>0.05</td>
<td>0.40</td>
</tr>
</tbody>
</table>

\[
\mu_{xy} = \sum_{x=0}^{2} \sum_{y=0}^{2} xy p(x, y)
\]

\[
\mu_x = (0)(0.35) + (1)(0.25) + (2)(0.40) = 1.05
\]

\[
\mu_y = (0)(0.35) + (1)(0.35) + (2)(0.30) = 0.95
\]

\[
\text{Cov}(X, Y) = 0.65 - (1.05)(0.95) = -0.3475.
\]

Correlation

Covariance has the disadvantage that it doesn’t let you compare the strength of relationship between two pairs of variables, because the units are in terms of multiplied units for the two means.

Correlation is a normalized value of covariance (always between -1 and 1) measured as a pure number.

\[
\text{Correlation}(X, Y) = \rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}
\]

Alternate formulation (Pearson’s \( r \)):

\[
r = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{x_i - \mu_x}{\sigma_x} \right) \left( \frac{y_i - \mu_y}{\sigma_y} \right)
\]

Monte Hall example: You pick a door, he shows you another door that loses. Should you switch to third door? Yes. You had 1/3 chance of picking right door. If you did, which door he picked is irrelevant, and you still have right door. But if you picked wrong door (2/3 probability), then third door is right one. So third door is right 2/3 of the time.

I have 3 cards: one is red on one side and green on the other, one is red on both sides, and one is green on both sides. I select one at random and show you that one side is red. What is the probability that the other side is red as well? Answer: 2/3. Why? Look at the three red sides. 2/3 of those have red on the other side. Its a conditional probability: Given I have a red side, what is the probability for the other side?

If you plot a 2D graph of the points \((X - \mu_x)(Y - \mu_y)\) where \(X\) and \(Y\) are the values of that given point, then you get positive values in the upper right and lower left quadrants, and negative values in the upper left and lower right quadrants.

The table gives percentages for the given values of \(X\) and \(Y\) (that is, 5% of the time the value of \(X\) and \(Y\) were both 0, 10% of the time, \(X\) was 0 and \(Y\) was 1). These are measures for two types of defect. The question is, how strong a relationship is there between the occurrence of the two types of defect? The answer is that there is not much relationship.

If the variables are commute time and # of dogs owned, you get really weird units (dogs*minutes) for the covariance value. So you can’t use this to compare to the strength of relationship between commute time and salary which is measured in ($ * minutes).
Correlation Example

From the example on covariance, we computed $\text{Cov}(X, Y) = -0.3475$, $\mu_X = 1.05$, $\mu_Y = 0.95$.

\[
\sigma_x^2 = \sum_{x=0}^{2} x^2 p_X(x) - \mu_x^2 \\
= 0^2(0.35) + 1^2(0.25) + 2^2(0.40) - 1.05^2 \\
= 0.7475
\]

\[
\sigma_y^2 = \sum_{y=0}^{2} y^2 p_Y(y) - \mu_y^2 \\
= 0^2(0.35) + 1^2(0.35) + 2^2(0.30) - 0.95^2 \\
= 0.6475
\]

\[
\rho_{X,Y} = \frac{-0.3475}{\sqrt{(0.7475)(0.6475)}} = -0.499
\]

This value of -0.499 indicates a modest inverse correlation.

Measurement Error (1)

Any physical measurements contain two types of errors:

- systematic error or bias [accuracy]
- random error [precision]

We can only reduce systematic error by use of external information.

We measure random error by use of the standard deviation, $\sigma$.

Example: Weigh a rock five times and get five different values.

Measurement Error (2)

We can reduce random error by repeated measurement.

- Compute the mean of the measurements.
- We can compute $\sigma$, the expected amount of error in one measurement.
- The sample mean uncertainty, or standard error, is $\sigma/\sqrt{n}$.

We see that we can reduce the (random) error on the mean by taking more measurements, but there are diminishing returns.

Bernoulli Distribution

- **Bernoulli Trial**: An event occurs with probability $p$.
  Score 1 if it succeeds, 0 if not.
  - $\mu_X = p$.
  - $\sigma_X^2 = p(1-p)$. 

I am going to describe some distributions and their properties. I picked these because they pop up often enough in various work that computer scientists do. What do you need to know? You need to know what your community expects members to know!
Binomial Distribution

- The number of successes out of \( n \) Bernoulli trials.
  - A function of parameters \( n \) (number of trials) and \( p \) (probability of success).
  - Note that we assume that the trials are independent.
- The probability of \( x \) successes:
  \[
P(X = x) = \frac{n!}{x!(n-x)!} P^x (1-P)^{n-x}
\]
- Mean and variance for a binomial distribution:
  \[
  \mu_x = np \\
  \sigma^2_x = np(1-p)
  \]
- We can draw a sample and back-fit the estimated mean and variance for the Bernoulli trials.

Poisson Distribution

When \( n \) is big and \( p \) is small, the binomial distribution is pretty much the same for all \( np \) (that is, \( np \) matters, not \( n \) and \( p \) individually in this situation). Same distribution for \((2,000,000*0.005)\) and \((1,000,000*0.01)\).

- Use the Poisson distribution.
- Define \( \lambda = np \)
- \( \mu_x = \lambda \)
- \( \sigma^2_x = \lambda \)

Exponential Distribution

Used to model waiting time if the events follow a Poisson process.

\[
f(x) = \lambda e^{-\lambda x} \\
\mu_x = 1/\lambda \\
\sigma^2_x = 1/\lambda^2
\]

Poisson Process: The number \( X \) of events that occur in time interval of length \( t \) has poisson distribution with mean \( \lambda t \).

- Memoryless Property. If we go time \( y \) with no event, the probability of the event occurring in the next time interval (of any length) is the same as it was at the beginning.

Geometric Distribution

Given a series of Bernoulli trials each with probability of success \( p \), let \( X \) represent the number of trials up to and including the first success.

- \( X \) is a discrete random variable.
- \( X \) has a geometric distribution with parameter \( p \).

\[
P(X = x) = p(1 - p)^{x-1} \\
\mu_x = 1/p \\
\sigma^2_x = \frac{1 - p}{p^2}
\]

An interesting variation is: How many tries will you need to make before the probability of (at least one) success is greater than some threshold?
Uniform Distribution

This is what a person usually means when they carelessly say “I pick something at random.”

Normal Distribution (1)

Continuous distribution.

\[ N(\mu, \sigma^2) : f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

Implications of being a normal distribution:
- 68% within \( \mu \pm \sigma \)
- 95% within \( \mu \pm 2\sigma \)
- 99.7% within \( \mu \pm 3\sigma \)

**Z-score**: A measure of position within a population for a value in standard units

\[ Z = \frac{X - \mu}{\sigma} \]

Normal Distribution (2)

We often want to know the area under the curve up to a given z-score.
- Want to know probability of an event occurring within some range of the mean.
- Examples: What is the probability that a battery lasts 42-52 hours? What is the 40th percentile for battery life?
- Calculating confidence intervals
- Measurement error is normally distributed

What Distribution Fits My Data?

- Understand the original population
- Look at histogram
- Probability plot
- Samples from normal distribution rarely contain outliers
- If the sample has outliers on one side, can try using a lognormal distribution.
**Probability Plot (1)**
(Quantile-Quantile Plot)

- Pick a distribution (eg., normal distribution)
- Sort data
- For each observation, compute its point in the 2-d plot
  - X-coordinate is its expected quantile in the distribution
  - Y-coordinate is the observed value
- If the result is a straight line, then the data follow that distribution

**Probability Plot (2)**

<table>
<thead>
<tr>
<th>i</th>
<th>$q_i$</th>
<th>$y_i$</th>
<th>$x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0625</td>
<td>-0.19</td>
<td>-1.535</td>
</tr>
<tr>
<td>2</td>
<td>0.1875</td>
<td>-0.14</td>
<td>-0.885</td>
</tr>
<tr>
<td>3</td>
<td>0.3125</td>
<td>-0.09</td>
<td>-0.487</td>
</tr>
<tr>
<td>4</td>
<td>0.4375</td>
<td>-0.04</td>
<td>-0.157</td>
</tr>
<tr>
<td>5</td>
<td>0.5625</td>
<td>0.04</td>
<td>0.157</td>
</tr>
<tr>
<td>6</td>
<td>0.6875</td>
<td>0.09</td>
<td>0.487</td>
</tr>
<tr>
<td>7</td>
<td>0.8125</td>
<td>0.14</td>
<td>0.885</td>
</tr>
<tr>
<td>8</td>
<td>0.9375</td>
<td>0.19</td>
<td>1.535</td>
</tr>
</tbody>
</table>

For normal distribution, calculate position with this approximation:

$$x_i = 4.91[q_i^{0.14} - (1 - q_i)^{0.14}]$$

**Probability Plot (3)**

This looks pretty good in that the points are all near the line. So you could say that it is reasonable to treat the sample as though it is drawn from a normal population. Perhaps you would be a bit nervous by what looks like a sinusoidal shape, indicating a potential bias. Ideally, they deviations from the line would look like random error (which itself is a normal distribution!).

**Central Limit Theorem**

For "large enough" samples, the distribution of the sample mean and sample deviation are normal.
- Even if the orginal distribution is not normal.
- This means we can compute various probabilities regarding the mean using z-table.
- The normal distribution for this sample mean is $N(\mu, \sigma^2/n)$, so we can estimate them.
- $N > 30$ is good enough (less if the underlying population is symmetrical).
Confidence Intervals (1)

Each sample mean is an estimate of the population mean
- Point estimate: A single number as the "best" estimate

It is not possible to get a perfect estimate of population mean from a finite number of samples
- How far off is it likely to be from the true value?
- The best we can get is a probabilistic bound

Confidence Intervals (2)

Goal: Get bounds \( c_1 \) and \( c_2 \) such that the population mean is in the interval \( (c_1, c_2) \) with probability \( 1 - \alpha \)
- Interval \( (c_1, c_2) \) is the confidence interval
- \( \alpha \) is the significance level
- \( 100(1 - \alpha)\% \) is the confidence level

Examples: Typically 90, 95, or 99% confidence levels, equivalently 0.1, 0.05, or 0.01 \( \alpha \) (significance) levels

Computing a Confidence Interval (1)

For 90% confidence levels, find the 5- and 95-percentiles of the sample means. How?
- Central Limit Theorem tells us that the value for the sample mean is normally distributed with \( \bar{X} = \mu \) and \( \sigma_{\bar{X}} = \sigma / \sqrt{n} \) for \( n \) the sample size.
- Look up a \textit{z-score} from a \textit{z} table for \( z = \alpha / 2 \).
- The confidence interval is then \( \bar{X} \pm z\sigma_{\bar{X}} \).
- But we don’t know \( \sigma \) (and therefore, don’t know \( \sigma_{\bar{X}} \)). So we use \( s \) for the sample.

Example Computing CI

I want to know the 95% confidence interval for a sample of 100 values with mean \( \bar{X} = 12.0 \) and standard deviation \( s = 0.1 \).
- Look up in the \textit{z} table the score that cuts off 2.5% of the area in the left tail.
- This is 0.0250 in the table. (See Table A.2, inside front cover of Navidi.) That is at row -1.9 and column 0.06 for a \textit{z-score} of -1.96
- Equivalently, look for a value of 0.9750 in the table. That is at row 1.9 and column 0.06 for a \textit{z-score} of 1.96.
- \( \sigma_{\bar{X}} = 5 / \sqrt{100} = 0.01 \).
- The 95% confidence interval is \( 12.0 \pm 1.96(0.01) = 12.0 \pm 0.0196 \), or (11.9804, 12.0196).
Philosophy

- A level 100(1 - α)% confidence interval will in the long run succeed in covering the population mean 100(1 - α)% of all the times that it is used.
- Alternatively, we are 100(1 - α)% **confident** that this interval includes the population mean.
  - Not 100(1 - α)% **probability** that the mean falls within the confidence interval, since both \( \mu \) and \( \bar{X} \) are fixed.
- What confidence interval to use?
  - Tradeoff between precision (amount of information) and probability of being wrong.
  - Depends in large part on the loss incurred if wrong
  - Sometimes 80% confidence is enough, or 99% confidence is not enough
  - 95% is a commonly used value.

How Big a Sample?

Say we want a fixed confidence and a fixed confidence interval size.

- Then we must take a sample that is big enough.
- Get the z-score for the \( \alpha \) value. (Ex: 1.96 for 95% confidence interval.)
- Get an estimate of standard deviation \( s \) (perhaps from smaller sample – say \( s = 6.83 \)).
- Decide the interval size \( i \). Ex: \( i = 0.50 \)
- Solve for \( n \) in \( zs / \sqrt{n} = i \) or \( (1.96)(6.83) / \sqrt{n} = 0.50. \)
- \( n = 716.83 \approx 717. \)

One Sided Intervals

Sometimes we only care that the true mean is at least (or at most) some value.

- Only care about one tail.
- Don’t need to divide \( \alpha \) by 2.

Now, for a 95% confidence interval, the z-score is only 1.645 instead of 1.96.

Sampling Distributions

The sampling distribution only approximates a normal curve until \( n \) gets big enough, over about 30.

Before that, the “tails” of the distribution are too big, so the estimated confidence interval is too small.

You can calculate the confidence intervals more precisely, but you need a separate table for every \( n \) value (because every curve is different).

Usually there is a table to look it up in, or better yet, the statistical software will calculate it for you.
**Student’s t Distribution**

Use for a sample of size \( n < 30 \), when the **population** has a normal distribution(!)

Degrees of freedom (DoF): In this case, \( n - 1 \).

- Given the mean, and the values of \( n - 1 \) of the items from the sample, we know the value of the \( n \)th item.

Don’t use this if there are outliers in the sample (because then the population probably is not normally distributed).

See Table A.3 (also, inside back cover).

---

**Example P. 347-8**

5 sample values: 56.3, 65.4, 58.7, 70.1, 63.9

To calculate 95% confidence interval:

- \( n = 5 = 4 \) degrees of freedom.
- \( \alpha \) value of 0.025 (column) and 4 DoF (row) yields 2.776 from Table A.3.

\[
-2.776 < \frac{\bar{X} - \mu}{s/\sqrt{n}} < 2.776
\]

\[
\frac{\bar{X} - 2.776 \sqrt{\frac{n}{s^2}}}{\frac{n}{s^2}} < \mu < \frac{\bar{X} + 2.776 \sqrt{\frac{n}{s^2}}}{\frac{n}{s^2}}
\]

- \( \bar{X} = 62.88 \), \( s = 5.4838 \), \( n = 5 \).
- Confidence interval is 62.88 ± 6.81 or (56.07, 69.69).

---

**Unpaired Observations**

This is the classic t-test.

- We have a collection of observations from \( P_1 \) and another collection of observations from \( P_2 \).
- We want to know if the two populations have the same mean.
- \( n_x \) samples from \( P_1 \) and \( n_y \) samples from \( P_2 \).

What number of degrees of freedom \( \nu \)??

\[
\nu = \left( \frac{s_x^2}{n_x - 1} + \frac{s_y^2}{n_y - 1} \right)^2 \left( \frac{(s_x^2/n_x)^2}{n_x - 1} + \frac{(s_y^2/n_y)^2}{n_y - 1} \right)
\]

---

**Unpaired Example**

- \( \bar{X} = 83.2 \), \( s_x = 5.2 \), \( n_x = 6 \)
- \( \bar{Y} = 71.3 \), \( s_y = 3.1 \), \( n_y = 10 \)

\[
\nu = \frac{(5.2^2/6 + 3.1^2/10)^2}{(5.2^2/6)^2/5 + (3.1^2/10)^2/9} = 7.18 \approx 7
\]

- From the t-table we get a value of 2.365 for 7 DoF at 95% confidence.
- Our confidence interval is therefore \( \bar{X} - \bar{Y} \pm 2.365 \sqrt{s_x^2/6 + s_y^2/10} = 11.9 \pm 5.53 \).
- Since this does not include 0, the populations are different (with 95% confidence).
Paired Observations

Are two systems different?
- The same set of workloads are run on each system
- So there are is an observation of each workload on each system
- Paired observations
- \((5.4, 19.1), (16.6, 3.5), (0.6, 3.4), (1.4, 2.5), (0.6, 3.6), (7.3, 1.7)\)
- Sample mean: -0.32
- Sample variance: 81.62
- Sample standard deviation: 9.03
- Confidence interval: \(-0.32 \pm t_{(3.69)} = -0.32 \pm t(3.69)\)
- For 5 DOF, at \(\alpha = 0.05\), the \(t\) value is 2.015
- Therefore, the 90% confidence interval is
  \[-0.32 \pm (2.015)(3.69) = (-7.75, 7.11)\].

Hypothesis Testing (1)

One fundamental experimental question: What is the parameter value?
- Use a confidence interval to express our level of uncertainty

Another fundamental experimental question: Are two populations different?
- Does a subpopulation perform better?
- Is one algorithm better than another?
- A confidence interval doesn’t express the level of certainty that one value is different from another.

Null Hypothesis \((H_0)\): Two populations have same mean
Alternate Hypothesis \((H_1)\): Two pops have different means

Hypothesis Testing (2)

Rejection of the Null Hypothesis: Do so if you are “confident” that the means are different
- A double negative.

Significant Difference: Two means are different with a certain reference confidence (probability).

We assume the null hypothesis is true until proven otherwise
- A \(P\)-value measures the plausibility of \(H_0\).

Doing the Test (1)

Example: Take 50 samples, \(\bar{X} = 92, s = 21\). Is this significantly different from 100?
- What is the probability that this sample comes from a population with mean 100?

The null distribution is a normal distribution using the null hypothesis mean and the standard error \(21 / \sqrt{50} = 2.97\), or \(\bar{X} \sim N(100, 2.97^2)\).

Compute the z-score
\[ z = \frac{92 - 100}{2.97} = -2.69.\]
Doing the Test (2)

Row -2.6, column .09 yields a value of 0.0036. That is, the probability of this sample arising by chance is 0.36%.

We can reject the null hypothesis with over 99% confidence.

- This is statistically significant at the 95% and 99% confidence levels.
- This is not statistically significant at the 99.9% confidence level.

Philosophy and Hypothesis Testing

The null hypothesis either is true or else it is not true.

- This reality is probably external to our knowledge.
- But it is fixed.

Therefore, we can’t talk about the probability of the null hypothesis being true.

- We can talk about the probability of a sample giving such a low mean.

Statistical Significance

The term statistical significance has a particular meaning

- Be very careful of how you use the term “significant” in any technical writing!
- Being statistically significant does not necessarily mean the difference is important. It only means that we are reasonably certain that the difference exists.

Small-Sample Tests

If the sample is small, and we do not know $\sigma$, use $s$ and the $t$-distribution.

If we do know $\sigma$, then we can use $z$-scores (because then $X$ is a random variable with normal distribution).
Are 2 Populations Different? (1)

Given sufficiently large samples, we want to know if the difference between the means of two samples is significantly different.

Example

\[ \bar{X} = 0.37, s_X = 0.25, n_x = 544. \]
\[ \bar{Y} = 0.40, s_Y = 0.26, n_y = 581. \]

Calculate the standard deviation

\[ \sqrt{0.25^2/544 + 0.26^2/581} = 0.01521. \]

\[ \bar{X} - \bar{Y} \sim N(0, 0.01521^2) \]

\[ z = \frac{0.03}{0.01521} = -1.97 \]

Are 2 Populations Different? (2)

From the z-table at row -1.9 and column 0.07, we get a P-value of 0.0244 for each tail, or an aggregate probability of being too far off of 4.88%.

Is this significant?

For small samples, use the same idea with a t-table.

Only valid if the populations are normal.

If in doubt, can use a nonparametric test.

Chi-squares

We will determine significance using the following formula:

\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_o} \approx 14.41 \]

across the table cells where \( f_o \) is the observed frequency of the cell and \( f_e \) is the expected frequency of the cell.

Some observations:

- Relationships have strength, and significance (confidence)
- Order of rows and columns should be irrelevant
- Strength should not depend on absolute values. Doubling the number of observations in all cells should not affect strength.
Testing Significance

We scale by degrees of freedom, calculated as \((R - 1)(C - 1)\) for \(R\) the number of rows and \(C\) the number of columns.

Now we can look it up in the chi-square table.
- We find for 6 degrees of freedom at 95% confidence a value of 12.592. This is significant at 95% confidence.
- At 99% confidence the value is 16.812.
- If we doubled the number of observations at every cell, the value for \(\chi^2\) would double, and now it would be significant at 99% confidence.
- Warning: The “strength” has not increased!

Hypothesis Test Errors

There is no “right” \(\alpha\) value. When do we switch from plausible to inplausible?
- But sometimes we have to make a decision. Reject a shipment? Recalibrate an assembly line?
- When a decision is made based on a cutoff \(\alpha\) value, it is called a fixed-level test.
- If \(P \leq \alpha\), we reject the null hypothesis.

There are two ways that a hypothesis test can be wrong.
- **Type I error**: Reject \(H_0\) when it is true.
- **Type II error**: Accept \(H_0\) when it is false.
- The probability of a type I error is never greater than \(\alpha\).
- The smaller we make the probability of a type I error, the greater the probability of a type II error.

Correlation and Confounding

Correlation measures the **strength** of the relationship between two variable.
- It is a pure number, so can be compared to correlations between other variables

If variables A and B are correlated, and variable A is also correlated to C, then B is correlated to C.
- Thus, correlation doesn’t imply causal relation
- This phenomenon is called **confounding**
- Good experimental design can minimize confounding

Regression Models (1)

A regression model predicts a random variable as a function of other variables.
- Response variable
- Predictor variables, predictors, factors
- All must be quantitative variables to do calculations

What is a good model?
- Minimize the distance between predicted value and observed values

Try #1: Use the mean value of the predictor variable as the response value.
- High error
Regression Models (2)

Try #2: Linear model: \( \hat{y} = b_0 + b_1 x \) where \( y \) is the response variable, \( x \) is the predictor variable, and \( b_0, b_1 \) are regression parameters (determined by the data)

Error: \( e_i = y_i - \hat{y} \) for the \( i \)th observation.

Least-squares: Minimize the sum of squared errors, subject to the constraint that the mean error is zero.

Parameter Values

To get the least squares values for \( b_0, b_1 \), use:

\[
\begin{align*}
  b_1 &= \frac{\sum xy - n \bar{x} \bar{y}}{\sum x^2 - n \bar{x}^2} \\
  b_0 &= \bar{y} - b_1 \bar{x}
\end{align*}
\]

- Disk I/Os and processor times for 7 program runs: \((14, 2), (16, 5), (27, 7), (42, 9), (39, 10), (50, 13), (83, 20)\)
- From the data, plug into the formulas to get \( b_1 = 0.2438, b_0 = -0.0083 \).
- So, CPU time is predicted as \(-0.0083 + 0.2438(# \text{ of disk I/Os})\)

Allocation of Variation (1)

The purpose of a model is to predict the response with minimum variability.

- No parameters: Use mean of response variable
- One parameter linear regression model: How good?

Allocation of Variation (2)

A measure of error without regression would be squares of differences from the mean

- Also called the total sum of squares (SST)
- Is a measure of \( y \)'s variability: variation
- \( \text{SST} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n \bar{y}^2 = \text{SSY} - \text{SS0} \)
- SSE with regression = \( \sum (y_i - \hat{y}_i)^2 \).
- The difference between SST and SSE (with regression) is the sum of squares explained by the regression, or SSR.
- The fraction of variation explained is called the coefficient of determination, \( R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\text{SST} - \text{SSE}}{\text{SST}} \).
- Note this is the square of the coefficient of correlation

So SST is the “worst” error, error without regression.

SSY: Sum of squares of the \( y \) values: \( y = \sum y_i^2 \).
SS0: Sum of squares of \( \bar{y} = \sum y_i \).
Allocation of Variation Example

From previous example, \( R^2 = \frac{SSR}{SST} = 0.9715 \). Thus, 97% of the variation in CPU requirements is explained by Disk I/O requirements.

Question: What is cause and what is effect?

Confidence Intervals (1)

We can (and should) compute confidence intervals for \( b_0 \) and \( b_1 \) in the normal way.

- For the example, \( n = 7 \). Note that the degrees of freedom will be \( n - 2 \) since the values are partly fixed given that we have set values to \( b_0 \) and \( b_1 \).
- Using the table, a \( t \)-variate at 0.95-quantile and 5 degrees of freedom is 2.015.
- Calculating standard error for the sample, we get a confidence interval for \( b_0 \) of (-1.6830, 1.6663). We cannot reject the null hypothesis in the \( y \)-intercept.
- Calculating standard error for the sample, we get a confidence interval for \( b_1 \) of (0.2061, 0.2814). The slope seems to be really there.

Confidence Intervals (2)

- When calculating confidence intervals, the standard deviations have this in the denominator:
  \[
  \sum (x_i - \bar{x})^2. 
  \]
- When you design the experiment, you might be able to pick the \( x_i \) values.
- So you would like bigger range of \( x_i \) to increase the sum, and thus the denominator (to decrease the standard deviation and thus the confidence interval)

Assumptions (1)

Assumption: Linear relationship between \( x \) and \( y \)

If you plot your two variables, you might find many possible results:

- No visible relationship
- Linear trend line
- Two or more lines (piecewise)
- Outlier(s)
- Non-linear trend curve

Only the second one can use a linear model

The third one could use multiple linear models
Assumptions (2)

Assumption: Independent Errors
- Make a scatter plot of errors versus response value
- If you see a trend, you have a problem

Limits to Simple Linear Regression

Only one predictor is used.
The predictor variable must be quantitative (not categorical).
The relationship between response and predictor must be linear.
The errors must be normally distributed.

Visual tests:
- Plot x vs. y: Anything odd?
- Plot residuals: Anything odd?
- Residual plot especially helpful if there is a meaningful order to the measurements

Multiple Linear Regression

For predictor variables $x_1, x_2, \ldots, x_k$, use the linear model

$$ y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e. $$

The $b_i$s are fixed parameters, $e$ is the error term.

In matrix terms, $y = Xb + e$

Of key interest is finding values for the $b_i$s. This is (usually) an over-constrained set of equations.

Next questions are:
- Is the model significant?
- Are the parameter values significant?

Multiple Regression Example (1)

<table>
<thead>
<tr>
<th>CPU time</th>
<th>Disk I/Os</th>
<th>Memory Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
<td>144</td>
</tr>
<tr>
<td>9</td>
<td>42</td>
<td>190</td>
</tr>
<tr>
<td>10</td>
<td>39</td>
<td>210</td>
</tr>
<tr>
<td>13</td>
<td>50</td>
<td>235</td>
</tr>
<tr>
<td>20</td>
<td>83</td>
<td>400</td>
</tr>
</tbody>
</table>

The $b$ values are found to be such that we get:

CPU time = $-0.1614 + 0.1182(\text{# of disk I/O}) + 0.0265(\text{memory size})$
Multiple Regression Example (2)

- We can compute error terms from this (difference between regression formula and actual observations)
- We can compute the coefficient of determination SSR/SST to find that the regression explains 97% of the variation of CPU time.
- 80% confidence intervals for the parameters are (-2.11, 1.79), (-0.29, 0.53), and (-0.06, 0.11), respectively.
  - What does this mean?
- Yet, model seems to give good predictive ability.
  - Example: What is the predicted CPU time for 100 disk I/Os and a memory size of 550? It is 26.23 with confidence interval (19.55, 32.93) at 90% confidence. What does this mean?

Analysis of Variance (1)

- ANOVA lets us determine if explanation of variance given by the model is “good” or not.
- Specifically, we want to know if the following null hypothesis is correct:
  - $y$ does not depend upon any predictor $x_j$, that is, all of the $b_j$s are indistinguishable (with confidence) from zero.
- To do this, we want to compare SSR (sum of squares explained by regression) to SSE (sum of squared errors) and see if the ratio is “good enough”.
- An important part of calculating the ratio is the number of degrees of freedom for each term.
  - SSR has $k$ degrees of freedom.
  - SSE has $n - k - 1$ degrees of freedom.

ANOVA (2)

- Thus, the actual ratio calculated is $(SSR/k)/(SSE/(n - k - 1))$.
- The other thing we need is to know what is “good enough” on the ratio.
  - This depends on the amount of information we have.
  - Use the F table appropriate for the desired confidence.
  - Look at position $F[k, n - k - 1]$ in the table.
- For the Disk I/O example, the computed ratio is 75.40, and the F table gives 4.32 as the minimum acceptable ratio.
- Thus, we have very high confidence that the regression model has predictive ability.

Multicollinearity (1)

Dilemma: None of our parameters are significant, yet the model is!!

The problem is that the two predictors (memory and disk I/O) are correlated
- $R = .9947$

Next we test if the two parameters each give significant regression on their own.
- We already did this for the Disk I/O regression model, and found that it alone accounted for about 97% of the variance.
- We get the same result for memory size.

No parameter is significant. Bad news!

This issue of no parameter being significant, while the model is significant, will be explained in the following slides.
Multicollinearity (2)

Conclusion: Each predictor alone gives as much predictive power as the two together!

Moral: Adding more predictors is not necessarily better in terms of predictive ability (aside from cost considerations).

ANOVA for Categorical Variables

A.k.a: One-factor experiments

A common problem is to determine if groups are different.

- Do plumbers make more than electricians?
- Is system A, B, or C better on a performance metric?

Now the question becomes: Are the between-group variances sums of squares (BSS) more or less important than the within-group variances sums of squares (WSS)?

Again, DOF is important.

- BSS has \( k - 1 \) DOF
- WSS has \( n - k \) DOF

Calculate, and compare the ratio to the \( F \) table to determine if the differences are significant.

Curvilinear Regression (1)

A model is a model. You can do anything you want, then measure the errors.

What is natural?

Do a scatterplot of response vs. predictor to see if its linear.

Often you can convert to a linear model and use the standard linear regression.

- Take the log when the curve looks like \( y = bx^a \)
Curvilinear Regression (2)

Example: Amdahls law says I/O rate is proportional to the processor speed.

- $I/O \ rate = \alpha \ (CPU \ rate)^b$
- Take logs: $\log(I/O\ rate) = \log \alpha + b \log(MIPS\ rate)$
- Now can use standard linear regression

Common Mistakes with Regression

- Verify that the relationship is linear. Look at a scatter diagram.
- Don’t worry about absolute value of parameters. They are totally dependent on an arbitrary decision regarding what dimensions to use.
- Always specify confidence intervals for parameters and coefficient of determination.
- Test for correlation between predictor variables, and eliminate redundancy. Test to see what subset of the possible predictors is “best” depending on cost vs. performance.
- Don’t make predictions too far out of the measured range.

More Observations

Correlation measures strength of a relationship

We’d prefer correlation if we could use it

- Pearson’s $r$ is affected by order of the columns/rows

ANOVA can be used when the columns are ordinal or better.

- Can we do something to turn the Crosstabs example into something quantitative?

If both the rows and columns are ordinal, then we can measure correlation.

- Does increased education lead to higher salary?

Experimental Design

- “The goal of a proper experimental design is to obtain the maximum information with the minimum number of experiments.”
- Determine the affects of various factors on performance.
- Does a factor have significant effect?
- An experimental design consists of specifying
  - the number of experiments
  - the factor/level combinations of each experiment
  - the number of replications of each experiment
- Response Variable: The outcome of the experiment (performance measure).
Potential Pitfalls

All measured values are random variables. Must account for experimental error.

Must control for important parameters. Example: User experience in a user interface study.

Must be able to allocate variation to the different factors.

There is a limit to the number of experiments you can perform. Some designs give more information per experiment.

Types of Experimental Designs (1)

Simple designs
- One factor is varied at a time.
- Inefficient
- Does not capture interactions
- Not recommended

Full factorial design
- Perform experiment(s) for every combination of factors at every level.
- Many experiments required
- Captures interactions
- If too expensive, can reduce number of levels (e.g., $2^k$ design), number of factors, or take another approach.

Types of Experimental Designs (2)

Fractional factorial design
- Only run experiments for a carefully selected subset of the full factorial combinations

<table>
<thead>
<tr>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

$2^k$ Factorial Designs

A $2^k$ experimental design is used to determine the effect of $k$ factors, each of which have two alternatives or levels.
- Used to prune out the less important factors for further study
- Pick the highest and lowest levels for each factor
- Assumes that a factor's effect is unidirectional

Unidirectionality might not be true!
Example (1)

Start with a $2^2$ experimental design.

<table>
<thead>
<tr>
<th>Cache</th>
<th>Memory</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>75</td>
</tr>
</tbody>
</table>

Define two variables $x_A$ and $x_B$ as $x_A = -1$ for 4MB memory, and 1 for 16MB memory; $x_B = -1$ for 1KB cache, and 1 for 2KB cache.

Regression equation:

$$y = q_0 + q_Ax_A + q_Bx_B + q_{AB}x_Ax_B$$

Example (2)

For this example, we get 4 equations and 4 unknowns:

15 = $q_0 - q_A - q_B + q_{AB}$
45 = $q_0 + q_A - q_B - q_{AB}$
25 = $q_0 - q_A + q_B - q_{AB}$
75 = $q_0 + q_A + q_B + q_{AB}$

Solve to yield:

$$y = 40 + 20x_A + 10x_B + 5x_Ax_B.$$ 

Interpretation: Mean performance is 40 MIPS, the effect of memory is 20 MIPS, the effect of cache is 10 MIPS, and the interaction effect is 5 MIPS.

Example (3)

Sign table:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>160</td>
<td>80</td>
<td>40</td>
<td>20</td>
<td>5</td>
</tr>
</tbody>
</table>

Why pick -1 and 1?

- It makes the computation easy.
- Don’t need to “solve” equations, just sum!

Allocation of Variation (1)

The importance of a factor is measured by the proportion of the total variation in the response that is explained by the factor.

Sum of Squares Total (SST) or variation:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2.$$ 

$$SST = SSA + SSB + SSAB$$

$$= 2^2q_A^2 + 2^2q_B^2 + 2^2q_{AB}^2.$$ 

$$2100 = 1600 + 400 + 100.$$
The fraction of variation explained by A is $SS_A/SST$.

If one factor explains the vast majority of the variation, then the other can be ignored.

- In this example, memory explains 76% of variation, cache explains 19%.

And their interaction explains 5%.

We can generalize this to $k$ factors.

Example: memory size, cache size, number of processors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level -1</th>
<th>Level 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory (A)</td>
<td>4MB</td>
<td>16MB</td>
</tr>
<tr>
<td>Cache (B)</td>
<td>1KB</td>
<td>2KB</td>
</tr>
<tr>
<td>Processors (C)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cache</th>
<th>4MB</th>
<th>16MB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One P</td>
<td>Two P</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

Sign table:

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>46</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>58</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>86</td>
</tr>
</tbody>
</table>

Total: 320 80 40 160 40 16 24 9

$SST = 2^3(q^2_A + q^2_B + q^2_C + q^2_{AB} + q^2_{AC} + q^2_{BC} + q^2_{ABC})$

$= 8(10^2 + 5^2 + 20^2 + 5^2 + 2^2 + 3^2 + 1^2)$

$= 800 + 200 + 3200 + 200 + 32 + 72 + 8 = 4512.$

- Memory explains 800/4512 = 18%
- Cache explains 200/4512 = 4%
- # of processors explains 3200/4512 = 71%
2^k r Factorial Designs

Experiments are observations of random variables.

With only one observation, can’t estimate error.

If we repeat each experiment r times, we get 2^k r observations.

With a 2 factor model, we now have:

\[ y = q_0 + q_a x_a + q_b x_b + q_{ab} x_a x_b + \epsilon. \]

\[ \begin{array}{|c|c|c|c|c|} \hline & A & B & AB & y & \text{errors} \\ \hline 1 & -1 & -1 & 1 & (15, 18, 12) & 15 \pm 0.3 \\ 1 & 1 & -1 & -1 & (45, 48, 51) & 48 \pm 3.0 \\ 1 & -1 & 1 & -1 & (25, 28, 19) & 24 \pm 1.4 \\ 1 & 1 & 1 & 1 & (75, 75, 81) & 77 \pm 2.2 \\ \hline \end{array} \]

The SSE line is what is new here.

Factor A explains 79%, B explains 15.4%, AB explains 4%, and error explains about 1.5%.

We can also compute confidence intervals. The degrees of freedom are 2^k (r – 1).

- Factor A has 90% confidence interval 21.5 ± 1.92. So it is significant.

A is 5547/7032. B is 1083/7032.
2^{k-p} Fractional Factorial Designs (1)

Even with 2^k designs, the number of experiments can get out of hand for several factors.

We can get a lot of information with fewer experiments.
- If we pick the experiments carefully, we can get enough information to compute what we need most. (p is “suitably chosen”)
- Do the following set of 8 experiments on 7 factors.

<table>
<thead>
<tr>
<th>#</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

2^{k-p} Fractional Factorial Designs (2)

Selecting the Experiments

The selection of levels is important.
- The vectors need to be orthogonal so that the contributions of the effects can be determined.
- Each column sums to zero
- The sum of products of any two columns is zero
- There are 2^{k-p} rows

Sum up the column values times their coefficients and divide by 2^{k-p}.

Example (1)

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>317</td>
<td>101</td>
<td>35</td>
<td>109</td>
<td>43</td>
<td>1</td>
<td>47</td>
<td>3</td>
<td>Tot</td>
</tr>
<tr>
<td></td>
<td>39.6</td>
<td>12.6</td>
<td>4.4</td>
<td>13.6</td>
<td>5.4</td>
<td>0.1</td>
<td>5.9</td>
<td>0.4</td>
<td>Tot/8</td>
</tr>
</tbody>
</table>

Calculated variation by variable: Prop. to square of effect.
SST (A through G) = 427.22; (12.6 x 12.6)/427.22 = 37.2% 
A = 37%, B = 5%, C = 43%, D = 7%, E = 0%, F = 8%, G = 0%
Further experiments should be conducted only on A and C.

SST = 12.6^2 + 4.4^2 + 13.6^2 + 5.4^2 + 0.1^2 + 5.9^2 + 0.4^2 = 427.22
Example (2)

Big assumption: These experiments only provide so much information.
- The effects due to interaction are summed into the values calculated for the separate variables
- Example: $F = BC$
- The experiments are masking these "confounding" interactions
- If the interaction effects are small, it's OK

Half-replicate example (1)

5 variables (named A-E).
Want to do a half-replicate design ($2^4$ experiments).

Step 1: Set up sign table for first 4 variables, plus 4-way interaction (which is an unlikely affector).

Half-replicate example (2)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>ABCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>a</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ab</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>c</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ac</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>bc</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>d</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>ad</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>bd</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>abd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>cd</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>acd</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>bcd</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>abcd</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Half-replicate example (3)

Step 2: Use these signs to set values for E.
Also add E to the treatment name for each row with +.

When we run our 16 experiments:
- We set E high on experiments for rows where it is +,
- We set E low on experiments for rows where it is -.

Note a few selected (derived) interactions.
### Half-replicate example (4)

<table>
<thead>
<tr>
<th>T'ment</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E=ABCD</th>
<th>AB</th>
<th>CDE</th>
<th>ACDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>abe</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>c</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ace</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>bc</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>d</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ade</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>bde</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>abd</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>cd</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>acd</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>bcd</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>abcd</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Half-replicate example (5)

Aliases:
- A and BCDE have the same sign column.
- They are aliases
- So this column actually measures their sum
- This is OK if A is big and BCDE is small
- AB and CDE is another example.

### One-Factor Experiments (1)

Assume we want to investigate the effects of a single factor, with multiple levels and \( r \) observations at each such level.

\[
y_{ij} = \mu + \alpha_j + \varepsilon_{ij}
\]

where \( \mu \) is mean response, \( \alpha_j \) is the effect of alternative \( j \) and \( \varepsilon_{ij} \) is the error term.

We choose to scale things such that the \( \alpha_j \)'s sum to zero.

Compare the effects of 3 processors, observed 5 times each.

### One-Factor Experiments (2)

<table>
<thead>
<tr>
<th></th>
<th>144</th>
<th>101</th>
<th>130</th>
<th>120</th>
<th>144</th>
<th>180</th>
<th>176</th>
<th>211</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col Sum</td>
<td>872</td>
<td>816</td>
<td>1127</td>
<td>2815</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col Mean</td>
<td>174.4</td>
<td>163.2</td>
<td>225.4</td>
<td>187.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect</td>
<td>-13.3</td>
<td>-24.5</td>
<td>37.7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subtract 187.7 from each column mean.

Example 20.1, p329 from Jain. Compare the effects of 3 processors, observed 5 times each.
Errors and Variation (1)

For each observation, the error is the difference between the observation and the sum of the mean and alternative effect.

- Mean is 187.7, R's effect is -13.3, so the first error term is $144 - (187.7 - 13.3) = -30.4$.

If we sum the squares of all the error terms, we get 94,365.20.

If we sum the squares of all the responses, we get 105,357.3.

As usual, the total variation from the mean can be allocated to the effect of the factor and the effect of the error.

$$SSY = SS_0 + SSA + SSE$$

Variation is divided into explained and unexplained error:

$$SST = SSY - SS_0 = SSA + SSE$$

“Error” in this sense simply means the difference between the (one variable) prediction and the observation. It might be measurement error, but it might also be other factors that we are not considering.

Errors and Variation (2)

The error accounts for $94,365.20/105,357.3 = 89.6\%$ of the variation.

- The factor accounts for 10.4\% of the variation.
- Is this significant?

ANOVA (1)

We use ANOVA to determine if there is a meaningful difference due to this factor.

- Recall that we have to scale the variations by the degrees of freedom.

$$SSY = SS_0 + SSA + SSE$$

$$ar = 1 + (a - 1) + a(r - 1)$$

- We have $ar$ independent terms (for SSY)
- One mean (for SS0)
- $a$ experiments but only $a - 1$ independent ones since they add to zero
- $ar$ error terms, but since the $r$ errors for a given experiment add to zero, each of $a$ experiments has only $r - 1$ degrees of freedom.

ANOVA (2)

Calculate the ratio of SSA/DOF(a) and SSE/DOF(e)

- $(10,992.13/2)/(94,365.2/12) = 0.7$
- From the F table position [2, 12] we require a value of 2.81

Conclusion: This factor does not give a significant difference.
Two-Factor Full Factorial Design (1)

<table>
<thead>
<tr>
<th>Work</th>
<th>1</th>
<th>2</th>
<th>No</th>
<th>Sum</th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>54</td>
<td>55</td>
<td>106</td>
<td>215</td>
<td>71.7</td>
<td>-0.5</td>
</tr>
<tr>
<td>TECO</td>
<td>60</td>
<td>60</td>
<td>123</td>
<td>243</td>
<td>81.0</td>
<td>8.8</td>
</tr>
<tr>
<td>SIEVE</td>
<td>43</td>
<td>43</td>
<td>120</td>
<td>206</td>
<td>68.7</td>
<td>-3.5</td>
</tr>
<tr>
<td>DHRY</td>
<td>49</td>
<td>52</td>
<td>111</td>
<td>212</td>
<td>70.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>SORT</td>
<td>49</td>
<td>50</td>
<td>108</td>
<td>207</td>
<td>69.0</td>
<td>-3.2</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>255</td>
<td>260</td>
<td>568</td>
<td>1083</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>51.0</td>
<td>52.0</td>
<td>113.6</td>
<td>72.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effect</strong></td>
<td>-21.2</td>
<td>-20.2</td>
<td>41.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If there are any doubts, then:

- Check that the errors are normal (use quantile-quantile plot)
- Check confidence intervals on the parameters, especially mean and on effects of the supposed significant factors.

In this case, the quantile-quantile plot looks reasonable.

Two-Factor Full Factorial Design (2)

Variances:
- SSA is the sum of squares of column effects = 12,857.2
- SSB is the sum of squares of row effects = 308.40
- SSE can be computed from the table of errors, or as SST – SSA – SSB, = 236.80

Caches (A) explain 95.9% of variation, workloads (B) explain 2.3%, and errors explain 1.8%.

ANOVA

DOF
- For caches, 3 – 1 = 2, so 12.857.2/2 = 6428.6
- For workloads, 5 – 1 = 4 so 308.4/4 = 77.10
- For errors, 2 × 4 = 8 so 236.80/8 = 29.60

F-Table
- 6428.6/29.60 = 217.2, higher than the required 3.1
- 77.10/29.60 = 2.6, lower than the required 2.8
- This is good, workloads are not meant to be a confounding influence.

Precautions (1)

If there are any doubts, then:
- Check that the errors are normal (use quantile-quantile plot)
- Check confidence intervals on the parameters, especially mean and on effects of the supposed significant factors.

This is good, workloads are not meant to be a confounding influence.

Assumed to be categorical variables – otherwise we would have used regression.
Precautions (2)

In this case, we get confidence intervals of:
- Grand mean: 72.2 ± 2.6
- Cache means: All effects do not include zero in the confidence interval
- Workloads: Only TECO does not include zero in the confidence interval, indicating that they are indistinguishable
- This is good: Workload was meant not to confound the cache study

Technical Graphics (1)

See three books by Edward R. Tufte:
- *The Visual Display of Quantitative Information*
- *Envisioning Information*
- *Visual Explanations*

Technical Graphics (2)

Graphical displays of quantitative data should have Power, Truth, Grace
- Show the data
- Induce viewer to think about substance
  - (not the methodology, graphic design, graphical technique, etc.)
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent

Technical Graphics (3)

- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail
  - Overview to fine structure
- Serve one clear purpose
  - Description, exploration, decoration, ...
- Be closely integrated with statistical and verbal descriptions of the data set

The rest of the presentation consists of examples of good and bad instances of graphical presentation drawn from a variety of sources. Unfortunately, I have these only in hard-copy form, which I displayed using transparencies. For those who missed the presentation, the following notes might be of some value.
Notes on Graphical Presentation of Quantitative Information

Nearly all of you will be creating technical documents of some sort as part of your career. This talk focuses on a particular aspect: graphical presentation of quantitative information, which is a common component of technical documentation.

This topic brings together two themes regarding the presentation of quantitative information: (1) graphical (statistical) honesty: avoid lying with statistics; and (2) graphical stylistic competence.

In many respects, graphical stylistic competence is similar to technical writing competence. The biggest concerns are sticking to the point, clearly presenting the information, and avoiding excessive ink or distracting graphical elements.

Power: This is the presentation of significant information in a small space. Space in most technical documents is at a premium. If the graphic does not present more information, or better present information, in that space than could be done with words, then words should be used. A good graphic carries a lot of information for its space.

Truth: Do not distort the information. Even honest people can accidentally distort information with poor use of graphics. A typical example is the shadow/3D bars that are prevalent with Excel spreadsheet graphics. They make it difficult to read the height of the bars. Be careful about using 2-dimensional icons to represent one dimensional quantities, since the larger of two 2-dimensional objects is likely to take up too much visual attention for the relative quantities being presented. Avoid abusing ratios (see Jain’s discussion on ratio games.)

Grace: Graphics should have an attractive style. This does not mean that the style should be a distraction from the data. Rather, just as good style in writing means avoiding excessive wordiness, good style in graphics means using the minimum of ink necessary to convey the information.

With the advent of the WWW, everyone suddenly became a publisher. The cycle of graphical abuse repeated itself as inexperienced web page designers abused their new abilities. That is why so many web pages use, for example, motion and blinks, which are generally inappropriate. Ease of use for an advanced feature does not imply that the feature will be used wisely.

We see a similar problem with graphics tools that make it easy to use ill-advised 3D effects for the purpose of “sexing up” quantitative presentation of data. For example, 3D shadow boxes on bar charts and pie charts that make the underlying data far more difficult to interpret and therefore should be avoided.

Making an Argument

Research papers typically involve convincing the reader of something.
- “Making an argument” is the process for doing the convincing
- Depending on the topic, the reader might be more or less friendly or skeptical
- The argument should be thought of as a conversation with readers.
- The goal is to provide a chain of reasoning with appropriate support.
- To properly support requires anticipating what support is needed.

Some Comments on Technology

When personal computers first became popular, a primary reason was the ability to use document processing software to easily create documents. Unfortunately, early document processors (and even some modern programs such as MS Word) do a poor job of guiding users to proper style for things such as page layout and font use. It was a common thing in early documents for users to use excessive font changes and other graphical distractions. Over time, as the availability of advanced features became commonplace, most people learned not to be excessive in their use. A similar problem occurred when color graphics became widely available. The problem is that most people know little about page layout or the proper use of color. The trick is to refrain from abusing such capabilities.

Making an Argument

no notes
The Reasoning Chain (1)

- What is my claim?
- What reasons support my claim?
- What evidence supports my reasons?
- Do I acknowledge alternatives/complications/objections, and how do I respond?
- What warrant makes my reasons relevant to my claim?

There might be sub-claims and/or sub-reasons with their own evidence, warrants, etc.

- A reason, or even “evidence,” is merely a claim to a reader who doesn’t believe it.

The Reasoning Chain (1)

- Claims are based on reasons.
- Reasons are relevant to the claim because of a warrant.
- Reasons are based on evidence.
- All of this has to be understandable by the reader (anticipation of needed support).

Warrant Example (1)

We should leave because $2 + 2 = 4$.

A better example:
- We are facing significantly higher health care costs in Europe and North America... [claim]
- ... because global warming is moving the line of extended hard freezes steadily northward. [reason]
- When an area has fewer hard freezes, it must pay more to combat new diseases carried by subtropical insects no longer killed by hose freezes. [warrant]

Warrant Example (2)

A warrant says:
- If a general circumstance exists [an area has fewer hard freezes] ...
- ... then we can infer a general consequence [that area will have higher costs to combat new diseases].
**Warrant Example (3)**

To work, the reader has to agree with the various parts.

- That the warrant is true [fewer hard freezes in fact mean higher medical costs]
- That the reason is true [hard freezes are in fact moving north]
- The specific circumstance in the reason qualifies as a plausible instance of the general circumstance in the warrant [freezes moving north means the location in question gets fewer hard freezes]
- The specific consequence in the claim qualifies as a plausible instance of the general consequence in the warrant [higher costs to combat new diseases lead to higher overall healthcare costs]

**Warrant Diagram**

Warrant: Condition Consequence

(predictably leads to)

General General

Specific

Reason Argument: Specific Claim (lets us infer)

Any of these four requirements might need to be considered claims, with their own chains of argument to support them.

**Qualifying Claims**

- A bare (unqualified) claim is more forceful
- But if it overreaches its support, reader will lose trust
  - What is overreaching depends on the audience and circumstance
  - Need to anticipate and acknowledge their views
- If it has too much qualification, it has no meaning
- So you need to strike the right balance

**Assembling Reasons and Evidence**

- Evidence is what is considered to be a fact.
- If the reader doesn’t accept something as fact, then it is only a reason to them.

American higher education should review its “hands-off” policy toward off-campus drinking, [claim] because dangerous binge drinking has become a common behavior. [reason]
The injuries and death it causes have increased in frequency and intensity, not only at big “party” schools but among first-year students at small colleges. [evidence]
- Will the reader believe the evidence? It might need support.
Acknowledgement and Response

- Core of argument: Claims backed by reasons based on evidence.
- “Thickening” the core: (Anticipate and) respond to predictable questions and objections.
- Intrinsic soundness: Clarity of claim, relevance of reasons, quality of evidence.
- Extrinsic soundness: Alternatives -- different ways of framing the problem, overlooked evidence, what others have written.
- Predictable disagreements:
  - Causes in addition to the one you claim
  - Counterexamples
  - Disagreements on definitions (definitions are crucial warrants)

Warrants (1)

- A proverb can be a warrant (if the reader agrees with it).
- Problem: many warrants are particular to a smaller community, which the reader might not be party to.
- But readers “too” familiar with the warrant will find declaring it to be distracting.
- Ex: A whale is more closely related to a hippopotamus than to a cow [claim] because it shares more DNA with a hippopotamus. [Reason]
- Warrant: When a species shares more DNA with one species than it does with another, [condition] we infer [claim] because it shares more DNA with a hippopotamus.

Warrants (2)

- A biologist already knows this warrant and so doesn’t want to see it again.
- A non-biologist might need the warrant to understand why the reason applies to the claim.
- Warrants are often not expressed in condition-consequence form, making them harder to understand.
- Ex: Shared DNA is the measure of the relationship between species.

Warrant Example (1)

- Few doubt that when we expose children to examples of admirable behavior, we influence them for the better. How can we then deny that when they are constantly exposed to images of sadistic violence, they are influenced for the worse? [warrant]
- Violence among children 12-16 is rising faster than among any other age group. [reason] [Place evidence of this statement here if necessary.]
- We can no longer ignore the conclusion that TV violence is a destructive influence on our children today.

Problem: The specific instance, rising violence, is not a valid instance of children being exposed to images of sadistic violence. So the reason is irrelevant to the claim.
Warrant Example (2)

Let's restate that previous argument to identify the problem.

- When children are constantly exposed to images of sadistic violence... [general circumstance]
- ... violence among children 12-16 is rising faster than among any other age group. [specific reason]
- Oops... rising violence is not an instance of children being exposed to images of sadistic violence.

Warrant Example (3)

- ... when [children] are constantly exposed to images of sadistic violence, they are influenced for the worse [warrant]
- All our data show that violence among children 12-16 is rising faster than among any other age group. [Evidence here if needed.]
- This violence results from many factors [qualifier], but we can no longer ignore the conclusion that because television is the major source of children’s images of violence, [reason]
- it is a destructive influence on our children. [claim]

This links the evidence and claim to the warrant.

Warrant Example (4)

The warrant connection:

- When children are constantly exposed to images of sadistic violence... [general circumstance]
- Television is the major source of children’s images of violence... [specific instance]
- This specific instance relates to general circumstance

Warrant Example (5)

Full argument:

- When children are exposed to images of sadistic violence... [condition]
- ... that exposure influences them for the worse. [consequence]
- Television is the major source of children’s images of violence... [Specific instance of condition]
- ... it is a destructive influence on children. [specific claim]
Warrant Example (6)

- All the last slide does is indicate that the argument is **syntactically** correct.
- The reader might still reject the argument.
- Examples:
  - Most TV images are not “sadistic” but are cartoon violence.
  - Children are not constantly exposed to sadistic images on TV.

Challenging Other’s Warrants

- Your warrant might not be accepted
- Especially if your warrant conflicts with a warrant the reader already has
- Which means that you effectively have to challenge their existing warrant
- Reader might have a warrant coming from a religious view, a cultural value, or experience.
- Of course, those are hard to challenge!

Why is Web Design Hard?

Web design potentially carries all of the problems associated with:
- Writing
  - Being on message, clarity, brevity, affect, style, consistency, reducing cognitive burdens
- Programming interfaces
  - Usability, consistency, bugs
- Complex technology
- Color
- Visual presentation of quantitative data

You don’t have to give yourself all of these problems!

Some Principles

- What you do matters
  - You make many decisions that affect usability
- Users come to your site with prior usability experiences
  - That’s good when you can leverage their experience
  - That’s bad when you (try to) go against it
- If you aren’t a natural, study usability
  - Usability classes
  - Books and websites (http://www.useit.com)
Some Tools

- HTML (with CSS)
- PHP
- Content Management System (CMS)
  - Drupal

Navigation Issues (1)

The web is a confusing place, and users cling to the few standards they are familiar with.

In general, try not to abuse the user’s expectations

- Don’t change standard expectations for links
  - Be careful about the colors or other mechanism used to indicate a link
  - Don’t use your link color/signal for anything that is not a link
  - Underlining means link
  - Visual signal on mouse-over is not sufficient
  - Visited vs. unvisited link should look different
  - Don’t break the “back button”

Navigation Issues (2)

- Opening windows makes the user feel loss of control
- URLs
  - Used for “dead reckoning” navigation
  - Disadvantageous to make hard to type and remember
- Search engines can drop users anywhere within a site
  - Beware of “orphan” pages

Technology Issues

- More technology = more load time
- Frames are considered problematic
  - Hard to bookmark
  - Often break user’s navigation expectations
- Advertising
  - Users have been trained about what looks like an ad: position, style, motion
  - Users object to advertising
  - Conclusion: You don’t want to look like an ad!
- Motion
  - Users object to motion
  - Motion makes you look like advertising
- Search: Provide for sites that are big enough (100 pages?)
Quality of Information Issues

- Outdated information
  - Let users know how old your pages are by adding “last changed” date stamp
- Long scrolling pages vs. lots of links
- Stylistic (and usability) Consistency
  - Use widgets in standard ways
    - Ex: Selection menu
- Moving pages breaks URLs
- Headlines need to make sense out-of-context
  - Search engines
  - Automatically generated TOC

Some Good Ideas (1)

- Place your name and logo on every page and make the logo a link to the home page
- Write straightforward and simple headlines and page titles
- Structure the page to facilitate scanning and help users ignore large chunks of the page in a single glance
  - Ex: In a 2D table (like a big menu), users will naturally scan down columns, not across rows
- Use hypertext: structure content into overview page and focused topic pages
- Use relevance-enhanced image reduction when preparing small photos and images
- Use link titles to give users a preview of where link will take them

Some Good Ideas (2)

- Design important pages for users with disabilities (e.g., blindness)
- Do the same as everybody else, that’s how users “learn”
- Don’t make a link to the current page
- Date the content (last changed date)
- Make the site’s purpose clear: explain who you are and what you do
- Take advantage of the window title (search engines use this)
- Avoid horizontal scrolling
- Avoid fixing the font size
- Avoid big blocks of text
- Make mail-to links obviously mail-to links

Some Good Ideas (3)

- Beware email address collection bots
- Allocate your homepage real estate wisely
- Use color to distinguish visited and unvisited links
- Use graphics to show real content, not just to decorate your homepage
Publishing on the Web

How do E-journals rate in the publication hierarchy?
- Not so well – but improving

Why are not all journals now E-journals?
- Issues of physical print vs. computer
- Issues of reputation
- Issues of authoritativeness
- Issues of business model
- Issues of reliability (changing URLs)
  - DOIs are improving this

 Talks are Different from Papers

Less formal.
- Can include conjectures, predictions, observations, personal experience
- Can be less precise, incomplete, simplified

Personalized
- Develop a style for presenting

Listeners can not pause, review, or scan ahead
- Need to reinforce meaning
- Repeat key points in different ways

Designing a Talk (1)

Who is the audience?
- Background and sophistication
- Can include extra (optional) slides for additional background

A controversial title can attract attention

Include a title slide (with your name and affiliation)
- If appropriate, include a collaborator/support source slide

Designing a Talk (2)

Outline the talk at the beginning, progress indicators in the middle, summarize at the end
- Tell them what you are going to tell them
- Tell them
- Tell them what you told them

You might lose people along the way
- Build in re-entry points

Talks have many of the same issues as other forms of communication: You are communicating TO an AUDIENCE for a PURPOSE. Talks are more immediate, more scope for impact (good or bad).
Style

- Develop a style for your slides (that you can rely on across talks)
- Sparse vs. decorative
- Thorough presentation vs. signposts
- Regardless of style, keep words (and equations and numbers!) to a minimum.
  - Too much on a slide can distract an audience from your presentation
  - Too much math will intimidate many listeners
- Each slide should have one main idea
- Overlays can be used to build up ideas
  - When a development won’t fit onto one slide, repeat information as necessary

Judging Time

Different presentation styles make it hard to equate number of slides with time.

In part, it depends on whether the slides carry most of the talk or not.
- Write on blackboard or slide as you go?
- Signpost slides? Or complete?

Rule of thumb: 2 minutes per slide

I’m lucky to get through 10 slides in a 50 minute lecture

The better prepared, the more you can cover

Common Mistakes

- Talk is too long
- Too much text per slide
- Technique at the expense of presentation (Powerpoint abuse)
- Any sort of animation/motion/blinkning
- Too much detail (rely on the paper for details)
- Too little motivation and context
- Too much reliance on listener’s memory

Giving a Talk

Things you can control
- Don’t read from the slide!
- Talk to the audience, not the slide
- Don’t speak too quickly
- Don’t speak too quietly
- Don’t speak too long

Things that are hard to control
- Be relaxed. (Remember: The audience wants you to succeed.)
- Be dynamic and interesting
- Don’t speak in a monotone
- Don’t um and er (pause instead)
- Use eye contact with the audience
Preparing to Give a Talk

If you are inexperienced, write out what you want to say in notes.
- As best you can, don’t just read the notes

Can write notes to yourself on the backing paper of the slide

Emphasize just one or two key points that you want the audience to take home

Be sure to give a timed practice talk before the presentation, in addition to your own rehearsals (2-3 times)

But reading the notes is NOT the greatest sin!

Beamer is the nicest tool for combining notes with the slides that I know of!

Physical Considerations

- How big is the room?
  - Is the presentation visible throughout the room?
- How well can the talk be heard?
  - Will you use a microphone?
  - How loud do you need to speak?
- How will you point to the slides?
  - Hand? Pointing device?
- Presentation medium?
  - True slides
  - Transparencies
  - Computer presentation
- Blackboard/Whiteboard available?

Never used, no infrastructure.

Will there be a projector?

Lots of potential for failure. Who’s computer? But now this is standard.