CS5014:

Research Methods in Computer Science

Spring 2004

Instructor: Dr. Cliff Shaffer

Getting Started

The syllabus.

The three topic areas:

- "Acculturation:" How to thrive as a researcher
- Technical presentation
- Experimental design & analysis

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

- Papers
- Proposals
- People
- Programs

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 Be a Succeed in Graduate School” by Marie desJardins, `http://www.erg.sri.com/people/marie/papers/advice-summary.html`
- “Stanford Graduate School Survival Guide” `http://www-smi.stanford.edu/people/pratt/smi/advice.html`
- “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/~pattrsn/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 skill 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.

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.

Graduate Research: How to be Successful

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!

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 an advisor on one criteria, such as support, research area, or personality.

How to be successful: Pick the best compromise (for YOU!) of the following indicators.

- 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

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

Human-Computer Interaction

- Bowman, Ehrich, Harrison, Hix, McKrickard, North, Perez, Tatar, Quek

Bioinformatics (PSEs)

- Choi, Heath, Murali, Onufriev, Ramakrishnan, Shaffer, Watson, Zhang

Systems, Networking, Grids

- Back, Gracanin, Kafura, Ribbens, Sandu, Santos, Varadarajan

Software Engineering

- Arthur, Balci, Edwards, Bohner

Digital Libraries

- Fox

CS Faculty (1)

Sean Arthur: Software engineering (IV&V), translators

Osman Balci: Software engineering (IV&V), modeling and simulation

Doug Bowman: HCI, Virtual Environments

Steve Edwards: Software engineering, programming languages

Roger Ehrich: Image processing, HCI, computers in K12 education

Ed Fox: Multimedia, digital libraries, educational technology

Steve Harrison: HCI, CSCW

Lenny Heath: Algorithms, graph theory, bioinformatics

Hix: HCI, UI Evaluation

Dennis Kafura: Software engineering (OOP), distributed systems, grids

Scott McCrickard: HCI, peripheral interfaces

T.M. Murali: Bioinformatics, gene expression and microarray processing, algorithms

CS Faculty (2)

Chris North: HCI, information visualization

Alexey Onufriev: Bioinformatics, protein-folding models, physics-based models, HPC

Manuel Perez: HCI, user interface software, educational uses of computers

Naren Ramakrishnan: Recommender systems, PSEs, data mining

Cal Ribbens: HPC, grid computing, numerical methods

Adrian Sandu: HPC, computational science, air quality models

Eunice Santos: HPC, scientific computing, algorithms, bioinformatics

Cliff Shaffer: Data structures, PSEs, computational biology, computer-aided education

Deborah Tatar: HCI, cognition, computer aided education, CSCW

Srinidhi Varadarajan: Parallel computation (cluster), networking

Layne Watson: Numerical analysis (optimization), PSEs, computational biology, scientific computing

CS Faculty (3)

Godmar Back: Operating Systems

Shawn Bohner: Software engineering

Vicky Choi: Algorithms, computational biology

Denis Gracanin: Virtual reality, modeling and simulation

Francis Quek: HCI, computer vision

Joao Setubal: Bioinformatics, algorithms

Liqing Zhang: Bioinformatics

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 last year

Here its 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 emphesis on preparing for some exam

Risk: Not a clear termination point

Questions and comments?

Writing a Thesis

What's the big deal?

- Its long: Psychological factors
- Its long: Document management issues
- Its long: Project management issues
- Its 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 (cont)

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!

Thesis Defense

In our department, its 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 preceeded by a private discussion among the committee on student's record, groundrules, 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

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

Acculturation Tidbit: The role of gifts in American society

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

Writing Tools of the Trade

- Document processor/text editor
- Dictionary
- Thesarus
- Spelling/grammer checkers
- Style and usage guides
- Technical writing guides

General Advice

The purpose of writing is to convey something TO A READER.

- Your goal is to have an (appropriate) impact

Keep your writing simple. Aim for economy of words.

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

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.

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
- Emphesize 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, their own aircraft, their own numerical or security software, or other critical systems their lives depend on.

So don't try to design your own page layouts. Leave it to an expert.

Ideally, your software gives you a suitable page layout. If not, you probably will be following some specified 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 figures
- Table of contents, figures, etc
- Index
- Cross referencing
- Spell checker
- Style checker

Contrast MS Word to LaTeX for these tasks.

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."

General 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

Than, That

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.

Avoid using "very"

"In order to" can almost always drop "in order"

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

Hotspots (continued)

“Use” not “usage” or “utilize”.

Be careful of ambiguous pronouns.

Proper names are capitalized. “In Section 3 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.”

Be sure not to mis-spell ”analyze”!!

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

What Does a Researcher Do?

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)

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 after about six years

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.

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 document, even after deciding the content.

- Students? Industry professionals? Researchers?
- Readers in the field? Out of the field?
- Tutorial work? Research results? Review/Synthesis?

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

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

Author List

Most papers have multiple authors.

Who gets listed as an author?

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. Its not your right to decide.

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 and other 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

Typically write abstract **after** the paper is essentially complete

Journal paper-style abstracts are typically 200-300 words

Don't start with "In this paper..."

Note: 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

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

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

As appropriate, a standard analysis, or use of a standard testbed, is a good idea. 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 its 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

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 its 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)

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

Citations (cont)

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

- Its 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 processor system doesn't give you equivalent support for formatting reference lists, switch to a real document processing system.
- 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.

Reference Lists (cont)

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

J.M.A. Begole, C.A. Struble, C.A. Shaffer, and R.L. Smith, "System Resource Sharing for Synchronous Collaboration," *IEEE Transactions on Networking 9*, 6(Dec 2001) 833-843.

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

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)

Examples of Reference List Styles (cont)

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

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

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

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!

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 amiguous 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 definately 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

Some venues take longer than others

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

Each venue normally has specs to meet

- Where is it sent?
- number of copies of manuscript; paper vs. electronic submission
- Layout? These specs might be vague, or highly constrained
- Online submission is becoming popular

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

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

- 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
- 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

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.

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. 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 (cont)

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, and there is a 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

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

Is the presentation satisfactory?

- References appropriate
- English satisfactory, style satisfactory
- Sufficiently complete
- 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?

Why Write a Grant Proposal?

To the Government:

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

To Industry:

- "Classic" techical research proposal
- Equipment donation
- Charitable/community organization sponsorship

Foundation

- Highly targetted research
- Programmatic support
- Charitable/community organization support

Internal Proposal

- Technical research

How Proposals are Reviewed

Who does it depends on the funder

- Foundation: Board of directors
- Industry: Employees
- Some federal agencies: Agency employees or professional "handlers"
- "Classic" proposal solicitiation 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 (cont)

Typical process:

- Reviewers catagorize into bins (excellent, poor)
- Reviewer "panel" might synthesize individual reviews
- Admistrator 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: Definate yes, grey area, definate 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.

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

Strategies for Success (2)

Your original contact letter to a foundation or corporation can be very 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.

After identifying available funding organize your efforts into writing the proposal, marketing, and management.

Read and follow the instructions. With any funder, but especially with federal agencies, not following the rules exactly can ruin your chances, even if your proposal is a brilliant one.

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.

Strategies for Success (3)

When writing your proposal remember:

- get straight to the point- don't waste time or words
- don't rush writing the proposal- take your time and get it right
- be don't request more money than you really need
- be candid about asking for money
- never submit the same application twice
- never lie

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!

Strategies for Success (4)

Having clear performance standards should not be underestimated. Outline measurement indicators and determine result areas. A strong proposal proves that it is likely to achieve its goals.

Cooperation is a good idea. Many funders like applications involving more than one organization. Be sure that the grantees have both a formal and informal relationship before submitting your cooperative proposal.

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.

- CAS: Don't make yourself a liar
- CAS: Samauri Sword metaphor
- CAS: Be careful of commitments in budget re-negotiations

Strategies for Success (5)

Stay on schedule. 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.

Keep all of your standard information (resumes, community statistics) in a file that can be constantly updated. This will save you time and allow you to concentrate on the specifics of preparing your proposal.

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 (6)

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.

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: 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

Common Mistakes

- No goals
  - Each model is special purpose
  - Performance problems are vague when first presented
- Biased goals
- Unsystematic approach
- Analysis without understanding
  - People want guidance, not models
- Incorrect performance metrics
  - Want correct metrics, not easy ones
- Unrepresentative workload
- Wrong evaluation technique
  - Easy to become married to one approach
- Overlooking important parameters
- Ignoring significant factors
  - Parameters that are varied in the study are called factors.
  - There's no use comparing what can't be changed
- Inappropriate experimental design
- Inappropriate level of detail

Common Mistakes (cont)

- No analysis
- Erroneous analysis
- No sensitivity analysis
  - Measure the effect on changing a parameter
- Ignoring errors in input
- Improper treatment of outliers
  - Some should be ignored (can't happen)
  - Some should be retained (key special cases)
- Assuming no change in the future
- Ignoring variability
  - Mean is of low significance in the face of high variability
- Too complex analysis
  - Complex models are "interesting" and so get published and studied
  - Real world use is simpler
  - Decision makers prefer simpler models

Common Mistakes (cont)

- Improper presentation of results
  - The proper metric for analyst performance is number of analyses that helped decision makers.
- Ignoring social aspects
- Omitting assumptions and limitations

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 distastful 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 Systematic Approach

1. State goals and define the system
2. List services and outcomes
3. Select metrics
4. List parameters
5. Select factors to study
6. Select evaluation technique
7. Select workload
8. Design experiments
9. Analyze and interpret data
10. Present results

Selecting a Technique

Choices: Analytical Modeling, Simulation, Measurement

"Until validated, all evaluation results are suspect."

Validate one of these approaches by comparing against another.

Measurement results are just as susceptible to experimental errors and biases as the other two techniques.

Criteria

- Stage of analysis
- Time required
- Tools
- Accuracy
- Trade-off evaluation
- Cost
- Saleability

Performance

Throughput: The rate (requests per unit of time) at which requests can be serviced by the system.

- Throughput generally increases as the load initially increases.
- Eventually it stops increasing, and might then decrease.
- Nominal capacity is maximum achievable throughput under ideal workload conditions.
- Usable capacity is maximum throughput achievable without violating a limit on response time.
- Efficiency is ratio of usable to nominal capacity.

Example: Hashing systems. Knee capacity.

Utilization: Fraction of time the resource is busy servicing requests.

Bottleneck: Component with highest utilization. Improving this component often gives highest payoff.

Workloads

A workload is the requests made by users of the system under study.

- A test workload is any workload used in performance studies
- A real workload is one observed on a real system. It cannot be repeated.
- A synthetic workload is a reproduction of a real workload to be applied to the tested system

Examples (for CPU performance)

- Addition instruction
- Instruction mixes
- Kernels
- Synthetic programs
- Application benchmarks

What if you are an HCI researcher?

Selecting a Workload

1. Determine the services for the SUT
2. Select the desired level of detail
3. Confirm that the workload is representative
4. Is the workload still valid?

A real-world workload is not repeatable. Most workloads are models of real service requests.

Markov Models

Assume the next request depends only on the last request.

Much more powerful than purely random probability.

- Ex: There is a distribution of letters in English, but it is different for beginning letters in a word

Transition matrices

- Ex: typical distribution for some system is about 4 small packets followed by a single large packet.
- Random chance: probability of small is always .8, large is always .2.
- Markov model: Small follows small .75, large follows small .25. In contrast, small always follows large.

Clustering

Goal: To separate a population into groups with similar characteristics.

- Specifically, the goal is to minimize the within-group variance while maximizing the between-group variance.

Uses:

- Select representatives to simplify further processing
- Recommender systems, retrieval systems

Steps:

- (Random) sampling
- Parameter selection
- Transformation of parameters (log scale?), scale (normalize data)
- Cull outliers
- Select distance metric (e.g., euclidean distance)
- Clustering technique. Ex: minimum spanning tree
- Assign full population to clusters

Variable Types

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

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

Statistics: Basic Concepts

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

Random Variable: takes one of a specified set of values with specified probability.

Mean or Expected Value (for discrete variables):

$$\mu = E(x) = \sum_{i=1}^{n} p_i x_i$$

Variance:

$$\mathrm{Var}(x) = \sigma^2 = E[(x-\mu)^2] = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

Standard Deviation: $\sqrt{\mathrm{Var}(x)} = \sigma$

Coefficient of Variation: σ/μ

Basic Concepts (cont)

Covariance: Given random variables x and y with means μ_x and μ_y:

$$\begin{aligned}\text{Cov}(x,y) = \sigma_{xy}^2 &= E[(x-\mu_x)(y-\mu_y)] \\ &= E(xy) - E(x)E(y) \\ &= \frac{\sum(x_i-\mu_x)(y_i-\mu_y)}{N-1}\end{aligned}$$

Correlation: Normalized value of covariance (always between -1 and 1)

$$\text{Correlation}(x,y) = \rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

Alternate formulation (Pearson's r):

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

Mean, Median, Mode

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_α, the x value at which the CDF takes a value α is called α-quantile.

Mean or Expected Value: $\mu = E(x) = \sum_{i=1}^{n} p_i x_i$

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

Mode: The most likely value

Choosing Mean, Median, Mode

Often want to give a single number that "characterizes" a data set.

- Indices of Central Tendency
- Favorites are Mean, Median, Mode
- The term "average" has no meaning

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.) Probably use mean.

Are the data skewed? Use Median

- GTA's mistake
- Otherwise, can use mean or median

Indices of Dispersion

"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

There are different ways to measure variability

Range (max - min)

- poor, tends to be unbounded, unstable over a range of observations, susceptible to outliers

Variance (s^2), standard deviation (s)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

- standard deviation is in units of the mean

10- and 90-precentiles

Quartiles (box plots)

Histograms

Might like to describe the distribution of a set of measured observations.

- Eg. uniform distribution, normal distribution

Plotting a histogram:

- Need to compute min and max of data
- Need to determine a cell size
- Danger: Small cell size means few observations per cell, and large variation in obeservations per cell
  - Rule of thumb: Do not permit less than 5 observations in a cell
- Danger: Large cell size loses distribution information

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

See Jain p. 198–199

Sampling

Most CS research work involves samples from some "population."

- HCI subjects
- Performance experiments on workloads for programs, systems, networks

A fundamental goal of CS experimentation is to determine the mean (μ) and variance of some population characteristic (a parameter).

- We can only measure the mean characteristic for the sample ($\overline{x}$), not the population (a statistic)

The parameter value (population mean) is fixed.

The statistic (sample mean) is a random variable.

The values for the statistic come from some sampling distribution.

Hypothesis Testing

One fundamental experimental question: What is the parameter value?

Another fundamental experimental question: Are two populations different?

- Does a subpopulation perform better?
- Is one algorithm better than another?

Null Hypothesis: Two populations have the same mean

Rejection of the Null Hypothesis: Do so if you are "confident" that the means are different

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

Confidence Intervals

Each sample mean is an estimate of the population mean

It is not possible to get a perfect estimate of population mean from a finite number of samples

- The best we can get is a probabilistic bound

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
- α is the significance level
- $100(1-\alpha)$ is the confidence level
- $1-\alpha$ is the confidence coefficient

Examples: Typically 90 or 95% confidence levels, equivalently 0.1 or 0.05 significance levels

For 90% confidence levels, find the 5- and 95-percentiles of the sample means

Central Limit Theorem

If the n observations in a sample are independent, and come from the same population that has mean μ and standard deviation σ, then the sample mean (for large samples) is a randome variable whose distribution is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$.

The standard deviation of the sample mean is called the standard error.

Note the requirement: The population must be normally distributed.

We can compute a $100(1-\alpha)$ confidence interval for a sample mean as follows:

- $\overline{x} \pm (z_{1-\alpha/2}) s/\sqrt{n}$
- where s is the sample standard deviation, and
- $z_{1-\alpha/2}$ is the quantile from a unit normal variate (take from a table)

Example 12.4

Numbers: 1.9, 2.7, 2.8, 2.8, 2.8, 2.9, 3.1, 3.1, 3.2, 3.2, 3.3, 3.4, 3.6, 3.7, 3.8, 3.9, 3.9, 3.9, 4.1. 4.1, 4.2, 4.2, 4.4, 4.5, 4.5, 4.8, 4.9, 5.1, 5.1, 5.3, 5.6, 5.9

Sample size: 32

Sample mean $\overline{x}$: 3.90

Sample standard deviation s: 0.95

From a table:

- $(1 - 0.9)/2 = 5\% = 1.645$
- $2.5\% = 1.960$
- $0.5\% = 2.576$

Examples:

- $90\% : 3.90 \pm (1.645)(0.95)/\sqrt{32} = (3.62, 4.17)$
- $95\% : 3.90 \pm (1.960)(0.95)/\sqrt{32} = (3.57, 4.23)$
- $99\% : 3.90 \pm (2.576)(0.95)/\sqrt{32} = (3.46, 4.33)$

Sampling Distributions

The sampling distribution only approximates a normal curve when 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.

Usually there is a table to look it up in, or better yet, the statistical software will calculate it for you.

Test Against Zero

Are the differences in processor times for two algorithms significant?

- Testing if mean (differences) is zero. Check if zero is within confidence interval at desired confidence.
- Note that the number of samples is likely to be small. Take from table for correct t value.

Example

- Sample size is 7.
- Taking differences, sample mean is 1.03
- Alternate variance calculation: $\frac{1}{n}\sum X_i^2 - \overline{X}^2$
- Sample variance: 2.57
- Sample standard deviation: 1.60
- Confidence interval:
  $1.03 \pm t \times 1.60/\sqrt{7} = 1.03 \pm 0.605t$
- For 6 degrees of freedom, $t_{[0.995;6]} = 3.707$
- The 99% confidence interval is (-1.21, 3.27)

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

Example:

- (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: $-0.32 \pm t(3.69)$
- For 5 degrees of freedom, 0.95-quantile t value is 2.015
- 90% confidence interval is $-0.32 \pm (2.015)(3.69) = (-7.75, 7.11)$.

Unpaired Observations

This is the classic t-test.

- We have a collection of observations from P1,and another collection of observations from P2
- We want to know if the two populations have the same mean

Example:

- System A: 5.36, 16.57, 0.62, 1.41, 0.64, 7.26
- System B: 19.12, 3.52, 3.38, 2.50, 3.60, 1.74
- Mean A: 5.31, Variance A: 37.92
- Mean B: 5.64, Variance B: 44.11
- Mean difference: -0.33
- Standard deviation of mean difference: 3.698
- Effective number of degrees of freedom: 11.921
- t-value from table at .95-quantile for 12 DOF: 1.782 (note error in book)
- 90% confidence interval for difference: $-0.33 \pm (3.698)(1.782) = (-6.92, 6.26)$

Philosophy

What confidence interval to use?

- Depends on the loss if wrong
- Sometimes 50% confidence is enough, or 99% confidence is not enough

Regression Models

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

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

Error: $e_i = y_i - \widehat{y}$ for the ith 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:

$$b_1 = \frac{\sum xy - n\overline{xy}}{\sum x^2 - n(\overline{x})^2}$$

$$b_0 = \overline{y} - b_1\overline{x}$$

Example:

- 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 $-0.0083 + 0.2438$(# of disk I/Os)

Allocation of Variation

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?

SSE without regression would be squares of differences from the mean

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

Allocation of Variation Example

From previous example,
$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{199.84}{205.71} = 0.9715$. 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

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.

Degrees of Freedom:

$$\text{SST} = \text{SSY} - \text{SS0} = \text{SSR} + \text{SSE}$$

$$n - 1 = n - 1 = 1 + (n - 2)$$

Assumptions

Assumption: Linear relationship between x and y

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

1. No visible relationship
2. Linear trend line
3. Two or more lines (piecewise)
4. Outlier(s)
5. Non-linear trend curve

Only the second one can use a linear model

The third one could use multiple linear models

Assumption: Independent Errors

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

Assumption: The predictor variable x is nonstochastic and it is measured without any error.

Limits to Simple Linear Regression

Only one predictor is used.

The predictor variable must be quantitative (not catagorical).

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 $x_1, x_2, ..., x_k$, use the linear model

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

The b_is are fixed parameters, e is the error term.

In matrix terms, $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$

Of key interest is finding values for the b_is. This is (usually) an over-constrained set of equations. See Jain's formula for a method of getting values.

Next questions are:

1. Is the model significant?
2. Are the parameter values significant?

Multiple Regression Example (15.1)

Table 15.1 shows observations of CPU time, Disk I/O, and Memory Size.

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

CPU time = -0.1614 + 0.1182(# of disk I/O) + 0.0265(memory size)

We can compute the error terms from this (difference between the regression formula and the actual observations)

We can compute the coefficient of determination SSR/SST to find that the regression explains 97% of the variation of y.

The 90% 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, the model seems to give good predictive ability. For 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 (ANOVA)

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_js 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.

Thus, the actual ratio calculated is (SSR/k)/(SSE/($n - k - 1$)).

ANOVA (cont)

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

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

The problem is that the correlation between the two predictors (memory and disk I/O) is 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.

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 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

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 $y = bx^a$

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

- I/O rate = α(CPU rate)^{b_1}
- Taking logs we get log(I/O rate) = $\log \alpha + b_1$ log(MIPS rate).
- Using standard linear regression, we find that the regression explains 84% of the variation.

Outliers

"Any observation that is atypical of the remaining observations *may* be considered an outlier."

The key question is whether the outlier represents a correct observation of system behavior.

An outlier might make a big change in the analysis, even to the extent of changing the conclusion (i.e., significance at a given confidence level).

To identify outliers, look at a scatterplot of the data.

If the outlier is not clearly an erroneous observation, then might want to do analysis with and without the outlier(s) and report both.

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.

Crosstabulations

EMPLOYMENT STATUS	RACE			
	WHITE	BLACK	OTHER	
UNEMPLOYED	236.0	39.0	5.0	280.0
PARTTIME	84.0	8.0	8.0	100.0
FULLTIME	415.0	69.0	16.0	500.0
RETIRED	105.0	14.0	1.0	120.0
	840.0	130.0	30.0	1000.0

Observed

EMPLOYMENT STATUS	RACE			
	WHITE	BLACK	OTHER	
UNEMPLOYED	235.2	36.4	8.4	280.0
PARTTIME	84.0	13.0	3.0	100.0
FULLTIME	420.0	65.0	15.0	500.0
RETIRED	100.8	15.6	3.6	120.0
	840.0	130.0	30.0	1000.0

Expected

EMPLOYMENT STATUS	RACE		
	WHITE	BLACK	OTHER
UNEMPLOYED	**+0.8**	**+2.6**	**-3.4**
PARTTIME	**0.0**	**-5.0**	**+5.0**
FULLTIME	**-5.0**	**+4.0**	**+1.0**
RETIRED	**+4.2**	**-1.6**	**-2.6**

Differences

Is this significant?

Chi-squares

We will determine significance using the following fomula:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

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 irrelevent
- Strength should not depend on absolute values. Doubling the number of observations in all cells should not affect strength.

Testing Significance

Simlar to ANOVA, 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 χ^2 would double, and now it would be significant at 99% confidence.
- Warning: The "strength" has not increased!

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 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

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.

One-factor-at-a-time studies do not capture factor interactions.

Types of Experimental Designs

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 (ex: 2^k design), number of factors, or take another approach.

Types of Experimental Designs (cont)

Fractional factorial design

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

F1	F2	F3	F4
1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	1	2
3	1	3	2
3	2	1	3
3	3	2	1

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

Example

Start with a 2^2 experimental design.

Cache	Memory 4	Memory 16
1	15	45
2	25	75

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:

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

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

$$\begin{aligned} 15 &= q_0 - q_A - q_B + q_{AB} \\ 45 &= q_0 + q_A - q_B - q_{AB} \\ 25 &= q_0 - q_A + q_B - q_{AB} \\ 75 &= q_0 + q_A + q_B + q_{AB} \end{aligned}$$

Example (cont)

Solve to yield:

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

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.

Sign table:

I	A	B	AB	y
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	80	40	20	Total
40	20	10	5	Tota/4

Allocation of Variation

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:

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

$$\begin{aligned} SST &= SSA + SSB + SSAB \\ &= 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2. \end{aligned}$$

$2100 = 1600 + 400 + 100.$

The fraction of variation explained by A is SSA/SST.

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

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

General 2^k Factorial Designs

We can generalize this to k factors.

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

Factor	Level -1	Level 1
Memory (A)	4MB	16MB
Cache (B)	1KB	2KB
Processors(C)	1	2

	4MB		16MB	
Cache	One P	Two P	One P	Two P
1	14	46	22	58
2	10	50	34	86

$$\begin{aligned} SST &= 2^3(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2 \\ &= 8(10^2 + 5^2 + 20^2 + 5^2 + 2^2 + 3^2 + 1^2) \\ &= 800 + 200 + 3200 + 200 + 32 + 72 + 8 = 4512. \end{aligned}$$

Memory explains 18%, cache explains 4%, and number of processors explains 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:

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e.$$

Example

y			$\overline{y}$	errors		
15	18	12	15	0	3	-3
45	48	51	48	3	0	3
25	28	19	24	1	4	-5
75	75	81	77	-2	-2	4

Coefficients (divide by 4):
$q_0 = 41, q_A = 21.5, q_B = 9.5, q_{AB} = 5$

$$\begin{aligned} SST &= SSA + SSB + SSAB + SSE \\ 7032 &= 5547 + 1083 + 300 + 102 \end{aligned}$$

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 41 ± 1.92. So it is significant.

2^{k-p} Fractional Factorial Designs

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.
- Do the following set of 8 experiments on 7 factors.

#	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

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

$$q_j = \sum_i y_i x_{ji}$$

That is, sum up the column values times their coefficients and divide by 8.

Example

I	A	B	C	D	E	F	G	y
1	-1	-1	-1	1	1	1	-1	20
1	1	-1	-1	-1	-1	1	1	35
1	-1	1	-1	-1	1	-1	1	7
1	1	1	-1	1	-1	-1	-1	42
1	-1	-1	1	1	-1	-1	1	36
1	1	-1	1	-1	1	-1	-1	50
1	-1	1	1	-1	-1	1	-1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Tot
39.6	12.6	4.4	13.6	5.4	0.1	5.9	0.4	Tot/8

Calculated variation by variable:
$A = 37\%, B = 5\%, C = 43\%, D = 7\%, E = 0\%, F = 8\%, G = 0\%$

Further experiments should be conducted only on A and C.

Big assumption: These experiments only provide so much information.

- The effects due to interaction are in the values calcuated for the separate variables
- The experiments are masking these "confounding" interactions
- If the interaction effects are small, its OK

One-Factor Experiments

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

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

where μ is mean response, α_j is the effect of alternative j and e_{ij} is the error term.

We choose to scale things such that the α_j's sum to zero.

	144	101	130	
	120	144	180	
	176	211	141	
	288	288	374	
	144	72	302	Sum
Col Sum	872	816	1127	2815
Col Mean	174.4	163.2	225.4	187.7
Effect	-13.3	-24.5	37.7	0

Errors and Variation

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.

- SST = SSY − SS0 = SSA + SSE
- Variation is divided into explained and unexplained error

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

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.

$$
\begin{aligned}
SSY &= SS0 + SSA + SSE \\
ar &= 1 + (a-1) + a(r-1)
\end{aligned}
$$

- We have ar indpendent 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.

Calcuate 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

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

We can now compute error terms using the model

$$y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}$$

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, $2 \times 4 = 8$ 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

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 (Figure 21.1) looks reasonable.

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

Case Study

Problem: What factors affect the pedagogical effectiveness of algorithm visualizations?

Purvi Saraiya, Cliff Shaffer, Scott McCrickard, Chris North

Context:

- There are 10 Gazillion java applets for algorithm/data structure visualization
- There are 1 Gazillion papers about algorithm visualizations
- 10% of them did enough analysis to claim that the students "like" the visualization, but can determine no significant effect over lecture or book
- .1% are good enough and have conducted analysis to claim a significant effect

Identifying Factors

Used "Expert review panel" (ourselves) to evaluate a collection of heapsort visualizations

- Goal: Identify key factors, (expert judgement)
- Several visualizations of one algorithm
- Noted both good and bad points to each visualization
- Synthesized observations into feature list

Features

- High usability, no bugs
- Student data input
- Appropriate feedback messages (background appropriate)
- User control (vs. animation)
- Backing up
- State changes are clear
- Multiple views (physical vs. logical views)
- Window management/relationships
- Pseudocode
- Guided (questions to answer)

Some of these were determined to be "baseline" requirements, not to be directly studied for their significance.

Features of Interest

- Student input
- Sample given
- User control vs. animation
- Pseudocode
- Back button
- Guide

Experimental Conditions

Created our own Heapsort visualization for which we could vary factors

Who to study?

- CS classes
- Should know basic DS (lists, etc) and simple sorts
- CS2604: too much background? Must catch before they cover heapsort
- CS2704: too little background?
- Freshman courses: Definately too little background

Need to minimize the number of experiments, since we have access to about 100 students.

- Each "experiment" needs about 10-15 students

Key motivational concern:

- Want to demonstrate that something matters
- Want to have some form of a visualization that has a significant effect as compared to some other form of the visualization

Experiment 1

#	Sample	Input	Back	Pseudo	Guide
1	X				
2		X	X		
3		X	X	X	
4		X	X		X
5		X	X	X	X

Results:

- 66 students. 2 outliers excluded (no sorting knowledge)
- Version 1 better (not significant) than others, 3 worse (not significant) on procedural questions
- Versions 3 and 5 better (not significant) than others on conceptual questions

Expert conclusions:

- Didn't properly testing the effects of giving a sample input
- Neglected to test for user control (previously observed as a key factor)

Experiment 2

#	Sample	Input	Back	Pseudo	Guide	Control
1		X				X
2	X	X	X			X
3	X	X	X	X	X	X
4		X				

Results

- 44 students
- Procedural: 2 better than 1, 2 much better than 4
- Pseudocode did not seem to help
- Tracked time taken, that did not correlate to performance
- Students with low GPA did well (better than expected)
- Control is confirmed as an important feature

Final conclusions: Full control over a simple visualization with a good imput example is effective.

Notes on Graphical Presentation of Quantitative Information

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.

Graphical presentations of quantitative information should strive for three virtues:

Power: This is the presentation of significant information in a small space. Space in most techinical 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 accidently distort information with poor use of graphics. NEVER use the shadow/3D bars that are prevelent with Excel spreadsheet graphics, since they make it difficult to read the hight 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 relatively 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.

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 occured when color grapics 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.

With the advent of the WWW, everyone suddenly became a publisher. The cycle of graphical abuse returned 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.

Technical Graphics

See three books by Edward R. Tufte:

- *The Visual Display of Quantitative Information*
- *Envisioning Information*
- *Visual Explanations*

Graphical displays should

- 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
- 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

Talks are Different from Papers

Less formal.

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

Personalized

- Develop a style

Listeners can not pause, review, or scan ahead

- Need to reinforce meaning
- Repeat key points in different ways

Designing a Talk

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

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

Style

Develop a style

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

1. Talk is too long
2. Too much text per slide
3. Technique at the expense of presentation (Powerpoint abuse)
4. Any sort of animation/motion/blinking
5. Too much detail (rely on the paper for details)
6. Too little motivation and context
7. 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)

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?

Publishing on the Web

How do E-journals rate in the publication hierarchy?

- Not well

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)

Nielsen's Top 10 Mistakes in Web Design

See `www.useit.com` Alertbox columns.

Original List (1996)

1. Using Frames
2. Gratuitous use of bleeding-edge technology
3. Scrolling text, marquees, and constantly running animation
4. Complex URLs
5. Orphan pages
6. Long scrolling pages
7. Lack of navigation support
8. Non-standard link colors
9. Outdate information
10. Overly long download times

For Nielsen at least, this list has stood the test of time (1999).

More Web Design Mistakes

An additional list of common problems (1999)

1. Breaking or Slowing Down the Back Button
2. Opening New Browser Windows
3. Non-Standard Use of GUI Widgets
4. Lack of Biographies
5. Lack of Archives
6. Moving Pages to New URLs
7. Headlines That Make No Sense Out of Context
8. Jumping at the Latest Internet Buzzword
9. Slow Server Response Times
10. Anything That Looks Like Advertising

Some Good Ideas

1. Place your name and logo on every page and make the logo a link to the home page
2. Provide search if the site has more than 100 pages.
3. Write straightforward and simple headlines and page titles
4. Structure the page to facilitate scanning and help users ignore large chunks of the page in a single glance
5. Use hypertext: structure content into overview page and focused topic pages
6. Use relevance-enhanced image reduction when preparing small photos and images
7. Use link titles to give users a preview of where link will take them
8. Design important pages for users with disabilities (e.g., blindness)
9. Do the same as everybody else, that's how users "learn"

Other Advice

- 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
- 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