Psychology 205: Research Methods in Psychology
Review

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

November, 2016

Outline

Goals

1. Introduce you to fundamental skills in psychological research.
2. To facilitate your understanding of substantive courses in psychology.
3. To make you a better consumer of scientific information.
4. To improve your ability to write and read scientific papers.

Theory development and testing

1. Theories as organizations of observables
2. Constructs, latent variables and observables
    - Observables
        - Multiple levels of description and abstraction
        - Multiple levels of inference about observables
        - What is observed or not observed is part of theory
    - Latent Variables
        - Latent variables as the common theme of a set of observables
        - Central tendency across time, space, people, situations
        - Constructs as organizations of latent variables and observed variables

Observed Variables

X Y

$\boxed{X_1}$ $\boxed{Y_1}$

$\boxed{X_2}$ $\boxed{Y_2}$

$\boxed{X_3}$ $\boxed{Y_3}$

$\boxed{X_4}$ $\boxed{Y_4}$

$\boxed{X_5}$ $\boxed{Y_5}$

$\boxed{X_6}$ $\boxed{Y_6}$

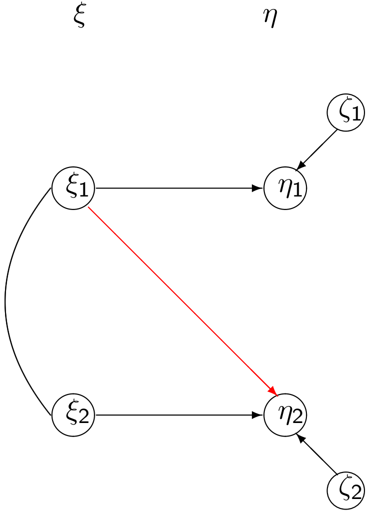
Latent Variables

ξ η

ξ_1 η_1

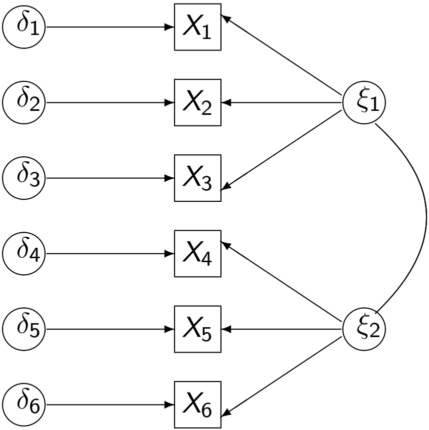
ξ_2 η_2

Theory: A regression model of latent variables

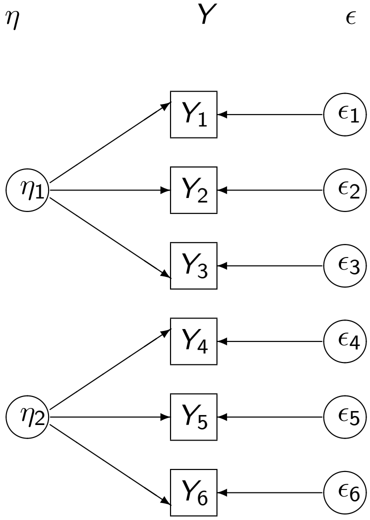


A measurement model for X – Correlated factors

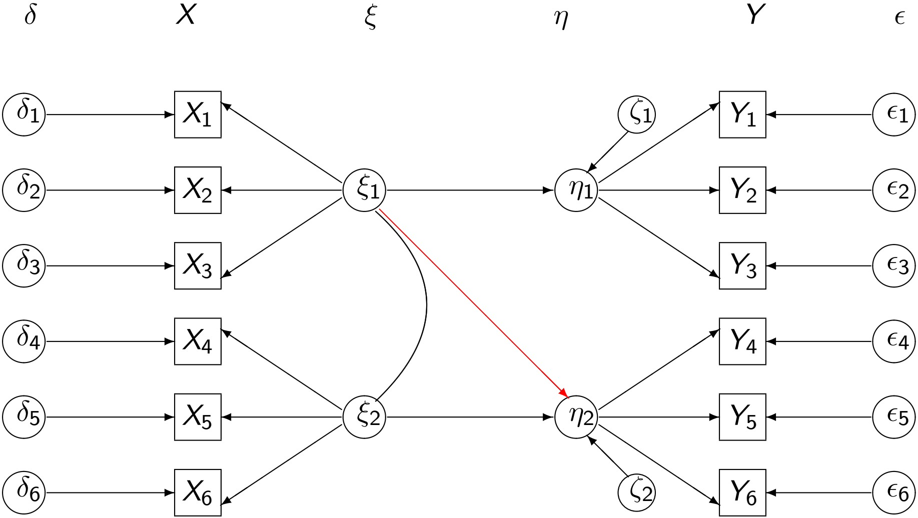
δ X ξ



A measurement model for Y - uncorrelated factors



A complete structural model



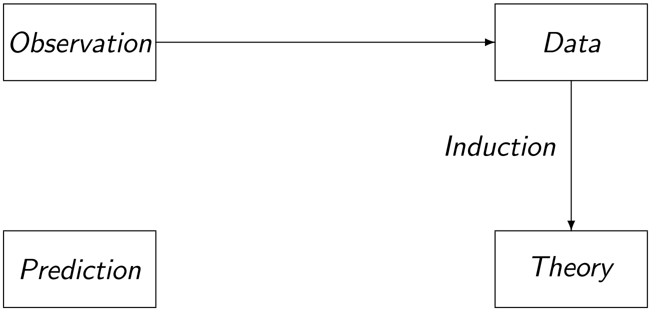
Reasoning in Research

1. Karl Popper and the testability of theory
    - The hallmark of science is the testability of theory
    - Non-testable theories are not science
    - "it must be possible for all empirical scientific system to be refuted by experience"
    - Theories are not shown to be correct, they are shown to be incorrect
2. Science is the process of asking questions that have answers (Rep. Rush Holt)
3. All models (theories) are wrong, but some are useful

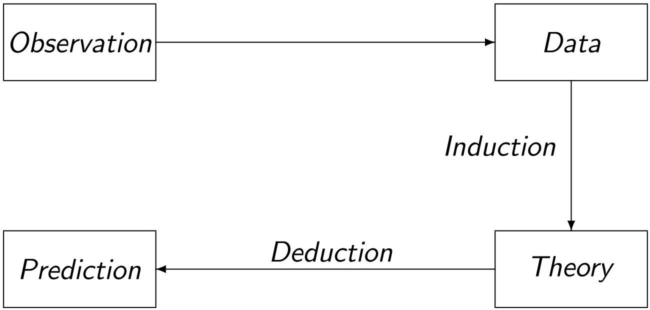
Reasoning in research

Reasoning in research

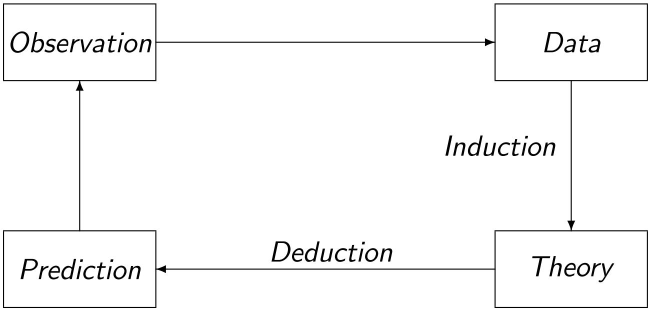
Reasoning in research



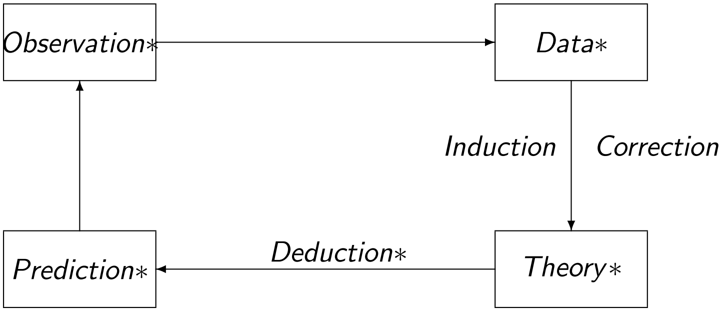
Reasoning in research



Reasoning in research



Reasoning in research



Deduction from Theory

1. Data –> Induction –> Theory
2. Theory –> Deduction –> Prediction/confirmation
3. Logical reasoning: Consider theory

If P, then Q

1. Appropriate Logical deductions
    - P –> Q Affirm the Antecedent
    - Not Q –> Not P Deny the Consequent
2. Incorrect logical deductions
    - Not P -> Not Q deny the antecedent
    - Q –> P affirming the consequent

J. Platt and Strong Inference (Science, 1964)

4 signs of strong science

1. Devising alternative hypotheses;
2. Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly is possible, exclude one or more of the hypotheses;
3. Carrying out the experiment so as to get a clean result;
4. Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain, and so on.

Platt and Strong Inference

"I will mention one severe but useful private test - a touchstone of strong inference - that removes the necessity for third-person criticism, because it is a test that anyone can learn to carry with him for use as needed.
It is our old friend the Baconian "exclusion," but I call it "The Question."
Obviously it should be applied as much to one's own thinking as to others'. It consists of asking in your own mind, on hearing any scientific explanation or theory put forward,
"But sir, what experiment could *dis*prove your hypothesis?"; or, on hearing a scientific experiment described, "But sir, what hypothesis does your experiment *dis*prove?" Platt, Science, 1964

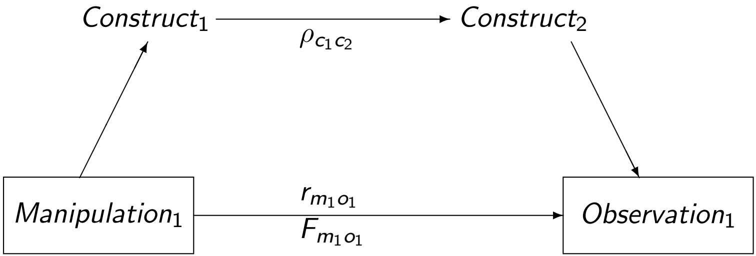
Reasoning in Research

1. Observe, Induce, Deduce, Predict, Observe
2. Disconfirm, don't confirm
3. Prune the tree of alternative explanations

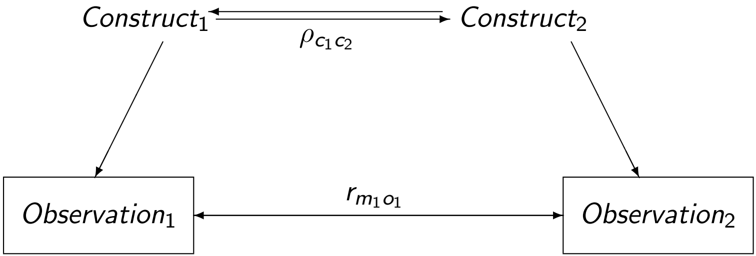
Theory and Theory Testing
Types of Designs

1. Experimental
   - Manipulation of at least one variable - Independent Variable (IV)
   - Can be subject variables (SV) if randomly assigned to conditons
   - Effect on (at least one) other variable - Dependent Variable (DV)
2. Correlational
   - Observation of the relationship between two variables
   - Typically with subject variables
   - Inability to determine causality
3. Quasi experimental
   - Field studies Direct relevance
   - Difficult to have appropriate controls

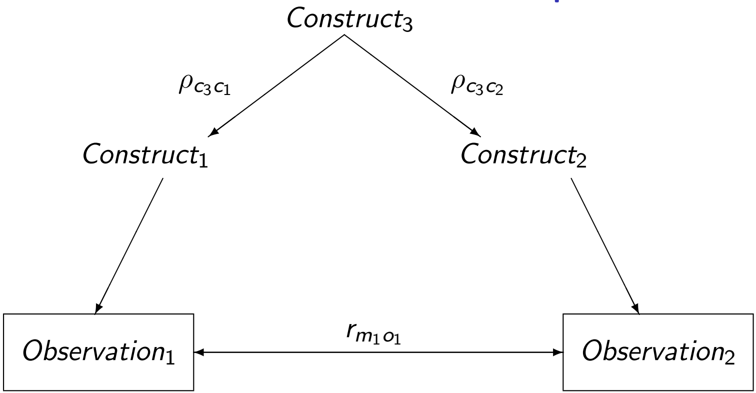
Theory and Theory Testing
Experimental Manipulations



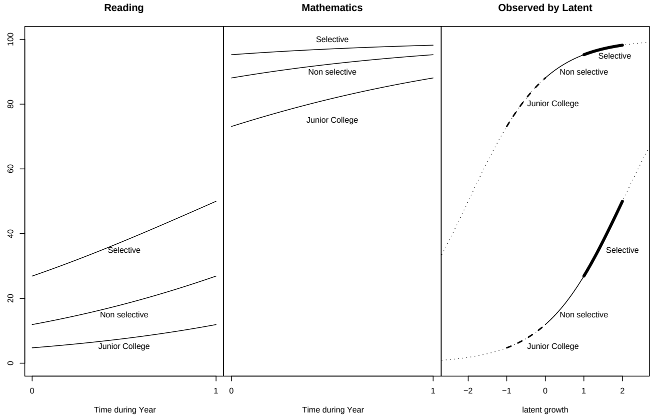
Theory and Theory Testing
Correlational Observations: Directional causality?



Theory and Theory Testing
Correlational Observations–The third variable problem



The effect of scaling upon the latent variable - observed variable relationship



Two classes of validity

1. Internal Validity: Is systematic error (bias) minimized
   - Have we controlled for confounds?
2. External Validity: Does the study actually study what is reported?
   - Will the results generalize?

Internal Validity

1. Are the results of the experiment/study due to the variables considered
   - What are the constructs
   - Do the measured variables measures those constructs
2. Are confounding variables controlled for?
   - What alternative explanations for the effect of the variables can you come up with
   - How do you control for them

Major Threats to internal validity

1. Within subject experiments
   - Fatigue
   - Practice
   - Boredom
   - Order effect
2. Between subject experiments
   - Subject differences
   - Many ways subjects can differ

Controlling for threats to internal validity within subjects

1. What are the obvious sources of error, and how to control them?
2. Order effects may be controlled by counterbalancing
   - But some order effects need long delays between trials (e.g., drug studies)

Practical problems and threats to internal validity

1. Manipulations affect more than the construct of interest
    - caffeine induces alertness and motor tremor
    - failure induces anxiety, depression, anger
    - practice leads to motivational changes as well as changes in skill
2. Observable variables reflect more than the construct of interest
    - self report of alertness reflects base line differences
    - cognitive performance–ability, motivation, training, practice
    - slowness of responding reflects caution as well as process speed

External Validity: Does the study actually study what is reported?

1. Do the effects generalize across other subjects
    - Are the effects true only for the type of subjects studied?
    - The case of WEIRD subjects?
2. Do the effects generalize across other conditions?
    - Are the effects true only for the specific situation studied?
    - How limited is the domain of generalization?
    - How do you know?

The many forms of the correlation coefficient

Table: A number of correlations are Pearson r in different forms, or with particular assumptions. If $r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$, then depending upon the type of data being analyzed, a variety of correlations are found.

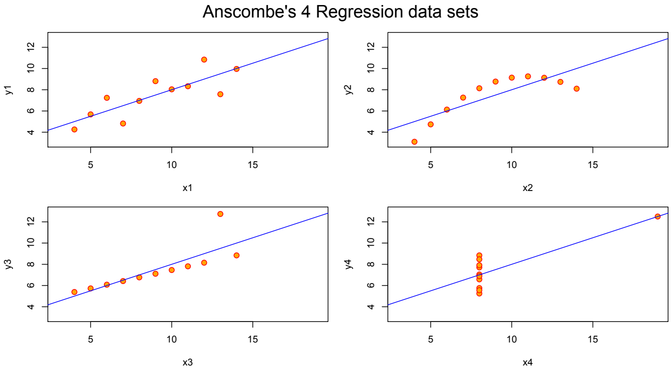
Coefficient	symbol	X	Y	Assumptions
Pearson	r	continuous	continuous	
Spearman	rho (ρ)	ranks	ranks	
Point bi-serial	r_{pb}	dichotomous	continuous	
Phi	ϕ	dichotomous	dichotomous	
Bi-serial	r_{bis}	dichotomous	continuous	normality
Tetrachoric	r_{tet}	dichotomous	dichotomous	bivariate normality
Polychoric	r_{pc}	categorical	categorical	bivariate normality

Many other statistics (F, d, b) may be expressed in terms of r

Table: Alternative Estimates of effect size. Using the correlation as a scale free estimate of effect size allows for combining experimental and correlational data in a metric that is directly interpretable as the effect of a standardized unit change in x leads to r change in standardized y.

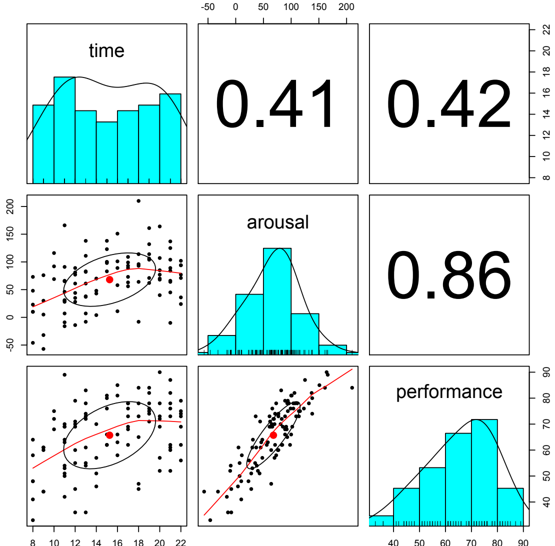
Statistic	Estimate	r equivalent	as a function of r
Pearson correlation	$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$	r_{xy}	
Regression	$b_{y.x} = \frac{Cxy}{\sigma_x^2}$	$r = b_{y.x} \frac{\sigma_y}{\sigma_x}$	$b_{y.x} = r \frac{\sigma_x}{\sigma_y}$
Cohen's d	$d = \frac{X_1 - X_2}{\sigma_x}$	$r = \frac{d}{\sqrt{d^2+4}}$	$d = \frac{2r}{\sqrt{1-r^2}}$
Hedge's g	$g = \frac{X_1 - X_2}{s_x}$	$r = \frac{g}{\sqrt{g^2+4(df/N)}}$	$g = \frac{2r\sqrt{df/N}}{\sqrt{1-r^2}}$
t - test	$t = \frac{d\sqrt{df}}{2}$	$r = \sqrt{t^2/(t^2+df)}$	$t = \sqrt{\frac{r^2 df}{1-r^2}}$
F-test	$F = \frac{d^2 df}{4}$	$r = \sqrt{F/(F+df)}$	$F = \frac{r^2 df}{1-r^2}$
Chi Square		$r = \sqrt{\chi^2/n}$	$\chi^2 = r^2 n$
Odds ratio	$d = \frac{ln(OR)}{1.81}$	$r = \frac{ln(OR)}{1.81\sqrt{(ln(OR)/1.81)^2+4}}$	$ln(OR) = \frac{3.62r}{\sqrt{1-r^2}}$
$r_{equivalent}$	r with probability p	$r = r_{equivalent}$	

Always plot your data!

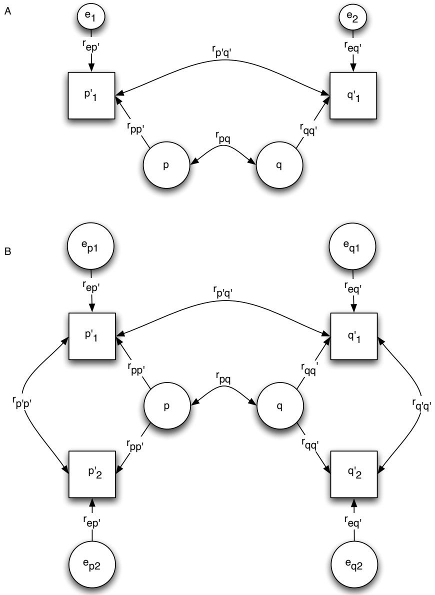


A Scatter Plot Matrix (SPLOM) plot

pairs.panels(my.data[c(4,7,9]) #include selected variables



Spearman's parallell test theory

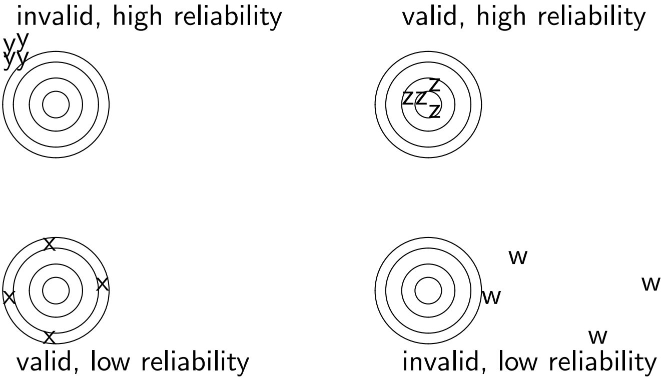


Types of reliability

- Internal consistency
  - α
  - $\omega_{hierarchical}$
  - ω_{total}
  - β
- Intraclass
- Agreement
- Test-retest, alternate form
- Generalizability

- Internal consistency
  - `alpha`, `score.items`
  - `omega`
  - `iclust`
- `icc`
- `wkappa`, `cohen.kappa`
- `cor`
- `aov`

Reliability and Validity



Statistical Description and Inference

1. Sampling theory: generalizing from a sample to a population
   - Sample estimates have sampling error
   - Need to consider both the sample estimate and the error of the estimate
2. Comparisons of group differences reflect real difference and sampling error
3. Data = Theory + Residual
4. Observed Variance = Explained Variance + Residual Variance

Generalization of results and threats to external validity-I

1. Limitations of generalization for subjects
2. Limits of generalization for conditions -interactions with other variables
3. limits of generalization for conditions – interactions with other variables
   - problems and benefits of interactions xy relationship depends upon z
   - interactions limit generalization
   - interactions test theoretical limits

Types of relationships and research designs

1. One Factor designs
   - Linear, monotonic, with or without inflection
   - Non-monotonic effects
2. Two factor designs
   - Simple additive effects
   - Additive and ordinal interactions
   - Disordinal interactions

Experimental Designs

1. Within Subjects
   - Controls for subject variability
   - Sensitive to within subject changes such as fatigue, learning, differential transfer
2. Between subjects
   - Controls for within subject changes
   - Sensitive to between subject variability
   - Effects due to subject selection, attrition, randomization
3. Mixed designs

Between Subject designs

1. Subject variables as threat to external validity
2. Confounded effects that can lead to subject variability
3. Randomization as a control
   - but does not guarantee control in any one study, just guarantees that the confounded variables have zero correlation in the long run.

Between Subject designs

Subject variables as threat to external validity

1. Ability
2. Practice
3. Motivation
4. Interest
5. Gender
6. Age
7. Culture

Between Subject designs

Confounded effects that can lead to subject variability

1. Time of day
   - Naturally occurring rhythms of alertness
   - Classroom effects
   - Fatigue
2. Time of week, month, season, year
3. Class schedules
   - Mid terms
   - Papers
   - Weather
4. Volunteer effects
5. Experimenter-Subject interactions

Randomization as a control

Only the *expected* values of groups are equal not the *observed* values

1. In any particular experiment, groups are not equivalent
2. Expected value of the (signed) group difference=0
3. Randomization does not introduce systematic bias

Types of Randomization

1. Subjects matched on variable of interest and then assigned to condition
2. Blocking to control for order effects
   - Ignores stable subject effects
   - Eliminates subject effects associated with time of appearance
3. Complete randomization
   - "failures" of randomization
   - Problems at the end of the experiment
4. Power is maximized with equal cell sizes
5. Randomization will tend not to produce equal size groups

Overview of the problem of inference

1. Theoretical problem: understanding the relationship between latent variables (constructs)
   - relationships among latent variables
   - relationships between latent variables and observed variables
2. Generalization of results and threats to external validity
3. Proper design maximizes internal validity

Types of Measures

1. Direct
   - Self report measures of desires, beliefs, knowledge
   - Peer and other ratings of behavior
2. Indirect
   - Reaction time as measures of:
     - implicit attitudes
     - cognitive availability
   - Psychophysiological measures of processing
     - EEG, MRI, SPEC, SC, HR, BP, etc.
3. Unobtrusive
   - Archival
   - Observational

Steps in correlational inference

1. Estimate the reliability of the variables
    - Magnitude of correlation is influenced by the reliability of the correlation
    - Varieties of reliability
    - (can you measure the same thing twice?)
2. Estimate the construct validity of the measures
    - Are you measuring what you think you are measuring?
    - Convergent, Discriminant, Incremental validity
3. Consider alternative explanatory variables

Pitfalls in Research

1. Investigator Effects
   - Investigator Paradigm effect
   - Investigator design effects
   - Investigator loose procedures effects
   - Investigator analysis effects
   - Investigator fraud effects
2. Experimenter Effects
   - Experimenter characteristics effects
   - Experimenter procedural effects
   - Experimenter data recording effects
   - Experimenter expectancy effects
   - Experimenter fraud effects
3. Recommendations
   - Tighten theory, design, execution
   - Consider statistical interpretation

Correlation and inverse probabilities

1. Does observing that B almost always happens when we do A imply that doing A almost always leads to B?
2. Examples:

Table: Examples of inverse probability problem

Observe	Cause ?
Auto Accidents	Drinking alcohol
Lung Cancer	Smoking
Pregnancy	Intercourse

3. Although strong association in one direction, how strong is the association in the other direction?
4. We need to know the base rates as well as the one cell

Correlation and inverse probabilities

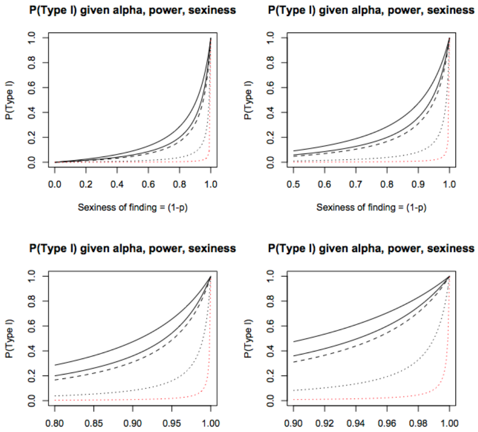
1. If one has disease B, then one tests A+ with p=.99
2. If one tests A-, then one has disease B with probability .01
3. 99% of people do not have the disease
4. If one tests A+, what the probability that they have disease B?

Science and error

1. Type 1 errors can happen to you (or me)!
   - Experiment wide error rate is a function of the number of tests run $=1-(1-\alpha)^n$
   - Bonferoni correction sets experiment wide error rate by using a correction for the number of tests $=\alpha/n$
   - This is somewhat conservative but better than pretending that type 1 errors don?t happen
   - Holm correction slightly more powerful
2. Type 2 errors happen due to lack of power
   - If the study is too small, important effects will probably not be detected
3. Type 3 error: asking the wrong question

As the surprisingness of the study increases, so does type I error rate

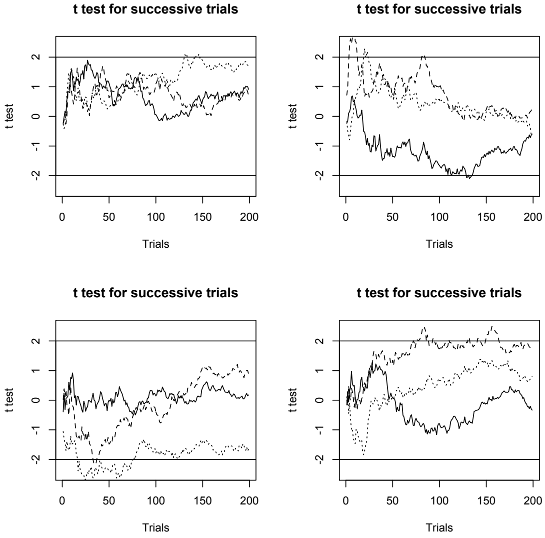
Type I Errors: It is not the power, it is the prior likelihood
dashed/dotted lines reflect alpha = .05, ,.01, .001 with power = 1



1. Extreme claims require extreme probabilities
2. Given that a finding is "significant", what is the likelihood that it is a Type I error?
3. Depends upon the prior likelihood (the 'sexiness') of the claim.

16 / 16

Trial by trial t-tests for true effect = 0 can lead to “significant” results



Always describe your data

R code

```
describe(my.data)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
snum	1	198	99.50	57.30	99.5	99.50	73.39	1	198	197	0.00	-1.22	4.07
sex	2	198	1.64	0.48	2.0	1.68	0.00	1	2	1	-0.59	-1.67	0.03
drug	3	198	0.52	0.50	1.0	0.52	0.00	0	1	1	-0.06	-2.01	0.04
time	4	198	15.02	4.46	15.0	15.01	5.93	8	22	14	0.02	-1.32	0.32
anxiety	5	198	4.00	1.99	4.0	3.98	1.48	-1	9	10	0.07	-0.30	0.14
impulsivity	6	198	4.90	2.04	5.0	4.91	1.48	0	10	10	-0.09	-0.43	0.14
arousal	7	198	63.18	42.93	63.5	64.47	43.00	-65	174	239	-0.25	-0.13	3.05
tension	8	198	23.67	12.95	21.5	22.46	12.60	4	61	57	0.74	-0.03	0.92
performance	9	198	63.40	10.96	65.0	64.06	10.38	33	86	53	-0.56	0.10	0.78
cost	10	198	1.00	0.00	1.0	1.00	0.00	1	1	0	NaN	NaN	0.00

Ethical Principals

1. Basic summary
   - Do no harm
   - Be honest
   - Be fair (would you be willing to switch roles with subject?)
2. Specific guidelines
   - American Psychological Association
   - National Academy of Sciences
   - Institutional Review Boards

Researching the literature

1. What has gone before - Science as an accumulation of knowledge
2. Original publications in peer reviewed journals
   - Produce new results based upon prior theory
   - Include references to prior work
3. Literature searching with database tools
   - Google Scholar

Measuring Individual Differences

1. The International Personality Item Pool by Lewis Goldberg
2. http://ipip.ori.org

International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences Mission Statement This IPIP Website is intended to provide rapid access to measures of individual differences, all in the public domain, to be developed conjointly among scientists worldwide. Later, the site may include raw data available for reanalysis; in addition, it should serve as a forum for the dissemination of psychometric ideas and research findings.*

Finding and labeling IPIP scales

At the link labeled "Multiple Constructs," there is a list of IPIP multi-scale inventories, including several based on either the lexically derived Big-Five factor structure or Costa and McCrae's Five-Factor Model. As is the case with most IPIP scales, these were developed by identifying IPIP items that, when summed into a scale, correlate highly with an existing measure.

The first three inventories, under the heading "The Big-Five Factor Structure," are IPIP measures designed to correlate with five-factor scales whose items are trait adjectives. The first of these is labeled "Big-Five 5 Broad Domains." If one follows its link labeled "Comparison Table," one will find descriptions of both 50- item and 100-item IPIP inventories designed to correlate highly with the five adjective markers described in the following article: Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological

Finding and labeling IPIP scales

Although none of the IPIP scales have official names, one should refer to them by the scales on which they are based. Hence, these scales could be labeled "the 50-item (or 100-item) IPIP representation of the Goldberg (1992) markers for the big-five factor structure" or something like that.

At the "Multiple Constructs" page, the next set of scales are labeled "Seven-Factor Scales" which refer to IPIP scales that were developed to measure the adjective scales constructed by Saucier (1997). His scales, which include the Big 5 plus Attractiveness and Negative Valence, are described in the link labeled "Comparison Table"

Surveys using Qualtrics

1. Qualtrics Research Suite (Qualtrics) is a cloud-based service which provides a comprehensive suite of tools to create web surveys and analyze collected data.
2. Qualtrics has been licensed by the College and unlimited access is available at no cost to all Weinberg faculty, staff, graduate students, and undergraduate students.
3. `http://www.weinberg.northwestern.edu/weinbergit/teaching-research/qualtrics/`

Using Qualtrics

1. What types of questions can I ask?
Qualtrics supports over 100 different question types, including:
Multiple choice Matrix table heat map Slider Text entry
2. Are there tools to help create a survey?
Qualtrics provides a library of surveys and answers that you can modify and add to your own survey library.
3. Can my library contain anything else?
Yes, your library can contain messages and media.
4. Does Qualtrics lock my survey into a template with rigid formatting requirements?
No. Qualtrics allows you to easily change the layout, position, and answer choices. You also have freedom to modify your text in a word processor-like interface. Additionally, you can create your own survey templates with a CSS editor.
5. Can I design surveys to only show some questions?

Analyzing data from Qualtrics

1. Qualtrics will export as a csv file (comma separated values)
2. This can be imported directly into R
3. How do to score scales http://personality-project.org/revelle/syllabi/205/scoring.pdf is a short tutorial for how to score the scales from the Qualtrics items.
4. Subsequent analyses are done as you did for your second paper.

Writing a scientific paper

1. Purpose is to add to the accumulated knowledge base
2. Reviews prior work
3. Methods are clearly stated so that others can replicate if they choose
4. Results are appropriately analyzed so that someone else would reach similar conclusions
5. Discussion links results to prior work and suggests future directions

Final Project

1. Conceptualize an interesting problem
2. Design a study to test a hypothesis concerning this problem
   - If doing a survey, consider items from the IPIP
   - If doing a survey, use Qualtrics
3. Execute the study following the design
4. Analyze appropriately
   - R is more userfriendly than you think
   - You already know much of what to do
   - WR and LC are available for help
5. Report in a scientific manner

Possibilities for further research

1. Advanced research courses
2. Summer research projects (see http://www.psychology.northwestern.edu/undergraduate/ and http://www.psychology.northwestern.edu/undergraduate/research/ ???
3. 397-399 independent studies
4. Honors research projects

Summer research

Each summer the Psychology Department offers two or more undergraduates a Benton J. Underwood Summer Research Fellowship. Students who accept these fellowships spend most of the summer working on research at Northwestern with a psychology professor. The exact schedule is worked out with the professor who supervises the research. Both current juniors and current sophomores can apply for this award; priority is given to current juniors. To apply for an Underwood Fellowship, follow these steps:
Choose a faculty member to supervise your research and talk with him or her about what you will be doing and what your time commitment will be. You should also talk with the faculty member about the need for Institutional Review Board approval for your planned project.

Summer research

Prepare an application in which you include (a) a statement describing your plans for this research (this can be the same proposal you submit to the university's grants committee); (b) a copy of your transcript (an unofficial transcript is fine); and (c) information about your general interests in psychology, your relevant course work, your previous research experience, and anything else that you think is relevant. Have the faculty member who will supervise your research write a confidential letter of support for your application.

The application and letter of support should be sent by email to Dr. Sara Broaders, s-broaders@northwestern.edu. The application deadline is always the same as the deadline for summer grants from Northwestern's Office of Undergraduate Research.

The psychology major at NU

1. Introductory Psychology
2. Methodology Sequence
   - 201: Statistics
   - 205: Research Methods
   - 3xx: Advanced research in a substantive area
3. Substantive courses
   - Personality/Social/Clinical ("Column A")
   - Cognitive/Physiological ("Column B')
   - Other broad courses ("Column C")
4. Advanced research courses ("Row 2")
5. Independent Study (399) and Honors (398)

The psychology major at NU