Correlations & Confounding Variables

Dewayne E Perry
ENS 623
Perry@ece.utexas.edu

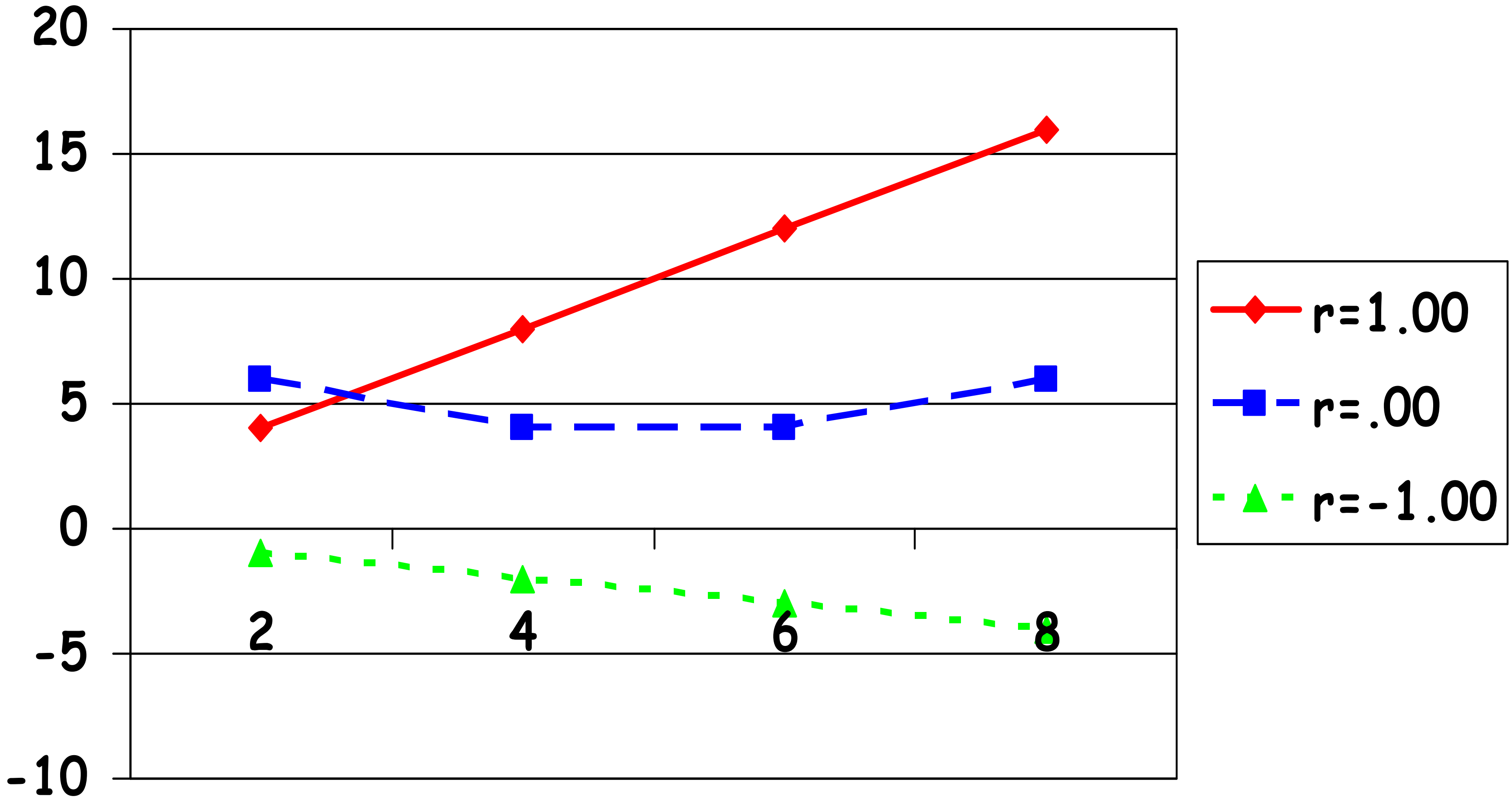
Pearson *r*

- Most widely used index of relationship
- Short for: *Karl Pearson's product moment correlation coefficient*
- Values ranges between -1.00 and +1.00
  - .00 means there is no relationship
  - +1.00 - a perfect positive linear relationship
  - -1.00 - a prefect negative linear relationship
- May be correlated even though scores do not agree

Pearson r

- Examples
  - (8, 6, 4, 2) and (16, 12, 8, 4) r = 1.00
  - (8, 6, 4, 2) and (6, 4, 4, 6) r = .00
  - (8, 6, 4, 2) and (-4, -3, -2, -1) r = -1.00
- Results are what one would expect if standard scored (Z-scored): $Z = X - \overline{X} \;/\; \sigma$
  - product moment correlation: $r_{xy} = \sum Z_x Z_y \;/N$
    - Z's are distances from mean called *moments*
    - multiplied by each other to form *products*

Pearson *r* Correlations



Interpretations

- Prime interpretation: the larger *r*, the higher the degree of linear relationship
- The square of r: the proportion of the variance shared by *X* and *Y*
  - Proportion of variance of *Y* scores attributable to variation in the *X* scores
  - $r^2 + k^2 = 1.00$
  - r^2 is the *coefficient of determination*
  - k^2 is the *coefficient of non-determinism*
  - Though useful, it is a poor reflection of the practical value of any given correlation
  - More useful in regression (discussed later)

Interpretations

- *r* as an indicator of practical importance
  - *Binomial effect-size display* (BESD) procedure
  - *Binomial* : research results cast as dichotomous
  - Introduced because
    - Interpretation is quite transparent
    - Applicable whenever *r* is used
    - Very conveniently computed
  - BESD question: what is the effect on the *success rate* of the new treatment
    - Displays the change attributable to treatment
    - Converts effect size *r* into a success rate via table lookup (RR Table 14.6)
      - ✓ *r=.30* , accounting for *9%* of the variance
      - ✓ shows an increase in the *success rate* from 35% to 65%
    - Short form: *r x 100* = percentage increase of success
    - [Insight based on 50-50 probability of treatment effect]
  - More clearly shows real-world importance of treatment than effect size estimates

Small but Important

- While effect size may be small, the practical importance may be large
  - May have important social, psychological or biological effects
- Another way to compute *r* (or *phi*)

$$r = \frac{\text{difference between cross products}}{\sqrt{\text{product of all marginal totals}}}$$

- Examples
  - Vietnam versus non-Vietnam veterans, 50% more likely to have an alcohol problem, *r=.0698*
  - Vietnam veterans about twice as likely to suffer depression as non-Vietnam, *r=.0597*
- Small effects. But can reflect effects of enormous consequence
  - Aspirin and heart attacks: *r=.0337*
  - But this translates in a significant number of lives

Spearman Rank Correlation

- ρ sometimes used as a quick index of correlation
  - Easy and painless to compute
  - Consider the following example (D is difference in rank)

X	Y	rank X	rank y	D	D-squared
6.8	79.713	2	1	1	1
12.2	47.691	1	2	-1	1
1.7	28.002	3	3	0	0
0.3	11.778	4	4	0	0

$$\rho = 1 - 6(2) / 4^3 - 4 = 1 - 12/60 = .80$$

- Nothing sacrosanct in scale used
  - Reduces skewedness
  - Choose for symmetry, lack of skewedness
  - Tends to increase accuracy of analysis
    - Sometimes leads to slightly higher *r*
    - Sometimes to lower
      - ✓ case of logarithmic transformations: .80 instead if .99 (RR 14.11)

Spearman Rank Correlation

- Most useful correlations are *product moment correlations*
- When data in rank form, apply *Spearman rho*
  - But nothing more than *Pearson r* computed on numbers that happen to be ranks
  - Ranks are more predictable
  - New ingredient: D – the difference between the ranks assigned to each pair of sampling units
- $\rho = 1 - 6\sum D^2 / N^3 - N$

Point Biserial Correlation

- Special case of product moment correlation *r*
  - One variable continuous,
  - One dichotomous,
    - with arbitrarily applied numeric values
    - Such as 0 and 1, or –1 and +1
- Example: M vs F on verbal skills
  - M=2,3,3,4 vs F=4,5,5,6
  - *X* is implicit in M/F, *Y* is explicit
  - Encode gender as 0,1
  - Y mean = 4, X mean = 0.5
  - X1 mean = 5, X2 mean = 3

Point Biserial Correlation

- $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\left(1/n_1 + 1/n_2\right)S^2\text{pooled}\right)}} = \dfrac{5-3}{\sqrt{\left(\left(1/4+1/4\right)\right)0.6667}} = 3.46$
- **Which at 6*df* is significant at *p* < .01, one-tailed test**
- ***r* = .816**
- **Significance test = size of effect X size of study**
  - Index for size of study varies with index of effect size:
    - Eg, *N*, *df*, square root of *N* or *df*
  - As either increases, significance test score increases
    - $$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{df}$$
  - First term is proportion of variance explained by r to the proportion not explained by 4 – ie signal to noise ratio

Phi Coefficient

- Another special case of the product moment correlation *r*
  - Both variables are dichotomous
  - Arbitrarily applied numeric values 0,1 or +-1
- Example – Dem/Rep answer Yes/No
  - D: 1Y, 4N vs R: 4Y, 1N
  - *r* = .60 for party membership and answer
  - If sample size not too small (*N* >20) and both variables are not too far from 50-50 split (no greater than 75/25), can use *t* test for significance
  - *t* = 2.12 which is *p* = .034, one-tailed
  - More common is chi-square test for significance of phi

$$\chi^2(1) = \phi^2 \times N$$

  - since *phi* = .60 and *N* = 10, *chi-square* = 3.60
  - which is significant at the .058 level

Curvilinear Correlation

- Sometimes predictions are not linear but curvilinear (quadratic - U shaped)
  - ★ higher (U shaped) or lower levels (upside down U) at ends
  - ★ eg, extreme levels of arousal associated with great/poor performance

5 Product-Moment Correlations

- Pearson *r*
  - both variables continuous
  - *t* test for significance
- Spearman rho
  - both variables ranked
  - *t* test for significance
  - or exact probability test if *N* is small ($N < 7$)
- Point biserial (*r-pb*)
  - one continuous, one dichotomous
  - *t* test for significance
- *Phi*
  - both variables dichotomous
  - *chi-square*, *t* and *Z* tests
- Curvilinear *r*
  - both continuous
  - *t* test

Comparing Correlations

- Primary question
  - often not so much about relationship
  - but about difference in such relationships
  - comparison of independent correlation coefficients
    - based on different independent subjects
  - comparison of non-independent correlation coefficients
    - based on the same subjects

Sampling Information

- Must describe sample sufficiently
  - To judge representativeness
  - To evaluate equivalence of different groups of participants
  - To assess whether participant variables have been controlled
  - Enough details to compare with other studies
- Representativeness
  - In many cases simply assumed
    - Eg, the populations fir the major categories
    - Assumed sample in Chicago behaves the same as sample in NYC or London
  - In surveys, representativeness critical
    - Eg, Roosevelt/Landon election
      - ✓ Predicted for Landon
      - ✓ Huge sample for prediction
      - ✓ BUT from car/telephone owners
      - ✓ Biased towards conservative and hence Landon

Non-probability Sampling

- Self selected sampling
  - Eg, in media: open invitations to respond to questions
  - Sampling limited to those who saw the request
    - Eg, people with computers
  - Unclear whom any the self-selected surveys represent
  - Slightly different: consumer's reports
    - Select population
    - Self-selected within that
- Haphazard sample
  - Recruiting in public space – eg, airports, malls
  - Difficult to replicate
  - Danger of biased samples
    - People who travel airlines don't go to laundromats, etc
    - But some topics don't make a difference: optical illusions
    - For social attitudes, bias could be pivotal
  - *NOT* random sampling

Non-probability Sampling

- Sample of opportunity
  - Convenience samples
    - Pool of participants who are available
      - Eg, this class
    - For the psychologist might be ok
  - The more the dependent variables are associated with variables other than the independent variables the more crucial representativeness becomes
  - A variation: participants from a hospital, school, clinic, project etc
    - May have more than enough people to meet requirements
    - And random selection from population
    - May need to know how representative population is

Non-probability Sampling

- Homogeneous, restricted, purposive sample
  - A specific subset of a convenience group
    - Can restrict homogeneity wrt feature shared
    - Eg, personal adjustment patterns in freshmen
    - But expected to represent all freshmen, not just Princeton or exclusively Caucasians
  - Eg, bias effects if too restrictive a sample
    - SAT national average: 906; Mississippi: 1001; NJ: 889
    - Only 3% take it in Mississippi; 65% in NJ
- Networking (snowballing) sample
  - Ask for references if do not have enough samples
  - Eg, networks of mothers who have small children
  - Can suffer from inbreeding, ie, too homogeneous
  - Vulnerable to contamination of the results if participants talk with each other about the experiment
- Systematic sample
  - Eg, first 50 people thru the door
  - But early arrivals may differ from latecomers in systematic ways

Probability Sampling

- Random sampling
  - Each member has an equal chance of being selected
  - large population often beyond scope of most researchers
  - From more limited populations is possible, but have to be careful
- Systematic sampling
  - Often hoped that systematic methods are unbiased and equivalent to random
    - Have to be careful that no bias introduced
  - Eg, every third person on list

Probability Sampling

- Stratified sampling
  - Usually limited opportunities, unless well-funded
  - Costly to represent/match target populations on all demographic and other variables
  - Physical scientist can assume 1 oz of silver representative worldwide
  - By psychologist studying female depression has to worry about a host of demographic and personal variables
- Cluster sample
  - Randomly target clusters of people
    - Eg, students in schools in a city
    - Then randomly within schools

Sample Sources

- Direct samples
  - Obtain data directly from people in sample
  - Experimental, quasi- and non-experimental commonly obtain data this way
- Archival samples
  - Use data already gathered and are a matter of record
    - Eg, actuarial records such as vital statistics, medical records, etc
  - Experimenter bias can have no influence on them
  - Disadvantages: forced to rely on accuracy and timeliness of data
    - Eg, income of 20 years ago not very useful today
    - On test records, have no control over qualifications of examiner
    - Or accuracy of scoring, administration, or interpretation
  - Advantages: not contaminated by the experiment – more than balances
  - Other disadvantages
    - Selective deposit, survival
    - Selective entry factors may distort

Sample Biases

- Volunteers
  - Accept or decline may bias sample
  - May not be representative of population
    - Unrepresentative characteristics
    - Could threaten generalizability
  - Ethical issues
    - Need sufficient info for informed decision
    - Foreknowledge can cause problems
  - Levels of volunteering
    - Anonymous opinion of social issue
    - Participation where no noxious effects
    - Out of their way, extra time, some discomfort
  - Declining may be a function of commitment
  - Recompense could effect level as well
  - Greater the sacrifice, fewer volunteers
  - Cannot know characteristics of non-volunteers

Sample Biases

- Rosenthal & Rosnow 1991
  - Maximum confidence: tend to be better educated, higher in social class, more intelligent, more approval motivated, and more sociable
  - Considerable confidence: see arousal, be unconventional, female, Jewish, non-authoritarian, nonconforming
  - Some confidence:from smaller towns, interested in religion, more altruistic, more self-disclosing, more maladjusted, younger

Sample Biases

- Biasing by selective attrition
  - At beginning sample may be representative
  - Attrition may cause non-representative
    - Dropouts are mainly women, middle aged, poor, etc
  - Obligation to exclude those no longer willing or able causes problems
    - What about people who do not respond to treatment
    - Eg, people who are not stressed by stress condition
  - Have we lost randomness as a result?
  - Rationale for exclusion should be made clear

Assignment

- Random
  - Selection brings into study, assignment places them in treatment
  - Does not solve problem of non-equivalent groups
  - Randomness by random number table simplest
- Systematic
  - Potential for bias always present
  - May have confounding variable present
  - Must convince two groups are equivalent

Assignment

- Sample size
  - As many as possible dwindles quickly to as many as feasible
  - Significant results can be obtained with 20-30 participants, 10 per treatment, provided
    - Distributions are reasonably normal
    - Statistical assumptions are met
  - Inconclusive results: sample size problem?
  - Large sample -> small differences could be significant
  - Power analysis: increase power by
    - Raising level of significance required
    - Reducing standard deviation
    - Increasing magnitude of effect by using strong treatments
    - Increasing the size of the sample
  - $p < .05$ usual desired level of significance
    - Depends on study context – not sufficient for life-threatening