Predicting Relations between Variables

http://philosophy.ucsd.edu/faculty/wuthrich/

12 Scientific Reasoning

Acknowledgements: Bill Bechtel

Where we have been

- Variables: things that vary
- Variables differ in the way they can be measured
- Types: nominal and ordinal vs. score variables
- Hypotheses: conjectures about the way some phenomenon behaves
  - $\Rightarrow$ need to be tested by predictions
- Hypotheses about how a single variable is distributed
  - Ages of people in food court
  - Amount of time animal spends foraging
  - Amount of time you spend studying

Hypotheses involving more than one variable

Many of the hypotheses of interest in science and in ordinary life involve relations between variables

- Amount of sleep and ability to recall
- Pressure, volume, and temperature of a gas
- SAT score and job performance
- Vitamin intake and health condition
- Sexual activity and sexually transmitted diseases
- Smoking and lung cancer
- Miles per gallon and horsepower of cars

The case against bread

- More than 98% of convicted felons are bread eaters.
- Fully half of all children who grow up in bread consuming households score below average on standardized tests.
- In the 18th century, when virtually all bread was baked in the home, the average life expectancy was less than 50 years.
- More than 90% of all violent crimes are committed within 24 hours of eating bread.
- Primitive tribal societies that have no bread exhibit a low incidence of cancer, Alzheimer’s, and Parkinson’s disease.

Ask yourself: are the statistics meaningful!

Correlational and causal hypotheses

Many of the hypotheses of interest in science and in ordinary life are causal:

- Red wine and (reduced) cholesterol levels
- Smoking and lung cancer
- Vitamin C and prevention of colds
- Fiscal policy and economic recovery
- Use of imagery and increased memory

But causal claims are more difficult to establish than correlational claims

$\Rightarrow$ so for now we focus on correlations only

Correlations and why they are interesting

- A correlational claim is a claim that the values on two variables vary systematically (though not necessarily in the same direction)
- Why care about correlations if they are not (known to be) causal?
  - They may indicate causal relations.
  - They can be used to make predictions about the value of one variable from the known value of another variable.

SAT and college grades

- Should the SAT be used as a (or maybe the) basis for admission to the University of California?
- If so, then it must be justified:
  - Does it predict success in college?
  - If it doesn’t, then it may be an inappropriate measure to use in judging admissions.
- Compare: basing admission to UC on
  - running speed for the mile
  - length of one’s index finger

Correlational research examples

- Is there a relationship between family income and grade point average?
- Is there a relationship between number of hours of part time employment and grade point average?
- Is there a relationship between education and income later in life?

From the general to the testable

Not all hypotheses relating variables are directly testable—hypotheses presented in general terms:

- Force is equivalent to mass times acceleration
- Entropy always increases
- Fitter people live longer
- Better education correlates with greater happiness
- Greater pollution corresponds to greater global warming
- Animals living in colder climates are larger

Testable predictions

- To test hypotheses such as these, we need to make specific predictions.
- Predictions which can be evaluated: must predict something we can detect and measure, either with our senses directly or via instruments.

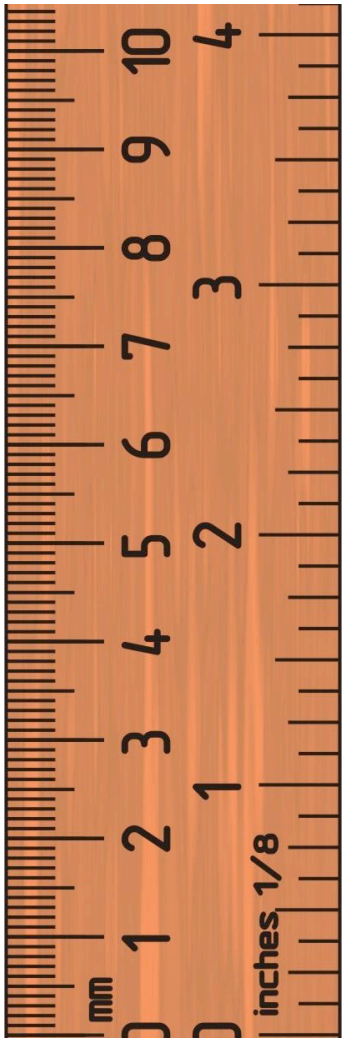
Operational ‘definitions’

- relate the variables used in the hypothesis to measurable variables
- variables such as force, entropy, memory ability, happiness, etc., are not directly measurable (observable)

$\Rightarrow$ We must specify a measurement procedure.

- The operational definitions of any non-observational terms are major auxiliary assumptions in any test of a hypothesis.

An example: distance



Inch: width of a grown man’s thumb; King Edward II (C14): the length of an inch shall be equal to three grains of barley, dry and round, placed end to end lengthwise

Foot: the name gives away its original reference; standardized to 12 inches

Yard: the length of a person’s belt; King Henry I (C13): distance from his nose to the thumb of his outstretched arm, which came to about 36 inches

The meter standard

Platinum-Iridium meter bar

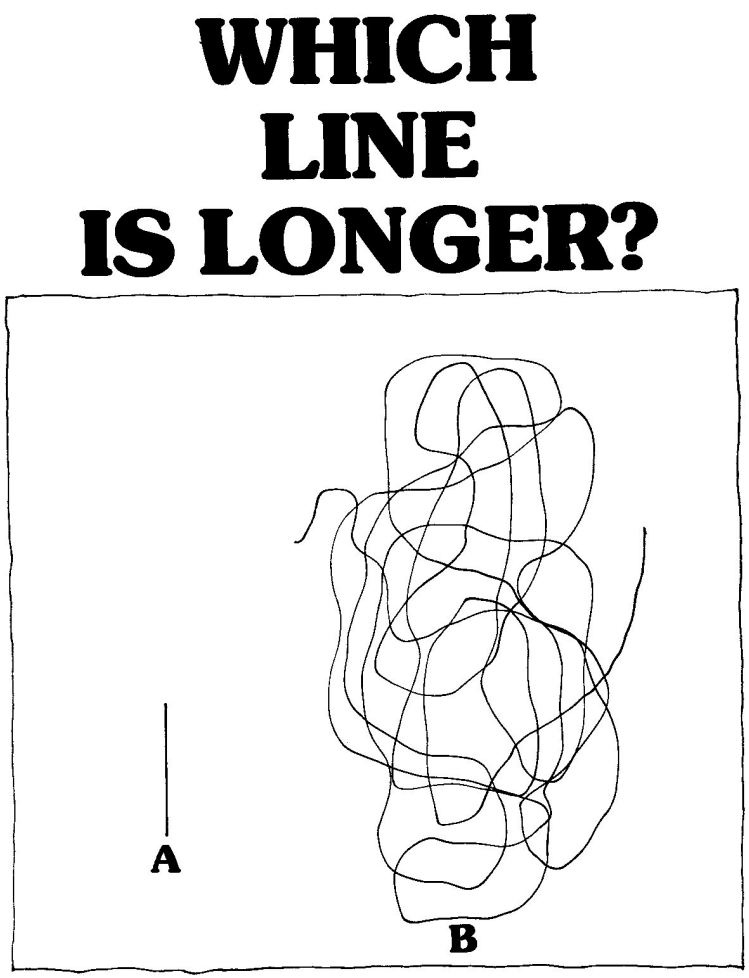
- Meter introduced by the French in 1791 as one ten-millionth of the distance from the equator to the north pole along a meridian through Paris
- Standard meter bar: A platinum bar with a rectangular cross section and polished parallel ends at a specific temperature
- 1859: J.C. Maxwell defined it in terms of the wavelength of the yellow spectral line of sodium
- 1892: A.A. Michelson 1,553,164.13 times the wavelength of cadmium red in air, at 760 mm of atmospheric pressure at 15°C.

- 1960: 1,650,763.73 vacuum wavelengths of light resulting from orange-red light, in a vacuum, produced by burning the element krypton (Kr-86)
- 1983: length of the path traveled by light in vacuum during 1/299,792,458 of a second

The case of IQ

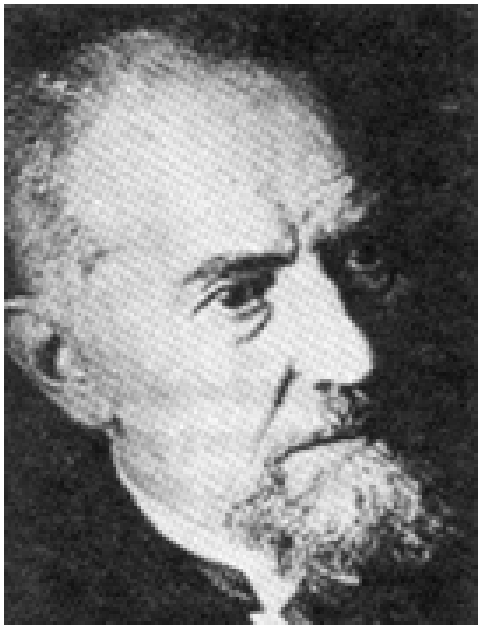
- In 1904 Alfred Binet (1857-1911) was commissioned by the French government to devise a test to differentiate children who would not do well in usual schools from those who would.
- The latter were to be assigned to special schools with greater individual attention but not disrupt the intellectually normal children.

It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which, is of the utmost importance for practical life. This faculty is judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. A person may be a moron or an imbecile if he is lacking in judgment; but with good judgment he can never be either. Indeed the rest of the intellectual faculties seem of little importance in comparison with judgment. (Binet and Simon (1916 [1973]): 42-43)



30 item test, with different questions typically solved by children at different ages:

- attend to simple instructions
- name parts of the body
- compare lengths and weights
- count coins
- assess which of several faces is ‘prettier’
- name objects in a picture
- remember digits from a list
- define words
- fill in the missing words in sentences



- Intelligence Quotient (William Stern, 1914): age level of test results/chronological age, normed to 100
- an 8 year old who passes the 10 year-old's test would have an IQ of 10/8 x 100

Lewis Terman at Stanford, 1916:

- Items on Binet's test didn't predict well in California, so revised the test $\Rightarrow$ Stanford-Binet
- Extended test to adults
- Introduced 'IQ'
- Studied, and promoted, 1000+ of those who scored high
- Advocated forced sterilization of 'feebleminded'

IQ joins the Army

- Entering World War I, the US army needed to assess the intelligence of recruits in order to assign them appropriately.
- The Stanford-Binet test required a skilled test administrator, which the army could not afford.
- Robert Yerkes et al. created a paper and pencil version.

But what does IQ measure?

Tests designed to predict success in specific functions.

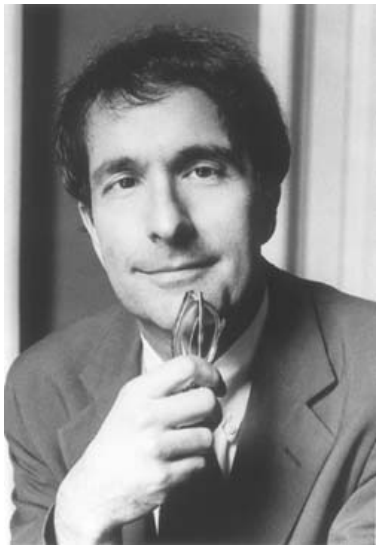
Charles Spearman:

- People who did well on IQ tests tended to do well in other intellectual activities.
- People who did poorly on IQ tests tended to do poorly on other intellectual activities.

$\Rightarrow$ General ability factor (g) + specific abilities

- Advocated that voting and procreation be restricted to those exceeding a base value for g.

Beyond IQ?



Howard Gardner: multiple intelligences

- verbal
- mathematical
- musical
- spatial
- kinaesthetic
- interpersonal (social skills)
- intrapersonal (self-understanding)

Robert Sternberg: three kinds of intelligence

- academic
- practical
- creative

Changing IQ

- In the early 1980s James Flynn, a New Zealand political scientist, discovered that the IQ in various groups of people increased on average 3 points per 10 years.
- Effect replicated throughout the industrialized world.
- Hidden by the fact that IQ tests are regularly renormed to keep the mean at 100.
- Increases apparently not linked to learning—greatest increases in non-verbal tests of intelligence.
- Explanation???

Construct validity

Question

Does the way you operationalized a variable really capture that variable?

- Does a ruler (grains of barley) really measure height?
- Does an intelligence test measure intelligence?
- Does a word-list test measure memory?

Operational definitions are not definitions

- An operational definition provides one way to measure a variable:
  - There will typically be alternatives.
  - The alternatives may not always agree.
- Even when construct validity is high, the operational definition does not provide necessary and sufficient conditions for the term.

Relating score variables

- same items measured on two score variables
- Is there any systematic relation between the score on one variable and the score on another?

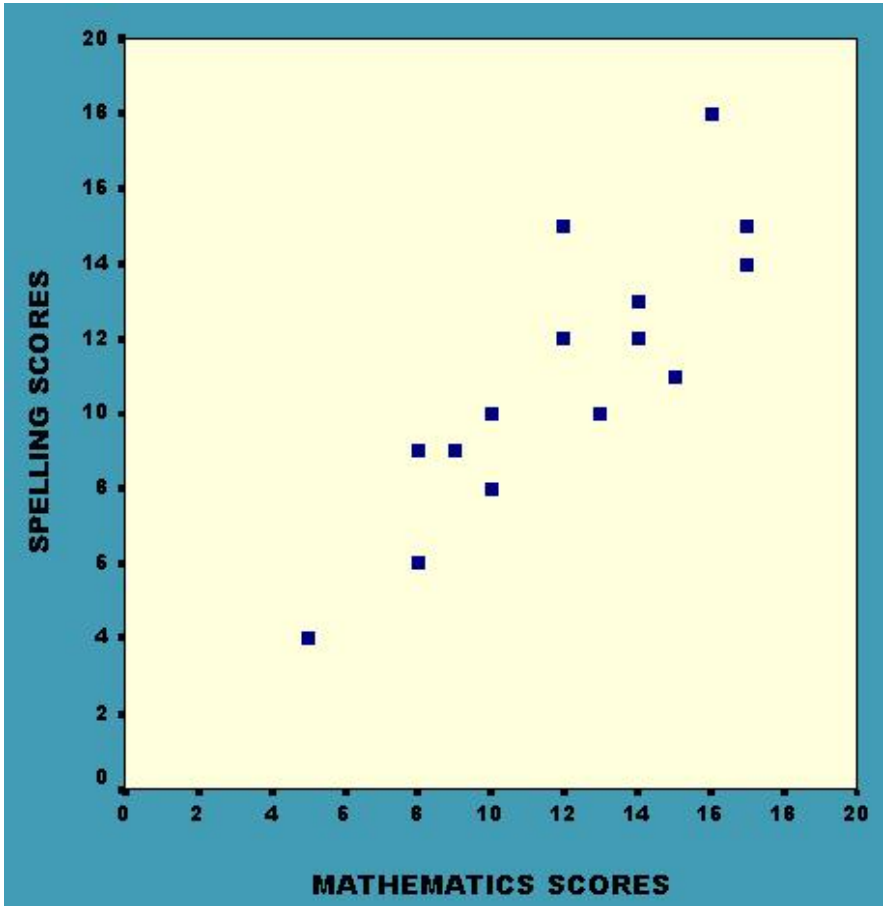
Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Spelling	15	14	15	12	6	4	8	9	9	12	18	13	10	10	11
Math	12	17	17	12	8	5	10	9	8	14	16	14	10	13	15

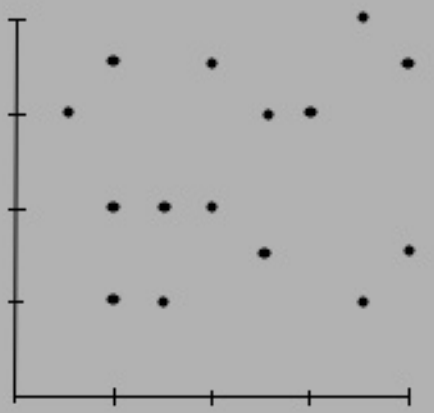
- Often it is difficult to determine if there is a regular pattern by just looking at scores (eyeballing the data).

$\Rightarrow$ important to graph or diagram the data

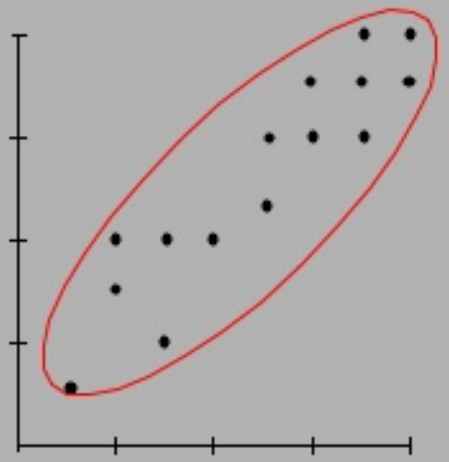
Scatterplots

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Spelling	15	14	15	12	6	4	8	9	9	12	18	13	10	10	11
Math	12	17	17	12	8	5	10	9	8	14	16	14	10	13	15

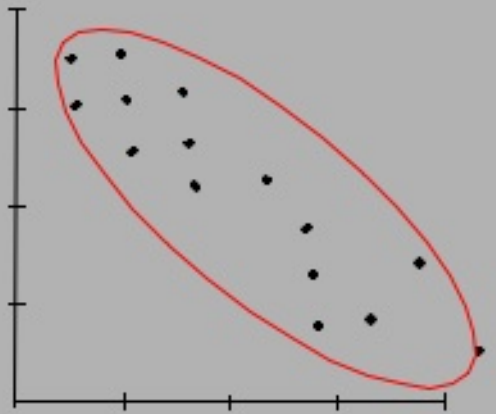




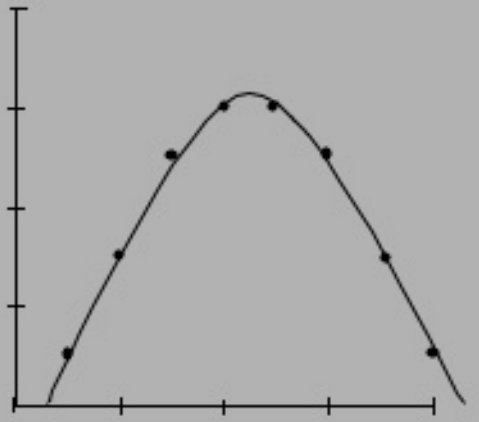
No correlation



Positive correlation



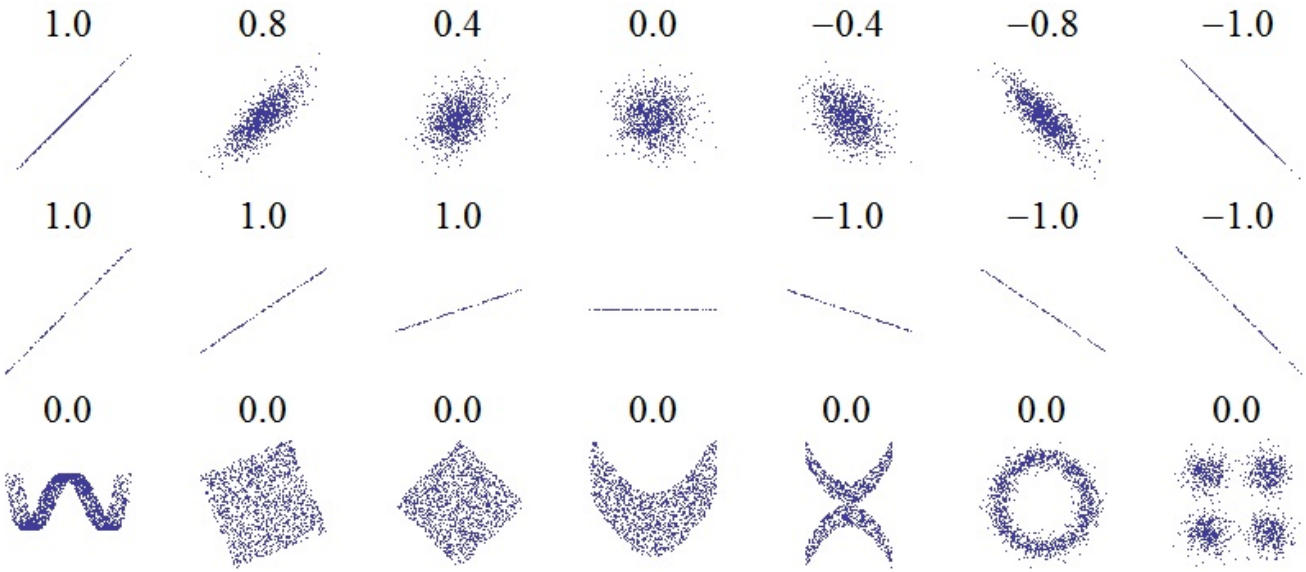
Negative correlation



Nonlinear correlation

Measuring correlation

- Karl Pearson (1857-1936), British statistician, developed a measure of correlation between two score varaibles (typically denoted X and Y), known as Pearson's Product-Moment Correlation (Coefficient), denoted r.
- $r = -1.0$ means perfect negative correlation, $r = 0$ no correlation and $r = 1.0$ perfect positive correlation.



r and Z

A Z-score for an individual is how many standard deviations that individual's score is from the mean of the population. Pearson's r can easily be calculated from the Z-score:

$$r = \frac{\sum_i Z_i^X Z_i^Y}{N}.$$

That the calculation is easier can be seen from the original formula necessary to calculate r directly from the scores:

$$r = \frac{\sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{N}}{\sqrt{\left(\sum_i X_i^2 - \frac{(\sum_i X_i)^2}{N}\right)\left(\sum_i Y_i^2 - \frac{(\sum_i Y_i)^2}{N}\right)}}.$$

Pearson correlation coefficient

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Spelling	15	14	15	12	6	4	8	9	9	12	18	13	10	10	11
Math	12	17	17	12	8	5	10	9	8	14	16	14	10	13	15

- Pearson's product-moment correlation $r = 0.857$.
- Note: positive value, very high

$\Rightarrow$ strong positive correlation

Measuring correlations

- Statistics was developed by real people
- Karl Pearson: reports that the first thing he could remember was sitting in a high chair sucking his thumb. Someone told him if he did not stop sucking it, the thumb would wither away. He examined his two thumbs carefully and concluded: "They look alike to me. I can't see that the thumb I suck is any smaller than the other. I wonder if she could be lying to me."

Karl Pearson

- Founded biometry: analyzing relations between variables characteristic of biological species
  - Traits of parents and offspring (including personality traits—temper, vivacity, assertiveness)
- Correlations between traits within an organism
  - Can you predict an individual's body-weight from knowing their arm length? Their temper from knowing their weight?
- Correlations between parents and offspring
  - Can you predict the eye color of offspring from the eye color of the parents?
  - Can you predict the IQ of offspring from knowing the IQ of their parents?

How (not?) to treat an up and coming star:

> *...Fisher...received an offer from Professor Pearson at the Galton Laboratory. Fisher's interests had always been in the very subjects that were of interest at the Galton Laboratory, and for five years he had been in communication with Pearson, yet during those years he had been rather consistently snubbed. Now Pearson made him an offer on terms which would constrain him to teach and to publish only what Pearson approved. It seems that the lover had at last been admitted to his lady's court—on condition that he first submit to castration. Fisher rejected the security and prestige of a post at the Galton Laboratory and took up the temporary job as sole statistician in a small agricultural research station [viz., Rothamsted Experimental Station] in the country. (Box, 1978, p. 61)*

All power corrupts! It is impossible to be a professor in charge of an important department, and the editor of an important journal, without being somewhat corrupted. We can now see that in both capacities Pearson made mistakes. He rejected lines of research which later turned out to be fruitful. He used his own energy and that of his subordinates in research which turned out to be much less important than he believed. (Haldane, 1957, p. 303)