Psychology 360: Personality Research
Part II: Psychometric Theory
Part I: Overview and the correlation coefficient

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

October, 2022

Outline

What is psychometrics?

In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be. (Thomsom, 1891)

Taken from Michell (2003) in his critique of psychometrics: Michell, J. The Quantitative Imperative: Positivism, Naïve Realism and the Place of Qualitative Methods in Psychology, Theory & Psychology, Vol. 13, No. 1, 5-31 (2003)

What is psychometrics?

The character which shapes our conduct is a definite and durable 'something', and therefore . . . it is reasonable to attempt to measure it. (Galton, 1884)

The history of science is the history of measurement" (J. M. Cattell, 1893)

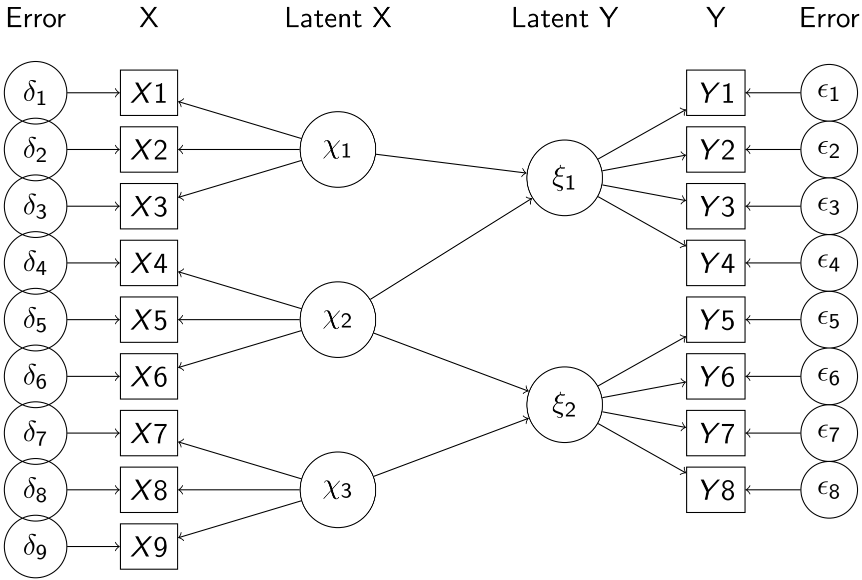
Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality (E.L. Thorndike, 1918)

What is psychometrics?

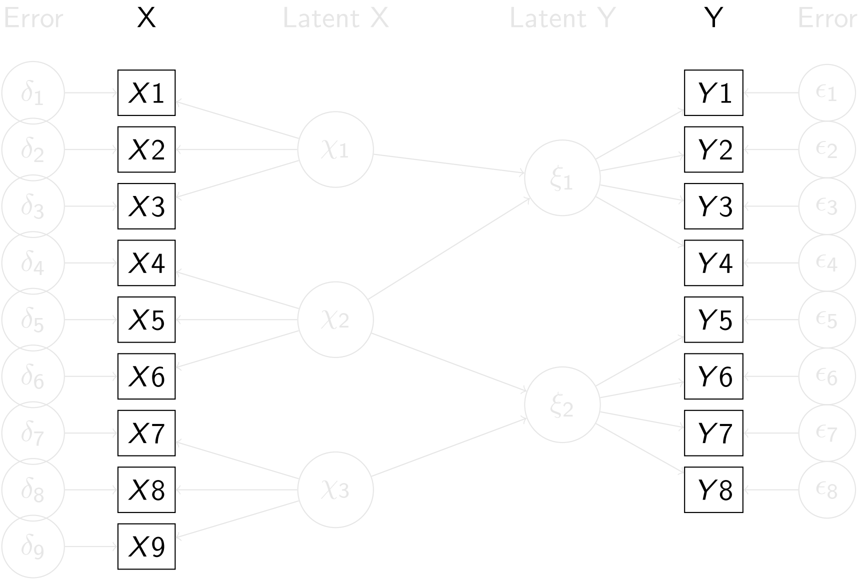
> *We hardly recognize a subject as scientific if measurement is not one of its tools (Boring, 1929)*
>
> *There is yet another [method] so vital that, if lacking it, any study is thought ... not be scientific in the full sense of the word. This further an crucial method is that of measurement. (Spearman, 1937)*
>
> *One's knowledge of science begins when he can measure what he is speaking about and express in numbers (Eysenck, 1973)*

Psychometrics: the assigning of numbers to observed psychological phenomena and to unobserved concepts. Evaluation of the fit of theoretical models to empirical data.

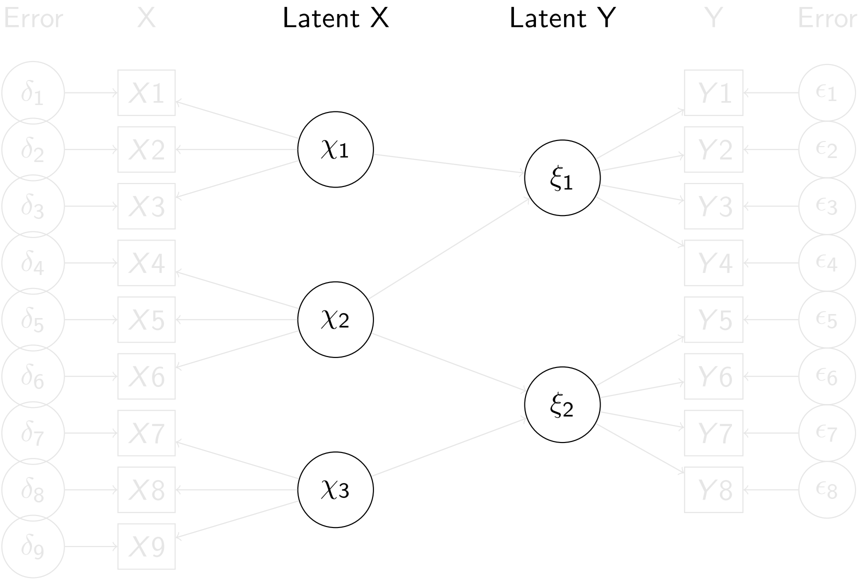
Psychometric Theory: A conceptual Overview



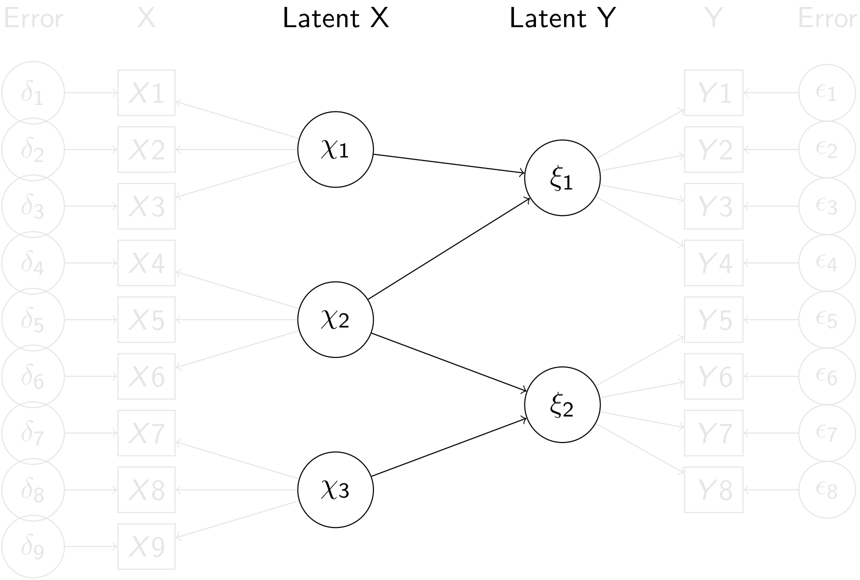
Observed Variables



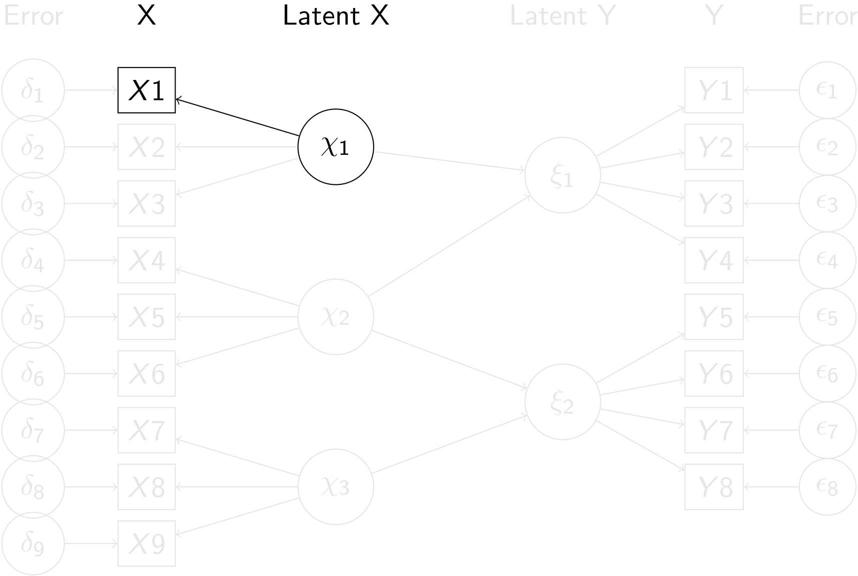
Latent Variables



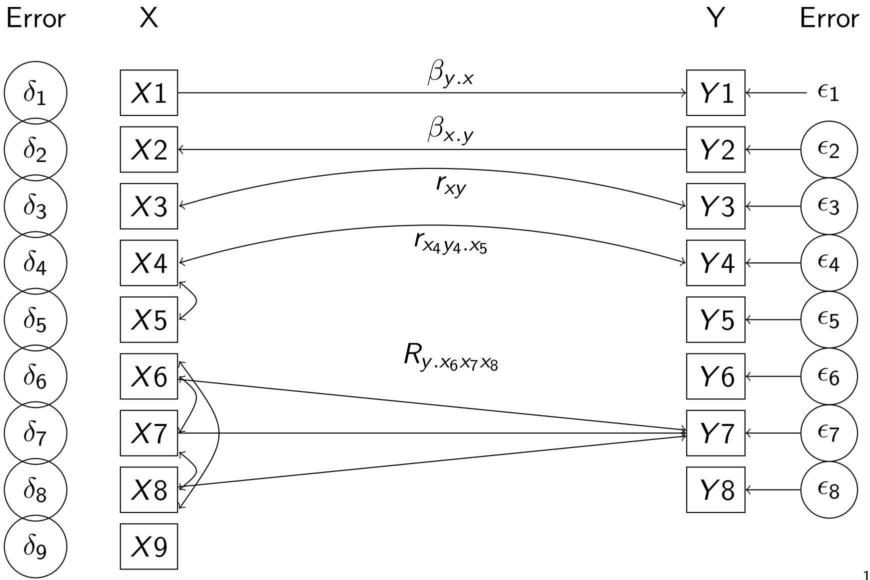
Theory


Error
X
Latent X
Latent Y
Y
Error
δ_1
δ_2
δ_3
δ_4
δ_5
δ_6
δ_7
δ_8
δ_9
X1
X2
X3
X4
X5
X6
X7
X8
X9
χ_1
χ_2
χ_3
ξ_1
ξ_2
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
ϵ_1
ϵ_2
ϵ_3
ϵ_4
ϵ_5
ϵ_6
ϵ_7
ϵ_8

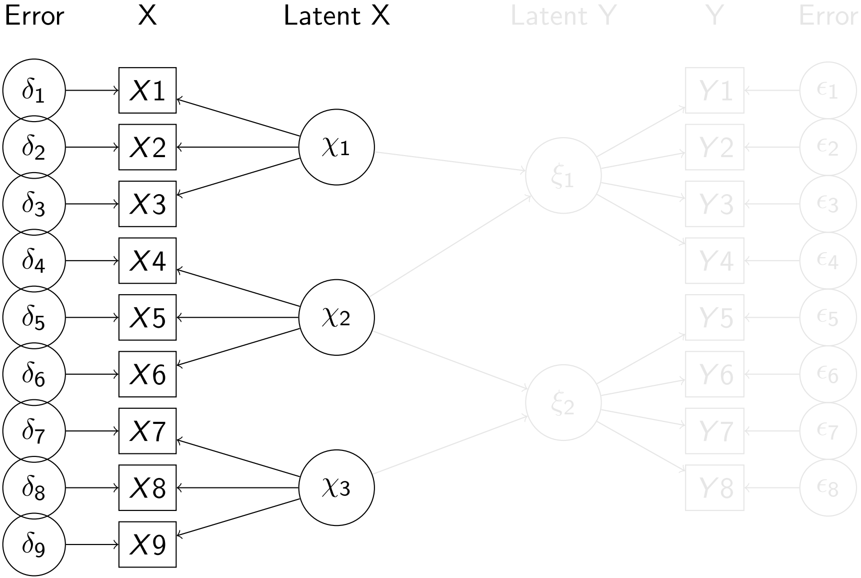
A theory of data and fundamentals of scaling



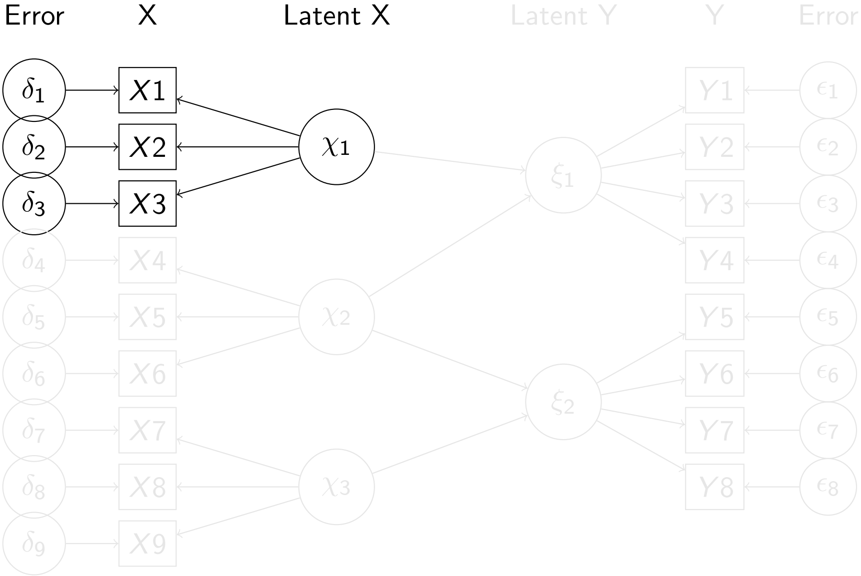
Correlation, Regression, Partial Correlation, Multiple Regression



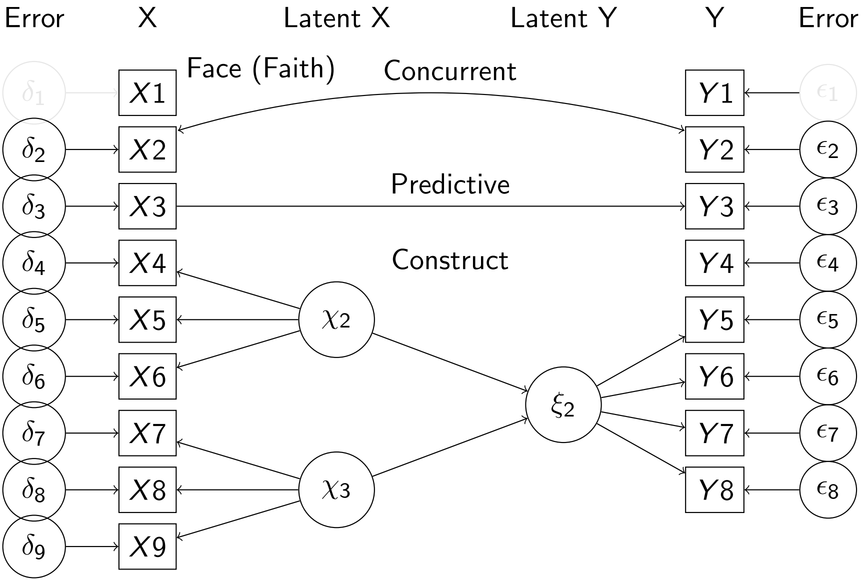
Measurement: A latent variable approach.



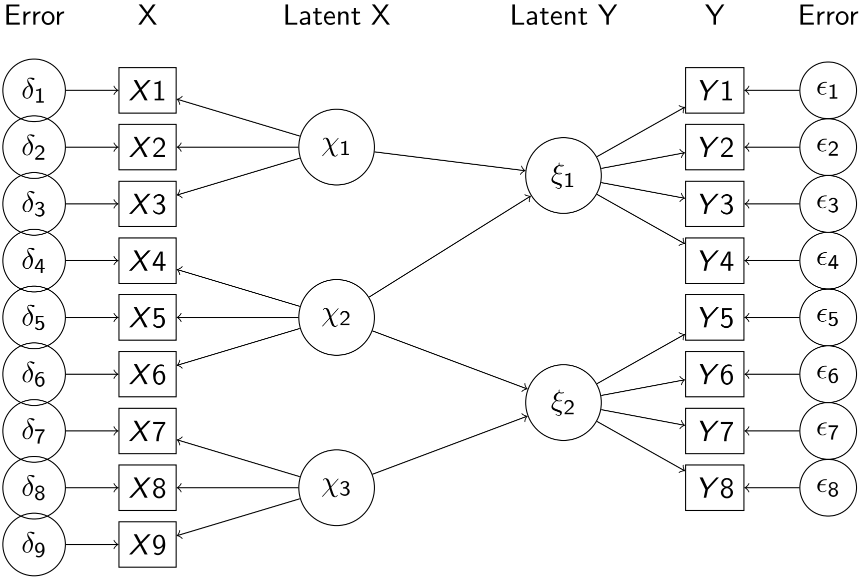
Reliability: How well does a test reflect one latent trait?



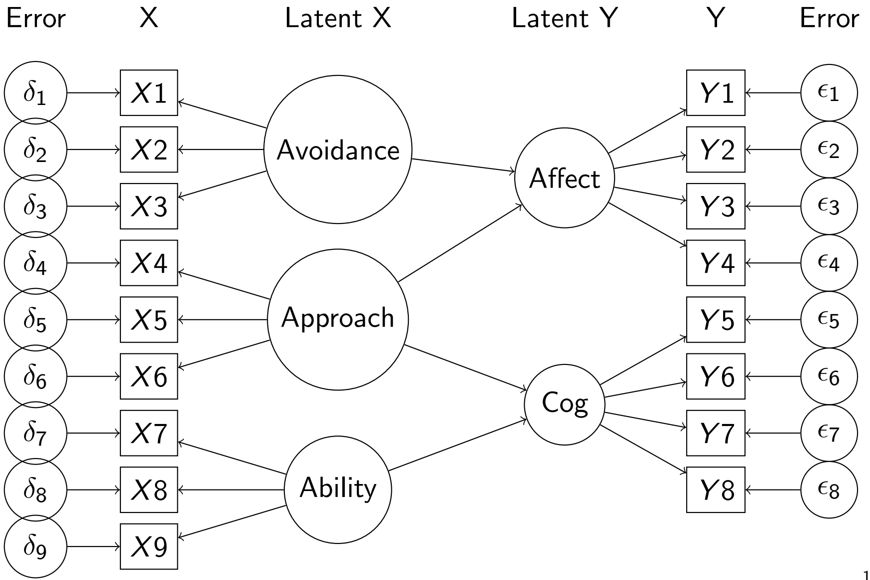
Face, Concurrent, Predictive, Consruct



Psychometric Theory: Data, Measurement, Theory



Psychometric Theory: Data, Measurement, Theory



Data = Model + Residual

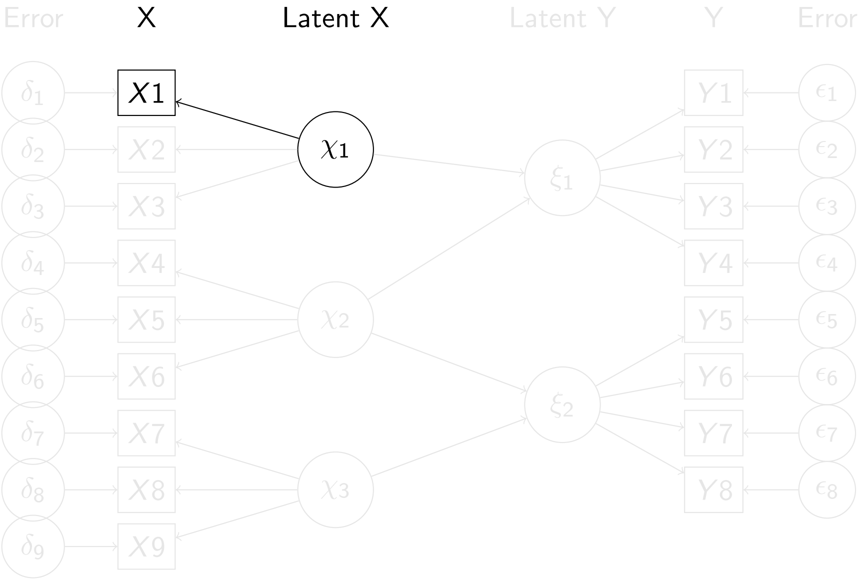
- The fundamental equations of statistics are that
  - Data = Model + Residual
  - Residual = Data - Model
- The problem is to specify the model and then evaluate the fit of the model to the data as compared to other models
  - Fit = f(Data, Residual)
  - Typically: $Fit = f(1 - \frac{Residual^2}{Data^2})$
  - $Fit = f(1 - \frac{(Data - Model)^2}{Data^2})$
- Even for something as simple as the mean is a model of the data. The residual left over after we remove the mean is the variance.

Psychometrics as model estimation and model fitting

We will explore a number of models

1. Modeling the process of data collection and of scaling
   - $X = f(\theta)$
   - How to measure X, properties of the function f.
2. Correlation and Regression
   - $Y = \beta X$
   - $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
3. Factor Analysis and Principal Components Analysis
   - $R = FF' + U^2 \quad R = CC'$
4. Reliability $\rho_{xx} = \frac{\sigma^2_\theta}{\sigma^2_X}$
5. Item Response Theory
   - $p(X|\theta, \delta) = f(\theta - \delta)$
6. Structural Equation Modeling
   - $\rho_{yy} Y = \beta \rho_{xx} X$

A theory of data and fundamentals of scaling



Consider the following numbers, what do they represent?

Table: Numbers without context are meaningless. What do these number represent? Which of these numbers represent the same thing?

2.7182818284590450908	3.141592653589793116
24	86,400
37	98.7
365.25	365.25636305
31,557,600	31,558,150
3,412.1416	.4046856422
299,792,458	$6.022141 * 10^{23}$
42	X

Scaling and assigning of numbers

1. Personality and psychometrics: two overlapping disciplines
2. The field of psychometrics is concerned with the theory and practice of assigning numbers in a systematic way.
3. This allows us to compare observations to each other
4. Formal models developed to allow for comparisons
5. Many assertions depend upon the scales we use

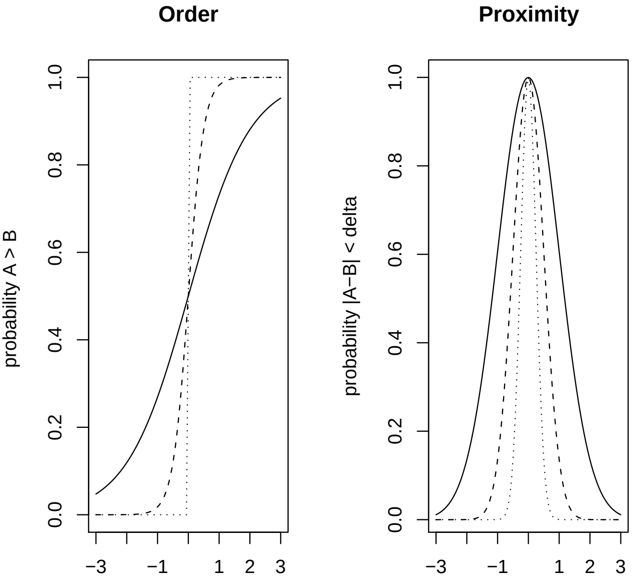
Clyde Coombs and the Theory of Data

A very abstract, but useful approach to the problem.

1. O = the set of objects
   - $O = \{o_i, o_j \dots o_n\}$
2. S = the set of Individuals
   - $S = \{s_i, s_j \dots s_n\}$
3. Two comparison operations
   - order ($x > y$)
   - proximity ($|x - y| < \epsilon$
4. Two types of comparisons
   - Single dyads
     - (s_i, s_j) (s_i, o_j) (o_i, o_j)
   - Pairs of dyads
     - $(s_i, s_j)(s_k, s_l)$ $(s_i, o_j)(s_k, o_l)$ $(o_i, o_j)(o_k, o_l)$

Coombs (1964)

2 types of comparisons: Monotone ordering and single peak proximity



Theory of Data and types of measures

Table: The theory of data provides a 3 x 2 x 2 taxonomy for various types of measures

Elements of Dyad	Number of Dyads	Comparison	Name
People x People	1	Order	Tournament rankings
People x People	1	Proximity	Social Networks
Objects x Objects	1	Order	Scaling
Objects x Objects	1	Proximity	Similarities
People x Objects	1	Order	Ability Measurement
People x Objects	1	Proximity	Attitude Measurement
People x People	2	Order	Tournament rankings
People x People	2	Proximity	Social Networks
Objects x Objects	2	Order	Scaling
Objects x Objects	2	Proximity	Multidimensional scaling
People x Objects	2	Order	Ability Comparisons
People x Objects	2	Proximity	Preferential Choice
People x Objects x Objects	2	Proximity	Individual Differences in Multidimensional Scaling

Tournaments to order people (or teams)

1. Goal is to order the players by outcome to predict future outcomes
2. Complete Round Robin comparisons
   - Everyone plays everyone
   - Requires $N * (N - 1)/2$ matches
   - How do you scale the results?
3. Partial Tournaments – Seeding and group play
   - World Cup
   - NCAA basketball
   - Is the winner really the best?
   - Can you predict other matches?

Friendship as proximity

1. Chess or football provides a ranking based upon an ordering relationship ($p_i > p_j$).
2. Alternatively, friendship groups are based upon closeness ($|p_i - p_j| < \delta$)
   2.1 Do you know person j?
   2.2 Do you like person j? or as an alternative:
   2.3 Please list all your friends in this class (and is j included on the list)
   2.4 Would you be interested in having a date with person j?
   2.5 Would you like to have sex with person j?
   2.6 Would you marry person j?
3. Typically such data will be a rectangular matrix for there are asymmetries in closeness.

Moh's hardness scale provides rank orders of hardness

Table: Mohs' scale of mineral hardness. An object is said to be harder than X if it scratches X. This provides a simple ordering. Also included are measures of relative hardness using a sclerometer (for the hardest of the planes if there is a ansiotropy or variation between the planes) which shows the non-linearity of the Mohs scale (Burchard, 2004).

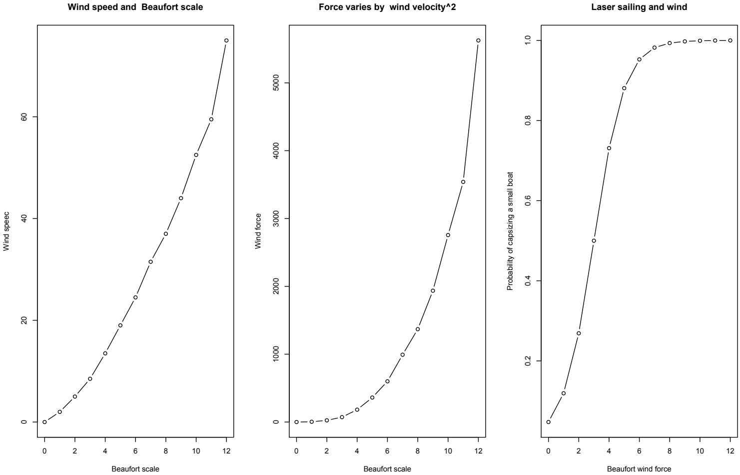
Mohs Hardness	Mineral	Scratch hardness
1	Talc	.59
2	Gypsum	.61
3	Calcite	3.44
4	Fluorite	3.05
5	Apaptite	5.2
6	Orthoclase Feldspar	37.2
7	Quartz	100
8	Topaz	121
9	Corundum	949
10	Diamond	85,300

Ordering based upon external measures

Table: The Beaufort scale of wind intensity is an early example of a scale with roughly equal units that is observationally based. Although the units are roughly in equal steps of wind speed in nautical miles/hour (knots), the force of the wind is not linear with this scale, but rather varies as the square of the velocity.

Force	Wind (Knots)	WMO Classification	Appearance of Wind Effects
0	Less than 1	Calm	Sea surface smooth and mirror-like
1	1-3	Light Air	Scaly ripples, no foam crests
2	4-6	Light Breeze	Small wavelets, crests glassy, no breaking
3	7-10	Gentle Breeze	Large wavelets, crests begin to break, scattered whitecaps
4	11-16	Moderate Breeze	Small waves 1-4 ft. becoming longer, numerous whitecaps
5	17-21	Fresh Breeze	Moderate waves 4-8 ft taking longer form, many whitecaps, some spray
6	22-27	Strong Breeze	Larger waves 8-13 ft, whitecaps common more spray
7	28-33	Near Gale	Sea heaps up, waves 13-20 ft, white foam streaks off breakers
8	34-40	Gale Moderately	high (13-20 ft) waves of greater length, edges of crests begin to break into spindrift, foam blown in streaks
9	41-47	Strong Gale	High waves (20 ft), sea begins to roll, dense streaks of foam, spray may reduce visibility
10	48-55	Storm	Very high waves (20-30 ft) with overhanging crests, sea white with densely blown foam, heavy rolling, lowered visibility
11	56-63	Violent Storm	Exceptionally high (30-45 ft) waves, foam patches cover sea, visibility more reduced
12	64+	Hurricane	Air filled with foam, waves over 45 ft, sea completely white with driving spray, visibility greatly reduced

The Beaufort scale is non-linear with force or probability of capsizing



Models of scaling objects

1. Assume each object $(a, b, ...z)$ has a scale value $(A, B, ...Z)$ with some noise for each measurement.
2. Probability of $A > B$ increases with difference between a and b
3. $P(A > B) = f(a - b)$
4. Can we find a function, f, such that equal differences in the latent variable (a, b, c) lead to equal differences in the observed variable?
5. Several alternatives
   - Direct scaling on some attribute dimension (simple but flawed)
   - Indirect scaling by paired comparisons (more complicated but probably better)

Scaling of Objects: O x O comparisons

1. Typical object scaling is concerned with order or location of objects
2. Subjects are assumed to be random replicates of each other, differing only as a source of noise
3. Absolute scaling techniques
   - Grant Proposals: 1 to 5
   - "On a scale from 1 to 10" this [object] is a X?
   - If A is 1 and B is 10, then what is C?
   - College rankings based upon selectivity
   - College rankings based upon "yield"
   - Zagat ratings of restaurants
   - A - F grading of papers

Absolute scaling: difficulties

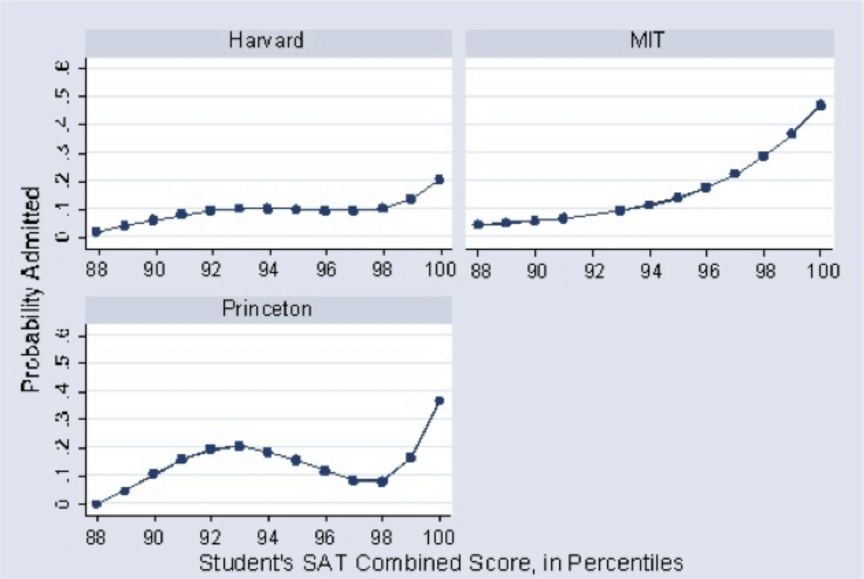
1. "On a scale from 1 to 10" this [object] is a X?
   - sensitive to context effects
   - what if a new object appears?
   - Need unbounded scale
2. If A is 1 and B is 10, then what is C?
   - results will depend upon A, B

Absolute scaling: artifacts

1. College rankings based upon selectivity
    - accept/applied
    - encourage less able to apply
2. College rankings based upon "yield"
    - matriculate/accepted
    - early admissions guarantee matriculation
    - don't accept students who will not attend
3. Proposed solution: college choice as a tournament
    - Consider all schools that accept a student
    - Which school does he/she choose?

Avery, Glickman, Hoxby & Metrick (2013)

A revealed preference ordering Avery et al. (2013)



A revealed preference ordering Avery et al. (2013)

A REVEALED PREFERENCE RANKING OF COLLEGES BASED ON MATRICULATION DECISIONS

Rank Based on Matriculation (with Covariates)	College Name	Theta	Implied Prob. of "Winning" vs. College Listed… 1 Row Below	Implied Prob. of "Winning" vs. College Listed… 10 Rows Below	Rank Based on Matriculation (no Covariates)
1	Harvard University	9.13	0.59	0.93	1
2	Caltech	8.77	0.56	0.92	3
3	Yale University	8.52	0.59	0.92	2
4	MIT	8.16	0.51	0.89	5
5	Stanford University	8.11	0.52	0.90	4
6	Princeton University	8.02	0.73	0.90	6
7	Brown University	7.01	0.56	0.78	7
8	Columbia University	6.77	0.54	0.73	8
9	Amherst College	6.61	0.51	0.71	9
10	Dartmouth	6.57	0.52	0.72	10
11	Wellesley College	6.51	0.53	0.71	12
12	University of Pennsylvania	6.39	0.56	0.71	11

Weber-Fechner Law and non-linearity of scales

1. Early studies of psychophysics by Weber (1834b,a) and subsequently Fechner (1860) demonstrated that the human perceptual system does not perceive stimulus intensity as a linear function of the physical input.
2. The basic paradigm was to compare one weight with another that differed by amount Δ, e.g., compare a 10 gram weight with an 11, 12, and 13 gram weight, or a 10 kg weight with a 11, 12, or 13 kg weight.
3. What was the Δ that was just detectable? The finding was that the perceived intensity follows a logarithmic function.
4. Examining the magnitude of the "*just noticeable difference*" or *JND*, Weber (1834b) found that

$$JND = \frac{\Delta Intensity}{Intensity} = constant. \tag{1}$$

Weber-Fechner Law and non-linearity of scales

1. An example of a logarithmic scale of intensity is the decibel measure of sound intensity.
2. Sound Pressure Level expressed in decibels (dB) of the root mean square observed sound pressure, P_o (in Pascals) is

$$L_p = 20Log_{10}\frac{P_o}{P_{ref}} \tag{2}$$

3. where the reference pressure, P_{ref}, in the air is $20\mu Pa$.
4. Just to make this confusing, the reference pressure for sound measured in the ocean is $1\mu Pa$. This means that sound intensities in the ocean are expressed in units that are 20 dB higher than those units used on land.

The Just Noticeable Difference in Person perception

1. Although typically thought of as just relevant for the perceptual experiences of physical stimuli, Ozer (1993) suggested that the JND is useful in personality assessment as a way of understanding the accuracy and inter judge agreement of judgments about other people.
2. In addition, Sinn (2003) has argued that the logarithmic nature of the *Weber-Fechner Law* is of evolutionary significance for preference for risk and cites Bernoulli (1738) as suggesting that our general utility function is logarithmic.

Money and non linearity

> *... the utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods already possessed .... if ... one has a fortune worth a hundred thousand ducats and another one a fortune worth the same number of semi-ducats and if the former receives from it a yearly income of five thousand ducats while the latter obtains the same number of semi-ducats, it is quite clear that to the former a ducat has exactly the same significance as a semi-ducat to the latter (Bernoulli, 1738, p 25).*

Implies a log function for utility.

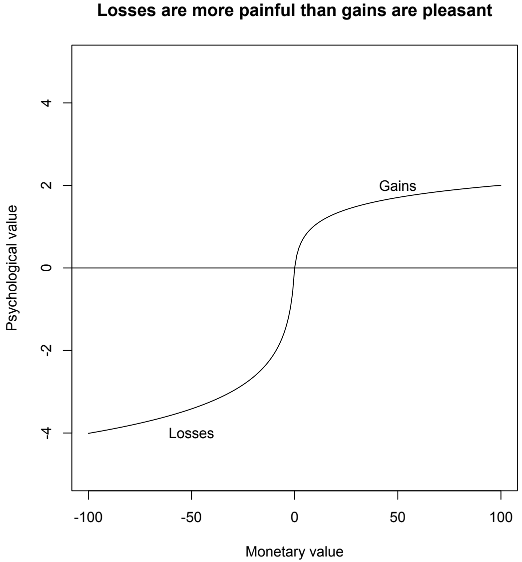
Econs and Humans

1. Simple expected value theory $\implies$ value = probability of event x value of event
2. Bernouli theory of expected utility came to dominate choice theory and is fundamental to economics
3. Studied by comparing gambles and showing utility is non linear with value
   - Would you rather have $80 or a 80% chance of $100 + 20% of $10?
   - expected value is 80 versus .8 * 100 + .2 * 10 = 82
4. Bernouli value (from Kahneman, 2011)

Wealth (millions)	1	2	3	4	5	6	7	8	9	10
Utility units	10	30	48	60	70	78	84	90	96	100

5. Reality meets mathematics. People are paid in $, not log($) But income is distributed in a lognormal distribution.

Kahneman and Tversky: Prospect Theory



Kahneman & Tversky (1979)
Better to skip lunch than be someone's dinner.

Four types of scales and their associated statistics

Table: Four types of scales and their associated statistics (Rossi, 2007; Stevens, 1946) The statistics listed for a scale are invariant for that type of transformation.

Scale	Basic operations	Transformations	Invariant statistic	Examples
Nominal	equality $x_i = x_j$	Permutations	Counts Mode χ^2 and (ϕ) correlation	Detection Species classification Taxons
Ordinal	order $x_i > x_j$	Monotonic (homeomorphic) x' =f(x) f is monotonic	Median Percentiles Spearman correlations*	Mhos Hardness scale Beaufort Wind (intensity) Richter earthquake scale
Interval	differences $(x_i - x_j) > (x_k - x_l)$	Linear (Affine) x' = a + bx	Mean (μ) Standard Deviation (σ) Pearson correlation (r) Regression (β)	Temperature (°F, °C) Beaufort Wind (velocity)
Ratio	ratios $\frac{x_i}{x_j} > \frac{x_k}{x_l}$	Multiplication (Similiarity) x' = bx	Coefficient of variation ($\frac{\sigma}{\mu}$)	Length, mass, time Temperature (°K) Heating degree days

The Beaufort wind speed scale is interval with respect to the velocity of the wind, but only ordinal with respect to the effect of the wind. The Richter scale of earthquake intensity is a logarithmic scale of the energy released but

Graphical and tabular summaries of data

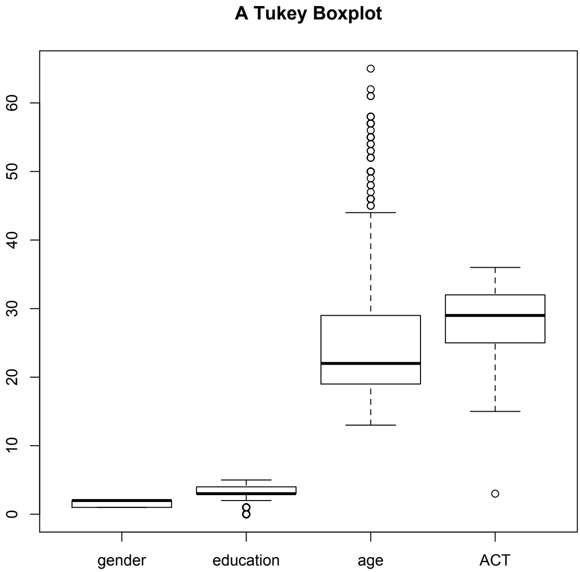
1. The Tukey 5 number summary shows the important characteristics of a set of numbers
   - Maximum
   - 75th percentile
   - Median (50th percentile)
   - 25th percentile
   - Minimum
2. Graphically, this is the box plot
   - Variations on the box plot include confidence intervals for the median

The summary command gives the Tukey 5 numbers

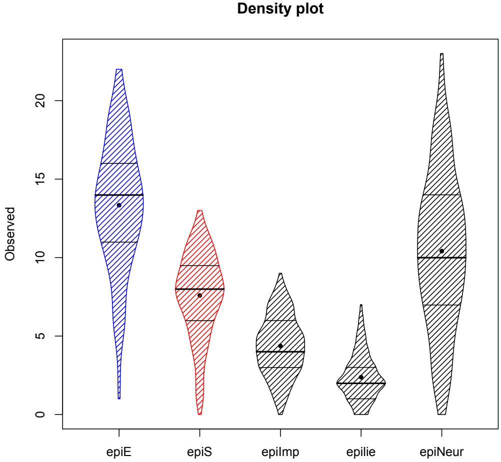
```
> summary(sat.act)

     gender        education          age             ACT
SATV             SATQ
 Min.    :1.000   Min.    :0.000   Min.    :13.00   Min.    : 3.00
Min.    :200.0   Min.    :200.0
 1st Qu.:1.000   1st Qu.:3.000   1st Qu.:19.00   1st Qu.:25.00
1st Qu.:550.0   1st Qu.:530.0
 Median :2.000   Median :3.000   Median :22.00   Median :29.00
Median :620.0   Median :620.0
 Mean   :1.647   Mean   :3.164   Mean   :25.59   Mean   :28.55
Mean   :612.2   Mean   :610.2
 3rd Qu.:2.000   3rd Qu.:4.000   3rd Qu.:29.00   3rd Qu.:32.00
3rd Qu.:700.0   3rd Qu.:700.0
 Max.    :2.000   Max.    :5.000   Max.    :65.00   Max.    :36.00
Max.    :800.0   Max.    :800.0
```

A box plot of the first 4 sat.act variables



A violin or density plot of the first 5 epi.bfi variables



violinBy(epi.bfi[1:5],main="A Tukey violin plot")

The `describe` function gives more descriptive statistics

```
> describe(sat.act)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
gender	1	700	1.65	0.48	2	1.68	0.00	1	2	1	−0.61	−1.62	0.02
education	2	700	3.16	1.43	3	3.31	1.48	0	5	5	−0.68	−0.07	0.05
age	3	700	25.59	9.50	22	23.86	5.93	13	65	52	1.64	2.42	0.36
ACT	4	700	28.55	4.82	29	28.84	4.45	3	36	33	−0.66	0.53	0.18
SATV	5	700	612.23	112.90	620	619.45	118.61	200	800	600	−0.64	0.33	4.27
SATQ	6	687	610.22	115.64	620	617.25	118.61	200	800	600	−0.59	−0.02	4.41

Multiple measures of central tendency

mode
: The most frequent observation. Not a very stable measure, depends upon grouping. Can be used for categorical data.

median
: The number with 50% above and 50% below. A powerful, if underused, measure. Not sensitive to transforms of the shape of the distribution, nor outliers. Appropriate for ordinal data, and useful for interval data.

mean
: One of at least seven measures that assume interval properties of the data.

Multiple ways to estimate the mean

Arithmetic mean $\bar{X} = X_. = (\sum_{i=1}^{N} X_i)/N$ `mean(x)`

Trimmed mean throws away the top and bottom t% of observations. This follows the principle that all data are normal at the middle. `mean(x,trim=.1)`

Winsorized mean Find the arithmetic mean after replacing the n lowest observations with the nth value, and the N largest values with the Nth largest. `winsor(x,trim=.2)`

Geometric Mean $\bar{X}_{geometric} = \sqrt[N]{\prod_{i=1}^{N} X_i} = e^{\Sigma(ln(x))/N}$ (The anti-log of the mean log score). `geometric.mean(x)`

Harmonic Mean $\bar{X}_{harmonic} = \frac{N}{\sum_{i=1}^{N} 1/X_i}$ (The reciprocal of the mean reciprocal). `harmonic.mean(x)`

Circular Mean $\bar{x}_{circular} = tan^{-1}\left(\frac{\sum cos(x)}{\sum sin(x)}\right)$ `circular.mean(x)` (where x is in radians)

circadian.mean `circular.mean(x)` (where x is in hours)

What is the "average" class size?

Table: Average class size depends upon point of view. For the faculty members, the median of 10 is very appealing. From the Dean's perspective, the faculty members teach an average of 50 students per calls. But what about the students?

Faculty Member	Freshman/ Sophmore	Junior	Senior	Graduate	Mean	Median
A	20	10	10	10	12.5	10
B	20	10	10	10	12.5	10
C	20	10	10	10	12.5	10
D	20	100	10	10	35.0	15
E	200	100	400	10	177.5	150
Total						
Mean	56	46	110	10	50.0	39
Median	20	10	10	10	12.5	10

Class size from the students' point of view.

Table: Class size from the students' point of view. Most students are in large classes; the median class size is 200 with a mean of 223.

Class size	Number of classes	number of students
10	12	120
20	4	80
100	2	200
200	1	200
400	1	400

Time in therapy

A psychotherapist is asked what is the average length of time that a patient is in therapy. This seems to be an easy question, for of the 20 patients, 19 have been in therapy for between 6 and 18 months (with a median of 12) and one has just started. Thus, the median client is in therapy for 52 weeks with an average (in weeks) $(1 * 1 + 19 * 52)/20$ or 49.4.

However, a more careful analysis examines the case load over a year and discovers that indeed, 19 patients have a median time in treatment of 52 weeks, but that each week the therapist is also seeing a new client for just one session. That is, over the year, the therapist sees 52 patients for 1 week and 19 for a median of 52 weeks. Thus, the median client is in therapy for 1 week and the average client is in therapy of $(52 * 1 + 19 * 52)/(52+19) =$ 14.6 weeks.

Does teaching effect learning?

1. A leading research team in motivational and educational psychology was interested in the effect that different teaching techniques at various colleges and universities have upon their students. They were particularly interested in the effect upon writing performance of attending a very selective university, a less selective university, or a two year junior college.
2. A writing test was given to the entering students at three institutions in the Boston area. After one year, a similar writing test was given again. Although there was some attrition from each sample, the researchers report data only for those who finished one year. The pre and post test scores as well as the change scores were as shown below:

Types of teaching affect student outcomes?

Table: Three types of teaching and their effect on student outcomes

School	Pretest	Posttest	Change
Junior College	1	5	4
Non-selective university	5	27	22
Selective university	27	73	45

From these data, the researchers concluded that the quality of teaching at the selective university was much better than that of the less selective university or the junior college and that the students learned a great deal more. They proposed to study the techniques used there in order to apply them to the other institutions.

Teaching and math performance

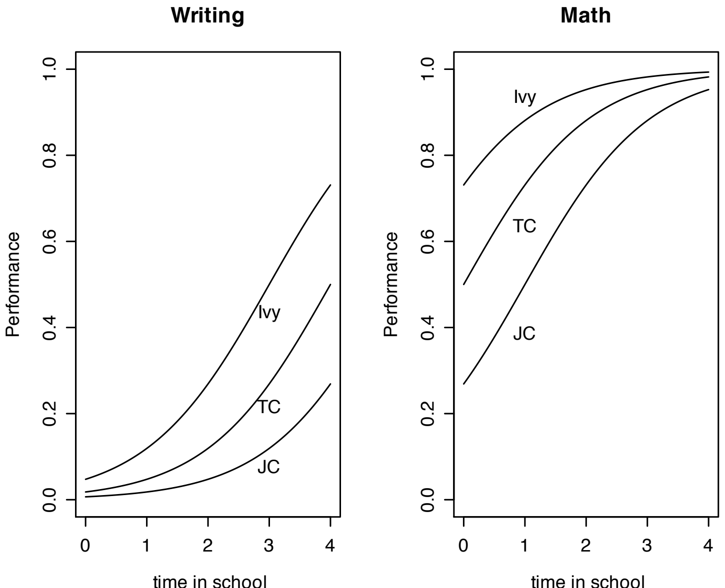
Another research team in motivational and educational psychology was interested in the effect that different teaching at various colleges and universities affect math performance. They used the same schools as the previous example with the same design.

Table: Three types of teaching and their effect on student outcomes

School	Pretest	Posttest	Change
Junior College	27	73	45
Non-selective university	73	95	22
Selective university	95	99	4

They concluded that the teaching at the junior college was far superior to that of the select university. What is wrong with this conclusion?

Effect of teaching, effect of students, or just scaling?

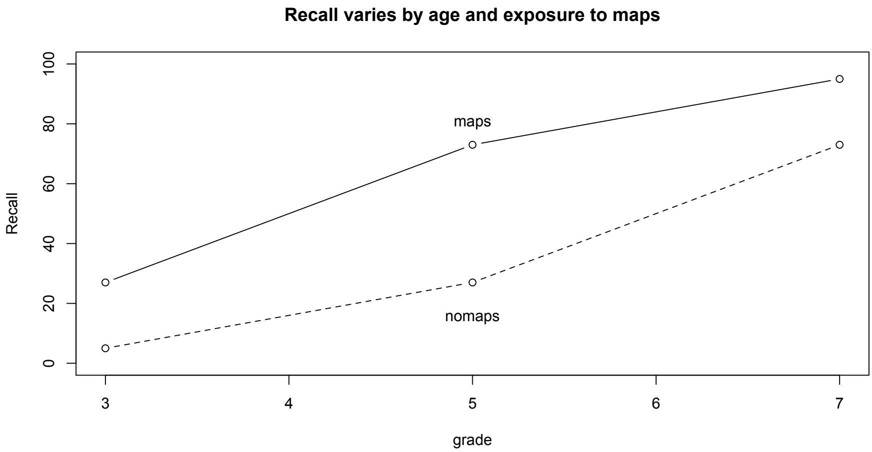


The problem of scaling is ubiquitous

1. A leading cognitive developmentalist believed that there is a critiical stage for learning spatial representations using maps. Children younger than this stage are not helped by maps, nor are children older than this stage.
2. He randomly assigned 3rd, 5th, and 7th grade students into two conditions (nested within grade), control, and map use. Performance was measures on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later.) Half the children were shown a map of the rooms before doing the task.
3. Their scores were

	No Map	Maps	Effect	
3rd grade	5	27	22	Too young
5th grade	27	73	46	Critical period
7th grade	73	95	22	Too old

Map use is most effective at a particular developmental stage

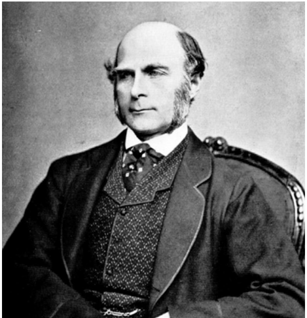


Correlation and Regression

1. Developed in 1886 by Francis Galton
   - Further developments by Karl Pearson and Charles Spearman
2. Correlation/regression are the root concept of psychometrics
   - Other statistics, including factor analysis are ways of partioning correlation matrices
   - Reliability theory is merely an application of factor analysis

Francis Galton 1822-1911

Francis Galton (1822-1911) was among the most influential psychologists of the 19th century. He did pioneering work on the correlation coefficient, behavior genetics and the measurement of individual differences. He introspectively examined the question of free will and introduced the lexical hypothesis to the study of personality and character. In addition to psychology, he did pioneering work in meteorology and introduced the scientific use of fingerprints. Whenever he could, he counted.



Karl Pearson 1857-1936

Carl (Karl) Pearson was among the most influential statisticians of the early 20th century. Founder of the statistics department at University College London. He developed the Pearson Product Moment Correlation Coefficient, its special case the ϕ coefficient, and the tetrachoric correlation. Major behavior geneticist and eugenicist.

Charles Spearman 1863-1945

Charles Spearman (1863-1945) was the leading psychometrician of the early 20th century. His work on the classical test theory, factor analysis, and the g theory of intelligence continues to influence psychometrics, statistics, and the study of intelligence. More than 100 years after their publication, his most influential papers remain two of the most frequently cited articles in psychometrics and intelligence.

Galton's height data

Table: The relationship between the average of both parents (mid parent) and the height of their children. The basic data table is from Galton (1886) who used these data to introduce reversion to the mean (and thus, linear regression). The data are available as part of the **UsingR** or **psych** packages.

```
> library(psych)
> data(galton)
> galton.tab <- table(galton)
> galton.tab[order(rank(rownames(galton.tab)),decreasing=TRUE),] #s
```

	child													
parent	61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	73.7
73	0	0	0	0	0	0	0	0	0	0	0	1	3	0
72.5	0	0	0	0	0	0	0	1	2	1	2	7	2	4
71.5	0	0	0	0	1	3	4	3	5	10	4	9	2	2
70.5	1	0	1	0	1	1	3	12	18	14	7	4	3	3
69.5	0	0	1	16	4	17	27	20	33	25	20	11	4	5
68.5	1	0	7	11	16	25	31	34	48	21	18	4	3	0
67.5	0	3	5	14	15	36	38	28	38	19	11	4	0	0
66.5	0	3	3	5	2	17	17	14	13	4	0	0	0	0
65.5	1	0	9	5	7	11	11	7	7	5	2	1	0	0
64.5	1	1	4	4	1	5	5	0	2	0	0	0	0	0
64	1	0	2	4	1	2	2	1	1	0	0	0	0	0

Galton's height data

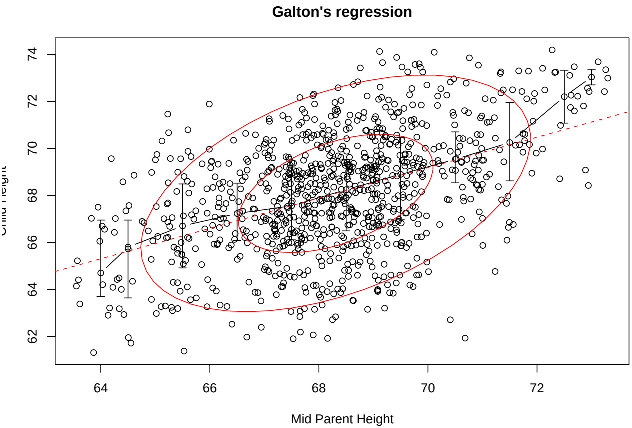


Figure: Galton's data can be plotted to show the relationships between mid parent and child heights. Because the original data are grouped, the data points have been *jittered* to emphasize the density of points along the median. The bars connect the first, 2nd (median) and third quartiles. The dashed line is the best fitting linear fit, the ellipses represent one and two standard deviations from the mean.

Bivariate Regression

X Y ϵ

$$\hat{y} = \beta_{y.x}x + \epsilon$$

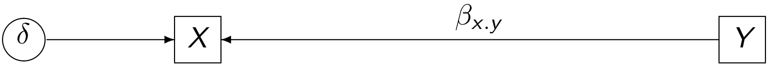
$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$$

Bivariate Regression

δ X Y ϵ

$$\hat{y} = \beta_{y.x}x + \epsilon$$

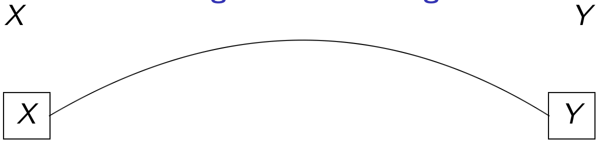
$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$$



$$\hat{x} = \beta_{x.y}y + \delta$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$$

Bivariate Correlation is the geometric average of the two regressions



$\hat{x} = \beta_{x.y}y + \delta$

$\hat{y} = \beta_{y.x}x + \epsilon$

$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$

$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}}$$

The variance and the variance of a composite

1. If $\mathbf{x_1}$ and $\mathbf{x_2}$ are vectors of N observations centered around their mean (that is, deviation scores) their variances are $V_{x1} = \sum x_{i1}^2/(N-1)$ and $V_{x2} = \sum x_{i2}^2/(N-1)$, or, in matrix terms $V_{\mathbf{x}1} = \mathbf{x_1'x_1}/(N-1)$ and $V_{\mathbf{x}2} = \mathbf{x_2'x_2}/(N-1)$.
2. The variance of the composite made up of the sum of the corresponding scores, $\mathbf{x}+\mathbf{y}$ is just

$$V_{(\mathbf{x1}+\mathbf{x2})} = \frac{\sum(x_i+y_i)^2}{N-1} = \frac{\sum x_i^2 + \sum y_i^2 + 2\sum x_i y_i}{N-1} = \frac{(\mathbf{x}+\mathbf{y})'(\mathbf{x}+\mathbf{y})}{N-1}. \quad (3)$$

Or, more generally,

$$\mathbf{S} = \begin{pmatrix} v_{x1} & c_{x1x2} & \cdots & c_{x1xn} \\ c_{x1x2} & v_{x2} & & c_{x2xn} \\ \vdots & & \ddots & \vdots \\ c_{x1xn} & c_{x2xn} & \cdots & v_{xn} \end{pmatrix}$$

Sums as matrix products

$$V_{\mathbf{X}} = \sum \frac{\mathbf{X}'\mathbf{X}}{N-1} = \frac{\mathbf{1}'(\mathbf{X}'\mathbf{X})\mathbf{1}}{N-1}.$$

$$V_{\mathbf{Y}} = \sum \frac{\mathbf{Y}'\mathbf{Y}}{N-1} = \frac{\mathbf{1}'(\mathbf{Y}'\mathbf{Y})\mathbf{1}}{N-1}$$

and

$$C_{\mathbf{XY}} = \sum \frac{\mathbf{X}'\mathbf{Y}}{N-1} = \frac{\mathbf{1}'(\mathbf{X}'\mathbf{Y})\mathbf{1}}{N-1}$$

Use R Continue with part II

Avery, C. N., Glickman, M. E., Hoxby, C. M., & Metrick, A. (2013). A revealed preference ranking of U.S. colleges and universities. *The Quarterly Journal of Economics*, *128*(1), 425–467.

Bernoulli, D. (1954/1738). Exposition of a new theory on the measurement of risk ("Specimen theoriae novae de mensura sortis," Commentarii Academiae Scientiarum Imperialis Petropolitanae 5, St. Petersburg 175-92.) translated by Louise C. Sommer. *Econometrica*, *22*(1), 23–36.

Burchard, U. (2004). The sclerometer and the determination of the hardness of minerals. *Mineralogical Record*, *35*, 109–120.

Coombs, C. (1964). *A Theory of Data*. New York: John Wiley.

Fechner, Gustav Theodor (H.E. Adler, T. (1966/1860). *Elemente der Psychophysik (Elements of psychophysics)*. Leipzig: Breitkopf & Hartel.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Ozer, D. J. (1993). Classical psychophysics and the assessment of agreement and accuracy in judgments of personality. *Journal of Personality*, *61*(4), 739–767.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. *Philisopical Transactions of the Royal Society of London. Series A*, *187*, 254–318.

Pearson, K. & Heron, D. (1913). On theories of association. *Biometrika*, *9*(1/2), 159–315.

Rossi, G. B. (2007). Measurability. *Measurement*, *40*(6), 545 – 562.

Sinn, H. W. (2003). Weber's law and the biological evolution of risk preferences: The selective dominance of the logarithmic utility function, 2002 geneva risk lecture. *Geneva Papers on Risk and Insurance Theory*, *28*(2), 87–100.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Weber, E. H. (1834b). *De pulsu, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae*. Leipzig: Kohler.

Weber, E. H. (1948/1834a). Concerning touch, 1834. In W. Dennis (Ed.), *Readings in the history of psychology* (pp. 155–156). East Norwalk, CT: Appleton-Century-Crofts Appleton-Century-Crofts Print.