Individual Differences and Correlations

I. Calculation of Scale or Test scores

II. Descriptive Statistics

III. Correlation

IV. Normal Scores

V. Normalized Scores

VI. Normed Scores

I. Calculation of Scale or Test Scores

Analyses performed on individual items from a scale or test (item analysis) or performed on a combination of items (validity, hypothesis tests)

Combination of items measuring the same underlying attribute (“scale score”, “test score”, “composite score”)

- Usually sum or mean of the items (usually equally weighted)
- Transformations are common before analyses (rescoring, reversing scores)
- Standardizing scores first

I. Calculation of Scale or Test Scores

Composite score can be average or sum of a set of questions from a measure

Example: five Likert-type questions about perceived stress combined to obtain a total scale score of perceived stress

$$\text{sumstress} = X_1 + X_2 + X_3 + X_4 + X_5$$

or

$$\text{avestress} = (X_1 + X_2 + X_3 + X_4 + X_5)/5$$

I. Calculation of Scale or Test Scores

Linear transformations: adding, subtracting, multiplying, dividing all cases in a data set by the same number

Example: dividing income by 1000 to express income in thousand dollar units

Counter examples (not linear): squaring all income scores on a test, taking the logarithm of reaction time, “normalizing” transformation to eliminate skew

I. Calculation of Scale or Test Scores

Linear transformations have no effect on statistical tests (e.g., correlations, t-tests)

So, sum and average composite scores will be equivalent (if no missing data) and choice of using a sum or average for scale scores is arbitrary

I. Calculation of Scale or Test Scores

Unweighted

$$\text{avestress} = (X_1 + X_2 + X_3 + X_4 + X_5)/5$$

Weighted

$$\text{avestress} = \left[\left(.1 \cdot X_1\right)0 + \left(.3 \cdot X_1\right) + \left(.2 \cdot X_1\right) + \left(.1 \cdot X_1\right) + \left(.3 \cdot X_1\right)\right] / 5$$

Unweighted used in nearly all research and practice applications

Weighting ok *if certain* that some items are more important than others

If all items good ones and highly related, little difference between weighted and unweighted

I. Calculation of Scale or Test Scores

In SPSS

```
sumstress = sum(Q1,Q2,Q3,Q4,Q5).
```

or Transformation -> Compute Variable…

Choose “Statistical” under Function Group, “Sum” under Functions and Special Variables, enter new variable name under Target Variable and click on items you wish to average to enter within parentheses on the right

```
avestress = mean(Q1,Q2,Q3,Q4,Q5).
```

or Transformation -> Compute Variable…

Choose “Statistical” under Function Group, “Mean” under Functions and Special Variables, enter new variable name under Target Variable and click on items you wish to sum to enter within parentheses on the right

I. Calculation of Scale or Test Scores

Careful that if any cases have missing data on any item, sums or averages will be calculated without those values

For averages, it is fairly standard practice (although not always optimal) to ignore missing values

For sums, an alternative is to compute sum by addition sign and dividing by the number of items:

```
sumstress = (Q1 + Q2 + Q3 + Q4 + Q5)/5.
```

Which requires that no values be missing for any case.

II. Descriptive Statistics

Mean and Variance

Mean

$$\bar{X} = \frac{\sum X}{N}$$

Variance

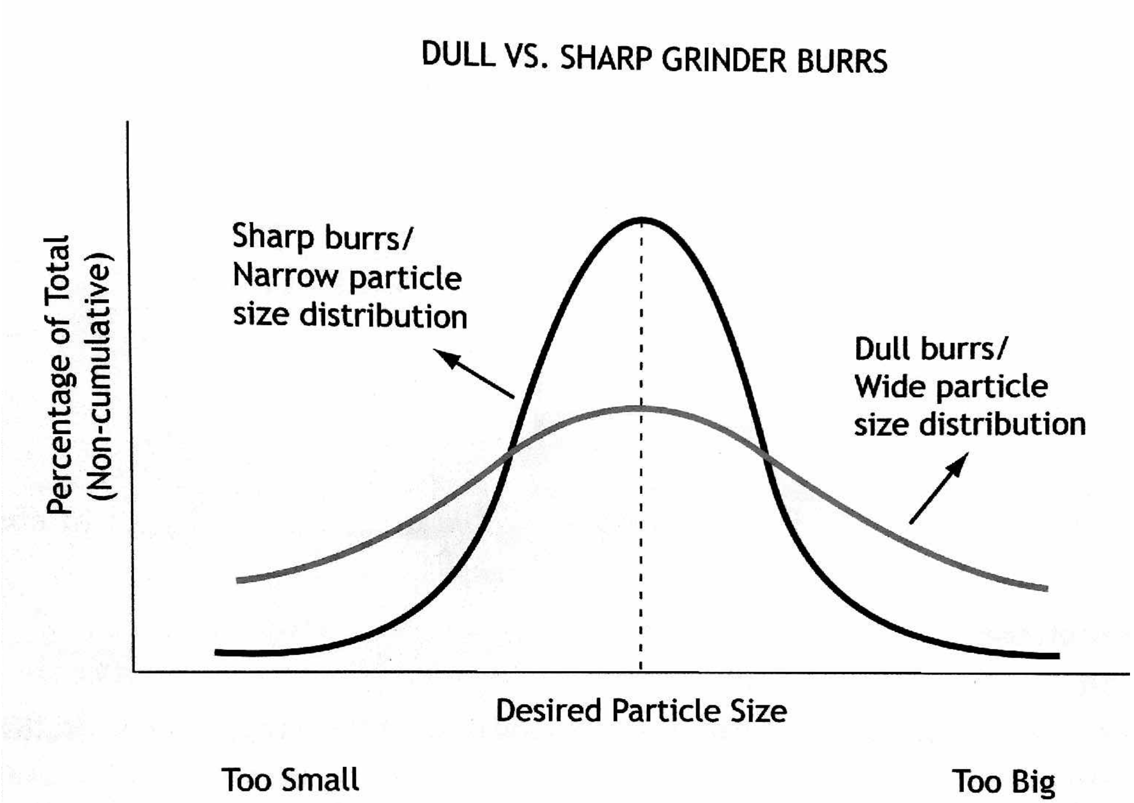
$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

Standard deviation

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

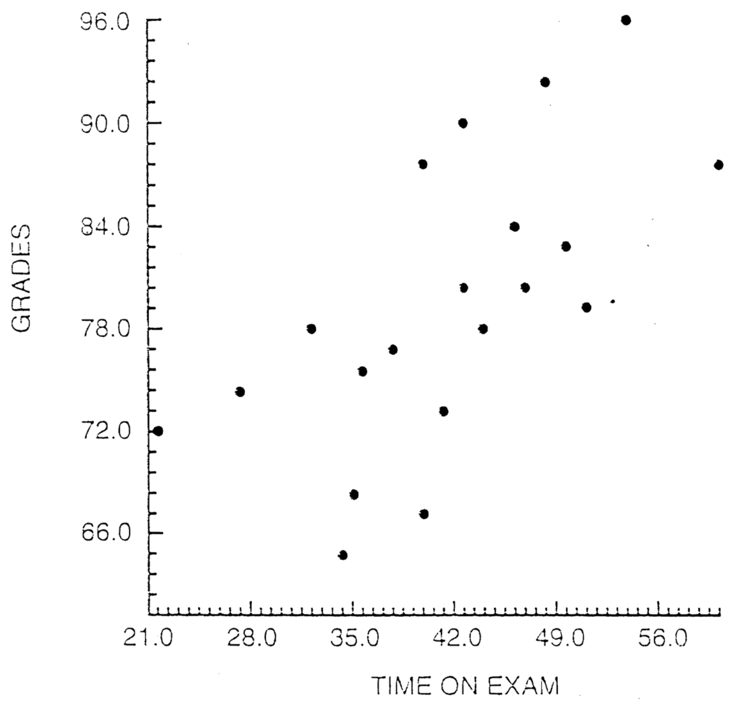
II. Descriptive Statistics

Coffee Grinds Frequency Histogram



III. Correlation

Hypothetical Data of Time Spent on Exam and Grade on Exam



III. Correlation

Correlations are on a standard metric

Possible range from -1.0 to +1.0

Positive value indicates higher values of Y when higher values of X observed

Example: SES and educational tests

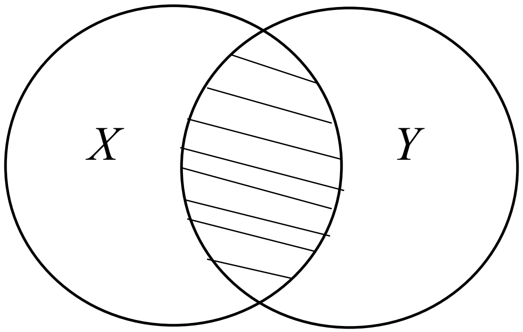
Negative correlation indicates lower values of Y when higher values of X observed

Example: driver age and risk of car accident

III. Correlation

Squaring correlation gives the percentage of shared variance between the two variables, r^2

Can be represented by a Venn diagram.



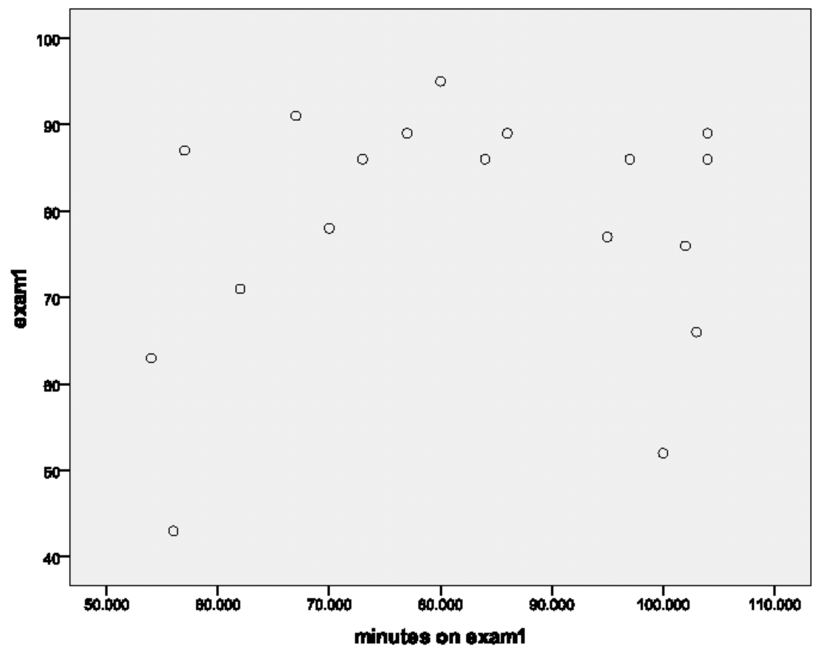
III. Correlation

Correlations can be misleading if

- Nonlinear relationship
- Outliers
- Restricted range

III. Correlation

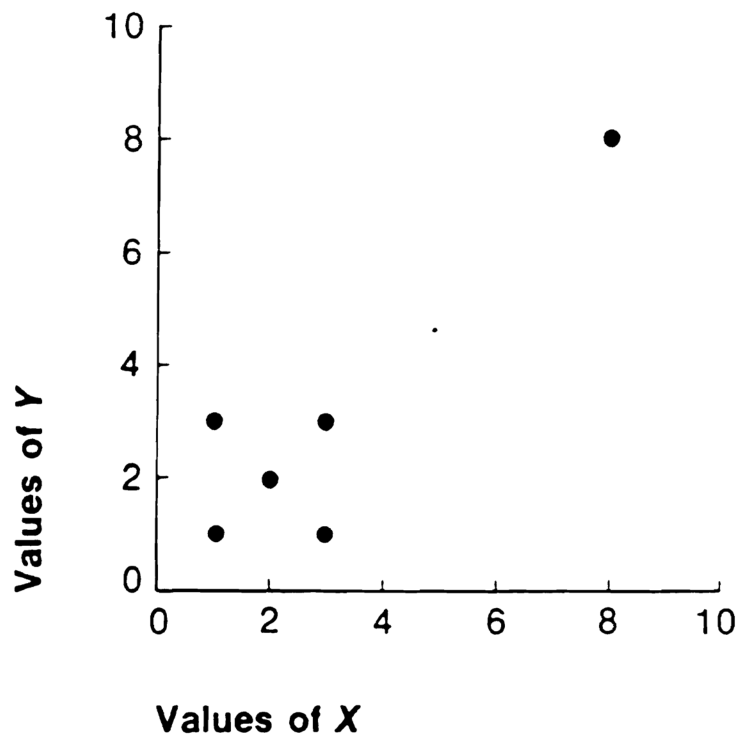
Nonlinear relationship problem



Real Data of Time Spent on Exam and Grade on Exam

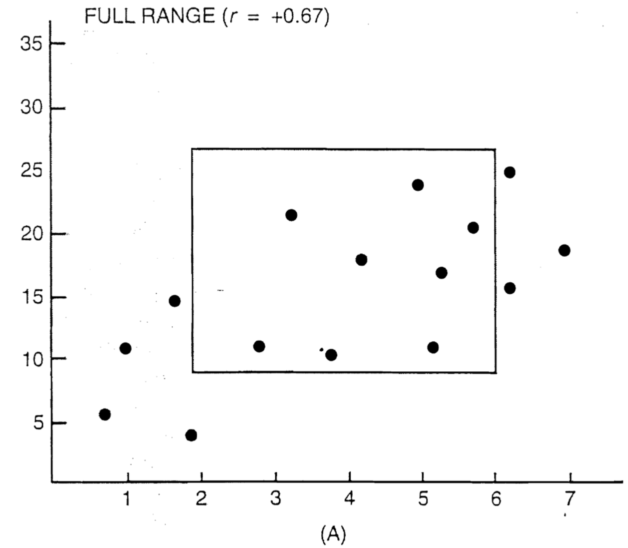
III. Correlation

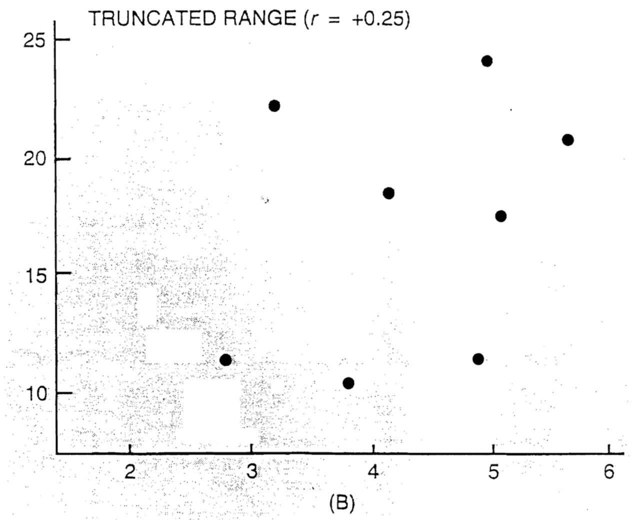
Outlier problem



III. Correlation

Restriction of range problem





III. Correlation

Correlation

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

Covariance

$$c_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N}$$

Standard Deviation of X

$$s_x = \frac{\sqrt{\sum (X - \bar{X})^2}}{N}$$

III. Correlation

Correlation

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{c_{xy}}{s_x s_y}$$

Covariance is the "unstandardized" version of the correlation

We standardize when we divide by the standard deviations

IV. Normal Scores

It is important to distinguish between *normal* scores, *normalized* scores, and *normed* scores

IV. Normal Scores

Normal (standardized) scores are also called z-scores

$$z = \frac{X - \bar{X}}{s}$$

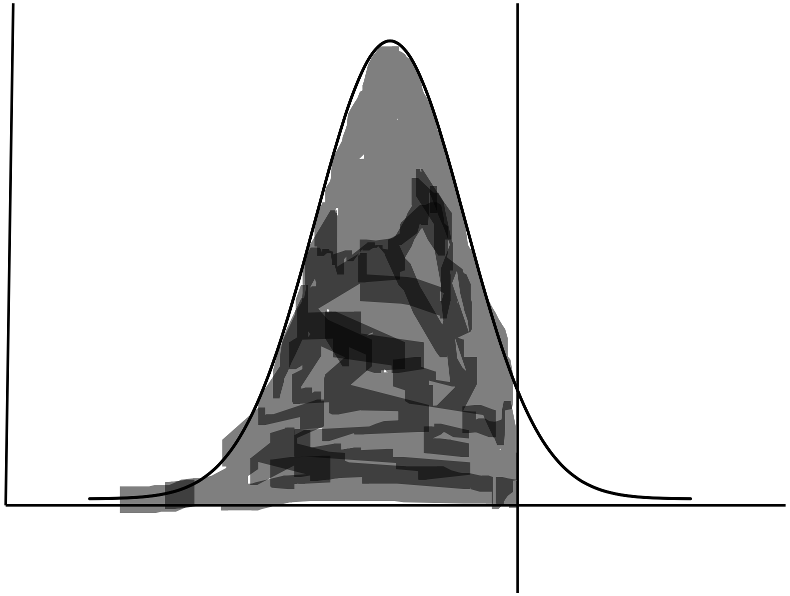
The mean is subtracted and the difference is divided by the standard deviation for the sample

Values have a mean of 0 and a standard deviation (and variance) of 1

Linear transformation, so statistical tests are not affected and there is a perfect correlation with raw scores

IV. Normal Scores

Standardized scores are useful because we can estimate the percentile (percentage of respondents at or below) an individual's score by using the properties of the normal curve



IV. Normal Scores

Variations on the standardized score include the following, computed by multiplying z-score by the desired standard deviation and adding the desired mean

Type	Mean	Standard Deviation	Use
IQ Scores	100	15	Intelligence and some aptitude tests
T Scores (or McCall T scores)	50	10	MMPI, some aptitude tests
SAT	500	100	Some aptitude tests
Stanine (standard nine)	5	2	Military; 1-9 values
GRE	150	8.75	Graduate record exam for grad school admission

V. Normalized Scores

In contrast to normal (standardized) scores, *normalized scores* are nonlinear transformations of the data

Nonlinear transformations alter the correlations and statistical tests, so caution is warranted

V. Normalized Scores

Common methods include

Percentile method

- Find percentiles for all cases in your sample
- Use the percentile to obtain z-score for all cases, z_n
- Compute the normalized score using desired mean (μ_n) and standard deviation (s_n)
- e.g., $T_{normalized} = z_n\left(s_n\right) + \mu_n$

Box-Cox transformation – raises the score to a power (λ) and divides by that number; λ can be estimated from the data

VI. Normed Scores

Normed scores use of percentiles from a comparison population

- Data are collected from a large and (hopefully) representative sample or samples to use as a reference
- Raw scores from a new sample are standardized according to the mean and standard deviation from comparison representative sample to obtain percentiles
- Norms are often established for particular groups, e.g., sex, race, age
- Example: Child height is compared to norms from the population, often computed by age and sex
- May or may not be based on standardized scores and/or normalized values