Individual Differences and Correlation

PSYC3302: Psychological Measurement and Its Applications

Mark Hurlstone
Univeristy of Western Australia

Week 2

Learning Objectives

- Foundations of psychological measurement:
  - variability
  - covariability
  - interpreting test scores

Individual Differences

- Individual differences are the currency of psychometric analysis
- It is assumed that individual differences in psychological attributes exist between people, and that these differences can be quantified
- To measure individual differences on some psychological attribute, we administer a psychological test or measure that is assumed to tap that attribute to a group of people
- The resulting collection of measurement or test scores constitutes a *distribution* of scores
- *Variability* is the term used to describe the differences among the scores in a distribution

Individual Differences

- Before we look at the concept of variability, let's first talk a bit more about distributions of scores

Frequency Distributions

- A teacher administers a 100-item multiple choice exam to the 25 students in her class
- The table on the right shows each student's *raw* score (number correct) on the test
- The collection of scores constitute a distribution of scores
- The scores can be distributed in various ways to help summarise the data

Student	Score (number correct)
Judy	78
Joe	67
Lee-Wu	69
Miriam	63
Valerie	85
Diane	72
Henry	92
Esperanza	67
Paula	94
Martha	62
Bill	61
Homer	44
Robert	66
Michael	87
Jorge	76
Mary	83
"Mousey"	42
Barbara	82
John	84
Donna	51
Uriah	69
Leroy	61
Ronald	96
Vinnie	73
Bianca	79

Frequency Distributions

- One way to distribute the scores is according to the frequency with which they occur
- In a *simple frequency distribution*, all scores are listed alongside the number of times each score occurred
- The scores could be listed in a table or illustrated graphically
- This approach is useful if we have a large number of scores (hundreds or thousands) but less useful with only 25 responses

Score	*f* (frequency)
96	1
94	1
92	1
87	1
85	1
84	1
83	1
82	1
79	1
78	1
76	1
73	1
72	1
69	2
67	2
66	1
63	1
62	1
61	2
51	1
44	1
42	1

Frequency Distributions

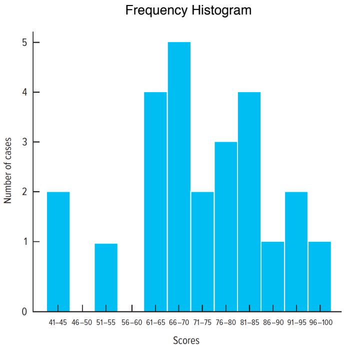
- Another kind of frequency distribution is a *grouped frequency distribution*
- In such distributions, test score intervals, known as *class intervals*, replace the raw test scores
- In the distribution on the right, the test scores have been grouped into 12 class intervals each equal to 5 points
- The data are now starting to look much more meaningful

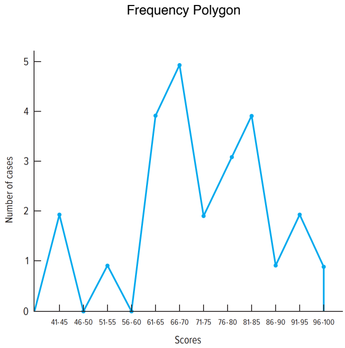
Class Interval	*f* (frequency)
95–99	1
90–94	2
85–89	2
80–84	3
75–79	3
70–74	2
65–69	5
60–64	4
55–59	0
50–54	1
45–49	0
40–44	2

The width of the class intervals used is a decision for the test user

Frequency Distributions

- Frequency distributions are more interpretable when conveyed graphically
- Two types of graphs that are useful for illustrating frequency distributions are *histograms* and *frequency polygons*





Frequency Distributions

- Frequency distributions can take on many different shapes
- One type of distribution—which you are already familiar with—that is of particular interest to measurement researchers is the normal distribution
    - notice that the distribution on the previous slide approximates a normal distribution (although it is somewhat skewed)
- We will talk more about the normal distribution and other types of distributions shortly

Variability and Distributions

- Many key ideas in psychological measurement depend upon our ability to detect and describe distributions of test scores
- A key goal of statistics is to describe distributions of scores in meaningful ways
- There are at least three kinds of information that are used to do this
  1. central tendency
  2. variability
  3. shape
- A good understanding of these concepts is crucial for our future discussion of reliability and validity concepts

Central Tendency

- One way to describe a distribution is through a measure of *central tendency*
- This is a statistic that describes the "typical" or average score between the extreme scores in a distribution
- There are various measures of central tendency (median and mode) but the mean, denoted as $\overline{X}$, is the most common:

$$\text{Mean} = \overline{X} = \frac{\sum X}{N}, \tag{1}$$

- where $\sum$ means to sum each individual score X in the distribution, and N represents the total number of scores

Central Tendency

Example: calculate the mean of the 25 raw class exam scores presented in the table presented earlier

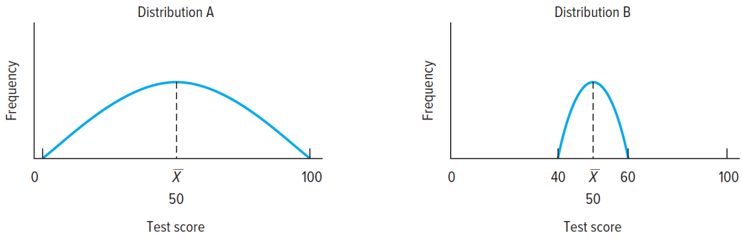
$$\overline{X} = \frac{78 + 67 + 69 + 63 + 85 + \cdots + 79}{25},$$

$$\overline{X} = \frac{1803}{25}$$

$$\overline{X} = 72.12$$

Variability

- The mean is a useful way of describing a distribution but we are more interested in quantifying the extent to which people in a distribution differ from one another—a concept known as *variability*
- Variability is an indication of how scores in a distribution are scattered or dispersed
- Two distributions can have the same mean but very different dispersions of scores:



Variability

- There are various statistical approaches to representing the variability in a distribution of scores (range, interquartile and semi-interquartile ranges, average deviation)
- We will focus on the *variance* and *standard deviation* since they are the most common metrics of variability, and they lie at the heart of psychometric analysis
- The variance and standard deviation reflect variability as the degree to which scores in a distribution deviate (i.e., differ) from the mean of that distribution

Variance

- The variance is the mean of the squared deviations between the scores in a distribution and their mean. To compute the variance we:
  1. calculate the deviation of each score from the mean X-$\overline{X}$
  2. square each deviation $(X$-$\overline{X})^2$
  3. sum the squared deviations and divide by the total number of scores in the distribution
- Formally, the variance is given by:

$$\text{Variance} = s^2 = \frac{\sum (X - \overline{X})^2}{N}, \tag{2}$$

- It represents the average degree to which people differ from each other

Variance

Example: calculate the variance of the 25 raw class exam scores presented in the table presented earlier

$$s^2 = \frac{[78 - 72.12]^2 + [67 - 72.12]^2 + [69 - 72.12]^2 + \cdots + [79 - 72.12]^2}{25}$$

$$s^2 = \frac{4972.64}{25}$$

$$s^2 = 198.91$$

Standard Deviation

- The standard deviation is closely related to the variance
- It is equal to the square root of the "average of the squared deviations about the mean"—viz. the variance
- Hence, the standard deviation is simply the square root of the variance:

$$\text{Standard deviation} = \sigma = \sqrt{s^2} = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}, \tag{3}$$

- Note that I am using σ to denote the standard deviation but the textbook denotes it as s

Standard Deviation

- The standard deviation is closely related to the variance
- It is equal to the square root of the "average of the squared deviations about the mean"—viz. the variance
- Hence, the standard deviation is simply the square root of the variance:

$$\text{Standard deviation} = \sigma = \sqrt{s^2} = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}, \qquad (3)$$

- Note that I am using σ to denote the standard deviation but the textbook denotes it as s

Standard Deviation

Example: calculate the standard deviation of the 25 raw class exam scores presented in the table presented earlier

$$\sigma = \sqrt{s^2} = \sqrt{198.91} = 14.1$$

Binary Items

- Sometimes psychological tests are based on dichotomous responses:
  - people may have to give *yes* or *no* answers to questions
  - people may have to *agree* or *disagree* with statements
  - people's responses to test items may be classified as *correct* or *incorrect*
- Negatively valenced responses (e.g., no, disagree, incorrect) are coded with a 0, whereas positively valenced responses (e.g., yes, agree, correct) are coded with a 1
- Like continuous items, we can calculate the mean, variance, and standard deviation for binary items

Binary Items

- The mean of a binary item is calculated in the same way as for a continuous item (see earlier), but we will denote the mean as p rather than $\overline{X}$
- For example, suppose that 10 people are asked a single question and eight people answer correct, whilst two answer incorrect
- The mean is the proportion p of positively valenced (correct) responses: 8 / 10 = .80
- The proportion of negatively valenced (incorrect) responses, denoted as q, is therefore equal to 1–p: 1 – .80 = .20

Binary Items

- The variance for a binary item is calculated as:

$$s^2 = pq \tag{4}$$

- For our example:

$$s^2 = (.80)(.20)$$

$$s^2 = .16$$

- The standard deviation is the square root of the variance:

$$\sigma = \sqrt{.16}$$

$$\sigma = .40$$

Variance vs. Standard Deviation

- Both the variance and standard deviation are important measures of variability in psychological measurement
- However, the standard deviation tends to be preferred on account that it is more "intuitive"
- This is because it reflects variability in terms of the raw deviation scores, whereas the variance reflects variability in terms of squared deviation scores
  - it makes intuitive sense to say that the students differed on average by 14 test score points
  - it does not make intuitive sense to say that the students differed from one another on average by 198 squared deviation points

Factors Affecting The Variance and Standard Deviation

- The size of the variance—and therefore the standard deviation—is affected by two things:
  1. the degree to which scores in a distribution differ from each other
  2. the metric of the scores in the distribution

1. The Degree to Which Scores in a Distribution Differ

- All else being equal, a larger variance and standard deviation indicates greater variability within a distribution of scores
- However, all else is not equal because the metric of scores in a distribution will also affect these variability indices

2. The Metric of the Scores in the Distribution

- By "metric of scores", we mean whether the scores in a distribution are derived from a "large-scale" or a "small-scale" measure
- The Scholastic Achievement Test (SAT) is an example of a large-scale measure—it produces large test scores that average about 1000
- Grade Point Average (GPA) is an example of a small-scale measure—it produces test scores that range between 0 and 4

2. The Metric of the Scores in the Distribution

- Suppose we have two distributions of scores—one derived from a large-scale measure (e.g., SAT) and one from a small-scale measure (e.g., GPA)
- Suppose that the scores in each distribution differ from each other to exactly the same degree
- The variance and standard deviation will nonetheless be larger for the large-scale than the small-scale measure
- This is because the former measure produces larger test scores that inflate these variability indices

Note:

- Caution should be exercised when comparing the variability indices of distributions of scores based on different metrics

2. The Metric of the Scores in the Distribution

- Suppose we have two distributions of scores—one derived from a large-scale measure (e.g., SAT) and one from a small-scale measure (e.g., GPA)
- Suppose that the scores in each distribution differ from each other to exactly the same degree
- The variance and standard deviation will nonetheless be larger for the large-scale than the small-scale measure
- This is because the former measure produces larger test scores that inflate these variability indices

Note:

- Caution should be exercised when comparing the variability indices of distributions of scores based on different metrics

Interlude: The N vs. $N - 1$ Controversy

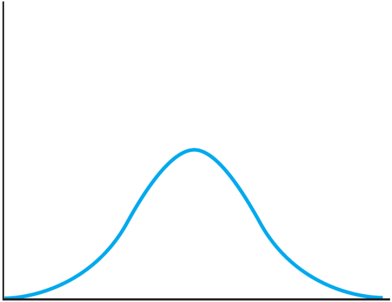
- There is some controversy in the statistics literature about whether the denominator in the calculation of the variance and standard deviation should be N or $N - 1$ (see p.45 of the textbook)
- The convention used in the textbook—and most other psychometrics texts—is the N convention, and that is what we will use
- The key thing you need to know is that SPSS uses the $N - 1$ convention, so bear in mind that it will produce slightly different values for the variance and standard deviation than hand calculations performed using the N convention

Distribution Shapes and Normal Distributions

- A distribution of scores is usually graphically represented by a curve
- The x-axis gives the test score values and the y-axis represents the frequency or proportion of those values in the distribution
- Distributions come in many different shapes
- One distribution that is fundamental to many statistical ideas and concepts is the *normal distribution*
- Many statistical procedures are based on the assumption that test scores are (approximately) normally distributed

Distribution Shapes and Normal Distributions

- The normal curve is a bell shaped smooth curve that is highest at its centre
- The curve is perfectly symmetrical, with no skewness

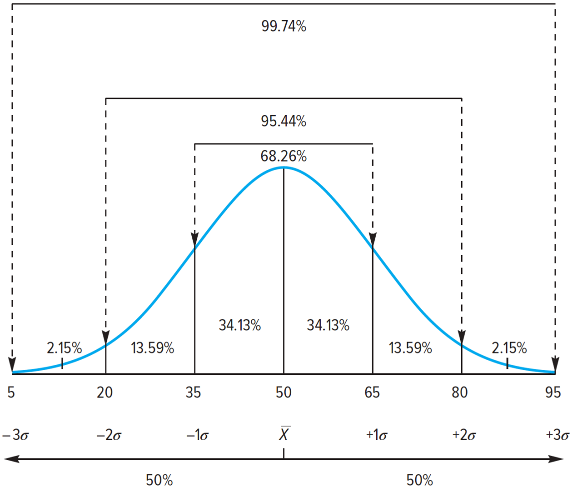


Normal (bell-shaped) curve

Distribution Shapes and Normal Distributions

- To illustrate the key properties of a normal distribution, let's consider a distribution of National Spelling Test Scores with $\overline{X} = 50$ and $\sigma = 15$

Distribution Shapes and Normal Distributions



Distribution Shapes and Normal Distributions

- 50% of the scores occur above the mean and 50% occur below the mean
- Approximately 34% of all scores occur between the mean and +1 standard deviation
- Approximately 34% of all scores occur between the mean and –1 standard deviation
- Approximately 68% of all scores occur between ± 1 standard deviations
- Approximately 95% of all scores occur between ± 2 standard deviations

Distribution Shapes and Normal Distributions

- 50% of the scores occur above the mean and 50% occur below the mean
- Approximately 34% of all scores occur between the mean and +1 standard deviation
- Approximately 34% of all scores occur between the mean and –1 standard deviation
- Approximately 68% of all scores occur between ± 1 standard deviations
- Approximately 95% of all scores occur between ± 2 standard deviations

Distribution Shapes and Normal Distributions

- 50% of the scores occur above the mean and 50% occur below the mean
- Approximately 34% of all scores occur between the mean and +1 standard deviation
- Approximately 34% of all scores occur between the mean and –1 standard deviation
- Approximately 68% of all scores occur between ± 1 standard deviations
- Approximately 95% of all scores occur between ± 2 standard deviations

Distribution Shapes and Normal Distributions

- 50% of the scores occur above the mean and 50% occur below the mean
- Approximately 34% of all scores occur between the mean and +1 standard deviation
- Approximately 34% of all scores occur between the mean and –1 standard deviation
- Approximately 68% of all scores occur between ± 1 standard deviations
- Approximately 95% of all scores occur between ± 2 standard deviations

Distribution Shapes and Normal Distributions

- 50% of the scores occur above the mean and 50% occur below the mean
- Approximately 34% of all scores occur between the mean and +1 standard deviation
- Approximately 34% of all scores occur between the mean and –1 standard deviation
- Approximately 68% of all scores occur between ± 1 standard deviations
- Approximately 95% of all scores occur between ± 2 standard deviations

Distribution Shapes and Normal Distributions

- The normal distribution is important because it helps simplify the interpretation of test scores
- For example, if a child obtains a raw spelling test score of 35—exactly 1 standard deviation beneath the mean—we know that only 16% of children have lower spelling test scores than this child
- This tells us that the child's spelling ability is "relatively" poor
- The characteristics of the normal distribution provide a convenient model for score interpretation

Distribution Shapes and Normal Distributions

- Few, if any, psychological tests yield "precisely" normal distributions of test scores
- Distributions of actual test scores are typically skewed to some degree

Distribution Shapes and Normal Distributions

- Despite this, scores on many psychological tests are often *approximately* normally distributed
- As a general rule (with many exceptions), the larger the sample size and the wider the range of abilities measured by a test, the more the curve of the test scores will approximate a normal curve
- A classic example of a normally distributed psychological characteristic is intelligence
- Many other psychological characteristics are approximately normal in distribution

Quantifying the Association Between Distributions

- Variability in measurement is an important concept
- However, equally important is the concept of *covariability*
- This refers to the degree to which two distributions of scores vary in a like manner
- Questions of covariability require that a person contribute scores on at least two variables
- For example, if we wanted to examine the association between GPA and IQ scores, we would need a sample in which each participant had taken an IQ test and obtained a GPA

Quantifying the Association Between Distributions

- To be useful, a measure of the covariability between two variables (distributions of scores) must:
  1. indicate the *direction* of the association
  2. indicate the *magnitude* of the association
- With regards direction, we want to know if:
  - relatively high scores on one variable are associated with relatively high scores on the second variable (positive association) or if ...
  - ... relatively high scores on one variable are associated with relatively low scores on the second variable (negative association)
- With regards magnitude, we want to known the strength of the association between two variables

Covariance

- A basic measure of covariability between two distributions of scores is the *covariance*. To calculate the covariance we:
  1. compute the deviation of each score from the mean for each distribution
  2. calculate the "cross-products" $(X\text{-}\overline{X})(Y\text{-}\overline{Y})$ of the deviation scores
  3. calculate the mean of the cross-products
- Formally, the covariance is given by:

$$c_{xy} = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{N}, \quad (5)$$

- Where $\overline{X}$ is the mean of the scores in variable X, and $\overline{Y}$ is the mean of the scores in variable Y

Covariance

Example: calculate the covariance of the following two distributions of scores

IQ (X): 80 110 120 90 130 100 ($\bar{X} = 105$)
GPA (Y): 2.5 2.8 3.2 2.9 3.6 3 ($\bar{Y} = 3$)

$$c_{xy} = \frac{(-25)(-0.5) + (-5)(-0.2) + (-15)(0.2) + \cdots + (-5)(0)}{6}$$

$$c_{xy} = \frac{12.5 + 1 + -3 + \cdots + 0}{6}$$

$$c_{xy} = \frac{31}{6}$$

$$c_{xy} = 5.17$$

Covariance

- The covariance provides clear information about the direction of an association between two variables
- If the covariance is a positive value—as in the example on the previous slide—then the association is a positive one
- If the covariance is a negative value then the association is a negative one
- Unfortunately, the covariance does not provide clear information about the magnitude of the association
- Why so?

Covariance

- The size of the covariance is influenced by the strength of the association between two variables
- However, it is also influenced by the metrics of the two variables
- The covariance between two variables that produce large scores ("large-scale" variables) will tend to be larger than a covariance that involves one or more variables that produce small scores ("small-scale" variables)
- Thus, a larger covariance between IQ and SAT scores than between IQ and GPA does not imply that the former association is stronger than the latter
- This is because IQ and SAT scores are both large-scale measures, whereas GPA is a small-scale measure

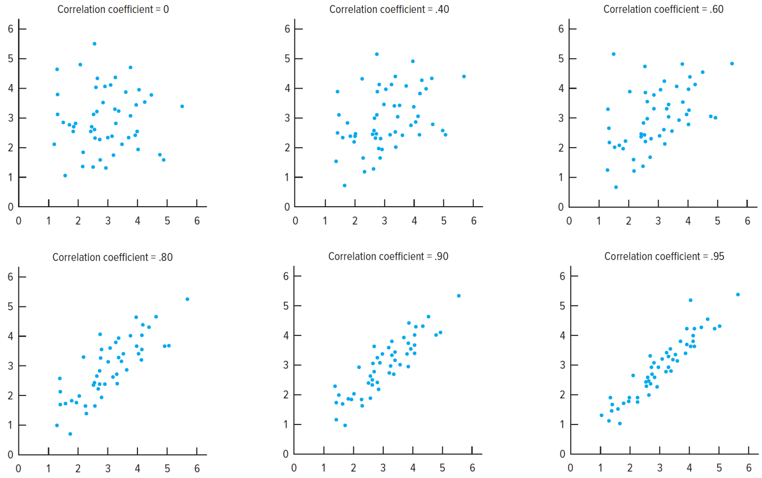
Correlation

- The correlation coefficient provides a clear representation of both the direction and the magnitude of an association between two variables
- Correlation coefficients are bounded within a range of –1 and +1
- A correlation with a value between 0 and +1 indicates a positive association between the two variables
- A correlation with a value between 0 and –1 indicates a negative association between the two variables

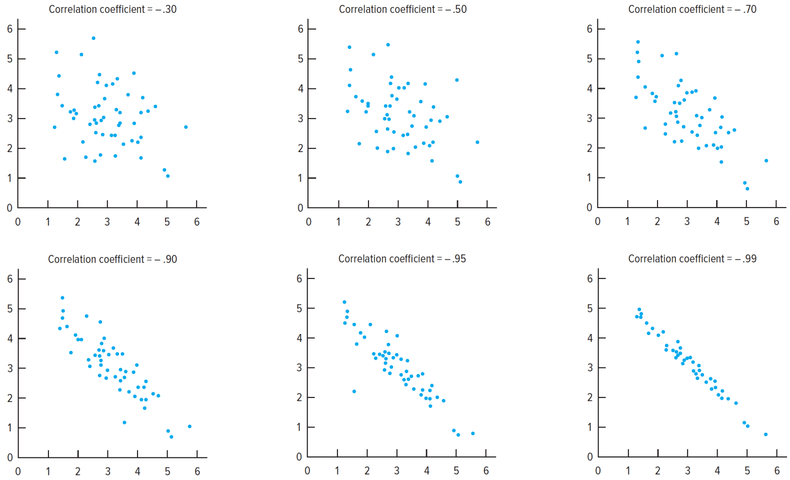
Correlation

- The main benefit of a correlation coefficient is that it clearly reflects the *magnitude* of the association between two variables
- The "boundedness" of correlation coefficients eliminates the influence of metric effects of the two variables on the strength of the association between them
- A correlation coefficient of a specific absolute value (e.g., r_{xy} = –.30 or r_{xy} = .30) represents the same magnitude of association, irrespective of the variables on which the correlation is based
- A large correlation coefficient reflects a strong association, whereas a small correlation coefficient reflects a weak association

Positive Correlations



Negative Correlations



Correlation

- The correlation coefficient is calculated by dividing the covariance c_{xy} by the product of the standard deviations of variables X (σ_x) and Y (σ_y):

$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y}, \tag{6}$$

- For our earlier example involving IQ and GPA, the correlation coefficient is:

$$r_{xy} = \frac{5.17}{(17.08)(.34)},$$

$$r_{xy} = \frac{5.17}{5.81},$$

$$r_{xy} = 0.89$$

Variance for "Composite Variables"

- Test scores are usually based on the sum of two or more items
- For example, the Beck Depression Inventory (BDI) consists of 21 items
- An individual's response to each item is scored on a scale from 0 to 3
- An individual's total score on the BDI is the sum of his/her scores across the 21 items—a score ranging from 0 to 63
- Such "summative scores" are known as *composite scores*

Variance for "Composite Variables"

- The variance of a composite variable can be calculated in the way as shown previously
- We simply sum the items to create our composite variable and then calculate the variance of the composite using equation 2
- However, you should know that the variance of a composite variable is determined by:
  1. the variance of each item within the composite, and
  2. the correlations among the items

Variance for "Composite Variables"

- Formally, the variance of a composite variable $s^2_{\text{composite}}$ containing two items summed together is computed as:

$$s^2_{\text{composite}} = s^2_i + s^2_j + 2\, r_{ij}\, \sigma_i\, \sigma_j, \tag{7}$$

- Where s^2_i is the variance of item i, and s^2_j is the variance of item j
- r_{ij} is the correlation between the two items
- σ_i is the standard deviation of item i, and σ_j is the standard deviation of item j

Variance for "Composite Variables"

Example: calculate the variance of a composite variable made up of the following two items

Item i: 1 2 1 2 3 5 (s_i^2 = 1.889; σ_i = 1.374)
Item j: 1 4 1 2 4 5 (s_j^2 = 2.472; σ_j = 1.572)

r_{ij} = 0.874

$$s^2_{\text{composite}} = 1.889 + 2.472 + 2(.874)(1.374)(1.572)$$

$$s^2_{\text{composite}} = 4.361 + 3.776$$

$$s^2_{\text{composite}} = 8.137$$

Variance for "Composite Variables"

- The important feature of this equation is that the total test score variance depends on item variability and the correlation between item pairs
- As the correlation between the items increases (and is positive), the magnitude of the corresponding composite score variance also increases
- If the two items are uncorrelated, the variance of the composite is simply the sum of the two items' variances

Variance for "Composite Variables"

- The preceding formula is used to calculate the variance of a composite variable created by summing only two items
- When we have more than two items, the variance for a composite variable is calculated by summing the item variance and inter-item covariance matrix
- It sounds horrible, but it's very straightforward—I'll show you how to do this in our lecture on reliability when we discuss a metric known as coefficient α
- Just focus on the calculation for the two item case for now

Interpreting Test Scores

- On most psychological tests, the raw scores on the test are not meaningful and interpretable
- Raw scores do not tell us if a particular score is relatively low, medium, or high
- For example, suppose a student obtains a score of 80 out of 100 on a class exam
- Ostensibly, this is a high score, but what if I told you that all the other students obtained scores $\geqslant 90$—now a score of 80 looks low, rather than, high
- The key point is that a person's test score only makes sense *relative* to the test scores of other people

Standard Scores

- A *standard score* is a raw score that has been converted from one scale to another scale
- The new scale has an arbitrarily set mean and standard deviation
- Raw scores are converted to standard scores because standard scores are more easily interpretable than raw scores
- With a standard score, the position of a test-taker's performance relative to other test-takers is readily apparent

Standard Scores: z Scores

- This type of standard scale score is based on a *zero plus or minus one scale*
- This is because it has a mean of 0 and a standard deviation of 1
- A z score indicates how many standard deviation units the raw score is below or above the mean of a distribution
- A z score is equal to the difference between a given raw score and the mean divided by the standard deviation:

$$z = \frac{X - \overline{X}}{\sigma}, \tag{8}$$

Standard Scores: z Scores

- Let's use the example of the normally distributed spelling data presented earlier and convert a raw score of 65 to a z score:

$$z = \frac{65 - 50}{15}$$

$$z = \frac{15}{15}$$

$$z = 1$$

- A raw score of 65 is therefore equal to a z score of +1 standard deviations above the mean

Standard Scores: z Scores

- Knowing the z-score provides context and meaning that is missing from the raw score
- For example, given our knowledge of areas under the normal curve, we know that only 16% of the other test-takers obtained higher scores, with 84% of test-takers obtaining lower scores
- As well as providing a context for interpreting scores on the same test, z scores provide a context for interpreting scores on different tests
- Suppose a child obtains a score of 42 on an arithmetic test and 24 on a reading test
- On the face of it, the child performed better on the arithmetic test than on the reading test

Standard Scores: z Scores

- However, suppose the child's z scores for the arithmetic and reading tests are –0.75 and +1.32, respectively
- Although the child has a higher raw score on the arithmetic test than the reading test, relative to the other students in his class he performed below average on the arithmetic test, and above average on the reading test
- With z scores, we can determine how much better the child performed on the reading test, and how much worse the child performed on the arithmetic test
- I will show you how to do this shortly when we discuss percentile ranks

Standard Scores: z Scores

- Although z scores are incredibly useful for imposing meaning on test scores, they may be difficult to comprehend for test-users and test-takers
- As a test-taker, how would you stomach being told that your score on your PSYC3302 exam was –1.5?
- It is confusing because not only is the score negative, it is expressed in decimal places
- We can get around this problem by converting z scores into T scores, which are more intuitive to understand

Standard Scores: T Scores

- This type of standard scale score is based on a *fifty plus or minus ten scale*
- This is because it has a mean of 50 and a standard deviation of 10
- The calculation of a T score is a two-step process:
  1. convert an individual's raw score into a z score
  2. convert an individual's z score using the following formula:

$$T = Z(10) + 50, \tag{9}$$

Standard Scores: T Scores

- Returning to our spelling score example, an individual with a score of 65 has a z score of +1
- This individual's T score would therefore be:

$$T = (1)(10) + 50$$

$$T = 10 + 50$$

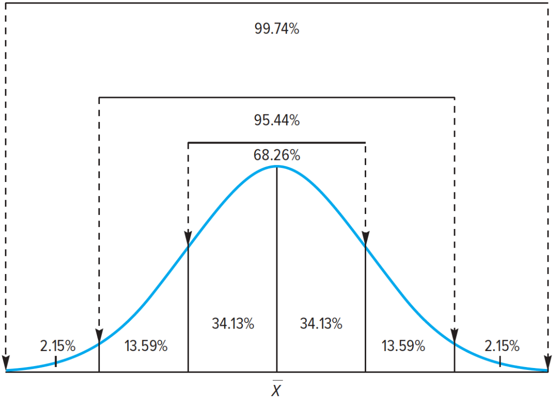
$$T = 60$$

- The T score tells us that the individual is 1 standard deviation above the mean on the spelling proficiency test

Standard Scores: T Scores

- The scale used for computing T scores ranges from 5 standard deviations below the mean to 5 standard deviations above the mean
  - a raw score that falls exactly at –5 standard deviations is equal to a T score of 0
  - a raw score that falls exactly at the mean is equal to a T score of 50
  - a raw score that falls exactly at +5 standard deviations is equal to a T score of 100
- The scale is therefore bounded between 0 and 100
- Thus, unlike z scores, T scores cannot take on negative values, which makes them more intuitive for test-takers

Standard Score Equivalents



σ units	-3σ	-2σ	-1σ		$+1\sigma$	$+2\sigma$	$+3\sigma$
z scores	−3	−2	−1	0	+1	+2	+3
T scores	20	30	40	50	60	70	80

Percentile Ranks

- Another common way of interpreting test scores is in terms of percentile ranks
- Percentile ranks indicate the percentage of scores that fall below and above a specific test score
- There are two ways to determine the percentile rank for an individual's test score
- Suppose we have access to the entire distribution of test scores
- An individual's percentile rank is the number of scores in the distribution that are lower than the individual's score divided by the total number of scores

Percentile Ranks

- For example, in the class exam data presented at the outset Barbara obtained an exam score of 82
- Barbara's score is higher than 14 out of 25 people's exam scores
- Her percentile rank is therefore: (14/25)(100) = 56%
- Thus, Barbara scored at the 56th percentile

Percentile Ranks

- Sometimes we do not have access to the entire distribution of scores
- However, if we know the mean and standard deviation—and that the underlying distribution of scores is relatively normal—we could compute a z score for an individual and link it to a percentile
- We can look up the z score in a standard normal distribution table detailing distances under the normal curve (look at the back of any good statistics textbook)
- The table tells us the percentage of cases that could be expected to fall above or below a particular standard deviation point (or z score)

Percentile Ranks

- Let's calculate the percentile rank for a raw score with a z score of +1.5
- The table tells us that the area of the normal curve between the mean and a z score of +1.5 is 43.32%
- But remember that 50% of the scores in a normal distribution fall between the mean and –3 standard deviations
- The person's percentile rank is therefore 50% + 43.32% = 93.32%
- This indicates that 93% of other test-takers scores fall below this person

Percentile Ranks

- What if the z score is negative?
- In this case, we subtract the value obtained from the table from 50%
- Suppose the z score is –1.5 rather than +1.5
- The percentile would be 50% – 43.32% = 6.68%
- This tells us that only 6.68% of the scores in a distribution fall below this person

Normalised Scores

- When a test developer produces a new test—e.g., of intelligence—it is desirable that the test yield normally distributed measurements
- Often this is to provide *norms* (interpretive guides) for test users
- These norms are based on a large *reference sample*—also known as a *normative sample*—thought to be representative of some population
- During test development, test constructors administer their test to the reference sample to obtain scores that can serve as an interpretative "frame of reference"

Normalised Scores

- What happens if the reference sample data is skewed?
- One option is to *normalise* the distribution
- This involves "stretching" the skewed curve into the shape of a normal distribution and creating a corresponding scale of standard scores (e.g., T scores)—known as a *normalised standard score scale*
- The textbook describes one (of several) procedures for normalising a distribution

Normalised Scores

- However, it is generally preferable to "fine-tune" a test to yield approximately normally distributed measurements than to normalise distributions
- This is because there are technical cautions to observe before normalising a skewed distribution
- For example, transformations should only be attempted when the test sample is large enough and representative enough
- If doubts exist about sample size and representativeness, failure to obtain normally distributed scores may not be the fault of the measuring instrument

A Final Note on Score Conversions

- You might be wondering whether it is legitimate to convert test scores using the procedures just described (z, T, and normalised scores, percentile ranks)
- These raw score conversion methods are based on sound statistical logic
- They do not represent attempts to "cook" the data—to paint a distorted picture of test scores
- On the contrary, the aim is to make the data more readily interpretable and meaningful for test-users and test-takers

Test Norms

- Earlier, we touched on the notion of a reference sample
- The point was made that the reference sample for a test must be representative of the population of interest
- For example, in IQ testing, you would want a reference sample that consists of people from a wide range of walks of life—e.g., gender, education, socioeconomic status, etc.
- Good IQ tests have reference samples that correspond to that countries bureau of statistics information
- If about 20% of the adult population has a university degree, then the test developer will want a reference sample with about 20% of people with a university degree

Probability vs. Non-Probability Sampling

- *Probability sampling* uses procedures that ensure a representative sample
- Random sampling is a type of probability sample, as you would expect a random sample from the population to be representative of the population
- However, random sampling rarely ever happens in practice
- Many test norms are based on *non-probability samples*
    - e.g., the normative sample may consist of all the people who have ever taken the test
    - obviously, there would be some self-selection bias in such a process

In Next Week's Lab ...

- Gentle introduction to SPSS
- Read the Andy Field textbook chapter "SPSS Environment" before attending your lab (located in the week 3 lab folder)

In Next Week's Lecture ...

- Introduction to reliability concepts:

  1. Theoretical basis of reliability
  2. Empirical estimates of reliability

References

Furr, M. R., & Bacharach, V. R. (2014; Chapter 3). Psychometrics: An Introduction (second edition). Sage.