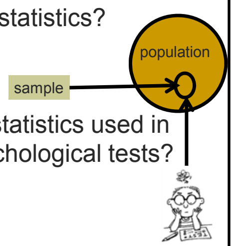
EDS 245: Psychology in the Schools

Stephen E. Brock, Ph.D., NCSP

Descriptive Statistics for Norm Referenced Tests

Small Group Discussion

- What are descriptive statistics?
- How are descriptive statistics used in norm referenced psychological tests?



Lecture Overview

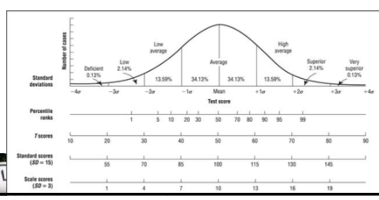
- Introduction
- Scales of Measurement
- Frequency Distribution
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- Norming a Test

Introduction

- Raw test data is rarely valuable to the school psychologist.
  - A raw score of 5 on the *WISC* Information subtest may mean different things for different students.
    - For example, a raw score of 5 for a 6-year-old will be suggestive of a different level of cognitive functioning than will the same score for a 7-year-old.
  - In addition, a raw score of 5 on one test will not have the same meaning as a raw score of 5 on another test.

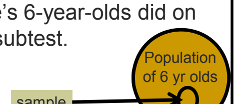
Introduction

- Thus, the raw scores obtained via psychological tests are most commonly interpreted by reference to ***norms*** and by their conversion into some relative reference score or ***derived score*** (a descriptive statistic).



Introduction

- ***Norms*** represent the test performance of individuals within a standardization sample.
  - For example, they document how well the standardization sample's 6-year-olds did on the *WISC* Information subtest.

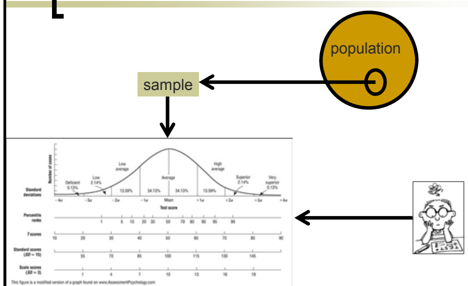


- What does – " a test is only as good as its standardization sample" mean?

Introduction: Why use derived scores

1. **Derived scores** are the descriptive statistics used to transform raw test data into a number that more precisely illustrates a student's exact position ***relative to individuals in the normative group***.
   - For example, at age 6, a raw score of 5 on the *WISC* Information subtest corresponds to a Scaled Score of 10. While at age 7, this same raw score corresponds to a Scaled Score of 6.
     - The WISC subtests have a mean of 10 and a standard deviation of 3

Introduction: Why use derived scores



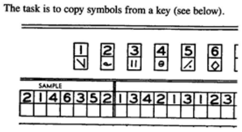
Introduction: Why use derived scores

2. **Derived scores** also provide comparable measures that allow direct comparison of a student's performance on different tests. Thus, allowing the psychologist to identify a relative pattern of ***unique strengths and weaknesses***.
   - For example, a Scaled Score of 10 on the Information Subtest (RS = 5) can be directly compared to a Scaled Score of 3 on the Coding Subtest (RS = 5).

Information & Coding Subtests

Coding

The task is to copy symbols from a key (see below).



Age 6
Raw Score = 5
Scaled Score = 3

Information (30 questions)

How many legs do you have?
What must you do to make water freeze?
Who discovered the North Pole?
What is the capital of France?

Age 6
Raw Score = 5
Scaled Score = 10

Introduction

- Understanding the conversion of raw scores into derived scores, and how they are used to describe a students performance relative to others, as well as a student's own unique pattern of strengths and weaknesses, requires knowledge of ***basic statistical concepts***.

Introduction

- Basic statistical concepts underlie the development and utilization of norms.
- It is critical that school psychologists, who use psychological tests, have a solid understanding of descriptive statistics.

"Data don't make any sense, we will have to resort to statistics."

Lecture Overview

- Introduction
- **Scales of Measurement**
- Frequency Distribution
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- Norming a Test

Scales of Measurement

Scale	Properties	E.G.
Nominal (to **name**)	Data represents qualitative or equivalent categories (not numerical).	Eye color Gender Race
Ordinal (to **order**)	Numerically ranked, but has no implication about how far apart ranks are.	Grades Grade Equalivants Percentile Ranks Stanines
Interval (**equal**)	Numerical value indicates rank and meaningfully reflects relative distance between points on a scale	Z Scores Standard Scores (Deviation IQ) T-Scores Scale Scores
Ratio (equal)	Has all the properties of an interval scale, and in addition has a true zero point.	Length Weight

Lecture Overview

- Introduction
- Scales of Measurement
- **Frequency Distribution**
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- Norming a Test

Frequency Distribution

- The first step in transforming the raw scores of a standardization sample into derived scores is to place the scores into a frequency distribution.

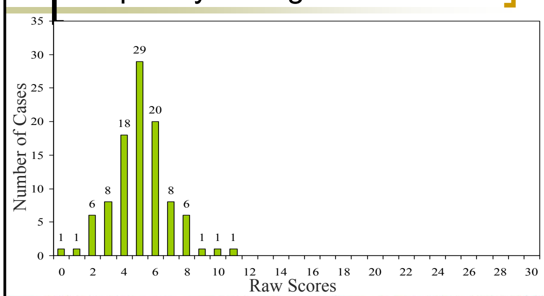
Frequency Distribution

- A sample of 100 carefully chosen 6-year-olds were given an 30 item Information test. Here are their raw scores in a frequency distribution.

RS	Freq.	RS	Freq.
0	1	7	8
1	1	8	6
2	6	9	1
3	8	10	1
4	18	11	1
5	29	12-30	0
6	20		

- Convert this distribution into an Frequency Histogram.

Frequency Histogram



Lecture Overview

- Introduction
- Scales of Measurement
- Frequency Distribution
- **Measures of Central Tendency**
- Measures of Variability
- The Normal Curve
- Norming a Test

Measures of Central Tendency

- Mode
  - Determined by looking at a set of scores and seeing which occurs most frequently.
  - Only appropriate statistic for nominal data.
- Median
  - The point above and below which 50% of the scores are found.
  - Appropriate for use when the data is ordinal.
- Mean
  - The arithmetic average of the scores
  - Appropriate for use when the data is interval or ratio.

Measures of Central Tendency

- What is the Mean, Median, and Mode of the data summarized in your Frequency Histogram?

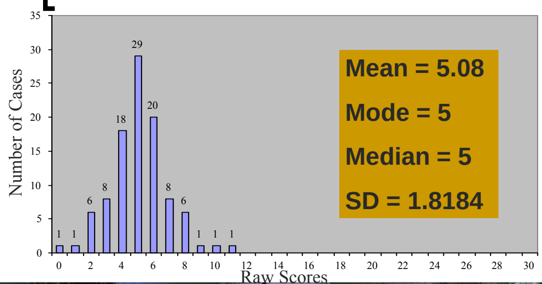
RS	Freq.
0	1
1	1
2	6
3	8
4	18
5	29
6	20

RS	Freq.
7	8
8	6
9	1
10	1
11	1
12-30	0

$$\text{Mean for grouped data} = \frac{\sum fX}{n}$$

f = frequency
x = score
n = total number of scores

Frequency Histogram



Lecture Overview

- Introduction
- Scales of Measurement
- Frequency Distribution
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- Norming a Test

Measures of Variability

- Range
  - The difference between the highest and the lowest score.
  - A quick estimate of variability.
- Standard Deviation (SD)
  - The square root of the variance.
  - The best estimate of variability.
  - Small SD indicates the scores are close together (little variability)
  - Large SD indicates the scores are far apart (large variability)

Range and Standard Deviation

- While range does give us information about the dispersion of scores, its is of limited usefulness.
- The most common measure of dispersion is the standard deviation (*SD*) and is required to obtain derived scores.
- An example…
  - A test developer is interested in standardizing a measure of Arithmetic for 6-0 year olds. To do this she collects data for 20 carefully selected children.
  - Results were as follows:

Range and Standard Deviation

Subject	Raw Score	Subject	Raw Score
1	64	11	60
2	48	12	43
3	55	13	67
4	68	14	70
5	72	15	65
6	59	16	55
7	57	17	56
8	61	18	64
9	63	19	61
10	60	20	60

Standard Deviation

$$\text{s.d.} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}}$$

where ΣX^2 = the sum of the squared score values

$(\Sigma X)^2$ = the square of the sum of all the scores

N = the total number of scores used in the computation

Standard Deviation

1. Add all the Arithmetic raw scores
   - 64 + 48 + … + 60 = **1208 (÷ 20= 60.4 = mean)**
2. Square all the scores, and add the squared values.
   - $64^2 + 48^2 + \ldots + 60^2$ = **73,894**
3. Square the sum obtained in step 1 (1208), and divide this value by the number of scores that were added to obtain the sum (20)
   - $1208^2 \div 20$ = 1,459,264/20 = **72,963**
4. Subtract the value obtained in step 3 (72,963) from the sum in step 2 (73,894)
   - 73,894 - 72,963 = **931**

Standard Deviation

5. Divide the value obtained in step 4 by N - 1 (in this example, it would be 20 - 1)* The resultant value is called the variance.
   - **931 ÷ 19 =** 49
6. Take the square root of the value obtained in step 5. This value is the standard deviation.
   - $\sqrt{49} = 7$

Mean = 60.4; Standard Deviation = 7, with this information raw scores can be converted into derived or standard scores. These scores express how many standard deviations a given raw score lies above or below the mean of the distribution.

* Dividing by "n - 1" yields the unbiased ***estimate*** of the of the ***population*** variance rather than the actual variance of the sample. If the actual sample variance is desired (or your scores reflect performance of a population), divide by N rather than n - 1.

Using Derived (or Standard) Scores

- You are now using the Information and Arithmetic tests that have just been developed. The 6-year-old student you are testing obtains a **Arithmetic Test raw score of 67** and a **Information Test raw score of 3**.
- How will you know what these two scores mean *relative to the standardization sample* and *relative to the student himself* (i.e., do either scores represent individual strengths or weaknesses)?
- The answer is to transform them into standard scores.

Arithmetic Test Standard Score

- Z Score Formula

$$\frac{X - \text{Mean}}{sd} = \text{Z Score}$$

$$\frac{67 - 60.4}{7} = +0.943$$

Information Test Standard Score

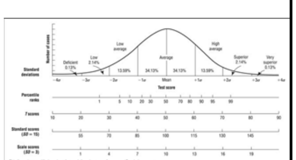
- Z Score Formula

$$\frac{X - \text{Mean}}{sd} = \text{Z Score}$$

$$\frac{3 - 5.08}{1.8184} = -1.144$$

Interpretation

- Arithmetic
  - Raw Score, 67
  - *Z* Score, +0.943
- Information
  - Raw Score, 3
  - *Z* Score, -1.144



Interpretation

- Typically, *Z* Scores are transformed into other standard scores to eliminate + and - signs and decimal points.
  - SS = $Mean_{ss} + (S_{ss})(z)$
    - **T Score = 50 + 10(*z*)**
      - Arithmetic RS of 67, 50 + 10(*0.943*) = 59.43 (67 = 59)
      - Information RS of 3, 50 + 10(*-1.144*) = 38.56 (3 = 39)
    - **Scaled Score = 10 + 3(*z*)**
      - Arithmetic RS of 67, 10 + 3(*0.943*) = 12.83 (67 = 13)
      - Information RS of 3, 10 + 3(*-1.144*) = 6.57 (3 = 7)
    - **Deviation IQ = 100 + 15(*z*)**
      - Arithmetic RS of 67, 100 + 15(*0.943*) = 114.15 (67 = 114)
      - Information RS of 3, 100 + 15(*-1.144*) = 82.84 (3 = 83)

Interpretation

- What if ….
  - Arithmetic RS = 57
  - Information RS = 10
- First obtain the z-scores

$$\frac{X - Mean}{sd} = Z$$

- Then convert them to scaled scores
  - SS = $Mean_{ss} + (S_{ss})(z)$
- Finally, make an interpretation of these data.

Interpretation

- All of these computations take a lot of time so test developers typically develop norm tables that take a given raw score and tell you what the corresponding derived score is.
- For example…

WISC-R Scaled Score Equivalents of Raw Scores

6 yrs. 0 mos. 0 days
through
6 yrs. 3 mos. 30 days

118

Table 19
Scaled Score Equivalents of Raw Scores

	VERBAL							PERFORMANCE						
Scaled Score*	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Scaled Score*	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Scaled Score*
1	—	—	—	0	—	0	1	0	—	—	0	0-1	—	1
2	0	—	—	1	—	1	2	1	—	—	1	2-3	—	2
3	1	0	0	2-3	0	—	3	2	0	0	2	4-5	0	3
4	—	—	—	4-5	1	2	4	3	—	1	3	6-7	1	4
5	2	1	1	6	2	—	5	4	1	—	4	8-9	2	5
6	—	—	2	7-8	3	3	6	—	—	2	5	10-12	3	6
7	3	2	3	9	4	—	7	5	2	—	—	13-16	4	7
8	—	3	—	10	—	4	8	6-7	3-4	3	6	17-20	5-6	8
9	4	4-5	4	11-12	5	5	9	8-9	5-6	4-5	7-8	21-24	7-9	9
10	5	6	5	13	6-7	6	10	10	7-9	6-8	9-11	25-28	10-11	10
11	—	7	6	14-15	8-9	7	11	11	10-12	9-11	12-13	29-32	12-13	11
12	6	—	—	16	10	8	12	12	13-14	12-14	14-16	33-36	14-15	12
13	7	8	7	17-18	—	9	13	13	15-17	15-17	17-18	37-40	16-17	13
14	8	9	—	19	11	10	14	14-15	18-20	18-21	19	41-43	18-19	14
15	9	10	8	20-21	12	11	15	16	21-22	22-25	20	44-45	20	15
16	10	11	—	22	13	12	16	17-18	23-25	26-28	21	46	21	16
17	11	12	9	23	14-15	13	17	19	26-27	29-30	22	47	22	17
18	12	13-14	10	24-25	16	14	18	20	28-29	31	23	48	23	18
19	13-30	15-30	11-18	26-64	17-34	15-28	19	21-26	30-48	32-62	24-33	49-50	24-30	19

*The scaled scores for the *WISC-R* tests extend from 1 to 19, which differs from the 0-20 range used for the 1949 *WISC*. See pages 21 and 23 for explanation.

Lecture Overview

- Introduction
- Scales of Measurement
- Frequency Distribution
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- Norming a Test

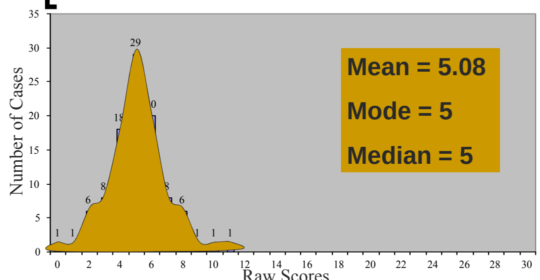
Variability and the Normal Curve

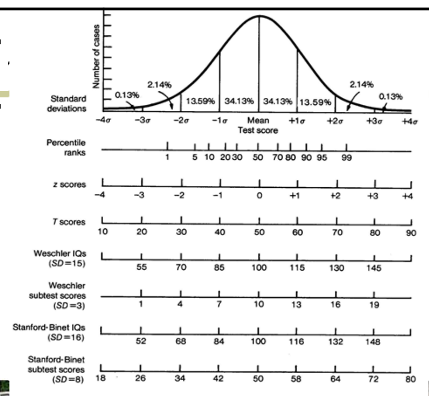
- When population scores on a particular characteristic are graphed, the shape of the "normal curve" resembles a bell.
- The majority of scores fall in the middle (near the mean), and a few scores fall at the extreme ends of the curve.
  - Extreme scores are very unusual!!!
- The height of a "normal curve" will be determined by the variability of the scores.

The Normal Curve

- If a variable is normally distributed it falls in a normal or bell shaped curve.
- Characteristics
  - 50% of scores are above/below the mean
  - Mean, median, mode have the same value (a reason for looking at all three)
  - Most scores are near the mean. Fewer scores are away from the mean.
  - The same number of scores are found + and - a standard deviation from the mean.

Frequency Histogram





The Use of Descriptive Statistics in Norming a Test

The following is a very a basic and simplified description of how descriptive statistics are used in the development of a norm referenced test.

These tests compare the performance of student you are testing to the performance of children in a standardization sample (How are they different from criterion referenced tests?).

The standardization sample provides an estimate of the performance of a population.

Thus, the obtained test score gives you an idea of how well a student did on a given test relative to a specific population (by comparing your testing subject to the standardization sample's performance).

Lecture Overview

- Introduction
- Scales of Measurement
- Frequency Distribution
- Measures of Central Tendency
- Measures of Variability
- The Normal Curve
- **Norming a Test**

The Use of Descriptive Statistics in Norming a Test

1. A test developer wants to construct a test that estimates the how individuals within a population perform in the area of short term auditory memory for numbers.

Memory for Numbers Test

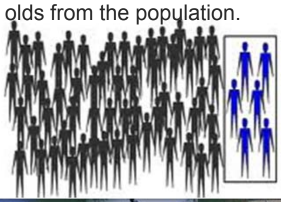
1.	23	35
2.	472	495
3.	8732	9436
4.	24372	38512
5.	629471	4938752
6.	2948576	8957383
7.	29485694	598327291
8.	485938569	3857265395

Directions
State: *I am going to say some numbers when I am done I want you to say them back to me.*
Read numbers at a rate of one digit per second.

Raw Score Range
0 to 16

The Use of Descriptive Statistics in Norming a Test

2. To begin norming the test she administers the *Memory for Numbers Test* (MNT) to a large and carefully selected sample of 6 year olds from the population.

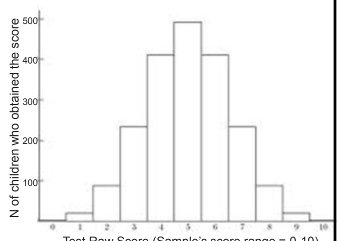


CALIFORNIA STATE UNIVERSITY

The Use of Descriptive Statistics in Norming a Test

3. Raw scores are placed in a frequency histogram.

The sample's performance on the test (the frequency with which specific raw scores are obtained) is "normally distributed. The mean, median, and mode area all about the same score.



CALIFORNIA STATE UNIVERSITY

The Use of Descriptive Statistics in Norming a Test

4. Raw scores from the sample are used to estimate the average (mean or $\bar{X}$) performance of the population's 6-year-olds on the *MNT*.

   The sum or all raw scores (or $\sum X$) is obtained and divided by the number of children who took the test (sample size or *N*).

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

CALIFORNIA STATE UNIVERSITY

The Use of Descriptive Statistics in Norming a Test

5. The test developer now knows the average score (let's pretend it is a raw score of 5 = $\bar{X}$) of this sample (and now has an estimate of the population's typical performance).
   - In other words, she now knows exactly how to describe the performance of children who obtain a raw score of 5 (these children are "average," their score falls at the 50th percentile rank).
   - She also knows that raw scores above 5 are above average and scores below 5 are below average.
     - But exactly how far above average is a raw score of 7??

The Use of Descriptive Statistics in Norming a Test

5. To determine exactly how far above or below average (or the mean) any other given raw score was, the test developer will need to determine how spread apart the scores of the sample are. In other words, she will need to determine how different the scores are from one another. Are they all very similar to each other (homogeneous) or are the all very different from each other (heterogeneous)?

   What descriptive statistic will answer this question?

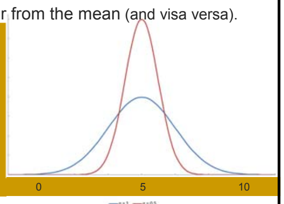
The Use of Descriptive Statistics in Norming a Test

6. The standard deviation is needed.
   - If the standard deviation relatively large, then relatively larger differences in raw scores will be required to be farther from the mean (and visa versa).

$$\sqrt{\frac{\Sigma(X - \bar{X})^2}{(n - 1)}}$$

where:

- X = each score
- $\bar{X}$ = the mean or average
- n = the number of values
- Σ means we sum across the values



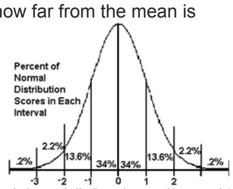
The Use of Descriptive Statistics in Norming a Test

7. With knowledge of the mean (e.g., 5) and standard deviation (let's pretend it is 1.8) of the sample's performance on the test we can determine exactly how far from the mean a given raw score is.
   What descriptive statistic do we need to do this?

The Use of Descriptive Statistics in Norming a Test

8. A Z-score is needed
   - With the Z-score we will now be able to determine exactly how far from the mean is a given raw score.

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



Calculating a Z Score: http://psych.colorado.edu/~mcclella/java/normal/normz.html
Interpreting a Z Score: http://www.zscorecalculator.com/index.php

The Use of Descriptive Statistics in Norming a Test

9. The test developer does not like the fact that Z-scores make use of "+" and "-" signs (as the can easily be missed). She would also like to get ride of decimal points. Thus, she transforms the Z-Scores into Scaled Scores
   - SS = $\text{Mean}_{ss} + (S_{ss})(z)$
     - **Scaled Score = 10 + 3(*z*)**
       - *MNT* RS of 7, 10 + 3(*1.111*) = 13.333 (7 = 13)
   - Develop a norm table for *MNT* Raw scores
     1. Compute Z-scores for each raw score
     2. Transform the Z-Scores to Scaled Scores

The Use of Descriptive Statistics in Norming a Test

Raw Score	Scaled Score	Raw Score	Scaled Score
0		6	
1		7	13
2		8	
3		9	
4		10	
5	10	11-16	

The Use of Descriptive Statistics in Norming a Test

First we need to know how far is each raw score from the sample's mean of 5?

Raw Score	Z-Score	Raw Score	Z-Score
0	-2.777	6	+0.555
1	-2.22	7	+1.111
2	-1.666	8	+1.666
3	-1.111	9	+2.222
4	-0.555	10-16	+2.777
5	0		

$$Z = \frac{\text{score} - \text{sample mean}}{\text{Standard deviation}} \longrightarrow Z = \frac{X - 5}{1.8}$$

The Use of Descriptive Statistics in Norming a Test

What are the scales scores for each raw score?

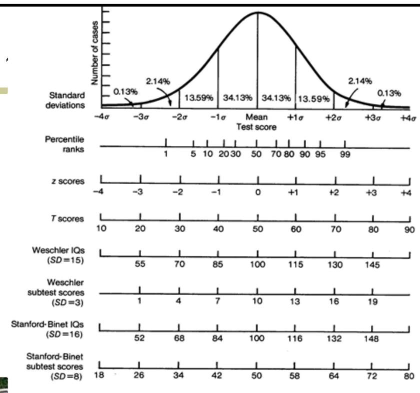
Raw Score	Scaled Score	Raw Score	Scaled Score
0		6	
1		7	13
2		8	
3		9	
4		10	
5	10	11-16	

Z-scores can be transformed into scaled scores

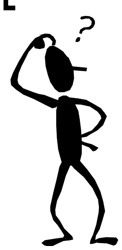
$SS = Mean_{SS} + (S_{SS})\ (z)$ $13.33 = 10 + (3)\ (1.111)$

The Use of Descriptive Statistics in Norming a Test

Raw Score	Scalded Score	Raw Score	Scaled Score
0	2	6	12
1	3	7	13
2	5	8	15
3	7	9	17
4	8	10-16	18
5	10		



Questions?



Next Week (11/13/12): Test Interpretation

No papers due

Optional reading Lyman (1998)