Analyzing Reliability and Validity in Outcomes Assessment (Part 1)

Robert Lingard and Deborah K. van Alphen

California State University, Northridge

Overview

- Introduction
- Fundamental Concepts
- Group Problem
- Statistical Preliminaries
- Software Tools
- Probability
- Inferential Statistics
- Hypothesis Testing
- Summary

Introduction – Basic Questions

- How can we make assessment easier?
  - Minimize the effort required to collect data
- How can we design an assessment to provide useful information?
  - Use tools to evaluate the approach
- How can we learn more from the assessment results we obtain?
  - Use tools to interpret quantitative data

Making Assessment Easier

- Depend on assessments at the College or University level.
- Assess existing student work rather than creating or acquiring separate instruments.
- Measure only a sample of the population to be assessed.

Designing Assessments to Produce Useful Information

- How should the students or student work to be measured be selected?
- Is the assessment approach stable and consistent?
- Does the assessment instrument measure what is intended to be assessed?

Learning More from Assessment Results

- Was the sample used representative and large enough?
- Were the instrument and process utilized dependable?
- Were the results obtained meaningful?
- Is a difference between two results significant?

Fundamental Concepts

- Sampling
- Reliability
- Validity
- Correlation

Definition of Terms Related to Sampling

- **Data**: Observations (test scores, survey responses) that have been collected
- **Population**: Complete collection of all elements to be studied (e.g., all students in the program being assessed)
- **Sample**: Subset of elements selected from a population
- **Parameter**: A numerical measurement of a population
- **Statistic**: A numerical measurement describing some characteristic of a sample

Sampling Example

There are 1000 students in our program, and we want to study certain achievements of these students. A subset of 100 students is selected for measurements.

Population = 1000 students

Sample = 100 students

Data = 100 achievement measurements

Methods of Sample Selection

- **Probability Sampling**: The sample is representative of the population
- **Non-Probability Sampling:** The sample may or may not represent the population

Probability Samples

- Random sample: Each member of a population has an equal chance of being selected.
- Stratified random sample: The population is divided into sub-groups (e.g., male and female) and a random sample from each sub-group is selected.
- Cluster sample: The population is divided into clusters and a random sample of clusters is selected.

Non-Probability Samples

- Convenience sample: The sample that is easily available at the time.
- Network sample: The sample is constructed by finding members of the population through the contacts of a known member.
- Quota sample: The sample takes available subjects, but it attempts to ensure inclusion of representatives from certain elements of the population.

Problems with Sampling

- The sample may not be representative of the population.
- The sample may be too small to provide valid results.
- It may be difficult to obtain the desired sample.

Reliability and Validity

- *Reliability* refers to the stability, repeatability, and consistency of an assessment instrument. (i.e., how good are the operational metrics and the measurement data).
- *Validity* refers to whether the measurement really measures what it was intended to measure (i.e., the extent to which an empirical measure reflects the real meaning of the concept under consideration).

Reliability

- Tests of stability: If an instrument can be used on the same individual more than once and achieve the same results, it is stable.
- Tests of repeatability: If different observers using the same instrument report the same results, the instrument is repeatable.
- Tests of internal consistency: If all parts of the instrument measure the same concept, it is internally consistent.

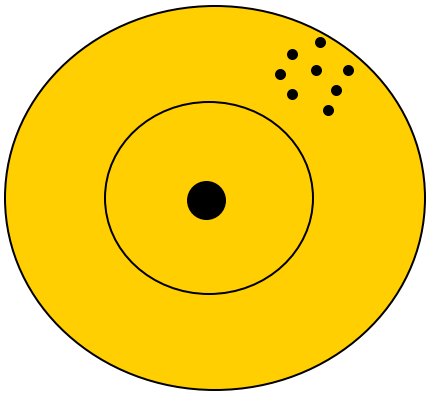
Validity

- Self-evident measures: Does the instrument appear to measure what it is supposed to measure.
  - Face validity: "It looks all right on the face of it."
  - Content validity: The validity is estimated from a review of literature on the topic or through consultation with experts.

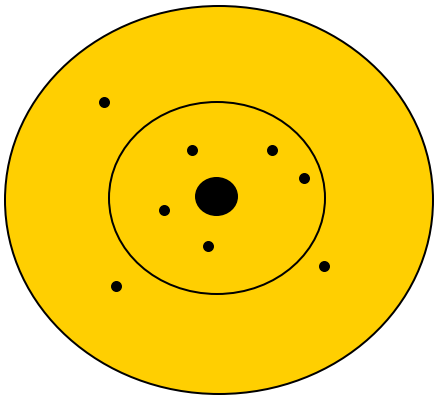
Validity (Cont'd)

- Pragmatic measures: These test the practical value of a particular instrument.
  - Concurrent validity: The results have high correlation with an established measurement.
  - Predictive validity: The results predicted actually occur.
  - Construct validity: When what you are attempting to measure is not directly observable, and the results correlate with a number of instruments attempting to measure the same construct.

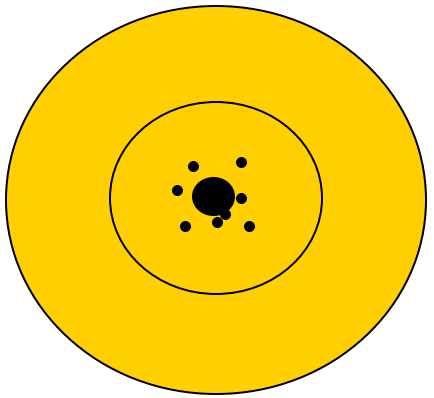
Reliability and Validity



Reliable but
not valid



Valid but
not reliable

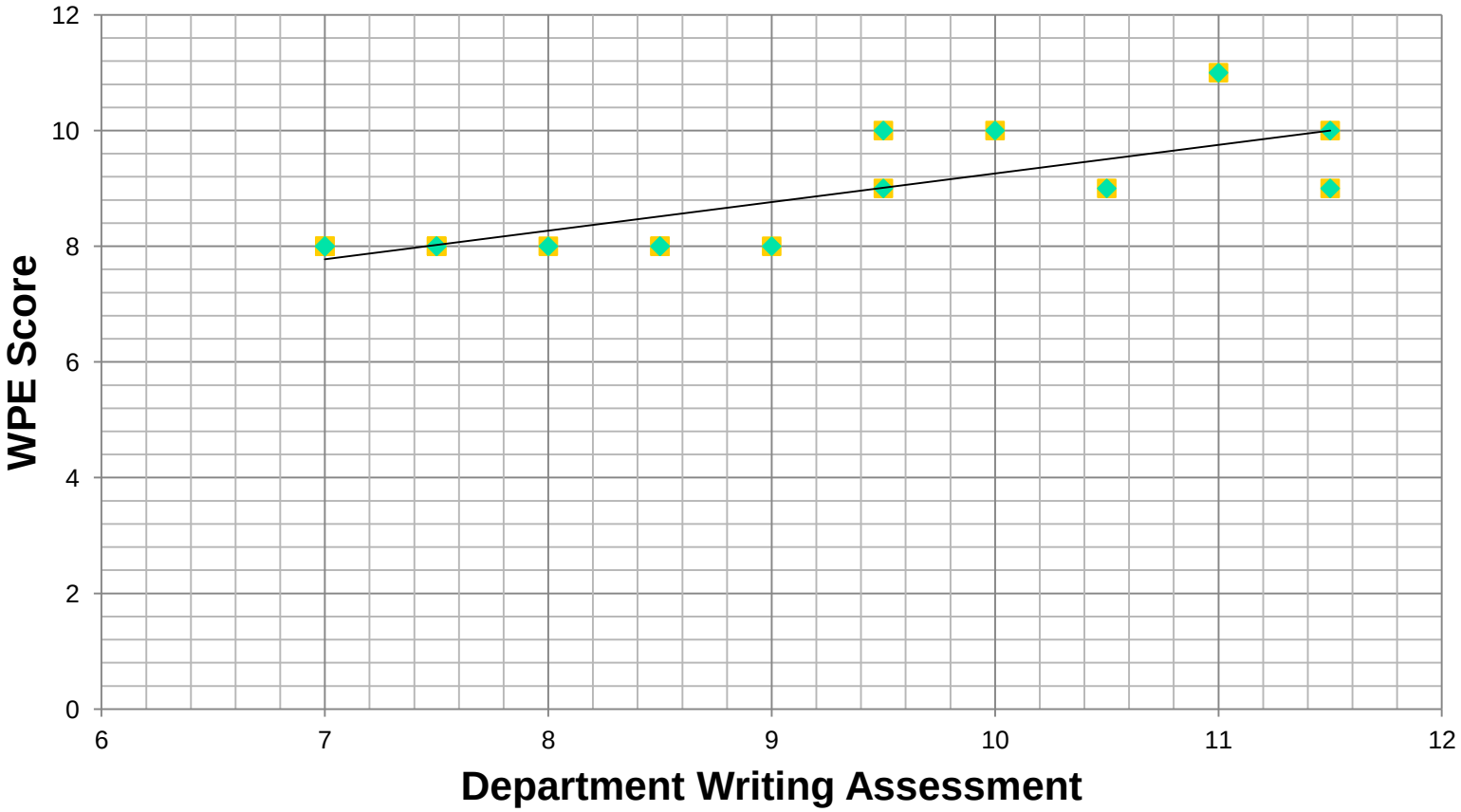


Reliable
& valid

Correlation

- Correlation is probably the most widely used statistical method to assess relationships among observational data.
- Correlation can show whether and how strongly two sets of observational data are related.
- Can be used to show reliability or validity by attempting to correlate the results from different assessments of the same outcome.

Example Correlation



Group Problem

- Assume your goal is to assess the written communication skills of students in your program. (Assume the number of students in the program is large and that you already have a rubric to use in assessing student writing.)
- Working with your group devise an approach to accomplish this task.

Group Problem (Cont'd)

- Specifically, who would you assess and what student produced work items would you evaluate, i.e., how would you construct an appropriate sample of students (or student work) to assess?
- Identify any concerns or potential difficulties with your plan, including issues of reliability or validity.
- What questions are you likely to have regarding the interpretation of results once the assessment is completed?

Objective of This Presentation

- Show how statistical analysis can be used to:
  - Simplify assessment using sampling
  - Measure reliability of assessment instruments and approaches
  - Measure validity of assessment results
  - Quantify program improvements through comparison of assessment results

Basic Statistical Concepts

- Central tendency for data
- Frequency distribution of data
- Variation among data

Measure of Central Tendency: Mean

- n = number of observations in a sample
- $x_1, x_2, \ldots, x_n$ denotes these n observations
- $\bar{x}$, the sample **mean**, is the most common measure of center
- $\bar{x}$ (a statistic) is the arithmetic mean of the n observations:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- μ represents the population mean, a parameter

Frequency Distribution of Data

The tabulation of raw data obtained by dividing the data into groups of some size and computing the number of data elements falling within each pair of group boundaries

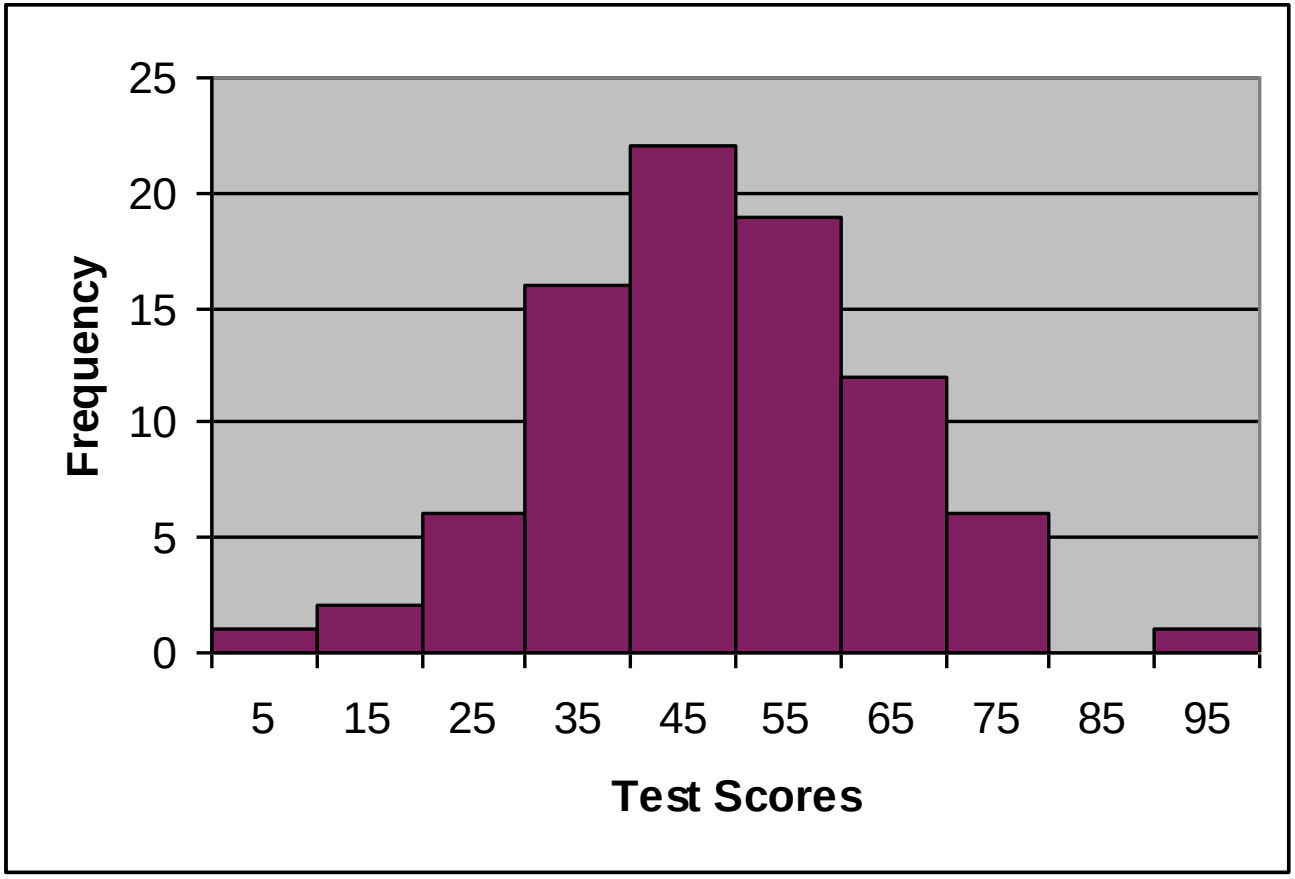
Frequency Distribution – Tabular Form

Group Interval	Frequency	Relative Frequency
0.00-9.99	1	1.18%
10.00-19.99	2	2.35%
20.00-29.99	6	7.06%
30.00-39.99	16	18.82%
40.00-49.99	22	25.88%
50.00-59.99	19	22.35%
60.00-69.99	12	14.12%
70.00-79.99	6	7.06%
80.00-89.99	0	0.00%
90.00-100.00	1	1.18%

Histogram

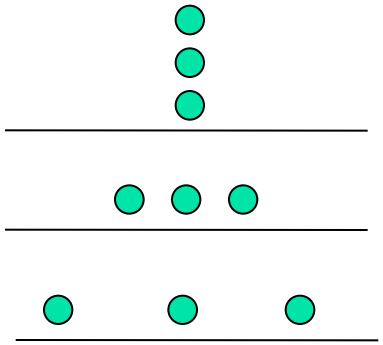
- A histogram is a graphical display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size. In the most common form of histogram, the independent variable is plotted along the horizontal axis and the dependent variable is plotted along the vertical axis.

Frequency Distribution -- Histogram



Variation among Data

- **The following three sets of data have a mean of 10:**
  - {10, 10, 10}
  - {5, 10, 15}
  - {0, 10, 20}



- **A numerical measure of their variation is needed to describe the data.**
- **The most commonly used measures of data variation are:**
  - Range
  - Variance
  - Standard Deviation

Measures of Variation: Variance

- **Sample of size n: x_1, x_2, …, x_n**
- **One measure of positive variation is** $(x_i - \bar{x})^2$
- **Definition of sample variance**

  (sample size = n): $$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$
- **Definition of population variance**

  (population size = N): $$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Measures of Variation: Standard Deviation

- **Sample Standard Deviation:** $$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- **Population Standard Deviation:** $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

- The units of standard deviation are the same as the units of the observations

Measures of Variation: Variance and Standard Deviation

The following data sets each have a mean of 10.

Data Set	Variance	Standard Deviation
10, 10, 10	(0+0+0)/2 = 0	0
5, 10, 15	(25 + 0 + 25)/2 = 25	5
0, 10, 20	(100 + 0 + 100)/2 = 100	10

↑

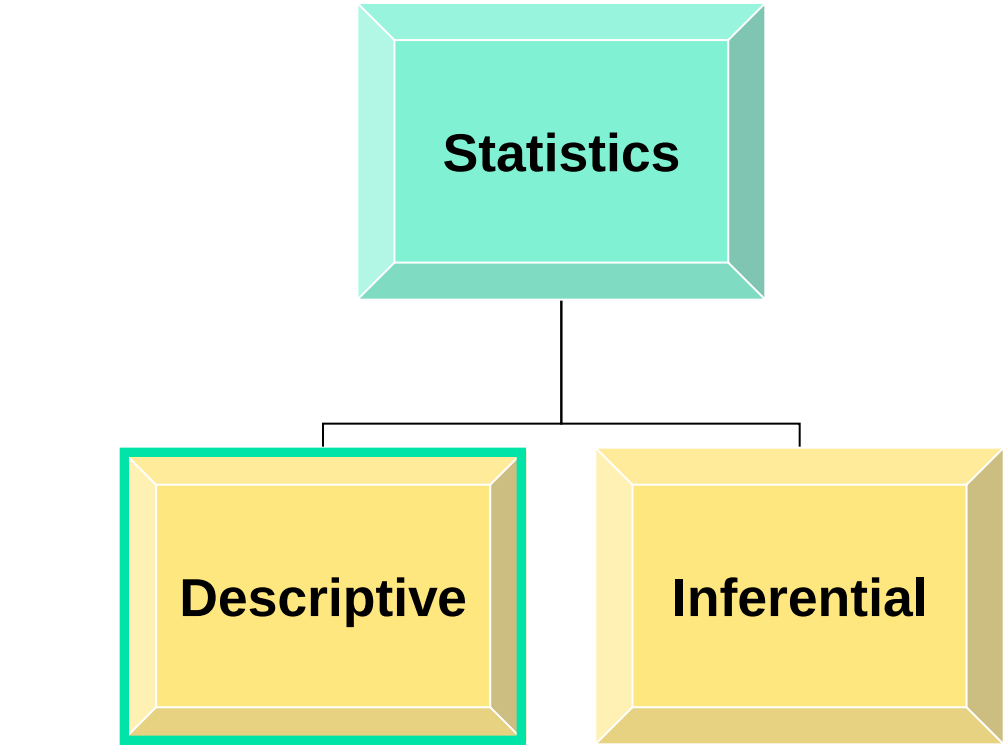
Good measure of variation

Analyzing Reliability and Validity in Outcomes Assessment (Part 2)

Robert Lingard and Deborah K. van Alphen

California State University, Northridge

Overview – Part 2



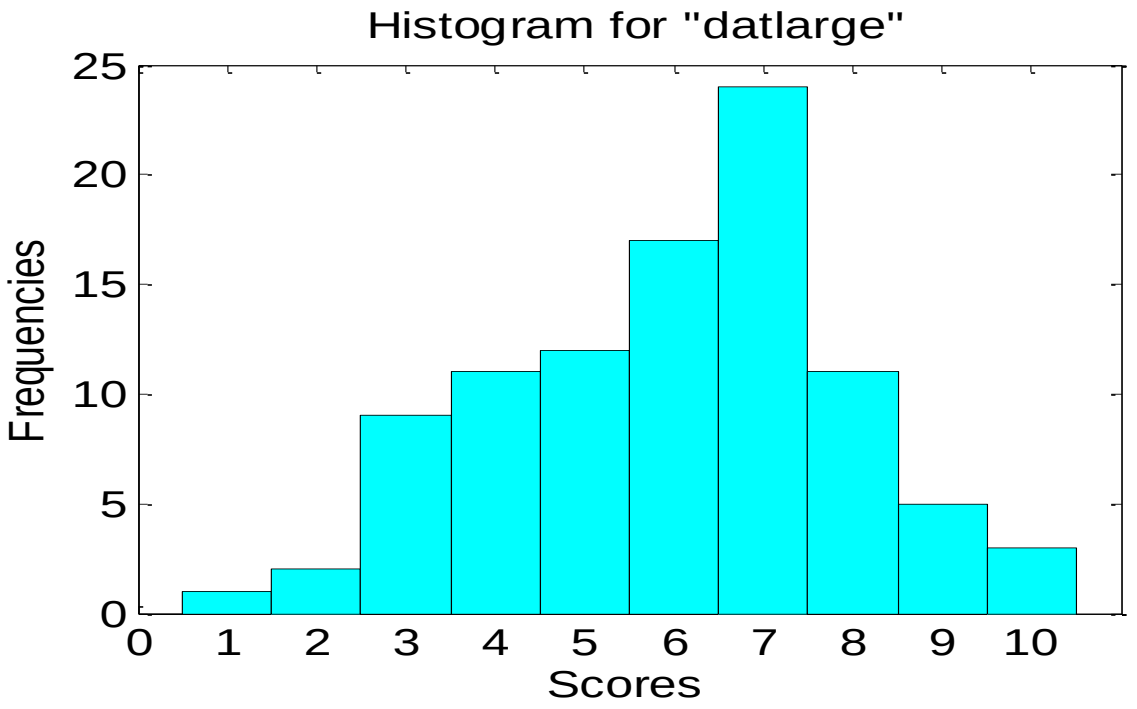
Software Perspective
& Assessment Applications

Software Tools for Statistical Analysis

- MATLAB, with the Statistics Toolbox ✶
  - "MATrix LABoratory," produced by *The Mathworks, Inc*.
  - Complete package includes many math functions, visualization tools, and special-purpose toolboxes
- Excel, with the Add-In: Analysis Toolpak
  - Part of *Microsoft Office Suite*
- Other choices: Minitab, SPSS, Mathematica, …

Using MATLAB to Generate a Histogram

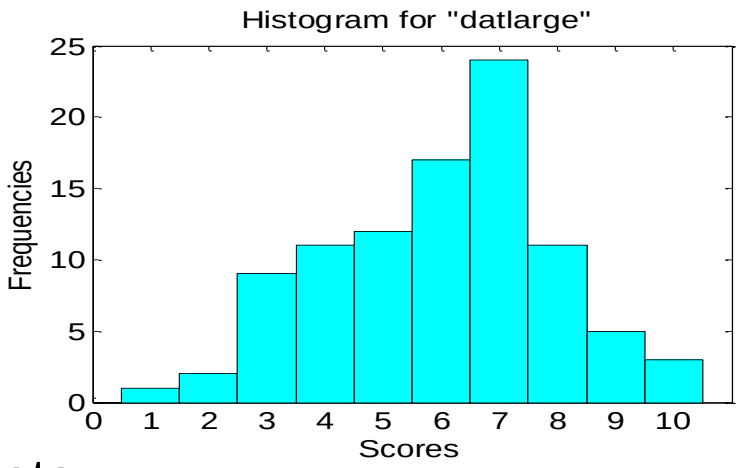
Set of 95 scores, ranging from 1 to 10




```
>> centers = 0 : 10;                    % bin centers at 0, 1, …, 10
>> hist(datlarge, centers)
```

- Constructs a histogram with bin centers specified in the vector *centers*

Using MATLAB to Calculate the Means and Standard Deviations




```
>> mean(datlarge)
ans =
    5.9895
```

```
>> std(datlarge)          % use for sample data
ans =
    1.9433
```

```
>> mle(datlarge)          % use for mean and st. dev. of population data
ans =
    5.9895    1.9331
```

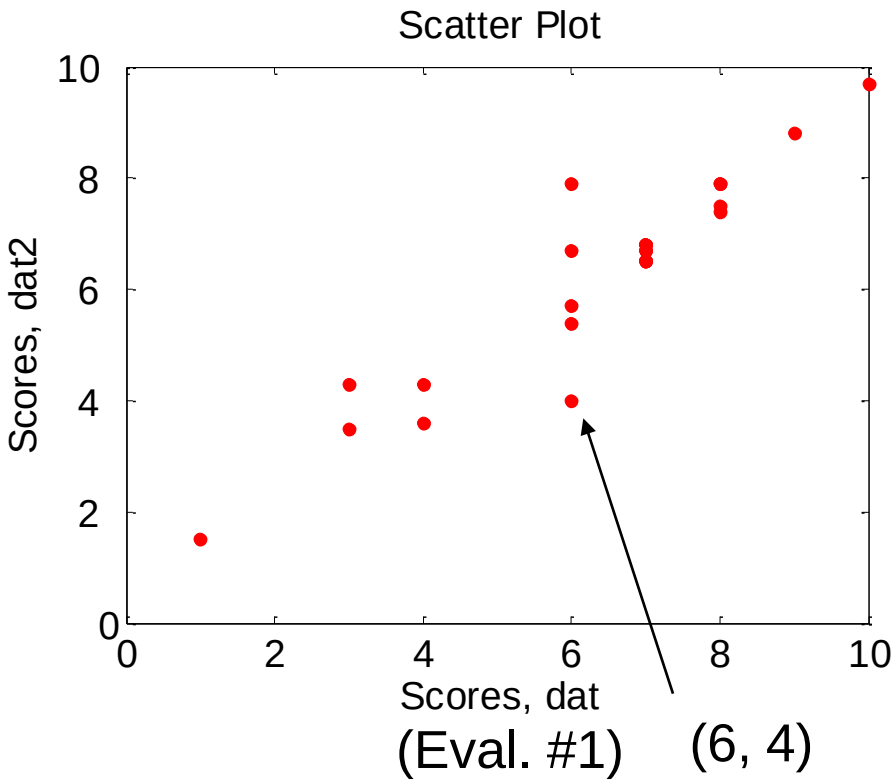
Reliability of Subjective Performance Data

Assessment Situation

Samples of students' writing were scored by two evaluators, using the same rubric, with scores from one evaluator in set "dat", and scores from the other in set "dat2"

Scatter plot: scores given to each sampled student by the two evaluators.

Correlating the scores:




```
>> R = corrcoef(dat,dat2)
R =
    1.0000    0.9331
    0.9331    1.0000
```

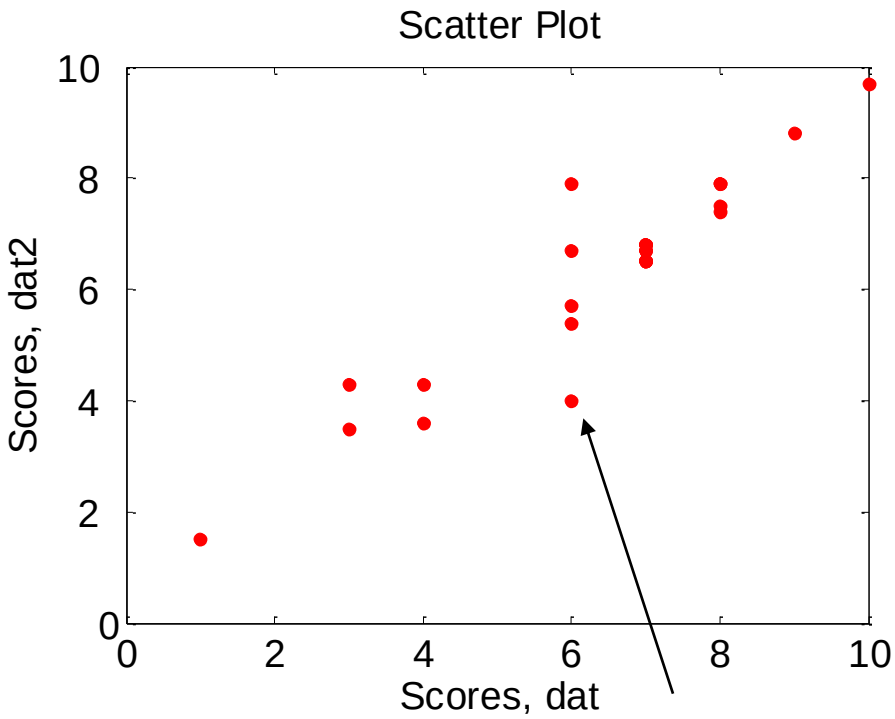
Correlation coefficient of .93 between the two data sets is an indicator of the pair-wise reliability of the data from the two evaluators.

Validity of Test Results

Assessment Situation

A sample of students have scores in set “dat” on one test (**assumed to be valid**), and scores in set “dat2” on another test measuring the same learning outcome.

Scatter plot: scores obtained by each sampled student on the 2 tests

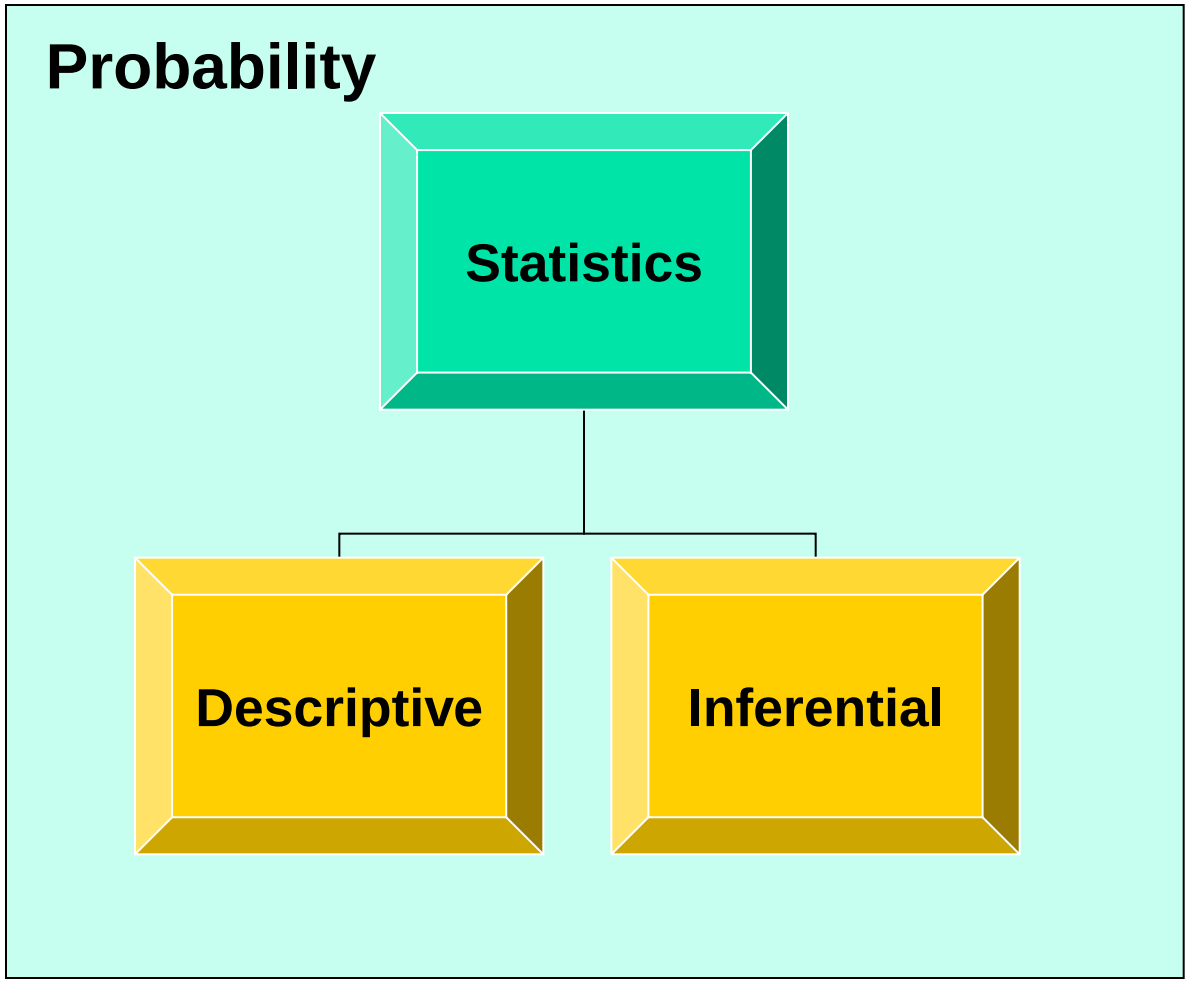


Correlating the scores:

```
>> R = corrcoef(dat,dat2)
R =
    1.0000    0.9331
    0.9331    1.0000
```

Correlation coefficient of .93 between the two data sets is an indicator of the validity the results in *dat2*.

Overview



Random Variables (RV's)

- Intuitive Definition: A random variable is a variable whose value is determined by the outcome of an experiment.
  - Often denoted by a capital bold letter, say **X**
  - Example**:**
    - The scores on an embedded assessment test question

Probability - Notation

- Notation: P(….) denotes the probability of the event described in the parentheses
- Example
  - If **X** is the RV denoting the scores on an embedded assessment test question
    - P(**X** > 8) denotes the probability of a student obtaining a score greater than 8

Probability Density Functions

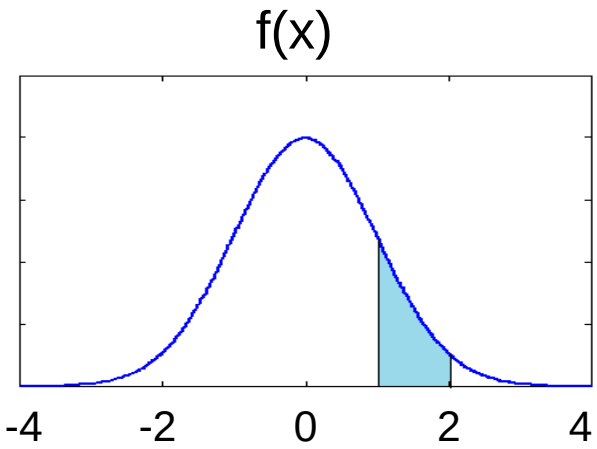
- The **probability density function** (pdf), f(x), for a random variable **X** gives the spread of probability on the real number line. Thus,

$$P(a < \mathbf{X} < b) = \int_a^b f(x)\,dx$$

(area under the pdf between a & b)

Example:

$P(1 < \mathbf{X} < 2)$



- Since the total probability is always 1, the area under the pdf is 1.

Gaussian Random Variables

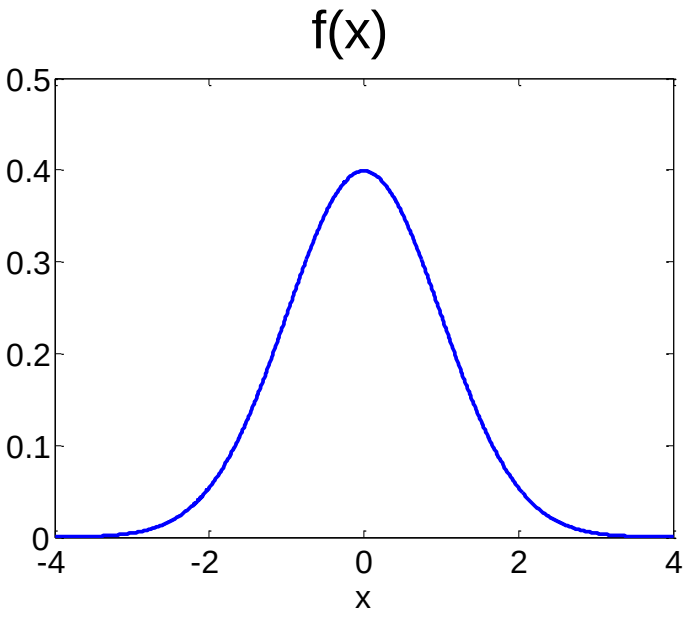
- Definition: Random variable **X** is **Gaussian (or Normal)**, with mean μ and variance σ^2, if its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

"Bell-shaped curve"

Mean: $\mu = 0$;

Standard Deviation: $\sigma = 1$



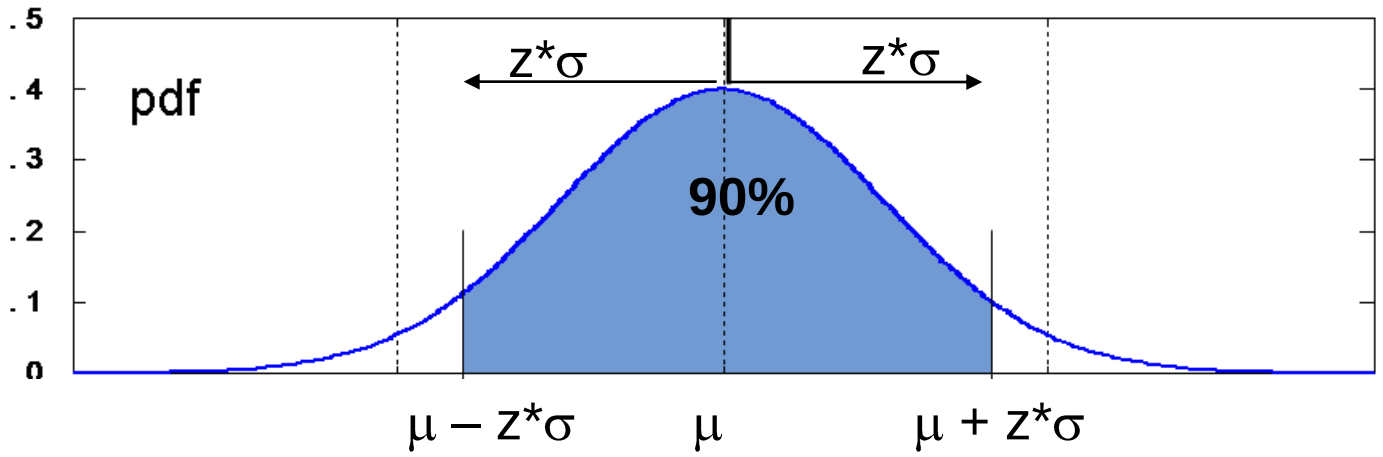
Critical Values for Gaussian RV's (Leading to "Confidence Intervals")

Goal: to form an interval, centered at the mean (μ), containing specific amounts of probability in the pdf of a Gaussian RV.

How many σ's away from the mean do we need to go?

Notation: z* is the number of standard deviations

Example: For a Gaussian RV, find an interval of values, centered at the mean, that contains **90%** of the probability in the pdf:

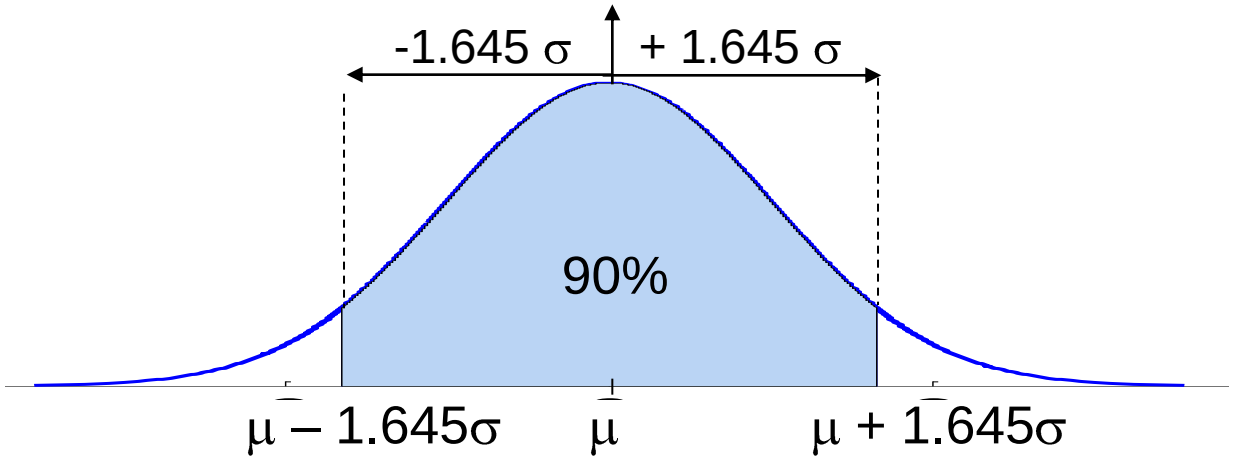


Critical Values for Gaussian RV's

Claim:

The number z* = 1.645 is the **critical value** required to contain 90% of the probability.

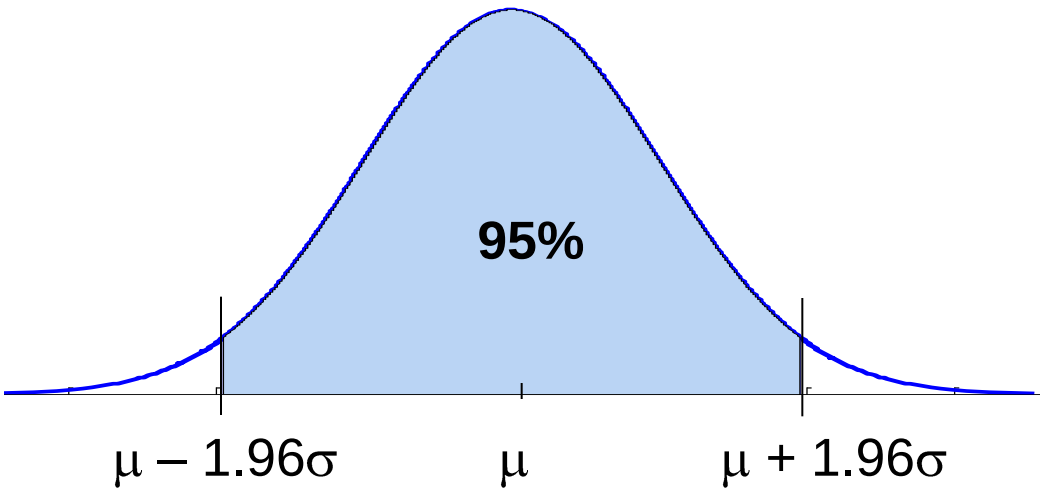
For **all Gaussian RV's**, 90% of the probability is contained within a distance of 1.645 standard deviations from the mean.



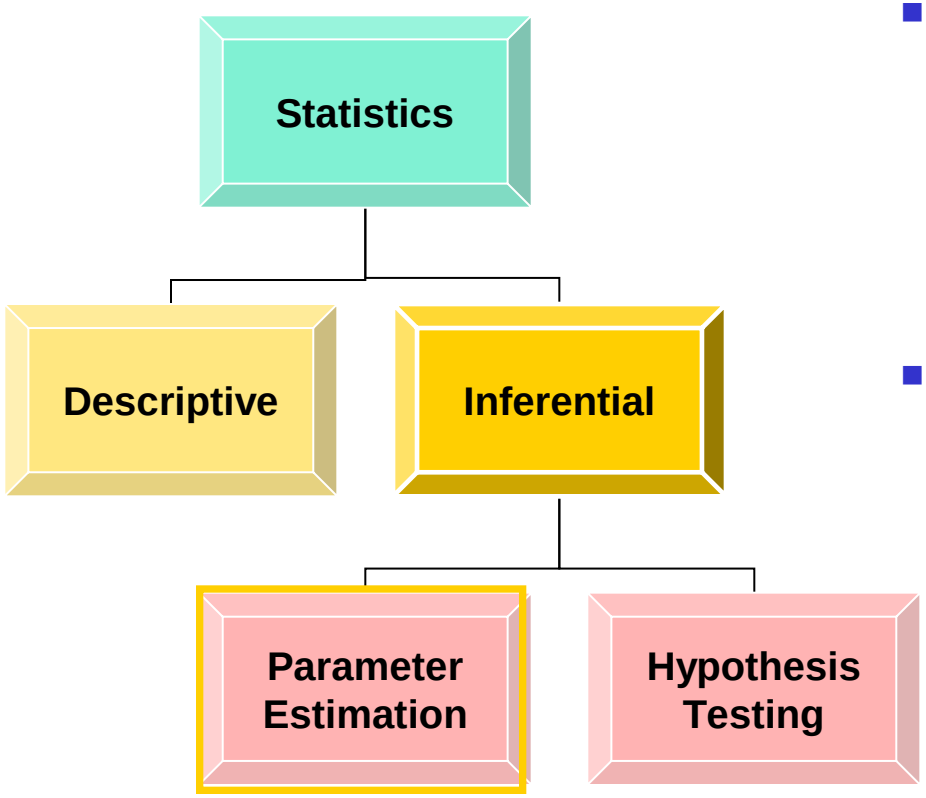
Commonly-used Critical Values

- Some commonly-used z* values, and the corresponding amounts of probability captured are:

z* values	↔	probability
1.280		80%
1.645		90%
1.960		95%
2.575		99%



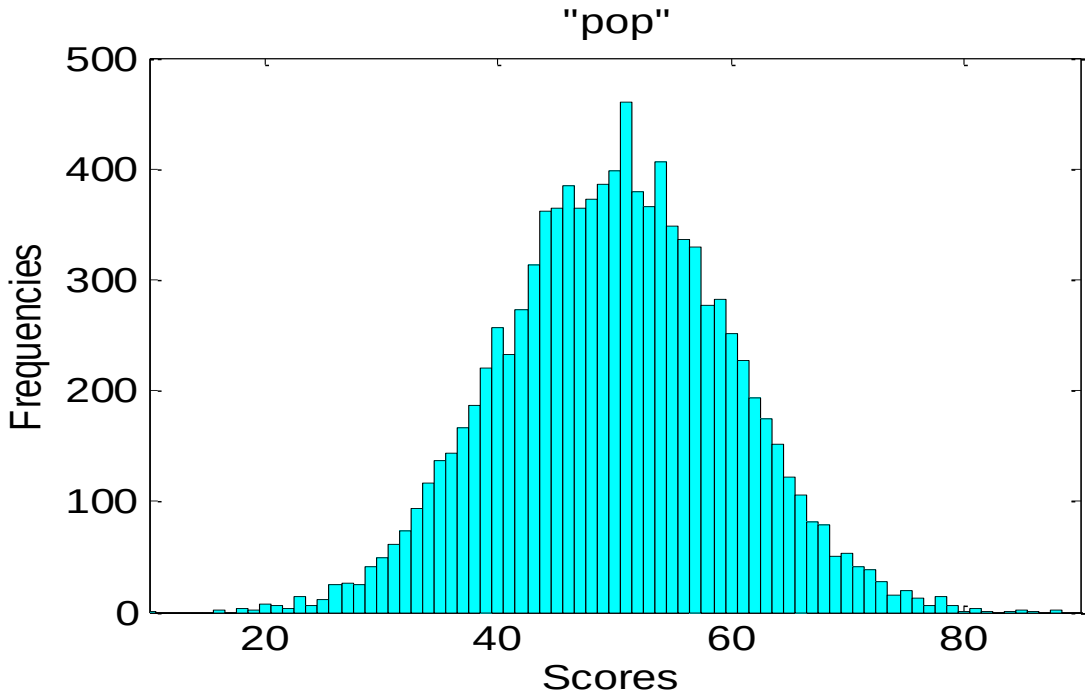
Inferential Statistics



- Given:
  - Random samples of data
  - Sample statistics: **mean**, median, …
- Goal:
  - Determine population parameters: **mean**, median, …
  - **Quantify the confidence** we have in the estimates
  - **Formulate hypotheses** to interpret the population data

Entire Population Distribution

- Consider: a normal population consisting of 10,000 academic scores
- The mean, μ, and standard deviation, σ, of the population **could** be found in MATLAB:



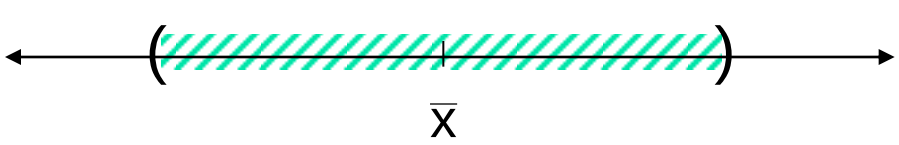
Problem: we don't want to collect/process this much data.

Solution: take a sample, and use the sample statistics to estimate the population parameters.

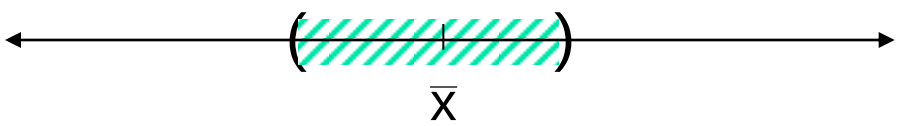
Point Estimates & Confidence Intervals

- Each sample statistic is a **point estimate** of the corresponding population parameter.
- A **confidence intervals**
  - quantify our level of confidence or belief in the point estimates;
  - show an interval of probable values for the population parameter;
  - indicate the precision of the point estimate.

Wide confidence interval



Narrow confidence interval



Confidence Intervals and Levels

- **Confidence level, %**: A measure of the degree of reliability of the confidence interval
  - A confidence level of 95% implies that 95% of all samples (taken from the population) would produce an interval that includes the population parameter being estimated
- Intuitively: the higher the confidence level, the more likely that the population parameter lies within the interval.

Example: Meaning of the 95% Confidence Interval About the Sample Mean

- Suppose that we take **multiple samples** from the population, and obtain the sample mean and standard deviation for each.
- Example: take 100 samples, each of size 100, from the population "pop"

  Sample 1: Calculate sample mean $\bar{x}_1$, sample st. dev. s_1

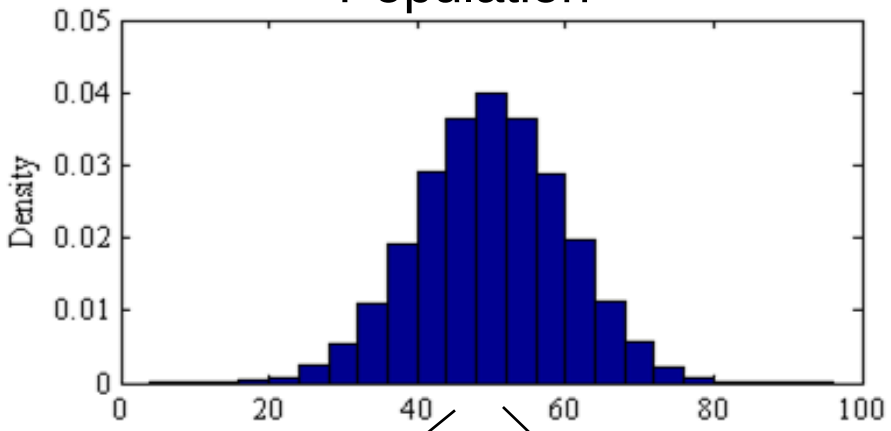
  ...

  Sample 100: Calculate sample mean $\bar{x}_{100}$, sample st. dev. s_{100}

  - Calculate the 95% confidence interval for each of the 100 sample means; and
  - Plot one horizontal line for each confidence interval (and thus for each sample)

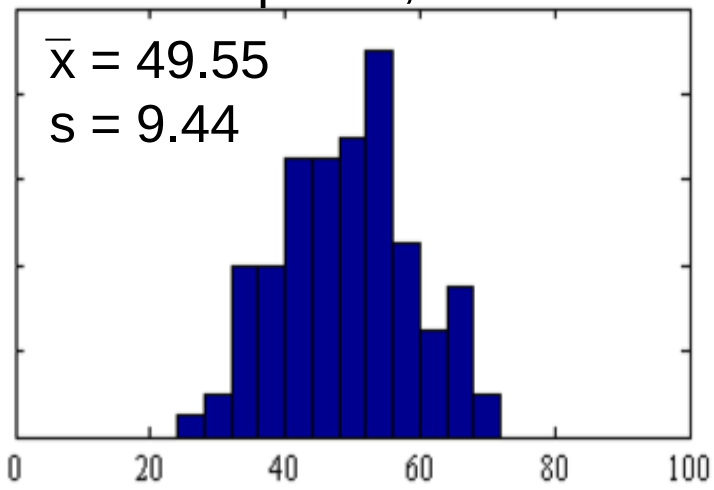
The Meaning of a Confidence Interval

Population



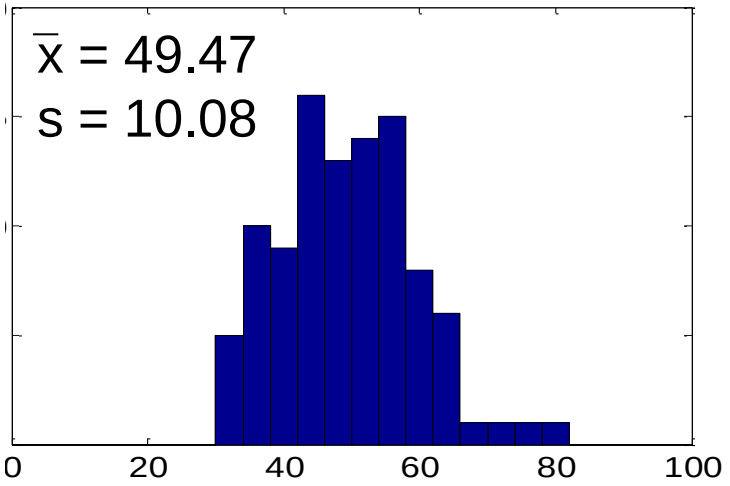
Normal
10,000 Scores
Mean: $\mu = 50.04$
S.D.: $\sigma = 10.01$

Sample #1, n = 100

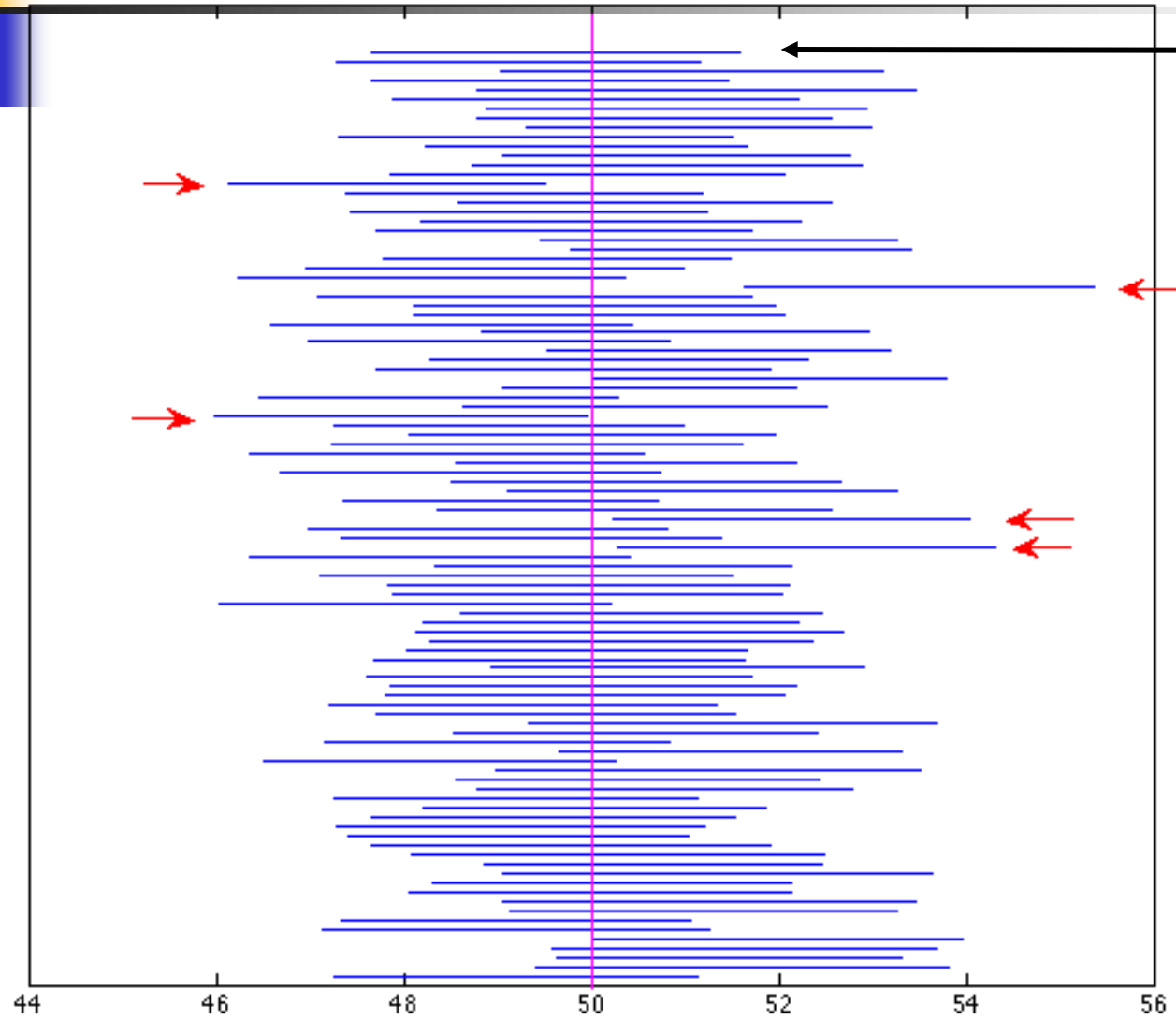


...

Sample #100, n = 100



95% Confidence Intervals (CI) About the Mean for Data: "pop"



1st sample:

$\bar{x}$ = 49.55

CI = (47.7, 51.4)

95 out of 100 confidence intervals include the population mean, μ = 50.04

Finding a Confidence Interval for the Mean

- For populations with:
  - normal distributions; or
  - non-normal distributions with large sample size (n ≥ 30)
- Finding the **95%** Confidence Interval for the Mean

(*)

- For a given sample of size n:
  - Compute $\bar{x}$ and s
  - Determine the 95% confidence interval using the equations for the endpoints in (*) above

Confidence Intervals – Arbitrary Confidence Levels

- To obtain a confidence interval with an arbitrary confidence level.
- Returning to endpoint equations in (*), replacing 1.96 by the more general z*:

Confidence Interval $\left(\bar{x} - \frac{z^* s}{\sqrt{n}}, \ \bar{x} + \frac{z^* s}{\sqrt{n}} \right)$

- Obtain the z* value from:

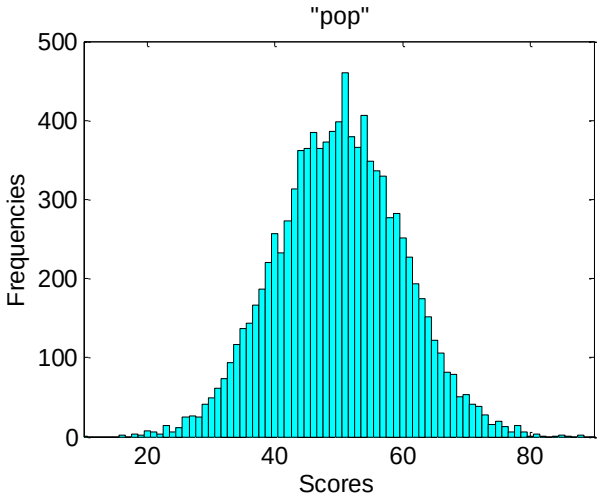
z* values	↔	Confidence Levels
1.280		80%
1.645		90%
1.960		95%
2.575		99%

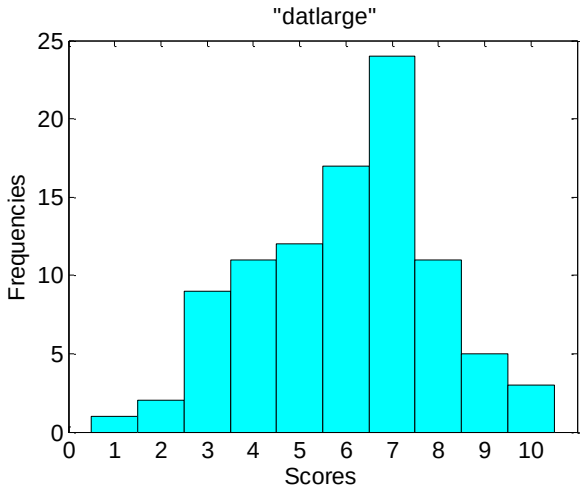
Data Sets for Sample Calculation - Confidence Intervals -

pop

Select sample

datlarge





Set of 10,000 scores, ranging from 10 to 90

Set of 95 scores, ranging from 1 to 10

MATLAB Example: 95% Confidence Intervals for *datlarge*

- The *normfit* MATLAB command finds the sample mean and sample standard deviation, and the 95% confidence interval about the mean.

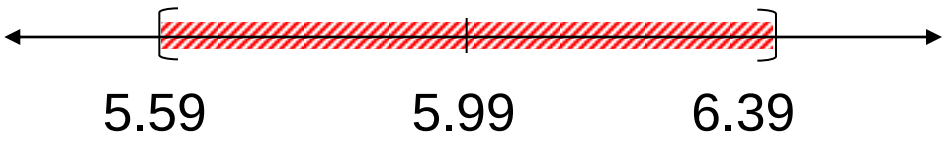
  [xbar, s, xbarci] = normfit(datlarge)

  xbar = sample mean (5.99)

  s = sample standard deviation (1.94)

  xbarci = confidence interval for sample mean: (5.59, 6.39)

95% Confidence Interval for mean:



Confidence Interval: Effect of Sample Size

- Let B be a bound on the "acceptable" estimation error – i.e., on the acceptable difference between the sample mean and the population mean.

- To ensure that: for (Conf. Level)% of the samples, the difference between the sample mean and population mean is no bigger than B:

  - the required sample size is:

$$n = \left[\frac{z^* s}{B} \right]^2$$

(use the z* associated with the desired confidence level)

Example: Selecting Sample Size (Based on data: "**pop**")

- Suppose that we want a 95% confidence that our estimate of the mean population score is within 2 points of the true mean population score
- Calculate n with: s = 10, B = 2, z* = 1.96

$$n = \left[\frac{z^* s}{B}\right]^2 = \left[\frac{1.96 * 10}{2}\right]^2 = 96.04 \Rightarrow n \geq 97$$

- Problem: s, the sample standard deviation is not available until the sample is selected.
- Alternative (Rule of Thumb): Divide the range (largest possible value minus smallest possible value) by 4, for a conservative estimate of s (usually too large)

Selecting Sample Size - Required n Values

90% Confidence Level

Bound, B

	Sample Variance, s							
B	**1**	**2**	**3**	**5**	**10**	**15**	**20**	**30**
1	3	11	25	68	271	609	1083	2436
2	1	3	7	17	68	153	271	609
3	1	2	3	8	31	68	121	271
5	1	1	1	3	11	25	44	98
10	1	1	1	1	3	7	11	25

80% Confidence Level

Bound, B

	Sample Variance, s							
B	**1**	**2**	**3**	**5**	**10**	**15**	**20**	**30**
1	2	7	15	41	164	369	656	1475
2	1	2	4	11	41	93	164	369
3	1	1	2	5	19	41	73	164
5	1	1	1	2	7	15	27	59
10	1	1	1	1	2	4	7	15

Verifying Alleged Improvements (Did We "Close the Loop?")

- In assessment, we may want to determine whether some change that we have implemented has resulted in program improvement.
- Example:
  - Find the difference in the mean scores on an assessment exam/question - one given before, and one after, a curriculum revision
- The statistic used as a point estimate for the difference in means for a population ($\mu_1 - \mu_2$) is the difference in means of samples of the two populations,

$$\bar{x}_1 - \bar{x}_2$$

Sampling Distribution of the Difference of Two Means

- If both population distributions are normal, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is normal.
- If both samples sizes are large, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be approximately normal irrespective of the two population distributions (CLT)
- Equation for the left and right end-points of the confidence interval:

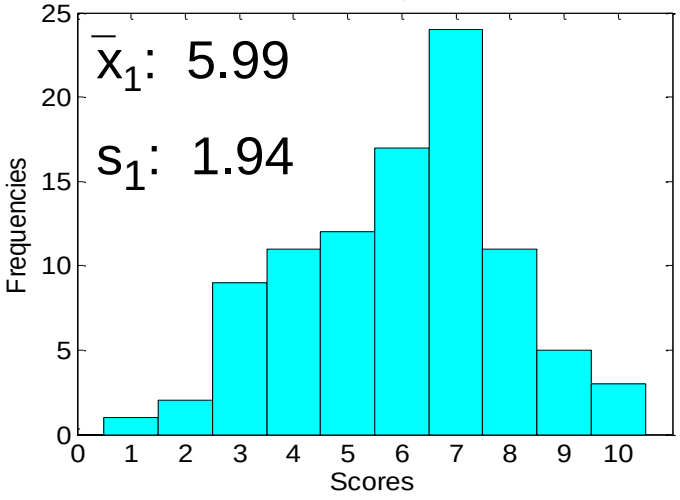
$$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Choose z* based on the desired confidence level.

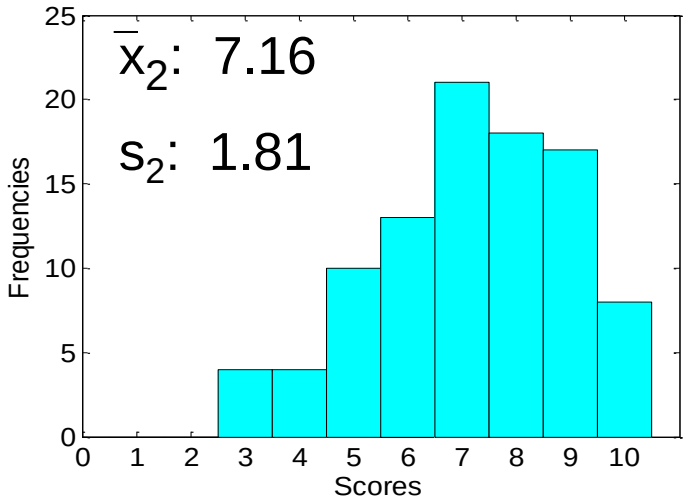
Example: Verifying Alleged Improvements

- Suppose that we want to estimate the difference between two sets of test scores – one given before a curriculum change, and one after.

datlarge: before



datlarge2: after



- Find the point estimate for the difference, and a 99% confidence interval.

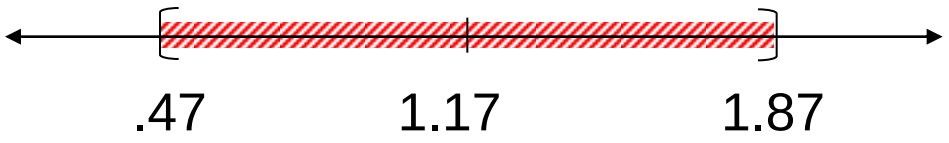
Example: Verifying Alleged Improvement

- Repeating: $\bar{x}_1$: 5.99 $\bar{x}_2$: 7.16; s_1: 1.94 s_2: 1.81
- Point Estimate for $\mu_2 - \mu_1$: $\bar{x}_2 - \bar{x}_1 = 7.16 - 5.99 = 1.17$
- Confidence Interval: $(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$= 1.17 \pm .7009$

99% confident: "after" – "before" = pos. $\Rightarrow$ "after" > "before" $\Rightarrow$ improvement

99% Confidence Interval for difference in means:



About the Assumptions …

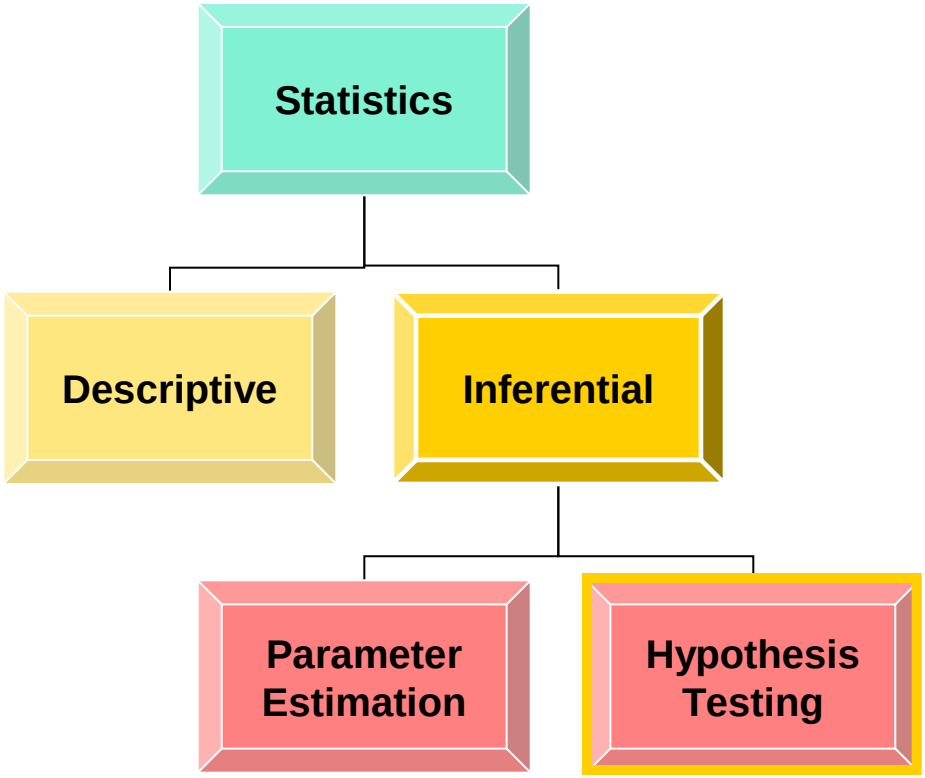
- We have assumed either:
  - normal population, or
  - non-normal population, with large sample size (n >30)

- Different formulas are available

  - For confidence intervals for the mean and the difference of two means if the underlying population is not normal and the sample size is small; and

  - For confidence intervals about different statistics (sample proportion, sample median, etc.)

Overview of Statistics



Definitions

- **Hypothesis**: A statement about the parameters of one or more populations
  - The mean of a set of scores for a certain population is greater than 75, or $\mu > 75$

- **Hypothesis Testing:** a method for deciding between **two competing hypotheses** using information from a random sample
  - $\mu = 75$ or $\mu > 75$

The Null and Alternative Hypotheses

- Null Hypothesis (H_0)
  - Statement that the value of a population parameter is **equal** to some claimed value.
  - Assertion that is initially **assumed to be true** (based on prior belief/measurement)

- Alternative Hypothesis (H_a)
  - Statement that the population parameter has a value that differs from that claimed by the null hypothesis, H_0.

Decision Making About the Hypotheses

- Decision: to reject or accept H_0
- If the data is "consistent" with the H_0, we conclude H_0 is true.
  - Otherwise: we conclude H_0 is false.
  - The data has to be **very inconsistent** with H_0 (highly unlikely) to warrant **rejecting H_0**.
- Criminal Trial Analogy: H_0 - The defendant is innocent.
  - The jury assumes the defendant is innocent until proven guilty.
  - The jury rejects H_0 only if there is compelling evidence against it.

How unlikely should the data be to warrant rejection of H_0?

- Significance level, α: our cut-off point for believing the data occurred just by chance, given H_0.
  - Often 5%, 1 %, or .1%
- p value: the probability that the observed sample data (or something even more extreme) could have occurred, given H_0.

Coin Toss Bet, H_0: The coin is fair.

Alpha: 5%

p = ½	1: heads	$1
p = ¼	2: heads	$2
p = 1/8	3: heads	$3
p = 1/16	4: heads	$4
p = 1/32	5: heads	$5

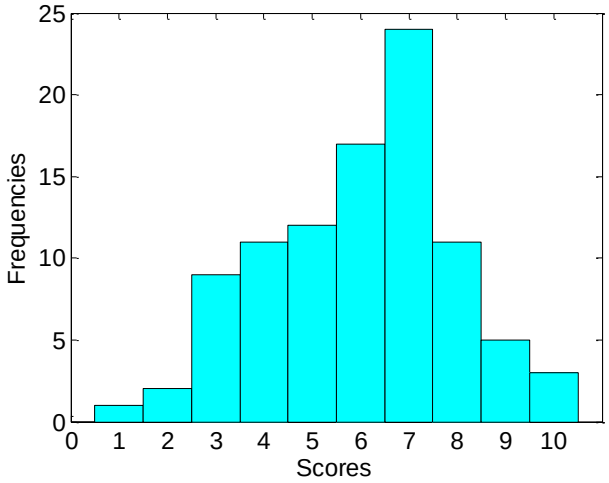
$p = 1/32 \approx 3\%$

$p < \alpha \Rightarrow$ Reject H_0

Hypothesis Testing in Assessment

- Situation: Historically, test scores for a particular learning outcome indicate $\mu = 5$. We "improve" the curriculum, and administer a test on that outcome to a sample of students. We would like to claim that the score has improved, based on our new test results.
- Hypothesis Testing:
  - H_0: $\mu = 5$; H_a: $\mu > 5 \Rightarrow$ (1-tailed test) $\alpha = .05$

"New test data", datlarge



Set of 95 scores, ranging from 1 to 10

Hypothesis Testing with MATLAB: ztest

- **ztest: Performs a hypothesis test on the mean of a normal population with known variance (or unknown variance if n large)***
- [h, p] = ztest(x, m, sigma, alpha, tail)
  - x = sample of data
  - m = mean μ under null hypothesis
  - sigma = variance of population (**use s if sigma unknown**)
  - Tail: 'right' for one- sided test (H_a: $\mu > \mu_o$)
  - **Output h = 1 $\Rightarrow$ reject H_0;** Output h = 0 $\Rightarrow$ accept H_0
  - Output p: the p-value for our data

---

* *ztest* also allows you to output the confidence interval about the population mean

Hypothesis Testing with MATLAB: ztest

- Returning to our example - Hypotheses: H_0: $\mu = 5$; H_a: $\mu > 5$
  - $\alpha = .05$
  - Tail = 'right', since H_a: u > 5

- MATLAB

```
>> std(datlarge)
ans =
   1.9433
>> [h   p] = ztest(datlarge,  5,   1.9433,  .05,  'right')
h =
      1
p =
      3.4751e-007
```

"Significant improvement"

($\Rightarrow$ reject H_0; performance did improve)

(prob. of sample data occurring if H_0 true)

Statistically Significant Improvements

- A result is said to be statistically significant if it is not likely to have occurred by chance.
- Since the p-value of a test is the probability that the data occurred by chance (given H_0), the smaller the p-value, the greater the statistical significance.

small p-value $\Rightarrow$ "statistically significance"

- Statistically significant at the 1% Level: The p-value of the data is less than .01.

$\Leftrightarrow$ Reject H_0 if $\alpha = .01$

Summary of Assessment Applications for Statistics – What Have We Shown?

- Use of descriptive statistics and data visualization with histograms
- Verification of **reliability** with
  - High correlation between 2 sets of scores
  - Lack of significant difference between 2 sets of scores
- Verification of **validity** with
  - High correlation between scores obtained from a new instrument and those from an established standard;
  - Lack of significant difference between scores obtained from a new instrument and those from an established standard

Summary of Assessment Applications for Statistics

- **Simplifying** Program Assessment with **Sampling**
  - Choosing an appropriate sample size for estimating the mean
  - Finding confidence levels for the (point) estimate
- **Claiming Improvement** – estimates of the difference in means
  - Finding confidence levels for the (point) estimate
  - Determining whether the difference is **significant**
- Applications discussed in the context of test scores also occur in the context of surveys

Questions??? & Contact Information

- Email: rlingard@csun.edu (Bob Lingard)

  dvanalphen@csun.edu (Debbie van Alphen)

"Not everything that counts can be counted, and not everything that can be counted counts."

(Sign hanging in Einstein's office at Princeton)

"98% of all statistics are made up." ~Author Unknown