Lecture 6: Finish Ch.4 (normal); Chapter 5, Section 1 Relationships (Categorical and Quantitative)

- z-scores
- Two- or Several-Sample or Paired Design
- Displays and Summaries
- Notation
- Role of Spreads and Sample Sizes

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-4)
  - Displaying and Summarizing
    - Single variables: 1 cat. (Lecture 5), 1 quantitative
    - Relationships between 2 variables
  - Probability
  - Statistical Inference

Definition

- ***z*-score**, or **standardized value**, tells how many standard deviations below or above the mean the original value ***x*** is:

$$z = \frac{\text{value}-\text{mean}}{\text{standard deviation}}$$

- Notation:
  - Sample: $z = \frac{x-\bar{x}}{s}$
  - Population: $z = \frac{x-\mu}{\sigma}$
- **Unstandardizing** *z*-scores:

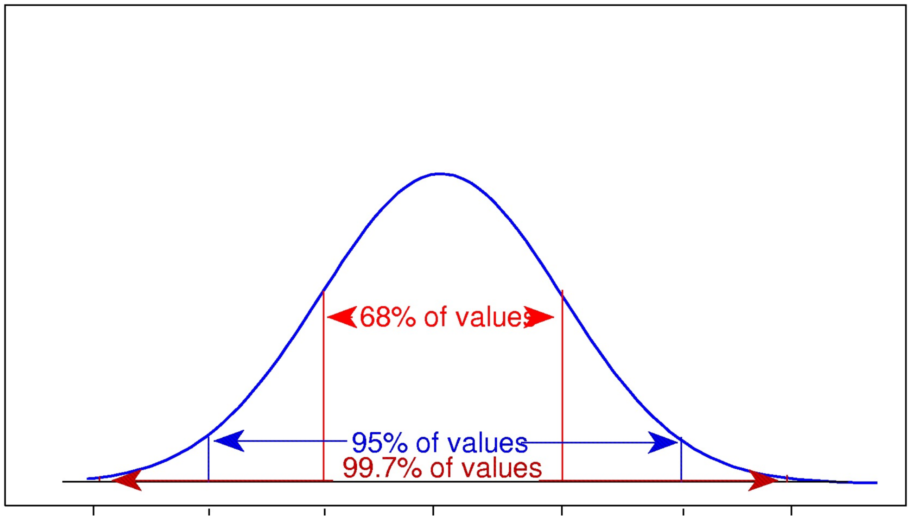
Original value *x* can be computed from *z*-score.

Take the mean and add *z* standard deviations:

$$x = \mu + z\sigma$$

Example: *68-95-99.7 Rule for z*

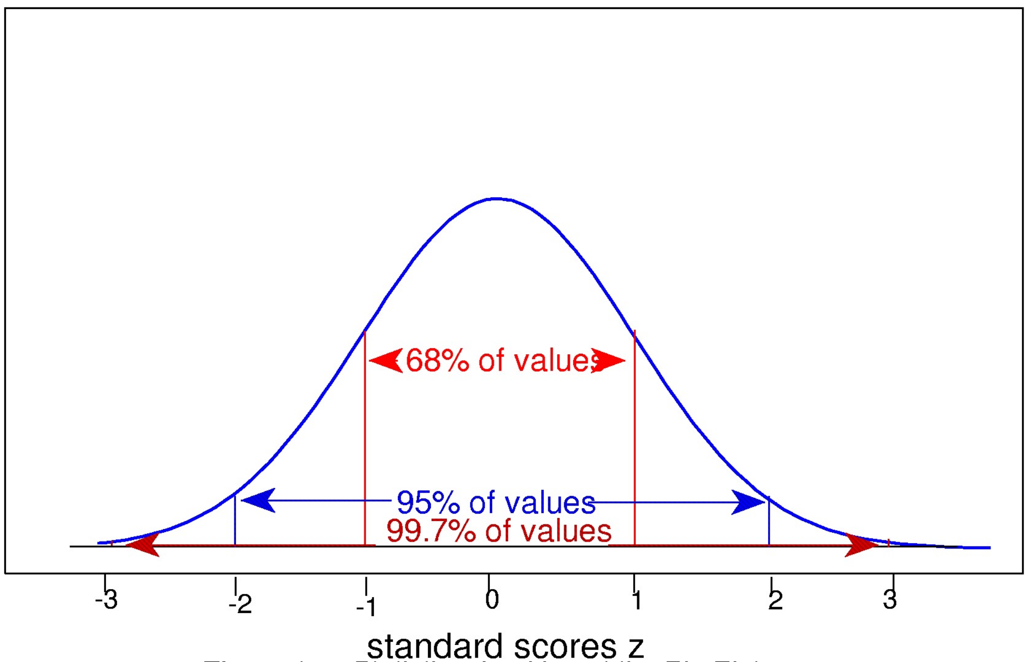
- **Background:** The 68-95-99.7 Rule applies to any normal distribution.
- **Question:** What does the Rule tell us about the distribution of standardized normal scores *z*?
- **Response:** Sketch a curve with mean__, standard deviation __:



68-95-99.7 Rule for z-scores

For distribution of standardized normal values z,

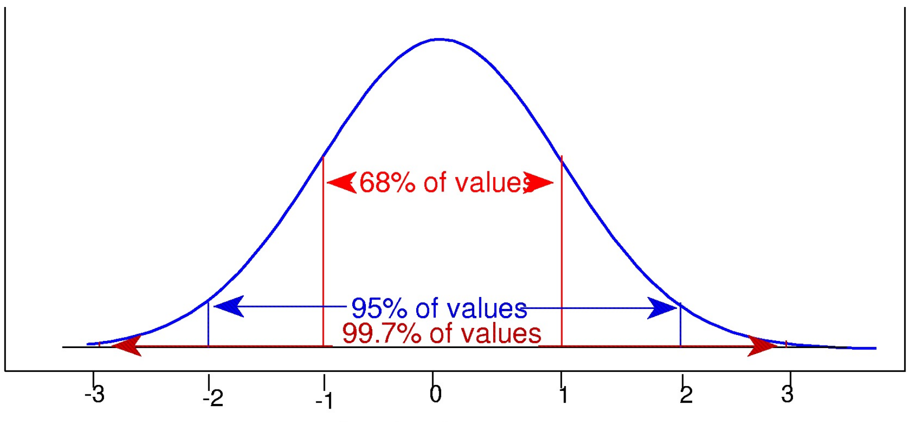
- 68% are between -1 and +1
- 95% are between -2 and +2
- 99.7% are between -3 and +3



Example: *What z-scores Tell Us*

- **Background:** On an exam (normal), two students' *z*-scores are -0.4 and +1.5.
- **Question:** How should they interpret these?
- **Response:**
  - -0.4: ________________________________
  - +1.5: ________________________________

    ____



Interpreting z-scores

This table classifies ranges of z-scores informally, in terms of being unusual or not.

Size of z	Unusual?
$\lvert z \rvert$ greater than 3	extremely unusual
$\lvert z \rvert$ between 2 and 3	very unusual
$\lvert z \rvert$ between 1.75 and 2	unusual
$\lvert z \rvert$ between 1.5 and 1.75	maybe unusual (depends on circumstances)
$\lvert z \rvert$ between 1 and 1.5	somewhat low/high, but not unusual
$\lvert z \rvert$ less than 1	quite common

Example: *Calculating and Interpreting z*

- **Background:** Molar lengths (mm) for 3 species are normal:
  - *Paranthropus Boisei* mean 11.5, standard deviation 0.8
  - *Early homo erectus*: mean 7.8, standard deviation 0.7
  - *Early homo:* mean 9.3, standard deviation 0.6
- **Question:** Anthropologists discovered a 9.7 mm molar. What's our best guess of which species it came from?
- **Response:**
  - *Paranthropus Boisei:*
  - *Early h. erectus:*
  - *Early homo:*

Practice: 4.48b p.119

Example: *z Score in Life-or-Death Decision*

- **Background:** IQs are normal; mean=100, sd=15. In 2002, Supreme Court ruled that execution of mentally retarded is cruel and unusual punishment, violating Constitution's 8th Amendment.
- **Questions:** A convicted criminal's IQ is 59. Is he borderline or well below the cut-off for mental retardation? Is the death penalty appropriate?
- **Response:** His *z*-score is ____________________

____

Practice: 4.55f p.121

Example: *From z-score to Original Value*

- **Background:** IQ's have mean 100, sd. 15.
- **Question:** What is a student's IQ, if z=+1.2?
- **Response:**

Practice: 4.48d p.119

Definition *(Review)*

- ***z*-score**, or **standardized value**, tells how many standard deviations below or above the mean the original value ***x*** is:

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

- Notation:
  - Sample: $z = \frac{x - \bar{x}}{s}$
  - Population: $z = \frac{x - \mu}{\sigma}$

- **Unstandardizing** *z*-scores:

Original value *x* can be computed from *z*-score.

Take the mean and add *z* standard deviations:

$$x = \mu + z\sigma$$

Example: *Negative z-score*

- **Background:** Exams have mean 79, standard deviation 5. A student's z score on the exam is -0.4.
- **Question:** What is the student's score?
- **Response:**

If z is negative, then the value x is below average.

Practice: 4.55h p.121

Example: *Unstandardizing a z-score*

- **Background:** Adult heights are normal:
  - Females: mean 65, standard deviation 2.8 (or 3)
  - Males: mean 70, standard deviation 3
- **Question:** Have a student report his or her z-score; what is his/her actual height value?
- **Response:**
  - Females: take $65+z(3)=$___
  - Males: take $70+z(3)=$___

Example: *When Rule Does Not Apply*

- **Background**: Students' computer times had mean 97.9 and standard deviation 109.7.
- **Question:** How do we know the distribution of times is not normal?
- **Response:**

Practice: 4.61a-b p.122

Looking Back: *Review*

- 4 Stages of Statistics
  - Data Production (discussed in Lectures 1-3)
  - Displaying and Summarizing
    - Single variables: 1 cat,1 quan (discussed Lectures 4-6)
    - Relationships between 2 variables:
      - Categorical and quantitative
      - Two categorical
      - Two quantitative
  - Probability
  - Statistical Inference

Single Quantitative Variables *(Review)*

- **Display:**
  - Stemplot
  - Histogram
  - Boxplot
- **Summarize**:
  - Five Number Summary
  - Mean and Standard Deviation

Add categorical explanatory variable →

display and summary of quantitative responses are **extensions** of those used for single quantitative variables.

Design for Categorical/Quantitative Relationship

- Two-Sample
- Several-Sample
- Paired

Looking Ahead: *Inference procedures for population relationship will differ, depending on which of the three designs was used.*

Displays and Summaries for Two-Sample Design

- **Display: Side-by-side boxplots**
  - One boxplot for each categorical group
  - Both share same quantitative scale
- **Summarize:** Compare
  - Five Number Summaries (looking at boxplots)
  - Means and Standard Deviations

Looking Ahead: *Inference for population relationship will focus on means and standard deviations.*

Example: *Formats for Two-Sample Data*

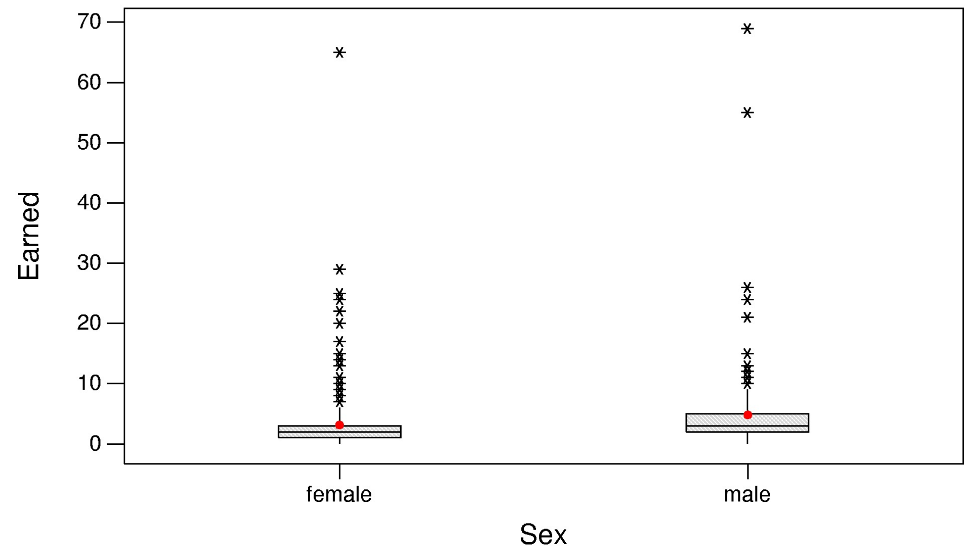
- **Background**: Data on students' earnings includes gender info:

MaleEarnings	FemaleEarnings
12	3
1	7
10	2
...	...

- **Question:** How else can we format the data?
- **Response:**

Example: *Display/Summarize for Two-Sample*

- **Background**: Earnings of sampled males and females are displayed with side-by-side boxplots.



- **Question:** What do the boxplots show?
- **Response:**
  - Center:
  - Spread:
  - Shape:

Example: *Summaries for Two-Sample Design*

- **Background**: Earnings of sampled males and females are summarized with software:

Descriptive Statistics: Earned by Sex

Variable	Sex	N	Mean	Median	TrMean	StDev
Earned	female	282	3.145	2.000	2.260	5.646
	male	164	4.860	3.000	3.797	7.657

Variable	Sex	SE Mean	Minimum	Maximum	Q1	Q3
Earned	female	0.336	0.000	65.000	1.000	3.000
	male	0.598	0.000	69.000	2.000	5.000

- **Question:** What does the output tell us?
- **Response:**
  - Centers:
  - Spreads:
  - Shapes:

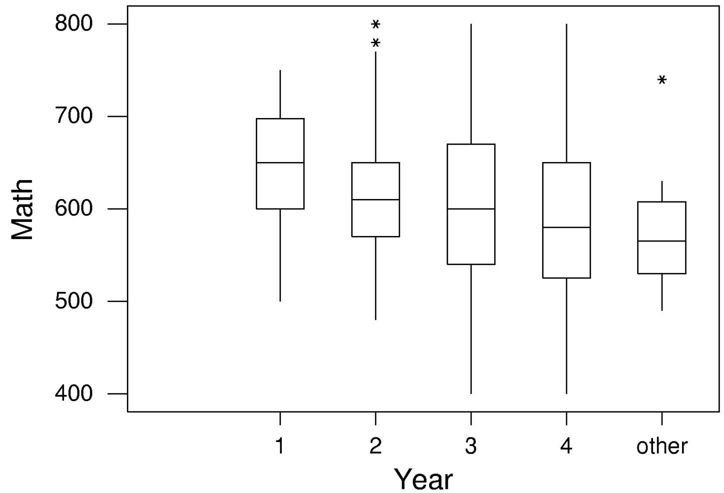
Practice: 5.83 p.210

Design for Categorical/Quantitative Relationship

- Two-Sample
- Several-Sample
- Paired

Example: *Several-Sample Design*

- **Background**: Math SAT scores compared for samples of students in 5 year categories.



- **Question:** What do the boxplots show?
- **Response:**

Looking Back: *(Sampling Design) Are there confounding variables/bias? These are all intro stats students…*

Practice: 5.10 p.146

Design for Categorical/Quantitative Relationship

- Two-Sample
- Several-Sample
- Paired

Display and Summaries for Paired Design

- **Display:** histogram of differences
- **Summarize:** mean and standard deviation of differences

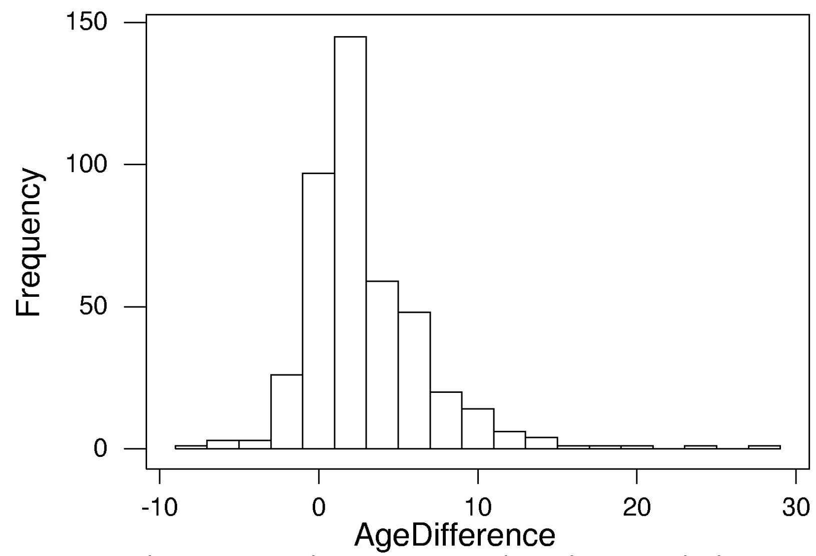
Example: *Paired vs. Two-Sample Design*

- **Background:** Comparing ages of surveyed students' parents to see if mothers or fathers are older.
- **Questions:**
  - Why is design paired, not two-sample?
  - How to display and summarize relationship between parent sex and parent age?
  - What results would you expect to see?
- **Responses:**
  - Paired because ________________________________
  - Display: ____________________________________
    Summarize: _________________________________
  - May suspect __________ tend to be older.

Practice: 5.13a p.147

Example: *Histogram of Differences*

- **Background:** Histogram of differences, father's age minus mother's age:



- **Question:** What does histogram show about relationship between parent sex and parent age?
- **Response:**
  - Center:
  - Spread:
  - Shape:

Practice: 5.13b-d p.147

Notation

- **Two-sample or Several-Sample Design:** extend notation for means and standard deviations with subscript numbers 1, 2, etc.
- **Paired Design:** indicate notation for differences with subscript "*d*"

Example: *Notation*

- **Background**: For a sample of countries, illiteracy rates are recorded for each gender group.
- **Question:** How do we denote the following?
  - Mean of illiteracy differences for sampled countries
  - Standard deviation of illiteracy differences for the sampled countries
- **Response:** (________ design)
  - Mean of illiteracy differences for the sampled countries:
  - Standard deviation of illiteracy differences for the sampled countries:

Example: *More Notation*

- **Background**: Records are kept concerning percentages of students at all private, state, and state-related schools receiving Pell grants.
- **Question:** How do we denote the following?
  - Mean percentages for the three types of school
  - Standard deviations of percentages for the three types of school
- **Response:**
  - Mean %'s for the three types of school:
  - Standard deviations of %'s for the three types of school:

Practice: 5.71b p.206

Sample vs. Population Differences

How different are responses for sampled groups?

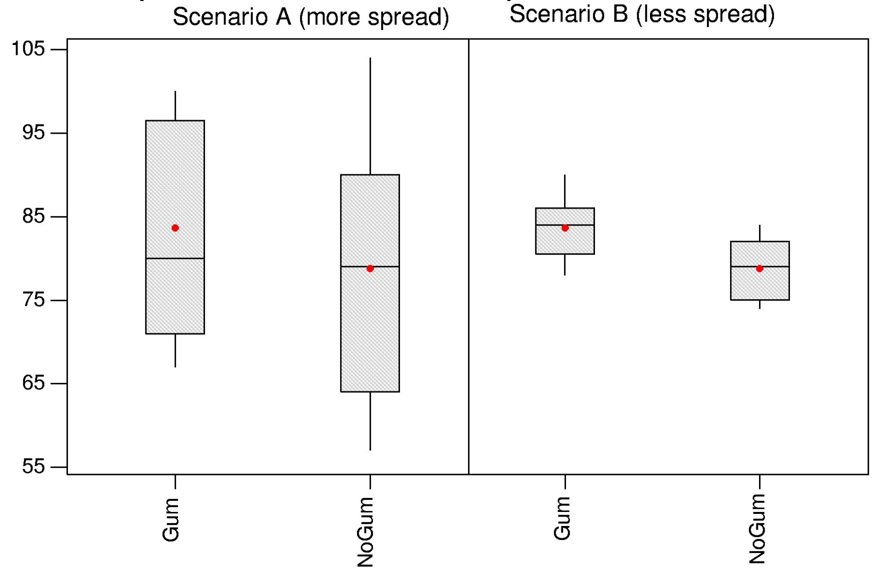
- **Centers:** First compare means/medians.
- **Spreads:** Differences appear more pronounced if values are concentrated around their centers.
- **Sample Sizes:** Differences are more impressive coming from larger samples.

Looking Ahead: *Inference comparing means will have us focus on centers, spreads, and sample sizes.*

Example: *Impact of Spreads on Perceived Difference between Means*

- **Background:** Experiment compared test scores for gum-chewers and non-chewers learning anatomy. Means: 83.6 (chewers), 78.8 (non-chewers)

One of these (left or right) represents the actual data.



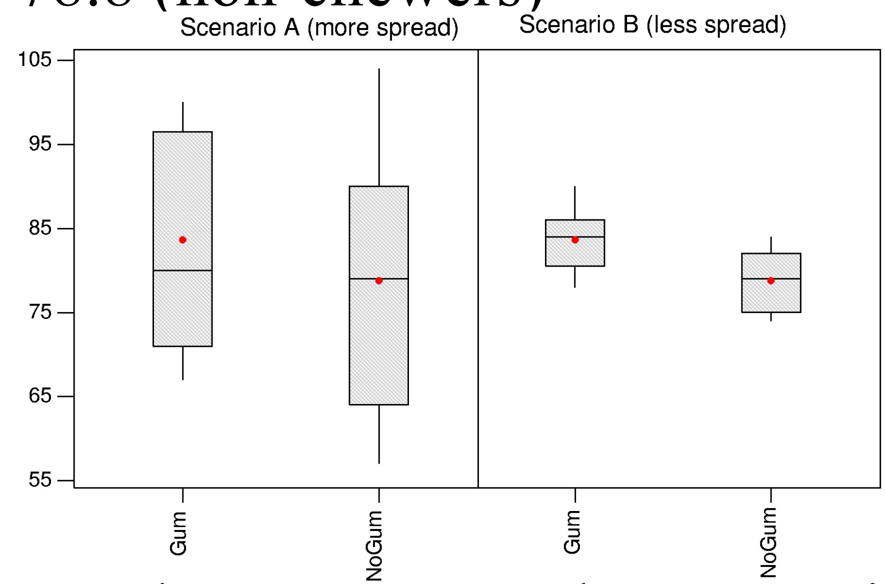
- **Question:** For which scenario (left or right) are you more convinced that chewing gum aids learning?
- **Response:**

Example: *Impact of Sample Size on Perceived Difference between Means*

- **Background:** Experiment compared test scores for gum-chewers and non-chewers learning anatomy. Means: 83.6 (chewers), 78.8 (non-chewers)
- **Question:** Which would convince you more that chewing gum aids learning: if data came from 56 students or 560 students?
- **Response:**

Example: *Impact of Study Design on Perceived Difference between Means*

- **Background:** Experiment compared test scores for gum-chewers and non-chewers learning anatomy. Means: 83.6 (chewers), 78.8 (non-chewers)



- **Question:** Are there concerns about experimenter effect, placebo effect, realism, ethics, compliance?
- **Response:**______________________is most worrisome.

Lecture Summary *(Normal Distributions)*

- **Standardizing:** z=(value-mean)/sd
- **68-95-99.7 Rule:** applied to standard scores z
- **Interpreting Standard Score z**
- **Unstandardizing:** x=mean+z(sd)

Lecture Summary
(Categorical and Quantitative Relationships)

- **Two- or Several-Sample Design**
  - **Format:** one column for each group or
    one column for each of two variables
  - **Display:** side-by-side boxplots
  - **Compare:** means and sd's or 5 No. Summaries
- **Paired Design:**
  - **Display:** Histogram of differences
  - **Summarize:** Mean and sd of differences
- **Notation:** Design? Sample or population?
- **How Different Are Sample Means?**
  - Impacted by spreads and sample sizes