Lecture 9: Chapter 5, Section 1
Relationships
(Categorical and Quantitative)

- 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,1 quan (discussed Lectures 5-8)
    - 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*

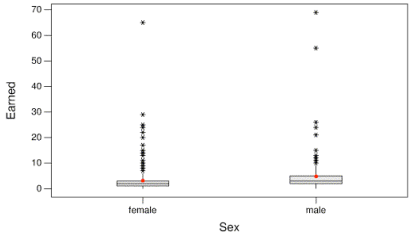
- **Background**: Data on students' earnings includes gender info:

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:

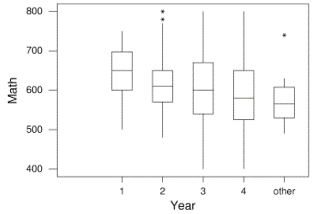
```
Descriptive Statistics: Earned by Sex
Variable   Sex              N       Mean     Median     TrMean      StDev
Earned     female         282      3.145      2.000      2.260      5.646
           male           164      4.860      3.000      3.797      7.657

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:

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…*

Display and Summaries for Paired Design

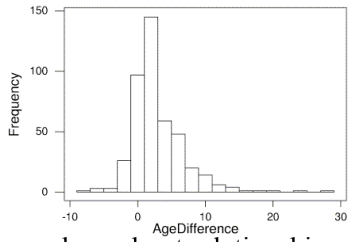
- **Display:** histogram of differences
- **Summarize:** mean and standard deviation of differences

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.

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:

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: *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:

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:

Sample vs. Population Differences

How different are responses for sampled groups?

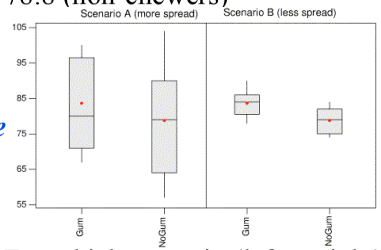
- **Centers:** First compare means/medians.
- **Spreads:** Differences appear more pronounced if values are concentrated around their centers.
- **Sample Sizes:** Differences are more impressive coming from larger samples.

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.



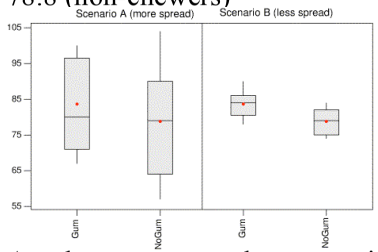
- **Question:** For which scenario (left or right) are you more convinced that chewing gum aids learning?
- **Response:**

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 Spreads/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:** Are there concerns about experimenter effect, placebo effect, realism, ethics, compliance?
- **Response**:_____is most worrisome.

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