# Notes 1: Descriptive Statistics
## Supplemental Presentation Notes

**Population vs. Sample Illustrated**

Population – Example: Graduate Class in Statistics with 10 students
Sample – Subset of population (less than entire population selected)
Census – All selected from Population

| | Population = Entire Class | Age | | Sample = Subset of Population (e.g., n = 4) | Age |
|---|---|---|---|---|---|
| Student | Name | | | | |
| 1 | Bryan | 47 | | Bryan | 47 |
| 2 | Marijke | 15 | | | |
| 3 | Gunther | 13 | | Gunther | 13 |
| 4 | Marlynn | 48 | | | |
| 5 | Bob | 73 | | | |
| 6 | James | 73 | | | |
| 7 | Diana | 67 | | Diana | 67 |
| 8 | Linda | 70 | | | |
| 9 | Gary | 40 | | Gary | 40 |
| 10 | Eric | 45 | | | |
| | | | | | |
| | Mean (Parameter) $\mu$ = | 49.10 | | Mean (Statistic) $\bar{X}$ or M = | 41.75 |

**Variables**

Identify variables in the following statements. Note difference between a variable and categories of a variable.

1. There is a difference in weight between males and females.
2. The age of a student is associated with reading readiness scores.
3. All students in the class are female.

Answers

1. Weight; Sex (male female)
2. Age; Reading readiness score
3. No variables

**Measurement**

Process of assigning labels to categories of a variable. (Do not assume numbers are involved; this process is not necessarily quantitative in nature.)

How would one measure student sex in this course?
How would one measure student age in this course?

**Scales of Measurement**

Nominal: categories only
> Student Sex
> Race
> Types of Flowers

Ordinal: categories with inherent rank
> I think this course is well organized.

| Strongly Disagree | Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|

Interval: categories with inherent rank and precise, equal intervals
> Time in infinity
> Distance in infinity
> No education related variables (that I can conceptualize)

Ratio: categories with inherent rank and precise, equal intervals; and also with a true starting or ending point (true zero)
> Weight in pounds
> Length of tabletop
> Time to complete lap or run 5 kilometers (can form ratios)
> > Runner A: 15:00
> > Runner B: 18:12
> > Runner C: 21:43
> > Runner D: 30:00
> > Runner E: 33:51
> > Runner F: 45:00
> > Runner G: 51:11

What scale is this variable?

> I think this course is well organized. (Circle number that best reflects your view.)

| Strongly Disagree | Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Qualitative vs. Quantitative Variable
> Qualitative = nominal (categorical variable)
> Quantitative = ratio, interval, and some ordinal variables

Examples:
1. If everyone buys a bag of apples, and we each count the number of whole apples in each of our bags, what is the scale of measurement for these counts? Is this qualitative or quantitative? (Work through logic of this example.)
2. Time required for individuals to complete a given task, such as washing dishes.
3. The classification of people into one of three student groups in high school (nerds, athletes, and others).
4. What about this, we take the classification used above (nerds, athletes, others) and provide a different label now with numbers, like this: group 1, group 2, and group 3?

Answers
1. Count of apples in a bag: ratio – quantitative
2. Ratio – quantitative
3. Categories without rank, therefore nominal
4. Nominal

**Independent (IV) and Dependent Variables (DV)**

How does one identify the IV? Is manipulation required?

Which are IV and DV for the following?

1. There is a difference in weight between males and females.
2. Students exposed to accelerated reader (AR) will have lower intrinsic motivation to read scores than students not exposed to AR.
3. The greater one's anxiety before a test, the lower will be one's test scores.
4. There is an association between test anxiety and academic self-efficacy.

Answers
1. IV = Sex (male/female), DV = Weight
2. IV = exposure to AR (exposed, not exposed), DV = motivation to read score
3. IV = test anxiety, DV = test score
4. Not enough information to determine whether test anxiety or academic self-efficacy is IV or DV

**Writing Hypotheses**

Qualitative IVs

Focus on group differences, not on variable relationships; identify all groups/categories of the IV involved and make clear which groups will have more or less of the DV for directional hypotheses.

Example
IV: Sex (males and females)
DV: Weight

Directional
    Generic: Group A will have more of the DV than Group B.
    Specific to Sex and Weight: Males will weigh more than females.

Non-directional
    Generic: There will be a ***difference*** in DV between Group A and Group B.
    Specific to Sex and Weight: There will be a difference in weight between males and females.

Null
    Generic: There will be ***no difference*** in DV between Group A and Group B.
    Specific to Sex and Weight: There will be no difference in weight between males and females.

Quantitative IVs

Focus on relationship between IV, not on group differences.

Example
IV: Weight of Car
DV: Miles per Gallon (MPG)

Review positive (time studying and test scores) and negative (car weight and MPG) relationships.

↑↑    ↑↓    ↓↓

Directional
> Generic: There is a positive (or negative) association between IV and DV
> > or
> Generic: The higher the IV, the higher (or lower) will be the DV.
> Specific to Weight and MPG: There is a negative association between car weight and MPG.
> > or
> Specific to Weight and MPG: The heavier the car, the lower will be MPG.

Non-directional
> Generic: There will be an association between the IV and DV.
> Specific to Weight and MPG: There will be an association between car weight and MPG.

Null
> Generic: There will be no association between the IV and DV.
> Specific to Weight and MPG: There will be no association between car weight and MPG.


Some Practice
- Review "Variables and Scales of Measurement" link and related practice exercise on course web page
- Review "Hypotheses" link and related practice exercise on course web page
- Write directional, non-directional, and null hypotheses for the following variable sets:

  1. IV = Class Size and DV = Student Achievement
  2. IV = Student sex (male/female) and DV = Verbal Scores
  3. Students whose parents are educators will earn higher scores on a test than students whose parents are not educators. (What are the IV and DV here? Rewrite as non-directional and null.)

Answers
> 1. IV = Class Size and DV = Student Achievement

Directional: The larger the class, the lower will be student achievement
Non-directional: There is an association between class size and student achievement
Null: There is no association between class size and student achievement

> 2. IV = Student sex (male/female) and DV = Verbal Scores

Directional: Females are expected to demonstrate greater verbal scores than area males.
Non-directional: There will be a difference in verbal scores between males and female.
Null:  There will be no difference in verbal scores between males and female.

> 3. Students whose parents are educators will earn higher scores on a test than students whose parents are not educators. (What are the IV and DV here? Rewrite as non-directional and null.)

IV = Parental occupation
DV = test scores

Non-directional: There will be a difference in test scores between students with parents who are educators and students whose parents are not educators.
Null: There will be no difference in test scores between students with parents who are educators and students whose parents are not educators.

Common Errors
- Confuse categories and variables (attempting to treat a variable category like a variable)
- Incorrect comparisons (comparing IV to DV rather than correctly comparing DV across categories of IV)
- Creating arbitrary group comparisons when IV is quantitative (comparing DV between higher and lower IV)
- Using relationship when difference should be focus, or vice versa (write about difference if IV is qual., write about relationship if IV is quant.)
- Each of the above represent poor or ambiguous wording

Examples of Problematic Wording

How should these be reworded?

a. "There will be a difference between student sex and reading test scores."

b. "There is no difference between parents' occupation (educator vs. non-educator) and test scores."

c. "There will be a difference between males and weight and between females and weight."

d. "Heavy cars will have lower MPG than light cars." (Note: IV = Car weight, DV = MPG)

e. "Sex will be positively related to reading scores."

f. "Type of instruction (cooperative learning or self-paced) affects student achievement."

   This wording suffers from the same problem:

   "Type of instruction (cooperative learning or self-paced) is related to student achievement."

   (Note: 30 to 70, pre-post for both groups)

Answers

a. "There will be a difference between student sex and reading test scores."

Problem: incorrect comparison of the IV to the DV
Revised:  There will be a difference in reading test scores between males and females.

b. "There is no difference between parents' occupation (educator vs. non-educator) and test scores."

Problem: incorrect comparison of the IV to the DV
Revised: Students' test scores will be similar whether their parents' occupation is educator or non-educator.

c. "There will be a difference between males and weight and between females and weight."

Problem: treats categories of a variable as if they are variables
Revised: There will be a difference in weight between males and females.

    d. "Heavy cars will have lower MPG than light cars." (Note: IV = Car weight, DV = MPG)

Problem: Arbitrarily created groups to compare (heavy vs. light)
Revised: The heavier the car the lower will be MPG. or There is a negative association between car weight and MPG.

    e. "Sex will be positively related to reading scores."

Problem: Incorrect usage of relationship when difference should be key
Revised: There will be a difference in reading scores between males and females.

    f. "Type of instruction (cooperative learning or self-paced) affects student achievement."

      This wording suffers from the same problem:

      "Type of instruction (cooperative learning or self-paced) is related to student achievement."

      (Note: 30 to 70, pre-post for both groups)


Coop Learn      pretest = 30     treatment for 6 weeks  posttest = 70
Self-paced       pretest = 30     treatment for 6 weeks  posttest = 70

Problem: Ambiguity – specify whether groups will differ
Revised: There will be a difference in student achievement between those who use coop learning and those who use self-paced instruction.


**Univariate (one variable) Summary Measures**

Central Tendency: Median

What is the median for these scores: 4, 2, 3, 1, 5
Median = 3

What is the median for these scores: 5, 2, 6, 1, 4
Median = 4;   1 2 *4* 5 6

What is the median for these scores: 1 2 3 4 5 6
Median =( 3+4)/2 = 3.5

What is the median for these scores: 5, 2, 6, 1, 4, 6
1 2 4 5 6 6
Median = 4.5