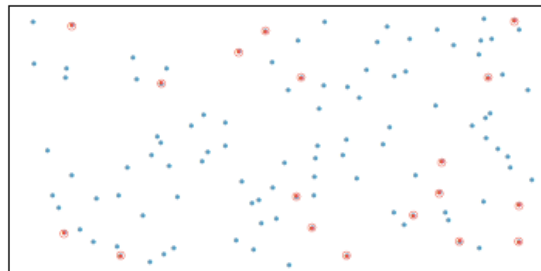


ECO239: Statistics I

Week 2

1. Use a sample to make inferences about the population



Population

- The complete set of all items that interest a researcher
- N

Examples?

Can you collect Population Data?

Sample

- A subset of population
- n
- Making inference = trying to “guess” what are the characteristics of population (N) from sample data (n).




Brain Storming

- Task: Find the average height of the students in Department of Economics (1 – 4 years (about 700 students)).
- Objective: Distribute HU ECON t-shirt for all the students, with proper XS, S, M, L, XL sizes.




- Population?
- Sample?
- How to sample?
- Let's TRY!
- n ???
- How to sample???



Data Entry

- [Height.xlsx](#)



- What could be the problem of this sample to make an inference about the population?
- How can you make it better?

Descriptive vs. Inferential Statistics

- Descriptive Statistics uses
 - ✓ Graphs, Charts
 - ✓ Numerical summaries (mean, median, variance, standard deviation...)

<= summarize data visually/numerically.

Data => Information

- Inferential Statistics
 - ✓ Based on sample statistics, try to estimate population parameters
 - ✓ Estimation
 - ✓ Hypothesis testing
 - ✓ Forecasting
 - ✓ Predictions

Information => Knowledge

Types of variables

```

graph TD
    A[all variables] --> B[numerical]
    A --> C[categorical]
    B --> D[continuous]
    B --> E[discrete]
    C --> F[Nominal]
    C --> G[ordinal]
  
```

Categorical Variable

- Observations belong to categories/groups

Nominal Categorical

Encoded Categories
No ordering, measurement, directions.
e.g. 1. Male, 2. Female
1. Yes, 2. No, 3. Don't know

Categorical Variables

Ordinal

Ranking, order, scale

e.g. 1. Strongly Agree, 2. Agree, 3. Neutral, 4. Disagree, 5. Strongly Disagree.

Numerical Variables

- Include discrete and continuous variables

Discrete Variable: Counting process

e.g. # brothers/sisters, #married/single, #cars

Continuous Variable: Measurement process

e.g. height, weight, income, distance,

Types of variables

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender:

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: (Q: How many hours do you usually sleep at night?)

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*
- bedtime: (Q: What time do you usually go to bed?
8-10, 10-12, 12-2, 2-4)

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*
- bedtime: *categorical, ordinal*

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*
- bedtime: *categorical, ordinal*
- countries: (Q: How many countries have you visited?)

Types of variables (cont.)

	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*
- bedtime: *categorical, ordinal*
- countries: *numerical, discrete*

Types of variables (cont.)

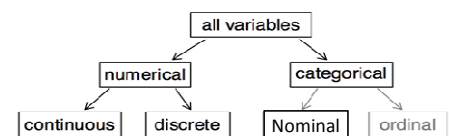
	gender	sleep	bedtime	countries
1	male	5	12-2	13
2	female	7	10-12	7
3	female	5.5	12-2	1
4	female	7	12-2	
5	female	3	12-2	1
6	female	3	12-2	9

- gender: *categorical*
- sleep: *numerical, continuous*
- bedtime: *categorical, ordinal*
- countries: *numerical, discrete*

Practice

What type of variable is a telephone area code?

- numerical, continuous
- numerical, discrete
- categorical, nominal
- categorical, ordinal



Practice

What type of variable is a telephone area code?

- (a) numerical, continuous
- (b) numerical, discrete
- (c) *categorical*
- (d) categorical, ordinal

In your project 1,

- You have to include questions to obtain at least two for each type of variable. (Total at least 8 questions)
 - categorical variable – nominal
 - categorical variable – ordinal
 - numerical variable – discrete
 - numerical variable - continuous

Describing Data Graphically

Options for **Categorical Variables**

- Frequency Distribution Table
- Bar Chart
- Pie Chart
- Parato Diagram

* Always consider what kind of graphs/tables describe your data the best, answer your question the best.

Tables/Graphs for Categorical Data

Frequency Distribution Table

⇒ Summarize Data by Category

Category	Frequency



Frequency Distribution Table

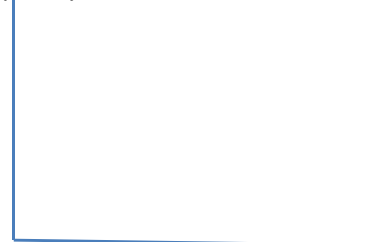
e.g. Which City are you from?

City


Graph for Categorical Variable

Bar Chart

Frequency




Category



Bar Chart

- [City](#)



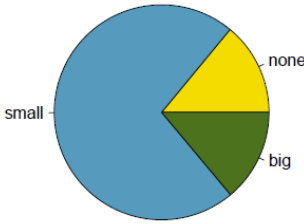

Q: If you have data for 2 years, what do you do?

[City](#)

Graphs for Categorical Data


Pie Chart

- Calculate % of each category and put them in a PIE according to their share.

Pie Chart

- [City](#)



- When do you think the pie charts are useful???

=> When the share of each category is your interest.

Graphs for Categorical Data

Parato Diagram

- Bar Chart (organized in descending order)
+ Cumulative polygon

Used to separate the “vital few” from “trivial many”

Procedures:

1. Re-order the category in descending order
2. Create a bar chart
3. Compute % in each category
4. Compute Cumulative %
5. Add Cumulative Polygon

e.g. Parato Diagram

- 400 defective cell phone has been returned to your office for the examination of the cases of the defect. Analyze what are the **vital causes** for **80% of the defects** which can be fixed at minimum cost. Fixing each costs your company \$100,000.

	Manufacturing Errors	Frequency
A.	Poor Alignment	223
B.	Missing Parts	25
C.	Cracked Case	21
D.	Paint Flaw	78
E.	Electrical Short	19
F.	Bad Weld	34
	TOTAL	400



Parato Diagram

- [Cell phone](#)
- In order to fix 80% of the problem, you have to fix
 - ✓ _Poor Alignment__
 - ✓ _Pain Flaw__
 - ✓ _Bad Weld__

Tables and Graphs for Numerical Data

Options for **Numerical Variables**

- Frequency Distribution & Cumulative Distribution
- Histogram
- Box Plot
- Ogive
- Stem-and-Leaf Display

Table for Numerical Data

- Frequency Distributions
- [Finalscore](#)

Interval (Class)	Frequency

Frequency Distribution Table: How to determine the classes?

- Step 1: sort raw data in ascending order (small-> large)
- Step 2: Find the range of data (100-0 = 100)
- Step 3: Determine the number of interval (classes) k
- Step 4: Compute interval width, w
- Step 5: Determine interval boundaries
- Step 6: Count observations & assign to each interval.

RULES!

1. Intervals should have the same width "w".

$$w = \frac{\text{Largest number} - \text{Smallest number}}{\text{\# of desired intervals, } k}$$

2. Use at least 5, but no more than 15-20 intervals
3. Intervals NEVER overlap
4. Round up the interval width to get desirable interval endpoints.

**Create Frequency Distribution Table for**

- [Finalscore](#)

Cumulative Distribution

Include

- ✓ Frequency (Count)
- ✓ Relative Frequency (% of Count)
- ✓ Cumulative Frequency (Cumulative Count)
- ✓ Relative Cumulative Frequency (% of Cumulative Count)

**Create Cumulative Distribution Table for**

- [Finalscore](#)