Empirical Loop

Descriptive Statistics

Collect Data

Research Design

Inferential Statistics

Hypothesis

1
Today’s Lecture

• The Rest of the Empirical Loop
  - Types of Data
  - Inferential vs. Descriptive Statistics

• Questions to Ask When Reading Research

• Descriptive Statistics
  - Frequency Distributions
Types of Data:

What are people’s favorite foods?

What are people’s letter grade?

How tall are people?
# Types of Data: Levels of Measurement

<table>
<thead>
<tr>
<th>Level</th>
<th>Properties</th>
<th>Observations Reflect</th>
<th>Example</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio/Interval</td>
<td>true zero equal intervals order classification</td>
<td>Measurable differences in total amount</td>
<td>Height</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Ordinal</td>
<td>order classification</td>
<td>Differences in degree</td>
<td>Grades</td>
<td>Ranked</td>
</tr>
<tr>
<td>Nominal</td>
<td>classification</td>
<td>Differences in kind</td>
<td>Favorite Food</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>
Nominal: Classification

What are people’s favorite foods?

45 people

42 people

3 people
## Types of Data: Levels of Measurement

<table>
<thead>
<tr>
<th>Level</th>
<th>Properties</th>
<th>Observations Reflect</th>
<th>Example</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio/Interval</td>
<td>true zero equal intervals order classification</td>
<td>Measurable differences in total amount</td>
<td>Height</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Ordinal</td>
<td>order classification</td>
<td>Differences in degree</td>
<td>Grades</td>
<td>Ranked</td>
</tr>
<tr>
<td>Nominal</td>
<td>classification</td>
<td>Differences in kind</td>
<td>Favorite Food</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>
Ordinal: Classification & Order

What are people’s letter grades?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>23</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
# Types of Data: Levels of Measurement

<table>
<thead>
<tr>
<th>Level</th>
<th>Properties</th>
<th>Observations Reflect</th>
<th>Example</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio/Interval</td>
<td>true zero equal intervals order classification</td>
<td>Measurable differences in total amount</td>
<td>Height</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Ordinal</td>
<td>order classification</td>
<td>Differences in degree</td>
<td>Grades</td>
<td>Ranked</td>
</tr>
<tr>
<td>Nominal</td>
<td>classification</td>
<td>Differences in kind</td>
<td>Favorite Food</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>
Interval/Ratio: Classification, Order, Equal Intervals, & True Zero

How tall are people?

<table>
<thead>
<tr>
<th></th>
<th>0'-1'</th>
<th>1'-2'</th>
<th>2'-3'</th>
<th>3'-4'</th>
<th>4'-5'</th>
<th>5'-6'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>24</td>
<td>26</td>
</tr>
</tbody>
</table>
Is there a difference between interval and ratio?

**Interval** (no true zero)
- IQ score
- Grade Point Average
- Fahrenheit Temperature

**Ratio** (true zero)
- Kelvin Temperature
- Income
- Family Size
Types of Data: Levels of Measurement

Approximately Ratio/Interval

- IQ score
- Grade Point Average

Is the difference between an IQ of 100 and 120 the same as that between 60 and 80?
FIGURE 1.2
Overview: types of data and levels of measurement.
Identify the following data as Nominal, Ordinal, or Interval/Ratio.

A) Ethnic Group
B) Age
C) Family Size
D) Academic Major
E) Political Preference
F) Cooking Time for Pasta
G) Parole Violations by Convicts
H) Freshman, Sophomore, Junior, Senior
I) SAT Score
J) Net Worth ($)
K) Favorite Sport
L) Gender
M) Position in line at a cashier
N) A Town’s Population
O) Car speed
P) Attitude towards statistics (Like, Neutral, Dislike)
Types of Data:
Continuous vs. Discrete

Interval/Ratio

Continuous  Discrete
Types of Data: Continuous vs. Discrete

**Continuous**-Numbers always separated by another number
- Fahrenheit Temperature
- Kelvin Temperature
- Grade Point Average

**Discrete**-Numbers separated by gaps
- Vocabulary Size
- Number of Correct Answers
- Number of Books Owned
Identify which of the following Interval/Ratio data are Continuous.

A) Age  
B) Family Size  
C) Cooking Time for Pasta  
D) Parole Violations by Convicts  
E) Number of Votes  
F) Money in a Bank Account  
G) A Town’s Population  
H) Car speed
Identify which of the following Interval/Ratio data are Continuous.

A) Age
B) Family Size
C) Cooking Time for Pasta
D) Parole Violations by Convicts
E) Number of Votes
F) Money in a Bank Account
G) A Town’s Population
H) Car speed
Interval/Ratio: Classification, Order, Equal Intervals, & True Zero

Continuous Measurements are always approximate

“She’s 5 feet tall.”

She’s between 4.5 and 5.5 feet tall.
Empirical Loop

Descriptive Statistics → Collect Data → Research Design → Hypothesis → Inferential Statistics

Collect Data → Research Design → Hypothesis

Descriptive Statistics → Inferential Statistics → Descriptive Statistics
Descriptive vs. Inferential Statistics

**Descriptive Statistics** - The area of statistics concerned with organizing and summarizing information about a collection of actual observations (i.e., your sample).

**Inferential Statistics** - The area of statistics concerned with generalizing beyond actual observations (i.e., making inferences about the population).
Descriptive vs. Inferential Statistics

**Descriptive Statistics**

- Summary Statistics (Percentages, Averages, Standard Deviations)
- Graphs
Descriptive vs. Inferential Statistics

Descriptive Statistics

The average temperature in my house is 5.1 degrees Fahrenheit hotter than it was 30 days ago.
Descriptive vs. Inferential Statistics

Descriptive Statistics

90% of people polled do not use public transportation.
Descriptive vs. Inferential Statistics

Descriptive Statistics

Sally completed the marathon in 2.26 hours.
Inferential Statistics- Generalizations beyond observations (too expensive or impossible to observe everything)

• Based on probability theory and largely invented in the 20th century
• “Statistically significant” differences
• “Confidence intervals”
Descriptive vs. Inferential Statistics

Inferential Statistics

The earth’s average near surface atmospheric temperature rose 1.1 (degrees Fahrenheit in the 20th century.)
Inferential Statistics

If these trends continue, the average global temperature will rise between 2.5 and 10.4 degrees Fahrenheit by 2100.
Inferential Statistics

Based on a poll of 5000 people, 90% of Americans do not use public transportation.
Descriptive vs. Inferential Statistics

Inferential Statistics

Regular exercise significantly ($p<.05$) reduces one’s risk of heart disease.
Which of the following statements are examples of inferential statistics?

A) On average students in my statistics class are 20 years old.
B) The population of the world now exceeds 6 billion (6,000,000,000).
C) Four years has been the most frequent term of office served by American presidents.
D) A recent poll indicates that 74% of all college students favor capital punishment.
Which of the following statements are examples of inferential statistics?

A) On average students in my statistics class are 20 years old.
B) The population of the world now exceeds 6 billion \((6,000,000,000)\).
C) Four years has been the most frequent term of office served by American presidents.
D) A recent poll indicates that 74% of all college students favor capital punishment.
Empirical Loop

- Descriptive Statistics
- Collect Data
- Research Design
- Inferential Statistics
- Hypothesis
Today’s Lecture

• The Rest of the Empirical Loop
  - Types of Data
  - Inferential vs. Descriptive Statistics

• Questions to Ask When Reading Research

• Descriptive Statistics
  - Frequency Distributions
Questions to Ask When Reading Research

1. Who produced and who interpreted the data?
2. Did they measure what they should have measured?
3. Did they use clear and reasonable definitions?
4. Is the research published in a peer-reviewed journal?
5. How did they get the data?

Largely based on material from:
I. Who produced/interpreted the data?

- Academics (Researchers at non-profit universities or research centers)
- Government Agencies (e.g., the census, FBI)
- For-profit corporations (e.g., drug companies, pollsters)
- Non-profit activists (e.g., Amnesty International, National Center for Public Policy Research)
I. Who produced/interpreted the data?

- What bias might the researchers/interpreters have?
- What are the consequences for the researchers/interpreters if they’re wrong?
- If the people interpreting the data are different than the people who produced the data, is it possible they’re misrepresenting the data?
I. Who produced/interpreted the data?

Proponents of the Bush administration’s 2003 tax cut proclaimed that on average, families receive a tax cut of $1000. Opponents of the cut countered that more than half of all families will receive a tax cut by less than $100.

Questions to Ask When Reading Research

1. Who produced and who interpreted the data?
2. Did they measure what they should have measured?
3. Did they use clear and reasonable definitions?
4. Is the research published in a peer-reviewed journal?
5. How did they get the data?
2. Did they measure what they should have measured?

- If you were trying to test the same hypothesis, what would you have done?

Does using marijuana, alcohol, and nicotine lead to using cocaine?
2. Did they measure what they should have measured?

“The CASA (Center for Addiction and Substance Abuse at Columbia) study establishes a clear progression that begins with gateway drugs and leads to cocaine use: nearly 90% of people who have ever tried cocaine used all three gateway substances [alcohol, marijuana, & cigarettes] first.”

2. Did they measure what they should have measured?

“...over 95% of all Volvos registered here [in the U.S.A.] in the last 11 years are still on the road.”

-1968 Volvo Advertisement

Questions to Ask When Reading Research

1. Who produced and who interpreted the data?
2. Did they measure what they should have measured?
3. Did they use clear and reasonable definitions?
4. Is the research published in a peer-reviewed journal?
5. How did they get the data?
3. Did they use clear and reasonable definitions?

Japan has 1 lawyer, “bengoshi,” for every 10,000 people.

The USA has 28 lawyers for every 10,000 people.

3. Did they use clear and reasonable definitions?

Japan has 32 law school graduates for every 10,000 people.

The USA has 28 lawyers for every 10,000 people.

3. Did they use clear and reasonable definitions?

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>If the her own health is endangered by the pregnancy</td>
<td>92%</td>
</tr>
<tr>
<td>If she became pregnant because of rape</td>
<td>84%</td>
</tr>
<tr>
<td>If there is a strong chance of serious birth defect</td>
<td>82%</td>
</tr>
<tr>
<td>If she is married and does not want more children</td>
<td>47%</td>
</tr>
<tr>
<td>If the family is poor and cannot afford more children</td>
<td>47%</td>
</tr>
<tr>
<td>If she does not want to marry the man</td>
<td>45%</td>
</tr>
<tr>
<td>If she wants it for any reason</td>
<td>45%</td>
</tr>
</tbody>
</table>

Questions to Ask When Reading Research

1. Who produced and who interpreted the data?
2. Did they measure what they should have measured?
3. Did they use clear and reasonable definitions?
4. Is the research published in a peer-reviewed journal?
5. How did they get the data?
4. Is the research published in a peer-reviewed journal?

Peer-Review

1. Scientists submit a manuscript of some research to a peer-reviewed journal.
2. Editors of the journal select other scientists to evaluate the research to ensure that it is quality work. The identity of the scientists who evaluate the work is usually kept secret from the people who produced the research.
3. The evaluators approve the manuscript, request improvements, or reject it.
4. Is the research published in a peer-reviewed journal?

Prominent Peer-Reviewed Journals
4. Is the research published in a peer-reviewed journal?

Prominent Peer-Reviewed Journals

PNAS

NEJM
About PNAS

Proceedings of the National Academy of Sciences of the United States of America

PNAS is one of the world's most-cited multidisciplinary scientific serials. Since its establishment in 1914, it continues to publish cutting-edge research reports, commentaries, reviews, perspectives, colloquium papers, and actions of the Academy. Coverage in PNAS spans the biological, physical, and social sciences. PNAS is published weekly in print, and daily online in PNAS Early Edition. The PNAS impact factor is 9.64 for 2006. PNAS is available by subscription.

PNAS is abstracted and/or indexed in: Index Medicus, PubMed Central, Current Contents, Medline, SPIN, JSTOR, ISI Web of Science, and BIOSIS.

http://www.pnas.org/misc/about.shtml
Recommend a JSM subscription to your library today!

The Journal of Sexual Medicine
Basic Research and Clinical Studies in Male and Female Sexual Function and Dysfunction

The Official Journal of the International Society for Sexual Medicine and the International Society for the Study of Women's Sexual Health

Edited by:
Irwin Goldstein

ISI Journal Citation Reports® Ranking: 2006: 4/55 (Urology & Nephrology)
Impact Factor: 4.676

http://jsm.issir.org/
Some search engines for scientific publications:

Google Scholar
http://scholar.google.com/

ISI Web of Knowledge
http://portal.isiknowledge.com/

PubMed
Questions to Ask When Reading Research

1. Who produced and who interpreted the data?
2. Did they measure what they should have measured?
3. Did they use clear and reasonable definitions?
4. Is the research published in a peer-reviewed journal?
5. How did they get the data?
5. How did they get the data?

An experiment, observational study, or guess?

• If observational, then what evidence do they have that there is a causal relationship?

• Do they provide “confidence intervals” or “p-values”? 
5. How did they get the data?

Guessed Statistics (Mitch Snyder, 1980s)

There are 2 to 3 million homeless in the USA.

5. How did they get the data?

Guessed Statistics (T. Cullen Davis, 1980s)

Public School’s Top Problems in 1940’s
1. Talking
2. Chewing Gum
3. Making Noise

Public School’s Top Problems in 1980’s
1. Drug Abuse
2. Alcohol Abuse
3. Pregnancy

5. How did they get the data?

Do they provide “confidence intervals” or “p-values”?

Source: http://www.galluppoll.com/content/?ci=1723&pg=1
5. How did they get the data?

“Results are based on telephone interviews with 1,000 national adults, aged 18 and older, conducted June 23-25, 2006. For results based on the total sample of national adults, one can say with 95% confidence that the margin of sampling error is ±3 percentage points. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.”

Source: http://poll.gallup.com/content/?ci=23494
5. How did they get the data?

An experiment, observational study, or guess?

- Do they provide “confidence intervals” or “p-values”?

Researchers found that exercise **significantly** \((p<.05)\) reduces one’s risk of heart disease.
5. How did they get the data?

An experiment, observational study, or guess?

- If observational, then what evidence do they have that there is a causal relationship?

"In typical beer-drinking cultures like Holland, Germany, England, Canada and Australia there is more violence than in Italy or Greece or Spain, where people drink wine."

The Mercury, 23 August, 1994, p.8
5. How did they get the data?

- Was the sample small or reasonably large?
- Was the sample biased or did the researchers randomly sample from a representative pool?
5. How did they get the data?

Beware the Pushy Fish Eater

“Men who eat a lot of fish are driven by ambition and the desire for success, British researchers claim.

Seven in 10 men who frequently eat canned tuna, sardines, salmon, mackerel or kippers admit to being ambitious, and one in two rate themselves as more successful than others.”

The Mercury, 7 May, 1991
5. How did they get the data?

Was the sample small or reasonably large?

250 people

125 women
125 men

6 eat a lot fish
(4 rated themselves as ambitious)

5. How did they get the data?

Was the sample small or reasonably large?

“Based on UCSD Healthy Relationship survey 12/05 (n= 1834 UCSD students).”
5. How did they get the data?

Was the sample biased or did the researchers randomly sample from a representative pool?

I**deal**: Randomly select participants from the population of interest.

R**arely** possible for research on humans.
5. How did they get the data?

The Infamous Literary Digest Poll of 1936

Before the 1936 presidential election between F. D. Roosevelt and Alf Landon, Literary Digest sent questionnaires to 10 million car and telephone owners and magazine subscribers. Based on the 23% of the surveys that were returned, LD predicted that Landon would win by a 3 to 2 margin. LD had been very successful at predicting previous election results.

Previous to the Digest poll, George Gallup surveyed 50,000 people and predicted that Roosevelt would win. Gallup was correct.

First Homework Assignment

http://www.cogsci.ucsd.edu/classes/WI09/COGS14/calendar.html
A few months ago, the science news website ScienceNOW published a report on differences in language comprehension between girls and boys:

He Heard, She Heard.

Read the article and answer the following questions about the study. If the article does not provide enough information to answer any of these questions, then simply say so and answer the questions you can.

A. Who did the research? What research institution are the researchers affiliated with? What biases might the researchers have?
B. Who is interpreting the research in the article? The researchers? Non-scientists? Both?
C. Is the research published in a peer-reviewed journal? If so, which one?
D. What kind of research is it? An experiment or observational study? Name the dependent and independent variables (if there are any independent variables). If it is an observational study, is there a good reason to think there might be a causal relation between the dependent variables?
E. Describe what type of data each independent and dependent variable is (i.e., nominal, ordinal, or interval/ratio)
F. How big was the sample size? Was it a random, representative sample? Do they report confidence intervals or p-values?
G. Do you find the research compelling? If so, why? If not, what would you have done differently?
Today’s Lecture

• The Rest of the Empirical Loop
  - Types of Data
  - Inferential vs. Descriptive Statistics

• Questions to Ask When Reading Research

• Descriptive Statistics
  - Frequency Distributions
Chapter 2

Data

Frequency Distributions

<table>
<thead>
<tr>
<th>Height</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-7 ft</td>
<td>16</td>
</tr>
<tr>
<td>3-5 ft</td>
<td>13</td>
</tr>
</tbody>
</table>

Graphs

Yes

No
Chapter 2: Frequency Distributions

1. Frequency Distributions
   - Ungrouped
   - Grouped
2. Outliers
3. Relative Frequency Distributions
4. Cumulative Frequency Distributions
5. Cumulative Relative Frequency Distributions
6. Percentile Ranks
Frequency: The number of times something happened
# Frequency Distribution

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>4</td>
</tr>
<tr>
<td>Tails</td>
<td>6</td>
</tr>
</tbody>
</table>
Ungrouped Frequency Distribution: You count the occurrence of every possible value.
Ungrouped Frequency Distribution

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>4</td>
</tr>
<tr>
<td>Tails</td>
<td>6</td>
</tr>
</tbody>
</table>
Height of 9 People at the Beach (in cm)

178  176  181  
181  176  185  
185  180  181
<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>176</th>
<th>177</th>
<th>178</th>
<th>179</th>
<th>180</th>
<th>181</th>
<th>182</th>
<th>183</th>
<th>184</th>
<th>185</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Grouped Frequency Distribution: You count the number of samples that fall within a range of values.

Only possible with Ordinal and Interval/Ratio Data
### Ungrouped Frequency Distribution

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>176</td>
<td>2</td>
</tr>
<tr>
<td>177</td>
<td>0</td>
</tr>
<tr>
<td>178</td>
<td>1</td>
</tr>
<tr>
<td>179</td>
<td>0</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
</tr>
<tr>
<td>181</td>
<td>3</td>
</tr>
<tr>
<td>182</td>
<td>0</td>
</tr>
<tr>
<td>183</td>
<td>0</td>
</tr>
<tr>
<td>184</td>
<td>0</td>
</tr>
<tr>
<td>185</td>
<td>2</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>175-179</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>Frequency</td>
<td>3</td>
</tr>
</tbody>
</table>
Guidelines for Frequency Distributions: Basic

1. Each observation should be included in one, and only one, class.

Example:

175-179, 180-184, 185-189

176-183, 180-185

175-179, 180-184

pg. 29
Guidelines for Frequency Distributions: Basic

2. List all classes, even those with zero frequencies.

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>176</th>
<th>177</th>
<th>178</th>
<th>179</th>
<th>180</th>
<th>181</th>
<th>182</th>
<th>183</th>
<th>184</th>
<th>185</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Guidelines for Frequency Distributions: Basic

3. All classes (with both upper and lower boundaries) should be equal width.

Example:

175-179, 180-184, 185-189

175-179, 180-185
Guidelines for Frequency Distributions: Optional

4. All classes should have both an upper boundary and a lower boundary.

Example:

175-179, 180-184, 185-189
below-174, 175-179, 180-184, 185-above
Guidelines for Frequency Distributions: \textbf{Optional}

5. Select the width of classes from convenient numbers, such as 1,2,3, ... 10, particularly 5 and 10 or multiples of 5 and 10.

\textbf{Example:}

175-179, 180-184, 185-189

170-180, 181-191, 192-202
Guidelines for Frequency Distributions: Optional

6. The lower boundary of each class should be a multiple of the class width.

Example:
175-179, 180-184, 185-189
177-181, 182-186, 187-191
Guidelines for Frequency Distributions: Optional

7. In general, aim for a total of approximately ten classes.

Example (too few observations!):
175-179, 180-184, 185-189
177-181, 182-186, 187-191
How do you construct frequency distributions?

Data from: Table 1.1, pg. 6

Guidelines: pg. 31
Constructing Frequency Distributions

1. Find the “data range,” the difference between the largest and smallest observations.

\[ 245 \text{ lbs.} - 133 \text{ lbs.} = 112 \text{ lbs.} \]
Constructing Frequency Distributions

2. Find the class width required to span the data range by dividing the range by the desired number of classes (usually 10).

\[
\frac{112 \text{ lbs.}}{10} = 11.2 \text{ lbs.}
\]
3. Round off to the nearest convenient width (e.g., 1, 2, 3, 5, 10).

\[
\frac{112 \text{ lbs.}}{10} = 11.2 \text{ lbs.}
\]

10
Constructing Frequency Distributions

4. Determine where the lowest class should begin (Ordinarily this should be a multiple of class width).

Class Width = 10 lbs.
Smallest Observation = 133 lbs.
130-
Constructing Frequency Distributions

5. Determine where the lowest class should end by adding the class width to the lower boundary and subtracting one “unit of measurement.”

Class Width = 10 lbs.
130-139
Constructing Frequency Distributions

6. Working upward, list as many equivalent classes (usually about 10) as are required to include the largest observation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Constructing Frequency Distributions

7. Indicate with a tally the class in which each observation falls.

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>130-139</td>
<td>III</td>
</tr>
<tr>
<td>140-149</td>
<td>I</td>
</tr>
<tr>
<td>150-159</td>
<td></td>
</tr>
<tr>
<td>160-169</td>
<td></td>
</tr>
<tr>
<td>170-179</td>
<td></td>
</tr>
<tr>
<td>180-189</td>
<td></td>
</tr>
<tr>
<td>190-199</td>
<td></td>
</tr>
<tr>
<td>200-209</td>
<td></td>
</tr>
<tr>
<td>210-219</td>
<td></td>
</tr>
<tr>
<td>220-229</td>
<td></td>
</tr>
<tr>
<td>230-239</td>
<td></td>
</tr>
<tr>
<td>240-249</td>
<td></td>
</tr>
</tbody>
</table>
Constructing Frequency Distributions

8. Replace the tally count for each class with a frequency and show the total of all frequencies.

```
<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>130-139</td>
<td>3</td>
</tr>
<tr>
<td>140-149</td>
<td>1</td>
</tr>
<tr>
<td>150-159</td>
<td>17</td>
</tr>
<tr>
<td>160-169</td>
<td>12</td>
</tr>
<tr>
<td>170-179</td>
<td>7</td>
</tr>
<tr>
<td>180-189</td>
<td>3</td>
</tr>
<tr>
<td>190-199</td>
<td>4</td>
</tr>
<tr>
<td>200-209</td>
<td>2</td>
</tr>
<tr>
<td>210-219</td>
<td>0</td>
</tr>
<tr>
<td>220-229</td>
<td>3</td>
</tr>
<tr>
<td>230-239</td>
<td>0</td>
</tr>
<tr>
<td>240-249</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
</tr>
</tbody>
</table>
```
Constructing Frequency Distributions

9. Supply headings for both rows (or columns) and a title for the table.

Weights of Male Statistics Students

<table>
<thead>
<tr>
<th>Weight (lbs)</th>
<th>130-139</th>
<th>140-149</th>
<th>150-159</th>
<th>160-169</th>
<th>170-179</th>
<th>180-189</th>
<th>190-199</th>
<th>200-209</th>
<th>210-219</th>
<th>220-229</th>
<th>230-239</th>
<th>240-249</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>1</td>
<td>17</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>53</td>
</tr>
</tbody>
</table>
Chapter 2: Frequency Distributions

1. Frequency Distributions
   - Ungrouped
   - Grouped
2. Outliers
3. Relative Frequency Distributions
4. Cumulative Frequency Distributions
5. Cumulative Relative Frequency Distributions
6. Percentile Ranks
Outliers: very extreme observations
Outliers: Very extreme observations

1. Check for accuracy
2. Exclude the observations?
3. Might enhance understanding
<table>
<thead>
<tr>
<th>Income</th>
<th>Age</th>
<th>Family Size</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$6450</td>
<td>20</td>
<td>2</td>
<td>2.30</td>
</tr>
<tr>
<td>$4820</td>
<td>19</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>$5650</td>
<td>61</td>
<td>3</td>
<td>3.56</td>
</tr>
<tr>
<td>$1720</td>
<td>32</td>
<td>6</td>
<td>2.89</td>
</tr>
<tr>
<td>$600</td>
<td>19</td>
<td>18</td>
<td>2.15</td>
</tr>
<tr>
<td>$0</td>
<td>22</td>
<td>2</td>
<td>3.01</td>
</tr>
<tr>
<td>$3482</td>
<td>23</td>
<td>6</td>
<td>3.09</td>
</tr>
<tr>
<td>$25,700</td>
<td>27</td>
<td>3</td>
<td>3.50</td>
</tr>
<tr>
<td>$8548</td>
<td>21</td>
<td>4</td>
<td>3.20</td>
</tr>
<tr>
<td>Income</td>
<td>Age</td>
<td>Family Size</td>
<td>GPA</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>$6450</td>
<td>20</td>
<td>2</td>
<td>2.30</td>
</tr>
<tr>
<td>$4820</td>
<td>19</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>$5650</td>
<td>61</td>
<td>3</td>
<td>3.56</td>
</tr>
<tr>
<td>$1720</td>
<td>32</td>
<td>6</td>
<td>2.89</td>
</tr>
<tr>
<td>$600</td>
<td>19</td>
<td>18</td>
<td>2.15</td>
</tr>
<tr>
<td>$0</td>
<td>22</td>
<td>2</td>
<td>3.01</td>
</tr>
<tr>
<td>$3482</td>
<td>23</td>
<td>6</td>
<td>3.09</td>
</tr>
<tr>
<td>$25,700</td>
<td>27</td>
<td>3</td>
<td>3.50</td>
</tr>
<tr>
<td>$8548</td>
<td>21</td>
<td>4</td>
<td>3.20</td>
</tr>
</tbody>
</table>
Chapter 2: Frequency Distributions

1. Frequency Distributions
   - Ungrouped
   - Grouped
2. Outliers
3. Relative Frequency Distributions
4. Cumulative Frequency Distributions
5. Cumulative Relative Frequency Distributions
6. Percentile Ranks
**Frequency:** The number of times something happened
Relative Frequency:

\[
\frac{\text{# of times something happened}}{\text{total # of things that happened}}
\]
## Frequency Distribution

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>4</td>
</tr>
<tr>
<td>Tails</td>
<td>6</td>
</tr>
</tbody>
</table>
### Relative Frequency Distribution (proportion)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>0.4</td>
</tr>
<tr>
<td>Tails</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Heads</td>
<td>40%</td>
</tr>
<tr>
<td>Tails</td>
<td>60%</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>Frequency</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Frequency</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Frequency</th>
<th>175-179</th>
<th>180-184</th>
<th>185-189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>30%</td>
<td>50%</td>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2: Frequency Distributions

1. Frequency Distributions
   - Ungrouped
   - Grouped
2. Outliers
3. Relative Frequency Distributions
4. Cumulative Frequency Distributions
5. Cumulative Relative Frequency Distributions
6. Percentile Ranks
### Weights of Male Statistics Students

<table>
<thead>
<tr>
<th>Weight (lbs)</th>
<th>130-139</th>
<th>140-149</th>
<th>150-159</th>
<th>160-169</th>
<th>170-179</th>
<th>180-189</th>
<th>190-199</th>
<th>200-209</th>
<th>210-219</th>
<th>220-229</th>
<th>230-239</th>
<th>240-249</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>1</td>
<td>17</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td>Cumulative Frequency</td>
<td>3</td>
<td>4</td>
<td>21</td>
<td>33</td>
<td>40</td>
<td>43</td>
<td>47</td>
<td>49</td>
<td>49</td>
<td>52</td>
<td>52</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>
**Frequency Distribution**

<table>
<thead>
<tr>
<th>Party</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republican</td>
<td>4</td>
</tr>
<tr>
<td>Democrat</td>
<td>4</td>
</tr>
<tr>
<td>Independent</td>
<td>1</td>
</tr>
</tbody>
</table>
### Weights of Male Statistics Students

<table>
<thead>
<tr>
<th>Weight (lbs)</th>
<th>130-139</th>
<th>140-149</th>
<th>150-159</th>
<th>160-169</th>
<th>170-179</th>
<th>180-189</th>
<th>190-199</th>
<th>200-209</th>
<th>210-219</th>
<th>220-229</th>
<th>230-239</th>
<th>240-249</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency</strong></td>
<td>3</td>
<td>1</td>
<td>17</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td><strong>Cumulative Frequency</strong></td>
<td>3</td>
<td>4</td>
<td>21</td>
<td>33</td>
<td>40</td>
<td>43</td>
<td>47</td>
<td>49</td>
<td>49</td>
<td>52</td>
<td>52</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td><strong>Relative Cumulative Frequency</strong></td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2: Frequency Distributions

1. Frequency Distributions
   - Ungrouped
   - Grouped
2. Outliers
3. Relative Frequency Distributions
4. Cumulative Frequency Distributions
5. Cumulative Relative Frequency Distributions
6. Percentile Ranks
Percentile Ranks

Percentile Rank of an Observation:
Percentage of observations in the entire distribution with similar or smaller values than that observation.
### Exact Percentile Ranks (Ungrouped Distribution)

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>176</th>
<th>177</th>
<th>178</th>
<th>179</th>
<th>180</th>
<th>181</th>
<th>182</th>
<th>183</th>
<th>184</th>
<th>185</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Exact percentile rank of observation 180 cm is 44th (4/9).

Exact percentile rank of observation 181 cm is 78th (7/9).
Approximate Percentile Ranks (Grouped Data)

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>175-179</th>
<th>180-184</th>
<th>185-189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Approx. percentile rank of observation 180 cm is 78th.

Approx. percentile rank of observation 181 cm is 78th.
“A new research paper by Ian Dew-Becker and Robert Gordon of Northwestern University, "Where Did the Productivity Growth Go?," gives the details. Between 1972 and 2001 the wage and salary income of Americans at the 90th percentile of the income distribution rose only 34 percent, or about 1 percent per year. So being in the top 10 percent of the income distribution, like being a college graduate, wasn't a ticket to big income gains.

But income at the 99th percentile rose 87 percent; income at the 99.9th percentile rose 181 percent; and income at the 99.99th percentile rose 497 percent. No, that's not a misprint.”

Today’s Lecture

• The Rest of the Empirical Loop
  - Types of Data
  - Inferential vs. Descriptive Statistics

• Questions to Ask When Reading Research

• Descriptive Statistics
  - Frequency Distributions