

Statistics Exercises Using

Adolescent Body Mass Index (BMI) Data

from OHSU’s Let’s Get Healthy

School Health Fairs

Developed by Jay Schauer

Wilsonville High School Science and Statistics Teacher

Let’s Get Healthy! Program Educator

Sponsored by Lisa Marriott, PhD

Associate Director of Let’s Get Healthy!

Table of Contents

1. What is Let’s Get Healthy (LGH)?
2. Justification of using Let’s Get Healthy for Statistics
3. Using Body Mass Index (BMI) data in the classroom.
4. Exercise 1 – Collecting Data: Observational Studies and Experiments
5. Exercise 2 – Using LGH Data for Descriptive Statistics
6. Exercise 3 – Using LGH Data for Confidence Intervals
7. Exercise 4 – Using LGH Data for Hypothesis Testing
8. Figure 1 – Tinker Plot comparison of Female and Male BMI Distributions
9. Table 1 – “Population” Data of Oregon 13 Year Old Female BMI
10. Table 2 - “Population” Data of Oregon 13 Year Old Male BMI
11. Student copy of Information and Questions for Exercise 1
12. Teacher Key and Comments for Exercise 1
13. Sample of Student work for Exercise 2
14. Student copy of Questions for Exercise 4
15. Teacher Key and Comments for Exercise 4

I. What is Let’s Get Healthy?

***About Let’s Get Healthy!***

Let’s Get Healthy! is an education and research exhibit from Oregon Health & Science University (OHSU) that allows participants to learn about their health while contributing to science. The exhibit is funded by the National Institutes of Health (NIH) and has been held at schools and communities throughout Oregon since 2007.

***About the Health Stations***

As part of Let’s Get Healthy!, participants can visit different stations to learn about their health. There are six data feedback health stations and four hands-on exhibits. The data feedback stations help participants learn about their diet, body composition, blood pressure, sun exposure, blood pressure and sleep patterns. Participants receive immediate results and recommendations from each station.

***About the Data collected from the Health Fairs***

LGH’s open-source database is populated with anonymous health-related data from each Let’s Get Healthy! Event. The data are available to researchers, communities, schools and students who use the data for school projects, community and workplace wellness policy decisions and scientific research. Requests are managed by our program staff and advisory board and provided in Excel or statistical formats (SPSS or SAS).

***About the Data for these Exercises***

The Excel data sheets that are included with these exercises come from four middle school fairs during the 2010-2012 school years. Fairs collected data from students at St. Helens Middle School, three West Linn-Wilsonville Middle schools, middle school students attending a fair in LaGrande and middle schools students attending a fair in Bend. These exercises only used data from 13 year olds of both genders who participated in the BMI station. Of the 6,000+ students who participated in these fairs there were 328 thirteen year old females with BMI data and 324 thirteen year old males with BMI data. These 652 individuals provide the basis for these exercises.

***Future Opportunities***

Let’s Get Healthy is working on making their database searchable for access by students and educators for additional research and data exercises through an extension of the program known as CHIDR (Community Health Interactive Data Resource) Chatter. This will allow individuals and groups to access all the variables collected at the fairs and not just BMI.

II. Justification of Exercises

Why use this data when there are already plenty of data sets to use in most statistics books and most of those books use the same data for multiple exercises similar to this collection of exercises?

1. The data is from a population (middle school adolescents) that is very relevant to the students in high school level classes. They are recently removed from this age group and may have siblings in this age group.
2. The data is relatively local. Let’s Get Healthy Fairs at this point have collected adolescent health data from four distinct regions of Oregon – St. Helens, Union County, West Linn-Wilsonville School District and Bend.
3. Adolescent health, and the subsequent implications for adult health, is a local, state and national conversation and concern. Our high school students have the opportunity to make choices that can affect their own health and they can be advocates for changes that affect public health.
4. The Let’s Get Healthy database has such a large sample of same age adolescents that it can be treated as a population that can be sampled from allowing teachers and students to compare sample statistics to population parameters.

III. Using Body Mass Index (BMI) data in the classroom - An Introduction and Caution.

Body Mass Index (BMI) is used as an indicator of healthy weight for a variety of reasons. BMI is calculated by dividing an individual’s weight (in kilograms) divided by their (height in m)2 and varies with age. BMI is used to delineate weight categories into underweight, healthy weight, overweight and obese. The BMI values associated with these weight categories varies by gender and with age. The values are abstract enough that unless students are calculating them in class (which they are not in these exercises) students within class should not be identified with any particular BMI or weight category.

Weight categories help health officials inform individuals of where their weight may be, where they may wish it to be and inform them of long term consequences of habits and choices they make that is related to their current and long term health.

Exercise 1: Observational Studies and Experiments

**Overview of Lesson Plan:**

This lesson is designed to introduce students to the Let’s Get Healthy program and the differences between Observational Studies and Experiments and what those differences mean for interpretation and application of findings from these types of data collection processes.

**Big Understanding:**

Observational Studies (like LGH) involves collecting data that already exists and may suggest relationships but can not be used to ascribe cause and effect. Experiments involve manipulating variables and apply treatments and can be use to assess cause and effect.

**Essential Questions:**

What is an observational study?

What is an experiment?

How do you identify the differences between observational studies and experiments?

How do observational studies lead to experiments?

**Grades:**

10th – 12th Grade

**Subjects:**

Advanced Placement Statistics or other Statistics Class

**Time Required:**

Approximately 15-45 minutes depending on whether it is used for an introductory or a review exercise, whether it is used for an individual or small group assignment and also how much discussion is used to follow up the exercise.

**Objectives:**

Upon completion of this exercise students will be able to:

* Define an observational study.
* Describe the distinguishing components of an experiment.
* Identify an investigation as either an observational study or an experiment.
* Suggest an experiment that could be created based on findings from an observational study.

**Materials:**

* “Let’s Get Healthy” Statistics lesson plan.
* TI-83, 84, or Inspire calculator.
* Student copy of Information and Questions for Exercise 1
* Teacher Key and Comments for Exercise 1

**Preparation:**

Copy enough sets of questions for your class or be prepared to project the descriptions and have students write the questions down.

**Activities/Procedures:**

**Step 1** – Pass out student sheet for this exercise.

**Step 2** –Have students complete sheet for Exercise one either individually, with a partner or a small group depending on instructor preference.

**Step 3** – Review answers with students with special emphasis on the experimental design question. This is an especially useful time to have students compare their designs with each other to identify strengths and weaknesses of their proposed experiments.

**Assessment/Reflection:**

Up to instructor. Assignment can be turned in, checked off or scored by student during discussion and review. Similar concepts are typically covered in multiple choice or open-ended form on most unit or cumulative tests dealing with this area of statistics.

**Resources:**

School adopted AP Statistics Textbook and associated sections on collecting data using observational studies and experiments.

**Standards:**

This lesson meets part of the Collecting Data syllabus portion of the Advanced Placement Statistics course.

Exercise 2: Applying Descriptive Statistics to LGH data

**Overview of Lesson Plan:**

This exercise uses a sample of data from the LGH 13 year old BMI data population and walks students through a variety of ways to describe that sample both numerically in text and graphically. It is designed to cover the majority of descriptive statistics in the AP Statistics curriculum. Students collect a random sample of data, create a frequency table, a cumulative percentage table, a histogram, a cumulative percentage plot, a back-to-back stem-and-leaf plot, a side-by-side bar graph, the five number summary and a box-and-whisker plot. They also discuss whether there may be groups that are under-represented in the data from the LGH health fairs.

**Big Understanding:**

Students should be able to understand which tables and which graphs are appropriate for describing which types of data.

**Essential Questions:**

How can you use a calculator to generate a set of random numbers for collecting a random sample of data?

How do you create a frequency table from a data set?

How do you create a cumulative percentage table from a frequency table?

How do you convert a frequency table into a histogram?

How do you convert a cumulative percentage table into a cumulative percentage graph?

How do you create a back-to-back stem-and-leaf plot from two samples?

How do you generate mean, standard deviation and the five number summary using a calculator?

How do you generate side-by-side box-and-whisker plots from two five number summaries?

What are some aspects of data collection that can lead to statistically unrepresentative samples?

**Grades:**

10th – 12th Grade

**Subjects:**

Advanced Placement Statistics or other Statistics Class

**Time Required:**

Up to 90 minutes. This exercise can be completed in one block class or two regular classes. It can be finished at home if necessary, but students often need help or confirmation throughout the exercise.

**Objectives:**

* Upon completion of this activity, students will be able to:
* Use a calculator to generate a set of random numbers for collecting a random sample of data.
* Create a frequency table from a data set.
* Create a cumulative percentage table from a frequency table.
* Convert a frequency table into a histogram.
* Convert a cumulative percentage table into a cumulative percentage graph.
* Create a back-to-back stem-and-leaf plot from two samples.
* Generate mean, standard deviation and the five number summary using a calculator.
* Generate side-by-side box-and-whisker plots from two five number summaries.
* Discuss aspects of data collection that can lead to statistically unrepresentative samples.

**Materials:**

* “Let’s Get Healthy” Statistics lesson plan.
* TI-83, 84, or Inspire calculator.
* Graph paper – 2 or 3 sheets per student
* Printouts of the 13 year old Female and Male BMI pages included with this document or Excel File – “LGH 13 Year Old BMI Data” or computers with Excel to access Excel files.
* Copy of student sample of exercise results.

**Preparation:**

Teacher should have data files available for students either electronically or hard copy. Teacher can either provide data sheet included with the lesson or have students record data on their own paper or mark their samples directly on hard copies of data sheets.

**Activities/Procedures:**

**Step 1**–“Collecting” a simple random sample from the “population”.

Option 1 - Have students use the first 30 females and the first 30 males from the table (already randomly sorted). The advantage of this option is that everyone's graphs should look the same and it is easier to check to make sure they are doing it correctly and they can work together.

Option 2 - Have students each generate their own random sample using the steps listed below.

Sample each of the populations by using the random integer function on the TI-83 or TI-84 calculator. You need a sample of 30 students. One easy way to make sure you don’t have repeats is to do the following:

* Press MATH Key
* Go to PRB menu
* Select 5:randInt(
* Enter 5 randInt(1,328,60) – this should ensure you have 30 values without repeats and unless you get numbers above 324, you can use these numbers for both male and female samples.
* Press the STO>
* And then put Ans> into L1
* Press STAT
* Use EDIT menu
* Select 1:Edit… - These are your random numbers write the first 30 numbers down, be sure to skip repeats. These are he numbers of the BMI’s you are going to use in your sample. If you have a number higher than 324 you will use that for the females and will need another number for the males.
* Create a two column table and record the BMIs for you female random sample in one column and the male random sample in the other column.

**Step 2** – Creating a Frequency Table

* Create a frequency table of 7 classes that fits all your BMI data for your gender (you will use the other gender data later). To get an appropriate range for each of your classes, take the range (maximum – minimum) of your data and divide by seven, round up to determine the class width for each grouping/class. Start with your lowest BMI and go up that width to start the next class.
* Example: Minimum BMI = 13.9 and Maximum BMI = 34.4, so range would be 34.4-13.9 = 20.5 which divided by 7 rounds up to 3 so the first class would be 13.9-16.8 because the next class would start at 13.9 + 3 = 16.9.
* Once you have your classes set up, record how many of that class you have in your sample of 30. See example if you are not sure of how this might look.

**Step 3** – Creating a Cumulative Relative Frequency Table

* Create a cumulative relative frequency table for the BMI data using your table from step two.
* Using the example from Step Two, the first class would be “Less than 16.9” and would include the count from 13.9 to 16.8 divided by the total sample (30). You may show this as proportion or percentage. The next class would be “Less than 19.9” and would include the combine counts from 13.9 to 19.8. divided by the total sample (30). Your last class should either be1.0 if you are using proportions or 100% if you are using percentages.

**Step 4** – Creating a Histogram.

* Use the information from the table in **Step 2** to create a histogram of the distribution of BMI from your sample.
* BMI will be your x-axis. You can use the zig-zag to truncate this axis so you don’t have a lot of blank space between zero and your lowest value.
* Frequency will be your y-axis and should start at zero.
* Be sure to label both axes.

**Step 5** – Creating a Cumulative Relative Frequency Graph

* Use the information from the table in **Step 3** to create a cumulative frequency plot (also called an ogive).
* BMI will be on the x-axis similar to the histogram.
* The y-axis will be cumulative relative frequency and should go from zero to 100% or 1.0 depending on whether you are using percent of proportion. Just be consistent.
* Be sure to label both axes.

**Step 6** – Why not a scatterplot?

* Explain why this data is not suitable for creating a scatterplot.
* Try to come up with one other continuous variable that could be used with BMI to generate a scatterplot

**Step 7** – Creating a Back-to-Back Stem-and-Leaf Plot

* Construct a back-to-back stem-and-leaf plot for the BMI by gender.
* Use stems that are 1 BMI unit each – with stems being the 10's and 1's units and the leaves being the 0.1 units.
* Be sure to label the top of your plot and provide a key. Ex. |13|9 = 13.9 BMI.
* Be sure to keep the spacing the same for each leaf.

**Step 8** – Creating a Categorical Bar Graph.

* Use the information from your stem and leaf plot to help you create a side-by-side bar graph that shows frequency of underweight, healthy weight, overweight and obese for both genders.
* The ranges for each gender are as follows:

Females: Underweight < 15.77 Males: Underweight <15.94

Healthy weight 15.77-23.29 Healthy weight 15.94-22.6

Overweight 22.30-27.18 Overweight 22.61-25.98

Obese >27.18 Obese >25.98

* Be sure to provide a key for the different shadings you use for the two genders.
* Be sure to label axes – x-axis should be “Weight Class” and y-axis should be “Number of 13 Year Olds”

**Step 9** -Summary Statistics

* Create a table for both genders that shows mean, standard deviation, and the 5-number summary (min, Q1, median, Q3 and max). These values can be generated with your calculator.
* Enter the female BMIs into L1 and the male BMIs into L2.
* Then use STAT with the CALC menu and run 1:1-Var Stats on L1 to get the values for females and on L2 to get the values for males.
* You can make the table horizontal or vertical.

**Step 10** – Generating Box-and-Whisker Plots

* Use the 5 number summaries for both genders from **Step 9** to create side-by-side vertical box-and-whisker plots.
* Be sure to label the y-axis and you can again truncate this axis with the zig between zero and your lowest number.
* Below the graph write a brief comparison of the distributions of your two samples.

**Step 11** – Assessing the data collection.

* Discuss whether you think the graphs in **Step 10** are likely to reflect the distribution of BMI for all the 13 year olds that attend all the schools where the Let’s Get Healthy Fairs were conducted.
* NOTE: Because these are one-day fairs and participation at any given station is self-slected, the data may not be perfectly representative of the population 13-year-olds in the four areas. Students with high absentee rates would be under-represented and students choosing to avoid the body composition station where BMI is calculated may also add statistical bias to the population data making certain portions of the distribution less well represented.

**Assessment/Reflection:**

There are a variety of ways that this assignment can be graded. Students can turn in assignment to be graded for completion and thoroughness or it can be assessed for credit as a class by going through a checklist. Questions assessing understanding of this material is typically part of unit tests or AP tests given in most courses.

**Resources:**

School adopted AP Statistics Textbook and associated sections on Descriptive Statistics.

**Standards:**

This lesson meets part of the Descriptive Statistics syllabus portion of the Advanced Placement Statistics course.

**Exercise 3:Introduction to Confidence Intervals using LGH Data**

**Overview of Lesson Plan:**

This lesson is designed to introduce students to the concept of confidence intervals for sample means and the reason for the sample size requirement when data is not normally distributed. The exercise uses data from the “Let’s Get Healthy” health fairs conducted in 2010-2012 school years from middle schools in Bend, St. Helens, West Linn – Wilsonville and Union County. Data on Body Mass Index (BMI) was collected from 328females and 324 males at these fairs. Information regarding adolescent BMI is valuable because it is a strong indicator of both short-term and long-term health concerns. This data is organized so that students can examine the distribution of BMI for each gender for these two “populations” which are both positively skewed. The exercise is designed so that individual students take small (n=10) and large (n=30) random samples from either the female or male population, generate confidence intervals for population means from the samples, provide a class plot of the confidence intervals compared to the known population mean and then compare the results of the two sample sizes in relation to the generated intervals containing the true population mean for these skewed populations.

**Big Understanding:**

The reason we check assumptions before we proceed with inferential statistics such as confidence intervals and hypothesis tests is that these assumptions are necessary for the validity of the statistics we generate.

**Essential Questions:**

Why is checking assumptions important for confidence intervals and hypothesis tests?

What are the assumptions for a one-sample *t* confidence interval?

What happens when the assumptions are violated – specifically, what happens when you use small samples to generate confidence intervals from a population that is not normally distributed?

**Grades:**

10th – 12th Grade

**Subjects:**

Advanced Placement Statistics or other Statistics Class

**Time Required:**

55 min – If students only generated one n=10 sample and confidence interval and one n=30 sample and confidence interval, then students and instructor combine intervals onto one graph followed by discussion.

**Objectives**

Upon completion of this exercise students will be able to:

* Generate random numbers for sample collection using a TI-83 or TI-84 calculator.
* Construct a confidence interval for a population mean using sample data.
* Interpret a confidence interval and interpret the level of confidence in context.
* Discuss the reasoning for the sample size requirement for confidence intervals when the characteristics of the population are unknown or are known to be skewed.
* Describe why adolescent BMI is a variable of interest for individuals and organizations working with public health issues.

**Materials:**

* “Let’s Get Healthy” Statistics lesson plan.
* TI-83, 84, or Inspire calculator.
* Figure 1. BMI distribution graph.
* Excel File – “LGH Statistics Exercise Data”
* Either computers with Excel to access Excel files or printouts of the Female and Male BMI pages included with this document

**Preparation:**

Teacher should have data files available for students either electronically or hard copy. Teacher can either provide data sheet included with the lesson or have students record data on their own paper or mark their samples directly on hard copies of data sheets.

**Activities/Procedures:**

**Step 1:** Introducing the population of interest. Show or pass out Figure 1. and have students describe the distribution of the BMI values for the graph that matches their gender. Students should at a minimum discuss shape, center and spread.

It may also be valuable to discuss that the data collected was collected from 9 different Oregon middle schools from eight different school districts which represent several different socioeconomic and cultural areas of the state. It is by no means a census of the students from these schools as some students were absent, did not have permission to participate, and many who did participate in the fair did not choose to participate in the Body Composition station. The instructor may wish to discuss whether there may be a statistical bias for or against a particular group of students because of this non-participation.

The instructor should emphasize that this does not represent an actual census of Oregon middle school students, but that the very large sample can be used to simulate a known population that is large enough to use for a sampling exercise.

**Step 2:** Introduce students to the random number generating function of their calculators and have them generate a sample of ten random numbers from 1-328 for females or 1-324 for males. Remind them not to use repeated numbers as we don’t want the same individual counted more than once.

On TI-83 or 84, the random number function is: MATH – PROB – RandInt(starting number, ending number, desired sample)

Note: If your students have never used this function on the calculator before, you may wish to have them generate a different list so they are not all starting in the same place. It is rather entertaining to have these students do RandInt(1,437,12) and have them all come up with the same 12 numbers! I have them enter the month of their birthday, year of their birth, and last 2 digits of their phone number to get them in different places.

**Step 3:** Have students locate the BMI values for the ten randomly selected individuals and then enter those values into L1 on their calculator.

**Step 4:** Introduce or review the assumptions necessary for generating confidence intervals from population means (t-interval).

* + Is the sample a simple random sample? In this case, YES.
  + Are the samples independent? YES, just be sure students don’t repeat numbers.
  + Is the population large relative to the sample? Is N>10n? YES, both females (N=328) and males (N=324) are greater than 10(n) = 10\*10 = 100.
  + Is the sample large (n>=30) or the population normally distributed? In what we are about to do NO!!. The populations are not normally distributed and we are about to use a sampleof 10 – we are going to do this so we can demonstrate why you SHOULD NOT proceed from this point.

**Step 5:** Have students generate a 95% confidence t-interval using the data they have entered in their L1.

**Step 6:** Have students then share their interval for their gender with the class. You can do this with a list and or with a graphic. The recommended graphic would have one graph for males and one for females. Student names are on the x-axis, Mean BMI is on the y-axis with a range of 12-42. For females a horizontal line should be drawn at the population mean BMI (20.25) and a similar horizontal line for the male graph at their population mean BMI (20.26). For visual demonstration purposes, show each students generated interval by plotting their lower and upper bounds and connecting them.

This should give you a good visual of the distribution of intervals and how many contain the true population values. If you have small classes, or a small gender representation, you may wish to only use one gender for the exercise or have students each generate multiple intervals.

Use the graphics to demonstrate what happens when small samples are taken from non-normal distributions. Results are likely to vary from class to class.

**Repeat Steps 2-6** with a sample that does meet the remaining assumption by using n=30.

This should generate fewer intervals that do not contain the true population parameter (as well as smaller intervals).

**Culmination:** Introduce or review interpreting confidence intervals and confidence level and the important difference in how those interpretations differ.

**Assessment/Reflection:**

Assessment for understanding of this material should come in the form of an AP-Type Free Response question using graphs and statistics and requiring students to check assumptions, generate a confidence interval and interpret both the interval and the level of confidence.

This exercise can also be modified to introduce or reinforce understanding of the Central Limit Theorem by generating more sample means and comparing the mean of those means to the population mean as well as the standard deviation of the sample means compared to the population standard deviation divided by the square root of the sample size.

Discussion of what the BMI values indicate could also be valuable to students and for interpreting the population graphs. Because BMI varies with age and gender the data was limited only to 13 year olds and was separated into male and female data sets. Conversions from BMI to weight categories is as follows:

Females: Underweight <15.77 Males: Underweight <15.94

Healthy weight 15.77-23.29 Healthy weight 15.94-22.6

Overweight 23.30-27.18 Overweight 22.61-25.98

Obese >27.18 Obese >25.98

This could lead to a discussion of whether this data suggests public health officials should be concerned about the weight and associated health of adolescents in Oregon.

**Resources:**

School adopted AP Statistics Textbook and associated sections on confidence intervals and central limit theorem.

**Standards:**

This lesson meets part of the Inferential Statistics syllabus portion of the Advanced Placement Statistics course.

Exercise 4: Statistical Hypothesis Testing with LGH data

**Overview of Lesson Plan:**

This exercise uses the Let’s Get Healthy BMI data and weight categories to run statistical hypothesis tests against regional and national data. It is best used as a cumulative review so that students are already familiar with the various hypothesis tests and are required to discern which test is appropriate for each of three different situations.

**Big Understanding:**

The main focus of this exercise is to illustrate which hypothesis tests are appropriate for which types of questions and which types of data. There is also an emphasis on interpreting p-values in relation to these tests.

**Essential Questions:**

When is a one proportion z-test appropriate?

When is a one sample t-test for means appropriate?

When is a Chi-Square Test for Goodness-of-Fit appropriate?

What does the p-value tell us regardless of which test is being used?

**Grades:**

10th – 12th Grade

**Subjects:**

Advanced Placement Statistics or other Statistics Class.

**Time Required:**

One to two 55-minute class periods depending on whether this exercise is used as a lead-in, follow-up or cumulative review.

**Objectives:**

Upon completion of this exercise, students will be able to:

* Identify which test is appropriate for which type of situation.
* Use their calculator to generate a test statistic for three different hypothesis tests.
* Interpret the p-value in relation to each of their three hypothesis tests.

**Materials:**

* + “Let’s Get Healthy” Statistics lesson plan.
  + TI-83, 84, or Inspire calculator.
  + Figure 1. BMI distribution graph.
  + Excel File – “LGH Statistics Exercise Data”
  + Either computers with Excel to access Excel files or printouts of the Female and Male BMI pages included with this document

**Preparation:**

Make copies of the 13-year old Female and Male BMI tables for each student and the student assignment sheet.

**Activities/Procedures:**

Step 1: Optional – Prior to assignment instructor may wish to review the types of hypothesis tests and how to identify for which type of situation each test is appropriate.

Step 2: Pass out assignment sheet and data sheets to students.

Step 3: There are several options or variations for this step depending on instructor preference and time allotted for this exercise. Students may be required to show all work as if the problem was an open-ended problem or calculate test statistics and p-values using their calculator along the lines of the time available for multiple-choice questions on the AP Test. You may also wish to have students work on their own or in small groups. One recommendation is to give them part of their time to work on their own and then group them compare to compare and discuss answers.

Step 4: Optional – Discuss the application of these results. How might public health officials use this data to inform and educate young people and community leaders about the status of 13-year old weight conditions in these areas of Oregon.

**Assessment/Reflection:**

Up to instructor. Assignment can be turned in, checked off or scored by student during discussion and review. Similar concepts are typically covered in multiple choice or open-ended form on most unit or cumulative tests dealing with this area of statistics.

**Resources:**

School adopted AP Statistics Textbook and associated sections on hypothesis tests – specifically one proportion z-test, one-sample t-test for means and Chi-Square test for Goodness-of-Fit.

**Standards:**

This lesson meets part of the Inferential Statistics syllabus portion of the Advanced Placement Statistics course.



Figure 1. Tinkerplot of BMI data from four combined Let’s Get Healthy Fairs in Oregon for 2010-2012. Distributions for both females (N=328) and males (N=324) are right skewed. Bar on x-axis represents medians (median BMI = 19.5 for females and 19.4 for males) and triangles represent means (mean BMI = 20.25 for females and20.26 for males).

Table 1. Female BMI data. This data represents all 13 year oldfemale students who participated in the Body Composition station at the middle school Let’s Get Healthy Fairs in 2010-2012 (N = 328). Data has been randomly sorted.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fem # | BMI | Fem # | BMI | Fem # | BMI | Fem # | BMI | Fem # | BMI | Fem # | BMI | Fem # | BMI | Fem # | BMI |
| 1 | 19.2 | 42 | 15.4 | 83 | 22.9 | 124 | 15.3 | 165 | 19.5 | 206 | 19.3 | 247 | 20.5 | 288 | 34.1 |
| 2 | 17.6 | 43 | 21.9 | 84 | 25.4 | 125 | 20.3 | 166 | 20.2 | 207 | 19.2 | 248 | 20.7 | 289 | 16.2 |
| 3 | 18.5 | 44 | 21.4 | 85 | 24.6 | 126 | 16.1 | 167 | 18.2 | 208 | 16.7 | 249 | 18.5 | 290 | 19.6 |
| 4 | 25.1 | 45 | 18.3 | 86 | 22.9 | 127 | 19.3 | 168 | 18.9 | 209 | 16.5 | 250 | 23.3 | 291 | 36.4 |
| 5 | 22.2 | 46 | 19.2 | 87 | 19 | 128 | 23.6 | 169 | 23.4 | 210 | 21.7 | 251 | 19.9 | 292 | 18.7 |
| 6 | 25.4 | 47 | 22.9 | 88 | 21.8 | 129 | 20.2 | 170 | 19.9 | 211 | 20.7 | 252 | 19.8 | 293 | 19.6 |
| 7 | 23.6 | 48 | 19.5 | 89 | 18.2 | 130 | 17.5 | 171 | 18.2 | 212 | 20.2 | 253 | 18.6 | 294 | 35.7 |
| 8 | 18 | 49 | 17.3 | 90 | 16.2 | 131 | 18.3 | 172 | 20.3 | 213 | 21.8 | 254 | 30 | 295 | 22.4 |
| 9 | 18.6 | 50 | 21.4 | 91 | 19.6 | 132 | 17.2 | 173 | 16.4 | 214 | 21.4 | 255 | 19.9 | 296 | 16.1 |
| 10 | 19.7 | 51 | 21.2 | 92 | 19.3 | 133 | 17.6 | 174 | 18.2 | 215 | 17.3 | 256 | 16.2 | 297 | 19 |
| 11 | 23.8 | 52 | 17 | 93 | 19 | 134 | 19.1 | 175 | 21.3 | 216 | 17.5 | 257 | 23.7 | 298 | 23.9 |
| 12 | 16.4 | 53 | 16.3 | 94 | 16.6 | 135 | 18.5 | 176 | 17.6 | 217 | 26.2 | 258 | 19.1 | 299 | 25.3 |
| 13 | 19.7 | 54 | 21.2 | 95 | 15 | 136 | 20.2 | 177 | 18.6 | 218 | 14.7 | 259 | 17.2 | 300 | 17.9 |
| 14 | 17.9 | 55 | 16.4 | 96 | 17.5 | 137 | 15.5 | 178 | 14.3 | 219 | 24 | 260 | 23.3 | 301 | 17.9 |
| 15 | 17.6 | 56 | 21.8 | 97 | 16.9 | 138 | 24.1 | 179 | 19.9 | 220 | 18.9 | 261 | 18.3 | 302 | 16.5 |
| 16 | 13.7 | 57 | 17 | 98 | 18.8 | 139 | 24.7 | 180 | 22.6 | 221 | 17.2 | 262 | 19 | 303 | 19.2 |
| 17 | 23.4 | 58 | 16.5 | 99 | 21.3 | 140 | 19.3 | 181 | 26.4 | 222 | 31.2 | 263 | 17.5 | 304 | 17.7 |
| 18 | 22.5 | 59 | 29.8 | 100 | 20.7 | 141 | 19.8 | 182 | 20.2 | 223 | 16.3 | 264 | 20.6 | 305 | 17.1 |
| 19 | 17.7 | 60 | 15.2 | 101 | 24 | 142 | 26.6 | 183 | 18.5 | 224 | 19.7 | 265 | 20.2 | 306 | 22 |
| 20 | 21.4 | 61 | 25.1 | 102 | 16.6 | 143 | 20.4 | 184 | 17.6 | 225 | 24.8 | 266 | 17.5 | 307 | 25.1 |
| 21 | 23 | 62 | 16.4 | 103 | 23.8 | 144 | 18.7 | 185 | 27.2 | 226 | 18 | 267 | 24 | 308 | 25.5 |
| 22 | 19 | 63 | 22.6 | 104 | 21.4 | 145 | 20.5 | 186 | 18.6 | 227 | 30.5 | 268 | 18.7 | 309 | 18.8 |
| 23 | 19.5 | 64 | 21.5 | 105 | 21.5 | 146 | 20.2 | 187 | 15.8 | 228 | 20.3 | 269 | 18.5 | 310 | 23.2 |
| 24 | 20.9 | 65 | 19.5 | 106 | 19.5 | 147 | 17.7 | 188 | 16.9 | 229 | 26.7 | 270 | 18.5 | 311 | 21.8 |
| 25 | 19.7 | 66 | 29.9 | 107 | 23.8 | 148 | 17.3 | 189 | 18.4 | 230 | 19.2 | 271 | 18.1 | 312 | 28.6 |
| 26 | 27.9 | 67 | 20.3 | 108 | 19.3 | 149 | 19.6 | 190 | 27.4 | 231 | 16.1 | 272 | 29.7 | 313 | 17.2 |
| 27 | 19 | 68 | 20.9 | 109 | 19.1 | 150 | 20.7 | 191 | 14.9 | 232 | 16.5 | 273 | 18.7 | 314 | 21.8 |
| 28 | 20.2 | 69 | 21.5 | 110 | 19.3 | 151 | 16.5 | 192 | 26.3 | 233 | 23.6 | 274 | 16 | 315 | 21.9 |
| 29 | 18.1 | 70 | 21.5 | 111 | 16.3 | 152 | 19.6 | 193 | 19.1 | 234 | 21.8 | 275 | 29.6 | 316 | 25.6 |
| 30 | 20.5 | 71 | 20.9 | 112 | 17.4 | 153 | 23.7 | 194 | 22.3 | 235 | 14.2 | 276 | 16.1 | 317 | 16.1 |
| 31 | 24.4 | 72 | 15.6 | 113 | 21 | 154 | 24 | 195 | 20.5 | 236 | 19.4 | 277 | 17.1 | 318 | 20.5 |
| 32 | 23.2 | 73 | 21.3 | 114 | 19.5 | 155 | 21 | 196 | 18.9 | 237 | 16.1 | 278 | 17.4 | 319 | 22.1 |
| 33 | 19.3 | 74 | 18.5 | 115 | 18.9 | 156 | 21.5 | 197 | 16.3 | 238 | 20.8 | 279 | 22.8 | 320 | 27.1 |
| 34 | 18.3 | 75 | 17.4 | 116 | 15.8 | 157 | 16.2 | 198 | 33.4 | 239 | 16.6 | 280 | 18.9 | 321 | 21.7 |
| 35 | 18.2 | 76 | 20.1 | 117 | 18.3 | 158 | 20.2 | 199 | 14.4 | 240 | 21.6 | 281 | 19.7 | 322 | 17.8 |
| 36 | 18.4 | 77 | 19.5 | 118 | 19.6 | 159 | 17.6 | 200 | 29.8 | 241 | 18 | 282 | 24.1 | 323 | 17 |
| 37 | 18.7 | 78 | 26.5 | 119 | 20.4 | 160 | 16.7 | 201 | 17.3 | 242 | 15.7 | 283 | 28.5 | 324 | 19.1 |
| 38 | 21.5 | 79 | 19.2 | 120 | 20.7 | 161 | 21.3 | 202 | 16.7 | 243 | 17.8 | 284 | 20.4 | 325 | 23 |
| 39 | 21 | 80 | 20.8 | 121 | 16.8 | 162 | 20.3 | 203 | 34.8 | 244 | 16.9 | 285 | 16.6 | 326 | 19 |
| 40 | 18 | 81 | 21.5 | 122 | 23.3 | 163 | 16.4 | 204 | 15.9 | 245 | 18.1 | 286 | 15.2 | 327 | 19.2 |
| 41 | 22.2 | 82 | 19.8 | 123 | 19.3 | 164 | 18.9 | 205 | 16.4 | 246 | 17.3 | 287 | 22.1 | 328 | 20.2 |

Table 2. Male BMI data. This data represents all 13 year oldmale students who participated in the Body Composition station at the middle school Let’s Get Healthy Fairs in 2010-2012 (N = 324). Data has been randomly sorted.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Male # | BMI | Male # | BMI | Male # | BMI | Male # | BMI | Male # | BMI | Male # | BMI | Male # | BMI | Male # | BMI |
| 1 | 16.8 | 41 | 24.5 | 81 | 23.6 | 121 | 19.6 | 161 | 20.3 | 201 | 23.9 | 241 | 19.3 | 281 | 26.4 |
| 2 | 18.3 | 42 | 23.2 | 82 | 20.1 | 122 | 14.9 | 162 | 21.7 | 202 | 21 | 242 | 17.6 | 282 | 21.3 |
| 3 | 24.6 | 43 | 21.3 | 83 | 18.3 | 123 | 21.5 | 163 | 34.6 | 203 | 15.4 | 243 | 17 | 283 | 19.5 |
| 4 | 20.5 | 44 | 30.7 | 84 | 20 | 124 | 21.9 | 164 | 16.4 | 204 | 16.3 | 244 | 14.7 | 284 | 21.6 |
| 5 | 18.2 | 45 | 18.6 | 85 | 16.4 | 125 | 22.5 | 165 | 19.5 | 205 | 18.4 | 245 | 19.2 | 285 | 16.8 |
| 6 | 21.2 | 46 | 21.7 | 86 | 23.5 | 126 | 19.2 | 166 | 17 | 206 | 25.5 | 246 | 19.9 | 286 | 16.8 |
| 7 | 32.4 | 47 | 19.4 | 87 | 25.4 | 127 | 21.4 | 167 | 21.2 | 207 | 17.6 | 247 | 24.4 | 287 | 20.1 |
| 8 | 26.2 | 48 | 17.5 | 88 | 21.5 | 128 | 22.8 | 168 | 14 | 208 | 17 | 248 | 27.3 | 288 | 20.3 |
| 9 | 15.1 | 49 | 23.2 | 89 | 16.7 | 129 | 20.4 | 169 | 15.5 | 209 | 23.3 | 249 | 21.7 | 289 | 18.2 |
| 10 | 17.3 | 50 | 18.6 | 90 | 17.2 | 130 | 18.7 | 170 | 16.6 | 210 | 24.7 | 250 | 18 | 290 | 17.1 |
| 11 | 26.9 | 51 | 23.7 | 91 | 18 | 131 | 28.7 | 171 | 21.1 | 211 | 26.6 | 251 | 22.4 | 291 | 17.4 |
| 12 | 20.6 | 52 | 17.9 | 92 | 19.8 | 132 | 23.8 | 172 | 25 | 212 | 19.9 | 252 | 19 | 292 | 19.8 |
| 13 | 21.5 | 53 | 13.9 | 93 | 24.9 | 133 | 23 | 173 | 17.7 | 213 | 27.8 | 253 | 16.7 | 293 | 14.5 |
| 14 | 18.1 | 54 | 21.4 | 94 | 21.8 | 134 | 17.8 | 174 | 19.9 | 214 | 18.8 | 254 | 22.1 | 294 | 17.3 |
| 15 | 19.6 | 55 | 23.5 | 95 | 17.5 | 135 | 23.8 | 175 | 19.6 | 215 | 19.2 | 255 | 17.1 | 295 | 21.3 |
| 16 | 18 | 56 | 18.5 | 96 | 22.9 | 136 | 15.3 | 176 | 32.1 | 216 | 18.7 | 256 | 16.7 | 296 | 16.9 |
| 17 | 17.4 | 57 | 21.6 | 97 | 18.2 | 137 | 15.7 | 177 | 24.2 | 217 | 20.3 | 257 | 16.2 | 297 | 19.4 |
| 18 | 16.2 | 58 | 24.1 | 98 | 16.8 | 138 | 22.9 | 178 | 15.7 | 218 | 17.5 | 258 | 17.7 | 298 | 22 |
| 19 | 15.8 | 59 | 20.9 | 99 | 21.2 | 139 | 16.7 | 179 | 16.5 | 219 | 16.6 | 259 | 17.1 | 299 | 20.1 |
| 20 | 18.9 | 60 | 21.9 | 100 | 18.3 | 140 | 19.8 | 180 | 17.9 | 220 | 27.3 | 260 | 17.5 | 300 | 18.8 |
| 21 | 19.5 | 61 | 17.5 | 101 | 17.1 | 141 | 23.1 | 181 | 17.7 | 221 | 18.1 | 261 | 19 | 301 | 19.4 |
| 22 | 18.6 | 62 | 21.5 | 102 | 21.9 | 142 | 20.7 | 182 | 16.8 | 222 | 18 | 262 | 19.9 | 302 | 17.7 |
| 23 | 19.9 | 63 | 26.6 | 103 | 18 | 143 | 21 | 183 | 18.6 | 223 | 22.4 | 263 | 18.6 | 303 | 31.7 |
| 24 | 24.6 | 64 | 19.2 | 104 | 19.7 | 144 | 16.3 | 184 | 25.1 | 224 | 20.7 | 264 | 17.6 | 304 | 18 |
| 25 | 19.4 | 65 | 16.2 | 105 | 16.2 | 145 | 17.4 | 185 | 24.4 | 225 | 18.3 | 265 | 22 | 305 | 23.6 |
| 26 | 21.5 | 66 | 19.8 | 106 | 21 | 146 | 17.9 | 186 | 18.7 | 226 | 17.8 | 266 | 22.6 | 306 | 18.5 |
| 27 | 19.9 | 67 | 19.1 | 107 | 24.2 | 147 | 19.4 | 187 | 25.8 | 227 | 16 | 267 | 22.8 | 307 | 17.8 |
| 28 | 23.6 | 68 | 19.3 | 108 | 24.8 | 148 | 17 | 188 | 17.4 | 228 | 19.1 | 268 | 31.6 | 308 | 15.1 |
| 29 | 21.5 | 69 | 15.8 | 109 | 27.9 | 149 | 25.1 | 189 | 20.9 | 229 | 20.6 | 269 | 17 | 309 | 14.7 |
| 30 | 19.8 | 70 | 18.7 | 110 | 18.6 | 150 | 26 | 190 | 18.4 | 230 | 14.8 | 270 | 27.2 | 310 | 18.2 |
| 31 | 20.1 | 71 | 19.3 | 111 | 19.7 | 151 | 18 | 191 | 30.6 | 231 | 18.3 | 271 | 17.9 | 311 | 19.2 |
| 32 | 19 | 72 | 16.9 | 112 | 18.9 | 152 | 20.6 | 192 | 17.4 | 232 | 16.7 | 272 | 17.2 | 312 | 17.2 |
| 33 | 19.8 | 73 | 16.9 | 113 | 20.3 | 153 | 15.7 | 193 | 22.3 | 233 | 16.1 | 273 | 18.3 | 313 | 20.7 |
| 34 | 20.4 | 74 | 18.1 | 114 | 22 | 154 | 30 | 194 | 17.5 | 234 | 20.8 | 274 | 19.9 | 314 | 17.5 |
| 35 | 33.5 | 75 | 20.1 | 115 | 22.2 | 155 | 18.7 | 195 | 17.2 | 235 | 31.4 | 275 | 17.4 | 315 | 17.8 |
| 36 | 17.8 | 76 | 13 | 116 | 15.1 | 156 | 21 | 196 | 16.1 | 236 | 21.7 | 276 | 16.2 | 316 | 19.6 |
| 37 | 26.5 | 77 | 18.5 | 117 | 18.3 | 157 | 27.6 | 197 | 17.3 | 237 | 22 | 277 | 26.5 | 317 | 19 |
| 38 | 30 | 78 | 18.9 | 118 | 18.1 | 158 | 25.8 | 198 | 18 | 238 | 32.1 | 278 | 16.1 | 318 | 18.4 |
| 39 | 19 | 79 | 19.3 | 119 | 20.1 | 159 | 19 | 199 | 23 | 239 | 21.5 | 279 | 20.3 | 319 | 19.5 |
| 40 | 17.9 | 80 | 17.4 | 120 | 22.6 | 160 | 24.3 | 200 | 19.2 | 240 | 20.7 | 280 | 22.6 | 320 | 21.1 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 321 | 17 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 322 | 37.2 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 323 | 16.3 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 324 | 22.5 |

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Class/Period \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Exercise 1: Observational Studies and Experiments

***About Let’s Get Healthy!***

Let’s Get Healthy! is an education and research exhibit from Oregon Health & Science University (OHSU) that allows participants to learn about their health while contributing to science. The exhibit is funded by the National Institutes of Health (NIH) and has been held at schools and communities throughout Oregon since 2007.

***About the Health Stations***

As part of Let’s Get Healthy!, participants can visit different stations to learn about their health. There are six data feedback health stations and four hands-on exhibits. The data feedback stations help participants learn about their diet, body composition including BMI, blood pressure, sun exposure, blood pressure and sleep patterns. Participants receive immediate results and recommendations from each station.

Questions:

1. Is Let’s Get Healthy an observational study or experiment?

2. How can you justify your answer to Question 1?

3. If the researchers found a strong apparent relationship between one of the variables from the sleep pattern station and one of the variables from the body composition station could they say that one causes the other? Why or Why Not?

4. Can you suggest an experiment to test the apparent relationship between a sleep pattern such as sleep duration and body composition such as BMI? Be sure to keep in mind the four key components of a well-designed experiment and ethical concerns of working with human subjects:

* 1. Identify and control for as many other variables as possible that are not your variables of interest.
  2. Decide how you will randomize your treatment and control groups.
  3. Determine what would be a suitable sample size for each group.
  4. Consider whether blocking is appropriate for your test subjects.
  5. Determine whether blinding or double-blinding are appropriate for this experiment.

TEACHER KEY AND COMMENTS

Exercise 1: Observational Studies and Experiments

***About Let’s Get Healthy!***

Let’s Get Healthy! is an education and research exhibit from Oregon Health & Science University (OHSU) that allows participants to learn about their health while contributing to science. The exhibit is funded by the National Institutes of Health (NIH) and has been held at schools and communities throughout Oregon since 2007.

***About the Health Stations***

As part of Let’s Get Healthy!, participants can visit different stations to learn about their health. There are six data feedback health stations and four hands-on exhibits. The data feedback stations help participants learn about their diet, body composition including BMI, blood pressure, sun exposure, blood pressure and sleep patterns. Participants receive immediate results and recommendations from each station.

Questions:

1. Is Let’s Get Healthy an observational study or experiment?

THIS IS AN OBSERVATIONAL STUDY NOT AN EXPERIMENT

2. How can you justify your answer to Question 1?

THIS IS AN OBSERVATIONAL STUDY BECAUSE DATA IS BEING COLLECTED THAT ALREADY EXISTS AND NO TREATMENTS ARE BEING APPLIED TO ANY OF THE SUBJECTS.

3. If the researchers found a strong apparent relationship between one of the variables from the sleep pattern station and one of the variables from the body composition station could they say that one causes the other? Why or Why Not?

NO, THEY COULD NOT. BECAUSE THIS IS AN OBSERVATIONAL STUDY AND THERE IS NO CONTROL OR TREATMENT, ANY APPARENT RELATIONSHIP CAN NOT BE IDENTIFIED AS CAUSE AND EFFECT – IT CAN BE SHOWN AS CORRLEATION BUT NOT CAUSATION. IT CAN HOWEVER SUGGEST EXPERIMENTS TO TEST THAT RELATIONSHIP.

4. Can you suggest an experiment to test the apparent relationship between a sleep pattern such as sleep duration and body composition such as BMI? Be sure to keep in mind the four key components of a well-designed experiment and ethical concerns of working with human subjects:

1. Identify and control for as many other variables as possible that are not your variables of interest.
2. Decide how you will randomize your treatment and control groups.
3. Determine what would be a suitable sample size for each group.
4. Consider whether blocking is appropriate for your test subjects.
5. Determine whether blinding or double-blinding are appropriate for this experiment.

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Class/Period \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Exercise 4: Statistical Hypothesis Testing with LGH data

**Part I.** Match the three scenarios with the three hypothesis tests that would be useful to test the questions that arise in each of those scenarios.

* Scenario 1 – A teen health advocacy group wants to know whether the average BMI for the Let’s Get Healthy (LGH) data differs from claimed or reported state or national averages for similar groups of adolescents.
* Scenario 2 – A public health official is curious whether the number of students in the various weight categories (underweight, healthy weight, overweight, obese) is different that what is reported nationally for those weight categories.
* Scenario 3 –A researcher trying to decide whether to add LGH health fairs to other areas of Oregon is interested in whether the proportion of 13-olds over the recommended upper limit for healthy weight from the current LGH health fairs is different than what public health data suggests is the case for the whole state of Oregon.
* Hypothesis Test 1 – One Proportion Z-test
* Hypothesis Test 2 – One Sample t-test for Means
* Hypothesis Test 3 – Chi-Square Goodness-of-Fit Test

**Part II.** One-Proportion Z-Test

Use data from all 652 students to compare proportion of students in the overweight and obese weight categories combined with the 24.3% value for this group reported by StateHealthFacts.org from the Henry J. Kaiser Family Foundation for Oregon in 2007.

**Part III.** One sample t-test for means

Collect a random sample of 30 students from either the female or male data and test whether this sample is lower than the appropriate national average BMI for adult females (26.5) or males (26.6) .

**Part IV.** Chi-square Test for Goodness-of-Fit

Use the data from the combined fairs to see whether this Oregon data has a different distribution than the national cut-offs (from cdc.gov) for the following weight categories: Underweight (less than or equal to 5th percentile), Healthy Weight (6th-84th percentile) and Overweight and Obese comined (greater than or equal to 85th percentile).