SeaGlide_TEAM	Detecting Bias	Lesson # 7 (60 minutes)

Description:

You work as a statistical analyst and fact checker for a magazine that publishes information form research studies. Your boss has asked you to check the validity of some of the studies that the magazine wants to publish by checking for bias. First, you need to familiarize yourself with types of bias. Then, you will read through a given description of each research study to determine what kind of bias is in each study, to explain why you selected that type of bias, and to determine if there is a way to remove the bias.

Students will be able to:

- Define types of bias
- Detect and identify bias in studies and experiments
- Complete Chi-Squared

Students will understand:

Students will understand that there are different types of bias and that they all can affect experimental outcome. They will be able to detect bias in a sample as well as identify which type of bias is affecting the experiment. Students will also have an understanding of how to compute chi-square which is an important aspect of experimental design.

Key Definitions & Concepts [1]:

- **Undercoverage Bias:** occurs when some members of the population are inadequately represented in the sample
- **Confirmation bias:** interpreting new evidence or information as confirmation of beliefs or theories that one already has.
- **Reporting bias:** selective revealing or suppression of information by subjects in the study
- **Self-Selection Bias:** any situation in which individuals select themselves into a group
- **Design bias:** a process where the researchers performing the research influence the results, in order to portray a certain outcome
- **Interviewer bias:** leading questions that influence respondents in such a way that it distorts the outcome.

Standards [Copied from: 2]:

CC.2.4.HS.B.4 Recognize and evaluate random processes underlying statistical experiments.
CC.2.4.HS.B.5 Make inferences and justify conclusions based on sample surveys, experiments, and observational studies.

	Background Information	

Prior Knowledge:

- All previous *Measurement and Data Analysis* lessons

Math Practices [Copied from: 3]:	**Core Ideas [Copied from: 4]:**	**Cross Cutting Concepts [Copied from: 5]:**
● Construct viable arguments and critique the reasoning of others. ● Use appropriate tools strategically.	● Information Processing ● Defining and Delimiting Problems	● Cause and effect ● Patterns

Possible Preconceptions/Misconceptions:

Students may not recognize certain types of bias as problematic. They may also not realize how many types of bias exist and how they can affect an experiment. Students may also struggle at first with the new notation for Chi-Square as it is a new equation being introduced. It is important for students to to understand that bias affects design and can be misleading. Bias in experimental design can lead to the dissemination of false information.

	Lesson Plan - 5E(+) Model	

Engage [6]:

Students will complete the *Review Quiz*, that has some of the keywords from the concept map from the first lesson in the *Measurement and Data Analysis* module. This forces students to recall important information that will set them up for success in the rest of the lesson. Students can also see how previous units are important later down the road and make conceptual connections. Provide the students with 5 minutes to initially fill out their *Review Quiz*. Then, allot 5 additional minutes for the students to reference their concept maps from the first lesson to correct their own responses as necessary. Take no more than 5 minutes to correct the *Review Quiz* as a whole class. Hence, this portion of the lesson should last less than 15 minutes.

Explore [1], [7]:

Part I:Introduction

Students will work on the *Types of Bias* worksheet to gain knowledge of the meaning of types of bias that show up in experimentation. Students will match the type of bias to its description. This helps prepare students for the next activity. After students complete the worksheet, the teacher should go over the answers with the class. It is important that students have the correct answers moving onto the next activity. This should take a total of 10 minutes.

Part II: *Benchmark Lesson*: Detecting Bias

Students can work in small groups of 2 or 3 to complete the *Detecting Bias* worksheet. This activity has an example of each type of bias from the above activity. This allows students to practice finding errors in experimental design and also see examples of how bias can affect an experiment. Students must also explain why that type of bias applies and how it might be fixed. This allows students to make connections to conducting a proper experiment. This connects to the previous lessons in the unit as students are learning about what makes a good data set, how to tell if it is skewed, what makes a good visual representation, and now what makes an experiment good or bad. This helps bring a full comprehensive understanding of how to perform valid statistical measures by connecting previous concepts to this current lesson. This activity should take 15 minutes.

Part III: *Investigation Lesson*: Chi-Squared Skittles

Students can either work on the *Chi-Squared Skittles* activity individually or in small groups. If students work in groups, each student should be given a packet of skittles, and they should be instructed to combine the counts for each color from each packet within the group. Students will work through a guided worksheet to calculate Chi-Square and determine if the number of each color of skittles in a packet is random or not. The worksheet gives them step-by-step instructions on how to calculate it as well as necessary definitions. If students are given pre-packaged small packs of skittles, the result should come out that the skittle allotment per color is random. If the teacher wishes, he/she can create packs of their own to give to students and add bias in when creating the packs. The teacher, at the end of the activity, should then promote a whole class discussion about the type(s) of bias present. The teacher can also choose to combine data from each group and calculate an overall Chi-Squared value for all of the data as a whole class. These modification suggestions would be best applicable if the class time is longer than a 60 minute session. This should take 20 minutes to complete without including either of the modification suggestions.

Explain:

Throughout the lesson students must explain their reasoning, especially in the *Detecting Bias* activity. They must explain why they chose that type of bias and make connects between the definitions and the application of such within a given scenario. They must also be able to explain a way to fix the bias, causing them to demonstrate a deeper understanding. The teacher should also be asking probing questions during completion of the worksheets that prompt the students to explain what they are doing and why. This allows for students to learn through inquiry and discovery, which ensures that they gain deeper understanding of the learning goals of this lesson.

Elaborate:

Students are connected to real world examples of how bias may affect an experiment as well as how and why bias needs to be checked. Students will be exposed to a mathematical way that can help determine bias as well. The lesson is set up where students are acting as fact checker, i.e. part of their job is to check the validity of the given studies. Students can see the importance of checking for bias in a study before it is published through the completion of this lesson.

Evaluate:

During whole class discussions and while the students are completing the worksheets, the teacher is able to check for surface level understanding and make sure that the class is all on the same page by listening to students' discussions and observing students' responses. These serve as the informal evaluations within this lesson. The instructor can gauge student progress through observing whether or not they are struggling to complete the activities. Formal evaluation can be done by checking for correctness in the students' worksheets.

Enrich:

This lesson can differentiated by being utilized as an introductory lesson for an AP biology class. Chi-Squared is an important concept in AP biology and also in biology labs. The class discussions can be further developed by introducing the concept of a null hypothesis, its significance, and how it can affect an experiment and/or its outcome. Also, Chi-Squared calculations can be completed for a biology experiment by solving for randomness of genetic traits within a fruit fly population instead of color within a skittles packet.

All associated documents are attached below

Reference *Annotated Bibliography* on the very last page of this packet

Name:_____ Date:_____

Review Quiz [6]

In your own words, define each of the following keywords.

Sample-

Population-

Biased sample-

Random sampling-

Stratified sampling-

Why is a proper sample important in running experiment?

Name:_____ Date:_____

Types of Bias [1]

You work as a statistical analyst and fact checker for a magazine that publishes information form research studies. Your boss has asked you to check the validity of some of the studies that the magazine wants to publish by checking for bias. First, you need to familiarize yourself with types of bias. Match the the type of bias with the description of the bias.

Types of Bias	**Description**
Undercoverage Bias	A) any situation in which individuals select themselves into a group
Confirmation Bias	B) occurs when some members of the population are inadequately represented in the sample.
Reporting Bias	C) a process where the researchers performing the research influence the results, in order to portray a certain outcome
Self-Selection Bias	D) selective revealing or suppression of information by subjects in the study
Design Bias	E) leading questions that influence respondents in such a way that it distorts the outcome.
Interviewer Bias	F) the tendency to interpret new evidence as confirmation of one's existing beliefs or theories.

Name:_____ Date:_____

Detecting Bias [1]

Read through a given description of each research study to determine what kind of bias is in each study, to explain why you selected that type of bias and, to determine if there is a way to remove the bias.

1. Participants in a vitamin supplement study are asked questions such as “Isn’t your skin so much healthier?”

2. A university is collecting data about grades and study habits. Advertisements are put up library and other common study locations so students can sign up for the study.

3. A researcher wants to show that their supplement promotes healthy skin and all of the participants chosen for the study already have healthy skin.

4. A poll was taken in a magazine asking the readers about whom they would want to be our next president. Subscribers to the magazine are primarily wealthy.

5. A researcher holds a belief that left-handed people are more creative than right-handed people. Whenever they encounter a person that is both left-handed and creative, they place greater importance on this "evidence" that supports what they already believe.

6. Individuals are participating in a research study about heart health. Researchers have no way of proving whether or not a participant is an active smoker and some participant may have chosen not to disclose that they smoke.

Name:_____ Date:_____

Chi-Squared Skittles [7]

Chi-Squared (pronounced k-eye) is a method to determine if experimental outcomes result by chance or if something else is affecting the results, i.e. a variable we are testing is affecting results but should not be, type(s) of bias is/are present, or an experimental manipulation occurred.

The magazine you work for wants to present a report proving that the colors in a skittles package are completely by chance (random). You need to verify this, so you chose to run a chi-square test. **Do not eat any of your skittles.**

We calculate Chi-Squared by the following equation:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

O = observed value

E = Expected value

Σ = summation

First, you want to calculate the expected value. We expect to have an equal number of each color. Hence, calculate **E =** $\frac{number\ of\ colors}{total\ skittle\ in\ your\ pack}$ for each color. Fill in the table for Expected with the calculated values. Then, in the Observed row, fill in the actual number of skittles for each color in your pack. Complete this process for each of the colors.

Individual	Green	Red	Orange	Purple	Yellow
Expected					
Observed					
O-E					
$(O-E)^2$					
$(O-E)^2 / E$					

Now sum all of the values in the last row to get a **chi-squared** value.

In order to determine whether or not the results are by chance, we need to compare the chi-square value to a **critical value**. To find the critical value, we first need to determine the **degrees of freedom**.

Degrees of freedom- how much a value can vary from what it is in the experiment.

***Degrees of freedom = number of possible outcomes minus 1*.**

How many degrees of freedom are there in this experiment?

We use this value to find the critical value by the table below. Use the 0.05 column to determine the critical value.

Degrees of freedom	Probability of exceeding the critical value				
	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588

[7]

If your chi-square number is less than the critical value then your experiment is by chance.

What is our critical value?_____

Are your results by chance? Why or why not?

Where else can Chi-Square be used?

Name:_____ANSWER KEY_____ Date:_____

Review Quiz [6]

In your own words, define each of the following keywords.

Sample: a small subset of a larger set of data

Population: the larger set from which a sample may be drawn

Biased sample: a sample that is collected in such a way that some members of the intended population are less likely to be included than others.

Random sampling: requires every member of the population to have an equal chance of being selected into the sample. In addition, the selection of one member must be independent of the selection of every other member

Stratified sampling: this method can be used if the population has a number of distinct “strata” or groups. In stratified sampling, you first identify members of your sample who belong to each group. Then you randomly sample from each of those subgroups in such a way that the sizes of the subgroups in the sample are proportional to their sizes in the population.

Why is a proper sample important in running experiment?

This is important so that you know that your data is accurate and is representative of the population as a whole.

Name:_____ANSWER KEY_____ Date:_____

Types of Bias [1]

You work as a statistical analyst and fact checker for a magazine that publishes information form research studies. Your boss has asked you to check the validity of some of the studies that the magazine wants to publish by checking for bias. First, you need to familiarize yourself with types of bias. Match the the type of bias with the description of the bias.

Types of Bias	**Description**
Undercoverage Bias B	A) any situation in which individuals select themselves into a group
Confirmation bias F	B) occurs when some members of the population are inadequately represented in the sample.
Reporting bias D	C) a process where the researchers performing the research influence the results, in order to portray a certain outcome
Self-Selection Bias A	D) selective revealing or suppression of information by subjects in the study
Design bias C	E) leading questions that influence respondents in such a way that it distorts the outcome.
Interviewer bias E	F) the tendency to interpret new evidence as confirmation of one's existing beliefs or theories.

Name:_____ANSWER KEY_____ Date:_____

Detecting Bias [1]

Read through a given description of each research study to determine what kind of bias is in each study, to explain why you selected that type of bias and, to determine if there is a way to remove the bias.

1. Participants in a vitamin supplement study are asked questions such as “Isn’t your skin so much healthier?”

   This is interviewer bias because the question is leading participants to talk about improvements in their skin. A better question would be, have you noticed any changes in your skin? Or more generally have you noticed any changes in your body?

2. A university is collecting data about grades and study habits. Advertisements are put up library and other common study locations so students can sign up for the study.

   This is Self-Selection bias because most likely students who study hard ae going to sign up especially because the flyers are in study areas. This can help be rectified by hanging posters in non-academic locations on campus.

3. A researcher wants to show that their supplement promotes healthy skin and all of the participants chosen for the study already have healthy skin.

   This is design bias. The experiment is designed so that, at the end of the study, everyone has healthy skin even though it may not be because of the supplement. This can be fixed by having a mix of people with different skin types.

4. A poll was taken in a magazine asking the readers about whom they would want to be our next president. Subscribers to the magazine are primarily wealthy.

   This is undercoverage bias since the majority of the people who will answer the poll are wealthy, and it does not poll those with lower income. This can be rectified by posting the poll in other magazines with more diverse subscribers.

5. A researcher holds a belief that left-handed people are more creative than right-handed people. Whenever they encounter a person that is both left-handed and creative, they place greater importance on this "evidence" that supports what they already believe.

   This is confirmation bias because his conclusions are based solely on their beliefs. This can be fixed by involving a mix of both right and left handed people in the study.

6. Individuals are participating in a research study about heart health. Researchers have no way of proving whether or not a participant is an active smoker and some participant may have chosen not to disclose that they smoke.

   This is reporting bias since some of the participants are not disclosing information. This can be rectified by asking participants to fill out a form asking health questions. This is hard to rectify completely because they could still lie on the form.

Name:_____ANSWER KEY_____ Date:_____

Chi-Squared Skittles [7]

Chi-Squared (pronounced k-eye) is a method to determine if experimental outcomes result by chance or if something else is affecting the results, i.e. a variable we are testing is affecting results but should not be, type(s) of bias is/are present, or an experimental manipulation occurred.

The magazine you work for wants to present a report proving that the colors in a skittles package are completely by chance (random). You need to verify this, so you chose to run a chi-square test. **Do not eat any of your skittles.**

We calculate Chi-Squared by the following equation

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

O = observed value

E = Expected value

Σ = summation

First, you want to calculate the expected value. We expect to have an equal number of each color. Hence, calculate **E =** $\frac{number\ of\ colors}{total\ skittle\ in\ your\ pack}$ for each color. Fill in the table for Expected with the calculated values. Then, in the Observed row, fill in the actual number of skittles for each color in your pack. Complete this process for each of the colors.

(The following is an example from a packet with 10 skittles)

Individual	Green	Red	Orange	Purple	Yellow
Expected	2	2	2	2	2
Observed	3	1	1	2	3
O-E	-1	1	1	0	-1
$(O-E)^2$	1	1	1	0	1
$(O-E)^2 / E$	0.5	0.5	0.5	0	0.5

Now sum all of the values in the last row to get a **chi-squared** value.

0.5+0.5+0.5+0+0.5 = 2

In order to determine whether or not the results are by chance, we need to compare the chi-square value to a **critical value**. To find the critical value, we first need to determine the **degrees of freedom**.

Degrees of freedom- how much a value can vary from what it is in the experiment.

***Degrees of freedom = number of possible outcomes minus 1*.**

How many degrees of freedom are there in this experiment?

5-1=4

We use this value to find the critical value by the table below. Use the 0.05 column to determine the critical value.

Degrees of freedom	Probability of exceeding the critical value				
	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588

[7]

If your chi-square number is less than the critical value, then your experiment is by chance.

What is our critical value?_____9.488_____

Are your results by chance? Why or why not?

Yes, because 2 is less than 9.488

Where else can Chi-Square be used?

It can be used in other experiments to see if something is affecting the outcome. (let students be creative they may list examples of other things that have to do with chance such as coin flip, dice roll, or deck of cards.)

Annotated Bibliography

[1] Research Bias. (n.d.). Retrieved from https://explorable.com/research-bias
This website was used for was used for adaptation in the Detecting Bias lesson in the Measurements and Data Analysis module to formulate definitions and examples in the Detecting Bias activity.

[2] Standards Aligned System. (n.d.). Retrieved from https://www.pdesas.org/
This website was used in each lesson in the Measurements and Data Analysis module to select proper Pennsylvania State standards, which are based in Common Core, that each lesson is centered around.

[3] Standards for Mathematical Practice. (n.d.). Retrieved from http://www.corestandards.org/Math/Practice/
This website used in every lesson in the Measurements and Data Analysis module to find Standards for Mathematical Practices that are applicable in each lesson.

[4] Nsta. (n.d.). Disciplinary Core Ideas. Retrieved from https://ngss.nsta.org/DisciplinaryCoreIdeasTop.aspx
This website was used in each lesson in the Measurements and Data Analysis module to select appropriate disciplinary core ideas set forth by the NSTA that are at the center of each lesson.

[5] Nsta. (n.d.). Crosscutting Concepts. Retrieved from https://ngss.nsta.org/CrosscuttingConceptsFull.aspx
This website was used in each lesson in the Measurements and Data Analysis module to selecting appropriate crosscutting concepts set forth by the NSTA that apply to each mathematics lesson

[6] Lane, D. M. (n.d.). [2.0]. Retrieved from http://onlinestatbook.com/Online_Statistics_Education.pdf
This online textbook was used for excerption within the Detecting Bias lesson plan as part of the Measurements and Data Analysis module. This reference aided in the completion of providing definitions for the Review Activity.

[7] Non-parametric between-subjects test for nominal data: Chi-square. (n.d.). Retrieved from http://jungminded.weebly.com/statistics/non-paramtric-between-subjects-test-for-nominal-data-chi-square
This image was excerpted from the reference for use within the Detecting Bias lesson in the Measurement and Data Analysis module and used in the Chi-Squared Skittles activity.