Day 1 - Statistics

Ryan Harrison

Agenda

- Goals for Statistics content
- Warm up Problems
- Statistics Background
  - Guidelines for Assessment and Instruction in Statistics Education (GAISE)
  - CCSSM progressions for statistics
- Emergency 911 Data Analysis Task
- Closure

Goals

- For the first week (and into the second week) of our PDs, we will spend roughly 1hr 30 minutes on statistics content each morning.
- I want to be as clear as possible about my plans and goals, as well as the team's goals for this time

Goals and Plans

- 1. A main focus of the summer PDs, specifically the morning statistics activities, is content knowledge.
- 2. Our grant is targeted at high school readiness, so I will focus on 6-12 content
- 3. Engage in 6-12 stats lessons, not just to 'learn' content – but to investigate content.
- 4. Attempt to give you experience with some statistics activities you may have experience with, and others you have not.
- 5. My intention is to expose you to 'statistical thinking' as much as possible…de-emphasizing computations, and encouraging statistical exploration

Warm Up Problems

- **<u>Question #1:</u>**

- Create a set of 7 data points with a mean of 5, median of -1 and mode of 2.

- Discuss: If all numbers remained the same, but the median was made to be smaller (for example it was changed to -15) and smaller, is there a point at which creating the 7 data points is impossible? Why or why not?

Warm Up Problems

- **<u>Question #2:</u>**
- Ten children take a 10-point test. The mean score is 7.
  - Is it possible for the median score to be 10, given that the mean score is 7? If so, give an example of how this could occur. If not, explain why not.
  - Is it possible for the median score to be 6 or less, given that the mean score is 7? If so, give an example of how this could occur. If not, explain why not.
  - Is it possible for the median score to be 3 or less, given that the mean score is 7? If so, give an example of how this could occur. If not, explain why not.

Z-scores Warm-Up

- In statistics, the standard score (z-score) is the signed (±) number of standard deviations an observation is above the mean.

- Thus, a positive standard score indicates an observation that is above the mean
- And a negative standard score indicates an observation that is below the mean.

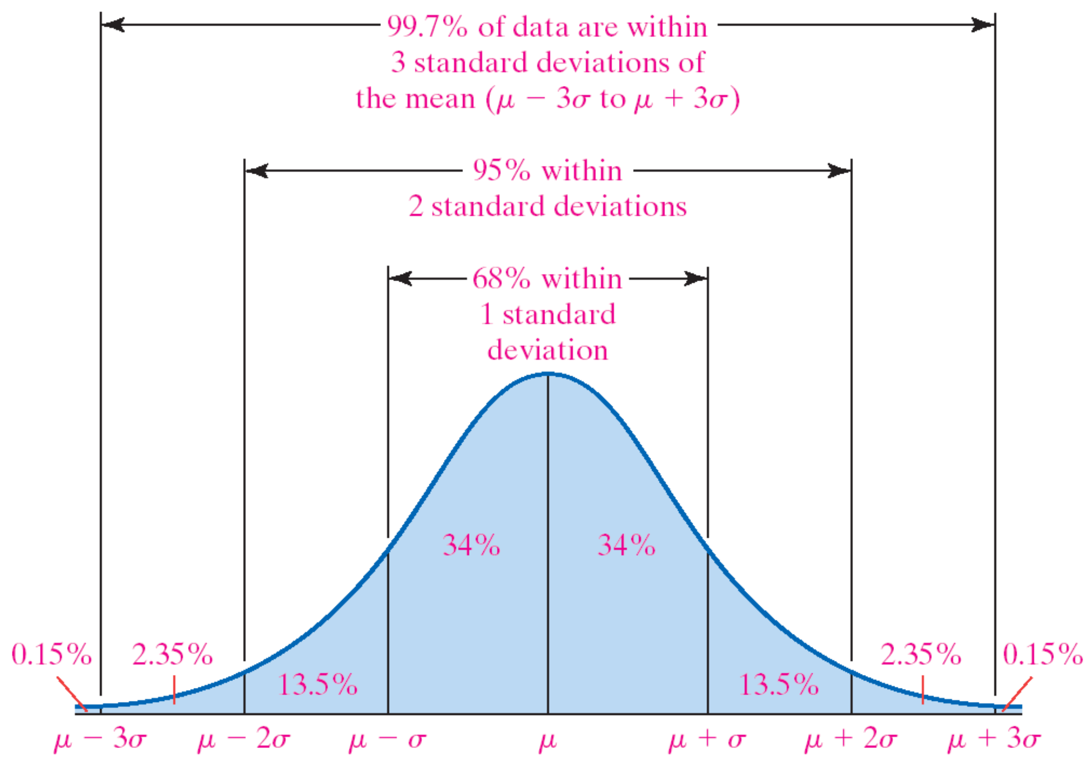
Z-scores Warm-Up

- It is a dimensionless quantity obtained by subtracting the overall mean from an individual raw score and then dividing the difference by the overall standard deviation.
- This conversion process is called standardizing or normalizing.

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation



Z-Scores Warm Up

- 1) A set of mathematics exam scores has a mean of 70 and a standard deviation of 8. A set of English exam scores has a mean of 74 and a standard deviation of 16. For which exam would a score of 78 have a higher standing? Why?

- 2) For a distribution of raw scores with a mean of 45, Steve calculates a Z-score for a raw score of 55 is calculated to be -2.00. Regardless of the value of the standard deviation, why must the Z-score that Steve calculated be incorrect?

Z-Scores Warm Up

- 3) A distribution of scores has a standard deviation of 10. Find the z-scores corresponding to the following values:
- a. A score that is 20 points below the mean
- b. A score that is 10 points below the mean
- c. A score that is 15 points above the mean
- d. A score that is 30 points below the mean

- 4) In a population of scores a raw score with the value of 83 corresponds to a Z of +1.00 and a raw score of 86 corresponds to a Z of +2.00. What is the mean and standard deviation of this population?

Z-Scores Warm Up

- 5) On a statistics exam, you have a score of 73. If the mean of the exam were 65 would you prefer the standard deviation of the scores to be 8 or 16? Why?

- 6) A normal distribution has a mean of 120 and a standard deviation of 20. For this distribution
- a. What score separates the top 16% of the scores from the rest?
- b. What score corresponds closely to the 99th percentile?
- c. What range of scores would form the middle 95% of this distribution?

History of Statistics

- Unlike the field of mathematics, which has texts such as the Rhind Mathematical Papyrus dating back to 2000 BC, the formal study of statistics began closer to the mid 1700s AD.
- Chris will discuss a little more about the history of probability later today, but the history of statistics and probability go hand-in-hand.

History of Statistics

- Wikipedia:
- By the 18th century, the term “statistics" designated the systematic collection of demographic and economic data by states.
- For at least two millennia, these data were mainly tabulations of human and material resources that might be taxed or put to military use.

History of Statistics

- Wikipedia:
- In the early 19th century, collection intensified, and the meaning of "statistics" broadened to include the discipline concerned with the collection, summary, and analysis of data

History of Statistics

- Wikipedia:
- The term "mathematical statistics" designates the mathematical theories of probability and statistical inference, which are used in statistical practice.

History of Statistics

- Wikipedia:
- The relation between statistics and probability theory developed rather late, however. In the 19th century, statistics increasingly used probability theory, whose initial results were found in the 17th and 18th centuries, particularly in the analysis of games of chance (gambling)

GAISE

- Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K -12 Curriculum Framework
- Recognition that Data Analysis and Statistical Literacy is one of the most commonly used mathematical skills in every day life
  - Understanding public opinion polls, making personal and professional decisions, and interpreting science are key skills every citizen must have proficiency in.

GAISE

- A major objective of statistics education is to help students develop statistical thinking.
- Statistical thinking, in large part, must deal with this omnipresence of variability; statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in the data.
- It is this focus on *variability in data* that sets apart statistics from mathematics.

GAISE

- Sources of variability:
- *Measurement Variability*—Repeated measurements on the same individual vary. Sometimes two measurements vary because the measuring device produces unreliable results, such as when we try to measure a large distance with a small ruler.

GAISE

- Sources of variability:
- *Natural Variability*—Variability is inherent in nature. Individuals are different. When we measure the same quantity across several individuals, we are bound to get differences in the measurements.

GAISE

- Sources of variability:
- *Induced Variability*—If we plant one pack of bean seeds in one field, and another pack of seeds in another location with a different climate, then an observed difference in growth among the seeds in one location with those in the other might be due to inherent differences in the seeds (natural variability), or the observed difference might be due to the fact that the locations are not the same.

GAISE

- Sources of variability:

- *Sampling Variability*—The value of a sample proportion will vary from sample to sample. This is called sampling variability.

GAISE

- Framework for Development of Statistical Thinking
- Levels ‘A’ ‘B’ and ‘C’
- 4 components to Statistical Problem Solving
  - ***I. Formulate Questions***
  - ***II. Collect Data***
  - ***III. Analyze Data***
  - ***IV. Interpret Results***

GAISE

– *I. Formulate Questions*

Process Component	Level A	Level B	Level C
I. Formulate Question	**Beginning awareness of the *statistics question distinction*** Teachers pose questions of interest Questions restricted to the classroom	**Increased awareness of the *statistics question distinction*** Students begin to pose their own questions of interest Questions not restricted to the classroom	**Students can make the *statistics question distinction*** Students pose their own questions of interest Questions seek generalization

GAISE

– *II. Collect Data*

Process Component	Level A	Level B	Level C
II. Collect Data	Do not yet *design for differences* Census of classroom Simple experiment	Beginning awareness of *design for differences* Sample surveys; begin to use random selection Comparative experiment; begin to use random allocation	Students make *design for differences* Sampling designs with random selection Experimental designs with randomization

GAISE

– *III. Analyze Data*

Process Component	Level A	Level B	Level C
III. Analyze Data	*Use* particular properties of *distributions* in the context of a specific example Display variability within a group Compare individual to individual Compare individual to group Beginning awareness of group to group Observe association between two variables	Learn to *use* particular properties of *distributions* as tools of analysis Quantify variability within a group Compare group to group in displays Acknowledge sampling error Some quantification of association; simple models for association	Understand and *use* *distributions* in analysis as a global concept Measure variability within a group; measure variability between groups Compare group to group using displays and measures of variability Describe and quantify sampling error Quantification of association; fitting of models for association

GAISE

– *IV. Interpret Results*

Process Component	Level A	Level B	Level C
IV. Interpret Results	Students do not look *beyond the data* No generalization beyond the classroom Note difference between two individuals with different conditions Observe association in displays	Students acknowledge that *looking beyond the data* is feasible Acknowledge that a sample may or may not be representative of the larger population Note the difference between two groups with different conditions Aware of distinction between observational study and experiment Note differences in strength of association Basic interepretation of models for association Aware of the distinction between association and cause and effect	Students are able to *look beyond the data* in some contexts Generalize from sample to population Aware of the effect of randomization on the results of experiments Understand the difference between observational studies and experiments Interpret measures of strength of association Interpret models of association Distinguish between conclusions from association studies and experiments

GAISE

- As we work through activities later this week, we may have time to look more closely at some examples of what Levels A, B, and C look like for each of the 4 processes
- We will do activities that involve/encourage thinking at all levels.

CCSSM Progressions

- These documents show the general trend of content development through the Common Core standards.
- I will not spend much time today speaking about these documents – because your homework assignment will be to read these over the course of this first week.

CCSSM Progressions

- We have planned out the first 6 days of the PD to focus more on content (statistics and probability is the theme for this summer).

- While we obviously will not go in depth into content that addresses **<u>all</u>** of the standards, we will try to track along the progression from grades 6 – 12, and push some of you to be exposed to new statistical ideas from the higher grades (even if you don't teach them)

CCSSM Progressions

- The team and I recognize that we have significant variability in the grades we teach. That is both a good thing, and a challenge.

- Our hope is that the activities we do are engaging, have entry points that are accessible to most, and give you opportunities to wrestle with statistical ideas that you may or may not have been exposed to before.

CCSSM Progressions

- I have designed my statistics activities based on existing lesson plans, that you will be given at the conclusion of the PD.

- The lessons are intended to be useful for you to engage in, but also potentially take back to your classroom/school. At least that is my hope.

911! Activity

- The accident last week at the Waterfront Amusement Park once again points to the necessity of formalizing a plan for rapid response to the park by EMS units.
- When the seat to a ride broke loose resulting in the serious injury to two area teens, once again EMS response time, critical to the well being of the injured, ended up unacceptably high.
- Park owners have charged that a lower response time would have resulted in fewer complications for the injured, and we agree. We urge the city council to initiate an investigation into the reasons for inconsistent response times to the area near the park.

911! Activity

- Page 2 of your handout displays the expectations for what needs to be produced from the data analysis process
- However, we're going to do this activity with a slight twist
- Some groups will be:
  - Neutral members of city council
  - Hired consultants representing 'Arrow Ambulance Service'
  - Hired consultants representing 'Metro Ambulances'

Your Task

- With your groups, analyze the data provided and include multiple numerical and graphical findings on a piece of chart paper. Also prepare a short narrative that could be read.
- There are no 'rules' regarding which data/facts that you use. However, you must have justification for the decisions you make.
  - If you opt to not include something because it is an 'outlier' – be prepared to defend that decision. The same applies to all other similar decisions.
- The goal of this is **<u>NOT</u>** to engage in debate about whose data is most persuasive.
- The goal of this **<u>IS</u>** to engage in debate over how things like bias and persuasion impact statistical decisions.

Discussion

- How does bias impact our decisions?
- Are there any facts that all 3 perspectives agree on? What factors were relevant?
- What data was included? Ignored? Why?

Discussion

- Statistics Quotes:
- Do not put your faith in what statistics say until you have carefully considered what they do not say. ~William W. Watt

- Statistics are used much like a drunk uses a lamppost: for support, not illumination. ~Andrew Lang

- Facts are stubborn things, but statistics are more pliable. ~Mark Twain