Session 686

Activities and Applets for Intuitive Understanding of SRS, Stratified, and Cluster Random Sampling

Beth Chance (bchance@calpoly.edu)
Cal Poly, San Luis Obispo, California
Ruth Carver (Ruth.Carver@germantownacademy.org)
Germantown Academy, Fort Washington, Pennsylvania

Turn and Talk (1 min each)

Introduce yourself to your neighbor on the right
- Where you are from
- Grade level you teach
- In what course do you teach about selecting samples from a population?
  - What do you most want students to understand?

About the Activity

- Educated citizens need to understand how random sampling enables us to make conclusions about a population (with a measure of uncertainty)
  - 7.SPA1: Understanding random sampling… Gauge variation in sample statistics
  - AP Statistics Exam often asks students about advantages and disadvantages of different sampling methods in specific contexts
- We wanted a way for them to easily **explore** the implications of different sampling methods
- **Simulation** often gives students a more concrete visualization of the theoretical results
 - Tactile, Computer

---

Outline

- **Introductions**

- **About the Activity**

- **Notice/Wonder**
  - **As you view the different simulations**
    - **What do you notice?**
    - **What do you wonder?**

- **TURN AND TALK**
  - **After each exploration, we will ask you to turn and talk with a neighbor about what you noticed and anything you are wondering about.**

- **Questions**

SAMPLING STARS IN THE SKY

- *Exploration 2.2.1 – Introduction to Statistical Investigations AP Version*
  - *Adapted From Petocz (1990), Sampling Space: Practical Experiments for Teaching Sampling*

Have you ever looked up into the night sky on a clear night and wondered how many stars there are in the sky?

How could you go about estimating the number of stars?
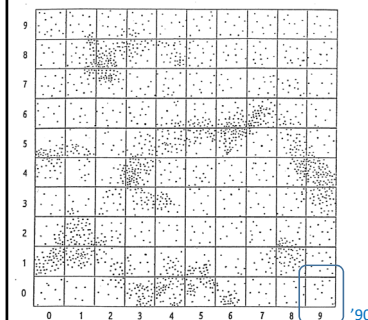
- Assume we have a photograph of part of the night sky at a given time
- The negative for the photograph was used, so the stars appear as black dots and the background sky is white
- We will identify each square by a two-digit number consisting of the number on the horizontal axis (0-9) followed by the number on the vertical axis (0-9).
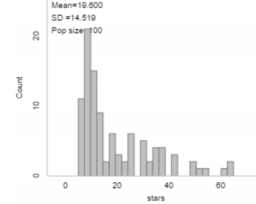
Goal: Estimate the number of stars in this section of the night sky.
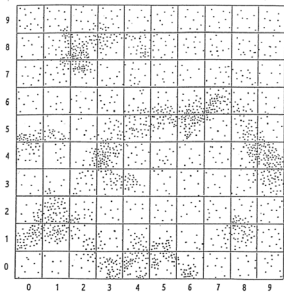
- ➢ Why not just count them?
- ➢ Alternative? Random sample?
  - ➢ *Sampling frame* = 100 squares
  - ➢ Randomly selecting 10 squares, count the stars in those 10 squares
- ➢ *Statistic* = Average number of stars per square
  - ➢ Then multiply by 100 to estimate total number of stars
- ➢ Will this "work"?

Sampling Stars

Population data:

Mean number of stars per square = 19.60 stars
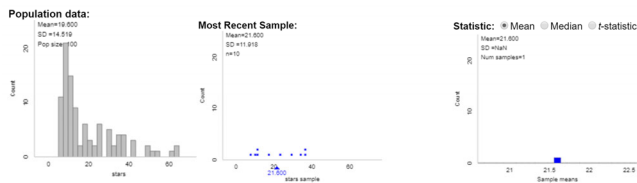**Std Dev = 14.519 stars**

Parameter = Average number of stars per square in this grid

Does random sampling work?
  Does the sample mean tend to be close to the population mean?

Explore this method by taking lots of samples from the same population

---

Some things to pay attention to in the Simulation

Three Different Graphs

Method #1: Simple Random Samples

Rossman/Chance One Mean Applet
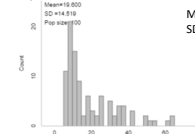
Notice/Wonder
  As you view the different simulations
    - What do you notice?
    - What do you wonder?
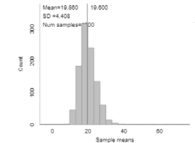   - Does the method seem to work?

TURN AND TALK

Population data:



Mean = 19.8
SD = 14.519

Mean = 19.9
SD = 4.408

Other Sampling Methods

- Would another method be more convenient?
- Would another method be more accurate?

Method #2: Stratified Random Sample
  Variable = "density"

If you were going to divide the star map into two strata, high density and low density, approximately how many squares would you classify as high density and how many as low density?

Just do a quick visual inspection, no need to actually count the number of stars in each square.
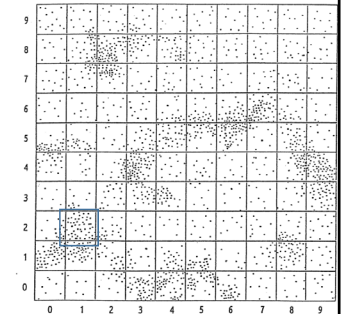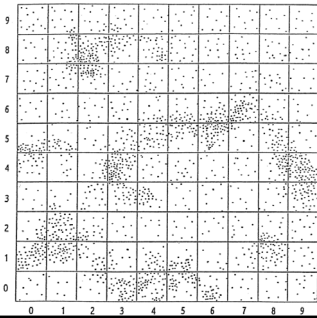
Method #2: Stratified Random Sample Variable = "density"

If you were going to divide the star map into two strata, high density and low density, approximately how many squares would you classify as high density and how many as low density?
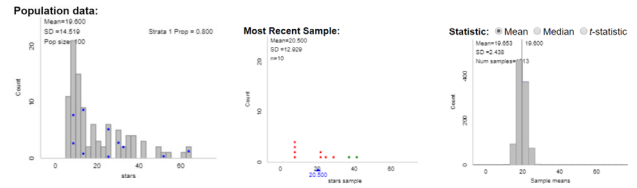
How does this suggest changing the sampling method? Why?

TURN AND TALK

Method #2: Stratified Random Sample *Rossman/Chance One Mean Applet* Variable = "density"
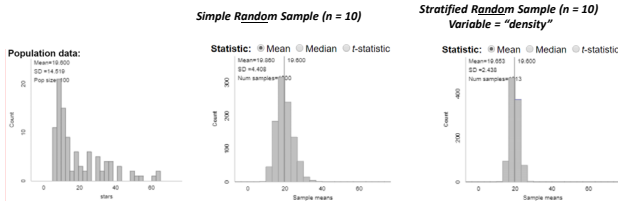
Population data:

Notice/Wonder
As you view the different simulations
- What do you notice?
- What do you wonder?
- Does the method seem to work?

TURN AND TALK

Comparison of Simple Random Sample (SRS) and Stratifying by Density

Simple Random Sample (n = 10)   *Stratified Random Sample (n = 10)* *Variable = "density"*

Population data:

Compare your results when stratifying by density and when taking simple random samples. Is one sampling method preferable over the other? Explain how so/not.

[*Hint:* Which changed the most, the center, shape, or variability? What are the implications of this change?]

TURN AND TALK


Method #3 *Rossman/Chance One Mean Applet*
Stratified Random Sample: Variable = "side" (left/right)

Let's look at another variable to to help understand how to choose a 'good' variable to stratify on.

Suppose instead of choosing density as our variable on which to stratify, we instead decided to split the star map vertically into two equal halves,
- where stars in the left 50 squares of the map (columns 0 – 4) would be in the left stratum and
- stars in the right 50 squares of the map (columns 5 – 9) would be in the right stratum.

Why might using which side the stars are in (left/right) as the variable to stratify on be more feasible than knowing the density of the squares in advance?
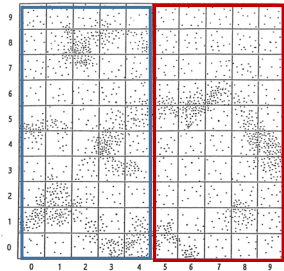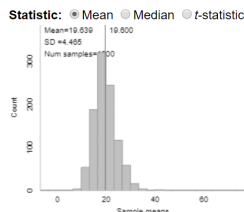

Does the choice of one variable provide a statistical advantage over the other as your choice of variable to stratify on? Explain.

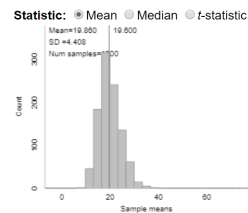*Stratified Random Sample (n = 10)* *Variable = "density"*
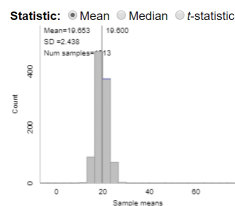
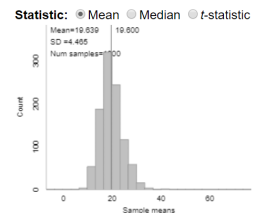*Stratified Random Sample (n = 10)* *Variable = "side"*

TURN AND TALK

Compare your results when **stratifying by side** and when taking a **simple random sample**. Is one sampling method preferable over the other? Explain how so/not.

Simple Random Sample (n = 10)   *Stratified Random Sample (n = 10)* *Variable = "side"*


If the population does not have obvious strata (subgroups that behave differently from each other on the variable(s) of interest), then a simple random sample ensures a high degree of representativeness and is simpler and less time consuming than obtaining a stratified random sample.

Method #4
Cluster Random Sample: Variable = "column"

Describe a method you could use to select a cluster random sample of 10 squares from the star map to estimate the mean number of stars in each square, using column as the cluster variable.

How does this suggest changing the sampling method? Why?
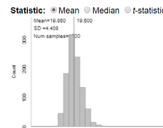*Predict the behavior of the sample means in this case*

TURN AND TALK
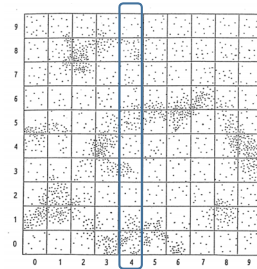
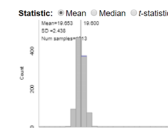**Method #4**
Cluster Random Sample: Variable = "column"
Compare your results when selecting a **cluster sample using the column** as the cluster variable with **stratifying by density** and with taking a **simple random sample**.

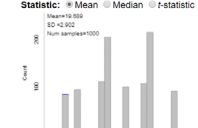**Is one sampling method preferable over the other? Explain.**

Simple Random Sample (n = 10)

Stratified Random Sample (n = 10)
Variable = "density"

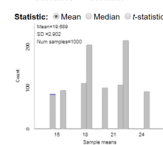*Cluster Random Sample (n = 10)*
Variable = "column"

TURN AND TALK

Method #5
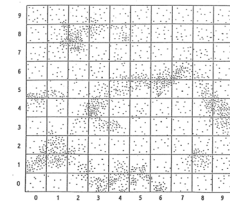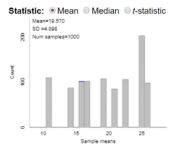**Cluster Random Sample: Variable = "row"**

Compare your results for the two different variables that you used to obtain a **cluster random sample** (**column** and **row**).

Does the choice of one variable provide a statistical advantage over the other as your choice of cluster variable?

Cluster Random Sample (n = 10)
Variable = "column"

Cluster Random Sample (n = 10)
Variable = "row"

Explain why this makes sense by looking at the rows and columns in the map.

TURN AND TALK

Comparing stratified sampling to cluster sampling

	Strata	Clusters
Is a listing of the whole population required?	Yes	No
Number of groups	Comparatively few strata	Typically a large number of clusters
Sizes of groups	Often large	Smaller
Measure units from	Every stratum	Only a simple random sample of clusters
Reason(s) for using	To ensure small strata are represented and to reduce the standard deviation of estimates	To get a random sample without having to list all units in the population

---

Thank You!
Questions?

Beth Chance (bchance@calpoly.edu)
Ruth Carver (Ruth.Carver@germantownacademy.org)
RossmanChance applets: www.rossmanchance.com/applets/
Introduction to Statistical Investigations (www.isi-stats.com/)
SBI blog (https://www.causeweb.org/sbi/)