Chpater 5 Moodle Questions

Rebekah Robinson

Tuesday, June 17, 2014

PART 1

# GC Baseball Simple Random Sample

You want to do a survey of members of the baseball team at Georgetown College to see how they feel about academic life at GC. You want to ensure that you take a *simple random sample*. There are several sampling strategies that you have in mind. Which one of the following strategies is a simple random sample?

You obtain a list of all GC baseball players and pick the first 12 names on the list.

You obtain a list of all GC baseball players and you select one of the first 5 names at random. Then, you select every fifth name on the list after that until you have 12 people selected.

You find out that there are 6 baseball players enrolled in each of the four sections of MAT 111. You write the sections on slips of paper - A, B, C, D - and place the slips of paper in a hat. You shuffle the papers and then select 2 slips of paper from the hat without replacement. The students in the two chosen sections make up your sample.

You find out that there are 20 JV players and 40 Varsity players on the baseball team at GC. To make your sample representative of this distribution, you randomly select 4 players from the JV team and 8 players from the Varsity team to be in your sample.

You obtain a list of all GC baseball players. You then write the name of each player on a slip of paper and put the papers in a hat. You thoroughly shuffle the names and then select 12 slips of paper from the hat without replacement.

# Kim Kardashian SRS

Suppose you are interested in the following

**Research Question:** On average, how do folks from imagpop feel about Kim Kardashian, on a 0 - 100 scale?

You decide to take a **simple random sample** of size 50 from this population. Which of the following lines of code will accomplish this?

RandomExp(imagpop, size=50)  
popsamp(imagpop, n=50)  
sample(imagpop, n=50)  
popsamp(imagpop, size=50)  
resample(imagpop, n=50)

# Kim Kardashian Sample Statistics

Suppose you are interested in the following

**Research Question:** On average, how do folks from imagpop feel about Kim Kardashian, on a 0 - 100 scale?

You decide to take a **simple random sample** of size 50 from this population. After drawing the sample, you compute the average temperature rating.

## min Q1 median Q3 max mean sd n missing  
## 0 8 81 95 100 55.8 42.52 50 0

Which of the following is the correct way to report your results?

mu=55.8 xbar=55.8 sigma=42.52 s=42.52

# What Type of Sample?

A random sample of adults that walk into Kroger on Wednesday mornings are stopped and asked if they would like to participate in a taste test to determine which soft drink is more popular - Coke or Pepsi.

What type of sample is this?

Simple Random Sample Stratified Sample Cluster Sample Volunteer Sample None of these

# Parameters vs. Statistics - Numerical Variable

The dataset imagpop contains information about 10000 people from an imaginary population. One of the variables is income, rounded to the nearest $100. You can see a numerical summary of the incomes of the population by using the function favstats:

favstats(~income,data=imagpop)

## min Q1 median Q3 max mean sd n missing  
## 200 19300 33600 54100 262200 40317 28739 10000 0

Suppose you take a simple random sample of size 100 from this population. You can see a numerical summary of the incomes of the sample by running the following code:

set.seed(2014)  
mysamp<-popsamp(n=100,pop=imagpop)  
favstats(~income,data=mysamp)

## min Q1 median Q3 max mean sd n missing  
## 2100 16775 27650 44650 138900 33337 24274 100 0

Match the following numbers with their correct descriptions.

mu - $40316.72 xbar - $33337 population size - 10000 sample size, n - 100 sigma - $28738.63 s - $24273.93

# Parameters vs. Statistics - Factor Variable

The dataset imagpop contains information about 10000 people from an imaginary population. One of the variables is cappun, which indicates whether the individual favors or opposes the death penalty. You can see the distribution of of this variable in the population in the row percents shown below:

rowPerc(xtabs(~cappun,data=imagpop))

## favor oppose Total  
## 29.76 70.24 100.00

Suppose you are interested in the following

*Research Question:* What proportion of individuals in the imagpop population oppose capital punishment.

Suppose you decide to take a simple random sample of size 300. You can see the distribution of cappun in the sample below:

set.seed(2014)  
mysamp<-popsamp(n=300,pop=imagpop)  
rowPerc(xtabs(~cappun,data=mysamp))

## favor oppose Total  
## 27.33 72.67 100.00

Match the following numbers with their correct descriptions.

p - 0.7024 (70.24%) phat - 0.7267 (72.67%) population size - 10000 sample size, n - 300 0.2976 (29.76%) 0.2733 (27.33%)

# Small Sample Size

The dataset imagpop contains information about 10000 people from an imaginary population. One of the variables is math, which indicates whether the individual was a mathematics major. You can see the distribution of math majors in the population in the row percents shown below:

rowPerc(xtabs(~math,data=imagpop))

## no yes Total  
## 95.37 4.63 100.00

Suppose you were interested in the following

**Research Question:** Are quantitative skills an important part of your job?

You decide to quickly take a simple random sample of size 10 from this population. You can see the distribution of math majors in the sample below:

set.seed(2014)  
smallsamp<-popsamp(n=10,pop=imagpop)  
rowPerc(xtabs(~math,data=smallsamp))

## no yes Total  
## 100 0 100

Which of the following statements best describes a problem that may arise in this scenario?

* Since 100% of the individuals in the sample are not math majors, this sample is not representative of the population. About 5% of the population are math majors. Our sample should reflect this distribution. This introduces bias into the results.
* The sampled individuals that were selected might all have jobs that require quantitative skills. This could introduce bias into the results because most jobs do not require quantitative skills.
* The researchers decided which individuals were part of the sample. This is an example of quota sampling. This introduces bias in the results because researchers may have chosen a group of people that are not representative of the population, even if this was done unintentionally.
* This was not really a simple random sample because an individual who was not a math major has a greater chance of being selected to be part of the sample than an individual who was a math major.

PART 2

# Types of Bias Matching

For each definition, identify the correct term for the type of bias being defined.

Participants respond differently from how they truly feel. - Response bias

The method for selecting the participants produces a sample that does not represent the population of interest - Selection bias

A representative sample is chosen, but a subset of the sample cannot be contacted or does not respond. - Non-response bias

# Potential Bias Matching I

For each of the following situations, match whether the potential bias is a **selection bias**, a **non-response bias**, a **response bias**, or **no potential bias**.

In a college town, college students are hired to conduct door-to-door interviews to determine whether city residents think there should be a law forbidding loud music at parties. - Response Bias

A "Tie of the Month" club sends a mail survey to a random sample of its subscribers asking them whether they would like to receive a new tie every other week rather than just once a month. - Non-response Bias

A random sample of adults that walk into Kroger on Wednesday mornings are stopped and asked if they prefer Coke or Pepsi. - Selection Bias

No Potential Bias

# Potential Bias Matching II

For each of the following situations, match whether the potential bias is a **selection bias**, a **non-response bias**, a **response bias**, or **no potential bias**.

In a study of women's opinions about community issues, investigators randomly selected a sample of households and interviewed a woman from each selected household. When no woman was present in a selected household, a woman from the house next door was interviewed instead. The survey was done between the hours of 8 AM and 3 PM. (Selection Bias)

A survey question asked of unmarried men was, "What is the most important feature you consider when deciding whether to date somebody?" The results were found to depend on whether the interviewer was male or female. (Response Bias)

A telephone survey of 500 residences was conducted. People refused to talk to the interviewer in 200 of the residences. (Non-Response Bias)

No Potential Bias

# Order Bias - Explanatory Variable

You are interested in studying order bias in questionnaires. Specifically, you want to know which of the following two questions (*Version A* or *Version B*) is more likely to elicit a response of "Hilary Clinton":

* (*Version A*) "If, in 2016, the following two persons were to receive the nomination of their party for President, whom would you be more inclined to support?"
* Hilary Clinton (Democratic Party)
* Rand Paul (Republican Party)
* (*Version B*) "If, in 2016, the following two persons were to receive the nomination of their party for President, whom would you be more inclined to support?"
* Rand Paul (Republican Party)
* Hilary Clinton (Democratic Party)

You have the resources to survey at least 2000 randomly-selected registered voters. Which of the following is the explanatory variable?

# Order Bias - Response Variable

Which of the following is the response variable?