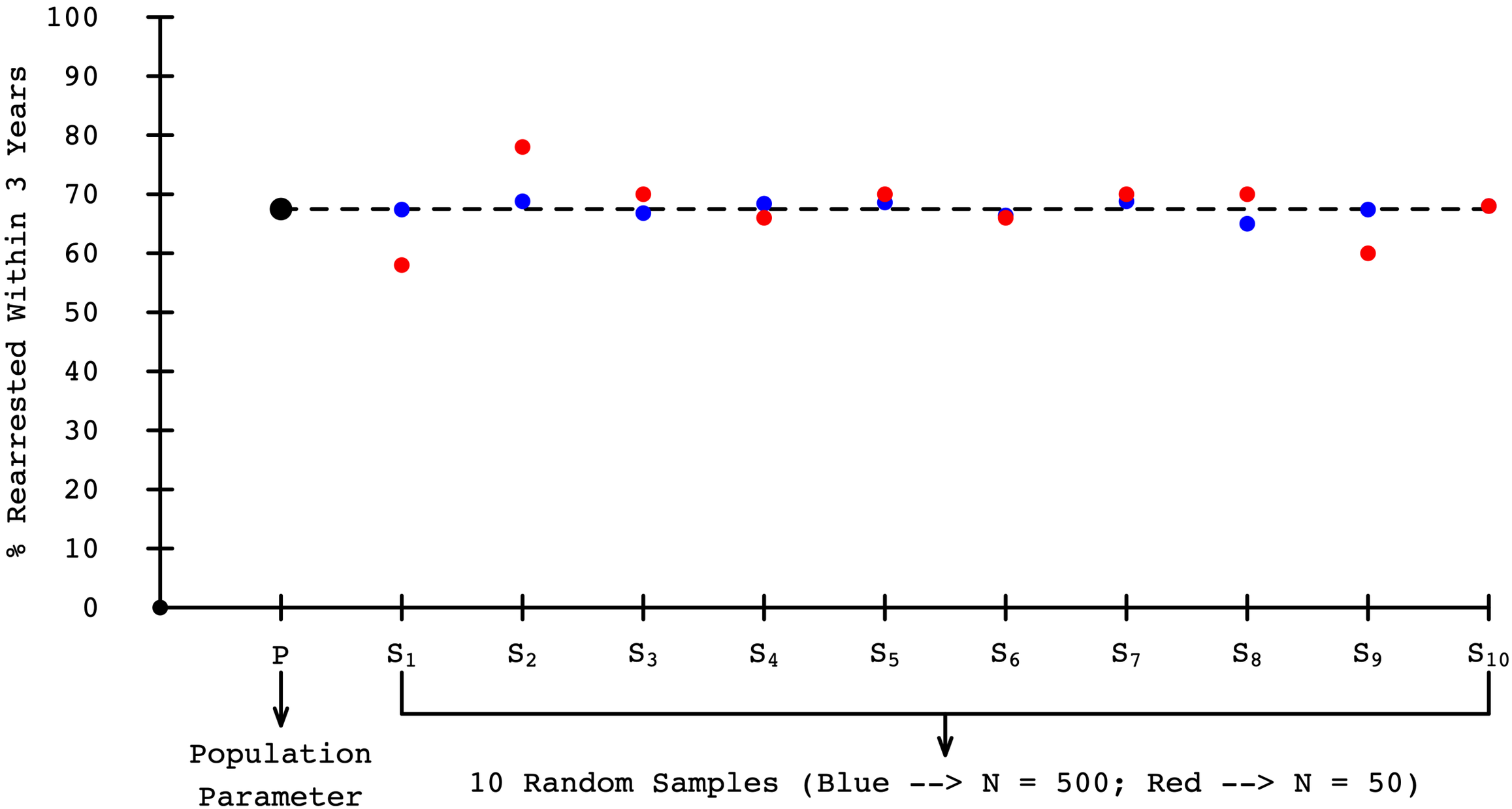
Lesson 3

Thursday 2/1/24

Recidivism Rate Estimates From 20 Simple Random Samples



Note: a simple random sample means that every case in the population has the same chance of being selected to be in the sample.

Simple Random Samples

- We have a well-defined, scientifically interesting population.
- It's not feasible to study the entire population.
- Each unit within the population has the same chance of being selected to be in the sample.
- Independence: the selection of any case for the sample has nothing to do with the selection of any other case.
- Sampling is carried out with replacement.

Example - Samples of Size 10

```
> population = c(rep("no",325000),rep("yes",675000))
> table(population)
population
    no    yes
325000 675000
>
> s1 = sample(population,size=10,replace=T)
> s1
 [1] "no"  "yes" "no"  "yes" "no"  "yes" "yes" "yes" "yes" "yes"
> s2 = sample(population,size=10,replace=T)
> s2
 [1] "yes" "no"  "yes" "yes" "yes" "yes" "no"  "no"  "yes" "yes"
> s3 = sample(population,size=10,replace=T)
> s3
 [1] "yes" "no"  "yes" "no"  "yes" "yes" "yes" "yes" "no"  "yes"
> s4 = sample(population,size=10,replace=T)
> s4
 [1] "yes" "yes" "no"  "yes" "no"  "yes" "yes" "yes" "yes" "yes"
> s5 = sample(population,size=10,replace=T)
> s5
 [1] "yes" "yes" "yes" "yes" "yes" "yes" "yes" "yes" "no"  "yes"
> |
```

Probability Samples

- All simple random samples are probability samples.
- There are probability samples that are not simple random samples.
- Simple random sample: every member of the population has the same chance of being selected for the sample, sampling is independent (the selection of one case has nothing to do with the selection of any other case).
- Probability sample: every member of the population has a known and non-zero chance of being selected for the sample.

A Brief Chat About Probability

- In everyday language, we often think of the word "probability" as synonymous with the chance, odds, or likelihood that something will happen.

- We can be a little more specific by thinking of probability in one of the following 2 ways: (1) a degree of belief; or (2) the limit of a relative frequency distribution.

- The second approach is the one to which we will refer most often in this class.

- In this case, we can think of "probability" as the long-run proportion of times something happens.

- Easy example: suppose there are 1M people released from prison and 675K of them are rearrested within 3 years of release. We can say that the probability that someone drawn at random from this population is rearrested within 3 years is 0.675 (675K/1M).

- More to come on probability ...

Different Types of Probability Samples

- Simple random samples
- Stratified samples
- Cluster samples
- Stratified cluster samples
- Stratified, multi-stage cluster samples

What these all have in common is that each member of the population has a known and nonzero chance of being selected for the sample.

↓

Generalizability

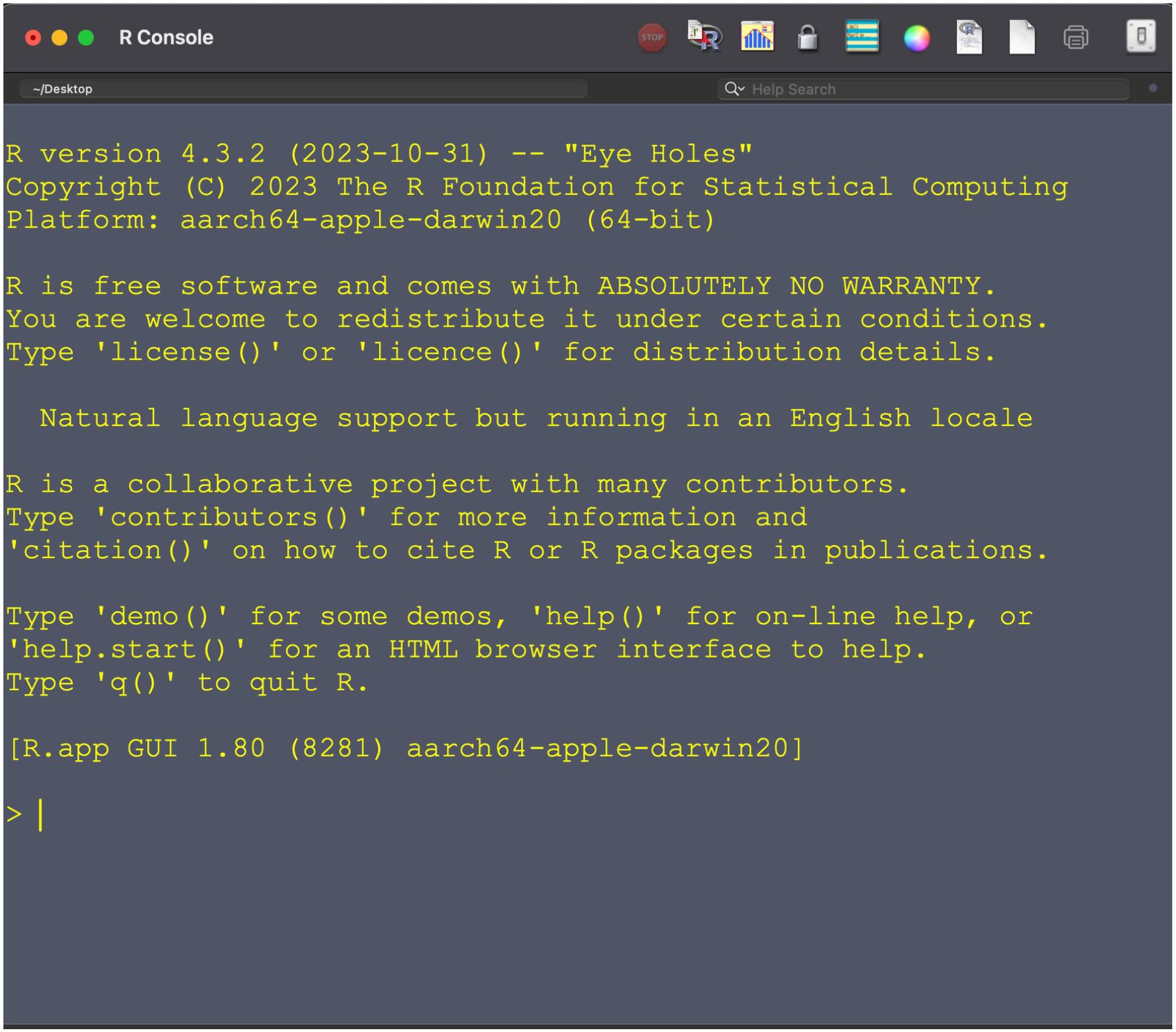
Nonprobability Samples

- Any sample that is not a probability sample
- Examples: we survey a class of students; CNN has a poll question on their website and they report the results on the evening news; you go over to the Stamp Union and ask a set of questions to the first 15 freshman and the first 15 seniors you can talk to.
- Different types of nonprobability samples: (1) convenience/availability samples; quota samples; (3) purposive/judgment samples (includes snowball samples).
- Why would we do this?

Introduction to R

- Website: https://www.r-project.org/
- Runs on Windows or MacOS
- Some people like to use RStudio with R (I use plain old R; Jane and Jordan like to use RStudio; they will go over it in discussion section).
- You can get it here: https://posit.co/products/open-source/rstudio/
- R is free.

Here's What it Looks Like



Let's Do A Couple of Calculations

Copy and paste into R:

```
x=2
y=3
x+y
x-y
x*y
x/y
x^y
factorial(x*y)
```

Output:

```
> x=2
> y=3
> x+y
[1] 5
> x-y
[1] -1
> x*y
[1] 6
> x/y
[1] 0.6666667
> x^y
[1] 8
> factorial(x*y)
[1] 720
>
```

A Grade in this Class

Copy and paste
into R:

```
exam1=77
exam2=81
exam3=80
assignment1=93
assignment2=87
assignment3=88
average.assignment=mean(assignment1,assignment2,assignment3)
average.assignment
formula1=0.25*exam1+0.25*exam2+0.25*exam3+0.25*average.assignment
formula1
formula2=0.2*exam1+0.2*exam2+0.2*exam3+0.4*average.assignment
formula2
```

And, Here are the Results!

```
> exam1=77
> exam2=81
> exam3=80
> assignment1=93
> assignment2=87
> assignment3=88
> average.assignment=mean(assignment1,assignment2,assignment3)
> average.assignment
[1] 93
> formula1=0.25*exam1+0.25*exam2+0.25*exam3+0.25*average.assignment
> formula1
[1] 82.75
> formula2=0.2*exam1+0.2*exam2+0.2*exam3+0.4*average.assignment
> formula2
[1] 84.8
>
```

So, this person would get a B using either formula (since an 82.8 rounds to an 83).