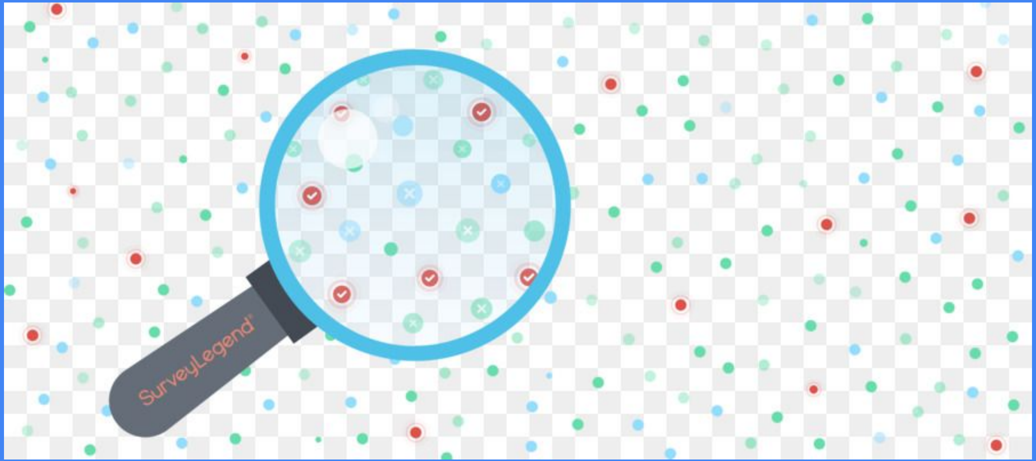
Lab 8: Distributions + simulations

November 30, 2020 - December 1, 2020



Review: Sampling

Sample distribution

Why we want to sample?

- “A random sample of the data can help us get some point estimates for population” parameters (mean, sd, variance, proportion).”
- a ***sampling distribution*** of our estimate allows us to “learn about the properties of the estimate, such as its distribution”.

Population Parameter	
Population mean	μ
Population standard deviation	σ
Population proportion	P
Population size	N
Population data value	X
Correlation coefficient	r

Random number Generation/seed

How does R get Random numbers to sample?

Truly random numbers are expensive

So R uses a “pseudo”random number generator.

This kind of generation is a deterministic (depends on the starting value) process that is almost the same as a true random process

The **seed** is the starting value that determines this sequence.

sample()

`x <- vector we would like to sample from`

`N <- how many values do we want in the sample`

`sample(x, N)`

We use the sample function because it is easier to work with a smaller sample than the whole data set

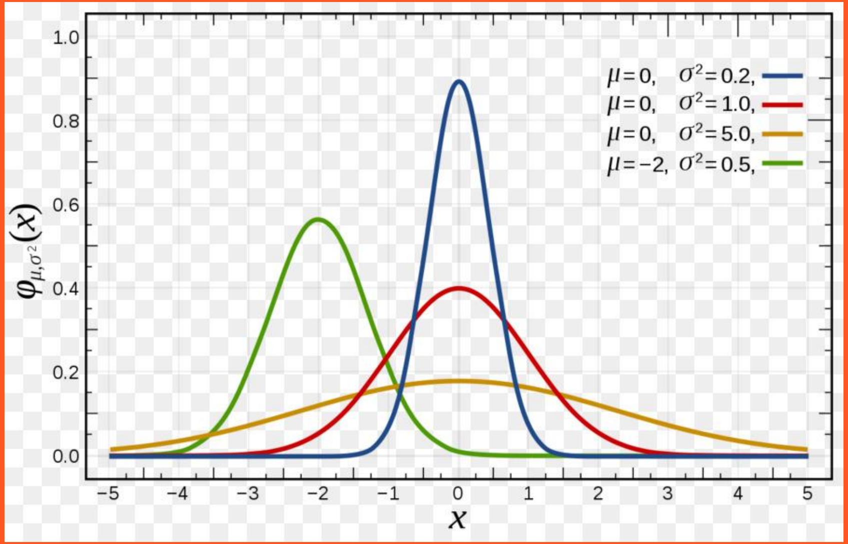
The output is a new vector of size N

Sampling With replacement / without replacement?

sampling **with replacement** means that if I'm drawing from a complete deck of 52 playing cards and I first draw the queen of spades, I put it back into the deck again (so the deck still has 52 cards) so I can still have a chance of drawing the queen of spades.

sampling **without replacement** means that I keep the card from the first draw and for the second draw I'm picking from an incomplete deck of 51 cards .

Simulating from distributions



Why Do We Simulate?

A simulation is a mathematical process that models random events.

- examine how methods work
- calculate things that are difficult to do mathematically.
- We want to use a pseudo-random procedure

All of these require sampling random variables from distributions

Common distributions

Can you think of some common distributions that we might want to draw from?

- Normal distribution
- t distribution
- uniform distribution
- Binomial distribution

Binomial

$$\boxed{\binom{n}{k}(1-p)^{n-k}(p)^{k} = P(x=k)}$$

R-code:

N <- 1000

n <- 10

p <- 0.5

dbinom(x=4, size=10, prob=0.5, log = FALSE)

rbiom(N, size = n, probability = p)

Normal

Formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x)$ = probability density function

σ = standard deviation

μ = mean

R-code:

```
N <- 1000
 m <-  0
 s <-  1
```

rnorm(N, mean = m, sd= s)

Uniform

PDF	$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$

R-code:

```
N <- 1000
b <- 1
a <- 0
runif(N, min = a , max = b)
```

Finding Power

Previously, you learned how to calculating power and sample size for t-tests of two means when the group sizes are equal.

What about when the group sizes differ?

Simulations allows us to calculate power in settings where this would otherwise be very difficult!

Finding power in R

(1) Create a loop that iterates the following steps (2-5) N times:
(2) Randomly draw observations for the 2 groups from normal distributions with different means but same sd.
(3) use t.test() to compare two sample means
(4) reject the equal sample mean hypothesis if p value is smaller than your predetermined alpha level (alpha = 0.05)
(5) count for rejections and add it to grand total number of rejections
(6) Outside of the Loop: number of rejections / N = power

Announcements

Lab 8 is the last lab assignment! (yay!!!!)

Next week: Review !!!!!!!

Come to lab next week with Questions!

Some review questions from the textbook: 2.8.2, 8.8, 9.6, 9.8

If there are topics/HW problems/labs/app ex./questions from midterm/chapters that you want to review email me before lab and I'll make sure to go over it.