

STAT 331 Lab 10

Instructions

Submit an HTML markdown document by the beginning of class.

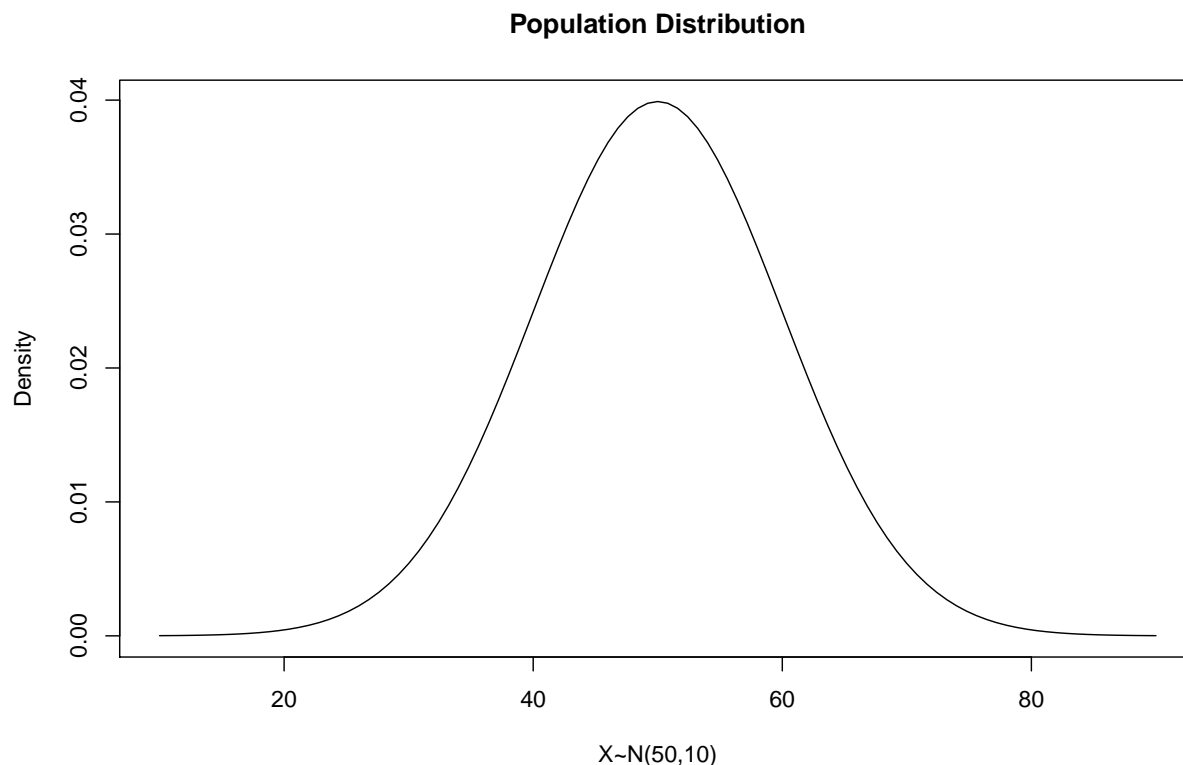
In this lab we'll use the probability distribution functions and graphics to further explore probability distributions.

Exercises

Recall that the Central Limit Theorem tells us about the distribution of the sample mean, which is called the sampling distribution. As sample size becomes large, the sampling distribution approaches a normal distribution, and the standard deviation of the sampling distribution decreases. The mean of the sampling distribution is equal to the population mean.

Not all statistics have such nice theory and machinery to take advantage of! In the following exercise, you will explore the sampling distributions of a few different statistics: median, standard deviation, range, and third quartile.

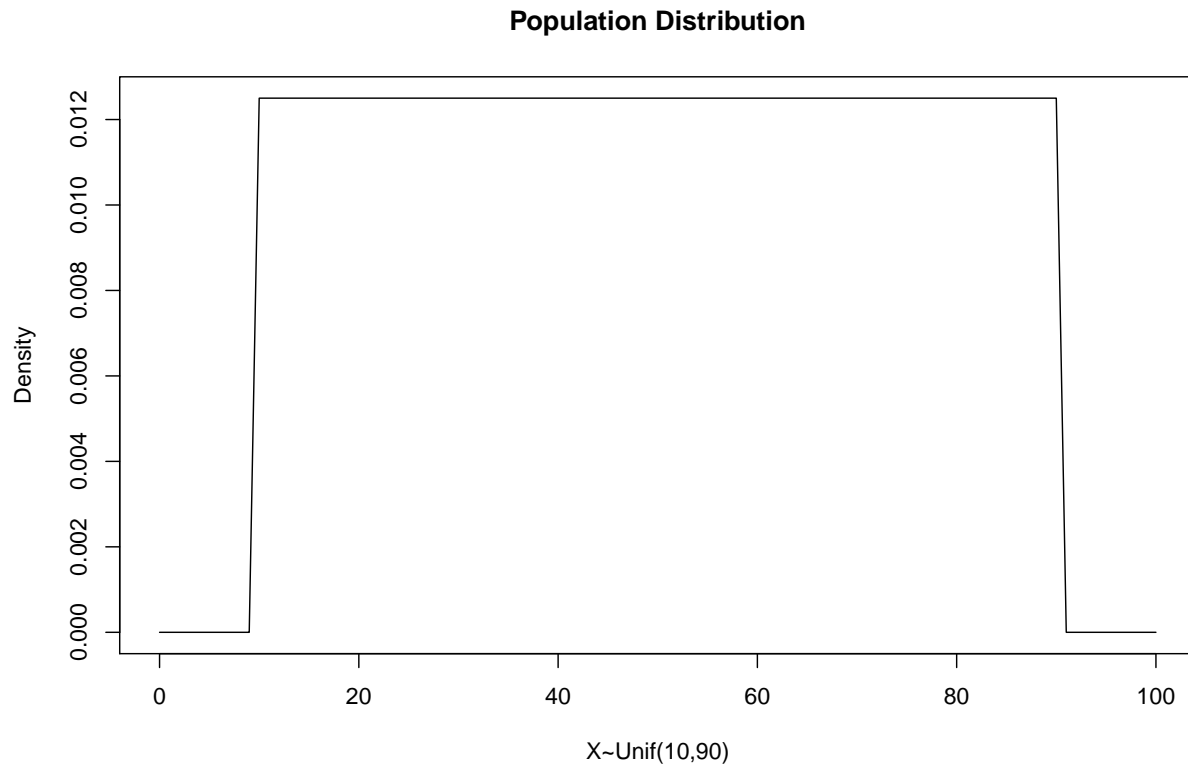
1. Suppose our population is Normal with mean 50 and standard deviation 10.



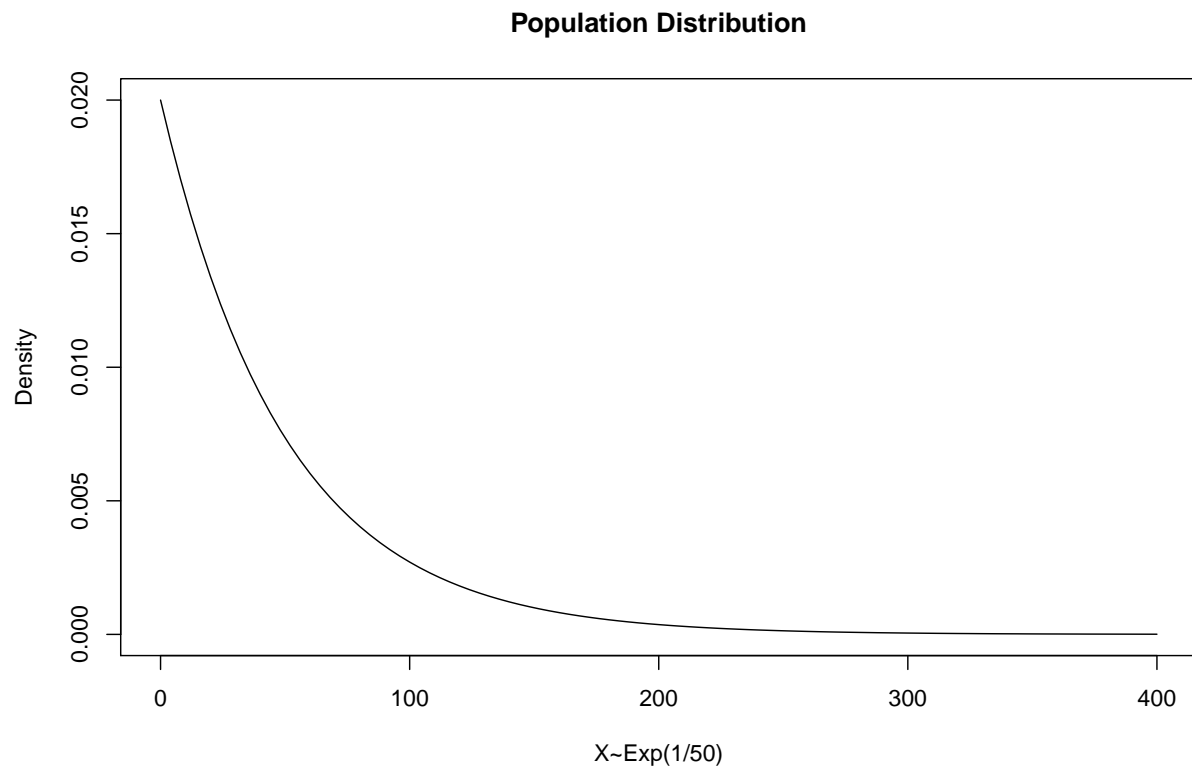
- a) Simulate 1000 samples of size 30 from this population.
- b) Recall that $\text{range} = \text{max} - \text{min}$. The `range()` function in R does NOT compute this. Use this function to compute the actual range:

```
myrange <- function(x) { max(x) - min(x) }
```

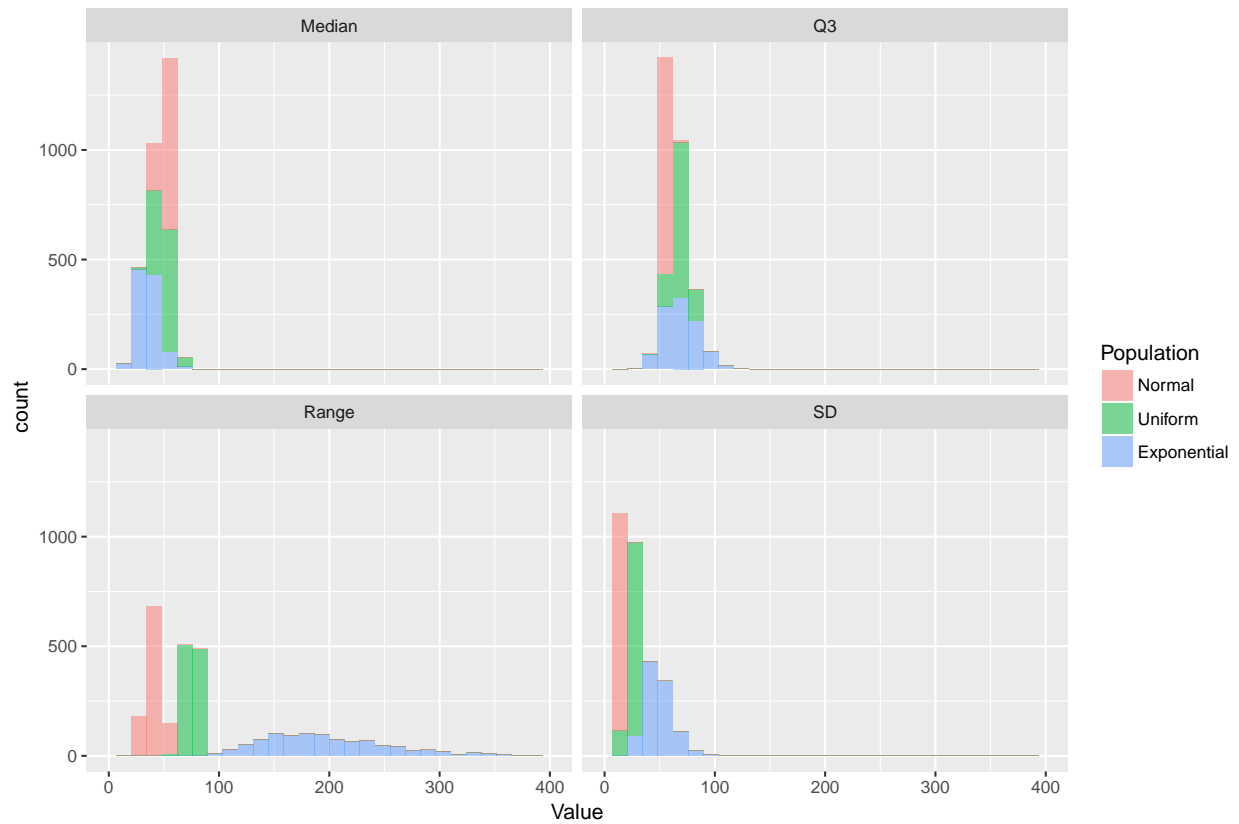
- c) Compute the median, standard deviation, range, and third quartile of each of these samples. Store them in separate vectors.
 - d) Plot the sampling distributions of these four statistics in (four separate graphs) of a single plotting window. Be sure to label things appropriately.
- 2) Repeat (a), (c), and (d) in 1) for a Uniform distribution with minimum of 10 and maximum of 90.



- 3) Repeat (a), (c), and (d) in 1) for an Exponential distribution with mean 50.



- 4) Using ggplot2, create 4 graphs in the same plotting window similar to part (d) in 1). However, in each of these 4 plots the sampling distributions from 1), 2), and 3) should be overlaid. The following is an example:



- 5) Based on all of your plots above, what can you conclude about the sampling distributions of these statistics as compared to the that of the sample mean? How sensitive are these sampling distributions to the underlying population distribution?

EXTRA CREDIT:

- 6) Convert your code for generating and plotting the data in 4) to a function that takes a single argument: `sample_size`. Create 4 sets of these 2x2 plot grids for 4 different sample sizes using this function: 5, 15, 30, 100. Comment on the differences between the different sample sizes.