Laboratory Exercise Week 7

Ryan Estes | 006, 10 pts

10/4/17

*Directions*:

* Write your R code inside the code chunks after each question.
* Write your answer comments after the # sign.
* To generate the word document output, click the button Knit and wait for the word document to appear.
* RStudio will prompt you (only once) to install the knitr package.
* Submit your completed laboratory exercise using Blackboard's Turnitin feature. Your Turnitin upload link is found on your Blackboard Course shell under the Laboratory folder.

For this exercise, you will need to use the package mosaic to find numerical and graphical summaries.

library(mosaic) # load the package mosaic to use its functions

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggformula

## Loading required package: ggplot2

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData

## Loading required package: Matrix

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.  
##   
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median,  
## prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

1. Consider the population of CEO salaries in the lesson this week.
   1. Select a random sample of 40 CEO's and compute the sample mean salary and sample standard deviation salary.
   2. Use the do() function to compute the sample mean of 100 randomly selected samples. Plot the histogram of these 100 sample means.
   3. Compute the average and standard deviation of these 100 sample means. Are the values what you expect from the properties of the sampling distribution.

### Code chunk

# start your code  
ceo\_salary <- read.csv("https://www.siue.edu/~jpailde/CEO\_Salary\_2012.csv")  
  
# i)  
(s40 <- sample\_n(ceo\_salary, size = 40))

## ï..Rank Name Company  
## 96 96 John C Lechleiter Eli Lilly & Co  
## 382 382 Andrï¿½s R Gluski AES  
## 444 444 David H Hannah Reliance Steel & Aluminum  
## 462 462 Daniel S Fulton Weyerhaeuser  
## 103 103 Mark C Pigott Paccar  
## 119 119 Sanjay K Jha Motorola Mobility Holdings  
## 221 221 Timothy J Naughton Avalonbay Communities  
## 14 14 Howard D Schultz Starbucks  
## 106 106 Lawrence J Ellison Oracle  
## 115 115 John P Daane Altera  
## 454 454 Edward J Bonach CNO Financial  
## 57 57 Scott D Sheffield Pioneer Natural Resources  
## 315 315 William D Johnson Progress Energy  
## 356 356 Charles W Shivery Northeast Utilities  
## 151 151 James M Cracchiolo Ameriprise Financial  
## 81 81 Gale E Klappa Wisconsin Energy  
## 377 377 Craig S Donohue CME Group  
## 491 491 Joseph W Brown MBIA  
## 266 266 Richard T O'Brien Newmont Mining  
## 219 219 William P Utt KBR  
## 376 376 Charles E Haldeman Jr Freddie Mac  
## 467 467 Denise L Ramos ITT  
## 360 360 Michael J Kowalski Tiffany & Co  
## 109 109 Gregory E Johnson Franklin Resources  
## 319 319 Ilene S Gordon Corn Products International  
## 309 309 Ralph Izzo Public Service Enterprise Group  
## 251 251 Arne N Sorenson Marriott International  
## 23 23 John B Hess Hess  
## 180 180 Kevin Burke Consolidated Edison  
## 273 273 John T Gremp FMC Technologies  
## 284 284 Thomas C Gallagher Genuine Parts  
## 183 183 Debra A Cafaro Ventas  
## 285 285 John A Hayes Ball  
## 380 380 Joseph R Ficalora New York Community Bancorp  
## 173 173 Ellen J Kullman EI du Pont de Nemours  
## 412 412 R Adam Norwitt Amphenol  
## 208 208 Paul C Varga Brown-Forman  
## 277 277 Murray D Martin Pitney Bowes  
## 60 60 Michael J Ward CSX  
## 297 297 W Craig Jelinek Costco Wholesale  
## Annual.pay Age  
## 96 15.570 58  
## 382 3.656 54  
## 444 2.325 60  
## 462 1.930 63  
## 103 15.110 58  
## 119 14.005 49  
## 221 7.865 50  
## 14 41.470 58  
## 106 14.890 67  
## 115 14.110 48  
## 454 2.076 58  
## 57 20.050 59  
## 315 5.105 58  
## 356 4.260 66  
## 151 11.170 54  
## 81 17.270 61  
## 377 3.830 50  
## 491 0.700 63  
## 266 6.520 57  
## 219 7.900 55  
## 376 3.870 64  
## 467 1.735 55  
## 360 4.185 59  
## 109 14.610 50  
## 319 5.025 58  
## 309 5.270 54  
## 251 6.965 53  
## 23 33.200 58  
## 180 9.535 61  
## 273 6.250 61  
## 284 6.050 64  
## 183 9.495 54  
## 285 6.020 46  
## 380 3.685 65  
## 173 9.730 56  
## 412 3.075 43  
## 208 8.490 48  
## 277 6.120 64  
## 60 19.835 61  
## 297 5.756 60

cat("A sample mean of 40 CEO salaries equates to: ")

## A sample mean of 40 CEO salaries equates to:

(s40.mu <- mean(~ Annual.pay, data = s40))

## [1] 9.467825

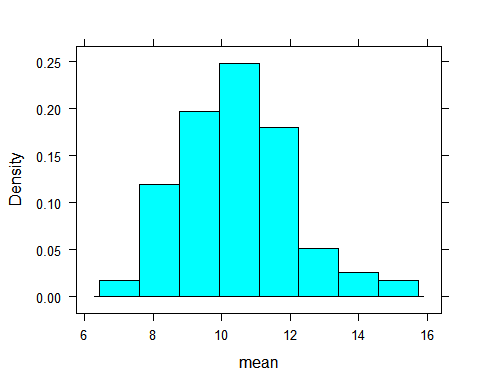
cat("A sample SD of 40 CEO salaries equates to: ")

## A sample SD of 40 CEO salaries equates to:

(s40.sigma <- sd(~ Annual.pay, data = s40))

## [1] 8.319481

# ii)  
s100 <- do(100)\*mean(~ Annual.pay, data = sample\_n(ceo\_salary, size = 40))  
histogram(~ mean, data = s100)



# iii)  
cat("A sample mean of 40 CEO salaries 100 times equates to: ")

## A sample mean of 40 CEO salaries 100 times equates to:

(s100.mu <- mean(~ mean, data = s100))

## [1] 10.42632

cat("A sample SD of 40 CEO salaries 100 times equates to: ")

## A sample SD of 40 CEO salaries 100 times equates to:

(s100.sd <- sd(~ mean, data = s100))

## [1] 1.681692

# By looking at the histogram of the data, the values are what I expect for the samples given.  
  
# last R code line

1. Rockwell hardness of pins of a certain type is known to have a mean value of 50 and standard deviation of 1.2.
   1. If the distribution is normal, what is the probability that the sample mean hardness for a random sample of 9 pins is at least 51?
   2. Without assuming population normality, what is the (approximate) probability that the sample mean hardness for a random sample of 40 pins is at least 51?

### Code chunk

# start your code  
  
  
  
# i)  
1 - pnorm(q = 51, mean = 50, sd = 1.622/sqrt(9)) #0.322%

## [1] 0.03218789

# ii)  
1 - pnorm(q = 51, mean = 50, sd = 1.2/sqrt(40)) #6.8x10^-8%

## [1] 6.804011e-08

# last R code line

1. Suppose that a random sample of size 64 is to be selected from a population with mean 40 and standard deviation 5.
   1. What are the mean and standard deviation of the sampling distribution of the sample mean?
   2. What is the approximate probability that the sample mean will be within 0.5 of the population mean?

### Code chunk

# start your code  
  
# i)  
mu <- 40  
sigma <- 5/sqrt(64) #0.625  
  
# ii)  
xpnorm(q = c(39.5, 40.5), mean = mu, sd = sigma) #0.576

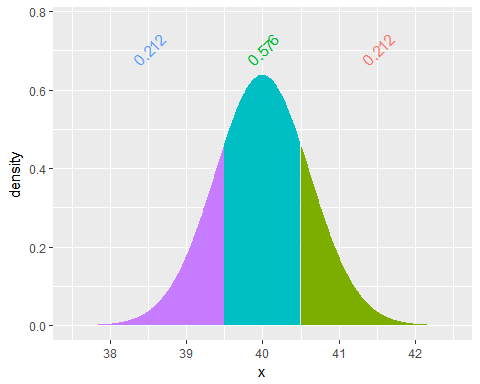
##

## If X ~ N(40, 0.625), then

## P(X <= 39.5) = P(Z <= -0.8) = 0.2119 P(X <= 40.5) = P(Z <= 0.8) = 0.7881

## P(X > 39.5) = P(Z > -0.8) = 0.7881 P(X > 40.5) = P(Z > 0.8) = 0.2119

##



## [1] 0.2118554 0.7881446

# last R code line