HW3

Bin Lin

2016-9-23

```
#install.packages(c('openintro','OIdata','devtools','ggplot2','psych','reshape2',
#                   'knitr','markdown','shiny'))
#devtools::install_github("jbryer/IS606")

library(IS606)
```

```
##
## Welcome to CUNY IS606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 3rd Edition. You can read this by typing
## vignette('os3') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='IS606') will list the demos that are available.
```

```
##
## Attaching package: 'IS606'
```

```
## The following object is masked from 'package:utils':
##
##     demo
```

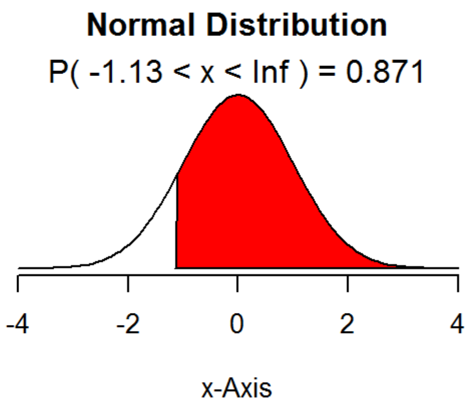
```
getwd()
```

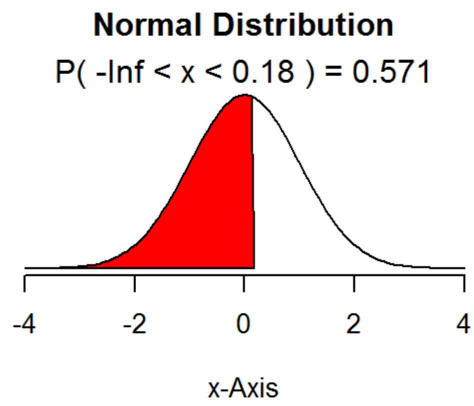
```
## [1] "C:/Users/blin261/Documents/R/DATA606"
```

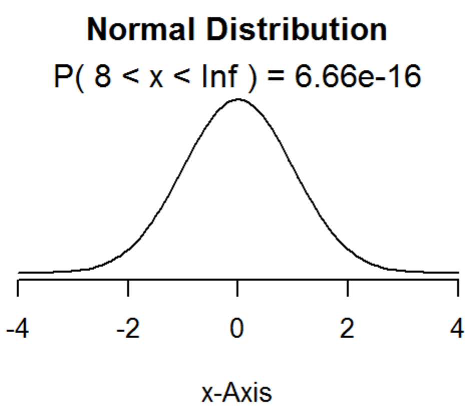
```
setwd('C:/Users/blin261/Documents/R/DATA606')
```

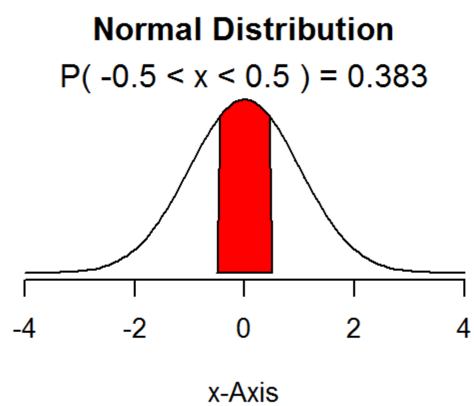
3.2 Area under the curve, Part II. What percent of a standard normal distribution N(?? = 0, ! = 1) is found in each region? Be sure to draw a graph. (a) Z > ???1.13 (b) Z < 0.18 (c) Z > 8 (d) |Z| < 0.5

```
par(mfrow=c(2, 2))
normalPlot(bounds = c(-1.13, Inf))
normalPlot(bounds = c(-Inf, 0.18))
normalPlot(bounds = c(8, Inf))
normalPlot(bounds = c(-0.5, 0.5))
```









3.4 Triathlon times, Part I. (a) Write down the short-hand for these two normal distributions. Men, Ages 30-34: N(?? = 4313, = 583) Women, Ages 25-29: N(?? = 5261, = 807)

b. What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you? Leo's z-score: (4948-4313)/583 = 1.089 Nary's z-score: (5513-5261)/807 = 0.312

c. Did Leo or Mary rank better in their respective groups? Explain your reasoning. Leo did much better. Because z-score is a standardized score. It describes the number of standard deviation it falls above or below the mean. Leo's z-score tell us, his result is 1.089 standard deviation above the mean, while mary's result is just 0.312 standard deviation above the mean. So Leo ranks better.

d. What percent of the triathletes did Leo finish faster than in his group?

```
pnorm(4948, mean = 4313, sd = 583)
```

```
## [1] 0.8619658
```

e. What percent of the triathletes did Mary finish faster than in her group?

```
pnorm(5513, mean = 5261, sd = 807)
```

```
## [1] 0.6225814
```

f. If the distributions of finishing times are not nearly normal, would your answers to parts

g. ◦ e. change? Explain your reasoning. Yes, except part b. Because if the distribution is not normal, we can still calculate the z-score, but it is actually pointless. All the techniques that is used in parts (c) through (e)are under the assumption the distribution is normal.
3.18 Heights of female college students.

h. The mean height is 61.52 inches with a standard deviation of 4.58 inches. Use this information to determine if the heights approximately follow the 68-95-99.7% Rule. Within 1 standard deviation (56.94, 66.1): 17/250 = 0.68 Within 2 standard deviation (52.36, 70.68): 24/25 = 0.96 Within 3 standard deviation (47.78, 75.26): 25/25 = 1 The probability is quite similar to 68-95-99.7%

i. Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.

The data appear to follow a normal distribution, because it is unimodal. The points on the normal probability plot also seem to follow a straight line.

3.22 Defective rate. A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others. (a) What is the probability that the 10th transistor produced is the first with a defect? 0.98^9*0.02 = 0.017

b. What is the probability that the machine produces no defective transistors in a batch of 100? 0.98^100 = 0.133

c. On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation? E(X) = 1/p = 1/0.02 = 50 Standard deviation = sqrt ((1-p)/p^2) = sqrt((1-0.02)/0.02^2) = 49.5

d. Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation? E(X) = 1/p = 1/0.05 = 20 Standard deviation = sqrt ((1-p)/p^2) = sqrt((1-0.05)/0.05^2) = 19.5

e. Based on your answers to parts (c) and (d), how does increasing the probability of an event affect the mean and standard deviation of the wait time until success? The increasing the probability of an event, the less the mean of the wait time until the even occur. The standard deviation of the wait time is also decreased.

3.38 Male children. While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids. (a) Use the binomial model to calculate the probability that two of them will be boys.

```
choose(3, 2) * 0.51^2 * 0.49
```

```
## [1] 0.382347
```

b. Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers from parts (a) and (b) match. BBG 0.51*0.51*0.49 = 0.127 BGB 0.51*0.49*0.51 = 0.127 GBB 0.49*0.51*0.51 = 0.127 Total: = 0.382

c. If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a).

```
choose(8, 3) * 0.51^3 * 0.49^5
```

```
## [1] 0.2098355
```

Because there are 56 possibilities of gettting 3 boys out of 8 children. To calculate the probability for each possibilities than sum up is a tedious process

3.42 Serving in volleyball. A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other. (a) What is the probability that on the 10th try she will make her 3rd successful serve?

```
choose(9, 2) * 0.15^3 * 0.85^7
```

```
## [1] 0.03895012
```

b. Suppose she has made two successful serves in nine attempts. What is the probability that her 10th serve will be successful? 0.15

c. Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be different. Can you explain the reason for this discrepancy? First of all, her serves are independent of each other, that means for each individual serve, the success rate is 0.15. Therefore, for part b, disregarding how many serve she made in the begining, she will still have 15% chance of making successful serve on her 10th try. However, for part a, because the question is asking the probability that she make her 3rd successful serve on the 10th try. That means on the 10th try, she will have to make the serve. The probability that is calculated is predetemined to take into consideration of that situation.