Assessment 04 - Continuous Probability

Gabriele Mineo - Harvard Data Science Professional

Distribution of female heights - 1

Assume the distribution of female heights is approximated by a normal distribution with a mean of 64 inches and a standard deviation of 3 inches. If we pick a female at random, what is the probability that she is 5 feet or shorter?

Instructions

- Use `pnorm` to define the probability that a height will take a value less than 5 feet given the stated distribution.

```
# Assign a variable 'female_avg' as the average female height.
female_avg <- 64

# Assign a variable 'female_sd' as the standard deviation for female heights.
female_sd <- 3

# Using variables 'female_avg' and 'female_sd', calculate the probability that a randomly selected fema
pnorm(5*12, female_avg, female_sd)
```

```
## [1] 0.09121122
```

Distribution of female heights - 2

Assume the distribution of female heights is approximated by a normal distribution with a mean of 64 inches and a standard deviation of 3 inches. If we pick a female at random, what is the probability that she is 6 feet or taller?

Instructions

- Use `pnorm` to define the probability that a height will take a value of 6 feet or taller.

```
# Assign a variable 'female_avg' as the average female height.
female_avg <- 64

# Assign a variable 'female_sd' as the standard deviation for female heights.
female_sd <- 3

# Using variables 'female_avg' and 'female_sd', calculate the probability that a randomly selected fema
1-pnorm(6*12, female_avg, female_sd)
```

```
## [1] 0.003830381
```

Distribution of female heights - 3

Assume the distribution of female heights is approximated by a normal distribution with a mean of 64 inches and a standard deviation of 3 inches. If we pick a female at random, what is the probability that she is between 61 and 67 inches?

Instructions

- Use `pnorm` to define the probability that a randomly chosen woman will be shorter than 67 inches.
- Subtract the probability that a randomly chosen will be shorter than 61 inches.

```
# Assign a variable 'female_avg' as the average female height.
female_avg <- 64

# Assign a variable 'female_sd' as the standard deviation for female heights.
female_sd <- 3

# Using variables 'female_avg' and 'female_sd', calculate the probability that a randomly selected fema
pnorm(67, female_avg, female_sd) - pnorm(61, female_avg, female_sd)
```

```
## [1] 0.6826895
```

Distribution of female heights - 4

Repeat the previous exercise, but convert everything to centimeters. That is, multiply every height, including the standard deviation, by 2.54. What is the answer now?

Instructions

- Convert the average height and standard deviation to centimeters by multiplying each value by 2.54.
- Repeat the previous calculation using pnorm to define the probability that a randomly chosen woman will have a height between 61 and 67 inches, converted to centimeters by multiplying each value by 2.54.

```
# Assign a variable 'female_avg' as the average female height. Convert this value to centimeters.
female_avg <- 64*2.54

# Assign a variable 'female_sd' as the standard deviation for female heights. Convert this value to cen
female_sd <- 3*2.54

# Using variables 'female_avg' and 'female_sd', calculate the probability that a randomly selected fema
pnorm(67*2.54, female_avg, female_sd) - pnorm(61*2.54, female_avg, female_sd)
```

```
## [1] 0.6826895
```

Probability of 1 SD from average

Compute the probability that the height of a randomly chosen female is within 1 SD from the average height.

Instructions

- Calculate the values for heights one standard deviation taller and shorter than the average.
- Calculate the probability that a randomly chosen woman will be within 1 SD from the average height.

```
# Assign a variable 'female_avg' as the average female height.
female_avg <- 64

# Assign a variable 'female_sd' as the standard deviation for female heights.
female_sd <- 3

# To a variable named 'taller', assign the value of a height that is one SD taller than average.
taller <- female_avg + female_sd

# To a variable named 'shorter', assign the value of a height that is one SD shorter than average.
shorter <- female_avg - female_sd
```

```r
# Calculate the probability that a randomly selected female is between the desired height range. Print
pnorm(taller, female_avg, female_sd) - pnorm(shorter, female_avg, female_sd)
```

```
## [1] 0.6826895
```

Distribution of male heights

Imagine the distribution of male adults is approximately normal with an expected value of 69 inches and a standard deviation of 3 inches. How tall is a male in the 99th percentile?

Instructions

- Determine the height of a man in the 99th percentile, given an average height of 69 inches and a standard deviation of 3 inches.

```r
# Assign a variable 'female_avg' as the average female height.
male_avg <- 69

# Assign a variable 'female_sd' as the standard deviation for female heights.
male_sd <- 3

# Determine the height of a man in the 99th percentile of the distribution.
qnorm(0.99, male_avg, male_sd)
```

```
## [1] 75.97904
```

Distribution of IQ scores

The distribution of IQ scores is approximately normally distributed. The expected value is 100 and the standard deviation is 15. Suppose you want to know the distribution of the person with the highest IQ in your school district, where 10,000 people are born each year.

Generate 10,000 IQ scores 1,000 times using a Monte Carlo simulation. Make a histogram of the highest IQ scores.

Instructions

- Use the function `rnorm` to generate a random distribution of 10,000 values with a given average and standard deviation.
- Use the function `max` to return the largest value from a supplied vector.
- Repeat the previous steps a total of 1,000 times.
- Plot the histogram of values using the function `hist`.

```r
# The variable `B` specifies the number of times we want the simulation to run.
B <- 1000

# Use the `set.seed` function to make sure your answer matches the expected result after random number
set.seed(1)

# Create an object called `highestIQ` that contains the highest IQ score from each random distribution
highestIQ <- rep(B, {

max(rnorm(10000,100,15))

})
```

```r
# Make a histogram of the highest IQ scores.
hist(highestIQ)
```