**Zachary Shin**

**October 1, 2020**

**Introduction to the Normal Distribution: Hand Span and Height Part II**

**As a result of completing this exercise you should be able to:**

* Explain what a *distribution* is
* Distinguish between a *sample* and a *population*
* Describe the *meaning of N(𝛍,𝛔)* and the standard normal distribution N(0,1)
* Construct *histograms* in R to describe distributions of quantitative data
* Understand the *meaning of the z-score*
* Use *the 68-96-99.7 rule* to approximate the area under the normal curve between any two data points or z-scores, or to the left of a given z-score, or to the right of a given z-score
* Find the *area under the normal curve* between data points or z-scores

**Instructions**

1. We will **use the dataset collected from previous ISAT 251 students called “ISAT251\_HandSpanHeight.csv”.** The data were collected by each student, he/she 1) recorded his/her own hand span and height and 2) found other 4 people who are outside of our ISAT 251 class to measure their hand span and height. To measure the hand span: 1) open the right hand as widely as possible and 2) measure the distance from the thumb to pinky using *centimeter* as the unit

The dataset includes the following variables or features:

* student: a student observer’s jmu eid (the id on his/her duke email)
* participant.id: the id of the participant (1 for the first participant, 2 for the second participant, and so on)
* gender: the participant’s gender (M as male, F as female)
* right.hand.span: the span of the participant’s right hand (from thumb to pinky) in **centimeters**
* height: the height of the participant in **centimeters**.

The top row, header, of csv file will be labeled like this (with one word in each cell):

student participant.id gender age right.hand.span height

1. Now it’s time to **bring the data into R**.
2. Go to the **Canvas** and download the data file **“ISAT251\_HandSpanHeight.csv”**. Move the file to a directory on your local machine that you can access using R and RStudio.
3. Open the RStudio Statistical Software. Use file.choose function to navigate and get the full path to your file where your hand span and height data stored in it. Then, store the path to a variable call myfile.

> myfile <- file.choose()

> myfile

**The full path to *your* file is: /Users/zshindc/Downloads/ISAT251\_HandSpanHeight.csv**

1. Create a new variable called handspan\_height which will hold every observation for records everyone collected outside of the class. The argument header=TRUE tells R that the first row of my data set contains variable or column names.

> handspan\_height <- read.csv(myfile,header=TRUE)

Once you’ve loaded the CSV file using read.csv, you can check to make sure it’s there with the head command, which pulls out the first six observations only.

> head(handspan\_height)

**Snapshot and paste your R code and the first six lines of your data set here:Text

Description automatically generated**

**Table

Description automatically generated**

1. Before you jump into your analyses, please use the summary function to get an overview of your data as well s to check whether there are weird data values or missing values

> summary(handspan\_height

**Snapshot and paste your R code and the summary of your data set here:Text

Description automatically generated**

A picture containing text, table

Description automatically generated

**Now, we are ready to work on exercises for the Normal Distributions**

1. Make a pdf plot of the height variable from **all observations (all students, regardless of gender) using a histogram with probability on the y-axis and height on the x-axis**. **Snapshot and paste your R code and the pdf plot here.**

Chart, scatter chart

Description automatically generatedText

Description automatically generated

1. Then, check whether the distribution of heights of **all observations** your class collected is normal by creating a **QQ plot** (also called a “**normal probability plot**”) -- and always include the qqline in your plot. **Snapshot and paste your R code and the plot here.**

Chart, scatter chart

Description automatically generated

A picture containing graphical user interface

Description automatically generated

1. Based on **those plots in question 2 and question 3**, is the height of all observations (regardless of gender) normally distributed?

**The height of all observations are normally distributed.**

1. Make a pdf of the height variable from **observations whose gender is M (male students only) using a histogram**. **Snapshot and paste your R code and the pdf plot here.**

*Please use the code below to create a variable,* handspan\_height.male*, to get all the observations with gender labeled as M. Then ,* handspan\_height.male *in your work for question 5 to question 7.*

> handspan\_height.male <- handspan\_height [which(handspan\_height $gender=='M'),]

1. Then, check whether the distribution of height of **male observations** your class observed is normal by creating a **QQ plot** (also called a “**normal probability plot**”) -- and always include the qqline in your plot. **Snapshot and paste your R code and the plot here**. (If needed, look in the table of contents for your book to find the chapter that contains the R code for doing this.**)**

A picture containing text

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

1. **Based on those plots in question 5 and question 6,** is the height of **male observations** normally distributed?

The height of male observations are normally distributed.

1. Before moving forward, **what does the z-score mean? Explain here**.

The z-score describes what the difference is between data point (x) and the mean of the distribution, scaled by how skinny or fat the bell curve is. Positive z-scores are associated with data points that are ABOVE the mean, and negative z-scores are associated with data points that are BELOW the mean.

1. What’s the formula for z-score: **write the formula out** (VARIABLES ONLY) and then **explain what each variable is**.
2. Please your information to fill the form below. Then, **use the formula for z-score** to compute the z-score for *each student’s height. P*lease calculate the sample mean () and sample standard deviation () of the heights in your data*. Then use the sample mean and sample standard deviations to calculate the z-score.* **Do NOT do the last two columns yet. You will be asked to come fill that in later in question 15.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Your eid | gender | Height (in centimeter) | z-score using  (mean, sd of the class data\*) | z-score using  (mean, sd of the American adults\*\*) | % of American adults\*\* with a height ***taller*** than you |
| shinzd | Male | 172.72 | -0.3757 | -0.503 | 69.1% |

Note:

\* If you are a male, please use the mean and standard deviation of **the male data from ISAT251\_HandSpanHeight.csv**; if you are a female, please use the mean and standard deviation of **the female data from ISAT251\_HandSpanHeight.csv**

\*\* If you are a male, please use the **American male adults’ mean and standard deviation** below; if you are a female please use the **American female adults’ mean and standard deviation** below.

**We are told that the height for all American male adults follows a normal model with a mean () of 175.5 cm and a standard deviation () of 7.4 cm -- or N(175.5, 7.4) – and the height for all American female adults follows a normal distribution with a mean () of 161.8 cm and a standard deviation () of 6.9 cm – or N(161.8, 6.9). These are the population means and population standard deviations.**

1. Using the N(161.8, 6.9) model, what is the **probability** of a *female adult* in the U.S. has a height as 165 cm? Please find out the exact probability using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer**.



46%

1. Using the N(175.5, 7.4) model that describes the distribution of height among all *male* American adults, what **proportion** of American *males* have a height **above** 178.5 cm?

2:5

* 1. Estimate what you think the answer should be using the 68-95-99.7 model. Draw a picture and show all your work. If you draw your picture by hand, take a photo of your picture and paste it here.

**68%**

**Letter

Description automatically generated**

* 1. Now determine the exact proportion using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer.**



1. Using the N(175.5, 7.4) model that describes the distribution of height among all *male* Americans adults, what proportion of American males have a height **lower** than 165.7 cm?
   1. Estimate what you think the answer should be using the 68-95-99.7 model. **Draw a picture and show all your work.** If you draw your picture by hand, take a photo of your picture and paste it here.

**95%**

**Text, letter

Description automatically generated**

* 1. Now determine the exact proportion using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer.**



1. Using the N(161.8, 6.9) model that describes the distribution of height among all American *female* adult, what **proportion** of American female have a height **between** 155 cm **and** 170 cm?
   1. Estimate what you think the answer should be using the 68-95-99.7 model. **Draw a picture and show all your work**. If you draw your picture by hand, take a photo of your picture and paste it here.

**77.4%**

**Letter

Description automatically generated**

* 1. Now determine the exact proportion using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer.**



1. Now go back to **question 10** and figure out the information in the **rightmost two columns** of the table using the population means ( and population standard deviations .
2. What **height** would some **male** have to have in order to have a height SHORTER than 90% of the population of male Americans?
   1. Estimate what you think the answer should be using the 68-95-99.7 model. **Draw a picture and show all your work.** If you draw your picture by hand, take a photo of your picture and paste it here.

**153.3 cm**

**Text, letter

Description automatically generated**

* 1. Now determine the exact height using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer.**



1. What **height** must some **female** have for that height to be TALLER than 90% of the population of American female adults?
   1. Estimate what you think the answer should be using the 68-95-99.7 model. Draw a picture and show all your work. If you draw your picture by hand, take a photo of your picture and paste it here.

**141.1 cm**

**Letter

Description automatically generated**

* 1. Now determine the exact height using proper R functions. **In addition to the answer, please snapshot and paste the codes you use to get the answer.**

