The Influence of Biological Sex on Human Body Part Ratios
How these Ratios Compare Between the Sexes

Kevin A. Black - kevin.black@wsu.edu *Washington State University Vancouver*

In this article we compare the *empirical characteristic function* (**??**) to a *moment-generating-functional form* to compute the proportion of hypotheses m that are rejected under the null hypothesis.

Here is a second paragraph of the abstract (if necessary), and with the pipe notation it doesn't break. Notice it still needs to be indented.

Generally, we write this abstract last. Often it is called the executive summary. It should succinctly summarize the entire document. You can include references such as this one to the Appendices section 5 if necessary.

Keywords: multiple comparisons to control; multivariate chi-square distribution; nonlinear growth curves; Richard's curve; simulated critical points

November 01, 2020

1 Introduction

As a customary portion of the STAT 419 course at Washington State University, the students involved were required to take part in a data collection/manipulation project involving the measurements of various body parts. Following the data collection, professor Monte Shaffer compiled all contributions made by the students into a single, pipe-delimited text file whose explicit content is to remain confidential and only used for the sake of addressing a variety of research questions.

After receiving the data set ($n = 428$) and cleaning up particular observations using a range of methodologies based on what I had in mind for how this data will be used, I determined that I would focus on what kind of influence biological sex has on certain human body part ratios. To supplement my inquiry of this data, I centered my research questions around particular measurements of the human body, such as height, arm span, foot length, distance from elbow to armpit, head height, just to name a few. Research questions pertaining to this study and supportive R code can be found below.

2 How does being male or female influence the ratios between certain body parts?

Making particular deductions throughout history regarding body measurements among males and females is more than likely due to the distinct patterns one can find when comparing the biological sexes. Those born as males or females have about as many differing features to one another than their similar features, both internally and externally. Internally speaking, men typically have deeper voices, a faster metabolism, and can easily build muscle mass, whereas women possess a much more complicated reproductive system, have the ability to breastfeed, and live longer on average (Wolchover 2011). What about externally? How does being male or female influence the ratios between certain body parts? The latter primary question will be elaborated through the following sub-questions, focusing on particular ratios that are likely to yield some insight on the different external measurements between males and females.

2.1 How does 'height' compare to 'arm span' between males and females?

Test

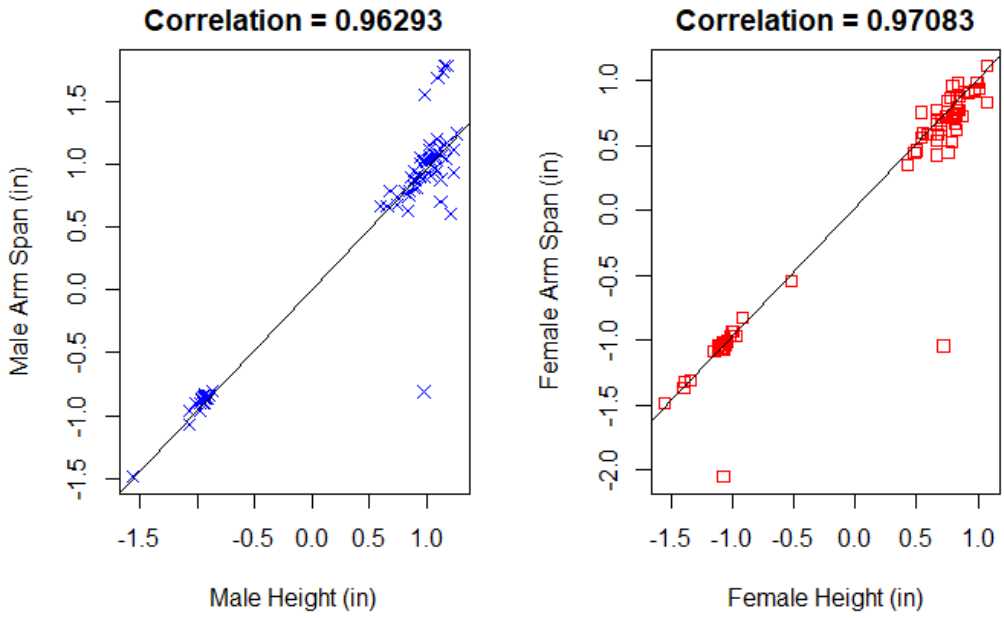


Figure 1: Plots and correlation values for males and females regarding height and arm span.

2.2 How does 'foot length' compare to the length of 'elbow to armpit' between males and females?

2.3 How many average 'head height' lengths are males and females relative to their respective, average 'height'?

3 Data Description

As mentioned in the introduction, the data set utilized throughout this research paper was supplied through the combined efforts of professor Monte Shaffer and the students of the STAT 419 class at Washington State University. Each student was tasked with recording distinct observations of 37 attributes from 10 different people, ideally with an even mix of males and females. For each observation (person), there are measurements of body parts, data collector/respondent identifiers ran through a MD5 hash function, and general information about each respondent. Body part measurements were recorded using body measuring tape in either inches or centimeters, but for the sake of consistency throughout the research paper, all centimeter values were converted to inches and will be treated as inches from here on.

These observations were recorded in early September 2020 and compiled by the instructor for our use in late October 2020. Given that these observations were recorded amid the COVID-19 pandemic, each student was required to make a simple, yet descriptive handout that would detail how one would go about recording their own body measurements and the necessary values to take note of. This would be an ideal situation of how observations were recorded and sent electronically by each respondent, but observations could also be taken in person given that the surveyor and respondent were comfortable being in close proximity of each other. An example of a two-page handout created by Kevin Black, as well as further information regarding the attributes of this data set, can be found in Appendix 5.1.2 and Appendix 5.2, respectively.

On the surface, the purpose for writing this research paper and collecting the necessary data can simply be attributed to project requirements for a university course. However, the deeper reasoning behind why this research paper was composed the first place was to give students a more thorough understanding of the data analytics process. More specifically, how to not just work with the data, but how to understand the data and derive effective questions, how to test for patterns and conclusions in a statistical, analytic environment, and how to exercise data provenance practices. This process will be very similar between all data focused projects in a data analyst's career, so this is a good starting point for garnering crucial experience.

4 Key Findings

5 APPENDICES

5.1 Data Provenance

5.1.1 Utilization of Data Provenance

While it could be seen as an application mainly used in large-scale data analytics projects, the multitude of steps to practice data provenance was also used to handle the vulnerable data used in this research. Auxiliary files such as the R project files can be found on the project repository through GitHub, while the specific functions used in this report can be found in the `functions-project-measure.R` file or in section 5.3. The data set itself was not saved to any location online due to privacy concerns, but it was organized and saved onto a local hard drive for immediate use.

The collection and general organization process for the data set is thoroughly described in sections 3 and 5.2. Cleaning for the actual substance of the data set was performed through the use of the `prepareMeasureData()` function, where the data set was manipulated to more better fit the aims of addressing the particular research questions for this report. Cleaning this data set involved multiple steps, such as: omitting rows containing NA values based solely on the body measurement columns, setting a consistent naming convention for gender and units, converting all body measurement values to inches, and scaling the body measurement data. Keeping the research questions in mind and despite the cleaning of all columns, the only body measurement fields utilized were *height*, *arm.span*, *foot.length*, *elbow.armpit*, and *head.height*.

Prior to `prepareMeasureData()`, the data is imported via the `read.file()` function where it checks to see if a cleaned version of the data set already exists, to which that particular file would take precedence and be used. If no clean version currently exists, the function goes through the cleaning process and saves the cleaned data set to the same directory as original data set. For unbiased samples of the data frame, a seed was set for reproducibility and 200 random observations were drawn without replacement.

The documented R code can be found in section 5.3. Key findings and visualizations of summary statistics were discussed previously throughout sections 2 through 4. Based on the points made in this section, its clear that data provenance was kept in mind for making sure there was a traceable history in the usage of this data set, whether to resolve potential issues or to cut down on access times.

5.1.2 Data Collection Handout

Figure 2: Handout Page 1

PROJECT 1: "HANDOUT"

This handout is a prelude to a data collection project comprised of having 10 individuals record numerous measurements around their body (measured in **centimeters**), along with general information that does not require measuring. Your name will be encrypted using a hash function known as MD5, so your measurements and identifying data will remain confidential throughout the project. Below, you will find the various measurements you need to take, which could take 20+ minutes. Covariates labeled with (left, right) means you will need to measure both the left and right appendages for that field, separated by a comma. Diagrams are shown on the next page, which will show you how to take the measurements yourself for each field and are annotated by the abbreviations listed in the table below.

AREA FOR RESPONDENT

COVARIATE	***MEASUREMENT***
Name (first name, last initial)	
Dominant writing hand	
Dominant eye for seeing	
Eye color	
Dominant swinging hand	
Age	
Gender	
Ethnicity	
Height *Standing height of the individual with no shoes on. H on diagram.*	
Head height *Height from the top of the head to below the chin. HH on diagram.*	
Head circumference Distance around the head, measured above the eyes/ears. HC on diagram.	
Hand length (left, right) *Length of the hand from the middle finger to the wrist (just below the palm). HL on diagram.*	
Hand width (left, right) *Width of a fully stretched hand from the pinky finger to the thumb. HW on diagram.*	
Hand to elbow (left, right) *Length from the middle finger to the elbow. HE on diagram.*	
Elbow to armpit (left, right) *Length from the elbow to the arm pit. EA on diagram.*	
Arm reach (left, right) *Length from floor to the extended arm maximum point, standing flatfooted. AR on diagram.*	
Arm span *Length from each middle finger, with fully extended arms. AS on diagram.*	
Foot length (left, right) *Length of the foot from the largest toe to the back of the heel. FL on diagram.*	
Floor to knee pit (left, right) *Distance from the floor to the knee pit. FK on diagram.*	
Floor to hip (left, right) *Distance from the floor to the hip. FH on diagram.*	
Floor to navel *Distance from the floor to the navel (belly button). FN on diagram.*	
Floor to armpit (left, right) *Distance from the floor to the arm pit. FA on diagram.*	

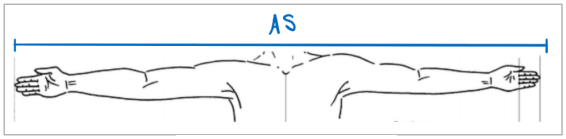
Kevin Black – kevin.black@wsu.edu

Figure 3: Handout Page 2

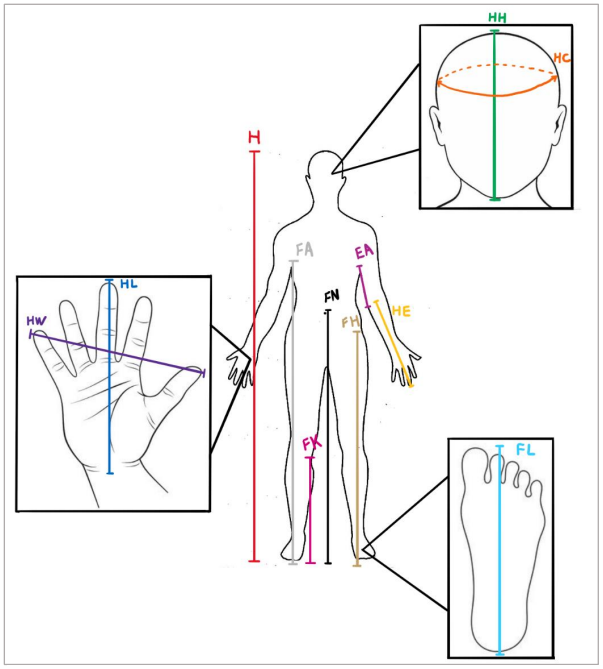
PROJECT 1: "HANDOUT"

DATA COLLECTOR USE ONLY

COVARIATE	*EVALUATION*
Quality (1-10)	
Minutes	
To be completed later?	
Notes	



How to measure arm span.



How to measure height, head height, head circumference, hand length, hand width, hand to elbow, elbow to armpit, foot length, floor to knee pit, floor to hip, floor to navel, and floor to armpit.

How to measure arm reach.

Kevin Black – kevin.black@wsu.edu

5.2 Data Set Explained

In addition to how the data set was described in Section 3, explanations regarding each attribute can be found below. After the collaborative effort of all students having their data compiled, the data set ended up having 428 total observations. However, the data set was filled with an enormous amount of NA values in the body measurements, potentially due to the time constraints of some students or lack of attempt to fill out all fields. As a result, running the function `complete.cases()` on the data set during the data cleaning process returned a data frame containing only 262 observations, about 61.21% of the original data set size. `complete.cases()` was used over `na.omit()` because the latter function would omit all rows containing NA values based on all columns, including the rows where the non-body measurement attributes had NA values, whereas the former function would omit NA values only for the specified body measurement columns.

Figure 4: Description of Each Field

Field	Description
data_collector	MD5 encryption of surveyor identifier.
person_id	MD5 encryption of respondent identifier.
side	Body part side being measured; left, right, or NA.
height	Standing height of the individual with no shoes on.
head.height	Height from the top of the head to below the chin.
head.circumference	Distance around the head, measured above the eyes/ears.
hand.length	Length of the hand from the middle finger to the wrist (just below the palm).
hand.width	Width of a fully stretched hand from the pinky finger to the thumb.
hand.elbow	Length from the middle finger to the elbow.
elbow.armpit	Length from the elbow to the arm pit.
arm.reach	Length from floor to the extended arm maximum point, standing flatfooted.
arm.span	Length from one middle finger to the other, with fully extended arms.
foot.length	Length of the foot from the largest toe to the back of the heel.
floor.kneepit	Distance from the floor to the knee pit.
floor.hip	Distance from the floor to the hip.
floor.navel	Distance from the floor to the navel (belly button).
floor.armpit	Distance from the floor to the arm pit.
units	Units used in measuring body parts; centimeters or inches.
writing	Dominant hand of respondent; left or right.
eye	Dominant eye of respondent; left, right, or both.
eye_color	Eye color of respondent.
swinging	Swinging hand or respondent; left or right.
age	Age of respondent.
gender	Specified gender of respondent; male or female (non-binary in one case).
quality	Quality of testing decided by surveyor; on a scale of 1-10.
minutes	Minutes spent performing measurements.
ethnicity	Ethnicity of respondent.
notes	Supplementary notes pertaining to observation.

5.3 R Code Used for Research

Below are the necessary functions required to produce the figures and findings referenced in this report.

```
prepareMeasureData = function(measure){
  # Cleaning: omit NA rows based measured values, not on $side
  measure = measure[complete.cases(measure[,4:26]),]

  # Cleaning: consistent naming convention
  measure$gender = factor(tolower(measure$gender))
  measure$gender[measure$gender=="f"] = "female"
  measure$gender[measure$gender=="m"] = "male"
  measure$units = factor(tolower(measure$units))
  measure$units[measure$units=="inches"] = "in"
  measure$units[measure$units=="inch"] = "in"
  measure$units[measure$units=="\"in\""] = "in"
  measure$units[measure$units=="cm"] = "in"

  # Converting cm to inches
  for(row in 1:nrow(measure)){
    if(measure[row,]$units=="cm"){
      measure[row,4:26] <- measure[row,4:26]/2.54
    }
  }

  # Scale data
  measure[,4:26] <- scale(measure[,4:26])

  return(measure)
}

read.file = function(path){
  tryCatch(
    expr = {
      # Open cleaned file if already available
      measure = utils::read.csv(paste0(path.to.secret,"measure-clean.txt"),
                                header = TRUE, quote = "", sep = "|")
      return(measure)
    },
    warning = function(w){
      # If no clean file, open original file
      measure = utils::read.csv(paste0(path.to.secret,"measure-students.txt"),
                                header = TRUE, quote = "", sep = "|")
      # Clean data
      measure = prepareMeasureData(measure)

      # Save cleaned data for later
      write.table(measure,paste0(path.to.secret,"measure-clean.txt"),sep="|",quote=FALSE)
      return(measure)
    }
  )
}
```

Below are the libraries and R code used to source functions, import measurement data, and perform analysis on said data.

```r
# Import necessary libraries
library(stats) # For cor()
library(devtools) # For source_url()

# Source cleaning function
path.hub = "https://raw.githubusercontent.com/KevnBlack/WSU_STATS419_FALL2020/"
source_url(paste0(path.hub,"master/functions/functions-project-measure.R"))
path.to.secret = "D:/School/Fall 2020/STAT 419/datasets/"
measure.df = read.file(path.to.secret) # Import data
set.seed(1) # Sample 200 observations from data frame
measure.sample = measure.df[sample(nrow(measure.df),200),]

# Isolate male and female data
measure.df.m = measure.sample[measure.sample$gender == "male",]
measure.df.f = measure.sample[measure.sample$gender == "female",]
cor.m = cor(measure.df.m$height.NA, measure.df.m$arm.span.NA)
cor.f = cor(measure.df.f$height.NA, measure.df.f$arm.span.NA)

# Sub-question 1
## Trend lines for plots
lm.m = lm(measure.df.m$height.NA ~ measure.df.m$arm.span.NA)
lm.f = lm(measure.df.f$height.NA ~ measure.df.f$arm.span.NA)

## Correlation Graphs
par(mfrow = c(1,2))
plot(measure.df.m$height.NA, measure.df.m$arm.span.NA,
     main = paste("Correlation =", round(cor.m,5)), pch = 4,
     xlab = "Male Height (in)", ylab = "Male Arm Span (in)", col = "blue")
abline(lm.m)

plot(measure.df.f$height.NA, measure.df.f$arm.span.NA,
     main = paste("Correlation =", round(cor.f,5)), pch = 0,
     xlab = "Female Height (in)", ylab = "Female Arm Span (in)", col = "red")
abline(lm.f)
```

```r
# Sub-question 2

# Sub-question 3
```

ENDNOTES

REFERENCES

Wolchover, Natalie. 2011. Men vs. women: Our key physical differences explained .

TABLE OF CONTENTS