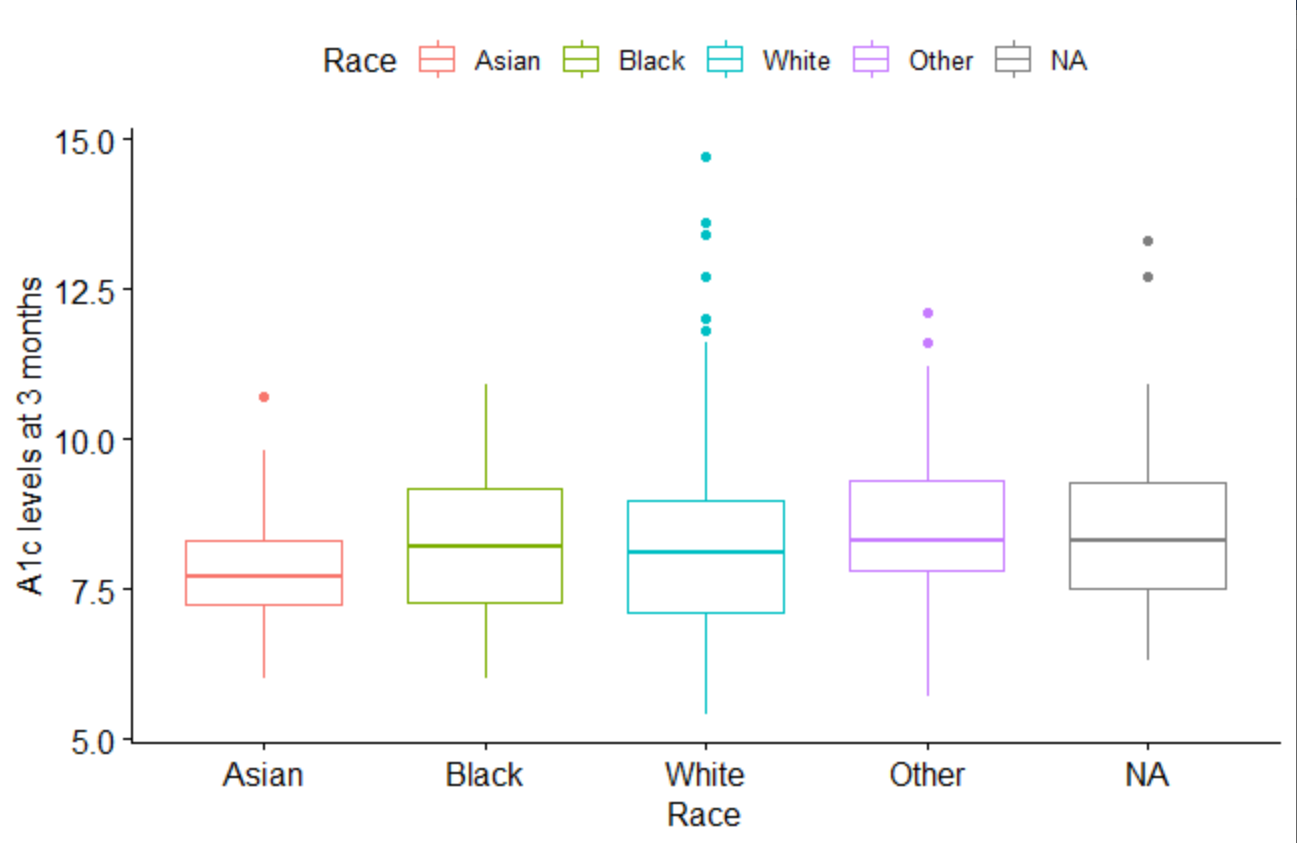
**Introduction:** Genetics can have an impact on the efficacy of a treatment. Our study aims to determine whether in a certain diabetes treatment study, the mean levels of hemoglobin A1c (HbA1c) measured after three months varied amongst different racial groups that participated in this study.

**Methods:** The quantitative (HbA1c) levels are reported in this study for each racial group. The sample size for Asian is n=27, Black is n=44, White is n=166, and Other is n=25. The mean (HbA1c) levels after three months in the study is reported for all races, along with standard deviation, and their confidence intervals (CI) of 95%. A graphically computed visual boxplot is created for each racial group to assess their distribution’s departure from normality. To conduct the correct mean test, equality of variances is first tested via Bartlett Test, with H01: Evidence does not suggest variances are different between groups, and HA1: Variance between groups is present. We shall test this at a significance level of α = 0.05. If α > 0.05, Null hypothesis (H01) is retained, but if α < 0.05, Null hypothesis is rejected in favor of the alternative hypothesis (HA1). Consequently, this will determine the use of either the unequal-variance ANOVA or equal-variance ANOVA for the mean test. One of these tests will be used to determine whether the means (statistically significantly) differ between each racial group. For our key mean test, our H02: μ1 = μ2 = μ3 = μ4, and HA2: The evidence suggests that at least two means differ between the groups. We shall test this at a significance level of α = 0.05. If α > 0.05, Null hypothesis (H02) is retained, but if α < 0.05, Null hypothesis is rejected in favor of the alternative hypothesis (HA2). The R statistical software is used for all measurement analysis.

To visually assess normality, the following function was used: *ggboxplot(data\_work, x = "Race", y = "A1c.value.3mo.post.DEAP", color = "Race", order = c("Asian", "Black", "White", "Other"), ylab = "A1c levels at 3 months", xlab = "Race").* To test the variance equality, the following function was used: *bartlett.test(data\_work$A1c.value.3mo.post.DEAP~data\_work$Race)*. For summarization, the following function is used*: summary\_IO\_1 <-summarize(data\_work[data\_work$Race=="",], sum(!is.na(A1c.value.3mo.post.DEAP)),mean(A1c.value.3mo.post.DEAP,na.rm=TRUE),sd(A1c.value.3mo.post.DEAP,na.rm=TRUE))print(summary\_IO\_1).*  For Confidence interval, the following function is used: *CI\_95\_IO <- t.test(data\_work$A1c.value.3mo.post.DEAP [data\_work$Race=="Black"],conf.level = 0.95*)print(CI\_95\_IO). For our main hypothesis test, the following function is used: *uneq\_var <- oneway.test(A1c.value.3mo.post.DEAP ~ Race, data=data\_work)print(uneq\_var).*

**Results:**

The box plot (Figure 1) shows that the distribution for each individual group appears to be relatively symmetric, because the medians are well-centered, and the ranges and quartiles are evenly distanced, so the Central Limit Theorem applies. The data for each group is summarized in Figure 2 and does not appear to show any significant difference in the means between any groups. The result from the Bartlett Test (K-squared = 10.358, df = 3, p-value = 0.01576) shows that that the p-value is less than 0.05, which tells us that that that variance between groups are present. Because of the normal distribution and unequal variances, the unequal-variance ANOVA test was performed for the mean hypothesis test and the results (F = 1.3611, num df = 3.00, denom df = 66.14, p-value = 0.2623) showed that the p-value is greater than 0.05, which inclines us to retain our final null hypothesis (H02), which states that there is no evidence suggesting any of the means differ between the racial groups. Hence, a multiple comparisons analysis is not necessary.

Figure 1: Box and wisker plot showing distribution of A1c Levels at 3 months for each racial Group.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Sample Size (n) | Mean A1c levels | Standard Deviation | Confidence interval (CI) |
| Asian | 27 | 7.951852 | 1.068561 | 7.529143, 8.374561 |
| Black | 44 | 8.261364 | 1.203525 | 7.895459, 8.627269 |
| White | 166 | 8.183735 | 1.618877 | 7.935647, 8.431823 |
| Other | 25 | 8.688 | 1.519408 | 8.060819, 9.315181 |

Figure 2: Data summarization table for each race group.

**Discussion:** Our statistical analysis showed that there is no evidence suggesting any of the mean HbA1c levels differ between the racial groups. Therefore, clinicians may not have to factor in race when determining this treatment’s scope of effectiveness. However, further studies should be done to confirm this conclusion, such as comparing mean HbA1c levels among racial groups over longer periods of time.

Biostats\_HW\_4\_R.R

2020-10-22

setwd("C:/Users/name\_/OneDrive/Desktop/Biostats\_543")  
getwd()

## [1] "C:/Users/ame\_/OneDrive/Desktop/Biostats\_543"

list.files()

## [1] "Name\_Covid\_19\_Script\_Assignment\_RStudio.R"  
## [2] "A1c\_data\_clean\_CSV.csv"   
## [3] "A1c\_data\_clean2\_HW\_3.csv"   
## [4] "A1c\_data\_clean2\_HW\_4.csv"   
## [5] "Biostats\_HW\_2\_RStudio.docx"   
## [6] "Biostats\_HW\_2\_RStudio.R"   
## [7] "Biostats\_HW\_4\_R.R"   
## [8] "Biostats\_HW\_4\_R.spin.R"   
## [9] "Biostats\_HW\_4\_R.spin.Rmd"   
## [10] "Covid\_19\_Script\_Assignment.docx"   
## [11] "Covid\_19\_Script\_Assignment.R"

data\_raw <- read.csv("A1c\_data\_clean2\_HW\_4.csv")  
View(data\_raw)  
data\_work <- data\_raw  
View(data\_work)  
  
library(dplyr) #calling "dplyr" package so that we can use summarize() function

##   
## 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

library(car) # for qqPlot() function

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

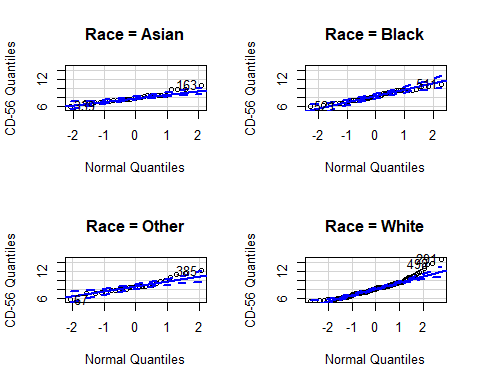
library(ggpubr) #for obtaining multi-group box-plots

## Warning: package 'ggpubr' was built under R version 4.0.3

## Loading required package: ggplot2

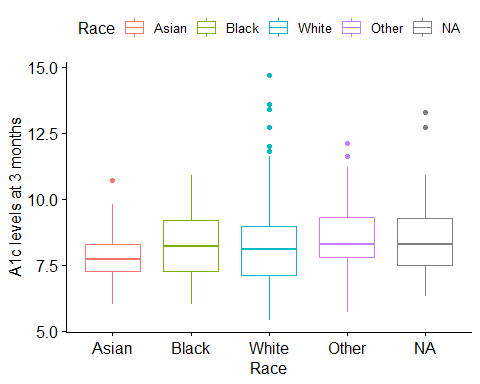
## Warning: package 'ggplot2' was built under R version 4.0.3

qqPlot(A1c.value.3mo.post.DEAP~Race,  
 main="QQ Plot of Day-30 CD-56 by Group",  
 xlab="Normal Quantiles",  
 ylab="CD-56 Quantiles",  
 data=data\_work)



ggboxplot(data\_work, x = "Race", y = "A1c.value.3mo.post.DEAP",   
 color = "Race",  
 order = c("Asian", "Black", "White", "Other"),  
 ylab = "A1c levels at 3 months", xlab = "Race")

## Warning: Removed 276 rows containing non-finite values (stat\_boxplot).



table(data\_work$Race)

##   
## Asian Black Other White   
## 46 77 56 296

#Summarize for "Asian"  
summary\_IY\_1 <- summarize(data\_work[data\_work$Race=="Asian",],sum(!is.na(A1c.value.3mo.post.DEAP)),mean(A1c.value.3mo.post.DEAP,na.rm=TRUE),sd(A1c.value.3mo.post.DEAP,na.rm=TRUE))  
print(summary\_IY\_1)

## sum(!is.na(A1c.value.3mo.post.DEAP))  
## 1 27  
## mean(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 7.951852  
## sd(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 1.068561

CI\_95\_IY <- t.test(data\_work$A1c.value.3mo.post.DEAP[data\_work$Race=="Asian"],conf.level = 0.95)   
print(CI\_95\_IY)

##   
## One Sample t-test  
##   
## data: data\_work$A1c.value.3mo.post.DEAP[data\_work$Race == "Asian"]  
## t = 38.668, df = 26, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 7.529143 8.374561  
## sample estimates:  
## mean of x   
## 7.951852

#Summarize for "Black"  
summary\_IO\_1 <- summarize(data\_work[data\_work$Race=="Black",],sum(!is.na(A1c.value.3mo.post.DEAP)),mean(A1c.value.3mo.post.DEAP,na.rm=TRUE),sd(A1c.value.3mo.post.DEAP,na.rm=TRUE))  
print(summary\_IO\_1)

## sum(!is.na(A1c.value.3mo.post.DEAP))  
## 1 44  
## mean(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 8.261364  
## sd(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 1.203525

CI\_95\_IO <- t.test(data\_work$A1c.value.3mo.post.DEAP[data\_work$Race=="Black"],conf.level = 0.95)   
print(CI\_95\_IO)

##   
## One Sample t-test  
##   
## data: data\_work$A1c.value.3mo.post.DEAP[data\_work$Race == "Black"]  
## t = 45.533, df = 43, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 7.895459 8.627269  
## sample estimates:  
## mean of x   
## 8.261364

#Summarize for "White"  
summary\_CY\_1 <- summarize(data\_work[data\_work$Race=="White",],sum(!is.na(A1c.value.3mo.post.DEAP)),mean(A1c.value.3mo.post.DEAP,na.rm=TRUE),sd(A1c.value.3mo.post.DEAP,na.rm=TRUE))  
print(summary\_CY\_1)

## sum(!is.na(A1c.value.3mo.post.DEAP))  
## 1 166  
## mean(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 8.183735  
## sd(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 1.618877

CI\_95\_CY <- t.test(data\_work$A1c.value.3mo.post.DEAP[data\_work$Race=="White"],conf.level = 0.95)   
print(CI\_95\_CY)

##   
## One Sample t-test  
##   
## data: data\_work$A1c.value.3mo.post.DEAP[data\_work$Race == "White"]  
## t = 65.132, df = 165, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 7.935647 8.431823  
## sample estimates:  
## mean of x   
## 8.183735

#Summarize for "Other"  
summary\_CO\_1 <- summarize(data\_work[data\_work$Race=="Other",],sum(!is.na(A1c.value.3mo.post.DEAP)),mean(A1c.value.3mo.post.DEAP,na.rm=TRUE),sd(A1c.value.3mo.post.DEAP,na.rm=TRUE))  
print(summary\_CO\_1)

## sum(!is.na(A1c.value.3mo.post.DEAP))  
## 1 25  
## mean(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 8.688  
## sd(A1c.value.3mo.post.DEAP, na.rm = TRUE)  
## 1 1.519408

CI\_95\_CO <- t.test(data\_work$A1c.value.3mo.post.DEAP[data\_work$Race=="Other"],conf.level = 0.95)   
print(CI\_95\_CO)

##   
## One Sample t-test  
##   
## data: data\_work$A1c.value.3mo.post.DEAP[data\_work$Race == "Other"]  
## t = 28.59, df = 24, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 8.060819 9.315181  
## sample estimates:  
## mean of x   
## 8.688

bartlett.test(data\_work$A1c.value.3mo.post.DEAP~data\_work$Race)

##   
## Bartlett test of homogeneity of variances  
##   
## data: data\_work$A1c.value.3mo.post.DEAP by data\_work$Race  
## Bartlett's K-squared = 10.358, df = 3, p-value = 0.01576

uneq\_var <- oneway.test(A1c.value.3mo.post.DEAP ~ Race, data=data\_work)  
print(uneq\_var)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: A1c.value.3mo.post.DEAP and Race  
## F = 1.3611, num df = 3.00, denom df = 66.14, p-value = 0.2623