Alexandrea Stylianou
Stats 543
Final Exam
Question 1
Grade sheet for Final Exam Problem One

Anova Test For Vitamin D Levels Among Four Age Groups

<u>Topic:</u>
Vitamin D is a fat soluble vitamin that is responsible for enhancing absorption of calcium,iron, magnesium, phosphate, and zinc. Vitamin D can be found in certain foods, or ingested as a supplement. However, synthesis of Vitamin D in the skin from sunlight is the major natural source of the Vitamin D. It was of interest to test the dietary intake of Vitamin D among 4 different age groups of healthy females.

<u>Population:</u>
Women who receive Vitamin D from dietary intake

<u>Research Question:</u>
Whether or or not the means of dietary intake of Vitamin D is equal among 4 different age groups of healthy women.

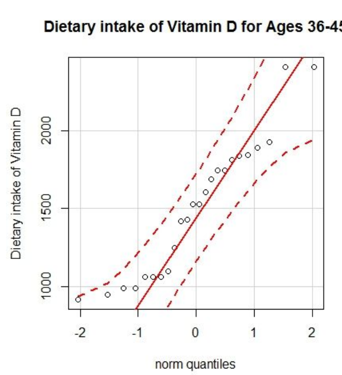
Methods

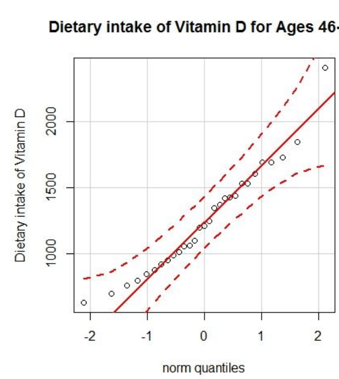
<u>Description of Outcome:</u> Outcome variable is the dietary intake of Vitamin D
<u>Description of Predictor:</u> Predictor variable is the age group of the women

<u>Description of Group:</u>
Group A: 25-35
Group B: 36-45
Group C: 46-55
Group D: 56- 66

<u>Description of Data Summary for Each Group:</u>
The sample is divided into four research groups, by ranges of age among healthy women and their Vitamin D intake

<u>Verification of Normality</u>
By the central limit theorem, if our sample size is above 30 the sample mean will be normally distributed. If the sample size is less than 30 we can assume normality by observation of the qqPlot for groups age 36-45 and 46-55:





Statement of Null Hypothesis:
There is no mean differences of the dietary intake of Vitamin D among four groups of healthy women

Statement of Alternative Hypothesis:
There is at least one mean difference of the dietary intake of Vitamin D among four groups of healthy women

Statistical Method for testing for testing the difference in means:
Anova F Test

Statistical Method for testing for testing equality of variances:
Bartlett Test

Method for Multiple comparisons:
Tukey Test

Decision Rule:
Reject H_O in favor of H_A if p-value is less than alpha (p-value<α) otherwise fail to reject the null H_o

Method of Computation:
R statistical software version 2.11.1

Significance Level:
α=0.05

Results

Data Summary:
Summary Table for Dietary Intake of Women Among 4 Age Groups

Group	*n*	Mean	Standard Deviation	95% CI
A	33	1482.2	479.15	1312.2510, 1652.0520
B	24	1508.2	439.79	1322.4610, 1693.8720

C	29	1254.5	403.88	1100.8530, 1408.1120
D	54	971.2	356.19	873.9835, 1068.4239

Normality is assumed per the CLT based on adequate sample size for each group, or the qqPlot

Sum of squares for Equal-Variance Case of Vitamin D intake among Healthy Women

Source	DF	Sum of Squares	Mean of Squares	F Ratio	p-value
Group	3	7567827	2522609	14.86	<0.0001
Error	136	23086764	169756		
Total	139	30,654,591			

Assumptions:
Sample is representative
Measurements of subjects are independent of each other
Measurements of each group is independent of each other
Sample is either large enough, or normality is assumed per observation of the qqPlot
Assume that the variances are equal based on the Bartlett test

Equality of the Variances:
Bartlett test p-value: 0.2798

Test Results:
The p-value for the Anova F test is less than the stated alpha level

Decision:
Since p-value < 0.05, we reject the H_O in favor of H_A

Mean Comparisons using Unadjusted Independent T-Test of dietary intake of Vitamin D of Healthy Women

Comparison	Difference	SE	95% CI	p-value
B-A	26.0152	122.540	-261.4867 313.5170	0.9953
B-C	253.6839	116.98	-42.0490 549.4165	0.1200
A-C	227.6688	112.17	-45.1070, 500.4445	0.1366
B-D	536.9630	102.02	274.0508, 799.8751	<0.0001

A-D	510.9478	96.47	274.1533, 747.7423	<0.0001
C-D	283.2791	89.30	36.55643, 530.0017	0.0174

Discussion

Evidence suggests that the greatest difference of means in Vitamin D intake of healthy women is between age groups of 36-45 and 56-66. Therefore, we can conclude that based on the data there is at least one difference in means between the two subject groups.Going forward, with this information health care professionals can appropriately assess the need for supplemental Vitamin D based on the age group of the female patient.

R-code

```
T3_Problem1<- read.csv(file.choose(),header=TRUE)
At<-T3_Problem1[T3_Problem1$Group=='A' & T3_Problem1$VitaminD,2]
Bt<-T3_Problem1[T3_Problem1$Group=='B' & T3_Problem1$VitaminD,2]
Ct<-T3_Problem1[T3_Problem1$Group=='C' & T3_Problem1$VitaminD,2]
Dt<-T3_Problem1[T3_Problem1$Group=='D' & T3_Problem1$VitaminD,2]
AT<-c(At) #33
BT<-c(Bt) #24 #small sample size, but normal qq plot
CT<-c(Ct) #29 #small sample size, but normal qq plot
DT<-c(Dt) #54
```

```
qqPlot(AT)
qqPlot(BT)
qqPlot(CT)
qqPlot(DT)
##############################CI#############################
CI(AT, ci=0.95)
#upper     mean    lower
#1652.052 1482.152 1312.251
CI(BT, ci=0.95)
#upper     mean    lower
#1693.872 1508.167 1322.46
CI(CT, ci=0.95)
#upper     mean    lower
#1408.112 1254.483 1100.853
CI(DT, ci=0.95)
#upper      mean     lower
#1068.4239  971.2037  873.9835


######################SD of Groups#######################
ddply(T3_Problem1, .(Group), summarize, VitaminD=sd(VitaminD))
#Group VitaminD
#1     A 479.1529
#2     B 439.7865
#3     C 403.8845
#4     D 356.1865
##############################bartlett test###########################
AlabelT <- rep("A", 33)
BlabelT <- rep("B", 24)
ClabelT <- rep("C", 29)
DlabelT <- rep("D", 54)
TreatT <- c(AlabelT, BlabelT, ClabelT,DlabelT)
ScoreT <- c(AT,BT,CT,DT)
bartlett.test(ScoreT~TreatT)
#Bartlett test of homogeneity of variances

#data:  ScoreT by TreatT
#Bartlett's K-squared = 3.8357, df = 3, p-value = 0.2798

#Decision: Uneuqal variances
####################################anova##############################
data1 <- data.frame(Treatment = TreatT, Score = ScoreT)

Score.aov <- aov( Score ~ Treatment, data=data1)
summary(Score.aov, order=TRUE)
#              Df   Sum Sq Mean Sq F value   Pr(>F)
#Treatment     3  7567827 2522609   14.86 2.01e-08 ***
#  Residuals   136 23086764  169756
#---
#  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
#########################################tukey###########################
Score.aov <- aov( Score ~ Treatment, data=data1)
TukeyHSD(Score.aov, order=TRUE)
#Fit: aov(formula = Score ~ Treatment, data = data1)

#$Treatment
#          diff       lwr       upr     p adj
#B-A   26.01515 -261.4867  313.51701 0.9953899
#C-A -227.66876 -500.4445   45.10695 0.1366191
#D-A -510.94781 -747.7423 -274.15332 0.0000006
#C-B -253.68391 -549.4165   42.04865 0.1200334
#D-B -536.96296 -799.8751 -274.05084 0.0000026
#D-C -283.27905 -530.0017  -36.55643 0.0174253
###########################################s. error###########################
B_A_standard_error_diff<-sqrt((439.7865^2/24) + (479.1529^2/33))
B_C_standard_error_diff<-sqrt((439.7865^2/24) + (403.8845^2/29))
A_C_standard_error_diff<-sqrt((479.1529^2/33) + (403.8845^2/29))
B_D_standard_error_diff<-sqrt((439.7865^2/24) + (356.1865^2/54))
A_D_standard_error_diff<-sqrt((479.1529^2/33) + (356.1865^2/54))
C_D_standard_error_diff<-sqrt((403.8845^2/29) + (356.1865^2/54))

B_A_standard_error_diff
B_C_standard_error_diff
A_C_standard_error_diff
B_D_standard_error_diff
A_D_standard_error_diff
C_D_standard_error_diff


#1     A 479.1529
#2     B 439.7865
#3     C 403.8845
#4     D 356.1865

std.error(BT,AT)
std.error(CT,AT)
std.error(DT,AT)
std.error(CT,BT)
std.error(DT,BT)
std.error(DT,CT)
#> std.error(BT,AT)
#[1] 89.77104
#> std.error(CT,AT)
#[1] 74.99947
#> std.error(DT,AT)
#[1] 48.47085
#> std.error(CT,BT)
#[1] 74.99947
#> std.error(DT,BT)
#[1] 48.47085
```

```
#> std.error(DT,CT)
#[1] 48.47085
```