Homework 2

Jackson Connor

Packages

```
library(tidyverse)
library(multcomp)
library(emmeans)
library(car)
```

Problem 1

A pharmaceutical manufacturer wants to investigate the bioactivity of a new drug. A completely randomized single-factor experiment was conducted with three dosage levels, and the following results were obtained.

Complete the ANOVA Table: ANOVA Table

	df	SS	MS	F-value	P-value
Bet_Dosage	2	450.7	225.33	7.036	0.0145
Error	9	288.2	32.03		
Total	11	738.9			

```
Observations=c(24,28,37,30,37,44,31,35,42,47,52,38)

Dosage=c(20,20,20,20,30,30,30,30,40,40,40,40)

data1<-data.frame(Observations, Dosage)

str(data1)
```

```
'data.frame':    12 obs. of  2 variables:
 $ Observations: num  24 28 37 30 37 44 31 35 42 47 ...
 $ Dosage      : num  20 20 20 20 30 30 30 30 40 40 ...
```

```
Dosage = factor(Dosage)

model1 <- aov(Observations ~ Dosage)

summary(model1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Dosage       2  450.7  225.33   7.036 0.0145 *
Residuals    9  288.2   32.03
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. Is there evidence to indicate that dosage level affects bioactivity?

A p-value of $0.0145 < 0.05$ supports the claim that dosage level affects bioactivity

b. Compute a 95 percent interval estimate of the mean of dosage level is 30g.

```
data1 |>
  group_by(Dosage) |>
  summarize(mean = mean(Observations))
```

```
# A tibble: 3 x 2
  Dosage  mean
   <dbl> <dbl>
1     20  29.8
2     30  36.8
3     40  44.8
```

$$\hat{L} = \frac{37+44+31+35}{4} = 36.75$$

$$\hat{V}(\hat{L}) = \sqrt{MSE * \sum_{i=1}^{a} \frac{C_i^2}{n_i}}$$

$$\sum_{i=1}^{a} \frac{C_i^2}{n_i} = \frac{1}{4}$$

```
qt(0.975,9)
```

[1] 2.262157

$$36.75 \pm (CV * \sqrt{MSE * \sum_{i=1}^{a} \frac{C_i^2}{n_i}}) = 36.75 \pm (2.262 * \sqrt{32 * \tfrac{1}{4}}) = 36.75 \pm 6.3979$$

(30.352, 43.148)

c. Compute a 99 percent interval estimate of the mean difference between dosages levels 20g and 40g.

$$\hat{L} = \frac{24+28+37+30}{4} - \frac{42+47+52+38}{4} = 29.75 - 44.75 = -15$$

$$\sum_{i=1}^{a} \frac{C_i^2}{n_i} = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{2}{4}$$

```
qt(0.995,9)
```

[1] 3.249836

$$-15 \pm (3.2498 * \sqrt{32 * \tfrac{2}{4}}) = -15 \pm 12.9992$$

(-27.9992, -2.0008)

d. Compute simultaneous 95% CIs for the difference in means for dosage levels 20g and 40g, based on Tukey's, Bonferroni's and Scheffe's Methods:

Bonferroni

$$t_{(1-\frac{\alpha}{2*g}),df_{error}} = \text{Bonferroni's CV}$$

$$g = \frac{a(a-1)}{2} = \frac{3(2)}{2} = 3$$

```
qt(1-(0.05/(2*3)), 9)
```

[1] 2.933324

$$\hat{L} \pm B * \sqrt{\hat{V}(\hat{L})} = \text{-15} \pm (2.933 * \sqrt{16}) = -15 \pm 11.732$$

(-26.732, -3.268)

Tukey's

3

```
qtukey(0.95, 3, 9)
```

[1] 3.948492

$Q = 3.948/\sqrt{2} = 2.792$

$HSD = Q * \sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})} = 2.792 * \sqrt{16} = 11.167$

$\hat{L} \pm HSD = -15 \pm 11.167$

(-26.167, -3.833)

Scheffe

$\sqrt{(a-1)f_{(1-\alpha),(a-1),df_{error}}} = $ Scheffe's CV

```
qf(0.95, 2, 9)
```

[1] 4.256495

$\sqrt{2 * 4.256} = 2.918$

$\hat{L} \pm S * \sqrt{\hat{V}(\hat{L})} = -15 \pm (2.918 * \sqrt{16}) = -15 \pm 11.672$

(-26.672, -3.328)

e. Use Tukey's HSD to compare pairs of treatment means.

	Difference	Significant (>HSD of 11.167)
20g-30g = |29.75-36.75|	7	N
20g-40g = |29.75-44.75|	15	Y
30g-40g = |36.75-44.75|	8	N

Problem 2

The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted, and the following data were collected.

You are given the following information:

$\bar{Y}_{(1.)}=2971.00, s(\bar{Y}_{(1.)})=56.65$ and MSE$=12825.7$

$\bar{Y}_{(2.)}=3156.25, s(\bar{Y}_{(2.)})=56.65$

$\bar{Y}_{(3.)}=2933.75, s(\bar{Y}_{(3.)})=56.65$

$\bar{Y}_{(4.)}=2666.25, s(\bar{Y}_{(4.)})=56.65$

a. Test the hypothesis that mixing techniques affect the strength of the cement.

```
Tensile_Strength=c(3129,3000,2865,2890,3200,3300,2975,3150,2800,2900,2985,3050,
                   2600,2700,2600,2765)

Mixing_Technique=c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)

data<-data.frame(Tensile_Strength,Mixing_Technique)

str(data)
```

```
'data.frame':   16 obs. of  2 variables:
 $ Tensile_Strength: num  3129 3000 2865 2890 3200 ...
 $ Mixing_Technique: num  1 1 1 1 2 2 2 2 3 3 ...
```

```
Mixing_Technique = factor(Mixing_Technique)

model <- aov(Tensile_Strength ~ Mixing_Technique)

summary(model)
```

```
                 Df Sum Sq Mean Sq F value   Pr(>F)
Mixing_Technique  3 489740  163247   12.73 0.000489 ***
Residuals        12 153908   12826
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b. Find a 95 percent confidence interval on the mean tensile strength of the Portland cement produced by mixing technique 3.

$\bar{y}_3 = 2933.7$, $s(\bar{y}_3) = 56.65$

$$\hat{V}(\hat{L}) = \sqrt{MSE * \sum_{i=1}^{a} \frac{C_i^2}{n_i}} = \sqrt{12825.7 * \frac{1}{4}} = 56.625$$

```
qt(0.975, 12)
```

```
[1] 2.178813
```

$2933.7 \pm (2.179 * 56.625) = \mathbf{(2810.31, 3057.09)}$

c. Find a 95 percent confidence interval on the difference in means for techniques 1 and 3 & 3 and 4.

$$L = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_3 + \mu_4}{2} = (\tfrac{1}{2})\mu_1 + (0)\mu_2 + (\tfrac{1}{2})\mu_3 + (\tfrac{-1}{2})\mu_3 + (\tfrac{-1}{2})\mu_4$$

$$\hat{L} = \frac{\hat{y}_1 + \hat{y}_3}{2} - \frac{\hat{y}_3 + \hat{y}_4}{2} = \frac{2971 + 2933.75}{2} - \frac{2933.75 + 2666.25}{2} = 152.375$$

$$\sum_{i=1}^{4} \frac{c_i^2}{n_i} = \frac{\left(\tfrac{1}{2}\right)^2}{4} + \frac{\left(\tfrac{1}{2}\right)^2}{4} + \frac{\left(\tfrac{-1}{2}\right)^2}{4} + \frac{\left(\tfrac{-1}{2}\right)^2}{4} = 0.25$$

$$\hat{V}(\hat{L}) = mse * \sum_{i=1}^{4} \frac{c_i^2}{n_i} = 12825.7 * 0.25 = 3206.425$$

```
qt(0.975,12)
```

```
[1] 2.178813
```

$152.375 \pm (2.179 * \sqrt{3206.425}) = 152.375 \pm 123.387$

$(28.988, 275.76)$

d. Compute simultaneous 95% CIs for the difference in means for techniques 1 and 3, based on Tukey's, Bonferroni's and Scheffe's Methods:

$\hat{L} = \hat{y}_1 - \hat{y}_3 = 2971.00 - 2933.75 = 37.25$

$$\hat{V}(\hat{L}) = mse * \sum_{i=1}^{4} \frac{c_i^2}{n_i} = 12825.7 * (\tfrac{1}{4} + \tfrac{1}{4}) = 6412.85$$

Tukey's

```
qtukey(0.95, 3, 12)
```

```
[1] 3.772929
```

$Q = 3.773/\sqrt{2}$ 2.668

$HSD = Q * \sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})} = 2.668 * \sqrt{6412.85} = 213.647$

37.25 ± 213.647

(-176.397, 250.897)

Bonferroni's

$t_{(1-\frac{\alpha}{2*g}),df_{error}} = $ Bonferroni's CV

$g = \frac{a(a-1)}{2} = \frac{4(3)}{2} = 6$

1-(0.05/12)

```
qt(.9958,12)
```

```
[1] 3.14839
```

$\hat{L} \pm B * \sqrt{\hat{V}(\hat{L})} = 37.25 \pm (3.148 * \sqrt{6412.85}) = 37.25 \pm 252.124$

(-214.874, 289.374)

Scheffe's

$\sqrt{(a-1)f_{(1-\alpha),(a-1),df_{error}}} = $ Scheffe's CV

```
qf(0.95, 3, 12)
```

```
[1] 3.490295
```

$\sqrt{3 * 3.49} = 3.236$

$\hat{L} \pm S * \sqrt{\hat{V}(\hat{L})} = 37.25 \pm (3.236 * \sqrt{6412.85}) = 37.25 \pm 259.130$

(-221.880, 296.380)

e. Compute Bonferroni's Minimum Significant Difference

$t_{(1-\frac{\alpha}{2*g}),df_{error}} = $ Bonferroni's CV

$g = \frac{a(a-1)}{2} = \frac{4(3)}{2} = 6$ $1 - \frac{0.05}{2*6} = 1 - .0042 = .9958$

```
qt(.9958, 12)
```

[1] 3.14839

$B_{CV} * \sqrt{\widehat{V}(\widehat{L})} = 3.14839 * 80.08 = 252.124$

f. Compute Tukey's Honest Significant Difference

$qtukey_{1-\alpha, df_{btwn}, df_{error}} = $ Tukey's CV

```
qtukey(0.95, 3, 12)
```

[1] 3.772929

$3.772929 \div \sqrt{2} = 2.668$

$Tukey_{CV} * \sqrt{\widehat{V}(\widehat{L})} = 2.668 * 80.08 = 213.647$

g. Compute Scheffe's Minimum Significant Difference

$\sqrt{(a-1)f_{(1-\alpha),(a-1),df_{error}}} = $ Scheffe's CV

```
qf(0.95,3,12)
```

[1] 3.490295

$S = \sqrt{3 * 3.490295} = 3.236$

$S_{CV} * \sqrt{\widehat{V}(\widehat{L})} \; 3.236 * 80.08 = 259.130$

Problem 3

A study is conducted to compare 4 menus in terms of numbers of calories ordered by restaurant customers (in 100s of calories). The treatments (menus) are (consider them increasing in order of information provided):
1) No Calories Reported 2) Calories Reported 3) Rank-Ordered Calories 4) Color-Coded Calories The sample sizes are all based on samples of $r = 20$ customers per menu. The sample means and estimated variance are: $\bar{Y}_{(1 \bullet)}$=17.6 $\bar{Y}_{(2 \bullet)}$=16.8 $\bar{Y}_{(3 \bullet)}$=16.0 $\bar{Y}_{(4 \bullet)}$=14.4 $\bar{Y}_{(\bullet \bullet)}$=16.2 s^2="MSE"=196.0. Give 3 orthogonal contrasts, and their estimates and estimated standard errors Contrast 1: Menu 1 vs Menus {2, 3, 4} Contrast 2: Menu 2 vs Menus {3, 4} Contrast 3: Menu 3 vs Menu 4

$$L_1 = (3)\mu_1 - \mu_2 - \mu_3 - \mu_4$$

$$\hat{L}_1 = (3)17.6 - (16.8 + 16 + 14.4) = 5.6$$

$$sd(\hat{L}_1) = \sqrt{196 * (\tfrac{3^2}{20} + \tfrac{1^2}{20} + \tfrac{1^2}{20} + \tfrac{1^2}{20})} = 10.844$$

$$L_2 = (0)\mu_1 + (2)\mu_2 + \mu_3 + \mu_4 \quad \hat{L}_2 = (0)17.6 + (2)16.8 - (16 + 14.4) = 3.2 \quad sd(\hat{L}_2) = \sqrt{196 * (\tfrac{2^2}{20} + \tfrac{1^2}{20} + \tfrac{1^2}{20})} = 7.668$$

$$L_3 = \mu_3 - \mu_4 \quad \hat{L}_3 = 16 - 14.4 = 1.6 \quad sd(\hat{L}_3) = \sqrt{196 * (\tfrac{1^2}{20} + \tfrac{1^2}{20})} = 4.427$$

Problem 4

An article in Environment International (Vol. 18, No. 4, 1992) describes an experiment in which the amount of radon released in showers was investigated. Radon enriched water was used in the experiment and six different Orifice diameters were tested in shower heads. The data from the experiment are shown in the following table.

```
radon_release<-c(80,83,83,85,75,75,79,79,74,73,76,
                 77,67,72,74,74,62,62,67,69,60,61,64,66)
diameter<-c(rep("0.37",4),rep("0.51",4),rep("0.71",4),
            rep("1.02",4),rep("1.40",4),rep("1.99",4))
radon_data<-data.frame(radon_release,diameter)
radon_data
```

```
   radon_release diameter
1             80     0.37
2             83     0.37
3             83     0.37
4             85     0.37
5             75     0.51
6             75     0.51
7             79     0.51
8             79     0.51
9             74     0.71
10            73     0.71
11            76     0.71
12            77     0.71
13            67     1.02
14            72     1.02
15            74     1.02
16            74     1.02
```

```
17              62      1.40
18              62      1.40
19              67      1.40
20              69      1.40
21              60      1.99
22              61      1.99
23              64      1.99
24              66      1.99
```

```
str(radon_data)
```

```
'data.frame':   24 obs. of  2 variables:
 $ radon_release: num   80 83 83 85 75 75 79 79 74 73 ...
 $ diameter     : chr  "0.37" "0.37" "0.37" "0.37" ...
```

a. Write a suitable model (Factor effects) for this experiment. Define all terms in the model. Include all assumptions of the model.

$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

Where μ is population mean

α_i is the treatment effect of the ith group

$\mu + \alpha_i = \mu_i$ is the mean of the ith group

i is the number of treatments, 1-6

j is the number of observations within each treatment, 1-4

ϵ_{ij} is the random error

Assumptions:

- All observations are independent from one another

- The population is approximately normally distributed

- Constant variance

b. Obtain point estimates and standard deviation of the radon released at each level.

```
radon_data |>
  group_by(diameter) |>
  summarize(point_estimate = mean(radon_release), sd = sd(radon_release))
```

```
# A tibble: 6 x 3
  diameter point_estimate    sd
  <chr>             <dbl> <dbl>
1 0.37               82.8  2.06
2 0.51               77    2.31
3 0.71               75    1.83
4 1.02               71.8  3.30
5 1.40               65    3.56
6 1.99               62.8  2.75
```

c. Does the size of the orifice affect the mean percentage of radon released? Use $\alpha = 0.05$

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

H_a: at least one μ is different from the rest

```
ano_mod <- aov(radon_release ~ diameter, data = radon_data)
summary(ano_mod)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
diameter     5 1133.4  226.68   30.85 3.16e-08 ***
Residuals   18  132.2    7.35
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of 3.16e-08 which is less than the alpha of 0.05, we reject the null hypothesis and have strong evidence to support the alternative, that the size of the orifice does affect the mean percentage of radon release.

d. Construct a normal probability plot of the residuals. What conclusion would you draw about the validity of the normality assumption?

```
norm <- lm(radon_release ~ diameter, data = radon_data)

plot(norm, which = c(2), sub.caption = "Problem 4D Jackson Connor")
```

Q–Q Residuals

Theoretical Quantiles
Problem 4D Jackson Connor
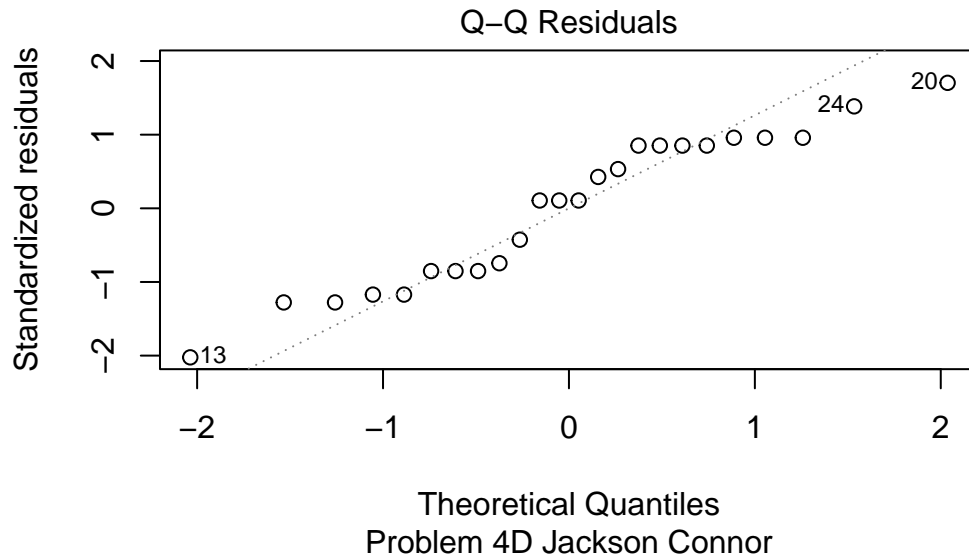
Based on the QQplot above, we can see that the observations mostly follow the line with no dramatic outliers, indicating linearity and that normality has been met.

e. Plot the residuals versus the predicted radon release. Comment on the plot.

```
plot(norm, which = c(1), sub.caption = "Problem 4E Jackson Connor")
```

Residuals vs Fitted
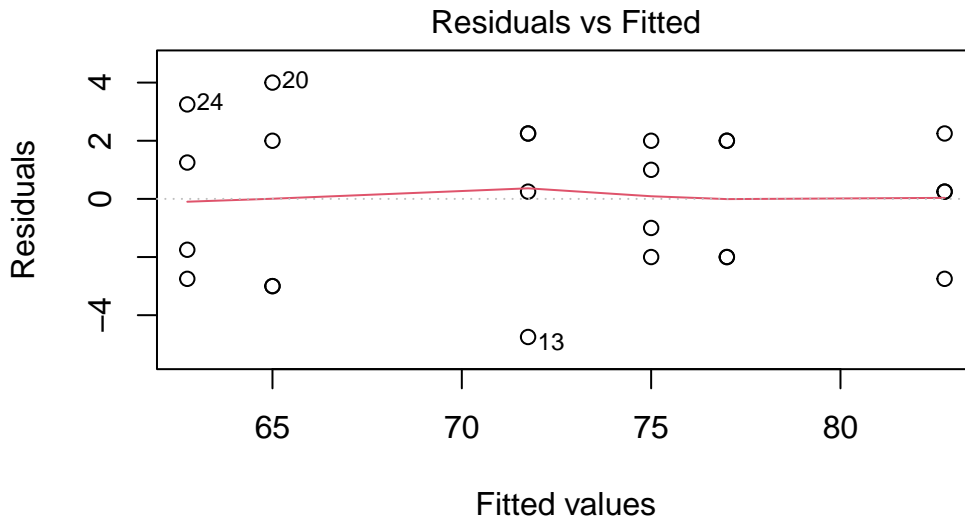
Problem 4E Jackson Connor

The observations are randomly distributed around the red line with no obvious pattern, indicating that constant variance assumption has been met.

f. Test by means of the Modified Leven's (Brown-Forsythe) and Bartlett's test whether the treatment error variances for the response variable are equal.

```
#Bartlett

bartlett.test(radon_release ~ diameter, data = radon_data)
```

```
	Bartlett test of homogeneity of variances

data:  radon_release by diameter
Bartlett's K-squared = 1.8214, df = 5, p-value = 0.8733
```

```
#Modified Levene's

leveneTest(radon_release ~ diameter, data = radon_data, center = median)
```

```
Warning in leveneTest.default(y = y, group = group, ...): group coerced to
factor.
```

13

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  5  0.8605 0.5261
      18
```

As both the Bartlett and Modified Levene's tests had p-values greater than 0.05 (0.8733 and 0.5261 respectively) we can assume equal variance.

g. Test the normality assumption using Shapiro-Wilk test.

```
resid <- residuals(ano_mod)

shapiro.test(resid)
```

```
    Shapiro-Wilk normality test

data:  resid
W = 0.93478, p-value = 0.1247
```

As our p-value is greater than 0.05 (0.1247>0.05), we can assume that the residuals follow an approximately normal distribution.

h. Use the Tukey's and Bonferroni's methods to make comparisons among the five treatments to determine specifically which treatments differ in radon release.

```
TukeyHSD(ano_mod)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = radon_release ~ diameter, data = radon_data)

$diameter
            diff        lwr         upr       p adj
0.51-0.37  -5.75 -11.841234   0.3412336 0.0707511
0.71-0.37  -7.75 -13.841234  -1.6587664 0.0084181
1.02-0.37 -11.00 -17.091234  -4.9087664 0.0002404
1.40-0.37 -17.75 -23.841234 -11.6587664 0.0000004
1.99-0.37 -20.00 -26.091234 -13.9087664 0.0000001
0.71-0.51  -2.00  -8.091234   4.0912336 0.8968057
1.02-0.51  -5.25 -11.341234   0.8412336 0.1153360
```

```
1.40-0.51 -12.00 -18.091234  -5.9087664 0.0000841
1.99-0.51 -14.25 -20.341234  -8.1587664 0.0000089
1.02-0.71  -3.25  -9.341234   2.8412336 0.5513482
1.40-0.71 -10.00 -16.091234  -3.9087664 0.0007059
1.99-0.71 -12.25 -18.341234  -6.1587664 0.0000650
1.40-1.02  -6.75 -12.841234  -0.6587664 0.0249971
1.99-1.02  -9.00 -15.091234  -2.9087664 0.0021152
1.99-1.40  -2.25  -8.341234   3.8412336 0.8432736
```

According to Tukey'sHSD, the differences are signficantly different between 0.71-0.37, 1.02-0.37, 1.40-0.37, 1.99-0.37, 1.40-0.51, 1.99-0.51, 1.40-0.71, 1.99-0.71, 1.40-1.02, and 1.99-1.02 at the significance level of 0.05.

```
pairwise.t.test(radon_release, diameter, p.adjust.method = "bonf")
```

```
	Pairwise comparisons using t tests with pooled SD

data:  radon_release and diameter

     0.37    0.51    0.71    1.02    1.40
0.51 0.11528 -       -       -       -
0.71 0.01143 1.00000 -       -       -
1.02 0.00029 0.20222 1.00000 -       -
1.40 4.3e-07 1.0e-04 0.00087 0.03654 -
1.99 6.9e-08 1.0e-05 7.7e-05 0.00270 1.00000

P value adjustment method: bonferroni
```

According to the Bonferroni method, there is significant difference between the means of 0.71-0.37, 1.02-0.37, 1.40-0.37, 1.99-0.37, 1.40-0.51, 1.99-0.51, 1.40-0.71, 1.99-0.71, 1.40-1.02, and 1.99-1.02 at the significance level of 0.05.