Stat 350

Homework #2 Name:\_\_Tuo Wang\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Due Tuesday March 1st by 12:00 pm**

1. The data in the table below resulted from an experiment that utilized a completely randomized design.

|  |  |  |
| --- | --- | --- |
| Treatment1 | Treatment2 | Treatment3 |
| 3.9 | 5.4 | 1.3 |
| 1.4 | 2.0 | 0.7 |
| 4.1 | 4.8 | 2.2 |
| 5.5 | 3.8 |  |
| 2.3 | 3.5 |  |

**N = 13 K =3**

**17.2 19.5 4.2 Sum = 40.9**

a) Use the information in the table above and the summary statistics below to find the three missing values in the ANOVA table. **Use the formulas in your notes (subtraction method does not count as formula) and show your work**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **n** | **Mean** | **Std. dev** |
| Treatment1 | 5 | 3.44 | 1.6087262 |
| Treatment2 | 5 | 3.9 | 1.3076697 |
| Treatment3 | 3 | 1.4 | 0.75498344 |

**ANOVA table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **df** | **SS** | **MS** | **F-Stat** | **P-value** |
| Treatments | 2 | ??? | 6.2101538 | ???? | 0.0753 |
| Error | 10 | 18.332 | ???? |  |  |
| Total | 12 | 30.752308 |  |  |  |

**SS(total) = SSE + SST = 30.752308 so SST= 30.752308 – SSE**

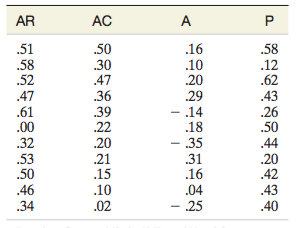
**Alternatively, SS(treatment) = ΣC2/n - G2/N = (17.22/5 + 19.52/5+4.22/3) – (40.92/13) = 12.42**

****

**MSE = = 18.332/(13-3)= 1.8332**

**F = MST / MSE = 6.2101538 / 1.8332 = 3.38760**

2. Complete textbook problem #10.34 on page 495 and the follow-up problem #10.57 on page 505. Conduct a complete analysis including stating the null and alternative hypothesis statements, defining all parameters, assumption checks and a post-hoc test if necessary. The data set is DRINKERS on STATCRUNCH chapter 10.



**4.84 2.92 0.7 4.4 sum of scores across groups=12.86**

**N=11\*4 = 44 K=4 sum of squared total = 5.941**

1, defining parameter:

µ= mean test score (the difference between the proportion of correct responses on the green list of words and the proportion of incorrect responses on the red list of words. )

2, Hypothesis :

Ho : µ1 = µ2 = µ3 = µ4

Ha: at least two of the 4 treatment means differ

3, assumption check:

AR: AC:

A: P:



From the QQplot we made, we assume that the data are randomly selected and independent from each other. Also the data is normally distributed because it follows the trend line.

4,ANOVA analysis

**The ANOVA Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F |
| Treatments | k-1 | ΣC2/n - G2/N | MST | MST/MSE |
| Error | N-k | By subtraction | MSE |  |
| Total | N-1 | ΣΣx2 – G2/N |  |  |

**The ANOVA Table’**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F |
| Treatments | 3 | 0.6456 | 0.7274 | 19.1421 |
| Error | 40 | 1.5367 | 0.038 |  |
| Total | 43 | 2.1823 |  |  |

SS(Total) = ΣΣx2 – G2/N = 5.491 – (12.86^2/44)= 5.591 - 3.758627= 2.1823

SS(Treatment) = ΣC2/n - G2/N ={ ( 4.84^2 + 2.92 ^2 + 0.7 ^2 + 4.4 ^2 ) /11 } – ( 12.86^2 /44) = 4.40425-3.758627 = 0.6456

SS(Error) = subtraction =1.5367

MST = SST/k-1 =2.1823/3=0.7274

MSE= SSE/N-k = 1.5367 /44-4 = 1.5367 /40= 0.038

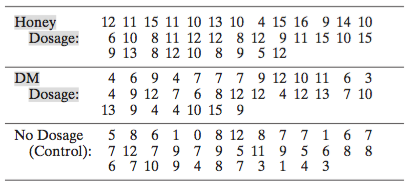
F = MST/MSE = 0.7274/0.038 = 19.1421

Find the critical value in the F table we got it = 2.84

Since F = 19.1421 > 2.84

We reject the null hypothesis. There is no evidence to suggest that the mean from 4 different treatments are the same.

3. Complete textbook problem #10.35 on page 496. Conduct a complete analysis including stating the null and alternative hypothesis statements, defining all parameters, assumption checks and a post-hoc test if necessary. See question 10.54 page 505 to help with your follow-up question. Dataset is HONEYCOUGH on STATCRUNCH chapter 10.



1, defining parameter :

µ=the mean of children’s cough symptom on scale (ranging from 0 – 30)

2, hypothesis :

Ho : µhoney = µDM = µControl

Ha: at least two of the 3 treatment means differ

3,assumption check

Honey: DM:



Control:



**We can say the data from each group is randomly, independently selected and normally distributed**

4,ANOVA analysis:

> dosage = c(12,11,15,11,10,13,10,4,15,16,9,14,10,6,10,8,11,12,12,8,12,9,11,15,10,15,9,13,8,12,10,8,9,5,12,4,6,9,4,7,7,7,9,12,10,11,6,3,4,9,12,7,6,8,12,12,4,12,13,7,10,13,9,4,4,10,15,9,5,8,6,1,0,8,12,8,7,7,1,6,7,7,12,7,9,7,9,5,11,9,5,6,8,8,6,7,10,9,4,8,7,3,1,4,3)

> Group = c(rep("Honey",33),rep("DM",35),rep("Control",37))

> plot(dosage ~ Group, data=migraine)



> aov(dosage ~ Group,data=migraine)

Call:

aov(formula = dosage ~ Group, data = migraine)

Terms:

**Group Residuals**

**Sum of Squares 328.8572 917.3714**

**Deg. of Freedom 2 102**

**Residual standard error: 2.998973**

>results=aov(dosage ~ Group,data=migraine)

> summary(results)

**Df Sum Sq Mean Sq F value Pr(>F)**

**Group 2 328.9 164.43 18.28 1.64e-07 \*\*\***

**Residuals 102 917.4 8.99**

**---**

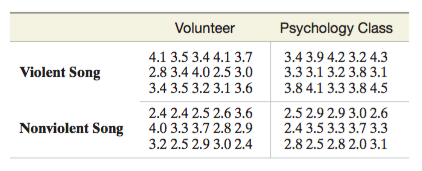
**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

>

**From the test result we can see that the F = 18.28 and P=1.64e-07,which < 0.05. So we can reject the null.**

**There is no evidence that the mean from three different groups(honey, DM, control) are the same**

4. Complete textbook example problem #10.96 on page 536. Conduct a complete analysis including stating the null and alternative hypothesis statements, defining all parameters, assumption checks and a post-hoc test if necessary.



1, defining parameter :

(a β)ij:the interaction effect of song and pook on vioence score

a=the true mean of the effect of types of songs played on violence score

β=the true mean of the effect of types of pool selected on violence score

2, hypothesis :

1. Ho : (a β)ij =0 Ha: Some qkj ≠0 (Interaction)

2. Ho : a1 = a2 =. . .ar= 0 Ha: Some ak≠ 0 (Factor A)

3. Ho : β1 = β2 =…. βc = 0 Ha: Some βj ≠0 (Factor B)

1: The factor of A and B do not interact to affect the response variable

2: The factor of A and B do interact to affect the response variable

3,assumption check

> data = read.csv("/Users/TariesW/Desktop/STAT 350/HW2no4.csv",header=T)

> data

song scores pool

1 violent 4.1 volunteer

2 violent 3.5 volunteer

3 violent 3.4 volunteer

4 violent 4.1 volunteer

5 violent 3.7 volunteer

6 violent 2.8 volunteer

7 violent 3.4 volunteer

8 violent 4.0 volunteer

9 violent 2.5 volunteer

10 violent 3.0 volunteer

11 violent 3.4 volunteer

12 violent 3.5 volunteer

13 violent 3.2 volunteer

14 violent 3.1 volunteer

15 violent 3.6 volunteer

16 violent 3.4 psychology class

17 violent 3.9 psychology class

18 violent 4.2 psychology class

19 violent 3.2 psychology class

20 violent 4.3 psychology class

21 violent 3.3 psychology class

22 violent 3.1 psychology class

23 violent 3.2 psychology class

24 violent 3.8 psychology class

25 violent 3.1 psychology class

26 violent 3.8 psychology class

27 violent 4.1 psychology class

28 violent 3.3 psychology class

29 violent 3.8 psychology class

30 violent 4.5 psychology class

31 nonviolent 2.4 volunteer

32 nonviolent 2.4 volunteer

33 nonviolent 2.5 volunteer

34 nonviolent 2.6 volunteer

35 nonviolent 3.6 volunteer

36 nonviolent 4.0 volunteer

37 nonviolent 3.3 volunteer

38 nonviolent 3.7 volunteer

39 nonviolent 2.8 volunteer

40 nonviolent 2.9 volunteer

41 nonviolent 3.2 volunteer

42 nonviolent 2.5 volunteer

43 nonviolent 2.9 volunteer

44 nonviolent 3.0 volunteer

45 nonviolent 2.4 volunteer

46 nonviolent 2.5 psychology class

47 nonviolent 2.9 psychology class

48 nonviolent 2.9 psychology class

49 nonviolent 3.0 psychology class

50 nonviolent 2.6 psychology class

51 nonviolent 2.4 psychology class

52 nonviolent 3.5 psychology class

53 nonviolent 3.3 psychology class

54 nonviolent 3.7 psychology class

55 nonviolent 3.3 psychology class

56 nonviolent 2.8 psychology class

57 nonviolent 2.5 psychology class

58 nonviolent 2.8 psychology class

59 nonviolent 2.0 psychology class

60 nonviolent 3.1 psychology class



> interaction.plot(data$song,data$pool,data$scores)

> interaction.plot(data$pool,data$song,data$scores)



> x = lm(scores ~ song+pool+song\*pool,data=data)

> anova(x)

Analysis of Variance Table

**Response: scores**

**Df Sum Sq Mean Sq F value Pr(>F)**

**song 1 5.8907 5.8907 26.1144 4.031e-06 \*\*\***

**pool 1 0.1307 0.1307 0.5793 0.4498**

**song:pool 1 0.3527 0.3527 1.5634 0.2164**

**Residuals 56 12.6320 0.2256**

---

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

> **the output of the ANOVA table we see that there is a significant interaction effect (F=1.5634 , p=0.2164>0.05). We conclude that there isn’t a interaction between song and pool). The test for the main effect of treatment (F=36.1144, p=4.031e-06<0.05) shows a significant song effect on the scores. Finally, the test for the main effect of pool (F=1.5634 p=0.2164 >0.05) tells us there is no significant pool(volunteer or psychology class) effect on scores.**

5. Suppose we wish to consider the effect of two factors on blood pressure. Factor A is diabetes with two levels and Factor B is weight with two levels. Ten diabetics (half of which are overweight) and ten non-diabetics (half of which are overweight) are randomly selected. None of the twenty participants was on medication for blood pressure. The diastolic blood pressure for the twenty subjects is given below. We are interested in the interaction of weight and diabetes.

|  |  |  |
| --- | --- | --- |
|  | **Normal Weight** | **Overweight** |
| **Non-diabetic** | 75,80,83,85,65 | 85,80,90,95,88 |
| **Diabetic** | 85,90,95,90,86 | 90,95,100,105,110 |

**Conduct a complete analysis of the data** to help the researchers answer their questions.

1, defining parameter :

a:= the true mean of the effect of disease (diabetic or not) on blood pressure

β= the true mean of the weight on blood pressure

2, hypothesis :

1. Ho : All (a β)ij =0 Ha: Some qkj ≠0 (Interaction)

2. Ho : a1 = a2 =. . .ar= 0 Ha: Some ak≠ 0 (Factor A)

3. Ho : β1 = β2 =…. βc = 0 Ha: Some βj ≠0 (Factor B)

3,assumption check

4

> data = read.csv("/Users/TariesW/Desktop/STAT 350/HW2no5.csv",header=T)

> data

disease blood.pressure weight

1 Non-diabetic 75 Normal Weight

2 Non-diabetic 80 Normal Weight

3 Non-diabetic 83 Normal Weight

4 Non-diabetic 85 Normal Weight

5 Non-diabetic 65 Normal Weight

6 Non-diabetic 85 Overweight

7 Non-diabetic 80 Overweight

8 Non-diabetic 90 Overweight

9 Non-diabetic 95 Overweight

10 Non-diabetic 88 Overweight

11 diabetic 85 Normal Weight

12 diabetic 90 Normal Weight

13 diabetic 95 Normal Weight

14 diabetic 90 Normal Weight

15 diabetic 86 Normal Weight

16 diabetic 90 Overweight

17 diabetic 95 Overweight

18 diabetic 100 Overweight

19 diabetic 105 Overweight

20 diabetic 110 Overweight

> par(mfrow=c(1,2))

> plot(blood.pressure ~ disease+weight,data=data)



> interaction.plot(data$disease,data$weight,data$blood.pressure)

> interaction.plot(data$weight,data$disease,data$blood.pressure)



>result = lm(blood.pressure ~ disease+weight+disease\*weight,data=data)

> anova(result)

Analysis of Variance Table

**Response: blood.pressure**

**Df Sum Sq Mean Sq F value Pr(>F)**

**disease 1 720.0 720.00 16.6186 0.0008787 \*\*\***

**weight 1 540.8 540.80 12.4824 0.0027638 \*\***

**disease:weight 1 0.8 0.80 0.0185 0.8936061**

**Residuals 16 693.2 43.33**

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**the output of the ANOVA table we see (F=0.0185, p=0.89>0.05). We conclude that there isn’t an interaction between weight and the disease(diabetic). The test for the main effect of treatment (F=16.61, p=0.0008787) shows a significant weight effect on the blood pressure. Finally, the test for the main effect of age (F=12.48, p=0. 0027638) tells us there is a significant disease(diabetic) effect on blood pressure.**