Stat 502 Exam 1

(note: expressing hypotheses in terms of effects (e.g. requires writing out the model and explaining what terms refer to).

1. A study was conducted to evaluate the durability of 4 types of paint used for highway lanes. Each of the 4 paints (p1, p2, p3, p4) were applied at random locations along a stretch of highway to complete 4 replications. The response is an index of durability. The data were as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Paint | | | |
| P1 | P2 | P3 | P4 |
| 28 | 21 | 26 | 16 |
| 35 | 36 | 38 | 25 |
| 27 | 25 | 27 | 22 |
| 21 | 18 | 17 | 18 |
|  |  |  |  |

The overall mean was 25.0 and the total sums of squares was 692.0

a) (5 pts) Show the computations to find the Sums of Squares for the Treatment effect

Each observation in the dataset can be referenced by two indicator subscripts,*i*and *j* as Yij where we use *Y* to indicate that it is a response variable.  The subscript *i*refers to the *i*th level of the treatment (we have 4 treatments for type of paint so *i* will take on the values 1,2 3, and 4)  The subscript *j*refers to the *j*th observation

Treatment SS = SStrt =∑i ni (Yi.¯ − Y¯..)2

We have treatment means:

|  |  |  |  |
| --- | --- | --- | --- |
| Y1. ¯ | Y2. ¯ | Y3. ¯ | Y4. ¯ |
| 27.75 | 25 | 27 | 20.25 |

So Treatment SS = 4 \* (27.75 – 25)^2 + 4 \* (25 – 25)^2 + 4 \* (27 – 25)^2 + 4 \* (20.25 – 25)^2 = 136.5

b) (15 pts) Construct the ANOVA table

**ANOVA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **df** | **SS** | **MS** | **F** |
| Treatment | *k* - 1 = 3 | 136.5 | MSTrt=SSTrt / dfTrt = 45.5 | F=MSTrt / MSError = 0.983 |
| Error | 15-3=12 | 555.5 | MSError=SSError / dfError = 46.292 |  |
| Total | *N* - 1 =15 | 692.0 |  |  |

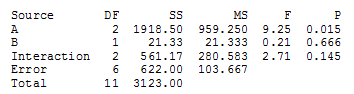
c) (5pts) Find the Fcritical value to test at a 95% confidence level.

Critical F is Fα =F(0.05, 3, 12) = 3.49

d) (5 pts) Conclude: \_\_\_\_\_\_\_\_\_\_\_ .

TheFcalculated < Fα so we **fail to** **reject H0**

2) (10 pts) The output below shows a two-factor ANOVA with no significant interaction and no significant main effect for Treatment B. Given that the researcher has expressed no interest in reporting the lack of significant effect of Treatment B or Interaction, complete the one-way ANOVA for treatment A.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F |
| A | 2 | 1918.50 | 959.25 | 7.164 |
| Error | 9 | 1205.1 | 133.9 |
| Total | 11 | 3123.6 |  |

3) (5X2=10 points) The interaction plots for the ANOVA1 and ANOVA2 outputs shown below are Plot1 and Plot2 respectively. State whether each of the following statements is TRUE or FALSE, and correct the statements that are false.

x

B1

Plot 1

A2

A1

B2

**ANOVA 1 : Yield versus Ingred1, ingred2**

Source DF SS MS F-Value P-Value

Ingred1 1 18.06 18.06 2.12 0.171

Ingred2 1 90.25 90.25 10.59 0.007

Ingred1\*ingred2 1 3.42 3.42 0.40 0.538

Error 12 102.30 8.52

Total 15 214.03

Observation: Table shows Ingred1 and interaction is not significant

Plot shows no effect of A, large effect of B and no interaction

**ANOVA 2**

X

B1

B2

A1

A2

Plot 2

**General Linear Model: weight versus school, gender**

Source DF SS MS F-Value P-Value

school 1 5.00 5.00 29.41 0.000

gender 1 0.24 0.24 1.42 0.250

school\*gender 1 5.61 5.61 33.05 0.000

Error 16 2.72 0.17

Total 19 13.57

Observation: Table shows that interaction is significant. School has large effect

Plot shows:

No effect of Factor A, a large effect of Factor B, with a very large interaction.

1. The factors A and B in Plot 1 are ingred1 and ingred2, respectively.

True.

Observation: Ingred1 and interaction is not significant

Plot shows no effect of A, large effect of B and no interaction

1. In ANOVA2, the factor school has a statistically significant main effect, and it corresponds to factor A in Plot2.

False. While the school is statistically significant it also has a large effect from the table. The plot however indicates that Factor A has no effect so school can’t correspond to Factor A.

Corrected statement will be: the factor school has a statistically significant main effect, and it corresponds to factor B in Plot2

1. The test statistic value for testing the main effect of factor Ingred2 is F=10.59, with numerator and denominator degrees of freedom equal to 1 and 12, respectively.

True.

***Multiple answers possible – no points taken off. The interaction may be retained, as would be the usual situation for a designed experiment. However, the other option is to remove the interaction term and ‘reduce’ the model to an additive effect only model. Model reduction is typically employed in observational studies where the objective is to identify significant predictors (as in MlR), whereas all the terms in the original model are retained in a designed experiment.***

1. The appropriate model for the effects of school and gender on weight is:

weight = mean + school + gender + error.

False. The model doesn’t contain the significant interaction term which must be added. The model should correspond to Yijk=μ..+αi+βj+(αβ)ij+ϵijk

weight = mean + school + gender + school\*gender + error.

1. The appropriate model for the effects of Ingred1 and Ingred2 on Yield is:

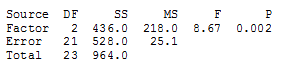
Yield = mean + Ingred1 + Ingred2 + error

False. The interaction and Ingred1 are not significant and should not be included. So Ingred1 should be removed. Therefore Yield = mean + Ingred2 + error

. ***Note: Multiple answers possible – no points taken off. Again, the interaction may be retained, as would be the usual situation for a designed experiment, but the other option is to remove the interaction term and ‘reduce’ the model to an additive effect only model. Model reduction is typically employed in observational studies where the objective is to identify significant predictors (as in MlR) as is not a necessary step to perform for a designed experiment..***

4) (10 pts) Using a balanced, one-factor ANOVA, the following are scores on a fifty-question vocabulary test administered to twenty-four (24) students after one year of studying a foreign language under one of three teaching methods (B1, B2, and B3). The ANOVA output is also provided. Perform a Tukey multiple comparison for the teaching method effect at a 0.05 alpha level of significance. Include a letter-labeling comparison of means in your conclusion.

|  |  |  |
| --- | --- | --- |
| Aural-Oral  Method (A)  B1 | Translation  Method (T)  B2 | Combined  Method (C)  B3 |
| 37  30  (etc)  37  28 | 27  24  (etc)    14  15 | 20  31  (etc)    23  18 |



The ranked sample means are:

In the Tukey procedure we compute a ‘yardstick’ value based on the MSError and the number of means being compared. If any two means differ by more than the Tukeyw value, then they are significantly different.

Tukey’s w value

w = qα(p,dfError)⋅sY¯

p = then number of treatment levels

We have

qα(p,dfError) = q.05(3, 21) = 3.58

r= number of replications = 8

sY¯ =standard error of a treatment mean =  = = 1.7713

Therefore w = 3.58 \* 1.7713 = 6.3413

Since *v=21* is not on table can use 20 or 24:

*v=*20: w *=* q0.05(3,21) = 3.58 = 3.58\*1.77 = 6.34

*v=*24: w *=* q0.05(3,21) = 3.53 = 3.53\*1.77 = 6.25

For the example the tukey analysis gives:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tukey w | 6.3413 |  |
| Method | A | C | T |
| Means | 30 | 21.5 | 20.5 |
|  | a | b | b |
|  |  |  |  |
|  |  | Is < w | Label |
| A-C | 8.5 | No | a to A and b to C |
| C-T | 1 | Yes | b to T |

5) A study is designed to compare three Companies A, B, and C, that produce heat sinks (a heat sink is a device for cooling electronic components). Each company makes two models of heat sink and they are interested in comparing these as well. To measure the effectiveness of the six heat sinks, each one is randomly applied to five identical central processing units (CPUs) and run under identical conditions for 10 minutes. The temperature of each CPU is then recorded.

1. (5 pts) Either write the statistical model for this experiment, or alternatively, you may write the model statement in SAS code that would be used for this situation.

This is a nested design situation where we are, using the subscript (i) to identify the Company α, and the subscript (j) to indicate the model β and where the statistical model is:

Yijk = μ.. + αi + βj(i) + ϵijk

where:

μ.. is a constant  
αi are constants subject to the restriction ∑αi=0  
βj(i) are constants subject to the restriction ∑jβj(i)=0 for all *i*  
ϵijk are independent *N*(0, σ2)  
*i* = 1, ... , *a*; *j* = 1, ... , *b*; *k* = 1, ... , *n (here a=3, b=2 and n=30 )*

We use the notation with parenthesis to indicate nesting: Model(Company).

The model statement in SAS will be:

model response = Company Model(Company)

1. (10 pts) Construct the first two columns (SS and df) of the ANOVA table for this experiment.

|  |  |  |
| --- | --- | --- |
| Source | DF | SS |
| Company | (a - 1) = 2 | )2 = )2 |
| Model(Company) | a(b - 1) = 3 | )2 = )2 |
| Error | ab(n - 1) = 6 \* 4 = 24 | (help: abn – ab)  )2 |
| Total | 29 | )2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Yijk−Y¯... | = | Y¯i..−Y¯... | + | Y¯ij.−Y¯i.. | + | Yijk−Y¯ij. |
| Total deviation |  | Company main effect |  | Specific *Model* effect when *Company* at the *i*th level |  | Residual |

6) Suppose a study was made of attendance of elementary, junior high, and senior high school students for four ethnic groups (I, II, III, and IV). The study was balanced and included n=24 students at each school level.

a) (10 pts) Portions of the ANOVA table from the analysis are given below. Complete the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ~~Source~~ | ~~DF~~ | ~~SS~~ | ~~MS~~ | ~~F~~ |
| ~~School Level (S)~~ | ~~(a - 1) = 2~~ | ~~900~~ | ~~450~~ | ~~4.875~~ |
| ~~Ethnicity (E)~~ | ~~(b - 1) = 3~~ | ~~750~~ | ~~250~~ | ~~2.708~~ |
| ~~S x E~~ | ~~(a - 1)(b - 1) = 6~~ | ~~1200~~ | ~~200~~ | ~~2.167~~ |
| ~~Error~~ | ~~71-6 = 65~~ | ~~6000~~ | ~~92.308~~ |  |

~~Total 71~~

DFE is incorrect

With n=24 in each school and three schools, N = 3\*24 = 72

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F |
| School Level (S) | (3-1) = 2 | 900 | (900/2) = 450 | (450/100) = 4.5 |
| Ethnicity (E) | (4-1) = 3 | (250\*3) = 750 | 250 | (250/100) = 2.5 |
| S x E | (3-1)\*(4-1) = 6 | 1200 | (1200/6) = 200 | (200/100) = 2.0 |
| Error | (72-11-1) = 60 | 6000 | (6000/60) = 100 |  |

b) (5 pts) Provide the Null and Althernative hypotheses for testing each main effect and interaction.

Main Effect of School Level S:

H0: μ1. = μ2. = μ3. HA: not all μi. are equal

Main Effect of Ethinicity E:

H0:μ.1 = μ.2 = μ.3 = μ.4  
HA: not all μ.j are equal

School × Ethinicity Interaction:

H0: there is no interaction    
HA: an interaction exists

The model if all the terms are significant will be:

Response = S + E + S\*E + error

c) (5 pts) What are critical F-values for each hypothesis test above in part b at alpha = 0.10 or 10%?

~~Critical F values are~~

~~School: F(0.9, 2, 65) = 2.38611~~

~~Ethinicity: F(0.9, 3, 65) = 2.17003~~

~~S\*E: F(0.9, 6, 65) = 1.86681~~

School Level: F0.10, 2,60 = 2.39

Ethnicity: F0.10, 3,60 = 2.18

Interaction: F0.10, 6,60 = 1.87

1. (5 pts) Does the F-test for the S x E interaction indicate that the attendance trend did not follow the same pattern for the four ethnic groups at a 10% level of significance?

Yes. We find that Fcalculated > Fcritical so we reject H0

We can conclude that the interaction is significant and therefore that the attendance trend did not follow the same pattern for the four ethnic groups at a 10% level of significance.

Yes. The F-test statistic for the interaction is 2.0 which exceeds the critical F-value of 1.87 at a 0.10 level of significance. ***Note: There were problems with interpreting the term “attendence trend” and no points were deducted accordingly.***