**BF[2] – Using Software**

**Type in your score here 🡪 \_\_\_34\_ out of 37 points possible**

1. (5 points) Ponder/Reflect Exercise – Reflect on what you have learned from this portion of the class. Examples of what you can do are: a brief outline of material covered, insights you gained from class or personal study, or items you feel that you need to follow up or work on. (3-5 sentences)

Doing a 2 way ANOVA can help test for interactions between different variables. If you use the interaction you can’t use the individual variables to show significance. This helps when doing multi-level tests with multi-factors.

2. Use the file snapbean data to conduct a two-way ANOVA in R. This experiment is evaluating whether the date of sowing and/or the variety of snapbean plant will affect the total yield of snapbeans. You should test if “Yield" (the response) is affected by “sowdate" (1=early,...,4=late), “variety" (1, 2, or 3), or the interaction of sowdate with variety. Conduct your analysis in R. To do the analysis for the main effects, sowdate and variety plus their interaction, please do the following:

a) (3 points) Get the interaction plots, Boxplots and means plots and describe what you see with each of them.

There appears to be at least a slight interaction between sowdate and variety. Sowdate 4 appears much higher on average than the other three. The variances appear to be equal.

b) (3 points) Check the assumption of equal variance and residuals being normally distributed.

The residuals appear to be normally distributed.

The variances appear to be equal for both effects.

c) (5 points) For each of the following main effects and interaction: i) state the null and alternative hypotheses, ii) give the test statistic, iii) give the degrees of freedom, iv) state the p-value, v) determine whether you should reject or not reject the null hypothesis, and vi) write a sentence which gives an appropriate conclusion.

|  |
| --- |
| Analysis of Variance Table  Response: soy$Yield  Df Sum Sq Mean Sq F value Pr(>F)  soy$sowdate 3 747.78 249.259 876.130 < 2.2e-16 \*\*\*  soy$variety 2 8.40 4.201 14.767 1.005e-05 \*\*\*  soy$sowdate:soy$variety 6 5.87 0.978 3.437 0.006645 \*\*  Residuals 48 13.66 0.284  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
|  |
|  |

Sowdate

1. Ho: µS1= µS2 = µS3 = µS4 Ha: at least one of the population means is different

2. Test Statistic F=876.130

3. Degrees of Freedom Num df=3 Den df=48

4. P-value Close to 0

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

Variety

1. Ho: µV1= µV2 = µV3 Ha: at least one of the population means is different

2. Test Statistic F=14.767

3. Degrees of Freedom Num df=2 Den df=48

4. P-value is close to zero

5. Since the P-value is below 0.05 so we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

Interaction

1. Ho: There is no interaction between Sowdate and variety Ha: There is an interaction

2. Test Statistic F=3.437

3. Degrees of Freedom Num df=6 Den df=48

4. P-value =0.007

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that there is an interaction between Sowdate and variety

d) (2 points) Based on the interaction, should either of the main effects be interpreted? Explain why or why not?

Because the interaction effect is significant the main effects should not be interpreted (or at least great caution should be used). A main effects should only be interpreted when all higher order interactions are not significant and the main effect is significant.

3. Use the file programmers data to conduct a two-way ANOVA in R. This experiment was run to see how the type of experience of computer programmers and/or the years of experience for programmers impacts their ability to accurately estimate the time needed (in programmer days) to complete a large systems project. The response variable “TimePredE” represents the difference between the actual time required to complete a large systems project and the programmer's estimated time. Note that all values are positive, meaning that every subject underestimated the length of the task, but larger values represent larger time-prediction errors. You should test if “TimePredE" (the response) is affected by “LrgSysExp” (no=experienced only with small systems, yes=experienced with large systems), “YearsOfExp” (less5 = less than 5 years, less10 = between 5 and 10 years, more10=more than 10 years), or the interaction of LrgSysExp with YearsOfExp. Conduct your analysis in SPSS. To do the analysis for the main effects, “LrgSysExp” and “YearsOfExp” plus their interaction, please do the following:

a) (3 points) Get the interaction plots, Boxplots and means plots and describe what you see with each of them.

There appears to be an interaction between “LrgSysExp” and “YearsOfExp”

The mean is higher for the no “LrgSysExp”

The mean is higher for the Less5 “YearsOfExp”

There appears NOT to be equal variances for both “LrgSysExp” and “YearsOfExp”

b) (3 points) Check the assumption of equal variance and residuals being normally distributed.

The residuals appear to be normally distributed.

The variances DON’T appear to be equal for both effects, since the largest standard deviation is more than twice the smallest standard deviation for both main effects.

c) (5 points) For each of the following main effects and interaction: i) state the null and alternative hypotheses, ii) give the test statistic, iii) give the degrees of freedom, iv) state the p-value, v) determine whether you should reject or not reject the null hypothesis, and vi) write a sentence which gives an appropriate conclusion.

Analysis of Variance Table

Response: dat$time\_pred\_e

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

dat$lg\_sys\_ex 1 34504 34504 358.59 2.469e-13 \*\*\*

dat$years\_of\_exp 2 41720 20860 216.79 2.540e-13 \*\*\*

dat$lg\_sys\_ex:dat$years\_of\_exp 2 24234 12117 125.93 2.614e-11 \*\*\*

Residuals 18 1732 96

---

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

YearsofExp

1. Ho: µYOE1= µYOE2 = µYOE3 Ha: at least one of the population means is different

2. Test Statistic F=216.789

3. Degrees of Freedom Num df=2 Den df=18

4. P-value Close to 0

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

LgSysEx

1. Ho: µLSE1= µLSE2 = µLSE3 Ha: at least one of the population means is different

2. Test Statistic F=358.588

3. Degrees of Freedom Num df=1 Den df=18

4. P-value is close to zero

5. Since the P-value is below 0.05 so we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

Interaction

1. Ho: There is no interaction between YearsofEx and LgSysEx Ha: There is an interaction

2. Test Statistic F=125.958

3. Degrees of Freedom Num df=2 Den df=18

4. P-value is close to zero

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that there is an interaction between YearsofEx and LgSysEx

d) (2 points) Based on the interaction, should either of the main effects be interpreted? Explain why or why not?

Because the interaction effect is significant the main effects should not be interpreted (or at least great caution should be used). A main-effects should only be interpreted when all higher order interactions are not significant, and the main effect is significant.

4) (6 points) Given the assumptions were not met in problem 3, do parts b and c again with a transformation.

Analysis of Variance Table

Response: log(dat$time\_pred\_e)

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

dat$lg\_sys\_ex 1 3.5005 3.5005 131.091 1.071e-09 \*\*\*

dat$years\_of\_exp 2 4.3798 2.1899 82.010 9.044e-10 \*\*\*

dat$lg\_sys\_ex:dat$years\_of\_exp 2 1.3565 0.6782 25.399 5.744e-06 \*\*\*

Residuals 18 0.4806 0.0267

---

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

LgSysEx

1. Ho: µYOE1= µYOE2 = µYOE3 Ha: at least one of the population means is different

2. Test Statistic F=131.091

3. Degrees of Freedom Num df=2 Den df=18

4. P-value Close to 0

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

YearsofExp

1. Ho: µLSE1= µLSE2 = µLSE3 Ha: at least one of the population means is different

2. Test Statistic F=82.010

3. Degrees of Freedom Num df=1 Den df=18

4. P-value is close to zero

5. Since the P-value is below 0.05 so we would reject the null hypothesis

6. Therefore, we have sufficient evidence that at least one of the population means is different

Interaction

1. Ho: There is no interaction between YearsofEx and LgSysEx Ha: There is an interaction

2. Test Statistic F=25.399

3. Degrees of Freedom Num df=2 Den df=18

4. P-value is close to zero

5. Since the p-value is less than 0.05, we would reject the null hypothesis

6. Therefore, we have sufficient evidence that there is an interaction between YearsofEx and LgSysEx