CSUEB – STAT 6305 – Winter 2017 - Prof Yan Zhou

Homework 6 - Henry Lankin

February 23, 2017

HW 6: 17.10, 17.11, 17.27, 17.28

17.10

The following study was designed to evaluate the effectiveness of four chemicals developed to control fire ants. The type of environmental conditions in which the chemical is placed might have an effect on the effectiveness of the treatment to kill fire ants. Thus, the researcher randomly selected five locations from a large selection of locations; each location representing a randomly selected environment. To reduce the effect of different colonies of fire ants and the type of mounds they inhabit, the researcher created 40 artificial fire ant mounds and populated them with 50,000 ants having similar ancestry. The researcher randomly assigned two mounds to each of the 20 treatment–location combinations. The number of fire ants killed during a 1-week period was recorded. The number of fire ants killed (in thousands) are given here.

a) Write an appropriate linear statistical model for this study. Identify all terms in your model.

Mixed effect factorial design model:

$$\boldsymbol{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}}$$

$$i = 1, 2, 3, 4;\ j = 1, 2, 3, 4, 5;\ k = 1, 2$$

$y_{ij}k$ – the number of fire ants killed (in thousands) representing the observed response variable: 40 observations from 2 replications of the 20 location-chemical treatment combinations

τ_i – the random effect due to the location treatments: 5 treatment levels chosen from a continuous distribution of locations

β_i – the fixed effect due to the chemical treatments: 4 treatment levels

$\tau\beta_{ij}$ – the random interaction effect: 20 interaction levels

ε_{ij} – random error associated with each location-chemical combination: 40 residual errors

b) Compute the sum of squares for this experiment and report this value in an AOV table. Be sure to include the expected mean squares column in the AOV table.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Model	19	200.1027500	10.5317237	30.42	<.0001
Locations	4	3.811500	0.952875	0.71	0.6020
Chemicals	3	180.132750	60.044250	44.59	<.0001
Locations*Chemicals	12	16.158500	1.346542	3.89	0.0037
Error	20	6.9250000	0.3462500		
Corrected Total	39	207.0277500			

The expected mean square of the model is $SS_{model} = 200.103$.

17.11

Refer to Exercise 17.10. Perform an analysis of variance. Draw your conclusions, using $\alpha = 0.05$.

Analysis of variance test for the locations-chemicals interaction:

Hypotheses:

$$H_0: {\sigma_{\tau\beta}}^2 = 0$$

$$H_a: {\sigma_{\tau\beta}}^2 > 0$$

Test statistic:

$$F = \frac{MSLC}{MSE} = \frac{1.3465}{0.3463} = 3.89$$

Rejection region:

At $\alpha = 0.05$, we have $F_{12,20}(0.05) = 2.28$. We reject H_0 for $F > 2.28$.

Conclusion:

Since $F = 3.89 > 2.28$, we reject H_0 and conclude that the interaction between the chemicals and locations significantly affects the number of ants killed.

Analysis of variance test for the locations factor:

Hypotheses:

$$H_0: {\sigma_\tau}^2 = 0$$

$$H_a: {\sigma_\tau}^2 > 0$$

Test statistic:

$$F = \frac{MSL}{MSLC} = \frac{0.9529}{1.3465} = 0.71$$

Rejection region:

At $\alpha = 0.05$, we have $F_{4,12}(0.05) = 3.259$. We reject H_0 for $F > 3.26$.

Conclusion:

Since $F = 0.71 < 3.259$, we fail to reject H_0 and conclude that the chemicals factor does not significantly affect the number of ants killed.

Analysis of variance test for the chemicals factor:

Hypotheses:

$$H_0: {\sigma_\beta}^2 = 0$$

$$H_a: {\sigma_\beta}^2 > 0$$

Test statistic:

$$F = \frac{MSC}{MSLC} = \frac{60.0443}{1.3465} = 44.59$$

Rejection region:

At $\alpha = 0.05$, we have $F_{3,12}(0.05) = 3.49$. We reject H_0 for $F > 3.49$.

Conclusion:

Since $F = 44.59 > 3.49$, we reject H_0 and conclude that the chemicals factor significantly affects the number of ants killed.

Overall, since the interaction is significant, we conclude that we may only discuss the interaction effect on mean number of ants killed.

17.27

The two most crucial factors that influence the strength of solders use in cementing computer chips into the mother board of the guidance system of an airplane are identified as the machine used to insert the solder and the operator of the machine. Four solder machines and three operators were randomly selected from the many machines and operators available at the company's plants. Each operator made two solders on each of the four machines. The resulting strength determinations of the solders are given here.

a) Write a model for this study. Include all terms and conditions placed on the terms in the model.

Random effect factorial design model:

$$\boldsymbol{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}}$$

$$i = 1, 2, 3, ;\ j = 1, 2, 3,4;\ k = 1, 2$$

$y_{ij}k$ – the solder strength, our response variable: 24 observations from 2 replications of the 12 operator-machine treatment combinations

τ_i – the random effect due to the operator treatments: 3 treatment levels chosen from an essentially continuous distribution of operations

β_i – the random effect due to the machine treatments: 4 treatment levels chosen from an essentially continuous distribution of machines

$\tau\beta_{ij}$ – the random interaction effect: 12 interaction levels

ε_{ij} – random error associated with each operator-machine combination: 24 residual errors

b) Present the AOV table for this study and include the expected mean squares.

Source	DF	Type III SS	Mean Square	EMS	F Value	Pr > F
Operator	2	160.333333	80.166667	${\sigma_\varepsilon}^2 + 2{\sigma_{\tau\beta}}^2 + 8{\sigma_\tau}^2$	10.77	0.0103
Machine	3	12.458333	4.152778	${\sigma_\varepsilon}^2 + 2{\sigma_{\tau\beta}}^2 + 6{\sigma_\beta}^2$	0.56	0.6619
Operator*Machine	6	44.666667	7.444444	${\sigma_\varepsilon}^2 + 2{\sigma_{\tau\beta}}^2$	1.96	0.1507
Error	12	45.500000	3.791667	${\sigma_\varepsilon}^2$		

c) What conclusions can you make about the effect of machine and operator on the variability in solder strength?

Since the p-value for the interaction effect between the operator factor and machine factor is 0.1507, we conclude that the interaction does not significantly affect solder strength. From the p-value of 0.6619, we see that the machine factor also does not significantly affect solder strength. Lastly, the p-value of 0.0103 shows that the operator factor significantly affects solder strength.

17.28

Refer to Exercise 17.27.

a) Estimate the variance components in this study.

We have

$$MS_{operator} = 80.166667$$
$$MS_{machine} = 4.152778$$
$$MS_{operator*machine} = 7.444444$$
$$MS_{error} = 3.791667$$

and since both factors are random with $n = 2$, $a = 3$, $b = 4$,

$$MS_{operator} = {\sigma_\varepsilon}^2 + n{\sigma_{\tau\beta}}^2 + bn{\sigma_\tau}^2$$
$$MS_{machine} = {\sigma_\varepsilon}^2 + n{\sigma_{\tau\beta}}^2 + an{\sigma_\beta}^2$$
$$MS_{machine*operator} = {\sigma_\varepsilon}^2 + n{\sigma_{\tau\beta}}^2$$
$$MS_{error} = {\sigma_\varepsilon}^2$$

Thus,

$${\hat{\sigma}_\varepsilon}^2 = 3.79$$
$$7.44 = 3.79 + 2{\hat{\sigma}_{\tau\beta}}^2 \rightarrow {\hat{\sigma}_{\tau\beta}}^2 = 1.83$$
$$80.17 = 3.79 + 2(1.83) + (4)(2){\hat{\sigma}_\tau}^2 \rightarrow {\hat{\sigma}_\tau}^2 = 9.09$$
$$4.15 = 3.79 + 2(1.83) + (3)(2){\hat{\sigma}_\beta}^2 \rightarrow {\hat{\sigma}_\beta}^2 = -0.55$$

b) Proportionally allocate the sources of variability with respect to the total variability in solder strength.

Source of Variance	Estimator	Proportion of Total
Operator (O)	9.09	$\frac{9.09}{14.16} = 0.642$
Machine (M)	-0.55	$\frac{0.55}{14.16} = 0.039$
Operator*Machine (OM)	1.83	$\frac{1.83}{14.16} = 0.129$
Error	3.79	$\frac{3.79}{14.16} = 0.268$
Total	14.16	$\frac{14.16}{14.16} = 1.000$

c) Place a 95% confidence interval on the average solder strength.

Firstly, we have

$$\bar{y}_{...} = 207.292$$

$$df_{approx} = \left\lfloor \frac{(MSO + MSM - MSOM)^2}{(MSO)^2/(a-1) + (MSM)^2/(b-1) + (MSOM)^2/[(a-1)(b-1)]} \right\rfloor$$

$$= \left\lfloor \frac{(80.17 + 4.15 - 7.44)^2}{(80.17)^2/(3-1) + (4.15)^2/(4-1) + (7.44)^2/[(3-1)(4-1)]} \right\rfloor$$

$$= 1$$

$$df_{approx} = 1$$

$$SE(\hat{\mu}) = \sqrt{\frac{MSO + MSM - MSOM}{abn}}$$

$$= \sqrt{\frac{801.7 + 4.15 - 7.44}{3 \cdot 4 \cdot 2}}$$

$$= 1.790$$

$$SE(\hat{\mu}) = 1.790$$

The confidence interval is given by,

$$\bar{y}_{...} \pm t_{\alpha/2, df_{approx}} \cdot SE(\hat{\mu}) = 207.292 \pm 12.706 \cdot 1.790$$

Therefore the 95% confidence interval is $[184.548, 230.036]$.

SAS code:

```
*** Excersice 17.10

* input data;
data chemicalants;
        input Locations$        Chemicals$      NumberKilled;
        cards;
1       1       7.2
1       1       9.6
2       1       8.5
2       1       9.6
3       1       9.1
3       1       8.6
4       1       8.2
4       1       9
5       1       7.8
5       1       8
1       2       4.2
1       2       3.5
2       2       2.9
2       2       3.3
3       2       1.8
3       2       2.4
4       2       3.6
4       2       4.4
5       2       3.7
5       2       3.9
1       3       9.5
1       3       9.3
2       3       8.8
2       3       9.2
3       3       7.6
3       3       7.1
4       3       7.3
4       3       7
5       3       9.2
5       3       8.3
1       4       5.4
1       4       3.9
2       4       6.3
2       4       6
3       4       6.1
3       4       5.6
4       4       5
4       4       5.4
5       4       6.5
5       4       6.9
;
run;

/*proc print data=chemicalants;
run;*/

* fixed, random effect model using proc glm;
proc glm data=chemicalants;
        class locations chemicals;
        model numberkilled = locations | chemicals / solution;
        random locations locations*chemicals / test;
        means chemicals / lsd;
        means chemicals*locations / lsd;
        output out=residuals r=res;
run;
quit;

* mixed effect model using proc mixed;
proc mixed data = chemicalants;
        class locations chemicals;
        model numberkilled = chemicals;
        random locations locations*chemicals;
run;
quit;


proc univariate normal data=residuals;
```

```
        var res;
run;
quit;


*** Excersice 17.27

* input data;
data solders;
        input Operator  Machine Response;
        cards;
1       1       204
1       1       205
2       1       205
2       1       207
3       1       211
3       1       209
1       2       205
1       2       210
2       2       205
2       2       206
3       2       207
3       2       210
1       3       203
1       3       204
2       3       206
2       3       204
3       3       209
3       3       214
1       4       205
1       4       203
2       4       209
2       4       207
3       4       215
3       4       212
;
run;

/*proc print data=solders;
run;*/

proc glm data = solders;
        class operator machine;
        model response = operator | machine  / solution;
        random operator machine operator*machine / test;
run;
quit;
```