HW 8

2024-10-28

Stat 5000 Homework #8
Fall 2024 due Fri, November 1st @ 11:59 pm Name: Sam Olson
Collaborators: **The Hatman**

Q1

A completely randomized two-factor experiment consisted of burning fuel with levels of two additives in a laboratory setting and determining the carbon monoxide (CO) emissions released. Eighteen batches of a standard fuel were available for this study. Two of the batches were randomly assigned to each of nine combinations of two additives corresponding to three levels of added ethanol (0.1, 0.2, or 0.3) and three air/fuel ratio settings (14, 15, or 16). Units for the ethanol levels were not reported. CO emission concentrations (g/meter3) were determined by burning the same amount of fuel from each of the 18 batches. The data are shown below and are located in the file emissions.txt.

Added	Air/Fuel Ratio		
Ethanol	14	15	16
0.1	66	72	68
	60	65	64
0.2	78	80	66
	81	81	69
0.3	90	75	60
	94	78	58

Figure 1: CocoMelon

(a)

Construct the full ANOVA table. Which factors or interactions have significant effects on CO concentrations in emissions? Interpret the results in the context of the study.

(b)

Partition the sum of squares for the ethanol effects, averaging across air/fuel ratio levels, into sums of squares for linear and quadratic components. The coefficients for these contrasts are (-1, 0, 1) and (-1, 2, -1). Is there a significant linear or quadratic effect in the model for the ethanol effects?

(c)

Partition the sum of squares for the air/fuel ratio effects, averaging across levels of ethanol, into sums of squares for linear and quadratic components. The coefficients for these contrasts are (-1, 0, 1) and (-1, 2, -1). Is there a significant linear or quadratic effect in the model for the air/fuel ratio effects?

(d)

Use Tukey's HSD method to make pairwise comparisons of the marginal means for the three ethanol values. Summarize the results in the context of the study.

(e)

Use Tukey's HSD method to make pairwise comparisons of the marginal means for the air/fuel ratio values. Summarize the results in the context of the study.

Q2

In a study of the effects of exposure to UV-B radiation on egg hatch rates for three species of frogs, eggs were collected from two different locations (Three Creek and Sparks Lake) and then subjected to UV-B radiation using three different kinds of filters. Thirty-six enclosures were constructed at each location. Within each location, four enclosures were randomly assigned to each of the 9 combination of the two factors: frog species (Hyla regilla, Rana cascade, and Bufo boreas) and type of radiation filters (none, UV-B transmitting, and UV-B blocking). One hundred and fifty eggs for the designated frog species were placed in each enclosure. The response is the percentage of eggs that failed to hatch in each enclosure. The data is posted in the frogeggs.txt file and displayed in the following tables:

Data for Three Creek Location

Type of Filter (Factor A)	Frog Species (Factor B)		
	Hyla regilla ($j = 1$)	*Rana cascade* ($j = 2$)	*Bufo boreas* ($j = 3$)
None ($i = 1$)	6.0	38.7	42.0
	4.7	44.0	50.7
	0.7	30.0	32.7
	5.2	38.7	44.0
UV-B Transmitting ($i = 2$)	0.9	28.7	47.3
	6.7	32.7	22.0
	2.7	36.0	37.2
	0.7	40.7	43.3
UV-B Blocking ($i = 3$)	4.7	25.3	18.7
	0.7	18.7	17.3
	4.7	21.3	16.0
	0.7	16.7	4.7

Data for Sparks Lake Location

Type of Filter (Factor A)	Frog Species (Factor B)		
	Hyla regilla ($j = 1$)	*Rana cascade* ($j = 2$)	*Bufo boreas* ($j = 3$)
None ($i = 1$)	1.5	36.7	54.0
	0.8	69.6	54.7
	2.9	39.3	48.0
	3.9	34.0	36.7
UV-B Transmitting ($i = 2$)	0.7	70.0	46.0
	2.1	54.0	46.7
	0.0	48.7	36.0
	1.4	51.3	35.3
UV-B Blocking ($i = 3$)	4.5	24.7	12.7
	0.0	25.3	17.3
	0.0	39.3	31.3
	0.0	32.7	17.3

Figure 2: CocoMelon

(a)

What is the treatment design and what is the experimental design in this study?

(b)

Consider the model $Y_{ijkl} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \beta_k + \epsilon_{ijkl}$ where $\epsilon_{ijkl} \sim N(0, sigma^2)$ are random errors, $\beta_k \sim N(0, \sigma^2)$ are random block effects corresponding to locations, and any random error is independent of any random block effect. Imposing the baseline constraints $\alpha_3 = \tau_3 = (\alpha\tau)_{13} = (\alpha\tau)_{23} = (\alpha\tau)_{33} = (\alpha\tau)_{31} = (\alpha\tau)_{32} = 0$ then interpret the following parameters in the context of the study:

i.

μ

ii.

α_1

iii.

τ_2

iv.

$(\alpha\tau)_{12}$

v.

$\mu + \alpha_1 + \tau_2 + (\alpha\tau)_{12}$

vi.

$(\alpha\tau)_{12} - (\alpha\tau)_{32} - (\alpha\tau)_{13} + (\alpha\tau)_{33}$

(c)

Examine the equal variance assumption. Summarize your findings and include supporting tables and/or figures.

(d)

Examine the normality assumption. Summarize your findings and include supporting tables and/or figures

(e)

Suppose that the diagnostics suggest the need for a transformation. Find which transformation of the responses is better, square root transformation, log transformation, or none? Summarize your findings and include supporting tables and/or figures.

(f)

For the best model specified in part (e), find the full ANOVA table. Summarize which factors and interactions are significant. Is there any evidence that the types of filter have different effects on egg hatch success? Explain.

(g)

For the best model specified in part (e): Examine a profile plot of the treatment means (do not hand it in), plotting the sample mean responses for the combinations of filters and frog species, averaging across locations. What does this plot suggest? Are your conclusions about interactions between types of filters and frog species supported by results in the ANOVA table?

Q3

The data shown in the table below are results from a study of amylace activity of malted wheat flour (Geddes, et al. 1941, Cereal Chem 18, 42-60.). Five factors, each at two levels, were examined:

Factor s: type/species of wheat Amber durum (1) hard red spring (2)

Factor p: wheat protein content low (1) high (2)

Factor m: wheat moisture content 40 percent (1) 44 percent (2)

Factor g: germination time 3 days (1) 5 days (2)

Factor k: kiln temperature rising 100F to 130F (1) constant at 100F (2)

Response: Amylace is a protein that helps you break down carbohydrates and starches into sugar, releasing carbon dioxide (CO_2) in the process. Amylase activity was measured by the amount of malt from each flour that was required to produce 204.7ml of CO_2. Measured amylase activity is reported in the data table in units of $Y = [0.6 + \log(\text{amount of malt})] \times 103$

Obs	species	protein	moisture	germination	kilntemp	activity
1	1	1	1	1	1	732
2	2	1	1	1	1	801
3	1	2	1	1	1	717
4	2	2	1	1	1	791
5	1	1	2	1	1	616
6	2	1	2	1	1	787
7	1	2	2	1	1	540
8	2	2	2	1	1	669
9	1	1	1	2	1	200
10	2	1	1	2	1	50
11	1	2	1	2	1	292
12	2	2	1	2	1	74
13	1	1	2	2	1	62
14	2	1	2	2	1	83
15	1	2	2	2	1	97
16	2	2	2	2	1	-9
17	1	1	1	1	2	744
18	2	1	1	1	2	732
19	1	2	1	1	2	713
20	2	2	1	1	2	746
21	1	1	2	1	2	569
22	2	1	2	1	2	785
23	1	2	2	1	2	486
24	2	2	2	1	2	544
25	1	1	1	2	2	253
26	2	1	1	2	2	91
27	1	2	1	2	2	265
28	2	2	1	2	2	147
29	1	1	2	2	2	80
30	2	1	2	2	2	80
31	1	2	2	2	2	102
32	2	2	2	2	2	-40

Figure 3: CocoMelon

(a)

The normal probability plot and table of estimates on the next page shows the values of main effects and interaction contrasts, for which the estimate of every contrast has the same variance. This information is used to determine which effects should be included in the analysis and which should be used to estimate the variance. Which effects appear to be large?

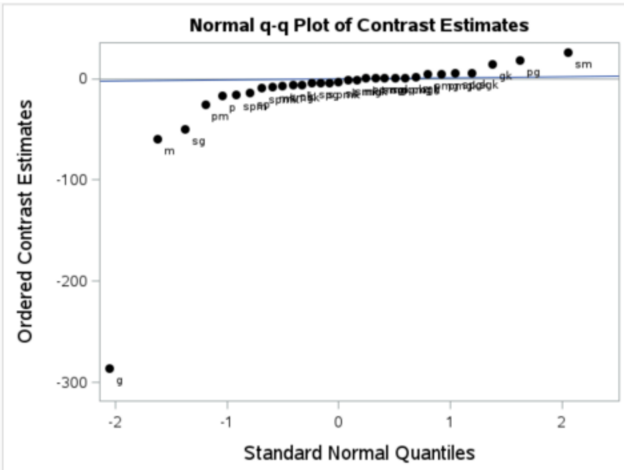

Figure 4: CocoMelon

Obs	Dependent	Parameter	Estimate
1	yield	g	-285.7812500
2	yield	m	-59.2812500
3	yield	sg	-50.4062500
4	yield	pm	-25.4687500
5	yield	p	-16.5937500
6	yield	spm	-15.4687500
7	yield	sp	-13.8437500
8	yield	spmk	-8.8437500
9	yield	mk	-8.5312500
10	yield	smgk	-7.5312500
11	yield	pk	-6.5937500
12	yield	k	-6.4062500
13	yield	spg	-4.4687500
14	yield	s	-4.2812500
15	yield	pmk	-4.0937500
16	yield	sk	-3.6562500
17	yield	smk	-1.5312500
18	yield	mgk	-0.9062500
19	yield	spmg	0.1562500
20	yield	smg	0.3437500
21	yield	spk	0.6562500
22	yield	spmgk	0.9062500
23	yield	pkg	1.0312500
24	yield	mg	1.9687500
25	yield	pmg	4.2812500
26	yield	pmgk	4.5312500
27	yield	spgk	5.4062500
28	yield	sgk	5.5937500
29	yield	gk	14.4687500
30	yield	pg	18.4062500
31	yield	sm	25.9687500

Figure 5: CocoMelon

(b)

Using least squares estimation to fit the model that includes all main effects and all interaction effects that were identified as "non-zero" by the analysis in part (a), (including all main effects in this model, regardless of whether the plot suggests they are significant or not, then the sum of sums of squares for the interaction contrasts that are not included in the model can be pooled to obtain a MSerror), the corresponding ANOVA table is provided below.

Examine the results of F-tests for terms kept in the model and summarize the results in the context of the study

(c)

Choose any significant two-way interaction for the model in part (b) and interpret it in the context of the study. Also interpret the significant three-way interaction for the model in part (b).

(d)

Comment on the normal probability plot of the residuals for the model in part (b), shown below.

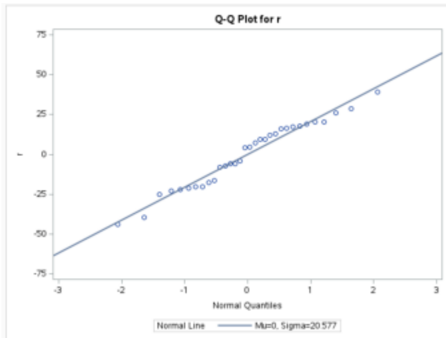

Figure 6: CocoMelon

(e)

Comment on the plot of the residuals versus the estimated mean yields for the model in part (b), shown below.

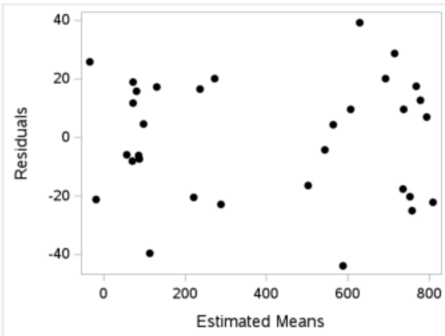

Figure 7: CocoMelon

(f)

Interpret the value of each of the estimated effects of the five factors on amylase activity, shown below. Keep in mind that low values of the response variable correspond to combinations of factors that produce 204.7 ml of CO2 with the least amount of malt.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-18.7187500	B	16.75233039	-1.12	0.2778
species 1	116.0625000	B	20.77870854	5.59	<.0001
species 2	0.0000000	B	.	.	.
protein 1	105.9375000	B	20.77870854	5.10	<.0001
protein 2	0.0000000	B	.	.	.
moisture 1	148.5000000	B	18.58504191	7.99	<.0001
moisture 2	0.0000000	B	.	.	.
germination 1	606.6250000	B	18.58504191	32.64	<.0001
germination 2	0.0000000	B	.	.	.
kilntemp 1	-16.1250000	B	13.14160917	-1.23	0.2348
kilntemp 2	0.0000000	B	.	.	.

Figure 8: CocoMelon