MATH 581: Experimental Design and Analysis

Sample MIDTERM

Name Key

Note: To receive full credits you need to show all your work. You may use two pages of notes, tables and a calculator but no other reference materials. Talking during the exam will be considered as cheating. Use separate papers if necessary. The exam is exactly 1 and a half hour (no exception).

1. A manufacturer has conducted an experiment to investigate the effect of the level of supply of raw material (Supply) and the ratio of its assignment (Ratio) to the two product manufacturing lines on the profit per unit of raw materials. The ultimate goal was to be able to choose the best ratio to match each day's supply of raw materials. The levels of supply of the raw material chosen for the experiment were 15, 18, and 21 tons. The levels of the ratio of allocation to the two product lines were ½, 1, and 2. The response was the profit (in cents) per unit of raw material supply obtained from a single day's production. Three replications of each combination were conducted in a random sequence. The data for the 27 days and the SAS (Proc GLM) output of ANOVA are shown in the followings.

Ratio of Raw Material Allocation	Raw Material Supply					
	15	$\bar{Y}_{ij.}$	18	$\bar{Y}_{ij.}$	21	$\bar{Y}_{ij.}$
½	22, 20, 21	21	21, 19, 20	20	19, 18, 20	19
1	21, 20, 19	20	23, 24, 22	23	20, 19, 21	20
2	17, 18, 16	17	21, 11, 20	17.3	20, 22, 24	22

SAS Output #1

Dependent Variable: profit

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	93.1851852	11.6481481	2.54	0.0482
Error	18	82.6666667	4.5925926		
Corrected Total	26	175.8518519			

R-Square	Coeff Var	Root MSE	profit Mean
0.529907	10.75500	2.143034	19.92593

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ratio	2	22.29629630	11.14814815	2.43	0.1166
supply	2	4.96296296	2.48148148	0.54	0.5917
ratio*supply	4	65.92592593	16.48148148	3.59	0.0255

Level of ratio	Level of supply	N	profit Mean	profit Std Dev

0.5	15	3	21.0000000	1.00000000
0.5	18	3	20.0000000	1.00000000
0.5	21	3	19.0000000	1.00000000
1	15	3	20.0000000	1.00000000
1	18	3	23.0000000	1.00000000
1	21	3	20.0000000	1.00000000
2	15	3	17.0000000	1.00000000
2	18	3	17.3333333	5.50757055
2	21	3	22.0000000	2.00000000

(a) Draw conclusions based on the analysis of variance shown above. Use 0.05 level of the significance

Two main effects (ratio and supply) are not significant at $\alpha = .05$ as p-values are greater than α.

The interaction effect is significant with $p = .0255 < .05$

(b) Identify the two best combinations of Supply and Ratio. Are theses two combinations significantly different? Use the Tukey procedure that limits the error rate of all pairwise comparisons of combinations to be 0.05.

$\bar{Y}_{22.} = 23$, $\bar{Y}_{33.} = 22$, $MSE = 4.59$

$D = \mu_{22} - \mu_{33}$ | $T = \frac{1}{\sqrt{2}} q(.95; 9, 18) = \frac{1}{\sqrt{2}}(4.96) = 3.51$

$\hat{D} = \bar{Y}_{22.} - \bar{Y}_{33.} = 1$ | $\hat{D} \pm T \cdot s(\hat{D}) = 1 \pm 3.51(\sqrt{3.06}) = 1 \pm 6.1$

$s^2(\hat{D}) = 2MSE/3 = 3.06$ | Two treatments are not significantly diff.

(c) Mr. Flippantly, the experimenter, made a mistake by fitting an ANOVA model without interactions and drew a wrong conclusion. Fill in the ANOVA table that Mr. Flippantly would have. What would be his conclusion?

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	27.26	6.815	1.01	> .1
Error	22	148.59	6.754		
Corrected Total	26	175.85			

2. Consider the data in problem 1. Since the Supply factor do not seem to be significant, the experimenter decided to consider the Ratio as only factor in the study (one-factor ANOVA). At this time, suppose that the three levels of ~~Supply~~ Ratio factor were chosen randomly from many different levels.

(a) Perform a hypothesis testing for the factor effect. Write the hypotheses carefully.

$Y_{ij} = \mu_i + \varepsilon_{ij}$, $\mu_i \sim N(\mu_., \sigma_\mu^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, μ_i & ε_{ij} indep.

$H_0: \sigma_\mu^2 = 0$ vs $H_1: \sigma_\mu^2 > 0$ | $F = \frac{MSTR}{MSE} = 1.75$

$SSTR = 22.3$; $MSTR = 22.3/2 = 11.15$

$SSE = 175.85 - 22.30 = 152.7$, $MSE = 152.7/24 = 6.36$; p-value > .1 $\Rightarrow$ We fail to reject H_0

$F(.975, 2, 24) = 4.32$

$F(.025, 2, 24) = \frac{1}{F(.975, 24, 2)} = \frac{1}{39.5}$

(b) Let σ_μ^2 and σ^2 be population variance for the factor effect and the error, respectively. Construct 95% confidence intervals for σ_μ^2/σ^2 and for σ^2.

95% CI for σ_μ^2/σ^2 : (L, U) $\approx [0, 7.57]$

$L = \frac{1}{9}\left[\frac{11.15}{6.36}\left(\frac{1}{4.32}\right) - 1\right] = -.066$

$U = \frac{1}{9}\left[\frac{11.15}{6.35}(39.5) - 1\right] = 7.57$

$\chi^2_{.975}(24) = 39.36$

$\chi^2_{.025}(24) = 12.4$

95% CI for σ^2 : $\left(\frac{24(6.36)}{39.36}, \frac{24(6.36)}{12.4}\right) = (3.88, 12.31)$

(c) Using the lower and upper bounds in (b). Construct an approximate 95% confidence interval for σ_μ^2.

CI for $\frac{\sigma_\mu^2}{\sigma^2} = (0, 7.57)$

~~CI for σ^2 = (3.88, 12.34)~~

$\hat{\sigma}^2 = MSE = 6.36$

Approx. 95% CI for σ_μ^2 : $(0,\ 7.57(6.36)) = (0, 48)$

3. The marketing research group of a corporation examined the public response to the introduction of a new TV game module by comparing weekly sales volumes (in thousand dollars) for three different store chains.

Week	Chain		
	1	2	3
1-Week	35 42 35	17 30 35 43	7 22 15
2-Week	30 48 38 26	22 28 40	12 19 20 23

As we have discussed in class, an appropriate regression model can be written as

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk}$$

where

$$X_{ijk1} = \begin{cases} 1 & if\ i=1 \\ -1 & if\ i=2 \end{cases},\quad X_{ijk2} = \begin{cases} 1 & if\ j=1 \\ 0 & if\ j=2 \\ -1 & if\ j=3 \end{cases},\quad X_{ijk3} = \begin{cases} 1 & if\ j=2 \\ 0 & if\ j=1 \\ -1 & if\ j=3 \end{cases}$$

The SAS output of the regression analysis is given below

SAS Output #2

The GLM Procedure

Dependent Variable: sales

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1434.869048	286.973810	4.23	0.0134
Error	15	1018.083333	67.872222		
Corrected Total	20	2452.952381			

R-Square	Coeff Var	Root MSE	sales Mean
0.584956	29.47320	8.238460	27.95238

Source	DF	Type I SS	Mean Square	F Value	Pr > F
X1	1	0.416017	0.416017	0.01	0.9386
X2	1	1321.142857	1321.142857	19.47	0.0005 ✓
X3	1	80.000649	80.000649	1.18	0.2948
X1X2	1	27.523810	27.523810	0.41	0.5338
X1X3	1	5.785714	5.785714	0.09	0.7743

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	27.87500000	1.81640968	15.35	<.0001 ✓
X1	-0.12500000	1.81640968	-0.07	0.9460
X2	8.54166667	2.56879121	3.33	0.0046 ✓
X3	2.75000000	2.56879121	1.07	0.3013
X1X2	1.04166667	2.56879121	0.41	0.6908
X1X3	0.75000000	2.56879121	0.29	0.7743

(a) Write the fitted regression line and calculate the estimated sales for 1-week and chain 2 under the regression model. Also calculate that under the cell means model. What did you find? Discuss.

$$\hat{Y}_{ijk} = 27.875 - .125\, X_{ijk1} + 8.54\, X_{ijk2} + 2.75\, X_{ijk3} + 1.04\, X_{ijk1} \cdot X_{ijk2} + 0.75\, X_{ijk1} X_{ijk3}$$

$X_1 = 1$
$X_2 = 0$
$X_3 = 1$

$$\hat{Y}_{12} = 27.875 - .125 + 2.75 + 0.75 = 31.25$$

Cell means $\hat{Y}_{12} = \dfrac{17+30+35+43}{4} = 31.25$

"Same"

(b) From the output above SAS Output #2, which factor seems to be the most significant? Write a reduced model with the factor of your choice only and perform a general linear F-test to see if the reduced model is acceptable.

Reduced model (X_2 only) : $Y_{ijk} = \mu_{..} + \beta_1 X_{ijk2} + \varepsilon_{ijk}$

$SSE(R) = 1018.08 + 0.416 + 80.0 + 27.52 + 5.79 = 1131.8$
$df = 19$ $(= 2452.95 - 1321.14)$

$$F^* = \frac{\dfrac{SSE(R) - SSE(F)}{19-15}}{\dfrac{SSE(F)}{15}} = \frac{\dfrac{1131.8 - 1018.08}{4}}{\dfrac{1018.08}{15}} = 0.42$$

df 4, 15

p-value > 0.5 (huge !!)

⇒ The reduced model is acceptable