Final Exam

Cody Frisby

April 26, 2016

1

- The model for this experiment can be described by

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl} \begin{cases} i = 1, 2 \\ j = 1, 2 \\ k = 1, 2 \\ l = 1, 2, 3, 4 \end{cases}$$

  where it is assumed that A, B, and C are fixed and ε_{ijkl} are NID(0, σ^2).

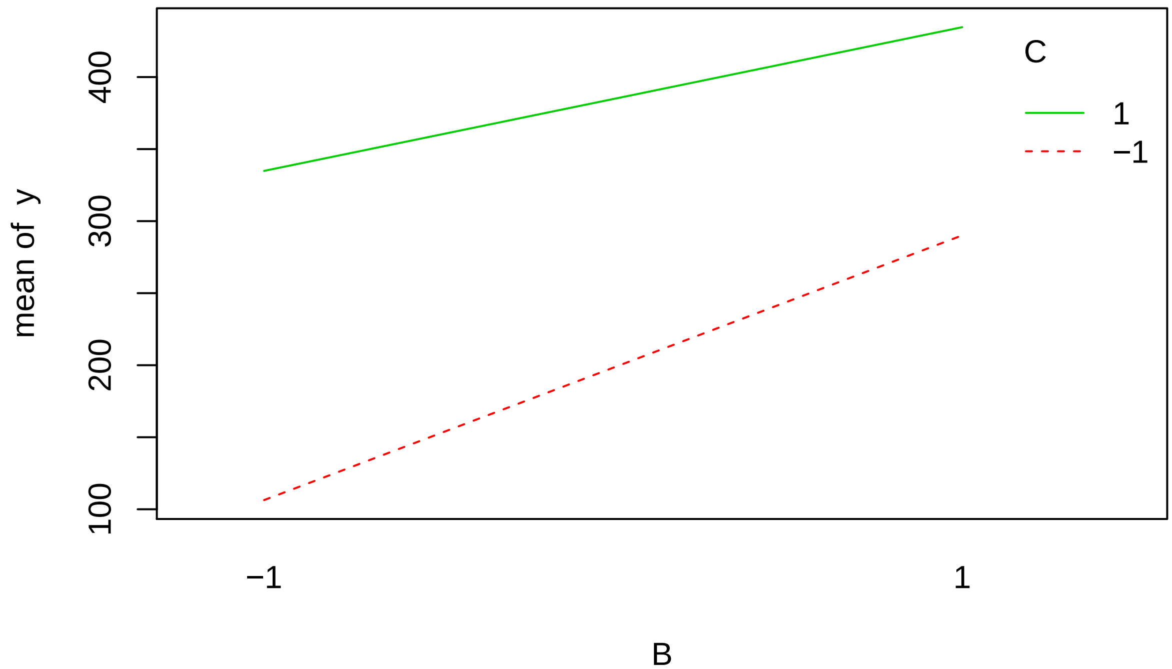
- The main and interaction effects are

Factor	effect
A	21.00
B	141.75
C	186.50
AB	6.75
AC	10.25
BC	-42.00
ABC	-3.75

- And the ANOVA table

Factor	Df	SS	MS	F0	p
A	1	3528.0	3528.00	10.3385	0.00370
B	1	160744.5	160744.50	471.0462	0.00000
C	1	278258.0	278258.00	815.4081	0.00000
AB	1	364.5	364.50	1.0681	0.31168
AC	1	840.5	840.50	2.4630	0.12965
BC	1	14112.0	14112.00	41.3538	0.00000
ABC	1	112.5	112.50	0.3297	0.57118
Error	24	8190.0	341.25	NA	NA
Total	31	466150.0	NA	NA	NA

- Here, I started to construct interaction plots of the effects of A, B, and C where I seperated the two plots into A low and A high, but seeing as there is no significant interaction effect from A, I will just plot the interaction between B and C.



There does not appear to be any evidence of a significant interaction effect from A. There is strong evidence that B & C have the largest effect on the response. And I would set B and C at their respective high levels to maximize the response. We could also set A at "high" since its effect is positive.

2

- A model that relates y to x_1 and x_2 that is supported by the design of this experiment is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,3 \\ j = 1,2,3 \\ k = 1,2 \end{cases}$$

  where the errors are assumed NID(0, σ^2). Also, the addition of a third factor level allows us to model the relationship between the response and design factors as quadratic.

- The linear and quadradic effects of our model, where A = **temp** and B = **pressure**, are

	Al	Bl	Aq	Bq	AlBl	AlBq	AqBl	AqBq
effects	24.75	10.65	5.05	11.95	9.25	3.45	8.55	7.75

- ANOVA table is displayed below

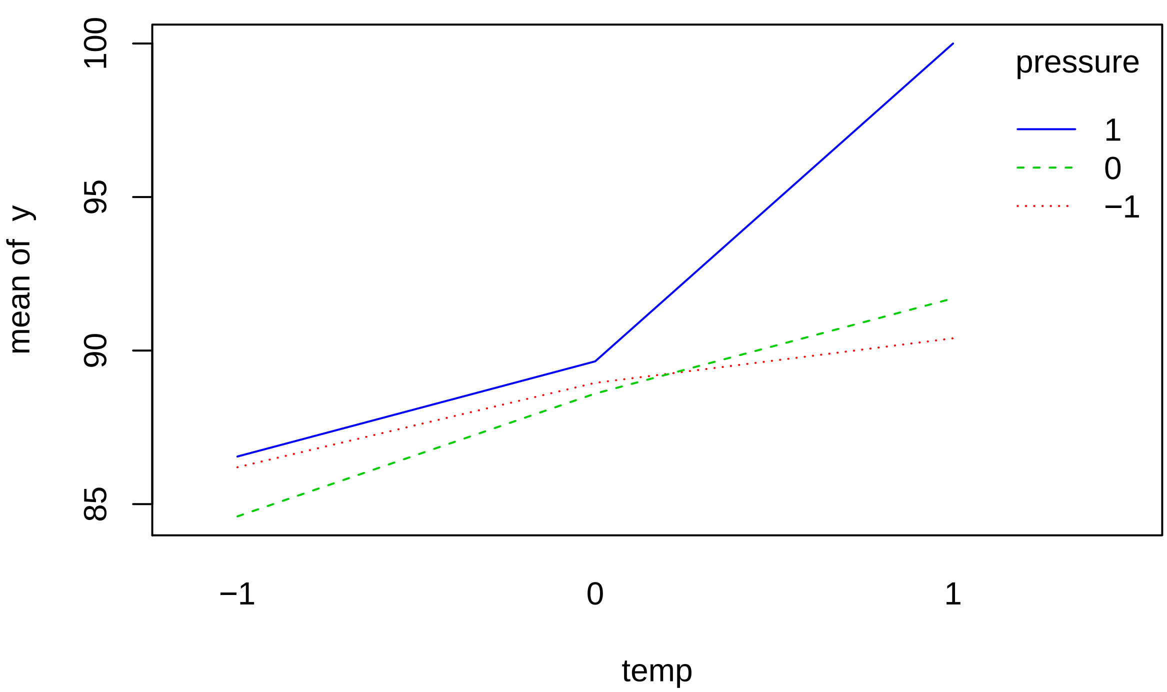
Source	SS	Df	MS	F_0	P-Value
temp	207.021	2	103.511	52.469	0
pressure	53.674	2	26.837	13.604	0.0019
temp:pressure	60.286	4	15.071	7.64	0.0057
Error	17.755	9	1.973		
Total	338.736	17			

and we conclude all effects are significant. Now let's take a look at the ANOVA with linear and quadratic effects.

Source	Sum Sq	Df	Mean Sq	F value	Pr(>F)
temp	204.188	1	204.188	103.4911	0
pressure	37.808	1	37.808	19.1627	0.002
$temp^2$	2.834	1	2.834	1.4364	0.261
$pressure^2$	15.867	1	15.867	8.0421	0.02
temp:pressure	42.781	1	42.781	21.6832	0.001
$pressure^2 : temp$	1.984	1	1.984	1.0056	0.342
$temp^2 : pressure$	12.184	1	12.184	6.1754	0.035
$pressure^2 : temp^2$	3.337	1	3.337	1.6913	0.226
Error	17.753	9	1.973		
Total	338.736	17			

And we conclude that all effects are significant *except* A_Q, $A_L B_Q$, and $A_Q B_Q$, at $\alpha = 0.05$.

- There doesn't appear to be much evidence for curvature with temperature. We can most likely estimate temp's effect effectively with only linear terms, thereby removing all quadratic temp terms. Looking at an interaction plot we can see that there is some curvature with pressure and definite interaction between pressure and temperature. The effect from pressure, 53.674, can be dicected into its linear and quadratic components. The proportion linear is $\frac{37.808}{53.674} = 0.7044006$. And the proportion quadratic would be, 0.2955994. At this point it would be up to the engineer to decide whether or not they are ok with only a linear term for pressure or if they feel like they need the quadratic term in the model, I would lean towards including it. Additionaly, for the interaction term, $\frac{42.781}{60.286} = 0.7096341$ can be explained from the linear term. Most of the remaining can be found in the **AqBl** term. Again, we may need to consult with the scienctist/engineer to understand whether or not previous theory indicates this type of interaction or if they are comfortable excluding/including this term from/in the model since A_Q is not significant by itself. Below, looking at the interaction plot provides some insight to the above analysis regarding linear and quadratic effects and their interactions.



3

- The model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, 3 \\ j = 1, 2, ..., 11 \end{cases}$$

  where the errors and τ_i are assumed to be independent random variables that are normally distributed with mean zero and where the variance of any observation is given by $V(y_{ij}) = \sigma_\tau^2 + \sigma^2$.

- Yes, the variability of flavor melt time is significant. Table 6 below contains the ANOVA results of this experiment.

Table 6: Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
flavor	2	173010	86505	12.76	9.799e-05
Residuals	30	203456	6782	NA	NA

- The point estimate for $\hat{\sigma}_\tau^2$ is computed by

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{8.6504939 \times 10^4 - 6781.8727}{11} = 7247.5515$$

- The variance of any observation melt time is estimated by

$$\hat{\sigma}_y^2 = \hat{\sigma}^2 + \hat{\sigma}_\tau^2 = 6781.8727 + 7247.5515 = 1.4029424 \times 10^4$$

- We can find a confidence interval for σ^2 by using the form

$$\frac{(N-a)MS_E}{\chi_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{(N-a)MS_E}{\chi_{1-(\alpha/2),N-a}}$$

  and the interval

$$[4330.7676174, 1.2117143 \times 10^4]$$

  is the 95% confidence interval for σ^2.

- To find a 95% confidence interval for $\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$ we find L and U

$$L = \frac{1}{n}(\frac{MS_{Treatments}}{MS_E}\frac{1}{F_{\alpha/2,a-1,N-a}} - 1)$$

$$U = \frac{1}{n}(\frac{MS_{Treatments}}{MS_E}\frac{1}{F_{1-\alpha/2,a-1,N-a}} - 1)$$

  and the interval is

$$\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2} \leq \frac{U}{1+U}$$

  and plugging into this formula we get

$$[0.158666, 0.9785746]$$

  and we conclude that variability between ice creams accounts for between 15.8 and 97.8 percent of the variability observed in melt times.

4

- The model is

$$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{cases}$$

where the α_i are fixed effecs such that $\sum \alpha_i = 0$ and $\gamma_j, (\alpha\gamma)_{ij}$, and ε_{ijk} are uncorrelated random variables have zero mean and variances: $V(\gamma_j) = \sigma_\gamma^2, V[(\alpha\gamma)_{ij}] = \sigma_{\alpha\gamma}$, and $V(\varepsilon_{ijk}) = \sigma^2$

- And computing the remaining ANOVA table using the unrestricted model we get

Source	SS	Df	MS	F_0	P-Value
Operator	39.27	2	19.635	7.2857	0.00481
Part	3935.96	9	437.3288889	162.2742	0
Operator:Part	48.51	18	2.695	5.2723	0
Error	30.67	60	0.5111667		
Total	4054.4	89			

where

$$F_0 = \frac{MS_B}{MS_{AB}}$$

and we conclude that all the effects are significant.

- The effect of operator is statistically significant, and we compute the confidence interval for the estimate of each operator using the formula

$$\sqrt{\frac{MS_{AB}}{bn}} = \sqrt{\frac{2.695}{30}} = 0.2997221$$

for the estimate of the standard error. And the critical value for our assumed normal distribution being $t_{\frac{\alpha}{2},abn-1} = 1.9869787$. And plugging into this formula we get

$$y_i \pm 0.2997221(1.9869787)$$

$$y_1 : [34.3044586, 35.4955414]$$

$$y_2 : [35.9044586, 37.0955414]$$

$$y_3 : [35.4044586, 36.5955414]$$

- Yes, the effect of part is statistically significant. And

$$E(MS_B) = \sigma^2 + n\sigma_{\alpha\gamma}^2 + an\sigma_\gamma^2$$

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\alpha\gamma}^2$$

are the expected mean square estimates for **part** and **part:operator** interaction respectively. The variance component for **part** is computed by

$$\hat{\sigma}_\gamma^2 = \frac{MS_B - MS_{AB}}{an}$$

$$48.2926543 = \frac{437.3288889 - 2.695}{3(3)}$$

And the variance component $\sigma_{\tau\beta}^2$ is estimated by

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n} = 0.7279444$$

5

- This is a **3** stage nested design with lawn nested in area and sites nested in lawn with all factors being randomly selected from the population.
- The statistical model is

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{(ijk)l} \begin{cases} i = 1, 2, ..., 8 \\ j = 1, 2, ..., 5 \\ k = 1, 2 \\ l = 1, 2 \end{cases}$$

  where ε_{ijkl} is assumed NIC(0,σ). It is also assumed that $Area_i \sim iidN(0, \sigma_A^2)$ and $Lawn_{j(i)} \sim iidN(0, \sigma_B^2)$ and $Site_{k(j(k))} \sim iidN(0, \sigma_C^2)$.

- Below I display a table containing the sum of squares, degrees of freedom, and mean squares from the experiment.

Factor	SS	df	ms
Area	1540	7	220.0
Lawn(Area)	640	32	20.0
Site(Lawn(Area))	400	40	10.0
Error	680	80	8.5
Total	3260	159	NA

- Since all factors are random, we extend similar ideas from 2 stage nested designs to this experiment where we have 3 stages, all random. For the factor area we have

$$\hat{\sigma}_\tau^2 = \frac{MS_{Area} - MS_{Lawn(Area)}}{bcn} = 10$$

  and similarly for Lawn(Area) we have

$$\hat{\sigma}_\beta^2 = \frac{MS_{Lawn(Area)} - MS_{Site(Lawn(Area))}}{cn} = 2.5$$

  and finally for Site we have

$$\hat{\sigma}_\gamma^2 = \frac{MS_{Site(Lawn(Area))} - MS_E}{n} = 0.75$$

- F ratios are found similarly to mixed models, and I also run an F test for significance with $\alpha = 0.05$, resulting in

Source	SS	Df	MS	E(MS)	F_0	P-Value
Area	1540	7	220	$\sigma^2 + 2\sigma_\gamma^2 + 4\sigma_\beta^2 + 20\sigma_\tau^2$	11	0
Lawn(Area)	640	32	20	$\sigma^2 + 2\sigma_\gamma^2 + 4\sigma_\beta^2$	2	0.01919
Site(Lawn(Area))	400	40	10	$\sigma^2 + 2\sigma_\gamma^2$	1.1765	0.26549
Error	680	80	8.5	σ^2		
Total	3260	159				

and we reject the null hypothesis that $\sigma_\tau^2 = 0$ and that $\sigma_\beta^2 = 0$

but we do not reject the hypothesis that $\sigma_\gamma^2 = 0$.

Appendix I

Showing My Work/R Code

I did all the calculations and work using R studio and knitr (using LateX) in an R markdown file. I display most the the coding here for reference.

```
############ NUMBER 1 ############
# 2^3 design
n <- 4; y <- c(425,426,1118,1283,1203,1396,1670,1807)
# build the design matrix
A <- c(-1,1,-1,-1,1,1,-1,1); B <- c(-1,-1,1,-1,1,-1,1,1)
C <- c(-1,-1,-1,1,-1,1,1,1);AB <- A*B; AC <- A*C; BC <- B*C;
ABC <- A*B*C; dm <- cbind(A,B,C,AB,AC,BC,ABC)
con <- apply(dm, 2, FUN = function(x) sum(x*y))
effects <- con/16
# effects
id <- c("(1)", "a", "b", "c", "ab", "ac", "bc", "abc")
ss <- round((con^2)/32, 3) # sum of squares
# note, sst and sse cannot be estimated percisely
sst <- 466150; sse <- sst - sum(ss); mse <- sse/24
f <- round(ss/mse,4); p <- round(pf(f,1,24, lower.tail = FALSE),5)
df1 <- data.frame(Factor = colnames(dm), effect = effects,
                  Df = c(rep(1,7)),SS = ss, MS = ss, F0 = f, p)
rownames(df1) <- NULL
df3 <- df1[-2]
df3 <- rbind(df3, data.frame(Factor = c("Error", "Total"), Df = c(24,31),
        SS = c(sse, sst), MS = c(mse, NA), F0 = c(NA, NA),
        p = c(NA, NA)))
########### NUMBER 2 ###########
y <- c(86.3,84,85.8,86.1,85.2,87.3,88.5,87.3,89,89.4,
       89.9,90.3,89.1,90.2,101.3,91.7,93.2,98.7)
temp <- factor(c(rep("low",6), rep("med",6), rep("high",6)))
pressure <- factor(rep(c(250,260,270), 6))
df2 <- data.frame(y, temp, pressure)
df <- data.frame(y, temp = c(rep(-1,6), rep(0,6), rep(1,6)),
                 pressure = rep(c(-1,0,1), 6))
fit <- aov(y ~ .*., df2)
# fitting the linear model
df$temp2 <- df$temp^2
df$pressure2 <- df$pressure^2
library(dplyr)
y <- df$y
y.i <- df %>%
  group_by(temp) %>%
  summarise_each(funs(sum), y)
y.j <- df %>%
  group_by(pressure) %>%
  summarise_each(funs(sum), y)
yij. <- df %>%
  group_by(temp, pressure) %>%
```

```
  summarise_each(funs(sum), y)
y.. <- sum(y)
a <- 3; b <- 3; n <- 2; N <- a*b*n
ssa <- (1/(b*n))*sum(y.i$y^2) - y..^2/N
ssb <- (1/(a*n))*sum(y.j$y^2) - y..^2/N
sssub <- (1/n)*sum(yij.$y^2) - y..^2/N
ssab <- sssub - ssa - ssb
sst <- round(sum(y^2) - y..^2/N, 3)
sse <- sst - ssab - ssa - ssb
# now to add squared terms
row.sums <- df %>%
  group_by(temp, pressure) %>%
  summarise_each(funs(sum), y)
y <- row.sums$y
id <- c("00", "01", "02", "10", "11", "12", "20", "21", "22")
#data.frame(id, y) # these are our summaries..
#... now to develop the matrix
Al <- c(rep(-1,3), rep(0,3), rep(1,3))
Bl <- rep(c(-1,0,1),3)
Aq <- c(rep(1,3), rep(-2,3), rep(1,3))
Bq <- rep(c(1,-2,1), 3)
AlBl <- Al*Bl; AlBq <- Al*Bq; AqBl <- Aq*Bl; AqBq <- Aq*Bq
dm <- cbind(Al,Bl,Aq,Bq,AlBl,AlBq,AqBl,AqBq)
rownames(dm) <- id
m.effects <- apply(dm,2, FUN = function(d) sum(d*y)/2) # effects
# now for the sum of squares...
C <- apply(dm, 2, function(x) sum(x^2)) #contrasts
ss <- round(2*(m.effects^2)/C, 3) # sum of squares
ss2 <- round(cbind(ssa,ssb,ssab,sse,sst), 3)
sse <- round(sst - sum(ss),3); mse <- round(sse/9, 3)
############### NUMBER 3 ######################
flavor1 <- c(924,876,1150,1053,1041,1037,1125,1075,1066,977,886)
flavor2 <- c(891,982,1041,1135,1019,1093,994,960,889,967,838)
flavor3 <- c(817,1032,844,841,785,823,846,840,848,848,832)
df <- stack(data.frame(flavor1, flavor2, flavor3))
names(df) <- c("y", "flavor")
df$flavor <- as.factor(df$flavor)
fit <- aov(y ~ flavor, df)
msa <- summary(fit)[[1]][1,3]; mse <- round(summary(fit)[[1]][2,3],4); n <- 11
sig.tau <- round((msa-mse)/n,4)
a <- 3
cv <- qchisq(c(0.025, 1 - 0.025), (n*a) - a, lower.tail = FALSE)
stat <- ((n*a - a)*mse) / cv
cv2 <- qf(c(0.025, 1 - 0.025), a-1, (a*n) - 1, lower.tail = FALSE)
L.U <- (1/n)*(((msa/mse)*(1/cv2)) - 1)
l.u <- L.U/(1 + L.U)
###########  NUMBER 4 ##############
# unrestricted mixed model for number 4
ssa <- 39.27; ssb <- 3935.96; ssab <- 48.51; sse <- 30.67
sst <- 4054.4
a <- 3; b <- 10; n <- 3
msa <- ssa/(a-1); msb <- ssb/(b-1)
msab <- ssab/((a-1)*(b-1)); mse <- sse/(a*b*(n-1))
```

```r
f.op <- round(msa/msab,4); f.part <- round(msb/msab,4)
f.inter <- round(msab/mse,4)
p.op <- round(1 - pf(f.op, a-1, (a-1)*(b-1)),5)
p.part <- round(1 - pf(f.part, b-1, (a-1)*(b-1)),5)
p.inter <- round(1 - pf(f.inter, (a-1)*(b-1), a*b*(n-1)),5)
##########  NUMBER 5  ##########
# nested model for number 5
ssa <- 1540; ssb <- 640; ssc <- 400; sse <- 680; sst <- 3260
SS <- c(1540,640,400,680,3260)
a <- 8; b <- 5; c <- 2; n <- 2
Factor <- c("Area", "Lawn(Area)", "Site(Lawn(Area))", "Error", "Total")
d <- c(a-1, a*(b-1), a*b*(c-1), a*b*c*(n-1), a*b*c*n - 1)
ms <- c(ssa/(a-1), ssb/(a*(b-1)), ssc/(a*b*(c-1)), sse/(a*b*c*(n-1)), NA)
t.2 <- data.frame(Factor,SS,df = d,ms)
f <- round(c(ms[1]/ms[2], ms[2]/ms[3], ms[3]/ms[4], NA, NA),4)
sig.tau <- (ms[1] - ms[2])/(b*c*n); sig.beta <- (ms[2] - ms[3])/(c*n)
sig.gam <- (ms[3] - ms[4])/n
fc <- round(pf(f, d[1:3], d[-1], lower.tail = FALSE),5)
t.3 <- cbind(t.2, F0 = f, p = fc)
```