

1. Prove that $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{X}\mathbf{B}^{-1})$:

$$\begin{aligned}
 \mathbf{a} \in \mathcal{C}(\mathbf{X}) &\iff \underset{n \times 1}{\mathbf{a}} = \mathbf{X}\mathbf{b} && \text{for some } \underset{p \times 1}{\mathbf{b}} \\
 &\iff \mathbf{a} = \underset{p \times p}{\mathbf{X}} \mathbf{I} \mathbf{b} && \text{for some } \mathbf{b} \\
 &\iff \mathbf{a} = \mathbf{X}\mathbf{B}^{-1} \underbrace{\mathbf{B}\mathbf{b}}_{p \times 1} && \text{treat as } \mathbf{X}\mathbf{B}^{-1} \text{ product a } p \times 1 \text{ vector} \\
 &\implies \mathbf{a} \in \mathcal{C}(\mathbf{P}_\mathbf{X})
 \end{aligned}$$

So $\mathcal{C}(\mathbf{X}) \subseteq \mathcal{C}(\mathbf{X}\mathbf{B}^{-1})$.

Then similarly,

$$\begin{aligned}
 \mathbf{g} \in \mathcal{C}(\mathbf{X}\mathbf{B}^{-1}) &\iff \mathbf{g} = \mathbf{X}\mathbf{B}^{-1}\mathbf{h} && \text{for some } p \times 1 \text{ vector } \mathbf{h} \\
 &\iff \mathbf{g} = \mathbf{X} \underbrace{\mathbf{B}^{-1}\mathbf{h}}_{p \times 1} && \text{treat as } \mathbf{X} \text{ product a } p \times 1 \text{ vector} \\
 &\implies \mathbf{g} \in \mathcal{C}(\mathbf{X})
 \end{aligned}$$

So $\mathcal{C}(\mathbf{X}\mathbf{B}^{-1}) \subseteq \mathcal{C}(\mathbf{X})$.

According to the results above, $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{X}\mathbf{B}^{-1})$.

2. (a) Rewrite these hypotheses in terms of the $\boldsymbol{\alpha}$ parameter vector.

$$H_0 : \boldsymbol{\alpha}_2 = \mathbf{0} \text{ vs. } H_A : \boldsymbol{\alpha}_2 \neq \mathbf{0}$$

- (b) If we want to fit a reduced model corresponding to the null hypothesis, we could use \mathbf{W}_1 because $\mathbf{W}\boldsymbol{\alpha} = \mathbf{W}_1\boldsymbol{\alpha}_1 + \mathbf{W}_2\boldsymbol{\alpha}_2$ reduces down to $\mathbf{W}_1\boldsymbol{\alpha}_1$ under the null hypothesis $H_0 : \boldsymbol{\alpha}_2 = \mathbf{0}$.
- (c) Provide a matrix \mathbf{C} for testing the main effect of temperature.
The main effect of temperature is

$$\frac{1}{2}(\mu_{11} + \mu_{21}) - \frac{1}{2}(\mu_{12} + \mu_{22}) = \begin{pmatrix} \frac{1}{2}, & -\frac{1}{2}, & \frac{1}{2}, & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{pmatrix}$$

So one possible matrix \mathbf{C} is $\begin{pmatrix} \frac{1}{2}, & -\frac{1}{2}, & \frac{1}{2}, & -\frac{1}{2} \end{pmatrix}$.

- (d) Provide a matrix \mathbf{A} so that the rank of \mathbf{B} is 4. Many answers are possible. One choice that works is

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

- (e) Provide a model matrix for a reduced model that corresponds to the null hypothesis of no temperature main effect.

$$\begin{aligned} \mathbf{W} = \mathbf{X}\mathbf{B}^{-1} &= \begin{bmatrix} \mathbf{1}_{2 \times 1} & & & \\ & \mathbf{1}_{3 \times 1} & & \\ & & \mathbf{1}_{4 \times 1} & \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{1}_{2 \times 1} & & & \\ & \mathbf{1}_{3 \times 1} & & \\ & & \mathbf{1}_{4 \times 1} & \\ 1 & -1 & 1 & -2 \end{bmatrix} \end{aligned}$$

By part (b), the model matrix of the reduced model is the first $p - q = 3$ columns of \mathbf{W} , i.e.,

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{1}_{2 \times 1} & & \\ & \mathbf{1}_{3 \times 1} & \\ & & \mathbf{1}_{4 \times 1} \\ 1 & -1 & 1 \end{bmatrix}.$$

- (f) Find the error sum of squares for the reduced and full models.
 $SS_{Reduced} = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{\mathbf{W}_1})\mathbf{y} = 185.28$
 $SS_{Full} = \mathbf{y}'(\mathbf{I} - \mathbf{P}_{\mathbf{W}})\mathbf{y} = 12$
- (g) Find the degrees of freedom associated with the sums of squares in part (f).
 $df_{Reduced} = \text{rank}(\mathbf{I} - \mathbf{P}_{\mathbf{W}_1}) = 10 - 3 = 7$
 $df_{Full} = \text{rank}(\mathbf{I} - \mathbf{P}_{\mathbf{W}}) = 10 - 4 = 6$
- (h) Compute the F -statistic for testing the null hypothesis of no time main effect using the sums of squares and degrees of freedom computed in parts (f) and (g).
 From the code and output below,

$$F = \frac{(SSE_{Reduced} - SSE_{Full}) / (df_{Reduced} - df_{Full})}{SSE_{Full} / df_{Full}} = 86.64$$

```
> X=model.matrix(~ 0 + factor(c(rep(1, 2), rep(2, 3), rep(3, 4), 4)))
> B=rbind(cbind(diag(3), 0), .5 * c(1, -1, 1, -1))
> W=X %%% solve(B)
> W1=W[, -4]
> proj=function(x) {
+   x %%% MASS::ginv(t(x) %%% x) %%% t(x)
+ }
> Pw=proj(W)
> Pw1=proj(W1)
> y=c(3, 5, 11, 13, 15, 5, 6, 6, 7, 16)
> I10=diag(rep(1, length(y)))
> SS.red=t(y) %%% (I10 - Pw1) %%% y
> SS.full=t(y) %%% (I10 - Pw) %%% y
> (SS.red - SS.full) / SS.full * 6
```

3. The easiest way to complete parts (a) through (d) is to write a few lines of SAS code as on slide 64 of slide set 8. For example,

```
proc glm;
  class source percent;
  model lconc = source percent source*percent /ss1 ss2 ss3;
  lsmeans source percent;
run;
```

Below you can see some other ways to do the computations.

```
<define functions needed for obtaining ANOVA table>
> ##projection matrix
> proj=function(X){X%*%MASS::ginv(t(X)%*%X)%*%t(X)}

> ##com: give one complete row in ANOVA table
> com=function(y,A,B){
+   ss=t(y)%*%(proj(A)-proj(B))%*%y
+   df=floor(Matrix::rankMatrix(A)-Matrix::rankMatrix(B))
+   ms=ss/df
+   f=ms/0.01351574
+   p=pf(f,df,17,lower.tail=FALSE)
+   return(round(c(ss,df,ms,f,p),4))
+ }

<data set>
> d = data.frame(pigs, lconc = log(pigs$conc), perc = factor(pigs$percent))

<generate model matrix>
> x1=matrix(rep(1,nrow(d)))
> xa=model.matrix(~0+d$source)
> xb=model.matrix(~0+d$perc)
> xab=model.matrix(~0+d$perc:d$source)
```

(a) ANOVA Table with Type 1:

```
> src1=com(d$lconc,cbind(x1,xa),x1)
> prct1=com(d$lconc,cbind(x1,xa,xb),cbind(x1,xa))
> srcprct1=com(d$lconc,cbind(x1,xa,xb,xab),cbind(x1,xa,xb))
> anova1=rbind(src1,prct1,srcprct1,error,total)
> rownames(anova1)=c("source|1","percent|1,source",
+ "sourceXpercent|1,source,percent","error","corrected total")
> colnames(anova1)=c("SS","df","MS","F","p-value")
```

```
> anova1
```

	SS	df	MS	F	p-value
source 1	0.63010	2	0.31510	23.3113	< 0.0001
percent 1,source	0.31740	3	0.10580	7.8269	0.0017
source×percent 1,source,percent	0.07510	6	0.01250	0.9259	0.5011
error	0.22977	17	0.01352		
corrected total	1.25237	28			

<Another way>

```
> o=lm(lconc~source+perc+source*perc, data=d)
> anova(o)
```

Comment: The arguments A and B in function “com” are matrices such that $\mathcal{C}(B) \subset \mathcal{C}(A)$ and calculate the sum of squares with $\mathbf{y}'(\mathbf{P}_A - \mathbf{P}_B)\mathbf{y}$.

(b) ANOVA Table with Type 2:

```
> src2=com(d$lconc,cbind(x1,xb,xa),cbind(x1,xb))
> prct2=com(d$lconc,cbind(x1,xa,xb),cbind(x1,xa))
> srcprct2=com(d$lconc,cbind(x1,xa,xb,xab),cbind(x1,xa,xb))
> anova2=rbind(src2,prct2,srcprct2,error,total)
> rownames(anova2)=c("source|1,percent","percent|1,source",
> "source×percent|1,source,percent","error","corrected total")
> colnames(anova2)=c("SS","df","MS","F","p-value")
> anova2
```

	SS	df	MS	F	p-value
source 1,percent	0.76480	2	0.38240	28.2914	< 0.0001
percent 1,source	0.31740	3	0.10580	7.8269	0.0017
source×percent 1,source,percent	0.07510	6	0.01250	0.9259	0.5011
error	0.22977	17	0.01352		
corrected total	1.38707	28			

<Another way>

```
car::Anova(o, type="II")
```

(c) You can use simple R code below that makes use of the joint_tests function in the emmeans package to get the Type III sums of squares ANOVA table.

```
> d = data.frame(pigs, lconc = log(pigs$conc), perc = factor(pigs$percent))
> o = lm(lconc ~ source + perc + source:perc, data =d)
> joint_tests(emmeans(o, c("source", "perc")), test = "F")
model term   df1 df2 F.ratio p.value
source          2   17  30.256 <.0001
perc            3   17   8.214  0.0013
source:perc     6   17   0.926  0.5011
```

Comments: ANOVA Table with Type 3 obtained by using the function “Anova” in car packages does not match with the table defined in slide 60 of set 8. Thus, we can get the sums of squares, $SS(\text{source}|1, \text{percent}, \text{source:percent})$ and $SS(\text{percent}|1, \text{source}, \text{source:percent})$, by applying the approach in the slides 70~74 of set 8.

As following the notation defined in (d), we can remove the main effect of source from the cell-means model.(see slide 72 of set 8)

$$\bar{\mu}_{1.} = \bar{\mu}_{2.} = \bar{\mu}_{3.} \iff \begin{cases} \mu_{24} = \mu_{11} + \mu_{12} + \mu_{13} + \mu_{14} - \mu_{21} - \mu_{22} - \mu_{23} \\ \mu_{34} = \mu_{11} + \mu_{12} + \mu_{13} + \mu_{14} - \mu_{31} - \mu_{32} - \mu_{33} \end{cases}$$

From the above, we can get the reduced matrix by removing two columns corresponding to μ_{24} and μ_{34} in cell-means model matrix and replacing the rows corresponding to μ_{24} and μ_{34} with $(1, 1, 1, 1, -1, -1, -1, 0, 0, 0)$ and $(1, 1, 1, 1, 0, 0, 0, -1, -1, -1)$, respectively. The function “src.red” below R code generates the reduced matrix of source.

Similarly, we can remove the main effect of percent.

$$\bar{\mu}_{.1} = \bar{\mu}_{.2} = \bar{\mu}_{.3} = \bar{\mu}_{.4} \iff \begin{cases} \mu_{32} = \mu_{11} - \mu_{12} + \mu_{21} - \mu_{22} + \mu_{31} \\ \mu_{33} = \mu_{11} - \mu_{13} + \mu_{21} - \mu_{23} + \mu_{31} \\ \mu_{34} = \mu_{11} - \mu_{14} + \mu_{21} - \mu_{24} + \mu_{31} \end{cases}$$

From the above, we can get the reduced matrix by removing three columns corresponding to μ_{32} , μ_{33} and μ_{34} in cell-means model matrix and replacing the rows corresponding to μ_{32} , μ_{33} and μ_{34} with $(1, -1, 0, 0, 1, -1, 0, 0, 1)$, $(1, 0, -1, 0, 1, 0, -1, 0, 1)$ and $(1, 0, 0, -1, 1, 0, 0, -1, 1)$, respectively. The function “prct.red” below generate the reduced matrix of percent.

```
<A Model Matrix for Model with 1, percent, source*percent>
> src.red=function(x,dat){
+   new.ab=x[,c(-8,-12)]
+   for(i in 1:nrow(x)){
+     if(dat[i,1]=="soy"&dat[i,2]=="18")
+       {new.ab[i,]=c(rep(1,4),rep(-1,3),rep(0,3))}else
+     if(dat[i,1]=="skim"&dat[i,2]=="18")
+       {new.ab[i,]=c(rep(1,4),rep(0,3),rep(-1,3))}
+   }
+   return(new.ab)
+ }
> xs.red=src.red(xab,d)
> anova(lm(d$lconc~0+xs.red),lm(d$lconc~0+xab))
> src3=c(0.81788,2,0.81788/2,30.256,2.507e-06)

<A Model Matrix for Model with 1, source, source*percent>
> src3=c(0.81788,2,0.81788/2,30.256,2.507e-06)
> prct.red=function(x,dat){
+   new.ab=x[,c(-10,-11,-12)]
+   for(i in 1:nrow(x)){
+     if(dat[i,1]=="skim"&dat[i,2]=="12"){new.ab[i,]=c(1,-1,0,0,1,-1,0,0,1)}else
+     if(dat[i,1]=="skim"&dat[i,2]=="15"){new.ab[i,]=c(1,0,-1,0,1,0,-1,0,1)}else
+     if(dat[i,1]=="skim"&dat[i,2]=="18"){new.ab[i,]=c(1,0,0,-1,1,0,0,-1,1)}
+   }
+   return(new.ab)
+ }
> xp.red=prct.red(xab,d)
> anova(lm(d$lconc~0+xp.red),lm(d$lconc~0+xab))
> prct3=c(0.33304,3,0.33304/3,8.2137,0.001348)
```

ANOVA Table with Type 3:

	SS	df	MS	F	p-value
source 1,percent,source×percent	0.81788	2	0.4089400	30.2560	< 0.0001
percent 1,source,source×percent	0.33304	3	0.1110133	8.2137	0.0013
source×percent 1,source,percent	0.07510	6	0.01250	0.9259	0.5011
error	0.22977	17	0.01352		
corrected total	1.45579	28			

- (d) Let μ_{ij} be a mean concentration of free plasma leucine for i source of protein and j , where $i = 1$ for fish meal, $i = 2$ for soybean meal and $i = 3$ dried skim milk and $j = 1, \dots, 4$ for 9%, 12%, 15% and 18%, respectively. From the code below, each μ_{ij} can be estimated like in table.

	$j=1$ (9%)	$j=2$ (12%)	$j=3$ (15%)	$j=4$ (18%)
$i=1$ (fish)	3.24526	3.43011	3.43461	3.47529
$i=2$ (soy)	3.53845	3.67962	3.66940	3.75887
$i=3$ (skim)	3.56054	3.76485	3.90463	4.09101

LSMeans for source

$$\begin{aligned} \text{fish} : \frac{\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{13} + \hat{\mu}_{14}}{4} &= \frac{3.24526 + 3.43011 + 3.43461 + 3.47529}{4} = 3.39632 \\ \text{soy} : \frac{\hat{\mu}_{21} + \hat{\mu}_{22} + \hat{\mu}_{23} + \hat{\mu}_{24}}{4} &= \frac{3.53845 + 3.67962 + 3.66940 + 3.75887}{4} = 3.66159 \\ \text{skim} : \frac{\hat{\mu}_{31} + \hat{\mu}_{32} + \hat{\mu}_{33} + \hat{\mu}_{34}}{4} &= \frac{3.56054 + 3.76485 + 3.90463 + 4.09101}{4} = 3.83026 \end{aligned}$$

LSMeans for percent

$$\begin{aligned} 9\% : \frac{\hat{\mu}_{11} + \hat{\mu}_{21} + \hat{\mu}_{31}}{3} &= \frac{3.24526 + 3.53845 + 3.56054}{3} = 3.44808 \\ 12\% : \frac{\hat{\mu}_{12} + \hat{\mu}_{22} + \hat{\mu}_{32}}{3} &= \frac{3.43011 + 3.67962 + 3.76485}{3} = 3.62486 \\ 15\% : \frac{\hat{\mu}_{13} + \hat{\mu}_{23} + \hat{\mu}_{33}}{3} &= \frac{3.43461 + 3.66940 + 3.90463}{3} = 3.66955 \\ 18\% : \frac{\hat{\mu}_{14} + \hat{\mu}_{24} + \hat{\mu}_{34}}{3} &= \frac{3.47529 + 3.75887 + 4.09101}{3} = 3.77506 \end{aligned}$$

- (e) Since the model that percent is treated like a quantitative variable is the reduced model of the cell-means model, we can conduct the lack of fit test of the reduced model compared to the cell-means model. From the code below and error in ANOVA table,

$$F = \frac{(SSE_{Reduced} - SSE_{Full}) / (df_{Reduced} - df_{Full})}{SSE_{Full} / df_{Full}} = \frac{(0.26291 - 0.22977) / (23 - 17)}{0.22977 / 17} = 0.4087.$$

Since the corresponding p-value is 0.8631, we can conclude that this model fit adequately relative to the cell-means model at significant level $\alpha=0.05$.

```
> o = lm(lconc ~ source + perc + source:perc, data =d)
> o1=lm(lconc~source+percent+source*percent,data=d)
> anova(o1,o)
```

- (f) The reduced model in (e) can be represented as $y_{ijk} = \mu + \alpha_i + \beta x_{ij} + \gamma_i x_{ij} + \epsilon_{ijk}$. Based on this model, the estimated linear relationship for each source is following.

i. fish

$$(\hat{\mu} + \hat{\alpha}_1) + (\hat{\beta} + \hat{\gamma}_1) * x_{1j} = 3.1164 + 0.0211x_{1j}$$

ii. soy

$$(\hat{\mu} + \hat{\alpha}_2) + (\hat{\beta} + \hat{\gamma}_2) * x_{2j} = (3.1164 + 0.2517) + (0.0211 + 0.0006)x_{2j} = 3.3681 + 0.0217x_{2j}$$

iii. skim

$$(\hat{\mu} + \hat{\alpha}_3) + (\hat{\beta} + \hat{\gamma}_3) * x_{3j} = (3.1164 - 0.0672) + (0.0211 + 0.0369)x_{3j} = 3.0492 + 0.058x_{3j}$$