hw_lect13-1

Consider The following LSD :

$$\begin{array}{cc} & \xrightarrow{k} \\ i \downarrow & \begin{pmatrix} 111 & 122 & 133 \\ 221 & 232 & 213 \\ 331 & 312 & 323 \end{pmatrix} \end{array}$$

j = middle

a) Use $\sum_j (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}) = 0$ to show That The cross-term $\sum'_{ijk} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots})$ is zero.

$\sum'_{ijk} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}) =$

$= \boxed{(\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 1 \cdot} - \bar{Y}_{\cdots}) + (\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 2 \cdot} - \bar{Y}_{\cdots}) + (\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdots})}$

$+ \quad 2 \quad 2 \quad 2 \quad 3 \quad 2 \quad 1$

$+ \quad 3 \quad 3 \quad 3 \quad 1 \quad 3 \quad 2$

$$(\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots}) \underbrace{\left[(\bar{Y}_{\cdot 1 \cdot} - \bar{Y}_{\cdots}) + (\bar{Y}_{\cdot 2 \cdot} - \bar{Y}_{\cdots}) + (\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdots}) \right]}_{\sum_j^3 (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}) = 0}$$

(*) FYI, proof, below

Similarly for The 2nd and 3rd rows. $= 0$.

b) Now, to illustrate The importance of LSD, consider The following non-LSD. It's hard to show generally That something is <u>not</u> zero; but provide some argument for why The above cross-term is not zero.

$$\begin{pmatrix} 111 & 132 & 133 \\ 221 & 222 & 213 \\ 331 & 312 & 323 \end{pmatrix}$$

$= \boxed{(\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 1 \cdot} - \bar{Y}_{\cdots})} + (\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdots}) + (\bar{Y}_{1\cdot\cdot} - \bar{Y}_{\cdots})(\bar{Y}_{\cdot 3 \cdot} - \bar{Y}_{\cdots})$

$+ \quad 2 \quad 2 \quad 2 \quad 2 \quad \boxed{2 \quad 1}$

$+ \quad 3 \quad 3 \quad \boxed{3 \quad 1} \quad 3 \quad 2$

blue terms $= (\bar{Y}_{\cdot 1 \cdot} - \bar{Y}_{\cdots}) \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}) = 0$, but The rest don't group into anything That involves $\sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}) = 0$ or $\sum_j (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}) =$ or etc.

$$\sum_j (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}) = \sum_j \bar{Y}_{\cdot j \cdot} - \sum_j \bar{Y}_{\cdots} = \frac{1}{3} \sum_j Y_{\cdot j \cdot} - \frac{1}{3^2} 3\, Y_{\cdots} = \frac{1}{3} Y_{\cdots} - \frac{1}{3} Y_{\cdots} = 0$$

hw-lect 13-2

Consider the model $y_{ij} = \mu + \tau_i + \beta_j + (\gamma)_{ij} + \epsilon_{ij}$.

As I explained in class, y_{ij} simply does not have enough df to estimate everything. But let's try anyway! Recall that we begin by taking derivatives of $SSE = \sum_{ij} (y_{ij} - \mu - \tau_i - \beta_j - \gamma_{ij})^2$, with respect to μ, τ_i, β_j, γ_{ij}, and setting the results to zero at $\hat{\mu}$, $\hat{\tau}_i$, $\hat{\beta}_j$, $\hat{\gamma}_{ij}$. Show that the deriv. w.r.t. γ_{ij} leads to an equation which implies $\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij} = 0$.

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + \epsilon_{ij}$$

$$SSE = \sum_{ij} (y_{ij} - \mu - \tau_i - \beta_j - \gamma_{ij})^2$$

$$\left.\frac{\partial}{\partial \gamma_{ij}}\right|_{\wedge} = \underbrace{y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \hat{\gamma}_{ij}}_{\hat{\epsilon}_{ij} \text{ (or } e_{ij})} = 0 \Rightarrow \underline{e_{ij} = 0} \Rightarrow MSE = 0$$

$\Downarrow$

Can't estimate σ_ϵ^2 with MSE

Just compare the above with what happens if there is a $\underline{k}$ index:

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

$$SSE = \sum_{ijk} (y_{ijk} - \mu - \tau_i - \beta_j - \gamma_{ij})^2$$

$$\left.\frac{\partial}{\partial \gamma_{ij}}\right|_{\wedge} = \sum_k (\underbrace{y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \hat{\gamma}_{ij}}_{e_{ijk}}) \Rightarrow \underline{\overline{e_{ij.}} = 0}$$

This time we $\underline{\underline{\text{can}}}$ estimate σ_ϵ^2 because $MSE \neq 0$.

hw_lect13-3

For the full (interactive) model $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$, show that

a) The sample mean (over i,j,k) of the predictions is equal to the sample mean (" ") of the observations.

b) The sample mean (" ") of the residuals is zero.

c) The sample covariance between predictions and residuals is zero.

defined as $\frac{1}{nab} \sum_{ijk} (\hat{y}_{ijk} - \bar{\hat{y}}_{\cdots})(e_{ijk} - \bar{e}_{\cdots})$

d),e) Same as a,b (Not c) but for reduced (additive) model.

Note/Hint: All of the above ask about sample stats. They are not asking about E[], V[] or cov[,]. So, just use the parameter estimates we have derived in the lecture, and simplify.

a) $\boxed{\bar{\hat{y}}_{\cdots}} = \frac{1}{nab} \sum_{ijk} \hat{y}_{ijk} = \frac{1}{nab} \sum_{ijk} \bar{y}_{ij\cdot} = \frac{1}{nab} \underbrace{\sum_{ij} \bar{y}_{ij\cdot}}_{ab\,\bar{y}_{\cdots}} \underbrace{\sum_k 1}_{n} = \boxed{\bar{y}_{\cdots}}$

b) $\boxed{\bar{e}_{\cdots}} = \frac{1}{nab} \sum_{ijk} e_{ijk} = \frac{1}{nab} \sum_{ijk} (y_{ijk} - \hat{y}_{ijk}) = \frac{1}{nab} \sum_{ijk} (y_{ijk} - \bar{y}_{ij\cdot})$

$= \frac{1}{nab} \left(\underbrace{\sum_{ijk} y_{ijk}}_{nab\,\bar{y}_{\cdots}} - \underbrace{\sum_{ij} \bar{y}_{ij\cdot}}_{ab\,\bar{y}_{\cdots}} \underbrace{\sum_k 1}_{n} \right) = \boxed{0}$

c) $= \frac{1}{nab} \sum_{ijk} (\hat{y}_{ijk} - \bar{\hat{y}}_{\cdots})(e_{ijk} - \bar{e}_{\cdots})$

$\bar{\hat{y}}_{\cdots} \to \bar{y}_{\cdots}$, $\bar{e}_{\cdots} \to 0$

$= \frac{1}{nab} \left[\sum_{ijk} \hat{y}_{ijk} e_{ijk} - \bar{y}_{\cdots} \underbrace{\sum_{ijk} e_{ijk}}_{nab\,\bar{e}_{\cdots} = 0} \right] = \frac{1}{nab} \sum_{ijk} \hat{y}_{ijk} e_{ijk}$

$$= \frac{1}{nab}\sum_{ijk}\hat{y}_{ijk}(y_{ijk}-\hat{y}_{ijk}) = \frac{1}{nab}\sum_{ijk}\bar{y}_{ij.}(y_{ijk}-\bar{y}_{ij.})$$

$$= \frac{1}{nab}\sum_{ij}\bar{y}_{ij.}\overbrace{\sum_{k}(y_{ijk}-\bar{y}_{ij.})}^{=0} = 0$$

d) $$\boxed{\bar{\hat{y}}_{...}} = \frac{1}{nab}\sum_{ijk}\hat{y}_{ijk} = \frac{1}{nab}\sum_{ijk}(\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...}) =$$

$$= \frac{1}{nab}\Big[nb\underbrace{\sum_{i}\bar{y}_{i..}}_{a\bar{y}_{...}} + na\underbrace{\cancel{\sum_{j}\bar{y}_{.j.}}}_{b\bar{y}_{...}} - \cancel{nab\bar{y}_{...}}\Big] = \boxed{\bar{y}_{...}}$$

e) $$\boxed{\bar{e}_{...}} = \frac{1}{nab}\sum_{ijk}e_{ijk} = \frac{1}{nab}\sum_{ijk}(y_{ijk}-\hat{y}_{ijk}) = \frac{1}{nab}\sum_{ijk}(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})$$

$$= \frac{1}{nab}(\cancel{nab\bar{y}_{...}} - \cancel{nab\bar{y}_{...}} - \cancel{nab\bar{y}_{...}} + \cancel{nab\bar{y}_{...}}) = \boxed{0}$$

Think of the residual plot

Moral: For the full model the residuals and the predictions are uncorrelated. But not so for the reduced (additive) model.

$$= \frac{1}{nab}\sum_{ijk}\hat{y}_{ijk}e_{ijk} = \frac{1}{nab}\sum_{ijk}\hat{y}_{ijk}(y_{ijk}-\hat{y}_{ijk})$$

$$= \frac{1}{nab}\Big[\sum_{ijk}(\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...})y_{ijk} - \sum_{ijk}(\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...})^2\Big]$$

$$= \frac{1}{nab}\Big[\cancel{\sum_{i}\bar{y}_{i..}\overbrace{\sum_{jk}y_{ijk}}^{bn\bar{y}_{i..}}} + \sum_{j}\cancel{\bar{y}_{.j.}\overbrace{\sum_{ik}y_{ijk}}^{an\bar{y}_{.j.}}} - \cancel{\bar{y}_{...}\overbrace{\sum_{ijk}y_{ijk}}^{abn\bar{y}_{...}}}$$

$$- \cancel{bn\sum_{i}(\bar{y}_{i..})^2} - \cancel{an\sum_{j}(\bar{y}_{.j.})^2} - \cancel{abn(\bar{y}_{...})^2}$$

$$- 2n\sum_{ij}\bar{y}_{i..}\bar{y}_{.j.} + 2bn\bar{y}_{...}\underbrace{\sum_{i}\bar{y}_{i..}}_{a\bar{y}_{...}} + \cancel{2an\bar{y}_{...}\underbrace{\sum_{j}\bar{y}_{.j.}}_{b\bar{y}_{...}}}\Big]$$

$$= 2\Big[(\bar{y}_{...})^2 - \frac{1}{ab}\sum_{ij}\bar{y}_{i..}\bar{y}_{.j.}\Big] \boxed{\neq 0}$$

hw-lect 14-1

Consider The model $Y_{ijk} = \mu + \alpha_i + (\alpha\beta)_{ij} + \epsilon_{ijk}$. $i = 1 \dots a$, $j = 1 \dots b$, $k = 1 \dots n$

a) Write The Least-squares equations (from $\partial/\partial\mu$, ...)

All of These should have hat on Them.

$$SSE = \sum_{ijk} \left(Y_{ijk} - \mu - \alpha_i - (\alpha\beta)_{ij}\right)^2$$

$$\frac{\partial}{\partial\mu} : \sum_{ijk}\left(Y_{ijk} - \mu - \alpha_i - (\alpha\beta)_{ij}\right) = 0 \rightarrow \overline{Y}_{\cdot\cdot\cdot} - \mu - \overline{\alpha}_{\cdot} - \overline{(\alpha\beta)}_{\cdot\cdot} = 0 \quad \text{I}$$

$$\frac{\partial}{\partial\alpha_i} : \sum_{jk}\left(Y_{ijk} - \mu - \alpha_i - (\alpha\beta)_{ij}\right) = 0 \rightarrow \overline{Y}_{i\cdot\cdot} - \mu - \alpha_i - \overline{(\alpha\beta)}_{i\cdot} = 0 \quad \text{II}$$

$$\frac{\partial}{\partial(\alpha\beta)_{ij}} : \sum_{k}\left(Y_{ijk} - \mu - \alpha_i - (\alpha\beta)_{ij}\right) = 0 \rightarrow \overline{Y}_{ij\cdot} - \mu - \alpha_i - (\alpha\beta)_{ij} = 0. \quad \text{III}$$

b) How many indep. equations are There in part a)?

I = 1 equ. $\qquad \sum_i \text{II} = \text{I}$.

II = $a - 1$ equs $\qquad \sum_j \text{III} = \text{II}$

III = $ab - (a-1) - 1$ $\qquad \sum_{ij} \text{III} = \text{I}$

$\therefore \quad 1 + a - 1 + ab - (a-1) - 1 = \underline{ab \text{ indep. equations.}}$

c) Given That The model has $1 + a + ab$ params, will The following Constraints be sufficient? $\alpha_{\cdot} = 0$ $\quad \alpha\beta_{i\cdot} = 0$ Explain.

Yes. $1 + a$ additional equs, is how many we need.

d) Write The df for each of The SS terms in The ANOVA decomp.

SST =	SSA +	SSAB +	SSE
$abn-1$	$a-1$	$ab-a$	$abn-ab$

hw- lect14-2

1

```
# For the data in problem 5.28 in book
# a) Produce the ANOVA Table for the full model (By hand).
# b) Make a qqplot of the residuals, and the residuals plots (vs. predicted,
# versus factor1, versus factor2, and versus factor3). By R.
# c) The plots in part b call for data transformation (of y). Try log(y) and sqrt(y)
# to see if the results are improved, but do not turn in these results. Instead
# let y --> y^{1.7}, and then repeat parts a and b.
# Hint in all of the above, use the formulas on page 185-186 in book.
# You may also need to make a 3d-array in R; if so, you can do it this way:
# t = numeric(a); tt = matrix(nrow=b,ncol=c) ; 3d-array = outer(t,tt)

  library(AlgDesign)
  a = 2          # levels of Booster
  b = 2          # levels of Washings
  c = 2          # levels of Formulation
  n = 2          # replications (Here, no need to block)
  design = gen.factorial(c(a,b,c,n), varNames=c("A","B","C","R"), factors="all")
  attach(design)

  y = c(6,6,3,4,  10,11,10,9,     # 1st replicate,
      5,5,2,1,  9,11,9,10 )     # second replicate
 cbind(A,B,C,R,y)                # Check.

 # Compute the conditional means:

  y.... = sum(y)               # formulas on p 185 use sums (not means)
  yi... = c( sum(y[A==1]), sum(y[A==2]) )
  y.j.. = c( sum(y[B==1]), sum(y[B==2]) )
  y..k. = c( sum(y[C==1]), sum(y[C==2]) )

    yij.. = matrix(nrow=a,ncol=b)
    for(i in 1:a){
    for(j in 1:b){
    yij..[i,j] = sum(y[A==i & B==j])
    }}
    yi.k. = matrix(nrow=a,ncol=c)
    for(i in 1:a){
    for(j in 1:c){
    yi.k.[i,j] = sum(y[A==i & C==j])
    }}
    y.jk. = matrix(nrow=b,ncol=c)
    for(i in 1:b){
    for(j in 1:c){
    y.jk.[i,j] = sum(y[B==i & C==j])
    }}

      t = numeric(a); tt = matrix(nrow=b,ncol=c) ; yijk. = outer(t,tt)
      for(i in 1:a){
      for(j in 1:b){
      for(k in 1:c){
      yijk.[i,j,k] = sum(y[A==i & B==j & C==k])
      }}}
```

```
# Get ready for ANOVA table:

dfA = (a-1)
dfB = (b-1)
dfC = (c-1)
dfAB = (a-1)*(b-1)
dfAC = (a-1)*(c-1)
dfBC = (b-1)*(c-1)
dfABC = (a-1)*(b-1)*(c-1)
dfE = a*b*c*(n-1)

SST = sum( y^2 ) - y....^2/(a*b*c*n)
SSA = sum(yi...^2)/(b*c*n) - y....^2/(a*b*c*n)
SSB = sum(y.j..^2)/(a*c*n) - y....^2/(a*b*c*n)
SSC = sum(y..k.^2)/(a*b*n) - y....^2/(a*b*c*n)

 temp = 0
 for(i in 1:a){
 for(j in 1:b){
 temp = temp + yij..[i,j]^2
 }}
 SSAB = temp/(c*n) - y....^2/(a*b*c*n) - SSA - SSB

 temp = 0
 for(i in 1:a){
 for(j in 1:c){
 temp = temp + yi.k.[i,j]^2
 }}
 SSAC = temp/(b*n) - y....^2/(a*b*c*n) - SSA - SSC

 temp = 0
 for(i in 1:b){
 for(j in 1:c){
 temp = temp + y.jk.[i,j]^2
 }}
 SSBC = temp/(a*n) - y....^2/(a*b*c*n) -  SSB - SSC

 temp = 0
 for(i in 1:a){
 for(j in 1:b){
 for(k in 1:c){
 temp = temp + yijk.[i,j,k]^2
 }}}
 SSABC = temp/n - y....^2/(a*b*c*n) - SSA - SSB - SSC - SSAB - SSAC - SSBC

 SSE = SST - SSA - SSB - SSC - SSAB - SSAC - SSBC - SSABC

MSA = SSA/dfA
MSB = SSB/dfB
MSC = SSC/dfC
MSAB = SSAB/dfAB
MSAC = SSAC/dfAC
MSBC = SSBC/dfBC
MSABC = SSABC/dfABC
MSE = SSE/dfE
```

```
FA = MSA/MSE
  FB = MSB/MSE
  FC = MSC/MSE
  FAB = MSAB/MSE
  FAC = MSAC/MSE
  FBC = MSBC/MSE
  FABC = MSABC/MSE

  pvA = pf(FA,dfA,dfE,lower.tail=F)
  pvB = pf(FB,dfB,dfE,lower.tail=F)
  pvC = pf(FC,dfC,dfE,lower.tail=F)
  pvAB = pf(FAB,dfAB,dfE,lower.tail=F)
  pvAC = pf(FAC,dfAC,dfE,lower.tail=F)
  pvBC = pf(FBC,dfBC,dfE,lower.tail=F)
  pvABC = pf(FABC,dfABC,dfE,lower.tail=F)

  table = rbind(
     cbind(dfA,dfB,dfC,dfAB,dfAC,dfBC,dfABC,dfE),
     cbind(SSA,SSB,SSC,SSAB,SSAC,SSBC,SSABC,SSE),
     cbind(MSA,MSB,MSC,MSAB,MSAC,MSBC,MSABC,MSE),
     cbind(FA,FB,FC,FAB,FAC,FBC,FABC,-99),        # -99 as a place holder.
     cbind(pvA,pvB,pvC,pvAB,pvAC,pvBC,pvABC,-99)
     )

  t(table)

# A     1   0.5625   0.5625   0.6000  4.608560e-01
# B     1  14.0625  14.0625  15.0000  4.721383e-03
# C     1 138.0625 138.0625 147.2667  1.968149e-06
# AB    1   0.5625   0.5625   0.6000  4.608560e-01
# AC    1   0.5625   0.5625   0.6000  4.608560e-01
# BC    1   5.0625   5.0625   5.4000  4.863085e-02
# ABC   1   0.5625   0.5625   0.6000  4.608560e-01
# E     8   7.5000   0.9375 -99.0000 -9.900000e+01

# All of the above agrees with R says:

  lm.1 = lm( y ~ A*B*C)
  summary.aov(lm.1)
#           Df Sum Sq Mean Sq F value   Pr(>F)
# A          1   0.56    0.56     0.6  0.46086
# B          1  14.06   14.06    15.0  0.00472 **
# C          1 138.06  138.06   147.3 1.97e-06 ***
# A:B        1   0.56    0.56     0.6  0.46086
# A:C        1   0.56    0.56     0.6  0.46086
# B:C        1   5.06    5.06     5.4  0.04863 *
# A:B:C      1   0.56    0.56     0.6  0.46086
# Residuals  8   7.50    0.94
```

```
# b)

 lm.1 = lm( y ~ A*B*C)                    # full model
 summary.aov(lm.1)                        # View with suspicision, because of
 yhat = predict(lm.1)                     # the following plots don't look good.
 par(mfrow=c(3,2),mar=c(2,2,2,2))
 qqnorm(lm.1$resid)
 plot(yhat, lm.1$resid) ; abline(h=0)
 plot(as.numeric(A),lm.1$resid); abline(h=0)   # w/o as.numeric() you'll get boxplots.
 plot(as.numeric(B),lm.1$resid); abline(h=0)
 plot(as.numeric(C),lm.1$resid); abline(h=0)

# All of these look bad. The qqnorm plot is not straight, and the residual plots
# all show non-constant variances. The recommended thing to do is transform data.
# Some standard transformations and sqrt and log, certainly when y is positive:

 lm.1 = lm( sqrt(y) ~ A*B*C)              # sqrt transform
#   lm.1 = lm( log(y) ~ A*B*C)
 summary.aov(lm.1)
 yhat = predict(lm.1)
 par(mfrow=c(3,2),mar=c(2,2,2,2))
 qqnorm(lm.1$resid)
 plot(yhat, lm.1$resid) ; abline(h=0)
 plot(as.numeric(A),lm.1$resid); abline(h=0)
 plot(as.numeric(B),lm.1$resid); abline(h=0)
 plot(as.numeric(C),lm.1$resid); abline(h=0)

# Both the sqrt and the log transform improve the qqplot a bit, but the
# residuals are still bad. Now you can try a weird transform:

 lm.1 = lm( (y)^1.7 ~ A*B*C)
 summary.aov(lm.1)
 yhat = predict(lm.1)
 par(mfrow=c(3,2),mar=c(2,2,2,2))
 qqnorm(lm.1$resid)
 plot(yhat, lm.1$resid) ; abline(h=0)
 plot(as.numeric(A),lm.1$resid); abline(h=0)
 plot(as.numeric(B),lm.1$resid); abline(h=0)
 plot(as.numeric(C),lm.1$resid); abline(h=0)

# This one improves the residual plots, but now the qqplot looks bad.
# Note that the F tests do differ in terms of which factors are significant.
# For example, the non-transformed data and the sqrt and log transformed data
# all lead to BC being a significant interaction. But the 1.7 transformed data
# leads to a model in which BC is not significant.
# It's possible that there exists yet another transformation that improves
# both the qqplot and the residual plots, but for now this is good enough.
```

hw-lect14-3

```
# 5.23
  rm(list=ls(all=TRUE))
  library(AlgDesign)

  y.m = matrix( c(
      109, 110, 108, 110,
      110, 115, 109, 108,
      110, 110, 111, 114,
      112, 111, 109, 112,
      116, 112, 114, 120,
      114, 115, 119, 117 ), ncol=4, byrow=T)
  y = as.vector(t(y.m))
```

By comparing the order of these y-values with the data given in the problem, you can see that the machine factor (A) is changing fastest, then replication (R), and then the slowest changing factor is operator (B). So, we need

```
  design = gen.factorial(c(4,2,3),varNames=c("A","R","B"), factors="all")
  attach(design)

  cbind(A,R,B,y)          # compare and confirm with data in the problem.

  # If you don't include the Replication factor in the model, then you will be
  # solving problem 5.8, i.e., developing a full model involving A and B:

  lm.1 = lm(y ~ A*B)
  summary.aov(lm.1)

#              Df Sum Sq Mean Sq F value   Pr(>F)
# machine       3  12.46    4.15   1.095 0.388753
# oper          2 160.33   80.17  21.143 0.000117 ***
# machine:oper  6  44.67    7.44   1.963 0.150681
# Residuals    12  45.50    3.79
```

But, here we want the replication to be a block factor. As far as R is concerned, you simply enter the R factor as a factor in the model. R has no way of knowing whether a factor is a treatment factor or a block factor - that difference arises in the way the experiment is performed, AND in the interpretation of the p-values in the anove table:

```
  lm.2 = lm(y ~ A*B + R)
  summary.aov(lm.2)

          Df Sum Sq Mean Sq F value   Pr(>F)
A          3  12.46    4.15   1.051 0.408659
B          2 160.33   80.17  20.291 0.000204 ***
R          1   2.04    2.04   0.517 0.487209
A:B        6  44.67    7.44   1.884 0.171630
Residuals 11  43.46    3.95
```

We can see that the operator factor (B) is still significant, and the machine factor (A) is not, just as in the unblocked/full model. This is consistent with the large p-value associated with the block factor, suggesting that there is no evidence for variability across blocks anyway (Although we remind ourselves that we should be cautious in interpretating this p-value).

hw-14-4

1

```
# Consider the data from example 9.1. The design is a factorial involving 3 factors
# A, B, and C, each with 3 levels. Consider only the first replicate.
# a) Fit a model that includes at most 2-way interactions, and produce the
# anova table. State your conclusions; which effects are significant, and
# which are not?
# b) Now consider only the portion of the data that would follow if we had
# followed a Latin Square Design, of the form
#      A
#    1,2,3
# B (2,3,1)                          # Note B = row factor.
#    3,1,2
# Develop an additive model, and and produce the anova table. State your
# conclusions; which effects are significant, and which are not?
# Are the conclusions in parts a and b consistent?
# c) For the LSD consider a non-additive model that includes A + B + C but
# also a single 2-way interaction term BC. Develop the model and produce
# the anova table. What comments can you make about this table?

  rm(list=ls(all=TRUE))
  library(AlgDesign)                  # for gen.factorial()

# a)
  nf = 3                             # Number of factors.
  p=3                                # Number of levels.
  design = gen.factorial(rep(p,nf), varNames=c("A","B","C"), factors="all")
  attach(design)
  y = c(-35,17,-39, -45,-65,-55, -40,20,15, # 1st replicate in Table 9.1
       110,55,90, -10,-55,-28, 80,110,110,
       4,-23,-30, -40,-64,-61, 31,-20,54)
  cbind(A,B,C,y)                          # Confirm with Table 9.1
  lm.1 = lm(y ~ A*B + A*C + B*C )             # same as A+B+C+ A*B + A*C + B*C
  summary.aov(lm.1)
#           Df Sum Sq Mean Sq  F value Pr(>F)
# A          2 480    240      0.506 0.620709
# B          2 36474  18237   38.481 7.86e-05 ***
# C          2 31634  15817   33.375 0.000131 ***
# A:B        4 3399   850      1.793 0.223418
# A:C        4 3729   932      1.967 0.192726
# B:C        4 7626   1906     4.023 0.044649 *
# Residuals  8 3791   474

# The (main) effects of B and C are statistically significant, but there is
# no evidence from this data that factor A has an effect on y. The interaction
# between B and C is also significant.

# This is not necessary for hw, but as a rule check the residual and qqplot:
  plot(lm.1$residuals, predict(lm.1)) ; abline(h=0) # Looks OK/random.
  qqnorm(lm.1$residuals)                        # Looks straight/normal.
```

```
# b) This time, only A and B have a factorial design, and C follows the pattern in
# the given LSD with B (A) as row (col) factors. Note: if the LSD is read-in row-wise,
# then, A changes faster than B:
  design = gen.factorial(rep(p,2), varNames=c("A","B"), factors="all")
  attach(design)
  C = as.factor(c(1,2,3,
                  2,3,1,
                  3,1,2))
  y = c(-35,  55, -30,        # Keep only the corresponding y's.
      -10, -64, -55,
       31,  20, 110)
  cbind(A,B,C,y)             # Confirm.
  lm.2 = lm(y ~ A + B + C)
  summary.aov(lm.2)
#          Df Sum Sq Mean Sq F value Pr(>F)
# A         2 260    130  0.591  0.6285
# B         2 14167  7083 32.181 0.0301 *
# C         2 10911  5455 24.785 0.0388 *
# Residuals 2 440    220
# The significant factors are B and C. This is the same conclusion that followed
# from fitting the model with 2-way interactions. This is not guaranteed, but in
# this case, the two models agree.
# If you're wondering why we didn't include interactions in this model, see next part.
# Again, not necessary for hw, but check these plots:
  plot(lm.2$residuals, predict(lm.2)) ; abline(h=0) # OK
  qqnorm(lm.2$residuals)                        # OK

# c)
  lm.3 = lm(y ~ A + B + C + B:C)
  summary.aov(lm.3)
#           Df Sum Sq Mean Sq
# A          2 260   130
# B          2 14167 7083
# C          2 10911 5455
# B:C        2 440   220
# This time, there is no SSE (or MSE) term, and so no tests can be performed.
# The same conclusion follows if you include A:C or A:B. This is happening
# because the LSD data simply does not have enough df to estimate all the
# parameters/effects. Of course, if we had replication, then we would have more
# df in the data; we are not doing that here because we have not completely
# learned how to do LSD with replication; maybe later. However, LSD designs
# do generally assume that the interactions are small, because in an LSD
# we cannot discriminate between a main effect (from A, B, or C), and an
# interaction (from AB, BC, AC). The effects one computes in LSD are some
# unknown linear combination of the true main effects and interaction effects.
# The inability do discriminate between two things already shows up in the
# anova tables above; note that SS(BC) in the LSD model with an interaction
# is equal to SSE of the factorial design, i.e., 440.
```

hw-lect 15-1

For $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$, consider The main effect for A, for B, and The interaction effect, all written in Yates' notation. We showed That The product of The contrast constants for A and for B is equal to The contrast constant of The interaction effect. Show That The product of The main effect for A, and for B, does NOT give The interaction effect.

$$\left.\begin{aligned} A &= \frac{1}{2n}\left[-(1) + a - b + ab\right] \\ B &= \frac{1}{2n}\left[-(1) - a + b + ab\right] \end{aligned}\right\} \quad A \times B = \frac{1}{4n^2}\Big[\big(-(1) + ab\big) + (a-b)\Big] \times \Big[\big(-(1) + ab\big) - (a-b)\Big]$$

$$A \times B = \frac{1}{4n^2}\left[\big(-(1) + ab\big)^2 - (a-b)^2\right] \neq AB = \frac{1}{2n}\left[+(1) - a - b + ab\right]$$

No squares.

hw-lect 15-2

In a 2^1 design with the model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $i = 1 \cdots, a$; $j = 1 \cdots, n$,

a) Starting from either $SS_A = \sum_{i,j}^{an} (\bar{y}_{i.} - \bar{y}_{..})^2$ or $\frac{1}{n}\sum_i y_{i.}^2 - \frac{1}{an} y_{..}^2$ show that SS_A is proportional to the square of a contrast in (1) and $\underline{a}$. what $\underline{\underline{is}}$ that contrast?

b) Can SSE be written in terms of a $(\text{contrast})^2$? provide some math/explanation.

a) $y_{ij} = y_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$

$$SS_A = \sum_{i,j}^{a,n} (\bar{y}_{i.} - \bar{y}_{..})^2 = n \sum_{i=1}^{a} \left(\frac{1}{n} y_{i.} - \frac{1}{an} y_{..} \right)^2 = \frac{1}{n} \sum_i \left(y_{i.} - \frac{1}{a} y_{..} \right)^2$$

$$= \frac{1}{n} \left[\sum_i y_{i.}^2 - \frac{2}{a} y_{..} \underbrace{\sum_i y_{i.}}_{y_{..}} + \frac{1}{a^2} y_{..}^2 \underbrace{\sum_i 1}_{1} \right]$$

$$SS_A = \frac{1}{n}\sum_i^a y_{i.}^2 - \frac{1}{na} y_{..}^2$$ ← You may start from here. 3.9, p.74

$$= \frac{1}{n} \left[y_{1.}^2 + y_{2.}^2 - \frac{1}{2} (y_{1.} + y_{2.})^2 \right] \quad (a = 2)$$

$$= \frac{1}{2n} \left(y_{1.}^2 + y_{2.}^2 - 2 y_{1.} y_{2.} \right) = \frac{1}{2n} \left(y_{2.} - y_{1.} \right)^2$$

$$= \frac{1}{2n} \left[a - (1) \right]^2$$

$\xrightarrow{\quad (1) \quad a \quad} A$; $(1) = y_{1.}$, $a = y_{2.}$

$$= \frac{1}{2n} (\text{Contrast}_A)^2 \quad \text{where } \text{contrast}_A = -(1) + a.$$

b) $SS_E = \sum_{i}^{a} \sum_{j}^{n} (y_{ij} - \bar{y}_{i.})^2 = \sum_{i,j}^{a,n} (y_{ij} - \frac{1}{n} y_{i.})^2$

$$= \sum_{i,j} y_{ij}^2 - \frac{2}{n} \sum_{i} y_{i.} \underbrace{\sum_{j} y_{ij}}_{y_{i.}} + \frac{1}{n^2} \sum_{i} y_{i.}^2 \underbrace{\sum_{j}}_{n}$$

$$= \sum_{i,j} y_{ij}^2 - \frac{1}{n} \sum_{i} y_{i.}^2$$ $\longleftarrow$ You may start from here. top of p. 68

$$= \sum_{j}^{n} (y_{1j}^2 + y_{2j}^2) - \frac{1}{n} (y_{1.}^2 + y_{2.}^2)$$

The only contrasts are $(1) - a$ and $-(1) + a$.

= cannot be written as $(\text{contrast})^2$

$\therefore SS_E$ can not be written in terms of a $(\text{contrast})^2$.

hw - lect 15-3

We have seen that for 2^2, the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i=1,2 \quad j=1,2 \quad k=1,\dots,n$$

has a decomposition wherein SSA, defined as

$$SSA = \sum_{i,j,k}^{22n} (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

can be written as $\frac{1}{4n}(\text{contrast } A)^2$

It can be shown that SSAB, defined as

$$SSAB = \sum_{i,j,k}^{22n} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

can be written in terms of $\frac{1}{4n}(\text{Contrast } AB)^2$. Here all you have to show is that the $i=1, j=1$ term in SSAB is equal to $\frac{1}{16n}(\text{Contrast } AB)^2$, ie. show $\sum_k (\bar{Y}_{11.} - \bar{Y}_{1..} - \bar{Y}_{.1.} + \bar{Y}_{...})^2 = \frac{1}{16n}(\text{Contrast } AB)^2$.

$$SSAB = \sum_{i,j,k}^{22n} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 = n \sum_{i,j} \left(\frac{1}{n} Y_{ij.} - \frac{1}{2n} Y_{i..} - \frac{1}{2n} Y_{.j.} + \frac{1}{4n} Y_{...} \right)^2$$

$$= \frac{1}{n} \left(Y_{11.} - \frac{1}{2} Y_{1..} - \frac{1}{2} Y_{.1.} + \frac{1}{4} Y_{...} \right)^2 + \cdots$$

$$= \frac{1}{n} \left[\left(Y_{11.} - \frac{1}{2}(Y_{11.} + Y_{12.}) - \frac{1}{2}(Y_{11.} + Y_{21.}) + \frac{1}{4}(Y_{11.} + Y_{12.} + Y_{21.} + Y_{22.}) \right) \right]^2 + \cdots$$

$$= \frac{1}{n} \left[\frac{1}{4} Y_{11.} - \frac{1}{4} Y_{12.} - \frac{1}{4} Y_{21.} + \frac{1}{4} Y_{22.} \right]^2 + \cdots$$

$$= \frac{1}{n} \frac{1}{16} \left[(1) - b - a + ab \right]^2 + \cdots$$

$$= \frac{1}{16n} \left[(1) - a - b + ab \right]^2 + \cdots$$

$$= \frac{1}{16n} \left[\text{Contrast } AB \right]^2 + \cdots$$

	A	B	AB
(1)	−	−	+
a	+	−	−
b	−	+	−
ab	+	+	+

hw-lect 15-4

For the data in problem 6.5,
a) Perform ANOVA By R.

```
  rm(list=ls(all=TRUE))
   nr = 4                    # number replicates.
      y.m = matrix(
      c(18.2, 18.9, 12.9, 14.4,
        27.2, 24.0, 22.4, 22.5,
        15.9, 14.5, 15.1, 14.2,
        41.0, 43.9, 36.3, 39.9),  nrow = 4, ncol= 4, byrow=T )
       y = as.vector(y.m)
       A = as.factor(rep(c(-1,+1,-1,+1),4))
        B = as.factor(rep(c(-1,-1,+1,+1),4))
        lm.1 = lm(y~A+B+A*B)
        summary.aov(lm.1)
#            Df Sum Sq Mean Sq F value   Pr(>F)
# A           1 1107.2  1107.2  185.25 1.17e-08 ***
# B           1  227.3   227.3   38.02 4.83e-05 ***
# A:B         1  303.6   303.6   50.80 1.20e-05 ***
# Residuals   12   71.7     6.0
```

b) Recompute the 4 SS values in the anova table of part a, this time by hand, using the y-totals (1), a, b, ab, and the contrasts.

```
# Totals:
    one = sum(y[A==-1 & B==-1])               # (1) = 64.4
    a =   sum(y[A==+1 & B==-1])               # a   = 96.1
    b =   sum(y[A==-1 & B==+1])               # b   = 59.7
    ab =  sum(y[A==+1 & B==+1])               # ab  = 161.1
# Contrasts:
    cont1 = ( -one + a - b + ab )/nr
    cont2 = ( -one - a + b + ab )/nr
    cont3 = ( +one - a - b + ab )/nr
    cont1^2/(4/nr)                      # 1107.226  = SS in printout above.
    cont2^2/(4/nr)                      # 227.2556
    cont3^2/(4/nr)                      # 303.6306
    # SSE can be computed either by subtraction from SST:
    sum((y - mean(y))^2) - ( cont1^2/(4/nr) + cont2^2/(4/nr) + cont3^2/(4/nr))
    # or from its definition:
    sum((y-predict(lm.1))^2)                # 71.7225
```

c) Under what conditions (bit size and cutting speed) would you operate this process? To answer that question, make comparative boxplots of y for each of the treatement combinations.

```
   boxplot(y[A==-1 & B==-1],y[A==+1 & B==-1], ylim=c(10,50))
   boxplot(y[A==-1 & B==+1],y[A==+1 & B==+1],add=T,col=2,boxwex=0.5)
```

The x axis refers to the 2 levels of A. And the 2 colors correspond to different levels of B. Two reduce vibration, it is clearly better to use A=-1 (smaller bit). As for the factor B (speed), given the big overlap between the boxplots corresponding to B=-1 and B=+1, when A=-1, as long as we use A=-1, B does not have have a significant effect on vibration.

d) How does the interaction term in the anova table manifest itself in the diagram of part c.

Imagine the same-colored boxplots are connected by a straight line. The big difference in the slope of the two lines (black and red), is consistent with the significant interaction term in the anova printout.