

c) Based on the above "hint," deduce how all of the other contrasts can be computed. Again, print them so the grader can see them.

```
design = gen.factorial(c(rep(2,k),n),varNames=c("A","B","C","D","R"))
attach(design)
all.contrasts = c(sum(A*y), sum(B*y), sum(C*y), sum(D*y), sum(A*B*y), sum(A*C*y), sum(A*D*y), sum(B*C*y), sum
(B*D*y), sum(C*D*y), sum(A*B*C*y), sum(A*B*D*y), sum(A*C*D*y), sum(B*C*D*y), sum(A*B*C*D*y) )

# 48.302 63.614 -57.540 31.324 30.946 -64.124 1.224 1.536 0.756
# -1.230 50.200 1.568 0.306 0.570 0.226
```

d) From the above contrasts, compute all the effects, and perform a t-test on each. Use the MSE value saved in part b. Be careful and clear on what df you use for the tests. Are the effects, the t-ratios, and the p-values equal to the "by R" results in part b?

```
all.effects = all.contrasts/(n*2^(k-1))
all.t = (all.effects - 0)/sqrt( MSE/ (n*2^(k-2) ))
all.pv = 2*pt(abs(all.t),(n-1)*(2^k),lower.tail=F) # df = N - p = n*2^k - 2^k

cbind(all.effects, all.t, all.pv)

# all.effects all.t all.pv
# [1,] 3.018875 29.9723026 1.740225e-15
# [2,] 3.975875 39.4736876 2.247046e-17
# [3,] -3.596250 -35.7046559 1.100354e-16
# [4,] 1.957750 19.4371332 1.485334e-12
# [5,] 1.934125 19.2025770 1.790240e-12
# [6,] -4.007750 -39.7901522 1.979949e-17
# [7,] 0.076500 0.7595151 4.585895e-01
# [8,] 0.096000 0.9531170 3.547091e-01
# [9,] 0.047250 0.4691123 6.453180e-01
# [10,] -0.076875 -0.7632382 4.564284e-01
# [11,] 3.137500 31.1500474 9.485296e-16
# [12,] 0.098000 0.9729736 3.450474e-01
# [13,] 0.019125 0.1898788 8.517922e-01
# [14,] 0.035625 0.3536958 7.281845e-01
# [15,] 0.014125 0.1402373 8.902228e-01
```

low-lat 17-1

Same as in part b.

e) In this problem the existence of replication allows us to estimate σ_{ϵ} (with MSE), which in turn allows us to perform tests. Suppose, we didn't have MSE. Use Daniel's idea to identify the significant effects. Are they the same effects as found via the F-test (in part a) and the t-tests (in part b)?

```
qqnorm(all.effects)
abline(0.05,0.1) # a visual fit to the linear portion of the qqplot.
```

Yes, the outlier effects here are the same as those found significant above.

hw-lect 17-2 in R

Today's questions about "How can one use Daniel's conjecture if all the effects are significant?" made me wonder! Recall that we can always arrange for nonsignificant effects to exist - just include factors in the model that are "random," i.e., unrelated to the response. Alternatively, one can use the idea behind the permutation/randomization test we studied early in the quarter. Let's see if we can, using the data in example 6_2 discussed in class.

- Compute the 15 ($2^4 - 1$) effects (excluding the intercept), and call them `eff_obs`.
- Randomly shuffle the `y` values `ntrial = 100` times, each time computing the 15 effects. Store these effects in a 100 x 15 matrix called `eff`

Each row of `eff` has some qqplot, and so we have a 100 qqplots which we can draw on a single figure, as a cloud of points, which display the region in the qqplot that is consistent with the null-hypothesis of no-effect. Each of these qqplots has some intercept and slope estimating the mean and the standard deviation of the population of effects.

- Compute the typical mean and typical standard deviation of the 100 qqplots, and call them `typ_mean` and `typ_sd`, respectively. You may use mean or median as a measure of "typical."

We want to compare the qqplot of the observed data with the cloud of null qqplots, and so we need to use a unique set of quantiles on the x-axis of the qqplot. Let's agree to use the quantiles of $N(\text{typ_mean}, \text{typ_sd}/\sqrt{15})$. (I think you will see the reason for $1/\sqrt{15}$ later, if not already.)

- Using the "by hand" method of making qqplots from `lab2`, make a single figure that shows the cloud of all 100 qqplots. Don't forget to use the quantiles we just agreed upon.
- Superimpose the qqplot of `eff_obs` (in red color), and draw a diagonal line (in red). (The reason we're drawing a diagonal line, i.e., intercept = 0, slope = 1, is that we've already decided what quantiles should be on the x-axis.)
- Comment on the figure. Remember why we did all this! Specifically, comment on whether the figure we have made is consistent with the 5 significant effects we selected in class.

I'll make a confession: Some facets of the figure we've made don't make sense to me. Either I've made a mistake and you guys will catch it; or we (including me) will learn something new. Let's see.

Soln a)

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

design = gen.factorial(2,4,varNames=c("A","B","C","D"), factors="all")
attach(design)          # Makes the names in design visible.
design                  # Table 6.10, p. 231.
y = c(45,71,48,65,68,60,80,65,43,100,45,104,75,86,70,96)

contr = as.character("contr.helmert")
lm.1 = lm(y~A*B*C*D, contrasts = list(A=contr,B=contr,C=contr,D=contr))
eff_obs = 2*lm.1$coef[2:2^4]
```



b)

```
ntrial = 100
eff = matrix(nrow=ntrial, ncol=2^4-1)
for(trial in 1:ntrial){
  lm.1 = lm( sample(y,length(y), replace=F)~A*B*C*D,
    contrasts = list(A=contr,B=contr,C=contr,D=contr))
  eff[trial,] = 2*lm.1$coef[2:2^4]
}
```

c)

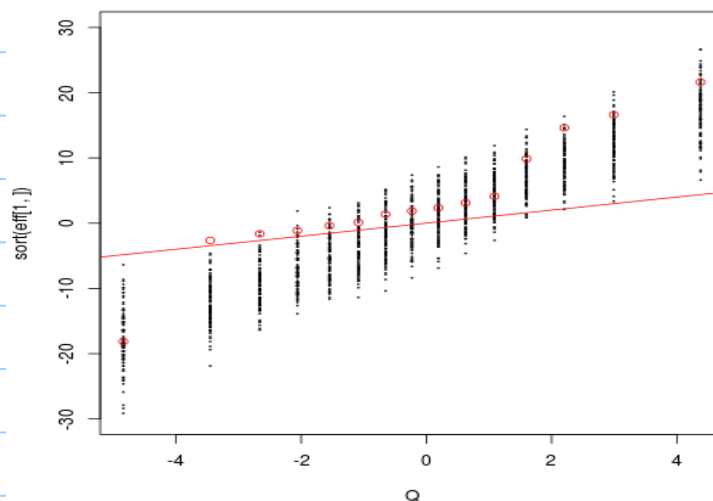
```
typ_mean = mean(apply(eff,1,mean))
typ_sd = mean(apply(eff,1,sd))
```

d)

```
n = 2^4 - 1
probs = seq(0.5/n, 1-0.5/n, length=n)
Q = qnorm(probs, typ_mean, typ_sd/sqrt(n)) # quantiles of the N() we agreed upon.
plot(Q, sort(eff[1,]), ylim=c(-30,30), cex=0.1)
for(trial in 2:ntrial)
  points(Q,sort(eff[trial,]), cex=0.1)
points(Q, sort(eff_obs), col=2)
abline(0,1, col=2)
```

f) The observed qqplot (red) falls within the cloud, with the top 4 effects close to the outer limits of the cloud. As such, that's consistent with what we got in class. And all of 5 effects we selected in class deviate significantly from the diagonal, and so this is all consistent with what we got.

The facet that doesn't make sense to me is that the slope of the cloud is not the same as the slope of diagonal line. It should be!



```
# 6.26 (6.24 in 7th ed)
rm(list=ls(all=TRUE))
library(AlgDesign)      # for gen.factorial()

design = gen.factorial(2,5,varNames=c("A","B","C","D","E"), factors="all")
# design = gen.factorial(2,4,varNames=c("A","B","C","D")) # -1/+1
attach(design)
y = c(7,9,34,55,16,20,40,60,
      8,10,32,50,18,21,44,61,
      8,12,35,52,15,22,45,65,
      6,10,30,53,15,20,41,63)

contr = as.character("contr.helmert")
lm.1 = lm(y~A*B*C*D*E, contrasts = list(A=contr,B=contr,C=contr,D=contr,E=contr))
summary.aov(lm.1)      # Full model gives SSE = 0
```

```
# a)
eff = 2*lm.1$coef
eff = eff[2:length(eff)]      # Exclude the grand mean.
as.matrix(eff,col=1)
sort(eff)
qqnorm(eff)      # B, A, C, AB
```

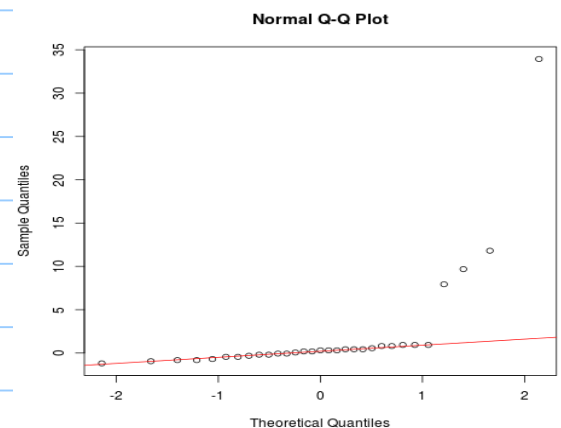
We don't even need to do the visual line-fitting business to see that the significant effects are B, A, C, and AB. But, if you want the line anyway, then either of these is good enough:

```
abline(0.2 , 0.7, col=2)
abline(median(eff), sd(eff)/sqrt(2^5-1), col=2)
```

```
b)
lm.2 = lm(y~A + B + C + A:B)
summary.aov(lm.2)
```

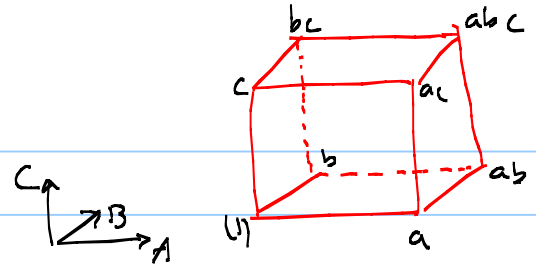
#		Df	Sum Sq	Mean Sq	F value	Pr(>F)
# A	1	1116	1116	382.3	< 2e-16 ***	
# B	1	9214	9214	3155.3	< 2e-16 ***	
# C	1	751	751	257.1	2.53e-15 ***	
# A:B	1	504	504	172.6	3.04e-13 ***	
# Residuals	27	79	3			

This anova table confirms that the effects selected according to Daniel's method are significant according the p-values as well.



hw lect 18-1

Consider a 2^3 design in 2 blocks.



- a) If the blocks are Block 1: $(1), ab, ac, bc$ Block 2: a, b, c, abc
 Then what effect is confounded with the block effect?
- b) What if the blocks are Block 1: $(1), a, bc, abc$ Block 2: b, ab, c, ac ?
- c) Find the blocks that will confound the A effect with the block effect.

a) If Block 1: $(1), ab, ac, bc$ Block 2: a, b, c, abc
 Then Block effect = $\frac{(1) + ab + ac + bc}{4} - \frac{\dots}{4}$
 $= -ABC \text{ Effect}$

∴ ABC confounded with Block

b) If Block 1: $(1), a, bc, abc$ Block 2: b, ab, c, ac
 Then Block effect = $\frac{(1) + a + bc + abc}{4} - \frac{\dots}{4}$
 $= BC \text{ Effect}$

∴ BC confounded with Block.

- c) To confound the A effect with block, we must form 2 blocks, one with $A = -1$, the other with $A = +1$. I.e.

Block 1: $\{(1), b, c, bc\}$, Block 2: $\{a, ab, ac, abc\}$

We can confirm:

Block effect = $\frac{1}{4} [(a + ab + ac + abc) - ((1) + b + c + bc)]$
 $= A \text{ effect.}$

hw-lect 18-2

Consider a 2^4 design, involving 4 blocks. a) Write the treatment combinations that must be in each of the 4 blocks (in the 1, a, b, ... notation) if we want to assure that

The ABC effect and the ACD effect are confounded with Block.

b) Do the same as part a, but this time confound ABC and BD. The +/- Table is on p. 247 in book.

Note that you get the same blocks in parts a and b. I.e. confounding ABC and ACD with blocks implies that the BD effect is confounded with block, too; which justifies things like $ABC \cdot ACD = BD$.

a) We must generate the blocks according to the +/- values of ABC and ACD in the +/- table for 2^4 .

$ABC = - \Rightarrow (1, ab, ac, bc, d, abd, acd, bcd)$

$ABC = + \Rightarrow a, b, c, abc, ad, bd, cd, abcd$

$ABC = - \text{ and } ACD = - \Rightarrow (1, ac, abd, bcd)$

= Block 1

$ABC = - \text{ and } ACD = + \Rightarrow ab, bc, d, acd$

= 2

$ABC = + \text{ , } ACD = - \Rightarrow b, abc, ad, cd$

= 3

$ABC = + \text{ , } ACD = + \Rightarrow a, c, bd, abcd$

= 4

order does not matter.

b) $ABC = - \text{ and } BD = - \Rightarrow ab, bc, d, acd$

= Block 1

$ABC = - \text{ and } BD = + \Rightarrow (1, ac, abd, bcd)$

= 2

$ABC = + \text{ , } BD = - \Rightarrow b, abc, ad, cd$

= 3

$ABC = + \text{ , } BD = + \Rightarrow a, c, bd, abcd$

= 4

order does not matter.

Note that in part b, Blocks 1 and 2 are a re-labeling of Blocks 2 and 1 in part a.

hwk18-3

In the $2^3, 4$ blocks example in the lecture above, I picked 3 specific block effects, and showed that they are equal to interaction effects. Firstly, I picked 3 block effects because that's the df of a factor with 4 levels (subject to 1 constraint). Now, suppose we pick some other set of 3 effects, e.g., $(B_1 - \bar{B}), (B_2 - \bar{B}), (B_3 - \bar{B})$, where $B_i = \text{sum (or avg.) of } y \text{ values when block factor} = i$, where $i=1, \dots, 4$. Note that I'm not considering $B_4 - \bar{B}$, because that effect is not independent of the 3 I have written. For the specific blocking shown above, show that the $(B_1 - \bar{B})$ is still related to the interaction effects. I'm intentionally not telling you what the relationship is - figure it out! Also, don't worry about over-all constants.

$$\begin{aligned}
 \underline{(B_1 - \bar{B})} &= B_1 - \frac{1}{4}(B_1 + B_2 + B_3 + B_4) = \frac{1}{4}[3B_1 - B_2 - B_3 - B_4] && \left. \begin{array}{l} B_i = \text{lecture} \\ \text{Yates' order} \end{array} \right\} \\
 &= \frac{1}{4}[3(1) + abc - (a + bc) - (b + ac) - (c + ab)] \\
 &= \frac{1}{4}[3(1) - a - b - ab - c - ac - bc + 3abc] \\
 &= \underline{AB + AC + BC} && \begin{array}{l} \text{The 3's here suggest the sum of the} \\ \text{interactions, and sure enough,} \\ \text{it's correct.} \end{array}
 \end{aligned}$$

ie. $(AB + AC + BC)$ is confounded with the block effect $(B_1 - \bar{B})$.

Moral: The specific form of effects (written in terms of conditional means of $Y_{ijk\dots}$) depends on the choice of the MC constraints. But, regardless of the choice, incomplete blocks lead to confounding.

hw-lect 18.4

This is related to 7.4

Consider the first replicate in problem 6.1.

a) Develop a full model, but with all the factors taking numeric (+1/-1) values, and make the qqplot of the effects. (This numerical condition is important in order to check the normality of the effects, i.e., Daniel's idea; if the factors are not numeric, then the effects that `lm()` returns depend on the contrast, and things get complicated. Check the example we did in class to demonstrate Daniel's idea.)

b) Based on the qqplot in part a, identify the significant effects, and develop a model, reporting the anova table. Use the principle of hierarchy, e.g., if AC is found to be significant, then you should include all lower-order terms in the model, i.e. A and C. (Also, although in this case, it will not matter, in general, the factors you use in `lm()` should be declared as factors in R, i.e., with `as.factor()`.)

c) Now, suppose the 8 runs cannot be done in one day, and that they must be done across 2 days. Use the +- table (NOT any other method we may have learned) to block the data so that the ABC effect is confounded with the block effect. Make the anova table for the full model including a block factor. Note that one of the effects is missing.

d) Again, use the qqplot of the effects to develop a blocked model. Keep the hierarchy principle in mind. Based on an anova table, is there any evidence that any of the effects are significant?

e) Finally, assume that the 8 runs must be done across 4 days, with the AC and BC effects confounded with block. Produce the anova table for a full model of the three treatment factors.

```
# a)
rm(list=ls(all=TRUE))
library(DoE.base)
library(AlgDesign)      # for gen.factorial()

design = gen.factorial(2,3,varNames=c("A","B","C")) # , factors="all")
attach(design)
y = c(22,32,35,55,44,40,60,39)
cbind(A,B,C,y)
lm.1 = lm(y ~ A*B*C)
summary.aov(lm.1)
eff1 = lm.1$coef[2:8] # Exclude intercept
sort(eff1)
qqnorm(eff1)          # Looks like the strongest effects
abline(-1,3.5, col=2) # are AC, B, C .
```

b) It looks like B, C, and AC are significant. So, the appropriate model is

#	Df	Sum Sq	Mean Sq	F value	Pr(>F)
# Af	1	3.1	3.1	0.062	0.8201
# Bf	1	325.1	325.1	6.401	0.0854 .
# Cf	1	190.1	190.1	3.743	0.1485
# Af:Cf	1	378.1	378.1	7.445	0.0720 .
# Residuals	3	152.4	50.8		

```
Af = as.factor(A) ; Bf = as.factor(B); Cf = as.factor(C)
lm.2 = lm(y ~ Af + Bf + Cf + Af:Cf)
summary.aov(lm.2)
```



c) You need to make a block factor, with ABC=1 runs in one group, and ABC=2 runs in the other. According to the +- table, the factor should be

```
D = as.factor(c(1,2,2,1,2,1,1,2))
cbind(A,B,C,D)      # visually confirm.
lm.2 = lm(y ~ D + A*B*C)
summary.aov(lm.2)
```

#		Df	Sum Sq	Mean Sq
# D		1	91.1	91.1
# A		1	3.1	3.1
# B		1	325.1	325.1
# C		1	190.1	190.1
# A:B		1	6.1	6.1
# A:C		1	378.1	378.1
# B:C		1	55.1	55.1

Note: no ABC effect (because it's confounded with)

```
# d)
eff2 = lm.2$coef[3:9] # Exclude intercept AND block effects.
sort(eff2)
qqnorm(eff2)
abline(0,4, col=2)
lm.3 = lm(y ~ D + A + B + C + A:C )
summary.aov(lm.3)
```

#		Df	Sum Sq	Mean Sq	F value	Pr(>F)
# D		1	91.1	91.1	2.976	0.2267
# A		1	3.1	3.1	0.102	0.7797
# B		1	325.1	325.1	10.616	0.0827 .
# C		1	190.1	190.1	6.208	0.1303
# A:C		1	378.1	378.1	12.347	0.0723 .
R# esiduals		2	61.3	30.6		

None of the effects are significant at the 0.05 level, but B and AC come close.

e) put AC=+ and BC=+ in level 1 of block, and AC=+ and BC=- in level 2 of block, etc.

```
D = as.factor(c(1,3,2,4, 4,2,3,1))
lm.4 = lm(y ~ D + A*B*C)
summary.aov(lm.4)
```

#		Df	Sum Sq	Mean Sq
# D		3	439.4	146.5
# A		1	3.1	3.1
# B		1	325.1	325.1
# C		1	190.1	190.1
# A:B:C		1	91.1	91.1

Note: AB, BC, and AC effects are missing. Know why.

#####

Alternatively for part d:

```
L = A*B*C      # With A,B,C, all taking -1/+1 values; this is the block factor that assures ABC is confounded with block.
lm.4 = lm(y ~ as.factor(L) + A*B*C) # Make sure Block factor is a factor.
```

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.