Math 3080 § 1.	**Peanut Oil Data: 2^5 Factorial**	Name: Example
Treibergs	**design with 1/2 Replication.**	April 22, 2010

Data File Used in this Analysis:

```
# Math 3080 - 1            Peanut Oil Data       April 22, 2010
# Treibergs
#
# From Navidi, "Principles of Statistics for Engineers and Scientists,"
# McGraw Hill 2010. Taken from a study "Application of Fractional
# Factorial Designs," (Quarterly Engineering, 1988) to compare factors
# that affect high pressure CO2 oil extraction from peanuts. The
# outcome is solubility for the peanut oil of the CO2 (mg oil/liter CO2)
#
#  Treatment
#     A        CO2 pressure
#     B        CO2 temperature
#     C        peanut moisture
#     D        CO2 flow rate
#     E        peanut particle size
#
"A"  "B"  "C"  "D"  "E" "Outcome"
-1   -1   -1   -1    1   29.2
 1   -1   -1   -1   -1   23.0
-1    1   -1   -1   -1   37.0
 1    1   -1   -1    1  139.7
-1   -1    1   -1   -1   23.3
 1   -1    1   -1    1   38.3
-1    1    1   -1    1   42.6
 1    1    1   -1   -1  141.4
-1   -1   -1    1   -1   22.4
 1   -1   -1    1    1   37.2
-1    1   -1    1    1   31.3
 1    1   -1    1   -1   48.6
-1   -1    1    1    1   22.9
 1   -1    1    1   -1   36.2
-1    1    1    1   -1   33.6
 1    1    1    1    1  172.6
```

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
  Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5537) powerpc-apple-darwin9.8.0]

> tt <- read.table("M3081DataPeanutOil.txt",header=TRUE)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'M3081DataPeanutOil.txt': No such file or directory
> tt <- read.table("M3081DataPeanutOil.txt",header=TRUE)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'M3081DataPeanutOil.txt': No such file or directory
> tt <- read.table("M3081DataPeanutOil.txt",header=TRUE)
> attach(tt)
> tt
    A  B  C  D  E Outcome
1  -1 -1 -1 -1  1    29.2
2   1 -1 -1 -1 -1    23.0
3  -1  1 -1 -1 -1    37.0
4   1  1 -1 -1  1   139.7
5  -1 -1  1 -1 -1    23.3
6   1 -1  1 -1  1    38.3
7  -1  1  1 -1  1    42.6
8   1  1  1 -1 -1   141.4
9  -1 -1 -1  1 -1    22.4
10  1 -1 -1  1  1    37.2
11 -1  1 -1  1  1    31.3
12  1  1 -1  1 -1    48.6
13 -1 -1  1  1  1    22.9
14  1 -1  1  1 -1    36.2
15 -1  1  1  1 -1    33.6
16  1  1  1  1  1   172.6
```

```
>#==========================================RUN ADDITIVE MODEL ANOVA===========
> f1 <- aov(Outcome~A+B+C+D+E);anova(f1)
Analysis of Variance Table

Response: Outcome
          Df  Sum Sq Mean Sq F value  Pr(>F)
A          1  9736.8  9736.8  7.8954 0.01848 *
B          1 10727.8 10727.8  8.6990 0.01455 *
C          1  1269.1  1269.1  1.0291 0.33428
D          1   303.6   303.6  0.2462 0.63048
E          1  1374.6  1374.6  1.1146 0.31592
Residuals 10 12332.2  1233.2
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

>#=======================SET UP CONTRASTS MATRIX================================
> E1<-rep(1,times=32);EA<-c(A,A);EB<-c(B,B);EC<-c(C,C);ED<-c(D,D)
> EE <- rep(c(-1,1),each=16);EAB<-EA*EB;EAC<-EA*EC;EAD<-EA*ED;EAE<-EA*EE
> EBC<-EB*EC;EBD<-EB*ED;EBE<-EB*EE;ECD<-EC*ED;ECE<-EC*EE;EDE<-ED*EE
> EABC<-EA*EBC;EABD<-EA*EBD;EABE<-EAB*EE;EACD<-EA*ECD;EACE<-EA*EC*EE
> EADE<-EA*EDE;EBCD<-EB*ECD;EBCE<-EBC*EE;ECDE<-EC*EDE;EABCD<-EAB*ECD
> EABCE<-EABC*EE;EABDE<-EAB*EDE;EACDE<-EA*ECDE;EBCDE<-EBC*EDE
> EABCDE<-EABC*EDE; EBDE<-EB*EDE
> labl<-c("1","a","b","ab","c","ac","bc","abc","d","ad","bd","abd","cd","acd","bcd",
 "abcd","e","ae","be","abe","ce","ace","bce","abce","de","ade","bde","abde","cde",
  "acde","bcde","abcde")

>#=============OBSERVE THAT ABCD  AND  E  ARE ALIASES (HAVE THE SAME CONTRAST)=====
> A*B*C*D-E
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> #====DEFINING FACTOR IS ABCDE BECAUSE TREATMENTS ARE THOSE WITH==================
> #==== ABCDE=1 ==================================================================

>#=============THE EXPERIMENTAL TREATMENTS ARE ALL TREATMENT COMBINATIONS WITH=====
>#=============DEFINING CONTRAST  ABCDE=1 ========================================

>#==================== "order" GIVES THE SUBSCRIPTS SO ABCDE[I] ARE INCREASING=====
>#==================== IN THIS CASE THE  ABCDE=-1 ARE THE FIRST SIXTEEN============
> I <- order(EABCDE)

>#======THE FIRST "BLOCK" ARE TREATMENTS WITH ABCDE=-1 OR HAVE EVEN NO. FACTORS IN==
>#======COMMON WITH THE DEFINING FACTOR  ABCDE====================================
> labl[I[1:16]]
 [1] "1"    "ab"   "ac"   "bc"   "ad"   "bd"   "cd"   "abcd"
 [9] "ae"   "be"   "ce"   "abce" "de"   "abde" "acde" "bcde"

>#====THE OTHER TREATMENT BLOCK IS THE ONE USED IN THIS 1/2 REPLICATE EXPERIMENT====
> labl[I[17:32]]
 [1] "a"     "b"     "c"     "abc"   "d"     "abd"   "acd"
 [8] "bcd"   "e"     "abe"   "ace"   "bce"   "ade"   "bde"
[15] "cde"   "abcde"
```

```
>#=================COMPUTE CELL TOTALS. EVERY OTHER CELL IS EMPTY (1/2 REP.!)======
> xtabs(Outcome~A+B+C+D+E)
, , C = -1, D = -1, E = -1
   B
A      -1     1
  -1   0.0  37.0
  1   23.0   0.0

, , C = 1, D = -1, E = -1
   B
A      -1     1
  -1  23.3   0.0
  1    0.0 141.4

, , C = -1, D = 1, E = -1
   B
A      -1     1
  -1  22.4   0.0
  1    0.0  48.6

, , C = 1, D = 1, E = -1
   B
A      -1     1
  -1   0.0  33.6
  1   36.2   0.0

, , C = -1, D = -1, E = 1
   B
A      -1     1
  -1  29.2   0.0
  1    0.0 139.7

, , C = 1, D = -1, E = 1
   B
A      -1     1
  -1   0.0  42.6
  1   38.3   0.0

, , C = -1, D = 1, E = 1
   B
A      -1     1
  -1   0.0  31.3
  1   37.2   0.0

, , C = 1, D = 1, E = 1
   B
A      -1     1
  -1  22.9   0.0
  1    0.0 172.6
```

```
>#===============THESE GIVE CELL SUMS. PICK OFF THE NONZERO ONES============
> VV<-as.vector(xtabs(Outcome~A+B+C+D+E))
> VV
 [1]   0.0  23.0  37.0   0.0  23.3   0.0   0.0 141.4  22.4   0.0
[11]   0.0  48.6   0.0  36.2  33.6   0.0  29.2   0.0   0.0 139.7
[21]   0.0  38.3  42.6   0.0   0.0  37.2  31.3   0.0  22.9   0.0
[31]   0.0 172.6

>#==========NONZERO CELLS ARE THOSE WITH  ABCDE=+1=========================
>#==========THEY ARE JUST THE 16 OBSERVED VALUES============================
> VY<-VV[I[17:32]];VY
 [1]  23.0  37.0  23.3 141.4  22.4  48.6  36.2  33.6  29.2 139.7
[11]  38.3  42.6  37.2  31.3  22.9 172.6


>#=============COMPUTE SST, CONTRASTS AND SUM SQUARES======================
> SST<-sum(VY*VY)-sum(VY)^2/16;SST
[1] 35744.02
> LA <- sum(EA[I[17:32]]*VY);SSA<-LA*LA/16;SSA
[1] 9736.756
> LB <- sum(EB[I[17:32]]*VY);SSB<-LB*LB/16;SSB
[1] 10727.78
> LC <- sum(EC[I[17:32]]*VY);SSC<-LC*LC/16;SSC
[1] 1269.141
> LD <- sum(ED[I[17:32]]*VY);SSD<-LD*LD/16;SSD
[1] 303.6306
> LE <- sum(EE[I[17:32]]*VY);SSE<-LE*LE/16;SSE
[1] 1374.556
> SSError<-SST-SSA-SSB-SSC-SSD-SSE;SSError
[1] 12332.16
```

```
>#==============BUILD ANOVA TABLE=========================================
> DF<-c(1,1,1,1,1,10,15);SS<-c(SSA,SSB,SSC, SSD,SSE,SSError,SST)
> MSE<-SSError/10;MS<-c(SSA,SSB,SSC, SSD,SSE,MSE,-1)
> F<-MS/MSE;F[6]<--1;F[7]<- -1
> P<-pf(F,1,10,lower.tail=FALSE)
> matrix(c(DF,SS,MS,F,P),ncol=5,
  dimnames=list(c("A","B","C","D","E","Error","Total"),c("DF","SS","MS","F","P(>F)")))

      DF         SS         MS          F      P(>F)
A      1  9736.7556  9736.7556  7.8954203 0.01847872
B      1 10727.7806 10727.7806  8.6990307 0.01454929
C      1  1269.1406  1269.1406  1.0291312 0.33427872
D      1   303.6306   303.6306  0.2462105 0.63048062
E      1  1374.5556  1374.5556  1.1146109 0.31592039
Error 10 12332.1563  1233.2156 -1.0000000 1.00000000
Total 15 35744.0194    -1.0000 -1.0000000 1.00000000


>#=========================COMPARE TO=====================================
> anova(f1)
Analysis of Variance Table

Response: Outcome
          Df  Sum Sq Mean Sq F value  Pr(>F)
A          1  9736.8  9736.8  7.8954 0.01848 *
B          1 10727.8 10727.8  8.6990 0.01455 *
C          1  1269.1  1269.1  1.0291 0.33428
D          1   303.6   303.6  0.2462 0.63048
E          1  1374.6  1374.6  1.1146 0.31592
Residuals 10 12332.2  1233.2
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
>#==================================MEANS=============================================
> model.tables(f1,type="means")
Tables of means
Grand mean

54.95625

 A
A
   -1     1
30.29 79.62

 B
B
   -1     1
29.06 80.85

 C
C
   -1     1
46.05 63.86

 D
D
   -1     1
59.31 50.60

 E
E
   -1     1
45.69 64.23

>#=======================TABLE OF EFFECTS & PCT. VARIABILITY "BY HAND"=======================
> EFF <- c(LA/16,LB/16,LC/16,LD/16,LE/16)
> PCT <- c(SSA/SST,SSB/SST,SSC/SST,SSD/SST,SSE/SST)
> matrix(c(EFF,PCT),ncol=2,dimnames=list(c("A","B","C","D","E"),c("Effect","Frac.Variability")))

    Effect Frac.Variability
A 24.66875      0.272402371
B 25.89375      0.300127988
C  8.90625      0.035506377
D -4.35625      0.008494585
E  9.26875      0.038455542


>#================R SQ = FRACTION OF VARIABILITY ACCOUNTED FOR BY MODEL=================
> 1-SSError/SST
[1] 0.6549869
```

```
>#=======TABLE OF CONTRAST COEFFICIENTS FOR TREATMENT COMBINATIONS OF THEIS DESIGN=====
> J<-I[17:32]
> matrix(c(EA[J],EB[J],EC[J],ED[J],EE[J],VY),ncol=6,
  dimnames=list(labl[J], c("A","B","C","D","E","Outcome")))
```

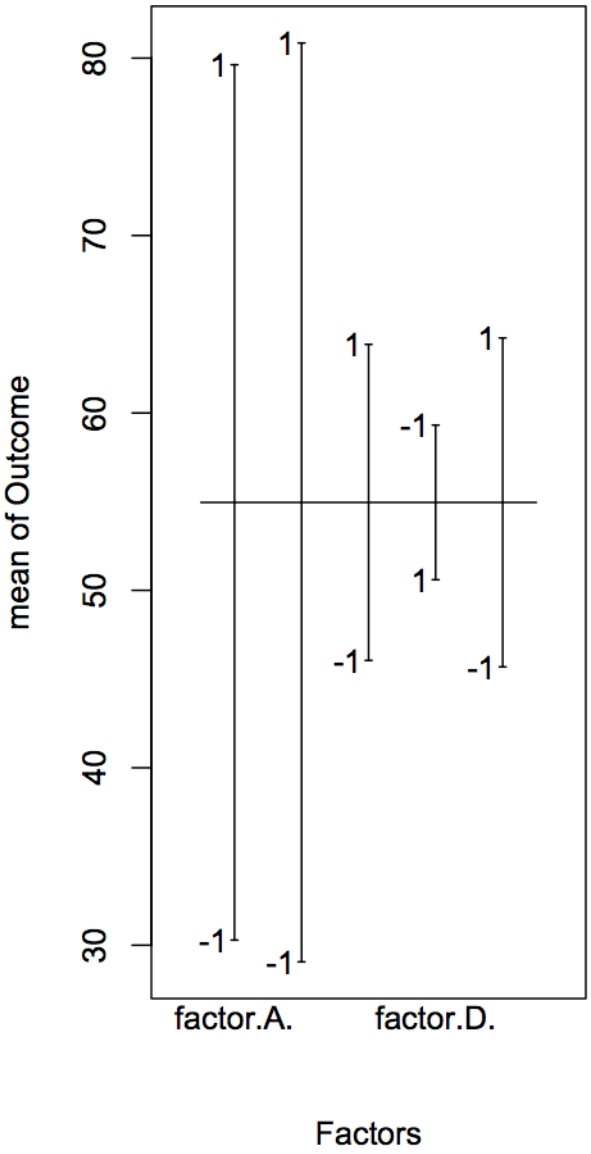
	A	B	C	D	E	Outcome
a	1	-1	-1	-1	-1	23.0
b	-1	1	-1	-1	-1	37.0
c	-1	-1	1	-1	-1	23.3
abc	1	1	1	-1	-1	141.4
d	-1	-1	-1	1	-1	22.4
abd	1	1	-1	1	-1	48.6
acd	1	-1	1	1	-1	36.2
bcd	-1	1	1	1	-1	33.6
e	-1	-1	-1	-1	1	29.2
abe	1	1	-1	-1	1	139.7
ace	1	-1	1	-1	1	38.3
bce	-1	1	1	-1	1	42.6
ade	1	-1	-1	1	1	37.2
bde	-1	1	-1	1	1	31.3
cde	-1	-1	1	1	1	22.9
abcde	1	1	1	1	1	172.6

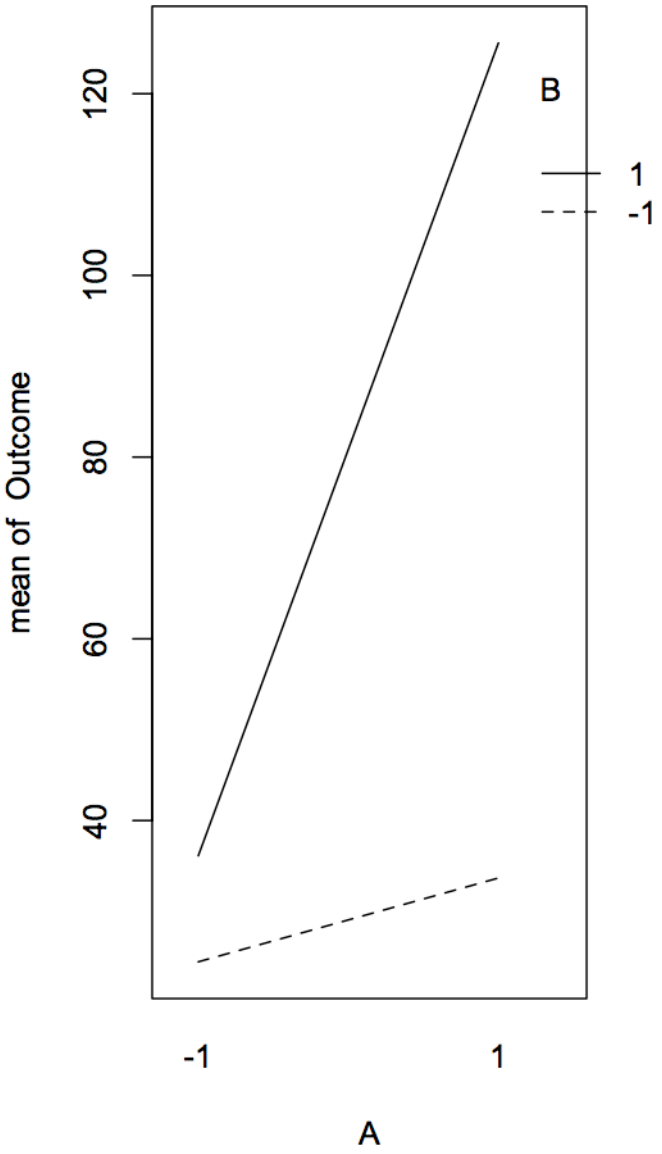
```
>
```

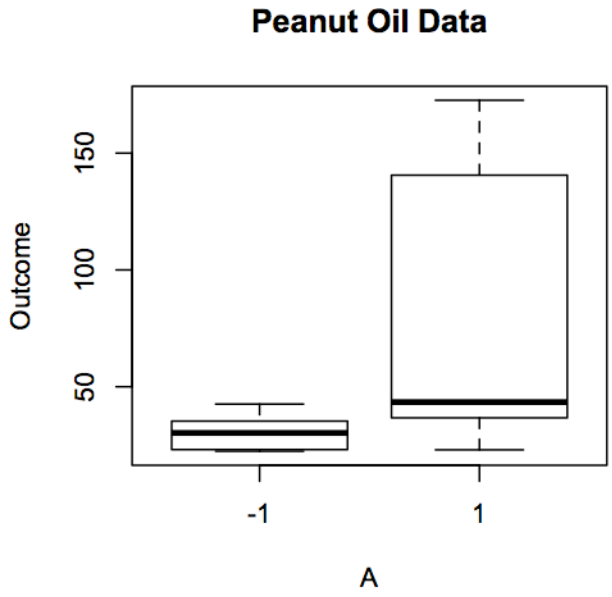
```
>#==============TABLE OF ALIASED PAIRS. FIRST IS ALIAS OF SECOND====================
>#==============SINCE FIRST * SECOND = ABCDE, THE DEFINING CONTRAST, BOTH===========
>#==============HAVE SAME CONTRAST RESTRICTED TO THE "EXPERIMENTAL BLOCK"===========
>#==============LISTING COL 2 BACKWARDS MATCHES UP THE ALIASED CONTRASTS============
>#==============THE SECOND COLUMN (ABCD=1) EXPERIMENTAL TREATMENT COMBOS============

> matrix(c(labl[I[1:16]],labl[J[seq(from=16,to=1,by= -1)]]),ncol=2)
```

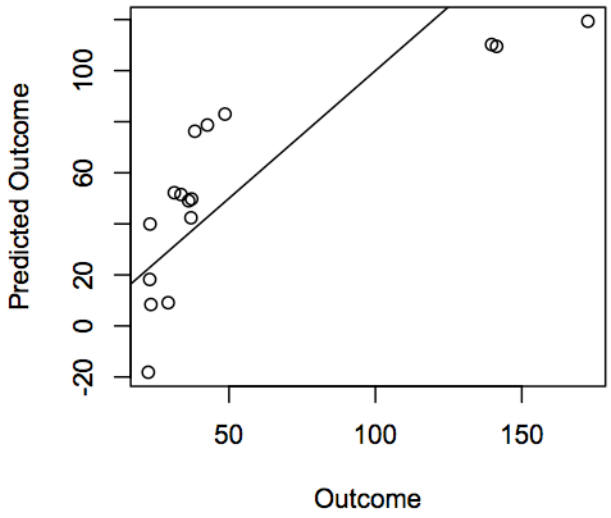
	[,1]	[,2]
[1,]	"1"	"abcde"
[2,]	"ab"	"cde"
[3,]	"ac"	"bde"
[4,]	"bc"	"ade"
[5,]	"ad"	"bce"
[6,]	"bd"	"ace"
[7,]	"cd"	"abe"
[8,]	"abcd"	"e"
[9,]	"ae"	"bcd"
[10,]	"be"	"acd"
[11,]	"ce"	"abd"
[12,]	"abce"	"d"
[13,]	"de"	"abc"
[14,]	"abde"	"c"
[15,]	"acde"	"b"
[16,]	"bcde"	"a"

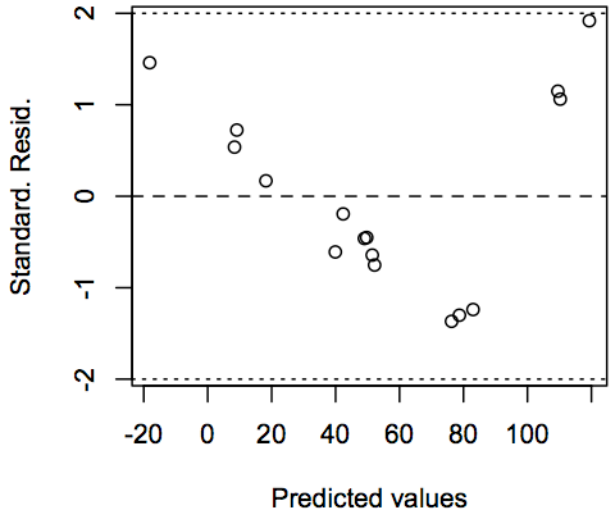
```
>#=============PLOT STANDARD DIAGNOSTICS & SHAPIRO-WILK TEST FOR NORMALITY=============
> layout(matrix(1:2,ncol=2))
> plot.design(data.frame(Outcome,factor(A),factor(B),factor(C),factor(D),factor(E)))
> interaction.plot(A,B,Outcome)
> layout(matrix(1:4,ncol=2))
> plot(Outcome~factor(A),xlab="A",main="Peanut Oil Data")
> plot(rstandard(f1)~fitted(f1),ylab="Standard. Resid.",xlab="Predicted values",
  ylim=max(abs(rstandard(f1)))*c(-1,1));abline(h=c(0,-2,2),lty=c(2,3,3))
> plot(fitted(f1)~Outcome,ylab="Predicted Outcome");abline(0,1)
> qqnorm(rstandard(f1),ylab="Standard. Resid.",ylim=max(abs(rstandard(f1)))*c(-1,1))
> abline(h=c(0,-2,2),lty=c(2,3,3));abline(0,1)
>
> shapiro.test(rstandard(f1))

Shapiro-Wilk normality test

data:  rstandard(f1)
W = 0.9403, p-value = 0.3523

>
```

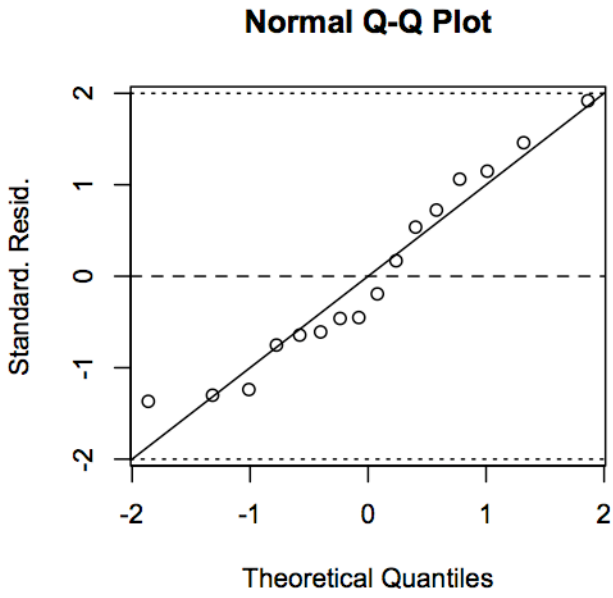

mean of Outcome
80
70
60
50
40
30
1
1
1
-1
1
-1
1
-1
-1
-1
factor.A.
factor.D.
Factors


mean of Outcome
120
100
80
60
40
B
1
-1
-1
1
A


Peanut Oil Data
Outcome
50
100
150
-1
1
A


Predicted Outcome
-20
0
20
60
100
50
100
150
Outcome


Standard. Resid.
-2
-1
0
1
2
-20
0
20
40
60
80
100
Predicted values


Normal Q-Q Plot
Standard. Resid.
-2
-1
0
1
2
-2
-1
0
1
2
Theoretical Quantiles