STAT 500

Additional Factorial Designs

Factorial Designs with Blocking

- RCBD with full factorial treatment design ($r = ab$ treatments)
- Different random assignment of units to treatments in each of the n blocks
  - One experimental unit for each block each treatment
  - Assume no interaction between block and treatment effects
  - Model: $Y_{ijk} = \mu + \beta_i + \alpha_j + \tau_k + (\alpha\tau)_{jk} + \epsilon_{ijk}$

Factorial Designs with Blocking

source of variation	degrees of freedom	Sums of Squares
Block	$n-1$	$ab\sum_{i=1}^{n}(\bar{Y}_{i\cdot\cdot}-\bar{Y}_{\cdots})^2$
Factor A	$a-1$	$nb\sum_{j=1}^{a}(\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdots})^2$
Factor B	$b-1$	$na\sum_{k=1}^{b}(\bar{Y}_{\cdot\cdot k}-\bar{Y}_{\cdots})^2$
A×B interaction	$(a-1)(b-1)$	$n\sum_j\sum_k(\bar{Y}_{\cdot jk}-\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2$
Error	$(ab-1)(n-1)$	subtraction
Corrected total	$abn-1$	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{\cdots})^2$

$$
\begin{aligned}
SS_{error} &= b\sum_i\sum_j(\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdot\cdot}-\bar{Y}_{\cdot j\cdot}+\bar{Y}_{\cdots})^2 \\
&\quad +a\sum_i\sum_j(\bar{Y}_{i\cdot k}-\bar{Y}_{i\cdot\cdot}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2 \\
&\quad +\sum_i^i\sum_j^k\sum_k(Y_{ijk}-\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdot k}-\bar{Y}_{\cdot jk}+\bar{Y}_{i\cdot\cdot}+\bar{Y}_{\cdot j\cdot}+\bar{Y}_{\cdot\cdot k}-\bar{Y}_{\cdots})^2
\end{aligned}
$$

Three Factor Experiments

- Factor A with a levels: $i = 1, \ldots, a$
- Factor B with b levels: $j = 1, \ldots, b$
- Factor C with c levels: $k = 1, \ldots, c$
- n replications for each treatment: $l = 1, \ldots, n$

$$Y_{ijkl} = \mu + \alpha_i + \tau_j + \delta_k + (\alpha\tau)_{ij} + (\alpha\delta)_{ik} + (\tau\delta)_{jk} + (\alpha\tau\delta)_{ijk} + \epsilon_{ijkl}$$

Three-Factor Experiment

- α_i = Factor A effect
- τ_j = Factor B effect
- δ_k = Factor C effect
- $(\alpha\tau)$ = interaction of Factors A and B
- $(\alpha\delta)$ = interaction of Factors A and C
- $(\tau\delta)$ = interaction of Factors B and C
- $(\alpha\tau\delta)$ = interaction of Factors A, B, and C

Example

- Experimental units: 36 water tanks containing minnow larvae
- Factor A = Nickel (3 levels), Factor B = Copper (2 levels), Factor C = Zinc (3 levels)
- Experimental units are randomly assigned to the 18 treatments = 2 units per treatment.
- Response Variable: protein content (μg/larva)

Three-Factor Experiment

source of variation	degrees of freedom	sums of squares
factor A	$a-1$	$nbc\sum_i(\bar{Y}_{i\cdots}-\bar{Y}_{\cdots})^2$
factor B	$b-1$	$nac\sum_j(\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdots})^2$
factor C	$c-1$	$nab\sum_k(\bar{Y}_{\cdot\cdot k}-\bar{Y}_{\cdots})^2$
interaction AB	$(a-1)(b-1)$	$nc\sum_i\sum_j(\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdots}-\bar{Y}_{\cdot j\cdot}+\bar{Y}_{\cdots})^2$
interaction AC	$(a-1)(c-1)$	$nb\sum_i\sum_k(\bar{Y}_{i\cdot k}-\bar{Y}_{i\cdots}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2$
interaction BC	$(b-1)(c-1)$	$na\sum_j\sum_k(\bar{Y}_{\cdot jk}-\bar{Y}_{\cdot j\cdot}-\bar{Y}_{\cdot\cdot k}+\bar{Y}_{\cdots})^2$
interaction ABC	$(a-1)(b-1)(c-1)$	$SS(ABC)$
error	$abc(n-1)$	$\sum_i\sum_j\sum_k\sum_l(Y_{ijkl}-\bar{Y}_{ijk\cdot})^2$
total	$abcn-1$	$\sum_i\sum_j\sum_k\sum_l(Y_{ijkl}-\bar{Y}_{\cdots})^2$

$$SS(ABC)=n\sum_i\sum_j\sum_k(\bar{Y}_{ijk\cdot}-\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdot k}-\bar{Y}_{\cdot jk}+\bar{Y}_{i\cdots}+\bar{Y}_{\cdot j\cdot}+\bar{Y}_{\cdot\cdot k}-\bar{Y}_{\cdots})^2$$

Three-Factor Experiment

- Main effects and simple effects are defined as before.
- Main effects are differences (or contrasts) between levels of one factor averaged over all levels of the other factors. Correspond to marginal means that are averages over "left out" factors and replicates.
  - e.g., difference between two copper levels averaged over all Zinc levels and Nickel levels
- Simple effects correspond to differences between levels of one factor at specific levels of the other factors.
  - e.g., difference between two copper levels at Zinc level 1 and Nickel level 1.

Three-Factor Experiment

- Interpretations of factor effects are complicated by the presence of three-factor interaction
- Example - A, B, C
  - AB interaction: interaction between factor A and factor B (averaging across the levels of factor C)
  - ABC interaction: nature of the AB interaction depends on the level of factor C
    * Remember 2-way interaction: Are the simple effects of A the same at every level of B?
    * Three-way interaction generalizes this: Are the 2-way A*B interaction effects the same for each level of C?

Three-Factor Experiment

Example: Is there 3-way interaction? The cell means:

level of C	level of B	Cell mean for: A=1	A=2
1	1	6	6
1	2	1	3
2	1	2	4
2	2	1	5

- When C = 1, the interaction effect for A*B is: (6-1)-(6-3)=2
- When C = 2, the interaction effect for A*B is: (2-1)-(4-5) = 2

Three-Factor Experiment

- Because the 2-way interaction A*B does not depend on the level of C, there is no 3-way interaction.
- Four-way interaction (if four factors): Are the 3-way A*B*C interaction effects the same for each level of D?
- Such concepts extend to many factors.
- Often (not always) magnitude of main effect $>$ that of 2-way interactions $>$ 3-way $>$ ... $>$ high order interaction

Factorial Designs with No Replication

- K factors with $j = 1, 2, ..., r_k$ levels for the kth factor
- Known as a $r_1 \times r_2 \times \cdots \times r_k$ factorial designs
- There are $r_1 \times r_2 \times \cdots \times r_k$ experimental units, and exactly one unit is assigned to each treatment.
- If all possible interactions are included in the model, there are no degrees of freedom left for computing MS_{error}
- We will only consider the special situation of K factors with exactly two levels for each factor - 2^K factorial designs

Factorial Designs
No Replication – 2^K Studies

Some special features

- All main effect and interaction contrasts have 1 d.f.
- Set up the model with
  - One column in the model matrix X for each main effect using +1/-1 coding
  - Interaction columns are obtained by multiplication of appropriate main effect columns
  - All columns of X are orthogonal

Factorial Designs
No Replication – 2^K Studies

- With no replication there are no degrees of freedom for error when you fit all possible main effects and interactions
- Pool sums of squares from non-significant interaction terms to obtain a mean square error value

Factorial Designs
No Replication – 2^K Studies

Determination of non-significant terms:

- Construct a normal probability plot: Order estimated effects (except the overall mean/intercept) from smallest to largest and plot against expected quantiles from a N(0,1) distribution
- All estimated effects must have same variance
- The slope of the line provides an estimate of the common standard deviation

Factorial Designs
No Replication – 2^K Studies

- Estimates far from the line indicate "non-zero" values that do not correspond to random error
- Estimates close to the line reflect random error; pool their corresponding sums of squares to obtain a mean square error value
- Use this estimate of σ^2 to test the significance of the "non-zero" contrasts, construct CI's, etc...

Chemical Process Study (Box, Hunter & Hunter)

- Examine the effects of four factors on a chemical process
  - Determine which factors affect the conversion percentage (response)
  - Look for interactions
- Factors
  - Factor 1: Amount of a catalyst (α)
    ($i = 1$) 10 lb (coded as 1)
    ($i = 2$) 15 lb (coded as -1)

- Factor 2: Temperature (τ)
  - ($j = 1$) 220^{o}C (coded as 1)
  - ($j = 2$) 240^{o}C (coded as -1)
- Factor 3: Pressure (γ)
  - ($k = 1$) 50 psi (coded as 1)
  - ($k = 2$) 80 psi (coded as -1)
- Factor 4: Concentration of substance to be converted (δ)
  - ($l = 1$) 10% (coded as 1)
  - ($l = 2$) 15% (coded as -1)

Chemical Process Study

- Randomization: Order in which various combinations of factor levels were run was randomized
- Data

Catalyst	Temp.	Pressure	Conc.	Y	Run order
10	220	50	10	71	8
15	220	50	10	61	2
10	240	50	10	90	10
15	240	50	10	82	4
10	220	80	10	68	15
15	220	80	10	61	9
10	240	80	10	87	1
15	240	80	10	80	13
10	220	50	15	61	16
15	220	50	15	50	5
10	240	50	15	89	11
15	240	50	15	83	14
10	220	80	15	59	3
15	220	80	15	51	12
10	240	80	15	85	6
15	240	80	15	78	7

Chemical Process Study

Model: $\mathbf{Y} = \mathbf{X}\beta + \epsilon$

$$\begin{pmatrix} Y_{1111} \\ Y_{2111} \\ Y_{1211} \\ Y_{2211} \\ Y_{1121} \\ Y_{2121} \\ Y_{1221} \\ Y_{2221} \\ Y_{1112} \\ Y_{2112} \\ Y_{1212} \\ Y_{2212} \\ Y_{1122} \\ Y_{2122} \\ Y_{1222} \\ Y_{2222} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Chemical Process Study

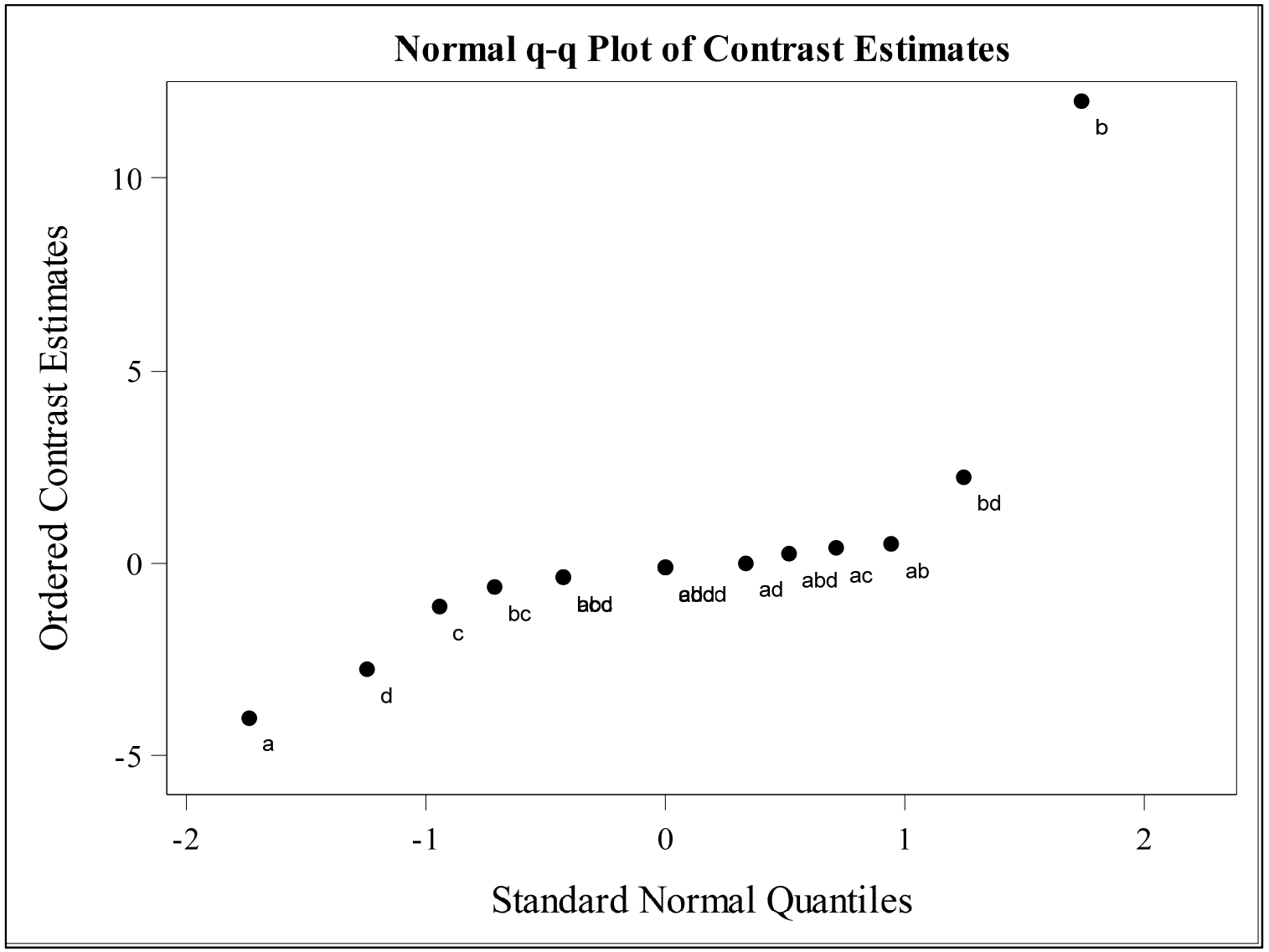
$$
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & 1 & -1 \\
-1 & -1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 \\
-1 & 1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 \\
1 & -1 & -1 & 1 & 1 \\
-1 & 1 & 1 & 1 & -1 \\
1 & -1 & -1 & -1 & -1 \\
-1 & 1 & 1 & -1 & 1 \\
-1 & 1 & -1 & 1 & 1 \\
1 & -1 & 1 & 1 & -1 \\
-1 & -1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & -1 \\
1 & 1 & 1 & -1 & -1 \\
-1 & -1 & -1 & -1 & 1
\end{pmatrix}
\begin{pmatrix}
\mu \\
\alpha_1 \\
\tau_1 \\
\gamma_1 \\
\delta_1 \\
(\alpha\tau)_{11} \\
(\alpha\gamma)_{11} \\
(\alpha\delta)_{11} \\
(\tau\gamma)_{11} \\
(\tau\delta)_{11} \\
(\gamma\delta)_{11} \\
(\alpha\tau\gamma)_{111} \\
(\alpha\tau\delta)_{111} \\
(\alpha\gamma\delta)_{111} \\
(\tau\gamma\delta)_{111} \\
(\alpha\tau\gamma\delta)_{1111}
\end{pmatrix}
+
\begin{pmatrix}
\epsilon_{1111} \\
\epsilon_{2111} \\
\epsilon_{1211} \\
\epsilon_{2211} \\
\epsilon_{1121} \\
\epsilon_{2121} \\
\epsilon_{1221} \\
\epsilon_{2221} \\
\epsilon_{1112} \\
\epsilon_{2112} \\
\epsilon_{1212} \\
\epsilon_{2212} \\
\epsilon_{1122} \\
\epsilon_{2122} \\
\epsilon_{1222} \\
\epsilon_{2222}
\end{pmatrix}
$$

Chemical Process Study

- Least Squares Estimates $\quad \widehat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$
  with covariance matrix $Var(\widehat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} = \frac{\sigma^2}{2^K}\,\mathbf{I}$
- Order the $2^K - 1$ estimates (other than the overall mean) and plot the i-th smallest against $q_i = \Phi(\frac{i-.375}{2^K-1+0.25})$ (Blom approximation)
- Points that deviate from a straight line pattern indicate significant effects
- Slope of line fit to the remaining points provides an estimate of $\sqrt{\sigma^2/2^K}$

Least Squares Estimates

$$
\begin{pmatrix}
\widehat{\alpha}_1 = -4.000 \\
\widehat{\delta}_1 = -2.750 \\
\widehat{\gamma}_1 = -1.125 \\
\widehat{(\tau\gamma)}_{11} = -0.625 \\
\widehat{(\alpha\tau\gamma)}_{111} = -0.375 \\
\widehat{(\tau\gamma\delta)}_{111} = -0.375 \\
\widehat{(\gamma\delta)}_{11} = -0.125 \\
\widehat{(\alpha\gamma\delta)}_{111} = -0.125 \\
\widehat{(\alpha\tau\gamma\delta)}_{1111} = -0.125 \\
\widehat{(\alpha\delta)}_{11} = 0.000 \\
\widehat{(\alpha\tau\delta)}_{111} = 0.250 \\
\widehat{(\alpha\gamma)}_{11} = 0.375 \\
\widehat{(\alpha\tau)}_{11} = 0.500 \\
\widehat{(\tau\delta)}_{11} = 2.250 \\
\widehat{\tau}_1 = 12.000
\end{pmatrix}
$$


Normal q-q Plot of Contrast Estimates
Ordered Contrast Estimates
10
5
0
-5
-2
-1
0
1
2
Standard Normal Quantiles
b
bd
ab
ac
abd
ad
abdd
abd
bc
c
d
a

Chemical Process Study
ANOVA for Selected Effects

source of variation	df	sums of squares	Mean Squares	F	P-value
Catalyst	1	256	256	136.5	$< .0001$
Temp.	1	2304	2304	1228.8	$< .0001$
Pressure	1	20.25	20.25	10.8	.0082
Conc.	1	121	121	64.5	$< .0001$
Temp$\times$Conc	1	81	81	43.2	$< .0001$
Residuals	10	18.75	1.875		
total	15	2801.00			

Chemical Process Study: SAS Code

```
/*  Program to analyze an unreplicated factorial experiment
    where each factor has two levels. This program is posted in the S
    as  process_FactExpNoRep.sas  It is applied to to the process dev
    data from Box, Hunter, and Hunter.          */

Data set1;
  input a b c d y order;
  a = (a-12.5)/2.5;
  b = (b-230)/10;
  c = (c-65)/15;
  d = (d-11)/1;
  ab=a*b; ac=a*c; ad=a*d; bc=b*c; bd=b*d; cd=c*d;
  abc=a*b*c; abd=a*b*d; acd=a*c*d; bcd=b*c*d; abcd=a*b*c*d;
datalines;
```

```
10 220 50 10 71  8
15 220 50 10 61  2
10 240 50 10 90 10
15 240 50 10 82  4
10 220 80 10 68 15
15 220 80 10 61  9
10 240 80 10 87  1
15 240 80 10 80 13
10 220 50 12 61 16
15 220 50 12 50  5
10 240 50 12 89 11
15 240 50 12 83 14
10 220 80 12 59  3
15 220 80 12 51 12
10 240 80 12 85  6
15 240 80 12 78  7
run;
```

```
/* Print the data file.  Values of the explanatory
    variables have been converted  to  1  or -1   */

    proc print data=set1;  run;

/*  Fit the  model that produces estimated effects.
    The least squares estimates are written to a temporary file */

proc glm data=set1;
  model y = a b c d ab ac ad bc bd cd abc abd acd bcd abcd;
  ods output ParameterEstimates=Parms;   run;
```

```
/*  Print the file containing the estimated coefficients    */

proc print data=parms; run;

/* Delete the intercept estimate  */

data parms; set parms;
  if(Parameter="Intercept") then delete;  run;

/* Create a normal probablity plot  */

proc univariate data=parms normal;
  var estimate;  id Parameter;
  qqplot;   run;
```

```
/*  Use SGPLOT to create a q-q plot with effects labels */

proc rank data=parms out=parms normal=blom;
  var estimate; ranks q; run;

proc sgplot data=parms;
  scatter x=q y=estimate /
     markerattrs=(size=10 symbol=CircleFilled color=black)
 datalabel=Parameter  datalabelattrs=(family=Ariel size=8)
 datalabelpos=bottomright;
  yaxis label="Ordered Contrast Estimates"
          labelattrs=(size=14) valueattrs=(size=12);
  xaxis label="Standard Normal Quantiles"
          labelattrs=(size=14) valueattrs=(size=12);
  title h=1.5 "Normal q-q Plot of Contrast Estimates";
  run;
```

```
/*  Fit a model containing only the main effects
    and the significant interactions identified
    in the normal probability plot.              */


 proc glm data=set1;
    class a b c d;
    model y = a b c d b*d / solution clparm;
    output out=set2 residual=r;   run;

  proc univariate data=set2 normal;
    qqplot;
    var r;   run;
```