Unreplicated 2^k Factorial Designs

- These are 2^k factorial designs with **one observation** at each corner of the "cube"
- An unreplicated 2^k factorial design is also sometimes called a "**single replicate**" of the 2^k
- These designs are very widely used
- Risks…if there is only one observation at each corner, is there a chance of unusual response observations spoiling the results?
- Modeling "noise"?

Spacing of Factor Levels in the Unreplicated 2^k Factorial Designs

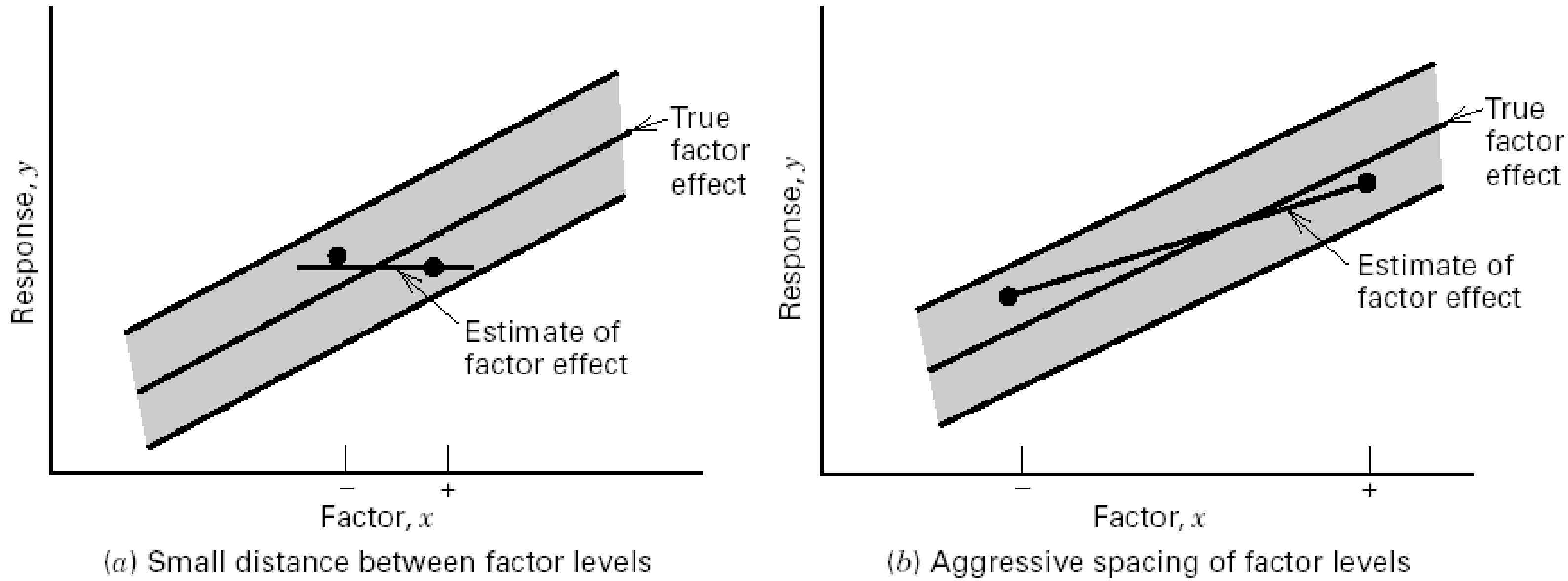


Figure 6-9 The impact of the choice of factor levels in an unreplicated design.

If the factors are spaced too closely, it increases the chances that the noise will overwhelm the signal in the data

More aggressive spacing is usually best

Unreplicated 2^k Factorial Designs

- Lack of replication causes potential **problems** in statistical testing
  - Replication admits an estimate of “pure error” (a better phrase is an **internal estimate** of error)
  - With no replication, fitting the full model results in zero degrees of freedom for error
- Potential **solutions** to this problem
  - Pooling high-order interactions to estimate error
  - **Normal probability plotting** of effects (Daniels, 1959)
  - Other methods…see text, pp. 234

Example of an Unreplicated 2^k Design

- A 2^4 factorial was used to investigate the effects of four factors on the filtration rate of a resin
- The factors are *A* = temperature, *B* = pressure, *C* = mole ratio, *D*= stirring rate
- Experiment was performed in a pilot plant

The Resin Plant Experiment

Table 6-10 Pilot Plant Filtration Rate Experiment

Run Number	Factor A	B	C	D	Run Label	Filtration Rate (gal/h)
1	—	−	−	−	(1)	45
2	+	−	−	−	*a*	71
3	−	+	−	−	*b*	48
4	+	+	−	−	*ab*	65
5	−	−	+	−	*c*	68
6	+	−	+	−	*ac*	60
7	−	+	+	−	*bc*	80
8	+	+	+	−	*abc*	65
9	−	−	−	+	*d*	43
10	+	−	−	+	*ad*	100
11	−	+	−	+	*bd*	45
12	+	+	−	+	*abd*	104
13	−	−	+	+	*cd*	75
14	+	−	+	+	*acd*	86
15	−	+	+	+	*bcd*	70
16	+	+	+	+	*abcd*	96

The Resin Plant Experiment

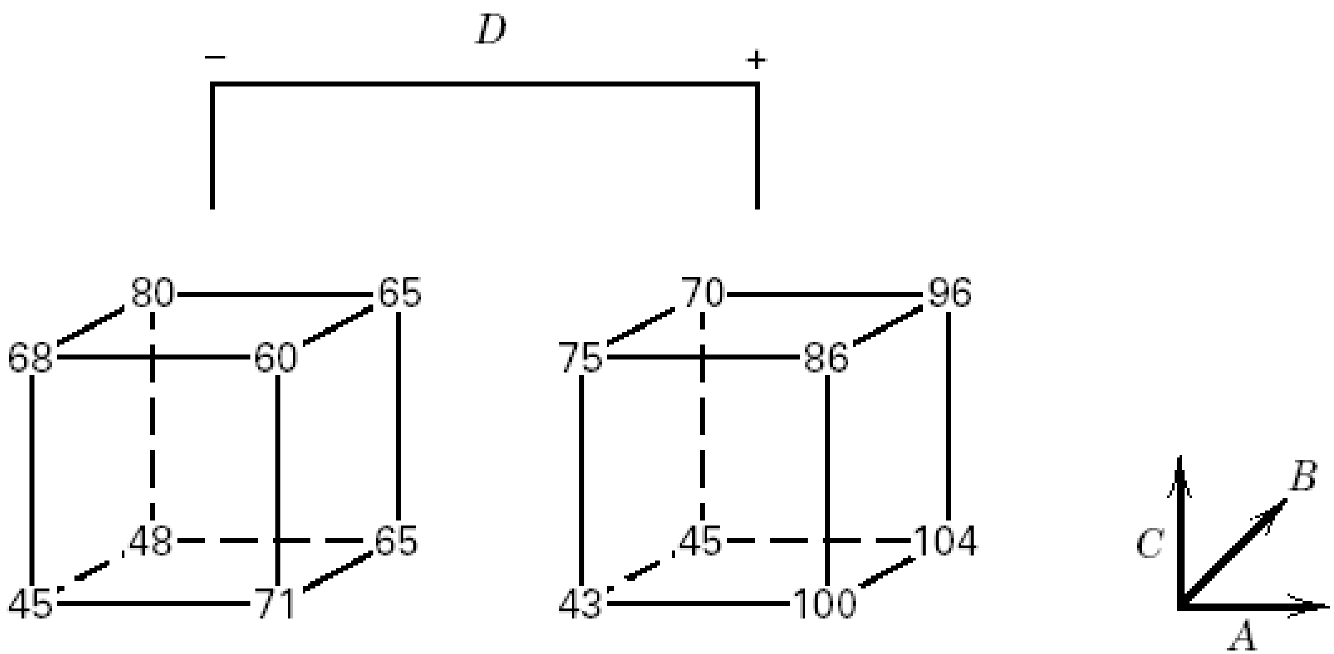


Figure 6-10 Data from the pilot plant filtration rate experiment for Example 6-2.

Main Effect

A	B	C	D
21.625	3.125	9.875	14.625

Two-way Interaction

AB	AC	AD	BC	BD	CD
0.125	-18.125	16.625	2.375	-0.375	-1.125

Three-way interaction

ABC	ABD	ACD	BCD
1.875	4.125	-1.625	-2.625

Four-way Interaction

ABCD
1.375

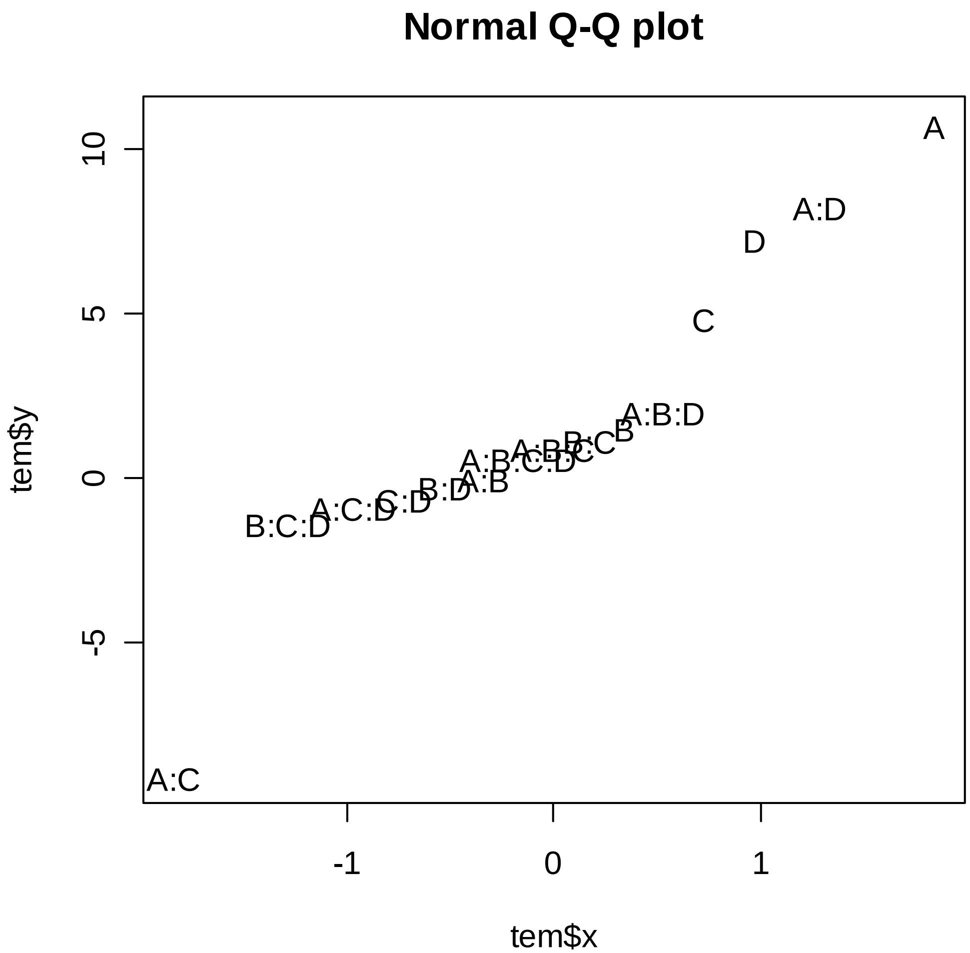
Analysis of Variance Table

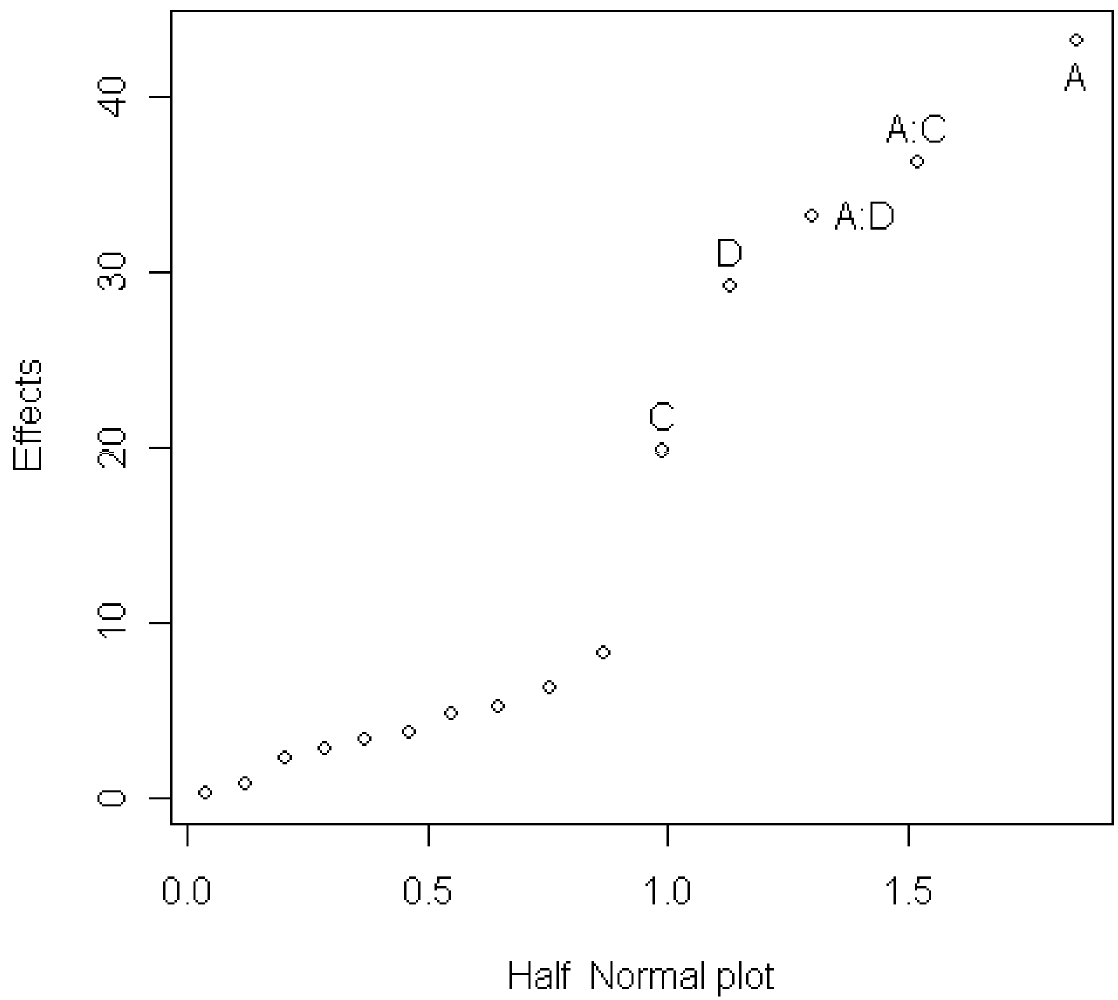
Response: FRate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	1870.56	1870.56		
B	1	39.06	39.06		
C	1	390.06	390.06		
D	1	855.56	855.56		
A:B	1	0.06	0.06		
A:C	1	1314.06	1314.06		
B:C	1	22.56	22.56		
A:D	1	1105.56	1105.56		
B:D	1	0.56	0.56		
C:D	1	5.06	5.06		
A:B:C	1	14.06	14.06		
A:B:D	1	68.06	68.06		
A:C:D	1	10.56	10.56		
B:C:D	1	27.56	27.56		
A:B:C:D	1	7.56	7.56		
Residuals	0	0.00			

What to do when there is no internal estimate of error?

- One approach to the analysis is to assume that certain high-order interactions are negligible and combine their mean squares to estimate the error.
- Occasionally real high-order interactions occur, then use the **Q-Q Normal Plot (or half-normal plot)** of the estimates of the effects. The effect that are negligible will tend to fall along a straight line whereas significant effects will not lie along the straight line.


Normal Q-Q plot
A
A:D
D
C
A:B:D
B
B:C
A:B
B:D
C:D
A:C:D
B:C:D
A:C
10
5
0
-5
-1
0
1
tem$y
tem$x

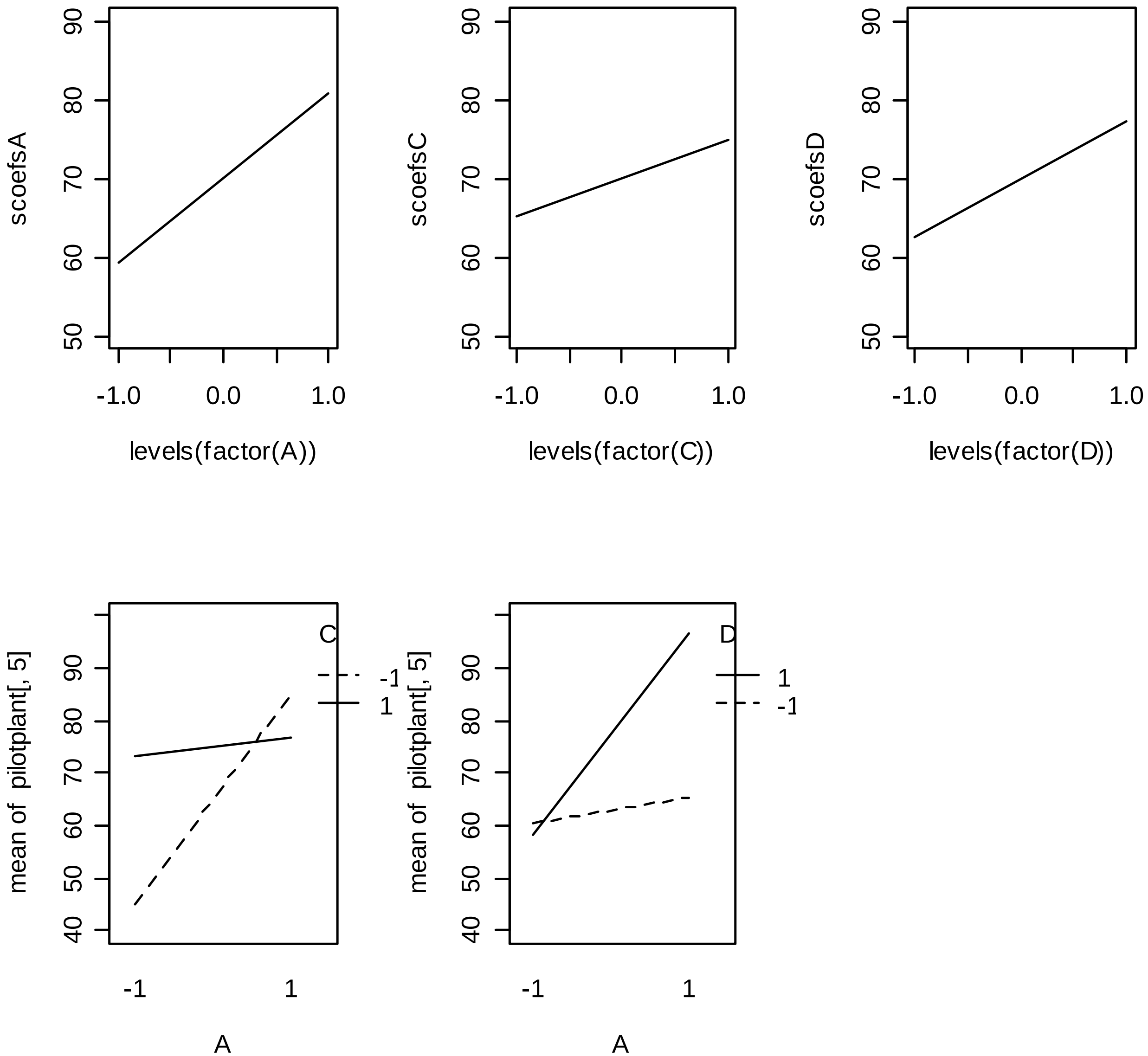


Effects
0
10
20
30
40
0.0
0.5
1.0
1.5
Half Normal plot
A
A:C
A:D
D
C

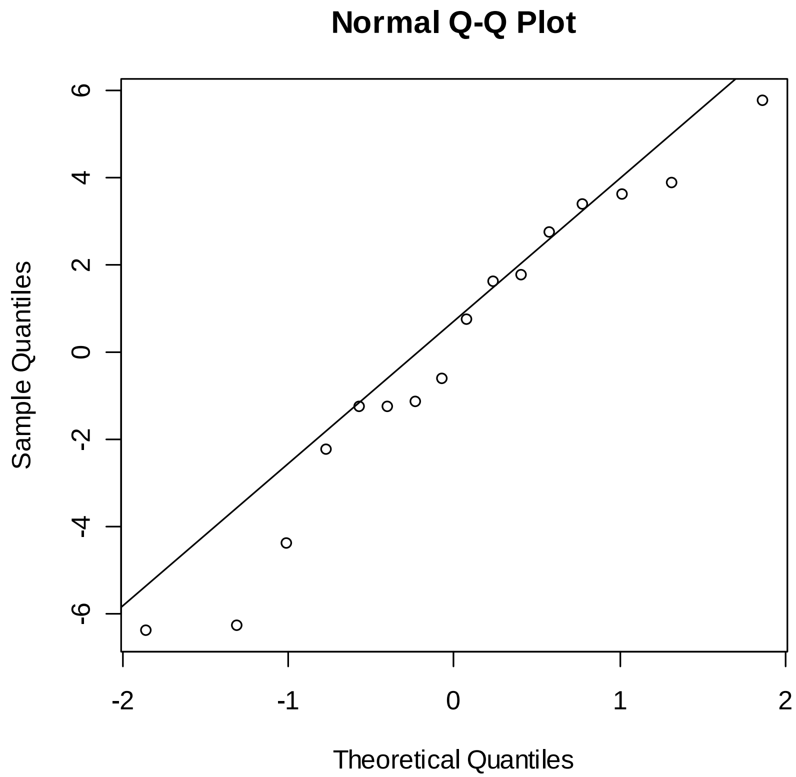
Analysis of Variance Table

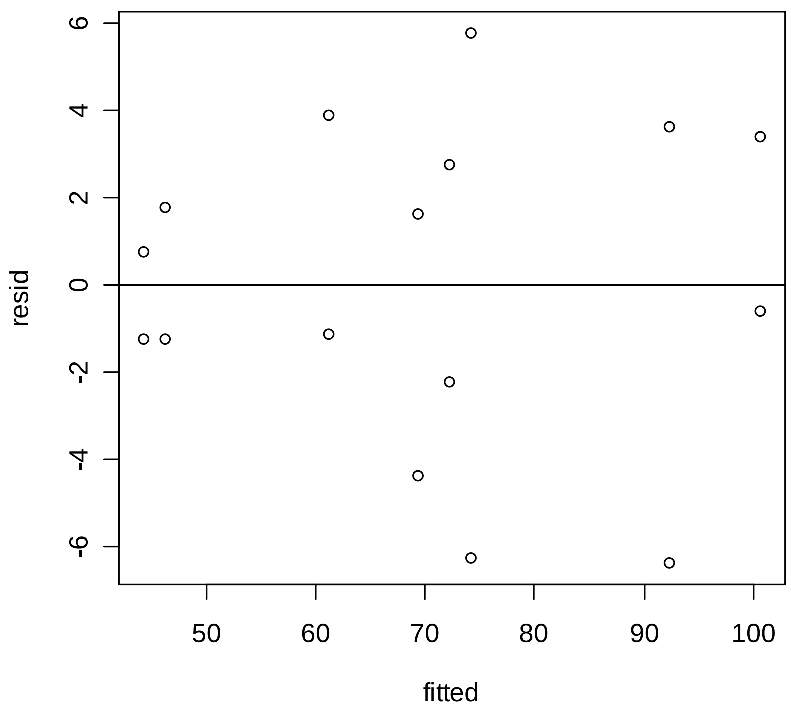
Response: FRate

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	1870.6	1870.6	83.368	1.67e-05
C	1	390.1	390.1	17.384	0.003124
D	1	855.6	855.6	38.131	0.000267
A:C	1	1314.1	1314.1	58.565	6.00e-05
A:D	1	1105.6	1105.6	49.273	0.000110
C:D	1	5.1	5.1	0.226	0.647483
A:C:D	1	10.6	10.6	0.471	0.512032
Residuals	8	179.5	22.4		


scoefsA
levels(factor(A))
scoefsC
levels(factor(C))
scoefsD
levels(factor(D))
mean of pilotplant[, 5]
A
C
-1
1
D
1
-1

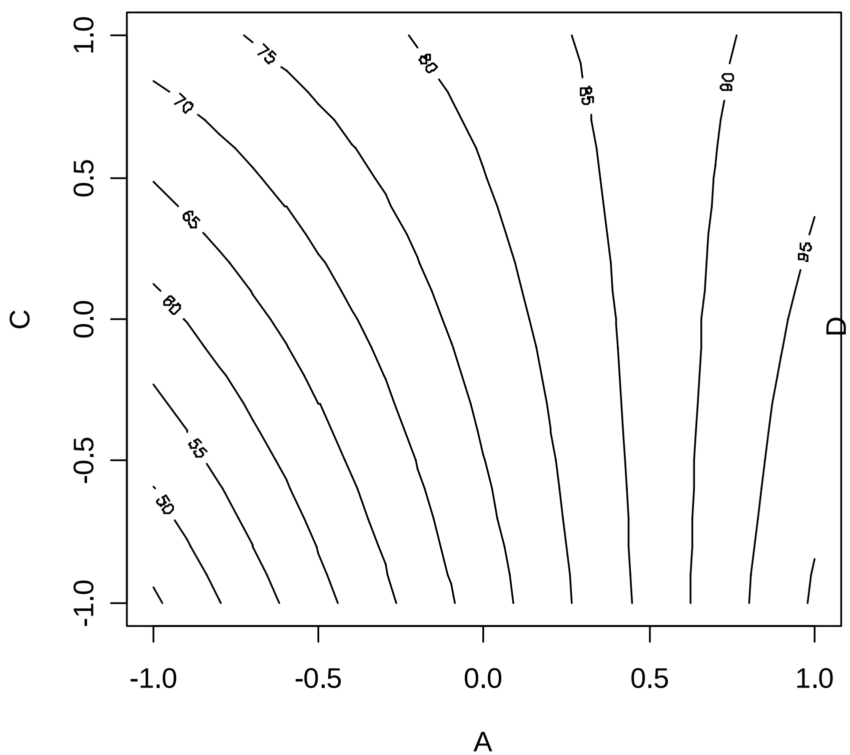
Final Model Checking



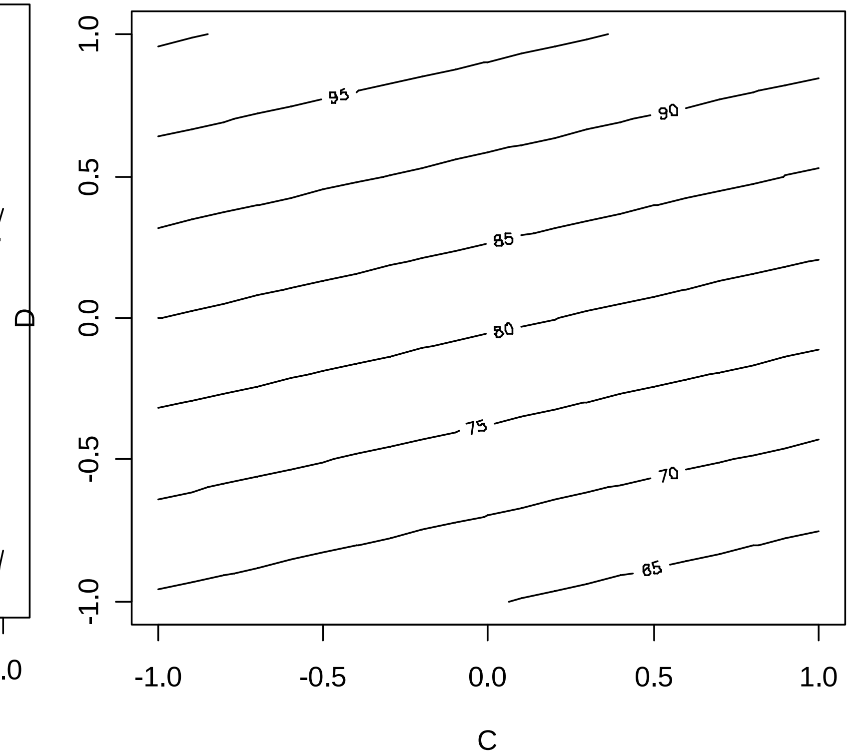


Model Interpretation – Response Surface Plots

contour plot with stirring rate D=1



contour plot with temperature A=1



With concentration at either the low or high level, high temperature and high stirring rate results in high filtration rates

The Drilling Experiment
Example 6-3, pg. 237

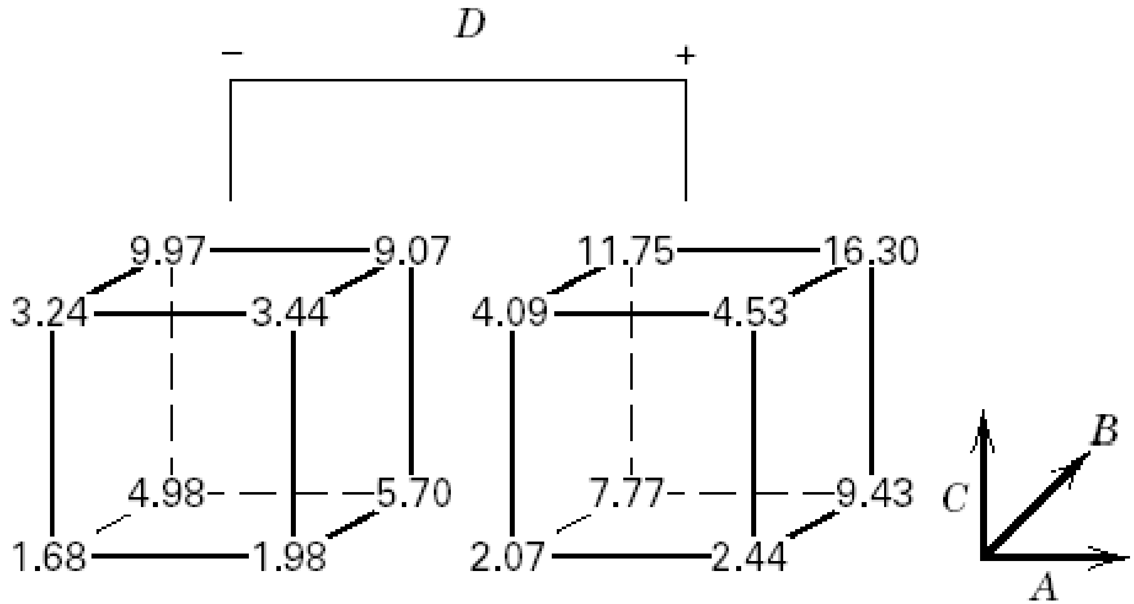
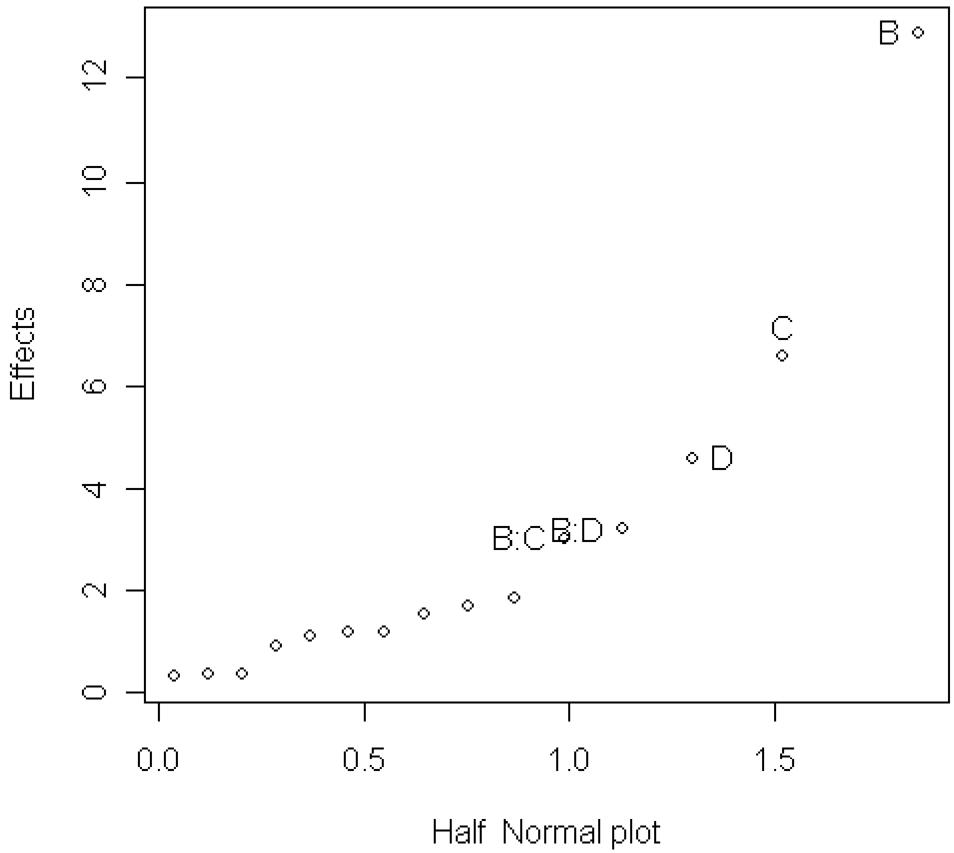


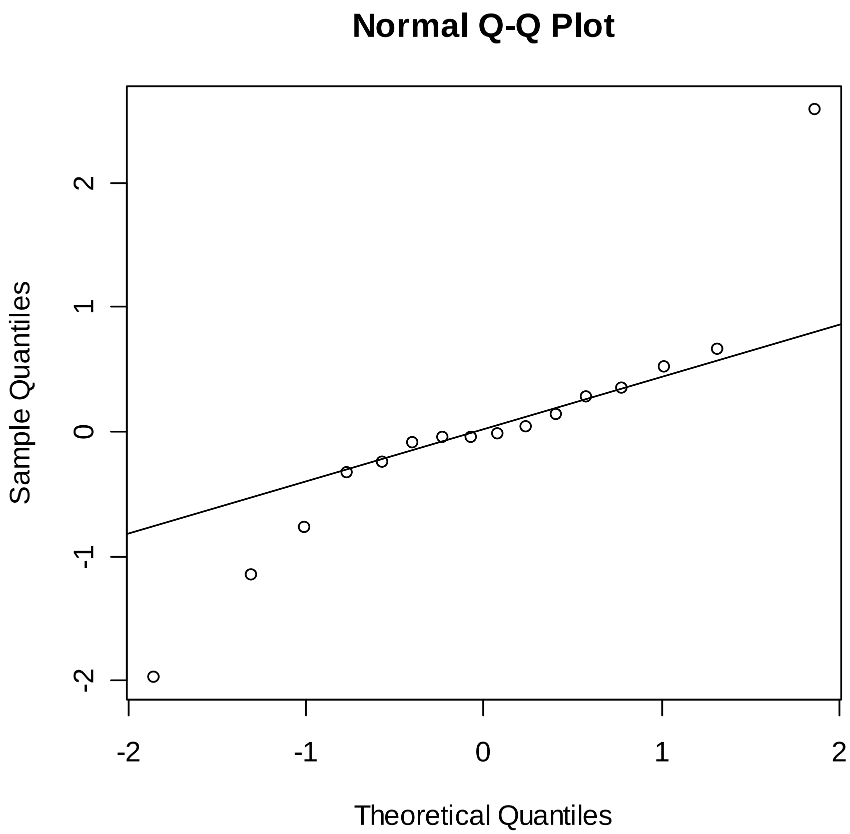
Figure 6-17 Data from the drilling experiment of Example 6-3.

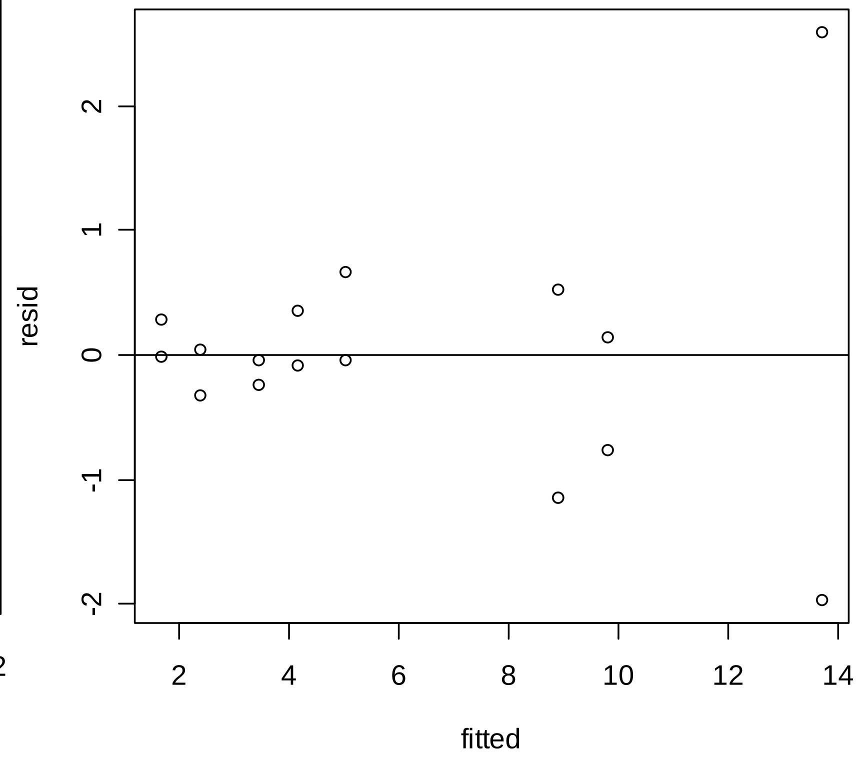
A = drill load, B = flow, C = speed, D = type of mud,
y = advance rate of the drill

Half Normal Probability Plot of Effects
The Drilling Experiment



Residual Analysis





Residual Plots

- The residual plots indicate that there are problems with the **equality of variance** assumption
- The usual approach to this problem is to employ a **transformation** on the response
- **Power family** transformations are widely used

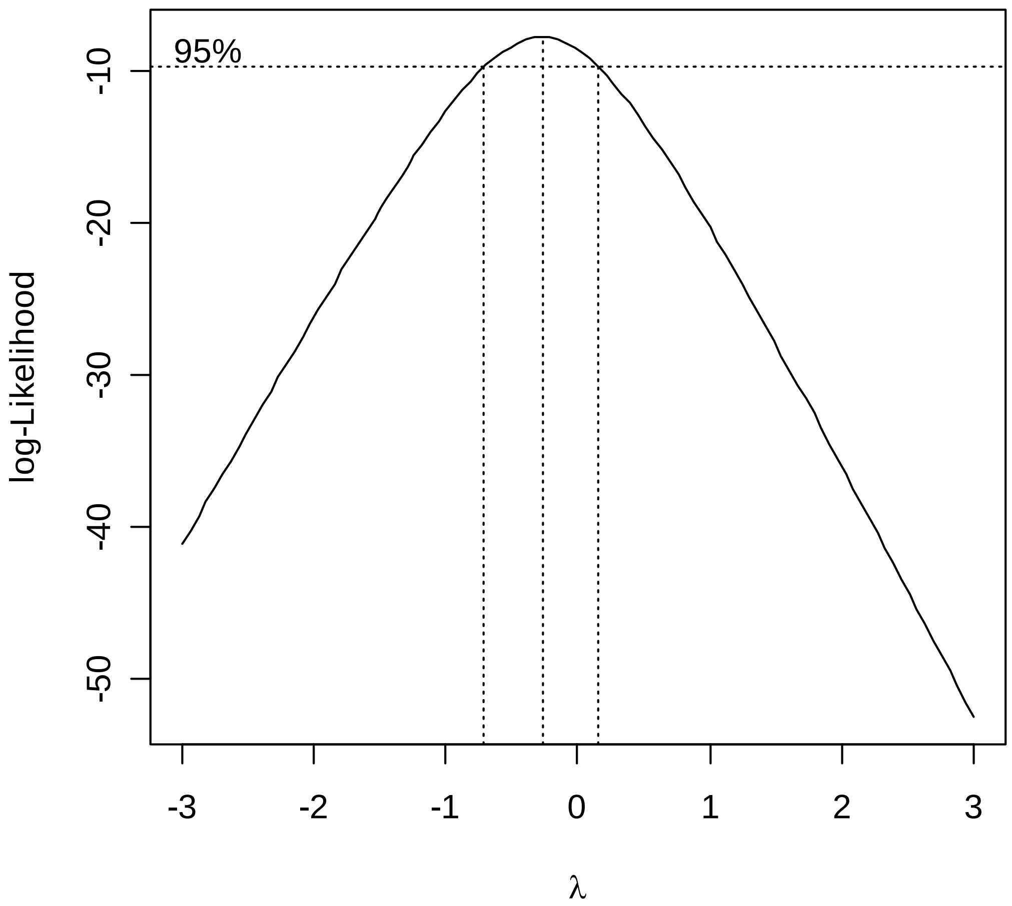
$$y^* = y^\lambda$$

- Transformations are typically performed to
  - Stabilize variance
  - Induce normality
  - Simplify the model

Selecting a Transformation

- **Empirical** selection of lambda
- Prior (theoretical) knowledge or experience can often suggest the form of a transformation
- **Analytical** selection of lambda…the Box-Cox (1964) method (simultaneously estimates the model parameters and the transformation parameter lambda)

Box-Cox Transformation

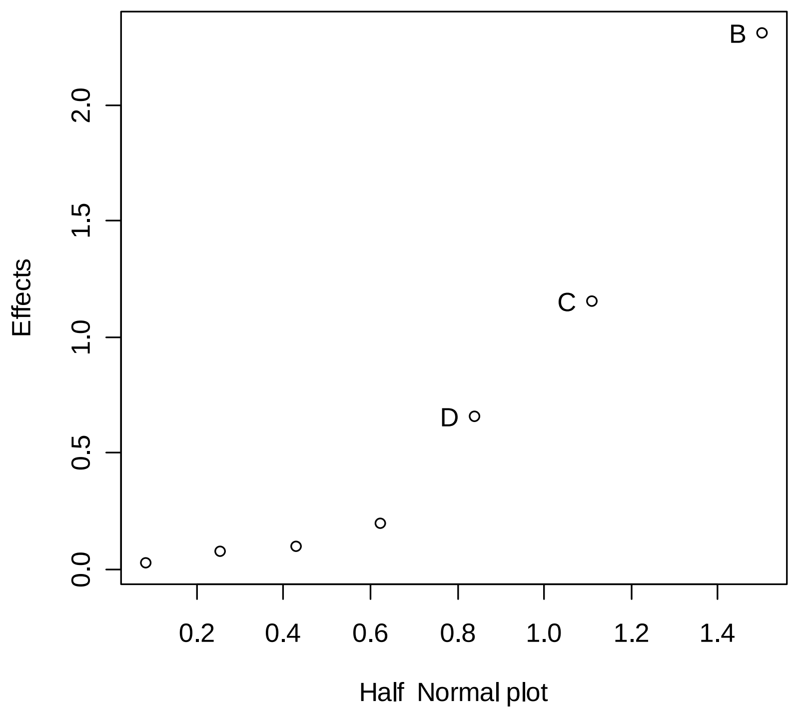


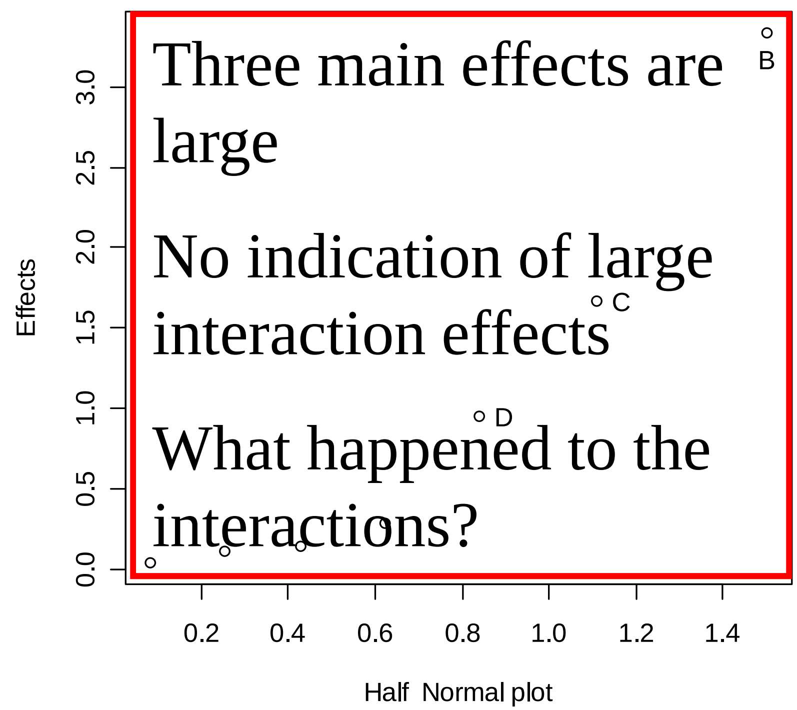
A **log** transformation is recommended

The procedure provides a **confidence interval** on the transformation parameter lambda

If unity is included in the confidence interval, no transformation would be needed

Effect Estimates Following the Log Transformation





Three main effects are large

No indication of large interaction effects

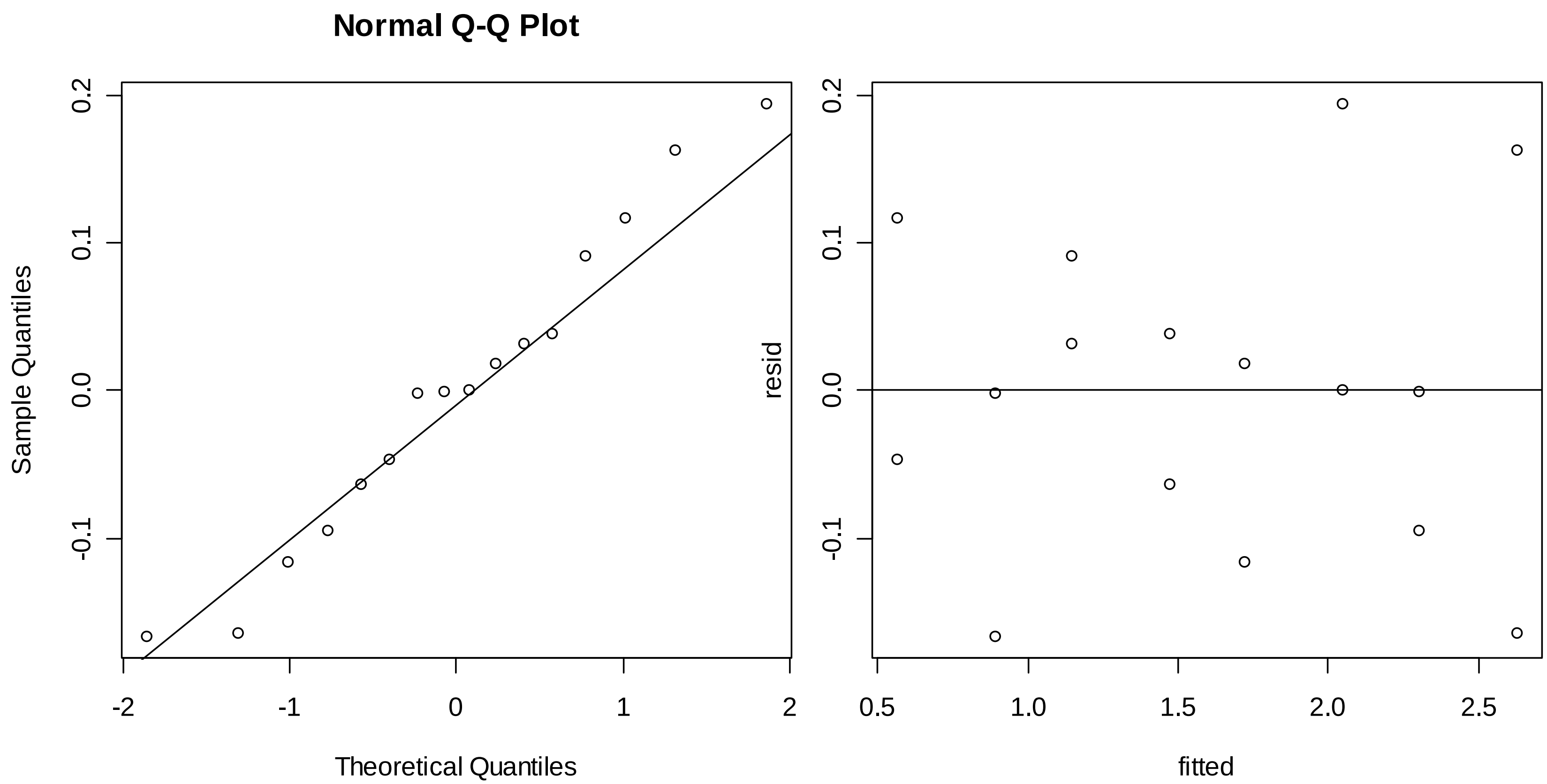
What happened to the interactions?

ANOVA Following the Log Transformation

Table 6-14 Analysis of Variance for Example 6-3 Following the Log Transformation

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
B (Flow)	5.345	1	5.345	381.79	<0.0001
C (Speed)	1.339	1	1.339	95.64	<0.0001
D (Mud)	0.431	1	0.431	30.79	<0.0001
Error	0.173	12	0.014		
Total	7.288	15			

Following the Log Transformation



The Log Advance Rate Model

- Is the log model “better”?
- We would generally prefer a **simpler model** in a transformed scale to a more complicated model in the original metric
- What happened to the interactions?
- Sometimes transformations provide insight into the underlying **mechanism**