CS 147:
Computer Systems Performance Analysis
Replicated Binary Designs

CS 147:
Computer Systems Performance Analysis
Replicated Binary Designs

Overview

$2^k r$ Designs

- $2^2 r$ Designs
- Effects
- Analysis of Variance
- Confidence Intervals
- Predictions
- Verification

Multiplicative Models

- Example

General $2^k r$ Designs

Overview
$2^k r$ Designs
$2^2 r$ Designs
Effects
Analysis of Variance
Confidence Intervals
Predictions
Verification
Multiplicative Models
Example
General $2^k r$ Designs

2^k Factorial Designs With Replications

- 2^k factorial designs do not allow for estimation of experimental error
  - No experiment is ever repeated
- Error is usually present
  - And usually important
- Handle issue by replicating experiments
- But which to replicate, and how often?

2^k Factorial Designs With Replications

- 2^k factorial designs do not allow for estimation of experimental error
  - No experiment is ever repeated
- Error is usually present
  - And usually important
- Handle issue by replicating experiments
- But which to replicate, and how often?

$2^k r$ Factorial Designs

- Replicate each experiment r times
- Allows quantifying experimental error
- Again, easiest to first look at case of only 2 factors

└ $2^k r$ Factorial Designs

$2^k r$ Factorial Designs

- Replicate each experiment r times
- Allows quantifying experimental error
- Again, easiest to first look at case of only 2 factors

$2^2 r$ Factorial Designs

- 2 factors, 2 levels each, with r replications at each of the four combinations
- $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$
- Now we need to compute effects, estimate errors, and allocate variation
- Can also produce confidence intervals for effects and predicted responses

$2^2 r$ Factorial Designs

- 2 factors, 2 levels each, with r replications at each of the four combinations
- $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$
- Now we need to compute effects, estimate errors, and allocate variation
- Can also produce confidence intervals for effects and predicted responses

Computing Effects for $2^2 r$ Factorial Experiments

- We can use sign table, as before
- But instead of single observations, regress off mean of the r observations
- Compute errors for each replication using similar tabular method
  - Sum of errors must be zero
  - $e_{ij} = y_{ij} - \hat{y}_i$
- Similar methods used for allocation of variance and calculating confidence intervals

Computing Effects for $2^2 r$ Factorial Experiments
- We can use sign table, as before
- But instead of single observations, regress off mean of the r observations
- Compute errors for each replication using similar tabular method
  - Sum of errors must be zero
  - $e_{ij} = y_{ij} - \hat{y}_i$
- Similar methods used for allocation of variance and calculating confidence intervals

The tabular method for errors is as follows: after computing the effects, multiply the effects by the sign table to get the estimated response. Enter that into the table and then subtractfrom each measured response to get errors.

Example of $2^2 r$ Factorial Design With Replications

- Same parallel system as before, but with 4 replications at each point ($r = 4$)
- No DLM, 8 nodes: 820, 822, 813, 809
- DLM, 8 nodes: 776, 798, 750, 755
- No DLM, 64 nodes: 217, 228, 215, 221
- DLM, 64 nodes: 197, 180, 220, 185

Example of $2^2 r$ Factorial Design With Replications

- Same parallel system as before, but with 4 replications at each point ($r = 4$)
- No DLM, 8 nodes: 820, 822, 813, 809
- DLM, 8 nodes: 776, 798, 750, 755
- No DLM, 64 nodes: 217, 228, 215, 221
- DLM, 64 nodes: 197, 180, 220, 185

$2^2 r$ Factorial Example Analysis Matrix

I	A	B	AB	y	Mean
1	-1	-1	1	(820,822,813,809)	816.00
1	1	-1	-1	(217,228,215,221)	220.25
1	-1	1	-1	(776,798,750,755)	769.75
1	1	1	1	(197,180,220,185)	195.50
2001.5	-1170.0	-71.00	21.5		Total
500.4	-292.5	-17.75	5.4		Total/4

$$q_0 = 500.40 \qquad q_A = -292.5$$
$$q_B = -17.75 \qquad q_{AB} = 5.4$$

2015-06-15

CS147
└ $2^k r$ Designs
　└ Effects
　　└ $2^2 r$ Factorial Example Analysis Matrix

$2^2 r$ Factorial Example Analysis Matrix

I	A	B	AB	y	Mean
1	-1	-1	1	(820,822,813,809)	816.00
1	1	-1	-1	(217,228,215,221)	220.25
1	-1	1	-1	(776,798,750,755)	769.75
1	1	1	1	(197,180,220,185)	195.50
2001.5	-1170.0	-71.00	21.5		Total
500.4	-292.5	-17.75	5.4		Total/4

$q_0 = 500.40$ $q_A = -292.5$
$q_B = -17.75$ $q_{AB} = 5.4$

Estimation of Errors for $2^2 r$ Factorial Example

- Figure differences between predicted and observed values for each replication:

$$
\begin{aligned}
e_{ij} &= y_{ij} - \hat{y}_i \\
&= y_{ij} - q_0 - q_A x_{Ai} - q_B x_{Bi} - q_{AB} x_{Ai} x_{Bi}
\end{aligned}
$$

- Now calculate SSE:

$$
\text{SSE} = \sum_{i=1}^{2^2} \sum_{j=1}^{r} e_{ij}^2 = 2606
$$

2015-06-15

CS147
└ $2^k r$ Designs
  └ Effects
    └ Estimation of Errors for $2^2 r$ Factorial Example

Estimation of Errors for $2^2 r$ Factorial Example

- Figure differences between predicted and observed values for each replication:

$$
\begin{aligned}
e_{ij} &= y_{ij} - \hat{y}_i \\
&= y_{ij} - q_0 - q_A x_{Ai} - q_B x_{Bi} - q_{AB} x_{Ai} x_{Bi}
\end{aligned}
$$

- Now calculate SSE:

$$
\text{SSE} = \sum_{i=1}^{2^2} \sum_{j=1}^{r} e_{ij}^2 = 2606
$$

Allocating Variation

- We can determine percentage of variation due to each factor's impact
  - Just like 2^k designs without replication
- But we can also isolate variation due to experimental errors
- Methods are similar to other regression techniques for allocating variation

Allocating Variation

- We can determine percentage of variation due to each factor's impact
  - Just like 2^k designs without replication
- But we can also isolate variation due to experimental errors
- Methods are similar to other regression techniques for allocating variation

Variation Allocation in Example

- We've already figured SSE
- We also need SST, SSA, SSB, and SSAB

$$\text{SST} = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$$

- Also, SST = SSA + SSB + SSAB + SSE
- Use same formulae as before for SSA, SSB, and SSAB

Variation Allocation in Example

- We've already figured SSE
- We also need SST, SSA, SSB, and SSAB

$$\text{SST} = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$$

- Also, SST = SSA + SSB + SSAB + SSE
- Use same formulae as before for SSA, SSB, and SSAB

Sums of Squares for Example

- $SST = SSY - SS0 = 1{,}377{,}009.75$
- SSA = 1,368,900
- SSB = 5041
- SSAB = 462.25
- Percentage of variation for A is 99.4%
- Percentage of variation for B is 0.4%
- Percentage of variation for A/B interaction is 0.03%
- And 0.2% (approx.) is due to experimental errors

Sums of Squares for Example

- $SST = SSY - SS0 = 1{,}377{,}009.75$
- SSA = 1,368,900
- SSB = 5041
- SSAB = 462.25
- Percentage of variation for A is 99.4%
- Percentage of variation for B is 0.4%
- Percentage of variation for A/B interaction is 0.03%
- And 0.2% (approx.) is due to experimental errors

Confidence Intervals for Effects

- Computed effects are random variables
- Thus would like to specify how confident we are that they are correct
- Usual confidence-interval methods
- First, must figure Mean Square of Errors

$$s_e^2 = \frac{SSE}{2^2(r-1)}$$

- $r-1$ is because errors add up to zero
  - $\Rightarrow$ Only $r-1$ can be chosen independently

Confidence Intervals for Effects
- Computed effects are random variables
- Thus would like to specify how confident we are that they are correct
- Usual confidence-interval methods
- First, must figure Mean Square of Errors
$$s_e^2 = \frac{SSE}{2^2(r-1)}$$
- $r-1$ is because errors add up to zero
  - $\Rightarrow$ Only $r-1$ can be chosen independently

Calculating Variances of Effects

- Variance (due to errors) of all effects is the same:

$$s_{q_0}^2 = s_{q_A}^2 = s_{q_B}^2 = s_{q_AB}^2 = \frac{s_e^2}{2^2 r}$$

- So standard deviation is also the same
- In calculations, use t- or z-value for $2^2(r-1)$ degrees of freedom

Calculating Variances of Effects

- Variance (due to errors) of all effects is the same:

$$s_{q_0}^2 = s_{q_A}^2 = s_{q_B}^2 = s_{q_AB}^2 = \frac{s_e^2}{2^2 r}$$

- So standard deviation is also the same
- In calculations, use t- or z-value for $2^2(r-1)$ degrees of freedom

Calculating Confidence Intervals for Example

- At 90% level, using t-value for 12 degrees of freedom, 1.782
- Standard deviation of effects is 3.68
- Confidence intervals are $q_i \mp (1.782)(3.68)$
- q_0 is (493.8,506.9)
- q_A is (-299.1,-285.9)
- q_B is (-24.3,-11.2)
- q_{AB} is (-1.2,11.9)

Calculating Confidence Intervals for Example

- At 90% level, using t-value for 12 degrees of freedom, 1.782
- Standard deviation of effects is 3.68
- Confidence intervals are $q_i \mp (1.782)(3.68)$
- q_0 is (493.8,506.9)
- q_A is (-299.1,-285.9)
- q_B is (-24.3,-11.2)
- q_{AB} is (-1.2,11.9)

Predicted Responses

- We already have predicted all the means we can predict from this kind of model
- We measured four, we can "predict" four
- However, we *can* predict how close we would get to true sample mean if we ran m more experiments

Predicted Responses

- We already have predicted all the means we can predict from this kind of model
- We measured four, we can "predict" four
- However, we *can* predict how close we would get to true sample mean if we ran m more experiments

Formula for Predicted Means

- For m future experiments, predicted mean is

$$\hat{y} \mp t_{[1-\alpha/2;2^2(r-1)]} s_{\hat{y}_m}$$

Where

$$s_{y_{\hat{y}_m}} = s_e \left(\frac{1}{n_{\text{eff}}} + \frac{1}{m} \right)^{1/2}$$

$$n_{\text{eff}} = \frac{\text{Total number of runs}}{1 + \text{sum of DFs of parameters used in } \hat{y}}$$

2015-06-15

Formula for Predicted Means

- For m future experiments, predicted mean is

$$\hat{y} \mp t_{[1-\alpha/2;2^2(r-1)]} s_{\hat{y}_m}$$

Where

$$s_{y_{\hat{y}_m}} = s_e \left(\frac{1}{n_{\text{eff}}} + \frac{1}{m} \right)^{1/2}$$

$$n_{\text{eff}} = \frac{\text{Total number of runs}}{1 + \text{sum of DFs of parameters used in } \hat{y}}$$

Example of Predicted Means

- What would we predict as confidence interval of response for no dynamic load management at 8 nodes for 7 more tests?

$$s_{\hat{y}_7} = 3.68 \left(\frac{1}{16/5} + \frac{1}{7} \right)^{1/2} = 2.49$$

- 90% confidence interval is (811.6,820.4)
- We're 90% confident that mean would be in this range

Example of Predicted Means

- What would we predict as confidence interval of response for no dynamic load management at 8 nodes for 7 more tests?

$$s_{\hat{y}_7} = 3.68 \left(\frac{1}{16/5} + \frac{1}{7} \right)^{1/2} = 2.49$$

- 90% confidence interval is (811.6,820.4)
- We're 90% confident that mean would be in this range

Visual Tests for Verifying Assumptions

- What assumptions have we been making?
  - Model errors are statistically independent
  - Model errors are additive
  - Errors are normally distributed
  - Errors have constant standard deviation
  - Effects of errors are additive
- All boils down to independent, normally distributed observations with constant variance

Visual Tests for Verifying Assumptions

- What assumptions have we been making?
  - Model errors are statistically independent
  - Model errors are additive
  - Errors are normally distributed
  - Errors have constant standard deviation
  - Effects of errors are additive
- All boils down to independent, normally distributed observations with constant variance

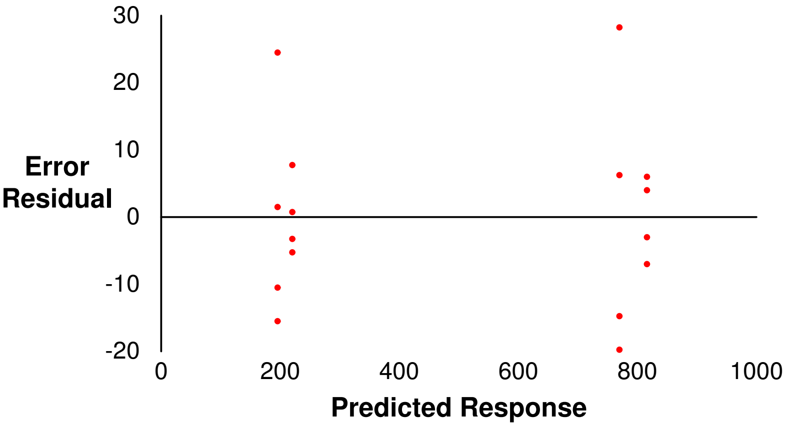
Testing for Independent Errors

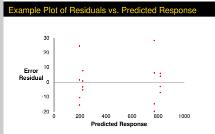
- Compute residuals and make scatter plot
- Trends indicate dependence of errors on factor levels
  - But if residuals order of magnitude below predicted response, trends can be ignored
- Usually good idea to plot residuals vs. experiment number

Testing for Independent Errors

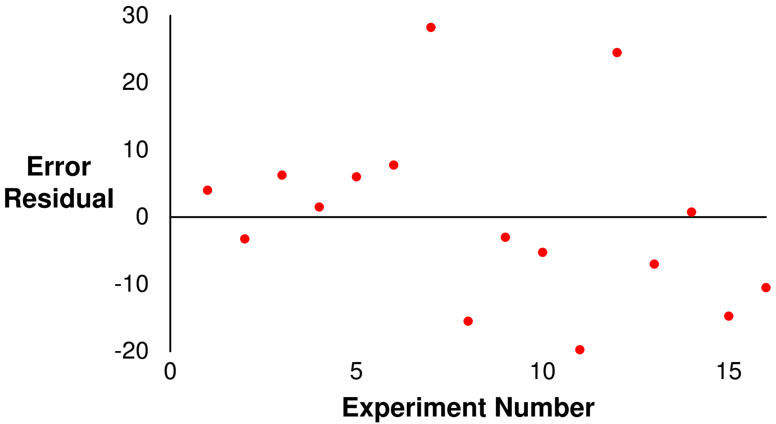
- Compute residuals and make scatter plot
- Trends indicate dependence of errors on factor levels
  - But if residuals order of magnitude below predicted response, trends can be ignored
- Usually good idea to plot residuals vs. experiment number

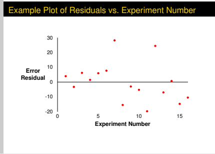
Example Plot of Residuals vs. Predicted Response





Example Plot of Residuals vs. Experiment Number





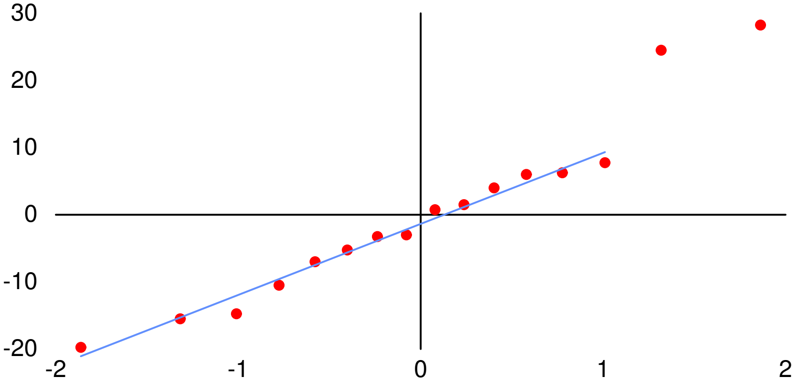
Testing for Normally Distributed Errors

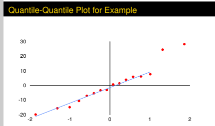
- As usual, do quantile-quantile chart against normal distribution
- If close to linear, normality assumption is good

Testing for Normally Distributed Errors

- As usual, do quantile-quantile chart against normal distribution
- If close to linear, normality assumption is good

Quantile-Quantile Plot for Example





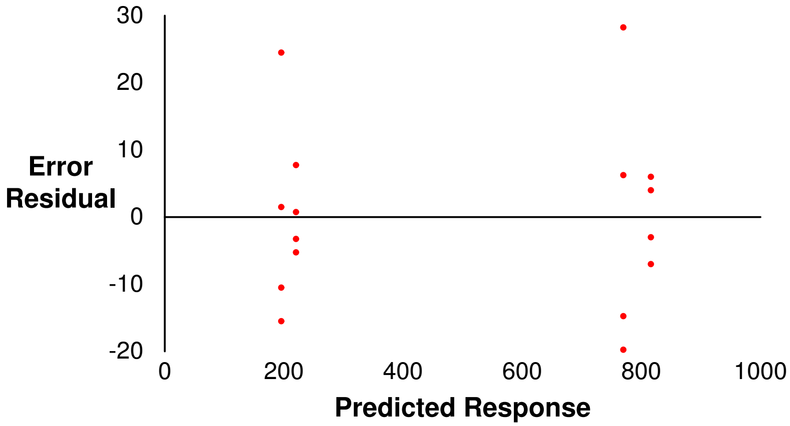
Assumption of Constant Variance

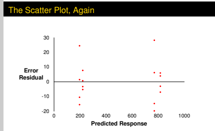
- Checking homoscedasticity
- Go back to scatter plot of residuals vs. prediction and check for even spread

Assumption of Constant Variance

- Checking homoscedasticity
- Go back to scatter plot of residuals vs. prediction and check for even spread

The Scatter Plot, Again





Multiplicative Models for 2^2r Experiments

- Assumptions of additive models
- Example of a multiplicative situation
- Handling a multiplicative model
- When to choose multiplicative model
- Multiplicative example

Multiplicative Models for 2^2r Experiments

- Assumptions of additive models
- Example of a multiplicative situation
- Handling a multiplicative model
- When to choose multiplicative model
- Multiplicative example

Assumptions of Additive Models

- Previous analysis used additive model:
  - $y_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e_{ij}$
- Assumes all effects are additive:
  - Factors
  - Interactions
  - Errors
- This assumption *must* be validated!

Assumptions of Additive Models
- Previous analysis used additive model:
  - $y_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e_{ij}$
- Assumes all effects are additive:
  - Factors
  - Interactions
  - Errors
- This assumption *must* be validated!

Example of a Multiplicative Situation

- Testing processors with different workloads
- Most common multiplicative case
- Consider 2 processors, 2 workloads
  - Use 2^2r design
- Response is time to execute w_j instructions on processor that requires v_i seconds/instruction
- Without interactions, time is $y_{ij} = v_i w_j$

Example of a Multiplicative Situation

- Testing processors with different workloads
- Most common multiplicative case
- Consider 2 processors, 2 workloads
  - Use 2^2r design
- Response is time to execute w_j instructions on processor that requires v_i seconds/instruction
- Without interactions, time is $y_{ij} = v_i w_j$

Handling a Multiplicative Model

- Take logarithm of both sides:

$$y_i j = v_i w_j$$

  so $\log y_{ij} = \log v_i + \log w_j$
- Now easy to solve using previous methods
- Resulting model is:

$$y = 10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_{AB}} 10^{e}$$

Handling a Multiplicative Model
- Take logarithm of both sides:
$$y_i j = v_i w_j$$
so $\log y_{ij} = \log v_i + \log w_j$
- Now easy to solve using previous methods
- Resulting model is:
$$y = 10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_{AB}} 10^{e}$$

Meaning of a Multiplicative Model

- Model is $10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_{AB}} 10^{e}$
- Here, $\mu_A = 10^{q_A}$ is inverse of ratio of MIPS ratings of processors; $\mu_B = 10^{q_B}$ is ratio of workload sizes
- Antilog of q_0 is geometric mean of responses:

$$\dot{y} = 10^{q_0} = \sqrt[n]{y_1 y_2 \cdots y_n}$$

where n = $2^2 r$

2015-06-15

Meaning of a Multiplicative Model

- Model is $10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_{AB}} 10^{e}$
- Here, $\mu_A = 10^{q_A}$ is inverse of ratio of MIPS ratings of processors; $\mu_B = 10^{q_B}$ is ratio of workload sizes
- Antilog of q_0 is geometric mean of responses:

$$\dot{y} = 10^{q_0} = \sqrt[n]{y_1 y_2 \cdots y_n}$$

where n = $2^2 r$

When to Choose a Multiplicative Model?

- Physical considerations (see previous slides)
- Range of y is large
  - Making arithmetic mean unreasonable
  - Calling for log transformation
- Plot of residuals shows large values and increasing spread
- Quantile-quantile plot doesn't look like normal distribution

When to Choose a Multiplicative Model?
- Physical considerations (see previous slides)
- Range of y is large
  - Making arithmetic mean unreasonable
  - Calling for log transformation
- Plot of residuals shows large values and increasing spread
- Quantile-quantile plot doesn't look like normal distribution

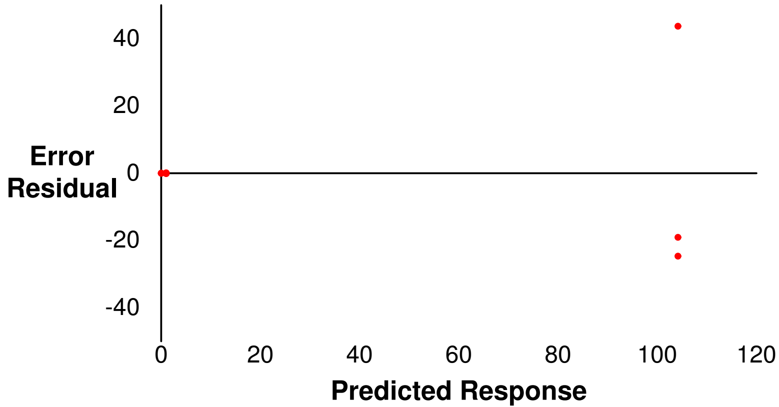
Multiplicative Example

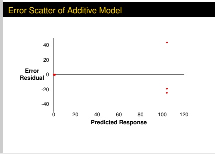
- Consider additive model of processors A_1 & A_2 running benchmarks B_1 and B_2:

y_1	y_2	y_3	Mean	I	A	B	AB
85.1	79.5	147.9	104.167	1	-1	-1	1
0.891	1.047	1.072	1.003	1	1	-1	-1
0.955	0.933	1.122	1.003	1	-1	1	-1
0.015	0.013	0.012	0.013	1	1	1	1
		Total		106.19	-104.15	-104.15	102.17
		Total/4		26.55	-26.04	-26.04	25.54

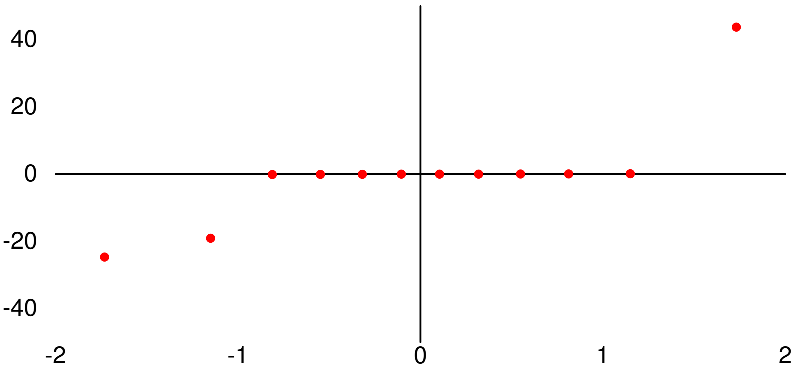
- Note large range of y values

Error Scatter of Additive Model



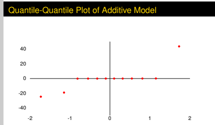


Quantile-Quantile Plot of Additive Model



2015-06-15

CS147
└Multiplicative Models
└Example
└Quantile-Quantile Plot of Additive Model

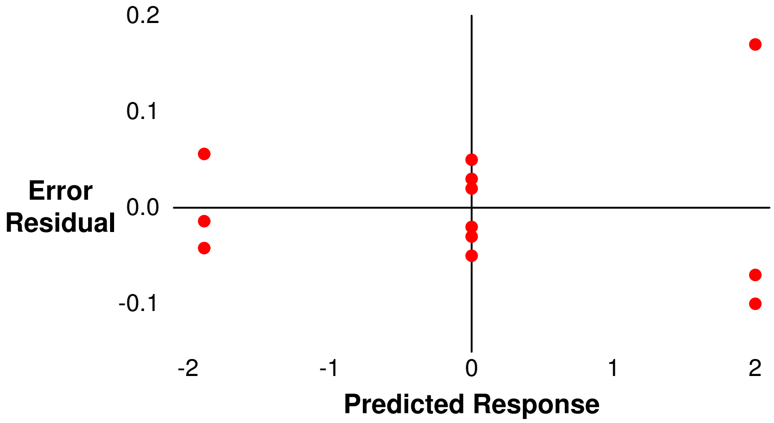


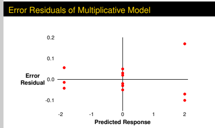
Multiplicative Model

- Taking logs of everything, the model is:

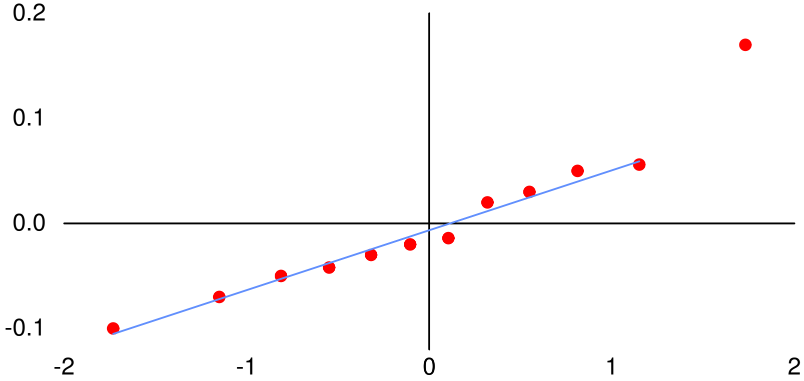
y_1	y_2	y_3	Mean	I	A	B	AB
1.93	1.9	2.17	2.000	1	-1	-1	1
-0.05	0.02	0.0302	0.000	1	-1	-1	
-0.02	-0.03	0.05	0.000	-1	1	-1	
-1.83	-1.9	-1.928	-1.886 1	1	1	1	
		Total		0.11	-3.89	-3.89	0.11
		Total/4		0.03	-0.97	-0.97	0.03

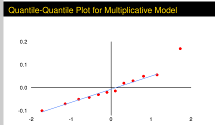
Error Residuals of Multiplicative Model





Quantile-Quantile Plot for Multiplicative Model





Summary of the Two Models

Factor	Additive Model				Multiplicative Model			
	Effect	Pct of Variation	Confidence Interval		Effect	Pct of Variation	Confidence Interval	
I	26.55		16.35	36.74	0.03		-0.02	0.07
A	-26.04	30.15	-36.23	-15.85	-0.97	49.85	-1.02	-0.93
B	-26.04	30.15	-36.23	-15.85	-0.97	49.86	-1.02	-0.93
AB	25.54	29.01	15.35	35.74	0.03	0.04	-0.02	0.07
e		10.69				0.25		

General $2^k r$ Factorial Design

- Simple extension of $2^2 r$
- See Box 18.1 in book for summary
- Always do visual tests
- Remember to consider multiplicative model as alternative

General $2^k r$ Factorial Design

- Simple extension of $2^2 r$
- See Box 18.1 in book for summary
- Always do visual tests
- Remember to consider multiplicative model as alternative

Example of $2^k r$ Factorial Design

Consider a $2^3 3$ design:

y1	y2	y3	Mean	I	A	B	C	AB	AC	BC	ABC
14	16	12	14	1	-1	-1	-1	1	1	1	-1
22	18	20	20	1	1	-1	-1	-1	-1	1	1
11	15	19	15	1	-1	1	-1	-1	1	-1	1
34	30	35	33	1	1	1	-1	1	-1	-1	-1
46	42	44	44	1	-1	-1	1	1	-1	-1	1
58	62	60	60	1	1	-1	1	-1	1	-1	-1
50	55	54	53	1	-1	1	1	-1	-1	1	-1
86	80	74	80	1	1	1	1	1	1	1	1
Total				319	67	43	155	23	19	15	-1
Total/8				39.88	8.38	5.38	19.38	2.88	2.38	1.88	-0.13

Example of $2^k r$ Factorial Design
Consider a $2^3 3$ design:

ANOVA for $2^3 3$ Design

- Percent variation explained:

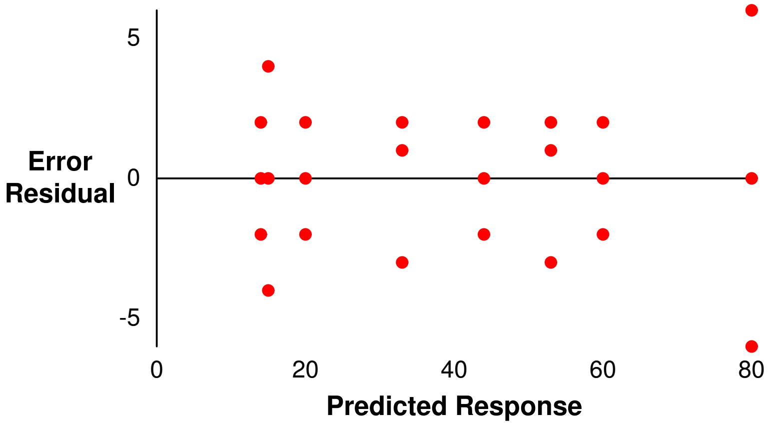
A	B	C	AB	AC	BC	ABC	Errors
14.1	5.8	75.3	1.7	1.1	0.7	0	1.37

- 90% confidence intervals

I	A	B	C	AB	AC	BC	ABC
38.7	7.2	4.2	18.2	1.7	1.2	0.7	-1.3
41.0	9.5	6.5	20.5	4.0	3.5	3.0	1.0

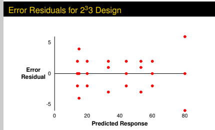
ANOVA for $2^3 3$ Design
Percent variation explained:
A B C AB AC BC ABC Errors
14.1 5.8 75.3 1.7 1.1 0.7 0 1.37
90% confidence intervals
I A B C AB AC BC ABC
38.7 7.2 4.2 18.2 1.7 1.2 0.7 -1.3
41.0 9.5 6.5 20.5 4.0 3.5 3.0 1.0

Error Residuals for $2^3 3$ Design

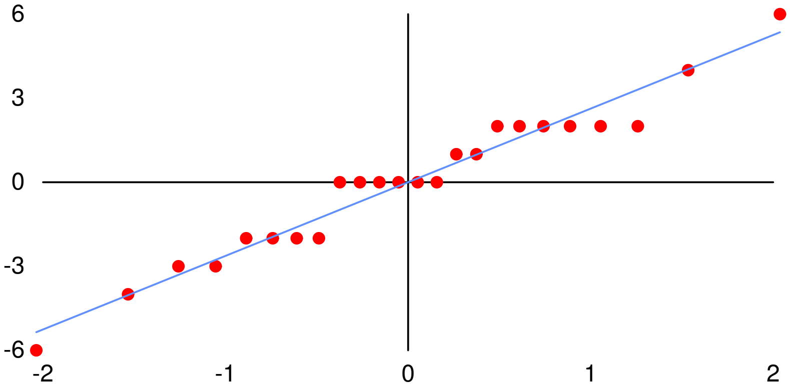


2015-06-15

CS147
└General $2^k r$ Designs
└Error Residuals for $2^3 3$ Design



Quantile-Quantile Plot for $2^3 3$ Design



2015-06-15

CS147
└General $2^k r$ Designs
└Quantile-Quantile Plot for $2^3 3$ Design

Quantile-Quantile Plot for $2^3 3$ Design