Analysis of Variance: Case Study
Bios 662

Michael G. Hudgens, Ph.D.

mhudgens@bios.unc.edu

http://www.bios.unc.edu/∼mhudgens

2008-10-28 15:15

Strategy for Analysis

1. Consider data generation mechanism
2. Analysis plan: specify model/assumptions; hyps to be tested; diagnostics to be performed
3. Summary statistics, tables, figures
4. Model fitting and diagnostics
5. Inference
6. Sensitivity analysis
7. Conclusions; summary; limitations (eg lack of power)

Hypothetical Example

- Survival times of patients following heart transplant surgery based on degree of mismatch (low, med, high) of the tissue type between donor and recipient

Analysis Plan

- ANOVA

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

$i = 1, 2, 3$ denoting low, med, high groups

$j = 1, 2, \ldots, n_i$ denoting j^{th} patient in i^{th} group

$n_1 = 14, n_2 = 13, n_3 = 12$

Y_{ij} survival time in days

μ_i mean survival time in patients with type i mismatch

Analysis Plan

- Primary hypotheses of interest: all pairwise comparisons

$$H_0 : \mu_1 = \mu_2, \qquad H_0 : \mu_1 = \mu_3, \qquad H_0 : \mu_2 = \mu_3$$

- Based on power considerations, decide $\alpha = .1$
- All pairwise comparisons using Tukey
- Further contrasts, group level means, etc will be considered hyp generating/exploratory; thus no multiplicity adjustment
- Diagnostics, remedial measures, sensitivity analyses

Data

j	Low	Medium	High
1	44	15	3.0
2	551	280	136.0
3	127	1024	65.0
4	1400	836	400.0
5	1000	51	10.4
6	700	600	39.4
7	550	250	33.4
8	300	200	48.4
9	47	22	13.5
10	26	71	34.5
11	50	62	35.5
12	10	69	45.5
13	70	13	
14	20		

Figure 1: Box plot with raw data

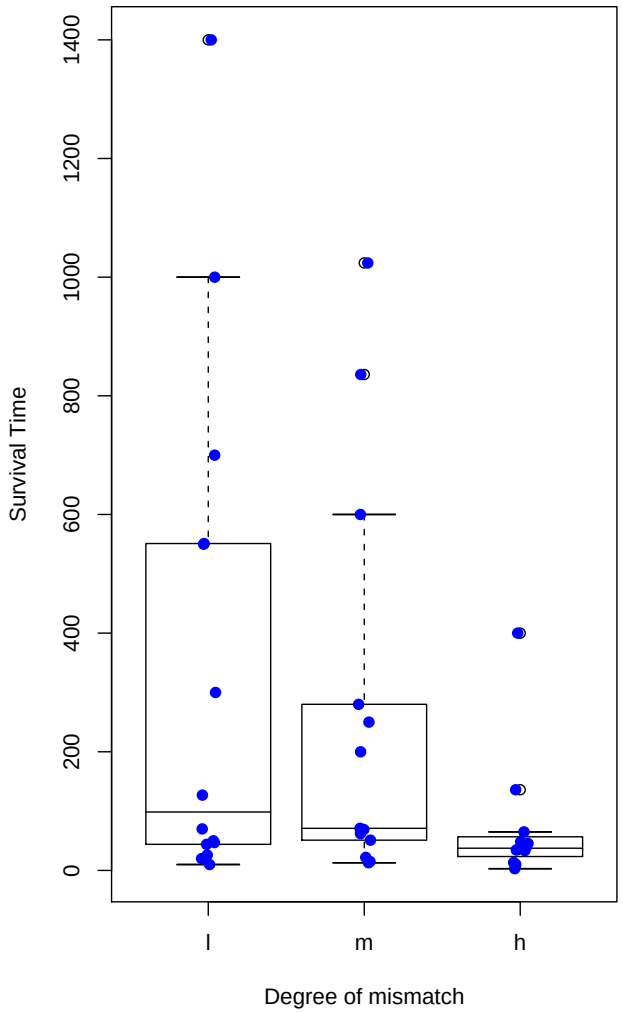
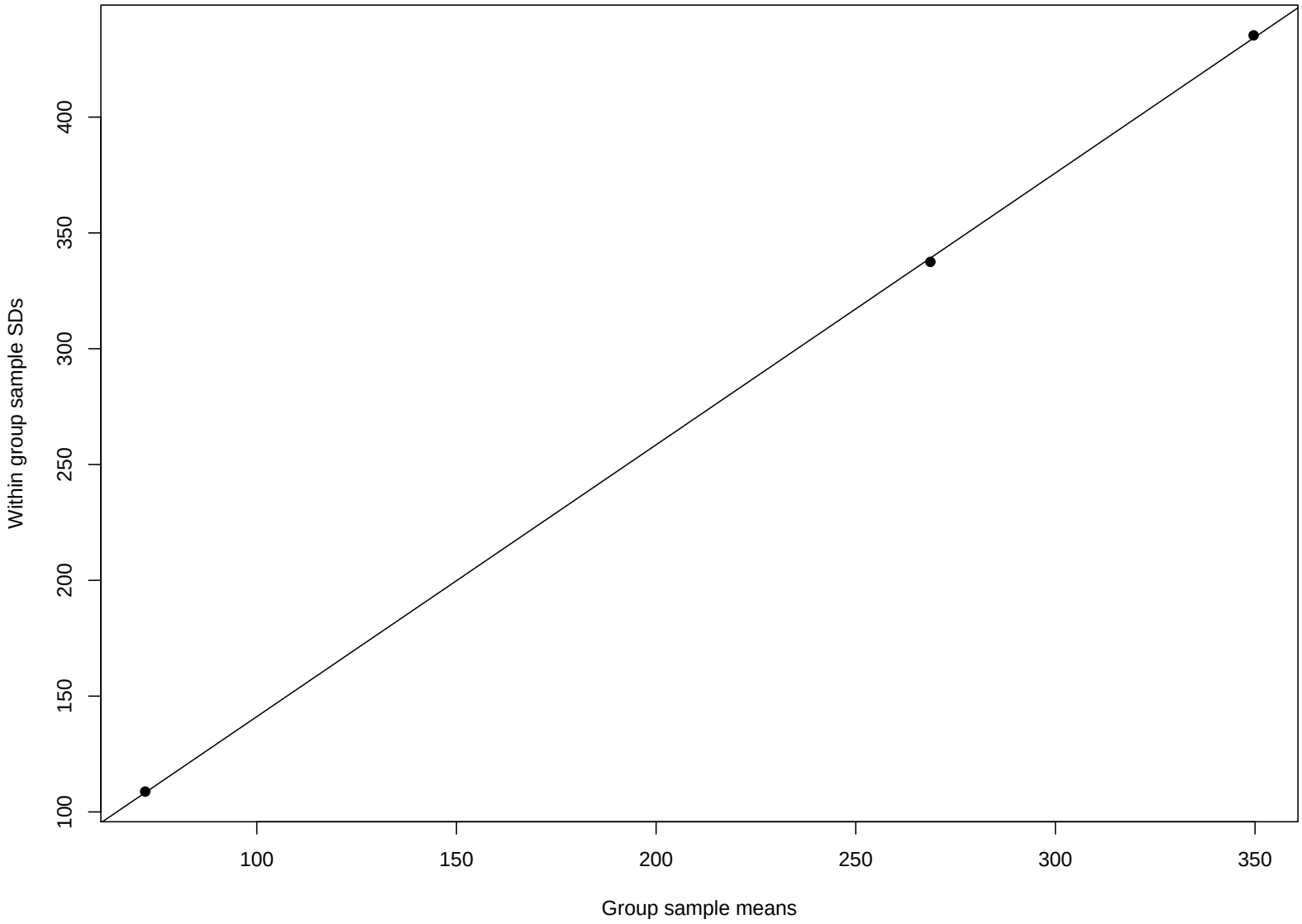


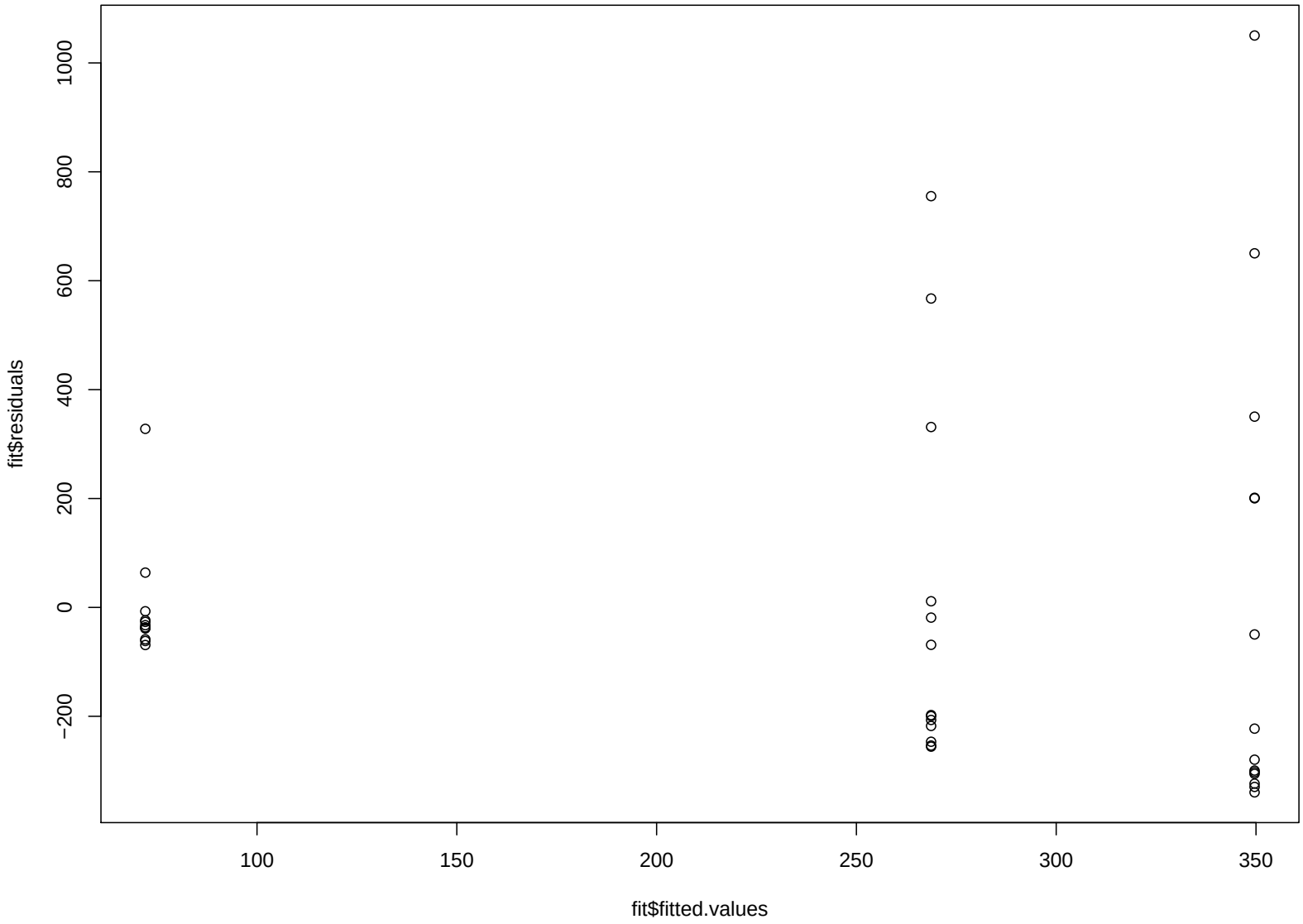
Table 1: Summary Statistics

	Low	Medium	High
n	14	13	12
Mean	349.6	268.7	72.1
Median	98.5	71.0	37.5
SD	435.31	337.51	108.82
(Min, Max)	(10,1400)	(13, 1042)	(3,400)

Homogeneity of Variance Plot



Residual Plot

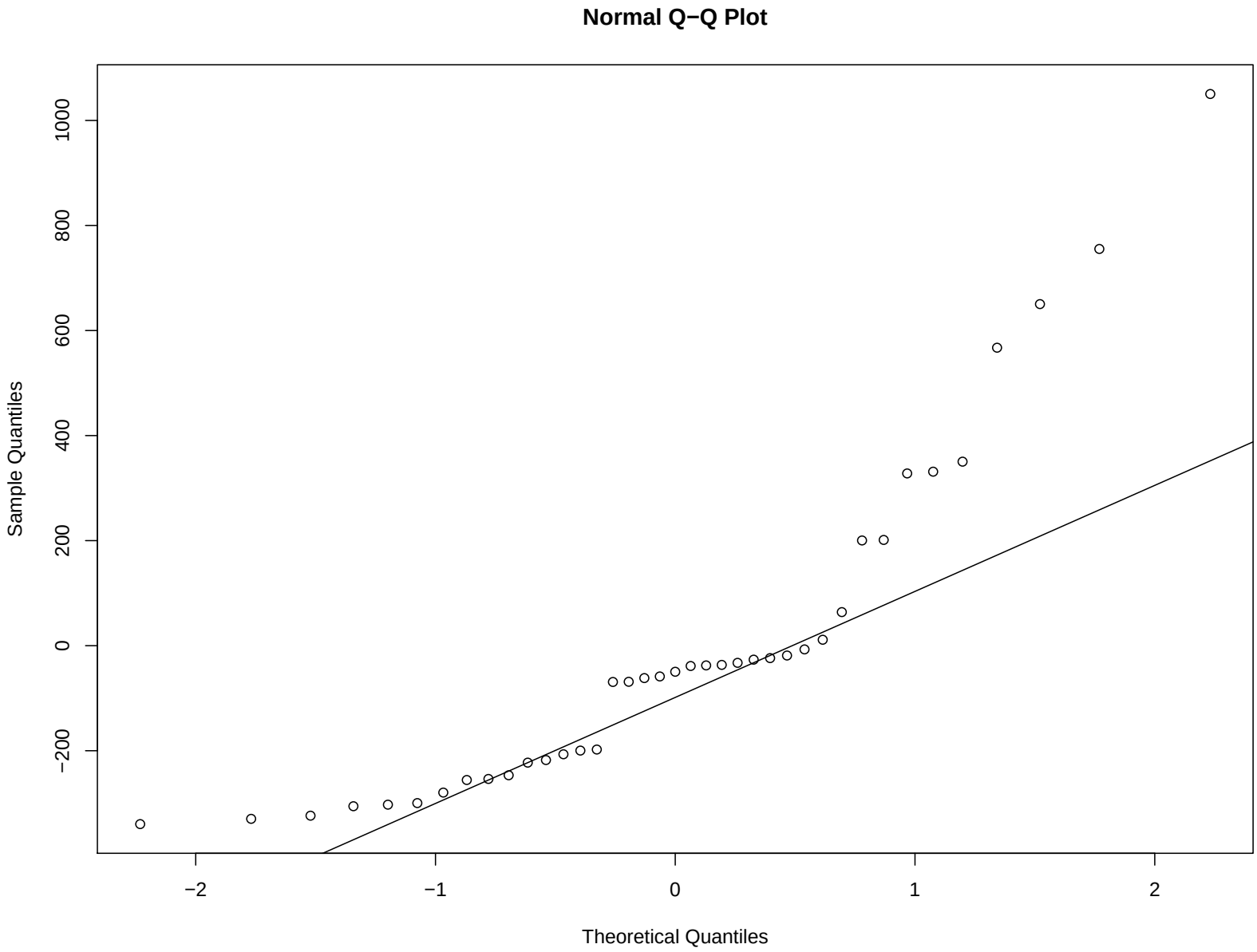


Modified Levene Test

Brown and Forsythe's Test for Homogeneity of surv Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
degree	2	452324	226162	2.46	0.0993
Error	36	3304203	91783.4		

QQ Plot


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−200
0
200
400
600
800
1000
−2
−1
0
1
2

Box-Cox Transformations

$$Y_\lambda = \begin{cases} k_1(Y^\lambda - 1) & \text{for } \lambda \neq 0 \\ k_2 \log(Y) & \text{for } \lambda = 0 \end{cases}$$

where

$$k_2 = \left(\prod_{i,j} Y_{ij} \right)^{1/N} \quad \text{and } k_1 = \frac{1}{\lambda k_2^{\lambda - 1}}$$

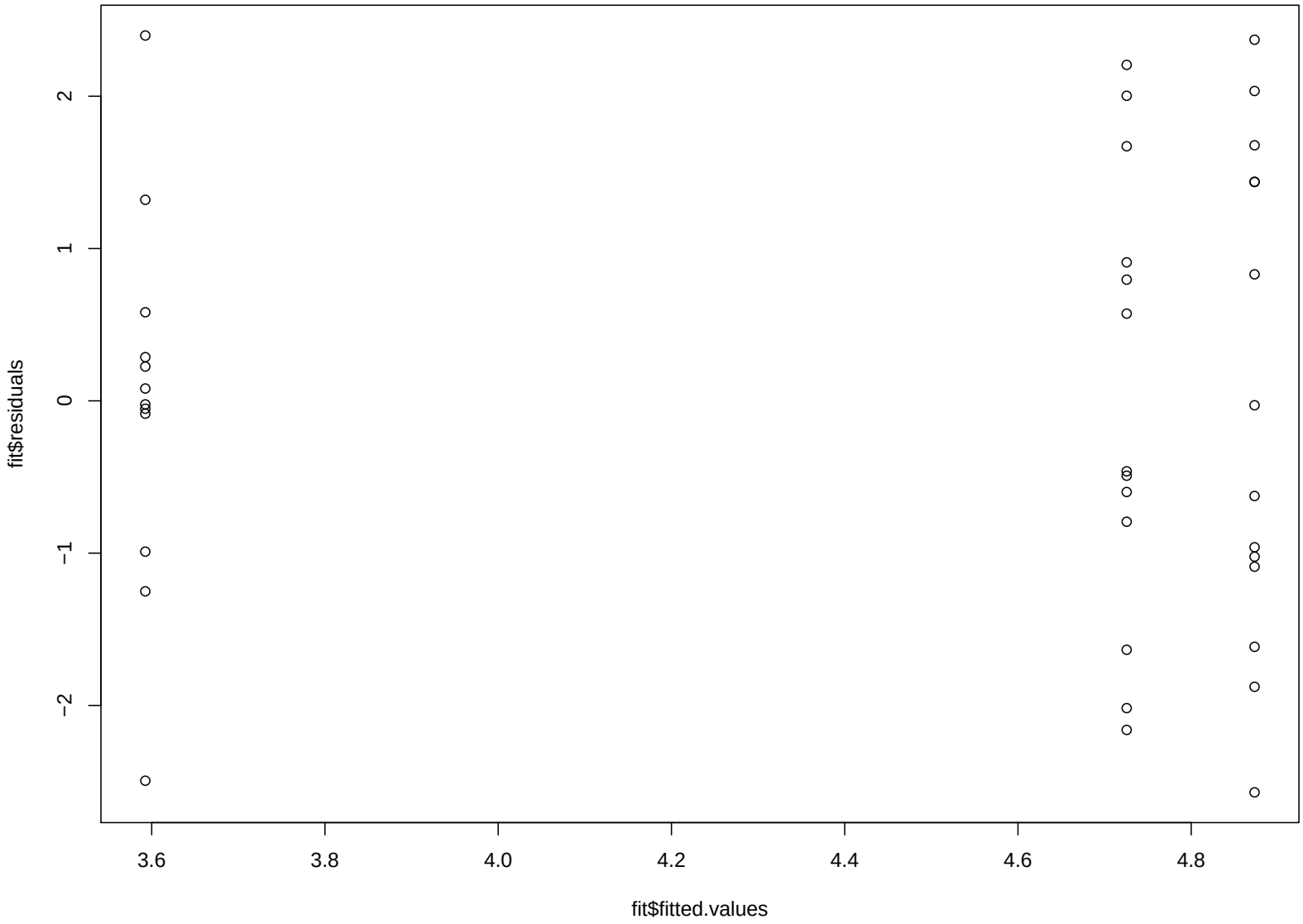
Box-Cox Transformation

Box-Cox Power (lambda)	Log Likelihood	Root mean squared error	.95 Confidence Interval
-1.0	-232.441	387.62	
-0.8	-216.424	257.07	
-0.6	-203.272	183.48	
-0.4	-193.809	143.95	
-0.2	-188.568	125.85	*
0.0	-187.530	122.54	<+
0.2	-190.213	131.27	
0.4	-195.970	152.15	
0.6	-204.212	187.96	
0.8	-214.467	244.48	
1.0	-226.364	331.69	

Box-Cox Transformation

- Choose log transformation
- Run diagnostics again

Residual Plot

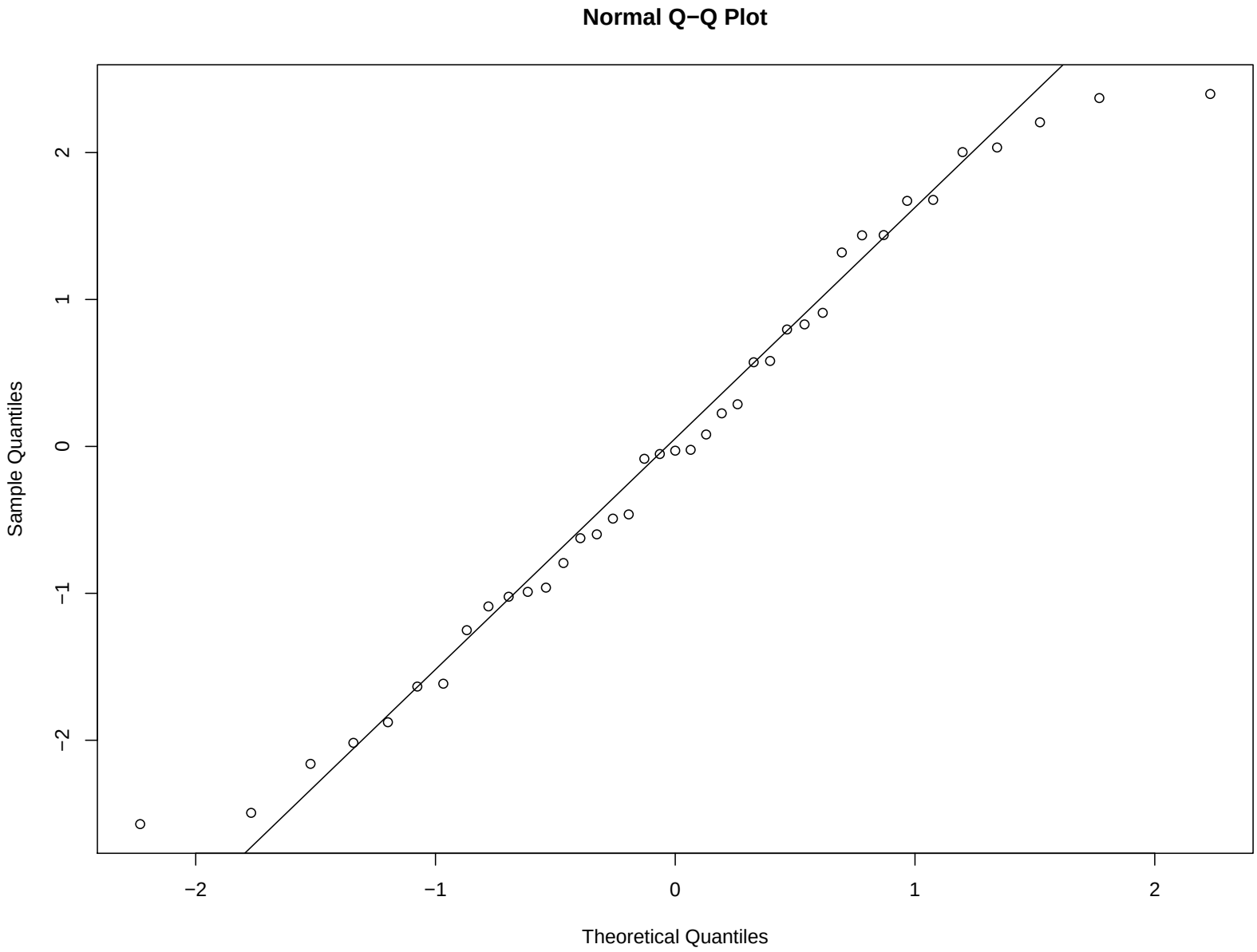


Modified Levene Test

Brown and Forsythe's Test for Homogeneity of lsurv Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
degree	2	2.2384	1.1192	1.52	0.2324
Error	36	26.5038	0.7362		

QQ Plot


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2

Tests for Normality

```
        Pearson's product-moment correlation

data:  qq$x and qq$y
sample estimates:
      cor
0.9882992

       Lilliefors (Kolmogorov-Smirnov) normality test

data:  fit$residuals
D = 0.0799, p-value = 0.766

       Shapiro-Wilk normality test

data:  fit$residuals
W = 0.9682, p-value = 0.3294
```

Box-Cox Transformation

- Diagnostics suggest adequate fit after log transformation
- Employ Tukey for all (3) pairwise comparisons of factor level means

Tukey Simultaneous 90% CIs

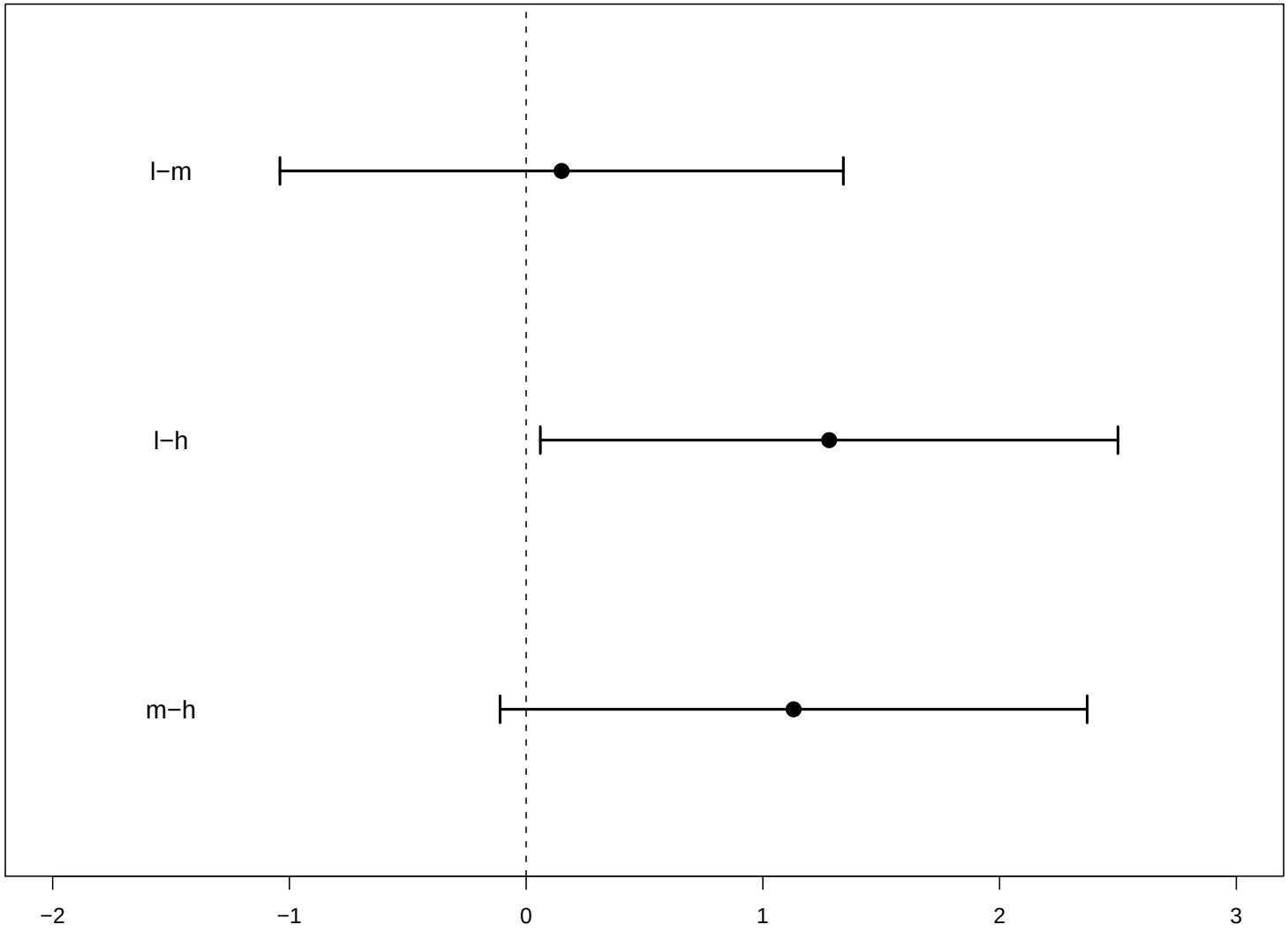
Tukey's Studentized Range (HSD) Test for lsurv

NOTE: This test controls the Type I experimentwise error rate.

Comparisons significant at the 0.1 level are indicated by ***.

degree Comparison	Difference Between Means	Simultaneous 90% Confidence Limits		
l - m	0.1476	-1.0444	1.3397	
l - h	1.2805	0.0630	2.4980	***
m - h	1.1329	-0.1061	2.3718	

Tukey Simultaneous 90% CIs



Sensitivity Analyses: Bonferroni

Bonferroni (Dunn) t Tests for lsurv

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Comparisons significant at the 0.1 level are indicated by ***.

degree Comparison	Difference Between Means	Simultaneous 90% Confidence Limits		
l - m	0.1476	-1.0969	1.3922	
l - h	1.2805	0.0094	2.5516	***
m - h	1.1329	-0.1606	2.4263	

Sensitivity Analyses: Wilcoxon Ranksum

```
        Wilcoxon rank sum test

data:  low.surv and med.surv
W = 93, p-value = 0.943
alternative hypothesis: true mu is not equal to 0


         Wilcoxon rank sum test

data:  low.surv and high.surv
W = 122, p-value = 0.05264
alternative hypothesis: true mu is not equal to 0


         Wilcoxon rank sum test

data:  med.surv and high.surv
W = 115, p-value = 0.04571
alternative hypothesis: true mu is not equal to 0
```

Conclusions

- Conclude there is evidence of a marginally sig difference in mean log survival times between between low and high (although not totally consistent across sensitivity analyses)
- How to quantify/interpret on log scale?
- Power of study to detect what size effects?

  Alternative analysis to test for trend?