Chapter 15

# Other Design and Analysis Topics

**Solutions**

**15.1S.** Reconsider the experiment in Problem 5.25. Use the Box-Cox procedure to determine if a transformation on the response is appropriate (or useful) in the analysis of the data from this experiment.



With the value of lambda near zero, and since the confidence interval does not include one, a natural log transformation would be appropriate.

**15.2S.** In Example 6.3 we selected a log transformation for the drill advance rate response. Use the Box-Cox procedure to demonstrate that this is an appropriate data transformation.



Because the value of lambda is very close to zero, and the confidence interval does not include one, the natural log was the correct transformation chosen for this analysis.

**15.3.** In Problem 8.22 a replicated fractional factorial design was used to study substrate camber in semiconductor manufacturing. Both the mean and standard deviation of the camber measurements were used as response variables. Is there any indication that a transformation is required for either response?



The Box-Cox plot for the Camber Average suggests a natural log transformation should be applied. This decision is based on the confidence interval for lambda not including one and the point estimate of lambda being very close to zero. With a lambda of approximately 0.5, a square root transformation could be considered for the Camber Standard Deviation; however, the confidence interval indicates that no transformation is needed.

**15.4.** Reconsider the photoresist experiment in Problem 8.23. Use the variance of the resist thickness at each test combination as the response variable. Is there any indication that a transformation is required?

Table P8.4 from Problem 8.23 presents the range, and not the variance. The variance must first be calculated, and the most appropriate model fit to the data. A Box-Cox plot is shown below.



With the point estimate of lambda near zero, and the confidence interval for lambda not inclusive of one, a log transformation would be appropriate. The parameters included in the model affect this plot.

**15.5S.** In the grill defects experiment described in Problem 8.41 a variation of the square root transformation was employed in the analysis of the data. Use the Box-Cox method to determine if this is the appropriate transformation.

The Box-Cox plot is shown below. Because the confidence interval for the minimum lambda does not include one, the decision to use a transformation is correct. Because the lambda point estimate is close to zero, the natural log transformation would be appropriate. This is a stronger transformation than the square root.



**15.6.** In the central composite design of Problem 11.11, two responses were obtained, the mean and variance of an oxide thickness. Use the Box-Cox method to investigate the potential usefulness of transformation for both of these responses. Is the log transformation suggested in part (c) of that problem appropriate?



The Box-Cox plot for the Mean Thickness model suggests that a natural log transformation could be applied; however, the confidence interval for lambda includes one. Therefore, a transformation would have a minimal effect. The natural log transformation applied to the Variance of Thickness model appears to be acceptable; however, again the confidence interval for lambda includes one.

**15.7.** Problem 12.8 suggests using the *ln*(*s*2) as the response (refer to part b). Does the Box-Cox method indicate that a transformation is appropriate?



Because the confidence interval for lambda does not include one, a transformation should be applied. The confidence interval does not include zero; therefore, the natural log transformation is inappropriate. With the point estimate of lambda at –1.17, the reciprocal transformation is appropriate.

**15.8.** Myers, Montgomery and Vining (2002) describe an experiment to study spermatozoa survival. The design factors are the amount of sodium citrate, the amount of glycerol, and equilibrium time, each at two levels. The response variable is the number of spermatozoa that survive out of fifty that were tested at each set of conditions. The data are in the following table. Analyze the data from this experiment with logistical regression. E – e

|  |  |  |  |
| --- | --- | --- | --- |
| Sodium Citrate | Glycerol | Equilibrium Time | Number Survived |
| – | – | – | 34 |
| + | – | – | 20 |
| – | + | – | 8 |
| + | + | – | 21 |
| – | – | + | 30 |
| + | – | + | 20 |
| – | + | + | 10 |
| + | + | + | 25 |

Minitab Output

**Binary Logistic Regression: Number Survi, Freq versus Sodium Citra, Glycerol, .**

Link Function: Logit

Response Information

Variable Value Count

Number Survived Success 168

Failure 232

Freq Total 400

Logistic Regression Table

Odds 95% CI

Predictor Coef SE Coef Z P Ratio Lower Upper

Constant -0.376962 0.110113 -3.42 0.001

Sodium Citrate 0.0932642 0.110103 0.85 0.397 1.10 0.88 1.36

Glycerol -0.463247 0.110078 -4.21 0.000 0.63 0.51 0.78

Equilbrium Time 0.0259045 0.109167 0.24 0.812 1.03 0.83 1.27

AB 0.585116 0.110066 5.32 0.000 1.80 1.45 2.23

AC 0.0543714 0.109317 0.50 0.619 1.06 0.85 1.31

BC 0.112190 0.108845 1.03 0.303 1.12 0.90 1.38

Log-Likelihood = -248.028

Test that all slopes are zero: G = 48.178, DF = 6, P-Value = 0.000

Goodness-of-Fit Tests

Method Chi-Square DF P

Pearson 0.113790 1 0.736

Deviance 0.113865 1 0.736

Hosmer-Lemeshow 0.113790 6 1.000

This analysis shows that Glycerol (*B*) and the Sodium Citrate x Glycerol (*AB*) interaction have an effect on the survival rate of spermatozoa.

**15.9S.** The sums of squares and products for a single-factor analysis of covariance follow. Complete the analysis and draw appropriate conclusions. Use *α* = 0.05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of | Degrees of | Sums of Squares and Products | | |
| Variation | Freedom | *x* | *xy* | *y* |
| Treatment | 3 | 1500 | 1000 | 650 |
| Error | 12 | 6000 | 1200 | 550 |
| Total | 15 | 7500 | 2200 | 1200 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Sums of Squares and Products | | |  | Adjusted |  |  |
| Source | *df* | *x* | *xy* | *y* | *y* | *df* | *MS* | *F*0 |
| Treatment | 3 | 1500 | 1000 | 650 | - | - |  |  |
| Error | 12 | 6000 | 1200 | 550 | 310 | 11 | 28.18 |  |
| Total | 15 | 7500 | 2200 | 1200 | 554.67 | 14 |  |  |
| Adjusted | Treat. |  |  |  | 244.67 | 3 | 81.56 | 2.89 |

Treatments differ only at 10%.

**15.10.** Find the standard errors of the adjusted treatment means in Example 15.5.

From Example 15.5,  , , 







**15.11.** Four different formulations of an industrial glue are being tested. The tensile strength of the glue when it is applied to join parts is also related to the application thickness. Five observations on strength (*y*) in pounds and thickness (*x*) in 0.01 inches are obtained for each formulation. The data are shown in the following table. Analyze these data and draw appropriate conclusions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Glue | Formulation |  |  |  |
| 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 |
| *y* | *x* | *y* | *x* | *y* | *x* | *y* | *x* |
| 46.5 | 13 | 48.7 | 12 | 46.3 | 15 | 44.7 | 16 |
| 45.9 | 14 | 49.0 | 10 | 47.1 | 14 | 43.0 | 15 |
| 49.8 | 12 | 50.1 | 11 | 48.9 | 11 | 51.0 | 10 |
| 46.1 | 12 | 48.5 | 12 | 48.2 | 11 | 48.1 | 12 |
| 44.3 | 14 | 45.2 | 14 | 50.3 | 10 | 48.6 | 11 |

From the analysis performed in *Minitab*, glue formulation does not have a statistically significant effect on strength. As expected, glue thickness does affect strength.

Minitab Output

**General Linear Model: Strength versus Glue**

Factor Type Levels Values

Glue fixed 4 1 2 3 4

Analysis of Variance for Strength, using Adjusted SS for Tests

Source DF Seq SS Adj SS Adj MS F P

Thick 1 68.852 59.566 59.566 42.62 0.000

Glue 3 1.771 1.771 0.590 0.42 0.740

Error 15 20.962 20.962 1.397

Total 19 91.585

Term Coef SE Coef T P

Constant 60.089 1.944 30.91 0.000

Thick -1.0099 0.1547 -6.53 0.000

Unusual Observations for Strength

Obs Strength Fit SE Fit Residual St Resid

3 49.8000 47.5299 0.5508 2.2701 2.17R

R denotes an observation with a large standardized residual.

Expected Mean Squares, using Adjusted SS

Source Expected Mean Square for Each Term

1 Thick (3) + Q[1]

2 Glue (3) + Q[2]

3 Error (3)

Error Terms for Tests, using Adjusted SS

Source Error DF Error MS Synthesis of Error MS

1 Thick 15.00 1.397 (3)

2 Glue 15.00 1.397 (3)

Variance Components, using Adjusted SS

Source Estimated Value

Error 1.397

**15.12.** Compute the adjusted treatment means and their standard errors using the data in Problem 15.11.





















The adjusted treatment means can also be generated in *Minitab* as follows:

Minitab Output

Least Squares Means for Strength

Glue Mean SE Mean

1 47.08 0.5355

2 47.64 0.5382

3 47.91 0.5301

4 47.43 0.5314

**15.13.** An engineer is studying the effect of cutting speed on the rate of metal removal in a machining operation. However, the rate of metal removal is also related to the hardness of the test specimen. Five observations are taken at each cutting speed. The amount of metal removed (*y*) and the hardness of the specimen (*x*) are shown in the following table. Analyze the data using and analysis of covariance. Use α=0.05.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Cutting | Speed | (rpm) |  |
| 1000 | 1000 | 1200 | 1200 | 1400 | 1400 |
| y | x | y | x | y | x |
| 68 | 120 | 112 | 165 | 118 | 175 |
| 90 | 140 | 94 | 140 | 82 | 132 |
| 98 | 150 | 65 | 120 | 73 | 124 |
| 77 | 125 | 74 | 125 | 92 | 141 |
| 88 | 136 | 85 | 133 | 80 | 130 |

As shown in the analysis performed in *Minitab*, there is no difference in the rate of removal between the three cutting speeds. As expected, the hardness does have an impact on rate of removal.

Minitab Output

**General Linear Model: Removal versus Speed**

Factor Type Levels Values

Speed fixed 3 1000 1200 1400

Analysis of Variance for Removal, using Adjusted SS for Tests

Source DF Seq SS Adj SS Adj MS F P

Hardness 1 3075.7 3019.3 3019.3 347.96 0.000

Speed 2 2.4 2.4 1.2 0.14 0.872

Error 11 95.5 95.5 8.7

Total 14 3173.6

Term Coef SE Coef T P

Constant -41.656 6.907 -6.03 0.000

Hardness 0.93426 0.05008 18.65 0.000

Speed

1000 0.478 1.085 0.44 0.668

1200 0.036 1.076 0.03 0.974

Unusual Observations for Removal

Obs Removal Fit SE Fit Residual St Resid

8 65.000 70.491 1.558 -5.491 -2.20R

R denotes an observation with a large standardized residual.

Expected Mean Squares, using Adjusted SS

Source Expected Mean Square for Each Term

1 Hardness (3) + Q[1]

2 Speed (3) + Q[2]

3 Error (3)

Error Terms for Tests, using Adjusted SS

Source Error DF Error MS Synthesis of Error MS

1 Hardness 11.00 8.7 (3)

2 Speed 11.00 8.7 (3)

Variance Components, using Adjusted SS

Source Estimated Value

Error 8.677

Means for Covariates

Covariate Mean StDev

Hardness 137.1 15.94

Least Squares Means for Removal

Speed Mean SE Mean

1000 86.88 1.325

1200 86.44 1.318

1400 85.89 1.328

**15.14S.** Show that in a single factor analysis of covariance with a single covariate a 100(1-*α*) percent confidence interval on the ith adjusted treatment mean is



Using this formula, calculate a 95 percent confidence interval on the adjusted mean of Machine 1 in Example 15.5.

The 100(1–*α*) percent interval on the ith adjusted treatment mean would be



where is an estimator of the ith adjusted treatment mean. The standard error of the adjusted treatment mean is found as follows:



Since the  and  are independent. From regression analysis, we have . Therefore,



Replacing  by its estimator *MSE*, yields

 or



Substitution of this result into  will produce the desired confidence interval. A 95% confidence interval on the mean of machine 1 would be found as follows:







Therefore, , where *μ*1 denotes the true adjusted mean of treatment one.

**15.15.** Discuss how the operating characteristic curves for the analysis of variance can be used in the analysis of covariance.

To use the operating characteristic curves, fixed effects case, we would use as the parameter Φ2,



The test has *a*–1 degrees of freedom in the numerator and *a*(*n*–1) –1 degrees of freedom in the denominator.

**15.16.** Three different Pinot Noir wines were evaluated by a panel of eight judges. The judges are considered a random panel of all possible judges. The wines are evaluated on a 100-point scale. The wines were presented in random order to each judge, and the following results were obtained.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Wine | | |
| Judge | 1 | 2 | 3 |
| 1 | 85 | 88 | 93 |
| 2 | 90 | 89 | 94 |
| 3 | 88 | 90 | 98 |
| 4 | 91 | 93 | 96 |
| 5 | 92 | 92 | 95 |
| 6 | 89 | 90 | 95 |
| 7 | 90 | 91 | 97 |
| 8 | 91 | 89 | 98 |

Analyze the data from this experiment. Is there a difference in the wine quality? Analyze the residuals and comment on model adequacy.

This is a repeated measures problem where the three Pinot Noirs are the treatments and the random Judges are the subjects. As described in the textbook, the analysis for a single factor repeated measures is the same as the randomized complete block design, RCBD, where the Judges are the random blocks.

The analysis below identifies a significant difference in the wine quality.

The residual plots do not identify any concerns with model adequacy.

Design Expert Output

**Response** **1** **Wine Quality**

**ANOVA for selected factorial model**  
 **Analysis of variance table [Classical sum of squares - Type II]**  
 **Sum of** **Mean** **F** **p-value**  
 **Source** **Squares** **df** **Square** **Value** **Prob > F**

Block 48.00 7 6.86

Model 186.33 2 93.17 44.98 < 0.0001 significant  
  *A-Pinot Noir* *186.33* *2* *93.17* *44.98* *< 0.0001*  
 Residual 29.00 14 2.07  
 Cor Total 263.33 23

Std. Dev. 1.44 R-Squared 0.8653  
 Mean 91.83 Adj R-Squared 0.8461  
 C.V. % 1.57 Pred R-Squared 0.6042

PRESS 85.22 Adeq Precision 11.751

**Treatment Means (Adjusted, If Necessary)**  
 **Estimated** **Standard**  
 **Mean** **Error**

1-Wine 1 89.50 0.51

2-Wine 2 90.25 0.51  
 3-Wine 3 95.75 0.51

**Mean** **Standard** **t for H0**   
 **Treatment** **Difference** **df** **Error** **Coeff=0** **Prob > |t|**

1 vs 2 -0.75 1 0.72 -1.04 0.3150

1 vs 3 -6.25 1 0.72 -8.69 < 0.0001  
 2 vs 3 -5.50 1 0.72 -7.64 < 0.0001











**15.17.** A transformation that stabilizes the variance of the response also may make the response distribution closer to normal.

**True** False

**15.18.** A primary reason for inequality of variance is non-normality of the response distribution.

**True** False

**15.19.** Inequality of variance in an experiment can be caused by operator fatigue, tool wear, or depletion of some chemical reagent.

**True** False

**15.20.** A general linear model had the following components: a response distribution, a linear predictor and a link function.

**True** False

**15.21.** A commonly-used link function for binomial data is the logistic link.

**True** False

**15.22.** The generalized linear model is actually a non-linear model.

**True** False

**15.28.** Use of a generalized linear model is often a good alternative to a data transformation on the response.

**True** False

**15.29.** A covariate is a nuisance factor that can be easily controlled by the experimenter.

True **False**

**15.30.** Analysis of covariance is an extension of the ANOVA intended to incorporate the effects of uncontrolled, but measured variables into the analysis.

**True** False

**15.31.** An experiment that includes a covariate can improve the precision of the comparisons between controlled factors.

**True** False