Handout 4: The Statistical Model

Example 4.1 Consider the following etch rate data. The goal is to make comparisons across the power settings.

|  |  |
| --- | --- |
| Outcomes in Run Order, i.e. random order | Outcomes in Standard Order |

Summaries for this data.



|  |  |
| --- | --- |
| Overall Average | Average for each power setting |

**The Statistical Model**

In the last handout, we carried out the ANOVA and intuitively obtained measures of error by computing the sums of squares “by hand.” These calculations are made much simpler using the framework for a general linear model. The general linear model approach does require the use of matrices and some linear algebra operations.

In the previous handout, the goal was to compare different levels of a single factor. One way to write the statistical model for our example is as follows:

,

where

* i=1, 2, 3, and 4 identifies the treatment level, and
* j=1, 2, 3, 4, 5 denotes the replicate.

Identify the meaning of each of the terms in the model.

:

:

:

:



**The Model in Matrix Notation**

The easiest way to work with such a model is through the use of matrices. The model for our simple example can be expressed in matrix form as follows.

which is equivalent to



Here, *y* represents the **response vector** and X represents the **design matrix**. The estimated model parameters can be obtained as follows -- this approach to estimating the model parameter is *very general* and in fact works for any linear model.

Compute





Problem: The columns of are not linearly independent; thus, the inverse of cannot be computed.

Solution: We need to re-parameterize the model so that the model parameters can be estimated. Consider the following re-parameterization. This re-parameterization utilizes the **“Sum to Zero”** restriction .

Given this parameterization,

Design matrix using the sum-to-zero parameterization.



Computing these estimates in Minitab

Select **Stat > ANOVA > General Linear Model**.

|  |  |
| --- | --- |
|  |  |



**Getting the Estimates via Linear Algebra**







**The following estimate are computed**



**Getting the Predicted Outcomes using our General Linear Model**

The **predicted response vector** from our simply contains the mean for each treatment level. Compute the predicted response vector for our data.

Predicted response, often identified as , is given by

For our design matrix and estimated vector of model coefficients we have the following predicted outcomes.

**Getting the Residuals, i.e. Errors, using our General Linear Model**

The amount of error present in our model is simply the difference between the original response vector and the predicted response vector. The **residual vector** is computed as follows.

Getting the residual vector for our example.

The sums-of-squared error can easily be computed as follows

The average amount of squared error can be computed as

For our example, we get

**The Standard Error of Model Estimates**

The estimate of is used in computing the standard errors of the estimated model parameters. The standard error estimates of our model parameters are necessary for hypothesis testing and for computing confidence intervals (which we used in the previous handout).

The variance-covariance matrix of the estimated parameter vector is given by

Computing the variance of the estimated model parameters for our example.

The standard errors for the estimated model parameters obtained from Minitab.



The diagonal elements of this matrix are the variances for each of the estimated model parameters. The off-diagonal elements are the covariances. A nonzero covariance implies that variation in of the estimated model parameters influences the variation in another. The **design matrix results in, i.e. causes, a covariance structure** amongst the .

This covariance structure implies that the design choice will impact our ability to estimate one treatment level effect independent of another. Designs that permit one treatment level effect to be estimated independent of another are often considered optimal designs.

Next, we will consider the standard error necessary to compare across treatment levels.

The following output from Minitab contains the standard error Difference quantities for the treatment level comparisons.





Next, we will confirm the calculations for one of these standard errors, say the comparison for Power Setting 180 against Power Setting 160.

* Computing the Difference of Means
* The above Difference in Means is simply a linear combination, say , of the estimated parameter vector, . In particular,

The associated SE of Difference can be computed using the fact that

with yields

The standard error is simply the square root of this quantity.

Example 4.2 A manufacturer of television sets is interested in the effect on tube connectivity of four different types of coating for color picture tubes. A completely randomized experiment is conducted and the following conductivity data are obtained.



Complete the following:

1. Write out the general linear model (using matrix notation) for this data. In particular, clearly specify the response vector **y**, the design matrix **X**, the vector of model parameters ****, and the error vector ****. You should use the set-to-zero parameterization for the model as in done in Minitab.

Consider the following Minitab output from an appropriate analysis.



1. Verify ❶. That is, use your design matrix **X** and the response vector **y** to compute the estimated model parameters.
2. Use your estimated model parameters to verify at least one of the Means given in ❷.



1. Compute the estimated residual vector, , or your model. (See page 7).
2. Use the estimated residual vector to verify ❸. In particular, compute the Adj MS value.
3. Use the following to verify ❹on the Minitab output.



1. Use your design matrix **X** and

to verify at least one of the quantities in ❺.

1. Consider the predicted or fitted value for Group 2 of 145.25, i.e. ❻, and it’s associated standard error value of 2.22, i.e. ❼.



What is the form of the vector for Group 2? Use the vector to verify the calculations for the fitted value and standard error provided above.

Consider the following output from the pairwise comparison portion of the Minitab output.



1. Verify the calculations for ❽. What is the appropriate **c** vector for this calculation?
2. Verify the calculations for ❾ using the same **c** vector from above.