

Chapter 213

Analysis of Two-Level Designs

Introduction

Several analysis programs are provided for the analysis of designed experiments. The GLM-ANOVA and the Multiple Regression programs are often used. This chapter describes a program to analyze very particular designs: two-level factorials (with an optional blocking variable) in which the number of rows is a power of two (4, 8, 16, 32, 64, 128, etc.) and there are no missing values.

Given that your data meet these restrictions, this program gives you a complete analysis including:

1. Analysis of the design itself.
2. List of confounding and aliasing patterns.
3. Analysis of variance table.
4. Tables of means and effects.
5. Probability plots of residuals and effects.
6. Two-way and cube plots of means and differences.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

These options specify the variables that will be used in the analysis. They also specify the type of analysis that will be performed.

Response Variable

Response Variable

Specifies the response (dependent) variable to be analyzed.

Block Variable

Block Variable

An optional variable containing the levels of the blocking factor. Note that block sizes must be a power of two.

Analysis of Two-Level Designs

Factor Specification

Factor Variables

At least two factor (categorical) variables must be specified. Each factor consists of a variable that contains a column of two unique values (two levels). The values may be text or numeric.

Error Estimation Options

Pooled Terms

Often, two-level designs do not provide a direct estimate of the mean square error (MSE). The F-tests in the analysis of variance require an estimate of the MSE, so this option lets you specify one.

This option provides a list of term numbers (separated by commas) that represent the terms that should be pooled (averaged) to form the estimated MSE. These should be determined from the probability plot of the effects and from the Sorted Means and Effects report. This is a list of the terms whose effect is small in absolute value.

This list is optional and may be left blank, in which case it will be ignored. Note that this list is also ignored when the Estimated MSE option is non-zero.

Estimated MSE

Often, two-level designs do not provide a direct estimate of the mean square error (MSE). The F-tests in the analysis of variance require an estimate of the MSE, so this option lets you specify one.

This option allows the direct specification of an MSE value. This value overrides the Pooled Terms option when it is nonzero. The degrees of freedom associated with the MSE are set to 99. This MSE value should be determined from the analysis of variance of previous experiments.

This value is optional and may be left at zero, in which case it will be ignored.

Error DF

Enter a value for the error degrees of freedom. This value is only used when the 'Estimated MSE' is non-zero.

Reports Tab

The following options control which plots and reports are displayed.

Select Additional Reports

Show Two-Way Tables

Display the Two-Way Tables of Means and Effects. These reports are useful in analyzing the two-way interactions.

Show Three-Way Tables

Display the Three-Way Tables of Means and Effects. These reports are useful in analyzing the three-way interactions.

Report Options

Confounding Size

Specify what order interactions are included in the report showing the confounding and aliasing patterns in the design.

Analysis of Two-Level Designs

Alpha Level

The value of alpha for the statistical tests. Usually, this number will range from 0.1 to 0.001. A common choice for alpha is 0.05, but this value is a legacy from the age before computers when only printed tables were available. You should determine a value appropriate for your particular study.

Report Options – Decimal Places

Decimals in Means

Specify the number of decimal places to use when displaying means and effects.

Decimals in Mean Squares

Specify the number of decimal places to use when displaying mean squares.

Plots Tabs

These options control the appearance of the two probability plots and the means plots that may be generated. More details on their interpretation will be contained in annotated output section presented later.

Select Plots

Show Probability Plots

Specify whether to display probability plots of the residuals and effects.

Show Means Plots

Specify whether to display plots of the means for all factors and two-way interactions.

Click the plot format button to change the plot settings.

Storage Tab

The residuals calculated for each row may be stored on the current database for further analysis. This option lets you designate where to store the residuals.

Data Storage Variables

Residuals

If a variable is specified here, the residuals are automatically stored in that variable. Note that any previous values in the variable are automatically replaced.

Example 1 – Running the Analysis of a Two-Level Design

This section presents an example of how to analyze data using this program. We will analyze the three-factor experiment given on page 320 of Box and Hunter (1978). These data are the results of a pilot plant study conducted to investigate the influence of temperature, concentration, and catalyst on chemical yield. The data are contained in the Box320 dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Analysis of Two-Level Designs window.

1 Open the Box320 dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Box320.NCSS**.
- Click **Open**.

2 Open the Analysis of Two-Level Designs window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Two-Level Designs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- Select the **Variables tab**.
- Double-click in the **Response Variable** box. This will bring up the variable selection window.
- Select **Yield** from the list of variables and then click **Ok**. “Yield” will appear in the Response Variables box.
- Double-click in the **Factor Variables** box. This will bring up the variable selection window.
- Select **Temp, Concentration, Catalyst** from the list of variables and then click **Ok**. “Temp-Catalyst” will appear in the Factor Variables box.

4 Specify the reports.

- Select the **Reports tab**.
- Select **All** in the Confounding Size list box.

5 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Design Information Section

Design Information Section			
Input Data			
Response:	Yield		
Rows:	16		
Reps:	2		
Blocks:	None		
Factor Symbol	Factor Name	Level One	Level Two
A(1)	Temp	160	180
B(2)	Concentration	20	40
C(3)	Catalyst	1	2
Design			
2/1 replication of 3 factors.			

This section describes the experimental design of the data.

Confounding / Alias Section

Confounding / Alias Section

Term No.	Terms Confounded
1	A
2	B
3	AB
4	C
5	AC
6	BC
7	ABC

This section reports confounding and aliasing information for each term (degree of freedom). In the present example, a complete replication is given so there is no confounding. Hence, each degree of freedom is associated with only one term in the analysis of variance model.

Means and Effects Section

Means and Effects Section

Term No.	Term Symbol	Mean -	Mean +	Estimated Effect	Standard Error
0	Grand Mean			64.25	0.71
1	A (Temp)	52.75	75.75	23.00	1.41
2	B (Concentration)	66.75	61.75	-5.00	1.41
3	AB	63.50	65.00	1.50	1.41
4	C (Catalyst)	63.50	65.00	1.50	1.41
5	AC	59.25	69.25	10.00	1.41
6	BC	64.25	64.25	0.00	1.41
7	ABC	64.00	64.50	0.50	1.41

This section reports on the estimation of the effects for each degree of freedom.

Term No.

This is an arbitrary identification number assigned to each degree of freedom in the model. This number is needed to correctly specify terms to be pooled as MSE.

Term Symbol

This is the letter that is assigned to each factor. Since we have three factors in this database, three letters (A, B, and C) are used. The names of the variables associated with a given letter are shown in parentheses.

Mean -

The average of all observations having the low value (-1) for this term. If you think of the low value as -1 and the high value as +1, then interaction terms (formed by multiplication) will also have only two possible values, -1 and +1.

Mean +

The average of all observations having the high value (+1) for this term. If you think of the low value as -1 and the high value as +1, then interaction terms (formed by multiplication) will also have only two possible values, -1 and +1.

Estimated Effect

The estimated effect value. This is equal to (Mean +) - (Mean -).

Analysis of Two-Level Designs

Standard Error

The estimated standard error of the above effect value. Note that this standard error only depends on the MSE, so it is constant for all terms. Remember, this value is not calculated from individual groups of data but from the MSE!

Sorted Means and Effects Section

Sorted Means and Effects Section					
Term No.	Term Symbol	Mean -	Mean +	Estimated Effect	Standard Error
0	Grand Mean			64.25	0.71
1	A (Temp)	52.75	75.75	23.00	1.41
5	AC	59.25	69.25	10.00	1.41
2	B (Concentration)	66.75	61.75	-5.00	1.41
4	C (Catalyst)	63.50	65.00	1.50	1.41
3	AB	63.50	65.00	1.50	1.41
7	ABC	64.00	64.50	0.50	1.41
6	BC	64.25	64.25	0.00	1.41

This is a sorted version of the report presented in the last section. The report is sorted by the absolute value of the Estimated Effect. This report is used with the Probability Plot of Effects to pick those effects that are not large enough to be important and thus can be pooled (averaged) into the MSE.

Analysis of Variance Section

Analysis of Variance Section						
Term No.	Term Symbol	DF	Mean Square	F-Ratio	Prob Level	Statistically Significant
1	A (Temp)	1	2116.0000	264.50	0.000000	Yes
2	B (Concentration)	1	100.0000	12.50	0.007670	Yes
3	AB	1	9.0000	1.13	0.319813	No
4	C (Catalyst)	1	9.0000	1.13	0.319813	No
5	AC	1	400.0000	50.00	0.000105	Yes
6	BC	1	0.0000	0.00	1.000000	No Test
7	ABC	1	1.0000	0.13	0.732810	No
	Error	8	8.0000			
	Total	15	2699.0000			

This section presents the analysis of variance table.

Term No.

The identification number of this degree of freedom.

Term Symbol

This is the symbol of this term. Refer to the Confounding / Alias Section for a list of all main effects and interactions associated with this term.

DF

The degrees of freedom. The number of observations used by the corresponding model term.

Mean Square

An estimate of the variation in the response accounted for by this term. The sum of squares divided by the degrees of freedom.

Analysis of Two-Level Designs

F-Ratio

The ratio of the mean square for this term and mean square error (MSE). This F-ratio tests the statistical significance of the effects associated with this term.

Prob Level

The probability of obtaining an F-ratio larger than that obtained by the analysis.

Statistically Significant

If the probability level is less than the value of alpha that was set, the term is designated as being statistically significant (Yes). If it is not less than alpha, the term is not statistically significant (No).

Two-Way Tables of Means and Effects

Means and Effects of Concentration by Temp

Concentration	Temp 160	Temp 180	Effect	Overall
20	56.00	77.50	21.50	66.75
40	49.50	74.00	24.50	61.75
Effect	-6.50	-3.50	1.50	-5.00
Overall	52.75	75.75	23.00	64.25

This report presents the two-way interaction means and effects. One report is displayed for each of the possible two-way interactions.

The four means in the upper left-hand corner (56.0, 77.5, 49.5, 74.0) are the individual means. For example, 56.0 is the average of all rows in which the Temp was 160 and Concentration was 20.

The fourth column (Effect) is the estimated effect for that row. In the first, second, and fourth rows, this is the difference between the two previous columns. For example, $77.50 - 56.00 = 21.50$. In the third row (Effect), the value (the effect of the interaction) is calculated as the difference between the two previous columns divided by 2. In this example, $((-3.50) - (-6.50))/2 = 1.50$.

The fifth column (Overall) gives the mean or effect for this row averaged across all column values. Thus, 66.75 is the average of all rows in which the value of Concentration was 20. Finally, 64.25 is the average of all rows in the experiment.

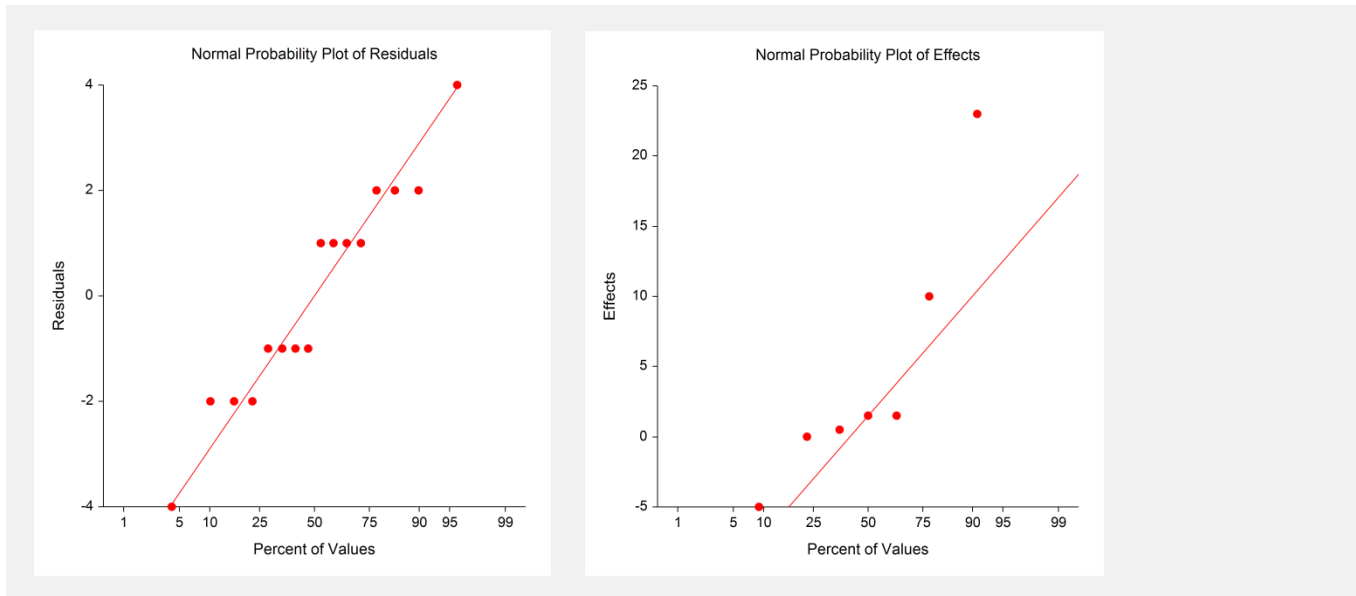
Three-Way Tables of Means

Means and Effects of Concentration by Temp by Catalyst

Concentration	Temp 160	Temp 160	Temp 180	Temp 180
	Catalyst 1	Catalyst 2	Catalyst 1	Catalyst 2
20	60.00	52.00	72.00	83.00
40	54.00	45.00	68.00	80.00

This report presents a three-way table of the means. For example, 60.00 is the average of all rows in which Temp was 160, Concentration was 20, and Catalyst was 1.

Probability Plots Section



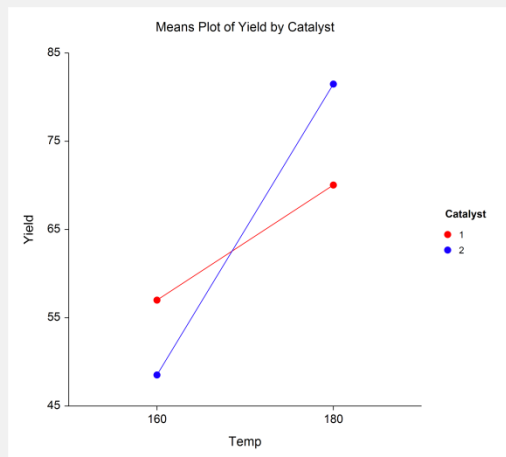
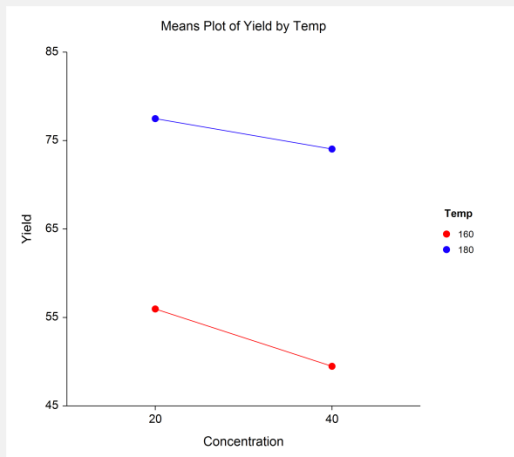
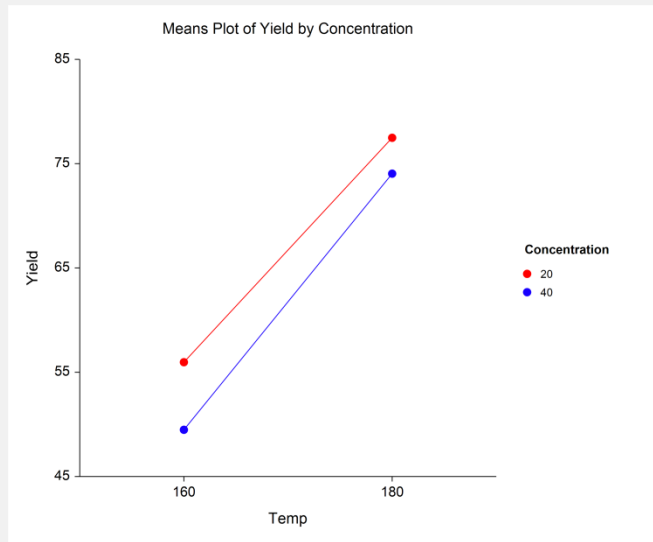
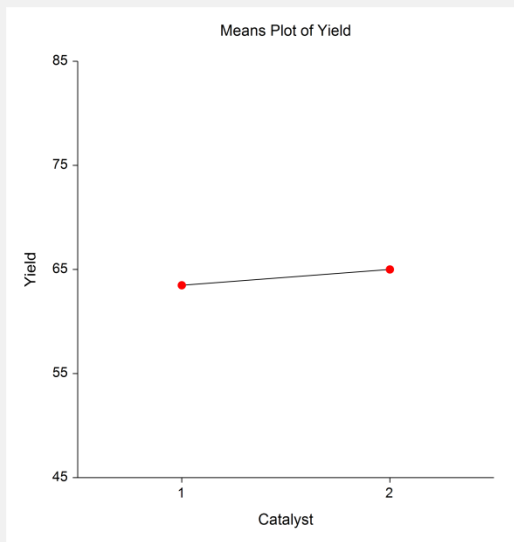
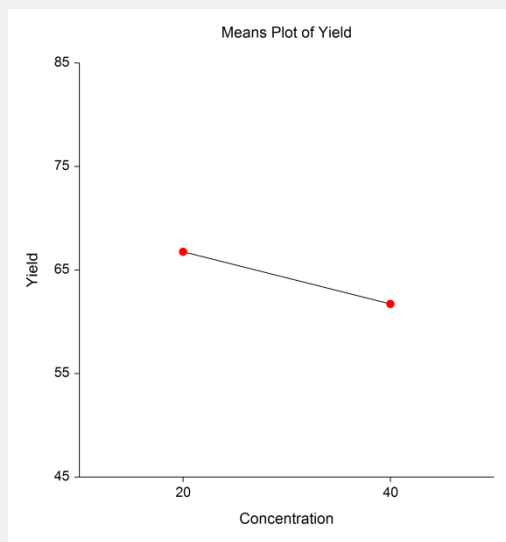
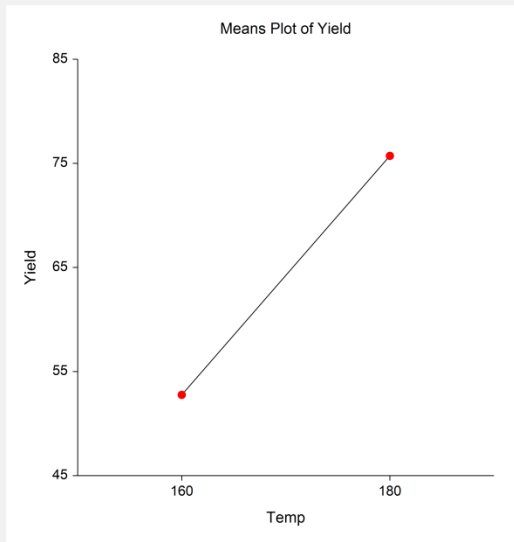
A normal probability plot of the residuals is supplied to allow you to study the distribution of the residuals. This plot will not be displayed if the residuals are all zero (which often occurs in designs like fractional factorials).

A normal probability plot of the effects is supplied to allow you to consider the relative sizes of the effects. If all terms are non-significant (and hence come from the normal distribution), these effects should fall along a straight line. When some of the effects are significant, they will fall off this line. The plot is useful for visually interpreting designs that do not supply an explicit estimate of the experimental error variance (such as fractional factorial designs).

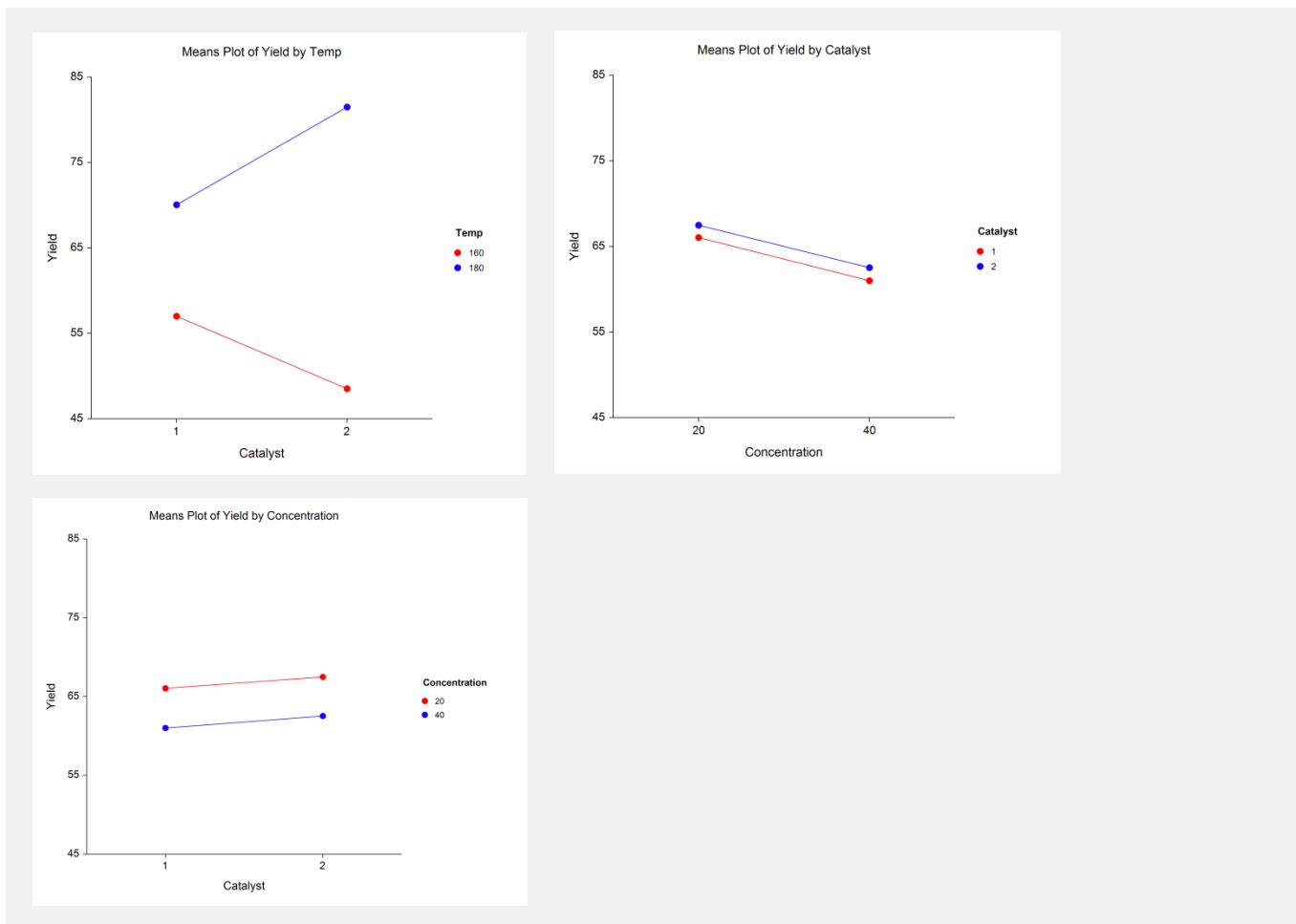
In our example, the first six points seem to fall along a straight line, while the final point falls off this line. This term is associated with Temp (factor A), as you can see from the Means and Effects report.

Analysis of Two-Level Designs

Means Plots Section



Analysis of Two-Level Designs



These plots display the means for all one-way and two-way interaction terms.

Example 2 – Analysis of a Two-Level Design

This section presents another example of how to analyze data using this program. We will analyze an eight-factor experiment given on page 402 of Box and Hunter (1978). These data are the results of an injection molding study. The data are contained in the Box402 dataset.

To run this example, open the Box402 dataset and load the completed template **Example 2** by clicking on Open Example Template from the File menu of the Analysis of Two-Level Designs window. Running this template will yield the following results.

Analysis of Two-Level Designs

Design Information Section

Design Information Section

Input Data

Response: Shrinkage
 Rows: 16
 Reps: 1
 Blocks: None

Factor Symbol	Factor Name	Level One	Level Two
A(1)	MoldTemp	1	2
B(2)	Moisture	1	2
C(3)	HoldPressure	1	2
D(4)	Thickness	1	2
E(5)	BoosterPressure	1	2
F(6)	CycleTime	1	2
G(7)	GateSize	1	2
H(8)	ScrewSpeed	1	2

Design

1/16 replication of 8 factors.

Defining Contrast

$i = BCDE = ACDF = ABEF = ABCG = ADEG = BDFG = CEFG = ABDH = ACEH = BCFH = DEFH = CDGH = BEGH = AFGH = ABCDEFGH$

Design Construction

Full model of the factors [A B C D].
 The remaining factors are aliased with interactions of this reduced model as follows:
 $E=BCD$ $F=ACD$ $G=ABC$ $H=ABD$

Notice the Defining Contrast and the Design Construction reports.

Confounding / Alias Section

Confounding / Alias Section

Term No.	Terms Confounded
1	A+CDF+BEF+BCG+DEG+BDH+CEH+FGH
2	B+CDE+AEF+ACG+DFG+ADH+CFH+EGH
3	AB+EF+CG+DH
4	C+BDE+ADF+ABG+EFG+AEH+BFH+DGH
5	AC+DF+BG+EH
6	BC+DE+AG+FH
7	G+ABC+ADE+BDF+CEF+CDH+BEH+AFH
8	D+BCE+ACF+AEG+BFG+ABH+EFH+CGH
9	AD+CF+EG+BH
10	BD+CE+FG+AH
11	H+ABD+ACE+BCF+DEF+CDG+BEG+AFG
12	CD+BE+AF+GH
13	F+ACD+ABE+BDG+CEG+BCH+DEH+AGH
14	E+BCD+ABF+ADG+CFG+ACH+DFH+BGH
15	AE+BF+DG+CH

This section reports confounding and aliasing information for each term (degree of freedom). Note that in this design, no two-way interactions are confounded with any of the main effects. Note, however, the all two-way interactions are confounded with each other.

Analysis of Two-Level Designs

Analysis of Variance Section

Analysis of Variance Section

Term No.	Term Symbol	DF	Mean Square	F-Ratio	Prob Level	Statistically Significant
1	A (MoldTemp)*	1	1.9600	1.39	0.260707	No
2	B (Moisture)*	1	0.0400	0.03	0.868897	No
3	AB*	1	1.4400	1.02	0.331610	No
4	C (HoldPressure)	1	121.0000	86.02	0.000001	Yes
5	AC*	1	3.2400	2.30	0.154993	No
6	BC*	1	0.1600	0.11	0.741747	No
7	G (GateSize)*	1	1.4400	1.02	0.331610	No
8	D (Thickness)*	1	0.3600	0.26	0.622099	No
9	AD*	1	0.6400	0.45	0.512767	No
10	BD*	1	1.4400	1.02	0.331610	No
11	H (ScrewSpeed)*	1	5.7600	4.09	0.065871	No
12	CD*	1	0.3600	0.26	0.622099	No
13	F (CycleTime)*	1	0.0400	0.03	0.868897	No
14	E (BoosterPressure)	1	57.7600	41.06	0.000034	Yes
15	AE	1	84.6400	60.17	0.000005	Yes
	Error	12	1.4067			
	Total	15	280.2800			

Terms marked with an '*' have been pooled to form the error term.

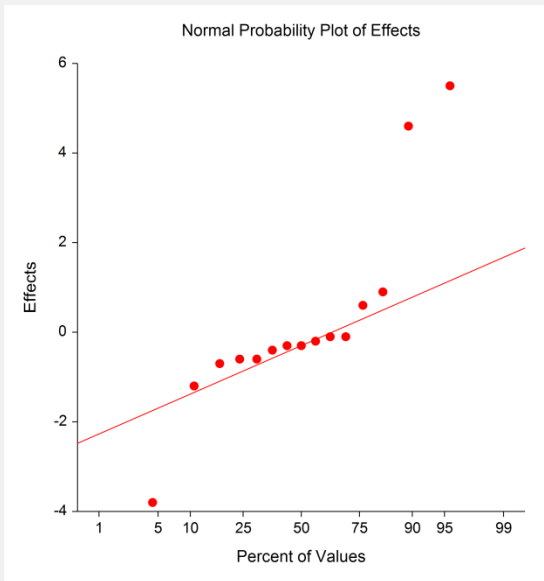
Now we see that HoldPressure, BoosterPressure, and at least one of the two-way interactions AE+BF+DG+CH are significant.

Means and Effects and Probability Plot Sections

Sorted Means and Effects Section

Term No.	Term Symbol	Mean -	Mean +	Estimated Effect	Standard Error
0	Grand Mean			19.75	0.30
4	C (HoldPressure)	17.00	22.50	5.50	0.59
15	AE	17.45	22.05	4.60	0.59
14	E (BoosterPressure)	21.65	17.85	-3.80	0.59
11	H (ScrewSpeed)*	20.35	19.15	-1.20	0.59
5	AC*	19.30	20.20	0.90	0.59
1	A (MoldTemp)*	20.10	19.40	-0.70	0.59
7	G (GateSize)*	19.45	20.05	0.60	0.59
3	AB*	20.05	19.45	-0.60	0.59
10	BD*	20.05	19.45	-0.60	0.59
9	AD*	19.95	19.55	-0.40	0.59
8	D (Thickness)*	19.90	19.60	-0.30	0.59
12	CD*	19.90	19.60	-0.30	0.59
6	BC*	19.85	19.65	-0.20	0.59
2	B (Moisture)*	19.80	19.70	-0.10	0.59
13	F (CycleTime)*	19.80	19.70	-0.10	0.59

Analysis of Two-Level Designs



From the probability plot you can see that three of the effects fall outside the range that would be expected if all effects come from the normal distribution. By looking at the Sorted Means and Effects report, we see that these three terms are numbers 4, 15, and 14. Hence, we decided to pool the rest of the terms to form an estimate of the experimental error variance (MSE) and rerun the program. We add the text 1,2,3,5,6,7,8,9,10,11,12,13 to the Pooled Terms option and rerun. The analysis of variance table above is produced. Note that without pooling these terms, the error DF would have been zero and no F-Ratios would have been generated.