JMP® ACADEMIC CASE STUDY

JMP039: Detergent Cleaning Effectiveness

Statistical Modeling – Analysis of Variance

Produced by

Kevin Potcner, JMP Global Academic Team
kevin.potcner@jmp.com

JMP STATISTICAL DISCOVERY

Detergent Cleaning Effectiveness
Statistical Modeling – Analysis of Variance

Key ideas:

This case study requires constructing an ANOVA-based statistical model to explore and describe the effect that multiple factors have on a response, as well as identifying conditions with the most and least impact.

Background

The effectiveness of a consumer product is of paramount importance, both to the company that manufacturers it and the consumer who uses it. This effectiveness determines the product's price point, customer satisfaction, and ultimately profitability for the company. One such consumer product used in almost every household is laundry detergent.

Detergent manufacturers are constantly working on developing new formulations that can perform better than the current one and be manufactured at a lower cost. The R&D team for one such manufacturer has developed a promising new laundry detergent formulation that can be produced with a 10% cost reduction.

To study the effectiveness of this new detergent formulation and to compare its performance to the current formulation, the following experiment was performed and the results stored in a JMP data table.

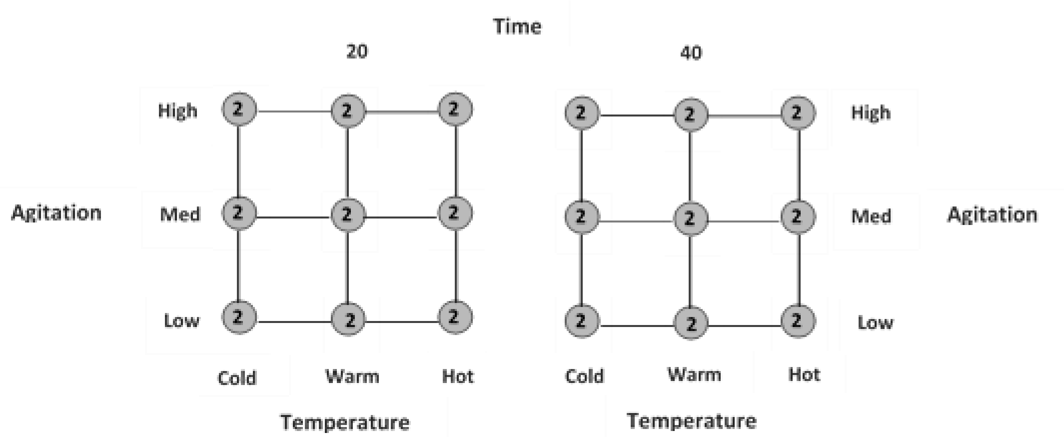
A set of cotton fabric specimens were prepared by being soiled with a dirt-based substance uniformly across the fabric. The test specimens were cut into two sub-specimens. One of them was washed using the current formulation of detergent and the other with the new formulation. After washing, each sub-specimen was measured using a reflective densitometer to obtain a brightness measure. The difference between the two brightness readings was used to create a metric that is a measure of the percent increase in brightness of the new formulation versus the current one. Specifically, the value 0 represents no difference in the brightness readings of the two sub-specimens, a value of 10.0 represents 10% more brightness in the sub-specimen washed with the new formulation compared to the current formulation, and a value of -10.0 represents the sub-specimen washed with the new detergent having a brightness reading 10% less than the one washed with the current formulation.

In order to study the effectiveness of the new formulation across a range of washing conditions, all combinations of three water temperatures (Cold, Warm, Hot), two washing times (20 minutes, 40 minutes), and three agitation levels (Low, Med, High) was used. This resulted in 3x2x3=18 experimental treatment combinations.

An important principle in designing experiments is to obtain an estimate of experiment error, which is an estimate of the variation that occurs between experimental units receiving the same treatment. The experimenters decided to replicate the experiment so that each treatment combination of Temp, Time, and Agitation was performed twice, resulting in 36 experimental runs, since (3x2x3)x2=36.

A diagram of the experiment is shown in Exhibit 1.

Exhibit 1 Experimental Design



The Task

The primary objectives of this experiment are to:

1. Compare the performance between the two formulations quantifying the cleaning effectiveness of the new formulation compared to the current.
2. Determine if the difference in cleaning effectiveness between the two formulations is consistent across all Temperatures, Agitation Levels, and Washing Times or if the difference is dependent upon the specific washing conditions.
3. Determine the specific temperatures, agitation levels, and washing times that result in the greatest difference in cleaning effectiveness between the two formulations. Determine the conditions that result in the least difference. Determine if there any conditions where there is either no difference or the new formulation performs worse.

The Data   Cleaning Effectiveness_1.jmp

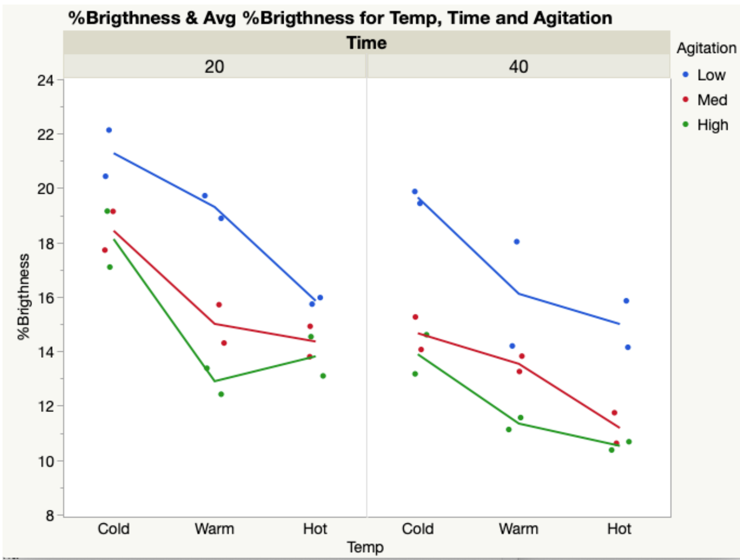
Stain	Stain type (Dirt)
Temp	Water temperature (Cold, Warm, Hot)
Time	Washing time (20 minutes, 40 minutes)
Agitation	Level of agitation (Low, Med, High)
Rep	Replicate of each experimental condition (1, 2)
%Brightness	The percent increase in brightness of the sub-specimen washed using the new formulation of the detergent versus the current formulation

Analysis

Graphical

We begin by visualizing the data. Exhibit 2 is a graph displaying the %Brightness values for all the 3x2x3=18 different experimental conditions (colored dots) along with the average values for the two replicates (shown in the colored lines).

Exhibit 2 Scatterplot with Mean Connect Lines



To create, Graph>Graph Builder. Select %Brightness as the Y variable, Temp as the X variable, Time as the group variable, and Agitation as the overlay variable. Select Points and Line of Fit in the graph palette.

Note: Any combination of roles for the experimental factors (e.g., X variable, group variable, and overlay variable) can be used. A good practice is to choose a layout that best communicates the features you wish to convey, such as assigning the X and overlay variables to the factors you think will be affected most.

A few features are seen in the graph. The %Brightness values are the highest for the Low Agitation level, while High Agitation has the lowest values. Results for the Med Agitation are between the two but closer to High Agitation. The Cold Temp values have the highest %Brightness values compared to the Warm and Hot Temp. The %Brightness values are also higher for 20-minute wash Time. It is important to refrain from reaching final conclusions just yet without more formal statistical analyses. Some of these differences we see in the graph may be random experimental variation rather than statistically significant.

Numerical summary

To summarize the experimental results numerically, a table of the mean and range for the 18 experimental conditions is shown in Exhibit 3.

Exhibit 3 Tabulate

			Mean	Range
Temp	**Time**	**Agitation**	**%Brigthness**	**%Brigthness**
Cold	20	Low	21.28	1.70
		Med	18.43	1.42
		High	18.12	2.06
	40	Low	19.65	0.44
		Med	14.66	1.20
		High	13.89	1.44
Warm	20	Low	19.30	0.83
		Med	15.01	1.41
		High	12.90	0.96
	40	Low	16.11	3.83
		Med	13.54	0.57
		High	11.34	0.44
Hot	20	Low	15.85	0.24
		Med	14.36	1.12
		High	13.81	1.44
	40	Low	15.00	1.71
		Med	11.18	1.12
		High	10.52	0.30

To create, Analyze>Graph Tabulate. Drag the experimental factors into the desired drop zones on the side. Drag Mean and Range into the column drop zones to create the desired table. Drag the %Brightness variable into the center of the table. Note: You may also use other table configurations. It is a good practice to choose a layout that best communicates the features you wish to convey.

This table provides numerical summaries of the experimental results. Though the graphical display is a much easier way to see and compare the results across the different experimental conditions, it is still very important to generate numerical summaries as they allow us to quantify those results.

Analysis of variance model

An analysis of variance (ANOVA) model is a standard technique used to compare means across various experimental conditions. Exhibit 4 shows the Effect Summary table from creating an ANOVA model that contains main effects for each of the three factors and the three possible two-way interactions between those factors.

Exhibit 4 Effect Summary Table

Effect Summary

Source	LogWorth	PValue
Agitation	8.351	0.00000
Temp	7.854	0.00000
Time	6.272	0.00000
Temp*Agitation	0.580	0.26322
Time*Agitation	0.371	0.42570
Temp*Time	0.345	0.45182

Remove Add Edit FDR

*To create, Analyze>Fit Model. Select %Brightness as the Y variable. Highlight the three experimental factors (Temp, Time, and Agitation) and choose Macros>Factorial to Degree. Note: Degree 2 is set as the default. It will fit a model containing the three main effects for Temp, Time, and Agitation, as well as the three possible two-way interactions (Temp*Time, Temp*Agitation, Time* Agitation).*

The Effects Summary table displays p-values corresponding to statistical tests evaluating the hypothesis that each of the six terms we added to the model are actually helpful in describing the data. That is, the table demonstrates if there is evidence to conclude that a term is statistically different than zero and thus should be included in the model. The LogWorth value is $-\log_{10}$(p-value). This transformation adjusts the p-values to provide a more appropriate scale for graphing. A value that exceeds 2 is significant at the 0.01 level because $-\log_{10}(0.01)=2$.

The standard approach to creating an ANOVA model is to reduce the model such that it only includes terms that are statistically significant, thus the terms that are useful in describing the features in the data. This process of reducing a model begins by examining the most complicated terms first. In this case, it means looking at the three two-way interactions. The p-values for each of these are 0.45182 for Temp*Time; 0.42570 for Time*Agitation; and 0.26322 for Temp*Agitation. They are all quite a bit larger than any standard significance level used (e.g., 0.01, 0.05, 0.10), which indicates that these interaction terms are not helpful at describing the features in the data. An ANOVA model without interactions terms can be interpreted as the effect that each one of the three experimental factors has on the response is similar across the levels of the other factors.

Exhibit 5 shows the same Effects Summary table with the three interaction terms removed.

Exhibit 5 Effect Summary Table

Effect Summary

Source	LogWorth	PValue
Agitation	9.447	0.00000
Temp	8.706	0.00000
Time	7.007	0.00000

Remove Add Edit Undo FDR

To create, highlight the three interaction terms and choose Remove.

The p-values for the three main effects are highly significant (<0.0000 for all). This demonstrates that there is a large amount of statistically significant evidence to indicate that the mean %Brightness values are not equal across the different levels of each factor.

Time has only two levels (20 minutes and 40 minutes). The hypothesis being tested in the Effects Summary table is:

$$\mathrm{H_0}: \mu_{Time\ 20} = \mu_{Time\ 40}$$

$$\mathrm{H_A}: \mu_{Time\ 20} \neq \mu_{Time\ 40}$$

The experimental factor Temp has three levels. The hypothesis being tested in the Effects Summary table is:

$$\mathrm{H_0}: \mu_{Temp\ Cold} = \mu_{Temp\ Warm} = \mu_{Temp\ Hot}$$

$\mathrm{H_A}$: $\mu_{Temp\ Cold}$, $\mu_{Temp\ Warm}$ and $\mu_{Temp\ Hot}$ are not all equal

Note that the alternative hypothesis $\mathrm{H_A}$ does not state that “all three means are not equal” but instead, “the three means are not all equal.” The significant result we have indicates only that at least one of the means is different. A further analysis would be needed to determine which ones are different.

The experimental factor Agitation also has three levels. The hypothesis being tested in the Effects Summary table is:

$$\mathrm{H_0}: \mu_{Ag\ Low} = \mu_{Ag\ Med} = \mu_{Ag\ High}$$

$\mathrm{H_A}$: $\mu_{Ag\ Low}$, $\mu_{Ag\ Med}$ and $\mu_{Ag\ High}$ are not all equal

Similar to the conclusion for the factor Temp, further analysis will be required to determine which ones are different.

Multiple comparisons

Exhibit 6 displays the results of the multiple comparisons for Temp using Student's t method.

Exhibit 6 Multiple Comparison (Student's t)

Multiple Comparisons for Temp

Student's t All Pairwise Comparisons

All Pairwise Differences

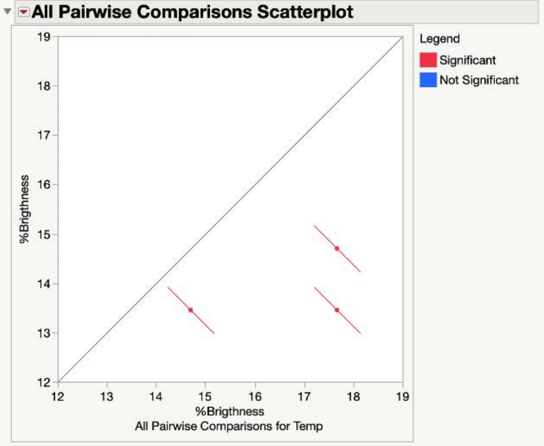
Temp	-Temp	Difference	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Cold	Warm	2.972035	0.4600926	6.46	<.0001*	2.032400	3.911669
Cold	Hot	4.216975	0.4600926	9.17	<.0001*	3.277340	5.156609
Warm	Hot	1.244940	0.4600926	2.71	0.0111*	0.305306	2.184575

All Pairwise Differences Connecting Letters

Temp				Least Squares Mean
Cold	A			17.671758
Warm		B		14.699723
Hot			C	13.454783

Levels not connected by same letter are significantly different.

All Pairwise Comparisons Scatterplot



To create, select Estimates>Multiple Comparison under the red triangle menu at the top of the output. Select the variable Temp and choose All Pairwise Comparisons – Student's t. Click OK. Then select All Pairwise Differences Connecting Letters under the red triangle next to Student's t All Pairwise Comparisons heading.

A set of statistical tests are conducted for each possible pairwise difference. These hypotheses can be written as:

$$H_0: \mu_{Temp\ Cold} = \mu_{Temp\ Warm}$$

$$H_A: \mu_{Temp\ Cold} \neq \mu_{Temp\ Warm}$$

$$H_0: \mu_{Temp\ Cold} = \mu_{Temp\ Hot}$$

$$H_A: \mu_{Temp\ Cold} \neq \mu_{Temp\ Hot}$$

$$H_0: \mu_{Temp\ Warm} = \mu_{Temp\ Hot}$$

$$H_A: \mu_{Temp\ Warn} \neq \mu_{Temp\ Hot}$$

The p-values for each are all statistically significant (<0.0001 for Cold vs. Warm, <0.0001 for Cold vs. Hot, and 0.0111 for Warm vs. Hot), indicating that we have statistical evidence suggesting that the mean

%Brightness for each temperature is different than the others. These differences are further described by 95% confidence interval estimates of the mean difference between each comparison. For example, a 95% confidence interval (CI) for $(\mu_{Temp\ Cold} - \mu_{Temp\ Hot})$ is [3.28, 5.16], which is the largest difference of the three. In other words, this CI means that one can have 95% confidence that the mean %Brightness at the Cold Temp is between 3.18 to 5.07 percentage units larger than that of the Hot Temp.

The All Pairwise Connecting Letters table is a visual way to display the comparisons that resulted in a statistically significant difference. Factor levels that share a letter are not statistically different. Here, since all the means were determined to be different, none of the Temp levels share the same letter.

The All Pairwise Comparison scatterplot is a visual way to display the results of these comparisons. The points are plotted at the coordinates of each pair of means. For example, the point in the lower right (17.7, 13.5) is the mean value for Cold Temp on the X axis and the mean value for Hot Temp on the Y axis. You can see the label for each by hovering the cursor over the point. The diagonal line represents the place where all of the means would be equal. The confidence interval for each pairwise comparison is shown as the red line extending from each point. If the confidence interval crosses the diagonal line, the pair being compared is not statistically significantly different and is color-coded blue. If the CI does not cross the diagonal, as seen here, the pair being compared is statistically different and is color-coded red. Note that the CI for the comparison between the Cold vs. Hot Temp is furthest from that diagonal line consistent with the CI for that difference being furthest from 0 as we saw earlier. The Warm vs. Hot Temp comparison has a CI closest to the diagonal line. The 95% confidence interval for the difference in those means is [0.305, 2.185] , the closest to 0, and has the largest p-value (0.0111).

Technical note: The p-value for the Warm vs. Hot comparison is greater than 0.01, though just barely. If we had chosen 0.01 as the significant level in the hypothesis test (i.e., 99% confidence level), we would not have concluded a statistically significant difference and the confidence interval displayed in the comparisons scatterplot would have crossed the diagonal line and been color-coded blue. This demonstrates that it is not uncommon in practice to find a significant result at one commonly used significance level but not at another, which is why it's important to be careful not to interpret statistical analysis results as a strict "Yes" or "No" binary decision regarding a hypothesis. Instead it's best to think of it more as a continuum of evidence that supports or does not support a hypothesis at a certain level of confidence.

Exhibit 7 displays the results of the multiple comparisons for Agitation using Student's t method.

Exhibit 7 Multiple Comparison (Student's t)

Multiple Comparisons for Agitation

Student's t All Pairwise Comparisons

All Pairwise Differences

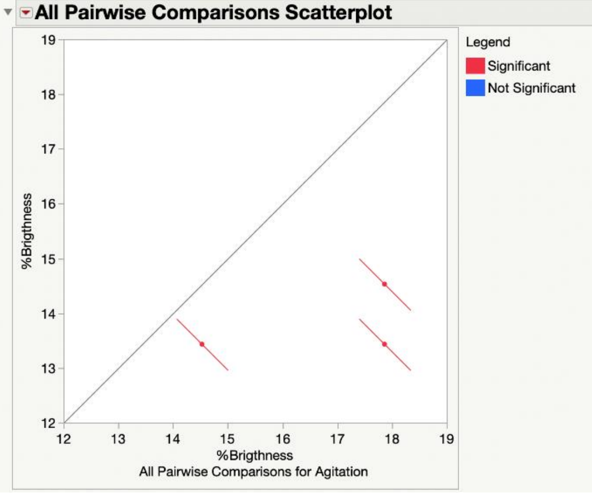
Agitation	-Agitation	Difference	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Low	Med	3.336637	0.4600926	7.25	<.0001*	2.397002	4.276271
Low	High	4.433826	0.4600926	9.64	<.0001*	3.494192	5.373461
Med	High	1.097189	0.4600926	2.38	0.0236*	0.157555	2.036824

All Pairwise Differences Connecting Letters

Agitation				Least Squares Mean
Low	A			17.865576
Med		B		14.528939
High			C	13.431750

Levels not connected by same letter are significantly different.

All Pairwise Comparisons Scatterplot



To create, select Estimates>Multiple Comparison under the red triangle menu at the top of the output. Select the variable Agitation and choose All Pairwise Comparisons – Student's t. Click OK. Then select All Pairwise Differences Connecting Letters under the red triangle next to the Student's t All Pairwise Comparisons heading.

These results (low p-values and CIs that do not cross the diagonal line) demonstrate that there is statistically significant evidence to conclude the mean %Brightness is different across all three levels of Agitation. Similar to the Hot vs. Warm Temp comparison, the statistical test for the Med vs. High Agitation comparison is not significant at the 99% confidence level but is at the 95% confidence level.

ANOVA model

To get a better understanding of the ANOVA model that we've fit to these data, we can view the equation for the model, as shown in Exhibit 8.

Exhibit 8 Predicted Expression

Prediction Expression

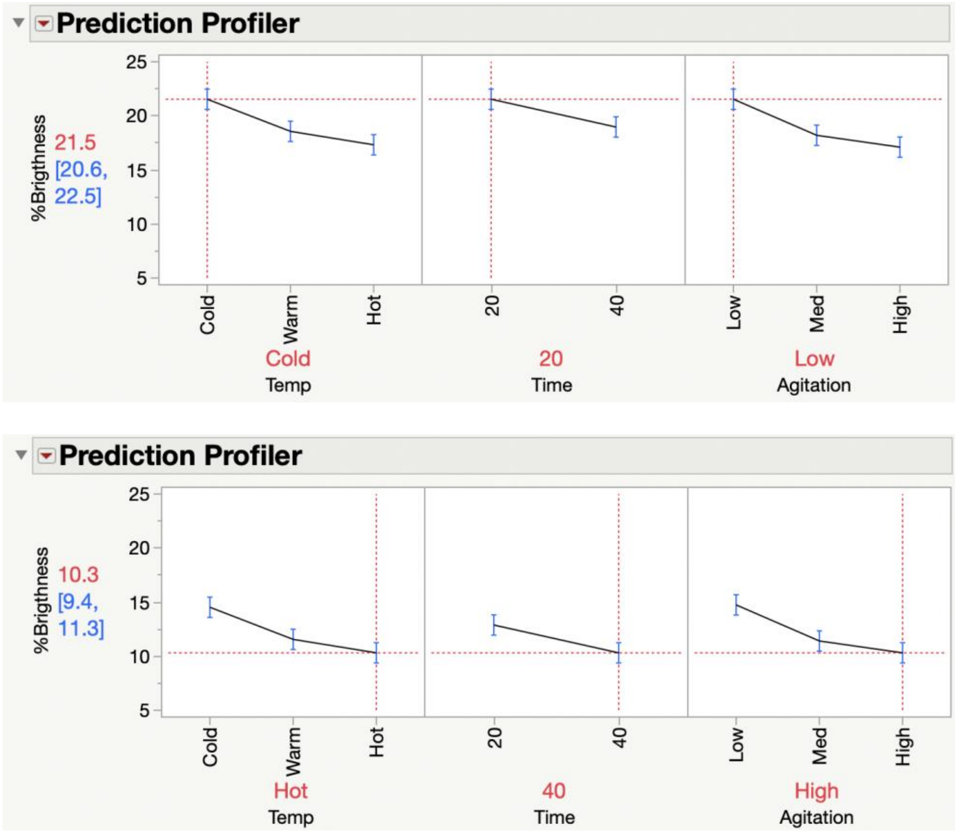
$$
\begin{aligned}
&15.275421658\\
&+\text{Match}\left(\text{Temp}\right)\begin{pmatrix} \text{"Cold"} & \Rightarrow 2.396336492 \\ \text{"Warm"} & \Rightarrow -0.575698179 \\ \text{"Hot"} & \Rightarrow -1.820638313 \\ \text{else} & \Rightarrow . \end{pmatrix}\\
&+\text{Match}\left(\text{Time}\right)\begin{pmatrix} \text{"20"} & \Rightarrow 1.2868758461 \\ \text{"40"} & \Rightarrow -1.286875846 \\ \text{else} & \Rightarrow . \end{pmatrix}\\
&+\text{Match}\left(\text{Agitation}\right)\begin{pmatrix} \text{"Low"} & \Rightarrow 2.5901543323 \\ \text{"Med"} & \Rightarrow -0.746482502 \\ \text{"High"} & \Rightarrow -1.84367183 \\ \text{else} & \Rightarrow . \end{pmatrix}
\end{aligned}
$$

To create, select Estimates>Show Predicted Expression under the red triangle at the top of the output.

The predicted value for each treatment combination is obtained through this equation.

A visual display of the model is typically the best way to "see" what the fitted model is. Exhibit 9 shows that visualization set at the highest predicted response in the first display and at the lowest predicted response in the second.

Exhibit 9 Prediction Profiler



To create, select Factor Profiler>Profiler under the red triangle at the top of the output. Drag the dotted red line to any of the factor settings to see the predicted value along with a 95% confidence interval for the mean response.

The predicted value for the Cold Temp, 20 minutes Time, and Low Agitation is 21.5 (the highest predicted value) with a 95% CI for the mean response of [20.6, 22.5].

The predicted value for the Hot Temp, 20 minutes Time, and High Agitation is 10.3 (the lowest predicted value) with a 95% CI for the mean response of [9.4, 11.3].

It is important to remember that %Brightness, the variable being analyzed, is a measure of the percent increase in brightness between one sub-specimen washed with the new formulation versus the other washed with the current formulation. Our analysis has shown that the Cold Temp, 20 minutes Time, and Low Agitation is the combination of conditions in which the new formulation has the most improvement from the current formulation. These data do not allow us to make a conclusion about which washing conditions resulted in the most effective and least effective cleaning for either of the formulations.

Custom comparisons

It can be informative to make a statistical comparison between specific experimental conditions of interest. For example, an examination of the hypothesis test and a confidence interval estimate for the difference between the two experimental conditions that produced the highest and lowest predicted response is shown in Exhibit 10.

Exhibit 10 Multiple Comparison (Student's t)

Multiple Comparisons for User-Defined Estimates

User-Defined Estimates

Temp	Time	Agitation	Estimate	Std Error	DF	Lower 95%	Upper 95%
Cold	20	Low	21.548788	0.46009262	30	20.609154	22.488423
Hot	40	High	10.324236	0.46009262	30	9.384601	11.263870

Student's t All Pairwise Comparisons

Quantile = 2.04227, DF = 30.0

All Pairwise Differences

Temp	Time	Agitation	-Temp	-Time	-Agitation	Difference	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Cold	20	Low	Hot	40	High	11.22455	0.7513281	14.94	<.0001*	9.690136	12.75897

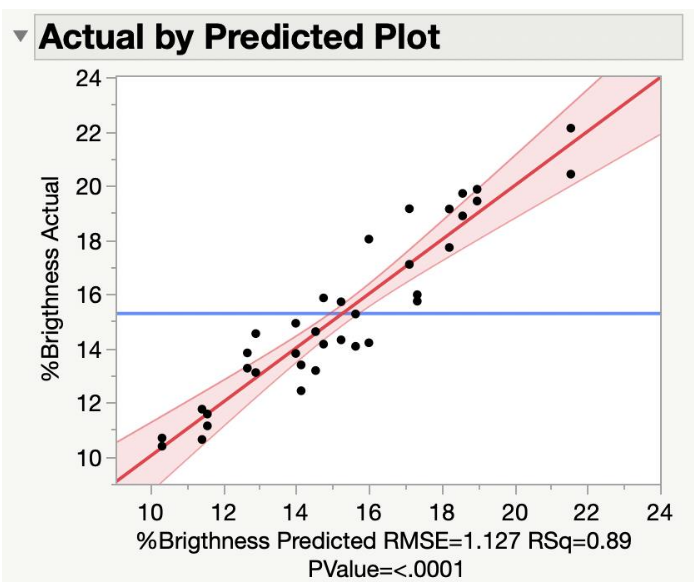
To create, select Estimates>Multiple Comparison under the red triangle at the top of the output. Choose User-Defined Estimates. Select the experimental conditions Temp=Cold, Time=20, Agitation=Low and click Add Estimates. Select the conditions Temp=Hot, Time=40, Agitation=High and click Add Estimate. Select All Pairwise Comparisons –Student's t for the Comparison. Click OK

Here we see that the difference in the mean response between these two experimental conditions is estimated to be 11.22 with a 95% confidence interval estimate for that difference of [9.69, 12.76].

Model performance and diagnostics

When building a statistical model as we've done here, it's important to evaluate how well that model fits the data. Exhibit 11 is one such tool to visualize that fit.

Exhibit 11 Actual by Predicted Plot



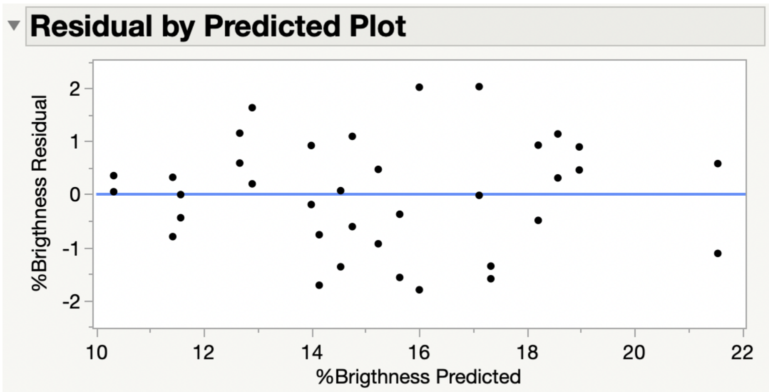
This plot will be displayed if either Effect Screening or Effect Leverage Personality was chosen from within the Fit Model platform. If not displayed, choose Row Diagnostics>Plot Actual by Predicted under the red triangle at the top of the top of the output.

The actual %Brightness values are plotted on the Y axis and the predicted %Brightness values on the X. The red line corresponds to the actual values being equal to the predicted values. The variation around that line provides a visual of the excess variation remaining in the data beyond the fitted model. The R-squared statistic is a numerical measure of how well the fitted model describes the variation in the data.

Here we see that our model accounts for 89% of the total variation in the data with 11% of the variation remaining unexplained.

Exhibit 12 is an alternative way to visualize the variation in the data remaining beyond what the model is able to account for.

Exhibit 12 A Few Words Describing Exhibit



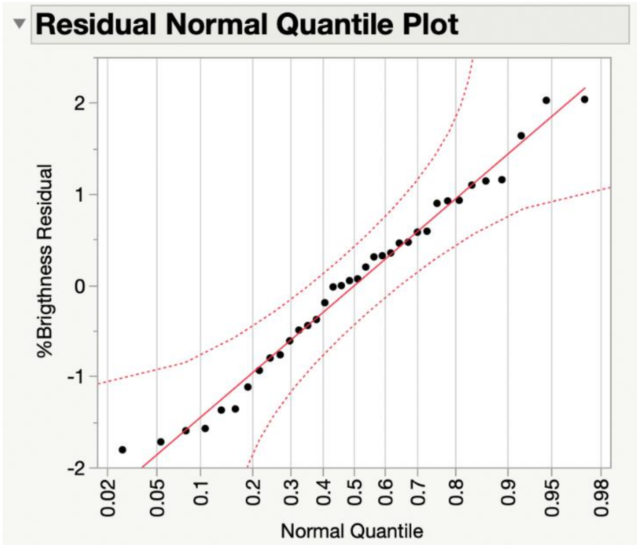
This plot will be displayed if either Effect Screening or Effect Leverage Personality was chosen from within the Fit Model Dialog. If not displayed, choose Row Diagnostics>Plot Residuals by Predicted under the red triangle at the top of the output.

The predicted %Brightness values are plotted on the X axis, and the residuals (Actual %Brightness: Predicted %Brightness) are plotted on the Y axis. Data points that fall on the horizontal line at 0 are observations where the predicted value is the same as the actual value. Data points above the line are observations where the actual %Brightness value is greater than the predicted value, and data points below the line are observations where the actual value is less than the predicted value. This graph is a convenient way to examine the variation around a fitted model and identify unusual observations regardless of how many terms are in a model and/or how complicated the form of the model is.

This graph is also important to consider when building an ANOVA model since it allows us to check one of the assumptions in the inferential techniques we've done thus far. The statistical tests and confidence intervals formed from our analysis are based on the assumption that the variance in the data is very similar across all possible treatments studied (i.e., homogeneity of variance), specifically that a common estimate of experimental variation is used to quantify the uncertainty for all the tests and confidence intervals. If there are substantial differences in variation across different experimental conditions, then some tests and confidence intervals are using an experimental error that is too large, while others are based upon an estimate that is too small. Examining this graph indicates that homogeneity of variance is an appropriate assumption. Alternative techniques exist to conduct these types analyses if homogeneity of variance is an issue. In addition, unequal variation across the different experimental conditions may, in fact, be a very interesting discovery revealing important information about the process under study.

Another assumption in the inferential techniques performed in an ANOVA model is that these residuals can be well modeled by a normal distribution. Exhibit 13 shows a normal quantile plot of the residuals as a way to evaluate this assumption.

Exhibit 13 Normal Quantile Plot of Residuals



To create, select Row Diagnostics>Plot Residuals by Normal Quantiles under the red triangle at the top of the output.

This plot shows that the normality assumption of the residuals is very reasonable.

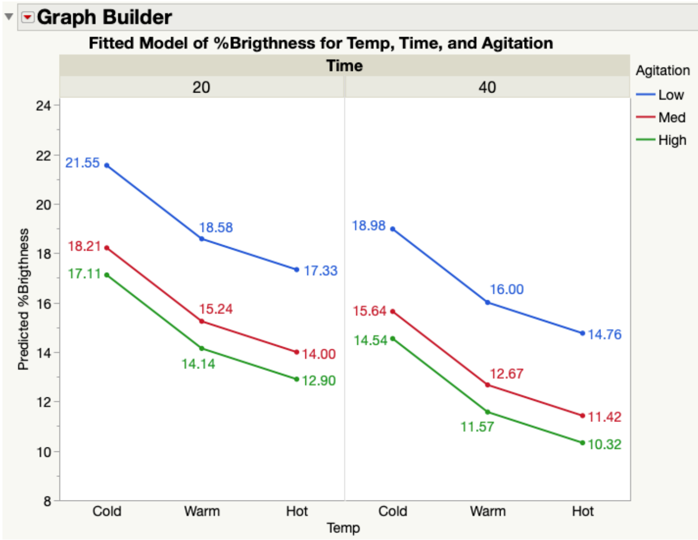
Since the assumptions of normality and homogeneity of a variance in the residuals are appropriate, we have no concern regarding the statistical tests and confidence interval estimates we've obtained earlier.

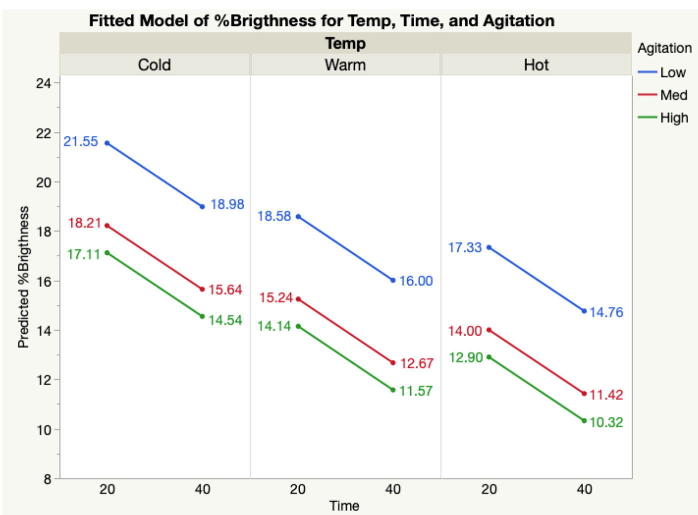
It is important to note, however, that inferences about means, as we're doing in this ANOVA model, are not very sensitive to the normality assumption. It is still important to check, since serious departures from normality may require alternative analysis techniques.

Graphical and numerical description of fitted model

When a statistical model is fit to data from an experiment, as was done here, it can be advantageous to create a visualization of that model. Exhibit 14 provides two such visualizations.

Exhibit 14 Fitted ANOVA Models



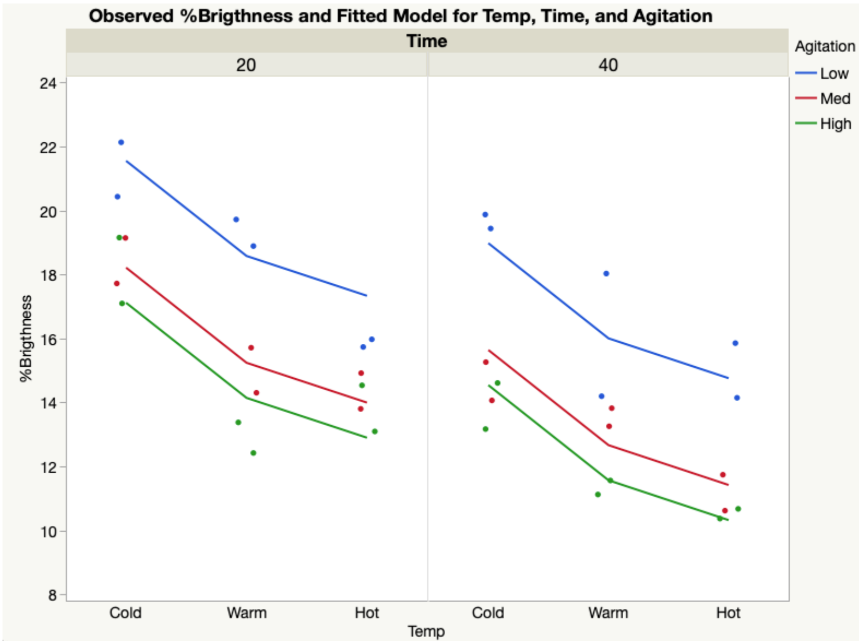


To create, save the Predicted Values from the Model by selecting Save Columns>Predicted Values from the top red triangle. Graph>Graph Builder. Place Predicted %Brightness as the Y variable, Temp as the X variable, Time as the group variable, and Agitation as the overlay variable. Select both Points and Line of Fit in the graph palette. To add the data labels, right-click on the variable and choose Label/Unlabel. Next, highlight all the data in the graph, right-click and choose Rows>Row Label.
Note: Any combination of choosing the roles for the experimental factors (e.g., X variable, group variable, and overlay variable) can be used. A good practice is to choose a layout that best communicates the features you wish to convey. For example, the second graph has Time as the X variable, which makes it easier to visualize the effect of Time on %Brightness.

These visualizations show the effect that each factor has on %Brightness. Notice that the effect of each factor is the same across any levels of the other factors, which is the case since there are no interaction terms in the model. This effect can be verified by calculating the difference between the predicted values for different experimental conditions. For example, the difference between the predicted values between 20 minutes Time and 40 minutes Time is 2.57 for any combination of Temp and Agitation.

Another common approach is to display the fitted model with the observed data, as seen in Exhibit 15.

Exhibit 15 Observed Response with Fitted ANOVA Model



To create, save the Predicted Values from the Model by selecting Save Columns>Predicted Values from the top red triangle. Graph>Graph Builder. Place both %Brightness and Predicted % Brightness as the Y variable, Temp as the X variable, Time as the group variable, and Agitation as the overlay variable. Select both Points and Line of Fit in the graph palette. Points and mean lines will be made for both the Actual %Brightness and Predicted %Brightness values. Right-click on the graph and select Customize. The points for the Predicted values can be removed by choosing transparency of 0 for each Agitation level. Desired colors, markers, and line styles can be chosen.

Note: Any combination of experimental factors to be used as the X variable and the two levels of the group variable can be used. A good practice is to choose a layout that best communicates the features you wish to convey, such as assigning the X and overlay variables to the factors you think will be affected most.

This graph provides a visualization of how the model is describing the observed data. Comparing this graph to Exhibit 2, where the lines plotted are at the sample means for the 18 different experimental conditions, we can see that our chosen model is essentially a simplified description of data, one that reduces the data to an equation that has only three terms to describe it (main effects for Temp, Time, and Agitation). You may note that for some specific experimental conditions, the fitted model is either entirely below or above the observed data. We had conducted statistical tests to determine if a more complicated model – one with interaction terms – was needed, but those tests were not statistically significant. If we conduct more experimental runs, however, we may detect interaction effects if any truly exist. But at this point, we don't have statistical evidence to support it, and our main effects model is a very reasonable description of our experimental results.

Summary

Statistical insights

Our analyses provided quite a bit of statistical evidence demonstrating an improvement in the new formulation. The resulting statistical model that we constructed provides an equation that we used to describe that improvement both visually and quantitatively. We learned that this improvement was not the same across the different washing conditions (Temp, Time, and Agitation). For example, the model helped us identify those conditions where the improvement was the most (Cold Temp, 20 minutes Time, and Low Agitation). For this condition, our predicted %Brightness is 21.5 with a 95% CI for the mean response of [20.6, 22.5]. The conditions we predicted to have the lowest improvement was for the Hot Temp, 20 minutes Time, and High Agitation. Here the predicted %Brightness is 10.3 with a 95% CI of [9.4, 11.3]. Based on these results, we can conclude that a consumer would experience an improvement in %Brightness between 9.4% to 22.5% using the new detergent formulation compared to the current one.

Implications and further study

As in all statistical analyses, it's important to not only identify statistically significant results but to compare these to what would be practically important. For example, if the objective was to develop a formulation where the improvement in %Brightness was at least 15% across all washing conditions, then we can see that was not achieved.

In addition, it is important to remember that the response variable %Brightness is a measure of the percent difference in the Brightness between the two formulations. Thus, we can only describe the difference in Brightness between the two formulations for the different washing conditions. We do not, for example, know from these data and resulting analyses which washing conditions produced the highest or lowest Brightness values for either detergent formulation.

To further explore the effectiveness of this new formulation, the R&D team had also conducted this same experiment using four additional stain types (Coffee, Grass, Grease, and Wine). The following exercises will ask you to build a new ANOVA-based statistical model incorporating the different stain types and then use that model to describe the performance of the new detergent.

Exercises

Use the data in file Cleaning Effectiveness_2.jmp data set to answer the following questions.:

These data contain the variables:

Stain	Stain type (Dirt, Coffee, Grass, Grease, Wine)
Temp	Water temperature (Cold, Warm, Hot)
Time	Washing time (20 minutes, 40 minutes)
Agitation	Level of agitation (Low, Med, High)
Rep	Replicate of each experimental condition (1, 2)
%Brightness	The percent increase in brightness of the sub-specimen washed using the new formulation of the detergent versus the current formulation

1. Create a graph that display the %Brightness for each possible treatment combinations. Describe some of the features the graph reveals.

2. Create a table displaying the mean and range %Brightness for each of the possible treatment combinations.

3. Build a linear model that describes the impact the experimental factors have on the %Brightness:

   a) Start by fitting a model that contains the four mains effects (Temp, Time, Agitation, and Stain) as well as the six possible two-way interactions (Temp*Time, Temp*Agitation, Temp*Stain, Time*Agitation, Time*Stain, Agitation*Stain).

   b) Remove the non-significant two-way interaction terms using a significance level of 0.10.

   c) Examine the tests for any of the main effects that are not included in any significant interaction terms still in the model. Remove any non-significant terms for these main effects.

4. Evaluate the assumptions of the ANOVA model (i.e., homogeneity of variance and normal distribution of the residuals).

5. Using the Prediction Profiler, determine which Stain is estimated to have the greatest improvement in the new formulation versus the current formulation. For which of the experimental conditions in Temp, Time, and Agitation is that improvement the most? Estimate through a 95% confidence interval the average percent increase in %Brightness in the new formulation compared to the current one.

6. Using the Prediction Profiler, determine which Stain is estimated to show the least improvement across the majority of experimental conditions for Temp, Time, and Agitation. Which experimental conditions for that Stain is estimated to have the lowest improvement? With a 95% confidence interval, estimate the average percent increase in %Brightness for that Stain. Are there any other Stains and experimental conditions where there is a comparable low level of improvement?

7. Conduct multiple comparisons to determine if there is any statistical difference between the nine different combinations of Temp and Agitation levels for the Stain identified in Exercise 6.

8. Conduct multiple comparisons across all the combinations of Stain, Temp, Time, and Agitation you identified in Exercise 6 as having comparably low %Brightness values. Describe your findings.

9. Evaluate the performance of the model by examining a plot of the actual versus predicted values and the R-squared statistic. How much of the total variation in the experimental data does the fitted ANOVA model account for?

10. Create a graph that displays both the fitted model as well as the observed %Brightness values. Create a similar graph that only shows the fitted model and includes the predicted values.

11. Provide a one-page executive summary of your conclusions by choosing only one visualization and writing no more than five bullet points. Do not use any statistical terminology. Do you have any recommendations for further study of the effectiveness of the new detergent formulation?