CAKE-BAKING EXPERIMENT

13. Brief Version of the Case Study

13.1 Problem Formulation

In this case study you will be involved in an experiment of baking a peanut butter carrot loaf. The recipe for the cake is described in Section 2. We will use the experiment to demonstrate two-way analysis of variance with three levels of temperature and three levels of baking time.

Baking conditions, such as time and temperature, are important concerns in developing a cake mix that will perform for a wide variety of users. In the recipe for our cake described in Section 2, the instructions say " Bake at 350 degrees F for 60 minutes". Because oven and timer settings vary, the mix must bake satisfactorily at ranges of temperatures and times.

The purpose of the experiment is to assess the impact of different combinations of the levels of temperature and baking time on the taste of the cake. Is the cake better if it is baked in slightly higher temperature or it is baked a little bit longer than stated in the instructions on the package? Is there any interaction between baking temperature and baking time? We will answer all the questions by conducting an experiment.

You will have to design the experiment, collect the data, enter the data into SPSS, carry out the statistical analysis, and formulate your conclusions. The data collection equipment used in the experiment is very simple and relatively cheap.

We will use the experiment to demonstrate two-way analysis of variance with three levels of temperature and three levels of baking time.

13.2   The Experiment Design

The quality of a cake is affected by several factors such as quality of ingredients, baking temperature and time, humidity of the room, taster, and other. The factors are displayed in the diagram below.
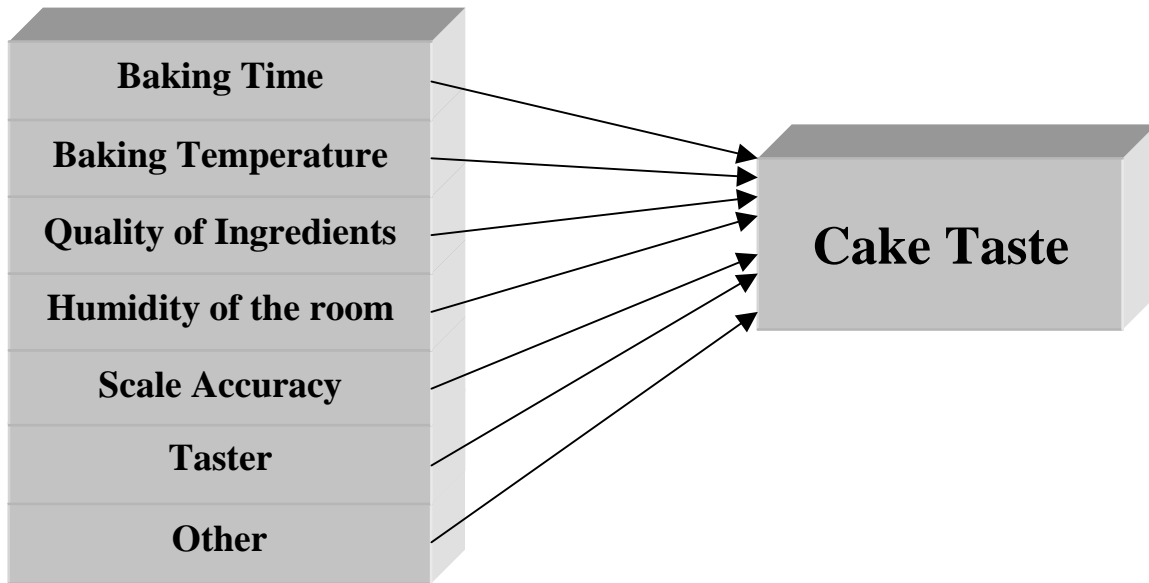
Only some of the factors can be controlled. For example, we cannot control the humidity of the room the cakes are baked. The ingredients also can be varying quality from package to package.

In our experiment we will consider the impact of two of the above factors: baking time and temperature on the taste of a cake made from a mix, introduced in Section 2. The responses are ratings of the taste of the cakes given by tasters.
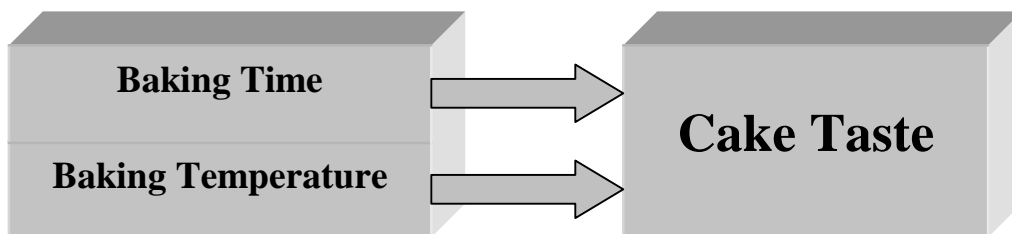
Assume that the range of temperatures to be studied is 300 F to 350 F and the range of times is 55 minutes to 65 minutes. A possible experimental strategy is to study the recommended times and temperatures and the extremes of the ranges. With this strategy, the three temperature levels are to be studied are 325 F, 350 F, and 375 F, and the three time levels to be studied are 55 minutes, 60 minutes, and 65 minutes.

The following table the nine possible combinations of three temperatures and three times in the cake-baking experiment presented in a form of a two-way table:

		Time (in minutes)		
		55	60	65
	325 F			
Temperature	350 F			
	375 F			

The responses are ratings of the taste of the cakes given by tasters. The tasters score the cakes on a seven-point scale, with 0 meaning well below average, 1 below average, 2 somewhat below average, 3 average, 4 somewhat above average, 5 above average, and 6 well above average.

This implies that 9x3=27 tasters and cakes are involved in the experiment. To avoid bias, the randomization process should be used in the experiment.

The process consists of the following steps:

1. Develop a protocol (that is, a detailed list of instructions) for preparing the cake mix and the oven so that the cakes are prepared under essentially identical conditions. The protocol (recipe) is given in Section 2.

2. Prepare 27 identical doughs according to the protocol specified in 1. Assign numerical labels 1, 2, …, 27 to the doughs and mark each of them clearly.

3. Assign randomly the 27 doughs to the nine treatment groups, three doughs in each group. This can be done by creating a deck of 27 cards, three for each of the nine combinations, and laying the well-shuffled cards down in a row or by using the table of random digits.

4. Bake the cakes at combinations chosen in random order. This is done for two reasons. First, if there are effects that carry over systematically from one baking to the next, moving systematically through the table of combinations introduces these carryover effects into the data. Second, the random selection of combinations creates a sound theoretical basis for the use of statistical inference methods in analyzing the data.

   Use an oven and a timer that provide extremely precise and accurate settings. Substantial error in the experimental equipment compromises the precision of the experimental findings.

5. The baked cakes are assigned to tasters at random by arranging the tasters' names in random order and giving one cake to each taster in that order. Offering cakes to the taste testers in a random order allows you to avoid carryover effects from test to test.

Make sure that the people who taste the cakes do not know the experimental conditions under which the cakes were prepared. A person who knows the experimental conditions can transmit clues to the taste tester that can seriously affect the results. The blinding of the conductor as well as the tester is called double blinding.

Recruit taste testers who can truly discriminate differences between cakes. Some people are much better tasters than others.

We will store our data in an SPSS worksheet with the three variables: time, temperature, and rating. These data are available in the SPSS file *cake.sav* located on the FTP server in the Stat337 directory.

The following is a description of the variables in the data file:

| Column	Name of Variable	Description of Variable
1	**TIME**	Baking time level (-1 if 55 minutes, 0 if 60 minutes and +1 if 65 minutes)
2	**TEMP**	Baking temperature (-1 if 325° F, 0 if 350° F, and +1 if 375° F)
3	**RATING**	An integer from 0 to 6

13.3   Describing and Displaying the Data

Before we will apply two-way analysis of variance to our data, we will use graphical summaries and plots to summarize and display the data. You can apply the same tools to your data.

In order to obtain a plot of mean taste scores versus temperature level by baking time in SPSS, click on *Line...* in the *Graphs* menu. In the Line menu choose Multiple as the type of a line chart you need. Fill out the displayed dialog box as follows:
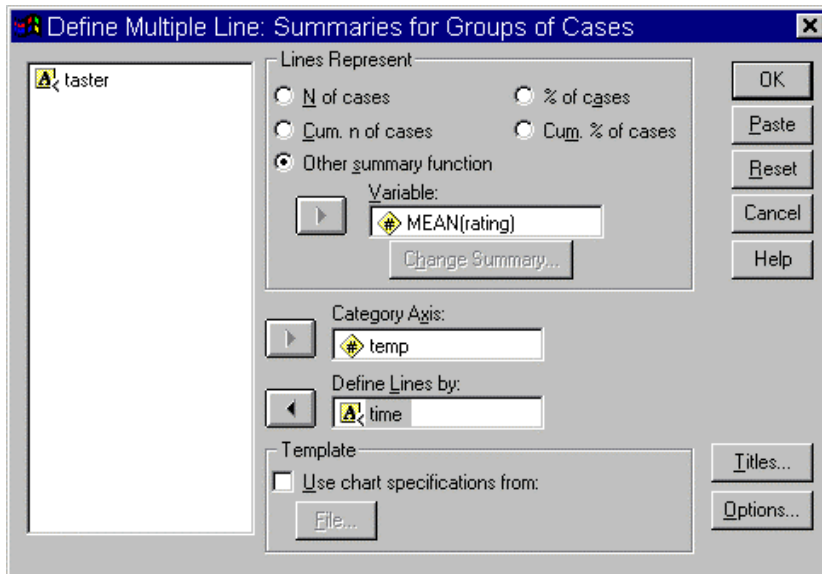

As a result, a plot of mean taste scores versus temperature level by baking time will be obtained. A similar plot can be produced with the GLM General Factorial…procedure.

Mean Ratings vs. Temperature Level by Time Level
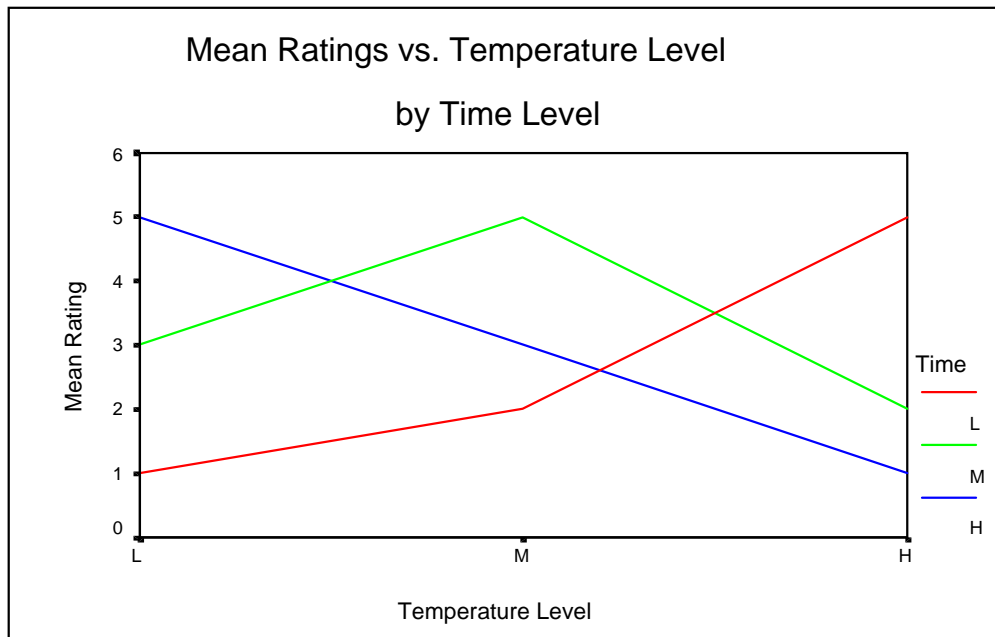
The three levels of temperature denoted in the data file by the numerical labels -1, 0, and +1 have been replaced by the labels L (low), M (medium), and H (high), respectively. The levels of time are plotted similarly, with L denoting the low (-1) level of time, M denoting the medium (0) level of time, and H denoting the high (+1) level of time.

Because the lines in the plot cross each other, there is interaction between time and temperature. The plot shows that the highest mean rating is achieved with the following combinations of time and temperature: (H, L), (M, M), and (L, H). Thus it is possible to compensate for a lower time by using a higher temperature and vice versa. The lowest average rating is achieved with the following combinations of time and temperature: (L, L) and (H, H). In other words, a combination of low time and temperature or high time and temperature produces an unsatisfactory cake.

In order to summarize the data set, we click on Compare Means in the Statistics menu, and then on Means…The displayed dialog box should be filled out as follows:
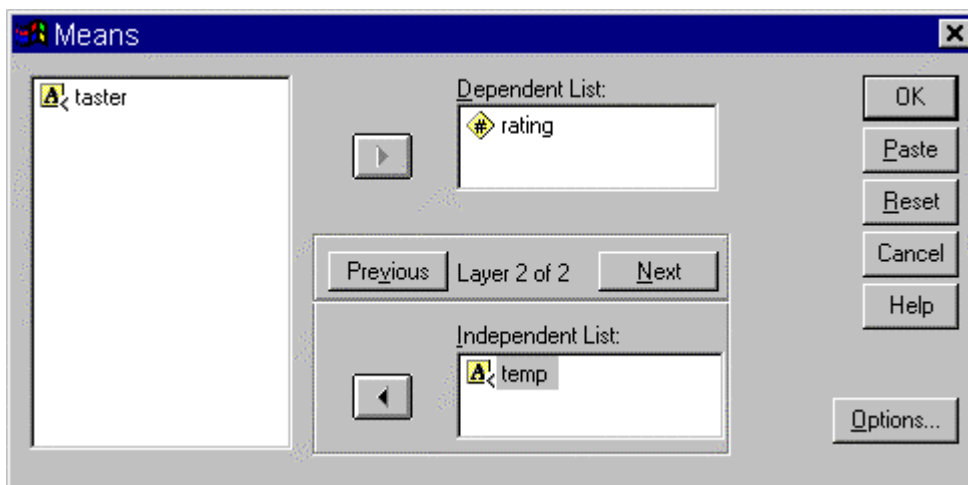
The two layers in the dialog box correspond to the two factors, time and temperature. The following table will be obtained:

Report

RATING

TIME	TEMP	Mean	N	Std. Deviation
-1	-1.00	1.0000	3	1.7321
	.00	2.0000	3	2.0000
	1.00	5.0000	3	1.0000
	Total	2.6667	9	2.2913
0	-1.00	3.0000	3	1.0000
	.00	5.0000	3	1.7321
	1.00	2.0000	3	1.0000
	Total	3.3333	9	1.7321
1	-1.00	5.0000	3	1.0000
	.00	3.0000	3	2.0000
	1.00	1.0000	3	1.0000
	Total	3.0000	9	2.1213
Total	-1.00	3.0000	9	2.0616
	.00	3.3333	9	2.1213
	1.00	2.6667	9	2.0000
	Total	3.0000	27	2.0000

The table shows the means of the three ratings for each combination of levels of the factors. The overall mean is 3, the average of all 27 ratings. The marginal means differ very little from the overall mean of 3, which is why the time and temperature main effects do not seem to be statistically significant. On the other hand, the means vary from a low of 1 to a high of 5.

The table shows that the highest mean rating is achieved with the following combinations of time and temperature: (+1, -1), (0, 0), and (-1, +1). The lowest average rating is achieved with the following combinations of time and temperature: (-1, -1) and (1, +1).

13.4   Two-Way Analysis of Variance

The cake-baking experiment is an example of a factorial experiment. A factorial experiment consists of several factors (baking time, baking temperature) which are set at different levels, and a response variable (taste score). The purpose of the experiment is to assess the impact of different combinations of the levels of baking time and temperature on the taste of the cakes.

The General Factorial Procedure available in SPSS 8.0 provides regression analysis and analysis of variance for one dependent variable by one or more factors or variables. The SPSS data file used for this study is available in the SPSS file *cake1.sav* located on the FTP server in the Stat337 directory. In the data file, variables include time, temperature and taste score. The two-predictor variables in this study, time level and temperature level, are categorical, which means they should be entered as factors in the GLM General Factorial procedure.

Analysis of variance allows us to test the null hypothesis that baking time and baking temperature have no impact on taste score. There are four sources of variation in the experiment: the main effects of *Time* and *Temperature*, the interaction effect, and the error variation. Corresponding to these four sources, there are three null hypotheses that may be tested:

1. H_0: No main effect of *Time*
2. H_0: No main effect of *Temperature*
3. H_0: No interaction effect between *Time* and *Temperature*

The GLM General Factorial procedure in SPSS produces the following output for the experiment:

Tests of Between-Subjects Effects

Dependent Variable: RATING

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	66.000a	8	8.250	3.908	.008
Intercept	243.000	1	243.000	115.105	.000
TIME	2.000	2	1.000	.474	.630
TEMP	2.000	2	1.000	.474	.630
TIME * TEMP	62.000	4	15.500	7.342	.001
Error	38.000	18	2.111		
Total	347.000	27			
Corrected Total	104.000	26			

a. R Squared = .635 (Adjusted R Squared = .472)

The table contains rows for the components of the model that contribute to the variation in the dependent variable. The row labeled *Corrected Model* contains values that can be

attributed to the regression model, aside from the intercept. The sources of variation are identified as *Time*, *Temp*, *Time*Temp* (interaction), and *Error*. *Error* displays the component attributable to the residuals, or the unexplained variation. *Total* shows the sum of squares of all values of the dependent variable. *Corrected Total* (sum of squared deviations from the mean) is the sum of the component due to the model and the component due to the error.

According to the output the model sum of squares is 66.000 and the error sum of squares is 38.000. The total sum of squares (corrected total) is 104.000. Notice a very small contribution of error in the total sum of squares. The p-value of the F-test for the model is reported as 0.008 indicating a sufficient evidence of an effect of at least one of the factors on the taste score.

The sum of squares for the time factor is estimated to be only 2.000, an extremely small value compared to the total sum of squares. The p-value of the F-test is reported as 0.630, indicating no evidence of effect of time on the taste score. Indeed, in all graphical displays and numerical summaries we found no evidence of time effect on the taste score.

The sum of squares due to temperature is also 2.000, a very small contribution in the total sum of squares of 104.000. The value of the F-statistic is 0.474 with the corresponding reported p-value of 0.630. Temperature main effects are not statistically significant.

The p-value of the interaction term *Time*Temperature* is equal to 0.001, indicating a strong evidence of an interaction between the two factors. Thus, in further analysis, the baking time effect should be compared at each level of baking temperature.

To further explore the interaction effects, we examine the profile plot displayed below.
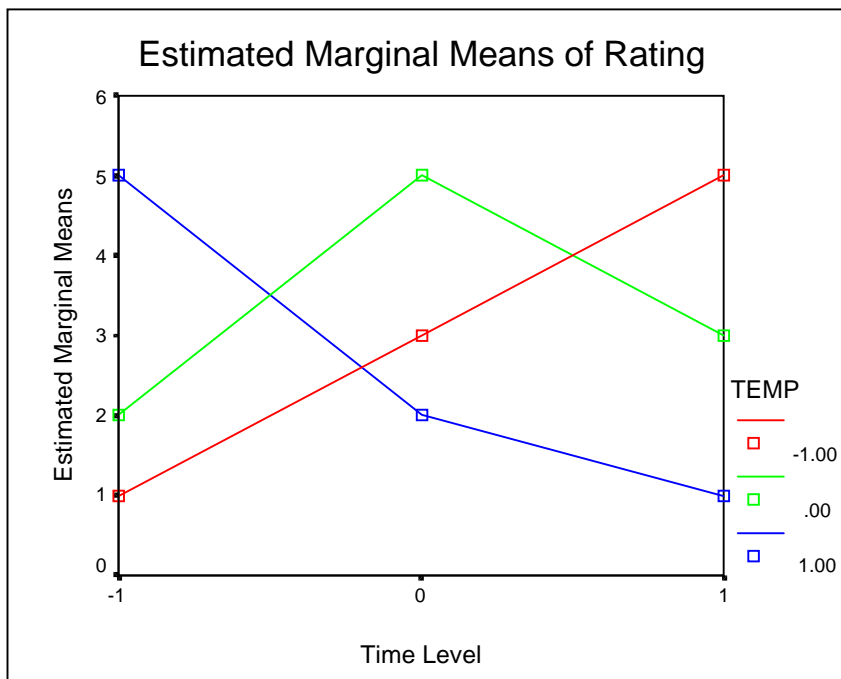
The plot is identical to the plot obtained in Section 13.3. The lines in the above graph cross each other, there is strong interaction between time and temperature. The strongest

interaction effects are shown for the time level -1 (low) and +1 (high) with temperature levels at -1 (low) and +1 (high). This corresponds to the point where the above graph displays the greatest degree of non-additivity.

The plot shows that the highest mean rating is achieved with the following combinations of time and temperature: (+1, -1), (0, 0), and (-1, +1). The lowest average rating is achieved with the following combinations of time and temperature: (-1, -1) and (1, +1). After the interaction between baking time and temperature is found to be significant, it is possible to discuss custom contrasts to find out which combinations of levels of the factors are different. The contrasts are discussed in Section 8.4.

13.5   Summary

The goal of the experiment is to study the impact of baking time and temperature on the taste of a cake made from a mix, introduced in Section 2. The factors are baking time and temperature. The responses are ratings of the cakes given by selected tasters.

In this experiment, three times and three temperatures are used, so it is a 3x3 factorial experiment, meaning two factors at three levels each. There are nine combinations of the levels of the factors, and because three taste testers evaluate cakes baked at a given combination, there are three replications. Overall, 9*3=27 tasters and cakes are involved in the experiment. To avoid bias, the 27 cakes are baked according to a random allocation of combinations.

The cake-baking experiment is an example of a factorial experiment. The GLM General Factorial Model in SPSS was used to analyze the data. The main effects of time and temperature were found to be statistically not significant. This should not be interpreted to mean that these factors are unimportant. The effects are small for the amount of deviation from the considered levels. Certainly it is possible to think of deviations that would produce stronger effects.

The analysis showed very strong interaction between time and temperature. The profile plots and numerical summaries show that the highest mean rating is achieved with the following combinations of time and temperature: (H, L), (M, M), and (L, H), where L, M, and H denote low, medium, and high level of a factor. Thus it is possible to compensate for a lower time by using a higher temperature and vice versa. The lowest average rating is achieved with the following combinations of time and temperature: (L, L) and (H, H). In other words, a combination of low time and temperature or high time and temperature produces an unsatisfactory cake.