PLANT-GROWTH EXPERIMENT

15. Brief Version of the Case Study

15.1   Problem Formulation

In the following study, you will be involved in the experiment of growing a plant of your choice. The experiment is designed such that the data can be collected with reasonably inexpensive measuring equipment. The data will be analyzed using SPSS. The experiment can be carried out in a team.

Your task is to examine and estimate the effects of seed type and amount of water on the growth of a particular type of plant. 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 plant-growth experiment will be performed in several versions, and you will see that the statistical model used depends heavily on how the experiment was carried out. In particular, we will consider the experiment when data for some of the factor-level combinations are not available (the plants did not shoot up). Moreover, various outcomes of the experiment will be discussed.

You will need 20 small flowerpots with identification tags, potting soil, seeds for three varieties of the same vegetable, meter stick, planting trowel, 1/4-cup measure, and a bucket of water.

After the data are collected, you will answer the following questions:

1.   What is the effect of seed variety on the plant growth? Does it appear that plants obtained from one seed variety tend to grow faster than the other plants regardless of the watering system used?

2.   Which watering system contributes the most to the growth of the plants?

3.   Is there any interaction between seed variety and watering system in their effect on the plant growth? How strong is the interaction?

15.2   Experiment Design

Plant growth is affected by several factors such as seed variety, amount of water, soil type, amount of light, temperature, humidity, and other. The factors are displayed in the diagram below.
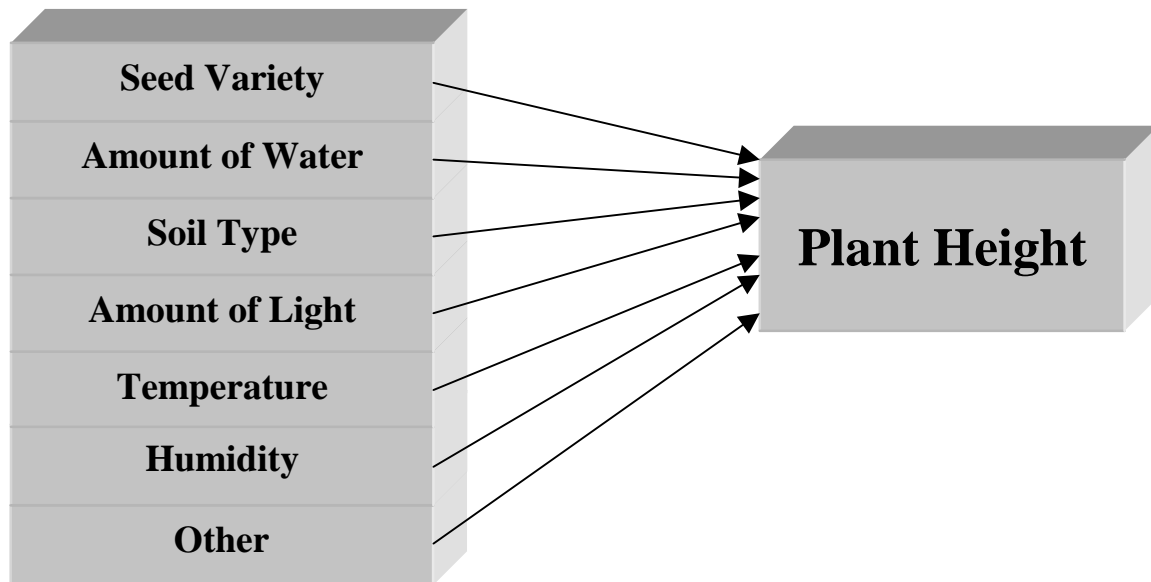
You will use two variables in the experiment: seed variety and amount of water. These variables are the factors in the experiment.
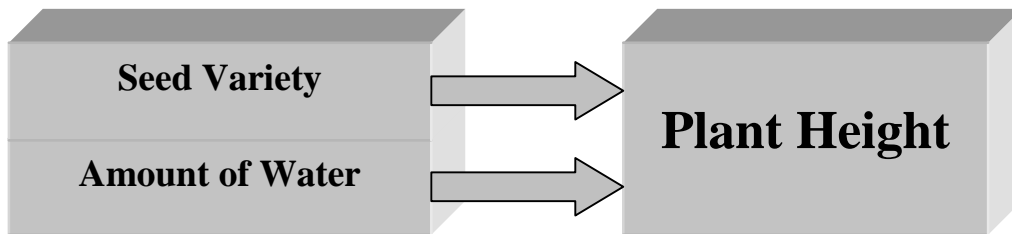
The seed factor will be three varieties of the same type of plant, and its levels will be denoted by 1, 2, and 3. Write the names of these three varieties in the table below.

Factor	Level	Description
Seed	1	
	2	
	3	
Water	1	
	2	
	3	
	4	

Now you should decide on four reasonable watering plans. An example of possible water-factor levels is:

Level 1 of water is 1/3 cup once a week.
Level 2 of water is 1/2 cup once a week.
Level 3 of water is 1/3 cup twice a week.
Level 4 of water is 1/2 cup twice a week.

Write your four levels of the water factor in the above table.

This experiment has two factors: seed (with three levels) and the amount of water (with four levels). This creates 3 x 4 = 12 factor-level combinations, as represented by the cells in the following table. The twelve combinations will be denoted by the letters A,..., L.

Symbol	Levels	General Description	Detailed Description
A	(1,1)		Seed variety 1 and water level 1
B	(1,2)		Seed variety 1 and water level 2
C	(1,3)		Seed variety 1 and water level 3
D	(1,4)		Seed variety 1 and water level 4
E	(2,1)		Seed variety 2 and water level 1
F	(2,2)		Seed variety 2 and water level 2
G	(2,3)		Seed variety 2 and water level 3
H	(2,4)		Seed variety 2 and water level 4
I	(3,1)		Seed variety 3 and water level 1
J	(3,2)		Seed variety 3 and water level 2
K	(3,3)		Seed variety 3 and water level 3
L	(3,4)		Seed variety 3 and water level 4

Your task will be to determine which of these twelve combinations produces the largest plants. As 24 pots are available, we shall have 2 pots for each treatment combination.

This is an example of a two-factor experiment with replication.

Seed type and water level are two factors affecting the growth of the plants that can be controlled. However, some of the other factors cannot be controlled. For example, we cannot control the amount of light coming through the window. The ideal situation would be for all 24 plants to receive the same amount of light, so any differences in plant growth will be due to the two controlled factors of water and seed variety.

To minimize the effect of uncontrollable factors, it is very important that the levels of the factors are assigned at random to the experimental units, the pots, in the study. Randomization is a technique for assigning treatment combinations to experimental units (in this case, pots). We will use randomization to decide the arrangement of the 24 pots in the window. Randomization gives each of the 24 pots an equal chance to be chosen to each of the 12 treatments.

In the first step of randomization, it is necessary to assign labels to the experimental units. Two digits are needed to label each of the 24 pots, so we use labels 01, 02, 03, ..., 24.

We assign the 24 pots to the twelve treatments, so that each treatment combination will be assigned randomly to exactly two pots.

Treatment Combination	Treatment Number
A (1,1)	1
B (1,2)	2
C (1,3)	3
D (1,4)	4
E (2,1)	5
F (2,2)	6
G (2,3)	7
H (2,4)	8
I (3,1)	9
J (3,2)	10
K (3,3)	11
L (3,4)	12

Now obtain a long sequence of random numbers between 1 and 12. The sequence can be obtained either from the table of random numbers (numbers different from the integers between 1 and 12 are disregarded) or by random number generation feature in the statistical software (integer uniform distribution with possible values between 1 and 12).

The first number in the sequence will assign a treatment combination to the first pot. Continue using the numbers until all the 24 pots have been assigned a treatment combination. Remember that each treatment is to be used only twice. That is, after two 1s have appeared, skip over the remaining 1s. Use the figure below to record the treatment-combination assignments as they are determined.

| Pot Number	Treatment Number	Random Numbers
1	7	7
2	4	4
3	11	11
4	8	8
5	10	10
6		
22	1	1
23	12	12
24	8	8

When you are finished, you should have 24 pots labelled as two 1s (A s), two 2s (B s),…,two 12s (L s). The random assignment of the experimental units (24 pots) to the treatment combinations is now complete.

15.3 Data Collection

We are now ready to plant the seeds. For each pot, make sure the proper treatment combination is being used. Label each pot with an identification tag that indicates the seed type and watering plan. Also, to minimize the effect of the uncontrollable factors, make the amount of soil and the position of the seed with regard to depth and distance from edge of pot as consistent as possible. Use a balance the measure the amount of soil used and meter stick to measure depth and distance.

Set the pots in their locations. You should make sure that all 24 plants receive approximately the same amount of light.

Water your plants according to the treatment combination assigned. Do not deviate from your set watering schemes, even if the plants do not appear healthy.

Straighten out the plant before the height measurement is taken
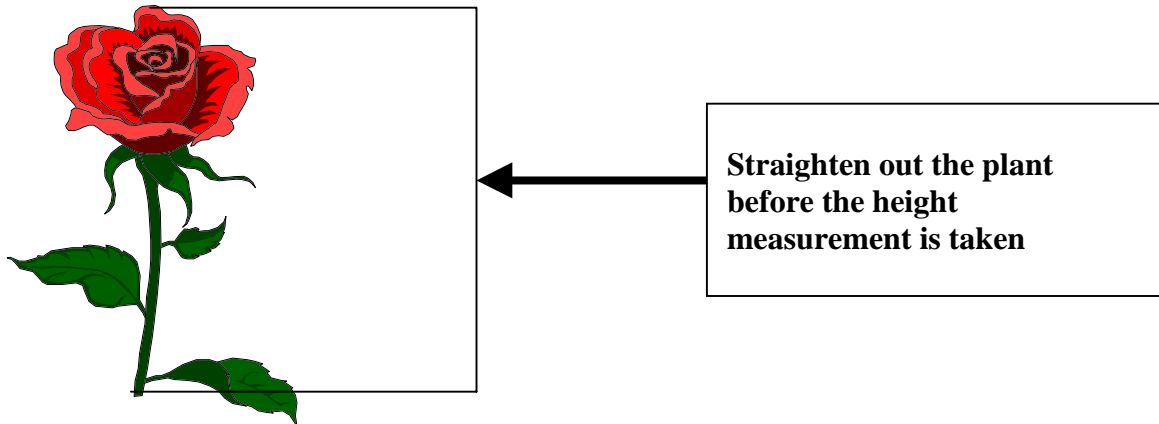
Think about how height will be measured and stick with that rule. For example, if your plant is of the droopy variety, then you might choose to straighten out the plant before the height measurement is taken. If so, you should do this every time the height is measured.

Our collected data in millimeters are displayed in the table below:

HEIGHT		WATERING PLAN			
		1	2	3	4
SEED	1	35	38	41	45
		37	38	39	43
	2	31	39	44	47
		33	37	40	45
	3	38	34	39	46
		38	36	37	44

We will store our data in an SPSS worksheet with the three variables: seed, water, and height. These data are available in the SPSS file *plant1.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	SEED	Seed Variety (an integer from 1 to 3)
2	WATER	Water level (an integer from 1 to 4)
3	HEIGHT	Height of plant (in millimeters)

15.4   Displaying Data

We will visualize the effects of seed type and watering plan on the plant growth by obtaining the plot of factor-level means versus watering plan by seed type.

SPSS produces the following line chart of mean height versus watering plan by seed type:
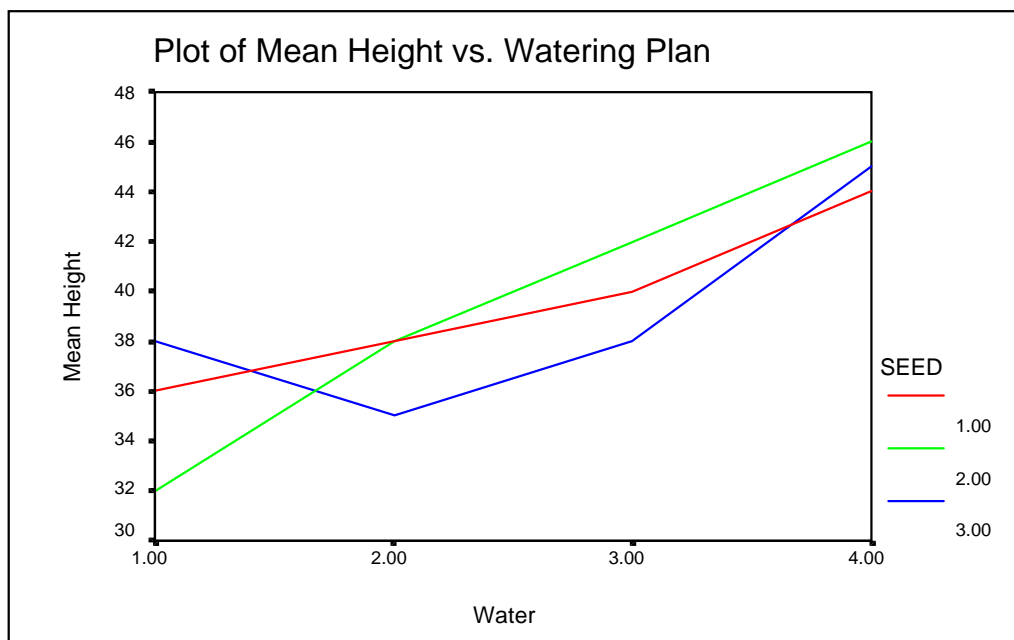
The lines in the graph are obtained by connecting the factor-level means for the four water levels. The plot indicates that the mean height increases with the water level for the seed type 1 and 2. On the other hand, the mean height decreased slightly from water level 1 to level 2 and increased sharply at water levels 2 and 3. The growth rate is uneven for different combinations of seed type and amount of water. The graph indicates a lack of additivity (interaction) between the means for the different seeds when taken across the water levels.

The seed 2 produces the shortest plants under the watering plan 1 but it surpasses the other seeds under the watering plans 2, 3, and 4. It looks that the plants obtained from the seed need more water than the other plants.

The seed 3 does not perform well under the watering plan 2 and 3, somehow increasing the frequency of watering and amount of water does not produce higher plants in this case.

The strongest interaction effect is shown for the water level 1 with seeds 2 and 3. This corresponds to the point where the above graph displays the greatest degree of non-additivity.

15.5   Two-Way Analysis of Variance

The plant-growth experiment is an example of a factorial experiment. A factorial experiment consists of several factors (seed, water) which are set at different levels, and a response variable (plant height). The purpose of the experiment is to assess the impact of different combinations of the levels of seed and water on plant height. Analysis of variance allows us to test the null hypothesis that seed and water have no impact on plant height.

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 *plant1.sav* located on the FTP server in the Stat337 directory. To produce the output for your data, select *SPSS Instructions* in the problem menu now. Here, we will display and analyze the output for our data.

Analysis of variance allows us to test the null hypothesis that seed and water have no impact on plant height. There are four sources of variation in the experiment: the main effects of *Seed* and *Water*, 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 *Seed*
2. H_0: No main effect of *Water*
3. H_0: No interaction effect between *Seed* and *Water*

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

Tests of Between-Subjects Effects

Dependent Variable: HEIGHT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	393.333ᵃ	11	35.758	16.503	.000
Intercept	37130.667	1	37130.667	17137.231	.000
SEED	1.333	2	.667	.308	.741
WATER	324.000	3	108.000	49.846	.000
SEED * WATER	68.000	6	11.333	5.231	.007
Error	26.000	12	2.167		
Total	37550.000	24			
Corrected Total	419.333	23			

a. R Squared = .938 (Adjusted R Squared = .881)

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 *Seed*, *Water*, *Seed*Water*, 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 393.333 and the error sum of squares is 26.000. The total sum of squares (corrected total) is 419.333. Notice a very small contribution of error term in the total sum of squares. The p-value of the F-test for the model is reported as zero indicating a sufficient evidence of an effect of at least one of the factors on the plant height.

The sum of squares for the seed factor is estimated to be only 1.333, an extremely small value compared to the total sum of squares. The p-value of the F-test is reported as 0.741, indicating a sufficient evidence of no effect of seed type on the plant height. Indeed, in all graphical displays and numerical summaries we found the plant seed not affected by the seed type.

The sum of squares due to water is 324.000, a very large contribution in the total sum of squares of 419.333. The value of the F-statistic is 49.846 with the corresponding reported p-value of zero. Water main effects are highly statistically significant.

The p-value of the interaction term Seed*Water is equal to 0.007, indicating a strong evidence of an interaction between the two factors. Thus, in further analysis, the water effect should be compared at each level of seed.

To further explore the interaction effects, we examine the table of estimated marginal means and the profile plot of the same values displayed below.

SEED * WATER

Dependent Variable: HEIGHT

| SEED | WATER | Mean | Std. Error | 95% Confidence Interval | |
				Lower Bound	Upper Bound
1.00	1.00	36.000	1.041	33.732	38.268
	2.00	38.000	1.041	35.732	40.268
	3.00	40.000	1.041	37.732	42.268
	4.00	44.000	1.041	41.732	46.268
2.00	1.00	32.000	1.041	29.732	34.268
	2.00	38.000	1.041	35.732	40.268
	3.00	42.000	1.041	39.732	44.268
	4.00	46.000	1.041	43.732	48.268
3.00	1.00	38.000	1.041	35.732	40.268
	2.00	35.000	1.041	32.732	37.268
	3.00	38.000	1.041	35.732	40.268
	4.00	45.000	1.041	42.732	47.268

The combination of seed 2 and water level 4 produces the highest plants. The combination of seed 2 and water level 1 produces the lowest plants. The pooled estimate of the standard deviation is 1.041.

Now we examine the interaction effects with a profile plot. A profile plot is a line plot in which each point indicates the estimated marginal mean of a dependent variable at one level of a factor. The plot for our data is displayed below.
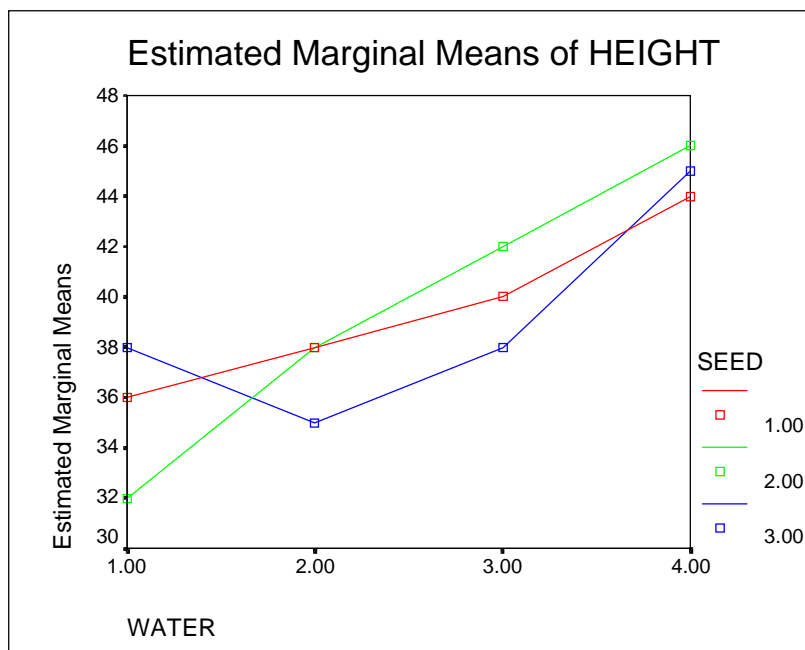
The above graph has already been displayed and discussed in Section 15.4. The graph indicates a lack of additivity (interaction) between the means for the different seeds when taken across the water levels.

The strongest interaction effect is shown for the water level 1 with seeds 2 and 3. This corresponds to the point where the above graph displays the greatest degree of non-additivity.

Since the hypothesis for main effects for water was strongly rejected, multiple comparisons might be considered for the means of the levels of the factor. The following SPSS output obtained with the GLM model shows the results of the multiple comparisons with the Tukey HSD procedure.

Multiple Comparisons

Dependent Variable: HEIGHT
Tukey HSD

(I) WATER	(J) WATER	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-1.6667	.850	.255	-4.1898	.8564
	3.00	-4.6667*	.850	.001	-7.1898	-2.1436
	4.00	-9.6667*	.850	.000	-12.1898	-7.1436
2.00	1.00	1.6667	.850	.255	-.8564	4.1898
	3.00	-3.0000*	.850	.019	-5.5231	-.4769
	4.00	-8.0000*	.850	.000	-10.5231	-5.4769
3.00	1.00	4.6667*	.850	.001	2.1436	7.1898
	2.00	3.0000*	.850	.019	.4769	5.5231
	4.00	-5.0000*	.850	.000	-7.5231	-2.4769
4.00	1.00	9.6667*	.850	.000	7.1436	12.1898
	2.00	8.0000*	.850	.000	5.4769	10.5231
	3.00	5.0000*	.850	.000	2.4769	7.5231

Based on observed means.
*. The mean difference is significant at the .05 level.

The above table shows significant differences in water main effects on the plant height across the water levels. As you can see, there are significant differences in the water main effects for any pair of water levels except for the levels 1 and 2.

15.6   Summary

The plant-growth experiment is an example of a factorial experiment. A factorial experiment consists of several factors (seed, water) which are set at different levels, and a response variable (plant height). The purpose of the experiment is to assess the impact of different combinations of the levels of seed and water on plant height. Analysis of variance allows us to test the null hypothesis that seed and water have no impact on plant height.

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 p-value of the F-test for the model is reported as zero indicating a sufficient evidence of an effect of at least one of the factors on the plant height.

The p-value of the F-test for the seed factor is reported as 0.741, indicating a sufficient evidence of no effect of seed type on the plant height. Indeed, in all graphical displays and numerical summaries we found the plant seed not affected by the seed type. The p-value of the F-test for the water factor is reported as zero. Thus, water main effects are highly statistically significant. The p-value of the interaction term Seed*Water is equal to 0.007, indicating a strong evidence of an interaction between the two factors.

 The combination of seed 2 and water level 4 produces the highest plants. The combination of seed 2 and water level 1 produces the lowest plants.