PLANT-GROWTH EXPERIMENT

9. Two-Way ANOVA with Missing Data

How is the height of plants affected by different combinations of water and seed when data for some of the factor-level combinations are not available? In the previous section, we studied the effect of the combining three different seed types with each of the four water levels on the height of plants. All combinations of seed type and water levels were included there.

Now suppose that some plants did not shoot up. What is even worse, the plants that did not shoot up come from the same level-factor combination. More precisely, all data in some level-factor combinations are missing. Assume that the data has the following form:

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

There are empty cells on the data. There is no data for the seed level 2 and water plan 1, (31 and 33 before) and for the seed level 3 and the water plan 4 (46 and 44 before). They are treated as missing data in this case study. The new data are saved in the SPSS file *plant4.sav*.

How will the tests for the presence of the main and interaction effects be affected with the new data? Before we use SPSS, let us try to determine how the changes in the data might effect the tests. The removed observations in the row 2 are the smallest observations for seed 2, and the removed observations in the row 3 are the largest observations for seed 3. Thus the change will make the differences among the average heights of the three seeds larger. As a consequence, there is stronger support for rejecting the null hypothesis of no main effect due to seed type, and therefore we expect smaller p-value for the test with the new data. The p-value for the full data was 0.741.

However, the Type III sum-of squares method is no longer appropriate in this case because there are empty cells in the design. In particular, the tests hypotheses for Type III sum of squares are not suitable for this set of data; they may not have meaningful interpretations. The Type IV sum of squares method is designed to construct test hypotheses that are meaningful in the presence of empty cells in the model. Type IV hypotheses compare the levels of one treatment by averaging over one or more common levels of the other treatments.

The GLM General Factorial Procedure in SPSS applied to the data with the Type IV sum of squares produces the following output.

Tests of Between-Subjects Effects

Dependent Variable: HEIGHT

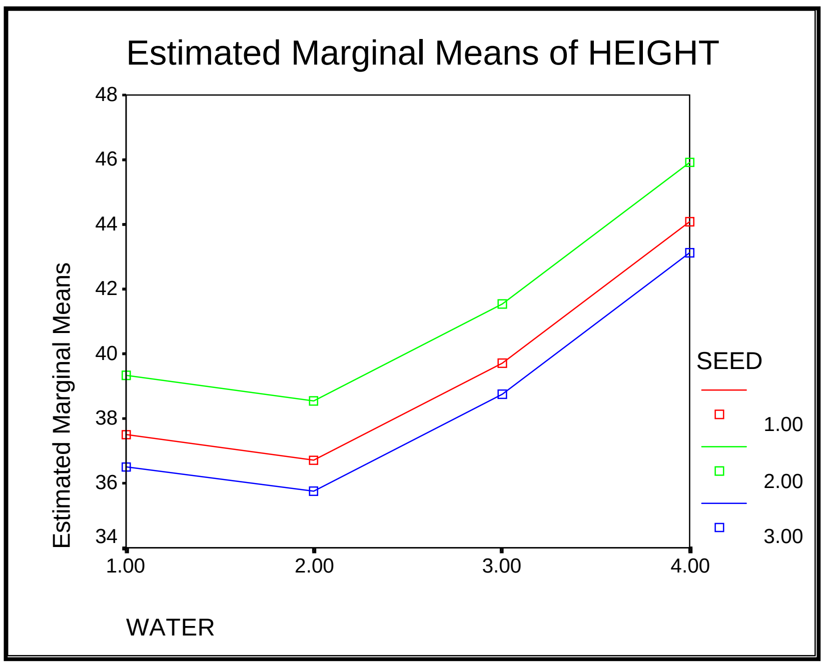
Source	Type IV Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	221.000[a]	9	24.556	11.162	.000
Intercept	31205.000	1	31205.000	14184.09	.000
SEED	24.600[b]	2	12.300	5.591	.023
WATER	121.200[b]	3	40.400	18.364	.000
SEED * WATER	15.543	4	3.886	1.766	.212
Error	22.000	10	2.200		
Total	31448.000	20			
Corrected Total	243.000	19			

a. R Squared = .909 (Adjusted R Squared = .828)

b. The Type IV testable hypothesis is not unqiue.

As you can see, the p-value of the test for equal main effects due to seed has decreased to 0.048. This is consistent with the above analysis. The p-value of the test of no interaction effect between seed and water is 0.212 indicating the presence of additivity (no interaction) in the model.

The profile plot shown below is consistent with the above conclusions.



The parallelism in this chart indicates that there is little or no interaction between seed variety and water. The plot shows also very strong main effects.