PLANT-GROWTH EXPERIMENT

11. Using Regression to Estimate the Factor Effects

The analysis in the previous section, using the general linear model, can be performed with a regression analysis. In this section we will use the regression analysis to estimate the effects of seed and water on the height of the plants.

The GLM model in the regression setting has the form:

$$(HEIGHT)_{ijk} = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 D_5 + \beta_6 D_1 D_4 + \beta_7 D_1 D_5 + \beta_8 D_2 D_4 + \\ + \beta_9 D_2 D_5 + \beta_{10} D_3 D_4 + \beta_{11} D_3 D_5 + (ERROR)_{ijk},$$

where $(\mathrm{HEIGHT})_{ijk}$ is the *k*th plant height observed for the *i*th level of water and *j*th level of seed type.

Three dummy variables, D_1, D_2, and D_3 are used to represent the four levels for water. The dummy variables are defined as follows: $D_i = 1$ when $(\mathrm{HEIGHT})_{ijk}$ is from level i+1 of the factor (i=1, 2, 3), and $D_i=0$ otherwise.

The dummy variables, D_4 and D_5 are used to represent the three levels of seed variety. Likewise for seed type: $D_j=1$ when $(\mathrm{HEIGHT})_{ijk}$ is from level j+1 of the factor (j=1, 2), and $D_j=0$ otherwise.

The data for the plant-growth experiment with the five dummy variables are stored in the SPSS file *plant2.sav* located on the FTP server in the Stat337 directory.

Consider the following null hypothesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11}.$$

This hypothesis is equivalent to saying that there is no difference in the means of the various factor-level combinations.

The regression output for the data in the file *plant2.sav* is displayed below.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	393.333	11	35.758	16.503	.000[a]
	Residual	26.000	12	2.167		
	Total	419.333	23			

a. Predictors: (Constant), D3D5, D3D4, D2D5, D2D4, D1D5, D1D4, D3, D2, D1, D5, D4

b. Dependent Variable: HEIGHT

As you can see the regression sum of squares of 393.333 is identical with the model sum of squares reported in the ANOVA output. The sum of squares includes the sum of squares for seed variety, water, and the interaction term seed*water. The analysis gives F=16.503 with a p-value of 0.000, which provides a strong rejection of the null hypothesis.

According to the table displayed below, the regression equation is

$$HEIGHT = 36.0 + 2D_1 + 4D_2 + 8D_3 - 4D_4 + 2D_5 + 4D_1D_4 - 5D_1D_5 + 6D_2D_4 - 4D_2D_5 + 6D_3D_4 - D_3D_5.$$

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	36.000	1.041		34.588	.000
	D1	2.000	1.472	.207	1.359	.199
	D2	4.000	1.472	.414	2.717	.019
	D3	8.000	1.472	.829	5.435	.000
	D4	-4.000	1.472	-.451	-2.717	.019
	D5	2.000	1.472	.226	1.359	.199
	D1D4	4.000	2.082	.264	1.922	.079
	D1D5	-5.000	2.082	-.331	-2.402	.033
	D2D4	6.000	2.082	.397	2.882	.014
	D2D5	-4.000	2.082	-.264	-1.922	.079
	D3D4	6.000	2.082	.397	2.882	.014
	D3D5	-1.000	2.082	-.066	-.480	.640

a. Dependent Variable: HEIGHT

The residual statistics is summarized in the following table:

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	32.0000	46.0000	39.3333	4.1354	24
Residual	-2.0000	2.0000	-3.6E-15	1.0632	24
Std. Predicted Value	-1.773	1.612	.000	1.000	24
Std. Residual	-1.359	1.359	.000	.722	24

a. Dependent Variable: HEIGHT