CLOUD SEEDING EXPERIMENT

9. The Effect of Outliers on the Inferences

We defined outliers as the observations that are more than 1.5IQRs from the box in the boxplot. SPSS defines outliers as cases with values between 1.5 and 3 box-lengths from the upper or lower edge of the box and extremes as cases with values more than 3 box-lengths from the upper or lower edge of the box.

In Section 4.1 we found that there are some outliers in both distributions in our data. In the following table, we summarize the information.

TREATMENT	OUTLIERS AND EXTREMES	
	OBSERVATION NUMBER	VALUE
UNSEEDED	3	372.40
	2	830.10
	1	1202.60
SEEDED	30	978.00
	29	1656.00
	28	1697.80
	27	2745.60

What is the effect of the observations on the t-test and confidence intervals discussed in the previous section? In order to answer the question, we will carry out the statistical analysis without the outlying observations and compare the results with those obtained in Section 8.

t-tests for Independent Samples without outliers and extremes (log scale)

Variable	Number of Cases	Mean	SD	SE of Mean
RAINFALL				
Unseeded	23	3.6529	1.414	0.295
Seeded	22	4.7198	1.361	0.290

Mean Difference = -1.0669
Levene's Test for Equality of Variances: F= .003 P= .959

t-test for Equality of Means

Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	-2.58	43	.013	.414	(-1.902, -.232)
Unequal	-2.58	43	.013	.414	(-1.901, -.233)

As you see, the value of the t-statistic has changed slightly. It was -2.54, it is -2.58 without the outlying observations. The p-value of the test is almost the same, it was 0.007 for the one-sided test, it is now 0.0013/2=0.0065. Moreover, the 95% confidence interval for the multiplicative effect has changed slightly, it was (-2.047, -.241) before, it is now (-1.902, -.232). The presence of outliers did not affect the final conclusions.