CLOUD SEEDING EXPERIMENT

14. Brief Version of the Case Study

14.1 Problem Formulation

The problem is based on the data from the *Cloud Seeding to Increase Rainfall* experiment discussed in your textbook, pages 54-55. This data are also available in the Excel file Case0301.xls located on the FTP server.

The experimental data are the result of a series of weather modification experiments conducted in south Florida from 1968 to 1972. These experiments were designed to test a hypothesis that massive injection of silver iodide into cumulus clouds can, under specified conditions, lead to cumulus growth, thereby increased precipitation. On each of 52 days that were deemed suitable for cloud seeding, a random mechanism was used to decide whether to seed the target cloud on that day or to leave it unseeded as a control. An airplane flew through the cloud in both cases, since the experimenters and the pilot were themselves unaware of whether on any particular day the seeding mechanism in the plane was loaded or not. Precipitation was measured as the total rain volume falling from the cloud base following the airplane seeding run, as measured by radar.

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

Column	Name of Variable	Description of Variable
1	RAIN	Total rain volume (acre-feet)
2	CODE	1 for unseeded days, 2 for seeded
	days.	

We will use SPSS to answer the following two questions using the data:

1. Did cloud seeding have an effect on rainfall in this experiment?

2. If cloud seeding did have an effect on rainfall, estimate the effect in terms of how many times the volume of rainfall produced by a seeded cloud is larger than the volume that would have been produced in the absence of seeding.

14.2 Data Collection and Study Design

The goal of the experiment is to establish the cause-and-effect relationship between cloud seeding and precipitation. However, the conclusions we are going to reach about the relationship do not depend on the data only, but also on the experiment design and the way the data were collected.

The observed response variable in the case study is the total rain volume falling from cloud base. This variable was measured by unique modified radar. The research gave some indication that the radar was in fact underestimating rainfall. Thus, the results of the case study about the effect of seeding on rainfall would be stronger rather than weaker if the errors of the rainfall evaluations could have been avoided.

Let us analyze now the way the experiment was conducted. In any experiment it is necessary to define the experimental unit upon which a treatment may be applied and the appropriate measurement is to be taken.

The experimental units in the cloud seeding experiment are isolated cumulus clouds in south Florida on a day that was deemed suitable for seeding.

The aircraft surveyed the experimental area for clouds that might meet some preset conditions. These conditions intended to make the units as homogeneous as possible. Then the researchers selected one of suitable clouds. The random mechanism was not used in the phase.

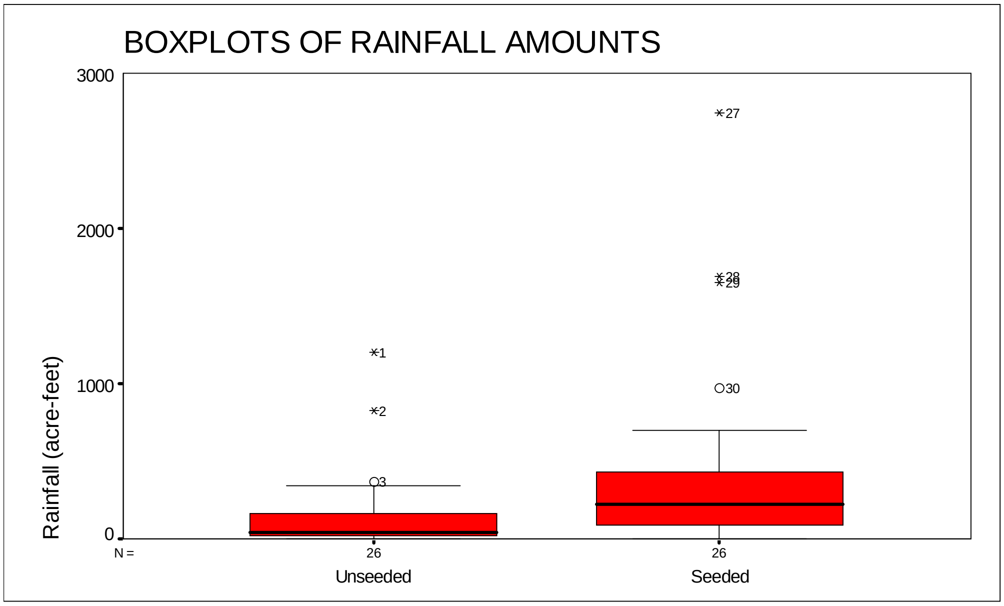
Once the target cloud was determined, a random mechanism was used to decide whether to seed the cloud on that day or to leave it unseeded as a control. The fact that the experiment was blind, that is the airplane crew was unaware of whether seeding was conducted or not, prevented the intentional or unintentional biases of the investigators from having a chance to make a difference in the results.

Thus the experiment is an example of a randomized experiment because the investigators controlled the assignment of experimental units (suitable clouds) to groups (seeded, unseeded) and used a chance mechanism to make the assignment. This study design enables us to draw causal inferences.

The clouds (experimental units) subjected to the treatment (seeding) or left as a control are not members of any well-defined population. Hence, the observed pattern cannot be inferred to hold in some general population.

14.3 Displaying and Describing Seeded and Unseeded Rainfalls

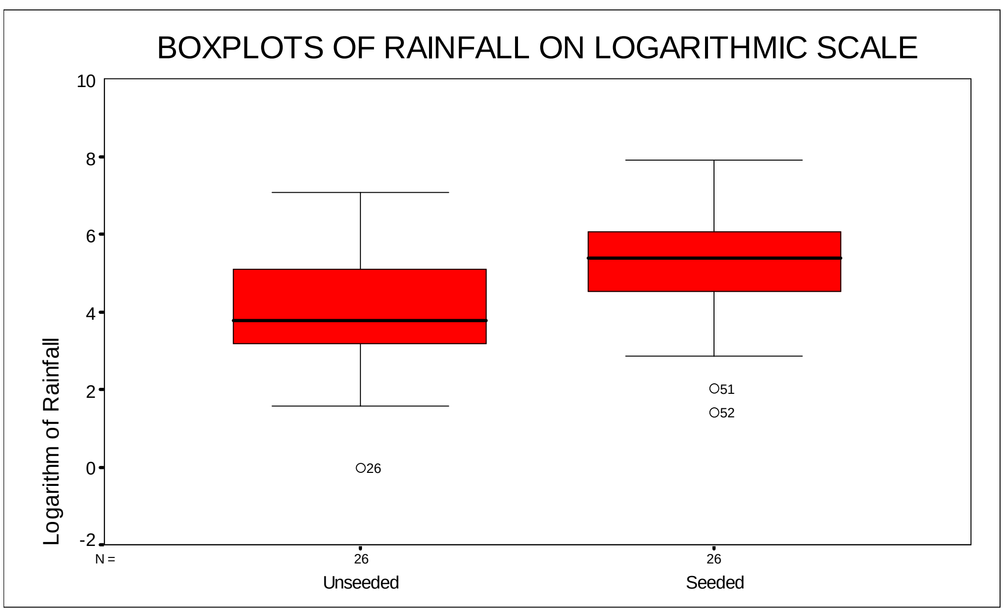
First we will obtain the boxplots of the rainfall amounts on the acre-feet scale for seeded and unseeded days. SPSS produces the following side-by-side boxplots rainfall amount on the acre-feet scale:



The boxplot for unseeded days is highly skewed to the right. It indicates that most of the time the rainfall volume from unseeded clouds is very small or zero. However, there are outliers and extreme observations in the distribution. The spread of rainfall amount is relatively small. The boxplot for seeded days is slightly skewed to the right. There is one outlier and three extreme observations in the distribution. The spread of rainfall amount is relatively large.

The side-by-side boxplots indicate that the rainfall tended to be larger on the seeded days. Both distributions are quite skewed, and more variability occurred in the seeded group than in the control group.

Now we will use SPSS to display and compare the distributions of the natural logarithm of rainfall for seeded and unseeded days.



The positions of the quartiles and whiskers indicate that on the logarithmic scale, both distributions are approximately symmetric, and have approximately the same spread. The boxplots confirm the conclusion we have reached before that the rainfall tended to be larger on the seeded days.

14.4 Describing Seeded and Unseeded Rainfalls

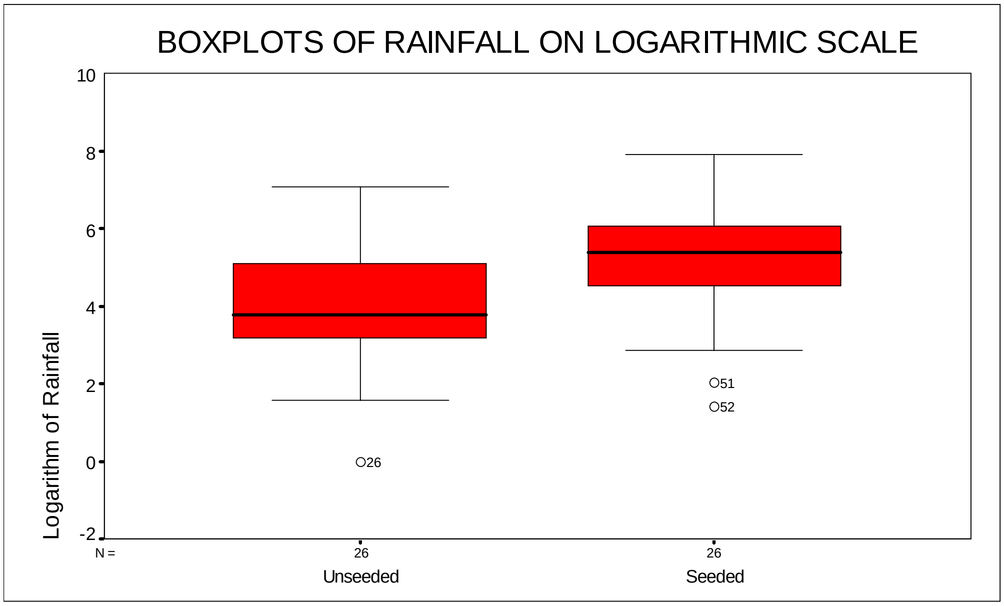
We will describe the data by obtaining the basic measures of center, spread, and shape for the distributions of rainfalls for unseeded and seeded days. The *Explore* command in SPSS produces the summary statistics for both distributions.

	STATISTICS	RAINFALLS	
		UNSEEDED	SEEDED
MEASURES OF CENTER	MEAN	164.5885	441.9846
	MEDIAN	44.2000	221.6000
	5% TRIMMED MEAN	120.7350	351.7201
	95% CI FOR MEAN	(52.1296, 277.0473)	(179.1260, 704.8433)
MEASURES OF SPREAD	STANDARD DEV.	278.4264	650.7872
	STANDARD ERROR	54.6039	127.6299
	VARIANCE	77521.26	423524.0
	IQR	159.6000	365.3250
	MINIMUM	1.0000	4.1000
	MAXIMUM	1202.600	2745.600
	RANGE	1201.600	2741.500
MEASURES OF SHAPE	SKEWNESS	2.7892	2.4352
	ST. ERROR SKEWNESS	0.4556	0.4556
	KURTOSIS	8.1731	6.0084
	ST. ERROR KURTOSIS	0.8865	0.8865
COUNT		26	26

The above numerical results confirm our conclusions reached in the previous section about the graphical displays for the data. All displayed measures of center indicate that the typical rainfall on the seeded days exceeds significantly the rainfall on the unseeded days. The measures of spread show that the spread of rainfall amounts is much larger on the seeded days. The distribution of rainfall amounts is more skewed on the unseeded days.

14.5 Displaying a Multiplicative Treatment Effect

The comparison of the boxplots for seeded and unseeded clouds can also give some insight in the nature of the seeding effect in the experiment. In Section 14.3 we obtained the following side-by-side boxplots of log-transformed observations for the two treatment groups.



As you can see, the distributions have approximately the same shape and spread. The appearance of the boxplot for seeded rainfalls suggests that it might be obtained by shifting the boxplot for unseeded rainfalls by a specific number of units up. It looks as if the seeding added the same rainfall amount to any cloud rainfall. Therefore, the additive treatment effect holds for the log-transformed data, and that translates into the multiplicative treatment effect on the original scale of measurement.

14.6 Making Inferences

Any inferences in this case should be stated in terms of treatment effects and causation, rather than differences in population means and association. In particular, we will test a null hypothesis of no seeding effect and obtain a confidence interval for the seeding effect.

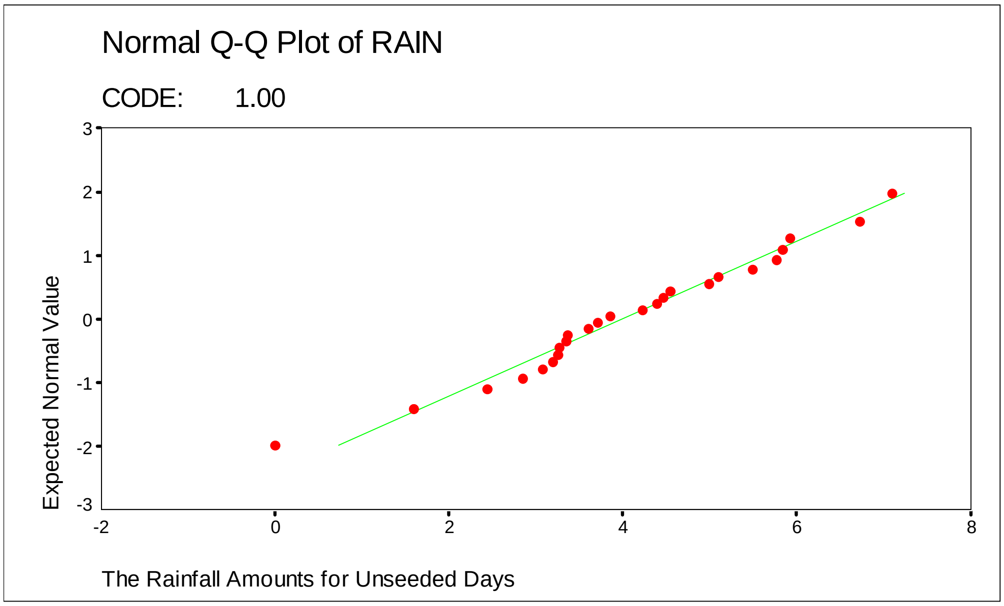
In order to see whether there is a treatment effect, we will use the t-test for two-sample problems in the way it would be used for a random sampling situation. The statistic can be accessed in SPSS by using *The Independent-Samples T Test* command. The procedure compares the means of one variable (rainfall amount) for two treatment groups (seeded, unseeded).

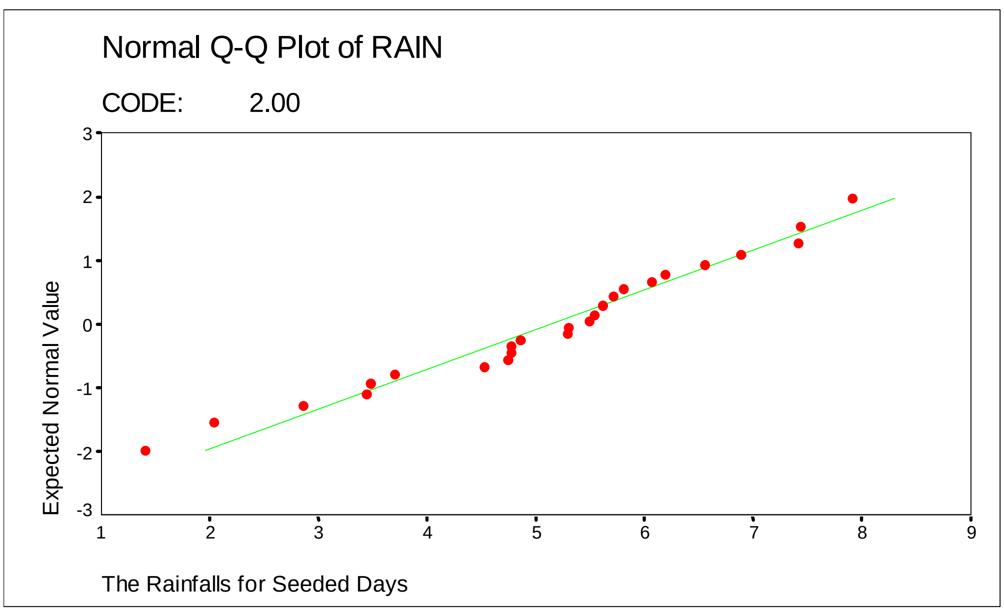
The assumptions of the t-tools (test and confidence interval) in the randomized experiment are that both treatment groups are independent of one another, and the treatment distributions are normal.

The design of the experiment does not provide any evidence that the independence assumption is violated. Is there any evidence that the assumption of normality is violated? The boxplots displayed in Section 14.3 indicate that the data for both distributions are skewed. Thus the assumption of normality is not justified. However, both distributions displayed on the logarithmic scale in Section 14.3 are approximately symmetric. Therefore, any inferences should be made after taking the logarithms of the rainfalls.

In order to determine whether or not a variable is normally distributed, you can use one of the two available procedures in SPSS: *Normal Q-Q* Plot or *Normal P-P* Plot. The *Normal Q-Q* plot plots the quantiles of a variable's distribution against the quantiles of the normal distribution. If the data come from a normal distribution, the plot should resemble a straight line.

The normal probability plot (Normal Q-Q plot) for each treatment group is displayed below. Remember that the unseeded days are coded as 1 and seeded days are coded as 2.





In each case the natural logarithms of data values were used instead of the original values themselves. Each of the above normal plots resembles a straight line and hence it supports the assumption of normality of the transformed observations.

The data do not provide any evidence that any of the assumptions necessary to apply the t-tools might be violated. Even if one of the assumptions is violated slightly, the robustness of the t-tools makes it possible to apply them in this case.

14.7 Test of Significance and Confidence Interval

SPSS produces the following output:

t-tests for Independent Samples

Variable	**Number of Cases**	**Mean**	**SD**	**SE of Mean**
RAINFALL				
Unseeded	26	3.9904	1.642	0.322
Seeded	26	5.1342	1.600	0.314

Mean Difference = -1.1438

Levene's Test for Equality of Variances: F= .058 P= .811

t-test for Equality of Means

Variances	**t-value**	**df**	**2-Tail Sig**	**SE of Diff**	**95% CI for Diff**
Equal	-2.54	50	.014	.450	(-2.047, -.241)
Unequal	-2.54	49.97	.014	.450	(-2047, -.241)

We test the null hypothesis that there is no seeding effect for the log-transformed observations (treatment effect is zero). In other words, we test the claim that there is no effect of cloud seeding on log rainfall.

A suitable tool to test the hypothesis of no treatment effect is the two-sample t-statistic. The *Independent-Samples T Test* procedure available in SPSS compares the additive effect of cloud seeding on the log rainfall. Denote by δ the additive effect of cloud seeding on the logarithm of rainfall. Thus the null and alternative hypotheses are

$H_0 : \delta = 0$ (no additive effect of cloud seeding on the log rainfall),

$H_a : \delta > 0$ (evidence of additive effect of cloud seeding on the log rainfall).

The value $\hat{\delta}$ defined as the difference between the average rainfall for unseeded clouds and the average rainfall for seeded clouds is an estimate $\hat{\delta}$ of δ. The t statistic under the null hypothesis has the form

$$t = \frac{\hat{\delta} - \delta}{SE(\hat{\delta})} = \frac{\hat{\delta} - 0}{SE(\hat{\delta})}.$$

The P-value of the two-sided t-test with the assumption of equal variances is obtained by SPSS as 0.014. Hence, one-sided p-value is 0.014/2 = 0.007. That means that there is convincing evidence to reject the null hypothesis of no effect of cloud seeding on log rainfall. The data support the claim that seeding causes the increase in rainfall.

The 95% confidence interval for the difference between the logarithms for seeded and unseeded rainfalls is 0.241 to 2.047.

We have applied the t-tools on the log scale because the log-transformed rainfalls have distributions that appear satisfactory for using the tools. Now we will transform our estimates back to the original scale.

According to the above output, the average seeded log rainfall minus the average unseeded log rainfall is 1.1438. The logarithm transformation enables us to obtain an estimate of the multiplicative effect of cloud seeding on rainfall.

In general, if Y_1, Y_2 are the responses to treatments 1 and 2, and $\overline{Z}_1, \overline{Z}_2$ are the observed averages for the two treatments, then

$$\exp(\overline{Z}_2 - \overline{Z}_1) \quad \text{estimates} \quad \frac{response(treatment\,2)}{response(treatment\,1)}.$$

In other words, the value of $\exp(\overline{Z}_2 - \overline{Z}_1)$ estimates how many times the response to treatment 2 is as large as the response to treatment 1. For details see your textbook, page 67-68.

In our case, the difference between the average rainfalls for seeded and unseeded clouds is $\overline{Z}_2 - \overline{Z}_1$ = 1.1438, and therefore exp(1.1438)=3.1384 is an estimate of the ratio of the responses. Thus the volume of rainfall produced by a seeded cloud is estimated to be 3.14 times as large as the volume that would have been produced in the absence of seeding.

The 95% confidence interval for the multiplicative treatment effect on the original scale is exp(0.241)=1.2720 to exp(2.047)=7.7425. Thus the treatment effect is estimated to be between 1.27 and 7.74 times.

14.8 Summary

The boxplots of the rainfalls for seeded and unseeded days reveal that the two distributions of rainfall are skewed. As the t-tools require the normality assumption be satisfied, they cannot be used on the original scale of measurement. However, the boxplots of the log-transformed data display symmetric distributions for seeded and unseeded days.

The comparison of the boxplots for seeded and unseeded observations for the log-transformed data reveals an additive treatment effect. The additive treatment effect for the log-transformed data can be converted into a multiplicative treatment effect for the data on the original scale of measurement. We estimated that the rainfall is 3.1 times as much when a cloud is seeded as when it is left unseeded.

The two-sample t-test can be used as an approximation to the randomization test. The null hypothesis about no additive treatment effect on the log scale can be back-transformed into the hypothesis about no multiplicative treatment effect for rainfalls on the original scale. The 95% confidence interval for the multiplicative treatment effect on the original scale is 1.2720 to 7.7425. Thus the treatment effect is estimated to be between 1.27 and 7.74 times.

The case study is an example of a randomized experiment. We used a random mechanism to decide whether to seed the target cloud on a given day or to leave it unseeded as a control. However, the clouds (experimental units) subjected to the treatment (seeding) were not selected from any well-defined population.

As the clouds were randomly allocated to the two treatment groups (seeded and unseeded), cause-and-effect conclusions can be drawn regarding the effect on the particular clouds selected. However, the observed pattern cannot be inferred to hold in some general population.