7. Checking the Assumptions Underlying Inferential Methods

7.1 What methods of inference should be used?

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 assesses the significance of the effect of the seeding on rainfall. As the clouds were randomly assigned to the two groups, any difference in response is due to treatment and not to other factors. A confidence interval based on the t distribution provides an interval estimate of the seeding effect.

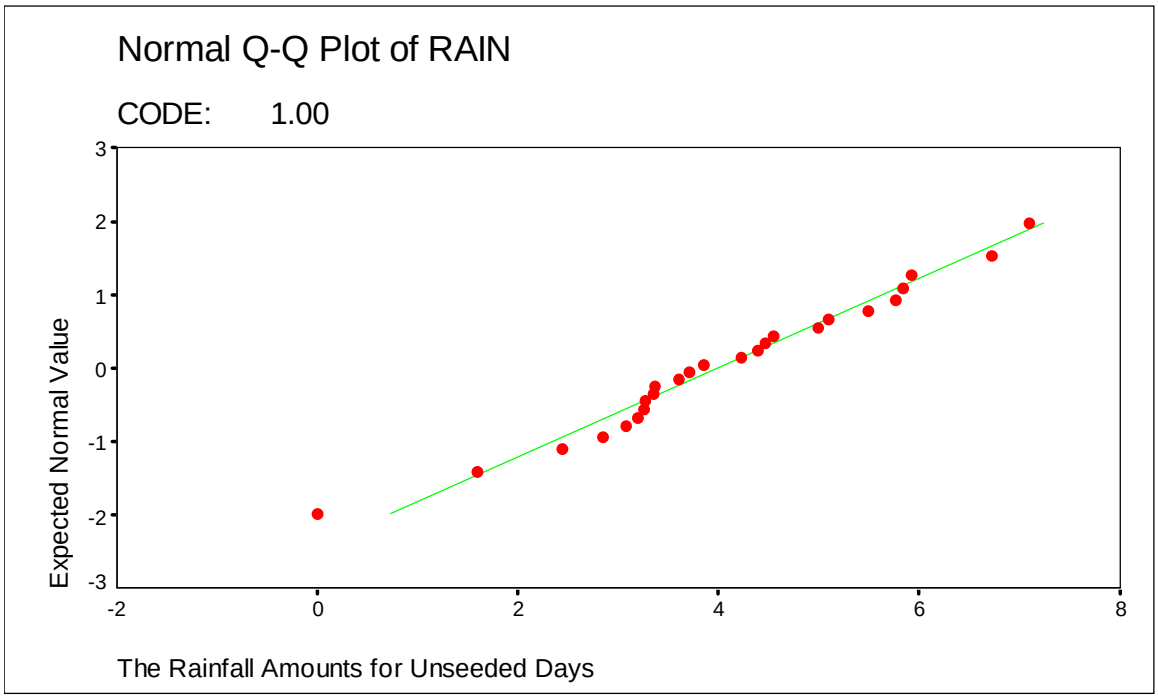
7.2 What assumptions should be satisfied and how to verify them?

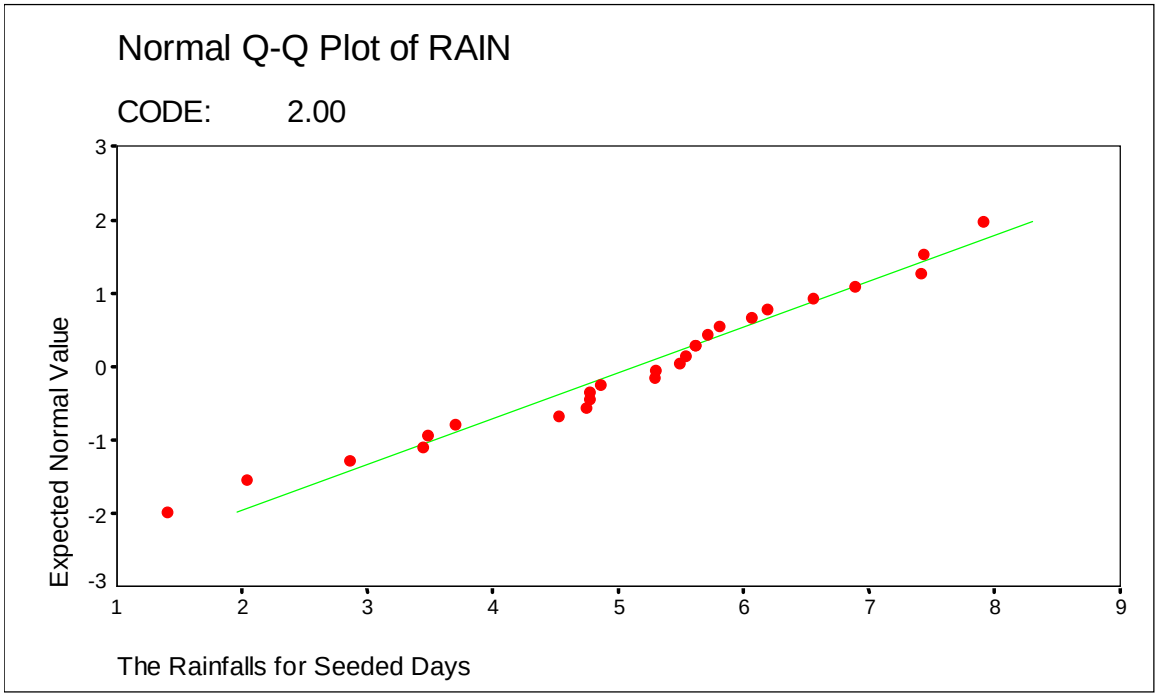
The assumptions of the t-tools (test and confidence interval) in the randomized experiment are that both treatment groups are independent of one another, and the treatment distributions are normal.

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 4.1 indicate that the data for both distributions are skewed. They are highly skewed for the unseeded clouds. Thus the assumption of normality is not justified. However, both distributions displayed on the logarithmic scale in Section 4.2 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.