Exercise 3

In this particular problem, we want to test if the daily number of deaths in various cities in Greece is equal to 18. This problem is a test for the population mean. I will check if the following assumptions are met for the use of this parametric test:

1. Our sample is random.
2. There are no extreme values in our data exceeding 10%.
3. Our data follows a normal distribution.

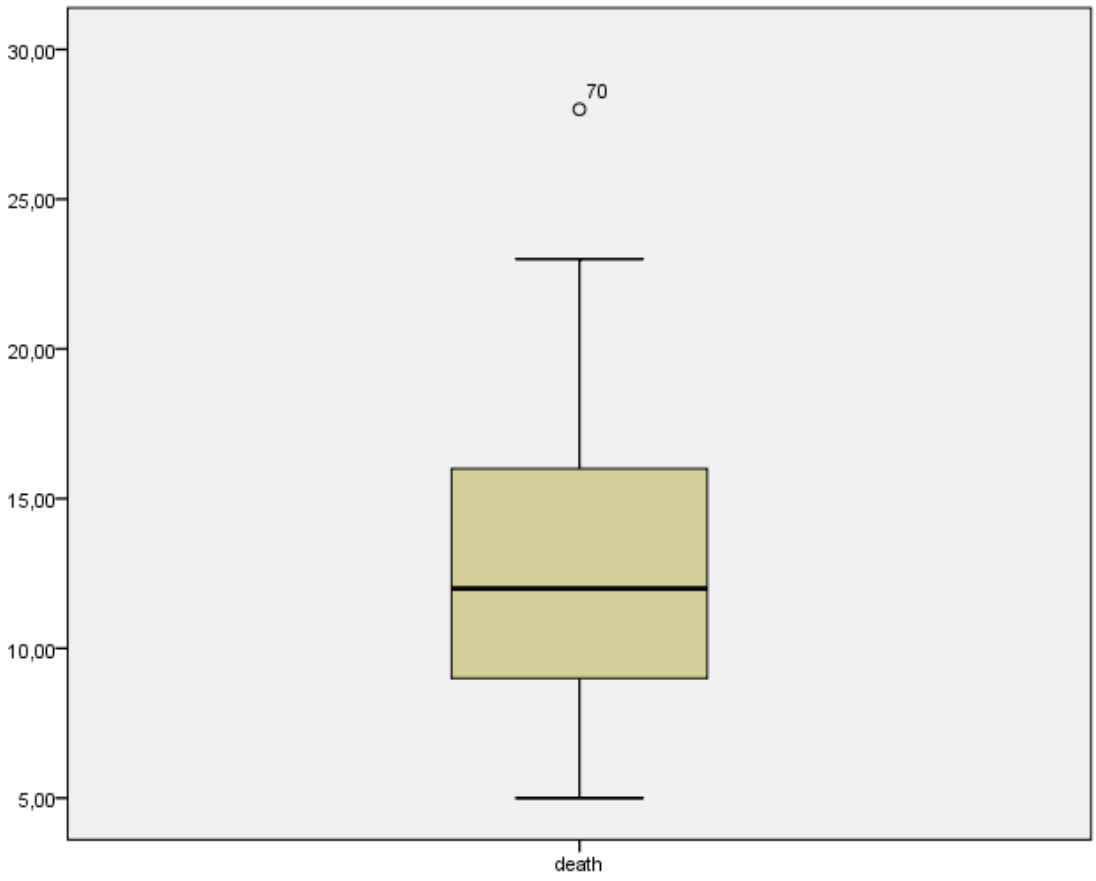
The first assumption relates to how we selected our sample, and it is satisfied. The check for extreme values was done using a boxplot, and it showed that there are 2 extreme values, observations with serial numbers 1 and 70, with death counts of 30 and 28, respectively (see boxplots 1, 2, 3). Since the percentage of extreme values does not exceed 10%, we proceed with further analysis, having excluded these two observations.

The Shapiro-Wilk test rejects the hypothesis that the sample values of the number of deaths come from a population that is reasonably described by a normal distribution (test statistic 0.948, p-value = 0.005). The transformation of the natural logarithm (applied to all observations) corrects the normality issue (see table 1).

Since all the assumptions are satisfied, we can use the parametric t-Test to test the hypothesis that the mean natural logarithm of the number of deaths is equal to the natural logarithm of 18, which is approximately 1.255. This test results in a statistically significant difference from 1.255 ($p < 0.001$). Additionally, a 95% confidence interval for the mean natural logarithm of deaths is (1.255-0.2121, 1.255-0.1325).

Boxplot 1, 2, 3


30,00
25,00
20,00
15,00
10,00
5,00
1
70
death


30,00
25,00
20,00
15,00
10,00
5,00
70
death

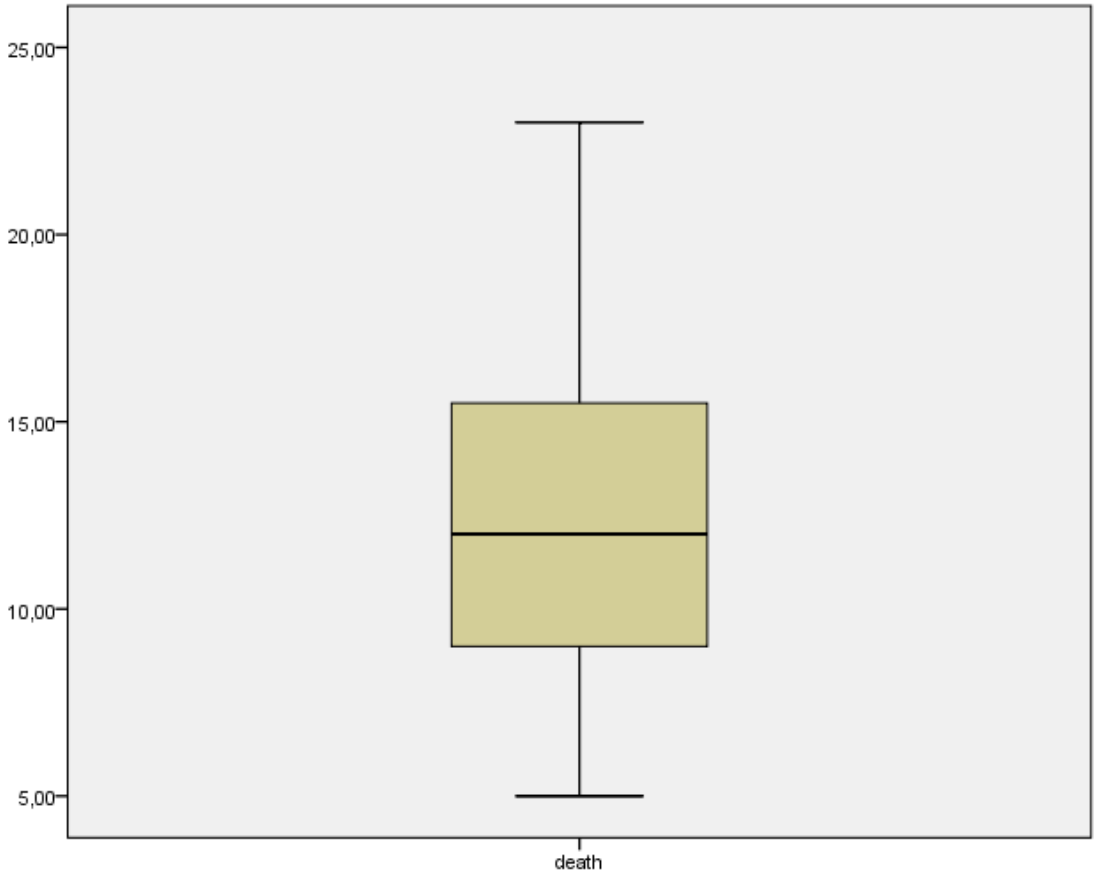


Table 1

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logarithmos	,062	74	,200*	,986	74	,610

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction