

Assignment 2: Statistics

1 Introduction

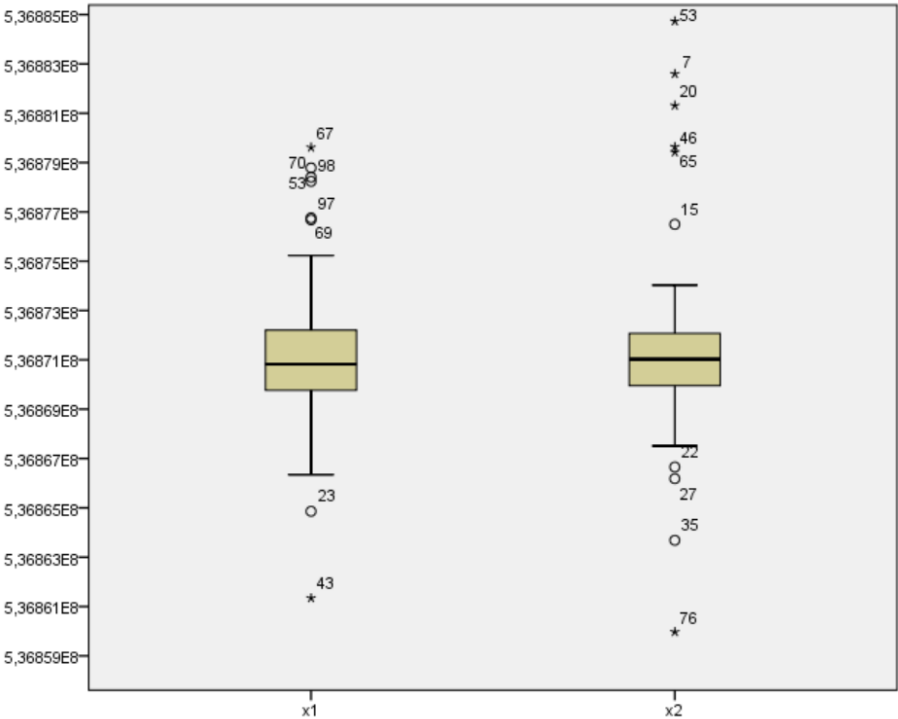
As a second assignment for the course Methods for Scientific Research, we have to fulfil four statistical exercises. Further details of which can be found in the assignment document available on Canvas. Also provided were the datasets to be used for the exercises, based on our student number (0529279) the set number calculator¹ assigned us with the sets 4, 1, 4, 3 for the four respective questions. All tests, charts and diagrams were generated using IBM's SPSS Statistics package version 23² released in 2015.

2 Question 1

The used dataset –set 4 for the first question– was gathered by way of an experiment measuring the resonance frequencies of two types of crystal oscillators (variables x1 and x2) needed to provide timing information for high performance hardware.

2.1 Diagram

The use of boxplots is a standardized way of displaying the distribution of data based various L-estimators such as the minimum, first quartile, median, third quartile and maximum.



Given boxplots show that both variables have an evenly distributed range.

¹<https://ai.vub.ac.be/~bart/statsnumbers.html>
²<https://en.wikipedia.org/wiki/SPSS>

Dataset x1 however has a wider distribution of its data with the maximum and minimum of the data being further away from the median than in dataset x2.

Explanation

The use of boxplots is a standardized way of displaying the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum. Between the first and third quartile lies 50 percent of all data with the median being the middle value of the whole set. Between the maximum and minimum lies about 99.75 percent of all data. Outliers are even further away of the median than the minimum and maximum and represent 0.25 percent of all data.

Using boxplots gives us a simple way of detecting certain characteristics of our data such as symmetry and distribution. It summarises our datasets in terms of median, first quartile, third quartile, maximum and minimum and eases comparison between the two.

2.2 Welch's T-test

Based on the experiment's description, we want to see if the same variable applied on two different populations (e.g. different groups of oscillator crystals) makes for a difference of the means of variables. Because the variance of both datasets differs significantly, a normal T-test would not be reliable in doing so. This is why we settle for Welch's T-test because it is robust even when the sample size or the variance is unequal. In this case, an independent T-test is the best option. Our null hypothesis in this case is that the means of both data collections do not differ.

ANOVA

frequency

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	186642792.2	1	186642792.2	1,508	,221
Within Groups	2.450E+10	198	123750522.7		
Total	2.469E+10	199			

Robust Tests of Equality of Means

frequency

	Statistic ^a	df1	df2	Sig.
Welch	1,508	1	135,500	,222
Brown-Forsythe	1,508	1	135,500	,222

a. Asymptotically F distributed.

Because the p-value in this case is greater than 0.05 (0.222), we cannot conclude that a significant difference exists and we accept our null hypothesis.

2.3 Crystal reliability

Because both dataset means do not differ significantly and are normally distributed, we can trust both crystal types x1 and x2.

3 Question 2

4 Question 3

5 Question 4