Exploring diffrences in Satellite's lifetime using permutation tests

Netta Shafir and Shaked Amar

August 2022

Introduction

During the cold war, one of the centers of the struggle between the United States and the Soviet Union has been known as the **"Space Race"** - the try of the two rivals to achive a supermacy in spaceflight capabilities. Since then, the presence of a country in the outer space is considered as main indicator for it's strength.

Today, a lot of countries has satellites in space for varied applications. Nevertheless, the majority of the satellites is still belong to few, and we were interested in the differences of the sattelites between those countries. In particular, we would like to test the hypothesis that there is a difference in the expected value lifetime of satellites of those countries. Under several assumptions, those kind of hypothesis can be tested within F-test as part of ANOVA (anylasis of variance), which are a generalization of the t-test in the case of more than two groups. Notwithstanding, an important assumption of the t-test and the F-test in our context of testing the hypothesis that two of more groups have the same mean, is that the response variable is Normaly distributed. As we'll see soon enough, this assumption doesn't hold in our case, and therefor we will conduct similar tests, which are not-parametric and therefor are free from any assumption about the exact distribution of the data. These tests involves random permutations.

The dataset

The dataset we investigates has been taken from Kaggle[1]. It contains data about all the satellites that launched since 1957 - The year in which the first satellite, Sputnik 1, was sent to space by the USSR. The variable **owner** refers the country that holds the satellite. The variables **launch_date** and **flight_ended** refer the dates in which the satellite was sent to space and got out of use, respectively. As for the **status** feature, it tells the current activity status of the satellite[2], so in particular a satellite whose status is **Operational** should not have any value in the **flight_ended** field. Therefor, we were intrested only in satellites their status is **decayed**, about which we could get the full number of days until out of use.

First exploration of the data

To minimize time-related influences, we focus on satelittes that launched in the last decade. Figure 1 shows the distribution of the satellites to countries in the last decade (not including satellites which are still operational).

We decided to focus on United States, Russia and China, for they are leading in the amount of satellites, in a noticeable gap from the other countries. Figure 2 shows the distribution of the days in space of the three countries. One can observe that it are not Normal.

Because the wide right tail of the distributions, we wanted to operate a concave transformation on the data for axamining their nature more carefully. Figue 3 shows the data after we applied a log transformations and a square root transformation on it. Now, the ditributions related to US and Russia seems like taken from a GMM (gaussina mixure model)[3]. A QQ-plot with comparison to the Standard Normal indeed shows that a few segments that unites with the $45°$ line and therfore have the same distribution as the Standard Normal, and another segments that create a straight line and therefore fits Normal distribution with other parameters.

Nevarthelss, although this result tells us about the nature of the disributions - they are still not normal, so the classic t-test and F-test won't bring any help here.

---

[1] https://www.kaggle.com/datasets/heyrobin/satellite-data-19572022

[2] As reported in Kaggle, the dataset gets updated mounthly.

[3] This might imply that the population of the satellites can naturally divided to few sub-populations. For example, a group of satellite that manufactured with flaws that the centroid of their disributed around small number of daysand another group of satellites that properly manufactured.

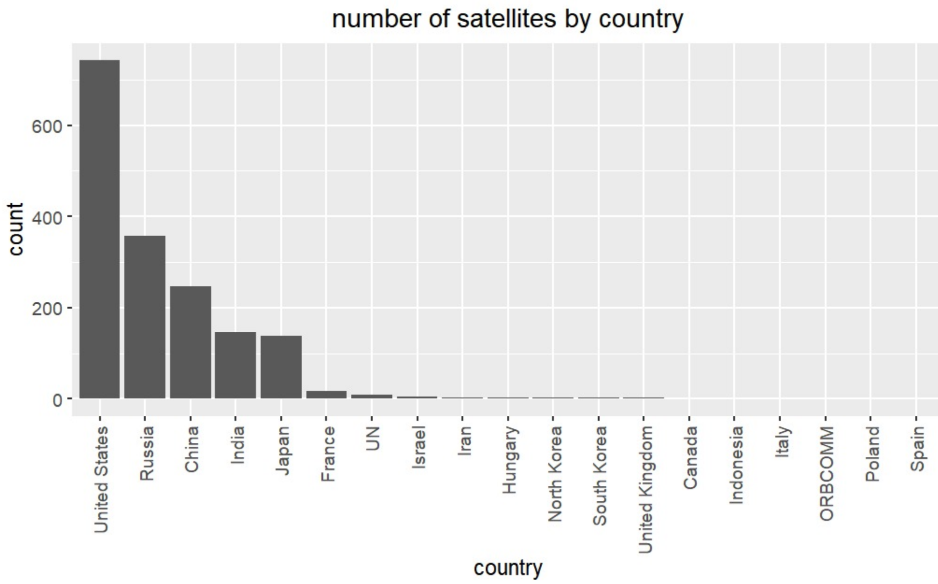

Figure : 1 The distribution of the satellites among the different countries.

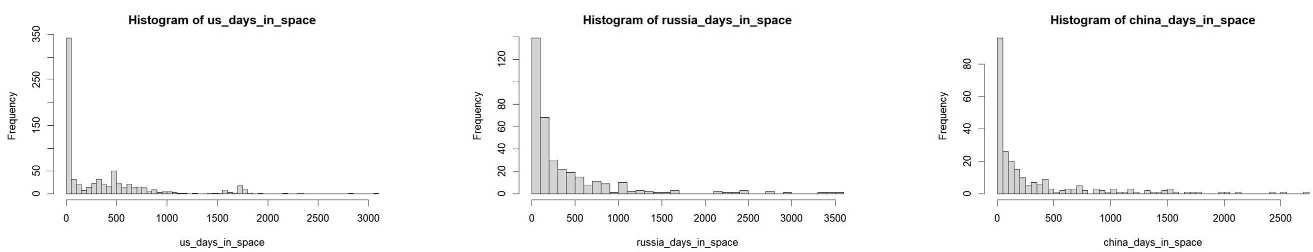

Figure : 2 The distribution of the days, for each country.

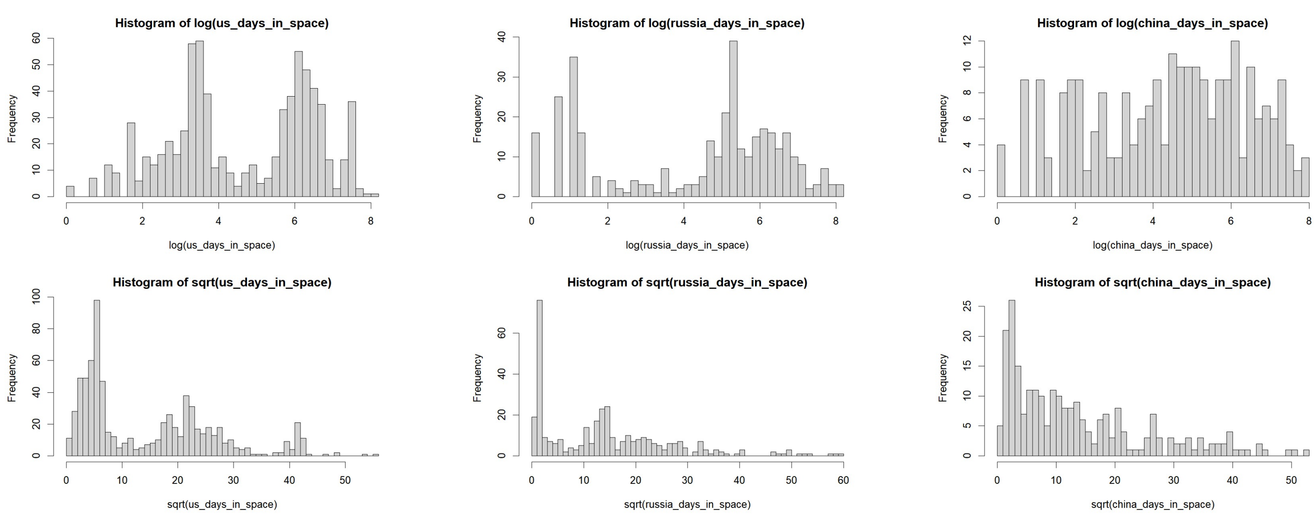

Figure : 3 The distirbutions of the days in space, for each country, after operating a log and square root transformations on them.

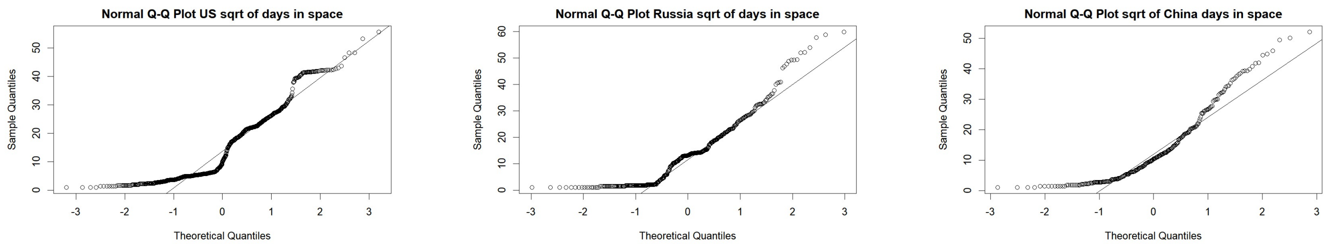

Figure : 4 QQ-plot of the sqare root of the three distributions, comparing to the Standard Normal. One can observe the straight segments in these graphs, which fits the Normal distribution. QQ-plot of the log transformation comparing to the Normal distribution gives similar results.

Methods

First, we will test the hypothesis that all the three countries has the same expected value. For that puprose, we will use a test that suggested in the paper by Anderson et al. (2001). This test is very similar to the regular F-test used in ANOVA, but without the assumption of the normality, what requires calculating the p-value throgh simulations, instead of the looking in a table of the F distribution.

The test statistic

In the case of regullar F-test of the one-way ANOVA, we have K groups, Normally disturbuted with the same variance, when the i'th group having n_i samples and expected value of μ_i. The null hypothesis is that $\mu_1 = \mu_2 = \ldots = \mu_K$, and under it the F statistic is calculated by:

$$F = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_j} \left(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}\right)^2 / (K-1)}{\sum_{i=1}^{K} \sum_{j=1}^{n_j} \left(\overline{y}_{i\cdot} - y_{ij}\right)^2 / (N-K)}$$

when N is the total number of samples, y_{ij} is the j sample in the i'th group, $\overline{y}_{i\cdot}$ is the mean of the i'th group and $\overline{y}_{\cdot\cdot}$ is the overall mean. Now, defining the following statistics:

$$SS_T = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij}^2 \qquad SS_W = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{N} d_{ij}^2 \epsilon_{ij} \qquad SS_A = SS_T - SS_W$$

When d_{ij} is any distance or similarity measure of our choice and ϵ_{ij} is an indicator of the event that the i'th and the j'th samples belongs to the same group. Then, the test statistic of our test is:

$$F_{pseudo} = \frac{SS_A / (K-1)}{SS_W / (N-K)}$$

And it can be shown that in the case in which d is chosen to be the Euclidiean distance, then F_{pseudo} is equals to the traditional parametric F statistic. We will use the Euclidean distance for d.

Calculating the p-value using permutations

Since the distribution of the F_{pseudo} test statistic under the null hypothesis is not known, it is necessary to use permutations for calculating it: Lets assume that the null hypothesis is true. Then, assuming that the observations are exchangeable[4], we get that there is no difference between the three groups. Thus, all the observations belongs to the same group, so the separation of the satellites to the three groups of US, Russia and China can be thought as a random shuffle. Furthermore, all the options of dividing all the satellites to three groups in those sizes (permutations) are equally probable. Hence, the distribution of F_{pseudo} can be numerically computed by calculating F_{pseudo} for every possible permutation of the satellites.

Yet, calculating all the possible permutations is quite expensive giving such large amount of samples, and hence not practical. Therefore we'll calculate an approximate p-value, using large amount of random permutations. This method is quite accepted and known as PERMANOVA. We chose a number of $1,000$ permutation, as recomended in the paper we

[4] This assumption will be discussed in the Discussion part later on.

read[5]. If marking with F^{π}_{pseudo} the test statistic calculated for a single random permutation, the p-value will then calculated by:

$$p = \frac{\left(\text{No. of } F^{\pi}_{pseudo} \geq F_{pseudo}\right)}{\left(\text{Total No. of } F^{\pi}_{pseudo}\right)}$$

Examining the pair-wise diffrences between the group

Then, we will conduct another three permutation tests, each of them will test the hypothesis that two of the three groups has the same expected value of lifetime. The tests statistic here will be just the difference in the means of two groups, and ame as before will be calculated for the orinigal groups - marked by t_{pseudo},and for each random permutation of dividing the total pool of samples to two groups - marked by t^{π}_{pseuso}. The p-value in those tests will be calculated by[6]:

$$p = 2 \cdot \frac{\left(\text{No. of } t^{\pi}_{pseudo} \geq |t_{pseudo}|\right)}{\left(\text{Total No. of } t^{\pi}_{pseudo}\right)}$$

The reason for this additional tests will be discussed in the Discussion part later on.

Results

The first test results can be shown if figure 5. The p-value, calculated as described above, stands on ????.

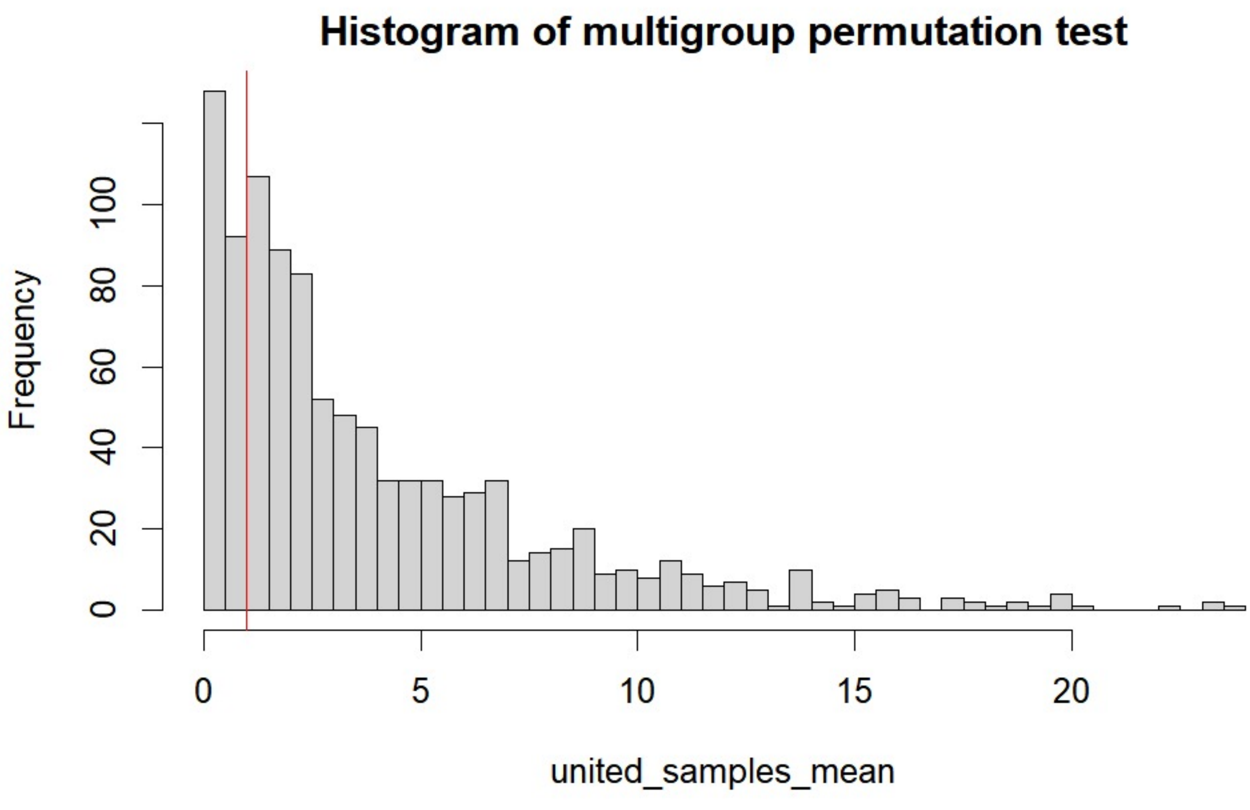

Figure : 5 The distribution of the F_{pseudo} under the null hypothesis of the first test. The value of the test statistic of the true groups are marked in red.

[5]Anderson et al. ,(2001) page .37

[6]The difference from the first test is that the support of the test statistic here is not necessarily positive. So the p-value - which is the propability of a random sample to be more extreme as the sample we actually got - should be calculated differently.

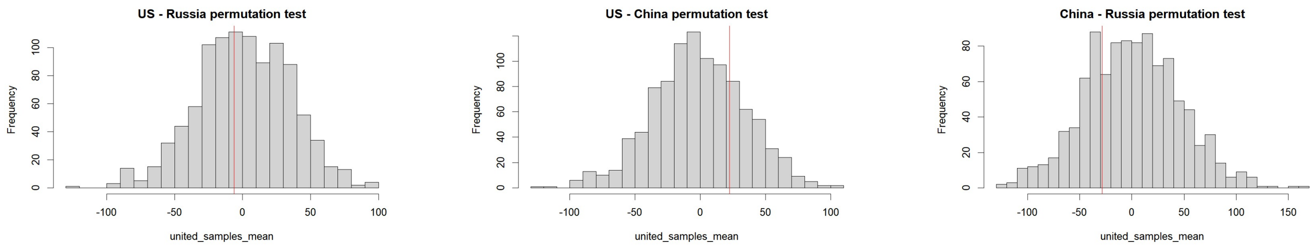

Figure : 6 The distributions of the t_{pseudo} test statistics of the permutation tests for checking pairwise differences between the groups.

As for the three permutations tests, the distributions of their t_{pseudo} statistic is shown in figure 6. The p-value of the three test is ???? for the US-Russia test, ???? for the US-China test and ???? for the China-Russia test.

All the p-values are pretty high, so the null hypothesis would not be rejected in tests of any accepted significance.

Discussion

Examining the model assumptions

As we wrote earlier, the only assumption of the first test (pseudo F-test, PERMANOVA) is that all of the observations are exchangeably under the null hypothesis. Exchanability means that the joint distribution of the observations (when thinking of them as random variables) does not affected from changing the order of the samples (Bernardo et al. 1996). In particular, this means that all the observations are identically distributed (but not necessarily independent, though[7]).

We tried to reduce the dependency between observations in the model by concentrating a single decade only, and from the nature of our data we think it is reasonable to assume that there is low dependency between it's observations. As for the distributions of the different groups, although looking somewhat similar from the histograms in figure 2, the standard deviations of the groups not as similar as we would expect: 573.9 for Russia, 489.1 for US and 512.6 for China, i.e. difference of almost 100 s.d. between Russia and US. Under the Assumptions section in Anderson et al., it is noted that although the first test checks differences in the expected values of the groups, it is sensitive to differences in the variances between the groups, so the p-value can be low even if the expected values not really differ. For that reason we wanted to use the pair-wise permutation tests: Even if the true distributions, or at least the variances, of the groups, are not equals (besides of their expected value), and would effect the first test, the pairwise regular permutation tests are less effected from differences in the variances of the groups and will catch biases in the first test. In addition, if the p-value was low, so we would tend to reject the null hypothesis, we would like to check more specifically which pair of the three has the difference, or maybe there is difference in of all the pairs.

Eventually, as shown above, the p-values of all of the tests were pretty high, so if we have to guess we would say that the three countries does not different in their outer space abilities, after all.

Refrences

1. Anderson, M.J. (2001), A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26: 32-46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

2. Bernardo, J. M. (1996). The concept of exchangeability and its applications. Far East Journal of Mathematical Sciences, 4, 111-122.

Appendix: The code

[7] In page 37 in Anderson et al. ,(2001) it written that the exchangeability assumption suggest that all the observations are independent, which is a stronger assumption that the known definition for exchanability we mentioned above, and therefore not true: As counterexample we can think of n r.v. normally disturbuted with correlation $\sigma^2 > 0$ between each two. It easy to see that this series meet the exchanability condition, altough it is not independent.