Analysis of silicone contamination of batches

Thom Hurks	*0828691*
Puck Mulders	*0737709*
Marijn van Knippenberg	*0676548*
Rik Coenders	*0777004*

May 25, 2016

Contents

1 Introduction

The health inspection has detected a silicone leakage in a machine used in a chemical production process. The leakage was detected in close vicinity of packed product, so the question is whether the product has been affected by the leakage. The machines are regularly inspected, and data on a batch produced before the previous inspection is available and is denoted as batch B0. Batches denoted B1 to B7 have been produced between the previous inspection and the current inspection and may have been affected by the leakage. The data consists of measurements of samples of units from the batches. The silicone that leaked is already part of the product itself, so it will always be present in the measured units to some extent. The goal of this report is to possibly identify the batches that were affected by the silicone leakage. In the first section we describe which methods we use to find the affected batches. In the next section we will describe what the results of these methods are. Lastly, we will conclude which of the batches are affected. In this section we will also describe what the weaknesses and risks are of our investigation.

2 Methods

To decide if any of the batches are affected, we perform both parametric and non-parametric tests. The parametric tests assume that the data is normally distributed, which means we must test for normality and if necessary transform the data. Non-parametric tests are generally less powerful than parametric tests, but can always be applied regardless of normality. With parametric tests one must be careful to draw conclusions about the original non-transformed data if the data needs to be transformed in order to meet the normality assumption.

2.1 Exploritorary data analysis

Before performing any parametric or non-parametric tests, we will first perform some standard statistical tests to gain some insight into the data. As a start, we will plot the data and calculate the mean, standard deviation, variance and median of each batch. After that, we will perform a Shapiro-Wilk test to see whether the non-transformed data is normal. We have formulated the null hypothesis as can be found in equation1:

$$\begin{cases} H_0 : B_i \sim \mathcal{N} & \text{for } 1 \leq i \leq 7 \\ H_1 : B_i \nsim \mathcal{N} & \text{for } 1 \leq i \leq 7 \end{cases} \tag{1}$$

We will reject the null hypothesis if the p-value < 0.01. After that, we will perform a Wald-Wolfowitz Runs Test which has as a null hypothesis that each element in the sequence is independently drawn from the same random distribution which is expected if none of the batches are affected by silicones. We reject the null hypothesis if $p < 0.01$.

2.2 Critical value

Most of our hypotheses include testing all seven possibly affected batches. Because of this reason, we have to adjust the commonly used critical value of 0.05 to make sure we do not reject the null hypothesis due to chance. Applying Bonferroni will give us a critical value of $\alpha = \frac{0,05}{7} = 0,00714$, but since this value is quite low we may fail to reject the null hypotheses in cases where we actually want to; our goal is to identify the affected batches, so we can be a bit more strict and raise the alpha value slightly. As such, we will set the critical value to 0.01.

2.3 Normal distribution

2.3.1 Investigating normality on transformed data

We will perform a log-normal transformation on the data to modify the data such that the form of the data will shape towards a normal distribution. To check whether this helped, we will plot the transformed data and perform a Shapiro-Wilk test on each of the batches, which can be found in equation 1.

We will reject the null hypothesis if the p-value < 0.01. Besides a Shapiro-Wilk test, we will perform skewness and kurtosis tests to test on where normality is violated. The critical value of a two-sided skewness test is 1,157, while the critical value of the kurtosis is 1,49. We will perform a Jarque-Bera test to see if it is appropriate to reject a normal distribution based on skewness and kurtosis. The null hypothesis is the same as with the Shapiro-Wilk test, which can be found in equation 1. We will reject normality if the p-value < 0.01.

2.3.2 T-test

Because batch 0 is produced before the previous inspection where no leakage was discovered, we know that the mean and variance of the amount of silicones found in each batch should be similar to the amount of silicone in batch 0. To check whether they are similar, we will perform a one-sided t-test, because a leakage will cause the outcome to become higher, while it would be acceptable if the average amount of silicones in a batch is lower than in batch 0. Therefore the hypothesis is the following:

$$\begin{cases} H_0 : \mu_0 \geq \mu_i & \text{for } 1 \leq i \leq 7 \\ H_1 : \mu_0 < \mu_i & \text{for } 1 \leq i \leq 7 \end{cases} \tag{2}$$

If the result of the t-test is less than 0.01, we will reject the null hypothesis.

2.3.3 F-test

Besides performing a t-test, we will also perform some homogeneity tests to see whether the data also share the same characteristics. To check if this is the case, we will perform a Barlett's test, which checks whether all the variances of the batches are equal to each other. The null hypothesis of that test is that all variances are equal, and thus that the sets are homogeneous. Besides that we will perform an F-test to see which batches have similar properties as batch 0.

2.3.4 Outliers

We will perform a Grubbs outlier test on the log-transformed to check whether there are any indications that one of the measurements has some incorrect value due to for example measurement errors. We will do this for both outliers on the left and right tail. Our null hypothesis is can be found in equation 3.We will reject the null hypothesis if the p-value < 0.01.

$$\begin{cases} H_0 : \text{Batch}_i \text{ contains no outliers} & \text{for } 1 \leq i \leq 7 \\ H_1 : \text{Batch}_i \text{ contains outliers} \quad \text{for } 1 \leq i \leq 7 \end{cases} \tag{3}$$

We will reject the null hypothesis if the p-value< 0.01.

2.4 Non-parametric tests

2.4.1 Serial correlation test on the means

We will perform a serial correlation test on the means of the batches. The serial correlation test is Von Neumann with lag-1 autocorrelation. The null hypothesis is that the means of the batches are random. If this indeed is the case, this is an indication that none of the batches is affected by the silicones. We will reject the null hypothesis if $p < 0.05$.

2.4.2 Kolmogorov-Smirnov Test

If the non-transformed data does not follow a normal distribution, we will also perform some non-parametric tests that do not assume a normal distribution. Such a parametric test is the Kolgomorov-Smirnov Test. For the two-sample version of this test, we need to make the following assumptions:

- The two samples are independent.
- The outcomes are ordinal or numerical.

Since the results of the samples do not depend on each other, the first assumption can be made. Second, the outcomes are all numerical so the second assumption can be made too. We perform a two-sided test to get a general view of whether non-parametric testing picks up on any differiations between the distributions of batch 0 and the other batches. Our hypothesis is the following: We will reject the null hypothesis when $p < 0.01$. Because batch 7 has a missing value, we have decided to impute the mean of batch 7 into the line with the missing value, to see whether this affects the test. The Kolmogorov-Smirnov test will not give the exact p-value due ties in the data, so it could be the case that the p-value retrieved from the test somewhat deviates from the real test. Because we have picked a somewhat high critical value, we will not accept batches that could be infected.

2.4.3 Wilcoxon Rank-Sum Test

A second non-parametric test we will perform is the Wilcoxon Rank-Sum Test, which needs the assumption that both distributions are from an ordinal distribution. This implies that the data cannot contain any ties. Since we have some ties in the data, we will not get the exact p-value. Since our value does not contain a substantial amount of ties, the bias of the p-value will not be extremely high. Because the critical value is already somewhat higher than the suggested Bonferroni value, the hypothesis test will not pass while it should have failed. The hypotheses of the Wilcoxon Rank-Sum test can be found in equation 4:

$$\begin{cases} H_0 : \text{Batch}_0 \text{ and batch}_i \text{ come from the same population} & \text{for } 1 \leq i \leq 7 \\ H_1 : \text{Batch}_0 \text{ and batch}_i \text{ come from a different population} & \text{for } 1 \leq i \leq 7 \end{cases} \tag{4}$$

This null hypothesis indicates that the change that the median of $batch_0$ is larger than the median of $batch_1$ is the same as that the median of $batch_0$ is smaller than the median of $batch_1$. We will reject the null hypothesis if p-value < 0.01.

3 Results

3.1 QQ Plot and Shapiro-Wilk

The first step is to generate a normal QQ plot for batch 0 as an easy visual confirmation of normality.

	batch	variance
1	B0	0.9400
2	B1	0.3147
3	B2	1.4468
4	B3	0.6830
5	B4	0.1558
6	B5	0.2116
7	B6	0.3832
8	B7	0.4372

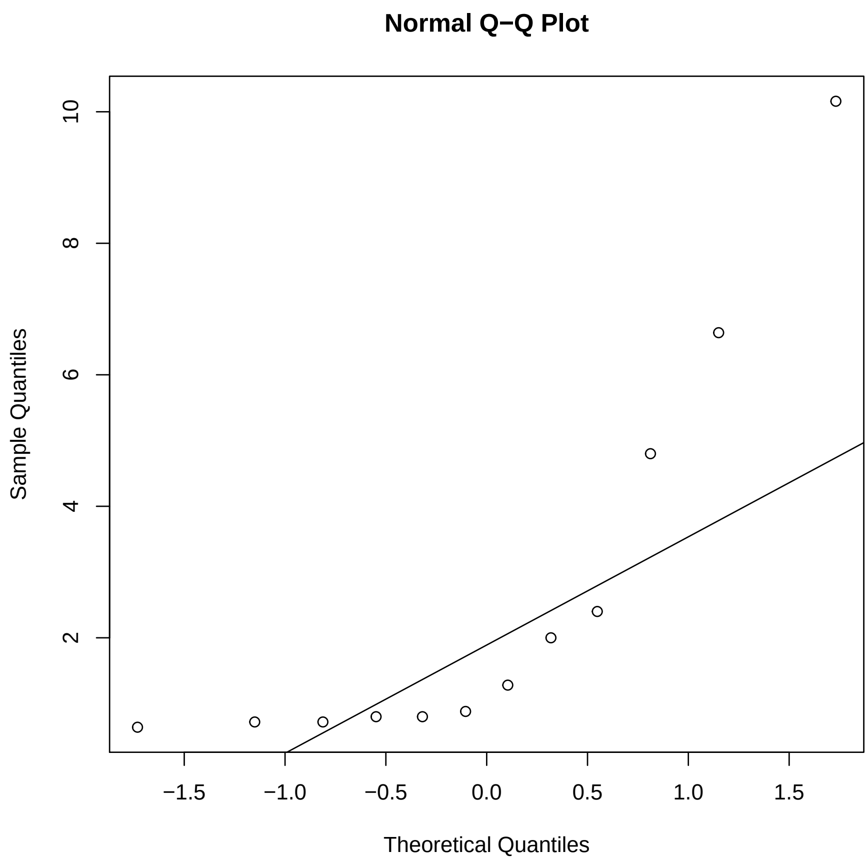

Figure 1: QQ Plot of batch 0

The normal QQ plot already seems to suggest that batch 0 is not normally distributed. When we also perform the Shapiro-Wilk normality test we can see the p-value $= 0.0013049 < 0.01$ so we can reject the null-hypothesis that the outcome is normally distributed.

Next we can generate a normal QQ plot for all batches 0 to 7. Of course generating this plot for multiple batches at once assumes that the units are somewhat consistent across different batches.

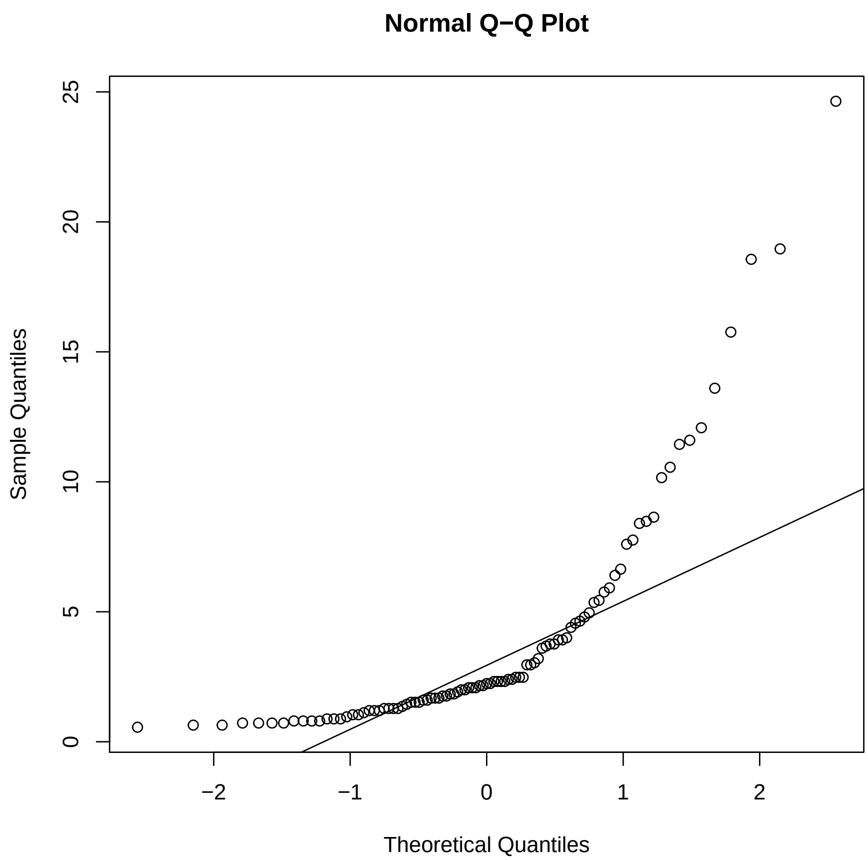

Figure 2: QQ Plot of batches 0 to 7

The normal QQ plot here too seems to suggest that the data is not normally distributed. We can again compute the Shapiro-Wilk test p-value for all batches to gather more information.

	p.value
B0	0.0013
B1	0.0113
B2	0.0442
B3	0.1218
B4	0.0759
B5	0.0020
B6	0.0006
B7	0.0865

We can reject the normality null hypothesis for batches zero, five and six with p-values respectively 0.0013049, 0.0019666 and 6.2737479×10^{-4}. We can also observe that for the other batches the p-values are not particularly high, the exception being batch 3 with a p-value of 0.1218305.

3.2 Normal distribution

3.2.1 Shapiro-Wilk on log-transformed data

In the previous section we observed that the data does not seem to follow a normal distribution. In order to perform parametric tests that assume normality, we must then transform the data. We will transform the data using the well-known log transformation and then check for each batch what the p-value of the Shapiro-Wilk test is in order to confirm that the transformation had the desired effect.

	batch	p.value
1	B0	0.0388
2	B1	0.1414
3	B2	0.6190
4	B3	0.6635
5	B4	0.8979
6	B5	0.0620
7	B6	0.2064
8	B7	0.9643

None of the batches reject normality, indicating that the Shapiro-Wilk test cannot be used to reject normality. We can generate a normal QQ-plot to get more insight into the log transformed batch 0.

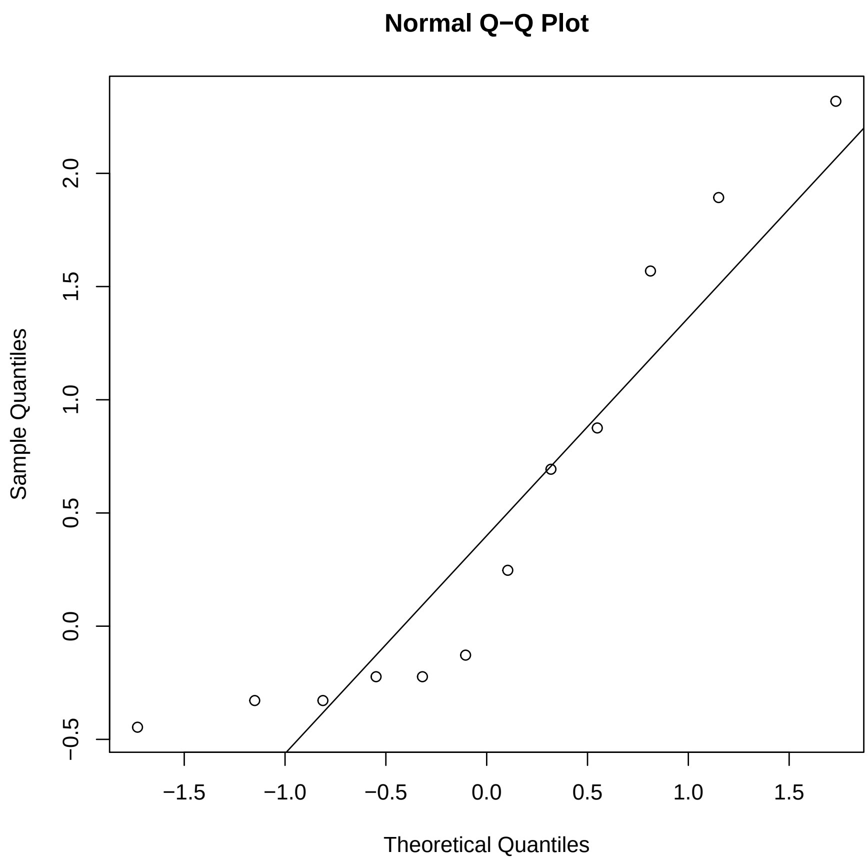

Figure 3: QQ Plot of the log of batch 0

3.3 Skewness and Kurtosis

After performing the Shapiro-Wilk test, we have performed skewness and kurtosis tests, to get some insights in what way the transformed data follows normality. In most cases the skewness is a bit too high, while in other cases the kurtosis is not in the range of

	kurtosis	skewness
B0	-0.703	0.862
B1	-0.773	0.720
B2	-1.306	-0.131
B3	-1.196	-0.062
B4	0.086	0.540
B5	1.123	1.282
B6	1.426	1.094
B7	-0.553	0.130

Kurtosis and skewness tests are also formalized in the Jarque-Bera test, which tests if the skewness is close to zero and the kurtosis close to three. The p-value of the Jarque-Bera test is above 0.01 for all batches, so we cannot reject the normality hypothesis based on skewness and kurtosis. Both the Shapiro-Wilk test and Jarque-Bera indicate that assuming normality on the batches is not rejected and therefore we will assume an underlying normal distribution for the following tests. It must also be noted that the t-test is robust against data that deviates somewhat from normality, so our assumption at least for that test is reasonable.

	Jarque.Bera
B0	0.4640
B1	0.5395
B2	0.6538
B3	0.6898
B4	0.7693
B5	0.2834
B6	0.3850
B7	0.8437

Kurtosis and skewness tests are also formalized in the Jarque-Bera test, which tests if the skewness is close to zero and the kurtosis close to three. The p-value of the Jarque-Bera test is above 0.01 for all batches, so we cannot reject the normality hypothesis based on skewness and kurtosis. Both the Shapiro-Wilk test and Jarque-Bera indicate that assuming normality on the batches is not rejected and therefore we will assume an underlying normal distribution for the following tests. It must also be noted that the t-test is robust against data that deviates somewhat from normality, so our assumption at least for that test is reasonable.It must also be noted that the t-test is robust against data that deviates somewhat from normality, so our assumption at least for that test is reasonable.

3.3.1 T-Test

As can be found in the table above containing the variance of the batches, we can see that the variances of the batches are not equal. We therefore have to perform a t-test that assumes unequal variances. Batch seven has one missing value, which should not be a problem for a t-test.
The results of the t-tests can be found in the following table:

	p.value
B1	0.1986
B2	0.0159
B3	0.0020
B4	0.9395
B5	0.0596
B6	0.2336
B7	0.0145

The results indicate that batch three has a mean that is significanlty different from the mean of the unaffected batch, since its p-value is less than 0.01. This could indicate that this batch is affected.

3.3.2 Homogeneity tests

Performing the Barlett's tests gives us a p-value of $4.5232441 \times 10^{-15} < 0.01$. This rejects the null hypothesis that all the batches have the same variance. After performing an F-test that compares all the batches with batch 0, we have found the following results:

	F.test
B1	0.083
B2	0.486
B3	0.605
B4	0.006
B5	0.020
B6	0.152
B7	0.239

These results indicate that batch four has different properties than the other batches.

3.3.3 Outliers

Both the extreme on the left and the right tail are not significantly big to reject them as an outlier. This indicates that no values can be indicated as errors due mistakes that people maked while measuring the amount of μg in the batches.

	min	p.value
B1	1.04	1.00
B2	0.72	0.91
B3	1.52	0.72
B4	0.56	0.80
B5	1.60	1.00
B6	0.96	1.00
B7	1.28	0.51

	max	p.value
B1	5.92	0.34
B2	24.64	0.67
B3	18.56	0.69
B4	2.24	0.14
B5	7.60	0.07
B6	8.64	0.05
B7	11.60	0.35

3.4 Non-parametric tests

3.4.1 Serial correlation tests with means

We have performed a Rank von Neumann Test for lag-1. The p-value is $0.490 < 0.05$ and we therefore cannot reject the null hypothesis. This indicates that no significant pattern can be found. This could indicate that the silicones have affected only one or two batches and not all batches from the point the leakage has started.

3.4.2 Wald-Wolfowitz Runs Test

We have performed a Wald-Wolfowitz Runs test. We see that the p-value $= 7.1225335 \times 10^{-5} < 0.05$ so we can reject that null hypothesis. This is an interesting result, because we know the data is obtained from the same source.

3.4.3 Kolmogorov-Smirnov Test

The results of the Kolmogorov-Smirnov test can be found in the following table.

	batch	p.value
1	B1	0.0996
2	B2	0.2485
3	B3	0.0337
4	B4	0.5176
5	B5	0.0337
6	B6	0.0996
7	B7	0.1134
8	B7*	0.0996

None of the batches violate the null-hypothesis, since all p-values are larger than 0.01. Note that the missing value in B7 is omitted in the test with B7 and replaced with the mean of all other values in B7 in the test with B7*.

3.4.4 Wilcoxon Rank-Sum Test

	batch	p.value
1	B1	0.1181
2	B2	0.0162
3	B3	0.0033
4	B4	0.8306
5	B5	0.0415
6	B6	0.1070
7	B7	0.0227
8	B7*	0.0201

We can observe that the p-value for batch 3 $= 0.0033165 < 0.01$ and so for batch 3 we reject the null hypothesis. The other batches do not deviate significantly from batch zero.

4 Conclusion