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. As a start, we will plot the data and perform a Shapiro-Wilk test to see whether the non-transformed data is normal. 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; if this is not the case, that could be an indicator that some of the batches are affected by silicones. We reject the null hypothesis if $p < 0.01$.

2.1 Critical value

Most of our hypothesis include multiple comparisons. Because of this reason, we have to adjust the critical value to make sure that it will not reject batches too soon. Applying Bonferroni will give us an $\alpha = \frac{0,05}{7} = 0,00714$, but since this value is quite low, it can be the case that too many batches will pass the test. Because we are dealing with dangerous materials, we will raise the alpha value slightly to prevent batches from passing the test even though they could be infected. To be a bit more strict, we will set the critical value to 0.01.

2.2 Normal distribution

To decide whether the data is normally distributed, we will plot the data to get an insight in how the data is distributed. We will then perform the Shapiro-Wilk tests on each of the batches to check whether the data is now normally distributed. We formulate the hypothesis as following:

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

We will reject H_0 if $p < 0,01$. Besides a Shapiro-Wilk test, we will perform skwewness and kurtosis tests to test on where normality is violated. The values for skewness should be between -0,03 and 0,3, while the values of kurtosis should be between -0,5 and 1,5. 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. If, however, the data is not normally distributed, we will apply a transformation to remove skewness and bring the data closer to a normal distribution. If a transformation is necessary, we must be careful to draw conclusions about the original untransformed data. 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.

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. We will also perform an F-test to see which batches have similar properties as batch 0.

2.3 Non-parametric tests

2.3.1 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 $p < 0.05$.

2.3.2 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:

$$\begin{cases} H_0 : F_0 = F_i & \text{for } 1 \leq i \leq 7 \\ H_1 : F_0 \neq F_i & \text{for } 1 \leq i \leq 7 \end{cases} \tag{3}$$

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.

2.3.3   Wilcoxon Rank-Sum Test

Another test to compare the clean batch with the other batches without any assumption of normality is the Wilcoxon Rank Sum Test. To apply this test, the following assumptions are made:

- The samples have been collected randomly from their respective populations.

- Sample pairs are independent.

- The observed variable is a continuous random variable.

- The underlying distributions from which the samples are derived are identical in shape.

We can make the first two assumptions based on the assignment description. Although the measurements themselves have limited precision, the underlying value is continuous, which justifies the third assumption. For the last assumption, we refer to the similarity of the histograms of the different batches.

2.4   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.

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.
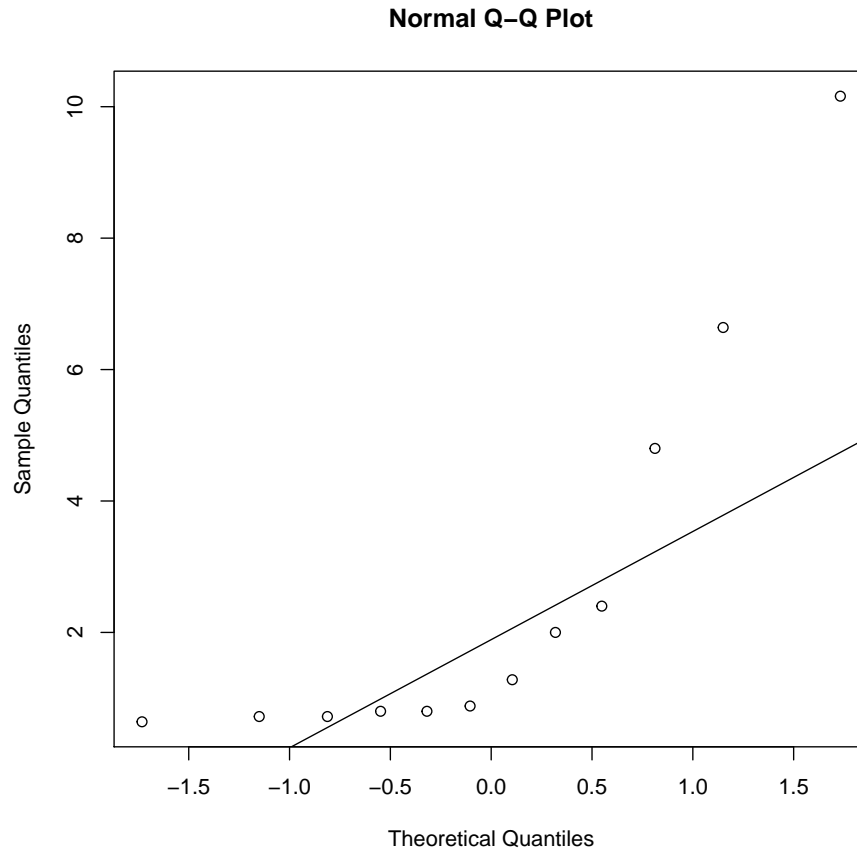
Figure 1: QQ Plot of batch 0

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 the other, possible affected, batches 1 to 7. Of course generating this plot for multiple batches at once assumes that the units are somewhat consistent across different batches.

6
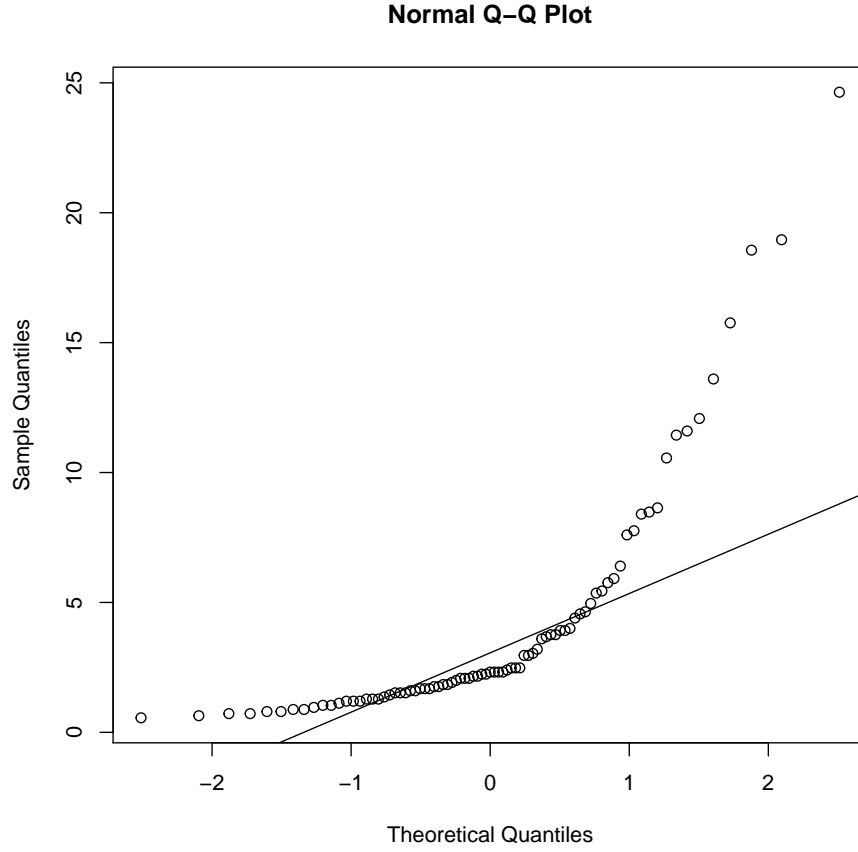
Normal Q–Q Plot

Figure 2: QQ Plot of batches 1 to 7

The normal QQ plot here too seems to suggest that the data is not normally distributed. When we again perform the Shapiro-Wilk normality test we can see the p-value $= 8.1716549 \times 10^{-12} < 0.01$ so we can reject the null-hypothesis that the outcome of batches 1 to 7 is normally distributed.

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.
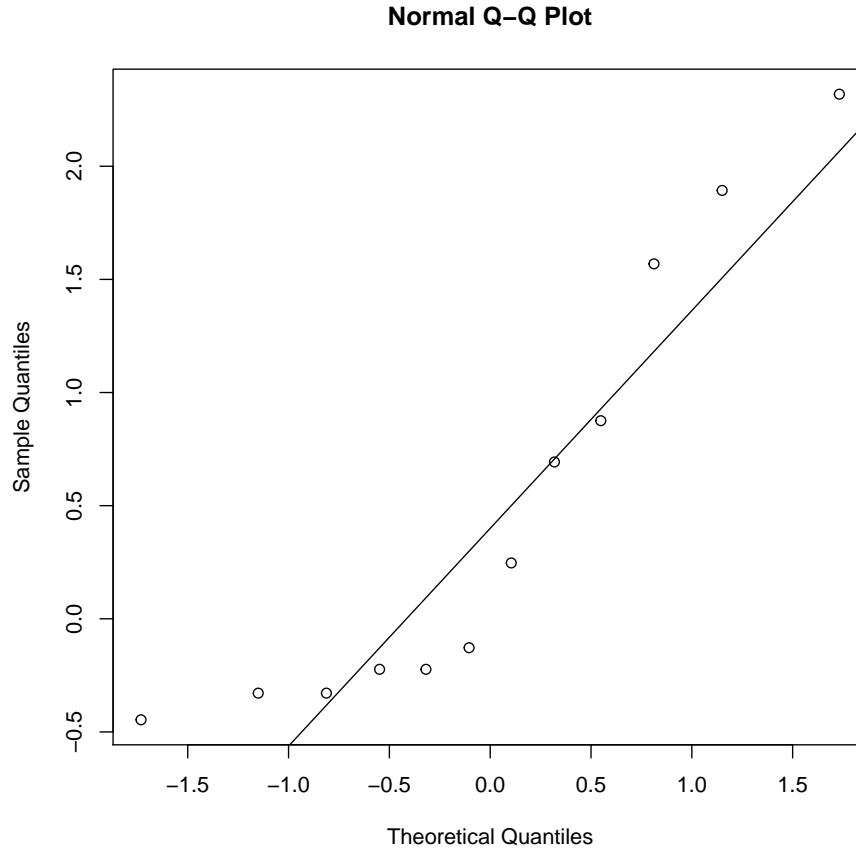
Normal Q–Q Plot

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

```
## Error in skewness(Bi, type = 2):  unused argument (type = 2)
## Error in characters | factors:  operations are possible only for
numeric, logical or complex types
```

This kind of test is 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

	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

hypothesis based on skewness and kurtosis. The data is skewed right, which is probably the reason that the Shapiro-Wilk test rejects the assumption of a normal distribution. The p-value of the Shapiro-Wilk test is not very low and the skewness, kurtosis and Jarque-Bera tests seem to suggest a close-to-normal shape. In combination with the fact that for all the other batches we cannot reject the normality null hypothesis, we will assume that batch 0 is normally distributed 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.

3.3.1   T-Test

To see which t-test variant we need to perform, we need to analyse the variance of all the log-transformed batches.

	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

The variance differs between batches, as can be seen in the preceding table. Because of that, we will 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:
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.

	batch	p.value
1	B1	0.1986
2	B2	0.0159
3	B3	0.0020
4	B4	0.9395
5	B5	0.0596
6	B6	0.2336
7	B7	0.0145

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.

3.4 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 $0.490 > 0.01$ 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 = $7.1225335 \times 10^{-5} < 0.01$ 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 Rank Serial Correlation Test

Here we perform the rank serial autocorrelation test at lag 1 using the rank von Neumann ratio. This tests the null hypothesis that the lag-k autocorrelation is 0 for all values of k greater than 0 (i.e., the time series is purely random). The procedure that we use emits some warnings because the dataset contains ties, and with this specific test that can make the p-value less accurate. With that warning in mind, we present the table containing the rank serial autocorrelation test p-values of each batch:

	batch	p.value
1	B0	0.9902
2	B1	0.1305
3	B2	0.2639
4	B3	0.3569
5	B4	0.6808
6	B5	0.2062
7	B6	0.4889
8	B7	0.3671

As one can observe, the p-values are all large, which means that we cannot reject the null hypothesis that the batches are random using this test.

3.4.4 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*.

3.4.5 Wilcoxon Rank-Sum Test

Since we reject batches if the p-value < 0.01, only batch 3 fails to pass the null hypothesis. The other batches do not deviate much from batch zero.

	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

3.5 Outliers

	min	p.value
B0	0.64	1.00
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
B0	10.16	0.25
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

4 Conclusion