Chapter 4 Homework

Cody Frisby

September 28, 2016

4.5

$$H_0 : \mu = 13.4 \times 1000 \text{ Å } vs \text{ } H_A : \mu \neq 13.4 \times 1000 \text{ Å}$$

Assuming the data follows a normal distrubution, running a t test for the above stated hypothesis, produces $t = -3.0905771$ with $p = 0.0129174$, thus rejecting the null hypothesis at $\alpha = 0.05$. A 99% percent confidence interval would be

$$[1.3392163 \times 10^4, 1.3400197 \times 10^4]$$

noticing that 13.4×10^3 is contained within this interval.

The normality assumption appears to be reasonable here. The Shapiro-Wilk test for normailty shows $p = 0.6331371$, thus we DO NOT reject the null hypothesis that the data is from the normal distribution. Additionally the normal quantile probability plot below indicates that the data does not stray too far from the line:
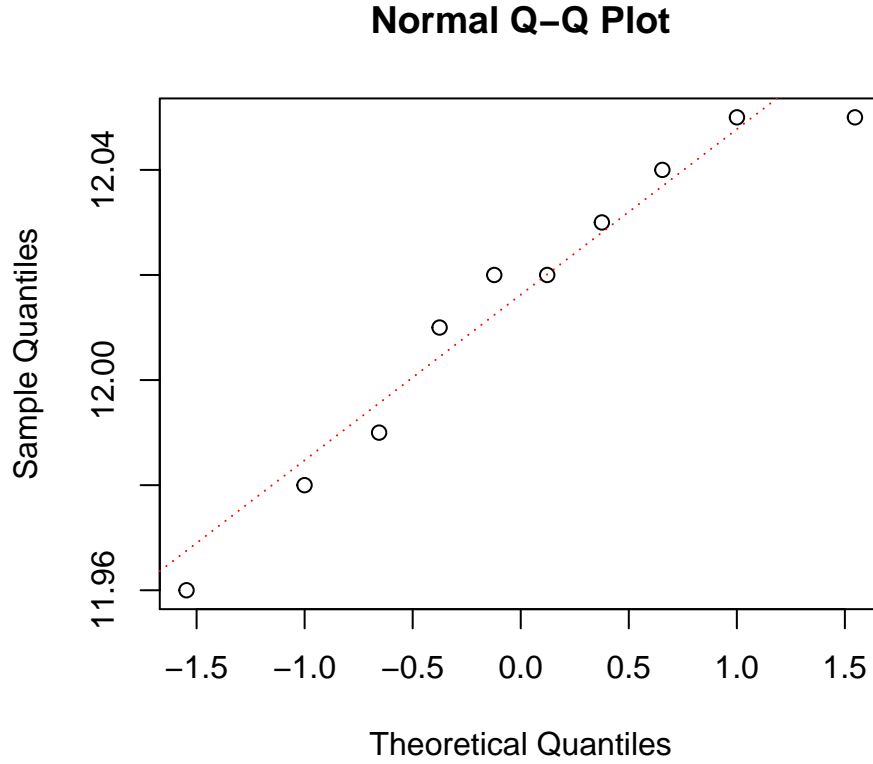
Normal Q–Q Plot
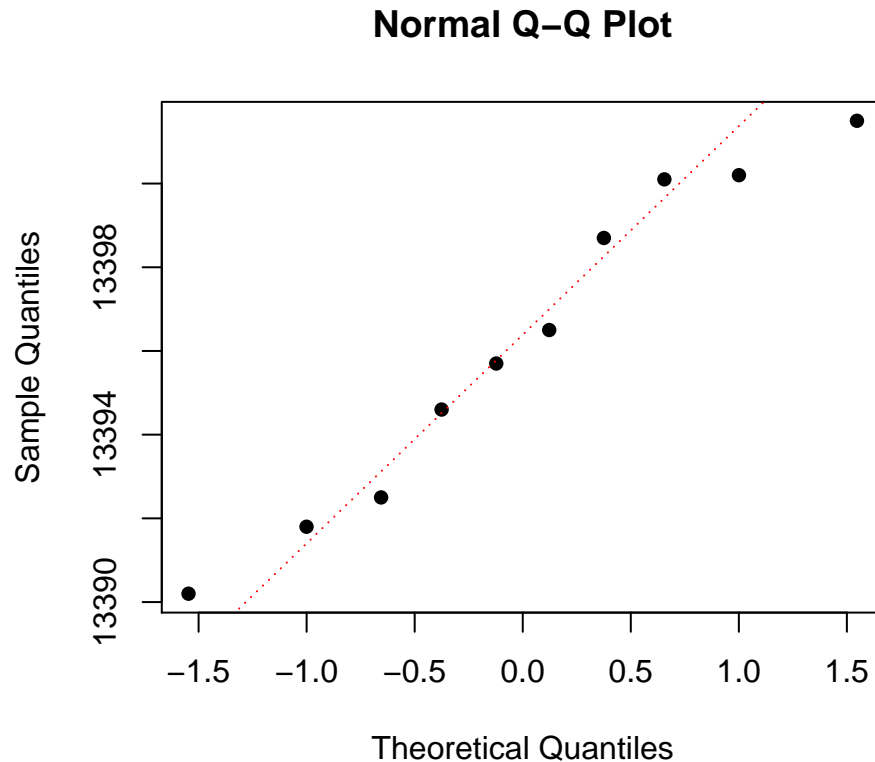
4.6

At $\alpha = 0.01$ it appears we cannot conclude that the mean does not equal 12, $p = 0.0758137$.

Two-sided 95% confidence interval is [11.9838852, 12.0461148]. And as can be seen, 12 is contained in this interval, confirming our conclusions above where we cannot reject the null hypothesis.

Assumption of normality does NOT appear to be violated:

Normal Q–Q Plot

[Normal Q-Q Plot with Sample Quantiles on y-axis (11.96, 12.00, 12.04) and Theoretical Quantiles on x-axis (-1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5)]

4.9

The test statistic for $H_0 : \mu = 12$, -6.9691253, and the confidence interval for the mean is $[9.7270192, 10.7917308]$

The hypothesis for the variance part of the problem can be stated by

$$H_0 : \sigma^2 = 1 \ vs \ H_A : \sigma^2 \neq 1$$

And the test statistic can be computed as

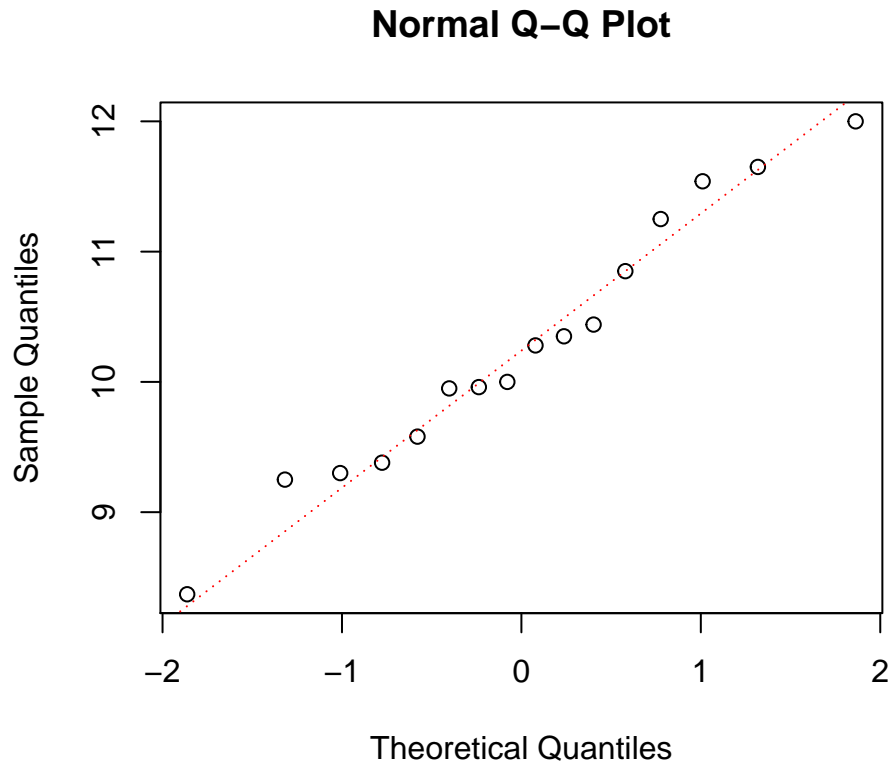$$T = (N-1)\frac{s^2}{\sigma_0^2} = (16-1)\frac{0.9980996}{1^2} = 14.9714937$$

Computing a 95% confidence interval for the standard deviation of our sample, assuming normality, is

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}} \leq \sqrt{\sigma^2} \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}$$

$$[0.7380026, 1.5462199]$$

An upper 95% confidence interval is equal to 1.4359391.

Assumption of normality does NOT appear to be violated:

Normal Q–Q Plot

4.11

Assuming normality, and testing that the mean measurements of the two technicians are equal, we get a p value of 0.9171497. There is NOT eveidence to reject the null hypothesis that the means are equal.

If the means were equal here, we might conclude (or at the very least be suspicious) that the two technicians do not measure the surface finish the same.

A 95% confidence interval from our test is [-0.127969, 0.1411833].

Looking at the two samples side by side might indicate that there is not a big difference between the two techs
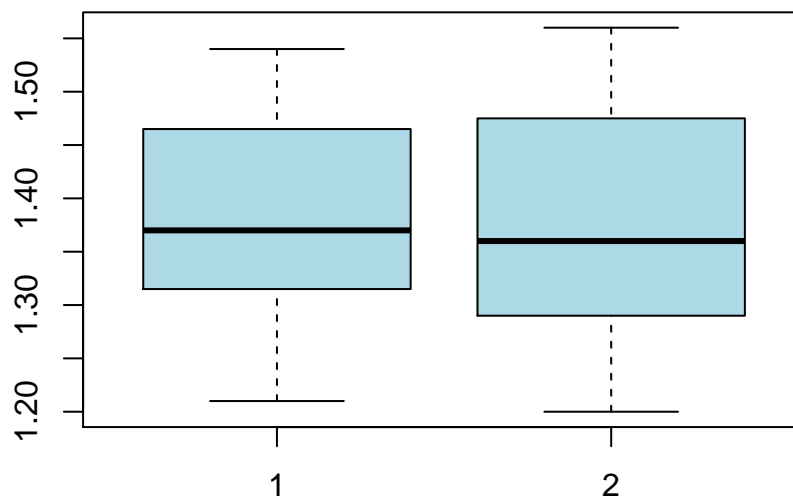
Visually, looking at the boxplot above, it does not appear that the variances are different. To be sure, running a variance test comparing the two samples we get a test statistic and p-value of 0.8456401 and 0.8540325 respectivley and we **DO NOT** reject.

The 95% confidence interal is [0.1652094, 4.8163185] and seeing as 1 is included in this interval, it confirms our conclusions.

And, like 4.9 above, a 95% confidence interval for σ^2 is [0.0039721, 0.0174361].

Assumption of normality does NOT appear to be violated:

Normal Q–Q Plot
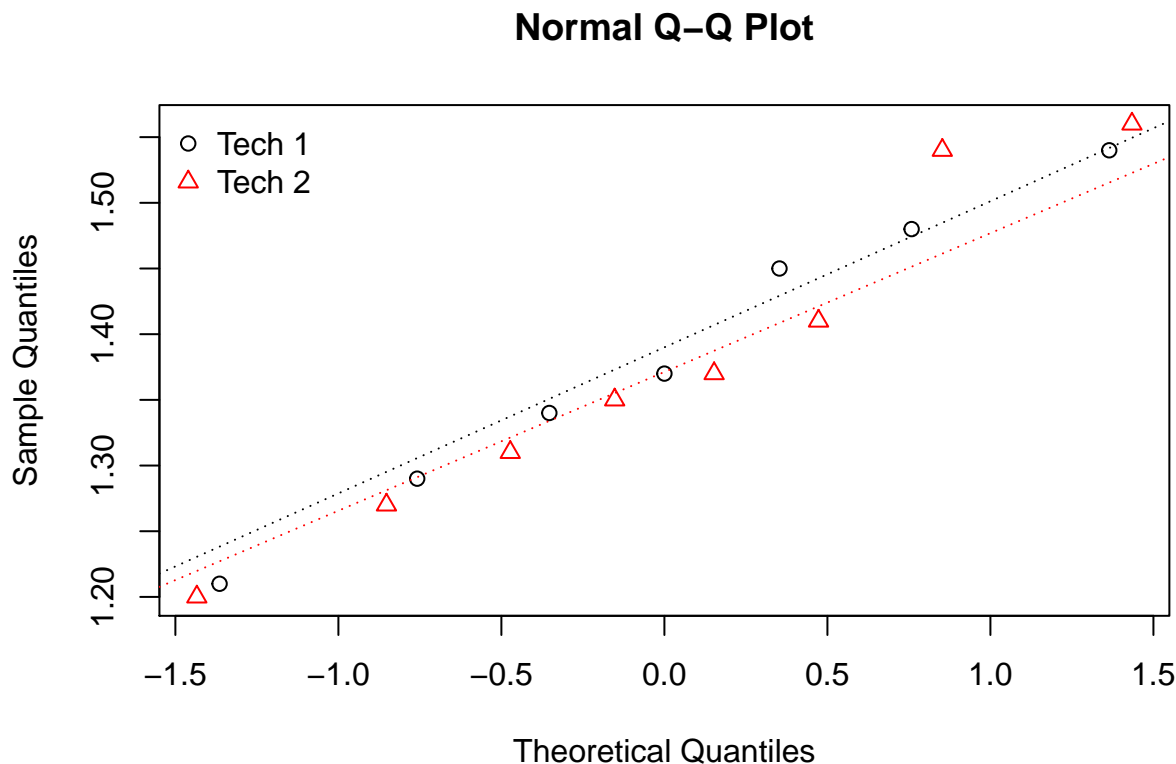
This plot can also be useful is acessing whether or not the variances are equal. The two lines are parallell and are approximately the same slope, demonstrating that the variances can be assumed equal.

4.12

First, we define the pooled variance for n_1 and n_2

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

and using this values as an estimate for the pooled variance we develop the CI for $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2}, n_1 + n_2 - 2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 + \bar{x}_2 - t_{\frac{\alpha}{2}, n_1 + n_2 - 2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

But here we have the formual for the CI if we are assuming **EQUAL** variances. Damn, unequal variances is a little more complex, we have to define another variable, v

$$v = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}} - 2$$

Now using this calculation in the above formula in place of S_p we get

$$\bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2}, n_1 + n_2 - 2} v \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2}, n_1 + n_2 - 2} v \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

4.13

$$H_0 : \mu_{salt} = \mu_{oil} \ vs \ H_A : \mu_{salt} \neq \mu_{oil}$$

Testing this hypothesis we get $t = -0.7707996$ the probability of this t statistic is 0.4508255. We **DO NOT** reject the null hypothesis.

A 95% confidence interval is [-6.7061516, 3.1061516]. Confirming the above conclusions, 0 is in this interval.

We cannot conclude that the variances are **NOT** equal. The confidence interval for this test is [0.205816, 3.3359952]. **1** is contained in this interval, thus we conclude there's isn't any evidence that the two variances are not equal.

The assumption of equal variances and normality are valid.

Normal Q–Q Plot
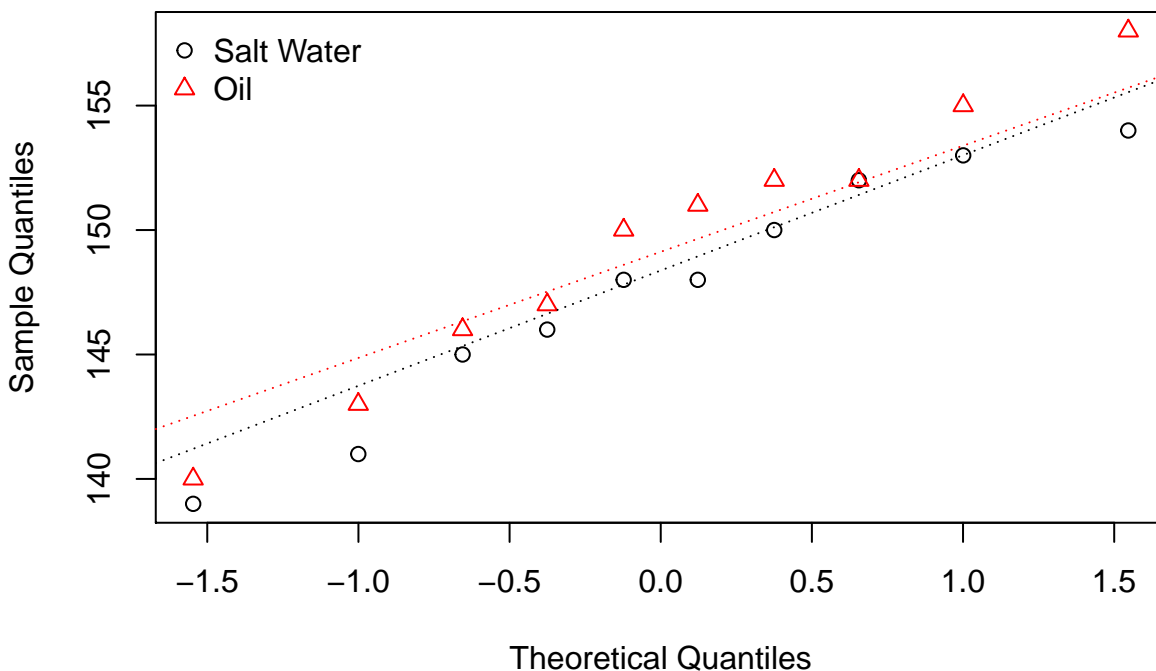
4.14

Here, our hypothesis is

$$H_0 : p = 0.1 \ vs \ H_A : p \neq 0.1$$

Computing the estimate for p, $\hat{p}$, is simply $\frac{18}{200} = 0.09$.
Using the normal approximate to the binomial

$$Z_0 = \frac{x - 0.5 - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{18 - 0.5 - 200 * 0.1}{\sqrt{200 * 0.1 * 0.9}} = -0.5892557$$

which is a well inside the critical value of 1.96.

Constructing a confidence interval we use the formula

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and we can see that 0.1 is inside the interval [0.0567146, 0.1232854].

4.15

Here our hypothesis is

$$H_0 : p = 0.08 \quad vs \quad H_A : p \neq 0.08$$

Computing the estimate for p, $\hat{p}$, is simply $\frac{65}{500} = 0.13$.
Using the normal approximate to the binomial

$$Z_0 = \frac{x - 0.5 - np_0}{\sqrt{np_0(1-p_0)}} = \frac{65 - 0.5 - 500 * 0.08}{\sqrt{500 * 0.08 * 0.92}} = 4.0387054$$

which is a well outside the critical value of 1.96.

Computing the pvalue for our statistic, 4.0387054, we get 2.6873509×10^{-5}. We reject the null hypothesis. A 95% upper confidence interval is 0.1547385.

4.17

In this example we have paired data. If we can still assume normality, we can run a paired t test to test our hypothesis

$$H_0 : \mu_1 - \mu_2 = 0$$

Computing a test statistic for test if the variances are equal can be accomplished with

$$F_0 = \frac{6.79}{6.18} = 1.0987055$$

and computing the probability of this value we get 0.5441956 and so we cannot reject the null hypothesis that the variances are equal.

Can we conclude that the new purification device has reduced the mean percentage of impurity?

$$t_0 = \frac{9.85 - 8.08}{\sqrt{\frac{6.79}{10} + \frac{6.18}{8}}} = 1.4691453$$

This does not appear to be a very large test statistic, but checking against the critical value under the t distribution with 16 degrees of freedom is 2.1199053 which is greater than 1.4691453 so we **DO NOT** reject the null.

4.19
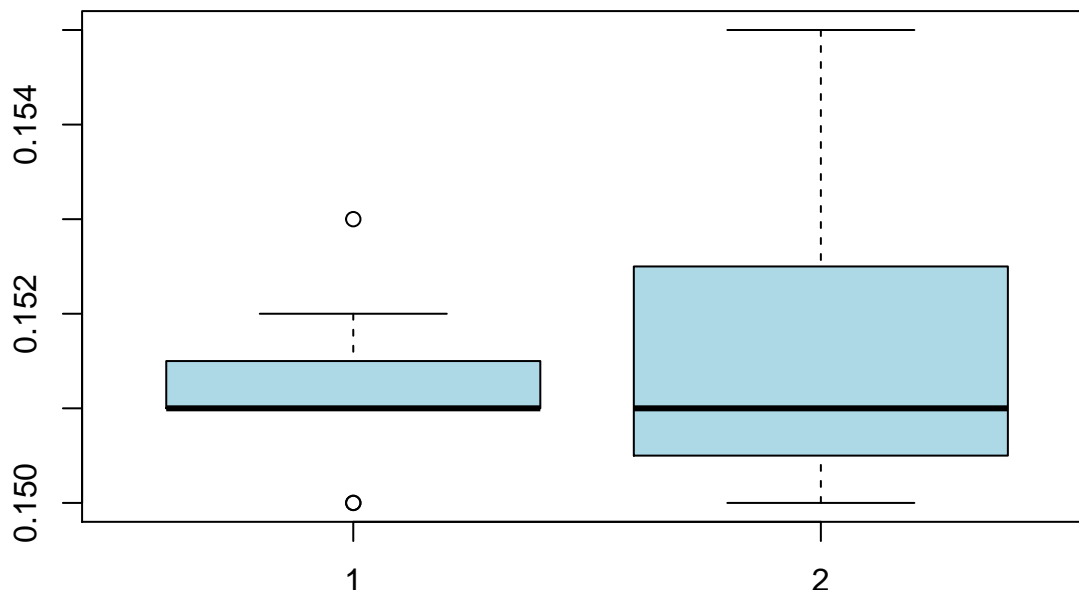
NO, there doesn't appear to be evidence that there's a difference between the *mean* measurements of the calipers. Although, there may be evidence to suggest their variances are NOT equal. From a quality control viewpoint, we may want to investigate the vernier calipers, or perhaps not use them at all.

4.26

If $X \sim \text{Poi}(\lambda)$ the $E[X] = \lambda$. And for large values of λ $X \sim N(\mu = \lambda, \sigma^2 = \lambda)$. Let Y = mean no. of defects per glass bottle, $\frac{X}{n}$. We now have a test that is similar to the test from 4.14.

Using this we test the hypothesis at $\alpha = 0.05$

$$H_0 : y = 0.15$$

$$Z_0 = \frac{x - 0.5 - y}{\sqrt{y(1 - y)}} = \frac{11 - 0.5 - 100 \times 0.15}{\sqrt{100(0.15)(1 - 0.15)}} = -1.2602521$$

which is well within the critical value of -2.5758293.

Computing the pvalue for our statistic, -1.2602521, we get 0.1037892. We **DO NOT** reject the null hypothesis. A 95% confidence interval is 0.0294049, 0.1905951.

4.29

Flow rate appears to have a small effect etch uniformity

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
rate	2	3.647778	1.8238889	3.585627	0.0533651
Residuals	15	7.630000	0.5086667	NA	NA

although we would not reject at $\alpha = 0.05$. Having a "small" p-value and by visual inspection of the box plot,

we might suspect that there is an effect and perhaps design some follow up experiments. Setting flow rate to 125 does appear to affect etch uniformity for the best (smaller values being better).
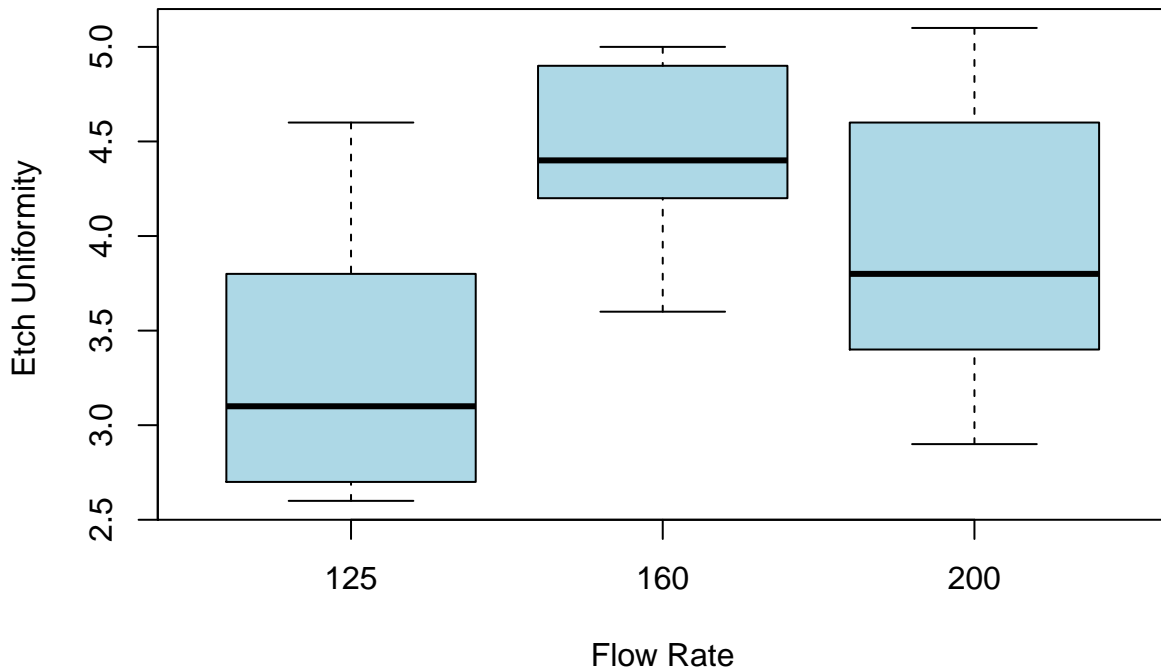
Taking a look at the models residuals here
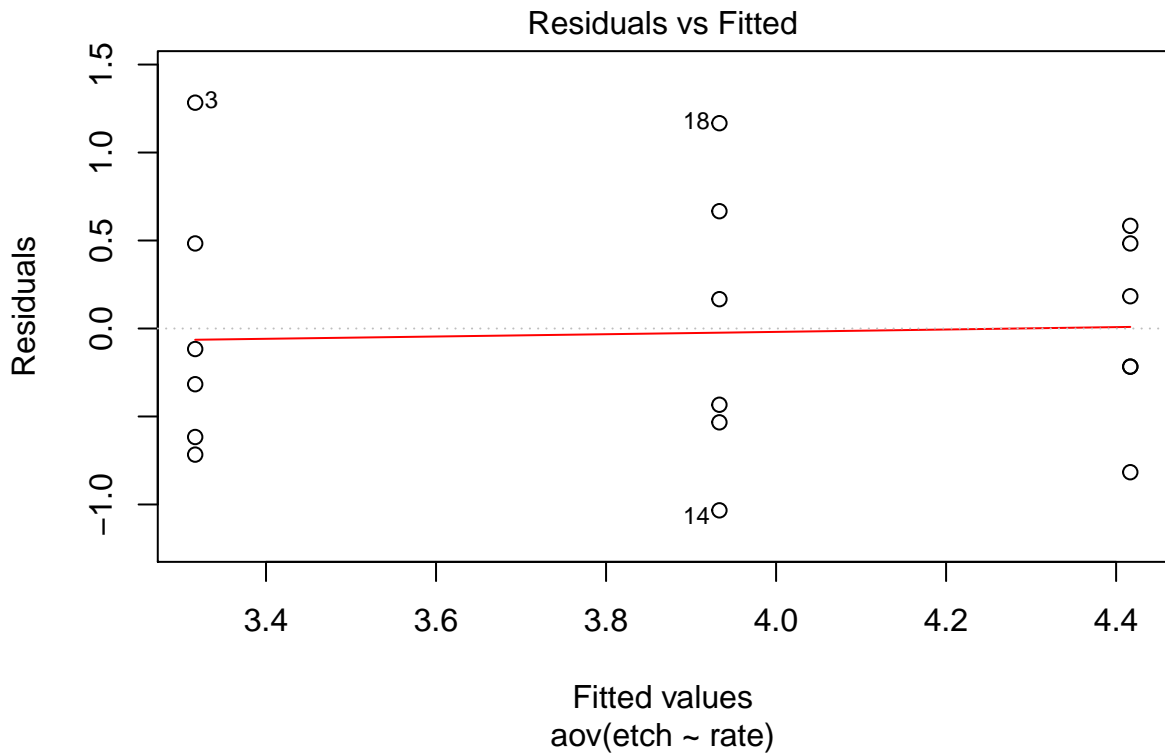
we see that the assumption of equal variances is probably OK. And the normality assumption can be assesed by visual inspection of the plot below.
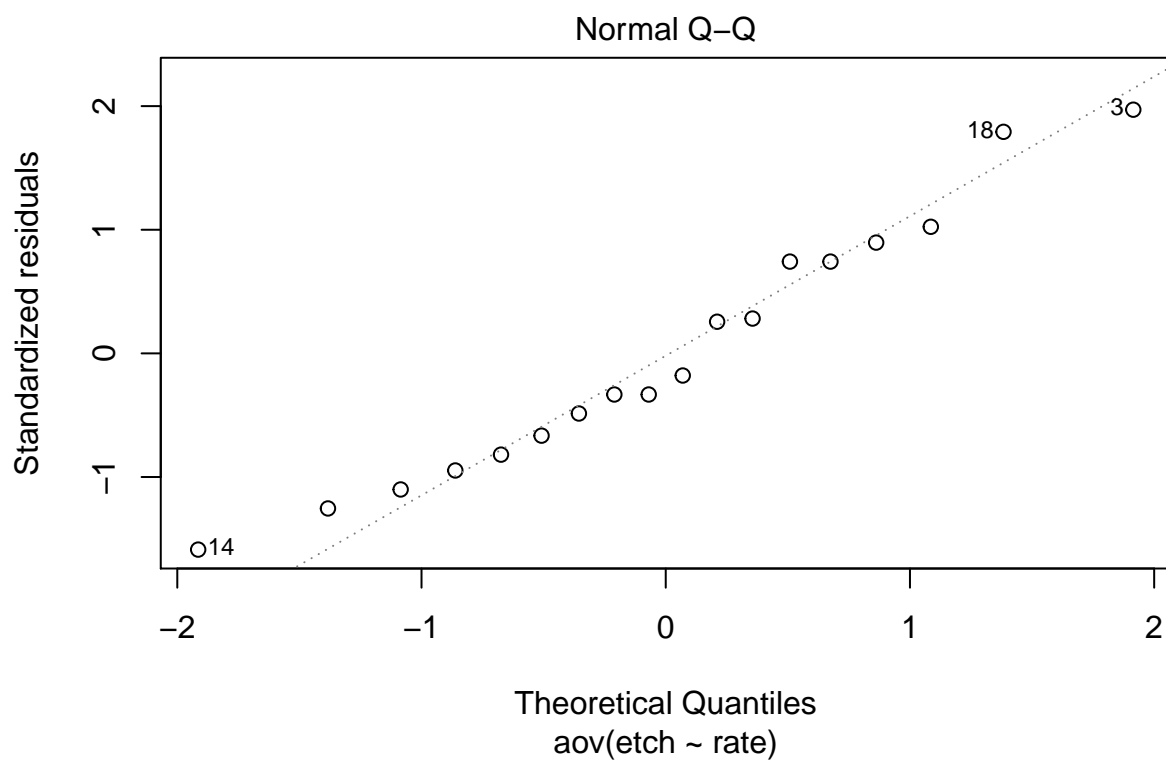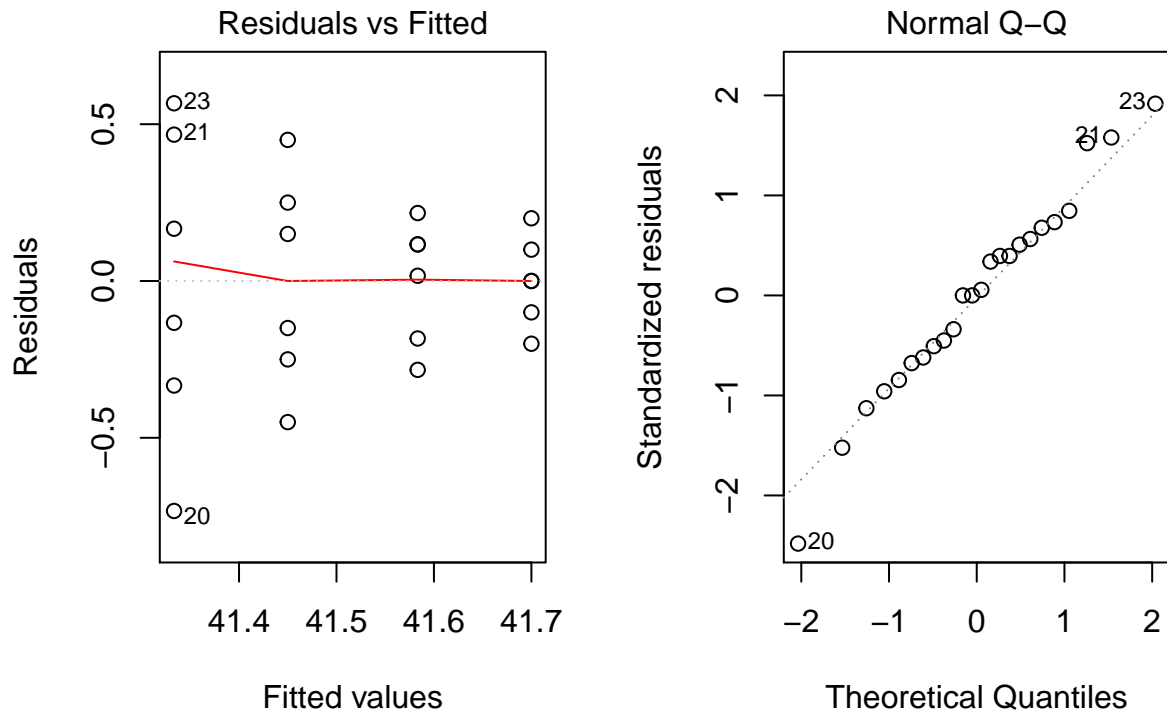
Normal Q–Q

Normality assumption is OK.

4.33

Firing temperature does not appear to have a large effect on density of the anode.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	3	0.4566667	0.1522222	1.45204	0.2576107
Residuals	20	2.0966667	0.1048333	NA	NA

The equality of variances assumption (left plot) appears to be an issue here. Normality assumption appears to be OK (right plot).
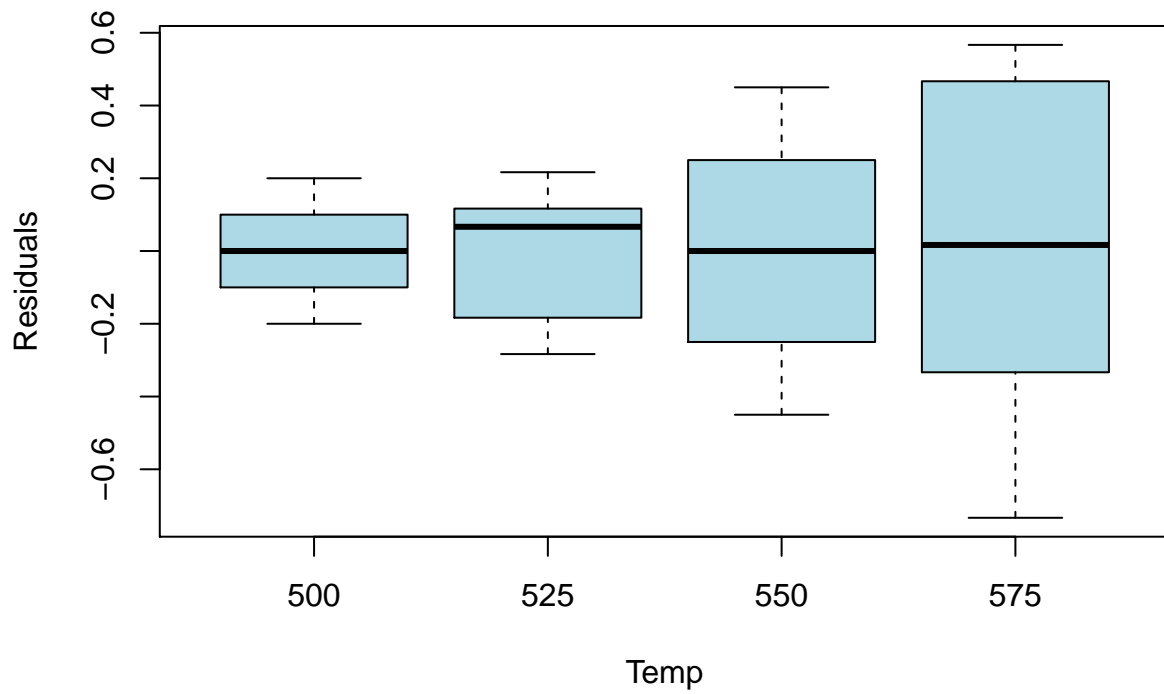
My suspicion is that the lower the baking temperature the better the anode life.
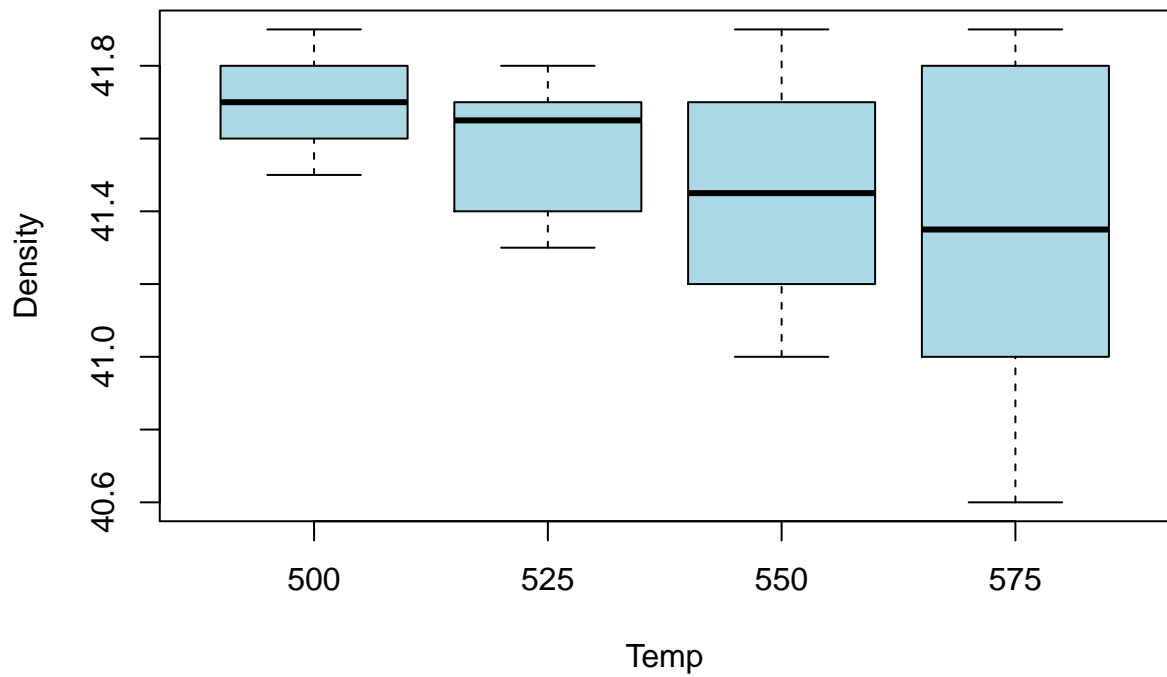
My suspicion seems valid after examing the box plots. Additionally, there may be less variability with the lower temps as well, which is a plus for quality.
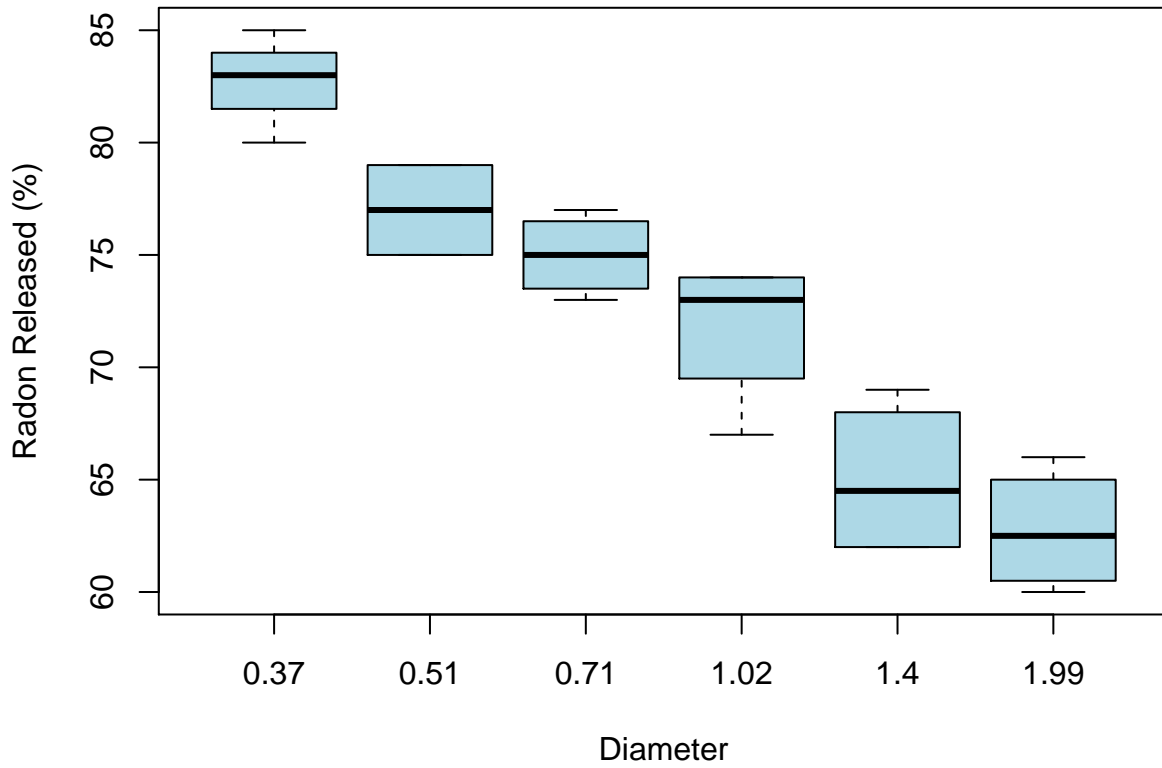
4.34

Taking a look at the residuals plotted by firing temp from the previous example

there does appear to be a trend in variability as the temp increases. We definitely would reccomend using 500.

4.35

First taking a look at a plot of the data:

We can immediately see a trend. Fitting an ANOVA model (assuming equal variance and normality) we get

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diam	5	1133.375	226.675000	30.8518	0
Residuals	18	132.250	7.347222	NA	NA

The size of the orifice appears to have a strong affect on mean percentage of radon released.

The wider the opening the less radon is released. If we wish to minimize the release of radon then we would want the orifice opening to be at least 1.4 if not 1.99.