Fluidics Comp 2.0

Ryan Gallagher

2025-01-21

Purpose

The lab uses a series of machines when processing samples into genetic data. The final step, the Scanner, offers quality metrics that are used to measure how successful a run was. The purpose of this report is to analyze sample data, plus their QC metrics, to determine what (if any) factors might influence these QC metrics.

Data

Variables This document will explore data from the `Day 4` sheet of the `Clinical QC.xlsx`. This sheet has Quality Control (QC) metrics from our Clinical QC database. The fields of interest are:

Field	Description
Plate #	Clinical batch of up to 11 samples and the CytoRef103 (a control on every batch)
Date	Either Start or End Date (Ask Dima)
Kit ID #	These are the reagents. The ID is the date that the reagents are received and then "*1" through "*_4" (they come in batches of 4).
Chip Lot #	This number is on the Kits, but supplied by Thermo. The Chip Lot # & the Kit ID # (before the "_") should give the same information.
Hybrid Machine	Either Sam or Frodo. Sam is mostly used.
Fluidics Machine	These machines wash the chips. We use 3,4,6,7,8. There are 4 modules per machine.
Fluidics Module	These are modules in the machine. There are 4 per.
Scanner	After being washed and stainedin the fluidics machines, these Scanners scan the chips and collect Copy Number data (SNPs/CN).
MapdQC	a metric for all probes used to determine copy number that is derived from both polymorphic (SNP) and non-polymorphic (CN) probes. Recommended threshold: MAPD <= 0.25
SnpQC	a metric for SNP probes that is derived from polymorphic (SNP) probes. Recommended threshold SnpQC >= 15.0.
Waviness QC	a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. Recommended threshold WavinessQC <= 0.12.
Pass/Fail	Based on QC metric thresholds in the Thermo manual. If any of the QC metrics fail, then this field is marked 'Fail'.

Distributions A contingency table is displayed which shows how many of each Fluidics Machine + Module combinations are represented in our data. We then run a Chi-Squared test to see if the selection of Machine + Module are indeed random. A p-value greater than 0.05 shows that the selection of Module is independent of the Machine selected.

Table 2: Contingency Table: Fluidics Machine and Module (Chi-squared p-value: 1)

	Module			
	1	2	3	4
3	32	33	33	33
4	37	36	37	37
6	21	21	23	23
7	28	28	28	27
8	25	24	25	25

Here, we check if the selection of `Scanner` is independent of the `Fluidics.Machine` used.

Table 3: Contingency Table: Scanner and Fluidics Machine (Chi-squared p-value: 0.0158)

	Fluidics Machine					
	3	4	6	7	8	Total
Thing1	57	85	51	68	63	**324**
Thing2	74	62	37	43	36	**252**
Total	**131**	**147**	**88**	**111**	**99**	**576**

We obtain a significant Chi-Squared p-value which indicates that there is an associated between `Fluidics.Machine` used and `Scanner` selected. Further, we can observe the residuals to determine which `Fluidics.Machines` are over/under represented per `Scanner`.

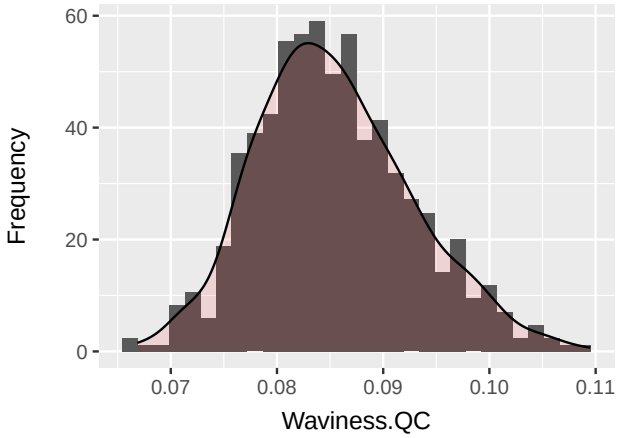
Table 4: Standardized Residuals for Scanner vs Fluidics Machine

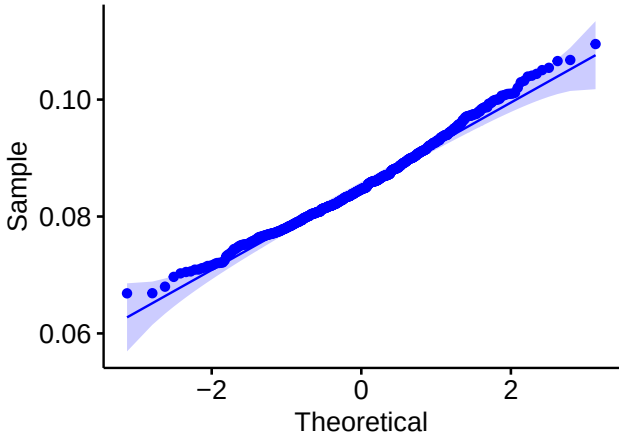
Scanner	Fluidics.Machine	Residual
Thing1	3	-1.9439915
Thing2	3	2.2042792
Thing1	4	0.2543090
Thing2	4	-0.2883594
Thing1	6	0.2132007
Thing2	6	-0.2417469
Thing1	7	0.7039588
Thing2	7	-0.7982143
Thing1	8	0.9799119
Thing2	8	-1.1111156

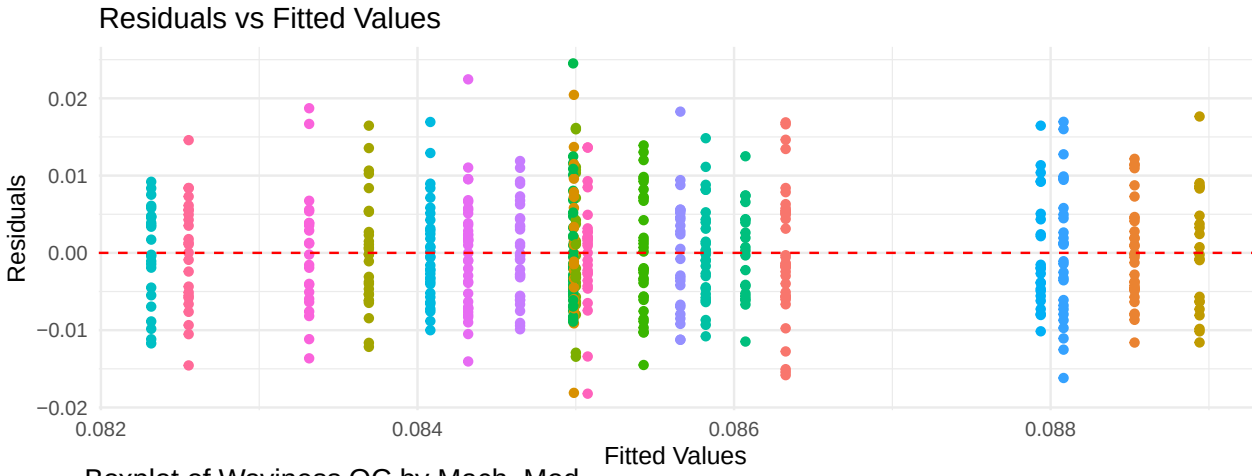
We find that Machine 3 samples are more often put onto Thing 2, and Machine 8 samples are more often put onto Thing 1.

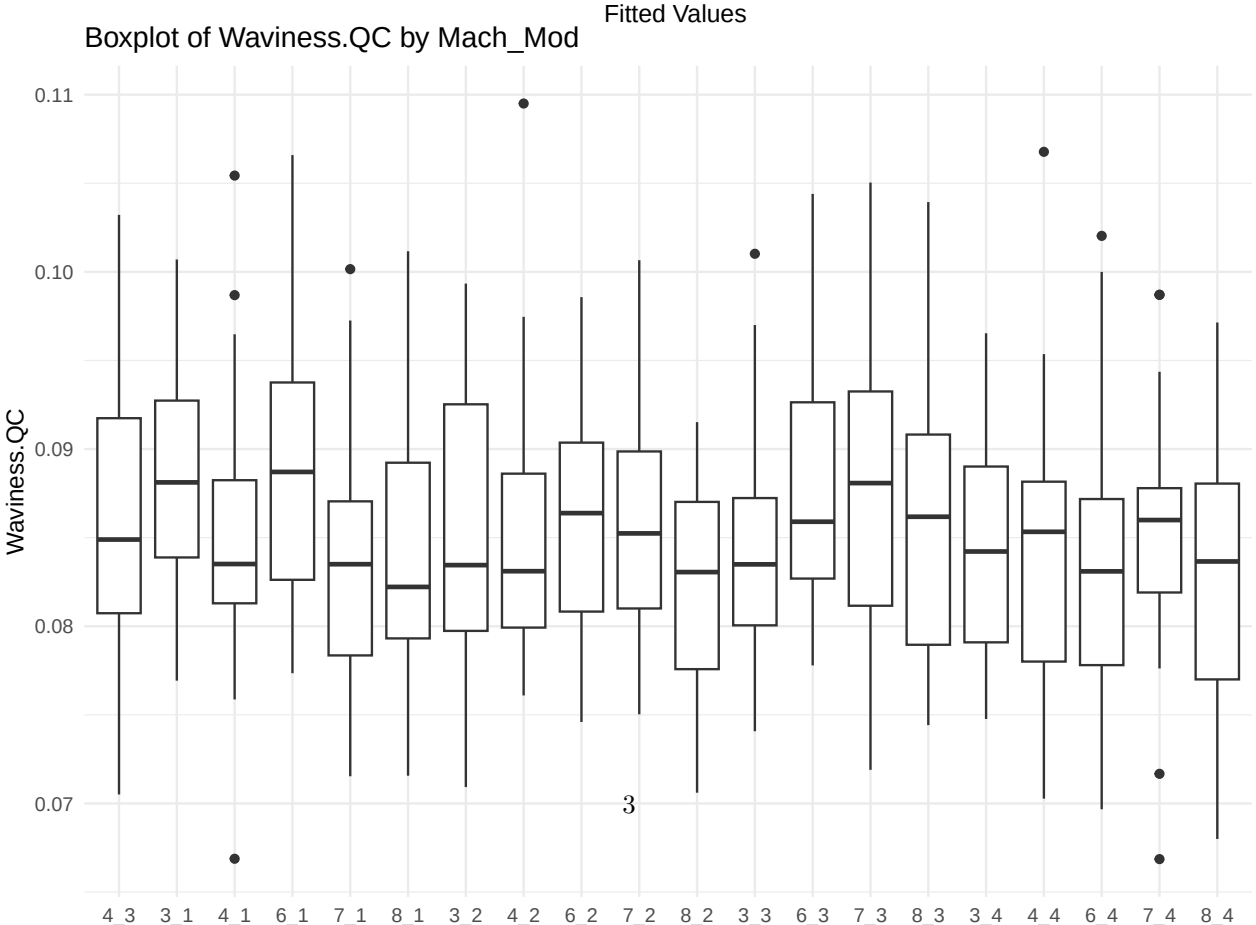
Univariate Analysis

Machine&Module ~ Waviness.QC







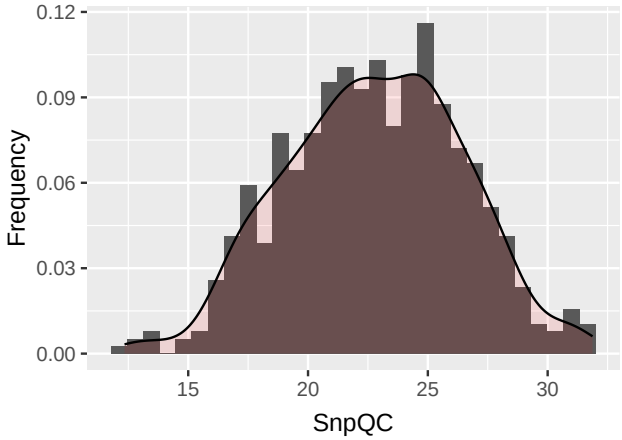


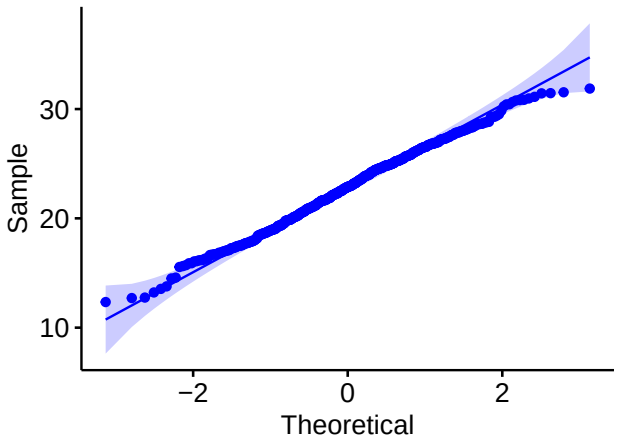
```
## # A tibble: 20 x 3
##    Mach_Mod   mean      sd
##    <fct>     <dbl>   <dbl>
##  1 4_3      0.0863 0.00842
##  2 3_1      0.0885 0.00644
##  3 4_1      0.0850 0.00733
##  4 6_1      0.0889 0.00784
##  5 7_1      0.0837 0.00723
##  6 8_1      0.0850 0.00814
##  7 3_2      0.0854 0.00772
##  8 4_2      0.0850 0.00737
##  9 6_2      0.0861 0.00601
## 10 7_2      0.0858 0.00641
## 11 8_2      0.0823 0.00637
## 12 3_3      0.0841 0.00629
## 13 6_3      0.0879 0.00732
## 14 7_3      0.0881 0.00866
## 15 8_3      0.0857 0.00736
## 16 3_4      0.0846 0.00610
## 17 4_4      0.0843 0.00725
## 18 6_4      0.0833 0.00787
## 19 7_4      0.0851 0.00702
## 20 8_4      0.0826 0.00706
```

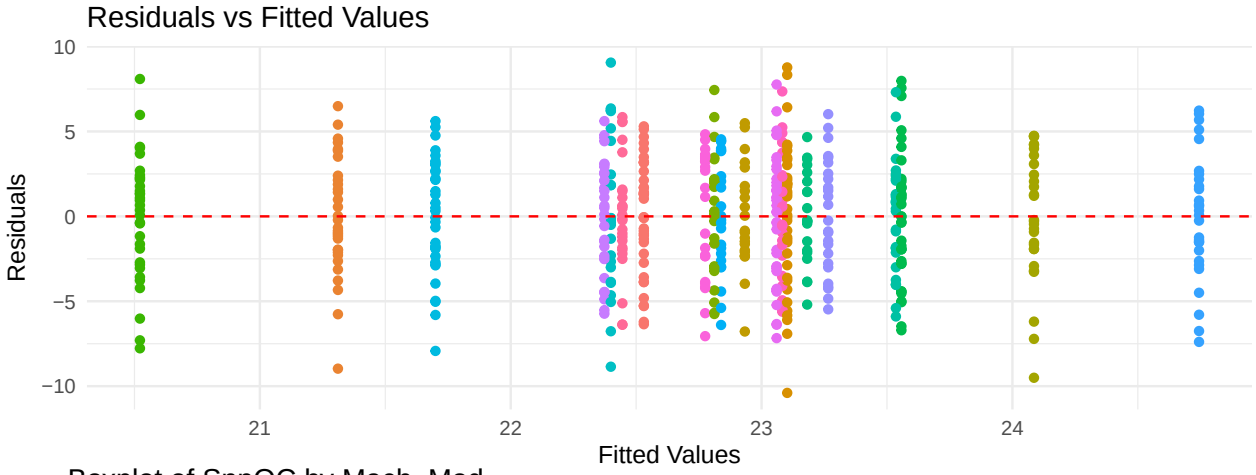
```
##              Df   Sum Sq   Mean Sq F value Pr(>F)
## Mach_Mod     19 0.001724 9.073e-05   1.723 0.0289 *
## Residuals   556 0.029271 5.265e-05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

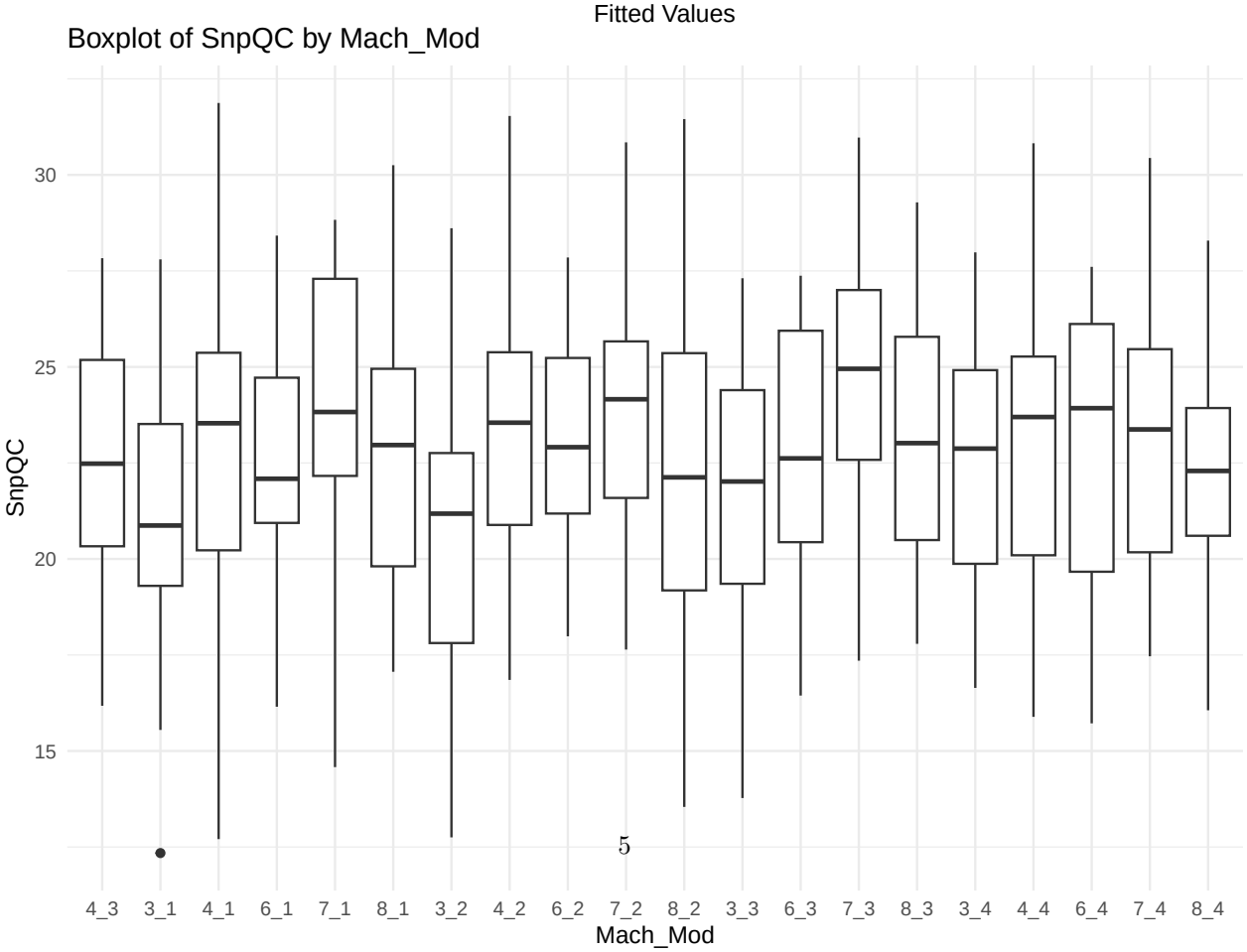
Our assumptions to conduct an ANOVA are satisfied. We find that the `Waviness.QC` metric is very slightly influenced by the `Fluidics.Machine+Module` chosen, though the effect size that is being detected is certainly small and inconsequential to the `Pass/Fail` requirement.

Machine&Module ~ SnpQC







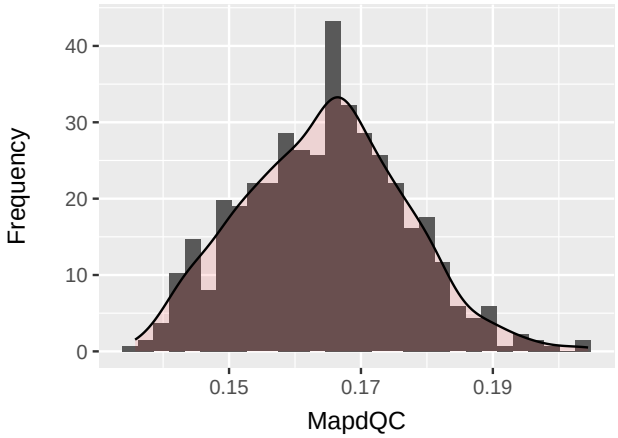


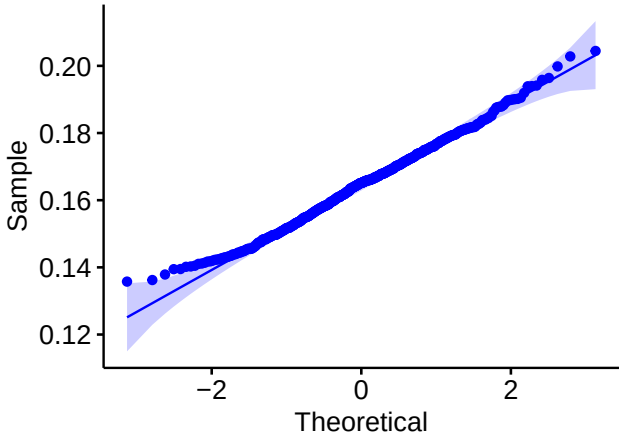
```
## # A tibble: 20 x 3
##    Mach_Mod  mean    sd
##    <fct>    <dbl> <dbl>
##  1 4_3       22.5  3.33
##  2 3_1       21.3  3.41
##  3 4_1       23.1  4.24
##  4 6_1       22.9  3.06
##  5 7_1       24.1  3.76
##  6 8_1       22.8  3.58
##  7 3_2       20.5  3.60
##  8 4_2       23.6  3.80
##  9 6_2       23.2  2.68
## 10 7_2       23.5  3.17
## 11 8_2       22.4  4.59
## 12 3_3       21.7  3.41
## 13 6_3       22.8  3.33
## 14 7_3       24.7  3.71
## 15 8_3       23.3  3.35
## 16 3_4       22.4  3.28
## 17 4_4       23.1  3.71
## 18 6_4       22.8  3.73
## 19 7_4       23.1  3.65
## 20 8_4       22.4  3.35
```

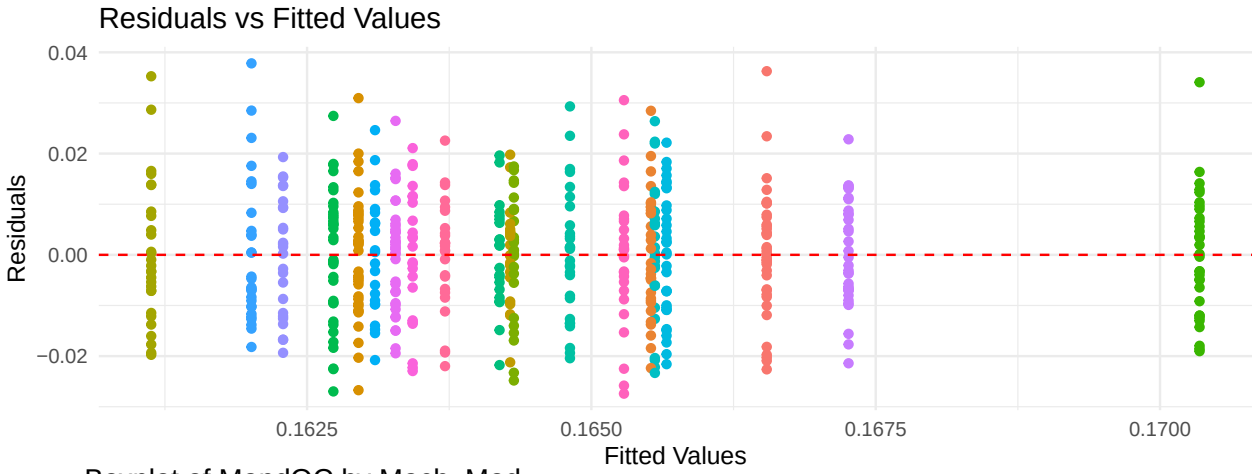
```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Mach_Mod     19    502   26.45   2.064 0.00517 **
## Residuals   556   7124   12.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

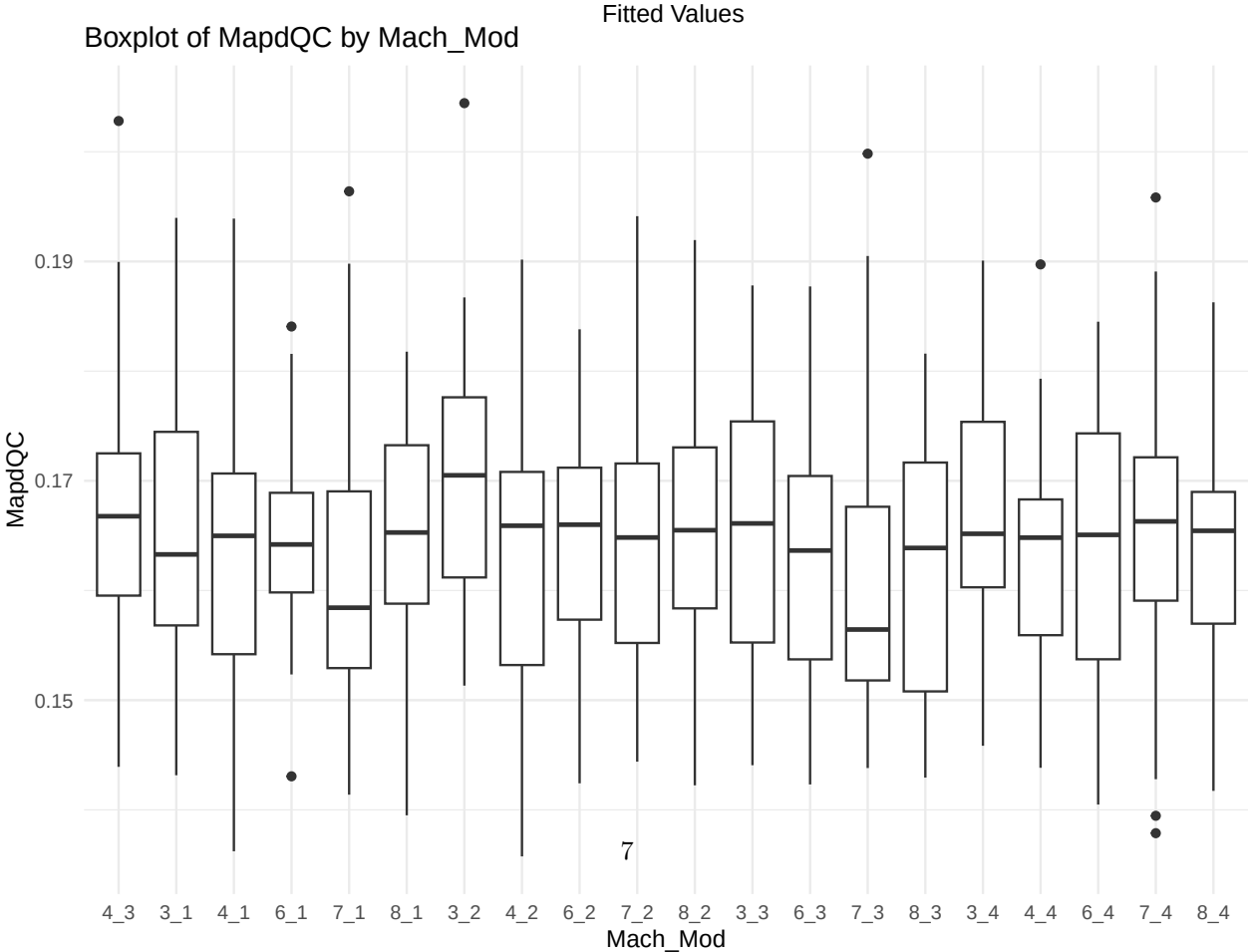
Our assumptions to conduct an ANOVA are satisfied. We find that Machine3, Module1 and Machine3, Module 2 show reasonable different mean values for SnpQC. With the threshold for this variable being 15, this range is still acceptable for these modules.

Machine&Module ~ MapdQC









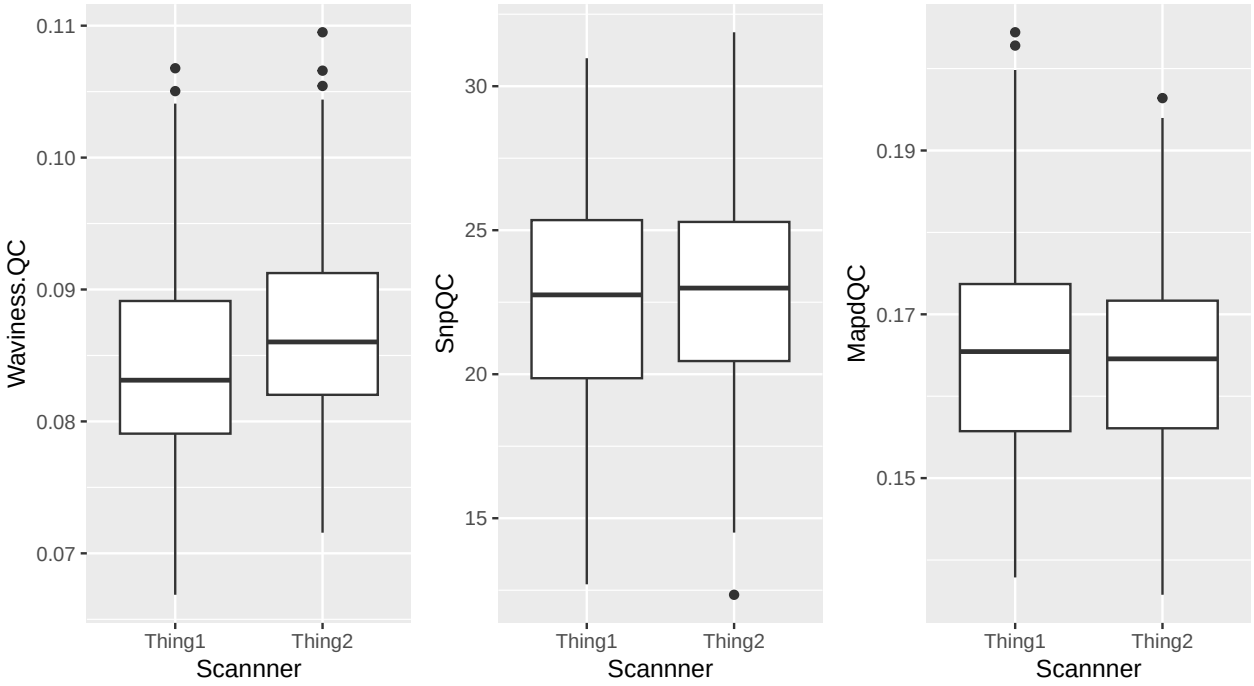
```
## # A tibble: 20 x 3
##    Mach_Mod  mean      sd
##    <fct>    <dbl>   <dbl>
##  1 4_3      0.167 0.0123
##  2 3_1      0.166 0.0118
##  3 4_1      0.163 0.0119
##  4 6_1      0.164 0.00961
##  5 7_1      0.161 0.0138
##  6 8_1      0.164 0.0121
##  7 3_2      0.170 0.0118
##  8 4_2      0.163 0.0130
##  9 6_2      0.164 0.0103
## 10 7_2      0.165 0.0129
## 11 8_2      0.166 0.0139
## 12 3_3      0.166 0.0124
## 13 6_3      0.163 0.0118
## 14 7_3      0.162 0.0141
## 15 8_3      0.162 0.0116
## 16 3_4      0.167 0.0104
## 17 4_4      0.163 0.0103
## 18 6_4      0.163 0.0132
## 19 7_4      0.165 0.0138
## 20 8_4      0.164 0.0108
```

```
##              Df  Sum Sq   Mean Sq F value Pr(>F)
## Mach_Mod     19 0.00262 0.0001379   0.932  0.543
## Residuals   556 0.08225 0.0001479
```

Our assumptions to conduct an ANOVA are satisfied. Our ANOVA test achieves an insignificant p-value, which would indicate that these groups have the same mean according to our data. Visually, it appears that Machine3, Module2 have present a higher mean MapdQC (though not by much). We would expect this after observing the SnpQC measurements since MapdQC and SnpQC are correlated variables (we see this in our data, as well as in the description of the variable).

Scanner

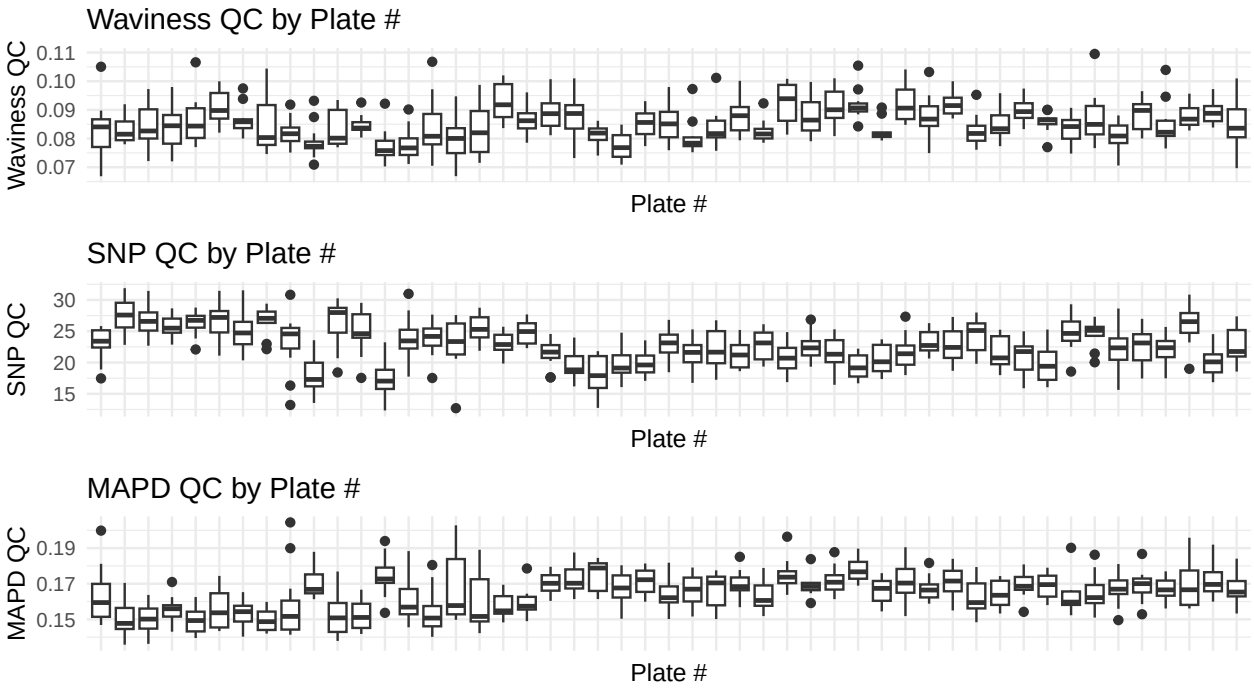
In a previous report using CytoRef103 data from 2024, we found that the `Scanner` is not a significant predictor of our QC variables. We explore that conclusion in this data. Recall that the sample size for `Thing 1` is 324 and `Thing 2` is 252.



We observe a consistent interpretation for `SnpQC` and `MapdQC`, however it now appears that `Waviness.QC` shows a distinct difference in distribution between both `Scanners`. We will observe this further in the following section.

Mixed Model

We fit a model that features our QC metric as the dependent variable. We set `Scanner`, `Machine+Module`, and `Position` as **fixed effects**, and `Plate #` as a **random effect**. We set `Plate #` as a random effect because there is natural variation in these metrics just based on the plates/batches run. That is displayed in the following:



Waviness.QC

```
model.wav = lmer((Waviness.QC) ~ Scannner + Mach_Mod + Position + (1|`Plate.#`), data=data)
summary(model.wav)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: (Waviness.QC) ~ Scannner + Mach_Mod + Position + (1 | `Plate.#`)
##    Data: data
##
## REML criterion at convergence: -3793.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3475 -0.6304 -0.1332  0.6312  3.3150
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  Plate.#  (Intercept) 1.140e-05 0.003376
##  Residual             3.933e-05 0.006272
## Number of obs: 576, groups:  Plate.#, 49
##
## Fixed effects:
##                  Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)     8.481e-02  1.356e-03  5.018e+02  62.546  < 2e-16 ***
## ScannnerThing2  2.026e-03  6.166e-04  5.434e+02   3.286  0.00108 **
## Mach_Mod3_1     1.679e-03  1.560e-03  4.997e+02   1.076  0.28224
## Mach_Mod4_1    -1.596e-03  1.521e-03  4.958e+02  -1.049  0.29467
## Mach_Mod6_1     2.204e-03  1.801e-03  5.114e+02   1.224  0.22156
## Mach_Mod7_1    -2.466e-03  1.662e-03  5.103e+02  -1.484  0.13847
## Mach_Mod8_1    -1.023e-03  1.723e-03  5.111e+02  -0.593  0.55314
## Mach_Mod3_2    -1.892e-03  1.584e-03  5.011e+02  -1.195  0.23281
## Mach_Mod4_2    -1.623e-03  1.509e-03  4.964e+02  -1.076  0.28259
## Mach_Mod6_2    -1.437e-04  1.781e-03  5.120e+02  -0.081  0.93575
## Mach_Mod7_2     6.925e-05  1.650e-03  5.098e+02   0.042  0.96655
## Mach_Mod8_2    -3.729e-03  1.752e-03  5.108e+02  -2.129  0.03375 *
## Mach_Mod3_3    -2.661e-03  1.624e-03  5.015e+02  -1.639  0.10193
## Mach_Mod6_3     8.642e-04  1.731e-03  5.088e+02   0.499  0.61773
## Mach_Mod7_3     1.649e-03  1.638e-03  5.099e+02   1.006  0.31466
## Mach_Mod8_3    -8.287e-04  1.717e-03  5.113e+02  -0.483  0.62949
## Mach_Mod3_4    -2.249e-03  1.614e-03  5.006e+02  -1.393  0.16418
## Mach_Mod4_4    -2.194e-03  1.503e-03  4.952e+02  -1.460  0.14504
## Mach_Mod6_4    -3.700e-03  1.739e-03  5.086e+02  -2.127  0.03387 *
## Mach_Mod7_4    -1.010e-03  1.640e-03  5.093e+02  -0.616  0.53835
## Mach_Mod8_4    -3.019e-03  1.699e-03  5.103e+02  -1.777  0.07625 .
## Position1       1.851e-03  1.346e-03  4.966e+02   1.375  0.16960
## Position2      -8.310e-04  1.366e-03  4.991e+02  -0.608  0.54323
## Position3       1.914e-04  1.374e-03  4.995e+02   0.139  0.88927
## Position4       1.580e-03  1.356e-03  4.991e+02   1.165  0.24441
## Position5       5.384e-04  1.338e-03  4.990e+02   0.403  0.68746
## Position6      -9.528e-04  1.351e-03  4.980e+02  -0.705  0.48085
## Position7       3.056e-03  1.358e-03  4.966e+02   2.251  0.02484 *
## Position8       1.129e-03  1.381e-03  4.979e+02   0.818  0.41399
## Position9       1.844e-03  1.392e-03  4.982e+02   1.324  0.18602
```

```
## Position10     -4.428e-05  1.362e-03  4.960e+02  -0.033  0.97408
## Position11      1.692e-03  1.326e-03  4.966e+02   1.276  0.20246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation matrix not shown by default, as p = 32 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

SnpQC

```
model.snp = lmer((SnpQC) ~ Scannner + Mach_Mod + Position + (1|`Plate.#`), data=data)
summary(model.snp)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: (SnpQC) ~ Scannner + Mach_Mod + Position + (1 | `Plate.#`)
##    Data: data
##
## REML criterion at convergence: 2832.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0637 -0.6320  0.0329  0.6658  2.5044
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Plate.#  (Intercept) 5.746    2.397
##  Residual             7.098    2.664
## Number of obs: 576, groups:  Plate.#, 49
##
## Fixed effects:
##                 Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)     23.06881    0.63880 325.37462  36.113  < 2e-16 ***
## ScannnerThing2   0.17351    0.26818 525.64750   0.647 0.517904
## Mach_Mod3_1     -1.57773    0.66343 497.83846  -2.378 0.017777 *
## Mach_Mod4_1      0.09081    0.64647 496.17410   0.140 0.888342
## Mach_Mod6_1     -0.66340    0.76913 503.50196  -0.863 0.388807
## Mach_Mod7_1      0.35028    0.70936 502.90853   0.494 0.621661
## Mach_Mod8_1     -0.19232    0.73588 503.37042  -0.261 0.793929
## Mach_Mod3_2     -2.35291    0.67404 498.47195  -3.491 0.000524 ***
## Mach_Mod4_2      0.79323    0.64141 496.47152   1.237 0.216778
## Mach_Mod6_2     -0.47612    0.76089 503.77624  -0.626 0.531769
## Mach_Mod7_2     -0.16071    0.70441 502.68120  -0.228 0.819621
## Mach_Mod8_2     -0.56188    0.74786 503.19111  -0.751 0.452814
## Mach_Mod3_3     -1.27194    0.69128 498.66263  -1.840 0.066365 .
## Mach_Mod6_3     -0.81209    0.73834 502.08223  -1.100 0.271911
## Mach_Mod7_3      1.09383    0.69909 502.70549   1.565 0.118292
## Mach_Mod8_3      0.01918    0.73307 503.43808   0.026 0.979136
## Mach_Mod3_4     -0.85349    0.68672 498.20565  -1.243 0.214508
## Mach_Mod4_4      0.51991    0.63869 495.92593   0.814 0.416025
## Mach_Mod6_4     -1.03607    0.74190 501.98031  -1.397 0.163176
## Mach_Mod7_4     -0.50922    0.69972 502.44185  -0.728 0.467101
```

```
## Mach_Mod8_4     -0.73789    0.72536 502.97230  -1.017 0.309515
## Position1       -0.28580    0.57192 496.52613  -0.500 0.617492
## Position2       -0.29466    0.58099 497.75293  -0.507 0.612256
## Position3        0.26576    0.58438 497.97516   0.455 0.649463
## Position4        0.11039    0.57654 497.75718   0.191 0.848233
## Position5        0.69038    0.56891 497.74363   1.214 0.225514
## Position6       -0.42944    0.57425 497.18520  -0.748 0.454913
## Position7       -0.22036    0.57705 496.52721  -0.382 0.702720
## Position8       -0.62511    0.58724 497.14137  -1.064 0.287629
## Position9       -0.09417    0.59207 497.27765  -0.159 0.873693
## Position10       0.77838    0.57875 496.25286   1.345 0.179264
## Position11       0.40631    0.56361 496.58477   0.721 0.471309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation matrix not shown by default, as p = 32 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

MapdQC

```
model.map = lmer((MapdQC) ~ Scannner + Mach_Mod + Position + (1|`Plate.#`), data=data)
summary(model.map)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: (MapdQC) ~ Scannner + Mach_Mod + Position + (1 | `Plate.#`)
##    Data: data
##
## REML criterion at convergence: -3268.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0248 -0.6433 -0.1281  0.5878  3.8696
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  Plate.#  (Intercept) 4.929e-05 0.007020
##  Residual             9.933e-05 0.009967
## Number of obs: 576, groups:  Plate.#, 49
##
## Fixed effects:
##                   Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)     1.648e-01  2.252e-03  4.269e+02  73.183   <2e-16 ***
## ScannnerThing2 -9.219e-04  9.939e-04  5.357e+02  -0.928   0.3540
## Mach_Mod3_1     1.398e-04  2.481e-03  4.988e+02   0.056   0.9551
## Mach_Mod4_1    -2.410e-03  2.418e-03  4.963e+02  -0.997   0.3194
## Mach_Mod6_1     2.065e-04  2.871e-03  5.070e+02   0.072   0.9427
## Mach_Mod7_1    -2.419e-03  2.648e-03  5.062e+02  -0.913   0.3614
## Mach_Mod8_1    -5.061e-04  2.747e-03  5.068e+02  -0.184   0.8539
## Mach_Mod3_2     4.919e-03  2.520e-03  4.998e+02   1.952   0.0515 .
## Mach_Mod4_2    -3.831e-03  2.399e-03  4.967e+02  -1.597   0.1109
## Mach_Mod6_2     5.426e-04  2.840e-03  5.074e+02   0.191   0.8486
```

```
## Mach_Mod7_2     1.353e-03  2.630e-03  5.059e+02   0.515   0.6071
## Mach_Mod8_2     3.257e-04  2.792e-03  5.066e+02   0.117   0.9072
## Mach_Mod3_3     9.224e-04  2.584e-03  5.000e+02   0.357   0.7213
## Mach_Mod6_3    -3.073e-04  2.757e-03  5.051e+02  -0.111   0.9113
## Mach_Mod7_3    -2.120e-03  2.610e-03  5.059e+02  -0.812   0.4171
## Mach_Mod8_3    -2.827e-03  2.737e-03  5.069e+02  -1.033   0.3020
## Mach_Mod3_4     2.466e-03  2.567e-03  4.994e+02   0.960   0.3373
## Mach_Mod4_4    -2.962e-03  2.389e-03  4.959e+02  -1.240   0.2157
## Mach_Mod6_4     3.451e-04  2.771e-03  5.049e+02   0.125   0.9009
## Mach_Mod7_4     1.408e-03  2.613e-03  5.055e+02   0.539   0.5903
## Mach_Mod8_4    -1.041e-03  2.708e-03  5.063e+02  -0.384   0.7009
## Position1       9.495e-04  2.139e-03  4.968e+02   0.444   0.6573
## Position2       3.652e-04  2.172e-03  4.986e+02   0.168   0.8665
## Position3      -1.064e-03  2.185e-03  4.989e+02  -0.487   0.6264
## Position4       2.076e-03  2.156e-03  4.986e+02   0.963   0.3361
## Position5      -1.058e-03  2.127e-03  4.986e+02  -0.497   0.6192
## Position6       1.111e-03  2.147e-03  4.978e+02   0.517   0.6051
## Position7       1.526e-03  2.158e-03  4.968e+02   0.707   0.4799
## Position8       6.145e-04  2.196e-03  4.977e+02   0.280   0.7797
## Position9       1.106e-03  2.214e-03  4.979e+02   0.499   0.6177
## Position10     -2.629e-03  2.165e-03  4.964e+02  -1.215   0.2251
## Position11      1.927e-03  2.108e-03  4.969e+02   0.914   0.3610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation matrix not shown by default, as p = 32 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```