passing plans analysis

Note: All analysis and plotting performed using R 2.14.0 for Windows 64 bit.[1]

I should note that all the tests performed have low power due to the small sample size.

Table 1 below presents the data for the ratio of measured to calculated doses for the passing plans.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| HT Teflon | HT Cheese | RA Teflon Including | RA Cheese Including | RA Teflon Excluding | RA Cheese Excluding |
| 0.99896 | *1.04525* | 0.99349 | 1.01322 | 0.99349 | 1.01322 |
| 0.98553 | 0.99157 | 1.00172 | *1.03931* | 0.99867 | 1.01583 |
| 1.00469 | 1.00583 | 0.99867 | 1.01583 | *0.95622* | 1.01890 |
| *1.02362* | 1.00883 | 1.01179 | *1.05953* | 0.99982 | *1.03144* |
| 1.00147 | 0.98986 | *0.95622* | 1.01890 | 1.00138 | 1.00760 |
| 0.98906 | 0.99980 | 0.99982 | *1.03144* | 1.00404 | 1.01506 |
| 0.99350 | 1.01278 | 1.00138 | 1.00760 | 0.99284 | 1.00068 |
| 0.98705 | 0.99750 | 1.00404 | 1.01506 | 1.00176 | 1.01503 |
| 0.99203 | 0.99096 | 0.99284 | 1.00068 | 0.98262 | 0.99538 |
|  |  | 1.00176 | 1.01503 |  |  |
|  |  | 0.98262 | 0.99538 |  |  |
|  |  | 0.99663 | 1.02826 |  |  |

Table 1. The ratios of measured to calculated point doses for each modality and phantom for the passing plans. Italicized values represent outlying data points.

The italicized values are henceforth referred to as the “outliers”. Note that these are also the points that are out of the acceptable range except for the “1.02362” data point in the HT Teflon dataset. I am not sure, but I think we attributed this to a high gradient region. I am not sure about including them or removing them (with or without keeping the pairing).

I look at the basic statistics of each dataset: number of samples, mean, standard deviation, variance, skewness, and kurtosis. Note: these are all sample statistics not population statistics. This is given in Table 2 below. Some of the plans for HT did not have both upper and lower ion chamber measurements. “including” and “excluding” refer to whether or not the corresponding plans were included or excluded.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | # Samples | Mean | Std. | Variance | Skewness[2] | Kurtosis[2] |
| HT Teflon | 9 | 0.99732 | 0.01184 | 0.00014 | 1.21200 | 0.70567 |
| HT Cheese | 9 | 1.00471 | 0.01727 | 0.00030 | 1.50906 | 1.46632 |
| RA Teflon including | 12 | 0.99508 | 0.01414 | 0.00020 | -1.82326 | 2.94706 |
| RA Cheese including | 12 | 1.02002 | 0.01753 | 0.00031 | 0.81338 | 0. 28838 |
| RA Teflon excluding | 9 | 0.99232 | 0.01502 | 0.00023 | -1.72500 | 1.82494 |
| RA Cheese excluding | 9 | 1.01257 | 0.01049 | 0.00011 | 0.01748 | -0.25774 |

Table 2. Statistical summary of the passing plan data sets.

Some points to consider.

1. The small sample sizes.
2. The standard deviation and variance are computed using N-1 as the denominator.
3. The corresponding variances between Teflon and Cheese phantoms can vary by factors of ~1.5 to ~2.
4. Overall, the skewness appears to deviate from 0.
5. Overall, the kurtosis appears to deviate leptokurticly from 0.

With the outliers and above points, I should consider testing for equivalent variances, normality, and alternative non-parametric tests.

Note on HT data: Removing the outliers reduces the skewness and kurtosis.

Note on RA data: Removing the outliers reduces the skewness and kurtosis.

Outlier points are not the same patient. So the outlier issue is not just a matter of dropping a patient and continuing on. Instead I have to accept them putting more weight on the non-parametric tests.

I look at normal probability qq-plots for each of the data sets to get a graphical sense of the data as well as qualitatively checking it for normality. Figures 1-6 give the qq plots for each dataset. The labeled purple points are the outliers. The label gives the corresponding measured to calculated ratio for the data point as listed in Table 1. The line drawn is such that it passes through the first and third quantile. Deviations from this line indicate deviations from normality. For the most part, the data follow a linear pattern with the exception of the outliers.

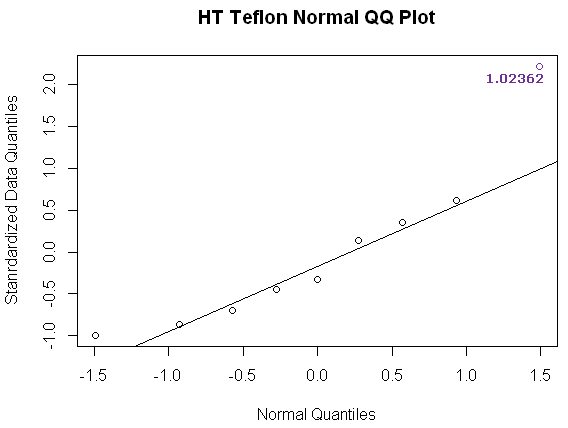


Figure 1. Normal QQ Plot for the HT Teflon dataset.

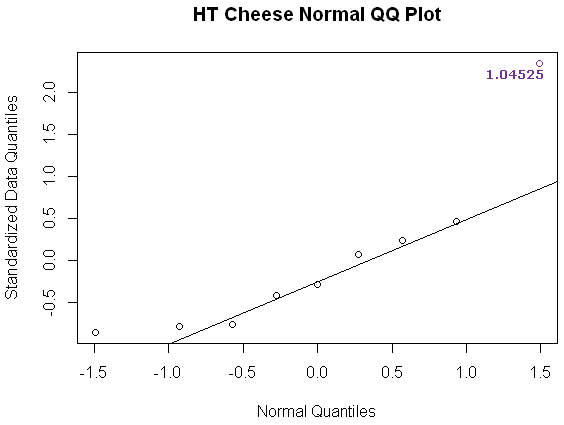


Figure 2. Normal QQ Plot for the HT Cheese dataset.

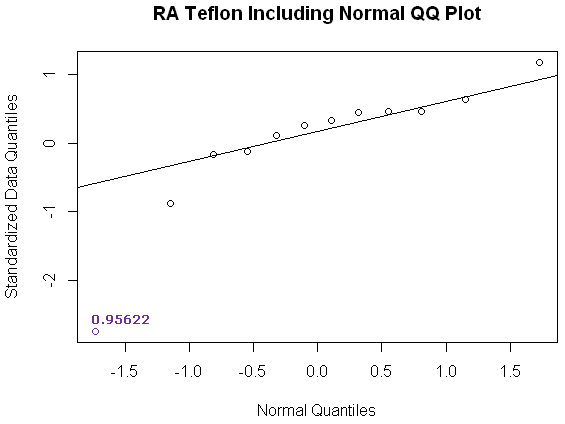


Figure 3. Normal QQ Plot for the RA Teflon Including dataset.

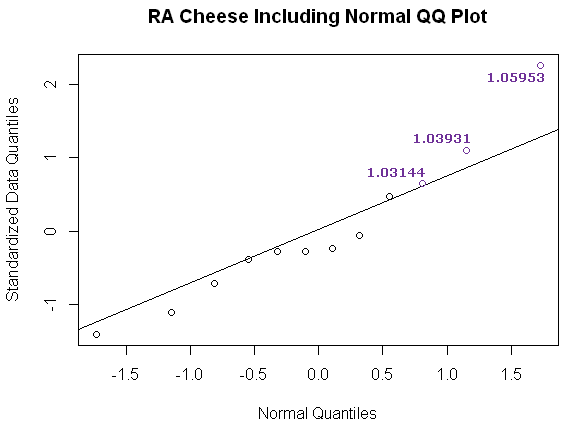


Figure 4. Normal QQ Plot for the RA Cheese Including dataset.

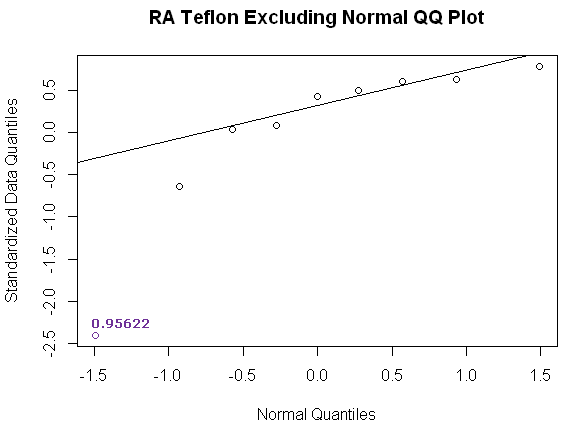


Figure 5. Normal QQ Plot for the RA Teflon Excluding dataset.

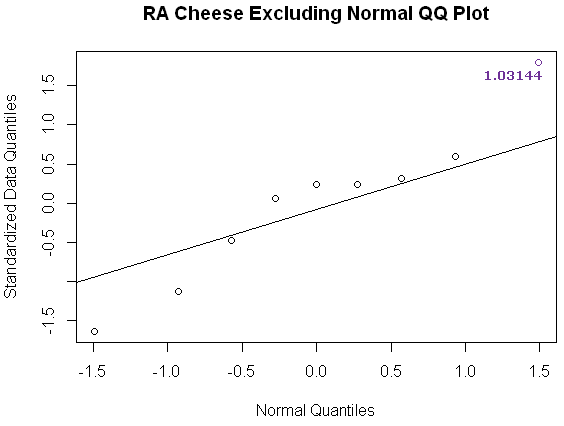


Figure 6. Normal QQ Plot for the RA Cheese Excluding dataset.

I continue with more formal normality testing. I use the Shapiro-Wilk[4,5] test. Table 3 below presents the results of the Shapiro-Wilk test for normality on our data sets assuming a significance level of α = 0.05. The table further indicates departure from normality. It is interesting to notice that we find that for HT, the Cheese phantom data is non-normal where as for RA, it is the Teflon phantom that is found to be non-normal.

|  |  |  |
| --- | --- | --- |
| Dataset | p-value | Normal? |
| HT Teflon | 0.12790 | Normal |
| HT Cheese | 0.02504 | Not Normal |
| RA Teflon including | 0.00564 | Not Normal |
| RA Cheese including | 0.47580 | Normal |
| RA Teflon excluding | 0.00470 | Not Normal |
| RA Cheese excluding | 0.69770 | Normal |

Table 3. Results of the Shapiro-Wilk test of normality on our data sets assuming a significance level of α = 0.05.

I test for equality of variances. I test for this using the Fligner-Killeen[6] test rather than the Bartlett[10,11] or F[ref] tests as the evidence for non-normality suggests departure from normality. Table 4 below summarizes the results of this testing for equality of variances. The table indicates that the paired data sets have statistically equivalent variances.

|  |  |  |
| --- | --- | --- |
| Dataset | p-value | Equality? |
| HT Teflon vs. Cheese | 0.3321 | Equal |
| RA including  Teflon vs. Cheese | 0.4305 | Equal |
| RA excluding  Teflon vs. Cheese | 0.8879 | Equal |

Table 4. Results of the Fligner-Killeen test for equality of variances on our data sets assuming a significance level of α = 0.05.

Based on the kurtosis, skewness, outliers, small sample size, Shapiro-Wilk test results suggest there may be some departure from normality in the data though the pairs all have similar variances. Due to the departure from normality, a paired t-test may not be appropriate; however, I include it for reference. A non-parametric alternative to a paired t-test is the Wilcoxon paired sign rank test[7]. However I must test if the distribution of the difference of the samples is symmetric. Table 5 below summarizes the results of this symmetry testing. It shows symmetry for the case of comparing the Teflon and Cheese phantoms on HT but for RA there is significant departure from symmetry. This indicates that for the RA case, the Wilcoxon paired sign rank test is not suggested, but may be used for the HT comparison. Thus I will be using the Wilcoxon paired sign test for HT and the paired sign test[8,9] for RA.

|  |  |
| --- | --- |
| Dataset | Skewness |
| HT Teflon vs. Cheese | -0.12451 |
| RA including  Teflon vs. Cheese | -1.22901 |
| RA excluding  Teflon vs. Cheese | -1.34749 |

Table 5. Results of the symmetry testing on the distribution of differences of corresponding Teflon and cheese samples.

Table 6 below summarizes the results of the testing between the corresponding Teflon and Cheese point dose comparisons by giving p-values for each of the performed tests with an assumed significance level of α = 0.05. I included the results of all three tests on all three comparisons. Each test set returned the same conclusion that for HT, we cannot reject the null hypothesis of the means being equivalent. But for RA, both including and excluding, that we can reject the null hypothesis that the means are equivalent. This suggests that for HT, the Teflon and Cheese phantoms were behaving similarly, but for RA, the two phantoms were not.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | t-test | signed rank test | sign test |
| HT Teflon vs. Cheese | 0.25730 | 0.39540 | 0.50780 |
| RA including  Teflon vs. Cheese | 0.00046 | 0.00037 | 0.00049 |
| RA excluding  Teflon vs. Cheese | 0.00867 | 0.00276 | 0.00391 |

Table 6. Results of the test between the corresponding Teflon and Cheese point dose comparisons by giving p-values for each of the performed tests assuming a significance level of α = 0.05.

failing plans analysis

Table 7 below presents the data for the ratio of measured to calculated doses for the failing plans.

|  |  |
| --- | --- |
| HT Teflon | HT Cheese |
| 0.90843 | 0.92368 |
| 0.91743 | 0.92966 |
| 0.92926 | 0.96096 |
| 0.92441 | 0.91733 |
| 0.93675 | 0.94345 |
| 0.92441 | 0.93178 |

Table 7. The ratios of measured to calculated point doses for each modality and phantom for the failing plans.

I look at the basic statistics of each dataset: number of samples, mean, standard deviation, variance, skewness, and kurtosis. Note: these are all sample statistics not population statistics. This is given in Table 8 below. The kurtosis suggests there may be some platykurtic departure from normality.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | # Samples | Mean | Std. | Variance | Skewness[2] | Kurtosis[2] |
| HT Teflon | 6 | 0.92345 | 0.00974 | 0.00010 | -0.19008 | -4.43264 |
| HT Cheese | 6 | 0.93448 | 0.01564 | 0.00024 | 0.56058 | -4.33113 |

Table 8. Statistical summary of the failing plan data sets.

I look at normal probability qq-plots for each of the data sets to get a graphical sense of the data as well as qualitatively checking it for normality. Figures 7-8 give the qq plots for each dataset.

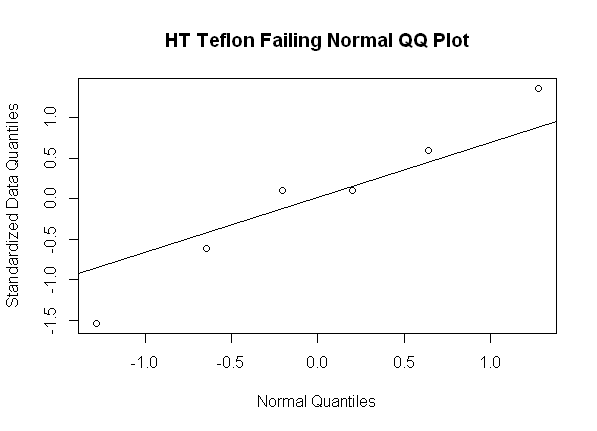


Figure 7. Normal QQ Plot for the HT Teflon Failing dataset.

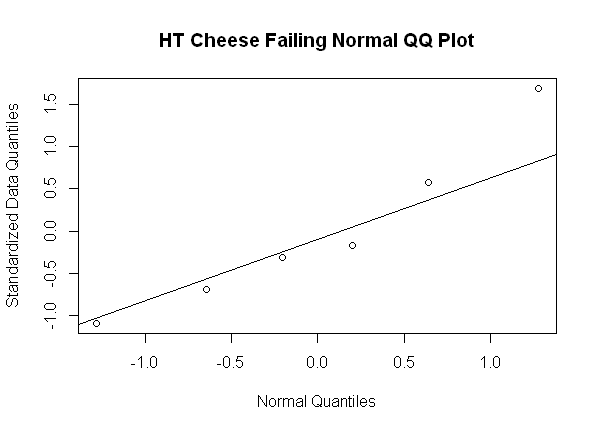


Figure 8. Normal QQ Plot for the HT Cheese Failing dataset.

I continue with more formal normality testing. I use the Shapiro-Wilk[4,5] test. Table 9 below presents the results of the Shapiro-Wilk test for normality on the failing data sets assuming a significance level of α = 0.05. The results indicates normality.

|  |  |  |
| --- | --- | --- |
| Dataset | p-value | Normal? |
| HT Teflon Failing | 0.9370 | Normal |
| HT Cheese Failing | 0.6082 | Normal |

Table 9. Results of the Shapiro-Wilk test of normality on our failing plan data sets assuming a significance level of α = 0.05.

Given the normality of the data, I continue on to test the equality of the variances. Given that the Shapiro-Wilk test suggests normality, I use the F[10] test for equality of variances as it is the most powerful test if the two samples are normally distributed. Table 10 below gives the result. The results suggest equality of variances.

|  |  |  |
| --- | --- | --- |
| Dataset | p-value | Equality? |
| HT Failing Teflon vs. Cheese | 0.3214 | Equal |

Table 10. Results of the F test of equality of variances on our failing plan data sets assuming a significance level of α = 0.05.

I compare the Teflon and Cheese phantom failing plan data sets using the paired t-test as it is the most powerful test when there is evidence supporting the normality of the data sets to be compared as well as the equality of their variances. Table 11 below gives the results. The results suggest the Teflon and Cheese phantoms had equal means suggesting that they behaved similarly.

|  |  |  |
| --- | --- | --- |
| Dataset | p-value | Equality? |
| HT Failing Teflon vs. Cheese | 0.08675 | Equal |

Table 11. Results of the t-test between the failing plan data sets assuming a significance level of α = 0.05.

References

1. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
2. Komsta, Lukasz and Novomestky, Frederick (2011). moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.12. <http://CRAN.R-project.org/package=moments>.
3. Gross, Juergen (2011). nortest: Tests for Normality. R package version 1.0. <http://CRAN.R-project.org/package=nortest>.
4. Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)", Biometrika, 52, 3 and 4, pages 591-611.
5. Razali, N.M. and Wah, Y.B (2011). “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests”, Journal of Statistical Modeling and Analytics, Vol. 2 No. 1 pages 21-33. <---- Possibly find a source with better English grammar
6. Fligner, Michael A., and Killeen, Timothy J. (1976). "Distribution-Free Two-Sample Tests for Scale", Journal of the American Statistical Association, 71, 210-213.
7. Wilcoxon, Frank (1945). “Individual comparison by ranking methods”, Biometrics 1, 80–83.
8. Gibbons, J.D. and Chakraborti, S. (1992). Nonparametric Statistical Inference. Marcel Dekker Inc., New York. <--- Not original work on sign test. Need to try to find it.
9. Arnholt, Alan T. (2010). BSDA: Basic Statistics and Data Analysis. R package version 1.0. <http://CRAN.R-project.org/package=BSDA>.
10. Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press.
11. sdf