Trans IMF 97(1) (2019). Post-print version. Accepted 30/11/18. Doi: 10.1080/00202967.2019.1555367

Statistical approaches in surface finishing. Part 2. Non-parametric methods for data analysis

J. R. Smith*[1] and C. Larson[2]

[1] *School of Pharmacy and Biomedical Sciences, University of Portsmouth, Portsmouth PO1 2DT, UK*
[2] *Glencree, Leics LE14 4PH, UK*
**For correspondence:* james.smith@port.ac.uk. J. R. Smith http://orcid.org/00000001-8805-3788

Abstract
This paper is the second of a short series of articles aimed towards describing some of the various statistical methods and approaches that have been used in surface finishing. The methods fall broadly into two areas: analysis and design-of-experiments. In the first paper, the subject was briefly introduced followed by a discussion of parametric statistics hypothesis testing. This second paper logically continues with descriptions of the various non-parametric tests and situations where these could be applied, all within the context of surface finishing applications.

Keywords: Statistics; Non-parametric; Mann-Whitney U test; Wilcoxon signed rank test; Kruskal-Wallis test; Dunn's test; Friedman test; Chi Square test; ranked data

Introduction

There has been a marked increase in the application of sophisticated statistical tools in both the design of materials finishing experiments and in the analysis of data in recent years.[1-4] Examples from the first category usually come under the design-of-experiments (DOE) type, which were briefly summarised in Part 1 of this series;[5] fundamental and methodological details describing the use of DOE will follow later in this series. Part 1 also contained a 'beginner's guide' in how to perform key parametric statistics tests using the stats software package SPSS.[5] Parametric tests assume data are independent, normally distributed and have homogeneity of variance.

This present paper will focus on how to perform hypothesis testing when the conditions for applying a parametric test are not met. For example, the data might not be normally distributed, perhaps due to a small sample size or the frequency distribution histogram might be skewed, or the data might be in an ordinal form (non-continuous, such as score data). There are a number of so-called non-parametric tests, the most important of which will be described here. The emphasis in this paper will be how to perform the tests manually, although the reader is referred to supplemental materials which provide instructions on how to perform each test using SPSS (doi: 10.1080/XXXXXXXX.XXXX.XXXXXXX; see Appendix).

The non-parametric tests discussed in this paper include: the Mann-Whitney U test, the Wilcoxon signed rank test, the Kruskal-Wallis test (and Dunn's test), the Friedman test and the Chi Square test (Table 1). Many of these have parametric equivalents, some of which were discussed in Part 1.[5]

Table 1. List of non-parametric statistical tests with parametric test equivalents and their requirements and assumptions.

Non-parametric test	Parametric test equivalent	Data type	Number of groups	Notes
Mann-Whitney U	Student's t-test	Interval or Ordinal	2	Compares difference in distributions or medians rather than means (Student's t-test). Often used as a *post hoc* to the Kruskal-Wallis test
Wilcoxon signed rank	Paired Student's t-test	Interval	2	
Kruskal-Wallis	One-way ANOVA	Interval or Ordinal	> 2	Mann-Whitney U applied as a *post hoc* test
Dunn's test	Student's t-test	Interval or Ordinal	≥ 2	*Post hoc* test (alternative to Mann-Whitney U) to the Kruskal-Wallis and *post hoc* test to the Friedman test
Friedman	Repeated measures One-way ANOVA	Interval or Ordinal	> 2	Dunn's test applied as a *post hoc* test
Chi Square	None	Nominal	≥ 2	Compares observed and expected values

Before going into detail on how to perform each of the chosen non-parametric tests, some selected examples from the literature where such tests have been applied to analyse real data from surface finishing/coatings experiments are considered (Table 2). The reader is reminded that the dependent variable (dv) is the measured property where the effect caused by the intervention is expected, for example, surface roughness, thickness, leach rate, hardness, % elongation at break, reflectance etc; this property is usually plotted on the y-axis of a graphical display. The independent variable (iv) is normally the group type, usually found on the x-axis, examples being coating type, alloy composition, cell type, exposure environment etc; this property can also be continuous data, *e.g.*, time. The type of non-parametric test to be employed depends on the type of data (interval, ordinal or nominal),[5] and the number of groups (Table 1).

The various tests are described and, for some, example analyses are given to assist the reader in understanding fully the application.

Figure 1 illustrates the conditions and choices associated with the different tests.

Mann-Whitney U test

The Mann-Whitney U test is the non-parametric equivalent of the Student's t-test,[5] which cannot be applied due to the data not being normally distributed.

Table 2. Examples of non-parametric statistical tests applied in surface finishing/coatings research publications.

Non-parametric test	Publication title	Dependent variable	Independent variable	Ref
Mann-Whitney U	Silver deposition on titanium surface by electrochemical anodizing process reduces bacterial adhesion of *Streptococcus sanguinis* and *Lactobacillus salivarius*	Adhesion of *Streptococcus sanguinis*	Coating and Ti substrate	-[6]
Mann-Whitney U	Shear bond strength between an indirect composite layering material and feldspathic porcelain-coated zirconia ceramics	Shear bond strengths	With and without opaque coating	-[7]
Wilcoxon rank sum	Evaluation of two lead-based paint removal and waste stabilization technology combinations on typical exterior surfaces	Average personal breathing space zone, conc level and separately, area/perimeter Pb-containing particulate levels	2 stabilisation technologies, 2 substrates (wood & brick); before application (n=15 for each substrate per technology) and after (n=75 for each substrate per technology)	-[8]
Wilcoxon rank sum	Photocatalytic and antimicrobial effects of interior paints	%Growth of bacteria (and separately fungi) on surface	4 paints containing various nano ZnO concentrations	-[9*]
Kruskal-Wallis	Osteogenic potential of human adipose-derived stromal cells on 3-dimensional mesoporous TiO2 coating with magnesium impregnation	Surface roughness	3 TiO_2 surfaces (nonporous, mesoporous & Mg-containing mesoporous)	-[10]
Kruskal-Wallis	Paclitaxel coating on the terminal portion of hemodialysis grafts effectively suppresses neointimal hyperplasia in a porcine model	Neointimal hyperplasia: graft area (H/G ratio) & %luminal stenosis	3 Paclitaxel coating groups	-[11]
Kruskal-Wallis	Development of Ti–C–N coatings with improved tribological behavior and antibacterial properties	%Surface covered by bacteria	2 bacteria strains, 3 Ti alloys	-[12]
Friedman	Gallium-modified chitosan/ poly(acrylic acid) bilayer coatings for improved titanium implant performances	Bacteria viability	Mg substrate, 2 coatings, both before & after incubation	-[13]
Friedman	Effect of surface protection, staining beverages and aging on the color stability and hardness of recently introduced uncoated glass ionomer restorative material	Colour value change	glass ionomer substrate & 2 coatings on same substrate, 3 different storage times	_[14]
Chi square	Estimating the potential of resource conservation in metal electroplating based on the stability index for the composition of solutions	Theoretically possible number of loadings of an electroplating bath	Concentration of the principal metal in the composition of the solution	-[15]

[*] Results of the antimicrobial and antifungal behaviour of interior coatings with nano ZnO were confirmed repeatedly, hence the Wilcoxon rank-sum test, which assesses whether two independent samples of observations come from the same distribution.[9]

Figure 1. Choices and requirements for hypothesis testing.

In materials finishing, for example, the Mann-Whitney U test could be applied to see whether the number of defects (*nd*; ordinal data, in this case) observed in micrographs (SEM/AFM images) of coatings exposed to two different environments (A and B) were statistically different (the alternative hypothesis, H_1;[5] the null hypothesis, H_0, being the number of defects after exposure to the two environments would be the same). An hypothetical example dataset (columns 1 and 2) is shown in Table 3 (note: the number of measurements per environment, *n*, does not have to be identical).

Table 3. Hypothetical example dataset for which the Mann-Whitney U test can be applied: The number of defects (*nd*; ordinal data, in this case) observed in micrographs (SEM/AFM images) of coatings exposed to two different environments (A and B).

nd_A (n_A=15)	nd_B (n_B=12)	Rank position	Ordered nd_A and nd_B	Rank position adjusted for ties	Rank position of nd_A	Rank position of nd_B
2	3	1	2	(1+2)/2=1.5	1.5	3
2	4	2	2	(1+2)/2=1.5	1.5	4
5	5	3	3	3	5.5	5.5
6	6	4	4	4	8	8
7	6	5	5	5.5	10	8
8	8	6	5	5.5	13	13
8	8	7	6	(7+8+9)/3=8	13	13
8	10	8	6	8	13	17.5
9	12	9	6	8	16	19.5
10	12	10	7	10	17.5	19.5
13	14	11	8	13	21	22.5
14	19	12	8	13	22.5	27
15	-	13	8	13	24	-
16	-	14	8	13	25.5	-
16	-	15	8	13	25.5	-
-	-	16	9	16	-	-
-	-	17	10	17.5	-	-
-	-	18	10	17.5	-	-
-	-	19	12	19.5	-	-
-	-	20	12	19.5	-	-
-	-	21	13	21	-	-
-	-	22	14	22.5	-	-
-	-	23	14	22.5	-	-
-	-	24	15	24	-	-
-	-	25	16	25.5	-	-
-	-	26	16	25.5	-	-
-	-	27	19	27	-	-
				Σ	217.5	160.5

It is convenient to first produce a column of rank positions increasing from 1 to *N*, in steps of 1, where *N* is the total number of measurements ($N = n_A + n_B = 27$; column 3, Table 3). Then, assign to each of these rank positions increasing nd_A and nd_B scores (from both of the groups; column 4). Where there is a duplication in the ordered nd_A and/or nd_B score (a 'tie', in column 4), the mean of the corresponding rank values should be recorded (column 5). The rank values are then placed in their original A and B groups (columns 6 and 7). The *U* statistic is then calculated for each group:

$$U_A = \sum A - \left[\frac{n_A(n_A+1)}{2}\right] = 217.5 - \left(\frac{15 \times (15+1)}{2}\right) = 217.5 - 120 = 97.5 \qquad \text{(Eq. 1)}$$

$$U_B = \sum B - \left[\frac{n_B(n_B+1)}{2}\right] = 160.5 - \left(\frac{12 \times (12+1)}{2}\right) = 160.5 - 78 = 82.5 \qquad \text{(Eq. 2)}$$

The smallest value of U_A or U_B is carried forward as the U_{Stat} statistic, therefore $U_{Stat} = 82.5$, which is then compared to critical values of the Mann-Whitney U statistic.[16] U_{crit} (at $\alpha = 0.05$, two-tailed test) = 49 (for $n_A = 15$, $n_B = 12$). Since $U_{Stat} > U_{crit}$ the null hypothesis is not rejected, therefore there is no statistical difference between the two groups.

Wilcoxon signed rank test

The Wilcoxon signed rank test is the non-parametric equivalent of the paired Student's t-test, where the same subject, sample or material is exposed to two conditions or treatments. For example, a coating could be leaching an antifouling agent and a difference between initial leach rate (r_o / µg cm^{-2} day^{-1}) and leaching after 6 months (r_1) is required for comparison; the before and after values will provide the pairings for a number of separate samples (say $N = 12$) (Table 4). The r_0 is noted to be skewed (not normally distributed; see Part 1 for normality test)[5] and so the paired Student's t-test cannot be used.

The differences and absolute differences (neglecting sign) between r_0 and r_1 are first calculated (columns 4 and 5, respectively, in Table 4). The absolute differences are then given a rank value (lowest to highest) as in the Mann-Whitney U test, and also noting a mean rank when two or more absolute differences are the same (column 6). It is convenient to make a 7th and 8th column (Table 4) noting the rank corresponding to the sign of the original ($r_0 - r_1$) subtraction; the sum of these ranks (Σrank+ and Σrank-) are then recorded. The smallest sum is referred to as the W_{Stat} value, which in this case = 4.5 (Σrank- < Σrank+). As in the Mann-Whitney U test, the critical value (W_{crit}) is looked up in tables[17] (using $\alpha = 0.05$ and the degrees of freedom, $df = N - 1 = 11$; two-tailed test), which is 10. Note, $df = N - 1$ in this test, and not the before and after cases (2) – 1 = 1. Also, for this test, W_{Stat} must be < W_{crit} to reject the null hypothesis (H_0, that there is no statistical difference between before and after measurements). Here, $W_{Stat} < W_{crit}$, therefore H_0 can be rejected: there is a statistical difference between before and after measurements (H_1).

Table 4. Hypothetical example dataset for which the Wilcoxon signed rank test can be applied: Initial leach rate (r_o / µg cm^{-2} day^{-1}, 'before') and leaching after 6 months (r_1, 'after') of an antifouling agent from a replicate coating samples (N = 12).

Coating	r_0	r_1	$r_0 - r_1$	ABS ($r_0 - r_1$)	Rank	Rank+	Rank-
1	35.6	24.0	11.6	11.6	12	12	-
2	29.5	19.6	9.9	9.9	9	9	-
3	31.2	21.0	10.2	10.2	11	11	-
4	27.3	17.3	10	10	10	10	-
5	21.5	18.9	2.6	2.6	7	7	-
6	22.3	21.9	0.4	0.4	2	2	-
7	15.8	15.9	-0.1	0.1	1	-	1
8	19.5	12.5	7	7	8	8	-
9	12.0	12.8	-0.8	0.8	3.5	-	3.5
10	8.7	7.0	1.7	1.7	5	5	-
11	17.3	14.8	2.5	2.5	6	6	-
12	16.8	16.0	0.8	0.8	3.5	3.5	-
					Σ	73.5	4.5

Kruskal-Wallis test

The Kruskal-Wallis test (also known as the Kruskal-Wallis H test) is the non-parametric equivalent of the Analysis-Of-Variance (ANOVA) test, described in Part 1,[5] and is used when normality has been violated or if the data is ordinal. As with the ANOVA, it is usually applied when there are three or more groups of data (although can be used for two groups). *Post hoc* (Latin. 'after this') testing is usually carried out using the previously described Mann-Whitney U test. For example, the Kruskal-Wallis test would state that there is a statistical difference ($H_{stat} > H_{crit}$) between at least one pair of groups within the comparison (perhaps five groups), but the Mann-Whitney U test used in *post hoc* will pin-point the pairs of groups causing the difference(s). An example in the materials finishing/coatings sector might be the comparison of skewed surface roughness (arithmetic roughness average)[18] data (RA_A, RA_B and RA_C; Table 5) between, say, 3 types of abrasive treatment (A-C), the first two groups having 6 samples, with, say, 5 in group 2). As with the Mann-Whitney U test, the first task is to assign a rank value (1 to 17) to every measurement regardless of which group the data item came from (column 4). If there are identical *RA* values, a mean rank should also be calculated (as previously described). Next, the ranked data should be placed in the same groups from where the unranked data originated. The sum of the ranks for each group is then used to calculate the *H* statistic, H_{Stat}:

$$H_{Stat} = \frac{12}{N(N+1)} \sum \frac{r_i^2}{n_i} - 3(N+1) = \frac{12}{17(17+1)} \times \left(\frac{51^2}{6} + \frac{19^2}{5} + \frac{83^2}{6}\right) - 3(17+1) = 10.86$$

(Eq. 3)

As with the previously mentioned tests, the critical value (H_{crit}) is looked up in tables, this time using a table of Chi Square (χ^2) values,[19] (using $\alpha = 0.05$ and df = number of groups – 1 = 2; two-tailed test), which is 5.991. For the Kruskal-Wallis test, H_{Stat} must be > H_{crit} (χ^2_{crit}) to reject the null hypothesis (H_0, that there is no statistical difference between surface roughness values). Here, $H_{Stat} > H_{crit}$, therefore H_0 can be rejected: there is a statistical difference between surface roughness values (H_1). As with the ANOVA, this difference means that there is a difference between at least two of the groups (A and B, B and C, or A and C), but the Kruskal-Wallis test will not make the distinction. *Post hoc* analysis, using the already described Mann-Whitney U test, will point to where these significant differences lie. Alternatively, the Dunn's multiple comparison test can be used for *post hoc* testing instead of the Mann-Whitney U test.

Table 5. Hypothetical example dataset for which the Kruskall-Wallis test can be applied: Skewed surface roughness (arithmetic roughness average)[18] data (RA_A, RA_B and RA_C) after treatment with 3 types of abrasive (A-C); r_A, r_B and r_C = rank roughness values.

RA_A	RA_B	RA_C	Rank position	Ordered RA_A, RA_B and RA_C*	r_A	r_B	r_C
59.6	30.3	115.1	1	30.3	6	1	17
63.4	58.7	98.6	2	42.6	7	5	15
45.6	42.9	78.5	3	42.9	4	3	10
90.5	42.6	100.3	4	45.6	13	2	16
80.8	70.3	79.9	5	58.7	12	8	11
78.1	-	95.6	6	59.6	9	-	14
-	-	-	7	63.4			
-	-	-	8	70.3			
-	-	-	9	78.1			
-	-	-	10	78.5			
-	-	-	11	79.9			
			12	80.8			
			13	90.5			
			14	95.6			
			15	98.6			
			16	100.3			
-	-	-	17	115.1			
				Σ	51	19	83

*No ties.

Dunn's multiple comparison test

The Dunn's test is often used as for *post hoc* analysis after the Kruskal-Wallis test. As with *post hoc* testing described after performing an ANOVA (a parametric test),[5] the Dunn's test compares three or more means to pinpoint which specific means are significant from the others. The Dunn's test divides the overall α = 0.05 level by the number of comparisons to produce a modified α value; for example, if there are 10 comparisons, a modified α of 0.005 should will result. This is considered more robust and will lessen the chance of type 1 errors (rejecting H_0 when in reality H_0 was correct – a 'false positive') and type 2 errors (retaining H_0 when in reality H_0 was false – a 'false negative') from being made.

Dunn's method has been used in a statistical analysis of the comparison between means of the cut amount for newly developed electrodeposited diamond scaler blades with very high abrasion resistance.[20] Four different diamond particle sizes were assessed.

Friedman test

The Friedman test is a non-parametric repeated-measures ANOVA test. Similar to the parametric repeated measures ANOVA, it is used to detect differences in treatments across multiple test attempts when the dependent variable being measured is ordinal. The procedure involves ranking each row together, then considering the values of ranks by columns. The test, along with the non-parametric *post hoc* Mann-Whitney test, has been used to calculate statistical significance in improving titanium implant performance when applying gallium-modified chitosan/poly (acrylic acid) bilayer coatings to a titanium surface.[13]

In a study of the effect of protective surface coatings, staining beverages and aging on the colour stability and hardness of recently introduced uncoated glass ionomer dental restorative material, Friedman's test was used to give a comparison between the colour changes (ΔE) values at different storage time periods. Dunn's test was used for pair-wise comparisons.[14]

A relevant, hypothetical example might be the comparison of five experts (n = 5) assessing the perceived brightness of three different coating panels (k = 3), with perceived brightness scores (1 = most dim, 5 = most bright) being assessed; the question would then be whether any of the coating panels produce consistently better or worse perceived brightness. Hypothetical data is shown in Table 6.

Firstly, a rank (r_j) is assigned to each expert's brightness score for each of the coating panels used (rank within each row); if there are identical brightness scores for panels used for a given expert, then a mean rank is given (Table 6). The sum of the squares of the brightness rank scores (ΣR_j^2) for each coating panel is then calculated and this is used to calculate the Q statistic, Q_{Stat}:

$$Q_{Stat} = \frac{12}{nk(k+1)}\sum_{j=1}^{k} R_j^2 - 3n(k+1) = \frac{12}{5\times 3(3+1)} \times (5.5^2 + 11^2 + 13.5^2) - (3\times 5\times(3+1)) = 6.7 \qquad \text{(Eq. 4)}$$

The critical value (Q_{crit}) is then looked up in tables, however, the tables to be consulted depend on whether n or k is large (*i.e.*, n > 15 or k > 4); if large, the probability distribution can be approximated to a Chi square (χ^2) distribution and Q_{crit} retrieved from tables of χ^2,[19] if small, then Q tables specifically prepared for the Friedman test[21] should be consulted to find Q_{crit}. In the current example, both n < 15 and k < 4 and so the latter tables should be consulted. For an α = 0.05 and for n = 5 and k = 3, Q_{crit} = 6.400.[21] For the Friedman test, Q_{Stat} must be > Q_{crit} to reject the null hypothesis (H_0, that there is no statistical difference between the perceived brightness values of each coating panel examined). Here, $Q_{Stat} > Q_{crit}$, therefore H_0 can be rejected: there is a statistical difference between the perceived brightness of coating panels examined (H_1). The Dunn's multiple comparison test then could be used to perform a *post hoc* analysis.

Table 6. Hypothetical example dataset for which the Friedman test can be applied: Perceived brightness scores (1 = most dim, 5 = most bright) given by five experts for three different coating panel samples.

	Perceived brightness score for each coating panel sample *k*			Rank position of perceived brightness score for each coating panel sample*		
Expert, *n*	*k* = 1	*k* = 2	*k* = 3	*k* = 1	*k* = 2	*k* = 3
1	4	5	5	1	2.5	2.5
2	3	4	5	1	2	3
3	2	4	4	1	2.5	2.5
4	4	5	5	1	2.5	2.5
5	3	3	4	1.5	1.5	3
			Σ	5.5	11	13.5
			mean	1.1	2.2	2.7

*Ties are allocated the mean of ranks that would have been assigned without ties.

Chi Square test

The Chi Square (χ^2, or Pearson's Chi Square or Chi Squared) test is probably one of the least used non-parametric tests in the physical sciences although is used ubiquitously in the social sciences. It is used for comparing differences between observed and expected nominal (categorised) data. For example, if 100 coins are tossed, and 73 coins landed "heads up", is this significant? Is it a significant deviation from the 50:50 expected? Are 73 heads more than would be expected due to chance? For this test, data should be in the form of frequency scores (not percentages, for example) and needs to be arranged in categories, with more than 20 total data items and at least four in each category. The test has been used in an analysis of the electroplating industry to assist in estimating the potential of conserving resources, *i.e.*, the metals, in the sector based on the relationship between the theoretically possible number of loadings of a bath and the concentration of the principal metal in the composition of the solution.[15] Another relevant example in the field of materials finishing/coatings, which has been explored in recent years in this journal,[22] and one of perhaps more insight to readers of this journal, might be whether the presence of a coating on a nickel substrate (*e.g.*, a jewellery item) had an effect on whether people reported an allergic response[22] (when questioned after a set period of time; example data are given in Table 7). Here, note the categories: coating-allergy (CA), coating-no allergy (CN), substrate-allergy (SA) and substrate-no allergy (SN); any continuous measure of degree of allergy (*i.e.*, non-binary) or coating thickness variations (currently in nominal form) would preclude the use of Chi Square test.

The null hypothesis (H_0) would be: The presence of a coating on nickel *does not* cause a reduction in reported allergic response; the alternative hypothesis (H_1) would be: The presence of a coating on nickel *does* cause a reduction in reported allergic response.

Assuming H_0 to be true, the proportion of people wearing a *coated* Ni jewellery item who reported an allergic response should be the same as the number of people wearing an *uncoated* Ni jewellery item who reported an allergic response.

Table 7. Hypothetical example dataset for which the Chi Square test can be applied: number of people reporting allergic reaction (A or N) to a Ni jewellery item (S) or the same surface which had been coated [with a layer to prevent allergic reaction] (C) questioned after a set period of time; *O* and *E* refer to observed and expected numbers of people.

Effect	Ni Substrate (S)	Coating on Ni (C)	Row total	Category combination	*O*	*E*	(*O-E*)	Yates' correction* (*O-E*) -0.5	$(O-E)^2/E$
Allergic reaction (A)	45	16	61	SA	45	25.5	19.5	19.0	14.16
No allergic reaction (N)	24	35	59	SN	24	34.5	-10.5	-11.0	3.51
-	-	-	-	CA	16	25.5	-9.5	-10.0	3.92
-	-	-	-	CN	35	34.5	0.5	0	0
Σ	69	51	120			120			21.59

*The Yates' correction subtracts 0.5 from the (*O-E*) values in a 2 × 2 table to reduce an error associated when using the χ^2 distribution to interpret χ_{Stat}^2 since an assumption is made that the discrete probability of observed binomial frequencies in the table can be approximated by the continuous the χ^2 distribution.[23]

The total proportion of people who reported an allergic response = 51/120 = 0.425 (Eq. 5)

The number of people expected (*E*) in each of the category combinations (CA, CN, SA and SN) can now be calculated:

CA = SA = 0.425 × (120/2) = 25.5 (Eq. 6)
CN = SN = (1-0.425) × (120/2) = 34.5 (Eq. 7)

These *E* values can then used with observed (*O*, original data) values (columns 5–10, Table 7) to calculate χ^2_{Stat}: (a Yates' correction is also applied for 2 × 2 categories, see Table 7).[23]

$$\chi^2_{Stat} = \sum \frac{(O-E)^2}{E} = 21.59 \qquad \text{(Eq. 8)}$$

As with previous tests, the critical value (χ_{crit}) is looked up in χ^2 tables,[19] (using $\alpha = 0.05$ and df = (number of rows – 1) × (number of columns – 1) = 1 × 1 = 1; two-tailed test), which is 3.841. For the Chi Square test, χ_{Stat} must be > χ_{crit} to not accept the null hypothesis (H_0, the presence of a coating on nickel does not cause a reduction in reported allergic response). Since $\chi_{\text{Stat}} > \chi_{\text{crit}}$, H_0 can be rejected: The presence of a coating on nickel does cause a reduction in cases of reported allergic response (H_1).

Summary

This paper is the second in a short series of articles that report the use of statistical methods in surface finishing. The focus in this paper has been on non-parametric hypothesis testing. Too often, parametric tests such as the Student's t-test or ANOVA are applied to data which are not normally distributed about the mean or exhibit heterogeneity of variance, and can result in type 1 and type 2 statistical errors (false positives and negatives, respectively). A variety of non-parametric methods (the Mann-Whitney U test, the Wilcoxon signed rank test, the Kruskal-Wallis test (and Dunn's test), the Friedman test and the Chi Square test) have been described in this paper, in the context of surface finishing using real and hypothetical examples.

References

1. M. I. Coskun, I. H. Karahan and Y. Yücel: *Electrochim. Acta*, 2014, **150**, 46–54.
2. Y. Yaghoubinezhad and A. Afshar: *J. Solid State Electrochem.*, 2015, **19**, (5), 1367–1380.
3. S. Pathak, N. K. Jain and I. A. Palani: *Trans. IMF*, 2016, **94**, (2), 64–69.
4. F. Pagnanelli, P. Altimari, M. Bellagamba, G. Granata, E. Moscardini, P. G. Schiavi and L. Toro: *Electrochim. Acta*, 2015, **155**, 228–235.
5. J. R. Smith and C. Larson: *Trans. IMF*, 2016, **94**, (6), 288–293.
6. Godoy-Gallardo M., Rodriguez-Hernandez A. G., Delgado L. M., Manero J. M., F. J. Gil and D. Rodrıguez: *Clin. Oral Implants Res.*, 2015, **26**, 1170–1179.
7. R. Fushiki, F. Komine, M. B. Blatz, M. Koizuka, K. Taguchi and H. Matsumura: *Clin. Oral Invest.*, 2012, **16**, 1401–1411.
8. A. E. Daniels, J. R. Kominsky and P. J. Clark: *J. Hazardous Mater.*, 2001, **B87**, 117–126.
9. L. Hochmannova and J. Vytrasova: *Progr. Org. Coatings*, 2010, **67**, 1–5.
10. F. Cecchinato, J. Karlsson, L. Ferroni, C. Gardin, S. Galli, A. Wennerberg, B. Zavan, M. Andersson and R. Jimbo: *Mat. Sci. Eng. C*, 2015, **52**, 225–234.
11. I. Baek, J. Hwang, J. Park, H. Kim, J.-S. Park and D. J. Kim: *J. Vascular Surg.*, 2015, **61**, (6), 1575–1582.
12. V. S. de Viteri, G. Barandika, R. Bayon, X. Fernandeza, I. Ciarsolo, A. Igartua, R. P. Tanoira, J. E. Moreno and C. P.-J. Peremarch: *J. Mech. Behav. Biomed. Mater.*, 2016, **55**, 75–86.
13. M. A. Bonifacio, S. Cometa, M. Dicarlo, F. Baruzzi, S. de Candia, A. Gloria, M. M. Giangregorio, M. Mattioli-Belmonte and E. De Giglio: *Carbohydrate Polymers* 2017, **166**, 348–357.
14. B. D. M. A. Hamid, G. M. Mahmoud, F. M. El-Sharkawy and E. A. A. Auf: Future Dental J., in press. doi: 10.1016/j.fdj.2018.05.004
15. S. V. Fadina, E. G. Vinokurov, T. F. Burukhina and V. A. Kolesnikov: *Theoret. Foundations Chem. Eng.*, 2014, **48**, (6), 842–847.
16. Real Statistics Using Excel. Mann-Whitney Table. http://www.real-statistics.com/statistics-tables/mann-whitney-table/. Accessed 29/11/2018.
17. Wilcoxon Signed-Ranks Table. http://www.real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/. Accessed 29/11/2018.
18. J. R. Smith, S. Breakspear and S. A. Campbell: *Trans. IMF*, 2003, **81**, (3), B55–B58.
19. Values of the Chi-squared distribution. https://www.medcalc.org/manual/chi-square-table.php. Accessed 29/11/2018.
20. T. Fujimura, H. Kaya and S. Kimura, A newly-developed electrodeposited diamond sealer *w*ith high abrasive resistance, *J. Periodontol.*, 1995, **66**, (10), 878–886.
21. Upper critical values for the Friedman Test. https://www.york.ac.uk/depts/maths/tables/friedman.pdf. Accessed 29/11/2018.
22. D. R. Gabe and C. Larson: *Trans. IMF*, 2017, **95**, (2), 71–72.
23. F. Yates: Contingency table involving small numbers and the χ^2 test, *Suppl. to J. Royal Stat. Soc.*, 1934, **1**, (2), 217–235. https://www.jstor.org/stable/2983604?seq=1#metadata_info_tab_contents. Accessed 29/11/2018.

Trans IMF 97(1) (2019). Post-print version. Accepted 30/11/18. Doi: 10.1080/00202967.2019.1555367

Supplementary material

Reproduced below from doi: 10.1080/XXXXXXXX.XXXX.XXXXXXX

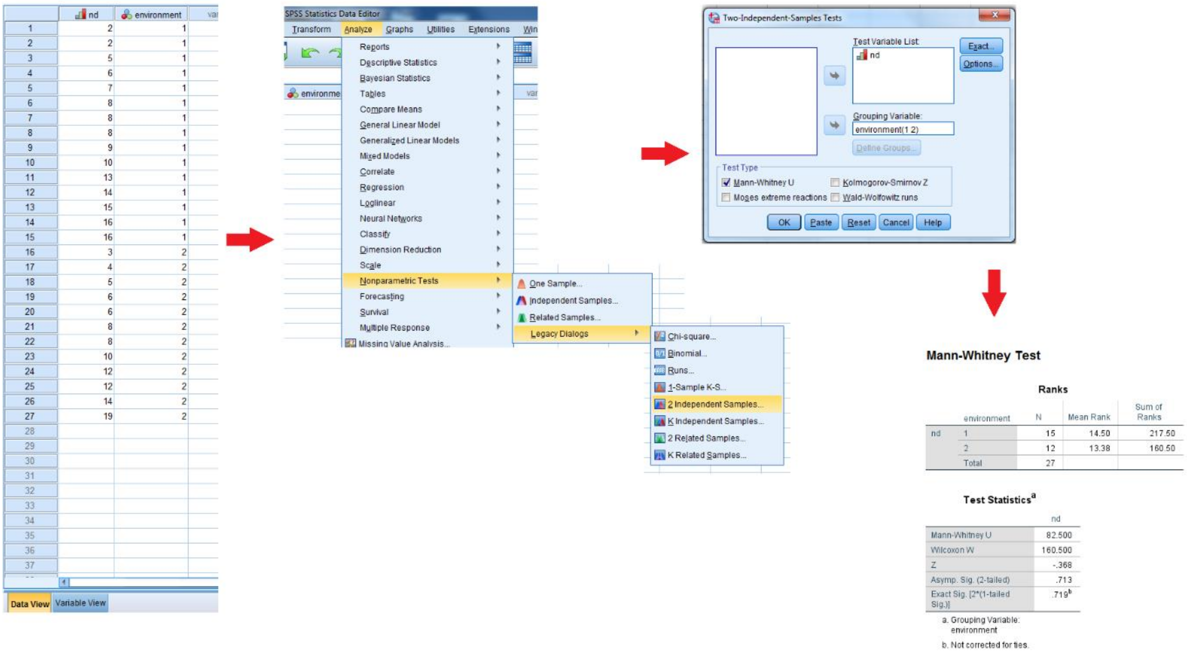

Figure S1. Mann-Whitney U SPSS Example.

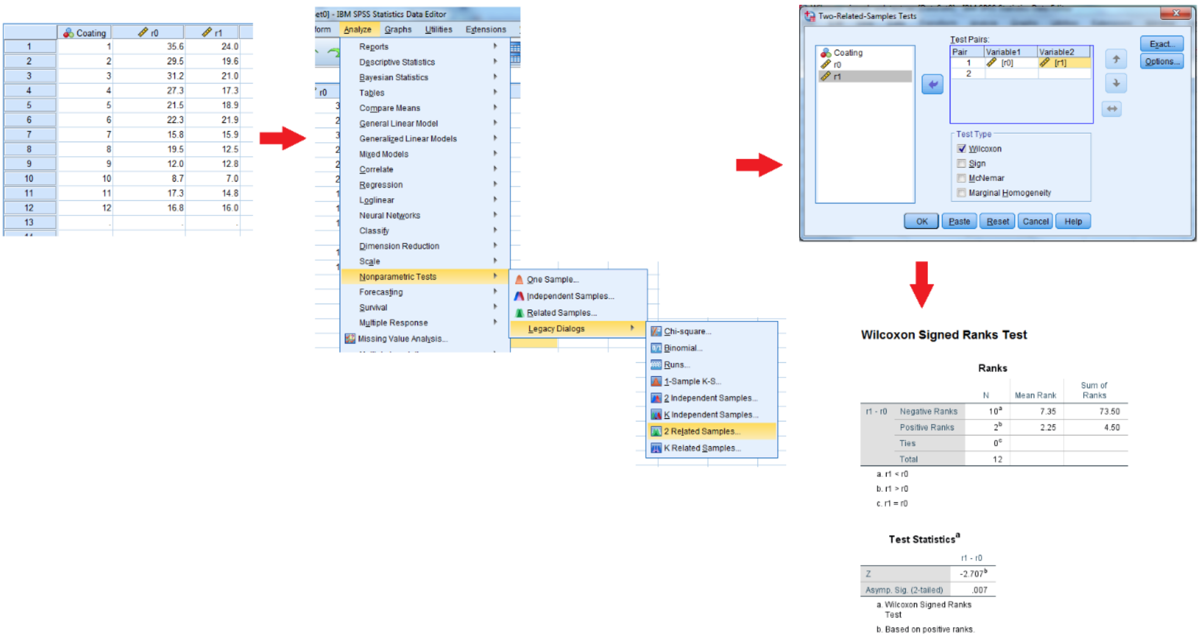

Figure S2. Wilcoxon signed rank SPSS Example.

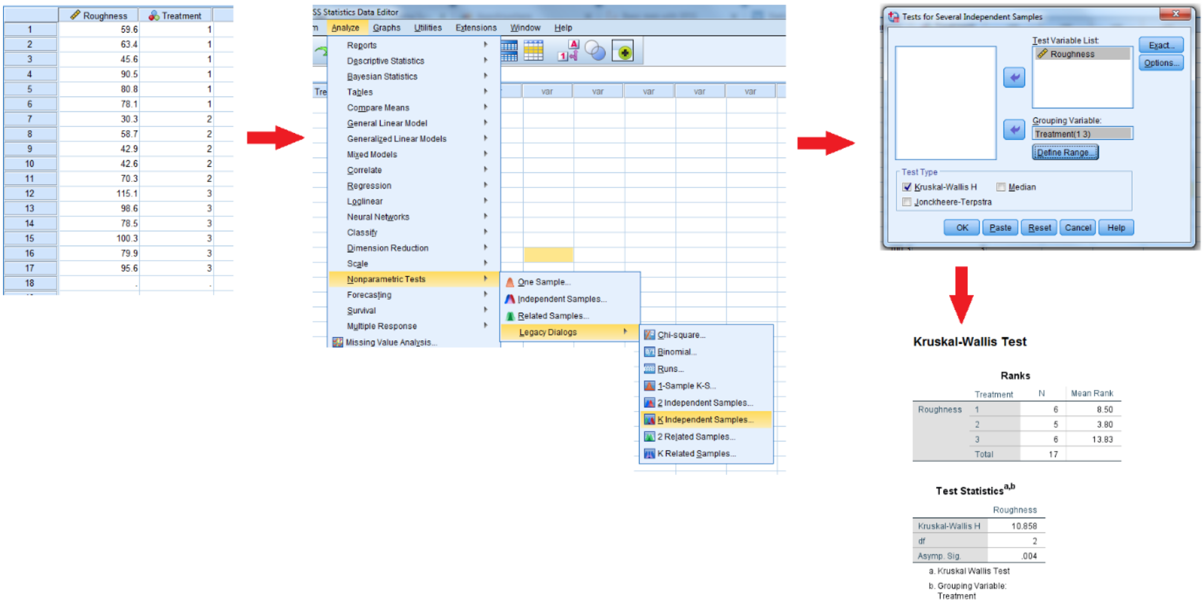

Figure S3. Kruskal-Wallis SPSS Example.