Statistical approaches in surface finishing. Part 1. Introductory review and parametric hypothesis testing

J. R. Smith^{1,*} and C. Larson²

¹ School of Pharmacy and Biomedical Sciences, University of Portsmouth, St Michael's Building, White Swan Road, Portsmouth PO1 2DT, UK

² Glencree, 6 Claxton Rise, Long Clawson, Leics LE14 4PH, UK

*Corresponding author: james.smith@port.ac.uk

Abstract

This paper is the first 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. This article introduces the subject, briefly reviewing the wide use of a number of experimental design tools in recent surface finishing research before starting with a discussion of parametric hypothesis testing, the simplest of the statistical methods.

Keywords: Statistics, Frequency distribution, Parametric, Normality, Heterogeneity, Variance, Student's t-test, ANOVA

Introduction

Statistical methods are essential for quantitative treatment of data and broadly fit into two areas: analysis and design-of-experiments. The first, more common, method describes data in terms of mean, median, standard deviation (σ) etc, and appropriate statistical tests (either parametric or non-parametric) can be used to check whether one group is significantly different from another other ($p < 0.05$; p meaning calculated probability), or not ($p > 0.05$).¹

A less familiar group of methods are those of the ‘design-of-experiments’ (DOE) type. Here, when a large number of variables are involved and wish to be controlled within an experiment, an approach is used to minimise the number of component experiments to be performed, often with ‘high’ and ‘low’ values set for each variable.²⁻⁴ The significance and ‘effect size’ of each variable (or combination of variables) can then be statistically ascertained. This approach needs to be carefully implemented (‘designed’) at the beginning of the set of measurements, and cannot be introduced after the event.

An early article in *Transactions*⁵ examined the subject of statistical process control, directed at helping to improve quality control in manufacturing processes. Standard deviation, tolerances and other basic terms used in statistical process control were explained and discussed. Since this publication,⁵ more than 20 years ago, there has been a marked increase in the use of quite sophisticated statistical tools in design of metal finishing experimentation and analysis of data obtained. Examples of the different design of experiments statistical tools used in metal finishing are widespread, and a short review of the use of some of these, both singly and in combination with other statistical tools, follows below.

In statistics, a central composite design (CCD) is an experimental design, useful in response surface methodology, for building a second order (quadratic) model for the response variable without needing to use a complete three-level factorial experiment. After the designed experiment is performed, linear regression is used, sometimes iteratively, to obtain results.

Response surface methodology (RSM) explores the relationships between several explanatory variables and one or more response variables. Essentially, RSM uses a sequence of designed experiments to obtain an optimal response. Using a second-degree polynomial model to do this often gets results, but firstly estimating a first-degree polynomial model can be easily done using a factorial experiment or a fractional factorial design (FFD). This is enough to show which explanatory variables affect the relevant response variable(s). Once the researcher believes that only significant explanatory variables are left, then a more complicated design, such as a CCD can be implemented to estimate a second-degree polynomial model.

Addach *et al.* used experimental strategies including a factorial portion of CCD and optimum paths coupled with the desirability function to optimise hardness and hydrogen content of chromium coating under pulse reverse electroplating.⁶ Poroeh-Seritan and colleagues used CCD and response surface methodology for statistical modelling and multi-response optimisation of a nickel electroplating process.⁷ Coşkuna *et al.* optimised electrodeposition concentrations for hydroxyapatite coatings on CoCrMo biomedical alloys by use of RSM and CCD. Data obtained from RSM were subjected to analysis of variance (ANOVA) and analysed using a second order polynomial equation.⁸

An experimental design for optimising the corrosion resistance of pulse reverse electrodeposited graphene oxide thin film by measuring linear polarisation resistance (LPR), by Yaghoubinezhad and Afshar used CCD;⁹ the best prediction model proposed a definitive linear relation without any significant interaction of the LPR by analysis of variance.

The optimal electroplating parameters for a pulse-current co-electrodeposition system of Au–Sn deposits in a non-cyanide electrolyte were investigated by Huang *et al.* using FFD and CCD coupled with RSM.¹⁰

RSM and optimisation has been used to study the influence of deposition parameters on the electrodeposition of Cu–Zn¹¹ and Cu–Co¹² alloys in citrate medium. RSM has also been

applied to experimental investigations on redefining the surface quality of bevel gears by pulsed electrochemical honing.¹³

Composition controlling for the non-anomalous plating of Co–Ni and Fe–Co,¹⁴ and Fe–Ni¹⁵ alloys using pulse-reverse electroplating was achieved through means of experimental strategies including FFD, path of steepest ascent and CCD coupled with RSM.

In statistical studies of Zn–Ni alloy electrodeposition using non-cyanide alkaline baths containing polyethyleneimine complexing agents, a 2-level FFD was employed with Pareto charts and normal probability plots of standardised effects to assess factors affecting %Ni deposited.²

In a study on pulsed electrodeposition of cobalt nanoparticles on copper,¹⁶ the influence of the operating parameters on size distribution and morphology was discovered *via* two partially superimposed factorial designs with two factors at two levels. The first factorial design investigated the effect of current density ($i = 10$ and 50 mA cm^{-2}) and discharged cobalt ($Q = 2.5 \times 10^{-3}$ and $1.0 \times 10^{-2} \text{ C}$); the second factorial design investigated the effect of cobalt concentration ($C_0 = 0.01$ and 0.1 M) for the same two levels of Q .

A factorial design 2^2 with 2 centre points was used to find the optimal current density and bath temperature for Ni–W–Fe electrodeposition.¹⁷ The influence of such variables on the cathodic current efficiency and polarisation resistance was obtained.

Another statistical tool in increasing use recently is the Taguchi experimental design approach; this has been employed in the electrodeposition of copper on titanium wires,¹⁸ and the optimisation of electroplating conditions of chromium from hexavalent (CrVI) baths.¹⁹

Increasingly now computerised analysis is coming to the aid of statistical tools. In a study of Ni-diamond composite coating, the prediction of the response variable (vol. % of diamond deposition) was obtained with the help of an empirical relation between the response variable and input variables using an Artificial Neural Network Model (ANN), and also through DOE.²⁰ A computational fluid dynamic (CFD) analysis of eductor agitation in an electroplating tank was carried out by Tong *et al.*²¹

A computerised numerical analysis of film thickness of electrodeposited paint coating for whole automotive bodies using a virtual surface model has been performed by Ohara *et al.*,²² and the calculated results verify the effectiveness of the virtual surface model. The tertiary current distributions on rotating electrodes have been studied using a computerised numerical simulation method.²³

Chemometrics has been successfully applied to functional chromium electroplating by pulse plating techniques by Imaz *et al.*,²⁴ and development of a novel electroplating tank layout by a three-dimensional simulation model, using commercial finite element software, and verification tests has been carried out by Laitinen *et al.*²⁵

This growing evidence of increasing wide usage of statistical tools to design surface finishing research experiments and analyse the data from them has led to a feeling that the time is right for illustrative articles with the emphasis on surface finishing examples, showing the most useful and appropriate tools to use.

In this short series of articles, both the analysis and DOE statistical methods will be examined, using examples from the field of surface finishing. Whilst the analysis methods (mean, standard deviation σ , Student's t-test, ANalysis-Of-Variance ANOVA) are generally well-known, they are revisited here for completeness and because there are some common misconceptions when applying these statistical tests. In addition, a variety of software platforms are now possible to perform such calculations (such as Microsoft Excel, SPSS and GraphPad Prism).

Parametric tests and hypothesis testing

This first article describes the use and concepts behind hypothesis testing of the parametric type (Fig. 1). In statistics, it is usual to speak in terms of a hypothesis: a ‘null hypothesis’ (H_0) being the prediction that there would be no statistical difference between two (or more) groups of data, with an ‘alternative hypothesis’ (H_1) stating that the values between the groups are different (not specifying whether greater or less than at this stage). H_0 can never be accepted or proved, but is either rejected (accept H_1) or not rejected, based on the likelihood of the event.

Parametric statistics is used to describe datasets that are independent, normally distributed and heterogeneous. The former condition is usually easy to decide based on experimental design (*i.e.*, data values in one group are not connected to those in another group), but the latter two need to be statistically tested (described below). In cases where some of these conditions are not met, either the data might be able to be transformed (*e.g.*, the logarithm of the data values might be normally distributed) or non-parametric tests might be applicable.

Data values can be divided into three types: interval (including ratios), ordinal and nominal. Examples of these might include, respectively, surface roughness (R_a , which is a continuous numerical property);^{26,27} compatibility with a particular solvent (an ordered ranking: poor, acceptable, good, excellent, *etc*); and electrolyte composition (with or without additive(s), where there is no rank order). Hypothesis testing using parametric methods can only be applied on interval data.

Sampling, frequency distributions and normality tests

Inherent in most experiments is the concept of sampling: only a fraction of the total population is included in a sample population of measurements. If a sufficient sample size of measurements is taken, however, a representative dataset may be obtained. Various types of frequency histogram may then result, not necessarily giving the well-known normal distribution (‘bell-shaped’ curve). This distribution has particular features and data must fit

this profile, as one of the conditions for the application of parametric hypothesis testing, such as the Student's t-test or ANOVA.

A normal distribution dataset can be constructed in MS Excel, for purpose of an example. If a mean and σ are entered into cells A1 and A2, and the function '=NORMINV(RAND(), \$A\$1, \$A\$2)' is entered into cell C1, a range of values fitting to a normal distribution can be generated by pulling down, perhaps to 200 rows, on the C1 bottom right square (Fig. 2a,b). A histogram of these data can then be plotted, also using MS Excel, by clicking on Data Analysis (on the Data tab in the Analysis group, once the Data Analysis add-in has been installed).²⁸ The data (C1:C200) and bin range (E1:E41) can then be added and the histogram plotted (Fig. 2c). A perfectly normal distribution has the following characteristics: mean = median = mode, symmetry about the centre point, 50% of values less than the mean, 50% greater than the mean and 68.26% of the data within 2 σ (1 σ each side) of the mean (and 99.74% of the data within 5 σ , $\pm 2.5 \sigma$, of the mean).

If the data shown in Fig. 2c represented particle or grain size distribution, *i.e.*, similar to the histograms shown in Ma *et al.*,²⁹ of nanocrystalline electrodeposited nickel and cobalt, was to be compared with a similar histogram of a different sample, then to use a Student's t-test, both histograms are required to be normally distributed. There are a number of ways of checking for normality. It is often convenient to perform statistical analyses in specialised software, such as SPSS or GraphPad Prism. Whilst many of these procedures can be performed in MS Excel, at reduced expense, the procedures can be fairly complicated. An example of the SPSS dataset used to compare two histograms of grain size data (auto-generated using the above normal distribution MS Excel routine) is shown in Fig. 3a. (A detailed documentation of screen output from MS Excel and SPSS is provided in the supplementary material, but is abbreviated in the Figures here for clarity; Figs. S1,S2) Notice the input data is contained in two columns: the first being the grain size (in nm, say) and the second being the sample number; a common mistake is to enter SPSS data in different columns for different groups. It is advisable to lay out the SPSS data in MS Excel and then copy this into SPSS. The 'VAR00001' and 'VAR00002' column names can be renamed to

say 'Grain_size' and 'Sample_number' and the data type set to 'Scale' (continuous, non-integer type) and 'Nominal' (group number, integer type), respectively (Fig. 3b).

By selecting 'Analyze', 'Descriptive Statistics' and 'Explore', and clicking on 'Normality plots with tests', the histograms and the results of the normality test (Fig. 4) and other tables and charts are displayed. A successful pass of the normality test is shown by the Shapiro-Wilk test (to check whether a sample came from a normally distributed population)³⁰ showing no statistical difference amongst the data in both groups ($p > 0.05$), which is the case in this example ($p = 0.811$ and 0.204 , for samples 1 and 2, respectively). It is also possible to check the Q-Q plot (a 'quantile-quantile' plot for graphically comparing two probability distributions) (Fig. 4):³¹ these should be linear if normality is observed. Normality plots may be skewed in negative (left) or positive (right) directions and so prevent the use of the Student's t-test and other statistical comparisons. SPSS provides skewness values for histograms ($+0.04$ and $+0.328$, for samples 1 and 2, respectively, in this example). The data may also have a kurtosis ('peakedness') value that takes it out of the normal-shaped profile; in this example, samples 1 and 2 have kurtosis values of -0.223 and $+0.225$, respectively, which is within the normal range (as reflected by the Shapiro-Wilk test and Q-Q plot observations, Fig. 4).

Tests for homogeneity and Student's t-test

The Student's t-test (and ANOVAs – for comparing more than two groups) assumes that data are independent, normally distributed and have homogeneity of variance (for parametric statistics to be applied). Homogeneity means the variances (σ^2) of populations from which the different samples are drawn are equal and can be checked using a Levene's test (or F-test). As in the Student's t-test, where a normality check is sometimes mistakenly not performed, some investigators are also guilty of not performing the Levene's test. It is important that the conditions (independence, normality and homogeneity) in which the Student's t-test is valid are tested. The same applies for the ANOVA.

To commence the Levene's test and Student's t-test, in SPSS, the data should not be split into the two groups (as it was in the previous section). The 'Independent Samples T Test' option should then be selected from under the 'Analyze' toolbar and then under 'Compare means'. Grain_size should be added as the test variable, and Sample_number as the grouping variable. It is then necessary to select groups 1 and 2 for the comparisons. The output, for the specific example chosen, is shown in the supplementary materials (Fig. S3). Firstly, in this example, the Levene's test has been passed ($p = 0.506$), *i.e.*, the heterogeneity data are not significant, and therefore the results from the Student's t-test can be examined. The p value for this test is $p = 0.848$, which is greater than the critical ('threshold') 0.05 value for significance (sometimes called alpha, α), and hence there is not a statistical difference between the two groups (*NS*, or $p > 0.05$) (supplementary materials, Fig. S3): H_0 , in this example, has not been rejected. The α value is often set to 0.05, but if a smaller margin of error is required, a lower value of α may be used, typically 0.01 or 0.001.

The terms 1- and 2-tail are sometimes used in statistics (supplementary materials, Fig. S3). The former means that only a less than or greater than mean condition is being considered. Usually, and especially when using data with normal distributions, it is more usual to consider whether differences both side of the mean are being investigated for significance testing.

ANOVA

The Student's t-test works for comparing two groups that are independent, normal and homogeneous (parametric). For three or more groups, which must also satisfy these conditions, an ANOVA must be used. Here, the ANOVA checks to see whether there are statistical differences among any of the groups. For example, if five grain size distributions were compared, there might be statistical differences between samples 2 and 4, and 3 and 5 only. The ANOVA reports whether a statistical difference has been found (H_1 accepted), and then a 'post hoc' (Latin: 'after this') analysis would need to be performed to identify which of the pairs were showing the statistical differences. Continuing with the above example for five grain size distribution samples (using similarly auto-generated data), each sample distribution is normally distributed, as identified from the Shapiro-Wilk test (supplementary materials,

Fig. S4). The ANOVA, which is also preceded by a Levene's test, was then instigated in the same manner as the Student's t-test, but selecting 'Analyze', 'Compare Means' and then 'One-Way ANOVA' (supplementary materials, Fig. S4). The one-way term, incidentally, refers to the influence of one categorical independent variable (what samples 1 and 2 represent) on one dependent variable (grain size); a two-way ANOVA might be used to study the influence of two separate additives on grain size, for example. The post-hoc test is then selected, and a typical test being the 'Tukey' test is chosen. This test calculates the minimum distance between two group means that must exist before the difference between the two groups is considered to be statistically significant. Under 'Options' the 'Homogeneity of Variance', 'Brown-Foresythe' and 'Welch' tests should also be selected (supplementary materials, Fig. S5). The former is a Levene's tests, based on median rather than mean values, that offers good robustness,³² and the Welch test is an unequal variance (differing σ values) t-test.³³ In this example, the data were found to be homogeneous ($p = 0.892$), however, no statistical differences were found between any of the pairs using the ANOVA ($p = 0.907$) (supplementary materials, Fig. S4) and so it would obviously be unnecessary to examine the post-hoc analysis table. Creating new sample groups using different means (not all 50 ± 5): 50 ± 5 , 50 ± 5 , 52 ± 5 , 52 ± 5 and 54 ± 5 , for groups 1 to 5, respectively, descriptive statistics were produced and normality observed in all cases ($p > 0.05$, supplementary materials, Fig. S5). The Levene's test showed that the data were homogeneous ($p = 0.825$), but this time, a significant difference was detected using the ANOVA ($p = 0.000$, displayed to 3 decimal places using the SPSS software; double clicking on the p value reveals 4.5×10^{-24} ; supplementary materials, Fig. S5). This means that there is a statistical difference between at least two of the pairs. The post-hoc test showed there to be statistical differences ($p < 0.05$) between all group pairs with the exception of 1 and 2 ($p = 1.000$), and 3 and 4 ($p = 0.689$, supplementary materials, Fig. S6).

Summary

In this first of a short series of articles on the use of statistical methods in surface finishing, after briefly reviewing the broad use of many of the tools available, hypothesis testing has been considered. The importance of making sure data are independent, normal and homogeneous when using parametric statistical methods such as the Student's t-test and ANOVA has been stressed.

Supplementary material

Supplemental data for this article can be accessed at doi:10.1080/00202967.2016.1232851

ORCID

J.R. Smith <http://orcid.org/0000-0001-8805-3788>

References

1. D.M. Lane, 'Online Statistics Education: An Interactive Multimedia Course of Study'. Rice University, University of Houston Clear Lake, and Tufts University. <http://onlinestatbook.com/2/index.html>. Accessed 30 Aug 2015.
2. K. Ramazan and Y. Ugur: 'Statistical studies of Zn-Ni alloy coatings using non-cyanide alkaline baths containing polyethyleneimine complexing agents', *Trans. IMF*, 2014, **92**, (5), 245–252.
3. N. Khanna and J.P. Davim: *Measurement*, 2015, **61**, 280–290.
4. J. Shur, T.G. Nevell, J.K. Shute and J.R. Smith: *Drug Develop. Ind. Pharm.*, 2008, **34**, (6), 559–568.
5. G. Morse: *Trans. IMF*, 1993, **71**, (2), 80–82.
6. H. Addach, P. Berçot, M. Rezrazi, M. De Petris-Wery and H.F. Ayedi: *Trans. IMF*, 2007, **85**, (4), 187–193.
7. M. Porocho-Seritan, S. Gutt, G. Gutt, I. Cretescu, C. Cojocar and T. Severin, *Chem. Eng. Res. Des.*, 2011, **89**, (2), 136–147.
8. M.i. Coşkun, I.H. Karahan, Y. Yücel: *Electrochim. Acta* 2014, **150**, 46–54.
9. Y. Yaghoubinezhad and A. Afshar: *J. Solid State Electrochem.*, 2015, **19**, (5), 1367–1380.
10. M.L. Huang, F.F. Huang, J.L. Pan and T.X. Zhang: *J. Mater. Sci.: Mater. Electron.*, 2014, **25**, (11), 4933–4942.
11. F.B.A. Ferreira, F.L.G. Silva, A.S. Luna, D.C.B. Lago and L.F. Senna: *J. Appl. Electrochem.*, 2007, **37**, (4), 473–481.

12. F.L.G. Silva, J.R. Garcia, V.G.M. Cruz, A.S. Luna, D.C.B. Lago and L.F. Senna: *J. Appl. Electrochem.*, 2008, **38**, 1763–1769.
13. S. Pathak, N.K. Jain and I.A. Palani: *Trans. IMF*, 2016, **94**, (2), 64–69.
14. A. Bai and C.-C. Hu: *Electrochimica Acta*, 2005, **50**, 1335–1345.
15. P. Tsay and C.-C. Hu: *J. Electrochem. Soc.*, 2002, **149**, (10), C492–C497.
16. F. Pagnanelli, P. Altimari, M. Bellagamba, G. Granata, E. Moscardini, P.G. Schiavi and L. Toro: *Electrochim. Acta*, 2015, **155**, 228–235.
17. A.L.M. Oliveira, J.D. Costa, M.B. de Sousa, J.J.N. Alvesa, A.R.N. Campos, R.A.C. Santana and S. Prasad: *J. Alloys Compounds*, 2015, **619**, 697–703.
18. J.L. Rosa, A. Robin, M.B. Silva, C.A. Baldan and P. Peres: *J. Mater. Proc. Technol.*, 2009, **209**, (3), 1181–1188.
19. A.P. Jadid, M. Pourjafar and A. Banaei: *Anal. Bioanal. Electrochem.*, 2014, **6**, (1), 16–27.
20. K. Ramanathan, V.M. Periasamy, U. Natarajan: *Portugaliae Electrochim. Acta*, 2008, **26**, (4), 361–368.
21. L.Z. Tong, K. Ohara, F. Asa, Y. Sugiura: *Trans. IMF*, 2010, **88**, (4), 185–190.
22. K. Ohara, L.Z. Tong, K. Sakita, Y. Hayashi: *Trans. IMF*, 2010, **88**, (6), 294–302.
23. L.Z. Tong: *Trans. IMF*, 2012, **90**, (3), 120–124.
24. N. Imaz, E. Garcia-Lecina, J.A. Diez, M. Ostra and M. Sarret: *Trans. IMF*, 2012, **90**, (5), 259–266.
25. K.T. Laitinen, J.M. Miettinen, M. Vauhkonen and R. Lappalainen: *Trans. IMF*, 2014, **92**, (5), 238–244.
26. H.I. Dawood, K.S. Mahammed, A. Rahmat and M.B. Uday: *Surf. Coat. Technol.*, 2015, **270**, 272–283.
27. J.R. Smith, S. Breakspear and S.A. Campbell: *Trans. IMF*, **81**, (3), 2003, B55–B58.
28. Available at: <https://support.microsoft.com/en-us/kb/214269>. Accessed 28 Aug 2015.
29. C. Ma, S.C. Wang and F.C. Walsh: *Trans. IMF*, 2015, **93**, (1), 8–17.
30. S.S. Shapiro, M.B. Wilk: *Biometrika*, 1965, **52**, (3–4), 591–611.
31. D. Scot: ‘Quantile-Quantile (q-q) Plots’, http://onlinestatbook.com/2/advanced_graphs/q-q_plots.html. Accessed 30 Aug 2015.
32. M.B. Brown, A.B. Forsythe: *J. Am. Stat. Assoc.*, 1974, **69**, 364–367.

33. B.L. Welch: *Biometrika*, 1947, **34**, (1–2), 28–35.

List of Figures

Fig. 1. Types of statistical tests used to compare groups of data.

Fig. 2. (a,b) Generating example data that fit a normal distribution in MS Excel (bin size = 1, in this example);²⁸ (c) resultant frequency distribution (histogram).

Fig. 3. (a) Layout of grouped data in SPSS, showing how to split data into groups ready for histogram plots; (b) the data types (grain size, scale/interval; sample number, nominal) are also shown; grain size (y-axis) is the dependant variable, sample number is the independent variable.

Fig. 4. (a,b) Histograms of grain size data, (c) results from the normality test (Shapiro-Wilk, no significant difference from normal distributions, $p > 0.05$) and (d) a linear Q-Q plot suggesting normality.

Figures

Fig. 1.

Fig. 2.

Fig. 3.

(c)

Tests of Normality							
Sample_number		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
1.00	Grain_size	.040	200	.200 [*]	.995	200	.811
2.00	Grain_size	.041	200	.200 [*]	.990	200	.204

^a. This is a lower bound of the true significance.
^a. Lilliefors Significance Correction

Fig. 4.

Supplementary materials for:

Statistical approaches in surface finishing. Part 1. Introductory review and parametric hypothesis testing

J. R. Smith^{1,*} and C. Larson²

¹ School of Pharmacy and Biomedical Sciences, University of Portsmouth, St Michael's Building, White Swan Road, Portsmouth PO1 2DT, UK

² Glencree, 6 Claxton Rise, Long Clawson, Leics LE14 4PH, UK

*Corresponding author: james.smith@port.ac.uk

Fig. S1. Generating example data that fit a normal distribution in MS Excel (bin size = 1, in this example); resultant frequency distribution (histogram).

Fig. S2. Layout of grouped data in SPSS, showing how to split data into groups ready for histogram plots. The data types (grain size, scale/interval; sample number, nominal) are also shown. Grain size (y-axis) is the dependant variable, sample number is the independent variable.

T-Test

Group Statistics

	Sample_number	N	Mean	Std. Deviation	Std. Error Mean
Grain_size	1.00	200	50.2492	5.11214	.36148
	2.00	200	50.1520	5.00082	.35361

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Grain_size	Equal variances assumed	.442	.506	.192	398	.848	.09718	.50568	-.89696	1.09131
	Equal variances not assumed			.192	397.807	.848	.09718	.50568	-.89696	1.09131

Fig. S3. Levene's test and Student's t-test results showing, respectively, no significance difference in homogeneity of variance ($p > 0.05$), but no significant difference in means ($p > 0.05$) between the two grain size groups.

Tests of Normality							
Sample_number		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
1.00	Grain_size	.040	200	.200 [*]	.995	200	.811
2.00	Grain_size	.041	200	.200 [*]	.990	200	.204
3.00	Grain_size	.051	200	.200 [*]	.994	200	.591
4.00	Grain_size	.033	200	.200 [*]	.994	200	.675
5.00	Grain_size	.034	200	.200 [*]	.996	200	.927

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

One-Way ANOVA

Dependent List: Grain_size

Factor: Sample_number

OK Paste Reset Cancel Help

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

☐ LSD ☐ S-N-K ☐ Waller-Duncan

☐ Bonferroni ☒ Tukey Type I/Type II Error Ratio: 100

☐ Sidak ☐ Tukey's-b ☐ Dunnett

☐ Scheffe ☐ Duncan Control Category: Last

☐ R-E-G-W F ☐ Hochberg's GT2

☐ R-E-G-W Q ☐ Gabriel

Test: ☒ 2-sided ☐ < Control ☐ > Control

Equal Variances Not Assumed

☐ Tamhane's T2 ☐ Dunnnett's T3 ☐ Games-Howell ☐ Dunnnett's C

Significance level: 0.05

Continue Cancel Help

One-Way ANOVA: Options

Statistics

☐ Descriptive

☐ Fixed and random effects

☒ Homogeneity of variance test

☒ Brown-Forsythe

☒ Welch

☐ Means plot

Missing Values

☒ Exclude cases analysis by analysis

☐ Exclude cases listwise

Continue Cancel Help

Oneway

Test of Homogeneity of Variances

Grain_size				
Levene Statistic	df1	df2	Sig.	
.278	4	995	.892	

ANOVA

Grain_size					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25.288	4	6.322	.254	.907
Within Groups	24723.746	995	24.848		
Total	24749.035	999			

Robust Tests of Equality of Means

Grain_size				
	Statistic ^a	df1	df2	Sig.
Welch	.251	4	497.466	.909
Brown-Forsythe	.254	4	993.894	.907

a. Asymptotically F distributed.

Fig. S4. *Previous page:* Tests for normality of each grain size group (passed, $p > 0.05$) and homogeneity (passed, $p > 0.05$) and one-way ANOVA (no significant difference, $p > 0.05$) between the groups. Tukey post-hoc test selected.

Descriptives

Descriptive Statistics										
Sample_number		N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
		Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
1.00	Grain_size	200	37.01	63.20	49.6590	5.23820	-.006	.172	-.173	.342
	Valid N (listwise)	200								
2.00	Grain_size	200	37.33	64.05	49.6501	5.04158	.031	.172	-.311	.342
	Valid N (listwise)	200								
3.00	Grain_size	200	39.62	65.00	51.8909	4.91210	.104	.172	-.017	.342
	Valid N (listwise)	200								
4.00	Grain_size	200	39.48	68.38	52.5518	5.23108	.298	.172	.045	.342
	Valid N (listwise)	200								
5.00	Grain_size	200	39.13	66.56	54.2841	4.93051	-.253	.172	-.118	.342
	Valid N (listwise)	200								

Tests of Normality							
Sample_number		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
1.00	Grain_size	.045	200	.200 [*]	.994	200	.539
2.00	Grain_size	.057	200	.200 [*]	.995	200	.682
3.00	Grain_size	.051	200	.200 [*]	.993	200	.457
4.00	Grain_size	.041	200	.200 [*]	.993	200	.429
5.00	Grain_size	.051	200	.200 [*]	.991	200	.246

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Test of Homogeneity of Variances

Grain_size

Levene Statistic	df1	df2	Sig.
.378	4	995	.825

ANOVA

Grain_size

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3152.897	4	788.224	30.632	.000
Within Groups	25603.168	995	25.732		
Total	28756.065	999			

Robust Tests of Equality of Means

Grain_size

	Statistic ^a	df1	df2	Sig.
Welch	30.845	4	497.405	.000
Brown-Forsythe	30.632	4	991.923	.000

a. Asymptotically F distributed.

Fig. S5. Tests for normality of each grain size group (passed, $p > 0.05$) and homogeneity (passed, $p > 0.05$) and one-way ANOVA between the groups. The ANOVA shows there to be a statistical difference between at least one of the pairs ($p < 0.05$).

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Grain_size

Tukey HSD

(I) Sample_number	(J) Sample_number	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	.00888	.50727	1.000	-1.3774	1.3951
	3.00	-2.23196*	.50727	.000	-3.6182	-.8457
	4.00	-2.89287*	.50727	.000	-4.2791	-1.5066
	5.00	-4.62512*	.50727	.000	-6.0114	-3.2389
2.00	1.00	-.00888	.50727	1.000	-1.3951	1.3774
	3.00	-2.24084*	.50727	.000	-3.6271	-.8546
	4.00	-2.90175*	.50727	.000	-4.2880	-1.5155
	5.00	-4.63401*	.50727	.000	-6.0203	-3.2478
3.00	1.00	2.23196*	.50727	.000	.8457	3.6182
	2.00	2.24084*	.50727	.000	.8546	3.6271
	4.00	-.66091	.50727	.689	-2.0472	.7253
	5.00	-2.39317*	.50727	.000	-3.7794	-1.0069
4.00	1.00	2.89287*	.50727	.000	1.5066	4.2791
	2.00	2.90175*	.50727	.000	1.5155	4.2880
	3.00	.66091	.50727	.689	-.7253	2.0472
	5.00	-1.73225*	.50727	.006	-3.1185	-.3460
5.00	1.00	4.62512*	.50727	.000	3.2389	6.0114
	2.00	4.63401*	.50727	.000	3.2478	6.0203
	3.00	2.39317*	.50727	.000	1.0069	3.7794
	4.00	1.73225*	.50727	.006	.3460	3.1185

*. The mean difference is significant at the 0.05 level.

Fig. S6. Post-hoc analysis (Tukey test, in this case) identifies which groups have statistical differences between means (here, all except groups 1 & 2 and 3 & 4) once the ANOVA has shown there to be a difference between at least one pair.