Subjective sensitivity data: Considerations to treat heteroscedasticity

Daniel Schmidt\(^1\)*, Andresa M.C. Germano\(^1\) and Thomas L. Milani\(^1\)

Abstract: In many clinical disciplines, especially in subjective sensitivity research, data are collected on the ratio scale. This is also true for vibration perception thresholds (VPTs). Hence, such data may be of heteroscedastic nature, where the amount of measurement error increases as the values increase. Heteroscedastic data is a great scientific concern, since many standard inferential statistical tests assume homoscedasticity and may, therefore, lead to incorrect interpretations. Unfortunately, only few studies even address heteroscedasticity, and even if they do, in many cases this is achieved inappropriately. Therefore, in this paper, we first show a simple approach for how to detect heteroscedasticity in VPTs measured at the plantar aspect of the hallux. Second, we show how to correct for heteroscedasticity by means of logarithmizing the raw data. Finally, we also demonstrate the importance of how data benefits from this transformation. Hence, if data takes benefit from the transformation, we recommend using logarithmized data to conduct further statistical approaches. We also interpret the results presented here. Furthermore, we briefly mention alternatives in case the transformation is not successful in eliminating heteroscedasticity.

Subjects: Neuroscience; Statistics & Probability; Health & Society; Health Conditions; Medicine

ABOUT THE AUTHOR
The research activities of the present group are related to the interaction of the human sensory and motor system. In particular, basic research related to plantar mechanoreceptors is one research focus (for example, determining sensory thresholds, influencing factors of cutaneous sensitivity). Furthermore, the contribution of plantar mechanoreceptors towards balance is another research emphasis. In this context, various experiments based on quasi-static and dynamic balance tasks are performed. Additionally, clinical studies with various patient groups (e.g. MCI, Parkinson) are implemented with focus on sensory and motor behavior.

PUBLIC INTEREST STATEMENT
Assessing skin sensitivity (e.g. how sensitive are you in terms of touch or vibration) is widely used, and very important in many clinical facilities. For example, in diseases like Diabetes mellitus or neuropathy, skin sensitivity is impaired. However, skin sensitivity also naturally decreases as we age. This pronounces the importance to accurately assess skin sensitivity in order to avoid false interpretations of test results. Among many other kinds of measurements, test results of skin sensitivity (such as vibration perception thresholds, VPTs), commonly exhibit a certain data pattern called “heteroscedasticity”. This means that the larger the VPTs (the more insensitive the person), the larger the variability, in other words, the error (and vice versa). When heteroscedasticity is not treated appropriately, the above-mentioned false interpretations of skin sensitivity may occur, which is problematic for clinicians, and patients. Since heteroscedasticity is still not considered appropriately in many cases, this paper clearly addresses how to deal with this problem.
Keywords: heteroscedasticity; log transformation; vibration perception thresholds; VPTs; skin sensitivity; ratio scale

1. Introduction
Skin is the largest organ in the human body and enables various external stimuli, such as touch or vibration, to be detected. This is achieved by mechanoreceptors located in the skin. For example, plantar mechanoreceptors are known to contribute to human balance regulation (Horak, Nashner, & Diener, 1990; Kavounoudias, Roll, & Roll, 1998; Peterka, 2018). Plantar sensitivity, however, depends on various factors like age (Gescheider, Bolanowski, Hall, Hoffman, & Verrillo, 1994) or skin temperature (Germano, Schmidt, & Milani, 2016; Schmidt, Germano, & Milani, 2017). Furthermore, the ability of balance regulation is impaired in various groups, such as patients with peripheral neuropathy, in which plantar sensitivity deteriorates (Inglis, Horak, Shupert, & Jones-Rycewicz, 1994). This highlights the clinical importance of skin sensitivity in order to detect certain diseases. Of course, the interpretation of the outcomes of skin sensitivity measurements must be correct in order to meet clinical requirements. Skin sensitivity may be assessed invasively inserting microelectrodes into afferent nerve fibers (microneurography). Other, more common approaches are known as subjective methods, such as determining vibration sensitivity (vibration perception thresholds, VPTs).

For vibration sensitivity, as well as in many other research disciplines, data are collected on the ratio scale. Such data may not only be expressed as whole numbers, but also as any number of decimal points (Safrit, 1989). In ratio scaled data, the amount of measurement error (or measurement difference, e.g. in a test-retest scenario) increases as the measured values increase (Atkinson & Nevill, 1998): The measurement error depends on the size of the measured parameter (Nevill & Atkinson, 1997). Such data will accumulate close to zero (without negativity). On the other hand, this kind of data is theoretically unbounded towards the positive direction, likely resulting in heteroscedastic patterns (Nevill & Atkinson, 1997). In contrast to homoscedastic data (especially when following normal distribution), heteroscedastic data should not be analyzed using most conventional statistical methods (Atkinson & Nevill, 1998), as the power and ability to control type I error probability might then be negatively affected (Wilcox, Peterson, & McNitt-gray, 2018). Conducting statistical tests on raw data may lead to substantial misinterpretations and error-prone conclusions.

Although VPT data are subjective and measured on the ratio scale, most VPTs are a) either still analyzed as homoscedastic data (e.g. Alfuth & Rosenbaum, 2011; Flondell et al., 2017; Jansson, Hakansson, Reinfeldt, Fröhlich, & Rahne, 2017; Meyer, Oddsson, & De Luca, 2004, to name a few), or b) sometimes transformed, however, without giving information regarding a validation of its benefit, or with incorrect justifications for performing a transformation. The aim of the present study, therefore, is first to examine whether VPTs exhibit heteroscedasticity in an appropriate and easy-to-understand way. We hypothesized the presence of heteroscedasticity in our vibration data. Second, in the case of heteroscedasticity, we also provide an example of how to correct for this in a simple way, and we demonstrate the necessity to check whether or not the correction was appropriate. Since the main clientele working with ratio-scaled data may not have a strong background in statistics, we deliberately focus on a simple and easy-to-conduct approach. However, we will briefly touch on current recommendations for alternatives in case the elimination of heteroscedasticity is not satisfactory with the simple approach we refer to in this study.

2. Methods

2.1. Subjects
Twenty-eight healthy and injury-free subjects of both genders (16 males, 12 females) participated in this study (mean±SD: 23.2 ± 2.6 yrs, 173.5 ± 8.0 cm, 66.9 ± 9.8 kg). Subjects had no signs of lower extremity pain during this study and were free of injuries for at least 6 months prior to testing. All
participants were also free of neurological diseases (e.g. diabetic neuropathy or Parkinson’s disease). Prior to testing, all subjects were informed about the aim and procedures of this study and gave their written informed consent. Participants were free to withdraw from the experiments at any time. This study was conducted according to the recommendations of the Declaration of Helsinki and was approved by the Ethics Committee of the faculty of the corresponding university.

2.2. Instrumentation
Cutaneous vibration perception thresholds (VPTs, in µm) were determined using a Tira Vib vibration exciter (model TV51075, Schalkau, Germany), powered by a Voltcraft oscillator (model FG 506, Hirschau, Germany). Before the measurements, the vertical movement of the contactor of the vibration exciter was laser calibrated in order to enable direct readings of the vibrating amplitude. The probe diameter was 7.8 mm, and its tip was placed 2 mm above the surrounding surface (Nurse & Nigg, 1999). The frequency of the vibrating contactor was 200 Hz, which is known to be the optimal stimulus to elicit Vater-Pacini-corpuscles (Gescheider et al., 1994; Verillo, 1985). Since skin temperature influences skin sensitivity (Germano et al., 2016; Schlee, Sterzing, & Milani, 2009; Schmidt et al., 2017), the surface of the vibration exciter could be heated to keep plantar temperatures constant. The vertical force applied from the subject’s foot towards the contactor was monitored via a force transducer and was kept within a range of ±0.5 N. Plantar temperatures (hallux) and room temperature were measured using a digital type-K-thermocouple (PeakTech 5135, Ahrensburg, Germany).

2.3. Testing procedure
Prior to testing, subjects went through an acclimatization period of 10 min to adjust to the room temperature (23 ± 2 °C, according to EN ISO/IEC 17,025). During the sensitivity tests, subjects closed their eyes and wore noise cancelling earphones (Quiet comfort 20i, Bose, Framingham, USA) in order to avoid distracting noise from the environment. All participants sat with ankle, knee, and hip-angles of approx. 90°, and rested their arms on top of their thighs, close to the abdomen. Subjects were instructed to sit as described above, but also to be comfortable to enable them to concentrate on detecting vibration stimuli but not on maintaining a certain posture. VPTs were measured similarly to a Method of Limits approach introduced by Mildren, Strzalkowski, and Bent (2016). First, three VPT-trials were collected at the Hallux (barefoot) of the dominant foot (test). After this, subjects waited for 45 min while sitting with their feet on top of the heatable aluminum platform. Finally, VPTs were measured in the same manner (retest). Plantar temperatures only slightly decreased from 26.5 ± 2.8 °C (test) to 25.5 ± 2.1 °C (retest).

2.4. Data analysis and statistics
In terms of VPTs, the mean±SD of the three trials was calculated. All further analyses were performed using R (The R Foundation for Statistical Computing, Vienna, Austria). To check for potential outliers, we used an adjusted boxplot rule as recommended by Wilcox (2013). As suggested (Nevill & Atkinson, 1997), heteroscedasticity was identified by plotting VPT differences (test-retest, raw data) against their mean. The Spearman-correlation (raw data) was also calculated between the absolute differences of the VPTs and their mean. Data was then treated by taking the natural logarithm of the raw data (Nevill & Atkinson, 1997; Wilcox et al., 2018). Bland-Altman plots were created, whereas the limits of agreement were calculated by multiplying 1.96 times the standard deviation of the mean difference.

3. Results and discussion
No outliers were detected by the adjusted boxplot rule. This is important when using means, as in our study, since outliers may considerably lower the power of e.g. difference tests (Wilcox et al., 2018). Raw VPTs (mean±SD) of the test and retest are presented in Table 1. A first quite simple sign towards heteroscedasticity was evident when plotting VPT differences (test-retest) against their mean (Figure 1). Smaller mean VPTs were associated with smaller error (differences), whereas greater mean VPTs were associated with greater error. Data points seem to be “funnel-shaped”, indicating that the measurement error increases as the measured values increase. Another indicator
of heteroscedasticity was that differences (not necessarily individual data from the test and retest, respectively) were not normally distributed (Shapiro-Wilk test: \( p = 0.002 \)). As the ultimate criterion, the Spearman-correlation between the absolute differences and their mean (Nevill & Atkinson, 1997) was significant (\( p = 0.004 \)), with a correlation coefficient of \( r = 0.530 \), indicating a moderate correlation (Mukaka, 2012). We used a Spearman correlation, since it does not require normally distributed data (Mukaka, 2012). This means that the error correlates with the magnitude of the VPT values, constituting heteroscedasticity. The correlation is not required to be significant or to present a high correlation coefficient. A small sample size may account for the lack of significance, but may still correspond to heteroscedastic data when the correlation is found to be positive (Nevill & Atkinson, 1997), even if of lower magnitude. It is then generally recommended to take the natural logarithm (hereafter referred to as log) of the raw data (Nevill & Atkinson, 1997).

To provide a first example regarding the benefit of transforming heteroscedastic data, we created a Bland-Altman plot (Figure 1) based on the raw data from Table 1. Bland-Altman plots

| Subject | VPT Test [\( \mu m \)] (mean\( \pm \)SD) | VPT Retest [\( \mu m \)] (mean\( \pm \)SD) |
|---------|----------------------------------------|----------------------------------------|
| 1       | 2.68 ± 0.86                            | 2.40 ± 1.15                            |
| 2       | 0.68 ± 0.22                            | 0.85 ± 0.05                            |
| 3       | 2.39 ± 0.12                            | 1.20 ± 0.08                            |
| 4       | 0.48 ± 0.11                            | 0.47 ± 0.08                            |
| 5       | 0.70 ± 0.04                            | 0.97 ± 0.19                            |
| 6       | 2.11 ± 0.45                            | 1.73 ± 0.15                            |
| 7       | 1.53 ± 0.82                            | 2.65 ± 0.31                            |
| 8       | 1.80 ± 0.11                            | 1.93 ± 0.46                            |
| 9       | 2.26 ± 0.57                            | 2.59 ± 0.24                            |
| 10      | 0.21 ± 0.04                            | 0.40 ± 0.08                            |
| 11      | 0.56 ± 0.11                            | 0.37 ± 0.04                            |
| 12      | 1.29 ± 0.03                            | 1.30 ± 0.98                            |
| 13      | 1.47 ± 0.21                            | 1.39 ± 0.20                            |
| 14      | 0.50 ± 0.19                            | 0.65 ± 0.33                            |
| 15      | 1.98 ± 0.44                            | 2.06 ± 0.53                            |
| 16      | 1.71 ± 0.17                            | 2.47 ± 0.33                            |
| 17      | 1.87 ± 0.52                            | 2.31 ± 0.61                            |
| 18      | 2.25 ± 0.28                            | 2.51 ± 0.14                            |
| 19      | 2.74 ± 0.90                            | 3.67 ± 0.18                            |
| 20      | 0.44 ± 0.22                            | 0.84 ± 0.40                            |
| 21      | 2.14 ± 0.43                            | 2.33 ± 0.48                            |
| 22      | 2.13 ± 0.33                            | 1.13 ± 0.12                            |
| 23      | 0.56 ± 0.16                            | 1.14 ± 0.07                            |
| 24      | 1.48 ± 0.50                            | 2.42 ± 0.33                            |
| 25      | 0.61 ± 0.07                            | 0.53 ± 0.13                            |
| 26      | 1.33 ± 0.22                            | 4.01 ± 0.38                            |
| 27      | 1.07 ± 0.47                            | 1.13 ± 0.47                            |
| 28      | 0.44 ± 0.04                            | 0.43 ± 0.04                            |
| All     | 1.41 ± 0.78                            | 1.64 ± 0.99                            |
are an important tool to quantify absolute reliability (Bruton, Conway, & Holgate, 2000). In accordance with Bland and Altman (Bland & Altman, 1986), and as can be observed in Figure 1, the limits of agreement were unacceptably far apart for small values, and unacceptably narrow for large values: For example, subject 28 obtained a VPT of 0.44 µm at the test (Table 1), hence data from this subject could range between 1.6 and −1.2 µm (0.44 + 1.131 µm and 0.44 −1.593 µm, respectively). This is not possible due to the ratio scale, which does not allow negative values. In conclusion, the absolute and fixed limits of agreement in Figure 1 do not consider ratio properties of data. Therefore, an appropriate interpretation of Bland-Altman plots based on heteroscedastic and untransformed data is not possible and may lead to overestimation of apparently poor reliability. Therefore, we calculated the natural logarithm (log transformation) of raw data from Table 1. When we then plot VPT differences (test-retest) against their mean, data no longer appear “funnel-shaped” (Figure 2). Instead, differences between tests and retests are then normally distributed (Shapiro-Wilk test: p = 0.724). Furthermore, the Spearman correlation coefficient decreased to close to zero (r = −0.020), indicating a negligible correlation (Mukaka, 2012). Additionally, the correlation turned out to be non-significant (p = 0.919). Hence, the error does not correlate with the magnitude of the values, and data is now of homoscedastic nature (Atkinson & Nevill, 1998). Please note that if the correlation coefficient is numerically reduced by the log transformation, irrespective of the sign, it is usually recommended to transform data logarithmically (Nevill & Atkinson, 1997). This was the case for our data. In other data sets (not related to this paper), we found that a log transformation did not always eliminate heteroscedasticity, as evident by an increasing (absolute) value of the Spearman correlation coefficient. In such cases, other approaches may be adequate (see last paragraph of discussion). Note, however, that most articles using log transformed data do not demonstrate whether or not the transformation induced any benefit toward eliminating heteroscedasticity. This is a clear limitation, and merely (log) transforming data prior to conducting statistical tests should not be regarded as a rigid or proper template. On the other hand, there are recent studies which define a certain \( r^2 \)-value as the “threshold” between homoscedastic vs. heteroscedastic data (Pérez-Castilla, Feriche, Jarić, Padial, & Garcia-Ramos, 2017; Plautard et al., 2017). However, not considering other parameters to identify heteroscedasticity may lead to restrictions when interpreting data.

Figure 1. Bland-Altman plot of raw, untransformed vibration perception thresholds (VPTs) based on data from Table 1. The upper limit of agreement (ULOA: 1.131 µm) equals the bias (−0.231 µm) plus the random error component (1.362 µm). The lower limit of agreement (LLOA: −1.593 µm) equals the bias (−0.231 µm) minus the random error component (1.362 µm).
Based on the log data (Figure 2), we again created a Bland-Altman plot. To be able to calculate on a ratio scale, we followed the recommendation as described in Nevill and Atkinson (1997) and calculated the antilogs:

$$\log(bias) = -0.144 \mu m$$

$$\text{antilog}(bias) = 0.866 \mu m$$

$$\log(\text{random error}) = 0.771 \mu m$$

$$\text{antilog}(\text{random error}) = 2.162 \mu m$$

After this, we calculated the ratio upper and lower limits of agreements (rULOA and rLLOA, respectively) as follows:

$$rULOA = \frac{\text{antilog}(bias)}{\text{antilog}(\text{random error})} = \frac{0.866 \mu m}{2.162 \mu m} = 0.401 \mu m$$

$$rLLOA = \frac{\text{antilog}(bias)}{\text{antilog}(\text{random error})} = \frac{0.866 \mu m}{2.162 \mu m} = 1.872 \mu m$$

Going back to our example, subject 28 (0.44 µm at test) could vary between 0.44 × 1.872 µm and 0.44 × 0.401 µm, which is between 0.82 and 0.18 µm (in 95 %). The limits of agreement are now more accurate, since they constitute a percentage change and account for the ratio properties. Therefore, a correct interpretation is now possible.

As mentioned earlier, many other standard inferential statistics assume homoscedasticity (Wilcox et al., 2018), something researchers may not always be aware of. These also include common statistical tests to detect differences (bias), such as t-tests (Atkinson & Nevill, 1998; Markowski & Markowski, 1990; Zimmerman, 2004). When considering different statistical procedures, the general class of analysis of variance (ANOVA), for example, assumes equality of variances within the groups (Zimmerman, 2004). In other words, it assumes that data is homoscedastic (Keselman et al., 1998; Kirchman, Sigda, Kapuscinski, & Mitchell, 1982). Nonparametric tests, like the Mann-Whitney-U test, account for violations of normal distribution, but not if the assumption of homoscedasticity is violated (Zimmerman & Zumbo, 1993). This is in line with other investigations (Keselman, Ragan, & Feir-Walsh, 1977; Pratt, 1964; Tomarken & Serlin, 1986). When executed on both raw and log-transformed data, the same tests for bias result in different p-values, which makes appropriate interpretations more difficult. Furthermore, conducting
statistical tests (which assume homoscedasticity) on heteroscedastic data, may indeed alter type I error rates (Zimmerman, 2004). In our data set (Table 1), p-values of the Wilcoxon test were slightly different from each other: 0.046 for log, and 0.045 for raw data (test vs. retest). In addition, conventional correlation analyses (e.g. Pearson correlation, intra-class-correlation (ICC)) are judged to be inappropriate when conducted on heteroscedastic data (Nevill, 1997). Chinn (1991) states that when data are analyzed on a log scale, all calculations should be performed on the log values. For our data set, the ICC coefficient was 0.821 based on raw, and 0.914 based on log data (both p<0.001, model 3,k). Heteroscedasticity also affects measures of variability and absolute reliability: The coefficient of variation (CV) assumes the presence of heteroscedasticity, whereas the standard error of measurement (SEM) assumes the absence of heteroscedasticity (Atkinson & Nevill, 1998). This highlights the importance of carefully analyzing data to be able to conduct adequate statistical tests and to interpret these tests appropriately.

Note again that it is important to check data for outliers. Since we detected no outliers in our data set, we did not include robust statistical methods. Many statistical measures or procedures are not robust against outliers (Wilcox, 2013). This includes the use of means and standard deviations, as well as conventional correlation analyses, such as Pearson or Spearman correlations. In such cases, it is recommended to use trimmed means and the standard error of trimmed means (Wilcox, 2013). An alternative to Spearman’s correlation, which may be distorted by properly placed outliers, are type O correlations. One type of these are e.g. skipped correlations, which downweight or eliminate values of low depth within the data cloud (Wilcox, 2013). Conducting a skipped correlation (center estimator: MCD) on our data resulted in the same interpretation and conclusion regarding the benefit of the logarithmic transformation.

As already mentioned, the present study shows one simple-to-conduct possibility to treat heteroscedasticity. This treatment is easy to understand, which is a clear advantage for the wide range of potential “users”. These may include clinicians, (general) practitioners, or physiotherapists who need to compare means of different groups, for example. Such a versatile clientele might exhibit restricted knowledge in terms of more sophisticated statistical analyses, or, access to appropriate software may simply be difficult. Therefore, we focused on a simple log transformation to treat heteroscedasticity.

In this regard, however, we must also mention that this simple approach may exhibit limitations. A logarithmic transformation does not always yield homoscedastic data. This might be true particularly when correlation coefficients decrease only slightly after logarithmic transformation. Additionally, recent studies showed that sometimes there is no normal distribution (of differences) after logarithmization, or that there are still outliers/a skewed distribution (Rasmussen, 2013; Wilcox et al., 2018). In these cases, when comparing group means, one recommendation is to test the assumption of equal variances by using e.g. the F-test. If there is no significance, one is usually guided to use Student’s t-test. However, this approach is not advisable, since in some instances the F-test fails to detect violations of the assumption of equal variances, or may falsely recommend of using the t-test (especially when sample sizes are not equal) (Hayes & Cai, 2007; Markowski & Markowski, 1990; Wilcox et al., 2018; Zimmerman, 2004). Instead, a rank-based test as introduced by Conover (1980) constitutes a more robust method. When comparing group variances, Levene’s test is also commonly implemented. However, in a more recent study this test was found not to be very robust (Hayes & Cai, 2007). Alternatively, the Brown-Forsythe test and the O’Brien test were found to constitute fairly robust tests (e.g. Howell, 1996), although they also have their weaknesses (Hayes & Cai, 2007). In an extensive simulation study performed by Hayes and Cai (2007), they found that so-called unconditional tests performed as good as or even superior to the conditional tests mentioned above (Brown-Forsythe test, O’Brien test, etc.). In particular, bootstrapping the sampling distribution of the separate variance t statistic (UBS), as an unconditional test, resulted in an overall superior performance in most of the simulated conditions (Hayes & Cai, 2007). The superior performance of unconditional separate-variance tests is particularly pronounced when the sample sizes are not equal (Zimmerman, 2004). Hence, these
methods are recommended when treating heteroscedasticity and when detecting differences between groups. As already mentioned, however, such approaches and recommendations require more in-depth statistical knowledge. Additionally, the fact that most statistical textbooks still recommend preliminary conditional tests to investigate the equality of variances contributes to this "dilemma".

4. Conclusion
In this paper, we demonstrated that vibration perception thresholds (VPTs) measured at the plantar aspect of the hallux exhibit heteroscedasticity. Heteroscedastic data is a great scientific concern, since many standard inferential statistical procedures and measures assume homoscedasticity and may, therefore, lead to incorrect interpretations and conclusions. Such measures include, for example, parametrical and non-parametrical tests to detect differences, correlation analyses, or Bland-Altman plots. Based on an existing data set, we corrected for heteroscedasticity by calculating the natural logarithm (log transformation) of raw data. We further provided examples of how our data set benefitted from this correction. In general, we recommend: First, to examine any type of data for heteroscedasticity, especially when measured on the ratio scale. Second, to use log data and to check if data took benefit of the log transformation. Finally, to use log transformed data for conducting further statistical approaches if data benefitted from the transformation. In case the log transformation does not yield homoscedastic data, other more sophisticated approaches based on unconditional tests are recommended.

Acknowledgements
Special thanks to the European Social Fund (ESF) for providing scholarship and to Lisa Peterson for proof reading.

Funding
The publication costs of this article were funded by the German Research Foundation/DFG and the Chemnitz University of Technology in the funding program Open Access Publishing [grant number ZY 1121/1-4].

Competing interests
The authors declare no competing interest.

Author details
Daniel Schmidt
E-mail: daniel.schmidt@hsw.tu-chemnitz.de
Andresa M.C. Germano
E-mail: andresa.germano@hsw.tu-chemnitz.de
Thomas L. Milani
E-mail: thomas.milani@hsw.tu-chemnitz.de
1 Department of Human Locomotion, Chemnitz University of Technology, Institute of Human Movement Science and Health, Reichenhainer Straße 31/33, Chemnitz 09126, Germany.

Citation information
Cite this article as: Subjective sensitivity data: Considerations to treat heteroscedasticity, Daniel Schmidt, Andresa M.C. Germano & Thomas L. Milani, Cogent Medicine (2019), 6: 1673086.

References
Alfluth, M., & Rosenbaum, D. (2011). Long distance running and acute effects on plantar foot sensitivity and plantar foot loading. Neuroscience Letters, 503, 58–62. doi:10.1016/j.neulet.2011.08.010
Atkinson, G., & Nevill, A. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine, 26, 217–238. doi:10.2165/00007256-199826040-00002
Blond, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet, 1, 307–310. doi:10.1016/S0140-6736(86)90837-8
Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? Physiotherapy, 86, 94–99. doi:10.1016/S0031-9408(05)61211-4
Chinn, S. (1993). Repeatability and method comparison. Thorax, 46, 454–456. doi:10.1136/thx.46.6.454
Conover, W. (1980). Practical nonparametric statistics (2nd ed.). New York: John Wiley.
Flandell, M., Rosén, B., Andersson, G., Schyman, T., Dahlen, L. B., & Björman, A. (2017). Vibration thresholds in carpal tunnel syndrome assessed by multiple frequency vibrometry: A case-control study. Journal of Occupational Medicine and Toxicology, 12, 34. doi:10.1186/s12995-017-0181-6
Germano, A., Schmidt, D., & Milani, T. (2016). Effects of hypothermically reduced plantar skin inputs on anticipatory and compensatory balance responses. BMC Neuroscience, 17, 1–14. doi:10.1186/s12868-016-0279-2
Gescheider, G. A., Bolanowski, S. J., Hall, K. L., Hoffman, K. E., & Verrillo, R. T. (1994). The effects of aging on information-processing channels in the sense of touch: 1. Absolute sensitivity. Somatosensory & Motor Research, 11, 345–357. doi:10.3109/0899029409028878
Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. British Journal of Mathematical and Statistical Psychology, 60, 217–244. doi:10.1348/000711007X62576
Horak, F. B., Nashner, L. M., & Diener, H. C. (1990). Postural strategies associated with somatosensory and vestibular loss. Experimental Brain Research, 82, 167–177. doi:10.1007/BF00230848
Howell, D. (1996). Statistical methods for psychology (4th ed.). Belmont, CA: Duxbury.
Inglis, J. T., Horak, F. B., Shupert, C. L., & Jones-Rycewicz, C. (1996). The importance of somatosensory information in triggering and scaling automatic postural responses in humans. Experimental Brain Research, 101, 159–164. doi:10.1007/BF00243226
Jansson, K.-J. F., Hakansson, B., Reinfeldt, S., Fröhlich, L., & Rahne, T. (2017). Vibrotactile thresholds on the mastoid and forehead position of deaf patients using
radiecoer B71 and B81. Ear and Hearing, 38, 714–723. doi:10.1097/AUD.0000000000000456
Kavounoudias, A., Roll, R., & Roll, J. P. (1998). The plantar sole is a “dynamometric map” for human balance control. Neuroreport, 9, 3247–3252. doi:10.1097/00001756-199810050-00021
Keselman, H., Rogan, J., & Feir-Walsh, B. (1977). An evaluation of some nonparametric and parametric tests for location equality. British Journal of Mathematical and Statistical Psychology, 30, 213–221. doi:10.1111/bmsp.1977.30.issue-2
Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research, 68, 350–386. doi:10.3102/00346543068003350
Kirchman, D. L., Sigda, J., Kapuscinski, R., & Mitchell, R. (1982). Statistical analysis of the direct count for enumerating bacteria. Applied and Environmental Microbiology, 44, 376–382.
Markowski, C. A., & Markowski, E. P. (1990). Conditions for the effectiveness of a preliminary test of variance. American Statistics, 44, 322–326.
Meyer, P. F., Oddsson, L. I. E., & De Luca, C. J. (2004). Reduced plantar sensitivity alters postural responses to lateral perturbations of balance. Experimental Brain Research, 157, 526–536. doi:10.1007/s00221-004-1868-3
Mildren, R. L., Straizkalkis, N. D. J., & Bent, L. R. (2016). Foot sole skin vibration perceptual thresholds are elevated in a standing posture compared to sitting. Gait & Posture, 43, 87–92. doi:10.1016/j.gaitpost.2015.10.027
Mukoko, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. Malawi Medical Journal, 24, 69–71.
Nevill, A. (1997). Why the analysis of performance variables recorded on a ratio scale will invariably benefit from a log transformation. Journal of Sports Sciences, 15, 457–458.
Nevill, A., & Atkinson, G. (1997). Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. British Journal of Sports Medicine, 31, 314–318. doi:10.1136/bjsm.31.4.314
Nurse, M., & Nigg, B. (1999). Quantifying a relationship between tactile and vibration sensitivity of the human foot with plantar pressure distributions during gait. Clinical Biomechanics, 14, 667–672. doi:10.1016/S0268-0033(99)00020-0
Perez-Castilla, A., Feriche, B., Janc, S., Padial, P., & Garcia-Ramos, A. (2017). Validity of a linear velocity transducer for testing maximum vertical jumps. Journal of Applied Biomechanics, 33, 388–392. doi:10.1123/jab.2016-0142
Peterka, R. J. (2018). Sensory integration for human balance control. In Handbook of clinical neurology - Balance, Gait, and Falls (1st ed.) (pp. 27–42). Elsevier B.V.
Plautard, M., Guilhem, G., Fohanno, V., Nordez, A., Cornu, C., & Guével, A. (2017). Comparison of two methodological approaches for the mechanical analysis of single-joint isoinertial movement using a customised isokinetic dynamometer. Sports Biomechanics, 17, 1–16. doi:10.1080/14763141.2016.1269187
Pratt, J. (1964). Robustness of some procedures for the two-sample location problem. Journal of the American Statistical Association, 59, 665–680.
Rasmussen, J. (2013). Data transformation, type I error rate and power. British Journal of Mathematical and Statistical Psychology, 42, 203–213. doi:10.1111/jbmsp.1989.42.issue-2
Safrit, M. (1989). An overview of measurement. In M. Safrit & T. Wood (Eds.), Measurement concepts in physical education and exercise science (pp. 3–20). Champaign: Human Kinetics.
Schlée, G., Sterzing, T., & Milani, T. L. (2009). Foot sole skin temperature affects plantar foot sensitivity. Clinical Neurophysiology, 120, 1548–1551. doi:10.1016/j.clinph.2009.06.010
Schmidt, D., Germano, A. M. C., & Milani, T. L. (2017). Effects of active and passive warming of the foot sole on vibration perception thresholds. Clinical Neurophysiology Practice, 2, 38–43. doi:10.1016/j.cnp.2016.12.005
Tømmeren, A., & Serlin, R. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin, 99, 90–99. doi:10.1037/0033-2909.99.1.90
Verillo, R. (1985). Psychophysics of vibrotactile stimulation. The Journal of the Acoustical Society of America, 77, 225–232. doi:10.1121/1.392263
Wilcox, R., Peterson, T. J., & McNitt-gray, J. L. (2013). Data analyses when sample sizes are small: Modern advances for dealing with outliers, skewed distributions, and heteroscedasticity. Journal of Applied Biomechanics, 34, 258–261. doi:10.1123/jab.2017-0269
Wilcox, R. R. (2013). Introduction to robust estimation and hypothesis testing (3rd ed.). San Diego: Academic Press.
Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. British Journal of Mathematical and Statistical Psychology, 57, 173–181. doi:10.1348/000711004X49222
Zimmerman, D. W., & Zumbo, B. D. (1993). Rank transformations and the power of the student T-Test for nonnormal populations with unequal variances. Canadian Journal of Experimental Psychology, 47, 523–539. doi:10.1037/h0078850
