INVITED LECTURE

THE METHODOLOGICAL PRINCIPLES OF USING ANALYSIS OF DIFFERENCES IN SCIENTIFIC RESEARCH IN PHYSIOTHERAPY

Doc. dr. sc. IGOR JELASKA

Universety of Split Faculty of Kinesiology

Abstract

INTRODUCTION: Goal of this paper was to review methodological usage principles of most frequently used analyses of differences in scientific research in physiotherapy.

DISCUSSION: In accordance with the aim, t-test and ANOVA's, both for independent samples and for repeated measures are discussed and highlighted from aspects of practice in physiotherapy. Appropriateness and limitations of methods are given and explained. **Conclusion:** Recommendation for appropriate reporting of t-test results in scientific research is given.

KEY WORDS: t-test, ANOVA, significance, assumptions

METODOLOŠKI PRINCIPI KORIŠTENJA RAZLIKA U ZNANSTVENIM ISTRAŽIVANJIMA U FIZIOTERAPIJI“

Kineziološki fakultet, Sveučilište u Splitu

Sažetak

UVOD: Cilj ovog rada je pregled metodoloških načela o korištenju najčešće korištenih analiza razlika u znanstvenom istraživanju u fizioterapiji. Razrada: U skladu s ciljem, t-test i ANOVA, oboje za nezavisne uzorke i za ponovljene mjere, opisani su i istaknuti iz aspekta prakse fizioterapije. Prikladnost i ograničenja metode su dati i objasniti. Zaključak: U radu je data preporuka za odgovarajuće izvještavanje o rezultatima t-testa u znanstvenom istraživanju.

KLJUČNE RIJEČI: t-test, ANOVA, značajnost, pretpostavke

Introduction

In scientific practice in physiotherapy, often researcher don't just look to describe and analyze one set of data obtained from precisely planned and realized measurements done on one homogenous group of examinees (1-6). Instead, two or three or even more groups of examinees are usually observed and compared (7-12).

Discussion

Usually, goal is to identify and from practical point of view explain „sources of variations“ if statistically significant differences between observed groups (i.e. independent samples) can be identified. Statistical analysis and appropriate hypothesis in above mentioned approaches are different and they depend on number of observed groups (Table 1). Similarly, measurements can be done on the same sample of examinees but in the different time points (i.e. dependent samples) (13-16). It is important to underline that measured variable (usually on interval or ratio scale) is called dependent variable or criterion or response while categorical variable is called independent variable or factor.

TABLE 1. Overview of statistical analysis and appropriate hypothesis in dependence of number of groups and sample type

	Dependent samples	**Independent samples**
2 groups	t test for dependent samples H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	t test for independent samples H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
k groups (k>2)	one way repeated measures ANOVA H_0: $\mu_1 = \mu_2 = ... = \mu_k$ H_1: $\mu_i \neq \mu_j$ for some i, j	one way ANOVA H_0: $\mu_1 = \mu_2 = ... = \mu_k$ H_1: $\mu_i \neq \mu_j$ for some i, j

As can be seen from table 1, T-tests are used to compare the population means from two different groups of data (T-test for independent samples) or population means of two measurements of the same group in two different time points. More precisely, they will point out if population means are significantly different from one another (H1) or if they are statistically/practically the same (H0). Furthermore, if dealing with more than 2 groups (k groups), ANOVA is used to compare the population means from k different groups of data (one way ANOVA or ANOVA for independent samples) or population means of k measurements of the same group in k different time points. As in T-test, ANOVA will identify if population means are significantly different from one another (H1) or if they are statistically/practically the same (H0). Level of statistical significance (type I error) is usually set at $\alpha = 5\%$.

If the means are significantly different or equivalently if hypothesis H0 is rejected, one can say that the independent variable, had a significant effect on the variable being measured, dependent variable (DV). Also, researcher must be aware that calculating effect size parameter is one of the most important outcomes of empirical studies. From practical point of view it is a measure of the practical significance of results. From scientific point of view, effect sizes can be used to determine the sample size for follow-up studies, or examining effects across studies (17-21).

Assumptions for appropriate usage of difference analysis

All statistical tests have assumptions for their appropriate usage. Relatively trivial assumption for independent samples analyses is independence of observations. Reporting it is usually skipped in scientific researches. Second, due to their influence on type I error, data have to be checked for significant outliers. Furthermore, for all above mentioned statistical analyses, assumptions are that dependent variable is approximately normally distributed within each group. That condition can be easily checked by using Kolmogorov Smirnov or Shapiro Wilks test. Both theory and practice agree that the t-test and ANOVA are robust tests with respect to the assumption of normality. More precisely, even distribution of dependent variables may deviate away from normality, it does not have a large influence on Type I error rates. The exception to this is independent samples analyses when the ratio of the size of the groups is greater than approximately 1.5 (22-25). If normality assumption is roughly violated, or data is purely nonparametric (for, example data is on ordinal scale) or ratio of group sizes is sufficiently large, transformations so the data becomes normally distributed can be applied. Also, non-parametric Mann-Whitney U Test can be applied due to the fact it that does not require normality of variables.

For independent samples, very important assumption is homogeneity of variances. Equivalently, variances of groups have to be ”almost“ equal. If variances of observed groups appear to be significantly unequal, this can affect the Type I error rate (26). Usually, the assumption of homogeneity of variance can be tested using Levene's Test of equality of variances. While testing for the homogeneity of variances, test statistics (F value) and a significance level (p-value) is calculated. As usual, if the significance level is greater than 0.05, variances of groups can be treated as equal. However, if $p < 0.05$, hypothesis of equality of group's variances is rejected. If researcher by using the Levene's Test, detect statistically significant differences between variances, correction by not using the pooled estimate for the error term for the t-statistic can be applied (26-29). Similarly, adjustments to the degrees of freedom using the Welch-Satterthwaite method can be done. Levene's test is usually integrated in any serious software package dealing with difference analysis. Similarly, in repeated measures ANOVA, assumption of sphericity have to be checked.

Appropriate Reporting of t- test results

Simple question appears: “How to provide optimal amount of information for readers to fully understand the results when independent t-test was applied?“. From practical point of view, researcher has to report: results of normality testing, results of equality of variances testing, both groups means and standard deviations, the actual t-test result and the direction of the difference. Additionally, researcher might also wish to include the difference between the groups along with the 95% confidence intervals. Hypothetical example is given of appropriate reporting t- test result: „...by using Kolmogorov Smirnov test, it was found that BMI was normally distributed for both experimental and control group ($p>0.20$ for both groups). Furthermore, homogeneity of variance was checked by Levene's Test for equality of variances. Therefore, an independent t-test was applied. BMI value of the control group (23.15 ± 2.52 kg/m2) were significantly higher than the experimental group (21.56 ± 1.79 kg/m2) ($t(64) = 3.012$; $p = 0.007$) with a difference of 1.59 (95% CI: 0.86 to 3.32) kg/m2. Cohen's d was chosen as effect size parameter and it was found to be moderate high (C-d=0.53).

Conclusion

T-test and ANOVA are frequently used in scientific research in physiotherapy. Knowing their appropriateness, limitations and optimal way to report results is an essential to easily “skip unnecessary problems” during process of publishing results of scientific research.

References

1. Vincent-Onabajo GO, Malgwi WS. Attitude of physiotherapy students in Nigeria toward persons with disability. Disability and health journal. 2015;8(1):102-8.
2. Lopez-Sendin N, Alburquerque-Sendin F, Cleland JA, Fernandez-de-las-Penas C. Effects of physical therapy on pain and mood in patients with terminal cancer: a pilot randomized clinical trial. Journal of alternative and complementary medicine. 2012;18(5):480-6.
3. Odebiyi OD, Aiyejusunle CB, Alonge EO, Olalekan TA. Comparison of the knowledge and perception of physiotherapy by medical students of the institutions with and without physiotherapy training programme. The Nigerian postgraduate medical journal. 2011;18(4):282-7.
4. Mangwani J, Giles C, Mullins M, Salih T, Natali C. Obesity and recovery from low back pain: a prospective study to investigate the effect of body mass index on recovery from low back pain. Annals of the Royal College of Surgeons of England. 2010;92(1):23-6.
5. Zoremba M, Dette F, Gerlach L, Wolf U, Wulf H. Short-term respiratory physical therapy treatment in the PACU and influence on postoperative lung function in obese adults. Obesity surgery. 2009;19(10):1346-54.
6. Pomeroy VM, Warren CM, Honeycombe C, Briggs RS, Wilkinson DG, Pickering RM, et al. Mobility and dementia: is physiotherapy treatment during respite care effective? International journal of geriatric psychiatry. 1999;14(5):389-97.
7. Schewitz J, Roos R, van Aswegen H, Manda S. The effect of two passive head-down tilt positions on diaphragm excursion in healthy adults: A preliminary study. Physiotherapy theory and practice. 2016;32(3):223-31.
8. Martinez-Gramage J, Merino-Ramirez MA, Amer-Cuenca JJ, Lison JF. Effect of Kinesio Taping on gastrocnemius activity and ankle range of movement during gait in healthy adults: A randomized controlled trial. Physical therapy in sport : official journal of the Association of Chartered Physiotherapists in Sports Medicine. 2016;18:56-61.
9. Samsson KS, Larsson ME. Physiotherapy triage assessment of patients referred for orthopaedic consultation - Long-term follow-up of health-related quality of life, pain-related disability and sick leave. Manual therapy. 2015;20(1):38-45.
10. Manca A, Limonta E, Pilurzi G, Ginatempo F, De Natale ER, Mercante B, et al. Ultrasound and laser as stand-alone therapies for myofascial trigger points: a randomized, double-blind, placebo-controlled study. Physiotherapy research international : the journal for researchers and clinicians in physical therapy. 2014;19(3):166-75.
11. Heredia-Rizo AM, Rodriguez-Blanco C, Oliva-Pascual-Vaca A, Torres-Lagares D, Albornoz-Cabello M, Pina-Pozo F, et al. Masticatory mechanosensitivity, mouth opening and impact of headache in subjects with a history of orthodontics use: a cross-sectional study. European journal of physical and rehabilitation medicine. 2014;50(4):411-8.
12. Guirro RR, Guirro EC, Alves de Sousa NT. Lack of maintenance of shortwave diathermy equipment has a negative impact on power output. Journal of physical therapy science. 2014;26(4):557-62.
13. Fuentes CJ, Armijo-Olivo S, Magee DJ, Gross DP. A preliminary investigation into the effects of active interferential current therapy and placebo on pressure pain sensitivity: a random crossover placebo controlled study. Physiotherapy. 2011;97(4):291-301.
14. Khamwong P, Nosaka K, Pirunsan U, Paungmali A. Reliability of muscle function and sensory perception measurements of the wrist extensors. Physiotherapy theory and practice. 2010;26(6):408-15.
15. Fifoot S, Wilson C, MacDonald J, Watter P. Respiratory exacerbations in children with cystic fibrosis: physiotherapy treatment outcomes. Physiotherapy theory and practice. 2005;21(2):103-11.
16. Almeida CC, Ribeiro JD, Almeida-Junior AA, Zeferino AM. Effect of expiratory flow increase technique on pulmonary function of infants on mechanical ventilation. Physiotherapy research international : the journal for researchers and clinicians in physical therapy. 2005;10(4):213-21.
17. Collins M, Carey TA. Identification of Real and Artifactual Moderators of Effect Size in Meta-Analysis. Multivariate behavioral research. 2015;50(1):109-25.
18. Shieh G. Confidence intervals and sample size calculations for the standardized mean difference effect size between two normal populations under heteroscedasticity. Behavior research methods. 2013;45(4):955-67.
19. So HC, Sham PC. Effect size measures in genetic association studies and age-conditional risk prediction. Human heredity. 2010;70(3):205-18.
20. Velicer WF, Cumming G, Fava JL, Rossi JS, Prochaska JO, Johnson J. Theory Testing Using Quantitative Predictions of Effect Size. Applied psychology = Psychologie appliquee. 2008;57(4):589-608.
21. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in psychology. 2013;4:863.
22. Stokes L. Sample size calculation for a hypothesis test. Jama. 2014;312(2):180-1.
23. McMillan GP, Hanson TE. Sample size requirements for establishing clinical test-retest standards. Ear and hearing. 2014;35(2):283-6.
24. Qiu P, Moeschberger ML, Cooke GE, Goldschmidt-Clermont PJ. Sample size to test for interaction between a specific exposure and a second risk factor in a pair-matched case-control study. Statistics in medicine. 2000;19(7):923-35.
25. Patrick DM. Sample size and diagnostic test evaluations. Sexually transmitted infections. 1998;74(1):78.
26. Edgington ES. The Assumption of Homogeneity of Variance for the T Test and Nonparametric Tests. The Journal of psychology. 1965;59:177-9.
27. Zimmerman DW. Consequences of choosing samples in hypothesis testing to ensure homogeneity of variance. The British journal of mathematical and statistical psychology. 2014;67(1):1-29.
28. Fang X, Li J, Wong WK, Fu B. Detecting the violation of variance homogeneity in mixed models. Statistical methods in medical research. 2014.
29. O'Neill ME, Mathews KL. Levene tests of homogeneity of variance for general block and treatment designs. Biometrics. 2002;58(1):216-24.