



# Comparison of p-value results between one versus two sample t testing: A case study

John Hart<sup>1,2</sup>

Correspondence: [jhartdc@yahoo.com](mailto:jhartdc@yahoo.com)



CrossMark

← Click for updates

<sup>1</sup>Greenville, South Carolina USA.

<sup>2</sup>Adjunct Faculty, Purdue University Global, USA.

## Abstract

The one-sample t test compares a sample to a known average. The standard deviation (SD) is known for the sample but not for the known average. This study compares p-values from one-sample versus two-sample t testing where SD is also known with the known average to see if the lack of information on SD makes a difference in p-values calculated with versus without the SD. Having confidence in a t test p-value is to have confidence in whether the difference in means happened by chance or not.

**Keywords:** Heart rate variability, biostatistics, t test

## Background

The one-sample t test is used to compare a sample mean to a known, usually published mean. The comparison does not include additional information for the known published mean like its sample size or standard deviation (SD) as the two-sample t test includes. This study sought to answer the author's question of whether it matters that SD is omitted in the one-sample t test. The author did not find any previous research in the literature on this question. The question was addressed by comparing p-value results where the SD *was* versus *was not* included.

Heart rate variability (HRV) is an evidence-based method of assessing the health and adaptability of a person's autonomic nervous system [1]. This is important because autonomic function plays a key role in promoting a long and healthy life [1]. HRV typically declines with age [2; Figure 1].

The author uses HRV to assess autonomic health in his chiropractic patients and in himself. He recently turned age 64 and was curious as to how a sample of his own HRV measurements compared to normative published data for his age group and younger age groups. The author was also curious as to whether the comparisons would be the same in *one* versus *two-sample* t testing when sample size and SD were available with the published norms. Since the one-sample t test requires only one parameter, the mean, comparison of one versus two-sample t tests can be made when the additional information is provided.

## Methods

The normative data used in this study included summary sta-

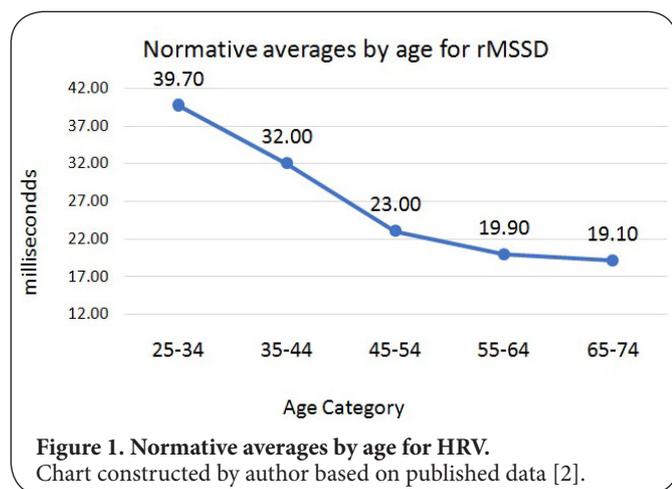
tistics of sample size, mean, and SD [2]. The study was declared Exempt by the Institutional Review Board at Purdue University Global. HRV was measured and recorded with the App, *Heart Rate Variability Logger* [3]. The App uses an earlobe sensor, the *Kyoto* ear clip which uses photoplethysmography (PPG) technology – lights that sense different tints of red with each heartbeat. The sensor then sends a Bluetooth signal to the App on a smartphone. The set-up has good agreement with standard ECG technology [4-5].

The HRV metric used was the standard root mean square of successive differences of time between heartbeats (rMSSD). The terms HRV and rMSSD are used interchangeably in this paper. For rMSSD a larger number is considered neurologically healthier than a smaller number.

The author's self-measurements were obtained every day except for three days from the day of his 64<sup>th</sup> birthday in November to the point at which analysis was completed (on 12-26-20) for a total of 47 measurements. The 1-minute recordings were obtained early in the morning in the supine position before getting out of bed. The author's data is referred to as *sample data*. Protocol for the normative published (comparative) data was similar: 5-10 minutes of pre-test rest, measured in the supine position, and ECG used for a 5-minute recording [2]. One-minute recordings for rMSSD show good agreement with 5-minute recordings [6].

## Analysis

Data were analyzed in Stata 12.1 (StataCorp, College Station,



TX) using one and two-sample t testing. The one-sample t test was performed comparing the sample data (author’s) with the published mean only (excluding the n and SD that were also available but saved for the two-sample t test). Also in the analysis:

- A two tailed p-value < the conventional alpha level of 0.05 was considered statistically significant.
- The unequal variances option for two-sample t testing was considered appropriate since the p-value for the variance ratio test (also performed in Stata) for each comparison (between the author’s HRV and published age group norms) was <0.05, indicating unequal variances.
- The t test assumes data exhibits a normal or near-normal distribution; or has a sample size of at least 30 [7]. All variables in the study had at least 30 observations.

## Results

Descriptive statistics are provided in **Table 1** and inferential statistics are in **Table 2**.

There were five age groups where p-values from the normative published HRV data were compared to the author’s HRV sample data using one and two-sample t testing (**Tables 1-2**). Four of these five (80%) comparisons showed p-value agreement between one and two-sample t testing (**Table 2**). The disagreement was observed for age group 35-44. This group had a relatively high SD (**Table 1**) and was calculated as an outlier and was the only outlier among the five SDs (from the five age groups) using the inter-quartile method of outlier detection. The upper fence in this analysis=33.4 while the SD for that age group=36.5 (a high outlier).

Out of curiosity, in a secondary analysis, a lower SD was used for the 35-44 age group based on the mean of the other four SDs. This mean SD was 13.2 which produced a statistically nonsignificant p-value in the two-sample t test (p=0.063) which then agreed with the two-sample t test for this age group.

**Table 1. Summary statistics by age category.**

Statistic	Descriptive Statistics by Age Group					
	64 (author)	25-34	35-44	45-54	55-64	65-74
n	47	330	292	235	183	84
mean	35.7	39.7	32.0	23.0	19.9	19.1
SD	7.9	19.9	36.5	10.9	11.1	10.7

n for author (age 64 column) is number of measurements (47 measurements over a 47 day period). n for the other age groups is number of persons in the published normative study. Mean is average HRV (rMSSD in milliseconds). SD is standard deviation. n and SD for normative data omitted for one-sample t test (in **Table 2**).

**Table 2. Inferential statistics (t test p-values) by age categories compared to the author’s sample.**

Test type	25-34	35-44	45-54	55-64	65-74
Two-sample, unequal variance	0.0129	0.1284	<0.0001	<0.0001	<0.0001
One-sample	0.0011	0.0021	<0.0001	<0.0001	<0.0001

The t test p-values reflect the difference in means between the author’s HRV (age 64 column) and the published normative data age groups. The bold highlighted p-value (in the age category of 35-44) is the one p-value in the study that showed disagreement between one and two-sample t testing.

## Discussion

There was good agreement, but not 100%, of p-values between one and two-sample t testing in this case study when comparing the author’s heart rate variability to published normative data.

The one comparative age group that showed the disagreement of p-values between the one and two-sample t tests (ages 35-44) also showed the largest standard deviation (and was an outlier) among SDs for the five comparative age groups. This could be the reason for the p-value disagreement in this age group. The other age groups had SDs that were closer to the SD in the author’s HRV data, 7.9 (as noted in **Table 1**).

There may be a validity question about comparing an individual’s sample (the author’s) to a group statistic (the means by age group in the published normative data). It should be noted though, that this was the author’s research question – as to how his HRV compared to others’ HRV at the group level.

The accuracy of the one-sample t test may be considered satisfactory. However, the resulting p-value should be viewed with some caution due to omission of sample size and standard deviation in the calculation. When these *are* provided, they should be included in the calculation, thereby changing the analysis into a two-sample test. A more reliable p-value allows for a more accurate and more confident estimation of whether the difference in means happened by chance alone.

## Competing interests

The author declares that he has no competing interests.

### Acknowledgment

After observing the p-value disagreement in the age group mentioned above, the author sent his results to one of his former statistics professors, Dr. James Higgins, Professor Emeritus, Kansas State University. The author is grateful to Professor Higgins for noting the relatively high standard deviation in the age group as the likely reason for the p-value disagreement.

### Publication history

Editor: Catherine Ortega, University of Texas Health Science Center, USA.

Received: 26-April-2021 Final Revised: 18-June-2021

Accepted: 23-June-2021 Published: 26-June-2021

### References

1. Usman Z, et al. Relation of high heart rate variability to healthy longevity. *The American Journal of Cardiology* 2010; 105(8):1181-1185.
2. Voss A, Schroeder R, Heitmann A, Peters A, Perz S. Short-term heart rate variability—influence of gender and age in healthy subjects. *PLoS One* 2015; 10(3): e0118308. doi:10.137.
3. Altini M. Heart rate variability Logger – app details. December 12, 2013. Accessed 1-1-19 at: <https://www.marcoaltini.com/blog/heart-rate-variability-logger-app-details>
4. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart rate variability recording with smart phone photoplethysmographic Polar H7 chest strap and electrocardiogram methods. *International Journal of Sports Physiology and Performance* 2017; 12(10)1-17.
5. Vescio B, Salsone M, Gambardella A, Quattrone A. Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings. *Sensors* 2018; 18: 844; doi:10.3390/s18030844.
6. Munoz ML, van Roon A, Riese H, Thio C, Oostenbroek E, Westrik I, et al. (2015) Validity of (Ultra-)Short Recordings for Heart Rate Variability Measurements. *PLoS One* 10(9): e0138921. doi:10.1371.
7. Devore J, Peck R. *Statistics: the exploration and analysis of data*. 4th ed. Pacific Grove, CA: Duxbury; 2001:410.

#### Citation:

Hart J. **Comparison of p-value results between one versus two sample t testing: A case study.** *J Med Stat Inform.* 2021; 9:1.

<http://dx.doi.org/10.7243/2053-7662-9-1>