XVII Congreso Galego de Estatística e Investigación de Operacións
Ourense, 23-25 de outubro de 2025

Normality: Its importance, Misconceptions, and Best Practices

Jorge Santos [1], Marília Pires [2] and Russell Alpizar Jara [3]

[1] Centro de Investigação em Matemática e Aplicações (CIMA) from Universidade de Évora
[2] Centro de Investigação em Matemática e Aplicações (CIMA) and Departamento de Matemática from Universidade de Évora
[3] Centro de Investigação em Matemática e Aplicações (CIMA) and Departamento de Matemática from Universidade de Évora

ABSTRACT

The t-test remains one of the most widely used tools in applied statistics, yet its assumptions are often misinterpreted. A recurring issue arises from oversimplified rules of thumb regarding normality. For the independent samples t-test, normality should be assessed separately for each group, which can be easily done with standard software. In contrast, for the paired samples t-test, the normality of the original variables is irrelevant; what truly matters is the distribution of the paired differences. Misunderstanding this distinction is common in portuguese textbooks and portuguese teaching materials, leading to systematic errors in applied research. In this work, we revisit these assumptions, clarify misconceptions, and discuss the role of the Central Limit Theorem in mitigating departures from normality in large samples. We also outline future directions, including simulations on small artificial datasets, to illustrate the practical impact of non-normality. Our goal is to provide a clearer framework of best practices, helping researchers and students alike avoid pitfalls and apply the t-test more reliably.

Keywords: Include here your keywords: normality, t-test, independence, correlation, CLT, BiVariate Normal

1. INTRODUCTION

In a textbook that we have been using for a couple of years one can read: “Considere-se agora o caso em que as duas amostras formam um par de observações ($X1i$; $X2i$), i=1,…,n, ou seja, trata-se de uma amostra emparelhada. Os pares de observações são independentes e retirados de populações **Normais**, com médias $\mu1$ e $\mu2$ e desvios padrão $\sigma1$ e $\sigma2$, respectivamente.” This is wrong, because normality is important at the series of the differences between $X1i$ and $X2i$, Di=$X1i$-$X2i$.

Some colleagues think that this is not a serious problema, since the difference of Normal Distributions is also Normal.

This reasoning is only valid in the theoretical field of Normal independent distributions or in the case of a bivariate normal distribution.

When it comes to real data it may happen that normality tests like Shapiro Wilk and Kolmogorov-Smirnov can’t reject the null hypothesis of normality of $X1$ and $X2$, but its difference D=$X1$-$X2$ can lead to a clear departure from Normality. We present an R Script as an example of this problem.

Some statistical software packages deal with it the proper way, but some very popular don’t, so we present some outputs for the two cases.

Specifically in SPSS one has to build a new variable D=$X1$-$X2$ and than analyse its normality by the usual nonparametric tests.

In the case of JASP it calculates the new variable D=$X1$-$X2$, and applies the Shapiro Wilk test to the differences.

2. SPSS RESULTS FOR AN ARTIFICIAL DATA SET

The independent samples t-test in SPSS has no problems and the outputs are clear as we can see in figure 1 where we test the two variables Weight Before and Weight After on a total of 17 individuals 8 females and 9 males.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Weigth Before	Equal variances assumed	0.007	0.934	26.207	15	<0.001	<0.001	20.19444	0.77057	18.55202	21.83687
	Equal variances not assumed			26.216	14.786	<0.001	<0.001	20.19444	0.77030	18.55052	21.83837
Weigth After	Equal variances assumed	0.093	0.764	25.056	15	<0.001	<0.001	20.30556	0.81039	18.57824	22.03287
	Equal variances not assumed			25.207	14.987	<0.001	<0.001	20.30556	0.80556	18.58842	22.02269

Figure 1: Results of the independent samples t-test in SPSS

In figure 2 we ignore the gender factor and we want to check if there was changes in the body weight after treatmant. We present the results of the dependent samples t-test in SPSS and it created a variable Pair 1, that is not tested automatically for normality. We have to create that variable by ourselves like we show in figure 3 an we called it Diff.

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	Weigth Before - Weigth After	9.94118	0.24254	0.05882	9.81648	10.06588	169.000	16	<0.001	<0.001

Figure 2: Results of the dependent samples t-test in SPSS

It would be good if SPSS automatically applied the normality tests to Pair 1. It would avoid people checking manually the normality assumptions.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weigth Before	0.266	17	0.002	0.792	17	0.002
Weigth After	0.275	17	0.001	0.769	17	<0.001
Diff	0.537	17	<0.001	0.262	17	<0.001

a. Lilliefors Significance Correction

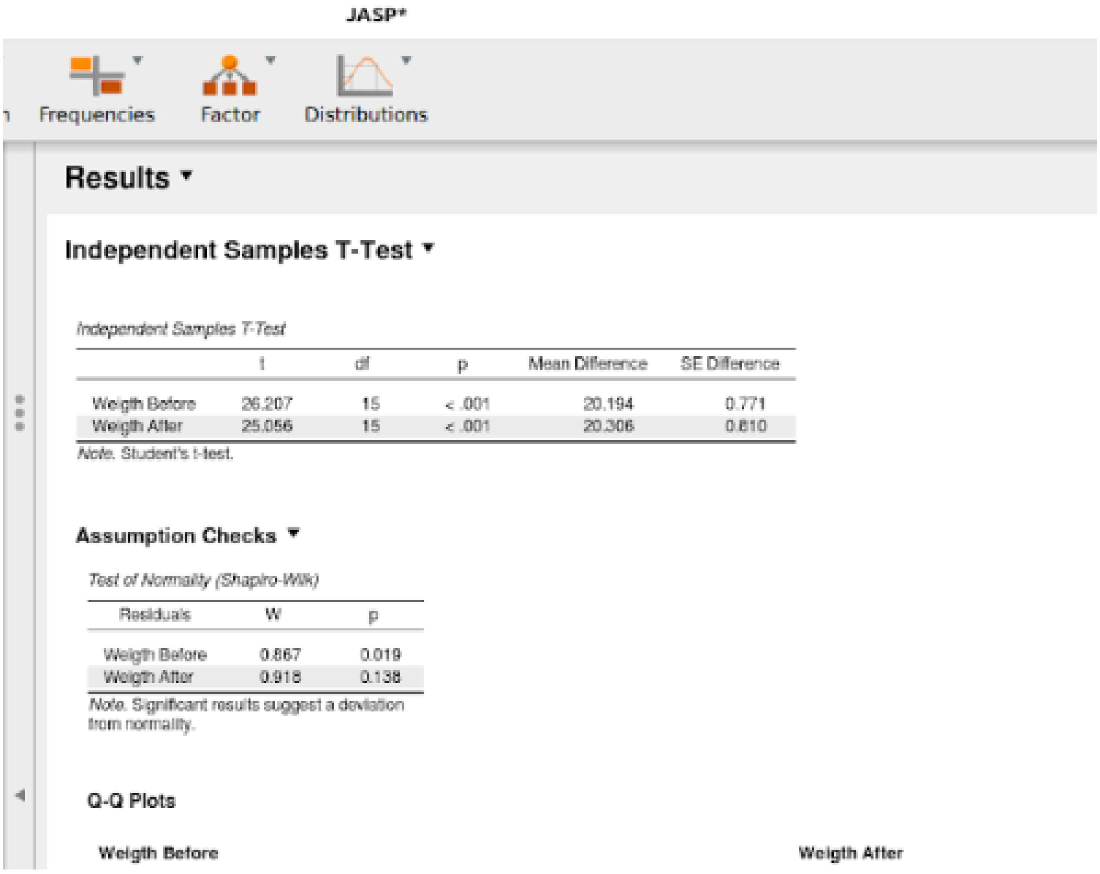
Figure 3: Results of the Normality tests in SPSS

3. JASP RESULTS FOR AN ARTIFICIAL DATA SET

In JASP normality is correctly assessed separately for each group, as shown in figure 4. An advantage over SPSS.

Results from the t test are similar, and there are clear difference between the weights of the two genders in both situations: Before and After treatment.

For the case of paired samples JASP just checks the normality assumption for the difference, and it is clearly better than in the SPSS counterpart.


JASP*

Frequencies Factor Distributions

Results

Independent Samples T-Test

Independent Samples T-Test

	t	df	p	Mean Difference	SE Difference
Weigth Before	26.207	15	< .001	20.194	0.771
Weigth After	25.056	15	< .001	20.306	0.810

Note. Student's t-test.

Assumption Checks

Test of Normality (Shapiro-Wilk)

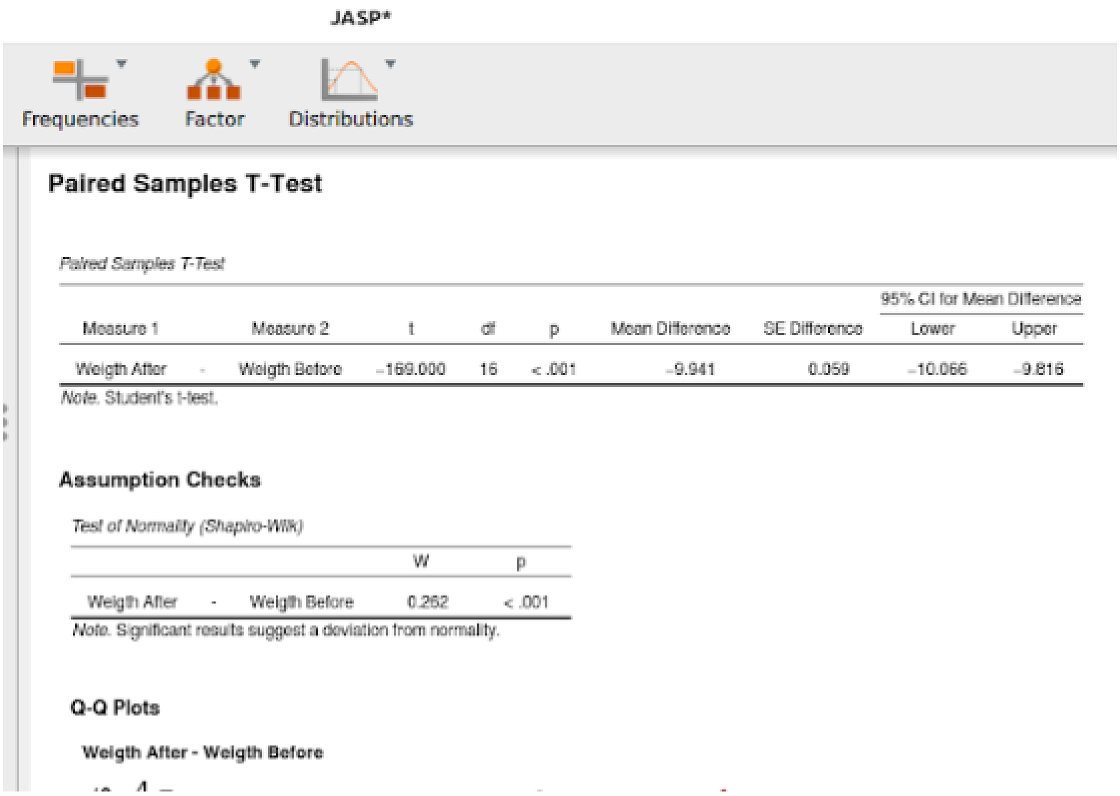
Residuals	W	p
Weigth Before	0.867	0.019
Weigth After	0.918	0.138

Note. Significant results suggest a deviation from normality.

Q-Q Plots

Weigth Before Weigth After

Figure 4: Results of the independent samples t-test in JASP


JASP*

Frequencies Factor Distributions

Paired Samples T-Test

Paired Samples T-Test

								95% CI for Mean Difference	
Measure 1		Measure 2	t	df	p	Mean Difference	SE Difference	Lower	Upper
Weigth After	-	Weigth Before	-169.000	16	< .001	-9.941	0.059	-10.066	-9.816

Note. Student's t-test.

Assumption Checks

Test of Normality (Shapiro-Wilk)

			W	p
Weigth After	-	Weigth Before	0.262	< .001

Note. Significant results suggest a deviation from normality.

Q-Q Plots

Weigth After - Weigth Before

Figure 5: Results of the paired samples t-test in JASP

4. AN ILLUSTRATION WITH A R SCRIPT

We prepared a case where the two variables X and Y whose normality can't be rejected, but their difference is clearly non-normal. This illustrates the need to test the difference, not the original variables.

The script was tested on Posit Cloud

```
set.seed(137);norm5=rnorm(22,5);norm6=rnorm(22,6) # Creates two normals with 22 observations
norm5D=sort(norm5,TRUE);norm6D=sort(norm6,TRUE)
norm5D24=c(norm5D,7,8);norm6D24=c(norm6D,7,8) # Creates two vars with 24 observations
shapiro.test(norm6D24)#p-value = 0.4477 Can't reject Normality
shapiro.test(norm5D24)#p-value = 0.7038  Can't reject Normality
shapiro.test(norm6D24-norm5D24)#p-value = 2.005e-05 Must reject Normality
```

5. CONCLUSIONS

Normality of differences is an important assumption for the paired t-test.

For small samples, if the normality assumption is violated, it can lead to incorrect conclusions.

For larger samples, the t-test can still be somewhat robust due to the Central Limit Theorem, but it's still good practice to check for normality.

If the normality assumption is in doubt, especially with small sample sizes, consider using a non-parametric test like the Wilcoxon signed-rank test, or a suitable variable transformation like the Box-Cox transformation.

For the paired samples t-test, the normality of the original variables is irrelevant; what truly matters is the distribution of the paired differences

. ACKNOWLEDGMENTS

This research has been partially supported by the Department of Mathematics, School of Sciences and Technology and by the Centro de Investigação em Matemática e Aplicações (CIMA), through the Project UIDB/04674/2020 of FCT-Fundação para a Ciência e a Tecnologia, Portugal.

REFERENCES

Afonso, A., Nunes, C. Estatística e Probabilidades: Aplicações e soluções em SPSS, Escolar Editora, 2011.

Agresti, A. An Introduction to Categorical Data Analysis, John Wiley and Sons, 2002.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, 211-252.

Rosner, B.A. Fundamentals of Biostatistics, Duxbury Press,1995.

https://www.jmp.com/en_nl/statistics-knowledge-portal/t-test/paired-t-test.html