

The Significance Testing of a Skewed Color-Imaging Data Set

Mohsen Mohammadzadeh, Maryam Mohammadzadeh* and Stephen Westland*

Department of Statistics, Tarbiat Modares University; Tehran / Iran

*School of Design; University of Leeds, Leeds / UK

Abstract

In color-imaging science researchers frequently conduct experiments that are evaluated by measuring a number of ΔE values. Often the effect of various experimental parameters is evaluated by comparing the means of two or more sets of ΔE values. Frequently, a t-test is used but correct application of the t-test requires that a number of assumptions (such as the data being normally distributed or the sample size is big enough) are satisfied. In this article, an improved T statistic has been used in a case study where the data are skewed and do not have a normal distribution while at the same time the sample size is small. The importance of this assumption comes to mind when we realize that most of the derived data from the color-imaging science are not applicable for t-test because of not following a normal distribution.

Introduction

There are many situations in color-imaging science where researchers conduct experiments that are evaluated by measuring a number of ΔE values. The effect of various experimental parameters is frequently evaluated by comparing the means of two or more sets of ΔE values. Frequently, a t-test is used but correct application of the t-test requires adherence of a number of assumptions (such as the data being normally distributed or the sample size being large). A concern about whether these assumptions are valid has led some researchers to use non-parametric techniques such as the Wilcoxon matched-pairs signed-rank test [1]. However, it needs to be understood that all statistical tests make some assumptions and require certain conditions be satisfied and the correct choice can be difficult to make. In this paper we consider how to correctly address this common problem for color scientists and make recommendations to allow use of the correct statistical method. Specifically, we discuss three modified versions of the T statistic by Johnson (1978), Sutton (1993) and Chen (1995) when the distribution of the data is asymmetric. These modifications are for one-sample t-tests. However, the color-imaging situation referred to above consists of two samples. Two-sample data can be analyzed using a one-sample t-test on the differences between the two sets; these differences can be computed when the two samples are of equal size and also are paired. It is important to note that this work therefore applies for one-sample t-tests or for situation where the data for a two-sample t-test are paired (this is, however, frequently the case for the type of color-imaging problem outlined earlier). In the two-sample test we check whether the two means are equal; in the one-sample test we check whether the mean of the differences is equal to zero.

These three different statistics (Johnson, Sutton and Chen) are described and the application of Chen's modification is demonstrated by a case study (based on a previously published experiment by one of the authors). Finally, example code is produced in MATLAB to allow any user to use this technique. The code is also available on-line for download.

Underlying Theory

It should be noted that the t-test shows more sensitivity to skewed distributions than to heavy-tailed distributions [2]. However, many efforts have been made to make changes to an asymmetrical distribution so that the t-test can be applied. For example, the T statistic can be modified by the Cornish-Fisher expansion to remove the skewness of the data [3] for one-sample t-tests. For very skewed distributions [4] a simple monotonic transformation method has been suggested by Hall [5]. The t-test can also lose accuracy when used for small samples. Another common problem is when we assume that the variances of two target populations are equal; the Behrens-Fisher problem [6]. This has a practical solution called the Welch's approximate test which is still not able to overcome the problem when the population is not normal (it cannot control the type-I error) [7]. We can attempt to overcome the problem of non-normality through non-parametric tests (for example, the Wilcoxon-Mann-Whitney test). Cases of non-normality become very important in the context of this paper because the practical data which are studied in color science are often not normally distributed.

The Student's t-test [8] uses the statistic $T = \sqrt{n}(\bar{X} - \mu_0)/S$ used to test the hypothesis $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ where μ_0 is zero in our case, μ is the mean of a normal population, \bar{X} , S and n are denoted for sample mean, standard deviation and size, respectively. (In our case, \bar{X} is considered to be the sample mean of the paired differences between the two data sets.) When the assumption of an underlying normal distribution is violated, this test is only robust provided that the population variance is finite and the sample size is sufficiently large (see e.g. [9]). Specifically, for the sample size n we should have:

$$n \geq \frac{z_{0.025}^2 s^2}{d^2}, \quad (1)$$

Where $z_{0.025}^2 = 1.96$ and the estimate of the mean within d (for e.g. 0.5) of the true mean with 95% confidence.

(Note that even alternative non-parametric tests have a requirement for symmetry and a large sample size does not rectify matters if that assumption is violated in a significant way. In fact, for asymmetrical distributions, both the sign test and the signed-

rank test become less accurate with an increase in sample size [10]. Thus the alternatives to the t-test with which most users of statistical procedures are best acquainted are of no real help when the data are from a skewed distribution.)

Early investigations of the effects of non-normality on the behavior of the Student's t-statistic indicated that the skewness of the data has a greater effect on the distribution of T when compared to kurtosis. Positive skewness in the distribution from which the observations arise results in the sampling distribution of T being negatively skewed. This leads to an inflated significance level for lower-tailed tests (i.e., reported p-values will be smaller than they should be) and a loss of power for upper-tailed tests (i.e., reported p-values will be too large) when the parent distribution of the data has positive skewness. Many users of statistical methods will routinely use the t-test without questioning the assumption of normality, when confronted with data from a skewed distribution, and others will use it hoping that the sample size is large enough so that its robustness will ensure that the result is reasonable. If the sample size is small and the parent distribution is as asymmetrical as an exponential distribution, then Johnson (1978) modified the Student's t-statistic T and proposed the t-statistic T_1 , thus

$$T_1 = T + \frac{\hat{\mu}_3}{6S^3\sqrt{n}} + \frac{\hat{\mu}_3}{3S^5}\sqrt{n}(\bar{X} - \mu_0)^2,$$

where $\hat{\mu}_3 = \sum(X_i - \bar{X})^3/n$ $\hat{\mu}_3 = \sum(X_i - \bar{X})^3/n$ is the sample third central moment and T_1 has approximately a Student t-distribution with $n-1$ degrees of freedom [3].

Sutton confirmed that the Johnson's procedure should be used instead of the t-test for one-tailed tests when the parent distribution is asymmetrical, because it reduces the probability of type I error in cases where the t-test has an inflated type I error rate and is more powerful in other situations [11]. However, Sutton pointed out that if the skewness is severe and the sample size is small, then Johnson's test can also be quite inaccurate. In all the cases discussed in Sutton's paper, the upper-tailed Johnson tests have lower type I error rates than specified, resulting in a loss of power. To improve the power, Sutton proposed a composite decision rule with a critical region consisting of the union of critical regions from Johnson's test and two bootstrap tests based on Johnson's modified T statistic. Obviously, the composite upper-tailed test will have a higher apparent power than Johnson's test and two other component tests, but this increased power is meaningful only if its type I error rate is less than or equal to the nominal significance level α .

In a simulation study Chen showed that, although Johnson's test generally appears to become more conservative with increasing skewness, for some strongly positively skewed distributions and small sample sizes, the type I error rate of Johnson's test can still be inflated [12]. Therefore, the composite test should be used with caution. Chen proposed a new test procedure with T statistic T_2 as follows

$$T_2 = T + \frac{\hat{\beta}}{6\sqrt{n}}\{1 + 2T^2\} + \frac{\hat{\beta}^2}{9n}\{T + 2T^3\}, \quad (2)$$

where $\hat{\beta} = n \sum(X_i - \bar{X})^3 / (n-1)(n-2)S^3$ is an unbiased estimate of the skewness coefficient and the decision rule for

$H_0: \mu = \mu_0$ testing against $H_1: \mu > \mu_0$ is $t_2 > z_{\alpha}$, where z_{α} is the $100(1 - \alpha)$ percentile of the standard normal distribution and α is the significant level. Chen showed that for a variety of positively skewed distributions with small sample sizes, this test is more accurate and more powerful than both Johnson's modified t-test and Sutton's composite test. So as a conclusion, we can say that Chen's test can be used as the latest improved version of t-test when the data are not normally distributed and are skewed to the right.

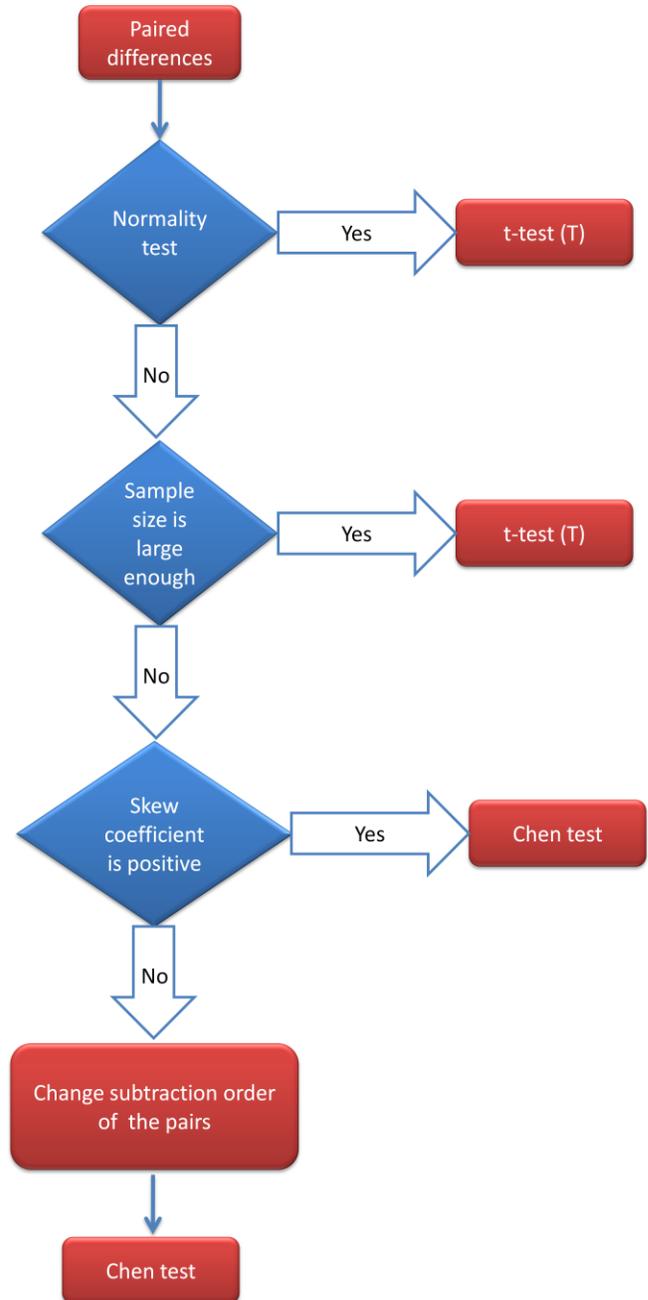


Figure 1: General statistical methodology flow chart for significance testing of a skewed color-imaging data set.

Figure 1 illustrates the algorithm that a color scientist needs to follow when dealing with possibly skewed non-normal data. The first step is to test the normality of the paired differences data (use of Q-Q plot or Shapiro-Wilks normality test is suggested). Next, the sample size should be calculated using formulae (1); by that, there would be only two methods for testing the hypothesis a) t-test if the sample size is big enough or b) Chen test with test statistic T_2 given in (2) if the sample size is small.

It should be noted that the Chen test is used only if the data are positively skewed; however, if the data are negatively skewed, a change of direction of skewness is trivial by changing the subtraction order of the paired data.

Case Study and Results

In this paper, data were obtained from a previous study [1] that was concerned with camera characterization using multispectral techniques. In that study, various parameters were tested and the performance evaluated by considering the ΔE values for a set of data (in this case, 40 samples randomly drawn from the NCS set). It is important to note that, in this case, the data are paired to allow for the possibility of one-sample t-tests (and alternatives). That is, the same 40 samples were used to evaluate all parameters. Two particular sets of parameters were selected (the actual details are not important) to yield two sets of forty ΔE values (denoted here as $\Delta E_{1,i}$ and $\Delta E_{2,i}$ for $i = 1, \dots, 40$) and we would like to explore whether their means are significantly different. In this study we would define a new variable which is the paired differences ($\Delta E_{1,i} - \Delta E_{2,i}$) thus $\bar{X} = \sum_i (\Delta E_{1,i} - \Delta E_{2,i}) / n$ which in this case n is equal to 40.

The two sets of data being considered had means of 4.38 and 3.76 and standard deviations of 3.85 and 2.46, respectively as shown in Table 1. Figure 2 shows a histogram of the paired differences between the two sets of data. It is evident that the differences are asymmetric and not normally distributed. Furthermore, they are positively skewed; the calculated skewness coefficient is 3.63 which indicate a right-skewed distribution. Figure 3 shows the normal Q-Q plot. The Q-Q plot is an illustration which acts as a test for normality [13]. For example, if all data points fall on the straight line indicated on the normal Q-Q plot, the data has a normal distribution whilst if the data does not fall on the indicated line then it is not normally distributed. This is just an illustrative way of testing the normality of a data set. On the other hand, we can also use the Shapiro-Wilks normality test that the differences X_i comes from a distribution in the normal family [14]. Note, however, that because code has been generated in MATLAB the Lilliefors test for normality has been used instead of the more popular Shapiro-Wilks test because the former is available in the statistics toolbox of MATLAB (and is also available in the Student Edition of MATLAB).

The Q-Q plot reveals that the data are not normally distributed, (since the points do not fall on the straight line shown). This is confirmed by the Lilliefors test which has a p-value less than 0.05. We now have to test whether a Student t-test is still possible even though the data are not normally distributed by virtue of a large sample size. Using Equation (1) we get a required sample size estimate of $n \geq 92.24 \approx 93$ which means that the sample size ($n=40$) of our data is not sufficiently large to allow application of the t-test. It is therefore necessary to use the Chen test instead of the t-test.

Table 2 shows the result of a one-sample t-test on the paired differences X_i for comparison (even though in this case it is invalid). The p-value of the t-test is higher than the significance level (of 0.05) which leads to the acceptance of the null hypothesis $H_0: \mu = 0$, where μ is the mean of the paired differences, that is the means of $\Delta E_{1,i}$ and $\Delta E_{2,i}$ are equal. While the p-value reported from Chen's t test is less than 0.05 marking the rejection of the null hypothesis. Thus the means are not the same; the mean of $\Delta E_{1,i}$ is bigger than the mean of $\Delta E_{2,i}$. In this case, it can be seen that the Student t-test gave an erroneous result.

Table 1: Table of descriptive statistics of the data

Data	Mean	Std
ΔE_1	4.39	3.85
ΔE_2	3.76	2.46
Paired differences	0.62	2.45

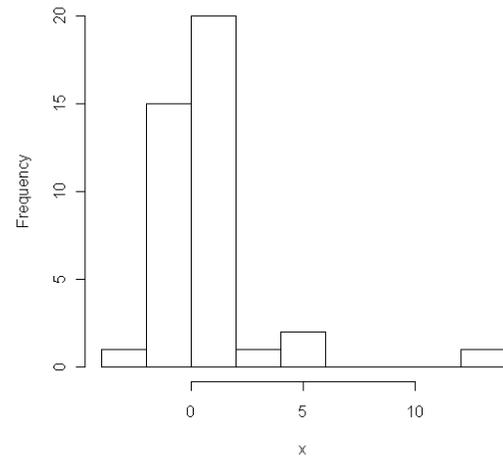


Figure 2: Histogram of paired differences

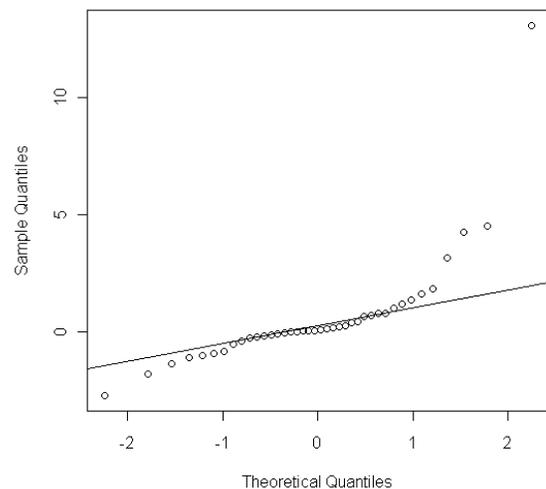


Figure 3: Normal Q-Q plot of the paired differences

In this case study, Chen's test is valid as the data are skewed and are not normally distributed.

Table 2: Table of student and Chen's t-statistics along with their p-values.

Test	T statistic	p-value
Student t-test	1.61	0.0580
Chen test	2.56	0.0052

Summary

Some researchers use the Student t-test even though the data do not satisfy the conditions required for the test to be valid. It is necessary to check the conditions of any statistical method before applying it to the data because the validity of the results of each method depends strictly on the legality of data conditions. In this paper, we highlight the conditions that need to be checked and provide an alternative test (Chen test) that should be used if the conditions are not met. We provide a case study that demonstrates the testing procedure. We note that Chen test is a one-sample t-test; however, we show that in the case where the data for a two-sample t-test are paired (which is common in color imaging; where, for example, the same samples are used to test various parameters) a one-sample test can be performed on the paired differences. MATLAB code is provided that performs the necessary condition checks and then performs a Student t-test or a Chen test accordingly. The MATLAB code is provided as an Appendix and can be downloaded from http://www.design.leeds.ac.uk/students/maryam_mohammadzadeh_darrodi.htm.

Acknowledgment

A valid statistical toolbox is required for using the MATLAB code attached in the Appendix.

Appendix

`function chen(data)`

```
%data should be a 2 columns matrix
x=data(:,1)-data(:,2);
H1='H1: mu1-mu2 > 0';
figure(20)
clf;
hold on;
subplot(1,2,1); hist(x);
xlabel('x')
ylabel('Frequency')
title('Histogram of data')
subplot(1,2,2); probplot(x);
hold off;

mu0=0;
n=length(x);
M=mean(x);
S=std(x);
beta=n*sum((x-M).^3)/((n-1)*(n-2)*S^3);
if beta<0
    x=-x; M=-M; beta=-beta; H1='H1: mu2-mu1 > 0';
end
disp(sprintf('\n'))
disp(sprintf('n=%d Mean=%d Std=%d',n,M,S))
disp(sprintf('Skewness Coef.=%d',beta))
```

`disp(H1)`

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Kolmogorov-Smirnov test for normality

```
[h,p,kstat] = lillietest(x,0.05,'norm',5e-4);
if p>=0.05
    N='can';
else
    N='can not';
end
disp(sprintf('\n'))
disp(sprintf('Lilliefors test (On the Komogorov-Smirnov test):
statistic ks=%d p_value=%d',kstat,p))
disp(sprintf('The distribution of the data %s be normal.',N))
%%%% plot
figure(2)
clf;
xx = -3:1:5;
F = cdfplot(x);
hold on
G = plot(xx,normcdf(xx),'r-');
set(F,'LineWidth',2)
set(G,'LineWidth',2)
legend([F G],...
'Empirical','Standard Normal',...
'Location','Best')
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
t=sqrt(n)*(M-mu0)/S;
ttest_p_value=1-tcdf(t,n-1);
disp(sprintf('\n'))
disp(sprintf('Paired t-test statistic t=%d
p_value=%d',t,ttest_p_value))
if (ttest_p_value<0.05)
    NN='is';
else
    NN='is not';
end
disp(sprintf('The paired t-test %s significant.',NN))
a=beta/(6*sqrt(n));
t2=t+a*(1+2*t^2)+4*a^2*(t+2*t^3);
p_value=1-normcdf(t2);
disp(sprintf('\n'))
disp(sprintf('Paired Chen test statistic t2=%d
p_value=%d',t2,p_value))
if (p_value<0.05)
    NN='is';
else
    NN='is not';
end
disp(sprintf('The paired Chen test %s significant.',NN))
```

References

- [1] V. Cheung, S. Westland, D. Connah et al., "A comparative study of the characterization of colour cameras by means of neural networks and polynomial transforms," *Coloration Technology*, 12, 19-25 (2004).
- [2] N. A. C. Cressie, and H. J. Whitford, "How to use the two sample t-test," *Biometrical Journal*, 2, 131-148 (1986).

- [3] N. J. Johnson, "Modified t tests and confidence intervals for asymmetrical populations," *Journal of American statistical association*, 73, 536-544 (1978).
- [4] R. R. Wilcox, "Comparing the means of two independent groups," *Biometrical Journal*, 7, 771-780 (1990).
- [5] P. Hall, "To remove the skewness by transformation," *J. Roy. Statist. Soc. B*, 54, 221-228 (1992).
- [6] S. H. Kim, and A. S. Cohen, "On the Behrens-Fisher problem: a review" *J. Educ. Behav. Statist*, 23, 356-377 (1998).
- [7] J. Aligna, T. C. Oshima, and W. Y. Lin, "Type I error rates for Welch's test and James's second-order test under non-normality and inequality of variances when there are two groups," *J. Educ. Behav. Statist*, 19, 275-292 (1994).
- [8] Student, "The probable error of a mean," *Biometrika*, 6, 1-25 (1908).
- [9] R. G. Miller, and Jr., [Beyond ANOVA, Basics of applied statistics] John Wiley, New York (1986).
- [10] J. P. C. Kleijnen, G. L. J. Kloppenburg, and F. L. Meeuwssen, "Testing the mean of an asymmetric population: Johnson's modified t test revisited," *Communications in statistics, part B-simulation and computation*, 15, 715-732 (1986).
- [11] C. D. Sutton, "Computer-intensive methods for tests about the mean of an asymmetrical distribution," *Journal of the American statistical association*, 88, 802-810 (1993).
- [12] L. Chen, "Testing the mean of skewed distribution," *Journal of the American statistical association*, 90, 767-772 (1995).
- [13] R. Nanadesikan, and M. B. Wilk, "Probability plotting methods for the analysis of data," *Biometrika (Boimetrika Trust)*, 55, 1-17 (1968).
- [14] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, 62, 399-402 (1967).

Author Biography

Maryam Mohammadzadeh graduated from the University of Shahid Beheshti Tehran, Iran in 2008 with a BSc degree in Statistics. Currently is a PhD student in the School of Design, University of Leeds working on Color semiotics in the process of product design under supervision of Professor Stephen Westland. She is a member of the Society of dyers and colorists of United Kingdom and also member of the Iranian Statistical Society.