Progress in Vibration and Acoustics, March 2015, Volume 3, Issue 1, 1-13 doi: 10.12866/J.PIVAA.2015.12.01

The effects of outliers on relative difference thresholds for automobile seat vibration with tarmac and pavè conditions

Progress in Vibration and Acoustics
ISSN : 2282 − 7668

Systematic Review

Graziella Aghilone[1] Massimo Cavacece[2]

[1] *Pharmacy and Medicine Faculty, University La Sapienza of Rome, Via Regina Margherita n.273, 00198 Rome (RM) Italy, e–mail: graziella.aghilone@uniroma1.it*

[2] *ASME Member, Department of Civil and Mechanical Engineering, University of Cassino and Lazio Meridionale, Via G. Di Biasio n.43, 03043 Cassino (FR) Italy, e–mail: cavacece@unicas.it*

Abstract In this research, we examine the methodologies proposed to analyze the data set in the presence of outliers. We review methods of determining outliers. We propose general principles for how to proceed. Most observations seem to follow a pattern or seem to satisfy a model. The outliers remain outside the pattern. Thus experimental data requires further investigation. [DOI:10.12866/J.PIVAA.2015.12.01] [1]

Keywords: Statistical approaches, Discomfort, Difference thresholds

1 Introduction

The verification of experimental conclusion plays a key role in scientific progress. It condenses science. Philosopher Karl Popper emphasizes its value and declares:

> We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated coincidence, but with events which, on account of their regularity and reproducibility, are in principle intersubjectively testable.

[1] Contributed by Technical Committee for publication in the Progress in Vibration and Acoustics. Manuscript receveid November 2, 2015; final manuscript revised November 9, 2015; pubblished online December 2, 2015.

Difference thresholds for whole–body vibration experienced on a car seat, investigated in the research [Mansfield and Griffin, 2000], are re–analyzed, evaluating the influence of outliers. Frequency weightings that have been established to predict the relative discomfort of different spectra can be used to compare the vibration discomfort in different vehicles and evaluate changes resulting from seats having different transmissibilities [ISO, 1997], [ISO, 1992], [BS7, 1989] and [BS6, 1987].

Genuine outliers are typically treated in one of the following ways:

1. keep the outlier and treat it like any other data point;
2. winsorize it or assign it lesser weight or modify its value so it is closer to the other sample values);
3. eliminate it from the sample. The danger of each of these methods is that they may produce poor estimates of parameters of interest.

Methods 1) and 3) may introduce statistical bias and may undervalue the outlier. Keep and treat the outliers like the other points may overvalue it and cause the estimate to vary drastically from the true population value [Ghosh and Vogt, 2012].

Below we investigate these methods, and make some proposals about what an outlier really is and how to treat it.

This paper is divided into two major subsections, statistical approaches and goodness of fit. In the first subsection we propose a comparison of statistical approaches to detect outliers. In the second section we propose goodness of fit by the two sample Kolmogorov–Smirnov test to validate the results.

2 Statistical approaches

In our study we describe re–analysis, replication and reproduction as methods of verifying experimental data. We propose three different methods as robust techniques.

2.1 Tukey's method

Box plots are an excellent tool for conveying location and variation information in data sets, particularly for detecting and illustrating location and variation changes between different groups of data. The standard boxplot rule, based on the upper Q_3 and lower Q_1 quartiles of the data distribution, offers an useful method of investigation to identify outliers.

The *interquartile range*, abbreviated IQR, is just the width in the box–and–whisker plot:

$$IQR = Q_3 - Q_1 \; . \tag{1}$$

The IQR can be used as a measure of how spread–out the values are. Statistics assumes that your values are clustered around some central value. The IQR analyses how the values are spread out. It can also be used to investigate when some of the other values are too far from the central value. Points too far away are called outliers, because they lie outside the range in which we expect them.

The IQR is the length of the box in your box–and–whisker plot. An outlier is any value that lies more than one and a half times the length of the box from either end of the box. If a data point is

$$\text{below} \quad (Q_1 - 1.5 \cdot IQR) \quad \text{or above} \quad (Q_3 + 1.5 \cdot IQR) \; , \tag{2}$$

it is viewed as being too far from the central values to be reasonable. The outliers are those points that don't seem to fit.

John Tukey, inventing the box–and–whisker plot in 1977, proposed one and a half times the width of the box to display these values. The $(1.5 \cdot IQR)$ is the demarcation line for outliers.

2.2 Grubbs' Test

Grubbs' test can be used to test the presence of one outlier and can be used with data that is normally distributed (except for the outlier).

Here we test the null hypothesis that the data has no outliers vs. the alternative hypothesis that there is one outlier. The ESD test should be used if there is the possibility of more than one outlier.

If you suspect that the maximum value in the data set may be an outlier you can use the test statistic

$$G = \frac{x_{max} - \bar{x}}{s} \ . \tag{3}$$

If you suspect that the minimum value in the data set may be an outlier you can use the test statistic

$$G = \frac{\bar{x} - x_{min}}{s} \ . \tag{4}$$

The critical value for the test is

$$G_{crit} = \frac{(n-1)\, t_{crit}}{\sqrt{n\,(n-2+t_{crit}^2)}} \ . \tag{5}$$

where t_{crit} is the critical value of the t distribution $[T \cdot (n-2)]$ and the significance level is α/n. Thus the null hypothesis is rejected if $G > G_{crit}$.

There is also a two–tailed version of the test where G is the larger of the two G values described above and G_{crit} is defined as above except that the significance level for t_{crit} is $\alpha/\left(2 \cdot n\right)$. Alternatively, G can be calculated using the formula

$$G = \frac{\max |x_i - \bar{x}|}{s} \ . \tag{6}$$

2.3 Winsorising

Winsorising or Winsorisation is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers.

The distribution of many statistics can be heavily influenced by outliers. A typical strategy is to set all outliers to a specified percentile of the data. Winsorised estimators are usually more robust to outliers than their more standard forms, although there are alternatives, such as trimming, that will achieve a similar effect.

A common procedure has been to replace any data value above the ninety–fifth percentile of the sample data by the ninety–fifth percentile and any value below the fifth percentile. The assumption seems to be that the outlier must be like other data. This suggests that the outlier value must be incorrect. The value is replaced by a more plausible value. The new value is a compromise. The outlier is not thrown out but it is a mediate value.

2.4 Testing of the cumulative distributions by two sample Kolmogorov–Smirnov test

Comparison of the cumulative distributions is obtained by using the basic approach of Kolmogorov Smirnov (K–S) test. The two sample K–S test is used to test whether two samples come from the same distribution.

Suppose X_1, X_2, ..., X_m are a sample from group 1 and Y_1, Y_2,..., Y_n are a sample from group 2. Suppose the cumulative distribution function for group 1 is $F(x)$. The first sample $F(x)$ has size m. The cumulative distribution function for group 2 is $G(y)$. The second sample $G(y)$ has size n. We assume both $F(x)$ and $G(y)$ are continuous distributions, meaning there are no ties in our data.

Key facts about the Kolmogorov–Smirnov test are following ones:

- The two sample Kolmogorov–Smirnov test is a nonparametric test that compares the cumulative distributions of two data sets.
- The test is nonparametric. It does not assume that data are sampled from Gaussian distributions or any other defined distributions.

The null hypothesis is that both groups were sampled from populations with identical distributions. We want to test between the following hypotheses:

$$\begin{cases} H_0: & F(x) = G(y) & \text{for all X} \\ H_1: & F(x) \neq G(y) & \text{for some X}, \end{cases} \tag{7}$$

Now we can evaluate whether there is a difference between the groups, not making any assumptions about the nature of the difference. The phases of test are the following ones:

✓ Define

$$D_{m,n} = \max_x \left[F(x) - G(y)\right] \ ; \tag{8}$$

✓ Calculate the critical value:

$$D_{m,n,\alpha} = c(\alpha)\sqrt{\frac{m+n}{mn}} \ , \tag{9}$$

The values of $c(\alpha)$ can be deduced from the Kolmogorov–Smirnov table.

✓ The null hypothesis is H_0: both samples come from a population with the same distribution. If $D_{m,n} > D_{m,n,\alpha}$, we reject the null hypothesis at significance level α.

3 Results

In this study we propose a re–analysis of results obtained in the research [Mansfield and Griffin, 2000]. We summarize the peculiar aspects that characterize experimental procedures (Table 1).

The relative difference thresholds, measured using using male and female subjects and four stimuli, are listed in Table 2 and in Table 3 [Mansfield and Griffin, 2000].

The comparison of statistical methods offers different values of outliers, represented in Table 4. We can estimate the location of the outliers (Fig.1 and Fig.2). In addition, we can deduce the comparison between our results, obtained by statistical methods, and ones deduced in the research [Mansfield and Griffin, 2000].

Obtained the values of outliers, the process of winsorising offers the following results:

Authors	Mansfield–Griffin
Female	20
Male	20
Number of trials	—
Step–size [dB]	0.25
reference stimulus [s]	10
pause [s]	2
test stimulus [s]	10
Frequency [Hz]	5, 20
Magnitude [ms^{-2}] r.m.s.	0.2, 0.4, 0.8 (weighted), pave

Table 1: Experimental procedures for relative difference thresholds

Subject	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
M1	17.8	22.9	27.6	27.2
M2	16.1	19.4	25.1	13.6
M3	11.6	13.3	9.5	13.9
M4	9.4	14.4	7.1	12.3
M5	16.7	9.5	6.8	15.8
M6	12.2	36.1	13.3	27.3
M7	6.0	7.1	5.9	14.3
M8	8.7	33.4	9.0	22.7
M9	6.5	15.1	13.1	10.8
M10	11.4	15.3	14.2	17.9

Table 2: Relative difference thresholds measured using male subjects and four stimuli

Subject	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
F1	15.5	15.6	14.7	11.7
F2	16.0	7.3	10.7	13.8
F3	13.5	12.1	10.3	11.4
F4	11.8	12.9	13.6	11.5
F5	18.5	14.8	8.4	17.4
F6	9.9	17.6	29.5	28.3
F7	32.3	11.4	30.4	29.5
F8	11.4	12.5	9.6	15.4
F9	22.2	8.4	13.0	12.4
F10	34.9	6.9	8.4	9.3

Table 3: Relative difference thresholds measured using female subjects and four stimuli

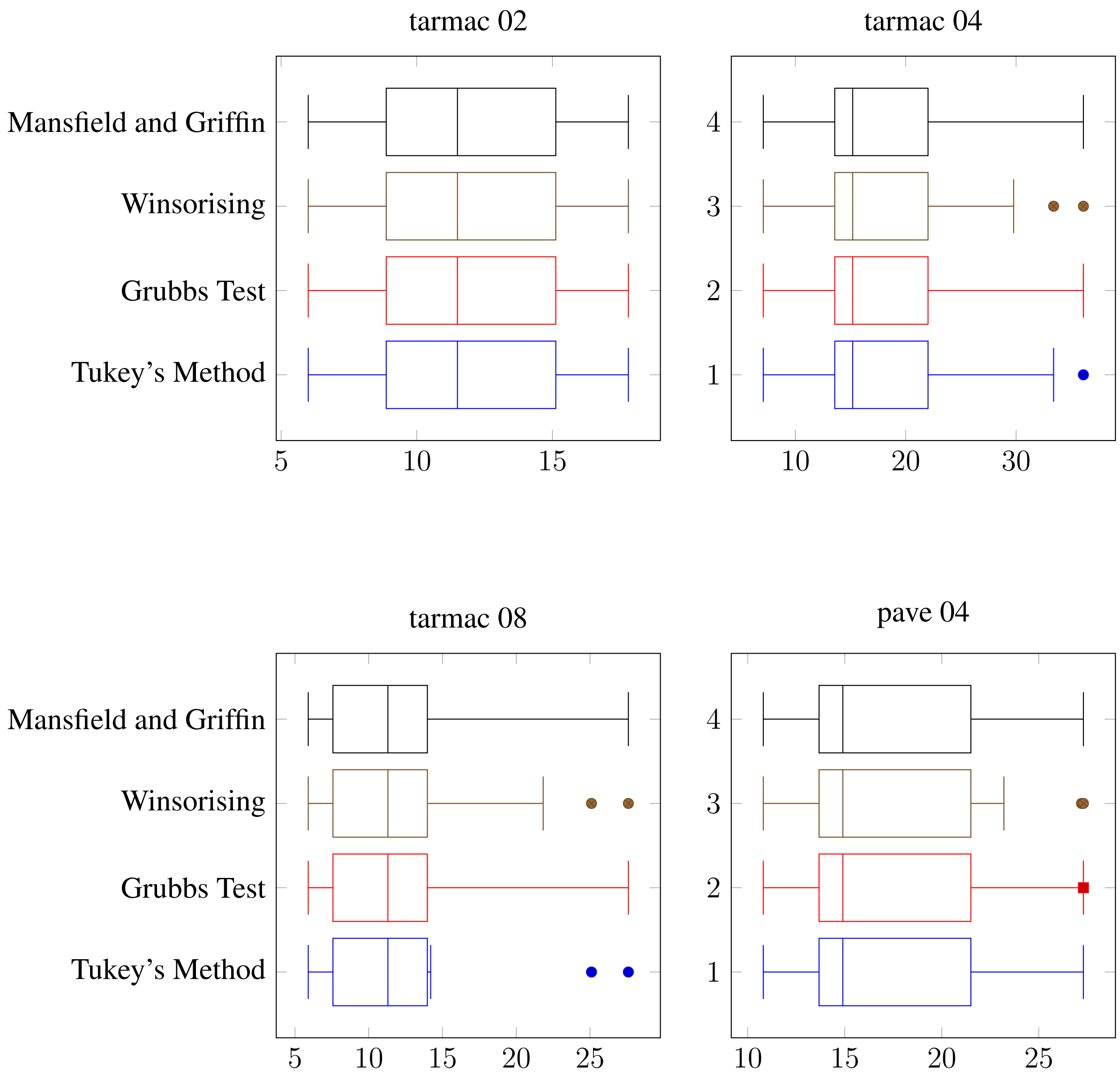

Figure 1: A comparison of statistical values for relative difference thresholds measured using male subjects

(i) data of relative difference thresholds measured using male and female subjects and four stimuli (Table 5 and Table 6);

(ii) mean values of data (i) (Fig.3 and Fig.4);

(iii) sample quantile of data (i) (Table 7).

Kolmogorov–Smirnov test allows us to compare the cumulative distributions of relative difference thresholds measured using male and female subjects and four stimuli, obtained by using the process of winsorising (Table 8, Table 9 and Table 10).

4 Discussion

Tukey's method and Grubbs' test offer important tools to detect outliers from experimental

	Male			Female		
Vibration stimuli	Tukey's method	Grubbs Test	process of Winsorising	Tukey's method	Grubbs Test	process Winsorising
Tarmac 0.2	—	17.8	—	—	34.8	32.3, 34.8
Tarmac 0.4	36.1	36.1	33.4, 36.1	—	17.6	—
Tarmac 0.8	25.1, 27.6	27.6	25.1, 27.6	29.5, 30.4	30.4	29.5, 30.4
Pave 0.4	—	27.3	27.2, 27.3	28.3, 29.5	29.5	28.3, 29.5

Table 4: Outliers for relative difference thresholds measured using male/female subjects and four stimuli

Subject	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
M1	17.8	22.9	21.82	23.20
M2	16.1	19.4	21.83	13.6
M3	11.6	13.3	9.5	13.9
M4	9.4	14.4	7.1	12.3
M5	16.7	9.5	6.8	15.5
M6	12.2	29.88	13.3	23.20
M7	6.0	7.1	5.9	14.3
M8	8.7	29.88	9.0	22.7
M9	6.5	15.1	13.1	10.8
M10	11.4	15.3	14.2	17.9

Table 5: Data of relative difference thresholds measured using male subjects and four stimuli obtained by the process of winsorising

Subject	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
F1	15.5	15.6	14.7	11.7
F2	16.0	7.3	10.7	13.8
F3	13.5	12.1	10.3	11.4
F4	11.8	12.9	13.6	11.5
F5	18.5	14.8	8.4	17.4
F6	9.9	17.6	19.41	19.03
F7	28.05	11.4	19.41	19.03
F8	11.4	12.5	9.6	15.4
F9	22.2	8.4	13.0	12.4
F10	28.05	6.9	8.4	9.3

Table 6: Data of relative difference thresholds measured using female subjects and four stimuli obtained by the process of winsorising

Stimulus	Quantile	Sample Quantiles 0%	25%	50%	75%	100%
Tarmac 0.2	Male	5.90	7.57	11.30	13.97	21.83
	Female	9.90	12.22	15.75	21.27	28.05
Tarmac 0.4	Male	7.10	13.57	15.20	22.02	29.88
	Female	6.90	9.15	12.30	14.32	17.60
Tarmac 0.8	Male	5.90	7.57	11.30	13.97	21.82
	Female	8.40	9.77	11.85	14.42	19.41
Pave 0.4	Male	10.80	13.67	14.90	21.50	23.20
	Female	9.30	11.55	13.10	16.90	19.03

Table 7: Sample quantile for relative difference thresholds measured using male/female subjects and four stimuli

Kolmogorov–Smirnov test p–value				
Stimuli	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
Tarmac 0.2	...	0.06964	0.8798	0.02837
Tarmac 0.4	0.06954	...	0.07522	0.8798
Tarmac 0.8	0.8798	0.05354	...	0.03423
Pave 0.4	0.02831	0.8798	0.09618	...

Table 8: Kolmogorov–Smirnov test for relative difference thresholds measured using male subjects and four stimuli

Kolmogorov–Smirnov test p–value				
Stimuli	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4
Tarmac 0.2	...	0.06382	0.1984	0.5705
Tarmac 0.4	0.06392	...	0.705	0.3073
Tarmac 0.8	0.08187	0.7049	...	0.2565
Pave 0.4	0.3071	0.3071	0.4053	...

Table 9: Kolmogorov–Smirnov test for relative difference thresholds measured using female subjects and four stimuli

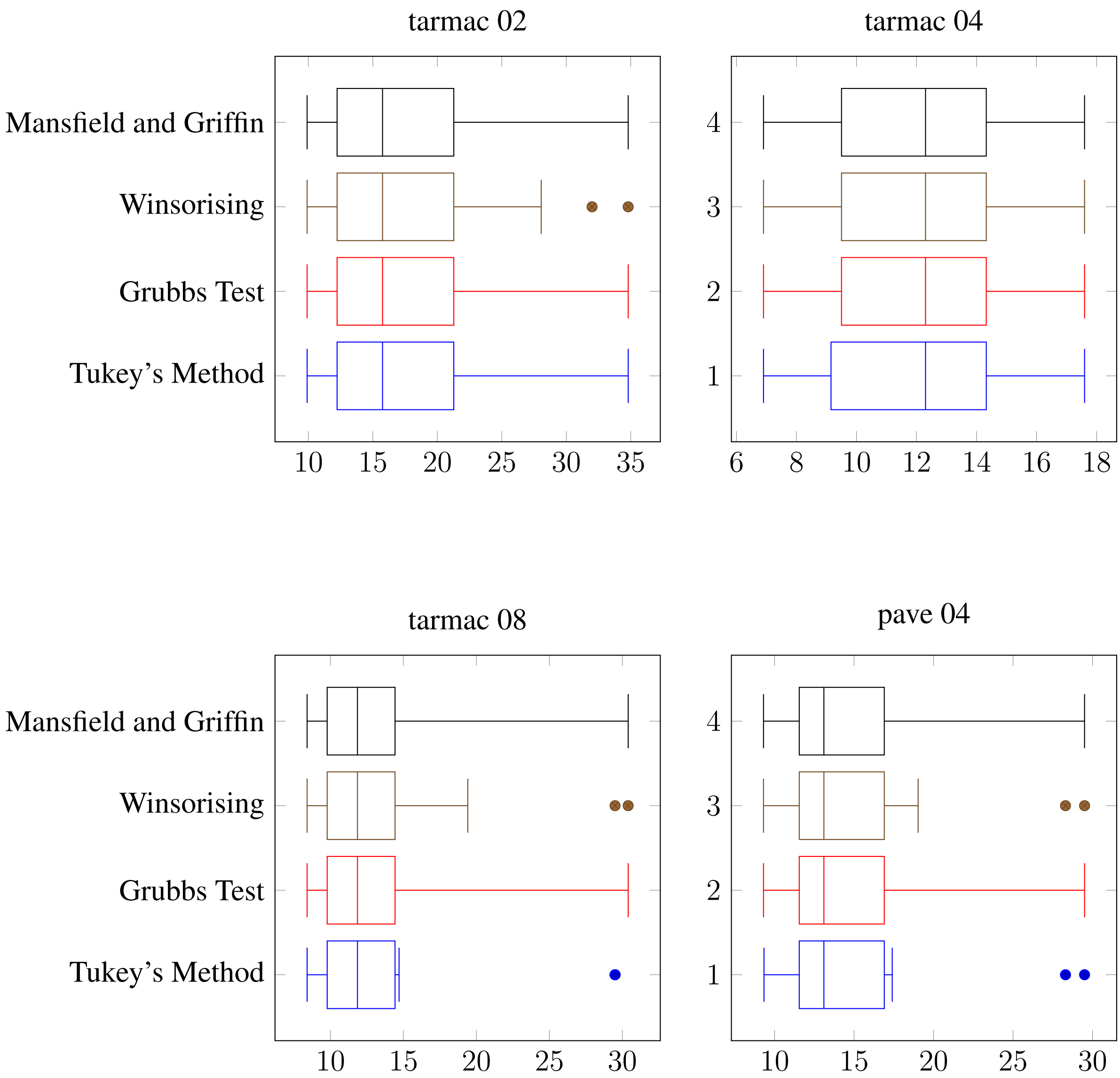

Figure 2: A comparison of statistical values for relative difference thresholds measured using female subjects

data. The comparison between Tukey's method and Grubbs' test can help the interpretation of the information obtained by the results.

The box–and–whisker plot is developed by the Tukey's method, evaluating the interquartile range (IQR) as measure of variability. The box–and–whisker plot is a standard technique for presenting a summary of the distribution of dataset. The concise representation provides an intuitive comprehension of a distribution. Overall, the simplicity of the box plot makes it an elegant method for the presentation of scientific data. An important limitation of Tukey's method is that the term $(Q_1 - 1.5 \cdot IQR)$ can become negative even positive values of data.

The Grubbs' test detects outliers from normal distributions. The result of Grubbs' test denotes that the data belongs to the core population. The typical aspects of Grubbs' test are following ones:

× The minimum and maximum values are the tested data.

× The mean and the standard deviation of the values in the data set must be calculated.

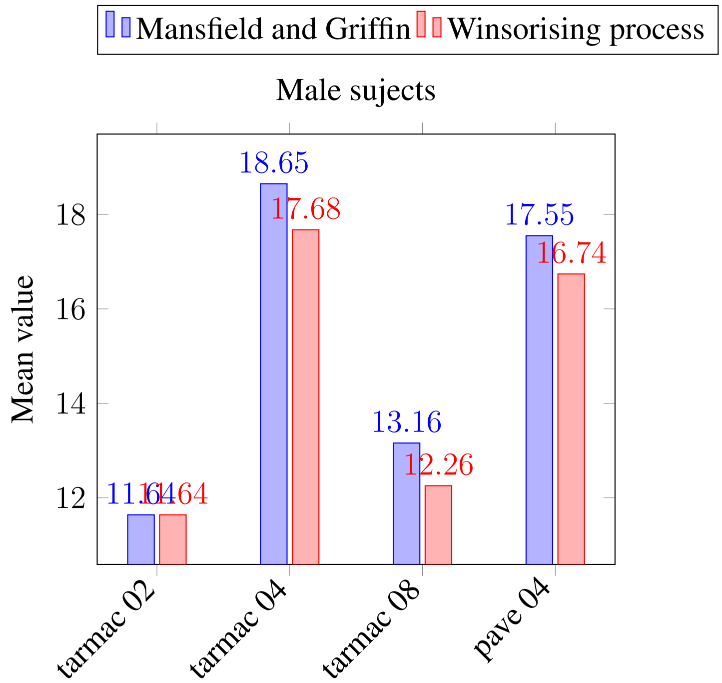

Figure 3: Mean values of relative difference thresholds measured using male subjects and four stimuli

× The test is based on the difference of the mean of the sample and the most extreme data considering the standard deviation.

There are the following limitations to Grubb's test:

- The investigated data, with outliers, must be nearly normally distributed.
- The Grubbs' test is valid for the detection of a single outlier. The test cannot be used for a second time on the same set of data.
- The Grubbs's test should be applied with caution. The same conclusion should be applied to all statistical tests used for rejecting data. There is a probability, equal to the significance

p–value Kolmogorov–Smirnov test Female subjects					
Stimuli	Tarmac 0.2	Tarmac 0.4	Tarmac 0.8	Pave 0.4	
	Tarmac 0.2	0.05381	0.7913	0.5966	0.1618
Male	Tarmac 0.4	0.9397	0.06964	0.1986	0.4497
Subjects	Tarmac 0.8	0.06964	0.8798	0.4055	0.1988
	Pave 0.4	0.9698	0.04125	0.1209	0.4057

Table 10: Kolmogorov–Smirnov test for relative difference thresholds measured using male and female subjects and four stimuli

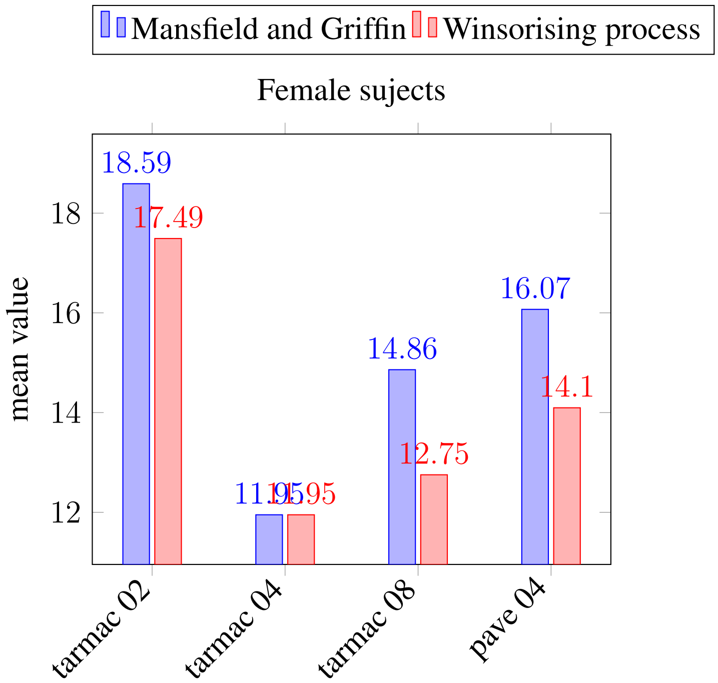

Figure 4: Mean values of relative difference thresholds measured using female subjects and four stimuli

level a ($a = 0.05$ at the 95% confidence level) that an outlier identified by the Grubbs' test actually is not an outlier.

- If the investigated samples have an asymmetric distribution (e.g. lognormal), then the Grubbs' test gives false results.

The methods of Tukey and of Grubbs identify outliers. However, the outliers are not cancelled. The process of winsorising replaces outliers with values more plausible. The process of winsorising is very simple and it offers a new data sequence. The new value, obtained by process of winsorising, is a compromise. To winsorize, one converts the value(s) of data points that are outlying data high to the value of the highest data point not considered to be an outlier. We have observed that, in the re–analysis of results [Mansfield and Griffin, 2000], the maximum values of whiskers and mean values of box–plot depend on spurious outliers (Fig.1 and Fig.2). We can affirm that the process of winsorising is more efficient than Tukey's method and Grubbs' method for the following reasons:

- it identifies the outliers;
- it reduce the effects of the outliers.

The knowledge of the outliers can indicate an important effect. The result of the psycho–physical response of the subject depends on the willingness or the capacity of the subject's cooperation. The situations psycho–physical contingents, the fluctuation of the level of sensitivity are critical elements in the experimental measurements. Vibrations may have been observed at different location on the body. It can be assumed that relative difference thresholds for whole–body will depend on the method used for their evaluation. Variations in the environment (e.g. noise, seating comfort) between two conditions may increase or decrease relative difference thresholds.

Another outcome can be the comparison of the relative difference thresholds measured using male, female subjects and four stimuli. Kolmogorov–Smirnov (K–S) test offers the comparison between measurements made using different stimuli and subjects. Before evaluating results, we observe important aspects of K–S test. If the outcome are categorical with many ties, we can not use the Kolmogorov–Smirnov test. We can use it only for ratio or interval data, where ties are rare. Interpreting the p–value is an another important question. Evaluating p–value, we can conclude that the two groups were or not sampled from populations with different distributions. The populations may differ in median, variability or the shape of the distribution.

The p–value of (K–S) test shows following remarks

- With regard to male subjects, we obtain different distribution between tarmac 02 and pave 04 conditions using male subjects (Tab.8). In this case we obtain p–value $= 0.02831 < 0.05$. In other cases we can accept the null hypothesis H_0 since p–value is greater than the default value $\alpha = 0.05$ of the level of significance.
- With regard to female subjects, we can accept always the null hypothesis (Tab.9).
- Comparing male and female subjects, we can not accept the null hypothesis H_0 between tarmac 08 (female subjects) and pave 04 (male subjects). In this case we obtain p–value $= 0.04125 < 0.05$ (Tab.10). In other cases we can adopt the null hypothesis H_0 since p–value is greater than the default value $\alpha = 0.05$ of the level of significance.

According to K–S test, the difference between two samples is not significant enough to say that they have different distribution. Most comparisons do not show any significant differences between measurements made using the different stimuli.

5 Conclusion

In order to avoid the clipping in signal processing of experimental data, the winsorising process is an excellent transformation of statistics that reduces the effect of spurious outliers.

In the experimental data it seem that some observations deviate markedly from other observations. An outlier may indicate bad data. If it is possible, we suggest that the outlying value should be corrected in the analysis. Outliers may be due to random variation or may indicate something scientifically interesting. In any event, we do not want to simply delete the outlying observation. However, if the data contains significant outliers, we propose the process of winsorising as robust statistical techniques.

K–S test compares pairs of data. The comparison of the relative difference thresholds measured using male and female subjects and four stimuli do not show any significant differences between measurements made using the different stimuli.

References

British standard guide to measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock., 1987.

Guide to safety aspects of experiments in which people are exposed to mechanical vibration and shock, 1989.

Mechanical vibration laboratory method for evaluating vehicle seat vibration. part 1: basic requirements. iso 10326-1, 1992.

Mechanical vibration and shock – evaluation of human exposure to whole-body vibration. part 1: General requirements. iso 2631-1., 1997.

D. Ghosh and A. Vogt. Outliers: An evaluation of methodologies. In *Joint Statistical Meetings*, pages 3455–3460, 2012.

N.J. Mansfield and M.J. Griffin. Difference thresholds for automobile seat vibration. *Applied Ergonomics*, 31:255–261, 2000.