Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs).

Jason J. Braithwaite

{Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham, UK}

Derrick G. Watson

{Department of Psychology, University of Warwick, UK}

Please cite as: Braithwaite, J.J., & Watson, D.G (2015). Issues surrounding the normalization and standardisation of skin conductance responses (SCRs). Technical Research Note. Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham©

The problem

It is not a viable or sound method to compare the raw Skin Conductance Responses (SCRs) from one individual, to those of another individual, and then derive interpretations over the true magnitude of emotional / autonomic response between them, to a given experimental manipulation. The magnitude of a basic numerical response is not the same as the magnitude of the emotional response - and it is the latter that we are really interested in. Consider the following scenario. Imagine a participant (Participant 1) who gives a response of '5' to a given stimulus presented on a computer screen, and another person (Participant 2) who gives a response of '10' to the same stimulus. The numerical response of Participant 2 is twice that of Participant 1 - ***but this does not mean that their emotional response is twice as large***! Therefore, although seductive, one cannot conclude that the second participant gave a larger ***emotional*** response. The reason for this is that we do not know, at this stage, what the overall responsiveness of each individual is. We must know this in order to directly compare the magnitude of responsiveness across individuals. Let us now imagine that Participant 1 produces non-specific SCRs (SCRs not connected to our experimental stimulation - those that represent the general background 'chatter' of their physiological system) of around '1' and Participant 2 gives non-specific SCR responses of around '9'. What this now means is that the difference between the stimulus of interest in the experiment, and the background 'chatter' of the electrodermal system not tied to experimental stimuli is '4' for Participant 1 and is '1' for Participant 2. Therefore, only by considering the full profile of individual responsiveness can we see that our initial view is reversed. It is the first participant who is providing a stronger response to the experimental stimulus. Not only do we need to consider the broader picture of electrodermal reactivity of each participant, but we also require a viable way to transform the data into a standardised range that represents the magnitude of responses accurately across individuals thus facilitating direct comparisons across those individuals (see Figure 1 for an illustration).

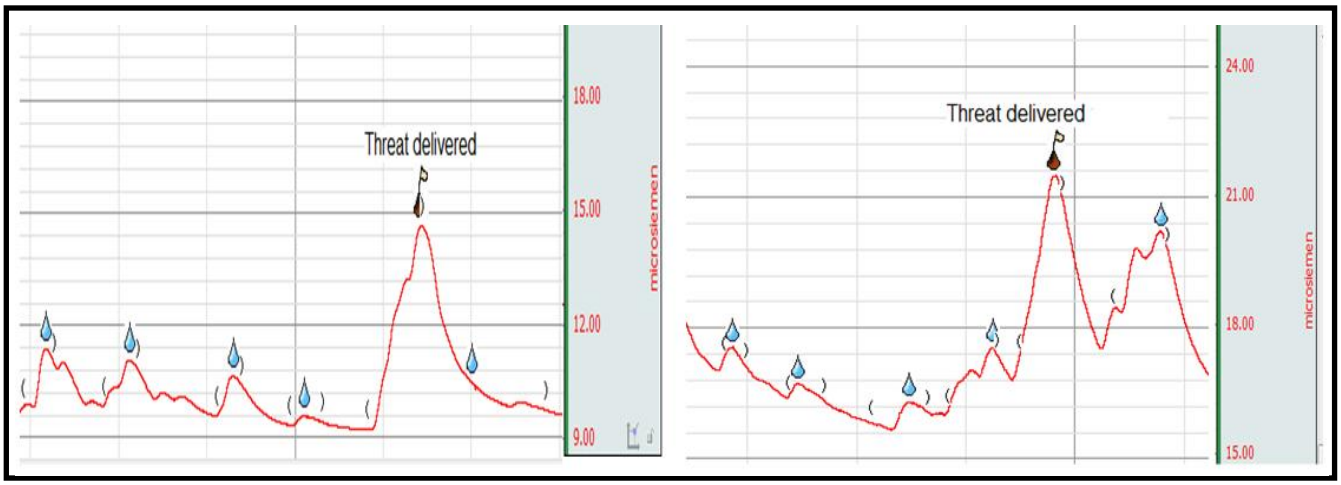


Figure 1. Separate signals (left and right) from different participants taking part in the same experiment. Background non-specific SCRs are denoted by the blue droplets. Event-related SCRs (a threat in the case of this example) are denoted by the red droplet with a flag. The issue of standardization relates to the question of which participant gave the largest emotional response to the threat? These signals can only be compared statistically once the scale of the SCRs have been transformed into a standardised range that takes the whole responsiveness of the individual into account. As discussed below, the most appropriate methods of standardisation are those which consider the whole signal (non-specific and event-related SCRs) and a calculation of mean reactivity to transform individual SCRs. A z-score or a ratio calculation could both be appropriate (discussed below).

Data Transformations: Normalizations & Standardizations

There is considerable confusion and varied practice in the field of SCR research surrounding the procedures of Normalization and Standardization. These components are not the same thing.

Normalization refers to the application of data transformations to correct for the presence of skew / kurtosis so that the SCRs are fit for parametric statistical analysis. Standard procedures might include; (i) taking the Log of the amplitudes; (ii) applying a square-root transformation to the SCR amplitudes; (ii) using the Log(SCR+1) routine applied to the SCR magnitudes. If these fail to normalize the data, then non-parametric analysis should be considered.

Standardization refers to correcting the SCR data so that they are in a standardised form for comparison across individuals. Standardization may not always be necessary, but it is difficult to think of a scenario when it might not be a prudent idea to transform SCRs from an individual, into a co-ordinate system that reflects the range of their capacity to respond per-se. The purpose of standardization is primarily to facilitate individual-difference comparisons where all SCRs have been transformed, relative to the participant's typical physiological responsiveness. In principle this is no different to the neurophysiologist wanting to compare individual differences in Event-Related Potentials (ERPs) but where the thickness of the skull, air-gap between brain and skull, and sensor position etc, may all vary slightly.

Standardization for Individual Differences

On occasion the researcher might want to transform SCRs or Skin Conductance Levels (SCL) into a standardised and corrected form (to correct for inter-individual variance) that can facilitate an individual-difference analysis. There is no universally agreed on method for this and there are pros and cons to every procedure. A number of suggestions have been made and used in the literature and their utility is still being debated. The most common are as follows:

SCL data: Range-Corrected Scores: Here one computes the minimum SCL during a baseline or rest period and a maximum SCL during the most arousing period. The participants SCL at any other time period in the study can then be delineated as a percentage of their individual range of psychophysiological response via the formula $(SCL - SCL_{min}) / (SCL_{max} - SCL_{min}$: Dawson et al., 2001).

SCR data: Proportion of Maximal Response: In the case of SCR data, we can assume the minimum to be zero and the maximum to be the result of a startle stimulus (i,e., surprising white noise stimuli played through earphones, hand-clap, balloon pop, and / or taking deep breaths). From this, each individual SCR can then be standardised for individual differences by dividing it by the participant's maximum SCR (a single value). This method is now viewed as controversial as maximal responses are unstable, inconsistent, just a single value, and it is not clear what they truly represent.

Transformations into standard values: Some studies recommend transforming SCRs into Z-scores (Ben-Shakhar, 1987; 1985; Bush et al., 1993). This requires the calculation of an overall mean and standard-deviation to be used instead of a hypothetical maximum (from the other methods above). Using a Z-score navigates around the problems associated with determining the maximum SCR response from range-corrected methods / maximal correction methods. Here each

raw SCR, a mean SCR value and standard deviation of SCRs, are used to compute the resultant Z-score which is normally distributed, has an average of 0 and a standard deviation of 1[1]. The advantage of this approach is that the resultant scores are based on mathematical factors that represent the participants' response level and not on unwarranted assumptions about unstable maximum SCRs. Such transformations also increase the power in analyses which is another advantage to their use (Ben-Shakhar, 1985; Bush, 1993). Transformations like the z-score, probably improve power by scaling the effects relative to a participants own variability or reactivity. In addition, an advantage to using the z-score method is that the standard deviation is a more stable estimate of dispersion than that utilised by maximum score / range methods which are more influenced by outliers. One worry about standardization methods is that they might distort the 'true' underlying pattern of results, altering the order of effects from different people or conditions. However, as Bush et al (1993) show across a host of simulations, differences (i.e., reversals) in the order of effects hardly ever occurred for significant effects, and when there was a difference it was the z-transformation that was much more likely than raw scores / means to reveal the correct order of effects.

One argument against the use of z-scores is that there can be a positive correlation between the size of the standard-deviation and the magnitude of raw SCRs - thus possibly leading to a bias in the transformed values (Stemmler, 1987). However, based on running a series of Monte-Carlo simulations, Bush et al (1993) repeatedly found that the z-score transformation was the best in terms of power relative to a host of alternatives. In addition, Ben-Shakhar, (1987) further suggested that another useful transformation might be to divide each raw SCR, by the participant's mean SCR thus providing a kind of 'standardised ratio' measure. This ratio method does not utilise the standard deviation and thus avoids all the limitations outlined by Stemmler (1987). In practical terms, z-scores are used extensively in data transformations in neuroscience and any biases that might be present are considered minor - the advantages far out-weigh any minor limitations. Nonetheless, the important issue here is that any transformation used by the researcher is one based on computed arithmetic means rather than those of theoretical maximum values which are based on various assumptions not all of which can be verified.

Normalizations and standardizations may not always be necessary. For example, if the data are normally distributed and have a homogeneous variability across individuals and conditions, then there appears to be no discernible benefit from running transformations on the data set - taking logs (for example) does not improve homogeneity under these circumstances (Bush et al., 1993). However, if variability is heterogeneous and differs across subjects and conditions, transformations can have a large and beneficial effect . The z-score works by equating variability across subjects and conditions, but these benefits appear to be maximal under conditions of non-normal distributions and heterogeneity of variance. Because SCR data are indeed inherently skewed and do produce non-normal distributions and substantially heterogeneous variability across participants - running transformations should be considered at least the norm, and at most a necessity.

[1] From here one can transform these Z-scores into T-scores, which have a mean of 50 and standard deviation of 10 (thus removing minus scores: see Boucsein, 2012).

A recommendation

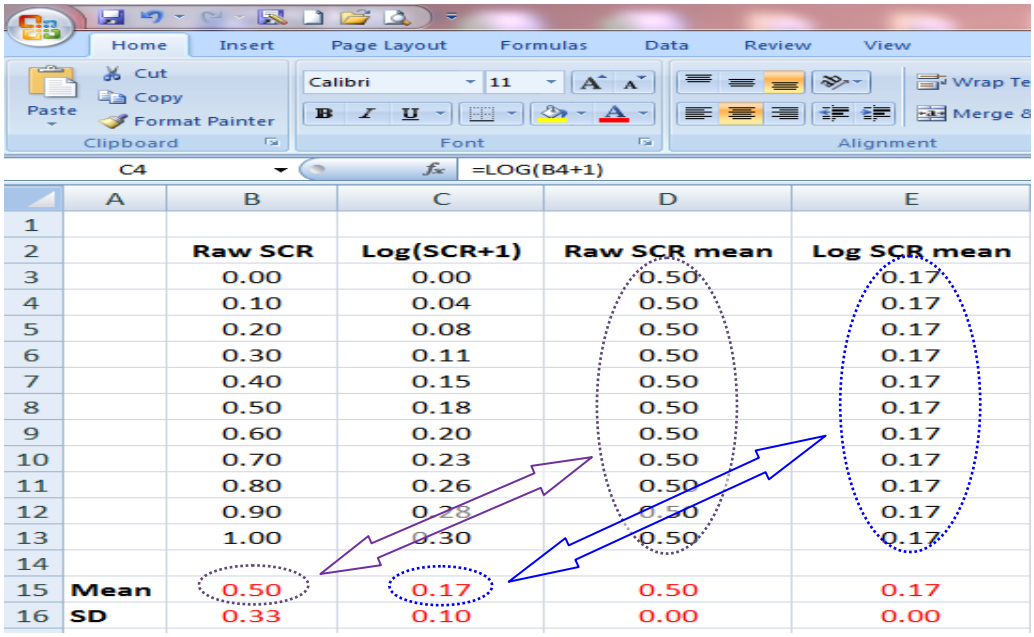
If a z-score or a ratio is to be used, it would be advisable to base the necessary descriptive statistics on data merged from all signals for a given participant. Therefore, if a basic experiment has three conditions, all three signals should be merged and the descriptive statistics ($\bar{X}$ and sd for z-scores, or $\bar{X}$ for ratio) for the standardisation process should be computed from these merged signals. Both event-related and non-specific SCRs would be included in this process to produce these overall descriptive statistics. This helps to reduce additional bias that may emerge due to extreme differences in numbers of SCRs across individuals and to ensure that the mean and standard deviation is derived from a sufficient number of individual SCRs to be an accurate representation of psychophysiological reactivity for the individual. In other words, each SCR counts as data point in the total sample (family) of SCRs from all conditions.

Data trimming & outliers

One suggestion for improving the power of an SCR analysis via z-score transformations has been to trim signals for outlier values before transforming the data (see Bush et al., 1993). Outliers can have a higher impact on data from skewed distributions. There are a number of methods available - but unfortunately all involve data loss. Arbitrary cut-off values (i.e., SCRs that are 3 standard deviations above / below the mean in a given cell) are not advisable. Bush et al (1993) suggested removing the highest and lowest SCR from each individual's data set, for each condition, can increase the power of analysis with both normal and non-normally distributed data - though such benefits are not huge. In practice the act of data trimming is extremely rare in SCR research and the true benefits debatable. Indeed, it might not always be feasible or practical for many experimental contexts including event-related paradigms. For example, maximum SCR trimming would not be advisable if a single presentation of a given stimulus occurs, and the prediction is the reaction might be greater for the content of that stimulus, relative to other stimuli (for whatever reason). In this circumstance, if the prediction is correct, the researcher would end up having to throw away the very SCR data of interest. Presenting the stimulus more than once does not navigate around this problem as one soon enters the realm of habituation effects where responses become attenuated as the physiological system becomes overly familiar with the repeated presentation of that specific content.

Other designs may lend themselves more towards the notion of SCR trimming though these decisions are very much down to the researcher and the justifications for carrying out such methods (or not as the case may be). Ambulatory measurements, where SCRs are measured over a longer period of time could be one where the rejection of the maximum / minimum SCR might be more practical - though again, this could be problematic for individuals producing fewer SCRs in the first place as resultant means and standard deviations would be based on fewer contributory SCRs.

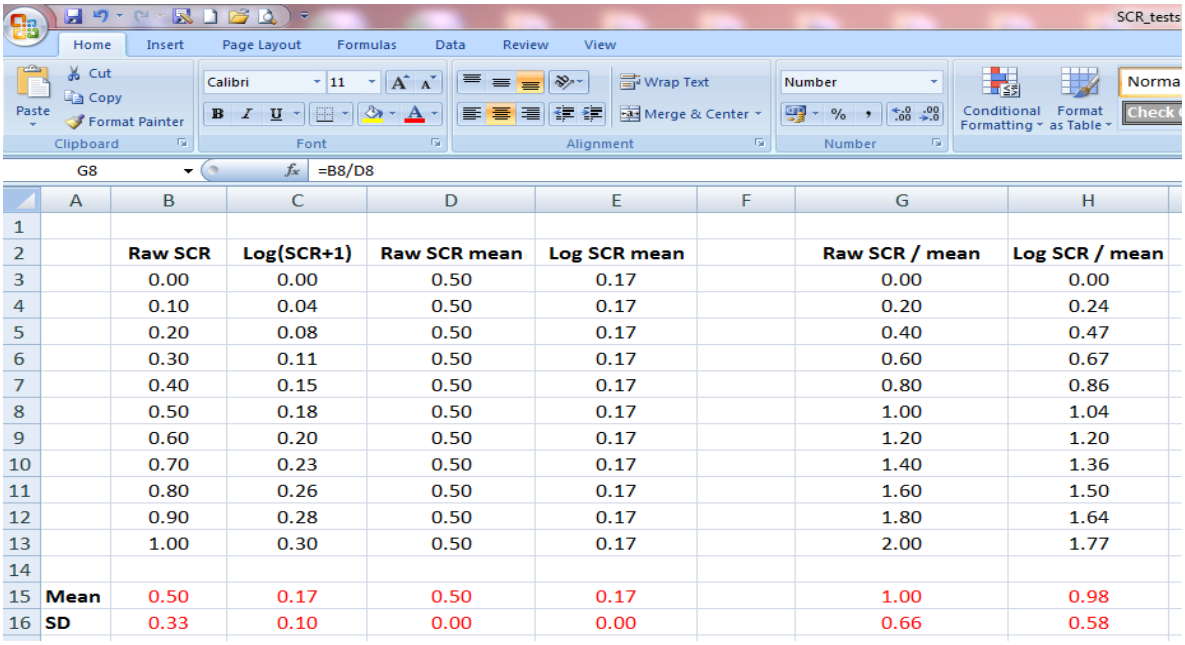
A Worked Example:



	A	B	C	D	E
1					
2		**Raw SCR**	**Log(SCR+1)**	**Raw SCR mean**	**Log SCR mean**
3		0.00	0.00	0.50	0.17
4		0.10	0.04	0.50	0.17
5		0.20	0.08	0.50	0.17
6		0.30	0.11	0.50	0.17
7		0.40	0.15	0.50	0.17
8		0.50	0.18	0.50	0.17
9		0.60	0.20	0.50	0.17
10		0.70	0.23	0.50	0.17
11		0.80	0.26	0.50	0.17
12		0.90	0.28	0.50	0.17
13		1.00	0.30	0.50	0.17
14					
15	**Mean**	0.50	0.17	0.50	0.17
16	**SD**	0.33	0.10	0.00	0.00

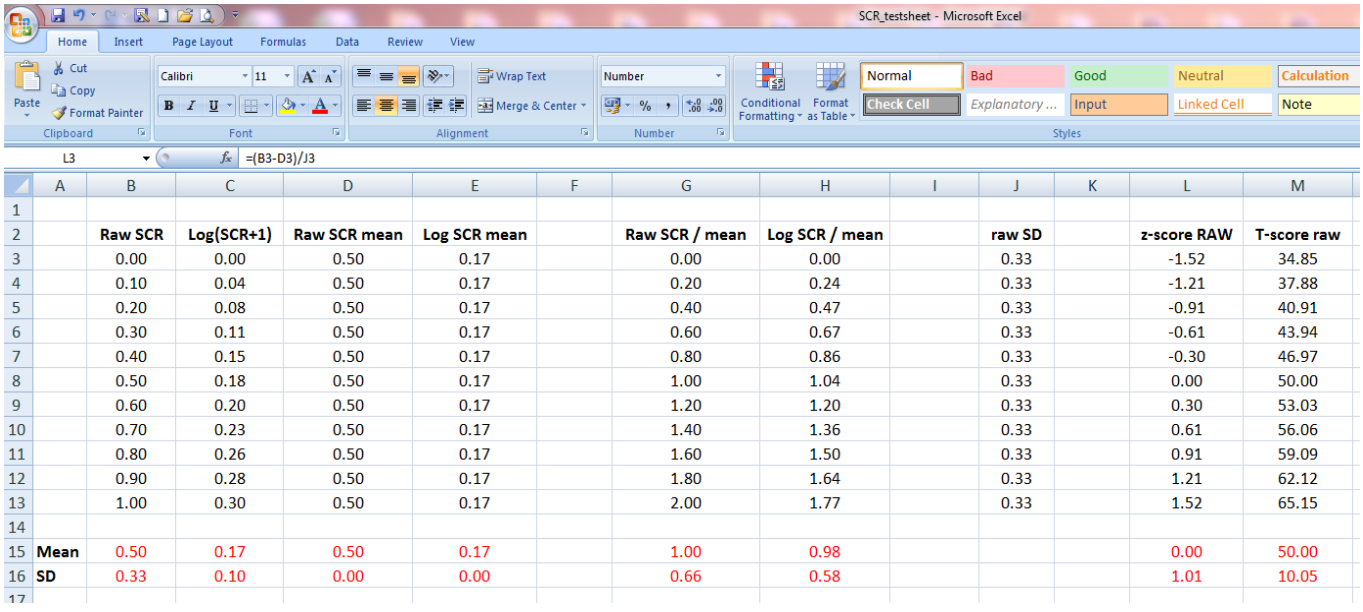
Figure 2. An initial example using fabricated data for illustration. Raw SCR magnitudes are shown in Column B. In Column C normalised {Log(SCR+1)} SCRs are shown (details in the equation editor as well)[2]. The overall mean and standard deviations for each column are given in red font at the bottom of the column. For transforming each SCR, these means are then repeated in Column D (and for the logged data Column E). Note the equation (Log(SCR+1) is applied to raw SCR magnitudes which include zero values, not SCR amplitudes which discard them (see Braithwaite et al., 2015; Dawson et al., 2001 for further discussion).

[2] Note - if the SCRs were amplitudes, then basic log-transformations or square-root transformations would be appropriate. The constant of '1' is added as with magnitude data a log of zero cannot be taken.



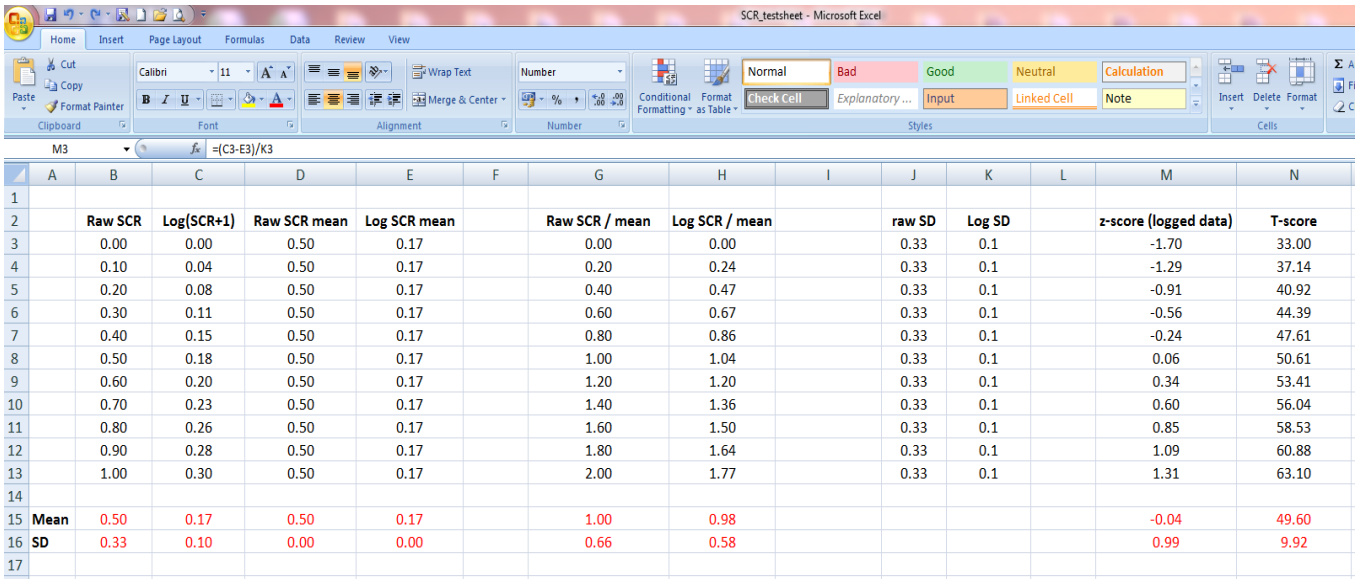
	A	B	C	D	E	F	G	H
1								
2		**Raw SCR**	**Log(SCR+1)**	**Raw SCR mean**	**Log SCR mean**		**Raw SCR / mean**	**Log SCR / mean**
3		0.00	0.00	0.50	0.17		0.00	0.00
4		0.10	0.04	0.50	0.17		0.20	0.24
5		0.20	0.08	0.50	0.17		0.40	0.47
6		0.30	0.11	0.50	0.17		0.60	0.67
7		0.40	0.15	0.50	0.17		0.80	0.86
8		0.50	0.18	0.50	0.17		1.00	1.04
9		0.60	0.20	0.50	0.17		1.20	1.20
10		0.70	0.23	0.50	0.17		1.40	1.36
11		0.80	0.26	0.50	0.17		1.60	1.50
12		0.90	0.28	0.50	0.17		1.80	1.64
13		1.00	0.30	0.50	0.17		2.00	1.77
14								
15	**Mean**	0.50	0.17	0.50	0.17		1.00	0.98
16	**SD**	0.33	0.10	0.00	0.00		0.66	0.58

Figure 3. Now with Columns G and H added which contain the ratio calculation of individual SCR / Mean SCR ($x / \bar{X}$) and done here for both Raw SCR (B) and Logged SCRs (C). So here, each individual SCR has been transformed into a ratio - and thus is a form of standardisation. The values in G and H correspond to the ratio correction suggested by Ben-Shakkar (1987) and is not dependent on any standard deviation calculation thus avoiding the issues raised by Stemmler (1987).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		**Raw SCR**	**Log(SCR+1)**	**Raw SCR mean**	**Log SCR mean**		**Raw SCR / mean**	**Log SCR / mean**		**raw SD**		**z-score RAW**	**T-score raw**
3		0.00	0.00	0.50	0.17		0.00	0.00		0.33		-1.52	34.85
4		0.10	0.04	0.50	0.17		0.20	0.24		0.33		-1.21	37.88
5		0.20	0.08	0.50	0.17		0.40	0.47		0.33		-0.91	40.91
6		0.30	0.11	0.50	0.17		0.60	0.67		0.33		-0.61	43.94
7		0.40	0.15	0.50	0.17		0.80	0.86		0.33		-0.30	46.97
8		0.50	0.18	0.50	0.17		1.00	1.04		0.33		0.00	50.00
9		0.60	0.20	0.50	0.17		1.20	1.20		0.33		0.30	53.03
10		0.70	0.23	0.50	0.17		1.40	1.36		0.33		0.61	56.06
11		0.80	0.26	0.50	0.17		1.60	1.50		0.33		0.91	59.09
12		0.90	0.28	0.50	0.17		1.80	1.64		0.33		1.21	62.12
13		1.00	0.30	0.50	0.17		2.00	1.77		0.33		1.52	65.15
14													
15	**Mean**	0.50	0.17	0.50	0.17		1.00	0.98				0.00	50.00
16	**SD**	0.33	0.10	0.00	0.00		0.66	0.58				1.01	10.05

Figure 4. A worked example of a z-score transformation on the raw SCRs (Column B). Column L contains the z-scores (see equation editor for formula) and Column M contains the T-scores - should those be required.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		Raw SCR	Log(SCR+1)	Raw SCR mean	Log SCR mean		Raw SCR / mean	Log SCR / mean		raw SD	Log SD		z-score (logged data)	T-score
3		0.00	0.00	0.50	0.17		0.00	0.00		0.33	0.1		-1.70	33.00
4		0.10	0.04	0.50	0.17		0.20	0.24		0.33	0.1		-1.29	37.14
5		0.20	0.08	0.50	0.17		0.40	0.47		0.33	0.1		-0.91	40.92
6		0.30	0.11	0.50	0.17		0.60	0.67		0.33	0.1		-0.56	44.39
7		0.40	0.15	0.50	0.17		0.80	0.86		0.33	0.1		-0.24	47.61
8		0.50	0.18	0.50	0.17		1.00	1.04		0.33	0.1		0.06	50.61
9		0.60	0.20	0.50	0.17		1.20	1.20		0.33	0.1		0.34	53.41
10		0.70	0.23	0.50	0.17		1.40	1.36		0.33	0.1		0.60	56.04
11		0.80	0.26	0.50	0.17		1.60	1.50		0.33	0.1		0.85	58.53
12		0.90	0.28	0.50	0.17		1.80	1.64		0.33	0.1		1.09	60.88
13		1.00	0.30	0.50	0.17		2.00	1.77		0.33	0.1		1.31	63.10
14														
15	Mean	0.50	0.17	0.50	0.17		1.00	0.98					-0.04	49.60
16	SD	0.33	0.10	0.00	0.00		0.66	0.58					0.99	9.92
17														

Figure 5. As above, a z-score transformation, but now for the Logged SCR data (Columns C, E and K).

Conclusion

The distributions of SCR data are typically skewed and have a heterogeneous variability across individuals and experimental conditions. As a consequence data transformations are required to normalise these influences for parametric analysis. In addition, for individual differences analysis, SCR reactivity requires a process of standardisation. This facilitates a sound comparison of the truer magnitude of the range of responsiveness, increases power in the analysis, and facilitates accurate correlations in SCR reactivity with other measures (i.e., questionnaires, other psychophysiological measures). Although many debates continue in the literature, contemporary studies are gravitating more towards transformations that use mean reactivity in some manner to transform raw SCRs for standardizations. Two main methods are z-score transformations and ratio transformations and both are methodologically superior to maximal response corrections which have been questioned on a number of grounds (Ben-Shakhar, 1987; 1985; Boucsein, 2012; Boucsein et al., 2012; Braithwaite et al., 2014; Bush et al., 1993; Dawson et al., 2001;).

References & Further Reading

Ben-Shakhar, G (1987). The correction of psychophysiological measures for individual differences in responsivity should be based on typical response parameters: A reply to Stemmler. *Psychophysiology, 24,* (2), 247-249.

Ben-Shakhar, G (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology, 22,* (3), 292-299.

Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., Dawson, M.E., & Filion, D.L (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*, 1017-1034.

Boucsein, W (2012) Electrodermal activity (2nd Ed). New York: Springer.

Braithwaite, J.J. et al (2015). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. Technical Report 2nd version: Selective Attention & Awareness Laboratory (SAAL) Behavioural Brain Sciences Centre, University of Birmingham (UK).

Bush, L.K., Hess, U., & Wolford, G (1993) Transformations for within-subject designs: A Monte-Carlo investigation. *Psychological Bulletin, 113*, (3), 566-579.

Dawson, M.E, et al (2001) The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, and G.B. Bernston, (Eds) Handbook of Psychophysiology (2nd Ed), 200–223. Cambridge Press, Cambridge.

Stemmler, G (1987). Standardization within subjects: A critique of Ben-Shakhar's conclusions. *Psychophysiology, 24*, (2), 243-246.