Robust Estimation of Event-Related Brain Potentials via 20% Trimmed Means

David M. Groppe1, Kevin Heins1, Marta Kutas1,2

1Department of Cognitive Science, 2Department of Neurosciences: University of California, San Diego

Background

QUESTION

Is a popular robust average, the 20% trimmed mean, a more reliable method for deriving ERPs than the conventional arithmetic mean? The 20% trimmed mean is equivalent to the arithmetic mean after discarding the highest 20% and lowest 20% of observations.

PRIOR RESEARCH

The electroencephalogram is most commonly analyzed by averaging epochs of EEG data time locked to some class of event to derive ERPs. This averaging is conventionally done with the arithmetic mean. However, the arithmetic mean can be highly variable and potentially misleading when data come from heavy tailed distributions that frequently produce extreme values or skewed distributions (e.g., lognormal distributions). For such populations, robust measures of central tendency like the 20% trimmed mean are more reliable and possibly more representative than the arithmetic mean (Wilcox, 2005). Some EEG researchers (Rousselet et al., 2008) have suggested that EEG distributions might often be heavy tailed and have advocated the use of the 20% trimmed mean for computing ERPs. However, in a small study of data from one electrode and three individuals, Leonowicz et al. (2005) found that such trimmed measures of central tendency did not generally improve upon the arithmetic mean for individual participant ERPs.

OUR STUDY

Using the 20% trimmed mean, we followed up on Leonowicz and colleagues’ work with a much larger number of data sets (63 data sets from 4 experiments) with greater scalp coverage (28 to 64 electrodes). Furthermore we analyzed the effect of trimmed averaging on grand average ERPs as well as on individual participant ERPs.

EEG Data

For all four experiments, EEG was recorded using a left mastoid reference at a sampling rate of 250 Hz, using an analog filter with a bandpass of 0.1-100 Hz. After recording, data were divided into approximately 1 second peri-stimulus epochs, re-referenced to the mastoid average, and lowpass filtered at 50 Hz. EEG artifacts were removed by discarding polluted epochs or by using independent components analysis. Subsequently, epochs were baseline-corrected by removing the mean EEG from -100 to 0 ms.

Exp. 1: Visual Oddball
8 Participants
20 to 30 Electrodes
6 Stimulus Categories (targets and standards at three levels of difficulty: easy, medium, & hard)

Exp. 2: Visual Oddball
16 Participants
64 Electrodes
4 Stimulus Categories (targets and standards at two levels of difficulty: medium & hard)

Exp. 3: Sentence Comprehension
16 Participants
64 Electrodes
4 Stimulus Categories (grammatical, ungrammatical, congruous, incongruous words)

Exp. 4: Textual Priming of Speech & Tone Counting
23 Participants
31 Electrodes
6 Stimulus Categories: Tones in counting task, Tones added to or replacing part of words that match or mismatch prime

Analysis

Reliability of Estimated Means:

For the arithmetic mean, 95% confidence interval ranges of 0.5 microvolts. ERP variability always tends to be smaller for the arithmetic mean. Red indicates significant differences (p<0.05).

Tables A & B: 95% confidence interval ranges from the arithmetic mean minus those from the 20% trimmed mean in ERP peak time windows. Differences are in microvolts. ERP variability always tends to be smaller for the arithmetic mean. Red indicates significant differences (p<0.05).

C: All Time Points: Exp 4

Summary & Conclusions

1. The 20% trimmed mean produces clearly less reliable ERPs for individual participants than the arithmetic mean, though the difference is small (95% confidence intervals differ around 0.15 microvolts). This is not surprising given that conventional artifact correction procedures trim extreme outliers.

2. The 20% trimmed mean sometimes produces more reliable grand average ERPs, though generally the arithmetic mean still performs better. Indeed, when averaging across ERP peak time windows, the arithmetic mean always tended to perform better. However, the difference between the two methods is still rather small (95% confidence intervals differed around 0.20 microvolts).

Contact: dgruppe@cogsci.ucsd.edu