

White Paper: October 2009

Title: Sample Size Analysis for Brainwave Collection (EEG) Methodologies

Author: Stephen F. Sands, Ph.D.

Summary

A common question in behavioral- and neuro-marketing research is, "what is the appropriate number of subjects needed to obtain a reliable result?" Traditional methods of market research use large numbers of respondents, and there seems to be general consensus that approximately 150-200 participants or more (depending on research objectives) are needed to obtain consistent results. With electroencephalogram methodologies (EEG) a much smaller sample size is needed to achieve a similar statistical threshold.

When the number of study participants is between 30 to 40 (per target demographic grouping), there is a less than 1% chance of error, and the associated Neuro-Engagement Factor™ (NEF) score portrays an accurate and significant rating for the media stimulus in question. Sands Research utilizes the less than 1% chance of error threshold for all studies. A larger sample size could be utilized to achieve an even smaller margin of error, say .25%, although that degree of threshold does not provide us with a significant amount of 'new' knowledge about the stimulus, nor is it financially efficient.

Below you'll find a detailed explanation and a sample study that illustrates our findings.

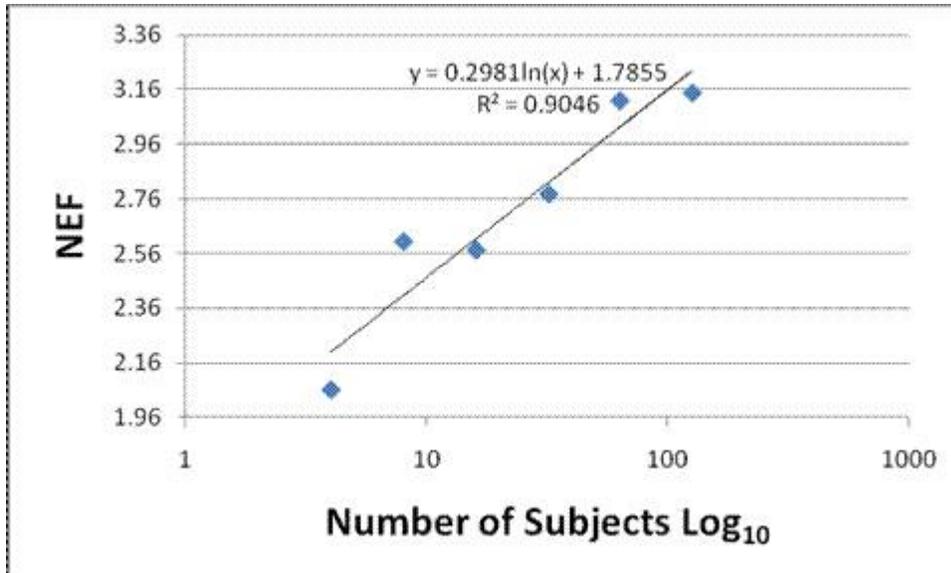
Research & Analysis

Perhaps a more formal way to frame this question is to place it in the context of statistical power analysis (Cohen, 1988). Given an acceptable statistical threshold such as 95% likelihood of being correct, we are really asking, "how many participants do we need to reach this threshold?" Power analysis is a formal method used to answer this question. A test's "power" is the ability to correctly detect an effect. In statistical terms this translates into the ability of a test to correctly reject the null hypothesis (1-b).

Currently, new biological measures are appearing in the field of market research. Often, the number of respondents employed can be an order of magnitude less than the familiar numbers employed in behavioral research. Due to the increased sensitivity of these EEG measures, it is argued that fewer participants are needed. These numbers are derived in the same consensual manner and have a history associated with levels needed to reach accepted statistical significance ($p < .05$). Traditional behavioral market researchers have viewed these participant levels with skepticism.

Although rarely performed, a power analysis is a simple way to answer these questions. Consequently, we have performed this analysis with our measure, the Neuro-Engagement-Factor™ (NEF), score. The NEF score is essentially a "Z" score derived from brain electrical activity. The more electrical activity detected, the higher the score. This measure is normalized against a baseline estimate of the brain background noise. The distribution of the NEF score is normally distributed allowing us to proceed with a traditional power analysis.

Shown below is an experiment in which a group of subjects have viewed a 30-second television commercial. The total number of respondents is 126. From this pool of respondents we perform a series of split-half segmentations. Each sample pool is split in half until the smallest subject pool size of 4 is reached.



The resulting function is then fitted with a power function ($R^2 = 0.9$ indicating an excellent fit). These data are best fit by a power function and indicate that it is possible to detect statistically significant results with small numbers of respondents. The example above is chosen to represent a simple binary choice between two conditions or the paired t-test scenario. The number of subjects to reach significance for this TV commercial would have been approximately 4 at $p < .05$ (NEF score of 1.96). This is the bare minimum to observe an effect. In our research we set our criterion at $p < .01$ or a 1% chance of an error. This translates to a NEF score of 2.76, requiring 32 subjects.

These findings show that approximately 30-40 subjects are required to achieve significance at $p < .01$.

Conclusion

Within our own normative database (all normalized at a less than 1% chance of error), we see the major findings change very little past our 30-40 sample size.

In short, Sands Research has isolated a very specific, scientifically-validated sample size that corresponds to a 99% significance threshold, or less than 1% chance of error.

Citation: Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.