For this statistic, we do not possess an obvious formula to decide whether the resulting statistic is significant (we cannot rely on the CLT, for example). To determine significance, we employ a bootstrapping procedure (also known as a permutation test), which we can perform even for the relatively complicated statistic. In this way, we may devise complicated measures of data and construct error bars or compute statistical significance, provided our computational resources are sufficient.

Summary

In this chapter, we considered scalp EEG data recorded from a single electrode during an auditory task. The task consisted of two conditions, and we sought to uncover the difference in the EEG responses between the two conditions. We began with a visual inspection of the EEG recordings from individual trials and from all trials, and concluded that the data were quite noisy; any evoked response due to the stimulus was not obvious in the single-trial data. To emphasize the evoked signal, we computed the ERP, which involved averaging the EEG signal across trials. By doing so, we uncovered interesting structure in condition A, but not much in condition B. We then developed two techniques to add error bars to an ERP. One technique relied on the central limit theorem, and the other technique involved a computationally expensive bootstrapping procedure. Both techniques suggested that the ERP in condition A differed significantly from zero following the stimulus at time 0.25 s. Finally, we assessed whether the two ERPs differed. We did so through visual inspection, by comparing the differences in the ERPs, and by computing a statistic and assessing its significance through a bootstrapping procedure. Using the last procedure, we concluded that the ERP in the two conditions significantly differed.

Problems

Please choose 3 questions below to answer. Due Oct 15, 2019.

2.1. Consider an alternative statistic to assess the difference in the ERPs between conditions A and B: the integrated area of the squared difference between the two ERPs. Compute this statistic for the data, and use a bootstrapping procedure to test the null hypothesis of no difference in this statistic between the two conditions.

2.2. Load the file `Ch2-EEG-2.mat`, available at

`http://github.com/Mark-Kramer/Case-Studies-Kramer-Eden`

into MATLAB. You'll find two variables:
`EEG` = the EEG data, consisting of 1,000 trials, each observed for 1 s;
`t` = the time axis, which ranges from 0 s to 1 s.
These data have a similar structure to the data studied in this chapter. To collect these data, a stimulus was presented to the subject at 0.25 s. Analyze these data for the presence of an evoked response. To do so, answer the following questions.

a. What is the time between samples (`dt`) in seconds?

b. Examine some individual trials of these data. Explain what you observe in pictures and words. From your visual inspection, do you expect to find an ERP in these data?

c. Compute the ERP for these data, and plot the results. Do you observe an ERP (i.e., times at which the 95% confidence intervals do not include zero)? Include 95% confidence intervals in your ERP plot, and label the axes. Explain in a few sentences the results of your analysis, as you would to a collaborator who collected these data.

2.3. Load the file `Ch2-EEG-3.mat`, available at

`http://github.com/Mark-Kramer/Case-Studies-Kramer-Eden`

into MATLAB. You'll find two variables:
`EEG` = the EEG data, consisting of 1,000 trials, each observed for 1 s;
`t` = the time axis, which ranges from 0 s to 1 s.
These data have a similar structure to the data studied in this chapter. To collect these data, a stimulus was presented to the subject at 0.25 s. Analyze these data for the presence of an evoked response. To do so, answer the following questions.

a. What is the time between samples (`dt`) in seconds?

b. Examine some individual trials of these data. Explain what you observe in pictures and words. From your visual inspection, do you expect to find an ERP in these data?

c. Compute the ERP for these data, and plot the results. Do you observe an ERP (i.e., times at which the 95% confidence intervals do not include zero)? Include 95% confidence intervals in your ERP plot, and label the axes. Explain in a few sentences the results of your analysis, as you would to a collaborator who collected these data.

2.4. Load the file `Ch2-EEG-4.mat`, available at

`http://github.com/Mark-Kramer/Case-Studies-Kramer-Eden`

into MATLAB. You'll find two variables:
`EEG` = the EEG data, consisting of 1,000 trials, each observed for 1 s;
`t` = the time axis, which ranges from 0 s to 1 s.
These data have a similar structure to the data studied in this chapter. To collect these data, a stimulus was presented to the subject at 0.25 s. Analyze these data for the presence of an evoked response. To do so, answer the following questions.

a. What is the time between samples (`dt`) in seconds?

b. Examine some individual trials of these data. Explain what you observe in pictures and words. From your visual inspection, do you expect to find an ERP in these data?

c. Compute the ERP for these data, and plot the results. Do you observe an ERP (i.e., times at which the 95% confidence intervals do not include zero)? Include 95% confidence intervals in your ERP plot, and label the axes. Explain in a few sentences the results of your analysis, as you would to a collaborator who collected these data.

2.5. In the previous problem, you considered the dataset `Ch2-EEG-4.mat` and analyzed these data for the presence of an ERP. To do so, you averaged the EEG data across trials. The results may have surprised you. Modify your analysis of these data in some way to better illustrate the appearance (or lack thereof) of an evoked response. Explain what's happening in these data, as you would to a colleague or experimental collaborator.

2.6. Load the file `Ch2-EEG-5.mat`, available at

```
http://github.com/Mark-Kramer/Case-Studies-Kramer-Eden
```

into MATLAB. You'll find two variables:
`EEG` = the EEG data, consisting of 1,000 trials, each observed for 1 s.
`t` = the time axis, which ranges from 0 s to 1 s.
These data have a similar structure to the data studied in this chapter. To collect these data, a stimulus was presented to the subject at 0.25 s. Analyze these data for the presence of an evoked response. To do so, answer the following questions.

a. What is the time between samples (`dt`) in seconds?

b. Examine some individual trials of these data. Explain what you observe in pictures and words. From your visual inspection, do you expect to find an ERP in these data?

c. Compute the ERP for these data, and plot the results. Do you observe an ERP (i.e., times at which the 95% confidence intervals do not include zero)? Include 95% confidence intervals in your ERP plot, and label the axes. Explain in a few sentences the results of your analysis, as you would to a collaborator who collected these data.

2.7. Compare the datasets `Ch2-EEG-3.mat` and `Ch2-EEG-4.mat` studied in the previous problems. Use a bootstrap procedure to test the hypothesis that the evoked response is significantly different in the two datasets.

2.8. The goal of this problem is to explore the central limit theorem. We informally defined the CLT in this chapter. Let's perform a numerical experiment to test it.

a. Generate a large set of independent random numbers. These are many options to choose from. For example, use MATLAB to produce 500 uniformly distributed random numbers:
```
R = rand(500,1);
```
Describe this list of random numbers. Include a histogram (with axes labeled).

b. Compute the mean of this list of numbers. What value do you find for the mean? Does the mean value make sense when compared to the histogram you created?

c. Generate another instance of the independent random numbers you chose in part (a) and compute its mean, Repeat 5,000 times, saving the result each time. In the end, you should have 5,000 values, each corresponding to a mean of a different realization of the variable `R`. Describe the distribution of these mean values of `R`. What do you find? Is your result consistent with the CLT? (*Hint*: Make a histogram and consider its shape . . .)

d. Repeat your analysis with a different choice for the independent random numbers. Note that MATLAB provides many different options. For your new choice of random numbers, describe a single instance of 500 values, as in part (a), and then describe the distribution of the mean of these random numbers, as in part (c).

Appendix: Standard Error of the Mean

Assume the values,

$$x = \{x_1, x_2, x_3, \ldots, x_K\},$$

are independent data drawn from a distribution with a theoretical mean μ and theoretical variance σ^2. The sample mean of x is

$$\bar{x} = \frac{1}{K} \sum_{k=1}^{K} x_k \,.$$

Since each of the values x_k is a sample of a random variable with a probability distribution, so is the sample mean. We are interested in computing the properties of the distribution of the sample mean of the original data.

First, we compute the expected value, or theoretical mean, of the sample mean,

$$E[\bar{x}] = \frac{1}{K} \sum_{k=1}^{K} E[x_k] = \frac{K\mu}{K} = \mu \,.$$