STAT 847: Midterm Project

DUE: Wednesday, March 6 2024 by 11:59pm EST

NOTES

Your midterm project must be submitted by the due date listed at the top of this document, and it must be submitted electronically in .pdf format via Crowdmark.

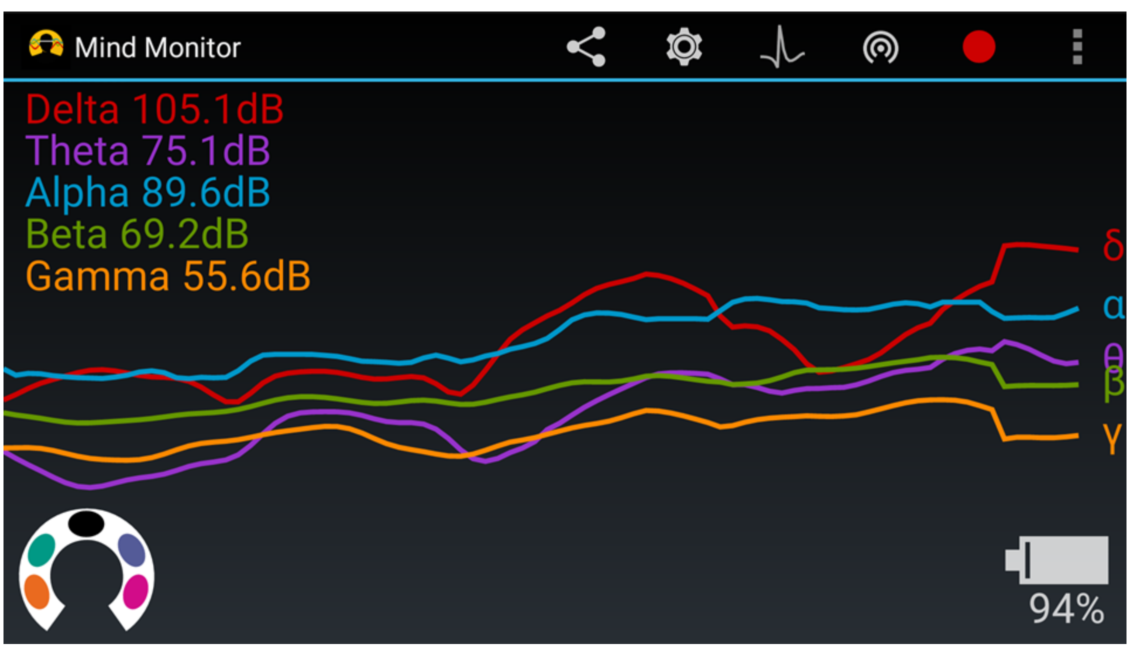
NOTICE THAT THE DUE DATE IS MARCH 6th, NOT FEB 28th.

Organization and comprehensibility is part of a full solution. Consequently, points will be deducted for solutions that are not organized and incomprehensible. Furthermore, if you submit your assignment to Crowdmark, but you do so incorrectly in any way (e.g., you upload your Question 2 solution in the Question 1 box), you will receive a 5% deduction (i.e., 5% of the assignment's point total will be deducted from your point total).

There are a total of 100 points possible.

For this project, you have the data from a collection of brainwave reading sessions that were collected from an informed, consenting participant.

These were collected by using a Muse 2 headband, and the Mind Monitor phone app. Both pictured below.



Several times per second. This headband collects electrical signal information and translates it into strength of different brainwaves at four locations (TP9, TP10 by the ears, and AF7, AF8 on the forehead). It also records head movements through an accelerometer and a gyroscope.

These headbands are usually used for guided meditation by playing one sound when the user's brainwaves are in a state of deep relaxation, and another sound when they are far from that state. In this case, the study participant wore the headband while working at a computer (active), napping (resting), or doing other activities like watching tv. **You can consider anything not resting to be an active state**

The detailed dataset has the measurements processed several times per second. The variables in the detailed dataset are as follows are as follows.

Variable Name	Description
See:	https://mind-monitor.com/Technical_Manual.php
TP9	Left Temporal-Parietal (Ear) sensor
AF7	Left Frontal-Prefrontal (Forehead) sensor
AF8	Right Frontal-Prefrontal (Forehead) sensor
TP10	Right Temporal-Parietal (Ear) sensor
Delta	1-4 Hz brainwave activity, absolute band power
Theta	4-8 Hz brainwave activity, absolute band power
Alpha	7.5-13 Hz brainwave activity, absolute band power
Beta	13-30 Hz brainwave activity, absolute band power
Gamma	30-44 Hz brainwave activity, absolute band power
RAW	Raw EEG sensor values in microvolts
AUX_RIGHT	Raw EEG sensor values in microvolts
Accelerometer_X	head tilt up/down in g
Accelerometer_Y	head tilt left/right in g
Accelerometer_Z	vertical motion in g
Gyro_X	change in rotation tilt left/right (roll)
Gyro_Y	change in rotation tilt up/down (pitch)
Gyro_Z	change in looking left/right (yaw)
HeadBandOn	1 for yes, 0 for no
HSI	horse shoe indicator (1 = strong connect, 2 = weak, 4 = none)
Battery	0-100 rounded to nearest 5
Elements	markers for blinks, jaw clenches, and disconnections
Sessionnum	Unique ID for session number
activity	Main activity. 'active' means various work and social things.
session_time	The total length of the session
time_in	Amount of time into the session of this measurement
Nblinks	total blinks in the this session
Njaw	total jaw clenches in this session
total_alpha	Alpha_TP9 + Alpha_TP10 + Alpha_AF7 + Alpha_AF8

See also: https://en.wikipedia.org/wiki/10%E2%80%9320_system_(EEG) for sensor location info.

There is also a summary dataset, made with the following code.

Get all the summary stuff

```
library(plyr)

dat_all = read.csv("Mind Monitor detailed data 2024-01-21.csv")

dat_summary = ddply(dat_all, "sessionnum", summarise,
```

```
      activity = activity[1],
      session_time = session_time[1],
      mean_alpha = mean(total_alpha, na.rm=TRUE),
      mean_beta = mean(total_beta, na.rm=TRUE),
      mean_gamma = mean(total_gamma, na.rm=TRUE),
      mean_delta = mean(total_delta, na.rm=TRUE),
      mean_theta = mean(total_theta, na.rm=TRUE),

      var_alpha = var(total_alpha, na.rm=TRUE),
      var_beta =  var(total_beta, na.rm=TRUE),
      var_gamma = var(total_gamma, na.rm=TRUE),
      var_delta = var(total_delta, na.rm=TRUE),
      var_theta = var(total_theta, na.rm=TRUE),

      blinks_minute = Nblinks[1]/session_time[1]*60,
      jaws_minute = Njaw[1]/session_time[1]*60,
      mean_pos_xy = mean(sqrt(Accelerometer_X^2 + Accelerometer_Y^2), na.rm=TRUE),
      mad_accel = mean(abs(Accelerometer_X^2 + Accelerometer_Y^2 + Accelerometer_Z^2 - 1), na.rm=TRUE),
      rmse_gyro = mean(sqrt(Gyro_X^2 + Gyro_Y^2 + Gyro_Z^2), na.rm=TRUE)
)
```

The summary dataset condenses each 20-80 minute session into a single row of summary statistics. The summary dataset has the following data dictionary.

Variable	Description
sessionnum	unique ID for that session
activity	main activity taken during session ‘active’ means various work or social things
session_time	length of session in seconds
mean_alpha	alpha wave activity, totaled across all sensors, average over time
var_alpha	alpha wave activity, totaled across all sensors, variance over time
blinks_minute	recorded blink events per minute
jaws_minute	recorded jaw clench events per minute
mean_pos_xy	horitzonalness of head position, averaged over time. 0 = upright, 1 = laying
mad_accel	mean absolute deviation from zero acceleration, in g. Higher = more movement
rmse_gyro	variance in gyroscope movements, in degrees/sec. Higher = more rotation

1. (20 points) Use the elements column of the **detailed dataset** to infer about the differences in blinking and jaw clenching between resting and non-resting states. Specifically, find differences in the distribution of the number of blinks and jaw clenches per minute during resting and active sessions.

2.(15 points) Use the data in the **summary dataset** to infer about the differences in brainwave activity between resting and non-resting states. Specifically, conduct a t-test of each of the ten brainwave variables to find any statistically significant differences. To get the p-value from a test, use `t.test(...)$p.value`.

3. (10 points) Are there any notable differences between the first 20% and the last 20% (by time_in) of active sessions? Look in the **detailed dataset** and isolate the first active session to find out. (The answer may be 'no'. In which case, mention some similarities). Compare at least three variables.
4. (10 points) Are there any notable differences between the first 20% and the last 20% (by time_in) of resting sessions? Look in the **detailed dataset** and isolate the first resting session to find out. (The answer may be 'no'. In which case, mention some similarities). Compare at least three variables.
5. (5 points) Use the **detailed dataset** Make a variable in the summary data that gets the Pearson correlation between Alpha_TP9 and Alpha_TP10 during a session. (That is, the alpha waves above the left and right ear, respectively). Compare the TP9-TP10 correlations during in resting and non-resting activities.
6. (5 points) The accelerometer always shows some acceleration because of gravity. (When completely upright and still, it shows accelerometer Z = 1, X = 0, Y = 0) You can use this to determine the relative position of the head, as well as how much the head moves. Use the acceleration data in the **detailed dataset** to infer about the differences in both the position of the head, and its movements, between resting and non-resting states.

e.g., you final answer should be something like "Head position was more ___________ when resting than when active, and we know this because the accelerometer ____________ instead of _______________", with code and supporting summary information.

7. (10 points) In the **detailed dataset**, plot the smoothed brainwaves for all five bands (alpha, beta, gamma, delta, and theta) session number 11 (resting, 3222 seconds long). Use any smoothing technique we discussed in class. Make a guess at when the patient fell asleep.

8.(10 points) Note that total_alpha and its similar functions have a flaw - measurements that were missing because of a bad connection are marked as zero, while true values are rarely zero. Find another measure using the **detailed dataset**, mean_alpha, that takes the average of **all the non-zero alpha measurements**. Take the mean of this new measurement for each session. For example, the average of mean_alpha for all of session 1 as a single value, all of session 2 as a single value, and so on.

9.(15 points) Find two interesting patterns in the data. These could be almost anything. You can use **either the detailed or the summary datasets**.