DATA ANALYTICS & MACHINE LEARNING

PREMANANDA INDIC, PH.D.

DEPARTMENT OF ELECTRICAL ENGINEERING

The University of Texas at TYLER
Center for Health Informatics & Analytics

ORS Research Design & Data Analysis Lab
Office of Research and Scholarship

WORKSHOP SCHEDULE

- WEEK1: DATA ANALYTICS
- WEEK2: FEATURE EXTRACTION
- WEEK3: MACHINE LEARNING

PREREQUISITE

- NO KNOWLEDGE OF PROGRAMMING
- NO KNOWLEDGE OF ANY QUANTITATIVE METHODS
- INTEREST IN RESEARCH (save $$$???, Predict Outcomes)

WHY MACHINE LEARNING ?

- An Ounce of Prevention Better Than a Pound of Cure

  – Benjamin Franklin (1730-1774)

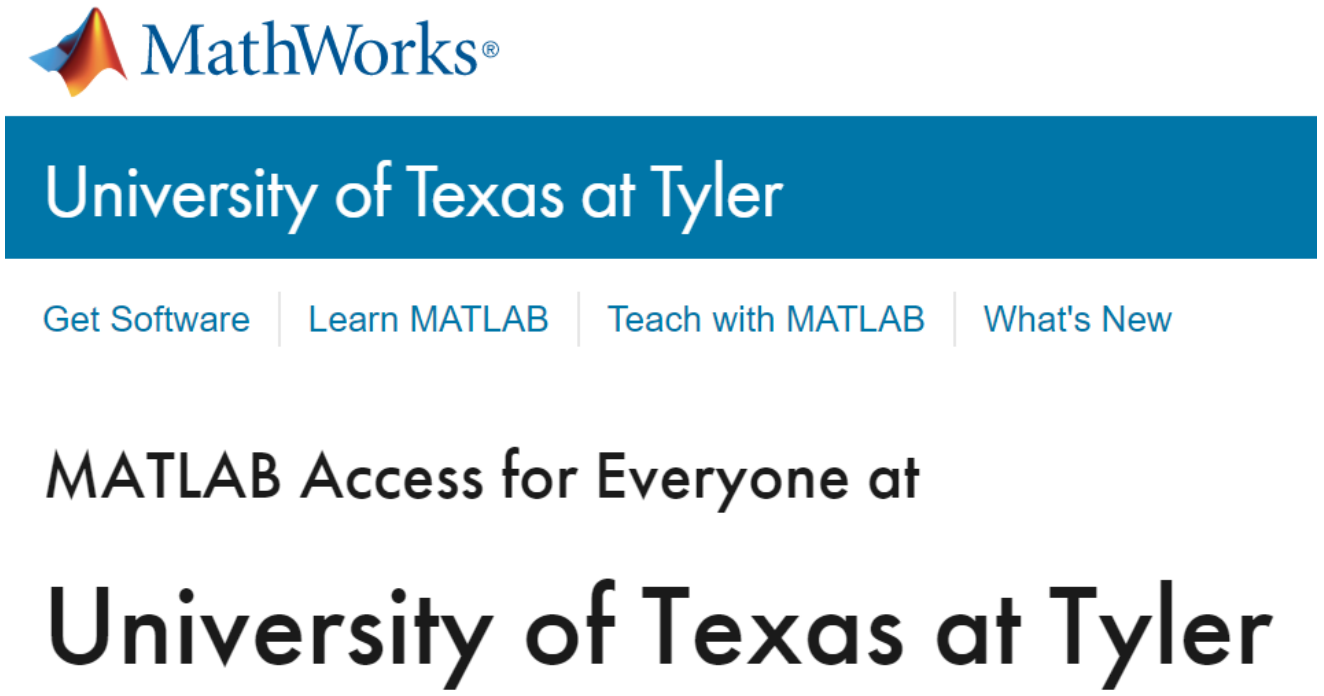
- Prevention Better Than Cure

  – Dutch Philosopher Desiderius Erasmus (1500)

- Precaution Better than Cure

  – Johann Wolfgang von Goethe (1749-1832)

ANALYSIS PLATFORM

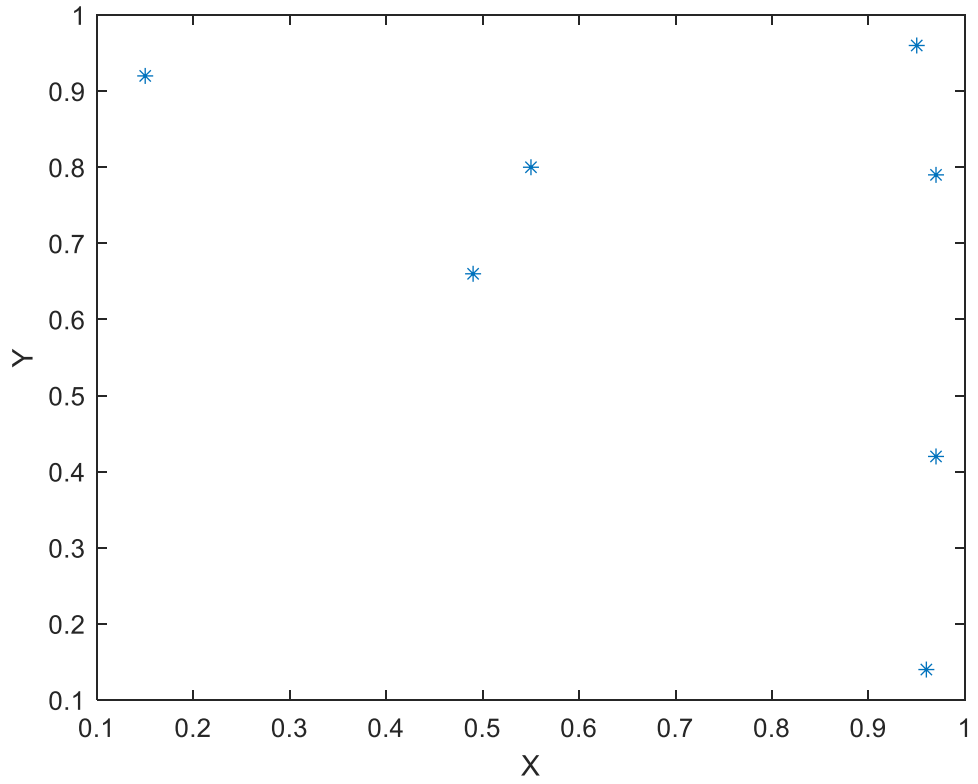


https://www.mathworks.com/academia/tah-portal/university-of-texas-at-tyler-1108545.html

Exercise 1

Plot X and Y in MATLAB

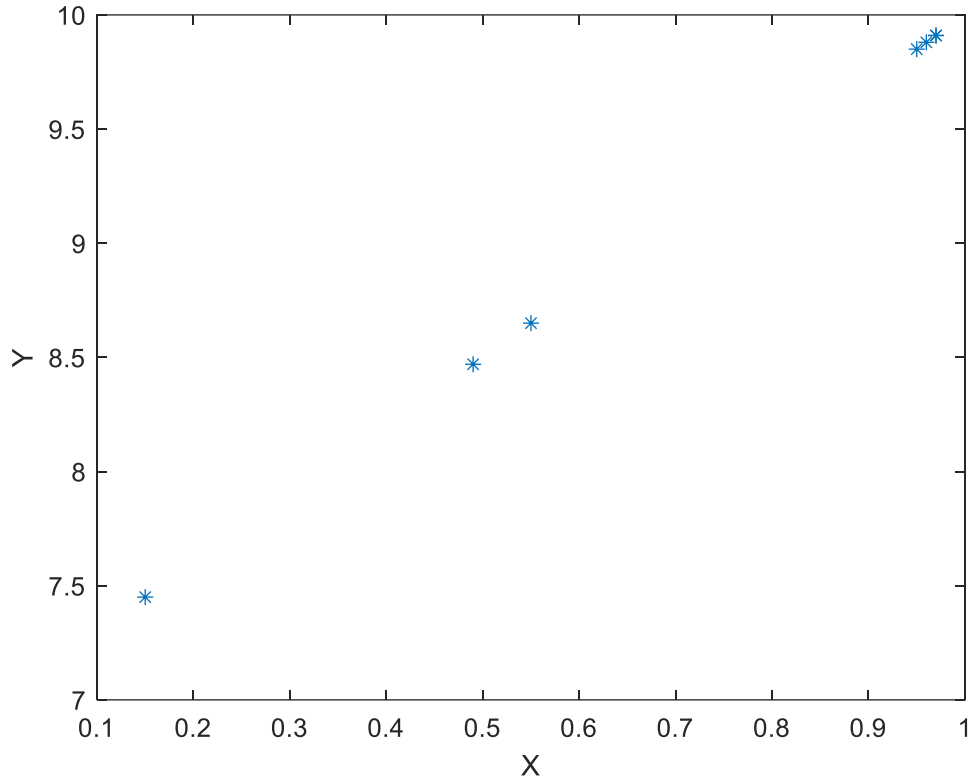
X	Y
0.55	0.80
0.96	0.14
0.97	0.42
0.15	0.92
0.97	0.79
0.95	0.96
0.49	0.66



Exercise 2

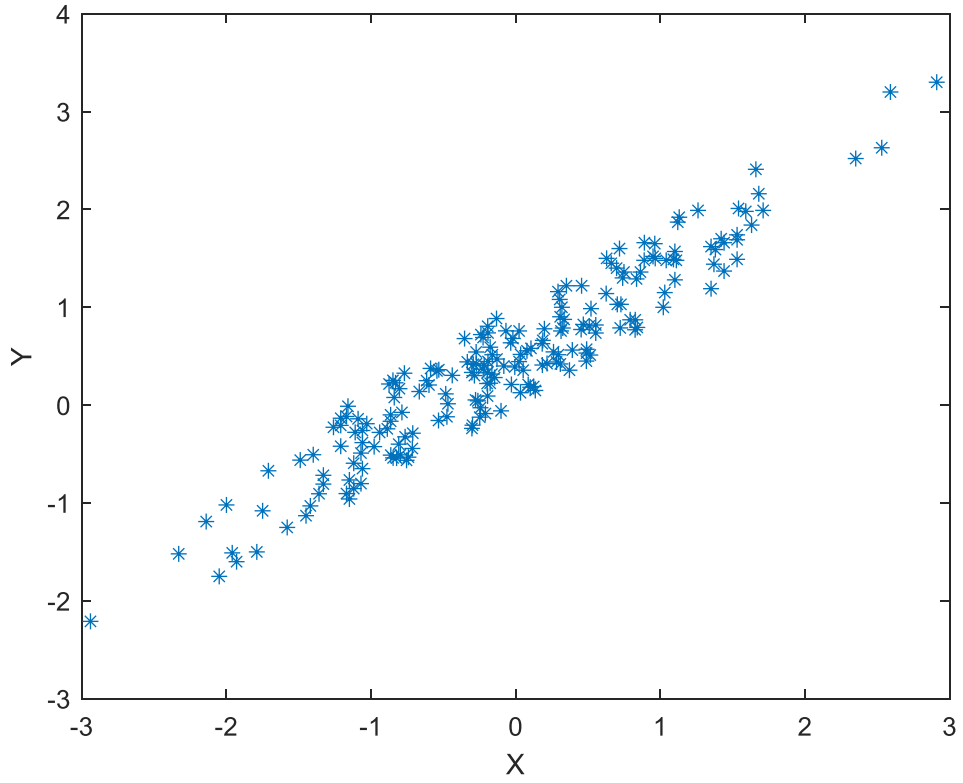
Plot X and Y in MATLAB

X	$\boldsymbol{Y = 3X + 7}$
0.55	
0.96	
0.97	
0.15	
0.97	
0.95	
0.49	



Exercise 3

Plot X and Y given in file dataA.xlsx in MATLAB



(BIG) DATA ANALYTICS

- OLD CONCEPT IN A NEW PHRASE

- HIGH PERFORMANCE COMPUTERS, HIGH RESOLUTION DATA, LARGE STORAGE CAPABILITY…….

PUBLICLY AVAILABLE DATA SET

- PHYSIOBANK : https://www.physionet.org/data/

- MIMICS II : https://mimic.physionet.org/

IDENTIFY THE IMAGES

IDENTIFY THE IMAGES

Tools

Items

What is missing ?

HYPOTHESIS

Scientific hypothesis, an idea that proposes a tentative explanation about a phenomenon or a narrow set of phenomena observed in the natural world. The two primary features of a scientific hypothesis are falsifiability and testability

Source: https://www.britannica.com/science/scientific-hypothesis

BIOMEDICAL DATA

- LINEAR VS NONLINEAR
- DETERMINISTIC VS STOCHASTIC
- STATIONARY VS NONSTATIONARY

Biomedical data are nonlinear, nonstationary and deterministic / stochastic in nature

Analytical tools are applicable only for linear, deterministic/stochastic and stationary

TYPES OF BIOMEDICAL DATA

- PHYSIOLOGICAL OR BEHAVIOURAL SIGNALS
- IMAGES
- GENOMES

TYPES OF BIOMEDICAL DATA

- PHYSIOLOGICAL OR BEHAVIOURAL SIGNALS
- IMAGES
- GENOMES

TYPES OF SIGNALS

- DISCRETE VS CONTINOUS

- EXAMPLES:

Blood Pressure, Heart Rate, Pulse Rate, SpO2, electrocardiogram, electroencephalogram.

GOALS OF ANALYSIS

- DIAGNOSTICS
- PREDICTION
- UNDERSTANDING FUNDAMENTAL PHYSIOLOGICAL MECHANISMS

TYPES OF ANALYSIS

- STATISTICAL (Mean, Variance, Skewness, Kurtosis)
- SPECTRAL (Amplitude, Frequency, Power)

TYPES OF ANALYSIS

- STATISTICAL (Mean, Variance, Skewness, Kurtosis)
- SPECTRAL (Amplitude, Frequency, Power)

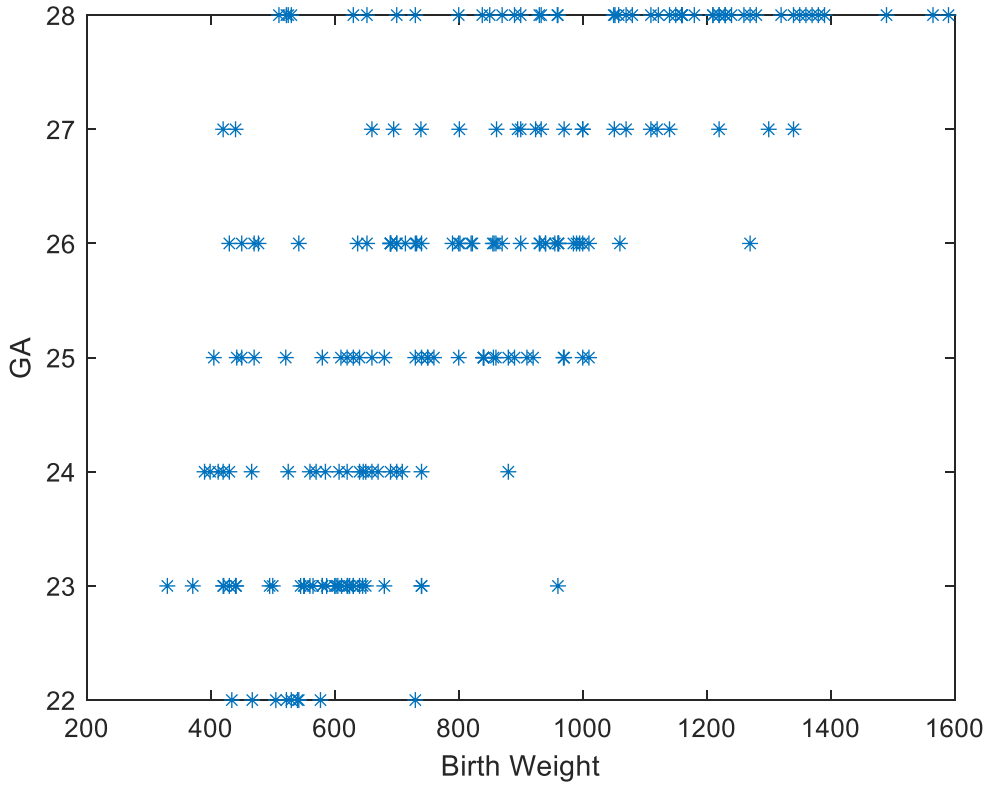
Exercise 4

To test the hypothesis that the birth weight of preterm infants associated with the gestational age.

Methods

Results

Conclusions



Correlation $r = 0.69$ $p < 0.05$

Exercise 5

To test the hypothesis that the more male infants are born premature than the female infants.

Methods

Results

Conclusions

Exercise 6

To test the hypothesis that the level of prematurity is more among male infants than the female infants.

Methods

Results

Conclusions

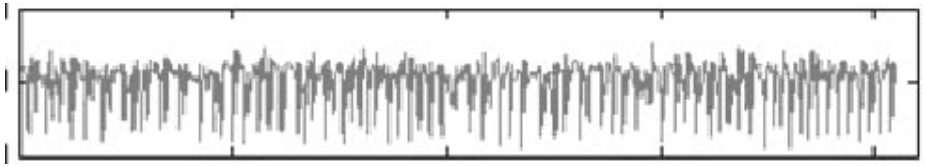
TYPES OF ANALYSIS

- STATISTICAL (Mean, Variance, Skewness, Kurtosis)
- SPECTRAL (Amplitude, Frequency, Power)

SIGNALS

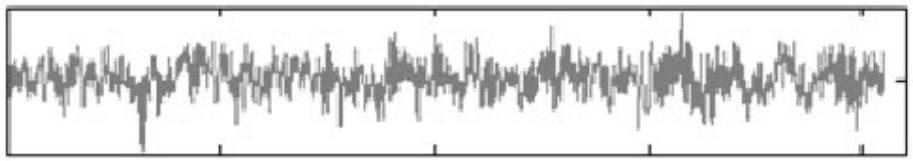
- Which of the given signal is regular ?
- Which of the given signal is normal?

(A)



EEG

(B)



SIGNALS

- Which of the given signal is regular ?

- Which of the given signal is normal?

(A)

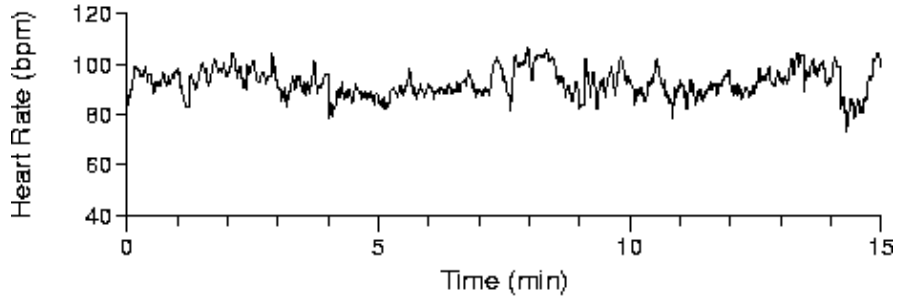
Respiration

(B)

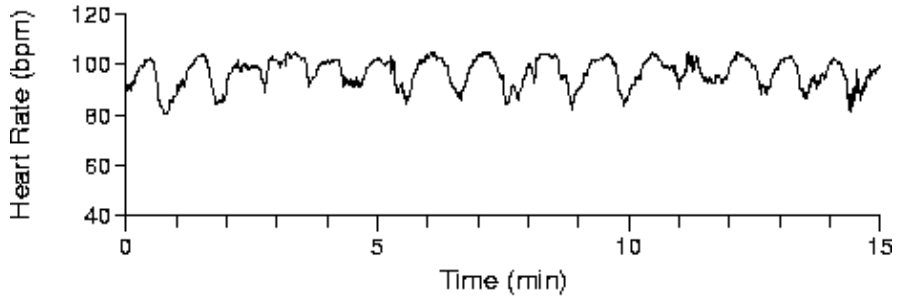
SIGNALS

- Which of the given signal is regular ?
- Which of the given signal is normal?

(A)



(B)



Heart Rate

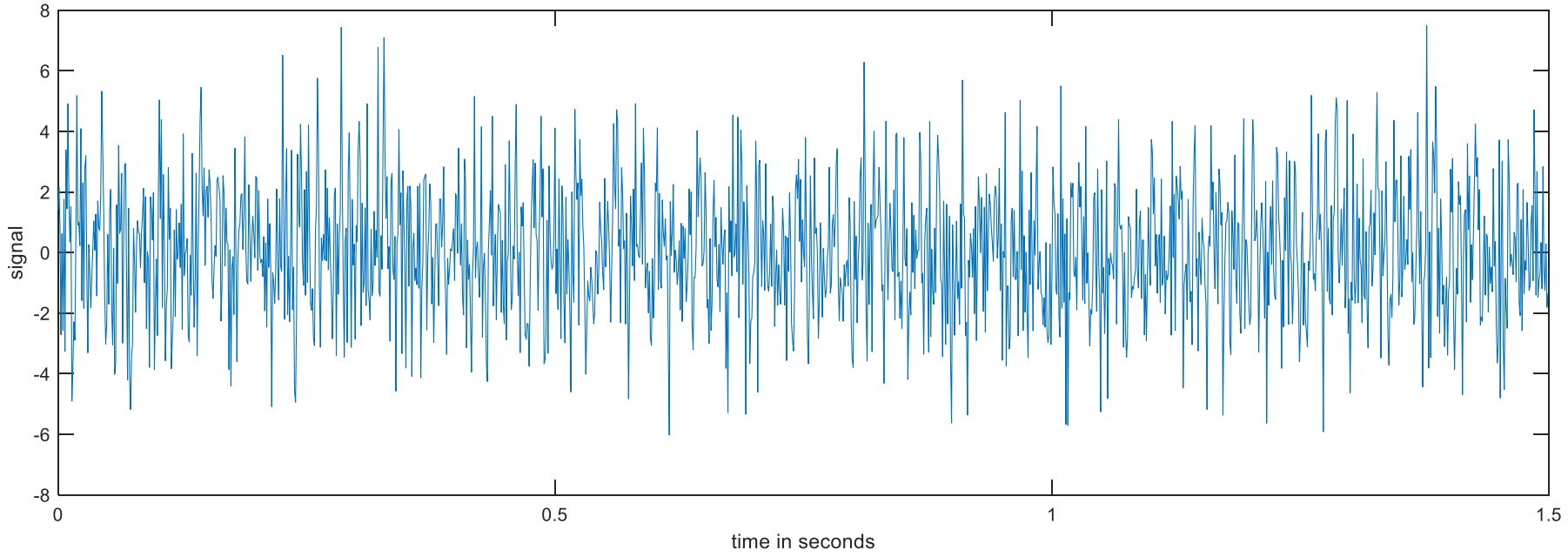
Exercise 7

To test the hypothesis that the given signal has a an unique frequency.

Methods

Results

Conclusions



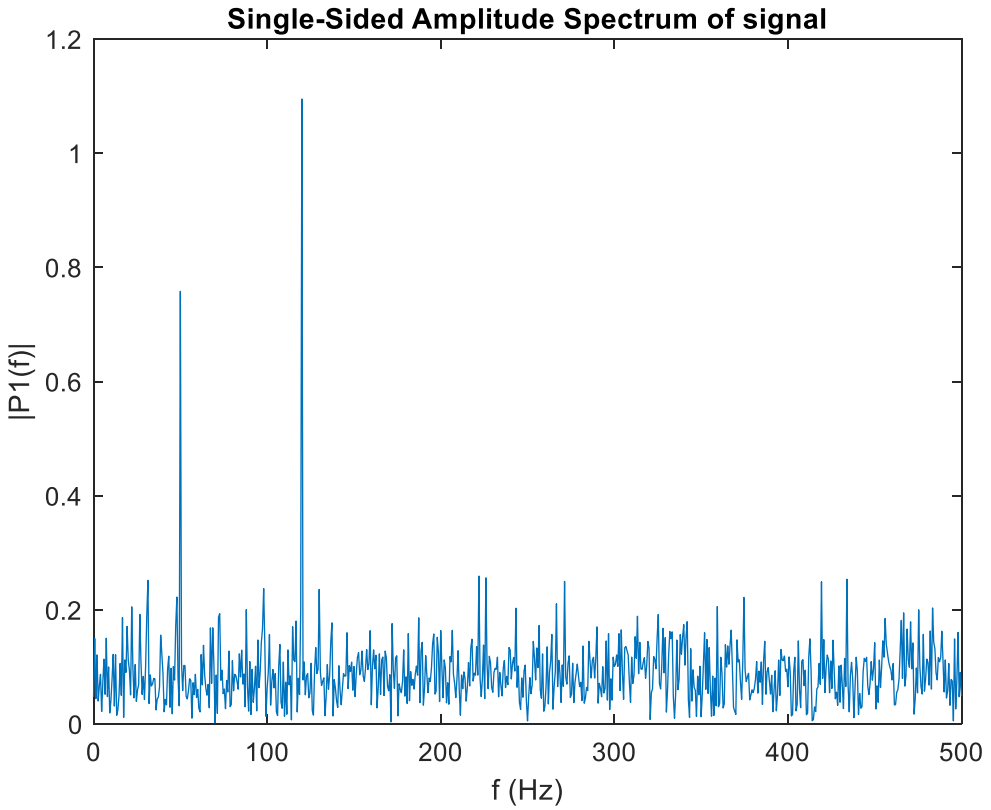
Exercise 7

To test the hypothesis that the given signal has an unique frequency.

Methods

Results

Conclusions



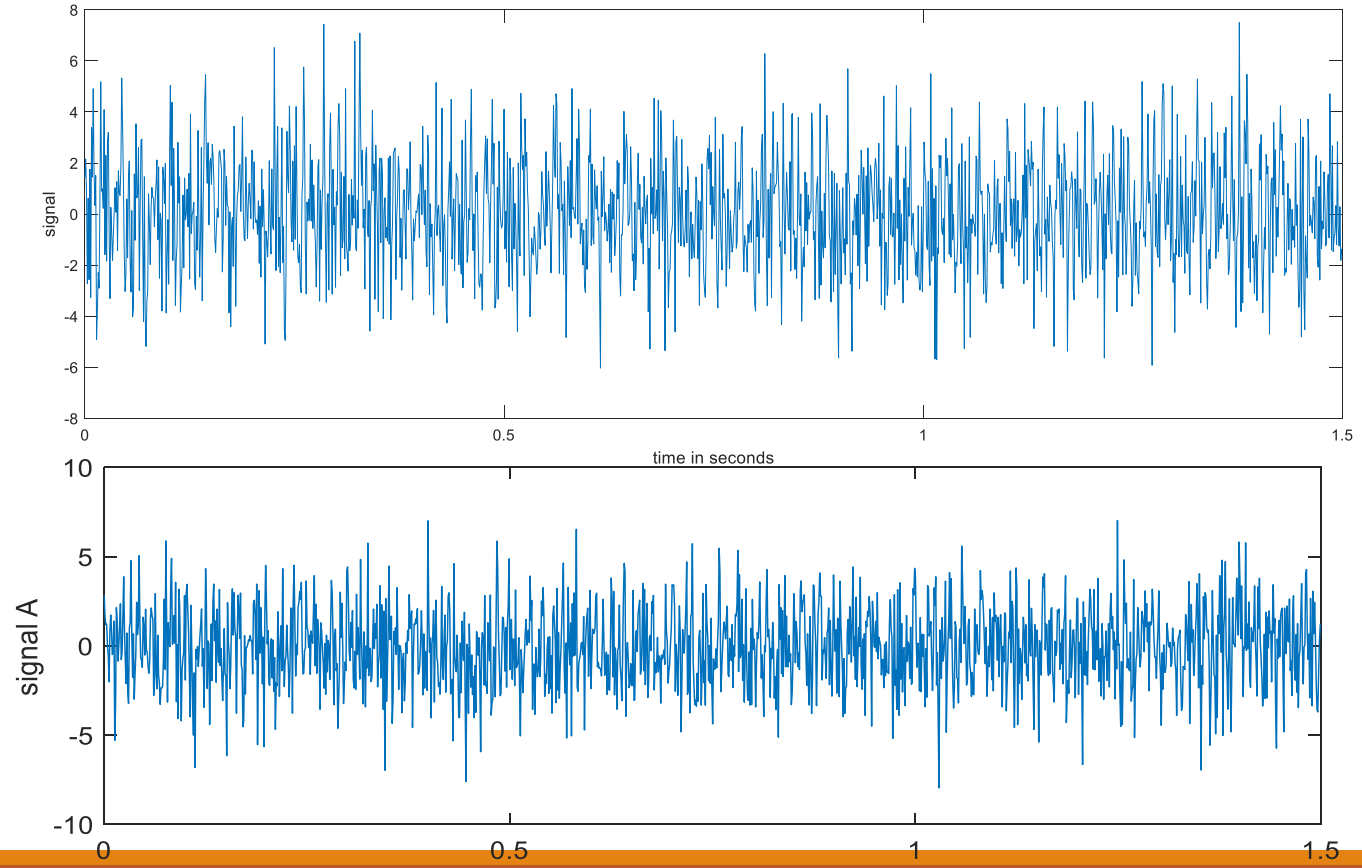
Exercise 8

To test the hypothesis that the previous signal has a relationship with signal A.

Methods

Results

Conclusions



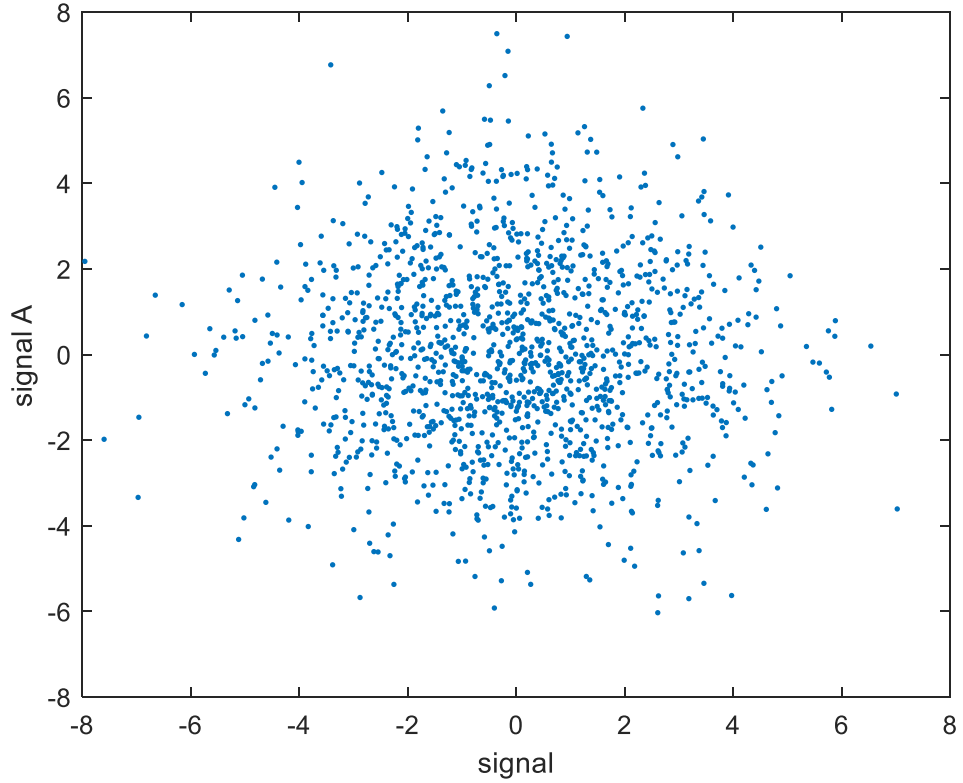
Exercise 8

To test the hypothesis that the previous signal has a relationship with signal A.

Methods

Results

Conclusions



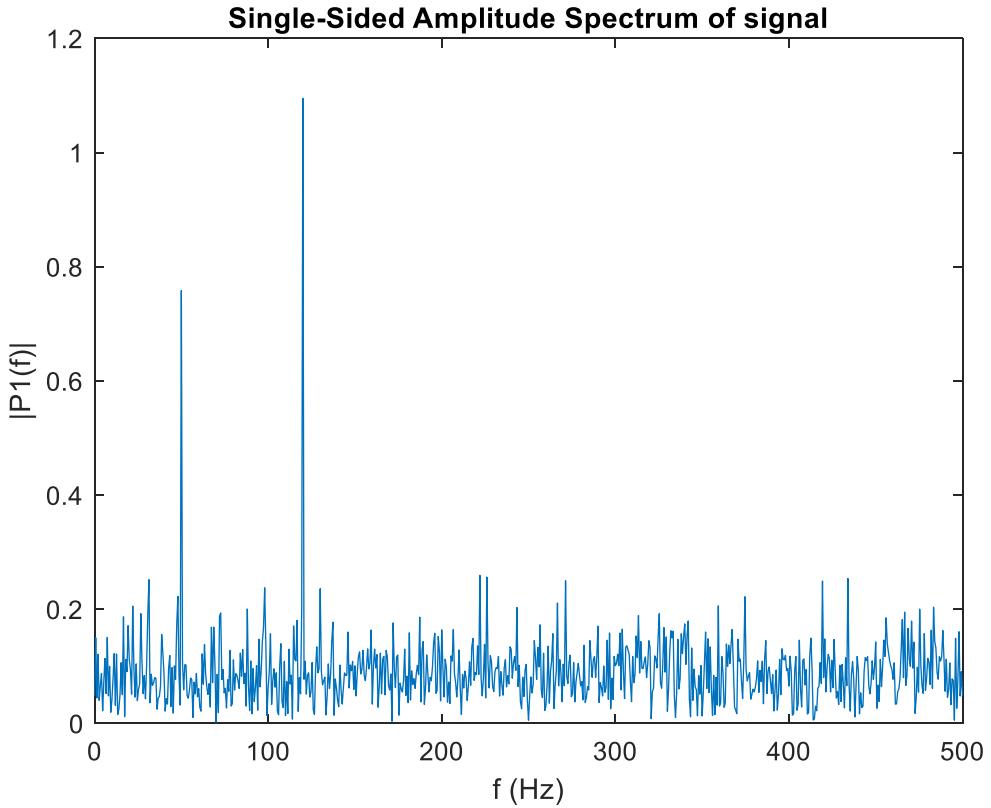
Exercise 8

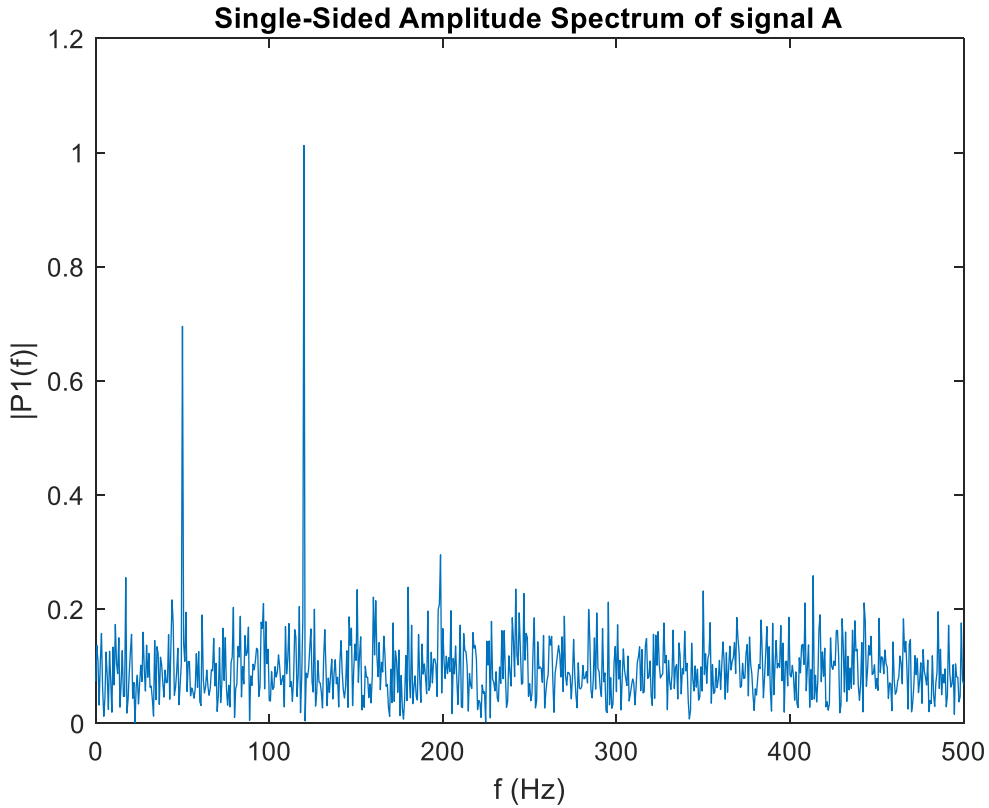
To test the hypothesis that the previous signal has a relationship with signal A.

Methods

Results

Conclusions





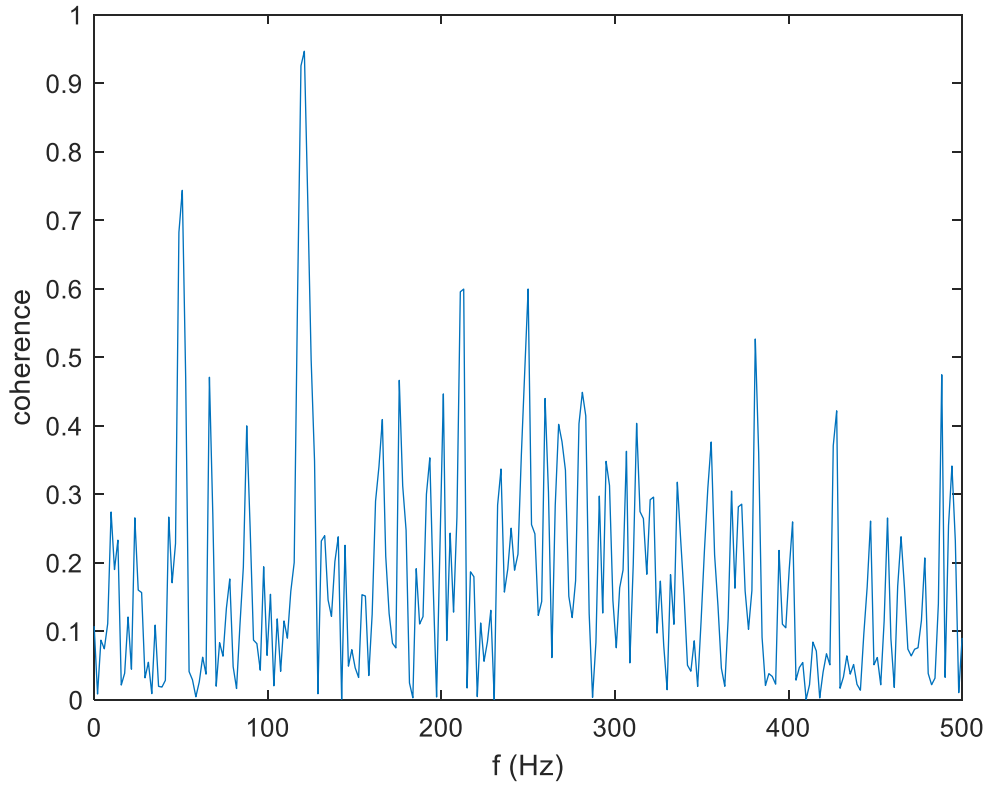
Exercise 8

To test the hypothesis that the previous signal has a relationship with signal A.

Methods

Results

Conclusions



BIOMEDICAL DATA

- LINEAR VS NONLINEAR
- DETERMINISTIC VS STOCHASTIC
- STATIONARY VS NONSTATIONARY

Pay Attention

Biomedical data are nonlinear, nonstationary and deterministic / stochastic in nature

Analytical tools are applicable only for linear, deterministic/stochastic and stationary

BIOMEDICAL DATA

- IDENTIFY OUTLIERS
- IDENTIFY NOISE

Pay Attention

PROJECT 1

- Test the hypothesis that apnea events in preterm infants occurs randomly

PROJECT 2

- Test the hypothesis that cortical regions have interactions and interactions are stronger in the adjacent regions

THANK YOU

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

SBIR: RAE (Realize, Analyze, Engage) - A digital biomarker based detection and intervention system for stress and carvings during recovery from substance abuse disorders.
PIs: M. Reinhardt, S. Carreiro, P. Indic

STARs Award

The University of Texas System
P. Indic (PI, UT Tyler)

Department of Veterans Affairs

Design of a wearable sensor system and associated algorithm to track suicidal ideation from movement variability and develop a novel objective marker of suicidal ideation and behavior risk in veterans.
Clinical Science Research and Development Grant (approved for funding),
P. Indic (site PI, UT-Tyler)
E.G. Smith (Project PI, VA)
P. Salvatore (Investigator, Harvard University)

Design of a wearable biosensor sensor system with wireless network for the remote detection of life threatening events in neonates.

National Science Foundation Smart & Connected Health Grant
P. Indic (Lead PI, UT-Tyler)
D. Paydarfar (Co PI, UT-Austin)
H. Wang (Co PI, UMass Dartmouth)
Y. Kim (Co PI, UMass Dartmouth)

Pre-Vent

National Institute Of Health Grant
P. Indic (Analytical Core PI, UT-Tyler)
N. Ambal (PI, Univ. of Alabama, Birmingham)

Wearable system for the detection of addiction
P. Indic (PI, UT-Tyler)
M. Reinhart (PI, ContinueYou, LLC
S. Carriero, (PI. Univ. of Mass. Med. School)