



Department of Electrical and Computer Engineering

ENEE2307 Probability and Statistical Engineering

– 1st semester 2019/2020

Project: Statistics about students' activities on Facebook

Deadline: 02/01/2020

Introduction

In this project we will derive some statistics about students' activities on Facebook. You will be collecting information about publishing posts on Facebook as well as liking, commenting, and sharing others posts on Facebook.

Phase One: Data Collection [each student works individually]

In this phase, you will be collecting information about your posts on Facebook. Collect information about at least 100 posts you published or shared on Facebook and list them in a table similar to the one shown next:

Table 1 : Activities on Facebook

i	Time in hours since the post has been published or shared	Number of likes	Number of Comments	Number of Shares
1				
2				
.				
.				
n				

Define the following random variable:

T : the time passed since you published or shared the post

L : the number of likes added to this post

C : the number of comments added to this post

S : the number of times this post has been shared

Phase Two: Computing Statistics [each student works individually]

In this phase, you will try to determine the relation between the defined random variable.

- **Derive Statistics about each of the Random Variables:**
 1. What is the sample mean of each of the random variables?
 2. What is the sample variance of each of the random variables?

- **Relation between Random Variables L (number of like) and C (number of comments):**
 1. Plot a scattering plot for the two random variables.
 2. Determine the correlation coefficient between these two random variables.
 3. Determine the best fitting line of the points $\langle L_i, C_i \rangle$.

- **Relation between Random Variables L (number of like) and S (number of shares):**
 1. Plot a scattering plot for the two random variables.
 2. Determine the correlation coefficient between these two random variables.
 3. Determine the best fitting line of the points $\langle L_i, S_i \rangle$.

- **Relation between Random Variables T (time passed since publishing/sharing the post) and Y (activities received for the post):**
 1. Define a new random variable **Y** such that $Y = L + C + S$
 2. Plot a scattering plot for the two random variables T and Y.
 3. Determine the correlation coefficient between these two random variables (T and Y)
 4. Determine the best fitting line of the points $\langle T_i, Y_i \rangle$.

Phase Three: Data Analysis [work with your partners in the group]

In this phase you will work with the other students in your group to compare the parameters and statistics computed in phase two. **Minimum group size is 4 and maximum group size is 8.** You may try to answer the following questions:

- 1- Is there a correlation between the random variables defined above?
- 2- Do you get the same correlation value compared to the other students in your group?
- 3- Does the best fit line represent accurately the relationship between the defined random variables?