Term Exam 2

(Linear Model)

1. Using the data provided in Bronchitis.csv generate a Linear Model, evaluate the model, and provide the evidence of the appropriateness of the model. The aim of the study was to define if the probability of having Bronchitis is influenced by smoking and/or pollution. The data are stored contains information on 212 participants.

(PCA)

2. 25 students aspiring to enter to Computer Science have been evaluated on 5 different subjects: Differential Geometry (DifGeo), Complex Analysis (CompAn), Algebra (Alg), Real Analysis (RealAn) and Statistics (Stat). These were the obtained scores:

ID	DifGeo	CompAn	Alg	RealAn	Stat
ST-01	36	58	43	36	37
ST-02	62	54	50	46	52
ST-03	31	42	41	40	29
ST-04	76	78	69	66	81
ST-05	46	56	52	56	40
ST-06	12	42	38	38	28
ST-07	39	46	51	54	41
ST-08	30	51	54	52	32
ST-09	22	32	43	28	22
ST-10	9	40	47	30	24
ST-11	32	49	54	37	52
ST-12	40	62	51	40	49
ST-13	64	75	70	66	63
ST-14	36	38	58	62	62
ST-15	24	46	44	55	49
ST-16	50	50	54	52	51
ST-17	42	42	52	38	50
ST-18	2	35	32	22	16
ST-19	56	53	42	40	32
ST-20	59	72	70	66	62
ST-21	28	50	50	42	63
ST-22	19	46	49	40	30
ST-23	36	56	56	54	52
ST-24	54	57	59	62	58
ST-25	14	35	38	29	20

The objective of this study is to obtain a global ranking of students for entry into the Computer Science, through a global score, extracted as a certain linear combination of grades in the five subjects tested. Use PCA to obtain that global score, sort the candidates and generate a formula to evaluate future candidates.

3. The following sample of a random vector (X, Y) has been obtained: (-3,7), (1,4), (0,6), (3,5), (1,9), (7,7), (6,9), (8,10), (7,12), (10,11).
    a. Perform a principal component analysis and a Regression on this sample.
    b. Represent a scatter plot, together with the first principal component and the regression line.