CH 5440 Multivariate Data Analysis

Assignment 1

1. The level of phytic acid in urine samples was determined by a catalytic fluorimetric (CF) method and the results were compared with those obtained using an established extraction photometric (EP) technique. The results, in mg/L, are the means of triplicate measurements, as shown in Table 2.
   (a) Is the new method (CF) a good substitute for the established method (EP) for measuring the level of phytic acid in urine? Use ordinary linear least squares regression to justify your conclusion.
   (b) Estimate the level of phytic acid in urine if EP measurement is 2.31 mg/l and provide 95% confidence intervals for these estimate.

Table 2. Comparison of CF versus EP

EP	CF	EP	CF
1.98	1.87	0.13	0.14
2.31	2.20	3.15	3.20
3.29	3.15	2.72	2.70
3.56	3.42	2.31	2.43
1.23	1.10	1.92	1.78
1.57	1.41	1.56	1.53
2.05	1.84	0.94	0.84
0.66	0.68	2.27	2.21
0.31	0.27	3.17	3.10
2.82	2.80	2.36	2.34

2. One of the consequences of global atmospheric rise in temperature that is predicted is the rise in sea level due to melting of glaciers, increase in precipitation etc. Annual average sea level data based on satellite measurements downloaded from EPA site are provided in file epa-sea-level.csv. Use OLS regression to estimate the annual rate of rise in sea level. If sea level rises by an additional 75cm from current levels, land extending up to 1km from Chennai coast will come under water. Estimate based on the regression model the number of years it will take for this to occur. Perform this analysis with (a) all the data and (b) data for the most recent 50 years. Compare your estimates with the report in TOI https://timesofindia.indiatimes.com/city/chennai/in-5-years-areas-100m-inland-may-go-under-the-sea/articleshow/78687890.cms)

3. Image analysis is used to identify defects in infrastructures such as bridges, roads or in manufactured products such as glass sheets, rolled steel sheets etc. One of the first steps in image analysis is annotation of the defect using an annotation tool such as CVAT, where each defect is marked using a polygon enclosing the defect. The corners of the polygon are pixels which are

indicated by the x and y coordinates of the pixel in the image. Annotated data for three defects identified in images taken of the surface of a concrete pillar of a bridge are given in the file defects_annotations_data.csv. We wish to determine whether the annotated defect is due to corrosion of the steel reinforcement bars. From previous experience, it is known that such defects will be aligned horizontally or vertically (since the reinforcement bars are embedded horizontally or vertically in the concrete. Determine which of the three annotated defects are due to corrosion of steel bars, and provide justification for your answer. Note that the measured coordinates of both x and y are noisy and can be assumed to have identical error variances.

4. Carbon-dioxide (CO2) is one of the major greenhouse gases that is implicated in the gradual warming of the earth's temperature. Measured concentrations of CO2 (in ppm) and atmospheric temperature (spatially and temporally averaged over a year) available from USEPA's Climate Change Indicators website (www.epa.gov/climate-indicators) between 1984 and 2014 is given in Table 1. The temperatures are deviation in deg F from the average temperature in the period 1901-2000. Climate models recommend that the global temperature increase should be kept below 2 deg C (3.6 deg F) by cutting down on CO2 emissions. Use OLS and TLS regression to estimate the maximum permissible level of CO2 in the atmosphere that can meet this goal. Assume that the level of CO2 increases linearly with time, estimate using the given data how many years it will take for CO2 to reach the maximum permissible. Note that this is a simplified analysis because other greenhouse gases such as methane, nitrous oxide, water vapour, etc. have not been considered.

Table 1. Measured average atmospheric CO2 concentration and temperature

Year	CO2	Temp (^{0}F)	Year	CO2	Temp (^{0}F)
1984	344.58	0.27	1999	368.33	0.792
1985	346.04	0.234	2000	369.52	0.756
1986	347.39	0.414	2001	371.13	0.972
1987	349.16	0.666	2002	373.22	1.08
1988	351.56	0.666	2003	375.77	1.098
1989	353.07	0.522	2004	377.49	1.026
1990	354.35	0.774	2005	379.8	1.17
1991	355.57	0.72	2006	381.9	1.098
1992	356.38	0.45	2007	383.76	1.098
1993	357.07	0.504	2008	385.59	0.972
1994	358.82	0.612	2009	387.37	1.134
1995	360.8	0.81	2010	389.85	1.26
1996	362.59	0.576	2011	391.63	1.026
1997	363.71	0.918	2012	393.82	1.116
1998	366.65	1.134	2013	396.48	1.188
			2014	398.61	1.332