**CH 5440 Multivariate Data Analysis**

Assignment 1

Due Date: 29/1/18

1. (a) Let and be a set of *N* measurements of two variables *x* and *y* which are linearly related. We are interested in determining the linear regression parameter *α* where *y = αx + β*. Assume that the measurements of x and y contain errors, with standard deviations and, respectively. (a) If the ratio of the error variances is known, derive the weighted TLS (WTLS) estimates of *α* and *β* in terms of . (b) How will the solution for *α* change if it is already known that the constant *β* is known to be 0?

*Note:* The WTLS regression problem when the error variances are known is the solution of the following minimization problem. Multiply the objective function by  and replace the ratio of the error variances by *λ*. Differentiate the objective function with respect to the decision variables and solve resulting set of nonlinear algebraic equations for obtaining the parameters *α* and *β*.

(b) Obtain the solution for the estimates of x and y for each case (OLS, IOLS, TLS) in terms of the regression parameters and measurements

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.
2. Is the new method (CF) a good substitute for the established method (EP) for measuring the level of phytic acid in urine? Justify your conclusion using linear regression between the two methods for different modelling assumptions regarding the accuracy of the respective measurement techniques.
3. Estimate the level of phytic acid in urine if EP measurement is 2.31 mg/l and CF measurement is 2.20 mg/l, for different modelling assumptions and provide 95% confidence intervals for these estimates, if possible.
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](http://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 1.5 deg C by cutting down on CO2 emissions. Using linear regression estimate the maximum permissible level of CO2 in the atmosphere that can meet this goal. Your estimate should be conservative (which implies that among all estimates based on different assumptions you should use the least value). Note that this is a simplified analysis because other greenhouse gases such as methane, nitrous oxide, water vapour, etc. have not been considered. In order to improve your model you are encouraged to use other reliable data sources you can find (cite the sources from where you obtain additional data).

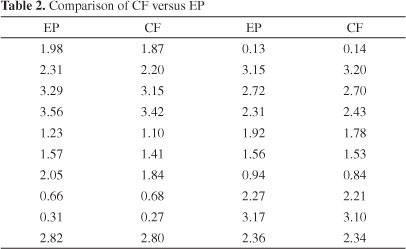


Table 1. Measured average atmospheric CO2 concentration and temperature

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | CO2 | Temp | Year | CO2 | Temp |
| 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 |