1)  **Linear Regression Review**

Consider the [climate dataset](https://sjsu.instructure.com/courses/1597069/files/79033776?wrap=1)[Download climate dataset](https://sjsu.instructure.com/courses/1597069/files/79033776/download?download_frd=1), perform the following tasks with linear regression.  Use the data before Year 2006 for training and the rest for testing.

1. Create a linear regression model that will predict **Temp**using all the other variables.   Evaluate the performance of this model using the testing set.  Create a scatter plot of the data with the linear regression model showing the R2 score.
2. Write the equation that represents the linear regression model in a).
3. Compute the correlations between all the variables.  Identify the ones that are highly correlated.
4. Create a new model after dropping the highly correlated variables (use 0.75 as the cutoff).  Keep the variables that appear first in the list (e.g.  keep CO2 instead of N2O if they are correlated).  Again, evaluate the model performance using the testing set and create a scatter plot of the data with the new linear regression model showing the R2 score.

(25 points)

2) **Poisson Regression Modeling**

For this problem, we'll perform Poisson regressions on a [sample asthma patient dataset](https://sjsu.instructure.com/courses/1597069/files/79034046?wrap=1)[Download sample asthma patient dataset](https://sjsu.instructure.com/courses/1597069/files/79034046/download?download_frd=1).  There are 4 quantities in this dataset: **gender**, **res\_inf**, **ghq12**and **attack**.   The **attack**variable is a **count**outcome variable (# of attacks per year) while the other 3 are predictor variables.  Perform the following tasks:

1. First, create 3 simple Poisson regression models with each of the 3 predictor variables to predict **attack**counts (i.e. one model to predict the attack count from **gender**, another from **res\_inf**etc).  Summarize, in a table, the significance of each of predictor variables in its corresponding model and comment on their significance.
2. Create another Poisson regression model which depends on all 3 predictor variables.  Evaluate the significance of each of the variables in this multivariate regression model.
3. Write down the equation that represents the Posson regression model in b).
4. Determine whether the data is overdispersed or not.
5. Assess the fitness of all four models using various metrics discussed in class.
6. Predict the # of **attacks**per year for a patient with **res\_inf** = yes and **ghq12**= 15.

(25 points)

3) **Logistic Regression Concepts**

The following table shows the partial output of a logistic regression model that was fitted on a data set on death due to a given disease:

==============================================================================  
                 coef     std err       z        P>|z|        [0.025    0.975]  
------------------------------------------------------------------------------  
Intercept    -15.3001      0.0015  
Factor 1       0.0018      0.0103  
Factor 2      -0.0061      0.0105  
Factor 3       0.0057      0.0028  
Factor 4       0.0066      0.0038  
Factor 5       0.0071      0.0199  
Factor 6       0.1113      0.0492  
Factor 7      -0.0098      0.0037  
Age            0.0686      0.0224  
==============================================================================

1. Write down the equation that represents the logistic regression model.
2. Complete the table above (i.e. typical summary output from a logistical regression model) by computing the following:
   * 1. z values
     2. P > |z| values
     3. Confidence intervals
3. Determine the significance of each of the factors.
4. Give brief interpretations of the **Age**and **Factor 7** coefficients.
5. Compute the odds ratios relating the additional risk of death associated with
   * 1. a 100 units decrease in **Factor 7**
     2. an additional year of age after adjusting for the other risk factors.
6. Predict the probability of death for a 50 year old woman with the following values for the different factors:

|  |  |
| --- | --- |
| Factor 1 | 125 |
| Factor 2 | 105 |
| Factor 3 | 235 |
| Factor 4 | 105 |
| Factor 5 | 12.5 |
| Factor 6 | 45 |
| Factor 7 | 475 |