STAT436 **Homework 3** Due Thursday January 23rd at 2.30pm

Directions: Answer the following exercises, showing all relevant work. Conclusions and justifications are to be given in clear detailed English. For each exercise and part, be sure to clearly write down all needed assumptions and requirements. Please type up your solutions or write ***very*** neatly.

1. In his Ph.D. thesis, Dr. Rick Linthurst of NCSU examined the dataset analyzed in Outputs A and B. His goal was to identify the important soil characteristics influencing (aerial) biomass production of the marsh grass *Spartina alterniflora* in the Cape Fear Estuary of North Carolina. Rick measured salinity (%), pH (i.e., acidity), K (potassium in ppm), Na (sodium in ppm) and Zn (zinc in ppm) in randomly selected plots of land taken from the estuary.

   (a) For the multiple linear regression (MLR) performed in Output A, write down the needed assumptions, and indicate how many plots were chosen for this study.

   (b) For the MLR in Output A, clearly interpret the estimated slope corresponding to the 'pH' term. Next, in this MLR model, do a test whether this slope is zero, indicating the hypotheses and justifying your conclusion.

   (c) From the above MLR model, test whether the 'salinity', 'K' and 'Zn' terms can simultaneously be dropped from the model. Clearly give the hypotheses, show your calculations, test statistic and justifications.

   (d) For the MLR in Output B, clearly interpret the estimated slope corresponding to the 'pH' term.

   (e) Assess how well you feel the MLR model in Output B fits these data. Be clear and detailed.

2. A researcher is trying to understand why statisticians call the model function

$$g(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

   "linear" but call the following model function *"nonlinear"*:

$$h(x) = \theta e^{-\varphi x}$$

   Help this researcher to understand these terms in the context of these examples by clearly explaining what makes a nonlinear model 'nonlinear' (and what makes a linear model 'linear'); provide all needed partial derivatives.

Output A

Regression Analysis: biomass versus salinity, ph, k, na, zn

```
The regression equation is
biomass = 1252 - 30.3 salinity + 306 ph - 0.285 k - 0.0087 na - 20.7 zn

Predictor      Coef  SE Coef      T      P
Constant       1252     1235   1.01  0.317
```

```
salinity      -30.29    24.03  -1.26  0.215
ph            305.53    87.88   3.48  0.001
k            -0.2851   0.3484  -0.82  0.418
na         -0.00867  0.01593  -0.54  0.589
zn           -20.68    15.05  -1.37  0.177

S = 398.274   R-Sq = 67.7%   R-Sq(adj) = 63.6%

Analysis of Variance
Source           DF        SS       MS      F      P
Regression        5  12984700  2596940  16.37  0.000
Residual Error   39   6186263   158622
Total            44  19170963
```

Output B

Regression Analysis: biomass versus ph, na

```
The regression equation is
biomass = - 476 + 405 ph - 0.0233 na

Predictor        Coef   SE Coef      T      P
Constant       -475.7     273.5  -1.74  0.089
ph             404.95     47.77   8.48  0.000
na         -0.023326  0.008655  -2.70  0.010

S = 394.853   R-Sq = 65.8%   R-Sq(adj) = 64.2%

Analysis of Variance
Source           DF        SS       MS      F      P
Regression        2  12622789  6311394  40.48  0.000
Residual Error   42   6548174   155909
Total            44  19170963
```