Stats 539 Homework 1: Due Thursday, Jan. 26 by 10:50am

1. Agresti Exercise 1.4 (p. 21)
2. Agresti Exercise 1.8 (p. 22)
3. Agresti Exercise 2.10 (p. 72)
4. Consider the following model: For $i = 1, \ldots, n$,

$$y_i = \beta_1 x_i + e_i,$$

where $E(e_i) = 0$ and $Var(e_i) = \sigma^2$.

   (a) Write this model in the matrix form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, writing out the dimensions and elements of $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\mathbf{e}$.

   (b) Derive the least squares estimate for β_1, $\hat{\beta}_1$.

   (c) Draw a plot (either by hand or with software) of $E(Y|x)$ versus x.

5. Patients with partial seizures were enrolled in a randomized clinical trial of the anti-epileptic drug, progabide. Participants were randomized to either progabide or a placebo. Prior to receiving treatment, baseline data on the number of seizures during the preceding 8-week interval were recorded. Counts of epileptic seizures during 2-week intervals before each of four successive post-randomization clinic visits were recorded. The main goal of the study was to compare the changes in the average rates of seizures in the two groups.

   Displayed below are the data for the first three and last three individuals in the data set, where the response listed in columns 4−8 is the *rate* of seizures per week for that individual at baseline (week 0), week 2, week 4, week 6, and week 8 (rate = count/8 for baseline measurement; rate = count/2 for post-baseline measurements):

```
> head(epilepsy,3)
    ID       trt   age   Week0 Week2 Week4 Week6 Week8
1     1   Placebo    31  1.375   2.5   1.5   1.5   1.5
2     2   Placebo    30  1.375   1.5   2.5   1.5   1.5
3     3   Placebo    25  0.750   1.0   2.0   0.0   2.5
> tail(epilepsy,3)
     ID         trt   age   Week0 Week2 Week4 Week6 Week8
57   57   Progabide    21  3.125   1.0   1.5   0.0   0.5
58   58   Progabide    36  1.625   0.0   0.0   0.0   0.0
59   59   Progabide    37  1.500   0.5   2.0   1.5   1.0
```

   Define variables: $Y_{ij} = j$th rate measurement on the ith individual,

$$T_i = \begin{cases} 1 & \text{individual } i \text{ was randomly assigned to Progabide} \\ 0 & \text{individual } i \text{ was randomly assigned to Placebo} \end{cases}$$

and

$$w_j = \text{week of the } j\text{th measurement (e.g., } w_1 = 0).$$

where $i = 1, \ldots, 59$, and $j = 1, \ldots, n$. (Note that we are treating week as a quantitative variable.)

(a) Is this study exploratory or confirmatory? Explain.

(b) For each of the following mean models expressed in scalar notation below,

1. write out the vector of regression parameters $\boldsymbol{\beta}$ (the elements will be symbols),
2. write out the rows of the design matrix $\mathbf{X}$ that correspond to the observations on individuals 3 and 59 (the elements will be numbers),
3. on a single well-labeled plot, draw two hypothetical mean response trajectories (assuming all elements of β are non-zero), one for each treatment (Placebo and Progabide), with time (in weeks) on the x-axis, and the mean rate of seizures on the y-axis.

i. $E(Y_{ij}) = \beta_1 + \beta_2 T_i$

ii. $E(Y_{ij}) = \beta_1 + \beta_2 w_j$

iii. $E(Y_{ij}) = \beta_1 + \beta_2 T_i + \beta_3 w_j$

iv. $E(Y_{ij}) = \beta_1 + \beta_2 w_j + \beta_3 T_i \times w_j$

v. $E(Y_{ij}) = \beta_1 + \beta_2 T_i + \beta_3 w_j + \beta_4 T_i \times w_j$

(c) Which of the models in part (b) is most appropriate for the main goal of the study? Explain.

(d) Suppose you fit a linear model to these data using ordinary least squares. What is the primary reason why inference on the linear model coefficients would be invalid? Explain.

(e) In mean model v. in part (b), if we treated week as a categorical variable, how many additional coefficients would we add?