STAT 645: Assignment 2

~~Due Friday, September 13~~, 11:55pm Central

Due date: Monday, September 7

1. With the `calcium` data in "calcium.txt," consider the `Decrease` variable as your response and `Treatment` as your treatment. In what follows, I have recoded `Treatment` to equal 0 for placebo and 1 for calcium treatment.

   (a) For the regression model

   $$\texttt{Decrease}_i = \beta_0 + \beta_1 \texttt{Treatment}_i + \epsilon_i,$$

   write down the model matrix.

   (b) Fit the above model, and report the coefficient estimates and standard errors.

   (c) Based on the model, what is the p-value for the null hypothesis of no treatment effect?

   (d) Now analyze the same data using a two-sample t-test, assuming equal variances. How do the results compare to those you obtained using the regression model?

   (e) Assuming that the ϵ_i are normally distributed, what is the estimated distribution of `Decrease` when `Treatment` $= 1$?

2. With the `onset` data in "onset_data.csv," conduct the following analysis.

   (a) Create side-by-side box plots comparing time to onset with (i) the `tx` variable and (ii) the `prior` variable. Comment.

   (b) Create a scatterplot of `onset` vs. `age`. Color code the points by `prior` status. Also, fit and overlay separate lowess curves, one each for `prior` $= 0$ and `prior` $= 1$.

   (c) Fit the regression model

   $$y_i = \beta_0 + \beta_1 \texttt{tx}_i + \beta_2 \texttt{prior}_i + \beta_3 \texttt{age}_i + \beta_4 (\texttt{prior} \times \texttt{age}) + \epsilon_i$$

   Interpret all coefficients and report their estimates and standard errors.

   (d) Use matrix manipulation using a design matrix to verify the estimates and standard errors from above.

   (e) What is a 95% confidence interval for the mean difference in onset times between the treatment and control groups, holding prior status and age constant?

   (f) What is a 95% confidence interval for the mean response of a treated individual, age 35, with no prior tumor incidence?

3. Suppose that $y_1, y_2, \ldots, y_n$ are *i.i.d.* realizations from the $N(0, \sigma^2)$ distribution. Derive the maximum likelihood estimator of σ^2.

4. Suppose the times to infection following exposure to a particular bacteria follow the gamma distribution with shape parameter α, scale parameter β, and *pdf*

   $$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}.$$

   Use the `nlm` function in `R` to compute the maximum likelihood estimates for the data in "gamma.csv."