MSiA 420, HW #1
See Canvas for Due Date

As with all HW (unless otherwise noted), upload your solutions for this assignment on Canvas, as a Word or pdf file, by the due date/time. For all problems for which you use R, include your R script in an appendix to your homework (clearly label which parts of the script correspond to which homework problems).

1) In class, we showed that the MLE of the mean of a random sample $\{y_1, y_2, \ldots, y_n\}$ from an $N(\mu, \sigma^2)$ population is exactly the sample average (i.e., $\hat{\mu} = \bar{y}$). Show that the MLE for the standard deviation is:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Hint: Find the partial derivative of the likelihood function with respect to σ, then set this equal to zero and solve for σ. The result will be a function of the unknown μ, but you can plug in the MLE of μ from class. Hence, we are really finding the *joint* MLEs of μ and σ.

2) (Problem 13.10 from KNN). In a study of enzyme kinetics, a postulated model relating the velocity of reaction (Y) to the concentration (X) is of the form

$$Y_i = \frac{\gamma_0 X_i}{\gamma_1 + X_i} + \varepsilon_i$$

The velocity at 18 different concentrations were recorded as part of a study, the data for which are listed the Prob 2 worksheet of HW1_Data.xls.

(a) When fitting nonlinear regression models, it is often important to have reasonably good initial guesses for the parameters. Towards this end, notice that without the error term the model is $Y'_i = \beta_0 + \beta_1 X'_i$, where $Y'_i = 1/Y_i$, $\beta_0 = 1/\gamma_0$, $\beta_1 = \gamma_1/\gamma_0$, and $X'_i = 1/X_i$. In light of this, fit a linear regression model to the transformed data with Y'_i the response and X'_i the predictor to obtain initial guesses of the form $\hat{\gamma}_0 = 1/\hat{\beta}_0$ and $\hat{\gamma}_1 = \hat{\beta}_1/\hat{\beta}_0$.
(b) Using the initial guesses from part (a), find the nonlinear least squares estimates of γ_1 and γ_0. Fit the model twice, using the two different R functions nlm() and nls().

3) Refer to the same data from Problem (2).

(a) Calculate the observed Fisher information matrix and the covariance matrix of the estimated parameter vector $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_0, \hat{\gamma}_1]^T$ using the Hessian produced by nlm(). Based on this, calculate the standard errors of the estimated parameters.

(b) Calculate the covariance matrix of $\hat{\boldsymbol{\gamma}}$ using the vcov() function applied to the output of nls(), and based on this calculate the standard errors of the estimated parameters. Do the results agree with Part (a)?

(c) Using the results of Part (a), calculate two-sided 95% CIs on the parameters γ_0 and γ_1. Compare this with the results of the confint.default() function applied to the output of nls().

4) This is a repeat of Problem (3), but using bootstrapping to calculate the standard errors and confidence intervals. You can use the boot() command in R (requires the boot package to be loaded with the library(boot) command). Use at least 20,000 bootstrap replicates.

(a) Calculate and plot bootstrapped histograms of $\hat{\gamma}_0$ and $\hat{\gamma}_1$, and calculate the corresponding bootstrapped standard errors.

(b) Calculate "crude" two-sided 95% CIs on γ_0 and γ_1 using the normal approximation to their bootstrapped distributions.

(c) Calculate the reflected two-sided 95% CIs on γ_0 and γ_1 (this corresponds to the type = "basic" option of the boot.ci() function).

(d) Do the CIs in part (c) agree with those in part (b)? Relate this to the histograms you see in part (a).

5) Use bootstrapping to calculate a two-sided 95% prediction interval on a "future" response Y^* at $X^* = 27$. Compare this to a two-sided 95% confidence interval on the predictable part $g(\mathbf{x}^*,\boldsymbol{\theta})$ of Y^* at $X^* = 27$. Which interval do you think better represents an interval that you would expect to contain the future response with roughly 95% chance? Explain

6) Use the AIC criterion to compare the model that you fitted in Problem (2b) with the alternative model $Y_i = \beta_0 + \beta_1\sqrt{X_i} + \varepsilon_i$. Which model does AIC suggest is the better model? Use the general expression given in the lecture notes for the AIC:

$$AIC = -\frac{2logf(\mathbf{y};\widehat{\boldsymbol{\theta}})}{n} + \frac{2p}{n}, \text{ with}$$
$$logf(\mathbf{y};\widehat{\boldsymbol{\theta}}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{SSE}{2\hat{\sigma}^2} \text{ and } \hat{\sigma}^2 = \frac{SSE}{n}$$

Note that the built-in log-likelihood values produced by both the lm() function and the nls() function are consistent with the above expression. However, their built-in values for AIC may use slightly different expressions, depending on whether they divide by n in the expression for AIC and whether they use $p = 2$ or $p = 3$ (the latter accounts for the estimation of σ^2). As long as you are consistent and include or exclude the estimation of σ^2 the same way for both models when determining p, it does not affect which model has a lower AIC.

7) Use n-fold cross-validation to compare the model that you fitted in Problem (2b) with the alternative model $Y_i = \beta_0 + \beta_1 \sqrt{X_i} + \varepsilon_i$. Which model does n-fold cross-validation suggest is the better model?

8) For the two models that you compared in Problems 6 and 7, construct plots of the residuals versus X. Based on the residual plots, does one model appear more appropriate than the other, and does this agree with your conclusions from Problems 6 and 7?