STAT 2131:
Applied Statistical Methods I
HW #4
Due Tuesday, September 15th

1. Consider the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$, and let $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ be the least squares estimates of the coefficients (β_0, β_1).

   (a) Derive the distribution of $\hat{\beta}$ (hint: it should be a 2-d normal distribution. Why?)

   (b) Suppose we observe a new covariate X_{n+1}. Assuming σ^2 is known, use your answer above to derive a 99% confidence interval for $E(Y_{n+1} \mid X_{n+1})$.

   (c) Now suppose σ^2 is unknown. If you simply replace σ^2 with $\{\hat{\sigma}^{(OLS)}\}^2$ in the interval you derived in (b), do you think the new interval is too wide, too narrow, or valid? That is, if you repeat the experiment many times, do you think this new interval will contain $E(Y_{n+1} \mid X_{n+1})$ more than, less than, or approximately 99% of the time? Explain.

   (d) Now derive a 99% confidence interval for Y_{n+1} assuming σ^2 is known. Is this interval larger or smaller than the one from part (b)? Does this agree with your intuition? Explain.