Statistics 111   **Quiz 21**

0.92 1. Consider the unit cube in 8 dimensions. What is the side length of a subcube that contains 50% of the total volume?

It is $\sqrt[8]{0.5} = 0.917$.

2. Describe how you would use 10-fold cross-validation to assess the predictive accuracy of a nonparametric regression. (5 points)

Randomly divide the sample into ten equal groups. Hold the first group out, fit the nonparametric regression model to the other nine groups combined. Use that fitted model to predict the hold-out group, and calculate the sum of the squared errors in your predictions on the hold-out group. Then put the hold-out group back with the others and remove the second group. Repeat the process, holding out each of the ten original groups in turn. Average the squared errors to estimate the squared error of your prediction when fitting the model to all of the data.

3. List all, and only, true statements. (7 points) A, B, F, G

   A. As the number of explanatory variables increases, predictive accuracy in regression gets much worse.
   B. When two of the explanatory variables are strongly correlated, one has multicollinearity.
   C. Interpolation is less reliable than extrapolation.
   D. Bin smoothing is more accurate than a running line smoother.
   E. A residual is the predicted value minus the observed value.
   F. As the number of independent variables increases, the number of possible models explodes.
   G. In high dimensions, one almost never has enough data.