Predicting Course Grades Using Midterm Scores
Part IV: Quantifying Uncertainty

Let's continue to analyze real, historical, student performance data from the class. The dataset we will use is in the *grades.csv* file. Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is students.

variable	description
midterm	students' scores in the midterm (from 0 to 100 points)
final	students' scores in the final exam (from 0 to 100 points)
overall	students' scores in the class overall (from 0 to 100 points)
gradeA	identifies students who earned an A or an A minus in the class

Table 1: Variables in "grades.csv"

In this problem set, we practice fitting a line to make predictions, creating scatter plots, adding the fitted line to the scatter plot, and constructing confidence intervals for our predictions.

As always, we start by loading and looking at the data:

```
## load and look at the data
grades <- read.csv("grades.csv") # reads and stores data
head(grades) # shows first observations
##   midterm final overall gradeA
## 1   79.25 47.00    69.2      0
## 2   96.25 87.75    94.3      1
## 3   58.25 37.75    62.0      0
## 4   54.50 62.00    72.4      0
## 5   83.00 39.75    72.4      0
## 6   41.75 49.50    59.5      0
```

1. First, let's fit the linear models that we will use to make predictions.

   a. Fit the following three linear models: (i) the linear model to predict final exam scores using midterm scores, (ii) the linear model to predict overall scores in the course using midterm scores, and (iii) the linear model to predict the probability of earning an A or an A- in the course using midterm scores. To specify the dataframe where the variables are stored, use the optional argument data, instead of using the $ operator for each variable. Store each fitted line in an object. Call the three objects *fit_final*, *fit_overall*, and *fit_gradeA*, respectively. Then, run the names of the objects, *fit_final*, *fit_overall*, and *fit_gradeA*, so that R will provide you with the contents of each object. Finish by writing the fitted line for each model. In other words, provide the formula $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X$ where you specify each term (i.e., substitute Y for the name of the outcome variable, substitute $\widehat{\alpha}$ for the estimated value of the intercept coefficient, substitute $\widehat{\beta}$ for the estimated value of the slope coefficient, and substitute X for the name of the predictor.) (10 points)

b. For each of the three linear models above, create a visualization of the relationship between X and Y and add the fitted line. (Hint: The functions plot() and abline() might be helpful here.) (R code only.) (10 points)

2. Now, let's use the fitted linear models to make some predictions.

a. Suppose that you earn 80 points in the midterm. Based on your performance in the midterm, what would be (i) your predicted final exam score, (ii) your predicted overall score, and (iii) your predicted probability of earning an A or A- in the course? Please show your calculations and then answer the question with a full sentence (including units of measurement). (5 points)

b. Because of potential noise in the data, there is some uncertainty around these predictions. Construct the 95% confidence interval for each of the three predictions using the function predict() and re-write your answers to the previous question accordingly. (25 points)