Note: This homework assignment covers Chapter 10.

Disclaimer: If you use R, include all R code and output as attachments. Do not just "write in" the R code you used. Also, don't just write the answer and say this is what R gave you. If my grader can't see how you got an answer, it is wrong. I want to see your code and your answers accompanying your code (like in the notes).

1. Physical fitness testing is an important aspect of athletic training. A common measure of the magnitude of cardiovascular fitness is the maximum volume of oxygen uptake during strenuous exercise. A study was conducted involving a random sample of $n = 24$ middle-aged men to determine the relationship between maximum oxygen uptake Y and the time required to complete a two-mile run (x, measured in seconds). Maximum oxygen uptake was measured with standard laboratory methods as the subjects performed on a treadmill. Here are the data from the study:

Max O_2	Time	Max O_2	Time	Max O_2	Time
42.33	918	36.23	1045	53.29	743
53.10	805	49.66	810	47.18	803
42.08	891	41.49	927	56.91	683
50.06	962	46.17	813	47.80	844
42.45	968	46.18	858	48.65	755
42.46	907	43.21	860	53.67	700
47.82	770	51.81	760	60.62	748
49.92	743	53.28	747	56.76	775

Consider the (population-level) model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

for $i = 1, 2, ..., 24$, as a model for these data.
(a) Use R to fit this model using least squares. What are the estimates of the population-level parameters β_0 and β_1? Does the straight-line regression model seem to be plausible (or does some other type of regression function seem better)? To answer this, superimpose the estimated model over the data.
(b) Does the time it takes to run two miles have a significant influence on maximum oxygen uptake in the population? Do an analysis to answer this question.
(c) Write a 90 percent confidence interval for the population mean maximum oxygen uptake for men who run two miles in 900 seconds. Clearly interpret the interval.
(d) Write a 90 percent prediction interval in the same setting as in part (c). Explain the difference between this interval and the confidence interval in part (c).

2. A rocket motor is manufactured by bonding together two types of propellants, an igniter and a sustainer. The shear strength of the bond Y (measured in psi) is thought to be a linear function of the age of the propellants x (measured in weeks) when the motor is cast. Consider the (population-level) model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

for $i = 1, 2, ..., 20$, as a model for the data on the next page.
(a) Use R to fit this model using least squares. What are the estimates of the population-level

Observation	Strength (Y)	Age (x)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.00
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2277.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

parameters β_0 and β_1? Does the straight-line regression model seem to be plausible (or does some other type of regression function seem better)? To answer this, superimpose the estimated model over the data.
(b) Does the shear strength of the bond appear to be linearly related to the age of the propellant in the population? Do an analysis to answer this question.
(c) Estimate, with a 95 percent confidence interval, the population mean shear strength of a motor made from propellant that is 20 weeks old. Interpret your interval.
(d) Calculate a 95 percent prediction interval in the same setting as in part (c). Interpret this interval clearly.

Extra Credit: Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Under these model assumptions, we claimed on pp 163 (notes) that

$$t = \frac{b_1 - \beta_1}{\sqrt{\frac{\text{MS}_{res}}{\text{SS}_{xx}}}} \sim t(n-2),$$

where b_1 is the least squares estimator of β_1, $\text{SS}_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, and MS_{res} is residual mean squares. Using this result, show how to obtain the $100(1-\alpha)$ percent confidence interval for the population slope parameter β_1; i.e.,

$$b_1 \pm t_{n-2,\alpha/2}\sqrt{\frac{\text{MS}_{res}}{\text{SS}_{xx}}}.$$

You will find previous confidence interval derivations (e.g., from Chapter 7) to be useful in getting started (perhaps also the prediction interval derivation on Problem 4 in HW6).