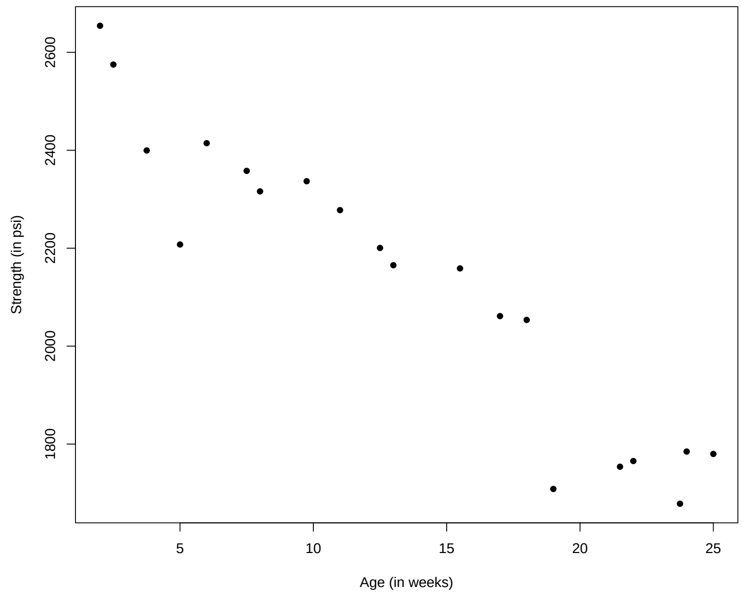
1. (a) Here is a scatterplot of the sample of motors and their data:



In the plot, we see that strength (Y) and age (x) are negatively linearly related. A straight-line relationship seems reasonable for the population of motors (at least, for those in this same age range).

(b) We will use R's `lm` function to calculate the least squares estimates of β_0 and β_1. Here is the R output:

```
> fit = lm(Strength ~ Age)
> fit

Coefficients:
(Intercept)          Age
    2625.39       -36.96
```

The least squares estimates are

$$\widehat{\beta}_0 \approx 2625.39$$
$$\widehat{\beta}_1 \approx -36.96.$$

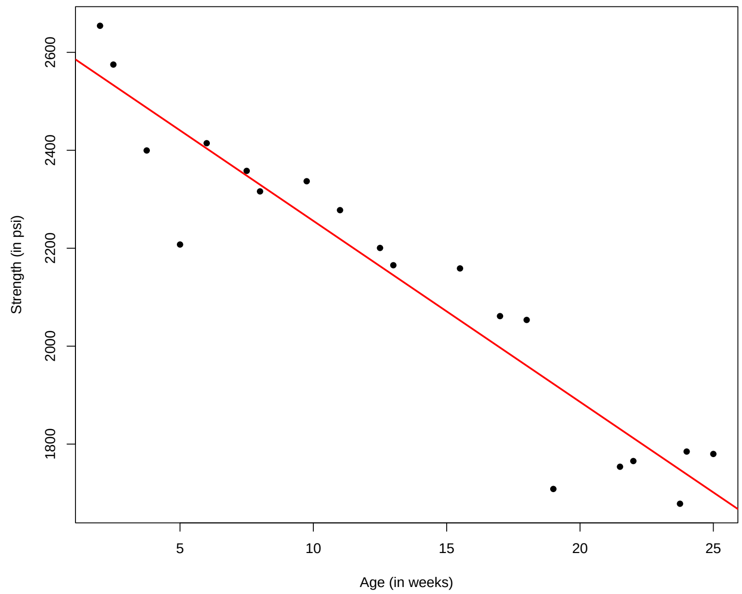
The equation of the least-squares regression line is

$$\widehat{Y} = 2625.39 - 36.96x \quad \Longleftrightarrow \quad \widehat{\texttt{Strength}} = 2625.39 - 36.96(\texttt{Age}).$$

The scatterplot with the least squares line superimposed is on the next page (top).

(c) We can get the 20 fitted values ($\widehat{Y}_i$) and the 20 residuals ($e_i = Y_i - \widehat{Y}_i$) in R. Here are the 20 fitted values:

```
> predict(fit)
      1       2       3       4       5       6       7       8       9      10
2052.48 1747.54 2329.69 1997.03 2440.58 1923.11 1738.30 2532.98 2348.17 2218.81
     11      12      13      14      15      16      17      18      19      20
2144.88 2486.78 1701.34 2265.01 1812.23 1960.07 2403.61 2163.36 2551.46 1830.71
```



Here are the 20 residuals:

```
> options(digits=3)
> residuals(fit)
      1       2       3       4       5       6       7       8       9      10      11
 106.22  -69.39  -13.69   64.27 -233.08 -214.81   46.40   42.02    9.73   58.89   20.32
     12      13      14      15      16      17      18      19      20
 -87.23   78.46   71.74  -46.93   93.43   10.79   37.14  102.74  -77.01
```

Verify the residuals sum to zero (up to rounding error):

```
> sum(residuals(fit))
[1] 1.07e-14
```

(d) We can use R to calculate a 95% confidence interval for β_1:

```
> options(digits=4)
> confint(fit,conf.level=0.95)
               2.5 %  97.5 %
(Intercept) 2530.12 2720.66
Age           -43.19  -30.73
```

From the output, we see a 95% confidence interval for β_1 is

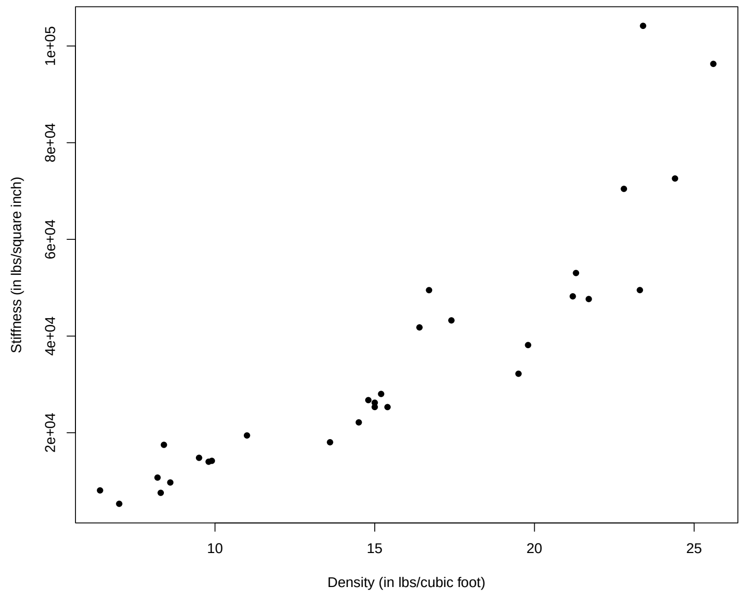
$$(-43.19, -30.73).$$

Note that this interval excludes 0 and includes only negative values. This suggests, at the 95% confidence level, the population mean strength $E(Y)$ and age are negatively related in the population of all motors; recall

$$Y = \beta_0 + \beta_1 x + \epsilon \iff E(Y) = \beta_0 + \beta_1 x.$$

Interpretation: For a one-week increase in the age of the propellants, we are 95% confident the population mean strength $E(Y)$ will decrease between 30.73 and 42.19 psi.

2. (a) Here is a scatterplot of the sample of particle boards and their data:



In the plot, we see that stiffness (Y) and density (x) are positively related. The relationship in the scatterplot looks mostly linear over smaller values of the density (x), but perhaps it is curved (i.e., quadratic) over the entire range when considering particle boards with larger densities?

(b) We use R's `lm` function to calculate the least squares estimates of β_0 and β_1. Here is the R output:

```
> fit = lm(Stiffness ~ Density)
> fit

Coefficients:
(Intercept)      Density
  -25433.84      3884.98
```

The least squares estimates are

$$\begin{aligned}\widehat{\beta}_0 &\approx -25433.84 \\ \widehat{\beta}_1 &\approx 3884.98.\end{aligned}$$

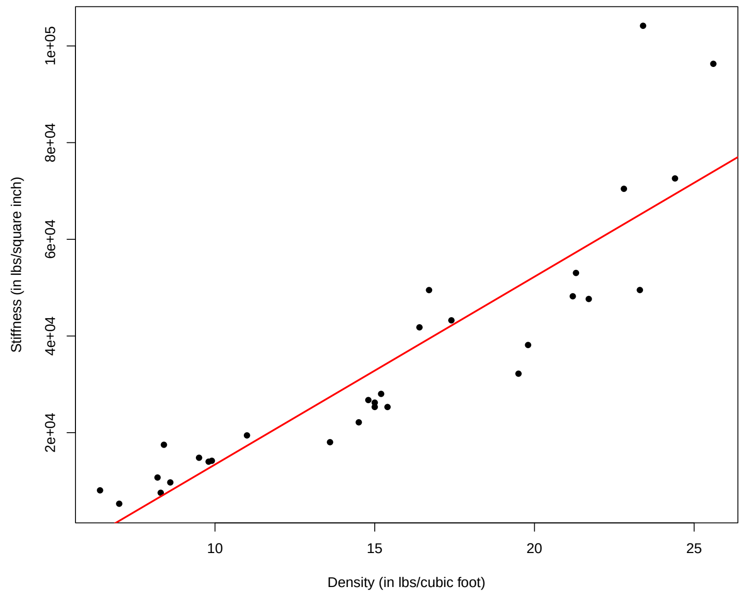
The equation of the least-squares regression line is

$$\widehat{Y} = -25433.84 + 3884.98x \quad \Longleftrightarrow \quad \widehat{\texttt{Stiffness}} = -25433.84 + 3884.98(\texttt{Density}).$$

The scatterplot with the least squares line superimposed is on the next page (top).

(c) We can use R's `predict` function to calculate a 95% confidence interval for the population mean stiffness $E(Y)$ when $x_0 = 15.0$ lbs/ft^3.

```
> predict(fit,data.frame(Density=15.0),level=0.95,interval="confidence")
       fit      lwr      upr
1 32840.86 28479.66 37202.06
```



The value `fit` in the R output (to 2 dp) is

$$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 = -25433.84 + 3884.98(15.0) = 32840.86.$$

The values `lwr` and `upr` give the lower and upper limits of the confidence interval for $E(Y)$.

Interpretation: We are 95% confident the population mean stiffness $E(Y)$ for all particle boards whose density is 15.0 lb/ft^3 is between 28479.66 and 37202.06 lb/in^2.

We can also use R's `predict` function to calculate a 95% prediction interval for Y^*, the stiffness of one particle board whose density is $x_0 = 15.0$ lbs/ft^3.

```
> predict(fit,data.frame(Density=15.0),level=0.95,interval="prediction")
       fit      lwr      upr
1 32840.86 8637.226 57044.49
```

The values `lwr` and `upr` give the lower and upper limits of the prediction interval for Y^*.

Interpretation: For a single particle board whose density is 15.0 lb/ft^3, we would predict its stiffness to be between 8637.226 and 57044.49 lbs/in^2 with probability 0.95.
(d) Confidence intervals for $E(Y)$ and prediction intervals for Y^* will be at their minimum length when $x_0 = \overline{x}$, the sample mean of the density observations $\overline{x} = 15.47$ lb/ft^3.

```
> mean(Density)
[1] 15.47
```

This makes sense intuitively. We have the best precision for our inferences nearest to the "center" of the data. As we attempt inference farther away from this center, the precision worsens.

R CODE:

```
# Problem 1
# Enter data
Strength = c(2158.70,1678.15,2316.00,2061.30,2207.50,1708.30,1784.70,2575.00,2357.90,2277.70,
    2165.20,2399.55,1779.80,2336.75,1765.30,2053.50,2414.40,2200.50,2654.20,1753.70)
Age = c(15.50,23.75,8.00,17.00,5.00,19.00,24.00,2.50,7.50,11.00,13.00,3.75,25.00,9.75,
    22.00,18.00,6.00,12.50,2.00,21.50)

# Create scatterplot
plot(Age,Strength,xlab="Age (in weeks)",ylab="Strength (in psi)",pch=16)

# Estimate the model
fit = lm(Strength ~ Age)
fit
plot(Age,Strength,xlab="Age (in weeks)",ylab="Strength (in psi)",pch=16)
# Superimpose least-squares line
abline(fit,lwd=2,col="red")

# Calculate 20 fitted values and 20 residuals
predict(fit)
residuals(fit)
sum(residuals(fit))

# Confidence intervals for regression parameters
confint(fit,conf.level=0.95)


# Problem 2
# Enter data
Stiffness = c(14814,14007,7573,9714,5304,43243,28028,49499,26222,26751,96305,72594,
    32207,70453,38138,17502,19442,14191,8076,10728,25319,41792,25312,22148,18036,
    104170,49512,48218,47661,53045)
Density = c(9.5,9.8,8.3,8.6,7.0,17.4,15.2,16.7,15.0,14.8,25.6,24.4,19.5,22.8,19.8,
    8.4,11.0,9.9,6.4,8.2,15,16.4,15.4,14.5,13.6,23.4,23.3,21.2,21.7,21.3)

# Create scatterplot
plot(Density,Stiffness,xlab="Density (in lbs/cubic foot)",
    ylab="Stiffness (in lbs/square inch)",pch=16)

# Estimate the model
fit = lm(Stiffness ~ Density)
fit
plot(Density,Stiffness,xlab="Density (in lbs/cubic foot)",
    ylab="Stiffness (in lbs/square inch)",pch=16)
# Superimpose least-squares line
abline(fit,lwd=2,col="red")

# Confidence interval for population mean
predict(fit,data.frame(Density=15.0),level=0.95,interval="confidence")
# Prediction interval
predict(fit,data.frame(Density=15.0),level=0.95,interval="prediction")
```