Inference with Simple Linear Regression

For this tutorial, we will continue to use the rocket propellant data. The textbook provides the description of the data as:

A rocket motor is manufactured by bonding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two types of propellant is an important quality characteristic. It is suspected that shear strength is related to the age in weeks of the batch of sustainer propellant.

Download the data file, `rocket.csv`, from Collab and read the data in.

```
Data<-read.csv("rocket.csv", header=TRUE)
head(Data)
```

```
##   Observation..i Shear.Strength..yi..psi. Age.of.Propellant..xi..weeks.
## 1              1                  2158.70                         15.50
## 2              2                  1678.15                         23.75
## 3              3                  2316.00                          8.00
## 4              4                  2061.30                         17.00
## 5              5                  2207.50                          5.50
## 6              6                  1708.30                         19.00
```

Notice there is an extra column for the observation number, and the names of the columns are long and complicated. So we remove the first column and rename the 2nd and 3rd columns

```
##remove first column
Data<-Data[,-1]
##rename the remaining 2 columns
names(Data)<-c("Strength", "Age")
head(Data)
```

```
##   Strength   Age
## 1  2158.70 15.50
## 2  1678.15 23.75
## 3  2316.00  8.00
## 4  2061.30 17.00
## 5  2207.50  5.50
## 6  1708.30 19.00
```

The data frame looks a lot neater now.

1. Hypothesis test for β_1 (and β_0)

Applying the `summary()` function to `lm()` gives the results of hypothesis tests for β_1 and β_0

```
result<-lm(Strength~Age, data=Data)
summary(result)
```

```
##
## Call:
## lm(formula = Strength ~ Age, data = Data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -215.98  -50.68   28.74   66.61  106.76
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2627.822     44.184   59.48  < 2e-16 ***
## Age          -37.154      2.889  -12.86 1.64e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.11 on 18 degrees of freedom
## Multiple R-squared:  0.9018, Adjusted R-squared:  0.8964
## F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10
```

Under coefficients, we can see the results of the hypothesis tests for β_1 and β_0. Specifically, for β_1:

- $\hat{\beta}_1 = -37.154$
- $se(\hat{\beta}_1) = 2.889$
- the test statistic is $t = -12.86$
- the corresponding p-value is $1.64e - 10$.

The null hypothesis for both tests is that the parameter is equal to 0. The p-value is computed with a two-sided alternative hypothesis.

2. Confidence interval for β_1 (and β_0)

To find the 95% confidence intervals for the coefficients

```
confint(result,level = 0.95)
```

```
##                   2.5 %    97.5 %
## (Intercept) 2534.99540 2720.6493
## Age          -43.22338  -31.0838
```

3. Confidence interval for mean response for given x

Suppose we want a confidence interval for the average strength of bonds for 10 week old propellants

```
newdata<-data.frame(Age=10)
predict(result,newdata,level=0.95,
        interval="confidence")
```

```
##        fit      lwr      upr
## 1 2256.286 2206.739 2305.834
```

4. Prediction interval for a response for a given x

For a prediction interval for the strength bonds for a 10 week old propellant

```
predict(result,newdata,level=0.95,
        interval="prediction")
```

```
##        fit      lwr      upr
## 1 2256.286 2048.385 2464.188
```