EDS241: Assignment template/example

Vincent Thivierge

01/18/2022

In this mock assignment, we use the preloaded "mtcars" data in R to investigate the relationship between vehicle miles per gallon (MPG), weight, and number of cylinders. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

1 Clean and plot data

The following code loads and cleans the data.

```
# Load data

data("mtcars")
raw_data <- mtcars

# Clean data

## Add model names as a column
## [this is just an example manipulation, I rarely assign rownames to a column]

clean_data <- tibble::rownames_to_column(raw_data, "model")
```

The code chunk below shows how to produce a scatter plot of MPG against weight.

```
# Plot 1

plot_1 <- ggplot(clean_data, aes(y=mpg, x = wt))+
  geom_point()+
  theme_cowplot(14)+
  labs(x = "Weight (1000 lbs)", y = "Miles per gallon")
```

Figure 1: MPG and vehicle weight

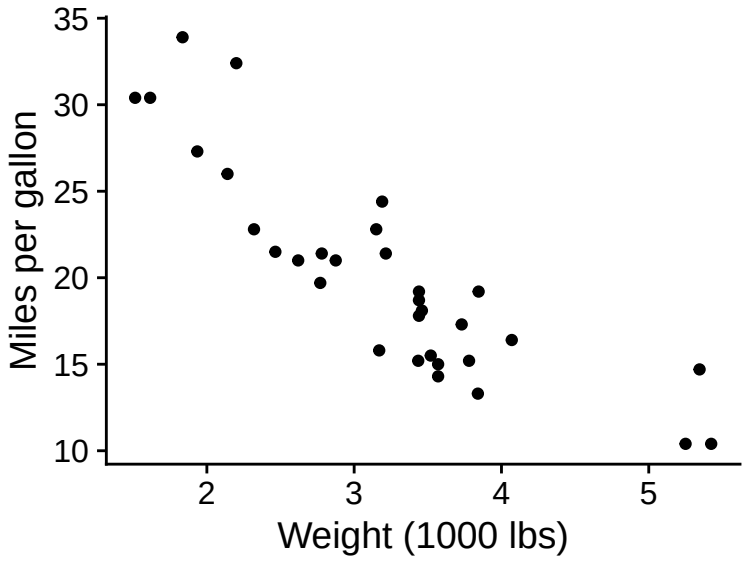

Figure 1 shows the expected negative relationship between vehicle weight and MPG.

2 Run and interpret regression models

In order to more formally analyze the relationship between MPG, vehicle weight, and cylinders we estimate the following regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \tag{1}$$

where Y_i is MPG for vehicle model i, X_{1i} is the vehicle weight, X_{2i} is the number of cylinders in the engine, and u_i the regression error term. We will consider a regression including only vehicle weight, and a regression including vehicle weight and number of cylinders.

In R, we run the following code:

```
model_1 <- lm(mpg ~ wt, data=clean_data)
model_2 <- lm(mpg ~ wt + cyl, data=clean_data)
```

Table 1 shows the estimated coeffients from estimating equation (1).

Table 1: MPG and vehicle weight

	MPG	
	(1)	(2)
Weight (1000 lbs)	-5.344^{***}	-3.191^{***}
	(0.559)	(0.757)
Cylinders		-1.508^{***}
		(0.415)
Observations	32	32
R^2	0.753	0.830
Note:	*p<0.1; **p<0.05; ***p<0.01	

In model (1), the estimated β_1 coefficient implies that a 1000 pound increase in vehicle weight reduces miles per gallon by 5.3 miles. Adding the number of cylinders in model (2) reduces $\hat{\beta}_1$ from -5.3 to -3.2.