Applying the Regression Anatomy Formula

w203: Statistics for Data Science

Open R and verify that the mtcars dataset is automatically loaded.

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

You are interested in the following model:

$$mpg = \beta_0 + \beta_1 hp + \beta_2 wt + u$$

Your task is to figure out what the ols estimate for β_1 would be, without conducting a multiple regression in R, instead only relying on simple regressions.

Applying the regression anatomy formula,

$$\hat{\beta}_1 = \frac{\hat{cov}(y, r_1)}{\hat{var}(r_1)}$$

where r_1 is the residual from a regression of hp on wt.

1. Use R to compute the residuals r_1, then compute the numerator of the regression anatomy formula.

```
y = mtcars$mpg
x1 = mtcars$hp
x2 = mtcars$wt
model = lm(x1~x2)
r1 = model$residuals
num = cov(y,r1)
num
```

```
## [1] -84.54563
```

2. Compute the denominator of the regression anatomy formula.

```
den = var(r1)
den
```

```
## [1] 2660.931
```

3. Explain, in a sentence or two, why this method works to compute β_1.

```
num/den
```

```
## [1] -0.03177295
```

The key point to understand is that the residuals from regressing hp on wt are the important signal. We're used to thinking of residuals as noise, but in this case they represent the unique variation in hp, and that's what ols regression needs to work.