Bivariate Uncertainty

Nicholas Ray

2023-09-22

This week we'll continue to use our toy example dataset, a random sample of 5 observations from the `mtcars` dataset pre-loaded in `R`. To generate this dataset, and again to fit a linear model between *mpg* and *hp*, I use the following commands. The summary output for the model is given below as well.

```
library(tidyverse)
set.seed(1)
data<-slice_sample(mtcars,n=5)
model<-lm(mpg~hp,data)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = data)
##
## Residuals:
## Pontiac Firebird   Hornet 4 Drive       Duster 360        Mazda RX4
##           1.0267           0.0892          -0.4944          -0.3108
##    Mazda RX4 Wag
##          -0.3108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.620484   0.937229  28.403 9.58e-05 ***
## hp          -0.048270   0.005882  -8.207  0.00379 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.707 on 3 degrees of freedom
## Multiple R-squared:  0.9574, Adjusted R-squared:  0.9431
## F-statistic: 67.35 on 1 and 3 DF,  p-value: 0.003787
```

Standard Errors

Since we are focusing on uncertainty this week, let's analytically confirm the standard errors for our slope given by `R`:

$$\text{s.e.}(\hat{\beta}_2) = \sqrt{\frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \tag{1}$$

$$= \sqrt{\frac{\frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-2}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \tag{2}$$

$$= \sqrt{\frac{\frac{(1.03)^2+(0.09)^2+(-0.49)^2+(-0.31)^2+(-0.31)^2}{3}}{(175-150)^2+(110-150)^2+(245-150)^2+(110-150)^2+(110-150)^2}} \tag{3}$$

$$= \sqrt{\frac{\frac{1.5}{3}}{12,850}} \tag{4}$$

$$= 0.0062. \tag{5}$$

As you can see, the standard error we computed by hand is very close to the one given by `R`, given rounding error.

Hypothesis Test

Now we may want to conduct statistical inference and see if our slope estimate is statistically significant. First, we will conduct a hypothesis test. Our test will be to see if our coefficient is statistically different from our specified null hypothesis, say 0. We will reject this null hypothesis and favor the alternative hypothesis ($\hat{\beta}_2 \neq 0$) if the absolute value of our t statistic is larger than the critical value of a Student t distribution (t_{crit}) at a certain confidence (significance) level, say 95%. Dougherty (2016, p. 548) states this t_{crit} is 3.182.

$$t = \frac{\hat{\beta}_2 - \beta_2^0}{\text{s.e.}(\hat{\beta}_2)} \tag{6}$$

$$= \frac{-0.048}{0.0062} \tag{7}$$

$$= -7.74 \tag{8}$$

$$|t| > 3.182 \tag{9}$$

Thus, we reject the null that the coefficient is 0 since $7.74 > 3.182$. In fact, the absolute value of our t statistic is greater the given t_{crit} value at 99% confidence (5.841), but not at the 99.9% confidence level (12.924). This is confirmed by the "stars'' given in our model output.

Confidence Interval

What would the confidence interval be for our coefficient? At 95% confidence, it would be:

$$\begin{aligned} CI &= \hat{\beta}_2 \pm (\text{s.e.}(\hat{\beta}_2) \times t_{crit}) && (10)\\ &= -0.048 \pm (0.0062 \times 3.182) && (11)\\ &= -0.048 \pm 0.02 && (12)\\ &= -0.068 < -0.048 < -0.028. && (13) \end{aligned}$$

This is confirmed by the following `R` output, excepting for rounding error:

```
confint(model)
```

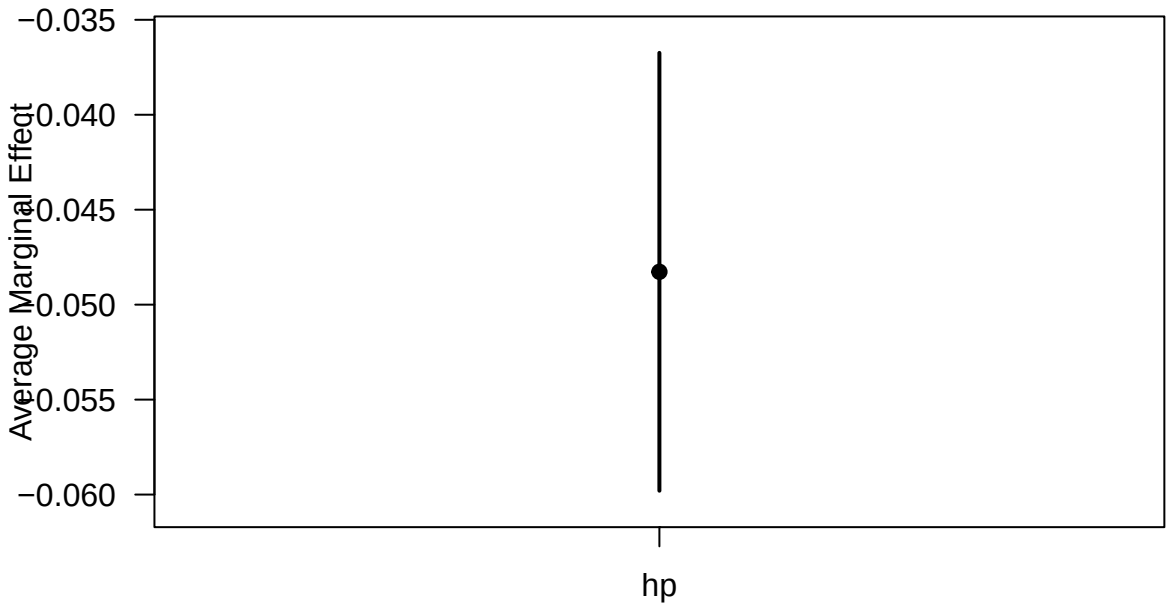
```
##                   2.5 %     97.5 %
## (Intercept) 23.63780249 29.6031664
## hp          -0.06698859 -0.0295512
```

We could also calculate the confidence interval and plot it using the `margins` package:

```
library(margins)
model %>% margins() %>% summary()
```

```
##  factor     AME     SE       z      p   lower   upper
##      hp -0.0483 0.0059 -8.2066 0.0000 -0.0598 -0.0367
```

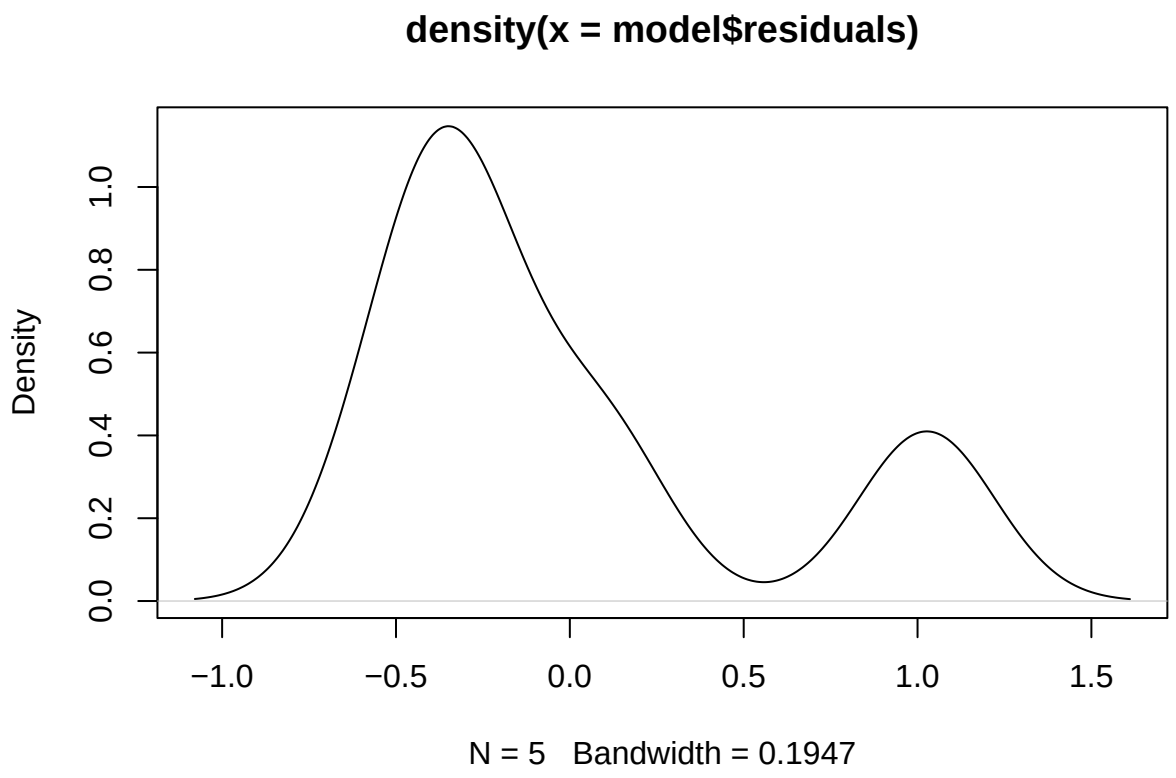
```
model %>% margins() %>% plot()
```



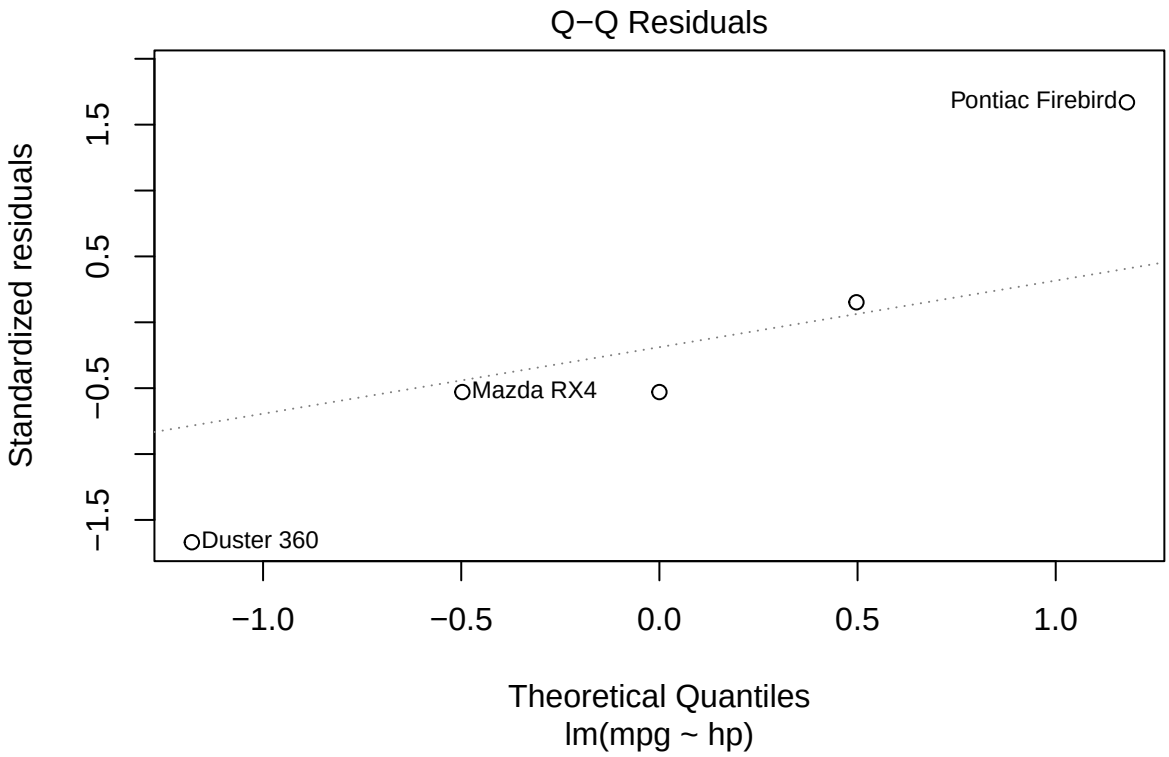
Looking at Residuals

To touch on the plots that Paul was talking about in class, we can look at the residuals for our model.

```
plot(density(model$residuals))
```




```
plot(model,2)
```


Q−Q Residuals
Standardized residuals
1.5
0.5
−0.5
−1.5
Pontiac Firebird
Mazda RX4
Duster 360
−1.0
−0.5
0.0
0.5
1.0
Theoretical Quantiles
lm(mpg ~ hp)