Multiple Regression Applications Solutions

Lt Col Ken Horton Professor Bradley Warner

11 August, 2020

1. The `mtcars` dataset contains average mileage (mpg) and other information about specific makes and models of cars. (This dataset is built-in to `R`; for more information about this dataset, reference the documentation with `?mtcars`).

   a. Build and interpret the coefficients of a model fitting `mpg` against displacement (`disp`), horsepower (`hp`), rear axle ratio (`drat`), and weight in 1000 lbs (`wt`).

```
cars_mod<-lm(mpg~disp+hp+drat+wt,data=mtcars)
summary(cars_mod)
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + drat + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5077 -1.9052 -0.5057  0.9821  5.6883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.148738   6.293588   4.631  8.2e-05 ***
## disp         0.003815   0.010805   0.353  0.72675
## hp          -0.034784   0.011597  -2.999  0.00576 **
## drat         1.768049   1.319779   1.340  0.19153
## wt          -3.479668   1.078371  -3.227  0.00327 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.602 on 27 degrees of freedom
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8136
## F-statistic: 34.82 on 4 and 27 DF,  p-value: 2.704e-10
```

$$\mathrm{E(mpg)} = 29.15 + 0.004 * \mathrm{disp} - 0.035 * \mathrm{hp} + 1.768 * \mathrm{drat} - 3.480 * \mathrm{wt}$$

Each coefficient represents the expected increase in `mpg` for a unit increase in the respective variable, leaving all other variables constant.

   b) Given your model, what is the expected mpg for a vehicle with a displacement of 170, a horsepower of 100, a `drat` of 3.80 and a wt of 2,900 lbs. Construct a 95% confidence interval and prediction interval for that expected mpg.

```
predict(cars_mod,newdata=data.frame(disp=170,hp=100,drat=3.8,wt=2.9),interval="confidence")
```

```
##        fit      lwr      upr
## 1 22.94652 21.76569 24.12735
```

```
predict(cars_mod,newdata=data.frame(disp=170,hp=100,drat=3.8,wt=2.9),interval="prediction")
```

```
##        fit      lwr      upr
## 1 22.94652 17.47811 28.41494
```

c. Repeat part (b) with a bootstrap.

```
set.seed(732)
results <- do(1000)*lm(mpg~disp+hp+drat+wt,data=resample(mtcars))
```

```
head(results)
```

```
##   Intercept          disp          hp      drat        wt    sigma r.squared
## 1  20.28185 -0.0017783926 -0.02620557 2.8897199 -2.016023 2.300320 0.8210210
## 2  33.66818 -0.0128734640 -0.01960700 0.3309526 -2.862143 2.102123 0.8760470
## 3  25.54583 -0.0001983653 -0.03743247 2.1905290 -2.392723 2.814197 0.7947977
## 4  33.22436  0.0112562822 -0.04151242 1.4796584 -4.557451 2.699223 0.8523512
## 5  25.96975 -0.0002882344 -0.03093247 2.3689688 -2.861715 2.670069 0.8643060
## 6  34.17742  0.0054509451 -0.04085537 0.9019272 -3.975107 2.562486 0.8377108
##          F numdf dendf .row .index
## 1 30.96393     4    27    1      1
## 2 47.70611     4    27    1      2
## 3 26.14436     4    27    1      3
## 4 38.96659     4    27    1      4
## 5 42.99429     4    27    1      5
## 6 34.84241     4    27    1      6
```

```
results %>%
mutate(pred=Intercept+disp*170+hp*100+drat*3.8+wt*2.9) %>%
cdata(~pred,data=.)
```

```
##         lower    upper central.p
## 2.5% 21.76703 24.15803      0.95
```

```
results %>%
mutate(pred=Intercept+disp*170+hp*100+drat*3.8+wt*2.9) %>%
cbind(resid=sample(cars_mod$residuals,size=1000,replace = TRUE)) %>%
mutate(pred_ind=pred+resid) %>%
cdata(~pred_ind,data=.)
```

```
##         lower    upper central.p
## 2.5% 19.42663 28.90335      0.95
```

2. Is that the best model for predicting mpg? With your group, try a variety of different models. You could explore higher order terms or even interactions. One place to start is my using the `pairs()` function on `mtcars` to plot a large pairwise scatterplot. How high could you get adjusted R-squared? Keep in mind that is only one measure of fit.

Answers will vary, but I tried this and got 0.8694.

```
summary(lm(mpg~disp+I(disp^2)+hp+I(hp^2)+wt,data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ disp + I(disp^2) + hp + I(hp^2) + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1591 -1.4907 -0.3903  1.5851  3.7795
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.440e+01  2.639e+00  16.823 1.71e-15 ***
## disp        -4.532e-02  2.131e-02  -2.127 0.043100 *
## I(disp^2)    8.844e-05  3.315e-05   2.668 0.012967 *
## hp          -8.652e-02  3.813e-02  -2.269 0.031813 *
## I(hp^2)      1.585e-04  8.932e-05   1.775 0.087666 .
## wt          -3.517e+00  8.874e-01  -3.963 0.000515 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.178 on 26 degrees of freedom
## Multiple R-squared:  0.8904, Adjusted R-squared:  0.8694
## F-statistic: 42.26 on 5 and 26 DF,  p-value: 1.129e-11
```

File Creation Information

- File creation date: 2020-08-11
- Windows version: Windows 10 x64 (build 18362)
- R version 3.6.3 (2020-02-29)
- `mosaic` package version: 1.7.0
- `tidyverse` package version: 1.3.0
- `openintro` package version: 2.0.0