Seif Kungulio
10/31/2024
Project 2
DATA 640
Section: 01W
Instructor: Chris Shannon
File Name: Project2_Kungulio_Seif.docx

1. Use the lm() function to perform a simple linear regression with the response mpg and the predictor hp.

   # Build the model
   simple_model = lm(mpg ~ hp, data = mtcars)

2. Is there a relationship between the target mpg and predator hp?

```
> #print out the model results
> summary(simple_model)

Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6657 -1.2794 -0.7977  0.8162  8.7977

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.297928   1.314115  22.295  < 2e-16 ***
hp          -0.064548   0.008429  -7.658 2.72e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.5 on 39 degrees of freedom
Multiple R-squared:  0.6006,    Adjusted R-squared:  0.5904
F-statistic: 58.65 on 1 and 39 DF,  p-value: 2.723e-09
```

   The simple linear regression model shows a significant relationship between “mpg” (response) and “hp” (predictor). This evident from the coefficient table in the simple model output:

   - t-value for hp: -7.658
   - p-value for hp: 2.723e-09

   The very small p-value indicates that “hp” is a statistically significant predictor of “mpg”.

3. How strong is the relationship between the response and predictor?

   The strength of the relationship is represented by the Multiple R-squared value of 0.6006. This indicates that approximately 60.06% of the variance in “mpg” is explained by the linear relationship with “hp”. The Adjusted R-squared value of 0.5904 is also a strong indicator of the model's effectiveness.

4. Is the relationship between mpg and hp positive or negative?

   The coefficient for hp is -0.064548, which is negative. This indicates that the relationship between “mpg” and “hp” is negative. As “hp” increases, “mpg” decreases.

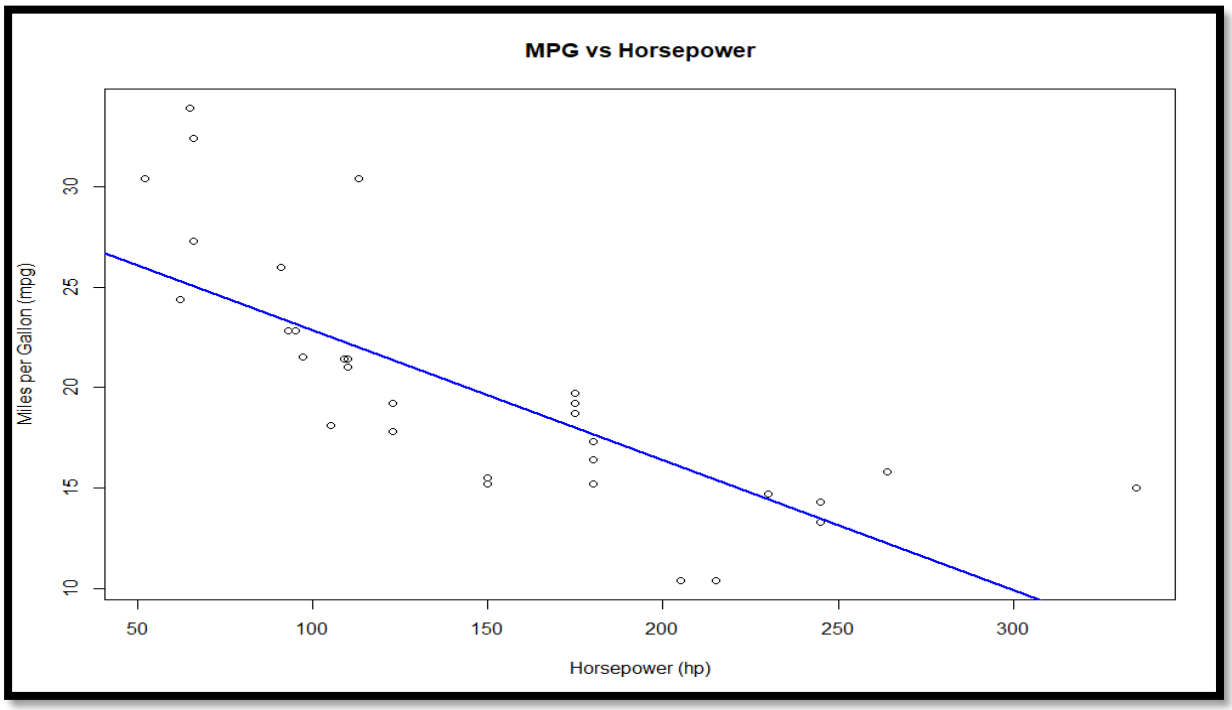
5. What is the predicted mpg associated with a horsepower (hp) of 100? What’s the 95% confidence interval for the predicted mpg?

```
> # Predict mpg for hp = 100 with confidence interval
> predicted_mpg <- predict(simple_model, data.frame(hp = 100),
+                          interval = "confidence", level = 0.95)
>
> # Print the predicted mpg
> predicted_mpg
       fit     lwr      upr
1 22.84317 21.5279 24.15844
>
```

   The predicted “mpg” for a car with “hp” of 100 is 22.84317, with a 95% confidence interval ranging from 21.5279 to 24.15844. This interval reflects the uncertainty around the predicted mean “mpg” value due to the variability in the data.

6. Plot the response and the predictor and add the regression line using abline().

```
# Plot the response and predictor with the regression line
plot(mtcars$hp, mtcars$mpg, main = "MPG vs Horsepower",
    xlab = "Horsepower (hp)", ylab = "Miles per Gallon (mpg)")
abline(simple_model, lwd = 2, col = "blue")
```



The plot of “mpg” vs. “hp” clearly indicates a negative relationship, with a regression line added using abline(). The plotted line confirms the decreasing trend of “mpg” as “hp” increases.

7. Perform a multiple linear regression with mpg as the response and the predictors cyl, disp, hp, wt, vs, and gear. Print out the results using summary() function.

```
> # Print the summary of the model
> summary(multiple_model)

Call:
lm(formula = mpg ~ cyl + disp + hp + wt + vs + gear, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5877 -1.8021 -0.3745  0.8538  6.3448

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.29705    6.58333   5.058 1.45e-05 ***
cyl         -0.74909    0.74433  -1.006  0.32133
disp         0.01733    0.01086   1.596  0.11981
hp          -0.03352    0.01381  -2.428  0.02061 *
wt          -3.85230    0.88681  -4.344  0.00012 ***
vs           0.70816    1.39192   0.509  0.61420
gear         1.14054    0.95373   1.196  0.24002
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.348 on 34 degrees of freedom
Multiple R-squared:  0.8433,    Adjusted R-squared:  0.8157
F-statistic:  30.5 on 6 and 34 DF,  p-value: 2.568e-12
```

8. Is there a relationship between the predictors and the response?

   The output of the multiple regression model suggests that there is a relationship between the predictors and mpg:
   - The Multiple R-squared is 0.8433, indicating that 84.33% of the variance in “mpg” is explained by the predictors.
   - The F-statistic has a p-value of 2.568e-12, indicating that at least one predictor has a statistically significant relationship with “mpg”.

9. Which predictors appears to have a statistically significant relationship to the response?

   In the multiple linear regression summary, predictors with p-values below 0.05 are statistically significant. They include:
   - hp (p-value = 0.02061)
   - wt (p-value = 0.00012)

   These predictors are marked with significance stars * indicating their statistical significance in predicting “mpg”.

10. Use * symbols to fit linear regression models with interaction effects between hp and wt. Does this interaction appear to be statistically significant?

    # Fit a model with interaction effects between hp and wt
    interaction_model <- lm(mpg ~ hp * wt, data = mtcars)

```
> # Print the summary of the interaction model
> summary(interaction_model)

Call:
lm(formula = mpg ~ hp * wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9170 -1.6420 -0.7411  1.4507  4.7557

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.66145    3.29797  14.755  < 2e-16 ***
hp          -0.11778    0.02277  -5.174 8.24e-06 ***
wt          -7.76370    1.12887  -6.877 4.13e-08 ***
hp:wt        0.02642    0.00673   3.926 0.000362 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.048 on 37 degrees of freedom
Multiple R-squared:  0.8703,    Adjusted R-squared:  0.8598
F-statistic: 82.77 on 3 and 37 DF,  p-value: < 2.2e-16
```

    The interaction model reveals a statistically significant relationship between “mpg” and the predictors “hp” and “wt”, as well as their interaction. This model accounts for a substantial portion of the variability in “mpg”, as evidenced by a strong R-squared value. The inclusion of an interaction term indicates that the combined effects of horsepower and weight are critical in understanding their impact on fuel efficiency.

    Specifically, the interaction term “hp:wt” is statistically significant, with:

    - Coefficient: 0.02642
    - t-value: 3.926
    - p-value: 0.000362

    The low p-value indicates that the interaction between “hp” and “wt” significantly influences the prediction of “mpg”. This finding suggests that changes in “mpg” depend not only on the individual effects of horsepower and weight but also on how these variables interact with each other.