Steven Shaw and Cooper Smith

Homework 1 report

Link to repo:
https://github.com/CAIS380-ML-S24/hw01-py-steven-shaw.git

Explain what your error rate indicates about each model. Do you think the model is a good fit? Why or why not?

- The error rate for the chocolate actual against predicted is 51.7%
  I calculated the MSE of the actual vs predicted and got 2699.6219904602667

Is there a relationship between the predictor and the response?

- With the P value being zero, this suggests that changes in horsepower are associated with changes in mpg, rejecting the null hypothesis of no relationship. While a p-value of zero typically means the probability is extremely low, not actually zero, it underscores the confidence in the observed association between these variables.

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

- The R2 value is 0.606. This indicates a moderately strong relationship between horsepower and mpg. The closer the R2 value is to 1, the stronger the relationship, so a value of 0.606 suggests that while horsepower is a significant predictor of mpg, there are other factors not included in the model that also affect mpg.

Is the relationship between the predictor and the response positive or negative?

- Given that the regression coefficient for horsepower is -0.1578, this indicates a negative relationship between horsepower and mpg in our model. Specifically, for each one-unit increase in horsepower, mpg is expected to decrease by 0.1578 units, holding all other variables constant. This result aligns with the general understanding that higher horsepower in vehicles is often associated with lower fuel efficiency.

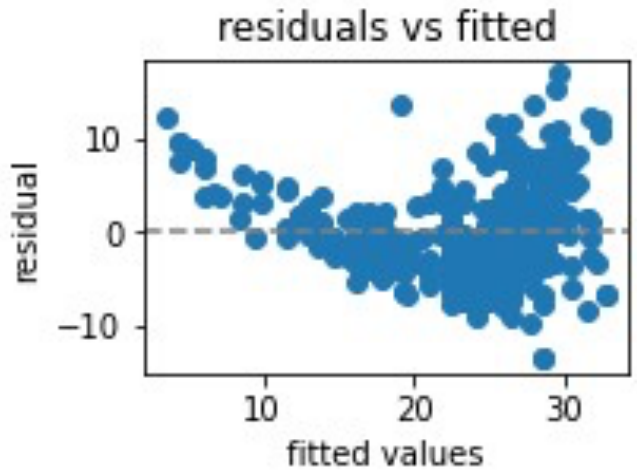
What is the predicted mpg associated with a horsepower of 98? What are the associated 95% confidence and prediction intervals?

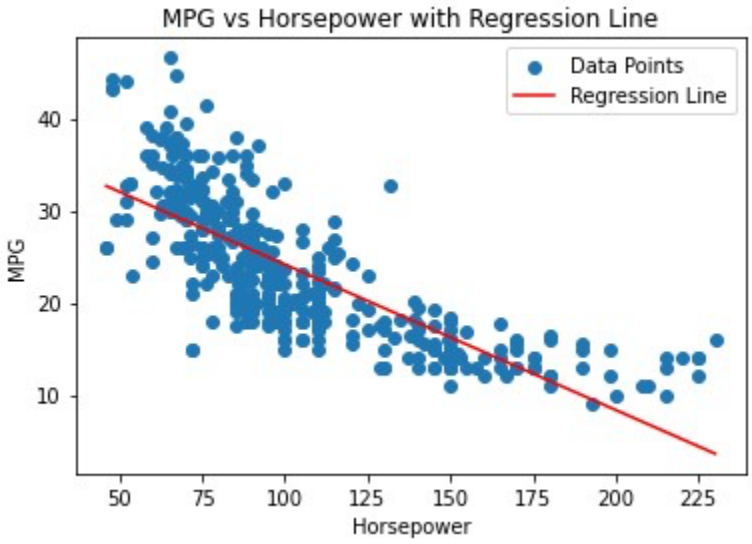
- The predicted mpg associated with a horsepower of 98 was 24.46707715.

The associated 95% confidence and prediction intervals are 95% Confidence Interval:

Mean: 24.46707,Mean_se: 0.251262 , obs_ci_lower: 14.809396 ,obs_ci_upper: 34.124758

Our Plots





Produce the diagnostic plots for your least squares regression fit. Do you see any issues with the fit?

- The plot of the residuals (the differences between actual and predicted values) against the predicted values shows a pattern that shouldn't be there if the model was a perfect fit. It looks like the model isn't capturing the true

relationship between horsepower and mpg entirely. The spread of the residuals isn't even, suggesting that the model's accuracy varies across different levels of horsepower. This could mean that a more complex model might work better for the data. More checks would be useful to figure out exactly what's going on and how to fix it.