Khyati Naik: Data 605 - HW11

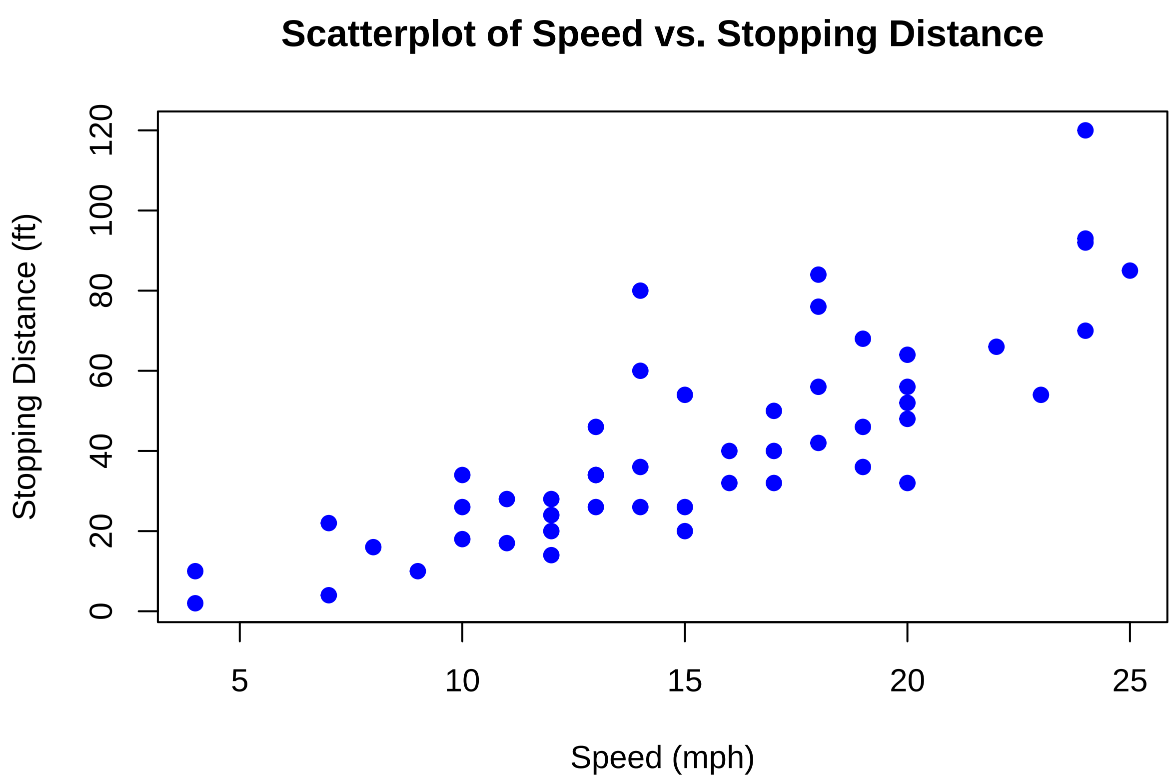
Nov 11, 2023

Contents

Question

Using the "cars" dataset in R, build a linear model for stopping distance as a function of speed and replicate the analysis of your textbook chapter 3 (visualization, quality evaluation of the model, and residual analysis.)

```r
# Scatterplot: Speed vs. Stopping Distance
plot(cars$speed, cars$dist, main = "Scatterplot of Speed vs. Stopping Distance",
     xlab = "Speed (mph)", ylab = "Stopping Distance (ft)", pch = 19, col = "blue")
```



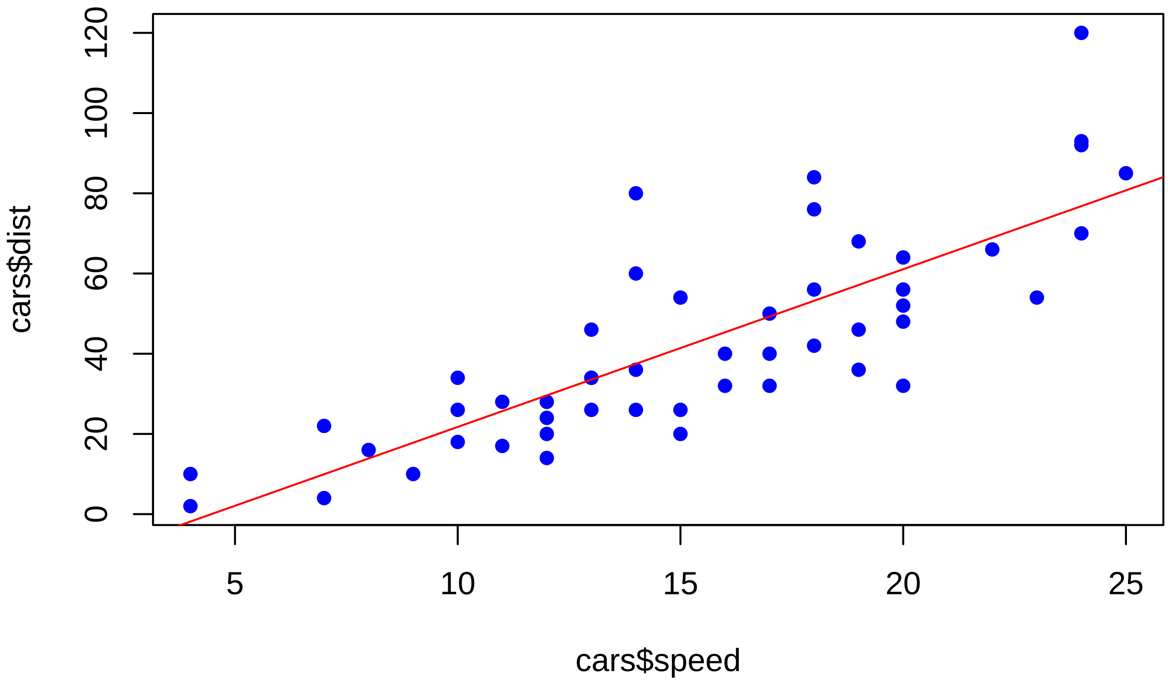
The plot shows a clear positive linear relationship. As speed increases, stopping distance tends to increase. It is evident that a linear regression model is appropriate for modeling this relationship.

```r
# Fit a Linear Model
model <- lm(dist ~ speed, data = cars)

# Summary of the Linear Model
summary(model)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

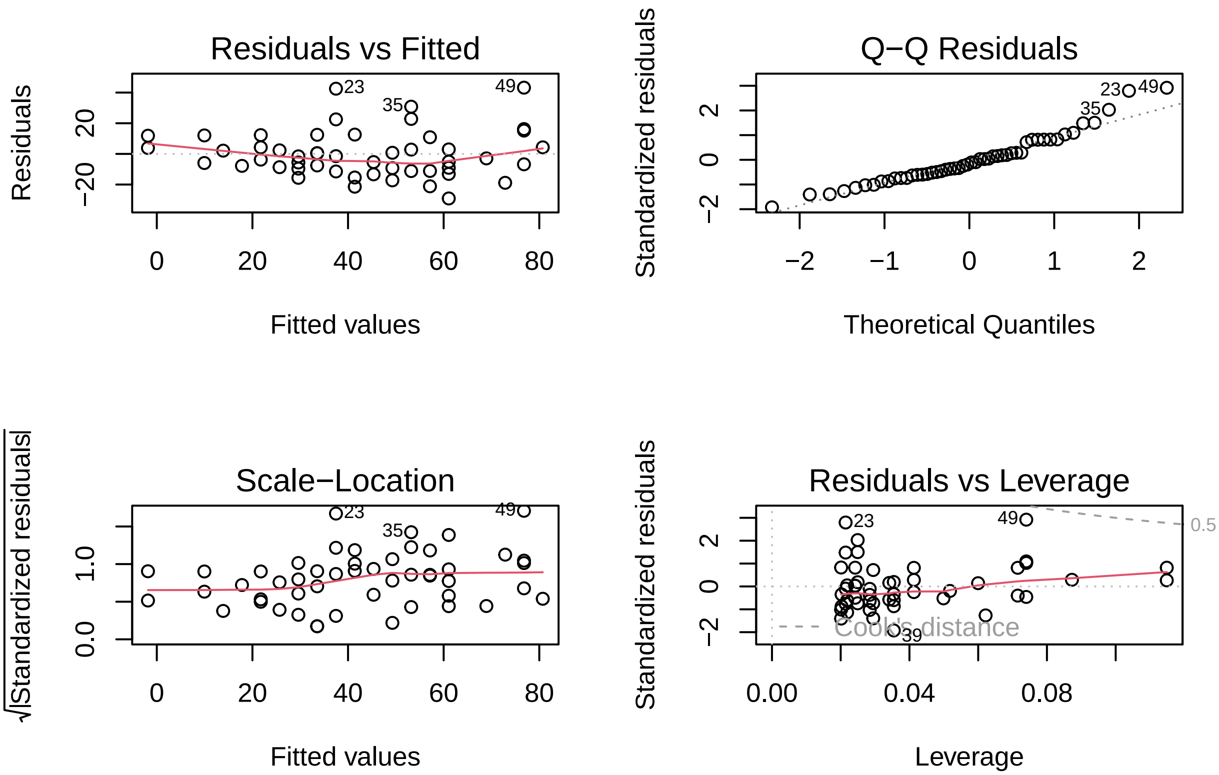
- The model relates "dist" (stopping distance) to "speed."
- The "speed" coefficient is approximately 3.9324, suggesting that for each 1 mph increase in speed, stopping distance increases by around 3.9324 feet.
- The "summary" function provides a summary of the linear model, including coefficients, p-values, and model fit statistics.

```r
# Overlay Regression Line on Scatterplot
# Create the scatterplot
plot(cars$speed, cars$dist, pch = 16, col = "blue")

# Overlay the regression line
abline(model, col = "red")
```



The regression line visually represents the linear relationship between speed and stopping distance as modeled by the linear regression.

```
# Residual Analysis Plots
par(mfrow = c(2, 2))
plot(model)
```



Four types of residual plots are generated:

- **Residuals vs. Fitted:** This plot checks for linearity. No specific patterns are observed, indicating that the linearity assumption holds.
- **Normal Q-Q:** This plot assesses the normality of residuals. The residuals generally follow a straight line, suggesting that the normality assumption is met.
- **Scale-Location:** This plot examines the constant variance assumption (homoscedasticity). Residuals are roughly spread evenly, indicating this assumption is reasonable.
- **Residuals vs. Leverage:** This plot identifies potential outliers or influential points. No extreme outliers are evident.

Conclusion:

- The linear model for stopping distance as a function of speed appears to be a good fit.
- The coefficient for speed is positive and statistically significant, suggesting that increasing speed leads to increased stopping distance.
- The R-squared value indicates that the model explains a substantial portion of the variance in stopping distance (approximately 65.11%).
- Residual analysis plots show that the model assumptions are reasonably met, which further supports the model's validity.

This analysis provides valuable insights into the relationship between speed and stopping distance, which can be essential for understanding and predicting vehicle safety.