predictive_analysis 2_roll_703

Priyanshu Dey 703

2026-02-05

Problem 4

4. Comparing Several Simple Linear Regressions

We analyze the `Boston` dataset to predict median home values (`medv`) using four distinct predictors: per capita crime rate (`crim`), nitrogen oxides concentration (`nox`), proportion of blacks (`black`), and lower status of the population (`lstat`).

(a) Regression Models Output

We load the necessary libraries

```
library(MASS)

data("Boston")
predictors = c("crim", "nox", "black", "lstat")
results = data.frame()

for(var in predictors){
  formula = as.formula(paste("medv ~", var))
  model = lm(formula, data = Boston)
  summ = summary(model)

  results = rbind(results, data.frame(
    Predictor = var,
    Intercept = round(coef(model)[1], 4),
    Coefficient = round(coef(model)[2], 4),
    Std_Error = round(summ$coefficients[2, 2], 4),
    R_Squared = round(summ$r.squared, 4),
    RSE = round(summ$sigma, 4)
  ))
}

results
```

```
##             Predictor Intercept Coefficient Std_Error R_Squared    RSE
## (Intercept)       crim   24.0331     -0.4152    0.0439    0.1508 8.4838
## (Intercept)1       nox   41.3459    -33.9161    3.1963    0.1826 8.3233
## (Intercept)2     black   10.5510      0.0336    0.0042    0.1112 8.6793
## (Intercept)3     lstat   34.5538     -0.9500    0.0387    0.5441 6.2158
```

(b) Best Fit Analysis

Based on the table above, the model using **`lstat`** gives the best fit.

- **R-Squared**: `lstat` has the highest of 0.5441, explaining approximately 54% of the variance in median home values.
- The next closest is nox at 0.1826 (18%).
- **Residual Standard Error (RSE)**: `lstat` also yields the lowest prediction error (RSE = 6.2158).

(c) Coefficient Interpretation

We examine the coefficients for usefulness and directionality:

1. **`lstat` (Coef: -0.95)**: This is the most useful predictor. The negative coefficient indicates that as the percentage of lower-status population increases, median home values decrease significantly.
2. **`nox` (Coef: -33.9161)**: The negative relationship suggests that higher pollution (nitrogen oxides) correlates with lower home values.
3. **`crim` (Coef: -0.4152)**: Crime rate shows a negative relationship with home value, though its explanatory power () is far lower than `lstat`.
4. **`black` (Coef: 0.0336)**: While the positive coefficient is statistically valid, this predictor has the lowest () among the four, making it the least useful single predictor in this comparison.