STT 380

In-Class Assignment 25

1. For the pizza dataset,
    a. Build a linear regression model for calories, based on fat.
        i. pizza <- read.csv('pizza.csv')
        ii. head(pizza)
        iii. summary(lm(data = pizza, cal ~ fat))
        iv. Call:
        v. lm(formula = cal ~ fat, data = pizza)
        vi.
        vii. Residuals:
        viii. Min 1Q Median 3Q Max
        ix. -0.60089 -0.36161 -0.08474 0.32265 0.83369
        x.
        xi. Coefficients:
        xii. Estimate Std. Error t value Pr(>|t|)
        xiii. (Intercept) 2.202558 0.057067 38.60 <2e-16 ***
        xiv. fat 0.052816 0.002579 20.48 <2e-16 ***
        xv. ---
        xvi. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
        xvii.
        xviii. Residual standard error: 0.4003 on 298 degrees of freedom
        xix. Multiple R-squared: 0.5846, Adjusted R-squared: 0.5832
        xx. F-statistic: 419.3 on 1 and 298 DF, p-value: < 2.2e-16
    b. How does the RMSE compare to that for the naïve model?
    c. Now, add moisture to the model. Has the model significantly improved? How can you tell?
        i. summary(lm(data = pizza, cal ~ fat + mois))
        ii. Adjusted r squared is now .99 so yes it improved

d. Now, add sodium to the model from (c). Has the model significantly improved? How can you tell?

    i. summary(lm(data = pizza, cal ~ fat + mois + sodium))

    ii. Yes r squared is now .9989

2. For the cars dataset, build a model to predict mpg based on acceleration.

    a. Build a confidence interval for the acceleration parameter based on t values

    b. Build a confidence interval for the acceleration parameter based on regular bootstrapping

    c. Build a confidence interval for the acceleration parameter based on Bayesian bootstrapping

    d. How do the CI's compare?

    e. Build a linear regression model, using acceleration and weight to predict mpg. Perform sensitivity analysis to determine the impact of each x variable on the model.