Stat 509: Statistics for Engineers

Homework Assignment 9

Make sure to include the R code and graphs used in your analyses.

1. In simple linear regression, the estimators of β_0 and β_1 obtained using the *least squares* method are *best* in what sense?

They are the estimators that minimize the sum of the squared residuals, making them the best estimators in terms of fitting the data.

2. What are the assumptions on the error term ϵ?

Linearity, Independent Samples, Variance of Errors, and Normality.

3. Under our assumptions on ϵ, Y_i follows what specific distribution?

Under these assumptions, Yi follows a normal distribution with mean B_0+B_1Xi and constant variance o^{-2}, where Yi is observed value of the dependent variable (y) corresponding to the ith value of the independent variable (x).

4. You have collected data on the fuel consumption in tons per hour (t/h) and the load weight in tons for 100 Boeing 777s. You wish to use this data to make inferences on the fuel consumption of a Boeing 777 for a given load weight.

   (a) Create a scatterplot of the data, making sure the *response* variable is on the y-axis. Does there appear to be a relationship between the weight of a load and the rate of fuel consumption?

Plot(fuel,load)

The scatterplot appears to be all over the place. I think there is little relationship between the two variables.

   (b) Fit a simple linear regression model to the data and add the corresponding regression line to your scatterplot. Visually, does your model appear to fit the data well?

Reg.model=lm(fuel,load)

Abline(reg.model,col="red",lwd=2)

Visually, the model does not appear to fit the data well. The line barely connects with any of the data.

   (c) What is your fitted regression equation (model)?

Summary(reg.model)

6.79+0.26*Load

   (d) Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ in the context of the problem.

B0 = 6.79

B1 = 0.26

   (e) Predict the fuel consumption of a Boeing 777 that has a load weight of 30 tons.

Predict(reg.model,data.frame(load=30))

14.69

   (f) What would the residual be in (e) if the fuel was actually consumed at a rate of 15 t/h.

15 – predict(reg.model,data.frame(load=30))
-0.31

   (g) Estimate the error variance σ^2.

Summary(reg.model)$sigma^2

4.35

   (h) Calculate and interpret a 99% confidence interval for β_1.

Confint(reg.model,level=0.99)[2,]

0.16-0.37

   (i) Based on the interval in (h), does it appear that fuel consumption is related to load weight?

Since the confidence interval for B1 does not contain 0, load weight is likely related to fuel consumption.

   (j) Conduct a hypothesis test at a 0.01 level of significance that $\beta_1 = 0$.

Summary(reg.model)$coefficients[2,"t value"]

6.58

(k) Based on the test in (j), does it appear that fuel consumption is related to load weight?

Since the p-value is less than 0.01, we reject the null hypothesis. It appears that fuel consumption is related to load weight.

(l) Do the confidence interval in (h) and hypothesis test in (j) lead to the same conclusion?

Both lead to the conclusion that fuel consumption is related to load weight.

(m) Calculate and interpret a 96% confidence interval for the average fuel consumption of a Boeing 777 that has a load weight in the 95*th* percentile.

Quantile_load = quantile(load,0.95)

Predict(reg.model,data.frame(load=quantile_load),interval="confidence",level=0.96)

13.72-15.40

(n) Calculate and interpret a 96% prediction interval for the fuel consumption of a Boeing 777 that has a load weight in the 95*th* percentile.

Quantile_load = quantile(load,0.95)

Predict(reg.model,data.frame(load=quantile_load),interval="prediction",level=0.96)

10.14-18.98

(o) Is the interval in (m) or (n) wider? Why?

The prediction interval is wider because it considers the uncertainty in individual predictions.

(p) Predict the fuel consumption of a Boeing 777 that has a load weight of 100 tons.

Predict(reg.model,data.frame(load=100))

33.11

(q) Calculate the correlation r between fuel consumption and load weight and describe the information it provides in the context of the problem.

Cor(load,fuel)

0.55

5. You are interested in the speed at which a particle travels through a certain type of solution. You take 40 measurements on particle speed, solution viscosity, and concentration of a certain substance within the solution. You would like to determine if the viscosity or substance concentration affect the speed of the particle.

(a) Fit a linear regression model with no higher-order terms to the data.

Reg.model=lm(Speed~Viscosity+Concentration)

Intercept = 49.78, Viscosity = -0.69, Concentration = -0.31

(b) Calculate R^2 and describe its meaning in the context of the problem.

Summary(reg.model)$r.squared

0.50

(c) Interpret the estimated regression coefficients in the context of the problem.

Summary(reg.model)

49.78 + 0.08*Viscosity + 0.13*Concentration

(d) At the 0.01 level of significance, does either variable appear helpful in predicting particle speed?

Anova_table = anova(reg.model)

Anova_table[,"Pr(>F)"]

1.01e-06 – 1.27e-01

Since the interval is above 0, Viscosity appears to be the more helpful variable.