Stat509 Fall 2014 Homework 8 Instructor: Peijie Hou 11/18/2014

Instruction: Quiz based on this homework will be given on 12/02/2014. Have a good Thanksgiving!

1. We will look at the purity of oxygen (in percent) produced in a chemical distillation process and the percentage hydrocarbons that are present in the main condenser of the distillation unit. The purity of the oxygen will be the response while the percentage of hydrocarbons will be the regressor. The data is given by

Hydrocarbon%: 7.78 0.99 1.02 1.15 1.29 1.46 1.36 0.87 1.23 1.55 1.4 1.19 1.115 0.98 1.01 1.11 1.2 1.26 1.32 1.43 0.95 Purity%: 4.54 90.01 89.05 91.43 93.74 96.73 94.45 87.59 91.77 99.42 93.65 93.54 92.52 90.56 89.54 89.85 90.39 93.25 93.41 94.98 87.33

You can use the following R code:

hydrocarbon<-c(0.99,1.02,1.15,1.29,1.46,1.36,0.87,1.23,1.55,1.4,1.19,1.115,0.98,1.01,1.11,1.2,1.26,1.32, 1.43, 0.95)purity<-c(90.01,89.05,91.43,93.74,96.73,94.45,87.59,91.77,99.42,93.65,93.54,92.52,90.56,89.54,89.85,90.39, 93.25,93.41,94.98,87.33) fit = lm(purity~hydrocarbon) #scatterplot with regression line superimposed plot(hydrocarbon,purity,xlab = "Hydrocarbon%",ylab = "Purity%",pch=16) abline(fit) #residual plot # Residual plot plot(fitted(fit),residuals(fit),pch=16, xlab="Fitted values",ylab="Residuals") abline(h=0) #QQ plot resid<-residuals(fit) qqnorm(resid);qqline(resid) #Find coefficient estimate summary(fit) #ANOVA table anova(fit)

- (a) Run the regression in R. Superimpose a fitted regression line on the scatter plot of Hydrocarbon% versus Purity%. Plot residuals vs x values, which assumption does this plot check? Plot QQ plot for residuals, which assumption does this plot check? Are the assumptions met?
- (b) Identify the sum of squares for the model from the ANOVA table. What does it measure?
- (c) Conduct a hypothesis test for testing $H_0: \beta_1 = 0$ against $H_a: \beta_1 \neq 0$. Calculate a 95% confidence interval for β_1 , interpret your result.
- (d) Write an interpretation of the estimated coefficient $\hat{\beta}_1$.
- (e) Write an interpretation of the coefficient $\hat{\beta}_0$.
- (f) What is the coefficient of correlation between hydrocarbon%(x) and purity%(Y)? What is the percentage of total sample variation explained by linear regression?
- (g) Predict the mean percentage of oxygen purity when 1.3% hydrocarbons is used. Calculate a 95% confidence interval for that predicted mean response by R, interpret your result.predict(fit,data.frame(hydrocarbon=1.3),level=0.95,interval="confidence")
- (h) Calculate a 95% prediction interval for a future observation when hydrocarbon percentage is 1.3% by R, interpret your result.

predict(fit,data.frame(hydrocarbon=1.3),level=0.95,interval="prediction")