LINEAR REGRESSION MODELS W4315

HOMEWORK 3 ANSWERS

February 20, 2010

Instructor: Frank Wood

1. (50 points) [1] Refer to **Copier maintenance** Problem 1.20.
a. Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.
b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the α risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of your test?
c. Are your results in parts (a) and (b) consistent? Explain.
d. The manufacturer has suggested hat the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?
e. Does b_0 give any relevant information here about the "start-up" time on calls-i.e., about the time required before service work is begun on the copiers at a customer location?

Answer:

a. First read the data into Matlab and estimate the coefficients of the regression line
pr1=textread('CH01PR20.txt');
X=pr1(:,2);
Y=pr1(:,1);
avgX = mean(X);
avgY = mean(Y);
$SXX = sum((X - avgX).^2)$;
SXY = sum((X-avgX).*(Y-avgY));
b1 = SXY/SXX;
b0 = avgY- b1*avgX;

[1] This is problem 2.5 in "Applied Linear Regression Models(4th edition)" by Kutner etc.

The estimated intercept of the regression line b_0= -0.5802 and the estimated slope b_1=15.0352. The estimated change in the mean service time when the number of copiers serviced increases by one is 15.0352.

Calculate 90% confident interval for b_1
$sse = sum((Y - b0 - b1 * X).^2)$
n=length(X)
mse=sse/(n-2)
sb1=sqrt(mse/SXX)
t=1.681071
confl=b1-t*sb1
confh=b1+t*sb1
We obtained MSE=79.4506, $S(b_1)$=0.4831, t(0.95, 43)=1.6811
The 90% confidence interval for b_1: (14.2231, 15.8474), which means if repeating this procedure on multiple samples, 90% of the time, the true parameter β_1 would fall inside of this interval.
b. $H_0 : \beta_1 = 0 \quad H_a : \beta_1 \neq 0$
The test statistic: $t = \frac{b_1 - 0}{s(b_j)} = \frac{15.0352}{0.4831} = 31.1232$
The decision rule: reject H_0 if $t > 1.681$, or equivalently, reject H_0 if the p-value< 0.1
The conclusion: there is convincing evidence to reject the hypothesis that there is no linear association between the number of copiers at a location of a call and the time spent by the serviceman for the call.
The p-value of the test is $P(t_{43} > 31.1232) < 0.000001$.
c. The results from part a and part b are consistent. The 90% confidence interval of β_1 does not include 0, so we expect that the hypothesis that $\beta_1 = 0$ at a 10% significance level will be rejected.
d. $H_0 : \beta_1 \leq 14 \quad H_a : \beta_1 > 14$
The test statistic $t = \frac{b_1 - 14}{s(b_j)} = \frac{15.0352 - 14}{0.4831} = 2.1428$
The decision rule: reject H_0 if $t > 1.681$, or equivalently, reject H_0 if $p < 0.05$
The p-value of the test is $P(t_{43} > 2.1428) = 0.0189$.
There is sufficient evidence to reject the hypothesis that the standard is being satisfied by Tri-City.
e. The intercept does not give relevant information on the start-up time for calls. The estimated coefficient b_0 is negative which does not provide meaningful information about the start-up time. A formal t test on the hypothesis that $\beta_0 = 0$ shows no evidence(p-value=0.84) to reject the hypothesis, that is, there is no enough evidence to indicate that the true parameterβ_0 is significantly different from zero.

2. (20 points) [2] Consider the test problem in a normal error regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where:

$$\beta_0 \text{ and } \beta_1 \text{ are parameters}$$
$$X_i \text{ are known constants}$$
$$\epsilon_i \text{ are independent } N(0, \sigma^2)$$

When testing whether or not $\beta_1 = 0$, why is the F test a one-sided test even though H_a includes both $\beta_1 < 0$ and $\beta_1 > 0$? [Hint: refer to the following problem]

Answer:

When testing whether or not $\beta_1 = 0$ for a simple linear regression model,
the test statistic $F^* = \frac{MSR}{MSE}$

$MSR = \frac{SSR}{1} = SSR = \sum_i(\hat{Y}_i - \bar{Y})^2 = \sum_i(b_0 - b_1 X_i - b_0 - b_1\bar{X})^2 = b_1^2 \sum_i(X_i - \bar{X})^2$
$E(SSR) = E[b_1^2 \sum_i(X_i - \bar{X})^2]$
$= \sum_i(X_i - \bar{X})^2 E(b_1^2)$
$= \sum_i(X_i - \bar{X})^2[Var(b_1) + (E(b_1))^2]$
$= \sum_i(X_i - \bar{X})^2[\frac{\sigma^2}{\sum_i(X_i-\bar{X})^2} + \beta_1^2]$
$= \sigma^2 + \beta_1^2 \sum_i(X_i - \bar{X})^2$

As we know, $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$, so $E(SSE) = \sigma^2(n-2)$ and $E(MSE) = \frac{E(SSE)}{n-2} = \sigma^2$
When $\beta_1 = 0, E(SSR) = \sigma^2, \frac{E(MSR)}{E(MSE)} = 1$. The test statistic F^* should be close to 1.
However, when $\beta_1 \neq 0$, regardless of whether $\beta_1 > 0$ or $\beta_1 < 0$, $E(SSR) > \sigma^2, \frac{E(MSR)}{E(MSE)} > 1$ and the F statistic should always be larger than 1. The test is one sided.
From another point of view, the test statistic F can also be written as $F = \frac{SSE_R - SSE_F}{df_R - df_F} \div \frac{SSE_F}{df_F}$ where SSE_R and SSE_F are respectively the sum of squared errors of the reduced model and the full model. df_R and df_F are the degrees of freedoms of the reduced model and the full model.
Even though H_a includes both $\beta_1 < 0$ and $\beta_1 > 0$, the reduced model for both cases is the

[2] This is problem 2.19 in "Applied Linear Regression Models(4th edition)" by Kutner etc.

same, i.e. $Y_i = \beta_0 + \epsilon_i$, therefore the SSE_R, SSE_F ,df_R, df_F and the test statistic for the two cases are also the same, the test is a one sided test.

3. (30 points) [3] Consider the same normal regression model as in problem 2.
a. When testing $H_0: \ \beta_1 = 5$ versus $H_a: \ \beta_1 \neq 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom df_R?
b. When testing $H_0: \ \beta_0 = 2, \ \beta_1 = 5$ versus H_a : not both $\beta_0 = 2$ and $\beta_1 = 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom df_R?

Answer:

a. The reduced model is $Y_i = \beta_0 + 5X_i + \epsilon_i$. The degrees of freedom df_R=n-1.
b. The reduced model is $Y_i = 2 + 5X_i + \epsilon_i$. The degrees of freedom df_R=n.

[3] This is problem 2.57 in "Applied Linear Regression Models(4th edition)" by Kutner etc.