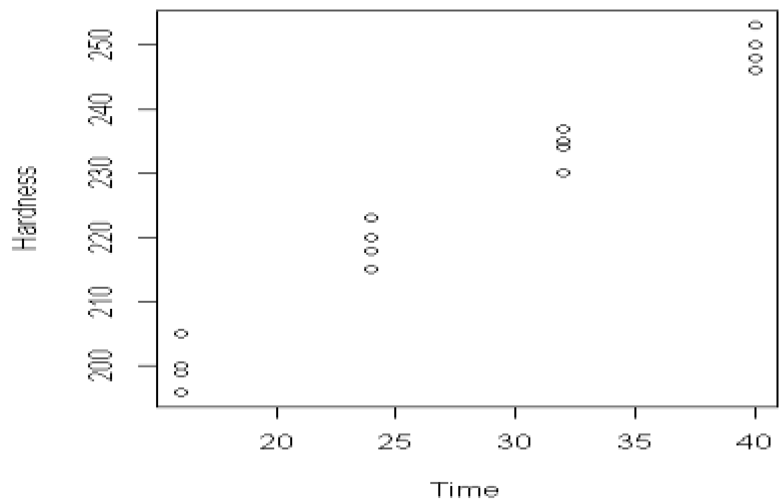
R Codes for Chapter-2

Scatter Plots

Before fitting the simple linear Regression Model, the scatter plot of the response vs Predictor is needed to be drawn to see the pattern of the relationship. The command "plot(X,Y)" can be used to draw the scatter plot.

E.g:- Plastic Hardness example.

```
plot(Time,Hardness) # Plot(X,Y)
```



Here we can see that the relationship between Hardness and Time is linear and so the simple linear model is appropriate.

Confidence and Prediction Intervals

Confidence Intervals for β_0 and β_1

Confidence intervals for β_0 and β_1 are obtained from the command "confint(NAME, level= 1-α)" and NAME is replace with the model assigned for the simple linear model. 95% is the default significance level and so "confint(NAME)" calculates 95% confidence intervals.

E.g:- Plastic Hardness Example.

```
> confint(SLR) # 95% (default significance level) confidence intervals for
                 beeta0 and beet1
                2.5 %    97.5 %
(Intercept) 162.9013 174.29875
Time          1.8405   2.22825
```

So 95% confidence intervals for β_0 and β_1 are (162.901, 174.299) and (1.840, 2.228) respectively.

```
> confint(SLR,level=0.99) # 99% confidence intervals for beeta0 and beet1
                  0.5 %     99.5 %
(Intercept) 160.690457 176.509543
```

```
Time            1.765287   2.303463
```

So 99% confidence intervals for β_0 and β_1 are (160.690, 176.510) and (1.765, 2.303) respectively.

Confidence interval for mean response $E(Y_h)$

First, a new data frame should be created for the value of the predictor variable.

```
> newx=data.frame(Time=20) # Creating a data frame with the new value of the
                             predictor.
> newx
  Time
   1   20
```

Then the following command calculates the confidence interval for the mean response.

```
predict.lm(SLR,newx,interval="confidence",level=0.95) #95% Confidence interval
                                                        at the new X level.
       fit      lwr      upr
1 209.2875 206.961 211.614
```

So the 95% confidence interval for mean response (Hardness) at Time = 20 mins is (206.961, 211.614) and the fitted value of the model is 209.2875.

Prediction interval for a new observation X_h

Here also the data frame for the new predictor value should be created.

```
> newpx=data.frame(Time=22)
> newpx
  Time
1   22
```

Then the following command calculates the corresponding prediction interval.

```
> predict.lm(SLR,newpx,interval="prediction",level=0.99) #99% Confidence
                                                  interval at the new X level.
       fit       lwr       upr
1 213.3562 203.3023 223.4102
```

So the 99% prediction Interval at time =22 mins is (203.3023, 223.4102) and the fitted value of th e model is 213.3562.

Hypotheses Tests

Test statistics and the p-values for the tests $H_0 : \beta_0 = 0 \;\; vs \;\; H_1 : \beta_0 \neq 0$ and $H_0 : \beta_1 = 0 \;\; vs \;\; H_1 : \beta_1 \neq 0$ can be obtained directly from the summary output of the simple linear Regression model.

```
> summary(SLR)

Call:
lm(formula = Hardness ~ Time, data = dataf)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.1500 -2.2188  0.1625  2.6875  5.5750 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 168.60000    2.65702   63.45  < 2e-16 ***
Time          2.03438    0.09039   22.51 2.16e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.234 on 14 degrees of freedom
Multiple R-squared:  0.9731,	Adjusted R-squared:  0.9712 
F-statistic: 506.5 on 1 and 14 DF,  p-value: 2.159e-12
```

So for the hypotheses tests $H_0 : \beta_0 = 0 \ \ vs \ \ H_1 : \beta_0 \neq 0$,

Test Statistic = T = 63.45, and

The P- value < 2e-16 (exact value is unknown),

and, for the hypotheses tests $H_0 : \beta_1 = 0 \ \ vs \ \ H_1 : \beta_1 \neq 0$,

Test Statistic = T = 22.51, and

The P- value = 2.16e-12.

NOTE: p-value for a one sided test is same as one-half of the p-value of a two-sided test.

Therefore, for the hypotheses tests $H_0 : \beta_0 = 0 \ \ vs \ \ H_1 : \beta_0 > 0$, (OR $H_1 : \beta_0 < 0$)

Test Statistic = T = 63.45, and

The P- value < 1e-16 (exact value is unknown),

and, for the hypotheses tests $H_0 : \beta_1 = 0 \ \ vs \ \ H_1 : \beta_1 > 0$, , (OR $H_1 : \beta_1 < 0$)

Test Statistic = T = 22.51, and

The P- value = 1.08e-12.

For all the other hypotheses (i.e., $H_0 : \beta_0 = \beta_{01} \ \ vs \ \ H_1 : \beta_0 \neq \beta_{01}$ and $\beta_{01} \neq 0, etc,$) the test statistics and the p- values should be calculated manually.

Analysis of Variance (ANOVA) Approach

The ANOVA table can be generated using

anova (MODEL NAME),

where the MODEL NAME is the name used when the simple linear regression model is fitted.

Ex. Plastic Hardness (continued….)

```
> anova(SLR)
Analysis of Variance Table

Response: Hardness
          Df Sum Sq Mean Sq F value    Pr(>F)    
Time       1 5297.5  5297.5  506.51 2.159e-12 ***
Residuals 14  146.4    10.5                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here note that the test statistic and the P-value for the F- test are 506.51 and 2.159e-12 respectively. These values can also be obtained from the SUMMARY output above.