R Lab 4. Inference for the Univariate Linear Regression

Let's work with textbook data on Copier Maintenance, problems 1.20, 2.5, etc. The data are available on our Blackboard, but also, on the internet in the public domain. Here is how we can read them in one step.

```
> C = read.table(url("http://statweb.lsu.edu/EXSTWeb/StatLab/DataSets/NKNWData/CH01PR20.txt"))
> head(C)
  V1 V2
1 97  7
2 86  6
3 78  5
4 10  1
5 75  5
6 62  4
> attach(C)
> X=V2; Y=V1;                    # Rename variables
> reg = lm(Y ~ X)                # Fit a regression model of Y on X
```

 # Summary of regression inference, including t-tests, analysis of residuals, F-test, and R^2.

```
> summary(reg)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.6309 -3.2500 -0.2383  4.0235  6.6309 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.3221     2.5644  -0.906    0.379    
X            14.7383     0.5193  28.383  4.1e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.482 on 16 degrees of freedom
Multiple R-squared:  0.9805,    Adjusted R-squared:  0.9793 
F-statistic: 805.6 on 1 and 16 DF,  p-value: 4.097e-15
```

 # Inference about regression coefficients

```
> names(reg)                     # There are several components of the output
 [1] "coefficients"  "residuals"     "effects"       "rank"          "fitted.values" "assign"        "qr"            "df.residual"  
 [9] "xlevels"       "call"          "terms"         "model"        

> reg$coefficients               # Here, we saved both coefficients, b0 and b1
(Intercept)           X 
  -2.322148   14.738255 
> slope = reg$coefficients[2]      # … and we can save just the slope b1, if we want
> slope
```

```
       X
14.73826
```

 # Confidence intervals for regression coefficients

```
> confint(reg)                        # By default, we get 95% confidence intervals for β0 and β1
                2.5 %    97.5 %
(Intercept) -7.758337  3.114041
X           13.637480 15.839030
> confint(reg,level=0.90)             # We can set the desired confidence level
                  5 %      95 %
(Intercept) -6.799213  2.154917
X           13.831693 15.644817
> confint(reg,level=0.99)             # A 99% confidence interval would be wider
                0.5 %    99.5 %
(Intercept) -9.812068  5.167772
X           13.221620 16.254890
```

 # Estimation of mean responses and prediction of individual responses

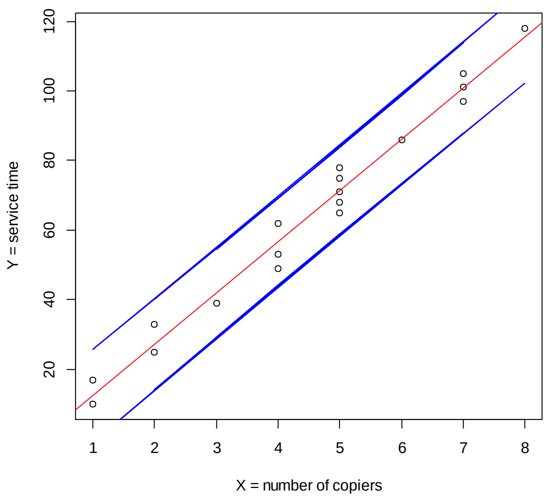
```
> predict(reg, data.frame(X=3))              # Predict the response Y for X=3
       1
41.89262
> predict(reg, data.frame(X=3), interval="confidence") # Confidence interval for E(Y) at X=3
       fit      lwr      upr
1 41.89262 39.11027 44.67496
> predict(reg, data.frame(X=3), interval="prediction") # Prediction interval for Y at X=3
       fit      lwr      upr
1 41.89262 31.99244 51.79279
```

 # Confidence band for the whole regression line

```
> n = length(Y)                     # sample size
> e = reg$residuals                 # residuals
> s = sqrt(sum(e^2)/(n-2))          # estimated standard deviation = root MSE
> s
[1] 4.48188
> W = sqrt( 2*qf(0.95,2,n-2) )
> W
[1] 2.69582
> Yhat = fitted.values(reg)         # Yhat = b0 + b1X
> Sxx = (n-1)*var(X)
> upper.band = Yhat + W*s*sqrt(1/n+(X-mean(X))^2/Sxx)
> lower.band = Yhat - W*s*sqrt(1/n+(X-mean(X))^2/Sxx)

> plot(X,Y,xlab="X = number of copiers",ylab="Y = service time")
```

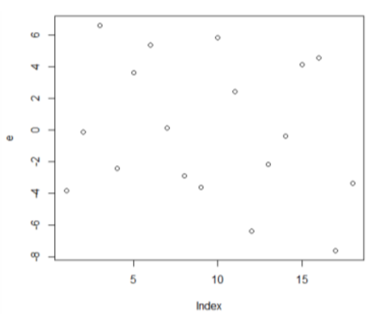
```
> abline(reg,col="red")
> lines(X,upper.band,col="blue")
> lines(X,lower.band,col="blue")
```



 # Regression residuals

```
> e = reg$residuals
```
We can save residuals for further analysis

```
> summary(e)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-7.6310 -3.2500 -0.2383   0.0000   4.0230   6.6310
> plot(e)
```




```
> sum(e)
[1] 9.436896e-16
```
We've proved that this sum is 0

```
> sum(e*X)
[1] 4.374279e-14
```
… and this is 0 as well

 # ANOVA: sums of squares and F-test

```
> anova(reg)
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X          1 16182.6 16182.6  805.62 4.097e-15 ***
Residuals 16   321.4    20.1
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
```