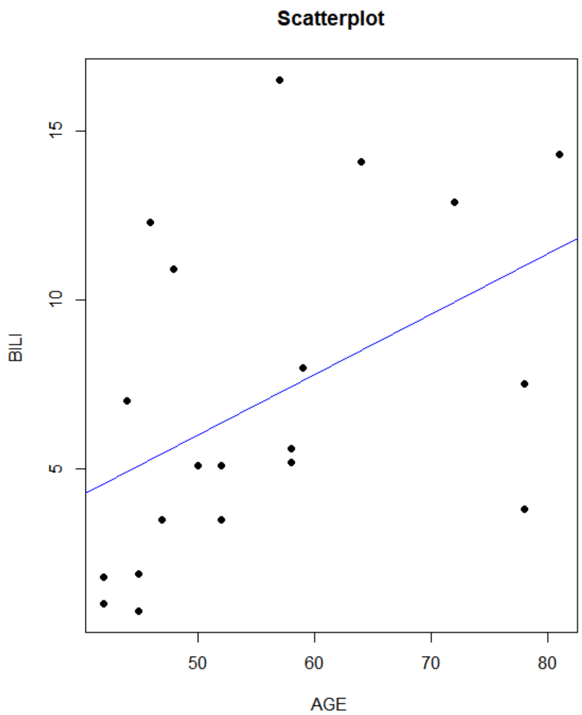
Ex 9.7.1

```
> hw971 <- read.csv("C:/Users/ma4310/EXR_C09_S07_01.csv")
> x <- hw971$AGE
> y <- hw971$BILI
> plot(x, y, main="Scatterplot",xlab="AGE ", ylab="BILI ", pch=19)
> # regression line
> abline(lm(y ~ x, data = hw971), col = "blue")
```




```
> print(summary(relation))
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
-7.203 -2.927 -1.529  2.812  9.263 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -2.9842     4.5962  -0.649   0.5244  
x             0.1793     0.0803   2.233   0.0385 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.414 on 18 degrees of freedom
Multiple R-squared:  0.2169,   Adjusted R-squared:  0.1734 
F-statistic: 4.987 on 1 and 18 DF,  p-value: 0.03848
> cor(y, x)
[1] 0.4657643
> cor.test(x,y)
```

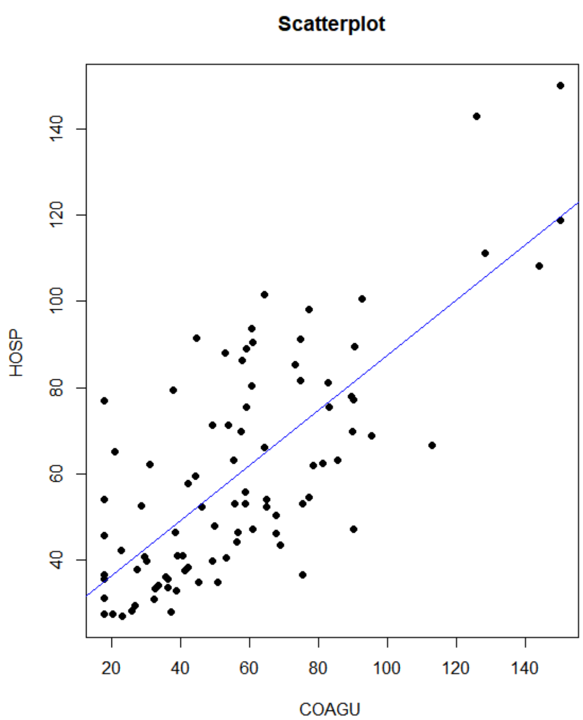
```
        Pearson's product-moment correlation

data:  x and y
t = 2.2331, df = 18, p-value = 0.03848
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.02927797 0.75306959
sample estimates:
      cor
0.4657643
```

Summary:

- `The regression line is y = -2.9842 + 0.1793 x`
- Sample correlation coefficient = `0.4657643`
- `H0 is rejected at significance level 0.05 since p-value = 0.03848`
- `95 percent confidence interval is (0.02927797, 0.75306959)`

Ex 9.7.2

```
> hw972 <- read.csv("C:/Users/ma4310/EXR_C09_S07_02.csv")
> x <- hw972$COAGU
> y <- hw972$HOSP
> plot(x, y, main="Scatterplot",xlab="COAGU ", ylab="HOSP ", pch=19)
> # regression line
> abline(lm(y ~ x, data = hw972), col = "blue")
```



```
> # correlation
> cor(y, x)
[1] 0.735034
> cor.test(x,y)

        Pearson's product-moment correlation

data:  x and y
t = 10.17, df = 88, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6227350 0.8176615
sample estimates:
     cor
0.735034
```

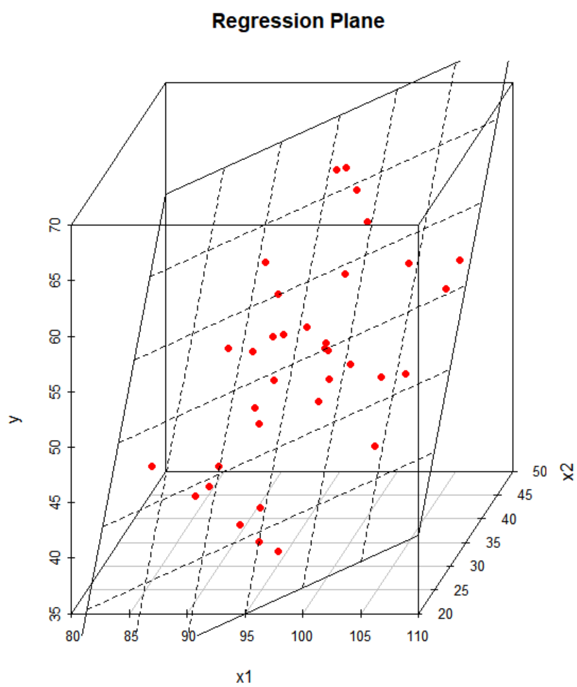
Summary

- Sample correlation coefficient = 0. 735034
- H0 is rejected at significance level 0.05 since p-value = 2.2e-16
- 95 percent confidence interval is (0.6227350, 0.8176615)

Ex 10.3.1

```
> hw1031 <- read.csv("C:/Users/ma4310/EXR_C10_S03_01.csv")
> x1 <- hw1031$X1
> x2 <- hw1031$X2
> y <- hw1031$Y
> # regression equation
> relation <- lm(y~x1+x2, data = hw1031)
> print(relation)

Call:
lm(formula = y ~ x1 + x2, data = hw1031)

Coefficients:
(Intercept)           x1           x2
   -31.4248       0.4732       1.0712

> print(summary(relation))

Call:
lm(formula = y ~ x1 + x2, data = hw1031)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6219 -0.9816  0.1288  1.2022  6.2728

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -31.42480    6.14747  -5.112 1.44e-05 ***
x1            0.47317    0.06117   7.736 8.05e-09 ***
x2            1.07117    0.06280  17.058  < 2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.064 on 32 degrees of freedom
Multiple R-squared:  0.9204,   Adjusted R-squared:  0.9155
F-statistic: 185.1 on 2 and 32 DF,  p-value: < 2.2e-16

> confint(relation, level=0.95) # CIs for model parameters
                  2.5 %      97.5 %
(Intercept) -43.9467785 -18.9028184
x1            0.3485846   0.5977639
x2            0.9432598   1.1990851

> library(scatterplot3d)
> scatterplot3d(x1,x2,y)
> plot3d <- scatterplot3d(x1,x2,y,angle=55, scale.y=0.7, pch=16, color ="red"
, main ="Regression Plane")
> plot3d$plane3d(relation, lty.box = "solid")
```



1) **Multiple coefficient of determination: 0.9204**
2) **beta0, beta1 and beta2 are statistically different from 0 at significance level 0.05 since in all 3 cases the p-value is below 0.05**
3) **the 95% confidence intervals are**
   beta1 CI: (0.3485846, 0.5977639)
   beta2 CI: (0.9432598, 1.1990851)

Ex 10.3.2

```
> hw1032 <- read.csv("C:/Users/ ma4310/EXR_C10_S03_02.csv")
> x1 <- hw1032$ADL
> x2 <- hw1032$MEM
> x3 <- hw1032$COG
> y <- hw1032$KBI
> # regressiuon equation
> relation <- lm(y~x1+x2+x3, data = hw1032)
> print(relation)

Call:
lm(formula = y ~ x1 + x2 + x3, data = hw1032)

Coefficients:
(Intercept)           x1           x2           x3
    40.4908       0.2162       0.5547       0.1210
> print(summary(relation))

Call:
lm(formula = y ~ x1 + x2 + x3, data = hw1032)

Residuals:
    Min      1Q  Median      3Q     Max
-42.037 -10.535  -1.503   9.213  43.151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  40.4908    10.1030   4.008 0.000121 ***
x1            0.2162     0.1168   1.851 0.067273 .
x2            0.5547     0.1300   4.267 4.65e-05 ***
x3            0.1210     0.3003   0.403 0.687978
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.26 on 96 degrees of freedom
Multiple R-squared:  0.282,	Adjusted R-squared:  0.2596
F-statistic: 12.57 on 3 and 96 DF,  p-value: 5.315e-07
> confint(relation, level=0.95)
                  2.5 %     97.5 %
(Intercept) 20.43647272 60.5451146
x1          -0.01567293  0.4480273
x2           0.29662925  0.8126774
x3          -0.47511019  0.7170349
```

1) **Multiple coefficient of determination: 0.282**
2) **Only beta0 and beta2 are statistically different from 0 at significance level 0.05 since in these cases the p-value is below 0.05**
3) **the 95% confidence intervals are**
   beta1 CI: (-0.01567293, 0.4480273)
   beta2 CI: (0.29662925, 0.8126774)
   beta3 CI: (-0.47511019, 0.7170349)

QUIZ 6

A study to investigate the relationship between stress in a workplace and other variables, including the firm size X_1, the number of years of employment X_2, the salary (x1000) X_3 and the age X_4. The data relative to a sample of 15 workers is stored in the file data_q6b.csv.

1) Write the multiple regression equation.
2) Write the expression of the coefficient of determination and explain its significance.
3) Test the hypothesis problems H_0: beta_i = 0, vs. H_1: beta_i ≠ 0, for i=1,2,3,4, at significance level 0.05.
4) Write the 95\% confidence interval of beta_i, for i=1,2,3,4,

Quiz 6 solution

```
> data_q6 <- read.csv("C:/Users/dlabate/desktop/Teaching/ma4310/data_q6b.csv"
)
> relation <- lm(Y~X1+X2+X3+X4, data = data_q6)
> summary(relation)

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4, data = data_q6)

Residuals:
    Min      1Q  Median      3Q     Max 
-30.462 -17.109   2.376  12.886  39.515 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) -126.50532   32.28107  -3.919  0.00287 **
X1             0.17629    0.04009   4.397  0.00134 **
X2            -1.56295    2.01205  -0.777  0.45526   
X3             1.57454    0.44567   3.533  0.00542 **
X4             1.62929    0.62872   2.591  0.02688 * 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.03 on 10 degrees of freedom
Multiple R-squared:  0.8424,	Adjusted R-squared:  0.7794 
F-statistic: 13.37 on 4 and 10 DF,  p-value: 0.0005064
```

1) The regression line is**:** y= -126.51 + 0.18 x1 - 1.56 x2 + 1.57 x3 + 1.63 x4
2) The coefficient of determination is R-squared = 0.84. This indicates that 84% of the variability of y is explained by the variability of x1…x4
3) Hypothesis testing: H0: $\beta_i=0$ vs. H1: $\beta_i \neq 0$

   We reject H0 at significance level 0.01 for i=1,3,4 since the p-value is < 0.01

   We do not reject H0 at significance level 0.01 for i=2 since the p-value is > 0.01

```
> confint(relation)
                    2.5 %      97.5 %
(Intercept) -198.43202047 -54.5786250
X1             0.08695646   0.2656303
X2            -6.04608086   2.9201856
X3             0.58151343   2.5675622
X4             0.22841642   3.0301541
```

4) The 95% confidence interval of β_1 is [0.087,0.265]
   The 95% confidence interval of β_2 is [-6.046,2.920]
   The 95% confidence interval of β_3 is [0.582,2.568]
   The 95% confidence interval of β_4 is [0.228,3.030]