Part I - Testing for linear dependence. In a laboratory containing polarographic equipment six samples of dust were taken at various distances from the polarograph and the mercury content of each sample was determined. The following results were obtained:

Distance from polarograph,m	`1.4,3.8,7.5,10.2,11.7,15.0`
Mercury concentration, ng/g	`2.4,2.5,1.3,1.3,0.7,1.2`

The goal is to examine the possibility that the mercury contamination arose from the polarograph.

Task 1 *Produce a graph of the data representing the dependence of mercury concentration on distance from polarograph. Comment the choice of coordinates.*

Solution. The following three lines of R-code produced the graph that is placed below.

```
Dist=c(1.4,3.8,7.5,10.2,11.7,15.0)
Conc=c(2.4,2.5,1.3,1.3,0.7,1.2)
plot(Dist,Conc)
```

Since we analyze concentration as a function of distance it is natural to take distance as abscissa (independent variable) and concentration as ordinate (dependent variable or response).

Task 2 *Carry out the test if there is any indication of linear dependence between the variables. Comment on the strength of such dependence.*

The simplest way of examining the linear dependence is to compute the correlation coefficient:

```
cor(Conc,Dist)
%[1] -0.8569411
```

Its value being relative close to negative one suggest a linear inverse proportional relation. It is confirmed by a more complete analysis printed below

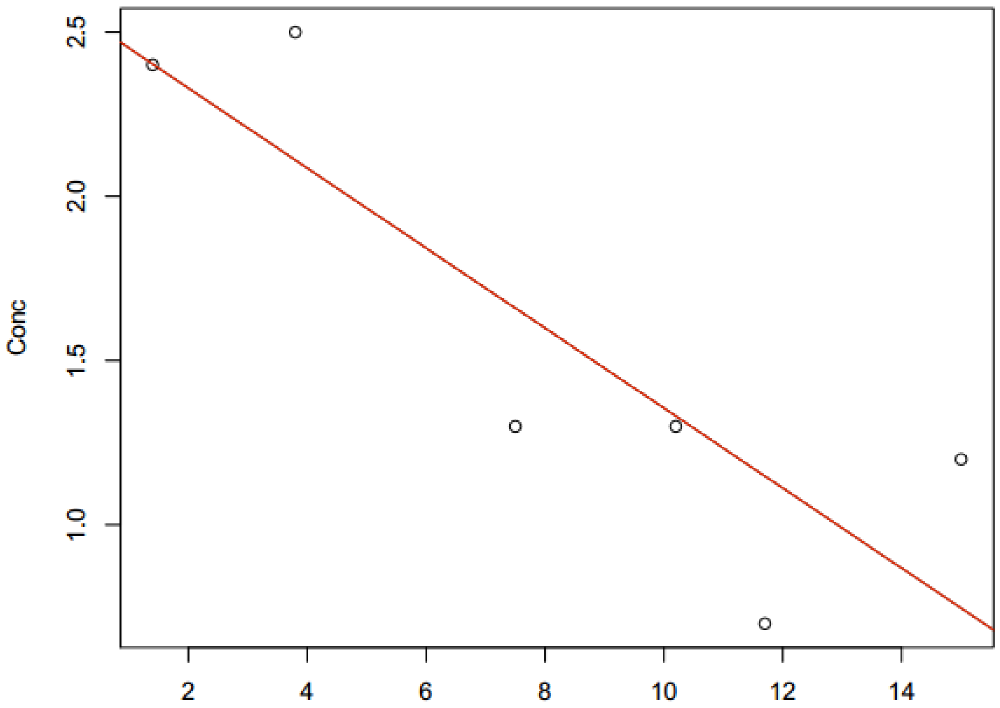
```
summary(lm(Conc~Dist))

Call:
lm(formula = Conc ~ Dist)

Residuals:
        1         2         3         4         5         6
-0.002432  0.389680 -0.359980 -0.031354 -0.448784  0.452870

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.5728     0.3468   7.419  0.00176 **
Dist         -0.1217     0.0366  -3.325  0.02923 *
---
Signif. codes:  0 Ô***Õ 0.001 Ô**Õ 0.01 Ô*Õ 0.05 Ô.Õ 0.1 Ô Õ 1

Residual standard error: 0.415 on 4 degrees of freedom
Multiple R-squared: 0.7343,Adjusted R-squared: 0.6679
F-statistic: 11.06 on 1 and 4 DF,  p-value: 0.02923
```

The p-value reported at the bottom is equal to the p-value of the test for non-zero correlation coefficient. Since it is smaller than 5% we conclude that there is a significant linear relation in the data.

Task 3 *Add the straight line that best fits the dependence.*

The R-function named `lm` performs regression analysis on the data. One of its results are the values of fitted coefficient (slope and intercept). This coefficients can be used to draw the linear fit to the data as follows. The result is presented on the graph.

```
regr=lm(Conc~Dist)
coef(regr)
abline(coef(regr),col="red")
```



Part II – Correlation coefficient. The response of a colorimetric test for glucose was checked with the aid of standard glucose solutions. Determine the correlation coefficient from the following data and comment on the result.

Glucose concentration, mM	Absorbance
0	0.002
2	0.150
4	0.294
6	0.434
8	0.570
10	0.704

Solution. The following produces the correlation coefficient

```
> Gluc = c(0,2,4,6,8,10)
> Absrb = c(0.002,0.15,0.294,0.434,0.57,0.704)
>
> cor.test(Absrb,Gluc)

        Pearson's product-moment correlation

data:  Absrb and Gluc
t = 105.1606, df = 4, p-value = 4.903e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9982632 0.9999812
sample estimates:
      cor
0.9998192
```

The Correlation is almost exactly one, which suggests a near-perfect linear relationship.

The p-value and confidence intervals are also highlighted, in purple.

Part III – Calibration and determination. The following results were obtained when each of a series of standard silver solutions was analysed by flame atomic-absorption spectrometry.

Concentration, ng/ml	Absorbance
10	0.251
15	0.390
20	0.498
25	0.625
30	0.763
0	0.003
5	0.127

Task 1 *Determine the slope and intercept of the calibration plot, and their confidence limits.*

```
Conc=c(0,15,20,25,30,0,5)
Abso=c(0.251,0.390,0.498,0.625,0.763,0.003,0.127)
Fit4 = lm(Abso ~ Conc)
```

```
Fit4
confint(Fit4)
summary(Fit4)
```

```
> Fit4

Call:
lm(formula = Absrb ~ Gold)

Coefficients:
(Intercept)         Gold
   0.256917     0.005349

> confint(Fit4)
                 2.5 %      97.5 %
(Intercept) 0.25108226 0.262751077
Gold        0.00520934 0.005488279
```

The regression equation is

Absrb.fitted = 0.257 + 0.0054 Gold

The 95% confidence intervals for the regression coefficients are

Intercept β_0: (0.2516,0.263)

Slope β_1 :(0.00520, 0.005488)

Part IV – The method of standard additions The gold content of a concentrated sea-water sample was determined by using atomic-absorption spectrometry with the method of standard additions. The results obtained were as follows:

Gold added, ng per ml of concentrated sample	Absorbance
30	0.413
40	0.468
50	0.528
60	0.574
70	0.635
0	0.257
10	0.314
20	0.364

```
> Gold = c(30,40,50,60,70,0,10,20)
> Absrb= c(0.413,0.468,0.528,0.574,0.635,0.257,0.314,0.364)
> Fit4 = lm(Absrb ~ Gold)
>
> summary(Fit4)

Call:
lm(formula = Absrb ~ Gold)

Residuals:
       Min          1Q     Median          3Q         Max
-0.0043810 -0.0031131  0.0000952  0.0036071  0.0036667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.256917   0.002384  107.75 4.31e-11 ***
Gold        0.005349   0.000057   93.84 9.87e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.003694 on 6 degrees of freedom
Multiple R-squared: 0.9993,     Adjusted R-squared: 0.9992
F-statistic:  8806 on 1 and 6 DF,  p-value: 9.866e-11
```

Part V – Comparing analytical methods An ion-selective electrode (ISE) determination of sulphide from sulphate-reducing bacteria was compared with a gravimetric determination. The result, obtained were expressed in milligrams of sulphide.

```
Sulphide (ISE method): 108,12,152,3,106,11,128,12,160,128
Sulphide (gravimetry): 105,16,113,0,108,11,141,11,182,118
```

Comment on the suitability of the ISE method for this sulphide determination (Al-Hitti, I. K., Moody, G.]. and Thomas, J. D. R. 1983. Analyst 108: 43).

Solution. We first plot the data with the regression fit.

```
ISE=c(108,12,152,3,106,11,128,12,160,128)
Grav=c(105,16,113,0,108,11,141,11,182,118)
plot(Grav,ISE)
regr=lm(ISE~Grav)
abline(coef(regr),col="blue")

a=coef(regr)[1]
b=coef(regr)[2]
```

Lab C part 5

```
> Fit5a

Call:
lm(formula = ISE ~ Grav)

Coefficients:
(Intercept)         Grav
     4.4837       0.9629
```

```
> Fit5b

Call:
lm(formula = Grav ~ ISE)

Coefficients:
(Intercept)          ISE
```

```
    0.4195          0.9766
```

There is a disparity in what both linear models predict.

A predicted value of ISE based on a known value of GRAV can be easily determined.

Grav = 4.4837 + 0.9629(10)

ISE.fitted = 4.4837 + 0.9629 Grav

Suppose Grav is 10. The fitted value of ISE, using Fit5a = 14.1127

Let’s use Fit5b to predict a value for ISE = 14. If we are to believe Fit5a, the answer should be roughly 10.

Using Fit5b, we get a predicted value of 14.20196 for grav.fitted.

Hence the use of Linear models is invalid when comparing methods of measurement.