Estimation of the intercept β_0 and slope β_1

The subsequent equations are used to *estimate* the intercept $\hat{\beta}_0$ and the slope $\hat{\beta}_1$ of a population regression model $y_i = \beta_0 + \beta_1 \cdot x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, 0.2)$:

$$\hat{\beta}_1 = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad (1)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} \quad (2)$$

The code below generates a set of $n = 50$ random normally distributed numbers for the independent variable x_i and the dependent variable y_i using the equation $y_i = 1 + 2 \cdot x_i + \epsilon_i$:

```
x <- rnorm(50)
y <- 1 + 2*x + rnorm(50,sd=0.2)
```

Note that the expected intercept is $\beta_0 = 1$ and the expected slope is $\beta_1 = 2$. The regression disturbances have a standard deviation of $\sqrt{Var(\epsilon)} = 0.2$

Subsequently the model is estimated with R's `lm( )` function and the summary statistics of the estimated model are listed:

```
my.lm <- lm(y~x)
summary(my.lm)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4200 -0.1482 -0.0083  0.1355  0.4493
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.98630    0.02834   34.80   <2e-16 ***
## x            1.95529    0.02596   75.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1993 on 48 degrees of freedom
## Multiple R-squared:  0.9916,Adjusted R-squared:  0.9914
## F-statistic:  5672 on 1 and 48 DF,  p-value: < 2.2e-16
```

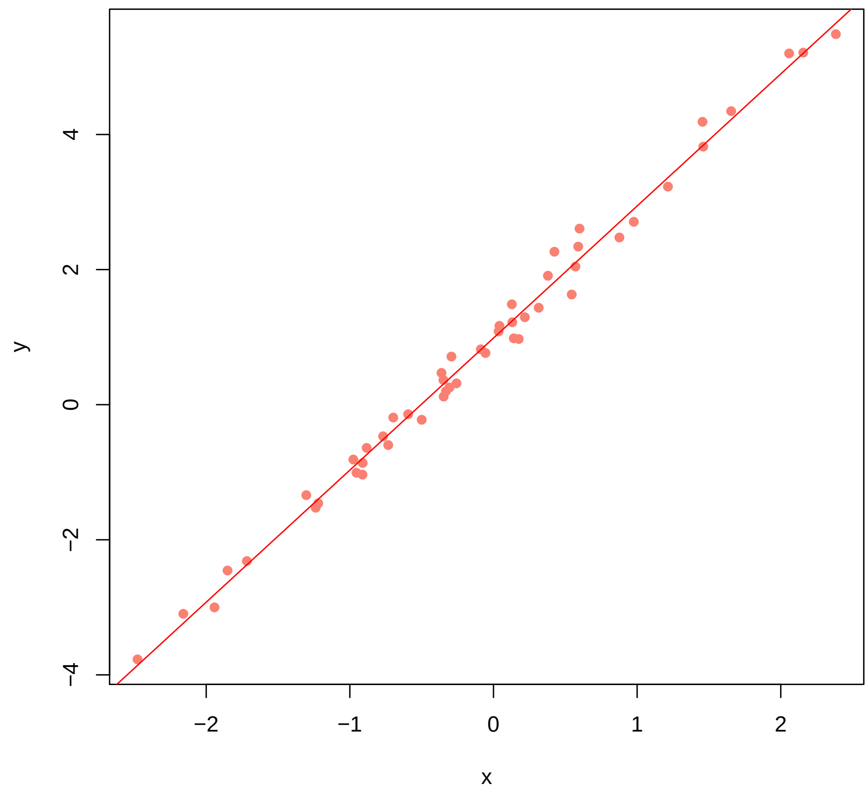

Figure 1: Scatterplot and Regression Line

The data can be displayed in a scatter-plot and the estimated regression line can be superimposed on top of it. The R commands generating the scatter-plot are shown here:

```
plot(y~x, pch=16, col="salmon")
abline(my.lm, col="red")
```