Inferences involving the Model Parameters

We previously considered the following observations $Y_1, Y_2, \ldots, Y_n$ following linear model $E(Y) = \beta_0 + \beta_1 x$:

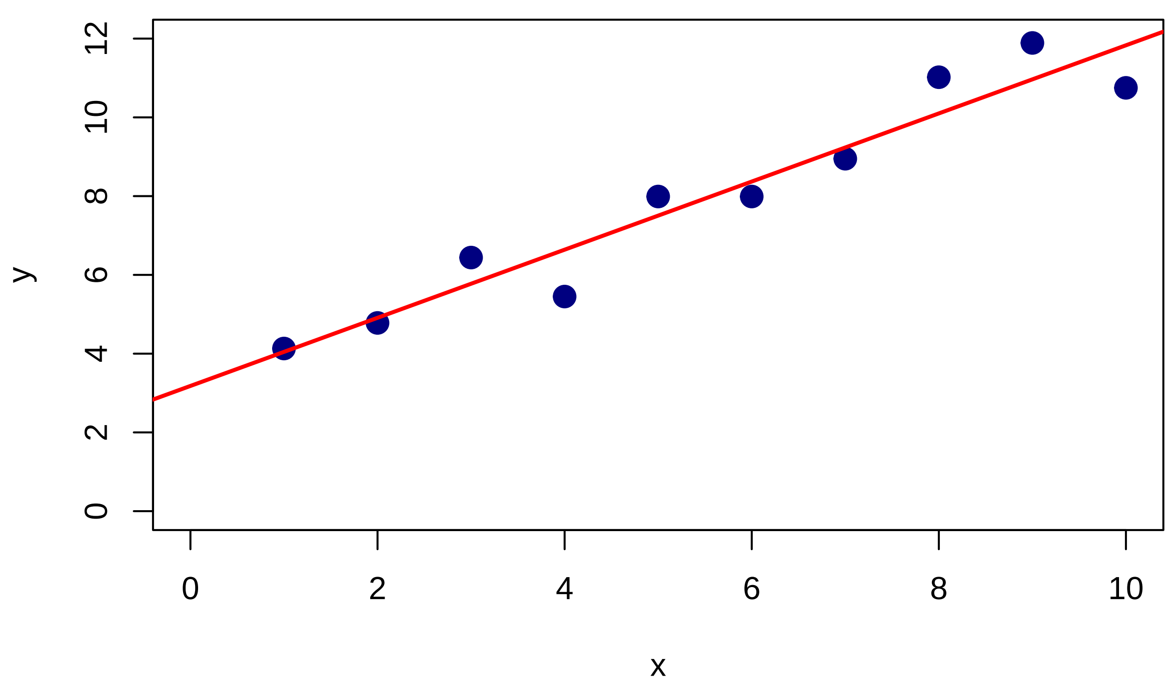
x	1	2	3	4	5	6	7	8	9	10
y	4.13	4.78	6.44	5.45	7.99	7.99	8.95	11.02	11.89	10.75

In the previous example, we fit the least squares model

$$\hat{y} = 3.18 + 0.865x$$

The following code plots each point (x, y) in the plane as well as the least squares model line. We can plot the fitted line, as well as the observed data using the **abline** function, which plots a line with given slope and intercept.

```
x <- 1:10
y <-  c(4.13,  4.78,  6.44,  5.45,  7.99,  7.99,  8.95, 11.02, 11.89, 10.75)
plot(x, y, xlim = c(0,10), ylim = c(0,12), col = "navy", pch = 19, cex = 1.5)
abline(3.18, 0.865, col = "red", lwd = 2)
```



A Hypothesis Test Involving β_0

Suppose that we wanted to test the hypothesis $H_0 : \beta_0 = 2$ against the alternative hypothesis $H_a : \beta_0 \neq 0$. Since we do not know the population variance σ^2, we need to estimate it using $S^2 = SSE/(n-2)$, which can be calculated using the following code. (Note that this repeats our earlier calculation of SSE.)

```
# Get number of observations.
n <- length(x)

# Calculate necessary sums.
sumx <- sum(x = x)
sumy <- sum(x = y)
sumxy <- sum(x = x*y)
sumxx <- sum(x = x^2)
sumyy <- sum(y^2)

# Calculate Sxy, Sxx, and Syy.
Sxy <- sumxy - 1/n*sumx*sumy
Sxx <- sumxx - 1/n*sumx^2
Syy = sumyy - sumy^2/n

# Recalculate least squares estimate of slope and y-intercept.
hb1 <- Sxy/Sxx
hb0 <- (sumy - hb1*sumx)/n

# Calculate SSE.
SSE <- Syy - hb1*Sxy

# Calculate sample variance.
sampleVar <- SSE/(n-2)
```

This yields $s^2 = 0.6514626$.

We need the value c_{00} to calculate the test statistic. We can do so using the following code.

```
c00 <- sumxx/(n*Sxx)
```

This yields $c_{00} = 0.4666667$.

Putting everything together we calculate the value of the test statistic T using the formula

$$t = \frac{\hat{\beta}_0 - 2}{s\sqrt{c_{00}}} = \frac{3.1806667 - 2}{\sqrt{(0.6514626)(0.4666667)}},$$

which can be evaluated using the following code.

```
tval <- (hb0 - 2)/sqrt(sampleVar*c00)
```

This gives $t = 2.1413082$.

Note that $t_{\alpha/2} = t_{0.025} = 2.3060041$; since our observed value of T does not exceed this value, we cannot reject the null hypothesis. Moreover, we can calculate the attained level of significance for this two-sided test:

```
pval <- 2*pt(q = tval, df = n-2, lower.tail = FALSE)
```

This yields p-value 0.0646446.

Calculating a confidence interval

We can also calculate a two-sided $100(1-\alpha)\%$ confidence interval for β_0 using the formula

$$\hat{\beta}_0 \pm t_{\alpha/2} S \sqrt{c_0 0}.$$

We can calculate the bounds of this confidence interval for $\alpha = 0.05$ using the following code.

```
# Radius/half-width of confidence interval
rad <- sqrt(sampleVar*c00)*qt(p = 0.025, df = n - 2, lower.tail = FALSE)

# Lower and upper limit of the confidence interval.
lb <- hb0 - rad
ub <- hb0 + rad
```

This yields the confidence interval

$$1.9091905 < \beta_0 < 4.4521428.$$

We can confirm that we should not reject H_0 since the value of β_0 under the null hypothesis ($\beta_0 = 2$) belongs to this confidence interval.