Inferences involving the Linear Functions of the Parameters

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

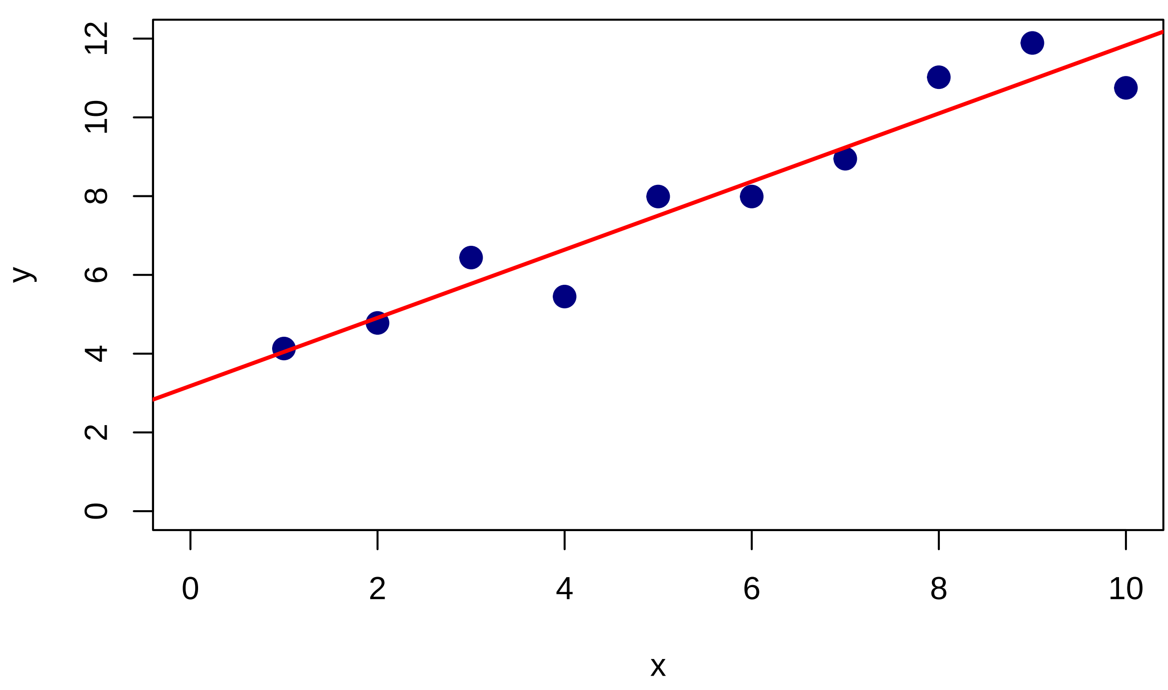
x	1	2	3	4	5	6	7	8	9	10
y	4.13	4.78	6.44	5.45	7.99	7.99	8.95	11.02	11.89	10.75

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 Confidence Interval for Expected Value

We want to calculate a $100(1-\alpha)\%$ confidence interval for $E(Y) = \beta_0 + \beta_1 x^*$. This is a special case of $a_0\beta_0 + a_1\beta_1$ with $a_0 = 1$ and $a_1 = x^*$. Thus, we can use the formula for the general confidence interval, specialized for this particular choice of a_0, a_1.

As an example, let's calculate a 90% confidence interval for $E(Y)$ at $x^* = a_1 = 1$ for this sample. Using the formula derived earlier, the desired confidence interval is given by the formula

$$\hat{\beta}_0 + \hat{\beta}_1 \pm t_{0.05} S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}},$$

where $t_{0.05}$ is based on $n - 2 = 8$ degrees of freedom.

We calculated many of the quantities in the formula earlier, which we load using the following code.

```r
# Model parameters.
hb0 <- 3.18
hb1 <- 0.865

# Recalculate mean, Sxx.
n <- length(x)
xbar <- mean(x)
Sxx <- sum(x^2) - n*xbar^2

# Estimate of the variance.
s <- sqrt(0.6515)
```

Finally, we obtain the confidence by substituting $x^* = 1$ using the following code.

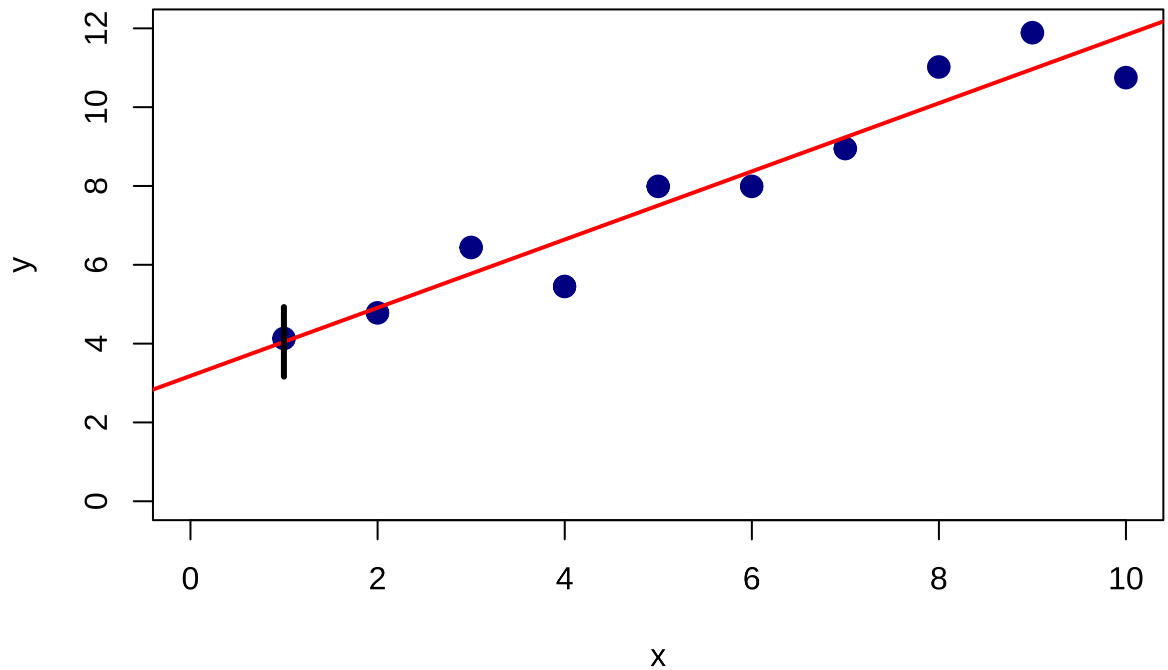
```r
xstar <- 1

# Calculate the radius/half-width of the interval.
rad <- qt(p = 0.05, df = n - 2, lower.tail = FALSE)*s*sqrt(1/n + (xstar - xbar)^2/Sxx)

# Calculate confidence interval.

lb <- hb0 + hb1*xstar - rad
ub <- hb0 + hb1*xstar + rad
```

This yields the confidence interval

$$3.1628142 < E(Y) < 4.9271858.$$

We can superimpose this confidence interval on our earlier plot of the observations and least squares line using the following code.

```r
# Earlier plot.
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)
# Superimpose confidence interval.
lines(c(1,1), c(lb, ub), col = "black", lwd = 3)
```



Note that the observed value of Y corresponding to $x^* = 1$ lies within this confidence interval.

Prediction Intervals

Suppose that we want to predict the value of Y for a given value of $x = x^*$. We know that a $100(1-\alpha)\%$ *prediction* interval for Y is given

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}.$$

Suppose that we want to predict the value of Y for $x = 1$. For example, this would occur if we run the underlying experiment again using the independent variable value $x = 1$ (we've already generated one observation $Y = 4.13$ corresponding to $x = 1$). We can generate a 90% prediction interval for Y using the following code.

```
# Choose x and alpha.
xstar <- 1
alpha <- 0.1

# Calculate radius of the prediction interval.
prad <- qt(p = alpha/2, df = n -2, lower.tail = FALSE)*s*sqrt(1 + 1/n + (xstar - xbar)^2/Sxx)

# Calculate endpoints of the prediction interval.
plb <- hb0 + hb1*xstar - prad
pub <- hb0 + hb1*xstar + prad
```

This gives the prediction interval

$$2.3039986 < Y < 5.7860014,$$

with width 3.4820029. Note that the code for generating the prediction interval is nearly identical to that used to calculate the confidence interval for $E(Y)$ at $x = 1$; the only difference is the presence of the 1-term in the square root of the formula for the prediction interval. We should note further that the prediction interval is significantly wider than the confidence interval for $E(Y)$.

Plotting Confidence and Prediction Bands

The formulas for confidence interval for $E(Y)$ at x^* and prediction interval for Y at x^* can be evaluated for a *set* of x^* with minor modification to our earlier code. For example, the following code generates the prediction intervals for Y for each x^* in the set $\{0, 0.25, 0.5, \ldots, 11.75, 12\}$.

```
# Choose x and alpha.
xstar <- seq(from = 0, to = 12, by = 0.25)
alpha <- 0.1

# Calculate radius of the prediction interval.
prads <- qt(p = alpha/2, df = n -2, lower.tail = FALSE)*s*sqrt(1 + 1/n + (xstar - xbar)^2/Sxx)

# Calculate endpoints of the prediction interval.
plbs <- hb0 + hb1*xstar - prads
pubs <- hb0 + hb1*xstar + prads
```

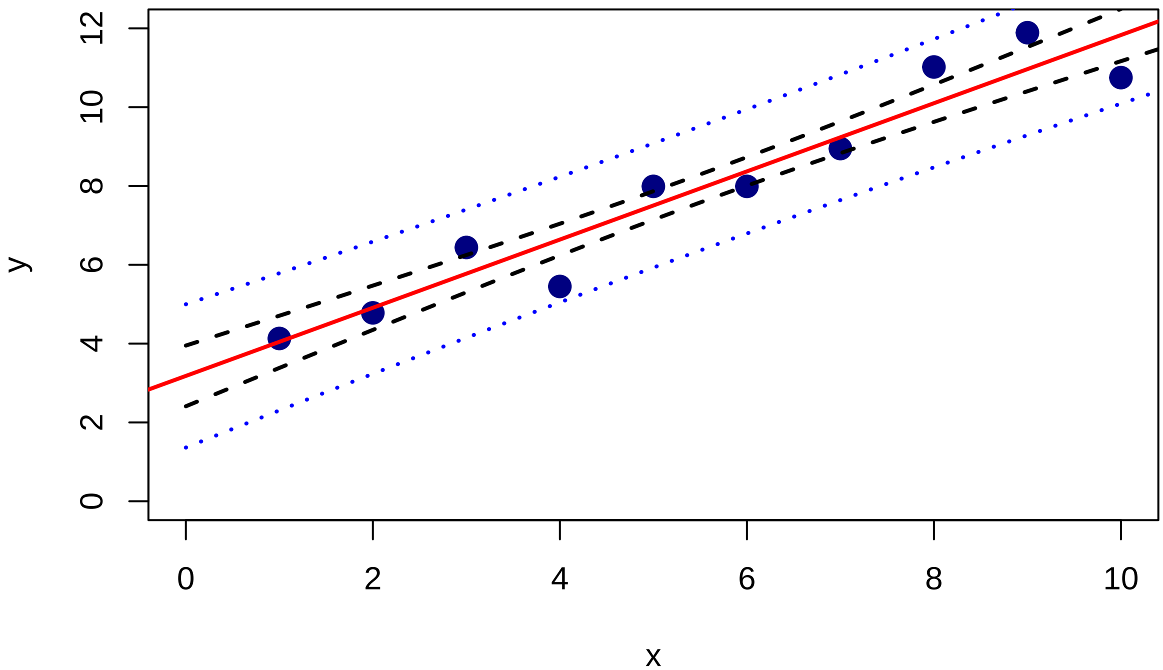
The only difference between this code and that used to calculate the prediction at a single point earlier, is that we now evaluate the formula for the endpoint with x^* given as an input vector (rather than a single value). The output of this code is vectors **plbs** and **pubs** storing the lower and upper bounds, respectively, of the prediction interval for each value of x^* upper bound.

Similarly, we can generate confidence intervals for $E(Y)$ at each x^* in the same set using the following code.

```
rads <- qt(p = alpha, df = n - 2, lower.tail = FALSE)*s*sqrt(1/n + (xstar - xbar)^2/Sxx)
lbs <- hb0 + hb1*xstar - rads
ubs <- hb0 + hb1*xstar + rads
```

We can use these calculated intervals to plot confidence and prediction bands for the least-squares line.

```
# Earlier plot.
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)

# Plot curve indicating lower/upper bounds on confidence intervals.
lines(xstar, lbs, col = "black", lty = 2, lwd = 2)
lines(xstar, ubs, col = "black", pch = "o", lty = 2, lwd = 2)

# Plot prediction band.
lines(xstar, plbs, col = "blue", pch = "o", lty = 3, lwd = 2)
lines(xstar, pubs, col = "blue", pch = "o", lty = 3, lwd = 2)
```



0
2
4
6
8
10
12
y
0
2
4
6
8
10
x