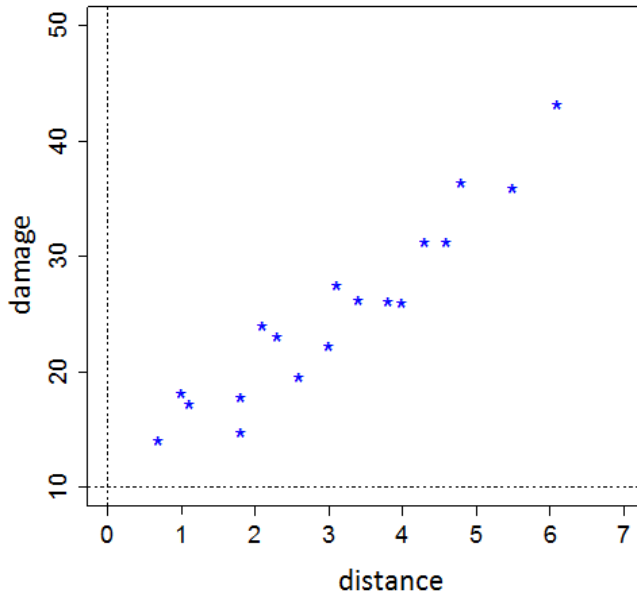


In-class-exercise week 8
Topic: Correlation & Simple Regression

Problem 1

The damage of cells due to inflammation depend on the distance to the source (data made up).



We have used R to modeled the data with a linear regression::

Call:

```
lm(formula = damage ~ dist, data = fire)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.1689	-2.1322	-0.1006	2.1287	3.5850

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.292	1.408	7.312	1.75e-06 ***
dist	4.820	0.406	11.874	2.40e-09 ***

Residual standard error: 2.637 on 16 degrees of freedom

a)

- What is the estimate for the intercept, what is the estimate for the slope?

estimated intercept =10.9

estimated slope=4.8

- Write down the estimated linear relationship describing how the damage depend on the

distance.

$$\hat{y} = 10.9 + 4.8 \cdot x \quad \text{with } y=\text{damage, } x=\text{distance}$$

- Draw the corresponding line in the scatterplot.
- What damage does your model predict for the distance 5? Mark the prediction in the plot.

$$\hat{y}(5) = 10.9 + 4.8 \cdot 5 = 34.5$$

b) Why do we have 16 degree of freedoms?

We have $n=18$ observations (count the points in the scatterplot) and we estimate $p=2$ model coefficients \rightarrow degree of freedom $= n-p=16$

c) Use the R output above to construct an 95% confidence interval for the slope.

$$4.8 \pm t_{16}^{97.5\%} \cdot 0.4 \approx 4.8 \pm 2 \cdot 0.4 = [4, 5.6]$$

d) How much does the damage change if the distance change from 1 to 2 or from 4 to 5?

We can use the slope to answer the question. If the distance change by 1 unit, the outcome changes by the slope, here 4.8.