Biostatistics

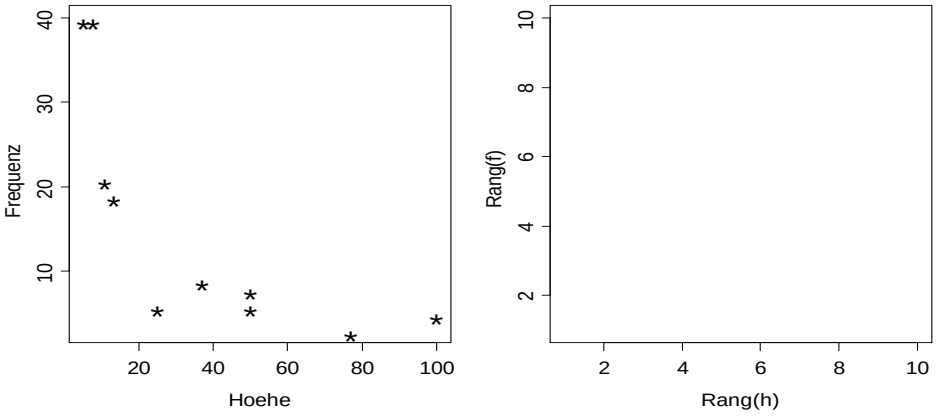
**In-class exercise 8**
**Topic: Correlation & Linear regression**

**Problem 1**: In a science article (Science, 164 (1969), p.1513) a study was presented that investigated the association between the hight of a waterfall and the frequence of vibrations in the ground.

| Name | h: Hoehe | f: Frequenz | Rang(f) | Rang(h) |
|---|---|---|---|---|
| Lower.Yellowstone | 100 | 5 | | |
| Yosemite | 77 | 3 | | |
| Canadian.Niagara | 50 | 6 | | |
| American.Niagara | 50 | 8 | | |
| Upper.Yellowstone | 37 | 9 | | |
| Lower.Gullfoss | 25 | 6 | | |
| Firehole | 13.3 | 19 | | |
| Godafoss | 10.9 | 21 | | |
| Upper.Gullfoss | 7.7 | 40 | | |
| Fort.Greeley | 5.2 | 40 | | |

The following graph shows a scatterplot of frequence vs hight.

a) Describe the form of the association.



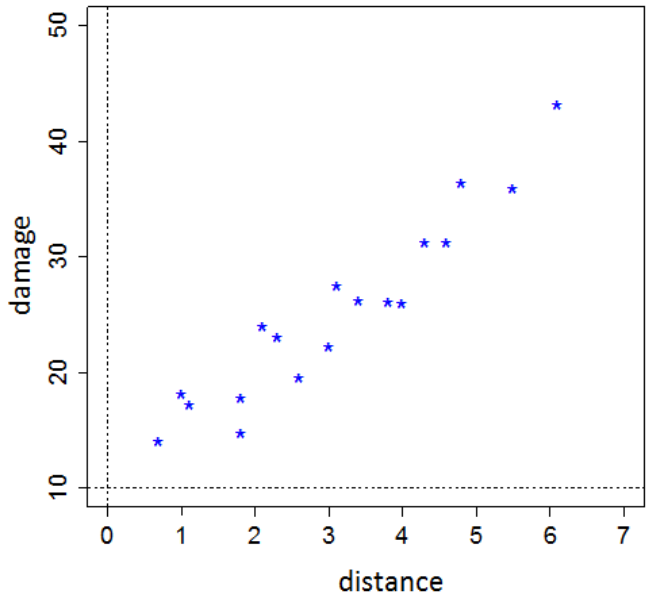

b) Is the Pearson correlation appropriate here? Give reasons.

c) Determine in the table the missing ranks and visualize the relationships as scatter plot.

d) What can you learn from the absolute value and the sign of the correlation?

**Problem 2:**

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?

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

- Draw the corresponding line in the scatterplot.

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