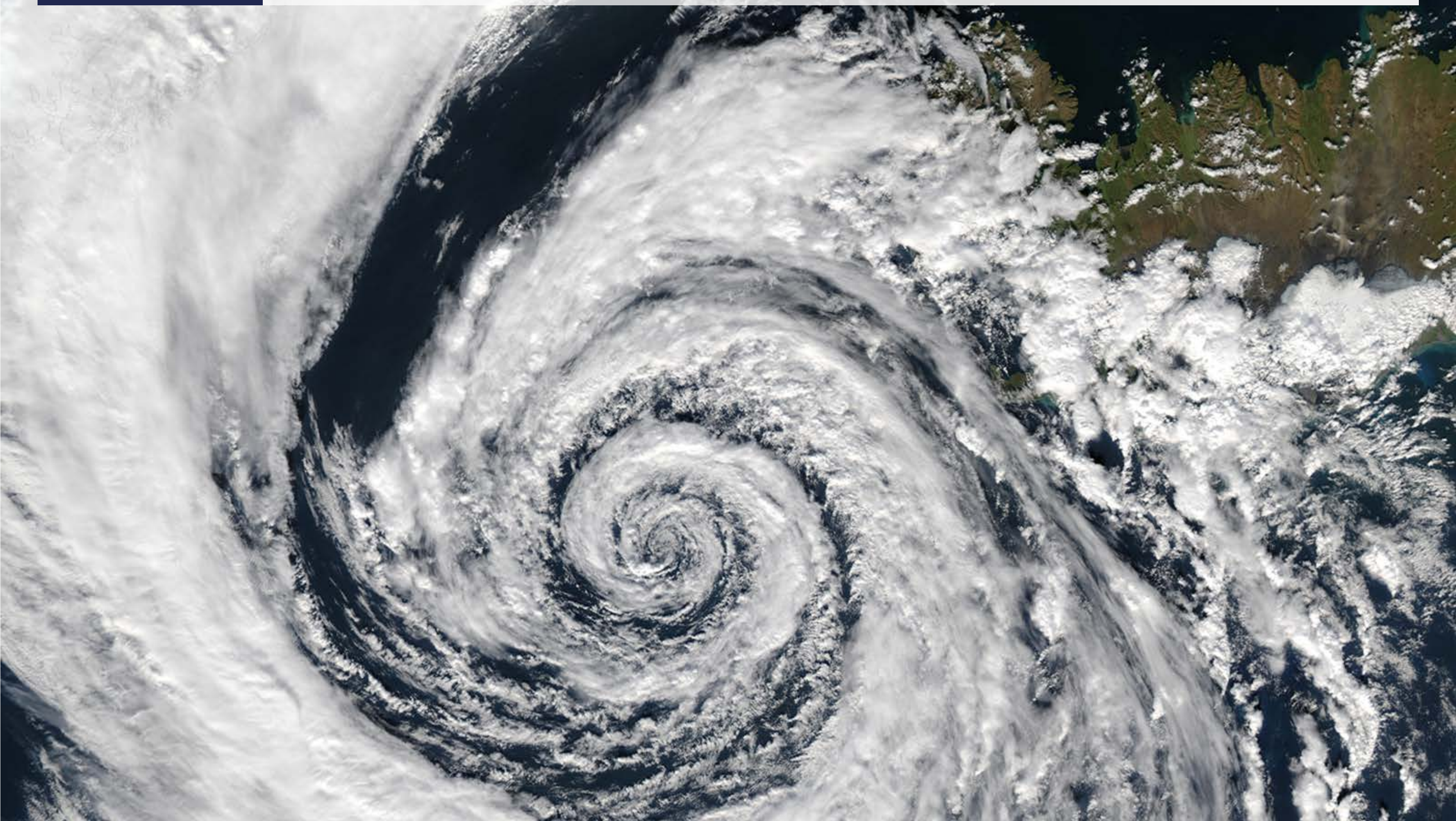
# Statistics

## Maximum Likelihood Estimate and Linear Regression

Name teacher

TU Delft



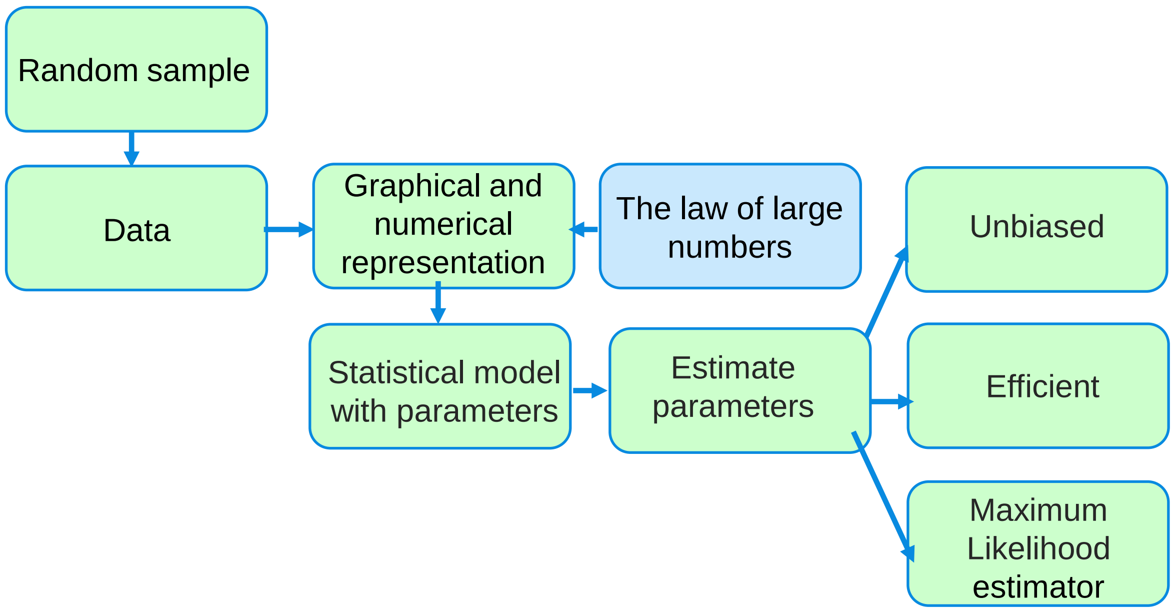
# Learning objective

After this class you can:

- Apply the Maximum Likelihood Principle in various settings
- Derive the Maximum Likelihood Estimate for model parameters
- Set up a linear regression model and estimates its parameters

Book: Chapters 21 and 22

# Statistics


Random sample
Data
Graphical and numerical representation
The law of large numbers
Statistical model with parameters
Estimate parameters
Unbiased
Efficient
Maximum Likelihood estimator

# Before this class (week 3.7 lesson 1):

Watch prelecture *‘Maximum Likelihood Principle’*

Book: Sections 21.1 and 21.2

# Programme

- Likelihood and loglikelihood
- Obtain maximum likelihood estimate
- Exercises
- Linear Regression
- Exercises

## Question from prelecture: What is the maximum likelihood estimate for passing probability of men?

A) 0.375

B) 0.607

C) 0.393

D) 0.647

**Consider two dice: die R has 5 red sides and 1 white, die W has 5 white side and 1 red. We choose one die. We throw it until first red appears. This is repeated 2 more times. The results: 3rd throw, 5th throw and 4th throw show first red. Which die do you think we threw?**



A) Die R

B) Die W

C) Equally likely

# Likelihood and maximum likelihood estimate

**Definition:**

Let $x_1, x_2, \ldots, x_n$ be a realization of a random sample from a distribution characterized by a parameter $\theta$.

Distribution is discrete with p.m.f. $p_\theta$: the *likelihood* is

$$L(\theta) = P(X_1 = x_1, \ldots, X_n = x_n) = p_\theta(x_1) \cdots p_\theta(x_n).$$

Distribution is continuous with p.d.f. $f_\theta$: the *likelihood* is

$$L(\theta) = f_\theta(x_1) \cdots f_\theta(x_n).$$

The *maximum likelihood estimate* of $\theta$ is the value $t = h(x_1, \ldots, x_n)$ that maximizes $L(\theta)$.

The corresponding random variable $T = h(X_1, \ldots, X_n)$ is the *maximum likelihood estimator* for $\theta$.

# Maximum Likelihood Estimate

**How to obtain the maximum likelihood estimate for $\theta$?**

- Compute the likelihood $L(\theta)$
- Compute the loglikelihood $\ell(\theta) = \ln(L(\theta))$
- Differentiate $\ell(\theta)$ with respect to $\theta$
- Solve $\ell'(\theta) = 0$: this yields $\hat{\theta}$
- Check whether it is a maximum, e.g. check $\ell''(\hat{\theta}) < 0$
- If this is the case, then $\hat{\theta}$ is the MLE for $\theta$

# Example: exponential distribution

Let $x_1, \ldots, x_n$ be a dataset from a $Exp(\lambda)$ distribution.

Then:
$$\begin{aligned}
L(\lambda) &= \ldots \\
\ell(\lambda) &= \ldots \\
\ell'(\lambda) &= \ldots \\
\ell'(\lambda) &= 0, \ \text{if } \lambda = \ldots \\
\ell''(\lambda) &= \ldots \\
\ell''(\hat{\lambda}) &= \ldots
\end{aligned}$$

Thus $\hat{\lambda} = \frac{1}{\bar{x}_n}$ is MLE for $\lambda$.

# Two exercises

During WW II many areas of London were hit by German bombs. The following table shows the number of hits in 576 parts (squares Of length ¼ km) of South London:

| Number of hits | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Number of squares | 229 | 211 | 93 | 35 | 7 | 0 | 0 | 1 |

Model the hits by Poisson distribution with parameter μ. Compute the MLE for μ.

Repeat the previous exercise if the data were corrupted:

| Number of hits | 0 or 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Number of squares | 440 | 93 | 35 | 7 | 0 | 0 | 1 |

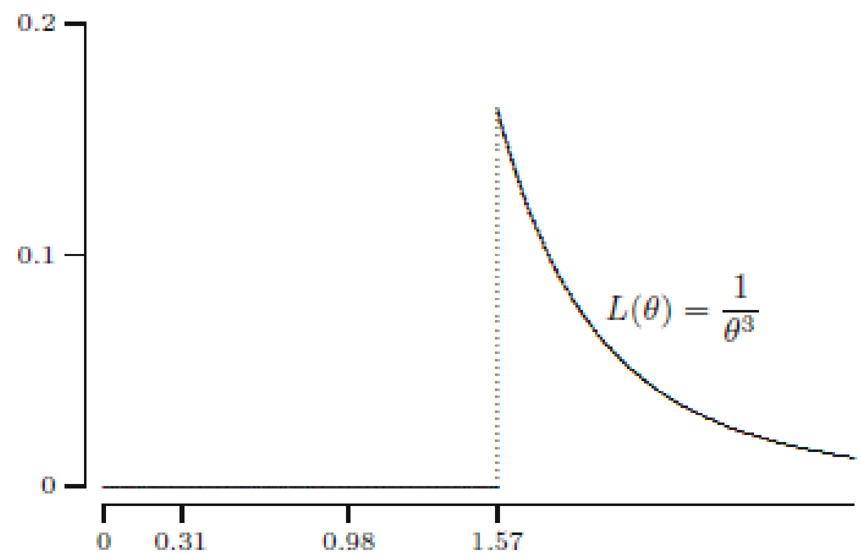
# Caution!

The stepwise approach sometimes fails!

Suppose the dataset $x_1 = 0.98, x_2 = 1.57$, and $x_3 = 0.31$ is the realization of a random sample from a $U(0, \theta)$ distribution.

Then $L(\theta) = \ldots$

Thus $\hat{\theta} = \max\{x_1, x_2, x_3\} = 1.57$

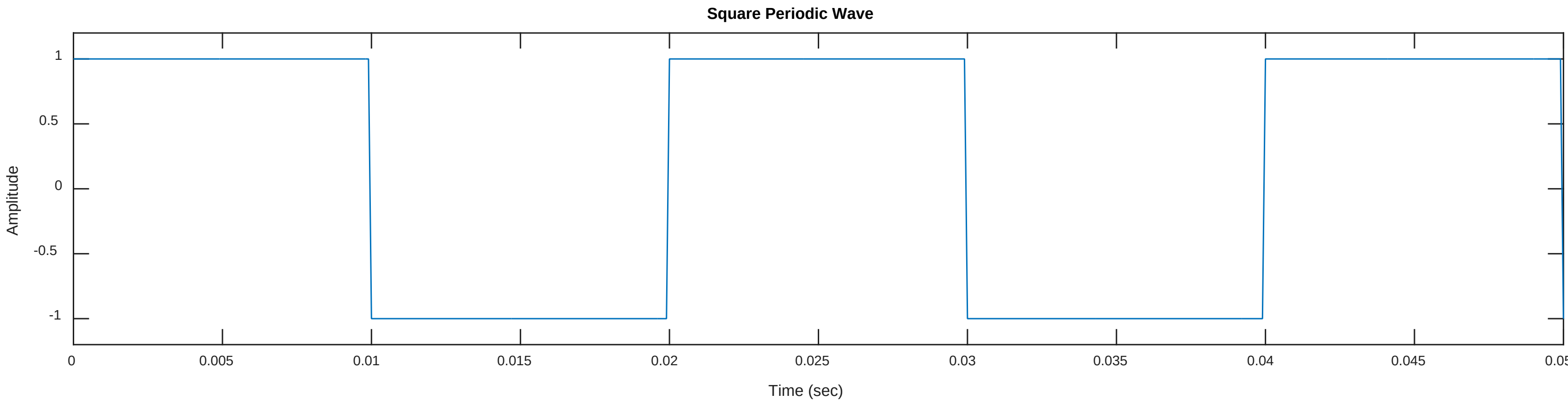


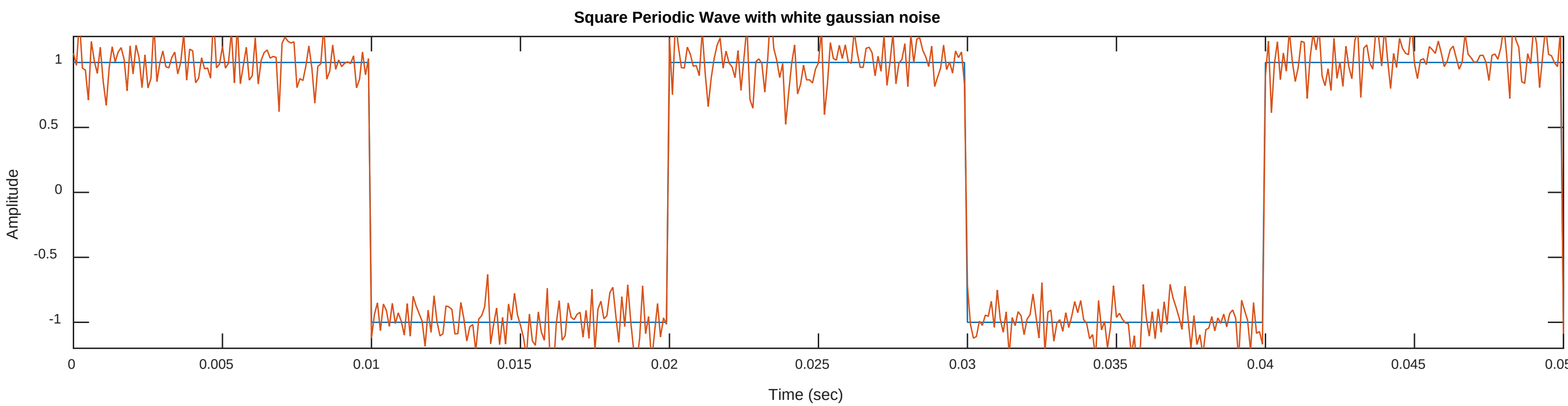
# Signal with noise

Let us take a square periodic wave signal $x$.

Suppose we have two independent white gaussian noise components

We receive $y = x + n_1 + n_2$ with $n_1, n_2 \sim N(0, \sigma^2)$.

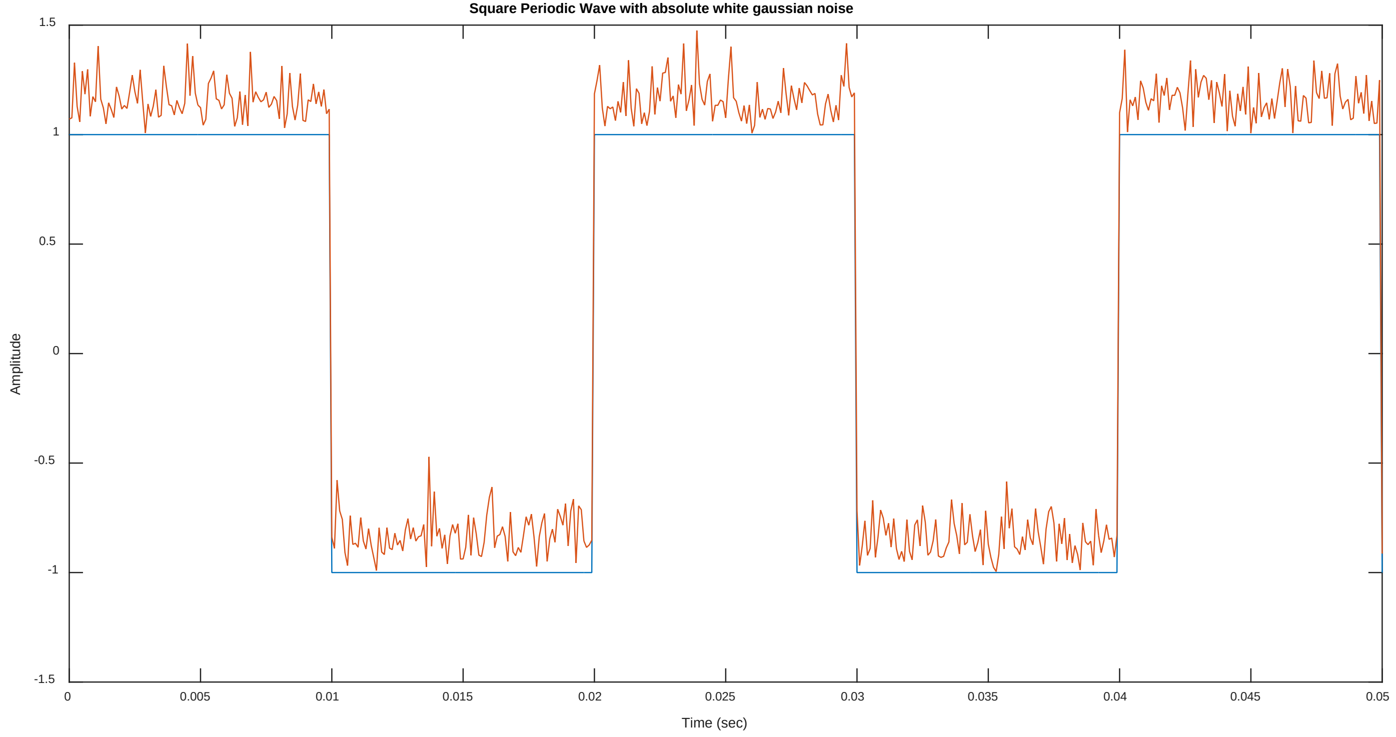


# Signal with noise

Now consider the absolute noise $N = \sqrt{n_1^2 + n_2^2}$ on the signal.
This has a so-called Rayleigh distribution with parameter $\sigma$!

Its density is given by $f_N(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$



# Two Exercises

Let $x_1, \ldots, x_n$ be a dataset from a distribution with density

$$f_\theta(x) = \begin{cases} \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Calculate the likelihood $L(\theta)$ and the maximum likelihood estimate $\hat{\theta}$ for $\theta$.

---

Estimate the standard deviation in the two noise signals $n_1, n_2 \sim N(0, \sigma^2)$ when we have the following dataset for the absolute noise:
1.77, 1.78, 1.36, 0.82, 1.06, 1.04, 2.01, 1.06, 0.88, 0.43

# Two exercises

Let $x_1, \ldots, x_n$ be a dataset from a distribution with density $f_\delta$:

$$f_\delta(x) = \begin{cases} e^{-(x-\delta)} & x \geq \delta \\ 0 & x < \delta. \end{cases}$$

Draw the likelihood $L(\delta)$ and determine the MLE for $\delta$.

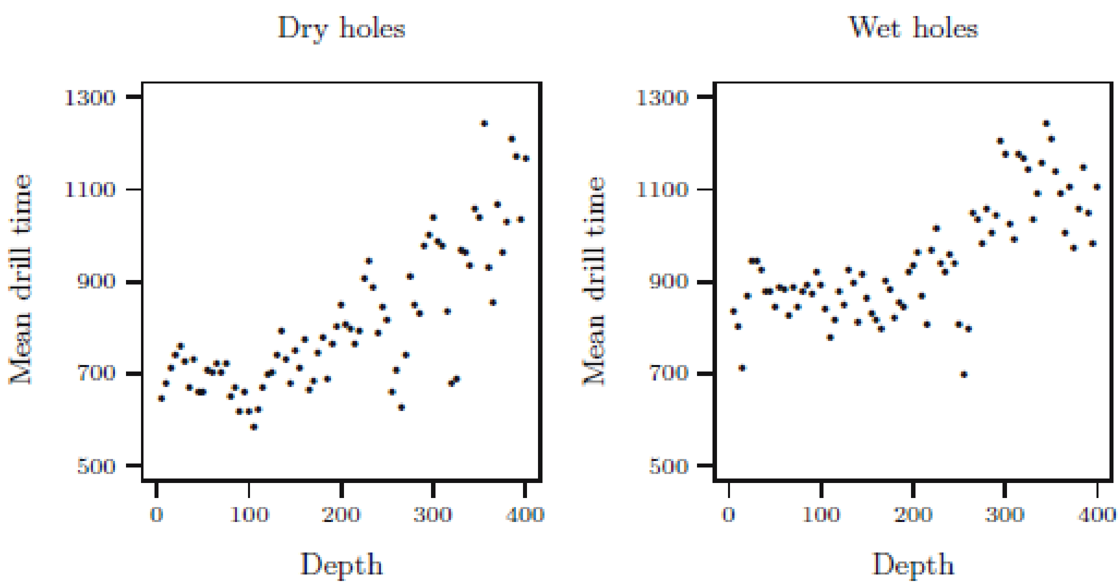
---

Let $x_1, \ldots, x_n$ be a realization of a random sample from a $N(\mu, \sigma^2)$ distributed RV.
Determine the MLE for $\mu$ and $\sigma$.

# Scatterplot

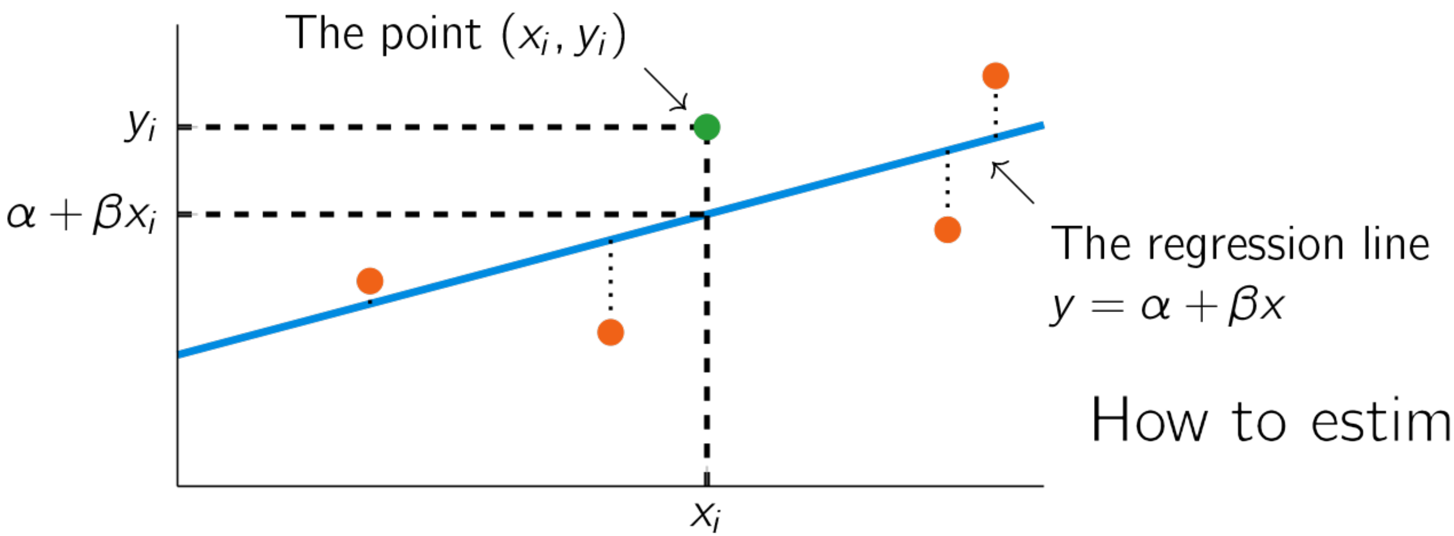
**Definition:**
If two variables $x$ and $y$ are measured on the same objects, the dataset $(x_1, y_1), \ldots, (x_n, y_n)$ is called a *bivariate dataset*.
A plot of the points $(x_i, y_i)$ is called a *scatterplot*.

Two examples: Mean drill time versus depth for “dry” and “wet” drilling



# Regression line

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a bivariate dataset.
*Simple linear regression model*: Assume that $x_1, \ldots, x_n$ are nonrandom and $y_1, \ldots, y_n$ realizations from random variables $Y_i = \alpha + \beta x_i + U_i$, where $U_i$ ('errors') are independent RVs with $\mathrm{E}[U_i] = 0$ and $\mathrm{Var}(U_i) = \sigma^2$.

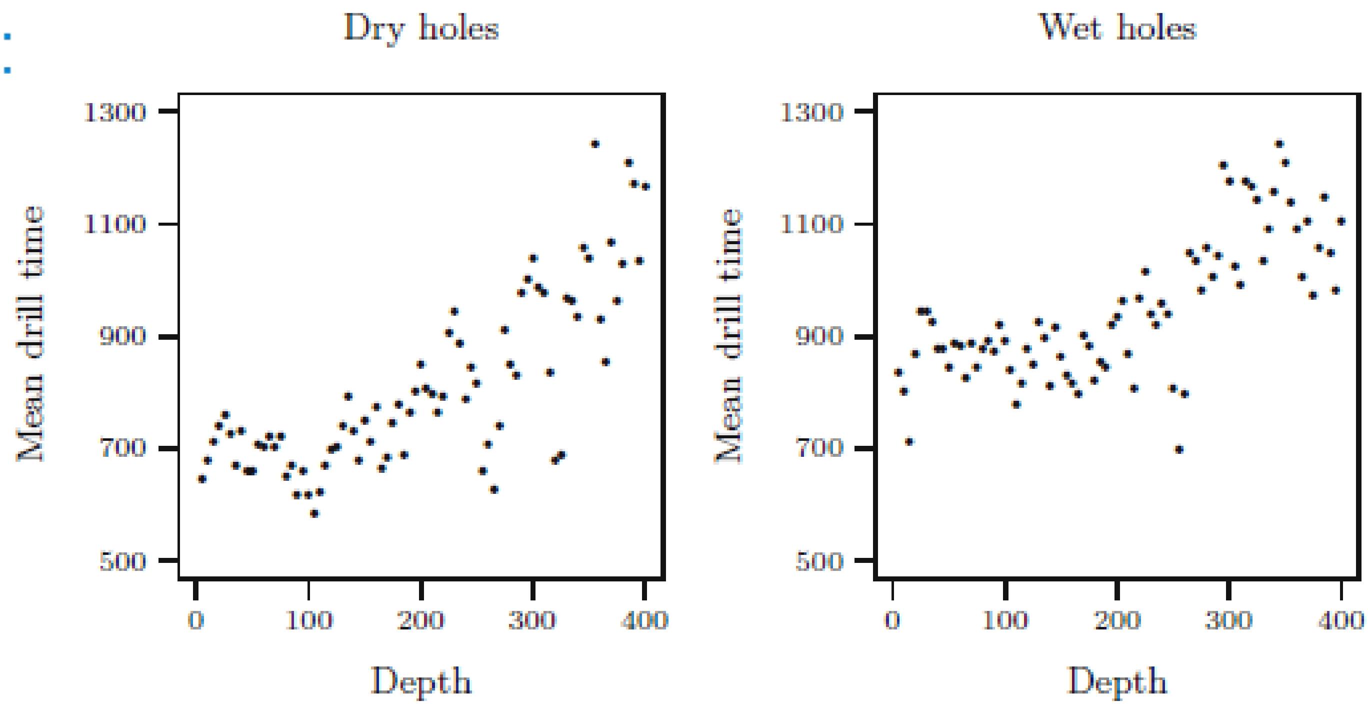



How to estimate $\alpha$ and $\beta$?

Recall the drilling dataset.

Estimate $\alpha_{\text{wet}}, \beta_{\text{wet}}, \alpha_{\text{dry}}$ and $\beta_{\text{dry}}$ by eye.

Then:




A) $\alpha_{\text{wet}} > \alpha_{\text{dry}}$ and $\beta_{\text{wet}} > \beta_{\text{dry}}$

B) $\alpha_{\text{wet}} < \alpha_{\text{dry}}$ and $\beta_{\text{wet}} > \beta_{\text{dry}}$

C) $\alpha_{\text{wet}} > \alpha_{\text{dry}}$ and $\beta_{\text{wet}} < \beta_{\text{dry}}$

D) $\alpha_{\text{wet}} < \alpha_{\text{dry}}$ and $\beta_{\text{wet}} < \beta_{\text{dry}}$

# MLE for regression line (errors normally distr.)

Suppose the $U_i$ have a $N(0, \sigma^2)$ distribution. Then the $Y_i$ have a $N(\alpha + \beta x_i, \sigma^2)$ distribution.
Using the maximum likelihood procedure we can show that the likelihood $L(\alpha, \beta, \sigma)$ attains its maximum when

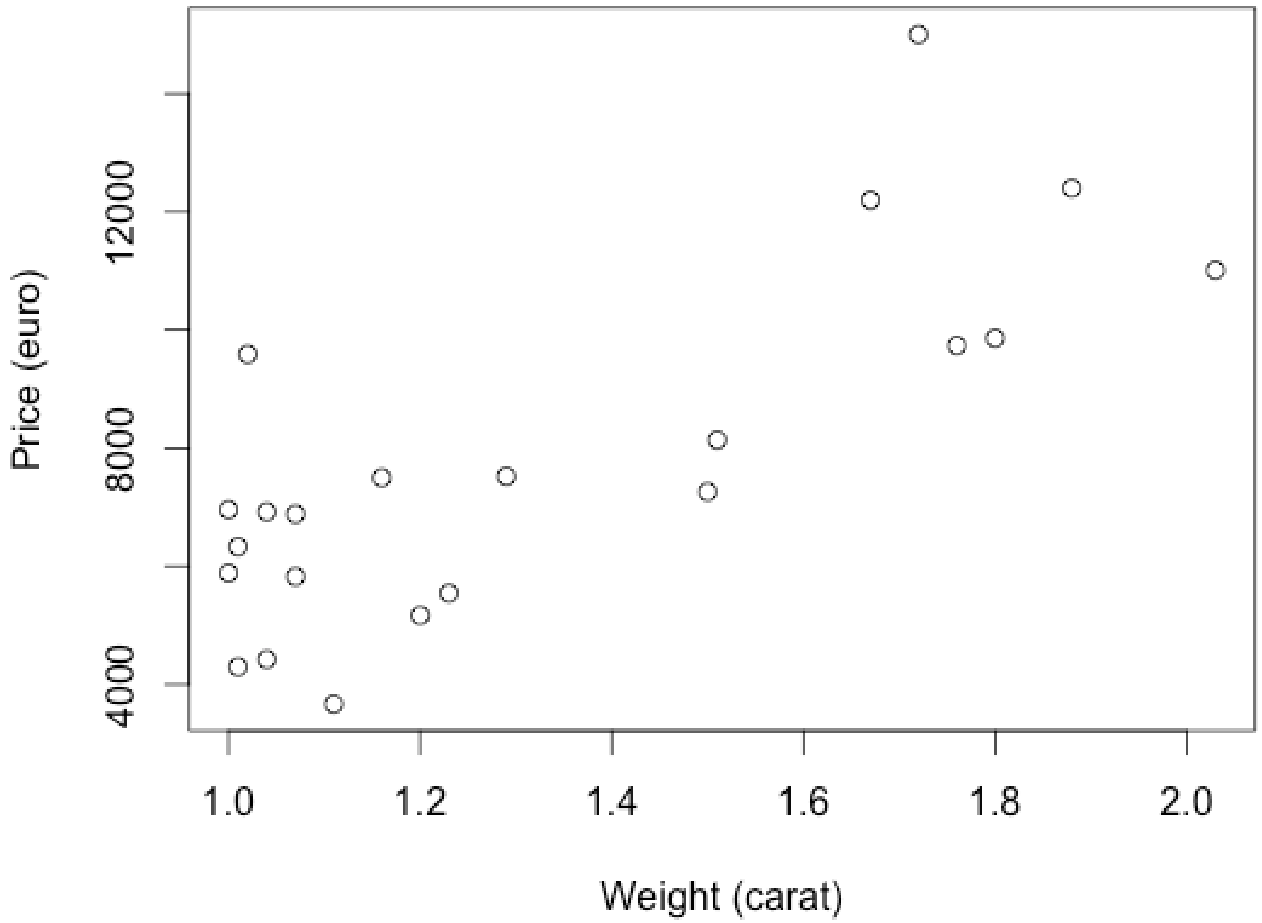
$$\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

is minimal. This yields the following MLE estimates:

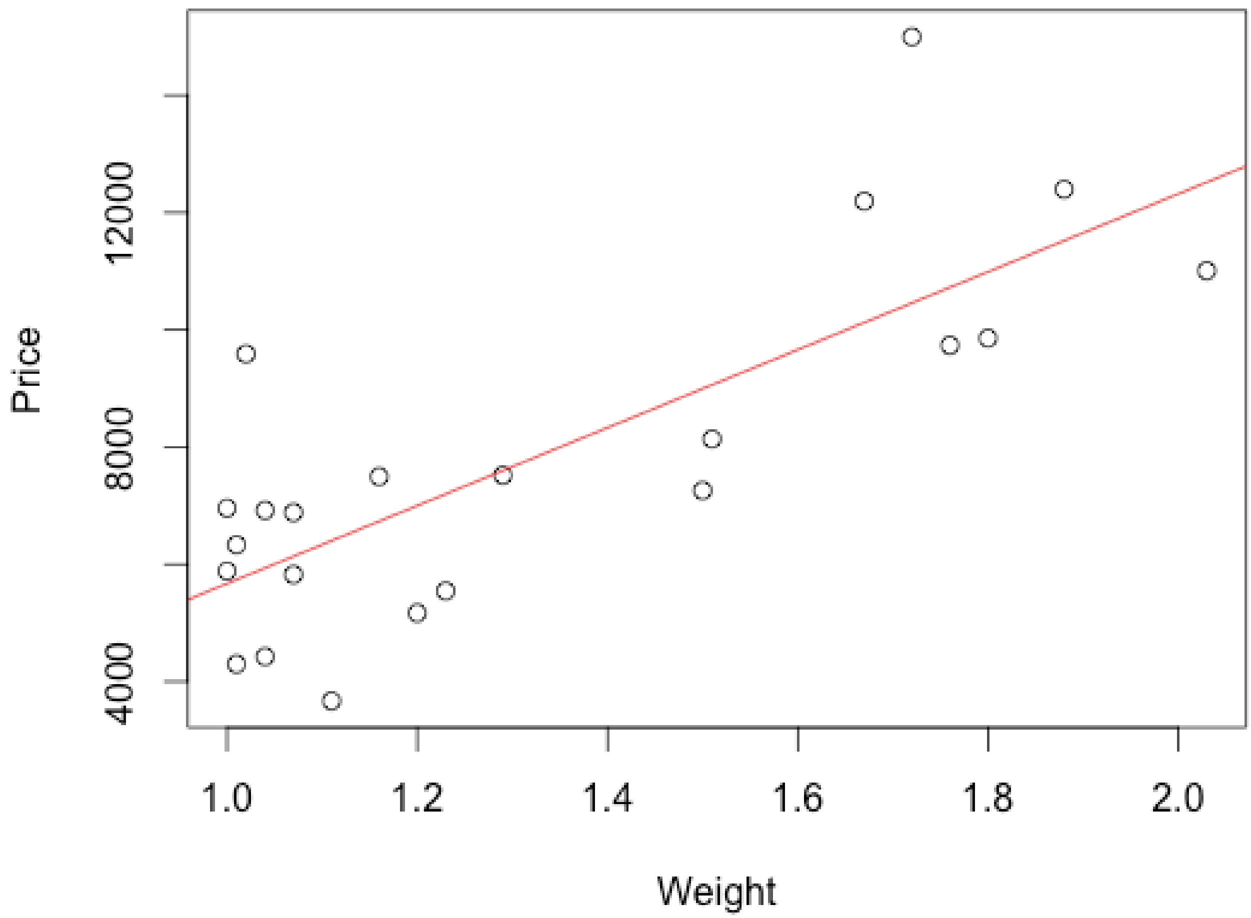
$$\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n,$$

$$\hat{\beta} = \frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{n \sum_i x_i^2 - (\sum_i x_i)^2}.$$

# Regression line: example



# Regression line: example

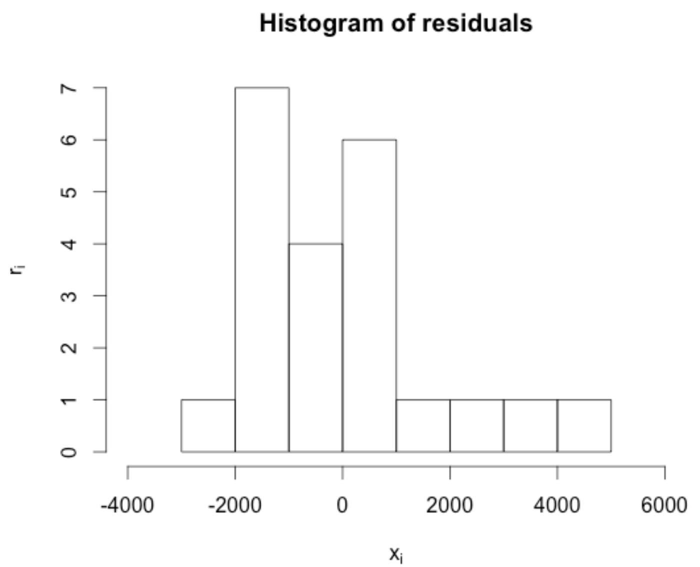


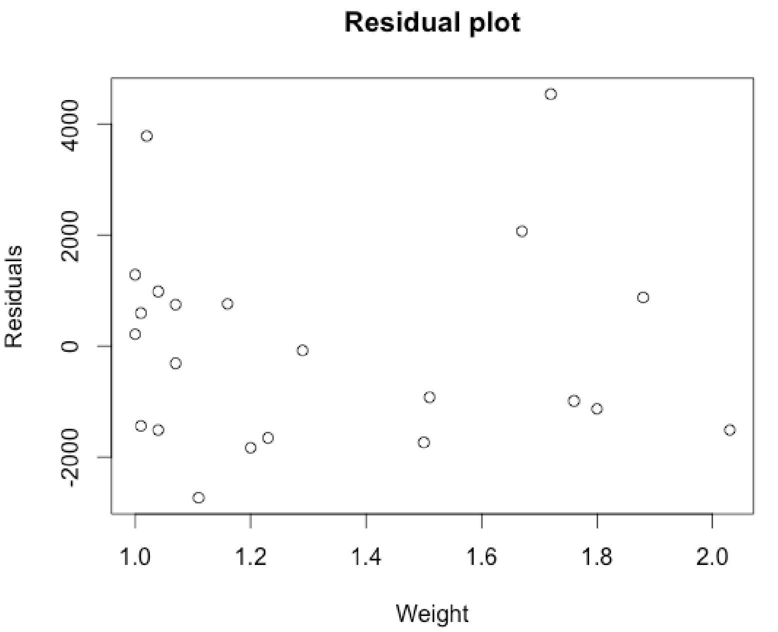

$$\hat{\alpha} = -976$$

$$\hat{\beta} = 6648$$

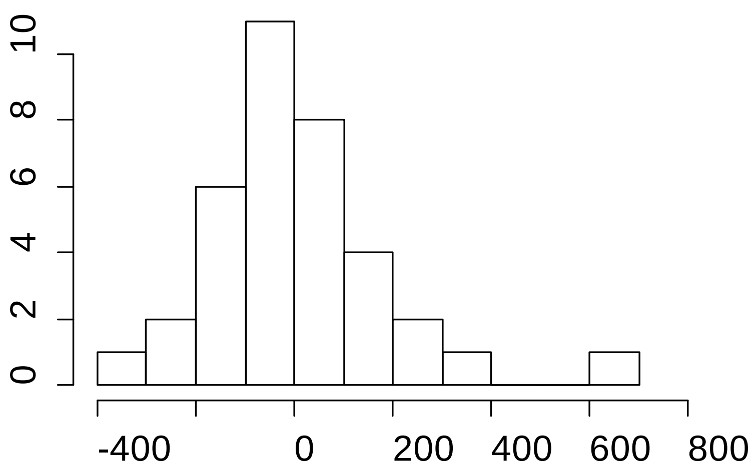
# Residuals

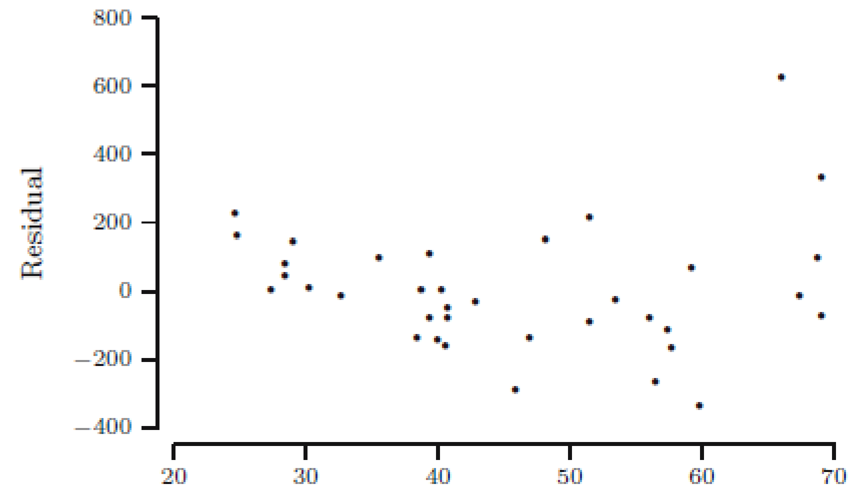
How to check whether the $U_i$ have a $N(0, \sigma^2)$ distribution?
Consider the *residuals* $r_i = y_i - \hat{\alpha} - \hat{\beta}x_i$.
The $r_i$ should look like a sample from a normal distribution.
Furthermore, a scatterplot of the $r_i$ versus the $x_i$ should not show any trend or pattern.



# Consider the following residuals.
# Is normality reasonable?
# And do they exhibit any pattern/trend?




A) Normality reasonable, no trend

B) Normality not reasonable, no trend

C) Normality reasonable, but trend

D) Normality not reasonable, and trend

# Method of least squares

What if the $U_i$ do **not** have a $N(0, \sigma^2)$ distribution?
(Of course, still $\mathrm{E}[U_i] = 0$ and $\mathrm{Var}(U_i) = \sigma^2$.)

Use the *method of least squares*: find $\alpha$ and $\beta$ such that

$$\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

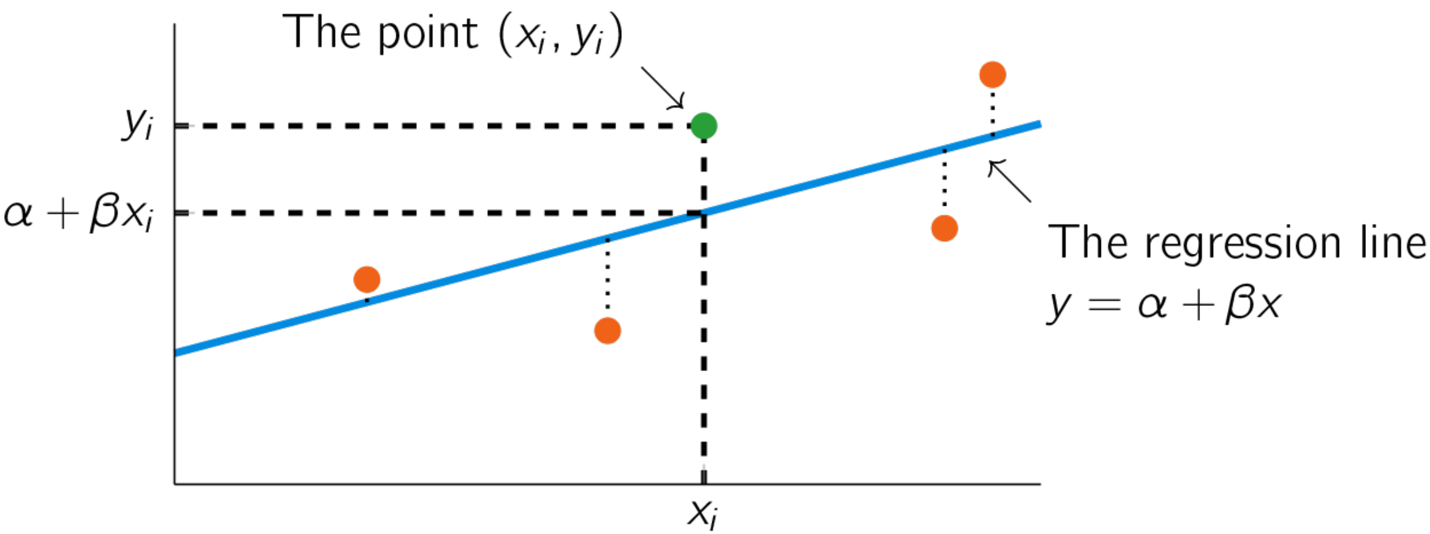
is minimal. We get the *least squares estimates*:

$$\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n,$$

$$\hat{\beta} = \frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{n \sum_i x_i^2 - (\sum_i x_i)^2}.$$

NB: they coincide (in case of normality) with the MLE estimates.

# Why ‘least squares’?




The least squares estimates minimize the sum of the squared distances between the points $y_i$ and their prediction $\alpha + \beta x_i$

# Two exercises

Suppose you have the following bivariate dataset

$$(1, 3.1), (1.7, 3.9), (2.1, 3.8)(2.5, 4.7), (2.7, 4.5).$$

Determine the least squares estimates for the regression line. You may use that $\sum x_i = 10, \sum y_i = 20, \sum x_i^2 = 21.84$ and $\sum x_i y_i = 41.61$.
Draw also the scatterplot and regression line in one figure.

---

Given a bivariate dataset $(x_1, y_1), \ldots, (x_n, y_n)$, consider a regression model without intercept, i.e. assume $Y_i = \beta x_i + U_i$. Derive the least squares estimate for $\beta$.

# For next class (week 3.8 lecture 1):

Complete MyStatlab assignments and book exercises

Watch prelectures *'Confidence intervals'*

Book: 23.1, 23.2, 23.4, 24.1, 24.2, 24.3, 24.4

After this class you can:

- Create and interpret confidence intervals in various settings
- Determine the sample size to achieve a given confidence level

# Statistics

Good luck!