Crash Course in Statistics
ZNZ 2026
V

Christoph Luchsinger and Zofia Baranczuk

Based on Script by Daniel J. Stekhoven

Tests and linear models

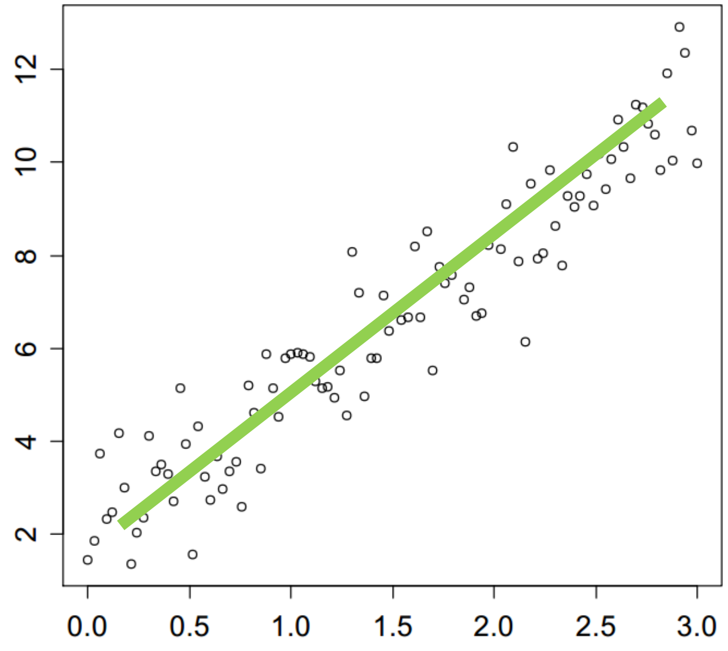
- Regression
- Diagnostics of regression
- R output of regression
- Multiple testing

Requirements for linear regression

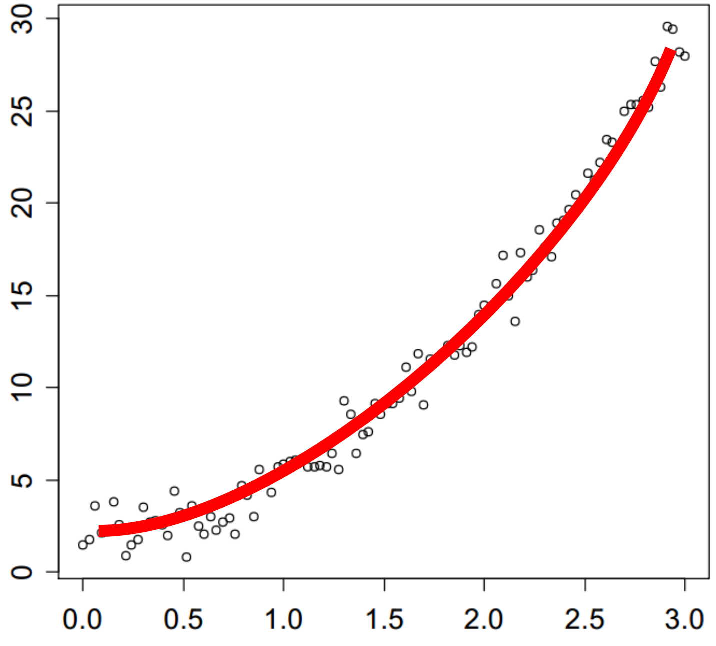
1. **Linearity** – Y can be explained using a linear combination, e.g., $Y = \beta_0 + \beta_1 \cdot X + \varepsilon$
2. **Constant variance** – error has constant variance (independent of X)
3. **Normality** – error ε needs to be normally distributed

If these assumptions are strongly violated the model is not valid

1. Linearity: Scatter plot for simple linear regression



OK

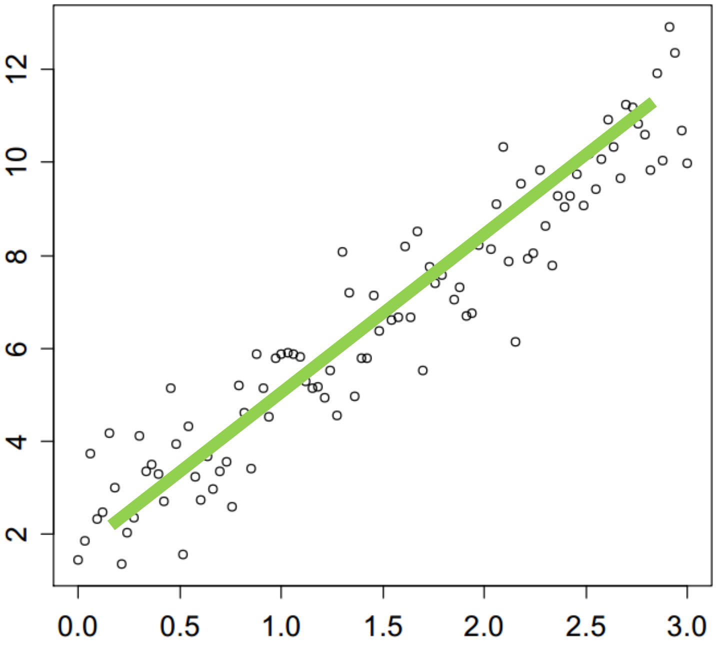


Systematic error

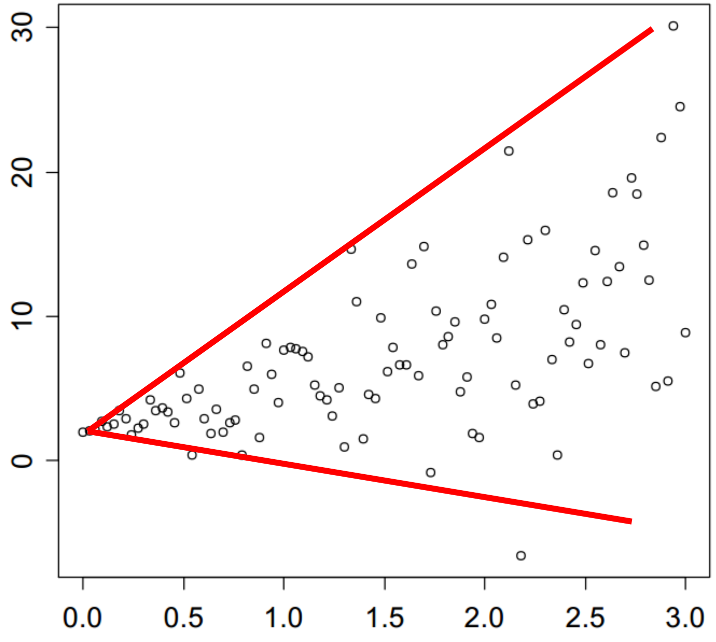
Curvature:

$$y = b_0 + b_1 x + b_2 x^2$$

2. Constant Variance? Scatter plot for simple linear regression

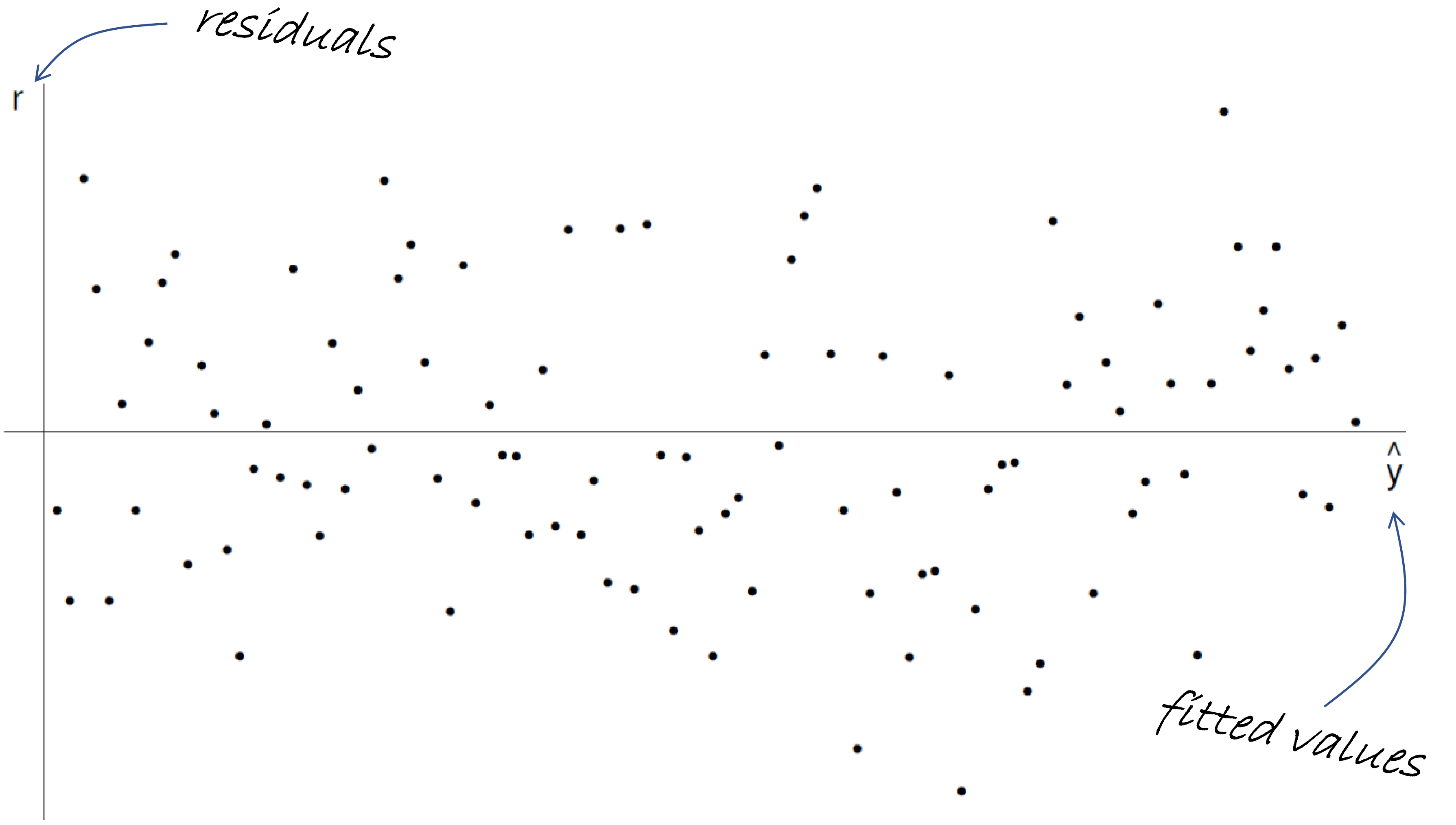


OK

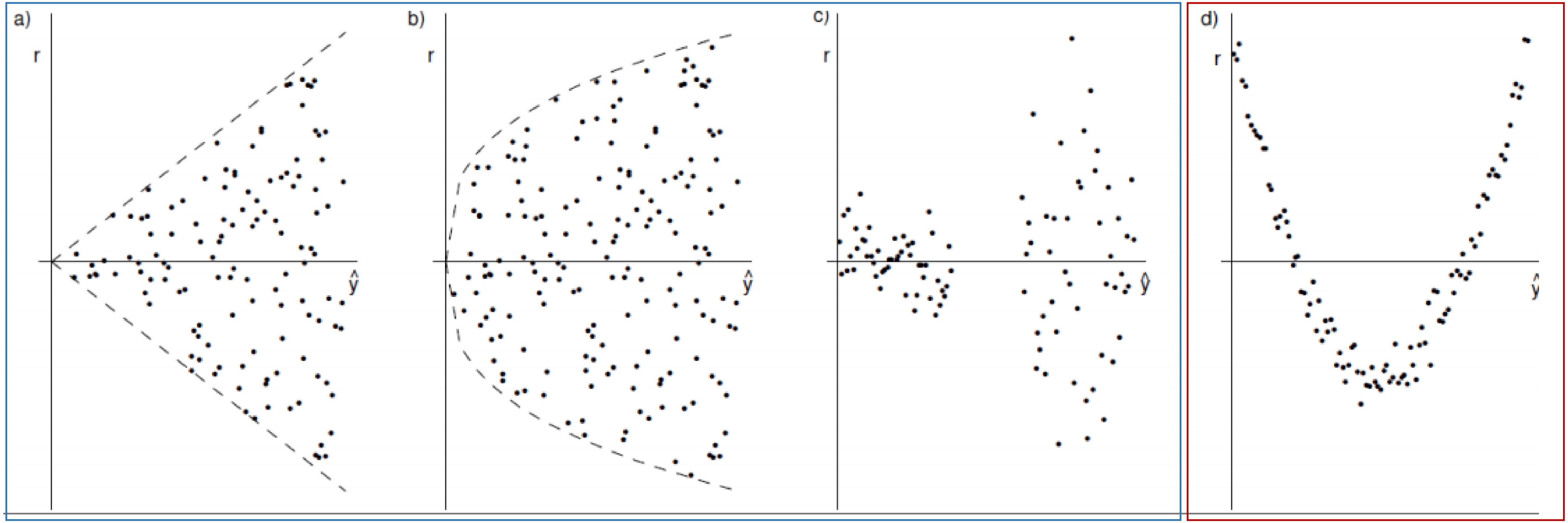


Error variance is not constant

Constant variance: Example for a good TA-plot (Tukey-Anscombe-Plot)



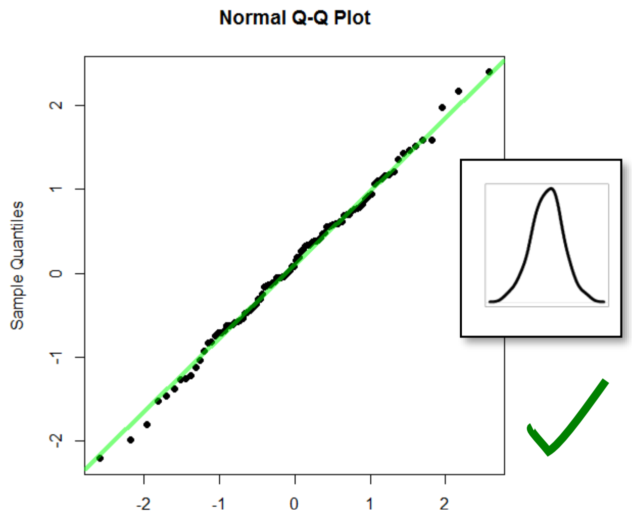
Constant variance: Examples for bad TA-plots



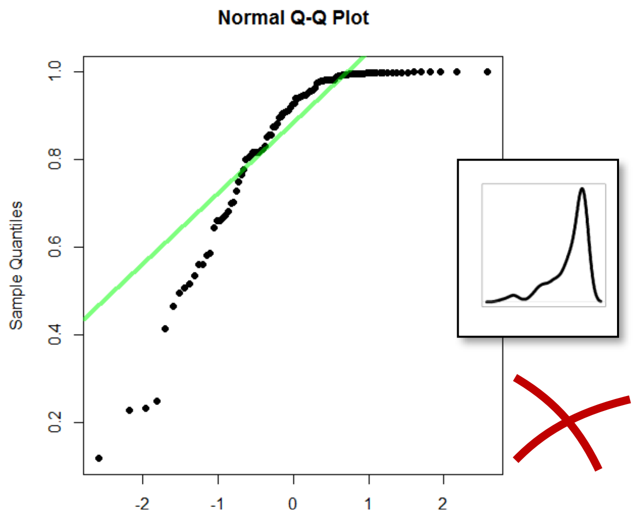
Error variance not constant

Systematic error

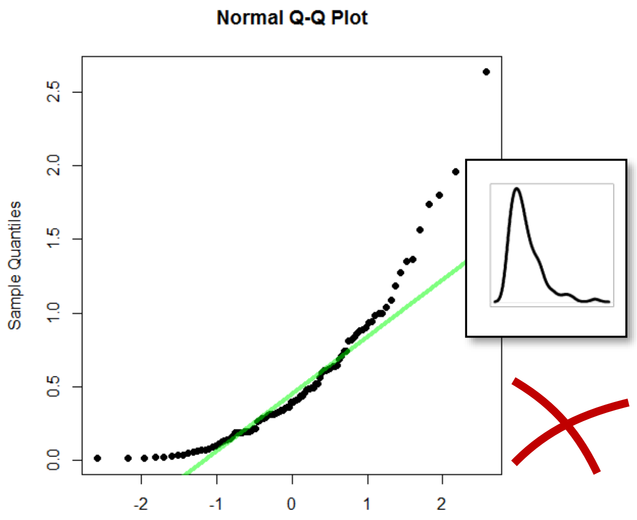
3. Normality: QQ-Plot



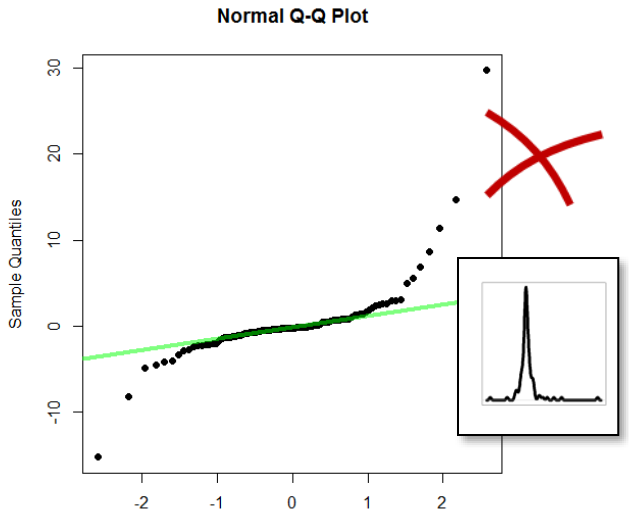
normal



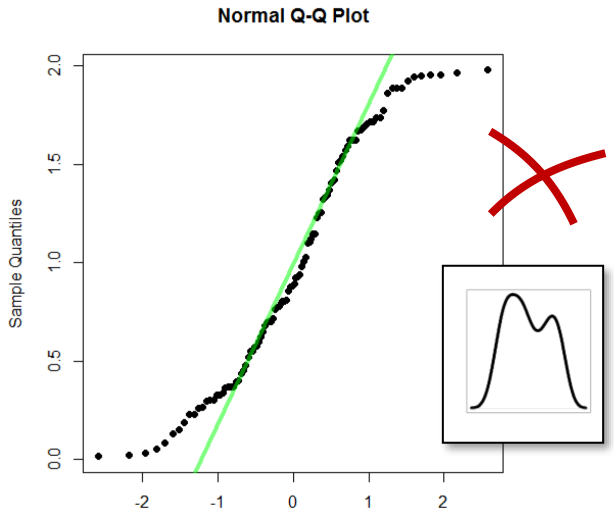
left skewed



right skewed

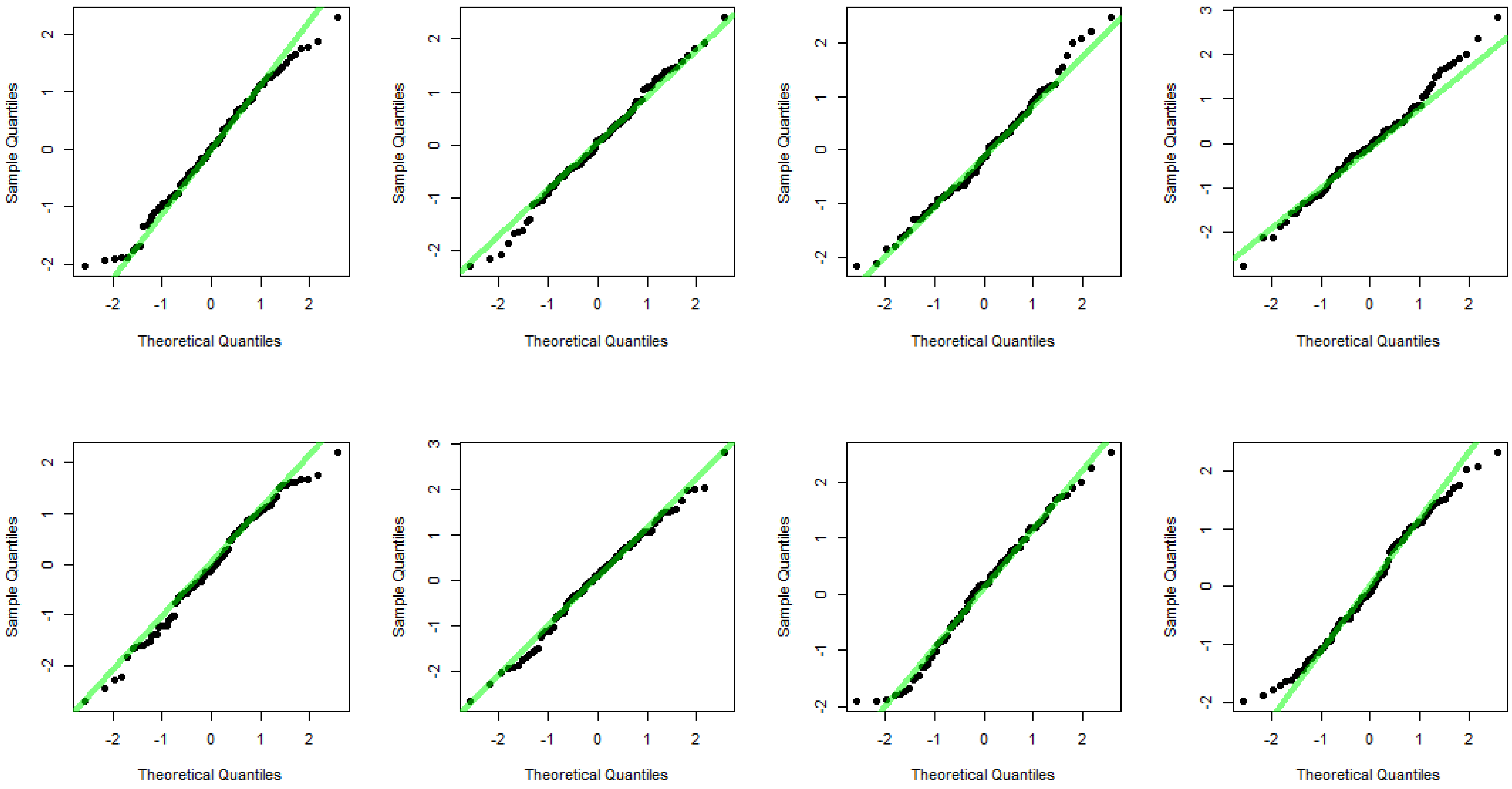


heavy tailed



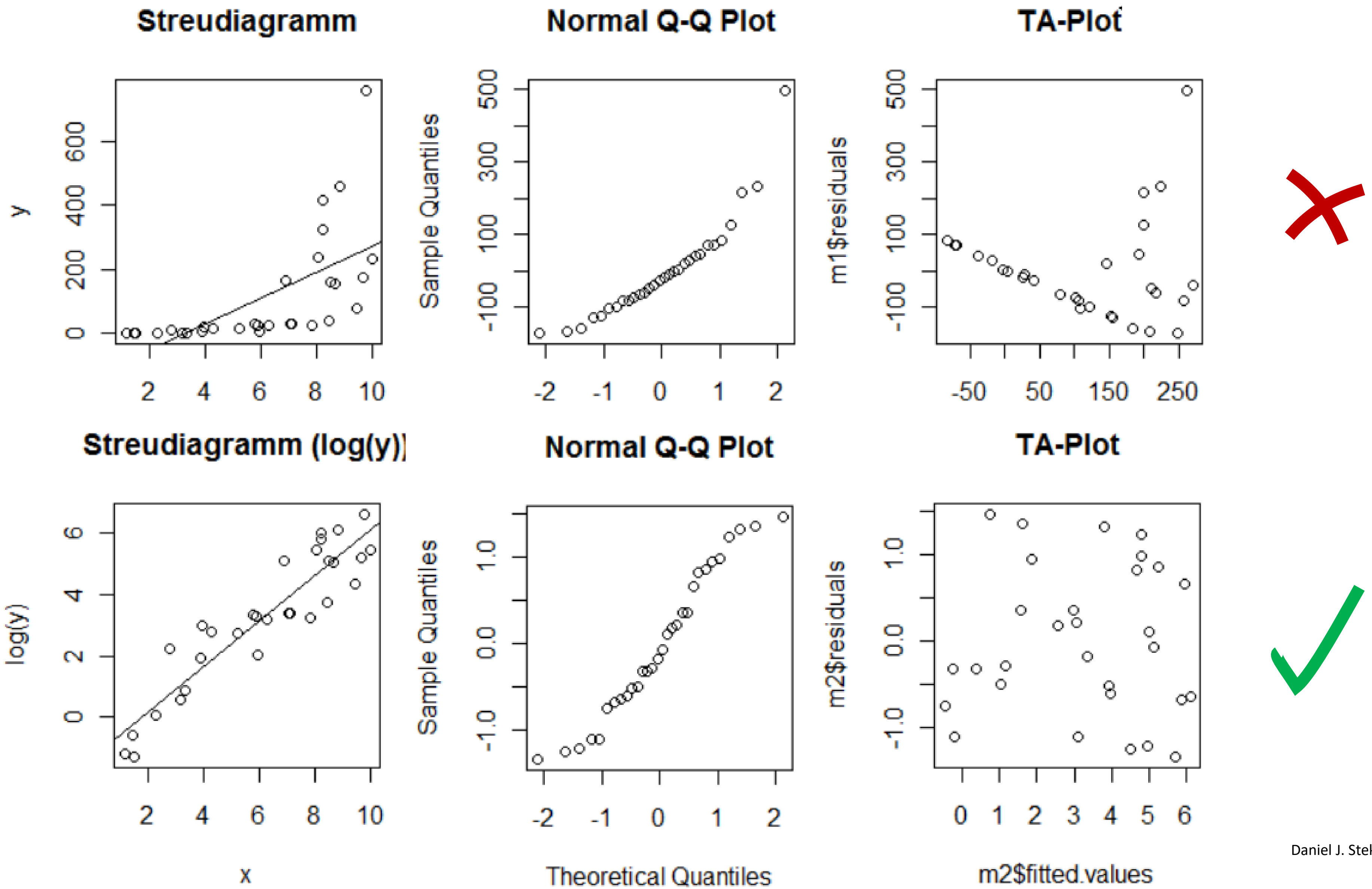
short tailed

QQ-Plots: what are «good» QQ-Plots



- $n = 100, R_i \sim \mathcal{N}(0,1)$

…what if residual analysis is haywire (drunter & drüber)



Aerobic performance

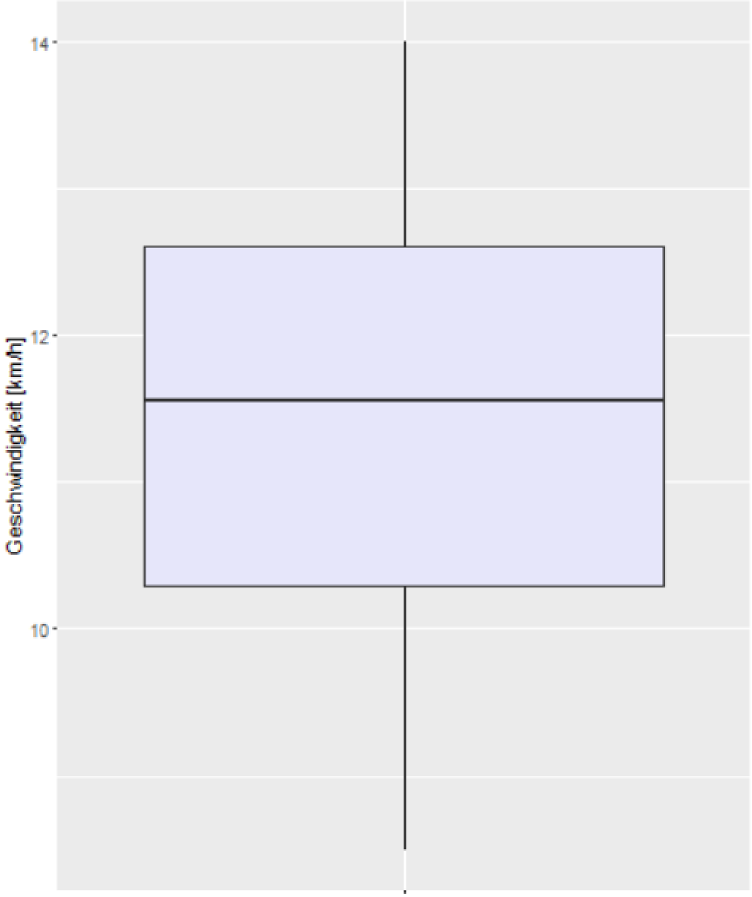
- VO_2max: amount of oxygen, the body can absorb per kg mass and minute

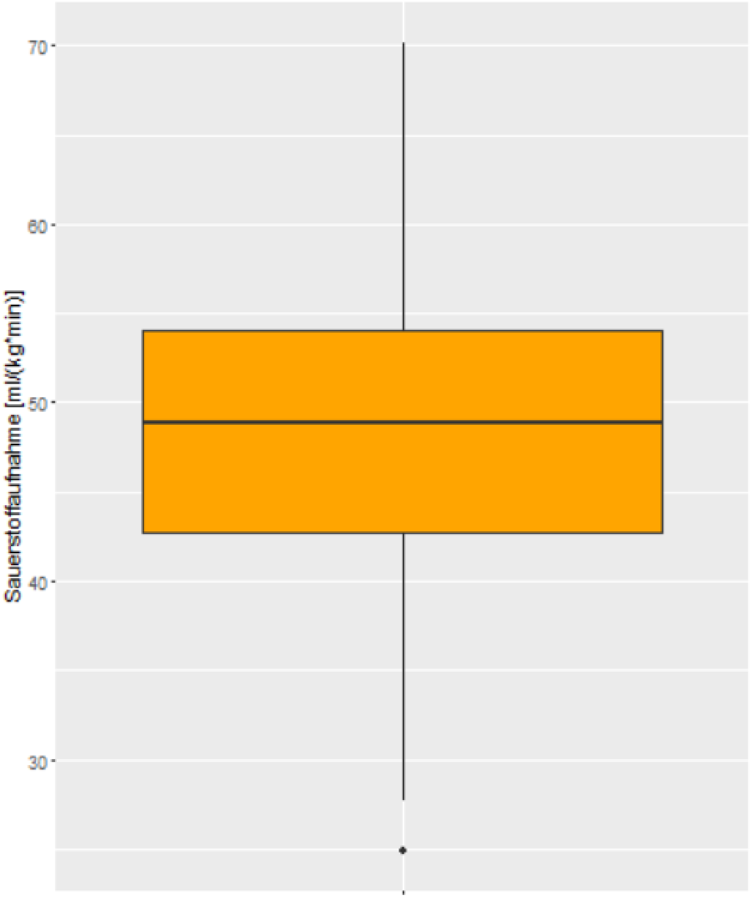
- Test is expensive and effortful
- Not meant for the broad community

- Alternative?

Léger et al., 1983

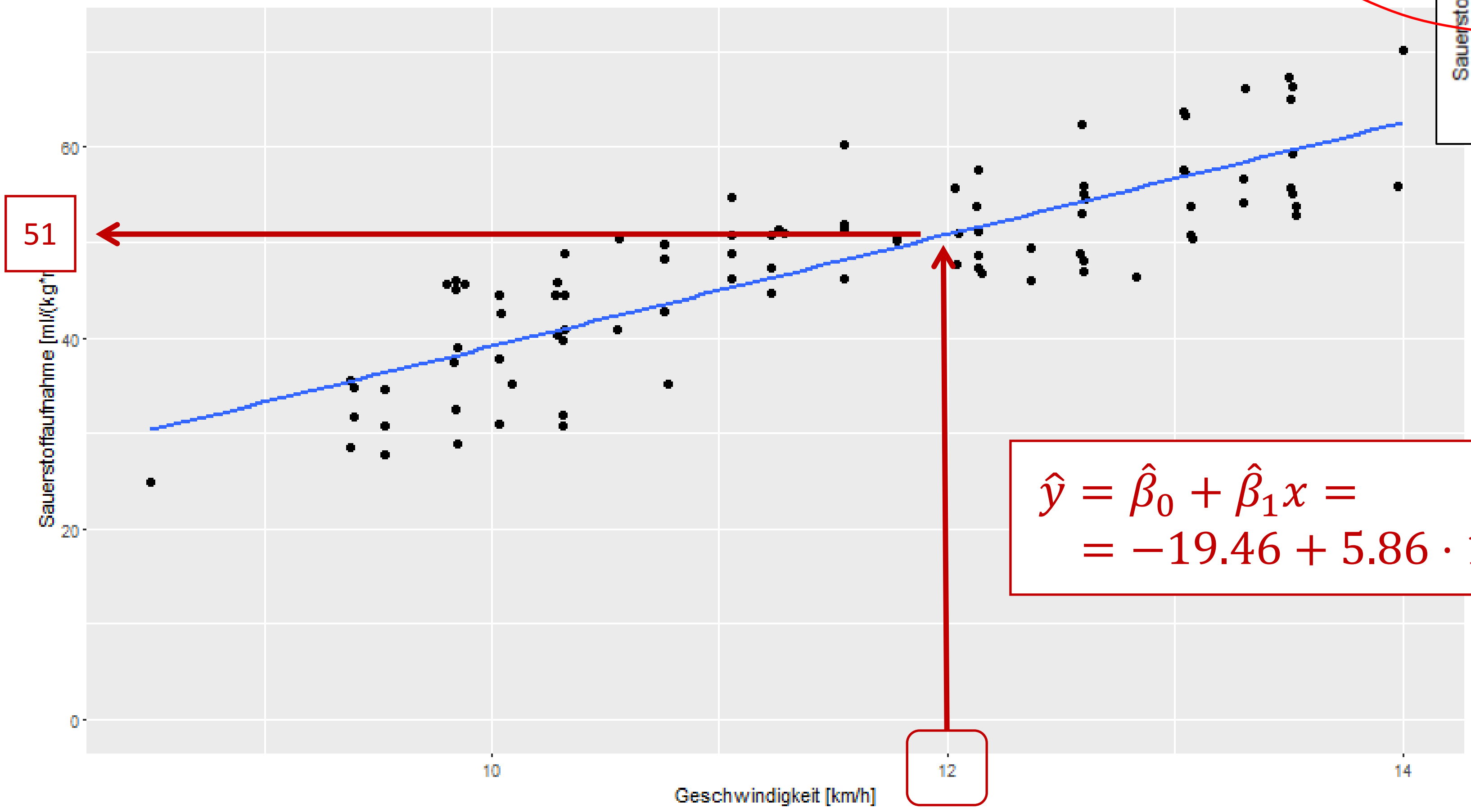
- 91 subjects, 20m-shuttle-test and VO_2max measurement

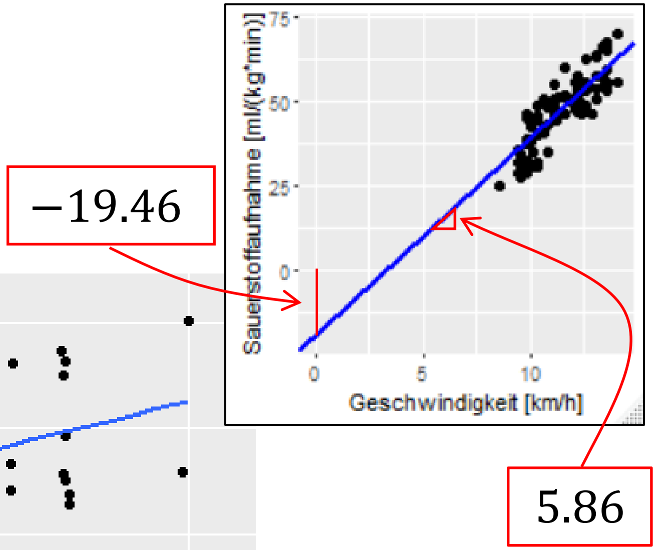




Aerobic Performance

- $\hat{\beta}_0 = -19.46$
- $\hat{\beta}_1 = 5.86$
- $\hat{\sigma} = 5.4$





$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \\ = -19.46 + 5.86 \cdot 12 \approx 51$$

Linear regression in R

- Model: $Y_i = \beta_0 + \beta_1 x_i + E_i, E_i \sim \mathcal{N}(0, \sigma^2)\ i.i.d.$
- Model: $Y_i = \mathbf{-19.46} + \mathbf{5.86} \cdot x_i + E_i, E_i \sim \mathcal{N}(0, \mathbf{5.43}^2)\ i.i.d$

Standard error of $\widehat{\beta}_1$
approx. 95%-CI:
$5.86 \pm 2 \cdot 0.41$
exact 95%-CI:
$5.86 \pm 1.99 \cdot 0.41$

$t_{89}; 0.975$

```
> fit <- lm(vo2max ~ vmax, data = dat)
> summary(fit)

Call:
lm(formula = vo2max ~ vmax, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-10.2230  -4.3976  -0.2016   4.7026  12.0348

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.4582     4.7239  -4.119  8.5e-05 ***
vmax          5.8566     0.4082  14.347  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.433 on 89 degrees of freedom
Multiple R-squared:  0.6981, Adjusted R-squared:  0.6948
F-statistic: 205.8 on 1 and 89 DF,  p-value: < 2.2e-16
```

Observed Test statistic t
in the test:
$\mathcal{H}_0: \beta_1 = 0$ vs $\mathcal{H}_A: \beta_1 \neq 0$

P-value:
Assume $\beta_1 = 0$; what is the probability of t or an even more extreme value?

Degrees of freedom:
$n - k = 91 - 2 = 89$

Multiple testing correction

- Significance level α
  - Probability to reject H_0 (*alarm goes off*), given it is true (*there is no fire*).

Is there a fire?	
Yes	No
TP ere's a fire, larm goes off!	FP There's no fire, the alarm goes off!
FN ere's a fire, m does'nt go off!	TN There's no fire, the alarm doesn't go off!

  - Traditionally set to 0.05 (5%)
- If a test is performed on this level under H_0, we have a 5% chance that it gives a wrong result…

Multiple testing correction

- Let's test 100 samples versus a reference with $\alpha = 0.05$ each (1 out of 20, 1/20)
- Assume that $\mathcal{H}_0$ is always true → there's never a difference!
- The chance of observing at least one significant result is

  1-P[no significant result]

- No significant result means, the test will not reject, given $\mathcal{H}_0$ is true – the chance for that is 1-α
- Doing 100 tests the chance of no significant result becomes

  $(1-\alpha)^{100} = (0.95)^{100} = 0.0059$

…and thus the chance to observe at least one significant result is

1-0.0059 = 0.9941

Multiple testing correction

- This is corrected by adjusting the significance levels of the test (or equivalently, the resulting p-values)

- Methods are:
  - Bonferroni Correction (very stringent)
  - Benjamini-Hochberg (less stringent, better for large scale data)
  - …

Multiple testing correction

- Example from before using Bonferroni:
  - P[at least one significant result] = 1-(1-α)^{100}
  - Bonferroni: divide α by the number of tests

$$\tilde{\alpha} = \frac{\alpha}{n}$$

  - then

$$1-(1-\tilde{\alpha})^{100} = 1-(1-0.0005)^{100} = 0.0488$$

this is quite tough

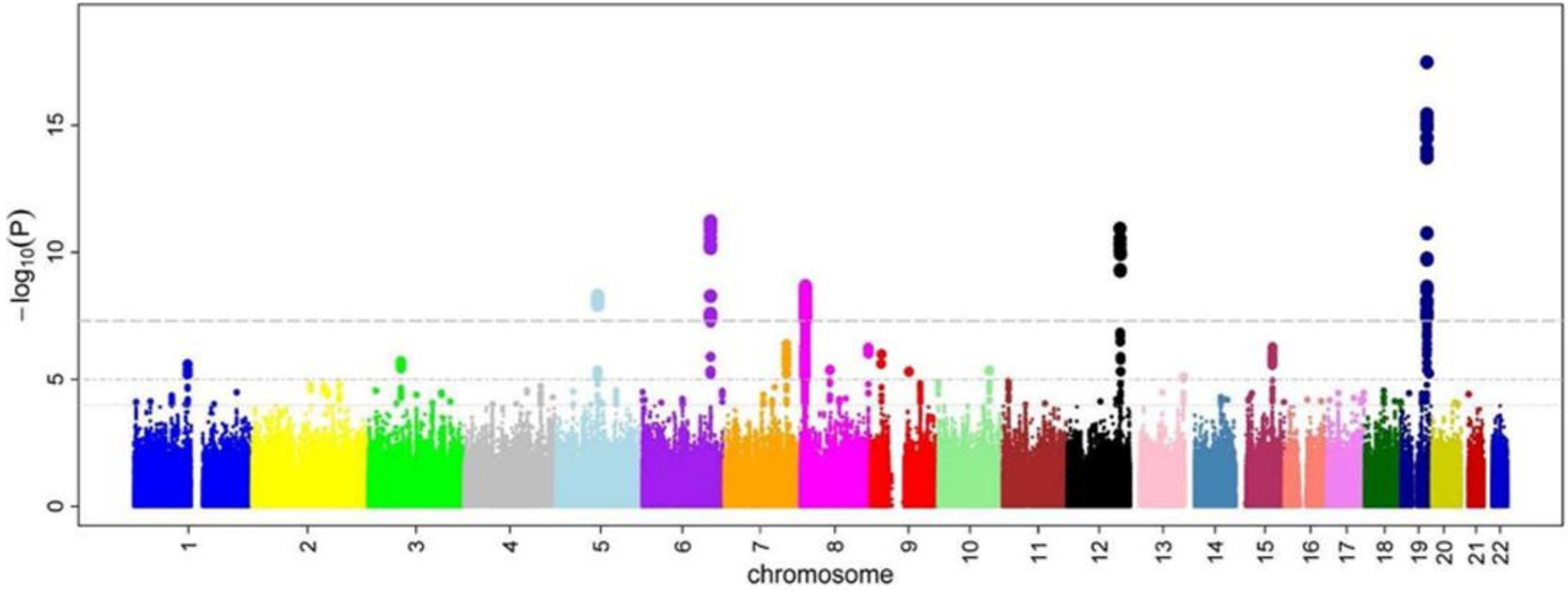
Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Full text is freely available on Wikipedia

Manhattan plot



Summary

- Make sure that your model is the right one
- If not either transform data or change the model
- Multiple testing is ruining everyone's day … but we need to do it

- Now solve exercise 2 and 3, you will need:
- In 2) qt(0.025, 50)=-2.008559 and qt(0.05, 50)=-1.675905
- In 3) qt(0.025, 7)=-2.364624 and qt(0.05, 7)=-1.894579