Common mistakes in experimental design and analysis of pig trials: an overview.

B. Ampe and C. De Cuyper[1]

EAAP 2024
04-09-2024

ILVO

ILVO

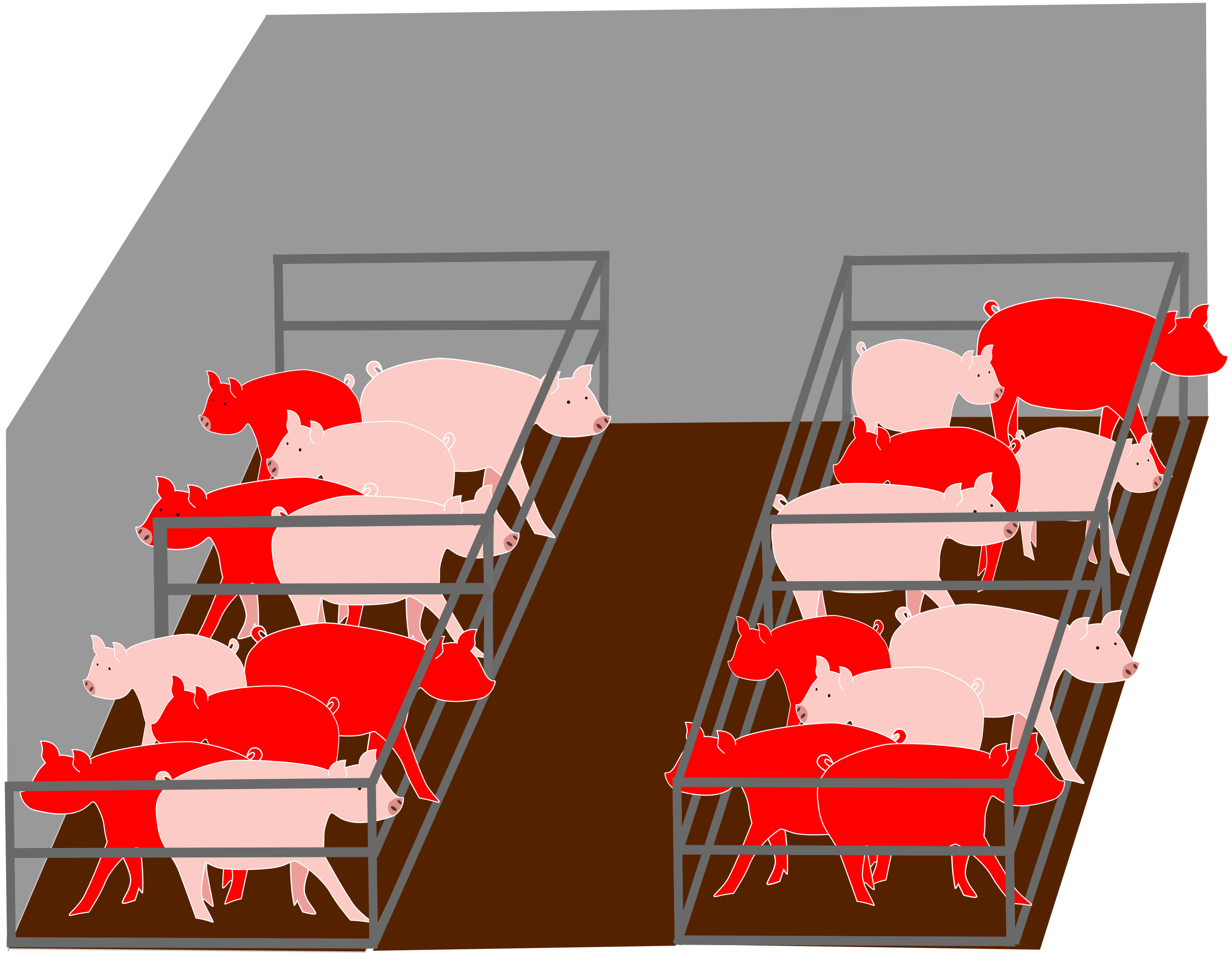
Outline

1. Design terminology
2. Blocking
3. Graphs vs Statistical analysis
4. Diagnostic plots
5. Different or equivalent
6. Interactions
7. Wrong interpretations

Experimenal vs observational unit

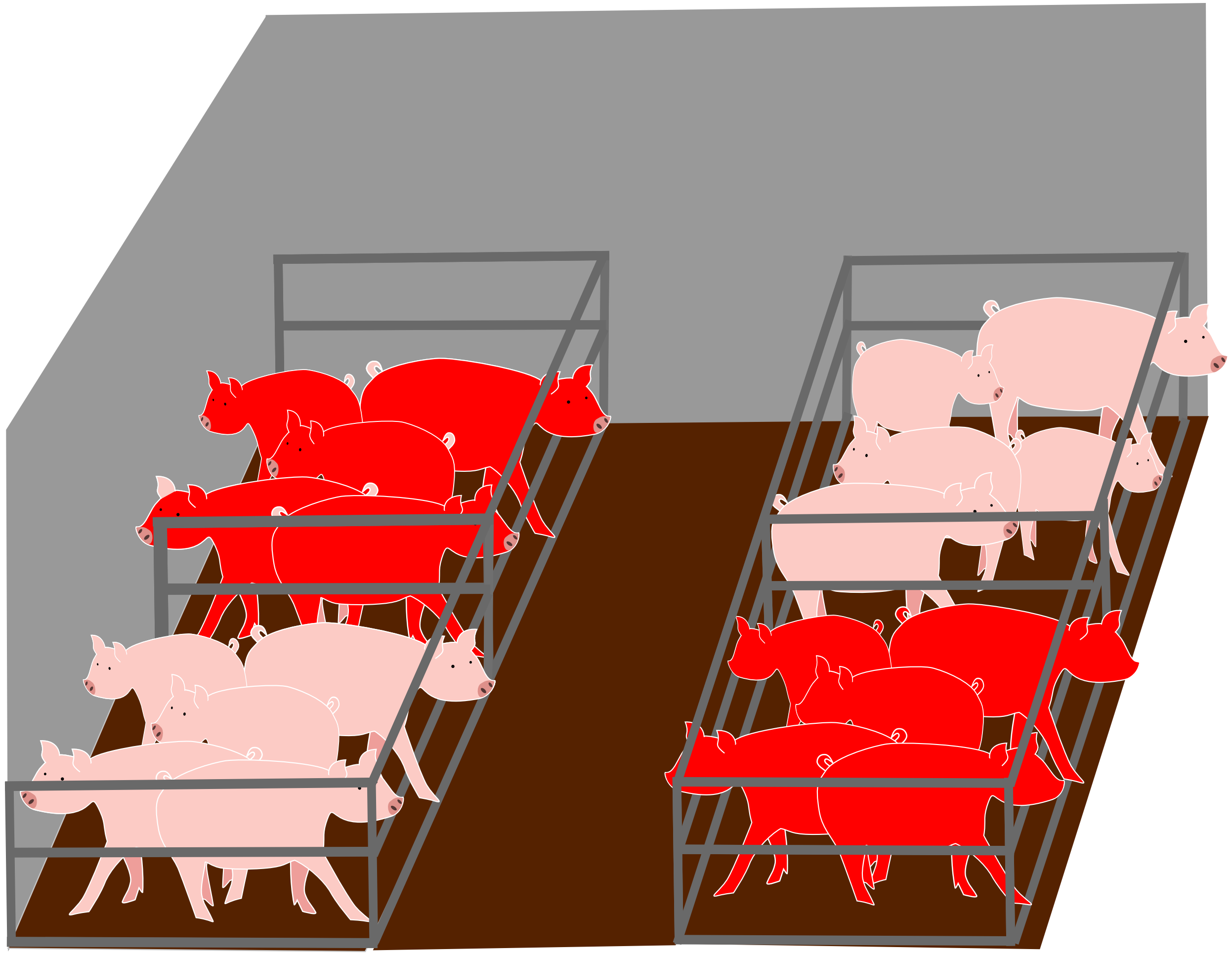
- Experimental unit: The unit to which the treatment is randomly applied.
- Observational unit: The unit on which the response is measured

Experimenal vs observational unit: example 1



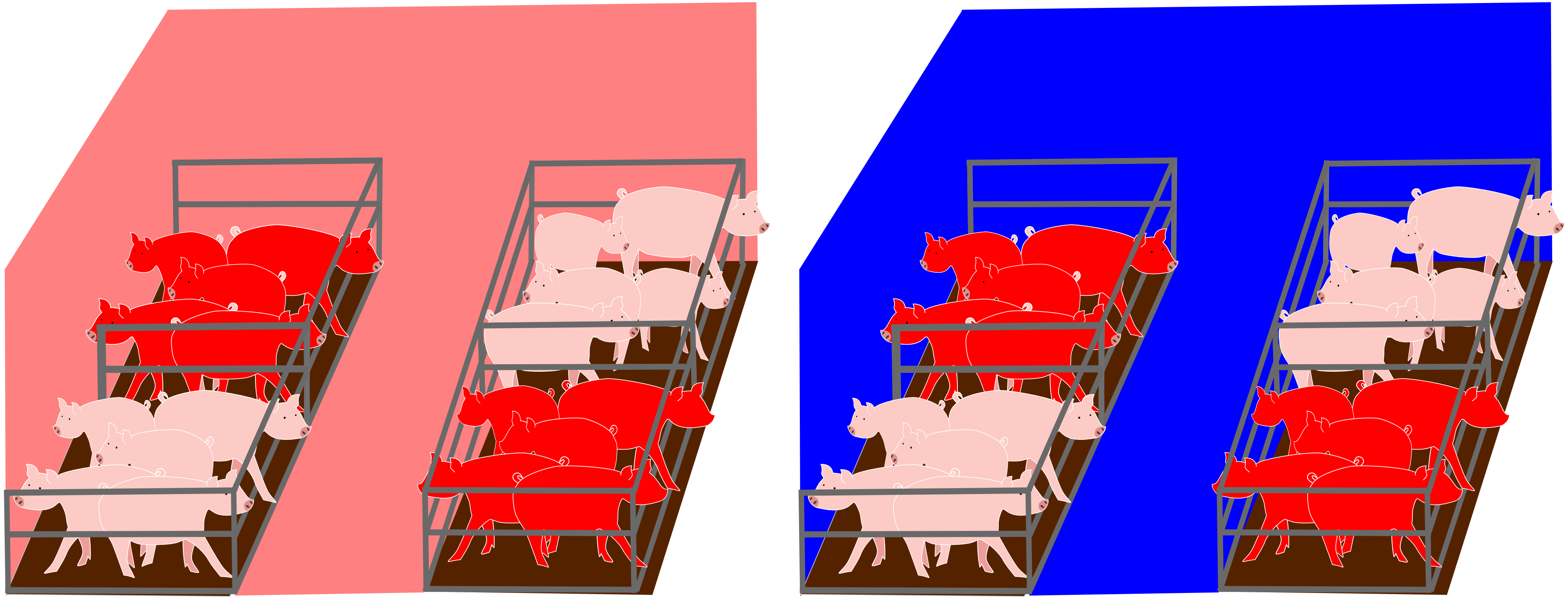
ILVO

Experimenal vs observational unit: example 2



ILVO

Experimenal vs observational unit: example 3

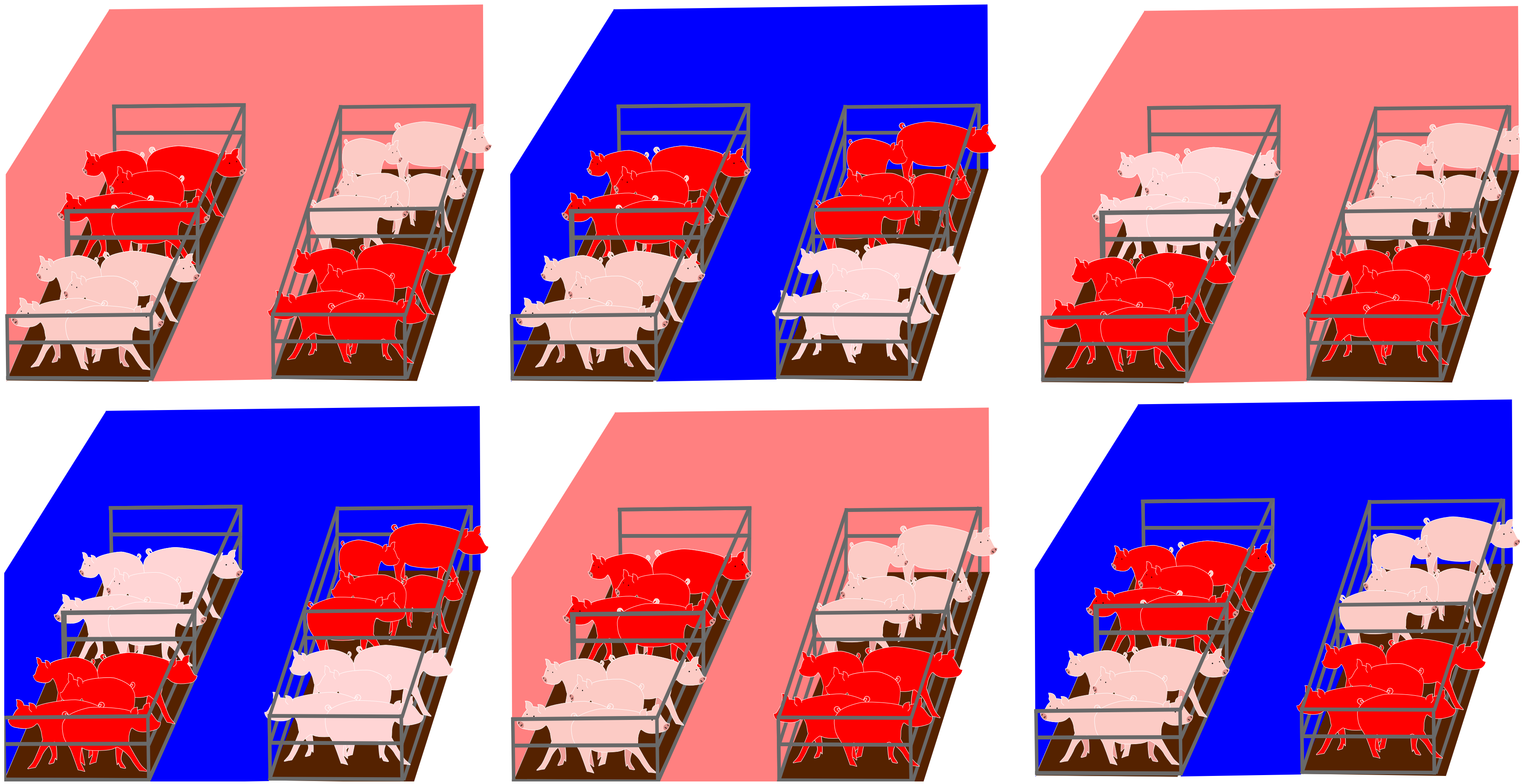


replicate vs repeated measurement

- **Replication** involves running the same study on **different** subjects, independently but under identical conditions

- **Repeated** measures involves measuring the **same** subjects multiple times

A heatstress example:



ILVO

“To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination, he may be able to say what the experiment died of.”

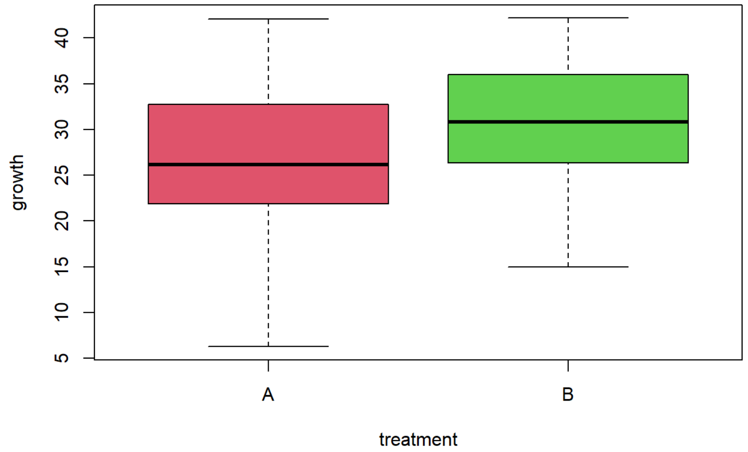
R.A. Fisher, Indian Statistical Congress, Sankhya, ca 1938

An example of an experiment AB

- Test the difference in growth of **food A vs B**
- Weight at start and end: **growth of each animal**
- **24 pens** available, 5 animals/pen = **120 animals**
- Pens blocked according to start weight: **low, mean and high** (8 pens each)
- Within each weight group: **randomization** for treatment on **pen-level** (balanced)

Analysis experiment AB (1)

- **Simple linear model (ANOVA):**

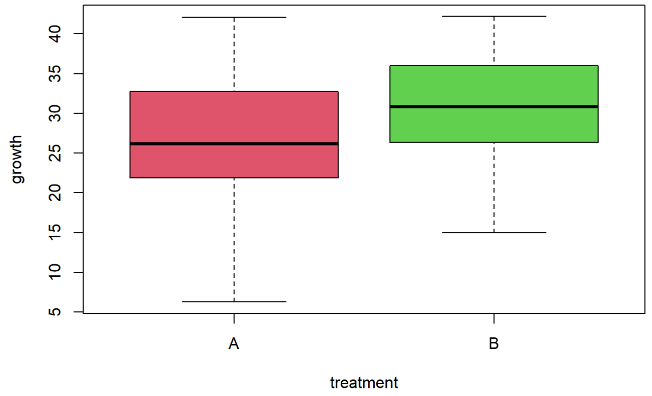



```
Call:
lm(formula = growth ~ treatment, data = data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.1645     0.9414  27.794  < 2e-16
treatmentB    4.7095     1.3313   3.537 0.000579

Residual standard error: 7.292 on 118 degrees of
freedom
Multiple R-squared:  0.09588, Adjusted R-squared:
0.08822
F-statistic: 12.51 on 1 and 118 DF,  p-value:
0.0005786
```

ILVO

Analysis experiment AB (2)

- **Simple linear model (ANOVA):**




```
Anova Table (Type III tests)

Response: growth
            Sum Sq  Df F value    Pr(>F)    
(Intercept)  41075   1 772.481 < 2.2e-16 ***
treatment      665   1  12.513 0.0005786 ***
Residuals     6274 118
```

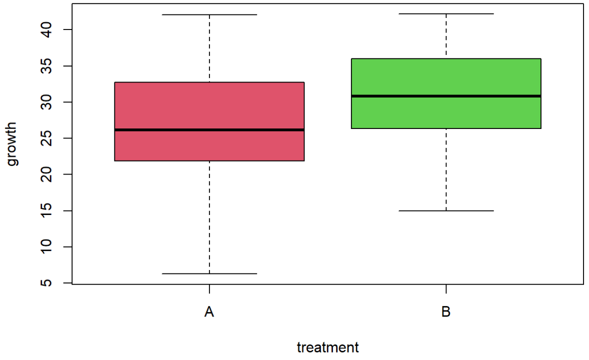
Simple linear model (ANOVA) on pen-means:

```
Anova Table (Type III tests)

Response: growth
            Sum Sq Df F value    Pr(>F)    
(Intercept) 8215.0  1 181.741 4.123e-12 ***
treatment    133.1  1   2.944    0.1002    
Residuals    994.4 22
```

ILVO

Analysis experiment AB (3)

- **Simple linear model (ANOVA):**




```
Anova Table (Type III tests)

Response: growth
             Sum Sq  Df F value    Pr(>F)
(Intercept)  41075   1 772.481 < 2.2e-16 ***
treatment      665   1  12.513 0.0005786 ***
Residuals     6274 118
```

linear MIXED model (ANOVA) pen included as random:

```
Analysis of Deviance Table (Type III Wald F
tests with Kenward-Roger df)

Response: growth
                  F Df Df.res    Pr(>F)
(Intercept) 181.741  1     22 4.123e-12 ***
treatment     2.944  1     22    0.1002
```

ILVO

Conclusion:

NEVER IGNORE DESIGN:

Analysis on level of experimentational unit!

Analysis experiment AB (4)

linear MIXED model pen included as random:

```
Analysis of Deviance Table (Type III Wald F tests
with Kenward-Roger df)

Response: growth
                  F Df Df.res    Pr(>F)
(Intercept) 181.741  1     22 4.123e-12 ***
treatment     2.944  1     22    0.1002
```

linear MIXED model pen included as random and weight group (block):

```
Analysis of Deviance Table (Type III Wald F tests
with Kenward-Roger df)

Response: growth
                   F Df Df.res    Pr(>F)
(Intercept) 108.1645  1     20 1.629e-09 ***
treatment     5.7537  1     20 0.0263211 *
block        11.4981  2     20 0.0004742 ***
```

ILVO

Conclusion:

NEVER IGNORE DESIGN:

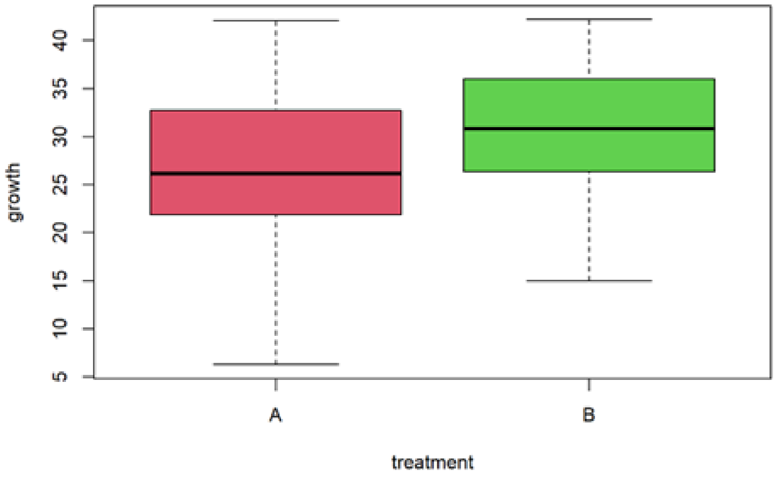
In case of blocking:

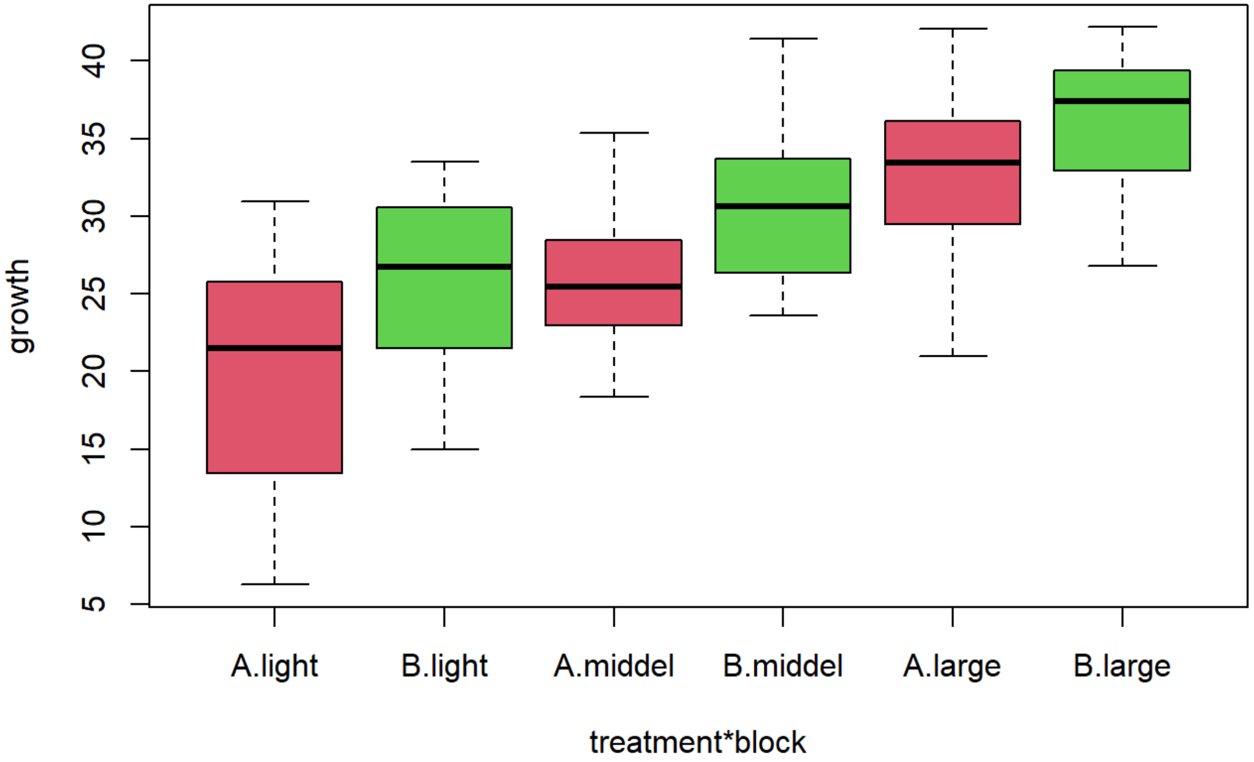
include block in your model!

ILVO

Analysis experiment AB (4b)

weight group included:






```
Analysis of Deviance Table (Type III Wald
F tests with Kenward-Roger df)

Response: growth
                   F Df Df.res    Pr(>F)
(Intercept) 108.1645  1     20 1.629e-09
treatment     5.7537  1     20 0.0263211
block        11.4981  2     20 0.0004742
```

“Block what you can and randomize what you cannot.”

Box, Hunter and Hunter 1978

ILVO

Another Example analysis: dataset

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Adapted from Anscombe, Francis J. (1973). Graphs in statistical analysis. *The American Statistician*, **27**, 17–21. doi: 10.2307/2682899.

Summary statistics

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

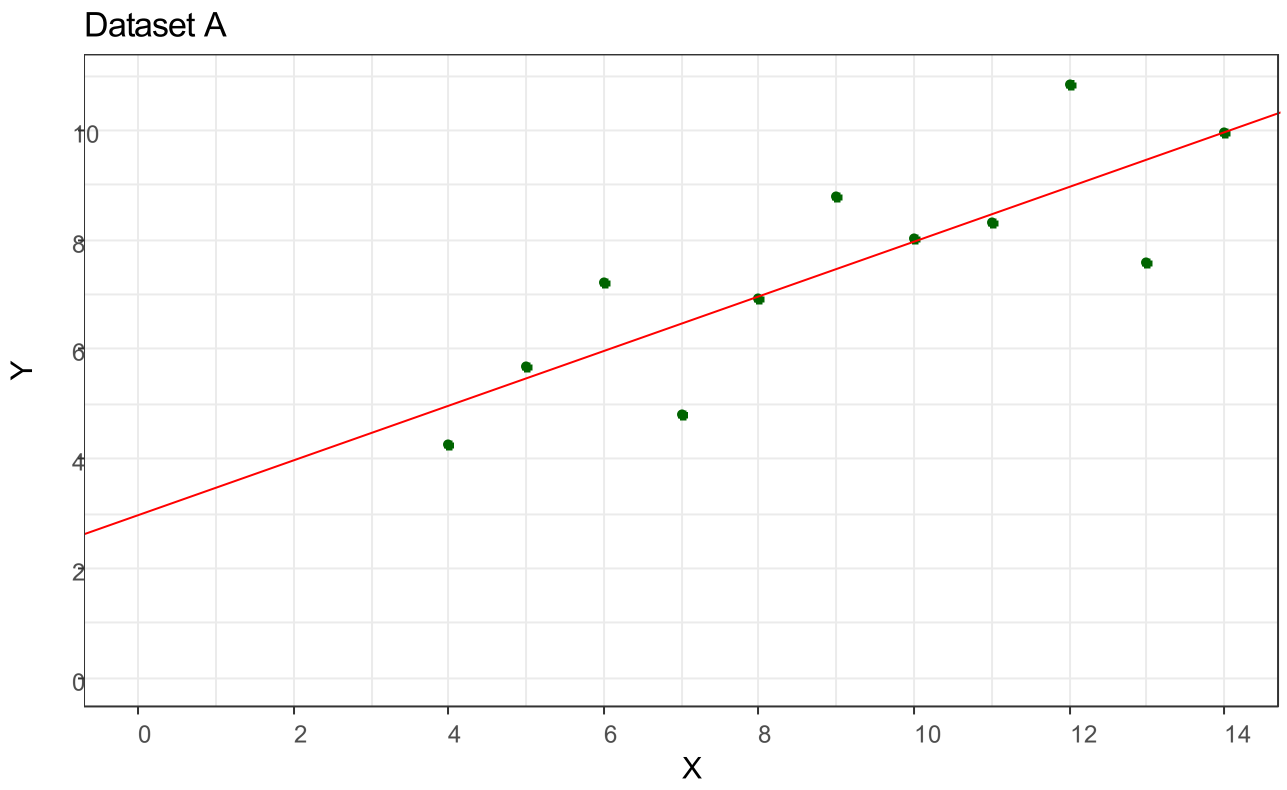
```
Variable: X
  mean        sd  n
A     9 3.316625 11
B     9 3.316625 11
C     9 3.316625 11
D     9 3.316625 11

Variable: Y
      mean       sd  n
A 7.500909 2.031568 11
B 7.500909 2.031657 11
C 7.500000 2.030424 11
D 7.500909 2.030579 11
```

Linear regression…

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

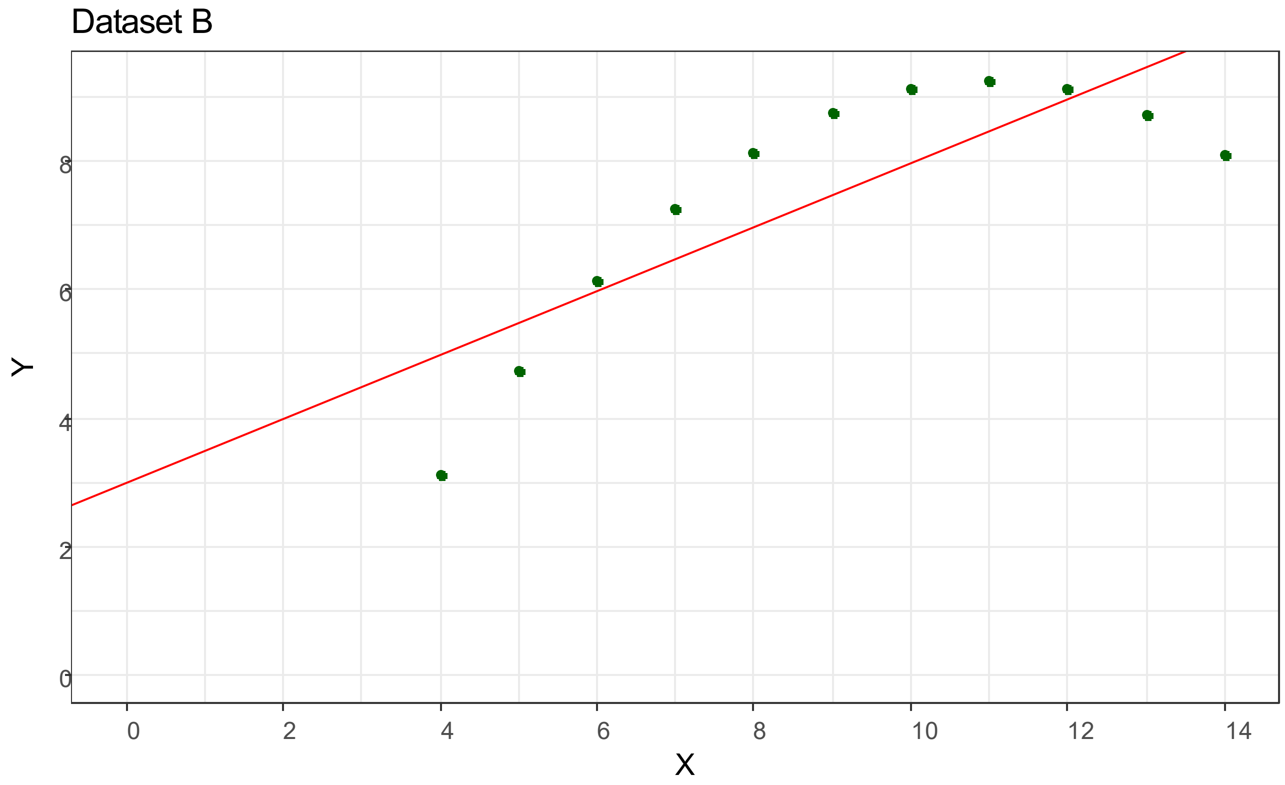
```
Call:
lm(formula = Y ~ X, data = data)

Coefficients:
             Estimate   s.e t value Pr(>|t|)
(Intercept)     3.00   1.12   2.667   0.025 *
X               0.50   0.11   4.239   0.002 **

Residual standard error: 1.237 on 9 degrees of
freedom
Multiple R-squared:  0.6662,
Adjusted R-squared:  0.6292
F-statistic: 17.97 on 1 and 9 DF,
p-value: 0.002179
```

Let’s compare the results with the data: Dataset A

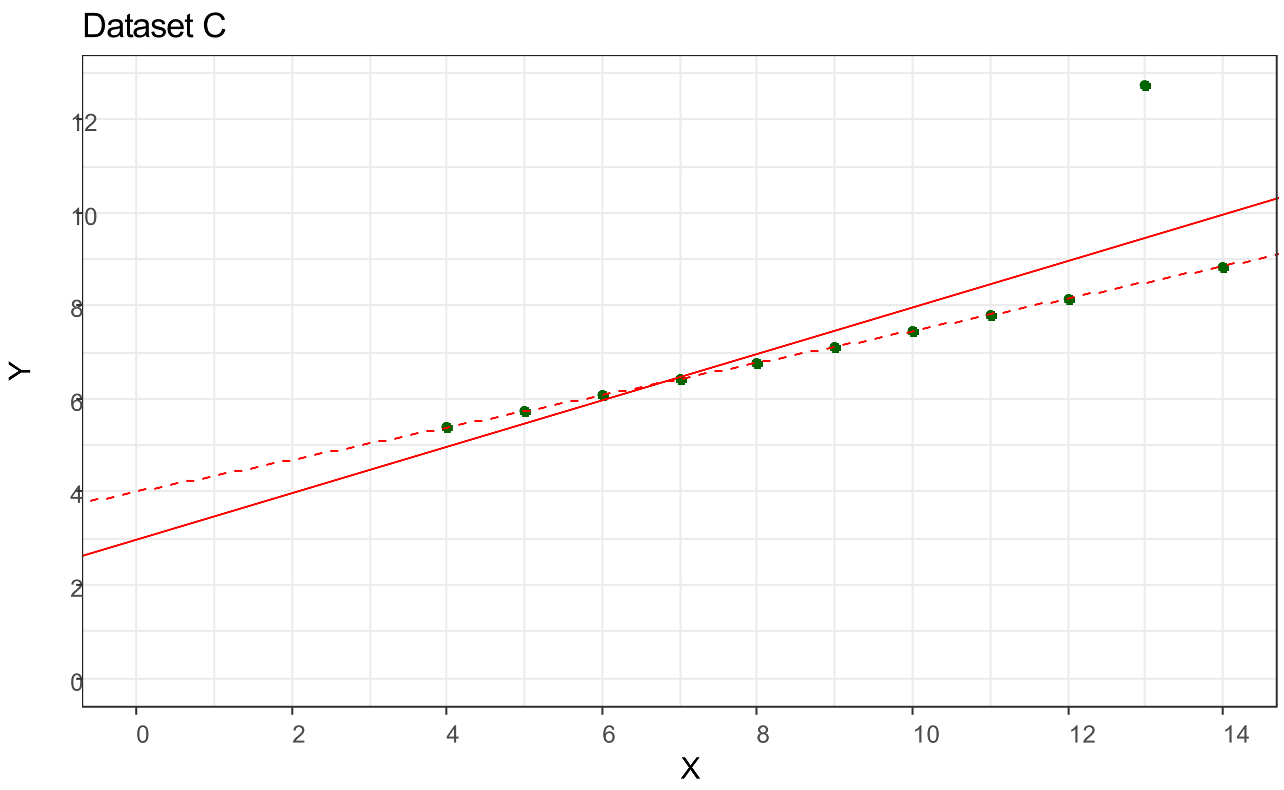


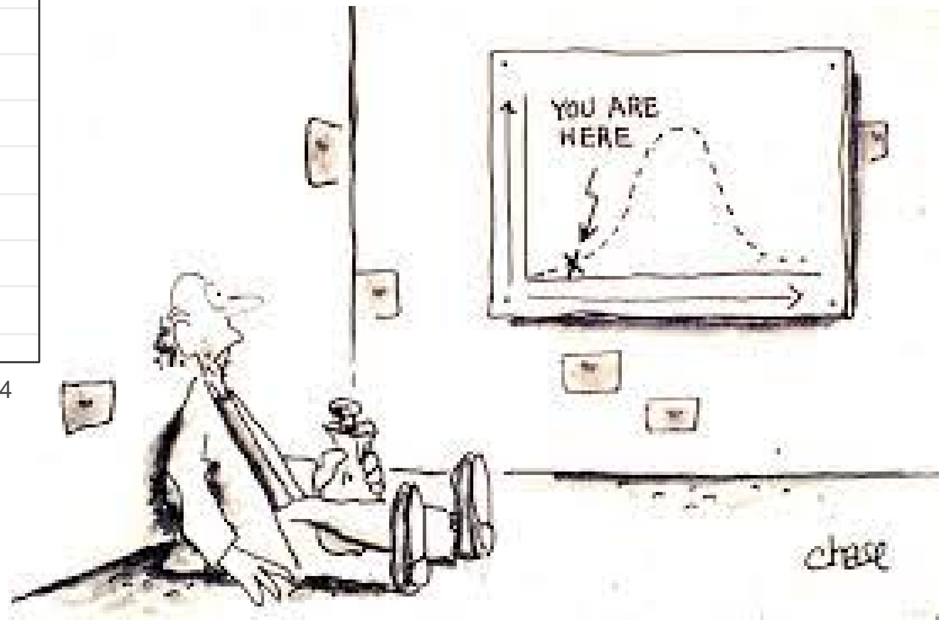

```
Call:
lm(formula = Y ~ X)

Coefficients:
              Estimate   s.e
(Intercept)     3.00    1.12
X               0.50    0.11
```

Let’s compare the results with the data: Dataset B

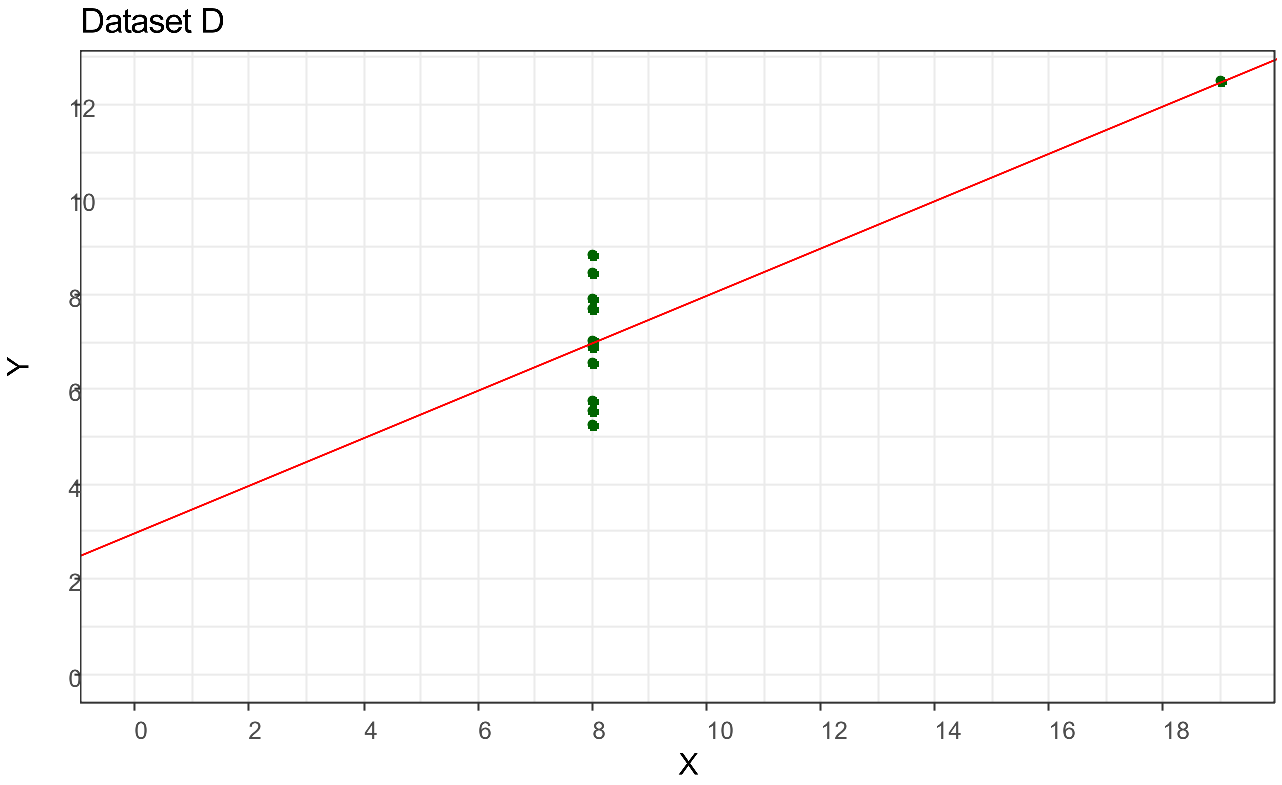


Let’s compare the results with the data: Dataset C

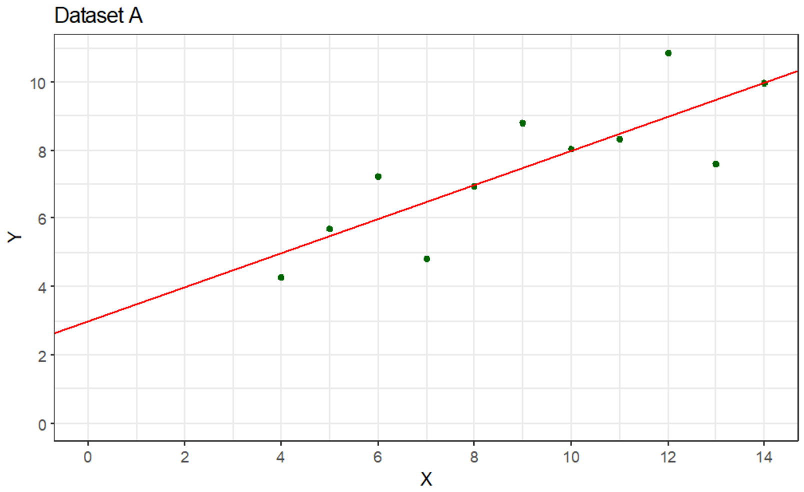


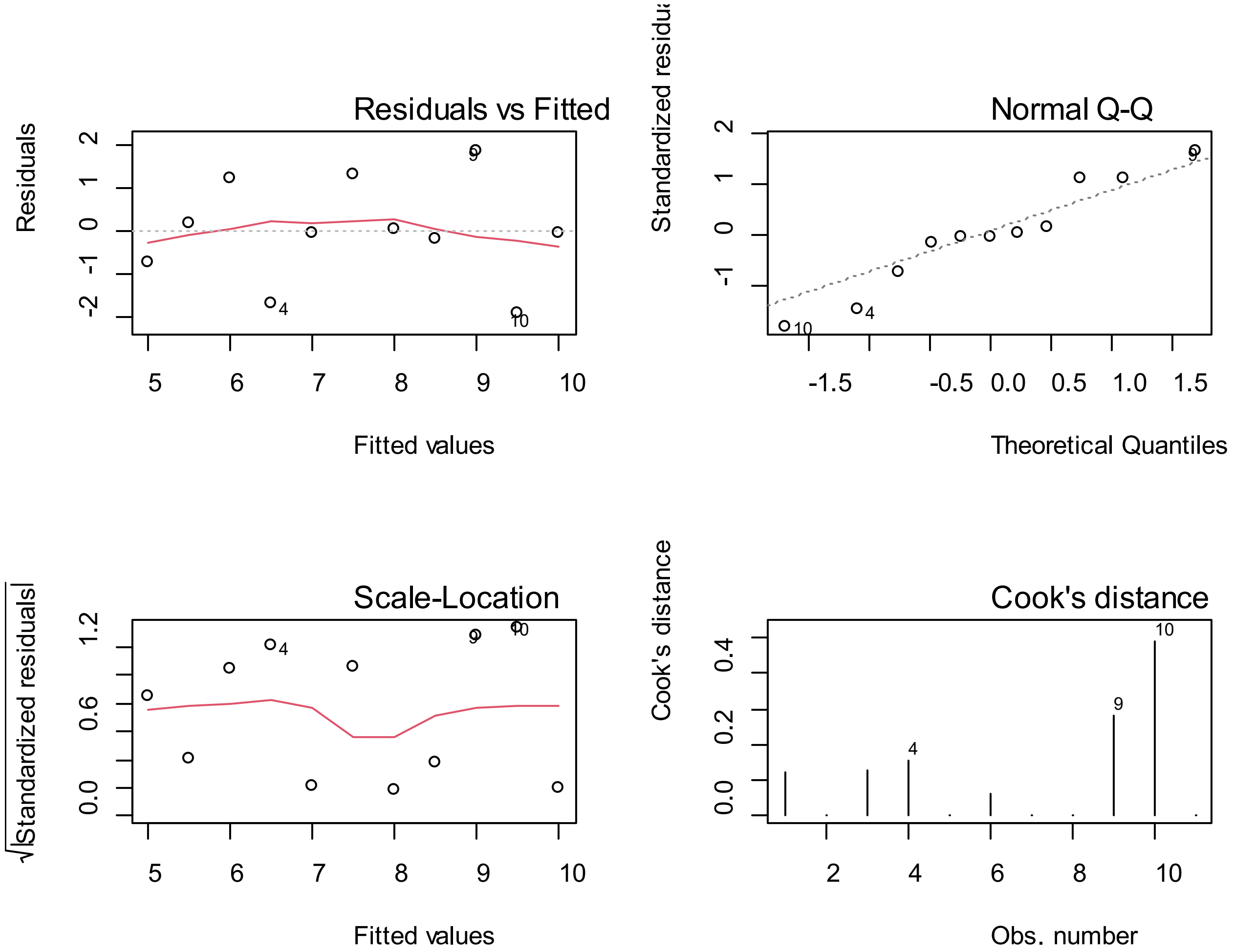


Let’s compare the results with the data: Dataset D

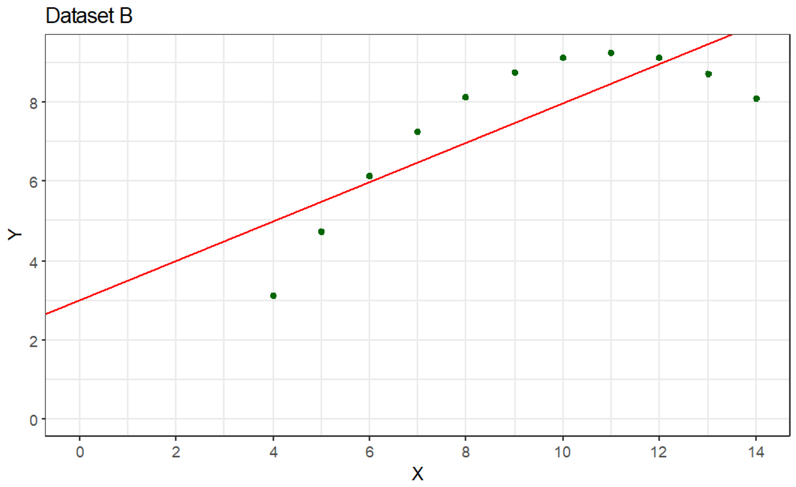


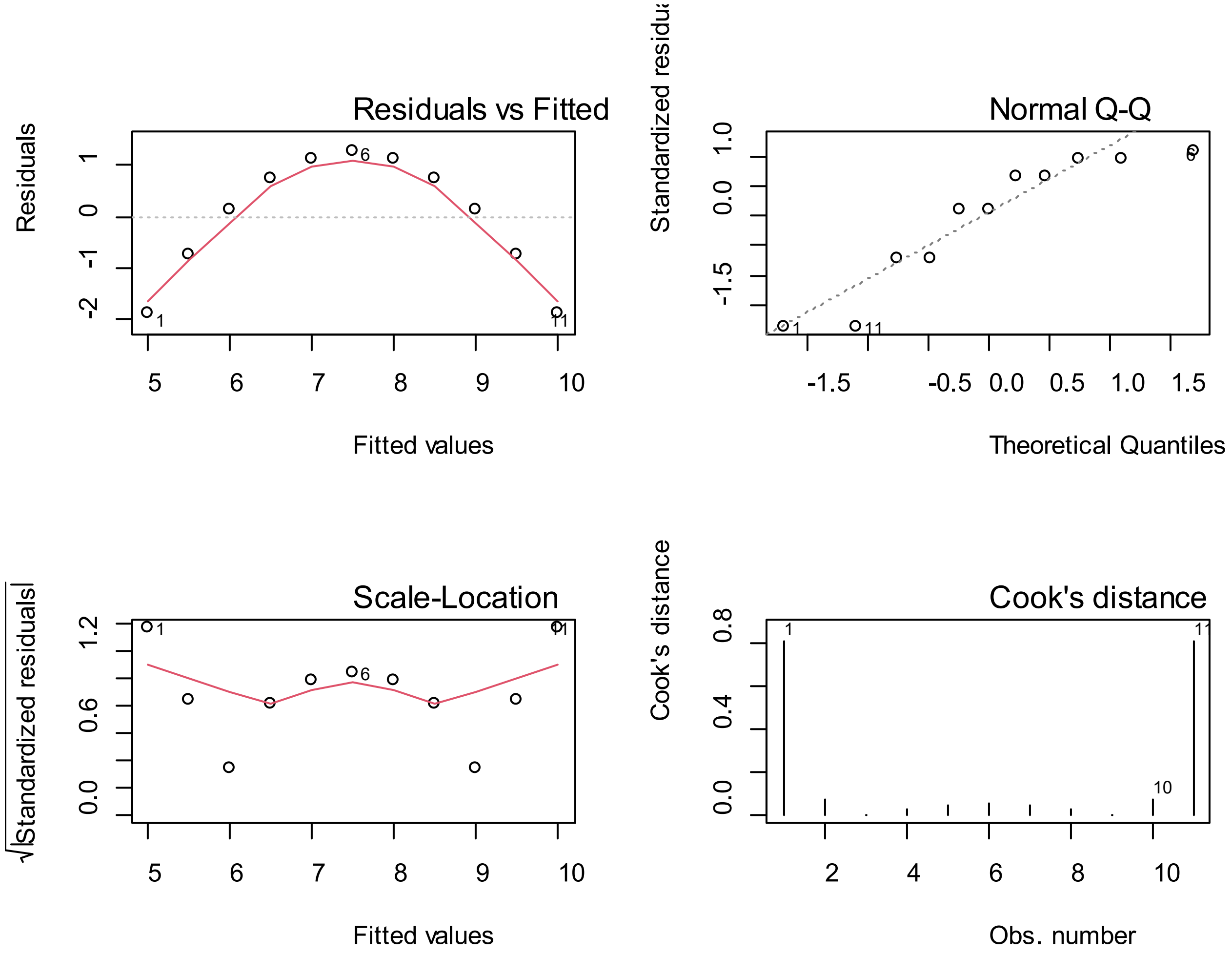
Let’s inspect diagnostic plots: Dataset A



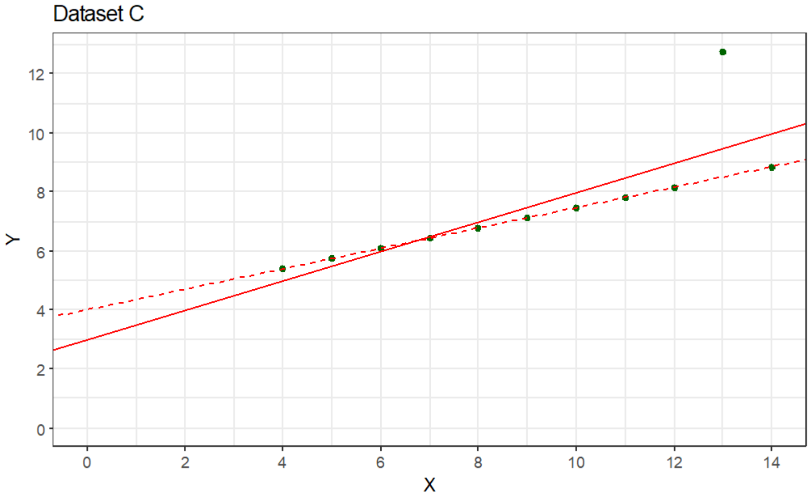


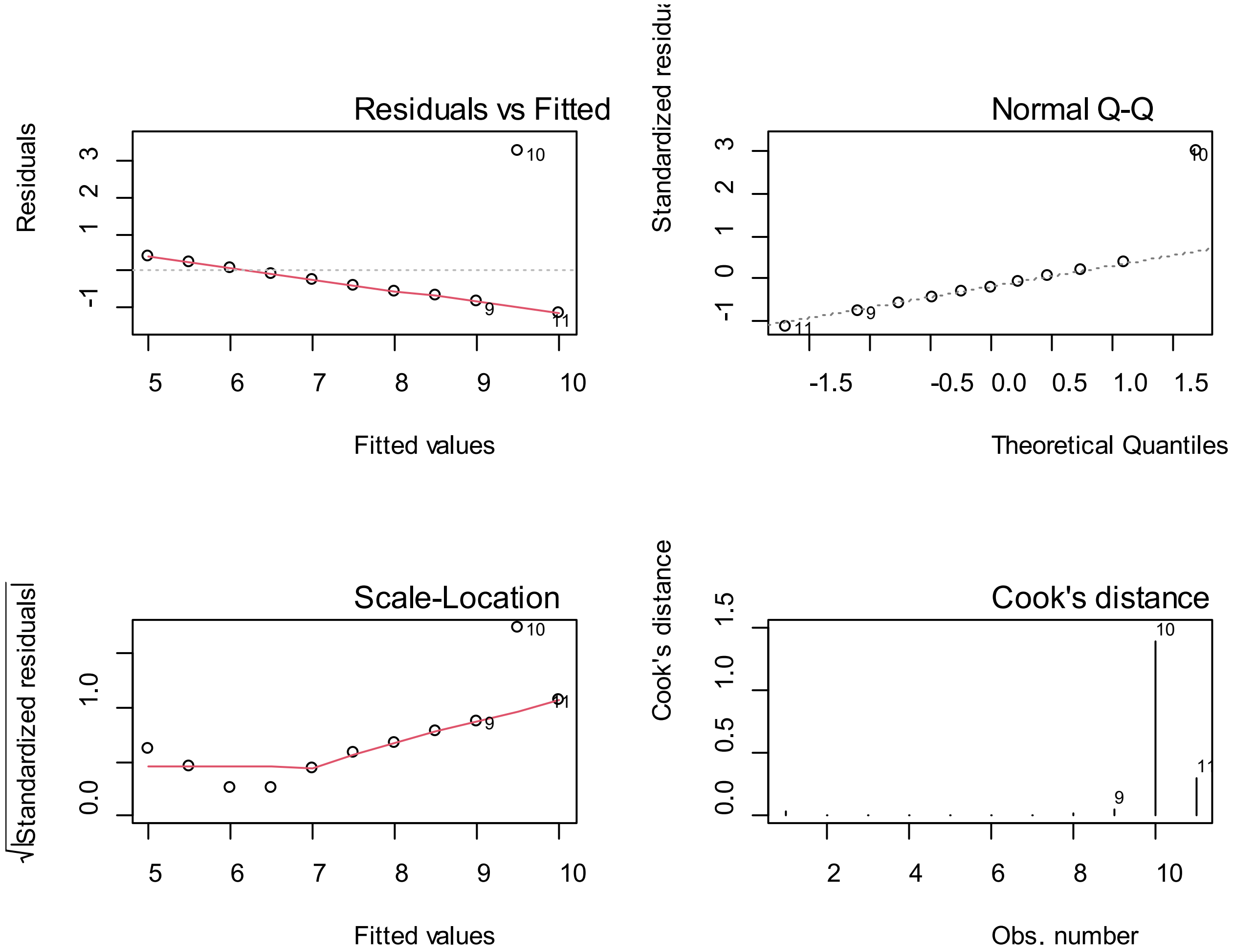
Let's inspect diagnostic plots: Dataset B



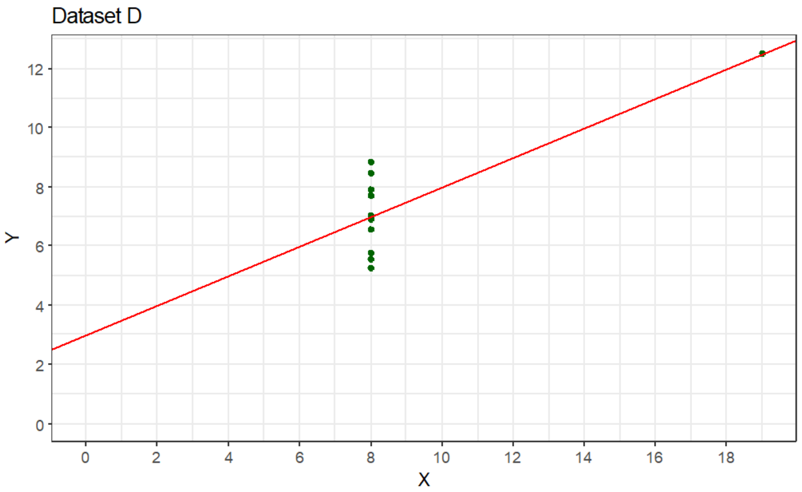


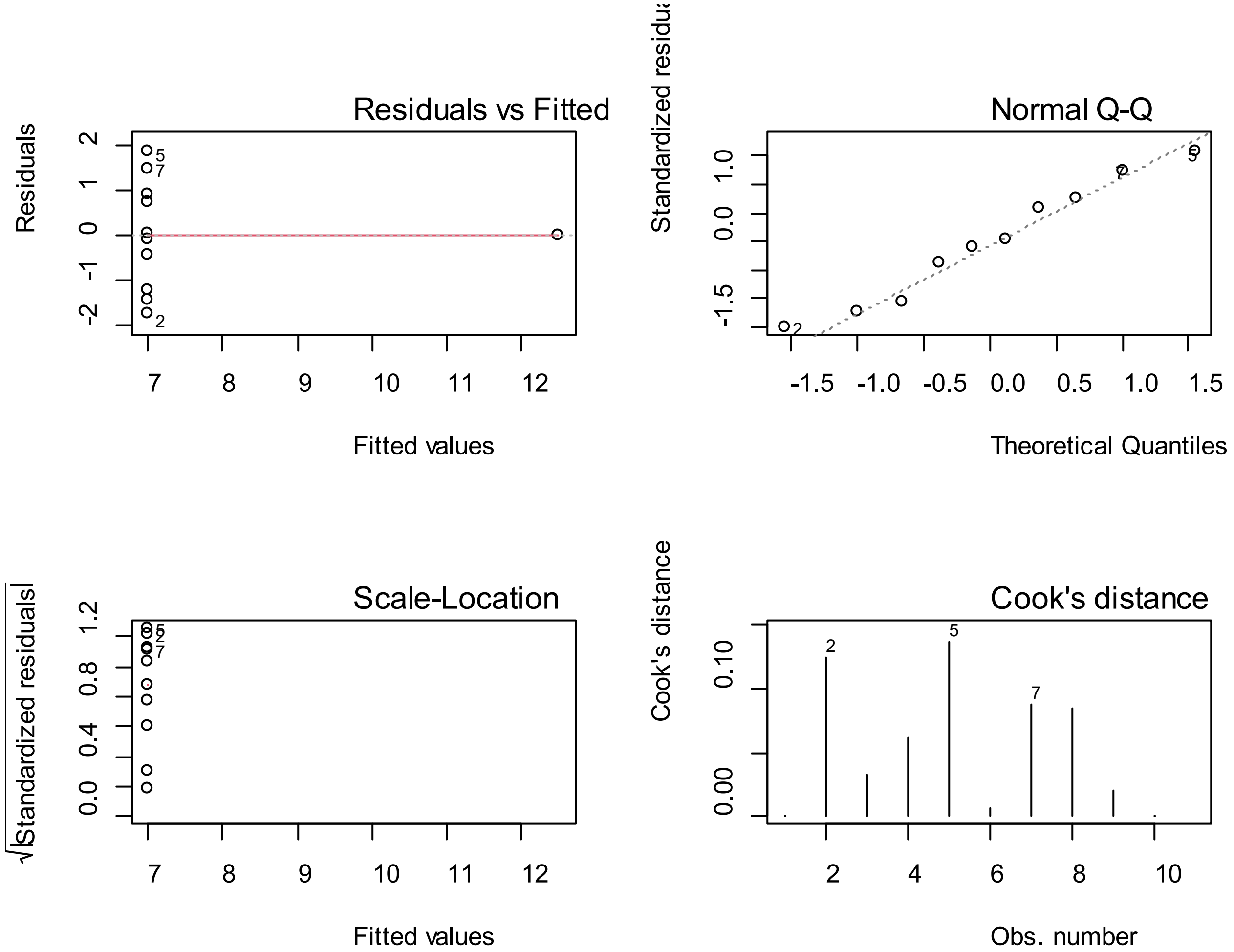
Let’s inspect diagnostic plots: Dataset C





Let's inspect diagnostic plots: Dataset D





Graphs: Conclusion

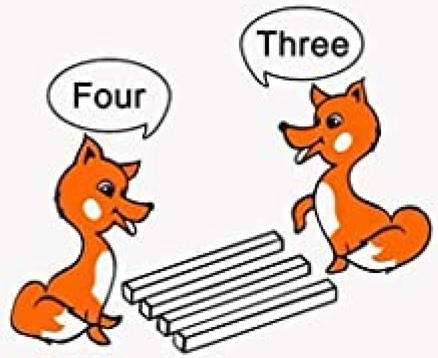
- Always create graphs to see if your model describes the data.
- A model can give a perfect fit but still be non-sense.
- Using diagnostic graphs can help you detect inappropriate models

Papillae Piglets (fictitious example)

- Skin tone is determined by the number of papillae:
  - Determination: Manually count on histological slice
  - New technique: Image analysis via machine-learning

  - Objective research: Compare the two techniques,

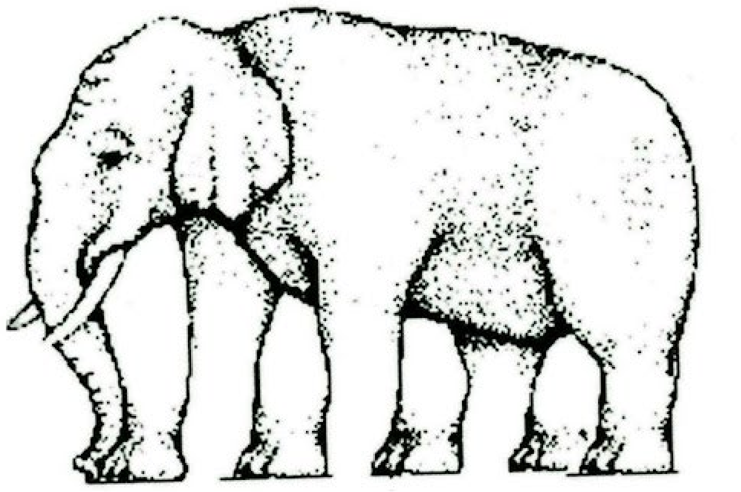
  can the new technique replace manual counts?



ILVO

Papillae Piglets: design 1

- Experimental design:
  - 5 histological slices: each manual count and AI count
- Analysis:
  - Paired t-test
- Results:
  - P-value: 0.20 -> no significant difference
- Conclusion:
  - The two techniques are the same!



ILVO

Papillae Piglets: design 2

- Experimental design:
  - 14 histological slices: each manual count and AI count
- Analysis:
  - Paired t-test
- Results:
  - P-value: 0.0453 -> significant difference
- Conclusion:
  - The two techniques are different!

Papillae piglets: overall conclusion?

- Where is the reasoning error??
  - The research question was actually whether the two techniques are the same…
  - But: statistical test for difference
  - No significant difference does not mean that they are equal!

ILVO

Papillae Piglest: equivalence test?

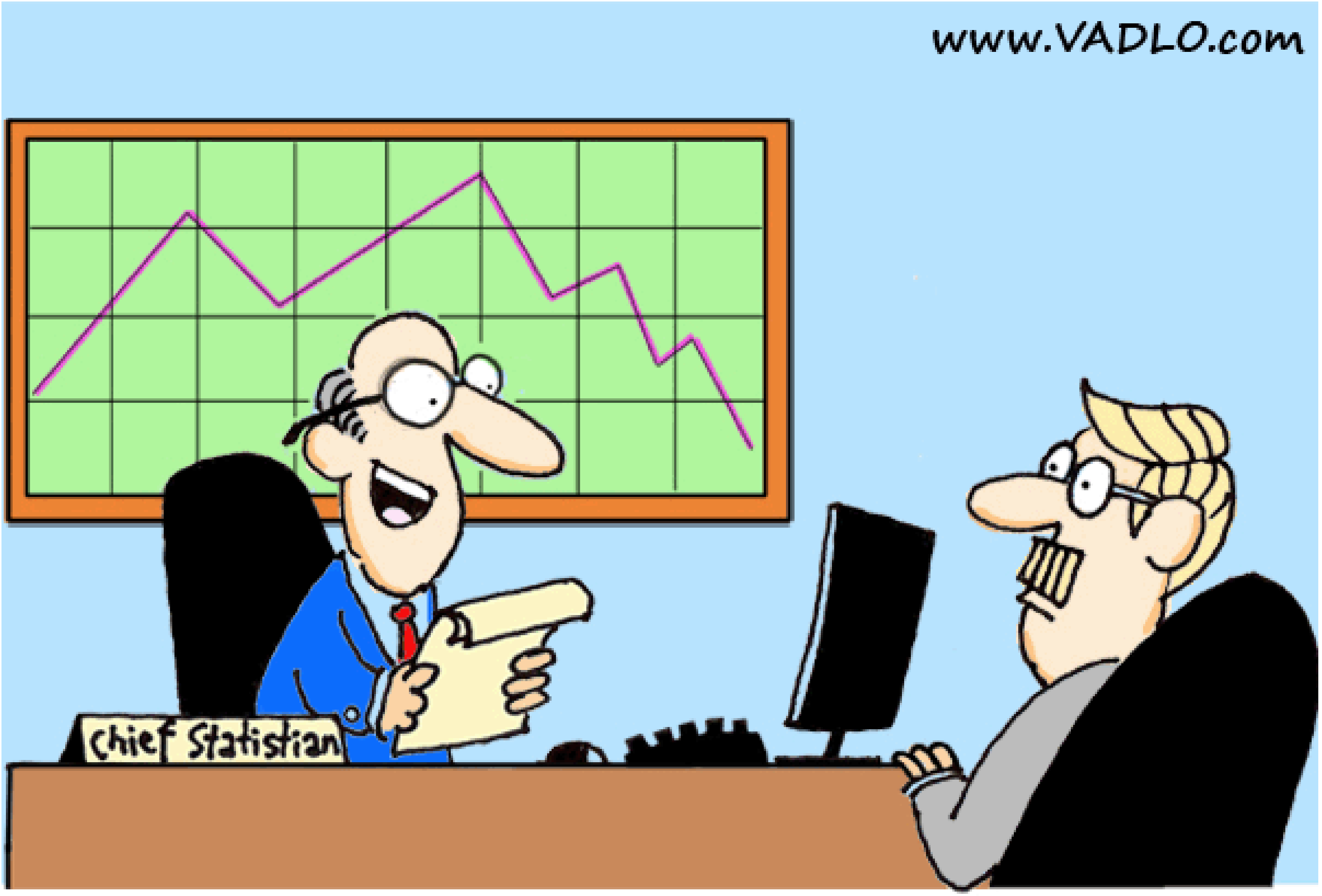
- Solution: equivalence testing!
  - First determine equivalence limits: In this case, experts determined that this is 100 (on an average of /- 1000)
  - Determine if CI for the difference is within the equivalence limits

Papillae Piglets: output R

```
Paired t-test data:
Mar_Mir_pair$GemA.A[1:5] and Mar_Mir_pair$GemA.B[1:5]
t = 1.5406, df = 4, p-value = 0.1983
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-61.2744 214.0424
sample estimates: mean of the differences
76.384
```

```
Paired t-test data:
Mar_Mir_pair$GemA.A and Mar_Mir_pair$GemA.B
t = 2.2142, df = 13, p-value = 0.0453
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
1.135941 92.304059
sample estimates: mean of the differences
46.72
```

ILVO



"I can prove it or disprove it! What do you want me to do?"

ILVO

Papillae Piglets: equivalence test

- Design: 5 histological slices:
  - 95% CI for the difference:

    [-61.3 ; 214.0]
  - Conclusion: the two techniques may not be considered equivalent
- Design:14 histological slices:
  - 95% CI for the difference:

    [1.1 ; 92.3]
  - Conclusion: both techniques are considered to be equivalent (difference is less than 100).

ILVO

Papillae piglets: equivalence

- Design:14 histological slices
- Analyse: equivalence test
- Results:
  - Manual: 1059 (sd 79)
  - Machine learning: 1012 (sd 67)
- 95% CI for the difference:

  [1.1 ; 92.3]
- Conclusion: both techniques are considered to be equivalent (difference is less than 100).

ILVO

Equivalence test: Conclusion

- Absence of a significant difference does not mean that two groups are equal (power may be too low)
- Equivalence can be tested with an equivalence test (usually many observations required) with predefined equivalence limits.
- Two groups can be significant different and yet equivalent (over-powered with an irrelevant difference)

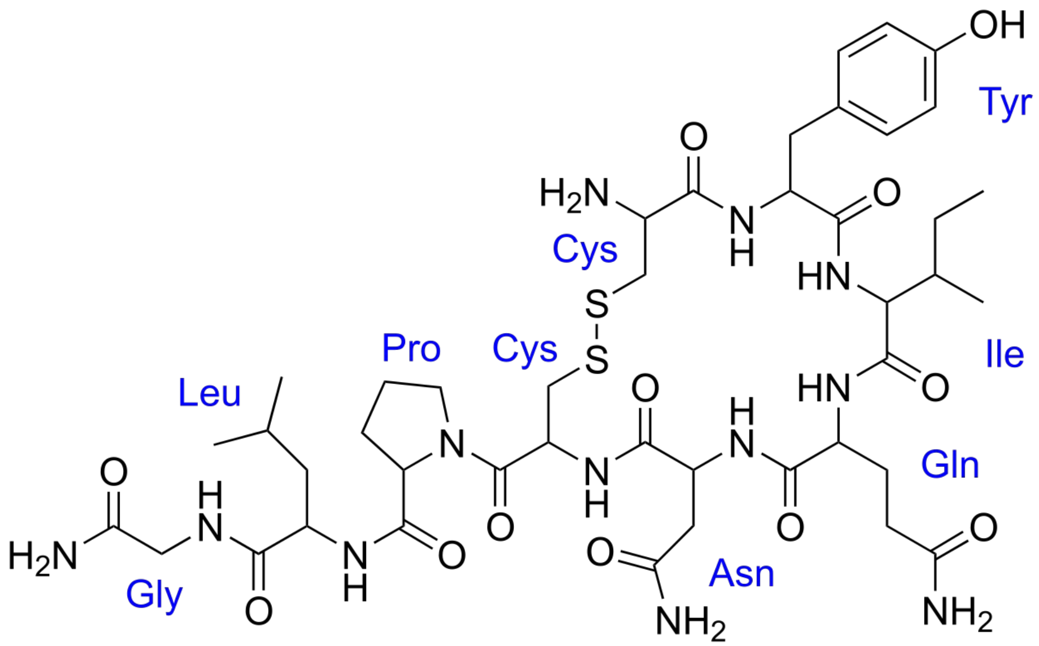
ILVO

The effect of experimental hormone OXY on the growth of piglets and the high risk of interactions with the environment (fictitious example)

a critical look at the misuse/misinterpretation of linear models with interactions

Hormone OXY

- Action in mammals:
  - Experimental growth-like hormone.
  - Does it work or not in piglets?

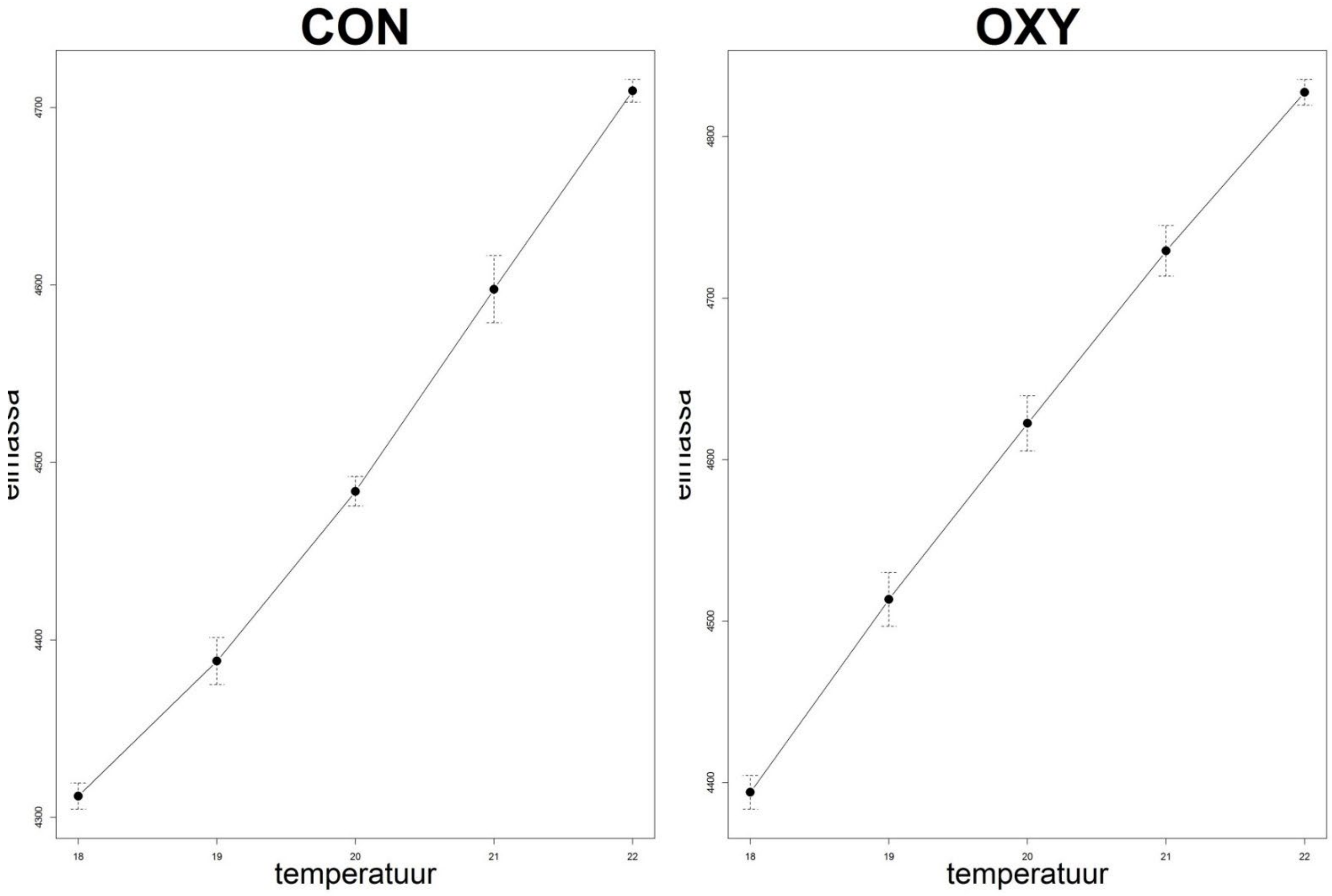


Hormone OXY

- Research question:
  - Has hormone OXY a positive effect on the growth of piglets
- Experimental design:
  - 40 piglet pairs ad random assigned to OXY or control group:
  - 5 environmental temperatures: 18, 19, 20, 21, 22°C
  - Separate housing and temperature control for each piglet pair
  - Total growth (g) during predefined period

Hormone OXY

- Results:



ILVO

Hormone OXY

- Results:
  - Temp significant ($p<0{,}001$)
  - Effect hormone OXY not significant ($p=0{,}75$)
  - Interaction Temp: OXY not significant ($p=0{,}18$)

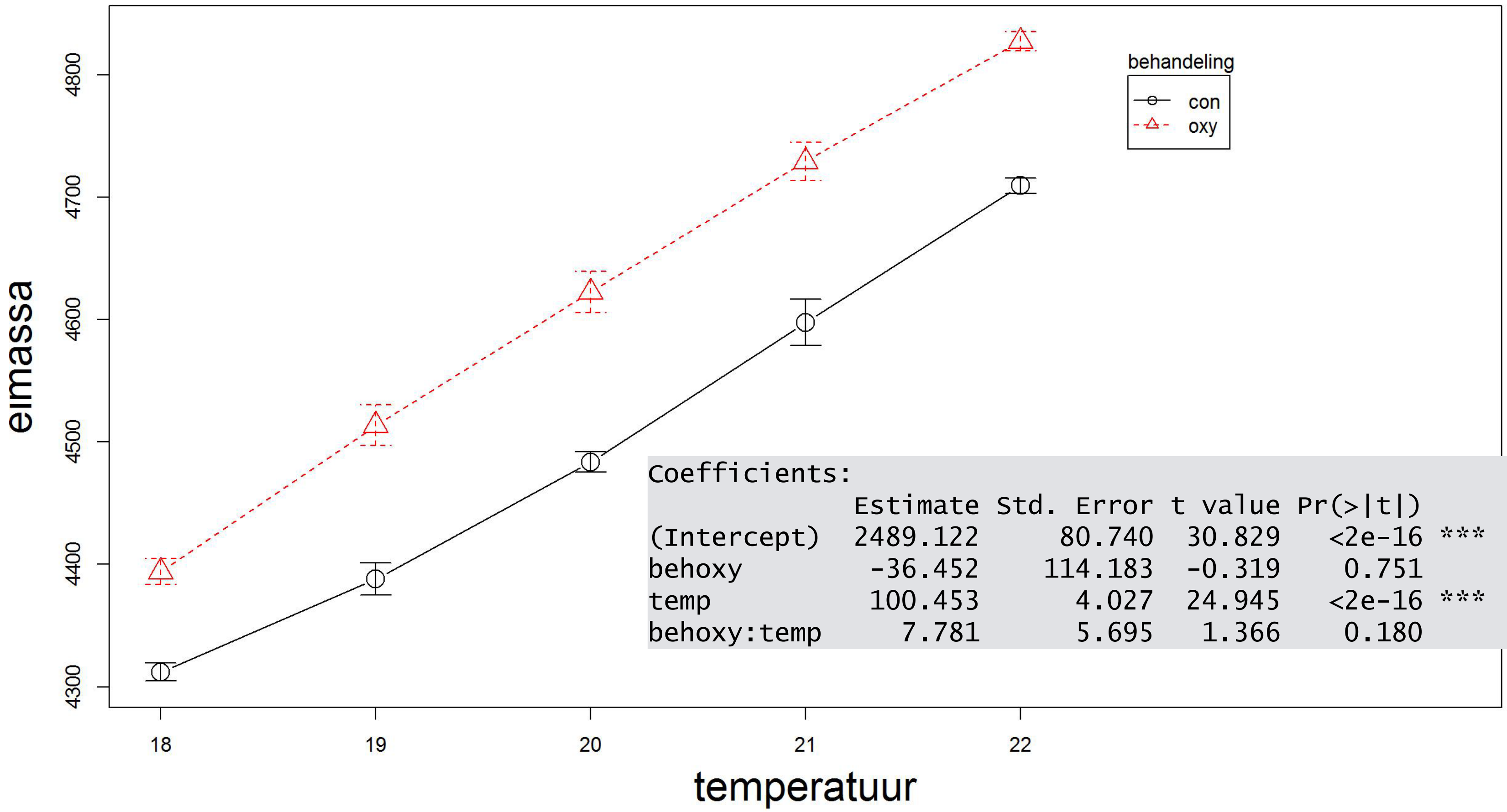
- Conclusion:
  - Hormone OXY does not improve the growth of piglets.

Critical look at the analyzes?

The conclusion was wrong!

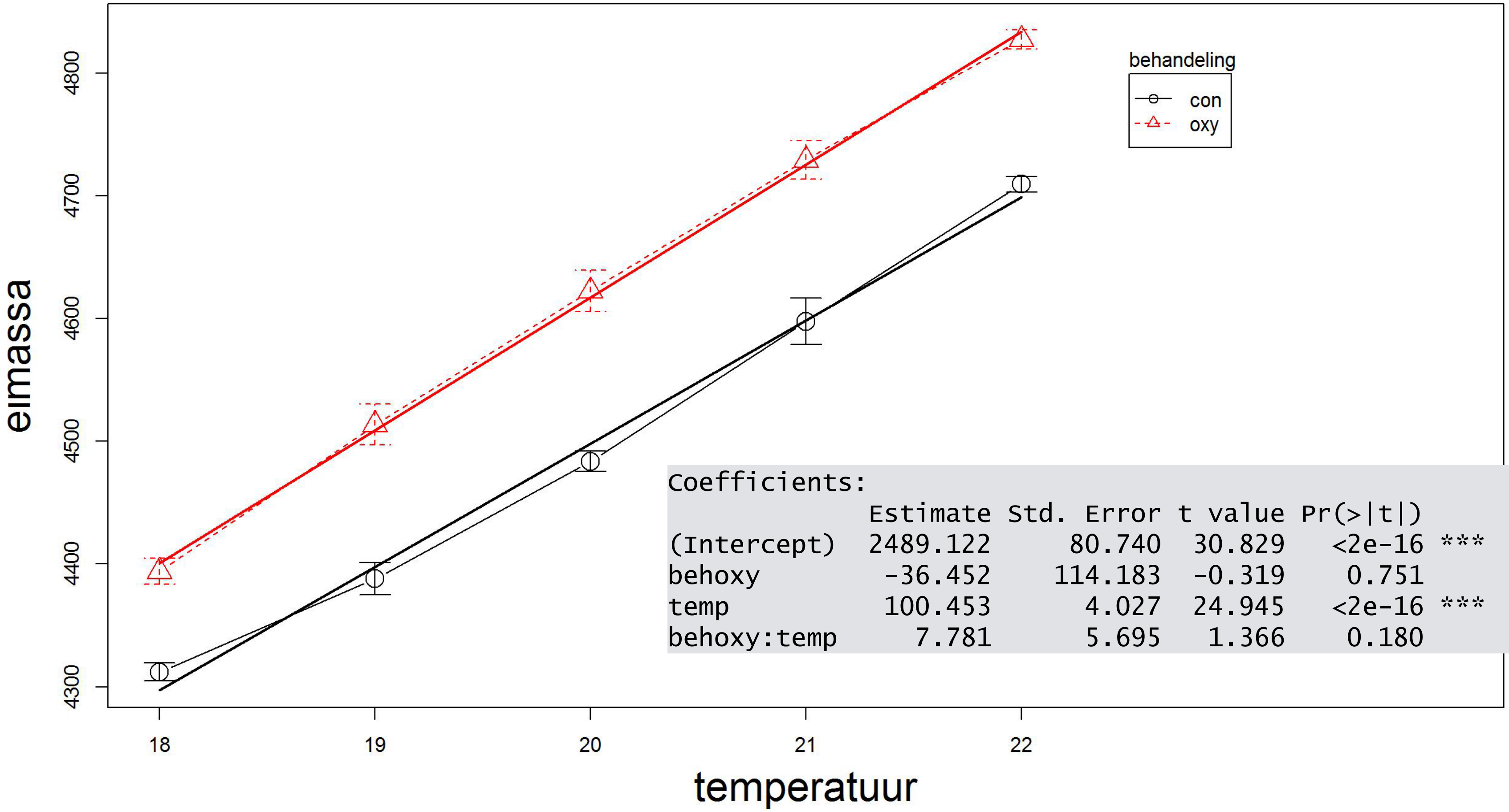
ILVO

Hormone OXY: graph

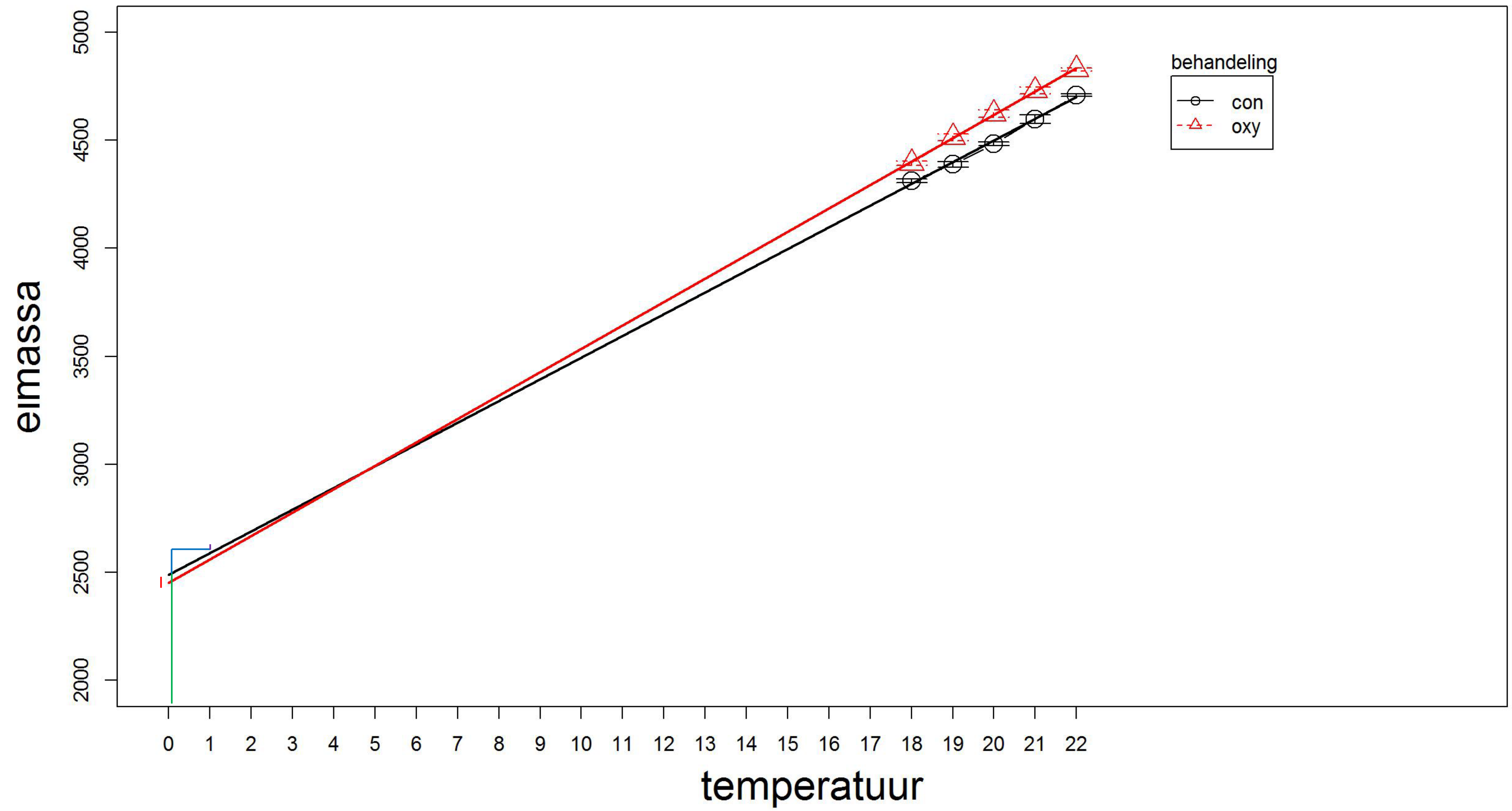



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2489.122     80.740  30.829   <2e-16 ***
behoxy        -36.452    114.183  -0.319    0.751    
temp          100.453      4.027  24.945   <2e-16 ***
behoxy:temp     7.781      5.695   1.366    0.180    
```

Hormone OXY: graph




```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2489.122     80.740  30.829   <2e-16 ***
behoxy         -36.452    114.183  -0.319    0.751    
temp           100.453      4.027  24.945   <2e-16 ***
behoxy:temp      7.781      5.695   1.366    0.180    
```

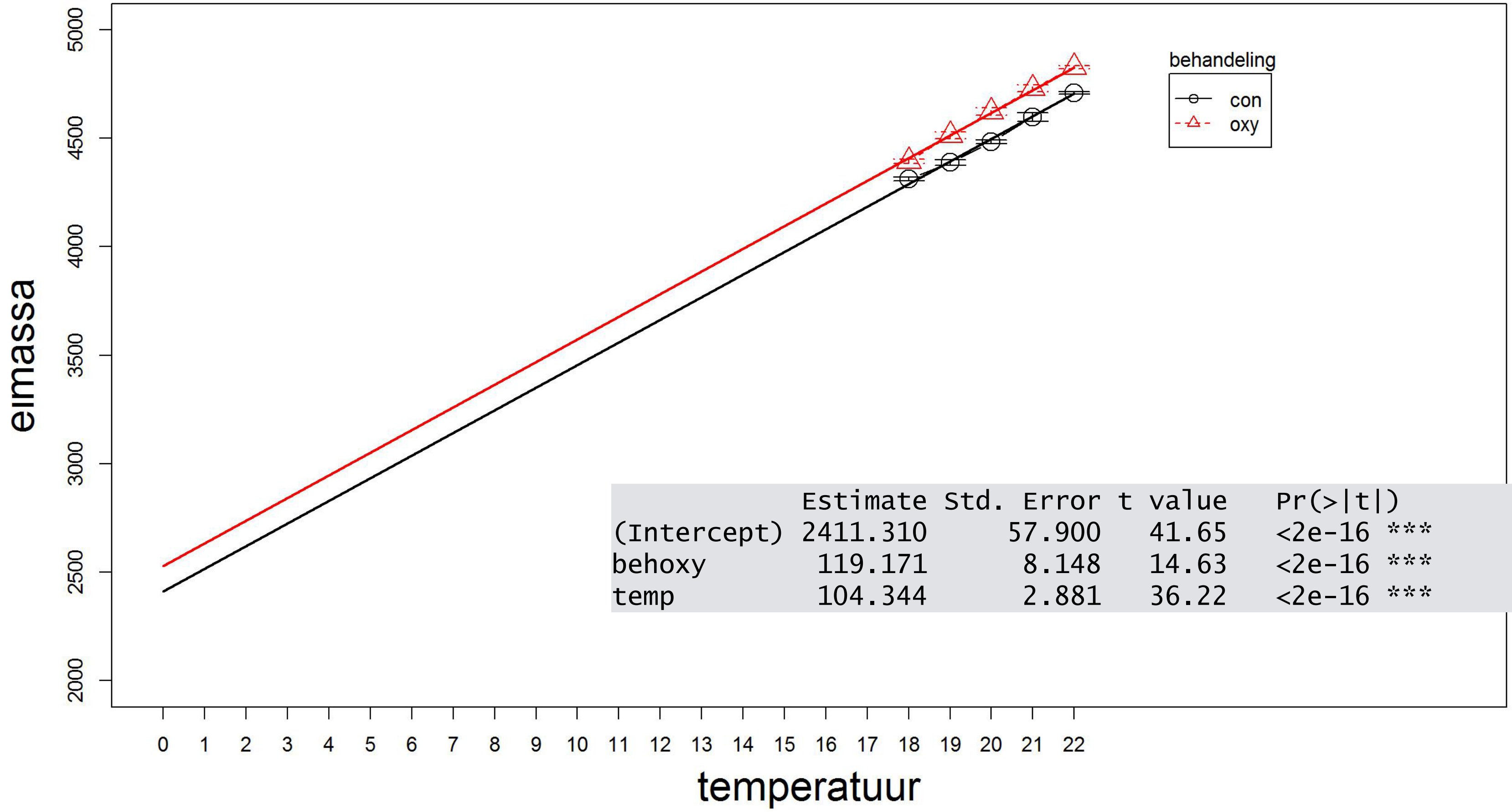



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2489.122     80.740  30.829   <2e-16 ***
behoxy        -36.452    114.183  -0.319    0.751    
temp          100.453      4.027  24.945   <2e-16 ***
behoxy:temp     7.781      5.695   1.366    0.180    
```

growth=Intercept+temp+treat+treat*temp

ILVO

OXY: only main effects (no interaction)



	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2411.310	57.900	41.65	<2e-16	***
behoxy	119.171	8.148	14.63	<2e-16	***
temp	104.344	2.881	36.22	<2e-16	***

Hormone OXY: only main effects

- Results:
  - Positive effect of temp ($p<0,001$)
  - Positive effect hormone OXY ($p<0,001$)

- Conclusion:
  - Hormone OXY has a positive effect on the growth of piglets!!!

Main conclusion:

- If an interaction in the model:

NEVER LOOK AT THE MAIN EFFECTS SEPARATELY

Correlation vs Causation

- Correlation: the amount of forest fires increases alongside an increase in people buying ice cream.

- Does this mean that people buying ice cream are causing the fires?

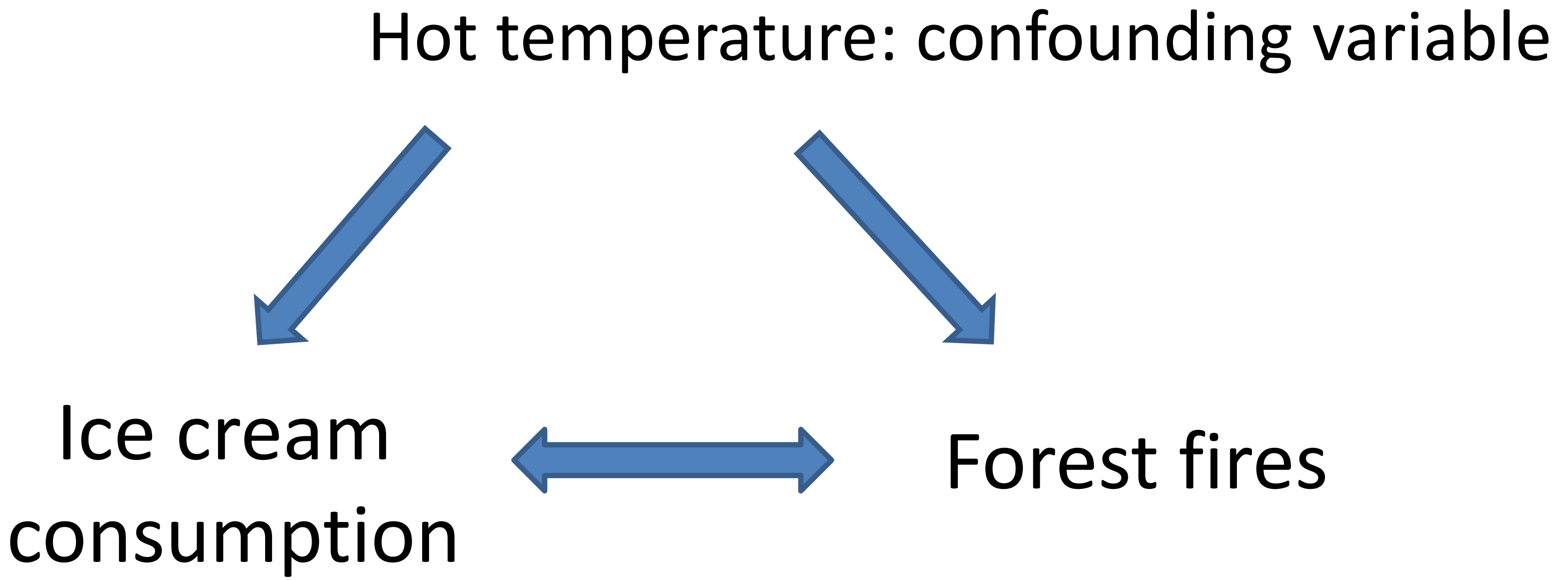
Observational vs Experimental

- Observational: the amount of forest fires increases alongside an increase in people buying ice cream. This is a correlation

- Experimental: force an increase in ice cream consumption in some regions and observe the number of fires compared to the control regions…

Confounding variable

- Correlation: the amount of forest fires increases alongside an increase in people buying ice cream.




Scientists tested a frog. They cut it's legs off and said, "Jump!" The frog didn't jump. Scientists therefore concluded that when frogs lose their legs they become deaf.
AMERICASBESTPICS.COM

Thanks to:

Everyone who did something wrong in the past!

ILVO

Questions?

Research Institute for
Agriculture, Fisheries and Food
Scheldeweg 68
9090 Melle – Belgium
T + 32 (0)9 272 26 00

dier@ilvo.vlaanderen.be
www.ilvo.vlaanderen.be

ILVO

Flanders
is agriculture and fisheries

ILVO