Standard Error

What is the **standard error**?

Typical size of an error that one would make when estimating something

Using a specific experimental design

With a specific measurement strategy

With a specific analysis strategy

In a specific population

Typical: If the experiment were repeated many times in exactly the same way

If Po-Kai and I **each** do the same experiment

Do our estimates have different standard error?

If I estimate +3 bpm and he estimates +6 bpm?	**NO**
If I calculate $s^2 = 12$ and he calculates $s^2 = 10$?	**NO**
If I measure for 30s and he measures for 60s?	**YES**
If I measure for 60 people and he measures for 61?	**YES**
If I do an **indirect** design and he does a **direct** design?	**YES**
If he goes to STA 120 and I use PLS205 for my subjects?	**YES**
If I estimate $\hat{\mu}_{SIT}$ and he estimates $\hat{\delta}$?	**YES**

Standard Error

What are the two formulas for the standard error?

Hint: It is the same for an estimate of a **mean** or a **treatment effect**

$$\sigma_r(\hat{\delta}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

Direct:

$$\sigma_r(\hat{\mu}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

Indirect: Sqrt of: Sum of the (Standard Error)^2's of each direct estimate

Effect = Difference between two means: $\sigma_r(\hat{\delta}) = \sqrt{\sigma_r^2(\hat{\mu}_B) + \sigma_r^2(\hat{\mu}_A)}$

Mean = Average of multiple means

Standard Error

What is the formula for **estimating** the standard error from data?

Hint: The calculation has 2 steps and uses sample size(s) twice

1) Use your replicates to calculate s^2

2) Plug s^2 into formulas for standard error

Direct: 1) Calculate s_δ^2 from replicates: $\hat{\delta}_i$

2) SE = $\sqrt{\frac{s_\delta^2}{n}}$

Indirect: 1) Calculate s_{pooled}^2 from replicates: $\hat{\mu}_{ij}$

2) SE = $\sqrt{\frac{s_{pooled}^2}{n_1} + \frac{s_{pooled}^2}{n_2}}$

Sample size is used:

1) In **denominator** of s^2 “Degrees of Freedom”

2) In **denominator** SE calculation # of replicates

Source	Df	SumSq	MeanSq	F-value	p-value
Treatment	dfT	SST	$\frac{SST}{dfT}$ = MST	F = $\frac{MST}{MSE}$	p
Error	dfE	SSE	$\frac{SSE}{dfE}$ = MSE		

Keys to the ANOVA

F-value: Are the treatment mean estimates more variable than we expect?

df1 and **df2**: Control the “we expect” above

Higher df1 => more accurate estimate of treatment variance

Higher df2 => more accurate estimate of EU variance

Both lead to higher power

MSE: s^2_{pooled}, basis of all confidence intervals of specific treatment effects

Outcome of the ANOVA

p-value: Answer to question: Are any means different?

By itself, it does not tell you **which treatments** are interesting

We still need to estimate **contrasts, or treatment effects** among pairs of treatments

Dunnett, Tukey

Just a starting point: Is it worth doing the work of reporting contrasts?

ANOVA tables in R

Source	Df	SumSq	MeanSq	F-value	p-value
Treatment	dfT	SST	$\frac{SST}{dfT}$ = MST	F = $\frac{MST}{MSE}$	p
Error	dfE	SSE	$\frac{SSE}{dfE}$ = MSE		

lm() model:

R Code | Start Over

```
anova_table = anova(yield_model)
print(anova_table)

```

```
Analysis of Variance Table

Response: Yield
           Df Sum Sq Mean Sq F value    Pr(>F)
Nitrogen    4  92042 23010.6  13.868 1.901e-07 ***
Residuals 45  74664  1659.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

lmer() model:

```r
anova(germinants_model,ddf = 'Kenward-Roger')
```

```
Type III Analysis of Variance Table with Kenward-Roger's method
     Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Days 4439.8  887.95     5    24  37.678 1.332e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What can go wrong?

~~Estimates are biased~~

CI too small

CI too big

Assumptions for calculating
Confidence Intervals

1) EU are independent

Count n for SE and *df*

2) μ_{ij} and ϵ_{ij} are Normally distributed

T, F, Dunnett, Tukey distributions

Confidence Intervals and p-values

3) $\sigma^2_{\mu_i}$ and σ^2_m are the same across groups

Pooling deviations to calculate
s^2_{pooled}, maximizing *df*

What can go wrong?

~~Estimates are biased~~

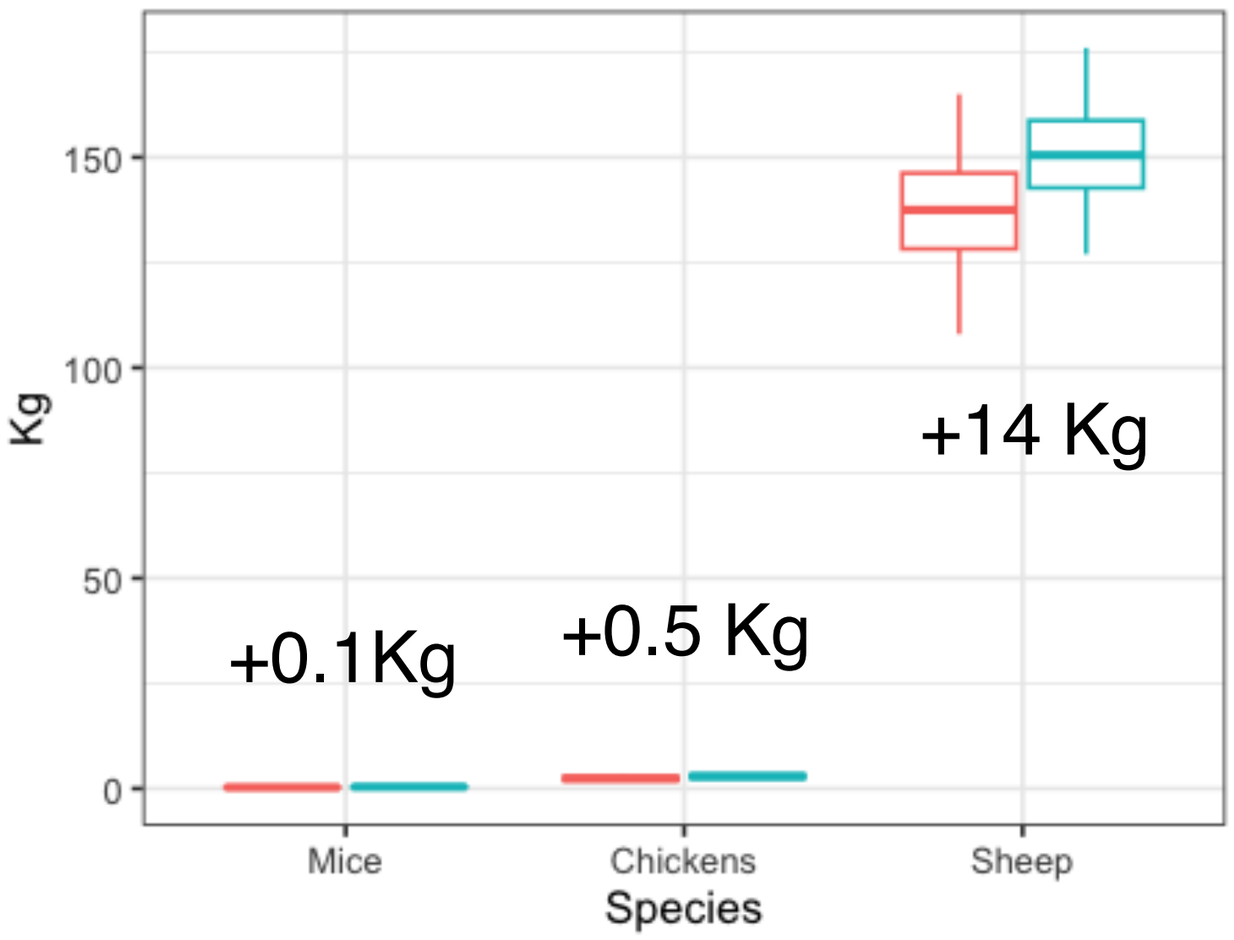
CI too small

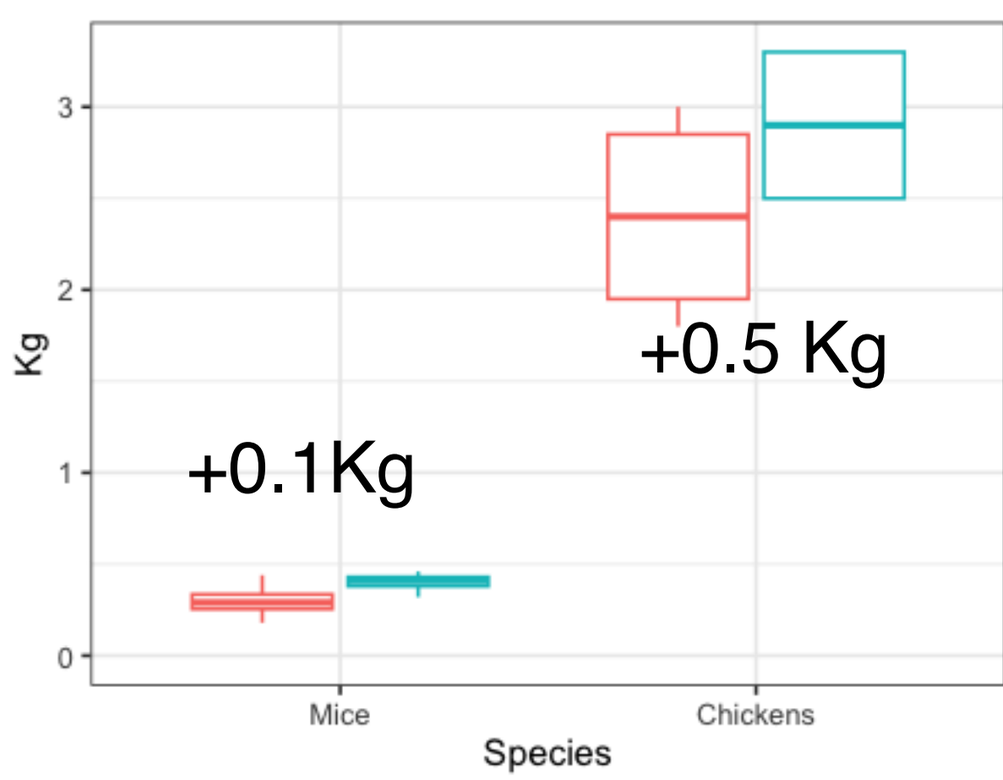
CI too big

Assumptions for calculating Confidence Intervals	Solutions
1) EU are independent	Randomization
Count n for SEM and *df*	Declare EUs (lmer)
2) μ_{ij} and ϵ_{ij} are Normally distributed	Check with QQplot
T, F, Dunnett, Tukey distributions	Use a different model
Confidence Intervals and p-values	glm(), glmer(), brm()
3) $\sigma^2_{\mu_i}$ and σ^2_m are the same across groups	Data transformations
Pooling deviations to calculate s^2_{pooled}, maximizing *df*	

An experiment tested 2 Feed treatments on 3 animal species:

Final Weight (Kg) was measured on each of 4 animals per feed per species



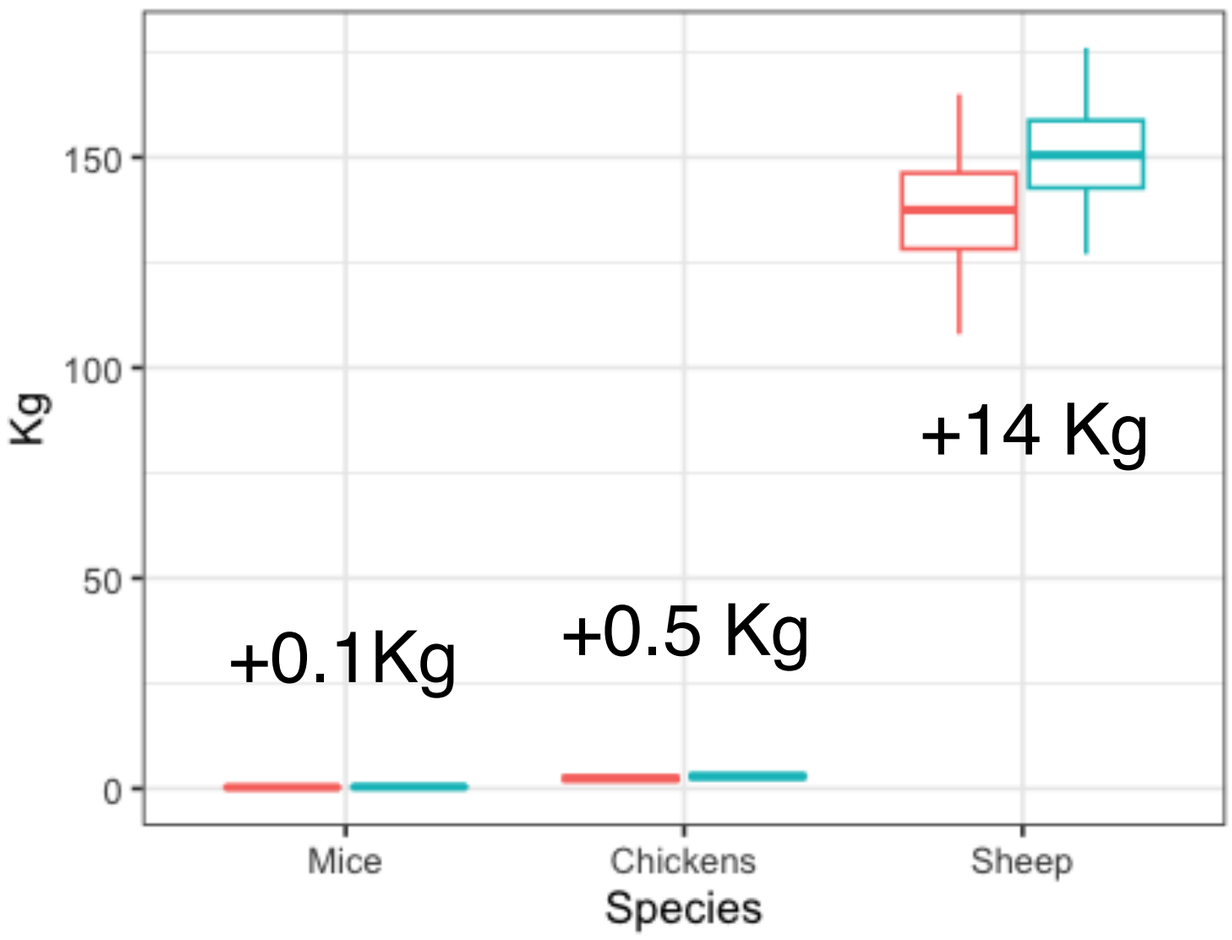


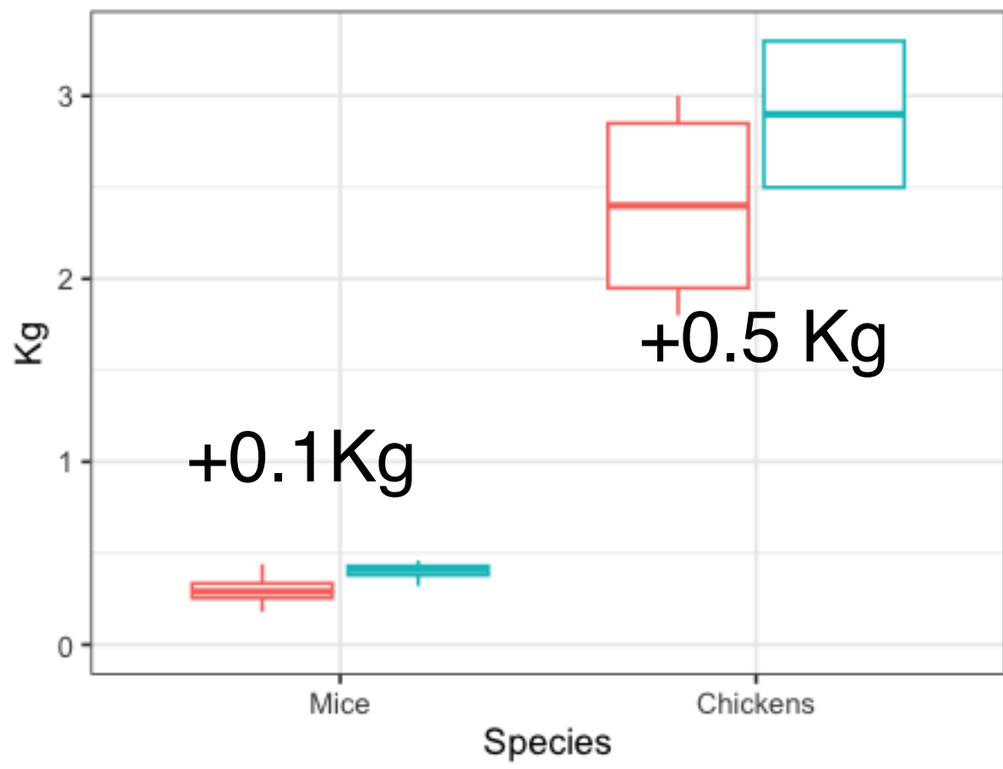
Do you think we can confidently say that the Vitamin treatment increased final weight?

```
Species = Mice:
 contrast           estimate  SE df t.ratio p.value
 Vitamin - Control       0.1 8.9 18   0.011  0.9912

Species = Chickens:
 contrast           estimate  SE df t.ratio p.value
 Vitamin - Control       0.5 8.9 18   0.056  0.9558

Species = Sheep:
 contrast           estimate  SE df t.ratio p.value
 Vitamin - Control      14.0 8.9 18   1.572  0.1333
```

An experiment tested 2 Feed treatments on 3 animal species:

Final Weight (Kg) was measured on each of 4 animals per feed per species





Problem: s^2_{pooled} is bad for each comparison

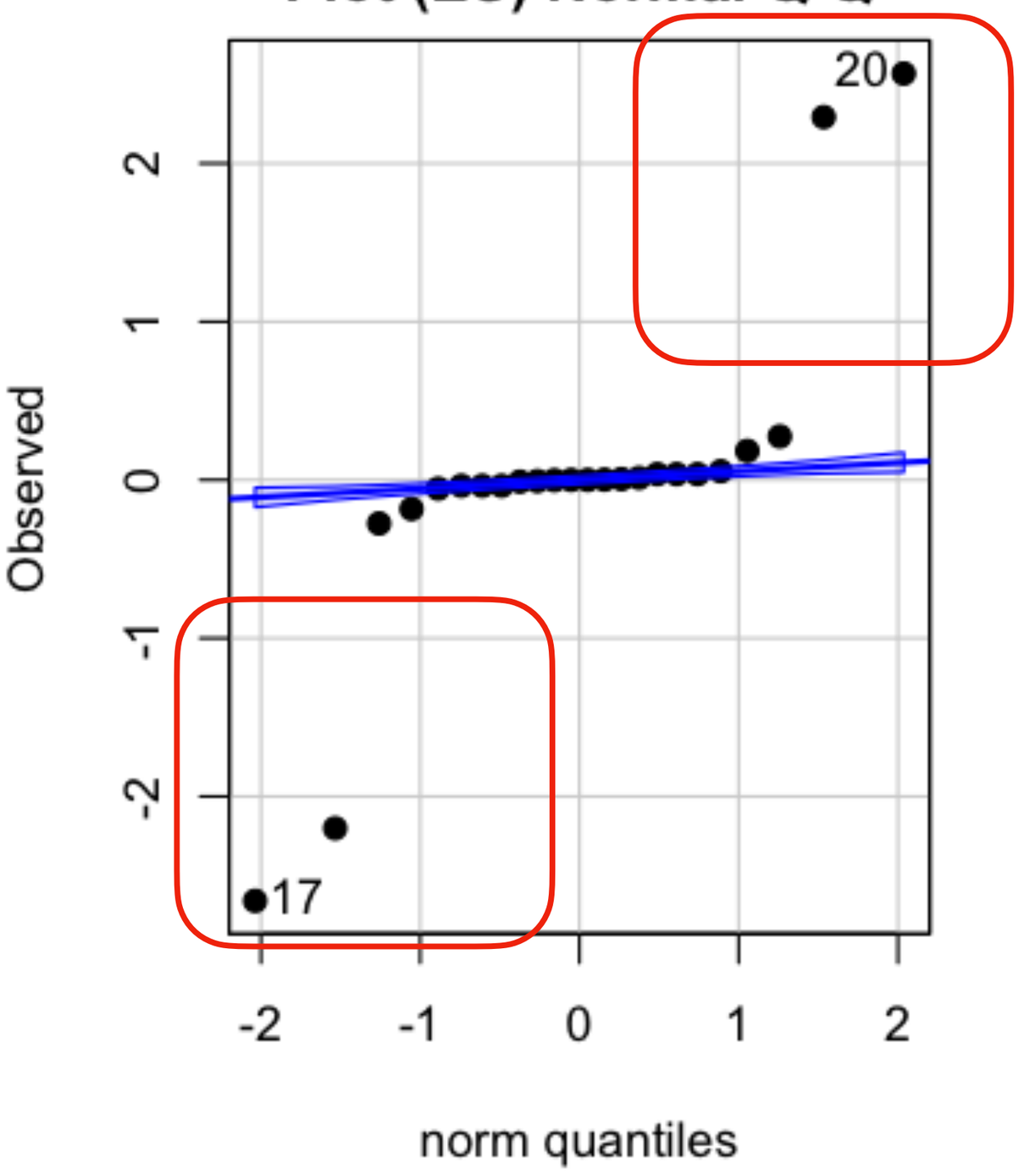
$$\text{SE (Mouse)} = \sqrt{\frac{s^2_{pooled}}{4} + \frac{s^2_{pooled}}{4}}$$

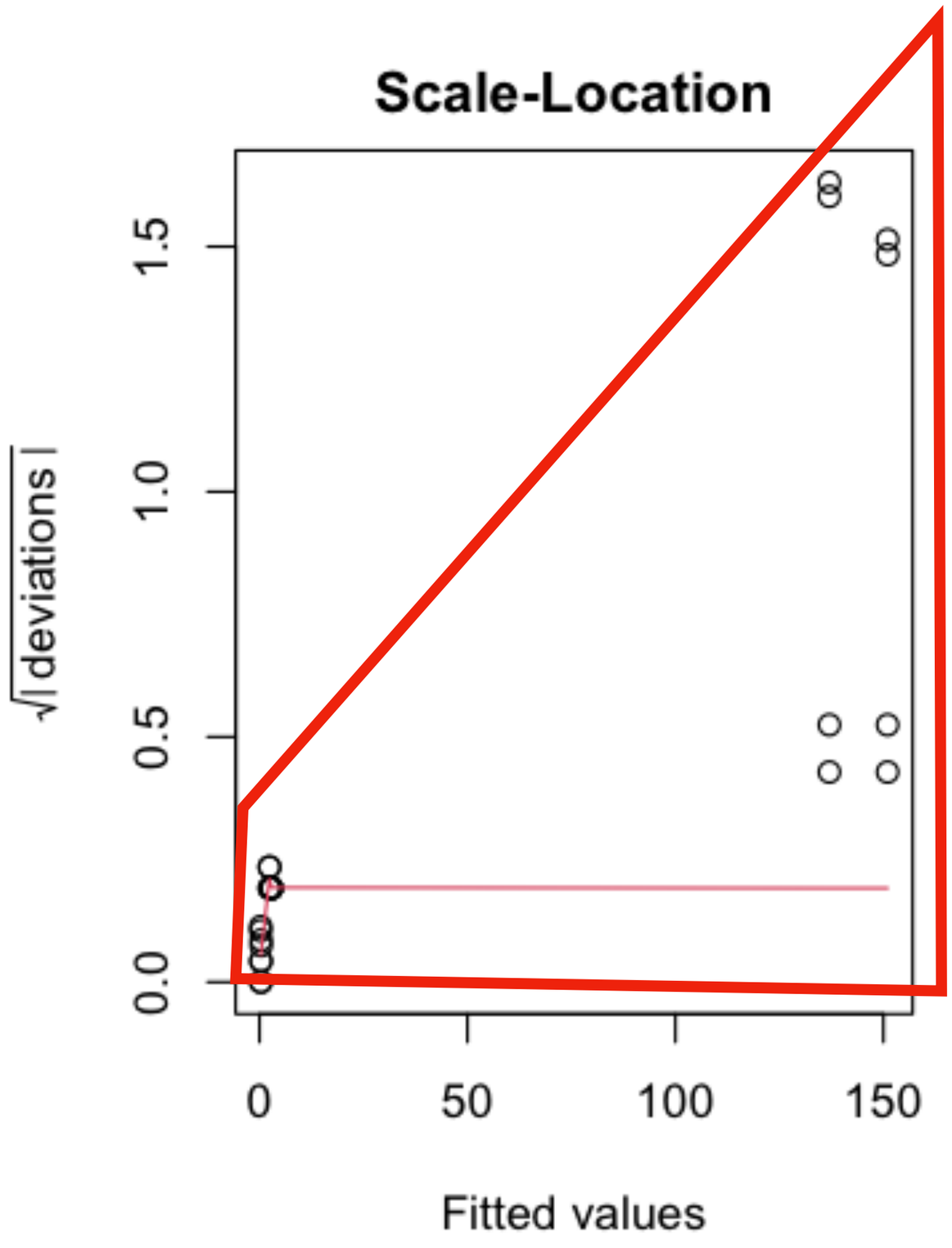
too big

s^2_{pooled} (average s_i^2)

$$\text{SE (Sheep)} = \sqrt{\frac{s^2_{pooled}}{4} + \frac{s^2_{pooled}}{4}}$$

too small





Treatment mean estimate

“residuals” in top-right and bottom-left means too many outliers

Triangular shape:

Big residuals (y) for big treatments (x)

“residuals” = “errors” = deviations of $\hat{\mu}_{ij}$ from $\hat{\mu}_i$

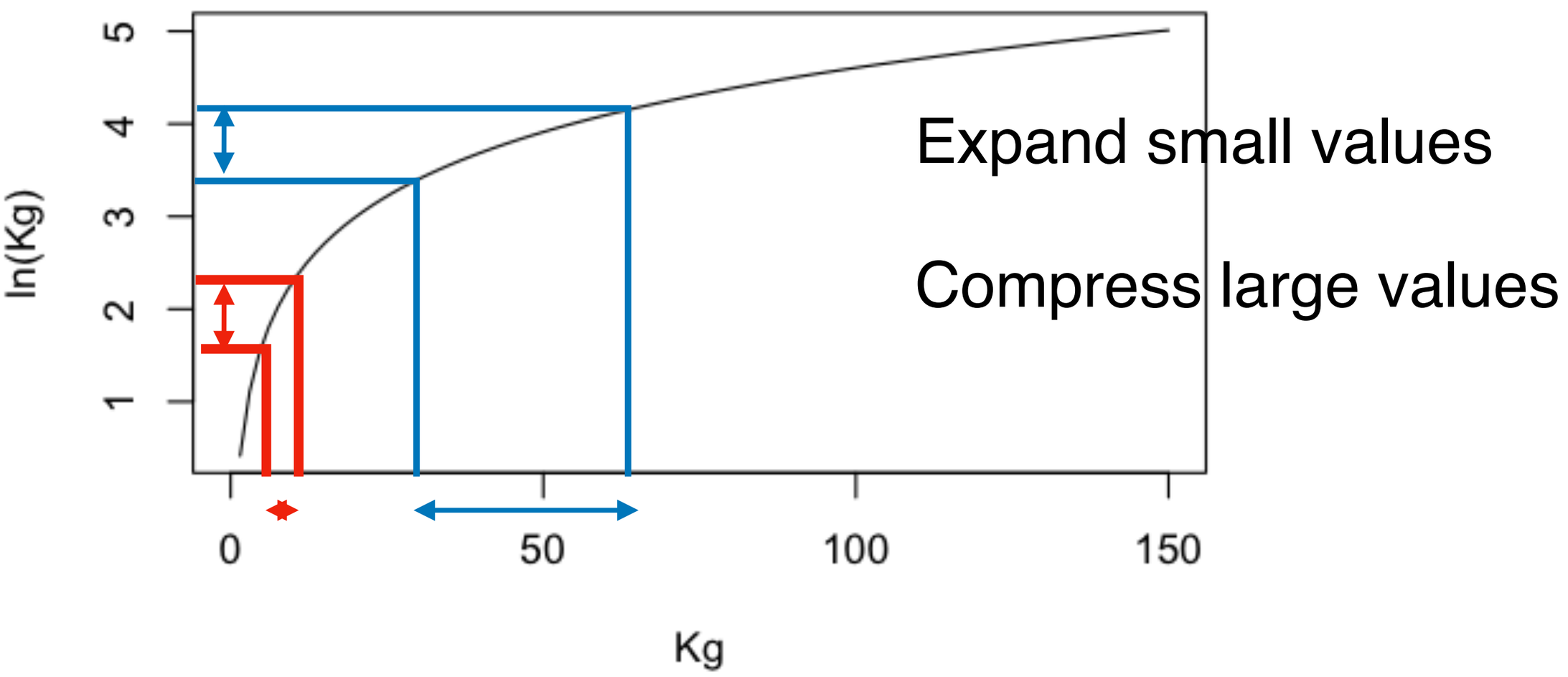
(Partial) Solution: Data transformations

1) Log-transformation:

Replace your data with log(data)

lm(log(Weight)~Species + Treatment)

Continue with normal analysis



Useful when:

all $y_{ij} > 0$

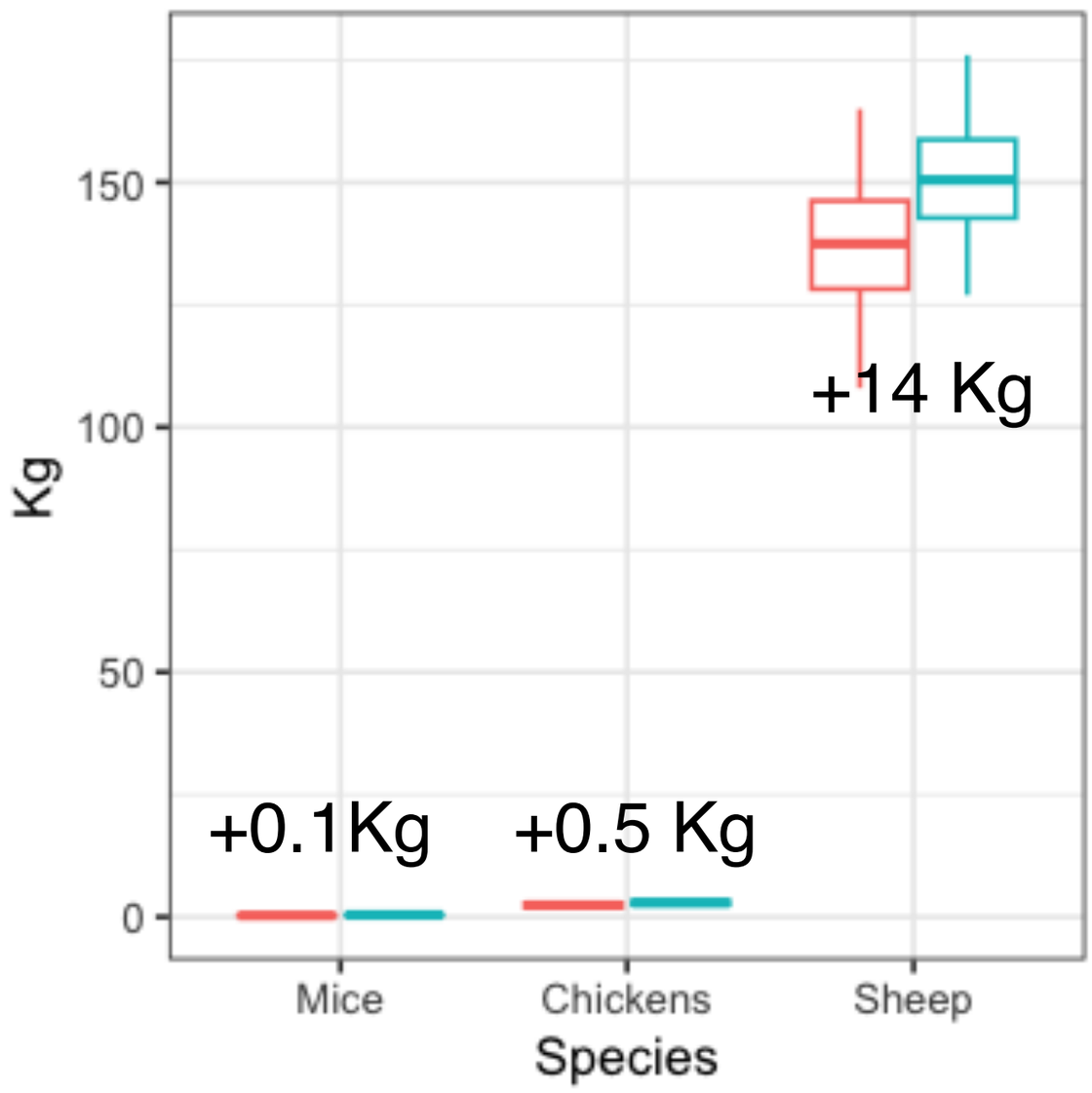
y covers a large range

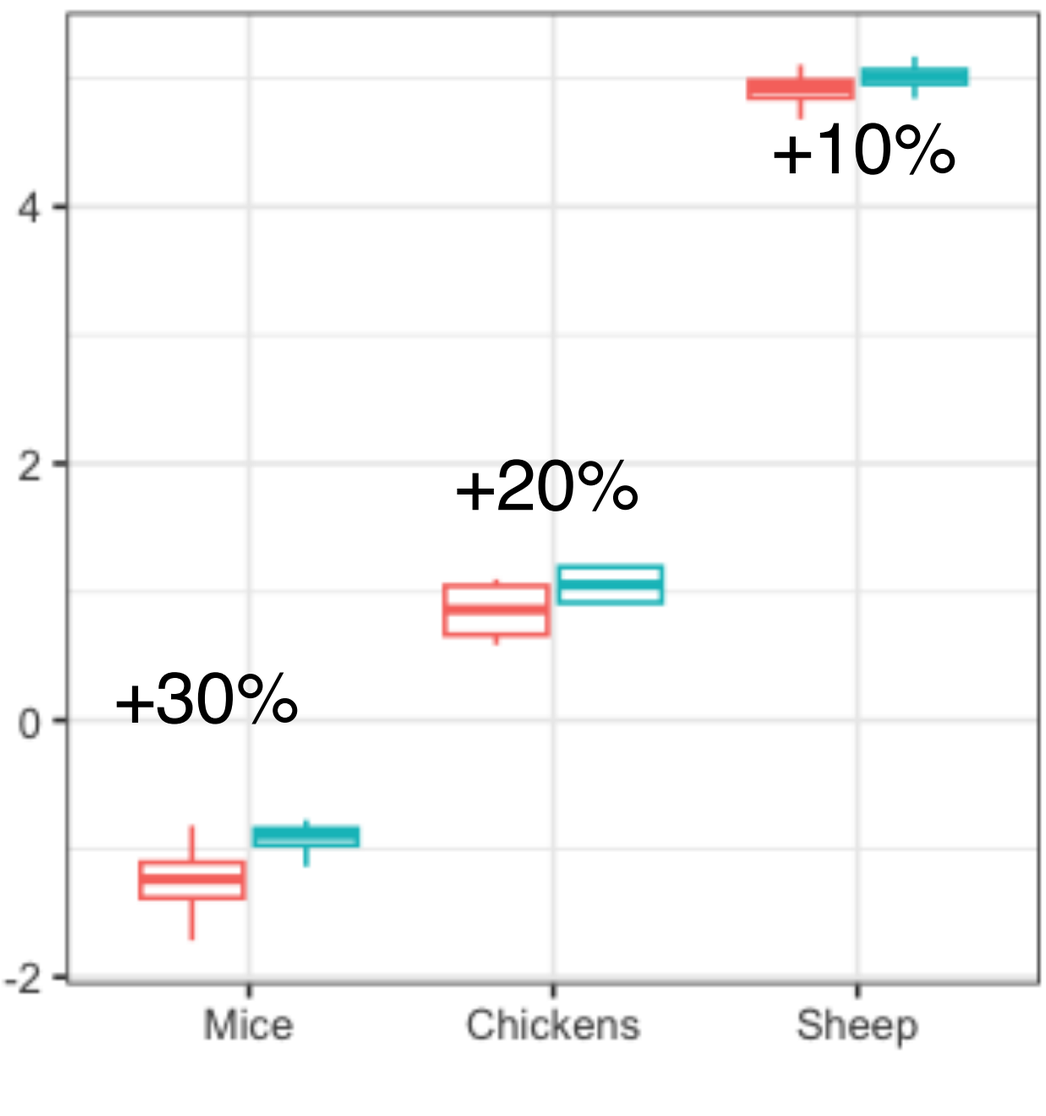
σ_y proportional to the mean in S/L plot

We are interested in **relative changes**

(Partial) Solution: Data transformations

1) Log-transformation: $y_{ij} \longrightarrow ln(y_{ij})$






```
Species = Mice:
 contrast          estimate    SE df t.ratio p.value
 Vitamin - Control    0.328 0.157 18   2.095  0.0506

Species = Chickens:
 contrast          estimate    SE df t.ratio p.value
 Vitamin - Control    0.203 0.157 18   1.294  0.2119

Species = Sheep:
 contrast          estimate    SE df t.ratio p.value
 Vitamin - Control    0.102 0.157 18   0.650  0.5239
```

Interpretation of treatment effects:

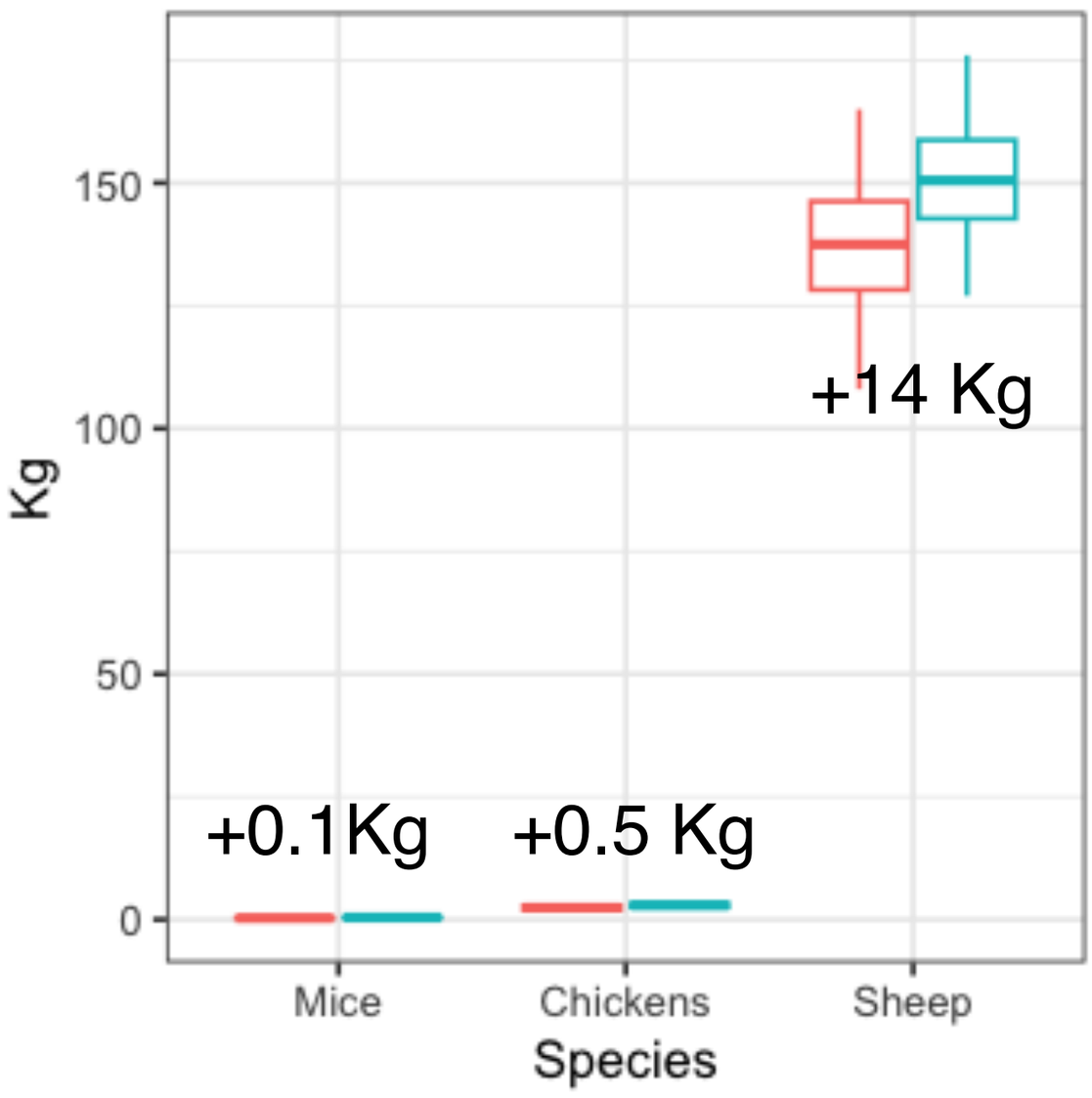
0.1 ~ 10% increase relative to control: $\frac{\mu_V - \mu_C}{\mu_C}$

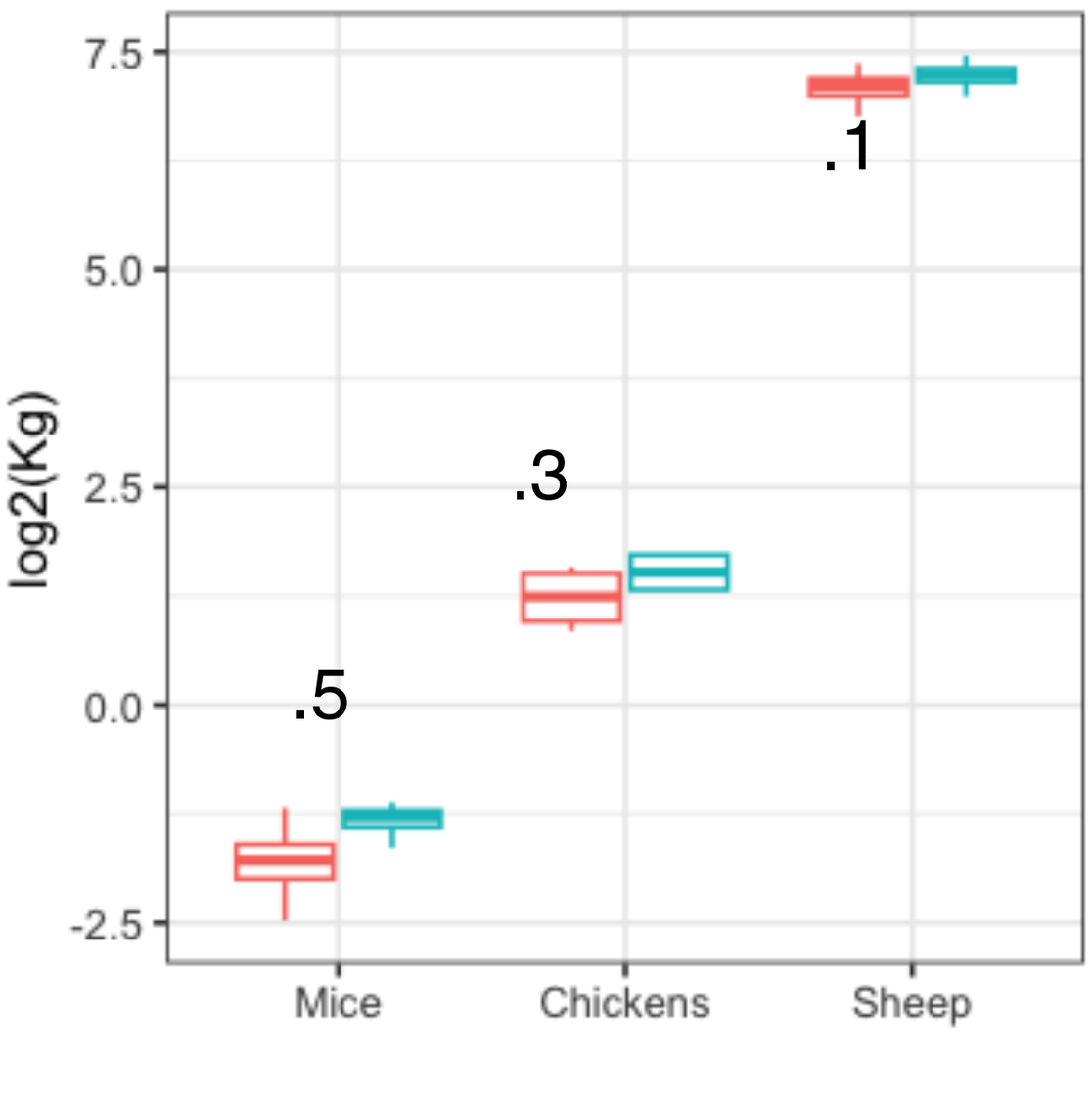
0.2 ~ 20% increase relative to control

Approximation breaks down for larger effects

(Partial) Solution: Data transformations

1) Log-transformation: $y_{ij} \longrightarrow \log_2(y_{ij})$






```
Species = Mice:
 contrast           estimate    SE df t.ratio p.value
 Vitamin - Control     0.474 0.226 18   2.095  0.0506

Species = Chickens:
 contrast           estimate    SE df t.ratio p.value
 Vitamin - Control     0.293 0.226 18   1.294  0.2119

Species = Sheep:
 contrast           estimate    SE df t.ratio p.value
 Vitamin - Control     0.147 0.226 18   0.650  0.5239
```

Interpretation of treatment effects: # **doublings** of control

0.5 = Vitamin is 2^0.5 times control = 1.4x

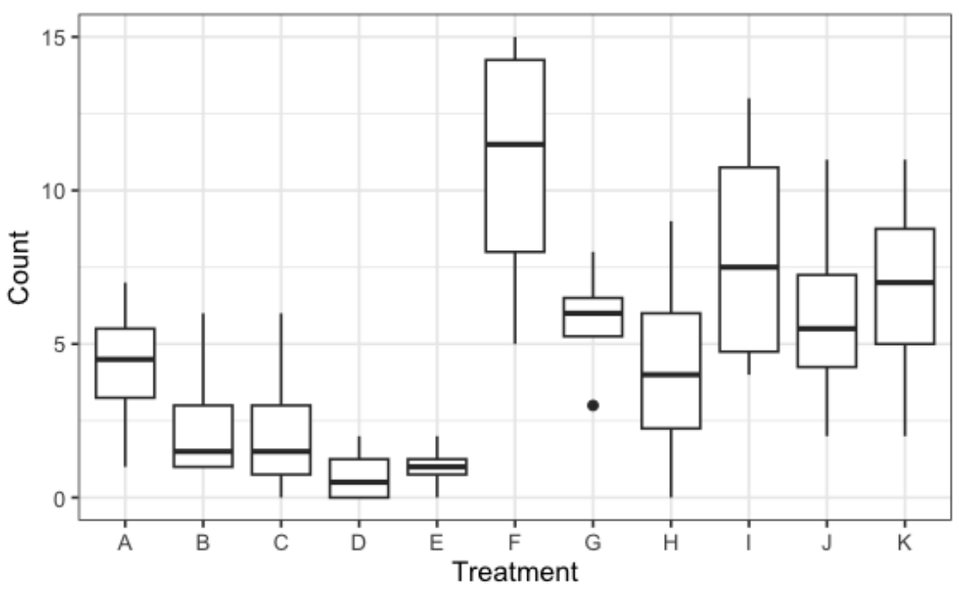
1 = Vitamin is 2 times control = 2x

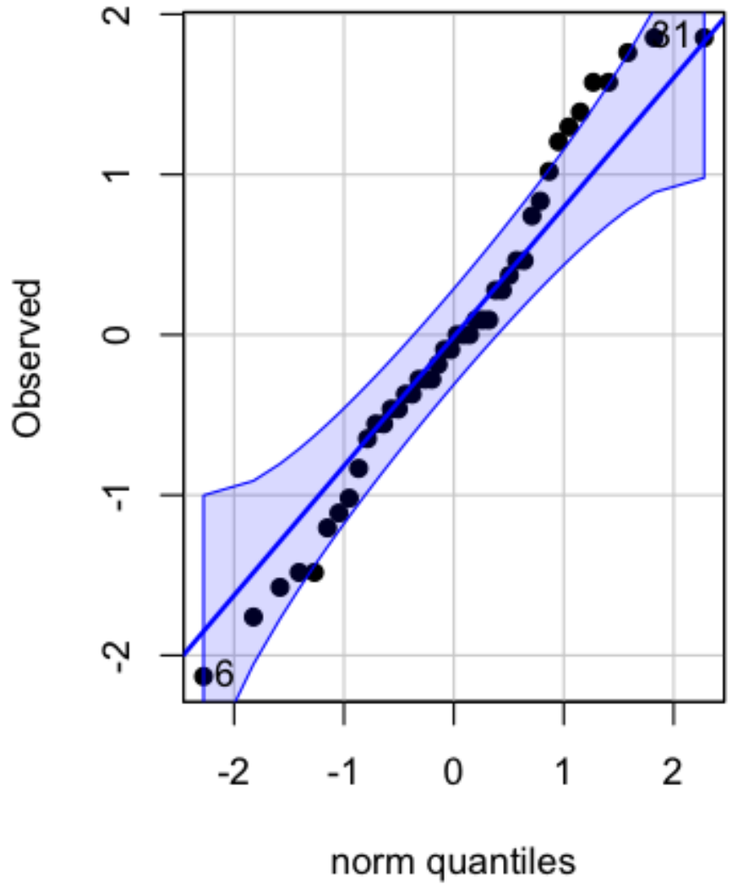
Useful for bigger effects

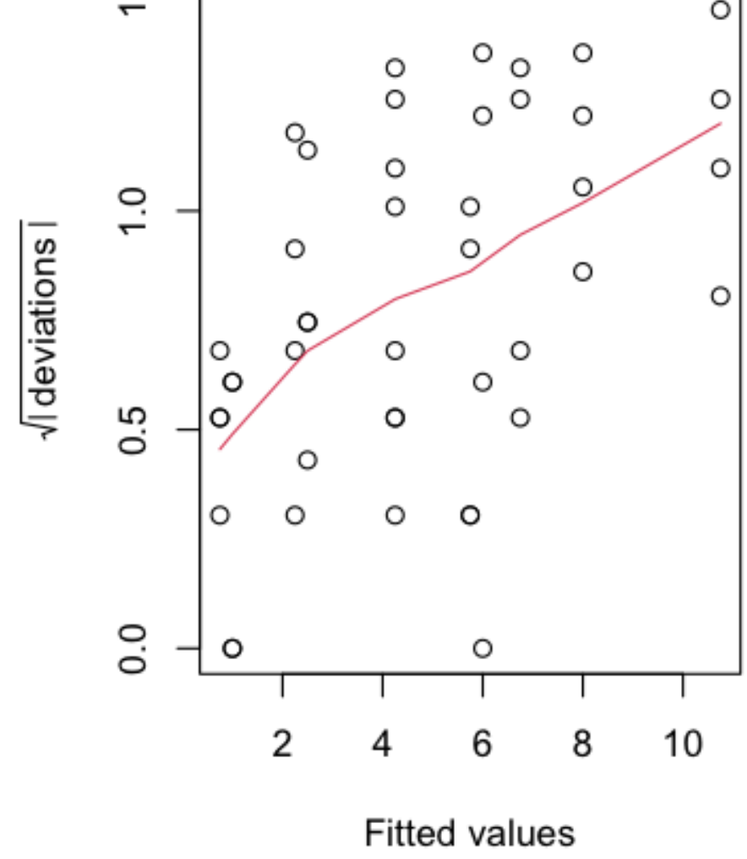
Other transformations

2) sqrt transformation y -> sqrt(y + 1/2)

bugs/plant







Useful when:

all $y_{ij} \geq 0$

y are counts

σ_y^2 proportional to the mean in S/L plot

Problem:

No interpretation to effect sizes on transformed scale

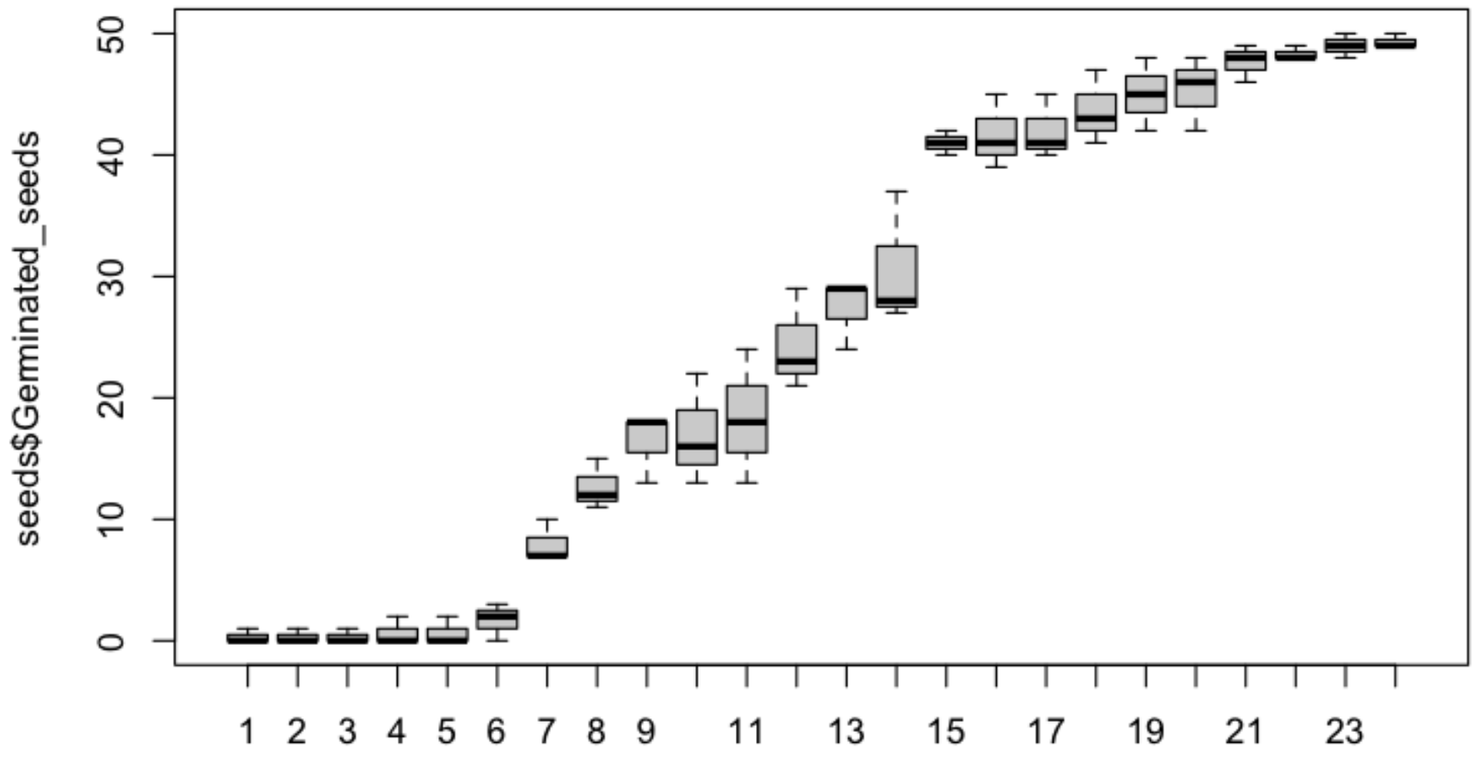
Report:

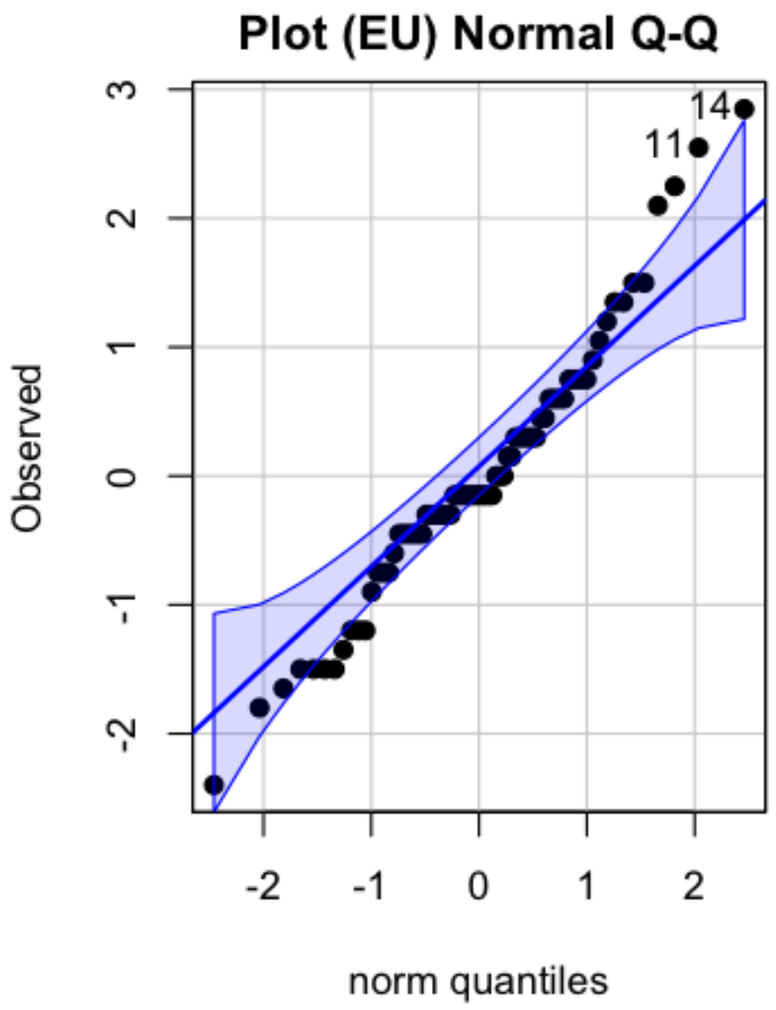
Hypothesis tests

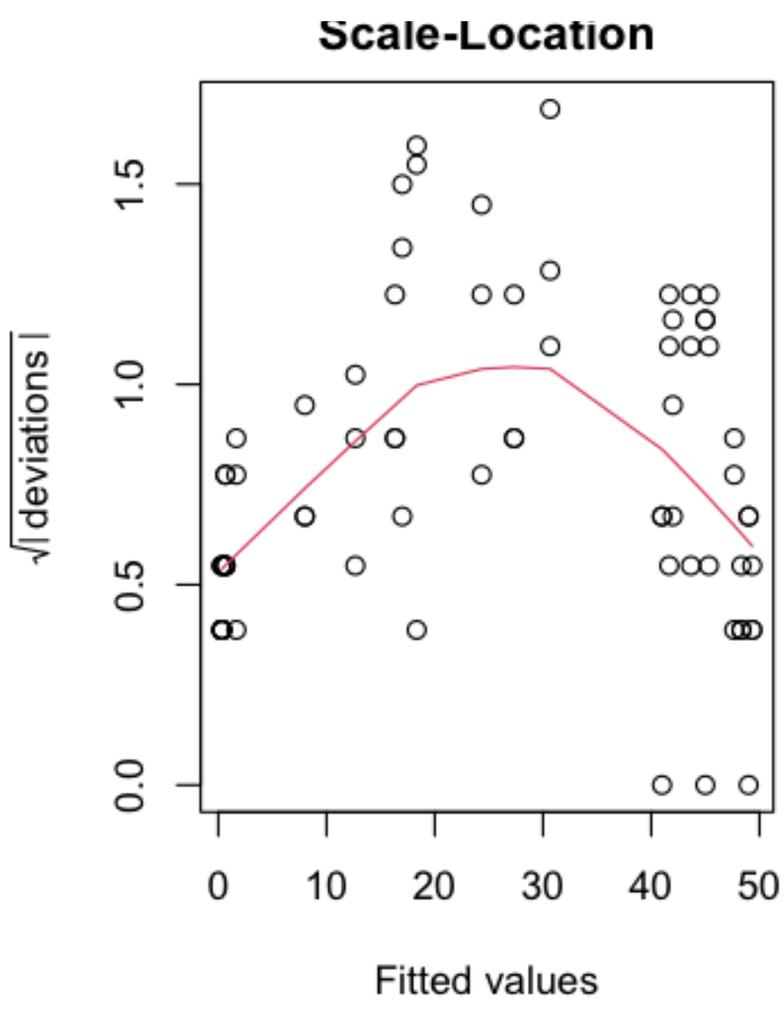
De-transformed means

Other transformations

3) Logit transformation







Useful when:

y_{ij} are proportions between 0 and 1

Interpretation on transformed scale:

Effects are fold-changes in **odds** of event

$$\frac{\text{Prob(yes)}}{\text{Prob(no)}}$$

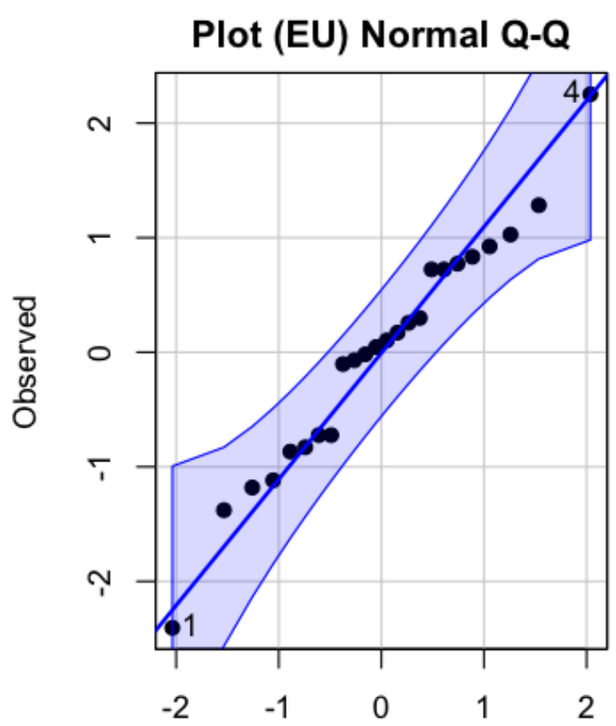
Other transformations

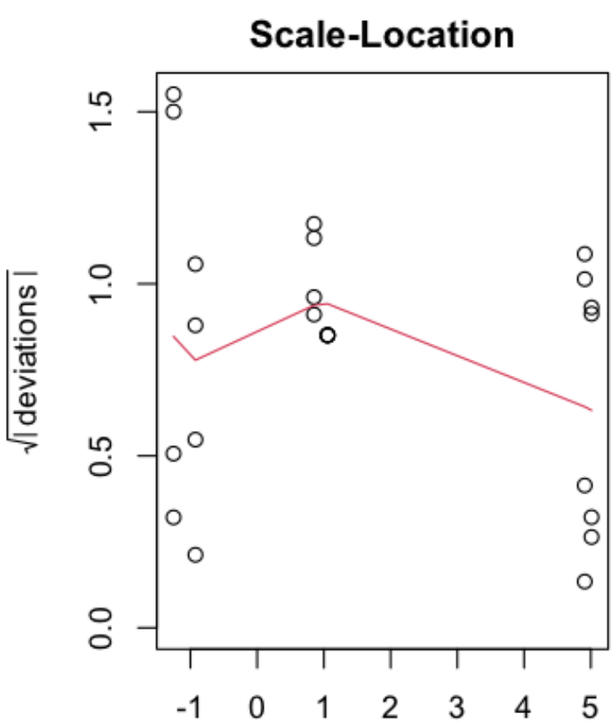
4) Box-Cox transformation

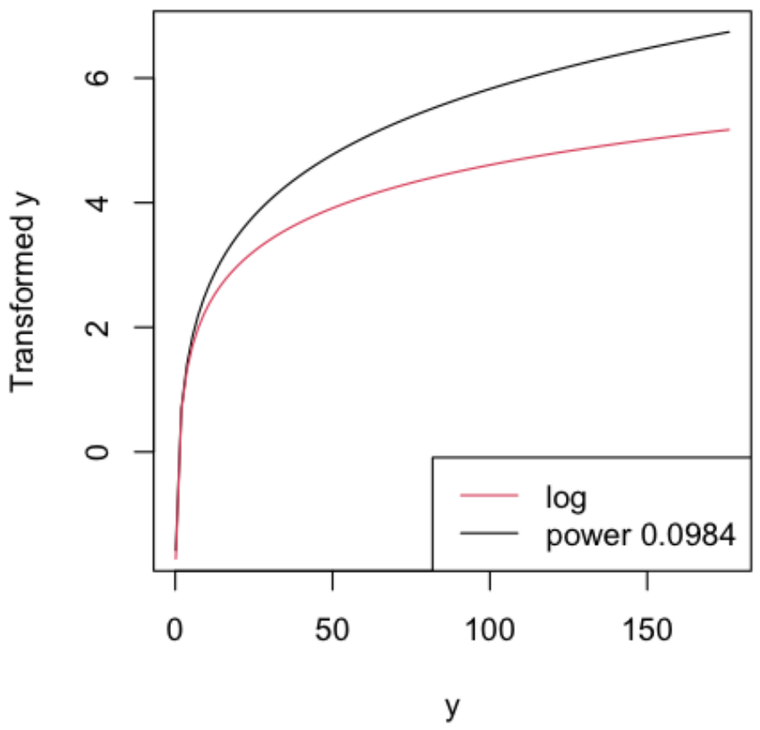
“Best” for satisfying homoscedasticity assumption

No interpretation on transformed scale

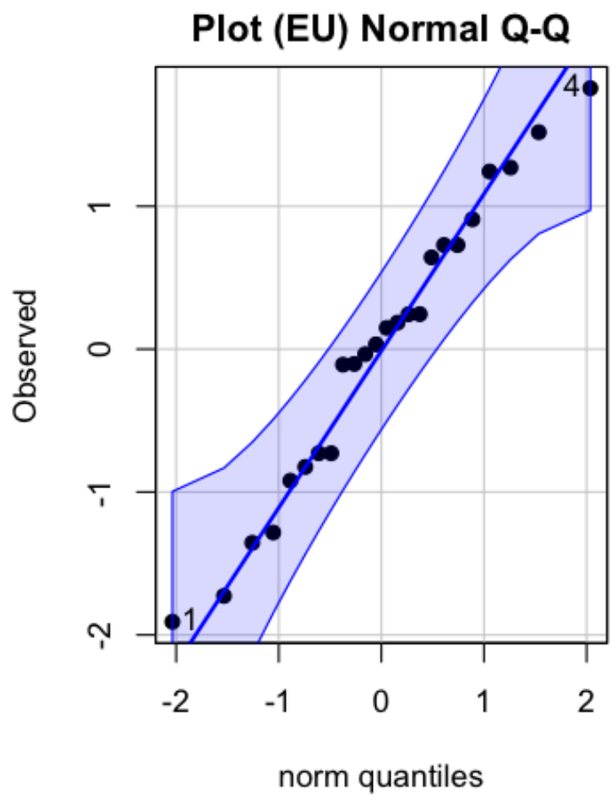
Log

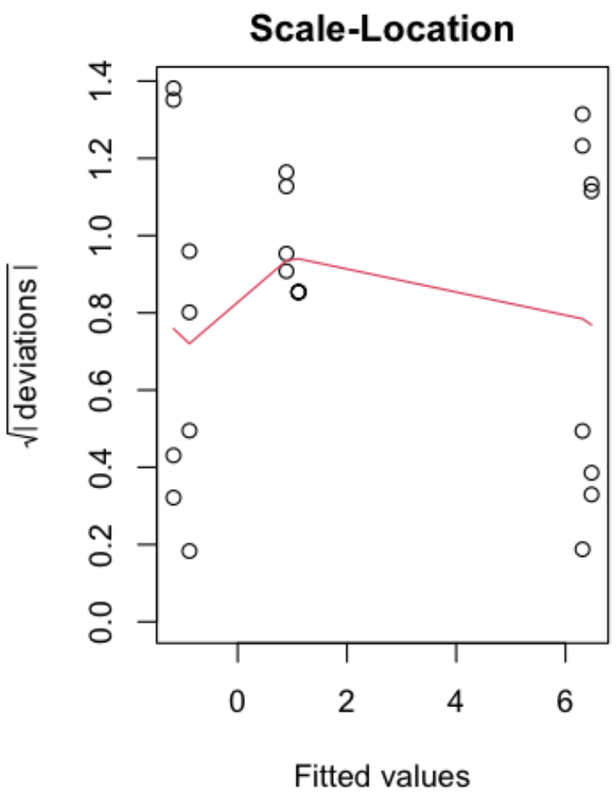






Power





Why transform?

1) Better satisfy model assumptions

Otherwise, statistics are unreliable

2) Alternate measurement scales are OK (or better)

^{o}F -> C $[H^{+}]$ -> pH Length -> Area

3) We care about non-additive effects

% change, fold-change

Rules:

1) Choose by diagnostic plots or theory

Not because it gives you $p < 0.05$!

2) Transform data before analysis

a) Report **effects** on transformed scale

b) Report **de-transformed** means + CIs