How to design better experiments

Mridul K. Thomas & Ravi Ranjan

https://www.linkedin.com/in/mridulkthomas/
https://bsky.app/profile/raviranjan.bsky.social

What we want you to take from this workshop

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.

What we want you to take from this workshop

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.
2. Consider what **parameter**/quantity is most useful to estimate, and what **precision** you need.

What we want you to take from this workshop

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.
2. Consider what **parameter**/quantity is most useful to estimate, and what **precision** you need.
3. Think carefully about your **model** and **error**.

What we want you to take from this workshop

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.
2. Consider what **parameter**/quantity is most useful to estimate, and what **precision** you need.
3. Think carefully about your **model** and **error**.
4. When planning your experiment, **simulate** data from your model and **compare precision** under different designs.

What we want you to take from this workshop

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.
2. Consider what **parameter**/quantity is most useful to estimate, and what **precision** you need.
3. Think carefully about your **model** and **error**.
4. When planning your experiment, **simulate** data from your model and **compare precision** under different designs.
5. Check **robustness** by changing assumptions and simulating again.

Talking about experimental design

Experimental unit: the lowest level that you manipulate

Level/Treatment: the value of a driver/factor used in the experiment

Design space: the set of all possible levels

Replicates: experimental units that are subjected to identical levels

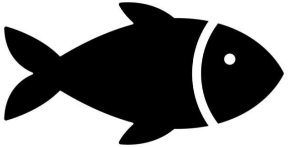
Suppose you are interested in how a toxin affects a fish species

Experimental unit: individual fish

Level/Treatment: toxin concentrations in the experiment

Design space: all possible toxin concentrations you could use

Replicates: fish that are given identical toxin levels



Classical power analysis for an ANOVA with 2 groups

What you need:

1. Sample size per group
2. Probability of incorrectly rejecting the null hypothesis when it is true (false positive)
3. Within-group standard deviation
4. Smallest effect size you are interested in

Classical power analysis for an ANOVA with 2 groups

What you need:

1. Sample size per group
2. Probability of incorrectly rejecting the null hypothesis when it is true (false positive)
3. Within-group standard deviation
4. Smallest effect size you are interested in

What you get out:

Probability of correctly rejecting the null hypothesis when the null hypothesis is false.

Simulation-based power analysis

Classical power analysis becomes harder for more complicated models*

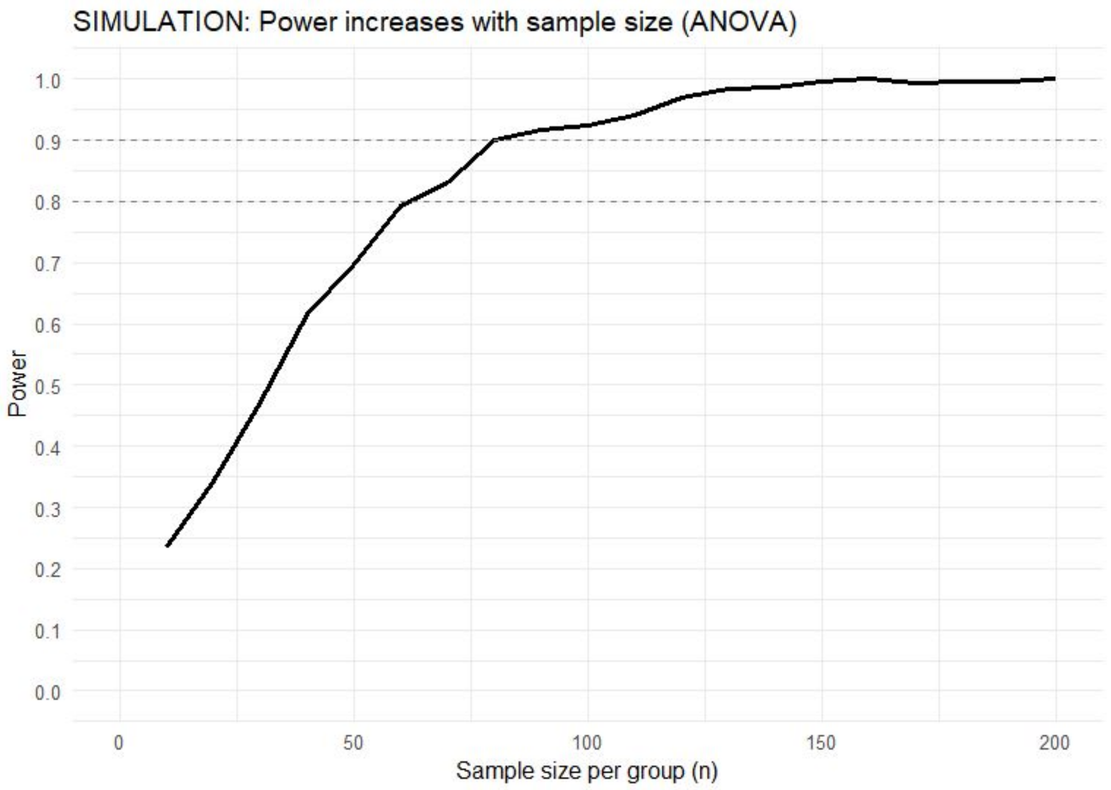
Simulations are **easy**, **flexible** and **powerful**.

1. Decide on the model you will fit to the data
2. Make up numbers for the parameters
3. Run it 1000+ times
4. Calculate % of times you get a $p < 0.05$. This is the power.

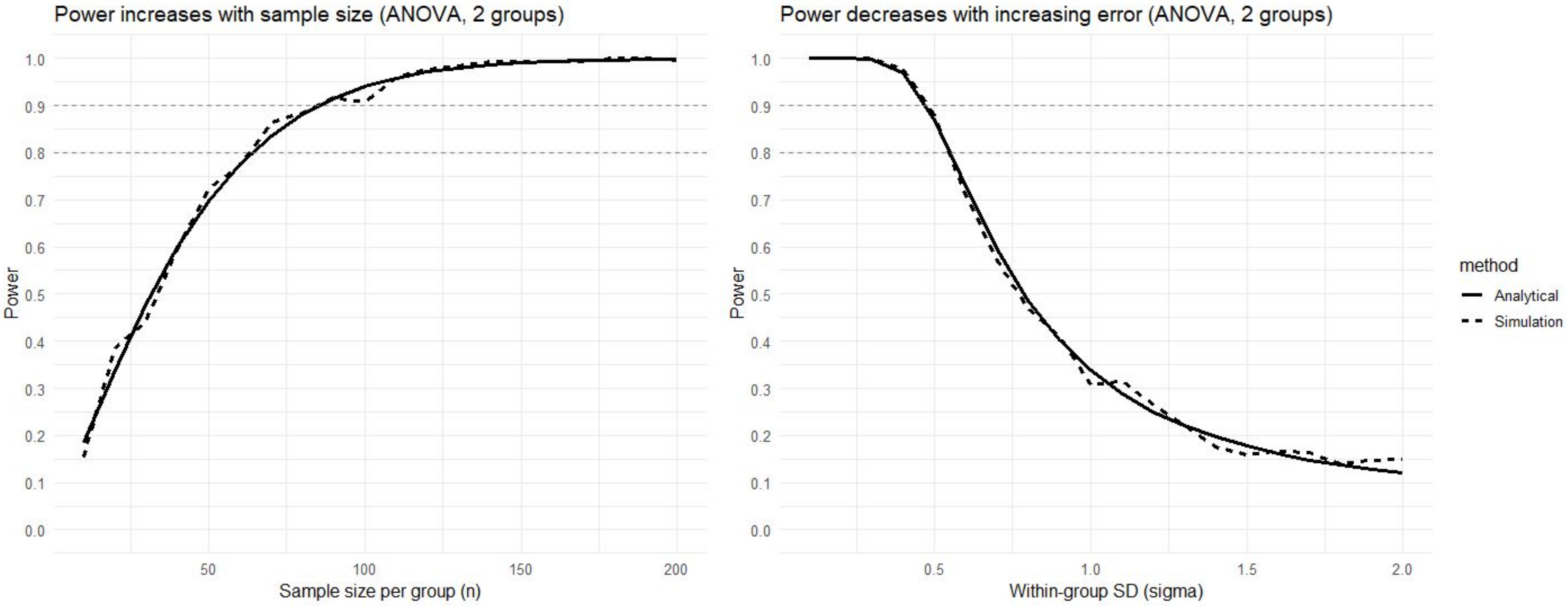
*Resources
https://cran.r-project.org/web/packages/pwrss/vignettes/examples.html
https://stats.stackexchange.com/questions/643560/is-it-possible-to-calculate-the-power-for-a-specific-beta-coefficient-in-a-multi

Simulation-based power analysis

1. Sample size per group: X axis
2. Probability of incorrectly rejecting the null hypothesis when it is true (false positive): 0.05
3. Within-group standard deviation: 1
4. Smallest effect size we are interested in: 0.5



Simulations give you the same results as the classical approach



But power analysis **focuses on the wrong thing**

For almost any experiment, we probably **know that the true value is different from zero**.

A statistically significant result is just about using enough experimental units.

It is a game, a **dead end**.

What should we be trying to do instead?

Simulations can be used for **precision analysis/planning**

Simulation-based precision analysis

The typical model you will use when comparing two groups is:

$$\mathbf{Y = \alpha + \beta X + \epsilon}$$

Y = response
X = predictor (group identity, 0/1)
α = intercept (mean of first group)
β = difference between groups
ϵ = error (usually normally distributed with a mean of zero)

If you are unfamiliar with this: it is just adding three numbers!

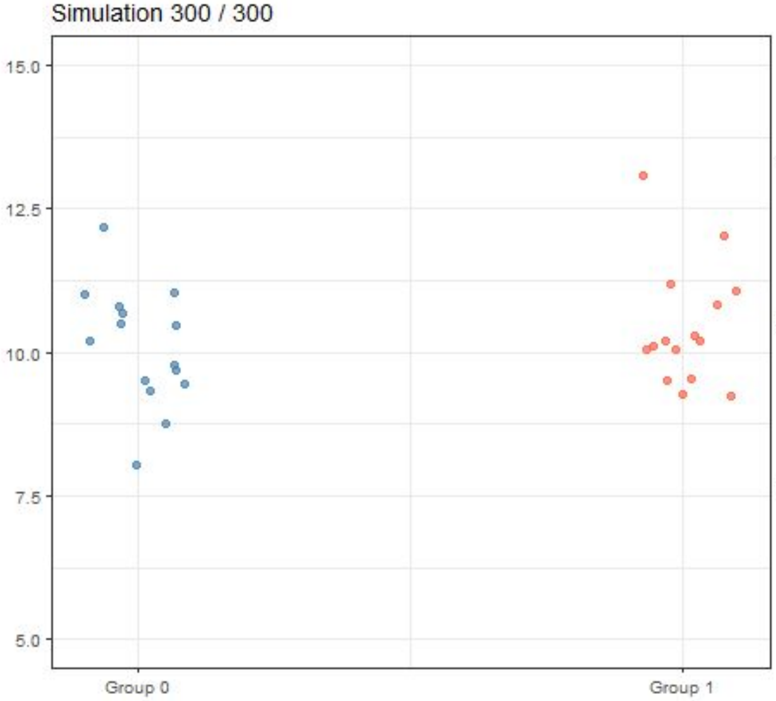
Simulation-based precision analysis

The typical model you will use when comparing two groups is:

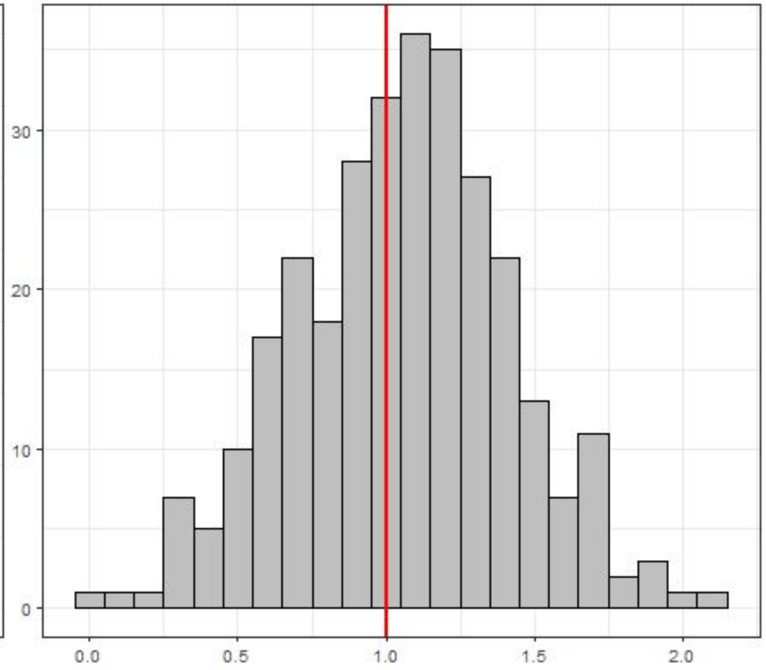
$$\mathbf{Y = \alpha + \beta X + \epsilon}$$

1. Choose values for α, β and the standard deviation of ϵ.
    a. Note that ϵ is random noise. We add it to reflect natural variation.
2. Also choose the sample size per group
3. Simulate data 1000+ times
4. Fit your model and estimate α and β in each of them.
5. How much variation is there in α and β ?

[final slide of animation] Simulation-based power analysis

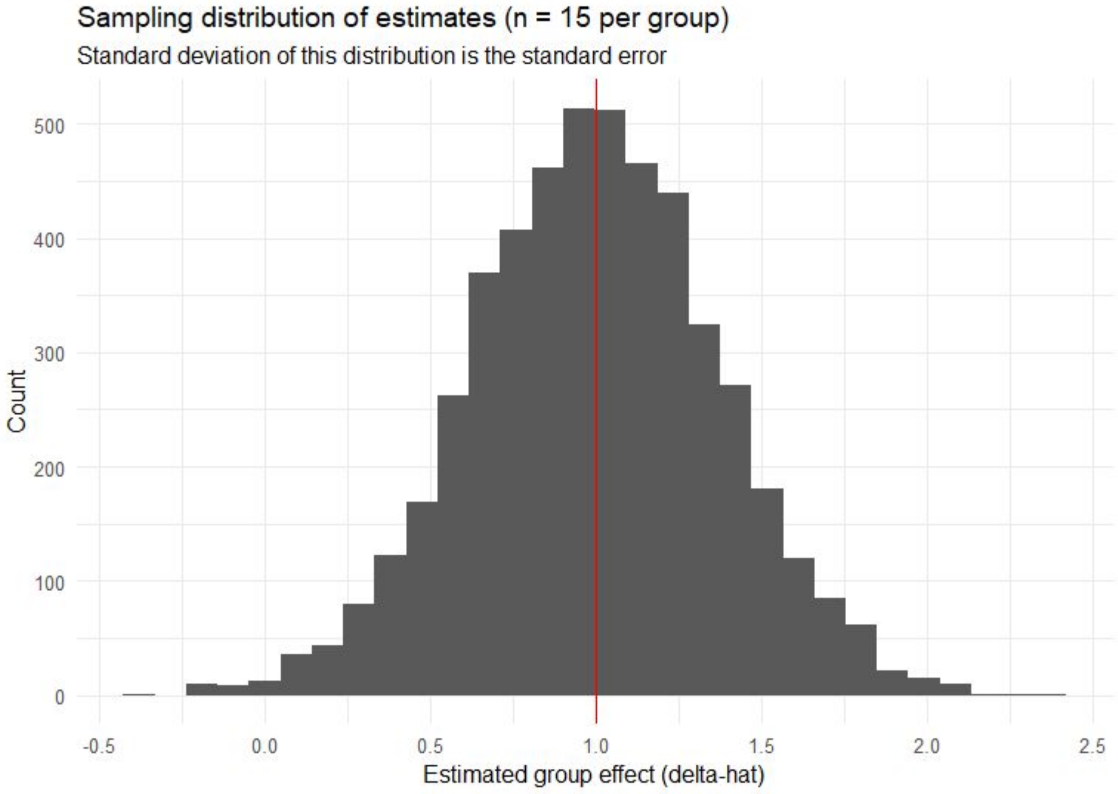


Simulated data



Estimated difference between groups (β)

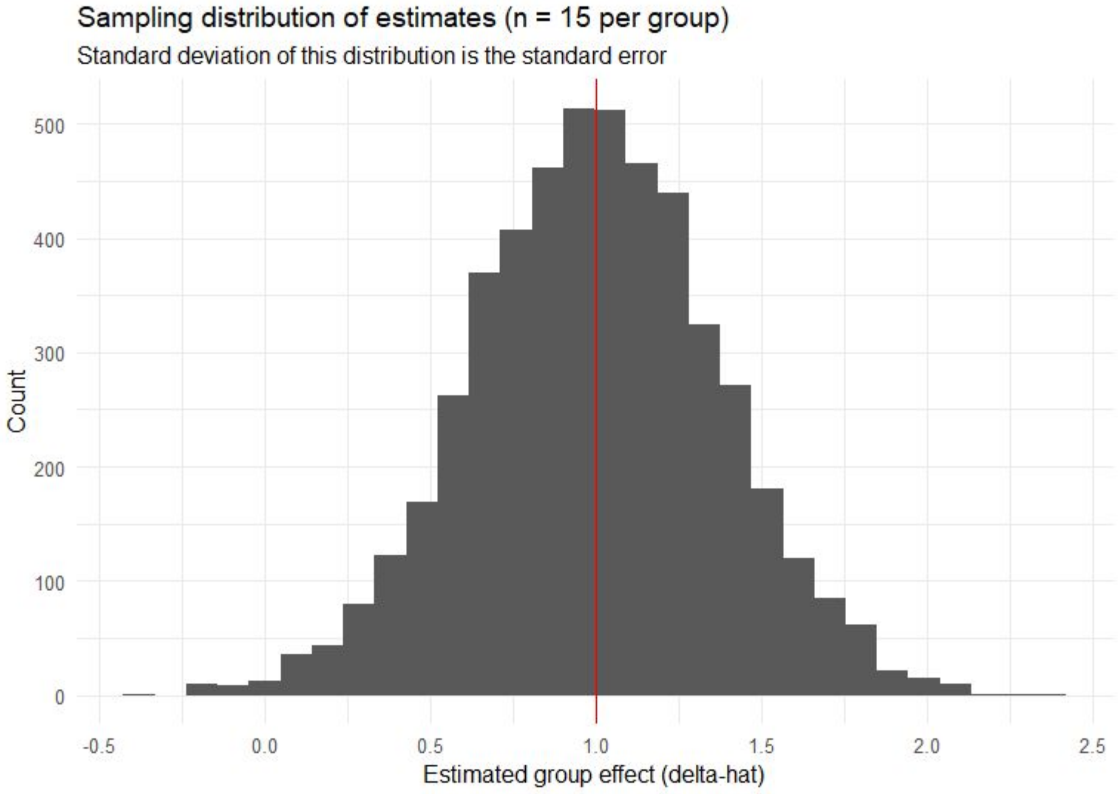
Expected SE is 0.37, but how does this change with sample size?



R demo

R script #3

What if we are wrong about the error we assumed?



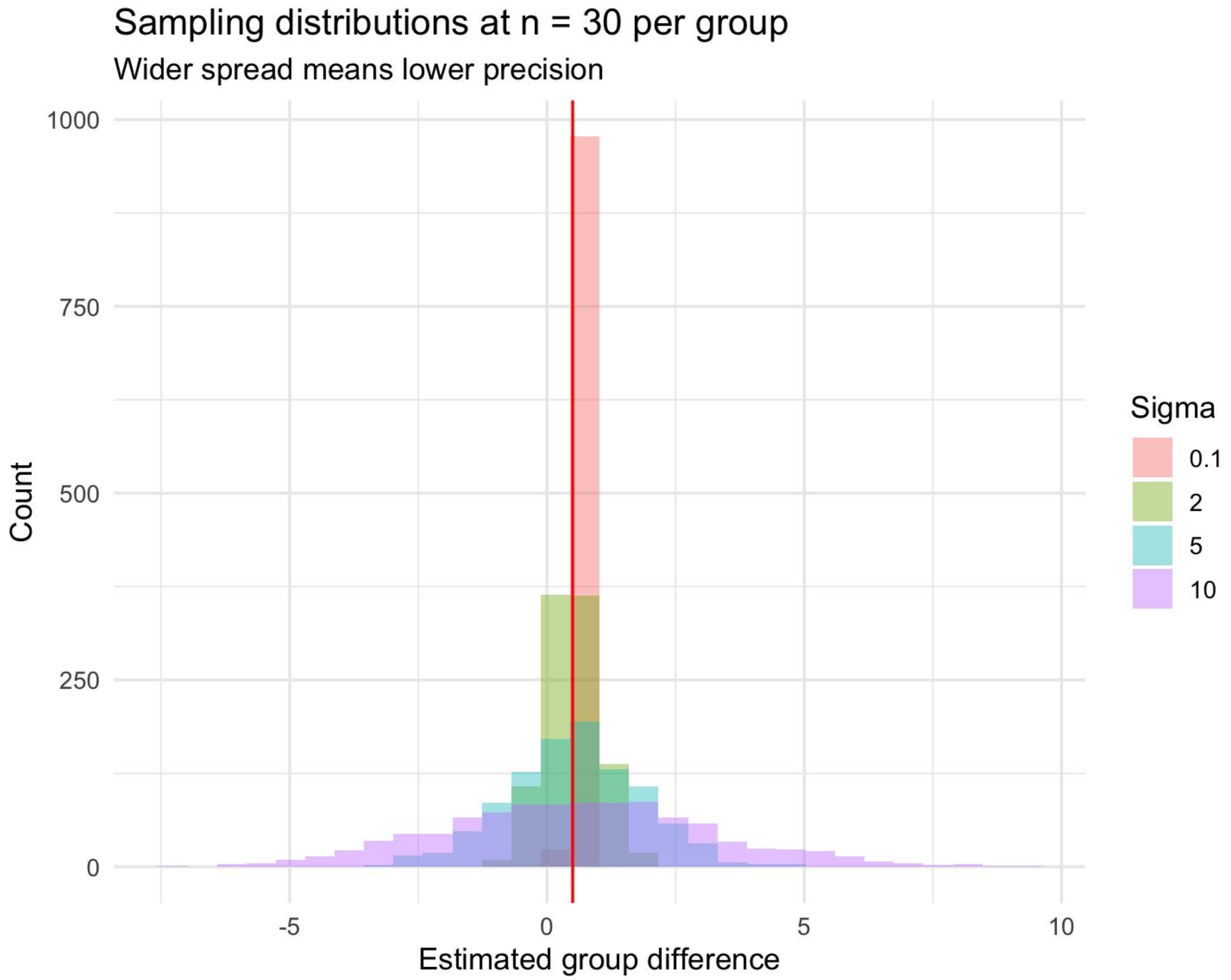
ACTIVITY - 5 minutes

TAKE A SCREENSHOT NOW

Working individually:

Figure out how precision changes with error, keeping the number of points constant.

Use **R script #4**


Sampling distributions at n = 30 per group
Wider spread means lower precision
1000
750
500
250
0
Count
-5
0
5
10
Estimated group difference
Sigma
0.1
2
5
10

Precision is good, but **what are we trying to actually achieve?**

What are we trying to achieve?

1. **Testing** a scientific hypothesis or establish **causation** -> most scientists think they are doing this
2. **Predicting** something of importance
3. **Estimating** an input for a mathematical model (to understand or predict)
4. **Exploring** how some change affects some property of interest

What is a **hypothesis**?

To us

1. In **science**: an explanation that relates two phenomena via a mechanism.

2. In **statistics**: a mathematical statement that can be evaluated (often as TRUE/FALSE) with data.

What is a **scientific** hypothesis?

Let’s work through turning a general claim into a scientific hypothesis.

What is a **scientific** hypothesis?

1. Biodiversity is related to the environment

What is a **scientific** hypothesis?

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature

What is a **scientific** hypothesis?

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature

What is a **scientific** hypothesis?

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature
4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster

What is a **scientific** hypothesis?

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature
4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster
5. The number of species per unit area increases with temperature with a slope of ¾ because (i) metabolic rates increases with temperature with a slope of ¾, (ii) higher metabolism means more individuals are born, and (iii) more individuals means that more diversity can survive.

We would call 4 & 5 scientific hypotheses

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature
4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster
5. The number of species per unit area increases with temperature with a slope of ¾ because (i) metabolic rates increases with temperature with a slope of ¾, (ii) higher metabolism means more individuals are born, and (iii) more individuals means that more diversity can survive.

The same statistical model would work for #3 onwards

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature
4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster
5. The number of species per unit area increases with temperature with a slope of ¾ because (i) metabolic rates increases with temperature with a slope of ¾, (ii) higher metabolism means more individuals are born, and (iii) more individuals means that more diversity can survive.

We might use different experiments for #2, 3, 4, and 5

1. Biodiversity is related to the environment
2. The number of species in an ecosystem changes with temperature
3. The number of species per unit area increases with temperature
4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster
5. The number of species per unit area increases with temperature with a slope of ¾ because (i) metabolic rates increases with temperature with a slope of ¾, (ii) higher metabolism means more individuals are born, and (iii) more individuals means that more diversity can survive.

We might use different experiments for #2, 3, 4, and 5

2. The number of species in an ecosystem changes with temperature

3. The number of species per unit area increases with temperature

- Manipulate 2 temperatures
- Measure number of species

We might use different experiments for #2, 3, 4, and 5

4. The number of species per unit area increases with temperature because higher temperatures cause individuals to grow faster

- Manipulate 2 temperatures
- Measure number of species AND individual growth rate

We might use different experiments for #2, 3, 4, and 5

5. The number of species per unit area increases with temperature with a slope of ¾ because (i) metabolic rates increases with temperature with a slope of ¾, (ii) higher metabolism means more individuals are born, and (iii) more individuals means that more diversity can survive.

- Manipulate >2 temperatures
- Measure number of species AND metabolic rate AND birth rate
- Ideally *manipulate* metabolic and birth rates somehow

Group activity - 5 MINUTES

1. **Refine a question** as much as possible.
2. **Design an experiment** to answer/test it.
   - What experimental levels will you use?
   - How many experimental units at each level?

TAKE A SCREENSHOT NOW

A. How does biodiversity affect ecosystem functioning?
B. Eutrophication destabilises ecosystems
C. Warming changes biodiversity.

Have one person write down your answer to share later.

Is your question/hypothesis precise enough to test well?

Testing a scientific hypothesis in one study is *rare*. Prediction is hard too.

We need mathematical models to understand and predict. This is how we predict future CO2 levels and biogeochemical cycling

Precisely estimating parameters for these mathematical models is a valuable goal for experiments.

Note: precise estimates help with prediction, exploration - and even testing hypotheses, often via meta-analysis.

What makes experiments useful for mathematical models?

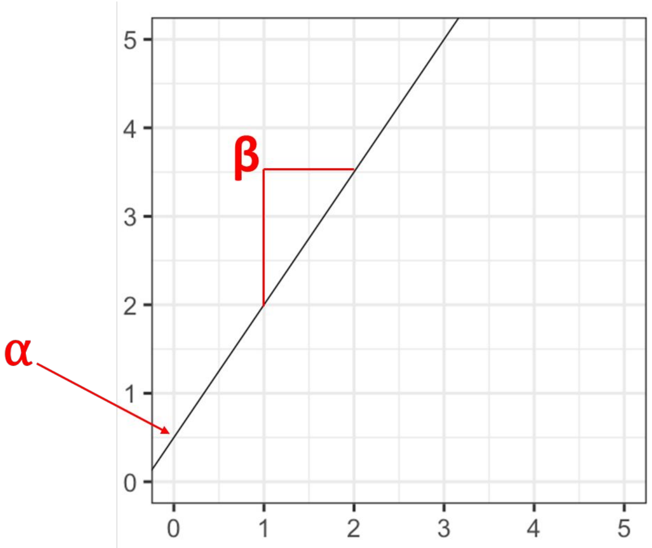
1. Regression is more useful than ANOVAs if you are dealing with continuous variables e.g. temperature, nutrients, diversity, density, etc.
2. When comparing two groups/species, consider comparing their model parameters e.g. feeding rate. There is often an underlying continuous gradient.
3. Understand the appropriate function/equation. Is it a straight line? Is it saturating? Is it unimodal? What are the parameters of this function?
4. Think about the level of parameter precision needed. Would a 20% error matter for the mathematical model? What about 50%?

Simulations can tell you what you can achieve - and where you can reduce effort!

Simulation-based precision analysis for regression

Same model as ANOVA - small differences in meaning.

$$\mathbf{Y = \alpha + \beta X + \epsilon}$$



Y = response
X = predictor (continuous value)
α = intercept
β = slope
ϵ = error (usually normally distributed with a mean of zero)

If you are unfamiliar with this: this is the equation of a straight line (+ some noise)

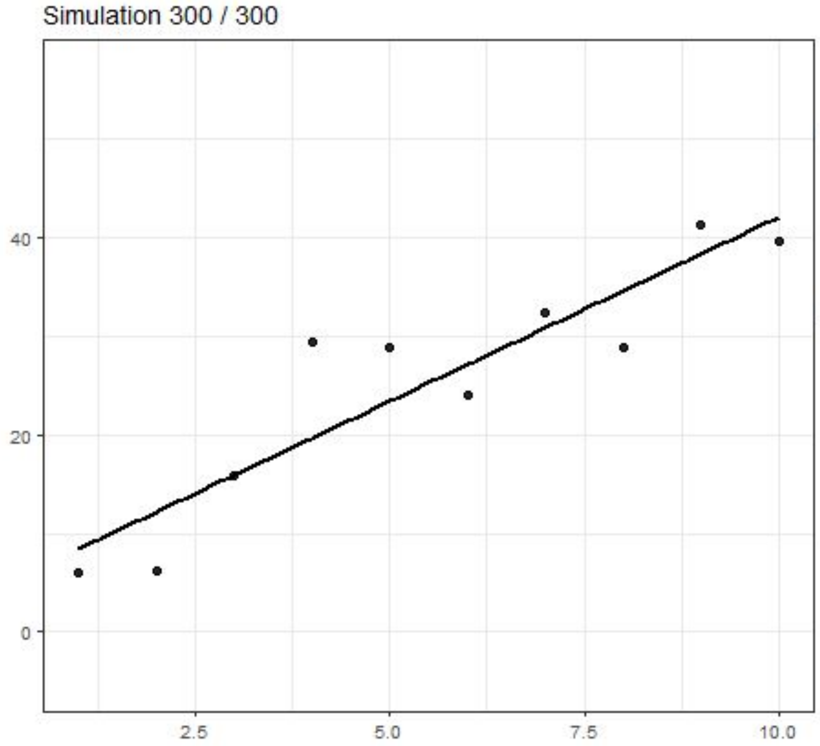
Simulation-based precision analysis for regression

The typical model you will use when estimating change across a gradient is:

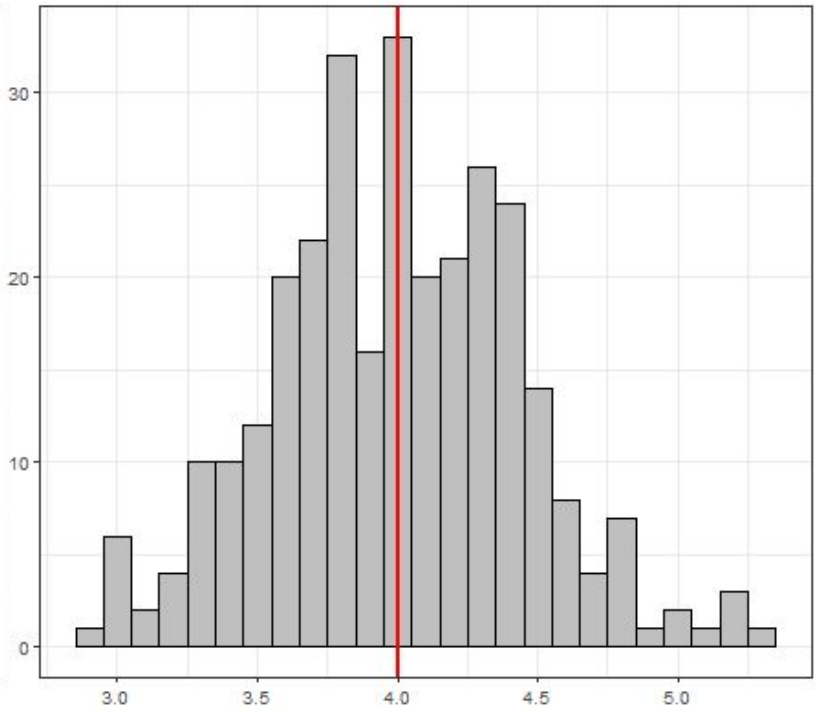
$$\mathbf{Y = \alpha + \beta X + \epsilon}$$

1. Choose values for α, β and the standard deviation of ϵ.
   a. Note that ϵ is random noise. We add it to reflect natural variation.
2. Also choose the ~~sample size per group~~ levels of X you will use in the experiment
3. Simulate data 1000+ times
4. Fit your model and estimate α and β in each of them.
5. How much variation is there in α and β ?

[final slide of animation] Simulation-based precision analysis



Simulated data and fitted regression line



Estimated slope (β)

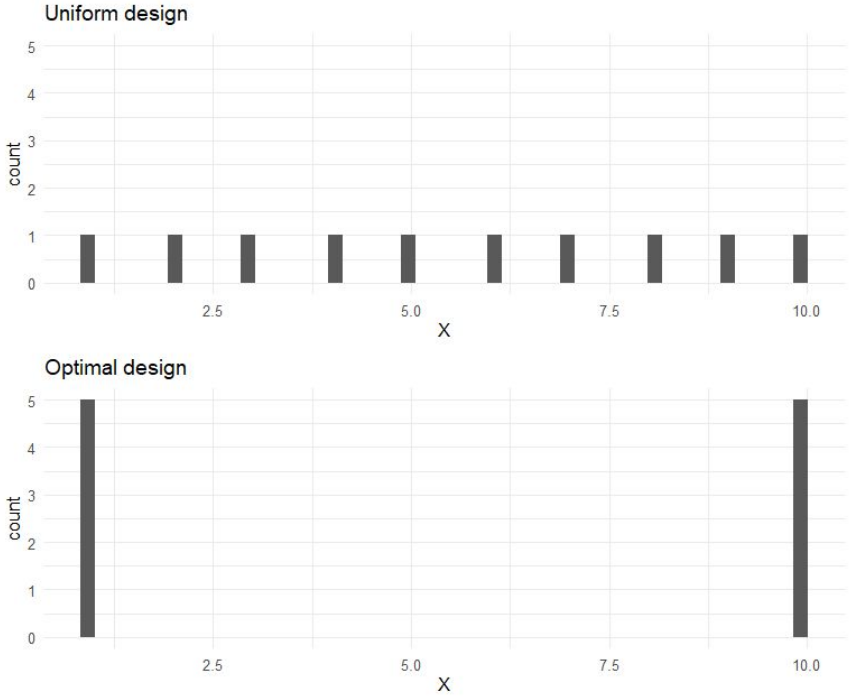
You have **one more choice** in regression designs

What will your experimental levels be?

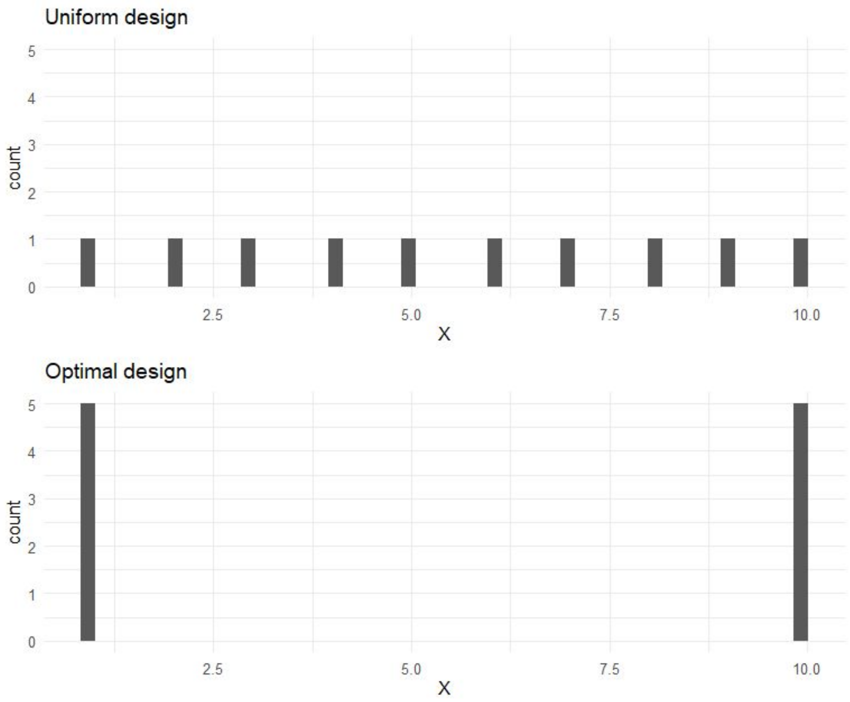
i.e.

Where are you going to put your points?

Here are two designs. One gives us more precise estimates.



Here are two designs. One gives us more precise estimates.

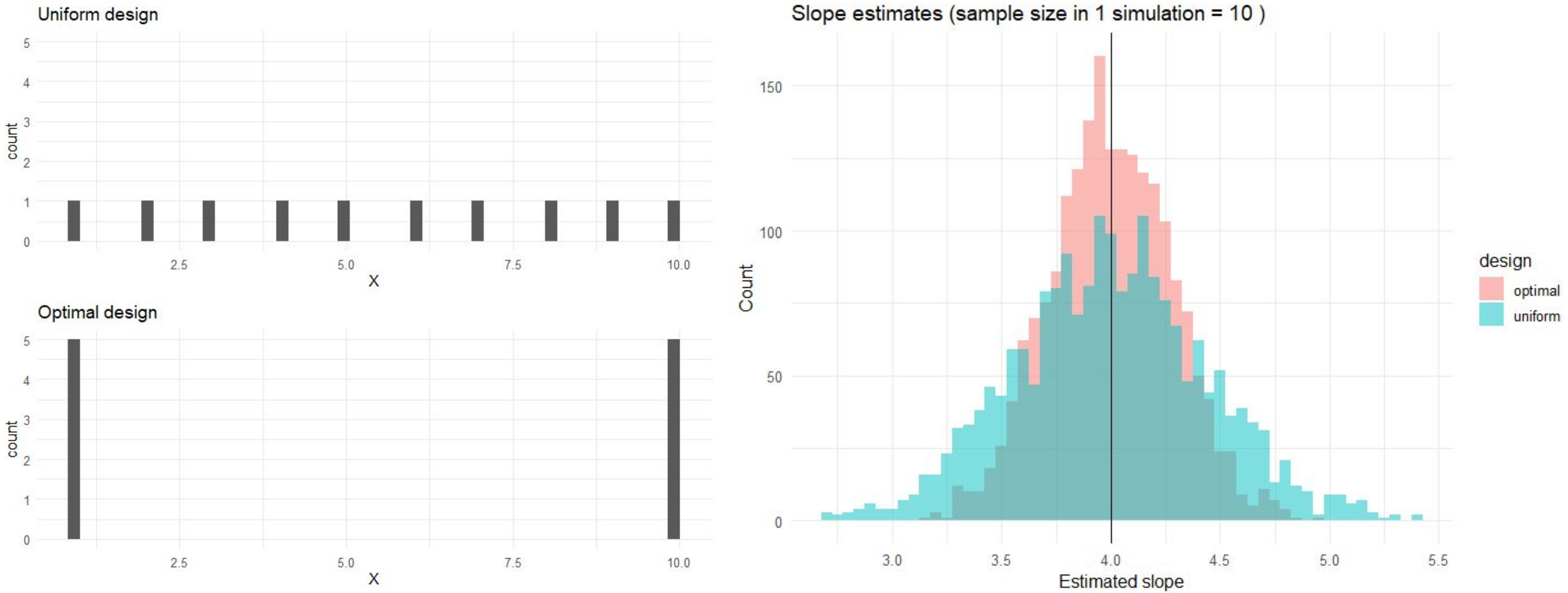


Why?

After I show the result, think for a minute.

Type your answer in chat but **do not hit enter till I say.**

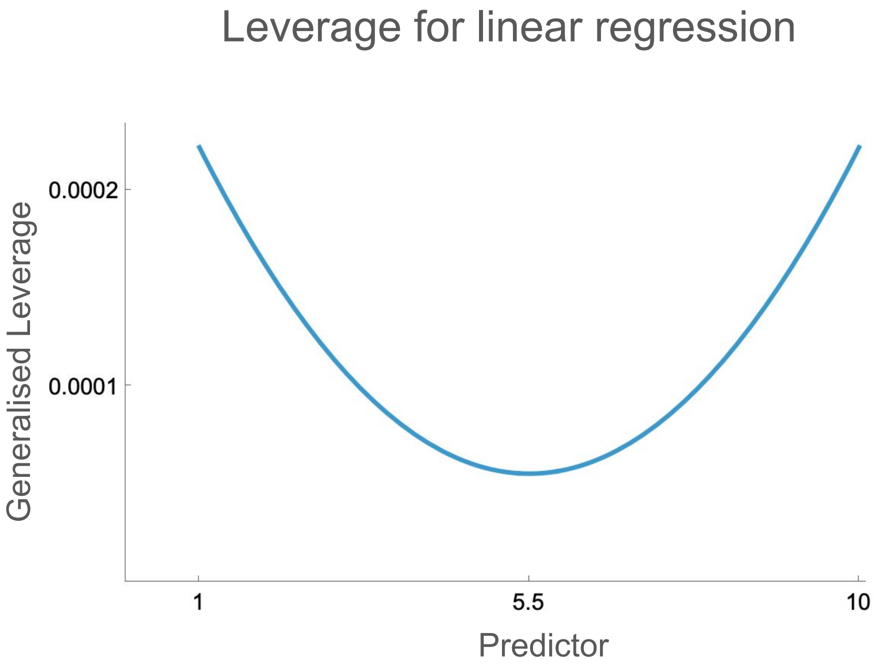
Why is the extreme/optimal design more precise?



(For later) You can do this yourself with **R script #5**

Leverage tells you how valuable a point is.

- All points are not equally valuable.
- Generalised Leverage: measures relative influence of a point on the fit.
- Combines parameter sensitivities:
  - How sensitive is the function output to a change in a parameter?
  - Changes with the predictor value
- Good diagnostic tool, but how do we develop a design?

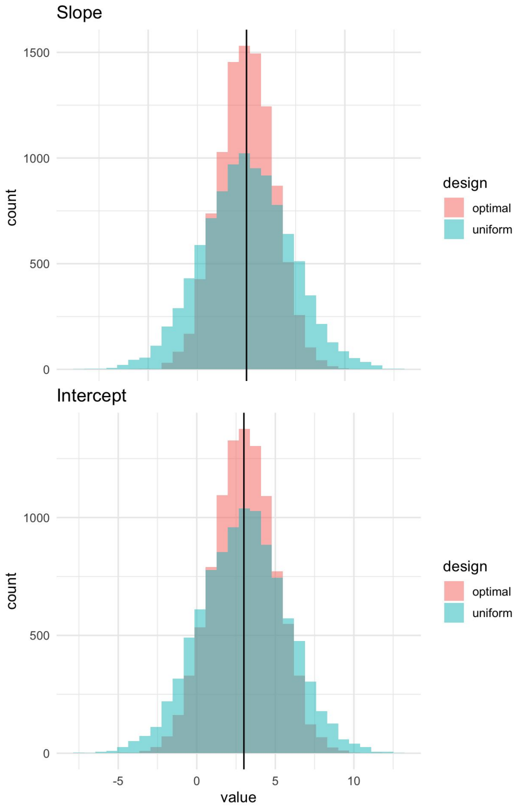


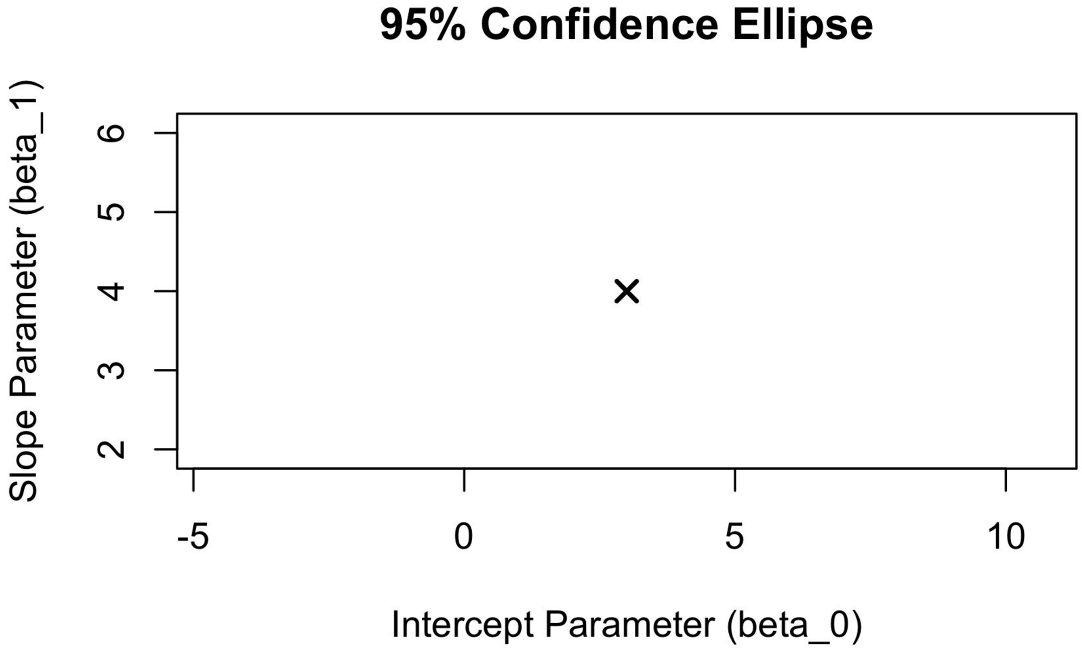
Burch et al. 2012. Metrika

Optimal designs

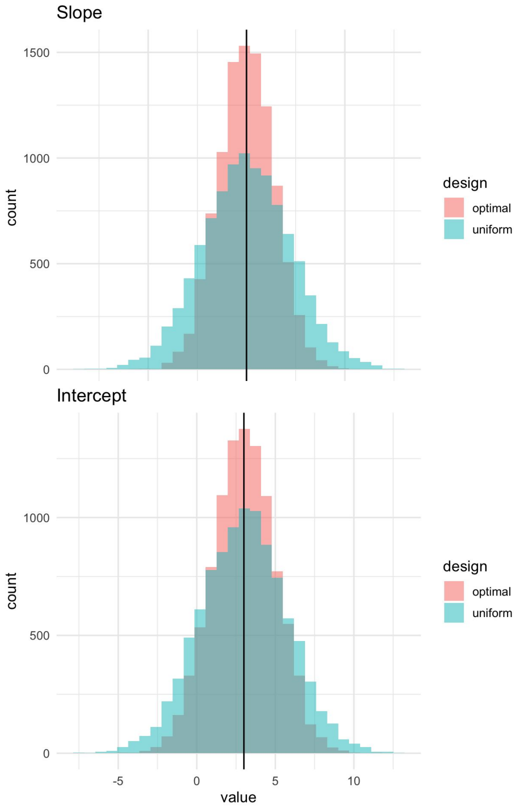
- The mathematically best design possible.
- What is the best design given:
  - objective of experiment : parameter fits vs predictive power.
  - a function
  - range of predictors
  - number of measurements
- Objective today: reduce parameter uncertainty.

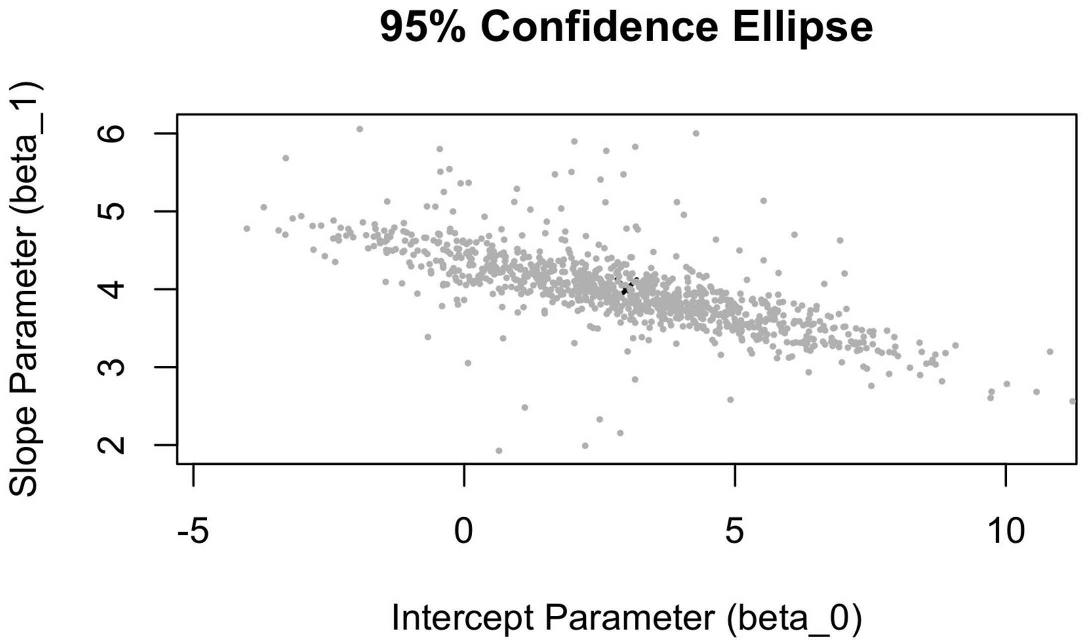
Confidence ellipses: a measure of parameter uncertainty



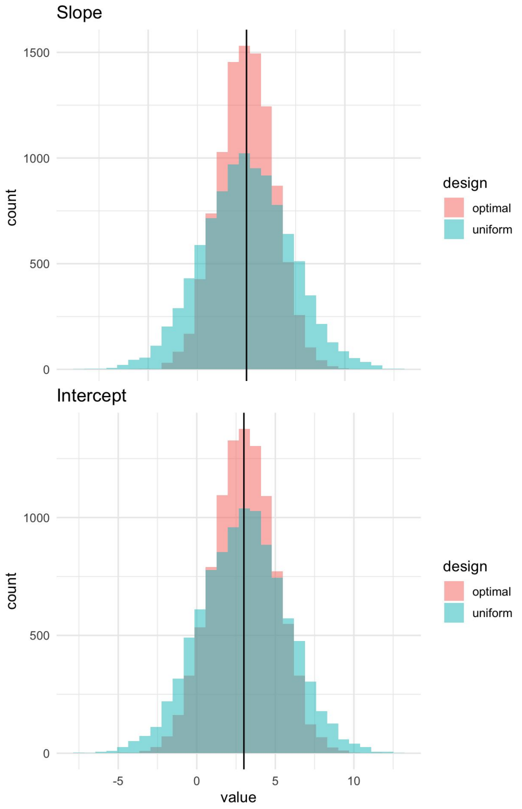


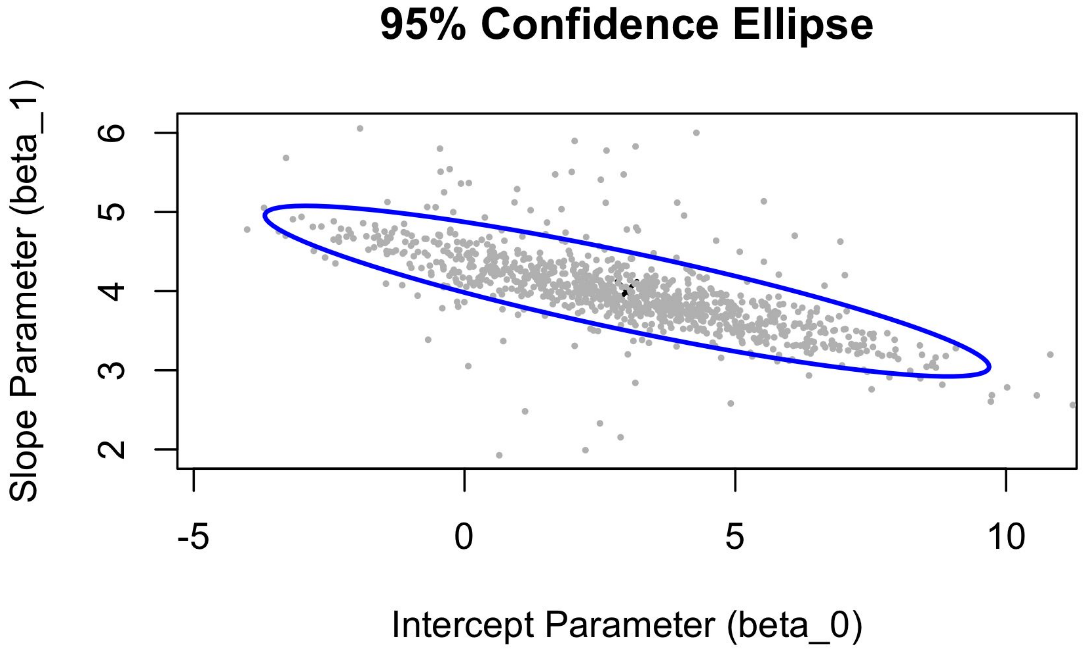
Confidence ellipses: a measure of parameter uncertainty





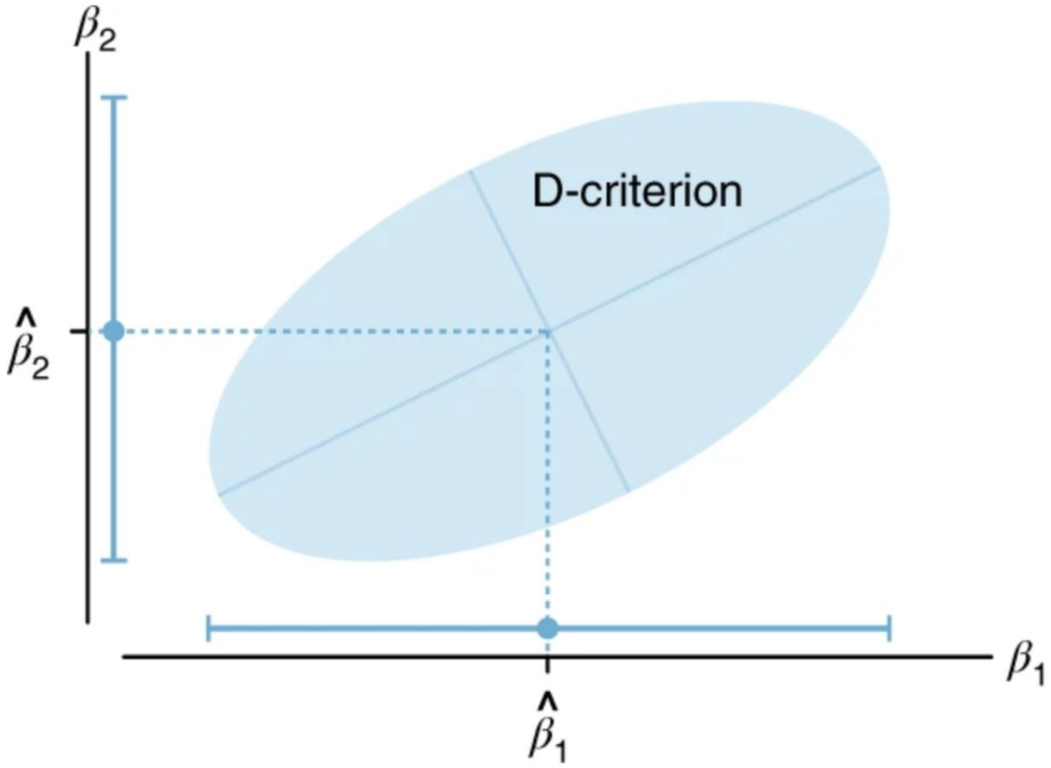
Confidence ellipses: a measure of parameter uncertainty





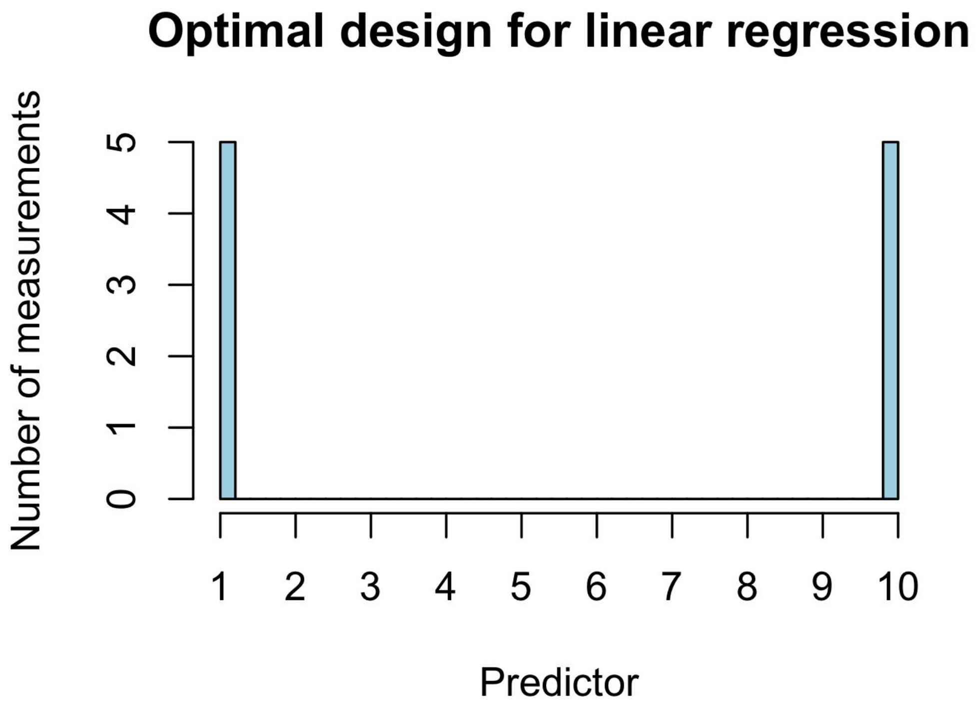
The D-optimality criterion maximizes the the information content of the parameter estimates.

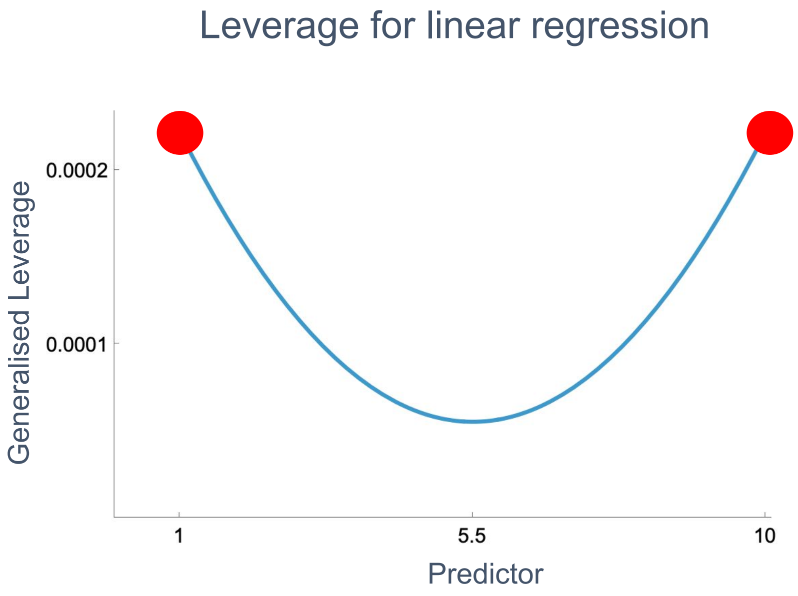
The confidence ellipse of a regression model with two parameters, β_1 and β_2.



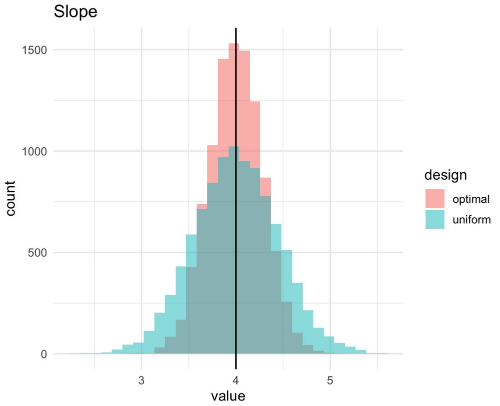
Smucker et al. 2018 Nature Methods

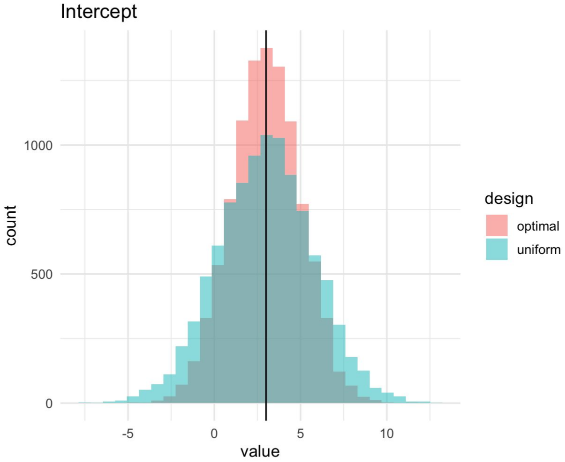
Optimal design: linear regression

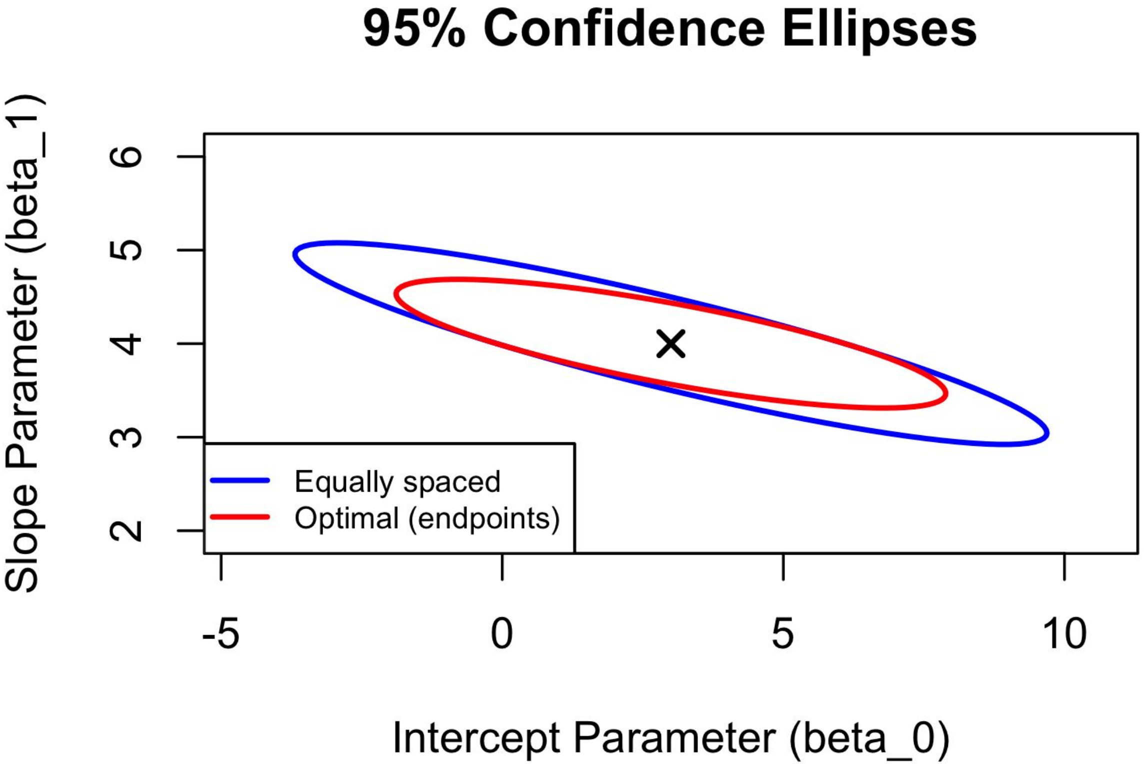




Optimal design: linear regression

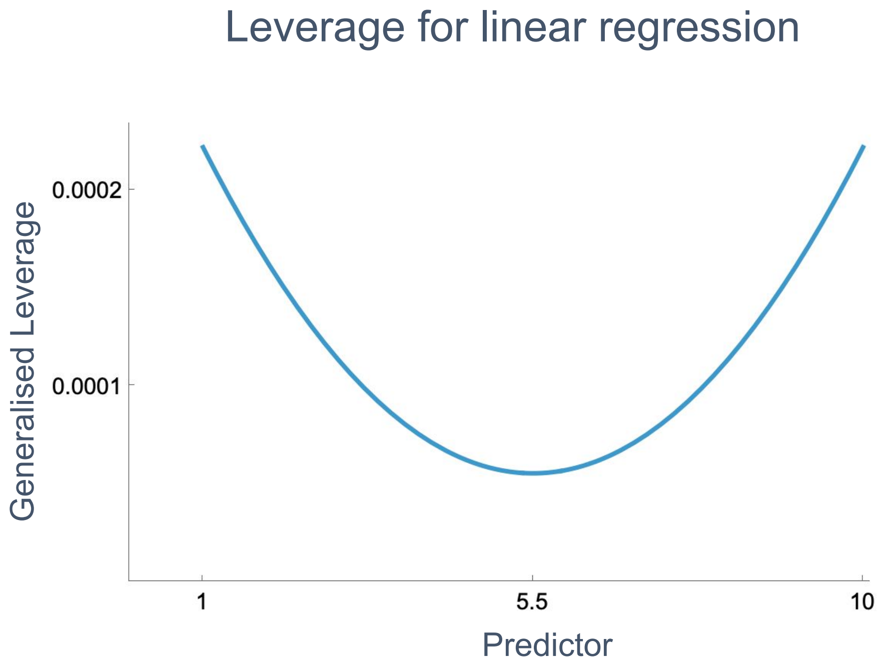






Should you put all the points at the extremes?

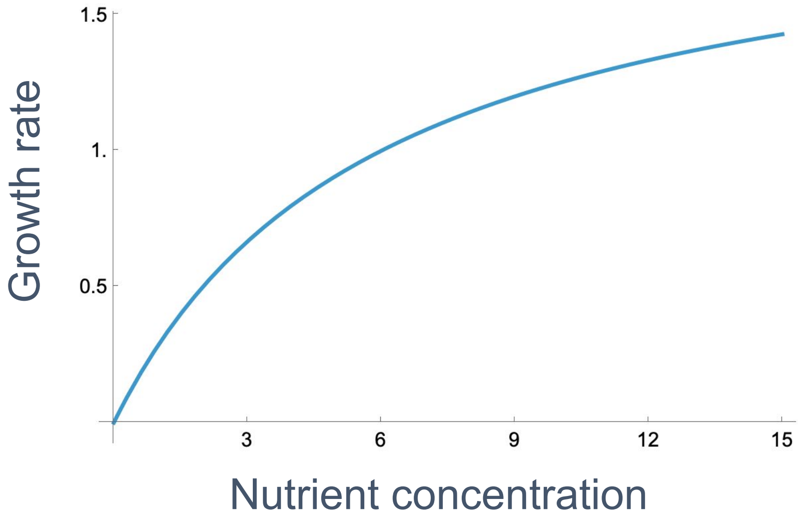
- No.
- Advantage:
  - Highest parameter precision possible
- Disadvantage:
  - No knowledge of the center
- What if the function is non-linear in the middle?
- Almost nothing is linear in ecology.



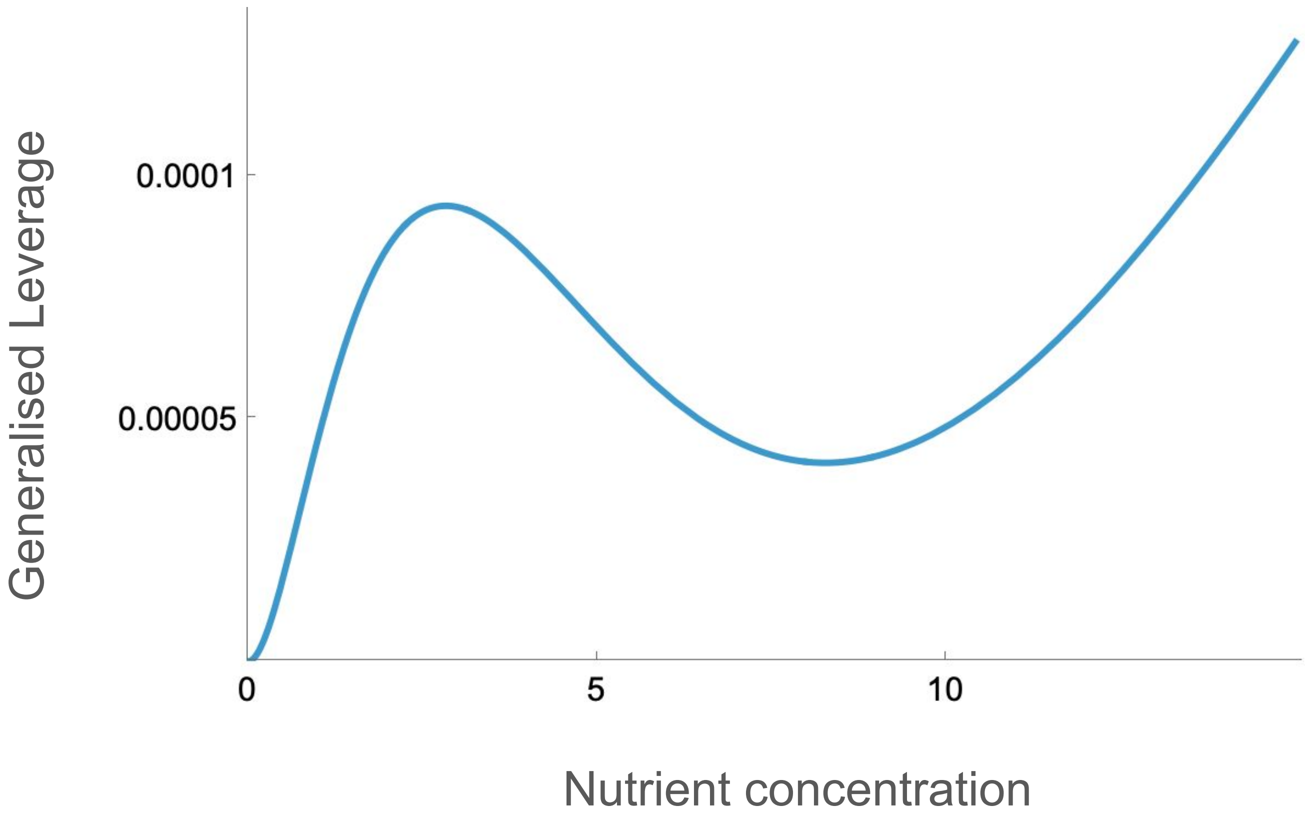
EXERCISE: Nonlinear regression
10 minutes

TAKE A SCREENSHOT NOW

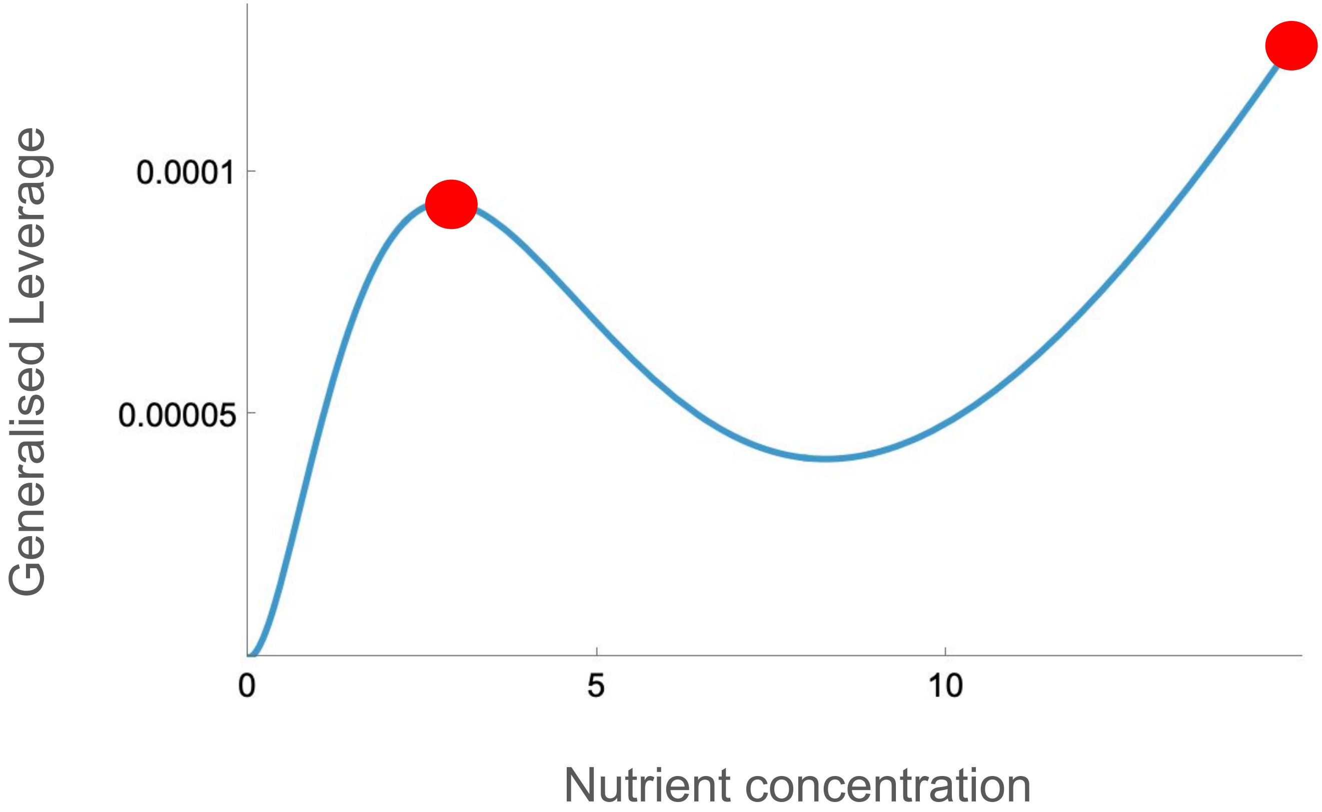
- Monod function: $\mu = \mu_{max} S/(K_S+S)$
- Parameters: $\mu_{max} = 2$, $K_S = 6$
- Nutrient range: 0-15
- Number of points: 16
- Your task:
    - Come up with 2 designs
    - Compare their precision using **R script #7**



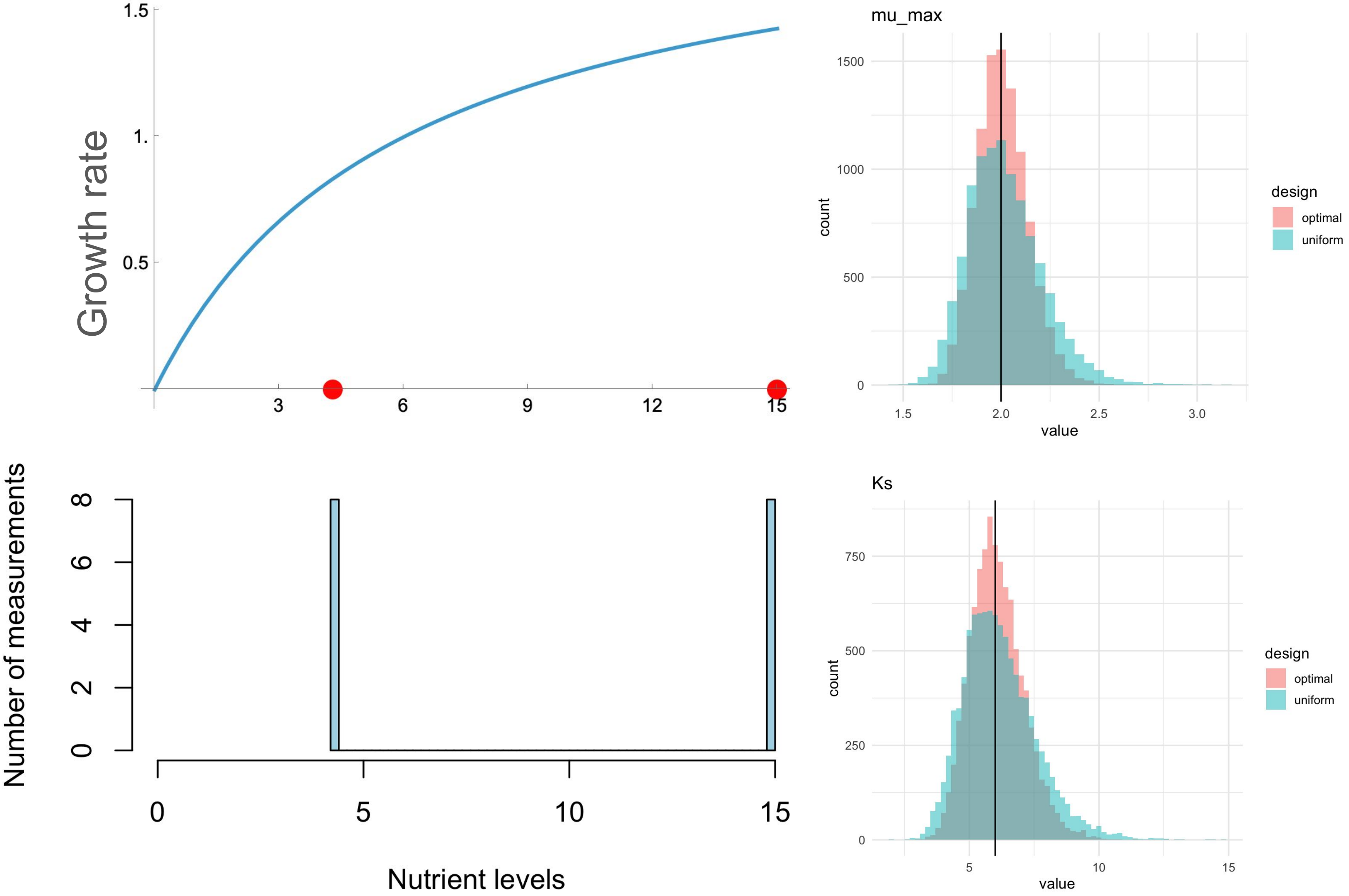
Where are the most valuable points for a Monod curve?



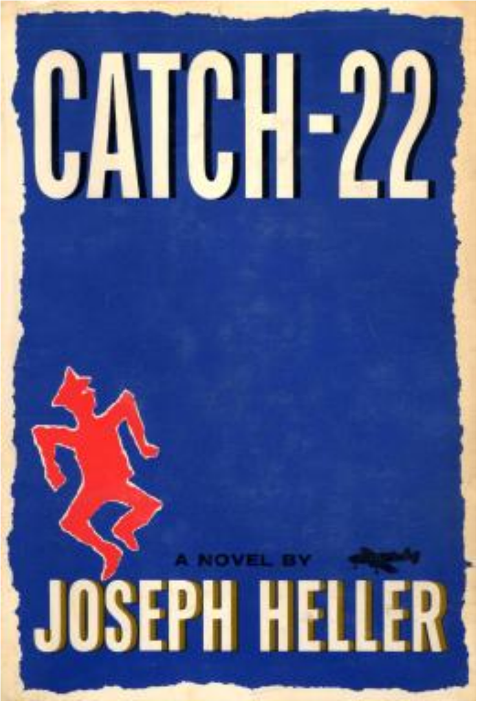
Where are the most valuable points for a Monod curve?



Optimal design: Monod

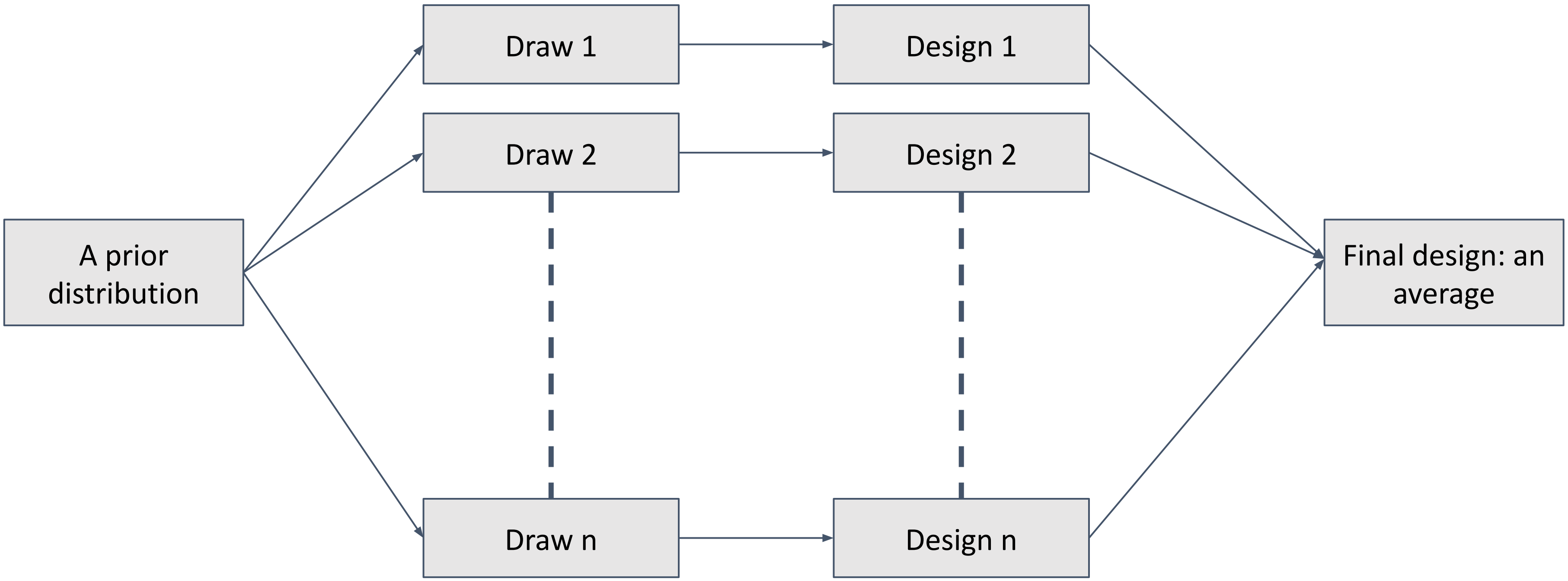


There are challenges.

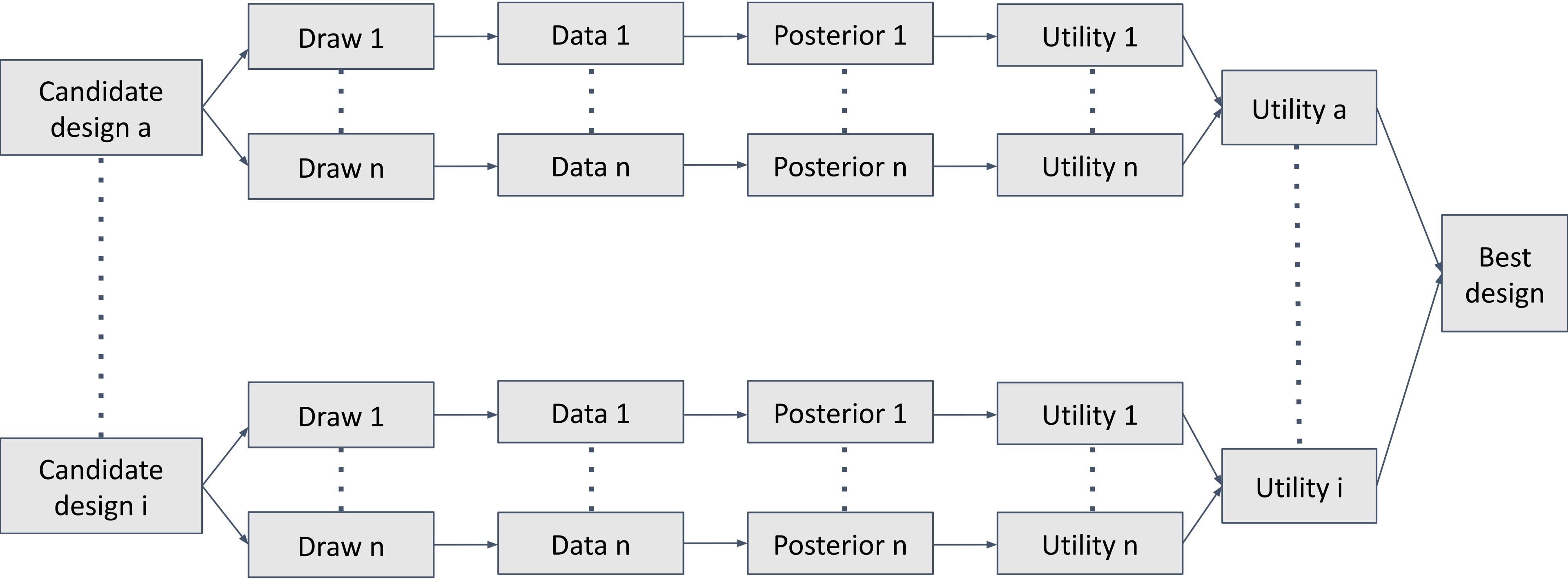


- For non-linear regressions, the optimal design depends on the parameter values.
- A Catch-22 arises:
  - the optimal experimental design cannot be derived without the parameters and
  - the parameters cannot be derived without the optimal design.
- What happens if you calculate a design with the wrong set of parameters?

Solution 1: pseudo-Bayesian design



Solution 2: fully Bayesian design



EXERCISE: Nonlinear regression

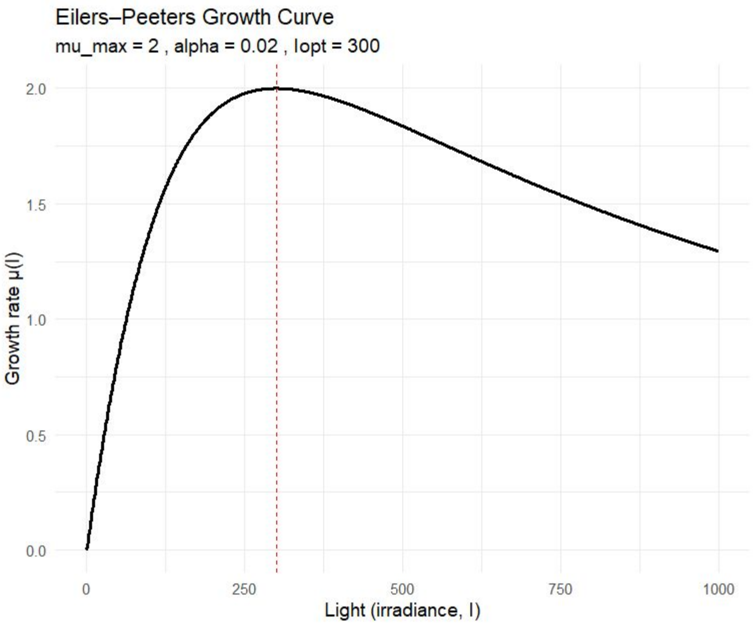
8 minutes

TAKE A SCREENSHOT NOW

- Eilers-Peeters function

$$\mu(I) = \frac{\mu_{\max} I}{\frac{\mu_{\max}}{\alpha I_{\mathrm{opt}}^2} I^2 + \left(1 - \frac{2\mu_{\max}}{\alpha I_{\mathrm{opt}}}\right) I + \frac{\mu_{\max}}{\alpha}}$$

- $\mu_{max} = 2$, $\alpha = 0.02$, Iopt = 300
- Light range: 0-1000
- Number of points: 10
- Your task:
    - Come up with 2 designs
    - Compare them using **R script #8**
    - **THEN in pairs, compare your designs**



Extensions that are easy to implement

1. Multiple regression
2. Mixed models (add more groups or individuals per group?)
3. GLMs like logistic regression and non-Gaussian errors

Basically: **if you can write the model equation, it’s easy to simulate**.

Conclusions

1. **Defining a clear, achievable, useful objective** is the most important thing you can do to improve your experiment.
2. Consider what **parameter**/quantity is most useful to estimate, and what **precision** you need.
3. Think carefully about your **model** and **error**.
4. When planning your experiment, **simulate** data from your model and **compare precision** under different designs.
5. Check **robustness** by changing assumptions and simulating again.

Conclusions

You don’t need to do an optimal design - but you can use simple simulations to do a *better* design.

Questions?

Thank you!

Mridul K. Thomas (mridul.thomas@unige.ch)

Ravi Ranjan (ravi.ranjan@utexas.edu)