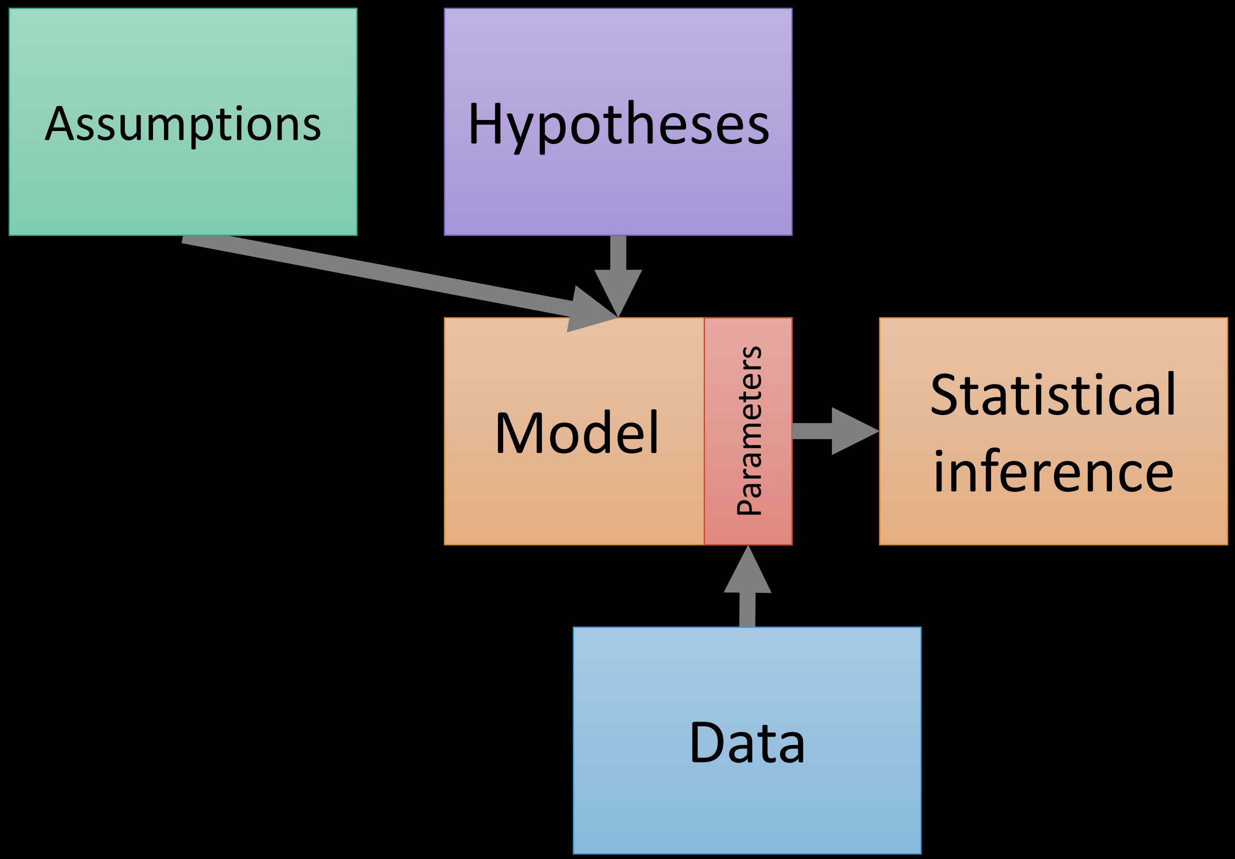
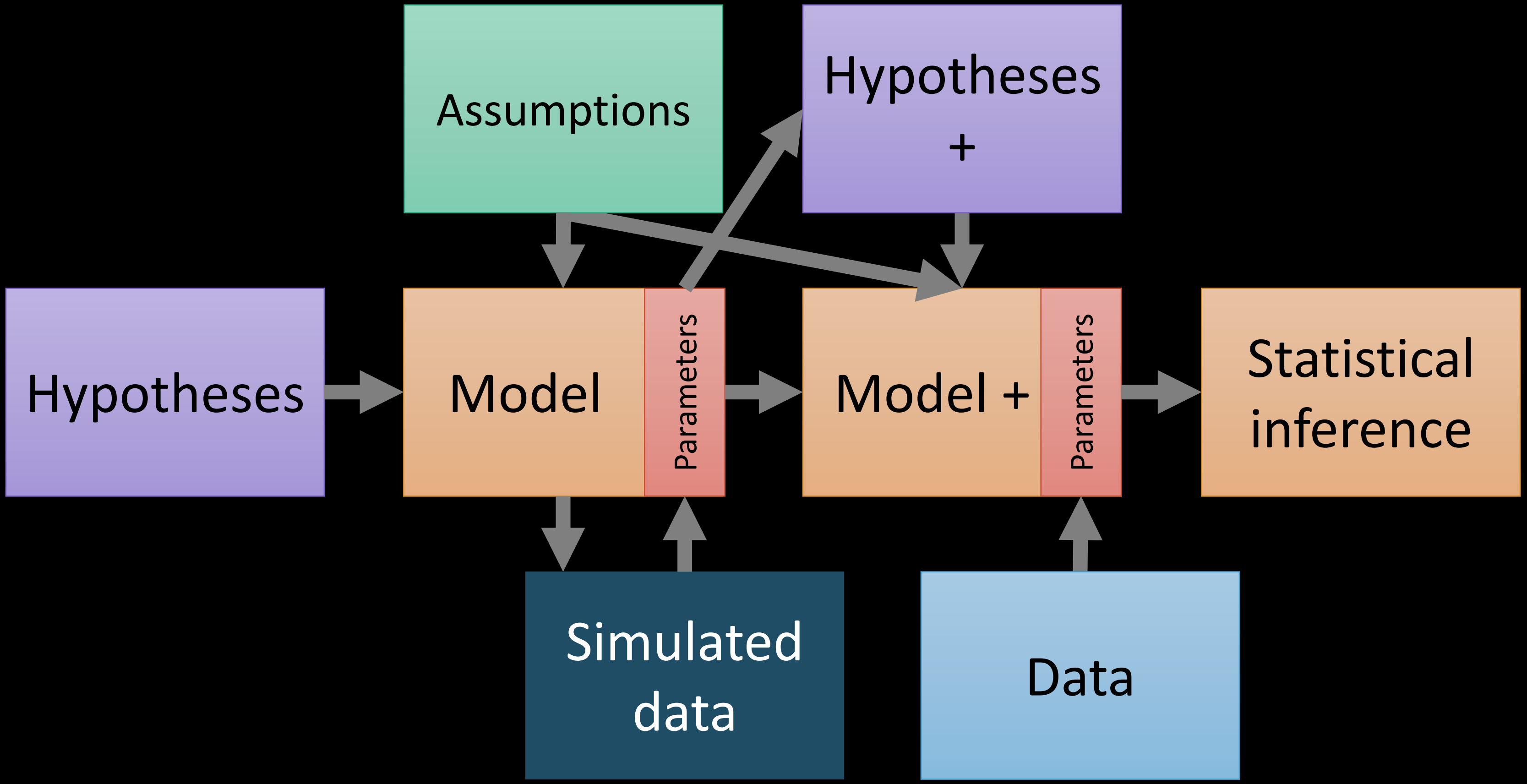
Simulating data: Theory

Without simulating data...



Why simulate data?...



Why simulate data?...

- Develop the model and hypotheses:
  - Useful for pre-registration/registered reports.
- Tests assumptions:
  - Do your stats tell you what you need to know?
  - Impossible analysis?
- Power analyses (when power cannot be computed directly):
  - Mixed-effects models.
  - Uncontrollable covariance between predictors.
  - Generalised linear models.

Why simulate data?...

- Develop model and hypotheses:
  - Useful for pre-registration/registered reports.
- Tests assumptions:
  - Do your stats tell you what you need to know?
  - Impossible analysis?
- Power analyses (when power cannot be computed directly):
  - Mixed-effects models.
  - Uncontrollable covariance between predictors.
  - Generalised linear models.

Reduce stress

What is a model?

Parametric equations

- Describe how the DV is <u>expected</u> to change in response to the IV.

Probability distributions

- 'Shapes' that describe the spread of data points… around <u>expectation</u>.
  - Noise or error.
  - Random effects.

"All models are wrong, but some are useful."
George Box

Choosing a parametric equation

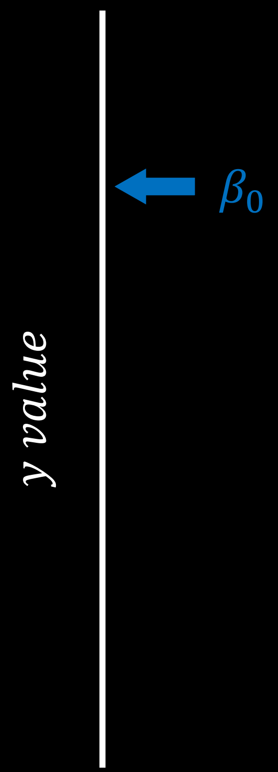
Expected relationship between variables:

- What do you think happens in the real world?
- What range of values can the IV and DV take?

Many options…

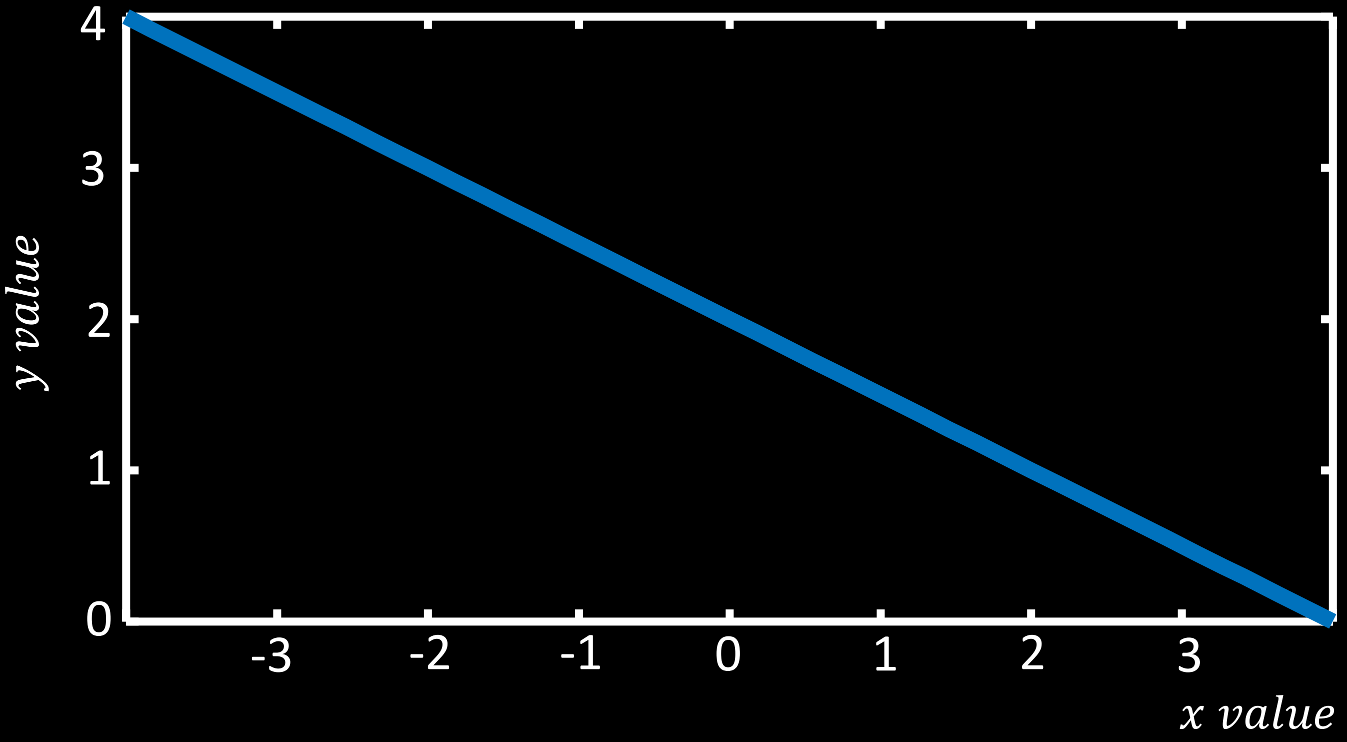
- Mean
- Linear
- Polynomial
- Exponential
- Sigmoid
- Power

Mean



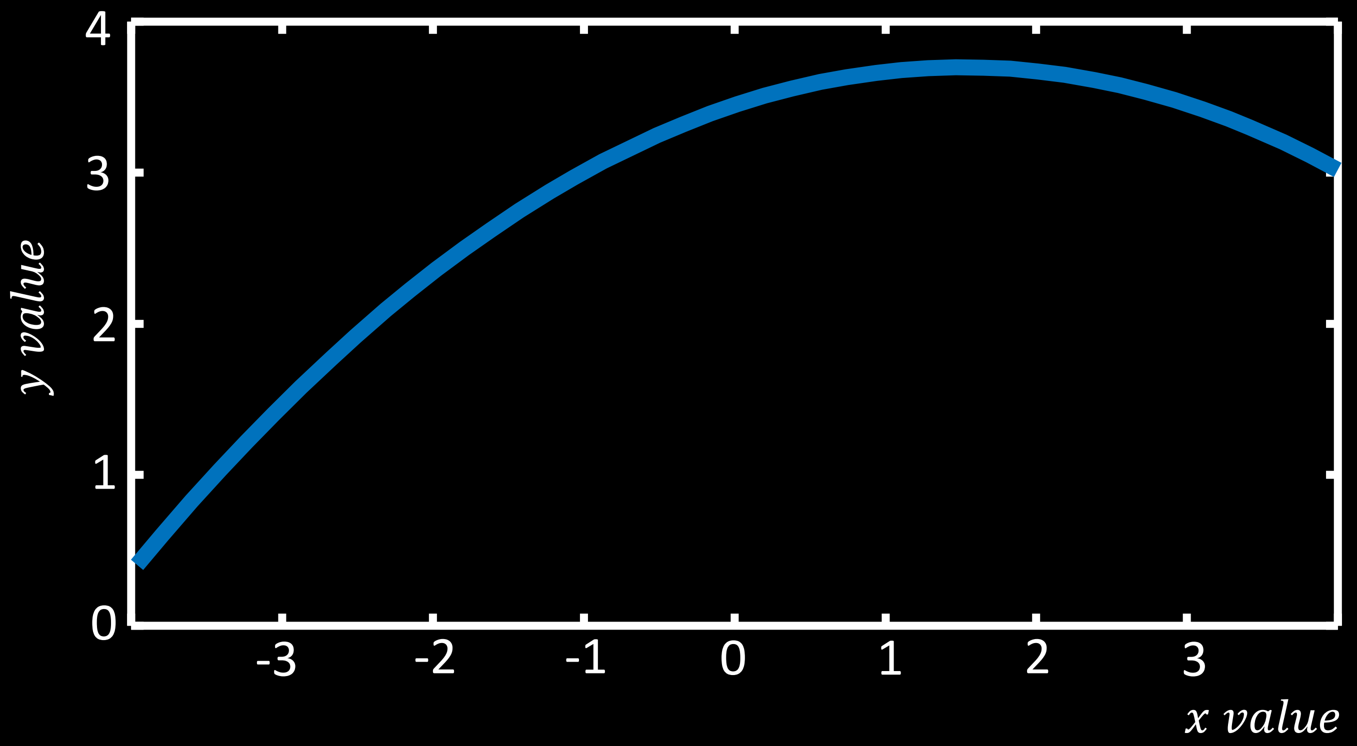
$$y = \beta_0 + \varepsilon$$

Linear



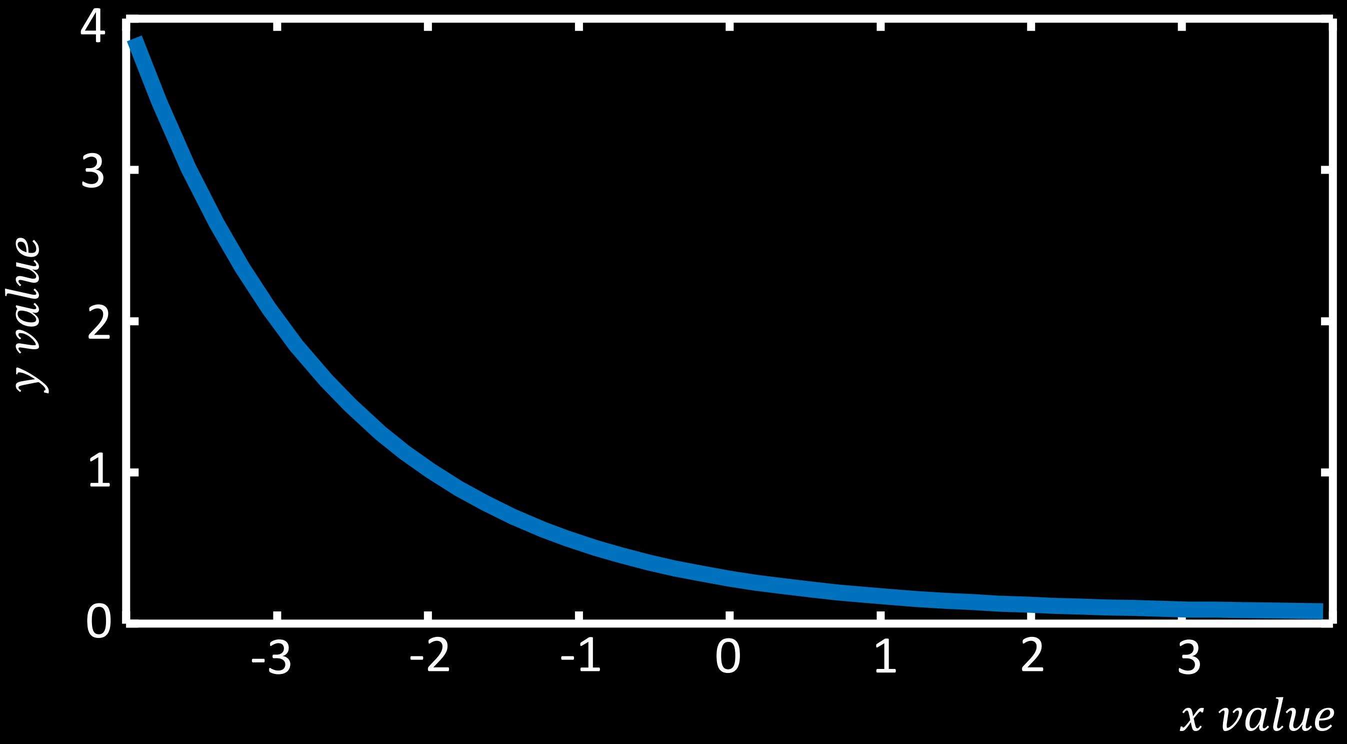
$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Polynomial



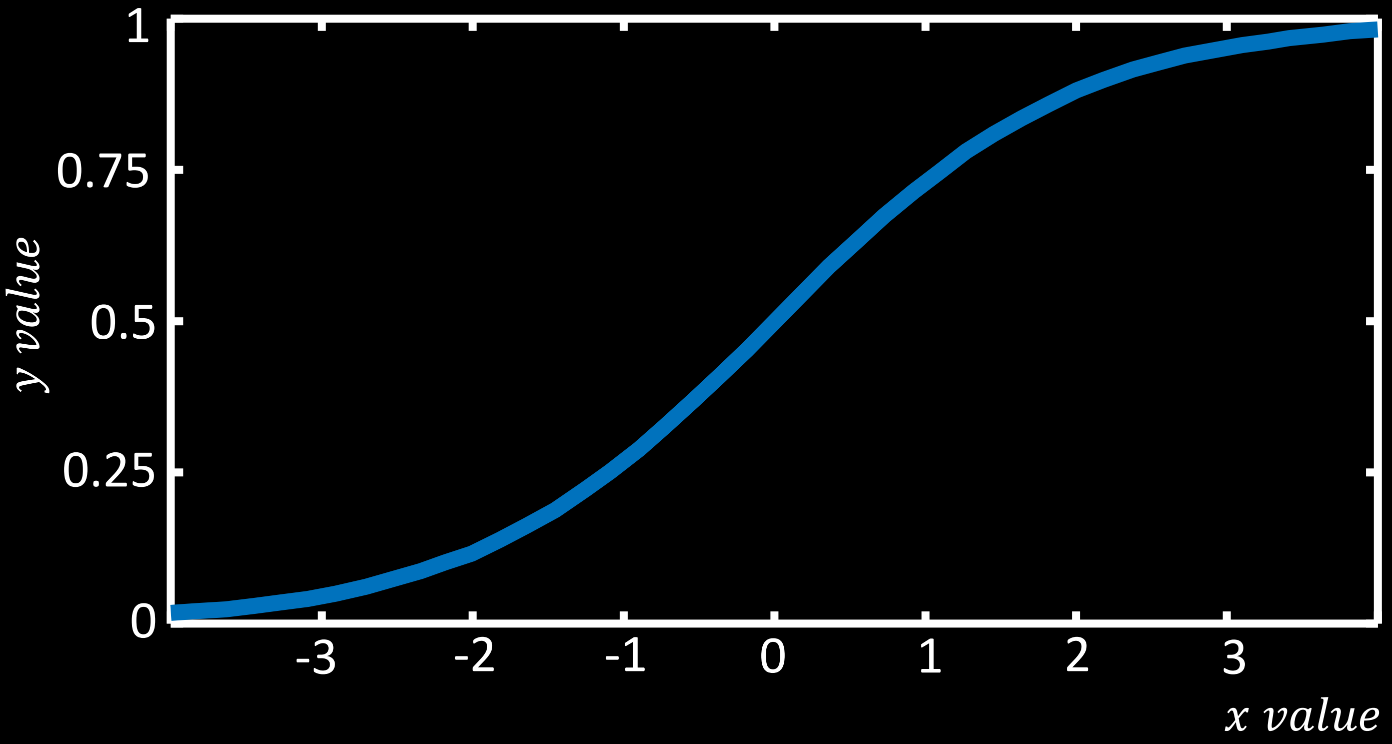
$$y = \beta_0 + \beta_{1,1}x_1 + \beta_{1,2}x_1^2 + \varepsilon$$

Exponential



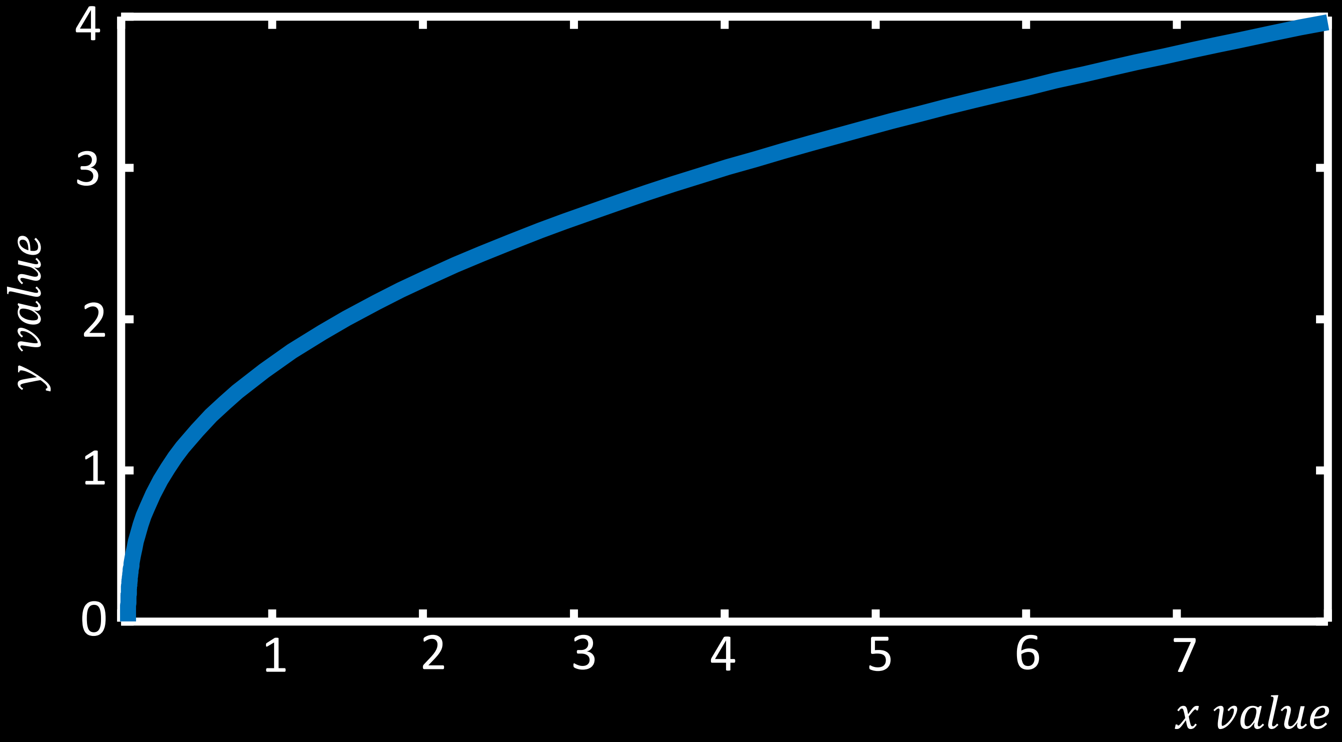
$$y = c + m \cdot \exp(\beta_0 + \beta_1 x_1) + \varepsilon$$

Sigmoid



$$y = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1))} + \varepsilon$$

Power (log-log analyses)



$$y = c + (\beta_0 + \beta_1 x_1)^m + \varepsilon$$

Choosing a probability distribution

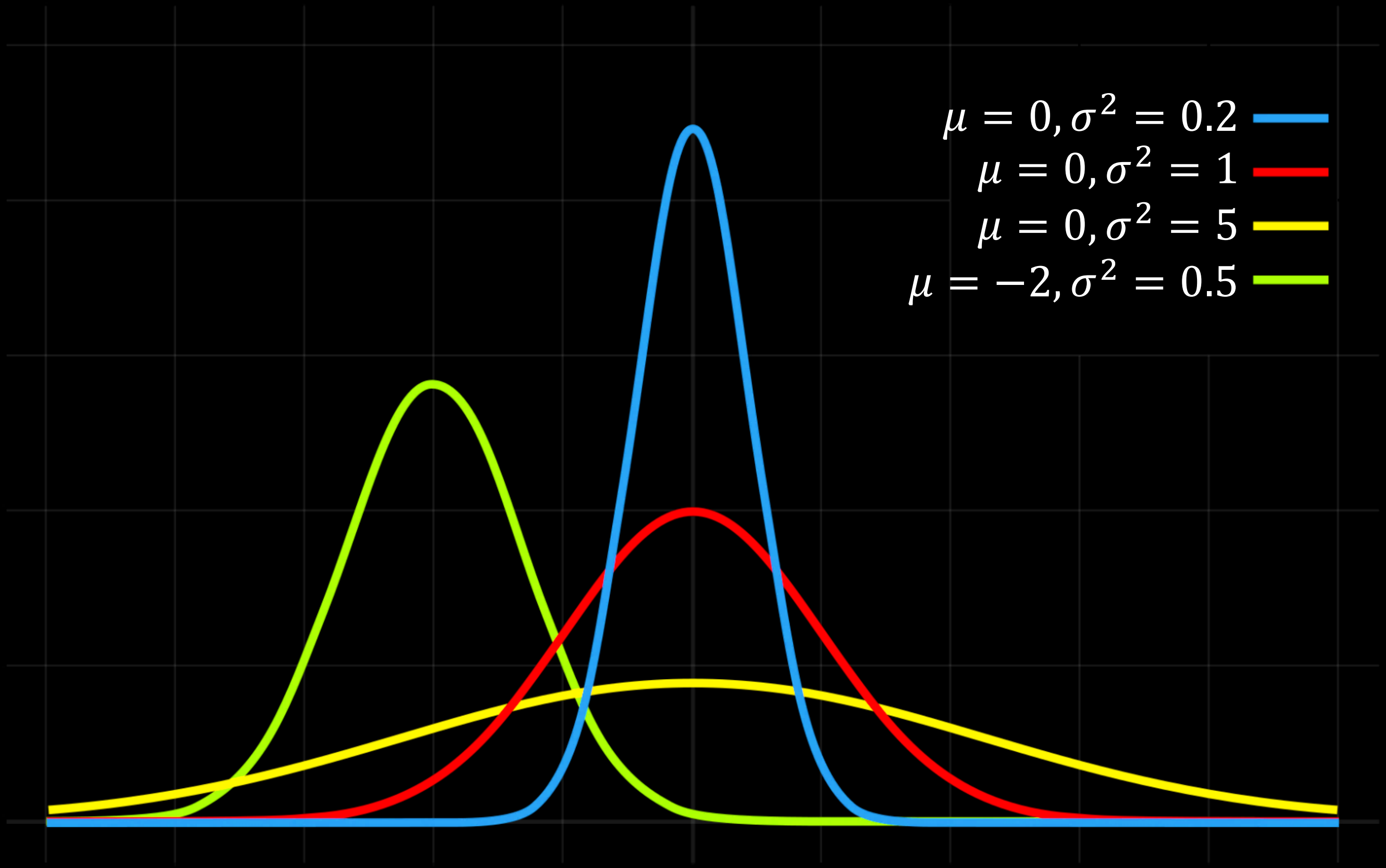
Spread of random factors around what is expected:

- What do you think happens in the real world?
- What range of values can the IV and DV take?

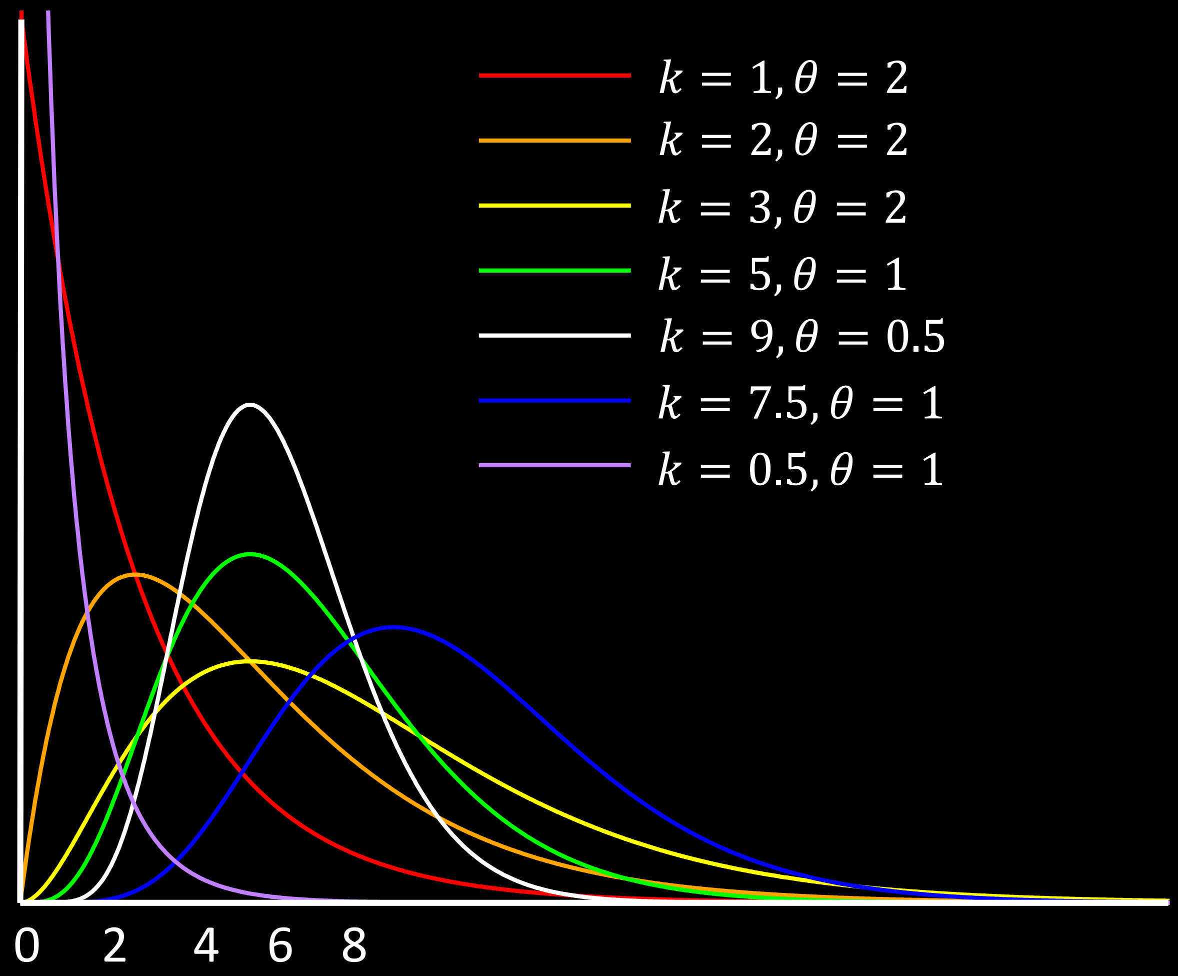
Many options…

- Normal
- Gamma
- Beta
- Etc…

Normal distribution



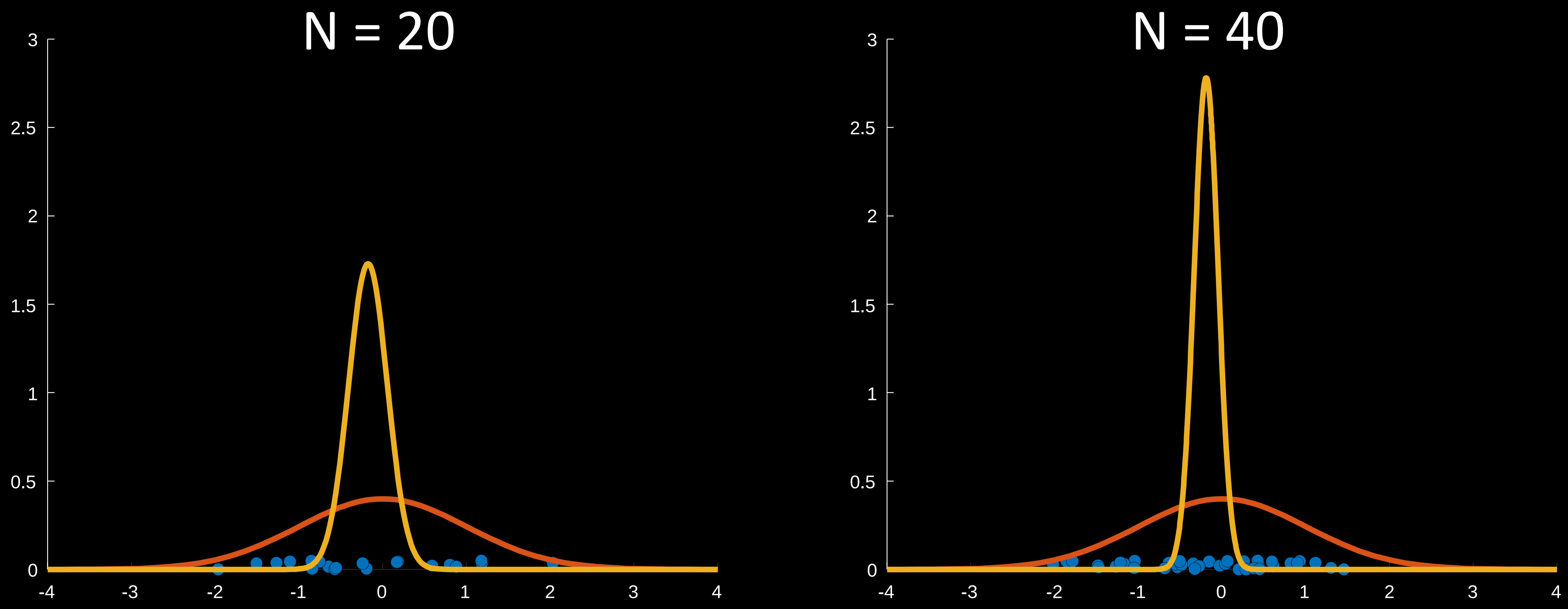
Gamma distribution



Beta distribution

Estimating power from simulations

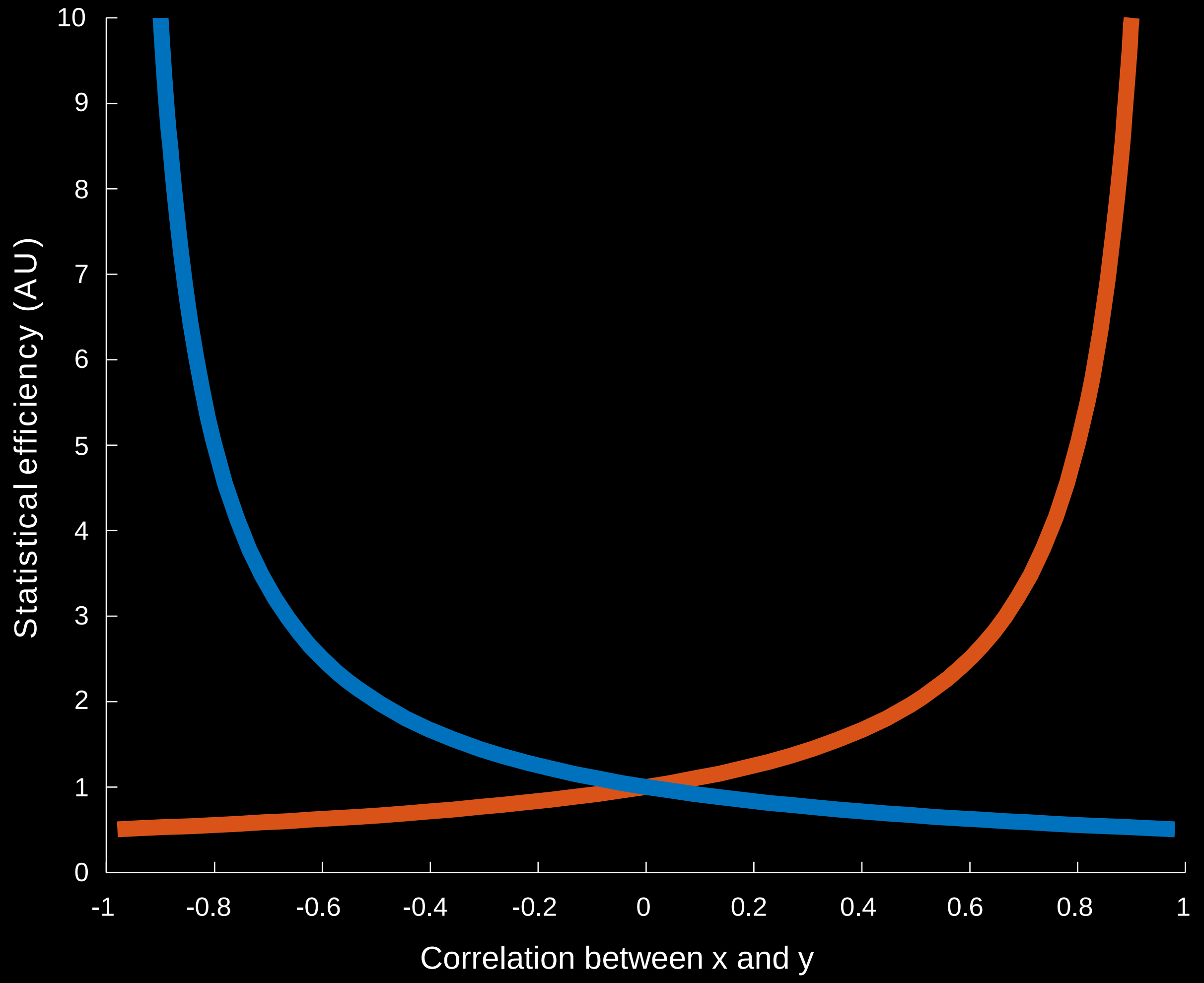
The sampling distribution



http://www.ltcconline.net/greenl/java/Statistics/clt/cltsimulation.html

Estimating power from simulations

Design efficiency



Efficiency finding a difference between X and Y.

Efficiency finding a difference between X+Y and zero.