Lecture 12: Analysis of model output

Complex Systems 530
3/12/20

Common tools for analyzing model output

- Visualization
- Descriptive statistics - means, standard deviations, etc. are often used to give a sense of the model output distributions
- Statistical analysis on model output - danger zone! This can be useful, but be careful

Visualization of ABM results

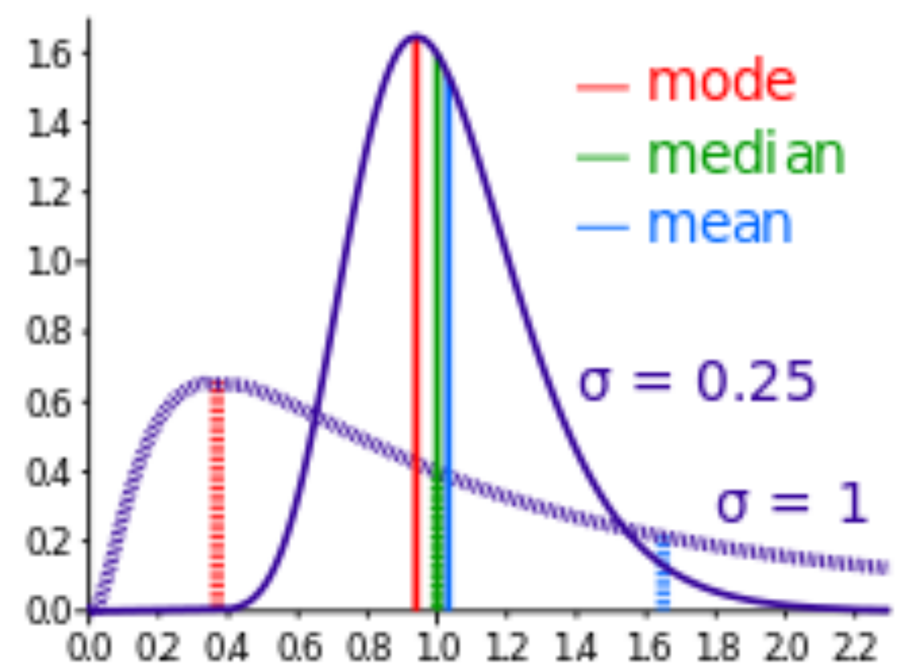
- Effective visual comparisons of model outcomes are often equally or more important and effective than statistical descriptions

- This is especially true when trying to capture “weird” behavior of a model or interesting qualitative features of its behavior

- Scatterplots, box-and-whisker, histograms, heat maps, etc.

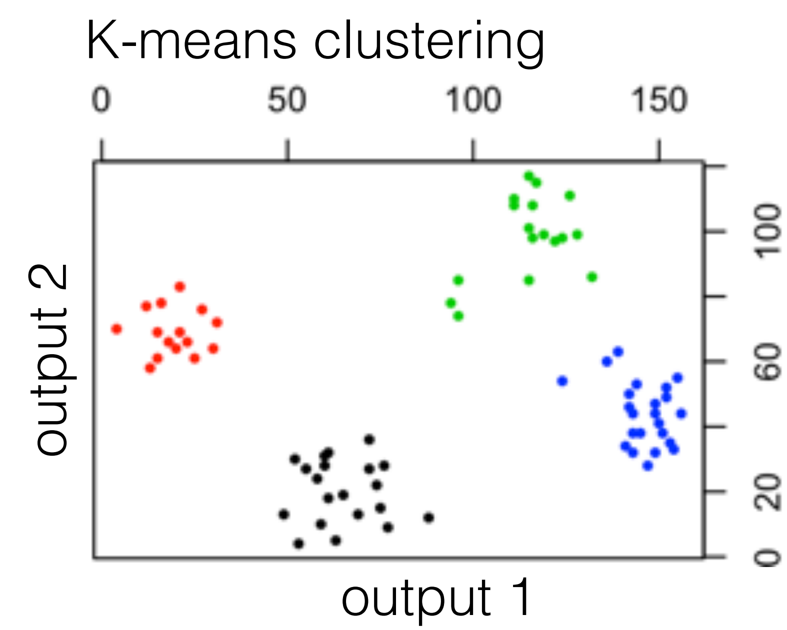
Descriptive statistics

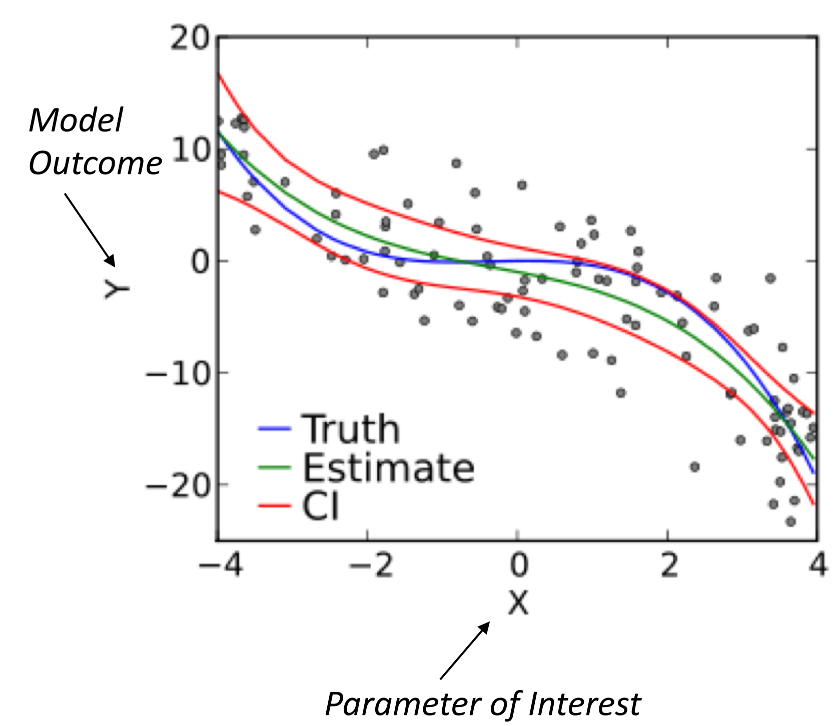
- Mean
- Standard deviation/variance
- Mode
- Median
- Proportions
- etc!



More advanced statistical analysis of ABM/model output

- Often will see use of:
  - t-tests, ANOVA, non-parametric versions of these, etc.
  - Analysis of trends, regression methods, clustering approaches, etc.





Statistical analysis on model runs/output

- Commonly used
- Be very careful with this! Think about what you are trying to do and what it means
- Often see p-values interpreted similarly to real-world data, sometimes/often in ways that don’t make sense
- Often papers will talk about an effect in the model being ‘significant’ as indicating that it has real-world impact, or that the effect is more ‘real’ if it is statistically significant

First, some definitions

- **P value**: the probability of obtaining test results at least as extreme as the results observed in the test, assuming that the null hypothesis is correct

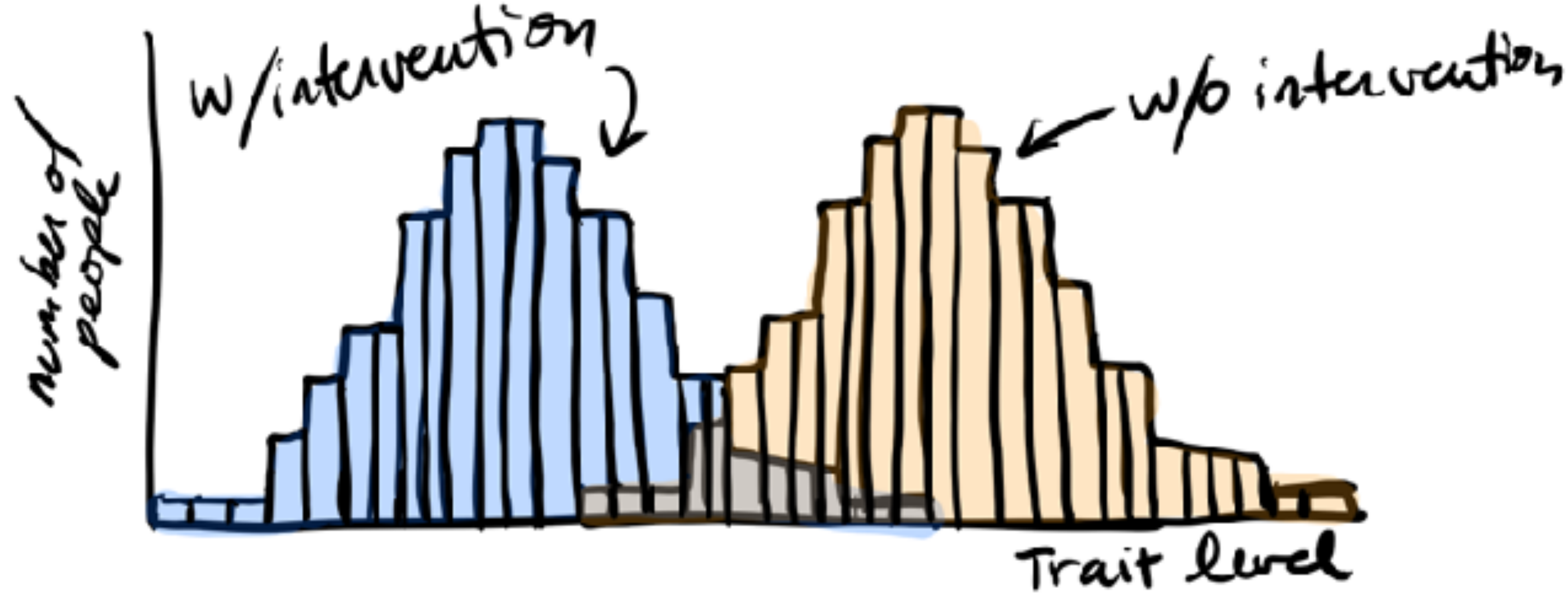
First, some definitions

- **Confidence interval (CI)**: a range of values which is likely to contain the population parameter of interest
  - E.g. a 95% CI means that if you sampled the same population and calculated the CI many times, it would contain the true parameter value (e.g. population mean) 95% of the time
  - Note this does not mean that 95% of the population is in this interval!
  - (Sidenote: using frequentist versions of confidence bounds here but we'll explore Bayesian versions later on!)

Example: measuring & analyzing real world data

- Suppose we have a group of people (or cells, etc.) with some trait (e.g. cholesterol levels, gene exp.)

- Do an intervention/change conditions and then re-measure the trait (or test two groups of individuals, etc.)



Example: measuring & analyzing real world data

- One common analysis approach: the t-test! (e.g. a paired t-test)
    - Note assumptions of independence, normality
- Null hypothesis: the two means are equal (difference between means = 0)

Example: measuring & analyzing real world data

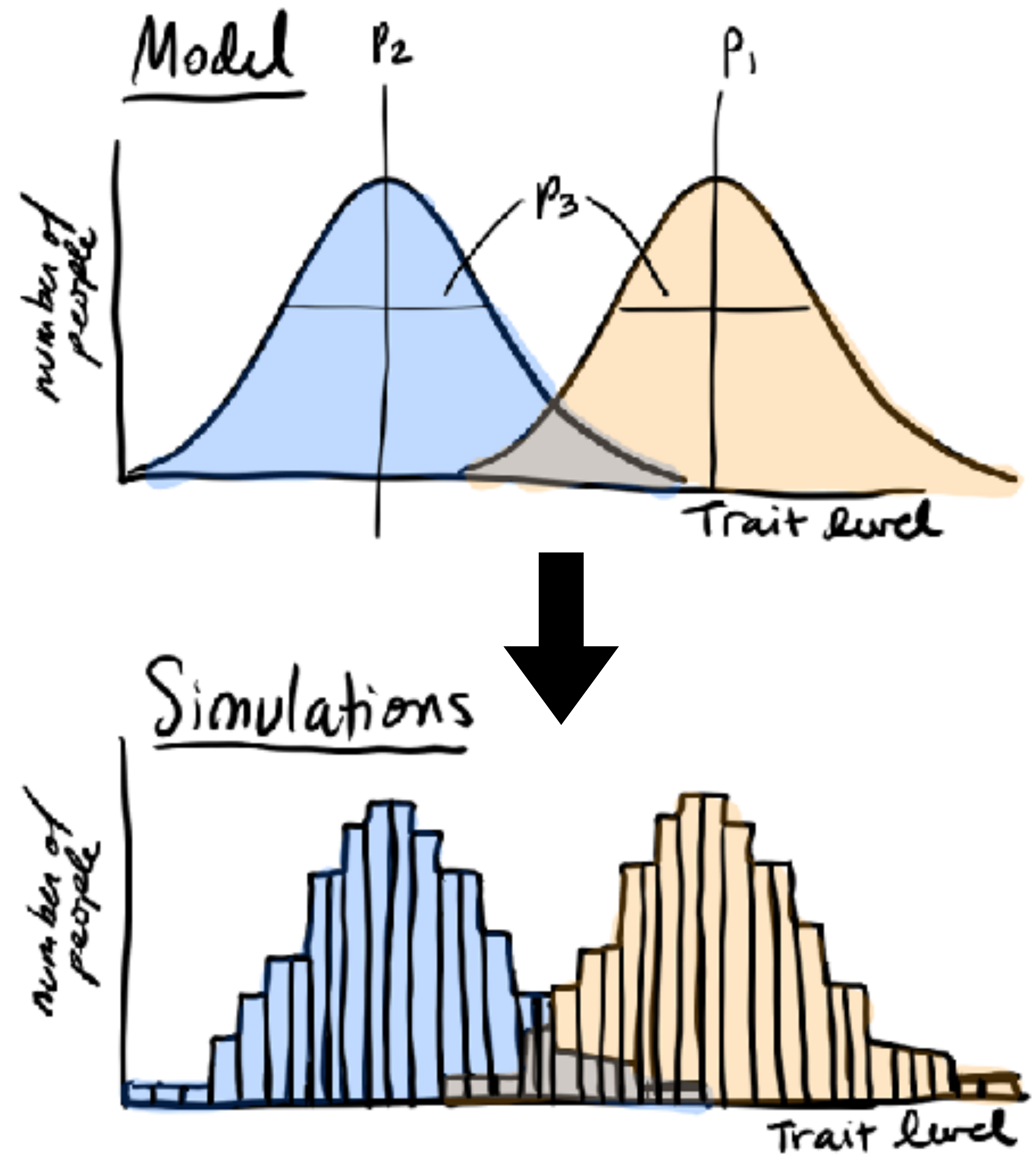
- Results are often reported by giving the two means and a p-value
- What does this p-value mean?
- Why do we need a p-value or some other measure of uncertainty?

Example: modeling the system

- Now suppose we model this system as follows:
  - Each individual or agent is modeled as having a trait value, which is drawn from a normal distribution, with mean p_1 and variance p_3
  - After the intervention, we each individual redraws their trait value from a new normal distribution, with mean p_2 and variance p_3

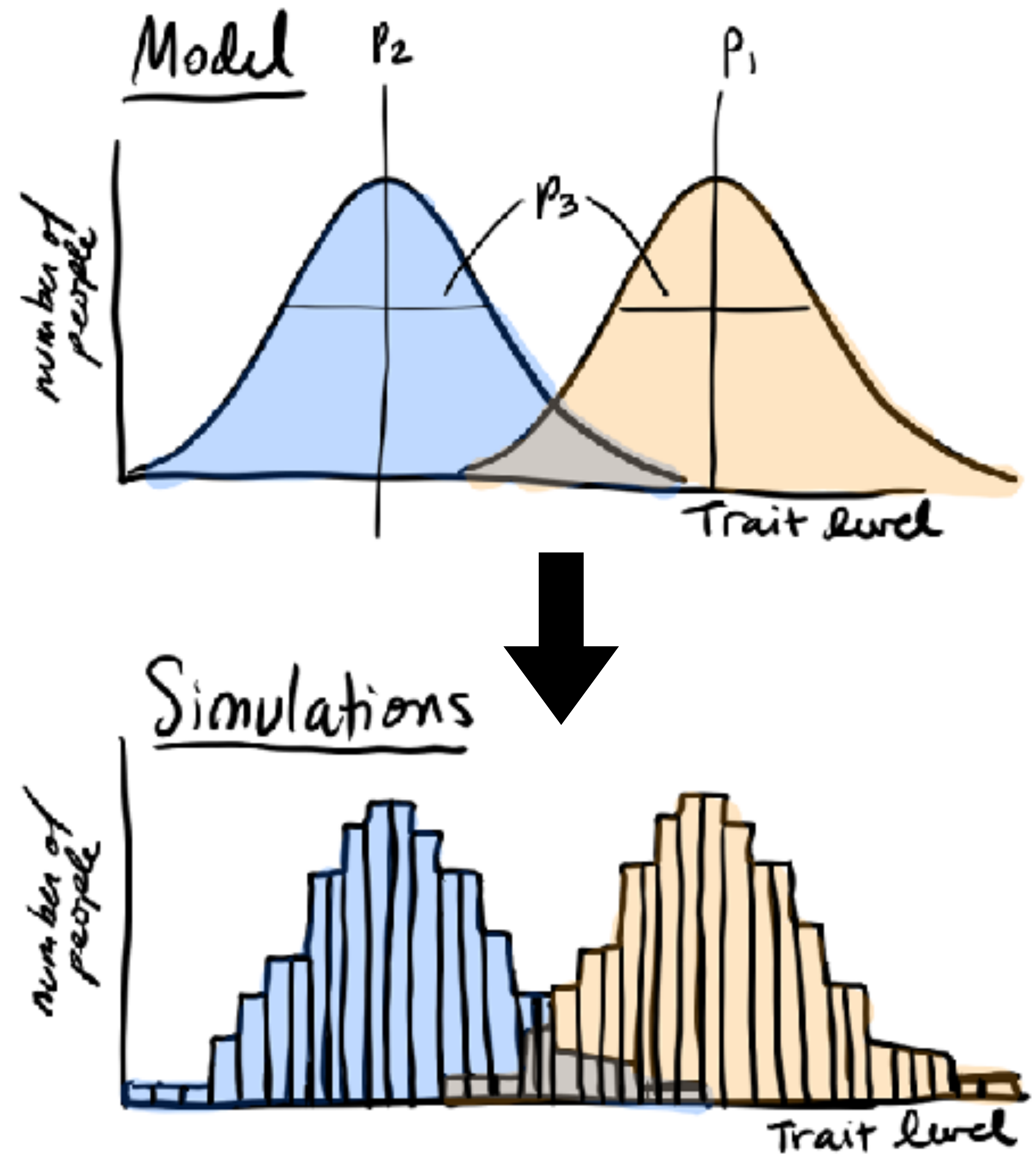
Example: modeling the system

- We want to understand what our model says about how effective this intervention is
- Suppose we generate simulations from our model, and we sweep over values of p_2 (e.g. from $p_2 = p_1$ to $p_2 = 0$)
- So, say we test the set of values: p2 = {0, 0.1p_1, 0.2 p_1, …, 0.9 p_1, p_1}



Example: modeling the system

- How should we analyze this model output data?
- What would a t-test tell us?
- What would the p-value mean?
- Is there uncertainty in whether the means are different?
- How will the results change if we change the number of simulations we do?



Example: modeling the system

- Let’s try it out!

Example: modeling the system

- What if we had a more complex model of this system, where a person/agent's trait value was determined by a more detailed, mechanistic model of the processes involved?
- Would there be uncertainty in whether there is an effect of changing particular parameters?
- What sources of uncertainty might there be?

Potential issues with statistical analyses of model output

- **Model misspecification** (discuss fitting a linear trend to model output data)

- **Interpretation** - statistical significance from model output often doesn't tell us anything about the actual effect "significance" in the real world

- Others?

Potential issues with statistical analyses of model output

- Typically, the p-values are telling us more about whether we have taken sufficiently many numerical samples (i.e. model runs) to observe the difference between conditions, and less about whether there actually is one

- If you run 1000 simulations and get a p-value of 0.06, you can often run millions of simulations and get the p-value as small as you want! Sample size issues mean something different here

Potential issues with statistical analyses of model output

- Sometimes there actually won't be a difference for a particular output as you change conditions (talk about an example model for this)—but usually you don't need stats analysis to tell you this (if you understand your model)

Potential issues with statistical analyses of model output

- Effect sizes, slopes, etc. are often more informative
- Confidence intervals can help to get a sense of the distribution (although similar issues to p-values apply)

Uses for statistical tests of model output

- Descriptive stats to understand and highlight model behavior—more advanced stats on model output can be useful, but remember what they mean
- Evaluate uncertainty due to finite number of numerical simulations
- Simulate the data collection process in the real world, and see what standard statistical models would say if the real world was the model.

Quick note

- Statistical analysis of model outputs/simulations
  vs.
  Using a model as a statistical model to connect with real world data (parameter estimation)

- These are different! We will talk more about parameter estimation in the next couple weeks

Epistemic vs. aleatory uncertainty

- **Aleatory uncertainty** - uncertainty due to randomness/stochasticity/error, etc. Can be process-related noise, e.g. stochastic mixing of indivdiuals in a population, measurement error, e.g. variation in a given assay, etc. (comes from the Latin *alea*, meaning dice)/

- **Epistemic uncertainty** - systematic or model uncertainty, often characterized as things we could in principle know but do not—e.g. we might not know the true underlying process for a system, resulting in multiple alternate models (or potentially a need to measure further)