Hierarchical Bayesian Analysis

LSST Discovery Fellowship Program Day 4

Greg Gilbert | LSST Discovery Workshop | 23 May 2025

Modeling choices

Physical
- What processes do you include?
- What approximations do you make?

Statistical
- Are data i.i.d.?
- Is there correlated noise?
- Do you account for data collection?

Model specification
- Parameterization
- Priors
- Convergence criteria

Sampler
- Grid search
- Maximum likelihood
- Markov Chain Monte Carlo
- Nested Sampling

Bayes Theorem

Likelihood

Prior

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)}$$

Posterior

Evidence

Hierarchical Bayesian Modelings self-consistently **modifies the prior**

Hierarchical Bayes Theorem

Parameters

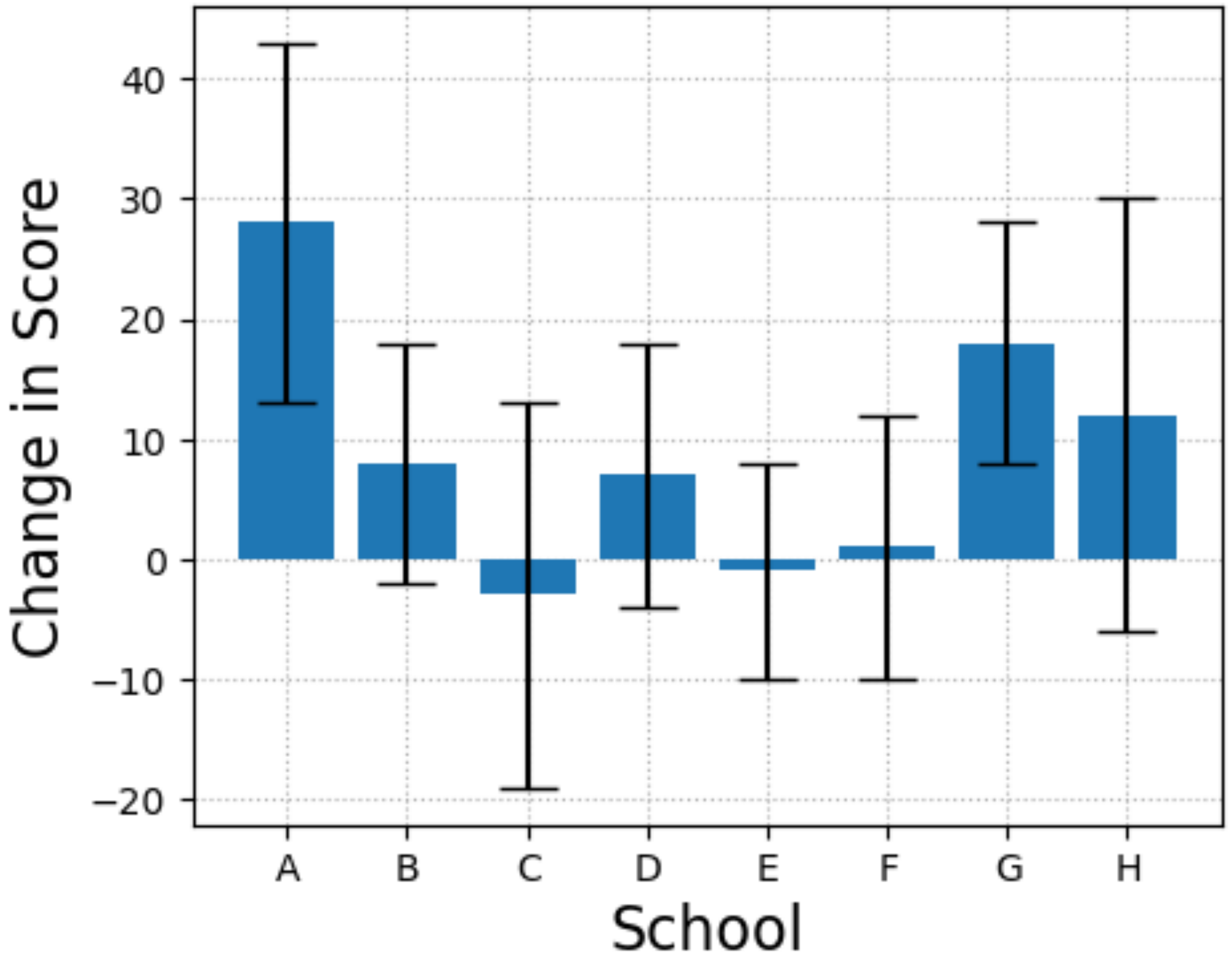
Hyper-parameters

$$P(\theta, \alpha \mid D) = \frac{P(D \mid \theta)P(\theta \mid \alpha)P(\alpha)}{P(D)}$$

The prior can be thought of as the **population-level distribution**

The Eight Schools Problem

The set-up: students from eight schools have participated in a test-prep program. The mean score improvement ΔS and uncertainty on the mean σ_μ for each school are recorded.

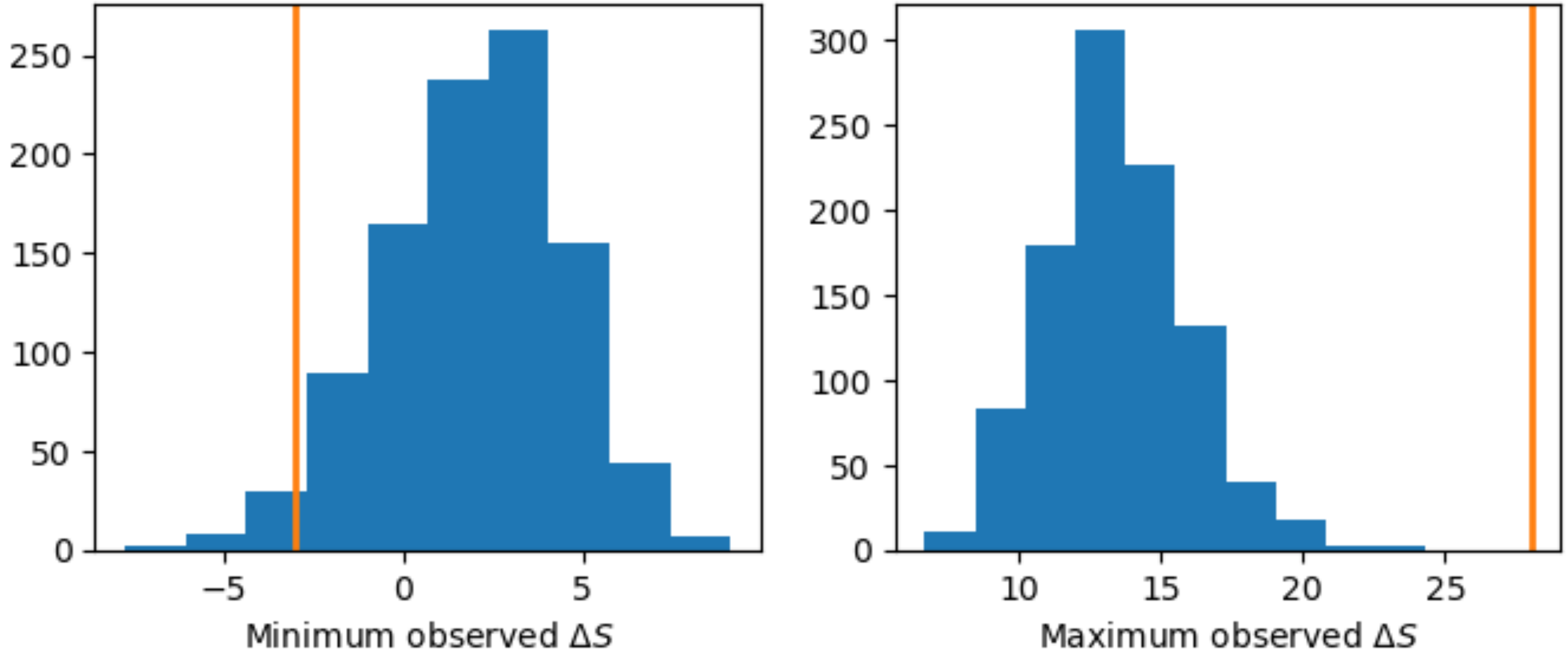


$$\mu = [28, 8, -3, 7, -1, 1, 18, 12]$$
$$\sigma_\mu = [15, 10, 16, 11, 9, 11, 10, 18]$$

Question: can the measured effect size for School A (28 pts) be attributed to the test-prep program?

The Eight Schools Problem

Question: can the measured effect size for School A (28 pts) be attributed to the test-prep program?



Running 1000 bootstrap trials suggests NO

Modeling Options

Independent: Each school is analyzed separately

Pooled: All schools are analyzed in one group

Hierarchically: The relationships between groups are considered

Independent

μ = [28, 8, -3, 7, -1, 1, 18, 12]

σ_μ = [15, 10, 16, 11, 9, 11, 10, 18]

Pooled

$$\Delta S = 7.7 \pm 4.1$$

Hierarchical Eight School Model

$\alpha_\mu \sim \text{Normal}(\mu, \sigma)$

$\alpha_\sigma \sim \text{Half-Cauchy}(\beta)$

The hyper-parameters $\alpha \equiv \{\alpha_\mu, \alpha_\sigma\}$ describe the population distribution (i.e. the prior)

$\Delta S_i \sim \text{Normal}(\alpha_\mu, \alpha_\sigma)$

Each school's $\theta_i \equiv \{\Delta S\}_i$ is drawn from a Gaussian described by α

We will simultaneously and self-consistently infer the population hyper-parameters α and the individual member values θ

Hierarchical Eight School Model