# Misc Review

## Personal review guide: resources + conceptual progression for this assignment

Below is a concise roadmap of **what to review** and **how the ideas develop logically** across Q1–Q3. The goal is to reinforce both the statistical theory and the modeling workflow that underlies your draft, not just the specific calculations.

---

## 1. Key topics to review (beyond the textbook chapters you already used)

### Hierarchical / multilevel modeling (Bayesian perspective)

Focus on: - exchangeability - partial pooling / shrinkage - population vs unit-level inference - predictive distributions

Good references: - Gelman et al., *Bayesian Data Analysis (BDA3/BDA4)*
- Ch. 5–7: hierarchical models, shrinkage, partial pooling
- Ch. 14–15: posterior predictive checks and model criticism
- McElreath, *Statistical Rethinking* (very intuitive treatment of hierarchical thinking)

Why this matters here:
You repeatedly reason about whether inference targets

$$p(\beta_i \mid y) \quad \text{or} \quad p(\lambda, \tau^2 \mid y),$$

i.e., **unit vs population parameters** (Q1–Q2).

---

### Mixture models and marginal likelihood thinking

Review: - integrating out random effects - predictive distributions - exchangeability arguments

Good references: - Hoff, *A First Course in Bayesian Statistical Methods* - Bernardo & Smith, *Bayesian Theory* (conceptual foundations)

Why:
Helps justify the mixture interpretation

$$f(y \mid \lambda) = \int f(y \mid \theta) g(\theta \mid \lambda)\, d\theta,$$

which underlies your Q1 conclusion.

---

#### Variance–mean relationships / heteroscedastic regression

This is **directly relevant to Q3**.

Review: - modeling $\mathrm{Var}(Y \mid X)$ as a function of $\mu$ - variance-stabilizing transformations - weighted least squares - variance functions in GLMs

Good references: - Carroll & Ruppert, *Transformation and Weighting in Regression* - McCullagh & Nelder, *Generalized Linear Models* - Pregibon's work on parameterized link functions

Why:
Your likelihood assumes

$$\mathrm{Var}(Y_{i,j} \mid \cdot) = \sigma_i^2 \mu_{i,j}^{2\theta},$$

so $\theta$ is literally a **variance-function parameter**.
Q3 is essentially a variance-function adequacy check.

---

#### Model checking and diagnostics

Review: - posterior predictive checks - residual plots for hierarchical models - graphical diagnostics vs formal tests - sensitivity analysis

Good references: - Gelman et al., BDA: posterior predictive checking chapter - Cook & Weisberg, *Residuals and Influence in Regression*

Why:
Q3 is fundamentally **model criticism**, not parameter estimation.

---

## 2. Conceptual development / progression of the assignment

Here is the logical chain of ideas.

---

#### Step 1 — Identify the scientific estimand (Q1)

Ask: > What quantity is scientifically meaningful?

Two possibilities:

- unit-specific: $p(\theta_i \mid y)$
- population/distributional: $p(\lambda, \tau^2 \mid y)$

Because lake **condition is population-level**, the primary target is distributional.
Therefore the mixture interpretation is most natural.

Core idea: > inference target determines modeling interpretation

---

**Step 2 — Summarize inference appropriately (Q2)**

Once the target is known:

If population-level: summarize

$$(\lambda, \tau^2), \quad p(\beta^* \mid y).$$

If unit-level: summarize $\{\beta_i\}$.

Thus you report: - hyperparameter summaries - predictive distributions - lake comparisons

Core idea: > summaries must align with the estimand

---

**Step 3 — Check model assumptions (Q3)**

Now shift from **estimation** to **adequacy**.

The model assumes

$$Y_{i,j} = \mu_{i,j} + \sigma_i \mu_{i,j}^{\theta} \varepsilon_{i,j}.$$

Fixing $\theta = 1$ imposes a specific mean–variance relationship:

$$\mathrm{Var}(Y_{i,j} \mid \cdot) = \sigma_i^2 \mu_{i,j}^2.$$

To assess this:

1. Compute standardized residuals
$$r_{i,j}^{(1)} = \frac{y_{i,j} - \hat{\mu}_{i,j}}{\hat{\sigma}_i \hat{\mu}_{i,j}}.$$

2. Check:

   - residual vs $\hat{\mu}$
   - $|r|$ vs $\hat{\mu}$
   - Q–Q plot
   - posterior predictive fit

3. Estimate implied $\theta$ via
$$\log\left(\frac{(y - \hat{\mu})^2}{\hat{\sigma}^2}\right) = c + 2\theta \log(\hat{\mu}) + \varepsilon.$$

4. Decide:

   - slope $\approx 2 \Rightarrow \theta \approx 1 \Rightarrow$ adequate

   - slope $\neq 2 \Rightarrow \theta \neq 1 \Rightarrow$ estimate $\theta$

   - strong between-lake variation $\Rightarrow$ consider $\{\theta_i\}$

Core idea: > embed the special case $\theta = 1$ inside a larger family and test adequacy

This mirrors parameterized link-function logic.

---

## 3. Big-picture synthesis

The full Bayesian workflow is:

1. Define the scientific estimand
2. Build a hierarchical model
3. Perform inference
4. Summarize parameters consistent with the estimand
5. Critique likelihood/variance assumptions
6. Expand the model only if diagnostics demand it

So the assignment teaches:

inference → interpretation → model criticism → refinement

which is exactly modern Bayesian practice.

---

## 4. Quick self-checklist

Before submitting, ask:

- Does my inference target match the biology?
- Do my summaries match that target?
- Did I check residuals and predictive fit?
- Did I justify any model expansion empirically?

If yes, your reasoning is statistically coherent.

---

## Bottom line

Review: - hierarchical Bayes, - mixture/marginal thinking, - variance-function modeling, - residual diagnostics / posterior predictive checks,