STAT7630: Bayesian Statistics
Lecture Slides # 17

Normal Hierarchical Models with Predictors
Chapter 17 (Normal) Hierarchical Models with Predictors

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: November, 2024

Outline

Hierarchical Models with Predictors

- **Hierarchical regression models** extend traditional regression by incorporating the grouping structure of hierarchical data while including predictor variables to enhance predictive accuracy.
- **Focus:** We revisit the Cherry Blossom Road Race dataset.
    - In Chapter 15, we analyzed net race time as a function of age using a "complete-data" normal regression model.
    - This model ignored the hierarchical structure of the data and failed to adequately capture the relationship between race time and age.
- **Advancement:**
    - We now introduce more sophisticated hierarchical models that explicitly account for the grouping structure present in the dataset.
    - These models better reflect the underlying relationship between race time and age.

Outline

Hierarchical Model with Varying Intercepts

- This model allows each group (e.g., each runner) to have a unique intercept, β_{0j}, for $j = 1, \ldots, n$:

$$Y_{ij} \mid \beta_{0j}, \beta_1, \sigma_y \sim N(\mu_{ij}, \sigma_y^2), \quad \text{where } \mu_{ij} = \beta_{0j} + \beta_1 X_{ij}.$$

- **Key assumptions:**
  - The slopes, β_1, of the group-specific regression lines are identical across all groups.
  - The regression lines for different groups are parallel on a graph.
- **Interpretation:**
  - Intercepts, β_{0j}, differ across groups, capturing variation in overall performance (e.g., some runners are inherently faster or slower).
  - The rate of change in expected times with respect to age, β_1, is the same for all runners.
- **Limitation:**
  - Assuming a constant slope across groups may not fully align with reality, as individual aging effects could vary.

Parameters in the Hierarchical Model with Varying Intercepts

- **Model Parameters:**
  - β_{0j}: Group-specific intercept for runner j, capturing their baseline performance level.
  - β_1: Global coefficient of age, representing the rate of change in race time with respect to age, assumed constant across all runners.
  - σ_y: Measure of within-group variability, quantifying how race times for a runner deviate from their true regression line.
- **Interpretation of σ_y:**
  - Describes the spread of the error terms, i.e., deviations of observed race times from the predicted times based on the runner's regression line.
  - This variability is assumed to be the same for all runners, reflecting a consistent level of uncertainty within groups.

Layer 2 of the Hierarchical Model: Varying Intercepts

- The second layer of the model specifies the distribution of the group-specific intercepts, β_{0j}:

$$\beta_{0j} \mid \beta_0, \sigma_0 \overset{\text{ind}}{\sim} N(\beta_0, \sigma_0^2).$$

- **Parameter interpretations:**
  - β_0: Global average intercept, representing the mean baseline performance across all runners.
  - σ_0: Between-group variability in β_{0j}, measuring the extent of variation in baseline speeds among runners.
- **Visual interpretation:**
  - σ_0 quantifies how vertically separated the runner-specific regression lines are on a graph.
  - Larger σ_0 indicates greater variation in baseline performance across runners.

Hierarchical Model with All Priors

- **Data model (within-runner regression):**

$$Y_{ij} \mid \beta_{0j}, \beta_1, \sigma_y \sim N(\mu_{ij}, \sigma_y^2), \quad \mu_{ij} = \beta_{0j} + \beta_1 X_{ij}.$$

- **Group-level model (variability in baseline speeds between runners):**

$$\beta_{0j} \mid \beta_0, \sigma_0 \overset{\text{ind}}{\sim} N(\beta_0, \sigma_0^2).$$

- **Priors on global parameters:**
  - $\beta_0 \sim N(m_0, s_0^2)$: Prior on the global intercept.
  - $\beta_1 \sim N(m_1, s_1^2)$: Prior on the global slope.
  - $\sigma_y \sim \text{Exp}(\ell_y)$: Prior on the within-runner variability.
  - $\sigma_0 \sim \text{Exp}(\ell_0)$: Prior on the between-runner variability.
- **Structure:** This hierarchical model combines within-group regression with between-group variability, anchored by priors on the global parameters.

Estimating the Model

- The model is estimated by simulating from the posterior distributions using the `stan_glmer` function from the `rstanarm` package (refer to `R` example).
- **Key results:**
  - The 80% credible interval for β_1 is:

$$(1.02, 1.58).$$

  - Since the credible interval contains only positive values, it indicates that:
    - **Runners slow down on average as they age.**
- **Comparison with complete pooling model:**
  - In the complete pooling model, the credible interval for β_1 included 0.
  - This result conflicted with expectations and demonstrated the limitations of the complete pooling approach.

Variation Among Runners

- The variation in intercepts (β_{0j}) among runners can be visualized (see the next 3 slides):
  - Compare credible intervals for β_{0j} values of runners 4 and 5.
  - Overlay posterior draws of their estimated regression lines.
- **Key observation:**
  - Runner 4 has a lower baseline speed (slower) compared to runner 5, as indicated by their respective β_{0j} values and regression lines.
- **Visualization for all runners:**
  - The runner-specific models for all 36 runners can be plotted, showing the distribution of baseline speeds and regression trends across the population (refer to `R` plot).

Posterior Plausible Models

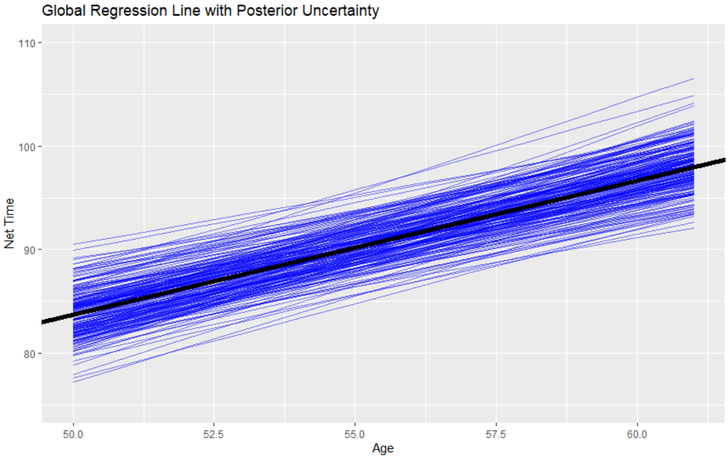

Figure 1: 200 posterior plausible global model lines, $\beta_0 + \beta_1 X$, for the relationship between running time and age.

Posterior Plausible Models for Runners 4 & 5

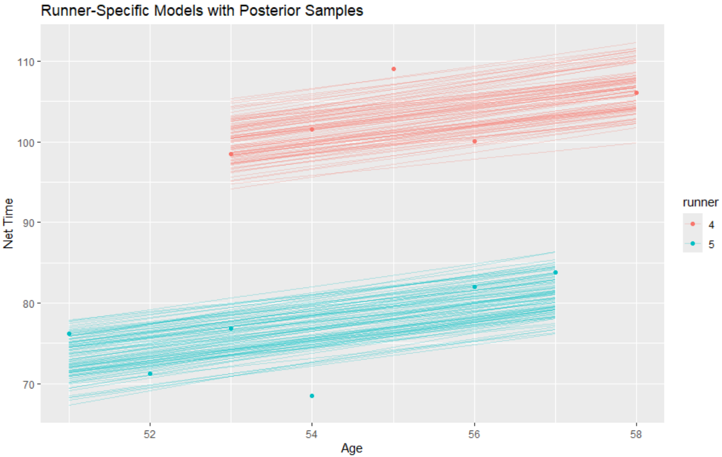

Figure 2: 100 posterior plausible models of running time by age, $\beta_{0j} + \beta_1 X$, for subjects $j \in \{4, 5\}$.

Posterior Models for all 36 Runners

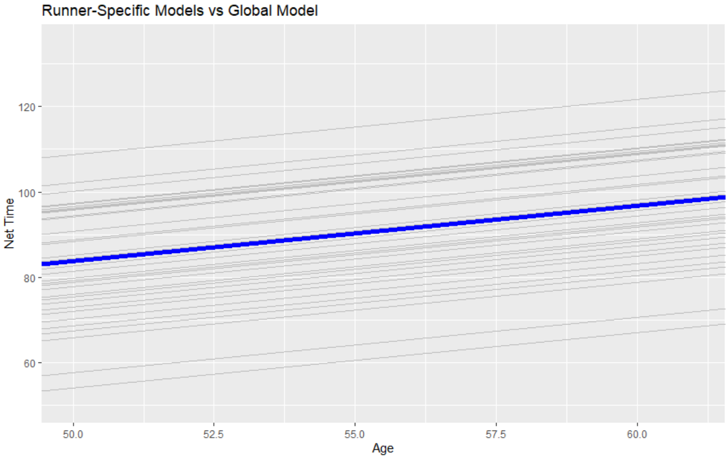

Figure 3: The posterior median models for our 36 runners j as calculated from the hierarchical random intercepts model (gray), with the posterior median global model (blue).

Examining Sources of Variability

- **Comparing σ_y and σ_0:**
  - σ_0: Variation in race times **between runners**.
  - σ_y: Variation in race times **within the same runner**.
- **Estimates:**
  - $\widehat{\sigma}_0 = 13.3$ (between-runner variability).
  - $\widehat{\sigma}_y = 5.25$ (within-runner variability).
- **Proportion of variance due to between-runner differences:**

$$\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_0^2 + \widehat{\sigma}_y^2} = \frac{13.3^2}{13.3^2 + 5.25^2} = 0.867.$$

  - Approximately 86.7% of the total variation in race times is attributable to differences between runners.

Outline

Hierarchical Model with Varying Intercepts and Slopes

- The earlier model assumed that all runners share the same slope, β_1, representing the rate at which race time changes with age.
- This assumption likely does not reflect reality:
  - Some runners slow down rapidly with age.
  - Others slow down gradually.
  - Some may even improve with age (see `R` plots in the next slide for evidence).
- **Advancement: The Varying Intercepts and Slopes Model:**
  - This model allows each runner to have:
    - A unique intercept, β_{0j}, reflecting their baseline performance.
    - A unique slope, β_{1j}, capturing their individual rate of change in race time with age.
- **Benefit:** By introducing varying slopes, the model better captures the heterogeneity in how runners' performances change over time.

Models for each of 36 Runners

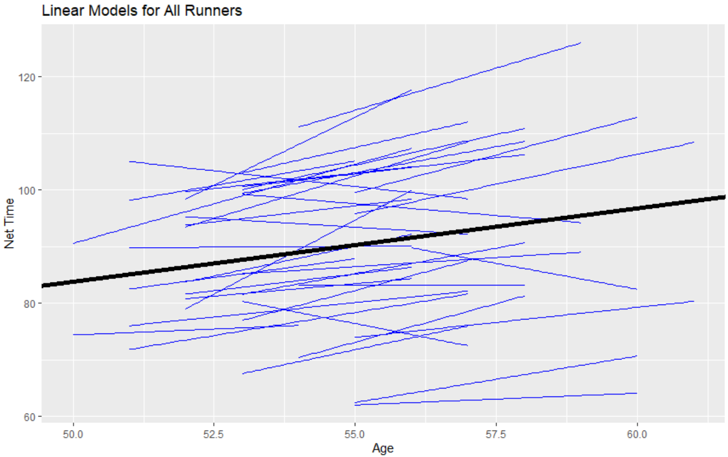

Figure 4: Observed trends in running time versus age for the 36 subjects (blue) along with the posterior median model (black).

Varying Intercepts and Slopes Model

- **Data model (within-runner regression):**

$$Y_{ij} \mid \beta_{0j}, \beta_{1j}, \sigma_y \sim N(\mu_{ij}, \sigma_y^2), \quad \mu_{ij} = \beta_{0j} + \beta_{1j} X_{ij}.$$

- **Group-level model (joint distribution of intercepts and slopes):** $\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} \Big| \beta_0, \beta_1, \Sigma \sim N\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \Sigma \right).$
- **Priors on global parameters:**
  - $\beta_0 \sim N(100, 10^2)$: Prior on the global intercept.
  - $\beta_1 \sim N(2.5, 1^2)$: Prior on the global slope.
  - $\sigma_y \sim \text{Exp}(0.072)$: Prior on the within-runner variability.
  - Σ: Covariance matrix for β_{0j} and β_{1j}, often modeled using a decomposition such as:
    - Variances: σ_0^2 (for β_{0j}) and σ_1^2 (for β_{1j}).
    - Correlation: ρ between β_{0j} and β_{1j}.
- **Structure:** This model captures both the variability in runners' baseline performance (intercepts) and their rates of change with age (slopes), as well as the relationship between these parameters.

Covariance Matrix of the β's

- **Covariance matrix Σ:**

$$\Sigma = \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix}.$$

- **Interpretation of elements:**
  - σ_0^2: Variance of the intercepts β_{0j}, capturing variability in baseline performance across runners.
  - σ_1^2: Variance of the slopes β_{1j}, representing variability in the effect of age on race time.
  - ρ: Correlation between β_{0j} and β_{1j}.

Covariance Matrix of the β's

- **Implications of strong correlation ($|\rho|$ close to 1):**
  - If β_{0j} and β_{1j} are strongly correlated, runners with particularly fast baselines (low β_{0j}) or slow baselines (high β_{0j}) are likely to have a pronounced effect of age on race time (very negative or very positive β_{1j}).
  - The precise interpretation of this correlation depends on the sign of β_{1j}:
    - Positive β_{1j}: Slower runners might improve more gradually with age.
    - Negative β_{1j}: Faster runners might slow down more dramatically with age.

Variance Components of the β's

- Variance components in Σ split the variability in runners' regression lines into:

$$\pi_0 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_1^2}, \quad \pi_1 = \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2}.$$

- **Interpretation:**
  - π_0: Proportion of variability due to intercept differences (β_{0j}), reflecting baseline performance differences.
  - π_1: Proportion of variability due to slope differences (β_{1j}), capturing aging trends.
- **Implications:**
  - Large π_0: Variation is mainly from intercepts (e.g., baseline speeds).
  - Large π_1: Variation is primarily from slopes (e.g., aging trends).

Outline

Posterior Simulation

- **Posterior analysis:**
  - Performed using the `stan_glmer` function.
  - The model includes 78 parameters, resulting in a slower computation time.
- **Posterior median model:**
  - The overall posterior median regression line is:

$$\widehat{Y} = 18.5 + 1.32 \times \text{age}.$$

  - This is similar to the random intercepts model but allows for greater flexibility through runner-specific parameters.
- **Advantages of varying intercepts and slopes:**
  - The model allows for runner-specific regression lines, incorporating unique β_{0j} and β_{1j} parameters for each runner.
  - Visualization of these runner-specific models (refer to `R` examples and plots) demonstrates the added nuance and variability captured by this approach.

Plots for Two Example Runners

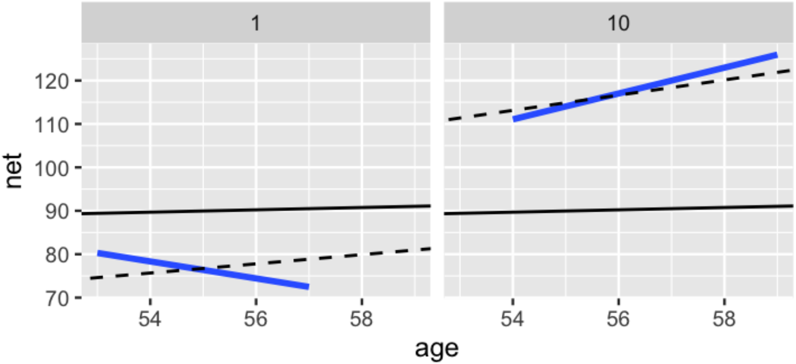

Figure 5: Posterior median relationships for runners 1 and 10 from the hierarchical model (dashed), contrasted with no-pooling (blue) and complete-pooling (black) models.

Shrinkage in the Hierarchical Model

- **Observation:** (Refer to plots for runners 1 and 10)
  - The no-pooling model's regression line for a runner (blue line) is **shrunk** toward the overall regression line from the complete-pooling model (solid black line).
  - This produces the estimated regression line from the hierarchical model (dashed black line).
- **Rationale for shrinkage:**
  - The hierarchical model assumes that information from other runners (captured by the complete-pooling model) informs the estimated regression line for runner j.
  - Data from a single runner, particularly with limited observations, may not fully describe that runner's true trend line.
  - Incorporating information from the broader population balances individual-level and group-level variability.

Shrinkage in the Hierarchical Model

- **Connection to Bayesian inference:**
    - Shrinkage embodies the Bayesian principle of balancing information:
        - Observed data provide specific details for a runner.
        - The population-level trend (akin to a prior) provides additional context, especially for runners with few data points.
- **Conclusion:** Shrinkage reflects the hierarchical model's ability to combine individual and group information to produce more robust estimates.

Outline

Model Selection

- **Model choices:**
  1. Complete pooling.
  2. No pooling.
  3. Varying intercepts.
  4. Varying intercepts and slopes.
- **Guidance for selection:**
  - Use intuition and context to inform the decision.
  - Formally evaluate model fit using `pp_check`.
  - Compare prediction accuracy using:
    - Prediction summary output.
    - Expected Log Predictive Density (ELPD) values.

Model Selection

- **Practical comparison:**
  - Refer to `R` example: Examine the criteria to evaluate the trade-off between the "varying intercepts" and "varying intercepts and slopes" models.
  - Consider whether the added complexity of varying slopes provides significant improvement in predictive accuracy.
- **Conclusion:** The best model balances fit quality and complexity, providing accurate predictions without unnecessary overfitting.

Posterior Prediction of Race Time for a New Individual

- **Prediction for a new individual:**
    - Use the `posterior_predict` function with the chosen hierarchical model to predict race time for a new individual at a specified age.
- **Prediction for individuals in the sample:**
    - Predict race time for an individual in the sample at an age not observed in the data.
    - Example: Predict the race time at age 61 for:
        - Runner 1.
        - Runner 10.
        - A new runner, Miles.

Posterior Prediction of Race Time for a New Individual

- **Precision of predictions:**
  - Predictions for Runner 1 and Runner 10 leverage their observed data, resulting in higher precision.
  - Predictions for Miles (a new runner) rely solely on population-level information, leading to much less precision (see `R` plots in the next slide for comparison).
- **Further exploration:**
  - Refer to Section 17.7 for an application using the Spotify dataset.
  - Example: Predict a song's danceability based on its genre and valence (mood). Explore this on your own for additional insights.

Posterior Predictive Models for 3 Runners

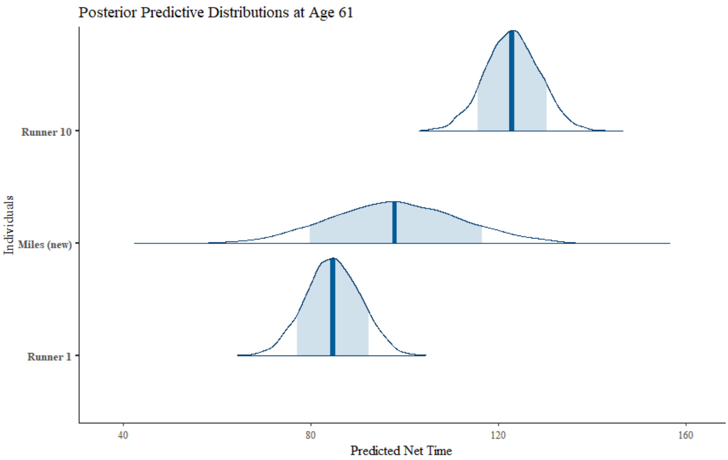

Figure 6: Posterior predictive models for the net running times at age 61 for sample runners 1 and 10, as well as Miles, a runner that wasn't in our original sample.