Model checking

After fitting the model, we can use the posterior to simulate synthetic data and compare to the data used to fit the model. Discrepancies can suggest paths to improve the model.

$$y_{sim} \sim P(y_{sim}|y) = \sum_{\theta} P(y_{sim}|\theta)P(\theta|y)$$

For each value of the parameters ($\theta_i = \{a_i, b_i, \sigma_i\}$) we can simulate a synthetic dataset y_{sim} and compare to the observed data y.

$$y_{sim} = Normal(a_i + b_i x, \sigma_i)$$

Step by step for posterior simulations

1. Extract the **posterior samples** for the parameters a, b, σ from the fitted model.

2. For each set of parameter values (a_i, b_i, σ_i):

  - Compute the predicted outcome: $y_{pred} = a + bx$.

  - Add random noise to y_{pred}, where the noise is drawn from a normal distribution with mean 0 and standard deviation σ_i.

This gives the synthetic data y_{sim}.

3. Compare the synthetic data y_{sim} to the observed data y.

  - Compute summary statistics (e.g., mean, variance, **quantiles**) for both y_{sim} and y.

  - If the summary statistics are similar for y_{sim} and y, this suggests that the model is a good fit to the data.

4. Repeat steps 2-3 for all sets of parameter values to get a distribution of summary statistics for the synthetic data.

5. Compare the distribution of summary statistics for the synthetic data to the corresponding summary statistics for the observed data. If they are similar, this suggests that the model is a good fit to the data. If they are not similar, this suggests that the model may need to be improved.