MISSING DATA

BST228

Speaker notes

No notes on this slide.

THE MISSING DATA PROBLEM

We want to learn about parameters $\boldsymbol{\theta}$ from observed data $\mathbf{y}_{\text{obs}}$, acknowledging missing data $\mathbf{y}_{\text{mis}}$. The posterior is

$$p\left(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}\right) \propto p\left(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right).$$

How do we account for $\mathbf{y}_{\text{mis}}$?

Speaker notes

- Sometimes we can’t or choose to observe all the data, e.g., due to respondents not answering survey questions, budget constraints, or individuals dropping out of clinical studies.
- Handling missing data is all about still making valid inferences in light of this challenge.

MISSING DATA AS PARAMETERS

We treat missing data $\mathbf{y}_{\text{mis}}$ like any other parameter

$$p\left(\boldsymbol{\theta}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}\right) \propto p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right).$$

Speaker notes

- Treating the missing data $\mathbf{y}_{\text{mis}}$ like any other parameter of the model is the general idea.

MISSING DATA COMPONENTS

1. Complete data likelihood $p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right)$.
2. Observation indicator $\mathbf{I}$ for each data point; 1 if observed, 0 if not.
3. Missingness model $p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}, \boldsymbol{\phi}\right)$ with parameters $\boldsymbol{\phi}$.
4. Joint posterior $p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right)$.

Speaker notes

- The complete data likelihood $p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right)$ is the likelihood we could evaluate if we had both the observed data $\mathbf{y}_{\text{obs}}$ and the missing data $\mathbf{y}_{\text{mis}}$.
- We treat the observation indicator $\mathbf{I}$ as fully observed—we generally know if we have the data or not.
- The missingness model captures how and why the data might be present or absent.
- If we can sample from the joint posterior, we can infer the parameters $\boldsymbol{\theta}$ we are interested in and ignore the imputed missing data $\mathbf{y}_{\text{mis}}$ and parameters of the ϕ of the missingness model.

MISSING DATA MECHANISMS

- Missing Completely at Random (MCAR): Data are missing independent of $\mathbf{y}_{\text{obs}}$ and $\mathbf{y}_{\text{mis}}$.
- Missing at Random (MAR): Conditional on $\mathbf{y}_{\text{obs}}$, data are missing independent of $\mathbf{y}_{\text{mis}}$.
- Missing not at Random (MNAR): Missingness depends on $\mathbf{y}_{\text{mis}}$.

Speaker notes

- MCAR is easy to analyze but often not realistic.
- MAR allows us to learn about $\mathbf{y}_{\text{mis}}$ based on $\mathbf{y}_{\text{obs}}$ and makes inference feasible.
- MNAR leaves us with a complex problem that can generally not be solved without strong assumptions.

EXAMPLE: TEST SCORES

Test scores for 1,000 students will be measured for two years, but 200 values are missing in the second year.

A: Students who scored lower in the first year are less likely to accept the invitation to take the test in the second year.

Speaker notes

- We consider the same dataset (1,000 test scores in the first year and 800 test scores in the second year) but with different scenarios for why the data are missing.

EXAMPLE: TEST SCORES

Test scores for 1,000 students will be measured for two years, but 200 values are missing in the second year.

B: Budget constraints in the second year require reducing the sample size. A random subset comprising 800 students are selected to take the test.

Speaker notes

No notes on this slide.

EXAMPLE: TEST SCORES

Test scores for 1,000 students will be measured for two years, but 200 values are missing in the second year.

C: Students who feel they didn’t do well on the test do not hand in their work.

Speaker notes

No notes on this slide.

RESULTS FROM POLL FOR MISSINGNESS MECHANISMS

Scenario	MCAR	MAR	MNAR
A: not returning for second test	8%	**62%**	31%
B: budget constraints	**85%**	8%	8%
C: not handing in poor results	0%	15%	**85%**

Speaker notes

- Our online poll correctly identifies the missing data mechanisms.
- Students not returning for the second test is data MAR because the missingness depends on the observed test scores from the previous year.
- Budget constraints lead to data MCAR because the budget has nothing to do with the test scores (unless the budget affects teaching).
- Not handing in poor results leads to data MNAR because the missingness depends on the data we are trying to collect.

THE DATA ARE …

- A: missing at random, $p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \phi\right) = p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \phi\right)$
- B: missing completely at random, $p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \phi\right) = p\left(\mathbf{I} \mid \phi\right)$
- C: missing not at random, $p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \phi\right) = p\left(\mathbf{I} \mid \mathbf{y}_{\text{mis}}, \phi\right)$

Speaker notes

- **A**: For data MAR, we can drop $\mathbf{y}_{\text{mis}}$ from the missingness model, but we must retain the conditioning on $\mathbf{y}_{\text{obs}}$. Missingness parameters ϕ could capture coefficients of a logistic regression model for $\mathbf{I}$ given previous test scores $\mathbf{y}_{\text{obs}}$.
- **B**: For data MCAR, we do not condition on any data in the missingness model. Missingness parameters ϕ might represent the available budget.
- **C**: For data MNAR, we have to condition on the missing data $\mathbf{y}_{\text{mis}}$ we are trying to collect—which greatly complicates our analysis. Missingness parameters ϕ could capture coefficients of a logistic regression model for $\mathbf{I}$ given unobserved test scores $\mathbf{y}_{\text{mis}}$.

WHY DOES THIS MATTER?

We use the following random effects model for test scores:

$$
\begin{aligned}
\mu_t &\sim \text{Normal}\left(0, 1\right) \quad \text{for } t \in \{1, 2\} \\
a_i &\sim \text{Normal}\left(0, \kappa^2\right) \quad \text{for } i \in \{1, \ldots, 1000\} \\
y_{it} &\sim \text{Normal}\left(a_i + \mu_t, \sigma^2\right),
\end{aligned}
$$

where μ is the mean score for each year and a_i is the skill of each student.

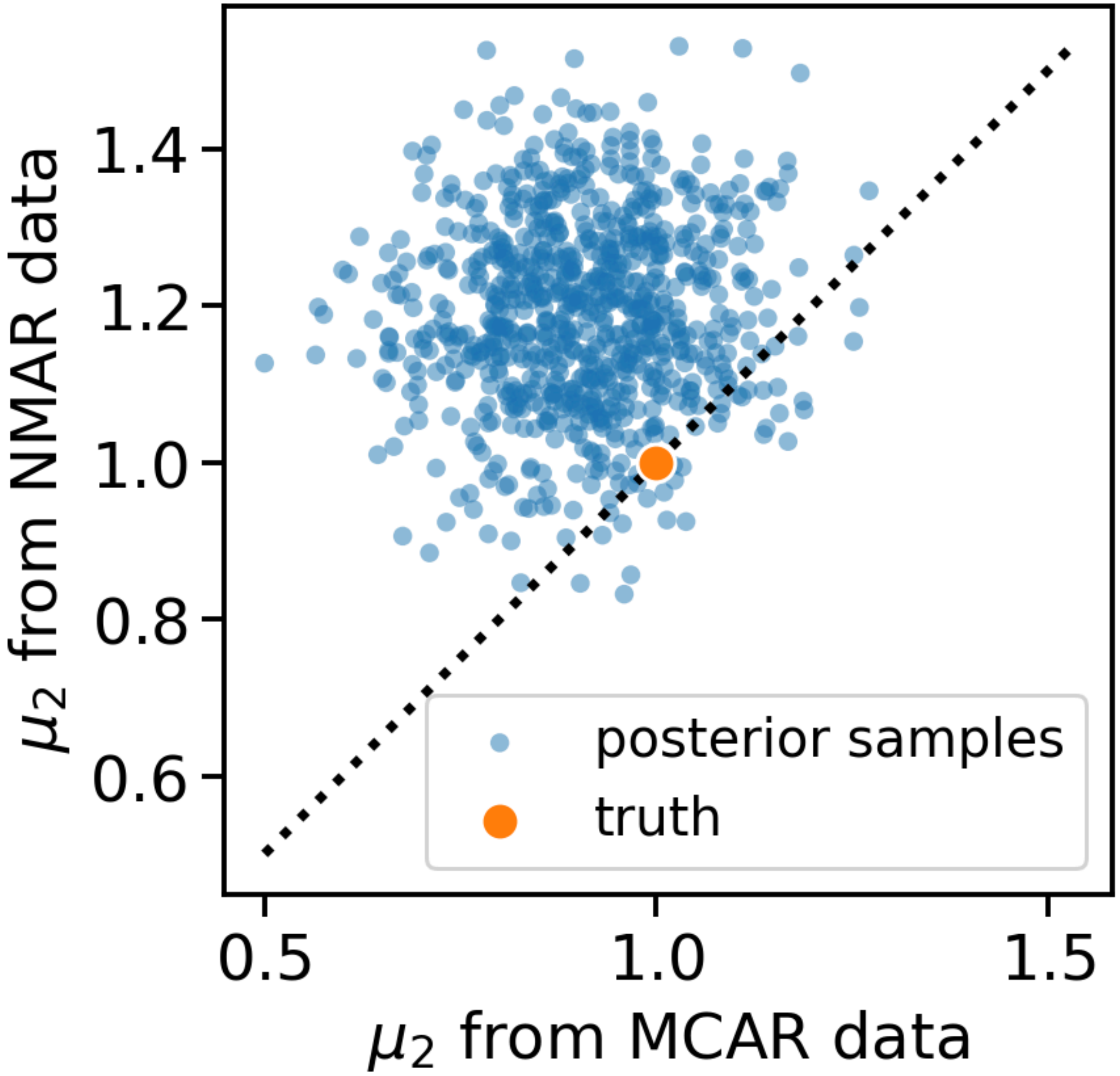
Speaker notes

- Test scores y_{it} depend both on the overall score μ_t for the test in year t (a measure of how easy the test is) and the individual student effect a_i for student i (a measure of how skilled the student is at answering test questions).
- The scale parameter κ captures the variance in student abilities. If κ is small, all students get similar scores. If κ is large, there is substantial variation in ability.

C: Students who feel they didn’t do well on the test do not hand in their work.

Speaker notes

- Under scenario C, what inferences might we observe if we naively don’t account for the missingness mechanism?



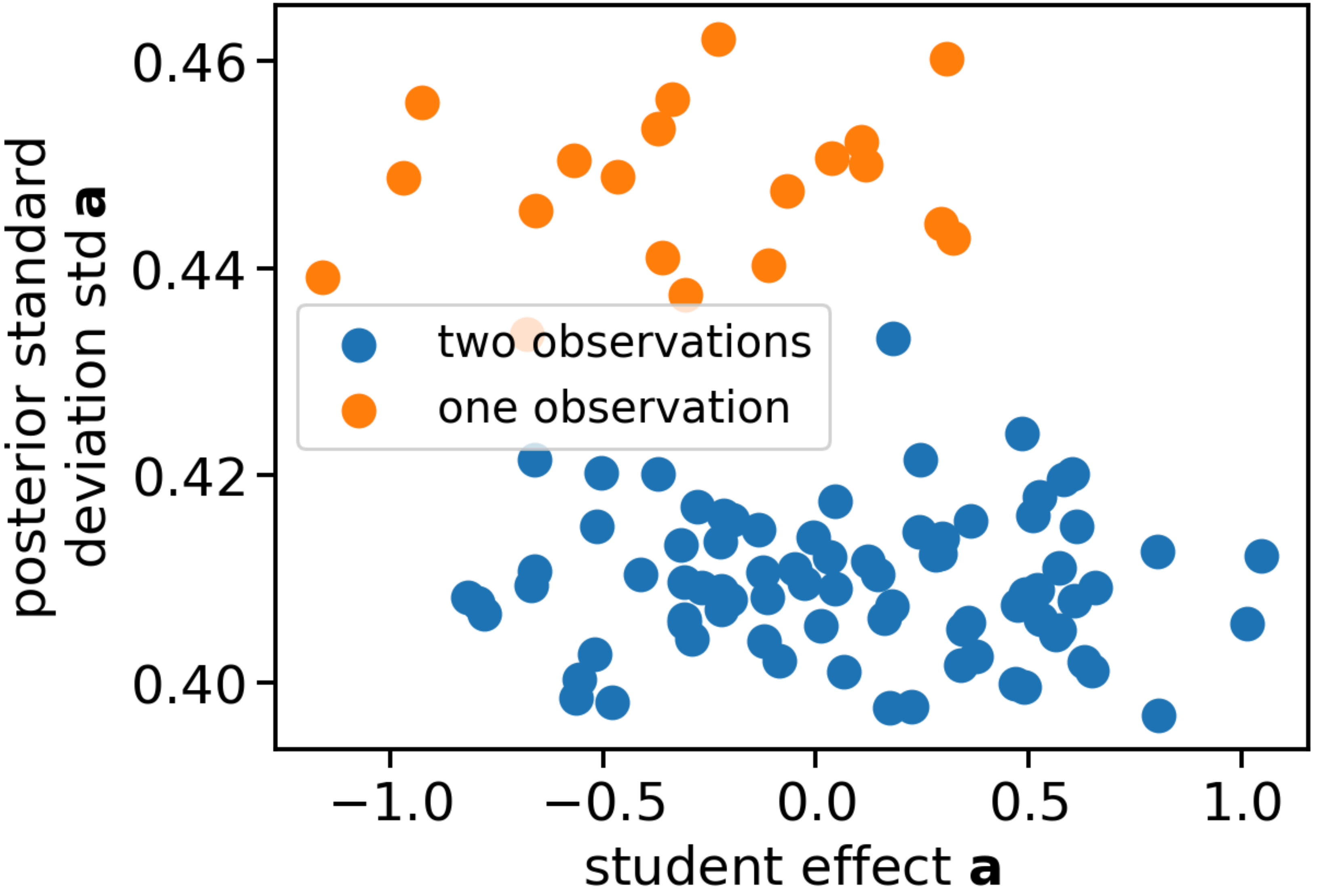
Speaker notes

- Blue scatter are posterior samples for a naive model that does not account for the missingness. Samples along the horizontal axis come from scenario B (MCAR due to budget constraints) and samples along the vertical axis come from scenario C (NMAR due to students not handing in results).
- Mean score μ_2 for the second year is overestimated because low scores are missing from the dataset.

A: Students who scored lower in the first year are less likely to accept the invitation to take the test in the second year.

Speaker notes

- Under scenario A, what inferences might we observe if we naively don’t account for the missingness mechanism?



Speaker notes

- The figure shows a scatter of posterior standard deviation for the inferred student effects $\boldsymbol{a}$ against the true student effect (which we know because the data are simulated).
- We make two observations: (a) are larger for students who did not complete the second test. (b) students who did not complete the second test have lower ability.
- Unlike in the NMAR setting, results for both $\boldsymbol{\mu}$ and $\mathbf{a}$ are unbiased. Intuitively, that's because row means of $\mathbf{y}$ (where entries are available) are unbiased estimators of student ability and columns means of $\mathbf{y}$ (where entries are available) are unbiased estimators of overall exam score.

JOINT VS MARGINAL POSTERIOR

We can sample …

- from the joint posterior

$$p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right) \propto p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \boldsymbol{\phi}\right) \\ \times p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}, \boldsymbol{\phi}\right)$$

- or from the marginal posterior

$$p\left(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right) = \int d\boldsymbol{\phi}\, d\mathbf{y}_{\text{mis}}\, p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right).$$

Speaker notes

- We can always sample from the joint posterior $p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right)$, but this is often challenging because the parameters can be high-dimensional, especially the missing data $\mathbf{y}_{\text{mis}}$.
- It is typically preferable to sample from the marginal posterior after integrating out the missing data $\mathbf{y}_{\text{mis}}$ and parameters of the missingness model $\boldsymbol{\phi}$. The multi-dimensional integral is also tricky, however.

IGNORABILITY FOR DATA MAR

Suppose data are MAR such that $p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \boldsymbol{\phi}\right) = p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi}\right)$ and the parameters of our model $\boldsymbol{\theta}$ and of the missingness model $\boldsymbol{\phi}$ are independent a priori, i.e., $p\left(\boldsymbol{\theta}, \boldsymbol{\phi}\right) = p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\phi}\right)$. Then, using the expression for the joint posterior from the previous slide,

$$\begin{aligned} p\left(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right) &= \int d\boldsymbol{\phi}\, d\mathbf{y}_{\text{mis}}\, p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right) \\ &\propto \int d\boldsymbol{\phi}\, d\mathbf{y}_{\text{mis}}\, p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi}\right) p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\phi}\right). \end{aligned}$$

The integral with respect to $\mathbf{y}_{\text{mis}}$ and $\boldsymbol{\phi}$ is separable such that

$$\begin{aligned} p\left(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}\right) &\propto \left[\int d\boldsymbol{\phi}\, p\left(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\phi}\right)\right] \left[\int d\mathbf{y}_{\text{mis}}\, p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)\right] \\ &\propto \int d\mathbf{y}_{\text{mis}}\, p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) \\ &\propto p\left(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right). \end{aligned}$$

The second line follows because the first integral does not depend on $\boldsymbol{\theta}$ and can be absorbed by the proportionality. The third line follows by the law of total probability.

The missing data model does not appear in the posterior, and we say the missing data mechanism is *ignorable*. This means we do not need to model $p\left(\mathbf{I} \mid \ldots\right)$, but it does *not* mean that we can ignore that data are missing.

Speaker notes

No notes on this slide.

EVALUATING THE OBSERVED-DATA LIKELIHOOD

Assuming an ignorable missingness mechanism, we seek to evaluate the observed data likelihood

$$p\left(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}\right) = \int d\mathbf{y}_{\text{mis}}\, p\left(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}\right).$$

Speaker notes

- The observed data likelihood $p\left(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}\right)$ looks straightforward, but the integral can be challenging if there are dependencies between observed and missing data.
- If $\mathbf{y}_{\text{obs}}$ and $\mathbf{y}_{\text{mis}}$ are independent conditional on the parameters $\boldsymbol{\theta}$, however, go straight ahead and run your model without further concerning yourself with missing data. This is exactly what happened in the student test score example in scenario A.

DATA AUGMENTATION IN THEORY

If the integral with respect to $\mathbf{y}_{\text{mis}}$ is not tractable, we treat $\mathbf{y}_{\text{mis}}$ as a latent variable and iteratively …

1. sample parameters $\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}$ from the posterior given complete data, including imputations $\mathbf{y}_{\text{mis}}$.
2. impute missing data $\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\theta}$ given observed data and model parameters.
3. repeat, starting at 1.

Speaker notes

- For non-tractable marginalization, we can sample from the posterior predictive distribution $\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\theta}$ to obtain an “imputed” dataset, i.e., a dataset that is consistent with our model.
- We then sample parameters $\boldsymbol{\theta}$ treating $\mathbf{y}_{\text{mis}}$ as if it was observed.
- What algorithm is this? It is a Gibbs sampler because we iteratively sample from the conditionals.

DATA AUGMENTATION IN PRACTICE

- Pick initial parameters $\boldsymbol{\theta}$, e.g., based on a complete case analysis.
- Pick initial imputations $\mathbf{y}_{\mathrm{mis}}$, e.g., based on mean imputation.
- Run the sampler.
- Check for convergence using MCMC diagnostics.
- Use $\boldsymbol{\theta}$ samples for analysis and posterior predictions.
- Use $\mathbf{y}_{\mathrm{mis}}$ samples to investigate missing data.

Speaker notes

- In practice, we need to initialize the algorithm with sensible starting values. For parameters $\boldsymbol{\theta}$, this can be achieved by using a complete case analysis (discarding all records that do not have complete data). For missing data $\mathbf{y}_{\mathrm{mis}}$, we can use a simple imputation method like replacing all missing values by the mean of observed values.
- After running the sampler, we need to check for convergence and discard burn-in samples because the sampler needs to converge to the target distribution from the initial guesses.

MISSING DATA ALWAYS REQUIRES A MODEL

To sample from $\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\theta}$, we must have a model for all missing parts, e.g., missing covariates in a regression model.

Speaker notes

No notes on this slide.

EXAMPLE: DIABETES

A diabetes dataset comprises 442 observations of disease progression $\mathbf{y}$ with a design matrix $\mathbf{X}$ comprising ten features for each observation. We assume 10% of predictors are MCAR.

Speaker notes

- If we conducted a complete-case analysis, we would have only ~150 records (34%) despite the missingness only being 10%.
- This problem becomes more severe the more features there are in the design matrix because every single one of them has to be observed to be included in a complete-case analysis.
- While a complete-case analysis is not technically wrong here because data are MCAR, we lose a lot of statistical power.

COMPLETE DATA MODEL

$$
\begin{aligned}
\text{mean of covariates } \boldsymbol{\mu} &\sim \mathsf{Normal}\left(0, 100\right) \\
\text{scale of covariates } \boldsymbol{\kappa} &\sim \mathsf{Normal}_{+}\left(0, 2\right) \\
\text{covariates } \mathbf{x}_i &\sim \mathsf{Normal}\left(\boldsymbol{\mu}, \boldsymbol{\kappa}^2\right) \\
\\
\text{regression coefficients } \boldsymbol{\theta} &\sim \mathsf{Normal}\left(0, 100\right) \\
\text{observation noise scale } \sigma &\sim \mathsf{Normal}_{+}\left(0, 2\right) \\
\text{outcomes } \mathbf{y} &\sim \mathsf{Normal}\left(\mathbf{X}\boldsymbol{\theta}, \sigma^2\right)
\end{aligned}
$$

Speaker notes

- We develop a model for every data element, including
  - the outcomes $\mathbf{y}$ in the second block: linear regression conditional on regression coefficients $\boldsymbol{\theta}$, features $\mathbf{X}$, and noise variance σ^2
  - features $\mathbf{X}$ in the first block: independent normal model conditional on population mean $\boldsymbol{\mu}$, e.g., mean age or blood sugar, and scale $\boldsymbol{\kappa}$, e.g., age or blood sugar variation.
- Beware of change in notation: We previously used $\mathbf{y}$ to denote all data. Here we use $\mathbf{y}$ for outcomes and $\mathbf{X}$ for features.
- We could implement a Gibbs sampler and iterate through all parameters. But we want to apply our Bayesian analysis skills to real-world problem rather than implement Gibbs samplers. Let’s use Stan.

```
data {
    int n, p;
    vector [n] y;
    matrix [n, p] X_obs;
}

transformed data {
    int n_mis = 0;
    for (i in 1:n) {
        for (j in 1:p) {
            n_mis += is_nan(X_obs[i, j]);
        }
    }
}
```

Speaker notes

- As `data`, we declare number of observations `n`, number of features `p`, outcomes `y`, and observed design matrix `X_obs`. Missing elements of `X_obs` are encoded by setting them to `nan`.
- The `transformed data` block is executed once per program to evaluate deterministic transformations of the data. Here, we simply count the number of missing elements `n_mis` in the observed design matrix `X_obs`.

```
parameters {
    vector [p] mu;
    vector<lower=0> [p] kappa;
    vector [p] theta;
    real<lower=0> sigma;
    array [n_mis] real X_mis;
}
```

Speaker notes

- We declare population mean `mu` and scale `kappa` for each column of the design matrix.
- The regression coefficients `theta` and `sigma` are standard for linear regression models.
- Finally, we have an array of real numbers `X_mis` with size equal to the number of missing elements `n_mis`.

```
1
2  transformed parameters {
3      matrix [n, p] X;
4      {
5          int k = 1;
6          for (i in 1:n) {
7              for (j in 1:p) {
8                  if (is_nan(X_obs[i, j])) {
9                      X[i, j] = X_mis[k];
10                     k += 1;
11                 } else {
12                     X[i, j] = X_obs[i, j];
13                 }
14             }
15         }
16     }
17 }
```

Speaker notes

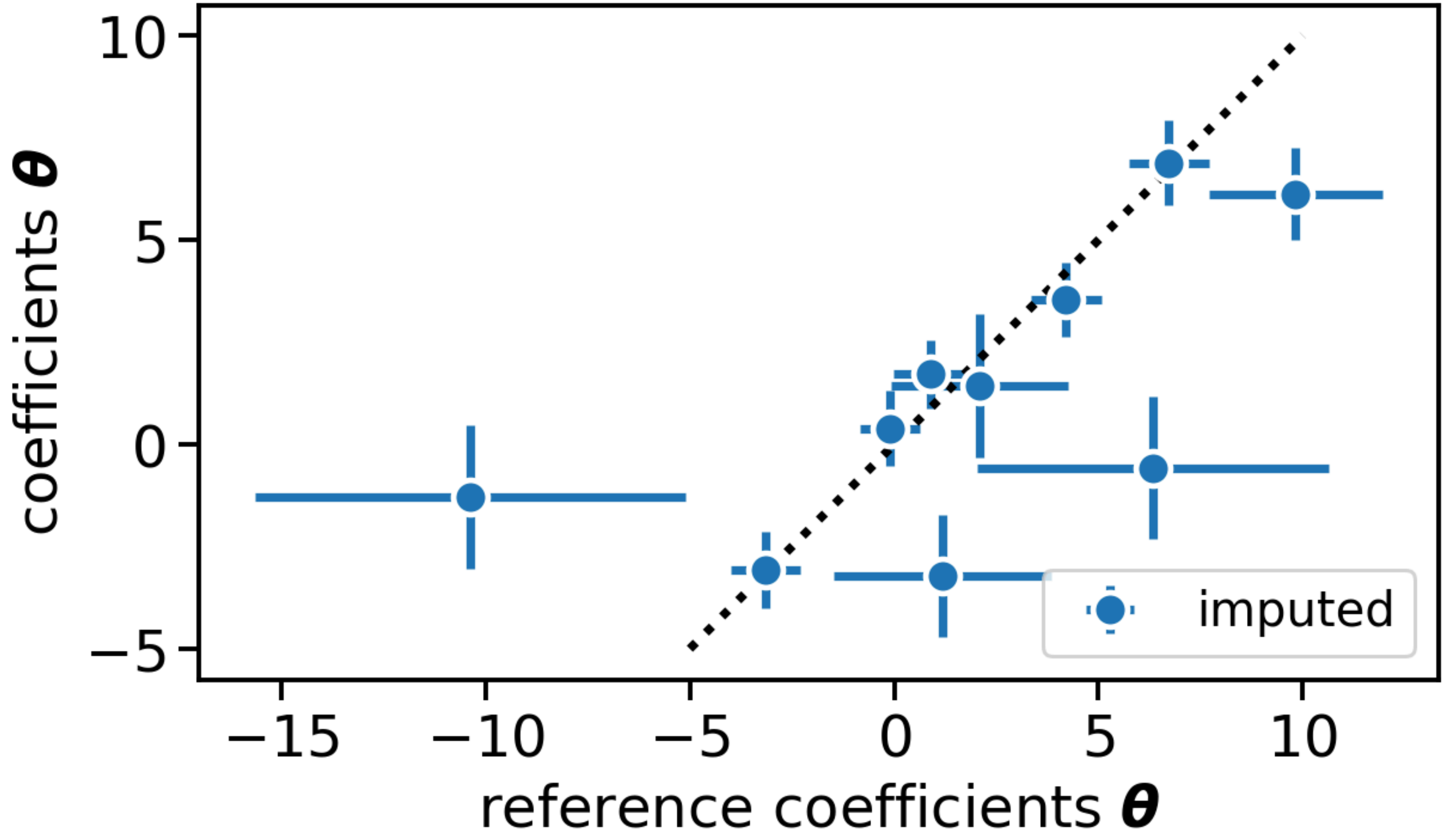
- In the transformed parameters, we construct the imputed design matrix `X` by copying from `X_obs` if data are available (line #12) or using a latent missing data parameter `X_mis` (#9).
- We use `k` to keep track of the next element we should take from `X_mis`.
- Everything is wrapped in braces `{ ... }` to hide the variable `k` from the `transformed parameters` block. Stan does not allow discrete parameters and complains if `k` is declared outside braces.

```
model {
    mu ~ normal(0, 100);
    kappa ~ normal(0, 2);
    for (i in 1:n) {
        X[i] ~ normal(mu, kappa);
    }

    theta ~ normal(0, 100);
    sigma ~ normal(0, 2);
    y ~ normal(X * theta, sigma);
}
```

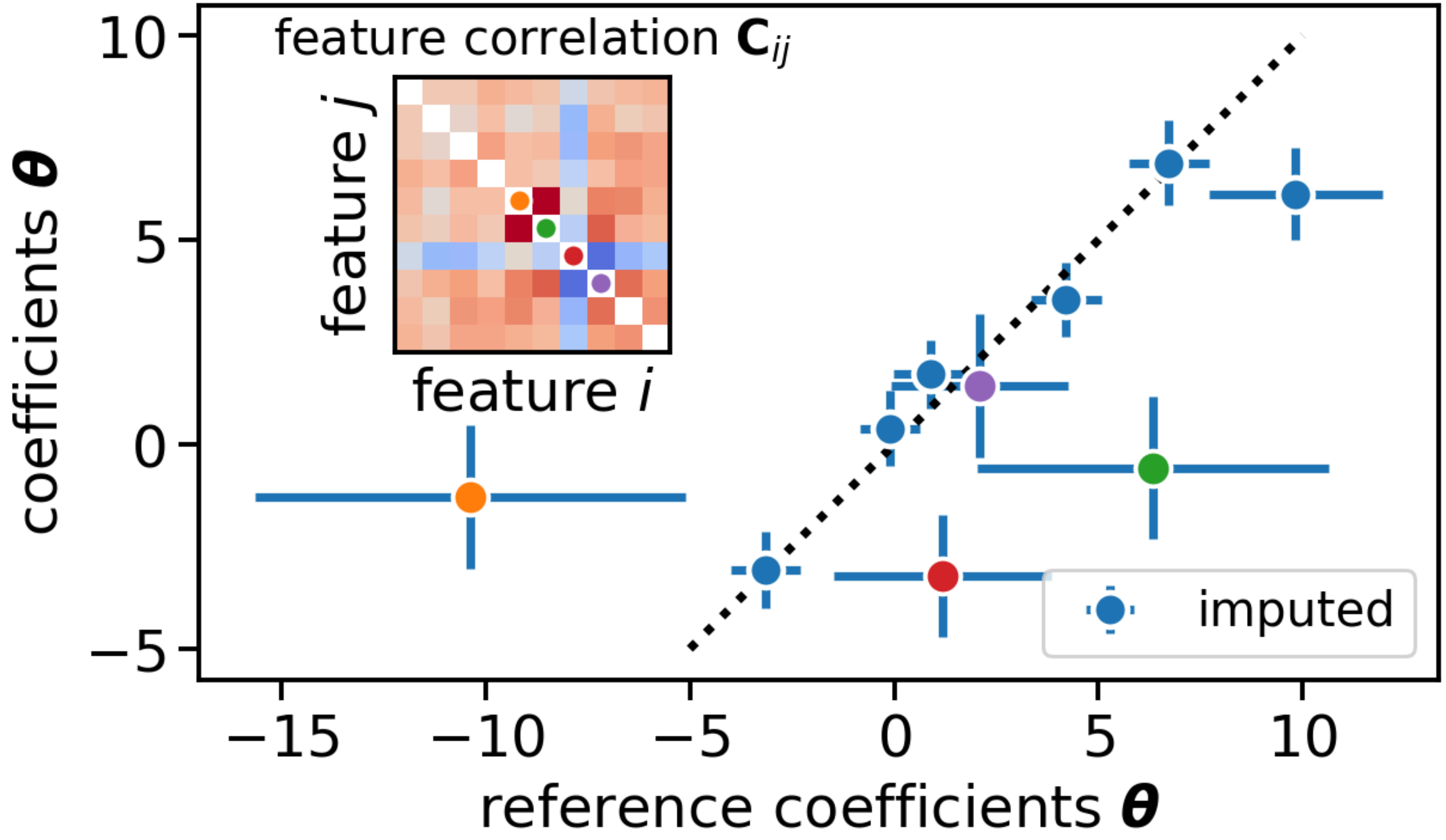
Speaker notes

- Finally, we declare the model in two blocks just like the model definition a few slides earlier.



Speaker notes

- We run the inference twice: (a) on the observed data with elements MCAR, (b) on the complete data to which we have access because we artificially removed elements from the design matrix for this example.
- We plot regression coefficients inferred using observed data against regression coefficients inferred on the complete dataset. Error bars are posterior standard deviations.
- They largely agree, but there are a few disagreements.



Speaker notes

- Why is this? Our missing data model is poor: (1) assumed a normal model which may not be appropriate for some features (like sex), (2) features are correlated and an independent normal model for each feature is not appropriate.
- 2nd is problematic: If a person suffers from diabetes, several blood tests are likely abnormal and features are correlated.
- We further investigate by plotting the feature correlation as an inset heatmap. We use colored markers on the diagonal of the correlation matrix to identify features and also color the corresponding regression coefficients.
- We observe that coefficients corresponding to correlated features exhibit disagreements. The issue is likely due to us neglecting feature correlation in the missing data model.

Speaker notes

No notes on this slide.

EXAMPLE: 1988 ELECTION

In 1988, Democrat Dukakis was far ahead of Republican Bush in the polls but decisively lost the election. What happened?

- **Data**: Surveys of public opinion from 51 national polls conducted by nine polling organizations over six months.
- **Goal**: Understand temporal evolution of vote intention for subgroups of the population.
- **Challenge**: Not all questions were asked in all surveys and some questions were not answered.

BUILDING AN IMPUTATION MODEL

- There are two extremes: complete pooling (impute all surveys together using the same model) and no pooling (impute all surveys independently—not possible for missing questions).
- We use a hierarchical imputation model: imputations are informed by the specific survey when possible and by the population if not.
- This can address both questions not answered (within-survey imputation) and questions not asked (between-survey imputation).

Speaker notes

No notes on this slide.

HIERARCHICAL MODEL

- y_{siq} is response of individual j in survey s to questions q.
- μ_{sq} is the mean response to question q in survey s.
- $\mathbf{x}_s$ is a feature vector for survey s, e.g., time.
- $\boldsymbol{\beta}$ is a coefficient *matrix*, capturing dependence of responses on features.
- $\boldsymbol{\Psi}$ is a covariance matrix for responses, e.g., opinion of Dukakis may be anticorrelated with opinion of Bush.

$$\begin{aligned}
\mathbf{y}_{si} &\sim \text{Normal}\left(\boldsymbol{\mu}_s, \boldsymbol{\Psi}\right) \\
\boldsymbol{\mu}_s &\sim \text{Normal}\left(\boldsymbol{\beta}\mathbf{x}_s\beta, \sigma^2\mathbf{I}_Q\right)
\end{aligned}$$

Speaker notes

- We have a *matrix* of coefficients because the mean response $\boldsymbol{\mu}_s$ is itself a vector rather than a scalar. The second line is a latent regression model with multivariate outcome.
- We ignore the missing data mechanism because the majority of missingness is due to some surveys not asking certain questions rather than people not answering questions they were asked.

GIBBS SAMPLER

- Sample missing answers in $\mathbf{y}$.
- Sample answer covariance $\boldsymbol{\Psi}$.
- Sample latent survey means $\boldsymbol{\mu}$.
- Sample between-survey variance σ^2.
- Sample regression matrix $\boldsymbol{\beta}$.

Speaker notes

- We can implement a custom Gibbs sampler or … use Stan.

RECAP

- Missing data are just more parameters.
- Distinction between MCAR, MAR, and NMAR.
- Ignorability of the missingness *mechanism* for MAR but *not* the missignness itself.
- Imputation using a Gibbs sampler.
- Stan model for imputed data.

Speaker notes

No notes on this slide.