5. Bayesian Statistical Paradigm

Frequentist Paradigm Tenets

A parameter such as μ is the true value of a statistic in the full population of participants

The parameter is fixed in the full population, we compute an estimate from a sample

Estimates vary across different random samples we could potentially work with

Important Implications

Probability doesn't apply to a fixed parameter

Any statements about probability, precision, confidence intervals, etc. refer to estimates

Probability = long run frequency of estimates across many different samples of data

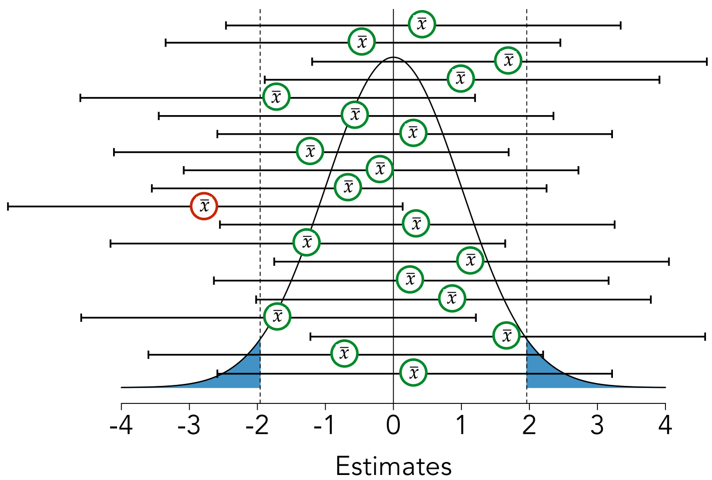
95% Confidence Intervals

The 95% confidence interval from a sample either contains the parameter or it does not

"Confidence" is the long-run frequency that such an interval contains the true parameter

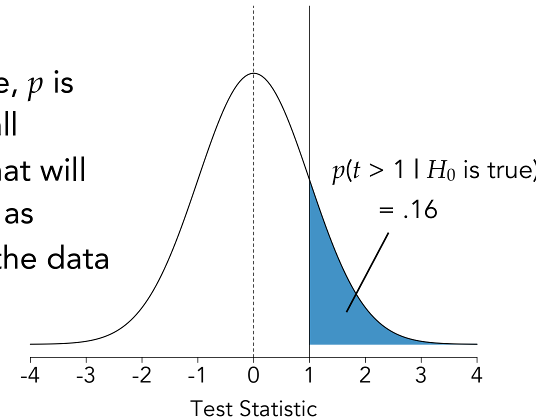
e.g., 95 out of 100 samples we could potentially work with should yield confidence intervals that include the true population parameter

Confidence Intervals From 20 Samples



Significance Tests And p-Values

Assuming H_0 is true, p is the proportion of all random samples that will give a test statistic as large as that from the data



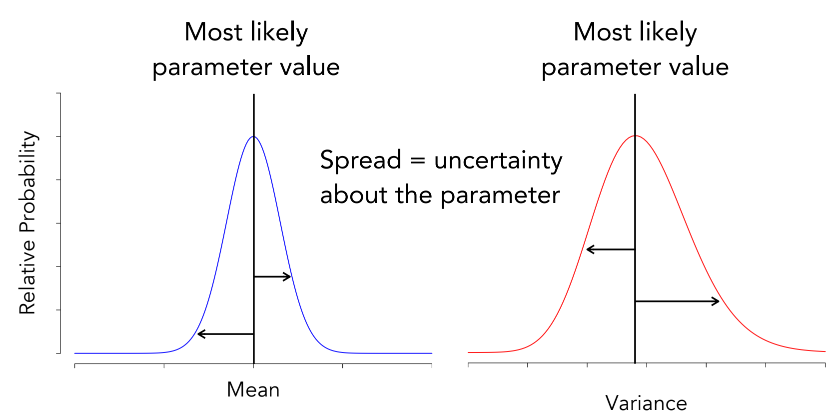
Bayesian Paradigm

The Bayesian framework defines a parameter such as μ as a variable that has a distribution

e.g., The mean is a normally distributed variable

The posterior distribution represents our knowledge about the potential values of the parameter after analyzing the data

Posterior Distributions Of Parameters



Important Implications

Probability applies to parameters

Probability = the degree of certainty about the parameter values after analyzing the data

e.g., The credible intervals gives a range in which 95% of the parameter values are likely to fall; the probability is the proportion of parameters above or below a particular value

Why Bayes?

Bayes is a direct estimation approach that is a flexible alternative to maximum likelihood

Bayesian estimation (MCMC) is the mathematical machinery behind multiple imputation

Often the only way to get good results for some analyses (e.g., mixes of categorical and continuous predictors, interactions, multilevel models)

Bayes' Theorem

Likelihood of the data, given an assumed parameter value (ML)

A priori distribution of the parameter

$$f(\theta|data) = \frac{f(prior) \times f(data|\theta)}{f(data)}$$

Probability of the parameter, given the data (the posterior)

Unnecessary scaling factor

Research Scenario

Two researchers use the Bayesian statistical paradigm to estimate the incidence of postpartum depression in the population of new mothers

The posterior distribution is proportion to the prior distribution times the likelihood of the data

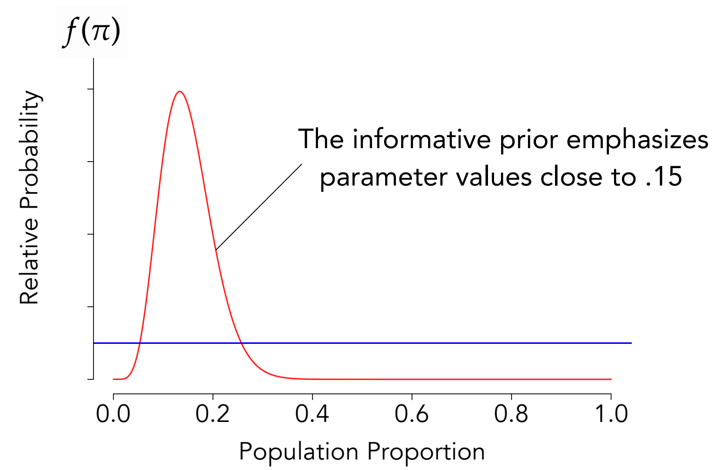
$$f(\pi|data) \propto f(\pi) \times f(data|\pi)$$

Prior Distributions

Researcher A uses a meta-analysis to specify an informative prior distribution where depressive rates of 15% are very likely

Researcher B specifies a non-informative prior distribution where all values of the population proportion are equally likely

Informative vs. Non-Informative Prior



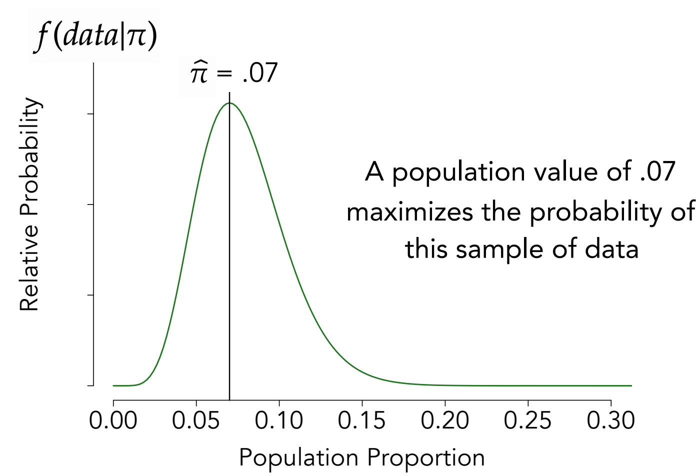
The Likelihood Of The Data

Researchers recruit 100 new mothers, seven of whom are diagnosed with clinical depression

The maximum likelihood estimate of the population proportion is $\hat{\pi}$ = .07

The likelihood function gives the relative probability of this data, given differed assumed values of the population parameter

Likelihood Function



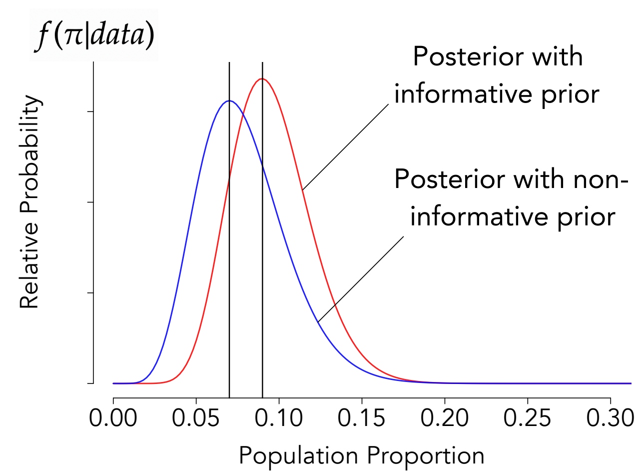
Posterior Distributions

The distribution of the parameter given the data — the posterior — blends information from the prior distribution and the likelihood

The non-informative prior has no influence because it weights all values of π the same

The informative prior acts like a secondary data source that combines with observed data

Posterior Distributions



Moving Forward With Prior Distributions

Software packages often default to non-informative prior distributions that have no (or minimal) impact on the analysis results

Priors for means and regression coefficients are usually flat like the previous example

Priors for variance parameters impart information, but their impact is only evident in small samples

A Simple Analysis Model

An empty regression model has two parameters, the mean and variance

$$Y_i = \mu + \varepsilon_i = E(Y) + \varepsilon_i$$
$$Y_i \sim N\left(E(Y), \sigma^2\right)$$

Y is normally distributed around a predicted value, $E(Y)$, which in this case is the mean

Some Recurring Notation

Predicted value

$$Y_i = \mu + \varepsilon_i = E(Y) + \varepsilon_i$$

$$Y_i \sim N\left(E(Y), \sigma^2\right)$$

Univariate normal curve

Spread

Center

Math Achievement Data

Math achievement data for 250 students

The data set includes pre-test and post-test math achievement scores and academic-related variables such as math self-efficacy, math anxiety, standardized reading scores, socio-demographic variables

math.dat

Variable	Name	Missing %	Scaling
Identifier variable	ID	0	Integer index
Gender	MALE	0	0 = female, 1 = male
Free or reduced lunch	LUNCHASST	4.3	0 = none, 1 = assistance
Achievement group	ACHIEVEGRP	2.0	1 = typically achieving, 2 = low achieving, 3 = learning disability
Standardized reading	STANREAD	10.0	Continuous
Math self-efficacy	EFFICACY	9.7	6-point ordinal scale
Math anxiety	ANXIETY	9.3	Continuous
Pre-test math achievement	MATHPRE	0	Continuous
Post-test math achievement	MATHPOST	18.0	Continuous

Substantive Example

Use the Bayesian framework to estimate the mean and variance of math post-test scores

$$MATHPOST_i = \mu + \varepsilon_i$$

The analysis assumes that math scores are normally distributed around the mean

Bayesian Estimation With MCMC

MCMC = Markov chain Monte Carlo

MCMC is an iterative procedure that estimates one parameter at a time, treating the current values of other parameters as known constants

Repeating the sequence many times gives a posterior distribution for each parameter

MCMC Recipe

Do for t = 1 to T iterations

1. Estimate the mean, given the current value of the variance

2. Estimate the variance, given the current value of the mean

Repeat

The Meaning Of Estimation

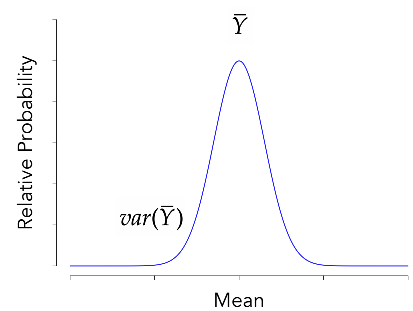
Parameters are variables that have a distribution

Bayesian estimation uses Monte Carlo computer simulation to "draw" new estimates from a distribution of plausible values

Estimation = random number generation

Conditional Distribution Of The Mean

$$f\left(\mu|\sigma^2, data\right) \propto f(\mu) \times f\left(data|\mu, \sigma^2\right) \propto N\left(\bar{Y}, \frac{\sigma^2}{N}\right)$$



The posterior distribution of μ is a normal curve centered at the mean of the sample data

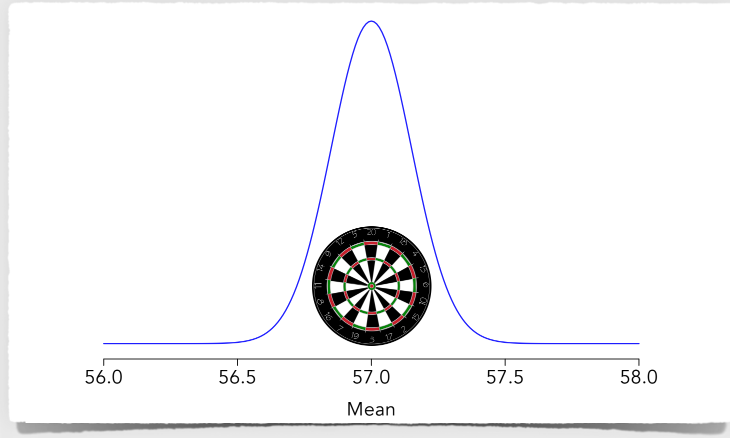
Monte Carlo Dart Game

Random number generation is akin to throwing a dart at a picture of a normal distribution

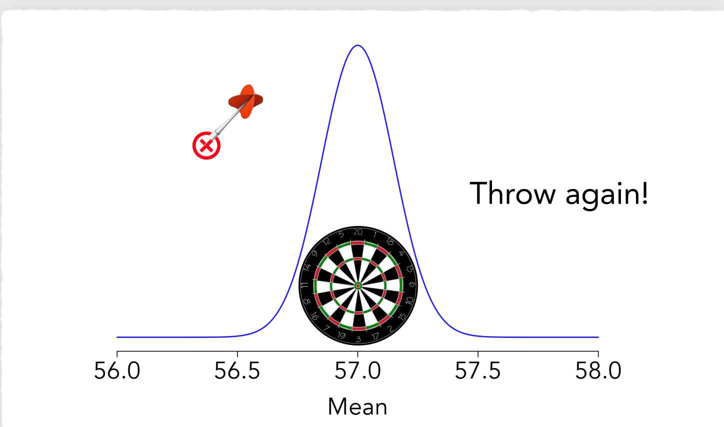
For any dart that lands under the distribution, its location on the horizontal axis is the value of the next parameter estimate

More darts should land in the peaked area of the curve, fewer darts should land in the tails

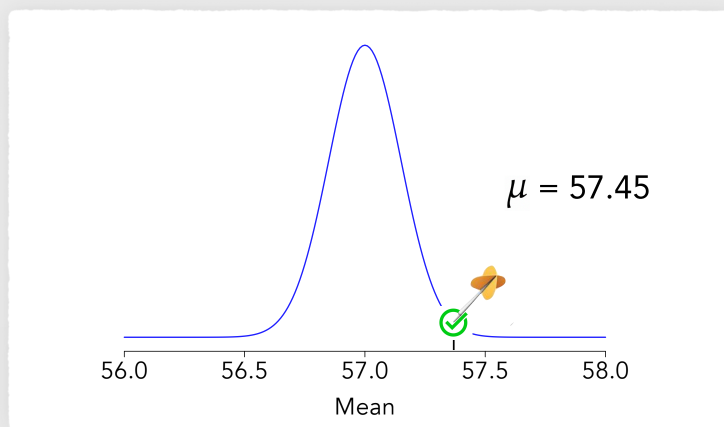
Monte Carlo Dartboard



Monte Carlo Dartboard

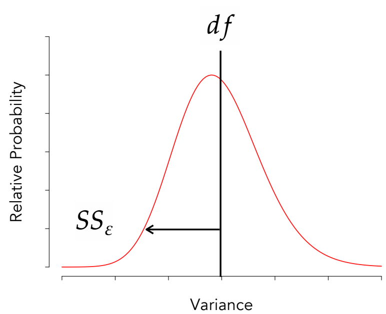


Monte Carlo Dartboard



Conditional Distribution Of The Variance

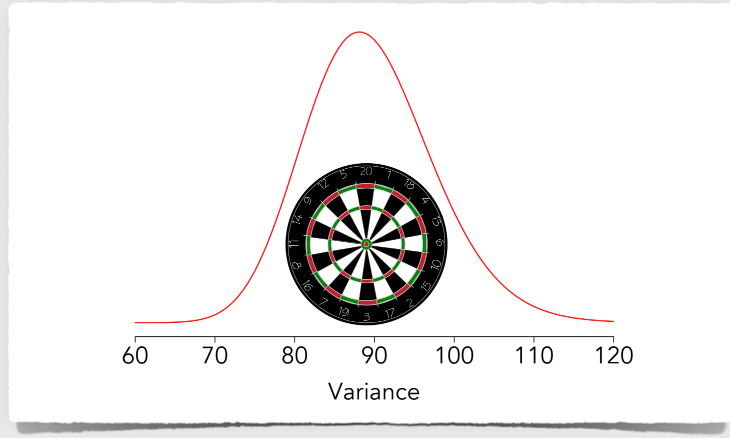
$$f\left(\sigma^2|\mu, data\right) \propto f\left(\sigma^2\right) \times f\left(data|\mu, \sigma^2\right) \propto IG\left(\frac{df}{2}, \frac{\sum(Y_i - \mu)^2}{2}\right)$$



The posterior distribution of σ^2 is a positively skewed inverse gamma distribution

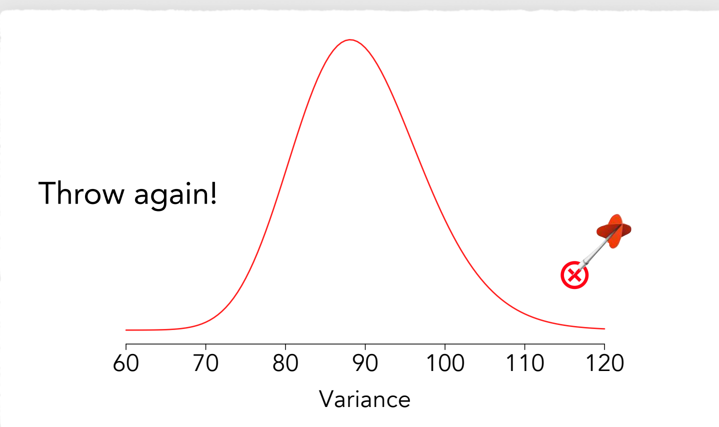
33

Monte Carlo Dartboard



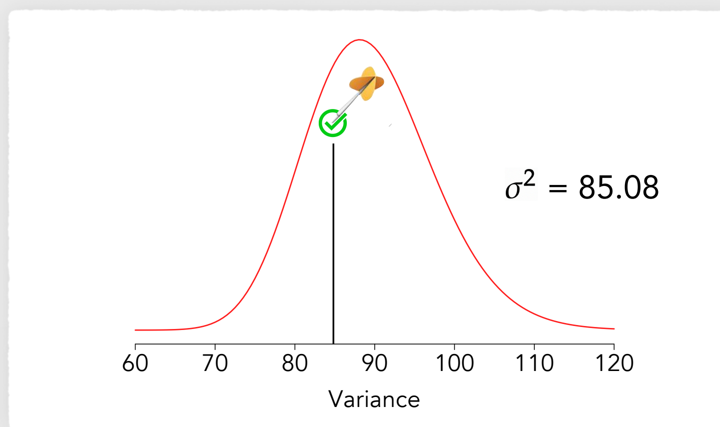
34

Monte Carlo Dartboard



35

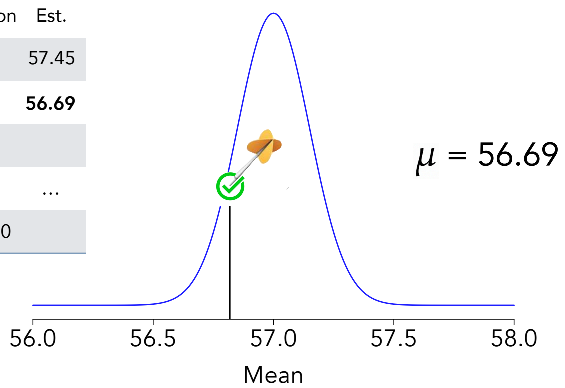
Monte Carlo Dartboard



36

Mean At Iteration Two

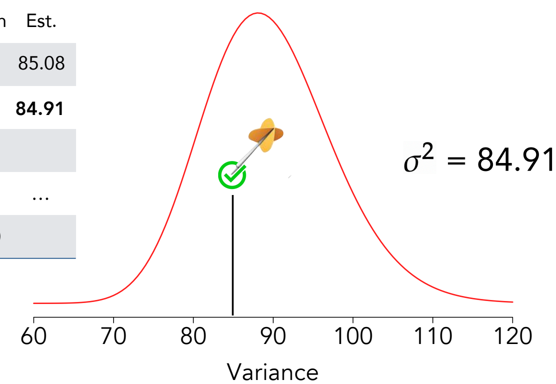
Iteration	Est.
1	57.45
2	**56.69**
3	
…	…
10,000	



37

Variance At Iteration Two

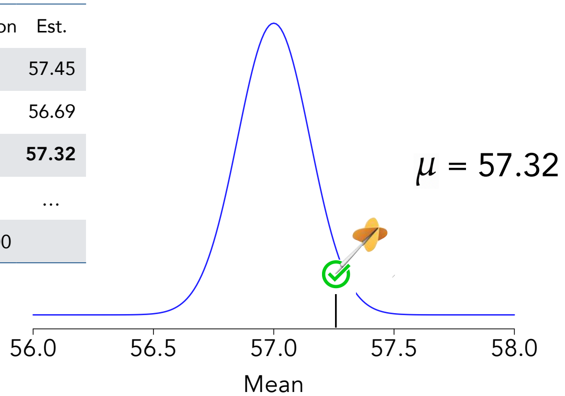
Iteration	Est.
1	85.08
2	**84.91**
3	
…	…
10,000	



38

Mean At Iteration Three

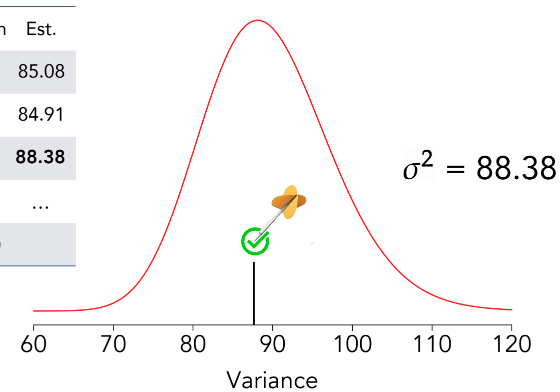
Iteration	Est.
1	57.45
2	56.69
3	**57.32**
…	…
10,000	



39

Variance At Iteration Three

Iteration	Est.
1	85.08
2	84.91
3	**88.38**
…	…
10,000	



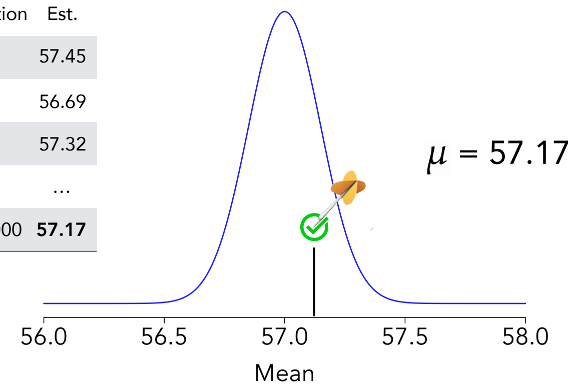
40

Iterate …

41

Mean At Iteration 10,000

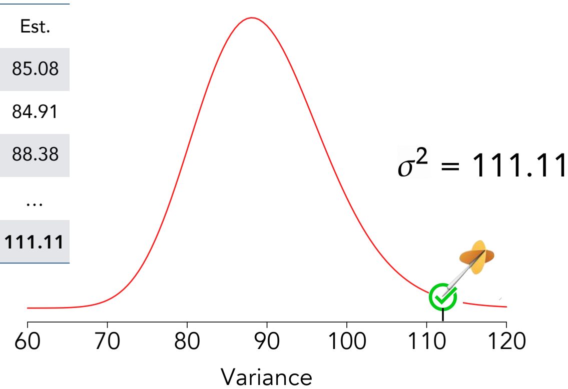
Iteration	Est.
1	57.45
2	56.69
3	57.32
…	…
10,000	**57.17**



42

Variance At Iteration 10,000

Iteration	Est.
1	85.08
2	84.91
3	88.38
…	…
10,000	**111.11**



43

Trace Plots

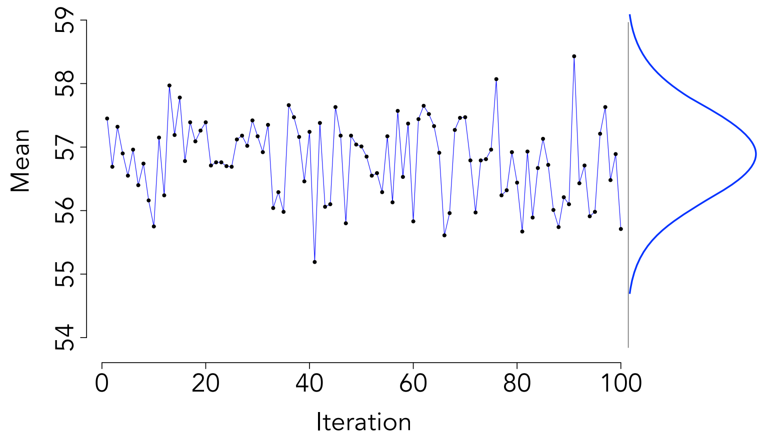
A trace plot is a line graph that displays the iterations on the horizontal axis and the parameter estimates on the vertical axis

Trace plots are important tools for evaluating whether the MCMC algorithm is working well

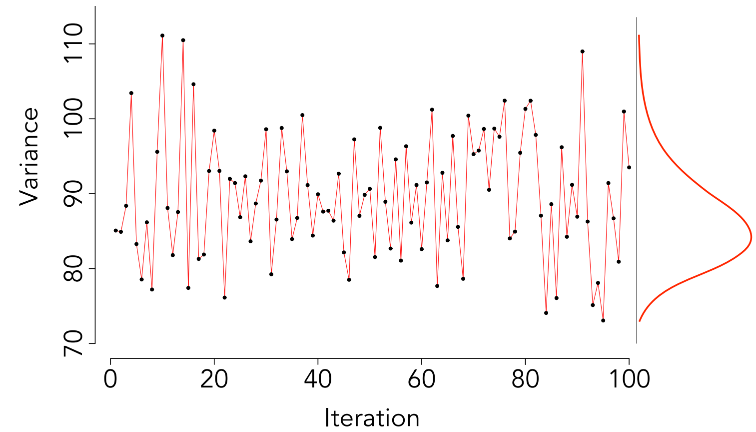
e.g., Does the algorithm achieve a steady state?

44

Trace Plot Of The Mean



Trace Plot Of The Variance



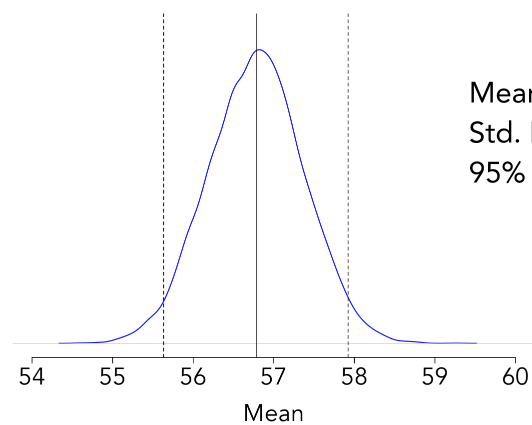
Summarizing The Posterior Distributions

Repeating the estimation steps for thousands of iterations gives a distribution for each parameter

Use descriptive statistics to summarize the center and spread of the posterior distribution

The mean and standard deviation are analogous to a point estimate and standard error

Posterior Distribution Of The Mean



Mean = 56.82
Std. Dev. = .63
95% C.I. = (55.57, 58.05)

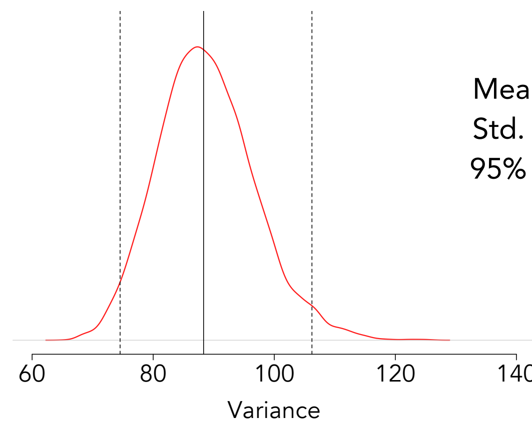
Interpretations Are Subtly Different!

The posterior mean, M_μ = 56.82, is the most likely parameter value (akin to a point estimate)

The posterior standard deviation, SD_μ = .63, quantifies uncertainty in the parameter after analyzing the data (akin to a standard error)

The credible interval indicates that 95% of the parameter values fall between 55.57 and 58.05

Posterior Distribution Of The Variance



Maximum Likelihood vs. Bayesian

Point estimates and standard errors are numerically similar to posterior means and standard deviations

	Maximum Likelihood		Bayesian Estimation	
	Estimate	Std. Error	Mean	Std. Dev.
Mean	56.83	0.63	56.82	0.63
Variance	82.76	8.12	84.00	8.34

Blimp Bayesian Analysis Script

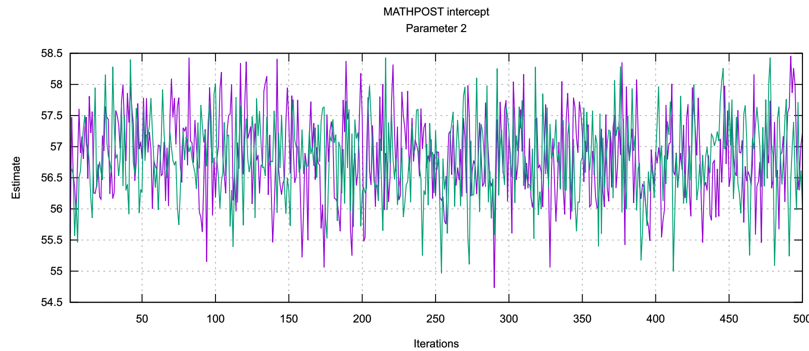
```
DATA: math.dat;
VARIABLES: id male lunchasst achievegrp stanread efficacy
    anxiety mathpre mathpost;
MISSING: 999;
MODEL: mathpost ~ ;
SEED: 90291;
BURN: 1000;
ITERATIONS: 10000;
CHAINS: 4 processors 4;
OPTIONS: psr;
```

Blimp Output

```
ANALYSIS MODEL ESTIMATES:

Missing outcome: mathpost

                              ------------------------------------------------------------
Parameters                    |   Mean    |  Median   |  StdDev   |Lower 2.5  |Upper 97.5|

Variances:                    |           |           |           |           |          |
  Residual Var.               |    84.001|    83.340|     8.337|    69.372|   101.917|
                              |----------|----------|----------|----------|----------|
Coefficients:                 |           |           |           |           |          |
  Intercept                   |    56.824|    56.827|     0.632|    55.569|    58.050|
                              |----------|----------|----------|----------|----------|
Standardized Coefficients:    |           |           |           |           |          |
                              |----------|----------|----------|----------|----------|
Proportion Variance Explained |           |           |           |           |          |
  by Fixed Effects            |     0.000|     0.000|     0.000|     0.000|     0.000|
  by Residual Variation       |     1.000|     1.000|     0.000|     1.000|     1.000|
                              |           |           |           |           |          |
                              ------------------------------------------------------------
                              Summaries based on 10000 iterations using 4 chains
```

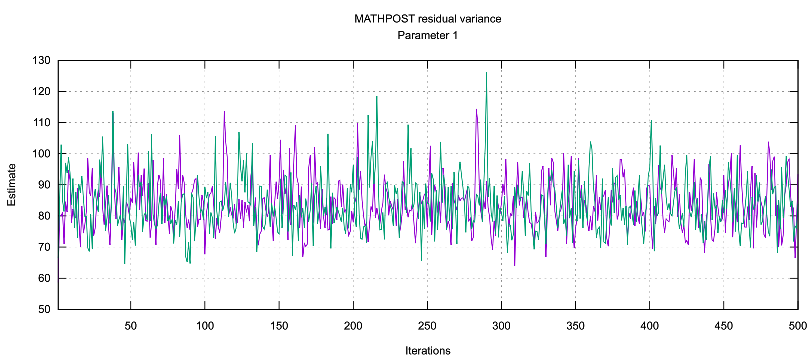

Blimp Trace Plots

Grand mean (intercept) estimates from 500 iterations



Blimp Trace Plots

Variance estimates from 500 iterations



Mplus Bayesian Analysis Script

```
DATA:
file = math.dat;
VARIABLE:
names = id male lunchasst achievegrp stanread efficacy anxiety mathpre mathpost;
usevariables = mathpost;
missing = all(999);
ANALYSIS:
estimator = bayes;
bseed = 90291;
fbiterations = 10000;
MODEL:
[mathpost]; mathpost;
OUTPUT:
tech8;
```

Mplus Output

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
Means					
MATHPOST	56.835	0.634	0.000	55.579	58.065
Variances					
MATHPOST	84.200	8.402	0.000	69.978	102.573