

BAYESIAN STATISTICS *CONCEPT* *AND* *BAYESIAN CAPABILITIES IN SAS*

Mark Janssens

I-BioStat / Hasselt University

PhUSE Annual Conference 15 Oct 2013



Bayesian Statistics

- ➡ • Concept
- Bayesian capabilities in SAS

Intro

- Classical inference

- “p-value”: How likely is my data, given the null-hypothesis?
- $P(\text{data} | \beta)$

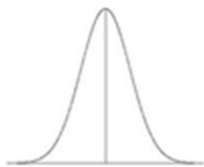
$$\tilde{\beta} \pm \text{SE}$$


- Bayesian inference

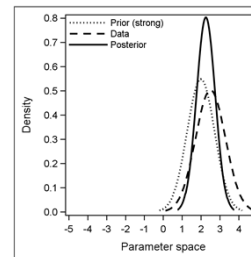
- How likely is my effect size β ? Can I include existing knowledge?

- $P(\beta | \text{data})$

$$\tilde{\beta} + \text{distribution}$$



- $P(\beta | \text{data}, \beta_0)$ with β_0 containing prior info on β



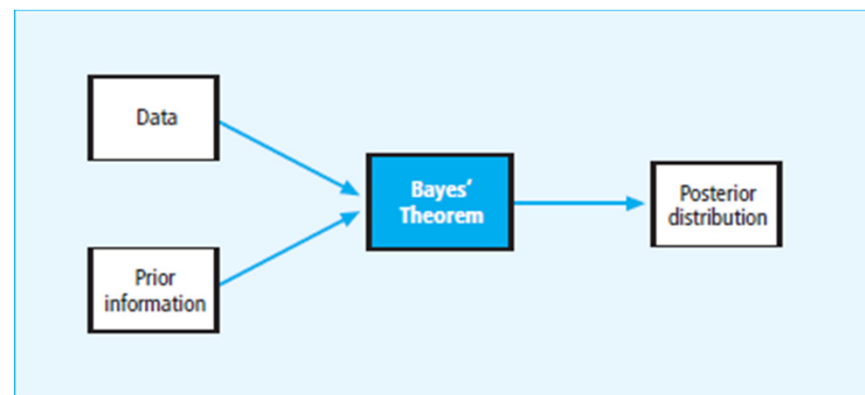
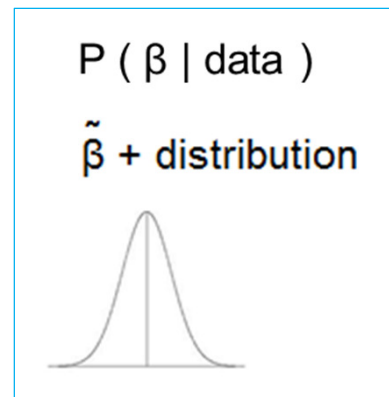
Definition

Bayesian methods are used to compute a

probability distribution of parameters in a statistical model,

using

observed data as well as **existing knowledge** about these parameters.



Use

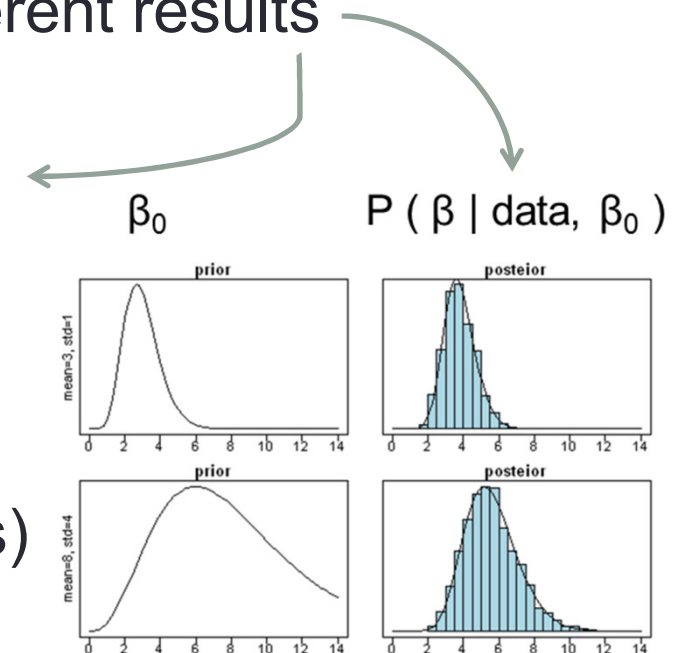
- Effect size β is influenced by “existing knowledge”

- Different prior information leads to different results

- Not used much in confirmatory phases

- Used in:

- Learning phases (dose finding)
- Drug effectiveness (health economics)
- ...

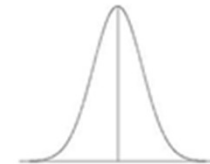


Example (health economics)

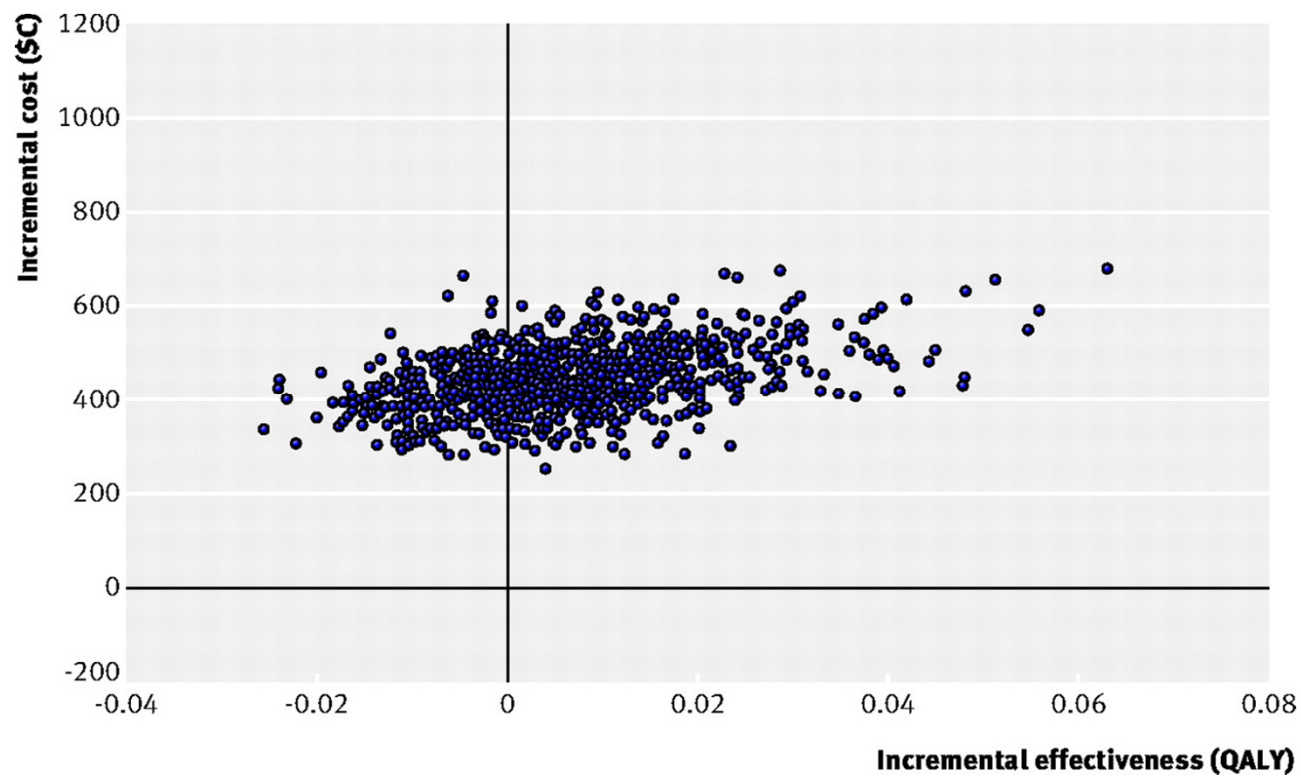
- **1** estimate for the relation between drug effectiveness and drug cost is not sufficient
- Probability estimates **s** are more adequate
- Strategies:
 - Deterministic:
 - Change one input parameters at a time
 - Build scenarios (best case / worst case)
 - Probabilistic:
 - Change multiple input parameters within “plausible ranges”
 - Replace point estimates of utilities & costs with probability distributions

$$\tilde{\beta} \pm SE$$


$\tilde{\beta}$ + distribution



Example (health economics)



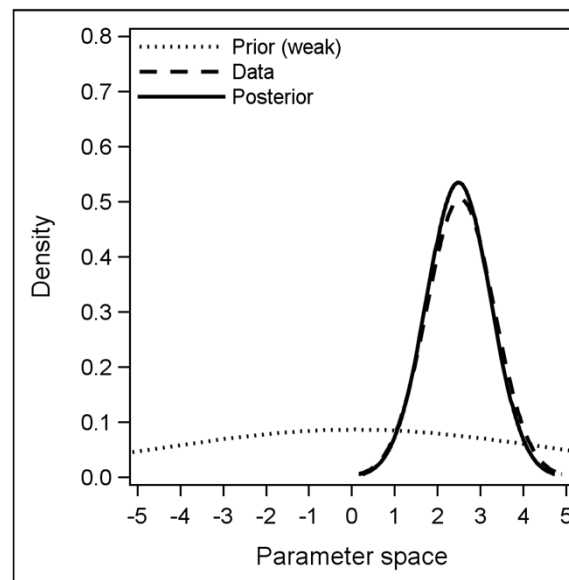
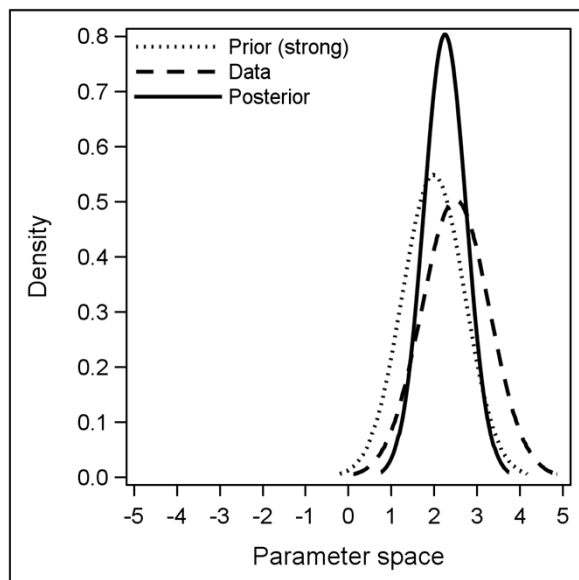
Population based screening for chronic kidney disease: cost effectiveness study (BMJ 2010;341:c5869)

Priors

- Conjugate prior

When the posterior distribution $f(\beta | X, \beta')$ is in the same family as the prior distribution $f(\beta')$, then the prior and posterior are called conjugate distributions

- Strong/Weak prior



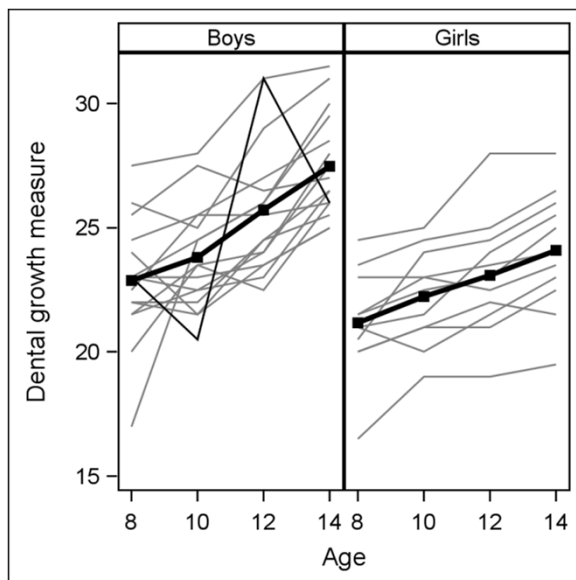
Bayesian Statistics

- Concept

➡ • Bayesian capabilities in SAS

Data

- Dental growth in 11 girls and 16 boys, measured at age 8, 10, 12, and 14

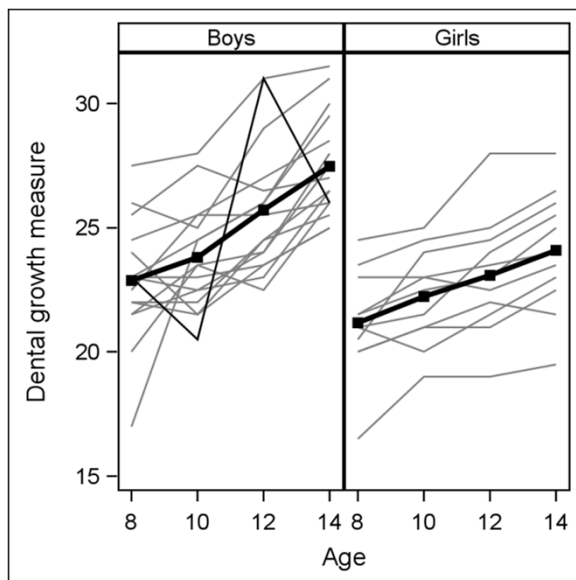


Response = growth increase of at least 10%

	Girls	Boys
Response=Yes	7 (64%)	13 (81%)
Response=No	4 (36%)	3 (19%)
	11 (100%)	16 (100%)

Data

- Dental growth in 11 girls and 16 boys, measured at age 8, 10, 12, and 14



Dental growth at age=14

Boys 27.5

Girls 24.0

Difference 3.5

(observed data)

Data

- Dental growth in 11 girls and 16 boys, measured at age 8, 10, 12, and 14

Odds Yes:No at age=14

Boys 13/3
Girls 7/4

Odds ratio 2.5

(observed data)

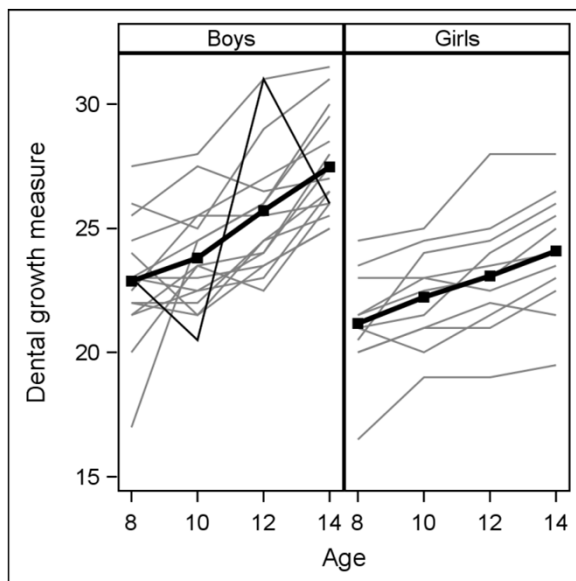
Response = growth increase of at least 10%

	Girls	Boys
Response=Yes	7 (64%)	13 (81%)
Response=No	4 (36%)	3 (19%)
	11 (100%)	16 (100%)

Models

- Is the *change from baseline* different between boys and girls?

- STATISTICAL MODEL 1 – LINEAR REGRESSION



- Continuous data
- Normal model

Models (see paper)

- Is the *response* different between boys and girls?
 - STATISTICAL MODEL 2 – LOGISTIC REGRESSION
 - STATISTICAL MODEL 3 – RANDOM EFFECTS LOGISTIC REGRESSION

Response = growth increase of at least 10%

	Girls	Boys
Response=Yes	7 (64%)	13 (81%)
Response=No	4 (36%)	3 (19%)
	11 (100%)	16 (100%)

- Binary data
- Binomial model

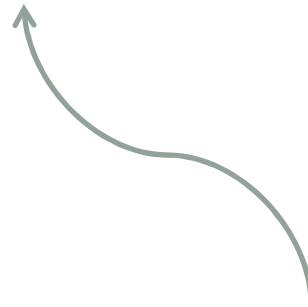
Linear Regression

- Is the *change from baseline* different between boys and girls?

$Y \sim \text{normal}(\mu; \sigma^2)$

$$\mu = \beta_0 + \beta_1 \text{YBASE} + \beta_2 \text{BOY}$$

- Direct likelihood
- Bayesian likelihood
- Bayesian likelihood incorporating prior evidence



PROC...
→ GENMOD
→ MCMC
→ MCMC

Direct Likelihood

$$Y \sim \text{normal}(\mu; \sigma^2)$$

$$\mu = \beta_0 + \beta_1 \text{ YBASE} + \beta_2 \text{ BOY}$$

```
proc genmod data=PERM.ANALYSIS_SET;
  where AGE=14;
  model Y = YBASE BOY / dist=normal;
run;
```

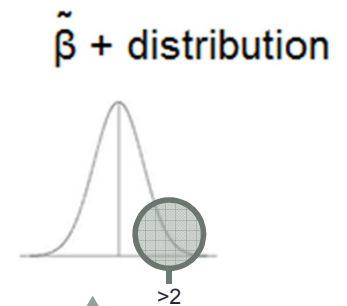
Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	13.4902	3.3826	6.8604	20.1201	15.90	<.0001
YBASE	1	0.5005	0.1575	0.1917	0.8093	10.09	0.0015
BOY	1	2.5305	0.7660	1.0292	4.0317	10.91	0.0010
Scale	1	1.8332	0.2495	1.4040	2.3935		

$$\tilde{\beta} \pm \text{SE}$$


The difference Boys vs Girls (β) is ~ 2.5 in this trial.

Bayesian likelihood

Based on this trial, what is the likelihood of $\beta > 2$?



- Option 1: PROC GENMOD

```
proc genmod data=PERM.ANALYSIS_SET;  
  where AGE=14;  
  model Y = YBASE BOY / dist=normal;  
  bayes nbi=1000 nmc=10000 thin=2 seed=159 cprior=jeffreys out=posterior;  
run;
```

- Option 2: PROC MCMC
next slide

Bayesian likelihood

- Option 2: PROC MCMC

```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=10000 thin=2 seed=159  
    monitor=(beta0-beta2 sigma2 beta2_gt_2);  
    where AGE=14;
```

- 1

```
parms beta0 13.49 beta1 0.50 beta2 2.53;  
parms sigma2 3.36;
```
- 2

```
prior beta0-beta2 ~ normal(mean = 0, var = 1000);  
prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);
```
- 3

```
mu = beta0 + beta1*YBASE + beta2*BOY;  
model Y ~ normal(mean = mu, var = sigma2);
```
- 4

```
beta2_gt_2 = beta2 > 2;
```

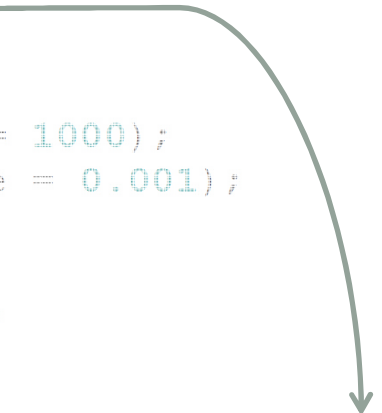
```
run;
```

Bayesian likelihood

- Option 2: PROC MCMC

```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=10000 thin=2 seed=159  
    monitor=(beta0-beta2 sigma2 beta2_gt_2);  
    where AGE=14;
```

```
1 parms beta0 13.49 beta1 0.50 beta2 2.53;  
   parms sigma2 3.36;  
  
2 prior beta0-beta2 ~ normal(mean = 0, var = 1000);  
   prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);  
  
3 mu = beta0 + beta1*YBASE + beta2*BOY;  
   model Y ~ normal(mean = mu, var = sigma2);  
  
4 beta2_gt_2 = beta2 > 2;  
  
run;
```



Maximum likelihood estimates
from PROC GENMOD.

Bayesian likelihood

- Option 2: PROC MCMC

```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=1000
      monitor=(beta0-beta2 sigma2 beta2_gt_2)
      where AGE=14;
```

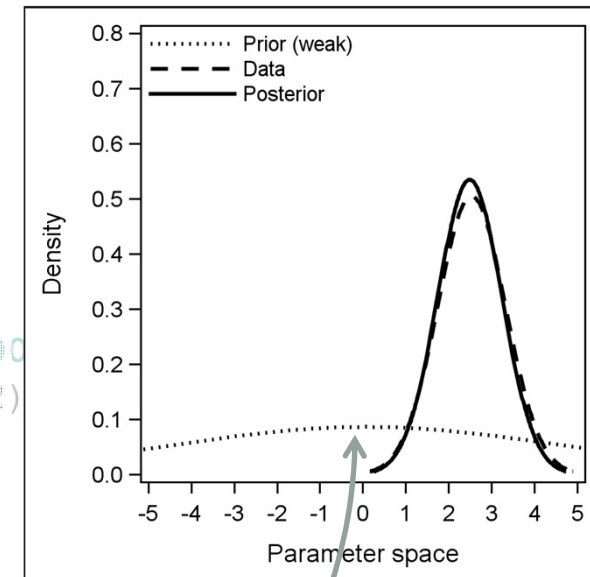
1 `parms beta0 13.49 beta1 0.50 beta2 2.53;`
`parms sigma2 3.36;`

2 `prior beta0-beta2 ~ normal(mean = 0, var = 1000);`
`prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);`

3 `mu = beta0 + beta1*YBASE + beta2*BOY;`
`model Y ~ normal(mean = mu, var = sigma2);`

4 `beta2_gt_2 = beta2 > 2;`

```
run;
```



Weak priors.
 Posterior point estimate will
 coincide with direct likelihood
 estimate.

Bayesian likelihood

- Option 2: PROC MCMC

```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=10000 thin=2 seed=159  
    monitor=(beta0-beta2 sigma2 beta2_gt_2);  
    where AGE=14;
```

```
1  parms beta0 13.49 beta1 0.50 beta2 2.53;  
    parms sigma2 3.36;
```

```
2  prior beta0-beta2 ~ normal(mean = 0, var = 1000);  
    prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);
```

```
3  mu = beta0 + beta1*YBASE + beta2*BOY;  
    model Y ~ normal(mean = mu, var = sigma2);
```

```
4  beta2_gt_2 = beta2 > 2;
```

```
run;
```


$$Y \sim \text{normal}(\mu; \sigma^2)$$
$$\mu = \beta_0 + \beta_1 \text{YBASE} + \beta_2 \text{BOY}$$

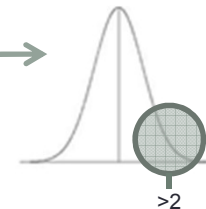
Bayesian likelihood

- Option 2: PROC MCMC

```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=10000 thin=2 seed=159  
    monitor=(beta0-beta2 sigma2 beta2_gt_2);  
    where AGE=14;
```

```
1  parms beta0 13.49 betal 0.50 beta2 2.53;  
    parms sigma2 3.36;  
  
2  prior beta0-beta2 ~ normal(mean = 0, var = 1000);  
    prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);  
  
3  mu = beta0 + betal*YBASE + beta2*BOY;  
    model Y ~ normal(mean = mu, var = sigma2);  
  
4  beta2_gt_2 = beta2 > 2;  
  
run;
```

$\tilde{\beta}$ + distribution



Uses β_2 posterior distribution.

Bayesian likelihood

Posterior Summaries						
Parameter	N	Standard		Percentiles		
		Mean	Deviation	25%	50%	75%
BETA0	5000	13.3587	3.5826	10.8873	13.3665	15.8217
BETA1	5000	0.5074	0.1666	0.3965	0.5112	0.6205
BETA2	5000	2.5018	0.8250	1.9534	2.4941	3.0815
SIGMA2	5000	4.0658	1.2536	3.2020	3.8352	4.6738
beta2_gt_2	5000	0.7326	0.4426	0	1.0000	1.0000

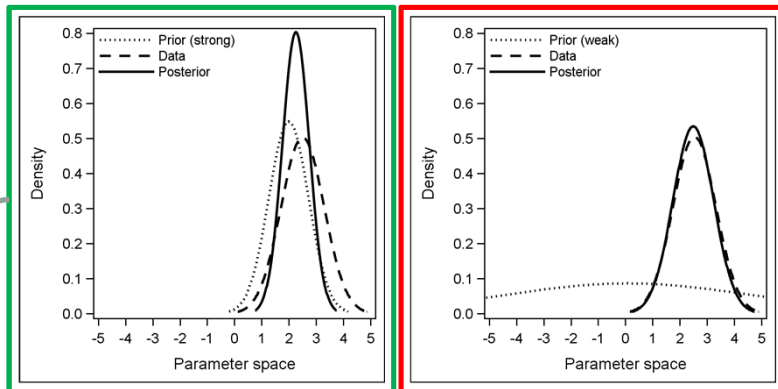
The difference Boys vs Girls (β) is ~ 2.5 in this trial.

The probability that the gender difference is at least 2, **based on the current data alone**, is 73%.

The gender effect was known to lie around 2.
Can we incorporate the existing knowledge?

Bayesian likelihood inc prior evidence

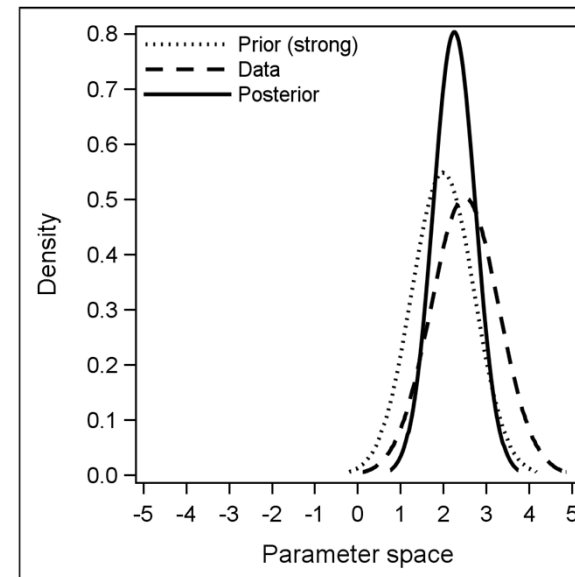
Only the prior statement changes...



```
proc mcmc data=PERM.ANALYSIS_SET nbi=1000 nmc=10000 thin=2 seed=159
    monitor=(beta0-beta2 sigma2 beta2_gt_2);
    where AGE=14;
    parms beta0 13.49 beta1 0.50 beta2 2.53;
    parms sigma2 3.36;
    prior beta0-beta1 ~ normal(mean = 0, var = 1000);
    prior beta2 ~ normal (mean = 2, var = 0.5);
    prior sigma2 ~ igamma(shape = 0.001, scale = 0.001);
    mu = beta0 + beta1*YBASE + beta2*BOY;
    model Y ~ normal(mean = mu, var = sigma2);
    beta2_gt_2 = beta2 > 2;
run;
```

Bayesian likelihood inc prior evidence

Parameter	N	Posterior Summaries				
		Mean	Standard Deviation	Percentiles		
				25%	50%	75%
BETA0	5000	12.9850	3.6821	10.4937	12.9318	15.4242
BETA1	5000	0.5306	0.1677	0.4204	0.5312	0.6435
BETA2	5000	2.2422	0.5540	1.8716	2.2367	2.6207
SIGMA2	5000	4.0383	1.2820	3.1455	3.8103	4.6395
beta2_gt_2	5000	0.6740	0.4688	0	1.0000	1.0000

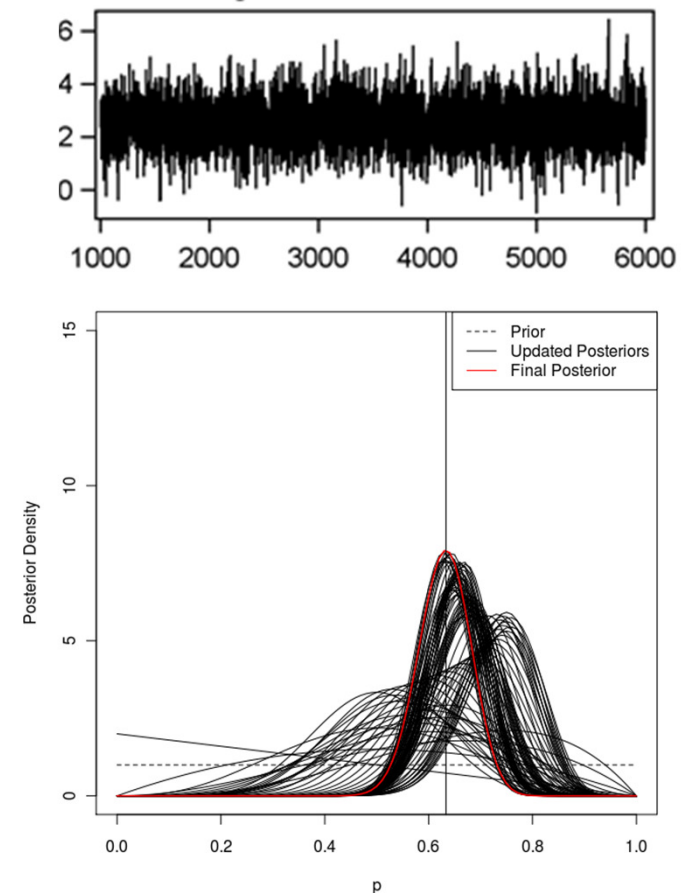


The posterior estimate for β_2 being greater than 2 should now be smaller than 73% (= result with weak prior).

The posterior estimate turns out to be 67%.

Diagnostics

- The posterior distribution is obtained through an iterative algorithm (Markov Chain Monte Carlo, or MCMC)
- Each MCMC step gives a value for your parameter(s)
- The posterior distribution is updated by each MCMC step
- If the updates do not reposition the posterior any longer, then posterior distribution is stationary
- MCMC convergence



Diagnostics

- How to assess MCMC convergence? Several tests, e.g.
 - **Geweke (standard in SAS)**

Tests whether the mean estimates have converged by comparing means from the early and latter part of the Markov chain.

 - z-test
 - high z is bad
 - **Gelman-Rubin (not standard in SAS)**

Uses parallel chains with dispersed initial values to test whether they all converge to the same target distribution.

 - variance ratio test
 - high R is bad

Diagnostics

- Tricks to speed convergence or to lower autocorrelation?
 - Center the data variables
 - Thin the chain
 - Block the model parameters, and/or
 - Reparameterize the model.

Take-away messages

- Maximum likelihood procedures such as PROC GENMOD provide readily available Bayesian functionality.
- More advanced statistical models can be fitted with PROC MCMC.
- With the introduction of the RANDOM statement in PROC MCMC, Bayesian random effect models have become easy to specify & run (see paper).
- Although not overly dealt with in this paper, Bayesian model fitting requires careful inspection of the model diagnostics, and advanced models require in-depth understanding of prior distributions (choice, construction, operational characteristics).