Bayes' Theorem

Let B_1, B_2,.....B_k be a set of mutually exclusive and exhaustive states.

Let A represent some event that happens.

Then, the probability of B_i on the condition that A occurs is given by Bayes' Theorem as:

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i) \times \Pr(B_i)}{\sum_{j=1}^{k} \Pr(A \mid B_j) \times \Pr(B_j)}$$

Example:

$Pr(B_i)$ values are the ***prior probabilities***

Let	B_1 = House Finch	$Pr(B_1) = 0.97$
	B_2 = Western Tanager	$Pr(B_2) = 0.02$
	B_3 = Ladder-backed Woodpecker	$Pr(B_3) = 0.01$

$Pr(B_i)$ is the probability of seeing species i on campus during spring migration

Let A = report from a student of seeing a bird with a red head on campus during spring migration

$Pr(A|B_i)$ are the ***likelihoods***. (Not to be confused with *maximum likelihood*)
Let $Pr(A|B_1) = 0.1$, $Pr(A|B_2) = 0.9$, $Pr(A|B_3) = 0.2$

$Pr(A|B_i)$ is the probability of a student giving report A if they have seen species B_i.

$Pr(B_i|A)$ are the ***posterior probabilities***
From Bayes Theorem, the posterior probability of a bird being a House Finch if a student gives report A is:

$Pr(B_1|A)$ = 0.1 * 0.97 / [0.1 * 0.97 + 0.9 * 0.02 + 0.2 * 0.01] = 0.097/0.117 = **0.829**

The probability of the bird being a Western Tanager is:

$Pr(B_2|A)$ = 0.9 * 0.02 / 0.117 = 0.018 /0 .117 = **0.154**

The probability of the bird being a Ladder-backed Woodpecker is:

$Pr(B_3|A)$ = 0.2 * .01 / .117 = .002 / .117 = **0.017**

Bayes, Thomas (b. 1702, London - d. 1761, Tunbridge Wells, Kent), mathematician who first used probability inductively and established a mathematical basis for probability. He set down his findings on probability in "Essay Towards Solving a Problem in the Doctrine of Chances" (1763), published posthumously in the *Philosophical Transactions of the Royal Society of London.* He was a Presbyterian minister in Tunbridge Wells from 1731. It is thought that his election to the Royal Society might have been based on a tract of 1736 in which Bayes defended the views and philosophy of Sir Isaac Newton. A notebook of his exists, and includes a method of finding the time and place of conjunction of two planets, notes on weights and measures, a method of differentiation, and logarithms.

Overview of Bayesian Analysis – most of the material is from SAS 9.2 Documentation

The most frequently used statistical methods are known as ***frequentist*** (or ***classical***) methods. These methods assume that unknown parameters are fixed constants, and they define probability by using limiting relative frequencies. It follows from these assumptions that probabilities are objective and that you cannot make probabilistic statements about parameters because they are fixed.

Bayesian methods offer an alternative approach; they treat parameters as random variables and define probability as "degrees of belief" (that is, the probability of an event is the degree to which you believe the event is true). It follows from these postulates that probabilities are subjective and that you can make probability statements about parameters. The term "Bayesian" comes from the prevalent usage of Bayes' theorem, which was named after the Reverend Thomas Bayes, an eighteenth century Presbyterian minister. Bayes was interested in solving the question of inverse probability: after observing a collection of events, what is the probability of one event?

Suppose you are interested in estimating θ from data $\mathbf{y} = \{y_1, \ldots, y_n\}$ by using a statistical model described by a density $p(\mathbf{y}|\theta)$. Bayesian philosophy states that θ cannot be determined exactly, and uncertainty about the parameter is expressed through probability statements and distributions. You can say that θ follows a normal distribution with mean 0 and variance 1, if it is believed that this distribution best describes the uncertainty associated with the parameter. The following steps describe the essential elements of Bayesian inference:

1. A probability distribution for θ is formulated as $\pi(\theta)$, which is known as the ***prior*** distribution, or just the prior. The prior distribution expresses your beliefs (for example, on the mean, the spread, the skewness, and so forth) about the parameter before you examine the data.
2. Given the observed data $\mathbf{y}$, you choose a statistical model $p(\mathbf{y}|\theta)$ to describe the distribution of $\mathbf{y}$ given θ.
3. You update your beliefs about θ by combining information from the prior distribution and the data through the calculation of the ***posterior*** distribution, $p(\theta|\mathbf{y})$.

Simply put, Bayes' theorem tells you how to update existing knowledge with new information. You begin with a prior belief $\pi(\theta)$, and after learning information from data $\mathbf{y}$, you change or update your belief about θ and obtain $p(\theta|\mathbf{y})$. These are the essential elements of the Bayesian approach to data analysis.

In theory, Bayesian methods offer simple alternatives to statistical inference—all inferences follow from the posterior distribution $p(\theta|\mathbf{y})$. In practice, however, you can obtain the posterior distribution with straightforward analytical solutions only in the most rudimentary problems. Most Bayesian analyses require sophisticated computations, including the use of simulation methods. You generate samples from the posterior distribution and use these samples to estimate the quantities of interest.

Prior Distributions

A prior distribution of a parameter is the probability distribution that represents your uncertainty about the parameter before the current data are examined. Multiplying the prior distribution and the likelihood function together leads to the posterior distribution of the parameter. You use the posterior distribution to carry out all inferences. You cannot carry out any Bayesian inference or perform any modeling without using a prior distribution.

Objective Priors versus Subjective Priors

Bayesian probability measures the degree of belief that you have in a random event. By this definition, probability is highly subjective. It follows that all priors are ***subjective priors***. Not everyone agrees with this notion of subjectivity when it comes to specifying prior distributions. There has long been a desire to obtain results that are objectively valid. Within the Bayesian paradigm, this can be somewhat achieved by using prior distributions that are "objective" (that is, that have a minimal impact on the posterior distribution). Such distributions are called ***objective*** or ***noninformative*** priors. However, while noninformative priors are very popular in some applications, they are not always easy to construct.

Bayesian Inference

Bayesian inference about θ is primarily based on the posterior distribution of θ. There are various ways in which you can summarize this distribution. For example, you can report your findings through point estimates. You can also use the posterior distribution to construct hypothesis tests or probability statements.

If you know the distributional form of the posterior density of interest, you can report the exact posterior point estimates. When models become too difficult to analyze analytically, you have to use simulation algorithms, such as the Markov chain Monte Carlo (MCMC) method to obtain posterior estimates.

Markov Chain Monte Carlo method

The Markov chain Monte Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. MCMC methods sample successively from a target distribution. Each sample depends on the previous one, hence the notion of the Markov chain. A Markov chain is a sequence of random variables, θ^{1}, θ^{2}, ..., for which the random variable θ^{i} depends on all previous θs only through its immediate predecessor θ^{i-1}. You can think of a Markov chain applied to sampling as a mechanism that traverses randomly through a target distribution without having any memory of where it has been. Where it moves next is entirely dependent on where it is now.

Algorithms that implement the MCMC method include the Metropolis, Metropolis-Hastings, and Gibbs sampler. (Note: Metropolis refers to American physicist and computer scientist Nicholas C. Metropolis).

Burn In and Thinning

Burn-in refers to the practice of discarding an initial portion of a Markov chain sample so that the effect of initial values on the posterior inference is minimized. For example, suppose the target distribution is $N(0, 1)$ and the Markov chain was started at the value 10^{6}. The chain might quickly travel to regions around 0 in a few iterations. However, including samples around the value 10^{6} in the posterior mean calculation can produce substantial bias in the mean estimate. In theory, if the Markov chain is run for an infinite amount of time, the effect of the initial values decreases to zero. In practice, you do not have the luxury of infinite samples. In practice, you assume that after t iterations, the chain has reached its target distribution and you can throw away the early portion and use the good samples for posterior inference. The value of t is the burn-in number.

With some models you might experience poor mixing (or slow convergence) of the Markov chain. This can happen, for example, when parameters are highly correlated with each other. Poor mixing means that the Markov chain slowly traverses the parameter space and the chain has high dependence. High sample autocorrelation can result in biased Monte Carlo standard errors. A common strategy is to ***thin*** the Markov chain in order to reduce sample autocorrelations. You thin a chain by keeping every kth simulated draw from each sequence. You can safely use a thinned Markov chain for posterior inference as long as the chain converges. It is important to note that thinning a Markov chain can be wasteful because you are throwing away a $\frac{k-1}{k}$ fraction of all the posterior samples generated. You always get more precise posterior estimates if the entire Markov chain is used. However, other factors, such as computer storage or plotting time, might prevent you from keeping all samples.

Advantages and Disadvantages

Bayesian methods and classical methods both have advantages and disadvantages, and there are some similarities. When the sample size is large, Bayesian inference often provides results for parametric models that are very similar to the results produced by frequentist methods. Some advantages to using Bayesian analysis include the following:

It provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. You can incorporate past information about a parameter and form a prior distribution for future analysis. When new observations become available, the previous posterior distribution can be used as a prior. All inferences logically follow from Bayes' theorem.

It provides inferences that are conditional on the data and are exact, without reliance on asymptotic approximation. Small sample inference proceeds in the same manner as if one had a large sample.

It obeys the likelihood principle. If two distinct sampling designs yield proportional likelihood functions for θ, then all inferences about θ should be identical from these two designs. Classical inference does not in general obey the likelihood principle.

It provides interpretable answers, such as "the true parameter θ has a probability of 0.95 of falling in a 95% credible interval."

It provides a convenient setting for a wide range of models, such as hierarchical models and missing data problems. MCMC, along with other numerical methods, makes computations tractable for virtually all parametric models.

There are also disadvantages to using Bayesian analysis:

It does not tell you how to select a prior. There is no correct way to choose a prior. Bayesian inferences require skills to translate subjective prior beliefs into a mathematically formulated prior. If you do not proceed with caution, you can generate misleading results.

It can produce posterior distributions that are heavily influenced by the priors. From a practical point of view, it might sometimes be difficult to convince subject matter experts who do not agree with the validity of the chosen prior.

It often comes with a high computational cost, especially in models with a large number of parameters. In addition, simulations provide slightly different answers unless the same random seed is used. Note that slight variations in simulation results do not contradict the early claim that Bayesian inferences are exact. The posterior distribution of a parameter is exact, given the likelihood function and the priors, while simulation-based estimates of posterior quantities can vary due to the random number generator used in the procedures.

[We will skip to the example (below) at this point, and return here after examining the example.]

Bayesian Analysis of Phylogenies – most of the material is from Hall, BG, 2004. Phylogenetic Trees Made Easy: A How To Manual, Sinauer Assoc., Sunderland, MA.

In phylogenetics, Bayesian analysis (B) is related to the Maximum Likelihood method (ML). You select a model of evolution, and the computer searches for the best trees relative to the model and the data (the alignment).

ML seeks the tree that maximizes the probability of observing the alignment given the tree.

B seeks the tree that maximizes the probability of the tree, given the alignment and the model of evolution. This "re-scales" likelihoods to true probabilities (sum over all trees = 1). This permits using probability to analyze the data.

ML seeks the single most likely tree. As it searches a "landscape" of possible trees, it continually seeks higher points on the landscape (more likely trees). If there is more than one hill, ML can get trapped on a peak, even though there may be a higher peak somewhere. ML cannot traverse the "valleys" to get to the higher peak.

B seeks the best set of trees, therefore B may consider the same tree many times. B searches the landscape by MCMC methods (the MrBayes program uses a Metropolis-coupled algorithm to implement the MCMC). The probability of trees usually cannot be calculated analytically. The MCMC method allows the calculation of probabilities by sampling the posterior distribution of tree probabilities.

More detail…….

B begins with either a randomly chosen or user-specified tree. This tree has a combination of branch lengths, nucleotide substitution parameters, and rate variation across sites parameter. This defines the initial state of a chain.

A new state (new tree) of the chain is proposed. This involves moving a branch or changing the length of a branch to create a modified tree.

The probability of the new state, given the old state, is calculated. A random number between 0 and 1 is drawn. If the calculated probability is greater than the random number the new tree is accepted, otherwise the tree remains the same. We have now completed one generation of the chain.

The new state involves moving a branch and/or changing the length of a branch. If the new state is more likely (given the model and data) it is more likely to be accepted. Over many generations, the program moves toward a more likely tree, but not every step yields a more likely tree. Some moves pick a less likely tree.

Eventually, the process converges on a set of trees that have very similar likelihoods. Accepting or rejecting a change becomes essentially random. The chain has converged on a **stable likelihood**. The frequency with which various trees are sampled is very close to the frequency of those trees in the likelihood distribution. The relative frequencies are the probabilities that a particular tree is the best tree.

Just as ML can get trapped on a peak that is not the highest peak (most likely), so can the B process get trapped. One way a B algorithm deals with this is to run several (often four) independent chains at once. Due to the random nature of the MCMC process, the chains quickly diverge.

One of the chains is designated as the “cold chain”, which will be the source of the final trees once it has reached the stable likelihood. Every generation, there is a probability for two chains to swap states. Thus, if the cold chain gets trapped on a suboptimal peak, it may escape by swapping with a chain that is on a higher peak. This does not guarantee you always get to the highest peak, but it does help.

It is not necessary to bootstrap a tree to estimate clade support. The probabilities determined from the MCMC process serve that purpose.

YOU MAY STOP READING HERE. Below is an example of Bayesian analysis run on SAS. We'll go over this in the discussion. If you are familiar with SAS, you might want to look through it before the discussion.

The data for this example are birth weights for 36 infants. Each baby is categorized by smoking status of mom during pregnancy: 1=nonsmoking; 2=smoke up to 1 pack/day; 3=smoke 1+ pack/day. There are 12 babies in each smoking group. SAS creates a dummy variable for each smoking category, plus one for the intercept.
Here's how the data look to SAS:

	Intercept	Smoking 1	Smoking 2	Smoking 3	Birth Weight
Nonsmoking Babies (11 more like this)	1	1	0	0	3515
1 pack/day Babies (11 more like this)	1	0	1	0	3444
1+ pack/day Babies (11 more like this)	1	0	0	1	2608

The model is: Birth Weight $= \beta_0 + \beta_1$ Smoking 1 + β_2 Smoking 2 + β_3 Smoking 3

β_0 is the intercept. We are interested here only in how the β parameters are estimated.

```
DATA BABYWT;
INPUT SMOKING WEIGHT @@; *The @@ allows multiple observations per line;
DATALINES;
1 3515 1 3420 1 3175 1 3586 1 3232 1 3884 1 3856 1 3941 1 3232 1 4054 1 3459 1 3998
2 3444 2 3827 2 3884 2 3515 2 3416 2 3742 2 3062 2 3076 2 2835 2 2750 2 3460 2 3340
3 2608 3 2509 3 3600 3 1730 3 3175 3 3459 3 3288 3 2920 3 3020 3 2778 3 2466 3 3260
;
PROC GLM;
TITLE2 'OLS - Ordinary Least Squares';
CLASS SMOKING;
MODEL WEIGHT = SMOKING / SS3 SOLUTION;
RUN;
PROC GENMOD;
TITLE2 'Likelihood';
CLASS SMOKING;
MODEL WEIGHT = SMOKING / dist=normal;
RUN;
PROC GENMOD;
TITLE2 'Bayes - Normal Prior';
CLASS SMOKING;
MODEL WEIGHT = SMOKING / dist=normal;
bayes seed=1 coeffprior=normal ;
RUN;
PROC GENMOD;
TITLE2 'Bayes - Uniform Prior';
CLASS SMOKING;
MODEL WEIGHT = SMOKING / dist=normal;
bayes seed=1 coeffprior=uniform OUTPOST = postsmoke;
RUN;
data bayes2;
set postsmoke;
if Iteration < 2026;
run;
title3 'First 25 values (out of 10000) from MCMC (after burn-in of 2000 iterations)';
proc print noobs;
var Iteration    Intercept    SMOKING1    SMOKING2    SMOKING3;
run;
```

The GLM Procedure

Dependent Variable: WEIGHT

Class Level Information		
Class	**Levels**	**Values**
SMOKING	3	1 2 3

Number of Observations Read	36
Number of Observations Used	36

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3127499.056	1563749.528	9.18	0.0007
Error	33	5619874.500	170299.227		
Corrected Total	35	8747373.556			

R-Square	Coeff Var	Root MSE	WEIGHT Mean
0.357536	12.53522	412.6733	3292.111

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SMOKING	2	3127499.056	1563749.528	9.18	0.0007

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	2901.083333	B	119.1285116	24.35	<.0001
SMOKING 1	711.583333	B	168.4731568	4.22	0.0002
SMOKING 2	461.500000	B	168.4731568	2.74	0.0099
SMOKING 3	0.000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Baby Birth weights - OLS , Likelihood, Bayes Likelihood

The GENMOD Procedure

Model Information	
Data Set	WORK.BABYWT
Distribution	Normal
Link Function	Identity
Dependent Variable	WEIGHT

Number of Observations Read	36
Number of Observations Used	36

Class Level Information		
Class	**Levels**	**Values**
SMOKING	3	1 2 3

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	33	5619874.5000	170299.2273
Scaled Deviance	33	36.0000	1.0909
Pearson Chi-Square	33	5619874.5000	170299.2273
Scaled Pearson X2	33	36.0000	1.0909
Log Likelihood		-266.3312	
Full Log Likelihood		-266.3312	
AIC (smaller is better)		540.6624	
AICC (smaller is better)		541.9527	
BIC (smaller is better)		546.9965	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	2901.083	114.0569	2677.536	3124.631	646.96	<.0001
SMOKING	1	1	711.5833	161.3008	395.4396	1027.727	19.46	<.0001
SMOKING	2	1	461.5000	161.3008	145.3563	777.6437	8.19	0.0042
SMOKING	3	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	395.1046	46.5635	313.6151	497.7682		

Note: The scale parameter was estimated by maximum likelihood

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Normal Prior

The GENMOD Procedure

Bayesian Analysis

Model Information	
Data Set	WORK.BABYWT
Burn-In Size	2000
MC Sample Size	10000
Thinning	1
Distribution	Normal
Link Function	Identity
Dependent Variable	WEIGHT

Number of Observations Read	36
Number of Observations Used	36

Class Level Information		
Class	**Levels**	**Values**
SMOKING	3	1 2 3

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**	
Intercept		1	2901.083	114.0569	2677.536	3124.631
SMOKING	1	1	711.5833	161.3008	395.4396	1027.727
SMOKING	2	1	461.5000	161.3008	145.3563	777.6437
SMOKING	3	0	0.0000	0.0000	0.0000	0.0000
Scale		1	395.1046	46.5635	313.6151	497.7682

Note: The scale parameter was estimated by maximum likelihood.

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Normal Prior

The GENMOD Procedure

Bayesian Analysis

Independent Normal Prior for Regression Coefficients		
Parameter	**Mean**	**Precision**
Intercept	0	1E-6
SMOKING1	0	1E-6
SMOKING2	0	1E-6

Algorithm converged.

Independent Prior Distributions for Model Parameters			
		Hyperparameters	
Parameter	**Prior Distribution**	**Shape**	**Inverse Scale**
Scale	Gamma	0.001	0.001

Initial Values of the Chain						
Chain	**Seed**	**Intercept**	**SMOKING1**	**SMOKING2**	**SMOKING3**	**Scale**
1	1	2880.014	723.5775	476.592	0	387.9367

Fit Statistics	
AIC (smaller is better)	540.662
AICC (smaller is better)	541.953
BIC (smaller is better)	546.996
DIC (smaller is better)	540.653
pD (effective number of parameters)	3.852

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Normal Prior

The GENMOD Procedure

Bayesian Analysis

Posterior Summaries						
				Percentiles		
Parameter	N	Mean	Standard Deviation	25%	50%	75%
Intercept	10000	2880.8	119.2	2802.7	2882.3	2959.9
SMOKING1	10000	718.6	169.7	606.1	717.9	830.7
SMOKING2	10000	475.1	169.4	364.3	475.0	584.3
Scale	10000	419.3	52.8488	381.7	413.5	451.1

Posterior Intervals					
		Equal-Tail Interval		HPD Interval	
Parameter	Alpha				
Intercept	0.050	2644.6	3112.3	2643.2	3110.4
SMOKING1	0.050	388.4	1049.4	389.2	1050.0
SMOKING2	0.050	143.0	814.6	145.1	816.1
Scale	0.050	331.9	540.7	321.3	523.2

Posterior Correlation Matrix				
Parameter	Intercept	SMOKING1	SMOKING2	Scale
Intercept	1.000	-0.700	-0.696	-0.047
SMOKING1	-0.700	1.000	0.478	0.024
SMOKING2	-0.696	0.478	1.000	0.009
Scale	-0.047	0.024	0.009	1.000

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Normal Prior

The GENMOD Procedure

Bayesian Analysis

Posterior Autocorrelations				
Parameter	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**
Intercept	0.6556	0.1057	-0.0035	-0.0002
SMOKING1	0.4796	0.0743	0.0024	0.0064
SMOKING2	0.4912	0.0882	-0.0019	-0.0016
Scale	0.0730	0.0122	-0.0104	0.0106

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-0.2395	0.8108
SMOKING1	0.9478	0.3432
SMOKING2	-0.5429	0.5872
Scale	1.7982	0.0722

Effective Sample Sizes			
Parameter	**ESS**	**Correlation Time**	**Efficiency**
Intercept	2225.9	4.4925	0.2226
SMOKING1	2842.7	3.5178	0.2843
SMOKING2	2725.5	3.6690	0.2726
Scale	8523.7	1.1732	0.8524

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Uniform Prior

The GENMOD Procedure

Bayesian Analysis

Model Information	
Data Set	WORK.BABYWT
Burn-In Size	2000
MC Sample Size	10000
Thinning	1
Distribution	Normal
Link Function	Identity
Dependent Variable	WEIGHT

Number of Observations Read	36
Number of Observations Used	36

Class Level Information		
Class	**Levels**	**Values**
SMOKING	3	1 2 3

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**	
Intercept		1	2901.083	114.0569	2677.536	3124.631
SMOKING	1	1	711.5833	161.3008	395.4396	1027.727
SMOKING	2	1	461.5000	161.3008	145.3563	777.6437
SMOKING	3	0	0.0000	0.0000	0.0000	0.0000
Scale		1	395.1046	46.5635	313.6151	497.7682

Note: The scale parameter was estimated by maximum likelihood.

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Uniform Prior

The GENMOD Procedure

Bayesian Analysis

Uniform Prior for Regression Coefficients	
Parameter	**Prior**
Intercept	Constant
SMOKING1	Constant
SMOKING2	Constant

Algorithm converged.

Independent Prior Distributions for Model Parameters			
		Hyperparameters	
Parameter	**Prior Distribution**	**Shape**	**Inverse Scale**
Scale	Gamma	0.001	0.001

Initial Values of the Chain						
Chain	**Seed**	**Intercept**	**SMOKING1**	**SMOKING2**	**SMOKING3**	**Scale**
1	1	2901.083	711.5833	461.5	0	387.3467

Fit Statistics	
AIC (smaller is better)	540.662
AICC (smaller is better)	541.953
BIC (smaller is better)	546.996
DIC (smaller is better)	540.683
pD (effective number of parameters)	3.884

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Uniform Prior

The GENMOD Procedure

Bayesian Analysis

Posterior Summaries						
				Percentiles		
Parameter	N	Mean	Standard Deviation	25%	50%	75%
Intercept	10000	2904.1	120.6	2824.9	2905.9	2983.2
SMOKING1	10000	706.1	171.7	590.7	706.2	817.5
SMOKING2	10000	458.7	172.3	345.4	457.1	572.4
Scale	10000	419.7	52.4838	382.6	414.3	450.9

Posterior Intervals					
		Equal-Tail Interval		HPD Interval	
Parameter	Alpha				
Intercept	0.050	2664.3	3141.0	2666.4	3142.3
SMOKING1	0.050	366.1	1052.3	372.1	1057.5
SMOKING2	0.050	119.0	802.1	108.4	789.8
Scale	0.050	332.1	538.2	322.7	523.9

Posterior Correlation Matrix				
Parameter	Intercept	SMOKING1	SMOKING2	Scale
Intercept	1.000	-0.705	-0.704	0.013
SMOKING1	-0.705	1.000	0.484	-0.021
SMOKING2	-0.704	0.484	1.000	-0.010
Scale	0.013	-0.021	-0.010	1.000

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Uniform Prior

The GENMOD Procedure

Bayesian Analysis

Posterior Autocorrelations				
Parameter	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**
Intercept	0.6686	0.1351	0.0219	-0.0081
SMOKING1	0.4980	0.1025	0.0095	-0.0185
SMOKING2	0.5020	0.1037	0.0149	0.0016
Scale	0.0807	-0.0165	-0.0135	0.0001

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-1.0668	0.2861
SMOKING1	1.0286	0.3037
SMOKING2	0.5529	0.5803
Scale	0.9814	0.3264

Effective Sample Sizes			
Parameter	**ESS**	**Correlation Time**	**Efficiency**
Intercept	1965.1	5.0887	0.1965
SMOKING1	2534.9	3.9450	0.2535
SMOKING2	2459.7	4.0656	0.2460
Scale	8264.9	1.2099	0.8265

Baby Birth weights - OLS , Likelihood, Bayes
Bayes - Uniform Prior

The GENMOD Procedure

Bayesian Analysis

Below are the First 25 values (out of 10000) of the parameters from MCMC (after burn-in of 2000 iterations). Notice this allows the calculation of probabilities. For example: " What is the probability a mom who smokes 1+ pack/day will have a baby weighing less than 2900 g?" Simply determine what proportion of the 8000 (10000 – 2000) Intercept values are less than 2900.

Iteration	Intercept	SMOKING1	SMOKING2	SMOKING3
2001	2759.16	795.9475	781.407	0
2002	2748.758	960.295	486.6837	0
2003	2763.78	672.6066	462.5101	0
2004	3077.469	591.2895	202.5567	0
2005	2908.902	826.9714	466.1523	0
2006	2833.719	794.0175	299.194	0
2007	2858.713	762.1387	511.7358	0
2008	2790.134	806.9293	762.6325	0
2009	2709.928	972.9401	758.457	0
2010	2785.718	827.1389	595.2235	0
2011	2787.232	1018.465	573.945	0
2012	2818.456	937.4654	418.3222	0
2013	2798.74	829.6059	720.569	0
2014	2895.175	726.943	514.0472	0
2015	2918.645	666.4316	381.4124	0
2016	2982.129	613.3209	569.4569	0
2017	2947.873	559.7913	282.2882	0
2018	3100.379	576.1784	203.7038	0
2019	3053.526	206.0228	140.4035	0
2020	3126.473	498.0764	312.586	0
2021	3085.487	529.0946	327.2274	0
2022	3024.261	568.6052	299.7321	0
2023	2915.458	742.4472	448.6989	0
2024	2865.359	690.9192	375.7048	0
2025	2936.446	749.8947	452.8128	0