



Bayesian Analysis

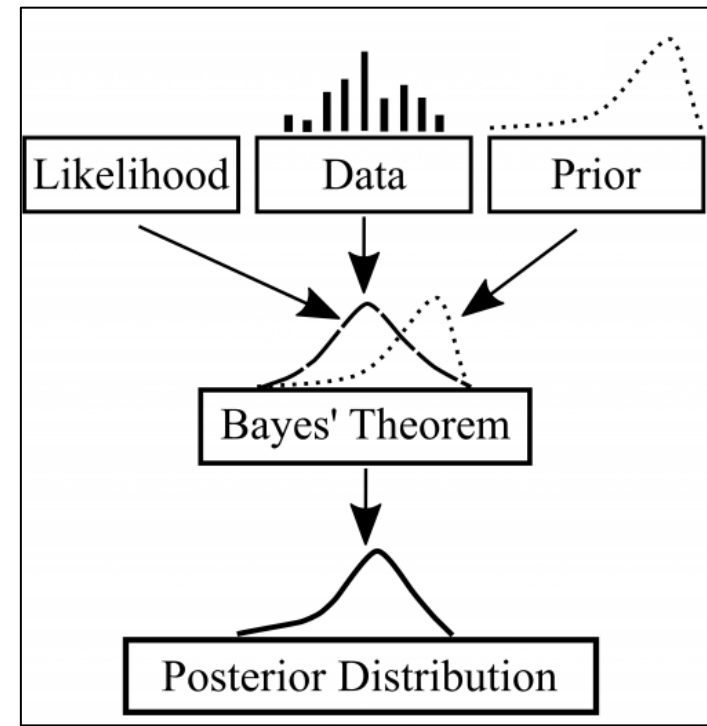
Module II: Leaves and Trees

Dr. Mark Williamson
DaCCoTA
University of North Dakota

Introduction



- Last time, we covered the history and basic theory of Bayesian Analysis, as well as comparison to Frequentist methods
- Today we'll cover more details about the computational process
- We will also look at examples in R, SAS, and SPSS for t-tests, ANOVA, and regressions



Theory

Bayes' Theorem

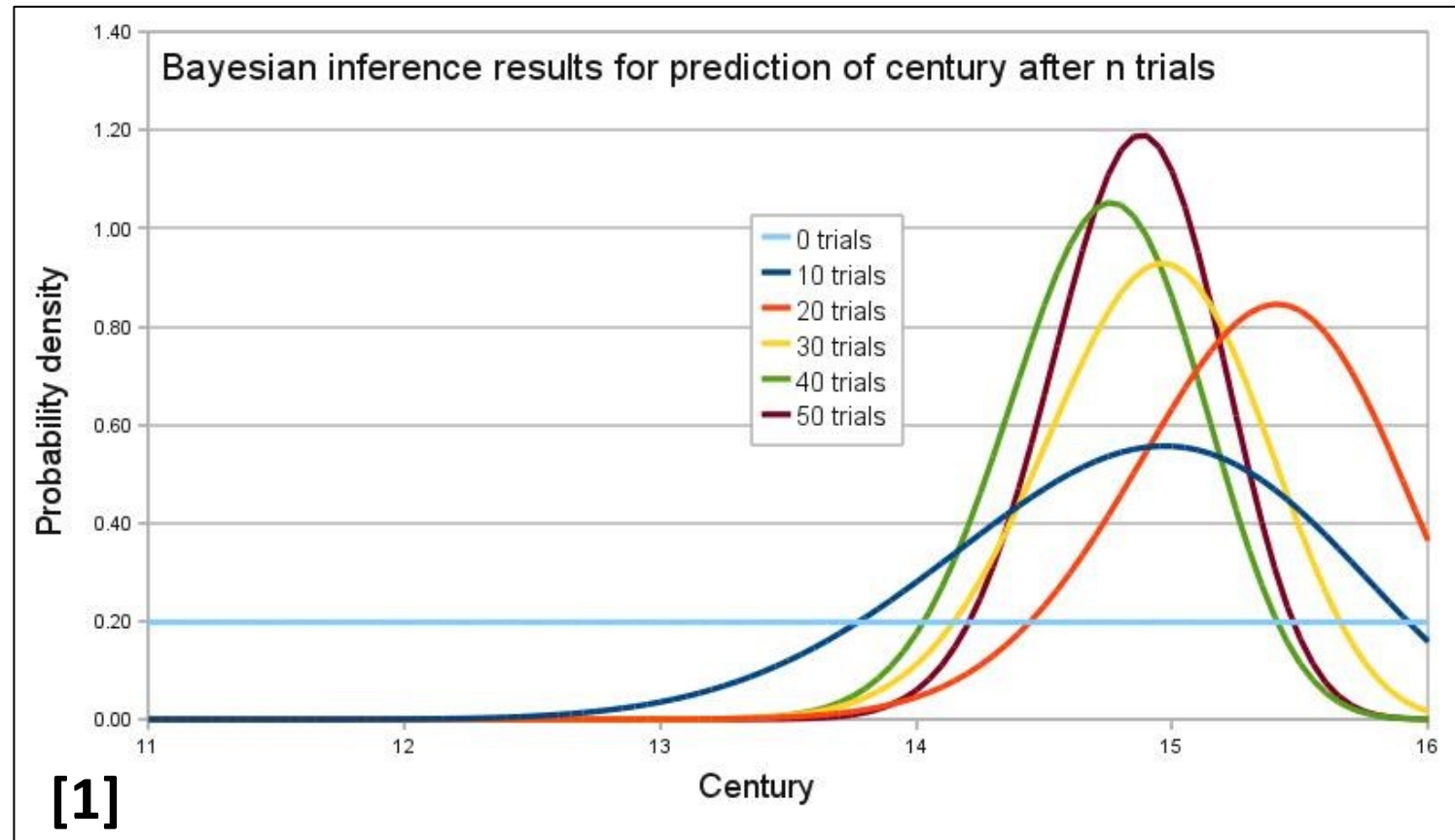
- Updates probabilities (degrees of belief) after obtaining new data

Bayes Theorem:

$$P(\text{hyp}|\text{data}) = \frac{P(\text{data}|\text{hyp}) \times P(\text{hyp})}{P(\text{data})}$$

Bayes Theorem:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$



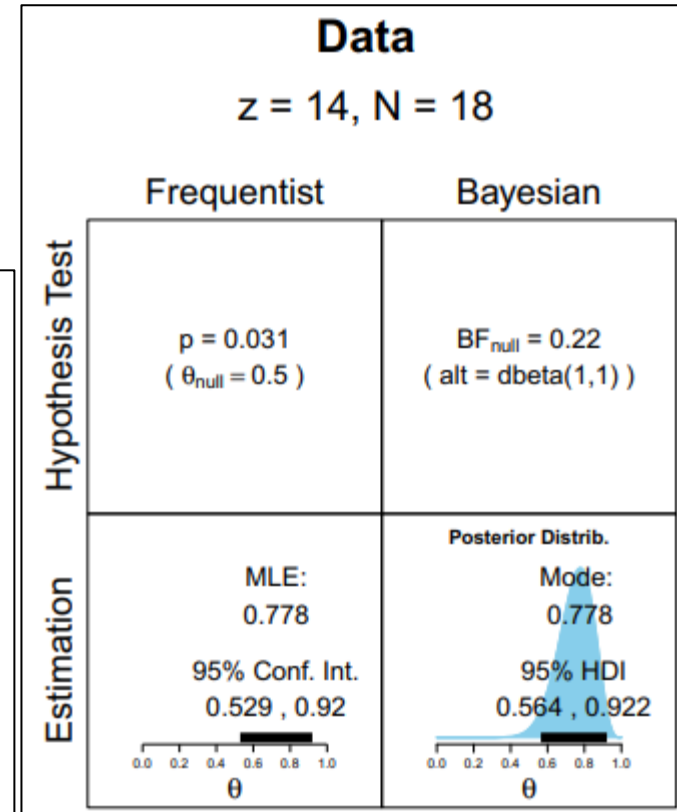
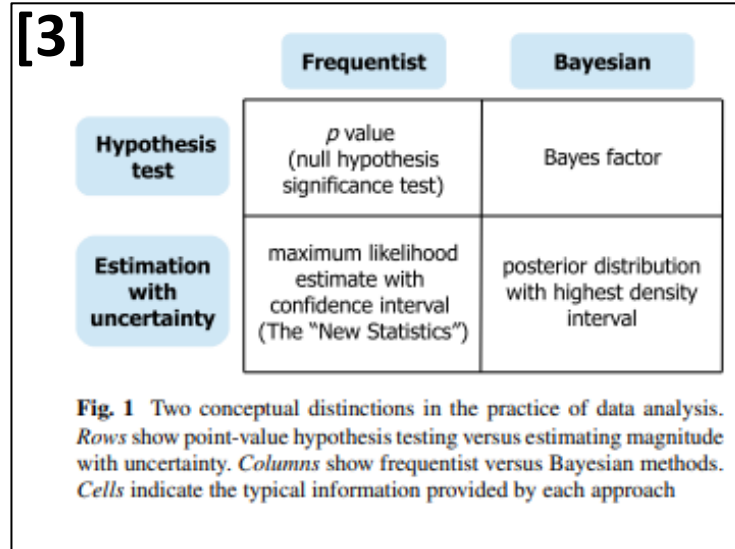
Descriptions

Hypothesis testing vs. parameter estimation [2,3]

Debate on the process [4]

Bayes Factor (K)

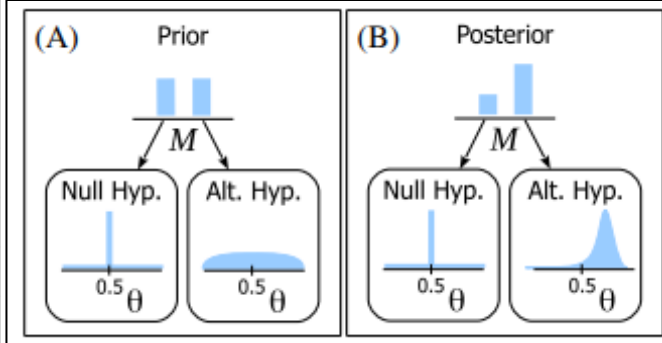
- Analogous to p-value or full vs reduced model
- Way to formally compare two competing models [5]
- Nesting not necessary
- Ratio of posterior odds for M1 to prior odds for M1 -> so factor >> 1 supports M1 over M2



Highest density interval analogous to MLE with 95% CI

[6]

K (exp notation)	K (std notation)	Strength
$< 10^0$	< 1	Negative (supports M2)
10^0 to $10^{1/2}$	1 to 3.16	Barely worth mentioning
$10^{1/2}$ to 10^1	3.16 to 10	Substantial
10^1 to $10^{3/2}$	10 to 31.62	Strong
$10^{3/2}$ to 10^2	31.62 to 100	Very strong
$> 10^2$	> 100	Decisive



Descriptions 2

Items needed for analysis

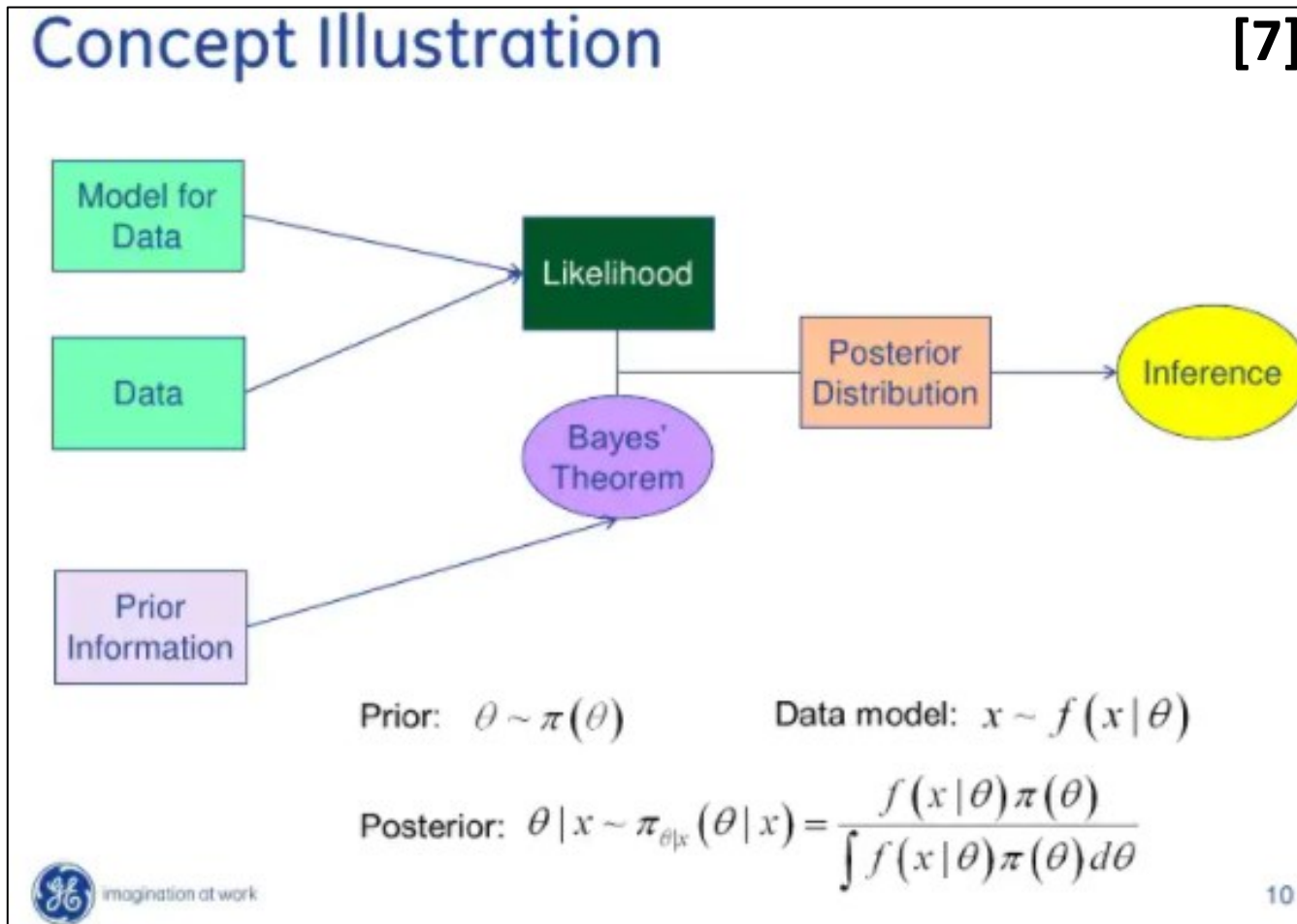
- Priors
- Model
- Data

Steps for analysis

- Covered in Module III

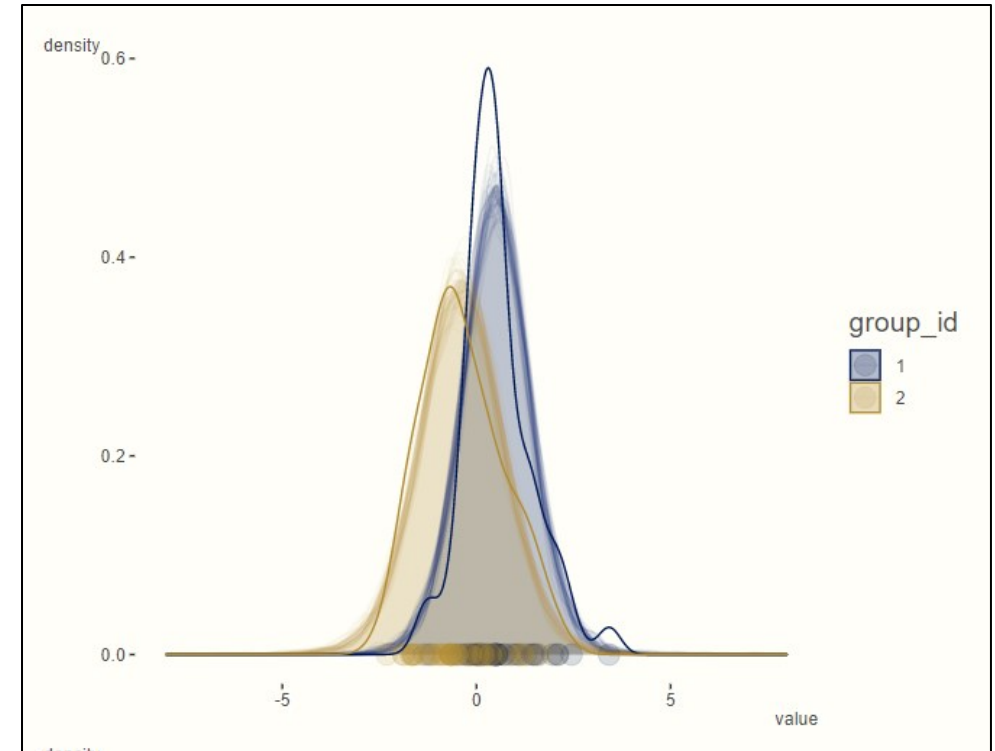
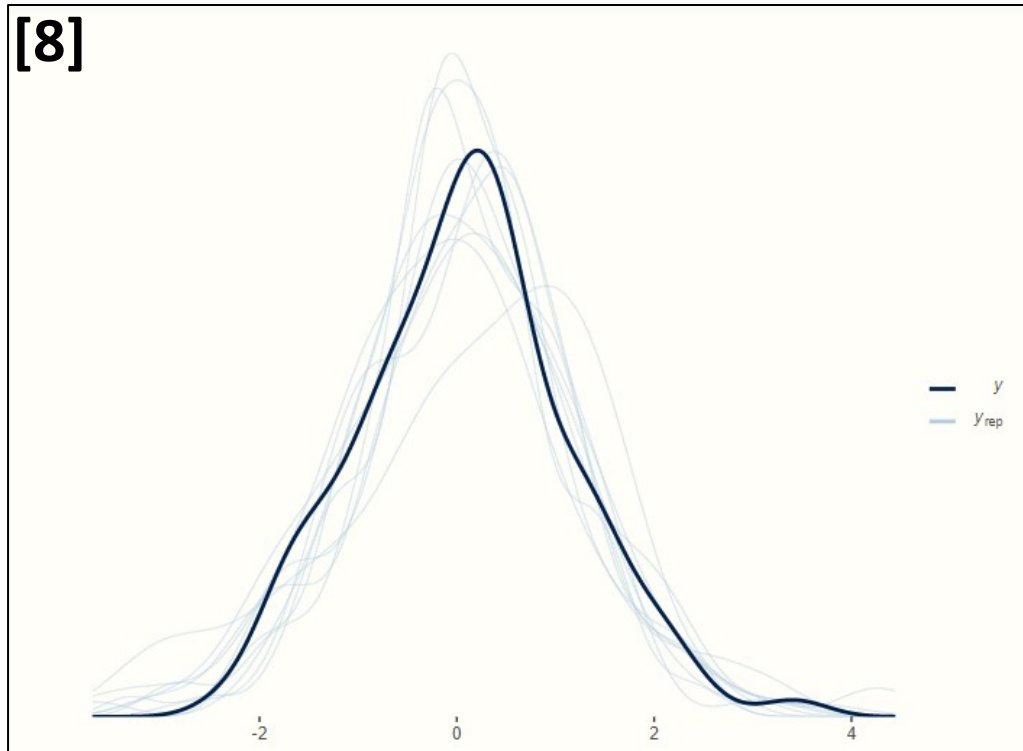
Posterior simulation details

- Covered in Module III



Specific tests

T-test



Specific tests

ANOVA

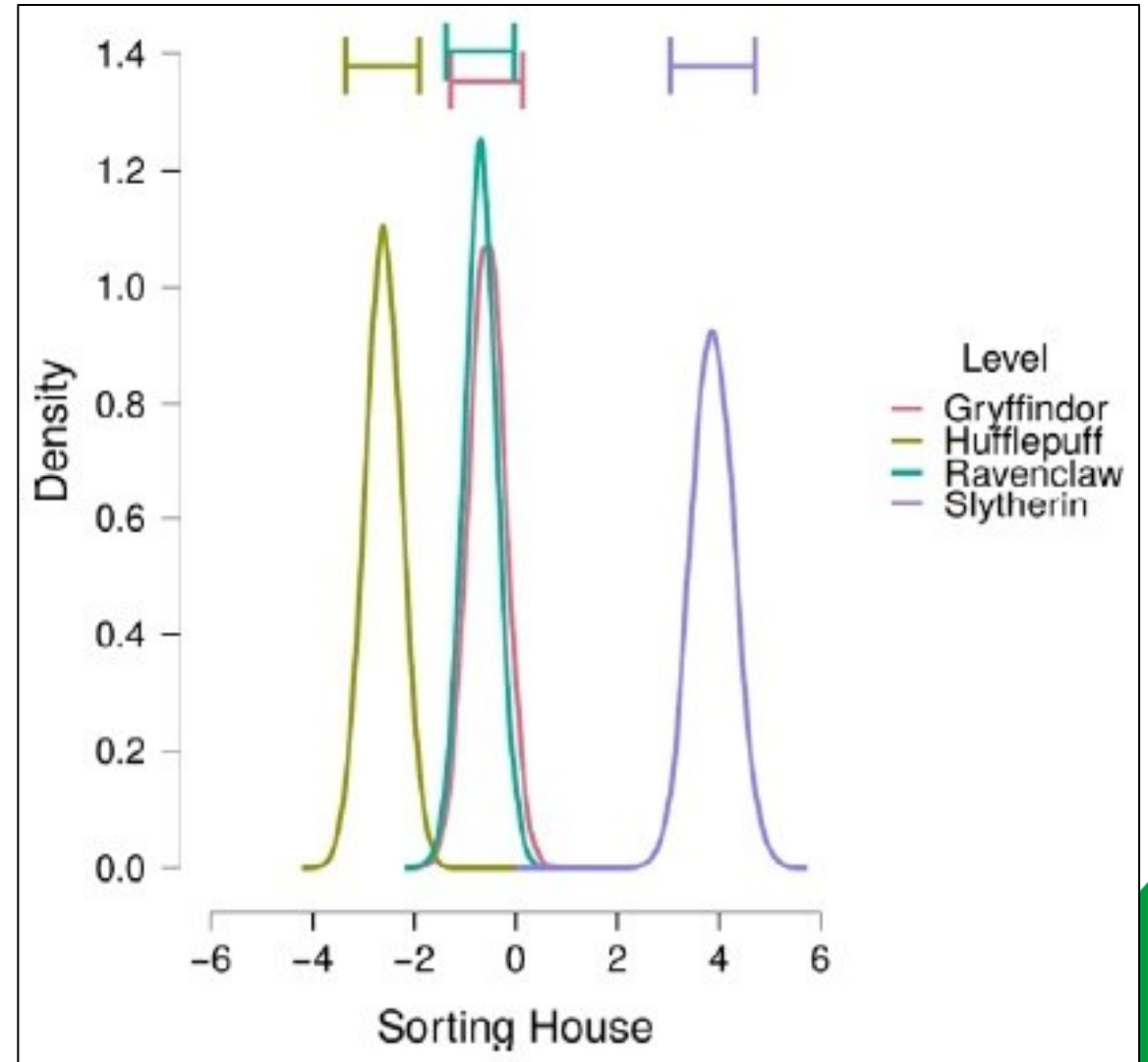
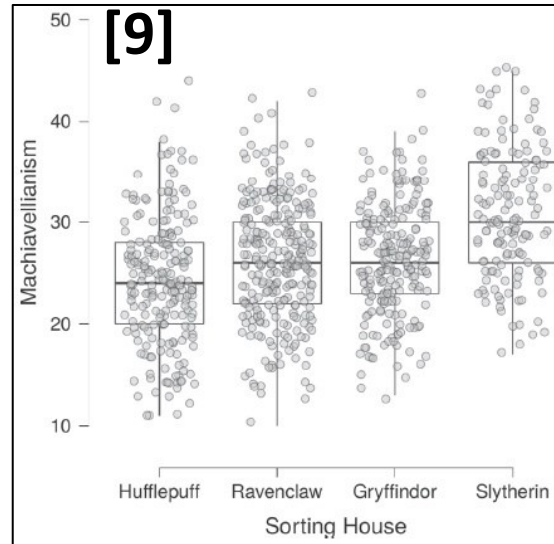


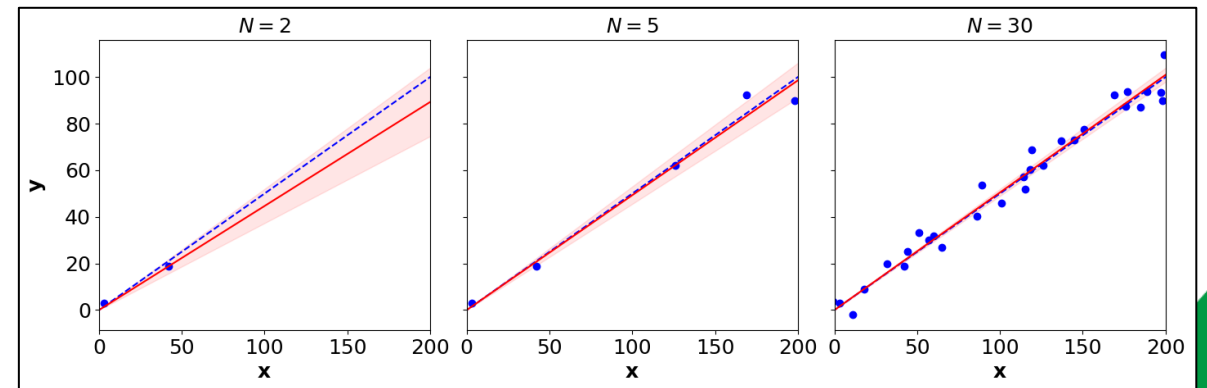
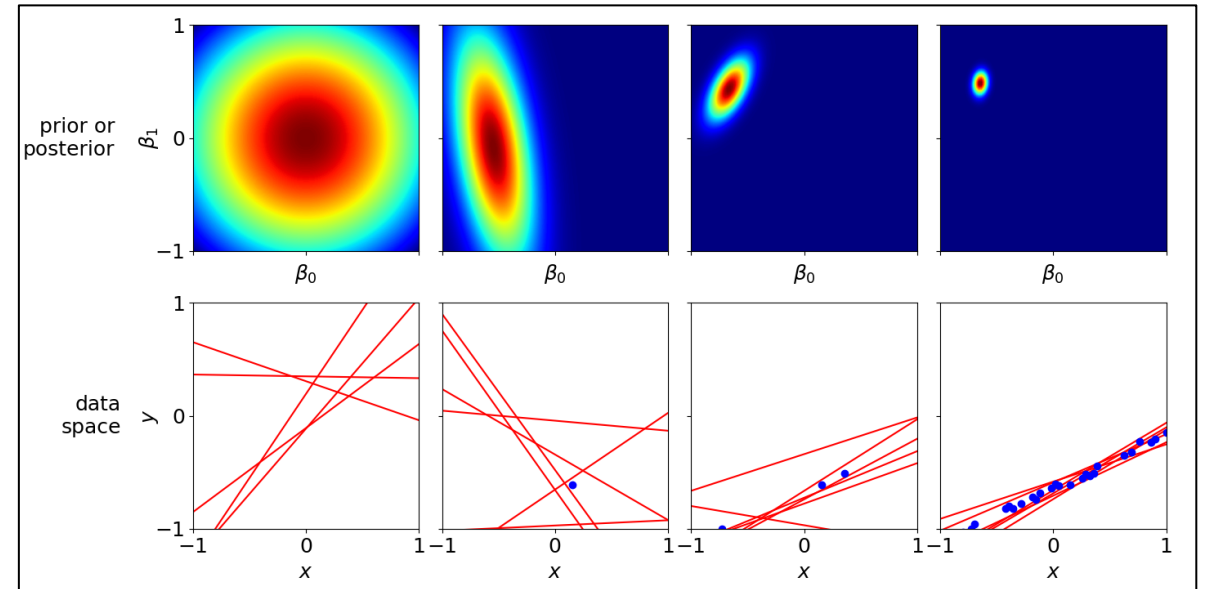
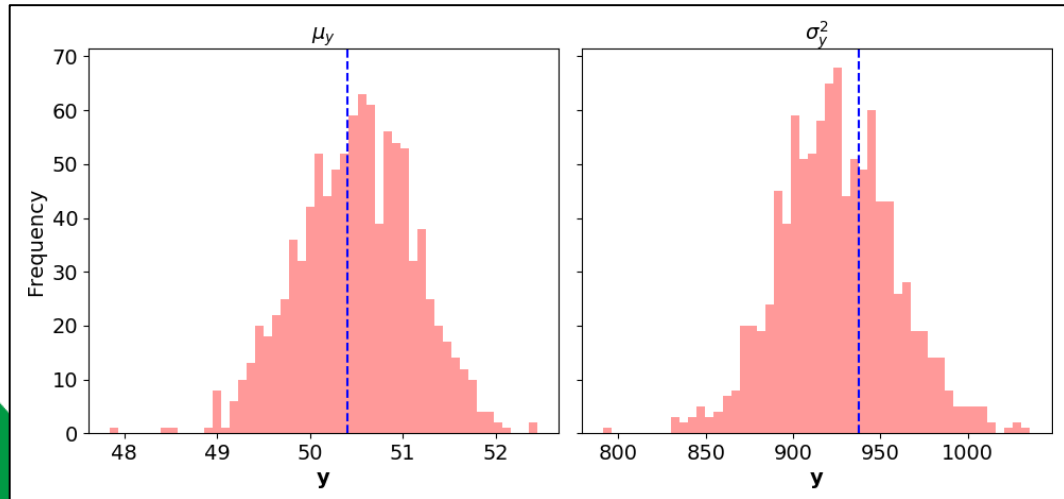
TABLE 5. POST HOC TEST FOR THE SORTING HOUSE DATA
TABLEAU 5. TEST POST HOC SELON L'AFFECTATION AUX MAISONS DE POUDLARD

Level 1	Level 2	Prior Odds	Posterior Odds	$BF_{10,U}$	error %
Gryffindor	Hufflepuff	0.414	27.2	65.6	5.73×10^{-5}
Gryffindor	Ravenclaw	0.414	0.0432	0.104	9.56×10^{-5}
Gryffindor	Slytherin	0.414	1.04×10^9	2.50×10^9	3.94×10^{-16}
Hufflepuff	Ravenclaw	0.414	15.5	37.3	7.57×10^{-8}
Hufflepuff	Slytherin	0.414	5.43×10^{16}	1.31×10^{17}	3.35×10^{-23}
Ravenclaw	Slytherin	0.414	5.30×10^9	1.28×10^{10}	6.36×10^{-16}

Specific tests

Regression

[10]
$$p(\beta | \mathbf{X}, y) \propto \underbrace{p(\mathbf{X}, y | \beta)}_{\text{likelihood}} \underbrace{p(\beta)}_{\text{prior}}$$



Examples setup

R _[11]	SAS _[12]	SPSS _[13]
Package BayesFactor <ul style="list-style-type: none"> • T-test • ANOVA • Simple Linear Regression • Multiple Linear Regression 	PROC GENMOD <ul style="list-style-type: none"> • T-test • ANOVA • Simple Linear Regression • Poisson Regression 	Bayesian Statistics <ul style="list-style-type: none"> • T-test • ANOVA • Simple Linear Regression

R-code available at: https://med.und.edu/daccota/_files/docs/berdc_docs/bayesian_analysis_2_r_code.txt

SAS-code available at: https://med.und.edu/daccota/_files/docs/berdc_docs/bayesian_analysis_2_sas_code.txt

Step-by-step Example 1.1

Standard versus Bayesian Analysis in R

T-test

```
library("BayesFactor")
```

```
#t-test
```

```
#Data formatting
```

```
data(sleep)
```

```
diffScores <- sleep$extra[1:10] - sleep$extra[11:20]
```

```
#Standard
```

```
t.test(diffScores)
```

```
#Bayes
```

```
bf_t <- ttestBF(x = diffScores)
```

```
bf_t #alternative model
```

```
1/bf_t #null model
```

One Sample t-test

data: diffScores

t = -4.0621, df = 9, p-value = 0.002833

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-2.4598858 -0.7001142

sample estimates:

mean of x

-1.58

Bayes factor analysis

[1] Alt., r=0.707 : 17.25888 ±0%

Against denominator:

Null, mu = 0

Bayes factor type: BFoneSample, JZS

Bayes factor analysis

[1] Null, mu=0 : 0.05794119 ±0%

Against denominator:

Alternative, r = 0.707106781186548, mu =/= 0

Bayes factor type: BFoneSample, JZS

Step-by-step Example 1.1

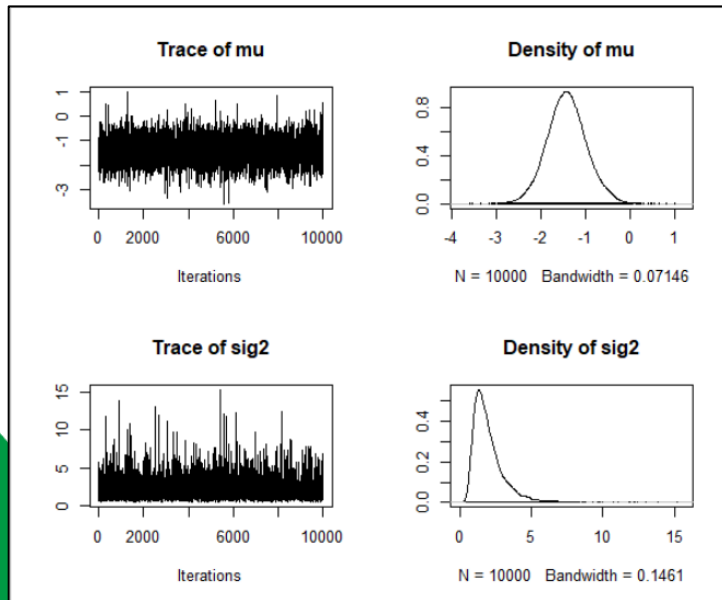
T-test cont.

#Chains

```
chains <- posterior(bf_t, iterations=1000)
summary(chains)
chains2 <- recompute(chains, iterations = 10000)
plot(chains2[,1:2])
```

#Multiple

```
bfInterval <- ttestBF(x = diffScores, nullInterval=c(-Inf,0))
allbf <- c(bf_t, bfInterval)
allbf
plot(allbf)
```

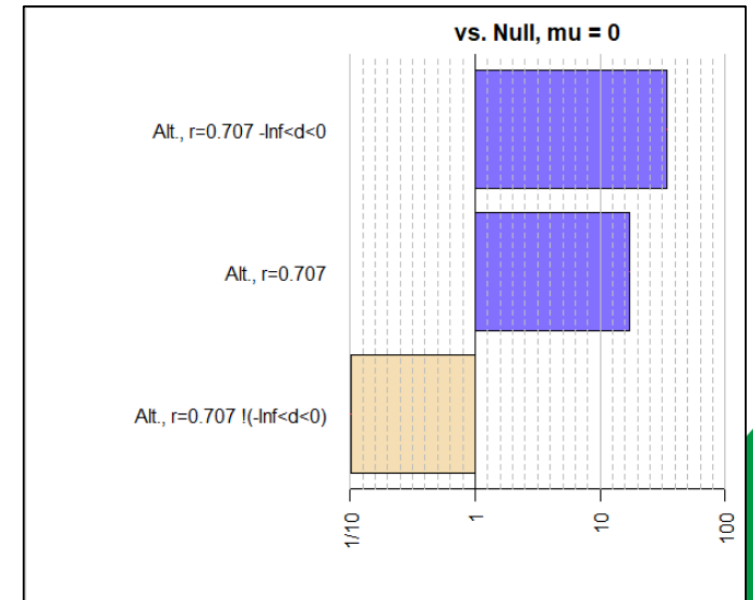


Bayes factor analysis

- [1] Alt., r=0.707 : 17.25888 ±0%
- [2] Alt., r=0.707 -lnf<d<0 : 34.41694 ±0%
- [3] Alt., r=0.707 !(-lnf<d<0) : 0.1008246 ±0.06%

Against denominator:
 Null, mu = 0

Bayes factor type: BFoneSample, JZS



Step-by-step Example 1.2

ANOVA

#ANOVA

#Data formatting

```
plot(chickwts$weight~chickwts$feed)
```

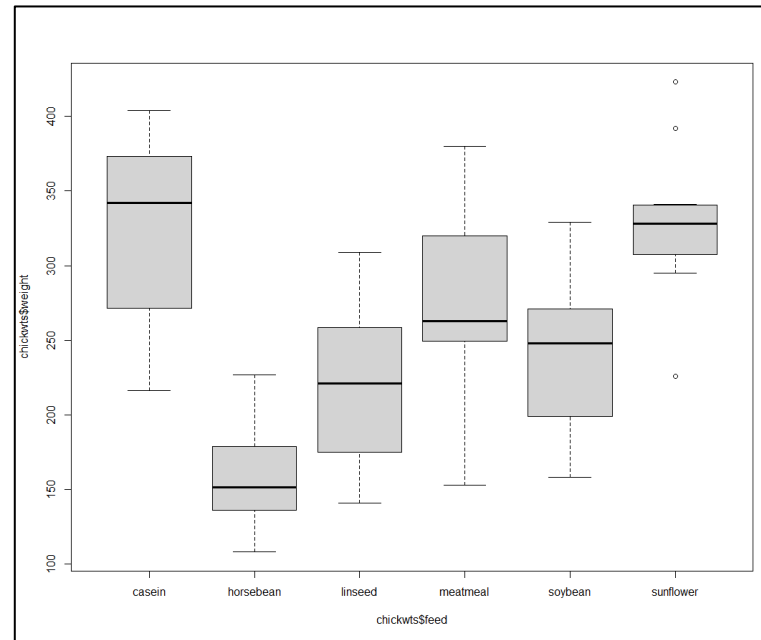
#Standard

```
summary(aov(weight~feed, data=chickwts))
```

#Bayes

```
bf_a <-anovaBF(weight~feed, data=chickwts)
```

```
bf_a
```



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
feed	5	231129	46226	15.37	5.94e-10 ***
Residuals	65	195556	3009		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bayes factor analysis

[1] feed : 14067867 ±0%

Against denominator:
Intercept only

Bayes factor type: BFlinearModel, JZS

Step-by-step Example 1.2

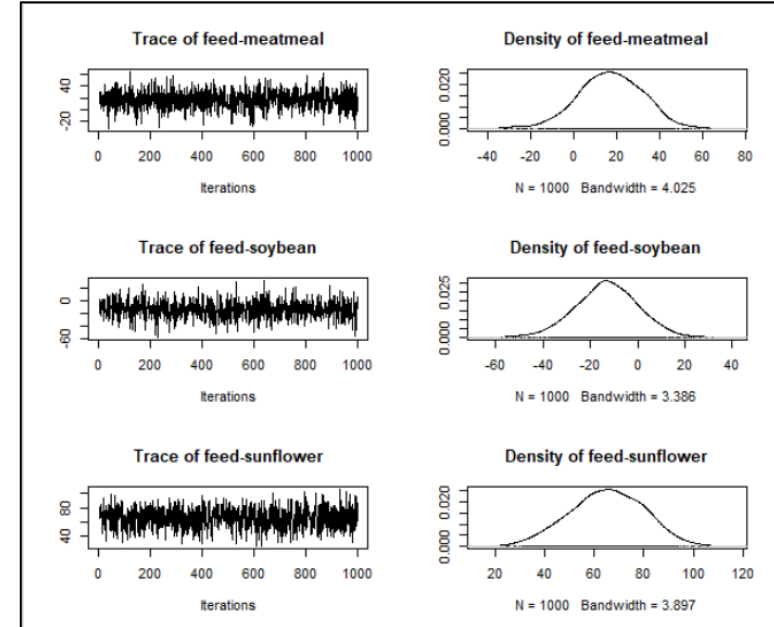
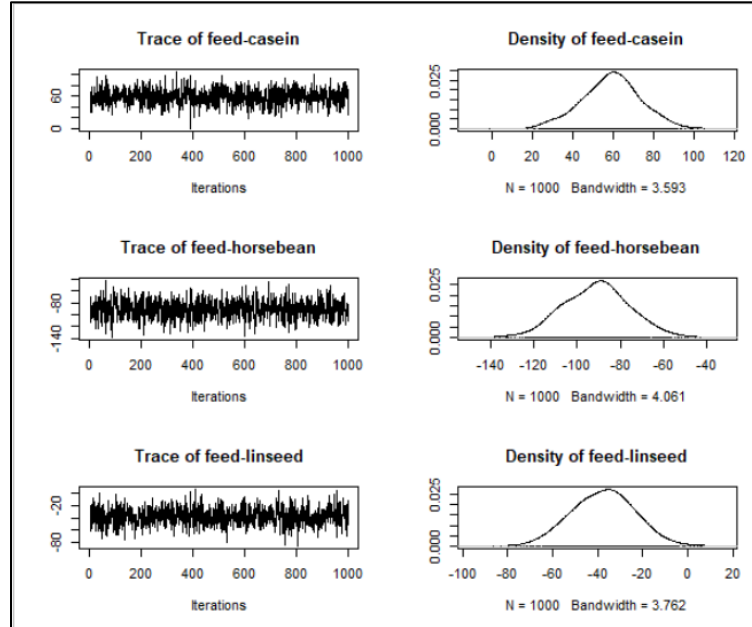
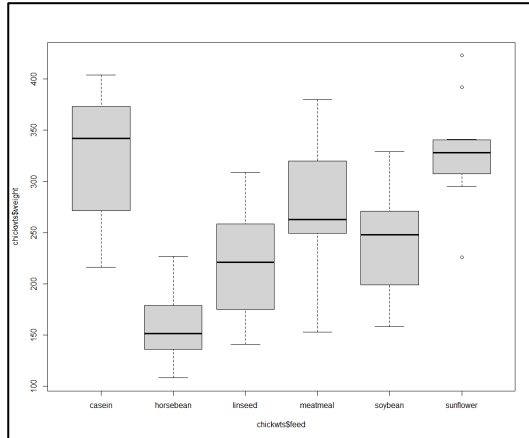
ANOVA cont.

#Chains

`chains <- posterior(bf_a, iterations=1000)`

`summary(chains)`

`plot(chains[,2:7])`



Iterations = 1:1000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

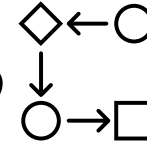
1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
mu	259.35	6.588	0.20832	0.21197
feed-casein	59.04	14.627	0.46255	0.46255
feed-horsebean	-90.99	15.253	0.48233	0.31331
feed-linseed	-37.51	14.129	0.44680	0.47422
feed-meatmeal	17.22	15.118	0.47808	0.47808
feed-soybean	-12.92	13.261	0.41937	0.42755
feed-sunflower	65.15	14.635	0.46281	0.48797
sig2	3138.97	571.281	18.06548	19.90381
g_feed	1.60	1.446	0.04572	0.04572

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu	246.9212	254.834	259.295	263.856	272.199
feed-casein	29.0577	50.076	59.399	68.161	87.889
feed-horsebean	-120.4751	-101.600	-90.733	-81.078	-60.935
feed-linseed	-65.4949	-47.219	-37.210	-27.900	-10.219
feed-meatmeal	-12.8971	7.509	17.343	27.791	45.609
feed-soybean	-39.2842	-21.363	-13.029	-4.321	13.042
feed-sunflower	35.7072	55.543	65.632	75.919	92.326
sig2	2231.9416	2727.824	3084.627	3440.836	4512.308
g_feed	0.3711	0.800	1.219	1.876	5.006

Step-by-step Example 1.3



Simple Linear Regression

#Simple Linear Regression

#Data formatting

data(attitude)

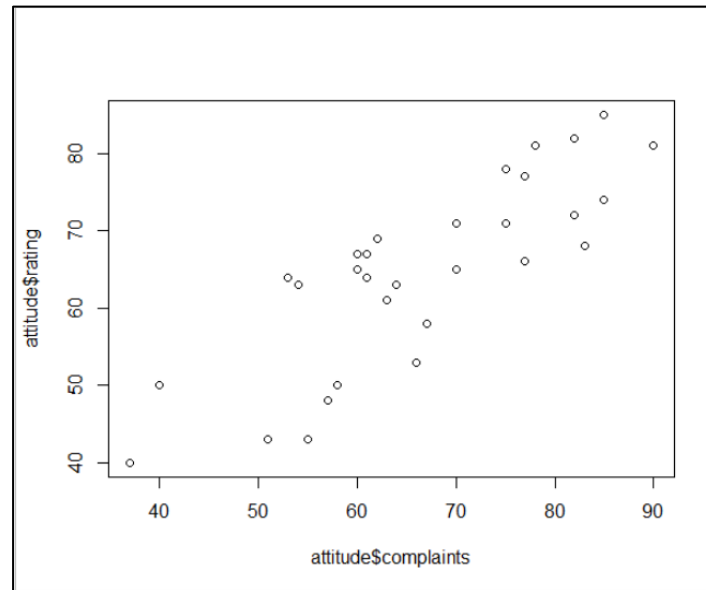
par(mfrow=c(1,1))

plot(x=attitude\$complaints, y=attitude\$rating)

#Standard

lm1 = lm(rating ~ complaints, data = attitude)

summary(lm1)



Call:

lm(formula = rating ~ complaints, data = attitude)

Residuals:

Min	1Q	Median	3Q	Max
-12.8799	-5.9905	0.1783	6.2978	9.6294

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.37632	6.61999	2.172	0.0385 *
complaints	0.75461	0.09753	7.737	1.99e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.993 on 28 degrees of freedom

Multiple R-squared: 0.6813, Adjusted R-squared: 0.6699

F-statistic: 59.86 on 1 and 28 DF, p-value: 1.988e-08

Step-by-step Example 1.3

Simple Linear Regression cont.

#Bayes

```
bf_r <- regressionBF(rating ~ complaints,
data=attitude)
```

bf_r

#Chains

```
chains <- posterior(bf_r, iterations=1000)
summary(chains)
```

Bayes factor analysis

 [1] complaints : **417938.6 ±0.01%**

Against denominator:

Intercept only

Bayes factor type: BFlinearModel, JZS

Iterations = 1:1000
 Thinning interval = 1
 Number of chains = 1
 Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
mu	64.6068	1.3873	0.043871	0.04387
complaints	0.7286	0.1081	0.003417	0.00390
sig2	54.2648	15.4479	0.488505	0.54308
g	6.8366	54.0517	1.709265	1.70926

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu	61.8160	63.6485	64.6576	65.5125	67.154
complaints	0.5051	0.6581	0.7256	0.8017	0.946
sig2	31.6604	43.3656	51.5495	62.0780	91.130
g	0.2130	0.6508	1.3513	3.3668	30.291

Step-by-step Example 1.4

Multiple Linear Regression

`#Multiple Linear Regression`

`#Data formatting`

`data(attitude)`

`#Standard`

`lm2 = lm(rating ~ ., data = attitude)`

`summary(lm2)`

Call:

`lm(formula = rating ~ ., data = attitude)`

Residuals:

Min	1Q	Median	3Q	Max
-10.9418	-4.3555	0.3158	5.5425	11.5990

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.78708	11.58926	0.9310	0.361634
complaints	0.61319	0.16098	3.809	0.000903 ***
privileges	-0.07305	0.13572	-0.538	0.595594
learning	0.32033	0.16852	1.901	0.069925 .
raises	0.08173	0.22148	0.369	0.715480
critical	0.03838	0.14700	0.261	0.796334
advance	-0.21706	0.17821	-1.218	0.235577

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.068 on 23 degrees of freedom

Multiple R-squared: 0.7326, Adjusted R-squared: 0.6628

F-statistic: 10.5 on 6 and 23 DF, p-value: 1.24e-05

Step-by-step Example 1.4

Multiple Linear Regression cont.

#Bayes

```
bf_r2 <- regressionBF(rating ~ ., data=attitude)
length(bf_r2) 63
head(bf_r2, n=6)
```

Bayes factor analysis

```
-----
[1] complaints           : 417938.6 ±0.01%
[2] complaints + learning : 207271.9 ±0%
[3] complaints + learning + advance : 88041.54 ±0%
[4] complaints + raises   : 77498.99 ±0%
[5] complaints + privileges : 75015.23 ±0%
[6] complaints + advance  : 72759.76 ±0%
```

Against denominator:

Intercept only

Bayes factor type: BFlinearModel, JZS

#Compare 5 models to best

```
bf_r2.2 <- head(bf_r2) / max(bf_r2)
bf_r2.2
plot(bf_r2.2)
```

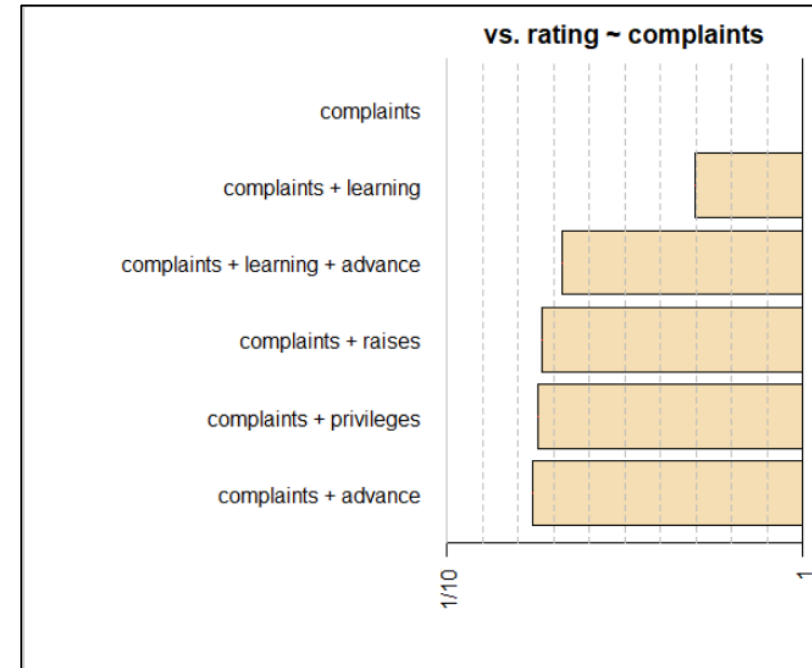
Bayes factor analysis

```
-----
[1] complaints           : 1 ±0%
[2] complaints + learning : 0.4959386 ±0.01%
[3] complaints + learning + advance : 0.2106566 ±0.01%
[4] complaints + raises   : 0.1854315 ±0.01%
[5] complaints + privileges : 0.1794886 ±0.01%
[6] complaints + advance  : 0.174092 ±0.01%
```

Against denominator:

rating ~ complaints

Bayes factor type: BFlinearModel, JZS



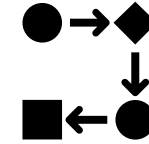
Assessment 1

qualtrics^{XM}



https://und.qualtrics.com/jfe/form/SV_eW1CNrs7cPhjqFE

Step-by-step Example 2.1

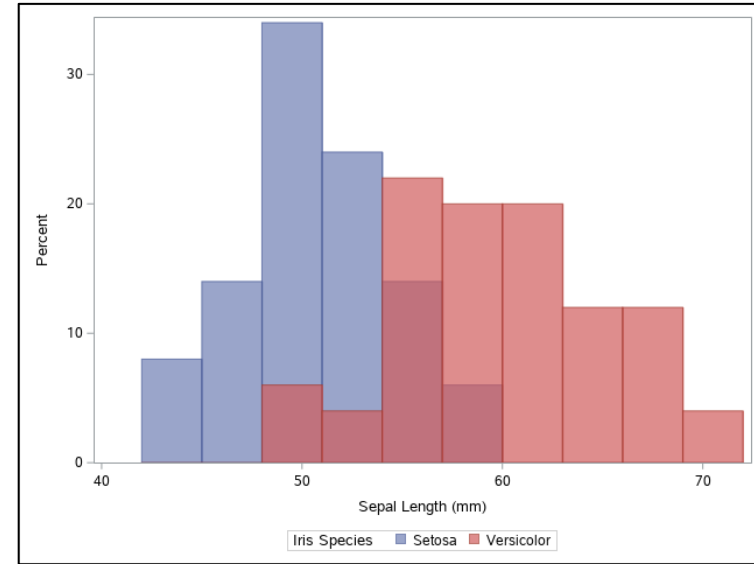


Standard versus Bayesian Analysis in SAS

T-test

```
DATA Iris; set sashelp.Iris;
PROC FREQ data=Iris;
    tables Species;
PROC MEANS data=Iris;
    where Species in ("Setosa" "Versicolor");
    var SepalLength;
    by Species;
```

```
PROC SGPLOT data=Iris;
    where Species in ("Setosa" "Versicolor");
    histogram SepalLength /group=Species transparency=0.30;
```



Iris Species=Setosa

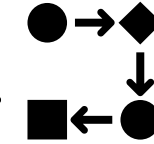
Analysis Variable : SepalLength Sepal Length (mm)				
N	Mean	Std Dev	Minimum	Maximum
50	50.0600000	3.5248969	43.0000000	58.0000000

Iris Species=Versicolor

Analysis Variable : SepalLength Sepal Length (mm)				
N	Mean	Std Dev	Minimum	Maximum
50	59.3600000	5.1617115	49.0000000	70.0000000

Iris Species				
Species	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Setosa	50	33.33	50	33.33
Versicolor	50	33.33	100	66.67
Virginica	50	33.33	150	100.00

Step-by-step Example 2.1



T-test cont.

PROC GENMOD data=Iris;
where Species in ("Setosa" "Versicolor");
class Species;
model SepalLength=Species;

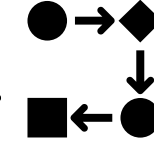
PROC GENMOD data=Iris;
where Species in ("Setosa" "Versicolor");
class Species;
model SepalLength=Species;
bayes seed=1 coeffprior=normal;

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	59.3600	0.6188	58.1472	60.5728	9203.20	<.0001
Species	Setosa	1	-9.3000	0.8751	-11.0151	-7.5849	112.95	<.0001
Species	Versicolor	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	4.3753	0.3094	3.8091	5.0257		

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Intercept	10000	59.3613	0.6312	58.9320	59.3711	59.7808
SpeciesSetosa	10000	-9.3018	0.8966	-9.9028	-9.2927	-8.6954
Dispersion	10000	19.9560	2.9351	17.8867	19.6865	21.7329

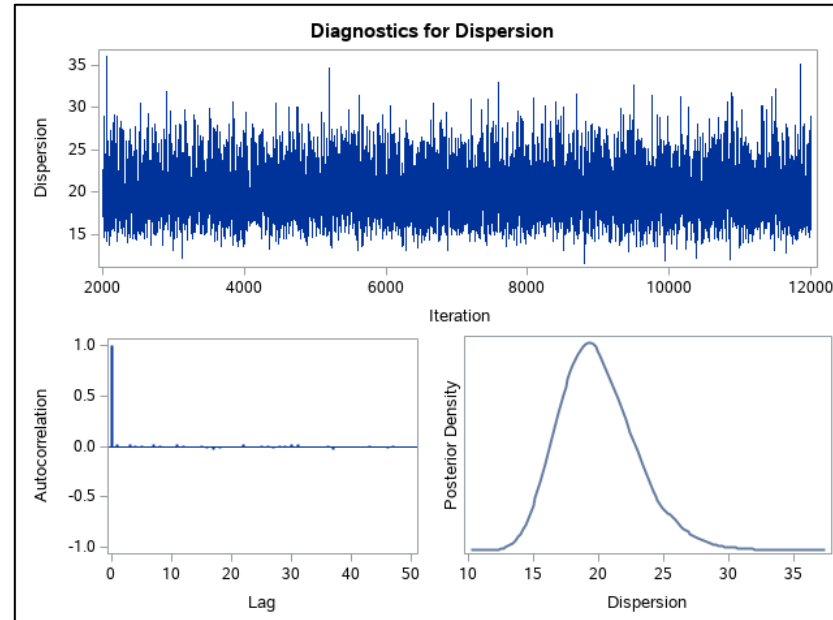
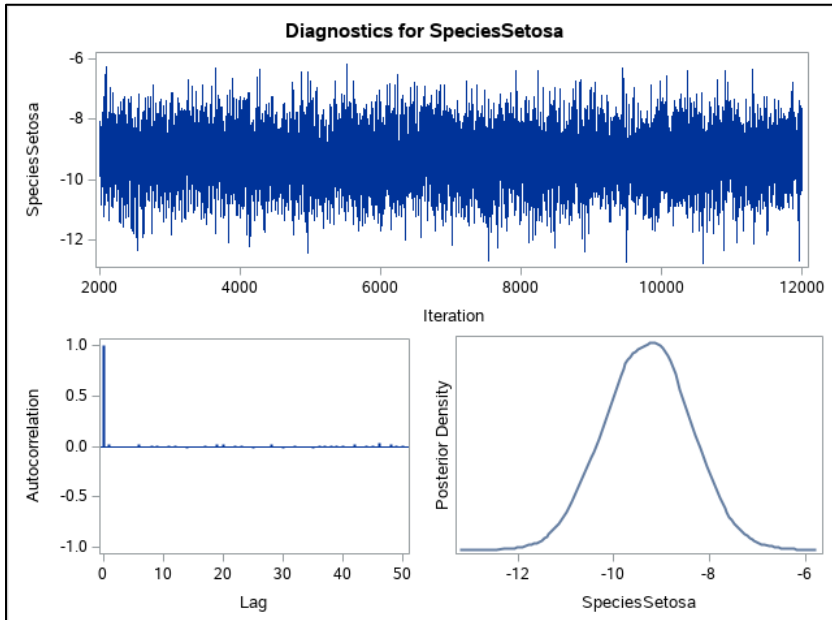
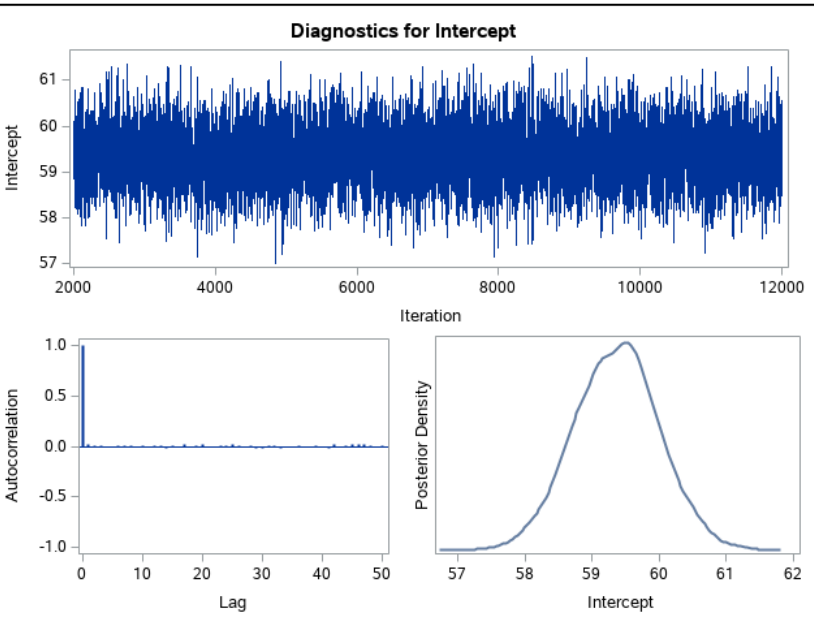
Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	58.1180	60.6000	58.1440	60.6221
SpeciesSetosa	0.050	-11.0525	-7.5646	-11.1072	-7.6300
Dispersion	0.050	15.0091	26.5130	14.8495	26.0907

Step-by-step Example 2.1

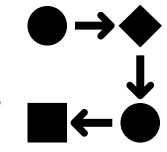


T-test cont. 2

```
PROC GENMOD data=Iris;  
  where Species in ("Setosa" "Versicolor");  
  class Species;  
  model SepalLength=Species;  
  bayes seed=1 coeffprior=normal;
```



Step-by-step Example 2.2

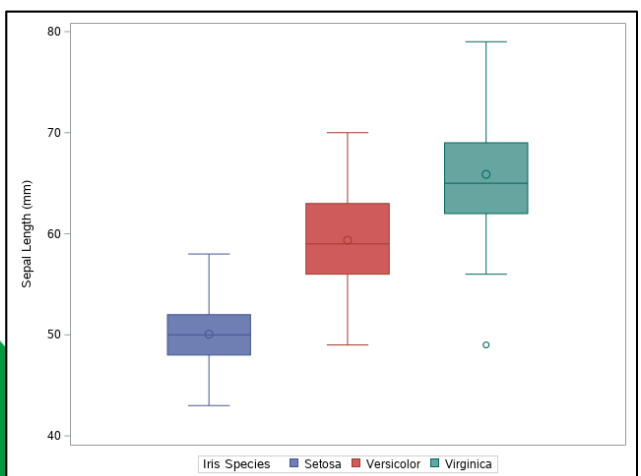


ANOVA

PROC SGPLOT data=Iris;
vbox SepalLength /group=Species;

PROC GENMOD data=Iris;
class Species;
model SepalLength=Species;

PROC GENMOD data=Iris;
class Species;
model SepalLength=Species;
bayes seed=1 coeffprior=uniform

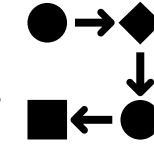


Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	65.8800	0.7207	64.4674	67.2926	8355.87	<.0001
Species	Setosa	1	-15.8200	1.0192	-17.8177	-13.8223	240.92	<.0001
Species	Versicolor	1	-6.5200	1.0192	-8.5177	-4.5223	40.92	<.0001
Species	Virginica	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	5.0962	0.2942	4.5509	5.7067		

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Intercept	10000	65.8905	0.7270	65.4044	65.8926	66.3677
SpeciesSetosa	10000	-15.8376	1.0319	-16.5273	-15.8303	-15.1419
SpeciesVersicolor	10000	-6.5454	1.0245	-7.2170	-6.5473	-5.8728
Dispersion	10000	26.8852	3.1557	24.6744	26.6780	28.8562

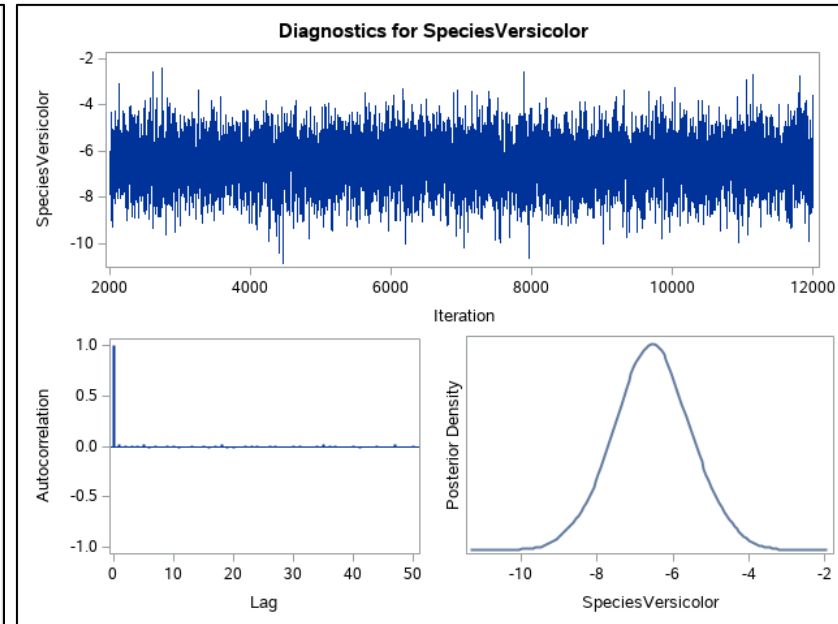
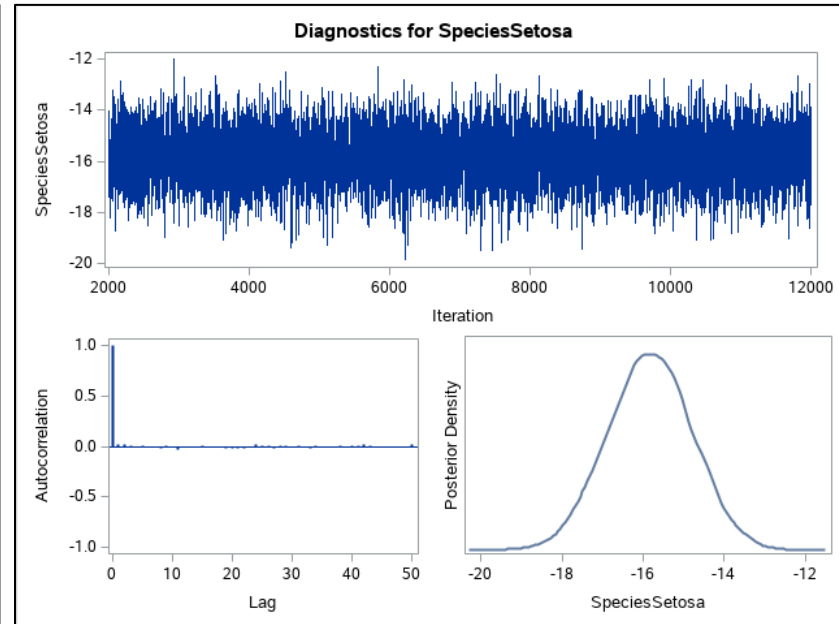
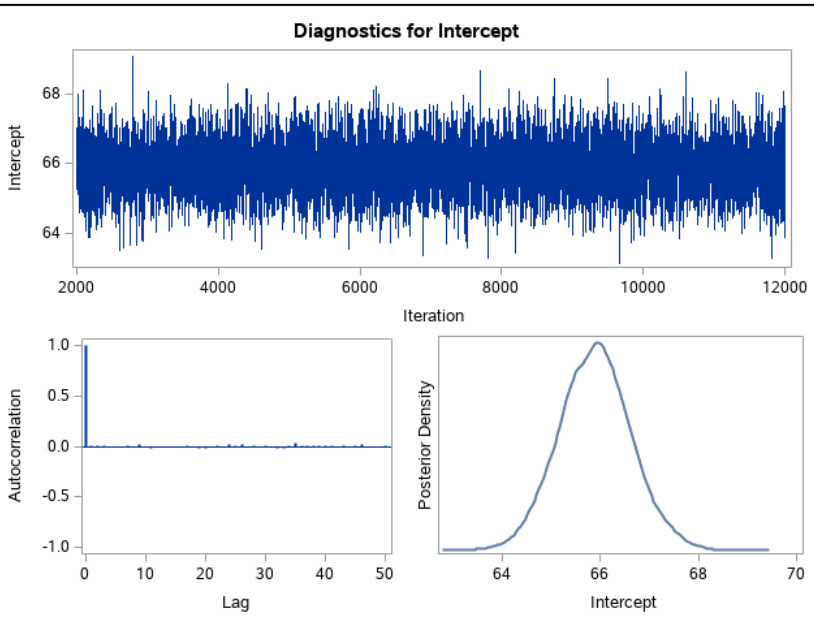
Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	64.4771	67.3561	64.4578	67.3284
SpeciesSetosa	0.050	-17.8639	-13.8305	-17.9082	-13.8751
SpeciesVersicolor	0.050	-8.5717	-4.5105	-8.6151	-4.5645
Dispersion	0.050	21.3747	33.7054	20.9688	33.1140

Step-by-step Example 2.2

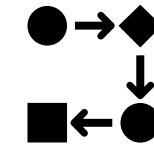


ANOVA cont.

```
PROC GENMOD data=Iris;  
class Species;  
model SepalLength=Species;  
bayes seed=1 coeffprior=uniform;
```



Step-by-step Example 2.3



Linear Regression

PROC GLIMMIX data=Liver;
model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Normal;

PROC GENMOD data=Liver;
model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Normal;

Pearson Chi-Square	4006.57
Pearson Chi-Square / DF	31.06

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BMI	1	129	0.04	0.8478
Age	1	129	5.81	0.0173
Time	1	129	0.37	0.5423
BCM	1	129	2.01	0.1585
AHBA	1	129	2.59	0.1103
Jaund	1	129	0.97	0.3268

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	129	4006.5654	31.0586
Scaled Deviance	129	136.0000	1.0543
Pearson Chi-Square	129	4006.5654	31.0586

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	10.0684	2.8684	4.4464	15.6904	12.32	0.0004
BMI	1	-0.0180	0.0912	-0.1968	0.1608	0.04	0.8435
Age	1	-0.0779	0.0315	-0.1396	-0.0162	6.13	0.0133
Time	1	-0.0167	0.0267	-0.0690	0.0355	0.39	0.5304
BCM	1	-1.4874	1.0213	-3.4892	0.5144	2.12	0.1453
AHBA	1	1.8310	1.1089	-0.3424	4.0043	2.73	0.0987
Jaund	1	0.9864	0.9761	-0.9266	2.8995	1.02	0.3122
Scale	1	5.4277	0.3291	4.8195	6.1126		

Step-by-step Example 2.3

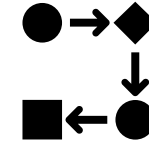
Linear Regression cont.

```
PROC GENMOD data=Liver;
  model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Normal;
  bayes seed=1 coeffprior=normal;
```

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Intercept	10000	10.0877	3.0033	8.0747	10.1044	12.1463
BMI	10000	-0.0188	0.0957	-0.0839	-0.0183	0.0454
Age	10000	-0.0776	0.0323	-0.0994	-0.0776	-0.0561
Time	10000	-0.0167	0.0277	-0.0352	-0.0168	0.00192
BCM	10000	-1.5052	1.0603	-2.2129	-1.5099	-0.7910
AHBA	10000	1.8320	1.1369	1.0794	1.8338	2.5821
Jaund	10000	0.9794	1.0128	0.2964	0.9762	1.6622
Dispersion	10000	31.5808	4.0150	28.7134	31.2603	34.0755

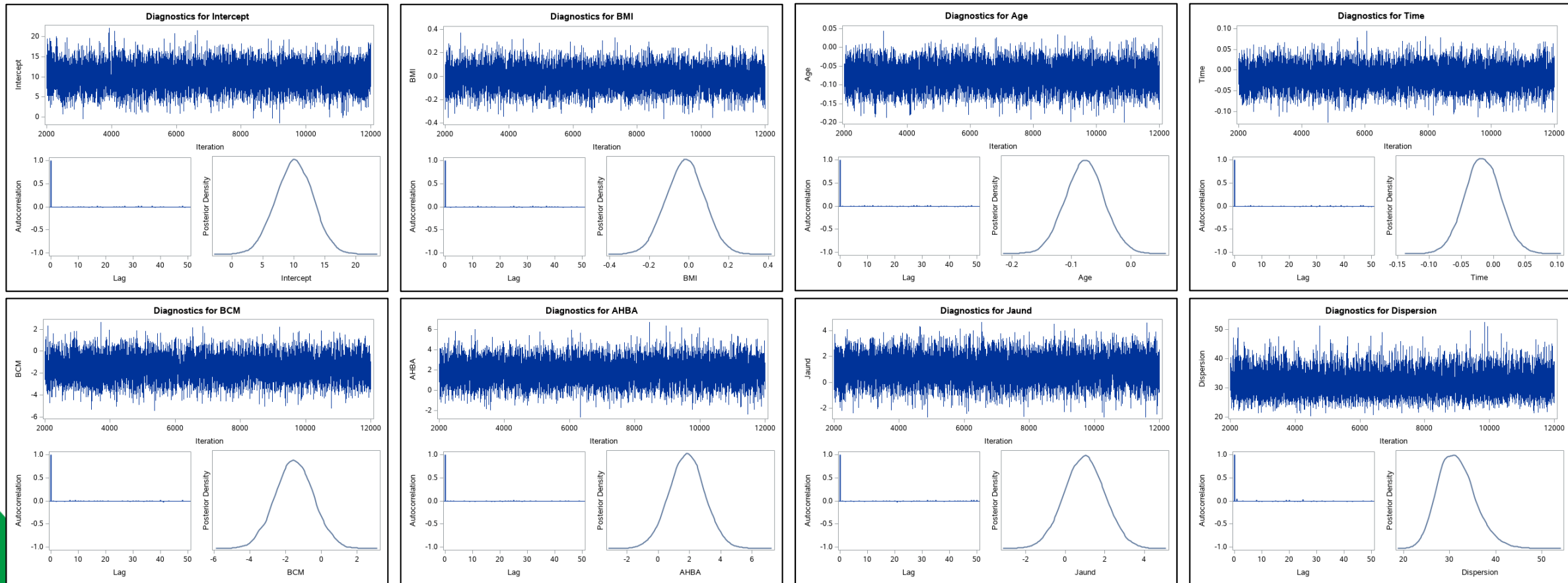
Parameter	Alpha	Posterior Intervals			
		Equal-Tail Interval		HPD Interval	
Intercept	0.050	4.1960	15.9416	4.4029	16.1206
BMI	0.050	-0.2038	0.1698	-0.2066	0.1663
Age	0.050	-0.1414	-0.0134	-0.1384	-0.0110
Time	0.050	-0.0707	0.0381	-0.0713	0.0371
BCM	0.050	-3.5754	0.5938	-3.5908	0.5647
AHBA	0.050	-0.4079	4.0623	-0.3635	4.0957
Jaund	0.050	-0.9916	2.9613	-1.0089	2.9212
Dispersion	0.050	24.6876	40.2912	24.1990	39.6035

Step-by-step Example 2.3



Linear Regression cont. 2

PROC GENMOD data=Liver;
model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Normal;
bayes seed=1 coeffprior=normal;



Step-by-step Example 2.4

Poisson Regression

```
PROC GLIMMIX data=Liver;
  model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Poisson;
```

```
PROC GENMOD data=Liver;
  model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Poisson link=log;
```

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	129	382.1373	2.9623
Scaled Deviance	129	382.1373	2.9623
Pearson Chi-Square	129	501.1176	3.8846

Pearson Chi-Square	501.12
Pearson Chi-Square / DF	3.88

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BMI	1	129	0.29	0.5889
Age	1	129	32.34	<.0001
Time	1	129	1.79	0.1833
BCM	1	129	11.67	0.0008
AHBA	1	129	14.94	0.0002
Jaund	1	129	6.31	0.0132

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	2.4508	0.2284	2.0032	2.8984	115.16	<.0001
BMI	1	-0.0044	0.0080	-0.0201	0.0114	0.29	0.5880
Age	1	-0.0135	0.0024	-0.0181	-0.0088	32.34	<.0001
Time	1	-0.0029	0.0022	-0.0072	0.0014	1.79	0.1810
BCM	1	-0.2715	0.0795	-0.4272	-0.1157	11.67	0.0006
AHBA	1	0.3215	0.0832	0.1585	0.4845	14.94	0.0001
Jaund	1	0.2077	0.0827	0.0456	0.3698	6.31	0.0120
Scale	0	1.0000	0.0000	1.0000	1.0000		

Step-by-step Example 2.4

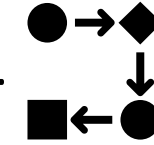
Poisson Regression cont.

```
PROC GENMOD data=Liver;
  model LC_nodes = BMI Age Time BCM AHBA Jaund / dist=Poisson link=log;
  bayes seed=1 plots=none coeffprior=normal;
```

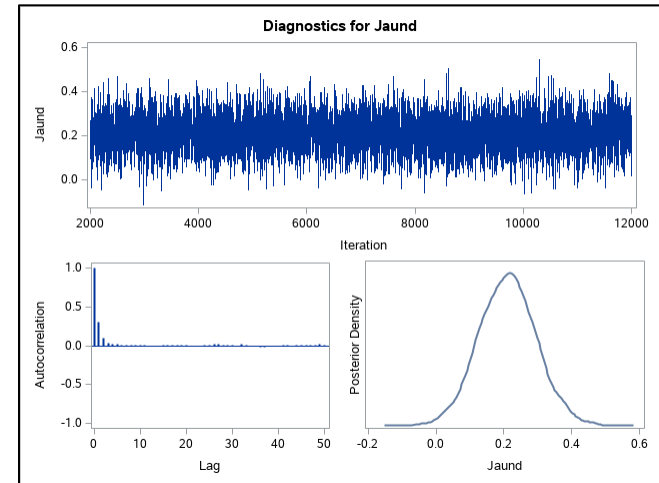
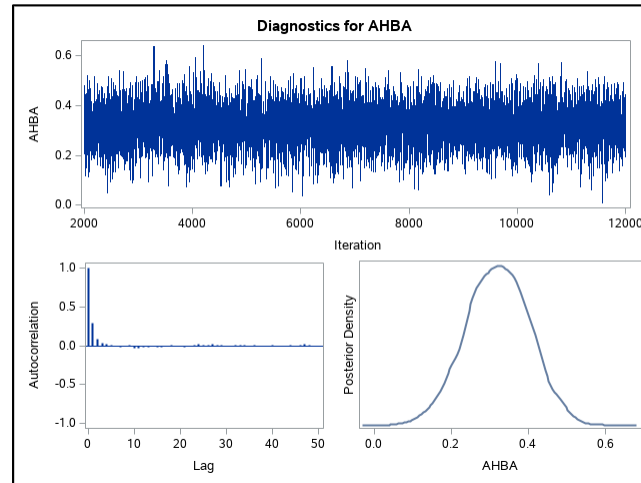
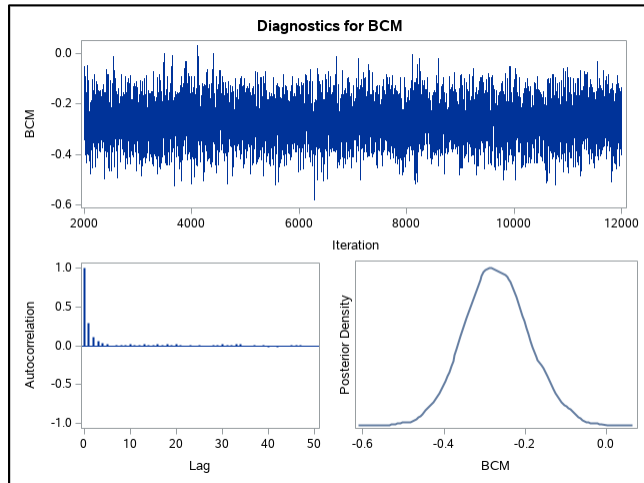
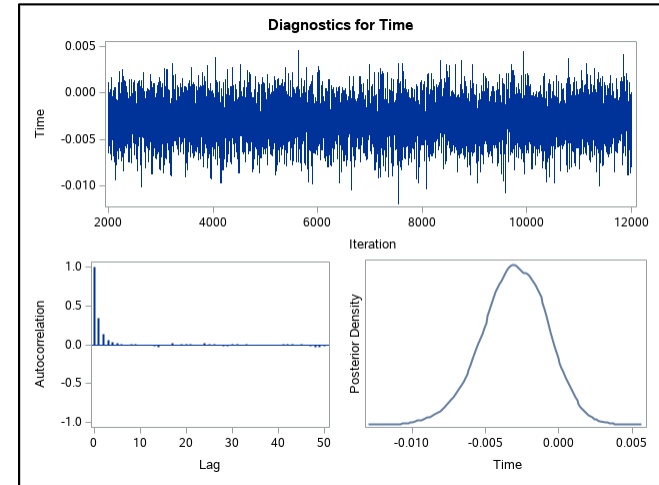
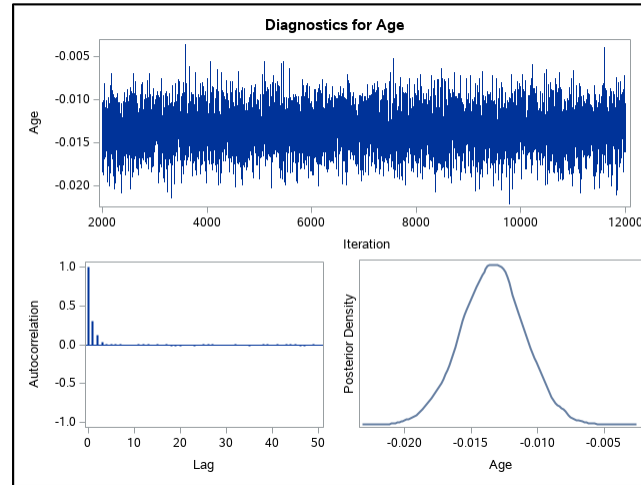
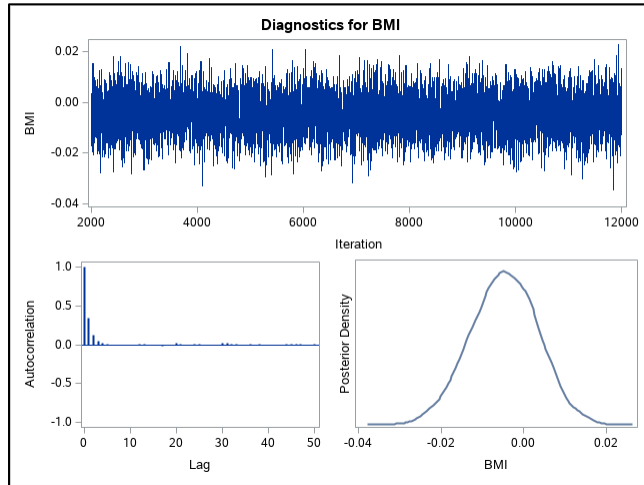
Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Intercept	10000	2.4483	0.2320	2.2903	2.4493	2.6093
BMI	10000	-0.00475	0.00809	-0.0101	-0.00466	0.000851
Age	10000	-0.0134	0.00237	-0.0150	-0.0134	-0.0118
Time	10000	-0.00303	0.00220	-0.00445	-0.00298	-0.00150
BCM	10000	-0.2703	0.0799	-0.3241	-0.2725	-0.2190
AHBA	10000	0.3202	0.0828	0.2642	0.3209	0.3775
Jaund	10000	0.2106	0.0838	0.1533	0.2111	0.2663

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	1.9903	2.9059	2.0289	2.9321
BMI	0.050	-0.0209	0.0108	-0.0211	0.0106
Age	0.050	-0.0181	-0.00870	-0.0184	-0.00908
Time	0.050	-0.00761	0.00105	-0.00745	0.00113
BCM	0.050	-0.4257	-0.1063	-0.4314	-0.1152
AHBA	0.050	0.1563	0.4804	0.1574	0.4811
Jaund	0.050	0.0450	0.3777	0.0468	0.3788

Step-by-step Example 2.4



Poisson Regression cont. 2



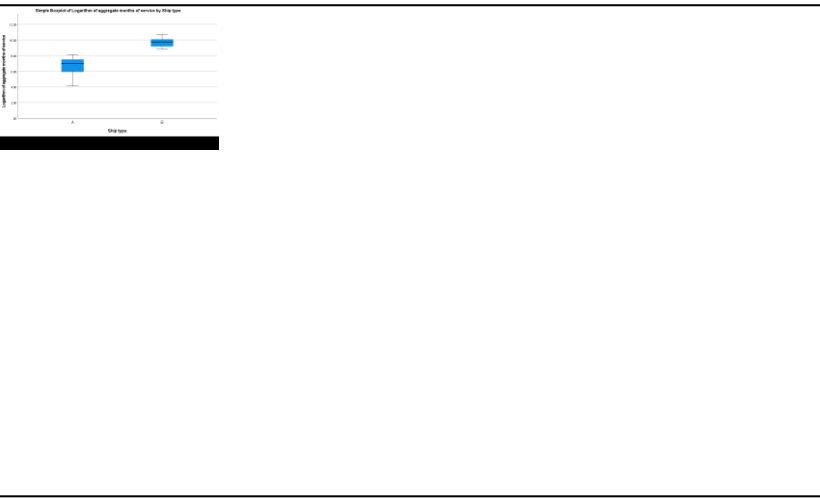
Step-by-step Example 3.1

Standard versus Bayesian Analysis in SPSS

T-test

The screenshot shows the SPSS interface with the 'ships.sav' dataset open. The 'Chart Builder' dialog is active, showing a 'Simple Boxplot of Logarithm of aggregate months of service' by 'Ship type'. The 'Variables' list includes 'Ship type [type]', 'Year of construction...', 'Period of operation...', 'Aggregate months...', 'Logarithm of aggregate months of service', and 'Number of damage incidents'. The 'Logarithm of aggregate months of service' variable is selected for the Y-axis, and 'Ship type' is selected for the X-axis. The 'Element Properties' tab is open, showing the 'Edit Properties of:' section with 'Box1' selected. The 'Axis Label' is 'Ship type'. The 'Categories' section shows 'Variable: Ship type', 'Sort by: Custom', and 'Direction: Ascending'. The 'Order' section shows categories 'A' and 'B'. The 'Small/Empty Categories' section has 'Show empty labeled categories' selected. The 'Collapse (sum) small categories' checkbox is unchecked, and the 'Collapse if % less than:' is set to 5. The 'Gallery' shows 'Simple Boxplot' selected.

type	construction	log_months_service	damage_incidents
1	1	60	0
2	1	60	0
3	1	65	3
4	1	65	4
5	1	70	6
6	1	70	18
7	1	75	0
8	1	75	11
9	2	60	39
10	2	60	29
11	2	65	58
12	2	65	53
13	2	70	12
14	2	70	44
15	2	75	0
16	2	75	18
17	3	60	1
18	3	60	1
19	3	65	0
20	3	65	1
21	3	70	6
22	3	70	2



Step-by-step Example 3.1

T-test cont.

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Bayesian Statistics' > 'Independent Samples Normal' is highlighted. The data editor shows a table with columns 'type' and 'const'.

Row	type	const
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	2	
10	2	
11	2	
12	2	
13	2	
14	2	
15	2	
16	2	
17	3	
18	3	
19	3	
20	3	
21	3	
22	3	

The screenshot shows the 'Bayesian Independent-Sample Inference' dialog box. The 'Test Variable(s)' field contains 'Logarithm of aggregate months o...'. The 'Grouping Variable' is set to 'type(1 2)'. The 'Bayesian Analysis' section has 'Use Both Methods' selected. The 'Define Groups' sub-dialog box is also visible, showing 'Use specified values' selected with Group 1 set to 1 and Group 2 set to 2.

Step-by-step Example 3.1

T-test cont. 2

Bayesian Independent

Group Statistics

	Ship type	N	Mean	Std. Deviation	Std. Error Mean
Logarithm of aggregate months of service	= A	7	6.5913	1.50027	.56705
	= B	7	9.6942	.68657	.25950

Bayes Factor Independent Sample Test (Method = Roudier)^a

	Mean Difference	Pooled Std. Error Difference	Bayes Factor ^b	t	df	Sig.(2-tailed)
Logarithm of aggregate months of service	3.1029	.62361	.011	4.976	12	<.001

a. Assumes unequal variance between groups.

b. Bayes factor: Null versus alternative hypothesis.

Posterior Distribution Characterization for Independent Sample Mean^a

	Mode	Posterior		95% Credible Interval	
		Mean	Variance	Lower Bound	Upper Bound
Logarithm of aggregate months of service	3.1029	3.1029	.583	1.5765	4.6294

a. Prior for Variance: Diffuse. Prior for Mean: Diffuse.

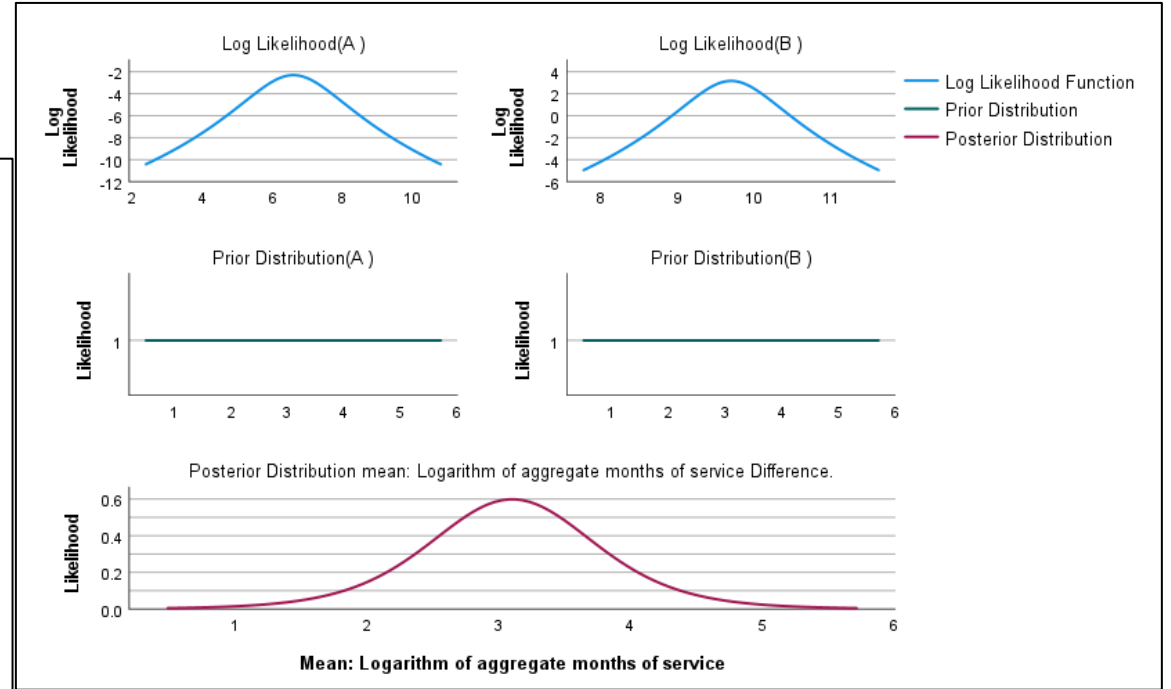


Table 1. Commonly used thresholds to define significance of evidence

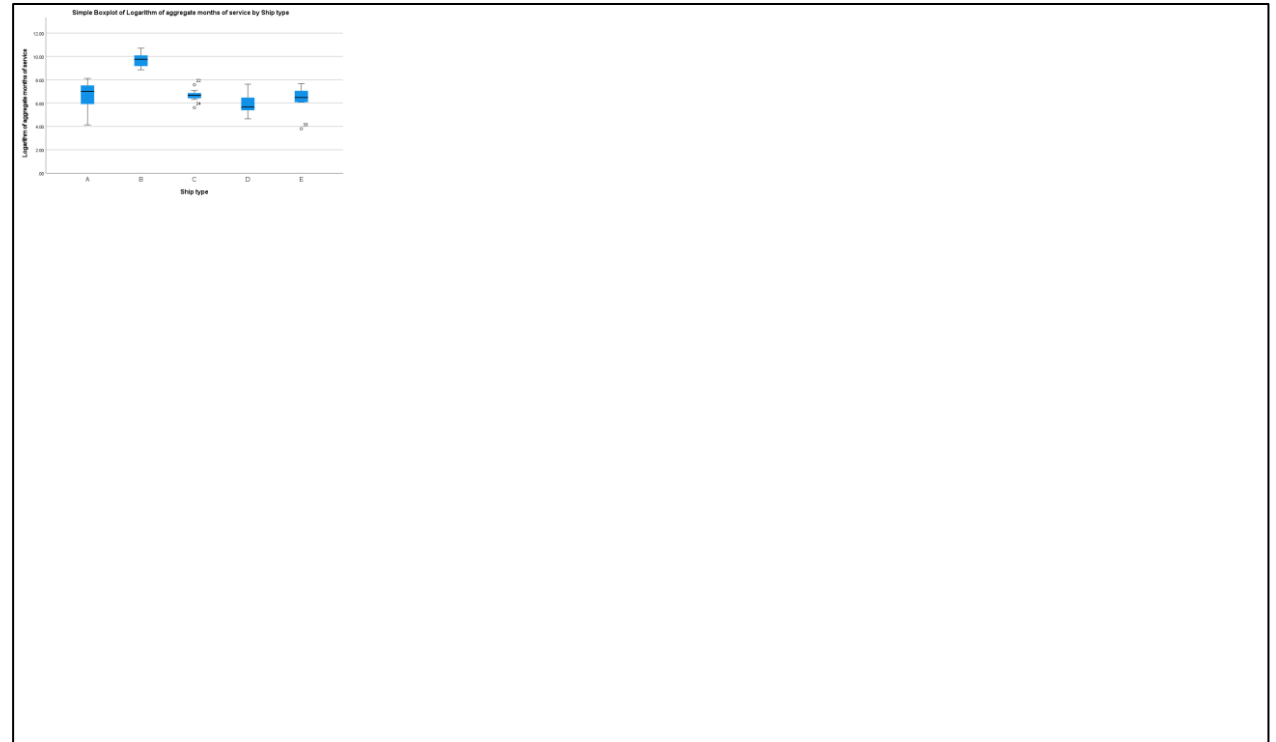
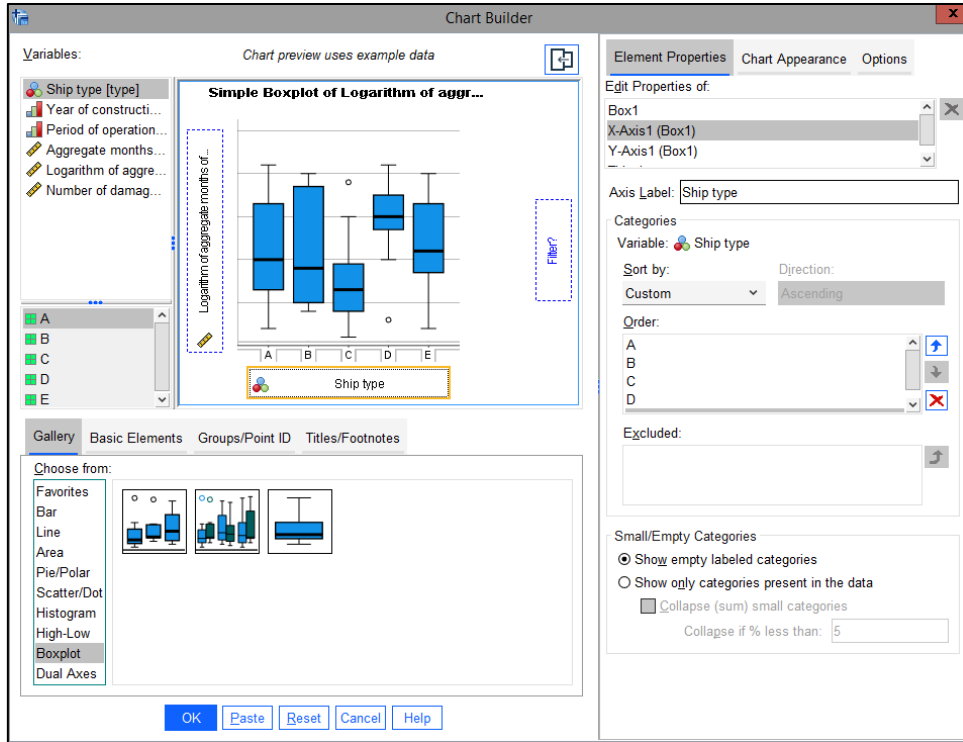
Bayes Factor	Evidence Category	Bayes Factor	Evidence Category	Bayes Factor	Evidence Category
>100	Extreme Evidence for H0	1-3	Anecdotal Evidence for H0	1/30-1/10	Strong Evidence for H1
30-100	Very Strong Evidence for H0	1	No Evidence	1/100-1/30	Very Strong Evidence for H1
10-30	Strong Evidence for H0	1/3-1	Anecdotal Evidence for H1	1/100	Extreme Evidence for H1
3-10	Moderate Evidence for H0	1/10-1/3	Moderate Evidence for H1		

H0: Null Hypothesis

H1: Alternative Hypothesis

Step-by-step Example 3.2

ANOVA



Step-by-step Example 3.2

ANOVA cont.

Bayes Factor	Evidence Category	Bayes Factor	Evidence Category	Bayes Factor	Evidence Category
>100	Extreme Evidence for H0	1-3	Anecdotal Evidence for H0	1/30-1/10	Strong Evidence for H1
30-100	Very Strong Evidence for H0	1	No Evidence	1/100-1/30	Very Strong Evidence for H1
10-30	Strong Evidence for H0	1/3-1	Anecdotal Evidence for H1	1/100	Extreme Evidence for H1
3-10	Moderate Evidence for H0	1/10-1/3	Moderate Evidence for H1		

Bayesian ANOVA

Logarithm of aggregate months of service	Sum of Squares	df	Mean Square	F	Sig.	Bayes Factor ^a
Between Groups	63.763	4	15.941	13.601	<.001	8073.163
Within Groups	33.988	29	1.172			
Total	97.751	33				

a. Bayes factor: JZS

Bayesian Estimates of Coefficients^{a,b,c}

Parameter	Mode	Posterior		95% Credible Interval	
		Mean	Variance	Lower Bound	Upper Bound
Ship type = A	6.591	6.591	.180	5.754	7.428
Ship type = B	9.694	9.694	.180	8.857	10.531
Ship type = C	6.631	6.631	.180	5.794	7.467
Ship type = D	5.954	5.954	.180	5.117	6.791
Ship type = E	6.264	6.264	.210	5.360	7.168

a. Dependent Variable: Logarithm of aggregate months of service

b. Model: Ship type

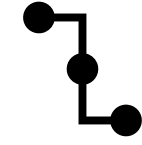
c. Assume standard reference priors.

Bayesian Estimates of Error Variance^a

Parameter	Mode	Posterior		95% Credible Interval	
		Mean	Variance	Lower Bound	Upper Bound
Error variance	1.096	1.259	.127	.743	2.118

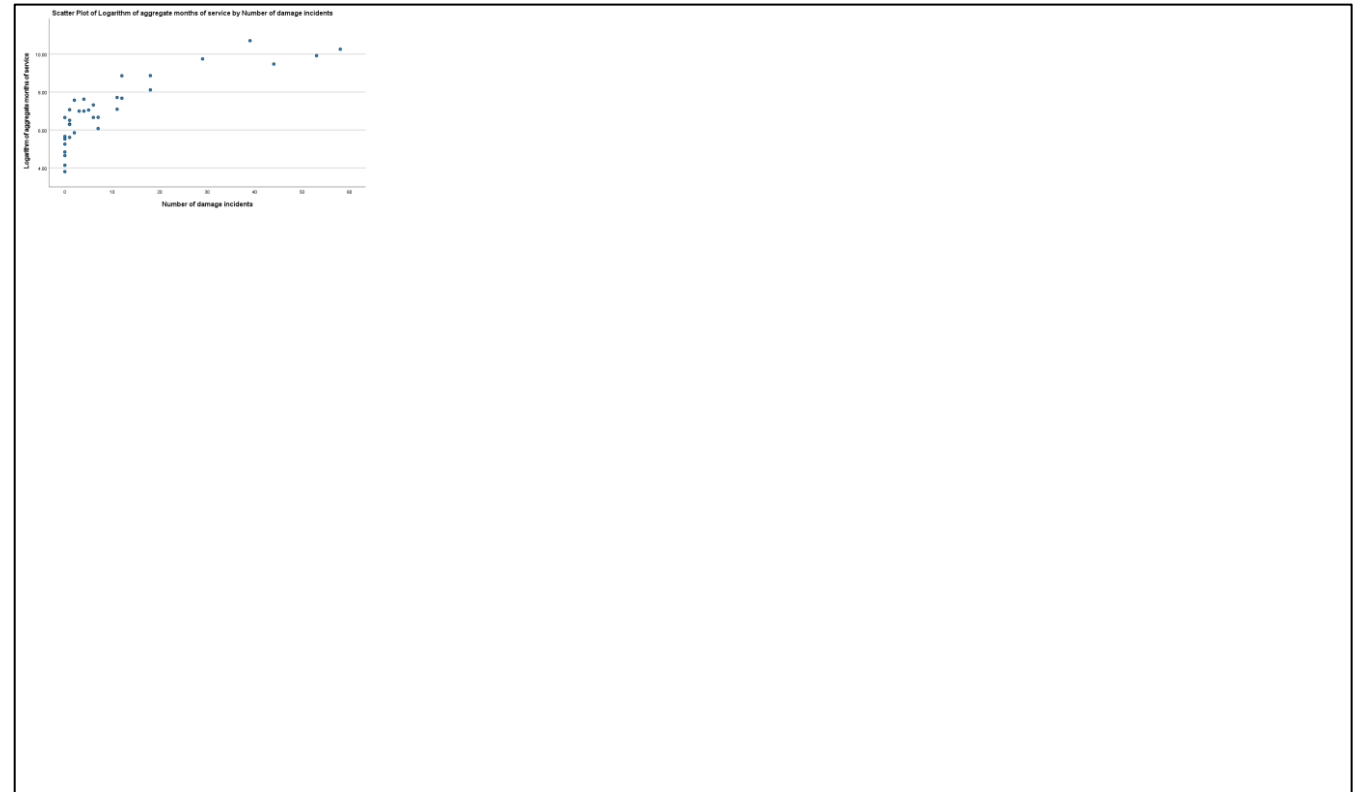
a. Assume standard reference priors.

Step-by-step Example 3.3



Simple Linear Regression

The screenshot shows the 'Chart Builder' window. On the left, the 'Variables' list includes 'Ship type [type]', 'Year of constructi...', 'Period of operation...', 'Aggregate months...', 'Logarithm of aggreg...', and 'Number of damag...'. The 'Logarithm of aggregate months of service' variable is assigned to the Y-axis, and 'Number of damage incidents' is assigned to the X-axis. The 'Statistics' section is set to 'Logarithm of aggregate months of service'. The 'Error Bars Represent' section has 'Confidence intervals' selected with a level of 95%. The 'Linear Fit Lines' section has 'Total' selected. The 'Gallery' shows 'Scatter/Dot' as the chosen chart type.



Step-by-step Example 3.3

Simple Linear Regression cont.

Bayesian Regression

ANOVA^{a,b}

Source	Sum of Squares	df	Mean Square	F	Sig.
Regression	67.562	1	67.562	71.613	<.001
Residual	30.190	32	.943		
Total	97.751	33			

a. Dependent Variable: Logarithm of aggregate months of service
b. Model: (Intercept), Number of damage incidents

Bayes Factor Model Summary^{a,b}

Bayes Factor ^c	R	R Square	Adjusted R Square	Std. Error of the Estimate
11836625.05	.831	.691	.682	.9713

a. Method: JZS
b. Model: (Intercept), Number of damage incidents
c. Bayes factor: Testing model versus null model (Intercept).

Bayesian Estimates of Coefficients^{a,b,c}

Parameter	Mode	Posterior		95% Credible Interval	
		Mean	Variance	Lower Bound	Upper Bound
(Intercept)	6.097	6.097	.043	5.688	6.507
Number of damage incidents	.091	.091	.000	.069	.113

a. Dependent Variable: Logarithm of aggregate months of service
b. Model: (Intercept), Number of damage incidents
c. Assume standard reference priors.

Bayesian Estimates of Error Variance^a

Parameter	Mode	Posterior		95% Credible Interval	
		Mean	Variance	Lower Bound	Upper Bound
Error variance	.888	1.006	.072	.610	1.651

a. Assume standard reference priors.

Bayes Factor	Evidence Category	Bayes Factor	Evidence Category	Bayes Factor	Evidence Category
>100	Extreme Evidence for H0	1-3	Anecdotal Evidence for H0	1/30-1/10	Strong Evidence for H1
30-100	Very Strong Evidence for H0	1	No Evidence	1/100-1/30	Very Strong Evidence for H1
10-30	Strong Evidence for H0	1/3-1	Anecdotal Evidence for H1	1/100	Extreme Evidence for H1
3-10	Moderate Evidence for H0	1/10-1/3	Moderate Evidence for H1		

Assessment 2



qualtrics^{XM}



https://und.qualtrics.com/jfe/form/SV_0ufeUVznLiqeZJc

Caveats and Concerns

- Not always a strong difference between Frequentist and Bayesian
- Most of what we've covered is relatively trivial
- Move from familiarity to unfamiliarity
- Choice of priors can be important (Module III)
- Need to check fit of posterior modeling (Module III)

Real World Examples

[15]

Brard C, Le Teuff G, Le Deley MC, Hampson LV. Bayesian survival analysis in clinical trials: What methods are used in practice? Clin Trials. 2017 Feb;14(1):78-87. doi: 10.1177/1740774516673362. Epub 2016 Oct 11. PMID: 27729499.

A systematic review of clinical trials using Bayesian survival analyses was performed through PubMed and Web of Science databases.

In total, 28 articles met the inclusion criteria, 25 were original reports of clinical trials and 3 were re-analyses of a clinical trial.

Few trials implemented a Bayesian survival analysis and few incorporated external data into priors.

Table 3. Specification of the Bayesian survival analysis, overall and according to its purpose (monitoring or final analysis only).

	Monitoring, N = 14 ^a	Final analysis only, N = 14	Total, N (%)
Bayesian modelling in RCTs (N = 21)	10	11	21
Survival at a fixed time point ^b	1	1	2 (10)
Hybrid approach ^c	5	5	10 (48)
Semi-parametric model ^d	1	2	3 (14)
One parameter exponential model	3	–	3 (14)
Log-normal model	–	2	2 (10)
Parametric model unspecified	–	1	1 (5)
Bayesian modelling in single-arm trials (N = 7)	4	3	7
Survival at a fixed time point ^b	2	–	2 (29)
One parameter exponential model	1	–	1 (14)
Log-normal model	–	2	2 (29)
Unspecified model of a 'continuous' time-to-event endpoint	1	1	2 (29)
Prior distribution of the treatment effect (RCT, n = 21)	10	11	21
Historical data + experts + non-informative	–	1	1 (5)
Historical data + archetypal + non-informative	1	–	1 (5)
Historical data only	–	2	2 (10)
Archetypal + non-informative	1	–	1 (5)
Non-informative only	4	6	10 (48)
Unspecified	4	2	6 (29)
Prior distribution of the model parameter (Non-RCT, n = 7)	4	3	7
Non-informative	–	2	2 (29)
Unspecified	4	1	5 (71)
Minimum important efficacy effect pre-specified			
Yes	11	5	16 (57)
No	3	9	12 (43)
Decision rule pre-specified			
Yes	8	1	9 (32)
No	6	13	19 (68)

^aAmong the 14 trials performing Bayesian survival analyses for trial monitoring, three also planned and conducted a Bayesian final analysis of the survival endpoint.

^bSurvival at a fixed time point: for example, progression-free survival at 2 years.

^cHybrid approach: this consists of updating a prior distribution for the logarithm of the hazard ratio with the partial likelihood estimate of this parameter obtained from a frequentist analysis.

^dBayesian semi-parametric model: this consists of using the full data likelihood under a survival regression model with the baseline hazard function specified. The values given in Boldface represent totals and subtotals for each main category.

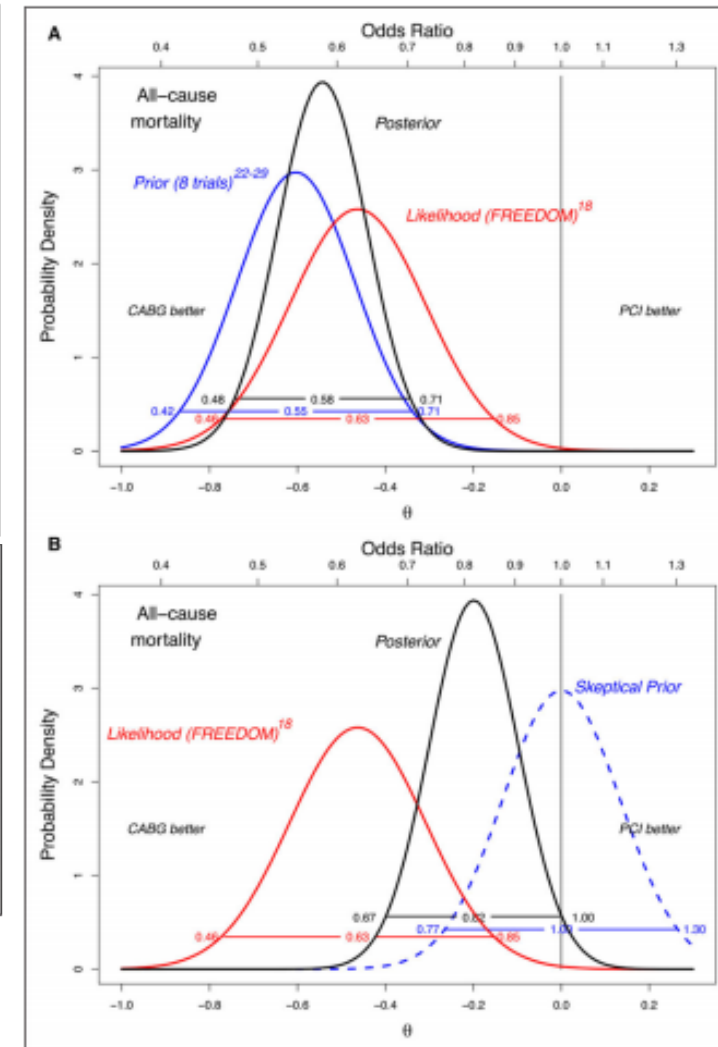
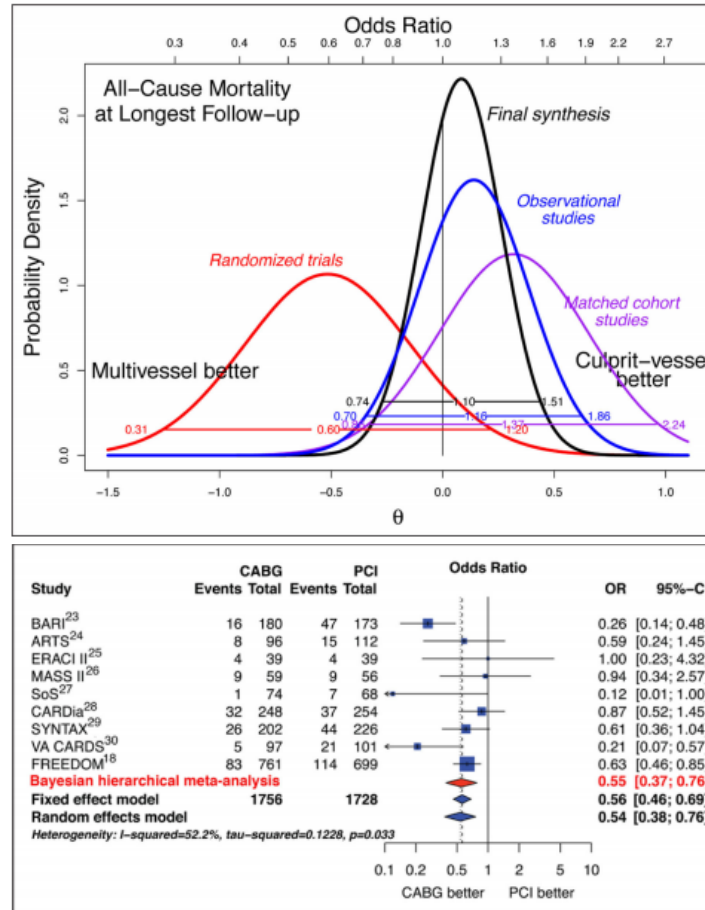
Real World Examples

[16]

Bittl JA, He Y. Bayesian Analysis: A Practical Approach to Interpret Clinical Trials and Create Clinical Practice Guidelines. *Circ Cardiovasc Qual Outcomes*. 2017;10(8):e003563. doi:10.1161/CIRCOUTCOMES.117.003563

In this review, we present gradually more complex examples, along with programming code and data sets, to show how Bayesian analysis takes evidence from randomized clinical trials to update what is already known about specific treatments in cardiovascular medicine.

To perform traditional meta-analyses, we use the open-source statistical program R and library package meta. To generate conjugate-normal models, we combine normal probability distributions from older trial data (prior) and new trial results (likelihood) to generate the posterior. To perform more complex computations, we use a version of BUGS called OpenBUGS that allows Markov chain Monte Carlo modeling to specify the posterior distribution.

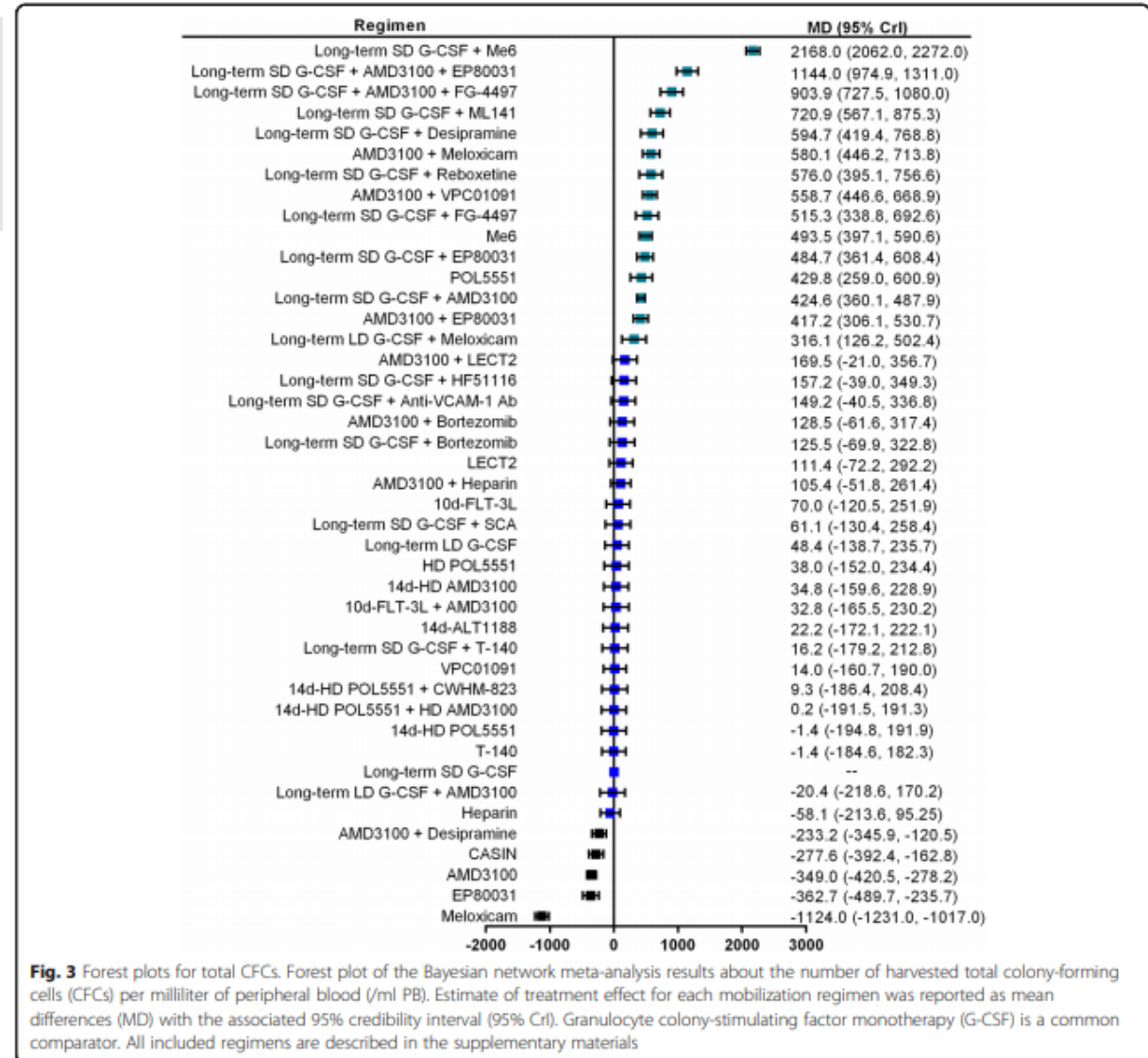
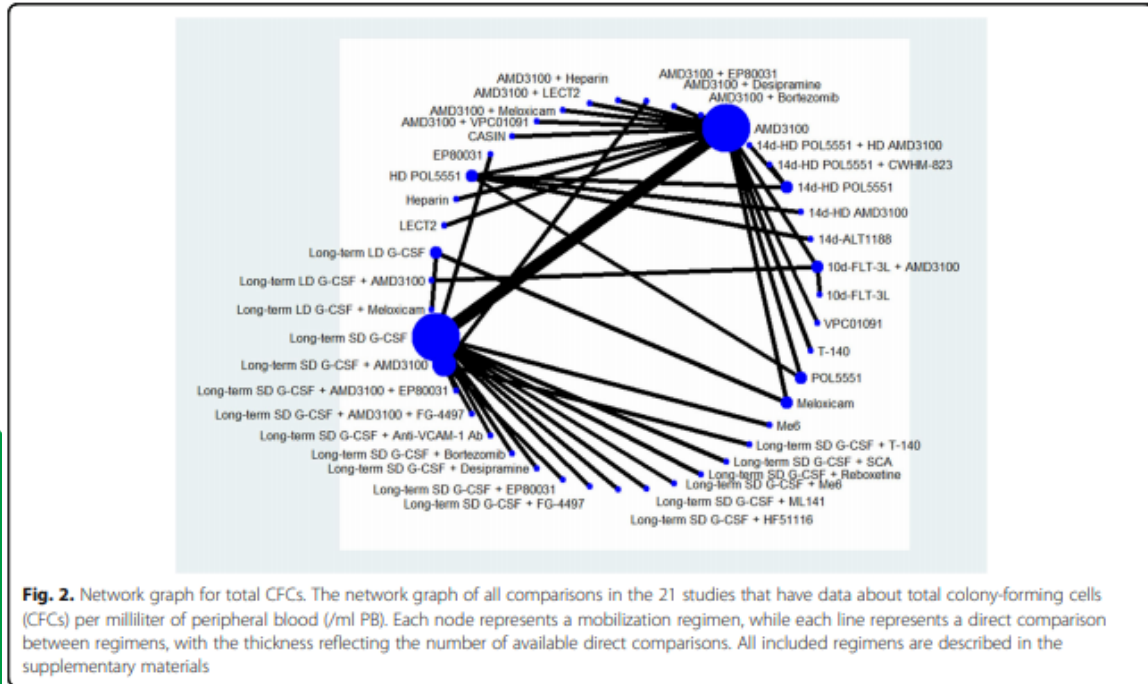


Real World Examples

[17]

Luo C, Wang L, Wu G, et al. Comparison of the efficacy of hematopoietic stem cell mobilization regimens: a systematic review and network meta-analysis of preclinical studies. *Stem Cell Res Ther.* 2021;12(1):310. Published 2021 May 29. doi:10.1186/s13287-021-02379-6

Bayesian network meta-analyses were performed following the guidelines of the National Institute for Health and Care Excellence Decision Support Unit (NICE DSU) with WinBUGS version 1.4.3.



Real World Examples

[18]

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5(3): e9490. doi:10.1371/journal.pone.0009490

FastTree recomputes all posterior distributions and recalculates the log likelihood of the tree at each site. FastTree then uses a Bayesian approach to select which rate to use at each site: FastTree maximizes $P(\text{rate}|\text{Site}) \propto Lk(\text{Site}|\text{rate}) \prod (\text{rate})$ where $\prod (\text{rate})$ is a gamma-distributed prior.

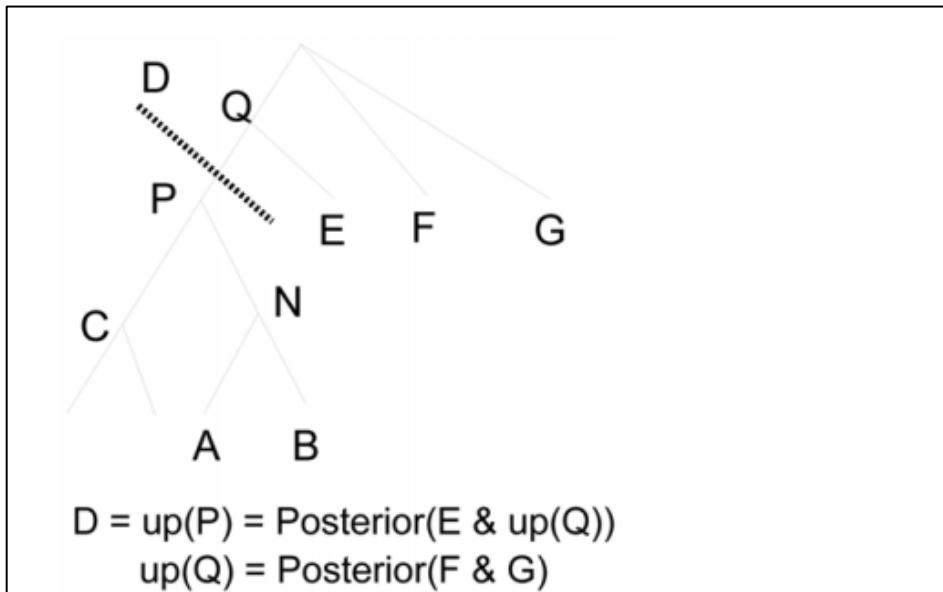


Figure 3. Traversing a tree with up-posteriors. FastTree optimizes the tree near node N by analyzing the posterior distributions for subtrees A, B, and C, as well as the “up-posterior” D.
 doi:10.1371/journal.pone.0009490.g003

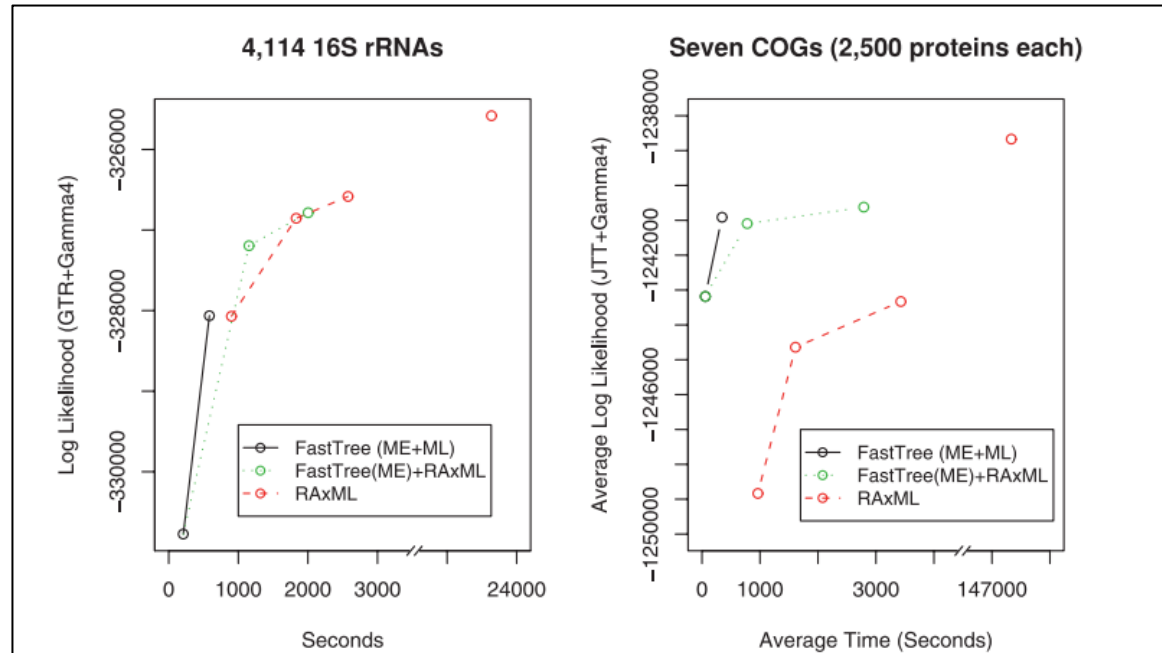


Figure 2. Likelihoods over time for genuine alignments. Each line shows the time it takes a different tool to reach a given likelihood. For the COG alignments, all times and likelihoods are averages over the seven alignments. For FastTree, we show the time and the improvement in likelihood for the minimum-evolution topology and the final (approximately-ML) topology. For RAxML, we show the maximum parsimony starting topology, the first two rounds of SPR moves, and the final topology (note the interrupted x axis). For RAxML with FastTree’s (minimum-evolution) starting tree, we show the starting topology and RAxML’s first two rounds of SPR moves.
 doi:10.1371/journal.pone.0009490.g002

Summary and Conclusion

- Bayesian analysis takes in prior beliefs (hypothesis) and evidence (data) to generate posterior beliefs (probability)
- The two major methods of using Bayesian analysis is 1) the Bayes factor for hypothesis testing (p-value analog) and 2) highest density interval for estimation with uncertainty (ML analog)
- Bayesian analysis can be run in R, SAS, and SPSS
- Tune in next time for deeper details into Bayesian analysis in Bayesian Analysis Module III: Deep Dive

References

- [1] https://www.digitalvidya.com/wp-content/w3-webp/uploads/2018/11/Bayesian_inference_archaeology_example.jpgw3.webp
- [2] <https://doingbayesiandataanalysis.blogspot.com/2016/12/bayesian-assessment-of-null-values.html>
- [3] <https://link.springer.com/content/pdf/10.3758/s13423-016-1221-4.pdf>
- [4] <https://stats.stackexchange.com/questions/246481/bayesian-parameter-estimation-or-bayesian-hypothesis-testing>
- [5] <https://people.stat.sc.edu/Hitchcock/stat535slidesday20.pdf>
- [6] https://en.wikipedia.org/wiki/Bayes_factor
- [7] <https://www.slideshare.net/ASQwebinars/general-bayesian-methods-for-typical-reliability-data-analysis>
- [8] <https://m-clark.github.io/models-by-example/bayesian-t-test.html>
- [9] <https://www.cairn.info/revue-l-annee-psychologique-2020-1-page-73.htm>
- [10] <https://gregorygundersen.com/blog/2020/02/04/bayesian-linear-regression/>
- [11] <https://cran.r-project.org/web/packages/BayesFactor/vignettes/manual.html>
- [12] <https://support.sas.com/resources/papers/proceedings14/SAS400-2014.pdf>
- [13] <https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=statistics-bayesian>
- [14] <https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=statistics-bayesian-independent-sample-inference>
- [15] <https://pubmed.ncbi.nlm.nih.gov/27729499/>
- [16] <https://pubmed.ncbi.nlm.nih.gov/28798016/>
- [17] <https://stemcellres.biomedcentral.com/track/pdf/10.1186/s13287-021-02379-6.pdf>
- [18] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009490>

Acknowledgements



- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY