

# Explicit Bayes: Working Concrete Examples to Introduce the Bayesian Perspective.

As published in Benchmarks RSS Matters, January 2015

<http://web3.unt.edu/benchmarks/issues/2015/01/rss-matters>

Jon Starkweather, PhD

Jon Starkweather, PhD  
jonathan.starkweather@unt.edu  
Consultant  
**Research and Statistical Support**



<http://www.unt.edu>



<http://www.unt.edu/rss>

RSS hosts a number of “Short Courses”.

A list of them is available at:

<http://www.unt.edu/rss/Instructional.htm>

Those interested in learning more about R, or how to use it, can find information here:

[http://www.unt.edu/rss/class/Jon/R\\_SC](http://www.unt.edu/rss/class/Jon/R_SC)

# Explicit Bayes: Working Concrete Examples to Introduce the Bayesian Perspective.

We use the term *explicit* because we are going to calculate these examples *by hand* with programming rather than simply loading a package and using functions to estimate parameters. The purpose of using these explicit methods is to hopefully convey a better understanding of what it means to *do* Bayesian statistics.

First, we must present a little bit about Bayesian statistics. Very, very briefly, Bayesian statistics requires three elements: a prior, likelihood, and a posterior. The prior is a distribution specified by the researcher which represents all *prior* information regarding the parameter the researcher is attempting to estimate. The prior represents an educated, best guess at the parameter (e.g. the mean of the prior) and the degree of certainty or confidence in that educated, best guess (e.g., the variance and shape of the prior distribution). The prior is specified before (i.e. *prior*) to data collection. The prior is then combined with the likelihood (a representation of the data at hand) to create a more informed, empirical distribution of the parameter being estimated. We call this last distribution the *posterior* distribution. The mean of the posterior is our estimate of the parameter. Interval estimates can then be calculated from the posterior which truly will represent the interval which contains the actual population parameter; we call those intervals *credible intervals* (rather than confidence intervals – which *do not* tell you the probability of the population parameter being contained in this interval).

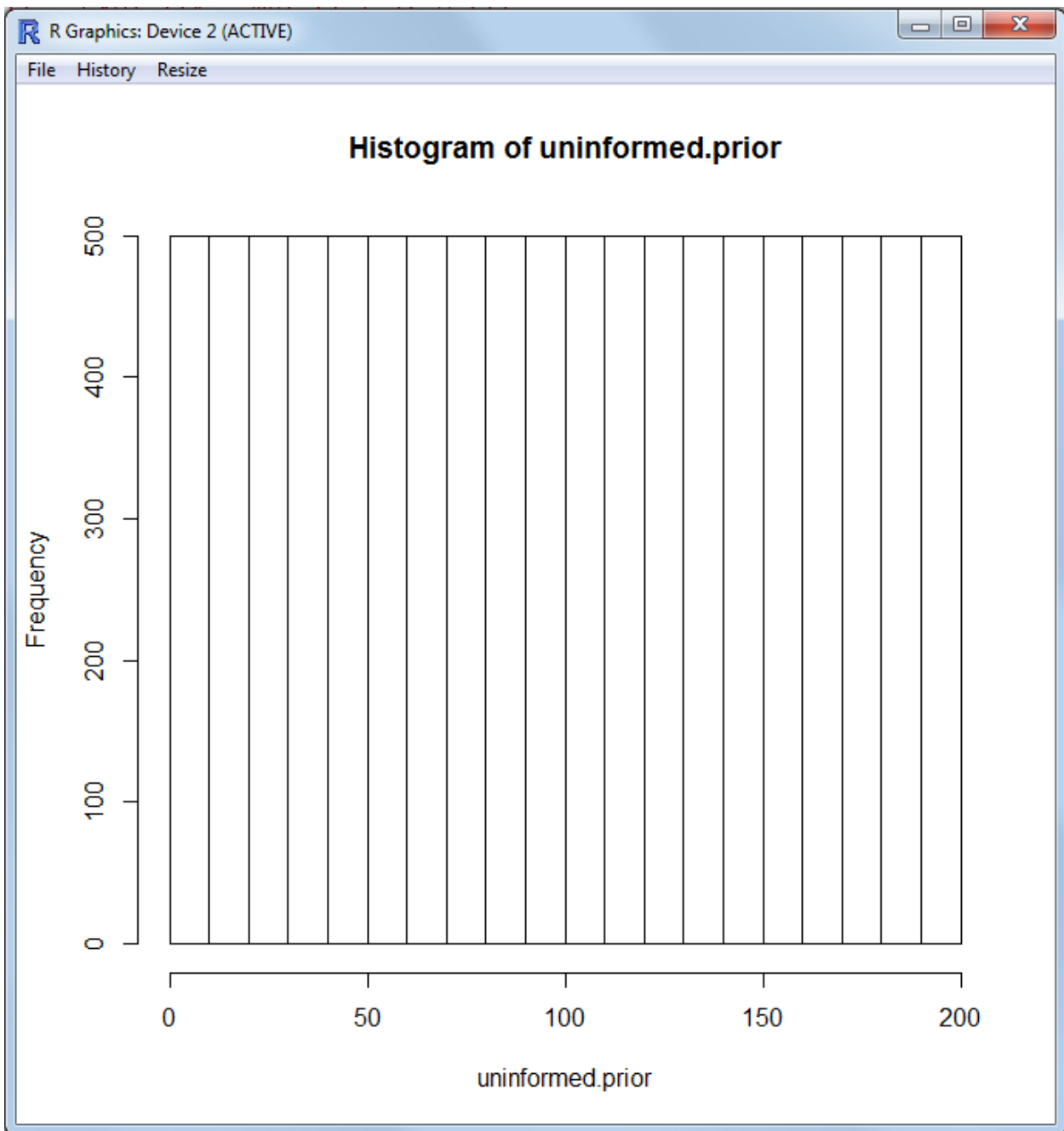
Let's say we want to estimate **the mean** IQ scores on the Wechsler Adult Intelligence Scale (WAIS) of a small town, X.Town, which has a population of 10000 individuals. Let's start by importing the X.Town data.

```
x.town.df <- read.table(
  "http://www.unt.edu/rss/class/Jon/ExampleData/X.Town.sample.txt",
  header = TRUE, sep = ",", na.strings = "NA", dec = ".")
nrow(x.town.df)
[1] 10000
```

We know from a mountain of normative data and prior research that the U.S. population distribution of WAIS scores has a mean ( $\mu$ ) of 100 and a standard deviation ( $\sigma$ ) of 15. This information represents a best case scenario; where we *know* the population distribution and that distribution is normally distributed with an identified mean and standard deviation. Generally, we would not have such great prior information; so consider an alternative where we have virtually no prior information except to know the WAIS questions / procedures which allow a possible score to range from 1 to 200. In such a case, our specification of a prior distribution would mean each score in that range is equally likely — which prompts us to specify a *uniform* distribution (i.e. a distribution in which each value has an equal probability of being represented). A uniform prior is also known as an un-informative or un-informed prior. In both examples below we are using a population of 10000 individuals.

```
uninformed.prior <- rep(seq(1:200), 50)
length(uninformed.prior)
[1] 10000
summary(uninformed.prior)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.00   50.75  100.50  100.50  150.20  200.00
```

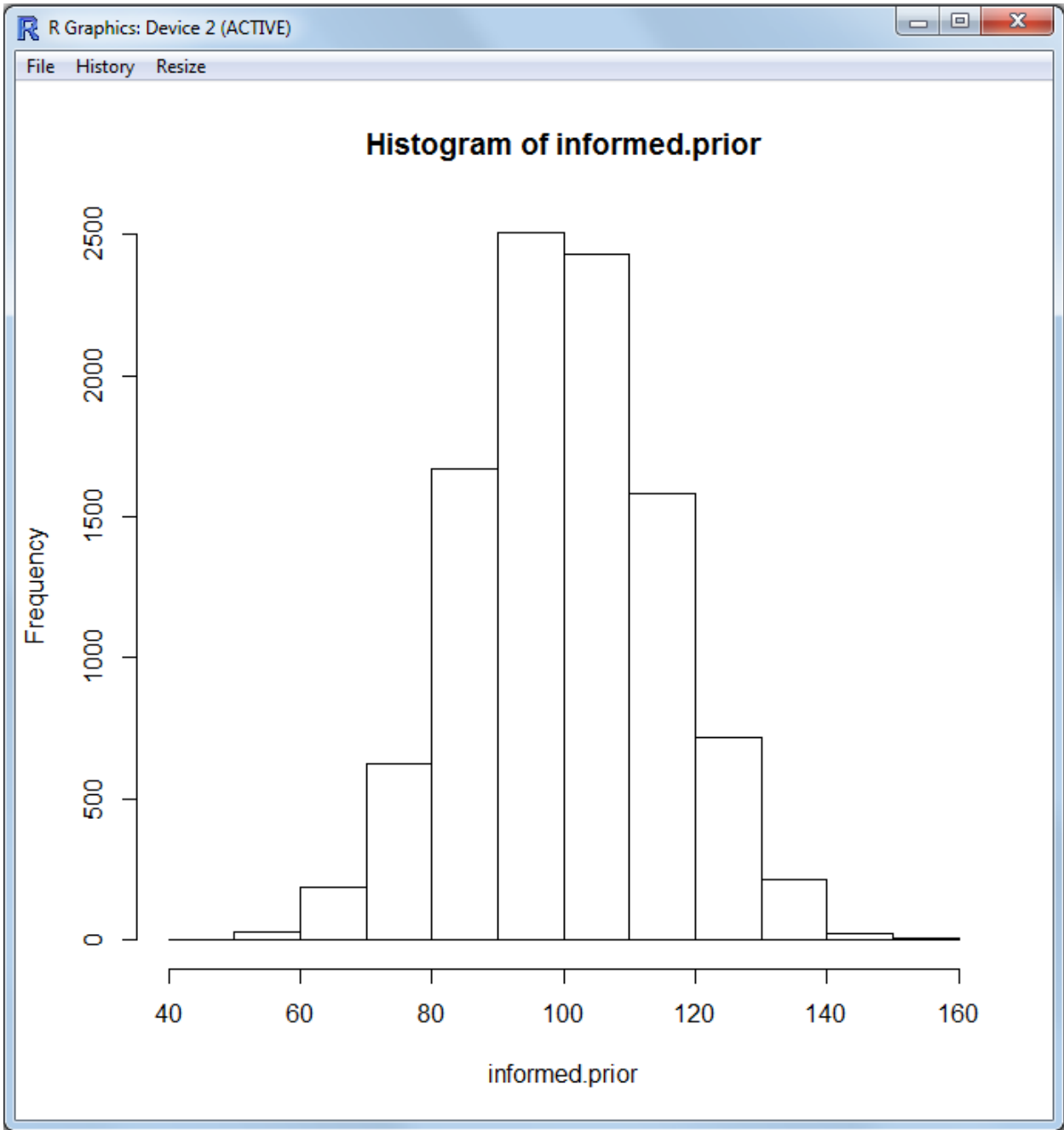
```
hist(uninformed.prior)
```



However, with the WAIS and the knowledge of the U.S. population, we can specify a Gaussian (i.e. normal) distribution as our prior.

```
informed.prior <- rnorm(10000, mean = 100, sd = 15)
length(informed.prior)
[1] 10000
summary(informed.prior)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 37.51  89.93  100.10  100.10  110.40  157.30
```

```
hist(informed.prior)
```



Clearly; the two example priors above are extremes (i.e. worst case and best case); there are a variety of other distributions which can be specified as priors (e.g. Cauchy, Poisson, beta, etc.) and the prior is not required to be symmetrical. For more information on the variety of distributions, see:

[http://en.wikipedia.org/wiki/List\\_of\\_probability\\_distributions](http://en.wikipedia.org/wiki/List_of_probability_distributions)

Our research questions are as follows: What is the mean WAIS score of the population ( $n = 10000$ ) of X.Town; and, does that mean differ from the larger (U.S.) population? In more precise terms, what is the population mean of X.Town WAIS scores and is that mean *larger* than the known U.S. population mean. To be clear, there are two populations we are referring to here; the population of X.Town ( $N = 10000$ ) and the larger population of the U.S.

It is unrealistic to think we would have all 10000 adult citizens' data from X.Town; we would generally have a sample of that town's data. Note; the 7th column of our X Town data file contains the WAIS scores. Here we randomly sample ( $n = 1000$ ) cases from the entire X.Town data ( $N = 10000$ ):

```
wais.sample <- sample(x.town.df[,7], 1000, replace = FALSE)
length(wais.sample)
[1] 1000
```

## 1 Traditional Frequentist Perspective: Null Hypothesis Significance Testing (NHST).

In a traditional *frequentist* setting, we would begin by simply calculating the sample mean as our best estimate of the entire X.Town population mean WAIS score:

```
M <- mean(wais.sample)
M
[1] 107.6305
```

and the standard error of that mean if we wanted confidence intervals for that estimate (of the entire X.Town's mean):

```
std.err <- sqrt(15^2 / length(wais.sample))
std.err
[1] 0.4743416
```

Then using an alpha value (e.g. 0.05) look up the associated critical value (i.e. +/-1.96) in a table; then calculate the lower and upper bounds of the confidence interval for our estimate (i.e. the confidence interval for the estimated mean of X.Town).

```
lower.bound <- (-1.96*std.err) + M
lower.bound
[1] 106.7008
upper.bound <- (+1.96*std.err) + M
upper.bound
[1] 108.5602
```

Then, we would run a one sample t-test using our random sample of X.Town adults' WAIS scores, comparing **the mean** of the sample scores ( $M$ ; as our best estimate of the entire X.Town's mean) to the mean of the U.S. population ( $\mu$ ); using the standard error of the mean ( $std.err$ ) and some pre-designated probability cutoff (e.g. 0.05) to determine statistical significance.

```
t.test(wais.sample, alternative = 'greater', mu = 100, conf.level = .95)
```

One Sample t-test

```

data: wais.sample
t = 17.0653, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 106.8944      Inf
sample estimates:
mean of x
 107.6305

```

It is important to recall (or review) what the above test is doing. We have drawn a random sample of data from X.Town and we are testing **the mean** of that sample against a known (U.S.) population mean to determine if the sample indeed comes from that population (i.e. the null hypothesis). Notice we are using the sample mean ( $n = 1000$ ) as a representation of the entire X.Town's WAIS scores ( $N = 10000$ ).

## 2 Bayesian Perspective: Bayesian Statistics; Bayesian Inference; Bayesian Parameter Estimation.

All three of the above terms are often used to refer to Bayesian data analysis. The examples below were all adapted from Kaplan (2014). Our example explores the normal prior for the normal sampling model in which the variance  $\sigma^2$  (sigma squared) is assumed to be known. Thus, the problem is one of estimating **the mean**  $\mu$  (mu). Let  $y$  denote a data vector of size  $n$  ( $y$  = the sample of 1000 WAIS scores). We assume that  $y$  follows a normal distribution shown with the equation below:

$$p(y|\mu, \sigma^2) = (1/\sqrt{2 * \pi * \sigma}) * \exp(-((y - \mu)^2)/(2 * \sigma^2)) \quad (1)$$

To clarify and show an example in R, we use the following:

```

mu <- 100
o <- 15
y <- wais.sample

```

We use the word 'output' to refer to  $p(y|\mu, \sigma^2)$  from above; which is read as the probability of  $y$ , given a mean of  $\mu$  ( $\mu$ ), and variance of sigma squared ( $\sigma^2$ ).

```

output <- (1/sqrt(2*pi*o)) * exp(-((y - mu)^2) / (2*o^2))
summary(output)
      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
0.000289 0.047630 0.078600 0.069690 0.096360 0.103000

```

Next, we specify the prior. We have plenty of confidence that our prior distribution of the mean is normal with its own mean and variance hyper-parameters,  $k$  and  $t^2$  (using  $t$  in R code to refer to tau:  $\tau$ ), respectively, which for this example are known. The prior distribution can be written as:

$$p(\mu|k, t^2) = (1/\sqrt{2 * \pi * t^2}) * \exp(-((\mu - k)^2)/(2 * t^2)) \quad (2)$$

The term  $p(\mu|k, t^2)$ , can be read as the probability of  $\mu$  given  $k$  and  $t^2$ .

```
k <- mean(y); k
[1] 107.6305
t <- sd(y); t
[1] 14.13976
n <- length(y); n
[1] 1000

prior.mean <- (1/sqrt(2*pi*t^2)) * exp(-((mu - k)^2) / (2*t^2))
prior.mean
[1] 0.02439102
```

Combine the prior information with the likelihood of the data (given the population variance; sigma squared  $[\sigma^2]$  and the sample size  $[n]$ ) to create the posterior distribution. Using some algebra, the posterior distribution can be obtained as:

$$p(\mu|y) \sim N\left[\frac{(k/t^2) + (n * \text{mean}(y)/\sigma^2)}{(1/t^2) + (n/\sigma^2)}, \frac{(t^2 * \sigma^2)/(\sigma^2 + (n * t^2))}\right] \quad (3)$$

Thus, the posterior distribution of  $\mu$  ( $\mu$ ) is normal with a mean:

```
posterior.mu <- ((k/t^2)+(n*mean(y)/o^2)) / ((1/t^2)+(n/o^2))
posterior.mu
[1] 107.6305
```

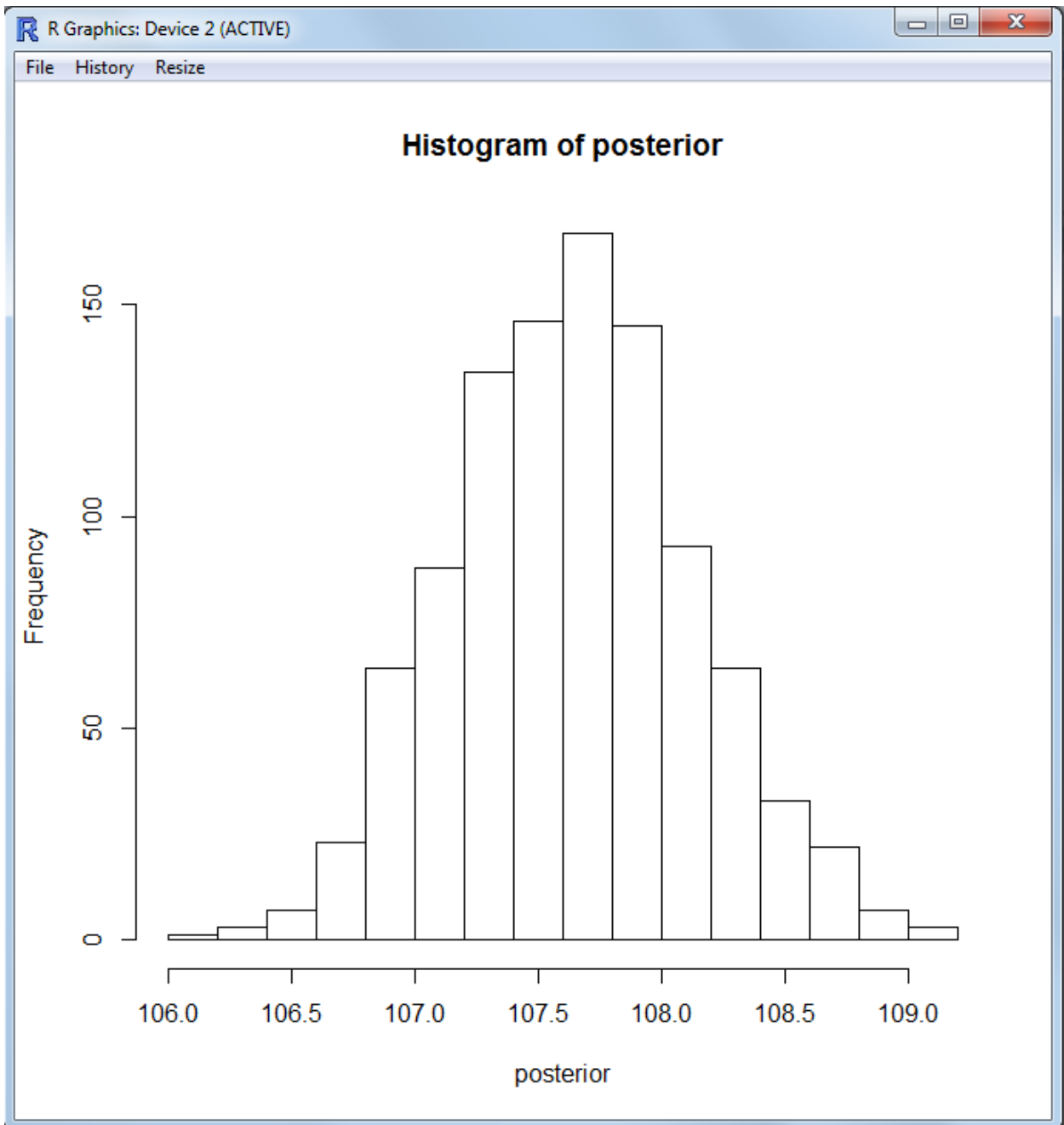
and variance:

```
posterior.o2 = (t^2*o^2)/(o^2+(n*t^2))
posterior.o2
[1] 0.2247471
```

So, the posterior distribution can be simulated using these two parameters (and  $n = 1000$ ); which in R, should be:

```
posterior <- rnorm(n = length(y), mean = posterior.mu,
                  sd = sqrt(posterior.o2))
hist(posterior)
```





In a traditional frequentist analysis, one would be required to report both the estimated mean (i.e. mean of the sample) and a confidence interval with lower and upper bounds of that mean. However, a frequentist confidence interval only tells us; if this same study was repeated 100 times, we would expect the sample mean to be between the upper and lower bounds 95 times (if using a 95% confidence interval). It **does not** tell us the probability of the population parameter being included in the interval. Here in the Bayesian setting, we use the posterior distribution and simply take the quantiles (i.e. probabilities) to compute the lower and upper bounds of a *credible interval* — which does give us the probability that the actual population parameter is included in this interval.

```
quantile(posterior, c(.05,.95))
5%      95%
```

106.8662 108.4625

It is critically important to recognize, the above example is **only** interested in estimating the mean of X.Town's WAIS scores. The example is NOT attempting to estimate the entire X.Town's distribution of WAIS scores. So let's compare the actual mean of X.Town's WAIS scores to the sample mean, and the mean of the posterior distribution (of course, in a real research situation you would not have the 'actual' parameter — i.e. mean of the entire population of X.Town).

```
mean(x.town.df$wais)
[1] 107.8662
mean(wais.sample)
[1] 107.6305
mean(posterior)
[1] 107.6389
```

Undoubtable readers will notice the virtually identical estimates provided by the mean of the posterior (i.e. Bayesian estimate) and simply the mean of the sample (i.e. frequentist estimate); and both of those are very, very close to the X.Town population mean. There are two very important reasons for this. First, the Bayesian and Frequentist methods will result in virtually the same parameter estimate(s) with large samples. The prior is weighted very lightly and the likelihood (a representation of the data at hand) contributes the bulk of the weight to the estimation when large samples are used in a Bayesian analysis. Second, the data used in the examples above is simulated data and a truly random sample ( $n = 1000$ ) was taken from the entire population ( $N = 10000$ ). Therefore, our results here have very low bias as a result of the truly random sample and the fact that 10% of the population was contained in the sample. Most research is not conducted on a truly random sample and very few research endeavors include 10% of the population as the sample.

Lastly, hypothesis testing and statistical significance are not foreign to the Bayesian perspective. For example, if one were interested in conducting a Bayesian  $t$ -test, you would use something called Bayes Factors which has been covered on the RSS Do-it-yourself Introduction to R web site<sup>1</sup> and specifically in Module 11<sup>2</sup>. Bayes Factors were also discussed in a previous RSS Matters article<sup>3</sup> (Adobe.pdf version<sup>4</sup>).

Until next time; *“knowledge is freedom and ignorance is slavery.”*

— The above quote is attributed to Miles Dewey Davis III (1926 - 1991):

[http://www.goodreads.com/author/quotes/54761.Miles\\_Davis](http://www.goodreads.com/author/quotes/54761.Miles_Davis)

### 3 Highly Recommended Reference

Kaplan, D. (2014). *Bayesian Statistics for the Social Sciences*. New York: The Guilford Press.

<sup>1</sup>[http://www.unt.edu/rss/class/Jon/R\\_SC/](http://www.unt.edu/rss/class/Jon/R_SC/)

<sup>2</sup>[http://www.unt.edu/rss/class/Jon/R\\_SC/Module10/BayesFactor.R](http://www.unt.edu/rss/class/Jon/R_SC/Module10/BayesFactor.R)

<sup>3</sup><http://web3.unt.edu/benchmarks/issues/2011/03/rss-matters>

<sup>4</sup>[http://www.unt.edu/rss/class/Jon/Benchmarks/BayesFactors\\_JDS\\_Mar2011.pdf](http://www.unt.edu/rss/class/Jon/Benchmarks/BayesFactors_JDS_Mar2011.pdf)

## 4 Other Important Resources

- Albert, J. (2007). *Bayesian Computation with R*. New York: Springer Science+Business Media, LLC.
- Berry, D. A. (1996). *Statistics: A Bayesian perspective*. Belmont, CA: Wadsworth Publishing Company.
- Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton, FL: Taylor & Francis Group, LLC.
- Bolker, B. M. (2008). *Ecological Models and Data in R*. Princeton, NJ: Princeton University Press.
- Bolstad, W. M. (2004). *Introduction to Bayesian statistics*. Hoboken, NJ: John Wiley & Sons, Inc.
- Broemeling, L. D. (2007). *Bayesian Biostatistics and Diagnostic Medicine*. Boca Raton, FL: Taylor & Francis Group, LLC.
- Congdon, P. (2005). *Bayesian Models for Categorical Data*. West Sussex, UK: John Wiley & Sons, Ltd.
- Congdon, P. (2006). *Bayesian Statistical Modeling*. West Sussex, UK: John Wiley & Sons, Ltd.
- Dey, D. K., Ghosh, S., & Mallick, B. K. (2011). *Bayesian Modeling in Bioinformatics*. Boca Raton, FL: Taylor & Francis Group, LLC.
- Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician*, 40, 1 - 5.
- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Gelman, A., & Meng, X. (2004). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. West Sussex, UK: John Wiley & Sons, Ltd.
- Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66, 8 - 38.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Science + Business Media, LLC.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. New York: Springer Science+Business Media, LLC.
- Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. West Sussex, UK: John Wiley & Sons,

Ltd.

Jeffreys, H. (1939). *Theory of Probability* (1st ed.). London: Oxford University Press.

Jeffreys, H. (1948). *Theory of Probability* (2nd ed.). London: Oxford University Press.

Koop, G. (2003). *Bayesian Econometrics*. Hoboken, NJ: John Wiley & Sons, Inc.

Koop, G., Poirier, D., & Tobias, J. (2007). *Bayesian Econometric Methods*. New York: Cambridge University Press.

Kruschke, J. K. (2011). *Doing Bayesian Data Analysis*. Burlington, MA: Academic Press.

Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Malden, MA: Blackwell Publishing.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction* (3rd ed.). New York: Oxford University Press Inc.

Link, W. A., & Barker, R. J. (2010). *Bayesian Inference with Ecological Applications*. London: Academic Press (Elsevier Ltd.).

Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer Science + Business Media, LLC.

Mallick, B., Gold, D. L., & Baladandayuthapani, V. (2009). *Bayesian Analysis of Gene Expression*. West Sussex, UK: John Wiley & Sons, Ltd.

Marin, J., & Robert, C. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York: Springer Science + Business Media, LLC.

McGrayne, S. B. (2011). *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven, CT: Yale University Press.

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. West Sussex, UK: John Wiley & Sons, Ltd.

Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225 - 237.

Sorensen, D., & Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York: Springer Science + Business Media, LLC.

Stone, L., D. (1975). *Theory of Optimal Search*. Mathematics in Science and Engineering, Vol. 118. New York: Academic Press, Inc.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer-Verlag, Inc.

Taroni, F., Bozza, S., Biedermann, A., Garbolino, P., & Aitken, C. (2010). *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. West Sussex, UK: John Wiley & Sons, Ltd.

Williamson, J. (2010). *In Defense of Objective Bayesianism*. Oxford, UK: Oxford University Press.

Woodworth, G. G. (2004). *Biostatistics: A Bayesian Introduction*. Hoboken, NJ: John Wiley & Sons, Inc.

This article was last updated on December 9, 2014.

This document was created using L<sup>A</sup>T<sub>E</sub>X