Bayes Factors for *t* tests and one way Analysis of Variance; in **R**

Dr. Jon Starkweather

It may seem like *small potatoes*, but the Bayesian approach offers advantages even when the analysis to be run is not complex. For instance, a traditional frequentist approach to a *t* test or one way Analysis of Variance (ANOVA; two or more group design with one outcome variable) would result in a *p* value which would be interpreted as the probability of the data (result) assuming the null hypothesis is true. Often, the *p* value's interpretation is abbreviated and it is interpreted as indicating empirical support for or against a null hypothesis. Of course, an effect size measure such as Cohen's *d* (Cohen, 1988) would also be computed to offer insight as to the magnitude of the effect. Wetzels, Matzke, Lee, Rouder, Iverson, and Wagenmakers (submitted) have advocated use of the Bayesian perspective for simple two or more group designs with the use of Bayes Factors. The advantage being that Bayes Factors "incorporate inferences about both the presence of effects, as well as their magnitude…" (Wetzels, et al., p. 1). Kass and Raftery (1995) define the Bayes factor as "a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (p. 777). Simply stated, the Bayes factor is a number, a ratio of one model's odds over the odds of another model. Another way of thinking about the meaning of a Bayes factor is that it is the resultant odds from dividing the likelihood of one model (e.g. the null hypothesis) by the likelihood of another model (e.g. the alternative hypothesis). At the risk of figuratively beating a dead horse; the Bayes factor can be thought of as the result of a *simple* fraction or division, in which the probability of one model (null) is divided by the probability of a second model (alternative). Therefore, its interpretation is very straightforward. For instance, a Bayes factor of 1.00 represents equal odds for either model (the null and alternative hypotheses), a Bayes factor greater than 1.00 represents evidence for the one model (e.g. the null hypothesis), and a Bayes factor less than 1.00 represents evidence for another model (e.g. the alternative hypothesis). The interpretation of magnitude for a Bayes factor, like traditional effect size estimates, involves some flexible categories (suggested by Jeffreys, 1961). For instance, a Bayes factor between (*roughly*) 1.00 and 3.00 (or between 1 and 0.30) represents scarce evidence, a Bayes factor between (roughly) 3.00 and 10.00 (or between 0.30 and 0.10) represents substantial evidence, a Bayes factor between (roughly) 10.00 and 30.00 (or between 0.10 and 0.03) represents strong evidence, and a Bayes factor between (roughly) 30.00 and 100.00 (or between 0.03 and 0.01) represents very strong evidence (Jeffreys). It is important to note; theoretically, there is no limit to the magnitude of a Bayes factor, Jeffreys suggested that a Bayes factor greater than 100.00 (or less than 0.01) would represent decisive evidence. So, the benefits of taking a Bayesian perspective (beyond the general reasons for choosing a Bayesian perspective over a frequentist perspective) are that in these simple situations, a Bayes factor is one number which is easily interpreted for both identifying an effect and measuring the magnitude of the effect. By contrast, the frequentist *p* value is easily confused, controversial, and would involve another statistic to express the magnitude of effect (i.e. effect size; e.g. Cohen's *d*).

Implementing the use of Bayes factors is very easy to do when working in R. The package 'BayesFactorPCL' (Morey & Rouder, 2010) provides functions for the computation of Bayes factors for

one sample or two sample *t* tests, as well as for one way ANOVA. The package is relatively new and is still being developed, so it is only available (for now) from R-Forge. However the functions for *t* tests and one way ANOVA are stable. The package authors are working on implementing a function for applying Bayes factors to regression and that is likely why the package has not yet been released to CRAN.

To explore some examples of Bayes factors analysis using the functions in the 'BayesFactorPCL' package, begin by importing some data from the web naming it 'example.1'. In R, load the 'foreign' library (necessary to import SPSS.sav files; which this example uses), then import the data, and then get a summary of the data if desired.

```
> library(foreign)
> example.1 <- read.spss("http://www.unt.edu/rss/class/Jon/R_SC/Module3/ExampleData1.sav",
+     use.value.labels=TRUE, max.value.labels=Inf, to.data.frame=TRUE)
> summary(example.1)
       ID              Sex           Age                       Ethnicity
 Min.   : 1.00   Female:40   Min.   :18.00   European American:33
 1st Qu.:14.25   Male  :14   1st Qu.:20.00   African American : 8
 Median :27.50               Median :21.00   Asian American   : 4
 Mean   :27.50               Mean   :21.04   Native American  : 3
 3rd Qu.:40.75               3rd Qu.:22.00   Canadian American: 4
 Max.   :54.00               Max.   :28.00   Other            : 2

       Cl_St            Fam_Income          Candy                 Stimuli
 Freshman : 2    55,001 to 65,000:29   Skittles:27   Printed            :18
 Sophomore: 6    45,001 to 55,000:15   None    :27   Spoken             :18
 Junior   :34    65,001 to 75,000: 5                 Printed and Spoken:18
 Senior   :12    35,001 to 45,000: 2
 Graduate : 0    75,001 to 85,000: 2
                 85,001 to 95,000: 1
                 (Other)         : 0
            Distraction     Recall1           Recall2          Skittles
 No Distraction    :18   Min.   : 4.00   Min.   : 3.68   Min.   :0.0
 Cell Phone Ring   :18   1st Qu.: 9.00   1st Qu.: 8.28   1st Qu.:0.0
 Light Bulb Failure:18   Median :13.00   Median :11.96   Median :0.5
                         Mean   :13.35   Mean   :12.08   Mean   :0.5
                         3rd Qu.:17.00   3rd Qu.:14.70   3rd Qu.:1.0
                         Max.   :25.00   Max.   :23.00   Max.   :1.0

    Printed            Spoken          CellPhone          LightBulb
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
 Mean   :0.3333   Mean   :0.3333   Mean   :0.3333   Mean   :0.3333
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

> |
```

Next, load the 'Rcmdr' and 'abind' packages using: library(Rcmdr) and library(abind) in the console. We will need these packages for the 'numSummary' functions (used below) which display descriptive statistics information in a cross-tabs manner.

```
> library(Rcmdr)
Loading required package: tcltk
Loading Tcl/Tk interface ... done
Loading required package: car
Loading required package: MASS
Loading required package: nnet
Loading required package: survival
Loading required package: splines

Rcmdr Version 1.6-4


Attaching package: 'Rcmdr'

The following object(s) are masked from 'package:tcltk':

    tclvalue

> library(abind)
>
```

Finally, we can load the 'BayesFactorPCL' library.

```
> library(BayesFactorPCL)
Loading required package: coda
Loading required package: lattice
Loading required package: mvtnorm
>
```

If you would like more information about the 'BayesFactorPCL' library, simply consult the help documentation, by typing: help(BayesFactorPCL) in the console.

t test example.

In order to see what we are using as an example, we can use the 'numSummary' function from the Rcmdr and abind packages. Our dependent variable is number of words recalled (Recall1) and our independent variable is type of candy given to participants (Candy) where some participants were given Skittles (Skittles) and some participants were given no candy (none).

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> help(BayesFactorPCL)
> numSummary(example.1$Recall1 , groups=example.1$Candy,statistics=c("mean", "sd"))
              mean        sd  n
Skittles   9.407407 3.201673 27
None      17.296296 4.204529 27
>
```

Next, we can use the 'tapply' function to calculate the variances (var) of each group and use the 'leveneTest' function to test the assumption of homogeneity of variances.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> tapply(example.1$Recall1, example.1$Candy, var, na.rm=TRUE)
 Skittles      None
10.25071  17.67806
> leveneTest(example.1$Recall1, example.1$Candy, center=median)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value  Pr(>F)
group   1   3.249 0.07726 .
       52
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

The output (above) shows that the variances (Skittles = 10.25, None = 17.68) are not significantly different ($p = 0.077$). Below a box and whisker plot displays fairly clearly how the groups differ.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> boxplot(example.1$Recall1 ~ example.1$Candy, col = "lightgreen")
>
```
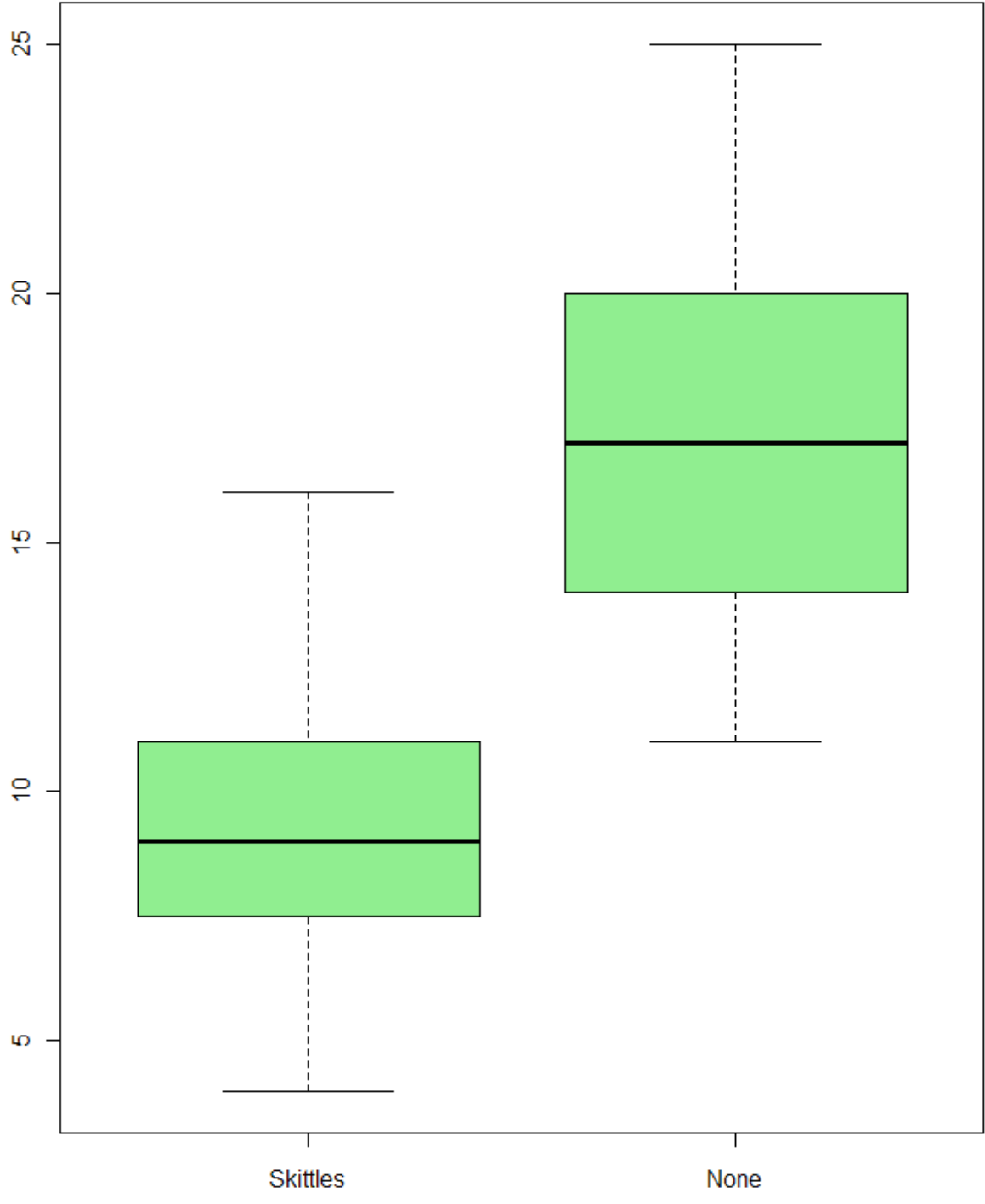
Next, we can conduct a traditional *t* test for comparison with the Bayes factor; and in R we need the actual *t* value to calculate the Bayes factor later.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> t.t1 <- t.test(Recall1~Candy, alternative="less", conf.level=.95, var.equal=TRUE, data=example.1)
> t.t1

        Two Sample t-test

data:  Recall1 by Candy
t = -7.7566, df = 52, p-value = 1.546e-10
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
    -Inf -6.18564
sample estimates:
mean in group Skittles     mean in group None
            9.407407              17.296296

>
```

Here, we see that participants in the Skittles group ($M = 9.41$, $SD = 3.20$) recalled significantly fewer words than participants in the group which received no candy ($M = 17.30$, $SD = 4.20$), $t(52) = -7.7566$, $p < .001$. Of course, we would want to take a look at the effect size, Cohen's *d* (1988), also called the Standardized Measure of Difference (SMD). To calculate the SMD in R, we need to first split the two groups' scores of the dependent variable.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> attach(example.1)
> x1 <- split(Recall1, Candy)
> x1
$Skittles
 [1]  9  9 11  8  7  6 10  8  9  8  8  9  5  4  5  7  7  8 15 15 16 11 10  9 14 13 13

$None
 [1] 19 18 18 16 15 15 17 17 18 14 13 14 11 12 12 12 14 13 24 25 25 20 20 19 21 22 23

> detach(example.1)
>
```

Next, load the 'MBESS' library which contains the 'smd' function. Then we can apply that function to our groups. Here (below) we see a substantially large effect size ($d = -2.11$).

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> library(MBESS)
> smd(x1$Skittles, x1$None)
[1] -2.111081
>
```

Next, we can conduct a Bayesian version of the Levene's test for homogeneity of variances; using the 'BayesFactorPCL' library function 'eqVariance.Gibbs'; which requires a matrix of data with each group as a column and each row a case.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> x2 <- cbind(x1$Skittles, x1$None)
> is.matrix(x2)
[1] TRUE
> |
```

The output for the 'eqVariance.Gibbs' function is quite long, but the key feature is the "$BF" is the Bayes Factor, which if greater than one indicates the groups' variances are equal.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> eqVariance.Gibbs(x2, iterations = 1000, whichModel = 2, M2.metrop.scale = 2)
  |===============================================================================================
Acceptance rates:
 decorr: 0.31 , g: 0.277 0.28
$chains
Markov Chain Monte Carlo (MCMC) output:
Start = 1
End = 1000
Thinning interval = 1
              mu1       mu2           g1            g2        IWMDE       sig2        sig2g decorr.acc g.acc1 g.acc2
   [1,]  8.772807  16.76373 -0.235543489  0.0000000000  7.546769e-03  19.287131  0.018904262          0      1      0
   [2,]  7.768447  16.03884 -0.235543489  0.0000000000  2.149459e+00  17.499977  0.039331409          0      0      0
   [3,]  9.479601  16.45494 -0.235543489  0.0000000000  1.248658e+00  13.809910  0.039288162          0      0      0
   [4,]  8.651388  17.41643 -0.235543489  0.0000000000  1.555051e+00  14.694846  0.046806277          0      0      0
```

Long section of output omitted to save space.

```
 [997,] 10.363959  16.93024 -0.070694771 -0.1509372321  2.334044e+00  20.740852  0.049645236          0      0      0
 [998,] 10.384451  17.66298 -0.125614376 -0.1363191001  7.440975e-01  15.741199  0.012870191          0      1      1
 [999,]  9.917560  17.85115 -0.125614376 -0.1363191001  2.522682e+00  10.941717  0.021798796          0      0      0
[1000,] 10.353809  17.75046 -0.125614376 -0.1363191001  1.893171e+00  14.297340  0.035625100          0      0      0

$BF
[1] 1.145594

$acc.rates
  decorr g.g.acc1 g.g.acc2
   0.310    0.277    0.280

>
```

Now, we can take the information from the traditional *t* test and conduct the Bayes factor analysis using the 'ttest.Quad' function from the 'BayesFactorPCL' library.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> t.b1 <- ttest.Quad(t = -7.7566, n1 = 27, n2 = 27, rscale = 1, prior.cauchy = TRUE)
> t.b1
[1] 3.427793e-08
> |
```

Notice the script above specifies the default for the prior; Cauchy, which is preferred (Rouder, Speckman, Sun, & Morey, 2009). When that argument is specified as FALSE, a normal prior distribution is applied. The normal prior applied to this (example) data changes the result very little (see below).

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> t.b2 <- ttest.Quad(t = -7.7566, n1 = 27, n2 = 27, rscale = 1, prior.cauchy = FALSE)
> t.b2
[1] 4.140999e-08
>
```

Use of either type of prior results in a very, very small Bayes factor, indicating decisive evidence for the alternative hypothesis (i.e. there is a difference between the groups in the number of words recalled and the magnitude of effect is 'decisive'). For more information on the 'ttest.Quad' function, simply type help(ttest.Quad) in the R console to bring up the function documentation.

```
R R Console
File  Edit  Misc  Packages  Windows  Help

> help(ttest.Quad)
>
```

```
R R Help on 'ttest.Quad'
File  Edit  View

ttest.Quad              package:BayesFactorPCL              R Documentation

A function to compute the Bayes factors for one- or two-sample designs

Description:

      This function computes the Bayes factor corresponding to a
      one-sample or two-sample t~test, using Gaussian quadrature
      integration.  See details about priors in the details.

Usage:

      ttest.Quad(t, n1, n2 = 0, rscale = 1, prior.cauchy = TRUE)

Arguments:

       t: The t value obtained from the corresponding t~test

      n1: Sample size for first group

      n2: Sample size for second group. If a one-sample test is
          desired, 'n1' should be 0.

  rscale: The prior scale (see Details below)
```

One way ANOVA example.

First, take a look at the variables of interest, here we are testing the number of words recalled (Recall1) of three distraction groups (No Distraction, Cell Phone Ring, & Light Bulb Failure).

```
R R Console
File Edit Misc Packages Windows Help

> numSummary(example.1$Recall1 , groups=example.1$Distraction,statistics=c("mean", "sd"))
                        mean       sd   n
No Distraction       15.00000 5.656854 18
Cell Phone Ring      11.38889 5.192063 18
Light Bulb Failure   13.66667 5.110543 18
>
> |
```

Next, as was done above, take a look at the variances of each group and evaluate the homogeneity of variance assumption.

```
R R Console
File Edit Misc Packages Windows Help

> tapply(example.1$Recall1, example.1$Distraction, var, na.rm=TRUE)
    No Distraction    Cell Phone Ring Light Bulb Failure
          32.00000           26.95752           26.11765
> leveneTest(example.1$Recall1, example.1$Distraction, center=median)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.0916 0.9127
      51
> |
```

A box-and-whisker plot shows how the groups' number of words recalled were distributed.

```
R R Console
File Edit Misc Packages Windows Help

> boxplot(example.1$Recall1 ~ example.1$Distraction, col = "lightgreen")
>
> |
```
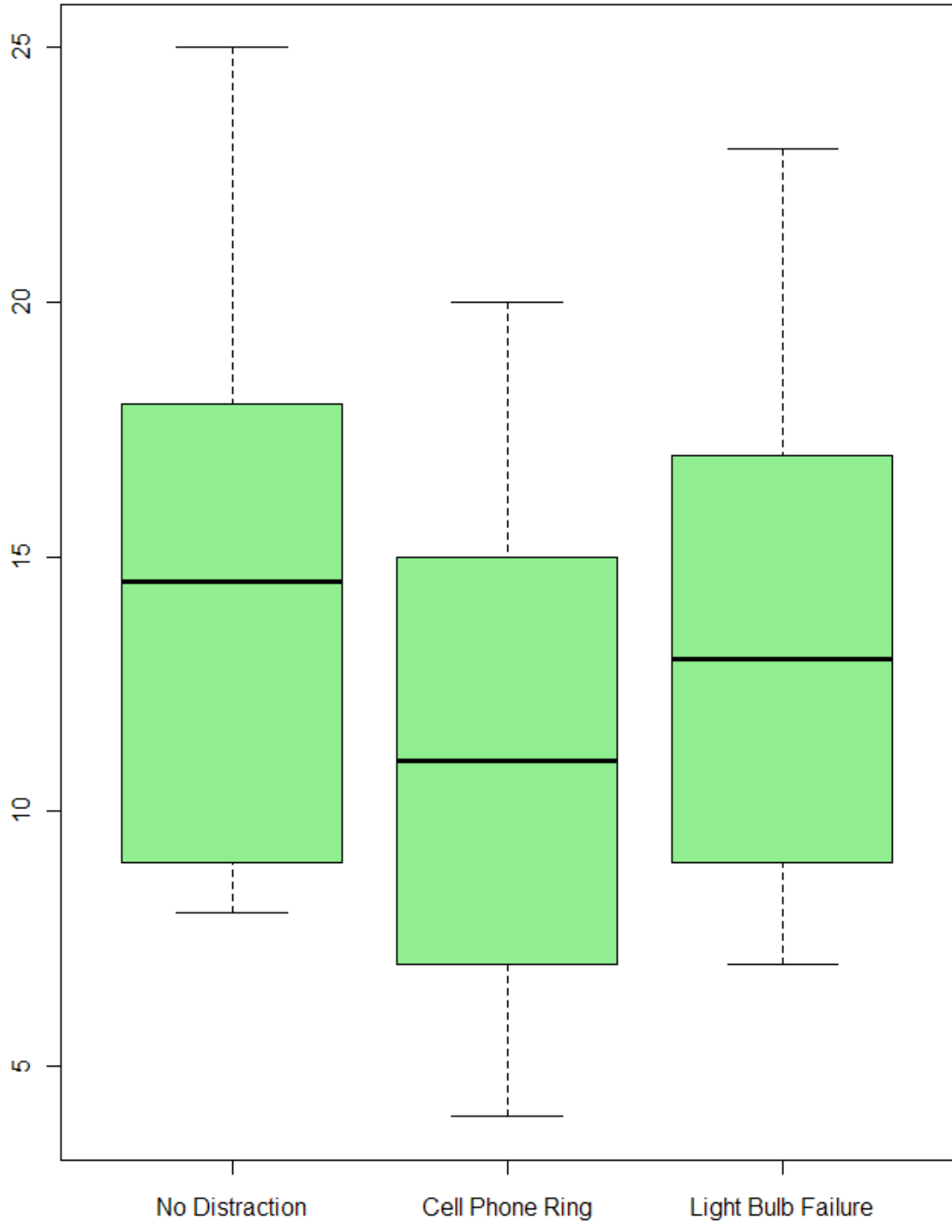
Next, we can conduct the traditional ANOVA. We see (below), there does not appear to be a significant difference in the number of words recalled among the groups.

```
> aov.t1 <- aov(Recall1 ~ Distraction, data=example.1)
> summary(aov.t1)
             Df  Sum Sq Mean Sq F value Pr(>F)
Distraction   2  120.04  60.019  2.1164 0.1309
Residuals    51 1446.28  28.358
>
```

Now, we can compute the Bayes factor using the 'oneWayAOV.Quad' function from the 'BayesFactorPCL' library. Notice the 'rscale = 1' argument which specifies a non-informative Jeffreys, standard multivariate prior; which is the default and recommended (Morey & Rouder, 2010; Rouder, Speckman, Sun, & Morey, 2009).

```
> aov.b1 <- oneWayAOV.Quad(F = 2.1164, N = 18, J = 3, rscale = 1)
> aov.b1
[1] 3.234245
>
```
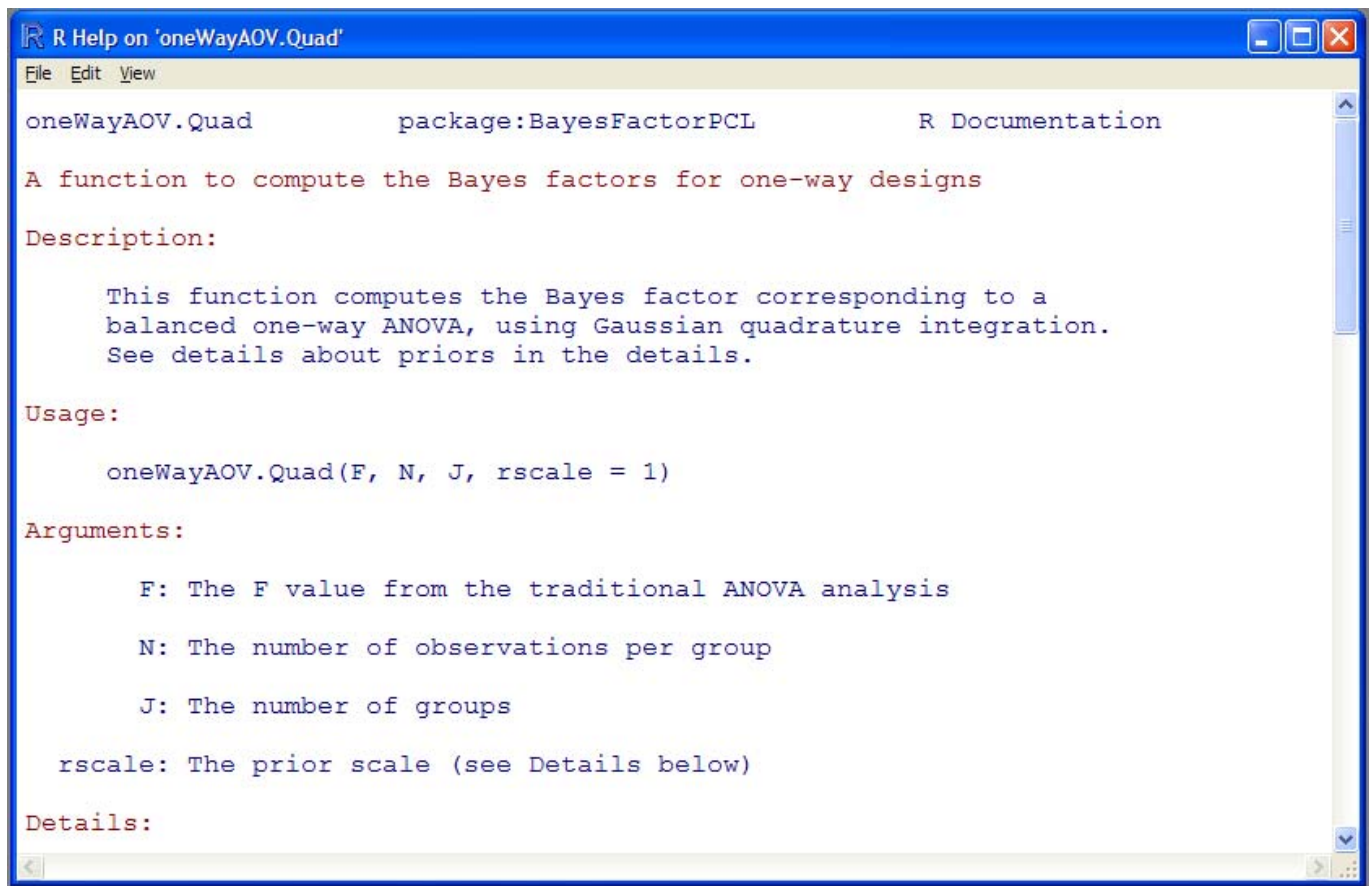
The 3.23 Bayes factor indicates there was 'substantial' evidence from this data, to support the null model or null hypothesis (i.e. consistent with the traditional approach results; there is no significant difference among the groups).

For more information on the 'oneWayAOV.Quad' function, simply type help(oneWayAOV.Quad) in the R console.

```
> help(oneWayAOV.Quad)
>
```

```
R Help on 'oneWayAOV.Quad'
File  Edit  View

oneWayAOV.Quad         package:BayesFactorPCL         R Documentation

A function to compute the Bayes factors for one-way designs

Description:

     This function computes the Bayes factor corresponding to a
     balanced one-way ANOVA, using Gaussian quadrature integration.
     See details about priors in the details.

Usage:

     oneWayAOV.Quad(F, N, J, rscale = 1)

Arguments:

       F: The F value from the traditional ANOVA analysis

       N: The number of observations per group

       J: The number of groups

   rscale: The prior scale (see Details below)

Details:
```

The 'LearnBayes' package, which is a companion for the book *Bayesian Computation with R*, both of which authored by Jim Albert (2010, 2007); also contains functions for computing Bayes Factors.

An Adobe.pdf version of this article can be found here.

Until next time, "*now all the criminals in their coats and their ties; are free to drink martinis and watch the sun rise…*" (Dylan & Levy, 1975)

References / Resources

Albert, J. (2007). *Bayesian computation with R.* New York: Springer Science+Business Media, LLC.

Albert, J. (2010). Package 'LearnBayes'. Available at CRAN: http://cran.r-project.org/web/packages/LearnBayes/index.html

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). Lawrence Erlbaum Associates.

Dylan, B., & Levy, J. (1975). *Hurricane.* Album: Desire.

Jeffreys, H. (1961). *Theory of probability.* Oxford, UK: Oxford University Press.

Geyer, C. J. (2010). Bayes factors via serial tempering. Available at: http://cran.r-project.org/web/packages/mcmc/vignettes/bfst.pdf

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*, 773 -- 795. Available at: http://www.stat.cmu.edu/~kass/papers/bayesfactors.pdf

Morey, R. D., & Rouder, J. N. (2010). Package 'BayesFactorPCL'. Available at: https://r-forge.r-project.org/projects/bayesfactorpcl/

Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. D. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225 -- 237. Available at: http://pcl.missouri.edu/node/32

Wetzels, R., Matzke, D., Lee, M., Rouder, J., Iverson, G., & Wagenmakers, E. (submitted). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. Available at: http://www.socsci.uci.edu/~mdlee/WetzelsEtAl2010.pdf