Homework 6
Syracuse University
IST 772
Summer 2021

```
library(BayesFactor)
library(BEST)
```

Question 1

```
# dimensions of the InsectSprays dataset
dim(InsectSprays)
```

```
## [1] 72  2
```

```
# preview the InsectSprays dataset
head(InsectSprays)
```

```
##   count spray
## 1    10     A
## 2     7     A
## 3    20     A
## 4    14     A
## 5    14     A
## 6    12     A
```

```
tail(InsectSprays)
```

```
##    count spray
## 67    13     F
## 68    10     F
## 69    26     F
## 70    26     F
## 71    24     F
## 72    13     F
```

```
# the dependent variable = count of insects killed
# independent variable = type of insect spray
# number of observations = 72
```

Question 2

```
# the between groups variance is 533.8
# the within groups variance is 15.4
```

Question 3

```
# calculate the F ratio
533.8 / 15.4
```

```
## [1] 34.66234
```

```
# the F ratio is very high indicating that it is very likely that the
# samples being compared are from different populations
```

```
# in this case, I would reject the null hypothesis that the samples
# were sampled from the same population
```

Question 4

```
# calculate the degrees of freedom between groups
# 6 groups - 1 grand mean = 5 df between groups
# 72 obs - 6 groups = 66 df within groups

# the reason that these add up to one less than the total number of
observations
# in the dataset is because we are always borrowing one degree of freedom
from
# the grand mean.
```

Question 5

```
# run the aov() command on the InsectSpray dataset
InsectSprayOut <- aov(formula = count ~ spray,
                      data = InsectSprays)

# show the results from the aov()
summary(InsectSprayOut)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## spray        5   2669   533.8    34.7 <2e-16 ***
## Residuals   66   1015    15.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# the null hypothesis is that the data was sampled from the same population.
# the alternative hypothesis is that the data was sampled from different
populations.
# Given F(5,66) = 34.7, p < 0.001, I would reject the null hypothesis.
# There is strong evidence that the data was sampled from different
populations.
```

Question 6

```
# load the bayes factor package
# already loaded earlier

# run the anovaBF() command on the InsectSprays dataset
BayesInsectSprayOut <- anovaBF(formula = count ~ spray,
                       data = InsectSprays)

# summarize the Bayes results
summary(BayesInsectSprayOut)
```

```
## Bayes factor analysis
## --------------
## [1] spray : 1.506706e+14 ±0%
```

```
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

```
# produce posterior distributions with the posterior command
# posteriors <- posterior(BayesInsectSprayOut, iterations = 10)

# could not run this due to computation issues

# the null hypothesis is that the groups were sampled from the same
population
# the alternative hypothesis is that the groups were sampled from different
populations
```

Question 7

```
# run bayesian t test on groups C and F
BESTmcmc(InsectSprays$count[InsectSprays$spray == 'C'],
        InsectSprays$count[InsectSprays$spray == 'F'])
```

```
## Waiting for parallel processing to complete...done.
```

```
## MCMC fit results for BEST analysis:
## 100002 simulations saved.
##          mean      sd median  HDIlo  HDIup Rhat n.eff
## mu1     1.972  0.6367  1.959  0.723  3.236    1 53305
## mu2    16.511  2.1539 16.490 12.274 20.818    1 56537
## nu     32.349 28.9820 23.627  1.035 90.308    1 18905
## sigma1  2.046  0.5684  1.964  1.057  3.200    1 30016
## sigma2  6.871  1.8073  6.572  3.870 10.554    1 33417
##
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
## 'n.eff' is a crude measure of effective sample size.
```

```
# the null hypothesis is that the means are not different
# the alternative hypothesis is that the means are different

# I would reject the null hypothesis because there the HDI does not
# cross through 0. There is a 95% chance that the mean difference
# in insects killed between spray C and spray F is between 0.7 and 3.3.
```