IST 777

Joe Hernandez

8/16/2019

<center>Homework #6</center>

1. Determine dependent variable and independent variable of InsectSprays data set

   ```
   > ?InsectSprays
   > str(InsectSprays)
   'data.frame':   72 obs. of  2 variables:
    $ count: num  10 7 20 14 14 12 10 23 17 20 ...
    $ spray: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1
   1 1 ...
   ```

   The data set contains two variables where the independent variable is the spray (type of spray) and the dependent variable is the count (insect count).

2. Run standard ANOVA to InsectSprays data set
   Given: Spray Mean Sq = 533.8; Residuals Mean Sq = 15.4
   The between-groups variance is 533.8 as it takes the scaled variance mean of the population and the within-groups variance is 15.4 as it takes the variance from within groups variability what is left over when all the systematic variance is removed.

3. By hand calculate F-ratio and interpret results
   F ratio is between-groups variance/within-gropus variance
   ```
   > 533.8 / 15.4
   [1] 34.66234
   ```

   F ratio is ~34.66
   With an F-ratio of ~ 34.662 I would reject the null hypothesis because it is significantly higher than even the default F value of 1.

4. Calculate degrees of freedom between groups and within groups
   Given: 12 observations six variables

   ```
   > aov(count ~ spray, data = InsectSprays)
   Call:
      aov(formula = count ~ spray, data = InsectSprays)

   Terms:
                     spray Residuals
   Sum of Squares  2668.833  1015.167
   Deg. of Freedom       5        66
   ```

```
Residual standard error: 3.921902
Estimated effects may be unbalanced
```

The degrees of freedom between groups is 5 and the within groups is 66.

These two values add up to one less than the total number of observations to account for an adjustment to lose information from calculating the mean in order to calculate variance and standard deviation

5. Execute aov on InsectSprays data set

```
> insectResults <- aov(count ~ spray, data = InsectSprays)
> summary(insectResults)
            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
```

With our null hypothesis as there is negligible difference between sprays and the insect count and the alternative suggesting there is. We see the probablility of F (<2e-16) is significantly low when compared to typical alpha levels such as 0.05.

6. Execute aov on InsectSprays data set

```
> insectBayesOut <- anovaBF(count ~ spray, data = InsectSprays)

|================================================================
====| 100%
> insectBayesOut
Bayes factor analysis
--------------
[1] spray : 1.506706e+14 ±0%

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

Produce posterior distributions and display the resulting HDI

```
> mcmcOut <- posterior(insectBayesOut, iterations = 10000)
0
%
|----|----|----|----|----|----|----|----|----|----|
**************************************************|
```

```
> summary(mcmcOut)

Iterations = 1:10000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

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

           Mean      SD Naive SE Time-series SE
mu        9.502 0.4778 0.004778       0.004778
spray-A   4.812 1.0411 0.010411       0.010351
spray-B   5.632 1.0554 0.010554       0.010554
spray-C  -7.155 1.0464 0.010464       0.010798
spray-D  -4.418 1.0344 0.010344       0.011176
spray-E  -5.778 1.0427 0.010427       0.010742
spray-F   6.907 1.0564 0.010564       0.010540
sig2     16.124 2.9335 0.029335       0.033941
g_spray   3.457 3.3683 0.033683       0.036124

2. Quantiles for each variable:

            2.5%     25%     50%     75%   97.5%
mu        8.5648   9.182   9.506   9.818  10.440
spray-A   2.7756   4.108   4.807   5.512   6.863
spray-B   3.5253   4.926   5.635   6.338   7.693
spray-C  -9.2622  -7.834  -7.150  -6.461  -5.111
spray-D  -6.4882  -5.117  -4.396  -3.726  -2.409
spray-E  -7.8105  -6.474  -5.785  -5.079  -3.742
spray-F   4.7822   6.216   6.911   7.611   8.970
sig2     11.4470  14.052  15.788  17.792  22.789
g_spray   0.8357   1.679   2.537   4.044  11.773
```

The null hypothesis is that the HDI for all spray variable will be identical with one another and the alternative hypothesis will have varying HDI values. As we look at the quantile data for each variable we see spray A, B, and F have HDI that overlap with one another whereas spray C, D, and E have their respective HDI that overlap with one another. It is evident in the following boxplot

```
> boxplot(as.matrix(mcmcOut[, 2:7]))
```
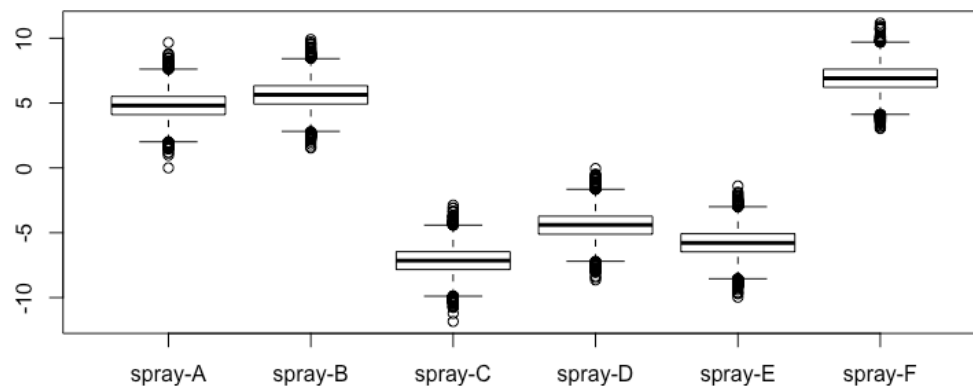
Lastly, the odds show a ~ 1.5:1 ratio in favor of the alternative hypothesis; however according to Kass and Raftery, this weak ratio is not worth mentioning.

7. Run Bayesian t-test comparing groups C and F

```
> BESTmcmc(InsectSprays[InsectSprays$spray == "C", 1],
InsectSprays[InsectSprays$spray == "F", 1])
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      1.978   0.6497  1.964  0.697   3.279 1.000 51381
mu2     16.502   2.1543 16.491 12.206 20.778 1.000 54288
nu      32.102  28.6371 23.534  1.023 89.298 1.001 18224
sigma1   2.062   0.5858  1.973  1.054   3.242 1.000 27029
sigma2   6.871   1.8230  6.563  3.862 10.542 1.001 31393

'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 results from the t-test shows that C and F are not identical due to the HDI lower and ranges not crossing one another.