Quinn Knudsen

Homework 6

Topic: **ANOVA & Experimental Groups**

1) The dependent variable (outcome) is the number of insects remaining. The independent variable (predicter) is the spray used. There are a total of 12 observations for each of the 6 spray conditions (72 in total). This is nice and timely because I just ran an analysis for James Madison University experimental biology lab and they decided to stop recording observations of rats that full recovered post surgery which was a nice little surprise when I found out the number of observations per treatment group was unequal (and inherently biased). ☹ Sigh

2) The Mean sq. is the sum of squares divided by the degrees of freedom (the variance). In this instance, the first line 533.8 is the between-groups variance and the second line 15.4 is the within groups variance.

3) The F-ratio is the ratio of the mean squares from the first line (between groups) to the mean squares of the second line (within groups). This would then look like the following:

> 533.8/15.4

[1] 34.66234

I would reject the null because the F-ratio value is clearly an extreme value that would fall in the tail of the F distribution. The associated p value helps establish this as well and we can confidently reject the null that the groups are all the same on average.

4) The degrees of freedom between groups is 6-1=5 and the degrees of freedom within groups is 72-6=66. The sum of these two values adds up to one less than the total number of observations due to an adjustment to consider the fact that, before we can calculate the standard deviation, we need to calculate the mean. We lose 1 degree of freedom for the mean.

5)

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

> 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

Ho: The type of spray does not impact the number of insects alive.

H1: Spray bottles will yield different number of insects remaining (a more informed person might have suggestions on which).

These results indicate we can reject the null hypothesis that the sprays yield equivalent impact of the number of insects leftover (34.7 p<2e-16).

More details can be found below although not asked for in the question specifically.

> TukeyHSD(insectResults, which = "spray")

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = count ~ spray, data = data)

$spray

diff lwr upr p adj

B-A 0.8333333 -3.866075 5.532742 0.9951810

C-A -12.4166667 -17.116075 -7.717258 0.0000000

D-A -9.5833333 -14.282742 -4.883925 0.0000014

E-A -11.0000000 -15.699409 -6.300591 0.0000000

F-A 2.1666667 -2.532742 6.866075 0.7542147

C-B -13.2500000 -17.949409 -8.550591 0.0000000

D-B -10.4166667 -15.116075 -5.717258 0.0000002

E-B -11.8333333 -16.532742 -7.133925 0.0000000

F-B 1.3333333 -3.366075 6.032742 0.9603075

D-C 2.8333333 -1.866075 7.532742 0.4920707

E-C 1.4166667 -3.282742 6.116075 0.9488669

F-C 14.5833333 9.883925 19.282742 0.0000000

E-D -1.4166667 -6.116075 3.282742 0.9488669

F-D 11.7500000 7.050591 16.449409 0.0000000

F-E 13.1666667 8.467258 17.866075 0.0000000

> summary(glht(insectResults, linfct = mcp(spray = "Tukey")))

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = count ~ spray, data = data)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

B - A == 0 0.8333 1.6011 0.520 0.995

C - A == 0 -12.4167 1.6011 -7.755 <1e-04 \*\*\*

D - A == 0 -9.5833 1.6011 -5.985 <1e-04 \*\*\*

E - A == 0 -11.0000 1.6011 -6.870 <1e-04 \*\*\*

F - A == 0 2.1667 1.6011 1.353 0.754

C - B == 0 -13.2500 1.6011 -8.276 <1e-04 \*\*\*

D - B == 0 -10.4167 1.6011 -6.506 <1e-04 \*\*\*

E - B == 0 -11.8333 1.6011 -7.391 <1e-04 \*\*\*

F - B == 0 1.3333 1.6011 0.833 0.960

D - C == 0 2.8333 1.6011 1.770 0.492

E - C == 0 1.4167 1.6011 0.885 0.949

F - C == 0 14.5833 1.6011 9.108 <1e-04 \*\*\*

E - D == 0 -1.4167 1.6011 -0.885 0.949

F - D == 0 11.7500 1.6011 7.339 <1e-04 \*\*\*

F - E == 0 13.1667 1.6011 8.223 <1e-04 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- single-step method)

6)

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

|============================================================================================================================| 100%

> insectBayesOut #very strong evidence from odds ratio

Bayes factor analysis

--------------

[1] spray : 1.506706e+14 ±0%

Against denominator:

Intercept only

---

Bayes factor type: BFlinearModel, JZS

## 3:1 not worth mentioning, 3:1-20:1 positive evidence for favored hypothesis

## 20:1 to 150:1 strong evidence

## 150:1 + very strong evidence for favored hypothesis

This falls into **the very strong evidence** category as the ratio exceeds 150:1.

> mcmcOut2 <-posterior(insectBayesOut,iterations=10000)

|----|----|----|----|----|----|----|----|----|----|

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*|

> boxplot(as.matrix(mcmcOut2[,2:7]))

> summary(mcmcOut2)

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 9.495 0.4739 0.004739 0.004739

spray-A 4.821 1.0275 0.010275 0.010275

spray-B 5.629 1.0384 0.010384 0.010512

spray-C -7.133 1.0498 0.010498 0.010995

spray-D -4.429 1.0411 0.010411 0.010572

spray-E -5.803 1.0422 0.010422 0.010945

spray-F 6.914 1.0432 0.010432 0.010985

sig2 16.092 2.9154 0.029154 0.034631

g\_spray 3.549 3.8524 0.038524 0.043280

2. Quantiles for each variable:

2.5% 25% 50% 75% 97.5%

mu 8.5605 9.178 9.492 9.811 10.436

spray-A 2.8116 4.135 4.813 5.514 6.860

spray-B 3.6192 4.935 5.627 6.329 7.680

spray-C -9.1660 -7.850 -7.132 -6.424 -5.091

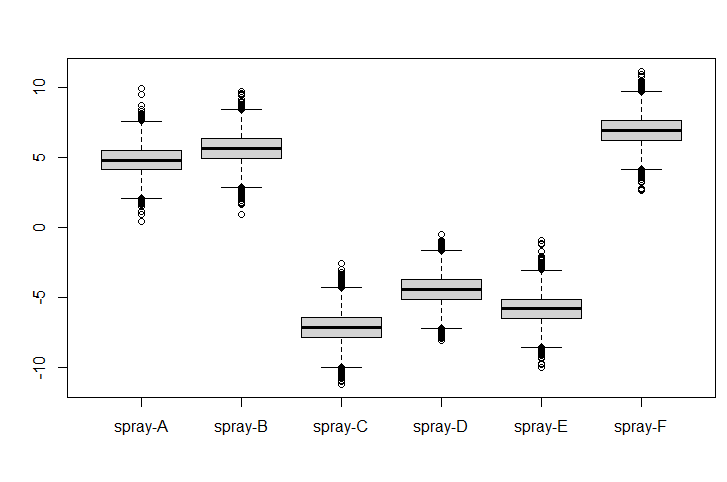
spray-D -6.4816 -5.121 -4.430 -3.735 -2.367

spray-E -7.8291 -6.497 -5.803 -5.108 -3.754

spray-F 4.8177 6.231 6.922 7.614 8.957

sig2 11.3574 14.040 15.738 17.788 22.852

g\_spray 0.8327 1.692 2.576 4.080 11.884



Ho: The type of spray does not impact the number of insects alive.

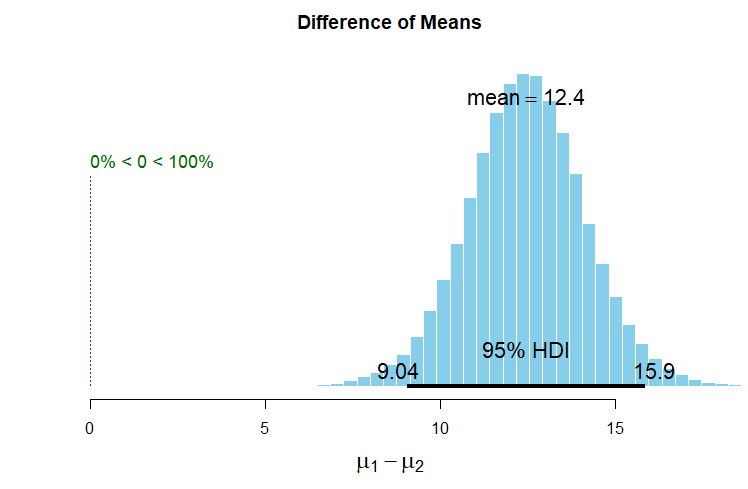
H1: Spray bottles will yield different number of insects remaining (a more informed person might have suggestions on which).

It is clear to see from the HDI summary and boxplot that spray A, B and F clearly kill less insects than spray C, D and F. Spray A,B and F are above the population mean, whereas spray C,D and E are below the populations mean (HDI does not overlap). If we were to compare A and C for example, a clear mean difference can be demonstrated (below).

> plot(BESTmcmc(data[data$spray=="A",1],

+ data[data$spray=="C",1]))

Waiting for parallel processing to complete...done.



7)

> t.test(data[data$spray=="F",1],

+ data[data$spray=="C",1])

Welch Two Sample t-test

data: data[data$spray == "F", 1] and data[data$spray == "C", 1]

t = 7.7484, df = 13.201, p-value = 2.876e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

10.52358 18.64308

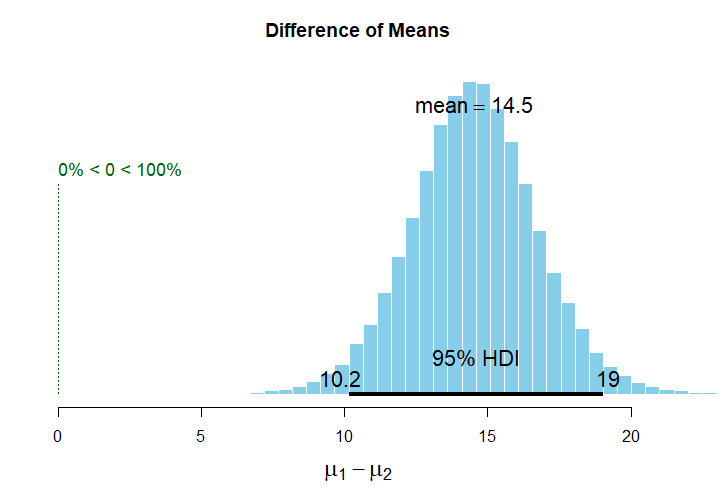
sample estimates:

mean of x mean of y

16.666667 2.083333

The t-test provides us with evidence to reject the null hypothesis that the mean of the spray C and spray F are equal based upon a p value below the set threshold of well below .05. Directionally, the confidence interval helps us make a determination that the mean of x is lower than y with a 95% range of 10.5 to 18.6. This supports the alternative hypothesis that spray F is greater than spray C.

The Bayesian method helps support our initial findings with the NHST by providing an estimated population difference of 14.5 and a 95% chance that the population mean difference between the two groups falls within the boundaries of the HDI (10.2 and 19). The HDI boundaries are clearly far above 0 as graphically illustrated below. The Bayesian method estimates 0% of the searches by the robot were conducted at a level below zero.



> summary(insectBEST)

mean median mode HDI% HDIlo HDIup compVal %>compVal

mu1 16.49 16.48 16.55 95 12.375 20.89

mu2 1.98 1.96 1.88 95 0.711 3.27

muDiff 14.51 14.50 14.34 95 10.220 19.07 0 100.0

sigma1 6.86 6.56 5.96 95 3.872 10.53

sigma2 2.06 1.97 1.81 95 1.050 3.23

sigmaDiff 4.80 4.52 4.14 95 1.477 8.61 0 99.9

nu 32.51 23.60 8.19 95 1.012 91.05

log10nu 1.34 1.37 1.50 95 0.542 2.07

effSz 3.00 2.98 3.00 95 1.473 4.53 0 100.0