Running & reporting paired t-tests in R

Nathan Brouwer brouwern@gmail.com @lobrowR

2017-10-26

This vignette will walk through how to carry out a paired t-test in R and report the results.

Create dataframe

We'll use some data from a paper by Faaborg et al on bird populations in Puerto Rico. We'll look at 9 species of warblers, and compare the number of birds captured in mist nests in 1991 and in 2005 to determine if on average birds are declining at this study site

```r
#make a vector of species names.
species <- c("OVEN","WEWA","NOWA",
             "BWWA","HOWA","AMRE",
             "CMWA","NOPA","PRWA")


#Number of birds of each species captured in 1991
N.1991 <- c(29, 6, 4, 60, 8, 19, 9, 7, 4)
N.2005 <- c(24, 5, 0, 16, 3, 9, 2, 5, 8)

#make dataframe
dat <- data.frame(species,
                  N.1991,
                  N.2005)
```

Take alook at the dataframe; we have 3 columns, one with the names of the 9 species, one with the number caught in 1991, and one with the nubmer caught in 2005

```r
head(dat)
```

```
##   species N.1991 N.2005
## 1    OVEN     29     24
## 2    WEWA      6      5
## 3    NOWA      4      0
## 4    BWWA     60     16
## 5    HOWA      8      3
## 6    AMRE     19      9
```

Paired t-test

Its a bit confusing, but there are multiple ways to do a paired t-test in R. (I can think about about 6, will focus on the 2 easiest ones). Paired t-tests are actually just a 1-sample t-test where the "1 sample" is a set of differnces between pairs of data points. Each one of our species has a pair of data points: abundance in 1991 and abundance in 2005. We can give R the raw data and t.test will calcaulte the differnce on the fly, or we can calculate the difference ourselvres. If we let R calcualte the difference, we **must** tell it that we are looking for a paired t-test by telling it "paired = TRUE". If we calcualte the difference ourselvees we **must** tell it we want a 1-sample t-test, which is done by giving it a mean value against which to test the null hypothesis ("mu = 0").

Paired t-test, Version 1

Paired t-test carried out by giving the t.test() function 2 columns from from a dataframe.

- Note there is no "~", just the name of each column, followed by a comma
- must include paired = TRUE

```r
t.test(dat$N.1991,    #column 1, then a comma
       dat$N.2005,    #column 2;
       paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  dat$N.1991 and dat$N.2005
## t = 1.7644, df = 8, p-value = 0.1157
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.523868 18.968313
## sample estimates:
## mean of the differences
##                8.222222
```

Paired t-test, Version 2

Paired t-test as a 1-sample t-test on the *difference* between two columns.

- First calcualte the difference between the columns
- T-test is given one column
- Note there is no "~", just the name of the column that has the differnces
- must set mu = 0
- there is NO "paired = TRUE"

```r
#make new column with the differencce between 1991 an 2005
dat$difference <- dat$N.1991 - dat$N.2005

#t.test() on difference
t.test(dat$difference,
       mu = 0)
```

```
##
##  One Sample t-test
##
## data:  dat$difference
## t = 1.7644, df = 8, p-value = 0.1157
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.523868 18.968313
## sample estimates:
## mean of x
##  8.222222
```

Reporting the results of a paired t-test

When we report a paired t-test we should give the p-value, the t statistic, and the degrees of freedom (df). NOte that for a paired t-test the df are equal to n-1, where n is the number of pairs in the data set (eg, the number of differences calculated), *not* the total number of seperate datapoinots.

We shoudl also report the effect size, which for a paried t-test is mean difference between the pairs; we should also report the 95% confidence interval for the effect size. Here, the mean difference is 8.2, which means on average there were 8 fewer individuals of each species captured in 2005 versus 1991. The 95 CI around this difference is large, from -2.5 to 19. Since it contains 0.0, the p value is greater than 0.05.

I would report the results of the t-test like this:

"There was a marginally significant difference in the number of birds of the 9 species captured in 1991 versus 2005 (paired t-test: t = 1.76, df = 8, p = 0.12). The mean difference in the number captured between years was 8.2 birds (95%CI: -2.5 to 19)."