

# Final

*Dave Anderson*

*April 28, 2019*

## 1

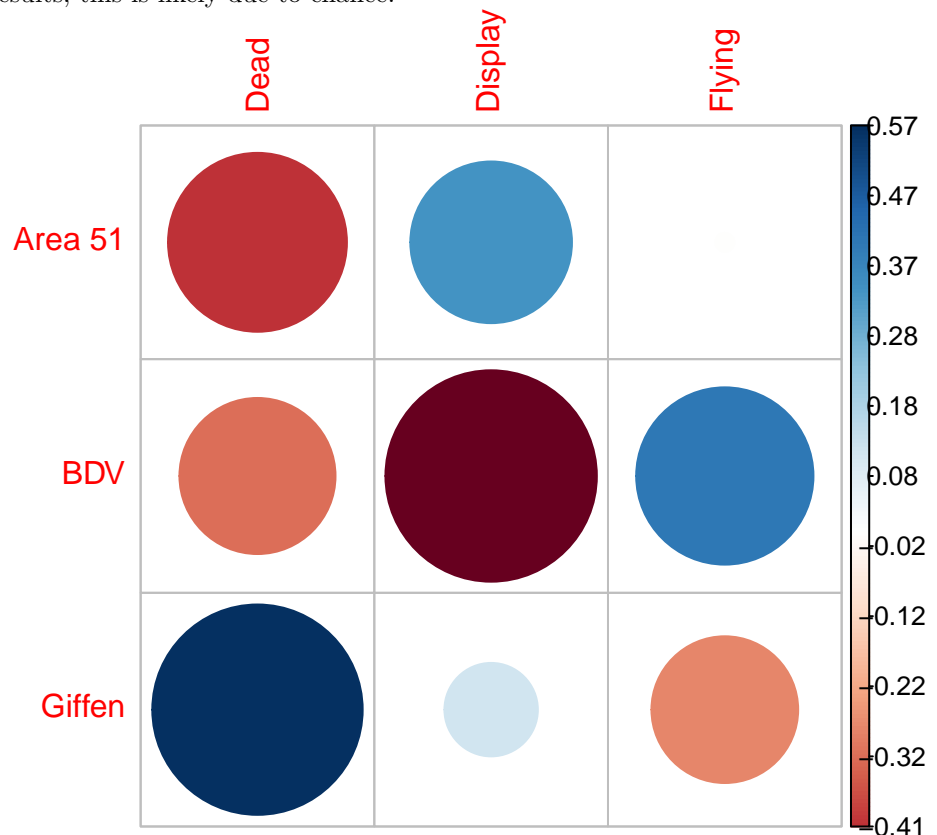
The following includes paired t-tests between the minority and white customers, as well as one to test between high income minority and white customers. The null hypothesis is  $H_0 : \mu_1 = \mu_2$  with the alternative being that minorities have a higher rejection rate than white customers,  $H_1 : \mu_1 > \mu_2$ . The tests both show evidence to reject the null hypothesis and support the claim that minorities have a higher rate of rejection at both class levels than whites. I also performed a permutation test for difference in means between the two groups. Using various numbers of resampling from the dataset, we still uptained extremely small p-values, suggesting that there is a difference in the means of the two groups. It is easy to look at the data and see that minorities have a generally higher rate of rejection from the banks. On the other hand, any statistician would be hesitant to draw conclusions from this data. There are only 20 banks, and we have no information about the number of customers in each group.

```
##
## Paired t-test
##
## data: acorn$MIN and acorn$WHITE
## t = 11.46, df = 19, p-value = 2.81e-10
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  18.04912      Inf
## sample estimates:
## mean of the differences
##          21.2565
##
## Paired t-test
##
## data: acorn$HIMIN and acorn$HIWHITE
## t = 8.9435, df = 19, p-value = 1.54e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.07999      Inf
## sample estimates:
## mean of the differences
##          16.215
##
## Paired t test power calculation
##
##          n = 20
##          d = 1.977454
## sig.level = 0.05
##          power = 1
## alternative = two.sided
##
## NOTE: n is number of *pairs*
```

```
## [1] 1e-04
```

## 2

At first glance, the schools appear to produce different quality ornithopters. I started by scaling the cells to the proportion of each schools totals. For example, 71% of Area 51's ornithopters are of "flying" quality. The proportions of quality for each school actually start to look similar. I decided to begin statistical testing with a chi-square test on the two-way table. The test produced a p-value of .86, so we don't have enough evidence to suggest that one school is larger in a certain category than another. In other words, the row and column variables are independent. If we plot the correlation of residuals, we can see that Giffen school tends to have more "dead" ornithopters compared to their expected. Area 51 has more display and BDV has more flying. Based on our chi-square results, this is likely due to chance.



I also ran a test on the proportions of flying ornithopters between the three schools. As you can see below, we again have a large p-value, indicating there isn't sufficient evidence to claim one school makes more flying ornithopters than the other schools.

```
##
## 3-sample test for equality of proportions without continuity
## correction
##
## data: c(89, 62, 119) out of c(124, 82, 170)
## X-squared = 0.86002, df = 2, p-value = 0.6505
## alternative hypothesis: two.sided
## sample estimates:
##   prop 1   prop 2   prop 3
## 0.7177419 0.7560976 0.7000000
```

### 3

```
##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  death
## X-squared = 133.41, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1286882 0.1869343
## sample estimates:
##   prop 1    prop 2
## 0.2656642 0.1078529

##
##      difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.4137712
##              n1 = 1197
##              n2 = 2012
##      sig.level = 0.05
##              power = 1
##      alternative = two.sided
##
## NOTE: different sample sizes
```

There are definitely more attacks in the United States, but we have no information about total beachgoers/surfers. I would assume there are more people active in the ocean in the United States than in Australia. Australia does have more coastline (US: 19,924 km, AU: 25,760 km), but the United States has a much larger population (US: 326,625,791, AU: 23,232,413). Interestingly, the estimated number of surfers for both countries is about 2.5 million. In terms of aggression and power of sharks, fatalities is most likely our best measure. Testing the proportions of fatalities in each country, we see that about 27% of Australian attacks are fatal, while only 11% of United States attacks are. The p-value of this proportion test shows there is evidence to suggest that Australian attacks are deadlier than American ones. With an alpha of .05, a power test suggests 100% power, indicating the probability of the two countries having different proportions and us failing to reject it is near 0. I believe there is strong evidence to suggest that sharks in Australia are deadlier.

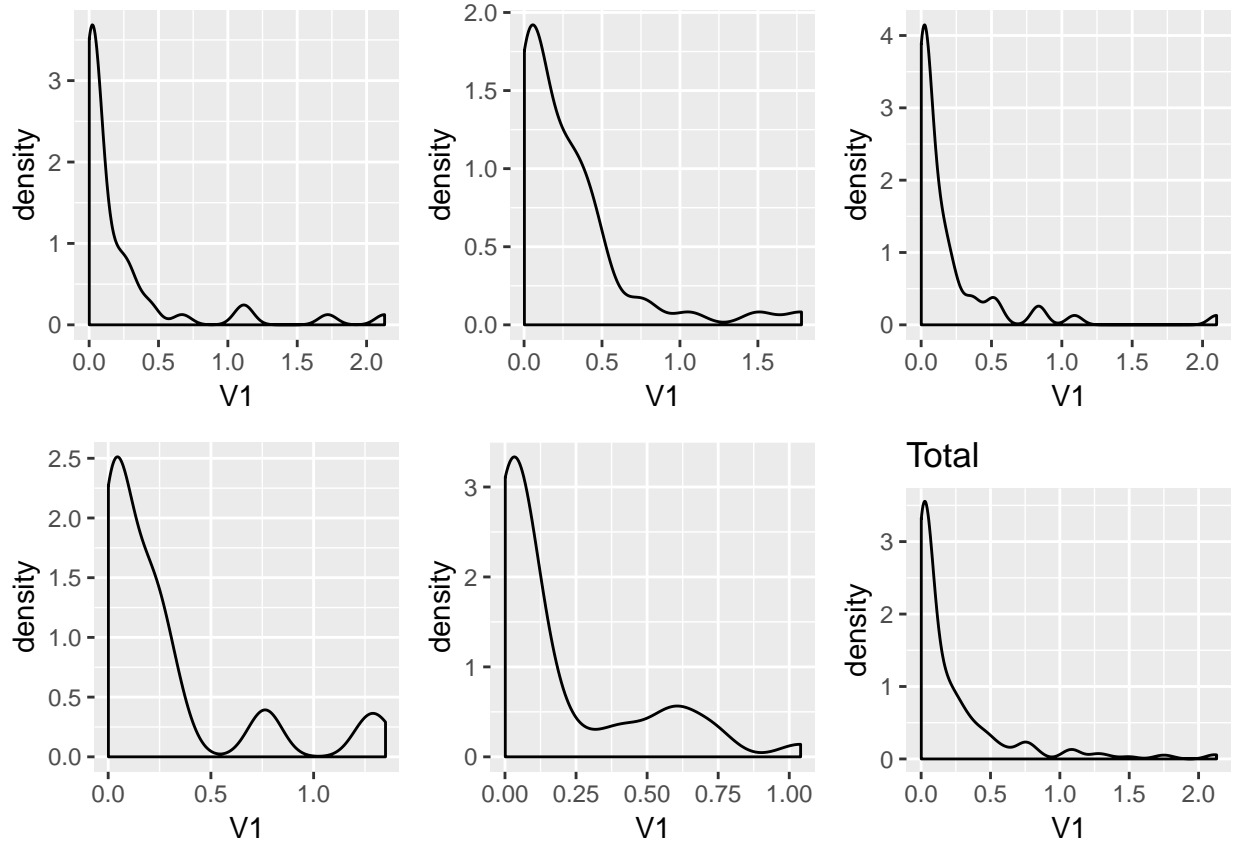
### 4

When testing the difference in proportions, it is tempting to subtract the two and use the difference as an “effect size”. As stated in the problem, this causes issues as two equal differences can have different levels of power associated with them. The arcsin transformation is calculated by taking  $\phi = 2\arcsin\sqrt{P}$ . The ES index is then calculated by taking the difference of the two transformations. This works because it produces effect sizes that do not depend on whether the proportions are in the middle of their ranges (.5) or not.

### 5

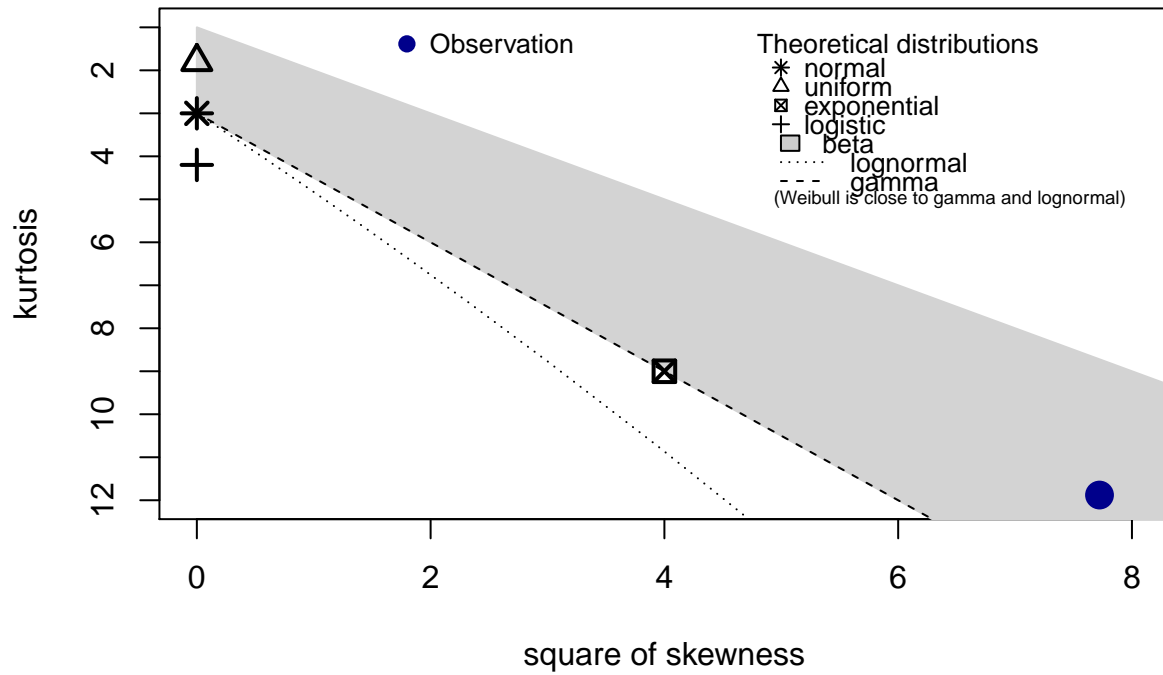
*a,b on paper* Comparing the five years, we see that 1962 had the most occurrences of rain. 1961 had the largest total rainfall of 13.197 inches, and the largest mean rainfall of 0.27 inches. We can see below that

each year has a heavy right tail, with most rainfalls being smaller. With the few number of observations, there are many bumps in the distributions. I have also displayed a distribution of all the years combined.



The distribution of the total rainfall is clearly non-normal. Based on our “descdist” plot of kurtosis and skewness, it doesn’t appear to be too close to the gamma distribution either. I believe it could be closer to an f or beta distribution.

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 0.001 max: 2.13
## median: 0.07
## mean: 0.2243921
## estimated sd: 0.3658212
## estimated skewness: 2.778925
## estimated kurtosis: 11.87935
```

6

*On Paper*