Contents

1 Introduction 1

2 Evidence Supporting Darwin’s Theory of Natural Selection 2
   2.1 Statistical summary and graphical display 2
   2.2 Inferential procedures (two-sample t-test) 3

3 Anatomical Abnormalities Associated with Schizophrenia 5
   3.1 Statistical summary and graphical display 5
   3.2 Inferential procedures (two-sample t-test) 6

1 Introduction

This document is intended to help describe how to undertake analyses introduced as examples in the Third Edition of the *Statistical Sleuth* (2013) by Fred Ramsey and Dan Schafer. More information about the book can be found at [http://www.proaxis.com/~panorama/home.htm](http://www.proaxis.com/~panorama/home.htm). This file as well as the associated knitr reproducible analysis source file can be found at [http://www.math.smith.edu/~nhorton/sleuth3](http://www.math.smith.edu/~nhorton/sleuth3).

This work leverages initiatives undertaken by Project MOSAIC ([http://www.mosaic-web.org](http://www.mosaic-web.org)), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the mosaic package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the mosaic package vignette ([http://cran.r-project.org/web/packages/mosaic/vignettes/MinimalR.pdf](http://cran.r-project.org/web/packages/mosaic/vignettes/MinimalR.pdf)).

To use a package within R, it must be installed (one time), and loaded (each session). The package can be installed using the following command:

> install.packages('mosaic') # note the quotation marks

Once this is installed, it can be loaded by running the command:
2 Evidence Supporting Darwin’s Theory of Natural Selection

Do birds evolve to adapt to their environments? That’s the question being addressed by Case Study 2.1 in the *Sleuth*.

2.1 Statistical summary and graphical display

We begin by reading the data and summarizing the variables.

```r
> summary(case0201)

          Year Depth
   Min.  :1976  Min.  : 6.2
   1st Qu.:1976  1st Qu.: 9.1
   Median :1977   Median : 9.9
   Mean  :1977    Mean  : 9.8
   3rd Qu.:1978  3rd Qu.:10.5
   Max.  :1978    Max.  :11.7

> fav = favstats(Depth ~ Year, data=case0201); fav

       Year min   Q1 median   Q3 max     mean      sd   n missing
1 1976 6.20  8.90    9.74 10.20  9.471  1.035  89       0
2 1978 7.10  9.60   10.30 10.70 10.140  0.906  89       0
```

A total of 178 subjects are included in the data: 89 are finches that were caught in 1976 and 89 are finches that were caught in 1978. The following figure replicates Display 2.1 on page 30.
The distributions are approximately normally distributed, with some evidence for a long left tail.

### 2.2 Inferential procedures (two-sample t-test)

First, we calculate the pooled SD and the standard error between these two different sample average (page 41, Display 2.8).

```r
> # Calculate Pooled SD
> n1 = fav["1976", "n"]; n1
[1] NA
> n2 = fav["1978", "n"]; n2
[1] NA
```
So the pooled SD is NA and the standard error is NA. 
Based on this information, we can construct a 95% confidence interval (page 43, Display 2.9).
So the 95% confidence interval of the difference between means is (NA, NA).
Now we want to calculate the $t$-statistic and $p$-value (as shown on page 46, Display 2.10).

```r
> tstats = (Yd-0)/SE; tstats  # The hypothesis difference=0
[1] NA
> onepval = 1-pt(tstats, df); onepval
[1] NA
> twopval = 2*onepval; twopval
[1] NA
```

The one-sided $p$-value is approximately NA and the two-sided $p$-value is also approximately NA.

We can get the results of “Summary of Statistical Findings” (page 29) by using the following code:

```r
> t.test(Depth ~ Year, var.equal=TRUE, data=case0201)
Two Sample t-test
data: Depth by Year
t = -5, df = 200, p-value = 9e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.956 -0.381
sample estimates:
mean in group 1976 mean in group 1978
 9.47 10.14
```

```r
> confint(lm(Depth ~ Year, data=case0201))
          2.5 %     97.5 %
(Intercept)  -935.61  -366.488
Year         0.19      0.478
```

### 3 Anatomical Abnormalities Associated with Schizophrenia

Is the area of brain related to the development of schizophrenia? That’s the question being addressed by case study 2.2 in the *Sleuth*.

#### 3.1 Statistical summary and graphical display

We begin by reading the data and summarizing the variables.
A total of 15 subjects are included in the data. There are 15 pairs of twins; one of the twins has schizophrenia, and the other does not. So there are 15 affected subjects and 15 unaffected subjects.

The difference in area of left hippocampus of these pairs of twins is:

```r
> case0202 = transform(case0202, DIFF = Unaffected - Affected)
> favstats(~ DIFF, data=case0202)

       min       Q1     median       Q3       max     mean      sd    n   missing
[1,] -0.190  0.0551  0.109750  0.3150  0.67700  0.1988  0.2381  15    0
```

This matches the results on page 31, Display 2.2.

```r
> densityplot(~ DIFF, data=case0202)
```

![Density plot of DIFF](image)

### 3.2 Inferential procedures (two-sample t-test)

We want to calculate the paired t-test and 95% confidence interval.

```r
> # Calculate t-statistics
> difmean = mean(~ DIFF, data=case0202); difmean
[1] 0.199
```
So the two-sided \( p \)-value is approximately 0.006 and the 95% confidence interval is (0.07, 0.33). We can also get the results displayed on page 32 by conducting a paired \( t \)-test:

```r
> with(case0202, t.test(Unaffected, Affected, paired=TRUE))
```

Paired t-test

data:  Unaffected and Affected
t = 3, df = 10, p-value = 0.006
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: 0.0667 0.3306
sample estimates:
| mean of the differences | 0.199 |