STAT 8010: HW3

March 10, 2020

Problem 1

This data set, *InvisibilityCloak.csv*, provides the number of mischievous acts committed by two groups of people, those with and those without and invisibility cloak. The variables in this data set are:

- **Participant**: Identification number of a participant.
- **Cloak**: Experimental group (0 = withouth a cloak of invisibility, 1 = with a cloack of invisibility).
- **Mischief**: the number of mischievous acts committed by a participant.

Suppose a researcher would like to examine if invisibility cloak affects the number mischievous acts committed.

Load the data

```
dat1 <- read.csv("./Data Sets/InvisibilityCloak.csv")
str(dat1)
```

```
## 'data.frame':    24 obs. of  3 variables:
##  $ Participant: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Cloak      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mischief   : int  3 1 5 4 6 4 6 2 0 5 ...
```

```
dat1$Cloak <- as.factor(dat1$Cloak)
```

a. State the null and alternative hypotheses

Let $\mu_1(\mu_2)$ be the average number of mischievous acts committed by a participant withouth (with) a cloak of invisibility

$H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$

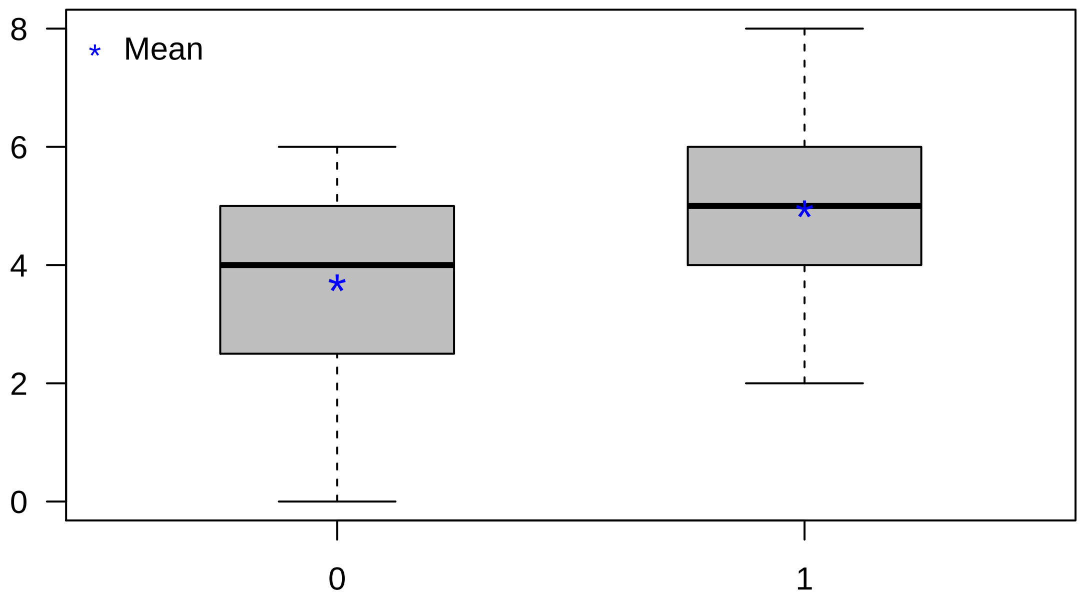
Data summary

```
library(dplyr)
summary1 <- dat1 %>%
select(Mischief, Cloak) %>%
group_by(Cloak) %>%
summarise(mean = mean(Mischief),
          sd = sd(Mischief),
          n = length(Mischief))
summary1
```

```
## # A tibble: 2 x 4
##   Cloak  mean    sd     n
##   <fct> <dbl> <dbl> <int>
```

```
## 1 0      3.75  1.91    12
## 2 1      5     1.65    12
```

```
boxplot(Mischief ~ Cloak, data = dat1, col = "gray",
        las = 1, boxwex = 0.5)
points(1, summary1$mean[1], pch = "*", cex = 1.5,
       col = "blue")
points(2, summary1$mean[2], pch = "*", cex = 1.5,
       col = "blue")
legend("toplef", legend = "Mean", pch = "*",
       col = "blue", bty = "n")
```



Test for $\sigma_1 = \sigma_2$

```
var.test(Mischief ~ Cloak, data = dat1)
```

```
##
##  F test to compare two variances
##
## data:  Mischief by Cloak
## F = 1.3417, num df = 11, denom df = 11, p-value = 0.6343
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3862357 4.6605462
## sample estimates:
## ratio of variances
##           1.341667
```

Pooled-t test

```
t.test(formula = Mischief ~ Cloak, data = dat1,
       var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  Mischief by Cloak
## t = -1.7135, df = 22, p-value = 0.1007
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.7629284  0.2629284
## sample estimates:
## mean in group 0 mean in group 1
##            3.75            5.00
```

Non pooled-t test

```
t.test(formula = Mischief ~ Cloak, data = dat1,
       var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  Mischief by Cloak
## t = -1.7135, df = 21.541, p-value = 0.101
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.764798  0.264798
## sample estimates:
## mean in group 0 mean in group 1
##            3.75            5.00
```

Problem 2

The data file, *Stereograms.csv*, records the time it took two groups of participants to see a figure hidden in a stereogram - one group received advance information about the scene, the other group did not. The variables in this data set are:

- **V1:**: Participant number.
- **fuseTime**: the time (in seconds) it took the participant to see the hidden figure.
- **condition**: experimental condition (NV = without information, VV = with information).
- **logFuseTime:**: the log transformation of the fuseTime.

Suppose a researcher would like to investigate whether providing advance information about the hidden figure shortens the time participant needs to see the figure.

Load the data

```
dat2 <- read.csv("./Data Sets/Stereograms.csv")
str(dat2)
```

```
## 'data.frame':    78 obs. of  4 variables:
##  $ V1         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ fuseTime   : num  47.2 22 20.4 19.7 17.4 14.7 13.4 13 12.3 12.2 ...
##  $ condition  : Factor w/ 2 levels "NV","VV": 1 1 1 1 1 1 1 1 1 1 ...
```

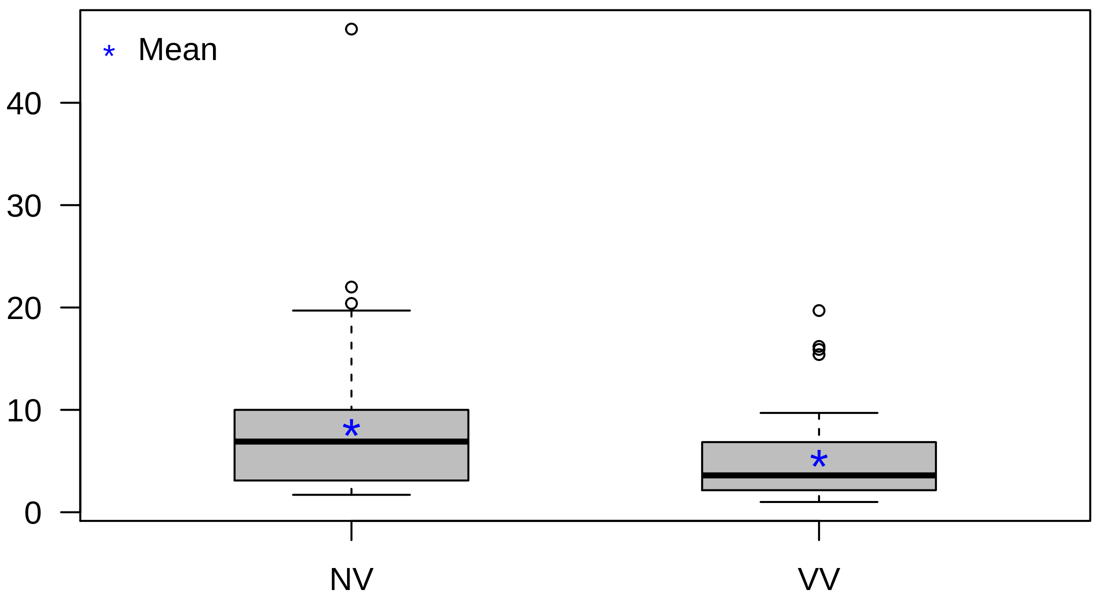
```
##  $ logFuseTime: num  3.85 3.09 3.02 2.98 2.86 ...
```

Data summary

```
summary2 <- dat2 %>%
select(fuseTime, logFuseTime, condition) %>%
group_by(condition) %>%
summarise(mean = mean(fuseTime),
          sd = sd(fuseTime),
          n = length(fuseTime),
          mean_log = mean(logFuseTime),
          sd_log = sd(logFuseTime))
summary2
```

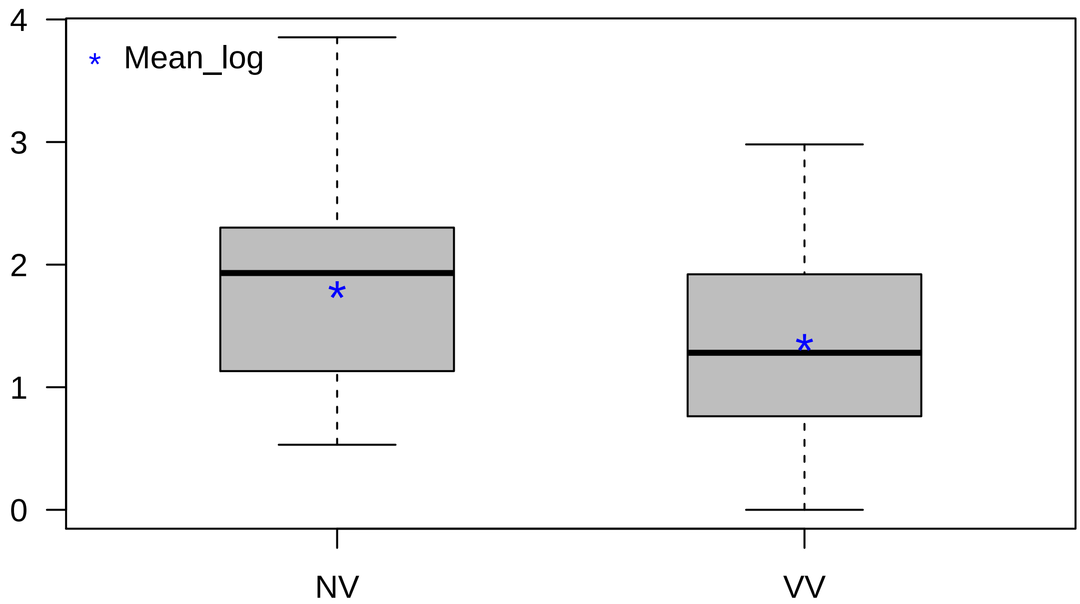
```
## # A tibble: 2 x 6
##   condition  mean    sd     n mean_log sd_log
##   <fct>     <dbl> <dbl> <int>    <dbl>  <dbl>
## 1 NV         8.56  8.09    43     1.82  0.814
## 2 VV         5.55  4.80    35     1.39  0.818
```

```
boxplot(fuseTime ~ condition, data = dat2,
        col = "gray", las = 1, boxwex = 0.5,
        outcex = 0.75)
points(1, summary2$mean[1], pch = "*", cex = 1.5,
       col = "blue")
points(2, summary2$mean[2], pch = "*", cex = 1.5,
       col = "blue")
legend("toplef", legend = "Mean", pch = "*",
       col = "blue", bty = "n")
```




```
boxplot(logFuseTime ~ condition, data = dat2,
        col = "gray", las = 1, boxwex = 0.5,
        outcex = 0.75)
points(1, summary2$mean_log[1], pch = "*",
       cex = 1.5, col = "blue")
```

```r
points(2, summary2$mean_log[2], pch = "*",
       cex = 1.5, col = "blue")
legend("toplef", legend = "Mean_log", pch = "*",
       col = "blue", bty = "n")
```



b. State the null and alternative hypotheses.

Let $\mu_1 (\mu_2)$ be the average time (log-transformed) for a participant without (with) information to see the hidden figure

$H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 > 0$

Test for $\sigma_1 = \sigma_2$

```r
var.test(logFuseTime ~ condition, data = dat2)
```

```
##
##  F test to compare two variances
##
## data:  logFuseTime by condition
## F = 0.99005, num df = 42, denom df = 34, p-value = 0.9665
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5114395 1.8741084
## sample estimates:
## ratio of variances
##          0.9900474
```

Pooled-t test

```r
t.test(formula = logFuseTime ~ condition,
       data = dat2, alternative = "greater",
       var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  logFuseTime by condition
## t = 2.319, df = 76, p-value = 0.01154
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.1213953       Inf
## sample estimates:
## mean in group NV mean in group VV
##         1.820010         1.389454
```

Non pooled-t test

```r
t.test(formula = logFuseTime ~ condition,
       data = dat2, alternative = "greater",
       var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  logFuseTime by condition
## t = 2.3178, df = 72.673, p-value = 0.01164
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.1210597       Inf
## sample estimates:
## mean in group NV mean in group VV
##         1.820010         1.389454
```

Problem 3

The file *WeightGain.csv* contains data from a study where weights of 16 participants before and after an eight-week period of 1000 excessive calorie intake were recorded. The variables in this data set are:

- **Weight Before**: Weight in pounds (lb) measured before eight weeks of excessive calorie intake
- **Weight After**: Weight in pounds (lb) measured after eight weeks of excessive calorie intake.
- **Difference**: Weight After - Weight Before

Suppose a researcher would like to investigate whether 1000 excess calorie intake per day over 8 weeks results in, on average, 16 pounds weight increase.

Load the data

```r
dat3 <- read.csv("./Data Sets/WeightGain.csv")
str(dat3)
```

```
## 'data.frame':    16 obs. of  3 variables:
##  $ Weight.Before: num  123 121 131 137 163 ...
##  $ Weight.After : num  136 129 145 146 174 ...
##  $ Difference   : num  13.2 8.58 14.08 8.58 10.56 ...
```

Data summary

```r
summary3 <- dat3 %>%
select(Weight.Before, Weight.After, Difference) %>%
summarise(mean_before = mean(Weight.Before),
          sd_before = sd(Weight.Before),
          n = length(Weight.Before),
          mean_after = mean(Weight.After),
          sd_after = sd(Weight.After),
          mean_diff = mean(Difference),
          sd_diff = sd(Difference))
summary3
```

```
##   mean_before sd_before  n mean_after sd_after mean_diff  sd_diff
## 1    144.6362  22.70488 16    155.045 21.43806  10.40875 3.840639
```

Paired t-test

```r
t.test(dat3$Weight.After, dat3$Weight.Before,
       paired = TRUE, mu = 16)
```

```
##
##  Paired t-test
##
## data:  dat3$Weight.After and dat3$Weight.Before
## t = -5.8233, df = 15, p-value = 3.355e-05
## alternative hypothesis: true difference in means is not equal to 16
## 95 percent confidence interval:
##   8.362218 12.455282
## sample estimates:
## mean of the differences
##                10.40875
```

One sample t-test for difference

```r
t.test(dat3$Difference, mu = 16)
```

```
##
##  One Sample t-test
##
## data:  dat3$Difference
## t = -5.8233, df = 15, p-value = 3.355e-05
## alternative hypothesis: true mean is not equal to 16
## 95 percent confidence interval:
##   8.362218 12.455282
## sample estimates:
```

```
## mean of x
##  10.40875
```