Write-up

1. Sampling

Power Tests

Given our power analysis, a group size of 12 was necessary for each combination of treatments.

```
pwr.anova.test(k = 4, f = 0.5, power = .8)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 4
##               n = 11.92611
##               f = 0.5
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

Random Number Generation

For this one, we're just generating a lot of pairs of numbers. Each pair will represent a city and household to get data from.

```
cities <- data.frame("City" = c("Vardo", "Hofn", "Helvig", "Bjurholm", "Blunduos",
                                "Helluland", "Harano","Reading", "Akkeshi",
                                "Nelson", "Birowa", "Shinobi", "Arcadia",
                                "Kiyobico", "Takazaki", "Nidoma", "Talu", "Pauma",
                                "Riroua", "Valai", "Gordes", "Maeva", "Kinsale",
                                "Colmar", "Vaiku", "Mahuti", "Eden"),
                     "Households" = c(1035, 1255, 836, 695, 797, 650, 753, 1182, 700,
                                      468, 805, 557, 1353, 926, 314, 698, 589, 745,
                                      831, 426, 433, 799, 231, 1631, 520, 1461, 838))

selections <- data.frame("City" = character(150), "Household" = numeric(150))

selections$City <- sample(cities$City, size = 150, replace = TRUE)
for(i in 1:150) {
  selections$Household[i] <- sample(seq_len(
    filter(cities, City == selections$City[i])$Households), 1)
}
```

2. Getting the Data ready

Data Import

Here, we just read in the data, focus on the procedural (exercise and stimulant) and blocking (age and gender) factors and the response variables, remove the NA's at the bottom of the list, and name the variables.

```
data_bp <- read_csv("Blood_Pressure_Data.csv")
```

```
## Parsed with column specification:
```

```
## cols(
##   ID = col_double(),
##   City = col_character(),
##   Household = col_double(),
##   Researcher = col_character(),
##   Names = col_character(),
##   Age = col_double(),
##   Gender = col_character(),
##   Exercise = col_character(),
##   Stimulant = col_character(),
##   `Blood Pressure (Start)` = col_double(),
##   `Blood Pressurer (End)` = col_double(),
##   `Blood Pressure (Difference)` = col_double()
## )
```

```
head(data_bp)
```

```
## # A tibble: 6 x 12
##      ID City  Household Researcher Names   Age Gender Exercise Stimulant
##   <dbl> <chr>     <dbl> <chr>      <chr> <dbl> <chr>  <chr>    <chr>
## 1     6 Biru~       252 Yoni Aboo~ Sann~    16 F      None     Olive Oil
## 2    55 Shin~        16 Jeremy Ph~ Emma~    20 F      5km out~ Olive Oil
## 3    57 Mahu~      1052 Ki Hyun P~ Ayak~    21 F      5km out~ Coca Lea~
## 4    45 Vardo       377 Ki Hyun P~ Dahl~    24 F      None     Coca Lea~
## 5    52 Haya~       105 Ki Hyun P~ Hall~    26 F      5km out~ Olive Oil
## 6    10 Nels~       102 Yoni Aboo~ Niam~    30 F      5km out~ Coca Lea~
## # ... with 3 more variables: `Blood Pressure (Start)` <dbl>, `Blood
## #   Pressurer (End)` <dbl>, `Blood Pressure (Difference)` <dbl>
```

```
data_bp <- data_bp[,6:12]
data_bp <- data_bp[-c(49,50,51),]
names(data_bp) <- c("B1", "B2", "A", "B", "BPS", "BPE", "BPD")

head(data_bp)
```

```
## # A tibble: 6 x 7
##      B1 B2    A                B              BPS   BPE   BPD
##   <dbl> <chr> <chr>            <chr>        <dbl> <dbl> <dbl>
## 1    16 F     None             Olive Oil      124   119    -5
## 2    20 F     5km outdoor run Olive Oil      124   122    -2
## 3    21 F     5km outdoor run Coca Leaves    126   124    -2
## 4    24 F     None             Coca Leaves    131   128    -3
## 5    26 F     5km outdoor run Olive Oil      128   126    -2
## 6    30 F     5km outdoor run Coca Leaves    130   128    -2
```

Data Coding

Once we switch from a numerical age and a character for gender to factors with (-1,0, and 1), we're going to be ready to do analysis. We'll also order the exercise and stimulant factors.

```
data_bp$B1 <- (rep(c(rep(-1,8), rep(0,8), rep(1,8)), 2))
data_bp$B2 <- (c(rep(-1, 24), rep(1,24)))

data_bp <- arrange(data_bp, B2, B1, A, B)

data_bp <- data_bp[,c(3,4,1,2,5,6,7)]
```

```r
head(data_bp)
```

```
## # A tibble: 6 x 7
##   A               B               B1    B2   BPS   BPE   BPD
##   <chr>           <chr>        <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 5km outdoor run Coca Leaves     -1    -1   126   124    -2
## 2 5km outdoor run Coca Leaves     -1    -1   130   128    -2
## 3 5km outdoor run Olive Oil       -1    -1   124   122    -2
## 4 5km outdoor run Olive Oil       -1    -1   128   126    -2
## 5 None            Coca Leaves     -1    -1   131   128    -3
## 6 None            Coca Leaves     -1    -1   130   131     1
```

3. Analyzing the Data

ANOVA

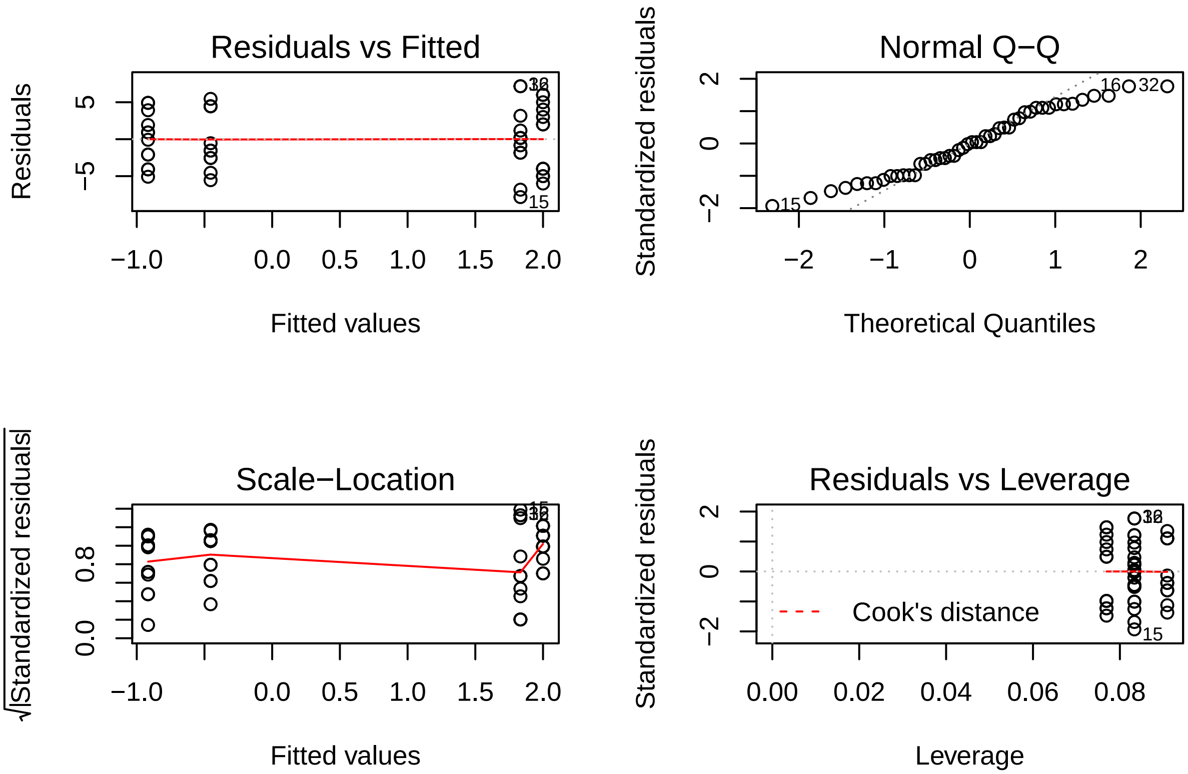
We're going to generate two different ANOVAs. The first `aov_bp` will just be the difference in blood pressure from start to finish as predicted by exercise(A), stimulant(B), and their interaction(A:B). The second will be the same but with the addition of blocking for age (B1) and gender (B2) as predictors.

```r
aov_bpd <- aov(BPD ~ A*B, data = data_bp)
aov_bpd_blocked <- aov(BPD ~ A*B + B1 + B2, data = data_bp)

summary(aov_bpd)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            1    2.1    2.08   0.116 0.7349
## B            1    0.3    0.29   0.016 0.8987
## A:B          1   81.0   80.98   4.514 0.0393 *
## Residuals   44  789.3   17.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow = c(2,2))
plot(aov_bpd)
```

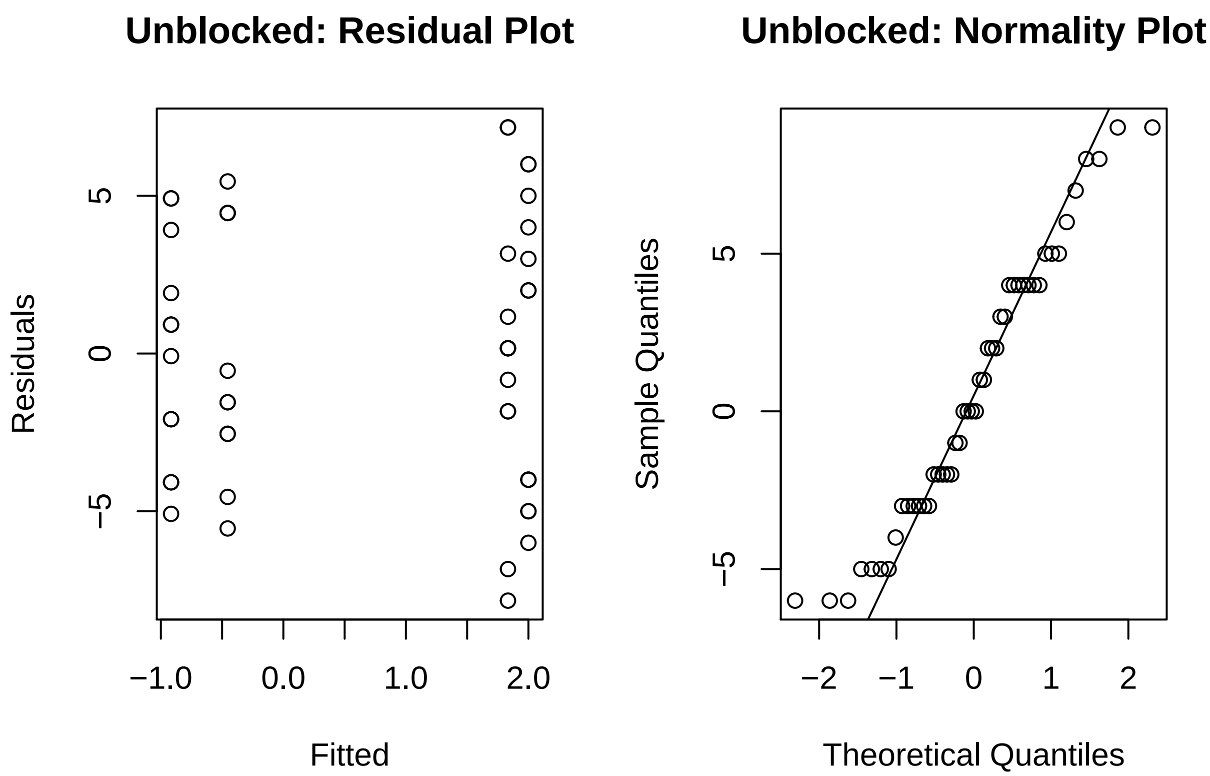



```
knitr::kable(data_bp %>% head(10))
```

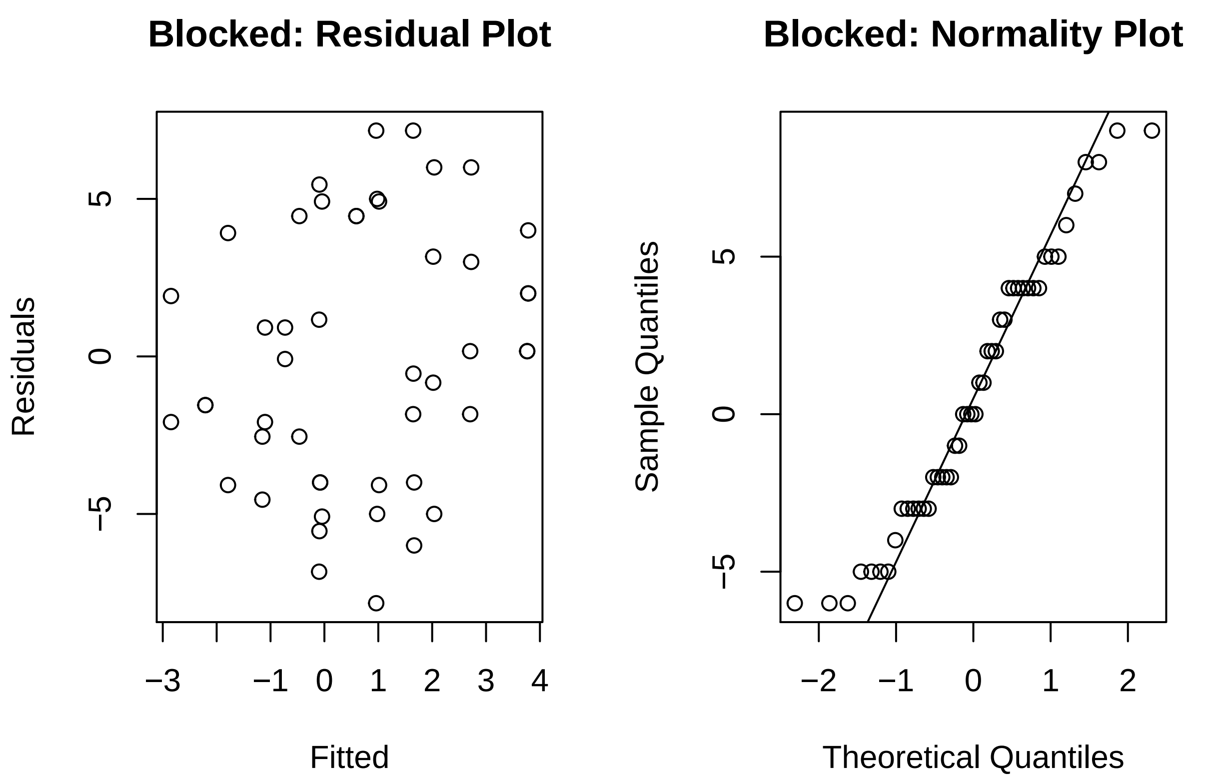
A	B	B1	B2	BPS	BPE	BPD
5km outdoor run	Coca Leaves	-1	-1	126	124	-2
5km outdoor run	Coca Leaves	-1	-1	130	128	-2
5km outdoor run	Olive Oil	-1	-1	124	122	-2
5km outdoor run	Olive Oil	-1	-1	128	126	-2
None	Coca Leaves	-1	-1	131	128	-3
None	Coca Leaves	-1	-1	130	131	1
None	Olive Oil	-1	-1	124	119	-5
None	Olive Oil	-1	-1	130	133	3
5km outdoor run	Coca Leaves	0	-1	133	140	7
5km outdoor run	Coca Leaves	0	-1	130	127	-3

Here are some graphs analyzing the equality of variance in our model, and the normality of our data.

```
par(mfrow = c(1,2))
plot(fitted(aov_bpd), resid(aov_bpd), main = "Unblocked: Residual Plot",
     xlab= "Fitted", ylab = "Residuals")
qqnorm(data_bp$BPD, main = "Unblocked: Normality Plot")
qqline(data_bp$BPD)
```



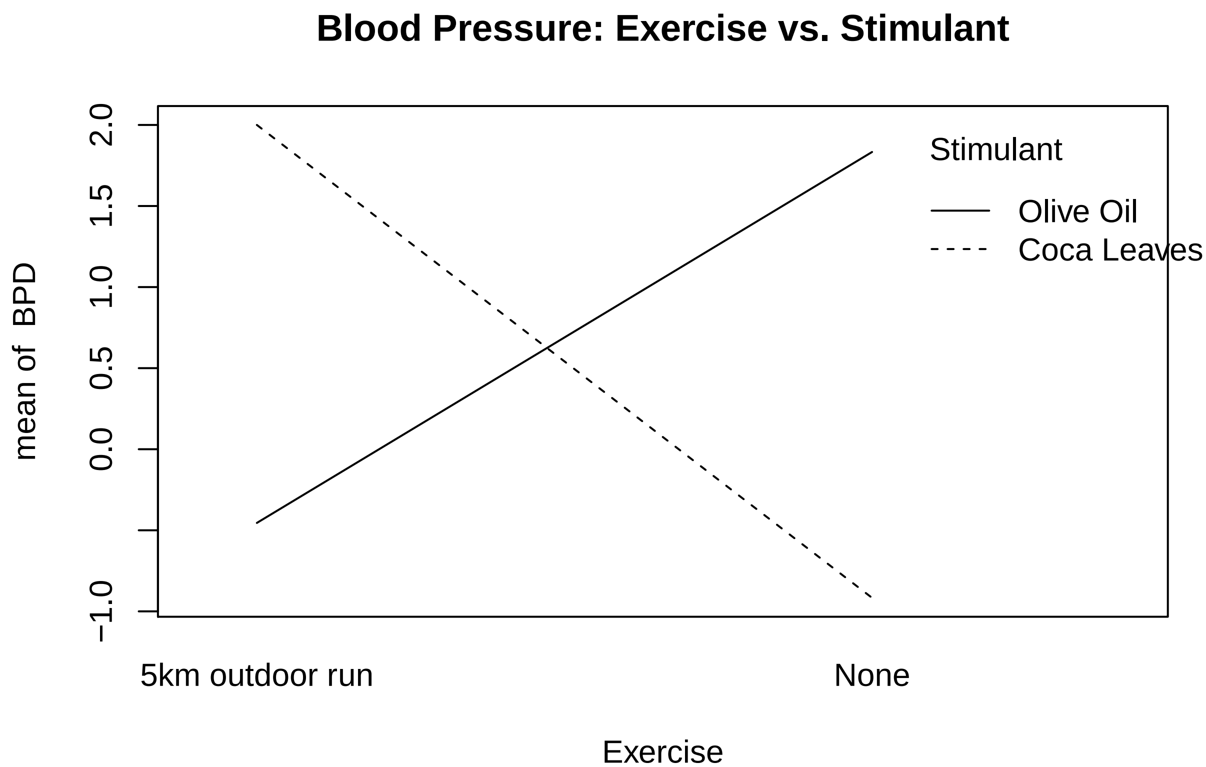

```r
par(mfrow = c(1,2))
plot(fitted(aov_bpd_blocked), resid(aov_bpd), main = "Blocked: Residual Plot",
     xlab= "Fitted", ylab = "Residuals")
qqnorm(data_bp$BPD, main = "Blocked: Normality Plot")
qqline(data_bp$BPD)
```



Interaction

This is just the plot of the interaction effects.

```r
with(data_bp, {interaction.plot(A, B, BPD, main = "Blood Pressure: Exercise vs. Stimulant",
                                xlab = "Exercise", trace.label = "Stimulant")})
```



Pairwise Comparisons

Here, we see the results of a Tukey test, which indicates that the difference of no particular pair of means rises to the level of significance.

```r
TukeyHSD(aov_bpd)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = BPD ~ A * B, data = data_bp)
##
## $A
##                            diff       lwr     upr     p adj
## None-5km outdoor run -0.4166667 -2.880783 2.04745 0.7348866
##
## $B
##                            diff       lwr     upr     p adj
## Olive Oil-Coca Leaves 0.1565217 -2.309737 2.62278 0.8988063
##
## $`A:B`
##                                                            diff       lwr
## None:Coca Leaves-5km outdoor run:Coca Leaves          -2.9166667 -7.443740
## 5km outdoor run:Olive Oil-5km outdoor run:Coca Leaves -2.4545455 -7.087387
```

```
## None:Olive Oil-5km outdoor run:Coca Leaves           -0.1666667 -4.693740
## 5km outdoor run:Olive Oil-None:Coca Leaves            0.4621212 -4.258365
## None:Olive Oil-None:Coca Leaves                       2.7500000 -1.866727
## None:Olive Oil-5km outdoor run:Olive Oil              2.2878788 -2.432608
##                                                           upr     p adj
## None:Coca Leaves-5km outdoor run:Coca Leaves         1.610407 0.3256711
## 5km outdoor run:Olive Oil-5km outdoor run:Coca Leaves 2.178296 0.4971963
## None:Olive Oil-5km outdoor run:Coca Leaves           4.360407 0.9996544
## 5km outdoor run:Olive Oil-None:Coca Leaves           5.182608 0.9936526
## None:Olive Oil-None:Coca Leaves                      7.366727 0.3944698
## None:Olive Oil-5km outdoor run:Olive Oil             7.008365 0.5715335
```