Statistics 101B Project – Caffeine & Attention

Group 4: Nicholas Cassol-Pawson, Max Chalekson, Romy Gou, Kanzah Jamil, Oliver Siu, Emma Vidal,

June 9, 2024

1. Sourcing

```
library(tidyverse)
library(DescTools)
library(car)
library(pwr)
island_caffeine <- read_csv("island_caffeine.csv")
```

2. Design of the Experiment

a) Design

```
latin <- function(n, random = FALSE){
  # generates a Latin Square of order n
  x <- matrix(LETTERS[1:n], n, n)
  for (j in 2:n) {
    x[, j] <- x[c(j:n, 1:(j - 1)), j]
  }
  if (random) {
    x <- x[sample(n), ]
    x <- x[, sample(n)]
  }
  x
}

latin(3)
```

```
##      [,1] [,2] [,3]
## [1,] "A"  "B"  "C"
## [2,] "B"  "C"  "A"
## [3,] "C"  "A"  "B"
```

```
set.seed(4)
latin(3, random = TRUE)
```

```
##      [,1] [,2] [,3]
## [1,] "B"  "A"  "C"
## [2,] "C"  "B"  "A"
## [3,] "A"  "C"  "B"
```

b) Sampling

```r
# island names
ironbark_names <- c("Hofn", "Vardo", "Helvig",
                    "Bjurholm", "Blonduos", "Helluland")
providence_names <- c("Hayarano", "Akkeshi", "Reading",
                      "Nelson", "Arcadia", "Kiyobico",
                      "Takazaki", "Shinobi", "Biruwa")
bonne_names <- c("Nidoma", "Colmar", "Riroua",
                 "Pauma", "Talu", "Valais",
                 "Kinsale", "Mahuti", "Vaiku",
                 "Eden", "Maeva", "Gordes")

# town house counts
ironbark_house <- c(937, 596, 483,
                    434, 431, 387)
providence_house <- c(521, 461, 714,
                      318, 1557, 520,
                      416, 358, 451)
bonne_house <- c(640, 2037, 462,
                 399, 483, 361,
                 429, 1017, 400,
                 523, 457, 403)

island_rng <- function(counts, town_names) {
  # generates and randomizes town and house number combinations
  town <- rep(town_names, counts)
  house <- numeric(0)
  for (i in counts) {
    house <- c(house, seq_len(i))
  }
  combo <- tibble(town, house)
  combo[sample(nrow(combo), nrow(combo)), ]
}

# generate rng objects
set.seed(4)
ironbark_rng <- island_rng(ironbark_house, ironbark_names)
providence_rng <- island_rng(providence_house, providence_names)
bonne_rng <- island_rng(bonne_house, bonne_names)

# index rng objects as necessary for viewing
ironbark_rng
```

```
## # A tibble: 3,268 x 2
##    town     house
##    <chr>    <dbl>
##  1 Vardo      591
##  2 Hofn       587
##  3 Blonduos   417
##  4 Helvig     262
##  5 Hofn        71
##  6 Hofn       684
```

```
##   7 Bjurholm   403
##   8 Bjurholm    22
##   9 Hofn       757
##  10 Hofn       698
## # i 3,258 more rows
```

```
providence_rng
```

```
## # A tibble: 5,316 x 2
##    town     house
##    <chr>    <dbl>
##  1 Reading    144
##  2 Reading    588
##  3 Arcadia   1109
##  4 Kiyobico   149
##  5 Takazaki   253
##  6 Shinobi    161
##  7 Arcadia   1307
##  8 Arcadia    221
##  9 Arcadia    658
## 10 Biruwa      51
## # i 5,306 more rows
```

```
bonne_rng
```

```
## # A tibble: 7,611 x 2
##    town   house
##    <chr>  <dbl>
##  1 Mahuti    68
##  2 Gordes   317
##  3 Riroua   418
##  4 Eden     399
##  5 Maeva    395
##  6 Riroua   416
##  7 Gordes   292
##  8 Colmar  1874
##  9 Colmar  1327
## 10 Vaiku    339
## # i 7,601 more rows
```

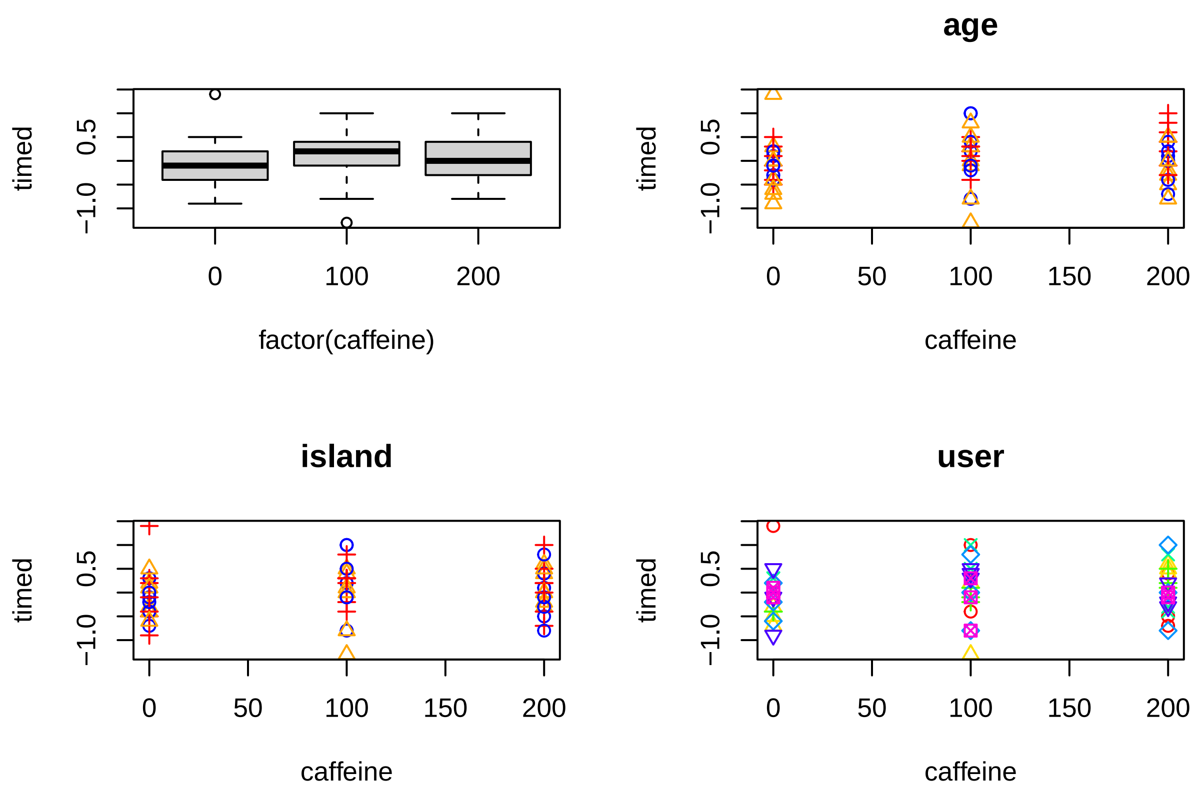
c) Experiment Data

```
# treatment, blocks, replicates
caffeine <- island_caffeine$caffeine
age <- island_caffeine$age_range
island <- island_caffeine$island
user <- island_caffeine$replicate

# response
timed <- island_caffeine$timed_diff
```

```r
# ANOVA
timed_aov <- aov(timed ~
                    factor(caffeine) +
                    factor(age) +
                    factor(island) +
                    factor(user))
```

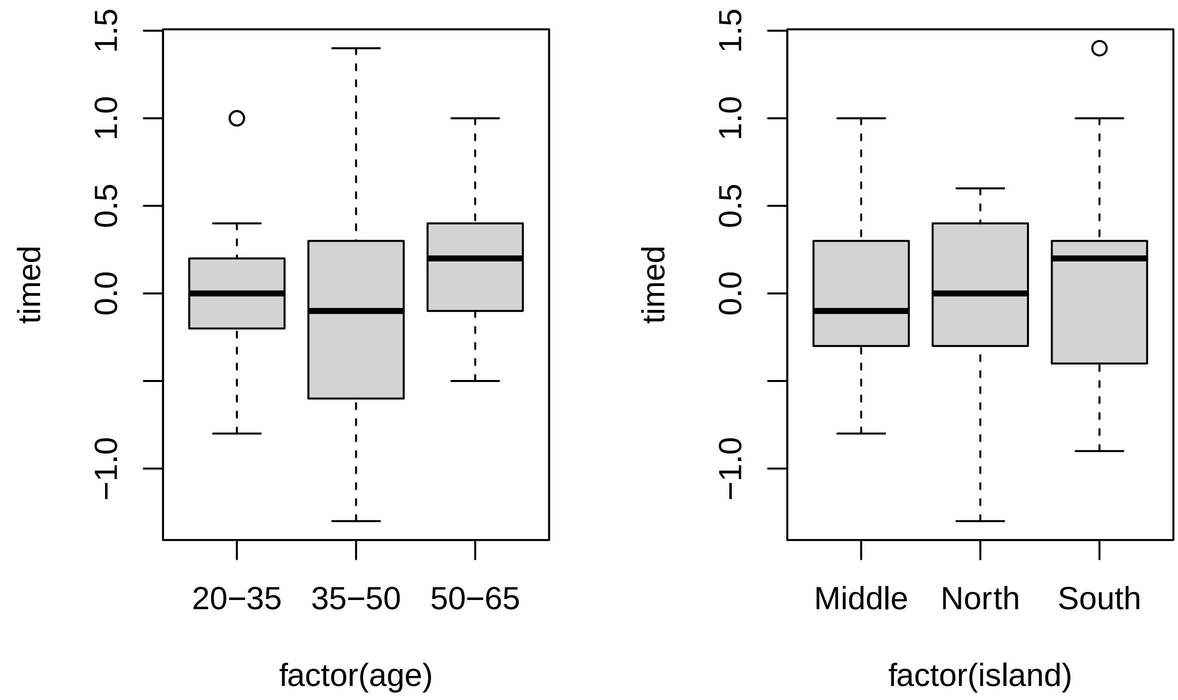
i) Loading Data Into Workspace

```r

age_index <- as.integer(factor(age))
island_index <- as.integer(factor(island))
user_index <- as.integer(factor(user))
caf_col <- c("blue", "orange", "red")

par(mfrow = c(2, 2))
plot(timed ~ factor(caffeine))
plot(timed ~ caffeine, main = "age",
     col = caf_col[age_index], pch = age_index)
plot(timed ~ caffeine, main = "island",
     col = caf_col[island_index], pch = island_index)
plot(timed ~ caffeine, main = "user",
     col = rainbow(7)[user_index], pch = user_index)
```

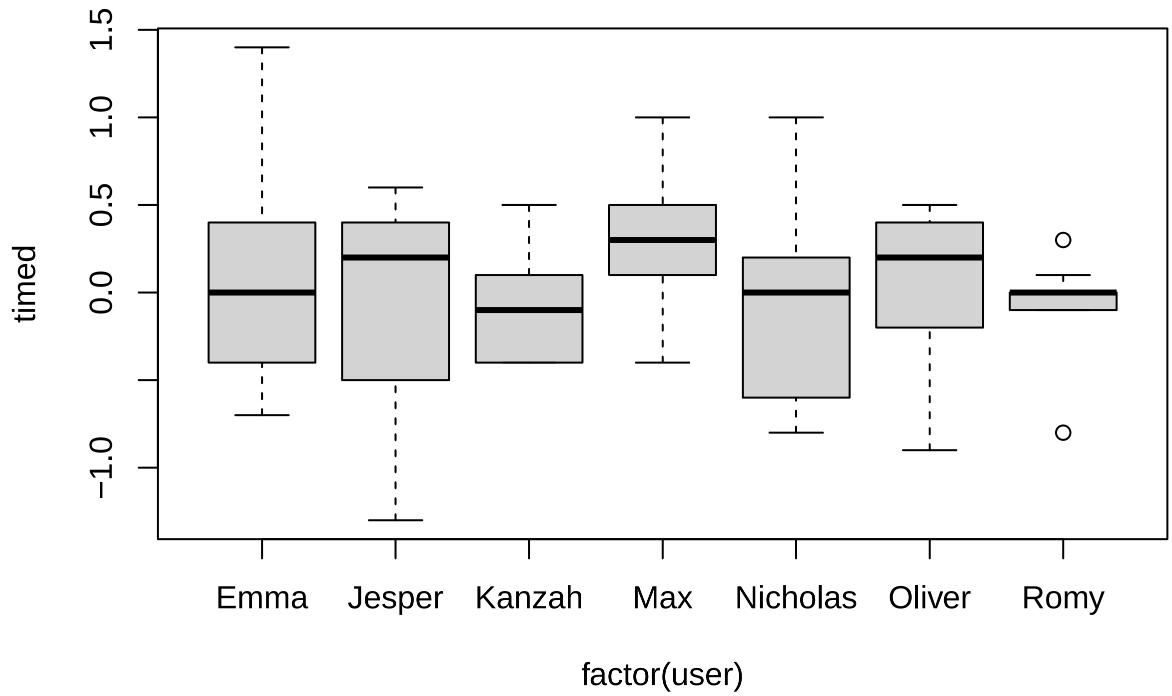


ii) Initial Plots

```r
par(mfrow = c(1, 2))
plot(timed ~ factor(age))
plot(timed ~ factor(island))
```




```r
par(mfrow = c(1, 1))
plot(timed ~ factor(user))
```



d) Sample Size and Power

```
mean0 <- mean(timed[caffeine == 0])
mean1 <- mean(timed[caffeine == 100])
mean2 <- mean(timed[caffeine == 200])

caffeine_range <- range(c(mean0, mean1, mean2))
caffeine_range
```

```
## [1] -0.07619048  0.09047619
```

```
d <- abs(diff(caffeine_range))
mse <- summary(timed_aov)[[1]][["Mean Sq"]][5]
caf_sd <- sqrt(mse)
f <- d / caf_sd

caffeine_min <- pwr.anova.test(k = 3, f = f, sig.level = 0.05, power = 0.8)
caffeine_min
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 3
```

```
##                n = 33.72766
##                f = 0.3133475
##        sig.level = 0.05
##            power = 0.8
##
## NOTE: n is number in each group
```

```
caffeine_pwr <- pwr.anova.test(k = 3, n = 21, f = f, sig.level = 0.05)
caffeine_pwr
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##                k = 3
##                n = 21
##                f = 0.3133475
##        sig.level = 0.05
##            power = 0.5751066
##
## NOTE: n is number in each group
```

```
caf_n <- caffeine_min$n
caf_pwr <- caffeine_pwr$power
```

i) Minimum and Maximum means The minimum and maximum means of the treatment group are $-0.0761905, 0.0904762$.

ii) Standard Deviation A preliminary estimate of σ^2 is obtained from MS_E which is $\hat{\sigma}^2 = 0.2829079$. Therefore, we have that $\sigma = 0.5318909$.

iii) Justifying Sample Size From `pwr.anova.test()`, $n = 33.72766$, so for 0.8 test power, $n = 34$ replicates are required which is a sample size of $N = 102$.

vi) Reporting Test Power Since there are $k = 3$ treatment groups and $n = 21$ observations per group, we have that the power of the test is 0.5751066.

3. Results and interpretation

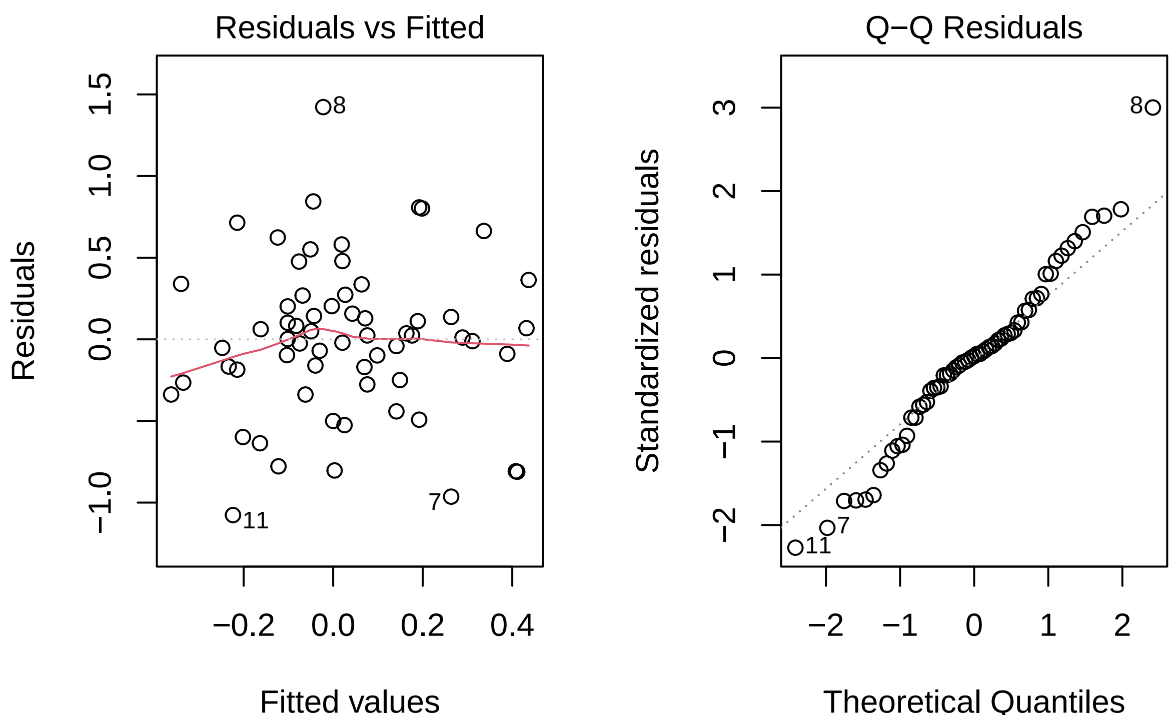
a) Summary

```
summary(timed_aov)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## factor(caffeine)  2  0.341  0.1706   0.603  0.551
## factor(age)       2  0.748  0.3740   1.322  0.276
## factor(island)    2  0.177  0.0883   0.312  0.733
## factor(user)      6  1.113  0.1855   0.656  0.685
## Residuals        50 14.145  0.2829
```

b) Diagnostic Plots

```r
par(mfrow = c(1, 2))
plot(timed_aov, which = c(1, 2))
```



c) Post-hoc Analysis

```r
TukeyHSD(timed_aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = timed ~ factor(caffeine) + factor(age) + factor(island) + factor(user))
##
## $`factor(caffeine)`
##                diff        lwr       upr     p adj
## 100-0    0.16666667 -0.2298128 0.5631461 0.5707420
## 200-0    0.14285714 -0.2536223 0.5393366 0.6613427
## 200-100 -0.02380952 -0.4202890 0.3726699 0.9884703
##
## $`factor(age)`
##                   diff        lwr       upr     p adj
## 35-50-20-35 -0.1428571 -0.5393366 0.2536223 0.6613427
## 50-65-20-35  0.1238095 -0.2726699 0.5202890 0.7324708
## 50-65-35-50  0.2666667 -0.1298128 0.6631461 0.2448666
##
## $`factor(island)`
```

```
##                     diff        lwr       upr     p adj
## North-Middle -0.02857143 -0.4250509 0.3679080 0.9834414
## South-Middle  0.09523810 -0.3012414 0.4917175 0.8312879
## South-North   0.12380952 -0.2726699 0.5202890 0.7324708
##
## $`factor(user)`
##                        diff        lwr       upr     p adj
## Jesper-Emma     -0.24444444 -1.0145704 0.5256815 0.9571271
## Kanzah-Emma     -0.21111111 -0.9812371 0.5590149 0.9791288
## Max-Emma         0.14444444 -0.6256815 0.9145704 0.9972302
## Nicholas-Emma   -0.18888889 -0.9590149 0.5812371 0.9882205
## Oliver-Emma     -0.10000000 -0.8701260 0.6701260 0.9996557
## Romy-Emma       -0.22222222 -0.9923482 0.5479038 0.9730211
## Kanzah-Jesper    0.03333333 -0.7367927 0.8034593 0.9999995
## Max-Jesper       0.38888889 -0.3812371 1.1590149 0.7130656
## Nicholas-Jesper  0.05555556 -0.7145704 0.8256815 0.9999891
## Oliver-Jesper    0.14444444 -0.6256815 0.9145704 0.9972302
## Romy-Jesper      0.02222222 -0.7479038 0.7923482 1.0000000
## Max-Kanzah       0.35555556 -0.4145704 1.1256815 0.7895800
## Nicholas-Kanzah  0.02222222 -0.7479038 0.7923482 1.0000000
## Oliver-Kanzah    0.11111111 -0.6590149 0.8812371 0.9993690
## Romy-Kanzah     -0.01111111 -0.7812371 0.7590149 1.0000000
## Nicholas-Max    -0.33333333 -1.1034593 0.4367927 0.8348711
## Oliver-Max      -0.24444444 -1.0145704 0.5256815 0.9571271
## Romy-Max        -0.36666667 -1.1367927 0.4034593 0.7650936
## Oliver-Nicholas  0.08888889 -0.6812371 0.8590149 0.9998261
## Romy-Nicholas   -0.03333333 -0.8034593 0.7367927 0.9999995
## Romy-Oliver     -0.12222222 -0.8923482 0.6479038 0.9989142
```