Attempting finality at individual level and group/level MHF COVID-19 Analyses

Zachary Levine

2020-09-18

Individual level variables of interest

1. Met minutes
2. Age

Group level variables of interest.

1. Volume of visits
2. Attendance
3. %Females

We'll be using the automatic ARIMA model fitting machinery from the *forecast* package to simplify this a bit.

```
library(lubridate)
#>
#> Attaching package: 'lubridate'
#> The following objects are masked from 'package:base':
#>
#>     date, intersect, setdiff, union
library(forecast)
#> Warning: package 'forecast' was built under R version 3.5.2
library(rstatix)
#>
#> Attaching package: 'rstatix'
#> The following object is masked from 'package:stats':
#>
#>     filter
```

First let's generate the three data frames of interest; mets, prescheduled appointments (volume), and noshows. Split each into only appointments in 2019 and only appointments in 2020.

```
##Change the year to select the year
splitbyyear <- function(df, year = "2019"){
  df[lubridate::year(df$Date) == year,]
}
mets2019 <- splitbyyear(RMHF::read_individual_data(fn = "mets.csv"), "2019")
mets2020 <- splitbyyear(RMHF::read_individual_data(fn = "mets.csv"), "2020")
##Order by date.
mets2019 <- mets2019[order(mets2019$Date),]
mets2020 <- mets2020[order(mets2020$Date),]

volume2019 <- splitbyyear(RMHF::read_individual_data(fn = "volume.csv"), "2019")
volume2020 <- splitbyyear(RMHF::read_individual_data(fn = "volume.csv"), "2020")
##Order by date.
```

```r
volume2019 <- volume2019[order(volume2019$Date),]
volume2020 <- volume2020[order(volume2020$Date),]
```

Now, let's split off into seperate analyses for each variable. For each, we ask two questions: 1. Does 2019 differ from 2019? 1. If that difference exists, is it before and after the implementation of COVID-19 social distancing in Ontario?

Individual level data - Met minutes, sex and age.

First, aggregate data on the daily level.

```r
##Should do the heavy lifting for us.
aggregate_date <- function(date, df, colselect = "Mets"){
  dfdate <- df[df$Date == date, colselect]
  return(mean(dfdate))
}
##Vectorized, gives data at the daily level.
meanmets2020 <- sapply(unique(mets2020$Date), aggregate_date, df = mets2020)
meanmets2019 <- sapply(unique(mets2019$Date), aggregate_date, df = mets2019)
meanage2020 <- sapply(unique(volume2020$Date), aggregate_date, df = volume2020, colselect = "Age")
meanage2019 <- sapply(unique(volume2019$Date), aggregate_date, df = volume2019, colselect = "Age")
```

We'd expect some autoregressive (ARIMA) structure in the data, so let's check for that.

```r
library(forecast)
auto.arima(meanmets2019)
#> Series: meanmets2019
#> ARIMA(0,0,1) with non-zero mean
#>
#> Coefficients:
#>          ma1      mean
#>       0.3202  616.9664
#> s.e.  0.1531   44.4507
#>
#> sigma^2 estimated as 51693:  log likelihood=-293.39
#> AIC=592.77   AICc=593.39   BIC=598.05
auto.arima(meanmets2020)
#> Series: meanmets2020
#> ARIMA(0,0,0) with non-zero mean
#>
#> Coefficients:
#>           mean
#>       645.8882
#> s.e.   33.6959
#>
#> sigma^2 estimated as 37505:  log likelihood=-213.41
#> AIC=430.83   AICc=431.24   BIC=433.76
auto.arima(meanage2019)
#> Series: meanage2019
#> ARIMA(0,0,0) with non-zero mean
#>
#> Coefficients:
#>          mean
#>       62.1731
#> s.e.   0.4840
```

```
#>
#> sigma^2 estimated as 11.25:  log likelihood=-123.06
#> AIC=250.13   AICc=250.4   BIC=253.83
auto.arima(meanage2020)
#> Series: meanage2020
#> ARIMA(0,0,0) with non-zero mean
#>
#> Coefficients:
#>          mean
#>       64.2391
#> s.e.   0.5903
#>
#> sigma^2 estimated as 19.17:  log likelihood=-155.86
#> AIC=315.73   AICc=315.96   BIC=319.71
```

ARIMA models mostly come up up as white noise processes. So we can't predict anything. Defaulting to anovas and t - tests.

Mets

```
##Normal stuff, t - test between data in different years.
library(rstatix)
t.test(meanmets2020, meanmets2019[1:length(meanmets2020)], paired = TRUE)
#>
#>  Paired t-test
#>
#> data:  meanmets2020 and meanmets2019[1:length(meanmets2020)]
#> t = 0.49691, df = 31, p-value = 0.6228
#> alternative hypothesis: true difference in means is not equal to 0
#> 95 percent confidence interval:
#>  -81.61794 134.20019
#> sample estimates:
#> mean of the differences
#>                26.29113
testdf <- data.frame("covidyear" = meanmets2020,
                     "covid" = c(rep(0,15), rep(1,17)))
anova_test(data = testdf,
           formula = covidyear ~ covid)
#> Coefficient covariances computed by hccm()
#> ANOVA Table (type II tests)
#>
#>   Effect DFn DFd     F     p p<.05   ges
#> 1  covid   1  30 1.443 0.239       0.046
```

Age

```
##Normal stuff, t - test between data in different years.
t.test(meanage2020, meanage2019[1:length(meanage2020)], paired = TRUE)
#>
#>  Paired t-test
#>
#> data:  meanage2020 and meanage2019[1:length(meanage2020)]
#> t = 2.1726, df = 46, p-value = 0.035
```

```
#> alternative hypothesis: true difference in means is not equal to 0
#> 95 percent confidence interval:
#>  0.1347006 3.5302182
#> sample estimates:
#> mean of the differences
#>                1.832459

##Help find the cut point for the intervention.
age2020map <- data.frame("Age" = meanage2020, "Date" = unique(volume2020$Date), stringsAsFactors = FALSE
age2020map <- age2020map[order(age2020map$Date),]
testdf <- data.frame("covidyear" = meanage2020,
                     ##Hardcoded groups, don't set.
                     "covid" = c(rep(0,23), rep(1,31)))
anova_test(data = testdf,
           formula = covidyear ~ covid)
#> Coefficient covariances computed by hccm()
#> ANOVA Table (type II tests)
#>
#>   Effect DFn DFd     F     p p<.05   ges
#> 1  covid   1  52 0.075 0.786       0.001
```

So there were no individual level differences between years. Or due to COVID. On to group - level stuff.

Group level variables of interest.

1. Volume of visits
2. Attendance
3. %Females

Volume of visits.

```
df <- RMHF::read_group_data()
df2019 <- as.numeric(df[df$Year == "2019",]$"Prescheduled appointments")
df2020 <- as.numeric(df[df$Year == "2020",]$"Prescheduled appointments")

t.test(df2019, df2020, paired = TRUE)
#>
#>  Paired t-test
#>
#> data:  df2019 and df2020
#> t = 0.48518, df = 10, p-value = 0.638
#> alternative hypothesis: true difference in means is not equal to 0
#> 95 percent confidence interval:
#>  -7.511284 11.693102
#> sample estimates:
#> mean of the differences
#>                2.090909
testdf <- data.frame("covidyear" = df2020,
                     "covid" = c(rep(0,6), rep(1,5)))
anova_test(data = testdf,
           formula = covidyear ~ covid)
#> Coefficient covariances computed by hccm()
#> ANOVA Table (type II tests)
```

```
#>
#>   Effect DFn DFd     F     p p<.05   ges
#> 1  covid   1   9 0.062 0.808       0.007
```

Attendance

```
df <- RMHF::read_group_data()
df2019 <- as.numeric(df[df$Year == "2019",]$"% of patients who were no-shows")
df2020 <- as.numeric(df[df$Year == "2020",]$"% of patients who were no-shows")

t.test(df2019, df2020, paired = TRUE)
#>
#>  Paired t-test
#>
#> data:  df2019 and df2020
#> t = -1.1189, df = 10, p-value = 0.2893
#> alternative hypothesis: true difference in means is not equal to 0
#> 95 percent confidence interval:
#>  -11.440414   3.791323
#> sample estimates:
#> mean of the differences
#>               -3.824545
testdf <- data.frame("covidyear" = df2020,
                     "covid" = c(rep(0,6), rep(1,5)))
anova_test(data = testdf,
           formula = covidyear ~ covid)
#> Coefficient covariances computed by hccm()
#> ANOVA Table (type II tests)
#>
#>   Effect DFn DFd     F     p p<.05   ges
#> 1  covid   1   9 3.226 0.106       0.264
```

%Females

```
df <- RMHF::read_group_data()
df2019 <- as.numeric(df[df$Year == "2019",]$"% Females")
df2020 <- as.numeric(df[df$Year == "2020",]$"% Females")

t.test(df2019, df2020, paired = TRUE)
#>
#>  Paired t-test
#>
#> data:  df2019 and df2020
#> t = 1.3529, df = 10, p-value = 0.2059
#> alternative hypothesis: true difference in means is not equal to 0
#> 95 percent confidence interval:
#>  -2.385561  9.760106
#> sample estimates:
#> mean of the differences
#>                3.687273
testdf <- data.frame("covidyear" = df2020,
                     "covid" = c(rep(0,6), rep(1,5)))
anova_test(data = testdf,
```

```
           formula = covidyear ~ covid)
#> Coefficient covariances computed by hccm()
#> ANOVA Table (type II tests)
#>
#>   Effect DFn DFd     F     p p<.05   ges
#> 1  covid   1   9 1.926 0.199       0.176
```

Correlations between attendance and met-minutes.

```
mets2020 <- RMHF::read_group_data()[RMHF::read_group_data()$Year == "2020",]$"Met-minutes"
mets2019 <- RMHF::read_group_data()[RMHF::read_group_data()$Year == "2019",]$"Met-minutes"
att2020 <- RMHF::read_group_data()[RMHF::read_group_data()$Year == "2020",]$"% of patients who were no-s
att2019 <- RMHF::read_group_data()[RMHF::read_group_data()$Year == "2019",]$"% of patients who were no-s

#The ith element in meanmets is the mean of the mets collected on ith day in that year.
cor(as.numeric(mets2019), as.numeric(att2019))
#> [1] -0.1939976
cor(as.numeric(mets2020), as.numeric(att2020))
#> [1] 0.1043858
```