In Turkey, some parts of society always compare Turkey to Germany and think that we are better than Germany for a lot of issues. The same applies to COVID-19 crisis management; is that reflects to true?

We will use two variables for compared parameters; the number of daily new cases and daily new deaths.First, we will compare the mean of new cases of the two countries. The dataset is attached in the repository.

#load and tidying the dataset

library(readxl)

deu <- read\_excel("covid-data.xlsx",sheet = "deu")

deu$date <- as.Date(deu$date)

tur <- read\_excel("covid-data.xlsx",sheet = "tur")

tur$date <- as.Date(tur$date)

#building the function comparing means on grid table

grid\_comparing <- function(column="new\_cases"){

table <-data.frame(

deu=c(mean=mean(deu[[column]]),sd=sd(deu[[column]]),n=nrow(deu)),

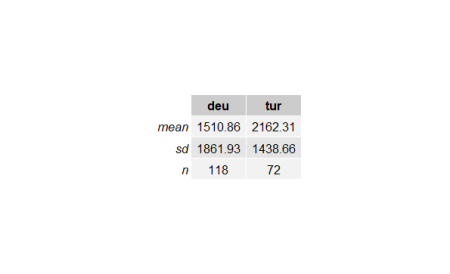
tur=c(mean=mean(tur[[column]]),sd=sd(tur[[column]]),n=nrow(tur))

) %>% round(2)

grid.table(table)

}

grid\_comparing()



Above table shows that the mean of new cases in Turkey is greater than Germany. To check it, we will inference concerning the difference between two means.

In order to make statistical inference for the , the sample distribution must be approximately normal distribution. If it is assumed that the related populations will not be normal, sample distribution is approximately normal only in the volume of relevant samples greater than 30 separately according to the **central limit theorem**. In this case, the distribution is assumed approximately normal.

If the variances of two populations and are known, **z-distribution** would be used for statistical inference. A more common situation, if the variances of population are unknown, we will instead use samples variances , and **distribution**.

When and are unknown, two situation are examined.

* : the assumption they are equal.
* : the assumption they are not equal.

There is a formal test to check whether population variances are equal or not which is a **hypothesis test for the ratio of two population variances**. A two-tailed hypothesis test is used for this as shown below.

The test statistic for :

The sample volumes n_1and n_2, **degrees of freedom of the samples** and . **F-distribution** is used to describe the sample distribution of

var.test(deu$new\_cases,tur$new\_cases)

# F test to compare two variances

#data: deu$new\_cases and tur$new\_cases

#F = 1.675, num df = 117, denom df = 71, p-value = 0.01933

#alternative hypothesis: true ratio of variances is not equal to 1

#95 percent confidence interval:

# 1.088810 2.521096

#sample estimates:

#ratio of variances

# 1.674964

At the %5 significance level, because **p-value(0.01933)** is less than 0.05, the null hypothesis(H_0) is rejected and we assume that variances of the populations are not equal.

Because the variances are not equal we use **Welch’s t-test** to calculate test statistic:

The degree of freedom:

Let’s see whether the mean of new cases per day of Turkey() greater than Germany(); to do that we will build the hypothesis test as shown below:

#default var.equal value is set to FALSE that indicates that the test is Welch's t-test

t.test(tur$new\_cases,deu$new\_cases,alternative = "g")

# Welch Two Sample t-test

#data: tur$new\_cases and deu$new\_cases

#t = 2.7021, df = 177.67, p-value = 0.00378

#alternative hypothesis: true difference in means is greater than 0

#95 percent confidence interval:

# 252.8078 Inf

#sample estimates:

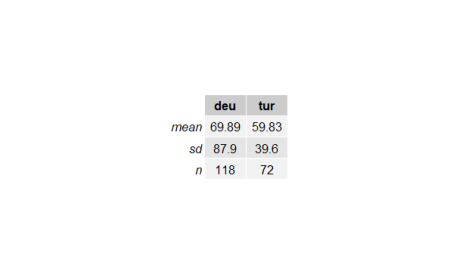
#mean of x mean of y

# 2162.306 1510.856

As shown above , at the %5 significance because the p-value(0.00378) is les than 0.05 the alternative hypothesis is accepted, which means in terms of controlling the spread of the disease, Turkey seems to be less successful than in Germany.

Another common thought in Turkish people that the health system in the country is much better than many European countries including Germany; let’s check that with daily death toll variable (new\_deaths).

grid\_comparing("new\_deaths")



It seems Turkey has much less mean of daily deaths than Germany. Let’s check it.

var.test(deu$new\_deaths,tur$new\_deaths)

# F test to compare two variances

#data: deu$new\_deaths and tur$new\_deaths

#F = 4.9262, num df = 117, denom df = 71, p-value = 1.586e-11

#alternative hypothesis: true ratio of variances is not equal to 1

#95 percent confidence interval:

# 3.202277 7.414748

#sample estimates:

#ratio of variances

# 4.926203

As described before, we will use Welch’s t-test because the variances are not equal as shown above(**p-value = 1.586e-11 < 0.05**).

t.test(deu$new\_deaths,tur$new\_deaths,alternative = "g")

# Welch Two Sample t-test

#data: deu$new\_deaths and tur$new\_deaths

#t = 1.0765, df = 175.74, p-value = 0.1416

#alternative hypothesis: true difference in means is greater than 0

#95 percent confidence interval:

# -5.390404 Inf

#sample estimates:

#mean of x mean of y

# 69.88983 59.83333

At %5 significance level, alternative hypothesis is rejected(**p-value = 0.1416 >0.05**). This indicates that the mean of daily deaths of Germany is not worst than Turkey’s.

**June 1** is set as the day of normalization by the Turkish government therefore many restrictions will be removed after that day. In order to check the decision, first, we will determine fit models for forecasting. To find the fit model we will build a function that compares [trend regression models](https://datageeek.wordpress.com/2020/03/11/trend-forecasting-models-and-seasonality-with-time-series/) in a plot.

models\_plot <- function(df=tur,column="new\_cases"){

df<- df[!df[[column]]==0,]#remove all 0 rows to calculate the models properly

#exponential trend model data frame

exp\_model <- lm(log(df[[column]])~index,data = df)

exp\_model\_df <- data.frame(index=df$index,column=exp(fitted(exp\_model)))

names(exp\_model\_df)[2] <- column

#comparing the trend plots

ggplot(df,mapping=aes(x=index,y=.data[[column]])) + geom\_point() +

stat\_smooth(method = 'lm', aes(colour = 'linear'), se = FALSE) +

stat\_smooth(method = 'lm', formula = y ~ poly(x,2), aes(colour = 'quadratic'), se= FALSE) +

stat\_smooth(method = 'lm', formula = y ~ poly(x,3), aes(colour = 'cubic'), se = FALSE)+

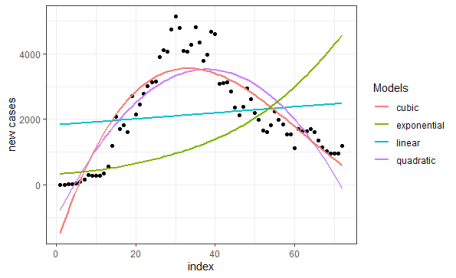
stat\_smooth(data=exp\_model\_df,method = 'loess',mapping=aes(x=index,y=.data[[column]],colour = 'exponential'), se = FALSE)+

labs(color="Models",y=str\_replace(column,"\_"," "))+

theme\_bw()

}

models\_plot()



As we can see from the plot above, the cubic and quadratic regression models seem to fit the data more. To be able to more precise we will create a function that compares **adjusted** .

#comparing model accuracy

trendModels\_accuracy <- function(df=tur,column="new\_cases"){

df<- df[!df[[column]]==0,]#remove all 0 rows to calculate the models properly

model\_quadratic <- lm(data = df,df[[column]]~poly(index,2))

model\_cubic <- lm(data = df,df[[column]]~poly(index,3))

#adjusted coefficients of determination

adj\_r\_squared\_quadratic <- summary(model\_quadratic) %>%

.$adj.r.squared

adj\_r\_squared\_cubic <- summary(model\_cubic) %>%

.$adj.r.squared

c(quadratic=round(adj\_r\_squared\_quadratic,2),cubic=round(adj\_r\_squared\_cubic,2))

}

trendModels\_accuracy()

#quadratic cubic

# 0.73 0.77

**The cubic trend regression model** is much better than the quadratic trend model for Turkeys spread of disease as shown above.

Now, let’s find should the normalization day(June 1) is true. In the following code chunk, we will try some index numbers to find zero new cases.

#forecasting zero point for new cases in Turkey

model\_cubic <- lm(formula = new\_cases ~ poly(index, 3), data = tur)

predict(model\_cubic,newdata=data.frame(index=c(77,78,79,80)))

# 1 2 3 4

#183.92149 111.23894 42.50292 -22.04057

As shown above, index 80 goes to negative, so it can be considered as the day of normalization. If we look at the dataset, we can see that day is June 1. So the government seems to be right about **the normalization calendar**.

You can do the same predictions for Germany using the functions we created before.