Online search trends and word-related emotional response during COVID-19 lockdown in Italy

Mahmoud M.Elsherif

Last Knitted: 2023-09-18

### Vignette Setup:

### Project/Data Title:

Online search trends and word-related emotional response during COVID-19 lockdown in Italy

Data provided by: Maria Montefinese

### Project/Data Description:

The strong and long lockdown adopted by the Italian government to limit the spread of the COVID-19 represents the first threat-related mass isolation in history that scientists can study in depth to understand the emotional response of individuals to a pandemic. Perception of a pandemic threat through invasive media communication, such as that related to COVID-19, can induce fear-related emotions (Van Bavel et al., 2020). The dimension theory of emotions (Osgood & Suci, 1955) assumes that emotive space is defined along three dimensions: valence (indicating the way an individual judges a stimulus; from unpleasant to pleasant), arousal (indicating the degree of activation an individual feels towards a stimulus; from calm to excited) and dominance (indicating the degree of control an individual feels over a given stimulus; from out of control to in control). Fear is characterized as a negatively valenced emotion, accompanied by a high level of arousal (Witte, 1992; Witte, 1998) and a low dominance (Stevenson, Mikel & James, 2007). This is generally in line with previous results showing that participants judged stimuli related to the most feared medical conditions as the most negative, the most anxiety-provoking, and the least controllable (Warriner, Kuperman & Brysbaert, 2013). Fear is also characterized by extreme levels of emotional avoidance of specific stimuli (Perin et al., 2015) and may be considered a unidirectional precursor to psychopathological responses within the current context (Ahorsu et al., 2020). dealing with fear in a pandemic situation could be easier for some people than others. Indeed, individual differences have been associated with behavioral responses to pandemic status (Carvalho Pianowski & Gonçalves, 2020).

To mitigate the effects of the COVID-19 on the mental health of individuals, it is imperative to evaluate their emotional response to this emergency. The internet searches are a direct tool to address this problem. In fact, COVID-19 has been reported to affect the content that people explore online (Effenberger et al., 2020), and online media and platforms offer essential channels where people express their feelings and emotions and seek health-related information (Kalichman et al., 2003; Reeves, 2001). In particular, Google Trends is an available data source of real-time internet search patterns, which has been shown to be a valid indicator of people’s desires and intentions (Payne, Brown-Iannuzzi & Hannay, 2017; Pelham et al., 2018). Therefore, the amount of searches related to COVID-19 on the internet revealed by Google Trends are an indicator of how people feel about concepts related to the COVID-19 pandemic. A change in online search trends reflects a change in participants’ interests and attitudes towards a specific topic. Based on the topic, the context (that is, the reasons for this change), and this mutated interest per se, it is possible to predict people’s behavior and affective response to the topic in question. In this study, our aim was to understand how emotional reaction and online search behavior have changed in response to the COVID-19 lockdown in the Italian population.

### Methods Description:

Data were collected in the period from 4 May to 17 May 2020, the last day of complete lockdown in Italy, from 71 native adult Italian speakers (56 females and 13 males; mean age (SD) = 26.2 (7.9) years; mean education (SD) = 15.3 (3.2) years). There were no other specific eligibility criteria. An online survey was conducted using Google Forms to collect affective ratings during the lockdown caused by the COVID-19 epidemic in Italy. In particular, we asked participants to complete the Positive and Negative Affect Schedule (PANAS, Terraciano, McCrae & Costa, 2003) and Fear of COVID-19 Scale (FCV-19S, Ahorsu et al., 2020) and judged valence, arousal, and dominance (on a 9-point self-assessment manikin, Montefinese et al., 2014) of words related or unrelated to COVID-19, as identified by Google search trends. The word stimuli consisted of 3 groups of 20 words each. The first group (REL+) consisted of the words showing the largest positive relation between their search trends and the search trend for COVID-related terms. On the contrary, the second group (REL-) consisted of the words showing the largest negative relation between their search trends and the search trend for COVID-related terms. In other words, the COVID-19 epidemic in Italy and the consequent increase in interest in terms related to COVID was related to a similar increase in interest for the REL+ words and a decrease in interest for the REL- words. The third group (UNREL) consisted of the words for which the search trend was unrelated to the search trend for the COVID-related terms.

### Data Location:

<https://osf.io/we9r4/>

DF <- import("montefinese\_data.csv")   
  
names(DF) <- make.names(names(DF),unique = TRUE)  
  
names(DF)[names(DF) == 'ITEM..ITA.'] <- "item"  
  
DF <- DF %>%  
 filter(StimType != "") %>%   
 filter(Measure == "Valence") %>% # only look at valence score   
 arrange(item) %>% #orders the rows of the data by the target\_name column  
 group\_by(item) %>% #group by the target name  
 transform(items = as.numeric(factor(item)))%>% #transform target name into a item  
 select(items, item, everything()  
 ) #select all variables from items and target\_name   
  
head(DF)

## items item ssID Gender Age Education Measure StimType Response  
## 1 1 affogare 1 F 36 21 Valence UNREL 2  
## 2 1 affogare 2 M 40 21 Valence UNREL 2  
## 3 1 affogare 3 F 29 21 Valence UNREL 1  
## 4 1 affogare 4 M 39 13 Valence UNREL 1  
## 5 1 affogare 5 F 27 16 Valence UNREL 1  
## 6 1 affogare 6 M 33 18 Valence UNREL 1

### Date Published:

2021-08-10

### Dataset Citation:

Montefinese M, Ambrosini E, Angrilli A. 2021. Online search trends and word-related emotional response during COVID-19 lockdown in Italy: a cross-sectional online study. PeerJ 9:e11858 <https://doi.org/10.7717/peerj.11858>

### Keywords:

Covid-19; Emotional response; Online search; Lockdown; Coping

### Use License:

CC-By Attribution 4.0 International

### Geographic Description - City/State/Country of Participants:

Italy

### Column Metadata:

metadata <- import("montefinese\_metadata.xlsx")  
  
flextable(metadata) %>% autofit()

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| ssID | Participant code | Numeric |
| Gender | Participants’ gender | Character |
| Age | Participants’ age | Numeric |
| Education | Participants’ years of education | Numeric |
| Measure | Questionnaires and ratings (PANAS, COVID-19 fear, valence, arousal, dominance) | Character |
| ITEM (ITA) | Test items and word stimuli | Character |
| Stim Type | Word condition (REL+, REL-, UNREL) | Character |
| Response | Participants’ scores to the questionnaires and ratings | Numeric |

### AIPE Analysis:

In this dataset, there are REL+ and REL- variables. In the REL+ condition, the words show the largest positive relation between their search trends and the search trend for the COVID-related terms. In the REL- condition, the words showed the largest negative relation between their search trends and the search trends for the COVID-related terms. The third group (UNREL) consisted in the words for which the search trend was unrelated to the search trend for the COVID-related terms.

#### Stopping Rule

What the usual standard error for the data that could be considered for our stopping rule using the 40% decile? Given potential differences in conditions, we subset the data to each condition to estimate separately.

### create subset for REL+  
DF\_RELpos <- subset(DF, StimType == "REL+")  
  
### create subset for REL-  
DF\_RELneg <- subset(DF, StimType == "REL-")  
  
### create subset for UNREL  
DF\_UNREL <- subset(DF, StimType == "UNREL")

# individual SEs for REL+ condition   
cutoff\_relpos <- calculate\_cutoff(population = DF\_RELpos,  
 grouping\_items = "item",   
 score = "Response",   
 minimum = min(DF\_RELpos$Response),  
 maximum = max(DF\_RELpos$Response))  
  
SE1 <- tapply(DF\_RELpos$Response, DF\_RELpos$item, function (x) { sd(x)/sqrt(length(x)) })  
SE1

## burro casa cioccolato computer corona famiglia febbre   
## 0.1557687 0.1766870 0.1537913 0.1998951 0.2176631 0.1481984 0.1481984   
## lavare libertà mondo muffin notizie peste salute   
## 0.1860313 0.1916883 0.2017999 0.1846399 0.1636991 0.1621336 0.1569829   
## salvare sole tempo termometro torta vaiolo   
## 0.1431403 0.1483131 0.1952478 0.1441268 0.1393278 0.1585099

cutoff\_relpos$cutoff

## 40%   
## 0.1564972

# individual SEs for REL- condition  
cutoff\_relneg <- calculate\_cutoff(population = DF\_RELneg,  
 grouping\_items = "item",   
 score = "Response",   
 minimum = min(DF\_RELneg$Response),  
 maximum = max(DF\_RELneg$Response))  
  
SE2 <- tapply(DF\_RELneg$Response, DF\_RELneg$item, function (x) { sd(x)/sqrt(length(x)) })  
SE2

## autobus costa dormire giacca hotel mangiare matrimonio   
## 0.17715152 0.18453238 0.16521553 0.15728951 0.17384120 0.14909443 0.20475544   
## motore palazzo pantalone piazza profumo ristorante sposa   
## 0.15085177 0.17073238 0.17481656 0.20640961 0.16138025 0.20118108 0.21208427   
## tram tumore uomo vacanza viaggio villaggio   
## 0.16872880 0.07001896 0.16715951 0.13096103 0.18557374 0.16663313

cutoff\_relneg$cutoff

## 40%   
## 0.166949

# individual SEs for UNREL condition  
cutoff\_unrel <- calculate\_cutoff(population = DF\_UNREL,  
 grouping\_items = "item",   
 score = "Response",   
 minimum = min(DF\_UNREL$Response),  
 maximum = max(DF\_UNREL$Response))  
  
SE3 <- tapply(DF\_UNREL$Response, DF\_UNREL$item, function (x) { sd(x)/sqrt(length(x)) })  
SE3

## affogare baco cannone cappio corridore dipendente   
## 0.1312204 0.1622734 0.2131240 0.1750919 0.1781088 0.1796297   
## disturbare fetore firmamento funerale gusto ladro   
## 0.1535516 0.1507766 0.1973986 0.1121214 0.1526631 0.1473738   
## malevolenza mestolo nettare oceano offendersi orgasmo   
## 0.1379994 0.1218824 0.1755445 0.1511708 0.1652155 0.1458857   
## perfezione tradire   
## 0.2210864 0.1158751

cutoff\_unrel$cutoff

## 40%   
## 0.1510131

# sequence of sample sizes to try  
samplesize\_values <- seq(25, 300, 5)  
  
# create a blank table for us to save the values in positive ----  
sim\_table <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF\_RELpos$item)))  
# make it a data frame  
sim\_table <- as.data.frame(sim\_table)  
  
# add a place for sample size values   
sim\_table$sample\_size <- NA  
sim\_table$var <- "Response"  
  
# make a second table for negative -----  
sim\_table2 <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF\_RELneg$item)))  
  
# make it a data frame  
sim\_table2 <- as.data.frame(sim\_table2)  
  
# add a place for sample size values   
sim\_table2$sample\_size <- NA  
sim\_table2$var <- "Response"  
  
# make a second table for unrelated -----  
sim\_table3 <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF\_UNREL$item)))  
  
# make it a data frame  
sim\_table3 <- as.data.frame(sim\_table3)  
  
# add a place for sample size values   
sim\_table3$sample\_size <- NA  
sim\_table3$var <- "Response"  
  
# loop over sample size  
for (i in 1:length(samplesize\_values)){  
   
 # related positive temp variables ----  
 temp\_RELpos <- DF\_RELpos %>%   
 dplyr::group\_by(item) %>%   
 dplyr::sample\_n(samplesize\_values[i], replace = T) %>%   
 dplyr::summarize(se1 = sd(Response)/sqrt(length(Response)))   
   
 # put in table   
 colnames(sim\_table)[1:length(unique(DF\_RELpos$item))] <- temp\_RELpos$item  
 sim\_table[i, 1:length(unique(DF\_RELpos$item))] <- temp\_RELpos$se1  
 sim\_table[i, "sample\_size"] <- samplesize\_values[i]  
   
 # related negative temp variables ----  
 temp\_RELneg <-DF\_RELneg %>%   
 dplyr::group\_by(item) %>%   
 dplyr::sample\_n(samplesize\_values[i], replace = T) %>%   
 dplyr::summarize(se2 = sd(Response)/sqrt(length(Response)))   
  
 # put in table   
 colnames(sim\_table2)[1:length(unique(DF\_RELneg$item))] <- temp\_RELneg$item  
 sim\_table2[i, 1:length(unique(DF\_RELneg$item))] <- temp\_RELneg$se2  
 sim\_table2[i, "sample\_size"] <- samplesize\_values[i]  
   
 # unrelated temp variables ----  
 temp\_UNREL <-DF\_UNREL %>%   
 dplyr::group\_by(item) %>%   
 dplyr::sample\_n(samplesize\_values[i], replace = T) %>%   
 dplyr::summarize(se3 = sd(Response)/sqrt(length(Response)))   
  
 # put in table   
 colnames(sim\_table3)[1:length(unique(DF\_UNREL$item))] <- temp\_UNREL$item  
 sim\_table3[i, 1:length(unique(DF\_UNREL$item))] <- temp\_UNREL$se3  
 sim\_table3[i, "sample\_size"] <- samplesize\_values[i]  
}

#### Minimum Sample Size

Suggestions for REL+ Condition:

# multiply by correction   
cutoff <- quantile(SE1, probs = .4)  
  
final\_sample <-   
 sim\_table %>%  
 pivot\_longer(cols = -c(sample\_size, var)) %>%   
 dplyr::rename(item = name, se = value) %>%   
 dplyr::group\_by(sample\_size, var) %>%   
 dplyr::summarize(percent\_below = sum(se <= cutoff)/length(unique(DF\_RELpos$item))) %>%   
 dplyr::arrange(percent\_below) %>%   
 ungroup()

## `summarise()` has grouped output by 'sample\_size'. You can override using the  
## `.groups` argument.

flextable(final\_sample %>% head()) %>% autofit()

| sample\_size | var | percent\_below |
| --- | --- | --- |
| 25 | Response | 0.0 |
| 30 | Response | 0.0 |
| 35 | Response | 0.0 |
| 40 | Response | 0.0 |
| 50 | Response | 0.0 |
| 45 | Response | 0.1 |

final\_table\_pos <- calculate\_correction(  
 proportion\_summary = final\_sample,  
 pilot\_sample\_size = length(unique(DF\_RELpos$ssID)),  
 proportion\_variability = cutoff\_relpos$prop\_var  
 )  
  
flextable(final\_table\_pos) %>%   
 autofit()

| percent\_below | sample\_size | corrected\_sample\_size |
| --- | --- | --- |
| 80 | 100 | 66.65688 |
| 90 | 105 | 70.07476 |
| 90 | 105 | 70.07476 |
| 95 | 125 | 84.64245 |

Suggestions for REL- Condition:

cutoff <- quantile(SE2, probs = .4)  
  
final\_sample2 <-   
 sim\_table2 %>%  
 pivot\_longer(cols = -c(sample\_size, var)) %>%   
 dplyr::rename(item = name, se = value) %>%   
 dplyr::group\_by(sample\_size, var) %>%   
 dplyr::summarize(percent\_below = sum(se <= cutoff)/length(unique(DF\_RELneg$item))) %>%   
 dplyr::arrange(percent\_below) %>%   
 ungroup()

## `summarise()` has grouped output by 'sample\_size'. You can override using the  
## `.groups` argument.

flextable(final\_sample2 %>% head()) %>% autofit()

| sample\_size | var | percent\_below |
| --- | --- | --- |
| 25 | Response | 0.05 |
| 30 | Response | 0.05 |
| 35 | Response | 0.05 |
| 40 | Response | 0.05 |
| 50 | Response | 0.15 |
| 55 | Response | 0.15 |

final\_table\_neg <- calculate\_correction(  
 proportion\_summary = final\_sample2,  
 pilot\_sample\_size = length(unique(DF\_RELneg$ssID)),  
 proportion\_variability = cutoff\_relneg$prop\_var  
 )  
  
flextable(final\_table\_neg) %>%   
 autofit()

| percent\_below | sample\_size | corrected\_sample\_size |
| --- | --- | --- |
| 85 | 95 | 61.90142 |
| 85 | 95 | 61.90142 |
| 100 | 100 | 66.58936 |
| 100 | 100 | 66.58936 |

Suggestions for UNREL Condition:

cutoff <- quantile(SE3, probs = .4)  
  
final\_sample3 <-   
 sim\_table3 %>%  
 pivot\_longer(cols = -c(sample\_size, var)) %>%   
 dplyr::rename(item = name, se = value) %>%   
 dplyr::group\_by(sample\_size, var) %>%   
 dplyr::summarize(percent\_below = sum(se <= cutoff)/length(unique(DF\_UNREL$item))) %>%   
 dplyr::arrange(percent\_below) %>%   
 ungroup()

## `summarise()` has grouped output by 'sample\_size'. You can override using the  
## `.groups` argument.

flextable(final\_sample3 %>% head()) %>% autofit()

| sample\_size | var | percent\_below |
| --- | --- | --- |
| 25 | Response | 0.00 |
| 30 | Response | 0.00 |
| 35 | Response | 0.00 |
| 40 | Response | 0.15 |
| 45 | Response | 0.15 |
| 55 | Response | 0.20 |

final\_table\_unrel <- calculate\_correction(  
 proportion\_summary = final\_sample3,  
 pilot\_sample\_size = length(unique(DF\_UNREL$ssID)),  
 proportion\_variability = cutoff\_unrel$prop\_var  
 )  
  
flextable(final\_table\_unrel) %>%   
 autofit()

| percent\_below | sample\_size | corrected\_sample\_size |
| --- | --- | --- |
| 80 | 95 | 62.33224 |
| 85 | 110 | 73.33152 |
| 90 | 115 | 76.92278 |
| 95 | 130 | 87.59422 |

Based on these simulations, we can decide our minimum sample size by examining all three potential scores at 80% of items below the criterion, = 67, = 62, or = 62. These scores are all very similar, and we should select the largest one.

#### Maximum Sample Size

In this example, we could set our maximum sample size for 90% power which would equate to: = 70, = 67, or = 77.

#### Final Sample Size

The final sample size should be selected from the largest suggested sample size based on condition to ensure that all conditions are adequately measured.