EDA_Autumn.Lawrence

2024-27-24

Introduction: DEMENTIA

Dementia is the loss of cognitive functionnig and is more common in people as they age. It can range in its severity and often interferes with an person's daily activities.

A study was conducted over a span of 12 weeks observing 15 nursing home patients diagnosed with dementia. Researchers recorded the number of aggressive incidents each day and totaled the counts of aggressive incidents per patient. They categorized these days into two categories: "moon" days (full moon +/-1 day) and "other" days.

This project will explore the variables "moon" days and "other" days among diagnosed patients. The projects aims to determine if the type of day influences the rate of aggressive incidents with the observed patients.

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE,error=TRUE)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
dementia_moon <- read.delim("~/Desktop/Math130/Data/dementia_moon.txt", header =TRUE, sep ="\t" )
str(dementia_moon)
```

```
## 'data.frame':    15 obs. of  3 variables:
##  $ patient : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ aggmoon : num  3.33 3.67 2.67 3.33 3.33 3.67 4.67 2.67 6 4.33 ...
##  $ aggnother: num  0.27 0.59 0.32 0.19 1.26 0.11 0.3 0.4 1.59 0.6 ...
```

This data set has 15 observations for 3 variables. The 3 variables are aggmoon, aggother, and patients.

Univariate Exploration: Variables

Moon days

I want to find the aggression incidences that occurred on moon days per patient. First I will group the variable moon days with the corresponding patient.

```
a1<- by_patient <- group_by(dementia_moon, patient)
reframe(by_patient, aggmoon = mean(aggmoon, na.rm=TRUE))
```

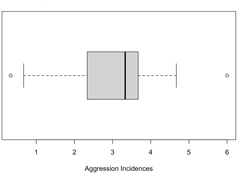
```
## # A tibble: 15 × 2
##    patient aggmoon
##      <int>   <dbl>
##  1       1    3.33
##  2       2    3.67
##  3       3    2.67
##  4       4    3.33
##  5       5    3.33
##  6       6    3.67
##  7       7    4.67
##  8       8    2.67
##  9       9    6
## 10      10    4.33
## 11      11    3.33
## 12      12    0.67
## 13      13    1.33
## 14      14    0.33
## 15      15    2
```

Since I know that we are measuring incidences of aggression, it is redundant to call the variable 'aggmoon'. I will shorten this to just 'moons'. This variable still represents the amount of aggression incidences, it now just indicates that it happened on days labeled moon days.

```
a2 <-rename(a1, moons = "aggmoon")
select(a2,-"aggother" )
```

```
## # A tibble: 15 × 2
## # Groups:   patient [15]
##    patient moons
##      <int> <dbl>
##  1       1  3.33
##  2       2  3.67
##  3       3  2.67
##  4       4  3.33
##  5       5  3.33
##  6       6  3.67
##  7       7  4.67
##  8       8  2.67
##  9       9  6
## 10      10  4.33
## 11      11  3.33
## 12      12  0.67
## 13      13  1.33
## 14      14  0.33
## 15      15  2
```

```
boxplot(dementia_moon$aggmoon, horizontal = TRUE, main="Distribution of aggressive incidences of dementia patient
s on Moon Days", xlab = "Aggression Incidences")
```



The boxplot demonstrates the distribution of the data for the variable: the minimum, maximum, mean, and robustic statistics. By using the summary() function we can see the values of each of these statistics. On average, patients had about 3 aggression incidences on moon days. Some outliers being that the minimum amount of incidences during moon days is 1 and the maximum being 6.

```
summary(dementia_moon$aggmoon)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.330   2.335   3.330   3.022   3.670   6.000
```

Other Days

I want to find the aggression incidences that occurred on other days than moon days per patient. First I will group the variable other days with the corresponding patient.

```
a3<- by_patient <- group_by(dementia_moon, patient)
reframe(by_patient, aggother = mean(aggother, na.rm=TRUE))
```

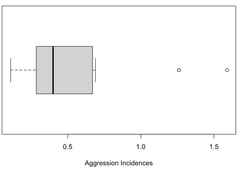
```
## # A tibble: 15 × 2
##    patient aggother
##      <int>   <dbl>
##  1       1    0.27
##  2       2    0.59
##  3       3    0.32
##  4       4    0.19
##  5       5    1.26
##  6       6    0.11
##  7       7    0.3
##  8       8    0.4
##  9       9    1.59
## 10      10    0.6
## 11      11    0.65
## 12      12    0.69
## 13      13    1.26
## 14      14    0.23
## 15      15    0.38
```

Since I know that we are measuring incidences of aggression, it is redundant to call the variable aggother. I will shorten this to just 'other'. This variable still represents the amount of aggression incidences, it now just indicates that it happened on days labeled other days. We also can see that the boxplot is negatively/left skewed, meaning that most data is lower values than higher values.

```
a4 <-rename(a3, other = "aggother")
select(a4,-"aggmoon")
```

```
## # A tibble: 15 × 2
## # Groups:   patient [15]
##    patient other
##      <int> <dbl>
##  1       1  0.27
##  2       2  0.59
##  3       3  0.32
##  4       4  0.19
##  5       5  1.26
##  6       6  0.11
##  7       7  0.3
##  8       8  0.4
##  9       9  1.59
## 10      10  0.6
## 11      11  0.65
## 12      12  0.69
## 13      13  1.26
## 14      14  0.23
## 15      15  0.38
```

```
boxplot(dementia_moon$aggother, horizontal = TRUE, main="Distribution of aggressive incidences of dementia patien
ts on Other Days", xlab = "Aggression Incidences")
```



The boxplot demonstrates the distribution of the data for the variable: the minimum, maximum, mean, and robustic statistics. By using the summary() function we can see the values of each of these statistics. On average, patients had virtually no aggression incidences on the days other than moon days. Some outliers being that the minimum amount of incidences during other days was virtually none and the maximum being 2. We also can see that the boxplot is positively/right skewed, meaning that most data are closer together and that the outliers are few and on the higher end.

```
summary(dementia_moon$aggother)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1100  0.2850  0.4000  0.5893  0.6700  1.5900
```

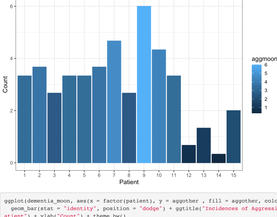
Patients

Is not a variable that we need to consider unless we were asking questions about individual patients. For example, "Was patient 10 on average have more aggression incidences on moon days or other days?"

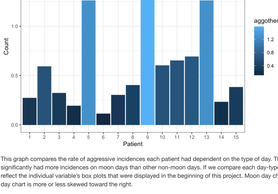
Bivariate Exploration

I want to determine which type of day did dementia patients had more incidences of aggression on: moon days or other days. To do this, I created a grouped bar chart.

```
ggplot(dementia_moon, aes(x = factor(patient), y = aggmoon , fill = aggmoon, colour = aggmoon)) +
  geom_bar(stat = "identity", position = "dodge") + ggtitle('Incidences of Aggression per Type of Day') + xlab('P
atient') + ylab('Count')+ theme_bw()
```




```
ggplot(dementia_moon, aes(x = factor(patient), y = aggother , fill = aggother, colour = aggother)) +
  geom_bar(stat = "identity", position = "dodge") + ggtitle('Incidences of Aggression per Type of Day') + xlab('P
atient') + ylab('Count') + theme_bw()
```



This graph compares the rate of aggressive incidences each patient had dependent on the type of day. This graph shows that dementia patients significantly had more incidences on moon days than other non-moon days. If we compare each day-types graph we will see that the bar charts reflect the individual variable's box plots that were displayed in the beginning of this project. Moon day chart is more left skewed and the Other day chart is more or less skewed toward the right.

Conclusion:

From looking at the variables individually we clearly see that moon days had a much more visible correlation with dementia patient's rate of aggressive incidences. There was a average of 3 incidences per patient on moon days. Meanwhile, on non moon days (other days) there were little to no incidences of aggression. The average incident per patient was below 1 on these days. From looking at the graphics, this leads me to conclude that the type of day does in fact have a correlation (if not influence) on the rate of incidents per dementia patient.