Module 3 Assignment - Airport Screening Analysis

Adam Kelly

University of South Florida

Author Note

Correspondence concerning this article should be addressed to Adam Kelly, 4202 E Fowler Ave, Tampa, FL 33620. E-mail: awkelly@mail.usf.edu

Module 3 Assignment - Airport Screening Analysis

Module 3 Assignment - Airport Screening Analysis

What is the Association Between the Number of Airport Screeners and Security Violations?

The association between these two variables is the assumption that the number of airport security screeners influences, in one way or another, the amount of security violations that were detected. Each of these 20 cases is an independent study with different amounts of airport screeners as well as violations. The goal of this assignment is to conduct analysis on these two variables and determine if there is any correlation between the number of screeners and the amount of violations detected.

```
df <- read.csv("https://www.dropbox.com/s/e51igg1rcp40kub/Module%203%20data.csv?dl=1", r
```

Calculating Pearson's Correlation Coefficient

```
dfpearson = cor(x = df$pre.boarding.screeners,y = df$security.violations.detected, metho
dfpearson
```

```
## [1] 0.8375321
```

Pearson's coefficient is a measure of the linear correlation of two values. When drawing a line of best fit on a scatterplot, Pearson's coefficient is the measure of how far the data points are from the line. In this case, the value of Pearson's coefficient, r, between the number of pre-boarding screeners and security violations is a value of 0.838. This implies that there is a very strong correlation between the number of pre-boarding screeners and the number of violations that were detected. This means that the number of violations detected is largely influenced by the number of screeners.

Calculating Spearman's Rank Coefficient

```
dfspearman = cor(x = df$pre.boarding.screeners,y = df$security.violations.detected, meth
dfspearman
```

```
## [1] 0.7575423
```

While Spearman's coefficient on the other hand is also a measure of correlation, it does
not take linearity into account. On a scatterplot, a line of best fit will be drawn, however it
does not necessarily have to be linear, and how far the data points are from the line
determines the correlation. In this case, the value of Spearman's coefficient is .758 indicating
again that there is a heavy correlation between the number of pre-boarding screeners and the
number of violations detected.

Scatterplot

As we can see, the scaterplot is quite linear and the line of best fit is very plain to see.
This plot shows us that there is indeed heavy correlation between the two variables tested
and it agrees with our earlier inferences alongside Pearson and Spearman's coefficients.

```
library(ggplot2)
```

```
dfscatter2 = ggplot(df, aes(x = df$pre.boarding.screeners,y = df$security.violations.det
dfscatter2
```

References

R (Version 3.6.1; R Core Team, 2019) and the R-packages *ggplot2* (Version 3.2.1; Wickham, 2016), and *papaja* (Version 0.1.0.9842; Aust & Barth, 2018)

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja
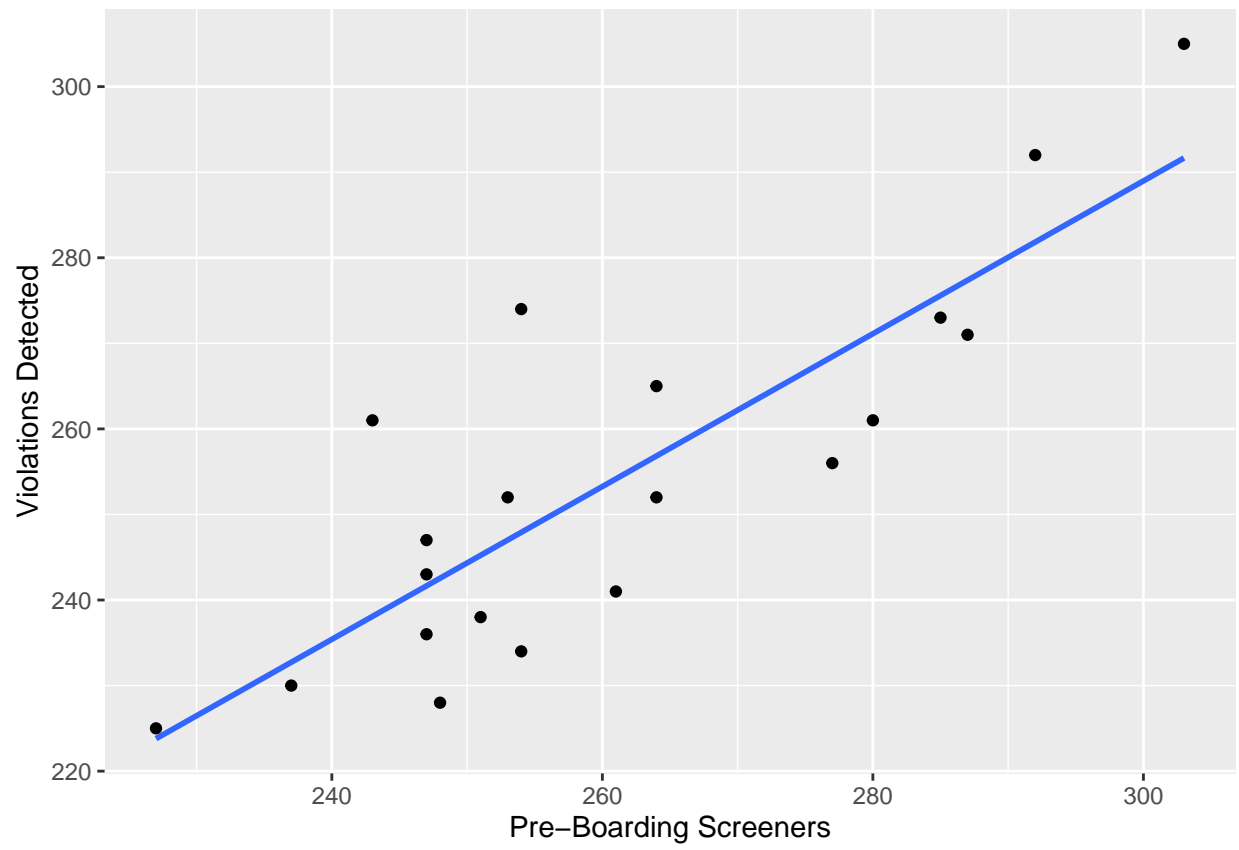
R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Figure 1