Validating the *FireEdge* Assessment

Robert W. Szarek

March 10, 2021

Overview and Context

This document was created as an R Markdown file from within R-Studio. The benefits of using an R Markdown file for scientific research is the ability to weave code and regular text into a single document, all the while providing anybody reviewing this file an opportunity to reproduce the analyses. The code included within the code chunks was written with R version 4.0.2.

This file walks the user through a validation study in connection with the *FireEdge* assessment, which is a cognitive and situation-based judgmenet test developed for the purposes of hiring entry-level firefighters. The project picks up post-exam development, with a study to help empirically demonstrate that test scores on the *FireEdge* correlate positively with overall job performance ratings for incumbents who participated in the study.

```
library(haven)
library(tidyverse)
```

Data Import

Once the required packages and settings have been loaded, we can proceed to import the dataset to begin running the various statistical analyses for the validation study.

```
val_data <- read_rds(file = "~/R-lang/FireEdge/data/FireEdge_data.Rds")
```

The original dataset includes item-level data, which we eliminate in the subsequent code chunk and focus on the overall composite score with the criteria of job performance, while also retaining some other nominal variables such as Race, Gender and Study location.

```
val_data_tidy <-
  val_data %>%
  select(
    Study,
    Gender,
    Race,
    Final_Score,
    Crit_Supv
  ) %>%
  rename(
    "JobPerformance" = Crit_Supv,
    "FireEdge_Score" = Final_Score
  )

str(val_data_tidy)
```

```
## tibble [339 x 5] (S3: tbl_df/tbl/data.frame)
##  $ Study         : chr [1:339] "Maryland" "Maryland" "Maryland" "Maryland" ...
```

```
##  $ Gender        : Factor w/ 2 levels "Male","Female": 1 2 2 2 1 2 2 1 2 2 ...
##  $ Race          : Factor w/ 7 levels "Black","Asian",..: 6 1 1 1 6 3 6 1 6 1 ...
##  $ FireEdge_Score: num [1:339, 1] 66.6 79.5 82.8 82.1 75.9 ...
##  $ JobPerformance: num [1:339] 45 NA 40 47 44 48 49 48 45 46 ...
```

The "chr" after the word Study indicates it is a character vector. We need to make it a factor instead as we know the three seperate states that took part in the validation study. The following code converts the Study variable from a character vector to a factor vector.

```
val_data_tidy <-
  val_data_tidy %>%
  mutate(Study = as_factor(Study))

str(val_data_tidy$Study)
```

```
##  Factor w/ 3 levels "Maryland","Colorado",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Descriptive Statistics

Next, we begin to investigate the descriptive statistics concerning the supervisor ratings of job performance. The code below creates a table of information broken down by Study location, such as the average job performance score, the standard deviation of the distribution, and the minimum and maximum scores.

```
val_data_tidy %>%
  group_by(Study) %>%
  summarise(
    .groups = "drop",
    RawTotal = n(),
    Missing = sum(is.na(JobPerformance)),
    True_Sample = RawTotal - Missing,
    Mean = mean(JobPerformance, na.rm = TRUE),
    SD = sd(JobPerformance, na.rm = TRUE),
    Min = min(JobPerformance, na.rm = TRUE),
    Max = max(JobPerformance, na.rm = TRUE)
  )
```

```
## # A tibble: 3 x 8
##   Study    RawTotal Missing True_Sample  Mean    SD   Min   Max
##   <fct>       <int>   <int>       <int> <dbl> <dbl> <dbl> <dbl>
## 1 Maryland       65       2          63  44.4  5.85    25    55
## 2 Colorado      205      13         192  45.4  6.30    26    62
## 3 Alabama        69      27          42  43.5  5.24    30    57
```

We see that the agency in Alabama had a total of 27 missing cases of supervisor ratings, while the Colorado agency had a total of 13 missing cases. The average scores are all consistent across the three study locations, as are the standard deviations, minimum and maximum scores. Next, we perform the same type of analysis for the *FireEdge* score.

```
val_data_tidy %>%
  group_by(Study) %>%
  summarise(
    .groups = "drop",
    RawTotal = n(),
    Missing = sum(is.na(FireEdge_Score)),
    True_Sample = RawTotal - Missing,
    Mean = mean(FireEdge_Score, na.rm = TRUE),
    SD = sd(FireEdge_Score, na.rm = TRUE),
```

```
    Min = min(FireEdge_Score, na.rm = TRUE),
    Max = max(FireEdge_Score, na.rm = TRUE)
  )
```

```
## # A tibble: 3 x 8
##   Study    RawTotal Missing True_Sample  Mean    SD   Min   Max
##   <fct>       <int>   <int>       <int> <dbl> <dbl> <dbl> <dbl>
## 1 Maryland       65       0          65  75.2  5.56  59.8  84.1
## 2 Colorado      205       0         205  77.6  5.71  49.4  85.6
## 3 Alabama        69       0          69  72.7  8.95  43.8  85.4
```

The means all fall in line quite closely, with an average score of around a 75.00 and a standard deviation around 6.5. Minimum and maximum scores were all quite consistent across the three studies. Once we have taken a look at these statistics, we create our finalized dataset by removing the missing data that we observed in the job performance ratings with the following code.

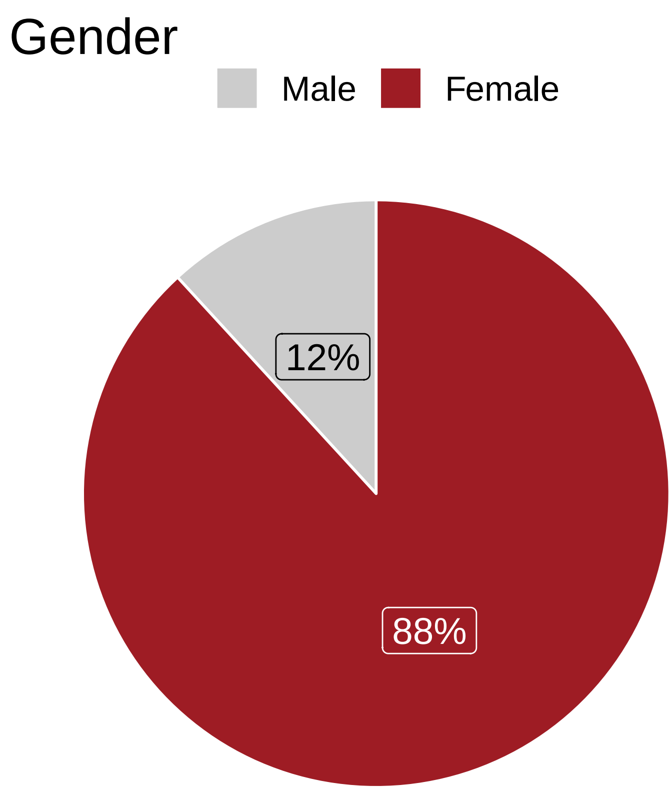
```
val_data_tidy <-
val_data_tidy %>%
  filter(!is.na(JobPerformance))
```

We remove the missing data points and determine we have a total of 297 rows of data to use for the validation study. Next, we create some simple pie charts to determine the number of participants based on Race, Gender and Study.

Descriptive Statistics for Study Sample

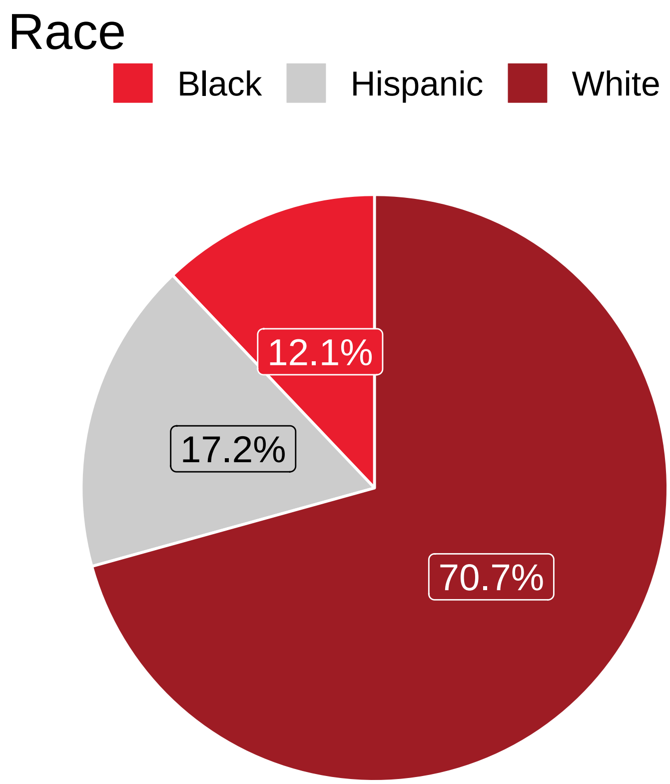
The first pie chart presents counts of Gender represented in the validaton study. Unfortunately, a general trend in firefighting is that males usually outweigh females by a considerable portion. This is reflected in the sample size we have obtained where there are significantly more males than females.

```
val_data_tidy %>%
  count(Gender) %>%
  filter(Gender == "Male" | Gender == "Female") %>%
  mutate(Percentage = scales::percent(n / sum(n))) %>%
  ggplot(aes(x = "", y = n, fill = Gender)) +
  geom_bar(stat = "identity",  color = "white", show.legend = TRUE) +
  coord_polar("y") +
  theme_void(base_size = 16) +
  scale_fill_manual(name= "", values = c("#CCCCCC", "#9E1C24")) +
    labs(x = "",
         y = "",
         title = "Gender") +
  geom_label(aes(label = Percentage), size=5, color = c("black", "white"),
             position = position_stack(vjust=0.5),show.legend = FALSE) +
  theme(legend.position = "top")
```



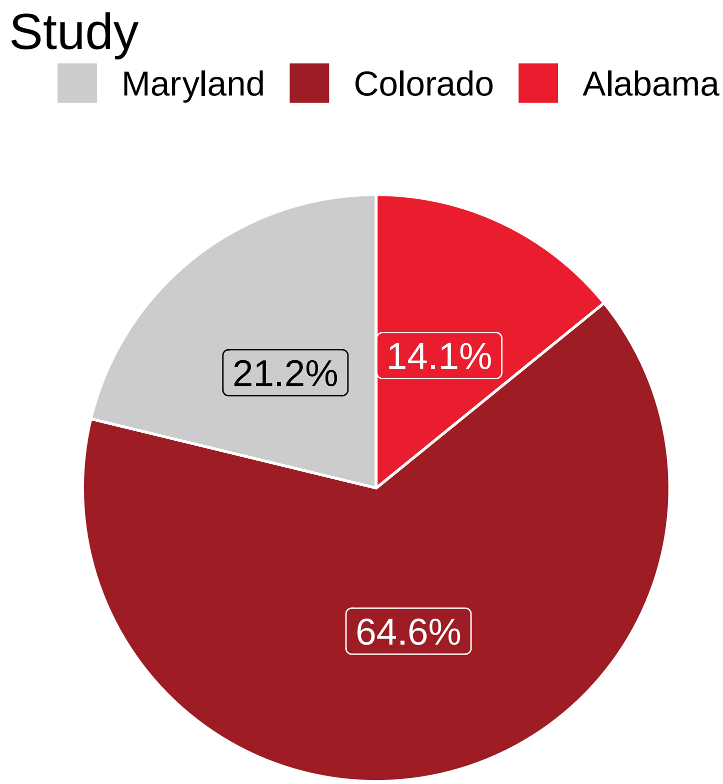
The second pie chart outlines Race counts for the validation study. We can see that minority representation rates are not as high as would be ideal, which would have been in the 20-25% range. However, we still have a robust representation of minority candidates for the study to ensure the sample represents the population of firefighters in general.

```
val_data_tidy %>%
  count(Race) %>%
  filter(Race == "Hispanic" | Race == "White" | Race == "Black") %>%
  mutate(Percentage = scales::percent(n / sum(n))) %>%
  ggplot(aes(x = "", y = n, fill = Race)) +
  geom_bar(stat = "identity",  color = "white", show.legend = TRUE) +
  coord_polar("y") +
  theme_void(base_size = 16) +
  scale_fill_manual(name= "", values = c( "#EA1D2E", "#CCCCCC", "#9E1C24")) +
    labs(x = "",
         y = "",
         title = "Race") +
  geom_label(aes(label = Percentage), size=5, color = c("white", "black", "white"),
             position = position_stack(vjust=0.5),show.legend = FALSE) +
  theme(legend.position = "top")
```



The final pie chart outlines the location from which participants were pulled. The majority of participants derived from a firefighting agency located in Colorado, with 64.6% of the total sample. Maryland was next, with 21.2% and Alabama was last, with 14.1%. As we do not expect the firefighting job to differ significantly across the nation, for the purposes of streamlining this validation study we will collapse all three agencies into a single study.

```
val_data_tidy %>%
  count(Study) %>%
  mutate(Percentage = scales::percent(n / sum(n))) %>%
  ggplot(aes(x = "", y = n, fill = Study)) +
  geom_bar(stat = "identity",  color = "white", show.legend = TRUE) +
  coord_polar("y") +
  theme_void(base_size = 16) +
  scale_fill_manual(name= "", values = c("#CCCCCC", "#9E1C24", "#EA1D2E")) +
    labs(x = "",
         y = "",
         title = "Study") +
  geom_label(aes(label = Percentage), size=5, color = c("black", "white", "white"),
             position = position_stack(vjust=0.5),show.legend = FALSE) +
  theme(legend.position = "top")
```



Correlation Coefficients

The following code chunk runs the Pearson Product-Moment Correlation Coefficient between the *FireEdge* test score and job performance across the 297 participants in the study. As we can see, the t-test was highly significant, with a t-score of $t(295) = 4.884$, $p < .001$. The correlation coefficient was .273, which is statistically significant.

```
cor.test(x = val_data_tidy$FireEdge_Score,
         y = val_data_tidy$JobPerformance,
         method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  val_data_tidy$FireEdge_Score and val_data_tidy$JobPerformance
## t = 4.8837, df = 295, p-value = 1.71e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1648158 0.3756179
## sample estimates:
##       cor
## 0.2734977
```

Linear Regression Model