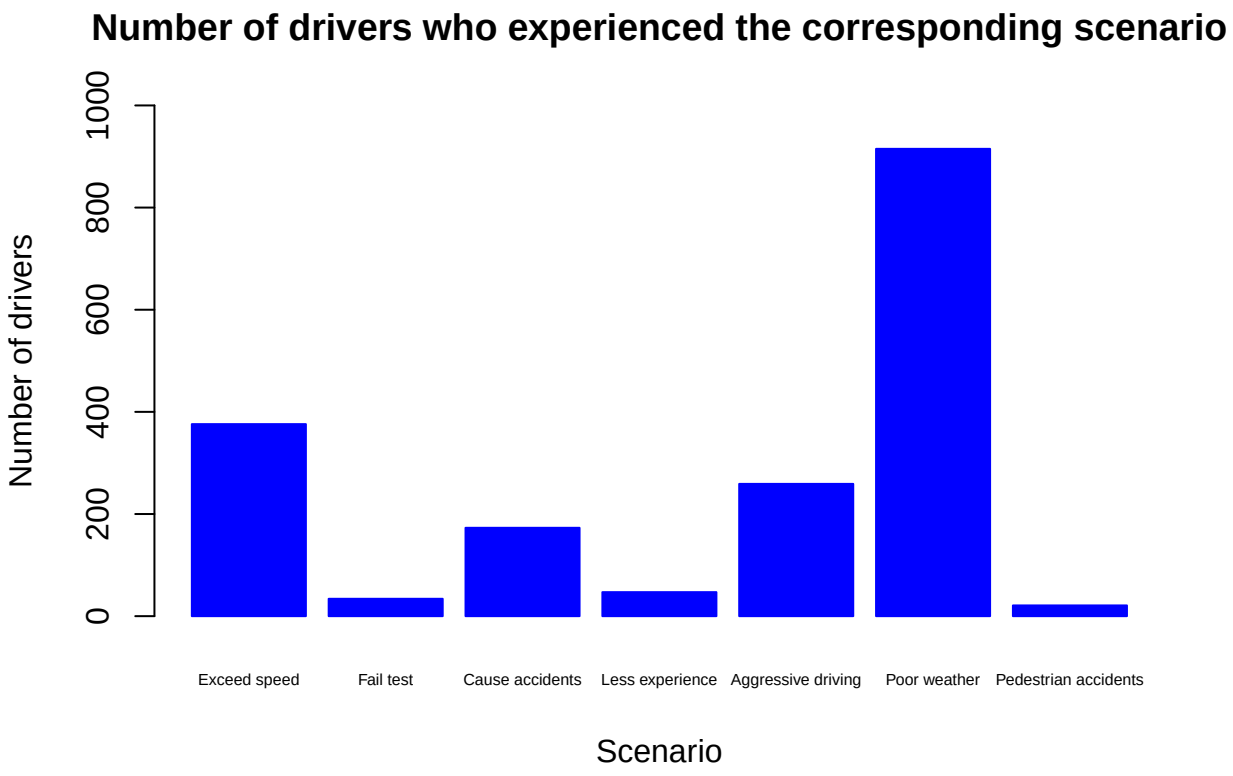
Survey for Drivers

Shu Dong

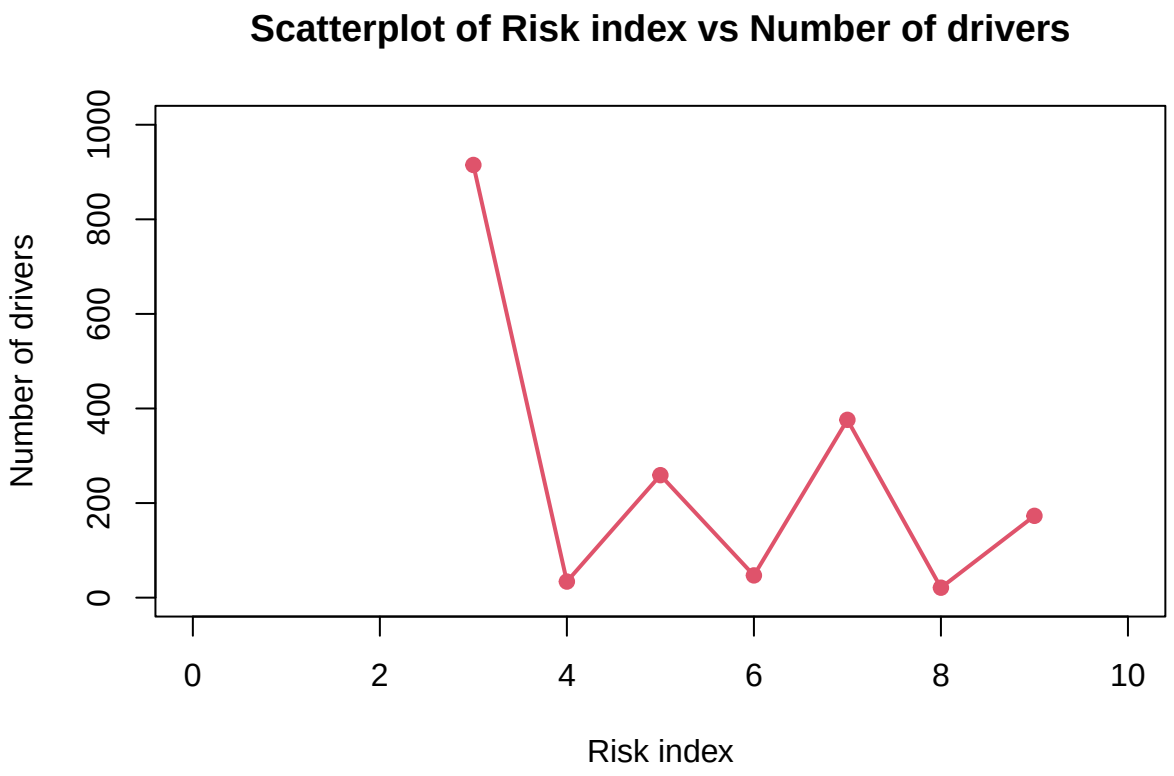
2022-12-06

```r
# Import the Survey data in R
survey = read.csv("SURVEY.csv", header = TRUE)
scenario = survey$Scenario
number_drivers = survey$Number.of.experience
risk_index = survey$Risk.Index
data_survey = data.frame(scenario, number_drivers, risk_index)
```

```r
# Plot the bar plot for the number of drivers
barplot(number_drivers,
        xlab = "Scenario",
        ylab = "Number of drivers", ylim = c(0, 1000),
        names.arg = scenario, cex.names = 0.53,
        col = "blue", border = "blue",
        main = "Number of drivers who experienced the corresponding scenario")
```



```r
# Scatterplot for the risk index vs number of drivers
plot(risk_index, number_drivers,
     ylab = "Number of drivers", ylim = c(0, 1000),
     xlab = "Risk index", xlim = c(0, 10),
     main = "Scatterplot of Risk index vs Number of drivers",
     pch = 19, col = 2)
lines(sort(risk_index), c(915, 34, 259, 47, 376, 21, 173),
      col = 2, lwd = 2, pch = 1)
```



```
# Let the risky scenario be treatment A and let risk index be treatment B
a = 7
b = 7
n = 1

y_xxx = (1 / (a * b * n)) * (915 + 34 + 259 + 47 + 376 + 21 + 173)
y_1xx = (1 / (b * n)) * 915
y_2xx = (1 / (b * n)) * 34
y_3xx = (1 / (b * n)) * 259
y_4xx = (1 / (b * n)) * 47
y_5xx = (1 / (b * n)) * 376
y_6xx = (1 / (b * n)) * 21
y_7xx = (1 / (b * n)) * 173
y_x1x = (1 / (a * n)) * 915
y_x2x = (1 / (a * n)) * 34
y_x3x = (1 / (a * n)) * 259
y_x4x = (1 / (a * n)) * 47
y_x5x = (1 / (a * n)) * 376
y_x6x = (1 / (a * n)) * 21
y_x7x = (1 / (a * n)) * 173
y_11x = (1 / n) * 915
y_21x = (1 / n) * 0
y_31x = (1 / n) * 0
y_41x = (1 / n) * 0
y_51x = (1 / n) * 0
y_61x = (1 / n) * 0
y_71x = (1 / n) * 0
y_12x = (1 / n) * 0
y_22x = (1 / n) * 34
y_32x = (1 / n) * 0
y_42x = (1 / n) * 0
y_52x = (1 / n) * 0
y_62x = (1 / n) * 0
y_72x = (1 / n) * 0
y_13x = (1 / n) * 0
y_23x = (1 / n) * 0
y_33x = (1 / n) * 259
y_43x = (1 / n) * 0
y_53x = (1 / n) * 0
y_63x = (1 / n) * 0
y_73x = (1 / n) * 0
y_14x = (1 / n) * 0
y_24x = (1 / n) * 0
y_34x = (1 / n) * 0
y_44x = (1 / n) * 47
y_54x = (1 / n) * 0
y_64x = (1 / n) * 0
y_74x = (1 / n) * 0
y_15x = (1 / n) * 0
y_25x = (1 / n) * 0
y_35x = (1 / n) * 0
y_45x = (1 / n) * 0
y_55x = (1 / n) * 376
```

```
y_65x = (1 / n) * 0
y_75x = (1 / n) * 0
y_16x = (1 / n) * 0
y_26x = (1 / n) * 0
y_36x = (1 / n) * 0
y_46x = (1 / n) * 0
y_56x = (1 / n) * 0
y_66x = (1 / n) * 21
y_76x = (1 / n) * 0
y_17x = (1 / n) * 0
y_27x = (1 / n) * 0
y_37x = (1 / n) * 0
y_47x = (1 / n) * 0
y_57x = (1 / n) * 0
y_67x = (1 / n) * 0
y_77x = (1 / n) * 173

SS_A = b * n * (((y_1xx - y_xxx)^2) + ((y_2xx - y_xxx)^2) +
                  ((y_3xx - y_xxx)^2) + ((y_4xx - y_xxx)^2) +
                  ((y_5xx - y_xxx)^2) + ((y_6xx - y_xxx)^2) +
                  ((y_7xx - y_xxx)^2))
SS_B = a * n * (((y_x1x - y_xxx)^2) + ((y_x2x - y_xxx)^2) +
                  ((y_x3x - y_xxx)^2) + ((y_x4x - y_xxx)^2) +
                  ((y_x5x - y_xxx)^2) + ((y_x6x - y_xxx)^2) +
                  ((y_x7x - y_xxx)^2))
SS_AB = n * (((y_11x - y_1xx - y_x1x + y_xxx)^2) +
  ((y_21x - y_2xx - y_x1x + y_xxx)^2) + ((y_31x - y_3xx - y_x1x + y_xxx)^2) +
  ((y_41x - y_4xx - y_x1x + y_xxx)^2) + ((y_51x - y_5xx - y_x1x + y_xxx)^2) +
  ((y_61x - y_6xx - y_x1x + y_xxx)^2) + ((y_71x - y_7xx - y_x1x + y_xxx)^2) +
  ((y_12x - y_1xx - y_x2x + y_xxx)^2) + ((y_22x - y_2xx - y_x2x + y_xxx)^2) +
  ((y_32x - y_3xx - y_x2x + y_xxx)^2) + ((y_42x - y_4xx - y_x2x + y_xxx)^2) +
  ((y_52x - y_5xx - y_x2x + y_xxx)^2) + ((y_62x - y_6xx - y_x2x + y_xxx)^2) +
  ((y_72x - y_7xx - y_x2x + y_xxx)^2) + ((y_13x - y_1xx - y_x3x + y_xxx)^2) +
  ((y_23x - y_2xx - y_x3x + y_xxx)^2) + ((y_33x - y_3xx - y_x3x + y_xxx)^2) +
  ((y_43x - y_4xx - y_x3x + y_xxx)^2) + ((y_53x - y_5xx - y_x3x + y_xxx)^2) +
  ((y_63x - y_6xx - y_x3x + y_xxx)^2) + ((y_73x - y_7xx - y_x3x + y_xxx)^2) +
  ((y_14x - y_1xx - y_x4x + y_xxx)^2) + ((y_24x - y_2xx - y_x4x + y_xxx)^2) +
  ((y_34x - y_3xx - y_x4x + y_xxx)^2) + ((y_44x - y_4xx - y_x4x + y_xxx)^2) +
  ((y_54x - y_5xx - y_x4x + y_xxx)^2) + ((y_64x - y_6xx - y_x4x + y_xxx)^2) +
  ((y_74x - y_7xx - y_x4x + y_xxx)^2) + ((y_15x - y_1xx - y_x5x + y_xxx)^2) +
  ((y_25x - y_2xx - y_x5x + y_xxx)^2) + ((y_35x - y_3xx - y_x5x + y_xxx)^2) +
  ((y_45x - y_4xx - y_x5x + y_xxx)^2) + ((y_55x - y_5xx - y_x5x + y_xxx)^2) +
  ((y_65x - y_6xx - y_x5x + y_xxx)^2) + ((y_75x - y_7xx - y_x5x + y_xxx)^2) +
  ((y_16x - y_1xx - y_x6x + y_xxx)^2) + ((y_26x - y_2xx - y_x6x + y_xxx)^2) +
  ((y_36x - y_3xx - y_x6x + y_xxx)^2) + ((y_46x - y_4xx - y_x6x + y_xxx)^2) +
  ((y_56x - y_5xx - y_x6x + y_xxx)^2) + ((y_66x - y_6xx - y_x6x + y_xxx)^2) +
  ((y_76x - y_7xx - y_x6x + y_xxx)^2) + ((y_17x - y_1xx - y_x7x + y_xxx)^2) +
  ((y_27x - y_2xx - y_x7x + y_xxx)^2) + ((y_37x - y_3xx - y_x7x + y_xxx)^2) +
  ((y_47x - y_4xx - y_x7x + y_xxx)^2) + ((y_57x - y_5xx - y_x7x + y_xxx)^2) +
  ((y_67x - y_6xx - y_x7x + y_xxx)^2) + ((y_77x - y_7xx - y_x7x + y_xxx)^2))
SS_Err = ((915 - y_xxx)^2) + ((34 - y_xxx)^2) + ((259 - y_xxx)^2) +
  ((47 - y_xxx)^2) + ((376 - y_xxx)^2) + ((21 - y_xxx)^2) + ((173 - y_xxx)^2)
SS_Tot = SS_A + SS_B + SS_AB + SS_Err
```

```r
# Construct the ANOVA table for two-way factorial design
df_A = a - 1
df_B = b - 1
df_AB = (a - 1) * (b - 1)
df_Err = a * b * (n + 1)
df_Tot =df_A + df_B + df_Err
MS_A = SS_A / df_A
MS_B = SS_B / df_B
MS_AB = SS_AB / df_AB
MS_Err = SS_Err / df_Err
F0_A = MS_A / MS_Err
F0_B = MS_B / MS_Err
F0_AB = MS_AB / MS_Err

ss = c(SS_A, SS_B, SS_AB, SS_Err, SS_Tot)
df = c(df_A, df_B, df_AB, df_Err, df_Tot)
ms = c(MS_A, MS_B, MS_AB, MS_Err)
F0 = c(F0_A, F0_B, F0_AB)

max_ln = length(ss)
name_row = c("Variation for A", "Variation for B", "Variation for AB",
             "Variation for Error", "Variation for Total")
name_col = c("Sum of Squares", "Degrees of Freedom", "Mean Squares", "F0")

ANOVA_data = setNames(data.frame(col1 = c(ss, rep(NA, max_ln - length(ss))),
                         col2 = c(df, rep(NA, max_ln - length(df))),
                         col3 = c(ms, rep(NA, max_ln - length(ms))),
                         col4 = c(F0, rep(NA, max_ln - length(F0))),
                         row.names = name_row), name_col)

library(knitr)
knitr::kable(ANOVA_data, "pipe")
```

	Sum of Squares	Degrees of Freedom	Mean Squares	F0
Variation for A	86230.49	6	14371.748	1.477608
Variation for B	86230.49	6	14371.748	1.477608
Variation for AB	838984.08	36	23305.113	2.396077
Variation for Error	953183.40	98	9726.361	NA
Variation for Total	1964628.46	110	NA	NA

```r
# Find the critical value
a_c = 7
b_c = 7
n_c = 1
df_AB_c = (a_c - 1) * (b_c - 1)
df_Err_c = a_c * b_c * (n_c + 1)
alpha_s = 1 - 0.95
qf(alpha_s, df_AB_c, df_Err_c, lower.tail = FALSE)
```

```
## [1] 1.53753
```