Data Trial

2025-09-03

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
data22 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_22.csv")
data23 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_23.csv")
data24 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_24.csv")

datacombined <- rbind(data22,data23,data24)

   # e.g., pitches <- read.csv("file.csv")
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
```

```r
library(tidyr)
library(ggplot2)

# --- Clean data ---
# pitches <- datacombined %>%
#   mutate(
#     ExitSpeed  = as.numeric(na_if(ExitSpeed,  "NULL")),
#     VExitAngle = as.numeric(na_if(VExitAngle, "NULL")),
#     HExitAngle = as.numeric(na_if(HExitAngle, "NULL"))
#   ) %>%
#   filter(!is.na(eventtype))  # drop rows with missing event type
#
# # --- Helper to build one plot for a given metric ---
# plot_mean_var <- function(df, metric, nice_label = metric) {
#   summ <- df %>%
```

```
#     filter(!is.na(.data[[metric]])) %>%
#     group_by(eventtype) %>%
#     filter(n() >= 1000) %>%                 # keep only eventtypes with >= 1000 rows
#     summarise(
#       mean = mean(.data[[metric]], na.rm = TRUE),
#       var  = var(.data[[metric]],  na.rm = TRUE),
#       .groups = "drop"
#     ) %>%
#     pivot_longer(cols = c(mean, var), names_to = "stat", values_to = "value")
#
#   ggplot(summ, aes(x = eventtype, y = value, fill = stat)) +
#     geom_col(position = "dodge") +
#     labs(
#       title = paste0(nice_label, ": mean & variance by eventtype (>=1000)"),
#       x = "Event Type", y = nice_label, fill = ""
#     ) +
#     scale_fill_manual(values = c(mean = "orange", var = "steelblue")) +
#     theme_minimal(base_size = 12) +
#     theme(axis.text.x = element_text(angle = 30, hjust = 1))
# }
#
# # --- Make the three separate plots ---
# p_vert  <- plot_mean_var(pitches, "VExitAngle", "VExitAngle")
# p_exit  <- plot_mean_var(pitches, "ExitSpeed",  "ExitSpeed")
# p_hexit <- plot_mean_var(pitches, "HExitAngle", "HExitAngle")
#
# # --- Print them ---
# p_vert
# p_exit
# p_hexit
```

```
RHH_GIDP_Clusters <- read.csv("RHH_GIDP_Clusters.csv")
View(RHH_GIDP_Clusters)
kruskal.test(VExitAngle ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  VExitAngle by as.factor(Cluster)
## Kruskal-Wallis chi-squared = 2900, df = 5, p-value < 2.2e-16
```

```
kruskal.test(ExitSpeed ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  ExitSpeed by as.factor(Cluster)
## Kruskal-Wallis chi-squared = 3828.9, df = 5, p-value < 2.2e-16
```

```
kruskal.test(HExitAngle ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
##
```

```
##  Kruskal-Wallis rank sum test
##
## data:  HExitAngle by as.factor(Cluster)
## Kruskal-Wallis chi-squared = 5487.1, df = 5, p-value < 2.2e-16
```

```r
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
leveneTest(VExitAngle ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    5    45.1 < 2.2e-16 ***
##       6742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
leveneTest(ExitSpeed ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    5  93.507 < 2.2e-16 ***
##       6742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
leveneTest(HExitAngle ~ as.factor(Cluster), data = RHH_GIDP_Clusters)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    5  123.53 < 2.2e-16 ***
##       6742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
pairwise.wilcox.test(RHH_GIDP_Clusters$VExitAngle,
                     RHH_GIDP_Clusters$Cluster,
                     p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
```

```
## data:  RHH_GIDP_Clusters$VExitAngle and RHH_GIDP_Clusters$Cluster
##
##   0       1       2       3       4
## 1 < 2e-16 -       -       -       -
## 2 < 2e-16 < 2e-16 -       -       -
## 3 3.2e-13 0.0023  < 2e-16 -       -
## 4 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -
## 5 < 2e-16 < 2e-16 < 2e-16 < 2e-16 2.9e-10
##
## P value adjustment method: BH
```

```
pairwise.wilcox.test(RHH_GIDP_Clusters$ExitSpeed,
                     RHH_GIDP_Clusters$Cluster,
                     p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  RHH_GIDP_Clusters$ExitSpeed and RHH_GIDP_Clusters$Cluster
##
##   0       1       2       3       4
## 1 < 2e-16 -       -       -       -
## 2 < 2e-16 < 2e-16 -       -       -
## 3 < 2e-16 2.8e-07 < 2e-16 -       -
## 4 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -
## 5 < 2e-16 < 2e-16 0.00065 < 2e-16 < 2e-16
##
## P value adjustment method: BH
```

```
pairwise.wilcox.test(RHH_GIDP_Clusters$HExitAngle,
                     RHH_GIDP_Clusters$Cluster,
                     p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  RHH_GIDP_Clusters$HExitAngle and RHH_GIDP_Clusters$Cluster
##
##   0       1       2       3       4
## 1 < 2e-16 -       -       -       -
## 2 < 2e-16 < 2e-16 -       -       -
## 3 < 2e-16 < 2e-16 < 2e-16 -       -
## 4 4.5e-12 < 2e-16 < 2e-16 < 2e-16 -
## 5 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16
##
## P value adjustment method: BH
```

```
install.packages("FSA")
```

```
##
## The downloaded binary packages are in
##  /var/folders/mq/7n54_6911q5g2kqjbjzj0kcm0000gq/T//Rtmp8fgLS4/downloaded_packages
```

```r
library(FSA)
```

```
## Registered S3 methods overwritten by 'FSA':
##   method       from
##   confint.boot car
##   hist.boot    car

## ## FSA v0.10.0. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.

##
## Attaching package: 'FSA'

## The following object is masked from 'package:car':
##
##     bootCase
```

```r
dunnTest(VExitAngle ~ Cluster, data = RHH_GIDP_Clusters, method = "bh")
```

```
## Warning: Cluster was coerced to a factor.

## Dunn (1964) Kruskal-Wallis multiple comparison

##   p-values adjusted with the Benjamini-Hochberg method.

##    Comparison          Z       P.unadj         P.adj
## 1       0 - 1  10.717736  8.403514e-27  1.050439e-26
## 2       0 - 2 -11.403136  4.033232e-30  5.499862e-30
## 3       1 - 2 -22.609084 3.527853e-113 5.879754e-113
## 4       0 - 3   6.674455  2.481522e-11  2.863294e-11
## 5       1 - 3  -2.724064  6.448403e-03  6.909003e-03
## 6       2 - 3  17.217407  1.966027e-66  2.949041e-66
## 7       0 - 4  34.248433 4.602126e-257 2.301063e-256
## 8       1 - 4  27.441631 8.744234e-166 2.623270e-165
## 9       2 - 4  43.481870  0.000000e+00  0.000000e+00
## 10      3 - 4  26.450090 3.640101e-154 9.100253e-154
## 11      0 - 5  31.350012 9.725301e-216 3.646988e-215
## 12      1 - 5  25.206991 3.357654e-140 7.194973e-140
## 13      2 - 5  39.444712  0.000000e+00  0.000000e+00
## 14      3 - 5  25.005903 5.273226e-138 9.887298e-138
## 15      4 - 5   2.365294  1.801575e-02  1.801575e-02
```

```r
dunnTest(ExitSpeed  ~ Cluster, data = RHH_GIDP_Clusters, method = "bh")
```

```
## Warning: Cluster was coerced to a factor.

## Dunn (1964) Kruskal-Wallis multiple comparison
##   p-values adjusted with the Benjamini-Hochberg method.
```

```
##    Comparison           Z       P.unadj         P.adj
## 1       0 - 1 -14.7179578  4.943652e-49  7.415478e-49
## 2       0 - 2  31.6328229 1.306574e-219 3.919721e-219
## 3       1 - 2  48.4755665  0.000000e+00  0.000000e+00
## 4       0 - 3  -8.3796633  5.307934e-17  6.634918e-17
## 5       1 - 3   4.5958183  4.310542e-06  4.618438e-06
## 6       2 - 3 -38.0156157 3.186398e-316 1.593199e-315
## 7       0 - 4  20.2562385  3.129794e-91  5.216324e-91
## 8       1 - 4  35.0879049 1.030862e-269 3.865733e-269
## 9       2 - 4  -9.5126387  1.858865e-21  2.534815e-21
## 10      3 - 4  26.9870563 2.096970e-160 4.493508e-160
## 11      0 - 5  25.8451642 2.757556e-147 5.170418e-147
## 12      1 - 5  38.1250823 4.922218e-318 3.691663e-317
## 13      2 - 5   0.4595983  6.458046e-01  6.458046e-01
## 14      3 - 5  31.5248810 3.962662e-218 9.906656e-218
## 15      4 - 5   8.2995546  1.045025e-16  1.205798e-16
```

```r
dunnTest(HExitAngle ~ Cluster, data = RHH_GIDP_Clusters, method = "bh")
```

```
## Warning: Cluster was coerced to a factor.
```

```
## Dunn (1964) Kruskal-Wallis multiple comparison
##   p-values adjusted with the Benjamini-Hochberg method.
```

```
##    Comparison          Z       P.unadj         P.adj
## 1       0 - 1 -34.083118 1.312193e-254 3.280481e-254
## 2       0 - 2 -34.588372 3.778829e-262 1.133649e-261
## 3       1 - 2  -5.403507  6.535044e-08  7.540436e-08
## 4       0 - 3 -58.680049  0.000000e+00  0.000000e+00
## 5       1 - 3 -32.086037 6.904462e-226 1.150744e-225
## 6       2 - 3 -23.669190 7.489791e-124 1.123469e-123
## 7       0 - 4   1.794096  7.279790e-02  7.279790e-02
## 8       1 - 4  32.463439 3.500361e-231 6.563177e-231
## 9       2 - 4  33.481294 9.022741e-246 1.933444e-245
## 10      3 - 4  55.702758  0.000000e+00  0.000000e+00
## 11      0 - 5 -42.550943  0.000000e+00  0.000000e+00
## 12      1 - 5 -19.827316  1.730282e-87  2.359476e-87
## 13      2 - 5 -14.422292  3.746882e-47  4.683603e-47
## 14      3 - 5   5.056457  4.271166e-07  4.576249e-07
## 15      4 - 5 -41.633108  0.000000e+00  0.000000e+00
```