

# Statistical-Analysis.r

55499

2025-10-03

```
#####  
# Load packages  
#####  
library(readxl)  
library(car)  
  
## Loading required package: carData  
  
library(dplyr)  
  
##  
## Attaching package: 'dplyr'  
  
## The following object is masked from 'package:car':  
##  
##      recode  
  
## The following objects are masked from 'package:stats':  
##  
##      filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union  
  
library(lme4)  
  
## Loading required package: Matrix  
  
library(emmeans)  
library(redres)  
library(geepack)  
  
## Warning: package 'geepack' was built under R version 4.3.3  
  
library(readr)  
library(tidyr)  
  
## Warning: package 'tidyr' was built under R version 4.3.3  
  
##  
## Attaching package: 'tidyr'
```

```

## The following objects are masked from 'package:Matrix':
##
##      expand, pack, unpack

library(openxlsx)

## Warning: package 'openxlsx' was built under R version 4.3.3

library(afex)

## Warning: package 'afex' was built under R version 4.3.3

## *****
## Welcome to afex. For support visit: http://afex.singmann.science/

## - Functions for ANOVAs: aov_car(), aov_ez(), and aov_4()
## - Methods for calculating p-values with mixed(): 'S', 'KR', 'LRT', and
  'PB'
## - 'afex_aov' and 'mixed' objects can be passed to emmeans() for
  follow-up tests
## - Get and set global package options with: afex_options()
## - Set sum-to-zero contrasts globally: set_sum_contrasts()
## - For example analyses see: browseVignettes("afex")
## *****

##
## Attaching package: 'afex'

## The following object is masked from 'package:lme4':
##
##      lmer

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

library(broom)
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##      select

```

```
##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##      geyser

library(lsmmeans)

## The 'lsmmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmmeans' objects and scripts to work with 'emmeans'.

library(effectsize)

## Warning: package 'effectsize' was built under R version 4.3.3

##
## Attaching package: 'effectsize'

## The following object is masked from 'package:mvtnorm':
##
##      standardize

library(tidyverse)

## Warning: package 'stringr' was built under R version 4.3.3

## — Attaching core tidyverse packages —————
tidyverse 2.0.0 —
## ✓ forcats   1.0.0      ✓ stringr   1.5.1
## ✓ lubridate 1.9.3      ✓ tibble    3.2.1
## ✓ purrr     1.0.2

## — Conflicts —————
tidyverse_conflicts() —
## ✗ tidyr::expand() masks Matrix::expand()
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ✗ tidyr::pack()   masks Matrix::pack()
## ✗ dplyr::recode() masks car::recode()
## ✗ MASS::select() masks dplyr::select()
## ✗ purrr::some()   masks car::some()
## ✗ tidyr::unpack() masks Matrix::unpack()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

library(ggpubr)
library(Factoshiny)
```

```

## Warning: package 'Factoshiny' was built under R version 4.3.3
## Loading required package: FactoMineR
## Warning: package 'FactoMineR' was built under R version 4.3.3
## Loading required package: shiny
## Loading required package: FactoInvestigate
## Warning: package 'FactoInvestigate' was built under R version 4.3.3
library(gridExtra)

##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine

library(simr)

## Warning: package 'simr' was built under R version 4.3.3
##
## Attaching package: 'simr'
##
## The following object is masked from 'package:stringr':
##
##     fixed
##
## The following object is masked from 'package:lme4':
##
##     getData

library(ggstatsplot)

## Warning: package 'ggstatsplot' was built under R version 4.3.3
## You can cite this package as:
##     Patil, I. (2021). Visualizations with statistical details: The
'ggstatsplot' approach.
##     Journal of Open Source Software, 6(61), 3167,
doi:10.21105/joss.03167

#####
# Load dataset
#####
df <- read_excel("DB Womens Soccer.xlsx", sheet = "gameData")
df$idNumeric <- as.numeric(as.factor(df$Athlete)) # create numeric
athlete ID

#####

```

```

# Split by halves and full match
#####
fullMatch <- df %>%
  filter(MatchPeriod == "Full") %>%
  arrange(idNumeric)

firstHalf <- df %>%
  filter(MatchPeriod == "1stHalf") %>%
  arrange(idNumeric)

secondHalf <- df %>%
  filter(MatchPeriod == "2ndHalf") %>%
  arrange(idNumeric)

#####
# Contextual categorical variables
#####
fullMatch <- fullMatch %>%
  mutate(MatchRecovery = case_when(
    DaysBetweenMatches <= 4 ~ "<=4 days",
    DaysBetweenMatches >= 5 & DaysBetweenMatches <= 7 ~ "5-7 days",
    DaysBetweenMatches >= 8 ~ ">=8 days"
  ))

fullMatch <- fullMatch %>%
  mutate(PointsDiffCat = case_when(
    PointDif < 3 ~ "Small",
    PointDif >= 3 & PointDif < 8 ~ "Moderate",
    PointDif >= 8 ~ "Large"
  ))

#####
# Descriptive statistics function
#####
descriptive_stats <- function(variable) {
  fullMatch %>%
    mutate(
      Position = factor(Position,
        levels = c("zagueira", "lateral", "meio-
campista", "atacante")),
      ReportCondition = factor(ReportCondition,
        levels = c("90+", "45+", "5+"))
    ) %>%
    group_by(ReportCondition) %>%
    summarise(
      Mean = mean(.data[[variable]], na.rm = TRUE),
      SD = sd(.data[[variable]], na.rm = TRUE),
      .groups = "drop"
    ) %>%
    arrange(ReportCondition) %>%

```

```

write_csv(paste0("summary_", variable, ".csv"))
}

#####
# Generalized Estimation Equations (GEE)
#####

dependentVars = c("DistanceRel") # Example: set dependent variables

for (depVar in dependentVars) {
  indepVar <- "PointsDiffCat" # Example: set independent variable

  formula <- as.formula(paste(depVar, " ~ Group * ", indepVar))

  # Exclude NAs
  data_temp <- fullMatch[!is.na(fullMatch[[indepVar]]), ]

  gee_model <- geeglm(
    formula,
    id = idNumeric,
    family = gaussian,
    constr = "independence",
    data = data_temp
  )

  print(anova(gee_model))

  # Post-hoc comparisons
  efGroup <- lsmeans(gee_model, "Group", adjust = "tukey")
  print(pairs(efGroup))

  efVar <- lsmeans(gee_model, indepVar, adjust = "tukey")
  print(pairs(efVar))

  formula_inter <- as.formula(paste("~ Group | ", indepVar))
  EF_inter <- lsmeans(gee_model, formula_inter, adjust = "tukey")
  print(pairs(EF_inter, by = "Group"))
  print(pairs(EF_inter, by = indepVar))
}

## Analysis of 'Wald statistic' Table
## Model: gaussian, link: identity
## Response: DistanceRel
## Terms added sequentially (first to last)
##
##              Df      X2 P(>|Chi|)
## Group          1 12.7083 0.0003640 ***
## PointsDiffCat   2 17.6538 0.0001467 ***
## Group:PointsDiffCat 2  1.1248 0.5698444

```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## NOTE: Results may be misleading due to involvement in interactions

## contrast      estimate    SE  df z.ratio p.value
## Reserva - Titular      3.8 1.08 Inf   3.525  0.0004
##
## Results are averaged over the levels of: PointsDiffCat

## NOTE: Results may be misleading due to involvement in interactions

## contrast      estimate    SE  df z.ratio p.value
## Large - Moderate    0.427 0.910 Inf   0.469  0.8860
## Large - Small       2.824 0.999 Inf   2.826  0.0131
## Moderate - Small    2.397 0.632 Inf   3.794  0.0004
##
## Results are averaged over the levels of: Group
## P value adjustment: tukey method for comparing a family of 3 estimates
## Group = Reserva:
## contrast      estimate    SE  df z.ratio p.value
## Large - Moderate   -0.152 1.196 Inf  -0.127  0.9911
## Large - Small      2.863 1.374 Inf   2.084  0.0931
## Moderate - Small    3.016 1.183 Inf   2.550  0.0290
##
## Group = Titular:
## contrast      estimate    SE  df z.ratio p.value
## Large - Moderate    1.006 1.118 Inf   0.900  0.6405
## Large - Small      2.784 1.288 Inf   2.160  0.0781
## Moderate - Small    1.778 0.656 Inf   2.709  0.0185
##
## P value adjustment: tukey method for comparing a family of 3 estimates
## PointsDiffCat = Large:
## contrast      estimate    SE  df z.ratio p.value
## Reserva - Titular    3.44 1.38 Inf   2.489  0.0128
##
## PointsDiffCat = Moderate:
## contrast      estimate    SE  df z.ratio p.value
## Reserva - Titular    4.60 1.30 Inf   3.542  0.0004
##
## PointsDiffCat = Small:
## contrast      estimate    SE  df z.ratio p.value
## Reserva - Titular    3.36 1.51 Inf   2.220  0.0264

#####
# Absolute descriptive summaries
#####
summaryAbs <- fullMatch %>%
  group_by(Group) %>%
  summarise(
    MeanTotalDist = mean(Distance, na.rm = TRUE),

```

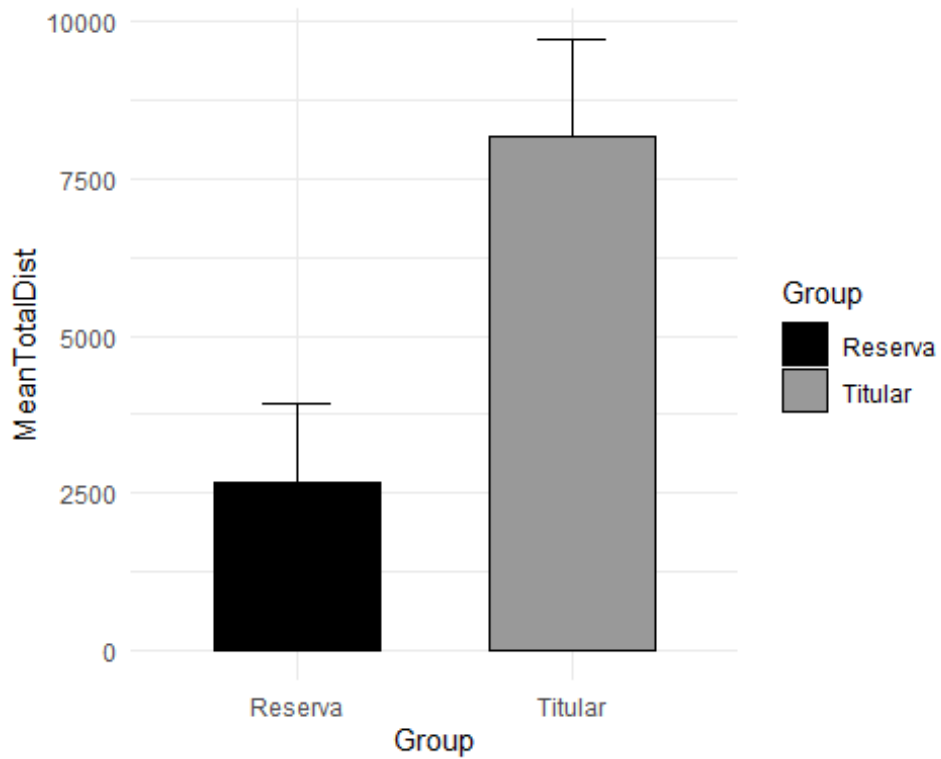
```

SDTotalDist = sd(Distance, na.rm = TRUE),
MeanZone1 = mean(Dist0to7, na.rm = TRUE),
SDZone1 = sd(Dist0to7, na.rm = TRUE),
MeanZone2 = mean(Dist7to13, na.rm = TRUE),
SDZone2 = sd(Dist7to13, na.rm = TRUE),
MeanZone3 = mean(Dist13to19, na.rm = TRUE),
SDZone3 = sd(Dist13to19, na.rm = TRUE),
MeanZone4 = mean(Dist19to23, na.rm = TRUE),
SDZone4 = sd(Dist19to23, na.rm = TRUE),
MeanZone5 = mean(DistAbove23, na.rm = TRUE),
SDZone5 = sd(DistAbove23, na.rm = TRUE),
MeanZone5Effort = mean(HighEfforts23, na.rm = TRUE),
SDZone5Effort = sd(HighEfforts23, na.rm = TRUE),
MeanPlayerLoad = mean(PlayerLoad, na.rm = TRUE),
SDPlayerLoad = sd(PlayerLoad, na.rm = TRUE),
MeanAcc = mean(ACCACIMA2, na.rm = TRUE),
SDAcc = sd(ACCACIMA2, na.rm = TRUE),
MeanDec = mean(DCCACIMA2, na.rm = TRUE),
SDDec = sd(DCCACIMA2, na.rm = TRUE)
)

#####
# Example plots
#####
ggplot(summaryAbs, aes(x = Group, y = MeanTotalDist, fill = Group)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.85),
width = 0.6, color = "black") +
  geom_errorbar(aes(ymin = MeanTotalDist, ymax = MeanTotalDist +
SDTotalDist),
                position = position_dodge(width = 0.8), width = 0.25) +
  scale_fill_manual(values = c("black", "gray60")) +
  theme_minimal()

```





```
# Boxplot function
create_boxplot <- function(y_var, y_label, title, show_legend = TRUE) {
  ggplot(fullMatch, aes(x = Position, y = !!sym(y_var), fill = Group)) +
    geom_boxplot() +
    labs(y = y_label, title = title) +
    scale_fill_manual(values = c("Starter" = "#000000", "Non-starter" =
"gray60")) +
    theme_minimal() +
    guides(fill = if (show_legend) guide_legend(title = NULL) else
"none")
}
```