Nonparametric Tests: Methods, Tests, and Effect Sizes

1. Nonparametric Tests: Design × Test × Effect Size

Design Type	Overall Test	Test Statistic	Overall Effect Size	Post Hoc Test	Pairwise Effect Size
Two Independent samples	Mann-Whitney U Test (Wilcoxon Rank-Sum Test)	U (or W)	Rank-biserial correlation(r_{rb})	-	-
Multiple Independent Samples	Kruskal-Wallis Test	H (χ^2 approximation)	Epsilon squared (ε^2, η^2[H])	Pairwise Mann – Whitney U tests	Rank-biserial correlation (r_{rb})
Two paired samples	Wilcoxon Signed-Rank Test	V (or Z)	Rank-biserial correlation (r_{rb})	-	-
Multiple paired samples	Friedman Test	χ^2_F	Kendall's W	Pairwise Wilcoxon signed-rank tests	Rank-biserial correlation (r_{rb})

2. Effect Size Definitions

2.1 Rank-Biserial Correlation (r_{rb})

Used for: Mann – Whitney U test, Wilcoxon signed-rank test.

Range: −1 to 1, Interpretation: 0 = complete overlap, ± 1 = complete dominance.

Magnitude guidelines : Small: $|r| \approx 0.10$, Moderate: $|r| \approx 0.30$, Large: $|r| \geqslant 0.50$.

2.2 Epsilon Squared (ε^2)

Used for: Kruskal – Wallis test.

Range: 0 to 1, Meaning: Proportion of rank variance explained by group membership.

Magnitude guidelines: Small: $\varepsilon^2 \approx 0.01$, Moderate: $\varepsilon^2 \approx 0.06$, Large: $\varepsilon^2 \geqslant 0.14$.

2.3 Kendall's W

Used for: Friedman test.

Range: 0 to 1, Meaning: Degree of agreement / consistency across conditions.

Interpretation: no systematic differences: W = 0 , perfect agreement: W = 1.

3. Table-Note

3.1 Two Independent Samples

A Mann – Whitney U test was conducted to compare differences between two independent groups. Effect size was quantified using the rank-biserial correlation (rrb) with 95% confidence intervals, calculated using the effectsize package in R.

3.2 Multiple Independent Samples

A Kruskal – Wallis test was performed to examine group differences across multiple independent samples. The overall effect size was estimated using epsilon squared (ε^2). Post hoc pairwise comparisons were conducted using Mann – Whitney U tests with Holm adjustment for multiple testing. Rank-biserial correlations with 95% confidence intervals were reported for all pairwise comparisons. All effect sizes were computed using the effectsize package in R.

3.3 Two Paired Samples

A Wilcoxon signed-rank test was conducted to assess differences between two paired conditions. Effect size was quantified using the rank-biserial correlation (rrb) with 95% confidence intervals, computed using the effectsize package in R.

3.4 Multiple Paired Samples

A Friedman test was conducted to examine differences across repeated-measures conditions.

Kendall’s W was reported as the overall effect size. Post hoc pairwise comparisons were performed using Wilcoxon signed-rank tests with Holm adjustment. Rank-biserial correlations with 95% confidence intervals were calculated for all pairwise comparisons using the effectsize package in R.

Appendix A

R-Script for Independent-Sample

```
# ------------------------------------------------------------
# Load required R packages
# ------------------------------------------------------------
library(effectsize)   # For effect size estimation and confidence intervals
library(rstatix)        # For nonparametric tests and post hoc comparisons
library(dplyr)            # For data manipulation (data wrangling)

############################################################
# PART 1: Two Independent Samples
# Nonparametric Test: Mann – Whitney U Test
# (also known as Wilcoxon Rank-Sum Test)
############################################################

# ------------------------------------------------------------
# Prepare dataset for two independent groups
# Dependent variable: Sepal.Length
# Grouping variable: Species (two independent groups)
# ------------------------------------------------------------
data_two <- iris %>%
  filter(Species %in% c("setosa", "versicolor")) %>%
  droplevels()   # Remove unused factor levels

# ------------------------------------------------------------
# Mann – Whitney U Test
# Purpose:
#     Test whether the distributions of two independent groups differ
# ------------------------------------------------------------
cat("=== Mann – Whitney U Test Results ===\n")
```

```
mw_test <- wilcox.test(
  Sepal.Length ~ Species,
  data = data_two,
  exact = FALSE,          # Normal approximation (large samples)
  conf.int = TRUE,       # Confidence interval for location shift
  conf.level = 0.95
)

# ------------------------------------------------------------
# Effect size for Mann – Whitney U Test
# Effect size measure: Rank-biserial correlation (independent)
# ------------------------------------------------------------
effect_two <- rank_biserial(
  Sepal.Length ~ Species,
  data = data_two,
  ci = 0.95,
  verbose = FALSE
)

# Effect size name: Rank-biserial correlation (r_rb)
# Range: −1 to 1
# Interpretation:
#    r = 0 → complete overlap between groups
# r = ±1 → complete separation between groups
# Meaning:
#    Difference in probability that a randomly chosen observation from one group exceeds the other
```

```
############################################################
# PART 2: Multiple Independent Samples
# Nonparametric Test: Kruskal – Wallis Test
############################################################

# ------------------------------------------------------------
# Prepare dataset for multiple independent groups
# Dependent variable: Sepal.Length
# Grouping variable: Species (three independent groups)
# ------------------------------------------------------------
data_multi <- iris %>%
  select(Sepal.Length, Species) %>%
  droplevels()

# ------------------------------------------------------------
# Kruskal – Wallis Test
# Purpose:
#    Test whether at least one group distribution differs
# ------------------------------------------------------------
cat("=== Kruskal – Wallis Test Results ===\n")

kw_test <- kruskal.test(
  Sepal.Length ~ Species,
  data = data_multi
)

print(kw_test)

# ------------------------------------------------------------
# Overall effect size for Kruskal – Wallis Test
```

```
# Effect size measure: Epsilon squared (ε²)
# -----------------------------------------------------------
kw_effectsize <- kruskal_effsize(
  data = data_multi,
  Sepal.Length ~ Species,
  ci = 0.95
)

print(kw_effectsize)

# Effect size name: Epsilon squared (ε²)
# Often reported as η²[H] or ε²
# Range: 0 to 1
# Interpretation (common guidelines):
#     ε² ≈ 0.01 → small
#  ε² ≈ 0.06 → moderate
#  ε² ≥ 0.14 → large
# Meaning:
#    Proportion of rank variance explained by group membership

# -----------------------------------------------------------
# Post hoc pairwise comparisons
# Test: Mann–Whitney U (Wilcoxon rank-sum)
# Multiple comparison correction: Holm adjustment
# -----------------------------------------------------------
cat("\n=== Post hoc Pairwise Wilcoxon Tests (Holm-adjusted) ===\n")

pairwise_test <- data_multi %>%
  pairwise_wilcox_test(
    Sepal.Length ~ Species,
```

```
    p.adjust.method = "holm"
  )

print(pairwise_test)

# ------------------------------------------------------------
# Effect sizes for pairwise comparisons
# Effect size measure: Rank-biserial correlation (independent)
# ------------------------------------------------------------

# Generate all pairwise combinations of groups
group_pairs <- combn(levels(data_multi$Species), 2, simplify = FALSE)

pairwise_effectsize <- lapply(group_pairs, function(g) {

  # Extract observations for each independent group
  x <- data_multi$Sepal.Length[data_multi$Species == g[1]]
  y <- data_multi$Sepal.Length[data_multi$Species == g[2]]

  # Compute rank-biserial correlation
  es <- rank_biserial(
    x, y,
    ci = 0.95,
    verbose = FALSE
  )

  es_df <- as.data.frame(es)

  # Assemble results
  data.frame(
```

```
      group1 = g[1],
      group2 = g[2],
      rank_biserial = es_df[[1]],
      CI_low = es_df$CI_low,
      CI_high = es_df$CI_high
    )
}) %>%
  bind_rows()

pairwise_effectsize

# Effect size name: Rank-biserial correlation (independent)
# Interpretation:
#     |r| ≈ 0.10 → small
# |r| ≈ 0.30 → moderate
# |r| ≥ 0.50 → large
# CI interpretation:
#     Narrow CI → precise dominance estimate
# CI including 0 → weak or ambiguous group difference
```

Appendix B

R-Script for Paired-Sample

```
# ------------------------------------------------------------
# Load required R packages
# ------------------------------------------------------------
library(effectsize)    # For effect size estimation and confidence intervals
library(rstatix)         # For nonparametric tests and post hoc procedures
library(dplyr)              # For data manipulation (data wrangling)
library(tidyr)              # For reshaping data (wide <-> long)

############################################################
# PART 1: Two Paired Samples
# Nonparametric Test: Wilcoxon Signed-Rank Test
############################################################

# ------------------------------------------------------------
# Read paired-sample dataset (wide format)
# Each row represents one subject
# Time2 and Time3 are two paired measurements
# ------------------------------------------------------------
data_two <- read.csv(
  "/Users/apple/Documents/教师/20-教学笔记/多元统计/假设检验/非参数检验-两个配对样本
-案例数据集.csv"
)

# Select the two paired variables
data_two <- data_two %>%
  select(Time2, Time3)

# ------------------------------------------------------------
```

```r
# Wilcoxon Signed-Rank Test
# Purpose:
#     Test whether the median of paired differences equals zero
# ------------------------------------------------------------
w_test <- wilcox.test(
  data_two$Time2,
  data_two$Time3,
  paired = TRUE,          # Paired samples
  exact = FALSE,          # Normal approximation (large sample)
  conf.int = TRUE,        # Confidence interval for location shift
  conf.level = 0.95
)

print(w_test)


# ------------------------------------------------------------
# Effect size for paired Wilcoxon test
# Effect size measure: Rank-biserial correlation (paired version)
# ------------------------------------------------------------
effect_two <- rank_biserial(
  data_two$Time2,
  data_two$Time3,
  paired = TRUE,          # Paired-sample version
  ci = 0.95,              # 95% confidence interval
  verbose = FALSE
)

print(effect_two)


# Effect size name: Rank-biserial correlation (r_rb, paired)
```

```
# Interpretation:
#      r =  ± 1 → all paired differences in the same direction
# r = 0 → symmetric differences
# Advantage:
#      Nonparametric
# Distribution-free
# Directly interpretable as dominance probability



############################################################
# PART 2: Multiple Paired Samples
# Nonparametric Test: Friedman Test
############################################################

# ------------------------------------------------------------
# Read wide-format repeated-measures dataset
# Each row = one subject
# Each column (Time1, Time2, Time3) = one condition
# ------------------------------------------------------------
data_wide <- read.csv(
  "/Users/apple/Documents/教师/20-教学笔记/多元统计/假设检验/非参数检验-多个配对样本
-iris.csv"
)

# Select repeated-measures variables
data_wide <- data_wide %>%
  select(Time1, Time2, Time3)

# ------------------------------------------------------------
```

```
# Convert wide format to long format (required for Friedman test)

# ID = subject identifier

# ------------------------------------------------------------

data_long <- data_wide %>%

  mutate(ID = row_number()) %>%          # Subject ID

  pivot_longer(

    cols = starts_with("Time"),

    names_to = "Time",                   # Within-subject factor

    values_to = "Value"                  # Measurement

  )


# ------------------------------------------------------------

# Friedman Test

# Purpose:

#    Test whether distributions differ across ≥3 paired conditions

# ------------------------------------------------------------

cat("=== Friedman Test Results ===\n")


friedman_test <- friedman.test(

  Value ~ Time | ID,

  data = data_long

)


print(friedman_test)


# ------------------------------------------------------------

# Overall effect size for Friedman test

# Effect size measure: Kendall's W

# ------------------------------------------------------------

cat("\n=== Overall Effect Size: Kendall's W ===\n")
```

```
friedman_es <- kendalls_w(
  Value ~ Time | ID,
  data = data_long,
  ci = 0.95
)

print(friedman_es)

# ------------------------------------------------------------
# Post hoc pairwise comparisons
# Test: Wilcoxon signed-rank test (paired)
# P-value adjustment: Holm correction
# ------------------------------------------------------------
cat("\n=== Post hoc Pairwise Wilcoxon Tests (Holm-adjusted) ===\n")

pairwise_test <- data_long %>%
  pairwise_wilcox_test(
    Value ~ Time,
    paired = TRUE,
    p.adjust.method = "holm"    # Holm or Bonferroni
  )

print(pairwise_test)

# ------------------------------------------------------------
# Effect sizes for pairwise comparisons
# Effect size measure: Rank-biserial correlation (paired)
# ------------------------------------------------------------
```

```r
# Generate all pairwise combinations of conditions
time_levels <- unique(data_long$Time)
time_pairs <- combn(time_levels, 2, simplify = FALSE)

pairwise_effectsize <- lapply(time_pairs, function(tp) {

  # Extract paired observations for each condition
  x <- data_long %>% filter(Time == tp[1]) %>% pull(Value)
  y <- data_long %>% filter(Time == tp[2]) %>% pull(Value)

  # Compute paired rank-biserial correlation
  es <- rank_biserial(
    x, y,
    paired = TRUE,
    ci = 0.95,
    verbose = FALSE
  )

  es_df <- as.data.frame(es)

  # Assemble results
  data.frame(
    time1 = tp[1],
    time2 = tp[2],
    rank_biserial = es_df[[1]],
    CI_low = es_df$CI_low,
    CI_high = es_df$CI_high
  )
}) %>%
  bind_rows()
```

pairwise_effectsize

Effect size name: Rank-biserial correlation (paired)

Interpretation:

|r| ≈ 0.10 → small

|r| ≈ 0.30 → moderate

|r| ≥ 0.50 → large

CI meaning:

Reflects uncertainty of dominance probability

CI = [−1, −1] or [1, 1] indicates perfect separation