

Statistical Analysis of Nodal Network Structures

2025-03-22

<https://github.com/stevehaworth02/uchicago>

Dataset Information

Let's read in two datasets containing graph theory measures of patients who were sleeping while undergoing EEG. The target variable is Dreams, a binary (0,1) value that indicates if a patient reported a conscious experience. The data is messy, let's read it in, inspect it, and do some pre-processing. We will need a non-parametric test to compare the two groups, assumption violations are showcased on GitHub.

```
# https://github.com/stevehaworth02/uchicago
suppressPackageStartupMessages(library(readr))
df1 <- read_csv("C:/Users/User/Desktop/group1.csv", show_col_types = FALSE)
df2 <- read_csv("C:/Users/User/Desktop/group2.csv", show_col_types = FALSE)
#head(df1, 3);
head(df2, 3)
```

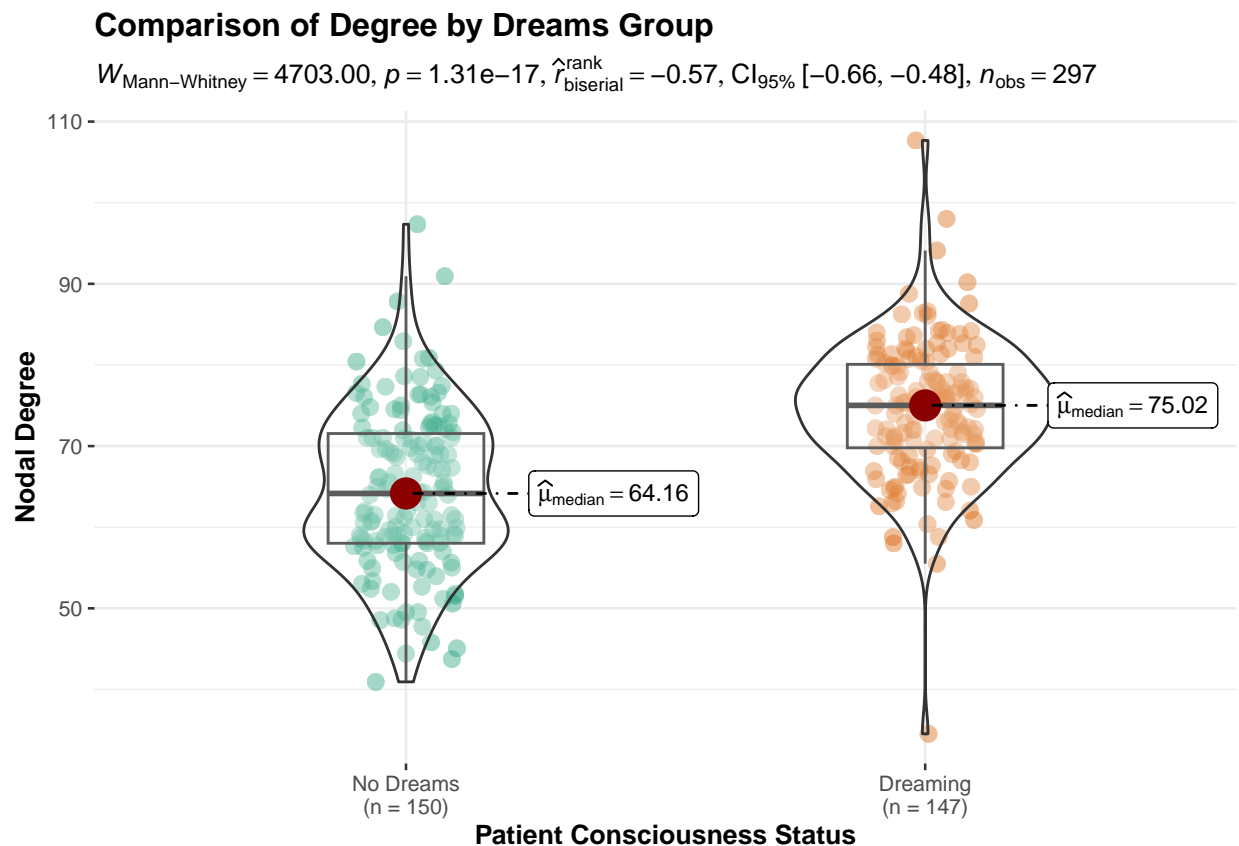
```
## # A tibble: 3 x 10
##   Subject Band threshold wGlobEff wTrans mean_wBetw bTrans mean_bBetw Dreams
##   <chr>    <chr>      <dbl>    <dbl>  <dbl>      <dbl>  <dbl>      <dbl>  <dbl>
## 1 sub_2012 delta      0.202    0.125  0.105      58.3   0.439      151.     0
## 2 sub_2010 beta       0.181    0.164  0.125      132.   0.309      201.     0
## 3 sub_2010 alpha      0.294    NA     0.151      47.8   0.302      233.     0
## # i 1 more variable: degree <dbl>
```

```
process_combined_data <- function(df1, df2) {
  # Combine both data frames row-wise
  combined <- rbind(df1, df2)
  # Number of rows before dropping NA values
  n_before <- nrow(combined)
  # Remove rows with any NA values
  combined_clean <- tidyr::drop_na(combined)
  # Number of rows after dropping NA values
  n_after <- nrow(combined_clean)
  # Print the total number of rows dropped
  cat("Dropped", n_before - n_after, "rows containing NA values.\n")
  # Convert the Dreams column: 0 -> "No Dreams", 1 -> "Dreaming"
  combined_clean$Dreams <- factor(
    combined_clean$Dreams,
    levels = c(0, 1),
    labels = c("No Dreams", "Dreaming")
  )
  # Create a new column 'obs' to track observation number
```

```
combined_clean$obs <- seq_len(nrow(combined_clean))
# Return the cleaned, combined dataset
return(combined_clean)
}
# Example usage:
processed_data <- process_combined_data(df1, df2)
```

Dropped 103 rows containing NA values.

```
suppressPackageStartupMessages(library(ggstatsplot))
ggbetweenstats(
  data = processed_data,
  x = Dreams,          # 'Dreams' is a factor that indicates group membership
  y = degree,          # 'degree' is the graph measure to compare the two groups
  type = "nonparametric", # Nonparametric test is selected
  bf.message = FALSE,   # set to TRUE for postier probabilities
  xlab = "Patient Consciousness Status",
  ylab = "Nodal Degree",
  title = "Comparison of Degree by Dreams Group",
  messages = FALSE
)
```



Patients who reported dreaming had significantly lower nodal degree (a network connectivity measure) compared to those who did not, with a clinically meaningful effect, as evidenced by a large effect size ($r = -0.57$), non-overlapping confidence intervals (CI: $[-0.66, -0.48]$), and extreme statistical significance ($p < 0.05$)