Calculating D-Scores for Target and Control Conditions

This document demonstrates how to calculate D-scores using a formula for target and control conditions. We will use the following formula for both target and control D-scores with different inputs.

D-scores are calculated by including the number of productions, either target or control words, in the last 3 sessions.

Target_dscore_entry1 = Number of target words produced by the participant in the third-to-last session as part of the treatment study

Target_dscore_entry2 = Number of target words produced by the participant in the second-to-last session as part of the treatment study

Target_dscore_entry3 = Number of target words produced by the participant in the last session as part of the treatment study

control_dscore_entry1 = Number of control words produced by the participant in the third-to-last session as part of the treatment study

control_dscore_entry2 = Number of control words produced by the participant in the second-to-last session as part of the treatment study

control_dscore_entry3 = Number of control words produced by the participant in the last session as part of the treatment study

Target D-Score Formula:
if((stdev([target_dscore_entry1],[target_dscore_entry2],[target_dscore_entry3])) > 0,
((mean([target_dscore_entry1], [target_dscore_entry2], [target_dscore_entry3]))
/(stdev([target_dscore_entry1], [target_dscore_entry2], [target_dscore_entry3]))),
(((mean([target_dscore_entry1],[target_dscore_entry2], [target_dscore_entry3]))/0.577)))

Control D-Score Formula:
if((stdev([control_dscore_entry1],[control_dscore_entry2],[control_dscore_entry3])) > 0,
((mean([control_dscore_entry1], [control_dscore_entry2], [control_dscore_entry3]))
/(stdev([control_dscore_entry1], [control_dscore_entry2], [control_dscore_entry3]))),
(((mean([control_dscore_entry1],[control_dscore_entry2],[control_dscore_entry3]))/0.577)))

Target D-Score Example Calculation

Using the values:
[target_dscore_entry1] = 1, [target_dscore_entry2] = 1, [target_dscore_entry3] = 3

Step 1: Calculate the mean:
mean([1, 1, 3]) = (1 + 1 + 3) / 3 = 5 / 3 = 1.6666

Step 2: Calculate the standard deviation:
stdev([1, 1, 3]) = sqrt(((1-1.67)^2 + (1-1.67)^2 + (3-1.67)^2) / 2) = 1.1547

Step 3: Apply the formula:
Since stdev > 0, we use:
target_dscore = mean / stdev = 1.6666 / 1.1547 = 1.4433

Control D-Score Example Calculation

Using the values:
[control_dscore_entry1] = 0, [control_dscore_entry2] = 0, [control_dscore_entry3] = 0

Step 1: Calculate the mean:
mean([0, 0, 0]) = (0 + 0 + 0) / 3 = 0

Step 2: Calculate the standard deviation:
stdev([0, 0, 0]) = 0 (since all values are the same)

Step 3: Apply the formula:
Since stdev = 0, we use the alternative formula:
control_dscore = mean / 0.577 = 0 / 0.577 = 0

Data Transformation Guide

Kimberly Leon

2024-12-10

Introduction

This document serves as a guide and resource for demonstrating code examples and workflows for this dataset. It is intended to illustrate the process of filtering and transforming to analyze this data in R. The example below shows how to transform the dataset from long format to wide format for a potential analysis focusing on several variables of interest. If you have questions or need further clarification, feel free to reach out at kimberlyleon@arizona.edu

Getting Started

Load libraries

```
# Load the dataset
data <- read.csv("dataset.csv")

# View the structure of the data with glimpse
glimpse(data)  # Use glimpse for a concise overview
```

```
## Rows: 392
## Columns: 79
## $ record_id                          <chr> "S01 (S1_40129)", "S01 (S1_401~
## $ vault_event                        <chr> "Vault", "Pre-delay", "Treatme~
## $ starting_treatment_age             <int> 29, NA, NA, NA, NA, 31, NA, NA~
## $ participant_language_status        <int> 0, NA, NA, NA, NA, 2, NA, NA, ~
## $ participant_ethnicity              <int> 0, NA, NA, NA, NA, 0, NA, NA, ~
## $ participant_ethnicity_spec         <lgl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ participant_sex                    <int> 1, NA, NA, NA, NA, 1, NA, NA, ~
## $ vault_spi_parent_eth               <int> 3, NA, NA, NA, NA, 3, NA, NA, ~
## $ vault_spi_parent_eth_other         <lgl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ participant_race                   <int> 5, NA, NA, NA, NA, 5, NA, NA, ~
## $ participant_race_spec              <lgl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ vault_spi_parent_race              <int> 7, NA, NA, NA, NA, 7, NA, NA, ~
## $ participant_parent_age             <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ vault_spi_parent_gender            <int> 2, NA, NA, NA, NA, 1, NA, NA, ~
## $ vault_spi_parent_gen_other         <lgl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ vault_maternal_education_level     <int> 4, NA, NA, NA, NA, 5, NA, NA, ~
## $ vault_delay_condition              <int> 0, NA, NA, NA, NA, 0, NA, NA, ~
## $ vault_studycondition               <int> 2, NA, NA, NA, NA, 2, NA, NA, ~
## $ vault_phase                        <int> 1, NA, NA, NA, NA, 1, NA, NA, ~
## $ responder_status                   <int> 1, NA, NA, NA, NA, 0, NA, NA, ~
## $ vocab_predelay_pretreatment_calc_e <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ delay_rate_words_learned_e         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ vocab_predelay_pretreatment_calc_r <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ delay_rate_words_learned_r         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
```

```
## $ vocab_predelay_posttreatment_calc_e     <int> NA, NA, 177, NA, NA, NA, NA, 3~
## $ vocab_pretreatment_posttreatment_calc_e <int> NA, NA, 177, NA, NA, NA, NA, 3~
## $ treatment_rate_words_learned_e          <dbl> NA, NA, 16.75, NA, NA, NA, NA,~
## $ vocab_pretreatment_posttreatment_calc_r <int> NA, NA, -17, NA, NA, NA, NA, -~
## $ treatment_rate_words_learned_r          <dbl> NA, NA, -1.61, NA, NA, NA, NA,~
## $ vocab_posttreatment_followup_calc_e     <int> NA, NA, 123, NA, NA, NA, NA, 8~
## $ rate_postttreatment_followup_calc_e     <dbl> NA, NA, 26.11, NA, NA, NA, NA,~
## $ vocab_posttreatment_followup_calc_r     <int> NA, NA, 119, NA, NA, NA, NA, 2~
## $ rate_postttreatment_followup_calc_r     <dbl> NA, NA, 25.27, NA, NA, NA, NA,~
## $ target_dscore                           <dbl> NA, NA, 1.443376, NA, NA, NA, ~
## $ control_dscore                          <dbl> NA, NA, 0.000000, NA, NA, NA, ~
## $ eng_mcdi_norms                          <int> NA, 1, NA, 1, 1, NA, 1, NA, 1,~
## $ eng_mcdi_calc_age                       <int> NA, 29, NA, 32, 33, NA, 31, NA~
## $ eng_mcdi_percentile                     <chr> "", "<5th", "", "10th-15th", "~
## $ mcdi_ext_norms                          <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_calc_age_2                         <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_ext_vocab_total                    <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_ext_vocab_percentile               <chr> "", "", "", "", "", "", "", ""~
## $ eng_mcditotal_express                   <int> NA, 47, NA, 224, 347, NA, 105,~
## $ eng_mcditotal_recep                     <int> NA, 242, NA, 225, 344, NA, 306~
## $ mcdi_total_express_spa                  <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_total_recep_spa                    <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_norms_eng                      <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_norms_spa                      <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_calc_age_eng                   <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_calc_age_spa                   <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_percentile_eng                 <chr> "", "", "", "", "", "", "", ""~
## $ mcdi_bil_percentile_spa                 <chr> "", "", "", "", "", "", "", ""~
## $ mcdi_total_express_bil                  <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_total_recep_bil                    <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_ext_norms                      <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_ext_calc_age                   <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_ext_vocab_total                <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ mcdi_bil_ext_vocab_percentile           <chr> "", "", "", "", "", "", "", ""~
## $ asrs_version                            <chr> "", "", "", "", "", "", "", ""~
## $ asrs_age                                <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ asrs_raw                                <chr> "", "", "", "", "", "", "", ""~
## $ asrs_tscore                             <chr> "", "", "", "", "", "", "", ""~
## $ asrs_percentile                         <chr> "", "", "", "", "", "", "", ""~
## $ asrs_classification                     <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ kabc_version                            <chr> "", "", "", "", "", "", "", ""~
## $ kabc_age                                <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ kabc_sum_scaled                         <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ kabc_nvi_standard                       <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ bayley_age                              <int> 29, NA, NA, NA, NA, 32, NA, NA~
## $ bayley_raw                              <int> 73, NA, NA, NA, NA, 67, NA, NA~
## $ bayley_scaled_score                     <int> 11, NA, NA, NA, NA, 8, NA, NA,~
## $ bayley_comp_standard_score              <int> 105, NA, NA, NA, NA, 90, NA, N~
## $ bayley_percentile                       <int> 63, NA, NA, NA, NA, 25, NA, NA~
## $ bayley_test_version                     <int> 1, NA, NA, NA, NA, 1, NA, NA, ~
## $ evt_version_dropdown                    <int> NA, NA, NA, NA, 2, NA, 2, NA, ~
## $ evt_age                                 <int> NA, NA, NA, NA, 32, NA, 32, NA~
## $ evt_raw                                 <int> NA, NA, NA, NA, 39, NA, 7, NA,~
## $ evt_standard_score                      <int> NA, NA, NA, NA, 120, NA, 77, N~
```

```
## $ evt_percentile                          <chr> "", "", "", "", "91", "", "6",~
```

Initial Data Exploration

```
# Summary of the dataset
summary(data)
```

```
# Check unique values in the vault_event column
unique(data$vault_event)
```

```
## [1] "Vault"                    "Pre-delay"
## [3] "Treatment"                "Post-treatment"
## [5] "Follow-up"                "Immediate Pre-treatment"
```

Transform Data

Let's say we are interested in keeping some exploratory variables but want to compare the MCDI total expressive and receptive vocabulary at Post-treatment. We can filter the data using the vault_event variable to filter variables under the 'Vault' event and 'Post-treatment' event. Then, we can transform it.

```
# Filter and transform the Post-treatment data into wide format
post_treatment_data <- data %>%
  filter(vault_event == "Post-treatment") %>%
  pivot_wider(
    id_cols = c(record_id),
    names_from = vault_event,  # Create wide format columns based on the event
    values_from = eng_mcditotal_express,# Only include the relevant variable
    names_glue = "{.value}_{vault_event}") #Append vault_event to the variable name

# Filter the Vault data for specific columns of interest
vault_data <- data %>%
  filter(vault_event == "Vault") %>%
  select(
    record_id, delay_rate_words_learned_e,
    starting_treatment_age, vault_maternal_education_level,
    participant_race, participant_ethnicity, participant_sex,
    vault_studycondition, responder_status )

# Join the two datasets together on record_id

#This is our dataset of interest
final_data <- vault_data %>%
  left_join(post_treatment_data, by = "record_id")
```

This dataset will now have 1 row per participant, which may be needed for certain analyses.

```
# Preview the transformed data
head(final_data)
```

```
##        record_id delay_rate_words_learned_e starting_treatment_age
## 1 S01 (S1_40129)                         NA                     29
## 2 S02 (S1_40231)                         NA                     31
## 3 S03 (S1_50328)                         NA                     30
## 4 S04 (S1_50429)                         NA                     29
## 5 S05 (S1_40524)                         NA                     26
## 6 S06 (S1_50631)                         NA                     31
```

```
##   vault_maternal_education_level participant_race participant_ethnicity
## 1                              4                5                      0
## 2                              5                5                      0
## 3                              4                5                      0
## 4                              4                5                      0
## 5                              7                5                      1
## 6                              7                5                      1
##   participant_sex vault_studycondition responder_status
## 1               1                    2                1
## 2               1                    2                0
## 3               0                    1                0
## 4               0                    1                1
## 5               1                    1                1
## 6               0                    1                1
##   eng_mcditotal_express_Post-treatment
## 1                                  224
## 2                                  140
## 3                                  121
## 4                                   34
## 5                                  353
## 6                                  121
```

Let's make sure the data is in the correct format.

```
# Check the structure of the final dataset
str(final_data)
```

```
## 'data.frame':    68 obs. of  10 variables:
##  $ record_id                          : chr  "S01 (S1_40129)" "S02 (S1_40231)" "S03 (S1_50328)" "S04
##  $ delay_rate_words_learned_e         : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ starting_treatment_age             : int  29 31 30 29 26 31 41 25 32 44 ...
##  $ vault_maternal_education_level     : int  4 5 4 4 7 7 7 2 1 5 ...
##  $ participant_race                   : int  5 5 5 5 5 5 6 5 5 5 ...
##  $ participant_ethnicity              : int  0 0 0 0 1 1 0 1 1 1 ...
##  $ participant_sex                    : int  1 1 0 0 1 0 0 1 1 0 ...
##  $ vault_studycondition               : int  2 2 1 1 1 1 2 1 2 1 ...
##  $ responder_status                   : int  1 0 0 1 1 1 1 1 1 1 ...
##  $ eng_mcditotal_express_Post-treatment: int  224 140 121 34 353 121 456 53 25 518 ...
```

```
# Save the transformed data to a CSV file
write.csv(final_data, "final_data.csv", row.names = FALSE)
```