

Lab #4 (model)

Jerid Francom

9/12/2021

Contents

1	Read and inspect	1
2	Single variables	2
3	Multiple variables	2
4	Assessment	2
5	References	2

1 Read and inspect

```
cedel <- read_rds(file = "data/derived/cedel_dataset.rds") # read cedel_dataset.rds
glimpse(cedel)
```

```
## Rows: 1,931
## Columns: 7
## $ filename      <fct> EN_WR_6_20_3_1_CJB, EN_WR_7_16_2_6_RM, EN~
## $ sex           <fct> Female, Female, Female, Female, Female, F~
## $ proficiency   <fct> Lower beginner, Lower beginner, Lower beg~
## $ placement_test_score_percent <dbl> 14.0, 16.3, 18.6, 18.6, 20.9, 20.9, 20.9,~
## $ num_tokens    <int> 81, 108, 208, 30, 510, 142, 287, 164, 66,~
## $ num_types     <int> 45, 69, 91, 20, 200, 71, 139, 113, 41, 46~
## $ ttr           <dbl> 0.556, 0.639, 0.438, 0.667, 0.392, 0.500,~
```

In the CEDEL2 dataset (`cedel`) there are 1,931 compositions and seven variables. The variables `filename`, `sex`, and `proficiency` are categorical variables and the variables `placement_test_score_percent`, `num_tokens`, `num_types`, and `ttr` are continuous variables.

- `filename` = The record number for each composition
- `sex` = The biological sex of each L2 Spanish learner
- `proficiency` = A standardized learner level associated with each learner
- `placement_test_score_percent` = The score the learner received on a standardized placement test.
- `num_tokens` = The number of words in each composition
- `num_types` = The number of unique words in each composition.
- `ttr` = The Type/ Token Ratio for each composition

- 2** **Single variables**
- 3** **Multiple variables**
- 4** **Assessment**
- ...
- 5** **References**