Lecture 6 problem set

INSERT YOUR NAME HERE

INSERT DATE

Contents

Required reading and instructions

Required reading

- GW 15.1-15.2
- Optional: GW 15.3-15.5
- Optional: GW 10, GW 20.6-20.7,

General instructions The purpose of this problem set is to familiarize yourself with a new dataset, the National Longitudinal Study of 1972 (NLS-72). NLS is a nationally representative, longitudinal study of 12th graders in 1972 with follow-up surveys throughout their postsecondary years. You will be using the Postsecondary Education Transcript File of the NLS-72, which contains information on transcripts from NLS-72 senior cohort members who reported attending a postsecondary institution after high school.

Load library and data

You'll need to load the `tidyverse`, `haven` and `labelled` libraries in order to load and work with the NLS data. If these packages are not yet installed, then you must install before you load. Install in "console" rather than .Rmd file

- Generic syntax: `install.packages("package_name")`
- Install "haben": `install.packages("haven")`

Note: when we **load** package, name of package is not in quotes; but when we **install** package, name of package is in quotes:

- `install.packages("tidyverse")`
- `library(tidyverse)`

```
library(tidyverse)
library(haven)
library(labelled)
```

```
rm(list = ls()) # remove all objects

nls_crs<- read_dta(file="https://github.com/ozanj/rclass/raw/master/data/nls72/nls72petscrs_v2.dta", en
```

Step 1: Investigate Variables

1. Use `typeof`, `class`, `str`, and `attributes` functions to investigate the following variables: crsgrada, crsgradb, gradtype, crsecred.

Step 2: Create New Variables

1. `crsgrada` is the variable for letter course grades. Create a factor version of the `crsgrada` variable. Hint: knowing what class the variable is currently and investigating the variable using `table()` will be helpful to creating the new factor version. Retain the new factor version variable in the nls_crs dataframe using the variable name `crsgrad_fac`. Check that this new variable is a factor class.

2. Create a numeric course grade version of the `crsgrada_fac` variable named `numgrade` with the following numeric values based on attribute levels from `crsgrada_fac` Hint: use `mutate()` and `recode()`. Retain this new `numgrade` variable.

   - A+= 4; A=4; A-=3.7; B+=3.3; B=3; B-=2.7; C+=2.3; C=2; C-=1.7; D+=1.3; D=1; D-=.7; F=0; E=0; WF=0
   - All other letter grades should have missing values for `numgrade`
   - When recoding to missing use `NA_real_` rather than `NA` due to `recode()` needing a double type/numeric class value to recode and `NA` is a logical)

3. `gradtype` is a labelled class variable for the type of grade given for each course. Retrieve the variable label and value labels for `gradtype`. Get a count of `gradtype` showing the values and the value labels. Now, get another count by filtering for observations associated with "{MISSING}".

4. `crsgradb` is the variable for numerical course grades. There are several issues with this variable. First, missing observations for `crsgradb` are currently `999` and `999.999`. The variable also has values greater than 4 (problematic when the highest possible grade A+ = 4). Create and retain a new `crsgradb_v2` variable that replaces all values greater than 4 for `crsgradb` to `NA` (Hint: you can use the `mutate` and `if_else()` functions to either replace the value to NA or keep the current value of the variable based on whether the expression you specify evaluates to `TRUE` or `FALSE`. See below...

ANSWER PROVIDED FOR YOU

```
table(nls_crs$crsgradb)

nls_crs<- nls_crs %>%
  mutate(crsgradb_v2= ifelse(crsgradb>4, NA, crsgradb))
```

5. `crsecred` is the variable for how many total credits were possible for each course. Missing observations for `crsecred` are currently `999` and `999.999`. Using code similar to Question 5, create and retain a new `crsecred_v2` variable that replaces values of `999` and `999.999` to `NA`.

6. Create a "final" numerical grade variable named `numgrade_v2` that has the numeric values below based on attribute levels from `crsgrada_fac` **when** `gradtype` indicates letter grades were used and crsecred_v2 is not missing. Additionally, `numgrade_v2` should equal the value of `crsgradb_v2` **when** `gradtype` indicates numeric grades were used and crsecred_v2 is not missing. Hint: use `mutate()` and `case_when()`.

   - A+= 4; A=4; A-=3.7; B+=3.3; B=3; B-=2.7; C+=2.3; C=2; C-=1.7; D+=1.3; D=1; D-=.7; F=0; E=0; WF=0