

Lab 10

Stat 427

Fall 2020

Instructions

Complete all questions. To prepare for the randomly collected lab, follow the instructions on the class website to prepare the work for submission. These submission rules will apply to all labs throughout the semester.

Data validation and cleaning

- (1) Read in the dataset `shoes_tracker.csv` from the website (where you found this document). Name the object whatever you want, and *look* through the dataset by either using `head()` or `tail()` or with `View()`
- (2) The variable `product.category` must not be missing and `supplier.country` must be either GB or US. To find this, use some of the validation methods.
- (3) How many observations have missing `Product_Category`?

How many observations have invalid `Supplier_Country`?

-
- (4) Create a table for `Supplier_Name` and `Supplier_ID`. The data should meet the following requirements:
`Supplier_Name` must be 3TOP Sports, or Greenline Sports Ltd.
`Supplier_ID` must be 2963 14682

What invalid data exist for `Supplier_Name` and `Supplier_ID`?

-
- (5) Use all of the validation methods to see how they work. `contents()` and `summary()` are great but make sure to use the other methods too just to see how they work

Combining datasets

- (6) Read in all of the following data files: `mnth72007.csv`, `mnth82007.csv` and `mnth92007.csv`. Create a new dataset by merging the three datasets by the *Custom ID* variable. Create two different datasets, one that is sorted and one that is not sorted.