

p8105_hw1_tz2472

tianwei zhao

9/26/2021

Problem 1

This problem is intended to emphasize variable types and introduce coercion; some awareness of how R treats numeric, character, and factor variables is necessary for working with these data types in practice.

Create a data frame comprised of:

a random sample of size 10 from a standard Normal distribution a logical vector indicating whether elements of the sample are greater than 0 a character vector of length 10 a factor vector of length 10, with 3 different factor “levels” Try to take the mean of each variable in your dataframe. What works and what doesn’t?

```
my_dataframe <- tibble(
  random_sample = rnorm(10),
  log_vector = random_sample > 0,
  cha_vector = c("one", "two", "three", "four", "five", "six", "seven", "eight", "nine", "ten"),
  fac_vector = factor(c("big", "big", "big", "medium", "medium", "medium", "small", "small", "small", "small"))
)
```

```
mean(pull(my_dataframe, random_sample)) #find mean of the random sample
```

```
## [1] -0.3858217
```

```
mean(pull(my_dataframe, log_vector)) #find mean of the logical vector
```

```
## [1] 0.3
```

```
mean(pull(my_dataframe, cha_vector)) # find mean of the character vector
```

```
## Warning in mean.default(pull(my_dataframe, cha_vector)): argument is not numeric
## or logical: returning NA
```

```
## [1] NA
```

```
mean(pull(my_dataframe, fac_vector)) # find mean of the factor vector
```

```
## Warning in mean.default(pull(my_dataframe, fac_vector)): argument is not numeric
## or logical: returning NA
```

```
## [1] NA
```

We can find the mean of random sample and logical vector, but we can not find the mean of the character vector and factor vector.

```
num_log_vector = as.numeric(pull(my_dataframe, log_vector))
num_cha_vector = as.numeric(pull(my_dataframe, cha_vector))
num_fac_vector = as.numeric(pull(my_dataframe, fac_vector))
```

In some cases, you can explicitly convert variables from one type to another. Write a code chunk that applies the `as.numeric` function to the logical, character, and factor variables (please show this chunk but not the output). What happens, and why? Does this help explain what happens when you try to take the mean? The logical vector can be converted into numeric vector as 0 or 1, since R regards False as 0 and True as 1, and when calculating the mean, R automatically calculates the mean of those 0s and 1s. And the mean of it has real numerical value that how likely the sample is true.

The character vector can not be converted into numeric vector since the character vector does not have an intrinsic numerical value, therefore, R can not calculate the mean of it.

The factor vector can be converted into numeric vector since the levels are converted into number as 1,2,3, indicating three different levels. But those mean of levels does not have numerical meanings, thus R also can not calculate the mean of it.