HW_2_ozone

Darrell Sonntag

2024-01-20

knitr options are located here:

https://yihui.org/knitr/options/

Read in ozone data into a data.frame called ozone.data

- Select "House.Number","Visit","Location","ppm","LOD ppm"
- Filter out the rows with House.Number =='BLANK'
- Create a new column called "O3.below.det" that contains True or FALSE,
  - TRUE values are when the "ppm" values are below the detection limit
  - Values that are below the detection limit have a "<" sign
  - Use the function str_detect() (it's the tidyverse version of grepl in base R)
- Create a new variable "O3.ppb" from the ppm variable, convert it to a numeric variable, and multiply it by 1000
- Create a new variable called O3.estimate that is the "O3.ppb", or if it is below the detection limit, it is the value of "LOD ppm"*1000
  - We are imputing (substituting) the missing values with the limit of detention as the maximum possible value
  - Use the if_else() function

```
#| warning: False

ozone.data<- read_excel("..\\CE594R_25_data_science_class\\data\\Ozone Data_corrected.xlsx",
  select("House.Number","Visit","Location","Date Sampled","ppm","LOD ppm") %>%
  filter(House.Number!='BLANK') %>% ## also could use drop_na() from tidyr
  mutate(O3.below.det = str_detect(ppm,"<")) %>%
  mutate(O3.ppb = as.numeric(ppm)*1000) %>%
  rename(O3.LOD.ppm = "LOD ppm") %>%
  mutate(O3.estimate = if_else(O3.below.det,O3.LOD.ppm*1000,O3.ppb)) %>% #  Or use the repla
```

```
  mutate(House.Number = factor(House.Number)) %>%
  mutate(Visit = factor(Visit)) %>%
  mutate(Location = factor(Location))
```

```
New names:
* `Result` -> `Result...9`
* `Result` -> `Result...13`
* `` -> `...18`
* `` -> `...25`
* `` -> `...26`

Warning: There was 1 warning in `mutate()`.
i In argument: `O3.ppb = as.numeric(ppm) * 1000`.
Caused by warning:
! NAs introduced by coercion
```

```
str(ozone.data)
```

```
tibble [90 x 9] (S3: tbl_df/tbl/data.frame)
 $ House.Number: Factor w/ 32 levels "H02","H03","H04",..: 3 3 1 1 2 2 4 4 5 5 ...
 $ Visit       : Factor w/ 4 levels "V1","V2","V3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Location    : Factor w/ 2 levels "In","Out": 2 1 2 1 1 2 1 2 1 2 ...
 $ Date Sampled: POSIXct[1:90], format: "2022-07-28" "2022-07-28" ...
 $ ppm         : chr [1:90] "0.04" "<0.006" "3.7999999999999999E-2" "<0.006" ...
 $ O3.LOD.ppm  : num [1:90] 0.00525 0.0052 0.00551 0.00548 0.00531 ...
 $ O3.below.det: logi [1:90] FALSE TRUE FALSE TRUE TRUE FALSE ...
 $ O3.ppb      : num [1:90] 40 NA 38 NA NA 32 NA 20 NA 31 ...
 $ O3.estimate : num [1:90] 40 5.2 38 5.48 5.31 ...
```

Read in the house.info data, and join it to the ozone.data using 'House.Number', into a new data.frame called ozone.data.info

```
house.info <- read_excel("..\\CE594R_data_science_R\\data\\IndoorAir_houseinfo.xlsx")

ozone.data.info <- ozone.data %>%
                    left_join(house.info,by=c('House.Number'))
```

Create a summary table called ozone.data.summary

- Return the total number of measurements, and the number of measurements above the level of detection
- the mean, median, min, and max of "O3.ppb"
- measurements grouped by:
  - Location
  - Type of Air Conditioner
    - Note, because there are spaces in `Type of Air Conditioner` refer to the column using the back apostrophe
    - Or rename it with a name with no spaces
- display the ozone.data.summary in a formatted table using kable or tt

```
ozone.data.summary <- ozone.data.info %>%
                      group_by(Location,`Type of Air Conditioner`) %>%
                      summarize(total_n=n(), valid_n = sum(!(O3.below.det)),
                                mean_O3= mean(O3.ppb,na.rm=T), median_O3 = median(O3.ppb,na.r
```

```
`summarise()` has grouped output by 'Location'. You can override using the
`.groups` argument.
```

```
kable(ozone.data.summary,caption = "Ozone summary",digits=2)
```

Table 1: Ozone summary

Location	Type of Air Conditioner	total_n	valid_n	mean_O3	median_O3	min_O3	max_O3
In	Central	25	1	10.00	10	10.0	10
In	Evaporative	20	18	18.54	19	4.8	26
Out	Central	25	25	32.12	33	18.0	47
Out	Evaporative	20	20	28.60	29	17.0	40

Make a wide table with O3.estimate in two separate columns for In and Out locations

Calculate a new variable called the Indoor/Outdoor ratio which is the indoor O3 concentration divided by the outdoor O3 concentration for each visit.

Display the equation of the Indoor/Outdoor (I/O) ratio using LaTeX in an equation block. Include the subset the 3 in O3.

We used the following equation to calculate the indoor/outdoor ratio.

$$Indoor/Outdoor\ O_3\ ratio = \frac{Indoor\ O_3\ Concentration}{Outdoor\ O_3\ Concentration}$$

```
names(ozone.data.info)
```

```
 [1] "House.Number"              "Visit"
 [3] "Location"                  "Date Sampled"
 [5] "ppm"                       "O3.LOD.ppm"
 [7] "O3.below.det"              "O3.ppb"
 [9] "O3.estimate"               "City"
[11] "Zipcode"                   "Type of Air Conditioner"
```

```
ozone.wide <- ozone.data.info %>%
              select(-c(ppm,O3.LOD.ppm,O3.below.det,O3.ppb)) %>%
              pivot_wider(names_from = Location,values_from = O3.estimate) %>%
              mutate(In_Out_ratio = In/Out)
```

Summarize the Indoor/Outdoor O3 ratio by the type of air conditioning type (Central and Evaporative), including the number of measurements, the mean, median, min and max.

```
In.Out.summary <- ozone.wide  %>%
                  group_by(`Type of Air Conditioner`) %>%
                  summarize(total_n=n(), valid_n = sum(!is.na(In_Out_ratio)),
                            mean= mean(In_Out_ratio ,na.rm=T), median = median(In_Out_ratio
                            min = min(In_Out_ratio,na.rm=T), max=max(In_Out_ratio,na.rm=T))

kable(In.Out.summary,caption = "Indoor Outdoor Ratio Ozone summary",digits=2)
```

Table 2: Indoor Outdoor Ratio Ozone summary

Type of Air Conditioner	total_n	valid_n	mean	median	min	max
Central	25	25	0.13	0.13	0.06	0.29
Evaporative	20	20	0.58	0.63	0.13	0.94

Embed the number and the mean indoor/outdoor ration for the two types of air conditioners in a sentence or two. Round the mean values to two decimal points.

The indoor/outdoor ratio for the 25 visits in the Central Air homes was 0.13.

The indoor/outdoor ratio for the 20 visits in the Evaporative Cooler homes was 0.58.

Knit the document to an html document. Make sure to suppress any warning messages in any of the code-chunks

Write the ozone.wide to ozone.I.O.csv

```
write.csv(ozone.wide,"..//CE594R_data_science_R//data//ozone.I.O.csv",row.names=F)
```

HW submission

- Push the .qmd file to your github repository
- Provide a link to your repository in the assignment submission