

Brilliant Cassowary's Coffee Research

Preregistration of analyses

Nidhi Soma (ns848) Joice Chen (jc3528) Jinpeng Li (jl3496)
Stephen Syl-Akinwale (sis33)

Inserting the dataset on Coffee

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
coffee_df<-read_csv("data/GACTT_RESULTS_ANONYMIZED_v2.csv")
```

```
Rows: 4042 Columns: 113
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (44): Submission ID, What is your age?, How many cups of coffee do you t...
```

```
dbl (13): Lastly, how would you rate your own coffee expertise?, Coffee A - ...
```

```
lgl (56): Where do you typically drink coffee? (At home), Where do you typic...
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hypothesis 1

Younger people 18 - 25 prefer their coffee with lighter roasts.

Analysis: Run a linear (or logistic, if that fits better) regression where we input each age category as a discrete variable and output a prediction of the probability of preferring lighter roasts (refer to HW6 ex 2). We consider lighter roasts to be Light, Nordic, or Medium according to the roast_level column. The first age-group, which is the 18-25 age group, will be the reference variable, so we can see if the coefficients for the other age groups is not 0, so they factor into a lighter roast level preference. .

Hypothesis 2

People who have higher levels of education are less likely to prefer their coffee with added sweetener.

Analysis: Run a linear regression where we input education level as a discrete variable and output a prediction of the probability of answering TRUE to the variables starting with sugar_sweetener_added. The first group “less than high school” degree will be used as the reference variable.