Class 29

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-11-11

Lab 02

- Open RStudio and start a new project in a folder called `lab02`

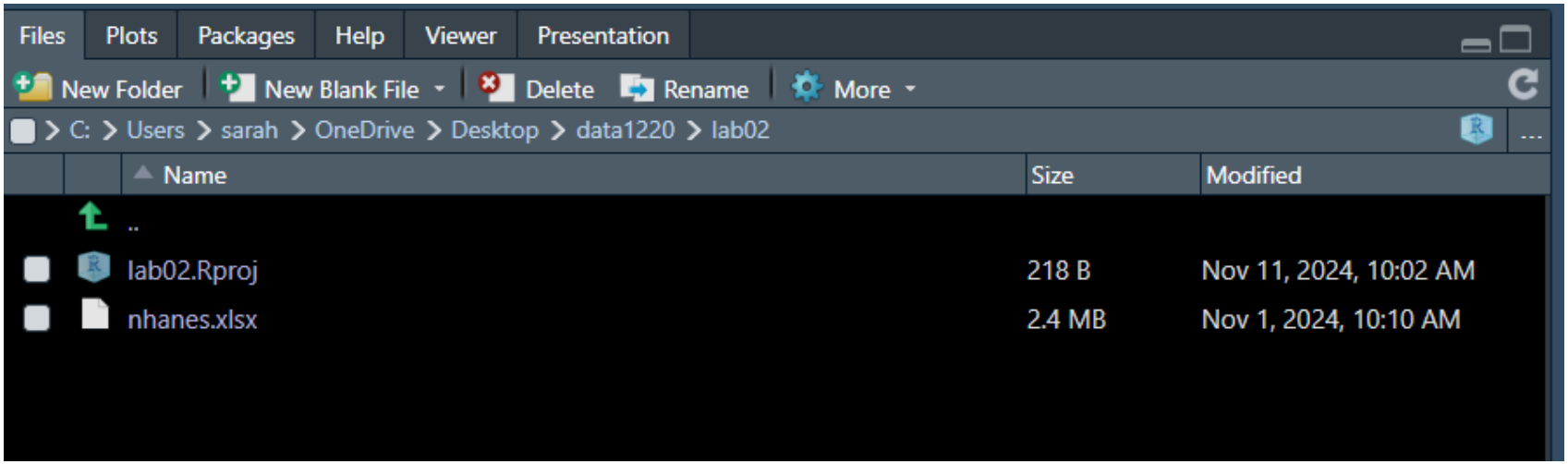
Lab 02

- Open RStudio and start a new project in a folder called `lab02`
- Go to File > New File > R Script. Save the file as `lab02.R`.

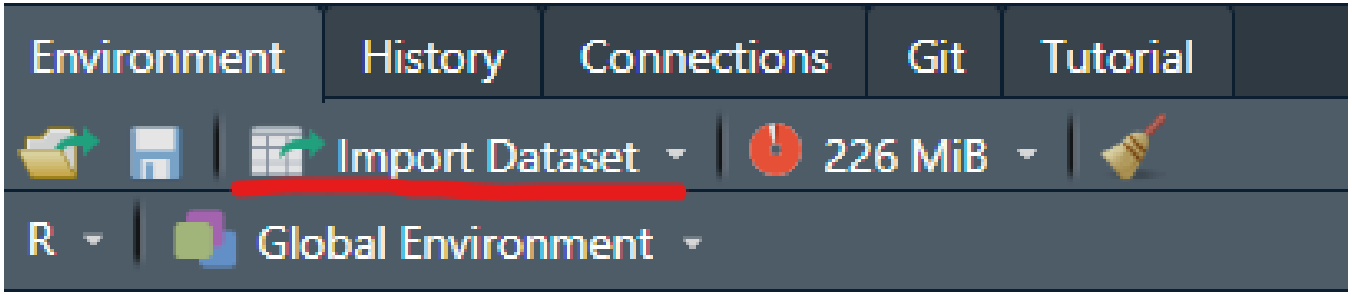
Lab 02

- Open RStudio and start a new project in a folder called `lab02`
- Go to File > New File > R Script. Save the file as `lab02.R`.
- Download the file `nhanes.xlsx` from Canvas under Lab 1 or find it in your `lab01` project folder.

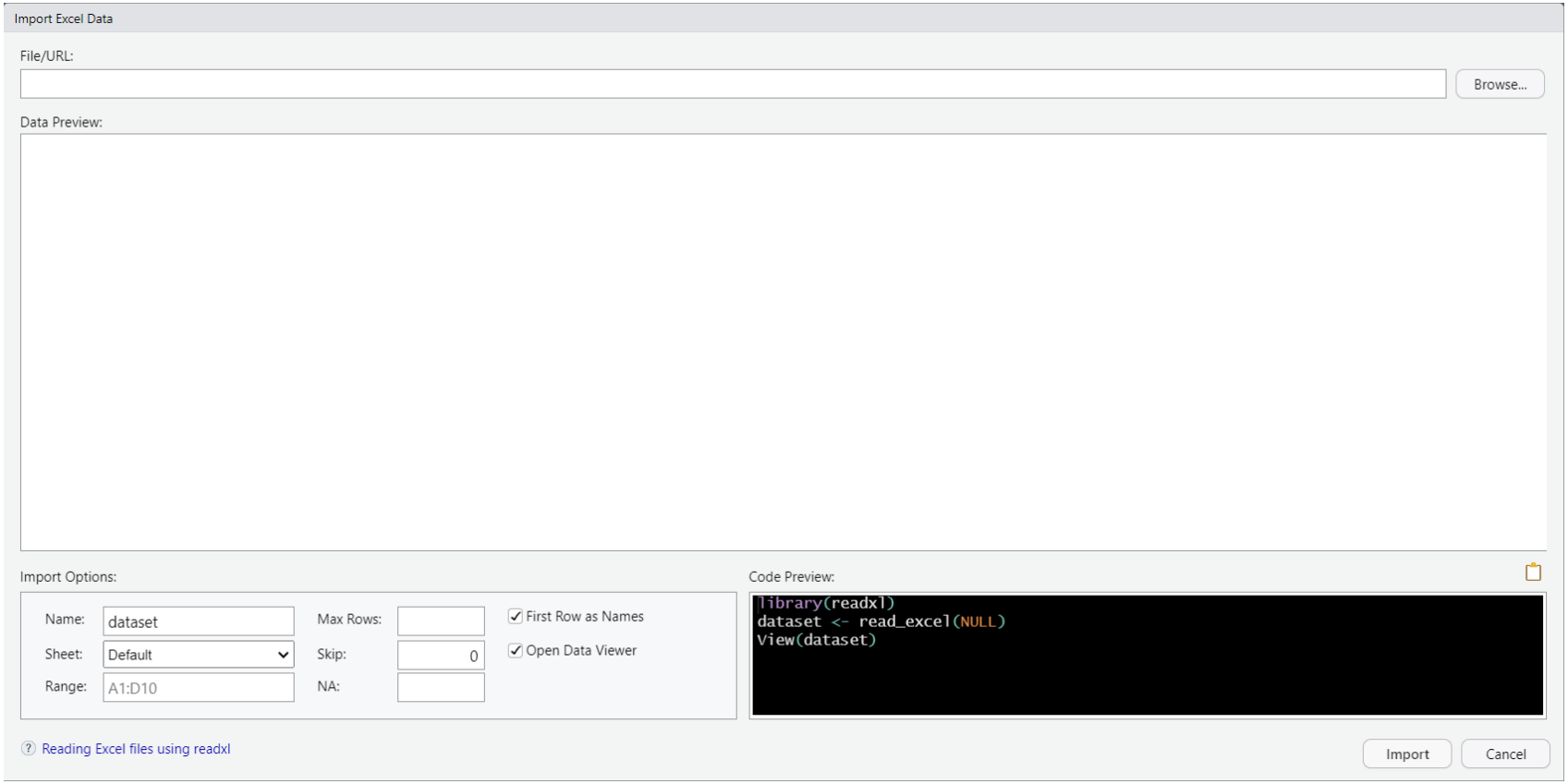
Files View


Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Delete
Rename
More
C: > Users > sarah > OneDrive > Desktop > data1220 > lab02
Name
Size
Modified
..
lab02.Rproj
218 B
Nov 11, 2024, 10:02 AM
nhanes.xlsx
2.4 MB
Nov 1, 2024, 10:10 AM

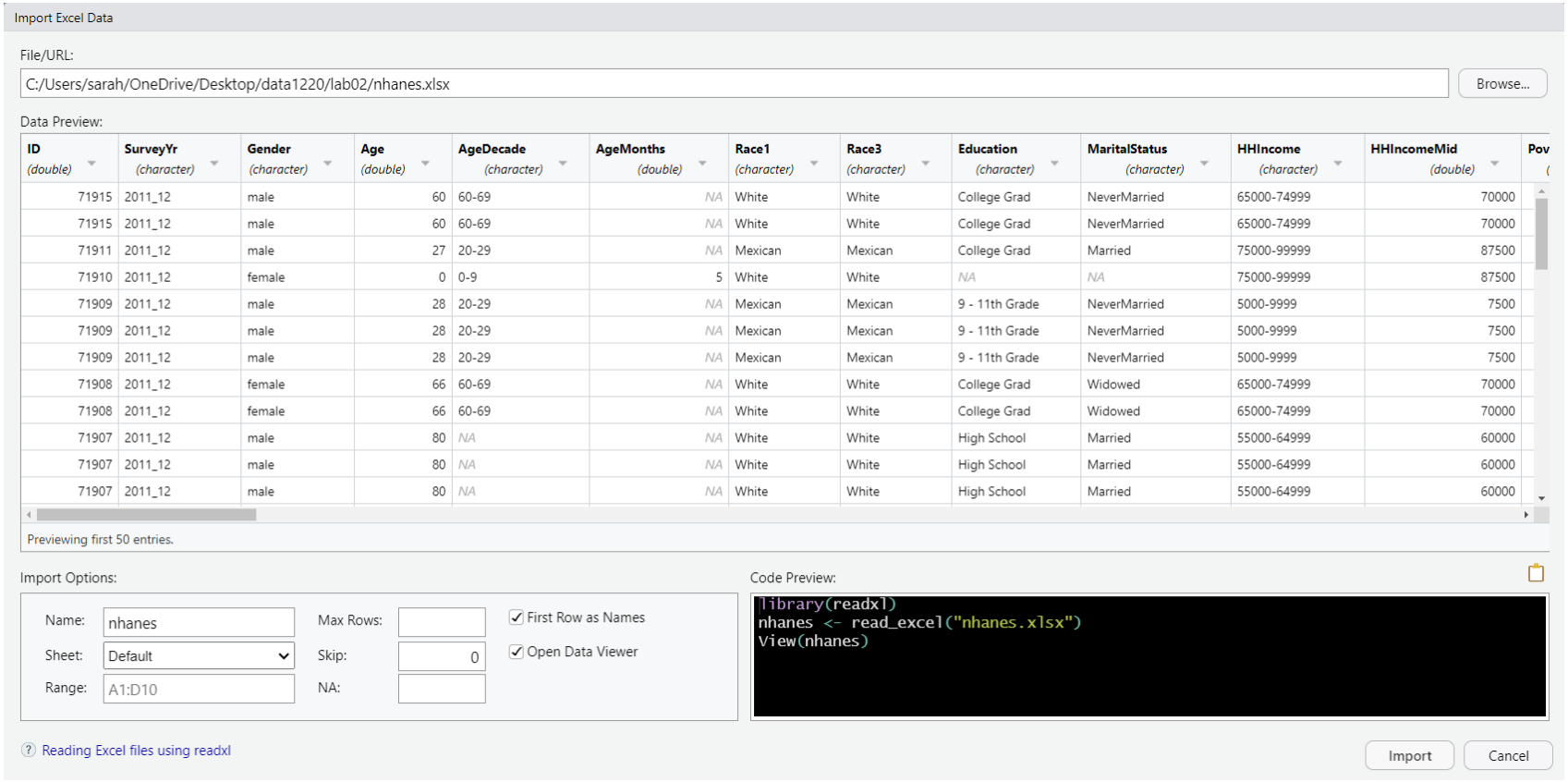
Importing a Dataset



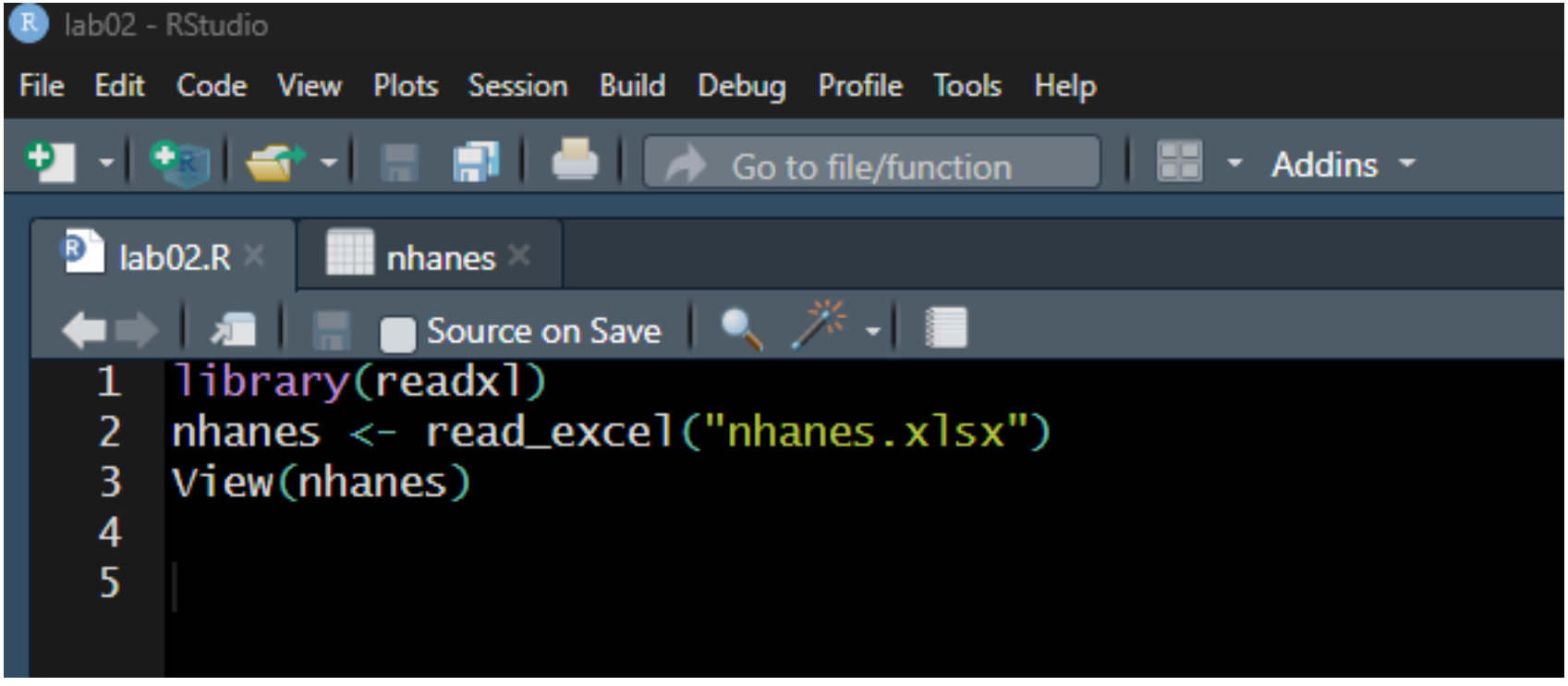
Importing Files - Excel



Importing Files - Excel



Importing Files - Excel (Code)



Load packages

- `readxl`: `read_xlsx()` function
- `Hmisc`: `describe()` function
- `dplyr`: `summarize()` and `select()` functions

Codebook

- `Gender`: male or female
- `AlcoholYear`: count of the number of days in a year the subject drinks alcohol

Research Question

How many days in a year do men and women drink on average?

Research Question

How many days in a year do men and women drink on average?

- Point estimates by gender ($\bar{x}_m$ and $\bar{x}_f$, s_m and s_f) and sample sizes (n_m and n_f)

Research Question

How many days in a year do men and women drink on average?

- Point estimates by gender ($\bar{x}_m$ and $\bar{x}_f$, s_m and s_f) and sample sizes (n_m and n_f)
- Sampling distributions for sample statistics $\bar{x}_m \sim \text{N}\left(\mu_m, SE_{\bar{x}_m}\right)$ and $\bar{x}_f \sim \text{N}\left(\mu_f, SE_{\bar{x}_f}\right)$

Research Question

How many days in a year do men and women drink on average?

- Point estimates by gender ($\bar{x}_m$ and $\bar{x}_f$, s_m and s_f) and sample sizes (n_m and n_f)
- Sampling distributions for sample statistics $\bar{x}_m \sim \text{N}\left(\mu_m, SE_{\bar{x}_m}\right)$ and $\bar{x}_f \sim \text{N}\left(\mu_f, SE_{\bar{x}_f}\right)$
- Confidence intervals for population means $\bar{x}_m \pm T^*_{\text{df}} \times \text{SE}_{\bar{x}_m}$ and $\bar{x}_f \pm T^*_{\text{df}} \times \text{SE}_{\bar{x}_f}$

Research Question

On average, do men and women drink the same number of days out of the year?

Research Question

On average, do men and women drink the same number of days out of the year?

- Point estimate of difference ($\bar{x}_m - \bar{x}_f$), standard deviations (s_m and s_f) and sample sizes (n_m and n_f)

Research Question

On average, do men and women drink the same number of days out of the year?

- Point estimate of difference $(\bar{x}_m - \bar{x}_f)$, standard deviations (s_m and s_f) and sample sizes (n_m and n_f)
- Sampling distribution $\bar{x}_m - \bar{x}_f \sim \text{N}\left(0, \text{SE}_{\bar{x}_m - \bar{x}_f}\right)$ for null hypothesis $H_0 : \mu_m - \mu_f = 0$

Research Question

On average, do men and women drink the same number of days out of the year?

- Point estimate of difference $(\bar{x}_m - \bar{x}_f)$, standard deviations (s_m and s_f) and sample sizes (n_m and n_f)
- Sampling distribution $\bar{x}_m - \bar{x}_f \sim \mathsf{N}\left(0, \mathsf{SE}_{\bar{x}_m - \bar{x}_f}\right)$ for null hypothesis $H_0 : \mu_m - \mu_f = 0$
- Test statistic t for observed $\bar{x}_m - \bar{x}_f$ under null distribution $\bar{x}_m - \bar{x}_f \sim \mathsf{N}\left(0, \mathsf{SE}_{\bar{x}_m - \bar{x}_f}\right)$

Research Question

On average, do men and women drink the same number of days out of the year?

- Point estimate of difference $(\bar{x}_m - \bar{x}_f)$, standard deviations (s_m and s_f) and sample sizes (n_m and n_f)
- Sampling distribution $\bar{x}_m - \bar{x}_f \sim \mathsf{N}\left(0, \mathsf{SE}_{\bar{x}_m - \bar{x}_f}\right)$ for null hypothesis $H_0 : \mu_m - \mu_f = 0$
- Test statistic t for observed $\bar{x}_m - \bar{x}_f$ under null distribution $\bar{x}_m - \bar{x}_f \sim \mathsf{N}\left(0, \mathsf{SE}_{\bar{x}_m - \bar{x}_f}\right)$
- P-value for test statistic t from Student's t distribution with degrees of freedom $\mathsf{df} = \min\left(n_m, n_f\right) - 1$

Research Question

What proportion of days in a year do men and women drink on average?

Research Question

What proportion of days in a year do men and women drink on average?

- Point estimates by gender ($\hat{p}_m$ and $\hat{p}_f$) and sample sizes (n_m and n_f)

Research Question

What proportion of days in a year do men and women drink on average?

- Point estimates by gender ($\hat{p}_m$ and $\hat{p}_f$) and sample sizes (n_m and n_f)
- Sampling distributions for sample statistics $\hat{p}_m \sim \text{N}\left(p_m, SE_{\hat{p}_m}\right)$ and $\hat{p}_f \sim \text{N}\left(p_f, SE_{\hat{p}_f}\right)$

Research Question

What proportion of days in a year do men and women drink on average?

- Point estimates by gender ($\hat{p}_m$ and $\hat{p}_f$) and sample sizes (n_m and n_f)
- Sampling distributions for sample statistics $\hat{p}_m \sim \text{N}\left(p_m, SE_{\hat{p}_m}\right)$ and $\hat{p}_f \sim \text{N}\left(p_f, SE_{\hat{p}_f}\right)$
- Confidence intervals for population means $\hat{p}_m \pm Z^* \times \text{SE}_{\hat{p}_m}$ and $\hat{p}_f \pm Z^* \times \text{SE}_{\hat{p}_f}$

Research Question

On average, do men and women drink the same proportion of days out of the year?

Research Question

On average, do men and women drink the same proportion of days out of the year?

- Point estimate of difference $(\hat{p}_m - \hat{p}_f)$ and sample sizes (n_m and n_f)

Research Question

On average, do men and women drink the same proportion of days out of the year?

- Point estimate of difference $(\hat{p}_m - \hat{p}_f)$ and sample sizes (n_m and n_f)
- Sampling distribution $\hat{p}_m - \hat{p}_f \sim \text{N}\left(0, \text{SE}_{\hat{p}_m - \hat{p}_f}\right)$ using $\hat{p}_{\text{pool}}$ for null hypothesis $H_0 : \mu_m - \mu_f = 0$

Research Question

On average, do men and women drink the same proportion of days out of the year?

- Point estimate of difference $(\hat{p}_m - \hat{p}_f)$ and sample sizes (n_m and n_f)
- Sampling distribution $\hat{p}_m - \hat{p}_f \sim \text{N}\left(0, \text{SE}_{\hat{p}_m - \hat{p}_f}\right)$ using $\hat{p}_{\text{pool}}$ for null hypothesis $H_0 : \mu_m - \mu_f = 0$
- Test statistic Z for observed $\hat{p}_m - \hat{p}_f$ under null distribution $\hat{p}_m - \hat{p}_f \sim \text{N}\left(0, \text{SE}_{\hat{p}_m - \hat{p}_f}\right)$

Research Question

On average, do men and women drink the same proportion of days out of the year?

- Point estimate of difference $(\hat{p}_m - \hat{p}_f)$ and sample sizes (n_m and n_f)
- Sampling distribution $\hat{p}_m - \hat{p}_f \sim \text{N}\left(0, \text{SE}_{\hat{p}_m - \hat{p}_f}\right)$ using $\hat{p}_{\text{pool}}$ for null hypothesis $H_0 \colon \mu_m - \mu_f = 0$
- Test statistic Z for observed $\hat{p}_m - \hat{p}_f$ under null distribution $\hat{p}_m - \hat{p}_f \sim \text{N}\left(0, \text{SE}_{\hat{p}_m - \hat{p}_f}\right)$
- P-value for test statistic Z from the standard normal Z distribution