

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

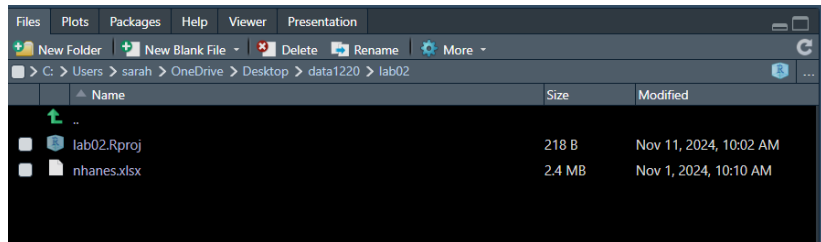
Lab 02

- ▶ 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.

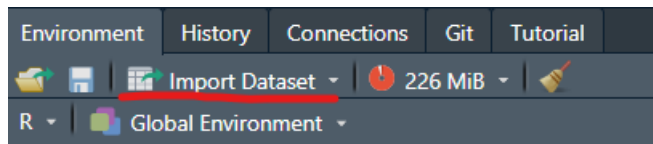
Lab 02

- ▶ 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



Importing a Dataset



Importing Files - Excel

Import Excel Data

File/URL: Browse...

Data Preview:

Import Options:

Name: <input type="text" value="dataset"/>	Max Rows: <input type="text"/>	<input checked="" type="checkbox"/> First Row as Names
Sheet: <input type="text" value="Default"/>	Skip: <input type="text" value="0"/>	<input checked="" type="checkbox"/> Open Data Viewer
Range: <input type="text" value="A1:D10"/>	NA: <input type="text"/>	

Code Preview:

```
library(readxl)
dataset <- read_excel(NULL)
View(dataset)
```

? Reading Excel files using readxl

Import Cancel

Importing Files - Excel

Import Excel Data

File/URL:
C:/Users/sarah/OneDrive/Desktop/data1220/lab02/nhanes.xlsx Browse...

Data Preview:

ID (double)	SurveyYr (character)	Gender (character)	Age (double)	AgeDecade (character)	AgeMonths (double)	Race1 (character)	Race3 (character)	Education (character)	MaritalStatus (character)	HHIncome (character)	HHIncomeMid (double)	Pov
71915	2011_12	male	60	60-69		NA	White	White	College Grad	NeverMarried	65000-74999	70000
71915	2011_12	male	60	60-69		NA	White	White	College Grad	NeverMarried	65000-74999	70000
71911	2011_12	male	27	20-29		NA	Mexican	Mexican	College Grad	Married	75000-99999	87500
71910	2011_12	female	0	0-9		5	White	White	NA	NA	75000-99999	87500
71909	2011_12	male	28	20-29		NA	Mexican	Mexican	9 - 11th Grade	NeverMarried	5000-9999	7500
71909	2011_12	male	28	20-29		NA	Mexican	Mexican	9 - 11th Grade	NeverMarried	5000-9999	7500
71909	2011_12	male	28	20-29		NA	Mexican	Mexican	9 - 11th Grade	NeverMarried	5000-9999	7500
71908	2011_12	female	66	60-69		NA	White	White	College Grad	Widowed	65000-74999	70000
71908	2011_12	female	66	60-69		NA	White	White	College Grad	Widowed	65000-74999	70000
71907	2011_12	male	80	NA		NA	White	White	High School	Married	55000-64999	60000
71907	2011_12	male	80	NA		NA	White	White	High School	Married	55000-64999	60000
71907	2011_12	male	80	NA		NA	White	White	High School	Married	55000-64999	60000

Previewing first 50 entries.

Import Options:

Name: Max Rows: ☒ First Row as Names

Sheet: Skip: ☒ Open Data Viewer

Range: NA:

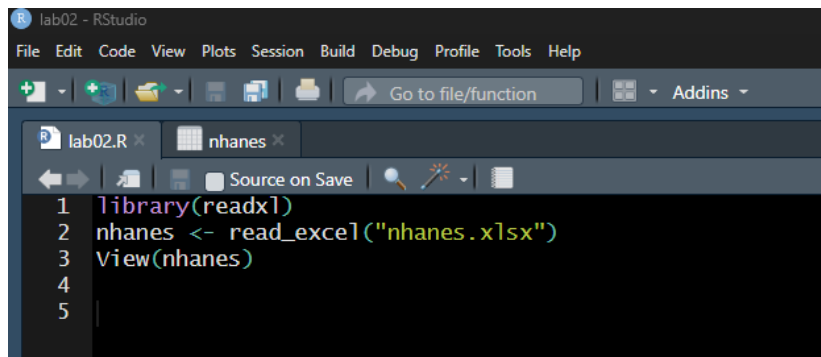
Code Preview:

```
library(readxl)
nhanes <- read_excel("nhanes.xlsx")
view(nhanes)
```

Import Cancel

Reading Excel files using readxl

Importing Files - Excel (Code)



The screenshot shows the RStudio environment. The title bar reads "lab02 - RStudio". The menu bar includes "File", "Edit", "Code", "View", "Plots", "Session", "Build", "Debug", "Profile", "Tools", and "Help". The toolbar contains icons for file operations and a search bar labeled "Go to file/function". The file explorer shows two open files: "lab02.R" and "nhanes". The source editor displays the following R code:

```
1 library(readxl)
2 nhanes <- read_excel("nhanes.xlsx")
3 View(nhanes)
4
5
```

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
 $\bar{x}_m \sim N(\mu_m, SE_{\bar{x}_m})$ and $\bar{x}_f \sim N(\mu_f, SE_{\bar{x}_f})$

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
 $\bar{x}_m \sim N(\mu_m, SE_{\bar{x}_m})$ and $\bar{x}_f \sim N(\mu_f, SE_{\bar{x}_f})$
- ▶ Confidence intervals for population means $\bar{x}_m \pm T_{df}^* \times SE_{\bar{x}_m}$ and $\bar{x}_f \pm T_{df}^* \times SE_{\bar{x}_f}$

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 $\bar{x}_m - \bar{x}_f \sim N(0, SE_{\bar{x}_m - \bar{x}_f})$ 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 $\bar{x}_m - \bar{x}_f \sim N(0, SE_{\bar{x}_m - \bar{x}_f})$ for null hypothesis $H_0: \mu_m - \mu_f = 0$
- ▶ Test statistic t for observed $\bar{x}_m - \bar{x}_f$ under null distribution $\bar{x}_m - \bar{x}_f \sim N(0, SE_{\bar{x}_m - \bar{x}_f})$

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 $\bar{x}_m - \bar{x}_f \sim N(0, SE_{\bar{x}_m - \bar{x}_f})$ for null hypothesis $H_0: \mu_m - \mu_f = 0$
- ▶ Test statistic t for observed $\bar{x}_m - \bar{x}_f$ under null distribution $\bar{x}_m - \bar{x}_f \sim N(0, SE_{\bar{x}_m - \bar{x}_f})$
- ▶ P-value for test statistic t from Student's t distribution with degrees of freedom $df = \min(n_m, n_f) - 1$

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
 $\hat{p}_m \sim N(p_m, SE_{\hat{p}_m})$ and $\hat{p}_f \sim N(p_f, SE_{\hat{p}_f})$

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
 $\hat{p}_m \sim N(p_m, SE_{\hat{p}_m})$ and $\hat{p}_f \sim N(p_f, SE_{\hat{p}_f})$
- ▶ Confidence intervals for population means $\hat{p}_m \pm Z^* \times SE_{\hat{p}_m}$ and $\hat{p}_f \pm Z^* \times SE_{\hat{p}_f}$

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 $\hat{p}_m - \hat{p}_f \sim N(0, SE_{\hat{p}_m - \hat{p}_f})$ using \hat{p}_{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 $\hat{p}_m - \hat{p}_f \sim N(0, SE_{\hat{p}_m - \hat{p}_f})$ using \hat{p}_{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 $\hat{p}_m - \hat{p}_f \sim N(0, SE_{\hat{p}_m - \hat{p}_f})$

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 $\hat{p}_m - \hat{p}_f \sim N(0, SE_{\hat{p}_m - \hat{p}_f})$ using \hat{p}_{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 $\hat{p}_m - \hat{p}_f \sim N(0, SE_{\hat{p}_m - \hat{p}_f})$
- ▶ P-value for test statistic Z from the standard normal Z distribution