



Data Science in Intro & Intermediate Statistics



August 13, 2021

you...

- know R
- are familiar with R Markdown
- are interested in integrating R into your course(s)

Tools

1. RStudio Cloud
2. Shiny Apps
3. learnr Tutorials

Demo:

<http://shinyed.github.io/intro-stats/>

Diagnostics for simple linear regression

Select a trend:

- ☒ Linear up
- ☐ Linear down
- ☐ Curved up
- ☐ Curved down
- ☐ Fan-shaped

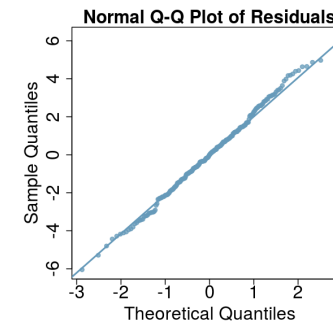
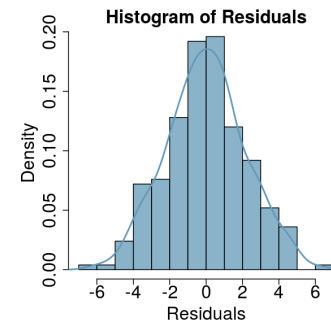
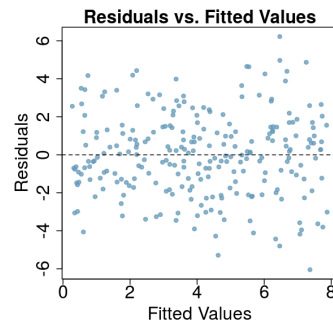
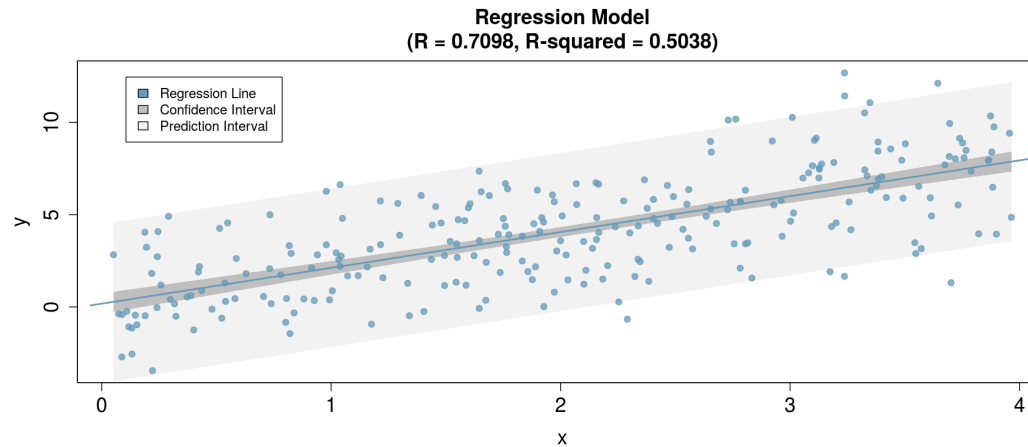
☐ Show residuals

This app uses ordinary least squares (OLS) to fit a regression line to the data with the selected trend. The app is designed to help you practice evaluating whether or not the linear model is an appropriate fit to the data. The three diagnostic plots on the lower half of the page are provided to help you identify undesirable patterns in the residuals that may arise from non-linear trends in the data.

[View code](#)

[Check out other apps](#)

[Want to learn more for free?](#)



Good!

- No need to install or interact with R/RStudio
- Students can get experience with code without typing code
- Gives them an easy "copy-paste" resource

Not so good...

- Apps need to be hosted somewhere
- Many students using app at once = high computing demand
- Somewhat hard to create (for now)



- **learnr** is an R package that makes it easy to create interactive tutorials from R Markdown documents.
- Tutorials can include:
 - Narrative, figures, illustrations, and equations
 - Code exercises (R code chunks that users can edit and execute directly)
 - Multiple choice questions
 - Videos (YouTube, Vimeo)
 - Interactive Shiny components
- learnr is on CRAN

```
install.packages("learnr")
```


Introduction to Statistical Inference: 2 - Randomization Test

Completing a randomization test:

gender discrimination

Distribution of statistics

Why 0.05?

What is a p-value?

Summary of gender discrimination

Congratulations!

Start Over

Gender discrimination hypotheses

Which of the following null and alternative hypotheses are appropriate for the gender discrimination example described in the previous lesson?

- ☐ H0: gender and promotion are unrelated variables. HA: men are more likely to be promoted.
- ☐ H0: gender and promotion are unrelated variables. HA: women are more likely to be promoted.
- ☐ H0: men are more likely to be promoted. HA: gender and promotion are unrelated variables.
- ☐ H0: women are more likely to be promoted. HA: gender and promotion are unrelated variables.

Submit Answer

Summarizing gender discrimination

As the first step of any analysis, you should look at and summarize the data. Categorical variables are often summarized using proportions, and it is always important to understand the denominator of the proportion.

Do you want the proportion of women who were promoted or the proportion of promoted individuals who were women? Here, you want the first of these, so in your R code it's necessary to group by `gender` **before** you calculate the proportions!

The discrimination study data are available in your workspace as `gender_discrimination`.

- Using the `count()` function, tabulate the variables `gender` and `decision`.
- Group the data by `gender`.
- Calculate the proportion of those who were and were not promoted in each gender and call this variable `prop`.

R code

↺ Start Over

💡 Hints

▶ Run Code

```
1 gender_discrimination %>%
2   count(____, ____ ) %>%
3   group_by(____) %>%
4   mutate(____ = ____ / ____)
```

Components of a `learnr` tutorial

YAML

Start with a YAML, just like in R Markdown:

```
---  
title: "Starting with Data"  
output:  
  learnr::tutorial:  
    progressive: true  
    allow_skip: true  
runtime: shiny_prerendered  
---
```

1. Create a new RMarkdown file
2. Select from Template
3. Choose the Interactive Tutorial template from **learnr**
4. Start editing!

Narrative

- R Markdown style section and subsection headings with ##, ###, etc.
- Text, figures, illustrations, and equations.
- Videos: supported services include YouTube and Vimeo

Hello, and welcome to **Getting Started with Data**!

In this tutorial we will take you through concepts and R code that are essential **for** getting started with data analysis.

Scientists seek to answer questions using rigorous methods and careful observations. These observations form the backbone of a statistical investigation and are called data. Statistics is the study of how best to collect, analyze, and draw conclusions from data. It is helpful to put statistics **in** the context of a general process of investigation:


- **Step 1**: Identify a question or problem.
- **Step 2**: Collect relevant data on the topic.
- **Step 3**: Analyze the data.
- **Step 4**: Form a conclusion.


Multiple choice questions


```
quiz(  
  question("What position is the letter A in the english alphabet?",  
    answer("8"),  
    answer("14"),  
    answer("1", correct = TRUE),  
    answer("23"),  
    incorrect = "See [here](https://en.wikipedia.org/wiki/English_alphabet) and try again.",  
    allow_retry = TRUE  
  ),  
  
  question("Where are you right now? (select ALL that apply)",  
    answer("Planet Earth", correct = TRUE),  
    answer("Pluto"),  
    answer("At a computing device", correct = TRUE),  
    answer("In the Milky Way", correct = TRUE),  
    incorrect = paste0("Incorrect. You're on Earth, ",  
                        "in the Milky Way, at a computer.")  
  )  
)
```

Code exercises - rendered

R code

 Start Over

 Hints

 Run Code

```
1 gender_discrimination %>%  
2   count(____, ____)%>%  
3   group_by(____)%>%  
4   mutate(____ = ____ / ____)
```

Code exercises - code

```
```{r gender-promoted, exercise=TRUE}
gender_discrimination %>%
 count(█, █) %>%
 group_by(█) %>%
 mutate(█ = █ / █)
```

```{r gender-promoted-hint-1}
gender_discrimination %>%
 count(gender, decision) %>%
 group_by(█) %>%
 mutate(█ = █ / █)
```

```{r gender-promoted-hint-2}
gender_discrimination %>%
 count(gender, decision) %>%
 group_by(gender) %>%
 mutate(█ = █ / █)
```
```

Code exercises - solution

Solution [Copy to Clipboard](#)

```
1 # Calculate the observed difference in promotion rate
2 diff_orig <- gender_discrimination %>%
3   # Group by gender
4   group_by(gender) %>%
5   # Summarize to calculate proportion promoted
6   summarize(prop_promoted = mean(decision == "promoted")) %>%
7   # Summarize to calculate difference
8   summarize(stat = diff(prop_promoted))
9
10 # See the result
11 diff_orig
```

R code [Start Over](#) [Solution](#) [Run Code](#)

```
1 # Calculate the observed difference in promotion rate
2 diff_orig <- gender_discrimination %>%
3   # Group by gender
4   group_by(____) %>%
5   # Summarize to calculate proportion promoted
6   summarize(prop_promoted = mean(decision == "promoted")) %>%
7   # Summarize to calculate difference
8   ____ (stat = ____ (____))
9
10 # See the result
11 diff_orig
```

Continue

Q: How do I share with my students?

- Deploy on
 - shinyapps.io (variety of pricing plans available)
 - RStudio Connect (free for academic use, requires setup)
- Essential reading:
 - Publishing learnr Tutorials on shinyapps.io by Angela Li
 - Teach R with learnr: a powerful tool for remote teaching by Allison Horst
 - See the publishing instructions on the learnr website for step-by-step instructions

What are my resources?

- Does your university have server time for hosting **learnr** tutorials?
- If not, does your university have funding sources to host these things elsewhere? (e.g. shinyapps.io)
- Can you acquire funding for RStudio Cloud? Can you charge students to use it?
- How much do **you** enjoy creating/teaching with R resources?

What are my learning objectives?

- If **learning R** is one of them...

...students should probably work with R through RStudio!

- If **understanding how statisticians use code** is one of them ...

... consider **learnr** tutorials with pre-supplied code or RS Cloud.

- If **software is not a learning objective**...

... consider using R as a back-end only, to make your own life easier.

Thank you!

 atheobol@calpoly.edu

 @allisontheobold

 @atheobold