

STAT-462:003 2021: Homework 1

THIS IS DUE SUNDAY JAN 24th AT 23.59PM

STRUCTURE

Every homework has a space to say hi and tell me how things are going, plus the structure:

1. Theory questions
2. A mini R-question
3. Reflection/reading/feedback

If you submit your attempt, I will provide the worked answers. You do not have to get all the answers correct to get 4/4.

4/4 All parts meaningfully attempted. You included your workings and thoughts.

3/4 All parts attempted at some level. Just an answer, no workings.

2/4 Missing sections or superficial attempt

1/4 Any engagement at all

THEORY

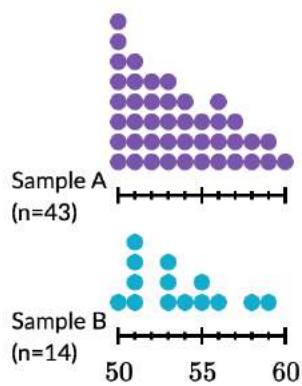
The aim of the theory section this week is to see how much you remember of the pre-requisites so that I don't pitch the course too low or too high. If you need a refresher, you might find these resources useful:

- STAT-200 online textbook: <https://online.stat.psu.edu/stat200/>
- Kahn Academy distributions: <https://www.khanacademy.org/math/ap-statistics/density-curves-normal-distribution-ap>
- Kahn Academy hypothesis tests: <https://www.khanacademy.org/math/ap-statistics/tests-significance-ap>

1. What pre-req did you take for this course and when did you take it? (STAT-200, AP-STAT etc)
2. The wingspans of a population of birds are normally distributed with a mean of 14.1 and a standard deviation of 1.7. Find:
 - a. The wingspan that is greater than 90% of the population.
 - b. The wingspan that is less than 20% of the population.
 - c. The limits of the central 95% of the wingspan.

3. Describe the Central Limit Theorem and why it is important

4. Here are two different samples drawn from two different populations.
Which sample satisfies the normal condition for performing a T test and why?



- a. A only
- b. B only
- c. Both A and B
- d. Neither

5.

Catalina read a report saying that the average daily high temperature in Buenos Aires a few years ago was 18.2°C . She believes that last year was warmer than this. To test her belief, she takes a simple random sample of 20 days from last year and records the daily high temperature on each chosen day.

Let μ represent the average daily high temperature in Buenos Aires last year.

Which of the following is an appropriate set of hypotheses for Catalina's significance test?

Choose 1 answer:

☐ (A) $H_0 : \mu = 18.2^{\circ}\text{C}$
 $H_a : \mu < 18.2^{\circ}\text{C}$

☐ (B) $H_0 : \mu > 18.2^{\circ}\text{C}$
 $H_a : \mu = 18.2^{\circ}\text{C}$

☐ (C) $H_0 : \mu = 18.2^{\circ}\text{C}$
 $H_a : \mu > 18.2^{\circ}\text{C}$

☐ (D) $H_0 : \mu > 18.2^{\circ}\text{C}$
 $H_a : \mu \neq 18.2^{\circ}\text{C}$

R PROGRAMMING

ACTIVITY 1

R is a programming language commonly used by statisticians and scientists across the world.

When I say R is a programming language, I mean that it is a collection of commands that you can type into the computer in order to analyse and visualise your data. Think of it literally like a language like Spanish or Hindi, so learning it means learning vocabulary and grammar to communicate with your computer. It also means it will get easier with experience and practice..

Read this blog about why R might be a useful tool:

<https://www.psychologicalscience.org/observer/why-you-should-become-a-user-a-brief-introduction-to-r>

Why do you think that R might be useful for a course on regression or for your future career? Is there anything you are especially excited to learn about? (or continue learning about)

R AND R STUDIO

When you install R on your computer, you are essentially teaching the computer to “speak in R” with some very basic Notepad-like software where you could enter the commands.

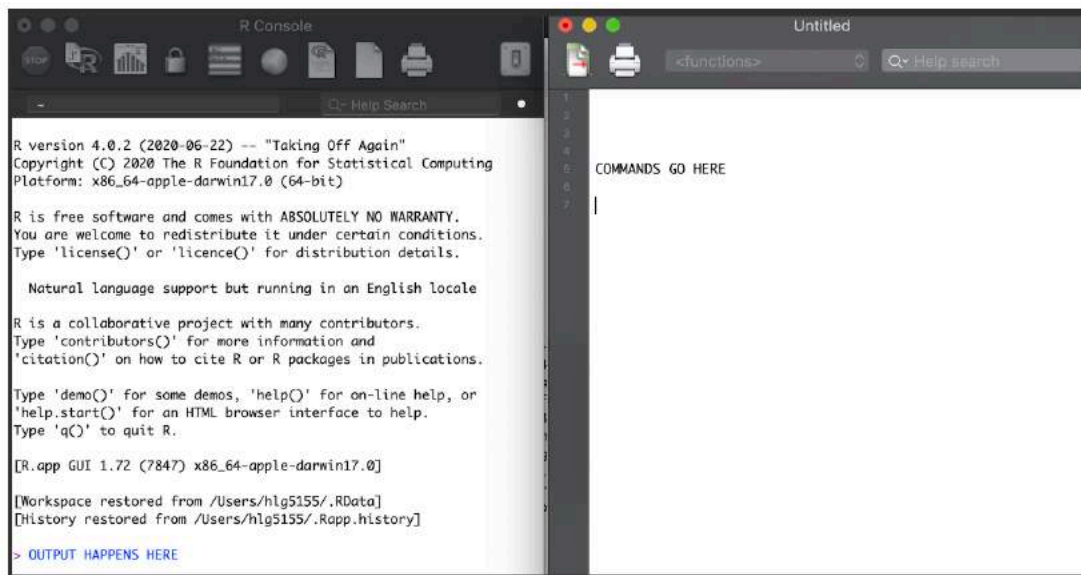


Figure 1.1: The basic R screen

More recently, R-studio has been designed as a piece of software to make it easier to programme in R. It's what Microsoft Word is compared to notepad - lots more functionality and things to click. For example, you can easily see help files, run code, see your output and link to all the website/dashboard builders.

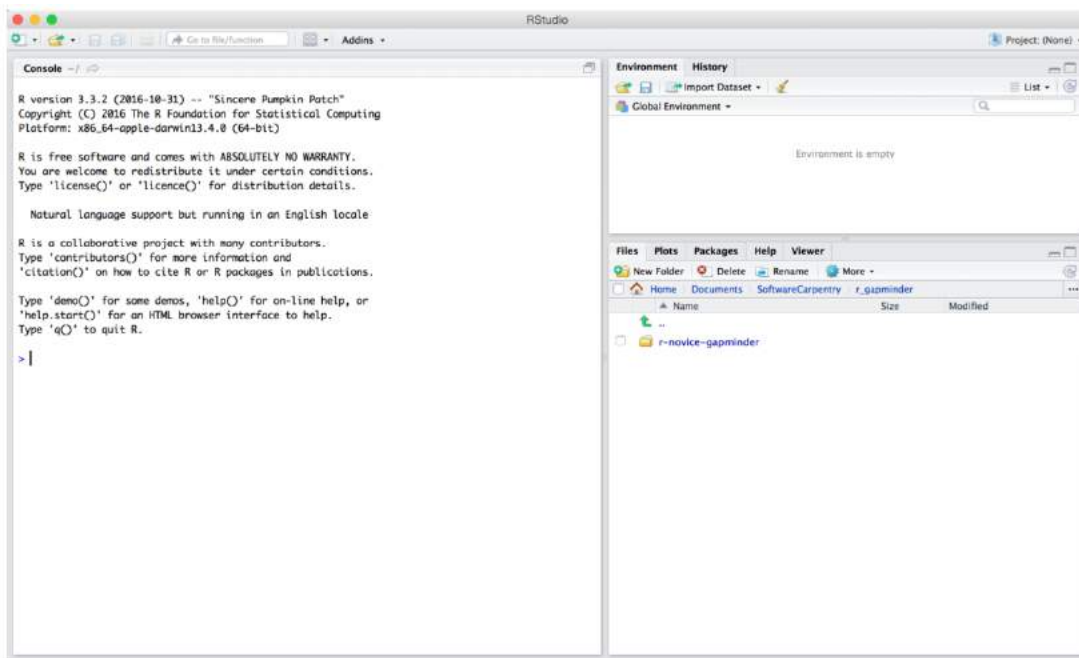


Figure 1.2: The basic R-studio screen. Here is the R-Studio website where you can explore some of its features: <https://rstudio.com/products/rstudio/features/>

For this course we need both installed on your computer.

This is week 1 and as we are remote, it is important to make sure that R and R-Studio work on your machine.

If there are issues, don't panic! We do have online and remote options, so talk to Dr Greatrex

ACTIVITY 2A – IF YOU DO NOT HAVE R/R-STUDIO ON YOUR COMPUTER

If you do not have R or R-studio on your machine, follow these instructions to install

<https://psu.instructure.com/courses/2115020/pages/how-to-install-or-update-r-and-r-studio>

ACTIVITY 2A – IF YOU DO ALREADY HAVE R/R-STUDIO ON YOUR COMPUTER

It is important that you update R and R studio to the most recent version.

These instructions might help: <https://uvastatlab.github.io/phdplus/installR.html#updateR>

ACTIVITY 2B

Now open **R-Studio**.

Click on the Help Tab and select “About R”.

Take a screenshot so that I can see both the R version AND the R-studio version (should be R version 4.0.3 and R-Studio version 1.3.1093 or later)

Add the screenshot here

REFLECTION

Fill in the introductory discussion board here:

https://psu.instructure.com/courses/2115020/discussion_topics/13706484