The IPSUR package: an Introduction to Probability and Statistics Using R

G. Jay Kerns, Youngstown State University

National Institute of Standards and Technology

Outline

Contents

1 Introduction

Who are we talking about?

- Course Title: Probability and Statistics (4 s.h.)
  - Approx. equal time for each of Probability and Statistics
- Students:
  - Primarily Engineering majors, also Mathematics, Computer Science, Physical Sciences
  - Lot of commuters, most own a computer, many are employed
- Challenges
  - Don't have time to stick around on campus
  - Don't want to buy student versions of proprietary software
  - Aren't keen on command line tools

2 What is IPSUR?

What is IPSUR?

- It's a book (PDF).

  `http://ipsur.r-forge.r-project.org/`

- It's a monstrous Sweave file.
- It's an R package that includes
  - PDF as a vignette
  - data as `.RData` file
  - solution manuals, answer keys
- Covers standard topics from introductory course with Calculus as prerequisite.

Why would anybody want IPSUR?

- Books are expensive.
- Books are HEAVY!
- Books are dated.
  - content
  - students

3 How does IPSUR work?

Using IPSUR in the classroom

- Students: install the `IPSUR` package
  - load the package with `library(IPSUR)`
  - read the book with `read(IPSUR)`
  - get the data for exercises with `data(IPSUR)`
- Instructors: use mine OR tweak it a bit OR both
  - download the `IPSUR` source package
  - choose a random seed (shhhhh!)
  - generate PDF via `Sweave` with new data, exercises, solution manuals, answer keys
  - distribute the book, data, and solution manuals (keep the answers)

6. Let X and Y denote the weights in pounds of male and female cheetahs, respectively. Assume that the distributions of X and Y are $\texttt{norm}(\texttt{mean} = \mu_X, \texttt{sd} = \sigma_X)$ and $\texttt{norm}(\texttt{mean} = \mu_Y, \texttt{sd} = \sigma_Y)$. Further assume that $\sigma_X = \sigma_Y$.

   Given the following data:

   For X:

   ```
   [1] 498.86 499.71 501.79 500.08 506.91 499.14 499.08 504.41
   ```

   For Y:

   ```
   [1] 499.51 493.91 503.47 500.47 504.15 500.72 501.26 502.95 496.03 499.51
   ```

   (a) Define the test statistic and critical region that has an $\alpha = 0.025$ level for testing $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$. Sketch a figure illustrating the critical region.

   (b) Calculate the value of the test statistic and state your conclusion.

   (c) Find a 97.5% one-sided confidence interval for $\mu_X - \mu_Y$. Be sure to address all five (5) salient features of confidence intervals. (PANIC)

   (d) Include a statistically appropriate visual display for these two datasets on the same graph, and discuss the relationship between the graph and your conclusion.

   (a) Define the test statistic and critical region that has an $\alpha = 0.01$ level for testing $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$. Sketch a figure illustrating the critical region.

   (b) Calculate the value of the test statistic and state your conclusion.

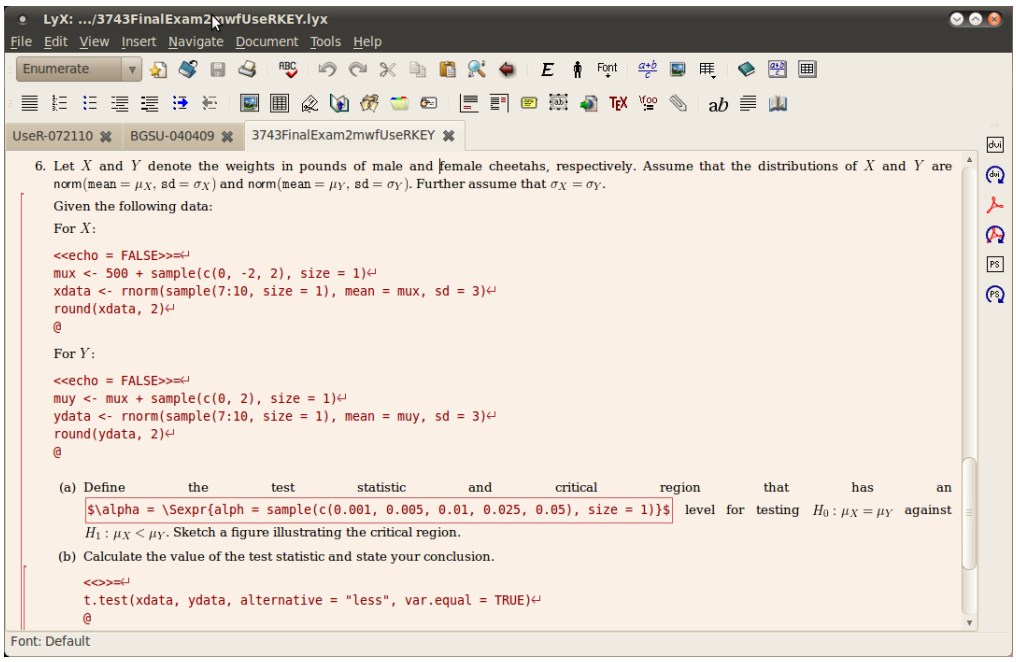
   ```
   > t.test(xdata, ydata, alternative = "less", var.equal = TRUE)

           Two Sample t-test

   data:  xdata and ydata
   t = 0.1527, df = 16, p-value = 0.5597
   alternative hypothesis: true difference in means is less than 0
   95 percent confidence interval:
        -Inf 2.491573
   sample estimates:
   mean of x mean of y
    500.1439  499.9435
   ```

   (c) Find a 99% one-sided confidence interval for $\mu_X - \mu_Y$. Be sure to address all five (5) salient features of confidence intervals. (PANIC)

   ```
   > t.test(xdata, ydata, alternative = "less", var.equal = TRUE,
   +     conf.level = 1 - alph)$conf.int
   [1]     -Inf 3.590793
   attr(,"conf.level")
   [1] 0.99
   ```



4 Advantages/Disadvantages

Why `IPSUR` is good

- It's FREE. (hard copy not bad, either)
  - CourseCompass, Blackboard, WebAssign, MyStatLab
- Binaries distributed via `CRAN` and/or R-Forge.
  - only need working R installation and Internet connection
  - platform independent
- It's a package
  - automated package CHECKs
  - students have immediate access to every datum, every line of code
  - easy to fix typos

Why `IPSUR` is not so good

- It's FREE.

- Quality control, i.e. peer review
- Real data
-

Why `IPSUR` is not so good

- It's FREE.
- Quality control, i.e. peer review
- Real data
- Grading SUCKS.

Recent Experiences

5 Concluding Remarks

Summary

- R is Great!
- The `IPSUR` package is an attempt to harness R's power to disseminate statistical content to students.
- Since it's FREE, you are FREE to make it better!

References

For Further Reading

References

[1] G. Jay Kerns, with contributions by Theophilius Boye and Tyler Drombosky *The* `RcmdrPlugin.IPSUR` *Package* Version 0.1-6, 2008.

[2] John Fox *The R Commander: A Basic-Statistics Graphical User Interface to R. Journal of Statistical Software.* 14(9):1–42, 2005.

[3] John Fox *Extending the R Commander by "Plug-In" Packages. R News.* 3:46–51, 2007.

[4] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.Other Resources

- IPSUR website: `http://ipsur.r-forge.r-project.org/`
- R and R Commander Instructions