CS 251 Statistical Computing

HOP 10: R for statistical project

8/28/2020 Developed by Aya Khalil

7/7/2019 Reviewed by

School of Technology and Computing @City University of Seattle (CityU)

**Before You Start**

* If you already finished this module through any CityU Technology Institute (TI) courses,  
  just skim this module and skip it.
* Version numbers may not match with the guide. But that should be fine.  
  If given the option to choose between stable release (long-term support) or most recent, please choose the stable release.
* This guide targets Windows OS users. So, MacOS users may have different commands to input in the shell/terminal.
* We cannot explain every step. **This cookbook always needs your own creative judgement.**
* **For your working directory, use your course number.** The hands-on tutorial may use a different course number as an example.

**Learning Outcomes**

* Poisson Distribution

**Resource**

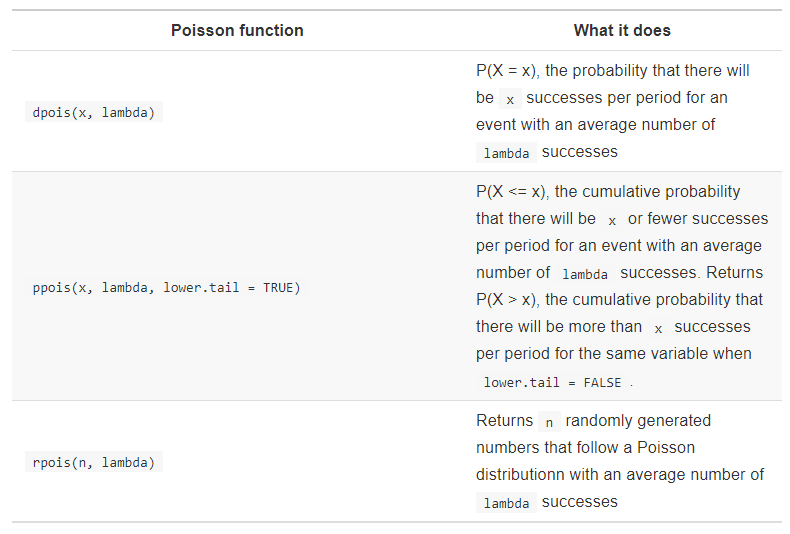
* Poisson Distribution: <https://bookdown.org/gabriel_butler/ECON41Labs/tutorial-5-the-poisson-distribution.html>
* Poisson Distribution: <http://www.r-tutor.com/elementary-statistics/probability-distributions/poisson-distribution>
* Pipes in R: <https://www.datacamp.com/community/tutorials/pipe-r-tutorial>

**Section 1: Poisson Distribution**

The Poisson distribution is the probability distribution of independent event occurrences in an interval. If λ is the mean occurrence per interval, then the probability of having x occurrences within a given interval is:



R has several built-in functions for the Poisson distribution. They’re listed in a table below along with brief descriptions of what each one does.



We will begin with rpois(). First, we will use it to make one (n = 1) randomly generated observation of a random variable that follows the Poisson distribution and has an average number of 10 successes (lambda = 10) per period.

**Setup Working Environment for Module9**

1. Open VS Code.

* **online student:** Open CS251 \_Fall\_2020/**ON**/FirstnameLastname /. ( File > Open )
* **onsite student:** Open CS251 \_ Fall \_2020/**IN**/FirstnameLastname. ( File > Open )

1. Then, create the “**Module10**” directory in the VSCode.

>>>mkdir Module10

1. In module10 project folder, create new file Rpois.R

Type the following in Rpois.R file

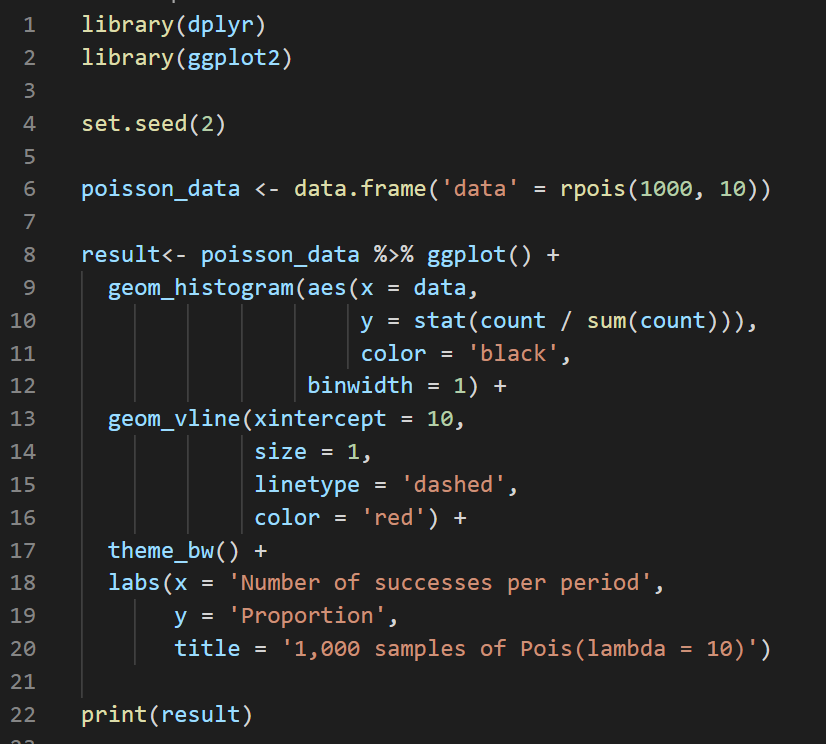


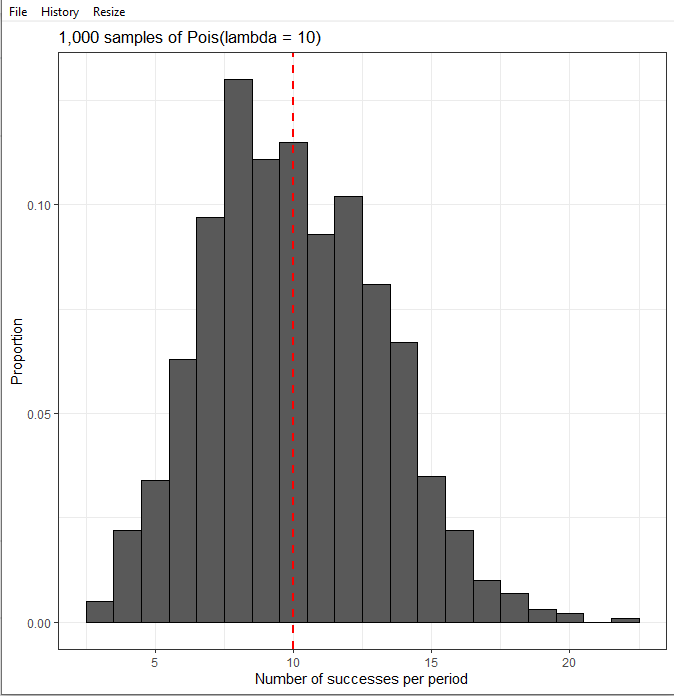
The output will be a single number and it will change every time you run that line.

This single observation isn’t very interesting on its own because there’s nothing we can say about it that hasn’t already been said.

So, what we’ll do with rpois() is generate some data that we can use to plot this distribution so we can see what it looks like. We’re going to generate 1,000 random observations with the same value for lambda.

* Type the following in Rpois.R file



Output: 

**You can visit the following website:** <https://www.datacamp.com/community/tutorials/pipe-r-tutorial> to learn more about the **pipe operator %>%**

**Ex2:**

Data from the maternity ward in a certain hospital shows that there is a historical average of 4.5 babies born in this hospital every day. What is the probability that 6 babies will be born in this hospital tomorrow?

First, let’s calculate the theoretical probability of this event using dpois(). The number of successes we’re considering is 6, so we will set x = 6. Additionally, this historical average of 4.5 babies per day is our value for lambda, so we will set lambda = 4.5.

* In Module10 folder, create Dpois.r file
* Type the following in Dpois.r file

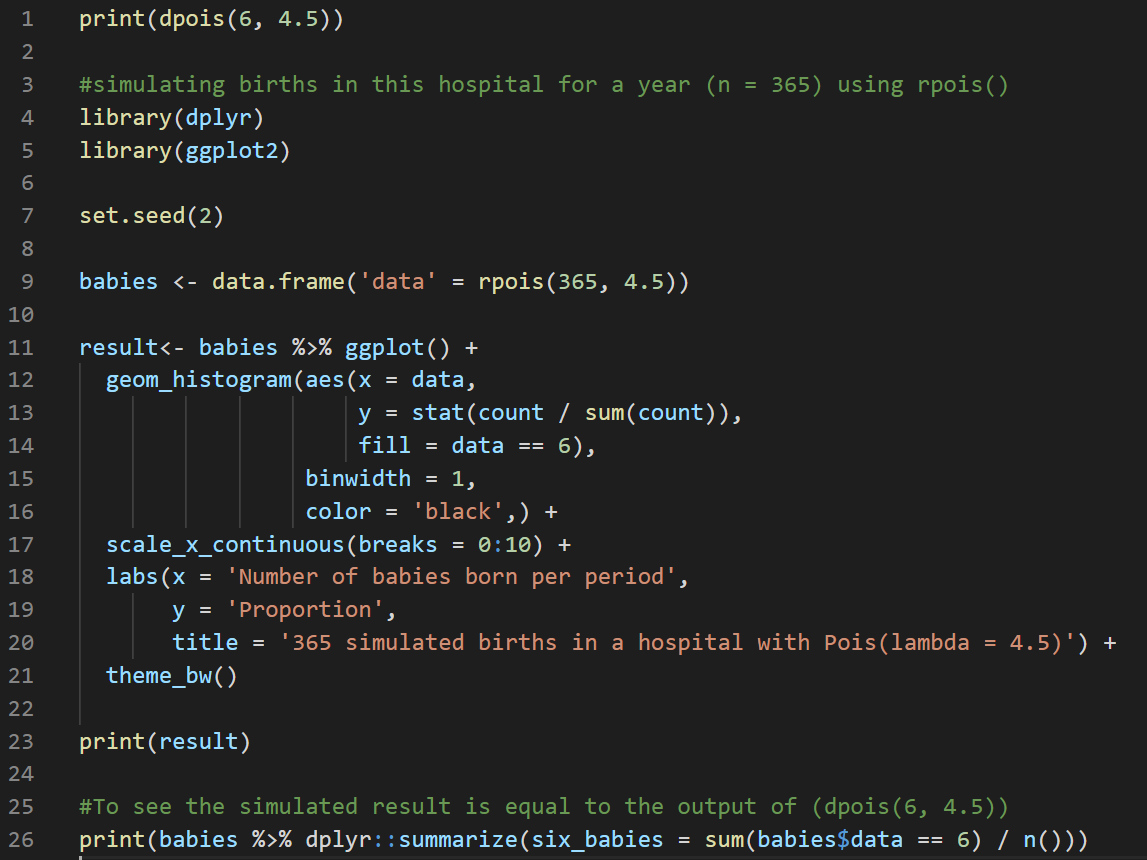


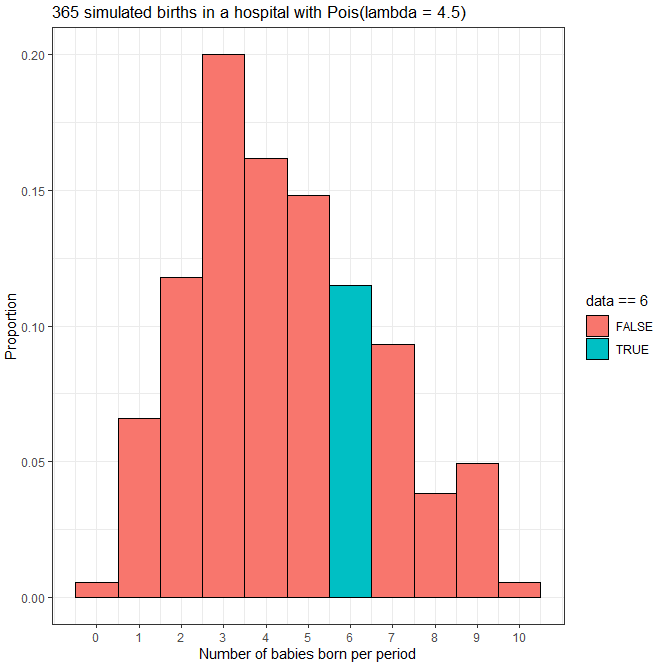
Output: 

The output means that the theoretical probability of 6 babies being born tomorrow if the historical average is 4.5 is about 13%.

Now let’s try simulating births in this hospital for a year (n = 365) using rpois() and compare the proportion of days in which there were 6 births to the theoretical probability we calculated above. We will also visualize this result.

* Type the following in Dpois.r file



Output: 



The simulated result of about 11.5% is pretty close to our theoretical probability of about 13%.

**Ex3:**

Data from the maternity ward in a certain hospital shows that there is a historical average of 4.5 babies born in this hospital every day. What about the probability of more than 6 babies being born?

* In Module10 folder, create Ppois.r file
* Type the following in Ppois.r file

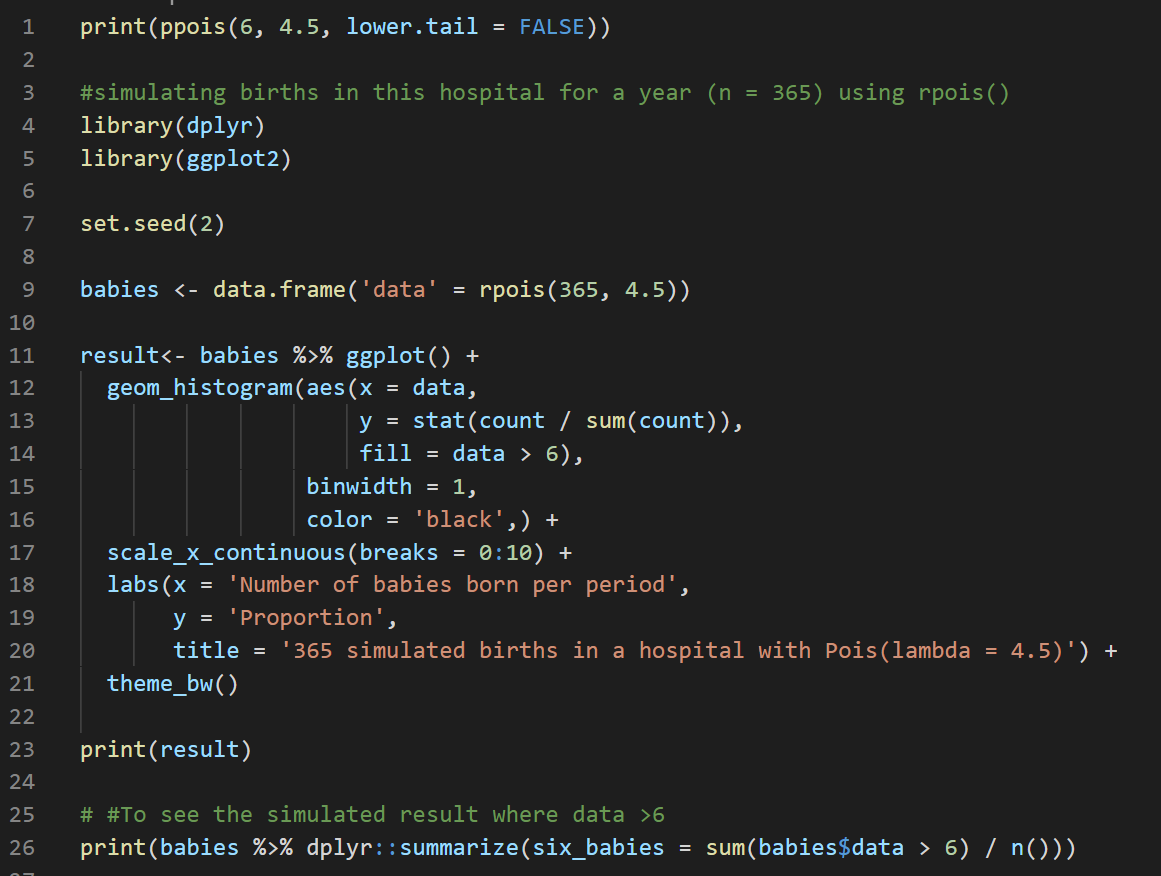


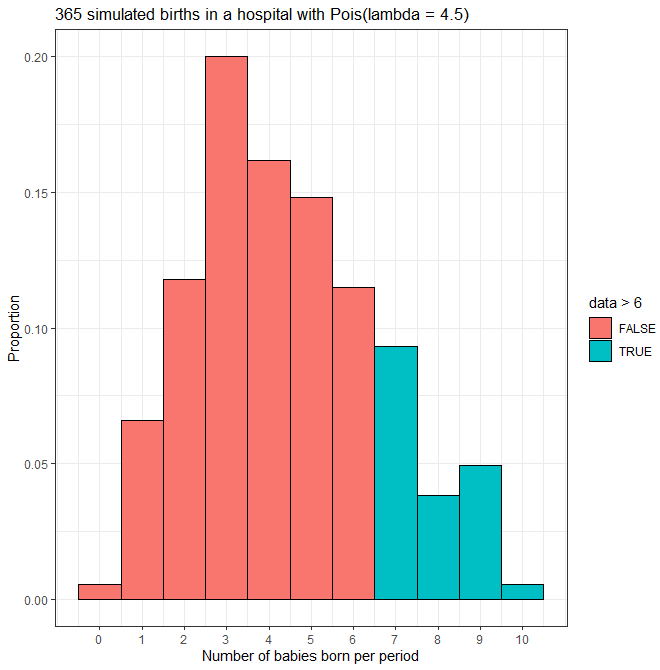
Output: 

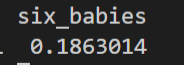
This theoretical probability is about 16.9%. Remember that cumulative probability functions in R calculate P(X > x) when lower.tail = FALSE. Here it calculated P(X > 6) = P(X >= 7).

Now let’s try simulating births in this hospital for a year (n = 365) using rpois() and compare the proportion of days in which there were more than 6 births to the theoretical probability we calculated above. We will also visualize this result.

* Type the following in Ppois.r file



Output: 



The simulated proportion of about 18.6% is pretty close to the theoretical proportion we calculated above.

**Push your work to GitHub**

**Make sure you are in**

Onsite students: CS251\_ Fall \_2020/**IN**/FirstnameLastname

Online students: CS251\_ Fall \_2020/**ON**/FirstnameLastname

Run the following commands to push your work to the GitHub repository:

Open the terminal from the VSCode by hit the **control + ~** key and type the following command:

>>> git add .

>>> git commit -m “Submission for Module 10”

>>> git push origin YOUR\_BRANCH\_NAME

Note: you should change the YOUR\_BRANCH\_NAME to your own branch name. It should be firstname-lastname (e.g. maria-gracia).