

Nuevos desafíos: El uso de R en cursos de grado

Adriana Pérez, M. Soledad Fernández, José Crespo, Martín Graziano,
Gerardo Cueto, Adelia González Arzac, Alejandra Pérez,
Vanessa Fernández, M. José Corriale, Oscar Lehmann
Universidad de Buenos Aires, Argentina

INTRODUCCIÓN

- Hoy en día no se pone en discusión la necesidad de incluir un software estadístico en los cursos de grado de estadística. Sin embargo, no existe el mismo consenso entre el uso de programas con una interfaz amigable versus programas basados en un **lenguaje de programación**, como R (R Core Development Team, 2017).
- Numerosas evidencias indican que las **simulaciones** promueven el aprendizaje activo y facilitan el abordaje de conceptos estadísticos abstractos (Mills, 2002, Batanero, 2009). El uso de simulaciones es sumamente sencillo en R
- Los métodos basados en la **expectativa** (como efectuar una pregunta a los estudiantes antes de su aprendizaje) han mostrado ser eficaces en promover el cambio cognitivo (Liu, 2010)
- En el presente trabajo proponemos una **actividad de anticipación y simulación computacional** desarrollada en entorno R. Aborda el concepto de variabilidad por muestreo y el desempeño de una prueba en relación al error tipo I y II bajo distintos escenarios, y está destinada a un curso de estadística inferencial. Sus objetivos son: facilitar la comprensión de conceptos estadísticos complejos mediante simulaciones y familiarizar al estudiante con el uso de R

MÉTODOS y RESULTADOS

El código completo de la actividad se encuentra disponible en <https://gist.github.com/adriana000/cf92afc4415d5a3d82ed309a38bd6e32> Solo es necesario el paquete base de R. Cada paso de la actividad es precedido por un espacio de discusión grupal sobre los resultados esperados, para luego contrastarlos con los resultados de las simulaciones. Aquí se muestran algunas líneas y resultados esperables

```
#1.Planteamos una variable X con distribución normal con media
30 y desvío estándar 10. ¿Qué aspecto se espera?
curve(dnorm(x,m=30,sd=10),from=0,to=60,main="distribución Normal")

#2.Simulamos una muestra de 20 observaciones de dicha distribución.
¿Qué aspecto se espera?
s1<-rnorm(n=20, mean = 30, sd = 10)

#3.Describimos la forma de la distribución. ¿Parece normal?
hist(s1)

#4.Estudiamos el ajuste a distribución normal mediante la prueba
de Shapiro-wilks (Ho: X ~ normal).¿Qué resultado esperamos?
shapiro.test(s1)

#5.Y si repetimos la simulación? ¿Qué esperamos?
s2<-rnorm(n=20, mean = 30, sd = 10)

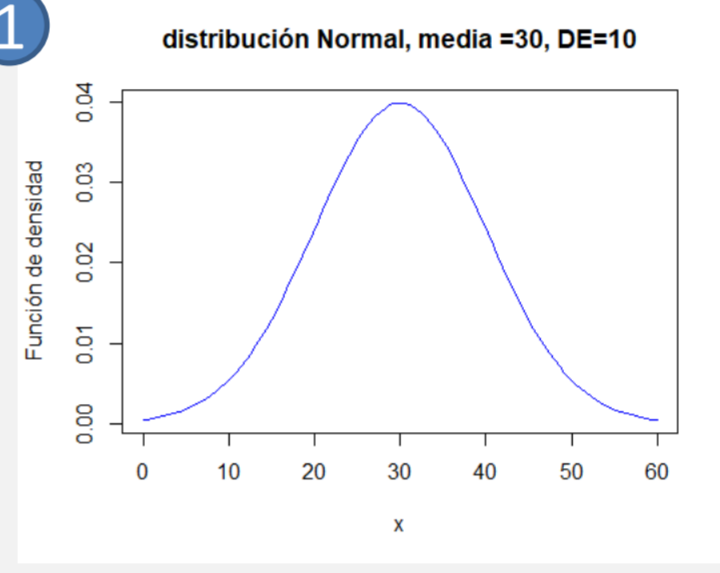
#6.¿Cómo repetir esto muchas veces (ie 1000)?
data<-data.frame()
for(i in 1:1000) {
  sim<-rnorm(20, 30, 10)
  test.normal<-shapiro.test(sim)
  p_Shapiro<-c(i, test.normal$p.value)
  data<-cbind(rbind(data, p_Shapiro))
}
colnames(data)<-c("simulacion", "p_Shapiro")

#7.¿Qué proporción de las veces esperamos rechazar
equivocadamente Ho? ¿Y qué ocurrió?
data$PH <- ifelse(data$p_Shapiro < 0.05,
"se rechaza normalidad", "no se rechaza normalidad")
tab<-table(data$PH) #tabla de frec
prop.table(tab) #tabla de prop

#8.¿Qué esperamos que ocurra si la distribución no es normal?
#Uniforme: runif(n, min, max)
#lognormal: rlnorm(n, mean, sd)

#9. ¿Y si en vez de 20 tenemos 10 datos? ¿Cómo cambian en cada caso:
#a) aspecto gráfico; b) error tipo I, c) error tipo II
```

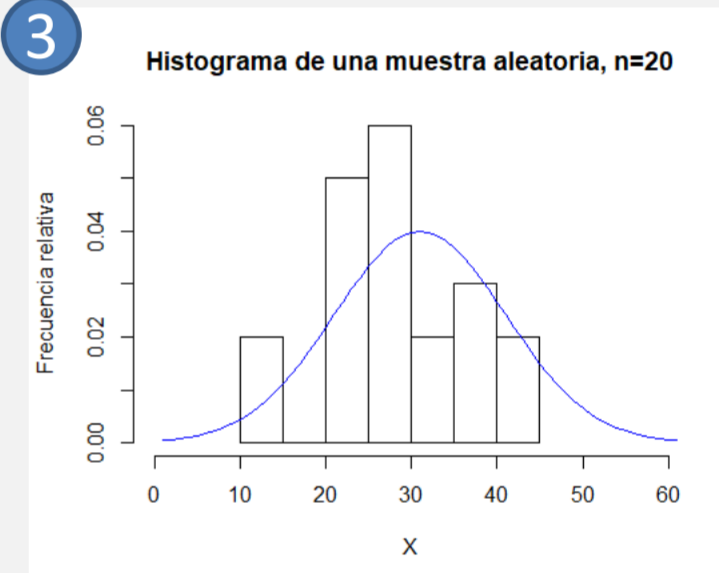
1 distribución Normal, media =30, DE=10



2

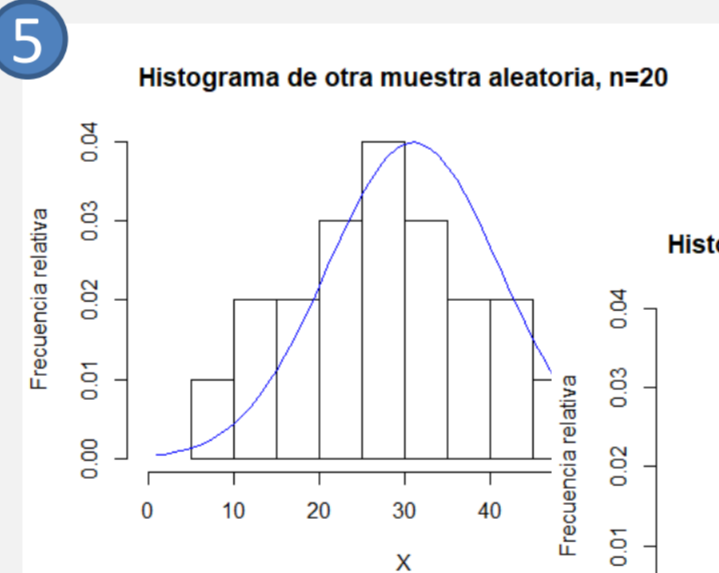
```
[1] 31.81 33.64 8.22 41.42
[5] 28.55 18.73 37.48 34.03
[9] 33.39 46.24 23.77 29.87
[13] 33.07 29.22 28.85 34.04
[17] 41.90 37.39 24.61 25.99
```

3 Histograma de una muestra aleatoria, n=20

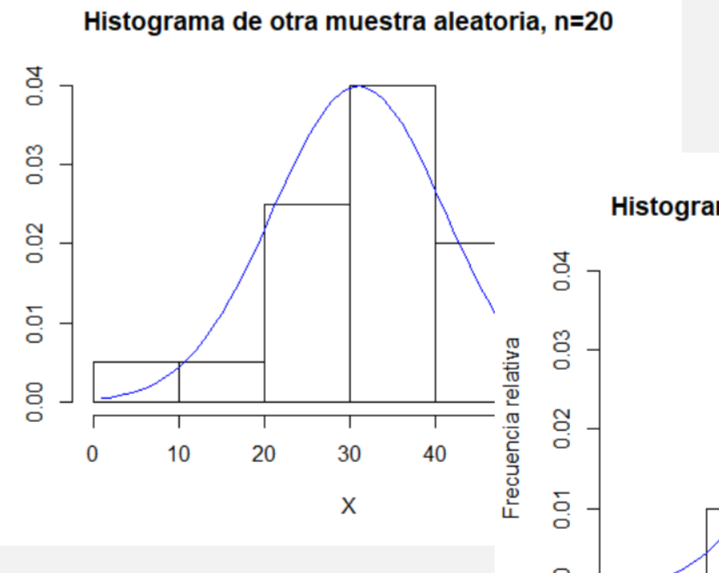


4 Shapiro-wilk normality test
data: s1
W = 0.96429, p-value = 0.6326

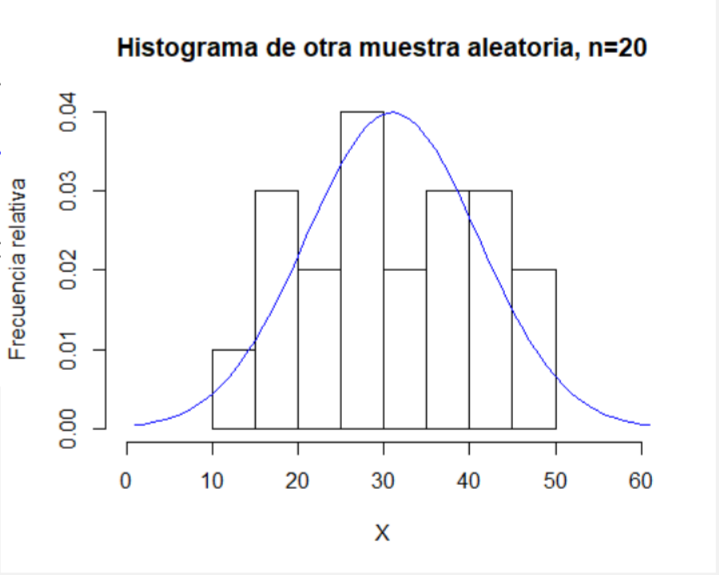
5 Histograma de otra muestra aleatoria, n=20



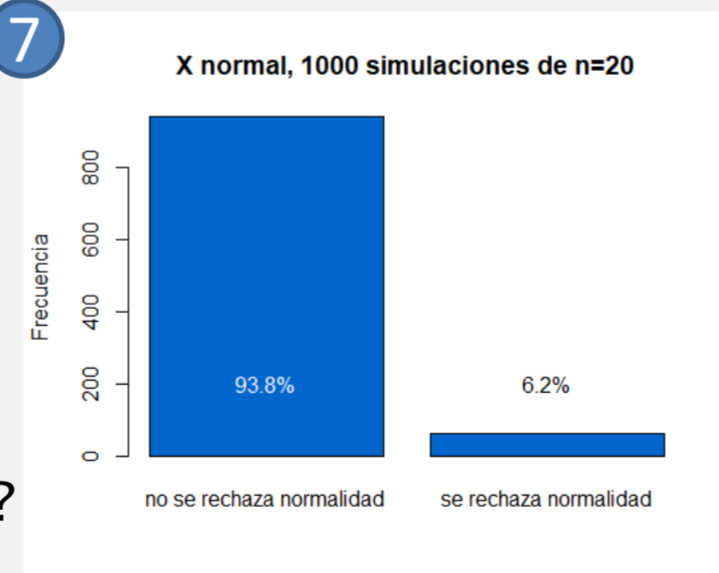
5 Histograma de otra muestra aleatoria, n=20



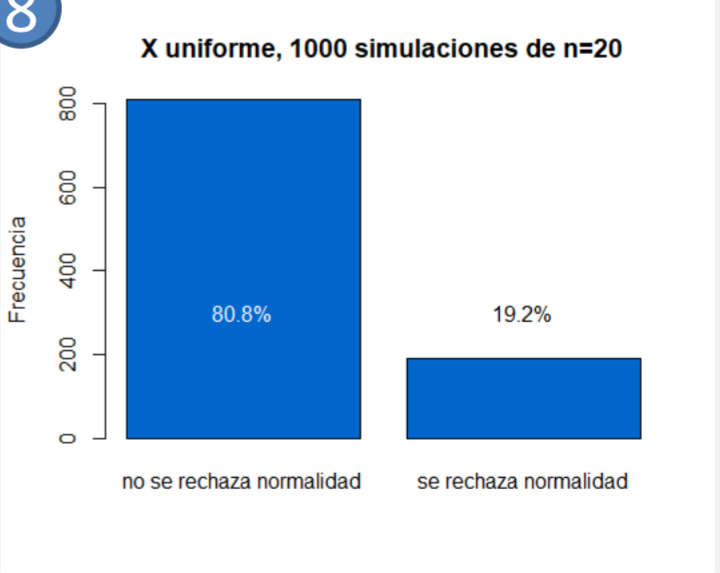
5 Histograma de otra muestra aleatoria, n=20



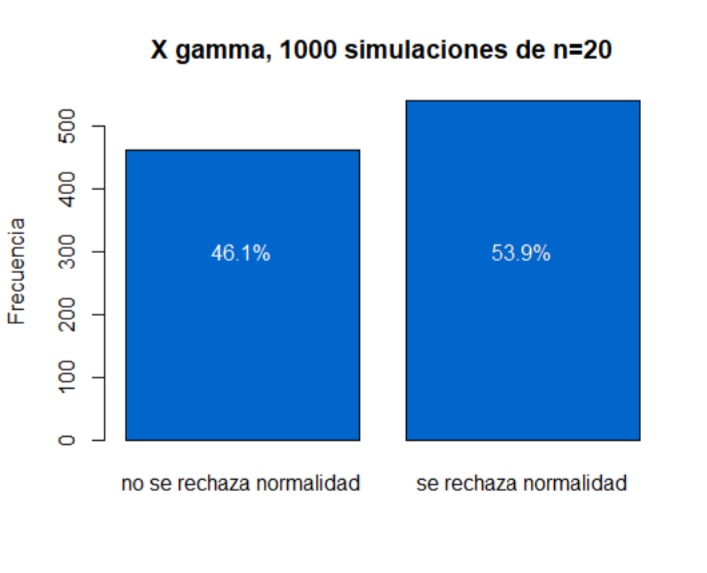
7 X normal, 1000 simulaciones de n=20



8 X uniforme, 1000 simulaciones de n=20



X gamma, 1000 simulaciones de n=20



CONCLUSIONES

- Se presentó una actividad que mediante la confrontación de ideas previas con simulaciones computacionales propicia el cambio conceptual y el aprendizaje activo de conceptos estadísticos complejos (variación por muestreo, error tipo I y tipo II, efecto sobre éstos del escenario alternativo a Ho y del tamaño muestral, entre otros)
- La actividad puede ser complementaria al enfoque tradicional basado en desarrollos teóricos
- Es sumamente flexible y puede ser adaptada a otras situaciones de aprendizaje
- Si bien R tiene una curva de aprendizaje más lenta que los programas con opciones cerradas, consideramos que los beneficios justifican su uso, particularmente en el uso de simulaciones

REFERENCIAS

Batanero, C. (2009). La simulación como instrumento de modelización en probabilidad. *Revista Educación y Pedagogía*, 15(35), 37-54.
Liu, T. C. (2010). Developing Simulation-based Computer Assisted Learning to Correct Students' Statistical Misconceptions based on Cognitive Conflict Theory, using "Correlation" as an Example. *Journal of Educational Technology & Society*, 13(2).

Mills J, D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1).
R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>