

Probabilidad y estadística con R

ldurazo

28/11/20

Esta libreta de R contiene una serie de ejercicios con el fin de practicar lo aprendido en la materia de matemáticas para ciencia de datos (probabilidad y estadística) en el lenguaje R.

Ejercicio con distribución binomial. Un equipo de desarrollo con un sistema de integración continua está experimentando fallas esporádicas en sus pruebas unitarias. Se sabe que hay una condición de carrera que aunque difícil de detectar, ocurre con una probabilidad de 7.5% en cada corrida.

- 1) Con qué probabilidad fallará 3 veces seguidas cuando Luis quiera subir sus cambios al repositorio, después de intentar subir 3 cambios?

$P(x = 3)$ si $n = 3$ y $p = 0.075$

```
# 1)
dbinom(3, 3, 0.075)
```

```
## [1] 0.000421875
```

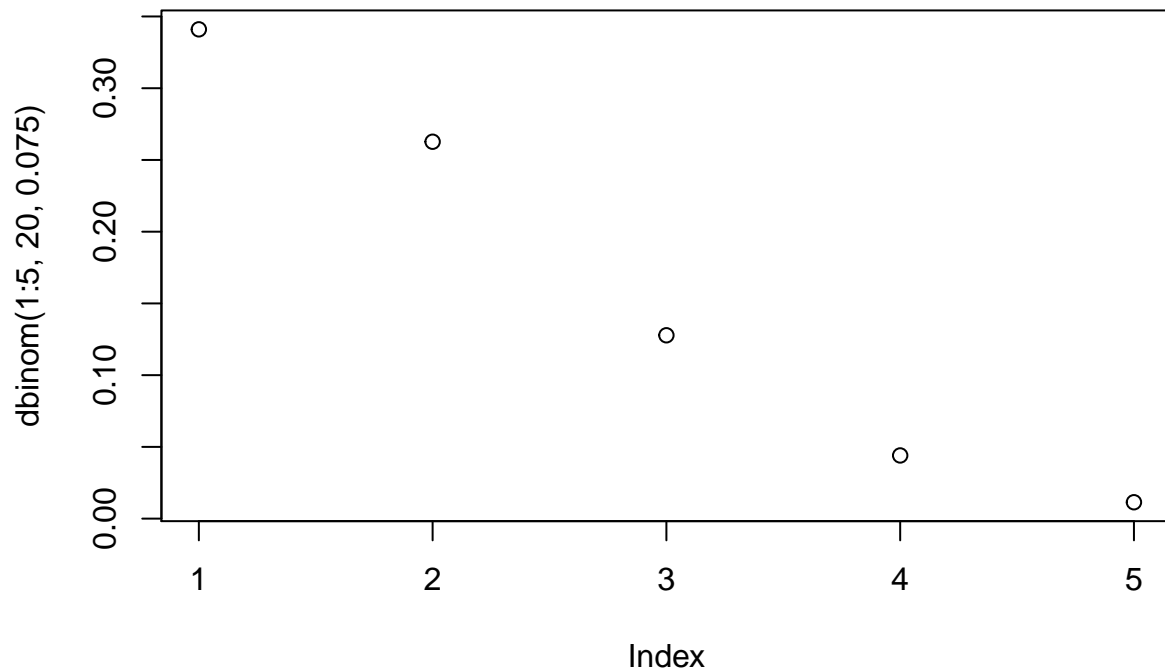
- 2) Con qué probabilidad fallará entre 1 y 5 veces seguidas cuando Luis quiera subir sus cambios al repositorio, después de intentar subir 20 cambios?

$P(1 \leq x \leq 5)$ si $n = 20$ y $p = 0.075$

```
# 2)
dbinom(1:5, 20, 0.075)
```

```
## [1] 0.34102340 0.26268019 0.12779036 0.04403587 0.01142552
```

```
plot(dbinom(1:5, 20, 0.075))
```



```
pbinom(0, 20, 0.075, lower.tail = FALSE) - pbinom(4, 20, 0.075, lower.tail = FALSE)
```

```
## [1] 0.7755298
```

- 3) Con qué probabilidad no fallará ninguna vez, si en promedio, Luis ejecuta 10 construcciones en el contenedor integración continua en cada ciclo de desarrollo diario?

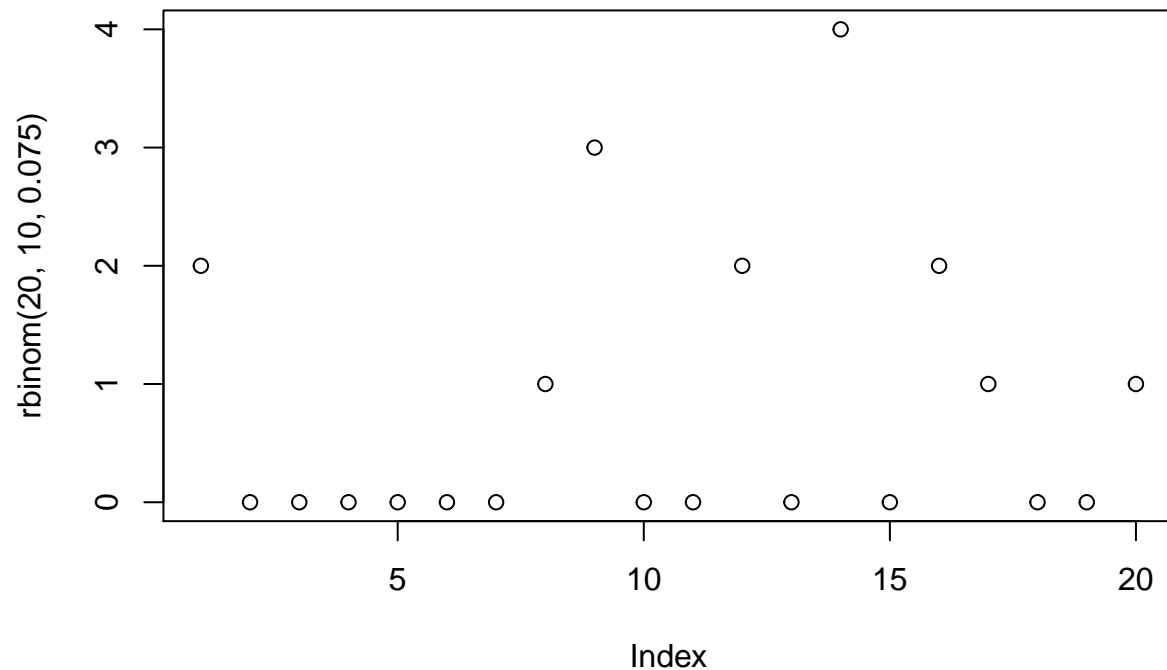
```
# 3)
1 - dbinom(0, 10, 0.075)
```

```
## [1] 0.5414177
```

- 4) Simula como se vería para desarrollador en un equipo de 20 personas de manera aleatoria si cada uno ejecuta 10 construcciones sobre integración continua por día, cuantas fallas podríamos observar?

Si queremos ver un único ejercicio de simulación, veamos el siguiente resultado:

```
plot(rbinom(20, 10, 0.075))
```



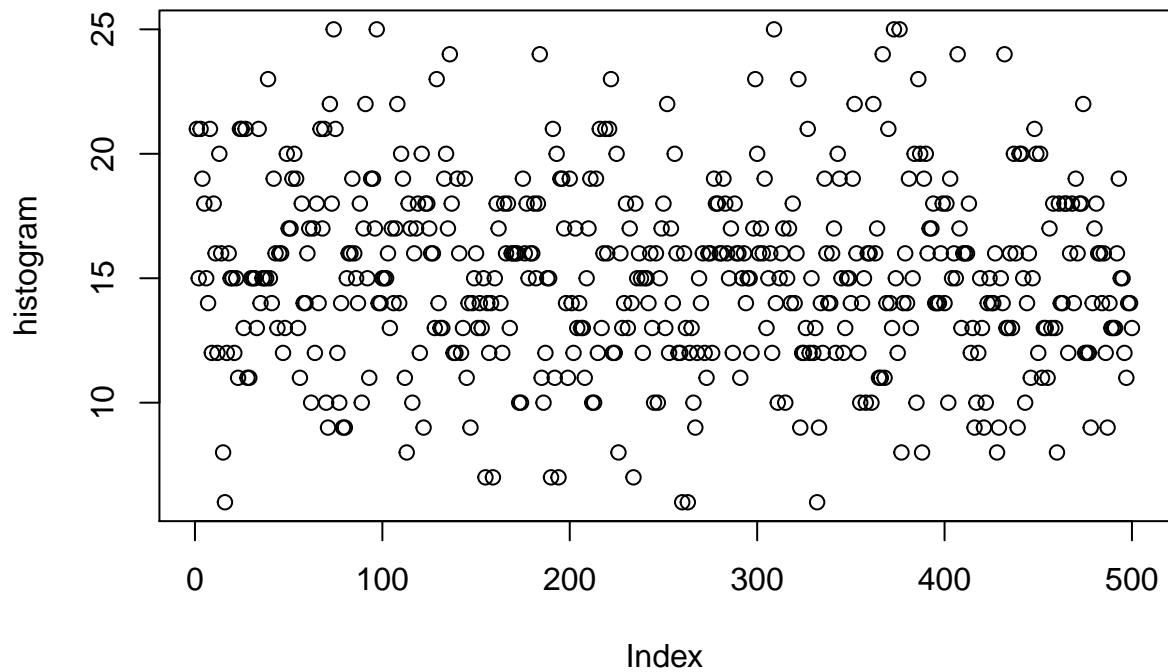
Sin embargo, esto representa una única simulación. Veamos que pasa si ejecutamos la simulación 500 veces, obtendremos un histograma del total de fallas por simulación.

nuestras m_1, m_2, \dots, m_{500} todas son iguales a 20. Nuestra $K = 200$. y la probabilidad persiste en 0.075.

```

histogram <- vector()
totalCount <- 0
for (i in 1:500) {
  eventCount <- 0
  rbinom_result <- rbinom(20, 10, 0.075)
  for (j in rbinom_result) {
    eventCount <- eventCount + j
  }
  histogram <- append(histogram, eventCount)
  totalCount <- totalCount + eventCount
}
plot(histogram)

```



Para un total en esta ejecución de:

```
print(totalCount)
```

```
## [1] 7557
```

```
print(totalCount / 500)
```

```
## [1] 15.114
```