

# Биостатистика и анализа података

Марија Цупарић

Пети час



# Геометријска случајна величина: $X \in \mathcal{G}(p)$

## Пример

- ★ Предатор у природи (нпр. птица) хвата плен (нпр. инсекте). Сваки покушај хватања је независан, а вероватноћа успешног хватања износи 0.8. Посматра се број успешних хватања пре него што предатор први пут не успе да ухвати плен.
- ★ Број гађања у мету пре него што се погоди

У свакој етапи имамо два исхода: „успех” и „неуспех”. Етапе су међусобно независне и вероватноћа успеха  $p$  је иста у свакој етапи. Бројимо експерименте док се не постигне први успех. Закон расподеле је

$$p(x; p) = P\{X = x\} = (1 - p)^{x-1}p, x = 1, 2, 3, \dots$$

- ★  $EX = \frac{1}{p}, DX = \frac{1-p}{p^2}$
- ★ може дефинисати и као број неуспеха док се не појави први успех и тада је закон расподеле

$$p(x; p) = P\{Y = x\} = P\{X = x + 1\} = (1 - p)^x p, x = 0, 1, 2, \dots$$

- ★ **R:** `rgeom(sample size, probability)` генерисање величине  $Y$



**Пример** Птица предатор лови инсекте. Сваки покушај хватања је независан, а вероватноћа успешног хватања износи 0.8.

- Колика је вероватноћа да први инсект који птица није успела да ухвати буде четврти по реду?
- Колика је вероватноћа да је птица успешно ухватила мање од 3 инсекта пре првог неуспеха?
- Одредити очекивану вредност и дисперзију броја успешно ухваћених инсеката пре првог неуспеха.

$Y$  — број успешно ухваћених инсеката пре првог промашаја,  $Y \in \mathcal{G}(0.2)$

- $P\{Y = 3\} = 0.8^3 \cdot 0.2 = 0.1024$
- $P\{Y < 3\} = P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\}$   
 $= 0.2 + 0.2 \cdot 0.8 + 0.2 \cdot 0.8^2 = 0.488$
- $EY = E(X - 1) = EX - 1 = \frac{1}{0.2} - 1 = 4,$   
 $DY = D(X - 1) = DX = \frac{1 - 0.2}{0.2^2} = 20$



```
> dgeom(3,0.2) #resenje dela a)
> pgeom(2,0.2) #resenje dela b)
```

Интерпретација резултата у претходном примеру:

- ★ Очекивана вредност је 4, што значи да очекујемо да ће птица ухватити 4 инсекта пре него што први пут промаши
- ★ Дисперзија је 20, односно стандардно одступање  $\sqrt{20} \approx 4.47$ , што је велика вредност.

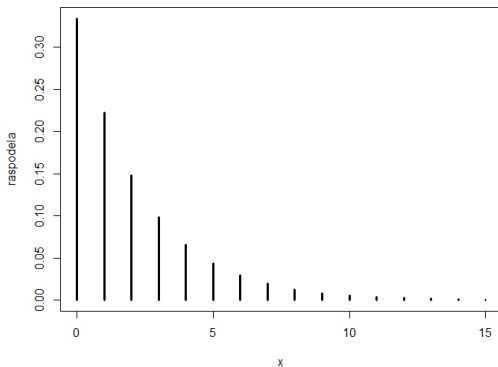
Иако просечно птица ухвати 4 инсекта, број успеха може доста да варира. Велика дисперзија проистиче из мале вероватноће успеха.

То значи да у пракси, у једном низу хватања, птица може ухватити много мање или много више инсеката од 4 пре него што први пут промаши.



R

```
> plot(0:15, dgeom(0:15, prob = 1/3), xlim = c(0, 15),
      type = 'h', axes = TRUE, xlab = "x",
      ylab="raspodela", lwd=3)
```



## Пример

- ★ Број заражених јединки у узорку популације
- ★ Број птица које прелете посматрачку тачку током једног сата

Број догађаја који се догоде у неком одређеном интервалу (времена) се често моделира Пуасоновом расподелом. Закон расподеле је

$$p(x; \lambda) = P\{X = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, 2, \dots$$

- ★  $EX = \lambda, DX = \lambda$
- ★ R: `rpois(sample size, parameter)`



**Пример** У једној популацији просечан број заражених јединки у узорку износи 4.

- а) Колика је вероватноћа да у узорку неће бити ниједне заражене јединке?
- б) Колика је вероватноћа да ће бити најмање 2 заражене јединке?
- в) Одредити очекивану вредност и дисперзију броја заражених јединки у узорку.

$X$  — број заражених јединки,  $X \in \mathcal{P}(4)$

а)  $P\{X = 0\} = \frac{4^0 e^{-4}}{0!} = e^{-4} \approx 0.0183$

б)  $P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - e^{-4} - \frac{4 \cdot e^{-4}}{1!} = 0.9084$

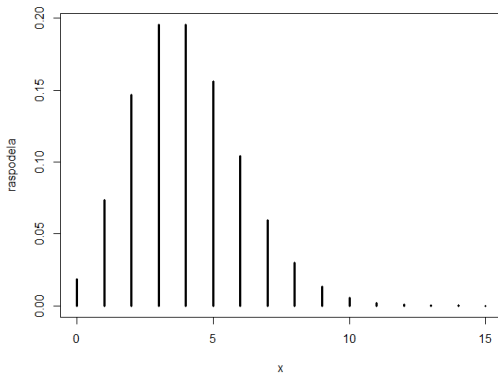
в)  $EX = 4, DX = 4$



```
> dpois(0,4) #resenje dela a)
> 1-ppois(1,4) #resenje dela b)
```

R

```
> plot(0:15, dpois(0:15, lambda = 4),xlim = c(0, 15), type =  
      'h', axes = TRUE, xlab = "x", ylab="raspodela", lwd=3)
```



**Пример** У једном језеру посматра се појава паразита на рибама. Из ранијих истраживања познато је да је просечан број риба заражених одређеним паразитом једнак 4 на 3000 уловљених риба. Одредити вероватноћу да ће се највише једна заражена риба појавити у узорку од 1500 риба.

$X$  – број заражених риба у узорку од 3000 риба,  $X \in \mathcal{P}(4)$

$X^*$  – број заражених риба у узорку од 1500 риба,  $X^* \in \mathcal{P}(2)$

$$\begin{aligned} P\{X^* \leq 1\} &= P\{X^* = 0\} + P\{X^* = 1\} \\ &= e^{-2} + 2e^{-2} \approx 0.406 \end{aligned}$$



```
> ppois(1, lambda = 2)
[1] 0.4060058
```

Одредити број заражених риба  $c$  за који је вероватноћа да ће их бити најмање  $c$  у узорку од 3000 риба једнака 0.2.

$$\begin{aligned} P\{X \geq c\} &= 0.2 \iff P\{X < c\} = 0.8 \\ &\iff P\{X \leq c-1\} = 0.8 \\ &\iff F(c-1) = 0.8 \\ &\iff \sum_{k=0}^{c-1} \frac{4^k e^{-4}}{k!} \geq 0.8 \end{aligned}$$



```
qpois(0.8, lambda = 4)
[1] 6
```

$$c-1 = 6 \implies c = 7$$

Приметимо да `ppois(6, lambda = 4)` даје као резултат 0.889326 (а не 0.8)!

# Рачунање биномних вероватноћа преко Пуасонових

- ★ Уколико је  $n$  велико, а  $p$  такво да је  $np \leq 5$ , биномне вероватноће могу приближно да се израчунају коришћењем Пуасонових

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k e^{-np}}{k!}$$

**Пример** У једној популацији јединки процењује се да је вероватноћа да случајно изабрана јединка буде заражена одређеном болешћу  $p = 0.002$  (0.2%). Колика је вероватноћа да су три насумично изабране јединке заражене?

$X$  – број заражених јединки,

$X \in \mathcal{B}(1000, 0.002)$

$np = 2 < 10 \implies X \sim \mathcal{P}(2)$

$$\begin{aligned} P\{X = 3\} &= \binom{1000}{3} 0.002^3 \cdot 0.998^{997} \\ &\approx \frac{2^3 e^{-2}}{3!} = 0.1804 \end{aligned}$$

```
> dbinom(3,1000,0.002)
[1] 0.1806277
> dpois(3,2)
[1] 0.180447
```

# Расподела (апсолутно) непрекидне случајне величине

- ★ Непрекидна случајна величина може узети било коју вредност из неког интервала или уније интервала на реалној правој, а појединачне вредности узима са вероватноћом 0.

**Дефиниција:** Случајна величина  $X$  је апсолутно непрекидна случајна величина ако постоји ненегативна функција  $f$  дефинисана на  $\mathbb{R}$  таква да се функција расподеле  $F$  случајне величине  $X$  може представити у облику

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}.$$

Функција  $f$  се назива функција густине случајне величине  $X$ .

Ако  $f(x)$  представимо као једну криву,  $F(x)$  је заправо површина испод криве до тачке  $x$ .



Из дефиниције следи да је

$$\star f(x) = \frac{dF(x)}{dx}$$

$$\star \int_{-\infty}^{\infty} f(x) dx = 1 = F(\infty) - F(-\infty), \text{ односно целокупна површина испод криве је једнака } 1;$$

$$\star \text{ Вероватноћа да } X \text{ узме вредност између било које две вредности } a \text{ и } b, P\{a < X < b\} \text{ једнака је површини испод графика функције } f \text{ од } a \text{ до } b$$

$$\begin{aligned} P\{a < X < b\} &= P\{a \leq X < b\} \\ &= P\{a < X \leq b\} \\ &= P\{a \leq X \leq b\} \\ &= F(b) - F(a) \\ &= \int_a^b f(x) dx \end{aligned}$$



**Пример.** Претпоставимо да је први позив у полицијску станицу упућен у произвољном тренутку између 7:30 и 8:00 часова. Одредити вероватноћу да је он приспео између 7:35 и 7:45.

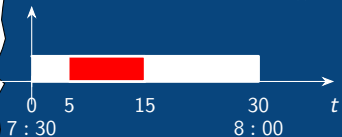
$X$  - време првог позива



свака тачка из интервала  $(0,30)$  једнако вероватна  $\implies$  густина се може представити као права линија која формира правоугаоник са једном страницом 30, а друга страница се одређује тако да површина буде 1

$$\int_0^{30} f(x) dx = 1 \iff 30a = 1 \iff a = \frac{1}{30} \iff f(x) = \begin{cases} \frac{1}{30}, & x \in (0, 30), \\ 0, & \text{иначе.} \end{cases}$$

$$\implies F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < 0, \\ \int_0^x \frac{1}{30} dt = \frac{x}{30}, & x \in [0, 30), \\ \int_0^{30} \frac{1}{30} dt = 1, & x \geq 30. \end{cases}$$



$$P\{5 < X < 15\} = \int_5^{15} \frac{1}{30} dt = \frac{10}{30}.$$

Ако постоји функција две променљиве тако да за свако  $(x, y) \in \mathbb{R}^2$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du,$$

тада је  $f$  заједничка густина непрекидног случајног вектора  $(X, Y)$ .

- ★ Све особине густине се преносе и на дводимензиони случај.
- ★ Маргинална густина  $f_X(x) = \int_{-\infty}^{\infty} f(x, v) dv$

**Дефиниција:** Случајне величине  $X$  и  $Y$  су независне ако важи

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}, \quad \forall x, y.$$

Еквивалентно, независност се може испитивати и преко следеће две формуле

$F(x, y) = F_X(x)F_Y(y)$ ,  $F_X, F_Y$  маргиналне функције расподела

$f(x, y) = f_X(x)f_Y(y)$ ,  $f_X, f_Y$  маргиналне густине.



# Математичко очекивање апсолутно непрекидне случајне величине

Нека је случајна величина  $X$  задата густином расподеле  $f(x)$ . Математичко очекивање случајне величине  $X$  је

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Ако важи  $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ , онда  $EX$  постоји.

- ★ очекивање се може видети као тачка баланса графика
- ★ код симетричних расподела та тачка ће се налазити на средини графика (оса симетрије графика)
- ★ особине очекивања су исте као у дискретном случају

Ако је  $h$  нека функција, тада

$$Eh(X) = \int_{-\infty}^{\infty} h(x)f(x)dx.$$



# Дисперзија апсолутно непрекидне случајне величине

Нека је случајна величина  $X$  задата густином расподеле  $f(x)$ . Дисперзија случајне величине  $X$  је

$$D(X) = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx.$$

- ★ параметар облика
- ★ ако већина тачака тежи да буде близу средине онда је дисперзија мала у супротном је велика
- ★ ако је дисперзија велика, онда су тачке на графику раширене
- ★ особине дисперзије су исте као у дискретном случају



**Пример.** Дата је функција  $f(x) = \begin{cases} a(1-x)^2, & x \in [0, 1]; \\ 0, & \text{иначе.} \end{cases}$

а) Израчунати константу  $a$  за коју је  $f(x)$  густина расподеле (вероватноћа) неке случајне величине  $X$ .

б) Одредити функцију расподеле те случајне величине.

а)

$$1) \ a > 0$$

$$\implies a = 3$$

$$2) \ 1 = \int_0^1 a(1-x)^2 dx = a \int_0^1 t^2 dt = a \frac{1}{3}$$

б)

$$F(x) = \begin{cases} 0, & x < 0, \\ \int_0^x 3(1-t)^2 dt, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases} = \begin{cases} 0, & x < 0, \\ -(3x - 3x^2 + x^3), & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$



в) Израчунати  $P\{X > \frac{1}{3}\}$ .

г) Одредити очекивање и дисперзију те случајне величине.

$$в) P\{X > \frac{1}{3}\} = 1 - P\{X \leq \frac{1}{3}\} = 1 - F_X(\frac{1}{3}) = \frac{8}{27}$$

Други начин:

$$P\left\{X > \frac{1}{3}\right\} = \int_{\frac{1}{3}}^1 f(t) dt = \int_{\frac{1}{3}}^1 3(1-t)^2 dt = \int_0^{\frac{2}{3}} 3u^2 du = \frac{8}{27}$$

$$г) EX = \int_0^1 x \cdot 3(1-x)^2 dx = \frac{1}{4}$$

$$DX = E(X - EX)^2 = E\left(X - \frac{1}{4}\right)^2 = \int_0^1 \left(x - \frac{1}{4}\right)^2 3(1-x)^2 dx = \frac{3}{80}$$



# Униформна случајна величина

## $X \in \mathcal{U}(a, b), a < b$

### Пример

- ★ Време првог позива у полицијској станици у првих пола сата радног времена (7:30, 8:00)
- ★ Претпоставимо да биљке једне врсте достижу одређену висину. Због природних варијација (количина светлости, воде и хранљивих материја),  $X$  представља стварну висину биљке која варира унутар интервала типичних вредности за ту врсту.

★ густина  $f(x; a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{иначе.} \end{cases}$

★ функција расподеле  $F(x; a, b) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & x \in [a, b], \\ 1, & x > b. \end{cases}$

★  $EX = \frac{a+b}{2}, DX = \frac{(b-a)^2}{12}$

★ користи се када све тачке интервала имају једнаку вероватноћу

★ R: `runif(sample_size, parameters)`

# Експоненцијална случајна величина

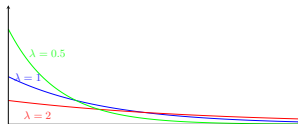
## $X \in \mathcal{E}(\lambda), \lambda > 0$

### Пример

- ★ Време преживљавања јединки у популацији код којих је вероватноћа угинућа иста у сваком тренутку (нема ефекта старења)
- ★ Ако доласци јединки на посматрачку тачку (нпр. птице на хранилицу) наступају случајно и независно, онда су времена између долазака независне експоненцијалне случајне величине

### ★ густина

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{иначе.} \end{cases}$$



### ★ функција расподеле

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & \text{иначе.} \end{cases}$$



★  $EX = \frac{1}{\lambda}, DX = \frac{1}{\lambda^2}$

★ Ако  $X \in \mathcal{E}(\lambda)$ , онда  $Y = \lambda X \in \mathcal{E}(1)$ :

$$P\{Y \leq y\} = P\{\lambda X \leq y\} = P\left\{X \leq \frac{1}{\lambda}y\right\} = 1 - e^{-y}, y \geq 0.$$

★ одсуство памћења:  $P\{X > t + s | X > s\} = P\{X > t\}$  одакле  
 следи да  $X - s | X > s \in \mathcal{E}(\lambda)$

★ R: `rexp(sample size, parameter)`



**Пример.** Време преживљавања јединки једне врсте у природи (у година-ма) има експоненцијалну  $\mathcal{E}(1)$  расподелу.

Одредити вероватноћу да јединка преживи између 1 и 3 године.

$$P\{1 \leq X \leq 3\} = F(3) - F(1) = 1 - e^{-3} - 1 + e^{-1} \approx 0.318$$



```
> pexp(3,rate = 1) - pexp(1,rate = 1)
[1] 0.3180924
```

Одредити број година  $t$  за које је познато да је вероватноћа да јединка живи више од  $t$  година једнака 0.2.

$$P\{X > t\} = 0.2 \iff 1 - F(t) = 0.2 \iff F(t) = 0.8 \iff t = F^{-1}(0.8)$$



```
> qexp(0.8,rate = 1)
[1] 1.609438
```

# Нормална случајна величина

$$X \in \mathcal{N}(m, \sigma^2), m \in \mathbb{R}, \sigma^2 > 0$$

- ★ Једна од најважнијих расподела у науци, медицини и инжењерству.

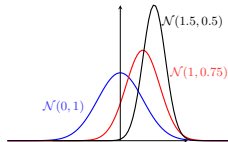
## Пример

- ★ Скоро свако физичко мерење има насумичне грешке које се моделују нормалном расподелом
- ★ Ако се посматра одређена особина у популацији (нпр. висина биљака или маса јединки), често се уочава да те вредности имају (приближно) нормалну расподелу.

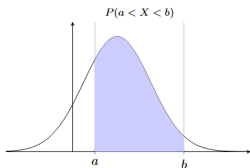
★ густина

$$f(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

★  $EX = m, DX = \sigma^2$



- ★ График сваке нормалне расподеле је симетрична, звонаста крива чија је средина једнака  $m$ , док дисперзија  $\sigma^2$  одређује облик криве



Површина осенченог дела једнака је

$$\begin{aligned} P\{a < X < b\} &= \int_a^b f(t) dt \\ &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(t-m)^2}{2\sigma^2}} dt \end{aligned}$$

Овај интеграл не може се одредити, већ се само за конкретне  $a$  и  $b$  може приближно израчунати

- ★ Ако  $X \in \mathcal{N}(m, \sigma^2)$ , онда  $Z = \frac{X-m}{\sigma} \in \mathcal{N}(0, 1)$  и она се назива стандардна нормална расподела:

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} = P\left\{\frac{X-m}{\sigma} \leq z\right\} \\ &= P\{X \leq m + z\sigma\} = F_X(m + z\sigma) \\ f_Z(z) &= F'_Z(z) = F'_X(m + z\sigma) \cdot \frac{d(m + z\sigma)}{dz} \\ &= F'_X(m + z\sigma) \cdot \sigma = \sigma \cdot f_X(m + z\sigma). \end{aligned}$$

- ★ R: `rnorm(sample size, parameters)`



## Читање таблица стандардне нормалне расподеле

Функција стандардне нормалне расподеле  $P\{Z < x\}$

| $x$ | 0.00   | 0.01   | 0.02   | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| ... | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| ... | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  | .....  |
| 3.8 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| 3.9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Ако је  $-4 < x < 4$ , онда се  $P\{Z < x\}$  чита из таблице

$$P\{Z < 1.75\} = P\{Z < 1.7 + 0.05\} = 0.9599$$

$$P\{Z < 0.39\} = P\{Z < 0.3 + 0.09\} = 0.6517$$



Случајна величине  $Z$  има стандардну нормалну расподелу

- ★ Вероватноћу догађаја  $P\{Z < x\}$  рачунамо функцијом `pnorm(x)`.

```
R > pnorm(1.25)
0.8943502
```

- ★ Вредност  $x$  за коју је  $P\{Z < x\} = p$  рачунамо функцијом `qnorm(p)`.

```
R > qnorm(0.95)
1.644854
```

- ★ Вредност густине расподеле у тачки  $x$ ,  $f(x)$ , рачунамо функцијом `dnorm(x)`.

```
R > dnorm(0.20)
0.3910427
```

- ★ Случајни узорак обима  $n$  (тј.  $n$  реализација експеримента случајне величине  $Z$ ) добијамо функцијом `rnorm(n)`.

```
R > rnorm(5)
-1.31419850 -0.03626391 0.91705320 0.61283009 -0.70323477
```



**Пример** Нека  $X$  означава висину биљака једне врсте (у cm). Претпоставимо да је висина нормално распоређена са средњом вредношћу 50 cm и стандардном девијацијом 0.2 cm.

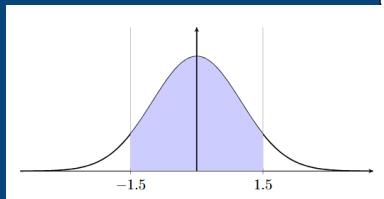
- а) Одредити вероватноћу да висина биљке одступа од средње вредности за више од 0.3 cm.

Пошто је  $X \in \mathcal{N}(50, 0.2^2)$ , онда  $Z = \frac{X-50}{0.2} \in \mathcal{N}(0, 1)$ .

Тражимо

$$\begin{aligned} P\{|X - 50| > 0.3\} &= P\left\{|Z| > \frac{0.3}{0.2}\right\} \\ &= P\{|Z| > 1.5\} \\ &= 1 - P\{-1.5 \leq Z \leq 1.5\} \\ &= 1 - P\{Z \leq 1.5\} + P\{Z < -1.5\} \\ &= 1 - P\{Z \leq 1.5\} + 1 - P\{Z \leq 1.5\} \\ &= 2(1 - P\{Z \leq 1.5\}) \\ &= 2(1 - 0.9332) = 0.1336 \end{aligned}$$

Дакле, око 13.36% висина биљака одступа за више од  $\pm 0.3$  cm од очекиване.

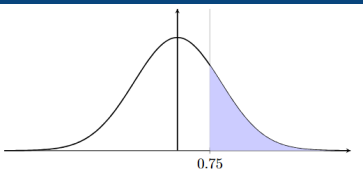


```
R > pnorm(1.5)
0.9331928
```

б) Одредити вероватноћу да висина биљке буде већа од 50.15 cm.

Треба одредити

$$P\{X > 50.15\} = P\left\{\frac{X - 50}{0.2} > \frac{50.15 - 50}{0.2}\right\} = P\{Z > 0.75\}.$$



```
> pnorm(0.75)
0.7733726
```

Дакле

$$P(X > 50.15) = 1 - 0.7734 = 0.2266.$$

У R-у се може рачунати и директно, без стандардизације (програм сам стандардизује):



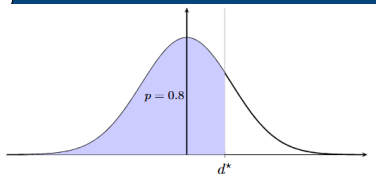
```
> 1-pnorm(50.15,50,0.2)
0.2266274
```

Дакле, око 22.66% висина биљака већа од 50.15cm.

- в) Наћи вредност  $d$  такву да висина биљке буде мања од  $d$  са вероватноћом 0.8.

Заправо треба одредити  $d$  такав да  $P\{X < d\} = 0.8$ .

$$P\left\{\frac{X - 50}{0.2} < \frac{d - 50}{0.2}\right\} = P\left\{Z < \frac{d - 50}{0.2}\right\} = 0.8,$$



```
> qnorm(0.8)
0.8416212
```

Дакле,  $d = 50 + 0.2 \cdot 0.8416212 \approx 50.168\text{cm}$ .

Или директно, без стандардизације:



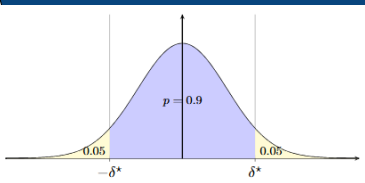
```
> qnorm(0.8, 50, 0.2)
50.16832
```

Дакле, са вероватноћом 0.8, висина биљке биће мањи од 50.168cm.

г) Колико одступање од висине 50cm не прелази 90% биљака?

Заправо треба одредити  $\delta$  такав да  $P(|X - 50| \leq \delta) = 0.9$ .

$$P\{|X - 50| \leq \delta\} = 0.9 \iff P\left\{|Z| \leq \frac{\delta}{0.2}\right\} = 0.9.$$



Са графика видимо да је  
 $P\{Z \leq \delta\} = 0.95$



```
> qnorm(0.95)  
1.644854
```

Дакле,  $\delta = 0.2 \cdot 1.644854 \approx 0.329\text{cm}$ .

Односно, 90% биљака имаће висину између 49.671cm и 50.329cm.