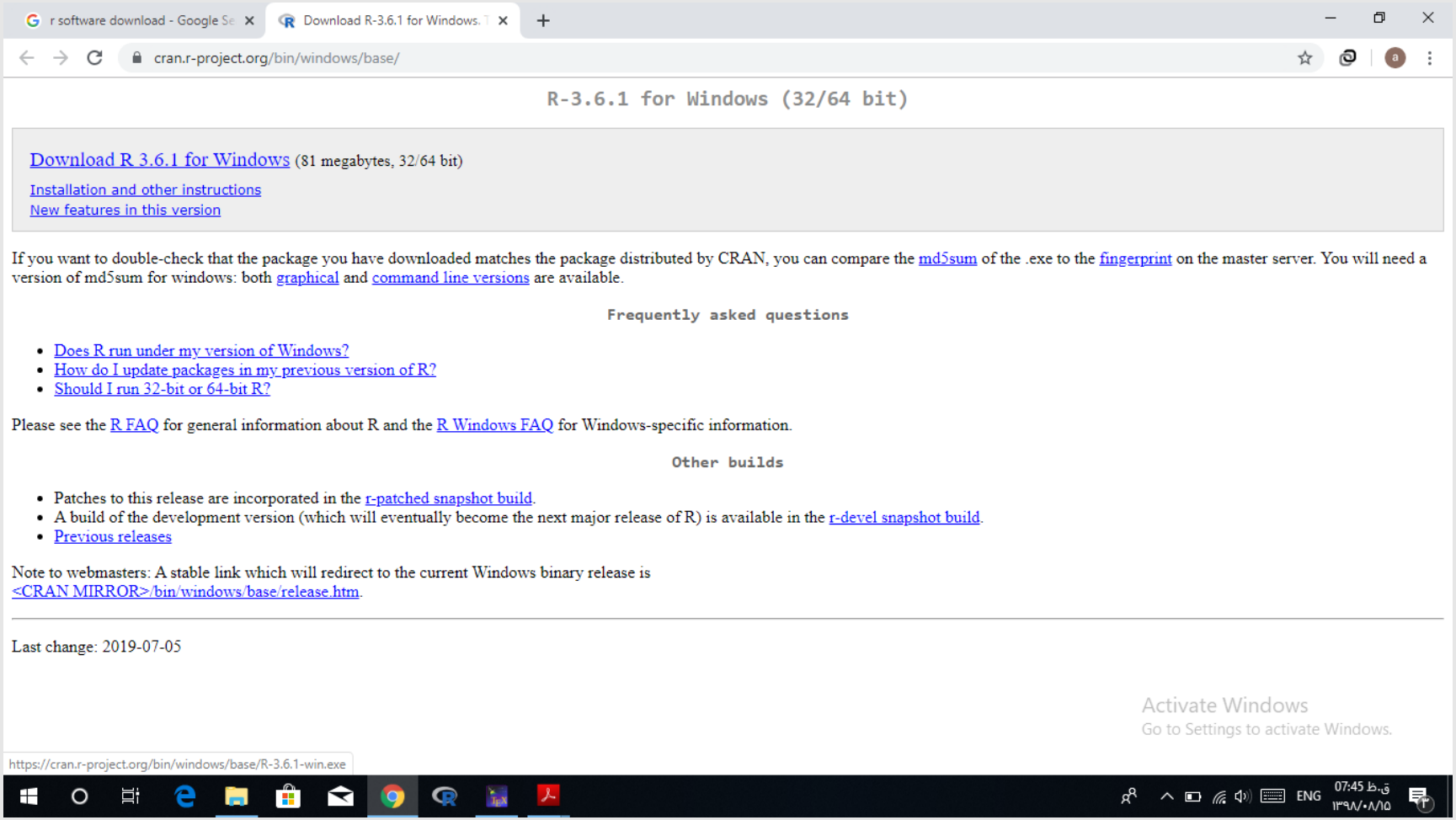
معرفی برخی توزیع‌های آماری در R

E-mail: a.saadatinik@stu.umz.ac.ir

University of Mazandaran

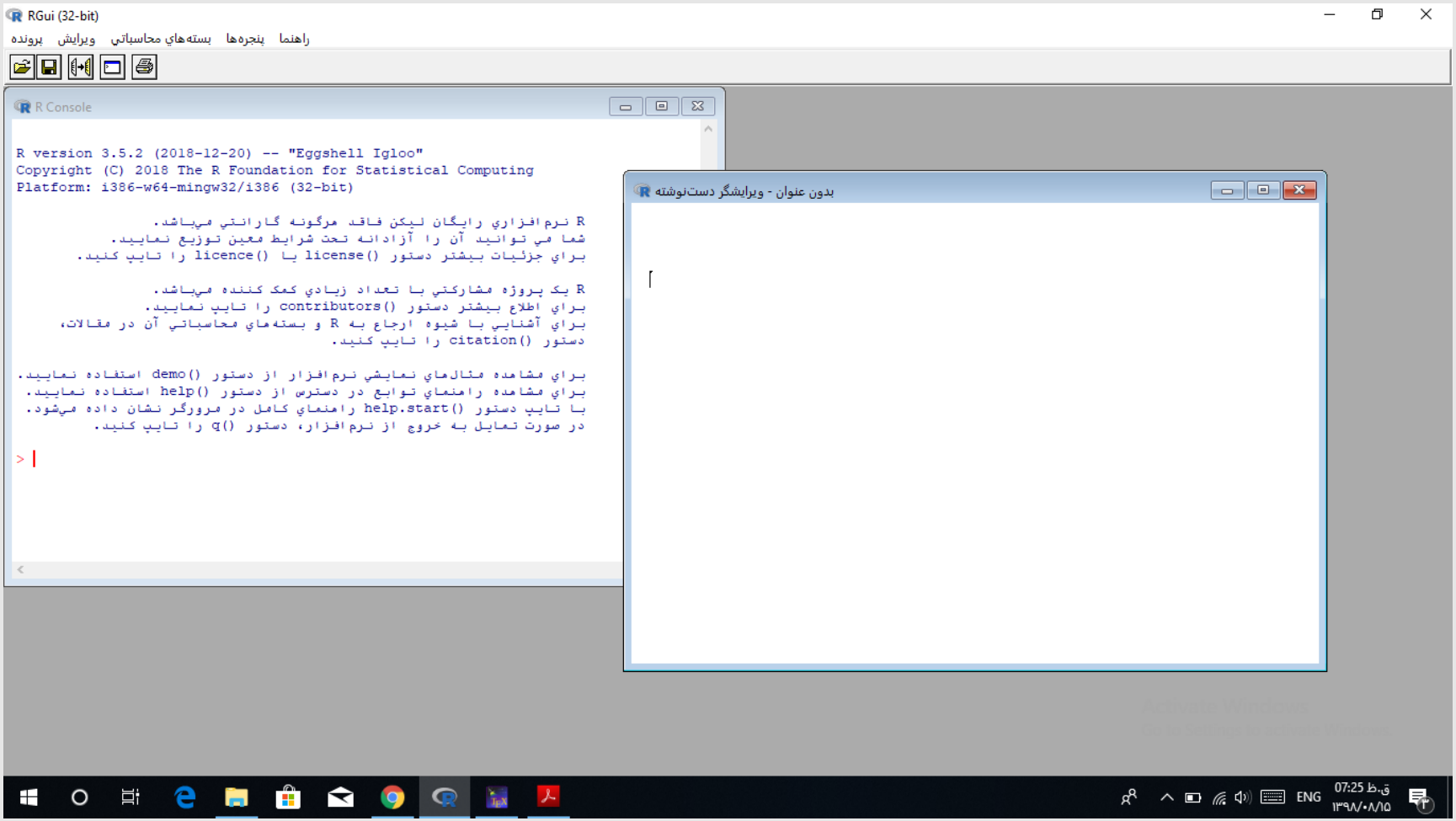
Statistical Analysis with
R Commander Training

نصب R

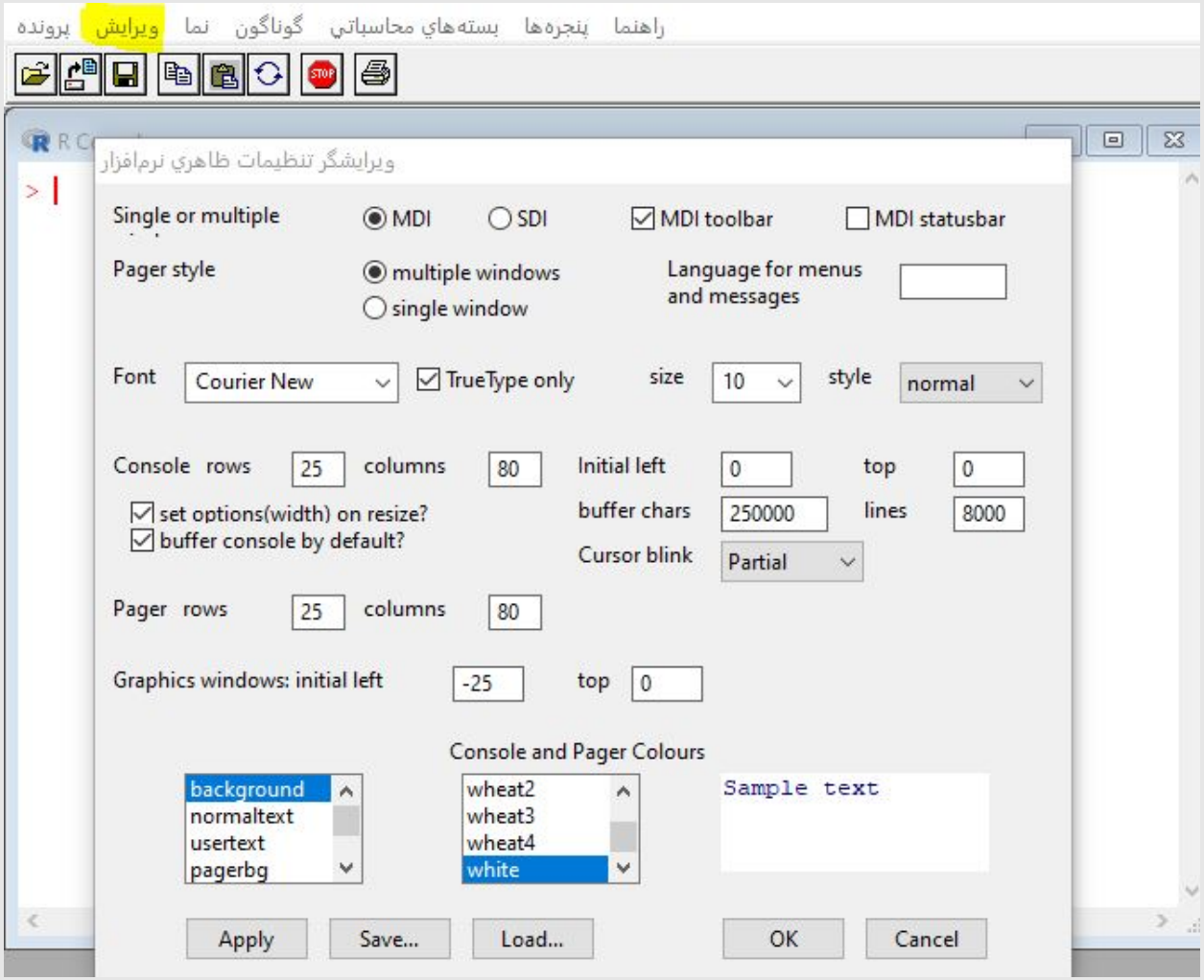


ابتدا فایل اجرایی نرم‌افزار را از آدرس سایت `https://cran.r-project.org/bin/windows/base/` دانلود کرده و با دو بار کلیک متوالی روی آن نصب می‌شود. پس از پایان عملیات نصب روی کامپیوتر شما یک آیکون به شکل حرف **R** قرار می‌گیرد. با دوبار کلیک روی آن، صفحه‌ای باز می‌شود که R Console نام دارد.

محیط R


RGui (32-bit)
پرونده ویرایش بسته‌های محاسباتی پنجره‌ها راهنما
R Console
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)
R نرم‌افزاري رايگان ليكن فاقد هرگونه گارانتي مي‌باشد.
شما مي‌توانيد آن را آزادانه تحت شرايط معين توزيع نماييد.
براي جزئيات بيشتر دستور license() يا licence() را تايپ كنيد.
R يک پروژه مشاركتي با تعداد زيادي كمک كننده مي‌باشد.
براي اطلاع بيشتر دستور contributors() را تايپ نماييد.
براي آشنايي با شيوه ارجاع به R و بسته‌هاي محاسباتي آن در مقالات،
دستور citation() را تايپ كنيد.
براي مشاهده مثال‌هاي نمايشي نرم‌افزار از دستور demo() استفاده نماييد.
براي مشاهده راهنماي توابع در دسترس از دستور help() استفاده نماييد.
با تايپ دستور help.start() راهنماي كامل در مرورگر نشان داده مي‌شود.
در صورت تمايل به خروج از نرم‌افزار، دستور q() را تايپ كنيد.
>
بدون عنوان - ویرایشگر دست‌نوشته
ENG
07:25 ق.ظ
۱۳۹۸/۰۸/۱۵

تنظیمات ظاهری:



چند نکته مهم در رابطه با زبان R

- نسبت به اندازه حروف حساس است.
- برای معرفی متغیرها میتوان از ترکیب حروف، اعداد و نمادهای "." و "_" استفاده نمود.
- اسم متغیرها نباید با اعداد و "_" شروع شوند.
- برای نسبت دادن یک اسم به یک شی میتوان از عملگرهای "="، "<-" یا "->" استفاده نمود.
- ابتدای هر عبارت در محیط R Console با خط فرمان ">" آغاز می‌شود و تا زمانیکه عبارت تکمیل نشده است، هر خط با علامت "+" شروع می‌شود.

زبان R به عنوان یک ماشین حساب قوی ریاضی و آماری

```
> 3+7
[1] 10

> 2*5+3
[1] 13

> 3 * (2 + 2)^2 - 5
[1] 7
```

ترتیب عملگرهای ریاضی به ترتیب به صورت پرانتز، توان، ضرب و تقسیم و جمع و تفریق می‌باشد.

نام دادن به داده‌ها

می‌توان داده‌ها را نام‌گذاری کرد. این نام‌گذاری متغیرها را در اختیار ما قرار می‌دهد.
بعضی نام‌ها قبلا در R ساخته شده‌اند:

```
> pi
[1] 3.141593

>  approx.pi <- 22/7
> approx.pi
[1] 3.142857


> a <- 2
> print(a)
[1] 2
```

بردارها: توابع c() و scan()

یک بردار، دنباله‌ای از مقادیر هم نوع است.

```
> x <- c(2, 7, 8, 10, 15, 30)
> x
[1]  2 7 8 10 15 30

> x=scan()
> 1: 2 7 8 10 15 30
> 7:
> Read 6 items
```

تابع $c()$ مقادیر یک بردار را با همان ترتیبی که وارد شده است بر می‌گرداند.

$x[i]$، iامین عنصر بردار x است، $x[c(i,\ j)]$، iامین و jامین عنصر x را نمایش می‌دهد، $x[-k]$ برداری است شامل همه عناصر بردار x بجز kامین عنصر. دستور $length(x)$ طول بردار x است:

```
> x[6]
[1] 30

> x[c(2, 3)]
[1] 7 8

> x[-2]
[1] 2 8 10 15 30

> length(x)
[1] 6
```

معرفی نماد

مجموعه تهی	NULL	c()
داده گمشده	NA	
مبهم	NaN	$\frac{0}{0}$
بینهایت	Inf	$\frac{1}{0}$

مثال:

```
x <- c(2, 5, 7, NA, 4, 5, 10)
[1] 2 5 7 NA 4 5 10

mean(x)
[1] NA

mean(x, na.rm=TRUE)
[1] 5.5
```

محاسبات برداری

عملگرها بر روی بردارها به‌صورت جفتی عمل می‌کنند:

```
> x <- c(2, 7, 8, 10, 15, 30)
> y <- c(3, 5, 8, 9, 11, 9)
> x+y
[1]  5 12 16 19 26 39

>  2*y
[1]  6 10 16 18 22 18
```

توابع بر روی بردارها

ورودی بسیاری از توابع، بردارها هستند.

$max()$ و $min()$ به ترتیب مقادیر ماکزیمم و مینیمم یک بردار را مشخص می‌کنند

$mean()$ و $median()$ به ترتیب میانگین و میانه یک بردار را مشخص می‌کنند

$sd()$ و $var()$ به ترتیب انحراف معیار و واریانس یک بردار را مشخص می‌کنند

$sum()$ مجموع یک بردار را مشخص می‌کند

توابع بر روی بردارها

تابع	شرح	مثال
$prod()$	حاصلضرب	$> prod(c(۲, ۵))$ $[۱]$ ۱۰
$sort()$	مرتب کردن اعداد بصورت صعودی	$> sort(c(۲, ۵, ۰))$ $[۱]$ ۰ ۲ ۵
$sqrt()$	ریشه دوم عدد	$> sqrt(۹)$ $[۱]$ ۳
$summary()$	یافتن میانه، حداقل، حداکثر، چارک اول و چارک سوم	–
$sum()$	مجموع	$sum(c(۲, ۳, ۵, ۷))$ $[۱]$ ۱۷

رسم نمودار در R

یکی از پرکاربردترین توابع رسم نمودار تابع plot() است که بصورت زیر استفاده می‌شود.

plot(x, y, type="", lwd=2, xlim="", ylim="", main="", pch=, col=)

شناسه type نوع نمودار را مشخص می‌کند:

p	l	c	O	H	s
نقطه	خط	خط بدون نقطه	خط و نقطه	خطوط عمودی	خطوط پلکانی

lwd: ضخامت خطوط
xlim: حدود محور x
ylim: حدود محور y
main: عنوان نمودار

pch: نوع نقاط (۱:۲۵)
xlab: برچسب محور x
ylab: برچسب محور y
col: رنگ نمودار (۱:۶۵۷)

تابع curve()

یکی دیگر از توابع رسم نمودار تابع curve() می‌باشد که برای رسم منحنی $y = f(x; .)$ استفاده می‌شود.

curve(f(x), from, to, ...)

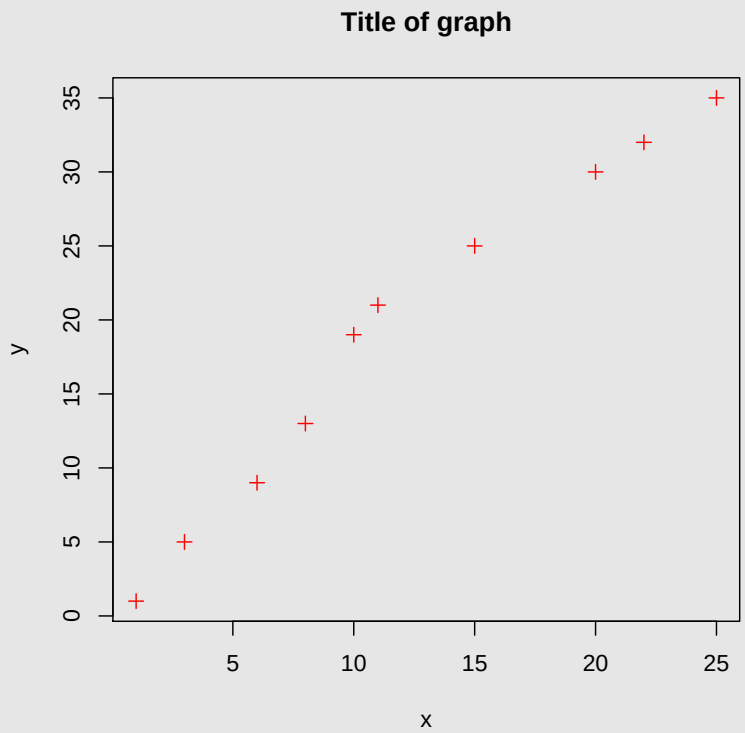
from: نقطه شروع محور x
to: نقطه پایان محور x

نکته: برای اضافه کردن منحنی $y = g(x)$ به منحنی قبلی از دستور زیر استفاده می‌شود:

curve(g(x), from, to, add=TRUE)

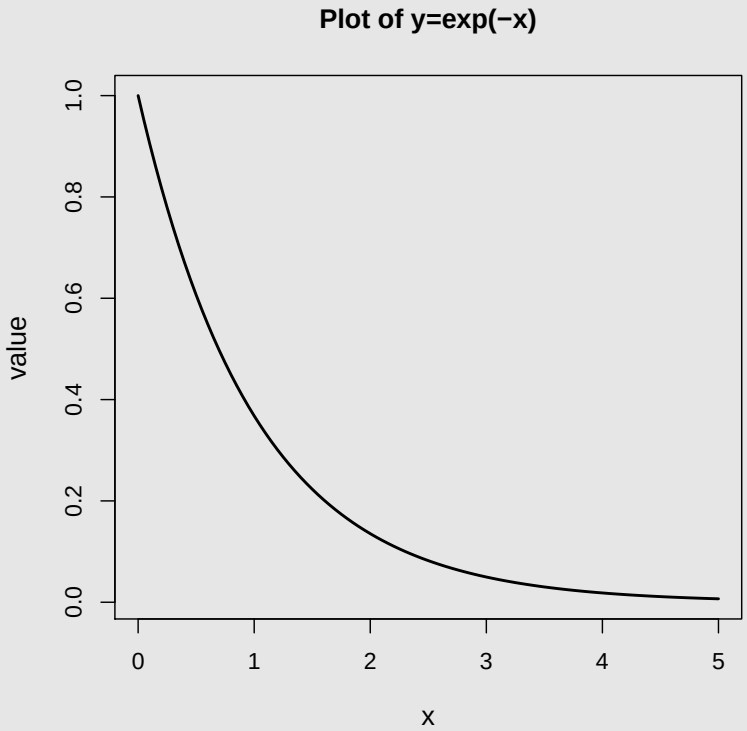
مثال.

```
> x=c(1, 3, 6, 8, 10, 11, 15, 20, 22, 25)
> y=c(1, 5, 9, 13, 19, 21, 25, 30, 32, 35)
> plot(x, y, col="red", pch=3, main="Title of graph")
```



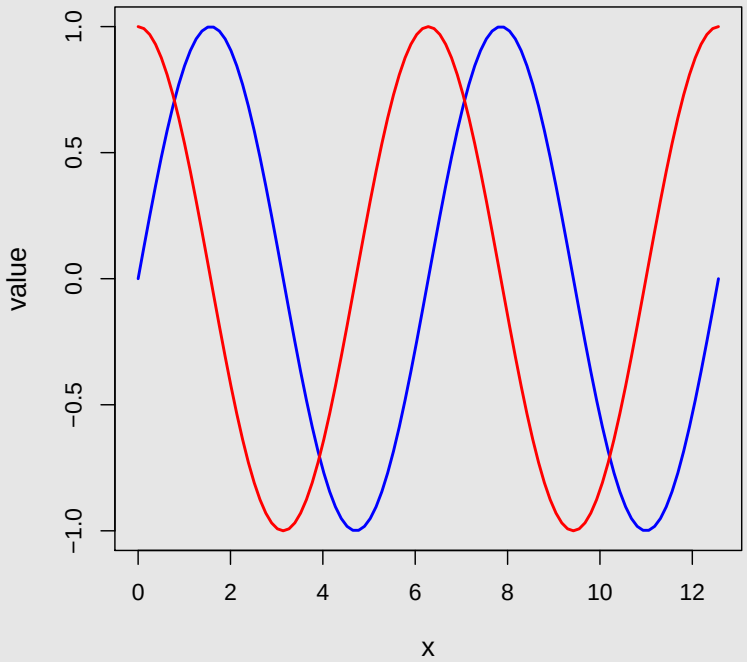
مثال. رسم منحنی $y = e^{-x}$ در بازه $(0, 5)$.

```
curve(exp(-x), 0, 5, type="l", lwd=2, ylab="value", main="Plot
of y=exp(-x)", cex.lab=1.2)
```



مثال. رسم منحنی $f(x) = \sin(x)$ و $g(x) = \cos(x)$ با استفاده از تابع curve در بازه $(0, 4\pi)$.

```
curve(sin(x), 0, 4*pi, col="blue", ylab="value")

curve(cos(x), 0, 4*pi, col="red", add=TRUE)
```



توزیع‌های گسسته در R

توزیع	نام انگلیسی	نام توزیع در R
دوجمله‌ای	binomial	binom(...,n,p)
دوجمله‌ای منفی	negative binomial	nbinom(..., r, p)
هندسی	geometric	geom(..., p)
پواسون	poisson	pois(..., p)

توزیع‌های پیوسته در R

توزیع	نام انگلیسی	نام توزیع در R
یکنواخت	uniform	unif(...,a,b)
نرمال	normal	norm(..., mu, sigma)
گاما	gamma	gamma(..., alpha, beta)
نمایی	exponential	exp(..., theta)
کی-دو	chi-squared	chisq(..., r, theta)
بتا	beta	beta(..., a, b)
وایبل	weibull	weibull(..., alpha, beta)

انجام محاسبات براساس توزیع‌های آماری در R

رابطه (فرمول)	محاسبه	تابع
$f_X(x)$	تابع احتمال یا چگالی احتمال (density)	ddist(x,par1,par2,...)
$P(X \leq x)$	تابع توزیع احتمال تجمعی Probability distribution function	pdist(x, par1, par2,...)
$F^{-1}(p)$	چندک های توزیع احتمال (quantiles)	qdist(p, par1, par2,...)
	تولید عدد تصادفی از توزیع دلخواه random variable	rdist(n, par1, par2,...)

dist: اسم توزیع در R

توزیع دوجمله‌ای (Binomial Distribution)

متغیر تصادفی X توزیع دوجمله‌ای دارد و می‌نویسیم $X \sim B(n, p)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x; n, p) = \binom{n}{x} p^x \, (1-p)^{n-x} \qquad x = 0, 1, \ldots, n, \qquad 0 < p \leq 1.$$

توزیع دو جمله‌ای در R

تابع احتمال:

dbinom(x, n, p)

تابع توزیع احتمال:

pbinom(q, n, p)

محاسبه چندک:

qbinom(p, n, p)

تولید داده تصادفی:

rbinom(n, n, p)

به عنوان مثال. اگر $X \sim B(۵, ۰/۵)$ باشد مطلوب است محاسبه

الف) $P(X = ۲)$ ب) $P(X < ۳)$ ج) $P(X \geq ۳)$.

برنامه R

الف)

```
dbinom(2, 5, 0.5)
[1] 0.3125
```

ب)

```
pbinom(2, 5, 0.5)
[1] 0.5
```

ج)

```
1-pbinom(2, 5, 0.5)
[1] 0.5
```

مثال. یک نمونه تصادفی ۱۰ تایی از توزیع $B(۳,\ ۰/۵)$ تولید کرده و مجموع، میانگین و انحراف معیار آن را حساب کنید.

```
> x=rbinom(10, 3, 0.5)
  x
[1] 2 2 3 1 2 2 0 1 2 1

> sum(x)
[1] 16

> mean(x)
[1] 1.6

> sd(x)
[1] 0.843274
```

توزیع دوجمله‌ای منفی (Negative Binomial Distribution)

متغیر تصادفی X توزیع دوجمله‌ای منفی دارد و می‌نویسیم $X \sim NB(r, \theta)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x; r, \theta) = \binom{x + r - ۱}{x} \theta^x \, (۱ - \theta)^r,$$

که در آن $r > ۰$,$۰ < \theta \leq ۱$,تعداد موفقیت ها $- \{۰, ۱, ۲, \ldots\} \in x$.

توزیع دو جمله‌ای منفی در R

تابع احتمال:

dnbinom(x, r, theta)

تابع توزیع احتمال:

pnbinom(q, r, theta)

محاسبه چندک:

qnbinom(p, r, theta)

تولید داده تصادفی:

rnbinom(n, r, theta)

توزیع هندسی (مدل شکست)

متغیر تصادفی X توزیع هندسی (مدل شکست) دارد اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;\theta) = \theta\,(۱-\theta)^x, \quad x = ۰, ۱, ۲, \ldots, \quad ۰ < \theta \leq ۱.$$

توزیع هندسی در R

تابع احتمال:

dgeom(x, theta)

تابع توزیع احتمال:

pgeom(q, theta)

محاسبه چندک:

qgeom(p, theta)

تولید داده تصادفی:

rgeom(n, theta)

توزیع پواسن (Poisson Distribution)

متغیر تصادفی X توزیع پواسن با پارامتر λ دارد و می‌نویسیم $X \sim Pois(\lambda)$ اگر و تنها اگر توزیع احتمالش به صورت زیر باشد.

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = ۰, ۱, ۲, \ldots, \quad \lambda > ۰.$$

توزیع پواسون در R

تابع احتمال:

dpois(x, lambda)

تابع توزیع احتمال:

ppois(q, lambda)

محاسبه چندک:

qpois(p, lambda)

تولید داده تصادفی:

rpois(n, lambda)

مثال. در توزیع پواسن با پارامتر $\lambda = ۳$ مطلوب است محاسبه $P(X < ۵)$.

برنامه R

داریم

$$P(X < ۵) = P(X \leq ۴)$$

احتمال فوق مقدار تابع توزیع تجمعی توزیع $Pois(۳)$ به ازای مقدار $x = ۴$ می‌باشد. بنابراین داریم:

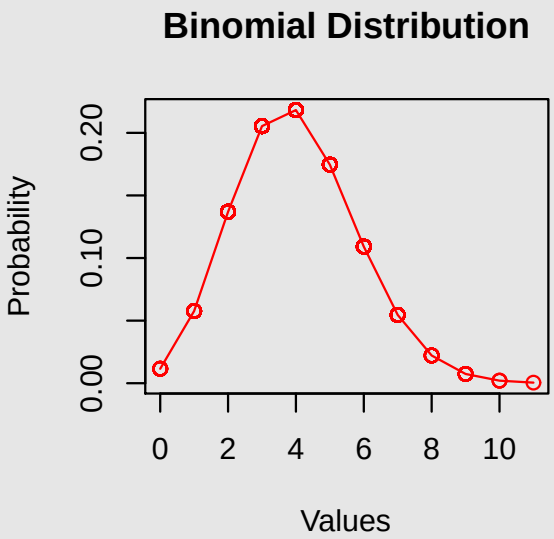
```
> ppois(4, 3)
[1] 0.8152632
```

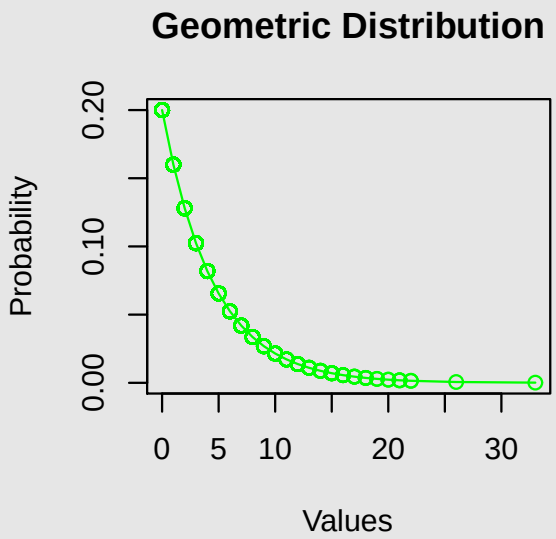
روش دوم:

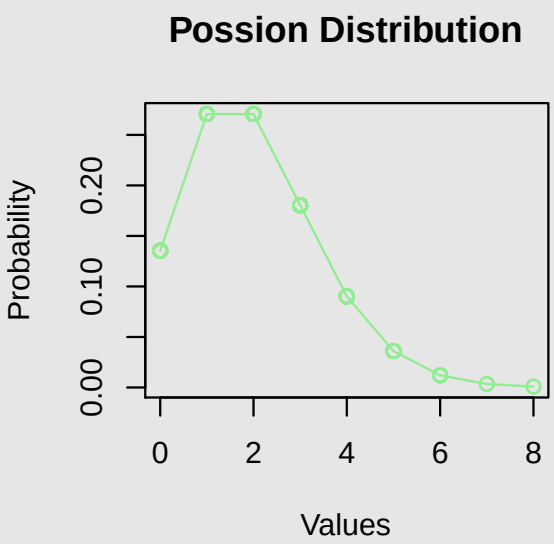
```
> sum( dpois(0:4, 3) )
[1] 0.8152632
```

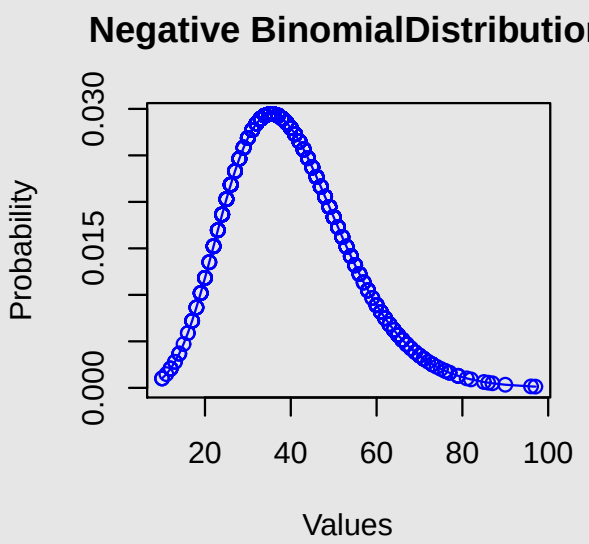
منحنی تابع جرم احتمال توزیع‌های گسسته

```
par(mfrow=c(2,2))
x=rbinom(n = 1000,size = 20,prob = 0.2)
z=sort(x)
y=dbinom(z,size = 20,prob = 0.2)
plot(z,y,type = 'o',col='red',ylab='Probability', main = 'Binomial Distribution')

x=rgeom(n = 1000,prob = 0.2)
z=sort(x)
y=dgeom(z,prob = 0.2)
plot(z,y,type = 'o',col='green',ylab='Probability', main = 'Geometric Distribution')

x=rpois(n = 1000,lambda = 2)
z=sort(x)
y=dpois(z,lambda =2)
plot(z,y,type = 'o',col='lightgreen',ylab='Probability', main = 'Possion Distribution')

x=rnbinom(n = 1000,size = 10,prob = 0.2)
z=sort(x)
y=dnbinom(z,size = 10,prob = 0.2)
plot(z,y,type = 'o',col='blue',ylab='Probability', main = 'Negative BinomialDistribution')
```


Binomial Distribution
Probability
0.00
0.10
0.20
0
2
4
6
8
10
Values


Geometric Distribution
Probability
0.00
0.10
0.20
0
5
10
20
30
Values


Possion Distribution
Probability
0.00
0.10
0.20
0
2
4
6
8
Values


Negative BinomialDistribution
Probability
0.000
0.015
0.030
20
40
60
80
100
Values

تعریف توزیع یکنواخت پیوسته (Uniform Distribution)

گوییم متغیر تصادفی X دارای توزیع یکنواخت در فاصله (a, b) است و می‌نویسیم $X \sim U(a, b)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x) = \frac{۱}{b-a}, \qquad a < x < b$$

توزیع یکنواخت در R

تابع چگالی احتمال:

dunif(x, a, b)

تابع توزیع:

punif(q, a, b)

محاسبه چندک:

qunif(p, a, b)

تولید داده تصادفی:

runif(n, a, b)

توزیع نرمال (Normal Distribution)

گوییم متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 است و می‌نویسیم $X \sim N(\mu, \sigma^2)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \qquad -\infty < x < \infty,$$

که در آن $\mu \in (-\infty, \infty)$ ، $\sigma > 0$.

توزیع نرمال در R

تابع چگالی احتمال:

dnorm(x, mu, sigma)

تابع توزیع:

pnorm(q, mu, sigma)

محاسبه چندک:

qnorm(p, mu, sigma)

تولید داده تصادفی:

rnorm(n, mu, sigma)

مثال ۱. متغیر تصادفی X توزیع نرمال با میانگین $\mu = ۱۵$ و واریانس $\sigma^۲ = ۱۲$ دارد. مطلوب است
الف) $P(X < ۱۰)$ ب) $P(X > ۱۶)$ ج) $P(۱۲ < X < ۲۰)$

برنامه R الف

```
> pnorm(10, mean=15, sd=sqrt(12) )
[1] 0.07445734
```

روش دوم: داریم

$$P(X < ۱۰) = P(\frac{X - ۱۵}{\sqrt{۱۲}} < \frac{۱۰ - ۱۵}{\sqrt{۱۲}})$$
$$= P(Z < -۱/۴۴) = \Phi(-۱/۴۴)$$

که در آن $\Phi(-۱/۴۴)$ تابع توزیع نرمال استاندارد در نقطه $-۱/۴۴$ می‌باشد. داریم:

```
> pnorm(-1.44 )
[1] 0.07445734
```

ب) داریم

$$P(X > ۱۶) = ۱ - P(X < ۱۶)$$

برنامه R ب

```
> 1-pnorm(16, mean=15, sd=sqrt(12) )
[1] 0.386415
```

روش دوم: داریم

$$\begin{aligned} P(X > ۱۶) &= ۱ - P(X < ۱۶) \\ &۱ - P(\frac{X - ۱۵}{\sqrt{۱۲}} < \frac{۱۶ - ۱۵}{\sqrt{۱۲}}) \\ &= ۱ - P(Z < ۰/۲۸) = ۱ - \Phi(۰/۲۸) \end{aligned}$$

داریم:

```
> 1-pnorm(0.28 )
[1] 0.386415
```

ج) $P(۱۲ < X < ۲۰) = P(X < ۲۰) - P(X < ۱۲)$

برنامه R ج

```
> pnorm(20, 15, sqrt(12) ) - pnorm(12, 15, sqrt(12) )
[1] 0.7323045
```

روش دوم: داریم

$$\begin{aligned} P(۱۲ < X < ۲۰) &= P(\frac{۱۲ - ۱۵}{\sqrt{۱۲}} < \frac{X - ۱۵}{\sqrt{۱۲}} < \frac{۲۰ - ۱۵}{\sqrt{۱۲}}) \\ &= P(-۰/۸۶ < Z < ۱/۴۴) \\ &= P(Z < ۱/۴۴) - P(Z < -۰/۸۶) \\ &= \Phi(۱/۴۴) - \Phi(-۰/۸۶) \end{aligned}$$

داریم:

```
> pnorm(1.44) - pnorm(-0.86)
[1] 0.7323046
```

مثال ۲. میزان رشد یک نوع نهال طی یک ماه، متغیر تصادفی نرمال با میانگین ۱۵ سانتی‌متر و انحراف معیار ۱/۲ سانتی‌متر است.
الف) احتمال اینکه نهالی از این نوع طی یک ماه بیش از ۱۷ سانتی متر رشد کند چقدر است؟
ب) ۹۰٪ نهال ها حداکثر چه رشدی دارند؟

الف: $P(X > ۱۷) = ۱ - P(X < ۱۷)$

برنامه R الف

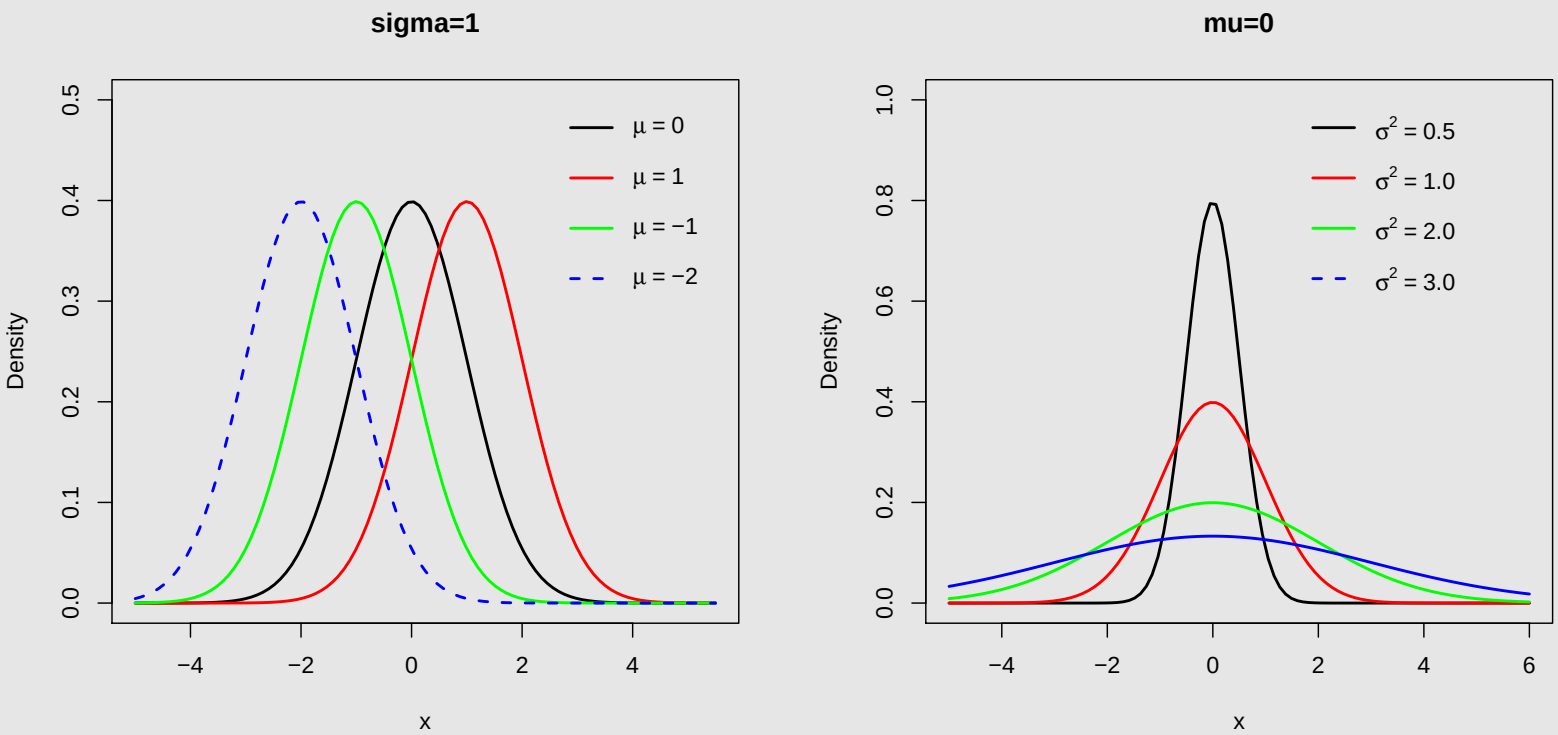
```
> 1 - pnorm(17, 15, 1.2 )
[1] 0.04779035
```

ب: $P(X < x) = ۰/۹ \Rightarrow x = ?$

برنامه R ب

```
> qnorm(0.9, 15, 1.2 )
[1] 16.53786
```

شکل: نمودار تابع چگالی احتمال توزیع نرمال به ازای مقادیر مختلف پارامترها



برنامه R منحنی تابع چگالی احتمال توزیع نرمال ($\sigma = ۱$)

```
> curve(dnorm(x,mean=0,sd=1),-5,5.5,col='black',type="l",lwd=2,lty=1
, ylim=c(0,1), ylab='Density',main="mu=0")

> curve(dnorm(x,mean=1,sd=1),-5,5.5,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(dnorm(x,mean=-1,sd=1),-5,5.5,col='green',type="l",lwd=2,lty=1,add=TRUE)

> curve(dnorm(x,mean=-2,sd=1),-5,5.5,col='blue',lty=2,type="l",lwd=2,add=TRUE)
```

برنامه R منحنی تابع چگالی احتمال توزیع نرمال ($\mu = ۰$)

```
> curve(dnorm(x,mean=0,sd=0.5),-5,6,col='black',type="l",lwd=2,lty=1,
,ylim=c(0,0.5),ylab='Density',main="sigma=1")

> curve(dnorm(x,mean=0,sd=1),-5,6,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(dnorm(x,mean=0,sd=2),-5,6,col='green',type="l",lwd=2,lty=1,add=TRUE)

> curve(dnorm(x,mean=0,sd=3),-5,6,col='blue',lty=1,type="l",lwd=2,add=TRUE)
```

تعریف توزیع گاما (*Gamma Distribution*)

گوییم متغیر تصادفی X دارای توزیع گاما با پارامترهای α و β است و می‌نویسیم $X \sim \Gamma(\alpha, \beta)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1}\, e^{-\frac{x}{\beta}}, \qquad x > 0,\ \alpha,\ \beta > 0$$

که در آن $\Gamma(\alpha)$ تابع گاما نام دارد که بصورت زیر تعریف می‌شود.

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\, e^{-x}\, \mathrm{d}x \qquad \alpha > 0$$

توزیع گاما در R

تابع چگالی احتمال:

```
dgamma(x, shape=alpha, scale=beta)
```

تابع توزیع:

```
pgamma(q, shape=alpha, scale=beta)
```

محاسبه چندک:

```
qgamma(p, shape=alpha, scale=beta)
```

تولید داده تصادفی:

```
rgamma(n, shape=alpha, scale=beta)
```

در صورتی که توزیع $\Gamma(r,\ \lambda)$ با تابع چگالی زیر مد نظر باشد از دستورات زیر استفاده می‌شود:

$$f_X(x) = \frac{۱}{\Gamma(r)}\, x^{r-۱}\, e^{-\lambda x}, \qquad x > ۰,\ r, \lambda > ۰$$

توزیع گاما در R

تابع احتمال:

```
dgamma(x, shape=r, scale=1/lambda)
```

تابع توزیع احتمال:

```
pgamma(q, shape=r, scale=1/lambda)
```

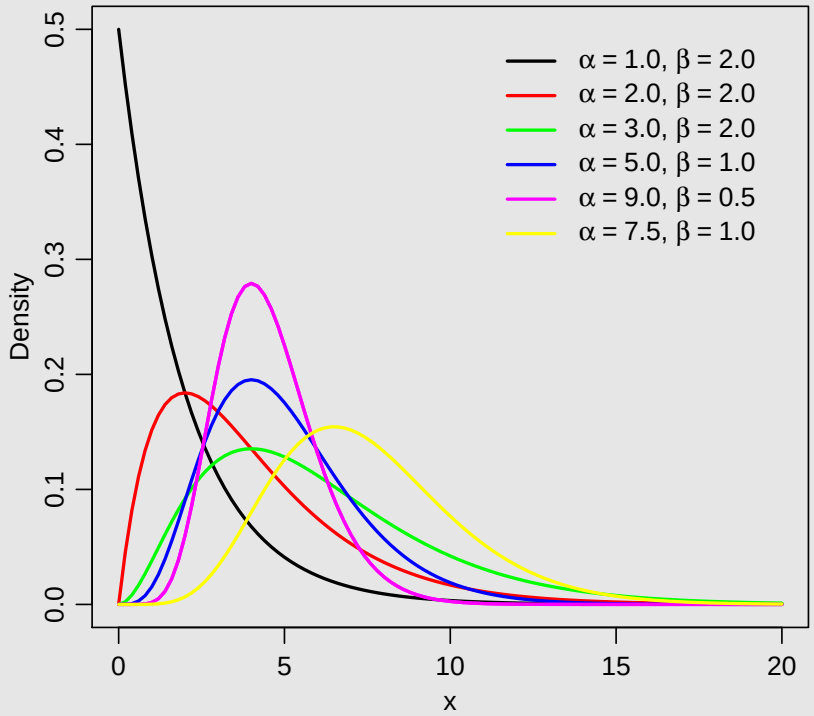
محاسبه چندک:

```
qgamma(p, shape=r, scale=1/lambda)
```

تولید داده تصادفی:

```
rgamma(n, shape=r, scale=1/lambda)
```

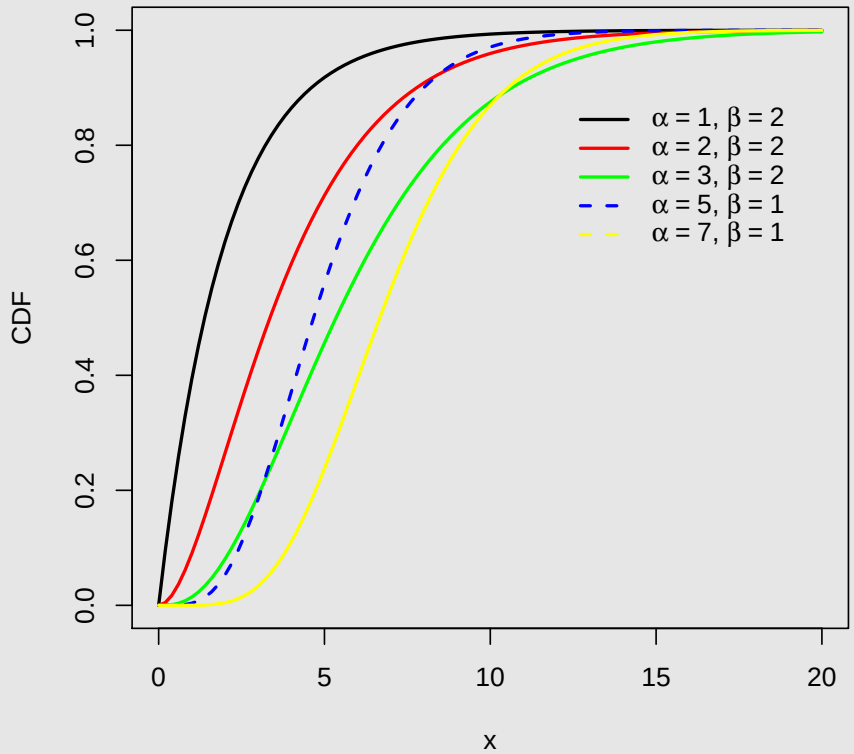
شکل: نمودار تابع چگالی احتمال توزیع گاما به ازای مقادیر مختلف پارامترها



برنامه R منحنی تابع چگالی احتمال توزیع گاما

```
> curve(dgamma(x,shape=1,scale=2),0,20,col='black', type="l", lwd=2, lty=1
,ylim=c(0,.5), ylab='Density')

> curve(dgamma(x, shape=2,scale=2),0,20,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(dgamma(x,shape=3,scale=2),0,20,col='green',type="l",lwd=2,lty=1,add=TRUE)

> curve(dgamma(x,shape=5,scale=1),0,20,col='blue',lty=2,type="l",lwd=2,add=TRUE)

> curve(dgamma(x,shape=7,scale=1),0,20,col='yellow',lty=1,type="l",lwd=2,add=TRUE)

> curve(dgamma(x,shape=9,scale=.5),0,20,col=5,lty=1,type="l",lwd=2,add=TRUE)
```

شکل: نمودار تابع توزیع احتمال توزیع گاما به ازای مقادیر مختلف پارامترها



برنامه R منحنی تابع توزیع احتمال توزیع گاما

```
> curve(pgamma(x,shape=1,scale=2),0,20,col='black',type="l",lwd=2,lty=1
,ylim=c(0,1), ylab='CDF')

> curve(pgamma(x, shape=2,scale=2),0,20,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(pgamma(x,shape=2,scale=2),0,20,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(pgamma(x,shape=3,scale=2),0,20,col='green',type="l",lwd=2,lty=1,add=TRUE)

> curve(pgamma(x,shape=5,scale=1),0,20,col='blue',lty=2,type="l",lwd=2,add=TRUE)

> curve(pgamma(x,shape=7,scale=1),0,20,col='yellow',lty=1,type="l",lwd=2,add=TRUE)
```

توزیع نمایی (Exponential Distribution)

توزیع نمایی با پارامتر θ را با نماد $Exp(\theta)$ نمایش می‌دهند. لذا

$$f(x) = \theta\, e^{-\theta\, x}, \qquad x > 0,\ \theta > 0$$

که در آن $\mathbb{E}(X) = \frac{1}{\theta}$.

توزیع نمایی در R

تابع چگالی احتمال:

dexp(x, scale=theta)

تابع توزیع:

pexp(q, scale=theta)

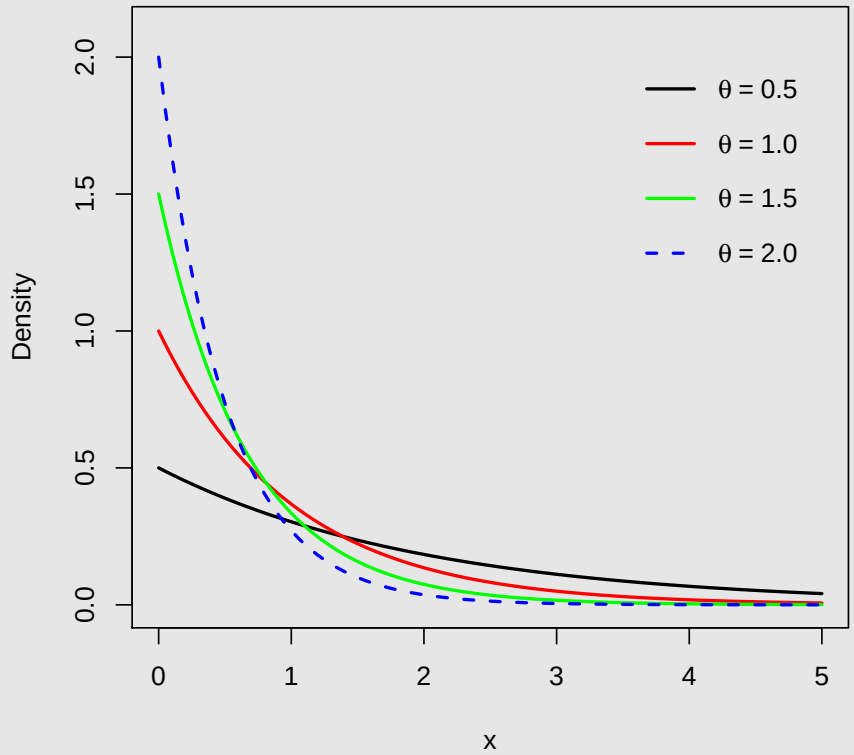
محاسبه چندک:

qexp(p, scale=theta)

تولید داده تصادفی:

rexp(n, scale=theta)

شکل: نمودار تابع چگالی احتمال توزیع نمایی به ازای مقادیر مختلف θ



با توجه به نمودار، رفتار منحنی تابع چگالی نمایی به ازای مقادیر مختلف θ همواره نزولی می‌باشد.

برنامه R منحنی تابع چگالی احتمال توزیع نمایی

```
> curve(dexp(x,0.5),0,5,col='black',type="l",lwd=2,lty=1,ylim=c(0,2.1),ylab='Density')
> curve(dexp(x, 1), 0, 5, col='red', type="l", lwd=2, lty=1, add=TRUE)
> curve(dexp(x, 1.5), 0, 5, col='green', type="l", lwd=2, lty=1, add=TRUE)
> curve(dexp(x, 2), 0, 5, col='blue', lty=2, type="l", lwd=2, add=TRUE)
```

مثال . طول عمر هر ذره از یک ایزوتوپ رادیواکتیو X (برحسب روز) متغیر تصادفی نمایی با میانگین ۴٫۲ روز است.
احتمال آنکه یک ذره از این ایزوتوپ بیش از ۷ روز عمر کند، چقدر است؟ نیم عمر این ایزوتوپ چقدر است؟

$$\frac{۱}{\theta} = ۴٫۲ \quad \Rightarrow \theta = \frac{۱}{۴٫۲} = ۰٫۲۳۸۰$$
$$P(X > ۷) = ۱ - P(X \leq ۷)$$

برنامه R

i. `> 1-pexp(7, 0.2380)`
`[1] 0.1888757`

ii. `> qexp(0.5, 0.2380)`
`[1] 2.911219`

توزیع وایبل (Weibull Distribution)

گوییم متغیر تصادفی X دارای توزیع وایبل با پارامترهای λ و k است و می‌نویسیم $X \sim We(k, \lambda)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x; k, \lambda) = \frac{k}{\lambda^k} x^{k-1} e^{-(\frac{x}{\lambda})^k}, \qquad x > 0, \lambda > 0, k > 0$$

توزیع وایبل در R

تابع چگالی احتمال:

dweibull(x, shape=k, scale=lambda)

تابع توزیع:

pweibull(q, shape=k, scale=lambda)

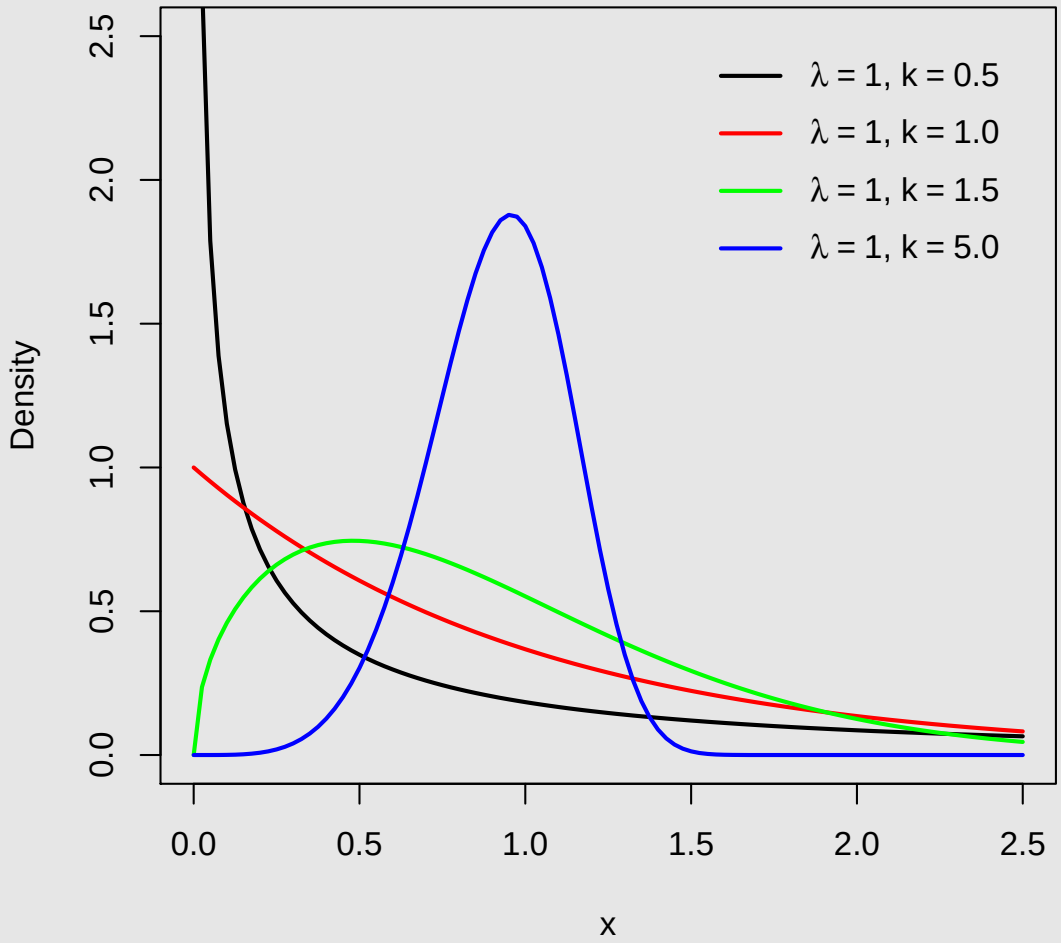
محاسبه چندک:

qweibull(p, shape=k, scale=lambda)

تولید داده تصادفی:

rweibull(n, shape=k, scale=lambda)

شکل: نمودار تابع چگالی احتمال توزیع وایبل به ازای مقادیر مختلف پارامترها



برنامه R منحنی تابع چگالی احتمال توزیع گاما

```
> curve(dweibull(x,shape=0.5,scale=1),0,2.5,col='black',type="l",lwd=2,lty=1,
, ylim=c(0,2.5), ylab='Density')

> curve(dweibull(x,shape=1,scale=1),0,2.5,col='red',type="l",lwd=2,lty=1,add=TRUE)

> curve(dweibull(x,shape=1.5,scale=1),0,2.5,col='green',type="l",lwd=2,lty=1,add=TRUE)

> curve(dweibull(x,shape=5,scale=1),0,2.5,col='blue',lty=1,type="l",lwd=2,add=TRUE)
```

توزیع پارتو (Pareto Distribution)

گوییم متغیر تصادفی X دارای توزیع پارتو با پارامترهای α و β است و می‌نویسیم $X \sim Pa(\alpha, \beta)$ هرگاه تابع چگالی X به صورت زیر باشد.

$$f(x; \alpha, \beta) = \frac{\alpha\, \beta^{\alpha}}{x^{\alpha+1}}, \qquad x \geq \beta,\ \alpha,\ \beta > 0$$

تابع توزیع تجمعی و تابع چندک X نیز به‌ترتیب به صورت زیر است:

$$F(x;\ \alpha,\ \beta) = 1 - (\frac{\beta}{x})^{\alpha}, \qquad x \geq \beta$$

و

$$F^{-1}(p) \ = \ \frac{\beta}{(1-p)^{1/\alpha}}, \qquad 0 < p < 1.$$

معرفی توزیع پارتو در R

از آنجایی که توزیع پارتو در R تعریف نشده است، بایستی توابع چگالی، توزیع، چندک و تابع تولید عدد تصادفی را معرفی کنیم.

معرفی توزیع پارتو

تابع چگالی احتمال:

```
> dpareto <- function(x, alpha, beta) alpha*beta^alpha/x^(alpha+1)
```

تابع توزیع:

```
> ppareto <- function(q, alpha, beta)   1-(beta/x)^alpha
```

محاسبه چندک:

```
> qpareto <- function(p, alpha, beta)   beta/(1-p)^(1/alpha)
```

تولید داده تصادفی:

```
> rpareto <- function(n, alpha, beta){
                p=runif(n)
                x=beta/(1-p)^(1/alpha)
                x
                                    }
```

مثال ۹. فرض کنید متغیر تصادفی X دارای توزیع پارتو با پارامترهای $\alpha = ۲$ و $\beta = ۱$ است.

الف) هریک از مقدارهای $P(X < ۳)$ و $P(۱ < X < ۲)$ را به دست آورید.

ب) برای هریک از مقدارهای $p = ۰/۸, ۰/۹۹$ مقدار x_p را طوری تعیین کنید که $P(X < x_p) = p$.

مثال ۹. فرض کنید متغیر تصادفی X دارای توزیع پارتو با پارامترهای $\alpha = ۲$ و $\beta = ۱$ است.

الف) هریک از مقدارهای $P(X < ۳)$ و $P(۱ < X < ۲)$ را به دست آورید.

ب) برای هریک از مقدارهای $p = ۰/۸, ۰/۹۹$ مقدار x_p را طوری تعیین کنید که $P(X < x_p) = p$.

برنامه R الف

i)
```
> ppareto(3, 2, 1)
[1] 0.88
```

```
> ppareto(2, 2, 1)-ppareto(1, 2, 1)
[1] 0.75
```

برنامه R ب

ii)
```
> qpareto(0.8, 2, 1)
[1] 2.2360
```

```
> qpareto(0.99, 2, 1)
[1] 10
```