



مقدمه ای بر تحلیل آماری با نرم افزار R

مدرس:

پیمان نیک چی

فارغ التحصیل کارشناسی ارشد آمار از دانشگاه تهران

Faradars.org

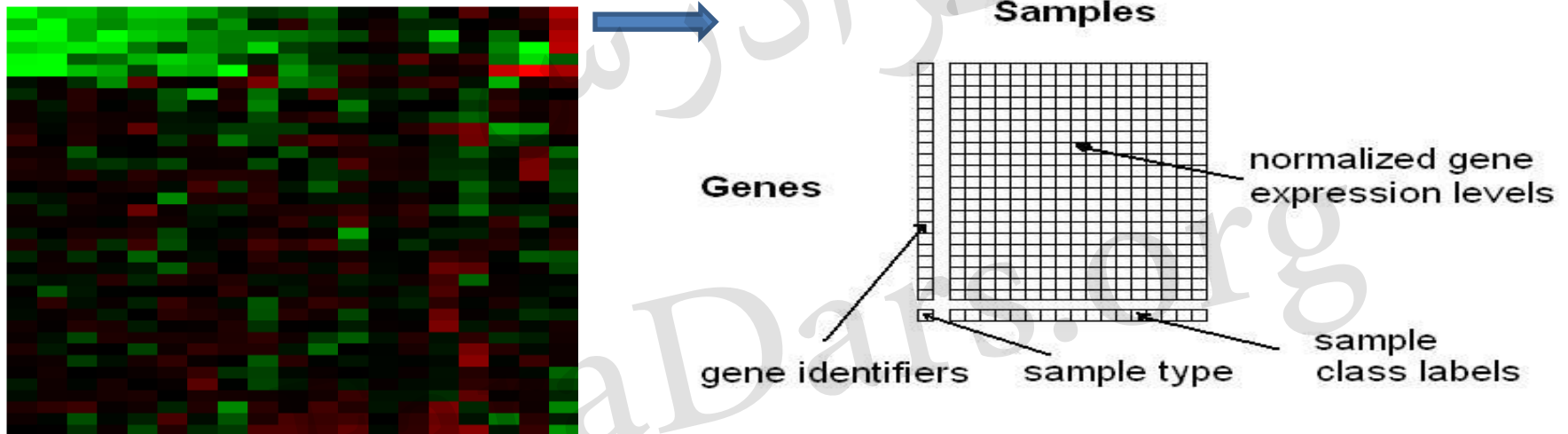


سرفصل مباحث جلسه هفتم

- نرمال سازی استاندارد
 - تبدیلات باکس کاکس
 - نرمال سازی بر اساس چندک
 - نرمال سازی به روش صیقل دادن با میانه
 - نرمال سازی به روش مقیاس گذاری
 - نرمال سازی به روش دامنه میان چارکی
- نرمال سازی ناپارامتری مبتنی بر رگرسیونی
- حل چند مثال در نرم افزار



مقدمه





نرمال سازی استاندارد

در این روش ابتدا میانگین را محاسبه و از تمامی داده ها کسر می کنیم. سپس تمامی داده ها را بر انحراف معیار تقسیم می کنیم.

2	4	4
5	4	14
4	6	8
3	5	8
3	3	9

3.33	1.15
7.66	5.50
6	2
5.33	2.51
5	3.46

-1.3	0.6	0.6
-2.6	-3.6	6.3
-2	0	2
-2.3	-0.3	2.6
-2	-2	4

-1.2	0.6	0.6
-0.5	-0.6	1.1
-1	0	1
-0.9	-0.1	1.1
-0.6	-0.6	1.2



تبدیلات باکس کاکس

- جورج باکس (George Box) و سر دیوید کاکس (Sir David Cox) خانواده ای از تبدیلات توانی را معرفی کردند که به کمک آن ها می توان داده های غیر نرمال را با تقریب بسیار خوبی نرمال کرد.
- برای این کار از تبدیل زیر استفاده می شود:

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(X) & \lambda = 0 \end{cases}$$



تبدیلات باکس کاکس

- تابع چگالی احتمال توزیع نرمال

$$X \sim N(\mu, \sigma^2)$$

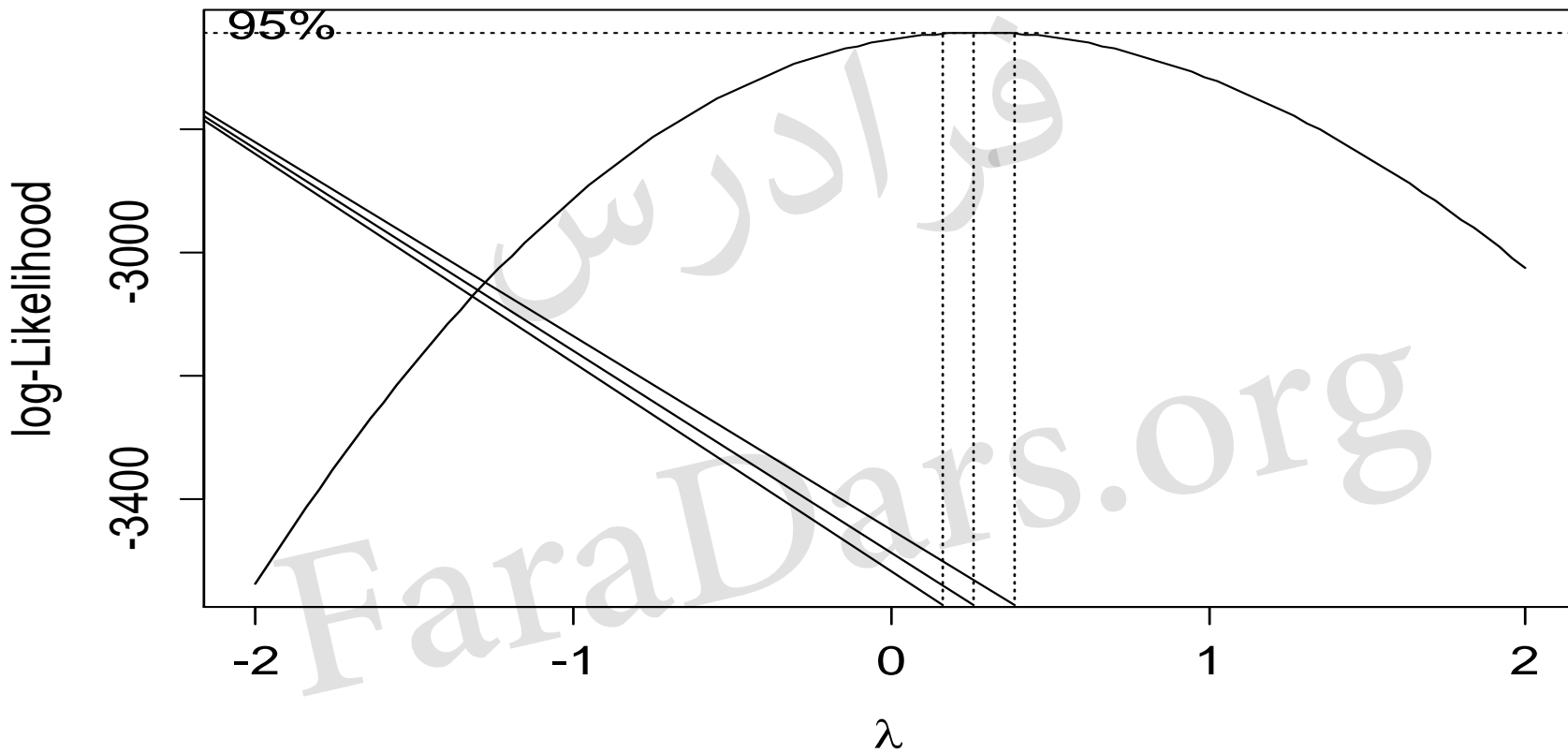
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x - \mu)^2\right)$$

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$L = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$



تبدیلات باکس کاکس





نرمال سازی بر اساس چندک

Original	Ranked	Averaged	Re-ordered
2 4 4	2 3 4	3 3 3	3 5 3
5 4 14	3 4 8	5 5 5	8 5 8
4 6 8	3 4 8	5 5 5	6 8 5
3 5 8	4 5 9	6 6 6	5 6 5
3 3 9	5 6 14	8 8 8	5 3 6



صیقل دادن با میانه

4	3	6	4	7	4
8	1	10	5	11	8
6	2	7	8	8	7
9	4	12	9	12	9
7	5	9	6	10	7



0	-1	2	0	3
0	-7	2	-3	3
-1	-5	0	1	1
0	-5	3	0	3
0	-2	2	-1	3

Faradars.org



صیقل دادن با میانه

0	-1	2	0	3
0	-7	2	-3	3
-1	-5	0	1	1
0	-5	3	0	3
0	-2	2	-1	3
0	-5	2	0	3



0	4	0	0	0
0	-2	0	-3	0
-1	0	-2	1	-2
0	0	1	0	0
0	3	0	-1	0



صیقل دادن با میانه

0	3	0	0	0	0
0	-3	0	-3	0	0
0	0	-1	2	-1	0
0	-1	1	0	0	0
0	2	0	-1	0	0
0	0	0	0	0	0



صیقل دادن با میانه

4	3	6	4	7
8	1	10	5	11
6	2	7	8	8
9	4	12	9	12
7	5	9	6	10

0	3	0	0	0
0	-3	0	-3	0
0	0	-1	2	-1
0	-1	1	0	0
0	2	0	-1	0

-

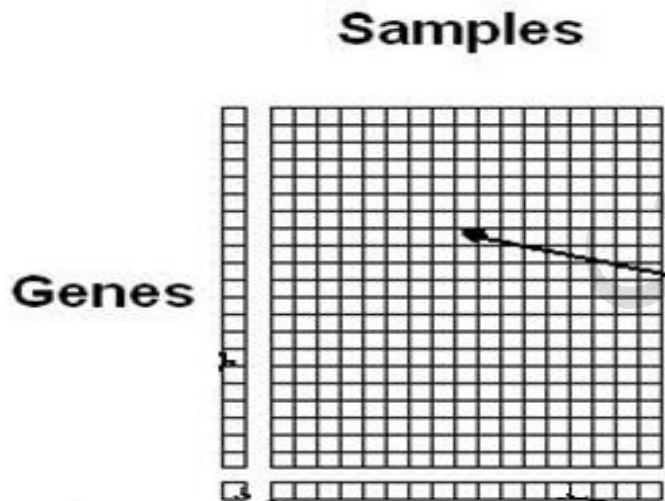
4	0	6	4	7
8	4	10	8	11
6	2	8	6	9
9	5	11	9	12
7	3	9	7	10

=

4.2
8.2
6.2
9.2
7.2



مقیاس گذاری

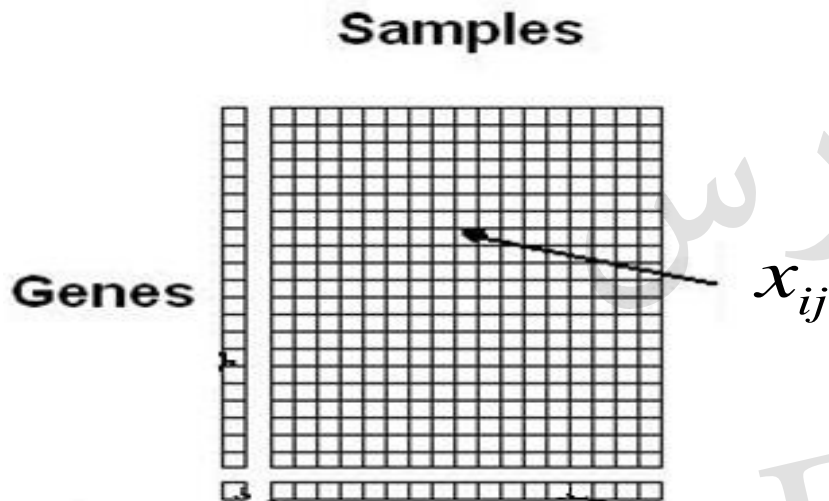


$$f(x_{ij}) = x_{ij} \times \frac{x_{ref}}{x_i}$$

FaraDars.org



دامنه میان چارکی



$$f(x_{ij}) = (x_{ij} + m_x - m_i) \times \frac{D_x}{D_i}$$



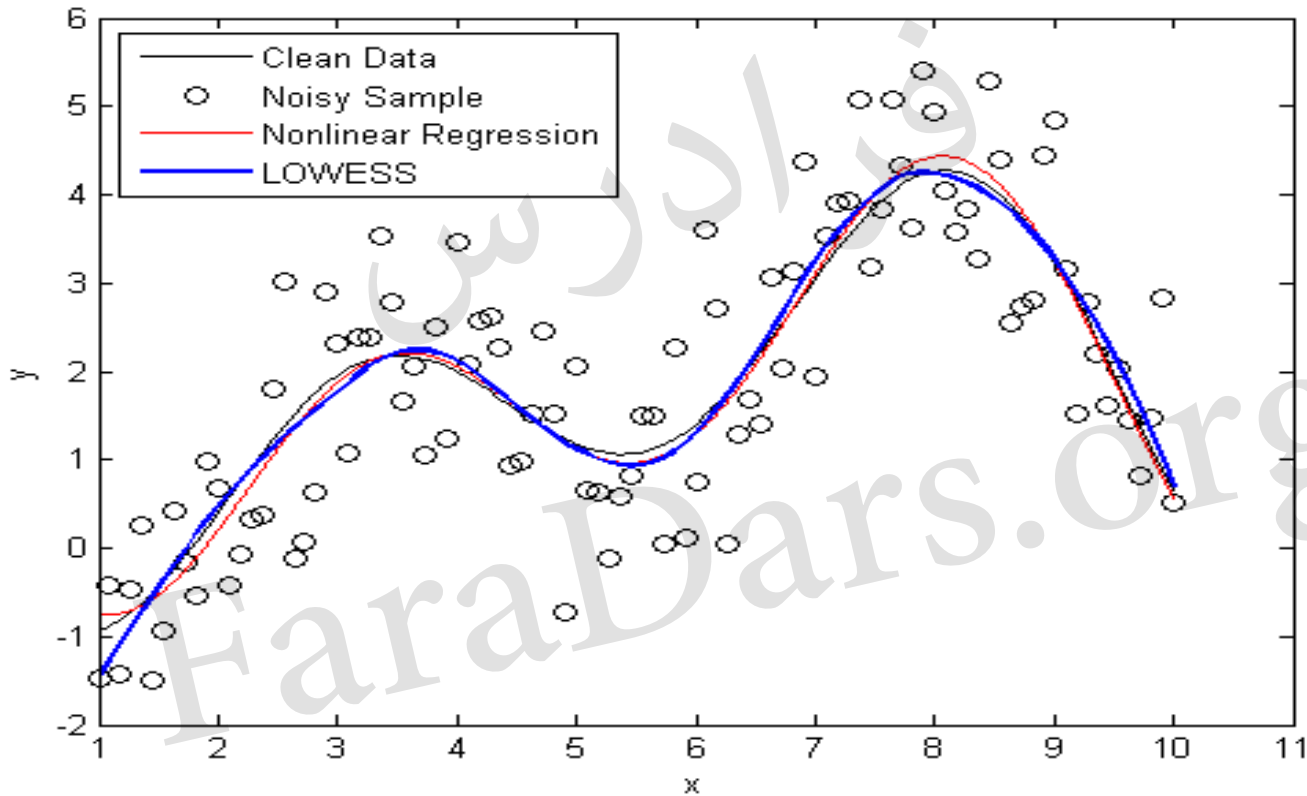
ناپارامتری مبتنی بر رگرسیون

- این روش ها مبتنی بر مدل های رگرسیون ناپارامتری هستند.
- این روش زمانی مورد استفاده قرار می گیرد که ارتباط غیر خطی بین نمونه ها وجود داشته باشد.

FaraDars.org



ناپارامتری مبتنی بر رگرسیون



این اسلاید ها بر مبنای نکات مطرح شده در فرادرس
«آموزش برنامه نویسی R و نرم افزار R Studio»
تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید
faradars.org/fvr9311