

تمرین‌های درس داده‌کاوی

مهدی محمدی - شماره دانشجویی: ۴۰۱۲۰۹۴

مصطفی سیلو - شماره دانشجویی: ۴۰۱۰۸۲۵۴

فصل ۲

تمرین ۱

دیتاستی از اطلاعات خودروهای دسته دوم

ردیف	تعداد سیلندر (X)	قیمت (هزار دلار) (Y)	نوع خودرو (C)
۱	۶	۱۵	Truck
۲	۶	۱۴	Mini-van
۳	۸	۳۲	Truck
۴	۸	۴۵	Truck
۵	۴	۴۰	Truck
۶	۶	۳۵	Mini-van
۷	۸	۲۹	Mini-van
۸	۸	۱۷	Truck
۹	۴	۱۱	Mini-van
۱۰	۸	۵۵	Truck

محاسبه میانگین و واریانس به تفکیک هر کلاس بازاری هر متغیر

تعداد سیلندر (X)	قیمت (هزار دلار) (Y)	تعداد نمونه (N)	
۷	۳۴	۶	Truck (T)
۲.۳۳	۲۰۸.۶۶		
۶	۲۲.۲۵	۴	Mini-van (MV)
۲	۱۰۰.۷۵		

محاسبه

$$S_X = \sqrt{\frac{VAR(X_T)}{N_T} + \frac{VAR(X_{MV})}{N_{MV}}} = \sqrt{\frac{2.33}{6} + \frac{2}{4}} = 0.9425144$$

$$S_Y = \sqrt{\frac{VAR(Y_T)}{N_T} + \frac{VAR(Y_{MV})}{N_{MV}}} = \sqrt{\frac{208.66}{6} + \frac{100.75}{4}} = 7.74365$$

$$\frac{|Mean(X_T) - Mean(X_{MV})|}{S_X} = \frac{|7 - 6|}{0.9425144} = 1.060991747$$

$$\frac{|Mean(Y_T) - Mean(Y_{MV})|}{S_Y} = \frac{|34 - 22.25|}{7.74365} = 1.51737229$$

اگر مقدار آستانه را ۱.۲ تعریف کنیم، بنابراین برای ویژگی تعداد سیلندر چون عدد محاسبه شده برای آن کمتر از مقدار آستانه است، بنابراین می‌توان برای کاهش بعد آن را حذف کرد

تمرین ۲

دیتاستی از اطلاعات خودروهای دسته دوم

ردیف	تعداد سیلندر (x)	قیمت (هزار دلار) (y)	نوع خودرو (c)
۱	۶	۱۵	Truck
۲	۶	۱۴	Mini-van
۳	۸	۳۲	Truck
۴	۴	۱۱	Mini-van

محاسبه فاصله بین دو نمونه:

$$D_{ij} = \sqrt{\sum_{k=1}^n \left(\frac{x_{ik} - x_{jk}}{\max_k - \min_k} \right)^2}$$

محاسبه تشابه بین دو نمونه:

$$S_{ij} = e^{-\alpha D_{ij}}$$

محاسبه آنتروپی:

$$Entropy = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N [S_{ij} \times \log_{10} S_{ij} + (1 - S_{ij}) \times \log_{10}(1 - S_{ij})]$$

فاصله	تشابه
$D_{12} = \sqrt{\left(\frac{6-6}{8-4}\right)^2 + \left(\frac{15-14}{32-11}\right)^2} = 0.0476$	$S_{12} = e^{-0.5 \times 0.0476} = 0.976$
$D_{13} = \sqrt{\left(\frac{6-8}{8-4}\right)^2 + \left(\frac{15-32}{32-11}\right)^2} = 0.951$	$S_{13} = e^{-0.5 \times 0.951} = 0.621$
$D_{14} = \sqrt{\left(\frac{6-4}{8-4}\right)^2 + \left(\frac{15-11}{32-11}\right)^2} = 0.535$	$S_{14} = e^{-0.5 \times 0.535} = 0.765$
$D_{23} = \sqrt{\left(\frac{6-8}{8-4}\right)^2 + \left(\frac{14-32}{32-11}\right)^2} = 0.992$	$S_{23} = e^{-0.5 \times 0.992} = 0.608$
$D_{24} = \sqrt{\left(\frac{6-4}{8-4}\right)^2 + \left(\frac{14-11}{32-11}\right)^2} = 0.52$	$S_{24} = e^{-0.5 \times 0.52} = 0.771$
$D_{34} = \sqrt{\left(\frac{8-4}{8-4}\right)^2 + \left(\frac{32-11}{32-11}\right)^2} = 1.414$	$S_{34} = e^{-0.5 \times 1.414} = 0.493$

آنتروپی کل: ۱.۴

حذف ویژگی (ستون) تعداد سیلندر

فاصله	تشابه
$D_{12} = \sqrt{\left(\frac{6-6}{8-4}\right)^2} = 0$	$S_{12} = e^{-0.5 \times 0} = 1$
$D_{13} = \sqrt{\left(\frac{6-8}{8-4}\right)^2} = 0.5$	$S_{13} = e^{-0.5 \times 0.5} = 0.778$
$D_{14} = \sqrt{\left(\frac{6-4}{8-4}\right)^2} = 0.5$	$S_{14} = e^{-0.5 \times 0.5} = 0.778$
$D_{23} = \sqrt{\left(\frac{6-8}{8-4}\right)^2} = 0.5$	$S_{23} = e^{-0.5 \times 0.5} = 0.778$
$D_{24} = \sqrt{\left(\frac{6-4}{8-4}\right)^2} = 0.5$	$S_{24} = e^{-0.5 \times 0.5} = 0.778$
$D_{34} = \sqrt{\left(\frac{8-4}{8-4}\right)^2} = 1$	$S_{34} = e^{-0.5 \times 1} = 0.606$

آنتروپی بعد از حذف ویژگی تعداد سیلندر: ۱.۲۱

حذف ویژگی (ستون) قیمت

فاصله	تشابه
$D_{12} = \sqrt{\left(\frac{15-14}{32-11}\right)^2} = 0.476$	$S_{12} = e^{-0.5 \times 0.476} = 0.976$
$D_{13} = \sqrt{\left(\frac{15-32}{32-11}\right)^2} = 0.809$	$S_{13} = e^{-0.5 \times 0.809} = 0.667$
$D_{14} = \sqrt{\left(\frac{15-11}{32-11}\right)^2} = 0.19$	$S_{14} = e^{-0.5 \times 0.19} = 0.909$
$D_{23} = \sqrt{\left(\frac{14-32}{32-11}\right)^2} = 0.857$	$S_{23} = e^{-0.5 \times 0.857} = 0.651$
$D_{24} = \sqrt{\left(\frac{14-11}{32-11}\right)^2} = 0.142$	$S_{24} = e^{-0.5 \times 0.142} = 0.931$
$D_{34} = \sqrt{\left(\frac{32-11}{32-11}\right)^2} = 1$	$S_{34} = e^{-0.5 \times 1} = 0.606$

آنتروپی بعد از حذف ویژگی قیمت: ۱.۱۳

ردیف	آنتروپی	فاصله تا آنتروپی کل	توضیحات
۱	۱.۱۳	۰.۲۷	آنتروپی بعد از حذف ویژگی قیمت
۲	۱.۲۱	۰.۱۹	آنتروپی بعد از حذف ویژگی تعداد سیلندر

تمرین ۳

ماتریس داده:

$$X = \begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix}$$

محاسبه ماتریس کواریانس:

$$\begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ -5 & 4 \\ -4 & 5 \\ 5 & -4 \end{bmatrix}$$

$$\Sigma = \frac{1}{4-1} \times \left(\begin{bmatrix} 4 & -5 & -4 & 5 \\ -5 & 4 & 5 & -4 \end{bmatrix} \times \begin{bmatrix} 4 & -5 \\ -5 & 4 \\ -4 & 5 \\ 5 & -4 \end{bmatrix} \right) = \frac{1}{3} \times \begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix} = \begin{bmatrix} 27.3333 & -26.6667 \\ -26.6667 & 27.3333 \end{bmatrix}$$

$$|\Sigma - \lambda I| = \left| \begin{bmatrix} 27.3333 & -26.6667 \\ -26.6667 & 27.3333 \end{bmatrix} - \lambda \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = \begin{vmatrix} 27.3333-\lambda & -26.6667 \\ -26.6667 & 27.3333-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (27.3333-\lambda)^2 - (26.6667)^2 = \lambda^2 - 54.6666\lambda + 35.9964 = 0$$

$$\Rightarrow \begin{cases} \lambda_1 = 54.0000 \\ \lambda_2 = 0.6666 \end{cases}$$

$$\Sigma v = \lambda v \Rightarrow \begin{bmatrix} 27.3333 & -26.6667 \\ -26.6667 & 27.3333 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 54.0000 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\begin{aligned} 27.3333 \times v_1 - 26.6667 \times v_2 &= 54.0000 \times v_1 \\ -26.6667 \times v_1 + 27.3333 \times v_2 &= 54.0000 \times v_2 \end{aligned}$$

$$-26.6667 \times v_2 = (54.0000 - 27.3333) \times v_1 = 26.6667 v_1$$

$$v_1 = -v_2$$

Normalization:

$$\boxed{\begin{aligned} v_1 &= \frac{-1}{\sqrt{(-1)^2 + 1^2}} = \frac{-1}{\sqrt{2}} \\ v_2 &= \frac{1}{\sqrt{(-1)^2 + 1^2}} = \frac{1}{\sqrt{2}} \end{aligned}}$$

$$\Sigma v = \lambda v \Rightarrow \begin{bmatrix} 27.3333 & -26.6667 \\ -26.6667 & 27.3333 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.6666 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\begin{aligned} 27.3333 \times v_1 - 26.6667 \times v_2 &= 0.6666 \times v_1 \\ -26.6667 \times v_1 + 27.3333 \times v_2 &= 0.6666 \times v_2 \end{aligned}$$

$$-26.6667 \times v_2 = (0.6666 - 27.3333) \times v_1 = -26.6667 v_1$$

$$v_1 = v_2$$

Normalization:

$$\boxed{\begin{aligned} v_1 &= \frac{-1}{\sqrt{(-1)^2 + (-1)^2}} = \frac{-1}{\sqrt{2}} \\ v_2 &= \frac{-1}{\sqrt{(-1)^2 + (-1)^2}} = \frac{-1}{\sqrt{2}} \end{aligned}}$$

تمرین ۴

شناسه	مقدار ویژگی	کلاس
۱	۱	A
۲	۳	B
۳	۷	A
۴	۸	A
۵	۹	A
۶	۱۱	B
۷	۲۳	B
۸	۳۷	A
۹	۳۹	B
۱۰	۴۵	A
۱۱	۴۶	A
۱۲	۵۹	A

در نظر گرفتن هر یک از مقادیر داده‌ها به عنوان بازه

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	(۰ و ۲]
۲	۳	B	(۲ و ۵]
۳	۷	A	(۵ و ۷.۵]
۴	۸	A	(۷.۵ و ۸.۵]
۵	۹	A	(۸.۵ و ۱۰]
۶	۱۱	B	(۱۰ و ۱۷]
۷	۲۳	B	(۱۷ و ۳۰]
۸	۳۷	A	(۳۰ و ۳۸]
۹	۳۹	B	(۳۸ و ۴۲]
۱۰	۴۵	A	(۴۲ و ۴۵.۵]
۱۱	۴۶	A	(۴۵.۵ و ۵۲]
۱۲	۵۹	A	(۵۲ و ۶۰]

مرحله ۱-۱:

	A	B	
(۰ و ۲]	$A_{11}=1$ $E_{11}=0.5$	$A_{12}=0$ $E_{12}=0.5$	$R_1=1$
(۲ و ۵]	$A_{21}=0$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

$$X^2 = \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۲-۱:

	A	B	
[۲و۵)	A ₁₁ =0 E ₁₁ =0.5	A ₁₂ =1 E ₁₂ =0.5	R ₁ =1
[۵و۷.۵)	A ₂₁ =1 E ₂₁ =0.5	A ₂₂ =0 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۳-۱:

	A	B	
[۵و۷.۵)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[۷.۵و۸.۵)	A ₂₁ =1 E ₂₁ =1	A ₂₂ =0 E ₂₂ =0.1	R ₂ =1
	C ₁ =2	C ₂ =0	N=2

$$X^2 = \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} = 0.2 < 2.706$$

مرحله ۴-۱:

	A	B	
[۷.۵و۸.۵)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[۸.۵و۱۰)	A ₂₁ =1 E ₂₁ =1	A ₂₂ =0 E ₂₂ =0.1	R ₂ =1
	C ₁ =2	C ₂ =0	N=2

$$X^2 = \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} = 0.2 < 2.706$$

مرحله ۵-۱:

	A	B	
[۸.۵و۱۰)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[۱۰و۱۷)	A ₂₁ =0 E ₂₁ =0.5	A ₂₂ =1 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۱-۶:

	A	B	
[۱۰۹۱۷)	A ₁₁ =0 E ₁₁ =0.1	A ₁₂ =1 E ₁₂ =1	R ₁ =1
[۱۷۹۳۰)	A ₂₁ =0 E ₂₁ =0.1	A ₂₂ =1 E ₂₂ =1	R ₂ =1
	C ₁ =0	C ₂ =2	N=2

$$X^2 = \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} = 0.2 < 2.706$$

مرحله ۱-۷:

	A	B	
[۱۷۹۳۰)	A ₁₁ =0 E ₁₁ =0.5	A ₁₂ =1 E ₁₂ =0.5	R ₁ =1
[۳۰۹۳۸)	A ₂₁ =1 E ₂₁ =0.5	A ₂₂ =0 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۱-۸:

	A	B	
[۳۰۹۳۸)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[۳۸۹۴۲)	A ₂₁ =0 E ₂₁ =0.5	A ₂₂ =1 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۱-۹:

	A	B	
[۳۸۹۴۲)	A ₁₁ =0 E ₁₁ =0.5	A ₁₂ =1 E ₁₂ =0.5	R ₁ =1
[۴۲۹۴۵.۵)	A ₂₁ =1 E ₂₁ =0.5	A ₂₂ =0 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۱-۱۰:

	A	B	
[۴۲ و ۴۵.۵)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[۴۵.۵ و ۵۲.۵)	A ₂₁ =1 E ₂₁ =1	A ₂₂ =0 E ₂₂ =0.1	R ₂ =1
	C ₁ =2	C ₂ =0	N=2

$$X^2 = \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} = 0.2 < 2.706$$

مرحله ۱-۱۱:

	A	B	
[۴۵.۵ و ۵۲.۵)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[۵۲.۵ و ۶۰)	A ₂₁ =1 E ₂₁ =1	A ₂₂ =0 E ₂₂ =0.1	R ₂ =1
	C ₁ =2	C ₂ =0	N=2

$$X^2 = \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} + \frac{(1 - 1)^2}{1} + \frac{(0 - 0.1)^2}{0.1} = 0.2 < 2.706$$

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	[۰ و ۲)
۲	۳	B	[۲ و ۵)
۳	۷	A	[۵ و ۱۰)
۴	۸	A	
۵	۹	A	
۶	۱۱	B	[۱۰ و ۳۰)
۷	۲۳	B	
۸	۳۷	A	[۳۰ و ۳۸)
۹	۳۹	B	[۳۸ و ۴۲)
۱۰	۴۵	A	[۴۲ و ۶۰)
۱۱	۴۶	A	
۱۲	۵۹	A	

مرحله ۲-۱:

	A	B	
[٠٩٢)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[٢٩٥)	A ₂₁ =0 E ₂₁ =0.5	A ₂₂ =1 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ٢-٢:

	A	B	
[٢٩٥)	A ₁₁ =0 E ₁₁ =0.75	A ₁₂ =1 E ₁₂ =0.25	R ₁ =1
[٥٩١٠)	A ₂₁ =3 E ₂₁ =2.25	A ₂₂ =0 E ₂₂ =0.75	R ₂ =3
	C ₁ =3	C ₂ =1	N=4

$$X^2 = \frac{(0 - 0.75)^2}{0.75} + \frac{(1 - 0.25)^2}{0.25} + \frac{(3 - 2.25)^2}{2.25} + \frac{(0 - 0.75)^2}{0.75} = 4 > 2.706$$

مرحله ٣-٢:

	A	B	
[٥٩١٠)	A ₁₁ =3 E ₁₁ =1.8	A ₁₂ =0 E ₁₂ =1.2	R ₁ =3
[١٠٩٣٠)	A ₂₁ =0 E ₂₁ =1.2	A ₂₂ =2 E ₂₂ =0.8	R ₂ =2
	C ₁ =3	C ₂ =2	N=5

$$X^2 = \frac{(3 - 1.8)^2}{1.8} + \frac{(0 - 1.2)^2}{1.2} + \frac{(0 - 1.2)^2}{1.2} + \frac{(2 - 0.8)^2}{0.8} = 5 > 2.706$$

مرحله ٤-٢:

	A	B	
[١٠٩٣٠)	A ₁₁ =0 E ₁₁ =0.66	A ₁₂ =2 E ₁₂ =1.33	R ₁ =2
[٣٠٩٣٨)	A ₂₁ =1 E ₂₁ =0.33	A ₂₂ =0 E ₂₂ =0.66	R ₂ =1
	C ₁ =1	C ₂ =2	N=3

$$X^2 = \frac{(0 - 0.66)^2}{0.66} + \frac{(2 - 1.33)^2}{1.33} + \frac{(1 - 0.33)^2}{0.33} + \frac{(0 - 0.66)^2}{0.66} = 3.0178 > 2.706$$

مرحله ۵-۲:

	A	B	
[۳۰۹۳۸)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[۳۸۹۴۲)	A ₂₁ =0 E ₂₁ =0.5	A ₂₂ =1 E ₂₂ =0.5	R ₂ =1
	C ₁ =1	C ₂ =1	N=2

$$X^2 = \frac{(1 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(1 - 0.5)^2}{0.5} = 2 < 2.706$$

مرحله ۶-۲:

	A	B	
[۳۸۹۴۲)	A ₁₁ =0 E ₁₁ =0.75	A ₁₂ =1 E ₁₂ =0.25	R ₁ =1
[۴۲۹۶۰)	A ₂₁ =3 E ₂₁ =2.25	A ₂₂ =0 E ₂₂ =0.75	R ₂ =3
	C ₁ =3	C ₂ =1	N=4

$$X^2 = \frac{(0 - 0.75)^2}{0.75} + \frac{(1 - 0.25)^2}{0.25} + \frac{(3 - 2.25)^2}{2.25} + \frac{(0 - 0.75)^2}{0.75} = 4 > 2.706$$

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	[۰۵)
۲	۳	B	
۳	۷	A	[۵۱۰)
۴	۸	A	
۵	۹	A	
۶	۱۱	B	[۱۰۳۰)
۷	۲۳	B	
۸	۳۷	A	[۳۰۹۴۲)
۹	۳۹	B	
۱۰	۴۵	A	[۴۲۹۶۰)
۱۱	۴۶	A	
۱۲	۵۹	A	

مرحله ۳-۱:

	A	B	
$[\circ 9 \Delta)$	$A_{11}=1$ $E_{11}=1.6$	$A_{12}=1$ $E_{12}=0.4$	$R_1=2$
$[\Delta 9 \circ)$	$A_{21}=3$ $E_{21}=2.4$	$A_{22}=0$ $E_{22}=0.6$	$R_2=3$
	$C_1=4$	$C_2=1$	$N=5$

$$X^2 = \frac{(1 - 1.6)^2}{1.6} + \frac{(1 - 0.4)^2}{0.4} + \frac{(3 - 2.4)^2}{2.4} + \frac{(0 - 0.6)^2}{0.6} = 1.8750 < 2.706$$

مرحله ۳-۲:

	A	B	
$[\Delta 9 \circ)$	$A_{11}=3$ $E_{11}=1.8$	$A_{12}=0$ $E_{12}=1.2$	$R_1=3$
$[\circ 9 \circ)$	$A_{21}=0$ $E_{21}=1.2$	$A_{22}=2$ $E_{22}=0.8$	$R_2=2$
	$C_1=3$	$C_2=2$	$N=5$

$$X^2 = \frac{(3 - 1.8)^2}{1.8} + \frac{(0 - 1.2)^2}{1.2} + \frac{(0 - 1.2)^2}{1.2} + \frac{(2 - 0.8)^2}{0.8} = 5 > 2.706$$

مرحله ۳-۳:

	A	B	
$[\circ 9 \circ)$	$A_{11}=0$ $E_{11}=0.5$	$A_{12}=2$ $E_{12}=1.5$	$R_1=2$
$[\circ 9 \Delta)$	$A_{21}=1$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=1.5$	$R_2=2$
	$C_1=1$	$C_2=3$	$N=4$

$$X^2 = \frac{(0 - 0.5)^2}{0.5} + \frac{(2 - 1.5)^2}{1.5} + \frac{(1 - 0.5)^2}{0.5} + \frac{(1 - 1.5)^2}{1.5} = 1.3333 < 2.706$$

مرحله ۳-۴:

	A	B	
$[\circ 9 \Delta)$	$A_{11}=1$ $E_{11}=1.6$	$A_{12}=1$ $E_{12}=0.4$	$R_1=2$
$[\Delta 9 \circ)$	$A_{21}=3$ $E_{21}=2.4$	$A_{22}=0$ $E_{22}=0.6$	$R_2=3$
	$C_1=4$	$C_2=1$	$N=5$

$$X^2 = \frac{(1 - 1.6)^2}{1.6} + \frac{(1 - 0.4)^2}{0.4} + \frac{(3 - 2.4)^2}{2.4} + \frac{(0 - 0.6)^2}{0.6} = 1.8750 < 2.706$$

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	[۰۹۵)
۲	۳	B	
۳	۷	A	[۵۹۱۰)
۴	۸	A	
۵	۹	A	
۶	۱۱	B	[۱۰۹۴۲)
۷	۲۳	B	
۸	۳۷	A	
۹	۳۹	B	
۱۰	۴۵	A	[۴۲۹۶۰)
۱۱	۴۶	A	
۱۲	۵۹	A	

مرحله ۴-۱:

	A	B	
[۰۹۵)	A ₁₁ =1 E ₁₁ =1.6	A ₁₂ =1 E ₁₂ =0.4	R ₁ =2
[۵۹۱۰)	A ₂₁ =3 E ₂₁ =2.4	A ₂₂ =0 E ₂₂ =0.6	R ₂ =3
	C ₁ =4	C ₂ =1	N=5

$$X^2 = \frac{(1 - 1.6)^2}{1.6} + \frac{(1 - 0.4)^2}{0.4} + \frac{(3 - 2.4)^2}{2.4} + \frac{(0 - 0.6)^2}{0.6} = 1.8750 < 2.706$$

مرحله ۴-۲:

	A	B	
[۵۹۱۰)	A ₁₁ =3 E ₁₁ =1.71	A ₁₂ =0 E ₁₂ =1.29	R ₁ =3
[۱۰۹۴۲)	A ₂₁ =1 E ₂₁ =2.29	A ₂₂ =3 E ₂₂ =1.71	R ₂ =4
	C ₁ =4	C ₂ =3	N=7

$$X^2 = \frac{(3 - 1.71)^2}{1.71} + \frac{(0 - 1.29)^2}{1.29} + \frac{(1 - 2.29)^2}{2.29} + \frac{(3 - 1.71)^2}{1.71} = 3.9630 > 2.706$$

مرحله ۳-۴:

	A	B	
[۱۰۹۴۲)	A ₁₁ =1 E ₁₁ =2.29	A ₁₂ =3 E ₁₂ =1.71	R ₁ =4
[۴۲۹۶۰)	A ₂₁ =3 E ₂₁ =1.71	A ₂₂ =0 E ₂₂ =1.29	R ₂ =3
	C ₁ =4	C ₂ =3	N=7

$$X^2 = \frac{(1 - 2.29)^2}{2.29} + \frac{(3 - 1.71)^2}{1.71} + \frac{(3 - 1.71)^2}{1.71} + \frac{(0 - 1.29)^2}{1.29} = 3.9630 > 2.706$$

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

بازه	کلاس	مقدار ویژگی	شناسه
[۰۱۰)	A	۱	۱
	B	۳	۲
	A	۷	۳
	A	۸	۴
	A	۹	۵
[۱۰۹۴۲)	B	۱۱	۶
	B	۲۳	۷
	A	۳۷	۸
	B	۳۹	۹
[۴۲۹۶۰)	A	۴۵	۱۰
	A	۴۶	۱۱
	A	۵۹	۱۲

مرحله ۱-۵:

	A	B	
[۰۱۰)	A ₁₁ =4 E ₁₁ =2.78	A ₁₂ =1 E ₁₂ =2.22	R ₁ =5
[۱۰۹۴۲)	A ₂₁ =1 E ₂₁ =2.22	A ₂₂ =3 E ₂₂ =1.78	R ₂ =4
	C ₁ =5	C ₂ =4	N=9

$$X^2 = \frac{(4 - 2.78)^2}{2.78} + \frac{(1 - 2.22)^2}{2.22} + \frac{(1 - 2.22)^2}{2.22} + \frac{(3 - 1.78)^2}{1.78} = 2.7125 > 2.706$$

مرحله ۵-۲:

	A	B	
(۱۰۴۲)	$A_{11}=1$ $E_{11}=2.29$	$A_{12}=3$ $E_{12}=1.71$	$R_1=4$
(۴۲۶۰)	$A_{21}=3$ $E_{21}=1.71$	$A_{22}=0$ $E_{22}=1.29$	$R_2=3$
	$C_1=4$	$C_2=3$	$N=7$

$$X^2 = \frac{(1 - 2.29)^2}{2.29} + \frac{(3 - 1.71)^2}{1.71} + \frac{(3 - 1.71)^2}{1.71} + \frac{(0 - 1.29)^2}{1.29} = 3.9630 > 2.706$$

چون در مرحله پنجم هیچ یک از X^2 کوچکتر از ۲.۷۰۶ نشد، بنابراین هیچ ادغامی صورت نمی‌گیرد.

نتیجه نهایی:

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	(۱۰۴۲)
۲	۳	B	
۳	۷	A	
۴	۸	A	
۵	۹	A	
۶	۱۱	B	(۱۰۴۲)
۷	۲۳	B	
۸	۳۷	A	
۹	۳۹	B	
۱۰	۴۵	A	(۴۲۶۰)
۱۱	۴۶	A	
۱۲	۵۹	A	

تمرین ۵

میانگین: ۲۹

میان: ۲۵

مد: ۲۵ و ۳۵

واریانس: ۱۸۰.۵۷۱

انحراف معیار: ۱۳.۴۳۷

چارک اول: ۲۰

چارک سوم: ۳۵

انحراف چارکی: ۷.۵

آستانه برای پیدا کردن داده‌های پرت: [۲.۱۲۶ و ۵۵.۸۷۳]

اعدادی که در این بازه قرار ندارند را می‌توان به عنوان داده‌های پرت در نظر گرفت.

دسته‌بندی به روش Sturge's Rule:

$$K = 1 + (3.3 \times \log_{10} n) = 1 + (3.3 \times \log_{10} 28) = 5.77562 \approx 6$$

$$R = \max - \min = 70 - 3 = 67$$

$$D = \frac{R}{K} = \frac{67}{6} = 11.166 \approx 11$$

{3, 13, 15, 16, 16}, {19, 20, 20, 21, 22}, {22, 25, 25, 25, 25}, {30, 33, 33, 35}, {35, 35, 35, 36, 40},
{45, 46, 52, 70}

مد را برای هر دسته

تمرین ۶

نرمال‌سازی به روش اول:

کمینه (min): ۷.۸ کمینه جدید (new_min): -۱

بیشینه (max): ۴۲.۵ بیشینه جدید (new_max): +۱

نرمال‌سازی به روش Z-score:

میانگین: ۲۷.۱۱۶۶ انحراف معیار: ۱۰.۹۲۶۳

سن	درصد چربی	درصد چربی (نرمال شده به روش اول)	درصد چربی (نرمال شده به روش Z-score)
۲۳	۹.۵	-0.902017291	-2.144104145
۲۳	۲۶.۵	0.077809798	-0.253882686
۲۷	۷.۸	-1	-2.333126291
۲۷	۱۷.۸	-0.423631124	-1.221231315
۳۹	۳۱.۴	0.360230548	0.290945852
۴۱	۲۵.۹	0.043227666	-0.320596384
۴۷	۲۷.۴	0.129682997	-0.153812138
۴۹	۲۷.۲	0.11815562	-0.176050037
۵۰	۳۱.۲	0.34870317	0.268707953

سن	درصد چربی	درصد چربی (نرمال شده به روش اول)	درصد چربی (نرمال شده به روش Z-score)
۵۲	۳۴.۶	0.544668588	0.646752245
۵۴	۴۲.۵	1	1.525149276
۵۴	۲۸.۸	0.21037464	0.001853159
۵۶	۳۳.۴	0.475504323	0.513324848
۵۷	۳۰.۲	0.291066282	0.157518455
۵۸	۳۴.۱	0.515850144	0.591157496
۵۸	۳۲.۹	0.446685879	0.457730099
۶۰	۴۱.۲	0.925072046	1.380602929
۶۱	۳۵.۷	0.608069164	0.769060692

واریانس سن (S_x): ۱۶۵.۰۲۴۶۹۱۴

واریانس درصد چربی (S_y): ۸۱.۳۰۴۴۴۴۴

کوواریانس (S_{xy}): ۹۴.۶۸۸۸۸۸۸۹

ضریب همبستگی (r_{xy}): ۰.۸۱۷۴۶۱۶۹۷

با توجه به اینکه ضریب همبستگی یک عدد مثبت است، بنابراین زمانی که متغیر سن افزایش یابد، متغیر درصد چربی نیز افزایش می‌یابد و بالعکس.