

تمرین‌های درس داده‌کاوی

مهدی محمدی – شماره دانشجویی: ۴۰۱۱۲۰۹۴

مصطفی سبیلو – 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 بازی هر متغیر

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

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	آنتروپی تا آنتروپی کل	توضیحات
۱	۰.۲۷	۰.۱۳	آنتروپی بعد از حذف ویژگی قیمت
۲	۰.۱۹	۰.۲۱	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:

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

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

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

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

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

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

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

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

$$v_1 = -v_2$$

Normalization:

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

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

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

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

$$v_1 = v_2$$

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	[۰۹۲)
۲	۳	B	[۲۹۵)
۳	۷	A	[۵۹۷.۵)
۴	۸	A	[۷.۵۹۸.۵)
۵	۹	A	[۸.۵۹۱۰)
۶	۱۱	B	[۱۰۰۱۷)
۷	۲۳	B	[۱۷۹۳۰)
۸	۳۷	A	[۳۰۹۳۸)
۹	۳۹	B	[۳۸۹۴۲)
۱۰	۴۵	A	[۴۲۹۴۵.۵)
۱۱	۴۶	A	[۴۵.۵۹۵۲)
۱۲	۵۹	A	[۵۲۹۶۰)

: مرحله ۱-۱

	A	B	
[۰۹۲)	$A_{11}=1$ $E_{11}=0.5$	$A_{12}=0$ $E_{12}=0.5$	$R_1=1$
[۲۹۵)	$A_{21}=0$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

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

: مرحله ۲-۱

	A	B	
[V ₉ Δ)	A ₁₁ =0 E ₁₁ =0.5	A ₁₂ =1 E ₁₂ =0.5	R ₁ =1
[Δ ₉ V.Δ)	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	
[Δ ₉ V.Δ)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[V.Δ ₉ Λ.Δ)	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	
[V.Δ ₉ Λ.Δ)	A ₁₁ =1 E ₁₁ =1	A ₁₂ =0 E ₁₂ =0.1	R ₁ =1
[Λ.Δ ₉ I _o)	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	
[Λ.Δ ₉ I _o)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[I _o Δ ₉ V)	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 ٦-١ :

	A	B	
[١٠٩]٧)	$A_{11}=0$ $E_{11}=0.1$	$A_{12}=1$ $E_{12}=1$	$R_1=1$
[٧٩٣]	$A_{21}=0$ $E_{21}=0.1$	$A_{22}=1$ $E_{22}=1$	$R_2=1$
	$C_1=0$	$C_2=2$	$N=2$

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

مرحله ٧-١ :

	A	B	
[١٧٩٣]	$A_{11}=0$ $E_{11}=0.5$	$A_{12}=1$ $E_{12}=0.5$	$R_1=1$
[٣٠٩٣٨]	$A_{21}=1$ $E_{21}=0.5$	$A_{22}=0$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

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

مرحله ٨-١ :

	A	B	
[٣٠٩٣٨]	$A_{11}=1$ $E_{11}=0.5$	$A_{12}=0$ $E_{12}=0.5$	$R_1=1$
[٣٨٩٤٢]	$A_{21}=0$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

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

مرحله ٩-١ :

	A	B	
[٣٨٩٤٢]	$A_{11}=0$ $E_{11}=0.5$	$A_{12}=1$ $E_{12}=0.5$	$R_1=1$
[٤٢٩٤٥.٥]	$A_{21}=1$ $E_{21}=0.5$	$A_{22}=0$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

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

: مرحله ۱-۱

	A	B	
[۱۲۹۱۵.۵)	$A_{11}=1$ $E_{11}=1$	$A_{12}=0$ $E_{12}=0.1$	$R_1=1$
[۱۵.۵۹۵۲.۵)	$A_{21}=1$ $E_{21}=1$	$A_{22}=0$ $E_{22}=0.1$	$R_2=1$
	$C_1=2$	$C_2=0$	$N=2$

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

: مرحله ۱-۲

	A	B	
[۱۵.۵۹۵۲.۵)	$A_{11}=1$ $E_{11}=1$	$A_{12}=0$ $E_{12}=0.1$	$R_1=1$
[۵۲.۵۹۶۰)	$A_{21}=1$ $E_{21}=1$	$A_{22}=0$ $E_{22}=0.1$	$R_2=1$
	$C_1=2$	$C_2=0$	$N=2$

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

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

شناشه	مقدار ویژگی	کلاس	بازه
۱	۱	A	[۰۰)
۲	۳	B	[۲۹۵)
۳	۷	A	[۵۹۱۰)
۴	۸	A	[۵۹۱۰)
۵	۹	A	[۵۹۱۰)
۶	۱۱	B	[۱۰۹۳۰)
۷	۲۳	B	[۱۰۹۳۰)
۸	۳۷	A	[۳۰۹۳۸)
۹	۳۹	B	[۳۸۹۴۲)
۱۰	۴۵	A	[۴۲۹۶۰)
۱۱	۴۶	A	[۴۲۹۶۰)
۱۲	۵۹	A	[۴۲۹۶۰)

: مرحله ۱-۳

	A	B	
[o9Y)	A ₁₁ =1 E ₁₁ =0.5	A ₁₂ =0 E ₁₂ =0.5	R ₁ =1
[Y9ω)	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

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

:٣-٢ مرحله

	A	B	
[٢٩٥)	$A_{11}=0$ $E_{11}=0.75$	$A_{12}=1$ $E_{12}=0.25$	$R_1=1$
[٥٩١٠)	$A_{21}=3$ $E_{21}=2.25$	$A_{22}=0$ $E_{22}=0.75$	$R_2=3$
	$C_1=3$	$C_2=1$	$N=4$

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

:٣-٣ مرحله

	A	B	
[٥٩١٠)	$A_{11}=3$ $E_{11}=1.8$	$A_{12}=0$ $E_{12}=1.2$	$R_1=3$
[١٠٩١٠)	$A_{21}=0$ $E_{21}=1.2$	$A_{22}=2$ $E_{22}=0.8$	$R_2=2$
	$C_1=3$	$C_2=2$	$N=5$

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

:٤-٢ مرحله

	A	B	
[١٠٩١٠)	$A_{11}=0$ $E_{11}=0.66$	$A_{12}=2$ $E_{12}=1.33$	$R_1=2$
[١٠٩١٠)	$A_{21}=1$ $E_{21}=0.33$	$A_{22}=0$ $E_{22}=0.66$	$R_2=1$
	$C_1=1$	$C_2=2$	$N=3$

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

:٥-٢ مرحله

	A	B	
[١٠٩١٠)	$A_{11}=1$ $E_{11}=0.5$	$A_{12}=0$ $E_{12}=0.5$	$R_1=1$
[١٠٩١٠)	$A_{21}=0$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=0.5$	$R_2=1$
	$C_1=1$	$C_2=1$	$N=2$

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

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
[۰۹۵)	A	۱	۱
	B	۳	۲
[۵۹۱۰)	A	۷	۳
	A	۸	۴
	A	۹	۵
[۱۰۹۳۰)	B	۱۱	۶
	B	۲۳	۷
[۳۰۹۴۲)	A	۳۷	۸
	B	۳۹	۹
[۴۲۹۶۰)	A	۴۵	۱۰
	A	۴۶	۱۱
	A	۵۹	۱۲

مرحله ۱-۳:

	A	B	
[۰۹۵)	$A_{11}=1$ $E_{11}=1.6$	$A_{12}=1$ $E_{12}=0.4$	$R_1=2$
[۵۹۱۰)	$A_{21}=3$ $E_{21}=2.4$	$A_{22}=0$ $E_{22}=0.6$	$R_2=3$
	$C_1=4$	$C_2=1$	$N=5$

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

:۲-۳ مرحله

	A	B	
[۱۰۹۳]	$A_{11}=3$ $E_{11}=1.8$	$A_{12}=0$ $E_{12}=1.2$	$R_1=3$
[۱۰۹۴]	$A_{21}=0$ $E_{21}=1.2$	$A_{22}=2$ $E_{22}=0.8$	$R_2=2$
	$C_1=3$	$C_2=2$	$N=5$

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

:۳-۴ مرحله

	A	B	
[۱۰۹۴]	$A_{11}=0$ $E_{11}=0.5$	$A_{12}=2$ $E_{12}=1.5$	$R_1=2$
[۳۰۹۴۲]	$A_{21}=1$ $E_{21}=0.5$	$A_{22}=1$ $E_{22}=1.5$	$R_2=2$
	$C_1=1$	$C_2=3$	$N=4$

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

:۴-۵ مرحله

	A	B	
[۳۰۹۴۲]	$A_{11}=1$ $E_{11}=1.6$	$A_{12}=1$ $E_{12}=0.4$	$R_1=2$
[۴۲۹۶۰]	$A_{21}=3$ $E_{21}=2.4$	$A_{22}=0$ $E_{22}=0.6$	$R_2=3$
	$C_1=4$	$C_2=1$	$N=5$

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

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

بازه	کلاس	مقدار ویژگی	شناسه
[۰۹۵)	A	۱	۱
	B	۳	۲
[۵۹۱۰)	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 ۱-۴

	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 ۳-۴

	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 ۱-۴ :

	A	B	
[۱۰۹۴۲)	$A_{11}=1$ $E_{11}=2.29$	$A_{12}=3$ $E_{12}=1.71$	$R_1=4$
[۱۴۲۹۶۰)	$A_{21}=3$ $E_{21}=1.71$	$A_{22}=0$ $E_{22}=1.29$	$R_2=3$
	$C_1=4$	$C_2=3$	$N=7$

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

پس از این مراحل، بازه‌ها به شکل زیر تغییر می‌کند

شناسه	مقدار ویژگی	کلاس	بازه
۱	۱	A	[۰۹۱۰)
۲	۳	B	
۳	۷	A	
۴	۸	A	
۵	۹	A	
۶	۱۱	B	[۱۰۹۴۲)
۷	۲۳	B	
۸	۳۷	A	
۹	۳۹	B	
۱۰	۴۵	A	[۱۴۲۹۶۰)
۱۱	۴۶	A	
۱۲	۵۹	A	

مرحله ۱-۵ :

	A	B	
[۰۹۱۰)	$A_{11}=4$ $E_{11}=2.78$	$A_{12}=1$ $E_{12}=2.22$	$R_1=5$
[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	
[۱۰۹۴۲)	$A_{11}=1$ $E_{11}=2.29$	$A_{12}=3$ $E_{12}=1.71$	$R_1=4$
[۱۴۲۹۶۰)	$A_{21}=3$ $E_{21}=1.71$	$A_{22}=0$ $E_{22}=1.29$	$R_2=3$
	$C_1=4$	$C_2=3$	$N=7$

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

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

نتیجه نهایی:

باشه	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
[۱۴۲۹۶۰)	A	۴۵	۱۰
	A	۴۶	۱۱
	A	۵۹	۱۲

تمرین ۵

میانگین: ۲۹

میانه: ۲۵

مد: ۳۵ و ۲۵

واریانس: ۱۸۰.۵۷۱

انحراف معیار: ۱۳.۴۳۷

چارک اول: ۲۰

چارک سوم: ۳۵

انحراف چارکی: ۷.۵

آستانه برای پیدا کردن داده‌های پرت: [۵۵.۸۷۳ و ۱۲۶]

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:

$$\text{کمینه جدید} : (\text{min}) \quad 7.8$$

$$\text{بیشینه جدید} : (\text{max}) \quad ۴۲.۵$$

نرمال‌سازی به روش Z-score:

$$\text{میانگین: } ۲۷.۱۱۶ \quad \text{انحراف معیار: } ۱۰.۹۲۶$$

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

سن	درصد چربی (Z-score)	درصد چربی (نرمال شده به روش اول)	درصد چربی
۵۲	۰.۶۴۶۷۵۲۲۴۵	۰.۵۴۴۶۶۸۵۸۸	۳۴.۶
۵۴	۱.۵۲۵۱۴۹۲۷۶	۱	۴۲.۵
۵۴	۰.۰۰۱۸۵۳۱۵۹	۰.۲۱۰۳۷۴۶۴	۲۸.۸
۵۶	۰.۵۱۳۳۲۴۸۴۸	۰.۴۷۵۵۰۴۳۲۳	۳۳.۴
۵۷	۰.۱۵۷۵۱۸۴۵۵	۰.۲۹۱۰۶۶۲۸۲	۳۰.۲
۵۸	۰.۵۹۱۱۵۷۴۹۶	۰.۵۱۵۸۵۰۱۴۴	۳۴.۱
۵۸	۰.۴۵۷۷۳۰۰۹۹	۰.۴۴۶۶۸۵۸۷۹	۳۲.۹
۶۰	۱.۳۸۰۶۰۲۹۲۹	۰.۹۲۵۰۷۲۰۴۶	۴۱.۲
۶۱	۰.۷۶۹۰۶۰۶۹۲	۰.۶۰۸۰۶۹۱۶۴	۳۵.۷

واریانس سن (S_x): ۱۶۵.۰۲۴۶۹۱۴

واریانس درصد چربی (S_y): ۸۱.۳۰۴۴۴۴۴

کوواریانس (S_{xy}): ۹۱۴.۶۸۸۸۸۸۸۹

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.