



سوال ۱) با فرض نرمال بودن داده های $\{x_1, x_2, \dots, x_n\}$ زیر و با کمک درست نمایی بیشینه پارامترهای (μ, σ^2) را مشخص نمایید.

پاسخ)

$$f_X(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$L(X; \mu, \sigma^2) = \prod_{i=1}^n f_X(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_1^n (x_i - \mu)^2}{2\sigma^2}\right)$$

$$\log L(X; \mu, \sigma^2) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \frac{\sum_1^n (x_i - \mu)^2}{2\sigma}$$

$$\begin{cases} \frac{\partial}{\partial \sigma} \log L = -\frac{n}{\sigma} + \frac{\sum_1^n (x_i - \mu)^2}{\sigma^3} \\ \frac{\partial}{\partial \mu} \log L = \frac{\sum_1^n (x_i - \mu)}{\sigma^2} \end{cases}$$

$$\hat{\mu}, \hat{\sigma}^2 \Rightarrow \begin{cases} -\frac{n}{\hat{\sigma}} + \frac{\sum_1^n (x_i - \hat{\mu})^2}{\hat{\sigma}^3} = 0 \\ \frac{\sum_1^n (x_i - \hat{\mu})}{\hat{\sigma}^2} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\sigma}^2 = \frac{\sum_1^n (x_i - \hat{\mu})^2}{n} \\ \hat{\mu} = \frac{\sum_1^n x_i}{n} \end{cases}$$

سوال ۲) جدول احتمالی سه طرفه را برای ویژگی های X, Y, Z نشان داده شده در جدول زیر در نظر بگیرید.

	Z=f		Z=g	
	Y=d	Y=e	Y=d	Y=e
X=a	۵	۲	۶	۹
X=b	۵	۲	۴	۳
X=c	۱۵	۲	۱۷	۷

مقدار χ^2 را برای همبستگی بین X و Z محاسبه کنید. در سطح معناداری ۱۰ درصد بررسی کنید که آیا آنها وابسته هستند یا مستقل؟ برای یافتن مقادیر بحرانی از جدول ۱۰-۳ کتاب [مرجع](#) استفاده نمایید.

پاسخ)

O(X=x,Z=z)	Z=f	Z=g	S(X=x)
X=a	5+2	6+9	22
X=b	5+2	4+3	14
X=c	15+2	17+7	41
S(Z=z)	31	46	N=77

$$e(X = x, Z = z) = \frac{S(X = x) * S(Z = z)}{N}$$

E(X=x,Z=z)	Z=f	Z=g	S(X=x)
X=a	$\frac{22 * 31}{77}$	$\frac{22 * 46}{77}$	22
X=b	$\frac{14 * 31}{77}$	$\frac{14 * 46}{77}$	14
X=c	$\frac{41 * 31}{77}$	$\frac{41 * 46}{77}$	41
S(Z=z)	31	46	N=77

$$\chi^2 = \sum_{i=1}^{m*n} \frac{(o_i - e_i)^2}{e_i}$$

$$\chi^2 = \frac{(7 - 8.86)^2}{8.86} + \frac{(7 - 5.64)^2}{5.64} + \frac{(17 - 16.51)^2}{16.51} + \frac{(13 - 13.14)^2}{13.14} + \frac{(7 - 8.36)^2}{8.36} + \frac{(24 - 24.49)^2}{24.49} = 1.229$$

$$df = (m - 1)(n - 1) = (3 - 1)(2 - 1) = 2$$

$$p - value = (\text{جدول}) \Rightarrow \text{مستقلند}$$

سوال ۳) ماتریس کوواریانس را با جزییات برای مثال زیر محاسبه نمایید:

	x_1	x_2	x_3
X_1^T	۲	۱	۲
X_3^T	۰	۳	۲
X_3^T	۱	۴	۳

نکته: سطر ها نشان دهنده نمونه ها و ستون ها مشخصه ها هستند.

ج: ستون age را به سه دسته تقسیم کنید.

پاسخ)

$$D \Rightarrow \mu = \left[\frac{3}{3}, \frac{8}{3}, \frac{7}{3} \right] \Rightarrow D - \mu = \begin{bmatrix} 1 & -\frac{5}{3} & -\frac{1}{3} \\ -1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{4}{3} & \frac{2}{3} \end{bmatrix}$$

$$\Sigma = \frac{1}{m} (D - \mu)^T (D - \mu) = \frac{1}{3} \begin{bmatrix} 1 & -1 & 0 \\ \frac{5}{3} & \frac{1}{3} & \frac{4}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{5}{3} & -\frac{1}{3} \\ -1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{4}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & 0 \\ -\frac{2}{3} & \frac{14}{9} & \frac{4}{9} \\ 0 & \frac{4}{9} & \frac{2}{9} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & 0 \\ \frac{2}{3} & \frac{14}{9} & \frac{4}{9} \\ 0 & \frac{4}{9} & \frac{2}{9} \end{bmatrix}$$

سوال ۴: نمونه داده های زیر را در نظر بگیرید:

age	sex	bmi	children	smoker	region	charges
19	female	16	0	no	south	16884
18	male	31	0	yes	south	26874
32	female	27.9	3	yes	east	21984
25	female	22.3	0	no	north	1725
34	male	31.4	1	yes	east	22584
42	male	37	2	no	north	35284

الف: ستون های عددی را با روش z_score یا min-max نرمال کنید.

ب: ستون های غیر عددی را به روش مناسب عددی کنید.

ج: ستون age را به سه دسته تقسیم کنید.

پاسخ)

الف):

	age	bmi	children	charges
mean	28.33333	27.2	1	20889.17
St.dev	9.352362	0.744446	1.264911	11228.6
min	18	16	0	1725
max	42	37	3	35284

$$n_x = \frac{x - \min}{\max - \min}$$

$$z = \frac{x - \text{mean}}{\text{st.dev}}$$

روش Min-max:

age	bmi	children	charges
0.0416666 67	0	0	0.45171191
0	0.7142857 14	0	0.749396585
0.5833333 33	0.5666666 67	1	0.603683066
0.2916666 67	0.3	0	0
0.6666666 67	0.7333333 33	0.333333	0.621562025
1	1	0.666667	1

روش Z-score:

age	bmi	children	charges
-0.99797	-1/50447	-0/79057	-0/35669
-1.10489	0/510447	-0/79057	0/532999
0.392058	0/09403	1/581139	0/097504
-0.35642	-0/65821	-0/79057	-1/70673
0.605908	0/564178	0	0/150939
1.461306	1316415	0/790569	1/281979

(ب)

sex	smoker	Region-S	Region-N	Region-E
0	0	1	0	0
1	1	1	0	0
0	1	0	0	1
0	0	0	1	0
1	1	0	0	1
1	0	0	1	0

ج) دو روش وجود دارد:

دسته های هم تعداد:

$$k = \frac{N}{3} = \frac{6}{3} = 2$$

$$bins: \{A: \{18,19\}, B: \{25,32\}, C: \{34,42\}\}$$

دسته های هم طول:

$$l = \frac{\max - \min}{3} = \frac{42 - 18}{3} = 8$$

$$bins: \{A: [18 - 26), B: [26 - 34), C: [34 - 42]\}$$

$$bins: \{A: \{18,19,25\}, B: \{32\}, C: \{34,42\}\}$$