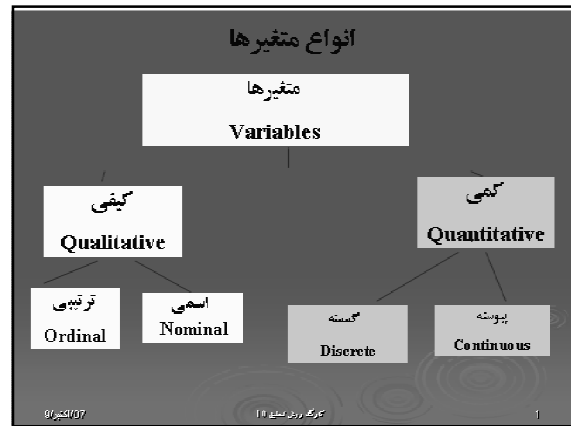


روش های آمار توصفی

مروری بر

دکتر پدا... محرابی
 دانشکده بهداشت - دانشگاه علوم پزشکی شهید بهشتی
ymehrab@gmail.com



انواع متغیرها:

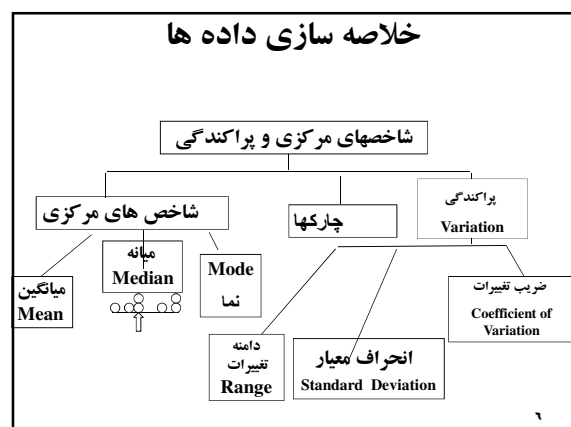
مستقل
وابسته
زمینه ای
مداخله گر

پیش بینی بروز دیابت در جمعیت بالای ۲۰ سال

متغیر	مستقل	وابسته	کیفی		مقیاس
			نسبی	رتبه ای	
مشخصات					
دیاره					
سختی					
سین اوست آیا؟					
بسته های دارویی مصرف					
کتابچه سوزن					
مصرف انسولین					
شماره سوزن					
mmHg					
kg/m					
IFG or IGT					
تایید بیمارستانی و مروری					
تعداد سوزن					
mg/dl					
mg/dl					
mg/dl					
HDL					
تصویلات					
cm					

پیش بینی بروز دیابت در جمعیت بالای ۲۰ سال

متغیر	مستقل	وابسته	کیفی		مقیاس
			نسبی	رتبه ای	
مشخصات					
دیاره					
سختی					
سین اوست آیا؟					
بسته های دارویی مصرف					
کتابچه سوزن					
مصرف انسولین					
شماره سوزن					
mmHg					
kg/m					
IFG or IGT					
تایید بیمارستانی و مروری					
تعداد سوزن					
mg/dl					
mg/dl					
mg/dl					
HDL					
تصویلات					
cm					



Mean میانگین

میانگین نمونه ای

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

۷

Mean میانگین

میانگین نمونه ای

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

۸

Mean میانگین

میانگین نمونه ای

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

تحت تاثیر داده های پرت قرار می گیرد

۹

Median میانه

۱۰

Median میانه

تعداد مشاهدات فرد : داده ها : 2 8 5 4 1

داده های مرتب شده : 1 2 4 5 8

↑
میانه

۱۱

Median میانه

تعداد مشاهدات فرد : داده ها : 2 8 5 4 1

داده های مرتب شده : 1 2 4 5 8

↑
میانه

تعداد مشاهدات زوج : داده ها : 2 8 1 4 3 8

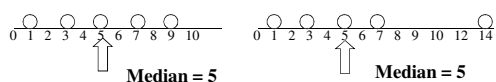
داده های مرتب شده : 1 2 3 4 8 8

↙ ↘
میانه (Median) = (3+4)/2 = 3.5

۱۲

Median میانه

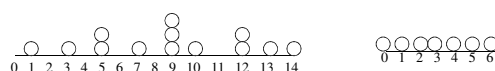
تحت تاثیر داده های پرت قرار نمی گیرد



۱۳

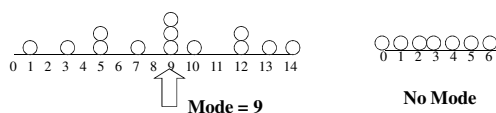
Mode نما

داده یا داده هایی که بیشترین فراوانی را داشته باشند



۱۴

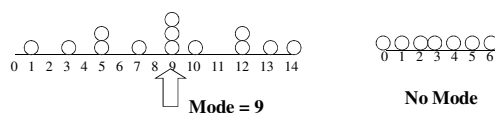
Mode نما



۱۵

Mode نما

تحت تاثیر داده های پرت قرار نمی گیرد



۱۶

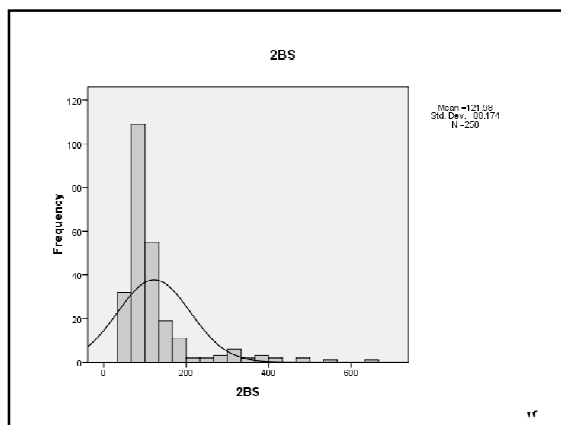
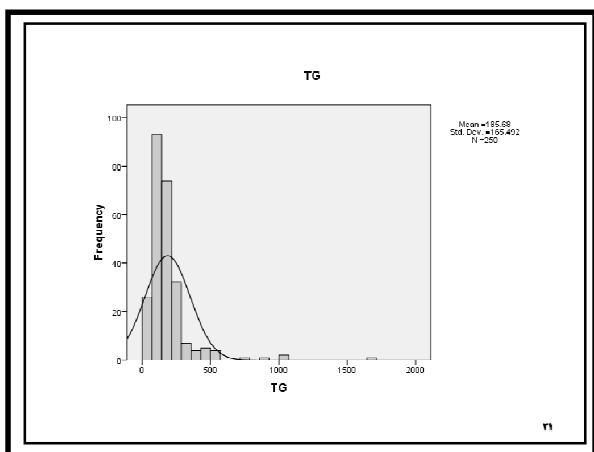
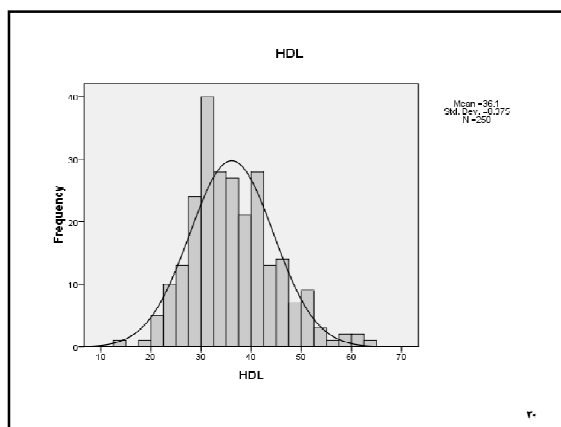
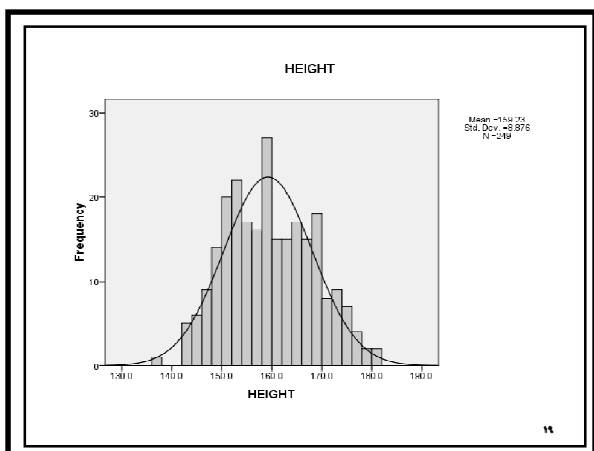
برای نمایش مرکز
مشاهدات کدام شاخص
مناسب تر است ؟

۱۷

مثال

		WEIGHT	HEIGHT	HDL	TG	2BS
N	Valid	250	249	250	250	250
Mean		67.5	159.2	36.1	185.6	121.9
Median		67.2	158.5	35.0	148.5	97.0
Std. Deviation		12.40	8.8	8.3	165.5	88.2
Range		63.5	44.0	49	1645	619
Minimum		38.5	137.0	14	43	34
Maximum		102.0	181.0	63	1688	653

۱۸



شکل داده ها

Symmetric or skewed تقارن یا چولگی توزیع

Left-Skewed
Mean Median Mode

Symmetric
Mean = Median = Mode

Right-Skewed
Mode Median Mean

برای نمایش مرکز مشاهدات کدام شاخص مناسب تر است؟

اگر توزیع داده ها تقریباً متقارن باشد
میانتکین ، میانه و نما تقریباً مشابه هستند

اگر توزیع داده ها نامتقارن باشد یا داده های پرت داشته باشیم
میانه مناسبتر است

برای داده های چند نمایی بهتر است
شاخصها برای هر زیر گروه محاسبه شود

چندک ها

Quantiles

Centiles صدک ها

Tentiles دهک ها

Quintiles پنجهک ها

Quartiles چارک ها

.....

چارک ها

داده ها را از نظر فراوانی به چهار قسمت مساوی تقسیم می کند

25%	25%	25%	25%
Q_1	Q_2	Q_3	

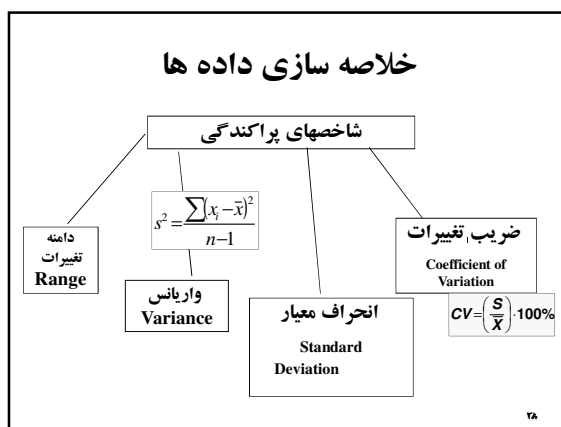
$$Q_i = \frac{i(n+1)}{4}$$

Data in Ordered Array: 11 12 13 16 16 17 18 21 22

Position of $Q_1 = \frac{1(9+1)}{4} = 2.50$ $Q_1 = 12.5$

چندک ها

Quantiles					
	WEIGHT	HEIGHT	SYSTOLIC BP	TG	CHOLESTEROL
Median	67.250	158	120.00	148.50	192.50
Percentiles					
5	47.000	145	95.00	63.00	126.10
10	51.550	148	100.00	71.00	146.10
20	56.000	151	105.00	94.00	157.00
25	58.000	152	110.00	102.00	163.00
30	60.000	153	110.00	110.30	168.00
40	64.000	156	110.00	129.40	179.40
50	67.250	158.5	120	148.5	192.5
60	70.000	161	120.00	169.00	200.60
70	74.000	164	130.00	198.00	219.70
75	76.125	166	136.25	209.00	226.00
80	78.000	168	140.00	223.00	231.00
90	83.900	171	150.00	302.10	251.90
95	90.000	174	170.00	434.50	271.00



مثال : فشار خون در دو گروه از افراد مورد بررسی به صورت زیر بوده است

گروه I : 90 130 180 80 190 110

گروه II : 110 140 135 130 120 145

مثال : فشار خون در دو گروه از افراد مورد بررسی به صورت زیر بوده است

گروه I : 90 130 180 80 190 110

گروه II : 110 140 135 130 120 145

$$\bar{X}_1 = \frac{780}{6} = 130$$

پراکندگی زیاد

$$\bar{X}_2 = \frac{780}{6} = 130$$

پراکندگی کم

آیا میانگین به تنهایی می تواند معرف خصوصیت اطلاعات این دو گروه باشد؟

دامنه تغییرات Range

$$Range = x_{Max} - x_{Min}$$

براهنگی رید

80 90 110 130 170 190

Range = 190 - 80 = 110

براهنگی کم

110 120 130 145

Range = 145 - 110 = 35

۳۱

انحراف معیار Standard Deviation

• مهم ترین شاخص پراکندگی
• پراکندگی حول میانگین را نشان می دهد

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

انحراف معیار نمونه

۳۲

محاسبه انحراف معیار نمونه

X	10	12	14	15	17	18	18	24	جمع
									128

sum=128 n = 8 Mean =16

۳۳

محاسبه انحراف معیار نمونه

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

X	10	12	14	15	17	18	18	24	جمع
									128
(X - \bar{X})	-6	-4	-2	-1	1	2	2	8	0

n = 8 Mean =16

۳۴

محاسبه انحراف معیار نمونه

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

X	10	12	14	15	17	18	18	24	جمع
									128
(X - \bar{X})	-6	-4	-2	-1	1	2	2	8	0
(X - \bar{X}) ²	36	16	4	1	1	4	4	64	130

n = 8 Mean =16

۳۵

محاسبه انحراف معیار نمونه

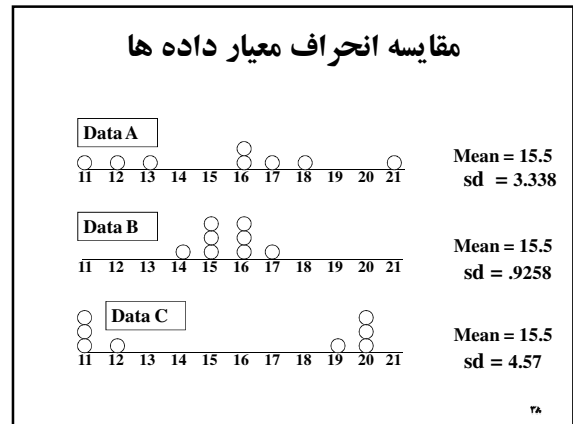
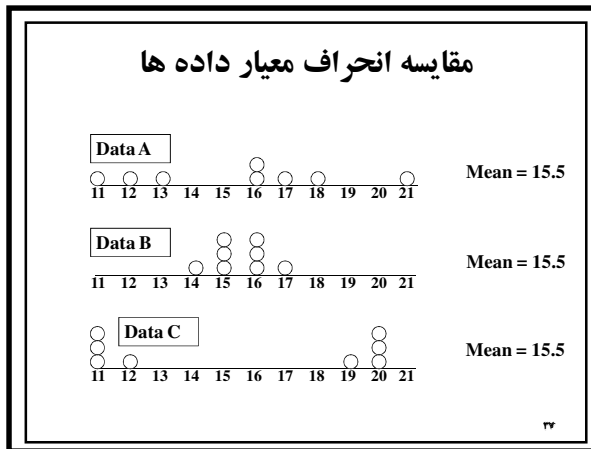
$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

X	10	12	14	15	17	18	18	24	جمع
									128
(X - \bar{X})	-6	-4	-2	-1	1	2	2	8	0
(X - \bar{X}) ²	36	16	4	1	1	4	4	64	130

n = 8 Mean =16

$$s = \sqrt{\frac{130}{8-1}} = \sqrt{18.57} = 4.31$$

۳۶



SYSTOLIC BP						
Age Groups	N	Mean	Std. Deviation	Minimum	Maximum	Range
<35	54	109.54	11.421	90	140	50
35-44	60	112.08	17.303	85	180	95
45-54	52	125.87	20.430	100	185	85
55-64	54	136.30	22.554	100	180	80
>=65	30	134.33	23.034	100	210	110
Total	250	122.30	21.848	85	210	125

۳۹

Coefficient of Variation ضریب تغییرات

> پراکندگی نسبت به میانگین را نشان می دهد
 > مقایسه پراکندگی دو یا چند گروه
 > مقایسه پراکندگی دو متغیر با واحد اندازه گیری متفاوت

$$CV = \left(\frac{S}{\bar{X}} \right) \cdot 100\%$$

۴۰

Coefficient of Variation ضریب تغییرات

$\bar{X} = 3000 \quad Sd = 300$

وزن نوزادان

$\bar{X} = 40Kg \quad Sd = 2$

وزن افراد ۱۲ ساله

۴۱

Coefficient of Variation ضریب تغییرات

$\bar{X} = 3000 \quad Sd = 300$

وزن نوزادان

$\bar{X} = 3000 \text{ gr} \quad Sd = 300 \text{ gr} \quad CV = \frac{300}{3000} \times 100 = 10\%$

وزن نوزادان

$\bar{X} = 40Kg \quad Sd = 2$

وزن افراد ۱۲ ساله

$\bar{X} = 40kg \quad Sd = 2kg \quad CV = \frac{2}{40} \times 100 = 5\%$

وزن افراد ۱۲ ساله

۴۲

ضرب تغییرات متغیرهای زیر را محاسبه کنید. کدام یک پراکندگی بیشتری دارد؟

SYSTOLIC BP						
Age Groups	N	Mean	Std. Deviation	Minimum	Maximum	Range
<35	54	109.54	11.421	90	140	50
35-44	60	112.08	17.303	85	180	95
45-54	52	125.87	20.430	100	185	85
55-64	54	136.30	22.554	100	180	80
≥65	30	134.33	23.034	100	210	110
Total	250	122.30	21.848	85	210	125

تمرین: در یک بررسی داده های زیر به دست آمده است
شاخصهای مرکزی و پراکندگی را برای آنها محاسبه نمایید:

2 6 5 3 9 1 4 10

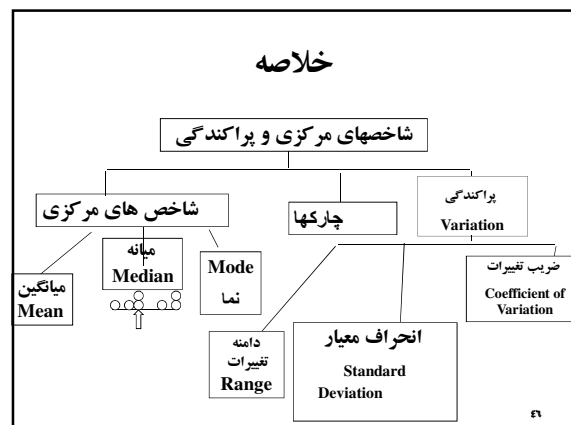
تمرین: در یک بررسی داده های زیر به دست آمده است
شاخصهای مرکزی و پراکندگی را برای آنها محاسبه نمایید:

2 6 5 3 9 1 4 10

$$R = 10 - 1 = 9 \quad \bar{X} = \frac{\sum x}{n} = \frac{40}{8} = 5$$

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{72}{7}} = 3.2$$

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{72}{7} = 10.3$$

$$CV = \frac{S}{\bar{X}} \times 100 = \frac{3.2}{5} \times 100 = 64$$


میانگین وزنی Weighted Mean

n_1	n_2	n_k	تعداد در گروه j
\bar{X}_1	\bar{X}_2	\bar{X}_k	میانگین گروه j

$$\bar{X} = \frac{1}{n} \sum_{j=1}^K n_j \bar{X}_j$$

$$n = \sum_{j=1}^k n_j$$

نسبت وزنی

n_1	n_2	n_k	تعداد در گروه j
p_1	p_2	p_k	نسبت در گروه j

$$P = \frac{1}{n} \sum_{j=1}^K n_j p_j$$

$$n = \sum_{j=1}^k n_j$$

اثر تغییر داده ها بر میانگین و انحراف معیار

$$Y = a + bX$$

$$\bar{Y} = a + b\bar{X}$$

$$S_Y = bS_X$$

$$S_Y^2 = b^2 S_X^2$$

۴۹

برخی خواص نماد سیگما

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

$$\sum_{i=1}^n cx_i = c \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n (x_i y_i) \neq \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

$$\sum_{i=1}^n \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i}$$

۵۰

