

شاخص های مرکزی و پراکندگی

جلسه سوم



○ شاخص های مرکزی و پراکندگی هم در داده های خام و هم در داده های طبقه بندی شده قابل محاسبه هستند.

بخش اول: شاخصهای مرکزی و پراکندگی در داده های خام

در پایان این بخش دانشجو باید بتواند:

- شاخص های مرکزی را در داده های خام محاسبه نماید
- شاخص های پراکندگی را در داده های خام محاسبه نماید.

شاخصهای مرکز یا معیارهای تمایل به مرکز :

○ منظور از معیار تمایل به مرکزی یا اندازه‌گیری گرایش مرکزی یک گروه داده، عددی است که توسط آن عدد، مرکزیت آن گروه داده مشخص می‌گردد.

○ مهمترین معیارهای تمایل مرکزی عبارت است از :

• میانگین حسابی

• میانه

• نما

میانگین حسابی (MEAN)

- ساده ترین و در بسیاری موارد مهمترین معیار تمایل به مرکز میانگین حسابی است و آن عبارت است از مجموع داده ها تقسیم بر تعداد آنها .

$$\mu = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{N} (X_1 + X_2 + \dots + X_N)$$

میانگین حسابی...

مثال : فرض کنید سن های ۶ کودک شرکت کننده در برنامه مراقبت بهداشتی به شکل زیر می باشد. میانگین سنی آنها را محاسبه کنید ؟

۱ ، ۲ ، ۲ ، ۳ ، ۲ ، ۴

$$\text{میانگین سنی} = \frac{1 + 2 + 2 + 3 + 2 + 4}{6} = \frac{14}{6} = 2/3$$

میانگین حسابی...

○ عیب بزرگ میانگین حسابی این است که متأثر از تک تک داده هاست و در صورتیکه داده های پرت و یا به اصطلاح اثرگذار در داده باشد به شدت و میانگین حسابی اثر می گذارد و بنابراین نمی تواند معیار خوبی برای مرکزیت داده باشد.

مثال : در مثال قبل فرض کنید که سن یک کودک به جای مثلاً ۲ سال ۱۱ سال باشد مشاهده می کنید که این یک عدد چه تأثیری بر میانگین می گذارد. این داده را به اصطلاح اثرگذار می گویند.

$$\text{میانگین سنی} = \frac{1, 2, 2, 3, 2, 4}{6} = \frac{1 + 2 + 2 + 3 + 11 + 4}{6} = \frac{23}{6} \cong 4$$

میانه (MEDIAN)

اگر داده ها همگن نباشد و داده پرت وجود داشته باشد، بهتر است از دیگر معیارهای تمایل به مرکز مانند میانه استفاده نمود.

○ میانه عبارت است از عددی که نیمی از داده ها کوچکتر از آن و نیم دیگر بزرگتر از آن باشند

مثال : میانه داده های ۶ ، ۴ ، ۹ ، ۱۶ ، ۱۵ ، ۱۸ ، ۲۱ به شکل زیر محاسبه می شود .

(۱) داده ها را مرتب می کنیم

(۲) فرد و یا زوج بودن تعداد داده ها را مشخص می کنیم ($n = 7$)

در این صورت $md = 15$ می باشد .

اگر تعداد داده ها زوج باشد میانگین دو عدد وسط میانه است.

میانہ...

مثال : میانہ داده های ۲۷ ، ۲۳ ، ۲۴ ، ۱۹ ، ۲۲ ، ۱۸ ، ۱۶ ، ۱۵ ، ۱۲ و ۲۸ عبارت است از

$$\text{Md} = \frac{19 + 22}{2} = 20.5$$

میانہ...

○ اگر حجم داده زیاد باشد، روش قبل کارآمد نخواهد بود. در این شرایط بهتر است از فرمول زیر استفاده کنیم.

$$\frac{(n + 1)}{2}$$

حال اگر تعداد داده ها به علاوه ۱ را بر ۴ تقسیم کنیم چارکها بدست میآیند. این چارک ها برای رسم نمودار جعبه ایی ضروری هستند. چارک ها داده ها را به چهار بخش مساوی تقسیم میکنند.

چار کھا...

$$\frac{n + 1}{4}$$

○ چارک اول

$$\frac{2(n + 1)}{4}$$

○ چارک دوم یا میانہ

$$\frac{3(n + 1)}{4}$$

○ چارک سوم

چندک ها

○ اگر مخرج فرمول قبل را 10 قرار دهیم دهک ها، اگر 100 قرار دهیم صدک ها بدست می آیند.

○ مثال: دهک 3 و 5 و 9 اینگونه بدست می آیند.

$$\frac{3(n+1)}{10}$$

$$\frac{5(n+1)}{10}$$

$$\frac{9(n+1)}{10}$$

○ مثال: صدک 30 و 50 و 75 اینگونه بدست می آیند.

$$\frac{30(n+1)}{100}$$

$$\frac{50(n+1)}{100}$$

$$\frac{75(n+1)}{100}$$

نمودار جعبه ایی

○ نمودار جعبه ای نموداری تصویری است که داده ها را بر اساس پنج مقدار نمایش می دهد. این مقادیر عبارتند از:

کوچک ترین داده

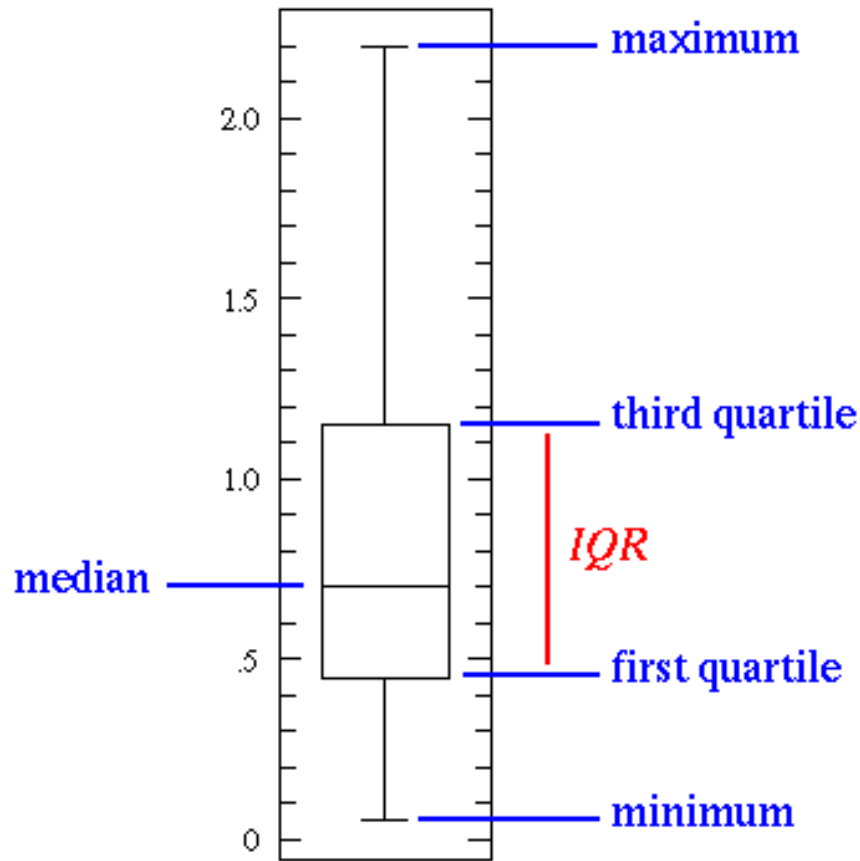
چارک اول

میانه

چارک سوم

بزرگ ترین داده

نمودار جعبه ایی...



نما (MODE)

- معیار دیگر تمایل مرکزی است. کاربرد آن نسبت به دو معیار دیگر کمتر ولی در مواردی که شیوع یک بیماری در سن خاص مدنظر باشد بهترین معیار است؛ عبارت است از داده ای که بیشترین فراوانی را دارد یک گروه داده ممکن است نما یا مد نداشته باشد مثل داده ها ۹، ۱۰، ۱۲، ۶، ۷ و ۸ که نما ندارد، یا ممکن است دو مدی و یا بیشتر باشد. مثال ۱۱، ۱۱، ۱۰، ۹، ۶، ۷، ۸ و ۸ که دارای دو نمای و یا مد ۸ و ۱۱ است.

شاخص های پراکندگی

هروقت یک شاخص مرکزی به ما دادند برای اینکه ببینیم این شاخص مرکزی چقدر دلچسب است شاخص پراکندگی را در نظر می گیریم تا ببینیم چقدر به این شاخص اعتماد داشته باشیم

○ دامنه

○ واریانس

○ انحراف استاندارد

○ ضریب تغییرات

دامنه (RANGE)

- تفاوت بین بزرگترین و کوچکترین داده.
- به میزان زیادی تحت تأثیر داده های پرت قرار می گیرد.
- برای داده های متقارن بدون هیچ برون هشتی مناسب است.

$$R = X_{\max} - X_{\min}$$

واریانس (VARIANCE)

○ منظور از واریانس، بررسی پراکندگی داده های کمی در اطراف میانگین داده ها است .

○ این پراکندگی ها هر چقدر در اطراف و نزدیک میانگین قرار بگیرند ، شرایط مطلوب تر خواهد بود.

$$\delta^2 = \frac{\sum (x - \bar{x})^2}{n}$$

انحراف استاندارد (SD)

- انحراف معیار نمونه ریشه دوم واریانس نمونه می باشد.
- واحدها، واحدهای اصلی هستند
- انحراف متوسط داده‌ها از میانگین خود را اندازه می گیرد.
- همچنین به میزان زیادی تحت تأثیر داده های پرت قرار دارد.

$$\delta = \sqrt{\delta^2}$$

ضریب تغییرات

- اگر هدف مقایسه پراکندگی توزیع وزن بدن در دو جامعه باشد که در یکی واحد اندازه گیری کیلوگرم و در دیگری پوند است، چون انحراف معیار برای جامعه اول برحسب کیلوگرم و در جامعه دوم برحسب پوند بیان خواهد شد در نتیجه این مقایسه بر اساس انحراف معیار منطقی نخواهد بود.
- به منظور رفع این اشکال از نسبت انحراف معیار به میانگین که معمولاً بصورت درصد بیان می شود استفاده می گردد. کمیت حاصل را که یک مشخص کننده نسبی است با CV نشان می دهیم.

ضریب تغییرات...

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

اگر میانگین و انحراف معیار درجه حرارت بدن برای افراد یک جامعه به ترتیب اعداد $36/9$ و $0/18$ درجه باشد و میانگین و انحراف معیار تعداد ضربان نبض به ترتیب 78 و 9 بار در دقیقه باشد، برای محاسبه ضریب تغییرات خواهیم داشت:

$$CV_1 = \frac{0.18}{36.9} \times 100 = 0.5$$

$$CV_2 = \frac{9}{78} \times 100 = 11.5$$

سوال:

جدول را کامل کنید.

	Mode	median	mean	range	δ^2	δ	CV
۱, ۲, ۲, ۳							
۲, ۳, ۳, ۴							
۲۰, ۳۰, ۳۰, ۴۰							

سوال:

نمودار جعبه ایی داده های زیر را رسم کنید. صدک ۶۸ و دهک ۷ را بیابید.

43 ,45 ,53 ,56, 56, 57, 58, 66, 67, 73, 74, 79, 80, 80, 81 ,81, 81,

82, 83, 83, 84 , 88, 91 ,91 ,91 ,92,92,97 ,99 ,99, 100, 100 ,102

,102 ,102 102 ,103,104,107 ,108 ,109, 113 ,118 ,118 ,121,

123, 126, 128, 137, 138, 139, 144, 147 ,147,156,162, 174 ,

178, 179, 184, 191, 198, 124, 214 ,243 ,249, 249 ,329 ,380

,403 ,511 ,522 , 598

بخش اول: شاخصهای مرکزی و پراکندگی در داده های خام

در پایان این بخش دانشجو باید بتواند:

- شاخص های مرکزی را در داده های طبقه بندی محاسبه نماید
- شاخص های پراکندگی را در داده های طبقه بندی محاسبه نماید.

میانگین

در مطالعاتی که نتیجه مشاهدات بصورت کمیت پیوسته است، چون معمولاً جدول توزیع فراوانی بصورت گروه بندی شده ارائه می شود، محاسبه میانگین تنها بطور تقریب امکان پذیر می گردد.

روش محاسبه میانگین مانند قبل است با این تفاوت که مقدار متوسط هر گروه (یعنی مرکز هر دسته) را در فراوانی های متناظر آن گروه ضرب و سپس به تعداد کل فراوانی ها تقسیم می کنیم.

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

مثال

حدود دسته ها	f_i	مرکز دسته (x_i)
90-100	15	95
100-110	57	105
110-120	82	115
120-130	155	125
130-140	105	135
140-150	86	145
150-160	33	155
160-170	29	165
170-180	21	175
180-190	5	185
190-200	16	195
جمع	604	

اطلاعات داده شده در جدول زیر مربوط به فشار خون سیستولیک ۶۰۴ نفر مرد ۳۵ سال به بالای چند روستای شهرستان رودسر در سال ۱۳۵۰ است . میانگین آنرا حساب کنید.

میانہ

در این حالت ابتدا فراوانی های تجمعی را محاسبه می کنیم (F_i معرف فراوانی تجمعی X_i). سپس ردیفی که مقدار فراوانی تجمعی آن برابر یا بزرگتر از $(N+1) / 2$ باشد بعنوان ردیف میانہ انتخاب می کنیم.

مثال

تعداد دندان های پوسیده X_i	فراوانی	فراوانی تجمعی F_i
0	70	70
1	36	106
2	29	135
3	19	154
4	15	169
5	14	183
6	10	193
7	6	199
8	4	203
9	1	204
10	1	205
جمع	205	

جدول مربوط به داده های
تعداد دندان های پوسیده
شیری ۲۰۵ نفر دانش آموز
۱۲ ساله می باشد، میانه آنرا
بیابید.

واریانس

برای محاسبه واریانس در داده های طبقه بندی از فرمول زیر استفاده میکنیم:

$$\sigma^2 = \frac{\sum f_i x_i^2 - \left(\frac{(\sum f_i x_i)^2}{\sum f_i}\right)}{\sum f_i}$$

مثال

واریانس را برای داده های جدول زیر بیابید.

کرانه های گروه	N_i	X_i	$N_i X_i$	$N_i X_i^2$
90-100	15	95	1425	135375
100-110	57	105	5985	628425
110-120	82	115	9430	1084450
120-130	155	125	19375	2421875
130-140	105	135	14175	1913625
140-150	86	145	12470	1808150
150-160	33	155	51115	792825
160-170	29	165	4785	789525
170-180	21	175	3675	643125
180-190	5	185	925	171125
190-200	16	195	3120	608400
جمع	604	-	80480	10996900

$$\begin{aligned} \sigma^2 &= \frac{10996900 - \left(\frac{(80480)^2}{604}\right)}{604} \\ &= 452.55 \end{aligned}$$

انحراف استاندارد

این شاخص در داده های طبقه بندی نیز با جذر واریانس برابر است.