

شخص های مرکزی و پراکنده

شخص‌های توصیفی

✓ برای توصیف هر چه بهتر متغیرهایی که داده‌های آنها کمی هستند، از دو گروه از شخص‌های مرکزی و شخص‌های پراکنده استفاده می‌کنیم.

✓ اگر بخواهیم نحوه تمرکز (مرکزیت) داده‌ها را توصیف نمایم از شخص‌های دسته اول و اگر چنانچه بخواهیم نحوه پراکندگی داده‌ها را توصیف نمایم از شخص‌های دسته دوم استفاده می‌کنیم.

شخص‌های توصیفی

شخص‌های پرانندگی

- واریانس
- انحراف معیار
- دامنه تغییرات
- ضریب تغییرات
- ضریب چولگی

شخص‌های مرکزی

- میانگین
- میان
- نما
- خدک‌ها

شاخص های مرکزی

✓ شاخص های **عددی** هستند که میزان گرایش به مرکز داده ها را مشخص می کنند.

➤ میانگین حسابی (Mean):

یکی از مهمترین شاخص های مرکزی است که از رابطه زیر به دست می آید:

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

ویژگی‌های میانگین:

✓ همیشه مجموع اختلاف داده‌ها از میانگین حسابی، صفر است.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

✓ هرگاه به هر یک از داده‌ها عدد ثابتی را مانند a اضافه یا کم کنیم، میانگین داده‌های حاصل، برابر با

میانگین داده‌های قبلی به اضافه (یا منهای) عدد ثابت a خواهد بود.

$$Y_i = X_i \pm a \rightarrow \bar{Y} = \bar{X} \pm a$$

ویژگی‌های میانگین:

✓ هرگاه عدد ثابت a را در هر یک از داده‌ها ضرب کنیم، میانگین داده‌های حاصل، برابر میانگین داده‌های قبلی ضرب در عدد ثابت a خواهد بود.

$$Y_i = aX_i \Rightarrow \bar{Y} = a\bar{X}$$

✓ در مورد تقسیم هم به صورت ضرب عمل می‌کنیم.

$$Y_i = \frac{X_i}{a} \rightarrow \bar{Y} = \frac{\bar{X}}{a}$$

ویژگی‌های میانگین:

✓ 5- اگر یک مجموعه n_1 تایی از داده‌های یک صفت با میانگین $\bar{X}_1$ و یک مجموعه n_2 تایی از داده‌ها با میانگین $\bar{X}_2$ داشته باشیم، آنگاه میانگین کل این دو مجموعه برابر است با:

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

✓ این موضوع برای k جمعیت و به صورت زیر قابل تعمیم است:

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_k \bar{X}_k}{n_1 + n_2 + \dots + n_k}$$

شاخص های مرکزی

✓ میانه (Median):

- ✓ عددی است در بین داده ها که اگر آنها را **مرتب** کرده باشیم، نیمی از داده ها کمتر از آن واقع شوند. به عبارت دیگر داده ای که در وسط قرار می گیرد و آن را با Md نشان می دهند.
- البته باید به تعداد داده ها که **فرد یا زوج** باشند نیز توجه داشت.

شاخص های مرکزی

✓ نما (Mode):

✓ داده ای است از بین داده ها که فروانی آن بیشتر از سایر داده ها باشد و آن را با Mo نمایش

می دهیم.

شاخص های مرکزی

✓ مثال: شاخص های مرکزی را برای داده های جدول زیر به دست آورده ایم.

تعداد دندانها	فراوانی f_i	$x_i \cdot f_i$	F_i
۰	۲	$۰ \times ۲ = ۰$	۲
۱	۹	$۱ \times ۹ = ۹$	۱۱
۲	۱۸	$۱۸ \times ۲ = ۳۶$	۲۹
۳	۱۰	$۱۰ \times ۳ = ۳۰$	۳۹
۴	۷	$۴ \times ۷ = ۲۸$	۴۶
۵	۴	$۴ \times ۵ = ۲۰$	۵۰
	۵۰	۱۲۳	

✓ در این مثال، $n=50$ است.

✓ میانگین 2.46

✓ میانه 2

✓ بقیه خدک ها را می توان به همین صورت به دست آورد.

شاخص های مرکزی

✓ مثال: برای داده های جدول فراوانی زیر شاخص های مرکزی را به دست آورده ایم.

حدود طبقات	فراوانی	نماینده		Fi
۱ - ۳	۸	۲	$۸ \times ۲ = ۱۶$	۸
۳ - ۵	۵	۴	$۴ \times ۵ = ۲۰$	۱۳
۵ - ۷	۱۳	۶	$۱۳ \times ۶ = ۷۸$	۲۶
۷ - ۹	۱۵	۸	$۱۵ \times ۸ = ۱۲۰$	۴۱
۹ - ۱۱	۹	۱۰	$۱۰ \times ۹ = ۹۰$	۵۰
	۵۰		۳۲۴	

✓ میانگین 6.48؛ که از تقسیم 324 بر 50 به دست آمده است.

✓ همچنین 7.5 و در طبقه چهارم است

✓ که آن را از رابطه زیر به دست آورده ایم:

$$Mo = l_i + \frac{D_1}{D_1 + D_2} \times C \Rightarrow ۷ + \frac{۲}{۲ + ۶} \times ۲ = ۷/۵$$

Li = کران پائین طبقه ای است که فراوانی مطلق آن ماکزیمم است. (طبقه نمادار)

D_1 = تفاضل فراوانی مطلق طبقه نمادار با طبقه قبل از آن است،

D_2 = تفاضل فراوانی مطلق طبقه نمادار با طبقه بعد از آن است.

C = فاصله طبقات

شاخص های پراکندگی

✓ شاخص هایی عددی هستند که میزان پراکندگی داده ها حول میانگین را نشان می دهند.

➤ مثال: دو گروه از بیماران را از نظر مصرف تعداد قرص های مسکن در یک ماه مورد بررسی قرار داده ایم که

نتایج در دو جدول زیر آمده است.

X_i	f_i	X_i	f_i
۲	۲	۴	۵
۳	۳	۵	۱۰
۴	۱۵	۶	۷۰
۵	۱۰	۷	۱۰
۶	۱۰	۸	۵
۷	۴۰		
۸	۱۵		
۹	۳		
۱۰	۲		

✓ برای این دو گروه، میانگین، میانه و نماییکسان و برابر 6 است.

✓ در این دو مجموعه داده، تفاوت هایی را مشاهده می کنید که با

شاخص های مرکزی قابل بیان نیستند.

شاخص های پراکندگی

✓ واریانس (Variance):

✓ واریانس یکی از معیارهای بسیار مناسب در بیان میزان پراکندگی داده ها است. زیرا برخلاف بعضی از شاخص ها، از همه داده ها برای محاسبه آن استفاده می شود. واریانس را می توان از یکی از دو رابطه زیر به دست آورد:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1)$$

$$S^2 = \overline{x^2} - \bar{x}^2 \quad (2)$$

شاخص های پراکندگی

✓ مثال: برای داده های زیر واریانس را به دست می آوریم. 3 2 4 3 6 3 5 6

✓ ابتدا میانگین داده ها را حساب می کنیم و سپس از فرمول (۱) واریانس را محاسبه می کنیم، داریم:

$$\bar{x} = \frac{1}{8} (3 + 2 + 4 + 3 + 6 + 3 + 5 + 6) = \frac{32}{8} = 4$$

$$S^2 = \frac{1}{8} [(3-4)^2 + (2-4)^2 + (4-4)^2 + (3-4)^2 + (6-4)^2 + (3-4)^2 + (5-4)^2 + (6-4)^2] = \frac{1}{8} (16) = 2$$

شاخص های پرانندگی

✓ مثال: برای داده های جدول توزیع فراوانی زیر واریانس را محاسبه کنید.

X_i	f_i	$f_i x_i$	$f_i x_i^2$
۰	۳	۰	$۰ \times ۳ = ۰$
۱	۷	۷	$۱ \times ۷ = ۷$
۲	۲۱	۴۲	$۴ \times ۲۱ = ۸۴$
۳	۶	۱۸	$۹ \times ۶ = ۵۴$
۴	۷	۲۸	$۱۶ \times ۷ = ۱۱۲$
۵	۶	۳۰	$۲۵ \times ۶ = ۱۵۰$
جمع	۵۰	۱۲۵	۴۰۷

ابتدا با تشکیل دو ستون آخر جدول مقادیر زیر را به دست می آوریم:

$$\bar{X} = \frac{۱۲۵}{۵۰} = ۲/۵ \quad \overline{X^2} = \frac{۴۰۷}{۵۰} = ۸/۱۴$$

پس واریانس از فرمول (۲) به صورت زیر بدست می آید.

$$S^2 = ۸/۱۴ - (۲/۵)^2 = ۱/۸۹$$

شاخص های پراکندگی

✓ خواص واریانس:

$$x_1 = x_2 = \dots = x_n \Rightarrow S_x^2 = 0$$

✓ اگر همه داده ها با هم برابر باشند واریانس داده ها صفر است و برعکس

✓ واریانس داده ها مقداری مثبت است. $S^2 \geq 0$

$$Y_i = a \pm X_i \rightarrow S_Y^2 = S_X^2$$

✓ اگر به داده ها عدد ثابتی مانند a را اضافه یا کم کنیم، واریانس تغییری نمی کند.

$$Y_i = ax_i \Rightarrow S_Y^2 = a^2 S_X^2$$

✓ اگر داده ها را در عدد ثابتی مانند a ضرب کنیم، واریانس آنها در مجذور a ضرب می شود.

$$Y_i = \frac{X_i}{a} \rightarrow S_Y^2 = \frac{S_X^2}{a^2}$$

✓ به همین صورت در مورد تقسیم بر عدد ثابتی مانند a ، داریم:

شاخص های پرآلودگی

✓ انحراف معیار (std Deviation) :

✓ انحراف معیار از جذر مثبت واریانس به دست می آید. $s = \sqrt{s^2}$

➤ انحراف معیار مثال قبل به این صورت به دست می آید. $s = \sqrt{19/57} = 4/42$

شاخص های پراکندگی

✓ ضریب تغییرات

✓ ضریب تغییرات داده ها که آن را با CV نشان می دهیم از نسبت انحراف معیار به میانگین به دست می آید

$$CV = \frac{S}{\bar{x}} \times 100 \quad \text{و به صورت زیر می نویسیم.}$$

✓ مثال: اگر در یک نمونه، میانگین داده ها 5 و واریانس داده ها 4 باشد؛ ضریب تغییرات داده ها به صورت زیر به

$$CV = \frac{2}{5} \times 100 = 40 \quad \text{دست می آید:}$$

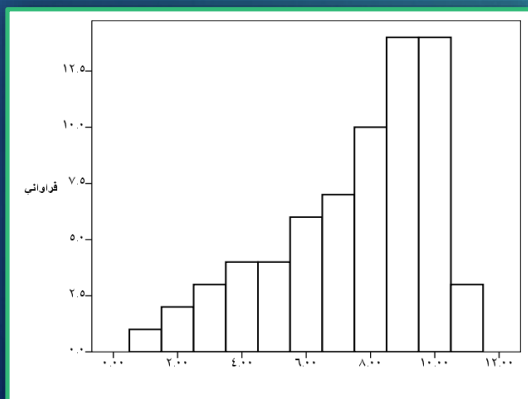
شاخص های پراکندگی

✓ تقارن و چولگی (Skewness):

✓ ممکن است دو مجموعه داده، از لحاظ کرایش به مرکز و پراکندگی تفاوت چندانی با هم نداشته باشند ولی از نظر تقارن با یکدیگر متفاوت باشند. شاخص های دیگری برای اندازه گیری میزان تقارن یا عدم تقارن داده ها وجود دارند.

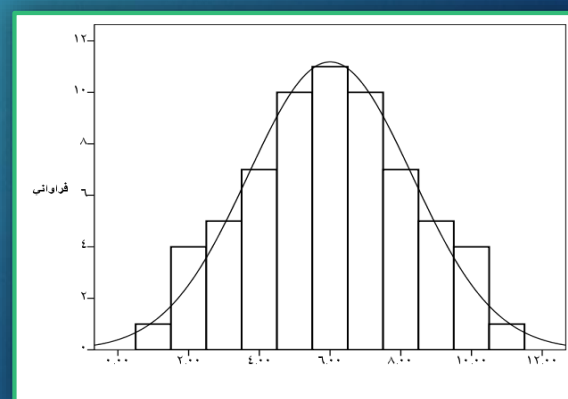
عدم تقارن (چولگی)

تقارن در توزیع داده ها



توجه داشته باشید:
در توزیع های متقارن
تقریباً میانگین، میانه و
نما بر هم منطبق اند.

$$Mo = \bar{x} = Md$$



شاخص های پراکندگی

✓ ضریب چولگی پیرن

✓ یکی از شاخص های تعیین میزان تقارن و عدم تقارن داده ها، ضریب چولگی پیرن است.

اگر $\bar{x}$ میانگین، Md میانه و S انحراف معیار یک مجموعه داده باشند؛ ضریب چولگی که منسوب به پیرن است،

از رابطه زیر بدست می آید:

$$SK = \frac{3(\bar{x} - Md)}{S} \quad \text{ضریب چولگی}$$

شاخص های پرانندگی

✓ مثال: اگر میانگین، میانه و انحراف معیار داده ها را به صورت مقابل در اختیار داشته باشیم؛

$$S = ۴/۴۲$$

$$\bar{X} = ۱۸/۲۷$$

$$Md = ۱۹/۴$$

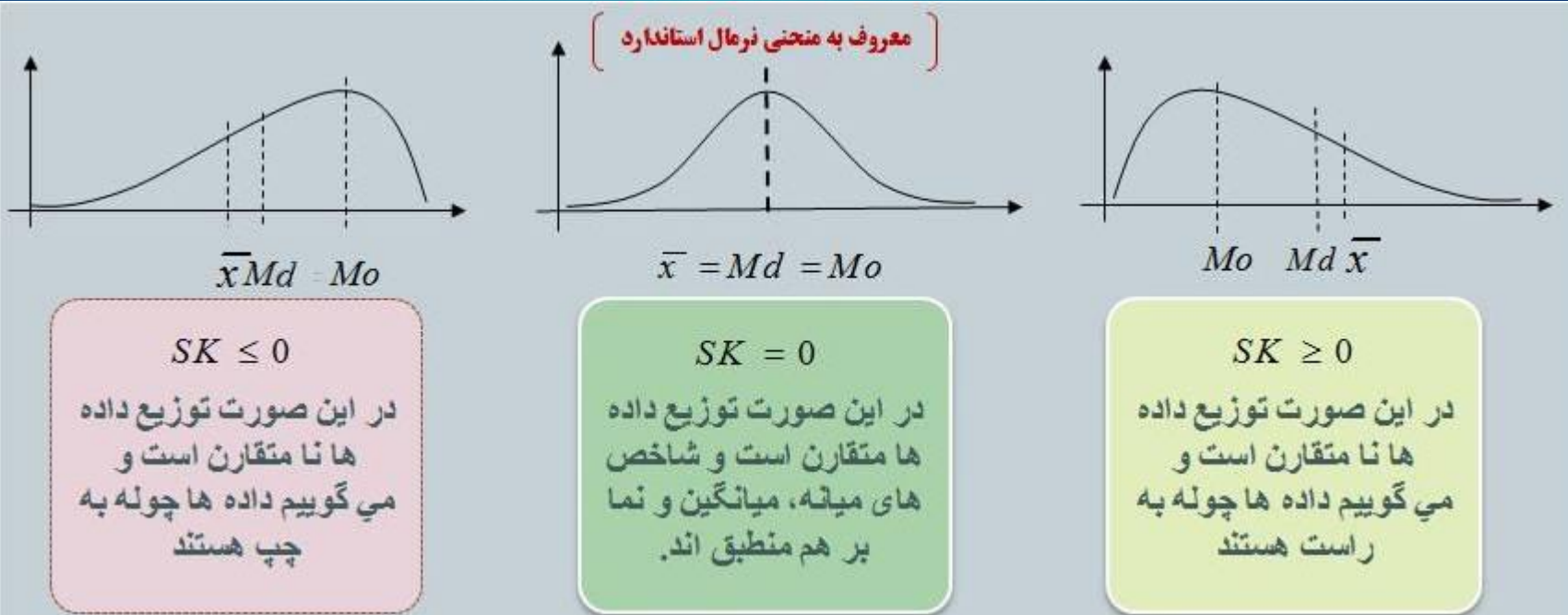
ضریب چولگی به صورت زیر به دست می آید.

$$Sk = \frac{۳(\bar{X} - Md)}{S} = \frac{۳(۱۸/۲۷ - ۱۹/۴)}{۴/۴۲} = -۰/۷۷$$

ضریب چولگی پیرسن

✓ تفسیر چولگی

✓ ضریب چولگی ممکن است مقداری منفی، مثبت یا صفر باشد. بر این اساس وضعیت قرار گرفتن سه شاخص مهم مرکزی، میانگین، میانه و نماراد حالت های مختلف ببینید.



✓ اگر داده‌ها تقریباً متقارن باشند، در صورتی که میانگین و انحراف معیار را در اختیار داشته باشیم، می‌توان به طور تجربی، فاصله‌های زیر را برای داده‌ها در نظر گرفت:

در حدود ۶۸ درصد داده‌ها در فاصله $(\bar{x} - s, \bar{x} + s)$ قرار دارند.

در حدود ۹۵ درصد داده‌ها در فاصله $(\bar{x} - 2s, \bar{x} + 2s)$ قرار دارند.

در حدود ۹۹ درصد داده‌ها در فاصله $(\bar{x} - 3s, \bar{x} + 3s)$ قرار دارند.

✓ مثال: توزیع فشار خون بیماران دیابتی تقریباً متقارن است. اگر میانگین فشار سیستولیک آنها 15 و واریانس 2.25 باشد، تقریباً 68 درصد این بیماران فشار خونشان در چه حدودی است؟

✓ تقریباً 68 درصد فشار خون در فاصله (13.5 و 16.5) قرار دارند، زیرا:

$$(\bar{x} - S, \bar{x} + S) = (15 - 1/5, 15 + 1/5) = (13/5, 16/5)$$