



دوره آمار توصیفی

یادگیری خیلی ساده، خیلی عمیق



دکتر عقیوب نژاد

راه‌های ارتباطی



Nogamcenter.com



۰۹۰۲۲۵۴۴۶۰۰



Nogamcenter

جایگاه روش تحقیق و آمار

تا اینجا ما با استفاده از روش‌های پژوهش (پژوهش‌های کمی مانند پیمایشی، آزمایشی، پس‌رویدادی و ...) داده‌ها را گردآوری کرده‌ایم. این داده‌ها خام هستند و ما باید به آنها معنا بدهیم و آنها را تحلیل کنیم. این کار توسط آمار انجام می‌شود.

روش‌های پژوهش (گردآوری داده‌های خام)



تحلیل و معنادهی به داده‌ها



محاسبه شاخص‌های آماری در سطح گروه نمونه n (آمار توصیفی):



استنباط پارامترهای جامعه آماری N (آمار استنباطی)

تست بزینم ۱۰۵

ارشد ۹۵: کدام مورد هدف از محاسبه آمارها (statistics) براساس داده‌های حاصل از نمونه است؟

1. برآورد ویژگی‌های جامعه از طریق تعمیم به نمونه موردنظر
2. توصیف ویژگی‌های نمونه و جامعه بر اساس آمارها
3. کاهش حجم داده‌ها و افزایش دقت شاخص‌های آماری
4. توصیف نمونه و فرض آزمایی پیرامون ویژگی‌های جامعه

شیوه‌های خلاصه کردن داده‌ها در آمار توصیفی

- آمار توصیفی
- تنظیم و طبقه بندی داده‌ها در **جدول توزیع فراوانی**
 - خلاصه کردن و توضیح داده‌ها به وسیله **رسم نمودارها**
 - خلاصه کردن داده‌ها در شاخص‌های مرکزی
 - خلاصه کردن داده‌ها در **شاخص‌های پراکندگی**

جداول توزیع فراوانی

تصور کنید یک سری عدد داریم که شامل نمره‌های ۲۰ نفر در یک کلاس است به ترتیب زیر:

۱۱ ۱۲ ۱۳ ۶ ۱۲ ۱۳ ۹ ۱ ۲ ۱۴ ۱۴ ۵ ۱۴ ۱۵ ۱۶ ۱۷ ۱۸ ۱۶ ۲۰ ۱۰

مراحل ترسیم جدول توزیع فراوانی:

$$R = X_H - X_L + 1$$

$$R = (20 - 1) + 1 = 20$$

❖ محاسبه دامنه تغییرات (R)

$$K = 1 + 3/3 \text{ Log} n$$

❖ انتخاب تعداد طبقات

$$i = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}} \rightarrow i = \frac{R}{k}$$

❖ محاسبه فاصله طبقات (i)



شروع طبقات معمولاً از کوچک‌ترین عدد است. کوچک‌ترین عدد، ۱ می‌باشد پس طبقه اول می‌شود ۱ تا ۴

طبقات	f_i
۱۷-۲۰	۳
۱۳-۱۶	۸
۹-۱۲	۵
۵-۸	۲
۱-۴	۲

$$\varepsilon f_i = n$$



فراوانی نسبی

طبقات	f_i	f/n
۱۷-۲۰	۳	۰/۱۵
۱۳-۱۶	۸	۰/۴
۹-۱۲	۵	۰/۲۵
۵-۸	۲	۰/۱
۱-۴	۲	۰/۱

$$\sum \frac{f}{n} = 1$$



فراوانی درصدی

$$\mathcal{E} \frac{f}{n} \times 100 = 100$$

طبقات	f_i	f/n	$f/n \times 100$
۱۷-۲۰	۳	۰/۱۵	۱۵
۱۳-۱۶	۸	۰/۴	۴۰
۹-۱۲	۵	۰/۲۵	۲۵
۵-۸	۲	۰/۱	۱۰
۱-۴	۲	۰/۱	۱۰



تست بزنیم ۱۰۶

فراوانی نسبی طبقه‌ای از داده‌ها $۰/۲۵$ می‌باشد. اگر مجموع فراوانی‌ها برابر ۲۰ باشد، فراوانی مطلق و فراوانی درصدی این طبقه را محاسبه کنید.

۱. $۲۵ - ۵$

۲. $۲۰ - ۱۰$

۳. $۳۰ - ۱۵$

۴. $۲۵ - ۱۰$



$$\frac{f}{n} = 0/25 \longrightarrow \frac{f}{20} = 0/25 \longrightarrow f = 5$$

$$\epsilon \frac{f}{n} \times 100 \longrightarrow 0/25 \times 100 = 25$$



فراوانی تراکمی

طبقات	f_i	f/n	$f/n \times 100$	cf
۱۷-۲۰	۳	۰/۱۵	۱۵	۲۰
۱۳-۱۶	۸	۰/۴	۴۰	۱۷
۹-۱۲	۵	۰/۲۵	۲۵	۹
۵-۸	۲	۰/۱	۱۰	۴
۱-۴	۲	۰/۱	۱۰	۲

فراوانی تراکمی طبقه آخر همیشه برابر است با مجموع فراوانی‌ها

$$cf = \varepsilon f = n \text{ طبقه آخر}$$



تست بنزیم ۱۵۷

دکتری ۹۹: درجدول فروانی زیر، تعداد داده‌ها چقدر است؟

فروانی نسبی	فروانی تجمعی	فراوانی مطلق	طبقات
		۲	۱-۵
۰/۳	۵		۶-۱۰
			۱۱-۱۵

۷ .1

۱۰ .2

۱۵ .3

۱۷



$$\frac{f}{n} = 0/3 \longrightarrow n = 10$$

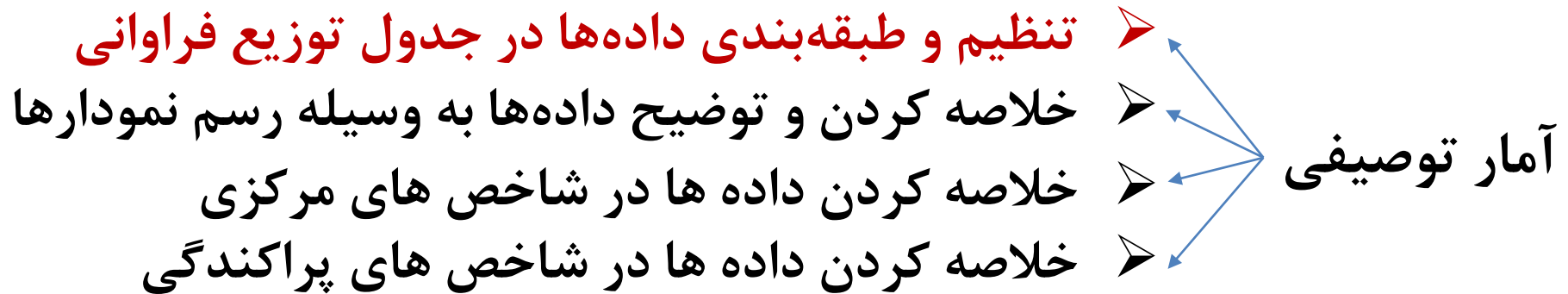


نقطه میانی

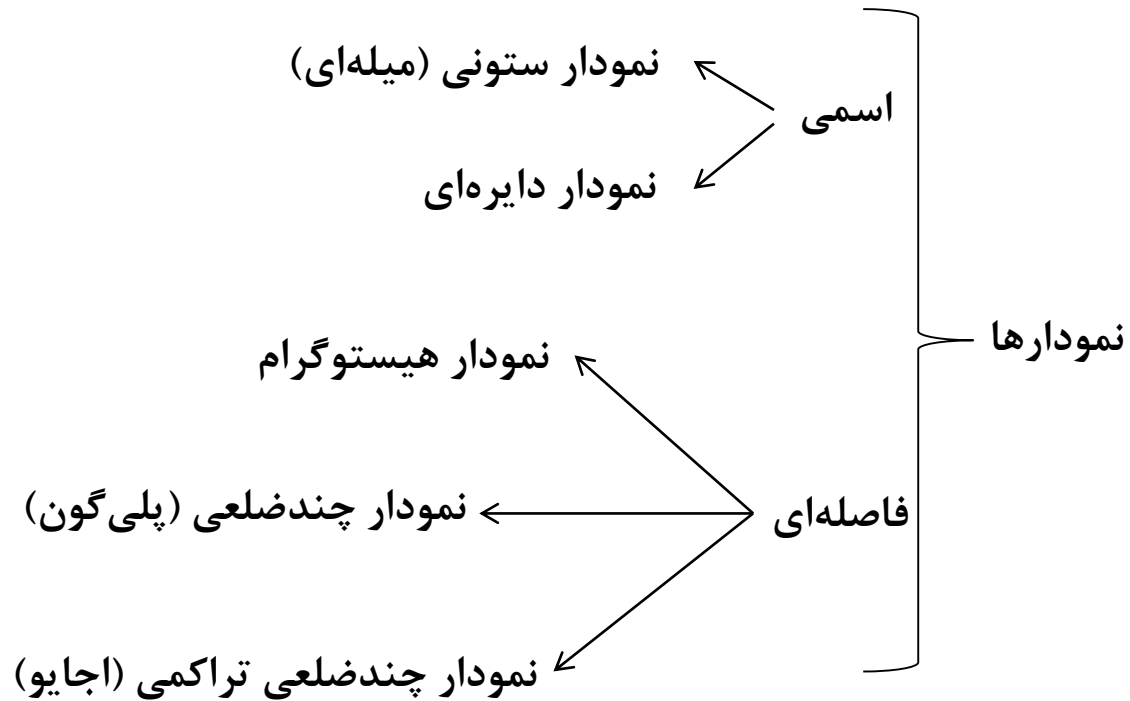
طبقات	f_i	f/n	$f/n \times 100$	cf	x_i
۱۷-۲۰	۳	۰/۱۵	۱۵	۲۰	۱۸/۵
۱۳-۱۶	۸	۰/۴	۴۰	۱۷	۱۴/۵
۹-۱۲	۵	۰/۲۵	۲۵	۹	۱۰/۵
۵-۸	۲	۰/۱	۱۰	۴	۶/۵
۱-۴	۲	۰/۱	۱۰	۲	۲/۵



شیوه‌های خلاصه کردن داده‌ها در آمار توصیفی



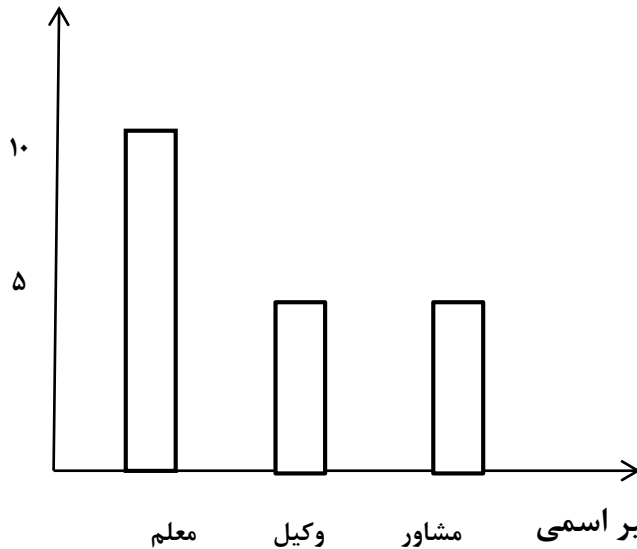
نمودارها



نمودار ستونی

نکته: در نمودار ستونی، میله‌ها یا ستون‌ها از هم جدا هستند. چراکه متغیر در نمودار ستونی، اسمی است و همانطور که قبلاً گفتیم در سطح اسمی بین متغیرها هیچ ارتباطی وجود ندارد.

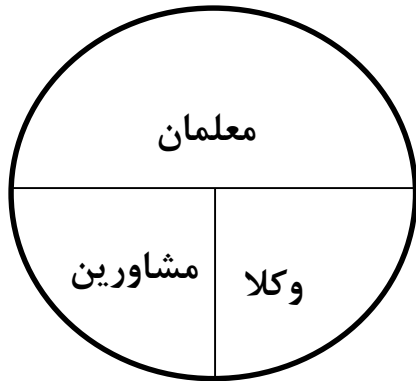
f (فراوانی)



نکته: در نمودار ستونی، روی محور عمودی فراوانی و روی محور افقی سطح طبقات یا ارزش‌های آن نوشته می‌شود.



ویژگی‌های جمعیت شناختی



فرمول محاسبه زاویه متغیرها در نمودار دایره-
$$\alpha = \frac{f}{n} \times 360$$

ای

در این فرمول f فراوانی مطلق هر طبقه

و n تعداد گروه نمونه می‌باشد.



تست بنزیم ۱۰۸

زاویه اختصاصی داده شده به طبقه‌ای از داده‌ها برابر با ۴۵ درجه می‌باشد. اگر فراوانی تراکمی طبقه آخر ۳۶ نفر باشد، فراوانی مطلق این طبقه را محاسبه کنید.

۴.۱

۴/۵.۲

۴/۷۵.۳

۵.۴



نمودار هیستوگرام

طبقات	f_i	cf	x_i
۱۷-۲۰	۳	۲۰	۱۸/۵
۱۳-۱۶	۸	۱۷	۱۴/۵
۹-۱۲	۵	۹	۱۰/۵
۵-۸	۲	۴	۶/۵
۱-۴	۲	۲	۲/۵

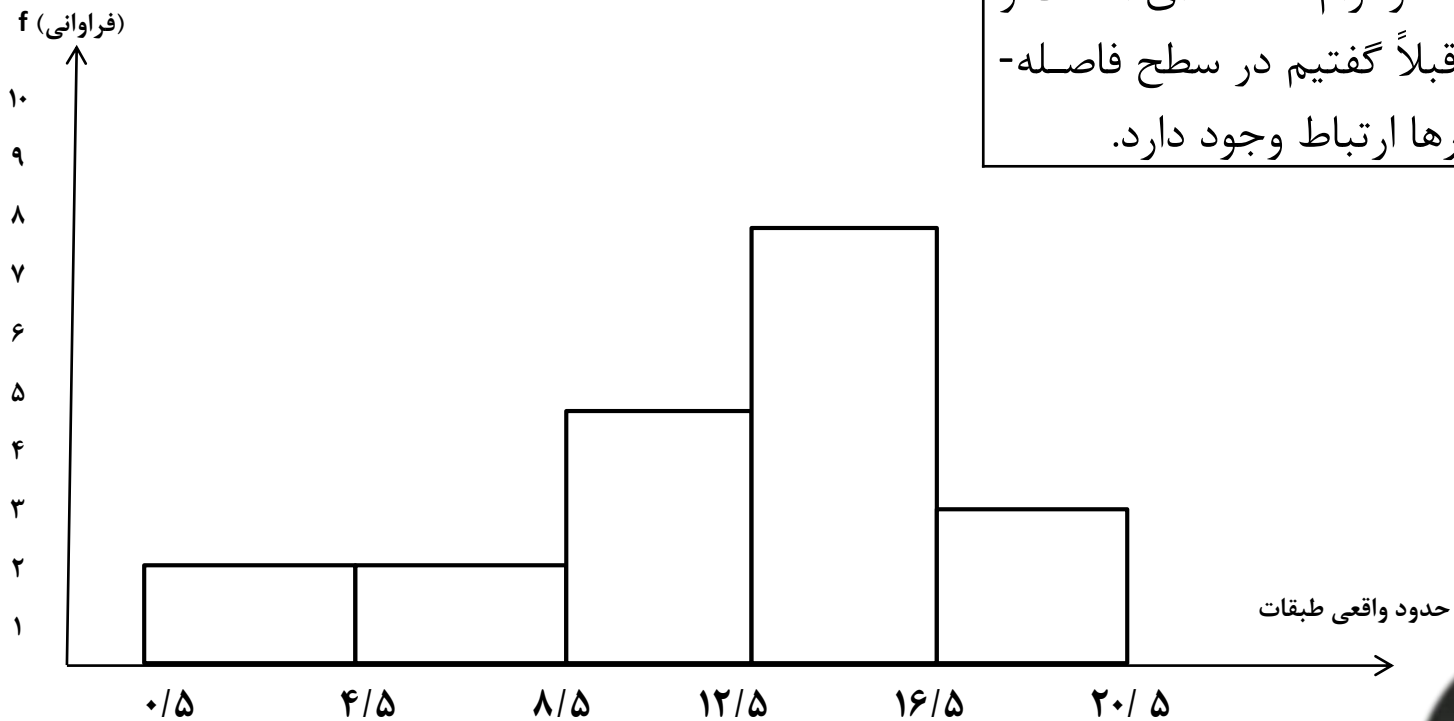
✓ در محور y ها (محور عمودی) فراوانی

✓ محور x ها (محور افقی) حدود واقعی طبقات



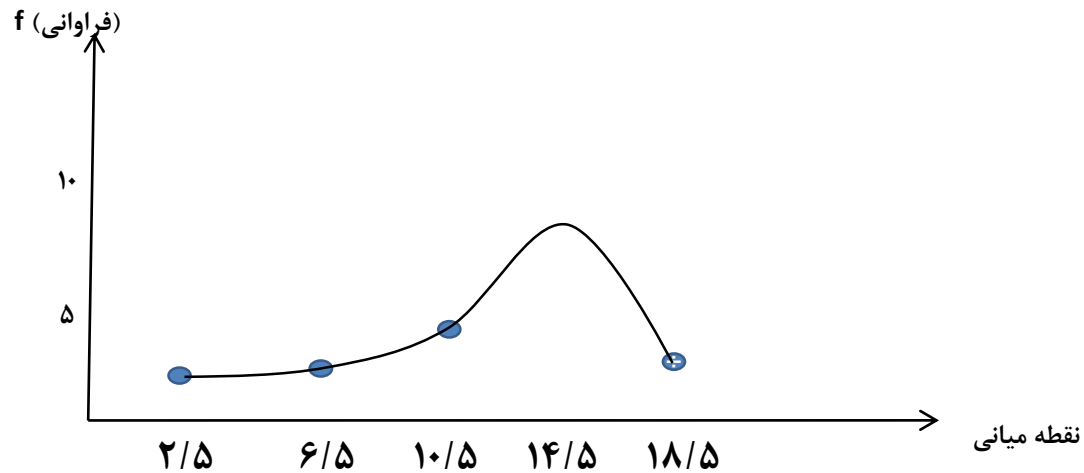
طبقات	f_i	cf	x_i
۱۷-۲۰	۳	۲۰	۱۸/۵
۱۳-۱۶	۸	۱۷	۱۴/۵
۹-۱۲	۵	۹	۱۰/۵
۵-۸	۲	۴	۶/۵
۱-۴	۲	۲	۲/۵

نکته: چرا ستون‌ها در نمودار هیستوگرام به هم متصل هستند؟ چون متغیر در نمودار هیستوگرام فاصله‌ای است و همانطور که قبلاً گفتیم در سطح فاصله‌ای بین متغیرها ارتباط وجود دارد.



نمودار چندضلعی (پلی گون)

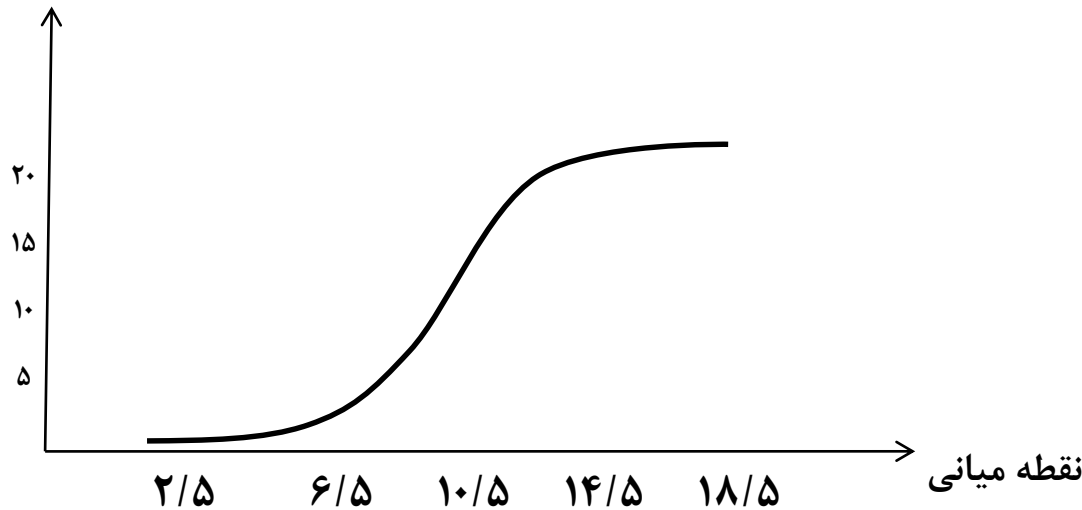
- پرکاربردترین نمودار؛ شکل توزیع / امکان مقایسه دو دسته داده



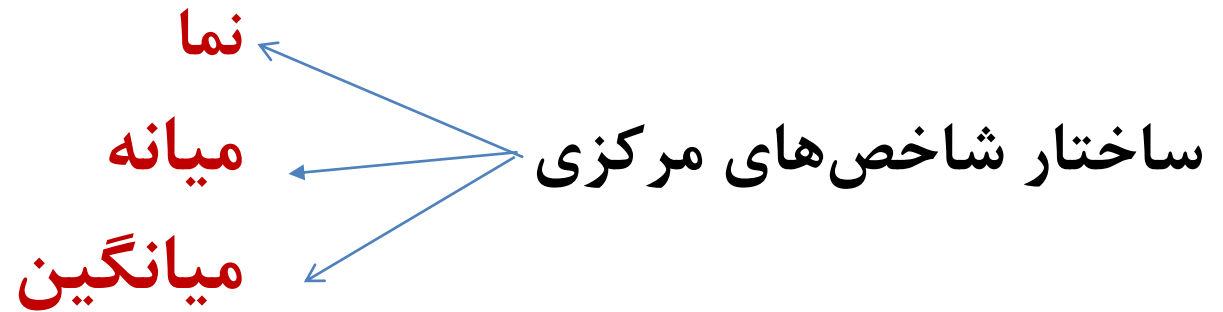
نمودار چند ضلعی تراکمی (اجایو)

- در این نمودار هم مانند نمودار چندضلعی روی محور X نقطه میانی نوشته می شود اما روی محور Y در این نمودار فراوانی تراکمی را باید بنویسیم.

فراوانی تراکمی



شاخص‌های مرکزی



نما (مد) MO

نما: شاخصی که **بیشترین فراوانی** را در توزیع دارد.

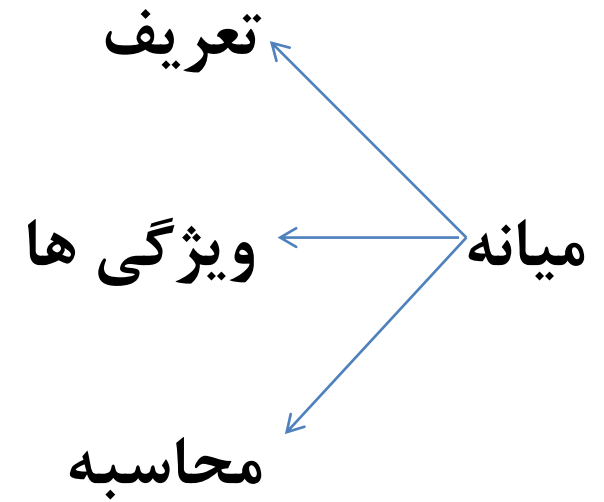
➤ نما سریع ترین و بی ثبات ترین شاخص مرکزی است.

➤ در محاسبه نما **حجم داده ها** اهمیت دارد.

➤ تنها شاخص مرکزی است زمانی که داده ها **در سطح اسمی** هستند.

۲ ۲ ۲ ۳ ۴ ۵





تعریف: شاخصی که توزیع را از وسط به دو نیم تقسیم می کند.

میانه

جایگاه عدد

بهترین شاخص مرکزی برای داده‌های **سطح رتبه‌ای**

شاخص مرکزی مناسب برای داده‌های دارای پرتی یا چولگی

مثال: ۱۱۵ ۱۱۴ ۱۱۳ ۱۱۲ ۱۱۱ ۱۱۰ ۱۰۹ ۲ ۱

مجموع قدر مطلق انحراف نمرات از میانه کوچک‌تر، مساوی
مجموع قدر مطلق انحراف داده‌ها از هر عدد دیگری درون
توزیع است.

$$\varepsilon |x - MD| \leq \varepsilon |x - C|$$



مجموع قدر مطلق انحراف نمرات از میانه کوچک تر، مساوی مجموع قدر مطلق انحراف داده ها از هر عدد دیگری درون توزیع است.

$$\varepsilon |x - MD| \leq \varepsilon |x - C|$$

$$MD=3$$

$$5 \quad 4 \quad 3 \quad 2 \quad 1$$

$$\varepsilon(x - MD) = 0 \quad \leftarrow \quad \bullet = (5-3) + (4-3) + (3-3) + (2-3) + (1-3)$$

$$\varepsilon|x - MD|=6 \quad \leftarrow \quad |5-3| + |4-3| + |3-3| + |2-3| + |1-3|$$

$$\varepsilon|x - C|=7 \quad |5-4| + |4-4| + |3-4| + |2-4| + |1-4|$$

$$\varepsilon |x - MD| \leq \varepsilon |x - C|$$

$\varepsilon|x - MD|$ همیشه کمینه است یعنی همیشه کمترین حالت ممکن است.



دکتری وزرات بهداشت ۹۵: اگر در مشاهدات آماری، یک داده پرت (Out Lier) وجود داشته باشد نتیجه بارز آن روی کدام شاخص آماری است؟

1. نما

2. میانه

3. میانگین

4. چارک اول و سوم



❖ وقتی میانه تکرار ندارد.

ابتدا داده‌ها را از کوچک به بزرگ مرتب کنیم.
اعداد را شمارش کرده و داده وسطی را انتخاب کنیم.

۱ ۲ ۳ ۴ ۵

اگر اعداد زوج باشند از دو عدد وسط میانگین می‌گیریم:

۱ ۲ ۳ ۴ ۵ ۶
۱ میانه می‌شود ۳/۵.



اگر میانه بیش از دو بار تکرار داشت از فرمول زیر استفاده می کنیم:

$$\text{تعداد میانه روی خط و قبل از خط} + \frac{\text{حد پایین میانه}}{\text{تعداد تکرار میانه}}$$

مثال: ۱۰ ۱۱ ۱۲ ۱۲ ۱۲ ۱۲ ۱۳ ۱۳ ۱۴

$$11/5 + \frac{3}{4}$$



تست بزنیہ ۱۱۰

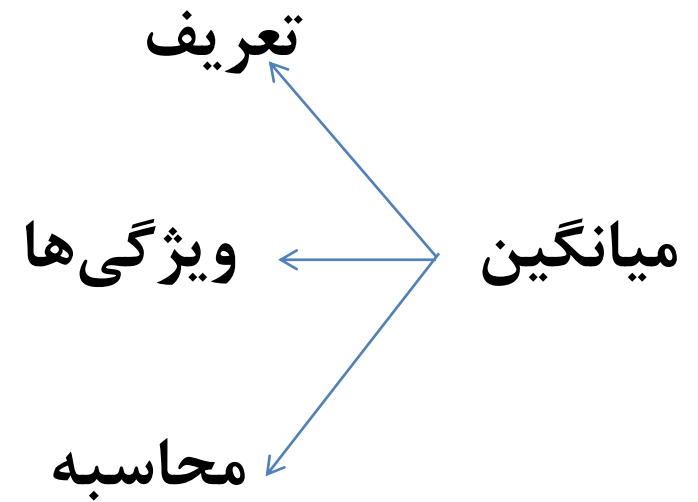


4 5 6 6 6 6 6 7 7 7

$$5/5 + \frac{3}{5}$$



میانگین



بهترین و دقیق ترین شاخص مرکزی

در محاسبه آن ارزش تمام داده ها در نظر گرفته می شود.

ما چند نوع میانگین داریم که $\bar{x}$ یک نوع میانگین است که به آن میانگین

حسابی گفته می شود.

$$\bar{x} = \frac{\sum x}{n} \longrightarrow \sum x = \bar{x} \cdot n$$

هر چیزی که بر n تقسیم شود نوعی میانگین است.



تست بزنیم ۱۱۱



میانگین چهار متغیر تصادفی $(x+2)$ ، $(7-x)$ ، $(3+2x)$ و $(x-4)$ برابر با $42/5$ است. مقدار x را محاسبه کنید.



میانگین چهار متغیر تصادفی $(x+2)$ ، $(7-x)$ ، $(3+2x)$ و $(x-4)$ برابر با $42/5$ است. مقدار x را محاسبه کنید.



- $(x+2) + (3+2x) + (7-x) + (x-4) = 42.5 \times n$
- $n = 4$
- $3x+8 = 42.5 \times 4$
- $3x+8 = 170$
- $3x = 162 \rightarrow x = 54$



اگر میانگین حسابی x_1, x_2, x_3, x_4, x_5 برابر با ۱۲ باشد و
میانگین حسابی x_1, x_2, x_3, x_4 برابر با ۱۰ باشد x_5 را محاسبه
کنید:



- $\bar{x} = \frac{\epsilon x}{n} \rightarrow 12 = \frac{\epsilon x}{5} \rightarrow \epsilon x = 60$

- $\bar{x} = \frac{\epsilon x}{n} \rightarrow 10 = \frac{\epsilon x}{4} \rightarrow \epsilon x = 40$

- $\rightarrow 60 - 40 = 20$



1. اگر تمام داده‌های یک توزیع را در یک عدد ثابت جمع، تفریق، ضرب یا تقسیم کنیم برای شاخص‌های مرکزی هم همان اتفاق خواهد افتاد.

اگر میانگین $(x_1, \dots, x_n)$ برابر با ۲۰ باشد میانگین $\left(\frac{x_1}{2} + 3 \dots \dots \frac{x_n}{2} + 3\right)$ را محاسبه کنید



دکتری ۹۸:

اگر میانه داده‌های $x_1, x_2, \dots, x_n$ مساوی ۴ باشد، آنگاه میانه داده‌های $2x_1 - 1, 2x_2 - 1, \dots, 2x_n - 1$ ، کدام است؟

۱. ۸

۲. ۹

۳. ۱۶

۴. ۷



مجموع انحراف نمرات از میانگین همیشه برابر با صفر است:

$$\varepsilon(x - \bar{x}) = 0$$

مثال: $\bar{x} = 3 \rightarrow 1 \quad 2 \quad 3 \quad 4 \quad 5$ همه داده‌ها را از $\bar{x}$ کم کرده و بعد آنها را با هم جمع می‌کنیم:

$$(1-3) + (2-3) + (3-3) + (4-3) + (5-3) = 0$$

• به میانگین مرکز ثقل داده‌ها نیز گفته می‌شود.



تست بزنیم ۱۱۴



دکتری ۹۹: دریک نمونه با ۱۵ مشاهده، $\sum(x - 15) = -1$ ، $\sum(x - 12) = 2$ ، $\sum(x - 14) = 0$ ، و $\sum(x - 18)^2 = 398$ است. میانگین این توزیع کدام است؟

۱. ۱۲

۲. ۱۴

۳. ۱۸

۴. ۱۶



فرمول محاسبه میانگین از نما و میانه

$$\bar{x} = \frac{3 MD - MO}{2}$$

تست ۱۱۵: در توزیعی از نمرات، میانگین و میانه به ترتیب ۲ و ۳ می‌باشد. نما را محاسبه نمایید.

- $\bar{x} = \frac{3 MD - MO}{2} \rightarrow$

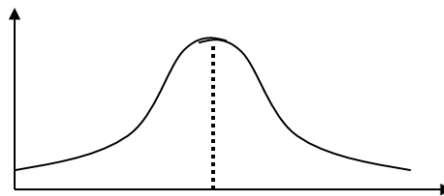
- $2 = \frac{3 \times 3 - MO}{2} \rightarrow$

- $2 = \frac{9 - MO}{2} \rightarrow$

- $4 = 9 - MO \rightarrow MO = 5$



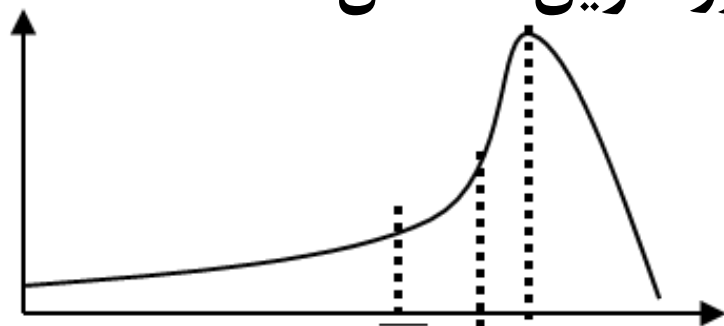
- در یک توزیع نرمال میانگین همیشه برابر است با میانه برابر است با نما پس وقتی هر سه شاخص مرکزی با هم برابر هستند توزیع نرمال است (چولگی برابر صفر).



$$\bar{x} = MD = MO$$

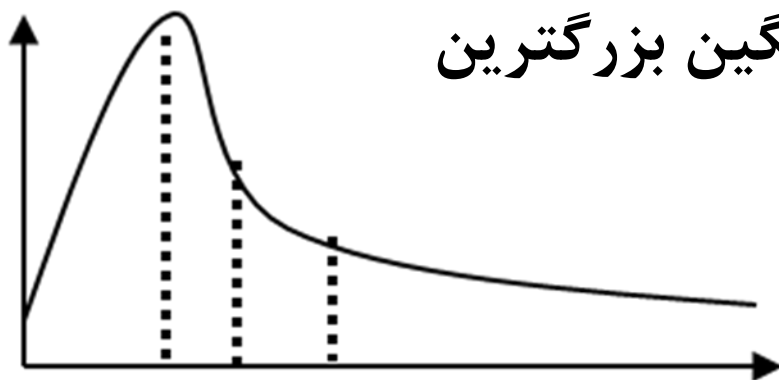


در توزیع دارای کجی منفی (چپ): نما بزرگترین شاخص



$$\bar{x} < MD < MO$$

در توزیع با کجی مثبت (راست): میانگین بزرگترین شاخص است.



M m $\bar{x}$

$$\bar{x} > MD > MO$$



شاخص‌های پراکندگی

چرا باید از شاخص‌های پراکندگی استفاده کنیم؟



شاخص‌های پراکندگی



- ◀ دامنه تغییرات
- ◀ انحراف چارکی
- ◀ انحراف متوسط
- ◀ واریانس
- ◀ انحراف استاندارد



دامنه تغییرات (R):

• سریع ترین و بی ثبات ترین شاخص پراکندگی

• فرمول دامنه تغییرات: $R = (X_h - X_l) + 1$

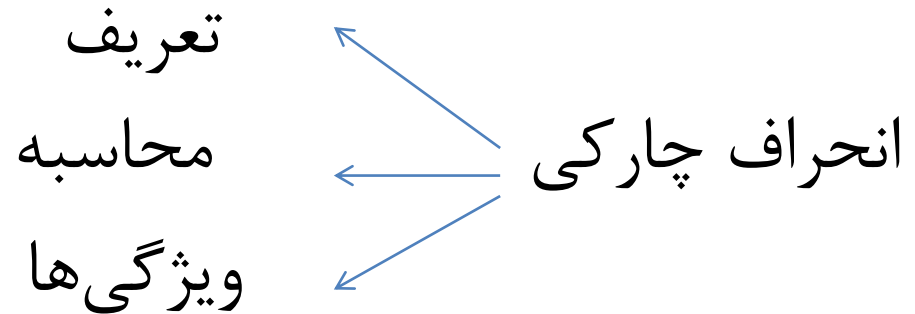


دکتری علوم شناختی ۹۹: دامنه تغییرات برای کدامیک از داده‌ها مناسب‌تر است؟

۱. اسمی
۲. رتبه‌ای
۳. فاصله‌ای
۴. نسبی



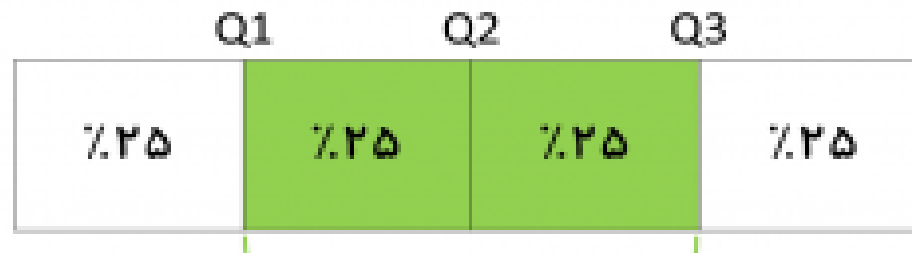
انحراف چارکی (Q)



تعریف انحراف چارکی

اگر یک توزیع را به چهار قسمت مساوی تقسیم کنیم هر قسمت آن یک چارک

نام دارد:



محاسبه انحراف چارکی (Q)

انحراف چارکی به ما می گوید به طور متوسط چارک ها از یکدیگر چقدر پراکندگی دارند در واقع پراکندگی بین چارک ها را بررسی می کند.

$$Q = \frac{Q_3 - Q_1}{2} \quad \text{فرمول انحراف چارکی:}$$

انحراف چارکی توزیع مقابل را محاسبه کنید:

۱ ۲ ۳ ۴ ۵ ۶ ۷

$$Q_1 = 2/5, Q_3 = 5/5 \rightarrow Q = \frac{Q_3 - Q_1}{2} \rightarrow Q = \frac{5/5 - 2/5}{2} = 1/5$$



ویژگی‌های انحراف چارکی

1. انحراف چارکی زمانی بهترین شاخص پراکندگی است که شاخص مرکزی مناسب برای توزیع میانه باشد.

وقتی توزیع دارای چولگی باشد شاخص مرکزی مناسب میانه و شاخص پراکندگی مناسب انحراف چارکی است.



تست بزنیم ۱۱۷



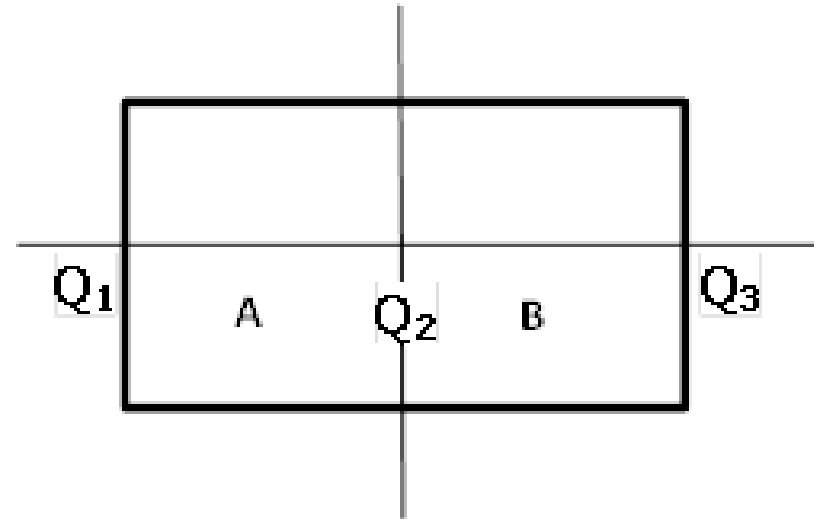
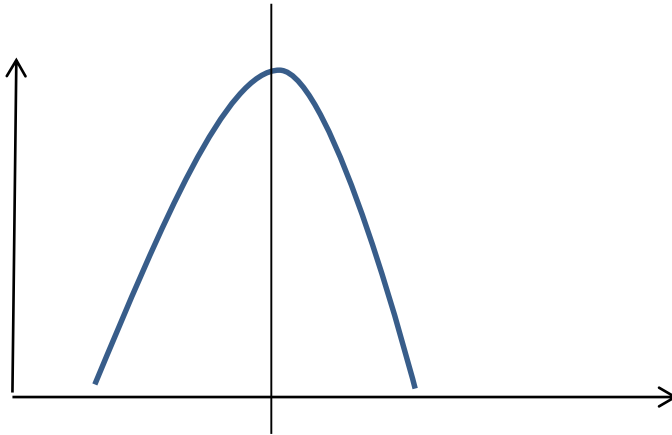
در جامعه‌ای که اکثریت افراد آن را سالمندان تشکیل می‌دهند، بهترین شاخص پراکندگی برای توزیع سنی کدام است؟

1. واریانس
2. دامنه تغییر
3. انحراف چار کی
4. انحراف استاندارد



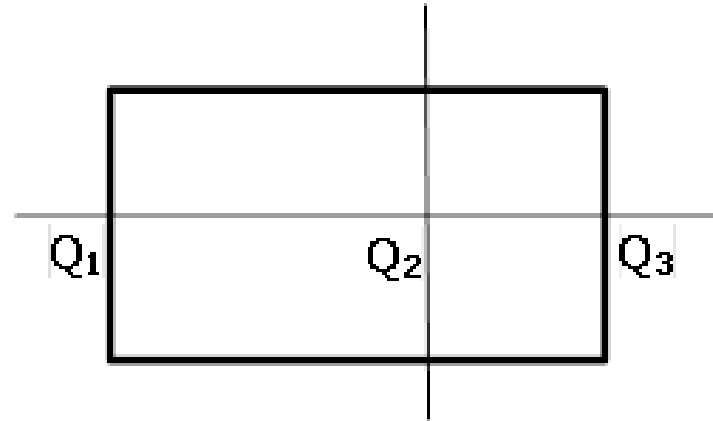
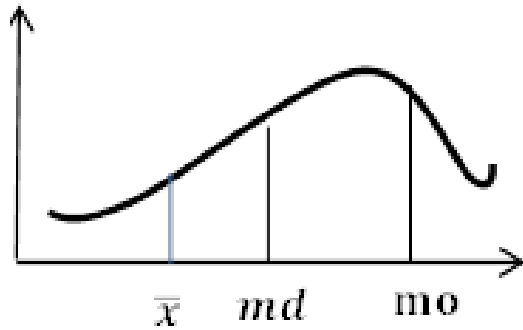
ویژگی‌های انحراف چارکی

۲. از چارک‌ها در ترسیم نمودار جعبه‌ای استفاده می‌شود. اگر A با B برابر باشد توزیع ما یک توزیع نرمال است به عبارت دیگر اگر $Q_2 - Q_1 = Q_3 - Q_2$ باشد توزیع نرمال است.



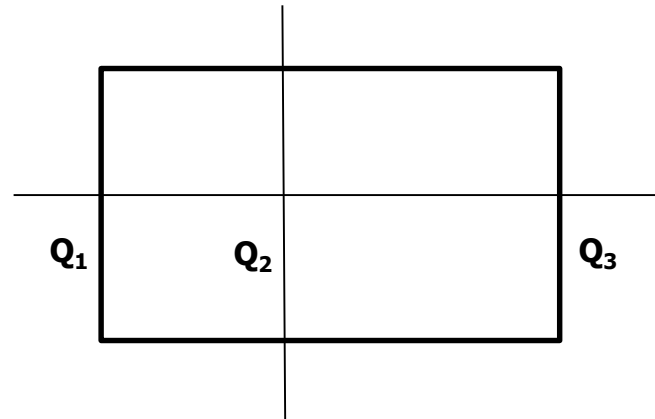
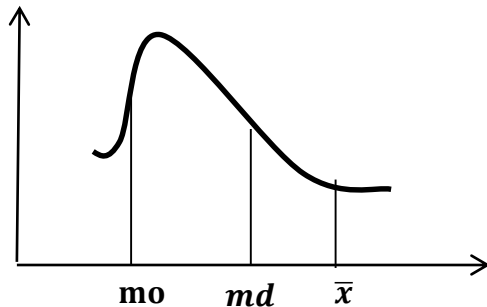
توزیع دارای کجی منفی:

$$Q_3 - Q_2 < Q_2 - Q_1$$



توزیع دارای کجی مثبت:

$$Q_3 - Q_2 > Q_2 - Q_1$$



تست بزنیم ۱۱۸



دکتری ۹۶

کدام یک از شاخص‌های آماری، از نمودار جعبه‌ای (Box Plot) قابل استخراج نیست؟

(۲) چارک سوم

(۴) کجی

(۱) انحراف متوسط

(۳) دامنه تغییرات



۳. با استفاده از چارک‌ها می‌توان داده‌های پرت را تشخیص داد.

برای پیدا کردن داده پرت از این قاعده پیروی کنید:

۱- دامنه چارکی را محاسبه کنید: $D = Q_3 - Q_1$

۲- دامنه چارکی را در عدد $1/5$ ضرب کنید.

۳- Q_3 را به اضافه D $1/5$ کنید و Q_1 را منهای $1/5$ کنید



تست بزنیم ۱۱۹



دکتری ۹۴

در صورتی که در یک توزیع، $Q_1 = 50$ ، $Q_2 = 65$ و $Q_3 = 70$ باشد کدام یک از اعداد زیر پرت (outlier) نیست؟

۱۸ (۲)

۱۰۷ (۴)

۱۵ (۱)

۹۸ (۳)



تست بزنیم ۱۲۰



ارشاد ۹۹: در نمودار جعبه‌ای کدام ملاک برای تشخیص داده‌های پرت در نظر گرفته می‌شود؟

$$Q_3 + 1/5(Q_3 - Q_1) < X < Q_1 - 1/5(Q_3 - Q_1)$$

$$Q_3 + 2(Q_3 - Q_1) < X < Q_1 - 1/5(Q_3 - Q_1)$$

$$Q_2 + 2S_x \leq X < Q_2 - 2S_x$$

$$Q_2 + 1/96S_x < X < Q_2 - 1/96S_x$$



تست بزنیم ۱۲۱



دکتری ۹۸:

اگر بزرگ‌ترین عدد توزیع فراوانی در عدد ۵ ضرب شود و اعداد تکراری نداشته باشیم، کدام شاخص پراکندگی توزیع تغییر نمی‌کند؟

(۲) دامنه تغییرات (R)

(۱) انحراف معیار (S)

(۴) دامنه میان چارکی (IQR)

(۳) واریانس (S^2)



شاخص‌های پراکندگی



- دامنه تغییرات
- انحراف چارکی
- انحراف متوسط
- واریانس
- انحراف استاندارد



انحراف متوسط

$$\begin{array}{ccccc} ۱ & ۲ & ۳ & ۴ & ۵ \\ (5-3)+(4-3)+(3-3)+(2-3)+(1-3) \end{array}$$

$$\varepsilon(x - \bar{X}) = 0$$

$$|(1 - 3)| + |(2 - 3)| + |(3 - 3)| + |(4 - 3)| + |(5 - 3)|$$

$$\varepsilon|x - \bar{X}|$$

مجموع قدر مطلق انحراف نمرات از میانگین

$$AD: \frac{\varepsilon|x - \bar{X}|}{n}$$

میانگین مجموع قدر مطلق انحراف نمرات از میانگین



شاخص‌های پراکندگی



- دامنه تغییرات
- انحراف چارکی
- انحراف متوسط
- واریانس
- انحراف استاندارد



تعریف و محاسبه واریانس



$$1 \quad 2 \quad 3 \quad 4 \quad 5 \rightarrow \bar{x} = 3$$

مجموع انحراف نمرات از میانگین برابر با صفر است که در انحراف متوسط بری حل این مشکل، قدر مطلق گرفتیم و در واریانس مجذور می‌کنیم:

$$(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 \rightarrow 10$$

مجموع مجذور انحراف نمرات از میانگین

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

واریانس: میانگین مجموع مجذور انحراف نمرات از میانگین



تست بزنیم ۱۲۲

کارشناسی ارشد ۹۷:

در یک نمونه n تایی، عبارت $(n-1)s_x^2$ با کدام مورد برابر است؟

$$\sum_{i=1}^n x_i^2 - \sum_{i=1}^n \bar{x} \quad (۳)$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 \quad (۴)$$

$$\sum_{i=1}^n (x_i - Md)^2 \quad (۱)$$

$$\sum_{i=1}^n x_i^2 - Md \quad (۲)$$



تست بزنیم ۱۲۳



در صورتی که واریانس توزیع نمرات یک آزمون مساوی صفر شود می توان نتیجه گرفت:

۱. نمره تمام آزمونی ها مساوی بوده است.
۲. تفاوت بین نمرات آزمودنی ها نزدیک به صفر است.
۳. اختلاف بین نمرات آزمودنی ها مساوی عدد ثابتی است.
۴. همه آزمودنی ها نمره صفر گرفته اند.



چند نکته مهم:



- ۱- واریانس ۳ عدد متوالی همیشه برابر با $0/66$ است.
- ۲- واریانس ۵ عدد متوالی همیشه برابر با ۲ است.
- ۳- واریانس ۷ عدد متوالی همیشه برابر با ۴ است.



چند نکته مهم:

در اغلب موارد واریانس گروه نمونه (شاخص واریانس) از پارامتر واریانس کوچکتر است. گاهی قصد برآورد واریانس جامعه آماری را از روی واریانس گروه نمونه داریم. شاخص واریانس یعنی واریانس گروه نمونه را داریم و از روی آن واریانس جامعه آماری برآورد می کنیم. وقتی قصد برآورد واریانس جامعه آماری را از روی واریانس گروه نمونه داشته باشیم فرمول واریانس یک تغییر جزئی خواهد داشت، در مخرج n را منهای یک می کنیم تا مخرج کوچک شده و خارج قسمت (واریانس) بزرگتر شود تا کمی به واریانس جامعه آماری نزدیکتر شود:

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



تست بزنیم ۱۲۴



کارشناسی ارشد علوم تربیتی ۱۴۰۰: برآورد اریب‌دار (سوگیر) واریانس
نمونه‌ای به حجم ۵ نفر، ۱۰ می‌باشد، برآورد بدون اریب (بدون سوگیری)
واریانس متناظر آن چند است؟

۱. ۵/۲

۲. ۵/۷

۳. ۱۲/۵

۴. ۱۵



چند نکته مهم



➤ جمع و تفریق داده‌ها هیچ تأثیری بر شاخص‌های پراکندگی ندارد اما ضرب و تقسیم می‌تواند بر این شاخص‌ها تأثیر بگذارد. ضرب و تقسیم داده‌ها در واریانس (یک عدد ثابت) باعث می‌شود برای محاسبه واریانس جدید، واریانس قبلی را در مجذور آن عدد ضرب یا تقسیم کنیم:

$$S^2_{\text{جدید}} = S^2_{\text{قدیم}} \times (\text{عدد ثابت})^2$$



تست بزنیم ۱۲۵



میانگین و انحراف استاندارد مجموعه‌ای از داده‌ها به ترتیب برابر با ۱۰ و ۵ است. میانگین و واریانس توزیع زیر را محاسبه کنید:

$$(2x_1 - 3 \dots 2x_n - 3)$$



چند نکته مهم



- گاهی باید واریانس را با داشتن مجموع نمرات و مجموع مجذور نمرات محاسبه کنیم. فرمول آن به شرح زیر است:

$$S^2 = \frac{\text{مجموع مجذور نمرات} - \frac{\text{مجموع نمرات}^2}{n}}{n}$$



تست بزنیم ۱۲۶



در صورتی که مجموع نمرات و مجموع مجذور نمرات برای ۱۰ نفر به ترتیب ۱۴۰ و ۲۰۰۰ باشد واریانس را محاسبه کنید:



انحراف استاندارد (انحراف معیار)



- انحراف استاندارد؛ جذور واریانس

$$s = \sqrt{S^2}$$

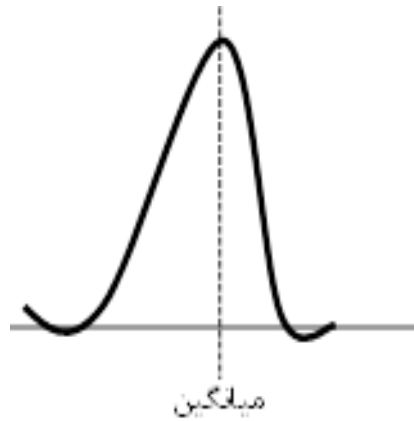
- دقیق‌ترین و بهترین شاخص پراکندگی



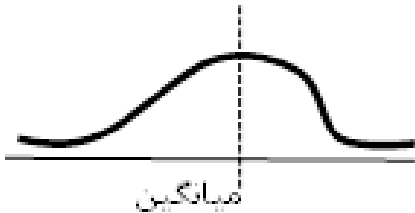
نکات مهم انحراف معیار



در توزیع‌ها انحراف معیار شاخص کشیدگی در نظر گرفته می‌شود. در یک توزیع نرمال، انحراف معیار اندازه متوسطی دارد.



هر چقدر انحراف استاندارد کوچک‌تر باشد یعنی پراکندگی نمرات از میانگین کمتر است و نمرات دور میانگین جمع می‌شود در نتیجه توزیع بلند و کشیده می‌شود.



هر چقدر انحراف معیار بزرگ باشد یعنی پراکندگی نمرات از میانگین زیاد است بنابراین توزیع پهن و کوتاه خواهد شد.



وقتی برای توزیع‌های طبقه‌بندی شده انحراف استاندارد محاسبه می‌کنیم چون داده‌ها هویت فردی خود را از دست داده‌اند انحراف استاندارد نیز توأم با خطا خواهد بود پس لازم است انحراف استاندارد را در مواقعی که توزیع فراوانی‌ها طبقه‌بندی شده است اصلاح نماییم تا خطا کمتر شود. به فرمول اصلاح استاندارد برای توزیع طبقه‌بندی تصحیح شپرد گفته می‌شود:

$$s_c = \sqrt{s^2 - \frac{i^2}{12}}$$

فرمول تصحیح شود



تست بزنیم ۱۲۷



دکتری ۹۷

اگر تمام مقادیر توزیع متغیر X در $\frac{1}{S_x}$ ضرب شوند، واریانس و انحراف معیار نمره‌های جدید برابر چند است؟

$$\frac{S_x}{S_x^2} - S_x^2 \quad (۴)$$

$$\frac{1}{2} - 1 \quad (۳)$$

$$1 - 1 \quad (۲)$$

$$S_x - S_x^2 \quad (۱)$$



تست بزنیم ۱۲۸



کارشناسی ارشد ۹۴

اگر واریانس نمره ریاضی ۶۴ باشد و از هر نفر ۲ نمره کم شود، در این صورت انحراف معیار کدام است؟

۱۶ (۱)	۸ (۲)	۶۴ (۳)	۱۲۸ (۴)
--------	-------	--------	---------



گشتاورهای میانگین



$$\text{گشتاور اول} = \frac{\varepsilon (x - \bar{x})^1}{n}$$

$$\text{گشتاور دوم} = \frac{\varepsilon (x - \bar{x})^2}{n}$$

$$\text{گشتاور سوم} = \frac{\varepsilon (x - \bar{x})^3}{n}$$

$$\text{گشتاور دوم} = \frac{\varepsilon (x - \bar{x})^4}{n}$$



شاخص ضریب تغییرات:

انحراف استاندارد تقسیم بر میانگین.

تست بز نیم ۱۲۹

دکتری ۹۴:

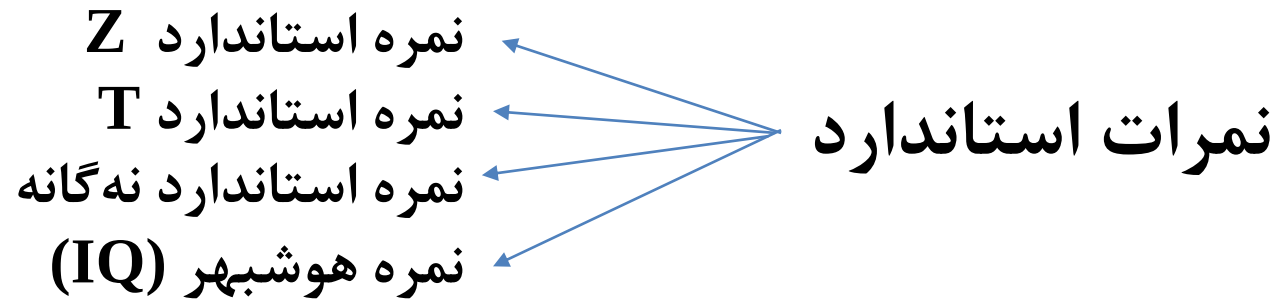
در صورتی که به تمام مقادیر یک توزیع عدد ثابت c ($c > 0$) اضافه شود، کدام شاخص کاهش می یابد؟

(۱) انحراف استاندارد (S_x)	(۲) ضریب تغییرات (CV)
(۳) میانگین ($\bar{x}$)	(۴) دامنه میان چارکی (IQR)



شاخص‌هایی برای مقایسه گروهی نمرات

رتبه درصدی



رتبه درصدی

فردی با نمره خام ۴۵ رتبه ۵۶ درصدی دارد. این یعنی این فرد با نمره خام ۴۵ از ۵۶ درصد افراد عملکرد بهتری داشته یا برابر بوده است.

پس رتبه درصدی یعنی درصد افرادی که فردی با یک نمره خام مشخص از آنها بهتر عمل کرده یا برابر بوده است.

فرد A رتبه درصدی ۸۰ و فرد B رتبه درصدی ۴۰ دارد. نمی‌توان گفت رتبه درصدی فرد A دو برابر فرد B است چون سطح رتبه درصدی در سطح اندازه-گیری رتبه‌ای مطرح می‌شود و در مقیاس رتبه‌ای فاصله بین رتبه‌ها نه مشخص است نه ثابت.

پس فقط می‌توانیم بگوییم فرد A بهتر از فرد B عمل کرده است.



محاسبه رتبه درصدی



$$PR = \frac{\frac{f}{2} + cf}{n} \times 100 \quad \text{فرمول رتبه درصدی}$$

با توجه به جدول زیر رتبه درصدی عدد ۵۰ را محاسبه کنید:

x	۴۸	۴۹	۵۰	۵۱	۵۲
f	۳	۳	۲	۶	۶



نمرات استاندارد



- نمره استاندارد Z
- نمره استاندارد T
- نمره استاندارد نه گانه
- نمره هوشبهر (IQ)

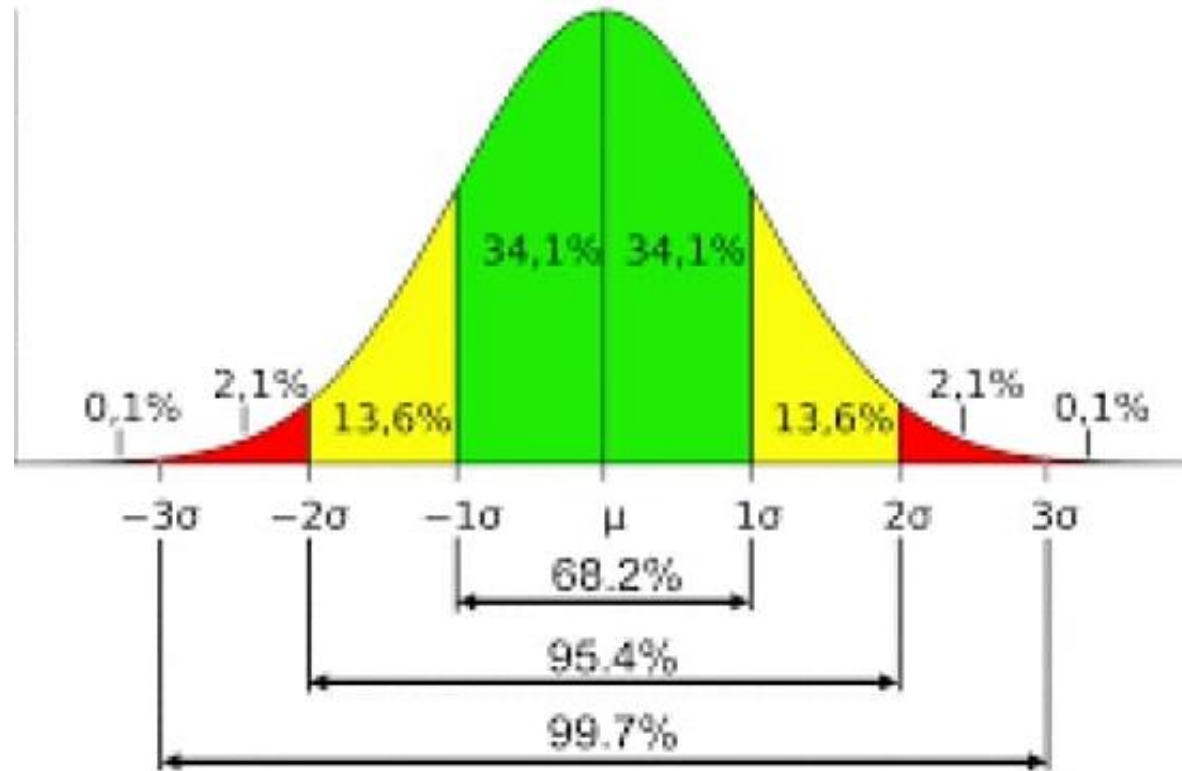


نمره استاندارد Z

$$Z = \frac{x - \bar{x}}{s}$$



نمره‌های استاندارد Z روی منحنی نرمال نمرات Z قابل تفسیر هستند.



۱- زنگوله‌ای شکل و متقارن است.

۲- سطح زیر منحنی نرمال همیشه ثابت است.

۳- میانگین منحنی نمرات استاندارد Z برابر با صفر است. هر انحراف استاندارد، ۱ نمره است.

۴- در میانگین منحنی Z دو نقطه عطف داریم که بین -1 تا $+1$ است.

۵- منحنی نرمال Z هیچ گاه به محور X ‌ها نمی‌چسبد یعنی از $Z = -\infty$ شروع می‌شود تا $Z = +\infty$ ؛ اما در محاسبات بالاتر از $z = +3$ و پایین‌تر از $z = -3$ را صفر در نظر می‌گیرند.



انتقال نمرات در منحنی نرمال یک انتقال خطی است.

یعنی برای هر نمره خام، یک Z محاسبه می‌شود. اگر شکل توزیع نمرات خام نرمال باشد شکل توزیع نمرات Z هم نرمال است و اگر شکل توزیع نمرات کجی داشته باشد شکل توزیع نمرات Z هم کجی دارد بنابراین شکل توزیع نمرات Z دقیقاً تابع شکل توزیع نمرات خام است.



یک نکته مهم



هر شکل توزیع از نمرات دو ویژگی کاملاً برجسته دارد:

۱- کشیدگی. ۲- کجی یا چولگی.

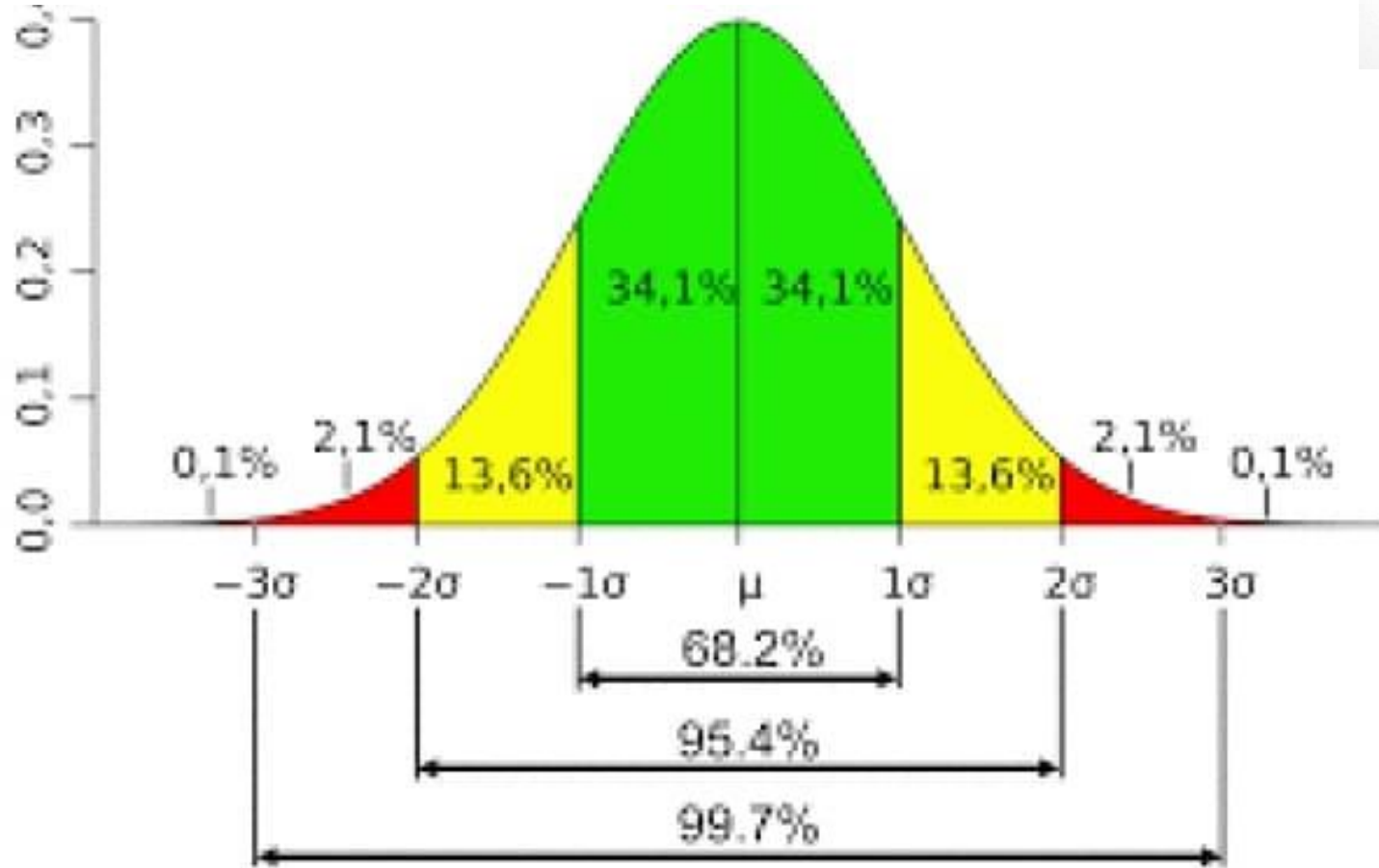
شاخص کشیدگی انحراف معیار است. هر چه انحراف معیار کمتر باشد توزیع بلندتر و کشیده تر است.

فرمول محاسبه کجی

$$SK = \frac{\bar{x} - MO}{s}$$



نحوه محاسبه رتبه درصدی از روی نمره Z



تست بزنیم ۱۳۰



رتبه درصدی فردی با نمره خام ۷۵ را در شرایطی محاسبه کنید که میانگین و انحراف استاندارد نمرات در یک گروه ۲۵ نفره به ترتیب ۶۵ و ۵ باشد.

$$Z = \frac{x - \bar{x}}{s} \rightarrow Z = \frac{75 - 65}{5} = 2$$

طبق جدول منحنی نرمال Z ، نمره $Z = 2$ برابر است با ۹۷/۶۲



تست بزنیم ۱۳۱



فاصله بین کدام یک از نمرات استاندارد زیر بیشتر است؟

1. $Z = -1$ تا $Z = +1$

2. $Z = 0$ تا $Z = +1$

3. $Z = 0$ تا $Z = +2$

4. $Z = +2$ تا $Z = +3$



نمره استاندارد T

$$T = 10 Z + 50$$



تست بنزیم ۱۳۲



فردی با نمره خام ۶۸ از ۸۴/۱۳ افراد گروه عملکرد بهتری داشته
است. نمره T را محاسبه کنید.



نمره استاندارد نه گانه

نمره استاندارد نه گانه : $2Z + 5$

نمره نه گاه فردی با رتبه درصدی $۸۴/۱۳$ چند است؟



نمره هوشبهر (IQ)



$$IQ = 15 Z + 100$$

