

فصل  
چهارم

## کار با داده‌های آماری

## درس ۱ گردآوری داده‌ها

## تعاریف:

## جامعه و نمونه

**جامعه آماری:** مجموعه‌ای از افراد یا اشیاء است که درباره اعضای آن می‌خواهیم موضوع یا موضوعاتی را مطالعه کنیم. یا به عبارت دیگر جامعه مجموعه‌ای از اشیاء است که در یک یا چند صفت مشترک هستند، و برحسب موضوع مورد مطالعه تعیین می‌شود.

مثال ۱: بانوان ورزشکار استان اردبیل  
موضوع مورد مطالعه: مدت زمان ورزش بانوان در هفته

مثال ۲: محصولات کشاورزی استان اردبیل  
موضوع مورد مطالعه: انواع محصولات

**اندازه جامعه:** تعداد اعضای جامعه را اندازه جامعه می‌گویند.

**نمونه آماری:** زیر مجموعه‌ای از جامعه آماری است که بتواند بیانگر جامعه باشد.

**عمل نمونه‌گیری:** مهم‌ترین بخش آمار را عمل نمونه‌گیری می‌گویند. برای آن که نمونه بتواند به درستی نمایانگر خصوصیت تمام جامعه باشد، باید به اندازه کافی بزرگ باشد و هم چنین اعضای نمونه باید تصادفی انتخاب شوند.

**اندازه نمونه:** تعداد اعضای نمونه را اندازه نمونه می‌گویند.

**نمونه تصادفی:** زیر مجموعه‌ای از جامعه آماری است که اعضای آن به طور تصادفی انتخاب می‌شوند.

**روش انتخاب نمونه تصادفی ( منظور از اعضای نمونه باید تصادفی انتخاب شوند )**

۱- هر یک از اعضا امکان حضور داشته باشند.

۲- قبل از انتخاب نمونه نتوانیم، با اطمینان بیشتر درباره حضور یا عدم حضور عده‌ای در نمونه قضاوت کنیم.

اعداد تصادفی: نمونه گیری تصادفی ساده به روش های مختلف انجام می گیرد ، یکی از آن ها استفاده از ماشین حساب است .

شیوه تولید اعداد تصادفی در ماشین حساب: پس از روشن کردن ماشین حساب ، مراحل زیر را طی کنید.

- ۱- دکمه inv (در برخی shift در برخی دیگر ۲ndf) را فشار دهید، (ماشین حسابهایی که این دکمه را ندارند، از مرحله دوم شروع کنید)
- ۲- دکمه RAN# را فشار دهید. پس از انجام این عمل ، عددی که همواره غیر منفی و کوچکتر از یک می باشد به دست می آید، این عدد عدد تصادفی می باشد.

نمونه استفاده از عدد تصادفی:

برای این منظور اعداد تولید شده را در اندازه جامعه ضرب می کنیم و اگر عدد حاصل قسمت اعشاری داشت، قسمت اعشاری را حذف کرده و یک واحد به آن اضافه می کنیم مثلاً اگر جامعه ما ۱۰۰ عضو داشته باشد و اعداد تصادفی تولید شده به وسیله ماشین حساب به صورت زیر باشد، در این صورت خواهیم داشت:

عدد تصادفی	۰/۰۳۴	۰/۰۲۹	۰/۹۵۵	۰/۷۰۲
۱۰۰× عدد تصادفی	۳/۴	۲/۹	۹۵/۵	۷۰/۲
شماره انتخاب شده	۴	۳	۹۶	۷۱

داده: نتایج حاصل از اندازه گیری یا بررسی نمونه را داده می گویند.

روش های جمع آوری داده ها:

- ۱- از طریق پرسش ( پرسش نامه کتبی، پرسش شفاهی از اشخاص (مصاحبه)) مثلاً: " زمستان فصل مورد علاقه بیشتر مردم برای مسافرت است " ، " پر اید ماشین محبوب ایرانیان برای استفاده است " بهترین روش جمع آوری داده ها استفاده از روش پرسش نامه است.
- ۲- از طریق مشاهده و ثبت وقایع: مثلاً: " افرادی که در ساعت خاص وارد یک ساختمان می شوند "، " بررسی استقبال مردم از نمایشگاه "
- ۳- از طریق آزمایش: مثلاً: " اثر نور خورشید بر رشد گیاهان "، " تأثیر موسیقی بر هوش دانش آموزان "
- ۴- استفاده از داده های از پیش تهیه شده: " بررسی تخلفات رانندگی یک راننده "، " ارتباط بین نمره درس ریاضی و معدل کل دانش آموز "

طراحی پرسش نامه: قبل از آن که هر گونه سوالی نوشته شود،

- ۱- محتوای پرسشنامه باید سازماندهی شود. ۲- هدف بررسی باید در نظر گرفته شود (چه می خواهید بدانید و چرا؟)
- ۳- فهرستی از عناوینی که باید راجه به آن ها اطلاعات جمع آوری کنید، تهیه کنید.
- ۴- از جمع آوری داده ها و اطلاعات اضافی که مورد نیاز نمی باشد، خودداری کنید.
- ۵- از سوالات ساده و واضح استفاده کنید. (استفاده از عباراتی چون بزرگ، کوچک که معیار مشخصی ندارند خوداری کنید تا از سوالات چند برداشت نشود)
- ۶- سعی کنید از سوالاتی که پاسخ آن ها تا حد امکان یک کلمه ای یا اعداد هستند استفاده شود.
- ۷- از سوالات هدایت کننده استفاده نکنید.

۸- سعی کنید از سوالات با پاسخ های چند گزینه ای استفاده کنید. در پایان پرسشنامه از پاسخ دهندگان تشکر کنید.  
**بررسی آماری:** روش جمع آوری داده ها بدون دخالت در تغییر وضع آن ها بررسی آماری نام دارد.

آمارگیری: گردآوری داده ها به یکی از روش های ممکن  
 آمارگیر: کسی که آمارگیری را انجام می دهد.

**پارامتر جامعه:** یک مشخصه عددی است که توصیف کننده جنبه ای خاص از جامعه است و در صورتی که داده های کل جامعه در اختیار باشند قابل محاسبه است. مثلاً اگر داده های مربوط به تک تک کوهنوردان را داشته باشیم، یعنی به داده های جامعه دسترسی داریم. نسبت مردان در کل جامعه کوهنوردان، معرف یک پارامتر است.

اگر داده های بعضی از کوهنوردان را داشته باشیم؛ یعنی داده های نمونه را در اختیار داریم. نسبت مردان کوهنورد به این داده های نمونه ای را، آماره (مقدار آماره) گویند. آماره ها از یک نمونه به نمونه دیگر تغییر می کنند؛ این در حالی است که پارامترهای جامعه همیشه ثابت اند، چرا؟  
 در بسیاری از موارد، آمارگیری از کل جامعه امکان پذیر نیست. بنابراین علی رغم اینکه پارامتر دارای مقدار ثابتی است، این مقدار مجهول است و به همین دلیل از آماره ها برای تخمین پارامترها استفاده می کنند.

**آماره نمونه:** مشخصه ای عددی که توصیف کننده جنبه ای خاص از نمونه است و از داده های نمونه به دست می آید.

**مثال:** اداره کشاورزی استان خوزستان در حال ارزیابی هندوانه های آماده برداشت است. در این بررسی، هندوانه ها همان واحدهای آماری هستند. اگر پژوهشگران وزن هندوانه ها را مورد بررسی قرار دهند، متغیر، «وزن» آنهاست. وزن یک متغیر کمی است، زیرا با مقادیر عددی ارائه می شود. اگر وزن تک تک هندوانه های این زمین بررسی شود، سرشماری از جامعه انجام داده ایم (که امکان پذیر نیست). متوسط وزن تمامی هندوانه های قابل برداشت در این زمین، «پارامتر» است.

حال فرض کنیم پژوهشگران تصمیم دارند بر اساس معیار «مزه» هندوانه‌ها را مورد بررسی قرار دهند. در این حالت مزه هندوانه‌ها را می‌توان به سه دسته تقسیم کرد: بد، قابل قبول و خوب. حال که می‌خواهیم مزه هندوانه‌ها را امتحان کنیم، مطالعه به بخشی از کل هندوانه‌ها محدود می‌شود. در اینجا متغیر «مزه» متغیری کیفی است. از آنجا که نمی‌توانیم تمام هندوانه‌ها را مزه مزه کنیم، تنها بخشی از هندوانه‌ها مورد مطالعه قرار می‌گیرند؛ پس باید «نمونه» بگیریم. نسبت هندوانه‌های با مزه «خوب» در نمونه، یک «آماره» است.

### متغیرهای تصادفی

**متغیر تصادفی:** همان موضوعی است که روی جامعه یا نمونه مورد مطالعه قرار می‌گیرد. و این موضوع از فردی به فرد دیگر تغییر می‌کند. مثال: فرض کنید می‌خواهیم بدانیم دانش آموزان یک کلاس چه قدر پول همراه دارند، برای این که بتوانیم مثال را تعقیب کنیم، یک نمونه تصادفی ۸ تایی از این کلاس انتخاب می‌کنیم، که عبارتند از:

$A = \{\text{محمد، امین، شهاب، موسی، جمشید، تقی، اکبر، رحیم}\}$  اگر پول‌های همراه این افراد را معلوم کنیم، مثلاً مجموعه بالا به مجموعه زیر تبدیل می‌شود.  $B = \{500, 1200, 1300, 1500, 2000, 2500, 600, 1500\}$  آنچه که برای ما مهم است مجموعه  $B$  می‌باشد. متغیر تصادفی مجموعه هشت فرد را تبدیل به مجموعه‌ای از اعداد کرد. یعنی هم چنانچه می‌بینیم فرد برای ما مهم نیست بلکه آن چه مهم است اطلاعاتی است که در ارتباط با متغیر تصادفی مورد مطالعه این فرد در اختیار ما قرار داده شده است.

**انواع متغیرهای تصادفی:** به دو دسته، تقسیم می‌شوند، الف) متغیر تصادفی کمی (ب) متغیر تصادفی کیفی

**متغیر تصادفی کمی:** متغیرهایی هستند که قابل اندازه‌گیری باشند، (یعنی می‌توانیم به آن‌ها عدد نسبت دهیم) مانند قد، وزن، سن و...  
**متغیر تصادفی کیفی:** متغیرهایی هستند که قابل اندازه‌گیری نباشند، (یعنی نمی‌توانیم به آن‌ها عدد نسبت دهیم) مانند گروه خونی، جنسیت افراد، کیفیت کالای تولید شده....

**انواع متغیر تصادفی کمی:** به دو دسته، تقسیم می‌شوند: الف) متغیر کمی پیوسته (ب) متغیر کمی گسسته

**متغیر کمی پیوسته:** متغیرهای کمی هستند که اگر دو مقدار  $a, b$  را بتواند اختیار کنند هر مقدار بین آن‌ها را نیز بتوانند اختیار کنند. مانند قد دانش آموزان یک کلاس، اگر قد یک دانش آموز  $150\text{cm}$  و قد دانش آموز دیگر  $151\text{cm}$  باشد، وزن دانش آموزان دیگر می‌تواند بین  $150\text{cm}$ ،  $151\text{cm}$  باشد.

**متغیر کمی گسسته:** متغیرهای کمی هستند که پیوسته نباشد. مانند تعداد بیماران مراجعه کننده به یک بیمارستان.

**نکته:** متغیرهای تصادفی پیوسته عموماً از راه اندازه‌گیری (با ابزار) به دست می‌آیند، مثل: قد، وزن، درجه حرارت. متغیرهای تصادفی گسسته عموماً از راه شمارش به دست می‌آیند. تعداد غایبین کلاس، تعداد دندانهای فاسد یک کودک.

کانال تلگرام: @mathvalizadeh

انواع متغیر تصادفی کیفی: به دو دسته، تقسیم می شوند: الف) متغیر کیفی ترتیبی (ب) متغیر کیفی اسمی

متغیر کیفی ترتیبی: متغیرهای کیفی هستند که در آن ها نوعی ترتیب طبیعی وجود دارد، مثل: مراحل زندگی انسان که شامل نوزادی، کودکی، نوجوانی، جوانی، بزرگسالی و پیری می باشد. یا حروف الفبای فارسی که به ترتیب شامل الف، ب، ... می باشد.

متغیرهای کیفی اسمی: متغیرهای کیفی هستند که ترتیبی نباشند. مثل: گروه خونی افراد، رنگ ماشین، جنسیت افراد، نوع کشت.

## سوال:

نوع متغیرهای تصادفی را در هر یک از عبارات زیر مشخص نمایید.

نوع متغیر	متغیر تصادفی	نوع متغیر	متغیر تصادفی
	تعداد نامه های یک صندوق		متغیری که معمولاً از نوع تعداد می باشد.
	وزن نامه های یک صندوق		غذای مورد علاقه
	زمان انتظار بیمار در مطب		نمره درس آمار
	تعداد بیماران مراجعه کننده		تعداد سوالات امتحانی درس آمار
	جنسیت افراد یک شهر		مدت زمان معاینه یک فرد
	میزان تحصیلات ( دیپلم، کارشناسی، ...)		مراحل تحصیل
	وضعیت مسکن (مستأجر، مالک)		متغیری که هر نوع عدد اعشاری یا صحیح را می توان به آن ها اختصاص داد.
	میزان مالیات پرداختی سالانه		نوع کشت ( برنج، گندم و..)
	میزان آلودگی هوا		تعداد مکالمات تلفنی
	مقاومت یک ترانزیستور		طول مکالمات تلفنی
	تعداد شکایات رسیده به یک پاسگاه		گنجایش آب یک تانکر
	وضعیت تأهل		روزهای هفته

- می خواهیم مدت زمانی که دانش آموزان کلاس شما صرف مطالعه کتاب های غیر درسی می کنند بررسی کنیم.
- الف) در این بررسی جامعه را مشخص نمایید.
- ب) برای این بررسی یک نمونه مشخص نمایید.
- ج) از چه روشی برای جمع آوری اطلاعات استفاده می کنید. (د) اندازه این جامعه چه قدر است؟
- ه) اندازه نمونه مورد بررسی چه قدر است؟ آیا این تعداد برای بررسی مورد نظر مناسب است؟
- و) متغیر مورد مطالعه در این مسأله چیست؟ (ز) متغیر مورد مطالعه از چه نوعی است؟

## کار در کلاس

۱. چه راه دیگری برای آمارگیری طول قد دانش آموزان یک مدرسه پیشنهاد می کنید؟

۲. فرض کنید زمان لازم را برای گردآوری تمامی داده‌های دانش آموزان در اختیار نداشته باشید. اگر بخواهیم نمونه‌ای را انتخاب و آمارگیری کنیم، چه راهی پیشنهاد می کنید که نمونه به صورت تصادفی انتخاب شود؟

## تمرین

کدام روش جمع‌آوری داده‌ها برای موارد زیر مناسب است؟ یک دلیل برای انتخاب خود ذکر کنید.

۱. میزان رضایت مشتریان بانک از نحوه برخورد و رسیدگی به درخواست‌های آنها.

۲. سن همه دانش آموزان مدرسه بر حسب ماه در پایه دهم.

۳. تعداد سرنشینان خودروهای سواری در یکی از محورهای خروجی شهر.

## کار در کلاس

الف) کدام روش برای جمع‌آوری هر یک از داده‌ها مناسب است؟

۱. تعداد قلم‌های هر دانش آموز در یک کلاس.

۲. ساعات خواب دانش آموزان کلاس درس شما در شب گذشته.

۳. طول قد دانش آموزان یک کلاس.

ب) می‌خواهیم طول قد دانش‌آموزان یک کلاس یا مدرسه را به یکی از سه روش زیر آمارگیری کنیم. هر یک از این روش‌ها محدودیت‌هایی دارند. چگونه می‌توان این محدودیت‌ها را از بین ببرد؟

پرسش‌نامه: اگر تعداد واحدهای نمونه زیاد باشد، این روش زمان‌بر است.  
مشاهده: اگر به‌دقت زیادی نیاز داریم، مناسب نیست.  
دادگان‌ها: همیشه اطلاعات ثبتي را در اختیار آمارگیر قرار نمی‌دهند.

### کار در کلاس

یک شبکه تلویزیونی می‌خواهد نسبت دارندگان تلویزیون در شیراز را، که برنامه جدید این شبکه را حداقل یک‌بار در هفته تماشا می‌کنند، بداند. بدین منظور یک گروه ۱۰۰۰ نفری از دارندگان تلویزیون را در این شهر بررسی می‌کند.

الف) داده‌ها و متغیرهایی را که بررسی می‌شوند، مشخص کنید.

جواب: داده‌ها اطلاعات گروه ۱۰۰۰ نفری دارندگان تلویزیون در شیرازند، و متغیر، تماشای تلویزیون است که پاسخ آن «تماشا می‌کند» یا «تماشا نمی‌کند» افراد مورد بررسی است.

ب) آیا این داده‌ها یک نمونه‌اند؟ جامعه آماری کدام است؟

پ) متغیر کمی است یا کیفی؟

ت) چند متغیر کمی را که ممکن است در اینجا جالب باشد، مشخص کنید.

جواب: سن، درآمد، ....

ث) نسبت افرادی در نمونه که برنامه جدید را تماشا می‌کنند، آماره است یا پارامتر؟ (تعداد اعضای مورد نظر تقسیم بر تعداد کل اعضای یک مجموعه را نسبت می‌گوییم.)

## کار در کلاس

نوع متغیر داده‌های زیر را مشخص کنید:

- الف) محسن، محمود، محمد و میثم همگی اسامی مذکر هستند.
- ب) در یک دبیرستان ۳۱۹ دانش‌آموز فارغ‌التحصیل وجود دارد. احمد رتبه بیست و پنجم، رضا رتبه نوزدهم، صادق رتبه دهم و جواد رتبه چهارم را کسب کرده است و می‌دانیم که رتبه یک، بالاترین است.
- پ) دمای بدن ماهی‌های قزل‌آلای رودخانه هراز (برحسب درجه سلسیوس).
- ت) طول ماهی‌های قزل‌آلا در رودخانه هراز.

## تمرین

- داده‌های زیر مربوط به یک نماینده مجلس است. در هر یک از سؤالات زیر نوع داده‌ها را مشخص کنید.
- الف) نام نماینده حسین ایرانی است.
- ب) این نماینده ۵۸ سال سن دارد.
- پ) سال‌هایی که این نماینده در مجلس انتخاب شده است، ۱۳۸۶، ۱۳۹۰ و ۱۳۹۴ است.
- ت) مجموع حقوق این نماینده در سال گذشته ۶۰۰,۰۰۰,۰۰۰ ریال بوده است.
- ث) این نماینده در حال بررسی لایحه پیشنهادی حفاظت از منابع آبی کشور است. گزینه‌های موردنظر: حمایت کامل، حمایت، بی‌طرف، مخالف و کاملاً مخالف است.
- ج) وضعیت تأهل این نماینده: متأهل
- چ) می‌گویند این نماینده در رأی‌گیری لایحه مرتبط با آموزش عمومی، هفتمین نفری است که از آن حمایت کرده است.



## تمرین

۱. فرق بین داده و متغیر چیست؟

۲. داده‌های در سطح اسمی، کمی هستند یا کیفی؟

۳. فرق بین آماره و پارامتر چیست؟

۴. در یک جامعه آماری، آیا ممکن است که یک پارامتر تغییر کند؟ اگر سه نمونه با اندازه یکسان از یک جامعه داشته باشیم، آیا می‌توان سه مقدار متفاوت از یک آماره به دست آورد؟

۵. در یک مطالعه از ۱۲۶۱ مشتری غذاخوری‌های گیاه‌خوار، سؤال شده است که برای کدام وعده غذایی

(ناهار یا شام) غذا سفارش داده‌اند؟

الف) متغیر را مشخص کنید.

ب) این متغیر کمی است یا کیفی؟

پ) جامعه آماری در اینجا چیست؟

۶. موضوعات زیر مرتبط است. متغیرهای آنها را در چهار مقیاس: اسمی، ترتیبی، فاصله‌ای و نسبی دسته‌بندی کنید.

الف) مدت زمان پاسخ‌گویی به سؤالات یک امتحان

ب) زمان اولین کلاس

پ) رشته تحصیلی

ت) مقیاس ارزیابی تحصیلی: ضعیف، معمولی و خوب

ث) نمره آخرین آزمون (از ۱۰۰ امتیاز)

ج) سن دانش‌آموز

## درس ۲

## معیارهای گرایش به مرکز

شاخص های که تمرکز داده ها را نشان می دهند را شاخص های مرکزی می گویند. مهمترین شاخص های مرکزی عبارتند از میانگین، مد و میانه.

میانگین: اصلی ترین و مورد استفاده ترین شاخص مرکزی، میانگین است. میانگین یعنی معدل داده ها که آن را با  $\bar{x}$  نشان می دهند. یعنی اگر داده ها به صورت  $x_1, x_2, x_3, \dots, x_n$  باشند، در این صورت میانگین آن ها از رابطه زیر به دست می آید.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

## کار در کلاس

۱. میانگین داده های  $10, 20, 30, 40, 50$  چقدر است؟
۲. اگر میانگین داده های  $10, x, 40$  برابر  $30$  شود مقدار  $x$  چقدر است؟
۳. میانگین اعداد  $1, 2, 3, 4, 5$  چقدر است؟
۴. میانگین اعداد  $2, 4, 6, 8, 5$  چقدر است؟
۵. میانگین اعداد  $5, 6, 7, 8, 9$  چقدر است؟
۶. میانگین اعداد  $50, 60, 70, 80, 90$  چقدر است؟
۷. آیا می توانید چند قاعده کلی از تمرین های قبل درباره خواص میانگین ذکر کنید.

**دورافتاده:** مقداری متفاوت با سایر مقادیر داده‌هاست. معمولاً مقدار آن بسیار بزرگ‌تر یا بسیار کوچک‌تر از بقیه داده‌هاست.

حال اگر یک میلیارد در با درآمد ماهیانه یک میلیارد ریال به انجمن خیریه دبیرستان ما بیاید، میانگین درآمد حضار چه تغییری می‌کند؟ (بیشتر مردم به او ثروتمند می‌گویند. آمارشناسان او را دورافتاده می‌نامند.) درآمد او میانگین را تا حدود ۱۴۸ میلیون ریال در ماه بالا می‌آورد (دقیقاً ۱۴۸/۶۲۵) و بر اساس شیوه تخمین گذشته، خیرین معادل ۱۷۸ میلیون ریال (دقیقاً ۱۷۸/۳۵) را به دبیرستان کمک خواهند کرد! که غیر واقعی به نظر می‌رسد یا امکان محقق شدن آن ضعیف است.

## نکاتی در مورد میانگین

۱- عموماً برای نشان دادن مجموع داده ها از حرف یونانی  $\Sigma$  (سیگما) استفاده می شود. مثلاً  $x_1 + x_2 + x_3 + x_4 + x_5 = \sum_{i=1}^5 x_i$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{لذا}$$

۲- با توجه به فرمول میانگین، می توان رابطه زیر را نوشت.  $x_1 + x_2 + x_3 + x_4 + \dots + x_n = n\bar{x}$  (یعنی اگر میانگین و تعداد داده ها مشخص باشند، مجموع داده ها از رابطه فوق به دست می آید)

۳- هرگاه داده های  $x_1, x_2, x_3, \dots, x_n$  تشکیل تصاعد عددی دهند، میانگین آن ها از رابطه زیر به دست می آید.  $\bar{x} = \frac{x_1 + x_n}{2}$

۴- میانگین وزن دار: هرگاه عددهای  $x_1, x_2, x_3, \dots, x_n$  دارای فراوانی های  $f_1, f_2, f_3, \dots, f_n$  باشند، در این صورت برای محاسبه

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + f_4 x_4 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n f_i x_i}{N}$$

میانگین از رابطه زیر استفاده می شود،  $N$  در این رابطه

کل داده هاست) که به این میانگین، میانگین وزن دار می گویند.

- در هر جامعه آماری فقط یک میانگین وجود دارد.

• کلیه مقادیر متغیر در مقدار میانگین تأثیر دارند.

- مجموع جبری تفاضل داده ها از میانگین صفر است.  $\sum_{i=1}^n (x_i - \bar{x}) = 0$

- هرگاه داده ها را با عدد ثابت  $b$  جمع کنیم، میانگین جدید برابر میانگین قبلی به اضافه  $b$  خواهد بود. یعنی  $\overline{x+b} = \bar{x} + b$

- هرگاه داده ها را در عدد ثابت  $a$  ضرب کنیم، میانگین جدید  $a$  برابر میانگین قبلی خواهد بود. یعنی  $\overline{ax} = a \times \bar{x}$

- اگر داده ها با هم برابر باشند، میانگین آن ها نیز همان مقدار مشترک خواهد بود.

کانال تلگرام: @mathvalizadeh

روش سریع برای محاسبه میانگین: برای محاسبه میانگین  $x_1, x_2, x_3, \dots, x_n$  میانگین داده ها را  $y$  تخمین می زنیم و اختلاف

$y$  را با هر یک از داده ها به ترتیب  $x'_1, x'_2, x'_3, \dots, x'_n$  قرار می دهیم سپس میانگین مورد نظر از فرمول زیر به دست می آید.

$$\bar{x} = y + \frac{\sum_{i=1}^n x'_i}{n}$$

## سوال:

میانگین داده های ۱، ۲، ۳، ۴، ۵ را به دست آورید.

میانگین ۵ داده برابر ۸ می باشد، مجموع آن ها را به دست آورید.

هرگاه میانگین اعداد ۱، ۳،  $a$ ، ۵ برابر ۲ باشد، مقدار  $a$  را بیابید.

میانگین ۵ داده آماری برابر ۸ و میانگین ۱۰ داده آماری دیگر برابر ۵ است. میانگین کل این ۱۵ داده آماری چه قدر است؟

میانگین جدول توزیع داده های زیر را به دست آورید.

$x_i$	۱	۳	۴	۶
$f_i$	۵	۱	۳	۵

بین  $x, y$  رابطه  $y_i = 7x_i - 3$  برقرار است. اگر میانگین  $x_i$  ها برابر 7 باشد، میانگین  $y_i$  ها چه قدر خواهد شد.



**میانہ:** عبارت است از مقداری که تعداد داده های بعد از آن برابر تعداد داده های قبل از آن باشد.

**روش به دست آوردن میانہ:** ابتدا داده ها را به طور صعودی مرتب می کنیم سپس اگر:

الف) تعداد داده ها فرد باشد، داده ای که در وسط قرار می گیرد، برابر میانہ است.

ب) تعداد داده ها زوج باشد، نصف مجموع دو داده ای که در وسط قرار گرفته اند، برابر میانہ است.

**روش دوم برای به دست آوردن میانہ:** ابتدا داده ها را به طور صعودی مرتب می کنیم، که  $\frac{n+1}{2}$  = شماره ردیف میانہ

### نکاتی در مورد میانہ

۱- در هر جامعه آماری فقط یک میانہ وجود دارد.

۲- میانہ نسبت به اندازه داده های حساسیت نشان نمی دهد، به عبارت دیگر میانہ به بزرگی و یا کوچکی داده ها کاری ندارد.

۳- میانہ در مقایسه با مد، اطلاعات بیشتر و مفید تری را در زمینه تمایل به مرکز هر توزیع آرایه می دهد.

۴- چنانچه داده ها یی که در ابتدا یا انتهای توزیع واقع شده اند، از لحاظ بزرگی و کوچکی، با سایر داده ها به طور قابل ملاحظه ای تفاوت داشته باشند، از میانہ به عنوان شاخص استفاده می کنیم.

۵- اگر فراوانی مقادیر به یک نسبت کم یا زیاد شود، مثلاً نصف یا دو برابر شود، میانہ تغییر نخواهد کرد.

۶- اگر داده ها را  $k$  برابر کنیم، میانہ نیز  $k$  برابر می شود.



### تمرین

۱. میانہ داده های ۱, ۶۸, ۲, ۸۶, ۱۴, ۱۰ چقدر است؟

۲. میانۀ داده‌های ۱, ۹۹, ۲, ۶۸, ۱۴, ۸۶, ۱۰, ۱۱ چقدر است؟

### کار در کلاس

با سه واژه: داده دورافتاده، میانه، و میانگین، ۵ جمله زیر را کامل کنید.

۱. آنچه اکثر مردم «حد وسط» می‌نامند، نزد آمارشناسان به ..... معروف است. برای محاسبه ..... به داده‌های خود به‌عنوان فهرستی از اعداد نگاه کنید؛ همه اعداد را باهم جمع کنید و بر تعدادشان تقسیم کنید.
۲. .... در واقع نقطهٔ وسطی فهرست اعداد مرتب‌شده است. نیمی از اعداد مقادیر بالاتر از ..... و نیمی دیگر مقادیر پایین‌تر از ..... قرار دارند.
۳. زمانی با ..... مواجه هستیم که مشاهده‌ای داشته باشیم که از الگوی داده‌هایمان پیروی نکند.
۴. وقتی با ..... مواجه هستید، معمولاً بازتاب بهتری از داده‌ها می‌دهد تا .....
۵. به‌طور کلی، برای برنامه‌ریزی و تصمیم‌گیری، ..... بهتر از ..... است.

مد: داده‌ای است که بیشترین فراوانی (تکرار) را در میان داده‌ها دارد.

#### نکاتی در مورد مد

- ۱- مد شاخصی است که ممکن است منحصر به فر نباشد، مثلاً ۱, ۲, ۳, ۴, ۱, ۲, ۵, ۷, ۵, ۵ دارای دو مد ۱, ۵ است. این قبیل جامعه‌ها را دو مده می‌گویند.
- ۲- جامعه ممکن است چند مدی باشد. لذا مد در این جامعه‌ها شاخص معتبری نیست. زیرا وجود چندین مد، نشان دهنده این است که جامعه ما یکدست نیست.
- ۳- اگر فراوانی مقادیر به یک نسبت کم یا زیاد شود، مثلاً نصف یا دو برابر شود، مد تغییر نخواهد کرد.
- ۴- اگر داده‌ها را  $k$  برابر کنیم، مد نیز  $k$  برابر می‌شود.

## درس ۳

## معیارهای پراکندگی

مقدمه: می دانیم که بهترین شاخص مرکزی میانگین است، ولی این شاخص همواره جوابگوی نیاز ما نیست، در این صورت می توان جوامع آماری را از نظر پراکندگی مورد بررسی و مقایسه قرار داد (منظور از پراکندگی یعنی این که داده ها از مرکز خود چه قدر دور هستند)، به عنوان مثال جدول زیر را در نظر بگیرید که در آمد ماهانه (بر حسب هزار تومان) سه نفر از جوامع مختلف را نشان می دهد.

	جامعه A	جامعه B	جامعه C
درآمد نفر اول	۳۰۰	۲۵۰	۷۰۰
درآمد شخص دوم	۳۰۰	۳۰۰	۱۵۰
درآمد شخص سوم	۳۰۰	۳۵۰	۵۰
میانگین	۳۰۰	۳۰۰	۳۰۰

همانطور که در جدول فوق دیده می شود، میانگین در هر سه جامعه برابر سیصد هزار تومان می باشد، ولی جوامع از لحاظ در آمد با همدیگر متفاوتند.

در جامعه A همه افراد جامعه در آمدشان یکسان و برابر می باشد و اختلاف در آمد افراد از میانگین صفر است.

در جامعه B بین در آمد افراد اختلاف وجود دارد ولی اختلاف آن ها نسبت به میانگین زیاد نیست، یعنی جامعه از حیث صفت مورد مطالعه نسبتاً یکدست (همگن) می باشد.

در جامعه C بین در آمد افراد اختلاف فاحشی وجود دارد و اختلاف در آمد افراد نسبت به میانگین زیاد است.

بنابراین فقط با دانستن میانگین نمی توان، در مورد جامعه قضاوت کرد، لذا استفاده از شاخص های دیگر ضرورت دارد که چگونگی پراکندگی داده ها را به صورت کمی ظاهر سازند، این شاخص ها به شاخص های پراکندگی معروفند.

مهم ترین شاخص های پراکندگی عبارتند از: دامنه تغییرات، دامنه چارک ها، واریانس و انحراف معیار

**دامنه تغییرات:** اختلاف بین بزرگترین و کوچک ترین مقدار داده را دامنه تغییرات می گویند و با  $R$  نشان می دهند، که

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



## نکاتی در مورد دامنه تغییرات

- ۱- دامنه تغییرات یک معیار سریع برای به دست آوردن پراکندگی بین داده ها است ولی معیار خوبی نیست. زیرا فقط بزرگترین داده و کوچک ترین داده را در نظر می گیرد و اطلاعات داده هایی را که بین بزرگترین و کوچکترین داده می باشند را در نظر نمی گیرد.
- ۲- اگر همه داده ها با هم برابر باشند، دامنه تغییرات برابر صفر می شود و اگر دامنه تغییرات صفر باشد در این صورت همه داده ها با هم برابر هستند.
- ۳- اگر همه داده ها با عددی جمع یا تفریق شوند، در این صورت در دامنه تغییرات، تغییری ایجاد نمی شود.
- ۴- اگر همه داده ها  $k$  برابر شوند، ( در  $k$  ضرب یا تقسیم شوند )، در این صورت دامنه تغییرات،  $k$  برابر می شود، یعنی (دامنه تغییرات قبلی)  $\times k =$  دامنه تغییرات جدید
- ۵- واحد اندازه گیری دامنه تغییرات با واحد اندازه گیری داده ها یکسان است.

## انحراف از معیار:

می خواهیم کلاس های دهم سه دبیرستان را بر اساس نتایج آزمون جامعی که هم زمان بین دانش آموزان ممتاز برگزار شده است، رتبه بندی کنیم. از هر دبیرستان ۱۰ نفر به تصادف انتخاب شده اند. نمرات آزمون جامع هر سه کلاس از ۱۲۰ نمره است.

داده های مربوط به سه کلاس دهم:

الف) {۶۵, ۷۵, ۷۳, ۵۰, ۶۰, ۶۴, ۶۹, ۶۲, ۶۷, ۸۵}

ب) {۸۵, ۷۹, ۵۷, ۳۹, ۴۵, ۷۱, ۶۷, ۸۷, ۹۱, ۴۹}

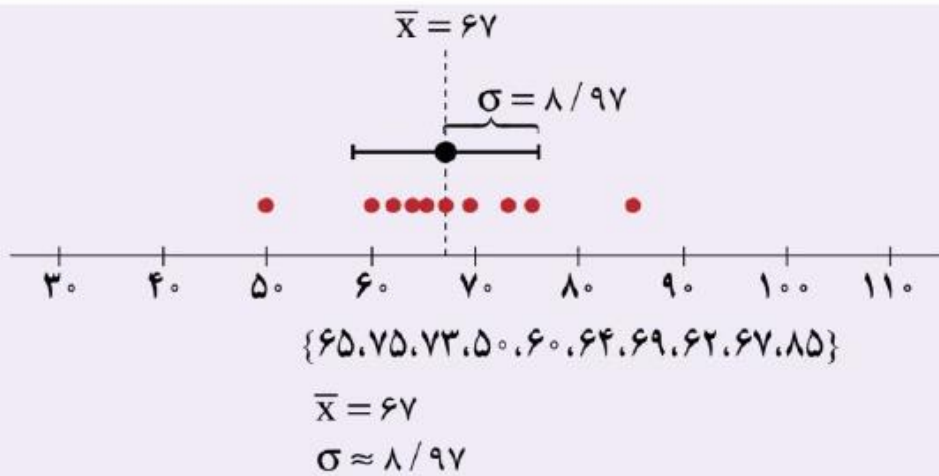
پ) {۴۳, ۵۱, ۵۳, ۱۱۰, ۵۰, ۴۸, ۸۷, ۶۹, ۶۸, ۹۱}

میانگین هر سه کلاس برابر ۶۷ است. پس معیار میانگین اطلاع دقیقی از تفاوت سه کلاس به ما نمی دهد به دنبال معیاری هستیم که میزان پراکندگی نمرات از میانگین را بیان کند که به این منظور از انحراف معیار استفاده می کنیم.

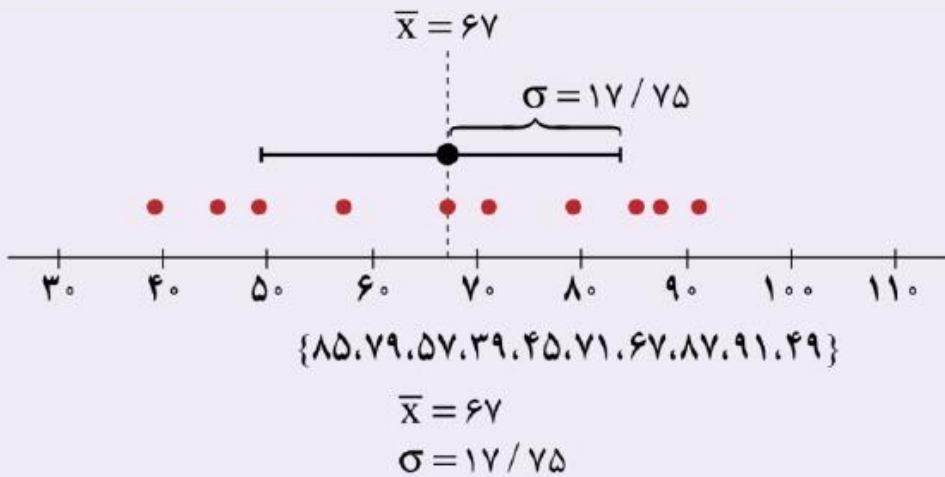
اگر  $n$  را به صورت  $x_1, x_2, \dots, x_n$  داشته باشیم، انحراف معیار آنها را با نماد  $\sigma$  (سیگما) نشان می دهند، که به صورت زیر تعریف می شود:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

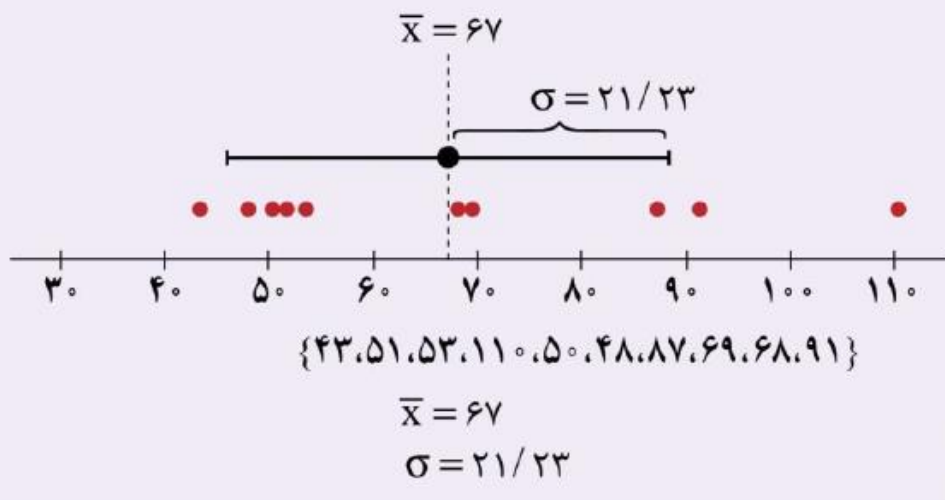
کلاس الف



کلاس ب



کلاس پ



مثلاً اگر والدین برای ثبت نام فرزند بازیگوش خود در سال آینده بخواهند بر اساس این نتایج یکی از این مدرسه ها را انتخاب کنند، مدرسه ای بهتر است که پراکندگی نمرات در آن کمتر است در مقابل برای فرزندی که به قصد شرکت در المپیاد می خواهد مدرسه را انتخاب کند، مدرسه ای بهتر است که پراکندگی نمرات در آن بیشتر باشد.

توجه: جدول زیر یک قانون در آمار است:

$(\bar{x} - 3\sigma, \bar{x} + 3\sigma)$	تقریباً ۹۹/۹ درصد از مشاهدات بین سه برابر انحراف معیار از میانگین قرار دارند
$(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$	تقریباً ..... مشاهدات بین ..... انحراف معیار از میانگین قرار دارند
.....	تقریباً ..... مشاهدات بین یک برابر انحراف معیار از میانگین قرار دارند

می خواهیم با همان مثال خیریه درس قبل بحث را ادامه دهیم. اگر انحراف معیار مجموعه داده ها کوچک باشد، بدین معناست که درآمد همه افراد به هم نزدیک است؛ اگر انحراف معیار بزرگ باشد، بدین معناست که درآمد افراد آن انجمن بسیار متفاوت است. انحراف معیار درآمد اعضا به این صورت محاسبه می شود:

مشاهدات		انحراف مشاهدات از میانگین		مربع انحراف مشاهدات از میانگین	
قبل از ورود میلیاردر	بعد از ورود میلیاردر	قبل از ورود میلیاردر	بعد از ورود میلیاردر	قبل از ورود میلیاردر	بعد از ورود میلیاردر
۴۰	۴۰				
۱۲	۱۲				
۲۸	۲۸				
۳۲	۳۲				
۳۰	۳۰				
۲۲	۲۲				
۲۵	۲۵				
	۱۰۰۰				

کانال تلگرام: @mathvalizadeh

انحراف معیار قبل و بعد از ورود میلیاردر اختلاف زیادی دارد نه؟ برای رفع این مشکل از معیاری به نام دامنه میان چارکی استفاده می کنیم. چارک اول در واقع میانه داده های مرتب شده قبل از میانه و چارک سوم میانه داده های مرتب شده بعد از میانه است. نمادهای  $Q_1, Q_2, Q_3$  به ترتیب برای چارک اول و دوم و سوم استفاده می شود.

چارک ها

اگر جامعه آماری به چهار قسمت مساوی تقسیم شود، چارک های اول تا سوم مشخص می شود  
 چارک اول ( $Q_1$ ): میانه نیمه اول داده ها (میانه داده های قبل از میانه اصلی) را چارک اول می گویند. (به عبارت دیگر، مقداری است که از ۲۵ درصد داده ها بزرگ تر و از ۷۵ درصد داده ها کوچک تر است).  
 چارک دوم ( $Q_2$ ): همان میانه اصلی داده ها می باشد (به عبارت دیگر، مقداری است که از ۵۰ درصد داده ها بزرگ تر و از ۵۰ درصد داده ها کوچک تر است).  
 چارک سوم ( $Q_3$ ): میانه نیمه دوم داده ها (میانه داده های بعد از میانه اصلی) را چارک سوم می گویند (به عبارت دیگر، مقداری است که از ۷۵ درصد داده ها بزرگ تر و از ۲۵ درصد داده ها کوچک تر است).

جدول زیر را برای مثال خیرین مدرسه تکمیل کنید.

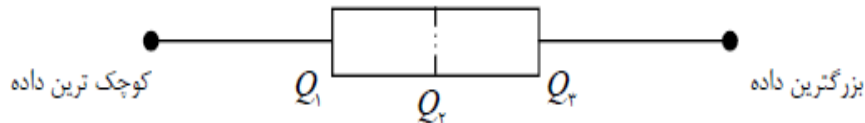
$\sigma$ انحراف معیار	$IQR = Q_3 - Q_1$ دامنه میان چارکی	$Q_3$ چارک سوم	$Q_2$ چارک دوم	$Q_1$ چارک اول	
					درآمد ماهیانه
					اعضای خیریه
					درآمد بعد از ورود میلیاردر

این اعداد امکان مشاهده چند چیز را به شما می دهند: میلیاردر مسبب درهم ریختگی میانگین درآمد و انحراف معیار بوده است. اما میانه و دامنه میان چارکی همچنان ثابت ماندند. این یکی از دلایلی است که ما می گوئیم، اگر داده دور افتاده داشته باشیم، از میانه (و در نتیجه دامنه میان چارکی) به جای میانگین استفاده کنید، در این صورت نتیجه بهتری از داده ها به دست می آورید. زمانی که تعداد داده ها زیاد باشد، به نحو دیگری می توان از تفاوت معیارهای گرایش به مرکز و پراکندگی نتیجه بگیرد که داده دور افتاده وجود داشته است.

## نمودار جعبه ای:

نمودارهایی که تاکنون شناختیم برای مقایسه داده ها بسیار مفید بودند اما نمودار جعبه ای بهتر از بقیه نمودارها پراکندگی داده ها را نشان می دهد. و این نمودار، یک نمودار تصویری می باشد که داده ها را بر اساس پنج مقدار زیر نشان می دهد.

(۱) کوچکترین داده (۲) چارک اول (۳) چارک دوم (میانه) (۴) چارک سوم (۵) بزرگترین داده



**نکته:** محل قرار گرفتن  $Q_2$ ، بستگی به داده ها دارد، لذا از لحاظ موقعیت قرار گرفتن، لزومی ندارد درست در وسط جعبه باشد، می تواند داخل جعبه مستطیلی تغییر نماید.

## مثال:

چارک اول، دوم و سوم داده ها را در هر یک از داده های زیر به دست آورید.

الف) ۲۵، ۱۱، ۱۵، ۲۳، ۵، ۱۲، ۶ (ب) ۵، ۱۷، ۱۸، ۱۱، ۱۳، ۸، ۷، ۲۰

تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان عبارتند از: ۱۹، ۳۱، ۲۵، ۱۸، ۳۲، ۴۳، ۴۱، ۳۴، ۱۶، ۲۷، ۱۴، ۲۳، ۱۵، ۱۰، ۱۲

الف) نمودار جعبه ای داده های فوق را رسم نمایید.

ب) چند درصد از داده ها در داخل جعبه قرار دارد؟

ج) چند درصد از داده ها در هر یک از دنباله ها قرار دارد؟

دلیل دیگر استفاده از میانه و دامنه میان چارکی تفسیر ساده آنها در مقابل تفسیر میانگین و انحراف معیار است. اگر نتایج مطالعات مربوط به سرطان را مطالعه کنیم، اولین چیزی که در گزارش‌ها مشاهده می‌کنیم، مشخصات عمومی بیماران مورد مطالعه است: آنها چند سال سن دارند؟ نسبت مردان به زنان چقدر است؟ چند نفر بیماریشان زود تشخیص داده شده و در مراحل ابتدایی است و چند نفر بیماریشان عود کرده است؟ اگر در گزارش، میانگین و انحراف معیار را ذکر کنیم، هر خواننده‌ای می‌تواند از روی اطلاعات داده شده با توجه به دو قانون ذکر شده نحوه توزیع سنین بیماران را به دست آورد. اما نکته در اینجاست که، آنها این کار را نخواهند کرد. به ندرت می‌توانید یک پزشک متخصص سرطان بسیار پرمشغله را در حال فکر کردن به این مسئله ببینید که، «خوب میانگین  $64/3$  و انحراف معیار  $9/8$  است؛  $68$  درصد از بیماران در فاصله یک برابر انحراف معیار از میانگین هستند؛ این یعنی،  $9/8 \pm 64/3$ ، که می‌شود، یک لحظه صبر کنید، ماشین حساب کجاست؟» شما تنها می‌توانید با نگاهی سریع به میانه و دامنه میان چارکی تصور خوبی نسبت به توزیع داده‌های پیش رویتان به دست آورید. به عبارت دیگر، میانه و دامنه میان چارکی در توصیف مجموعه داده‌ها بسیار مفیدند و این دقیقاً همان کاری است که ما از آنها انتظار داریم انجام دهند:

تمام آمارهایی که در اینجا ذکر شد (میانگین‌ها، میانه‌ها، انحراف‌های معیار، دامنه‌های میان چارکی) تحت عنوان آمار توصیفی شناخته می‌شوند.

### کار در کلاس

جملات زیر را کامل کنید :

۱. میانگین‌ها و میانه‌ها برای توصیف مجموعه داده‌ها مفیدند. .... و ..... انواعی از معیارهای گرایش به مرکزی هستند.
۲. شما معمولاً نه تنها معدل یک مجموعه را می‌خواهید، بلکه میزان تغییرات حوالی آن نقطه را هم نیاز دارید که آن معیار ..... است.
۳. معیار پراکندگی که معمولاً با میانگین بیان می‌شود، ..... است.
۴. معیار پراکندگی که معمولاً با میانه بیان می‌شود، ..... نام دارد.

۵. .... و ..... اطلاعات سریعی درباره داده‌ها بدون نیاز به هرگونه محاسبه می‌دهند.
۶. آماری که برای توصیف یک مجموعه داده، میانگین‌ها و میانها، انحرافات معیار و دامنه‌های میان چارکی به کار می‌رود، ..... دارد.
۷.  $50^\circ$  درصد داده‌ها قبل از ..... و  $50^\circ$  درصد داده‌ها بعد از ..... قرار دارند.
۸.  $75^\circ$  درصد داده‌ها قبل از ..... یا بعد از ..... قرار دارند.
۹.  $25^\circ$  درصد داده‌ها قبل از ..... یا بعد از ..... قرار دارند.
۱۰.  $50^\circ$  درصد داده‌ها بین ..... و ..... قرار دارند.
۱۱. تقریباً  $96^\circ$  درصد مشاهدات در فاصله ..... از میانگین هستند.
۱۲. تقریباً  $68^\circ$  درصد مشاهدات در فاصله ..... از میانگین هستند.

### تمرین

۱. یک نمونه ۲ تایی از بین اعداد ۱ تا ۶ انتخاب کنید. اگر بخواهید این نمونه حتماً تصادفی باشد، چه راهی را پیشنهاد می‌کنید. اگر بخواهید اعضای انتخابی این نمونه تصادفی، تکراری نباشند، چگونه این کار را انجام می‌دهید؟
۲. سؤال ۱ را برای اعداد ۱ تا ۳۶ تکرار کنید. (راهنمایی: می‌توانید از پرتاب دو تاس به عنوان روشی برای انتخاب اعداد ۱ تا ۳۶ استفاده کنید)

۳. مجموعه افراد فامیل درجه اول و درجه دوم خود را بنویسید و آنها را شماره گذاری کنید. سپس یک نمونه ۴ تایی از این جامعه انتخاب کنید. موضوع های زیر را در نظر بگیرید. درباره بهترین روش جمع آوری داده برای این فرضیه ها تصمیم بگیرید:

الف) بیشتر مردم فکر می کنند «حداکثر سرعت در اتوبان ها باید تعیین شود».

ب) آبی، رنگ مورد علاقه بیشتر مردم برای ماشین است.

پ) در زمان مطالعه، گوش دادن به موسیقی کلاسیک به یادگیری کمک می کند.

ت) بیشتر تصادفات اتومبیل ها را رانندگان با سن کمتر از ۲۵ سال مرتکب می شوند.

ث) رژیم گرفتن، باعث کاهش هوش می شود.

۴. دو موضوع برای هر یک از روش های جمع آوری داده بیان کنید.

الف) از طریق مصاحبه ب) از طریق مشاهده پ) از طریق دادگان ث) از طریق پرسش نامه

۵. کدام یک از نمونه گیری های زیر، یک نمونه گیری تصادفی است. در هر یک واحدهای آماری، جامعه و نمونه را مشخص کنید.

الف) با تمام پلیس های یک پاسگاه پلیس برای پیدا کردن نظر آنها راجع به تخلفات مصاحبه شد.

ب) با بچه هایی که وارد یک پارک بازی می شدند، پنج در میان مصاحبه شد تا وسیله بازی مورد علاقه کودکان بررسی شود.

پ) برای بررسی PH شامپوهای تولیدی یک کارخانه، شامپوها را صد در میان مورد آزمایش قرار دادیم.

۶. برای تحقیقات و بررسی های زیر جامعه را مشخص کرده و روش نمونه گیری پیشنهاد کنید.

الف) پیدا کردن درصد بیکاران استان مازندران. ب) مردان کدام شامپو را بیشتر ترجیح می دهند؟



۷. می خواهیم مدت زمانی را که دانش آموزان کلاس شما در طول یک هفته صرف مطالعه کتاب‌های غیردرسی می کنند، آمارگیری کنیم.

(الف) در این آمارگیری جامعه را مشخص کنید.

(ب) یک روش نمونه گیری برای انتخاب نمونه معرفی کنید.

(پ) از چه روشی برای گردآوری اطلاعات استفاده می کنید؟

(ت) اندازه (تعداد) این جامعه چقدر است؟

(ث) اندازه نمونه مورد بررسی چقدر است؟ آیا این تعداد برای بررسی مورد نظر مناسب است؟ در صورت مناسب نبودن، روش نمونه گیری و نمونه خود را اصلاح کنید.

(ج) متغیر مورد مطالعه در این مسئله چیست؟ توضیح دهید.

(چ) متغیر مورد مطالعه از چه نوعی است؟

۸. نوع هر یک از متغیرهای زیر را مشخص و بهترین مقیاس اندازه گیری آنها را ذکر کنید.

(الف) رنگ اتومبیل‌های موجود در یک نمایشگاه اتومبیل (ب) درجه حرارت کلاس شما در روزهای سال

(پ) گنجایش آب یک باری مخزنی (تانکر) (ت) تعداد شکایات رسیده شده به یک پاسگاه پلیس

(ث) درآمد دانشجویان شاغل به کار (ج) وضعیت تأهل کارمندان یک شرکت

(چ) سن دانشجویان شرکت کننده در یک دوره هنری

۹. میزان پرداخت حقوق در یک شرکت خصوصی در جدول زیر آمده است :

سِمَت	مدیر شرکت	معاون	حسابدار	بازاریاب	۲ نفر فروشنده	منشی	مستخدم
حقوق میلیون ریال	۱۰۰	۶۰	۳۰	۲۰	۲۰	۱۲	۸

مدیر شرکت در آگهی دعوت به همکاری در روزنامه اعلام می دارد که میانگین پرداخت حقوق در این شرکت بیشتر از ..... ریال است. آیا این رقم می تواند دلیلی برای پرداخت حقوق بالاتر در این شرکت باشد؟ چرا؟ کدام یک از معیارهای گرایش به مرکز برای نشان دادن وضع پرداختی کارکنان این شرکت مناسب تر است؟

۱۰. نمرات درس فیزیک دانش آموزی در طول سال برابر است با:

۱۷	۱۹	۱۸	۷	۱۸	۱۹
----	----	----	---	----	----

الف) میانگین و میانگین را برای نمرات این دانش آموز حساب کنید.

ب) کدام یک از شاخص‌های فوق، بیانگر وضع این دانش آموز در درس فیزیک است؟

پ) اگر معلم درس فیزیک این دانش آموز، برای جبران نمره ۷، امکان امتحان مجدد را به او بدهد، برای اینکه میانگین وی در این درس بیشتر از ۱۸ شود، او در این امتحان چه نمره‌ای باید کسب کند؟

۱۱. دو دسته داده زیر قیمت کالایی را در دو بازار جداگانه برحسب هزار ریال نشان می‌دهد.

۸	۱۳	۹	۱۲	۱۰	۱۱	۱۲	۹	۱۰	۱۱	۱۰	بازار الف
۱۰	۱۳	۸	۱۰	۹	۱۱	۱۰	۱۰	۱۱	۹	۱۰	بازار ب

الف) در کدام دسته پراکندگی بیشتر است؟

ب) دامنه تغییرات را محاسبه کنید.

پ) آیا دامنه تغییرات با پراکندگی مشاهده شده در «الف» همخوانی دارد؟

ت) ترجیح می‌دهید از کدام بازار خرید کنید؟ چرا؟

ث) اگر داده‌ها را در اختیار نداشته باشید، فقط به صرف داشتن دامنه تغییرات می‌توانید تصمیم‌گیری کنید؟

سوالات تکمیلی:

یک کارخانه سازنده اتومبیل های شخصی در نظر دارد، اتومبیل های جدیدی را طراحی نماید. برای این منظور اتومبیلهایی را که از یک چهارراه عبور می کردند، مورد بررسی قرار داد. اطلاعات به دست آمده به شرح زیر است.

نوع سرنشین	۱ سرنشین	۲ سرنشین	۳ سرنشین	۴ سرنشین	۶ سرنشین
تعداد اتومبیل	۱۲	۶	۴	۲	۱

الف) به طور متوسط هر اتومبیل چند سرنشین دارد؟ (ب) در این بررسی مد چند است؟

برای داده های جدول زیر میانه و مد را به دست آورید.

داده	۱	۲	۶	۷
فراوانی	۲	۳	۴	۶

در یک مجموعه اندازه مرتب شده، با ۸۰ داده داریم،  $x_{۴۰} = ۶۰$ ،  $x_{۳۳} = ۳۳$ ،  $x_{۲۷} = ۲۷$ ،  $x_{۲۳} = ۲۳$  میانه داده ها را بیابید.

نمره درس آمار و مدلسازی دوازده دانش آموز یک کلاس به صورت زیر داده شده است:

۸، ۱۱، ۷، ۱۲، ۱۰، ۱۳، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹

میانه و مد نمرات را به دست آورید و سپس بیان کنید کدام یک از این شاخص ها گویا تر است.

برای داده های جدول زیر مقدار  $a, b, c, d$  را به دست آورید.

$x_i$	۱	$b$	$c$	$d$
$x_i - \bar{x}$	$a$	۲	۳	۴

معدل نمرات درس آمار و مدلسازی سه کلاس،  $A = ۱۶$  کلاس  $B = ۱۶/۵$  و  $C = ۱۷$  کلاس، می باشد، اگر تعداد دانش آموزان کلاس ها به ترتیب ۳۰ و ۳۲ و ۳۵ نفر باشد، معدل نمرات سه کلاس روی هم چه قدر است؟

میانگین نمرات دانش آموزان یک کلاس ۲۰ نفره مساوی ۱۰ است. اگر نمره یکی از دانش آموزان به ۲۰ تغییر یابد، میانگین نمرات حداکثر چند نمره تغییر می کند.

میانگین نمره درس ریاضی دو یک کلاس ۱۶ و میانگین نمرات کلاس دیگر که جمعیت آن دو برابر جمعیت این کلاس است. ۱۰ می باشد، معدل نمره دو کلاس روی هم چه قدر است؟

اگر میانگین داده های  $x_1, x_2, x_3, \dots, x_n$  برابر ۲۰ باشد، میانگین داده های  $\frac{1}{4}x_1 - 1, \frac{1}{4}x_2 - 1, \frac{1}{4}x_3 - 1, \dots, \frac{1}{4}x_n - 1$  کدام است؟

الف) ۹      ب) ۲۰      ج) ۱۹      د) ۱۵

اگر  $\mu$  میانگین اعداد طبیعی از ۱ تا ۸۵ باشد، مقدار عبارت  $\sum_{i=1}^{85} (x_i - \mu)$  کدام است؟

الف) -۱۱      ب) ۰      ج) ۱      د) ۴۳

اگر میانگین داده های  $x_1, x_2, x_3, \dots, x_n$  برابر  $\bar{x}$  باشد، میانگین داده های  $x_1 + \bar{x}, x_2 + 2\bar{x}, x_3 + 3\bar{x}, \dots, x_n + n\bar{x}$  کدام است؟

الف)  $\frac{(n+1)\bar{x}}{2}$       ب)  $2\bar{x}$       ج)  $\frac{(n+3)\bar{x}}{2}$       د)  $\frac{n\bar{x}}{2}$

اگر میانگین داده های  $2n, n+2, \dots, n+1, n$  برابر  $\bar{x}$  باشد، میانگین داده های  $2n - (n+1), (n+2) - 2, (n+1) - 1, n - 1$  کدام است؟

الف)  $\bar{x} - \frac{n}{2}$       ب)  $\bar{x}$       ج)  $\bar{x} - \frac{n+1}{2}$       د)  $\bar{x} - \frac{n+2}{2}$



یادداشت: