



بخش آموزش رسانه تفریحی سنتر

کلیک کنید  www.tafrihicenter.ir/edu

 نمونه سوال  گام به گام

 امتحان نهایی  جزوه

 دانلود آزمون های آزمایشی

متوسطه اول : هفتم ... هشتم ... نهم

متوسطه دوم : دهم ... یازدهم ... دوازدهم

www.tafrihicenter.ir

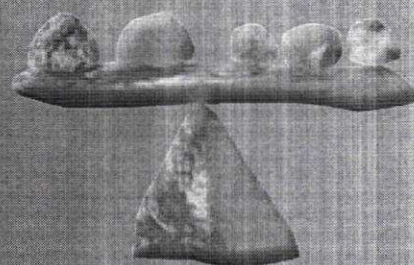
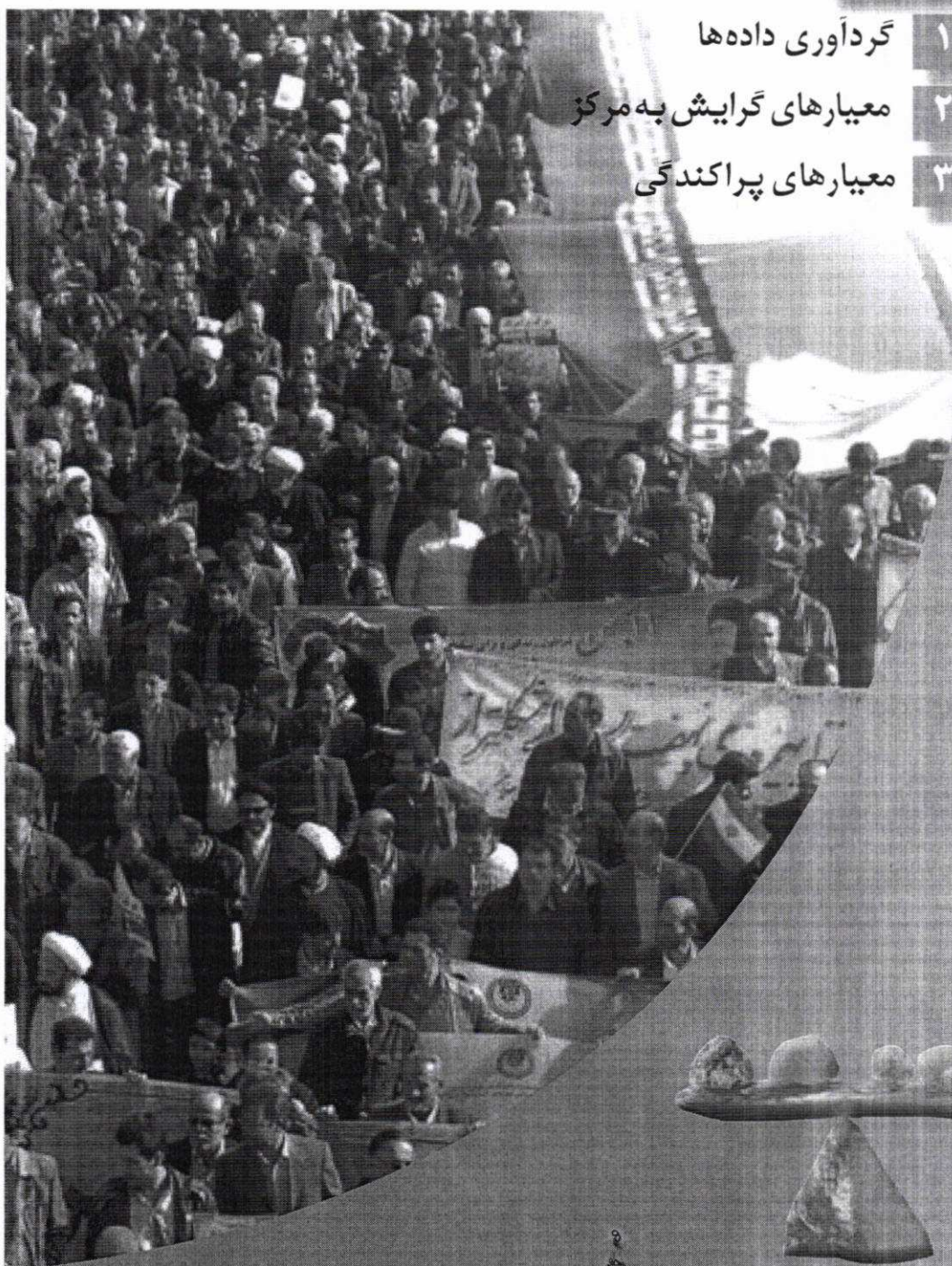
کار با داده‌های آماری

فصل چهارم

درس ۱ گردآوری داده‌ها

درس ۲ معیارهای گرایش به مرکز

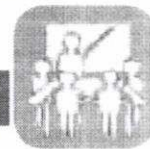
درس ۳ معیارهای پراکندگی



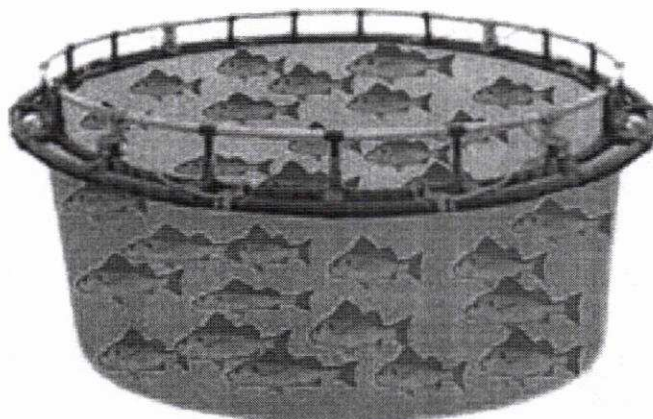
درس ۱

گردآوری داده‌ها

فعالیت



می‌خواهیم وزن ماهی‌های یک حوضچه پرورش ماهی را به منظور فروش آنها تخمین بزنیم. ابتدا از قسمت عمیق و در مرحله بعد، از قسمت کم عمق ۵ ماهی صید می‌کنیم.



انتخاب ۵ ماهی از قسمت عمیق



انتخاب ۵ ماهی از قسمت کم عمق



۱. آیا انتخاب ۵ ماهی از قسمت عمیق، تخمین خوبی از وزن ماهی‌های حوضچه به ما می‌دهد؟ یعنی آیا می‌توان برای فروش آنها اعلام آمادگی کرد؟ **خیر**

۲. به نظر شما کدام تخمین بیش از مقدار واقعی است و کدام یک کمتر از مقدار واقعی؟

تخمین از قسمت عمیق بیشتر - از قسمت کم عمق کمتر
از مقدار واقعی است.

از قفسه ها مختلف
مناسب کتابت

۳. اگر شما امکان صید ۵ ماهی داشتید، چگونه آنها را انتخاب می کردید تا منجر به تخمین بهتری از وزن ماهی های حوضچه می شد؟ فرض کنید ماهی ها، همانند شکل در حوضچه پخش شده اند و تحرک زیادی ندارند.

۴. اگر از نحوه پخش شدن ماهی ها اطلاعی نداشتیم، بهتر بود ۵ ماهی را چگونه انتخاب می کردیم؟ آیا انتخاب تصادفی چند بخش از حوضچه (مثلاً زمانی که آن را شطرنجی کرده ایم) به ما کمک می کند ماهی های انتخابی معرف بهتری از کل ماهی ها باشند؟ **انتخاب تصادفی مناسب کتابت.**

داده ها واقعیت هایی درباره یک چیزاند که در محاسبه، استنباط، یا برنامه ریزی به کار می روند. واحد آماری به هر یک از افراد یا چیزهایی می گویند که داده های مربوط به آنها در یک بررسی آماری گردآوری می شود.

مجموعه کل واحدهای آماری را جامعه آماری می نامند.

هر زیرمجموعه از جامعه آماری را که با روش مشخصی انتخاب شده باشد، یک نمونه می نامند. نمونه ای را که در آن، همه اعضای جامعه، شانس انتخاب یکسان در نمونه را داشته باشند نمونه تصادفی می نامند.

در مثال حوضچه ماهی، هر ماهی درون حوضچه یک واحد آماری است. به کل ماهی های حوضچه که عبارت است از مجموعه همه واحدهای آماری جامعه گفته می شود. اگر وزن تک تک ماهی ها را در اختیار داشته باشیم داده های جامعه را داریم. وزن نمونه ۵ ماهی از قسمت کم عمق معرف داده های یک نمونه پنج تایی است. اگر ۵ ماهی با یک روش تصادفی از حوضچه استخراج شود، عملاً یک نمونه تصادفی ۵ تایی از حوضچه در اختیار داریم.

از قدیم گفته اند که
مشت نمونه خروار است،
ولی نه هر مشت. چرا؟
اگر نمونه تصادفی
انتخاب شود
آنگاه نمونه
معرف جامعه
خواهد بود.

خواندنی

تفاوت زیادی بین عدد در ریاضی و داده در آمار وجود دارد. به عبارت دیگر عدد ۵۰ یک مفهوم در ریاضی دارد و داده ای که مقدار آن ۵۰ است، علاوه بر مقدار آن حاوی اطلاعات زیادی است. به عنوان مثال این داده می تواند متوسط تلفات روزانه جاده های کشور در یکی از سال های اخیر باشد.



تهیه کننده:

فعالیت



می خواهیم طول قد دانش آموزان یک مدرسه را گردآوری کنیم. برای این منظور چه راهی پیشنهاد می کنید؟
اندازه گیری قد نامی دانش آموزان مدرسه

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

اگر قرار شد آمارگیر باشیم، می توانیم جدولی به صورت زیر تکمیل کنیم.
مثالی از جدول طراحی شده برای ثبت داده ها

تعداد دانش آموزان	جوب خط برای شمارش	اندازه طول قد
		کوتاه تر از ۱۴۰ سانتی متر
		۱۴۰-۱۴۹ سانتی متر
		۱۵۰-۱۵۹ سانتی متر
		۱۶۰-۱۶۹ سانتی متر
		۱۷۰ سانتی متر یا بلندتر

چگونه مطمئن می شوید که دانش آموزی از قلم نیفتاده است؟ چه راهکاری برای این منظور پیشنهاد می کنید؟

آمارگیری زحمت زیادی برای آمارگیر دارد. آیا راه حل ساده تری برای انجام آن دارید؟ یکی از مرسوم ترین روش های آمارگیری، استفاده از پرسش نامه است. پرسش نامه شبیه همان جدولی است که هنگام ثبت نام در مدرسه، شما یا والدین، آن را تکمیل کرده اید. پرسش نامه را می توانند واحدهای جامعه یا نمونه تکمیل کنند.



مثالی از پرسش‌نامه طراحی شده

سلام، می‌خواهیم طول قد دانش‌آموزان مدرسه را آمارگیری کنیم.
لطفاً یکی از گزینه‌ها را انتخاب کنید.
طول قد شما چقدر است؟

- ☐ کوتاه‌تر از ۱۴۰ سانتی‌متر
- ☐ ۱۴۰-۱۴۹ سانتی‌متر
- ☐ ۱۵۰-۱۵۹ سانتی‌متر
- ☐ ۱۶۰-۱۶۹ سانتی‌متر
- ☐ ۱۷۰ سانتی‌متر یا بلندتر

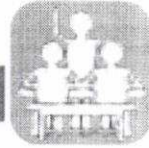
خواندنی

برای به انجام رساندن یک آمارگیری
باید پاسخ سؤال‌های زیر را بدانید:

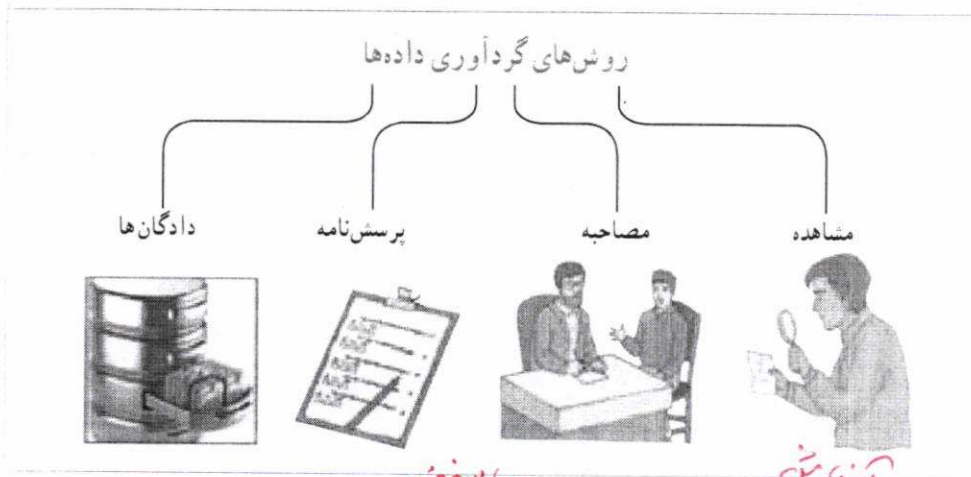
۱. می‌خواهید چه چیزی پیدا کنید؟ چرا؟
۲. چه داده‌هایی را باید گردآوری کنید؟ چگونه تصمیم می‌گیرید؟
۳. کدام راه، بهترین راه گردآوری داده‌ها است؟ چرا؟ آیا به ابزار خاصی نیاز دارید؟
۴. به چه مقدار داده نیاز دارید؟ چگونه تصمیم می‌گیرید؟
۵. داده‌های شما باید چقدر دقت داشته باشند؟ چرا؟
۶. داده‌هایتان را چگونه ثبت می‌کنید؟ چرا این روش را انتخاب کرده‌اید؟
۷. داده‌ها را چگونه ارائه می‌کنید؟ چرا؟
۸. آیا داده‌ها از الگوی خاصی پیروی می‌کنند؟ داده‌ها بیانگر چه هستند؟
۹. چه نتیجه‌گیری یا پیشگویی خاصی می‌توانید از داده‌ها ارائه کنید؟
۱۰. آیا نتایج، با آنچه انتظار داشتید، تطبیق دارد؟
۱۱. چگونه نتایج کار را ارائه می‌کنید؟ برای چه کسانی ارائه می‌کنید؟
۱۲. با توجه به نتایج به دست آمده، آیا می‌توانید سؤالات دیگری را نیز بررسی کنید، به نظر شما در مرحله گردآوری داده، به کدام یک از سؤالات فوق باید پاسخ داده شود؟

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان



۱. چه راه دیگری برای آمارگیری طول قد دانش آموزان یک مدرسه پیشنهاد می کنید؟ *پرسش از تعدادی از دانش آموزان*
۲. فرض کنید زمان لازم را برای گردآوری تمامی داده های دانش آموزان در اختیار نداشته باشید. اگر بخواهیم نمونه ای را انتخاب و آمارگیری کنیم، چه راهی پیشنهاد می کنید که نمونه به صورت تصادفی انتخاب شود؟ *از هر کلاس چند دانش آموز به طور تصادفی انتخاب می کنیم.*



۱. مشاهده: گردآوری داده ها بدون نیاز به فرد پاسخ دهنده، مانند شمارش تعداد وسایل نقلیه عبوری از یک تقاطع در هر ساعت یا اندازه گیری وزن محصولات یک باغ میوه.



تهیه کننده:

شعارهای برتر
سال جهانی آمار



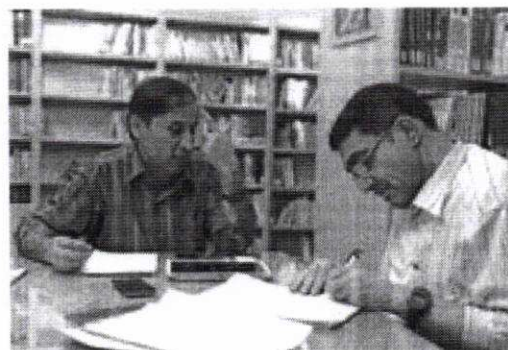
آمار، تصویر دیروز،
آینه امروز، دورنمای فردا

آمار، سنجش گذشته؛
شناخت حال؛
درک آینده



۲. پرسشنامه: مجموعه سؤالات از پیش

تعیین شده که توسط تعدادی پاسخ دهنده تکمیل می شود. این روش مرسوم ترین ابزار گرفتن اطلاعات از مردم است. مرکز آمار ایران هر ۱۰ سال یکبار با استفاده از پرسشنامه اطلاعات تمامی خانوارهای ساکن در ایران را جمع آوری می کند. به این فرایند، سرشماری نفوس و مسکن می گوئیم.



۳. مصاحبه: معمولاً بین دو نفر صورت

می گیرد، یکی مصاحبه گر (همان آمارگیر) و دیگری مصاحبه شونده یا پاسخ گو است. مثلاً اگر بخواهیم درباره مسائل فرهنگی کاهش شدآمد (ترافیک) پژوهش کنیم، مصاحبه از صاحب نظران راه حل مناسبی برای گردآوری داده ها است. این روش بیشتر زمانی استفاده می شود که آمارگیر اطلاع کافی از تمامی پاسخ های ممکن را ندارد.

۴. دادگانها: شامل مجموعه ای از اطلاعات ذخیره شده اند. در بسیاری از موارد، داده ها را می توان از

اطلاعاتی که قبلاً ذخیره شده اند، به دست آورد. اگر قرار است تحقیقی در مورد نمره های دروس ریاضی استان ها انجام شود، اطلاعات ثبتی اداره کل آموزش و پرورش راه گشا خواهد بود. از سوی دیگر به دلیل تولید داده ها به صورت خودکار، در بسیاری از مؤسسات و سامانه ها، استفاده از این روش برای گردآوری داده ها به سرعت رواج یافته است.

تمرین



کدام روش جمع آوری داده ها برای موارد زیر مناسب است؟ یک دلیل برای انتخاب خود ذکر کنید.

۱. میزان رضایت مشتریان بانک از نحوه برخورد و رسیدگی به درخواست های آنها. *پرسشنامه*
۲. سن همه دانش آموزان مدرسه بر حسب ماه در پایه دهم. *دادگانها*
۳. تعداد سرنشینان خودروهای سواری در یکی از محورهای خروجی شهر. *مشاهده*



کار در کلاس



مشاهده

مشاهده

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

۱. تعداد قلم‌های هر دانش‌آموز در یک کلاس.

۲. ساعات خواب دانش‌آموزان کلاس درس شما در شب گذشته.

۳. طول قد دانش‌آموزان یک کلاس.

مشاهده

ب) می‌خواهیم طول قد دانش‌آموزان یک کلاس یا مدرسه را به یکی از سه روش زیر آمارگیری کنیم.

هریک از این روش‌ها محدودیت‌هایی دارند. چگونه می‌توان این محدودیت‌ها را از بین برد؟
نمونه برداری کنیم.



پرسش‌نامه: اگر تعداد واحدهای نمونه زیاد باشد، این روش زمان‌بر است.

مشاهده: اگر به‌دقت زیادی نیاز داریم، مناسب نیست.

دادگان‌ها: همیشه اطلاعات ثبتی را در اختیار آمارگیر قرار نمی‌دهند.



پرسش‌نامه: نمونه‌گیری مشاهده: اندازه‌گیری
استخراج از وسایل نقلیه
دادگان‌ها: توجه به صاحبان اطلاعات

آمارگیری را می‌توان به روش‌هایی بسیار سریع‌تر یا کم‌هزینه‌تر مانند آمارگیری پستی، تلفنی، اینترنتی یا پیامکی انجام داد. همچنین می‌توان با ایزاری نظیر گوگل فرم یک پرسش‌نامه طراحی کرد، و آن را به نشانی نمونه انتخابی ارسال کرد و نتایج را از گوگل فرم بازبایی کنیم.

خواندنی



فعالیت

قرار است درباره افرادی که از کوه دنا بالا رفته‌اند، پژوهشی آماری انجام دهیم. واحدهای آماری این پژوهش، همه افرادی هستند که توانسته‌اند به قله برسند. هدف از این پژوهش می‌تواند فرهنگی یا علمی باشد. بسته به نوع پژوهش، یک یا چند ویژگی این افراد (مانند طول قد یا جنسیت) مورد نیاز است. به هر یک از این ویژگی‌ها که مورد پژوهش قرار می‌گیرد متغیر می‌گویند. سایر متغیرها می‌توانند مواردی مانند: سن، وزن، ملیت، میزان تحصیلات و درآمد باشند. متغیرهای مورد بررسی در یک پژوهش ممکن است کمی یا کیفی باشند.

تهیه‌کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

متغیر: هر ویژگی از اشخاص یا اشیا که قرار است بررسی شود.
متغیر کمی: متغیرهایی هستند که مقادیر عددی می‌گیرند و برای آنها عملیات ریاضی از قبیل جمع، تفریق و معدل‌گیری قابل انجام است.
متغیر کیفی: متغیرهایی هستند که صرفاً برای دسته‌بندی افراد یا اشیا در گروه‌ها به کار می‌روند و لزوماً مقدار عددی نمی‌گیرند.

در مثال کوهنوردان دنا، سن، وزن، قد و درآمد یک کوهنورد متغیرهای کمی هستند. متغیرهای کیفی معمولاً از نوع مشاهدات غیر عددی‌اند و در مثال کوهنوردان دنا، جنسیت و ملیت را در بر می‌گیرند. به عنوان مثال جنسیت برای دسته‌بندی افراد به مرد و زن استفاده می‌شود.

پارامتر جامعه: یک مشخصه عددی است که توصیف‌کننده جنبه‌ای خاص از جامعه است و در صورتی که داده‌های کل جامعه در اختیار باشند قابل محاسبه است. مثلاً اگر داده‌های مربوط به تک‌تک کوهنوردان را داشته باشیم، یعنی به داده‌های جامعه دسترسی داریم. نسبت مردان در کل جامعه کوهنوردان، معرف یک پارامتر است.

اگر داده‌های بعضی از کوهنوردان را داشته باشیم؛ یعنی داده‌های نمونه را در اختیار داریم. نسبت مردان کوهنورد به این داده‌های نمونه‌ای را، آماره (مقدار آماره) گویند. آماره‌ها از یک نمونه به نمونه دیگر تغییر می‌کنند؛ این در حالی است که پارامترهای جامعه همیشه ثابت‌اند، چرا؟ **چون جامعه یقیناً یک چیز است.** در بسیاری از موارد، آمارگیری از کل جامعه امکان‌پذیر نیست. بنابراین علی‌رغم اینکه پارامتر دارای مقدار ثابتی است، این مقدار مجهول است و به همین دلیل از آماره‌ها برای تخمین پارامترها استفاده می‌کنند.

آماره نمونه: مشخصه‌ای عددی که توصیف‌کننده جنبه‌ای خاص از نمونه است و از داده‌های نمونه به دست می‌آید.

مسئله: اداره کشاورزی استان خوزستان در حال ارزیابی هندوانه‌های آماده برداشت است. در این بررسی، هندوانه‌ها همان واحدهای آماری هستند. اگر پژوهشگران وزن هندوانه‌ها را مورد بررسی قرار دهند، متغیر، «وزن» آنهاست. وزن یک متغیر کمی است، زیرا با مقادیر عددی ارائه می‌شود. اگر وزن تک‌تک هندوانه‌های این زمین بررسی شود، **سرشماری** از جامعه انجام داده‌ایم (که امکان‌پذیر نیست). متوسط وزن تمامی هندوانه‌های 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.

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

محدودیت‌های سرشماری

عبارت است از:

۱. هزینه زیاد
۲. زمان بر بودن
۳. خطای بیشتر در گردآوری داده‌ها
۴. امکان استفاده در بررسی‌های مخرب (واحدهای انتخاب شده را از بین می‌برند) ندارد.



کار در کلاس

یک شبکه تلویزیونی می خواهد نسبت دارندگان تلویزیون در شیراز را، که برنامه جدید این شبکه را حداقل یک بار در هفته تماشا می کنند، بداند. بدین منظور یک گروه ۱۰۰۰ نفری از دارندگان تلویزیون را در این شهر بررسی می کند.

الف) داده ها و متغیرهایی را که بررسی می شوند، مشخص کنید.

جواب: داده ها اطلاعات گروه ۱۰۰۰ نفری دارندگان تلویزیون در شیرازند، و متغیر، تماشای تلویزیون است که پاسخ آن «تماشا می کند» یا «تماشا نمی کند» افراد مورد بررسی است.

ب) آیا این داده ها یک نمونه اند؟ جامعه آماری کدام است؟ **بله - جامعه کل ساکن شیراز هستند.**

پ) متغیر کمی است یا کیفی؟ **کمی**

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

جواب: سن، درآمد، **و غیره**

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

مقیاس های اندازه گیری

داده ها را به دو گروه کمی و کیفی تقسیم کردیم. از نگاهی دیگر، می توان متغیرهای داده ها را در چهار مقیاس اندازه گیری دسته بندی کرد. اندازه گیری در تعریف به معنی ایجاد تفکیک بین افراد یا اشیا است. دو نوزاد دو قلو را نام گذاری می کنیم تا آنها را تفکیک کنیم، در واقع اندازه گیری کرده ایم. بسته به دقتی که این اندازه گیری صورت می گیرد آن را به چهار مقیاس اسمی، ترتیبی، فاصله ای و نسبتی تقسیم می کنیم. هدف بررسی این مقیاس ها بیانگر نوع محاسبه ای است که برای این داده ها مناسب است؛ نظیر: ترتیب، محاسبه اختلاف و نسبت گرفتن.

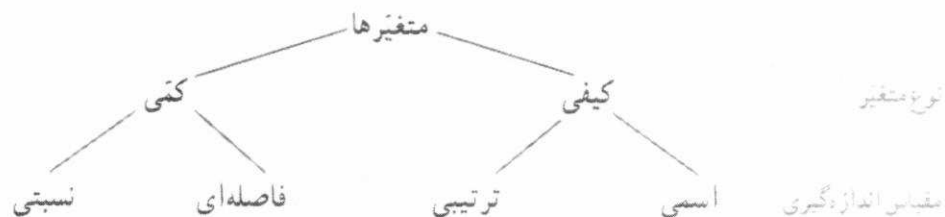
نسبتی: این مقیاس برای داده هایی است که قابل مرتب کردن هستند، اختلاف بین مقادیر داده ها، و نسبت مقادیر داده ها نیز بامعنا است. اغلب متغیرهای فیزیکی مانند نمره، وزن و قد دانش آموزان و متغیرهایی که با واژه تعداد شروع می شوند در این مقیاس اندازه گیری می شوند. در این مقیاس صفر به معنی نبود ویژگی در فرد یا شیء است.

فاصله ای: این مقیاس به دلیل استفاده از لوازم یا قواعد دقیق اندازه گیری ویژگی افراد یا اشیا به دقت اندازه گیری می شود. به بیان دیگر مقیاس فاصله ای برای داده هایی است که قابل مرتب کردن هستند و همچنین، اختلاف بین مقادیر داده ها بامعناست. مانند درجه حرارت در شهرهای مختلف برحسب سلسیوس. مقادیری که به دو نفر یا دوشیء داده می شود صرفاً بیانگر فاصله بین آنهاست. در نتیجه صفر در این مقیاس قراردادی است. مثلاً اگر دمای بوشهر ۲۰ و تهران ۱۰ درجه سلسیوس باشد نمی توان گفت دمای تهران دو برابر بوشهر است ولی اختلاف دما ۱۰ می باشد.

ترتیبی: این مقیاس با استفاده از الفاظ، ضمن ایجاد تفکیک بین افراد و اشیا، ارجحیت نیز قائل می شود. مقیاس ترتیبی برای متغیرهایی است که قابل مرتب کردن هستند؛ در عین حال محاسبه اختلاف بین مقادیر داده ها، یا امکان پذیر نیست و یا بی معناست، مانند رتبه دانش آموزان در یک کلاس. اگر رتبه های اول تا سوم معدل ۱۹، ۱۸، ۱۶ کسب کرده باشند رتبه ۳، ۲، ۱ می دهیم و توجه نمی کنیم که اختلاف نمرات آنها چقدر است.

اسمی: این مقیاس برای متغیرهایی است که شامل نام ها، برجسب ها و گروه ها می شود. در اینجا هیچ معیاری که با آن بتوان داده ها را از کوچک به بزرگ مرتب کرد وجود ندارد. مانند گروه خونی انسان ها و شماره دانش آموزی. کدهای عددی در این مقیاس به واقع عدد نیستند بلکه صرفاً برای گروه بندی به کار می رود.

اگر فقط یک متغیر از داده‌ها اندازه‌گیری شده باشد، به جای ذکر «مقیاس متغیر» از واژه «مقیاس داده‌ها» استفاده می‌کنند.



شعارهای برتر
سال جهانی آمار



بهترین برنامه‌ریزی
مستلزم بهترین آمار

با آمار بهتر بفهمیم.
بهتر تصمیم بگیریم

کار در کلاس



نوع متغیر داده‌های زیر را مشخص کنید:

- الف) محسن، محمود، محمد و میثم همگی اسامی مذکر هستند. **اسمی (کیفی)**
- 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. **کمی (نسبتی)**
- پ) دمای بدن ماهی‌های قزل‌آلای رودخانه هراز (برحسب درجه سلسیوس) **فاصله‌ای (کمی)**
- ت) طول ماهی‌های قزل‌آلا در رودخانه هراز **نسبتی (کمی)**

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. **اسمی (کیفی)**
- ب) این نماینده ۵۸ سال سن دارد. **کمی (نسبتی)**
- پ) سال‌هایی که این نماینده در مجلس انتخاب شده است، ۱۳۸۶، ۱۳۹۰ و ۱۳۹۴ است. **کمی (فاصله‌ای)**
- ت) مجموع حقوق این نماینده در سال گذشته ۶۰۰,۰۰۰,۰۰۰ ریال بوده است. **کمی (نسبتی)**
- 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. **کمی (ترتیبی)**
- ج) وضعیت تأهل این نماینده: متأهل **اسمی (کیفی)**
- چ) می‌گویند این نماینده در رأی‌گیری لایحه مرتبط با آموزش عمومی، هفتمین نفری است که از آن حمایت کرده است. **کمی (ترتیبی)**

آمار چیست؟

علوم تجربی نظیر کشاورزی و پزشکی نیازهای اساسی بشر را رفع می‌کنند. در این علوم یقین کامل برای حل مسئله وجود ندارد، بسیار کاربردی‌اند و مورد استفاده قرار می‌گیرد. یک پزشک معمولاً به صورت دقیق بیماری را تشخیص نمی‌دهد و داروی بیماری نیز به همین وضع دچار است. ولی پزشک بر اساس تجربه حکم به بیماری می‌دهد و دارو تجویز می‌کند و در اکثر مواقع نیز نتیجه می‌گیرد. از سوی دیگر در علوم ریاضی روابط به صورت صد درصد حاکم هستند و هیچ شک و شبهه‌ای به آن راه ندارد. وجود رابطه بین برخی پدیده‌ها در علوم تجربی باعث کشف حقایقی است که موجب پیشرفت آنها می‌شود. معمولاً روابط حاکم بر علوم تجربی را نمی‌توان به صورت ریاضی بیان کرد. علم آمار راهی برای بیان ریاضی چنین پدیده‌هایی است.



فعالیت

به نظر شما یک شهروند در زندگی روزمره خود از اطلاعات بیان شده در دو مثال زیر، چه استفاده‌ای می‌کند؟ این اطلاعات در رسانه‌ها منتشر شده‌اند.

فصل ۱ - آمار و جمعیت
چهار آمار در زمینه‌های مختلف است.

■ سلامت

- بیشترین آسیب‌دیدگی در منازل، افتادن یا زمین خوردن است.
- پنج درصد افراد به واکنش آنفولانزا واکنش شدید نشان می‌دهند.
- افراد سیگاری دو برابر دیگران در معرض سرطان قرار دارند.
- مصرف روزانه ۵ نوع میوه و سبزی بیشتر ویتامین‌های مورد نیاز بدن را تأمین می‌کند.

حمل و نقل

- اولین دلیل تماس با امداد خودرو فراموش کردن کلید داخل خودرو است.
 - بیشتر راننده‌های مرد و زن یک کشور خارجی، در پشت چراغ قرمز به ترتیب با بینی خود کلنجار می‌روند و در آینه نگاه می‌کنند.
 - راندن بین خطوط راهنمایی در اتوبان‌ها باعث کاهش ۳۰ درصدی شدآمد می‌شود.
 - متوسط تعداد کشته‌های تصادفات حدود ۵۰ نفر در روز اعلام شده است.
- به کمک روش‌های آمار و اطلاعات*
فصل ۱ - آمار و جمعیت
- چگونه این اطلاعات را به دست آورده‌اند؟ آیا تعریف دقیق کلمات پررنگ شده را حدس می‌زنید؟ جامعه، و پارامتر یا نمونه و آماره را در هریک از مثال‌ها مشخص کنید.

تصمیم‌گیری یکی از مهم‌ترین جنبه‌های زندگی ماست. ما بر اساس اطلاعاتمان و ارزش‌هایمان تصمیم‌گیری می‌کنیم. روش‌های آماری برای بررسی این اطلاعات به ما کمک می‌کنند. به علاوه، آمار در شرایطی که با عدم

تهیه کننده:

قطعیت در تصمیم‌گیری روبه‌رویم، هم به یاری ما می‌آید. چگونه به اطلاعات گزارش‌شده در فعالیت قبل دست‌یافته‌اند؟ به‌عنوان مثال، اگر درصدد برآورد نسبت افرادی هستیم که به واکسن آنفلوآنزا واکنش شدید نشان می‌دهند، بدون تزریق بر روی همهٔ افرادی که مایل به انجام آن هستند، آمار روش‌های مناسبی را پیش رویمان می‌گذارد. روش‌های آماری ما را قادر می‌سازند تا با نگاه کردن به اطلاعات به دست آمده از مجموعهٔ کوچکی از افراد یا اقلام، برای گروه‌های بزرگ‌تری از این افراد یا اقلام تصمیم‌گیری کنیم. شیوهٔ تحلیل داده‌ها، در کنار قواعد استنباطی، موضوعات اصلی مطالعات آماری را تشکیل می‌دهند.

شعارهای برتر
سال جهانی آمار



آمار، جهت‌نمای
توسعهٔ پایدار

آمار، راهنمای ما
در تصمیم‌گیری
و برنامه‌ریزی صحیح

به مطالعهٔ نحوه گردآوری، سازمان‌دهی، تحلیل و تفسیر داده‌ها برای استخراج اطلاعات و تصمیم‌گیری، آمار گفته می‌شود.

گفتنی است که روش‌های آماری به‌تنهایی نمی‌توانند معجزه کنند؛ این روش‌ها می‌توانند به ما کمک کنند تا تصمیم‌هایی بگیریم؛ اما نه هر تصمیمی. به یاد داشته باشید که حتی یک روش آماری مناسب، نمی‌تواند دقیق‌تر یا صحیح‌تر از داده‌ها و حقایق اصلی باشد. در نهایت، نتایج آماری باید توسط فردی که نه تنها روش‌ها، بلکه موضوع مورد بحث را کامل درک کرده باشد، تفسیر شود.

تمرین

۱. فرق بین داده و متغیر چیست؟
۲. داده‌های در سطح اسمی، کتی 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 اندازهٔ 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 وعدهٔ غذایی (ناهار یا شام) غذا سفارش داده‌اند؟

الف) متغیر را مشخص کنید. **همیشه ناهار ۵ شام ۵**

ب) این متغیر کتی است یا کیفی؟ **کیفی**
پ) جامعهٔ آماری در اینجا چیست؟ **کل مشتریان غذاخوری که گیاه‌خوار هستند.**

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

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

الف) مدت زمان پاسخ‌گویی به سوالات یک امتحان *نسبی*

ب) زمان اولین کلاس *فاصله‌ای*

پ) رشته تحصیلی *اسمی*

ت) مقیاس ارزیابی تحصیلی: ضعیف، معمولی و خوب *ترتیبی*

ث) نمره آخرین آزمون (از ۱۰۰ امتیاز) *نسبی*

ج) سن دانش‌آموز *نسبی*

لقمان و مرد پیاده

خواندنی

روزی لقمان در کنار چشمه‌ای نشسته بود. مردی که از آنجا می‌گذشت از لقمان پرسید: چند ساعت دیگر به ده بعدی خواهیم رسید. لقمان گفت: راه برو، آن مرد پنداشت که لقمان نشنیده است. دوباره سؤال کرد: مگر نشنیدی؟ پرسیدم: چند ساعت دیگر به ده بعدی خواهیم رسید؟ لقمان گفت: راه برو، آن مرد پنداشت که لقمان دیوانه است. برای همین راه خود را گرفت و رفت. زمانی که چند قدمی راه رفت، لقمان به بانگ بلند گفت: ای مرد، یک ساعت دیگر بدان ده خواهی رسید. مرد گفت: چرا اول نگفتی؟ لقمان گفت: چون راه رفتن تو را ندیده بودم، نمی‌دانستم تند می‌روی یا کند. حالا که دیدم دانستم که تو یک ساعت دیگر به ده خواهی رسید. در این داستان ساده و قدیمی تمام اصول آماری رعایت شده است. چرا؟ نکته ظریف این داستان این است که لقمان فقط می‌گوید، راه برو و توضیح دیگری نمی‌دهد. لقمان نمی‌گوید که می‌خواهم راه رفتن تو را ببینم تا از روی آن بگویم چه مدت طول می‌کشد تا به ده برسی، زیرا لقمان فکر می‌کند این اطلاع ممکن است در راه رفتن آن مرد اثر بگذارد و در نتیجه سرعتی که لقمان تخمین می‌زند، سرعت واقعی راه رفتن آن فرد نباشد و در نتیجه زمانی را که تخمین خواهد زد، مدت زمان دقیقی نباشد.

درس ۲

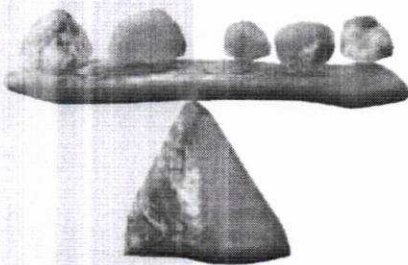
معیارهای گرایش به مرکز

فعالیت



امسال خیرین مدرسه چقدر به مدرسه کمک می‌کند؟ مدیر مدرسه بر اساس اطلاعات سال‌های گذشته خود می‌گوید: معمولاً خیرین، به‌طور متوسط، ۱۰ درصد درآمد سالانه خود را به این امر اختصاص می‌دهند. فرض کنید درآمد ماهیانه حضار در انجمن خیریه این دبیرستان در سال جاری به ترتیب حروف الفبا به‌صورت زیر باشد:

درآمد (میلیون ریال)	نجمیه	سبحان	رسول	حسنا	جوانه	احمد	آرمان
۴۰	۱۲	۲۸	۳۲	۳۰	۲۲	۲۵	



س برای پاسخ به سؤال طرح شده باید میانگین این اعداد را محاسبه کنیم. میانگین همان چیزی است که ما به آن معدل می‌گوییم. برای محاسبه آن شما تمامی درآمدها را باهم جمع و بر تعداد افراد حاضر تقسیم می‌کنید. اگر تمامی درآمدها را باهم جمع کنید (۱۸۹)، و بر تعداد افراد تقسیم کنید (۷)، به عدد میانگین ۲۷ میلیون ریال در ماه می‌رسید. در نتیجه ۱۰ درصد درآمد سالانه برابر است با ۳۲/۴ است.

اگر n داده به صورت $x_1, x_2, \dots, x_n$ داشته باشیم، میانگین آنها را با نماد $\bar{x}$ نشان می‌دهیم، که به صورت زیر تعریف می‌شود:

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

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

کار در کلاس



$$\bar{x} = \frac{10 + 20 + 30 + 40 + 50}{5} =$$

۱. میانگین داده‌های ۱۰، ۲۰، ۳۰، ۴۰، ۵۰ چقدر است؟

۲. اگر میانگین داده‌های ۱۰، ۴۰، ۳۰ شود مقدار x چقدر است؟

۳. میانگین اعداد ۱، ۲، ۳، ۴، ۵ چقدر است؟

۴. میانگین اعداد ۲، ۴، ۶، ۸، ۵ چقدر است؟

۵. میانگین اعداد ۵، ۶، ۷، ۸، ۹ چقدر است؟

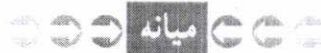
۶. میانگین اعداد ۵۰، ۶۰، ۷۰، ۸۰، ۹۰ چقدر است؟

۷. آیا می‌توانید چند قاعده کلی از تمرین‌های قبل درباره خواص میانگین ذکر کنید.

میانگین
جمع داده‌ها
تعداد داده‌ها

حال اگر یک میلیارد در با درآمد ماهیانه یک میلیارد ریال به انجمن خیریه دبیرستان ما بیايد، 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.

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



در مثال خیریه، داده دور افتاده باعث اشتباه ما در تخمین متوسط داده‌ها شد. میانگین مرسوم‌ترین معیار گرایش به مرکز است که گاهی ممکن است ما را به اشتباه بیندازد، ولی می‌توان از معیار دیگری نیز برای بیان متوسط درآمد استفاده کرد. برای این منظور از میانه که داده وسطی داده‌های مرتب شده است، می‌توان استفاده کرد. پس از مرتب کردن داده‌ها، داده‌ای که تعداد داده‌های بعد از آن با تعداد داده‌های قبل از آن برابر است، میانه است. اگر تعداد داده‌ها زوج باشد، میانه برابر میانگین دو داده وسطی مرتب شده است.

برای محاسبه میانه داده‌ها، قبل از ورود میلیارد، داده‌ها را از کوچک به بزرگ مرتب می‌کنیم. وسط (میانه) ۷ عدد چهارمین عدد است، پس میانه داده‌ها عدد ۲۸ است که با میانگین داده‌ها تفاوت زیادی ندارد.

ردیف	۱	۲	۳	۴	۵	۶	۷
درآمد مرتب شده	۱۲	۲۲	۲۵	۲۸	۳۰	۳۲	۴۰

حال فرض کنید میلیارد در خیریه حضور دارد. میانه برای ۸ عدد وجود ندارد یا به عبارتی ۸ عدد، عدد وسط ندارند. پس بین داده چهارم و پنجم میانگین می‌گیریم. جایگاه چهارم درآمد ۲۸ و جایگاه پنجم آن ۳۰ و در نتیجه میانه جدید درآمد ۲۹ است.

مد، داده‌ای است که بیشترین فراوانی را دارد. فروشندگان پوشاک از معیار گرایش به مرکز مد بسیار استفاده می‌کنند، آنها با آمارگیری‌های خود درمی‌یابند که چه نوع پوشاکی مورد پسند مصرف کنندگان است. از همان پوشاک برای فروش سفارش می‌دهند. در رأی‌گیری‌ها، اساس تصمیم‌گیری مد است، چون موضوعی که بیشترین فراوانی را داشته باشد، انتخاب می‌شود. در انتخاب رئیس جمهور، نامزدی انتخاب می‌شود که بیشترین فراوانی (رأی) را داشته باشد. برای محاسبه مد فقط کافی است فراوانی داده‌ها را با هم مقایسه کنیم و داده با بیشترین فراوانی مد است. مد ممکن است منحصر به فرد نباشد.

بیشتر افراد معتقدند که عدد ۲۹ میلیون ریال گویای مقدار صحیح متوسط درآمد افراد است و عدد ۱۴۸ میلیون ریال کاملاً بی ربط است. پس ما به یک نتیجه می‌رسیم: اگر در داده‌هایمان، دور افتاده وجود داشت - دقیقاً مانند زمانی که یک میلیارد در هوس کار خیر می‌کند - باید از میانه استفاده کنیم.



تمرین

اینجا داده‌ها را مرتب می‌کنیم.

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

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

نکته اصلی در اینجا است: اگر در آمار در جایی به یک نتیجه شسته و رفته برخوردید، خیلی احتیاط

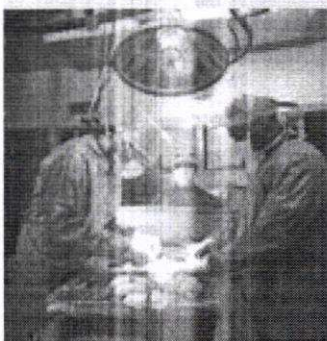
کنید. به دست آوردن «یک استنباط درست از داده‌ها تنها چیزی نیست که شما در یک بررسی آماری

خواهان آن هستید: خواسته دیگر ما برنامه‌ریزی و تصمیم‌گیری است».

تصور کنید که به جای انجمن خیریه با یک بیمارستان سر و کار داریم و داده‌ها به جای مبالغ درآمد،

هزینه‌های جراحی هستند. فرض کنید به جای میلیارد، با یک بیمار روبه‌رویم که گرفتار بعضی از مشکلات

بعد از عمل است و مجموع هزینه‌هایش بالغ بر ۲۵۰ میلیون ریال است.



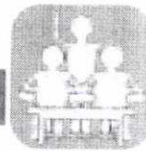
بیمار	هزینه‌های جراحی میلیون ریال	هزینه‌های جراحی مرتب شده
الف	۲	۱
ب	۵	۲
پ	۴۵	۳
ت	۷	۴
ث	۳۵	۵
ج	۳۰	۶
ح	۵۰	۷
ز	۲۵۰	۸

میانه مناسب‌تر است.

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

میانگین هزینه‌های جراحی این بیمارستان برابر ۵۳ میلیون ریال و میانه آن برابر ۳۲/۵ میلیون ریال است. اگر شما مدیر این بیمارستان بودید، کدام عدد برایتان از همه مهم‌تر بود؟ عدد ۳۲/۵ میلیون نشان دهنده هزینه‌های معمول نگهداری یک بیمار است، اگر بخواهیم بودجه سال بعد را با فرض این عدد بنویسیم، به احتمال زیاد با کسر بودجه مواجه می‌شویم. با قدری تفکر به میانگین و میانه، دلیل اینکه افراد از بیمه تأمین اجتماعی استفاده می‌کنند، روشن می‌شود؛ (میانه هزینه‌های سالیانه درمان معمولاً بسیار کمتر از حق بیمه پرداختی است، اما با نگاه به میانگین سالیانه هزینه‌ها، حساب می‌کنم که معامله خوبی کرده‌ام) و من همیشه کمربند ایمنی را می‌بندم، اگرچه میانه تعداد زخمی‌ها در هر سفر با خودرو صفر باشد. این نتیجه را نیز به یاد داشته باشید: گاهی هیچ آمار درست یا غلطی وجود ندارد، و همه چیز به اینکه شما چرا می‌خواهید از آنها استفاده کنید، بستگی دارد.



کار در کلاس

- با سه واژه: داده دورافتاده، میانه، و میانگین، ۵ جمله زیر را کامل کنید.
 - آنچه اکثر مردم «حد وسط» می‌نامند، نزد آمارشناسان به **میانه** معروف است. برای محاسبه **میانه**... به داده‌های خود به عنوان فهرستی از اعداد نگاه کنید؛ همه اعداد را باهم جمع کنید و بر تعدادشان تقسیم کنید.
 - میانه**... در واقع نقطه وسطی فهرست اعداد مرتب‌شده است. نیمی از اعداد مقادیر بالاتر از **میانه**... و نیمی دیگر مقادیر پایین‌تر از **میانه**... قرار دارند.
 - زمانی با **داده دورافتاده** مواجه هستیم که مشاهده‌ای داشته باشیم که از الگوی داده‌هایمان پیروی نکند.
 - وقتی با **داده دورافتاده** مواجه هستید، **میانه**... معمولاً بازتاب بهتری از داده‌ها می‌دهد تا **میانگین**...
 - به‌طور کلی، برای برنامه‌ریزی و تصمیم‌گیری، **میانه**... بهتر از **میانگین**... است.

شخصی دست راست خود را در یخچال و دست چپ خود را در فر کرده است. زمانی که از او درباره احساسش سؤال می‌شود پاسخ می‌دهد: «به‌طور متوسط خوب هستم». نتیجه اخلاقی این لطیفه این است که یک عدد به تنهایی، معمولاً مجموعه‌ای از داده‌ها را به خوبی توصیف نمی‌کند. بنابراین، اندیشه خوبی است که علاوه بر معیار گرایش به مرکز معیاری را نیز گزارش دهیم که نوسان و تغییر داده‌ها را هم بیان کند.

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

میانگین موزون:

شما قطعاً شنیده‌اید که برخی از نمرات شما ضریب دارند. مثلاً اگر چهار بار از شما آزمون گرفته باشند، نمره آخرین آزمون را ممکن است در ۳ ضرب کنند، یعنی به آن ضریب ۳ بدهند. پس اگر نمرات شما به صورت زیر باشند:

آزمون اول	دوم	سوم	چهارم
۱۶	۱۵	۱۸	۱۷

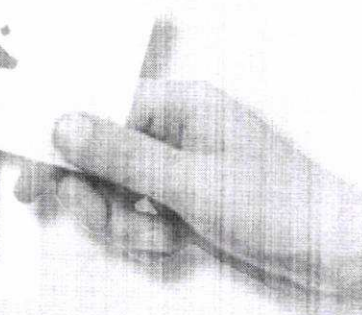
نمره‌ای که برای این درس در کارنامه شما خواهد آمد میانگین ۶ عدد است یعنی نمره درس مذکور عبارت است از:

$$\frac{16+15+18+17+17+17}{6} = 16.66$$

میانگین نمره‌ها بدون احتساب ضریب ۱۶/۵ می‌شود.

از میانگین موزون برای معدل دروسی که با واحدهای مختلف (تعداد ساعات متفاوت) در طول سال تحصیلی ارائه می‌شود نیز استفاده می‌شود.

خواندنی



درس ۳

معیارهای پراکندگی

فعالیت



می‌خواهیم کلاس‌های دهم سه دبیرستان را بر اساس نتایج آزمون جامعی که هم‌زمان بین دانش‌آموزان ممتاز برگزار شده است، رتبه‌بندی کنیم. از هر دبیرستان ۱۰ نفر به تصادف انتخاب شده‌اند. نمرات آزمون جامع هر سه کلاس از ۱۲۰ نمره است.

داده‌های مربوط به سه کلاس دهم:

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

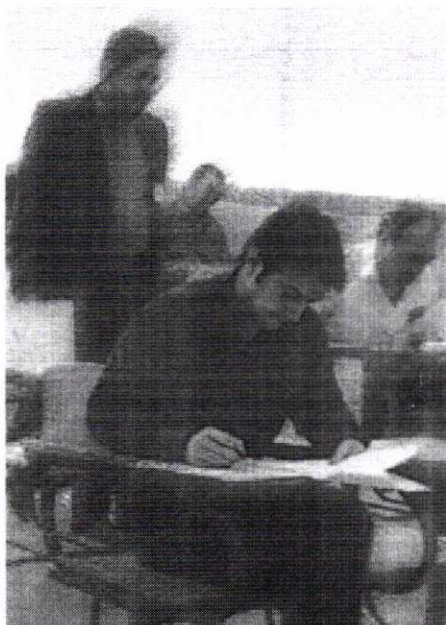
$$\bar{x} = 67$$

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

$$\bar{y} = 67$$

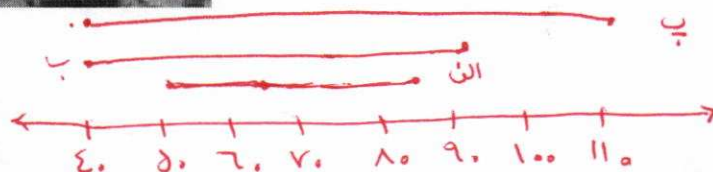
پ) {۴۳, ۵۱, ۵۳, ۱۱۰, ۵۰, ۴۸, ۸۷, ۶۹, ۶۸, ۹۱}

$$\bar{z} = 67$$



میانگین سه کلاس را محاسبه کنید. به نظر شما پراکندگی نمرات در کدام کلاس بیشترین و در کدام کلاس کمترین است؟ برای پاسخ به این سؤال، داده‌ها را روی سه محور موازی نمایش دهید. کدام دبیرستان نتایج بهتری با اطلاعات داده شده کسب کرده است؟ چرا؟ مثلاً اگر والدین برای ثبت نام فرزند بازگوش خود در سال آینده بخواهند بر اساس این نتایج یکی از مدرسه‌ها را انتخاب کنند، مدرسه‌ای بهتر است که پراکندگی نمرات در آن کمتر است در مقابل برای فرزندی که به قصد شرکت در المپیاد می‌خواهد مدرسه را انتخاب کند، مدرسه‌ای بهتر است که پراکندگی نمرات در آن بیشتر باشد. چرا؟

۱۱
۱۰
۹
۸
۷
۶
۵
۴



بهترین دانش‌آموز است.

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

یک معیار معمول سنجش میزان پراکندگی که نشان دهنده تغییرات داده‌های مورد مطالعه است، انحراف معیار است. انحراف معیار را با به کارگیری یک فرمول از داده‌ها به دست می‌آوریم (اختلاف هر عدد با میانگین داده‌ها را حساب کنید؛ به توان ۲ برسانید؛ میانگین اعداد حاصل را به دست آورید؛ و سپس جذر بگیرید.) توان دوم انحراف معیار را واریانس می‌نامند. داده‌ها در شکل زیر رسم شده‌اند. همچنین انحراف معیار نمرات هر سه کلاس محاسبه و در شکل نشان داده شده است.

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

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

کار در کلاس



۱. اگر داده‌ها با هم برابر باشند، انحراف معیار برابر با صفر خواهد شد؛ چرا؟
۲. انحراف معیار نمرات آزمون جامع سه کلاس را می‌توانید با تکمیل جدول زیر محاسبه کنید.

مشاهدات	انحراف مشاهدات از میانگین	انحراف مشاهدات از میانگین به توان ۲
۶۵	۶۵ - ۶۷	$(-۲)^2$
۷۵		
۷۳		
۸۰		
۶۰		
۶۴		
۶۹		
۶۲		
۶۷		

مشاهده می‌شود که انحراف معیار نه تنها نظر شما در خصوص کلاس با بیشترین پراکندگی نمرات را تأیید می‌کند، بلکه به شما امکان مقایسه کتی آنها را نیز می‌دهد. حال می‌توانید درستی پاسخ‌های خود را بررسی کنید.

۸۵

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

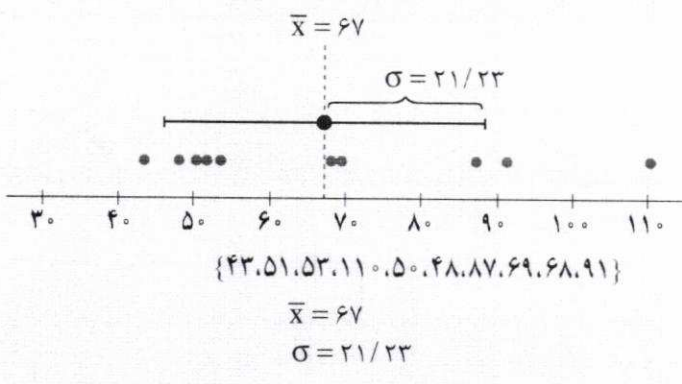
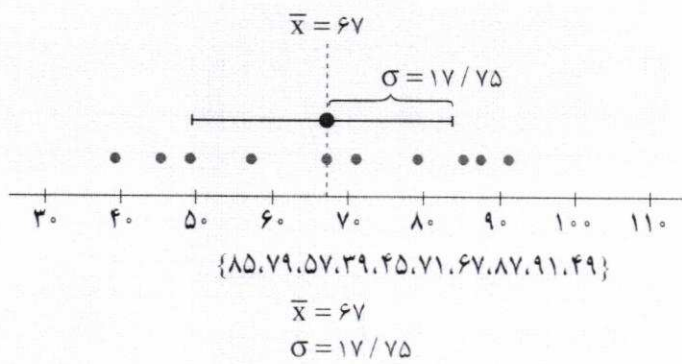
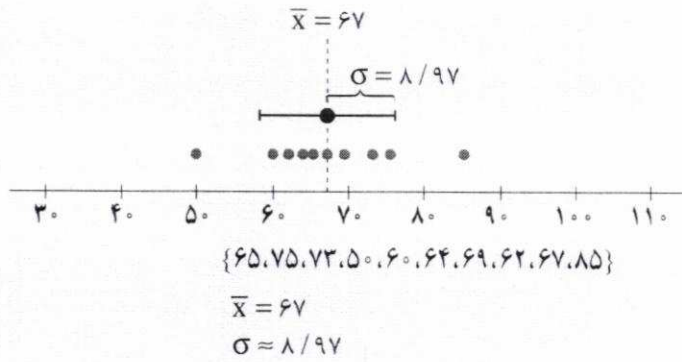
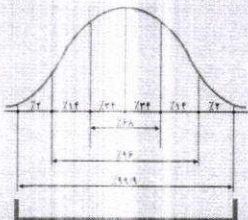
جدول باید
در صفر
دارد به
که ایراد دارد.



کلاس الف

پراکندگی در خم بهنجار
(منحنی نرمال)

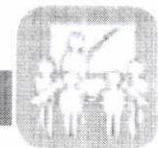
اگر تعداد داده‌ها زیاد باشد، بافت‌نگاشت (هستوگرام) آنها را می‌توان به صورت یک خم تصور کرد. اگر این خم به صورت یک «زنگ» باشد، به آن خم بهنجار گفته می‌شود. خداوند بسیاری از پدیده‌ها را بهنجار آفریده است. مثلاً طول قد یا وزن دانش‌آموزان بهنجاراند. یعنی بافت‌نگاشتی زنگی شکل دارند. خواص این خم کمک زیادی برای تصمیم‌گیری بر اساس داده‌ها یا کسب اطلاعات از آنها می‌کند. به یاد دارید که گفته شد اگر σ انحراف معیار داده‌ها باشد چند درصد داده‌ها به طور تقریبی بین میانگین که در واقع گرانگه یا مرکز نقل بافت‌نگاشت است و ضرابی از انحراف معیار قرار دارد. این اطلاعات را به طور مختصر در شکل زیر ملاحظه می‌کنید (مقادیر روی نمودار تقریبی هستند):



کلاس ب

کلاس ج

فعالیت



حتماً بر روی قوطی‌ها یا پاکت‌های مواد غذایی یا بهداشتی اعدادی به صورت زیر مشاهده کرده‌اید. به نظر شما هر یک از این اعداد چه چیزی را نشان می‌دهند.

250 ± 10 گرم

عدد ۲۵۰ وزن خالص ماده غذایی است که قرار است به دست مصرف کننده برسد. ولی معمولاً این چنین نیست. کمی خطا اجتناب ناپذیر است. این خطا چقدر است و چگونه آن را محاسبه کنیم؟ اگر تعداد زیادی از محتوای پاکت‌ها یا قوطی‌های یک تولیدی را وزن و انحراف معیار آنها را محاسبه کنیم، دو برابر آن عدد ۱۰ است. یعنی عدد ۱۰ دو برابر انحراف معیار مواد غذایی در بسته‌بندی‌های مختلف است. به عبارت دقیق‌تر اگر ما ۱۰۰ قوطی حاوی این ماده غذایی را وزن کنیم، انتظار داریم حداقل ۹۶ تا از آنها وزنی بین ۲۴۰ تا ۲۶۰ گرم داشته باشند. درج چنین اعدادی بر روی کالاها از الزامات استاندارد هر کشوری است. چگونه به عدد ۹۶ رسیده‌اند؟

مجموع سؤالات متن حاشیه صفحه ۱۰۵

برای پی بردن به میزان تغییراتی که داریم، معمولاً با محاسبه مقدار انحراف از میانگین اطلاعات دقیق‌تری در خصوص میانگین و انحراف معیار به دست خواهیم آورد. برای این منظور از یک قانون مشهور در آمار استفاده می‌کنیم: تقریباً ۹۶ درصد از داده‌ها بین «دو انحراف معیار از میانگین» هستند. این جمله یعنی اگر بازه‌ای تعریف کنیم که ابتدای آن $\bar{x} - 2\sigma$ و انتهای آن $\bar{x} + 2\sigma$ باشد، تقریباً ۹۶ درصد داده‌ها در این فاصله قرار دارند این مانند آن است که بگوییم، ۴ درصد نمره‌ها بیش از دو برابر انحراف معیار از میانگین فاصله دارد. البته مشابه همین قانون در شکل‌های قبل برای یک برابر انحراف معیار نمایش داده شده است. در این حالت قانون قبل به «تقریباً ۶۸ درصد از مشاهدات بین یک انحراف معیار از میانگین هستند»، تغییر می‌کند. دو برابر انحراف از میانگین را روی همان شکل رسم کنید.

بازه سنی چیست؟

کار در کلاس



جدول زیر را تکمیل کنید.

$(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$	تقریباً ۹۹/۹ درصد از مشاهدات بین سه برابر انحراف معیار از میانگین قرار دارند
$(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$	تقریباً ۹۶ درصد مشاهدات بین دو برابر انحراف معیار از میانگین قرار دارند
$(\bar{x} - \sigma, \bar{x} + \sigma)$	تقریباً ۶۸ درصد مشاهدات بین یک برابر انحراف معیار از میانگین قرار دارند

فعالیت



می‌خواهیم با همان مثال خیریه درس قبل بحث را ادامه دهیم. اگر انحراف معیار مجموعه داده‌ها 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 می‌شود:

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

به کمک جدول زیر این کار را انجام دهید.

انحراف منهای از میانگین		انحراف منهای از میانگین		انحراف منهای از میانگین	
قبل از ورود میلیارد	بعد از ورود میلیارد	قبل از ورود میلیارد	بعد از ورود میلیارد	قبل از ورود میلیارد	بعد از ورود میلیارد
۴۰	۴۰				
۱۲	۱۲				
۲۸	۲۸				
۳۲	۳۲				
۳۰	۳۰				
۲۲	۲۲				
۲۵	۲۵				
	۱۰۰۰				

۸,۵۸

زمانی که میلیارد وارد انجمن خیریه می شود، انحراف معیار داده ها برابر است با ۱۰,۱۸. این دو مقدار با هم اختلاف زیادی دارند. چرا و چه باید کرد؟ زمانی را به خاطر بیاورید که ما معیار گرایش به مرکز را برای این داده ها محاسبه کردیم. ما برای رفع آن مشکل به جای میانگین، میانه را به کار بردیم. مشابه همان کار را می توانیم انجام دهیم. معیار پراکندگی که تعریف می شود، انحراف معیار نیست. ما آن را دامنه میان چارکی، که با IQR نشان می دهیم، می نامیم. «میانه» جایی در وسط داده ها است؛ به طور مشابه، چارک ها هم یک چهارم و سه چهارم مشاهدات هستند. به عبارت ساده تر اگر برای داده های مرتب شده قبل از میانه، یک میانه حساب کنیم، همان چارک اول است و به همین صورت میانه داده های بعد از میانه، چارک سوم خواهد بود. برای سادگی از نمادهای Q_1 ، Q_2 و Q_3 به ترتیب برای چارک اول، میانه (چارک دوم)، و چارک سوم استفاده می شود.

جدول زیر را برای مثال خیرین مدرسه تکمیل کنید.

Q_1	Q_2	Q_3	$IQR = Q_3 - Q_1$	σ
چارک اول	میانه	چارک سوم	دامنه میان چارکی	انحراف معیار
			۱۰	
			۱۲,۵	

درآمد ماهیانه
اعضای خیریه
درآمد بعد از
ورود میلیارد

این اعداد امکان مشاهده چند چیز را به شما می دهند: میلیارد در هم 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 بگیرید که داده دور افتاده وجود داشته است.

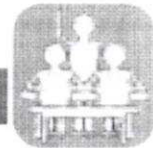
شعارهای برتر
سال جهانی آمار



آمارهای دقیق
بشنوانه توسعه پایدار
آمار، ابزاری علمی
در توسعه مبتنی بر
داناتی

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 بیمارشان زود تشخیص داده شده و در مراحل ابتدایی است و چند نفر بیمارشان 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 اینجا است که، آنها این کار را نخواهند کرد. به ندرت می‌توانید یک پزشک متخصص سرطان بسیار پر مشغله 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. ... میانگین ... و ... میانه ... 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 **انحراف از میانگین** است.
 ۳. معیار پراکندگی که معمولاً با میانگین بیان می‌شود، ... **انحراف معیار** است.
 ۴. معیار پراکندگی که معمولاً با میانه بیان می‌شود **دامنه میان چارکی** دارد. **IQR**
 ۵. **میانه**، ... و **دامنه میان چارکی** اطلاعات سریعی درباره داده‌ها بدون نیاز به هرگونه محاسبه می‌دهند.
 ۶. آماری که برای توصیف یک مجموعه داده، میانگین‌ها و میانه‌ها، انحرافات معیار و دامنه‌های میان چارکی به کار می‌رود، **آمار توصیفی** دارد.
 ۷. 50 درصد داده‌ها قبل از ... **میانه** ... و 50 درصد داده‌ها بعد از ... **میانه** ... قرار دارند.
 ۸. 75 درصد داده‌ها قبل از ... **چارکی اول** ... یا بعد از ... **چارکی اول** ... قرار دارند.
 ۹. 25 درصد داده‌ها قبل از ... **چارکی اول** ... یا بعد از ... **چارکی دوم** ... قرار دارند.
 ۱۰. 50 درصد داده‌ها بین ... **چارکی اول** ... و ... **چارکی سوم** ... قرار دارند.

تهیه کننده:

شایسته است تمرینی انجام شود که در آن لازم است دانش
میان چارگی نیز محاسبه شود.

دو آنرا مقایسه

۱۱. تقریباً ۹۶ درصد مشاهدات در فاصله ... از میانگین هستند.

۱۲. تقریباً ۶۸ 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 می کنید. اگر بخواهید اعضای انتخابی این نمونه تصادفی، تکراری نباشند، چگونه این کار را انجام می دهید؟ **استداده از تاس پرتاب مجدد در صورت تکرار**

۲. سؤال ۱ را برای اعداد ۱ تا ۳۶ تکرار کنید. (راهنمایی: می توانید از پرتاب دو تاس به عنوان روشی برای انتخاب اعداد ۱ تا ۳۶ استفاده کنید) **دو تاس همزمان پرتاب می کنیم.**

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 ت) از طریق پرسش نامه

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 باسگاه 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 تناسبات تولیدی یک کارخانه، شامپوها را صد در میان مورد آزمایش قرار دادیم.

۵. 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 کتاب های غیر درسی می کنند، آمارگیری کنیم.

الف) در این آمارگیری جامعه را مشخص کنید.

این تمرین به
دو تمرین جداگانه
تبدیل می شود.

شعارهای برتر
سال جهانی آمار



آمار دقیق،

برنامه ریزی بهتر،

آینده روشن



آمار سیمای واقعیت،

ابزار مدیریت

حل تمرین ۲ به بعد پشت برگه

۳

اسامی آنها را روی کاغذهای یکسان می نویسیم و سپس به قید قرعه ۴ نفر آنها را انتخاب می کنیم .
کمی روشن نیز استفاده از اعداد تصادفی است .

بقیه تمرین ۳ که به نظری رسیدیم این تمرین مرتبط است .
(الف) پرسش نامه (ب) پرسش نامه (پ) مشاهده (آزمایش)
(ت) دادگان (ث) آزمایش (مشاهده)

۴

(الف) بررسی علل موفقیت مدیر (برساک) - بررسی عوامل موفقیت یک فرد در تکنولوژی سراسری
(ب) بررسی میزان استفاده از کمربند ایمنی در جاده ها توسط رانندگان
بدون زقار و زدن آموزن در هنگام وقت استراحت آنها
(پ) بررسی میزان تصادف در یک نقطه ی خاصی در شهر - بررسی عزات در آموزن رشته ی ریاضی در دهه ۷۰
(ت) نظر خواهی در مورد (برای) آموزشگاه توسط مدیر - نظر خواهی در مورد رضایت مشتریان از برخورد کارکنان
فروشگاه

۵

نمونه	جامعه	واحد آماری	روش آماری	
—	مجموعی کل پرسش ها پاسگاه	هر یک از پرسش ها پاسگاه	ایجاد عمل سرشماری است و نمی تواند عنوان گیری باشد.	(الف)
بچه ها که با آنها مواجهه شد.	کل بچه ها وارد شده به پارک	هر یک از بچه ها وارد شده به پارک	عنوان گیری تصادفی (منظم، سیستماتیک)	(ب)
شامپوها بر ساشه	کل شامپوها تولیدی	هر یک از شامپوها	عنوان گیری تصادفی (منظم)	(ج)

۶

(الف) مجموعی کل ساکنین استان مازندران متقدما - تهیه ی پرسش نامه و توزیع آن بین تعداد معینی از
خانوارهای این شهرها انتخاب (تصادفی) که بلاک متوالی آنها
مضرب ۳ باشد .

(ب) مجموعی کل مراکز کشور که شامپو مصرف می کنند .
تهیه ی پرسش نامه و توزیع آن به متوالی شامپوهای تکمیل

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

(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.

(ج) 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 اتومبیل های 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

(ج) سن دانشجویان شرکت کننده در یک دوره هنری

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

سنت	مدیر شرکت	معاون	حسابدار	بازاریاب	۲ نفر فروشنده	منشی	مسئول
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 ۲۵۷ ریال است. آیا این رقم می تواند دلیلی برای پرداخت حقوق بالاتر از 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:

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

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

(ب) کدام یک از شاخص های 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 نمره ای باید کسب کند؟

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

بازار ۱	۱۰	۱۱	۱۰	۹	۱۲	۱۱	۱۰	۱۲	۹	۱۳	۸
بازار ۲	۱۰	۹	۱۱	۱۰	۱۱	۹	۱۰	۱۲	۱۰	۱۳	۸

$$S_1, S_2 = ?$$

(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 می دهید از کدام بازار خرید کنید؟ چرا؟

(ث) اگر داده ها را در اختیار نداشته باشید، فقط به صرف داشتن دامنه تغییرات می توانید تصمیم گیری کنید؟

قبل از خرید دامنه تغییرات را بررسی کنید.

بسیار کم فرد دارد.
داده دور افتاده دارد.

$$a = \text{نمره}$$

$$18 = \frac{17+19+18+a+18+19}{6}$$

$$\rightarrow a \geq 17$$

الف) مجموعی کل دانش آموزان کلاس

ب) انتخاب تعدادی از دانش آموزان به طور تصادفی (مثلاً دانش آموزان که شماره ردیف آنها در لیست کلاس عدد فرد باشند)

○ ساعت ۱

○ ساعت ۲

○ ساعت ۳

⋮

پ) پرسش نامه شامل تعداد ساعت معین قید شده

ت) تعداد دانش آموزان کلاس

ث) تعداد دانش آموزان انتخاب شده، هر چه جامعه کوچک باشد بهتر است سرشاری را جابجایی نمونه گیری کنیم.

ج) میزان ساعت مطالعه کتب غیر درسی

چ) کمی، کیفی، با مقیاس نسبی

ج) کیفی نسبی

چ) کیفی نسبی

تهیه کننده:

گروه ریاضی مقطع دوم متوسطه، استان خوزستان

الف) کیفی نسبی

ب) کیفی فاصله ای

پ) کیفی نسبی

ت) کیفی نسبی

ث) کیفی نسبی