

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

آمار و مدل سازی

سال سوم آموزش متوسطه

رشته علوم تجربی

وزارت آموزش و پرورش
سازمان پژوهش و برنامه‌ریزی آموزشی

برنامه‌ریزی محتوا و نظارت بر تألیف : دفتر تألیف کتاب‌های درسی ابتدایی و متوسطه نظری

نام کتاب : آمار و مدل‌سازی - ۲۵۸/۵

مؤلفان : شهرناز بخشعلی زاده، عین‌الله پاشا و آرش رستگار

آماده‌سازی و نظارت بر چاپ و توزیع : اداره کل نظارت بر نشر و توزیع مواد آموزشی

تهران : خیابان ایرانشهر شمالی - ساختمان شماره ۴ آموزش و پرورش (شهید موسوی)

تلفن : ۹-۸۸۸۳۱۱۶۱، دورنگار : ۰۹۲۶۶۰۸۸۳، کدپستی : ۱۵۸۴۷۴۷۳۵۹،

وبسایت : www.chap.sch.ir

رسام : هدیه بندار

صفحه‌آرا : خالد قهرمانی‌دهبکری

طراح جلد : محمدحسن معماری

ناشر : شرکت چاپ و نشر کتاب‌های درسی ایران : تهران - کیلومتر ۱۷ جاده مخصوص کرج - خیابان ۶۱ (داروبخش)

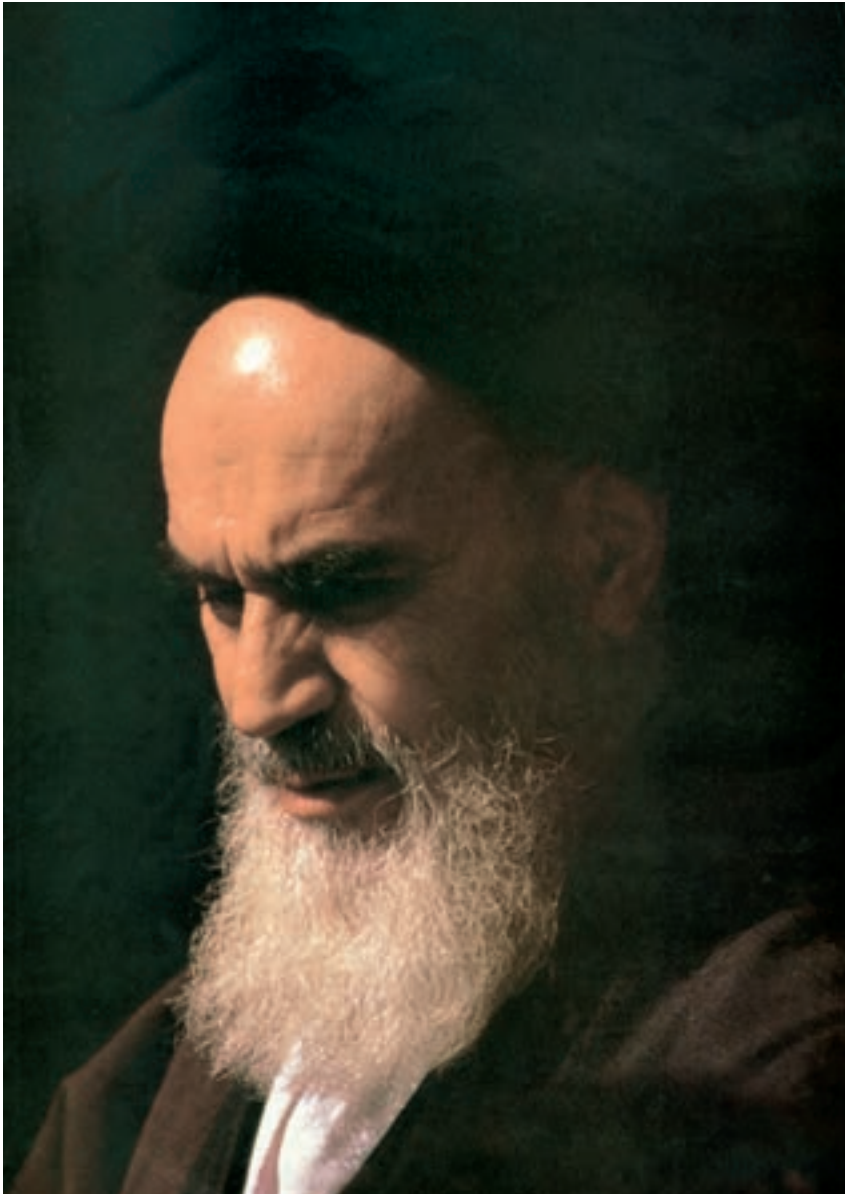
تلفن : ۵-۴۴۹۸۵۱۶۱، دورنگار : ۰۴۴۹۸۵۱۶، صندوق پستی : ۱۳۹-۳۷۵۱۵

چاپخانه : شرکت چاپ و نشر کتاب‌های درسی ایران «سهامی خاص»

سال انتشار و نوبت چاپ : چاپ هفدهم ۱۳۹۵

حق چاپ محفوظ است.

شابک ۵-۷۳۸-۰۵-۹۶۴ ISBN 964-05-0738-5

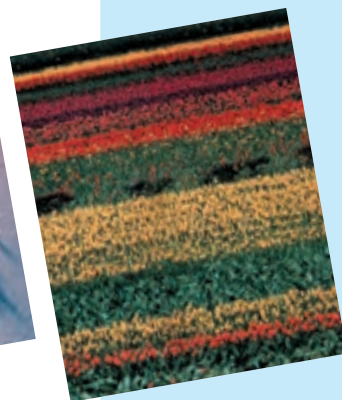


امام به بحث و مذاکره در مسایل علمی و بحث و مذاکره در مجلس درس
علاقه فراوانی داشتند. اگر بر جلسه درس سکوت حکمفرما می شد، می فرمودند:
مجلس درس است، نه مجلس روضه. سخن بگویید و اعتراض کنید.

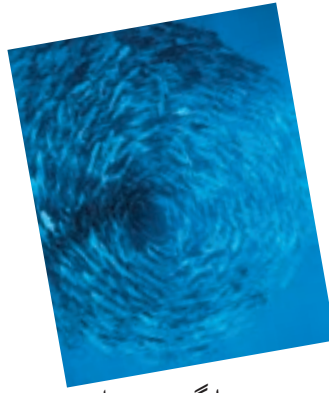
فهرست



۱	سخنی با خواننده
۳	فصل اول — اندازه گیری و مدل سازی
۴	اندازه گیری
۷	مدل سازی
۱۵	فصل دوم — جامعه و نمونه
۱۵	جامعه
۱۷	نمونه
۲۴	نمونه تصادفی
۲۷	روش های جمع آوری داده ها
۳۳	فصل سوم — متغیرهای تصادفی
۳۴	متغیر تصادفی
۳۷	انواع متغیرهای تصادفی
۴۳	فصل چهارم — دسته بندی داده ها و جدول فراوانی
۴۴	جدول فراوانی



- ۴۶ دسته بندی
- ۵۳ انواع فراوانی
- ۵۸ چند تذکر درباره تشکیل جدول فراوانی
- ۶۳ معرفی و آشنایی با نرم افزار آماری MINITAB
-
- ۷۷ فصل پنجم — نمودارها و تحلیل داده ها
- ۷۸ نمودار میله ای
- ۸۲ نمودار مستطیلی
- ۸۸ نمودار چندبر فراوانی
- ۹۲ نمودار دایره ای
- ۹۶ نمودار ساقه و برگ
- ۱۰۴ رسم نمودارها با استفاده از MINITAB
-
- ۱۱۳ فصل ششم — شاخص های مرکزی
- ۱۱۴ مد
- ۱۱۶ میانه
- ۱۲۰ نمودار جعبه ای
- ۱۲۵ میانگین



۱۳۰

میانگین وزن دار

۱۴۱

از ماشین حساب چگونه استفاده کنیم؟

۱۴۳

فصل هفتم — شاخص‌های پراکندگی

۱۴۴

پراکندگی

۱۴۵

دامنه تغییرات

۱۴۶

چارک‌های بالا و پایین

۱۴۸

واریانس

۱۵۳

انحراف معیار

۱۵۷

ضریب تغییرات

۱۶۲

پراکندگی در منحنی نرمال

۱۶۳

محاسبات شاخص‌های عددی در MINITAB

۱۶۷

فصل هشتم — همبستگی و رگرسیون

۱۶۷

جامعه‌های چند متغیری

۱۶۸

نمودار پراکنش

۱۷۳

همبستگی

۱۷۹

خط رگرسیون

۱۸۶

پروژه

۱۹۹

منابع و مراجع

سخنی با خواننده

می گویند هستی موضوع هیچ علمی به تنهایی نیست. این گفته بدین معنی است که پرداختن به زمینه های یک علم، فقط گوشه های بسیار کوچکی از هستی را برای ما روشن می کند. برای آن که دامنه شناخت و وسعت یابد، لازم است چند شاخه علمی به هم گره بخورد و از روش های یکی، در دیگری استفاده شود. ریاضیات به عنوان یک علم پایه ای و زیربنایی حضور خود را در بسیاری از علوم به اثبات رسانده است. آمار نیز از یک دیدگاه به عنوان علمی مستقل و از دیدگاهی دیگر به عنوان شاخه ای از ریاضیات، در علوم کاربردی جایگاه ویژه خود را دارد.

به علاوه در دنیای کنونی، سرعت دگرگونی به حدی است که عمر مفید اطلاعات به یک دهه هم نمی رسد. بنابراین انباشتن ذهن با اطلاعاتی که به سرعت رو به زوال است جز اتلاف وقت و انرژی حاصلی نخواهد داشت. به همین علت است که در برنامه های اخیر دولت، به جای انتقال دانش، بر تولید دانش تکیه شده است. برای آن که بتوانیم به این مقصود برسیم، لازم است روش ها و عادت های آموزشی خود را تغییر دهیم و به روش هایی توسل جویم که ما را در رسیدن به تولید دانش و شناخت بهتر از محیط اطراف، کمک کند. یکی از اساسی ترین این روش ها آن است که سطح توانایی خود را در مدل سازی پدیده های مورد مطالعه افزایش دهیم. مدل سازی به بیان ساده یعنی بیان مسئله و پدیده مورد مطالعه به زبان یک علم. هدف عمومی این درس ایجاد زمینه هایی در دانش آموزان است که آن ها را قادر سازد علاوه بر انجام اعمال ریاضی کاربرد ریاضیات را در زندگی روزمره مشاهده نموده و به بررسی مسائل محیط اطراف بپردازند. معمولاً مسائلی که در آمار و مدل سازی مطرح هستند به گونه ای هستند که افق های تازه ای را در برابر دانش آموزان باز می نمایند و در نتیجه دانش آموزان از وسعت دید بیش تری برخوردار می شوند. به طور کلی این درس توانایی های زیر را در دانش آموزان تقویت می نماید:

- ۱- پرورش قدرت مشاهده
- ۲- افزایش دقت دانش آموزان و عادت به یک نظم فکری
- ۳- پرورش قدرت استدلال و استفاده از تکنیک های آماری به عنوان روش های استدلال و تفسیر
- ۴- پرورش قدرت تجزیه و تحلیل داده ها و فرضیه ها و تشخیص مناسب بودن آن ها
- ۵- پرورش تفکر نقاد و قدرت بیان و ارائه نتایج حاصل از استدلال و تفسیر
- ۶- پرورش روحیه همکاری و توانایی ها و قابلیت های زبانی و مهارت های برقراری ارتباط در زندگی اجتماعی

و یادگیری های بعدی

- ۷- پرورش قدرت حل مسائل در ارتباط با محیط اطراف
 - ۸- آشنایی با ابزار تکنولوژی و استفاده از آن ها در آموزش
- قسمت هایی از کتاب که با ترام صورتی مشخص شده برای مطالعه در نظر گرفته شده است و جزو برنامه اجباری نمی باشد. همچنین فصل هشتم به صورت آزمایشی به کتاب افزوده شده است.
- معلمین با در نظر گرفتن توانایی دانش آموزان و در دست داشتن وقت کافی می توانند این فصل را در برنامه تدریس خود بگنجانند.

من پرچم هدایتم و پناهگاه پرهیزگاری و جایگاه سخاوت و دریای
جود و جوانمردی کوه خرد. من صدیق اکبرم و علم خدا هستم
و قلب آگاه خدایم و زبان گویای خدا و چشم خدا و دست خدا
هستم. من ریسمان استوار خدایم و کلمه تقوایم و امین راز خدایم.
من حجت عظمی و آیت کبری و نمونه والا و دروازه پیامبر مصطفی
هستم. من میزان اعمالم.



اندازه‌گیری و مدل سازی

آیا فکر می‌کنید میزان تحصیلات افرادی که با آنها تماس دارید در علاقه شما به ادامه تحصیل نقش تعیین‌کننده‌ای دارند؟ مسلماً بی‌تأثیر نیست. چگونه می‌توان تصور خوبی از بیشتر افرادی که روزانه با آنها برخورد می‌کنید در مورد تحصیلات عالی بدست آورد؟

آیا تاکنون از خود پرسیده‌اید روزانه با چه کسانی سر و کار دارید؟ این افراد چه سن و سالی دارند؟ شغل آن‌ها چیست و میزان تحصیلات آن‌ها چقدر است؟ برای چند روز متوالی فهرستی از این افراد تهیه کنید، اطلاعاتی درباره شغل، میزان تحصیلات و سن آن‌ها جمع‌آوری کنید.

قبل از آن که این فهرست را تهیه کنید چه اعدادی در ذهن شما مجسم شده است؟ پاسخ شما به این سؤال‌ها در ابتدا ممکن است حتی به صورت کمی (عددی) نباشد، مثلاً ممکن است بگویید: روزانه با تعداد کمی از افراد سر و کار دارم، اکثر آن‌ها باسواد و جوان هستند. این پاسخ‌ها برای ترسیم دورنمای اولیه از اطلاعات مورد نظر کفایت می‌کند، ولی به هیچ وجه از نظر مطالعات علمی سودمند نیستند. این اطلاعات باید با معیارهای مناسب اندازه‌گیری و به صورت اعداد و ارقام بیان شوند. برای انجام بررسی‌های آماری به اندازه‌گیری نیاز داریم.

اندازه‌گیری

ما در بخش
کشاورزی در تولید
گندم در سال‌های
اخیر پیشرفت زیادی
داشته‌ایم.

فکر می‌کنید فردی کوتاه قد یا بلند قد هستید؟ هنگام پاسخ، قدتان را با چه کسانی مقایسه می‌کنید. اگر در چین و یا اتریش زندگی می‌کردید جواب‌های متفاوتی به این سؤال می‌دادید. اگر بخواهید در مورد قدتان دقیق‌تر صحبت کنید اولین قدم اندازه‌گیری آن است.

طول قد شما چقدر است؟ قد را با کدام واحد اندازه

می‌گیرید؟ متر؟ سانتی‌متر؟ میلی‌متر؟

اگر بتوانیم اطلاعات خود را به زبان اعداد و ارقام بیان کنیم،

کار مفیدتری انجام داده‌ایم. چون اطلاعات کمی (اطلاعاتی که به صورت اعداد و ارقام بیان شدند) قابل تفسیر نیستند و برای همه کس یک معنای واحد دارد، اما اطلاعات کیفی (توصیفی) این طور نیستند ممکن است شنونده برداشت‌های متفاوتی از آن داشته باشد. تیترا خبری بالا را در نظر بگیرید.

این جمله خبری خوشحال‌کننده است ولی برای مسئولان که باید دربارهٔ واردات گندم تصمیم بگیرند کافی نیست. باید منظور از سال‌های اخیر معلوم باشد آیا منظور دو سال گذشته است و یا ۶ سال گذشته پیشرفت در تولید گندم چقدر بوده است؟ آیا به میزانی بوده است که ما را از واردات آن بی‌نیاز کند و یا اگر تا آن حد نبوده است به چه میزانی کمبود داریم. ما نمی‌توانیم به کشورهای صادرکننده بگوییم مقداری گندم به ما بفروشید. باید وزن و یا حتی نوع آن کاملاً مشخص باشد.

اولین اقدام در رسیدن به اطلاعات عددی اندازه‌گیری است. انتخاب معیار اندازه‌گیری برای بعضی از موضوعات از قبیل، قد، وزن، میزان تحصیلات و درجه حرارت آسان است و با وسایلی که می‌شناسیم می‌توانیم آن‌ها را اندازه‌گیری کنیم.

هر چند انتخاب معیار مناسب برای اندازه‌گیری ممکن است ابتدا آسان به نظر بیاید ولی برای برخی از موضوعات، کار آسانی نیست. به نظر می‌رسد به آسانی می‌توان قد، وزن، سن و میزان تحصیلات افرادی را که با آن‌ها سر و کار دارید، اندازه‌گیری کنید ولی میزان درآمد آن‌ها به این سادگی نیست. (امتحان کنید) مثلاً مردم زیاد مایل نیستند که دربارهٔ درآمد خود با دیگران صحبت کنند، بنابراین اگر شما بخواهید از میزان درآمد افراد اطلاعاتی کسب کنید، مشکلاتی خواهید داشت. در بعضی مواقع مردم کم‌گویی می‌کنند، یعنی درآمد خود را کم‌تر از آن چیزی که هست بیان می‌کنند و در بعضی مواقع زیادگویی می‌کنند. شاید به بعضی‌ها برخورد کنید که اصلاً نمی‌دانند میزان درآمد آن‌ها چقدر است ممکن است یک میزان تقریبی برای آن بیان کنند. در هر صورت شما سعی کنید درآمد افراد خانواده، برخی از افراد فامیل،

در نماز خم ابروی تو در یاد آمد ...

آشنایان و دوستان را معلوم کنید. در عمل خواهید دید که دست‌یابی به میزان درآمد به سادگی اندازه‌گیری قد نیست.

از این‌ها مشکل‌تر اندازه‌گیری موضوعاتی است که دسترسی به آن‌ها مشکلاتی دارد. برای مثال ممکن است شما درباره‌ی هوش هم‌کلاسی‌های خود قضاوت کرده باشید. بعضی‌ها را باهوش و برخی دیگر را دارای هوش متوسط ارزیابی کرده باشید؛ ولی اگر بخواهید مقایسه دقیق‌تری انجام دهید، لازم است معیاری برای سطح برخورداری از هوش در اختیار داشته باشید تهیّه این معیارها و استاندارد کردن آن‌ها کار افراد متخصص است. آن‌ها به‌نحوی و وسایل اندازه‌گیری را طراحی می‌کنند تا مطمئن شوند این وسیله همان چیزی را اندازه می‌گیرد که مورد نظر آن‌هاست.



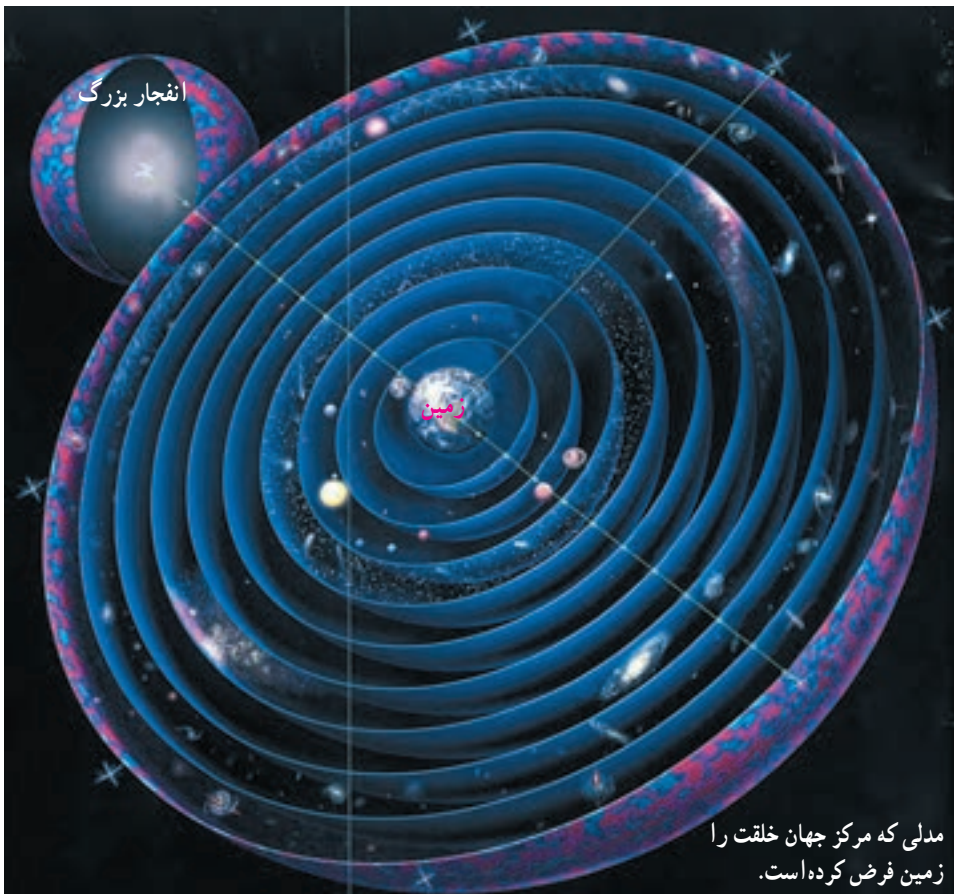
فرض کنید هنوز وسیله‌ای برای اندازه‌گیری وزن درست نشده است و ما می‌خواهیم سنگینی افراد و یا اشیا را اندازه‌گیری کنیم. ممکن است این‌طور استدلال کنیم هر شیئی که جای بیش‌تری اشغال کند (حجم آن بیش‌تر باشد) سنگین‌تر است و لذا برای محاسبه سنگینی اشیا، آن‌ها را در ظرف مدرج پر از آب قرار دهیم و به این ترتیب حجم آن را حساب کنیم. شاید این‌طور اندازه‌گیری کردن تا حدودی ما را به واقعیت نزدیک کند ولی می‌دانیم که دو مفهوم حجم و وزن دو مقوله جدا از هم هستند. می‌دانیم می‌توان دو قطعه هم وزن از طلا چنان ساخت که یکی حجم تراز دیگری به نظر برسد. بنابراین رسیدن به وزن از طریق حجم کار درستی نیست. در تعیین ظرفیت هواپیما و آسانسورها وزن اهمیت دارد نه حجم. درباره هوش اگر مواظب نباشیم شاید حافظه فرد و یا سایر قابلیت‌های دیگر ذهنی او اندازه‌گیری شود. در این کتاب، وارد مبحث ساختن معیارها، برای اندازه‌گیری نمی‌شویم و به آنچه که تاکنون آشنا شده‌ایم و به کار بسته‌ایم قناعت می‌کنیم.

فعالیت

کاغذی در کلاس خود به دیوار نصب کنید و در مقابل آن بایستید، بالای سر خودتان را روی کاغذ علامت بزنید و اندازه قد و نام خود را در کنار آن یادداشت کنید. این ورقه می‌تواند به عنوان یادگاری از شما در این کلاس باقی‌بماند و سایرین هم در سال‌های آینده از آن تبعیت کنند. این اعداد ممکن است مبنایی برای مقایسه قد دانش‌آموزان در سال‌های مختلف تحصیل در این مدرسه باشد.

مدل سازی

کیپلر ستاره‌شناس بزرگ در مورد خلقت جهان چنین گفته است: «خداوند جهان را به زبان اعداد خلق کرده است». شاید سخن او به این معنی است که هر آن چه که خداوند آفریده است به زبان ریاضی قابل توضیح و تفسیر است. مثلاً کاری که خود کیپلر در طول ۲۳ سال کار مداوم انجام داد، بیان ریاضی علت و چگونگی حرکت سیارات در قالب سه قانون کیپلر است. البته برای رسیدن به این سه قانون مجبور شد فرض‌هایی را در نظر بگیرد. مثلاً سیارات را به صورت نقطه در نظر گرفت. در حرکت، مسیرها و اندازه‌گیری‌ها از تقریب‌های مناسب استفاده کرد. همه ما زمین را به شکل کره در نظر می‌گیریم و حال آن که واقعیت امر این نیست. کره، ساده‌ترین و نزدیک‌ترین شیء ریاضی است که زمین به آن شباهت دارد. همین در نظر گرفتن شکل زمین به صورت کره نوعی مدل‌سازی ریاضی است. وقتی که کیپلر قوانین حرکت سیارات را در قالب سه قانون خود با استفاده از نمادها و اصطلاحات و قوانین ریاضی بیان کرد در واقع حرکت سیارات را مدل‌سازی ریاضی کرد.



مدلی که مرکز جهان خلقت را زمین فرض کرده است.

شما اگر به بازی‌های کودکان توجه کنید تمام آن‌ها به نوعی مدل‌های زندگی بزرگسالان است. وقتی کودکی عروسکی را در آغوش دارد همان رفتار و حرکاتی را دارد که مادران به هنگام در آغوش گرفتن فرزندان از خود بروز می‌دهند. وقتی کودکی اتومبیل کوچکی را روی زمین حرکت می‌دهد، ساده‌ترین شکلی است که در دنیای کودکان می‌توان اتومبیل‌رانی را مدل‌سازی کرد. زیبایی، خلاقیت و اهمیت این نوع بازی‌ها در این است که اولاً اسباب‌بازی‌ها حتی الامکان ساده باشند و ثانیاً کودکان بتوانند آنچه را که در ذهن آن‌ها می‌گذرد بهتر نشان دهند. ما اگر بتوانیم با مفاهیم ریاضی و یا آمار پدیده‌ها را توضیح دهیم یک مدل‌سازی ریاضی و یا آماری کرده‌ایم.

بیان مسئله به زبان ریاضی را مدل‌سازی ریاضی می‌گوییم. هرچقدر مفاهیم ریاضی به کار برده ساده‌تر و ابتدایی‌تر و نتیجه کار به پدیده مورد نظر نزدیک‌تر باشد، مدل‌سازی، با ارزش‌تر است.

ما در این درس به نوعی خاص از مدل‌سازی ریاضی که براساس آمار و مفاهیم آن بنا شده است، اشاره خواهیم کرد. در واقع خواهید دید هر یک از مفاهیمی را که در آمار معرفی می‌کنیم ابزاری است برای مدل‌سازی پدیده‌های تحت مطالعه.

مثلاً قد افراد یک پدیده طبیعی است که بنا بر ضرورت در مسائل مختلف در نظر گرفته می‌شود. وقتی شما قد یک فرد را اندازه می‌گیرید در واقع نزدیک‌ترین و ساده‌ترین مفهوم ریاضی به قد را که همان طول قد، که یک عدد است در نظر گرفته‌اید. پس از این اندازه‌گیری‌ها شما با افراد و سایر مشخصات آن‌ها کاری ندارید، فقط این اعداد هستند که از این پس موضوع مطالعه شما خواهند بود. وقتی یک طراح از یک چهره طراحی می‌کند در واقع مدلی از این چهره می‌سازد این مدل می‌تواند خیلی ساده باشد، مثلاً از چند خط راست تشکیل شده باشد مانند شکل الف.



ب



الف

اگر بتوانیم از خطوط منحنی و برخی از مهارت‌های طراحی هم استفاده کنیم مدل پیشرفته‌تری خواهیم داشت که به شکل واقعی نزدیک‌تر است مانند شکل ب. همین‌طور اگر ادامه دهیم و اجازه داشته باشیم از رنگ و سایر وسایل نقاشی استفاده کنیم به مدل کامل‌تری دست خواهیم یافت.

این تقریبات متوالی در اندازه‌گیری طول هم رخ می‌دهد. اگر فقط بخواهیم قد فردی را وجب کنیم (صرف‌نظر از اینکه اندازه دست افراد متفاوت است)، مثلاً خواهیم گفت قد همکلاسی شما کمی بیش‌تر از هفت وجب خودش است. چون مقدار کمی بیش‌تر را نمی‌توانیم برحسب وجب بیان کنیم، خواهیم گفت قد این فرد هفت وجب است. و این یک مدل‌سازی ساده از قد است. ولی اگر قد همین فرد را با یک خط‌کش مدرج اندازه‌گیری کنید خواهید گفت قد این فرد ۱۵۴ سانتی‌متر است.

مدل‌سازی اول خیلی ساده و ابتدایی بود و لذا کار دقیق نمی‌توان روی آن انجام داد. مثلاً شما لباسی را تصور کنید که برحسب وجب برای کسی دوخته شده است. اما مدل‌سازی دوم دقیق است و لباسی که از این طریق دوخته می‌شود قابل قبول‌تر است. البته ما در اندازه‌گیری قد نتوانسته‌ایم کار را تمام کنیم، زیرا از واحدهای کوچک‌تر از سانتی‌متر صرف‌نظر کرده‌ایم. البته دو یا سه میلی‌متر در مقابل یک متر محسوس نیست. لذا صرف‌نظر کردن از آن‌ها لطمه زیادی به مدل نخواهد زد.

ولی مواردی هست که به‌سادگی نمی‌توان از این جزئیات گذشت. شما وزن خود را برحسب کیلوگرم اندازه‌گیری می‌کنید. اگر بخواهید دقت زیاد داشته باشید از نیم‌کیلو هم استفاده خواهید کرد. از این رو وزن افراد معمولاً به‌صورت اعداد طبیعی (مثلاً ۵۳ کیلو، ۶۴ کیلو) و یا به‌صورت اعشار (مثلاً ۵۴/۵ کیلو و ۷۲/۵ کیلو) است. به ندرت از واحدهای گرم در اندازه‌گیری وزن بزرگسالان استفاده می‌شود. زیرا در مقابل چند ده کیلو، چند گرم به حساب نمی‌آید. اما در مسابقات کشتی از این واحدهای جزئی یعنی گرم هم نمی‌توان گذشت. در شرایط مساوی کسی برنده است که وزن او ولو چند گرم، کم‌تر باشد. در مسابقات شنا به واحدهای دهم ثانیه هم متوسل می‌شوند. اگر روزی وسایل اندازه‌گیری دقیق‌تر شود شاید به واحد کوچک‌تر هم متوسل شوند ولی شاید هیچ‌گاه نتوانند مقدار دقیق زمانی را که یک شناگر ۵۰ متر را طی کرده است اندازه‌گیری کنند. وسایل اندازه‌گیری را همیشه می‌توانند دقیق‌تر کنند ولی هیچ‌گاه دقت به آن اندازه‌ای نخواهد رسید که خطای اندازه‌گیری را صفر کند.

خطای اندازه‌گیری همان تفاضل مقدار واقعی و مقدار اندازه‌گیری شده است. این خطا لزوماً از واحد اندازه‌گیری کمتر است.

چگونه می‌توانیم این پدیدهٔ عدم دسترسی به مقدار واقعی را مدل‌سازی کنیم؟ فرض کنید مدت زمانی که یک دوندۀ فاصله ۱۰۰ متر را دویده است برابر ۷ ثانیه اندازه‌گیری شده باشد. می‌دانیم مدت‌زمان مذکور ممکن است دقیقاً برابر ۷ ثانیه نباشد، پس مرتکب خطایی که اندازهٔ آن از یک ثانیه کم‌تر است شده‌ایم، ولی واقعاً چقدر کم‌تر، نمی‌دانیم! از این رو اگر مدت زمانی که دوندۀ ۱۰۰ متر را بر حسب ثانیه دویده است با متغیر T نمایش دهیم، می‌توان مدل زیر را برای آن پیشنهاد کرد:

$$T = 7 + E^*$$

که در آن قدرمطلق E از یک ثانیه که واحد اندازه‌گیری ماست کم‌تر است، E ممکن است مثبت و یا منفی باشد.

به عنوان مثال دیگر فرض کنید وزن فردی برابر ۶۴/۵ کیلو گزارش شده باشد. چه مدلی ممکن است برای اندازه‌گیری بکار رفته باشد؟
فرض کنیم P وزن این فرد باشد، در این صورت

$$P = 64.5 + E$$

پس در این جا واحد اندازه‌گیری نیم کیلو در نظر گرفته شده است.

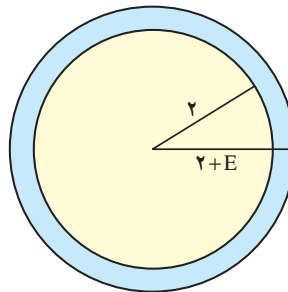
پس قدرمطلق E از ۰/۵ کیلو و یا ۵۰۰ گرم کم‌تر است.

اگر یک پدیده تابعی از چند متغیر باشد، خطای آن متغیرها چگونه در پدیدهٔ نهایی ظاهر می‌شود؟ مثلاً فرض کنید شعاع دایره‌ای به صورت $R = 2 + E$ مدل‌سازی شده باشد. می‌خواهیم مدلی برای مساحت این دایره بنویسیم.

می‌دانیم مساحت دایره برابر πR^2 است. بنابراین اگر فرض کنیم:

$$S = \text{مساحت دایره}$$

$$\begin{aligned} S &= \pi(2 + E)^2 \\ &= \pi(4 + E^2 + 4E) \end{aligned}$$

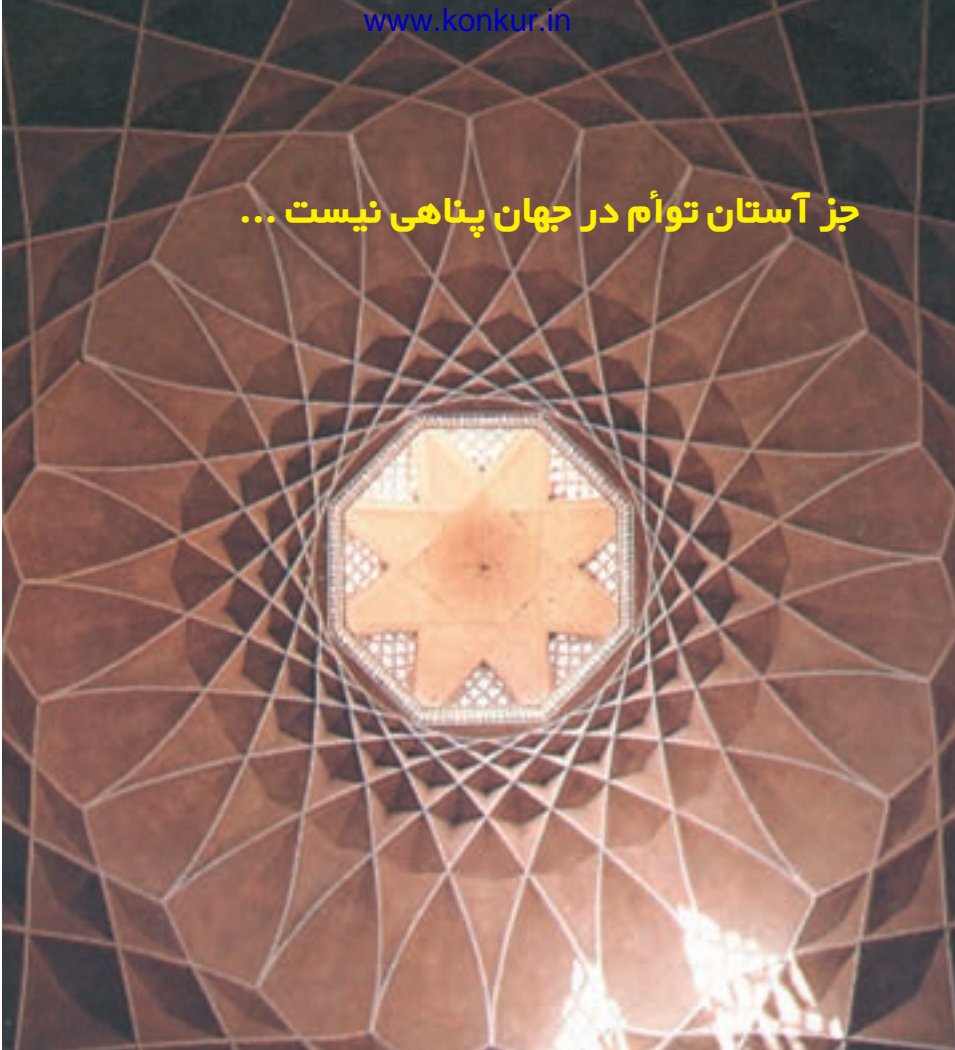


آن‌گاه

می‌توان از جملاتی که شامل توان دوم یا بالاتر از خطاست صرف‌نظر کرد.

* E ابتدای ERROR به معنای خطا می‌باشد.

جز آستان توأم در جهان پناهی نیست ...



پس داریم :

$$S \approx 4\pi + 4\pi E$$

اگر $4\pi E$ را برابر E_1 فرض کنیم مدل مساحت دایره عبارت خواهد بود از :

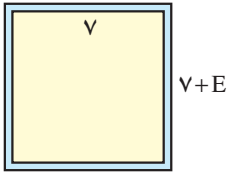
$$S \approx 4\pi + E_1$$

که 4π مساحت دایره‌ای به شعاع ۲ است.

می‌بینیم که خطای مساحت 4π برابر از خطای شعاع بزرگتر است. پس ممکن است از واحد سطح بزرگتر شود. برای جلوگیری از این اتفاق باید دقت اندازه‌گیری طول شعاع را بیشتر کنیم. می‌توان با اندازه‌گیری دقیق‌تر E را چنان کوچک کرد که E_1 از واحد سطح کمتر باشد.

حال به مثال دیگری می‌پردازیم. اگر ضلع مربعی به صورت $E + \gamma$ باشد، مساحت آن با چه

تقریبی در دسترس است؟



فرض کنید S برابر مساحت مربع باشد، در این صورت :

$$S = (V+E)^2 = 49 + 14E + E^2$$

با صرف نظر کردن از جملات مرتبه بالا داریم :

$$\begin{aligned} S &\approx 49 + 14E \\ &= 49 + E_1 \end{aligned}$$

اگر بخواهیم بینیم محیط مربع با چه تقریبی در دسترس است، این طور می نویسیم.

فرض کنید P محیط مربع باشد، در این صورت :

$$\begin{aligned} P &= 4(V+E) = 28 + 4E \\ &= 28 + E_2 \end{aligned}$$

ما در عمل محیط مربع را برابر ۲۸ خواهیم گرفت ولی می دانیم که محیط مربع با ۲۸ تفاوتی دارد. اگر قرار باشد این تفاوت را به حساب بیاوریم باید از برابری $P = 28 + E_2$ استفاده کنیم.

خطای ناشی از اندازه گیری اضلاع بیشتر در محیط مربع اثر می گذارد یا در مساحت آن؟

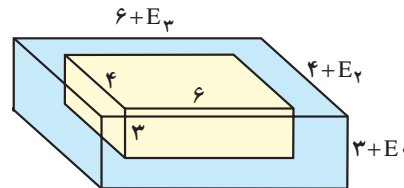
در بعضی موارد پدیده مورد بررسی متأثر از چند عامل است که هر یک با خطای متفاوتی اندازه گیری می شوند.

مثلاً اگر اضلاع یک مکعب مستطیل را به صورت زیر داشته باشیم :

$$\text{ارتفاع} = L_1 = 3 + E_1$$

$$\text{عرض} = L_2 = 4 + E_2$$

$$\text{طول} = L_3 = 6 + E_3$$



حجم این مکعب از چه مدلی بیروی می کند؟ فرض کنید حجم مکعب برابر V باشد، در این صورت :

$$\begin{aligned} V &= L_1 L_2 L_3 \\ &= (3 + E_1)(4 + E_2)(6 + E_3) \\ &\approx 72 + 24E_1 + 18E_2 + 12E_3 \end{aligned}$$

در این مدل E_i ها از واحد کوچک ترند پس حاصل ضرب آن ها خیلی کوچک خواهد بود، پس می توانیم از آن هم صرف نظر کنیم. بنابراین یک مدل ساده قابل قبولی به دست خواهد آمد. به پدیده های اطراف خود توجه کنید و بینید تابع چه متغیرهایی هستند و سعی کنید اثر خطای اندازه گیری در هر یک از متغیرها را در کل پدیده بررسی کنید.

در دو مثال قبل دیدیم که چگونه خطای اندازه‌گیری در مساحت و محیط مربع و حجم مکعب مستطیل «منتشر» می‌شود. این روابط به ما کمک می‌کنند تا برای جلوگیری از خطاهای بزرگ در محصول نهایی محاسبات، از ابتدا پیش‌گیری‌های لازم را به‌عمل آوریم.

تمرین

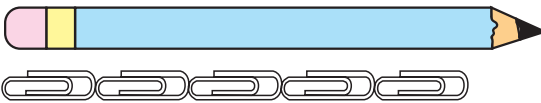


۱- دماسنجی بردارید و با اندازه‌گیری دمای اتاق مدلی برای گرمای آن بنویسید. این مدل تا چه مدت اعتبار دارد؟ آیا از همین مدل برای ۸ ساعت دیگر نیز می‌توانید استفاده کنید؟ توضیح دهید.

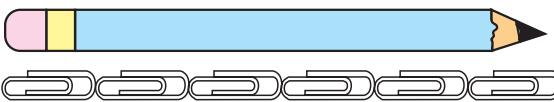
۲- طول مداد و قطر قاعده آن را برحسب میلی‌متر اندازه‌گیری کنید و مدلی برای حجم مداد بنویسید. (Eهایی که برای طول و قطر قاعده به کار می‌برید یکی نیستند، توضیح دهید.)
۳- طولی را برحسب سانتی‌متر اندازه‌گیری کرده‌اید و مدل آن را به صورت $L = 2 + E$ نوشته‌اید. اگر این طول را برحسب میلی‌متر اندازه‌گیری کنید آیا مدل $L = 20 + E$ برای آن مناسب است یا خیر؟ توضیح دهید.

۴- مدلی برای وزن خودتان پیشنهاد کنید. با استفاده از ابزار اندازه‌گیری وزن، درستی آن را تحقیق کنید.

۵- مدلی برای اندازه‌گیری طول مداد مطابق با هر یک از شرایط زیر بنویسید.



الف



ب



ج

خدایی که آفریدگان را از هیچ پدید آورد. الگویی نداشت تا به کار برد، و نه مقیاسی از آفریننده‌ای پیش از خود، تا بدان دستور کار کند، و آفریدگان اعتراف دارند بدین حقیقت که سر اسر ناتوان و فقیرند و نیازمند و حقیر؛ و اوست که باید بر آنان رحمت آرد، و به قدرت خود برپایشان دارد. به ما آن نشان داد که دیدیم، به حکم ضرورت آشکار است. که این نشانه‌ها بر شناخت او دلیلی استوار است و در آنچه آفریده، آثار صنعت و نشانه‌های حکمت او پدیدار است. چنانکه هر چه آفریده، او را برهانی است، و بر قدرت و حکمت او نشانی. و گرچه آفریده‌ای باشد خاموش، بر تدبیر او گویاست، و بر وجود پدیدآورنده دلیلی رساست.

جامعه و نمونه

چند ستاره در آسمان می توان یافت؟ آیا تا به حال در یک شب صاف که ماه در آسمان نیست به آسمان نگاه کرده اید؟ شاید فکر کرده اید که تعداد ستاره ها آن قدر زیاد است که هرگز نمی توان آن ها را شمرد. اما باید بدانید که تعداد ستارگانی که می توان بدون چشم مسلح دید بیش از ۳۰۰۰ نیست. اگر با یک دوربین به آسمان نگاه کنید تعداد آن ها تا ۵۰۰۰۰ افزایش می یابد و اگر با یک تلسکوپ بزرگ به آسمان نظر کنید تعداد ستارگانی که در دیدرس شما هستند تا ۵۰۰,۰۰۰,۰۰۰ افزایش می یابند. این عدد آن قدر بزرگ است که اگر شما در تمام طول عمر هرشب تمام شب را به شمردن ستارگان بگذرانید، عمر شما کفاف نخواهد داد که این تعداد ستاره را بشمارید.

فکر می کنید ستاره شناسان چگونه به چنین عدد بزرگی رسیده اند؟ آنان این شمارش را با نمونه گیری انجام داده اند. می توانیم آسمان را به چندین بخش مساوی تقسیم کنیم و تعداد ستارگان را در برخی از این نواحی بشماریم. براساس این اعداد، می توان تعداد تقریبی ستاره ها را محاسبه کرد.

جامعه

در فصل قبل با مسائل مربوط به اندازه گیری و مدل سازی های آن به طور مقدماتی آشنا شدید. حال به موضوع قد افراد باز می گردیم. شما قبلاً قد بعضی از همکلاسی های خود را اندازه گرفته اید. اگر بخواهید راجع به قد افراد مطالعه کنید، چه اقداماتی انجام خواهید داد؟ خوب است که در یک کار علمی و دقیق، کلمات دارای معنای روشن و تعریف شده ای باشد. آیا می دانید منظور از مطالعه قد چیست؟ پاسخ این سؤال را به تدریج در طول این درس خواهید فهمید ولی مقدمتاً بیان می کنیم که منظور از مطالعه قد افراد، آن است که بدانیم چند درصد افراد قد کوتاه و چند درصد دارای قد معمولی و چند درصد افراد قد بلند هستند و یا اصلاً می خواهیم بدانیم که چه افرادی را قد کوتاه و چه افرادی را قد بلند می گویند. اندازه قد بیش تر مردم چقدر است؟ و سؤالاتی از این قبیل. بلافاصله ممکن است بر این سؤالات این اشکال را وارد کنید که سؤال های بالا برای سنین مختلف متفاوت

است. معیار قد کوتاهی برای نوجوانان با معیار قد کوتاهی برای جوانان و یا بالاتر تفاوت می‌کند. پس نمی‌توانیم بدون محدود کردن افراد، جواب مشخصی به سؤال‌های بالا بدهیم. مگر افراد را مشخص کنیم و برای آن‌ها شرایطی تعیین کنیم. مثلاً بگوییم منظور ما مطالعه قد نوجوانان پسر است (پسران بین ۱۴-۱۱ سالگی) با این شرایط مجموعه‌ای از افراد مشخص داریم که مطالعه قد آن‌ها معقول به نظر می‌رسد و اشکالات مطرح شده در بالا را ندارد. چنین مجموعه‌ای را یک جامعه آماری می‌گوییم.

جامعه آماری مجموعه‌ای از افراد یا اشیا است که دربارهٔ اعضای آن می‌خواهیم موضوع و یا موضوعاتی را مطالعه کنیم.

- مثالهایی از جامعه آماری و ویژگی‌های مورد بررسی آن‌ها می‌توانند به صورت زیر باشند:
- محصولات کشاورزی استان مازندران و موضوع مورد مطالعه، انواع محصولات.
 - واردات و صادرات کشور و موضوع مورد مطالعه، نوع کالاهای وارد شده یا صادر شده.
 - مجموعهٔ دانش‌آموزانی که در سال ۱۳۷۹ در مقطع راهنمایی در شهر تهران ثبت نام کردند و موضوع مورد مطالعه، معدل سال قبل آن‌ها.
 - مجموعهٔ پزشکان متخصص قلب و عروق و موضوع مورد مطالعه، تعداد بیمارانی که به علت ناراحتی‌های قلبی در مقطع معینی از سال به آن‌ها مراجعه کرده‌اند.





– مجموعه کارگاه‌های خدمات فنی استان مرکزی و موضوع مورد مطالعه، تعداد کارگران شاغل در آن‌ها.

– مجموعه افراد جویای کار در سه ماهه اول سال ۱۳۷۹ و موضوع مورد مطالعه، میزان تحصیلات آن‌ها.

– مجموعه فارغ‌التحصیلان دانشگاه‌های ایران در سال ۱۳۷۸ و موضوع مورد مطالعه، طول دوره تحصیل دانشگاهی آن‌ها.

– مجموعه دبیران استان خراسان و موضوع مورد مطالعه، سابقه تدریس آن‌ها.

– مجموعه دبیران ریاضی استان خراسان و موضوع مورد مطالعه، سابقه تدریس آن‌ها. همانطوری که ملاحظه می‌کنید بعضی از جامعه‌های آماری خود زیر مجموعه‌ای از جامعه‌های آماری دیگر هستند. مثلاً مجموعه دبیران ریاضی استان خراسان زیر مجموعه‌ای از دبیران این استان است.

نمونه

ما در مطالعه قد گفتیم که می‌خواهیم بدانیم چند درصد افراد با در نظر گرفتن معیار قد بلندی در جامعه مورد بررسی، قد بلند هستند. یک راه حل ابتدایی آن است که تمام افراد را اندازه بگیریم و بعد ببینیم چند نفر قد بلند هستند و از این‌جا درصد مورد نظر را محاسبه کنیم. مثلاً اگر جامعه ما ۶۰۰ نفر عضو داشته باشد و ۶۰ نفر قد بلند باشند آن‌گاه ده درصد افراد این جامعه قد بلند هستند.



اگر تمام افراد جامعه را مورد مطالعه قرار دهیم می‌گوییم سرشماری کرده‌ایم.

معمولاً در سرشماری‌ها با مشکلاتی مواجه هستیم، مهم‌ترین آن‌ها عبارتند از:

- در دسترس نبودن تمام اعضای جامعه
- وقت‌گیر بودن دسترسی به تمام اعضای جامعه
- گران تمام شدن بررسی تمام اعضای جامعه
- از بین رفتن جامعه در برخی از مطالعات

این مشکلات سبب می‌شود تا سعی کنیم از راه میان‌بر برویم و به‌جای آن که تمام اعضای جامعه را مورد مطالعه قرار دهیم بخشی از آن را که با دقت و مطالعه لازم انتخاب شده است، بررسی کنیم. البته در این صورت بخشی از اطلاعات را از دست داده‌ایم ولی در مقابل، مشکلاتی را که در بالا به آن‌ها اشاره کردیم از پیش‌رو برداشته‌ایم. این بخش کوچک از جامعه آماری را نمونه می‌گوییم.

نمونه زیر مجموعه‌ای از جامعه آماری است.

فرض کنید بخواهید برخی از ویژگی‌های ماهی‌اوزن برون را در دریای خزر مطالعه کنید. در این صورت به تمام ماهی‌های دریای خزر دسترسی نخواهید داشت، لذا لازم است از طریق مطالعه نمونه، این بررسی انجام شود.

در بعضی از کارخانه‌ها بخشی به نام بخش کنترل کیفی کالا وجود دارد که کالاها را از لحاظ مرغوبیت و یا سالم و معیوب بودن بررسی می‌کند. فرض کنید یک کارخانه تولید لامپ، بخواهد ولتاژی را که لامپ تولیدی آن کارخانه تحمل می‌کند بررسی کند. آزمایشی که برای تعیین سطح تحمل لامپ از آن استفاده می‌شود به گونه‌ای است که لامپ در این آزمایش می‌سوزد. در این مثال خاص سرشماری یعنی از بین بردن تمام محصولات کارخانه. پس برای جلوگیری از چنین زیانی مجبوریم از نمونه استفاده کنیم.

در ادبیات ما ضرب‌المثل «مشت نمونه خروار است» به خوبی موضوع جامعه و ضرورت استفاده از نمونه را نشان می‌دهد. در این جا مسأله عبارت است از خرید برنج برای مصرف یک سال خانواده. فرض کنید میزان مصرف ۱۰۰ کیلو باشد. قبل از خرید باید به قسمی از کیفیت برنج موجود در انبار اطمینان حاصل کنیم. برای رسیدن به این هدف قسمتی از آن که در حد یک مشت برنج است آزمایش می‌شود و براساس آن تصمیم‌گیری می‌شود. در این شرایط برنج موجود در انبار جامعه و یک مشت



برنجی که انتخاب شده است، نمونه است.

ملاحظه می‌کنید که این نمونه نسبت به آن جامعه خیلی کوچک است. آیا نتیجه‌ای که از بررسی این نمونه کوچک حاصل می‌شود قابل اطمینان است؟ پاسخ به این سؤال مثبت است. زیرا فرض بر این است که برنج موجود در انبار همگی از یک نوع بوده و محصول یک دست است. اگر غیر از این باشد و برنج‌های موجود در انبار از انواع مختلف در گونی‌های متفاوت روی هم انبار شده باشند، این یک مشت نمی‌تواند معیاری برای قضاوت باشد.

عمل نمونه‌گیری مهم‌ترین بخش آمار را تشکیل می‌دهد. یک نمونه گروه کوچکی از اعضا است که به نحوی انتخاب شده‌اند که نمایانگر خصوصیات عده بزرگ‌تری که «جامعه» نام دارد، باشد. برای آن که نمونه به‌درستی نمایانگر خصوصیت تمام جامعه باشد، باید به اندازه کافی بزرگ باشد. اگر تعداد اعضای نمونه خیلی کوچک باشد ممکن است اعضای آن نتوانند نمایندگان خوبی برای اعضای جامعه باشند. همچنین اعضای نمونه باید تصادفی انتخاب شوند. یعنی انتخاب آن‌ها نباید از قانون خاصی پیروی کند.

اندازه جامعه و اندازه نمونه: گفتیم جامعه یک مجموعه است که می‌خواهیم موضوعاتی را در آن مطالعه کنیم. این مجموعه اعضای دارد. تعداد این اعضا در بررسی‌های آماری نقش مؤثری دارد. اگر جامعه در ارتباط با کاری که می‌خواهیم انجام دهیم کوچک باشد، بیشتر اوقات به‌صورت سرشماری مطالعه را انجام می‌دهیم ولی اگر بزرگ باشد همان‌طوری که گفتیم لازم است مطالعات از طریق نمونه‌گیری دنبال شود. از این رو تعداد اعضای جامعه را به‌عنوان پارامتری که باید مد نظر قرار داد مطرح می‌کنند. این تعداد را اندازه جامعه می‌گویند.

تعداد اعضای جامعه را اندازه جامعه می‌گوییم.

بعضی از جامعه‌ها به سبب ویژگی خاص آن‌ها کوچک‌اند مثلاً مجموعه افراد زال مو (افرادى که موهای آن‌ها از بدو تولد سفید است نه آنکه سفید شده باشد). این یک مجموعه نسبتاً کوچک است و از جمله مواردی است که دسترسی به تمام آن‌ها امکان‌پذیر نیست. مجموعه ماهی‌های کفال دریای خزر، مجموعه‌ای است بزرگ و باز هم دسترسی به همه آن‌ها غیر ممکن است.

مجموعه دانش‌آموزان کشور، مجموعه‌ای بزرگ و ناهمگون است. زیرا آن‌ها را می‌توان

به دسته‌های پسر، دختر و یا از نظر مقطع تحصیلی به ابتدایی، راهنمایی و دبیرستان تفکیک کرد. حتی می‌توان از نظر محل جغرافیایی تحصیل نیز آن‌ها را تقسیم‌بندی کرد. اندازه جامعه در نمونه‌ای که می‌خواهیم انتخاب کنیم تأثیر دارد. اگر جامعه بزرگ باشد یعنی اگر اندازه آن زیاد باشد، اغلب لازم است نمونه‌ای که از آن انتخاب می‌شود بزرگ باشد، از این رو تعداد اعضای نمونه هم یکی دیگر از پارامترهایی است که در یک بررسی آماری باید مورد توجه قرار گیرد. این تعداد را اندازه نمونه می‌گوییم.

تعداد اعضای نمونه را اندازه آن نمونه می‌گوییم.

هرچه جامعه بزرگ‌تر باشد، برای آن که نمونه بتواند بیان‌گر جامعه باشد، باید نمونه بزرگ‌تر باشد. همچنین اگر جامعه از گوناگونی بیش‌تری برخوردار باشد برای حصول اطمینان از حضور نماینده‌ای از این گوناگونی‌ها در نمونه، باز هم لازم است نمونه بزرگ باشد. اگر اندازه نمونه مناسب نباشد نتیجه‌ای که از آن گرفته می‌شود معتبر نخواهد بود. مثلاً در شهری با جمعیت ۱۰ میلیون نفر اگر از ۱۰۰ نفر درباره یک موضوع اجتماعی سؤال شود، نتیجه‌ای که از آن حاصل می‌شود نمی‌تواند قابل تعمیم به جامعه باشد. بنابر ضرب‌المثل معروف «با یک گل بهار نمی‌شود» نمی‌توان همواره با شواهد کم حکم کلی کرد. همان‌طوری که گفتیم اندازه جامعه ممکن است خیلی بزرگ باشد، در بعضی مواقع حتی ممکن است نامتناهی باشد و یا آن‌قدر بزرگ باشد که ما آن را نامتناهی فرض کنیم، مطالعه این قبیل جامعه‌ها مشکلات خاص خود را دارد و چون ما هنوز در ابتدای راه هستیم فقط جامعه‌هایی را مورد مطالعه قرار می‌دهیم که اندازه آن‌ها متناهی باشد.



بحث کلاسی

در زیر چند نمونه‌گیری مطرح می‌شود. علت تناسب یا عدم تناسب این نمونه‌ها را در کلاس به بحث بگذارید. همچنین روشی برای نمونه‌گیری مناسب ارائه دهید.

۱- موضوع مورد مطالعه: بررسی سطح نمرهٔ ریاضیات در دورهٔ دبستان.
نمونه: دانش‌آموزان یکی از مناطق آموزش و پرورش که حرف اول اسم کوچکشان با «ذال» شروع می‌شود.

۲- موضوع مورد مطالعه: نظرسنجی در مورد اینکه کدام یک از زن و شوهر نقش مهم‌تری در ادارهٔ خانواده دارند.

نمونه: دبیران ریاضی دبیرستان‌های پسرانه شیراز.
۳- موضوع مورد مطالعه: مطالعه آماری تعداد فرزندان هر خانواده در شهر تهران.

نمونه: ساکنین یکی از مناطق تهران که به تصادف انتخاب می‌شوند.
۴- موضوع مورد مطالعه: رضایت مشتریان از محصولات یک کارخانه.
نمونه: خانواده‌های کارکنان آن کارخانه.

تمرین

۱- فرض کنید بخواهید میزان آلودگی هوا را در شهر خود بررسی کنید. جامعه را معرفی کرده و توضیح دهید که نمونه از این جامعه به چه صورتی انتخاب می‌شود.

۲- می‌خواهیم مدت زمانی را که طول می‌کشد تا افراد یک شهر از منزل به سر کارشان بروند، مطالعه کنیم، جامعه را معرفی کنید و دربارهٔ نمونه‌ای که باید انتخاب شود صحبت کنید.

۳- از یک تولیدکننده مرکبات می‌خواهید ۵۰ جعبه پرتقال بخرید. قبلاً می‌خواهید دربارهٔ اندازه پرتقال‌ها مطمئن شوید. آیا بررسی پرتقال‌هایی که روی جعبه چیده شده است کفایت می‌کند؟

۴- از سه موضوع مورد مطالعه زیر دربارهٔ کلاسی که در آن درس می‌خوانید، کدام یک از تنوع بیش‌تری برخوردار است.

موضوع مورد مطالعه:

۱- معدل

۲- قد

۳- سن

در این تمرین، نمونه‌ی مربوط به بررسی کدام یک از ویژگی‌ها می‌تواند کوچک باشد؟
۵- پنج خانوار از بین خانوارهای همکلاسی‌های خود انتخاب و تعداد افراد هر خانوار را تعیین کنید. با توجه به اطلاعات بالا به سؤال‌های زیر پاسخ دهید:

– بیش‌تر خانوارها چند نفره هستند؟

– خانوارهایی که بیش‌ترین تعداد عضو را دارند در نظر بگیرید. این خانوارها چند نفره هستند؟
قبل از این که اطلاعات را جمع‌آوری کنید حدس شما درباره‌ی پاسخ سؤال‌های بالا چیست؟
پاسخ خود را یادداشت کنید و آن را با نتیجه حاصل از اطلاعات بالا مقایسه کنید. آیا جمع‌آوری اطلاعات براساس حدس و گمان نتایج مستندی به دست می‌دهد؟

۶- سن پدر ۸ نفر از همکلاسی‌های خود را بی‌رسید. حال به سؤال‌های زیر پاسخ دهید:

– چند نفر از آن‌ها بین ۲۰ تا ۳۵ سال سن دارند؟

– چند نفر از آن‌ها بین ۳۵ تا ۴۵ سال سن دارند؟

– چند نفر از آن‌ها بیش از ۴۵ سال سن دارند؟

قبل از اینکه سن دبیران خود را بی‌رسید چه پاسخی برای سؤال‌های بالا داشتید، حال حدس خود را با این پاسخ‌ها مقایسه کنید و دلایل یکسانی و عدم یکسانی آن‌ها را جستجو کنید.

۷- ۱۵ نفر از همکلاسی‌های خود را انتخاب کنید و از آن‌ها درباره‌ی تعداد محصلین در خانوارشان سؤال کنید. این تعداد را یادداشت کنید و به سؤال‌های زیر پاسخ دهید.

– بیش‌ترین تعداد محصل در این خانوارها چقدر است؟

– بیش‌تر این خانوارها چند محصل دارند؟

۸- «خانواده‌ها کوچک‌تر شده‌اند. سال‌ها قبل متوسط تعداد فرزندان خانواده‌ها ۵ بوده است ولی اکنون ۲ است.» فرض کنید از شما خواسته شده تا یک نمونه ۱۰۰ تایی از خانواده‌ها را در نظر گرفته و درستی خبر فوق را تحقیق کنید. در این مطالعه: جامعه و نمونه را تعریف کنید.

۹- ترس از بی‌کاری برای بخشی از مردم وجود دارد. بررسی وضعیت شغلی کارمندان شاغل در یک شرکت خصوصی، نشان داد از هر ۵ نفر، ۱ نفر نگران از دست دادن شغل خود است. در این بررسی، جامعه و نمونه را مشخص کنید.

۱۰- «کارمندان بیمارستان‌ها ترجیح می‌دهند به‌جای نوبت روز، در نوبت شب کار کنند.»

نظر ۳۰ پرستار شاغل در نوبت شب چند بیمارستان راجع به کار کردن در نوبت شب پرسیده شد. پاسخ آن‌ها به قرار زیر است:

۱۱ نفر نوبت روز را ترجیح دادند. ۱۶ نفر نوبت شب را ترجیح دادند. ۲ نفر بی‌تفاوت بودند.

از این بررسی می‌توان دید که پرستاران ترجیح می‌دهند نوبت شب کار کنند. در این بررسی و نتیجه‌گیری چه ایرادهایی وجود دارد؟



نمونه تصادفی

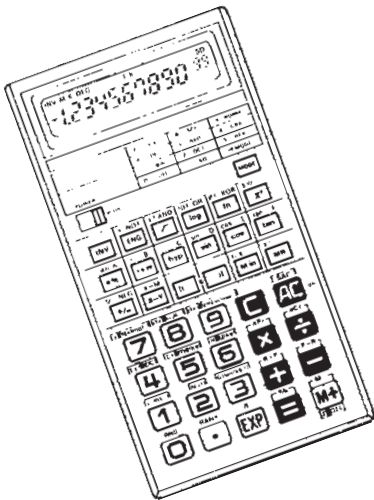
نمونه تصادفی ساده: کلاس خود را به عنوان جامعه آماری در نظر بگیرید. اندازه آن چقدر است؟ یک نمونه با اندازه ۸ از آن انتخاب کنید. چگونه این نمونه را انتخاب کردید؟ یکی از راه‌هایی که ممکن است به کار برده باشید، آن است که ۸ نفر اول را به ترتیب حروف الفبا از روی دفتر کلاس انتخاب کرده باشید. ممکن است ۸ نفری را که اطراف شما در کلاس می‌نشینند، انتخاب کرده باشید. بعضی مواقع ممکن است روش انتخاب نمونه مشکل‌ساز باشد. ابتدا یادآوری می‌کنیم که هدف ما از نمونه و مطالعه آن شناخت جامعه است. پس نمونه باید به‌قسمی انتخاب شود که بتواند «بیانگر» جامعه باشد. با این مقدمه فرض کنید بخواهید دربارهٔ قد همکلاسی‌های خود مطالعه کنید. حال اگر افراد تیم بسکتبال کلاس را به‌عنوان نمونه انتخاب کرده باشید، نمونه خوبی دربارهٔ قد همکلاسی‌های خود ارائه نکرده‌اید. آیا می‌توانید بگویید چرا؟ دلایل خود را توضیح دهید.

باید روش انتخاب نمونه به‌گونه‌ای باشد که :

- امکان انتخاب هر فرد به‌عنوان عضوی از نمونه امکان‌پذیر باشد.
- قبل از انتخاب نمونه، توانیم با اطمینان بیش‌تر دربارهٔ حضور و یا عدم حضور عده‌ای در نمونه قضاوت کنیم. به عبارت دیگر، هر فرد برای شرکت در نمونه همان‌قدر سهم داشته باشد که دیگران دارند.

نمونه‌ای که ویژگی‌های بالا را داشته باشد نمونه تصادفی ساده می‌گوییم و برای دستیابی به آن‌ها روش‌هایی ذکر شده است که یکی از آن‌ها استفاده از اعداد تصادفی است که ذیلاً توضیح می‌دهیم.

اعداد تصادفی: این عنوان شاید بنا بر آن‌چه که تا به حال خوانده‌ایم، درست نباشد و در واقع همچنین است چون ما عدد تصادفی نداریم. آن‌چه که در این عنوان نهفته است این است که ما عددی را از بین مجموعه‌ای از اعداد با اتکا به روش‌های تصادفی انتخاب کنیم. منظور از روش‌های تصادفی، یعنی روشی که این اطمینان را به ما بدهد که سهم انتخاب عددی از سایرین بیش‌تر و یا کم‌تر نیست. مثلاً اگر از فردی بخواهید که عددی طبیعی از بین اعداد ۱۰ تا ۱۵ انتخاب کند، به ندرت ممکن است عدد انتخاب شده ۱۳ باشد. زیرا اعتقادات مردم دربارهٔ این عدد به گونه‌ای است که کم‌تر ممکن است به آن توجه کنند و معمولاً در انتخاب، اعداد ابتدایی و انتهایی سهم کم‌تری برای انتخاب شدن دارند. پس انتخاب تصادفی به روشی که اراده و سابقه ذهنی انسان در آن دخیل باشد، عملی نیست. از این رو با برنامه‌نویسی‌هایی که در ماشین‌های حساب و کامپیوترها شده است، می‌توان به اعدادی که خارج از کنترل، سلیقه و سابقه ذهنی انسان‌ها باشد دسترسی پیدا کرد. ما ذیلاً به روش دست‌یابی به اعداد تصادفی در ماشین حساب می‌پردازیم.



این روش در اکثر ماشین حساب‌ها یکسان به کار می‌رود. پس از روشن کردن ماشین حساب، دکمه INV (در برخی، دکمه SHIFT) را فشار دهید. سپس دکمه RAN# را فشار می‌دهیم. پس از انجام این اعمال، عددی مثلاً به صورت ۰/۳۵۷ در صفحه ماشین ظاهر می‌شود. حال اگر مراحل فوق را تکرار کنیم اعداد تصادفی دیگری به دست خواهد آمد. اعداد تصادفی که توسط ماشین حساب تولید می‌شوند همواره غیر منفی و کوچکتر از ۱ است.

فرض کنید عدد تصادفی انتخاب شده ۰/۲۷۳ باشد و اندازه جامعه‌ای که می‌خواهیم از آن نمونه را انتخاب کنیم ۱۵۰ باشد. در این صورت:

$$0/273 \times 150 = 40/95$$

قسمت اعشاری عدد را حذف کرده و یک واحد به آن اضافه کنید. ۴۱، یکی از اعدادی است که از بین اعداد از ۱ تا ۱۵۰ به تصادف انتخاب شده است.

انتخاب نمونه: به مسئله انتخاب ۸ نفر از بین همکلاسی‌های خود باز می‌گردیم. تعداد افراد

کلاس را می‌دانید، مثلاً ۳۵ نفر. به این افراد می‌توانید خودتان یک شماره از ۱ تا ۳۵ اختصاص دهید و یا آن که همان شماره‌ای را که در دفتر کلاس دارند، در نظر بگیرید. حال با استفاده از ماشین حساب یک عدد تصادفی تولید کنید. مثلاً ۰/۱۲۴. در این صورت با ضرب این عدد در ۳۵ خواهیم داشت $۰/۱۲۴ \times ۳۵ = ۴/۳۴$. با حذف قسمت اعشاری و اضافه کردن ۱ به ۴ فرد شماره ۵ به‌عنوان عضوی از نمونه انتخاب خواهد شد. این عمل را ۸ بار تکرار می‌کنیم. جدول زیر اعداد تصادفی که توسط کامپیوتر تولید شده است را نشان می‌دهد.

جدول ۱

شماره	۱	۲	۳	۴	۵	۶	۷	۸
عدد تصادفی	۰/۴۹۶	۰/۳۹۹	۰/۹۴۳	۰/۱۰۸	۰/۵۹۳	۰/۱۹۹	۰/۹۷۷	۰/۲۹۸
$\times ۳۵$	۱۷/۳۶	۱۳/۹۶۵	۳۳/۰۰۵	۳/۷۸	۲۰/۷۵۵	۶/۹۶۵	۳۴/۱۹۵	۱۰/۴۳
شماره انتخاب شده	۱۸	۱۴	۳۴	۴	۲۱	۷	۳۵	۱۱

با ضرب کردن در ۳۵ اعداد تصادفی نامنفی و کوچکتر از ۳۵ به‌دست آمده است. بنابراین، نفرات به شماره‌های ۱۸، ۱۴، و ... ۱۱ اعضای نمونه شما را تشکیل خواهند داد. تذکر: ممکن است در تولید اعداد تصادفی این حالت پیش بیاید که دو شماره یکسان به‌دست آید. در این صورت از یکی از این اعداد صرف‌نظر کرده و عمل را یک بار بیش‌تر تکرار می‌کنیم تا در نهایت ۸ عدد متمایز داشته باشیم.

آن چه که از نمونه موردنظر ماست خود افراد یا اشیای انتخاب شده نیستند. قبلاً در تعریف جامعه گفتیم جامعه مجموعه اشیایی است که در یک یا چند موضوع مشترک باشند. نمونه که زیرمجموعه‌ای از جامعه است نیز این ویژگی را دارد. حال که نمونه را انتخاب کرده‌ایم، باید سراغ موضوعات مورد توجه برویم. مثلاً اگر بخواهیم قد افراد کلاس را مطالعه کنیم، باید به اندازه‌گیری قد آن‌ها بپردازیم.

روش‌های جمع‌آوری داده‌ها

شما کلاس خود را به‌عنوان جامعه در نظر گرفتید ۸ نفر از همکلاسی‌های خود را به‌عنوان یک نمونه ۸ تایی با استفاده از اعداد تصادفی انتخاب کردید. موضوع مورد مطالعه شما قد همکلاسی‌ها بود. حال قد همکلاسی‌های خود را اندازه بگیرید. حاصل کار ۸ عدد خواهد بود. این اعداد را که حاصل اندازه‌گیری اعضای نمونه هستند، داده می‌گوییم.

نتایج حاصل از اندازه‌گیری و یا بررسی نمونه را داده می‌گوییم.



داده‌ها به چند طریق جمع‌آوری می‌شوند:

۱- استفاده از داده‌های از پیش تهیه شده

۲- از طریق پرسش: مستقیماً از اشخاص (شفاهی، مصاحبه)

: پرسش‌نامه کتبی

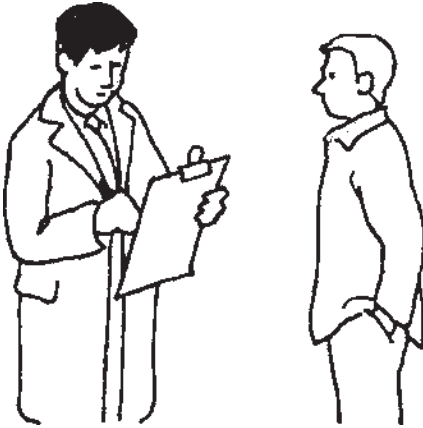


۴- از طریق انجام آزمایش



۳- از طریق مشاهده و ثبت وقایع





طراحی پرسش نامه: پرسیدن سؤال یکی از راه‌های متداول جمع‌آوری داده است که به دو صورت مصاحبه یا پرسش نامه کتبی می‌باشد. برای آن که مطمئن شویم از تمام افرادی که مورد بررسی هستند، سؤالاتی به طور یکسان پرسیده می‌شود، پرسش نامه‌ای به صورت کتبی تهیه می‌شود. از طرفی دیگر در مصاحبه اگر فرد مورد بررسی سؤال را به خوبی متوجه نشود مصاحبه‌گر می‌تواند توضیحات بیش‌تری بدهد.

– قبل از آن که هرگونه سؤالی نوشته شود،

محتوای پرسش نامه باید سازمان‌دهی شود.

– هدف بررسی باید در نظر گرفته شود. چه می‌خواهید

بدانید و چرا؟

– فهرستی از عناوینی که باید راجع به آن‌ها اطلاعات

جمع‌آوری کنید، تهیه کنید.

– از جمع‌آوری داده‌ها و اطلاعات اضافی که مورد نیاز

نمی‌باشند خودداری کنید.



– از سؤالات ساده و کاملاً واضح استفاده کنید. استفاده از عباراتی چون بزرگ، کوچک که

معیار مشخصی ندارند خودداری کنید تا از سؤالات چند برداشت نشود.

– سعی کنید از سؤالاتی که پاسخ آن‌ها «تا حد امکان» یک کلمه‌ای یا اعداد هستند استفاده شود.

– در مورد سؤالاتی که پاسخ‌دهنده ممکن است نخواهد جواب دقیق بدهد مانند سن، درآمد

و ... از پاسخ‌های پیش‌آماده شده چون: کمتر از ۱۶ سال

۱۶ الی ۲۵ سال

۲۶ الی ۳۵ سال

۳۶ الی ۴۵ سال

۴۶ الی ۵۴ سال

بالاتر از ۵۴ سال استفاده کنید.

– از سؤالات هدایت‌کننده استفاده نکنید. به‌طور مثال اگر می‌خواهید برنامه جدید مدرسه را

بررسی کنید، هرگز نپرسید: نظرتان را راجع به برنامه کامل شده مدرسه بیان کنید. عبارت «برنامه کامل

شده» به طور غیرمستقیم در خواننده نظر مثبت القا می کند.

– سعی کنید از سؤالات با پاسخ های چندگزینه ای استفاده کنید.

– اگر جوابی نظر یا عقیده راجع به موضوع یا محصولی هستید، پاسخ ها را به صورت گزینه های

کمی یا کیفی سطح بندی شده انتخاب کنید. مانند: بسیار خوب

خوب

متوسط

ضعیف

بسیار ضعیف

– دستورالعمل کامل و واضح پاسخ گویی به پرسش نامه را ضمیمه کنید.

– در پایان پرسش نامه از پاسخ دهندگان تشکر کنید.

برای به انجام رساندن یک بررسی آماری باید پاسخ سؤالی زیر را بدانید:

- ۱- می خواهید چه چیزی پیدا کنید؟ چرا؟
- ۲- چه داده هایی را باید جمع آوری کنید؟ چگونه تصمیم می گیرید؟
- ۳- کدام راه، بهترین راه جمع آوری داده است؟ چرا؟ آیا به وسیله و ابزار خاصی نیاز دارید؟
- ۴- به چه مقدار داده نیاز دارید؟ چگونه تصمیم می گیرید؟
- ۵- داده های شما باید چقدر دقت داشته باشند؟ چرا؟ آیا می توانید داده هایتان را آزمایش کنید؟
- ۶- داده هایتان را چگونه ثبت می کنید؟ چرا این روش را انتخاب کردید؟
- ۷- داده ها را چگونه ارائه می کنید؟ چرا؟
- ۸- آیا داده ها از الگوی خاصی پیروی می کنند؟ داده ها بیان گر چه می باشند؟
- ۹- چه نتیجه گیری یا پیش گویی می توانید از داده ها بکنید؟ آیا می توانید نتیجه یا پیش گویی خود را آزمایش کنید؟

۱۰- آیا نتایج، با آنچه انتظار داشتید تطبیق دارد؟

۱۱- چگونه نتایج کار را ارائه می کنید؟ برای چه کسانی ارائه می کنید؟

۱۲- با توجه به نتایج به دست آمده، آیا می توانید سؤالات دیگری را نیز مورد بررسی قرار دهید؟

به نظر شما در مرحله جمع آوری داده به کدامیک از سؤالات فوق باید پاسخ داده شود؟

فعالیت

ما معمولاً از یک مشکل زمانی آگاهی پیدا می‌کنیم که یک واقعه غیرقابل قبول یا غیرقابل انتظار گزارش داده شود و اگر این مشکل به اندازه کافی بزرگ باشد در روزنامه‌ها و یا دیگر وسایل ارتباط جمعی منعکس می‌شوند. به مثال زیر توجه کنید:

خبر زیر در یکی از روزنامه‌ها آمده است:

مادری با فرزندش مجبور شدند ۲۰ دقیقه صبر کنند تا از عرض یک اتوبان بگذرند.

اهالی اطراف یکی از بزرگراه‌ها از این که مجبورند زمان زیادی را منتظر بمانند تا با ایمنی از عرض اتوبان بگذرند خسته شده‌اند. این مشکل هفته گذشته به شهرداری منطقه پس از این که مادری مجبور شد ۲۰ دقیقه صبر کند تا با فرزندش از عرض اتوبان عبور کند، گزارش داده شد.....

فرض کنید، شما مسئول بررسی این مشکل هستید. ابتدا باید بدانید مشکل تا چه حد جدی است. سپس باید اطلاعاتی چون چند نفر در روز نیاز به عبور از عرض اتوبان در این محل دارند؟، به طور متوسط چند دقیقه باید صبر کنند تا به طور امن از عرض اتوبان بگذرند؟، آیا همیشه عبور از عرض اتوبان مشکل است یا گاه‌گاه؟، ... جمع‌آوری کنید.

پرسش‌نامه‌ای طرح کنید تا بتوانید این اطلاعات را جمع‌آوری کنید. (آیا به جز پرسش‌نامه می‌توان از روش دیگری استفاده کرد؟)

جامعه در این بررسی چیست؟ و نمونه خود را چگونه انتخاب می‌کنید.

چه راه‌حلهایی برای رفع این مشکل پیشنهاد می‌کنید؟

تمرین

- ۱- یک نمونه ۸ تایی از بین اعداد ۱ تا ۶۲۵ انتخاب کنید.
- ۲- یک نمونه ۶ تایی از بین اعداد طبیعی ۲۳ تا ۷۸ انتخاب کنید.
- ۳- مجموعه افراد فامیل درجه اول و درجه دوم خود را بنویسید و آن‌ها را شماره‌گذاری کنید. سپس یک نمونه ۴ تایی از این جامعه انتخاب کنید.
- ۴- موضوع‌های زیر را در نظر بگیرید. درباره بهترین روش جمع‌آوری داده برای آزمایش این فرضیه‌ها تصمیم بگیرید:

- بیش تر مردم فکر می کنند «حداکثر سرعت در اتوبان ها باید تعیین شود».
- آبی، رنگ مورد علاقهٔ بیش تر مردم برای ماشین است.
- در زمان مطالعه گوش دادن به موسیقی کلاسیک به یادگیری کمک می کند.
- بیش تر تصادفات اتومبیل ها را رانندگان با سن کم تر از ۲۵ سال موجب می شوند.
- رژیم گرفتن، موجب کاهش هوش می شود.
- ۵- دو موضوع بررسی برای هر یک از روش های جمع آوری داده بیان کنید.
- از طریق مصاحبه
- از طریق مشاهده
- از طریق مطالعهٔ داده های از پیش تهیه شده
- از طریق آزمایش

پروژه

- روی موضوعات زیر فکر کنید. برای هر موضوع که انتخاب می کنید، فهرستی از فرضیات در نظر بگیرید. یک پرسش نامه یا آزمایش طراحی کنید. نمونه ای برای بررسی انتخاب کنید و پرسش نامه را توزیع کنید. گزارشی از چگونگی و روش جمع آوری داده های خود تهیه کنید.
- اثر آلودگی روی کیفیت زندگی
 - در دسترس بودن پارک های بازی کودکان اطراف محل زندگی تان
 - زمانی را که کودکان در گروه های سنی مختلف صرف تماشا کردن تلویزیون می کنند.
 - آزمون هایی که در بعد از ظهر برگزار می شوند نتیجه بهتری از آزمون هایی که صبح ها برگزار می شوند دارند.
 - گل ها هر چه نور خورشید بیش تری دریافت کنند، سریع تر رشد می کنند.
 - هر چه وزن بیش تر باشد، سریع تر می توان وزن کم کرد.
 - گوش دادن به موسیقی کلاسیک، یادگیری را افزایش می دهد.
 - دانش آموزانی که نمره ریاضی بالاتر از ۱۸ دارند، نمرهٔ ادبیات بالای ۱۸ کسب می کنند.



و اگر به دیده دل بنگری بدانچه از بهشت برایت ستایند، دل بر کنی از آنچه در دنیا است، هر چند بدیع و زیباست؛ از خواهش‌های نفسانی، و خوشی‌های زندگانی و منظره‌های آراسته. و فکر تو در می‌ماند در آهنگ و آوای شاخه‌ها و برگ‌های درختان که در کنار نهرهای بهشتی است، و در آویختن خوشه‌های لؤلؤ آبدار بر شاخه‌های آن درختانی که ریشه‌های آن در پشته‌های مشک‌نمان است و رُستن این میوه‌های گوناگون، در غلافها و پوشش‌های درون. شاخه‌ها بی‌رنجی خم گردد و چنانکه چیننده آن خواهد در دسترس او بود. خدای به رحمت خود، ما و شما را از آنان گرداند که به دل کوشد تا خود را به منزل‌های نیکوکاران رساند.



متغیرهای تصادفی

مقدمه

تلویزیون یکی از رسانه‌های دیداری است که نقش بزرگی در فرهنگ هر جامعه‌ای دارد. تأثیر تلویزیون در آموزش مدرسه‌ای یک مسئله مهم است که می‌توان به کمک آمار آن را بررسی کرد. معلم بررسی نقش تلویزیون در تحصیل دانش‌آموزان را به‌عنوان پروژه گروهی پیشنهاد کرد. یک گروه از دانش‌آموزان روی این مسئله کار کردند. با هم روند کار علی یکی از اعضای گروه را مرور می‌کنیم:

علی برنامه‌های شبکه‌های سیمای جمهوری اسلامی ایران را به ۱۰ دسته تقسیم نمود که عبارتند از: فرهنگی و اجتماعی، ورزشی و تفریحات سالم، معارف اسلامی، اطلاعات عمومی، هنر و ادب، سیاسی، اقتصادی، علم و فن، ارزش‌های دفاع مقدس، و در آخر برنامه‌های تاریخی. او با یک نمونه‌گیری از برنامه‌های سه ماه متوالی، تعداد ساعتی را که هر شبکه به هر یک از موضوعات بالا اختصاص داده است، تخمین زد. علی برای این که مطمئن شود تعداد اعضای نمونه او کافی بوده است، بار دیگر نمونه‌ای به همان تعداد اول گرفت و نتایج به دست آمده را با نتایج اولیه مقایسه کرد. چون این نتایج اختلاف چندانی نداشتند نتیجه گرفت که نمونه او به تعداد کافی عضو داشته که بتواند نمایانگر تمام برنامه‌های تلویزیون باشد. اما این سؤال برای او پیش آمد که چرا نتایج دو نمونه‌گیری او «دقیقاً» منطبق بر هم نیستند؟ ما می‌دانیم که اطلاعات هر عضو نمونه، متغیری است که به عمل تصادفی نمونه‌گیری بستگی دارد. به همین دلیل موضوع مورد مطالعه را متغیر تصادفی می‌گوییم. چون مقدار آن پس از یک آزمایش تصادفی به دست می‌آید.

آنچه بین نتایج دو نمونه‌گیری تصادفی مناسب مشترک است الگوهای آماری است که در هر دو نمونه برقرار هستند. نتایج دو نمونه‌گیری تصادفی هرگز دقیقاً مساوی نیستند. حال دقت کنید به این که آیا علی واقعاً موضوع مورد سؤال را بررسی کرده است؟ بهتر بود علی پیش از جمع‌آوری داده‌ها با هم گروهی‌های خود مشورت می‌کرد!

در زیر میزان تولید برنامه در شبکه‌های سیمای جمهوری اسلامی ایران به نقل از گزارش‌های سالانه مرکز آمار آورده شده است.

جدول ۱ - میزان تولید برنامه در شبکه‌های سیمای جمهوری اسلامی ایران (ساعت)

شبکه پنجم (تهران)	شبکه چهارم	شبکه سوم	شبکه دوم	شبکه اول	جمع	سال و برنامه
۱۳۴۴	۱۱۴۱	۳۵۳۹	۲۵۹۶	۳۱۶۰	۱۱۸۷۶	۱۳۷۶
۵۹۴	۲۸۳	۱۱۲۳	۹۷۵	۱۸۳۱	۴۸۰۵	فرهنگی و اجتماعی
۳۶	۰	۲۰۷۶	۴۶۳	۲۱۰	۲۷۸۴	ورزش و تفریحات سالم
۶۳	۲۳۷	۱۰۴	۲۹۹	۲۵۷	۹۶۰	معارف اسلامی
۱۹۳	۲۰۹	۱۸۱	۴۲۶	۲۰۱	۱۲۰۹	اطلاعات عمومی
۲۸۵	۱۰۲	۱۸	۱۰۶	۰	۵۱۰	هنر و ادب
۲۸	۵۵	۲۹	۱۶۷	۱۷۲	۴۵۰	سیاسی
۸۷	۱۳۷	۰	۵۴	۱۵۹	۴۳۷	اقتصادی
۹	۱۹۷	۰	۷۰	۴۲	۳۱۸	علم و فن
۲۳	۲۰	۸	۲۷	۲۲۹	۳۰۷	ارزش‌های دفاع مقدس
۲۶	۱	۰	۹	۵۹	۹۶	تاریخی

متغیر تصادفی

در فصل قبل با جامعه آماری و نمونه آشنا شدید. دیدیم برای آن که نمونه بتواند بیان‌گر جامعه باشد، باید به دور از نظر شخصی و سلیقه‌های فردی انتخاب شود، و برای این منظور از اعداد تصادفی استفاده کردیم و گفتیم حال که نمونه به دست آمده است، باید موضوع و یا موضوعات مورد نظر را بررسی کنیم. این موضوع و یا موضوعات را **متغیر تصادفی** می‌نامیم. تا به حال با متغیرهای تصادفی زیادی سروکار داشته‌اید بدون آن که آن‌ها را با این نام بشناسید. قد دانش‌آموزان یک کلاس، سابقه کار دبیران ریاضی مدرسه شما، تعداد بیماران مراجعه‌کننده به یک درمانگاه، میزان درآمد افراد، تعداد کلماتی که در هر بیت از غزل‌های حافظ به کار رفته است، متغیرهای تصادفی هستند. حتی رنگ چشم افراد و یا گروه خونی متغیرهای تصادفی هستند. مثال‌های دیگری از متغیرهای تصادفی عبارتند از:

- درجه حرارت هوا در ساعت ۱۰ صبح روزهای آبان ماه سال جاری.
- تعداد افراد خانواده همکلاسی‌های شما.

– مرحله زندگی افرادی که برای گردش به یک پارک محلی آمده‌اند (مراحل زندگی عبارت است از: نوزادی، کودکی، نوجوانی، جوانی، میان‌سالی و پیری).

– گروه خونی همکلاسی‌های شما.

برخی از متغیرها قابل اندازه‌گیری‌اند، مانند قد، وزن، درجهٔ حرارت، درآمد خانوار، میزان بارندگی در سال در یک منطقه، میزان آلودگی هوا، شدت زلزله و غیره. آن‌چه که در این متغیرها مهم است آن است که اگرچه ما نمی‌دانیم که مقدار حاصل از اندازه‌گیری چه قدر خواهد شد ولی می‌دانیم عدد حاصل در مجموعه‌ای خواهد بود. مثلاً قد نوجوانان می‌تواند از ۱۲۰ سانتی‌متر شروع و به ۲۰۰ ختم شود. درجه حرارت در منطقه‌ای در فصل معینی باز دامنه معلومی دارد.

برخی دیگر از این متغیرها قابل شمارشند. مانند تعداد افراد خانواده، تعداد زنبورهای یک کندو، تعداد غایبین کلاس، تعداد تصادف‌های رانندگی در یک روز و در یک شهر، تعداد اتومبیل‌هایی که در یک روز به علت تخلف رانندگی جریمه شده‌اند و تعداد تخم‌مرغ‌هایی که فردی در طول یک ماه مصرف می‌کند. این قبیل متغیرها مانند حالت قبل این ویژگی را دارند که اگرچه تعداد مورد نظر از قبل برای ما معلوم نیست ولی می‌توانیم محدوده‌ای برای آن مشخص کنیم.

برخی از دیگر متغیرها قابل اندازه‌گیری نیستند مانند گروه خونی، بلکه این متغیرها به نوع و یا دسته خاصی تعلق دارند، در این‌جا هم اگرچه از قبل نمی‌توانیم بگویم که گروه خونی افراد انتخاب شده کدام نوع خواهد بود ولی مجموعهٔ انواع گروه خونی برای ما معلوم است (A, B, AB و O) مثال‌های دیگری از انواع متغیرهای غیرقابل اندازه‌گیری عبارتند از: کشت که ممکن است یکی از دو نوع دیم و یا آبی باشد، جنسیت افراد که ممکن است یکی از دو نوع زن و یا مرد باشد، کیفیت کالای تولید شده به وسیله کارخانه‌ای که ممکن است سالم و یا معیوب باشد، فردی از یک شهر ممکن است باسواد و یا بی‌سواد باشد و یا همچنین این فرد ممکن است شغل دولتی داشته باشد و یا شغل آزاد.



همان طوری که دیدیم متغیرها عموماً از دو نوع بودند یکی متغیرهایی که اندازه گیری می شدند و عدد به آن ها نسبت داده می شد که این متغیرها را متغیرهای کمی می گوئیم.

مثال هایی از متغیرهای کمی عبارتند از: وزن، فاصله، طول، جمعیت، خسارت مالی ناشی از تصادفات، تعداد کسانی که در یک کلاس معدل بالای ۱۵ دارند، تعداد درس هایی که شما بالاتر از ۱۵ دارید. از ویژگی های یک متغیر کمی آن است که می توانیم افراد را از نظر این متغیر مقایسه کنیم.

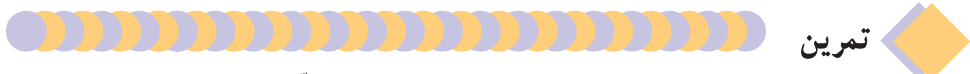
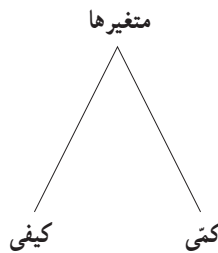
متغیرهای کمی متغیرهایی هستند که قابل اندازه گیری اند.

نوع دیگری از متغیرها، آن هایی هستند که اندازه گیری نمی شوند، فقط نوع آن ها تعیین می شود، مانند گروه خونی، کشت و مراحل زندگی یک فرد که این نوع متغیرها را متغیرهای کیفی می گوئیم.

متغیرهای کیفی متغیرهایی هستند که قابل اندازه گیری نیستند.

بنابراین ما متغیرها را از نظر این که قابل اندازه گیری باشند و یا نباشند به صورت زیر دسته بندی

کردیم.



تمرین

۱- چهار متغیر تصادفی نام ببرید و نوع هر یک را از نظر کمی و کیفی بودن تعیین کنید.

۲- فهرستی از متغیرهای تصادفی که روزانه با آن ها سروکار دارید تهیه کنید. آیا در این فهرست حجم غذای مصرفی را منظور کرده اید، آیا میزان کالری دریافتی از این غذا را در نظر گرفته اید، میزان پروتئین را چطور؟ مدتی را که صرف تماشای تلویزیون می کنید ثبت کرده اید، مدتی را که به ورزش می پردازید، مدتی را که به مطالعه درس های فردا می پردازید، مدتی را که مطالعه آزاد می کنید و... به نظر می رسد این فهرست بی انتها می باشد. ما به وسیله متغیرهای تصادفی زیادی احاطه شده ایم.

انواع متغیرهای تصادفی: در بخش قبلی متغیرهای تصادفی معرفی شدند و آن‌ها را به دو دسته عمده کمی و کیفی تقسیم کردیم. اگر مثال‌های مربوط به متغیرهای کمی را در نظر بگیرید ملاحظه خواهید کرد که بعضی از این متغیرها می‌توانند هر مقداری بین دو عدد را اختیار کنند، مثلاً وزن. اگر وزن فردی در کلاس شما ۶۰ کیلو و وزن یکی دیگر ۶۶ کیلو باشد، هر عددی را که بین ۶۰ و ۶۶ در نظر بگیرید (اعم از طبیعی و یا غیرطبیعی) دلیلی نداریم که وزن فردی نتواند برابر آن مقدار باشد. یعنی وزن می‌تواند به‌طور پیوسته تمام مقادیر واقع در بازه‌ای را بگیرد. این قبیل متغیرها را متغیرهای کمی پیوسته می‌گوییم.

متغیر پیوسته، یک متغیر کمی است که اگر دو مقدار a و b را بتواند اختیار کند هر مقدار بین آن‌ها را نیز بتواند اختیار کند.

در مقابل، متغیرهای کمی دیگری هستند که ویژگی بالا را ندارند، مثلاً تعداد غایبین کلاس که ممکن است امروز ۵ نفر و دیروز ۳ نفر باشد، در این کلاس ممکن است روزی تعداد غایبین برابر ۴ نفر باشد ولی نمی‌تواند این تعداد برابر $4/5$ شود. این قبیل متغیرها را متغیرهای کمی گسسته می‌گویند.

به متغیر کمی که پیوسته نباشد، گسسته می‌گوییم.

معمولاً متغیرهای گسسته از نوع تعداد هستند مانند تعداد تصادفات، تعداد کالاهای سالم، تعداد روزهای بارانی در یک شهر و غیره، ولی ممکن است متغیرهایی نیز مثال زد که از نوع تعداد نباشند. مثلاً در ساختمان‌سازی، معمولاً ساختمان را در چند طبقه بنا می‌کنند ولی گاهی اوقات ممکن است یک طبقه را کامل نسازند که در این صورت به آن نیم طبقه گفته می‌شود. از این رو اصطلاح «ساختمانی در دو و نیم طبقه و یا سه و نیم طبقه» رایج است، در این صورت اگر متغیر تصادفی شما تعداد طبقات باشد، این متغیر مقادیر ۱ و $1/5$ و ۲ و $2/5$ و غیره را می‌تواند اختیار کند.

– بیش‌تر میوه‌ها بر اساس وزن که یک متغیر پیوسته است خرید و فروش می‌شوند. میوه‌هایی را مثال بزنید که بر اساس وزن خرید و فروش می‌شوند.

متغیرهای کیفی را نیز می‌توان به دو دسته تقسیم کرد. قبلاً در ویژگی‌های متغیرهای کمی گفتیم که می‌توانیم افراد را از نظر متغیرهای کمی مقایسه کنیم. مثلاً بگوییم فردی از فرد دیگری بلند قدتر است، سنگین‌تر است، این کلاس بیش‌تر از کلاس دیگر غایب دارد. امروز آلودگی هوا کم‌تر از دیروز است و غیره.

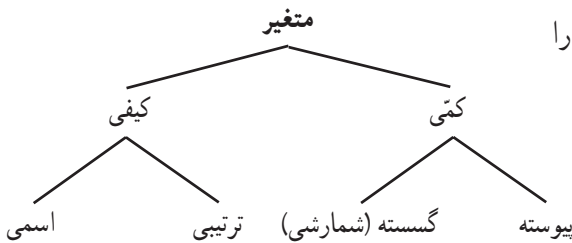
اما چنین مقایسه‌ای را به‌وسیله متغیرهای کیفی نمی‌توانیم انجام دهیم، دو فرد از نظر گروه

خونی مقایسه نمی‌شوند فقط می‌توان گفت که آیا این دو از نظر گروه خونی یکی هستند یا خیر. علت این امر در این است که در اعداد رابطه بزرگ تری و کوچک تری تعریف شده است اما مثلاً در گروه خونی چنین رابطه‌ای وجود ندارد. اما بعضی از متغیرهای کیفی از نوعی ترتیب طبیعی برخوردارند مثلاً مراحل زندگی که در آن مسلماً نوزادی قبل از کودکی، کودکی قبل از نوجوانی و الی آخر است و یا مراحل تحصیل که در آن‌ها دبستان قبل از راهنمایی و راهنمایی قبل از دبیرستان و غیره است. این قبیل متغیرها را متغیرهای کیفی ترتیبی می‌گوییم.

متغیرهایی کیفی‌ای که در آن‌ها نوعی ترتیب طبیعی وجود دارد متغیرهای کیفی ترتیبی می‌گوییم.

متغیرهایی کیفی نیز وجود دارند که در آن‌ها این ترتیب ملاحظه نمی‌شود، مانند گروه خونی، RH، خون. این قبیل متغیرها را متغیر کیفی اسمی می‌گوییم.

متغیر کیفی که ترتیبی نباشد متغیر کیفی اسمی می‌گوییم.



حال می‌توانیم تقسیم‌بندی متغیرها را به صورت روبرو تکمیل کنیم.

- غیر از مثال‌های ارائه شده در متن، برای هر یک از انواع متغیرها مثالی ارائه کنید.
- چند متغیر کیفی در توصیف میوه‌ها نام ببرید.



تمرین



- ۱- کدام یک از متغیرهای زیر گسسته و کدام یک پیوسته اند؟
 - تعداد مکالمات تلفنی یک اداره در یک روز.
 - طول مکالمات تلفنی یک اداره.
 - تعداد نامه‌های یک صندوق.
 - وزن نامه‌های موجود در یک صندوق.
 - زمانی که یک بیمار در اتاق انتظار مطب یک پزشک منتظر است.
 - تعداد بیماران مراجعه کننده به یک پزشک در طول روز.

- ۲- کدام یک از متغیرهای زیر کمی و کدام یک کیفی هستند؟
 - نوع تلفن مورد استفاده شهروندان (تلفن سیار و یا تلفن ثابت).
 - جنسیت افراد یک شهر (زن، مرد).
 - میزان تحصیلات افراد یک شهر (دیپلم، کارشناسی، کارشناسی ارشد، دکترا).
 - وضع سواد (باسواد، بی سواد).
 - وضعیت مسکن (صاحب مسکن، بدون مسکن ملکی).
 - میزان اجاره پرداختی به وسیله شهروندان.
 - میزان پرداخت مالیات سالانه ساختمان‌های مسکونی.
 - میزان آلودگی هوا.
 - میزان بارندگی در یک شهر در طول سال.

- ۳- نوع هر یک از متغیرهای زیر را مشخص کنید :
 - رنگ اتومبیل‌های موجود در یک نمایشگاه اتومبیل.
 - مقاومت یک ترانزیستور.
 - گنجایش آب یک تانکر.
 - تعداد شکایات رسیده شده به یک پاسگاه پلیس.
 - درآمد دانشجویان شاغل به کار.
 - وضعیت تأهل کارمندان یک شرکت.
 - سن دانشجویان شرکت کننده در یک دوره هنری.

۴ - کدام یک از نمونه‌گیری‌های زیر، یک نمونه‌گیری تصادفی است؟ در هر یک جامعه و نمونه را مشخص کنید.

- با تمام پلیس‌های یک پاسگاه پلیس برای پیدا کردن نظر پلیس‌های این پاسگاه راجع به تخلفات مصاحبه شد.

- با بچه‌هایی که وارد یک پارک بازی می‌شدند پنج در میان مصاحبه شد تا وسیله بازی مورد علاقه کودکان را مورد بررسی قرار دهیم.

- برای بررسی pH شامپوهای تولیدی یک کارخانه، شامپوها را صد در میان مورد آزمایش قرار دادیم.

۵ - برای تحقیقات و بررسی‌های زیر جامعه را مشخص کرده و روش نمونه‌برداری پیشنهاد کنید.

- پیدا کردن درصد بیکاران استان مازندران.

- مردان کدام شامپو را بیش‌تر ترجیح می‌دهند.

۶ - می‌خواهیم مدت زمانی را که دانش‌آموزان کلاس شما در طول یک هفته صرف مطالعه

کتاب‌های غیردرسی می‌کنند بررسی کنیم.

- در این بررسی جامعه را مشخص کنید.

- یک روش نمونه‌گیری برای انتخاب نمونه معرفی کنید.

- از چه روشی برای جمع‌آوری اطلاعات استفاده می‌کنید.

- برای این بررسی یک نمونه مشخص کنید.

- اندازه این جامعه چه قدر است؟

- اندازه نمونه مورد بررسی چه قدر است؟ آیا این تعداد برای بررسی مورد نظر مناسب است؟

در صورت مناسب نبودن، روش نمونه‌برداری و نمونه خود را اصلاح کنید.

- متغیر مورد مطالعه در این مسئله چیست؟ توضیح دهید.

- متغیر مورد مطالعه از چه نوعی است؟

پروژه

با افزایش جمعیت و با توجه به محدود بودن منابع آب، ضرورت صرفه‌جویی در مصرف

آب احساس می‌شود. برای مطالعه در این مسئله مدلی را معرفی کنید.

لقمان و مرد پیاده



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 و رفتن را پیشه کرد. زمانی که چند قدمی راه رفته بود، 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. حال که دیدم دانستم که تو یک ساعت دیگر به ده خواهی رسید.

در این داستان ساده و قدیمی تمام اصول آماری رعایت شده است.

جامعه؛ مجموعه افرادی است که از محل زندگی لقمان به ده می‌روند.

متغیر تصادفی؛ مدت زمانی که طول می‌کشد تا فردی به ده برسد.

نمونه؛ این مرد نمونه‌ای از جامعه افرادی است که به ده می‌روند و چند قدمی که مرد راه می‌رود

نمونه‌ای است از جامعه قدمهای فرد تا شهر تا لقمان بتواند سرعت شخص را تخمین بزند.

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.

سر لوحهٔ دین شناختن اوست؛
و درست شناختن او، باور داشتن
او؛ و درست باور داشتن او؛ یگانه
انگاشتن او؛ و یگانه انگاشتن او، او
را بسزا اطاعت نمودن، و بسزا
اطاعت نمودن او، صفتها را از او
زدودن، چه هر صفتی گواه است که
موصوف دو تاست و هر موصوف
نشان دهد که از صفت جداست،
پس هر که پاک خدای را با صفتی
همراه داند او را با قرینی پیوسته،
و آن که با قرینش پیوندد، دو تایش
دانسته؛ و آن که دو تایش خواند،
جزء جزء اش داند؛ و آن که او را
جزء جزء داند، او را نداند؛ و آن که
او را نداند در جهتش نشاند؛ و آن
که در جهتش نشاند، محدودش
انگارد؛ و آن که محدودش انگارد،
معدودش شمارد.

دسته‌بندی داده‌ها و جدول فراوانی

تاکنون با جامعه، نمونه، متغیر تصادفی و داده‌ها آشنا شدیم. پس از جمع‌آوری داده‌ها تعدادی عدد، مثلاً ۵۰ عدد در اختیار داریم. این اعداد داده‌های خام هستند. نگه‌داری این داده‌ها و استفاده از آن‌ها اگرچه جزئی‌ترین اطلاعات را دربر دارد ولی در مسایل بزرگ کنار آمدن با این همه اطلاعات جزئی غیرممکن و گیج‌کننده است. باید از بین این اطلاعات ریز و درشت یک طرح کلی از نمونه و نهایتاً از جامعه‌ای که این اعداد و ارقام از آن آمده‌اند، به‌دست آورد.

برای آن که موضوع مورد بحث ما روشن شود مثال زیر را در نظر بگیرید. شما در شهر و مکانی که زندگی می‌کنید اطراف خود را به‌خوبی می‌شناسید. نام کوچه‌ها، خیابان‌ها، مراکز خرید و حتی در بعضی از فروشگاه‌ها، مکان اجناس را نیز می‌دانید. حال اگر بخواهید محله خود را به کسی که با آن آشنایی ندارد، معرفی کنید ارائه این اطلاعات ریز نه مورد توجه است و نه آن که طرحی مناسب از محل زندگی شما به او می‌دهد. آن چه که با ارائه این اطلاعات نصیب شنونده خواهد شد، تصویری درهم از محل زندگی شماست. شما می‌توانید با ارائه شاخص‌های مناسبی از یک محله، چگونگی محل زندگی خود را معرفی کنید. مثلاً بگویید:



- مراکز خرید «متعدد» و در «دسترس» است.
- قیمت اجناس در آن «مناسب» است.
- مراکز تفریحی و فرهنگی «خوبی» دارد.
- میزان آلودگی هوا در آن «کم» است.
- سروصدای شهر در آن جا «زیاد» نیست.

این اطلاعات به شنونده احساسی از مکان زندگی شما می‌دهد و همین‌ها کفایت می‌کند. سایر اطلاعات جز آن که ایجاد اخلاص در سیستم اطلاع‌رسانی شما بکنند، فایده و تأثیر دیگری ندارد. در شاخص‌های معرفی شده بالا برخی از عبارات با نماد « \gg » مشخص شده‌اند. شما ممکن است ایراد بگیرید که اگر قرار است اظهار نظر علمی باشد باید با عدد و رقم همراه باشد. چگونه است که در

هیچ یک از این اظهارنظرها عدد و رقم دیده نمی شود. حق با شماست. ما باید با مطالعات قبلی تعریف کرده باشیم که مراکز فروش در «دسترس» هستند یعنی چه؟ میزان آلودگی هوا «کم» است یعنی چه؟ این عبارات بایستی به کمک اعداد و ارقام تعریف شوند. به خاطر سپردن اعداد و ارقام مخصوصاً زمانی که تعداد آن‌ها زیاد باشد، کار آسانی نیست. عبارات توصیفی را بهتر می توان به ذهن سپرد. از این رو برای انتقال سریع اطلاعات و نه برای منظورهای تخصصی، روش توصیفی مناسب است ولی این بیان توصیفی نمی تواند مبنایی برای تصمیم گیری های جدی شود. در حال ما ناگزیر هستیم برخی از اطلاعات را کنار بگذاریم. برخی دیگر را یک کاسه کنیم تا بتوانیم جامعه مورد مطالعه را معرفی کنیم. این معرفی شبیه آن است که با چرخ بال از بالای شهر به آن نگاه کنیم. از این طریق نه وارد جزئیات شده ایم و نه آن که ویژگی های اصلی شهر را نادیده گرفته ایم.

جدول فراوانی

معمولاً داده ها در قالب یک جدول داده می شوند. این داده ها به علت طبیعت ظاهری نامنظم آن ها گویای مطلبی درباره جامعه نیستند. برای آن که بتوان به آن ها نظم بهتری داد جدول های مناسبی تنظیم می شوند. برای آن که موضوع بحث را بهتر ادامه دهیم، به مثال های زیر توجه کنید:

مثال: برای مطالعه تعداد افراد خانوارهای یک شهرک، یک نمونه ۴۰ تایی از خانوارهای این شهرک را انتخاب کرده و از آن ها تعداد افراد خانوار را پرسیده ایم. جدول زیر به دست آمده است:

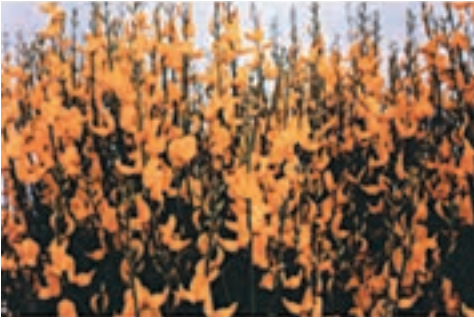
۳, ۴, ۲, ۱, ۲, ۵, ۵, ۳, ۲, ۱
 ۶, ۲, ۴, ۴, ۳, ۳, ۱, ۴, ۳, ۵
 ۷, ۳, ۲, ۴, ۵, ۵, ۴, ۲, ۶, ۱
 ۳, ۳, ۴, ۲, ۴, ۵, ۶, ۳, ۷, ۸

برای آن که بتوانیم با نظم بخشیدن به داده های موجود اطلاعات بهتری کسب کنیم آن ها را در جدولی به صورت زیر خلاصه می کنیم:

تعداد افراد خانوار	۱	۲	۳	۴	۵	۶	۷	۸	جمع
تعداد خانوارها	۴	۷	۹	۸	۶	۳	۲	۱	۴۰

این جدول نشان می دهد که در نمونه، ۴ خانوار ۱ نفره، ۸ خانوار ۲ نفره و فقط ۱ خانوار ۸ نفره بوده است.

چون نمونه را به گونه ای انتخاب کرده ایم که بیان گر جامعه باشد می توانیم بگوییم در این شهرک خانوارهای کمی ۸ نفره هستند و بیش تر خانوارها ۳ نفره هستند. بعد از آن خانوارهای ۴ نفره



بیشترین تکرار را دارند. اعداد سطر دوم در جدول صفحه قبل نشان می‌دهد که هر یک از اعداد سطر اول چند بار تکرار شده‌اند. این تکرارها را فراوانی مطلق می‌گوییم.

فراوانی مطلق داده x_i برابر تعداد دفعاتی است که آن داده تکرار شده است. فراوانی مطلق داده x_i را با نماد f_i نشان می‌دهیم.

در مثال صفحه قبل $f_1 = 4$ ، $f_5 = 6$ و $f_8 = 1$ است. در این جدول می‌بینیم که $f_1 + f_2 + \dots + f_8 = 40$. آیا می‌توانید توضیح دهید که این برابری همواره برقرار است؟ یعنی مجموع فراوانی‌ها برابر تعداد افراد و یا اشیایی است که مورد مطالعه قرار گرفته‌اند. البته می‌توانستیم بنا بر تقسیمات رایج، خانوارها را به صورت زیر دسته‌بندی کنیم:

- خانوار کم جمعیت؛ یعنی خانواری که جمعیت آن یک یا دو نفر باشد،
 - خانوار با جمعیت متوسط؛ یعنی خانواری که جمعیت آن ۳، ۴ و یا ۵ نفر باشد،
 - خانوار پر جمعیت؛ یعنی خانواری که جمعیت آن بیش از ۵ نفر باشد.
- در این صورت می‌توانستیم جدول زیر را برای داده‌های بالا تنظیم کنیم:

جمع	پر جمعیت	با جمعیت متوسط	کم جمعیت	نوع خانوار
۴۰	۶	۲۳	۱۱	تعداد خانوارها

این جدول نشان می‌دهد که بیش‌تر خانوارها دارای جمعیت متوسط هستند و تعداد خانوارهای پر جمعیت از تعداد خانوارهای کم جمعیت، کم‌تر است.

مجدداً در این جدول هم اعداد ۱۱، ۲۳ و ۶ را فراوانی مطلق می‌گوییم. در نمونه تحت بررسی خانواده کم جمعیت ۱۱ بار تکرار شده است.

برای توزیع کالا برگ جدول اول مناسب است و حال آن‌که در ساخت و ساز ساختمان‌های مسکونی اگر چه جدول اول هم می‌تواند مفید باشد ولی معمولاً جدول دوم مبنا قرار می‌گیرد.



دسته بندی

اولین اقدامی که در مطالعهٔ یک جامعه براساس داده‌ها انجام می‌دهیم، آن است که ببینیم آیا می‌توان جامعه را به چند دسته جدا از هم تفکیک کرد؟ البته ما هر مجموعه‌ای را می‌توانیم به چند دسته تقسیم کنیم، ولی منظور ما از تشکیل دسته آن است که اعضای موجود در یک دسته، آشنا و یک‌دست باشند. در واقع یک‌دسته همگن، همگون و متناسب داشته باشیم و تفاوت بسیار در آن دیده نشود. مثلاً از نظر آب و هوا، ایران را می‌توان به چند ناحیه مشخص تقسیم کرد که در هریک از این نواحی آب و هوا یکسان است. به‌طور مثال شمال ایران، استان‌های مازندران و گیلان علی‌رغم وسعت زیاد آن، دارای آب و هوای تقریباً یکسان است. از غرب گیلان تا شرق مازندران تفاوت زیادی در آب و هوا ملاحظه نخواهید کرد. شهرهای حاشیه کویر ایران نیز از لحاظ آب و هوا یکسانند. نواحی جنوبی ایران نیز آب و هوای خاص خود را دارد. غرب ایران آب و هوایی دارد که با آب و هوای شرق ایران متفاوت است. در هر حال ملاحظه می‌شود که ایران از لحاظ آب و هوا به چند دستهٔ مشخص تقسیم می‌شود. به‌همین ترتیب مردم ایران را می‌توان به نژادها و زبان‌های چندی دسته‌بندی کرد. این دسته‌بندی‌ها، جامعه ما را به چند جزء کوچک‌تر که مثل هم هستند تقسیم می‌کند. پس از این دسته‌بندی، مسئله مهم یافتن اندازهٔ این دسته‌هاست و مطالعهٔ جامعه روال آسان‌تری پیدا می‌کند. پس از این مقدمه نسبتاً طولانی به معرفی جدول فراوانی می‌پردازیم.

در مطالعه متغیرهای گسسته جدول فراوانی بدون دسته‌بندی کارآمد است، ولی اگر متغیر پیوسته باشد این روش برای منظم کردن داده‌ها عملی نیست. لذا برای تنظیم داده‌ها در این قبیل موارد مراحل زیر را طی می‌کنیم:

دامنه تغییرات: اولین نکته‌ای که در این مرحله مدنظر قرار می‌گیرد آن است که بینیم آیا متغیر در فاصله زیادی تغییر می‌کند یا در فاصله کمی قرار می‌گیرد. طول بازه‌ای را که متغیر در آن تغییر می‌کند «دامنه تغییرات» می‌گوییم. برای بیان روشن‌تر، فرض کنید:

کوچک‌ترین داده = a و بزرگ‌ترین داده = b

در این صورت تفاضل a از b یعنی $b - a$ را دامنه تغییرات می‌گویند و با علامت R نشان می‌دهیم.

$$R = b - a$$

بزرگی دامنه تغییرات نشان‌دهنده تفاوت زیاد در جامعه است. هرچقدر این دامنه کمتر باشد افراد جامعه از لحاظ این متغیر به هم نزدیک‌ترند. اگر دامنه تغییرات صفر باشد، تمام افراد باهم برابر و یکسانند. در زیر سن چند تن از معلمان یک دبیرستان آمده است:

۳۰، ۴۵، ۶۲، ۳۷، ۴۲، ۴۸

دامنه تغییرات را در این داده‌ها محاسبه می‌کنیم:

$$a = 30 \text{ و } b = 62$$

$$R = b - a = 62 - 30 = 32$$

پس:

فعالیت

دامنه تغییرات قد را در بین همکلاسی‌های خود حساب کنید.

در فعالیت بالا ممکن است این دامنه تغییرات را براساس قد تمام همکلاسی‌های خود حساب کنید (منظور فعالیت نیز همین است) ولی ممکن است شما فکر کنید که ما یک نمونه ۸ تایی از این جامعه داریم اگر دامنه تغییرات را در نمونه حساب کنیم چه خواهد شد؟ چه ارتباطی بین این دو دامنه تغییرات وجود دارد؟ برای روشن شدن مطلب به سؤال‌های زیر پاسخ دهید.

- ۱- از این دو دامنه تغییرات کدام بزرگ‌تر و کدام کوچک‌تر است؟
- ۲- آیا زیاد بودن دامنه تغییرات در نمونه، زیاد بودن دامنه تغییرات در جامعه را نتیجه می‌دهد؟
- ۳- آیا کوچک بودن دامنه تغییرات در نمونه، کوچک بودن دامنه تغییرات در جامعه را نتیجه

می‌دهد؟

اگر دامنه تغییرات به نظر کوچک باشد، دیگر نیازی به دسته‌بندی جامعه نیست زیرا کوچک بودن دامنه تغییرات می‌گوید که اندازه‌های اعضا به هم نزدیک‌اند و اختلاف آن‌ها قابل ملاحظه نیست پس با یک جامعه کاملاً یک‌دست سروکار داریم. اما در مواردی که تغییرات زیاد است، جامعه را نمی‌توانیم یک‌دست فرض کنیم و لذا باید آن را دسته‌بندی کنیم.

بحث کلاسی

– در کلاس خود متغیری را مثال بزنید که کلاس از این نظر یک‌دست باشد.
– در کلاس خود متغیری را مثال بزنید که کلاس از نظر این متغیر یک‌دست نباشد.

حال فرض کنید دامنه تغییرات زیاد باشد مانند میزان بارندگی در شهرهای مختلف ایران. این متغیر در کشور ما دامنه تغییرات زیادی دارد. دامنه تغییرات بارندگی در شمال کم است و به همین ترتیب در شهرهای حاشیه کویر. ابتدا دامنه تغییرات را به چند قسمت کوچک‌تر تقسیم می‌کنیم. سپس براساس این قسمت‌بندی، داده‌ها را در دسته‌های مختلف توزیع می‌کنیم. تعداد دسته‌ها باید به گونه‌ای باشد که در هر دسته تغییرات کم باشد. برای روشن بودن مطالب و ادامه بحث، موضوع را با یک مثال پی می‌گیریم.



داده‌های زیر یک نمونه ۲۵ تایی از نمرات ریاضی یک کلاس است.

۲۰، ۱۸/۵، ۱۹، ۱۵، ۱۳/۵، ۱۴/۲۵، ۱۰، ۱۱/۵، ۱۵/۵، ۱۲/۲۵، ۱۱، ۱۵/۵، ۱۷/۵

۱۵/۵، ۱۳/۵، ۱۴/۵، ۱۱/۵، ۱۱، ۱۶/۵، ۱۷/۲۵، ۱۸/۵، ۱۸، ۱۲/۵، ۱۳، ۱۴

در اینجا داریم:

$$a=10 \text{ و } b=20, \quad R=20-10= \boxed{10}$$

حال این دامنه تغییرات را چهار قسمت می‌کنیم. انتخاب این تعداد قانون و قاعده خاصی ندارد بلکه توصیه‌هایی درباره آن هست که به تدریج به آن‌ها اشاره می‌شود. با این تصمیم ما داده را به چهار دسته $\{17/5, 2\}$ و $[15, 17/5)$ و $[12/5, 15)$ و $[10, 12/5)$ تقسیم کرده‌ایم. در اینجا $(12/5, 15)$ یعنی اعدادی که از ۱۵ کوچکترند و بزرگتر یا مساوی ۱۲/۵ هستند. منظور ما از این تقسیم‌بندی آن است که مقادیر موجود در دسته اول یعنی $(10, 12/5)$ ، اگرچه متفاوت اند ولی از نظر ما یکی هستند و از تفاوت موجود بین آن‌ها می‌توان صرف‌نظر کرد. همین نکته یکی از توصیه‌های انتخاب تعداد دسته‌هاست. اگر از اختلاف بین ۱۰ و ۱۲/۵ بتوانیم چشم‌پوشی کنیم و این اختلاف به نظرمان بزرگ بیاید، باید فاصله‌ها را کوچک‌تر کنیم. پس باید تعداد آن‌ها را بیش‌تر کنیم. مثلاً داده‌ها را به جای چهار دسته به پنج دسته تقسیم کنیم.

مرکز دسته: داده‌ها را به نحوی دسته‌بندی کردیم که افرادی که در یک دسته قرار می‌گیرند از نظر اندازه متغیر مورد مطالعه تفاوت چندانی باهم نداشته باشند. در اصل ما می‌توانیم با کمی اغماض این افراد را دارای یک اندازه بدانیم. حال سؤال این است که این اندازه مشترک برای آن‌ها چه مقداری باشد، بهتر است. چون اختلاف‌ها را در هر دسته ناچیز فرض کرده‌ایم، هر یک از مقادیر حاصل از آن دسته را می‌توانیم برای هر یک از افراد آن دسته در نظر بگیریم ولی اگر بتوان کاری کرد که این یکسان‌سازی داده‌ها در دسته‌ها تأثیر کم‌تری در روند مطالعه بگذارند، مسلماً بهتر خواهد بود. از این‌رو مناسب‌ترین مقداری که می‌توانیم به هر یک از افراد یک دسته نسبت دهیم، مقدار میانی و یا وسط دسته است. این مقدار را مرکز دسته می‌گوییم. اگر دسته i ام به صورت $(a_i, b_i]$ باشد مرکز آن را که با نماد x_i نشان می‌دهیم به صورت زیر تعریف می‌شود.

$$x_i = \frac{b_i + a_i}{2}$$

تعریف: برای دسته $(a_i, b_i]$ را کران پایین، b_i را کران بالا و x_i را مرکز آن دسته

می‌نامیم.

در دسته دوم (۱۵، ۱۲/۵) مرکز دسته را محاسبه می‌کنیم:

$$x_i = \frac{۱۲/۵ + ۱۵}{۲} = \frac{۲۷/۵}{۲} = ۱۳/۷۵$$

ما اگر داده‌ها را در مثلاً ۴ دسته، دسته‌بندی کرده باشیم، به جای تعداد زیادی از داده‌ها ما فقط با ۴ عدد که همان مرکز دسته‌ها باشند سروکار داریم. البته چون قرار ما بر این است که داده‌های داخل یک دسته را برداریم و به جای آن‌ها مرکز دسته را قرار دهیم، این اعداد تکراری خواهند بود. داده‌های زیر نمرات آمار و مدلسازی یک نمونه ۸ نفری از دانش‌آموزان یک کلاس هستند.

$$۱۶, ۱۶/۵, ۱۸/۵, ۱۷, ۱۷/۵, ۱۸, ۱۹/۵, ۲۰$$

در این داده‌ها داریم:

$$a = ۱۶ \text{ و } b = ۲۰, \quad R = ۲۰ - ۱۶ = ۴$$

داده‌ها را در دو دسته، دسته‌بندی می‌کنیم. این دسته‌ها عبارتند از: (۱۶، ۱۸)، {۱۸، ۲}. در دسته اول، ۱۶ کران پایین دسته اول و ۱۸ کران بالای دسته اول می‌نامیم، به همین ترتیب در دسته دوم ۱۸ کران پایین دسته دوم و ۲۰ کران بالای دسته دوم می‌باشند. داده‌های موجود در هر دسته را در جدول زیر می‌نویسیم.

دسته‌ها	مرکز دسته‌ها	داده‌های هر دسته
۱۶ - ۱۸	۱۷	۱۶, ۱۶/۵, ۱۷, ۱۷/۵
۱۸ - ۲۰	۱۹	۱۸, ۱۸/۵, ۱۹/۵, ۲۰

حال از نقطه نظر ما داده‌هایی که در یک دسته قرار دارند، یکسان در نظر گرفته می‌شوند و مقدار مشترک آن‌ها همان مرکز دسته‌هاست. بنابراین، اطلاعات بالا به صورت زیر هموارتر می‌شود.

دسته‌ها	مرکز دسته‌ها	داده‌های هر دسته
۱۶ - ۱۸	۱۷	۱۷/۱۷, ۱۷/۱۷
۱۸ - ۲۰	۱۹	۱۹/۱۹, ۱۹/۱۹

پس در این مسئله ما با دو مقدار ۱۷ و ۱۹ سروکار داریم که هر یک از آن‌ها به تعدادی تکرار شده‌اند. در این مثال ملاحظه می‌کنید که مرکز دسته ممکن است یکی از داده‌ها نباشد.

تفاضل دو کران پایین متوالی یا دو کران بالای متوالی را طول آن دسته می‌نامیم.



تمرین

- ۱- در مثال نمرات آمار و مدلسازی طول دسته را حساب کنید.
 - همچنین تعداد تکرارهای مراکز دسته‌ها را حساب کنید.
 - مجموع این تکرارها چقدر است؟ این مجموع برابر چه چیزی است؟
- با توجه به تقسیمات آب و هوایی، ناحیه خشک به ناحیه‌ای گفته می‌شود که دارای بارش سالانه 30° تا 25° میلی متر است و ناحیه نیمه‌خشک ناحیه‌ای است که دارای بارش سالانه از 25° تا 600 میلی متر است و ناحیه مرطوب ناحیه‌ای است که بارش سالانه آن از 600 تا 2000 میلی متر است. به سؤالهای زیر پاسخ دهید :
- چند درصد مراکز استان‌های ایران در ناحیه خشک و نیمه مرطوب قرار دارند؟
 - حداکثر اختلاف میزان بارندگی در مراکز استان‌ها چقدر است؟

- جدول فراوانی را برای داده‌های زیر تشکیل دهید :
- اگر داده‌ها را در چهار دسته قرار دهیم، کران بالا و پایین دسته دوم را تعیین کنید.
- حداکثر اختلاف میزان بارندگی در هر دسته چقدر است؟ آیا این مقدار برای تمام دسته‌ها یکسان است؟
- در دسته سوم چند شهر قرار دارد؟ میزان بارندگی شهرهای این دسته حدوداً چقدر است؟ و کدام دسته بیشترین شهرها را دارد؟

مقدار بارندگی سالانه مراکز استان‌ها (میلی‌متر)

سال	مرکز استان
۳۲۸	تهران
۴۷۵	اراک
۱۷۳۴	رشت
–	ساری
۳۲۶	تبریز
۴۲۶	ارومیه
۴۰۴	کرمانشاه
۲۰۱	اهواز
۵۸۲	شیراز
۳۰۹	کرمان
۳۶۹	مشهد
۱۹۸	اصفهان
۹۹	زاهدان
۳۶۹	سنندج
۴۲۵	همدان
۴۲۶	شهرکرد
۷۶۲	خرم‌آباد
۶۳۵	ایلام
۱۲۶۵	یاسوج
۲۶۰	بوشهر
۳۱۲	زنجان
۱۶۰	سمنان
۱۰۵	یزد
۳۹۱	بندرعباس

سالنامه مرکز آمار سال ۱۳۷۱

انواع فراوانی

فراوانی مطلق: داده‌های زیر را که مربوط به ضریب هوشی ۷ دانش‌آموز پسر است در نظر

بگیرید :

۱۱۲, ۱۱۷, ۱۱۱, ۱۱۲, ۱۱۶, ۱۱۷, ۱۱۲

برخی از این داده‌ها تکرار شده‌اند. در جدول زیر این داده‌ها را با تعداد تکرارهای آن نوشته‌ایم :

داده‌ها	تعداد تکرارها
۱۱۱	۱
۱۱۲	۳
۱۱۶	۱
۱۱۷	۲

تکرار هر داده را فراوانی مطلق آن داده و جدول فوق را جدول فراوانی می‌گوییم. در مثال بالا فراوانی مطلق ۱۱۱ برابر ۱ و فراوانی مطلق ۱۱۲ برابر ۳ است. به همین ترتیب فراوانی‌های مطلق ۱۱۶ و ۱۱۷ به ترتیب برابر ۱ و ۲ می‌باشند.

در دسته‌بندی داده‌ها قرار گذاشتیم تمام داده‌های واقع در یک دسته را برابر مرکز آن دسته در نظر بگیریم. بنابراین مرکز دسته‌ها به تعداد اعضای که در آن دسته قرار دارند تکرار خواهند شد. یعنی فراوانی مرکز دسته‌ها برابر تعداد اعضای است که در آن دسته قرار گرفته‌اند. این فراوانی را فراوانی مطلق آن دسته می‌گوییم.



فراوانی مطلق دسته i ام را که با f_i نشان می‌دهیم برابر تعداد اعضای آن است که در دسته i ام قرار گرفته‌اند.

فعالیت

جدول فراوانی را برای وزن همکلاسی‌هایتان تشکیل دهید و مجموع فراوانی‌های مطلق را حساب کنید و تحقیق کنید این مجموع برابر چیست؟

فراوانی نسبی: دو جدول زیر را در نظر بگیرید:

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق
۱-۳	۲	۱۹
۳-۵	۴	۲۵
۵-۷	۶	۱۶
		۶۰

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق
۱-۳	۲	۶
۳-۵	۴	۱۰
۵-۷	۶	۴
		۲۰

اولین سؤالی که ممکن است به نظر خواننده بیاید این است که این دو جدول چگونه حاصل شده‌اند؟ معمولاً یکی از مطالعات آماری مقایسه دو نوع رفتار، بر روی یک مجموعه است. مثلاً فرض کنید دو روش تدریس داریم که روش اول را برای ۲۰ نفر و روش دوم را برای ۶۰ نفر دانش‌آموزان هم‌سطح اعمال کرده‌ایم، از این افراد امتحانی یکسان گرفته‌ایم که نمرات در این امتحان می‌تواند از ۱ تا ۷ تغییر کند. کسانی که نمره آن‌ها در (۱, ۳) است ضعیف و کسانی که نمره آن‌ها در (۳, ۵) است



متوسط و نمره کسانی که در [۵,۷] است قوی ارزیابی می‌شوند. حال می‌خواهیم عمل کرد این دو روش آموزش را با هم مقایسه کنیم. اگر تعداد دانش‌آموزان در دو کلاس برابر بودند، به‌سادگی می‌توانستیم این مقایسه را انجام دهیم و معلوم کنیم که با کدام روش آموزش موفق‌تر بوده است. ولی از آنجایی که تعداد دانش‌آموزان کلاس‌ها برابر نیستند، به‌سادگی نمی‌توانیم این مقایسه را به انجام برسانیم. برای حل این مشکل با فرض این که عمل کرد روش آموزش بستگی به تعداد دانش‌آموزان کلاس ندارد (چون اگر بدانیم که بستگی دارد این مطالعه نمی‌تواند مبنایی برای مقایسه موفقیت دو روش آموزش باشد. در این صورت توضیح دهید که تعداد افراد دو کلاس چگونه باید باشند؟) فرض می‌کنیم تعداد دانش‌آموزان هر کلاس برابر ۱۰۰ نفر باشد و با یک تناسب ساده تعداد دانش‌آموزان ضعیف، متوسط، قوی را در هر کلاس حساب می‌کنیم. مثلاً در دسته اول از کلاس اول محاسبات زیر را داریم:

$$\begin{array}{r} \text{تعداد دانش‌آموزان ضعیف} \\ ۶ \\ \hline ۱۰۰ \end{array} = \frac{x}{100} \quad x = 100 \times \frac{6}{100} = 30$$

همان‌طوری که قبلاً دیده‌اید، ۳۰ درصد دانش‌آموزان کلاس اول ضعیف ارزیابی شده‌اند. به همین ترتیب می‌توانیم برای کلاس دیگر و دسته‌های بعدی روش بالا را تکرار کنیم و در نتیجه جدول‌های زیر حاصل می‌شود.

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد
۱-۳	۲	۶	۳۰
۳-۵	۴	۱۰	۵۰
۵-۷	۶	۴	۲۰
جمع	-	۲۰	۱۰۰

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد
۱-۳	۲	۱۹	۳۱/۶
۳-۵	۴	۲۵	۴۱/۷
۵-۷	۶	۱۶	۲۶/۷
جمع	-	۶۰	۱۰۰

چون منظور، مقایسه دو آزمون بود جدول صفحه بعد را که خلاصه اطلاعات را دربر دارد در نظر می‌گیریم.

	درصد کلاس اول	درصد کلاس دوم
ضعیف	۳۰	۳۱/۶
متوسط	۵۰	۴۱/۷
قوی	۲۰	۲۶/۷

از مقایسه ستون‌های این دو جدول ملاحظه می‌شود که ترکیب دانش‌آموزان ضعیف در هر دو روش تقریباً ثابت بوده است (در حدود ۳۰، ۳۱ درصد) ولی روش دوم روی دانش‌آموزان متوسط موفق‌تر عمل کرده است. زیرا از درصد متوسط‌های کلاس اول کاسته شده ولی به درصد دانش‌آموزان قوی کلاس دوم اضافه شده است.

دیدیم که برای مقایسه لازم است تأثیر اختلاف نمونه را از بین ببریم برای این منظور تصمیم گرفتیم که اندازه نمونه را ۱۰۰ فرض کنیم که در این صورت درصد فراوانی‌ها را حساب کردیم و برای محاسبه این درصد دیدیم که باید عبارت $\frac{\text{فراوانی}}{\text{تعداد نمونه}}$ را در ۱۰۰ ضرب کنیم که آن را درصد فراوانی نسبی و نسبت $\frac{\text{فراوانی}}{\text{تعداد نمونه}}$ را فراوانی نسبی می‌گوییم.

اگر f_i فراوانی دسته i ام و تعداد داده‌ها n باشد کسر $\frac{f_i}{n}$ را فراوانی نسبی دسته i ام می‌گوییم.

جدول فراوانی موضوع مورد بحث به صورت زیر تکمیل می‌شود.

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد فراوانی نسبی	فراوانی نسبی
۱-۳	۲	۶	۳۰	۰/۳
۳-۵	۴	۱۰	۵۰	۰/۵
۵-۷	۶	۴	۲۰	۰/۲
جمع		۲۰	۱۰۰	۱

فعالیت

با همکاری دانش‌آموزان جدول فراوانی نمرات تمام دانش‌آموزان را برای هر درس تهیه کنید و در کلاس نصب نمایید.

دسته‌ها	فراوانی	فراوانی تجمعی: جدول فراوانی روبرو تعداد کالاهایی را که پس از فروش برای تعمیرات اولیه به کارخانه سازنده عودت داده شده‌اند، نشان می‌دهد. در این جدول ۳ کالا در فاصله ۱۰ تا ۱۳ ماه و ۶ کالا در فاصله ۱۳ تا ۱۶ ماه و ۷ کالا در فاصله ۱۶ تا ۱۹ ماه و ۴ کالا در فاصله ۱۹ تا ۲۲ ماه عودت داده شده‌اند.
۱۰-۱۳	۳	
۱۳-۱۶	۶	
۱۶-۱۹	۷	
۱۹-۲۲	۴	

چند کالا در فاصله ۱۰ تا ۱۹ ماه عودت داده شده است؟ فاصله ۱۰ تا ۱۹ به دسته‌های ۱۰ تا ۱۳، ۱۳ تا ۱۶ و ۱۶ تا ۱۹ تقسیم شده است. بنابراین تعداد کل کالاهایی که در فاصله ۱۰ تا ۱۹ عودت داده شده‌اند عبارت خواهد بود از:

$$۳+۶+۷=۱۶$$

این عدد را فراوانی تجمعی دسته ۱۹-۱۶ می‌گویند.

فراوانی تجمعی هر دسته برابر تعداد اشیایی است که مقدار آن‌ها از کران بالای آن دسته کمتراند.

مثلاً فراوانی تجمعی دسته اول عبارت است از تعدادی که از ۱۳ کم‌ترند و این برابر همان فراوانی مطلق دسته اول است. فراوانی تجمعی دسته دوم یعنی تعدادی که از ۱۶ کم‌ترند و این برابر ۳+۶ است. به این ترتیب می‌توانیم فراوانی تجمعی دسته‌ها را حساب کنیم. معمولاً ستونی به جدول فراوانی تحت عنوان فراوانی تجمعی اضافه می‌کنند. در مثال بالا جدول فراوانی را می‌توان به صورت روبرو تکمیل کرد.

دسته‌ها	فراوانی‌ها	فراوانی تجمعی
۱۰-۱۳	۳	۳
۱۳-۱۶	۶	۹
۱۶-۱۹	۷	۱۶
۱۹-۲۲	۴	۲۰
جمع	۲۰	

سؤال: در این جدول آیا همیشه دو عددی را که در نشان داده‌ایم باهم برابرند؟ چه دلیلی برای پاسخ خود دارید؟

محاسبه فراوانی تجمعی هر دسته، به صورت زیر است:

فراوانی تجمعی هر دسته برابر مجموع فراوانی‌های دسته‌های قبل و فراوانی آن دسته است.

چند تذکر درباره تشکیل جدول فراوانی

گفتیم که با بررسی اولیه در داده‌ها، با توجه به اندازه نمونه و بزرگی یا کوچکی دامنه تغییرات، تصمیم می‌گیریم که داده‌ها را در تعدادی دسته، دسته‌بندی کنیم، اگر این تعداد برابر k باشد با تقسیم R ، دامنه تغییرات بر k ، طول دسته‌ها به دست می‌آید. در بعضی مواقع ممکن است این دو عدد خارج قسمت خوبی به ما ندهند (اعشار زیاد داشته باشند و یا اصلاً قابل تقسیم نباشند). مثلاً فرض کنید $R = 29$ و $k = 7$ در این صورت طول دسته‌ها عبارت است از:

$$C = \frac{R}{k} = \frac{29}{7} = 4/1428$$

ابتدا توجه می‌کنیم که وقتی به اعشار متوسل می‌شویم، منظور ما این است که می‌خواهیم اختلاف را در حد یک‌دهم، یک‌صدم و یا حتی یک‌هزارم کنترل کنیم. اگر C را به صورت $4/14$ در نظر بگیریم حداکثر اختلاف $4/14$ با $\frac{29}{7}$ برابر یک‌صدم است و اگر به صورت $4/142$ در نظر بگیریم حداکثر اختلاف برابر یک‌هزارم است، آیا واقعاً ما به این همه دقت نیاز داریم؟ توجه کنید که ایده اصلی در تشکیل جدول فراوانی از بین بردن اختلافات جزئی و یک کاسه کردن داده‌های نزدیک به هم است. حال اگر بخواهیم در تعیین طول دسته‌ها به اعشار زیاد از حد متوسل شویم این ایده اصلی تشکیل جدول فراوانی را نادیده گرفته‌ایم، از این رو با در نظر گرفتن یک مقدار تقریبی مناسب برای C مسئله اعشار زیاد را حل می‌کنیم. ممکن است پیشنهاد کنید که در مثال بالا بهترین مقداری که برای C در نظر می‌گیریم عدد 4 باشد، توجه کنید 7 دسته به طول 4 ، فاصله‌ای به طول 28 را خواهند پوشاند که این از دامنه تغییرات کم‌تر است. دسته‌ها باید به نوعی باشند که مجموع طول آن‌ها از دامنه تغییرات کم‌تر نباشد.

در این مثال شما می‌توانید طول دسته‌ها را هر یک از مقادیر $4/2$ یا $4/5$ در نظر بگیرید، در این صورت فاصله‌ای که با $C = 4/2$ پوشانده می‌شود برابر $7 \times 4/2 = 29/4$ است که به اندازه $4/4$ از R بیش‌تر است. حال اگر دسته اول را شما از کوچک‌ترین داده شروع کنید دسته آخر به عددی ختم خواهد شد که $4/4$ از بزرگ‌ترین داده بیش‌تر است که البته اشکال چندانی ندارد ولی در ریاضیات توجه به زیبایی مهم است.

از زیباترین مفاهیم ریاضی موضوع تقارن است. از این رو برای زیبایی بیشتر بهتر است این مقدار را به دو قسمت مساوی تقسیم کنیم و اولین دسته را از $2/2$ کم‌تر شروع کرده تا آخرین دسته به $2/2$ بیش‌تر ختم شود.

در یک امتحان تستی که شامل 30 سؤال بوده است یک نمونه 20 نفری از دانش‌آموزان نمرات زیر را کسب کرده‌اند:

۱، ۸، ۹، ۱۶، ۷، ۲۴، ۱۳، ۲۵، ۲۹، ۲۸، ۲۹، ۳۰، ۱۲، ۵، ۲۵، ۱۷، ۲۷، ۱۴، ۱۹، ۱۲

این داده‌ها را در هفت دسته، دسته‌بندی کنید.

در این داده‌ها داریم، $a = 1$ ، $b = 3^\circ$ ، $R = 29$ پس $C = \frac{29}{V} \approx 4/2$ پس در این صورت دسته‌ها عبارتند از:

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق
$(0/8, 5)$	$2/9$	۱
$(5, 9/2)$	$7/1$	۴
$(9/2, 13/4)$	$11/3$	۳
$(13/4, 17/6)$	$15/5$	۳
$(17/6, 21/8)$	$19/7$	۱
$(21/8, 26)$	$23/9$	۳
$(26, 30/2]$	$28/1$	۵
	جمع	2°

اگر طول دسته‌ها را $4/5$ بگیریم خواهیم داشت:

$$4/5 \times 7 = 31/5$$

پس به اندازه $31/5 - 29 = 2/5$ واحد از دامنه تغییرات زیادت‌تر خواهیم داشت پس دسته اول

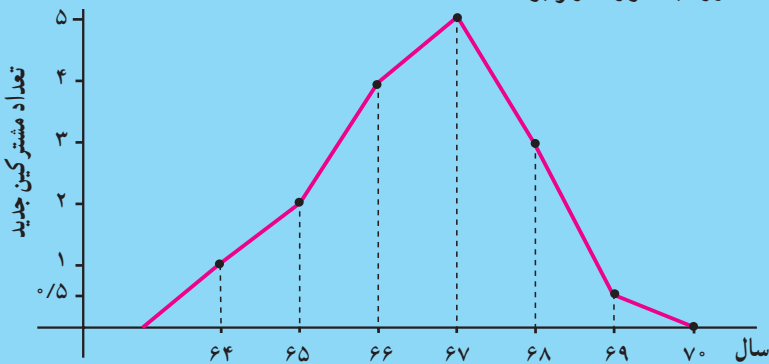
$$\text{را به جای } 1 \text{ از } 1 - \frac{2/5}{4} = -0/25 \text{ شروع می‌کنیم.}$$

ملاحظه می‌کنید که این دو جدول تفاوت‌هایی در ستون فراوانی مطلق و مرکز دسته‌ها دارند ولی این اختلاف‌ها در تفسیر داده‌ها و نتیجه‌گیری‌های آماری تأثیری نخواهد داشت.

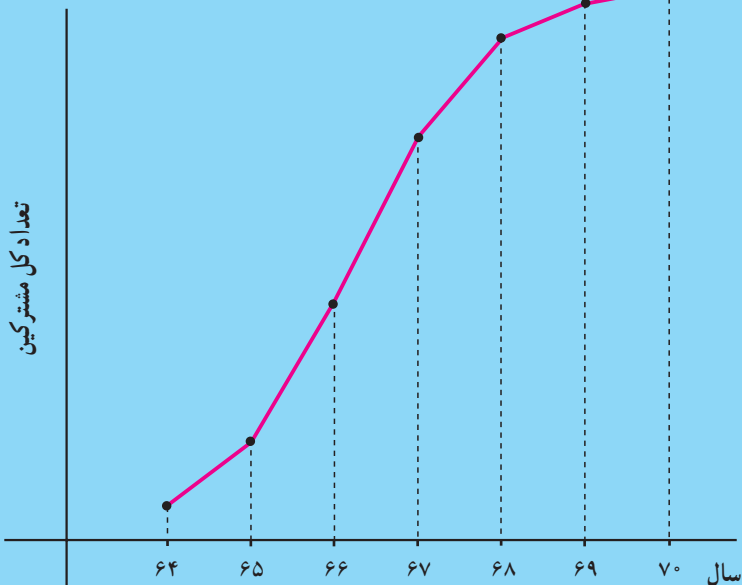
دسته‌ها	مرکز دسته‌ها	فراوانی
$-0/25, 4/25$	۲	۱
$4/25, 8/25$	$6/5$	۳
$8/25, 12/25$	۱۱	۴
$12/25, 16/25$	$15/5$	۲
$16/25, 20/25$	2°	۲
$20/25, 24/25$	$24/5$	۳
$24/25, 28/25$	۲۹	۵
	جمع	2°

بحث کلاسی

مدیر بانکی از هیأت مدیره برای بازدید شعبه دعوت کرد. این مدیر، تصمیم گرفت نموداری از وضعیت بانک به نمایش بگذارد. نموداری که او دربارهٔ تعداد مشترکین به دست آورد به صورت زیر بود.



این مدیر متوجه شد که از سال ۶۷ تعداد مشترکین جدید روبه کاهش گذاشته است و اتفاقاً این سال مصادف با سالی است که او مسئولیت این شعبه را پذیرفته بود. این وضعیت، وضعیت مناسبی نبود. او به فکر چاره افتاد و سرانجام نمودار زیر را رسم کرد و مورد تشویق قرار گرفت.



- می‌توانید بگویید برای رسم این نمودار از چه فراوانی استفاده شده است؟
- می‌توانید بگویید از کجا فراوانی‌های مطلق روبه کاهش گذاشته است؟

تمرین

۷-	۱۱	
۱۱		
		۲۵

۱- در دسته‌های جدول فراوانی و مرکز دسته‌ها داریم:

فاصله بین دو مرکز دسته متوالی = طول دسته

برابری بالا را با یک مثال نشان دهید. سپس با استفاده از این

برابری جدول روبرو را تکمیل کنید.

۲- می‌دانیم شکستن رمزها و کدها اهمیت زیادی در امنیت ملی

دارد.

حروف	درصد تکرار
E	۱۳
T	۹
A,O	۸
N	۷
I,R	۶/۵
S,H	۶
D	۴
L	۳/۵
C,U,M	۳
F,P,Y	۲
W,G,B	۱/۵
V	۱
K,X,J	۰/۵
Q,Z	۰/۲

ساده‌ترین نوع رمزی کردن یک پیغام این است که

هرحرف را با یک علامت جایگزین کنیم. معمولاً اولین

قدم در کشف این نوع رمز بررسی فراوانی هر علامت در

متن رمزنگاری شده است. زیرا معمولاً درصد حروفی که

در متون یک زبان خاص به کار رفته‌اند مشخصه‌ای از آن

زبان هستند. در جدول روبرو درصد فراوانی حروف در

متون زبان انگلیسی نمایش داده شده‌اند.

جدول درصد فراوانی حروف در متون زبان فارسی را به‌دست بیاورید.

۳- می‌خواهید چاپگر کامپیوتر خریداری کنید. چاپگرهای سوزنی هنوز هم طرفداران بسیاری

دارد. سرعت چاپ همیشه یک فاکتور مهم برای انتخاب بوده است. در مجله PC در مقاله ششم

بررسی سالانه چاپگر (جلد هشتم شماره ۱۹) سرعت ۳۱۲ نوع مختلف چاپگر آمده است. یک نمونه

تصادفی ۴۸ تایی از چاپگرهای سوزنی انتخاب شد. جدول صفحه بعد سرعت این چاپگرها را

برحسب کاراکتر در ثانیه نشان می‌دهد:

۲۴۷، ۱۹۵، ۱۴۵، ۹۱، ۱۰۰، ۱۱۱، ۲۵۸، ۹۹، ۵۳، ۶۰، ۱۲۴، ۱۳۳، ۱۷۳، ۴۶، ۱۱۵، ۹۳،
 ۱۱۱، ۱۳۸، ۱۶۳، ۱۰۵، ۱۴۱، ۱۳۸، ۱۶۳، ۱۰۵، ۱۴۱، ۱۳۸، ۱۵۴، ۱۲۸، ۱۰۲، ۳۷، ۸۶، ۷۲،
 ۱۷۳، ۱۵۷، ۱۵۰، ۱۳۳، ۸۹، ۸۹، ۱۴۷، ۱۱۱، ۱۵۷، ۱۴، ۱۵۰، ۱۰۰، ۱۳۱، ۱۳۶، ۱۵۴

الف - طول دسته‌ها را با در نظر گرفتن ۵ دسته به دست آورید.

ب - جدول فراوانی را با داشتن ۵ دسته تشکیل دهید. مرکز دسته‌ها را مشخص کنید.

۴- پرستاران یکی از بیمارستان‌های تهران اعتقاد دارند که نیاز به همکاران بیش‌تری در نوبت شب دارند. برای آن که حجم کارشان تخمین زده شود، نمونه‌ای از ۳۵ شب به‌طور تصادفی انتخاب شد و تعداد درخواست‌های بیماران در طول شب به‌قرار زیر ثبت شد:

۶۲، ۵۸، ۴۶، ۸۷، ۱۰۲، ۳۷، ۷۱، ۸۲، ۷۳، ۷۰، ۸۰، ۹۵، ۱۰۱، ۱۰۰، ۹۰، ۱۸، ۸۶، ۷۰، ۶۸،
 ۷۱، ۷۵، ۸۶، ۸۳، ۷۴، ۶۰، ۷۷، ۹۰، ۸۶، ۸۴، ۸۵، ۶۹، ۷۰، ۹۲، ۶۳

جدول فراوانی داده‌های فوق را برای پنج دسته تشکیل دهید. اگر ۷۵ درخواست در شب کار سنگین در نظر گرفته شود، آیا فکر می‌کنید، درخواست این پرستاران به حق است؟

پروژه

در این پروژه می‌خواهیم بدانیم نوجوانان روزانه با چه گروه‌های سنی بیش‌تر سر و کار دارند.

موضوع: بررسی سن جامعه خارج از مدرسه که شما در طی روز با آن‌ها سر و کار دارید (اعضای خانواده، فامیل، همسایگان، مغازه‌داران، راننده سرویس و ...). برای این بررسی حدود ۲۰ نفر از جامعه بالا را در نظر بگیرید. جدول فراوانی سن آن‌ها را تشکیل دهید. به سؤالات زیر پاسخ دهید:

- داده‌های خود را چگونه جمع‌آوری نموده‌اید؟
- بیش‌ترین تعداد افرادی که با آن‌ها سر و کار داشته‌اید، در چه گروه سنی قرار دارند؟
- این مجموعه چه درصدی از افرادی را تشکیل می‌دهند که با آن‌ها سر و کار داشته‌اید؟
- توزیع سن را از طریق تعیین درصد فراوانی هر یک از گروه‌های سنی مشخص کنید.

معرفی و آشنایی با نرم افزار آماری MINITAB



در سال‌های اخیر استفاده از کامپیوتر تقریباً در تمام جنبه‌های زندگی افزایش یافته است. استفاده از کامپیوتر زمان لازم برای تجزیه و تحلیل‌های کمی را کاهش داده است.

در سال‌های اخیر نرم‌افزارهای آماری بسیاری به بازار وارد شده است. چهار نرم‌افزار اصلی آماری عبارتند از: SPSS، SAS، MINITAB، BMDP. در این بخش خلاصه‌ای از چگونگی کار کردن با MINITAB برای حل مسائل آماری آورده شده است. این نرم‌افزار تحت ویندوز ۹۵ عمل می‌کند.

MINITAB برای ویندوز

ابتدا وارد محیط ویندوز شوید: روی نمایه‌آبی‌رنگ MINITAB دوبار کلیک کنید، تا وارد محیط MINITAB شوید. همین‌که وارد محیط MINITAB شدید، صفحه‌کار MINITAB_ UNTITLED را مشاهده می‌کنید، زیر آن لیست انتخاب PULL _ DOWN MENUS را مشاهده می‌کنید:

File, Edit, Manip, Calc, Stat, Graph, Editor, Window, Help

با کلیک کردن روی هریک از این انتخابات، لیست انتخابات در هریک نمایانگر می شود. به طور مثال اگر روی File کلیک کنید؛ لیست زیر نمایان شده و شما می توانید هر کدام را انتخاب کنید:

File		
<u>N</u> EW	Worksheet ...	Ctrl + N
<u>O</u> pen	Worksheet ...	Ctrl + O
<u>M</u> erg	Worksheet ...	
<u>Q</u> uery Data base [ODBC]	...	
<u>S</u> ave Worksheet		Ctrl + S
Save Worksheet <u>A</u> s		
Worksheet Description		
Open Graph		
Save Window As		
Other Files		→
<u>P</u> rint Window		Ctrl + P
Print Setup		
Restart Minitab		
Exit		

ورود داده ها

صفحه نمایش زیر لیست انتخابات به دو قسمت تقسیم شده است: Session و پنجره داده ها (Data Windows). از پنجره Session برای تایپ کردن فرمان های Minitab استفاده می شود. همچنین خروجی Minitab Output در این قسمت ظاهر می شود.

معمولاً داده ها را در قسمت پنجره داده ها (Data Windows) وارد می کنید. این قسمت شامل ستون هایی است که با C1 و C2 و ... نامگذاری شده اند. در Minitab داده ها همواره به شکل ستونی وارد می شوند. به طور مثال اگر داده های شما راجع به قد افراد است، این داده ها به صورت ستونی در یکی از ستون های C1 یا C2 یا ... وارد می شوند. داده های کیفی را نیز می توان با وارد کردن یک نام یا مقدار در هر خانه ستون ثبت نمود.

می توانید نام متغیر چون وزن، قد و ... را نیز وارد کنید. توجه کنید که هر خانه ستون شماره گذاری شده است. هر دفعه که داده ای را وارد می کنید، کلید ENTER یا RETURN را فشار دهید یا کلید فلش به پایین \downarrow را فشار دهید تا به خانه بعدی وارد شوید. (می توانید از کلیدهای جهت \leftarrow \rightarrow \uparrow \downarrow استفاده نمایید) همچنین می توانید با استفاده از ماوس و حرکت نشانگر خانه مورد نظر را انتخاب کنید. زمانی که ورود تمام داده ها تمام شد می توانید از لیست انتخاب، تحلیل

نمره امتحان	نام	مورد نیاز خود را انتخاب کنید.
۱۲	علی	تذکر: زمانی که داده‌ها را وارد می‌کنید از «،» استفاده
۱۷	حسن	نکنید برای مثال برای وارد کردن ۴۵،۷۶۳ باید ۴۵۷۶۳ تایپ
۱۴	بهزاد	شود.
۱۹	امیر	مثال ۱: فرض کنید می‌خواهید داده‌های روبرو را در
۲۰	جلال	Minitab وارد کنید :
۲۰	حمید	– ابتدا روی خانه C۱ کلیک کرده تایپ کنید NAME
۱۹	سعید	«نام» سپس به خانه زیر حرکت کنید (با استفاده از ماوس،
۱۸	احمد	کلیدهای جهتی، ENTER) و شروع به تایپ اسامی کنید. سپس
		روی C۲ کلیک کرده تایپ کنید SCORE «نمره» و ادامه دهید.
		پنجره داده‌ها در پایان به شکل زیر درخواهد آمد :

Data

	C1-T	C2	C3	C4	C5	C6	C7
↓	Name	Score					
1	ALI	12					
2	HASSAN	17					
3	BEHZAD	14					
4	AMIR	19					
5	JALAL	20					
6	HAMEED	20					
7	SAEED	19					
8	AHMAD	18					
9							

اگر در وارد کردن داده‌ها اشتباهی صورت گرفت، نشانگر را روی خانه موردنظر کلیک کرده کلید ENTER را فشار داده و سپس داده را وارد نمایید. اطلاعات جدید جایگزین اطلاعات قبلی می‌شود.

ذخیره کردن یک فایل Minitab

پس از ورود تمام داده‌ها می‌توانید فایل داده‌ها را برای استفاده در آینده ذخیره نمایید. به‌طور مثال می‌خواهید داده‌های مثال ۱ را ذخیره کنید :

قدم اول: روی لیست انتخاب، روی File کلیک کنید.

قدم دوم: روی Save Worksheet As کلیک کنید.

قدم سوم: نام فایل خود را در کادر File Name تایپ کنید. این نام معمولاً پسوند «.MTW» دارد به طور مثال فایل مثال را می توان SCORE.MTW نامید. سپس درایو مناسب را در قسمت Drives انتخاب کنید. برای مثال می توانید این فایل را در درایو A یا C و ... ذخیره نمایید، اگر بیش از یک دایرکتوری دارید، دایرکتوری موردنظر خود را انتخاب کرده سپس روی OK کلیک کنید. مطمئن شوید در کادر Save File As Type، لغت Minitab نشان داده شده باشد.

قدم چهارم: پس از انجام اعمال فوق روی OK کلیک کنید. در مثال بالا فایل فوق تحت نام SCORE.MTW ذخیره شد.

باز آوردن فایلی که قبلاً ذخیره شده است:

قدم اول: روی لیست انتخاب روی File کلیک کنید.

قدم دوم: روی Open Worksheet کلیک کنید.

قدم سوم: کادر Open Worksheet ظاهر می شود. درایوی را که فایل موردنظرتان در آن ذخیره شده است انتخاب کنید به طور مثال اگر درایو A انتخاب شود شما این عبارت را می بینید a:\ و نام تمام فایل های Minitab در درایو A زیر کادر File Name نمایان می شوند. فایل موردنظر را انتخاب کرده و روی OK کلیک کنید.

قدم چهارم: پس از کلیک کردن روی OK، فایل موردنظر بازمی گردد.

ذخیره سازی فایل خروجی Minitab

فایل خروجی Minitab شامل پاسخ ها و نتایج تجزیه و تحلیل های انجام شده است. شما می توانید فایل خروجی را برای استفاده در آینده به عنوان فایل Text ذخیره کنید. فرض کنید در مثال ۱، یک سری محاسبات آماری انجام شده است و نتایج را می خواهید در فایلی به نام Score.Txt ذخیره کنید:

قدم اول: روی File کلیک کنید.

قدم دوم: روی Save Window As کلیک کنید.

قدم سوم: در کادر File Name، نام فایل مورد نظر را تایپ کنید.

نام فایل خروجی معمولاً پسوند «.TXT» دارد که نشان دهنده فایل خروجی در Minitab است. درایو مورد نظر خود را در قسمت Drive انتخاب کرده و سپس در قسمت Directories، دایرکتوری موردنظرتان را انتخاب کنید.

اگر در فایل خروجی نمودار رسم شده است، مراحل فوق را برای ذخیره‌سازی انجام داده فقط به جای **TEXT** از «**MGF**» استفاده کنید.

چاپ یک فایل خروجی Minitab

برای چاپ یک فایل خروجی شامل نمودار، قدم‌های زیر را انجام دهید:
قدم اول: روی **File** کلیک کنید.

قدم دوم: روی **Print Window** کلیک کنید.

قدم سوم: اگر تمام فایل خروجی را می‌خواهید چاپ کنید روی **All** کلیک کنید، در غیر این صورت شماره صفحات موردنظر برای چاپ را در قسمت **FROM** (از) و **TO** (تا) تایپ کنید.
قدم چهارم: روی **OK** کلیک کنید.

انتخاب نمونه با استفاده از Minitab

از این نرم‌افزار می‌توان در انتخاب نمونه‌ای از یک جامعه استفاده نمود. فرض کنید در مثال ۱، از بین ۸ نفر می‌خواهید یک نمونه ۳ نفری انتخاب کنید و این نمونه را در ستون‌های **C۳** و **C۴** ثبت کنید. توجه کنید که داده‌های جامعه در ستون‌های **C۱** و **C۲** ثبت شده‌اند:

قدم اول: روی **Cal** در لیست انتخاب کلیک کنید.

قدم دوم: روی **Random Data** کلیک کنید.

قدم سوم: روی **Sample From Column** کلیک کنید.

قدم چهارم: در کادر کنار **Sample**، تایپ کنید: **۳** و در کادر **Samples** تایپ کنید: **C۱ - C۲**. این دستور به این معنی است که سه ردیف از داده‌های وارد شده در ستون‌های **C۱** و **C۲** را انتخاب کنید. سپس در کادر **Store Sample in** تایپ کنید **C۳ - C۴**. اگر تکرار در نمونه‌تان مجاز است کنار کادر **Sample With Replacement** علامت بزنید.

قدم پنجم: روی **OK** کلیک کنید.

خروج از Minitab

قبل از آن که از **Minitab** خارج شوید، مطمئن شوید تا تمام فایل‌های داده‌ها و خروجی ذخیره شده‌اند و سپس چنین عمل کنید:

قدم اول: روی **File** کلیک کنید.

قدم دوم: روی **Exit** کلیک کنید.

قدم سوم: به سؤالات **Minitab** پاسخ دهید.

فرمان‌های زبانی Minitab

به جای استفاده از ویندوز می‌توان از فرمان‌های زبانی Minitab استفاده نمود. پس از ورود به محیط Minitab نشان‌گر روی Data Window خواهد بود که نشان‌دهنده این امر است که شما می‌توانید از Windows استفاده کنید. برای تغییر این وضعیت چنین عمل کنید :

قدم اول: در کادر Session Window کلیک کنید. نشان‌گر از Data Window به Session Window انتقال پیدا می‌کند.

قدم دوم: روی Editor در لیست انتخاب کلیک کنید.

قدم سوم: روی Enable Command Language کلیک کنید.

پس از انجام این عملیات پیغام : >MTB نمایان می‌شود. که Minitab Prompt خوانده می‌شود. در این حالت کامپیوتر آماده دریافت فرمان‌های Minitab می‌باشد. پس از تایپ فرمان‌های Minitab پس از Prompt کلید Enter یا Return را فشار دهید. برای بازگشتن به وضعیت Data Window ، روی لیست انتخاب Editor را انتخاب کرده روی Disable Command Language کلیک کنید.

ورود داده‌ها

مثال ۲: فرض کنید می‌خواهید داده‌های مثال ۱ را با استفاده از فرمان‌های Minitab وارد کنید. می‌توانید از فرمان‌های SET یا READ استفاده کنید. از SET برای ورود داده‌ها یکی یکی استفاده می‌شود. از READ برای ورود داده‌های یک متغیر یا متغیرهای بیش‌تری استفاده می‌شود. فرمان Minitab که با Note شروع می‌شود فقط برای اطلاع رساندن به کاربر است و روی آن عمل نمی‌شود.

```
MTB > NOTE : DATA ON NAMES AND
```

```
SCORES OF 8 STUDENTS
```

ورود داده‌ها در Minitab با استفاده از فرمان‌های زبانی Minitab :

این فرمان ستون C۱ را ایجاد می‌کند وجود؛ در انتهای فرمان اشاره بر این دارد که فرمان ادامه دارد و اطلاعات بیش‌تری وارد خواهد شد. به همین جهت PROMPT ملاحظه شده >SUBC خواهد بود و یعنی زیر فرمان. (A6) FORMAT نشان‌دهنده این است که یک متغیر اسمی حداکثر با ۶ کاراکتر داریم. توجه داشته باشید طولانی‌ترین نام ۶ کاراکتر دارد.

```
MTB > SET C1 ;
```

```
SUBC > FORMAT (A6).
```

نقطه در پایان فرمان نشان‌دهنده این است که اطلاعات پایان یافته.

DATA > ALI ↵

DATA > HASSAN ↵

DATA > BEHZAD ↵

DATA > AMIR ↵

DATA > JALAL ↵

پایان ستون C۱

DATA > HAMEED ↵

ستون C۲ را آماده می‌کند. در این جا نیازی

DATA > SAEED ↵

به «؛» نمی‌باشد چون داده‌ها عدد هستند

DATA > AHMAD ↵

اعداد را می‌توان در یک ردیف نوشت ولی

DATA > END

نام‌ها باید یکی یکی وارد شوند.

MTB > SET C2

DATA > 12 17 14 19 20 20 19 18

DATA > END ↵

پایان ستون C۲

شما می‌توانید ستون‌های فوق را با استفاده از فرمان NAME، نام‌گذاری کنید :

MTB > NAME C1 'NAME'

MTB > NAME C2 'SCORE'

فرمان خروجی به قرار زیر است :

MTB > PRINT C1-C2

MTB > PRINT 'NAME' 'SCORE'

ذخیره‌سازی فایل ورودی

اگر تصمیم دارید از داده‌های وارد شده در Minitab بعدها استفاده کنید، باید آن‌ها را قبل از پایان دادن کار با Minitab، ذخیره کنید. برای ذخیره کردن فایل ورودی ابتدا باید فایل را نام‌گذاری کنید. فرض کنید می‌خواهید نام «SCORE» را به فایل موجود در مثال ۲ نسبت دهید. از هر یک از دو فرمان زیر می‌توان برای انجام این کار استفاده نمود. از این فرامین بسته به این که می‌خواهید روی درایو C یا A ذخیره کنید، استفاده کنید. توجه داشته باشید که نام فایل باید در بین ' ' قرار گیرد.

این فرمان فایل

MTB > SAVE 'SCORE'

SCORE. MTW را روی دیسک سخت ذخیره می کند.

این فرمان فایل

```
MTB > SAVE 'A: SCORE'
```

SCORE. MTW را روی دیسک Floppy در درایو A ذخیره می کند.

اگر خودتان پسوندی برای SCORE انتخاب نکرده باشید، Minitab پسوند MTW. را به نام فایل اختصاص می دهد.

مثال ۳: فایل SCORE را در دایرکتوری به نام STAT در درایو C ذخیره کنید :

```
MTB > SAVE 'C:\STAT\SCOR'
```

بازآوری فایل های ورودی ذخیره شده در Minitab

با استفاده از فرمان RETRIEVE فایل ورودی را که قبلاً ذخیره شده است می توانید فعال

کنید.

از این فرمان زمانی استفاده می شود که فایل روی درایو C در دیسک سخت ذخیره شده باشد.

```
MTB > RETRIEVE 'SCORE'
```

از این فرمان زمانی استفاده می شود که فایل روی درایو A روی Floppy ذخیره شده باشد.

```
MTB > RETRIEVE 'A: SCORE'
```

اگر فایل روی دایرکتوری خاصی در درایو C ذخیره شده باشد، باید دایرکتوری را مشخص

نموده به طور مثال برای بازآوری فایل ورودی مثال ۳ چنین عمل کنید :

```
MTB > RETRIEVE 'C:\STAT\SCORE'
```

در MINITAB ، فرمان با استفاده از چهار حرف اول شناخته می شود پس به جای استفاده از

RETRIEVE می توان از RETR استفاده نمود. فراموش نکنید نام فایل باید همواره بین ' ' قرار گیرد.

فرمان INFORMATION یا INFO به شما کمک می کند تا از آنچه در یک فایل باز یافته قرار

دارد اطلاع پیدا کنید :

```
MTB > INFO
```

ذخیره سازی یک فایل خروجی Minitab

از فرمان OUTFILE می توان در ذخیره سازی فایل خروجی که شامل پاسخها و تجزیه و

تحلیل داده هاست استفاده نمود. به طور مثال شما یک سری عملیات آماری روی داده های مثال ۲

انجام داده اید و می خواهید فایل خروجی با نام SCORE را ذخیره کنید. به یاد داشته باشید نام فایل

بین ' ' قرار می‌گیرد و این کار باید قبل از آن که هرگونه عملیات آماری انجام شود، پس از آمدن اولین PROMPT در MINITAB انجام شود :

این فرمان فایل خروجی را روی دیسک سخت ذخیره می‌کند.

```
MTB > OUTFILE 'SCORE'
```

این فرمان فایل خروجی را روی Floppy در درایو A ذخیره می‌سازد.

```
MTB > OUTFILE 'A: SCORE'
```

اگر می‌خواهید فایل خروجی در دایرکتوری خاصی چون STAT ذخیره شود چنین عمل کنید :

```
MTB > OUTFILE 'C:\STAT\SCORE'
```

پس از آن که کارتان با MINITAB تمام شد، از محیط MINITAB خارج شده و فایل SCORE.LIS را چاپ کنید یا در جای دیگر که لازم دارید کپی کنید. با اجرای دستور OUTFILE، پسوند LIS. بطور اتوماتیک توسط MINITAB به نام فایل اتصال پیدا می‌کند. اگر پسوند خاصی را در نظر دارید می‌توانید از آن نام ببرید :

```
OUTFILE 'SCORE.TXT'
```

یا

```
OUTFILE 'A: SCORE.TXT'
```

به این ترتیب فایل خروجی با نام SCORE.TXT ذخیره می‌شود.

چاپ فایل خروجی MINITAB

شما مستقیماً در محیط MINITAB نمی‌توانید یک فایل خروجی چون SCORE.LIS یا SCORE.TXT را چاپ کنید برای این کار یا باید ابتدا به محیط MINITAB FOR WINDOW رفته و همان‌طور که قبلاً گفته شد از FILE در لیست انتخاب استفاده کنید و فایل خروجی را همان‌طور که قبلاً توضیح داده شده است چاپ کنید. روش دیگر این است که ابتدا از محیط MINITAB خارج شوید و سپس فایل را چاپ کنید.

ورود داده‌ها با استفاده از فرمان READ

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید و دو یا بیش از دو متغیر کمی با تعداد مساوی داده برای هریک دارید، می‌توانید از READ استفاده کنید مثال ۴ نشان‌دهنده چگونگی این کار است.

وزن (کیلو)	قد (سانتی متر)
۸۱	۱۷۲
۶۱	۱۶۷
۵۵	۱۶۳
۹۵	۱۷۸
۶۸	۱۷۰
۶۵	۱۶۵

مثال ۴: فرض کنید می خواهید داده های روبرو را برای

قد و وزن شش نفر وارد کنید :

برای ورود این داده ها با استفاده از فرمان READ قد

و وزن هر فرد را در یک ردیف وارد کنید.

MTB > READ C1 C2 *

DATA > 172 81

DATA > 167 61

DATA > 163 55

DATA > 178 95

DATA > 170 68

DATA > 165 65

DATA > END

* این فرمان به MINITAB اعلام

می دارد که شما می خواهید مقادیر دو متغیر

را در دو ستون C1 و C2 وارد کنید.

انتخاب نمونه با استفاده از MINITAB

فرض کنید داده های ورودی شما در مثال ۲ متعلق به یک جامعه ۸ نفری باشد و شما می خواهید

مشاهده ای روی یک نمونه ۳ نفری داشته باشید و داده های این نمونه را در ستون های C3 و C4

ذخیره کنید. دستورهای زیر چگونگی انجام این کار را نشان می دهند.

MTB > NOTE : SELECTING A SAMPLE OF OBSERVATIONS

MTB > SAMPLE 3 FROM C1 - C2 PUT IN C3 - C4

MTB > PRINT C3 - C4

این فرمان داده های نمونه را در ستون های C3 و C4 به نمایش می گذارد.

خروج از MINITAB

برای خروج از MINITAB زمانی که از فرمان های زبانی MINITAB استفاده می کنید، باید

از دستور STOP پس از PROMPT در MINITAB استفاده نمود :

MTB > STOP

MINITAB از شما سؤالاتی می پرسد که باید به آن ها پاسخ بلی یا خیر دهید. پس از پاسخ دادن به تمام سؤالات از MINITAB خارج می شوید.

فرمان های دیگر در MINITAB

MTB > HELP HELP

از این فرمان می توان برای کمک گرفتن در MINITAB استفاده نمود.

MTB > HELP COMMANDS

این فرمان کمک در رابطه با فرمان های MINITAB ارائه می کند.

MTB > HELP OVERVIEW

از این فرمان می توان برای مرور MINITAB استفاده نمود.

MTB > COPY C1 TO C2

این فرمان مقادیر ستون C1 را در ستون C2 کپی می کند.

MTB > ERASE C2

این فرمان مقادیر موجود در ستون C2 را پاک می کند.

MTB > DELETE ROW 2 C1

این فرمان ردیف دوم (مقدار دوم) از ستون C1 را حذف می کند.

MTB > DELETE ROW 2 C1 - C2

این فرمان ردیف دوم در ستون های C1 و C2 را حذف می کند.

MTB > INSERT BETWEEN 2 and 3 C1 - C2

این فرمان ردیف جدید بین ردیف های دوم و سوم برای ستون های اول و دوم ایجاد می کند.

MTB > LET C1 (4) =10

این فرمان عدد 10 را در چهارمین ورودی ستون اول وارد می کند.

MTB > SORT C1 PUT IN C3

این فرمان داده های ستون اول را به ترتیب صعودی مرتب کرده و در ستون سوم می نویسد.

MTB > ADD C1 C2 PUT IN C4

این فرمان مقادیر متناظر را در ستون های اول و دوم باهم جمع می کند و در ستون چهارم قرار

می دهد.

MTB > SUBTRACT C2 FROM C1 PUT IN C5

این فرمان مقدار C۲ را از مقدار C۱ متناظر کم کرده و در ستون پنجم قرار می دهد.
MTB > MULTIPLY C1 BY C2 PUT IN C6

این فرمان مقادیر متناظر C۱ و C۲ را در هم ضرب کرده و در C۶ قرار می دهد.
MTB > DIVIDE C1 BY C2 PUT IN C7

این فرمان هر مقدار موجود در C۱ را بر داده متناظر در C۲ تقسیم می کند، نتیجه را در C۷ (ستون هفتم) قرار می دهد.

MTB > LET C8 = C1 * C2

این فرمان مقادیر C۱ و C۲ را در هم ضرب می کند و در C۸ (ستون هشتم) قرار می دهد.

MTB > LET C9 = C1 ** 2

این فرمان مقادیر C۱ را به توان ۲ می رساند و در C۹ (ستون نهم) قرار می دهد.

MTB > ADD 5 TO C1 PUT IN C10

این فرمان به هریک از مقادیر C۱ عدد پنج را اضافه کرده و در C۱۰ (ستون دهم) قرار می دهد.

MTB > SUBTRACT 8 FROM C1 PUT IN C11

این فرمان از هریک از مقادیر C۱ مقدار ۸ را کم کرده و در C۱۱ (ستون یازدهم) قرار می دهد.

MTB > MULTIPLY C1 BY 2 PUT IN C12

این فرمان مقادیر C۱ را در عدد ۲ ضرب کرده و در C۱۲ (ستون دوازدهم) قرار می دهد.

MTB > DIVIDE C1 BY 3 PUT IN C13

این فرمان مقادیر C۱ را بر ۳ تقسیم کرده و در C۱۳ (ستون سیزدهم) قرار می دهد.

فرمان HELP COMMAND توضیحات راجع به فرامین را نیز به همراه دارد.
به طور مثال HELP SET، اطلاعاتی راجع به فرمان SET و موارد استفاده آن به شما می دهد.

تمرین

۱- نام و قد و وزن همکلاسی های خود را روی یک دیسک Floppy ثبت کنید. سپس یک نمونه ۵ نفری با استفاده از فرمان های زبانی MINITAB و نیز MINITAB WINDOWS انتخاب کرده و داده های مربوط به این نمونه را چاپ کنید.

۲- جدول زیر نام ساعات کار و مزد پنج کارمند را نشان می دهد.

نام	ساعات کار	مزد (تومان)
علی	۴۲	۱۷۲۵۰
شیوا	۳۳	۱۴۸۳۰
زهرا	۲۸	۱۳۵۵۰
بهمن	۴۷	۱۷۹۰۰
اکبر	۴۰	۱۸۲۰۰

الف - با استفاده از MINITAB WINDOW داده‌های فوق را ثبت کنید. فایل فوق را با نام WORKER.MTW ذخیره کنید. از MINITAB خارج شوید. سپس دوباره وارد MINITAB شوید و فایل WORKER.MTW را دوباره باز کنید.

ب - داده‌های فوق را در ستون‌های C۱، C۲ و C۳ با استفاده از فرمان‌های زبانی MINITAB وارد کنید. فایل فوق را با نام WORKER.MTW ذخیره کنید. فرمان‌های زیر را امتحان کنید و نتیجه را بررسی کنید:

```

MTW > PRINT C1 - C3
MTB > NAME C1 'NAME' C2 'HOURS' C3 'SALARY'
MTB > PRINT. 'NAME' 'HOURS' 'SALARY'
MTB > LET K2 = SUM (C2)
MTB > PRINT K2
MTB > LET K3 = SUM (C3)
MTB > PRINT K3
MTB > PRINT C4
MTB > LET K4 = SUM (C4)
MTB > PRINT K4
MTB > LET C5 = C3 * C3
MTB > LET K5 = SUM (C5)
MTB > PRINT C2 - C5
MTB > PRINT K2 - K5
MTB > SAMPLE 2 FROM C1 - C3 PUT IN C6-C8
MTB > PRINT C6 - C8
MTB > SAVE 'ASSIGN1'
MTB > DIR
MTB > RETRIEVE 'ASSIGN1'
MTB > INFO
MTB > STOP

```

هیچ کس چیزی را در دل نمان
نکرد، جز که در سخنان
بی‌اندیشه‌اش آشکار گشت و
در صفحه‌رُخسارش پدیدار.

نمودارها و تحلیل داده‌ها

فکر می‌کنید کسب اطلاعات با استفاده از جدول آسان‌تر است یا نمودار و یا یک متن؟ شنیده‌ایم که می‌گویند: «یک تصویر خوب ارزش هزار کلمه را دارد.» با دیدن یک نمودار در یک نگاه می‌توان به بسیاری از ویژگی‌های مجموعه‌ای از اطلاعات پی‌برد. امروزه ارائه نموداری داده‌های آماری بسیار رایج است.

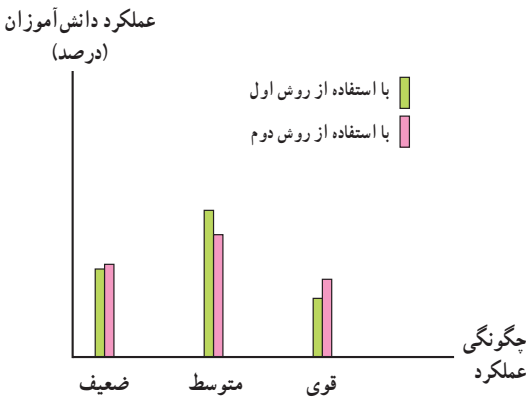
پوسترها، روزنامه‌ها، مجلات، تلویزیون از نمودار برای به تصویر درآوردن مسائل و ارائه آن‌ها به طریقی ملموس‌تر استفاده می‌کنند.

اگر به خاطر داشته باشید در فصل اول از شما پرسیدیم که روزانه با چه کسانی سروکار دارید؟ سن و سال آن‌ها چه قدر است؟ تا چه سطحی تحصیل کرده‌اند؟ یک جواب ساده به این سؤال می‌توانست به یکی از صورت‌های زیر باشد:

– با تعداد کمی از افراد که اکثر آن‌ها جوان و تحصیل کرده هستند.

– با تعداد زیادی از افراد که اکثر آن‌ها مسن و با تحصیلات پایین هستند.

– با تعداد زیادی از افراد که بیش‌تر آن‌ها تا حد دیپلم درس خوانده‌اند و میان‌سال هستند.



این سه پاسخ در یک جمله کوتاه توانسته‌اند دورنمایی از جمعیتی را که روزانه شما با آن‌ها سروکار دارید، ترسیم کنند. آیا سریع‌تر از بیان این جمله می‌توانستید همان جمعیت را معرفی کنید؟ یا در توفیق یکی از روش‌های آموزشی (مثالی که در فصل قبل بررسی شد) مطالعه جدول و توضیحات اگرچه ما را به اطلاعات عمیق‌تر می‌رساند ولی فعلاً لازم داریم در یک نگاه حقایق زیادی را دریا بایم. یکی از این روش‌ها رسم شکل روبرو است.

شکل ۱ – مقایسه عملکرد دانش‌آموزان دو کلاس در امتحان

نمودارها یا شاخص‌های هندسی وسیله‌ای سودمند برای به تصویر درآوردن و تجسم جامعه می‌باشند. در این فصل به معرفی برخی از این شاخص‌های هندسی می‌پردازیم.

جدول ۱- وضعیت بیمه دانش‌آموزان کلاس

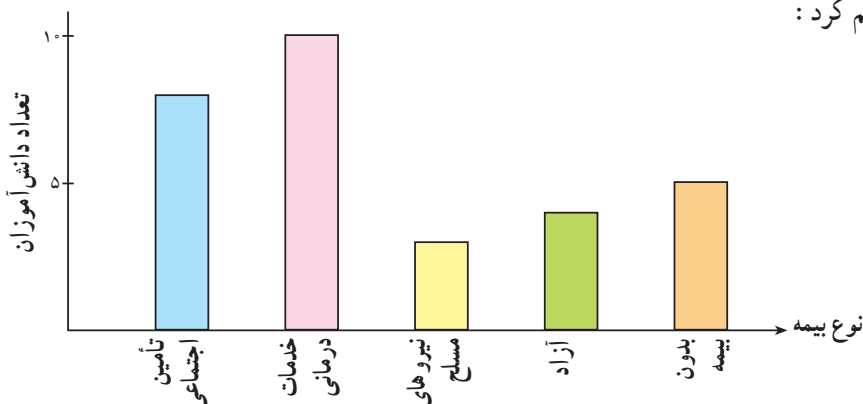
۸	بیمه تأمین اجتماعی
۱۰	بیمه خدمات درمانی
۳	بیمه نیروهای مسلح
۴	شرکت‌های بیمه آزاد
۵	دانش‌آموزانی که بیمه ندارند

نمودار میله‌ای

جدول ۱ نشان‌دهنده پاسخ‌های دانش‌آموزان کلاس درباره نوع بیمه درمانی مورد استفاده آن‌هاست.

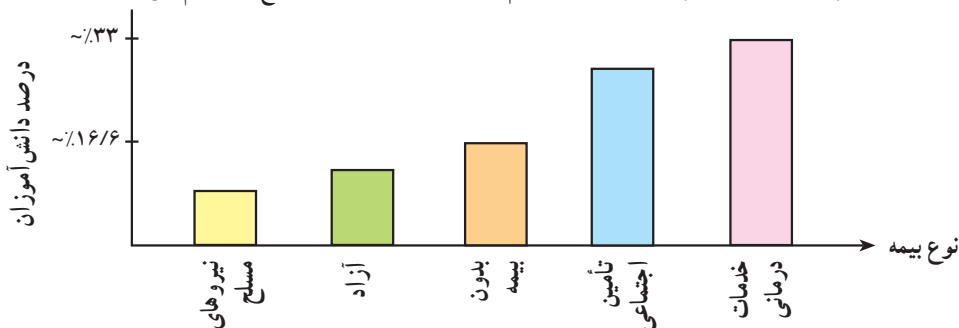
این اطلاعات را می‌توان در شکل زیر

مجسم کرد:



شکل ۲- نمودار توزیع دارندگان انواع بیمه در یک کلاس

این نمودار بیش‌تر برای متغیرهای گسسته و کیفی مناسب است. ترتیب قرارگرفتن میله‌ها اهمیت ندارد. آنچه که در این نمودار مهم است مقایسه فراوانی انواع بیمه است. از این رو اگر این نمودار را با درصد رسم کنیم، بهتر خواهد بود و هم‌چنین اگر به گونه‌ای رسم کنیم که فراوانی‌ها از بزرگ به کوچک و یا از کوچک به بزرگ رسم شوند، عمل مقایسه سریع‌تر انجام می‌شود.



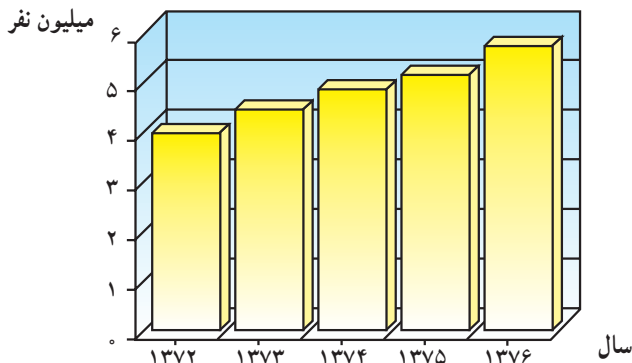
شکل ۳- نمودار توزیع دارندگان انواع بیمه در یک کلاس

کم بودن درصد استفاده کنندگان از بیمه نیروهای مسلح به این علت نیست که خدمات قابل قبولی ارائه نمی کند، بلکه به این علت است که افراد خاصی (نظامی) می توانند از این بیمه بهره بگیرند. هم چنین زیاد بودن درصد استفاده کنندگان از خدمات درمانی هم نباید به این نتیجه منتهی شود که خدمات آن مورد قبول افراد بیشتری است، این زیاد بودن به دلیل جمعیت کثیری است که از این بیمه استفاده می کنند. نمودار میله‌ای شامل قسمت‌های زیر است:

عنوان: زیر هر نمودار باید موضوع مورد مطالعه به طور خلاصه نوشته شود. در نمودار صفحه قبل عنوان نمودار عبارت است از: توزیع دارندگان انواع بیمه در کلاس. برچسب محورها: هر یک از محورها برچسبی دارند که مشخص کننده متغیری است که آن محور نشان می دهد. در شکل ۲، محور افقی مشخص کننده انواع بیمه‌ها و محور عمودی مشخص کننده تعداد استفاده کنندگان از این بیمه‌هاست.

مقیاس: مقیاس هر محور باید مشخص باشد. مثلاً در مثال بالا اگر بخواهیم اطلاعات مربوط به استفاده کنندگان از بیمه در سطح کشور را مطالعه کنیم در این صورت تعداد استفاده کننده برای برخی از بیمه‌ها از میلیون هم تجاوز می کند. پس مشخص کردن یک واحد برای هر فرد منجر به رسم میله‌های بسیار طولانی می گردد که در صفحه نخواهد گنجید. از این رو برای هر ۱۰۰ نفر یا هر ۱۰۰۰ نفر یک واحد انتخاب می شود که در این صورت باید در حاشیه نمودار این نکته تذکر داده شود. در نمودار زیر تعداد بیمه‌شدگان تحت پوشش سازمان تأمین اجتماعی را در سال‌های مختلف ملاحظه می کنید.





شکل ۴ - بیمه شدگان اصلی تحت پوشش سازمان تأمین اجتماعی

با توجه به نمودار بالا به سؤالات زیر پاسخ دهید :

- ۱- هر واحد روی نمودار نشان دهنده چند نفر می باشد؟
- ۲- در کدام سال نفرت بیش تری و در کدام سال نفرت کم تری تحت پوشش سازمان تأمین اجتماعی قرار گرفتند؟
- ۳- در کدام سال (ها) تعداد بیمه شدگان بین ۴ میلیون و ۵ میلیون است؟
- ۴- تقریباً چند نفر در سال ۱۳۷۵ بیش تر از سال ۱۳۷۳ زیر پوشش بیمه قرار گرفته اند؟
- ۵- آیا می توانید در مورد تعداد دارندگان بیمه در سال ۱۳۸۰ پیش بینی کنید؟

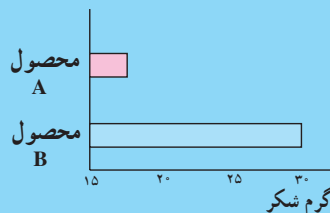
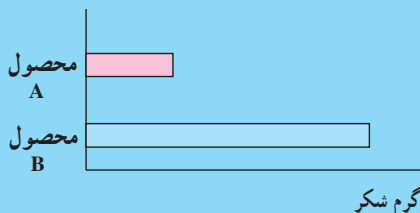
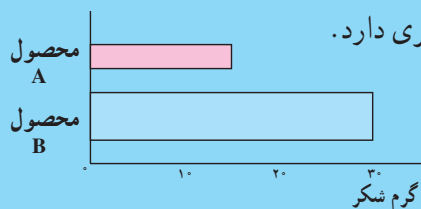
همان طوری که در نمودارها دیده اید برای رسم نمودار میله ای کافی است، متغیر تصادفی مورد مطالعه را روی محور xها مشخص کنیم و روی هر یک از آن ها میله ای رسم کنیم که طول میله متناسب با فراوانی آن مقدار باشد. طول میله ها بیانگر فراوانی و یا درصد فراوانی نسبی نظیر آن مقدار خواهد بود.

بحث کلاسی ۱

دو شرکت A و B نوعی شیرینی تولید می کنند. در نمودارهای زیر شرکت B

سعی دارد اعلام کند شیرینی های او شکر بیشتری دارد.

چرا این نمودارها گمراه کننده هستند؟



بحث کلاسی ۲

اطلاعات زیر را در نظر بگیرید :

	پیش‌دانشگاهی سوم دبیرستان دوم دبیرستان اول دبیرستان سوم راهنمایی دوم راهنمایی اول راهنمایی						
تعداد دختران بر حسب ۱۰۰۰ نفر	۸۴۹	۸۰۶	۷۴۷	۶۰۳	۵۰۳	۵۱۵	۲۹۴
	(برگرفته از آمار آموزش و پرورش، معاونت برنامه‌ریزی و نیروی انسانی)						

جدول ۲

- برای رسم نمودار میله‌ای به چه اطلاعاتی نیاز دارید؟
- نمودار میله‌ای اطلاعات به دست آمده را با استفاده از فراوانی مطلق و فراوانی نسبی رسم کنید.
- این دو نمودار چه تفاوت‌هایی دارند؟
- آیا این تفاوت‌ها تأثیری در تعبیر داده‌ها دارند یا خیر؟ توضیح دهید.



جدول ۳

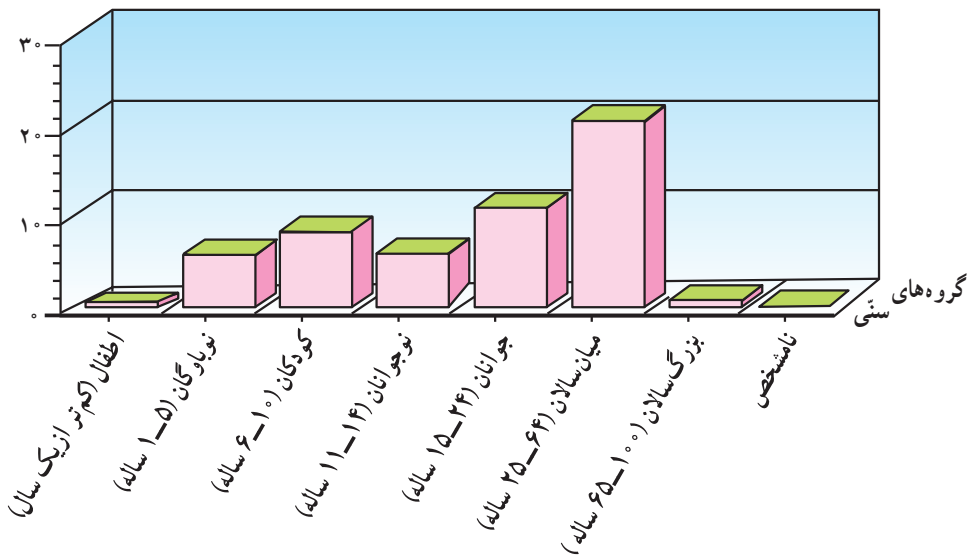
مرد و زن	گروه عمده سنی
۶۰۰۵۵۴۸۸	جمع
۱۰۲۰۹۳۶	اطفال (کمتر از یک سال)
۶۷۳۲۷۸۷	نوباوگان (۱-۵ ساله)
۸۷۶۹۷۳۷	کودکان (۱۰-۶ ساله)
۷۲۰۲۰۸۵	نوجوانان (۱۱-۱۴ ساله)
۱۲۳۳۷۵۲۹	جوانان (۱۵-۲۴ ساله)
۲۱۳۶۴۸۷۷	میان سالان (۲۵-۶۴ ساله)
۲۵۹۵۱۸۱	بزرگ سالان (۱۰۰-۶۵ ساله)
۳۲۳۵۶	نامشخص

نمودار مستطیلی

این نمودار برای متغیرهای کمی پیوسته مناسب است. در جدول زیر، اطلاعات مربوط به متغیر سن جمعیت ایران آمده است. می دانیم سن یک متغیر پیوسته است (در صورت لزوم می توانیم از واحدهای کوچکتر از قبیل روز و ساعت استفاده کنیم ولی از آن‌ها صرف نظر می شود) سن از نوع زمان است و لذا پیوسته است. ولی گاهی برحسب فاصله‌های سنی جمعیت را به دسته‌هایی که در جدول روپرو ملاحظه می کنید تقسیم می کنند. این دسته‌ها را گروه‌های سنی می نامند.

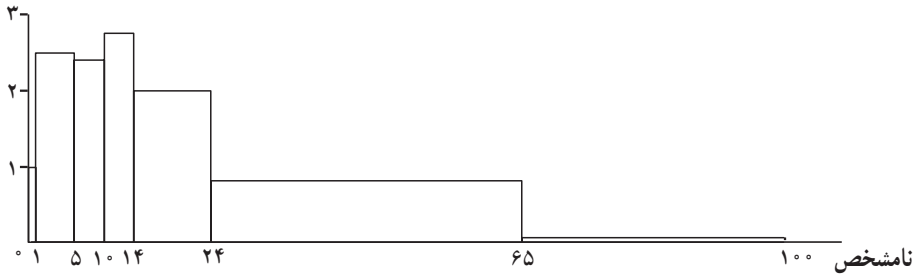
مأخذ: مرکز آمار ایران.

نمودار مربوط به این اطلاعات را ممکن است به صورت نمودار میله‌ای به شکل زیر رسم کنیم.



شکل ۵- نمودار توزیع جمعیت در گروه‌های مختلف سنی

از آنجایی که سن، یک متغیر پیوسته است، انتظار داریم این پیوستگی به طریقی در نمودار دیده شود از این رو از نمودار مستطیلی استفاده می کنیم. در این نمودار مستطیل‌هایی رسم می کنیم که قاعده آن‌ها روی محور x ها و برابر طول هر یک از دسته‌هاست و ارتفاع آن‌ها به موازات محور y ها و متناسب با فراوانی دسته‌هاست. مثلاً نمودار مستطیلی مربوط به داده‌های جدول بالا به صورت زیر است: (۵ - ۱ یعنی تا ۶ سال، پنج سال تمام)



شکل ۶ - نمودار توزیع جمعیت

تذکر: در نمودار مستطیلی مساحت مستطیل‌ها با یک‌دیگر مقایسه می‌شوند هرچقدر مساحت مستطیل بیشتر باشد تعداد کسانی که در آن دسته قرار می‌گیرند بیشتر است و در صورتی که طول دسته‌ها (قاعده مستطیل‌ها) با یک‌دیگر برابر باشند، ارتفاع مستطیل‌ها (فراوانی دسته‌ها) با یک‌دیگر مقایسه خواهند شد.

در این کتاب برای سهولت مقایسه‌ها، طول دسته‌ها را مساوی در نظر خواهیم گرفت.

نمودار مستطیلی نمایشی از داده‌های دسته‌بندی شده است که در آن سطح مستطیل‌ها متناسب با فراوانی دسته‌ها است.

بحث کلاسی

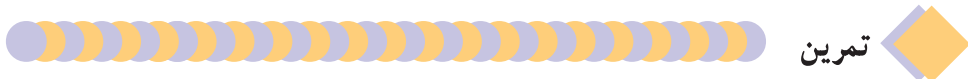
با توجه به نمودار جمعیت (شکل ۶) به سؤالات زیر پاسخ دهید.

– در یک نمودار مستطیلی، طول قاعده ستون‌ها نشان‌گر چیست؟ ارتفاع ستون‌ها نشان‌گر چیست؟

– اگر به‌طور متوسط دانش‌آموزان در سن ۱۵ سالگی در کلاس اول دبیرستان مشغول به تحصیل باشند، آیا می‌توانید با استفاده از نمودار فوق بگویید در این سال چند دانش‌آموز در کلاس اول دبیرستان مشغول به تحصیل هستند؟

– با استفاده از نمودار، جمعیت جوان و نوجوان کشور را تخمین بزنید.

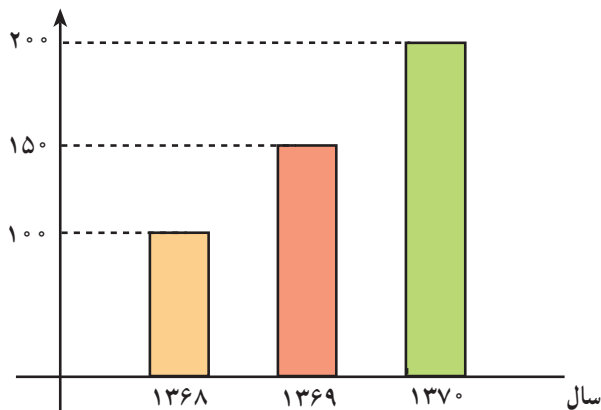
– همان‌طور که مشاهده می‌شود با استفاده از نموداری چون نمودار ستونی می‌توان به سؤال‌های مربوط به مجموعه‌ای از داده‌ها سریع‌تر پاسخ داد. اگر شما دست‌اندرکار طراحی و برنامه‌ریزی در امر آموزش باشید، با توجه به نمودار فوق، (شکل ۶) سؤال‌هایی طرح کنید و به آن‌ها در کلاس پاسخ دهید.



تمرین

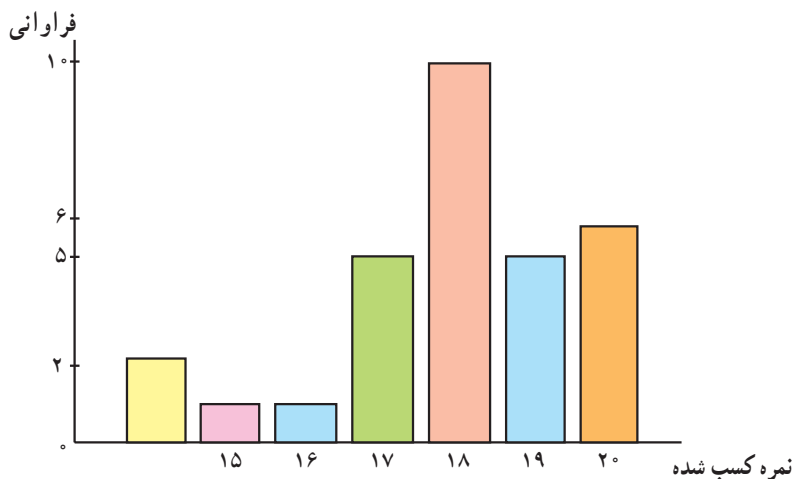
۱- نمودار زیر قسمتی از نمودار تولید یک کارخانه را نشان می‌دهد. آیا این نمودار نشان‌دهنده افزایش تولید در سال‌های ۱۳۶۸ تا ۱۳۷۰ است؟ این نمودار به همین شکل که هست، ما را دچار اشتباه می‌کند. چرا؟

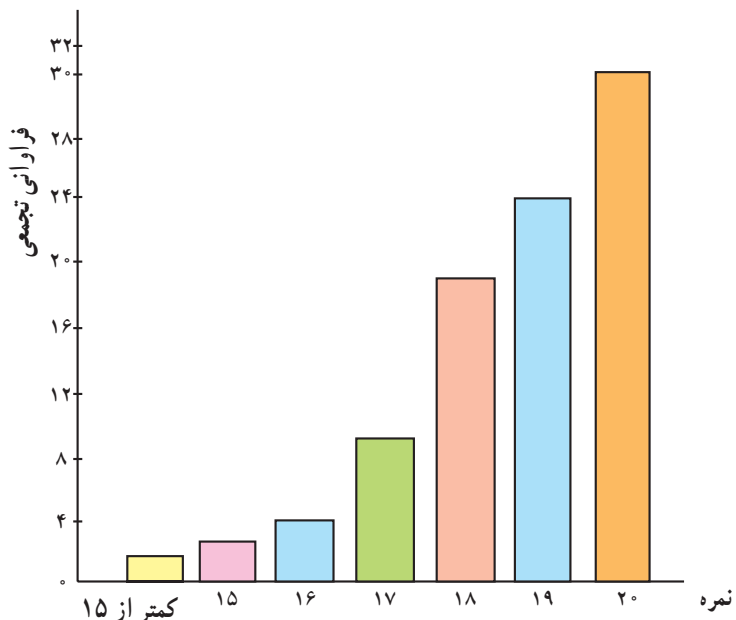
تعداد (برحسب ۱۰۰۰)



۲- در یک امتحان کتبی از درس کمک‌های اولیه، افرادی که نمره بیش‌تر از ۱۵ (از ۲۰) کسب کرده باشند، می‌توانند دوره عملی را شروع کنند. نمودارهای زیر نتیجه آزمون را نشان می‌دهد. (نمرات به شکل عدد صحیح ثبت شده‌اند.)

- با توجه به نمودارها تعداد افرادی را که می‌توانند در دوره شرکت نمایند مشخص کنید.
- کدام نمودار شما را سریع‌تر به پاسخ هدایت می‌کند.

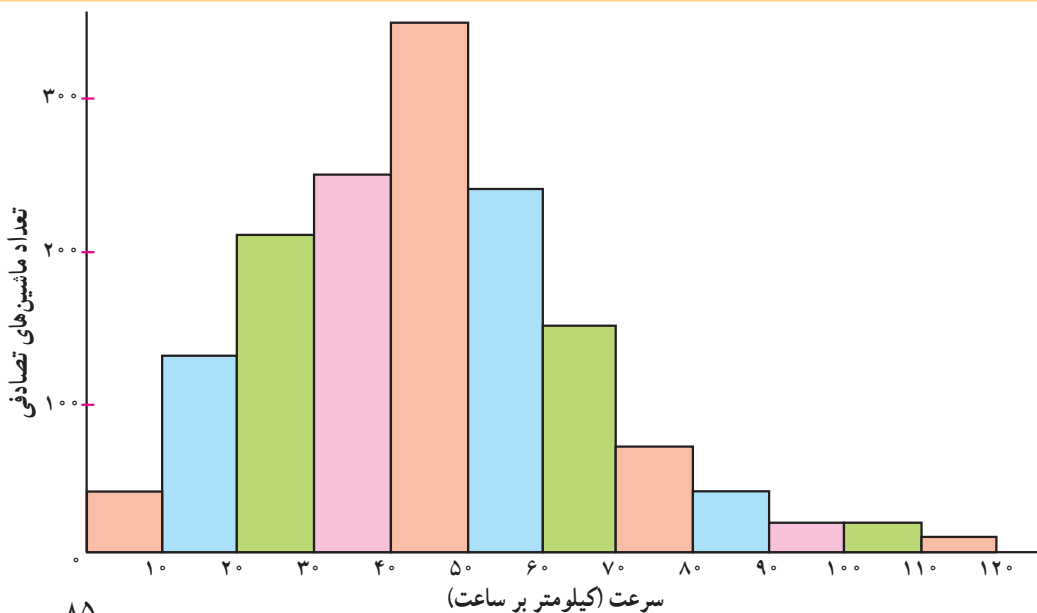




۳- جدول زیر نشان دهنده تعداد تصادف‌های اتومبیل‌ها و سرعت حرکت آن‌ها در زمان تصادف

است.

سرعت	کیلومتر	۱۰	۲۰	۳۰	۴۰	۵۰	۶۰	۷۰	۸۰	۹۰	۱۰۰	۱۱۰	۱۲۰
	ساعت	۱	۱	۲	۳	۶	۹	۷	۸	۹	۹	۱	۱
تعداد ماشین‌های تصادف کرده		۴۰	۱۳۰	۲۱۰	۲۵۰	۳۵۰	۲۴۰	۱۵۰	۷۰	۴۰	۲۰	۲۰	۱۰



همان طور که دیده می شود تعداد تصادف های اتومبیل هایی که با سرعت بالاتر از 90° کیلومتر در ساعت رفته اند کم تر است. پس :

«هرچه سریع تر بروید، مطمئن تر و امن تر است»

نتیجه گیری فوق چه اشکالی دارد؟ چرا این تصور ایجاد شده است؟ برای تحلیل صحیح، چه نموداری باید رسم کرد؟

۴-۵۰ دانش آموز یک کلاس در آزمون سنجش IQ (بهره هوشی) شرکت نمودند. نتایج به دست آمده به شرح زیر است :

۱۲۲ ، ۱۱۷ ، ۱۱۴ ، ۱۱۲ ، ۱۳۰ ، ۹۲ ، ۱۱۸ ، ۱۰۹ ، ۱۱۳ ، ۱۱۵ ،
۱۱۸ ، ۱۱۶ ، ۱۰۶ ، ۱۱۱ ، ۱۱۲ ، ۱۲۴ ، ۱۱۳ ، ۱۰۸ ، ۱۰۷ ، ۱۲۷ ،
۱۱۴ ، ۱۰۰ ، ۱۰۳ ، ۱۱۵ ، ۱۲۴ ، ۱۱۰ ، ۱۱۸ ، ۱۱۸ ، ۱۰۷ ، ۱۲۱ ،
۱۱۰ ، ۱۱۶ ، ۱۲۶ ، ۱۲۱ ، ۱۳۵ ، ۱۰۴ ، ۱۲۳ ، ۱۱۶ ، ۱۲۴ ، ۱۰۴ ،
۱۱۲ ، ۱۱۷ ، ۱۱۳ ، ۱۰۷ ، ۱۰۳ ، ۱۰۲ ، ۱۲۹ ، ۹۸ ، ۱۳۴ ، ۹۶

– جدول فراوانی را برای این داده ها تشکیل دهید. برای این جدول ۹ گروه با طول دسته مساوی در نظر بگیرید.

– نمودار مستطیلی این داده ها را رسم کنید.

– چه نتیجه ای می گیرید.

۵- با توجه به اطلاعات داده شده در جدول صفحه بعد :

– برای رسم نمودار میله ای برای اطلاعات داده شده چه مشکلاتی خواهید داشت؟

– برای رسم نمودار مستطیلی به چه اطلاعاتی نیاز دارید؟

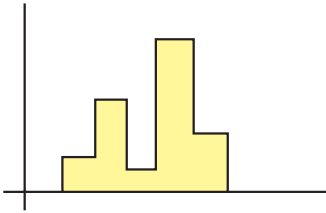
– نمودار مستطیلی مربوط به فضاهای سبز مراکز استان های ایران را با در نظر گرفتن ۴ دسته رسم کنید.

– نمودار مستطیلی رسم شده چه اطلاعاتی در اختیار شما قرار می دهد؟

مساحت بوستان‌های طبیعی، فضاهای سبز و ذخیره‌گاه‌های جنگلی تا پایان سال ۱۳۷۱ (هکتار)

ذخیره‌گاه‌های جنگلی	فضاهای سبز	بوستان‌های طبیعی	جمع کل عرصه‌های جنگلی	استان
۲۸۴۸۳	۲۳۳۹۰۴	۳۳۸۴۱	۲۹۶۲۲۸	جمع
۰	۱۲۵۳۰	۰	۱۲۵۳۰	تهران
۰	۸۵۰	۰	۸۵۰	مرکزی
۱۰۰۰	۳۳۳۵۶	۰	۳۴۳۵۶	گیلان
۸۷۴۵	۱۱۲۴۲۷	۸۶۳۱	۱۲۹۸۰۳	مازندران
۱۰۰	۸۶۰	۱۰۰	۱۰۶۰	آذربایجان شرقی
۰	۶۰۹	۰	۶۰۹	آذربایجان غربی
۳۰	۱۲۳۳	۰	۱۲۶۳	کرمانشاه
۴۳	۳۸۶۷	۱۲۰۰۰	۱۵۹۱۰	خوزستان
۱۵۰	۱۹۰۸	۹۱۶۵	۱۱۲۲۳	فارس
۳۰۰	۱۳۸۹	۰	۱۶۸۹	کرمان
۱۰۰۰۰	۳۱۳	۰	۱۰۳۱۳	خراسان
۲۰	۳۳۲۰	۰	۳۳۴۰	اصفهان
۵۰	۲۷۵۸۷	۰	۲۷۶۳۷	سیستان و بلوچستان
۳۰۰	۱۵۳۴	۱۰۰۰	۲۸۳۴	کردستان
۲۰۰	۱۱۷۴	۰	۱۳۷۴	همدان
۶۸۵	۱۳۴۷	۶۵۰	۲۶۸۲	چهارمحال و بختیاری
۷۵۰	۱۵۸۸	۸۵۰	۳۱۸۸	لرستان
۰	۲۱۰۱	۱۴۰۰	۳۵۰۱	ایلام
۱۰۰۰	۱۰۶۳	۴۵	۲۱۰۸	کهگیلویه و بویراحمد
۱۵۱۰	۱۸۹۱	۰	۳۴۰۱	بوشهر
۰	۳۹۶	۰	۳۹۶	زنجان
۶۰۰	۱۰۶۲	۰	۱۶۶۲	سمنان
۰	۵۳۲	۰	۵۳۲	یزد
۳۰۰۰	۲۰۹۶۷	۰	۲۳۹۶۷	هرمزگان

نمودار چندبر فراوانی



شکل ۸

جدول ۴

دسته‌ها	مرکز دسته‌ها	فراوانی
۱۰-۱۳	۱۱/۵	۳
۱۳-۱۶	۱۴/۵	۶
۱۶-۱۹	۱۷/۵	۷
۱۹-۲۲	۲۰/۵	۴
	جمع	۲۰

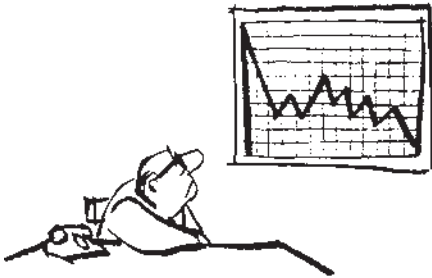
مجدداً به داده‌های مربوط به گروه سنی باز می‌گردیم. در رسم نمودار مستطیلی گفتیم که چون سن یک متغیر پیوسته است باید این پیوستگی را در نمودار آن نیز ببینیم. اما در نمودار مستطیلی فراوانی‌ها روی دسته‌ها تغییر نمی‌کنند و نمودار مستطیلی بیش‌تر شبیه به پلکانی به صورت روبرو است.

اگر بخواهیم تغییرات متغیر را بهتر نشان دهیم از نمودار چندبر فراوانی استفاده می‌کنیم.

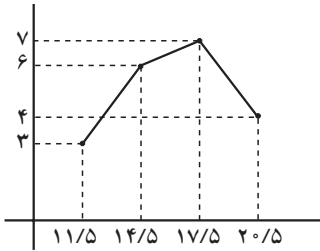
داده‌هایی که در فصل سوم مربوط به تعداد کالاهایی که پس از فروش برای تعمیرات اولیه به کارخانه سازنده عودت داده شده‌اند، در نظر بگیرید.

برای مثال فرض کنید بخواهیم یک نمودار چندبر

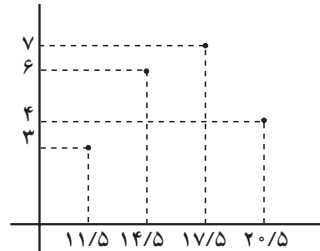
فراوانی برای جدول فراوانی روبرو رسم کنیم.



شکل ۷



شکل ۱۰



شکل ۹

ابتدا نقاط $(۱۱/۵, ۳)$ ، $(۱۴/۵, ۶)$ ، $(۱۷/۵, ۷)$ و $(۲۰/۵, ۴)$ را در صفحهٔ محورهای مختصات رسم می‌کنیم. توجه دارید که طول این نقاط مرکز دسته‌ها و عرض آن‌ها فراوانی همان دسته است. حال نقاط مذکور را متوالیاً به هم وصل می‌کنیم تا شکل ۱۰ حاصل شود.

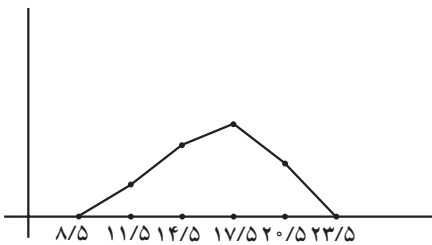
برای آن که سطح زیر این منحنی برابر مساحت نمودار مستطیلی شود (این نکته بعدها مورد استفاده زیاد قرار می‌گیرد) دو سر نمودار فوق را به محور x ‌ها به طریقی که توزیع داده می‌شود وصل می‌کنیم. فرض کنیم دو دستهٔ اضافی یکی قبل از دستهٔ اول و یکی بعد از دستهٔ آخر داشته باشیم، مسلماً فراوانی این دسته‌ها صفر است. پس نقاط نظیر این دسته‌ها عبارت خواهد بود از:

$$(۸/۵, ۰) \text{ و } (۲۳/۵, ۰)$$

(دقت کنید $۱۴/۵ - ۱۱/۵ = ۳$ طول دسته، پس $۱۱/۵ - ۳ = ۸/۵$ مرکز دسته فرضی ماقبل

دستهٔ اول و $۲۳/۵ = ۲۰/۵ + ۳$ مرکز دستهٔ فرضی آخر است) حال اگر این دو نقطه را به مجموعهٔ

نقاط قبلی اضافه کنیم نمودار چندبر فراوانی به صورت زیر تکمیل می‌شود.



شکل ۱۱

برای جدول ۴ هر سه نمودار میله‌ای، مستطیلی و چندبر فراوانی را روی یک دستگاه محورها رسم می‌کنیم تا ارتباط آن‌ها مشخص‌تر شود.

از روی شکل ۱۲ می‌توانید توضیح دهید که

چرا مساحت سطح زیر چندبر فراوانی و مساحت نمودار مستطیلی با هم برابر است؟ (به مساحت‌هایی که هاشور خورده است توجه کنید).

اگر به جای نمودار فراوانی، نمودار فراوانی

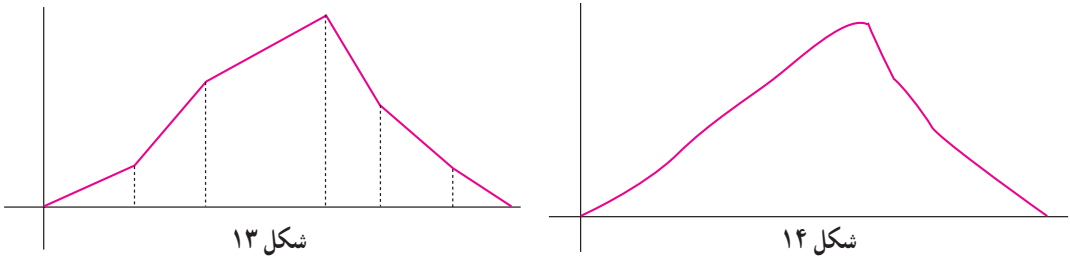
نسبی را رسم کنیم، اطلاعات منسجم‌تری در اختیار



شکل ۱۲

ما قرار می‌گیریم چون می‌توان فراوانی را با کل جامعه مقایسه کرد. نمودار چندبر فراوانی را می‌توان با داشتن فراوانی نسبی نیز رسم کرد و آن را چندبر فراوانی نسبی نامید.

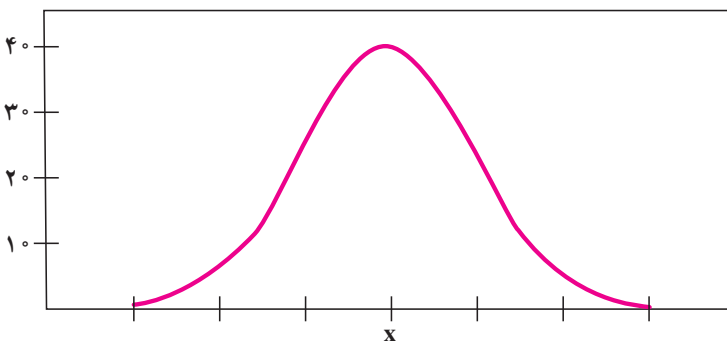
چندبر فراوانی نسبی برای داده‌های پیوسته مناسب است و می‌دانیم که ما بر اساس داده‌های حاصل از یک نمونه n تایی از یک جامعه آماری به نمودار چندبر فراوانی دست یافتیم. حال ممکن است سؤال شود که اگر ما اندازه نمونه را افزایش دهیم چه اتفاقی خواهد افتاد؟ چون تعداد داده‌ها زیاد می‌شود. بنا بر توصیه‌هایی که در جدول فراوانی به عمل آمد، تعداد دسته‌ها افزایش خواهد یافت. و در این صورت معمولاً طول دسته‌ها کاهش خواهد یافت و در نتیجه چندبر فراوانی از پاره‌خط‌های بیش‌تری تشکیل خواهد شد. اگر به همین ترتیب ادامه داده و تعداد داده‌ها را افزایش دهیم در نهایت به یک چندبر فراوانی دست خواهیم یافت که بیش‌تر به یک منحنی هموار شبیه است تا یک چندبر فراوانی. در شکل زیر مراحل شکل‌گیری چنین منحنی را ملاحظه می‌کنید.



شکل ۱۳

شکل ۱۴

این منحنی هموار، بیان‌کننده وضعیت متغیر در جامعه است. در اکثر مطالعات آماری این منحنی هموار شکل زیر را به خود می‌گیرد.



شکل ۱۵

این منحنی از منحنی‌های معروف در آمار است و در اکثر پدیده‌های طبیعی ظاهر می‌شود. ولی چون مقدمات کافی برای یادگیری آن را هنوز نمی‌دانیم به معرفی ظاهری آن پرداختیم. این منحنی متقارن و شبیه به زنگ است.

تمرین

۱- مدیر یک مؤسسه تاکسی سرویس مایل است بداند اتومبیل‌های این مؤسسه روزانه چند کیلومتر مسافت می‌کنند. برای این منظور کارکرد ۲۴ اتومبیل را یادداشت کرد. داده‌های زیر به دست آمدند.

۱۳۸	۱۰۷	۱۳۶	۱۲۸	۱۴۸	۱۱۸
۱۴۲	۱۲۹	۱۱۵	۱۲۳	۱۳۳	۱۲۳
۱۲۱	۱۲۸	۱۲۲	۱۴۴	۱۲۶	۱۳۵
۱۲۵	۹۸	۱۱۷	۱۵۳	۱۴۱	۱۲۶

برای این داده‌ها یک جدول فراوانی تنظیم کنید. داده‌ها را در شش دسته، دسته‌بندی کنید. برای نمایش این داده‌ها چه نموداری مناسب است. نمودار چندبر فراوانی را رسم کنید.

۲- نمودار روبرو نشان‌دهنده توزیع

سرعت باد در ۱۹ روز متوالی می‌باشد. با

استفاده از اطلاعات داده شده در نمودار، نمودار

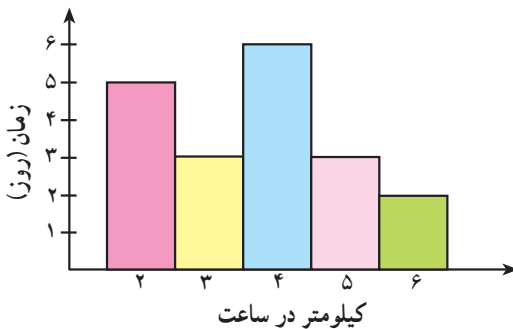
چندبر فراوانی را برای سرعت باد رسم کنید.

اگر حداقل سرعت باد لازم برای راندن

یک قایق بادی ۵ کیلومتر در ساعت سرعت

باد لازم باشد، چند روز برای راندن قایق بادی

مناسب است؟



شکل ۱۶

چرا در این مسئله، نمودار چندبر فراوانی مناسب‌تر از نمودار مستطیلی است؟





نمودار دایره‌ای

یکی دیگر از نمودارهایی که می‌تواند اطلاعات موجود در داده‌ها را به سرعت در معرض دید قرار دهد نمودار دایره‌ای است. فرض کنید متغیر تصادفی مورد مطالعه ما دارای k حالت باشد (مثلاً در مطالعه مقطع تحصیلی، متغیر تصادفی کیفی دارای ۴ حالت ابتدایی، راهنمایی، دبیرستان و دانشگاه است) فرض کنید فراوانی حالت اول f_1 و فراوانی حالت دوم f_2 و الی آخر باشد، در این صورت دایره‌ای به شعاع دلخواه را به وسیله زاویه‌های مرکزی به k قسمت تقسیم می‌کنیم به قسمی که اندازه زاویه مرکزی هریک از این قسمت‌ها متناسب با فراوانی آن قسمت باشد. پس زاویه مرکزی نظیر دسته اول عبارت است از:

$$\frac{360}{f_1 + \dots + f_k} \times f_1 = 360 \times \frac{f_1}{n}, \quad n = \text{تعداد کل داده‌ها}$$

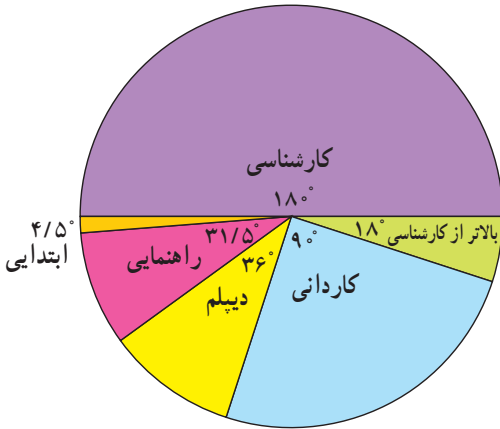
و به همین ترتیب می‌توانیم سایر زوایا را به دست آوریم. هریک از این قسمت‌ها در دایره نمایش دهنده سهم آن حالت از متغیر در جامعه است.

مثال: یک مؤسسه بزرگ ۴۰۰ نفر کارمند دارد که برحسب سطح سواد به صورت زیر توزیع

شده‌اند.

سطح سواد	راهنمایی ابتدایی	دیپلم	کاردانی	کارشناسی	کارشناسی بالاتر
تعداد کارمندان	۵	۳۵	۴۰	۱۰۰	۲۰۰

نمودار دایره‌ای مربوط به این داده‌ها به صورت زیر است :
 زاویه‌ها مطابق دستور به شرح زیر محاسبه شده‌اند :



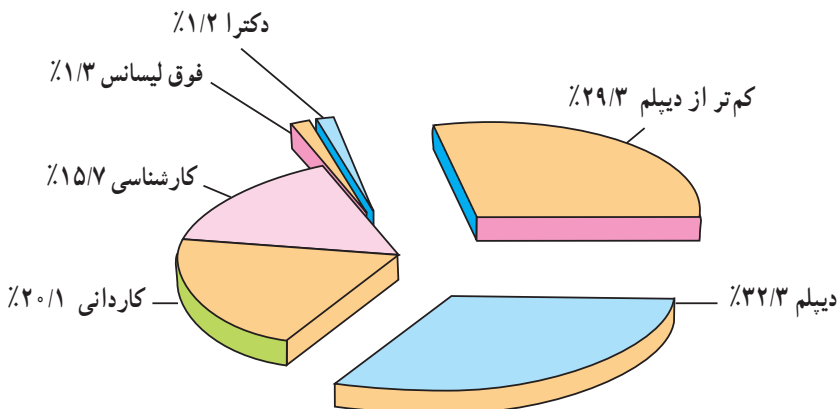
حالت	فراوانی	زاویه مرکزی برحسب درجه
ابتدایی	۵	$\frac{360}{400} \times 5 = 0/9 \times 5 = 4/5$
راهنمایی	۳۵	$0/9 \times 35 = 31/5$
دیپلم	۴۰	$0/9 \times 40 = 36$
کارردانی	۱۰۰	$0/9 \times 100 = 90$
کارشناسی	۲۰۰	$0/9 \times 200 = 180$
بالاتر از کارشناسی	۲۰	$0/9 \times 20 = 18$

در شکل ۱۶ دیده می‌شود که ناحیه مربوط به کارشناسی بزرگتر از سایر نواحی است و برابر نصف دایره می‌باشد. ناحیه مربوط به ابتدایی از همه کم‌تر است. ناحیه کارردانی نصف ناحیه کارشناسی است. در هر صورت بزرگی و کوچکی این ناحیه‌ها (در مقایسه با یک‌دیگر) بیان‌کننده فراوانی بیش‌تر و یا کم‌تر برای آن حالت از متغیر است.

جدول زیر توزیع کارکنان تابع قانون کار را در سال ۱۳۷۶ برحسب مدرک تحصیلی نشان می‌دهد.

نوع مدرک	کمتر از دیپلم	دیپلم	کارردانی	کارشناسی	کارشناسی ارشد	دکترا
تعداد	۱۵۲۳۹	۱۶۸۴۳	۱۲۵۳	۱۷۹۱	۱۷۲	۳۸

سالنامه آمار سال ۱۳۷۶



تمرین

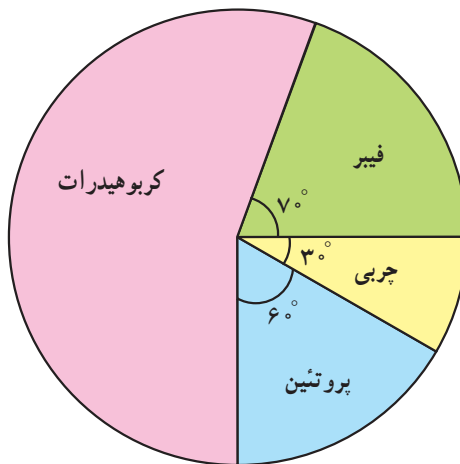
- ۱- نمودار دایره‌ای برای چه نوع متغیری مناسب است؟
- ۲- اگر فراوانی داده‌ها را دوبرابر کنیم آیا زاویه مرکزی عوض می‌شود؟ توضیح دهید.
- ۳- در یک نمودار دایره‌ای اگر یک زاویه مرکزی دوبرابر زاویه مرکزی دیگر باشد چه رابطه‌ای بین فراوانی‌های این دو زاویه وجود دارد؟
- ۴- آیا ترتیب کنار هم قرارگرفتن نواحی در نمودار دایره‌ای اهمیت دارد؟ توضیح دهید.
- ۵- گروه خونی همکلاسی‌های خود را بپرسید و برای اطلاعات جمع‌آوری شده نمودار دایره‌ای رسم کنید.
- ۶- با توجه به جدول زیر، نمودار دایره‌ای توزیع خانوارها برحسب جمعیت را رسم کنید.

نوع خانوار	کم جمعیت	با جمعیت متوسط	پرجمعیت	جمع
تعداد خانوار	۱۱	۲۳	۶	۴۰

- ۷- اگر شعاع دایره را تغییر دهیم آیا تفسیر جدیدی از توزیع داده‌ها به دست می‌آید یا خیر؟
- ۸- نمودار دایره‌ای زیر سهم وزنی ترکیبات تشکیل دهنده یک بسته غذای کنسرو شده را نشان می‌دهد.

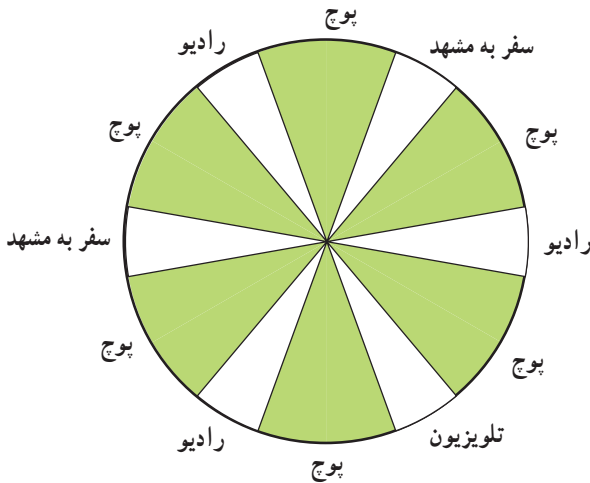
الف - چه کسری از این ترکیبات :

چربی است؟ کربوهیدرات است؟



- ب - چند گرم پروتئین در یک بسته ۳۶ گرمی از این محصول وجود دارد؟

۹- در یک مسابقه علمی، شرکت کنندگان پس از دادن پاسخ درست به سؤال با گرداندن یک صفحه گردان، جایزه دریافت می کنند. اگر نواحی رنگ شده 40° و نواحی بدون رنگ 20° باشند به سؤالات زیر پاسخ دهید:



- امکان بردن جایزه بیش تر است یا نبردن جایزه؟
- چند درصد شانس بردن رادیو وجود دارد؟
- چند درصد شانس بردن تلویزیون وجود دارد؟
- چند درصد شانس بردن سفر به مشهد وجود دارد؟

- چند درصد امکان نبردن جایزه وجود دارد؟
 - نمودار دایره ای نشان دهنده بردن جایزه را رسم کنید.

مخارج	درصد
غذا	۳۰٪
اجاره	۲۵٪
لباس	۱۵٪
کتاب	۱۰٪
دیگر مخارج	۲۰٪

۱۰- مخارج علی در سال اول دانشگاه در جدول روبرو آمده است:
 نمودار دایره ای هزینه های علی را رسم کنید.

۱۱- جدول زیر اطلاعات راجع به سطح زیر کشت و میزان تولید مرکبات کشور در سال ۷۳ را برای استان های ذکر شده نشان می دهد.

سطح زیر کشت (هکتار)

استان	نهال	بارور	میزان تولید (تن)	عمل کرد (کیلوگرم)
مازندران	۵۶۲۰	۷۶۷۷۵	۱۳۷۳۶۸۰	۱۷۸۹۲
فارس	۳۲۱۸	۳۹۱۵۱	۶۱۷۶۳۰	۱۵۷۷۶
کرمان	۱۶۹۱	۱۱۰۴۹	۹۴۸۷۶	۸۵۷۸
هرمزگان	۵۴۷۹	۲۴۲۶۰	۳۴۱۲۳۹	۱۳۸۴۹

آمارنامه کشاورزی، سال ۷۳- اداره کل آمار و اطلاعات، معاونت طرح و برنامه وزارت کشاورزی، نشریه شماره ۱۸، دی ماه ۷۴.
 نمودار دایره ای مربوط به میزان تولید مرکبات بر حسب تن را برای این استانها رسم کنید.

نمودار ساقه و برگ



آیا تاکنون فکر کرده‌اید که می‌توان از اعداد در تهیه یک نمودار استفاده کرد؟
آیا نموداری دیده‌اید که اعداد، شکل‌دهنده آن باشند؟

در یک بررسی که روی 150° نفر انجام شده بود اعلام شد که 45% صبحها قبل از ساعت ۸ صبح از خانه بیرون می‌روند. برای بررسی درستی این خبر، تحقیقی روی ۲۴ گروه هر یک شامل 150° نفر انجام شد. درصد افرادی که در هر گروه قبل از ساعت ۸ صبح از خانه بیرون می‌رفتند به شکل زیر گزارش داده شد:

۴۴	۴۵	۴۵	۴۶	۴۵	۴۶
۴۳	۴۷	۴۲	۴۴	۴۶	۴۴
۴۰	۴۷	۴۴	۴۸	۴۳	۴۵
۴۲	۴۵	۴۶	۴۳	۴۸	۴۷

همان طوری که می‌بینید تمام این اعداد در رقم ۴ مشترک هستند. یعنی تماماً اعداد طبیعی در فاصله 40° تا 50° هستند، از این رو بخش اصلی این اعداد را که همان 40° است می‌توانیم کنار بگذاریم و سایر ارقام را به صورت زیر بنویسیم:

$40 + 0, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5,$
 $6, 6, 6, 6, 7, 7, 7, 8, 8$

البته داده‌ها را در بالا از کوچک به بزرگ مرتب کرده‌ایم. یک راه ساده‌تر نوشتن اعداد فوق آن است که مثلاً عدد ۴۷ را به صورت ۴۷ بنویسیم (کمی فاصله بین ۴ و ۷). در این صورت عدد چهار را ثابت نگاه داشته و سایر ارقام را در مقابل آن به ترتیب می‌نویسیم. در مثال مورد بحث چون تعداد زیادی داده داریم که در عدد ۴ مشترک هستند می‌توانیم هر رقم را در یک ردیف به صورت روبرو بنویسیم.

شکل ۱۸

ملاحظه می‌کنید که این طور مرتب کردن اعداد، یک نمودار میله‌ای از داده‌ها مشخص می‌کند (برای آن که بهتر این نکته را ببینید می‌توانید صفحه کتاب را ۹۰ درجه دوران دهید). خوبی این نمودار در این است که تمام داده‌ها را دربر دارد. عدد چهار را که در واقع همان ۴۰ است و قسمت اصلی داده‌هاست ساقه می‌گوییم. سایر ارقام را که متصل به این ساقه هستند برگ می‌گوییم و شکلی که در صفحه قبل حاصل شد نمودار ساقه و برگ می‌گویند. در زیر شما چند نمونه از نمودارهای ساقه و برگ را ملاحظه می‌کنید. مثال: بعضی از مردم نسبت به گزیدگی پشه حساس‌اند. محل گزیدگی در این افراد تا مدتی همراه با خارش و سوزش است که پس از مدتی این آثار برطرف می‌شود. در این جا بخشهای مسأله آماری، عبارتند از:

جامعه: تمام کسانی که نسبت به گزیدگی پشه حساس‌اند.

متغیر تصادفی: مدت زمانی که طول می‌کشد تا آثار حساسیت محو شود.

نمونه: برای مطالعه این مدت زمان چون تمام افراد در اختیار ما نیستند لذا تعدادی از این افراد مثلاً ۴۰ نفر آنها را مورد مطالعه قرار داده‌ایم. داده‌های حاصل از این نمونه عبارتند از: (داده‌ها برحسب دقیقه اندازه‌گیری شده‌اند).

۱۰/۵	۱۱/۲	۹/۹	۱۵/۰	۱۱/۴	۱۲/۷	<u>۱۶/۵</u>	۱۰/۱
۱۲/۷	۱۱/۴	۱۱/۶	۶/۲	۷/۹	۸/۳	۱۰/۹	۸/۱
<u>۳/۸</u>	۱۰/۵	۱۱/۷	۸/۴	۱۲/۵	۱۱/۲	۹/۱	۱۰/۴
۹/۱	۱۳/۴	۱۲/۳	۵/۹	۱۱/۴	۸/۸	۷/۴	۸/۶
۱۳/۶	۱۴/۷	۱۱/۵	۱۱/۵	۱۰/۹	۹/۸	۱۲/۹	۹/۹

درخت دوستی بنشان که کام دل ببار آرد ...

۰۳
۰۴
۰۵
۰۶
۰۷
۰۸
۰۹
۱۰
۱۱
۱۲
۱۳
۱۴
۱۵
۱۶

برای یکنواختی، داده‌ها را به صورت یک عدد سه رقمی می‌نویسیم، مثلاً عدد $3/8$ را به صورت 038 و عدد $16/5$ را به صورت 165 می‌نویسیم، اگر عددی قسمت اعشاری نداشته مثلاً به صورت 12 بود آن را به صورت 120 می‌نویسیم. با این مقدمه نمودار ساقه و برگ را به صورت روبرو تشکیل می‌دهیم. ابتدا ساقه را تشکیل می‌دهیم، ساقه را به صورت دو رقم اول داده‌ها در نظر گرفته‌ایم. بنابراین ساقه به صورت روبرو است:

حال برای تشکیل برگ‌ها، هر داده را در محل مربوطه روی ساقه قرار می‌دهیم، مثلاً مقدار 112 (مربوط به داده $11/2$) را در محل 11 روی ساقه قرار می‌دهیم، و آن را به صورت 112 می‌نویسیم $11/5$ در همان محل و به صورت 115 نوشته می‌شود. پس این دو داده‌ی اخیر در محل 11 به صورت زیر نوشته می‌شوند:

۱۱ ۲۵

با ادامه این عمل برگ مربوط به اتصال 11 روی ساقه به صورت زیر تکمیل می‌شود:

۱۱ ۲۲۴۴۴۵۵۵۶۷

این سطر نشان می‌دهد که 9 داده داریم که با 11 شروع می‌شوند (داده‌های

$11/2, 11/2, 11/4, 11/4, 11/4, 11/4, 11/5, 11/5, 11/6, 11/7$)

با تکرار این عمل برای سایر مقادیر، نمودار ساقه و برگ به صورت زیر

تکمیل می‌شود:

۰۳ ۸
۰۴
۰۵ ۹
۰۶ ۲
۰۷ ۴۹
۰۸ ۱۳۴۶۸
۰۹ ۱۱۸۹۹
۱۰ ۱۴۵۵۹۹
۱۱ ۲۲۴۴۴۵۵۵۶۷
۱۲ ۳۵۷۷۹
۱۳ ۴۶
۱۴ ۷
۱۵ ۰
۱۶ ۵

کلید نمودار: $7/4 = 4/7$

در یک تست انگلیسی ۸۰ سؤاله نتایج زیر توسط دانش آموزان یک کلاس بدست آمده است.

یک نمودار ساقه و برگ برای داده‌های زیر رسم می‌کنیم:

۲۵	۳۰	۳۰	۳۷	۳۸	۳۸	۴۴	۴۴	۴۴
۴۹	۵۰	۵۰	۵۶	۵۷	۶۱	۳۱	۳۱	۳۲
۳۲	۳۲	۳۵	۳۵	۳۷	۳۹	۴۰	۴۰	۴۰
۴۰	۴۱	۴۲	۴۵	۴۵	۴۶	۴۶	۴۶	۴۷
۴۷	۴۷	۴۹	۵۰	۵۰	۵۰	۵۱	۵۲	۵۲
۵۲	۵۴	۵۶	۵۶	۵۶	۵۷	۶۱	۶۴	۶۶
۶۷	۷۱	۷۸						

در اینجا برای تشکیل ساقه‌ها از رقم دهگان و برای تشکیل برگها از رقم یکان استفاده می‌کنیم،

مثلاً عدد ۷۸ را به صورت

۷ ۸

می‌نویسیم که ۷ روی ساقه و ۸ روی برگها قرار دارد. با این ترتیب نمودار زیر به دست می‌آید:

ساقه	برگ
۲	۵
۳	۰ ۰ ۱ ۱ ۲ ۲ ۲ ۵ ۵ ۷ ۷ ۸ ۸ ۹
۴	۰ ۰ ۰ ۰ ۱ ۲ ۴ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۷ ۷ ۷ ۹ ۹
۵	۰ ۰ ۰ ۰ ۰ ۱ ۲ ۲ ۲ ۴ ۶ ۶ ۶ ۶ ۷ ۷
۶	۱ ۱ ۴ ۶ ۷
۷	۱ ۸

کلید نمودار: $۵۲ = ۲ = ۵$

برای انجام یک پروژه در کلاس دو هفته در نظر گرفته شده است. تعداد روزهایی که هر

دانش‌آموز صرف انجام پروژه کرده است در زیر آمده است:

۷	۷	۸	۸	۸	۸	۹	۹	۹	۹	۹
۹	۹	۹	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰
۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۱	۱۱	۱۱	۱۱	۱۱
۱۱	۱۱	۱۱	۱۱	۱۲	۱۲	۱۲	۱۲	۱۲	۱۳	۱۳
۱۴										

نمودار ساقه و برگ به این شکل خواهد بود:

۰	۷	۷	۸	۸	۸	۸				
۰	۹	۹	۹	۹	۹	۹	۹			
۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
۱	۱	۱	۱	۱	۱	۱	۱	۱		

کلید نمودار: $۱۰ = ۰ = ۱$

تمرین

ساقه	برگ
۱	۰ ۳ ۳ ۴
۲	۰ ۲ ۴ ۸ ۸
۳	۲

۱- نمودار ساقه و برگ روبرو داده شده است:
- داده‌های موجود در این نمودار را بنویسید.

ساقه	برگ
۱	۲ □ ۵ ۶
۲	۷ ۹
۳	۲ ۴ □

۲- در نمودار ساقه و برگ روبرو چه اعدادی
می‌توانند در □ قرار گیرند؟

۳- نمودار ساقه و برگ ارتفاع ۲۱۸ آتش‌فشان به شکل

۰	۲ ۵ ۶ ۶ ۶ ۷ ۸ ۹
۱	۰ ۱ ۳ ۶ ۷ ۷ ۹ ۹
۲	۰ ۰ ۰ ۱ ۱ ۲ ۲ ۲ ۴ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۷ ۷ ۸ ۸ ۹ ۹ ۹
۳	۰ ۱ ۱ ۲ ۲ ۴ ۴ ۵ ۵ ۵ ۶ ۶ ۶ ۷ ۸ ۹ ۹
۴	۰ ۱ ۱ ۱ ۲ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۶ ۷ ۸ ۸ ۹ ۹ ۹ ۹ ۹
۵	۰ ۰ ۱ ۱ ۲ ۲ ۲ ۳ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۶ ۶ ۷ ۷ ۷ ۹ ۹
۶	۰ ۰ ۱ ۱ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۶ ۷ ۷ ۷ ۸ ۸ ۹
۷	۰ ۰ ۰ ۰ ۱ ۱ ۱ ۲ ۳ ۳ ۴ ۵ ۵ ۵ ۶ ۷ ۸ ۸ ۹
۸	۱ ۲ ۲ ۲ ۲ ۳ ۳ ۳ ۵ ۶ ۷ ۹
۹	۰ ۰ ۰ ۱ ۲ ۳ ۳ ۴ ۴ ۵ ۵ ۶ ۷ ۷ ۹
۱۰	۰ ۱ ۱ ۲ ۲ ۳ ۳ ۴ ۴ ۵ ۶ ۸ ۹
۱۱	۰ ۱ ۱ ۲ ۳ ۳ ۴ ۶ ۶ ۹
۱۲	۱ ۱ ۲ ۴ ۴ ۴ ۵ ۶
۱۳	۰ ۳ ۴ ۷ ۸
۱۴	۰ ۰
۱۵	۶ ۶ ۷
۱۶	۲ ۵
۱۷	۲ ۹
۱۸	۵
۱۹	۰ ۳ ۳ ۷ ۹

زیر است.

(سانتی‌متر = ۳۰ = ۱ فوت)

واحد: ۱۰۰ فوت

فوت = ۹۰۰ = ۹

فوت = ۱۹۳۰۰ = ۳ = ۱۹

- ارتفاع بلندترین آتش فشان چه قدر است؟
 – چند آتش فشان ارتفاعی بین ۱۰۰۰۰ و ۱۲۰۰۰ فوت دارند؟
 – چند آتش فشان با ارتفاع ۴۳۰۰ فوت وجود دارد؟

۱۰	۵۰	۶۵	۳۳	۴– داده‌های روبرو، زمان صرف شده برحسب
۴۸	۵	۱۱	۲۳	دقیقه برای رسیدن به محل کار ۲۰ کارمند را نشان می‌دهد.
۳۷	۲۶	۲۶	۳۲	– نمودار ساقه و برگ این داده‌ها را رسم کنید.
۱۷	۷	۱۳	۱۹	
۲۹	۴۳	۲۱	۲۲	

۵– نمرات تست زبان دو کلاس (از ۱۰۰) در نمودار ساقه و برگ در زیر نشان داده شده

است :

- کم‌ترین نمره در هر یک از کلاس‌های دختران و پسران چند است؟
 – چند نفر در کلاس دختران و چند نفر در کلاس پسران نمره بالاتر از ۶۰ کسب کرده‌اند؟

– جدول فراوانی داده‌های فوق را با توجه به دسته‌بندی

- $F \leftarrow 0 \leq X < 60$
 $D \leftarrow 60 \leq X < 70$
 $C \leftarrow 70 \leq X < 80$
 $B \leftarrow 80 \leq X < 90$
 $A \leftarrow 90 \leq X \leq 100$

روبرو رسم کنید.

برگ	ساقه	برگ
کلاس پسران		کلاس دختران
	۲	۹
۸ ۰	۳	۴
۹ ۴	۴	۷ ۸
۹ ۱ ۰ ۰	۵	۰ ۱ ۸
۹ ۸ ۸ ۸ ۶ ۰ ۰	۶	۰ ۲ ۲ ۴ ۸ ۸ ۹
۹ ۸ ۸	۷	
۶ ۵ ۵ ۵	۸	۱ ۲ ۲ ۴ ۵ ۵ ۸ ۹
۸ ۷ ۷ ۴ ۴ ۱ ۰	۹	۳ ۴ ۴ ۴ ۷ ۷
۰	۱۰	۰ ۰

– اگر شاخص‌های A, B, C, D و F برای نتایج امتحان استفاده شود، نمودار میله‌ای را برای هر یک از کلاس‌ها رسم کنید.

– نمودار مستطیلی هر کلاس را رسم کنید.

– کدام کلاس درصد قبولی بالاتری دارد؟ (نمرات A, B, C و D قبول می‌باشند)

۶ – مربی تیم بسکتبال یک مدرسه جهت تشکیل تیم بسکتبال مدرسه، طول قد دانش‌آموزان دو کلاس را اندازه‌گیری نموده است. طول قد‌های دانش‌آموزان این دو کلاس در جدول‌های زیر آمده است:

کلاس الف	کلاس ب
۱۳۴, ۱۱۹, ۱۳۷, ۱۴۰, ۱۶۳,	۱۴۴, ۱۴۲, ۱۲۶, ۱۴۳, ۱۵۴,
۱۴۷, ۱۳۵, ۱۶۸, ۱۳۱, ۱۶۰,	۱۶۴, ۱۶۵, ۱۶۶, ۱۲۸, ۱۵۷,
۱۱۶, ۱۴۴, ۱۴۶, ۱۳۳, ۱۵۱,	۱۵۶, ۱۲۹, ۱۴۰, ۱۲۸, ۱۴۷,
۱۳۰, ۱۳۷, ۱۳۷, ۱۲۵, ۱۲۹,	۱۵۹, ۱۲۲, ۱۵۵, ۱۴۶, ۱۶۹,
۱۴۲, ۱۶۹, ۱۴۷, ۱۱۷,	۱۱۲, ۱۵۸, ۱۳۱, ۱۳۳, ۱۴۴

الف – نمودار ساقه و برگ برای هر یک از جدول‌ها را رسم کنید.

ب – کوتاه‌ترین دانش‌آموز در کدام کلاس است؟

روشن از پرتو رویت نظری نیست که نیست ...

- ج - بلندقدترین دانش آموز در کدام کلاس است؟
د - قد چند دانش آموز در کلاس الف کوتاه تر از 110° سانتی متر است؟
ه - اگر شرط عضویت در تیم بسکتبال مدرسه قد حداقل 120° سانتی متر باشد، چه تعداد از دانش آموزان هر کلاس واجد شرط عضویت در تیم بسکتبال می باشند؟

پروژه

به عنوان پروژه آماری، محمد می بایست یک نمونه 30° تایی از ماشین هایی که در یک پارکینگ در کنار یک ورزشگاه بزرگ پارک شده بودند انتخاب نموده و نام ماشین و سال ماشین را به دست آورد.

الف - اگر در این پارکینگ 950° ماشین پارک شده باشد، توضیح دهید که محمد با استفاده از اعداد تصادفی چگونه می تواند نمونه خود را انتخاب کند.

ب - پس از این که محمد نمونه خود را انتخاب کرد، او متوجه شد که ۸ ماشین پیکان، ۷ ماشین پژو، ۵ ماشین رنو، ۳ ماشین تویوتا، ۳ ماشین فولکس، ۳ ماشین دوو و یک ماشین بنز بود. جدول فراوانی داده های به دست آمده را تشکیل دهید و نمودار میله ای آن را رسم کنید.

- نمودار دایره ای این نمونه را بر اساس درصد فراوانی ماشین ها رسم کنید.

- در قسمت ب، چه نوع متغیری داریم؟ آیا می توانیم برای این متغیر نمودار مستطیلی رسم کنیم؟

- سال ماشین ها چه نوع متغیری است؟ آیا می توان برای این متغیر نمودار مستطیلی رسم کرد؟

رسم نمودارها با استفاده از MINITAB



در این بخش چگونگی رسم نمودار میله‌ای، مستطیلی، دایره‌ای و ساقه و برگ را خواهید آموخت.

نمودار دایره‌ای و نمودار میله‌ای

مثال ۱: جدول فراوانی رشته‌های تحصیلی ۲۵ دانشجو در دست است.

با استفاده از MINITAB نمودار میله‌ای و دایره‌ای این اطلاعات را رسم کنید.

حل: اگر از MINITAB WINDOWS استفاده می‌کنید مراحل زیر را دنبال کنید:

رشته تحصیلی	تعداد دانشجویان
اقتصاد	۶
ریاضی	۳
فیزیک	۶
شیمی	۲
ادبیات	۸

قدم اول: رشته‌های تحصیلی و فراوانی دانشجویان هر رشته را در ستون‌های C۱ و C۲ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

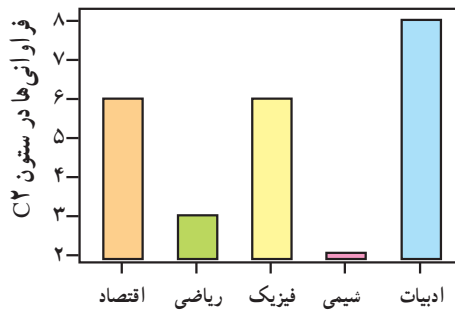
قدم سوم: روی CHART کلیک کنید.

قدم چهارم: در کادر، زیر Y، تایپ کنید C2 و زیر X، تایپ کنید C1
قدم پنجم: روی OK کلیک کنید.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، ابتدا رشته‌های تحصیلی و فراوانی دانشجویان را در ستون‌های C1 و C2 با استفاده از فرمان SET وارد کنید. توجه داشته باشید برای وارد کردن رشته‌ها باید از فرمان FORMAT استفاده کنید. سپس فرمان زیر را تایپ کنید:

MTB > CHART C2 * C1

بدون در نظر گرفتن این که از کدام طریق شما کار را به پایان رسانیده‌اید، نمودار میله‌ای زیر را خواهید داشت.



برای رسم نمودار دایره‌ای چنین عمل کنید:

قدم اول: رشته‌های تحصیلی و فراوانی دانشجویان هر رشته را در ستون‌های C1 و C2 وارد کنید.
قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی PIE CHART کلیک کنید.

قدم چهارم: روی CHART TABLE کلیک کنید. سپس کنار CATEGORIES IN

کادر، C1 تایپ کنید و در کنار FREQUENCIES IN در کادر C2 تایپ کنید.

قدم پنجم: روی OK کلیک کنید.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید ابتدا رشته‌های تحصیلی و فراوانی

دانشجویان را در ستون‌های C1 و C2 وارد کنید. سپس فرمان‌های زیر را تایپ کنید.

MTB > %PIE C1 ;

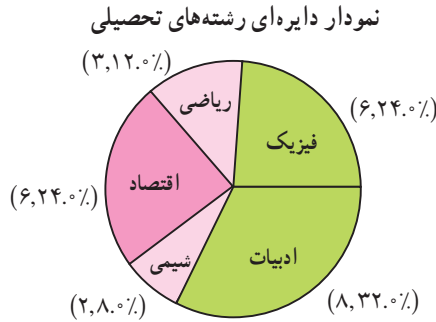
SUBC > COUNTS C2 .

اولین فرمان از MINITAB می‌خواهد نمودار دایره‌ای برای رشته‌های موجود در ستون C1

رسم کند و زیر فرمان COUNTS فراوانی هر یک از رشته‌ها را که در ستون C2 آمده است ارائه

می‌دهد.

بدون در نظر گرفتن این که شما از چه روشی کار را به پایان رسانیده‌اید، نمودار دایره‌ای زیر رسم می‌شود.



نمودار مستطیلی

مثال ۲: در جدول زیر قد 30° بازیکن بسکتبال به طور تصادفی انتخاب شده‌اند (برحسب اینچ، ۱ اینچ = $\frac{2}{5}$ سانتی متر است) آمده است.

۸۱	۸۴	۷۹	۷۶	۷۳	۷۴	۷۷	۸۲
۷۵	۸۱	۷۶	۷۶	۸۰	۸۲	۷۸	۷۲
۸۰	۸۳	۸۰	۷۷	۷۸	۷۸	۷۹	۸۴
۷۳	۸۶	۸۳	۷۹	۸۳	۷۹		

جدول فراوانی داده‌ها با فرض داشتن ۵ دسته به شکل زیر خواهد بود.

مرکز داده‌ها	f	قد برحسب اینچ
۷۳/۵	۴	۷۲_۷۵
۷۶/۵	۶	۷۵_۷۸
۷۹/۵	۱۰	۷۸_۸۱
۸۲/۵	۷	۸۱_۸۴
۸۵/۵	۳	۸۴_۸۷

طول دسته‌ها = ۳

برای رسم نمودار مستطیلی داده‌های فوق با استفاده از MINITAB چنین عمل کنید:
 قدم اول: داده‌ها را در ستون C۱ وارد کنید.
 قدم دوم: روی GRAPH در لیست انتخاب، کلیک کنید.

قدم سوم: روی HISTOGRAM کلیک کنید.

قدم چهارم: C1 را در کادر، زیر X تایپ کنید.

قدم پنجم: روی OPTION در پایین این پنجره کلیک کنید.

قدم ششم: روی دایره‌ای که در کنار MID POINT زیر TYPE OF INTERVALS قرار دارد کلیک کنید. سپس روی دایره که در کنار DEFINITION OF MIDPOINT/CUT POINT POSITION زیر MIDPOINT / CUT POINT POSITION قرار دارد کلیک کنید.

73.5:85.5/3. توجه داشته باشید که ۷۳/۵ مرکز دسته اول و ۸۵/۵ مرکز دسته آخر و

۳ طول دسته‌ها می‌باشد.

قدم هفتم: روی OK در پایین صفحه کلیک کنید. سپس مجدداً روی OK کلیک کنید.

نمودار مستطیلی ظاهر می‌شود.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، مرکز دسته‌ها را در ستون C1 با استفاده

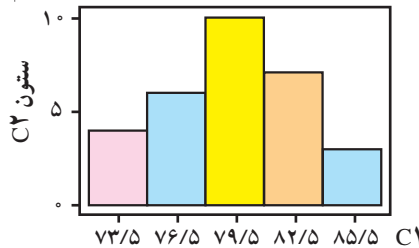
از فرمان SET وارد کنید، سپس فرمان‌های زیر را تایپ کنید :

MTB > HISTOGRAM C1 ;

SUBC > MIDPOINTS 73.5:85.5/3.

توجه داشته باشید از ";" و "." استفاده صحیح همان‌گونه که قبلاً توضیح داده شد، استفاده

شود. بدون توجه به روش استفاده شده، نمودار مستطیلی زیر رسم خواهد شد :



در مثال بالا دسته‌ها را خودمان انتخاب کردیم. اگر می‌خواهید MINITAB خودش دسته‌ها

و طول دسته‌ها را انتخاب کند در MINITAB WINDOWS قدم پنجم و ششم را اجرا

نکنید و در فرمان‌های زبانی MINITAB فقط HISTOGRAM C1 تایپ کنید و زیر

فرمان را تایپ نکنید.

اگر می‌خواهید نمودار چندبر فراوانی با استفاده از MINITAB WINDOWS رسم کنید، روی GRAPH در لیست انتخاب کلیک کنید و سپس HISTOGRAM را انتخاب کنید و زیر X در کادر، C۱ تایپ کنید. روی DISPLAY کلیک کنید. روی CONNECT کلیک کنید، روی OPTION در پایین کادر کلیک کنید، روی MIDPOINT/CUT POINT POSITIONS کلیک کنید، در کادر 73.5:85.5/3 تایپ کنید و در پایان روی OK دوبار کلیک کنید. MINITAB چندبر فراوانی را رسم خواهد کرد.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، فرمان‌های زیر را تایپ کنید :

```
MTB > HISTOGRAM C1 ;  
SUBC > MIDPOINTS 73.5:85.5/3 ;  
SUBC > CONNECT.
```

نمایش نمودار ساقه و برگ

نمرات زبان ۳۰ دانش‌آموز در یک مؤسسه زبان (از ۱۰۰ نمره) به‌قرار زیر است :

۷۵ ، ۵۲ ، ۸۰ ، ۹۶ ، ۶۵ ، ۷۹ ، ۷۱ ، ۸۷ ، ۹۳ ، ۹۵
۶۹ ، ۷۲ ، ۸۱ ، ۶۱ ، ۷۶ ، ۸۶ ، ۷۹ ، ۶۸ ، ۵۰ ، ۹۲
۸۳ ، ۸۴ ، ۷۷ ، ۶۴ ، ۷۱ ، ۸۷ ، ۷۲ ، ۹۲ ، ۵۷ ، ۹۸

برای نمایش نمودار ساقه و برگ این ۳۰ دانش‌آموز با استفاده از MINITAB WINDOWS مراحل زیر را انجام دهید :

قدم اول: داده‌ها را در ستون C۱ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی CHARACTER GRAPHS کلیک کنید.

قدم چهارم: روی STEM - AND - LEAF کلیک کنید.

قدم پنجم: C۱ را در کادر زیر Variable و ۱۰ را در کادر در کنار Increment تایپ کنید.

قدم ششم: روی OK کلیک کنید. نمودار ساقه و برگ نمایش داده خواهد شد.

اگر شما از فرمان‌های زبانی MINITAB استفاده می‌کنید ابتدا داده‌ها را با استفاده از فرمان

SET در C۱ وارد کنید و سپس فرمان‌های زیر را تایپ کنید :

```
MTB > STEM - AND - LEAF C1;  
SUBC > INCREMENT = 10.
```

در زیر فرمان INCREMENT ، ۱۰ فاصله بین دو ساقه متوالی در نمودار را نشان می‌دهد.

بدون در نظر گرفتن روش کار، نمودار ساقه و برگ زیر نمایش داده می‌شود:

STEM – AND – LEAF OF C1 N = 30

LEAF UNIT = 1.0

3 5 0 2 7

8 6 1 4 5 8 9

(9) 7 1 1 2 2 5 6 7 9 9

13 8 0 1 3 4 6 7 7

6 9 2 2 3 5 6 8

در این نمایش $N = 30$ تعداد داده‌ها را نشان می‌دهد. LEAF UNIT = 1.0 نشان می‌دهد که 0.1 ، 0.2 ، 0.3 ، 0.4 ، 0.5 و دومین عدد 0.52 ... خواهد بود. اگر Leaf Unit = 0.01 باشد، اولین عدد 5.0 و دومین عدد 5.2 و سومین عدد 5.7 ... خواهد بود.

از طرف دیگر اگر LEAF UNIT = 10 باشد، اولین عدد 50 ، دومین عدد 520 و سومین عدد 570 ... خواهند بود. عددی که داخل پرانتز در ستون اول، MINITAB چاپ می‌کند (در این جا 9) نشان دهنده ردیفی است که میانه در آن قرار دارد. عدد 9 نشان دهنده تعداد برگ‌های ساقه در این ردیف می‌باشد. اعداد بالا و پایین این عدد فراوانی تجمعی دسته‌های قبل و بعد از این دسته را نشان می‌دهند.



تمرین

۱- فرض کنید از ۵۰ دانش‌آموز دبیرستانی پرسیده شده است که سال چندم هستند. جدول زیر

نتیجه این بررسی را نشان می‌دهد.

اول	سوم	اول	اول	سوم	پیش‌دانشگاهی	دوم	اول
دوم	دوم	دوم	اول	دوم	دوم	اول	اول
پیش‌دانشگاهی	سوم	دوم	اول	اول	پیش‌دانشگاهی	سوم	سوم
سوم	سوم	پیش‌دانشگاهی	سوم	اول	دوم	دوم	پیش‌دانشگاهی
سوم	دوم	پیش‌دانشگاهی	اول	اول	پیش‌دانشگاهی	پیش‌دانشگاهی	اول
اول	دوم	پیش‌دانشگاهی	اول	سوم	سوم	دوم	پیش‌دانشگاهی
پیش‌دانشگاهی	دوم	پیش‌دانشگاهی	سوم	سوم	پیش‌دانشگاهی	پیش‌دانشگاهی	پیش‌دانشگاهی

– جدول فراوانی این داده‌ها را تشکیل دهید.

– نمودار میله‌ای و دایره‌ای این داده‌ها را رسم کنید. (با استفاده از MINITAB)

۲– داده‌های زیر وزن ۲۰ بسته پستی را برحسب کیلوگرم که در هفته گذشته از یکی از شعبه‌های اداره پست، ارسال شده‌اند نشان می‌دهد.

۱/۸	۷/۵	۸/۲	۳/۴	۵/۱	۹/۳	۱/۹	۲/۵
۷/۳	۵/۸	۶/۲	۸/۶	۲/۰	۶/۳	۸/۵	۰/۷
۳/۸	۷/۳	۵/۲	۳/۷				

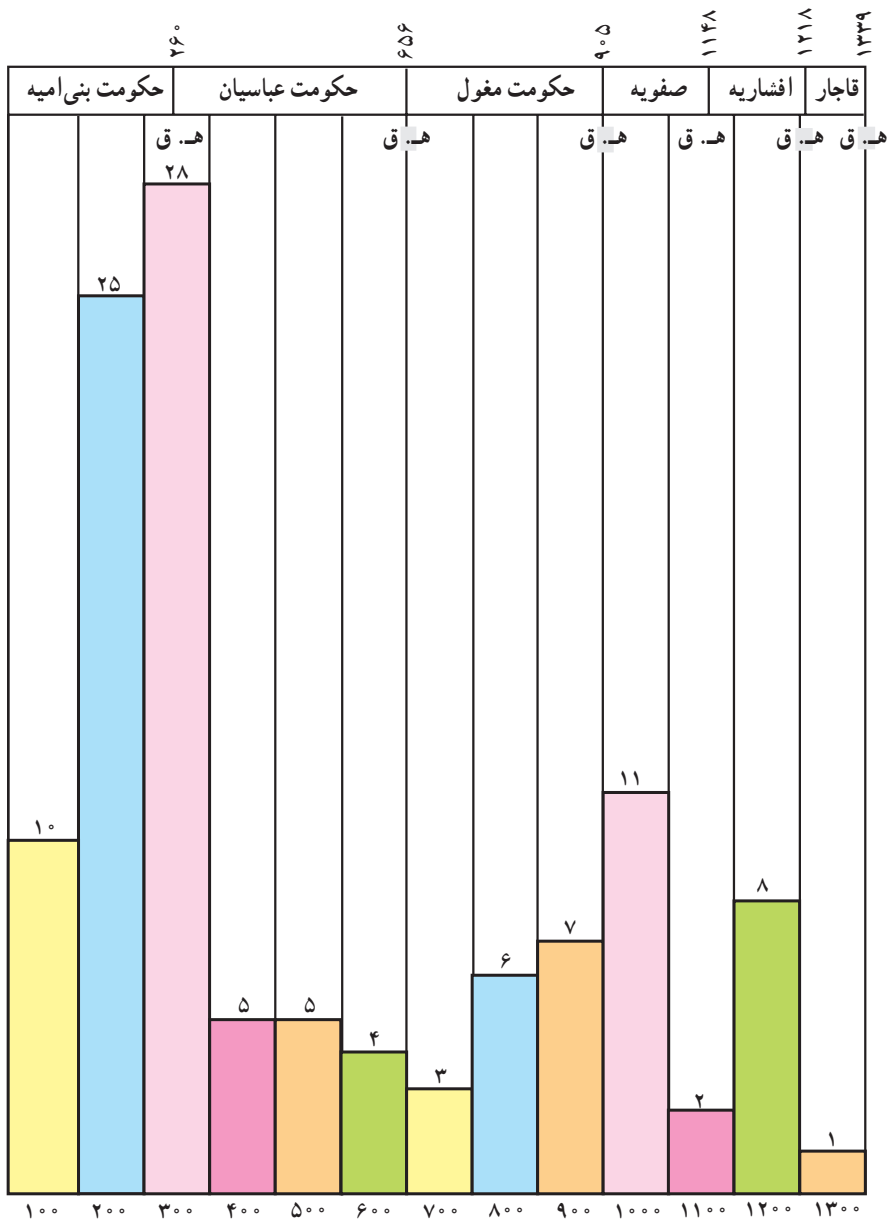
– با استفاده از MINITAB، نمودار ساقه و برگ این داده‌ها را با در نظر گرفتن $INCREMENT = 1.0$ رسم کنید. مقادیر برگ‌ها را مشاهده کنید.

۳– داده‌های زیر تعداد صفحه کلیدهای جمع شده را در یک شرکت الکترونیکی نشان می‌دهد.

۴۵	۵۲	۴۸	۴۱	۵۶	۴۶	۴۴	۴۲	۴۸
۵۳	۵۱	۵۳	۵۱	۴۸	۴۶	۴۳	۵۲	۵۰
۵۴	۴۷	۴۴	۴۷	۵۰	۴۹	۵۲		

– جدول فراوانی این داده‌ها را با داشتن ۴ دسته تشکیل دهید.

– نمودار مستطیلی و چندبر فراوانی این داده‌ها را با استفاده از MINITAB رسم کنید.



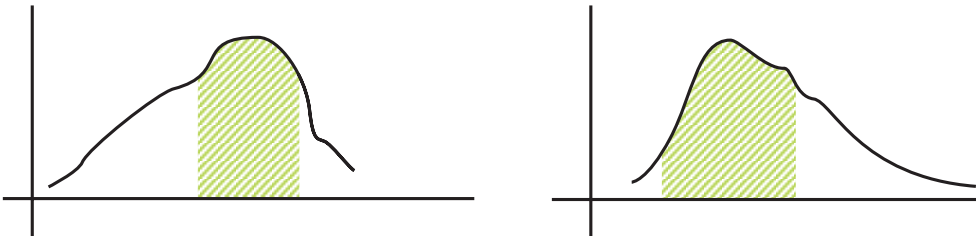
نقل از: نگارنامه تاریخی و زندگی‌نامه دانشمندان ریاضی و نجوم ایران در دوره اسلام

نمودار تعداد دانشمندان ریاضی و نجوم ایران در هر صد سال از قرن اول هجری تا پایان قرن ۱۳ هجری قمری

آفرینش را آغاز کرد و آفریدگان را به یکبار پدید آورد، بی آنکه اندیشه‌ای به کار برد، یا از آزمایشی سود برد یا جنبشی پدید آرد یا کسی را به خدمت گمارد. از هر چیز بهنگام پرداخت، و اجزای مخالف را با هم سازوار ساخت، و هر طبیعت را اثری بداد و آن اثر را در ذات آن نهاد. از آن پیش که بیافریند به آفریدگان دانا بود. بر آغاز و انجامشان بینا، و با سرشت و چگونگی آنان آشنا.

شاخص‌های مرکزی

در فصل‌های قبل اشاره کردیم که یک اظهارنظر را وقتی می‌توان علمی تلقی کرد که همراه با عدد و رقم باشد. تا به حال ما به‌جز دامنه تغییرات، کمیتی که بتواند جامعه را معرفی کند ارائه نکرده‌ایم، آن‌چه که تا به حال مطرح شده است، درباره کم کردن گوناگونی (از طریق دسته‌بندی آن‌ها) و یا نمایش هندسی داده‌هاست. در این فصل سعی می‌کنیم کمیت‌هایی را معرفی کنیم که بتوانند به‌صورت کمی، جامعه را معرفی نمایند. در رسم نمودارهای چندبر فراوانی و یا مستطیلی توجه کرده‌اید که معمولاً فراوانی داده‌ها در نقطه‌ای بیش‌تر از سایر نقاط است. به نمودارهای زیر توجه کنید:



در نمودارهای بالا در فاصله‌های هاشوردار فراوانی از سایر قسمت‌ها بیش‌تر است. در واقع به‌نظر می‌رسد که داده‌ها بیش‌تر در این قسمت متمرکز شده‌اند. ما می‌خواهیم شاخصی را معرفی کنیم که محل تمرکز داده‌ها را معرفی کند. از این‌رو این شاخص‌ها را شاخص‌های مرکزی می‌گوییم. البته تاکنون از این قبیل شاخص‌ها بدون آن‌که معرفی شده باشند استفاده می‌کردیم، مثلاً:

— اگر بخواهند وضعیت تحصیلی شما را بدانند، اولین اطلاعی که از شما می‌خواهند، معدل نمرات شماست.

— اگر بخواهند میزان علاقه شما را درباره زمینه‌های مختلف علمی بدانند از شما درباره تعداد کتاب‌هایی که در زمینه‌های مختلف مطالعه کرده‌اید، سؤال می‌شود. مثلاً می‌پرسند در چه زمینه‌ای بیش‌ترین تعداد کتاب را مطالعه کرده‌اید؟

– وضعیت تحصیلی شما در کلاس چگونه است؟ چه تعدادی از دانش‌آموزان کلاس را از نظر تحصیلی بهتر از خود ارزیابی می‌کنید. از چند نفر از دانش‌آموزان کلاس خود از نظر تحصیلی بهتر هستید؟ کدام یک از این دو عدد بزرگ‌تر است؟ مایلید کدام عدد بزرگ‌تر باشد؟ ملاحظه می‌کنید که این سؤال‌ها هر کدام به نوعی وضع تحصیلی شما را معرفی می‌کنند. ما در زیر به معرفی شاخص‌های نهفته در این سؤال‌ها می‌پردازیم. مهم‌ترین شاخص‌های مرکزی عبارتند از: مد، میانه و میانگین، که به معرفی آن‌ها می‌پردازیم.

خسروا گوی فلک در خم چوگان تو باد ...



مد

یک کارخانه اتومبیل‌سازی می‌خواهد برای کشوری اتومبیل‌های سواری تولید کند. مدیر کارخانه می‌خواهد بداند کدام نوع اتومبیل مشتری‌های بیش‌تری دارد.

اتومبیل‌های با ظرفیت ۲ نفر
اتومبیل‌های با ظرفیت ۴ نفر
اتومبیل‌های با ظرفیت ۶ نفر

مدیر این کارخانه برای آن‌که بتواند تصمیم بگیرد چاره‌ای جز نمونه‌گیری ندارد. برای این منظور در یک نمونه‌گیری، از ۱۱۴ اتومبیل در حال حرکت مشاهده شده است که:

۴۸ دستگاه دارای ۱ یا ۲ سرنشین بوده‌اند.
۳۸ دستگاه دارای ۳ یا ۴ سرنشین بوده‌اند.
۲۸ دستگاه دارای ۵ یا ۶ سرنشین بوده‌اند.

این اعداد و ارقام نشان می‌دهد که بیش‌تر اتومبیل‌ها کم سرنشین بوده‌اند. پس اتومبیل‌های با ظرفیت ۶ نفره در این کشور استفاده کم‌تری دارد. در این مثال اگر ماکزیمم تعداد سرنشین‌ها حساب شود، شاخصی به دست می‌آید که از نظر اقتصادی ممکن است به صرفه نباشد زیرا به علت وجود تعدادی

اتومبیل‌های با سرنشین‌های ۵ یا ۶ نفر ماکزیمم ۶ به دست می‌آید و لذا مدیر کارخانه تصمیم به تولید اتومبیل‌های بزرگ می‌گیرد که در این صورت هم از نظر قیمت و هم از نظر مصرف و هم از نظر تولید آلودگی، کاری زیان‌بار انجام شده است. ما برای مطالعات خود باید شاخص‌های مناسب را انتخاب کنیم.

مد، داده‌ای است که بیش‌ترین فراوانی را دارد.

در مثال بالا، نوع سرنشین کم، بیش‌ترین فراوانی (۴۸) را دارد پس مد مثال بالا اتومبیل کوچک است.

فروشنندگان پوشاک از شاخص مد بسیار استفاده می‌کنند. آن‌ها با بررسی‌های خود درمی‌یابند که چه نوع پوشاکی مورد پسند مصرف‌کنندگان است. از همان پوشاک برای فروش سفارش می‌دهند، چون در شهریور ماه خرید لوازم التحریر خواهان بیش‌تری دارد، از این نوع اجناس بیش‌تر تولید و عرضه می‌کنند. در اعیاد بیش‌تر به تولید شیرینی و پوشاک و سایر ملزومات می‌پردازند. در نزدیکی عید، فروشنده نگران مشتری‌هایی که ممکن است احتیاج به کالایی داشته باشند که در این روزها مصرف چندانی ندارد، نیست.

مد ممکن است منحصر به فرد نباشد. مثلاً داده‌های

۱، ۱، ۲، ۳، ۴، ۱، ۲، ۵، ۷، ۵، ۵

دارای دو مد ۵ و ۱ است. این قبیل جامعه‌ها را ۲ مدی می‌گویند. جامعه ممکن است چندمدی هم باشد. مد در این قبیل جامعه‌ها شاخص معتبری نیست. زیرا وجود چندین مد، نشان‌دهنده این واقعیت است که جامعه ما یک‌دست نیست (یادآوری می‌کنیم ما جامعه‌های یک‌دست را بررسی می‌کنیم) و لذا اگر مطالعات را دقیق‌تر کنیم، شاید بتوانیم جامعه را به دو بخش یک‌دست تفکیک کنیم. مثلاً ممکن است در بررسی قد افراد به ۲ مد برخورد کنیم که ممکن است نشان‌دهنده این حقیقت باشد که ما در نمونه‌گیری بزرگسالان و خردسالان را تفکیک نکرده‌ایم و لذا جامعه‌ای یک‌دست نداریم.

در رأی‌گیری‌ها، اساس تصمیم‌گیری مد است چون موضوعی که بیش‌ترین فراوانی را داشته باشد، انتخاب می‌شود. در انتخاب رئیس‌جمهور، 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.



میانه

قضاوت شما درباره کلاسی که نصف دانش آموزان نمره بالای ۱۷ دارند چیست؟ اگر نمره نصف دانش آموزان بالای ۱۳ باشد قضاوت شما چگونه خواهد بود؟ در حالت اول خواهید گفت کلاس، کلاس خوبی است و دانش آموزان و دبیران در آن خوب کار کرده اند. حتی نصف دیگر کلاس که نمره کم تر از ۱۷ دارند ممکن است تعدادی از نمره ها ۱۶، ۱۵ و غیره باشند که باز هم نشان دهنده وضع خوب کلاس است. اما در حالت دوم نصف دانش آموزان نمره زیر ۱۳ دارند که این نشانه ضعف در کلاس است و لذا باید درباره آن چاره اندیشی کرد. نمره ۱۷ در حالت اول و نمره ۱۳ در حالت دوم را میانه می گوئیم. میانه یکی از شاخص های مرکزی است که می توان درباره وضعیت جامعه از آن استفاده کرد.

پس از مرتب کردن داده‌ها، مقداری را که تعداد داده‌های بعد از آن با
تعداد داده‌های قبل از آن برابر است، میانه می‌نامیم.

آیا لزوماً داده‌ها باید به صورت صعودی مرتب شوند؟ آیا می‌توان داده‌ها را به صورت نزولی نیز
مرتب کرد؟

تعداد روزهایی که ۵ خانواده در تعطیلات عید به مسافرت رفته‌اند در زیر آمده است.

۱ ۷ ۶ ۴ ۳

می‌خواهیم میانه این داده‌ها را محاسبه کنیم. پس باید داده‌ها را مقایسه کنیم و لازم است آن‌ها
را مرتب کنیم.

۱ ۳ ۴ ۶ ۷

حال ملاحظه می‌شود که عدد ۴ در بالا دارای ویژگی زیر است:

تعداد داده‌های بعد از ۴ تا ۲

تعداد داده‌های قبل از ۴ تا ۲

این دو تعداد با هم برابرند لذا، میانه داده‌های بالا برابر ۴ است. می‌بینید که در این مثال نصف
تعداد داده‌ها که ۲/۵ است معنی ندارد.

آیا میانه همواره عضوی از مجموعه داده‌ها است؟ آیا می‌توانید مثالی بزنید که میانه عضوی از
مجموعه داده‌ها نباشد؟ در زیر ساعات مطالعه آزاد دانش‌آموزان در طول هفته آمده است.
مثال: میانه داده‌های زیر را تعیین کنید.

۲ ۶ ۹ ۳ ۴ ۱۰



برای محاسبهٔ میانه ابتدا داده‌ها را مرتب می‌کنیم.

۲ ۳ ۴ ۶ ۹ ۱۰

می‌بینیم که هیچ یک از داده‌ها ویژگی میانه را ندارد ولی اگر شما عددی بین ۴ و ۶ مثلاً $4/3$ را در نظر بگیرید، این عدد در تعریف میانه صدق می‌کند زیرا تعداد داده‌های بعد از $4/3$ برابر ۳ و تعداد داده‌های قبل از $4/3$ برابر با ۳ است. از این جا نتیجه می‌شود که شما هر عددی را بین ۴ و ۶ در نظر بگیرید در تعریف میانه صدق می‌کند. لکن برای آن که یکسان عمل نماییم در این حالت میانه را به صورت زیر به دست می‌آوریم:

$$\frac{4+6}{2} = 5$$

روش پیدا کردن میانه

۱- داده‌ها را مرتب کنیم.

۲- اگر تعداد داده‌ها فرد باشد، داده‌ای که در وسط قرار می‌گیرد برابر میانه است.

۳- اگر تعداد داده‌ها زوج باشد، نصف مجموع دو داده‌ای که در وسط قرار گرفته‌اند برابر میانه است.

دو دسته داده‌های زیر را در نظر بگیرید

۱، ۲، ۳

۰، ۲، ۱۰۰

هر دو دسته داده دارای میانه ۲ هستند ولی ملاحظه می‌کنید که تفاوت‌های اساسی بین آن‌ها وجود دارد. میانه فقط به این نکته توجه دارد که تعداد داده‌های بعد از آن برابر تعداد داده‌های قبل از آن باشد و به بزرگی و یا کوچکی داده‌ها کاری ندارد. به عبارت دیگر، میانه نسبت به اندازهٔ داده‌ها حساسیت نشان نمی‌دهد.

اگر میانه از مد کوچک‌تر باشد چگونه داده‌ها را تحلیل می‌کنید؟

اگر میانه از مد بزرگ‌تر باشد چه اطلاعاتی در مورد داده‌ها می‌دهد؟

آیا مد نسبت به اندازهٔ داده‌ها حساسیت نشان می‌دهد؟



تمرین

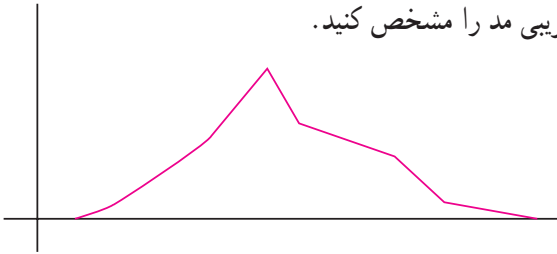


۱- میانه و مد را در داده‌های زیر حساب کنید.

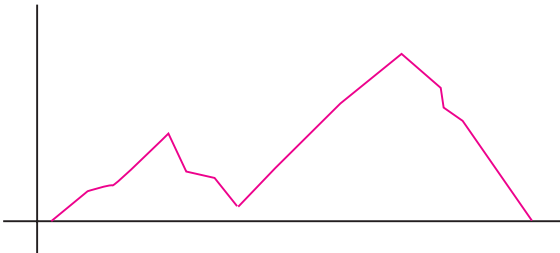
۵, ۱۵, ۴, ۰, ۱۲, ۱۵, ۵, ۶, ۵, ۰

۲- اگر داده‌ها را دو برابر کنیم چه تأثیری در مقدار میانه و مد حاصل می‌شود؟

۳- در نمودار زیر محل تقریبی مد را مشخص کنید.



۴- اگر در مطالعه قد، نمودار به شکل زیر باشد چه نتیجه‌ای می‌گیرید؟



نمودار جعبه‌ای

نمودارهایی که تاکنون شناخته‌ایم هر کدام به طریقی داده‌ها را نمایش می‌دادند. فکر می‌کنید کدام نمودار بهتر از بقیه پراکندگی داده‌ها را نشان می‌دهد؟ موارد استفاده هر یک از انواع نمودارها را بیان کنید. همان‌طور که می‌بینید این نمودارها برای مقایسه داده‌ها بسیار مفیدند ولی هیچ‌کدام به سؤالاتی از قبیل: آیا داده‌ها به هم نزدیک هستند؟ آیا داده‌ها بیش‌تر در اطراف میانگین متمرکزند یا بیش‌تر اطراف کم‌ترین داده یا بیش‌ترین داده متمرکزند؟ پاسخ نمی‌دهند.

فعالیت ۱

تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان عبارتند از:

۱۲, ۱۰, ۱۵, ۲۳, ۱۴, ۲۷, ۱۶

۳۴, ۴۱, ۴۳, ۳۲, ۱۸, ۲۵, ۳۱, ۱۹

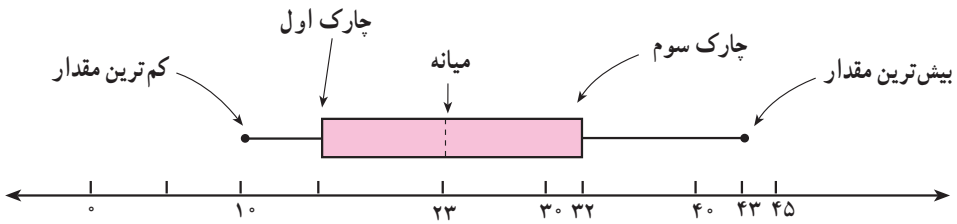
- میانه را پیدا کنید.
- اعدادی را که در نیمه قبل از میانه قرار دارند بنویسید.
- برای این اعداد، میانه را پیدا کنید.
- اعدادی را که در نیمه بعد از میانه قرار دارند بنویسید.
- برای این اعداد میانه را پیدا کنید.

میانه نیمه اول داده‌ها را چارک اول Q_1 و میانه نیمه دوم داده‌ها را چارک سوم Q_3 می‌نامند.

– چارک دوم یا Q_2 چه نام دارد؟

نمودار جعبه‌ای نموداری تصویری است که داده‌ها را براساس پنج مقدار نمایش می‌دهد. این مقادیر عبارتند از:

- ۱- کوچک‌ترین داده
- ۲- چارک اول
- ۳- میانه
- ۴- چارک سوم
- ۵- بزرگ‌ترین داده

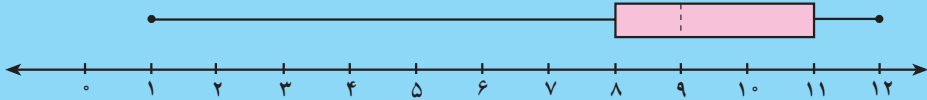


نمودار فوق نمودار جعبه‌ای داده‌های موجود در فعالیت ۱ می‌باشد.

– چند درصد از داده‌ها در داخل جعبه قرار دارد؟

– چند درصد از داده‌ها در هر یک از دنباله‌ها قرار دارند؟

فعالیت ۲



– در نمودار بالا میانه چند است؟

– چارک اول در این داده‌ها ۸ است، این عدد نشان‌دهنده چیست؟

– چارک سوم چند است؟ این عدد نشان‌دهنده چیست؟

– وجود میانه در سمت چپ جعبه نشان‌دهنده چیست؟

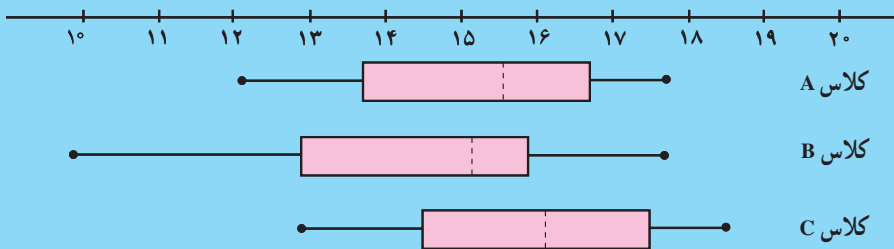
– بلندتر بودن دنباله سمت چپ نسبت به دنباله سمت راست نشان‌دهنده چیست؟



ای آفتاب آینه دار جمال تو ...

فعالیت ۳

نمودار زیر عمل کرد سه کلاس A، B و C را در امتحان ریاضی نشان می‌دهد. با توجه به نمودار به سوالات زیر پاسخ دهید.



- کدام کلاس بیشترین دامنه تغییرات را دارد؟
- میانه نمرات کدام کلاس از همه بیش‌تر است؟ میانه نمرات کدام کلاس از همه کم‌تر است؟
- پراکندگی نمرات کدام کلاس بیش‌تر از همه است؟
- این سه کلاس را با توجه به عمل‌کردشان در امتحان، از ضعیف‌ترین به قوی‌ترین مرتب کنید.

فعالیت ۴

آمار در پزشکی و بیمه کاربردهای بسیاری دارد. به طور مثال یک شرکت بیمه می‌خواهد از آماری که حاوی اطلاعاتی در مورد طول مدت بستری شدن پس از سکته قلبی در بیمارستان است استفاده کند و روش‌ها و برنامه‌های خود را برای بیمه نمودن بیماران مورد بررسی قرار دهد. داده‌ها به شرح زیر می‌باشند:

تعداد بیماران زن	تعداد بیماران مرد	طول مدت بستری شدن پس از سکته قلبی (روز)
۴	۶	۰
۲	۱۵	۱
۳	۴	۲
۴	۲	۳
۲	۵	۴
۱	۸	۵
۴	۱۶	۶
۲	۱۲	۷
۵	۹	۸
۵	۱۱	۹
۰	۱۰	۱۰
۰	۱۸	۱۱
۲	۷	۱۲
۳	۲	۱۳
۲	۵	۱۴
۰	۷	۱۵
۲	۱	۱۶
۰	۳	۱۷
۱	۱	۱۸
۰	۱	۱۹
۰	۱	۲۱
۱	۲	۲۲
۱	۰	۲۳
۱	۱	۲۴
۰	۱	۲۶
۱	۱	۲۸
۲	۰	۲۹
۰	۲	۳۷
۰	۱	۴۱
۱	۰	۴۵
۱	۰	۷۲

– نمودار جعبه‌ای را برای مدت بستری شدن زنان و مردان به طور جداگانه و برای کل بیماران رسم کنید. این نمودارها را مقایسه کنید و در کلاس به بحث بگذارید.

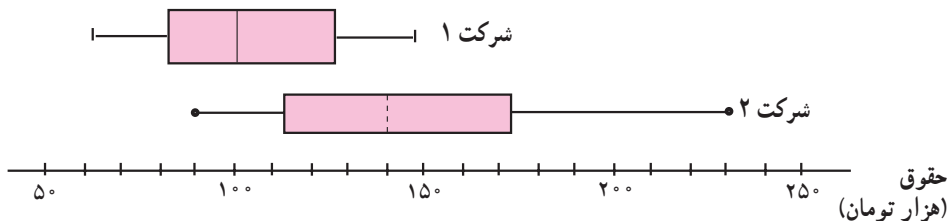
تمرین

۱- جهت تعیین مقاومت یک حیوان آزمایشگاهی در مقابل بیماری سل (باسیل تویرکلوسیس)، به ۷۳ حیوان این باسیل تزریق شد، پس از ۴۳ روز اولین حیوان مُرد. قوی‌ترین آن‌ها پس از ۵۹۸ روز مُرد. جدول زیر طول مقاومت این حیوان را در مقابل این باسیل نشان می‌دهد:

۴۳, ۴۵, ۵۳, ۵۶, ۵۶, ۵۷, ۵۸, ۶۶, ۶۷, ۷۳
 ۷۴, ۷۹, ۸۰, ۸۰, ۸۱, ۸۱, ۸۱, ۸۲, ۸۳, ۸۳
 ۸۴, ۸۸, ۹۱, ۹۱, ۹۱, ۹۲, ۹۲, ۹۷, ۹۹, ۹۹, ۱۰۰
 ۱۰۰, ۱۰۲, ۱۰۲, ۱۰۲, ۱۰۲, ۱۰۳, ۱۰۴, ۱۰۷
 ۱۰۸, ۱۰۹, ۱۱۳, ۱۱۸, ۱۱۸, ۱۲۱, ۱۲۳, ۱۲۶
 ۱۲۸, ۱۳۷, ۱۳۸, ۱۳۹, ۱۴۴, ۱۴۷, ۱۴۷, ۱۵۶
 ۱۶۲, ۱۷۴, ۱۷۸, ۱۷۹, ۱۸۴, ۱۹۱, ۱۹۸, ۲۱۴, ۲۱۴
 ۲۴۳, ۲۴۹, ۲۴۹, ۳۲۹, ۳۸۰, ۴۰۳, ۵۱۱, ۵۲۲, ۵۹۸

آیا نمودار ساقه و برگ برای توصیف این داده‌ها، نمودار مناسبی است؟ چرا؟ نمودار جعبه‌ای این داده‌ها را رسم کنید.

۲- نمودار زیر نشان‌دهندهٔ پرداخت حقوق کارمندان در دو شرکت می‌باشد.



«حقوق تقریباً ۵۰٪ کارمندان شرکت ۲ بیش‌تر از ماکزیمم حقوق کارمندان شرکت ۱ است.»

«حقوق تمام کارمندان شرکت ۲ بیش‌تر از حقوق ۵۰٪ از کارمندان شرکت ۱ است.»

— آیا ادعاهای فوق با نمودار مطابقت دارند؟

— از چنین نموداری چه استفاده‌هایی می‌توان کرد؟

دل سرپردهٔ محبت اوست ...

میانگین

شما تاکنون معدل خود را بارها حساب کرده‌اید، برای این کار نمرات را جمع کرده‌اید و بر تعداد آن‌ها تقسیم کرده‌اید. در واقع میانگین یعنی معدل داده‌ها. به‌طور صریح فرض کنید داده‌های ما به‌صورت:

$$x_1, x_2, \dots, x_n$$

باشد، در این صورت میانگین آن‌ها که با نماد \bar{x} نشان می‌دهیم عبارت است از:

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

فرض کنید میانگین نمرات ۷ درس شما ۱۵ شده باشد و میانگین نمرات همان ۷ درس برای یکی از همکلاسی‌های شما برابر ۱۴ باشد. پس مجموع نمرات شما در این هفت درس برابر $105 = 15 \times 7$ و مجموع نمرات همکلاسی شما برابر $98 = 14 \times 7$ است. یعنی اگر برای هر درسی

اسرار ازل را نه تو دانی و نه من ...

امتیازی بین ۰ تا ۲۰ (که همان نمرهٔ درس است) بدهند شما از این ۷ درس ۱۰۵ امتیاز و همکلاسی شما ۹۸ امتیاز آورده است. این دو امتیاز نشان می‌دهند که وضع تحصیلی شما از وضع تحصیلی همکلاسی شما بهتر است. اگر تعداد درس‌ها زیاد باشد، جمع امتیازها عدد بزرگی خواهد شد و مقایسه و تحلیل آن‌ها به راحتی در ذهن انجام نمی‌شود.

— اگر تعداد درس‌ها برابر نباشند، مثلاً شما ۱۱۰ امتیاز در ۸ درس و همکلاسی شما ۹۰ امتیاز در ۶ درس آورده باشد، کدام یک از شما دو نفر بهتر از دیگری عمل کرده‌اید؟

ظاهراً امتیاز شما بیش‌تر است ولی این دلیل بهتر بودن شما نیست، مسلماً اگر تعداد درس‌ها را بیش‌تر کنیم، امتیاز بالاتر خواهد رفت. برای آن‌که بتوانیم این مقایسه را انجام دهیم، باید ببینیم که اگر قرار باشد در تمام درس‌ها امتیازات برابر آورده باشید به قسمتی که جمع امتیاز ثابت باشد، آن امتیاز کدام است. مثلاً اگر برای هر یک از ۸ درس شما امتیاز a فرض کرده باشیم، جمع امتیازهای شما برابر $a + a + \dots + a = 8a$ خواهد بود که این مقدار باید برابر 110° شود. پس امتیاز مشترک درس‌های شما برابر $\frac{110^\circ}{8} = 13.75$ که همان میانگین نمرات شماست به دست می‌آید. امتیاز مشترک درس‌های همکلاسی شما برابر $\frac{90^\circ}{6} = 15$ است. این اعداد و ارقام نشان می‌دهند عموماً همکلاسی شما در درس‌های خود بهتر از شما عمل کرده است. نمرات شما بیش‌تر در اطراف 13.75 متمرکز شده‌اند و حال آن‌که نمرات همکلاسی شما در اطراف ۱۵ متمرکز شده‌اند. پس سطح نمرات همکلاسی شما بالاتر است.

جایی در توضیحات بالا گفتیم که نمرات شما در اطراف 13.75 متمرکز شده است، میانگین در مرکز داده قرار دارد، یعنی همان مقداری که داده‌ها از میانگین اضافی دارند، همان مقدار هم از میانگین کم دارند. این موضوع در قضیهٔ صفحهٔ بعد بیان شده است.

قضیه: اگر \bar{x} ، میانگین x_1, \dots, x_n باشد آن گاه مجموع اختلافات داده‌ها از میانگین برابر

است با صفر.

اثبات:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) \\ = x_1 + x_2 + \dots + x_n - \underbrace{\bar{x} - \bar{x} - \dots - \bar{x}}_{\text{تا } n} = x_1 + x_2 + \dots + x_n - \underbrace{(\bar{x} + \bar{x} + \dots + \bar{x})}_{\text{تا } n} = 0$$

زیرا می‌دانیم مجموع مقادیر برابر حاصل ضرب میانگین در تعداد آن‌ها است یعنی:

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

فعالیت

جدول زیر قد ۳۰ دانش‌آموز را برحسب سانتی‌متر نشان می‌دهد:

۱۷۷، ۱۷۵، ۱۳۷، ۱۵۵، ۱۵۰، ۱۶۰

۱۳۲، ۱۴۶، ۱۷۹، ۱۴۰، ۱۶۹، ۱۷۷

۱۴۱، ۱۴۸، ۱۳۰، ۱۷۶، ۱۳۵، ۱۳۰

۱۵۷، ۱۷۲، ۱۷۸، ۱۴۳، ۱۴۳، ۱۳۶

۱۳۲، ۱۶۶، ۱۳۰، ۱۵۱، ۱۴۵، ۱۷۸

– میانگین قد این دانش‌آموزان را حساب کنید.

– اگر به تمام این داده‌ها ۵ تا اضافه شود، چه تغییری در میانگین حاصل

می‌شود؟

– اگر از تمام این داده‌ها ۱۲ تا کم شود، چه تغییری در میانگین حاصل می‌شود؟

– با اضافه کردن a چه تغییری در میانگین حاصل می‌شود؟

– اگر تمام داده‌ها در ۲ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

– اگر تمام داده‌ها در $\frac{1}{2}$ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

– اگر تمام داده‌ها در $a > 0$ ضرب شوند چه تغییری در میانگین حاصل می‌شود؟

– اگر تمام داده‌ها در $a < 0$ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

– با توجه به نتایج به‌دست آمده در بالا، میانگین داده‌های زیر را به‌دست آورید.

الف – ۱۴۱۲، ۱۳۰۰، ۱۳۴۲، ۱۳۰۷، ۱۳۱۷، ۱۳۲۸

ب – ۸۰۳۸، ۱۰۰۳۸، ۳۰۳۸، ۴۰۳۸، ۱۰۳۸، ۲۰۳۸

در برخی از محاسبات ممکن است اعداد بسیار بزرگ باشند، جمع کردن این اعداد و تقسیم آن‌ها همراه با خطاهایی خواهد بود. شاید بگویید در عصر رایانه‌ها، بزرگی اعداد مشکلی نخواهد بود. ولی به هر حال این اعداد را باید به طریقی وارد کنیم که در این مرحله بزرگی آن‌ها، اشکالاتی را ایجاد خواهد کرد. اگر بتوانیم تدبیری بیندیشیم تا از اعداد کوچک‌تر استفاده کنیم، این مشکل را دور زده‌ایم. قضیه زیر راه حل مناسب این مسئله را ارائه می‌کند.

قضیه: اگر \bar{x} میانگین داده‌های x_1, \dots, x_n باشد آنگاه میانگین داده‌های $ax_1 + b, \dots, ax_n + b$ که با نماد $\overline{ax + b}$ نشان می‌دهیم عبارت است از:

$$\overline{ax + b} = a\bar{x} + b$$

اثبات:

$$\begin{aligned}\overline{ax + b} &= \frac{(ax_1 + b) + \dots + (ax_n + b)}{n} \\ &= \frac{a(x_1 + \dots + x_n) + (b + \dots + b)}{n} \\ &= a \frac{x_1 + \dots + x_n}{n} + \frac{nb}{n} \\ &= a\bar{x} + b\end{aligned}$$

مثال: میانگین داده‌های زیر را حساب کنید.

$$۱۷۵۰, ۱۷۶۰, ۱۷۶۶, ۱۷۵۵, ۱۷۸۰$$

حل: ملاحظه می‌کنیم که اعداد بالا به صورت زیر هستند.

$$۱۷۰۰ + ۱۰ \times ۵, ۱۷۰۰ + ۱۰ \times ۶, ۱۷۰۰ + ۱۰ \times ۶/۶, ۱۷۰۰ + ۱۰ \times ۵/۵, ۱۷۰۰ + ۱۰ \times ۸$$

پس $a = ۱۰$ و $b = ۱۷۰۰$ کافی است میانگین داده‌ها

$$۵, ۶, ۶/۶, ۵/۵, ۸$$

را حساب کنیم. میانگین این اعداد عبارت است از:

$$\frac{۵ + ۶ + ۶/۶ + ۵/۵ + ۸}{۵} = \frac{۳۱/۱}{۵} = ۶/۲۲$$

پس میانگین داده‌های اصلی عبارت است از:

$$۱۷۰۰ + ۱۰ \times ۶/۲۲ = ۱۷۰۰ + ۶۲/۲ = ۱۷۶۲/۲$$

مقایسه میانگین و میانه: میانگین شاخص خوبی برای نشان دادن مرکزیت داده‌هاست. در بعضی از مسائل این شاخص نمی‌تواند مؤثر باشد. مثلاً انتظار می‌رود در یک کلاس تمامی دانش‌آموزان نمره خوبی بگیرند. اگر نمرات خوب باشد مسلماً میانگین هم بالا خواهد بود ولی آیا بالا بودن میانگین

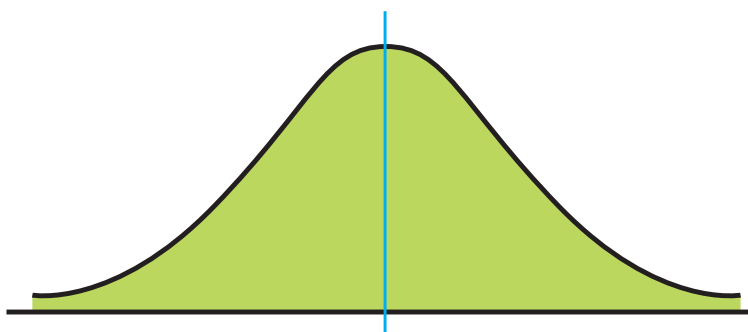
نمرات نشان دهنده وضع خوب کلاس است. ممکن است تعداد کمی نمره 20° گرفته باشند و تعداد زیادی نمره 10° یا 12° . وضع این کلاس رضایت بخش نخواهد بود برای آن که وضع کلاس را بتوانیم خوب ارزیابی کنیم حداقل نصف کلاس باید نمره خوبی آورده باشند.

آن چه نمره‌ای است که نمره نصف دانش آموزان کلاس از آن بیش تر است؟

– اگر میانه خیلی از میانگین کوچک تر باشد تعبیر آن چیست؟

– اگر میانه خیلی بزرگ تر از میانگین باشد تعبیر آن چیست؟

در فصل نمودارها به منحنی نرمال اشاره کردیم و گفتیم که این یک منحنی متقارن زنگی شکل به صورت زیر است.



از تقارن آن نتیجه می شود که میانه و میانگین آن برابرند و چون نقطه ماکزیمم هم دارند مد آن نیز برابر میانگین است.



میانگین وزن دار

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 باشد:

آزمون اول	آزمون دوم	آزمون سوم	آزمون چهارم
۱۶	۱۵	۱۸	۱۷

نمره‌ای که برای این درس برای شما در کارنامه خواهد آمد میانگین ۶ عدد زیر است:

۱۶	۱۵	۱۸	۱۷	۱۷	۱۷
----	----	----	----	----	----

یعنی نمره درس مذکور عبارت است از:

$$\frac{۱۶+۱۵+۱۸+۱۷+۱۷+۱۷}{۶} = \frac{۱۰۰}{۶} \approx ۱۶/۶۶$$

ابتدا ببینیم این ضریب چه اثری در نمره شما می‌گذارد. نمره بدون احتساب ضریب عبارت

است از:

$$\frac{۱۶+۱۵+۱۸+۱۷}{۴} = ۱۶/۵$$

اگر ضریب را بیش‌تر کنیم مثلاً نمره آخر را با ضریب ۵ در نظر بگیریم خواهیم داشت:

$$\frac{۱۶+۱۵+۱۸+۱۷ \times ۵}{۸} = ۱۶/۷۵$$

ملاحظه می‌شود که هر چه قدر ضریب نمره آزمون چهارم بالاتر رود نمره درس شما بیش‌تر

خواهد شد. حال ببینیم آیا همیشه این‌طور است. فرض کنید نمرات شما به صورت زیر باشد:

۱۶	۱۵	۱۸	۱۶
----	----	----	----

در این صورت نمره درس شما بدون احتساب ضریب عبارت است از:

$$\frac{۱۶+۱۵+۱۸+۱۶}{۴} = ۱۶/۲۵$$

حال اگر به نمره آزمون چهارم ضریب ۳ بدهیم، خواهیم داشت:

$$\frac{۱۶+۱۵+۱۸+۱۶+۱۶+۱۶}{۶} \approx ۱۶/۱۶$$

اگر ضریب آزمون چهارم ۵ باشد نمره درس شما عبارت است از:

$$\frac{۱۶+۱۵+۱۸+۵ \times ۱۶}{۸} = ۱۶/۱۲۵$$

این بار ملاحظه می‌کنید که هر چه قدر ضریب بالاتر می‌رود نمره شما کم‌تر می‌شود. می‌توانید

علت آن را توضیح دهید؟

در ازل پرتو حسنت ز تجلی دم زد ...

چگونه است که در مثال قبل با افزایش ضریب، نمره بالاتر می‌رفت ولی در این مثال با افزایش ضریب نمره کاهش می‌یابد؟

شاید دقت کرده باشید که در مثال اول نمرهٔ آزمون چهارم بیش از میانگین نمرات (بدون ضریب) شماسست. میانگین نمرات بدون ضریب برابر $16/5$ بود ولی نمرهٔ آزمون چهارم شما 17 بود. اما در مثال دوم میانگین نمرات بدون ضریب برابر $16/25$ و نمرهٔ آزمون چهارم شما 16 بود. هر چه قدر این نمره از میانگین فاصلهٔ بیش‌تری داشته باشد تأثیر ضریب آن محسوس‌تر است. این مطالب را در جدول زیر به وضوح می‌بینید.

نمرات	میانگین بدون ضریب	نمره چهارم با ضریب ۲	نمره چهارم با ضریب ۴
۱۵،۱۶،۱۴،۱۷	۱۵/۵	۱۵/۸	۱۶/۱۴
۱۵،۱۶،۱۴،۱۹	۱۶	۱۶/۶	۱۷/۲۸
۱۵،۱۶،۱۴،۱۲	۱۴/۲۵	۱۳/۸	۱۳/۲۸

ملاحظه می‌کنید اگر نمره زیر میانگین باشد و ضریب آن بالاتر رود نمره درس کاهش می‌یابد برعکس اگر نمره بالاتر از میانگین باشد و ضریب آن بالاتر رود نمره درس افزایش می‌یابد. معمولاً

نمراتی که اهمیت بیش‌تری دارند باید بتوانند جایگاه خود را در بین نمرات نشان دهند. سه نمره اول چون از بخش‌های کوچک کتاب امتحان گرفته شده است، انتظار می‌رود نمرات خوبی باشند ولی نمره آزمون چهارم چون از تمام کتاب بوده و از سؤال‌های ترکیبی (سؤالاتی که چند مفهوم را در هم ادغام کرده باشند) استفاده شده است، اهمیت آن بیش‌تر است این اهمیت از طریق ضریب اعمال می‌شود. حال به حالت کلی برمی‌گردیم، فرض کنیم برحسب ضروریاتی به هر آزمونی ضریب اختصاص داده شده باشد، مثلاً فرض کنید در چهار آزمون برگزار شده به ترتیب ضریب‌های ۱، ۲، ۳، ۳ اختصاص داده شده باشد و نمرات شما به ترتیب از آزمون اول تا چهارم برابر ۱۴، ۱۷، ۱۶، ۱۵ باشد، نمره این درس شما چه قدر خواهد بود؟ در واقع می‌خواهیم میانگین مقادیر زیر را حساب کنیم.

$$۱۴, ۱۷, ۱۷, ۱۶, ۱۶, ۱۶, ۱۵, ۱۵, ۱۵$$

پس نمره درس عبارت است از :

$$\frac{۱۴+(۱۷+۱۷)+(۱۶+۱۶+۱۶)+(۱۵+۱۵+۱۵)}{۱+۲+۳+۳}$$

$$= \frac{۱۴۱}{۹} \approx ۱۵/۶۶$$

این قبیل میانگین‌ها را میانگین‌های وزن‌دار (وزنی) می‌گویند.

در حالت کلی فرض کنید داده‌های x_1, \dots, x_n به ترتیب دارای ضریب‌های w_1, \dots, w_n باشد. این اطلاعات را برای سادگی می‌توانیم در جدول زیر خلاصه کنیم :

داده‌ها	x_1	x_2	\dots	x_n
ضریب (وزن)	w_1	w_2	\dots	w_n

در این صورت میانگین داده‌های بالا با احتساب ضرایب مربوطه به صورت زیر محاسبه می‌شود :

$$\text{میانگین وزنی} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n}$$

مثال: میانگین وزن‌دار را در جدول زیر حساب کنید.

داده	۱۲	۱۴/۵	۱۵	۹
وزن	۱	۳	۱	۳

حل:

$$\text{میانگین وزنی} = \frac{۱ \times ۱۲ + ۳ \times ۱۴/۵ + ۱ \times ۱۵ + ۳ \times ۹}{۱ + ۳ + ۱ + ۳}$$

$$= \frac{۹۷/۵}{۸} \approx ۱۲/۱۸$$

میانگین وزن‌دار را هم با همان نماد \bar{x} نشان می‌دهند و اگر بخواهند تمایز بین میانگین معمولی و میانگین وزن‌دار قائل شوند، میانگین وزن‌دار را با نماد \bar{x}_w نشان می‌دهند.

در مواردی که ما با میانگین وزن دار مواجه می‌شویم محاسبه میانگین در یک جدول فراوانی است. دیدیم که در یک جدول فراوانی قرار شد تمام افرادی که در یک دسته قرار دارند برابر مرکز دسته فرض کنیم. پس مرکز دسته به اندازه فراوانی آن دسته تکرار می‌شود، و این مانند آن است که ما به مرکز دسته وزنی برابر فراوانی آن دسته داده باشیم، پس اگر بخواهیم میانگین را حساب کنیم باید مرکز دسته‌ها را در فراوانی‌های نظیر، ضرب و با هم جمع کرده و پس از آن بر مجموع فراوانی‌ها (که همان تعداد داده‌هاست) تقسیم کنیم.

$$\bar{x} = \frac{f_1x_1 + \dots + f_kx_k}{f_1 + \dots + f_k}$$

میانگین در جدول فراوانی

این مفهوم میانگین برای محاسبات از روی جدول فراوانی مناسب‌تر است. داده‌های زیر را که ضریب هوشی ۳۴ دانش‌آموز پسر در سال اول دبیرستان را نشان می‌دهد در نظر بگیرید.

۸۸	۱۰۸	۱۰۸	۱۱۱	۱۱۱	۱۰۳	۱۰۸	۱۰۱	۸۸
۹۱	۱۰۳	۱۰۵	۱۰۵	۱۱۲	۹۶	۱۱۱	۱۰۷	۹۶
۹۲	۱۰۰	۱۱۲	۱۱۲	۹۲	۸۹	۹۵	۹۰	۱۰۵
۱۱۱	۹۵	۱۰۸	۷۷	۱۰۶	۹۰	۷۷		

ابتدا مستقیماً میانگین این داده‌ها را حساب کنید. سپس با تشکیل یک جدول فراوانی مناسب میانگین را به دست آورده و جواب‌ها را مقایسه کنید.
حل: میانگین داده‌های بالا عبارت است از:

$$\frac{\text{مجموع داده‌ها}}{۳۴} = \frac{۳۴۰۳}{۳۴} \approx ۱۰۰/۹$$

حال داده‌ها را در دسته‌هایی، دسته‌بندی می‌کنیم. ابتدا ملاحظه می‌کنیم که

$$a = ۷۷, b = ۱۱۲$$

پس $R = b - a = ۳۵$. این ۳۵ داده را در ۵ دسته، دسته‌بندی می‌کنیم.

$$C = \frac{۳۵}{۵} = ۷$$

طول دسته‌ها

	دسته‌ها	مرکز دسته‌ها	فراوانی‌ها
	۷۷-۸۴	۸۰/۵	۲
	۸۴-۹۱	۸۷/۵	۵
	۹۱-۹۸	۹۴/۵	۷
	۹۸-۱۰۵	۱۰۱/۵	۴
۱۳۳	۱۰۵-۱۱۲	۱۰۸/۵	۱۶
	جمع		۳۴

$$\bar{x}_w = \frac{2 \times 80/5 + 5 \times 87/5 + 7 \times 94/5 + 4 \times 101/5 + 16 \times 107/5}{34}$$

$$= \frac{3386}{34} \approx 99/58$$

ملاحظه می کنید آن چه که از روی جدول فراوانی به دست آمد با آن چه که مستقیماً از روی داده ها به دست آمد به اندازه $0/51 - 99/58 = 100/09 - 99/58$ اختلاف دارد. این اختلاف در مقایسه با میانگین داده ها که حدوداً برابر 100 است ناچیز است. قبلاً هم گفتیم در جدول بندی و یکسان سازی داده ها مقداری از اطلاعات از دست می رود ولی سهولت ها و تعبیرهایی که این روش برای ما ایجاد خواهد کرد به مراتب با ارزش تر از حفظ مقدار ناچیزی از اطلاعات است.

توجه ۱: اگر بخواهیم برابری $(x_1 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$ را در جدول فراوانی بررسی کنیم باید فراوانی ها را نیز در نظر بگیریم. یعنی در جدول فراوانی داریم:

$$f_1(x_1 - \bar{x}) + f_2(x_2 - \bar{x}) + \dots + f_n(x_n - \bar{x}) = 0$$

در مثال بالا داریم:

$$2(80/5 - 100/1) + 5(87/5 - 100/1) +$$

$$7(94/5 - 100/1) + 4(101/5 - 100/1) +$$

$$16(108/5 - 100/1)$$

$$= 39/2 - 63 - 39/2 + 5/6 + 134/4 = 1/4$$

البته این مقدار صفر نیست ولی به صفر بسیار نزدیک است آیا می توانید توضیح دهید که چرا

دقیقاً برابر صفر به دست نیامد؟

روشی سریع برای محاسبه میانگین: سعید نمرات زیر را در پایان نیمسال اول اتخاذ کرده است. او برای محاسبه سریع معدل نمرات خود، ابتدا معدل خود را حدس می زند مثلاً فرض می کند معدل او ۱۸ است. سپس از نمره اول شروع می کند و با معدل حدسی خود مقایسه می کند.

لیست نمرات	۱۵	۱۷	۱۵	۱۹	۲۰	۱۹	۱۳	۱۶	۱۸	۱۹
تفاضل از معدل تخمینی ۱۸	-۳	-۱	-۳	+۱	+۲	+۱	-۵	-۲	۰	+۱
جمع مقادیر مقایسه شده	-۳	-۳+(-۱) -۴	-۴+(-۳) -۷	-۷+۱= -۶	-۴	-۳	-۸	-۱۰	-۱۰	-۹

نمره اول ۱۵ است که نسبت به معدل تخمینی ۳،۱۸ نمره کم دارد، این تفاوت را با ۳- نشان می‌دهیم. حال نمره دوم را در نظر می‌گیریم. نسبت به معدل ۱۸ نمره دوم ۱ نمره کم است. او تاکنون ۴ نمره از معدل حدسی خود کم آورده است. همین روش را ادامه می‌دهد و در پایان او ۹ نمره از معدل کم آورده است. ۹ را بر تعداد نمرات تقسیم می‌کند و مقدار آن را از میانگین حدسی کم می‌کند، تا میانگین واقعی به دست آید.

$$\begin{aligned}\bar{x} &= 18 - \frac{9}{10} = 18 - 0.9 \\ &= 18 - 0.9 = 17.1\end{aligned}$$

در حالت کلی فرض کنید داده‌ها باشند و y میانگین تخمینی باشد.

$$\begin{aligned}A &= \frac{(x_1 - y) + \dots + (x_n - y)}{n} \\ &= \frac{x_1 + \dots + x_n - ny}{n} \\ &= \bar{x} - y\end{aligned}$$

پس:

$$\bar{x} = y + A$$

در مثال مقدار A را برابر $\frac{-9}{10}$ حساب کردیم و از رابطه $A = \bar{x} - y$ با داشتن A و y میانگین

واقعی \bar{x} را به دست آوردیم

فردی را شکنجه می‌دادند. شکنجه این فرد به این صورت بود که یکی از دست‌های او را در کوره‌ای با دمای 100° درجه سانتی‌گراد بالای صفر قرار دادند و دست دیگر او را در داخل مبردی با سرمای 100° درجه سانتی‌گراد قرار دادند. در این حال از او پرسیدند، در چه حالی هستی؟ او در پاسخ گفت: به طور متوسط احساس راحتی می‌کنم.

تمرین

۱- بهزاد در آزمایشگاه فیزیک، پس از انجام یک آزمایش به داده‌های زیر دست پیدا کرد:

۲۰۲۰، ۲۰۹۶، ۲۲۰۲، ۲۰۱۶، ۲۰۷۳، ۲۰۳۹

– میانگین و میانه و مد را برای داده‌های به دست آمده به دست آورید.

– پس از بررسی مجدد داده‌ها، بهزاد متوجه شد که به جای ۲۲۰۲ باید ۲۰۲۲ ثبت می‌شد.

او این تغییر را در لیست خود اعمال نمود.

– میانگین، میانه و مد را برای لیست جدید داده‌ها به دست آورید.

– کدام یک از شاخص‌های فوق با این تغییر، بیش‌تر از همه تغییر یافت؟

– عموماً کدام شاخص در اثر تغییرات، بیش‌تر از همه تغییرپذیر است؟ توضیح دهید.

۲- در یک امتحان ریاضی ۱۵ نفر نمرات زیر را کسب نمودند:

۰، ۱، ۳، ۱۴، ۱۴، ۱۵، ۱۶، ۱۶، ۱۷، ۱۷، ۱۸، ۱۸، ۱۸، ۱۹، ۲۰

– میانگین، میانه و مد را به دست آورید.

– کدام شاخص نمایان‌گر بهتری از وضع دانش‌آموزان این کلاس می‌باشد؟

۳- میزان پرداخت حقوق در یک شرکت خصوصی در جدول زیر آمده است:

حقوق هر نفر (تومان)	بسمت
۳۰۰,۰۰۰	مدیر شرکت
۲۰۰,۰۰۰	معاون
۸۵,۰۰۰	حسابدار
۶۰,۰۰۰	بازاریاب
۴۰,۰۰۰	منشی
۶۰,۰۰۰	۲ نفر فروشنده
۳۵,۰۰۰	مستخدم

مدیر شرکت در آگهی دعوت به همکاری در روزنامه اعلام می‌دارد که میانگین پرداخت حقوق

در این شرکت بیش‌تر از ۱۰۰,۰۰۰ تومان است. آیا این رقم می‌تواند دلیلی برای پرداخت حقوق بالا

در این شرکت باشد؟ چرا؟ کدام یک از شاخص‌ها برای نشان دادن وضع پرداختی کارکنان این شرکت

بهتر می‌باشد چرا؟

۴- میانگین ۱۰ عدد ۳, ۸, ۵, ۹, ۳, ۷, ۸, a, ۳, ۴ برابر ۵/۵ است.

الف - a را پیدا کنید.

ب - میانه را به دست آورید.

۵- مجموع هفت عدد متوالی برابر ۱۴۷ است. اگر میانگین این اعداد از میانه کم شود چه

اتفاقی می افتد؟ (بدون محاسبه)

۶- نمرات درس فیزیک دانش آموزی در طول سال برابر است با:

۱۷, ۱۹, ۱۸, ۷, ۱۸/۵, ۱۹

الف - میانه و میانگین، مد را برای نمرات این دانش آموز حساب کنید.

ب - کدام یک از شاخص های فوق، بیان گر بهتری از وضع این دانش آموز در درس فیزیک است؟

ج - اگر معلم درس فیزیک این دانش آموز، برای جبران نمره ۷، امکان امتحان مجدد را به او

بدهد، برای این که میانگین وی در این درس ۱۸/۵ شود، او در این امتحان چه نمره ای باید کسب

کند؟

۷- اگر در مجموعه اعداد $\{7\%, 11\%, 8\%, 6\%\}$ میانگین برابر میانه و مد باشد، x را پیدا کنید.

۸- یک کارخانه تولید مواد غذایی می خواهد محصولی را بسته بندی کند. در یک تحقیق،

مسئولین متوجه شدند که از ۲۲ نفر، پنج نفر ترجیح می دهند محصول در بسته بندی یکی ای ارائه

شود، پنج نفر ترجیح دادند محصول در بسته بندی دوتایی و سه نفر در بسته بندی سه تایی و شش نفر در

بسته بندی چهارتایی و سه نفر در بسته بندی شش تایی ارائه شود.

- میانگین، مد، میانه را برای تعداد محصول در هر بسته به دست آورید.

- به نظر شما این کارخانه باید از کدام شاخص در تصمیم گیری خود استفاده کند؟ چرا؟

۹- در بسیاری از دانشگاه ها، نمرات بر حسب A, B, C, D, F اعلام می شوند. در این سیستم

نمره گذاری A دارای وزن یا ارزش ۴، B دارای وزن یا ارزش ۳ و C دارای وزن یا ارزش ۲ و D

دارای وزن یا ارزش ۱ و F وزن یا ارزش صفر دارد.

جدول زیر دروس و تعداد واحدها و نمرات اکتساب شده توسط مریم را در دانشگاه نشان

می دهد :

– معدل مریم را در این ترم حساب کنید.

نمره	تعداد واحد	درس
B	۳	ریاضیات
A	۳	فیزیک
C	۲	ورزش
B	۲	تاریخ
C	۳	ادبیات
A	۱	آزمایشگاه

۱۰– یک باغبان می خواهد تعداد علف های هرز یک زمین بازی را تخمین بزند. او به طور

تصادفی ۱۰۰ ناحیه از این زمین را هر یک به مساحت 100 cm^2 انتخاب نمود و تعداد علف های هرز

هر ناحیه را شمرد. جدول زیر نتیجه این شمارش است.

تعداد علف های هرز	فراوانی نواحی
۰–۴	۱۸
۵–۹	۲۵
۱۰–۱۴	۳۲
۱۵–۱۹	۱۴
۲۰–۲۴	۷
۲۵–۲۹	۴

– فراوانی تجمعی نواحی را به دست آورید.

– نمودار میله ای را برای فراوانی تجمعی نواحی رسم کنید.

– میانگین تعداد علف های هرز را برای هر قسمت (هر ۱۰۰ سانتی متر مربع) تخمین بزنید.

– اگر مساحت زمین 8000 m^2 باشد، تعداد کل علف های هرز این زمین را تخمین بزنید.

۱۱- سن بازیکنان تیم ملی فوتبال یک کشور به شرح زیر است :

۲۷، ۲۴، ۲۶، ۲۶، ۲۹، ۱۹، ۳۱، ۱۸، ۲۳، ۲۲

۲۵، ۲۶، ۲۷، ۲۳، ۲۹، ۲۵، ۲۵، ۳۳

۳۱، ۲۱، ۲۶، ۲۵

– میانگین، مد و میانه سن بازیکنان این تیم را به دست آورید.

– تعداد بازیکنانی که سن آن‌ها بالاتر از میانگین است، بیش تر است یا تعداد بازیکنانی که سن

آن‌ها از میانگین کم تر است؟

– تعداد بازیکنانی که سن آن‌ها بالاتر از میانه است بیش تر است یا تعداد بازیکنانی که سن

آن‌ها از میانه کم تر است؟

– میانه و میانگین را در این بررسی مقایسه کنید. چه نتیجه‌ای می‌گیرید؟

۱۲- یک کارخانه سازنده اتومبیل‌های شخصی در نظر دارد اتومبیل‌های جدیدی را طراحی

کند. برای این منظور اتومبیل‌هایی را که از یک چهارراه عبور می‌کردند، مورد بررسی قرار داد.

اطلاعات به دست آمده به شرح زیر است :

۱۲ اتومبیل با ۱ سر نشین

۶ اتومبیل با ۲ سر نشین

۴ اتومبیل با ۳ سر نشین

۲ اتومبیل با ۴ سر نشین

۱ اتومبیل با ۶ سر نشین

– به طور متوسط هر اتومبیل چند سر نشین دارد؟

در این بررسی مد چند است؟

۱۳- داده‌های زیر مصرف برق ۵۰ خانوار ساکن تهران را نشان می‌دهد. (برحسب کیلووات ساعت)

۲۵، ۵، ۱۵، ۱۰، ۱۲، ۲۵، ۱۷، ۱۹، ۱۵، ۳۰

۲۵، ۳۵، ۴۷، ۵۵، ۶۶، ۶۸، ۹۸، ۸۸، ۶۸، ۷۵

۹۵، ۸۵، ۱۰۱، ۱۱۰، ۱۲۰، ۹۸، ۷۸، ۷۷، ۷۵، ۱۱۵

۶۶، ۴۵، ۶۸، ۷۵، ۹۳، ۸۸، ۱۲۰، ۱۴۰، ۱۳۵، ۱۵۰

۱۶۰، ۱۸۰، ۱۷۵، ۱۶۵، ۱۹۵، ۲۰۰، ۱۷۷، ۱۸۰، ۱۹۰، ۱۳۵

الف - آیا نمودار ساقه و برگ برای این داده‌ها مناسب است؟
 ب - آیا میانگین معیار خوبی برای نشان دادن مصرف خانوارهای تهرانی هست یا خیر؟ توضیح دهید.

ج - نمودار جعبه‌ای این داده‌ها را رسم کنید.
 د - تراکم مصرف بیش‌تر روی چه میزان از مصرف متمرکز شده است؟

۱۴- مدیر یک فروشگاه در نظر دارد برای حسابداران پای صندوق فروشگاهش برنامه‌ریزی کند. او می‌خواهد بداند در چه ساعتی چند حسابدار برای رسیدگی به مشتریانش لازم دارد. برای این منظور او باید بداند به‌طور متوسط طول صف در ساعات مختلف روز چه قدر است. فرض کنید در ساعت ۱۰ صبح به‌طور متوسط هر ۱۰ دقیقه ۶ مشتری به صندوق مراجعه می‌کنند و فرض کنید رسیدگی به هر یک از آن‌ها به‌طور متوسط ۲ دقیقه طول می‌کشد. مثلاً اگر ترتیب رسیدن مشتری‌های A, B, C, D, E, F مطابق جدول زیر باشد:

زمان	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
مشتری	A	B	C, D			E				F

برای پیدا کردن میانگین طول صف در هر دقیقه باید نمودار طول صف بر حسب نفر در هر دقیقه را رسم کنیم.



اما این شش نفر ممکن است با چندین الگوی زمانی متفاوت به صندوق‌ها برسند و رسم جدول بالا برای تمام الگوهای ممکن کاری بسیار دشوار و وقت‌گیر است. با توجه به الگوی بالا شما چه پیشنهادی برای کمک به مدیر فروشگاه می‌توانید بکنید؟

از ماشین حساب چگونه استفاده کنیم

پیدا کردن میانگین با استفاده از ماشین حساب: در یک بررسی از رانندگان در مورد تعداد تصادفاتی که در سال گذشته داشتند سؤال شد. اطلاعات به دست آمده در جدول زیر آمده است:

تعداد تصادفات	۰	۱	۲	۳	۴	۵	۶
تعداد رانندگان	۱۷	۱۳	۲۱	۴	۲	۲	۱

– چند راننده در این بررسی شرکت داشتند؟

◦

$$۱۷ \text{ } ۱۳ = \text{$$

$$۲۱ \text{$$

$$۴ \text{$$

$$۲ \text{$$

$$۲ \text{$$

$$۱ \text{$$

میانگین تعداد تصادفات برای هر راننده را به دست آورید:

◦

$$◦ \text{ } ۱۷ = \text{$$

$$۱ \text{ } ۱۳ = \text{$$

$$۲ \text{ } ۲۱ = \text{$$

$$۳ \text{ } ۴ = \text{$$

$$۵ \text{ } ۲ = \text{$$

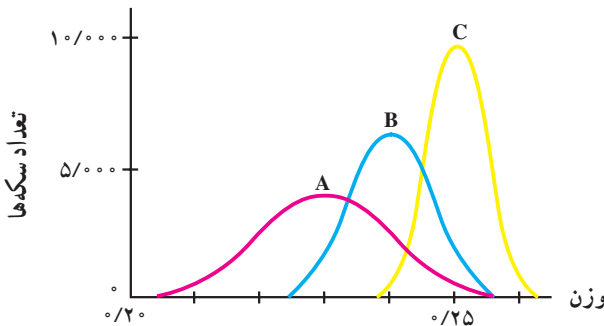
$$۶ \text{ } ۱ = \text{ } ۶۰ \approx ۱/۳۸$$

از فرشتگانی که در آسمانهایت جایشان دادی، و از زمین خود برترشان بردی. آنان از دیگر آفریدگانت تو را بهتر شناسند، و از عقاب تو بیشتر می‌هراسند، و به تو نزدیکترند و نه پراکنده‌گردش زمان. آنان با مرتبتی که از آن برخوردارند، و از منزلتی که نزد تو دارند، و یکدله تو را دوست دارند و تو را فراوان طاعت می‌گذارند، و اندک غفلتی در فرمان تو نیارند.



شاخص‌های پراکندگی

ضرابخانه سکه‌های دولتی تصمیم گرفته است که وزن پنج تومانی‌هایی که در جریان بازار قرار می‌دهد، روی نمودار بیاورد. برای این منظور، چندین سکه پنج تومانی که تازه ضرب شده‌اند به طور تصادفی انتخاب و به دقت وزن می‌شوند. همچنین به همان تعداد سکه پنج تومانی که پنج سال پیش به بازار عرضه شده بود و به همان تعداد سکه پنج تومانی که ۱۰ سال پیش به بازار عرضه شده بود به دقت وزن می‌شوند. سه نمودار A و B و C به دست می‌آید.



- فکر می‌کنید کدام نمودار مربوط به سکه‌های نو ارائه شده به بازار و کدام نمودار مربوط به سکه‌های پنج ساله و ده ساله است؟
- با گذشت زمان میانگین وزن سکه‌های پنج تومانی چه تغییری می‌کند؟
- با گذشت زمان مد وزن سکه‌های پنج تومانی چه تغییری می‌کند؟
- چاق و یا لاغر شدن نمودارها را چگونه تعبیر می‌کنید؟



پراکندگی

در فصل قبل با شاخص‌های مرکزی آشنا شدیم. مثلاً شما بر اساس میانگین زمان‌هایی که طول کشیده است تا در ۲۰ روز گذشته از خانه به مدرسه بروید، از منزل خارج شده‌اید. در بین راه اعدادی را که معرف این مدت زمان بوده‌اند در ذهن خود مرور می‌کنید. میانگین اعداد ۳۵ دقیقه شده است و شما درست ۳۵ دقیقه قبل از شروع کلاس‌ها از خانه خارج شده‌اید، ولی اعدادی مانند ۴۰، ۳۸، ۳۰، ۳۲ و یا حتی ۴۵ در بین اعداد دیده می‌شود. شما مایل نیستید که دیر به مدرسه برسید، به ویژه آن‌که روز امتحان هم باشد. یادآوری اعداد بالا آسودگی خیال را به هم می‌ریزد. از خود می‌پرسید نکند امروز یکی از آن روزهایی باشد که بیش از ۳۵ دقیقه طول بکشد. اگر کمی بیش‌تر طول بکشد، (در حد ۵ دقیقه) اشکال عمده‌ای نخواهد بود ولی تأخیر یک ربع و یا بیش‌تر مشکل‌ساز است. هرچه قدر در بین داده‌های شما، اعداد دورتر از میانگین وجود داشته باشد، ناراحتی شما بیش‌تر خواهد بود. اگر قبلاً این اعداد را بهتر و آروم‌تری می‌کردید و صرفاً میانگین را مدنظر قرار نمی‌دادید الان وضع بهتری داشتید، اگر از ابتدا درمی‌یافتید که اعداد تفاوت دارند می‌توانستید در این روز سرنوشت‌ساز تصمیم بهتری بگیرید. معیاری که می‌تواند این تفاوت‌ها و میزان آن‌ها و به‌خصوص دوری آن‌ها از میانگین را برای ما اندازه‌گیری کند شاخص پراکندگی نامیده می‌شود. ما به چند شاخص پراکندگی در زیر اشاره می‌کنیم:

دامنه تغییرات

دامنه تغییرات که در فصل دوم با آن آشنا شدید یکی از شاخص‌های پراکندگی است. هرچه قدر داده‌ها از هم دورتر باشند، دامنه تغییرات آن‌ها بیش‌تر خواهد بود. دیدیم اگر دامنه تغییرات برابر صفر باشد تمام داده‌ها با هم برابرند. یعنی هیچ تفاوتی بین آن‌ها موجود نیست. حال به دو دسته داده زیر که نمره‌های دانش‌آموزان در درس فیزیک در دو نمونه ۱۰ عضوی را به نمایش می‌گذارند توجه کنید.

الف) ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹

ب) ۱۰، ۱۵، ۱۴، ۱۴، ۱۵، ۱۵، ۱۵، ۱۵، ۱۴، ۱۹

دامنه تغییرات هر دو دسته یک اندازه است:

$$R = 19 - 10 = 9$$

ولی پراکندگی داده‌های دسته دوم کم‌تر است. در داده‌های دسته دوم ما عملاً اعداد ۱۴ و ۱۵ را داریم. به ۱۰ و ۱۹ بیش‌تر باید با شک و تردید نگریست. این دو مقدار ممکن است به خاطر اندازه‌گیری‌های نادرست حاصل شده باشند و یا ممکن است علل دیگری برای وجود آن‌ها در بین داده‌ها باشد. اما در داده‌های دسته اول این دو مقدار کاملاً طبیعی به نظر می‌رسند، زیرا مقادیر نزدیک به آن‌ها نیز داریم. با این توضیحات ما در داده‌های دسته دوم بیش‌تر روی ۱۴ و ۱۵ حساب می‌کنیم و این دو مقدار هم به هم خیلی نزدیک‌اند. این توضیح می‌رساند که دامنه تغییرات در بعضی مواقع به خوبی نمی‌تواند پراکندگی موجود در داده‌ها را نشان دهد. بنابراین به دنبال معیاری هستیم که بهتر از دامنه تغییرات بتواند پراکندگی و تفاوت‌ها را نشان دهد. قبل از آن که به معرفی معیار بعدی برویم ذکر یک نکته ضروری است که وقتی با دید مجرد ریاضی به دو دسته عدد بالا نگاه می‌کنید تفاوت هر دو را در حد ۹ می‌بینید، ولی با دید آماری مسئله کمی فرق می‌کند. در آمار حضور دو مقدار ۱۰ و ۱۹ را در دسته «ب» زیاد جدی نمی‌گیریم چون این‌ها مقادیر تک افتاده‌ای هستند. به نظر می‌رسد این دو مقدار به جامعه تعلق نداشته باشند. ما در این درس سعی می‌کنیم علاوه بر نقطه نظراتی که شما درباره مسائل اطراف خود دارید پایه‌های دید آماری را نیز ایجاد کنیم و آن را توسعه بخشیم.



چارک‌های بالا و پایین

گفتیم که دامنه تغییرات ممکن است در بعضی مواقع تعبیرهای نامناسب از جامعه ارائه کند، مثلاً در جامعه‌ای که عملاً داده‌ها به هم نزدیک‌اند، به علت وجود دو مقدار خیلی کوچک و خیلی بزرگ در جامعه، دامنه تغییرات عدد بزرگی به دست آید، و حال آن که جامعه از دید آماری جامعه متمرکزی باشد. تعداد بازدیدکنندگان از کتابخانه مدرسه در ۱۲ روز کاری در زیر آمده است.

۰, ۱, ۲, ۸, ۷, ۶, ۵, ۹, ۱۰, ۶, ۱۵, ۱۱

دامنه تغییرات این داده‌ها برابر ۱۵ است و به نظر زیاد می‌رسد ولی داده‌ها عموماً از ۵ تا ۱۱ پراکنده شده‌اند که پراکندگی آن‌ها زیاد نیست. از این رو برای حذف تأثیر داده‌های بزرگ و کوچک روی تعدادی از اعداد بالای داده‌ها و تعدادی از اعداد پایین داده‌ها را حذف می‌کنند. این که چه تعدادی حذف شود بستگی به تعداد داده‌ها و نزدیکی داده‌ها به هم دارد. در بعضی مواقع یک دهم داده‌ها از بالا و یک دهم داده‌ها از پایین را حذف می‌کنند و در برخی مواقع یک چهارم داده‌های بالایی و یک چهارم داده‌های پایینی را حذف می‌کنند و بعد از آن دامنه تغییرات داده‌های باقی‌مانده را محاسبه می‌کنند. به عنوان مثال در داده‌های بالا اگر یک چهارم داده‌ها را از بالا و پایین حذف کنیم به داده‌های زیر خواهیم رسید :

۰, ۱, ۲, ۵, ۶, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۵

بنابراین دامنه تغییرات این داده‌ها عبارت خواهد بود از $15 - 0 = 15$ است. اگر توجه کرده باشید در برخی از مسابقات امتیازهایی را که داورها می‌دهند، بیش‌ترین و کم‌ترین امتیاز را حذف می‌کنند و بر اساس امتیازهای باقی‌مانده امتیاز ورزشکار را تعیین می‌کنند.



تمرین

۱- دو کارخانه تولید کننده مواد غذایی A و B شکلات در بسته بندی ۴۸ گرمی می فروشند. پنج بسته شکلات به صورت تصادفی از یک فروشگاه مواد غذایی از دو محصول انتخاب شد و تمام وزن بسته ها به دقت اندازه گیری شد. نتیجه زیر به دست آمد:

A: ۴۸/۰۸, ۴۸/۳۲, ۴۷/۹۶, ۴۷/۸۴, ۴۷/۹۶

B: ۴۹/۱۶, ۴۸/۸۴, ۴۸/۸۸, ۴۹/۰۸, ۴۹

الف - کدام کارخانه شکلات بیش تری در بسته ها می فروشد؟ از چه شاخصی استفاده می کنید؟
ب - کدام کارخانه در توزیع شکلات یکنواخت تر عمل کرده است؟

۲- اگر دامنه تغییرات برابر صفر باشد در باره داده ها چه نتیجه ای می گیرید؟

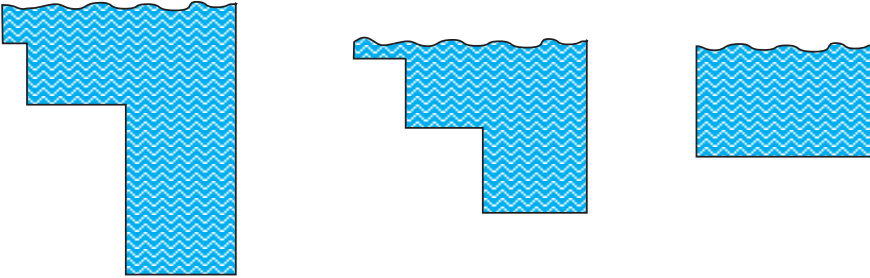
۳- تغییر واحد اندازه گیری چه تأثیری در دامنه تغییرات می گذارد؟ مثلاً اگر شما داده ها را برحسب متر اندازه گیری کرده باشید و سپس آن ها را برحسب سانتی متر به کار برید دامنه تغییرات چگونه تغییر می کند؟

۴- اگر به داده ها یک مقدار ثابت اضافه کنیم دامنه تغییرات چگونه تغییر می کند؟



واریانس

اگر شما خوب شنا نمی دانستید و می خواستید در استخری شنا کنید که عمق آن متغیر است برای این که اطمینان حاصل کنید که در خطر نخواهید بود چه اطلاعاتی را لازم دارید؟ اگر میانگین عمق استخر را به شما داده باشند و این میانگین برابر $1/5$ متر باشد، آیا برای شنا در این استخر اطمینان دارید؟ به حالات زیر توجه کنید.



در هر سه استخر بالا میانگین عمق برابر $1/5$ متر است. می خواهیم با یک شاخص عددی پراکندگی متغیر عمق را نسبت به عمق میانگین بررسی کنیم. اگر بدانیم که تغییرات عمق در اطراف عمق میانگین، یعنی $1/5$ متر زیاد نیست با اطمینان بیشتری در آن شنا خواهیم کرد. به شرط آن که این پراکندگی با یک شاخص عددی مناسب معرفی شده باشد.

دامنه تغییرات شاخص مناسبی است. اما در تصمیم گیری های کلان از ارزش آماری زیادی برخوردار نیست. زیرا ما به شاخص هایی نیاز داریم که هم پراکندگی داده ها و هم فراوانی آن ها را مدنظر قرار دهند.

دیدیم که پراکندگی یعنی این که داده ها از مرکز خود چه قدر دور هستند، پس یک راه ابتدایی ممکن است این طور به نظر برسد که تک تک مقادیر را از میانگین کم کنیم. این تفاضل ها را انحراف از میانگین می نامیم و سپس مجموع مقادیر حاصل را به دست آوریم، یعنی مجموع مقادیر

$$x_1 - \bar{x} \text{ و } x_2 - \bar{x} \text{ و } \dots \text{ و } x_n - \bar{x}$$

را حساب کنیم. قبلاً دیده ایم که مجموع مقادیر بالا برابر صفر است. لذا برای هر نوع داده ای اعم از آن که داده ها به میانگین نزدیک باشند و یا از آن دور باشند، آن چه که از این روش به دست می آید برابر صفر است. پس معیار مذکور نمی تواند معیار مفیدی باشد. البته علت آن که این مجموع برابر صفر شد به این خاطر است که برخی از داده ها از میانگین بزرگ تر و برخی دیگر کوچک ترند. در نتیجه مقادیر مثبت و منفی حاصل می شوند که مجموع آن ها یک دیگر را خنثی می کنند. برای جلوگیری از علامت های مثبت و منفی یکی از راه ها استفاده از قدر مطلق است. یعنی از مقادیر $|x_1 - \bar{x}|$ و \dots و $|x_n - \bar{x}|$

استفاده کنیم. این مقادیر می‌تواند معیاری برای پراکندگی ارائه کند ولی معمولاً کار کردن با قدر مطلق کار آسانی نیست و از این رو از توان دوم مقادیر بالا استفاده می‌کنیم یعنی مجموع مجذورات فاصله از میانگین را در نظر می‌گیریم. پس یک معیار پراکندگی ممکن است به صورت زیر باشد:

$$(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

ملاحظه می‌شود که اگر داده‌ها برابر باشند مقدار بالا برابر صفر است و بالعکس اگر مقدار بالا برابر صفر باشد داده‌ها برابر بوده و در نتیجه هیچ تفاوتی در بین آن‌ها نیست. اما اشکالی که دستور بالا دارد آن است که اگر تعدادی داده دیگر اضافه کنیم به مجموع بالا مقادیر مثبت دیگری اضافه خواهد شد و لذا این مجموع بزرگ‌تر خواهد شد. حال آیا درست است که بگوییم با اضافه شدن داده‌ها پراکندگی بیشتر می‌شود؟ مسلماً پاسخ منفی است، چون پراکندگی چیزی است که به اصل جامعه مربوط است و این نوع محاسبات نباید در مقدار آن‌ها تأثیر بگذارد. پس برای آن که تأثیر تعداد داده‌ها را از بین ببریم مجموع مذکور را بر تعداد آن‌ها تقسیم می‌کنیم و از این‌جا به معیار زیر می‌رسیم:

واریانس

واریانس برابر میانگین مجذور انحرافات از میانگین است و آن را با σ^2 نشان

می‌دهیم. بنابراین:

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

(σ از حروف کوچک یونانی است و سیگما خوانده می‌شود. حرف بزرگ آن Σ است)



تعداد ساعاتی که ۴ دانش‌آموز در طول هفته به ورزش اختصاص داده‌اند در زیر آمده است :
واریانس این داده‌ها را حساب کنید.

۱، ۵، ۷، ۹

ابتدا میانگین آن‌ها را حساب می‌کنیم.

$$\text{میانگین : } x = \frac{1+5+7+9}{4} = \frac{22}{4} = 5.5$$

سپس انحراف از میانگین را حساب می‌کنیم.

انحرافات از میانگین : $-4/5, -0/5, 1/5, 3/5$

مجدور انحرافات از میانگین : $20/25, 0/25, 2/25, 12/25$

$$\text{میانگین مجدور انحرافات از میانگین : } \frac{20/25 + 0/25 + 2/25 + 12/25}{4} = \frac{35}{4} = 8.75$$

$$\sigma^2 = 8.75 \text{ پس}$$

اگر داده‌ها برابر باشند واریانس آن‌ها صفر است و بالعکس.

اگر داده‌ها برابر باشند، میانگین آن‌ها نیز همان مقدار مشترک داده‌ها خواهد بود. پس انحرافات از میانگین برابر صفر است. لذا مجموع مجدورات انحراف صفر خواهد بود. پس واریانسی که از تقسیم این مجموع بر تعداد داده‌ها به دست می‌آید برابر صفر است. حال فرض کنید واریانس صفر باشد. پس باید مجموع مجدورات انحرافات صفر شده باشد. مجدور انحرافات عددی نامنفی است و چون مجموع آن‌ها صفر است لازم می‌آید هر یک از آن‌ها صفر باشند، چون در غیر این صورت مجموع مجدور انحرافات صفر نخواهد شد. با صفر شدن هر یک از انحرافات نتیجه می‌شود که تمام داده‌ها برابر میانگین هستند. پس مساوی‌اند.



نکاتی دربارهٔ واحد میانگین و واریانس: فرض کنید مطالعه‌ای دربارهٔ قد افراد انجام می‌دهید، و داده‌های شما برحسب سانتی متر اندازه‌گیری و بیان شدند. از آن جایی که میانگین به وسیله مجموع داده‌ها به دست می‌آید آن هم برحسب سانتی متر خواهد بود. یعنی واحد میانگین از نوع همان واحد داده‌هاست. اما در واریانس شرایط فرق می‌کند زیرا در این جا شما مقادیر $(x_i - \bar{x})$ را که برحسب سانتی متر است به توان ۲ می‌رسانید، پس مجذور انحرافات بر حسب سانتی متر مربع است و لذا مجموع آن‌ها و در نتیجه واریانس برحسب سانتی متر مربع خواهد بود. بنابراین واحد واریانس از نوع مجذور واحد متغیر است.

دستور دیگری برای محاسبه واریانس: دیدیم که برای محاسبه واریانس باید میانگین مجذور انحرافات از میانگین را حساب کنیم، یعنی

$$\sigma^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$$

محاسبه واریانس با استفاده از این دستور در بعضی مواقع مشکلاتی دارد.

می‌توان ثابت کرد عبارت بالا را می‌توان به صورت زیر نیز نوشت

$$\sigma^2 = \frac{1}{n} [x_1^2 + \dots + x_n^2] - \bar{x}^2$$

اثبات این دستور چندان مشکل نیست، توجه کنید که

$$\begin{aligned}\sigma^2 &= \frac{1}{n} \left[(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right] \\ &= \frac{1}{n} \left[(x_1^2 - 2x_1\bar{x} + \bar{x}^2) + \dots + (x_n^2 - 2x_n\bar{x} + \bar{x}^2) \right] \\ &= \frac{1}{n} \left[(x_1^2 + \dots + x_n^2) - 2(x_1 + \dots + x_n)\bar{x} + n\bar{x}^2 \right] \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - \frac{2(x_1 + \dots + x_n)}{n}\bar{x} + \bar{x}^2 \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - 2\bar{x} + \bar{x}^2 \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2\end{aligned}$$

توجه کنید در $x_1^2 + \dots + x_n^2$ ابتدا مقادیر x_i ها را به توان ۲ می‌رسانیم و سپس آن‌ها را با هم جمع می‌کنیم و در \bar{x}^2 ، \bar{x} (میانگین داده‌ها) را به توان دو می‌رسانیم. حال واریانس داده‌های زیر را حساب می‌کنیم.

۱، ۵، ۶، ۷، ۹

$$\bar{x} = \frac{1+5+6+7+9}{5} = \frac{28}{5} = 5.6 \quad \text{ابتدا میانگین را حساب می‌کنیم:}$$

$$\begin{aligned}x_1^2 + \dots + x_5^2 &= 1+25+36+49+81 \\ &= 192\end{aligned}$$

$$\sigma^2 = \frac{192}{5} - (5.6)^2 \quad \text{پس:}$$

$$= 38.4 - 31.36 = 7.04$$

با استفاده مستقیم از دستور واریانس نیز به همین نتیجه می‌رسیدیم

$$\begin{aligned}\sigma^2 &= \frac{(1-5.6)^2 + (5-5.6)^2 + (6-5.6)^2 + (7-5.6)^2 + (9-5.6)^2}{5} \\ &= \frac{21.16 + 0.36 + 0.16 + 1.96 + 11.56}{5} \\ &= 7.04\end{aligned}$$



انحراف معیار

این که واحد واریانس از نوع مجذور واحد متغیر است می تواند مشکلات و سوء تفاهماتی را در پی داشته باشد. فرض کنید شما قدها را برحسب متر اندازه گیری کرده باشید و همکلاسی شما قدها را برحسب سانتی متر، در این صورت واریانسی که همکلاسی شما حساب می کند ۱۰۰۰۰ برابر واریانسی است که شما حساب کرده اید. همکلاسی شما ممکن است این طور نتیجه بگیرد که پراکندگی در قد بسیار زیاد و چشم گیر است و حال آن که نظر شما ممکن است غیر از این باشد. برای آن که این اختلاف نظرها را از بین ببریم سعی می کنیم که تفاوت عمده در واحد واریانس و واحد میانگین را با جذر گرفتن از واریانس از بین ببریم. جذر واریانس را انحراف معیار می گوئیم.

انحراف معیار که با نماد σ نشان داده می شود برابر جذر واریانس است.

واحد انحراف معیار همان واحد متغیر است.

تعداد ساعت هایی در هفته که ۴ نفر به صورت داوطلبانه در یک بیمارستان خدمت می کنند در

۱، ۳، ۴، ۵، ۶

زیر آمده است. انحراف معیار داده ها را حساب کنید.

$$\bar{x} = \frac{1+3+4+5+6}{5} = \frac{19}{5} = 3.8$$

$$\sigma^2 = \frac{(-2/8)^2 + (-1/8)^2 + (0/2)^2 + (1/2)^2 + (2/2)^2}{5}$$

$$= \frac{7/84 + 0/64 + 0/04 + 1/44 + 4/84}{5}$$

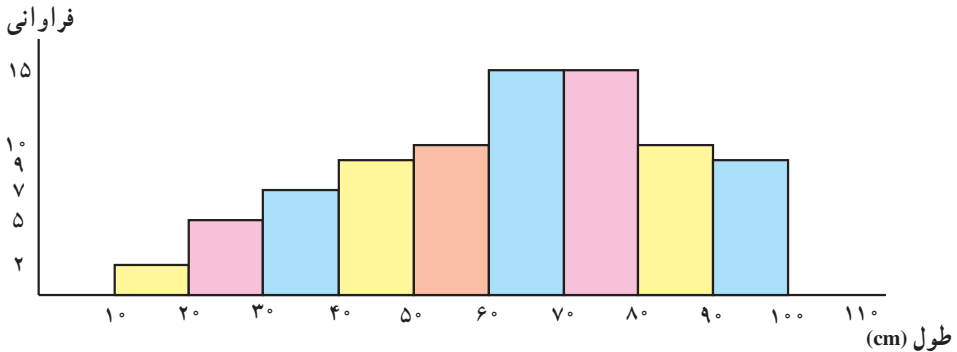
$$= \frac{14/80}{5} = 2/96$$

$$\sigma = \sqrt{3/96} \approx 1/7$$

در نتیجه :

تمرین

- ۱- نمودار زیر طول ۸۲ گیاه پرورش داده شده در یک آزمایشگاه را نشان می‌دهد.
 - طول چند گیاه بین ۵۰ تا ۶۰ سانتی‌متر است؟
 - طول چند گیاه بین ۷۰ تا ۹۰ سانتی‌متر است؟
 - میانگین و انحراف معیار طول گیاهان را به دست آورید.
 - توضیح دهید چه ویژگی نمودار بیان‌گر آن است که میانه و میانگین برابر نیستند.



- ۲- درجه حرارت بدن ۸ بیمار (به نزدیک‌ترین عدد صحیح گرد شده است) در زیر آمده است:

۳۸ ۳۸ ۳۹ ۳۹ ۴۰ ۴۰ ۴۱ ۴۱

میانگین و انحراف معیار را حساب کنید.

درجه حرارت‌های واقعی به قرار زیر می‌باشند:

$38/0$ $38/4$ $39/0$ $39/4$ $39/8$ $40/2$ $40/9$ $41/2$

- میانگین و انحراف معیار را حساب کنید.

- گرد کردن داده‌ها چه تأثیری در میانگین و انحراف معیار داشت؟

- آیا گرد کردن همواره تأثیری بر میانگین و انحراف معیار دارد؟

- ۳- در یک نمونه ۹۰۰ نفری از مردان درباره مسافتی که روزانه از منزل به محل کار طی می‌کنند سؤال شد. اعداد صفحه بعد نتیجه این بررسی را نشان می‌دهد.

درصد	مسافت طی شده (کیلومتر)
۱۷	کمتر از ۱
۳۱	۱ و کمتر از ۳
۳۸	۳ و کمتر از ۸
۷	۸ و کمتر از ۱۰
۵	۱۰ و کمتر از ۱۵
۲	۱۵ و کمتر از ۵۰
۱۰۰	

– نمودار مستطیلی این داده‌ها را رسم کنید.

– میانگین و انحراف معیار را برای این داده‌ها به دست آورید.

– اگر میانگین مسافت طی شده توسط زنان کارمند ۵ کیلومتر باشد، با فرض آن که انحراف

معیار هر دو گروه یکسان است، آیا می‌توانید بگویید که مردان فاصلهٔ بیش‌تری طی می‌کنند یا

زنان؟

۴– دو گروه داده به صورت زیر داریم :

گروه اول : ۲ ۳ ۴ ۵ ۶

گروه دوم : ۴ ۵ ۶ ۷ ۸

الف – واریانس هر گروه را حساب کنید.

ب – این دو گروه داده چه رابطه‌ای با هم دارند؟

ج – رابطه موجود بین دو گروه داده، چه اثری روی واریانس‌ها داشته است؟

د – با توجه به تعریف واریانس، نتیجهٔ حاصل از قسمت ج را توضیح دهید.

۵– دو گروه داده به صورت زیر داریم :

گروه اول : ۱ ۲ ۳ ۴ ۵

گروه دوم : ۳ ۶ ۹ ۱۲ ۱۵

الف – واریانس هر گروه را حساب کنید.

ب – این دو گروه چه رابطه‌ای با هم دارند؟

ج – رابطه موجود بین دو گروه داده، چه تأثیری روی واریانس داشته است؟

د – نتیجه حاصل از قسمت ج را توضیح دهید.

۶- در حالت کلی ثابت کنید که اگر داده‌ها را با عددی جمع کنیم، در واریانس آن‌ها تغییری حاصل نخواهد شد. به زبان نمادها $\sigma_{a+x}^2 = \sigma_x^2$

۷- در حالت کلی ثابت کنید که اگر داده‌ها را در عددی ضرب کنیم واریانس آن‌ها در مجذور این عدد ضرب خواهد شد. به زبان نمادها $\sigma_{ax}^2 = a^2 \sigma_x^2$

۸- با ترکیب دو تمرین بالا نتیجه بگیرید که $\sigma_{ax+b}^2 = a^2 \sigma_x^2$

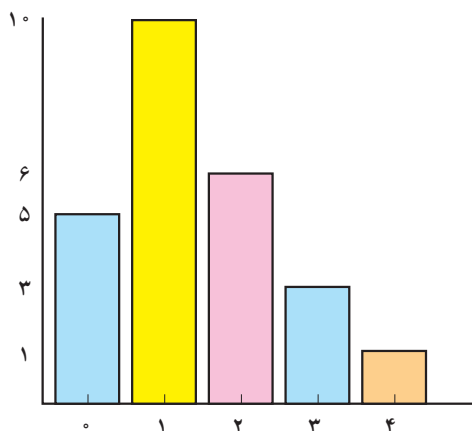
۹- کدامیک از برابری‌های زیر دربارهٔ انحراف معیار صحیح است؟

$$\sigma_{ax+b} = a\sigma_x \quad \sigma_{ax+b} = -a\sigma_x \quad \sigma_{ax+b} = |a|\sigma_x$$

۱۰- در یک بررسی تعداد فرزندان ۲۵ نفر از زنان مورد سؤال بود. نتیجه در نمودار زیر نشان داده شده است.

الف - نشان دهید متوسط تعداد فرزندان برابر $1/4$ است.

ب - نشان دهید انحراف معیار این داده‌ها تقریباً $1/6$ است.



- گروه ۲۵ نفری دیگری از زنان مورد بررسی قرار گرفتند. در این گروه متوسط تعداد فرزندان $2/4$ و انحراف معیار آن‌ها ۲ است. با توجه به اطلاعات به دست آمده در الف و ب، تفاوت بین تعداد فرزندان این دو گروه را بررسی کنید.

ضریب تغییرات

مسلماً قد همکلاسی‌های شما متفاوت است، همچنین وزن آن‌ها نیز چنین است ولی در سن همکلاسی‌های شما اختلاف زیادی نیست. یعنی انتظار می‌رود واریانس سن کم، ولی واریانس قد و وزن زیاد باشد. حال سؤال این است که تغییرات در قد همکلاسی‌ها بیش‌تر است یا در وزن آن‌ها. سن تقریباً ثابت است، زیاد متغیر نیست ولی می‌خواهیم بدانیم از قد و وزن کدام بیش‌تر متغیر است. اگر این دو صفت بر حسب یک واحد می‌بودند مقایسه واریانس آن‌ها می‌توانست جواب‌گوی این سؤال باشد ولی چون واحدها یکی نیستند مقایسه میزان پراکندگی این دو متغیر از طریق واریانس‌ها ممکن نیست. برای آن‌که بتوانیم به سؤال بالا جواب دهیم باید شاخصی معرفی کنیم که بدون واحد باشد تا امکان مقایسه فراهم آید. اگر به خاطر داشته باشید در جدول فراوانی برای آن‌که تأثیر اندازه نمونه را از بین ببریم به محاسبه درصد فراوانی نسبی پرداختیم. در این جا نیز برای از بین بردن واحد اندازه‌گیری از معیار ضریب تغییرات استفاده می‌کنیم.

ضریب تغییرات که با نماد CV نشان می‌دهیم عبارت است از خارج قسمت انحراف معیار بر میانگین



تذکر: از آنجایی که ضریب تغییرات معیاری برای میزان پراکندگی است باید مثبت باشد، چون پراکندگی منفی معنی ندارد. پس لازم است \bar{x} ، که در مخرج کسر بالا آمده است، مثبت باشد. برای اطمینان از مثبت بودن \bar{x} ، ضریب اطمینان را فقط برای داده‌های مثبت تعریف می‌کنیم.

ضریب تغییرات داده‌های زیر را محاسبه می‌کنیم:

۱، ۳، ۴، ۵، ۶

قبلاً دیدیم که $\bar{x} = 3/8$ و $\sigma = 1/7$ پس

$$CV = \frac{1/7}{3/8}$$

تذکر: هرچه قدر داده‌ها بزرگ‌تر باشند زمانی پراکندگی در آن‌ها محسوس است که تفاوت داده‌ها بیش‌تر باشد. ما اختلاف بین ۳ و ۳/۵ را به خوبی درک می‌کنیم ولی اختلاف بین ۳۰۰۰ و ۳۰۰۰/۵ چندان برای ما معنی‌دار نیست. شما در خرید اجناس ارزان قیمت انتظار تخفیف زیاد ندارید ولی در خرید اجناس گران قیمت از بحث و مجادله بر روی مبالغ کم خودداری می‌کنید. دلیل این امر آن است که در مقادیر بزرگ می‌توان از اختلاف‌های کم صرف‌نظر کرد. مثلاً از نظر آماری داده‌های زیر تقریباً با هم برابرند و تفاوت زیادی با هم ندارند.

۱۱۱، ۱۱۱/۵، ۱۱۲، ۱۱۲/۵

ولی داده‌های زیر با هم تفاوت دارند

۱، ۱/۵، ۲، ۲/۵

وقتی از ۱ به ۱/۵ می‌رویم پنجاه درصد به آن اضافه می‌شود ولی وقتی که از ۱۱۱ به ۱۱۱/۵ می‌رویم حدود چهارهزارم ۱۱۱ به آن اضافه می‌شود که ناچیز است. بنابراین پراکندگی به طریقی باید به نسبت بزرگی داده‌ها تعدیل شود. از این رو اگر ما انحراف معیار را بر \bar{x} تقسیم کنیم میزان پراکندگی را برای یک واحد از میانگین حساب کرده‌ایم و به این ترتیب تأثیر بزرگی داده‌ها را از بین برده‌ایم. با ملاحظه بالا می‌توان توصیف زیر را برای ضریب تغییرات به کار برد.

ضریب تغییرات عبارت است از میزان پراکندگی به ازای یک واحد از میانگین

تمرین

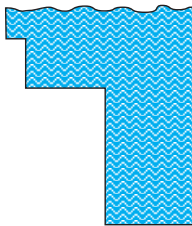
- ۱- اگر میانگین برابر ۴ و انحراف معیار برابر ۶ باشد ضریب تغییرات چه قدر است؟
- ۲- ضریب تغییرات سن دانش‌آموزان کلاس شما ۱۰ سال دیگر چه تغییری می‌کند؟ کم‌تر می‌شود یا بیش‌تر؟
- ۳- اگر 20% نمره هر دانش‌آموز به نمره او اضافه شود چه تأثیری روی ضریب تغییرات نمره‌ها حاصل می‌شود؟
- ۴- دو دسته داده زیر قیمت کالایی را در دو بازار جداگانه نشان می‌دهد.

۸, ۱۳, ۹, ۱۲, ۱۰, ۱۱, ۱۲, ۹, ۱۰, ۱۱, ۱۰

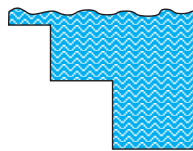
۱۰, ۱۳, ۸, ۱۰, ۹, ۱۱, ۱۰, ۱۰, ۱۱, ۹, ۱۰

- الف - در کدام دسته پراکندگی بیش‌تر مشهود است؟
- ب - دامنه تغییرات را در هر دو دسته حساب کنید.
- ج - آیا مقدار دامنه تغییرات با پراکندگی مشاهده شده در «الف» هم‌خوانی دارد؟
- د - ترجیح می‌دهید از کدام بازار خرید کنید؟ چرا؟
- ه - اگر داده‌های واقعی در اختیار شما نباشند، فقط به صرف داشتن دامنه تغییرات می‌توانستید تصمیم‌گیری کنید؟
- و - کدام یک از شاخص‌های پراکندگی در تصمیم‌گیری شما می‌تواند مؤثرتر باشد؟ این شاخص را معرفی کرده و محاسبه کنید.

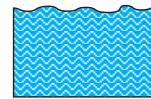
۵- در محله‌ای سه استخر شنا با مشخصات زیر موجود است :



میانگین عمق = ۱/۵ متر
انحراف معیار = ۱/۰۵



میانگین عمق = ۱/۵ متر
ضریب تغییرات = ۰/۵



میانگین عمق = ۱/۵ متر
انحراف معیار = ۰

- الف - برای فردی مبتدی با قد ۱/۴۵ متر شنا کردن در کدام یک از این استخرها امنیت بیش‌تری دارد؟ چرا؟
- ب - در کدام استخر با خطر بیش‌تری روبه‌رو است؟ چرا؟

۶- دو دسته داده زیر ارزش واردات برحسب کشورهای مبدأ را در سال‌های ۶۵ و ۷۰ نشان

می‌دهد. (میلیون ریال)

کشور و منطقه	۱۳۵۵	۱۳۶۰	۱۳۶۵	۱۳۶۶	۱۳۶۷	۱۳۶۸	۱۳۶۹	۱۳۷۰
امارات	۳۲۴۳	۳۲۳۷۱	۴۱۳۴۳	۲۷۴۰۸	۱۹۱۹۹	۶۸۶۷۵	۶۵۴۵۲	۹۵۶۶۱
اندونزی	۴۱	۳۴	۲۱۱	۱۱۰۵	۷۳۴	۵۰۱۵	۴۳۳۵	۶۹۱۱
بنگلادش	۱۰۵۰	۳۰۹۸	۲۲۲۸	۲۶۷۷	۱۱۸۶	۳۱۹۶	۳۳۲۴	۳۰۹۶
پاکستان	۴۴۶۸	۶۷۲۱	۷۰۴۱	۴۳۰۴	۲۷۴۲	۱۰۴۵۵	۳۲۹۶	۱۲۴۰۳
تایلند	۱۳۷۵	۱۹۸۷۵	۴۰۲۵	۹۵۶۹	۴۸۹۳	۱۳۴۳۳	۱۳۲۷۷	۷۲۰۶
تایوان	۹۲۰	۱۰۵۱۰	۷۳۳۰	۴۶۱۹	۴۷۵۴	۵۲۹۱	۲۸۹۸	۱۰۲۹۸
ترکیه	۲۹۳۷	۲۳۳۶۹	۴۵۲۱۶	۳۳۳۳۱	۲۹۵۲۹	۵۰۶۴۵	۴۸۹۳۶	۶۹۶۳۳
چین	۷۰۹۳	۹۶۳۱	۱۹۹۳	۳۷۴۷	۲۴۸۶	۳۴۶۰	۱۱۶۱۳	۱۹۲۶۲
ژاپن	۱۵۵۵۰۰	۱۲۹۶۰۶	۹۷۶۲۷	۷۴۰۶۸	۵۸۰۶۶	۷۰۴۳۹	۱۳۰۲۴۸	۲۱۶۹۲۵
سنگاپور	۲۷۷۳	۸۲۵۴	۵۷۰۰	۳۸۱۶	۵۳۲۸	۵۵۵۴	۷۴۴۶	۹۰۳۰
سوریه	۲۷۳	۱۷۸	۹۱۴	۲۵۸۲	۶۸۸۳	۸۸۷۸	۲۰۰۵	۹۰۷
عربستان	۲۲	۱۶۹۱	۲۰۶۶	۷۳۲	۹۱۰	۶۳۸۲	۴۶۱	۲۰۸۶

– تغییرات ارزش واردات در کدام سال بیش‌تر بوده‌است؟

۷- با مراجعه به جدول صفحه بعد به سؤالات زیر پاسخ دهید.

۱۰ مرکز استان را به تصادف انتخاب کنید و بر اساس میزان بارندگی سالانه در این مراکز،

میزان بارندگی سالانه در ایران در این سال را حدوداً به‌دست آورید.

– در این جامعه چند متغیر وجود دارد؟ نام ببرید و نوع آن‌ها را مشخص کنید.

– آیا میزان بارندگی سالانه در مراکز استان‌ها، به میزان بارندگی سالانه در ایران نزدیک است یا خیر؟

– برای اندازه‌گیری تفاوت‌ها از چه معیارهایی استفاده می‌کنید؟

– میزان بارندگی به طور متوسط در هر ماه در رشت چقدر است؟

سطر مربوط به بارندگی در استان گیلان (به مرکز رشت) معرف چه متغیری است؟ آن متغیر را

معرفی کنید و نوع آن را مشخص کنید.

(میلیمتر)

مقدار بارندگی سالانه مراکز استانها

۱۳۷۶	۱۳۷۵	۱۳۷۴	۱۳۷۳	۱۳۷۲	۱۳۷۰	۱۳۶۵	مرکز استان
۲۵۶	۳۰۳	۳۶۷	۳۵۶	۳۲۰	۳۲۰	۳۴۸	اراک
۳۴۲	۳۴۷	۲۳۰	۳۴۴	۳۲۸	۰۰۰	۰۰۰	اردبیل
۲۶۹	۲۶۳	۳۶۷	۴۸۶	۵۳۴	۲۶۹	۳۸۶	ارومیه
۲۱۱	۶۷	۲۰۷	۷۹	۷۷	۱۱۲	۱۳۸	اصفهان
۵۳۹	۱۵۲	۲۳۹	۲۰۶	۱۶۸	۲۹۲	۲۸۶	اهواز
۷۱۷	۵۷۸	۵۴۹	۸۶۱	۷۴۱	۶۷۶	۰۰۰	ایلام
۴۳۷	۱۶۷	۵۰۷	۷۰	۳۶	۲۴۴	۲۳۰	بندرعباس
۸۰۷	۲۱۷	۵۸۸	۲۸۷	۶۴	۱۷۱	۱۱۹	بوشهر
۲۰۲	۲۵۹	۱۵۸	۳۳۶	۳۳۹	۲۳۶	۳۲۳	تبریز
۲۴۴	۱۵۶	۳۱۹	۲۵۵	۲۱۹	۲۶۰	۲۹۲	تهران
۶۰۹	۴۲۸	۵۳۷	۵۸۹	۵۸۷	۵۱۶	۵۹۴	خرم‌آباد
۱۲۹۷	۱۰۱۵	۱۲۱۰	۱۳۱۹	۱۶۸۵	۱۲۹۸	۱۱۸۶	رشت
۱۴۹	۷۳	۱۹۰	۵۶	۵۴	۴۴	۴۲	زاهدان
۳۳۱	۲۴۹	۲۶۶	۳۳۵	۴۲۵	۲۸۶	۳۲۱	زنجان
۱۶۱	۱۲۸	۱۹۲	۱۲۲	۶۷	۱۹۴	۲۲۲	سمنان
۵۷۶	۴۱۳	۵۱۳	۵۶۱	۵۹۹	۳۸۲	۵۴۸	سنندج
۴۲۳	۲۵۸	۳۷۹	۳۶۸	۲۸۲	۴۹۱	۴۹۲	شهرکرد
۴۹۸	۲۰۲	۵۷۸	۴۵۱	۱۵۵	۲۷۲	۴۹۵	شیراز
۳۳۶	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	قزوین
۱۳۸	۷۸	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	قم
۱۷۹	۱۲۳	۲۵۶	۱۷۶	۵۵	۱۱۵	۱۵۹	کرمان
۴۸۸	۳۹۶	۴۳۳	۵۵۹	۶۵۲	۴۵۱	۵۶۹	کرمانشاه
۵۳۶	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	گرگان
۳۴۸	۱۷۰	۲۲۴	۱۶۷	۳۲۷	۳۲۶	۲۵۰	مشهد
۳۳۶	۲۱۳	۳۱۹	۴۳۸	۲۸۸	۲۹۲	۴۱۷	همدان
۱۱۹۸	۴۷۴	۱۰۲۵	۱۰۶۳	۴۱۷	۱۲۰۲	۱۲۲۸	یاسوج
۴۲	۲۹	۱۰۴	۳۴	۲۱	۶۰	۶۴	یزد

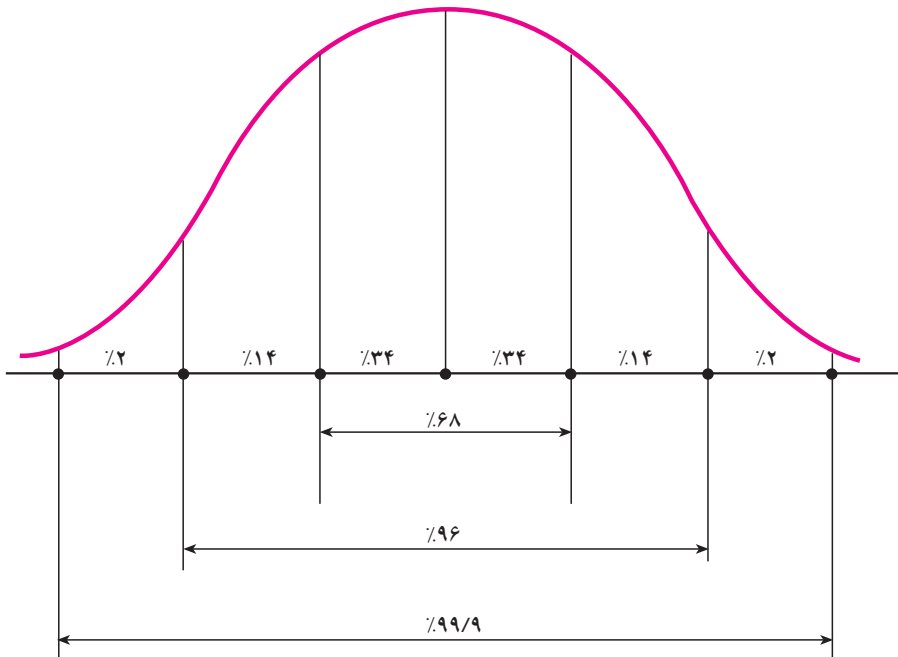
مأخذ: سازمان هواشناسی کشور.

پراکندگی در منحنی نرمال

قبلاً در فصل نمودارها و همچنین در فصل شاخص‌های مرکزی به منحنی نرمال اشاره کردیم. اگر σ انحراف معیار این جامعه باشد پراکندگی افراد در جامعه به صورت زیر است. برای درک بهتر موضوع ابتدا فرض می‌کنیم مثلاً میانگین برابر ۵ و انحراف معیار برابر ۲ باشد در این صورت می‌خواهیم بدانیم چند درصد افراد در فاصله‌ای به مرکز میانگین و به فاصله یک انحراف معیار از آن قرار دارند. در واقع می‌خواهیم بدانیم چند درصد افراد در فاصله $(2-5, 5+2)$ یعنی $(3, 7)$ قرار دارند.

در این منحنی نشان داده می‌شود که این درصد برابر 68% است.

همچنین درصد افرادی که در فاصله‌ای به مرکز میانگین و به طول دو برابر انحراف معیار از میانگین قرار دارند برابر 96% است و سرانجام تقریباً 100% افراد در فاصله‌ای به مرکز میانگین و به طول ۳ برابر انحراف معیار قرار دارند. این اطلاعات را به‌طور مختصر در شکل زیر ملاحظه می‌کنید (مقادیر روی نمودار تقریبی هستند):



محاسبات شاخص‌های عددی در MINITAB

– داده‌های زیر ساعتی را که ۱۰ دانشجوی (انتخاب شده به‌طور تصادفی) صرف تحقیق در طول هفته می‌کنند نشان می‌دهد:

۷,۱۴,۵,۰,۲,۷,۱۰,۴,۰,۸

با استفاده از MINITAB میانگین، میانه، دامنه تغییرات و انحراف معیار این داده‌ها را به‌دست آورید.

شما می‌توانید از Calc و Stat در لیست انتخاب استفاده نمایید.

Stat مقادیر چند شاخص عددی را هم‌زمان ارائه می‌دهد ولی Calc مقدار شاخص‌ها را یکی‌یکی به‌دست می‌آورد. با استفاده از MINITAB WINDOWS برای انتخاب Stat مراحل زیر را دنبال کنید:

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی Stat در لیست انتخاب، کلیک کنید.

قدم سوم: روی Basic Statistics کلیک کنید.

قدم چهارم: روی Descriptive Statistics کلیک کنید.

قدم پنجم: C1 را در کار زیر Variable تایپ کنید.

قدم ششم: روی OK کلیک کنید. نتایج و خروجی Minitab روی صفحه نمایان می‌شود.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، با استفاده از SET داده‌ها را در ستون

C1 وارد کنید. سپس فرمان زیر را تایپ کنید:

MTB > DESCRIBE C1

با اجرای این فرمان خروجی Minitab ظاهر می‌شود.

فعالیت

فعالیت مثال بالا را اجرا کنید و مقادیر به‌دست آمده را در جدول زیر بنویسید.

Descriptive Statistics

Variable	N	Mean	Median	Tr	Mean	Stder	SE	MEAN
hours	?	?	?	?	?	?	?	?
Variable	Min	Max	Q1	Q3				
hours	?	?	?	?				

همان طور که گفته شد می توان از دستور Calc نیز استفاده کرد، اگر از Minitab Windows استفاده می کنید به ترتیب زیر عمل کنید :

قدم اول: داده ها را در ستون C1 وارد کنید.

قدم دوم: روی Calc در لیست انتخاب، کلیک کنید.

قدم سوم: روی Column Statistics کلیک کنید.

قدم چهارم: در صفحه Column Statistics روی شاخصی که می خواهید محاسبه کنید کلیک کنید. به طور مثال برای پیدا کردن میانگین روی Mean کلیک کنید.

قدم پنجم: در کنار Input Variable در کادر، C1 را تایپ کنید.

قدم ششم: اگر می خواهید پاسخ را به عنوان یک عدد ثابت ذخیره کنید در کادر در کنار Store Result in، K1 را تایپ کنید. اگر از K1 قبلاً استفاده کرده اید K2 را تایپ کنید. اگر نمی خواهید نتیجه را ذخیره کنید، این مرحله را اجرا نکنید.

قدم هفتم: روی OK کلیک کنید. نتیجه روی صفحه نمایان خواهد شد.

اگر از فرمان های زبانی Minitab استفاده می کنید، با استفاده از SET داده ها را در ستون C1 وارد کنید. سپس فرمان زیر را تایپ کنید :

MTB > Mean C1

در پاسخ، Minitab خروجی زیر را ارائه می کند :

Column Mean

Mean of C1 =

به طور متشابه می توانید میانه، دامنه تغییرات و انحراف معیار را با کلیک کردن روی شاخص مورد نظر محاسبه کنید و فرمان های زبانی Minitab برای شاخص های فوق عبارتند از: MEDIAN C1، RANGE C1، STANDARD DEVIATION، که باید پس از PROMPT تایپ شوند.

نمودار جعبه ای

داده های زیر، درآمد هفتگی نمونه ۱۲ تایی از خانواده ها را نشان می دهد.

۲۳۰۰۰ ۱۷۰۰۰ ۳۲۰۰۰ ۶۰۰۰۰ ۲۲۰۰۰ ۵۲۰۰۰

۲۹۰۰۰ ۳۸۰۰۰ ۴۲۰۰۰ ۹۲۰۰۰ ۲۷۰۰۰ ۴۶۰۰۰

با استفاده از Minitab نمودار جعبه ای این داده ها را رسم کنید. اگر از Minitab Windows

استفاده می کنید، مراحل زیر را دنبال کنید :

قدم اول: داده ها را در ستون C1 وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی BOXPLOT کلیک کنید.

قدم چهارم: C1 را در کادر زیر Y تایپ کنید.

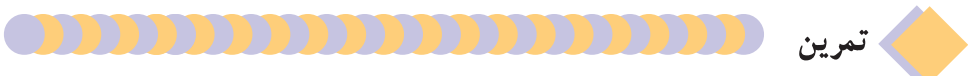
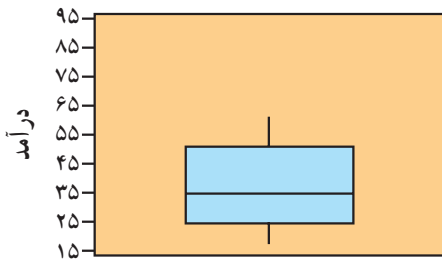
قدم پنجم: روی OK کلیک کنید. نمودار جعبه ای مورد نظر نمایان می شود. اگر از فرمان های

زبانی Minitab استفاده می کنید ابتدا با استفاده از فرمان SET داده ها را در ستون C1 وارد کنید.

سپس فرمان زیر را تایپ کنید :

MTB > Boxplot C1

با اجرای دستورات فوق نمودار جعبه ای زیر رسم می شود :



۱- نمودار ساقه و برگ زیر موجود است.

– میانگین، انحراف معیار این داده ها را با استفاده از Minitab به دست آورید.

– نمودار جعبه ای این داده ها را با استفاده از Minitab رسم کنید.

ساقه	برگ
۰	۳۶۹
۱	۲۸۵۱۰۵
۲	۵۱۶
۴	۱
۵	
۶	۲

و بدان! کسی که گنجینه‌های آسمان و زمین در دست اوست تو را در دعا رخصت داده و پذیرفتن دعایت را برعهده نهاده و تو را فرموده از او خواهی تا به تو دهد، و از او طلبی تا تو را بیمارزد. و میان تو و خود کسی را نگمارده تا تو را از وی بازدارد و از کسی ناگزیرت نکرده که نزد او برایت میانجی‌گری آرد، و اگر گناه‌کردی از توبهات منع ننموده و در کیفیت شتاب نفرموده و چون بدو باز گردی سرزنش نکند، و آنجا که رسوا شدنت سزااست پندهات را ندرد؛ و در پذیرفتن توبه بر تو سخت نگرفته و حساب گناهت را نکشیده، و از بخشایش نومیدت نگردانیده، بلکه بازگشتت را از گناه نیک شمرده و هر گناهت را یکی گرفته و هر کار نیکویت را ده به حساب آورده و در بازگشتت را برایت بازگذارده و چون بخوانیش آوایت را شنود، و چون راز خود را با او در میان نهی آن را داند، پس حاجت خود بدو نمایی و آنچه در دل داری پیش او بگشایی، و از اندوه خویش بدو شکایت کنی و خواهی تا غم تو را گشاید و در کارها یاریت نماید و از گنجینه‌های رحمت او آن را خواهی که بخشیدنش از جز او نیاید؛ از افزودن مدت زندگانی و تندرستی‌ها و در روزی‌ها فراوانی.

(تدریس این فصل اختیاری است)

همبستگی و رگرسیون

شهاب برای پروژه کلاس خود مایل بود تأثیر سطح تحصیلات والدین را بر موفقیت تحصیلی فرزندان مطالعه کند. در انجام این پروژه او با یک مشکل جدی روبرو بود و آن اینکه جامعه‌ای که با آن سروکار داشت شامل یک متغیر نبود و او نمی‌دانست که در این شرایط چگونه باید عمل کند. از این رو به معلم مراجعه کرد و معلم به او پاسخ داد: جامعه‌ای که می‌خواهی مطالعه کنی یک جامعه با دو متغیر است که موضوع درس امروز ماست.

ما در بررسی این قبیل جوامع نه تنها کارهایی را که تا بحال آموخته‌ایم به کار می‌بریم، بلکه به دو سؤال نیز سعی می‌کنیم پاسخ دهیم و آن اینکه این متغیرها چقدر به هم وابسته‌اند و دیگر آنکه اگر مقدار یکی از متغیرها را بدانیم آیا می‌توانیم مقدار متغیر دیگر را «حدس» بزنیم.

جامعه‌های چند متغیری

تاکنون جوامعی را بررسی کرده‌ایم که در آن‌ها فقط یک متغیر وجود داشت. شما درباره قد همکلاسی‌های خود مطالعه کرده‌اید. نمودارهای مربوطه، میانگین و واریانس آن را تعیین کرده‌اید. همین کار را درباره وزن همکلاسی‌های خود جداگانه انجام داده‌اید. ولی هیچ‌گاه این دو متغیر را با هم مطالعه نکرده‌اید. از تأثیرات آن‌ها بر روی یکدیگر اطلاعی ندارید، هنوز نمی‌دانید که اگر قد همکلاسی خود را بدانید آیا می‌توانید وزن او را «حدس» بزنید، یعنی به‌طور تقریبی بگویید که وزن او چقدر است؟ شما نمی‌دانید که آیا افراد قد بلندتر وزن بیشتری دارند یا خیر، تغییرات این دو متغیر با هم چگونه است؟ تنها کاری که کرده‌اید در یکی از تمرین‌ها تعیین کرده‌اید که از این دو متغیر کدامیک تغییرات بیشتری در مقایسه با دیگری دارد.

می‌توانیم مسائلی را مطرح کنیم که در آن‌ها تعداد متغیرها از دو تا هم بیشتر باشد. مثلاً در کشاورزی میزان محصول، میزان آب، میزان کود، نوع کود، میزان آفتاب و نوع خاک مؤثرند. ملاحظه می‌شود که در مسائل کشاورزی حداقل با ۶ متغیر سروکار داریم. بررسی این قبیل مسائل و

دستیابی به نوع تأثیر این متغیرها بر روی یکدیگر جزء مسائل عمده آماری است و معمولاً در بخش تحقیقات کشاورزی با این قبیل متغیرها سروکار دارند. ما در این فصل هم به جهت سادگی و هم به جهت آنکه بسیاری از مسائل را می‌توان به جامعه‌هایی با دو متغیر تجزیه کرد به بررسی جامعه‌های دو متغیری می‌پردازیم.

نمودار پراکنش

داده‌هایی را که تاکنون جمع‌آوری می‌کردید از نظر بگذرانید. به هر فرد یک عدد با یک صفت نسبت داده می‌شود، ولی اکنون به هر فرد دو مقدار نسبت داده می‌شود، زیرا ما دو متغیر داریم. برای هر متغیر یک مقدار اندازه گرفته می‌شود. برای آنکه مقادیر مربوط به متغیرها اشتباه نشود، داده‌ها را به صورت زوج مرتب می‌نویسیم. مثلاً اگر موضوع مورد مطالعه قد و وزن است و شما قد و وزن یکی از همکلاسی‌های خود را اندازه گرفته‌اید، داده‌ای به صورت مرتب شده (۵۲ و ۱۶۷) خواهید داشت که قد این همکلاسی را بر حسب سانتیمتر برابر ۱۶۷ و وزن او را بر حسب کیلو برابر ۵۲ است. نمایش (۵۲ و ۱۶۷) را یک «زوج مرتب» می‌نامند که نشان می‌دهد عددی که اول از سمت چپ نوشته شده نمایانگر قد و عددی که دوم نوشته شده نمایانگر وزن دانش‌آموز است.

فعالیت

قد و وزن ۱۰ نفر از همکلاسی‌های خود را اندازه بگیرید و داده‌های حاصل را به صورت زوج‌های مرتب نمایش دهید.



البته می‌توان داده‌ها را به صورت مرتب در دو ستون و یا دو سطر کنار هم نیز نشان داد. در یک مطالعه آماری خواسته‌اند تأثیر نوعی رژیم غذایی را روی موش‌های آزمایشگاهی بررسی کنند، از این رو وزن اولیه هر موش را اندازه گرفتند و پس از اعمال رژیم غذایی در مدت معینی، مجدداً وزن آن‌ها را اندازه گرفتند و داده‌ها عبارتند از:

(۲,۴) و (۳,۵) و (۱,۷) و (۲,۳) و (۱,۸)

که مؤلفه اول بیانگر وزن اولیه و مؤلفه دوم بیانگر وزن همان موش پس از اعمال رژیم غذایی است. نوع دیگری که می‌توان این داده‌ها را نشان داد چنین است:

شماره موش‌ها	۱	۲	۳	۴	۵
وزن اولیه	۱	۲	۱	۳	۲
وزن پس از اعمال رژیم غذایی	۸	۳	۷	۵	۴

ممکن بود این اطلاعات را به صورت ستونی هم بنویسیم.

شماره موش‌ها	وزن اولیه	وزن پس از رژیم
۱	۱	۸
۲	۲	۳
۳	۱	۷
۴	۳	۵
۵	۲	۴

در هر صورت ثبت داده‌ها باید به گونه‌ای باشد که اولاً داده‌های مربوط به یک فرد در کنار هم

باشند و ثانیاً مشخص باشد که کدام مقدار مربوط به کدام متغیر است.

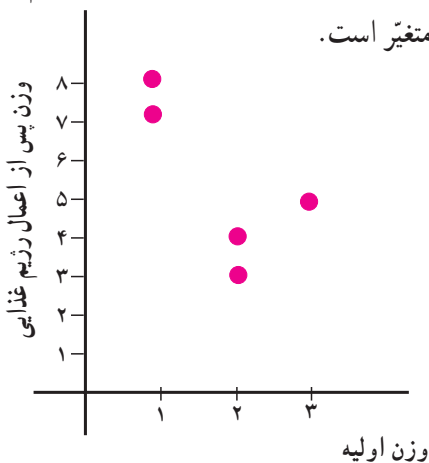
حال که داده‌ها را قرار شد به صورت زوج

مرتب نشان دهیم، می‌توانیم آن‌ها را به صورت

نقاطی در صفحه نمایش دهیم. شکل

روبرو داده‌های مربوط به وزن موش‌ها را نشان

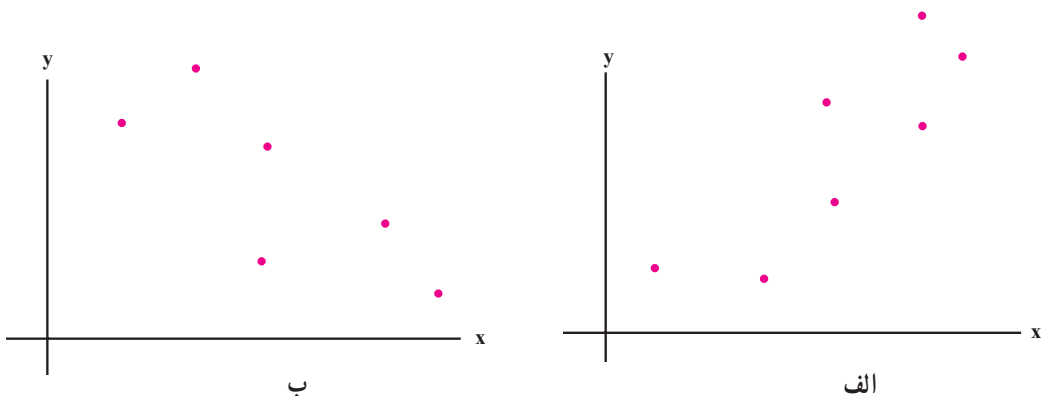
می‌دهد. این نمودار را نمودار پراکنش می‌گویند.



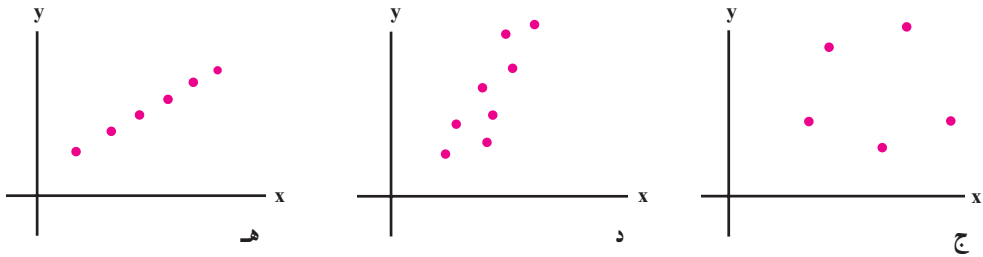


نمودار پراکنش مجموعه‌ای از نقاط در صفحه محورهای مختصات است که از رسم داده‌های مربوط به اندازه‌گیری در جامعه‌های دو متغیری به دست آمده است.

مطالعه این نمودار می‌تواند اطلاعاتی درباره جامعه در اختیار ما بگذارد. مثلاً نمودارهای زیر را در نظر بگیرید.



در نمودار «الف» وقتی مقادیر روی محور x ها افزایش می‌یابد، مقادیر روی محور y ها «عموماً» افزایش می‌یابد اما در نمودار «ب» برعکس است با افزایش مقادیر روی محور x ها، مقادیر روی محور y ها «عموماً» کاهش می‌یابد.



در نمودار «ج» تغییرات در مقدار x ها هیچ گونه اطلاعی درباره تغییرات y نمی‌دهد، یعنی رفتار مقادیر x رفتار مقادیر y را مشخص نمی‌کند. نمودار «د» علاوه بر آنکه نشان می‌دهد دو متغیر با هم افزایش و یا کاهش می‌یابند، با داشتن مقادیر x با دقت بیشتری می‌توان y را حدس زد. دقت حدس زدن مقادیر y در این نمودار از دقت حدس در نمودار «الف» و «ب» بیشتر است. در نمودار «ه» حدس مقادیر y با دقت بیشتری انجام می‌شود. برای روشن‌تر شدن این مفاهیم فعالیت زیر را انجام دهید.

فعالیت

نمودار پراکنش مربوط به داده‌های زیر را رسم کنید.

(۱,۲) و (۲,۳) و (۳,۴) و (۴,۵) و (۵,۶)

اگر نمودار را دقیق رسم کرده باشید متوجه خواهید شد که این نقاط روی یک خط راست قرار دارند، می‌توانید معادله آن خط را بیابید.

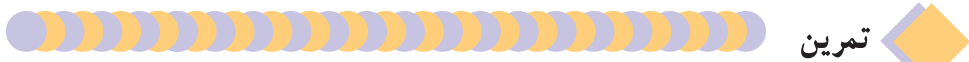
حال مطابق دستور العمل زیر مؤلفه دوم هر یک از این داده‌ها را تغییر دهید: برای هر نقطه سکه‌ای را پرتاب می‌کنیم اگر شیر آمد به y یک واحد اضافه می‌کنیم و اگر خط بیاید از y یک واحد کم می‌کنیم. با تغییرات حاصل شده نمودار پراکنش را رسم کنید. این فعالیت را تکرار کنید منتها این بار مقادیری را که اضافه و یا کم می‌کنید تغییر دهید.

چه تغییری در نمودار پراکنش مشاهده می‌کنید؟

وابستگی x و y چگونه تغییر می‌کند؟

در این فعالیت می‌بینید که هر چقدر وابستگی خطی y و x بیشتر باشد مقدار y حاصل از داده‌ها به مقدار y حاصل از معادله خط نزدیکتر است. یعنی با داشتن x مقدار y را با خطای کمتری می‌توانید تعیین کنید.

شما می‌توانید فعالیت بالا را به جای خط با هر منحنی دیگری تکرار کنید. ما در این فصل فقط با خط کار می‌کنیم.



تمرین

۱- تعداد ساعات صرف شده و نمره کسب شده در آزمون فارسی برای هر دانش آموز کلاس دوم یک دبیرستان به شکل زوج مرتب در زیر آورده شده است.

(۲, ۱۰) و (۳, ۱۰) و (۳, ۱۴) و (۴, ۱۰) و (۴, ۱۴)

(۵, ۱۴) و (۵, ۱۶) و (۶, ۱۲) و (۶, ۱۶) و (۶, ۱۸)

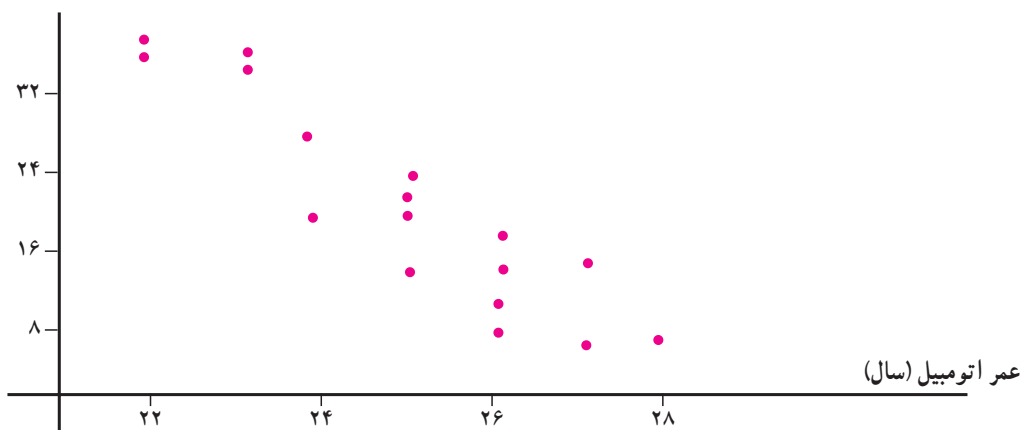
(۷, ۱۴) و (۷, ۱۸) و (۷, ۲۰) و (۸, ۱۶) و (۸, ۱۸)

نمودار پراکنش را برای نشان دادن ارتباط بین تعداد ساعات مطالعه و نمره آزمون رسم کنید.

چه نتیجه ای می گیرید.

۲- نمودار زیر قیمت و عمر تعدادی اتومبیل را نشان می دهد.

قیمت (۱۰۰۰۰۰۰ تومان)

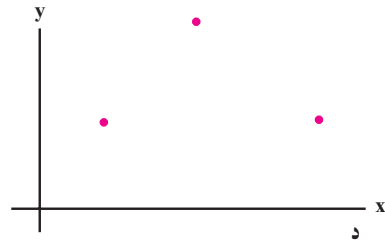
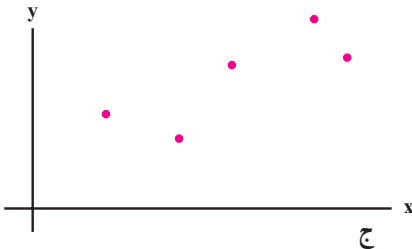
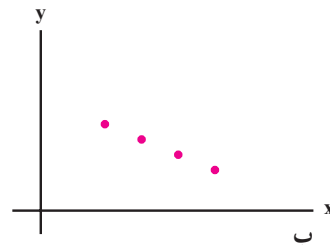
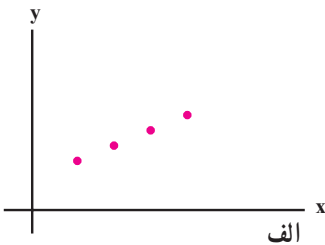


آیا ارتباطی بین این دو متغیر وجود دارد؟ توضیح دهید.



همبستگی

در فعالیت‌های انجام شده دیده‌اید که پس از تغییراتی که در داده‌ها داده‌اید در برخی از آن‌ها نقاط به شکل یک خط نزدیکتر و در برخی دیگر دورتر می‌شوند. اگر نقاط به شکل یک خط نزدیکتر باشند به کمک x با خطای کمتری می‌توان y را تعیین کرد و برعکس هرچقدر نقاط از خط دورتر باشند خطای تعیین y بیشتر خواهد بود. هر چقدر خطای تعیین y به کمک x کمتر باشد می‌نویسیم دو متغیر همبستگی خطی بیشتری دارند مثلاً در شکل‌های زیر



در «الف» و «ب» بدون خطا می‌توانیم y را به کمک خطی که از این نقاط می‌گذرد تعیین کنیم. پس همبستگی بین x و y خیلی زیاد است. ولی در شکل «ج» نمی‌توانیم خطی را تعیین کنیم که تمام نقاط روی آن باشند. به نظر می‌رسد این نقاط به نحوی از خط راست دور شده‌اند پس همبستگی آن به اندازه همبستگی در شکل‌های «الف» و «ب» نیست. ولی باز خطای تعیین y به کمک x زیاد نیست. اما این خطا در شکل «د» خیلی زیاد است.

ما معیاری می‌خواهیم که این همبستگی را برای ما اندازه‌گیری کند. دستوری که برای محاسبه همبستگی معرفی شده است با نام ضریب همبستگی و با نماد r ، به صورت زیر است:

$$r = \frac{\text{میانگین } (y \text{ ها}) - (\text{میانگین } x \text{ ها}) \times \text{مجموع حاصلضرب } x \text{ ها و } y \text{ ها}}{\sqrt{(\text{انحراف معیار } y \text{ ها}) \times (\text{انحراف معیار } x \text{ ها})}}$$

در زیر داده‌هایی مربوط به وزن اولیه و بعد از رژیم موش‌ها آمده است.

شماره موش‌ها	۱	۲	۳	۴	۵	جمع
وزن اولیه (x)	۱	۱	۲	۳	۲	۹
وزن بعد از اعمال رژیم غذایی (y)	۸	۳	۷	۶	۴	۲۸
حاصلضرب xها و yها	۸	۳	۱۴	۱۸	۸	۵۱

ضریب همبستگی بین وزن اولیه و وزن بعد از اعمال رژیم غذایی را محاسبه کنید.
در این مسئله $n=5$ و x ها را وزن اولیه و y ها را وزن بعد از اعمال رژیم غذایی گرفته‌ایم.
ابتدا به محاسبه میانگین x و y می‌پردازیم:

$$\bar{x} = \frac{9}{5} = 1/8, \quad \bar{y} = \frac{28}{5} = 5/6,$$

$$\begin{aligned} \text{واریانس } x\text{ها} &= \frac{(1-1/8)^2 + (1-1/8)^2 + (2-1/8)^2 + (3-1/8)^2 + (2-1/8)^2}{5} \\ &= \frac{0/64 + 0/64 + 0/04 + 1/44 + 0/04}{5} = \frac{2/8}{5} = 0/56 \end{aligned}$$

$$\begin{aligned} \text{واریانس } y\text{ها} &= \frac{(8-5/6)^2 + (3-5/6)^2 + (7-5/6)^2 + (6-5/6)^2 + (4-5/6)^2}{5} \\ &= \frac{5/76 + 6/76 + 1/96 + 0/16 + 2/56}{5} = \frac{17/2}{5} = 3/44 \end{aligned}$$

$$\frac{\text{مجموع حاصلضرب } x\text{ها و } y\text{ها}}{n} = \frac{51}{5} = 10/2$$

بنابراین:

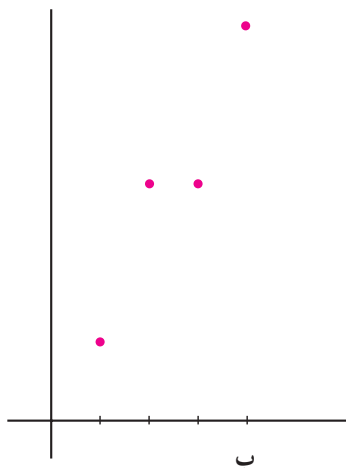
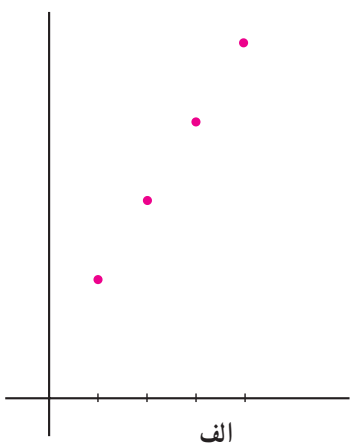
$$\begin{aligned} \text{ضریب همبستگی} &= \frac{10/2 - (1/8)(5/6)}{(\sqrt{0/56})(\sqrt{3/44})} \approx \frac{0/12}{1/39} \\ &\approx 0/09 \end{aligned}$$

حال این سؤال پیش می‌آید که آیا ضریب همبستگی $0/09$ نشان از همبستگی زیاد بین x و y است و یا خیر. برای آنکه بتوانیم به این سؤال پاسخ دهیم ضریب همبستگی را در برخی حالت‌های زیر حساب می‌کنیم:



x	۱	۲	۳	۴
y	۳	۵	۷	۹

x	۱	۲	۳	۴
y	۲	۶	۶	۱۰



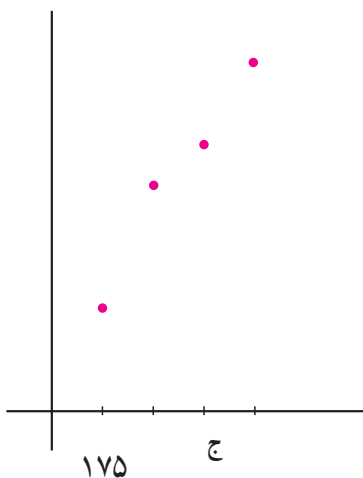
x	۱	۲	۳	۴
y	۲/۵	۵/۵	۶/۵	۸/۵

در شکل «الف» نقاط کاملاً روی خط قرار دارند پس باید بالاترین ضریب همبستگی را داشته باشند.

در شکل «ب» نقاط از خط دورتر شده‌اند پس باید ضریب همبستگی آن از ضریب همبستگی «الف» کمتر باشد.

در شکل «ج» نقاط از خط دور شده‌اند ولی نه به اندازه دوری حالت «ب»، پس باید ضریب همبستگی در این حالت از ضریب همبستگی در حالت «ب» بیشتر ولی از ضریب همبستگی حالت «الف» کمتر باشد.

حال با محاسبه صحت مطالب بالا را آزمایش می‌کنیم.



ضریب همبستگی در حالت «الف»

$$\bar{x} = \frac{10}{4} = 2.5, \quad \bar{y} = \frac{24}{4} = 6$$

$$x \text{ واریانس} = \frac{2/25 + 0/25 + 0/25 + 2/25}{4} = \frac{5}{4} = 1.25$$

$$y \text{ واریانس} = \frac{9 + 1 + 1 + 9}{4} = \frac{20}{4} = 5$$

$$\text{مجموع حاصلضرب } x \text{ ها و } y \text{ ها} = 3 + 10 + 21 + 36 = 70$$

$$\text{ضریب همبستگی} = \frac{\frac{70}{4} - (2.5)(6)}{\sqrt{1.25} \times \sqrt{5}} = \frac{17.5 - 15}{\sqrt{6.25}} = \frac{2.5}{2.5} = 1$$

ضریب همبستگی در حالت «ب»

$$\bar{x} = 2.5, \quad \bar{y} = \frac{24}{4} = 6$$

$$x \text{ واریانس} = 1.25,$$

$$y \text{ واریانس} = \frac{16 + 0 + 0 + 16}{4} = 8$$

$$\text{مجموع حاصلضرب } x \text{ ها و } y \text{ ها} = 2 + 12 + 18 + 40 = 72$$

$$\text{ضریب همبستگی} = \frac{\frac{72}{4} - (2.5)(6)}{\sqrt{1.25} \times \sqrt{8}} = \frac{3}{\sqrt{10}} \approx 0.9486$$

ضریب همبستگی در حالت «ج»

$$\bar{x} = 2.5, \quad x \text{ واریانس} = 1.25$$

$$y \text{ میانگین} = \frac{23}{4} = 5.75$$

$$\begin{aligned} y \text{ واریانس} &= \frac{(3/25)^2 + (0/25)^2 + (0/75)^2 + (2/75)^2}{4} \\ &= \frac{18/75}{4} = 4/6875 \end{aligned}$$

$$\frac{\text{مجموع حاصلضرب } x \text{ ها و } y \text{ ها}}{4} = \frac{2.5 + 11 + 19.5 + 34}{4} = \frac{67}{4} = 16.75$$

$$\text{ضریب همبستگی} = \frac{16.75 - (2.5)(5.75)}{\sqrt{1.25} \sqrt{4/6875}} = \frac{2/375}{\sqrt{5/8858}} \approx 0.9812$$

با مقایسه محاسبات در حالت های مختلف ملاحظه می شود که نتیجه محاسبات با انتظاری که از ضریب همبستگی داریم هماهنگی دارد.

البته ما هنوز به سؤالی که بعد از مثال از خود پرسیده بودیم پاسخ نداده ایم. سؤال این بود که ضریب همبستگی به اندازه 0.9 چه معنایی دارد. آیا این نشان از همبستگی زیاد است یا همبستگی کم. ما در این قسمت وارد تحلیل اندازه ضریب همبستگی نمی شویم فقط اضافه می کنیم که در عرف محاسبات آماری 0.9 ضریب همبستگی بالایی نیست.

ما تا بحال ضریب همبستگی هایی که محاسبه کردیم مثبت بودند، ممکن است سؤال شود آیا ممکن است ضریب همبستگی منفی هم باشد؟ پاسخ به این سؤال مثبت است. داده های زیر را در نظر بگیرید.

x	۱	۲	۳	۴	۵
y	۴	۳	۲	۱	۰

ضریب همبستگی این داده ها را محاسبه می کنیم.

البته با رسم نمودار پراکنش متوجه می شویم که این داده ها روی خط $y = -x + 5$ قرار دارند.

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3$$

$$\bar{y} = \frac{4+3+2+1}{5} = 2$$

$$x \text{ واریانس} = \frac{4+1+0+1+4}{5} = 2$$

$$y \text{ واریانس} = \frac{4+1+0+1+4}{5} = 2$$

$$\frac{\text{مجموع حاصلضرب } x \text{ ها و } y \text{ ها}}{5} = \frac{4+6+6+4+0}{5} = \frac{20}{5} = 4$$

$$\text{ضریب همبستگی} = \frac{4 - (3)(2)}{\sqrt{2} \times \sqrt{2}} = \frac{-2}{2} = -1$$



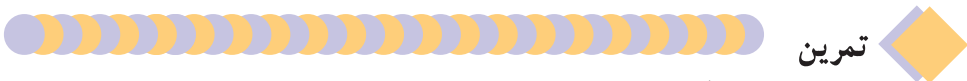
حتماً متوجه شده‌اید در شرایطی که تعیین y با کمترین خطا انجام شد (مقادیر y و x روی یک خط قرار دارند) ضریب همبستگی ۱ یا -۱ به دست آمد، در سایر شرایط ضریب همبستگی بین این دو مقدار است. در این مرحله ما قادر به اثبات ویژگی‌های ضریب همبستگی نیستیم فقط به ذکر آن‌ها می‌پردازیم.

– ضریب همبستگی همواره بین -۱ و +۱ قرار دارد.

– اگر مقادیر x و y روی خطی با شیب زاویه مثبت باشد، ضریب همبستگی x و y برابر +۱ است.

– اگر مقادیر x و y روی خطی با ضریب منفی زاویه قرار گیرد، ضریب همبستگی x و y برابر -۱ است.

- اگر ضریب همبستگی مثبت باشد با افزایش x ، مقادیر y نیز عموماً افزایش می‌یابد و بالعکس.
- اگر ضریب همبستگی منفی باشد با افزایش x ، مقادیر y عموماً کاهش می‌یابد و بالعکس.
- هر چقدر ضریب همبستگی به ۰ نزدیکتر باشد تعیین y با خطای بیشتری انجام می‌شود.
- هر چقدر ضریب همبستگی به ۱ یا -۱ نزدیکتر باشد تعیین y با خطای کمتری انجام می‌شود.



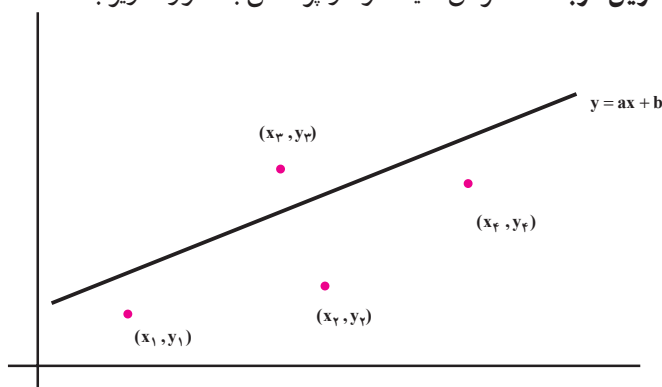
تمرین

- ۱ – ضریب همبستگی قد و وزن همکلاسی‌های خود را حساب کنید.
- ۲ – ضریب همبستگی بین نمره درس آمار و مدلسازی و نمره یکی دیگر از درس‌ها را در کلاس حساب کنید.
- ۳ – ضریب همبستگی بین معدل دانش‌آموزان و تعداد افراد خانواده آن‌ها را حساب کنید.
- ۴ – گمان می‌کنید ضریب همبستگی بین کدام دو درس در کلاس بیش از همه است؟ درستی گمان خود را چگونه تحقیق می‌کنید؟
- ۵ – در تمرین‌های ۱ تا ۴ به سؤال‌های زیر پاسخ دهید:
 - آیا می‌توانید علامت ضریب همبستگی را توجیه کنید؟
 - در کدام یک از موارد ضریب همبستگی به ۰ نزدیکتر است؟
 - آیا نتایج حاصل از تحلیل ضریب همبستگی‌ها معقول به نظر می‌رسد؟



خط رگرسیون

تا بحال صحبت از این بود که هر چقدر نقاط به خط نزدیک تر باشند، ضریب همبستگی از صفر دورتر است و به ۱ یا -۱ نزدیک تر خواهد بود. ما اگر معادله خطی را که از آن صحبت می کنیم داشته باشیم و بدانیم ضریب همبستگی هم در حد قابل قبولی است آن گاه می توانیم با داشتن x مقدار y را از روی خط تعیین کنیم و تا «حدودی» مطمئن باشیم که خطای زیادی مرتکب نشده ایم. اما در مسائل عملی ما این خط را نداریم، در این مبحث می خواهیم روش یافتن این خط را بیان کنیم. روش کمترین مربعات: فرض کنید نمودار پراکنش به صورت زیر باشد.



می خواهیم خطی به معادله $y = ax + b$ از بین این نقاط بگذرانیم به قسمتی که به همه نقاط «نزدیک» باشد. ما ذیلاً توضیح می دهیم که نزدیکی این خط به همه نقاط به چه معنی است؟ ما از این خط می خواهیم برای به دست آوردن y ها استفاده کنیم، مثلاً به جای y_1 می خواهیم از مقدار $ax_1 + b$ که از روی خط حاصل شده است استفاده کنیم، در این صورت مقدار خطایی که در این جابجایی مرتکب شده ایم عبارت است از $(ax + b - y_1)$ و همین طور برای سایر نقاط می توانیم

این خطاها را حساب کنیم. از آنجایی که تعداد این خطاها ممکن است مثبت و یا منفی باشد، با به توان دو رساندن آن‌ها همگی را مثبت خواهیم کرد، در واقع می‌خواهیم خط را چنان تعیین کنیم تا مجموع مجذور خطاها می‌نیم شود، یعنی می‌خواهیم a و b را چنان به دست آوریم که عبارت:

$$(ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + (ax_3 + b - y_3)^2 + \dots$$

کمترین مقدار ممکن را بگیرد. این روش را روش کمترین مربعات می‌گویند و خط حاصل را خط رگرسیون y نسبت به x می‌گویند. ما بدون آنکه وارد روش‌های ریاضی محاسبه a و b شویم فقط به ذکر دستور محاسبه آن‌ها می‌پردازیم.

محاسبه a و b به روش کمترین مربعات

$$a = r \frac{\sigma_y}{\sigma_x} = \text{ضریب همبستگی} \times \frac{\text{انحراف معیار } y}{\text{انحراف معیار } x}$$

$$b = \bar{y} - a\bar{x}$$

حال خط رگرسیون y نسبت به x را در مسأله وزن موش‌ها به دست می‌آوریم. قبلاً دیدیم که:

$$\bar{x} = 1/6, \quad \bar{y} = 5/4$$

$$x \text{ انحراف معیار} = \sqrt{0/6}, \quad y \text{ انحراف معیار} = \sqrt{3/44}$$

$$r = 0/11$$

$$a = 0/11 \frac{\sqrt{3/44}}{\sqrt{0/6}} \approx 0/263$$

بنابراین:

$$b = 5/4 - 1/6 = 4/97$$

پس معادله خط رگرسیون عبارت است از:

$$y = 0/263x + 4/97$$

حال اگر وزن اولیه موشی $1/5$ باشد، پس از اعمال رژیم غذایی وزن آن حدوداً چقدر خواهد بود؟ برای پاسخ به این سؤال از خط رگرسیون y نسبت به x استفاده می‌کنیم داریم $x = 1/5$ و پس

$$y = (0/263)(1/5) + 4/97 = 5/36$$

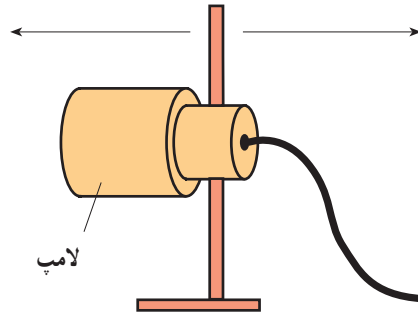
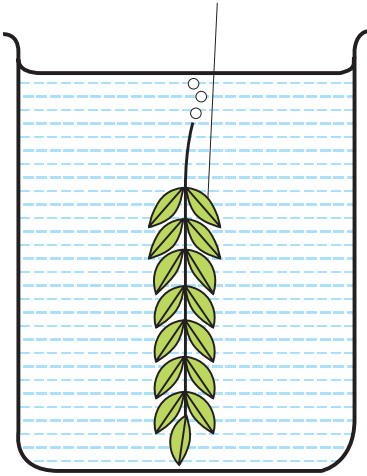
خط رگرسیون ابزاری است برای پیش‌بینی مقدار یک متغیر برحسب

متغیری که به آن وابسته است.

تمرین

۱- مهرناز آزمایشی را تدارک دیده است که اثر نور را بر سرعت تهیه غذا توسط گیاه از طریق

فتوسنتز بررسی می کند.



مهرناز لامپ را در ۱۰۰ متری گیاه قرار داده و تعداد حباب‌هایی را که توسط گیاه در یک دقیقه تولید شد، شمرد. سپس لامپ را نزدیک‌تر کرده و تعداد حباب‌های ایجاد شده توسط گیاه در دقیقه را شمرد و جدول زیر را تهیه نمود.

فاصله لامپ از گیاه بر حسب cm	تعداد حباب‌ها در دقیقه
۱۰۰	۱۰
۸۰	۲۰
۶۰	۳۲
۴۰	۳۷
۲۰	۳۷

– نمودار پراکنش این داده‌ها را رسم کنید.

– بدون محاسبه و به‌طور تقریبی، خط رگرسیون را رسم کنید.

– اگر لامپ را به ۱۰ سانتی‌متری گیاه منتقل کنیم، تعداد حباب‌ها در دقیقه را حدس بزنید.

– با استفاده از نتیجه حاصل شده جمله زیر را کامل کنید:

همان‌طور که فاصله لامپ، سرعت فتوسنتز

۲- نمودار پراکنش نتایج امتحانات درس ریاضی و درس فیزیک ۲۰ دانش آموز که به شرح زیر آمده است را رسم کنید.

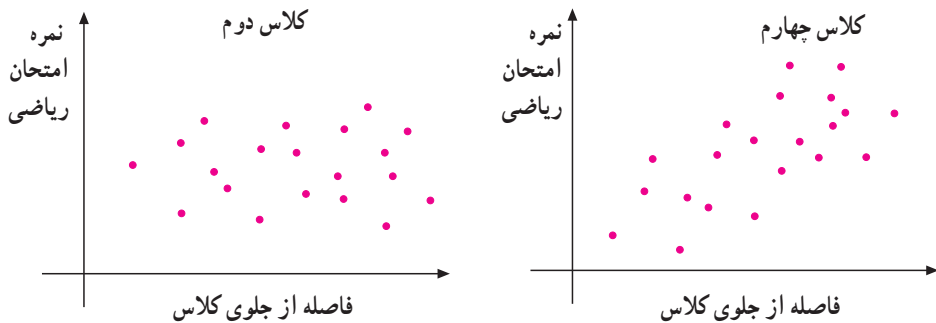
دانش آموز	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
نمره ریاضیات	۱۸	۸	۱۲	۱۸	۱۶	۶	۱۰	۶	۱۶	۱۰	۱۲
نمره فیزیک	۱۶	۱۰	۸	۱۸	۱۴	۱۰	۱۰	۶	۱۰	۱۴	۱۰

دانش آموز	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
نمره ریاضیات	۱۲	۱۰	۱۶	۱۸	۱۲	۶	۱۴	۱۴	۱۲
نمره فیزیک	۱۶	۱۲	۱۶	۱۲	۱۲	۸	۱۲	۱۴	۱۶

– معادله خط رگرسیون را به دست آورید.

– آیا ارتباطی بین نتایج دو تست وجود دارد؟

۳- برخی از دانش آموزان تصور می کنند هرکس که در ردیف های جلوی کلاس می نشیند، نمرات بهتری در درس ریاضی کسب می کند. یکی از معلمین تصمیم گرفت این موضوع را در دو کلاس خود مورد بررسی قرار دهد. نمودارهای زیر نتیجه این بررسی است.



درستی نظر دانش آموزان را در دو کلاس مورد بررسی قرار دهید. دو نمودار فوق را تفسیر کنید.

۴- آشپزی زمان لازم برای پختن گوشت غذا به وزن های مختلف را در جدول صفحه بعد ثبت نموده است.

- روی محور افقی با در نظر گرفتن یک سانتی متر نشان گر یک کیلوگرم و روی محور عمودی با در نظر گرفتن هر سانتی متر نشان گر ۳۰ دقیقه، نمودار پراکنش را برای نشان دادن ارتباط بین وزن گوشت و زمان لازم برای پختن آن رسم کنید.
- خط رگرسیون را رسم کنید.
- با استفاده از خط رگرسیون زمان لازم را برای پختن ۳ کیلوگوشت تخمین بزنید.

وزن گوشت (kg)	زمان بر حسب دقیقه
۲	۱۳۰
۳/۵	۲۰۰
۵	۲۱۵
۵/۵	۲۸۰
۶/۵	۲۷۰

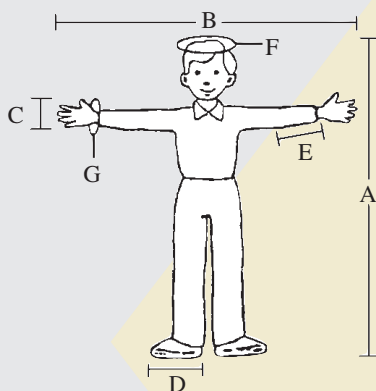
- ۵- خانم خانه‌داری از فرزندش که درس آمار خوانده بود خواست تا روی اندازه سیب‌زمینی‌های داخل کیسه‌های ۲ کیلویی بررسی انجام دهد. او هر یک از سیب‌زمینی‌ها را وزن کرده و جدول زیر را تهیه نمود:

تعداد سیب‌زمینی‌های موجود در کیسه							
۱۰	۱۳	۱۵	۱۴	۱۴	۱۸	۱۶	۲۰
میانگین وزن سیب‌زمینی‌ها بر حسب گرم							
۱۹۲	۱۷۴	۱۴۲	۱۳۶	۱۲۵	۱۱۰	۱۰۵	۷۲

- نمودار پراکنش داده‌ها را رسم کنید (۱ سانتی متر روی محور xها نشان دهنده یک سیب‌زمینی و ۱ سانتی متر روی محور yها نشان دهنده ۲۰ گرم)
- خط رگرسیون را رسم کنید.
- ارتباط بین تعداد سیب‌زمینی‌های موجود در یک بسته را با میانگین وزن بیان کنید.
- با استفاده از نمودار، میانگین وزن سیب‌زمینی‌ها را در بسته‌ای که ۱۷ سیب‌زمینی وجود دارد، تخمین بزنید.
- در یک بسته ۲ کیلوگرمی اگر میانگین وزن ۱۶۰ گرم باشد، تعداد سیب‌زمینی‌های بسته را (به نزدیک‌ترین عدد صحیح) تخمین بزنید.

پروژه ۱

در اندازه‌گیری زیر، کدامیک از اندازه‌ها با هم مرتبط هستند.



A : قد

B : فاصله بین دو دست

C : وجب

D : طول پا

E : میج تا آرنج

F : دور سر

G : دور میج دست

— اندازه‌های فوق را برای چند تن از همکلاسی‌هایتان به‌دست آورید.

— با استفاده از اندازه‌های به‌دست آمده معین کنید آیا جفت متغیری وجود دارد که

اندازه‌شان با هم مرتبط باشند.

اگر چنین است، خط رگرسیون را برایشان رسم کنید. گزارشی تهیه کنید و روی نتیجه

نظر بدهید. آیا با آنچه قبل از شروع پروژه حدس زدید هماهنگی دارد یا خیر؟ اگر داده‌ها را از

دو گروه زنان و مردان به‌دست آورید چه اتفاقی می‌افتد؟



پروژه ۲

می خواهیم بدانیم آیا ارتباطی بین موفقیت محصل در دروس ریاضی و دروس غیر ریاضی وجود دارد؟ برای این منظور یکی از دروس ریاضی و یکی از دروس غیر ریاضی خود را انتخاب کنید و با نمونه مناسبی از کلاس درس وجود نوع ارتباط را تحقیق کنید. ضریب تغییرات در کدام یک از این دروس بیش تر است. در صورت وجود ارتباط خطی معادله خط کم ترین مربعات را به دست آورید.

– نتایج کار خود را با نتایج به دست آمده از تحقیق گروه های دیگر مقایسه کنید. آیا نتایج به دست آمده یکسان است؟

منم که دیده به دیدار دوست کردم باز ...

پروژه

آیا بین اوقات فراغت دانش آموزان و عملکرد آن‌ها در مدرسه ارتباطی وجود دارد؟

معیارهای ارزشیابی پروژه

هر پروژه با ۶ معیار ارزشیابی می‌شود:

معیار ۱- تعریف و طراحی پروژه: در این قسمت پروژه باید کاملاً تعریف شده باشد، طرحی برای اجرا و به پایان رسانیدن پروژه ارائه شده باشد.

۳- امتیاز

به‌طور مثال در پروژه ضمیمه این کار در قسمت خلاصه و مقدمه انجام شده است. پروژه کاملاً تعریف شده و شمای کلی کار ارائه شده است.

معیار ۲- جمع‌آوری داده‌ها: در این قسمت روش جمع‌آوری داده‌ها، ارتباط داده‌های جمع‌آوری شده با موضوع پروژه، سازمان‌دهی داده‌های جمع‌آوری شده و کافی بودن داده‌ها از نظر کمی و کیفی مورد نظر است.

۵- امتیاز

معیار ۳- تجزیه و تحلیل داده‌ها: در این قسمت تجزیه و تحلیل داده‌ها از دو جنبه محاسبات و ارائه نموداری مورد ارزشیابی قرار می‌گیرد.

استفاده مناسب از نمودارها، دقت در محاسبات عددی و استفاده از تکنیک‌های مختلف

تجزیه و تحلیل داده‌ها مورد نظر می‌باشد.

۶- امتیاز

معیار ۴- ارزیابی: در این قسمت نتیجه‌گیری و ارزیابی نتایج حاصل از جمع‌آوری و تجزیه و تحلیل داده‌ها و نیز نقد پروژه توسط خود دانش‌آموز مورد توجه است.

۶- امتیاز

معیار ۵- برقراری ارتباط: چگونگی ارائه پروژه و استفاده صحیح از نمادها مورد بررسی می‌باشد.

۵- امتیاز

معیار ۶- همکاری و ارائه شفاهی پروژه در کلاس

۱۵- امتیاز

خلاصه

اغلب، دانش‌آموزان فکر می‌کنند، به دلیل آن که استراحت کافی ندارند، نمی‌توانند تلاش بیشتری در مطالعه داشته باشند. برخی فکر می‌کنند چرا همکلاسی‌هایشان نمرات بهتری کسب می‌کنند ولی آن‌ها نمی‌توانند نمره خوبی کسب کنند. برخی فکر می‌کنند اگر اوقات فراغت بیشتری داشتند می‌توانستند عمل کرد بهتری داشته باشند. سؤال این است که آیا ارتباطی بین اوقات فراغت دانش‌آموزان و عمل کرد آن‌ها در مدرسه وجود دارد؟ این سؤال از طریق این پروژه مورد بررسی قرار گرفت. نمونه آماری در این پروژه، دانش‌آموزانی هستند که به‌طور تصادفی از میان دانش‌آموزان مدارس بین‌الملل دختران و پسران تهران انتخاب شدند. با استفاده از محاسبات آماری، این پروژه به پایان رسید. نتیجه نشان داد که این ارتباط وجود ندارد.

فهرست محتوا

مقدمه	نتیجه محاسبات
جمع‌آوری داده‌ها	نتیجه‌گیری نهایی
تجزیه و تحلیل داده‌ها	ملحقات
محاسبات آماری	A : پرسش‌نامه
نمودارها	B : جداول
نتیجه‌گیری	C : محاسبه وقت بیکاری
بررسی مشکلات	منابع و مآخذ

هدف این پروژه، بررسی ارتباط بین اوقات فراغت و عمل کرد دانش‌آموزان در مدرسه می‌باشد. برای یک دانش‌آموز کاملاً عادی است که بشنود: «چرا وقت خودت را تلف می‌کنی؟» «چرا بیش‌تر درس نمی‌خوانی؟» و یا نصیحت‌هایی چون «بیش‌تر درس بخوان».

بیش‌تر دانش‌آموزان دوست ندارند نصیحت‌های خانواده‌هایشان را گوش کنند ولی آیا این دانش‌آموزان به اندازه کافی احساس مسئولیت می‌کنند و از وقت خود استفاده درست و لازم را دارند؟ آیا در برنامه‌ریزی برای اوقات خود معقول عمل می‌کنند؟ چقدر از زمان خود را صرف مطالعه می‌کنند و چقدر از آن را به‌هدر می‌دهند؟ ممکن است یک روز ۲ ساعت هدر شود، روز دیگر ۳ ساعت ... ولی در پایان هفته این ساعات باهم جمع می‌گردد. ساعتی که می‌شد بهره‌ی بیش‌تری از آن‌ها جست و عمل کرد بهتری داشت.

بیش‌تر خانواده‌ها می‌گویند: شما به اندازه کافی درس نمی‌خوانی و اگر بیش‌تر درس بخوانی نمره‌ی بهتری خواهی گرفت. آیا چنین است؟

این پروژه با توزیع پرسش‌نامه بین دانش‌آموزان آغاز گشت. در این رابطه داده‌ها جمع‌آوری شده و مورد ارزیابی و تجزیه و تحلیل قرار گرفتند. در این بررسی از محاسبات آماری، نمودارها و نیز محاسبه احتمالات شرطی استفاده شد.

جمع‌آوری داده‌ها

به دلیل مشکلات موجود برای جمع‌آوری داده در مدارس، دو مدرسه بین‌الملل تهران به‌عنوان جامعه آماری در نظر گرفته شد و ۴۰ دانش‌آموز به‌طور تصادفی از بین دانش‌آموزان این دو مدرسه به‌عنوان نمونه انتخاب شدند. دانش‌آموزان از کلاس‌های اول دبیرستان تا پیش‌دانشگاهی پسر و دختر، انتخاب شدند.

جمع‌آوری داده‌ها می‌توانست از دو طریق مصاحبه و پرسش‌نامه انجام شود ولی روش پرسش‌نامه به‌نظر مناسب‌تر آمد. در این نظرخواهی، اسامی خواسته نشده بود تا دانش‌آموزان بتوانند با راحتی خیال پاسخ‌های خود را بدهند. از طریق مصاحبه ممکن بود دانش‌آموزان پاسخ‌های درست ندهند و نیز زمان بیش‌تری مورد نیاز بود. در پرسش‌نامه از طول زمان و چگونگی گذراندن اوقات فراغت دانش‌آموزان سؤال شد و این که چه مقدار از این زمان فراغت صرف مطالعه می‌شود و معدل دانش‌آموزان چقدر است؟ یک نسخه از پرسش‌نامه و برخی جداول حاوی اطلاعات ضمیمه این گزارش است.

روند محاسبه اوقات بیکاری نیز در ضمیمه C آمده است. جداول ۱ و ۲ و ۳ و ۴ برخی از داده‌های جمع‌آوری شده را نشان می‌دهند.

جدول ۲: زمان صرف شده برای مطالعه بعد از مدرسه در اوقات فراغت (در هفته)

دانش آموزان شماره	زمان برحسب ساعت در هفته
۱	۰
۲	۵
۳	۲۸
۴	۶
۵	۶
۶	۵-۶
۷	۳
۸	۳
۹	۴
۱۰	۱۴
۱۱	۱۴-۱۶
۱۲	۴۵
۱۳	۵
۱۴	۸
۱۵	۱۲
۱۶	۱۰
۱۷	۳
۱۸	۷-۸
۱۹	۲۸
۲۰	۱۲-۱۵
۲۱	۲
۲۲	۲
۲۳	۲
۲۴	۲
۲۵	۲
۲۶	۱۴
۲۷	۴
۲۸	۴
۲۹	۸
۳۰	۱۰
۳۱	۱۰
۳۲	۱۲

جدول ۱: زمان صرف شده برای فعالیت‌های غیر از درس خواندن در هفته

دانش آموزان شماره	زمان برحسب ساعت در هفته
۱	۹
۲	۷
۳	۱۳
۴	۶
۵	۱۸
۶	۲/۵
۷	۴
۸	۴/۵
۹	۴۹
۱۰	۳۲
۱۱	۱۷
۱۲	۱۷
۱۳	۳
۱۴	۵
۱۵	۶
۱۶	۷۶
۱۷	۹
۱۸	۲
۱۹	۷
۲۰	۲۲
۲۱	۳۰
۲۲	۳۵
۲۳	۳۰
۲۴	۲۴
۲۵	۴۵
۲۶	۳۰
۲۷	۷۱
۲۸	۲۰
۲۹	۸
۳۰	۱
۳۱	۶
۳۲	۷

جدول ۴: زمان فراغت دانش آموزان

شماره دانش آموزان	ساعات فراغت
۱	۵۰/۷۵
۲	۳۵
۳	۴۲
۴	۳۹
۵	۵۰
۶	۳۳/۸۳
۷	۲۱
۸	۱۵
۹	۲۸
۱۰	۳۹/۵
۱۱	۳۹/۵
۱۲	۵۳/۵
۱۳	۶۳/۱۶
۱۴	۵۶/۵
۱۵	۵۹/۵
۱۶	۵۶/۸۳
۱۷	۳۴
۱۸	۴۶
۱۹	۳۷/۵
۲۰	۴۳
۲۱	۳۵/۵۸
۲۲	۳۸/۲۵
۲۳	۲۸/۵۸
۲۴	۵۲
۲۵	۲۵/۵
۲۶	۲۲/۵
۲۷	۳۸
۲۸	۲۸/۱۶
۲۹	۳۲/۵
۳۰	۳۱/۷۵
۳۱	۳۲/۵
۳۲	۲۵/۵

جدول ۳: معدل آخرین سال دانش آموزان

شماره دانش آموزان	معدل برحسب درصد
۱	۹۵
۲	۹۶/۷۵
۳	۹۷
۴	۸۸
۵	۸۶
۶	۸۳
۷	۹۴/۵
۸	۸۸/۶۳
۹	۷۰
۱۰	۸۶
۱۱	۸۹/۸
۱۲	۷۸/۵
۱۳	۷۱
۱۴	۸۶
۱۵	۷۸/۵
۱۶	۸۵
۱۷	۸۰
۱۸	۸۳
۱۹	۸۵
۲۰	۹۳
۲۱	۹۳
۲۲	۹۳
۲۳	۵۹/۵
۲۴	۶۵/۴
۲۵	۷۳/۵
۲۶	۷۸/۵
۲۷	۷۳/۵
۲۸	۷۳/۶
۲۹	۷۸
۳۰	۸۰
۳۱	۷۸
۳۲	۷۰

تجزیه و تحلیل داده‌ها

این قسمت از دو بخش محاسبات و نمودارها تشکیل شده است: ابتدا میانگین و انحراف معیار برای زمان فراغت و معدل هریک از دانش‌آموزان حساب شد. نمودارهای میله‌ای نیز رسم شدند.

صفحات ۱۹۲، ۱۹۳ و ۱۹۴.

ارزیابی مشکلات

در انجام این پروژه متأسفانه با مشکلاتی روبرو بودم که عبارتند از:

- به دلیل مشکل عدم آشنایی به زبان فارسی این بررسی محدود به مدارس بین‌الملل تهران شد.
- در این بررسی ۸ پاسخ‌نامه به دلیل مخدوش بودن پاسخ‌ها کنار گذاشته شد. بنابراین نمونه به ۳۲ کاهش یافت.
- محدودیت زمان

نتایج

قبل از آن که هرگونه نتیجه‌گیری شود، نتایج مشاهدات به قرار زیر ارائه می‌شوند.

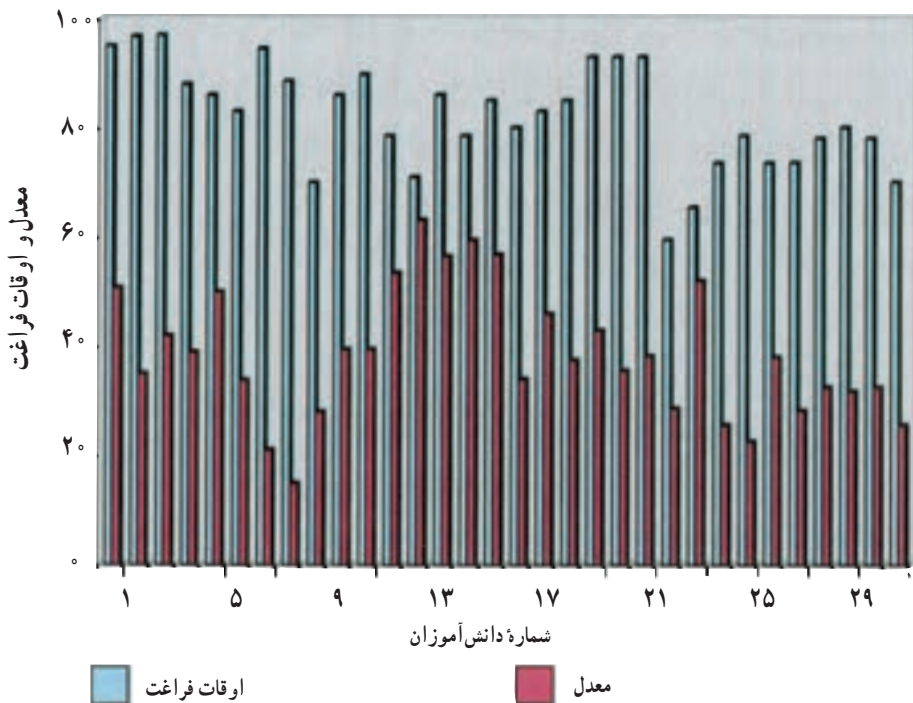
- محاسبات آماری روی معدل نشان دادند که ۷ نفر از ۳۲ نفر دانش‌آموزان عمل کرد بسیار عالی داشتند. ۱۱ نفر عمل کرد خوب و ۹ نفر عمل کرد قابل قبول، ۴ نفر ضعیف و یک نفر خیلی ضعیف بودند.
- محاسبات روی اوقات فراغت نشان دادند که ۶ نفر وقت آزاد بسیار داشتند و ۷ نفر بیش‌تر از کافی، ۱۴ نفر وقت آزاد کافی داشتند. ۴ نفر وقت کافی نداشتند و یک نفر اصلاً وقت آزاد نداشت.
- اطلاعات اضافی عبارتند از ۲۴ نفر از ۳۲ نفر از اوقات بیکاری خود خسته می‌شوند.
- ۹ نفر از ۳۱ نفر حداقل یک‌بار در طول تحصیل‌شان از یک درس نمره قبولی نیاورده‌اند.
- ۲۶ نفر از ۳۰ نفر از عمل کرد خود راضی نیستند و می‌خواهند آن را بهبود دهند.
- ۲۴ نفر از ۳۲ نفر عقیده داشتند که اگر وقت بیش‌تری صرف مطالعه کنند نتیجه بهتری خواهند داشت.

جدول ۱- جدول فراوانی اوقات بیکاری

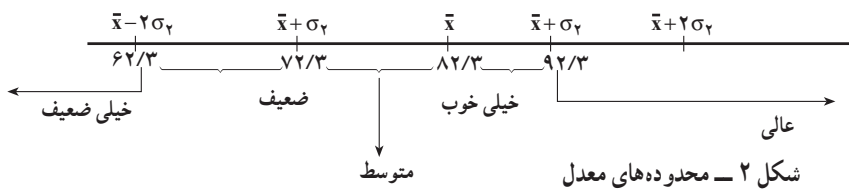
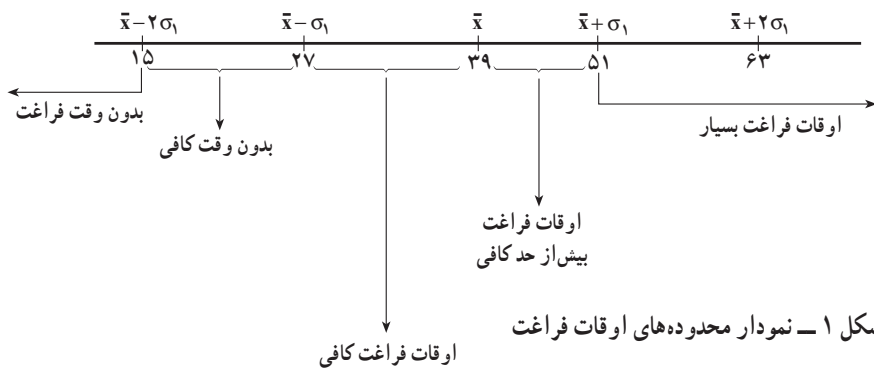
زمان فراغت	فراوانی	فراوانی × زمان فراغت	انحراف از میانگین	مجذور انحراف از میانگین	فراوانی × مجذور انحراف از میانگین
۱۵	۱	۱۵	-۲۴	۵۷۶	۵۷۶
۲۱	۱	۲۱	-۱۵	۲۲۵	۲۲۵
۲۲/۵	۱	۲۲/۵	-۱۶/۵	۲۷۲/۲۵	۲۷۲/۲۵
۲۵/۵	۲	۵۱	-۱۳/۵	۱۸۲/۲۵	۳۶۴/۵
۲۸	۱	۲۸	-۱۱	۱۲۱	۱۲۱
۲۸/۱۶	۱	۲۸/۱۶	-۱۰/۸۴	۱۱۷/۵۰۵۶	۱۱۷/۵۰۵۶
۲۸/۵۸	۱	۲۸/۵۸	-۱۰/۴۲	۱۰۸/۵۷۶۴	۱۰۸/۵۷۶۴
۳۱/۷۵	۱	۳۱/۷۵	-۷/۲۵	۵۲/۵۶۲۵	۵۲/۵۶۲۵
۳۲/۵	۲	۶۵	-۶/۵	۴۲/۲۵	۸۴/۵
۳۳/۸۳	۱	۳۳/۸۳	-۵/۱۷	۲۶/۷۲۸۹	۲۶/۷۲۸۹
۳۴	۱	۳۴	-۵	۲۵	۲۵
۳۵	۱	۳۵	-۴	۱۶	۱۶
۳۵/۵۸	۱	۳۵/۵۸	-۳/۴۲	۱۱/۶۹۶۴	۱۱/۶۹۶۴
۳۷/۵	۱	۳۷/۵	-۱/۵	۲/۲۵	۲/۲۵
۳۸	۱	۳۸	-۱	۱	۱
۳۸/۲۵	۱	۳۸/۲۵	-۰/۷۵	۰/۵۶۲۵	۰/۵۶۲۵
۳۹	۱	۳۹	۰	۰	۰
۳۹/۵	۲	۷۹	۰/۵	۰/۲۵	۰/۵
۴۲	۱	۴۲	۳	۹	۹
۴۳	۱	۴۳	۴	۱۶	۱۶
۴۶	۱	۴۶	۷	۴۹	۴۹
۵۰	۱	۵۰	۱۱	۱۲۱	۱۲۱
۵۰/۷۵	۱	۵۷/۷۵	۱۱/۷۵	۱۳۸/۰۶۲۵	۱۳۸/۰۶۲۵
۵۲	۱	۵۲	۱۳	۱۶۹	۱۶۹
۵۳/۵	۱	۵۳/۵	۱۴/۵	۲۱۰/۲۵	۲۱۰/۲۵
۵۶/۵	۱	۵۶/۵	۱۷/۵	۳۰۶/۲۵	۳۰۶/۲۵
۵۶/۸۳	۱	۵۶/۸۳	۱۷/۸۳	۳۱۷/۹۰۸۹	۳۱۷/۹۰۸۹
۵۹/۵	۱	۵۹/۵	۲۰/۵	۴۲۰/۲۵	۴۲۰/۲۵
۶۳/۱۶	۱	۶۳/۱۶	۲۴/۱۶	۵۸۳/۷۰۵۶	۵۸۳/۷۰۵۶
	۳۲	۱۲۳۴/۳۹			۴۳۴۶/۰۵۹۳

جدول ۲- جدول فراوانی برای معدل

معدل	فراوانی	معدل × فراوانی	انحراف از میانگین	مجذور انحراف از میانگین	مجذور انحراف از میانگین × فراوانی
۶۰	۱	۶۰	-۲۳	۵۲۹	۵۲۹
۶۵	۱	۶۵	-۱۸	۳۲۴	۳۲۴
۷۰	۲	۱۴۰	-۱۳	۱۶۹	۳۳۸
۷۱	۱	۷۱	-۱۲	۱۴۴	۱۴۴
۷۴	۳	۲۲۲	-۹	۸۱	۲۴۳
۷۶	۱	۷۶	-۷	۴۹	۴۹
۷۸	۲	۱۵۶	-۵	۲۵	۵۰
۷۹	۳	۲۳۷	-۴	۱۶	۴۸
۸۰	۲	۱۶۰	-۳	۹	۱۸
۸۳	۲	۱۶۶	۰	۰	۰
۸۵	۲	۱۷۰	۲	۴	۸
۸۶	۳	۲۵۸	۳	۹	۲۷
۸۸	۱	۸۸	۵	۲۵	۲۵
۸۹	۱	۸۹	۶	۳۶	۳۶
۹۰	۱	۹۰	۷	۴۹	۴۹
۹۳	۳	۲۷۹	۱۰	۱۰۰	۳۰۰
۹۵	۳	۲۸۵	۱۲	۱۴۴	۴۳۲
۹۷	۲	۱۹۴	۱۴	۱۹۶	۳۹۲
	۳۴	۲۸۱۶			۳۰۱۲



ارتباط بین معدل و اوقات فراغت



نتیجه گیری

با توجه به نمودارها و محاسبات انجام شده نتیجه می شود که ارتباط محسوسی بین مقدار اوقات فراغت و عمل کرد در مدرسه وجود ندارد. معدل دانش آموز ارتباطی به میزان اوقات فراغت ندارد. دانش آموزان بسیاری وجود دارند که وقت فراغت بسیار دارند ولی عمل کرد بالایی ندارند.

ضمیمه A: «پرسش نامه»

این پرسش نامه شامل سؤالاتی در زمینه چگونگی گذراندن اوقات فراغت شما، زمانی که صرف فعالیت های غیر مدرسه ای و زمانی که صرف مطالعه می شود می باشد. هدف این بررسی، پیدا کردن ارتباط بین اوقات فراغت دانش آموزان و عمل کرد آنها در مدرسه است. این پرسش نامه ناشناس است. بنابراین، لطفاً صادقانه پاسخ دهید.

با تشکر از همکاری شما

۱- ساعت شروع مدرسه و پایان مدرسه را بیان کنید (لازم به ذکر است در این دو مدرسه ساعات درسی به دلیل انتخاب واحدها متغیر است)

شنبه یکشنبه دوشنبه سه شنبه چهارشنبه پنجشنبه

ساعت شروع مدرسه

ساعت پایان مدرسه

۲- چقدر زمان لازم است تا از مدرسه به خانه برگردید؟

۳- معمولاً ساعت چند می خوابید؟

۴- چند ساعت بعد از مدرسه در هفته کار می کنید؟

۵- چند ساعت در هفته به تماشای تلویزیون یا انجام دیگر تفریحات می پردازید؟

۶- معدل شما در سال گذشته چند است؟

۷- آیا از درسی مردود شده اید؟

در این ضمیمه اطلاعات مربوط به ۲۰ دانش آموز آورده شده است.
جدول ۱: ساعات شروع و پایان مدرسه در طول روزهای هفته
ضمیمه B:

شماره دانش آموزان	شنبه		یکشنبه		دوشنبه		سه شنبه		چهارشنبه		پنجشنبه	
	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان
۱	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۱۵ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۲	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۳	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۴	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۵	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۶	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۱:۱۰ ب.ظ.
۷	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۸	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۹	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۱۰	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۱۱	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.
۱۲	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۱:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.	۹:۳۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۳	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۱۱:۱۰ ظ.ق.	۱۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۴	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۵	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۶	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۷	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۸	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۱۹	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۲:۳۰ ب.ظ.
۲۰	۸:۰۰ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۹:۱۵ ظ.ق.	۱۲:۳۰ ب.ظ.	۸:۰۰ ظ.ق.	۲:۳۰ ب.ظ.	۹:۱۵ ظ.ق.	۴:۰۰ ب.ظ.	۸:۰۰ ظ.ق.	۱۱:۰۰ ب.ظ.

*- ق. ظ یعنی قبل از ظهر و ب. ظ یعنی بعد از ظهر

جدول ۳: زمان لازم برای رسیدن از
خانه به مدرسه

جدول ۴: ساعت خواب دانش آموزان

شماره دانش آموزان	زمان (ساعت)
۱	قبل از ظهر ۱۲،۳۰
۲	بعد از ظهر ۱۱،۰۰
۳	بعد از ظهر ۱۱،۰۰
۴	قبل از ظهر ۱۲،۰۰
۵	قبل از ظهر ۱،۳۰
۶	بعد از ظهر ۱۱،۰۰
۷	بعد از ظهر ۹،۰۰
۸	بعد از ظهر ۹،۰۰
۹	بعد از ظهر ۱۰،۰۰
۱۰	بعد از ظهر ۱۱،۰۰
۱۱	بعد از ظهر ۱۱،۰۰
۱۲	قبل از ظهر ۱،۳۰
۱۳	قبل از ظهر ۲،۰۰
۱۴	قبل از ظهر ۲،۰۰
۱۵	قبل از ظهر ۳،۰۰
۱۶	قبل از ظهر ۱،۰۰
۱۷	بعد از ظهر ۱۰،۰۰
۱۸	قبل از ظهر ۱۲،۰۰
۱۹	بعد از ظهر ۱۱،۰۰
۲۰	بعد از ظهر ۱۱،۳۰
۲۱	بعد از ظهر ۱۱،۰۰
۲۲	بعد از ظهر ۱۱،۰۰
۲۳	بعد از ظهر ۱۰،۰۰
۲۴	بعد از ظهر ۱۱،۰۰
۲۵	بعد از ظهر ۹،۳۰
۲۶	بعد از ظهر ۱۰،۰۰
۲۷	بعد از ظهر ۱۱،۰۰
۲۸	بعد از ظهر ۱۰،۳۰
۲۹	بعد از ظهر ۱۰،۰۰
۳۰	بعد از ظهر ۱۰،۰۰
۳۱	بعد از ظهر ۱۰،۰۰
۳۲	بعد از ظهر ۹،۰۰

شماره دانش آموزان	زمان (ساعت)
۱	۰/۳۰
۲	۰/۳۰
۳	۰/۳۰
۴	۳/۳۰
۵	۰/۳۰
۶	۰/۴۵
۷	۰/۳۰
۸	۱/۳۰
۹	۰/۳۰
۱۰	۱/۰
۱۱	۱/۰
۱۲	۰/۱۵
۱۳	۰/۳۰
۱۴	۰/۳۰
۱۵	۰/۳۰-۱/۰
۱۶	۰/۲۰
۱۷	۰/۲۰-۰/۳۰
۱۸	۰/۱۵-۰/۳۰
۱۹	۰/۰۵
۲۰	۰/۳۵
۲۱	۰/۳۰
۲۲	۰/۴۰
۲۳	۰/۳۰
۲۴	۰/۳۰
۲۵	۰/۳۰
۲۶	۱/۳۰
۲۷	۰/۴۵
۲۸	۱/۳۰
۲۹	۱/۰
۳۰	۰/۴۵
۳۱	۱/۰
۳۲	۱/۰

ضمیمه C: محاسبه اوقات فراغت

اوقات فراغت دانش‌آموزان به صورت زیر محاسبه شده است. طول روز ۱۲ ساعت در نظر گرفته شده است. ساعت در هفته $۷ \times ۱۲ = ۸۴$

محاسبه اوقات فراغت دانش‌آموز شماره ۱ بنابر جدول (ضمیمه B): تعداد ساعتی را که دانش‌آموزان در مدرسه صرف می‌کنند می‌توان پیدا کرد: به طور مثال:

$$۶/۵ + ۷/۲۵ + ۷/۵ + ۶/۵ + ۶/۵ = ۳۴/۲۵$$

ساعت در هفته ۳۴/۲۵
با استفاده از جدول ۲ (ضمیمه B)، زمان لازم برای رسیدن به منزل در طول هفته عبارت است از:

روز ساعت

$$۰/۵ \times ۵ = ۲/۵$$

دانش‌آموز شماره ۱ معمولاً ساعت نیم بعد از نیمه شب می‌خوابد پس طول روز برای او ۵ ساعت در روز بیش‌تر از فرض ۱۲ ساعت است پس $۰/۵ \times ۷ = ۳/۵$ باید به زمان فراغت اضافه شود:

$$[۳/۵ + (۳۴/۵ + ۲/۵) - ۸۴] + ۳/۵ = ۵۹/۷۵$$

وقت بیکاری دانش‌آموز شماره ۱

منابع و مراجع

۱- سالنامه آمار کشور، سالهای ۱۳۷۱ و ۱۳۷۳ و ۱۳۷۶، مرکز آمار ایران

- 2- DUNCAN AND CHRISTINE GRAHAM, MAINSTREAM MATHEMATICS FOR GCSE, MACMILLAN PRESS ltd.
- 3- GEOFF BUCKWELL. GCSE CORE MATHEMATICS, MCMILLAN WORKOUT SERIES.
- 4- ALLAN G. BLUMAN, ELEMENTARY STATISTICS, Wm .C. BROWN PUBLISHER.
- 5- PERM S. MANN, INTRODUCTORY STATISTICS, JOHN WILEY & SONS.
- 6- BRASE & BRASE, UNDERSTANDABLE STATISTICS.
- 7- GEOFF BUCKWELL, MASTERING MATHEMATICS, MCMILLAN MASTER SERIES.
- 8- HAROLD R. JACOBS, MATHEMATICS A HUMAN ENDEAVOR, FREEMAN.
- 9- JAN DE LANGE JZN, HELEEN VERHANGE, DATA VISUALIZATION.
- 10- EWART SMITH, MATHS FOR INTERMEDIATE GNVQ.
- 11- B.L. AGARWAL, BASIC STATISTICS.
- 12- JAMES M. LANDWEHR & ANN E. WATKINS EXPLORING DATA.
- 13- ROBERT JOHNSON, ELEMENTARY STATISTICS, SIXTH EDITION DUXBURY PRESS.

- 14- JOHN E. FREUND, GARY A. SIMON, STATISTICS A FIRST COURSE FIFTH EDITION. PRENTICE - HALL INTERNATIONAL EDITIOW.
- 15- THE SCHOOL MATEMATICS PROJECT. STATISTICS, MATHEMATICS FOR A AND AS LEVEL, CAMBRIDGE UNIVERSITY PRESS.
- 16- MATHEMATICS PLUS. HARCOURT BRACE & COMPANY.
- 17- HOUGHTON MIFFLIN, UNIFIED MATHEMATICS, HOUGHTON MIFFLIN COMPANY / BOSTON.
- 18- SUE BRIGGS, HIGHER GCSE, OXFORD MATHEMATICS, OXFORD UNIVERSITY PRESS.

تعدادی از تصاویر از مجموعه مجلات NATIONAL GEOGRAPHIC انتخاب شده اند.



