

درس ۳

چرخه آمار در حل مسائل

دنیای امروز پر از اطلاعاتی است که ما در بخش‌های مختلف زندگی با آنها روبه‌رو می‌شویم. بنابراین، برای اینکه بتوانیم بهتر زندگی کنیم و دربارهٔ مسائل مختلف بهتر تصمیم بگیریم، لازم است بتوانیم این اطلاعات را به خوبی درک و تحلیل کنیم. درک صحیح شاخص‌ها، نمودارها، اصطلاحات و مفاهیم آماری به ما کمک می‌کند که تفسیر درستی از این اطلاعات داشته باشیم، محدودیت‌های نتایج به‌دست آمده را بشناسیم و در نهایت، با استدلالی درست، بهتر تصمیم بگیریم. در این درس می‌خواهیم از شکل‌گیری یک مسئله در محیط زندگی‌مان تا یافتن پاسخی برای آن، از آمار استفاده کنیم و آماری فکر کردن را تمرین کنیم.

فعالیت

❓ دانش‌آموزان یک مدرسه، روزانه زمان زیادی را به استفاده از تلفن‌های هوشمند اختصاص می‌دادند. مشاوران آموزشی مدرسه تصمیم گرفتند دسترسی دانش‌آموزان به تلفن‌های هوشمند را در جهت اهداف آموزشی مدرسه هدفمند کنند اما از میزان و چگونگی استفاده دانش‌آموزان از تلفن‌های هوشمند اطلاع دقیقی نداشتند و بنابراین، برنامه‌ریزی برای این هدف ممکن نبود. پس، قبل از هر چیز مسئله مشاوران آموزشی این بود که بفهمند: «دانش‌آموزان پایه‌های دهم، یازدهم و دوازدهم این مدرسه به‌طور متوسط روزانه چه مدت و برای چه فعالیت‌هایی از تلفن‌های هوشمند خود استفاده می‌کنند؟»

🔧 گروه اجراکنندهٔ این پروژه، پرسش‌نامه‌ای طراحی کردند و با انتخاب تصادفی ۲۰ نفر از هر پایه، از این نمونه از دانش‌آموزان خواستند که این پرسش‌نامه را تکمیل کنند:

- الف) به‌طور متوسط، روزانه چند دقیقه از تلفن هوشمند خود استفاده می‌کنید؟
ب) از میان گزینه‌های زیر سه فعالیتی را که بیش از دیگر گزینه‌ها با تلفن هوشمند خود انجام می‌دهید، علامت بزنید.
- خواندن کتاب الکترونیک
 - شبکه‌های اجتماعی
 - عکاسی
 - بازی
 - تماس تلفنی
 - تماشای فیلم یا گوش دادن به موسیقی

مشاوران مدرسه تصمیم داشتند با توجه به کمی بودن میزان استفاده از تلفن هوشمند، نتایج آن را به صورت میانه و دامنه میان چارکی^۱ گزارش کنند و درصد فراوانی هر یک از فعالیت‌ها را به صورت نمودار میله‌ای نمایش دهند.

● حدس می‌زنید دانش‌آموزان کلاس شما به طور متوسط روزانه چقدر و برای چه فعالیت‌هایی از تلفن‌های هوشمند خود استفاده می‌کنند؟

● اگر واحد اندازه‌گیری مدت استفاده از تلفن هوشمند (دقیقه) در پرسش اول مشخص نمی‌شد، ممکن بود چه مشکلی پیش بیاید؟

🇩🇪 داده‌ها پس از گردآوری در نرم‌افزار وارد شدند. اشتباه‌های تایی تا حد ممکن رفع و درستی داده‌های دورافتاده بررسی شد. پرسش‌نامه‌هایی که اطلاعات آنها اشتباه به نظر می‌رسید (مثلاً میزان استفاده از تلفن هوشمند بیش از ۲۴ ساعت در روز بود!) کنار گذاشته شدند.

● ممکن است برخی دانش‌آموزان نخواهند کسی از جزئیات فعالیت‌های آنها با تلفن هوشمند باخبر شود. در طراحی پرسش‌نامه چگونه می‌توانستیم با احترام به حریم خصوصی افراد، داده‌های گردآوری شده را به واقعیت نزدیک‌تر کنیم؟

📊 نتایج این نظرسنجی درباره فعالیت‌های انجام شده با تلفن هوشمند به صورت نمودار میله‌ای زیر گزارش شد. در نمونه مورد بررسی، میانه میزان استفاده از تلفن هوشمند ۷۰ دقیقه و دامنه میان چارکی آن ۳۵ (۱۰۰ = Q_3 , $Q_1 = 65$) دقیقه بود.



● با توجه به چارک‌های گزارش شده، (تقریباً) چند درصد از دانش‌آموزان به طور متوسط بیش از ۱۰۰ دقیقه در روز با تلفن هوشمند خود کار می‌کنند؟

● با توجه به نمودار میله‌ای، بیشترین و کمترین فراوانی مربوط به کدام فعالیت‌هاست؟

● چرا در این نمودار میله‌ای، مجموع درصد فراوانی‌ها از ۱۰۰ بیشتر می‌شود؟

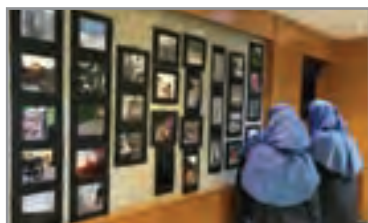
۱- پس از مرتب کردن داده‌ها، مقداری را که تعداد داده‌های بعد از آن با تعداد داده‌های قبل از آن برابر است، میانه می‌نامیم. در داده‌های مرتب شده، میانه داده‌های قبل از چارک اول (Q_1) و میانه داده‌های بعد از چارک سوم (Q_3) می‌نامیم. دامنه میان چارکی حاصل تفاضل چارک سوم و اول است که آن را با IQR (Interquartile Range) نمایش می‌دهیم.

💡 باید توجه داشته باشیم که این نتایج فقط در مورد دانش‌آموزان پایه‌های این مدرسه (نه دانش‌آموزان دیگر مدارس) و فقط در مورد چگونگی استفاده از (نه ابزارهای دیگر) صادق است.^۱

● با توجه به نمودار میله‌ای صفحه پیش، دانش‌آموزان متوسطه دوم این مدرسه بیشتر برای انجام دادن چه فعالیت‌هایی از تلفن هوشمند خود استفاده می‌کنند؟

● از میانه و دامنه میان چارکی گزارش شده برای زمان استفاده روزانه از تلفن‌های هوشمند چه نتیجه‌ای می‌گیرید یا به زبان آماری، آنها را چگونه تفسیر می‌کنید؟

❖ مشاوران آموزشی، به کمک دانش‌آموزان و اولیا با بررسی نتایج و فعالیت‌هایی که بیشترین فراوانی را داشتند و با در نظر گرفتن ویژگی‌های دانش‌آموزان مدرسه به این نتیجه رسیدند که (۱) تشکیل یک گروه آموزشی در یکی از شبکه‌های اجتماعی می‌تواند عملکرد تحصیلی و ارتباط عاطفی معلم‌ها و دانش‌آموزان را بهبود بخشد؛ (۲) معلم‌ها می‌توانند با تهیه فیلم‌های آموزشی جذاب، محتوای درسی را به صورت فیلم در اختیار دانش‌آموزان قرار دهند و دانش‌آموزان در ساعاتی از شبانه‌روز که آمادگی بیشتری دارند، فیلم‌ها را تماشا کنند و پرسش‌هایشان را در کلاس به صورت حضوری با معلم‌ها در میان بگذارند؛ (۳) با برگزاری دوره‌های آموزشی و نمایشگاهی از آثار دانش‌آموزان، می‌توان علاقه دانش‌آموزان به عکاسی و در نتیجه، استفاده از تلفن‌های هوشمند را هدفمند کرد.



❓ در این نظرسنجی فقط درباره مدت زمانی که دانش‌آموزان در طول روز با تلفن‌هایشان کار می‌کنند، پرسیدیم اما اگر بخواهیم برنامه‌ای داشته باشیم که با استفاده از آن عملکرد تحصیلی دانش‌آموزان بهتر شود، باید در یک نظرسنجی دیگر، مدت زمانی را که دانش‌آموزان صرف کارهای مختلف می‌کنند، به تفکیک این فعالیت‌ها بدانیم.

همان‌طور که در فعالیت بالا دیدید، حل کردن مسئله‌های مرتبط با آمار به صورت چرخه‌ای کامل شامل گام‌های ۱: بیان مسئله، ۲: طرح و برنامه‌ریزی، ۳: گردآوری و پاک‌سازی داده‌ها، ۴: تحلیل داده‌ها و ۵: بحث و نتیجه‌گیری است.



۱- هنگامی که بر اساس داده‌های یک نمونه آماری، برای جامعه آماری مرتبط با آن نتیجه‌گیری می‌کنیم، به اصطلاح می‌گوییم نتایج را تعمیم داده‌ایم.

کار در کلاس

اکنون که چرخه حل مسئله را می‌شناسیم، می‌توانیم در جدول زیر نام هر گام و ترتیب آن را مقابل توضیحاتش بنویسیم. (مشاهده نمونه حل‌شده در سطر سوم می‌تواند به ما کمک کند.)

ترتیب گام	نام گام	توضیح گام
		راهی برای رسیدن به پاسخ مسئله پیدا می‌کنیم. به نمونه‌گیری، شیوه اندازه‌گیری متغیر و چگونگی توصیف نتایج می‌اندیشیم.
		نتایج به‌دست‌آمده را تفسیر می‌کنیم و پاسخی برای پرسش اصلی پیدا می‌کنیم.
۱	بیان مسئله	مسئله‌ای را که در دنیای واقعی وجود دارد به صورت یک مسئله شفاف و دقیق آماری مطرح می‌کنیم.
		با استفاده از معیارها، نمودارها و مفاهیمی که آموخته‌ایم، نتایج را متناسب با هدف‌های کارمان، نوع متغیرها و ویژگی‌های داده‌ها گزارش می‌کنیم.
		داده‌ها را گردآوری می‌کنیم و تا حد ممکن از درستی آنها مطمئن می‌شویم.

گام‌های چرخه آمار در حل مسائل

پیش از هر چیز خوب است بدانیم «چرا فرایند حل مسئله را «چرخه» می‌نامیم». همان‌طور که در مثال «استفاده از تلفن هوشمند» دیدیم، گام‌های مختلف حل مسئله به ما کمک می‌کنند که برای پرسش اصلی پاسخی پیدا کنیم، اما اغلب پس از اجرای تمامی گام‌ها با روشن‌تر شدن ابعاد مختلف مسئله و براساس نتایج و تجربه به‌دست آمده، یک مسئله جدید دقیق‌تر شکل می‌گیرد. در واقع، برای پیدا کردن پاسخ پرسش اصلی تا رسیدن به یک جمع‌بندی منطقی، یک مسئله چندین بار بازمینی می‌شود. در این بخش، گام‌های چرخه حل مسئله را بیشتر بررسی می‌کنیم.

؟ گام اول (بیان مسئله)

طرح یک پرسش دقیق و شفاف مهم‌ترین گام رسیدن به پاسخ است. بنابراین، پیش از هر چیز باید مسئله به درستی درک شود و سپس به طور دقیق تعریف و بیان گردد؛ یعنی با توجه به اهداف، بودجه، زمان و دیگر شرایط موجود، جامعه آماری را محدود، و هدف مطالعه را مشخص کنیم.



مثال: فرض کنید می‌خواهیم میزان فعالیت بدنی پسران کرمانشاه را بدانیم و برای آن برنامه‌ای داشته باشیم. پیش از هر چیز باید بتوانیم مسئله را دقیق تعریف کنیم. پس از مشورت با کارشناسان این حوزه متوجه می‌شویم که با توجه به عواملی مانند سن، محل سکونت (از نظر نزدیکی به مکان‌های ورزشی)، الگوهای فرهنگی فعالیت بدنی (مانند ارزشمندی ورزش‌های پهلوانی و...) یا توانایی‌ها و ناتوانی‌های

جسمانی، میزان فعالیت بدنی این افراد بسیار متفاوت است؛ پس، بدون محدود کردن جامعه آماری نمی‌توانیم پاسخ مشخصی برای این مسئله داشته باشیم.

بدین ترتیب، مثلاً «بررسی متوسط فعالیت بدنی هفتگی در هوای آزاد در میان پسران ۱۳ تا ۱۵ ساله یک مدرسه مشخص در کرمانشاه» می‌تواند به عنوان یک مسئله مطرح شود. چند مثال دیگر بزنید.

بحث و گفت‌وگو

هر یک از عوامل نام برده، (سن، محل سکونت و...) چگونه می‌تواند بر فعالیت بدنی جامعه آماری مورد بررسی تأثیر بگذارد؟ چه عوامل دیگری به تعریف دقیق مسئله کمک می‌کند؟ گروه اجرا کننده این مطالعه باید شامل چه متخصصانی باشد و دانش هر یک چه نقشی در این مطالعه دارد؟ (مثلاً متخصصان تربیت بدنی چگونه می‌توانند به اندازه‌گیری متغیر «فعالیت بدنی» کمک کنند؟)

گام ۲ (طرح و برنامه‌ریزی)

اندازه‌گیری یا سنجش، اولین قدم برای یافتن داده‌ها و بررسی متغیر مورد نظر است. در اندازه‌گیری، در واقع سعی می‌کنیم اطلاعات توصیفی (کیفی) را تا حد ممکن به اطلاعات کمی (عددی) تبدیل کنیم. انتخاب معیار اندازه‌گیری مناسب برای متغیرهایی مانند قد، وزن و دما آسان است؛ زیرا با ابزارهایی آشنا می‌توانیم توصیف‌هایی مانند قد بلند، وزن مناسب یا هوای گرم را به کمیتی قابل فهم برای همه تبدیل کنیم. البته اندازه‌گیری همیشه این قدر هم آسان نیست. فرض کنید در مطالعه‌ای می‌خواهیم میزان «خوشحالی» افراد را اندازه‌گیری کنیم. چه پرسش‌هایی می‌تواند برای اندازه‌گیری این متغیر به ما کمک کند؟ آیا اندازه‌گیری دارایی‌ها، تحصیلات، دستاوردها، تعداد دوستان و... لزوماً مشخص می‌کند که هر فرد چقدر خوشحال است؟ احتمالاً این طور نیست. با توجه به اهمیت اندازه‌گیری می‌توانیم بفهمیم که حضور افراد متخصص در گروه هدایت‌کننده پژوهش چقدر مهم است. متخصصان، ابزارهای اندازه‌گیری یا پرسش‌نامه‌ها را طوری طراحی می‌کنند که با آنها بتوانیم همان متغیری را که مورد نظر است، اندازه بگیریم. بدون مشورت با این افراد، ممکن است به جای «خوشحالی» مثلاً «رفاه» یا به جای «هوش» مثلاً «حافظه» اندازه‌گیری شود.

در مرحله طرح و برنامه‌ریزی علاوه بر توافق درباره‌ی چگونگی اندازه‌گیری متغیرهای مورد نظر، درباره‌ی اندازه نمونه، چگونگی نمونه‌گیری و همچنین شیوه تحلیل داده‌ها تصمیم‌گیری می‌شود.



تعداد اعضای جامعه را اندازه جامعه و تعداد اعضای نمونه را اندازه نمونه می‌نامیم. هرچه پراکندگی متغیر مورد بررسی در جامعه بیشتر باشد، برای حصول اطمینان از وجود تنوع در نمونه، به اندازه نمونه بزرگ‌تری نیاز داریم.

کار در کلاس

۱. در هر قسمت، برای بررسی کدام یک از موارد به اندازه نمونه بزرگ تری نیاز است؟ توضیح دهید.

الف) سن دانش‌آموزان در کلاسی که در آن درس می‌خوانید.

ب) معدل سال گذشته دانش‌آموزان در کلاسی که در آن درس می‌خوانید.

ب) بررسی متوسط حقوق معلمان یک مدرسه دولتی

ب) بررسی متوسط حقوق کارکنان یک شرکت بزرگ خصوصی

۲. علی و امید در یک پروژه دانش‌آموزی می‌خواهند بررسی کنند که «نوجوانان منطقه ۲ آبادان در اوقات فراغت خود به چه فعالیت‌هایی مشغول‌اند و برای هر فعالیت چقدر وقت صرف می‌کنند.»

علی پیشنهاد می‌کند: «برای نمونه‌گیری به چند مکان تفریحی، فرهنگی، مذهبی و ورزشی برویم و در هر مکان چند نوجوان را به‌طور تصادفی انتخاب کنیم و از آنها بپرسیم به‌طور متوسط چند ساعت در هفته به این مکان می‌آیند.»

امید پیشنهاد می‌کند: «اگر اداره آموزش و پرورش منطقه موافقت کند، برای نمونه‌گیری فهرستی از دانش‌آموزان متوسطه اول و دوم منطقه ۲ آبادان تهیه کنیم. بعد، با اندازه نمونه کافی افرادی را به‌طور تصادفی انتخاب کنیم و از هر دانش‌آموز بپرسیم در اوقات فراغت خود چه فعالیت‌هایی انجام می‌دهد و به‌صورت متوسط چند ساعت در هفته به این فعالیت‌ها مشغول است.»



موارد زیر اشکالاتی است که در پیشنهادها امید و علی وجود دارد. در هر مورد، بنویسید که اشکال مربوط به پیشنهاد علی است یا امید.

الف) کسانی که اوقات فراغتشان را در خانه می‌گذرانند، شانس حضور در نمونه را ندارند.

ب) نوجوانان بازمانده از تحصیل، شانس حضور در نمونه را ندارند.

پ) اگر مثلاً به کتابخانه برویم و از نوجوانان بپرسیم به‌طور متوسط چند ساعت در هفته به این مکان می‌آیند، میانگین داده‌های به‌دست‌آمده (آماره نمونه) بزرگ‌تر از میانگین آن در جامعه آماری (پارامتر جامعه) می‌شود؛ چون نوجوانانی که اصلاً به کتابخانه نمی‌روند، در نظر گرفته نشده‌اند.

دو اشکال دیگر را که به نمونه‌گیری علی وارد است، مطرح کنید.



۳. برای بررسی وضعیت اجتماعی - اقتصادی خانوارهای یک شهر در کدام شیوه نمونه‌گیری، همه قشرهای جامعه شانس حضور ندارند؟ چرا؟

الف) انتخاب خانوارها براساس رقم اول تلفن خانه‌ها

ب) انتخاب خانوارها براساس رقم آخر تلفن خانه‌ها.

گام ۳ (گردآوری و پاک‌سازی داده‌ها)



در هر مطالعه ممکن است در مرحله اندازه‌گیری، گردآوری یا ثبت داده‌ها و یا وارد کردن داده‌ها در نرم‌افزار اشتباهی رخ دهد. با بررسی دقیق داده‌ها می‌توان برخی از این اشتباه‌ها را تصحیح کرد.

کار در کلاس

آنچه در جدول زیر نشان داده شده، بخشی از داده‌های گردآوری شده در یک پروژه دانش‌آموزی پیش از برطرف کردن اشتباهات است.

چون برخی نرم‌افزارها نمی‌توانند داده‌های غیر عددی را تحلیل کنند، برای متغیرهای کیفی کُد تعیین می‌کنیم؛ مثلاً در این داده‌ها برای جنسیت دختر کد ۱ و برای پسر کد ۲ تعیین شده است. در متغیر «ورزش مورد علاقه» نیز برای فوتبال، والیبال، بسکتبال و شنا به ترتیب کدهای ۱، ۲، ۳ و ۴ تعیین شده است. در این داده‌ها هر سطر نشان‌دهنده یک فرد و هر ستون نشان‌دهنده یک متغیر است.

نمره ادبیات	متوسط ساعات مطالعه آزاد در هفته	ورزش مورد علاقه	جنسیت	سن	شماره دانش‌آموز
۱۵	۲	۱	۱	۱۴	۱
۱۴	۰	۱	۲	۱۵	۲
۱۵	۳	۲	۱	۱۴	۳
+	۵	۲۲	۱	۱۵	۴
۱۷	۴۴	۲	۲	۵۱	۵
۱۸	۶	۳	۲	۱۴	۶
	۱۹	۷	۳	۱۵	۷
۱۷	۴	۴	۲	۱۴	۸
۱۵	۳	۳	۱	۱۴	۹
۱۰	۰	۲	۱	۱۴	۱۰
۱۲	۱	۲	۱	۱۴	۱۱
۱۴	۲	۱	۲	۱۴	۱۲
۲۰	۷	۴	۱	۱۵	۱۳
۱۸	۵	۱	۲	۱۳	۱۴

با توجه به جدول داده‌ها

الف) در متغیر سن، کدام داده احتمالاً اشتباه تایپ شده است؟ چگونه می‌توانستیم با مرتب کردن داده‌ها در نرم‌افزار، این داده را از میان تعداد بسیار زیادی سطر پیدا کنیم؟

ب) در داده‌های مربوط به کدام دانش‌آموز به دلیل جا افتادن کد جنسیت، داده‌های متغیرهای بعدی (ورزش مورد علاقه، متوسط ساعات مطالعه آزاد و نمره ادبیات) به سمت چپ جابه‌جا شده‌اند؟ ورزش مورد علاقه این دانش‌آموز چیست؛ نمره ادبیات او چند است.

پ) در متغیر «ورزش مورد علاقه» کدام عدد به اشتباه نوشته شده است؟

ت) در متغیر «متوسط ساعات مطالعه آزاد در هفته» کدام عدد با الگوی داده‌های دیگر تفاوت دارد؟ آیا می‌توانیم این عدد را بدون بررسی، حذف یا اصلاح کنیم؟

ث) در متغیر «نمره ادبیات» داده‌های مربوط به کدام دانش‌آموز با واحد اندازه‌گیری متفاوتی ثبت شده است؟ چگونه می‌توانستیم از این اتفاق جلوگیری کنیم؟

بحث و گفت‌وگو

اگر دانش‌آموزی در روز برگزاری آزمون ادبیات غایب باشد چه باید کرد؟ اگر بخواهیم نمره‌ای برای او در نظر بگیریم، این کار را چگونه انجام دهیم؟ در این باره چه پیشنهادی دارید؟
آیا با بررسی تمامی موارد ممکن می‌توانیم مطمئن باشیم که داده‌ها کاملاً درست ثبت شده‌اند؟ چرا؟

گام ۴ (تحلیل داده‌ها)

در این مرحله، داده‌ها را تحلیل می‌کنیم و نتایج را ارائه می‌دهیم. منظور از تحلیل داده‌ها در واقع صرفاً گزارش معیارها و ارائه نمودارها و دیگر نتایج آماری است. تفسیر این نتایج و نتیجه‌گیری از آنها در گام بعد انجام خواهد شد.
یادآوری: همان‌طور که می‌دانید، برای توصیف داده‌های کمی باید هم معیار گرایش به مرکز و هم معیار پراکندگی گزارش شود. معیارهای گرایش به مرکز مانند میانگین و میانه به ما کمک می‌کنند بدانیم داده‌ها در کجا متمرکزند و معیارهای پراکندگی به ما کمک می‌کنند بدانیم داده‌ها چگونه متراکم یا پراکنده شده‌اند. به یاد دارید که اگر داده‌دوره‌افتاده داشته باشیم، میانگین و انحراف معیار، معیارهای مناسبی برای توصیف داده‌ها نخواهند بود. چرا؟

کار در کلاس

با توجه به داده‌ها جدول زیر را کامل کنید.

متغیر	داده‌ها	معیارهای گرایش به مرکز		معیارهای پراکندگی		
		میانگین	میانه	IQR	SD (s)	دامنه پراکندگی
سن (سال)	۱۷ ۱۸ ۱۶ ۱۷ ۱۵ ۱۹ ۱۵ ۱۶ ۱۷ ۱۸ ۱۵ ۱۹					
میزان تماشای تلویزیون	۴ ۲۹ ۴ ۶ ۷ ۶/۵ ۷ ۸ ۶ ۳۰ ۴ ۵					

راهنمای جدول: میزان تماشای تلویزیون: متوسط ساعت‌های تماشای تلویزیون در هفته؛ *IQR: دامنه میان چارکی؛ **SD: انحراف معیار.

با توجه به جدول بالا، به پرسش‌های زیر پاسخ دهید.

الف) کدام معیار گرایش به مرکز و کدام معیار پراکندگی برای توصیف داده‌های متغیر «سن» مناسب‌اند؟ چرا؟

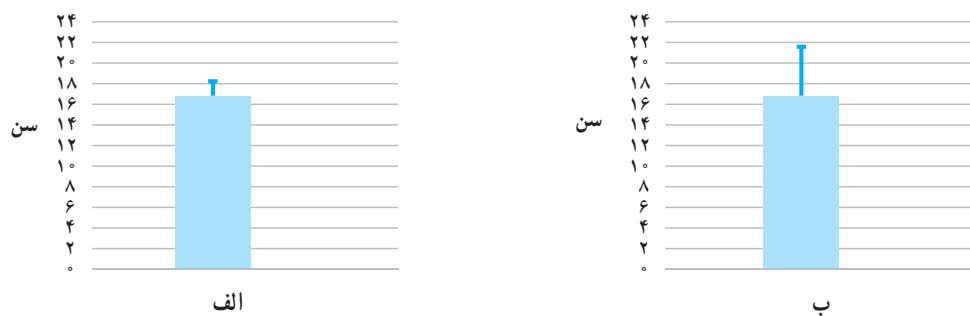
* IQR: Interquartile Range

** SD: Standard Deviation

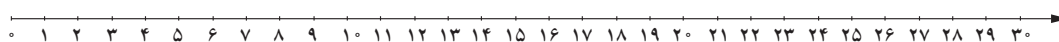
- ب) کدام معیار گرایش به مرکز و کدام معیار پراکندگی، برای توصیف داده‌های «ساعت‌های تماشای تلویزیون» مناسب‌اند؟ چرا؟
- پ) انحراف کدام داده‌ها از میانگین، باعث بزرگ شدن انحراف معیار متغیر «ساعت‌های تماشای تلویزیون» شده است؟ چگونه؟
- ت) اگر برای توصیف داده‌های متغیر «ساعت‌های تماشای تلویزیون» فقط میانگین گزارش می‌شد، خوانندگانی که به داده‌های خام دسترسی نداشتند چگونه گمراه می‌شدند؟

در داده‌هایی که میانگین و انحراف معیار شاخص‌های مناسبی برای توصیف هستند، می‌توانیم از نموداری استفاده کنیم که بلندی مستطیل آن نشان‌دهنده میانگین باشد و میله خطی آن، به اندازه انحراف معیار، روی مستطیل بالا آمده باشد.

- ث) با توجه به داده‌های جدول کار در کلاس در صفحه قبل، کدام یک از نمودارهای زیر نشان‌دهنده میانگین و انحراف معیار سن است؟ نمودار الف یا ب؟



اگر داده‌های دورافتاده داشته باشیم، نمی‌توانیم فقط به نمایش میانگین و انحراف معیار بسنده کنیم؛ زیرا گزارش میانگین می‌تواند گمراه‌کننده باشد. در این گونه موارد باید از نمودار جعبه‌ای استفاده کنیم. برای داده‌های «ساعت‌های تماشای تلویزیون» در مثال بالا، با استفاده از مقدار چارک‌های اول، دوم (میانه) و سوم، ابعاد جعبه (به ترتیب ابتدا، خط میانی و انتها) را روی محور زیر مشخص کنید. سپس، با توجه به کوچک‌ترین و بزرگ‌ترین داده‌ها، نمودار جعبه‌ای را رسم کنید.



- الف) کدام نمودار بهتر نشان می‌دهد که داده‌ها کجا متراکم‌تر و کجا پراکنده‌ترند: نمودار نمایش‌دهنده میانگین و انحراف معیار یا نمودار جعبه‌ای؟

- ب) اگر از الگوی توزیع داده‌ها و وجود داده‌های دورافتاده اطلاعی نداشته باشیم، کدام نمودار برای نمایش اطلاعات متغیرهای کمی مطمئن‌تر است؟

کار در کلاس

دو شرکت به نرگس پیشنهاد کار داده‌اند که یکی از آنها ۳۰ و دیگری ۳۱ کارمند دارد. میانگین حقوق ماهیانه کارمندان شرکت اول، تقریباً ۱/۶ میلیون تومان و میانگین حقوق ماهیانه کارمندان شرکت دوم تقریباً ۴ میلیون تومان است. او برای انتخاب یکی از این دو شرکت، از اعظم و مریم نظرخواهی می‌کند.

اعظم می‌گوید: «شرکت دوم را انتخاب کن؛ چون به طور متوسط هر ماه ۴ میلیون تومان درآمد خواهی داشت.»
مریم می‌گوید «به نظر من این اطلاعات کافی نیست. گزارش معیار گرایش به مرکز بدون معیار پراکندگی چه فایده‌ای دارد؟»
آنها تصمیم می‌گیرند شرایط را کمی بیشتر بررسی کنند پس، در هر شرکت چند نفر را به طور تصادفی انتخاب می‌کنند و اطلاعات زیر را گردآوری می‌کنند.

		حقوق کارمندان							
شرکت اول	۱	۱/۵	۱/۲	۱/۷	۱/۴	۱/۹	۳	۱/۱	۱/۸
شرکت دوم	۰/۷	۱	۸	۱۲/۱	۰/۸	۱/۶	۱۰/۱	۰/۹	۰/۸

الف) با استفاده از داده‌های جدول بالا، نمودار جعبه‌ای حقوق ماهیانه کارمندان این دو شرکت را روی یک محور رسم کنید.



ب) با توجه به نمودارهای قسمت الف، درستی یا نادرستی هر یک از عبارتهای زیر را مشخص کنید.

- حقوق تقریباً ۵٪ کارمندان شرکت دوم کمتر از کمترین حقوق کارمندان شرکت اول است.
- اگر نرگس سطح تحصیلات و تجربه متوسطی داشته باشد و این دو شرکت افراد را براساس شایستگی‌هایشان انتخاب کنند، شرکت اول برای او مناسب است.
- در شرکت دوم، میانگین، معیار گرایش به مرکز مناسبی است.

کار در کلاس

دو مدرسه متوسطه دوم، که تعداد دانش‌آموزانشان تقریباً برابر بود، در سال گذشته با تغییر برنامه‌ها و کادر آموزشی‌شان، وضعیت قبولی دانش‌آموزان خود را در ۱۵ دانشگاه برتر کشور بهبود دادند. مدرسه اول قبولی‌هایش ۲۰۰ درصد افزایش داشت و مدرسه دوم ۲۰ درصد. فرض می‌کنیم اولیا برای انتخاب مدرسه فرزندان خود محدودیتی ندارند.

الف) در این مسئله، متغیر «قبولی در ۱۵ دانشگاه برتر کشور» (قبول شدن / قبول نشدن) چه نوع متغیری است؟ کمی یا کیفی؟

ب) با توجه به این اطلاعات، کدام مدرسه را به اولیا پیشنهاد می‌کنید؟

پ) افزایش ۲۰۰ درصدی و ۲۰ درصد به چه معناست؟ توضیح دهید.

ت) فرض کنید در مدرسه اول، تعداد قبولی‌ها قبلاً ۱ نفر بوده و با تغییر کادر آموزشی ۳ نفر شده است (۲۰۰٪ افزایش) و در مدرسه دوم، تعداد قبولی‌ها ۷۰ نفر بوده و با تغییر کادر آموزشی ۸۴ نفر شده است (۲۰٪ افزایش). اکنون کدام مدرسه را پیشنهاد می‌کنید؟

ث) در گزارش متغیرهای کیفی، ارائه درصد بدون مشخص کردن تعداد، چگونه می‌تواند گمراه کننده باشد؟

برای توصیف داده‌های کیفی (اسمی یا ترتیبی) گزارش درصد باید همیشه با گزارش تعداد همراه باشد.

گام ۵ (بحث و نتیجه‌گیری) 

پس از تحلیل داده‌ها، باید بتوانیم با تفسیر نتایج، پاسخی برای مسئله اصلی پیدا کنیم. در تفسیر نتایج، توجه به محدودیت‌های مطالعه‌ای که انجام داده‌ایم بسیار مهم است. روش کار و محدودیت‌های آن باید چنان صادقانه گزارش شود که اگر افراد دیگری تصمیم به انجام دادن مطالعه‌ای در همان زمینه داشتند، با مشکلاتی مشابه مواجه نشوند. پیشنهادهای ما می‌تواند به این افراد کمک کند. اگر ضمن توجه به تمامی نکات نمونه‌گیری، داده‌ها را با روش‌های مناسبی تحلیل کرده باشیم، در بهترین حالت می‌توانیم نتایج را فقط به جامعه آماری مورد بررسی تعمیم دهیم؛ توجه کنید که اگر تمامی افراد جامعه آماری را بررسی نکرده‌ایم، نتایج ما قطعی نیستند؛ چون در نمونه‌گیری دیگری از همین جامعه، افراد دیگری در نمونه قرار می‌گیرند و نتایج متفاوتی به دست خواهد آمد. (از جامعه‌ای با اندازه N به چند طریق می‌توانیم نمونه‌هایی با اندازه n داشته باشیم؟)

یکپارچگی چرخه آمار در حل مسائل

همان‌طور که دیدید، در تمامی گام‌های حل مسئله مرتبط با آمار، لازم است مسئله مورد بررسی، متغیرها و عوامل مرتبط با آنها را خوب بشناسیم. این امر، بدون کار گروهی و همکاری کارشناسان زمینه پژوهش مورد بررسی، ممکن نیست. مثلاً اگر آمارگیران از اهداف کلی مطالعه باخبر نباشند یا فردی که داده‌ها را تحلیل می‌کند، ویژگی‌های جامعه آماری یا موضوع مورد بررسی را به خوبی



نشناسد، حتی اگر هر فرد به صورت جداگانه کار خود را خیلی خوب انجام دهد، نتیجه نهایی لزوماً مطلوب نخواهد بود. باید توجه داشته باشیم که اگر مطالعه در منطقه‌ای انجام می‌شود که ویژگی‌های آن را به خوبی نمی‌شناسیم، در تمام گام‌ها به دانش اهالی منطقه اولویت دهیم. کیفیت اجرای هر یک از گام‌های چرخه آمار، گام‌های دیگر را به شدت تحت تأثیر قرار می‌دهد. مثلاً برای نقد معیارهای گزارش شده در یک مطالعه، نه تنها می‌توانیم چرایی گزارش میانگین یا میانه و یا گزارش نشدن شاخص پراکندگی را نقد کنیم، بلکه باید درباره چگونگی نمونه‌گیری، شیوه اندازه‌گیری، نوع ورود داده‌ها و تمامی موارد گام‌های دیگر چرخه نیز تردید و پرسش داشته باشیم.

کار در کلاس



۱. برای احداث یک استخر در حاشیه یک کلان‌شهر، محل مناسبی انتخاب، و با همکاری گروهی متخصص، استخری مجهز و اصولی در آن محل ساخته شد. پس از بهره‌برداری از استخر، مشاهده شد که اغلب اوقات افراد مسن زیادی، که شناگر نبودند، به توصیه پزشکی در قسمت کم‌عمق در حال راه رفتن در آب‌اند و در قسمت عمیق گاهی فقط چند نفر شنا می‌کنند.

الف) آگاهی از میانگین سنی افراد منطقه احداث استخر، چگونه می‌توانست به اجرای این پروژه کمک کند؟

ب) چه کسانی می‌توانستند پیش از احداث این استخر، درباره فراوانی افرادی که در آن منطقه شناگر بودند، اطلاع بدهند؟

پ) چگونه مشورت با افراد بومی درباره اطلاعات قسمت الف و ب می‌توانست به متخصصان کمک کند که ابعاد قسمت کم‌عمق و عمیق استخر را متناسب با نیاز منطقه طراحی کنند؟

۲. هر یک از موارد زیر اجرای نادرست کدام گام است و بر کدام گام‌های دیگر اثر می‌گذارد؟ چگونه؟

الف) مسئله به صورتی بیان شده است که اجراکنندگان برداشت‌های متفاوتی از اهداف پژوهش دارند.

ب) اندازه‌گیری وزن افراد نمونه با دو واحد متفاوت (کیلوگرم و پوند) انجام شده است.

پ) تمامی داده‌های دورافتاده حذف شده‌اند.



ت) در پژوهشی با موضوع «بررسی قدمت بناهای مسکونی چند روستا»، مجریان پروژه جلساتی را با حضور کارشناسان برگزار می‌کنند و پس از تصمیم‌گیری درباره طراحی و برنامه‌ریزی پروژه، برای گردآوری داده‌ها چند نفر را استخدام می‌کنند. آمارگیرها بدون اطلاع از چگونگی شکل‌گیری مسئله و اهداف این پژوهش، برای پر کردن پرسش‌نامه‌ها به منازل مختلف مراجعه می‌کنند. هنگام وارد کردن داده‌ها در نرم‌افزار، مشاهده می‌شود که در چندین پرسش‌نامه در مقابل متغیر «سن بنا» نوشته‌اند: «فوت شده است».

ث) پس از گردآوری و تحلیل داده‌ها به شیوه‌ای که پیش‌تر در گام «طرح و برنامه‌ریزی» به تأیید رسیده بود، نتایج مورد انتظار به دست نیامد. به همین سبب، معیارهای دیگری برای گزارش نتایج انتخاب شده است.

ج) نتایج را فقط افراد متخصص آمار تفسیر کرده‌اند.

خواندنی

پادشاهی پسر خود را به جماعتی اهل هنر سپرده بود تا او را از علوم نجوم و رمل و غیره آموخته بودند و استاد تمام گشت با کمال کودکی. روزی پادشاه انگشتی در مُشت گرفت؛ فرزند خود را امتحان کرد که بیا بگو در مُشت چه دارم؟ پسر گفت: «آنچه داری گرد است و زرد است و مُجَوَّف (میان تهی) است.» پادشاه گفت: «چون نشان‌های راست دادی، پس حکم کن که آن چه چیزی باشد.» پسر گفت: «باید که غریب (غریب، الک) باشد.» پادشاه گفت: «از قوت تحصیل و دانش این چندین نشان‌های دقیق دادی که عقول در آن حیران شوند اما ندانستی که غریب در مُشت نگنجد؟»

برگرفته از فیه ما فیه — مولانا جلال‌الدین محمد

تمرین

۱. به نظر می‌رسد کودکانی که زبان مادری‌شان فارسی نیست، در دوره دبستان مشکلات بیشتری برای یادگیری مفاهیم درسی دارند. با انتخاب یک جامعه آماری محدود (از نظر پایه تحصیلی، جنسیت، منطقه، زبان مادری و...) و انتخاب متغیر مناسب، این مسئله را به صورت دقیق و شفاف بیان کنید.

۲. با توجه به تأثیر شیوه‌های مختلف حمل و نقل، بر محیط‌زیست، می‌خواهیم در یک کلاس ۲۸ نفره با انتخاب تصادفی ۹ نفر از دانش‌آموزان، از آنها بپرسیم در هفته گذشته، بیشتر به کدام یک از شیوه‌های زیر به مدرسه آمده‌اند؟

● ۱- پیاده یا با دوچرخه

○ ۲- با وسایل حمل و نقل عمومی

● ۳- با سرویس مدرسه یا خودروی شخصی.

الف) در این مطالعه، جامعه آماری، نمونه آماری، اندازه جامعه و اندازه نمونه را مشخص کنید.

ب) داده‌های این مطالعه را با چه روشی می‌توان گردآوری کرد؟ چه مشکلاتی ممکن است در گردآوری این داده‌ها رخ دهد؟

پ) متغیر تصادفی مورد بررسی این مطالعه چیست؟ نوع متغیر (کمی/کیفی) و مقیاس اندازه‌گیری (فاصله‌ای/نسبتی - اسمی/ترتیبی) آن را مشخص کنید.

ت) چه نمودارها و آماره‌هایی برای گزارش نتایج این مطالعه مناسب‌اند؟

ث) آیا می‌توانیم این نتایج را به شیوه رفت و آمد دانش‌آموزان این کلاس در کل سال تحصیلی تعمیم دهیم (مثلاً به فصل‌های مختلف یا زمان برگزاری آزمون‌های پایان نیم‌سال)؟ توضیح دهید.

ج) اگر ۹ نفر دیگر از این کلاس را به طور تصادفی انتخاب می‌کردیم، آیا لزوماً نتایج مشابهی حاصل می‌شد؟ توضیح دهید.

۳. در هریک از موارد زیر، علت عدم تناسب جامعه آماری با نمونه‌های انتخاب شده را توضیح دهید. برای بهبود نمونه‌گیری چه پیشنهادی دارید؟ انتخاب نادرست نمونه در هریک از موارد چگونه بر نتایج تأثیر می‌گذارد؟

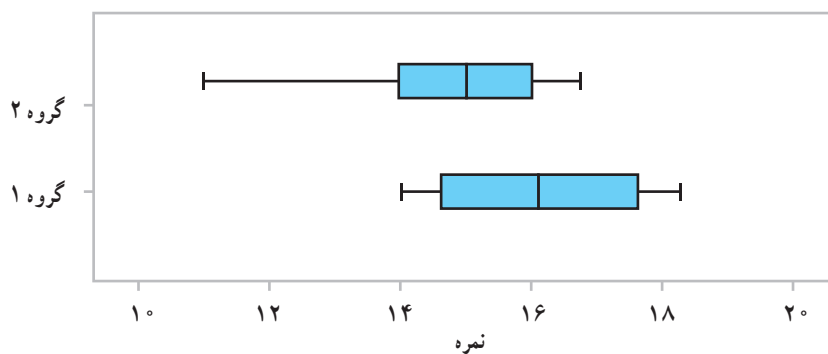
الف) مسئله: بررسی میزان رضایت شغلی در کارمندان یک کارگاه شبانه‌روزی

نمونه: انتخاب تصادفی تعدادی از مدیران ارشد شیفت روز

ب) مسئله: نظرسنجی از مادران یک شهر درباره میزان رضایتشان از برنامه‌های کودک تلویزیون
 نمونه: انتخاب تصادفی برخی منازل از تمامی مناطق شهر و مراجعه به آنها بین ساعت ۸ تا ۱۰ صبح

پ) مسئله: نظرسنجی از دانش‌آموزان پایه دوازدهم منطقه ۲ بوشهر درباره اینکه به طور متوسط چند ساعت از روز را به انجام دادن تکالیف مدرسه اختصاص می‌دهند.
 نمونه: دختران پایه دوازدهم یک مدرسه از این منطقه که به صورت داوطلبانه در این مطالعه مشارکت کرده‌اند.

۴. «مطالعات دانشمندان علوم اعصاب نشان می‌دهد که مغز انسان برای انجام دادن هم‌زمان چند فعالیت ساخته نشده است و هر بار که ما گمان می‌کنیم مشغول انجام دادن چند کار به صورت هم‌زمان هستیم، در واقع مغزمان دارد با سرعت از کاری به کار دیگر می‌پرد. مغز انسان تشنه خیرها و اتفاقات جدید است. وقتی مشغول مطالعه و یادگیری هستیم و ناگهان صدای دریافت یک پیام را از تلفن همراه خود می‌شنویم، مغز ما، که بی‌تابانه منتظر کشف آن اطلاع جدید است، دیگر روی یک کار متمرکز نیست. این مطالعات نشان می‌دهد که مثلاً اگر دانش‌آموزی هنگام مطالعه و یادگیری، هم‌زمان تلویزیون تماشا کند یا گفت‌وگویی را در یک شبکه اجتماعی دنبال کند، مطالب درسی به جای اینکه به قسمت حقایق و ایده‌ها برود، به قسمت مهارت‌ها فرستاده می‌شود.»
 سعید و سهراب پس از خواندن این مطلب در نشریه مدرسه تصمیم گرفتند نمره عملکرد تحصیلی ۲۵ دانش‌آموز پایه دوازدهم مدرسه را که هنگام مطالعه، تلفن همراه خود را در حالت بی‌صدا قرار می‌دهند (گروه یک)، با ۲۵ دانش‌آموز پایه دوازدهم دیگر که این کار را انجام نمی‌دهند (گروه دو) مقایسه کنند. نتایج این مطالعه به این شرح است:



الف) از مقایسه میانه گروه یک با چارک سوم گروه دو چه نتیجه‌ای می‌گیرید؟

ب) پراکندگی دو گروه را با استفاده از دامنه تغییرات و دامنه میان چارکی مقایسه کنید.

پ) از مقایسه کمینه گروه یک با چارک اول گروه دوم چه نتیجه‌ای می‌گیرید؟

ت) در کدام گروه گزارش میانگین و انحراف معیار می‌تواند گمراه‌کننده باشد؟

ث) در کدام گروه مقدار میانگین و میانه به هم نزدیک‌ترند؟

ج) نتایج این مطالعه را به چه جامعه‌ای می‌توان تعمیم داد؟

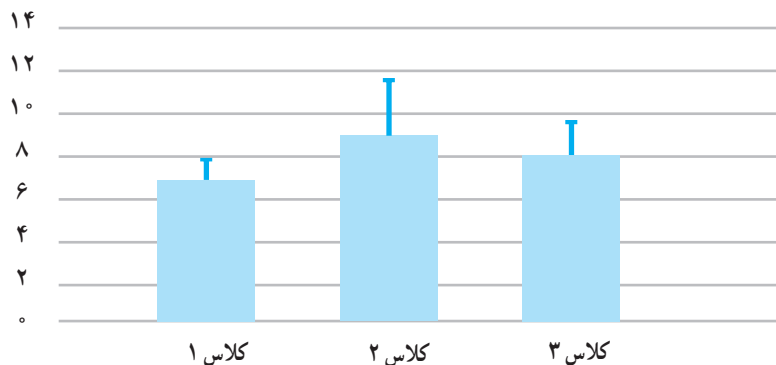
۵. در یک نظرسنجی، از ۱۵ دانش‌آموز علوم انسانی پایهٔ یازدهم یک مدرسه پرسیده‌ایم که به نظر آنها چند ساعت آموزشی در ماه برای برگزاری جلسات نقد کتاب یا فیلم مناسب است. داده‌های گردآوری شده به این شرح است:

۱۶ ۱۰ ۱۲ ۱۰ ۱۲ ۱۳ ۶ ۴ ۰ ۲۰ ۱۸ ۲۲ ۲۰ ۱۴ ۱۲

الف) این داده‌ها را طوری تغییر دهید که میانگین و انحراف معیار بیشتر شود اما میانه و چارک اول و چارک سوم تغییر نکند.

ب) فرض کنید می‌خواهیم دو نفر دیگر به این نمونه اضافه کنیم. داده‌های این دو نفر را طوری انتخاب کنید که میانگین و میانه تغییر نکند.

۶. با توجه به تأثیر کمیت و کیفیت خواب بر یادگیری، ندا و آوا در مطالعه‌ای با بررسی تمامی دانش‌آموزان سه کلاس پایهٔ دهم، میانگین میزان خواب این دانش‌آموزان را در هفتهٔ گذشته برحسب ساعت پرسیدند و ثبت کردند. آنها پس از گردآوری داده‌ها و بررسی درستی آن را به صورت نمودارهای زیر نمایش دادند. (در این نمودارها بلندی مستطیل، نشان‌دهندهٔ میانگین و میلهٔ خطا نشان‌دهندهٔ انحراف معیار است.)



الف) چه عواملی ممکن است هنگام گردآوری داده‌ها نتایج را از واقعیت دور کند؟ (هنگامی که دانش‌آموزان راجع به میانگین ساعت خوابشان در هفتهٔ گذشته می‌پرسیم آیا آنها می‌توانند به درستی به یاد آورند که هر شب چند ساعت خوابیده‌اند تا بتوانند میانگین را به درستی اعلام کنند؟ برای حل این مشکل چه راهکاری پیشنهاد می‌کنید؟)

ب) نوع (کمی/کیفی) و مقیاس اندازه‌گیری (فاصله‌ای/نسبتی - اسمی/ترتیبی) متغیر مورد بررسی در این مطالعه را مشخص کنید.

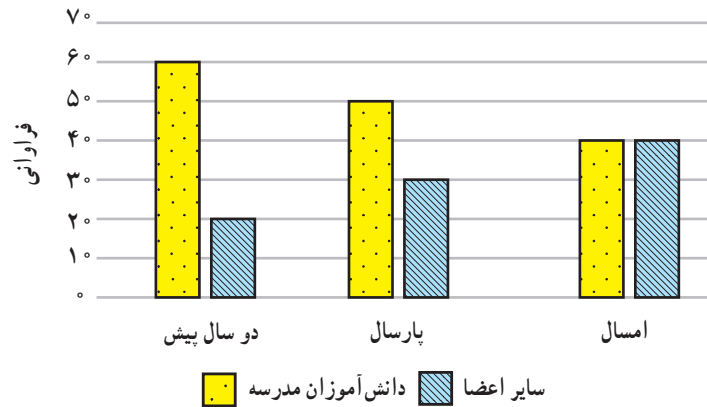
پ) با استفاده از اطلاعات نمودارها، نتایج را به صورت تقریبی در جدول زیر بنویسید.

	کلاس ۱	کلاس ۲	کلاس ۳
میانگین			
انحراف معیار			

ت) چه عواملی می‌تواند بر نتایج این مطالعه تأثیر بگذارد؟ چگونه می‌توانیم با بیان مسئله‌ای جدید، پاسخ دقیق‌تری برای این مسئله پیدا کنیم؟

ث) چه کسانی می‌توانند در اجرای بهتر این مطالعه به ما کمک کنند؟ چگونه؟

۷. مدرسه‌ای برای ترویج فرهنگ کتاب‌خوانی، از دو سال پیش به غیر از دانش‌آموزان مدرسه، از خارج از مدرسه نیز عضو می‌پذیرد. نمودار میله‌ای زیر نشان‌دهنده روند تغییرات فراوانی اعضای جدید کتابخانه در دو گروه مذکور است.



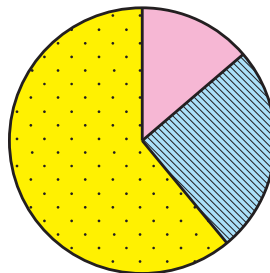
الف) سال گذشته چند نفر از خارج مدرسه عضو کتابخانه شده‌اند؟

ب) با توجه به روند این تغییرات، پیش‌بینی می‌کنید سال آینده چه تعداد از دانش‌آموزان این مدرسه عضو کتابخانه شوند؟

پ) با استفاده از نمودارهای میله‌ای داده شده، برای هر یک از این سه سال یک نمودار دایره‌ای جداگانه رسم کنید.

ت) به نظر شما چه عواملی موجب این روند تغییرات شده است؟

۸. فراوانی بازدیدکنندگان از یک سالن نمایش در گروه‌های سنی مختلف در نمودار دایره‌ای زیر نمایش داده شده است. (در این نمودار، ۵۴ درجه مربوط به گروه سنی کمتر از ۲۰ سال، ۹۰ درجه مربوط به گروه سنی ۲۰ تا ۴۰ سال و بقیه مربوط به گروه سنی ۴۰ تا ۶۰ سال است.)



۴۰ تا ۶۰ سال ۲۰ تا ۴۰ سال کمتر از ۲۰ سال

الف) با توجه به این نمودار، چند درصد افراد بازدیدکننده بین ۴۰ تا ۶۰ سال سن دارند؟

ب) مسئول فروش بلیت این سالن نمایش پس از مشاهده نتایج، آن را غیرواقعی دانست و گفت: «بیشتر بازدیدکنندگان این سالن کمتر از ۴۰ سال دارند.» از موارد زیر کدام می‌توانند نتایج این مطالعه را بدین شکل از واقعیت دور کرده باشد؟ توضیح دهید.

- بسیاری از افراد کمتر از چهل سال تمایلی به پاسخ‌گویی به سؤالات پرسشگر نداشته‌اند.
 - گردآوری داده‌ها در ساعات اداری انجام شده است و بازنشستگان بیشتری در نمونه قرار گرفته‌اند.
 - هنگام گردآوری داده‌ها، دانش‌آموزان بسیاری از طرف مدرسه برای بازدید حضور داشته‌اند.
۹. جدول زیر نشان‌دهنده تعداد تصادف خودروها و سرعت حرکت آنها در زمان تصادف است.

سرعت (کیلومتر در ساعت)	۱۰	۱۳	۲۱	۲۵	۳۵	۴۰	۷۰	۱۵۰	۲۴۰	۳۵۰	۴۰۰
تعداد خودروهای تصادف کرده	۱۰	۲۰	۴۰	۷۰	۱۵۰	۲۴۰	۳۵۰	۴۰۰	۷۰	۲۱	۱۳

همان‌طور که دیده می‌شود، تعداد تصادف خودروهایی که سرعتشان بیش از ۹۰ کیلومتر در ساعت بوده، کمتر است. پس:

«هرچه سریع‌تر بروید، مطمئن‌تر و امن‌تر است.»

نتیجه‌گیری بالا چه اشکالی دارد؟ چرا این تصور ایجاد شده است؟

پروژه

موضوعی دلخواه یا یکی از موضوعات زیر را انتخاب کنید و با استفاده از گام‌های چرخه آمار، نتایج حاصل را در قالب یک گزارش در کلاس ارائه کنید.

- بررسی ساعات روزانه حضور دانش‌آموزان در شبکه‌های اجتماعی
- نظرسنجی درباره تغییر قوانین حاکم بر کلاس یا مدرسه
- نظرسنجی درباره کلاس درسی که می‌توانید پرسش‌های خود را آزادانه در آن مطرح کنید.
- نظرسنجی درباره ساعات لازم برای تدریس ریاضی در کلاس دوازدهم انسانی
- بررسی ساعت‌های مطالعه آزاد دانش‌آموزان و مقایسه آن بین دو گروه از دانش‌آموزان
- بررسی فعالیت‌های فوق برنامه مورد علاقه دانش‌آموزان.

