

درس ۳

چرخه آمار در حل مسائل

دنیای امروز پر از اطلاعاتی است که ما در ابعاد مختلف زندگی با آنها روبه‌رو می‌شویم. بنابراین برای اینکه بتوانیم بهتر زندگی کنیم و در مورد 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، تفسیر درستی از این اطلاعات داشته باشیم، محدودیت‌های نتایج به‌دست آمده را بشناسیم و نهایتاً 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 داشتند دسترسی دانش‌آموزان به تلفن‌های هوشمند را در راستای اهداف آموزشی مدرسه هدفمند کنند. اما این مشاوران اطلاع دقیقی از میزان و چگونگی استفاده دانش‌آموزان از تلفن‌های هوشمند نداشتند و بدین ترتیب برنامه‌ریزی برای این هدف ممکن نبود. پس اولین مسئله مشاوران آموزشی این بود که بفهمند: «دانش‌آموزان پایه‌های دهم، یازدهم و دوازدهم این مدرسه به‌طور متوسط روزانه چقدر و برای انجام چه فعالیت‌هایی از تلفن‌های هوشمند خود استفاده می‌کنند؟»

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

- الف) به‌طور متوسط روزانه چند دقیقه از تلفن هوشمند خود استفاده می‌کنید؟
ب) تمامی فعالیت‌هایی را که در طول یک هفته حداقل یک‌بار با تلفن هوشمند خود انجام می‌دهید در گزینه‌های داده شده علامت بزنید. (می‌توانید بیش از یک گزینه را انتخاب کنید.)
- خواندن کتاب الکترونیک شبکه‌های اجتماعی عکاسی بازی تماس تلفنی گوش دادن به موسیقی یا تماشای فیلم

آنها 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 فراوانی انجام هر یک از فعالیت‌ها را به صورت نمودار میله‌ای نمایش دهند.

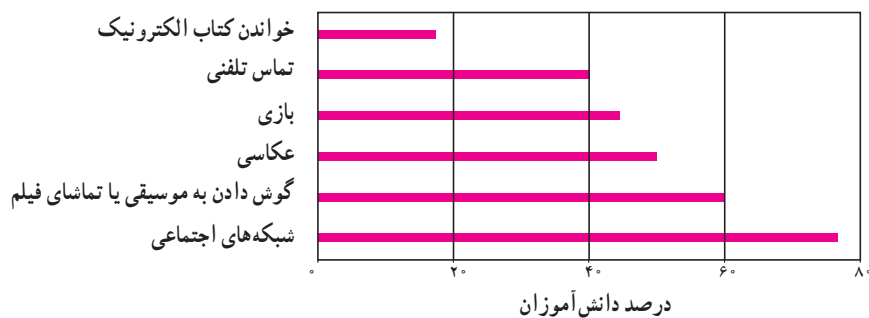
۱-۲- حدس می‌زنید دانش‌آموزان کلاس شما به طور متوسط روزانه چقدر و برای انجام چه فعالیت‌هایی از تلفن‌های هوشمند خود استفاده می‌کنند؟

۲-۲- اگر واحد اندازه‌گیری زمان استفاده از تلفن هوشمند (دقیقه) در پرسش اول مشخص نمی‌شد، ممکن بود چه مشکلی پیش بیاید؟

📷 پس از گردآوری این داده‌ها، اطلاعات در نرم‌افزار وارد شد. اشتباه‌های تاییبی 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 شد.

۱-۳- 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 به صورت نمودار میله‌ای زیر گزارش شد؛ میانه و دامنه میان چارکی میزان استفاده از تلفن هوشمند در نمونه مورد بررسی به ترتیب ۷۰ و ۳۵ ($Q_1 = 65$ و $Q_3 = 100$) دقیقه بود. در نمونه مورد بررسی میانه میزان استفاده از تلفن هوشمند ۷۰ دقیقه و دامنه میان چارکی آن ۳۵ ($Q_1 = 65$, $Q_3 = 100$) دقیقه بود.



۱-۴- 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؟

۲-۴- با توجه به نمودار میله‌ای، بیشترین و کمترین فراوانی مربوط به کدام فعالیت‌ها است؟

۳-۴- 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 میانه را چارک اول (Q_1) و میانه داده‌های بعد از میانه را 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) و فقط در مورد چگونگی استفاده از (نه ابزارهای دیگر) صادق است^۱.

۱-۵ با توجه به نمودار میله‌ای بالا، دانش‌آموزان متوسطه دوم این مدرسه، بیشتر برای انجام چه فعالیت‌هایی از تلفن هوشمند خود استفاده می‌کنند؟

۲-۵ 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 آن 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 ببخشد. (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 از معلم‌ها بپرسند. (۳) با برگزاری دوره‌های آموزشی و نمایشگاه‌هایی از آثار دانش‌آموزان، می‌توان علاقه دانش‌آموزان به عکاسی و در نتیجه استفاده از تلفن‌های هوشمند را هدفمند کرد.

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

همان‌طور که در فعالیت بالا دیدید، حل 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 اصطلاحاً 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.
		به تفسیر نتایج به‌دست‌آمده می‌پردازیم و پاسخی برای پرسش اصلی پیدا می‌کنیم.
۱	بیان مسئله	مسئله‌ای را که در دنیای واقعی وجود دارد به‌صورت یک مسئله شفاف و دقیق آماری مطرح می‌کنیم.
		با استفاده از شاخص‌ها، نمودارها و مفاهیمی که آموخته‌ایم نتایج را متناسب با هدف‌های کارمان، نوع متغیرها و ویژگی‌های داده‌ها گزارش می‌کنیم.
		به گردآوری داده‌ها می‌پردازیم 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 می‌کند پاسخی برای پرسش اصلی پیدا کنیم. اما اغلب پس از انجام 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 برای پرسش اصلی تا رسیدن به یک جمع‌بندی منطقی، یک مسئله چندین بار بازبینی می‌شود. در این بخش به بررسی بیشتر هر یک از گام‌های چرخه حل مسئله می‌پردازیم.

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 دقیقاً تعریف و بیان شود؛ یعنی با توجه به اهداف، بودجه، زمان و دیگر شرایط موجود، جامعه آماری را محدود کنیم و هدف مطالعه را مشخص نماییم.

مثال: فرض کنید می‌خواهیم میزان فعالیت بدنی پسران کرامت‌شاه را بدانیم و برای آن برنامه‌ای داشته باشیم. پیش از هر چیز باید بتوانیم مسئله را دقیق تعریف کنیم. با مشورت 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...) و یا توانایی‌ها

یا ناتوانی‌های جسمانی، میزان انجام 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 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؟

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 چرا که با ابزارهایی آشنا می‌توانیم توصیف‌هایی مانند قد بلند، وزن مناسب یا هوای گرم را به کمی قابل فهم برای همه تبدیل کنیم. اما 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 هر فرد چقدر خوشحال است؟ احتمالاً این طور نیست. با توجه به اهمیت اندازه‌گیری می‌توان فهمید حضور افراد متخصص در تیم هدایت‌کننده پژوهش چقدر حیاتی است. متخصصین، ابزارهای اندازه‌گیری یا پرسش‌نامه‌ها را طوری طراحی می‌کنند که بتوانیم همان متغیری را اندازه بگیریم که مورد نظر است. بدون مشورت 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 در مورد چگونگی اندازه‌گیری متغیرهای مورد نظر، درباره چگونگی نمونه‌گیری و همچنین شیوه تحلیل داده‌ها تصمیم‌گیری می‌شود.



تعداد اعضای جامعه را اندازه‌گیری و تعداد اعضای نمونه را اندازه‌گیری می‌نمایم. هرچه اندازه جامعه بزرگ‌تر باشد، برای آنکه نمونه بتواند به خوبی بیانگر ویژگی‌های جامعه باشد اندازه نمونه بزرگ‌تری لازم داریم و هر چه پراکندگی متغیر مورد بررسی در جامعه بیشتر باشد، برای حصول اطمینان از حضور تنوع در نمونه، اندازه نمونه بزرگ‌تری لازم داریم.

کار در کلاس

۱. در هر قسمت برای بررسی کدام یک از موارد، اندازه نمونه بزرگ‌تری لازم 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

الف) بررسی میانگین ساعت‌های مطالعه آزاد هفتگی دانش‌آموزان یک مدرسه

ب) بررسی میانگین ساعت‌های مطالعه آزاد هفتگی دبیران یک مدرسه

۲. علی و امید در یک پروژه دانش‌آموزی می‌خواهند با انجام مطالعه‌ای بررسی کنند نوجوانان منطقه ۲ اهواز در اوقات فراغت خود به چه فعالیت‌هایی مشغول‌اند و برای هر فعالیت چقدر وقت می‌گذارند.

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

امید پیشنهاد می‌کند برای نمونه‌گیری، به کمک اداره آموزش و پرورش منطقه، به شرطی که موافقت کنند، فهرستی از دانش‌آموزان متوسطه اول و دوم منطقه ۲ اهواز تهیه کنیم و با اندازه نمونه کافی افرادی را به تصادف 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 می‌پردازد.



موارد زیر اشکالاتی در مورد پیشنهادهای امید و علی است. در هر مورد بنویسید این اشکال مربوط به پیشنهاد علی است یا امید؟

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

ب) نوجوانان بازمانده از تحصیل شانس حضور در نمونه 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 که به نمونه‌گیری علی وارد است مطرح کنید.

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

الف) انتخاب خانوارها براساس رقم اول تلفن منازل

ب) انتخاب خانوارها براساس رقم آخر تلفن منازل



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

داده‌هایی که در شکل زیر نشان داده شده‌اند بخشی از داده‌های گردآوری شده در یک پروژه دانش‌آموزی پیش از برطرف کردن اشتباه‌ها 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	۱
۱۵	۲	۱	۱	۱۴	۱۴	۲
۱۴	۰	۱	۲	۱۵	۵	۳
۱۵	۳	۲	۱	۱۴	۱	۴
+	۵	۲۲	۱	۱۵	۴۵	۵
۱۷	۴۴	۲	۲	۵۱	۲	۶
۱۸	۶	۳	۲	۱۴	۱۷	۷
	۱۹	۷	۳	۱۵	۱۹	۸
۱۷	۴	۴	۲	۱۴	۳۴	۹
۱۵	۳	۳	۱	۱۴	۵۲	۱۰
۱۰	۰	۲	۱	۱۴	۱۲	۱۱
۱۲	۱	۲	۱	۱۴	۷	۱۲
۱۴	۲	۱	۲	۱۴	۳۵	۱۳
۲۰	۷	۴	۱	۱۵	۲۶	۱۴
۱۸	۵	۱	۲	۱۳	۲۱	۱۵

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) داده‌های مربوط به کدام دانش‌آموز به دلیل جا افتادن کد جنسیت، به سمت چپ جابه‌جا شده‌اند؟

پ) در متغیر «ورزش مورد علاقه» کدام عدد به اشتباه تایپ 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 چه پیشنهادی دارید؟

آیا با بررسی تمامی موارد ممکن می‌توانیم مطمئن باشیم داده‌ها کاملاً درست ثبت شده‌اند؟ چرا؟

گام ۴ (تحلیل داده‌ها)

در این مرحله به تحلیل داده‌ها و ارائه نتایج می‌پردازیم. منظور از تحلیل داده‌ها در واقع صرفاً گزارش شاخص‌ها 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 هر دو شاخص مرکزی و پراکندگی گزارش شوند. شاخص‌های مرکزی مانند میانگین و میانه به ما کمک می‌کنند بدانیم داده‌ها کجا متمرکز هستند و شاخص‌های پراکندگی به ما کمک می‌کنند بدانیم داده‌ها چگونه متمرکز یا پراکنده شده‌اند. به یاد دارید که اگر داده دورافتاده داشته باشیم میانگین و انحراف معیار، شاخص‌های مناسبی برای توصیف داده‌ها نخواهند بود. (چرا؟)

کار در کلاس

با توجه به داده‌ها جدول زیر را کامل کنید.

متغیر	داده‌ها	شاخص‌های مرکزی		شاخص‌های پراکندگی		
		میانگین	میانه	IQR	SD	دامنه
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 شاخص‌های مرکزی و پراکندگی برای توصیف داده‌های متغیر «سن» مناسب‌اند؟ چرا؟

ب) کدام شاخص‌های مرکزی و پراکندگی برای توصیف داده‌های متغیر «ساعت‌های تماشای تلویزیون» مناسب‌اند؟ چرا؟

پ) انحراف کدام داده‌ها از میانگین، باعث بزرگ شدن انحراف معیار متغیر «ساعت‌های تماشای تلویزیون» شده است؟ چگونه؟

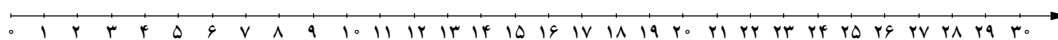
ت) اگر برای توصیف داده‌های متغیر «ساعت‌های تماشای تلویزیون» فقط میانگین گزارش می‌شد، خوانندگانی که به داده‌های خام دسترسی نداشتند چگونه گمراه می‌شدند؟

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

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



اگر داده‌های دورافتاده داشته باشیم، نمی‌توانیم تنها به نمایش میانگین و انحراف معیار بسنده کنیم زیرا گزارش میانگین می‌تواند گمراه‌کننده باشد؛ در این 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 کامل کنید.



الف) کدام نمودار بهتر نشان می‌دهد که داده‌ها کجا متراکم‌تر و کجا پراکنده‌تراند؟ 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؟

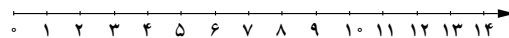
کار در کلاس

نرگس از دو شرکت پیشنهاد کار دارد که یکی از شرکت‌ها ۳۰ و دیگری ۳۱ کارمند دارد. میانگین حقوق ماهیانه کارکنان شرکت اول، تقریباً چهار میلیون تومان و میانگین حقوق ماهیانه کارکنان شرکت دوم تقریباً ۱/۶ میلیون تومان است. او برای انتخاب یکی از این دو شرکت از اعظم و مریم نظر می‌خواهد.

اعظم می‌گوید «شرکت اول را انتخاب کن چون به طور متوسط ماهی ۴ میلیون تومان درآمد خواهی داشت.»
مریم می‌گوید «به نظر من این اطلاعات کافی نیست. گزارش شاخص مرکزی بدون شاخص پراکندگی چه فایده‌ای دارد؟»
آنها تصمیم می‌گیرند شرایط را کمی بیشتر بررسی کنند. در هر شرکت چند نفر را به تصادف انتخاب می‌کنند و اطلاعات زیر را گردآوری می‌کنند.

شرکت اول	حقوق کارمندان								
	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 قبل درستی یا نادرستی هر یک از عبارتهای زیر را مشخص کنید.
- حقوق تقریباً ۵۰٪ کارکنان شرکت دوم کمتر از کمترین حقوق کارکنان شرکت اول است.
 - اگر نرگس سطح تحصیلات و تجربه متوسطی داشته باشد شرکت اول برای او مناسب است.
 - در شرکت دوم، میانگین، شاخص مرکزی مناسبی است.

کار در کلاس

دو مدرسه متوسطه دوم، که تعداد دانش‌آموزانشان تقریباً برابر بود، در سال گذشته با تغییر کادر آموزشی خود 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 والدین برای انتخاب مدرسه بچه‌های خود محدودیتی ندارند.

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

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

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

یادمان 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 مطالعه‌ای در این زمینه داشتند با مشکلاتی مشابه مواجه نشوند. ارائه پیشنهادها ما ممکن است به این افراد کمک کند. اگر به تمامی نکات نمونه‌گیری توجه و با روش‌های مناسبی داده‌ها را 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 نیست؛ 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 دیدیم، در تمامی گام‌های حل مسئله مرتبط با آمار، نیازمندیم 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 شاخص‌های 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، دیده شد که اغلب اوقات افراد مسن زیادی که شناگر نبودند به توصیه پزشکان در قسمت کم‌عمق در حال راه رفتن در آب بودند و در قسمت عمیق گاهی تنها چند نفر شنا می‌کردند.

الف) اطلاع از میانگین سنی افراد آن منطقه چگونه می‌توانست در اجرای این پروژه کمک کند؟

ب) چه کسانی می‌توانستند پیش از احداث این استخر، در مورد فراوانی افرادی که در این منطقه شناگرند اطلاع بدهند؟

پ) چگونه مشورت با افراد بومی در مورد 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 در مورد طرح 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» نوشته شده «فوت شده است».

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

ج) تفسیر نتایج فقط توسط افراد متخصص آمار انجام شده است.

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 انگستری 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 مُجَوَف (میان‌تهی) 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 ننگ‌جذ؟»
برگرفته از فیه ما فیه — مولانا جلال‌الدین محمد

تمرین

۱. به نظر می‌رسد کودکانی که زبان مادری‌شان فارسی نیست در مقطع 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 تصادفی 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 خودرو شخصی

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

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

پ) متغیر تصادفی مورد بررسی این مطالعه چیست؟ نوع متغیر و مقیاس اندازه‌گیری آن را مشخص کنید.

ت) چه نمودارها و چه 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 تصادف 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؟

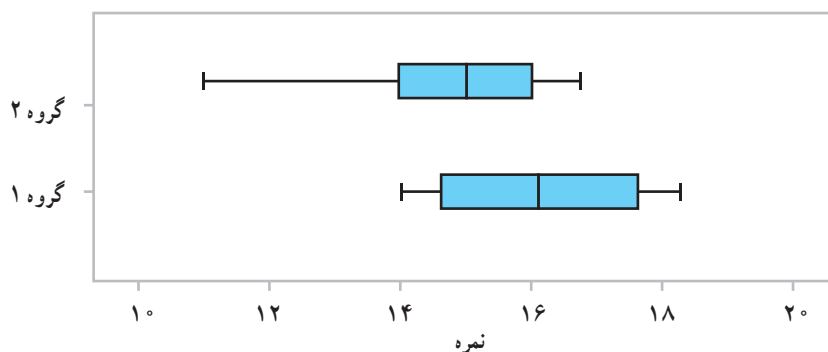
الف) مسئله: بررسی میزان رضایت شغلی در بین کارمندان یک کارگاه شبانه‌روزی

نمونه: انتخاب تصادفی تعدادی از مدیران ارشد شیفت روز

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

پ) مسئله: نظرسنجی از دانش‌آموزان پایه دوازدهم منطقه ۲ بوشهر درباره اینکه به صورت متوسط چند ساعت در روز را به انجام 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 هم‌زمان چند فعالیت ساخته نشده است و هر بار که ما گمان می‌کنیم مشغول انجام دادن چند کار به صورت هم‌زمان هستیم در واقع مغز ما دارد با سرعت از کاری به کار دیگر می‌پرد. مغز انسان تشنه خبرها و اتفاقات جدید است. وقتی مشغول مطالعه و یادگیری هستیم و ناگهان صدای دریافت یک پیام را از تلفن خود می‌شنویم مغز ما که بی‌تابانه منتظر کشف آن اطلاع جدید است و 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 نمی‌کنند (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 نزدیک‌تر هستند؟

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

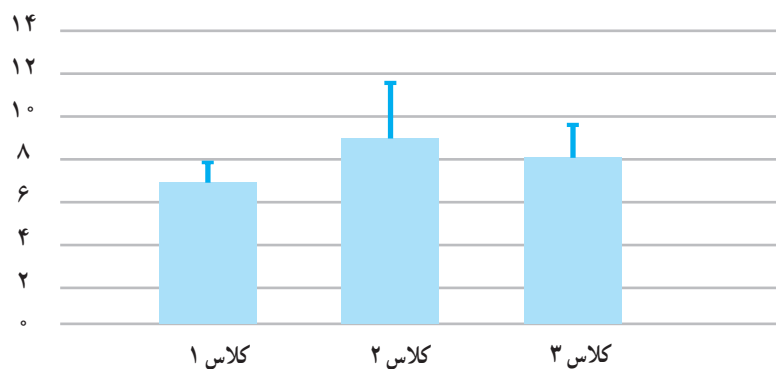
۵. در یک نظرسنجی از ۱۵ دانش آموز علوم انسانی پایه یازدهم یک مدرسه پرسیده ایم به نظر آنها چند ساعت آموزشی در ماه برای برگزاری جلسات نقد کتاب یا فیلم مناسب است. داده های جمع آوری شده به این شرح است :

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

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

ب) فرض کنید می خواهیم دو نفر دیگر به این نمونه اضافه کنیم. داده های 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 مطالعه ای با بررسی تمامی دانش آموزان سه کلاس پایه دهم میانگین ساعت خواب این دانش آموزان را در هفته گذشته برحسب ساعت ثبت کردند. آنها پس از گردآوری داده ها و بررسی درستی آن، نتایج را به صورت نمودارهای زیر نمایش دادند. (در این نمودارها بلندی مستطیل نشان دهنده میانگین و میله خط نشان دهنده انحراف معیار است.)



الف) چه عواملی ممکن است هنگام جمع آوری داده ها نتایج را از واقعیت دور کند؟

ب) نوع و مقیاس اندازه گیری متغیر مورد بررسی در این مطالعه را مشخص کنید.

پ) با استفاده از اطلاعات نمودارها، نتایج را به صورت جدول زیر گزارش کنید.

	کلاس ۳	کلاس ۲	کلاس ۱
میانگین			
انحراف معیار			

ت) چه عواملی می تواند نتایج این مطالعه را تحت تأثیر قرار دهد؟ چگونه می توانیم با بیان مسئله ای جدید پاسخ دقیق تری برای این مسئله پیدا کنیم؟

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

۷. مدرسه‌ای به منظور ترویج فرهنگ کتاب‌خوانی از دو سال پیش تاکنون 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.

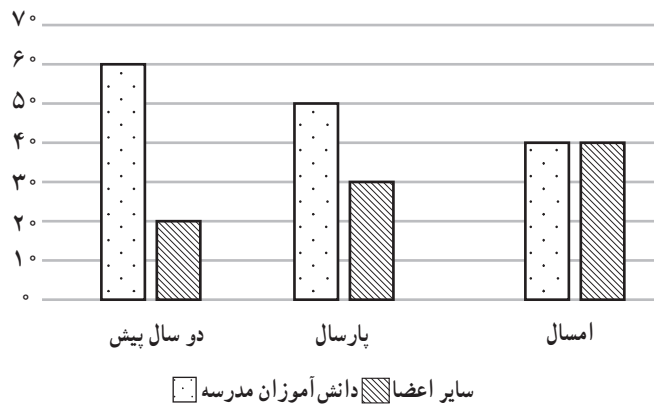
الف) سال گذشته چند نفر از خارج مدرسه عضو کتابخانه شده‌اند؟

ب) این کتابخانه سالانه مجموعاً چند عضو جدید می‌پذیرد؟

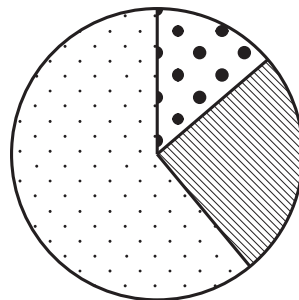
پ) با توجه به روند این تغییرات، پیش‌بینی می‌کنید سال آینده چه تعداد از دانش‌آموزان این مدرسه عضو کتابخانه شوند؟

ت) با استفاده از نمودارهای میله‌ای داده شده، برای این سه سال نمودار دایره‌ای رسم کنید.

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



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



● ۲۰ سال زیر ▨ ۲۰ تا ۴۰ سال □ ۴۰ تا ۶۰ سال

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

ب) مسئول فروش بلیط این سالن نمایش پس از مشاهده نتایج، آن را غیرواقعی دانست و گفت بیشتر بازدیدکنندگان این سالن زیر ۴۰ سال دارند. کدام یک از عوامل زیر می تواند نتایج این مطالعه را بدین شکل از واقعیت دور کرده باشد؟ توضیح دهید.

- بسیاری از افراد زیر چهل سال تمایلی به پاسخ گویی به سؤالات پرسش گر نداشتند.
- جمع آوری داده ها در ساعات اداری انجام شده است و بازنشستگان بیشتری در نمونه قرار گرفته اند.
- هنگام جمع آوری داده ها دانش آموزان بسیاری از طرف مدرسه برای بازدید حضور داشتند.

۹. جدول زیر نشان دهنده تعداد تصادف های اتومبیل ها و سرعت حرکت آنها در زمان تصادف است.

سرعت (کیلومتر در ساعت)	۱۰	۲۰	۳۰	۴۰	۵۰	۶۰	۷۰	۸۰	۹۰	۱۰۰	۱۱۰	۱۲۰
تعداد ماشین های 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 ۹۰ کیلومتر در ساعت رفته اند کمتر است. پس: «هرچه سریع تر بروید، مطمئن تر و امن تر است.»

نتیجه گیری فوق چه اشکالی دارد؟ چرا این تصور ایجاد شده است؟