

درس ۳ چرخه آمار در حل مسئله

- آمار مجموعه‌ای از روش‌ها که برای جمع‌آوری، تصمیم‌گیری، تجزیه و تحلیل و تفسیر داده‌ها برای رسیدن به هدف مشخص انجام می‌گیرد.
 - هر تحقیق و جمع‌آوری اطلاعات با طرح سوال یا مسأله‌ای آغاز می‌شود که تحقیقات برای پاسخگویی به آن انجام می‌شود.
 - حل مسائل آماری به صورت چرخه‌ای کامل است، که شامل گام‌های مشخص و مرتبی است.
- ۱- بیان، فهم و تعریف دقیق مسئله ۲- طرح و برنامه‌ریزی ۳- گردآوری داده‌ها ۴- تحلیل داده‌ها
- ۵- بحث و نتیجه‌گیری، تفسیر نتایج

گام‌های چرخه آمار در حل مسئله

چرا فرایند حل مسئله را «چرخه» می‌نامیم.

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 بر اساس 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 بازبینی می‌شود.

• گام اول (بیان مسئله)

طرح یک پرسش دقیق و شفاف مهم‌ترین گام رسیدن به پاسخ است. بنابراین، بیش از هر چیز باید مسئله به درستی درک شود و سپس به‌طور 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.

توجه: بیان مسئله حساس‌ترین و شاید به‌توان گفت مشکل‌ترین بخش از فرایند پژوهش را تشکیل می‌دهد. قبل از هر چیز باید درک درستی از مسئله داشته باشیم و با توجه به اهداف پژوهش، بودجه، ... جامعه آماری را انتخاب و محدود کنیم.

• گام دوم (طرح و برنامه‌ریزی)

برای یافتن داده‌ها و بررسی متغیرهای گام اول، اندازه‌گیری یا سنجش متغیرها می‌باشد. در اندازه‌گیری، در واقع سعی می‌کنیم اطلاعات توصیفی (کیفی) را تا حد ممکن به اطلاعات کمی (عددی) تبدیل کنیم.

تعریف

جامعه آماری : مجموعه‌ای از افراد یا اشیاء می‌نامیم که دربارهٔ اعضای آن می‌خواهیم موضوع یا موضوعاتی را مطالعه کنیم .

اندازه جامعه : تعداد اعضای جامعه را اندازه جامعه می‌گویند .

نمونه : زیرمجموعه‌ای از جامعه آماری است که مورد بررسی قرار می‌گیرد و به نحوی باید انتخاب شود که نمایانگر خصوصیات تمام جامعه باشد .

اندازه نمونه : تعداد اعضای نمونه اندازه نمونه می‌گوییم .

سرشماری : اگر تمام اعضای جامعه آماری را مورد مطالعه قرار دهیم به این عمل سرشماری می‌گوییم .

توجه : هر چه پراکندگی متغیر مورد بررسی در جامعه بیشتر باشد ، برای حصول اطمینان از وجود تنوع در نمونه ، به اندازه نمونه بزرگ‌تری نیاز است .

• گام سوم (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 . اطلاعات واقعیت‌ها و نتایجی را که از اندازه‌گیری و بررسی نمونه یا جامعه به دست می‌آوریم و در محاسبات ، استنباط‌ها یا برنامه‌ریزی‌ها از آنها استفاده می‌کنیم را « داده آماری » می‌گوییم .

روش‌های گردآوری داده‌ها

۱-مشاهد : جمع‌آوری داده‌ها « بدون نیاز به فرد پاسخگو » را مشاهده می‌گویند . مانند شمارش تعداد کتاب‌های یک کتابخانه
 ۲- پرسش‌نامه : مجموعه « سؤالات از پیش تعیین شده » که توسط افراد پاسخگو تکمیل می‌شود را پرسش‌نامه می‌گویند .
 ۳- مصاحبه : یکی از روش‌های آمارگیری است که معمولاً « بین دو نفر » صورت می‌گیرد که یکی مصاحبه‌کننده (آمارگیر) و دیگری مصاحبه‌شونده (پاسخگو) است .

۴- دادگان‌ها : منظور از دادگان‌ها « مجموعه اطلاعات ذخیره شده » است . در بسیاری از موارد اطلاعات مورد نیاز را می‌توان از سازمان‌ها و نهادهایی که قبلاً اطلاعات را ذخیره کرده‌اند ، به دست آورد این عمل سرعت جمع‌آوری اطلاعات را افزایش می‌دهد .

متغیرهای آماری

متغیر : در علم آمار به ویژگی‌ها یا مشخصه‌های واحدهای آماری که قرار است مورد بررسی قرار بگیرند متغیر می‌گویند .

متغیرها به دو گروه کمی و کیفی تقسیم می‌شوند .

متغیر کمی : متغیری که مقادیر عددی می‌گیرد و برای آن عملیات ریاضی جمع ، تفریق و ... قابل انجام است متغیر کمی نامیده می‌شود . مانند وزن افراد

متغیر کیفی : متغیری که لزوماً مقدار عددی نمی‌گیرد و برای دسته‌بندی افراد یا اشیاء در گروه‌های مختلف به کار می‌رود متغیر کیفی نامیده می‌شود . مانند گروه خونی افراد

• گام چهارم (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 .

تحلیل داده‌ها فرایند مهمی است که برای استخراج اطلاعات ضروری و سودمند ، جهت تصمیم‌گیری انجام می‌شود . منظور از تحلیل داده‌ها ۱- مرتب‌سازی آنها ۲- گزارش شاخص‌ها ۳- ارائه نمودارها و نتایج آماری است که برای ارائه نتایج آماری مورد نظر به ابزارهای ریاضی و روش‌های خاصی نیاز داریم .

توجه : 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 . اگر داده دور افتاده داشته باشیم ، میانگین و انحراف معیار ، معیارهای مناسبی برای توصیف داده‌ها نخواهد بود . زیرا گزارش آنها می‌تواند گمراه‌کننده باشد .

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

• گام پنجم (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 داده مطالعه‌ای در همان زمینه داشتند ، با مشکلاتی مشابه مواجه نشوند . همچنین از ویژگی‌های مهم نتیجه‌گیری می‌توان به ۱- انتخاب شاخص مناسب برای گزارش نتایج و ۲- تفسیر نتایج توسط افراد متخصص آمار اشاره کرد .

یادآوری

میانگین : اگر x_1, x_2, \dots, x_n تا داده آماری باشند ، میانگین آنها را با نماد \bar{x} نشان داده و به صورت زیر تعریف می‌کنیم

:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\text{مجموع تمام داده‌ها}}{\text{تعداد داده‌ها}}$$

ویژگی‌های میانگین

- ۱- میانگین منحصر به فرد است، یعنی فقط و فقط یک مقدار برای میانگین داریم.
 - ۲- میانگین می‌تواند متعلق به جامعه نباشد، یعنی؛ گاهی مقدار عددی میانگین با مقدار هیچ یک از داده‌ها برابر نیست.
 - ۳- میانگین همیشه بین کم‌ترین و بیش‌ترین مقدار داده‌هاست.
- انحراف داده‌ها از میانگین: اختلاف داده‌ها از میانگین را انحراف داده‌ها از میانگین می‌نامند. مجموع تمام داده‌ها از میانگین برابر صفر است.

مد: مد داده‌ای است که بیش‌ترین فراوانی یا بیش‌ترین تکرار را دارد.

میانه: میانه داده‌ای است که نصف داده‌ها از آن کم‌تر و نصف داده‌ها از آن بیش‌تر هستند یعنی همان داده وسط.

نکته

میانگین و میانه ممکن است عضوی از جامعه آماری نباشند اما مد حتماً عضوی از جامعه آماری است.

انحراف معیار

برای سنجش میزان پراکندگی داده‌ها از انحراف معیار استفاده می‌کنیم. برای به‌دست آوردن انحراف معیار اختلاف هر داده با میانگین را به توان ۲ می‌رسانیم و پس از میانگین‌گیری از آنها، جذر می‌گیریم.

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

یکپارچگی چرخه آمار در حل مسائل

در تمام گام‌های حل مسئله در چرخه آمار باید موارد زیر را در نظر گرفت:

- باید مسئله مورد بررسی، متغیرهای آماری و عوامل مرتبط با آنها را خوب بشناسیم.
- آمارگیران باید اهداف کلی مطالعه آماری و ویژگی‌های جامعه آماری را بشناسند.
- اگر در آمارگیری هر فرد به صورت جداگانه کار خود را خیلی خوب انجام دهد اما اگر مطالعه در منطقه‌ای انجام شود که آمارگیر ویژگی‌های آن را به خوبی نشناسد، نتیجه مطالعه مطلوب نخواهد بود.
- در تمام گام‌های حل مسئله باید به دانش اهالی منطقه و متخصص آن موضوع اولویت دهیم.

- کیفیت اجرای هر یک از گام‌ها در چرخه آمار ، گام‌های دیگر را به شدت تحت تأثیر قرار می‌دهد .

یادآوری

- **نمودار میله‌ای :** از دو محور عمودی و افقی تشکیل شده است . محور افقی نشان دهنده داده‌ها و محور عمودی هم نشان دهنده فراوانی داده‌هاست .

- **نمودار بافت نگاشت :** در این نمودار روی محور افقی ، مقادیر داده‌های کمی پیوسته یا دسته‌بندی شده مانند سرعت ، زمان ، نمرات امتحانی ، ... را می‌نویسیم و روی محور عمودی نیز فراوانی مربوط به هر بازه ثبت می‌شود .
توجه : برای متغیرهای اسمی ، نمودار میله‌ای بهتر از بافت نگاشت (هیستوگرام یا مستطیلی) است .
- **نمودار دایره‌ای :** نموداری است که از تقسیم دایره به چند قطاع (قسمتی از دایره که توسط دو شعاع کمائی از آن را محصور می‌کند) به دست می‌آید .

زاویه هر قطاع با توجه به فراوانی آن داده و با تناسب زاویه به روش زیر به دست می‌آید .

$$\frac{\text{فراوانی داده/آم}}{\text{تعداد کل داده‌ها}} = \frac{\text{زاویه قطاع داده/آم}}{360^\circ} \Rightarrow \frac{f}{n} = \frac{\theta}{360^\circ} \Rightarrow \theta = \frac{f}{n} \times 360^\circ$$

توجه : زمانی که درصد را گزارش می‌دهیم ، بهتر است از نمودار دایره‌ای استفاده کنیم تا بتوانیم مقایسه انجام دهیم ؛ البته به شرط آنکه بیش از ۶ مقدار نداشته باشیم .

تذکر : رسم نمودار میله‌ای و دایره‌ای برای متغیرهای کیفی مناسب‌تر از متغیرهای کمی است .

- **نمودار جعبه‌ای :** روش سودمندی برای نمایش دامنه‌ها و چارک‌های داده‌هاست . برای رسم نمودار جعبه‌ای به ۵ مقدار نیاز داریم :

$$1- \text{کوچک‌ترین داده (Min)} \quad 2- \text{چارک اول (} Q_1 \text{)} \quad 3- \text{میانه (} Q_2 \text{)} \quad 4- \text{چارک سوم (} Q_3 \text{)} \\ 5- \text{بزرگ‌ترین داده (Max)}$$

مزیت نسبت به بقیه نمودارها

در نمودار جعبه‌ای ، به طور هم‌زمان ، دامنه تغییرات ، دامنه میان چارکی و میانه نمایش داده می‌شود که برای مقایسه داده‌ها مناسب است . این نمودار برای مقایسه داده‌ها بسیار مناسب است .

مثال ۳۳ . نمودار جعبه‌ای داده‌های ۹، ۷، ۲، ۱۰، ۱، ۵ را رسم کنید .

تمرین ۲۳. با توجه به تأثیر شیوه‌های مختلف حمل و نقل، بر محیط زیست، می‌خواهیم در یک کلاس ۲۸ نفره با انتخاب تصادفی ۹ نفر از دانش‌آموزان، از آنها پرسیم 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- یا سرویس مدرسه یا خودرو شخصی

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

پاسخ: الف)

جامعه آماری	دانش‌آموزان کلاس
نمونه آماری	دانش‌آموزان انتخاب شده
اندازه جامعه	۲۸ نفر
اندازه نمونه	۹ نفر

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

پاسخ: پرسش شفاهی. چون ممکن است در روز جمع‌آوری اطلاعات، مشکلی (سرویس مدرسه نیامده باشد) پیش آمده باشد و دانش‌آموزان پیاده به مدرسه آمده باشند.

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

پاسخ: متغیر تصادفی، شیوه رفتن به مدرسه است و نوع متغیر کیفی و مقیاس اندازه‌گیری اسمی است.

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

پاسخ: نمودار دایره‌ای، میله‌ای برای این داده‌ها مناسب است و \bar{x} و s^2 نسبت آمارهای مناسب برای این نوع متغیر مناسب است.

ث) آیا می‌توانیم این نتایج را به شیوه رفت و آمد دانش‌آموزان این کلاس در کل سال تحصیلی تعمیم دهیم؟ توضیح دهید.

پاسخ: خیر، چون ممکن است در ایام امتحانات شرایط متفاوتی پیش بیاید.

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

پاسخ: خیر نتایج ممکن است مقداری تغییر کنند.

تمرین ۲۴. در هر یک از موارد زیر، علت عدم تناسب جامعه آماری یا نمونه‌های انتخاب شده را توضیح دهید. انتخاب نادرست

نمونه در هر یک 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

پاسخ : (۱) مدیران شیفت شب در نظر گرفته نشده‌اند . (۲) مدیران نمی‌توانند ، نماینده کل کارمندان محسوب شوند . برای رفع مشکل نمونه‌گیری از هر دو شیفت به تناسب و از تمامی کارمندان صورت گیرد .

ب) مسئله : نظرسنجی از مادران یک شهر درباره میزان رضایتشان از برنامه‌های کودک تلویزیون

نمونه : انتخاب تصادفی برخی منازل شهر و مراجعه به آنها بین ساعت ۸ تا ۱۰ صبح

پاسخ : ممکن است برخی از مادران در این ساعت سرکار رفته باشند و لذا اطلاعات از آنها دریافت نمی‌شود . برای رفع این مشکل کافی است مدت مراجعه را تغییر یا افزایش دهیم .

پ) مسئله : نظرسنجی از دانش‌آموزان پایه دوازدهم منطقه ۲ بوشهر درباره اینکه به‌طور متوسط چند ساعت از روز را به انجام دادن تکلیف مدرسه اختصاص می‌دهند .

نمونه : دختران پایه دوازدهم یک مدرسه از این منطقه که به صورت داوطلبانه در این مطالعه مشارکت کرده‌اند .

پاسخ : (۱) دانش‌آموزان پسر از این مطالعه حذف شده‌اند . (۲) نمونه‌گیری تصادفی نیست . نمونه انتخاب نمونه‌ای تصادفی و به تناسب دانش‌آموزان توصیه می‌شود .

تمرین ۲۵. در ک نظر سنجی ، از ۱۵ دانش‌آموز علوم انسانی پایه 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 .

پاسخ : جمع کل داده‌ها برابر ۱۸۹ و میانگین آنها ۱۲/۶ می‌باشد داده آخر را به یک عدد دور افتاده مانند ۵۰ تغییر می‌دهیم .

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 .

پاسخ : میانگین این داده‌ها ۱۲/۶ می‌باشد . لذا می‌توان دو داده طوری اضافه کرد که مثلاً ۳ واحد بیشتر و ۳ واحد کمتر از میانگین

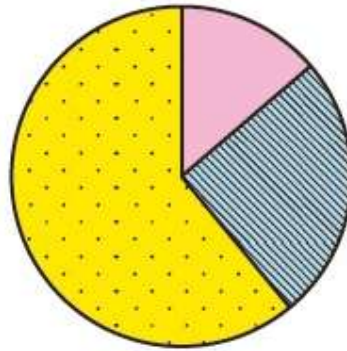
باشند (۹/۶ و ۱۲/۶) . بدین صورت میانگین تغییر نمی‌کند . از طرفی چون وسط بودن میانه نیز حفظ شده است ، لذا با این عمل میانه

نیز ثابت مانده است .

تمرین ۲۶. 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؟

پاسخ: $360^\circ - (90^\circ + 54^\circ) = 216^\circ \Rightarrow \frac{36^\circ}{100} = \frac{216}{x} \Rightarrow x = \frac{216 \times 100}{36} = 600$ درصد

ب) مسئولین 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.
- پاسخ: مورد اول، زیرا اولاً: در ساعات اداری افراد بازنشسته لزومی ندارد که در ادارات حاضر شوند. ثانیاً: افراد کمتر از ۲۰ سال سن معمولاً دانش‌آموز هستند که در این مطالعه سهم کمتری دارند.