



۷

۲۶

فصل اول: معادله درجه دوم

پاسخ سؤال‌های امتحانی



۳۷

۵۷

فصل دوم: تابع

پاسخ سؤال‌های امتحانی

۷۶

۹۴

فصل سوم: کار با داده‌های آماری

پاسخ سؤال‌های امتحانی



۱۰۲

۱۱۵

فصل چهارم: نمایش داده‌ها

پاسخ سؤال‌های امتحانی

۱۲۳

۷۰

۷۲

۱۲۹

۱۳۷

خلاصه فصل‌ها

امتحان‌های نیم‌سال اول

پاسخ‌نامه امتحان‌های نیم‌سال اول

امتحان‌های نیم‌سال دوم

پاسخ‌نامه امتحان‌های نیم‌سال دوم

کارهای دادهای آمار

سوم
فصل



جنس این فصل و فصل بعدی با جنس فصل‌های قبلی، از زمین تا آسمون فرق داره. کلاً آمار یه جور دیگه‌اس. بعضی از تعاریف این فصل مثل جامعه آماری، نمونه تصادفی، انواع متغیرها رو حتماً حتماً باید حفظ کنید. چون سؤال تعریفی هم از این فصل مطرح می‌شه. توی این فصل، مباحث مُد، میانه، میانگین، واریانس، انحراف معیار و دامنه تغییرات رو یاد می‌گیرین. خلاصه این‌جوری بگم که کلاً فصل متنوع و جذابه.

گردآوری داده‌ها - انواع متغیرها



فرض کنید یک استخر بسیار بزرگ پر از ماهی قزل‌آلا وجود دارد و می‌خواهیم در مدت زمان کوتاهی وزن این ماهی‌ها را تخمین بزنیم. خب معلوم است که اگر بخواهیم تک‌تک این ماهی‌ها را وزن کنیم، چندین ساعت و یا چندین روز طول خواهد کشید. پس بهتر است از هر قسمت استخر (کم‌عمق، عمق متوسط، پُرعمق) تعدادی ماهی را انتخاب کرده و وزن کنیم. مثلاً اگر کلاً ۱۰۰۰ ماهی وجود دارد انتخاب ۱۰ ماهی از هر قسمت استخر، منطقی به نظر می‌رسد. پس ما با بررسی وزن ۳۰ ماهی می‌توانیم به طور تقریبی، وزن تمام ماهی‌ها را حدس بزنیم. به این عمل، نمونه‌گیری تصادفی گفته می‌شود. حال با توجه به این مثال، یک سری تعاریف خیلی مهم را مطرح می‌کنیم.

جامعه آماری

به کل افراد، اشیا یا جانداران که آن‌ها را از نظر موضوعی خاص بررسی می‌کنیم، جامعه آماری می‌گوییم. در مثال بالا، کل ماهی‌های استخر، جامعه آماری ما می‌باشند. ضمناً به تعداد آن‌ها، اندازه جامعه (حجم جامعه) می‌گوییم. ضمناً به هر یک از افراد، اشیا یا جانداران موجود در جامعه، یک واحد آماری می‌گوییم.

سرشماری

اگر تمام اعضای جامعه آماری را مورد بررسی و مطالعه قرار دهیم اصطلاحاً می‌گوییم سرشماری کرده‌ایم. ولی سرشماری دارای یک سری مشکلات (محدودیت‌ها) است از جمله: ۱) وقت‌گیر بودن و هزینه‌بر بودن ۲) عدم امکان استفاده در بررسی‌های مُخْتَرَب ۳) در دسترس نبودن تمام اعضای جامعه ۴) کم و زیاد شدن اعضا در طول مدت سرشماری (مانند مرگ و میر یا زاد و ولد یا مهاجرت) ۵) خطای بیشتر در جمع‌آوری داده‌ها به خاطر به همین علت‌ها، در اغلب موارد به جای سرشماری از عمل نمونه‌گیری تصادفی استفاده می‌کنیم.

شگردد؛ من نفهمیدم «بررسی‌های مُقَرَّب» یعنی چی استاد؟

دبیر: بعضی وقت‌ها سرشماری کنیم، همه اعضای جامعه (واحد‌های آماری) نابود میشن. مثلاً تو کارخانه‌ها تولید لامپ، برای آن‌که مقاومت لامپ‌ها رو تفمین بزنند، به هر لامپ جریان برق وصل کرده و ولتاژ رو مرتباً زیاد می‌کنند تا لامپ بسوزه. حال آگه کارفونه بفواد این کار رو برای همه لامپ‌ها انجام بده دیگه لامپی باقی نمی‌مونه که بفواد بفروشه؛ پس میان به پای سرشماری، نمونه‌گیری می‌کنن.

نمونه تصادفی

قسمتی کوچک‌تر از جامعه آماری است که برای بررسی و مطالعه انتخاب می‌کنیم. عمل نمونه‌گیری، مهم‌ترین بخش علم آمار است. نمونه تصادفی باید دارای یک سری شرایط باشد که عبارت است از:

- ۱) اعضا به صورت کاملاً شانس و تصادفی انتخاب شوند؛ یعنی برای انتخاب آن‌ها نباید از هیچ قانون یا الگوی خاصی استفاده کنیم.
- ۲) اندازه نمونه (تعداد اعضای نمونه) باید به اندازه کافی بزرگ باشد؛ یعنی تعداد عضوهای نمونه باید متناسب با اندازه جامعه باشد.
- ۳) نمونه انتخابی باید بیان‌کننده خصوصیات اصلی جامعه باشد.

متغیر تصادفی

به موضوع مورد مطالعه، متغیر می‌گوییم. در مثال ماهی‌ها، موضوع مورد مطالعه، وزن ماهی‌ها بود پس متغیر تصادفی، همان وزن ماهی‌هاست. به این علت از کلمه متغیر تصادفی استفاده می‌کنیم که اولاً اعضا به صورت تصادفی و شانس انتخاب می‌شوند و ثانیاً موضوع مورد مطالعه، از یک عضو به عضو دیگر در حال تغییر است. (می‌دونید که متغیر یعنی تغییرکننده) مثلاً وزن ماهی‌ها معمولاً با هم متفاوت است (پس وزن، متغیر است).

داده‌های آماری

پس از مطالعه و بررسی متغیرهای تصادفی مربوط به یک جامعه یا نمونه، به نتایجی می‌رسیم که به آن‌ها داده‌های آماری می‌گوییم. این نتایج ممکن است عددی یا غیرعددی باشند. مثل اعداد وزن ماهی‌ها و یا رنگ خودروهای یک نمایشگاه.

آمارگیری

عبارت است از جمع‌آوری داده‌های آماری به یکی از روش‌های ممکن (با این روش‌ها همین الان آشنا می‌شید).

آمارگیر

کسی است که با روش‌های موجود در علم آمار، به بررسی و جمع‌آوری داده‌ها می‌پردازد.

روش‌های جمع‌آوری داده‌های آماری

برای رسیدن به داده‌های آماری چهار روش وجود دارد که عبارت‌اند از: ۱) مشاهده و آزمایش ۲) مصاحبه ۳) پرسش‌نامه ۴) دادگان (داده‌های از پیش تهیه‌شده)

۱) مشاهده و آزمایش

یعنی گردآوری داده‌ها بدون نیاز به تماس مستقیم، مانند شمارش تعداد ماشین‌هایی که در یک ساعت از تقاطعی مشخص عبور می‌کنند یا مانند بررسی رابطهٔ هوش و پیشرفت تحصیلی افراد.

۲) مصاحبه

همان پرسش شفاهی است و معمولاً بین دو نفر اتفاق می‌افتد. (یکی مصاحبه‌گر یا آمارگر و دیگری مصاحبه‌شونده). از این روش بیشتر زمانی استفاده می‌شود که آمارگر از تمامی پاسخ‌های طرف مقابل، اطلاعی ندارد. (گاهی اوقات، آمارگر نمی‌تونه همهٔ جواب‌های بعضی سوالات را درس بزنه پس نمی‌تونه پرسش‌نامهٔ کتبی تهیه‌کنه و می‌ره سراغ مصاحبه).

مثلاً اگر بخواهیم دربارهٔ مسائل فرهنگی کاهش ترافیک پژوهش کنیم، مصاحبه با صاحب‌نظران برای یافتن داده‌های مورد نیاز، بسیار راه‌گشا خواهد بود. (متن کتاب صفحه ۹۱)

۳) پرسش‌نامه

مجموعه‌ای از سؤالات است که از قبل، تهیه و تنظیم شده است. البته باید دقت کرد که سؤالات باید ساده و کاملاً واضح باشند تا از سؤالات، برداشت‌های مختلف نشود. ضمناً تا حد امکان باید از سؤالاتی استفاده کرد که جواب‌های کوتاه و یک‌کلمه‌ای داشته باشند یا به صورت تستی (چند گزینه‌ای) باشند و در نهایت این که نباید از سؤالات هدایت‌کننده استفاده کرد. مثلاً اگر از دانش‌آموزان یک مدرسه این‌گونه سؤال شود که: «با توجه به این که کیفیت آموزشی در چند ماه اخیر بسیار بهتر شده، نظر شما در مورد عملکرد دبیران چگونه است؟» در واقع به دانش‌آموزان تلقین شده که عملکرد دبیران خوب است و شاید دیگر هیچ انتقادی را مطرح نکنند.

۴) تذکر مرکز آمار ایران هر ۵ سال یک بار، با استفاده از پرسش‌نامه، اطلاعات نفوس و مسکن تمامی خانوارهای ایرانی را جمع‌آوری می‌کند. (متن کتاب، صفحه ۹۰)

۴) دادگان‌ها (داده‌های از پیش تهیه‌شده)

در بسیاری از مواقع می‌توانیم از داده‌هایی استفاده کنیم که قبلاً به دست آمده و ذخیره شده است. به عنوان مثال اگر بخواهیم در مورد رتبه‌های کنکور یک شهر در سال ۹۴ تحقیق کنیم می‌توانیم به بانک اطلاعاتی سازمان سنجش مراجعه کنیم. استفاده از این روش باعث صرفه‌جویی در وقت و هزینه خواهد شد.

نکته هر یک از روش‌های جمع‌آوری داده‌ها محدودیت‌هایی هم دارند که عبارت‌اند از:

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

مشاهده: اگر به دقت زیادی نیاز داشته باشیم، مشاهده خیلی مناسب نیست.

دادگان: سازمان‌ها همیشه اطلاعات ثبتی (ذخیره‌شده) را در اختیار آمارگیر قرار نمی‌دهند.

مثال در هر قسمت مشخص می‌کنیم که برای جمع‌آوری داده‌ها چه روشی مناسب‌تر است؟

الف) رضایت مشتریان یک بانک از نحوهٔ برخورد و رسیدگی کارمندان ← پرسش‌نامه

(مثلاً یکی از سوالات پرسش‌نامه می‌تونه این طوری باشه: سرعت عمل کارمندان این شعبه چگونه است؟ بسیار خوب - خوب - متوسط - بد - بسیار بد)

ب) علائمی که کودکان سرطانی بیمارستان محک داشته‌اند. ← دادگان‌ها (با مراجعه به پروندهٔ بچه‌ها)

ج) تعداد سرنشینان خودروهای سواری در یکی از محورهای خروجی شهر تهران در روز و ساعتی مشخص ← مشاهده

د) ساعات خواب دانش‌آموزان کلاس شما در شب گذشته ← مصاحبه

انواع متغیرها

به طور کلی متغیرها به دو دستهٔ کیفی و کمی تقسیم می‌شوند. به متغیرهایی کمی می‌گوییم که فقط مقادیر عددی را می‌پذیرند و ضمناً عملیات ریاضی از قبیل جمع و معدل‌گیری برای آن‌ها قابل تعریف و بامعنی باشد. مثل نمره، قد، وزن، شدت زلزله، درجهٔ حرارت، درآمد افراد و مثلاً جمع درآمد دو نفر یا معدل درآمد دو نفر بامعنی است.

متغیرهای کیفی متغیرهایی هستند که از جنس عدد یا غیر عددند و در حالتی که از جنس عدد هستند، جمع دو مقدار از آن‌ها بی‌معنی است. مانند گروه خونی، جنسیت افراد و رنگ ماشین‌ها. در واقع از این متغیرها برای دسته‌بندی افراد یا اشیاء در گروه‌ها استفاده می‌شود. مثلاً در یک گروه کوهنوردی با توجه به جنسیت افراد می‌توانیم آن‌ها را به دو دسته مردان و زنان تقسیم کنیم. البته رتبه کنکور بچه‌ها و یا شماره دانش‌آموزان در لیست حضور و غیاب با آن که عدد هستند، کیفی محسوب می‌شوند، چون جمع رتبه دو نفر یا معدل رتبه دو نفر بی‌معنی است. در علم آمار، اندازه‌گیری به معنی ایجاد تفکیک بین افراد یا اشیاء یا جانداران است. دو نوزاد دوقلو شبیه هم را نام‌گذاری می‌کنیم تا آن‌ها را تفکیک کنیم. در واقع نوعی اندازه‌گیری انجام داده‌ایم. به طور کلی اندازه‌گیری‌ها بسته به دقتی که دارند به چهار مقیاس اسمی، ترتیبی، فاصله‌ای و نسبتی تقسیم می‌شوند. این مقیاس‌ها نشان می‌دهند که چه نوع محاسبه‌ای برای داده‌های مختلف مناسب هستند؛ از قبیل: ترتیب، محاسبه اختلاف و نسبت گرفتن.

مقیاس‌های متغیرهای کیفی

متغیرهای کیفی با دو مقیاس اسمی و ترتیبی اندازه‌گیری می‌شوند:

۱-۱) مقیاس اسمی: این مقیاس برای متغیرهای غیر عددی است که شامل نام‌ها، برچسب‌ها و گروه‌ها می‌شوند. ضمناً در این مقیاس، متغیرها دارای ترتیب خاصی نیستند و نمی‌توانیم آن‌ها را از کوچک تا بزرگ، ضعیف تا قوی و ... مرتب کنیم. مانند رنگ چشم افراد، نوع گروه خونی، ورزش مورد علاقه و ... مثلاً شما نمی‌توانید از قبل بگویید یک فرد ناشناس چه نوع گروه خونی دارد ولی می‌دانیم گروه خونی او جزء یکی از گروه‌های A، B، AB و O است. ضمناً گروه‌های خونی، دارای ترتیب نیستند.

۱-۲) مقیاس ترتیبی: این مقیاس برای متغیرهای از جنس عدد یا غیر عدد است که در آن‌ها نوعی ترتیب طبیعی وجود دارد. (قابل مرتب‌کردن و مقایسه‌کردن هستند) و ضمناً محاسبه اختلاف هر دو مقدار از داده‌ها امکان‌پذیر نیست و یا اصلاً بی‌معنی است. مثلاً حروف الفبای فارسی یا مراحل زندگی انسان (نوزادی، کودکی، نوجوانی و ...) و یا رتبه کنکور دانش‌آموزان، همگی دارای نوعی ترتیب هستند پس مقیاس آن‌ها ترتیبی است.

مقیاس‌های متغیرهای کمی

متغیرهای کمی با دو مقیاس فاصله‌ای و نسبتی اندازه‌گیری می‌شوند.

۱-۳) مقیاس فاصله‌ای: این مقیاس برای متغیرهای از جنس عدد است که اولاً قابل مرتب‌کردن و مقایسه‌کردن باشند و ثانیاً اختلاف بین هر دو مقدار از داده‌ها با معنی باشد. مثل درجه حرارت شهرهای مختلف. مثلاً اگر در حال حاضر تهران ۳۰ درجه سانتی‌گراد و اهواز ۵۰ درجه باشد اختلاف این دو عدد با معنی است و نشان می‌دهد اهواز ۲۰ درجه گرم‌تر از تهران است. ضمناً درجه حرارت شهرها را می‌توان از کوچک به بزرگ مرتب کرد.

۱-۴) مقیاس نسبتی: این مقیاس برای متغیرهای از جنس عدد است که اولاً قابل مرتب‌کردن و مقایسه‌کردن باشند، ثانیاً اختلاف بین هر دو مقدار از داده‌ها با معنی باشد، ثالثاً نسبت (تقسیم) هر دو مقدار از داده‌ها نیز با معنی و دقیق باشد؛ مثل قد، وزن، درآمد افراد و شدت زلزله. مثلاً فرض کنید وزن علی، امیر و رضا به ترتیب ۸۸، ۸۰ و ۴۴ کیلوگرم باشد، در این صورت این اعداد قابل مرتب‌کردن هستند و اختلاف دویه‌دوی آن‌ها و نسبت دویه‌دوی آن‌ها با معنی است، یعنی می‌توان گفت وزن علی ۸ کیلو از وزن امیر بیشتر است و یا می‌توان گفت نسبت وزن علی به وزن رضا برابر با $\frac{88}{44} = 2$ است، یعنی وزن علی دقیقاً ۲ برابر وزن رضاست.

پس در مقیاس نسبتی، عدد صفر، واقعی است نه قراردادی؛ مثلاً وقتی می‌گوییم وزن، صفر است، یعنی هیچ وزنی وجود ندارد یا وقتی می‌گوییم درآمد صفر است، یعنی واقعاً هیچ درآمدی وجود ندارد ولی مقیاس اندازه‌گیری درجه هوای شهرها فاصله‌ای است؛ چون صفر در آن قراردادی است، یعنی اگر دمای هوای شهری صفر درجه باشد، به این معنی نیست که این شهر، دمای هوا ندارد.

مثال: نوع و مقیاس اندازه‌گیری متغیرهای زیر را تعیین کنید:

۱-۱) اسامی افراد یک گروه ← نوع: کیفی - مقیاس: اسمی

۱-۲) رتبه ورزشکاران ایرانی در مسابقات المپیک ← نوع: کیفی - مقیاس: ترتیبی

۱-۳) طول ماهی‌های یک رودخانه خاص ← نوع: کمی - مقیاس: نسبتی

۱-۴) وعده‌های غذایی ما در طول یک شبانه‌روز (صبحانه، نهار، عصرانه، شام) ← نوع: کیفی - مقیاس: ترتیبی

۱-۵) دمای بدن ماهی‌های قزل‌آلا در یک رودخانه خاص ← نوع: کمی - مقیاس: فاصله‌ای

جمع‌بندی و خلاصه مطالب اخیر به شکل زیر است:



تعریف علم آمار

به مطالعه نحوه جمع‌آوری، سازمان‌دهی (طبقه‌بندی)، تحلیل و تفسیر داده‌ها جهت استخراج اطلاعات و تصمیم‌گیری، آمار می‌گوییم. البته جور دیگری هم می‌توان در مورد علم آمار سخن گفت. در علوم تجربی مثل کشاورزی و پزشکی، بین بعضی از پدیده‌هایشان روابطی وجود دارد که معمولاً نمی‌توان آن‌ها را به صورت ریاضی بیان کرد. علم آمار راهی برای بیان ریاضی چنین پدیده‌هایی است. مثلاً علم آمار در پزشکی، مطرح می‌کند که ۵ درصد افراد به واکسن آنفلوآنزا واکنش شدید نشان می‌دهند و یا افراد سیگاری ۲ برابر بقیه در معرض سرطان هستند.

تعریف پارامتر و آماره

پارامتر یک مشخصه عددی است که بیان‌کننده ویژگی خاصی از کل جامعه است. مثلاً در یک گروه کوهنوردی بزرگ نسبت مردان به کل کوهنوردان (یعنی کسر $\frac{\text{تعداد مردان این جامعه}}{\text{تعداد کل این جامعه}}$) پارامتر نامیده می‌شود. مقدار پارامتر همیشه ثابت و بدون تغییر است. مثلاً اگر گروه کوهنوردی شامل ۱۰۰۰ نفر بوده که ۷۰۰ نفر آن‌ها مرد باشند، نسبت $\frac{۷۰۰}{۱۰۰۰}$ پارامتری است که بیان‌کننده ویژگی مرد بودن اعضا است. ولی معمولاً بررسی تمام اعضای جامعه، دشوار و گاهی غیرممکن است. به همین دلیل در اکثر موارد به جای پارامتر از آماره استفاده می‌کنیم. آماره هم یک مشخصه عددی است که بیان‌کننده ویژگی خاصی از نمونه است. در مثال کوهنوردان، اگر نمونه‌ای تصادفی از اعضا انتخاب کنیم، در این نمونه، نسبت تعداد مردان به تعداد کل اعضای نمونه، یک آماره می‌باشد. مثلاً فرض کنید از ۱۰۰۰ نفر اولیه ۸۰ نفر را تصادفاً انتخاب کرده‌ایم که ۵۰ نفر آن‌ها مرد هستند، در این صورت نسبت $\frac{۵۰}{۸۰} = \frac{\text{تعداد مردان این نمونه}}{\text{تعداد کل این نمونه}}$ آماره‌ای است که بیان‌کننده ویژگی مرد بودن اعضای نمونه است. البته اگر نمونه‌ای دیگر را انتخاب کنیم حاصل کسر $\frac{\text{تعداد مردان}}{\text{تعداد کل نمونه}}$ تغییر خواهد کرد.

نتیجه: مقدار پارامتر همیشه ثابت و بدون تغییر است ولی مقدار آماره ثابت نیست و از نمونه‌ای به نمونه دیگر متفاوت است. ولی چون تعداد اعضای جوامع آماری زیاد هستند، معمولاً مقدار پارامتر مجهول است (قابل محاسبه نیست) به همین علت از آماره‌ها برای تخمین پارامترها استفاده می‌کنند.

مثالی دیگر از پارامتر و آماره: فرض کنید در یک زمین کشاورزی، هندوانه‌ها آماده برداشت هستند. اگر متغیر مورد بررسی «وزن» هندوانه‌ها باشد، میانگین وزن تمام هندوانه‌ها پارامتر است. ولی چون محاسبه وزن تک‌تک هندوانه‌ها و سپس میانگین آن‌ها کاری بسیار وقت‌گیر است، نمونه‌های تصادفی از هندوانه‌ها انتخاب می‌شوند. مثلاً فرض کنید ۳ نمونه انتخاب شود که هر کدام شامل ۲۰ هندوانه است. اگر میانگین وزن ۲۰ هندوانه نمونه اول برابر ۸ کیلوگرم، میانگین وزن ۲۰ هندوانه نمونه دوم برابر ۷ کیلوگرم و میانگین وزن ۲۰ هندوانه نمونه سوم برابر ۹ کیلوگرم باشد، به این ۳ عدد، آماره می‌گوییم. حال به کمک این ۳ آماره می‌توانیم حدس بزنیم که میانگین وزن تمام هندوانه‌ها (پارامتر) ۸ کیلوگرم است. یعنی از چند آماره به یک پارامتر رسیده‌ایم. $(\text{میانگین} = \frac{۷+۸+۹}{۳} = ۸)$



تذکره: در مثال وزن هندوانه‌ها دیدید که هم در تعریف پارامتر و هم در آماره، میانگین (متوسط) وزن‌ها را در نظر گرفتیم ولی در مثال کوهنوردان نسبت $\frac{\text{تعداد مردان}}{\text{تعداد کل}}$ را هم در تعیین پارامتر و هم در تعیین آماره در نظر گرفتیم. برای محاسبه پارامتر و آماره به طور کلی در متغیرهای کیفی، نسبت‌ها را حساب می‌کنیم ولی در متغیرهای کمی، میانگین (متوسط) را در نظر می‌گیریم. (وزن هندوانه‌ها، متغیر کمی است ولی جنسیت افراد کوهنوردی، کیفی است).

مثال و پاسخ

مثال در یک دانشکده، نسبت دختران به کل دانشجویان $\frac{145}{300}$ می‌باشد. در کلاس فیزیک، نسبت دختران به کل افراد کلاس $\frac{25}{48}$ و در کلاس ادبیات نسبت دختران به کل افراد کلاس $\frac{19}{44}$ است. مشخص کنید از کسرهای ذکر شده، کدام پارامتر و کدام آماره است؟

پاسخ کسر $\frac{145}{300}$ پارامتر است؛ چون کل افراد دانشکده از نظر جنسیت مورد بررسی قرار گرفته‌اند ولی کسرهای $\frac{25}{48}$ و $\frac{19}{44}$ آماره هستند؛ چون دو نمونه از کل جامعه آماری مورد مطالعه قرار گرفته‌اند. (کلاس‌های فیزیک و ادبیات)

سؤال‌های امتحانی

(مدارس تهران)

۱- جمع‌آوری داده‌ها به کدام طریق، قابل قبول نیست؟

- (۱) مصاحبه (۲) مشاهده (۳) دادگان (۴) پرسش هدایت‌کننده

(مدارس تهران)

۲- بررسی کدام موضوع از طریق داده‌های از پیش تهیه‌شده (دادگان) صورت می‌گیرد؟

- (۱) سن ازدواج اشخاص (۲) علاقه اشخاص به انجام نوعی ورزش (۳) رابطه رژیم گرفتن با اختلالات عصبی (۴) رنگ مورد علاقه اکثر مردم برای خرید اتومبیل

(مدارس تهران)

۳- نوع متغیر و مقیاس میزان آلودگی هوا، کدام است؟

- (۱) کمی - نسبتی (۲) کمی - فاصله‌ای (۳) کیفی - اسمی (۴) کیفی - ترتیبی

(مدارس تهران)

۴- آلاینده‌گی هوا چگونه متغیری است؟ مقیاس اندازه‌گیری آن کدام است؟

- (۱) کمی - فاصله‌ای (۲) کمی - نسبتی (۳) کیفی - اسمی (۴) کیفی - ترتیبی

(مدارس تهران)

۵- در عبارت زیر، متغیر تصادفی کدام است؟

«آمارگری می‌خواهد در ساعت ۸ صبح ۲۱ فروردین ۹۵ درجه حرارت هوا را در شهر گیلان بررسی کند.»

- (۱) ۲۱ فروردین (۲) سال ۱۳۹۵ (۳) درجه حرارت (۴) شهر گیلان

(تمرین کتاب درسی)

۶- الف) داده‌های در سطح اسمی، کمی هستند یا کیفی؟ ب) داده‌های در سطح فاصله‌ای چه‌طور؟

۷- در یک جامعه آماری آیا ممکن است یک پارامتر تغییر کند؟ اگر ۳ نمونه با اندازه یکسان از جامعه داشته باشیم، آیا مقادیر آماره‌های این ۳ نمونه برابر است؟

(مشابه تمرین کتاب درسی)

۸- در یک مطالعه از ۱۸۰۰ مشتری یک رستوران گیاهی، سؤال شده که برای کدام وعده غذایی (صبحانه یا ناهار یا شام) سفارش داده‌اند؟ (الف) متغیر را مشخص کنید. (ب) این متغیر کمی است یا کیفی؟ مقیاس اندازه‌گیری آن را تعیین کنید.

(ج) جامعه آماری در این بررسی چیست؟ (د) در این مطالعه، سرشماری کرده‌ایم یا نمونه‌گیری؟

(متن کتاب درسی)

۹- کدام مورد زیر، از ویژگی‌های متغیرهای کمی است؟

- (۱) داده‌های مربوط به آن‌ها، همیشه اعدادی صحیح هستند. (۲) اعضای جامعه یا نمونه را می‌توان از لحاظ یک متغیر کمی با هم مقایسه کرد. (۳) اختلاف بین مقادیر آن‌ها، بی‌معنی است. (۴) از جنس عدد و رقم نیستند، بلکه شامل نام‌ها، برچسب‌ها و گروه‌ها هستند.

(متن کتاب درسی)

۱۰- کدام مورد زیر، از ویژگی‌های متغیر کیفی است؟

- (۱) قابل اندازه‌گیری یا شمارش بوده و اختلاف بین مقادیر آن‌ها بی‌معنی است. (۲) داده‌های مربوط به آن‌ها، همیشه اعداد طبیعی هستند. (۳) به دو دسته اسمی و ترتیبی تقسیم می‌شوند. (۴) به دو دسته فاصله‌ای و نسبتی تقسیم می‌شوند.

(متن کتاب درسی)

۱۱- از چه نوع متغیری و با چه مقیاسی برای شماره‌گذاری صندوق‌های شرکت کنندگان در امتحانات استفاده می‌شود؟

- (۱) کیفی اسمی (۲) کیفی ترتیبی (۳) کمی فاصله‌ای (۴) کمی نسبتی

(متن کتاب درسی)

۱۲- برای آن که نمونه به درستی نمایانگر خصوصیت تمام جامعه باشد لازم است که

- (۱) به اندازه کافی بزرگ باشد و اعضای نمونه به صورت تصادفی انتخاب شوند. (۲) به اندازه کافی کوچک باشد و اعضای نمونه به صورت تصادفی انتخاب شوند. (۳) اعضای آن با یک الگو و قانون خاصی انتخاب شوند. (۴) تمام افراد جامعه را برای بررسی انتخاب کنیم.

(مدارس تهران)

۱۳- نوع متغیر «میزان آب هدررفته از شیرهای خراب شهر» با نوع متغیر کدام گزینه متفاوت است؟

- (۱) مدت زمان رسیدن بسته‌های پستی
(۲) میزان آلودگی هوا
(۳) قد دانش‌آموزان یک کلاس
(۴) رتبه کشورها از نظر رفاه اقتصادی

(متن کتاب درسی)

۱۴- از متغیرهای ذکرشده در زیر، چند مورد آن‌ها متغیر کمی است؟

- رنگ مو - قد - وزن - میزان آلودگی هوا - مدت مکالمه - نوع تلفن (سیار یا ثابت) - سطح سواد (سیکل، دیپلم، فوق دیپلم، لیسانس و ...) - سن دانشجویان - گنجایش یک مخزن - وضعیت مسکن (مالک مستأجر)
(۱) ۴ (۲) ۵ (۳) ۶ (۴) ۷

(متن کتاب درسی)

۱۵- چند مورد از متغیرهای زیر، کیفی هستند؟

- مقاومت یک لامپ - تعداد بیماران یک بیمارستان - رتبه کنکور - جنسیت افراد - مراحل زندگی انسان - اندازه محیط شکل‌های هندسی - مزه غذا (خوب، متوسط، بد)
(۱) ۵ (۲) ۴ (۳) ۳ (۴) ۲

(مدارس تهران)

۱۶- در کدام گزینه، شماره‌های (۱) تا (۴) جدول به ترتیب از راست به چپ به درستی پر شده‌اند؟

سابقه کار دبیران	مراحل کشت محصولات کشاورزی	میزان سرعت ماشین‌ها	وضع تأهل	متغیر تصادفی
(۱)	(۲)	(۳)	(۴)	نوع متغیر و مقیاس آن

(۱) کتی نسبتی - کیفی اسمی - کتی فاصله‌ای - کیفی ترتیبی - کتی نسبتی - کیفی اسمی

(۳) کتی نسبتی - کیفی ترتیبی - کتی نسبتی - کیفی اسمی - کتی فاصله‌ای - کیفی اسمی - کتی فاصله‌ای - کیفی ترتیبی

(مشابه تمرین و کار در کلاس کتاب درسی)

۱۷- در هر قسمت، مشخص کنید کدام روش برای جمع‌آوری داده‌ها مناسب‌تر است؟

- (الف) تاریخ تولد همه دانش‌آموزان مدرسه‌ای خاص در پایه دهم
(ب) تعداد قلم‌های هر دانش‌آموز یک کلاس
(ج) نحوه استفاده جوانان از اوقات فراغت
(د) تعداد نوزادانی که در استان تهران در خرداد ماه ۹۴ متولد شده‌اند.
(ه) بررسی اطلاعات عقیدتی و علمی داوطلبان ورود به ارتش
(و) میزان رضایت مشتریان از عملکرد کارمندان یک اداره
(ز) دانشجویان یک دانشکده، شب گذشته چند ساعت خوابیده‌اند؟
(ط) مردان ایرانی، بیشتر از چه شامپویی استفاده می‌کنند.
(ح) درصد بیکاران استان خوزستان

(مشابه مثال کتاب درسی)

۱۸- در عبارت‌های زیر، جامعه آماری، نمونه تصادفی، متغیر تصادفی، پارامتر و آماره را مشخص کنید.

(الف) «به منظور تعیین وزن متوسط هندوانه‌های یک باغ، ۱۰۰ عدد هندوانه را بدون هیچ نظم خاصی جمع‌آوری و وزن می‌کنیم. سپس وزن متوسط کل هندوانه‌ها را تخمین می‌زنیم.»

(ب) «در یک کارخانه تولید هواپیماهای کنترل‌ی، روزانه ۱۰۰۰ هواپیما تولید می‌شود. می‌خواهیم بُرد پروازی آن‌ها را مورد بررسی قرار دهیم (بُرد زیاد، بُرد متوسط، بُرد کم). یک نمونه ۲۰ تایی از هواپیماها را به طور تصادفی از خط تولید برمی‌داریم. در این نمونه ملاحظه می‌کنیم که ۱۰ هواپیما بردشان زیاد، ۷ هواپیما بردشان متوسط و ۳ هواپیما بردشان کم است.»

۱۹- یک شبکه تلویزیونی می‌خواهد در یک شهر ۱,۰۰۰,۰۰۰ خانواری، نسبت تعداد خانواده‌هایی را که کانال خاصی را حداقل ۳ ساعت در هفته تماشا می‌کنند، به کل خانواده‌ها تخمین بزند. برای این منظور به طور تصادفی یک گروه شامل ۲۰۰۰ خانواده را مورد بررسی قرار می‌دهد:

(مشابه کار در کلاس کتاب درسی)

(الف) جامعه آماری، اندازه جامعه، نمونه تصادفی، متغیر تصادفی و اندازه نمونه تصادفی را مشخص کنید.

(ب) نوع متغیر موجود و مقیاس اندازه‌گیری آن را مشخص کنید.

(مشابه تمرین کتاب درسی)

۲۰- نوع متغیرها را در عبارات زیر مشخص کنید:

- (الف) رتبه دانش‌آموزان در مسابقات ورزشی
(ب) سن نمایندگان مجلس دوره یازدهم
(ج) گزینه‌های موردنظر برای پاسخ به یک سؤال: خیلی خوب، خوب، متوسط، بد، خیلی بد
(د) وضعیت تأهل دبیران یک آموزشگاه
(ه) زمان اولین کلاس دانشگاه‌های مختلف در روز شنبه
(و) مدت پاسخ‌گویی به سؤالات یک امتحان
(ز) رشته تحصیلی دانشجویان یک دانشگاه
(ح) نمره آخرین آزمون دانشگاه
(ط) مقیاس ارزیابی تحصیلی: ضعیف، معمولی، خوب
(ی) تاریخ تولد افراد (سال تولد)

(تمرین کتاب درسی)

۲۱- فرق بین داده و متغیر چیست؟ فرق بین آماره و پارامتر چیست؟ برای هر یک مثالی بیاورید.

۲۲- در مطالعه کدام یک از موارد زیر، نیازی به نمونه‌گیری نیست؟ (می‌توانیم سرشماری انجام بدهیم.)

(۱) بررسی وزن و طول ماهی‌های خاویار دریای خزر (۲) بررسی حداکثر ولتاژ قابل تحمل لامپ‌های تولیدی یک کارخانه

(۳) بررسی کیفیت یک نوع برنج موجود در انبار (۴) بررسی معدل ترم اول دانش‌آموزان یک کلاس

(برگرفته از متن کتاب درسی)

۲۳- به عبارت زیر دقت کنید:

«در یک مزرعه هندوانه، ۱۰,۰۰۰ هندوانه قابل برداشت هستند. اگر متغیر تصادفی، مزرعه هندوانه‌ها باشد (خوب، معمولی، بد) و یک نمونه ۵۰ تایی از هندوانه‌ها را به طور تصادفی انتخاب کنیم، در این نمونه، نسبت هندوانه‌های با مزرعه خوب برابر $\frac{1}{5}$ می‌باشد.»

(الف) نوع متغیر در متن بالا و مقیاس اندازه‌گیری آن چیست؟

(ب) نسبت $\frac{1}{5}$ آماره است یا پارامتر؟

(ج) آیا می‌توانیم پارامتر مربوط به نسبت هندوانه‌های با مزرعه بد به کل هندوانه‌ها را مستقیماً حساب کنیم؟

۲۴- کدام یک از نمونه‌گیری‌های زیر، یک نمونه‌گیری تصادفی است. در هر قسمت، جامعه، نمونه، متغیر تصادفی، اندازه جامعه و اندازه نمونه را مشخص کنید.

(تمرین کتاب درسی)

(الف) مصاحبه با تمام پلیس‌های یک پاسگاه در مورد تخلفات

(ب) با بچه‌هایی که وارد یک پارک بازی می‌شدند، پنج در میان مصاحبه شد تا وسیله بازی مورد علاقه‌شان را اعلام کنند.

(ج) برای بررسی pH شامپوهای تولیدی یک کارخانه، شامپوها را صد در میان مورد آزمایش قرار می‌دهیم.

۲ معیارهای گرایش به مرکز (حد وسط)

هدف از یادگیری این فصل، یافتن میانگین، میانه، مُد و نحوه استفاده از هر یک از آن‌هاست. در واقع میانگین، میانه و مُد معیارهایی هستند که محل تمرکز داده‌ها را نشان می‌دهند.

مُد یا نما

مُد، داده‌ای است که بیشترین فراوانی (تکرار) را داشته باشد. البته مُد ممکن است منحصر به فرد نباشد؛ یعنی یک جامعه آماری، می‌تواند ۲ یا چند مُد هم داشته باشد. مثلاً در داده‌های ۱, ۱, ۲, ۲, ۲, ۳, ۳, ۵, ۸، مُد برابر ۲ است، چون بیشترین تکرار را دارد ولی در داده‌های ۱, ۱, ۲, ۲, ۳, ۳, ۵, ۸، هم ۱ و هم ۲ هر دو مُد محسوب می‌شوند چون هر دو ۳ بار تکرار شده‌اند. (پس مُد، می‌تونه منمهر به فرد نباشه.) در انتخابات، مُد تعیین‌کننده نتیجه انتخابات است؛ یعنی موضوعی رأی می‌آورد که بیشترین افراد به آن رأی داده‌اند. تولیدکنندگان و فروشندگان لباس هم، اجناس خود را براساس مُد تولید کرده و به فروش می‌رسانند.

نکاتی در مورد مُد

۱ از مُد برای تحلیل متغیرهای کیفی استفاده می‌شود.

۲ با جابه‌جایی داده‌ها در کنار یکدیگر، مُد تغییری نمی‌کند.

۳ اگر یک جامعه آماری، چند مُد داشته باشد، مُد شاخص معتبری برای تحلیل داده‌ها نخواهد بود.

۴ لزومی ندارد که بزرگ‌ترین داده، مُد هم باشد. ملاک تشخیص مُد، فقط داشتن بیشترین فراوانی است.

مثال پاسخ

مثال در جدول مقابل، مقدار مُد کدام است؟

داده‌ها x_i	۱	۲	۳	۴
فراوانی f_i	۸	۱۰	۳	۱۶

پاسخ عدد ۴ شانزده بار تکرار شده و طبق جدول، فراوانی‌اش از همه بیشتر است، پس مُد برابر ۴ است.

میانگین

میانگین همان معدل داده‌هاست. اگر $x_1, x_2, \dots, x_n$ داده‌های آماری باشند، میانگین آن‌ها را با $\bar{x}$ (بخوانید $\bar{x}$ بار) نمایش داده و خواهیم نوشت:

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

به عنوان مثال میانگین داده‌های آماری ۱۰، ۱۰، ۲۵، ۳۶، ۷۰ و ۱ و ۳ برابر است با:

$$\bar{x} = \frac{1+3+10+10+25+36+70}{7} = \frac{155}{7} \approx 22.14$$

مثال و پاسخ

مثال میانگین اعداد $3 + 10x$ و $x - 6$ و $5x + 3$ برابر $\frac{2}{3}$ است. مقدار x را به دست آورید.

پاسخ $\bar{x} = \frac{2}{3} \Rightarrow \frac{5x + 3 + x - 6 + 3 + 10x}{3} = \frac{2}{3} \Rightarrow 16x = 2 \Rightarrow x = \frac{2}{16} = \frac{1}{8}$

مثال و پاسخ

مثال میانگین داده‌های $a, 4, 4, 3$ و 2 برابر 2 است. میانگین داده‌های $(100 + 6a), 4, 4, 3$ و 2 کدام است؟

پاسخ 18 (۱) 19 (۲) 20 (۳) 21 (۴)

$\bar{x} = 2 \Rightarrow \frac{2 + 3 + 4 + 4 + a}{5} = 2 \Rightarrow \frac{13 + a}{5} = 2 \Rightarrow 13 + a = 10 \Rightarrow a = 10 - 13 = -3$

گزینه (۲) صحیح است. $\bar{x} = \frac{2 + 3 + 4 + 4 + (100 + 6a)}{5} = \frac{13 + 100 + 6(-3)}{5} = \frac{95}{5} = 19$

اولیه جدید

خواص میانگین

- ۱ میانگین یک سری از داده‌ها، همیشه بین کوچک‌ترین و بزرگ‌ترین داده قرار دارد.
- ۲ تمام داده‌ها چه کوچک و چه بزرگ در مقدار میانگین تأثیرگذار هستند.
- ۳ در هر جامعه آماری، میانگین منحصر به فرد است، یعنی هر جامعه آماری فقط یک میانگین دارد.
- ۴ اگر در یک سری از داده‌های آماری با میانگین $\bar{x}$ به جای تک‌تک داده‌ها، عدد $\bar{x}$ را قرار دهیم، مجموع داده‌ها تغییری نخواهد کرد. مثلاً در داده‌های $3, 3, 3, 3, 3 \Rightarrow$ مجموع $= 15$ اگر به جای تمام داده‌ها 3 قرار دهیم، آن‌گاه: $3, 3, 3, 3, 3 \Rightarrow$ مجموع $= 15$ می‌بینید که مجموع داده‌ها تغییری نکرده است.
- ۵ اگر همه داده‌ها برابر باشند، میانگین آن‌ها برابر یکی از آن‌ها است؛ مثلاً میانگین داده‌های $8, 8, 8$ برابر 8 است.
- ۶ اگر فاصله داده‌ها از هم ثابت باشد، برای محاسبه سریع میانگین آن‌ها می‌توانیم کوچک‌ترین داده را با بزرگ‌ترین داده جمع کرده تقسیم بر 2 کنیم؛ مثلاً در داده‌های $5, 7, 9, 11, 13, 15, 17, 19, 21$ داده‌ها 2 تا 2 زیاد می‌شوند (پس فاصله بینشون ثابت)، لذا: $\bar{x} = \frac{5 + 21}{2} = 13$

نکته با توجه به تعریف میانگین، با یک طرفین وسطین کردن ساده می‌توان فرمولی برای مجموع داده‌ها به صورت زیر نوشت:

تعداد داده‌ها $\times$ میانگین = مجموع داده‌ها

مثال و پاسخ

مثال میانگین 10 داده آماری $32/5$ است. اگر داده‌های 35 و 40 را از آن‌ها کنار بگذاریم، میانگین داده‌های حاصل را به دست آورید.

پاسخ $325 = 32 \times 10$ تعداد $\times$ میانگین = مجموع داده‌ها

حال باید دو عدد 35 و 40 را از 325 کم کنیم. $\frac{250}{8} = 31/25$ میانگین جدید = $\frac{\text{مجموع جدید}}{\text{تعداد جدید}}$ $\Rightarrow 325 - 40 - 35 = 250$ مجموع جدید

در این صورت تعداد داده‌ها از 10 به 8 می‌رسد:

میانگین وزن دار (موزون)

اگر در یک سری از داده‌های آماری، همه یا بعضی از داده‌ها ضریب (تکرار یا فراوانی) داشته باشند، برای محاسبه میانگین تمام داده‌ها به صورت زیر عمل می‌کنیم:

با فرض آن که فراوانی (تکرار یا ضریب) داده‌های $x_1, x_2, \dots, x_n$ به ترتیب برابر با $f_1, f_2, \dots, f_n$ باشد، میانگین کل داده‌ها برابر است با:

داده‌ها x_i	x_1	x_2	$\dots$	x_n
فراوانی f_i	f_1	f_2	$\dots$	f_n

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{N}, N = f_1 + f_2 + \dots + f_n$$

میانگین وزن دار

مثال میانگین جدول فراوانی مقابل را به دست آورید.

داده‌ها x_i	۲	۴	۵
فراوانی f_i	۷	۲	۱

$\Rightarrow N = 7 + 2 + 1 = 10$

$\bar{x} = \frac{14 + 8 + 5}{10} = \frac{27}{10} = 2/7$

(این جدول نشون می‌ده که ۷ تا عدد ۲، ۲ تا عدد ۴ و یک عدد ۵ وجود داره)

مثال پاسخ

حدود دسته‌ها	۲-۴	۴-۶	۶-۸	۸-۱۰
فراوانی	۵	۱۲	۱	۲

مثال میانگین جدول فراوانی مقابل را به دست آورید.

مرکز دسته‌ها	۳	۵	۷	۹
فراوانی	۵	۱۲	۱	۲

پاسخ وقتی در یک جدول، حدود دسته‌ها داده می‌شود باید مرکز دسته‌ها

را به دست آوریم؛ مثلاً مرکز دسته (۲-۴) برابر است با $\frac{۲+۴}{۲} = ۳$ به

همین ترتیب تمام مرکزها را حساب کرده و خواهیم داشت:

$$\bar{x} = \frac{۱۵ + ۶۰ + ۷ + ۱۸}{۲۰} = \frac{۱۰۰}{۲۰} = ۵$$

مثال پاسخ

مثال نمرات ریاضی، فیزیک، شیمی و عربی یک دانش‌آموز به ترتیب ۱۴، ۱۵، ۱۸ و ۲۰ می‌باشد. اگر ضرایب این چهار درس

به ترتیب ۴، ۳، ۲ و ۱ باشد، میانگین کل این دروس را به دست آورید.

نمرات	۱۴	۱۵	۱۸	۲۰
ضرایب (فراوانی‌ها)	۴	۳	۲	۱

پاسخ می‌توانیم برای این مسئله جدول مقابل را

رسم کنیم (ضرایب همان فراوانی‌ها هستند).

$$\bar{x} = \frac{(۴ \times ۱۴) + (۳ \times ۱۵) + (۲ \times ۱۸) + (۱ \times ۲۰)}{۱۰} = \frac{۵۶ + ۴۵ + ۳۶ + ۲۰}{۱۰} = \frac{۱۵۷}{۱۰} = ۱۵.۷$$

مثله

پس از مرتب‌کردن داده‌های آماری، میانه عددی است که تعداد داده‌های قبل از آن با تعداد داده‌های بعد از آن مساوی است. به عبارت دیگر اگر داده‌ها متمایز باشند، میانه از ۵۰ درصد داده‌ها بزرگ‌تر و از ۵۰ درصد دیگر کوچک‌تر است. برای یافتن میانه ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم. اگر تعداد داده‌ها فرد باشد، عدد وسط، میانه است و اگر تعداد زوج باشد، میانگین دو عدد وسط، میانه است.

مثال پاسخ

مثال میانه داده‌های زیر را که مربوط به نمرات درس ریاضی دانش‌آموزان یک کلاس است، به دست آورید.

۴، ۷، ۷، ۳، ۱۲، ۱۱، ۱۷، ۱۵، ۱۴، ۱۷، ۱۹، ۱۴، ۱۰، ۹، ۵

مرتب‌کردن داده‌ها
از کوچک به بزرگ $\rightarrow ۳، ۴، ۵، ۷، ۷، ۹، ۱۰، ۱۱، ۱۲، ۱۴، ۱۴، ۱۵، ۱۷، ۱۷، ۱۹$

پاسخ

تعداد داده‌ها ۱۵ تا است، پس چون تعداد، عددی فرد است، داده وسط (داده هشتم) همان میانه است؛ یعنی: $x_8 = ۱۱$ = میانه (x_8 یعنی داده هشتم)

مثال پاسخ

۱، ۲، ۴، ۴، ۸، ۸، ۹، ۱۰، ۱۰، ۲۰، ۳۰، ۴۰

مثال میانه داده‌های مقابل را به دست آورید.

پاسخ داده‌ها خودشان به صورت مرتب‌شده می‌باشند و تعداد آن‌ها ۱۲ تا است، پس باید داده‌های ششم و هفتم را جمع و نصف کنیم:

$$\text{میانه} = \frac{x_6 + x_7}{۲} = \frac{۸ + ۹}{۲} = \frac{۱۷}{۲} = ۸.۵$$

خواص میانه

۱ شرط این که میانه، عضوی از داده‌ها باشد این است که تعداد داده‌ها فرد باشد. یعنی اگر تعداد داده‌ها زوج باشد، میانه ممکن است عضوی از داده‌ها باشد یا نباشد. (اگر دو عدد وسط مساوی باشند، میانه، عضوی از داده‌ها می‌شود.)

۲ به داده‌های ۱، ۳، ۵، ۸، ۲۰۰۰ و هم‌چنین داده‌های ۲۹۰۰، ۲۸۰۰، ۲۷۰۰، دقت کنید. در گروه اول عدد ۲۰۰۰ خیلی بزرگ‌تر از بقیه داده‌هاست و در گروه دوم عدد ۱ خیلی کوچک‌تر از بقیه داده‌هاست. به این‌گونه اعداد، داده‌های دورافتاده (پرت) می‌گوییم. در مواقعی که در یک سری از داده‌های آماری، داده یا داده‌های پرت وجود داشته باشد، میانگین، شاخص خوبی برای نشان دادن محل تمرکز داده‌ها نیست و باید از میانه استفاده کنیم. مثلاً در داده‌های ۱، ۳، ۵، ۸، ۲۰۰۰ و ۱ میانگین تقریباً برابر با ۴۰۳ است ولی میانه برابر با ۵ است. واضح است که عدد ۵ خیلی بهتر تمرکز و تراکم داده‌ها را نشان می‌دهد.

مثال و پاسخ

مثال برای داده‌های ۳۰۰۰۰، ۲۹۰۰۰، ۲۸۰۰۰، ۲۵۰۰۰، ۲۳۰۰۰، ۱۸۰۰۰، ۱۶۰۰۰، ۱۴۰۰۰، ۴ و ۳ چه شاخصی به عنوان شاخص حد وسط (محل تمرکز داده‌ها) مناسب‌تر است؟ مقدار آن را محاسبه کنید.

پاسخ داده‌های ۳ و ۴ دورافتاده (پرت) هستند؛ چون از بقیه داده‌ها خیلی خیلی کوچک‌ترند. لذا شاخص حد وسط، همان میانه است. ضمناً داده‌ها مرتب‌شده هستند و تعداد آن‌ها ۱۰ تا است. پس میانه برابر است با:

$$\text{میانه} = \frac{x_5 + x_6}{2} = \frac{18000 + 23000}{2} = 20500$$

خاصیت مشترک میانه، میانگین و مُد

در یک سری از داده‌ها، میانه، میانگین و مُد، نسبت به ۴ عمل اصلی (جمع، تفریق، ضرب و تقسیم) تأثیرپذیر هستند؛ یعنی هر کاری که روی داده‌ها انجام شود، همان کار را باید روی میانه، میانگین و مُد هم پیاده کنیم. مثلاً اگر تمام داده‌ها را در عدد ۱۰ ضرب کنیم، میانه، میانگین و مُد این داده‌ها نیز در عدد ۱۰ ضرب می‌شوند.

مثال و پاسخ

مثال اگر میانگین داده‌های (a, b, c) برابر ۱۲ باشد، میانگین داده‌های $(3a+1, 3b+1, 3c+1)$ را به دست آورید.

پاسخ اگر به داده‌های اولیه و داده‌های جدید کمی دقت کنیم، متوجه می‌شویم که هر یک از داده‌های اولیه، ابتدا در عدد ۳ ضرب و سپس با عدد ۱ جمع شده‌اند. پس برای میانگین هم، همین اتفاق رخ خواهد داد؛ یعنی:

$$3 \times 12 + 1 = 37 = \text{میانگین جدید}$$

مثال و پاسخ

مثال اگر میانه داده‌های $x_1, x_2, \dots, x_n$ برابر ۱۰ باشد، میانه داده‌های $\frac{x_1}{4} - 4, \frac{x_2}{4} - 4, \dots, \frac{x_n}{4} - 4$ را به دست آورید.

پاسخ کاملاً واضح است که هر یک از داده‌های اولیه ابتدا در $(\frac{1}{4})$ ضرب و سپس ۴ واحد از آن‌ها کم شده است. لذا همین بلاها را سر میانه هم می‌آوریم:

$$1 = 5 - 4 = \frac{1}{4} \times 10 - 4 = \text{میانه جدید}$$

سؤال‌های امتحانی

۲۵- جاهای خالی را با کلمات مناسب پُر کنید:

- (الف) آن چه اکثر مردم «حد وسط» می‌نامند، نزد آمارشناسان به معروف است.
- (ب) اگر در داده‌های آماری، داده دورافتاده داشته باشیم، باید برای تعیین حد وسط به سراغ برویم.
- (ج) وقتی با مواجه هستیم، معمولاً بازتاب بهتری از داده‌ها می‌دهد تا
- (د) در واقع نقطه وسط فهرست اعداد مرتب‌شده است. نیمی از اعداد بزرگ‌تر از و نیمی دیگر کوچک‌تر از هستند.
- (هـ) به طور کلی برای برنامه‌ریزی و تصمیم‌گیری، بهتر از است.

۲۶- در یک سری از داده‌های آماری، کم‌ترین داده ۱۲ و بیشترین داده ۲۴ است. میانگین تمام داده‌ها کدام عدد می‌تواند باشد؟

- ۱۱ (۱) ۱۸ (۲) ۲۵ (۳) ۲۸ (۴)

۲۷- میانگین اعداد $3 - 2x + 1$ و $7x + 1$ برابر ۱۶ است. مقدار x را به دست آورید.

۲۸- در داده‌های $(1, 5, 7, 8, x, 12, 14, 15)$ که از کوچک به بزرگ مرتب شده‌اند، میانه برابر با ۹ است. مقدار x را به دست آورید. (مدارس تهران)

۲۹- میانگین وزن ۱۰ نفر ۶۵ کیلوگرم است. دو نفر جمعاً به وزن ۱۴۲ کیلوگرم به این افراد اضافه می‌شوند. میانگین جدید برحسب گرم چه قدر است؟

- ۶۷۰۰۰ (۱) ۶۹۰۰۰ (۲) ۶۵۰۰۰ (۳) ۶۶۰۰۰ (۴) (مدارس تهران)

۳۰- اگر میانگین داده‌های $16, 9, 17, 13, 10, a, 10, 17, 11, 16$ برابر ۱۳/۱ باشد، میانه این داده‌ها را تعیین کنید. (برگرفته از متن کتاب درسی)

۳۱- میانه تعدادی داده برابر با ۷ است. به هر داده ۴ واحد اضافه می‌کنیم. میانه داده‌های جدید کدام است؟ (مدارس تهران)

- ۷ (۱) ۴ (۲) ۱۱ (۳) ۴ (۴) قابل محاسبه نیست.

۳۲- میانگین داده‌های $(x_1, x_2, \dots, x_n)$ برابر ۴ است. میانگین داده‌های $\frac{3}{4}x_1 + 7, \frac{3}{4}x_2 + 7, \dots, \frac{3}{4}x_n + 7$ کدام است؟ (مدارس تهران)

- ۱۳ (۱) ۱۲ (۲) ۱۱ (۳) ۱۰ (۴)

۳۳- میانۀ داده‌های $(x_1, x_2, \dots, x_n)$ برابر ۱۰ است. میانۀ داده‌های $-4x_1 + 5, -4x_2 + 5, \dots, -4x_n + 5$ کدام است؟ (مدارس تهران)

۲۰ (۱) ۱۰ (۲) ۳۰ (۳) ۱۵ (۴)

۳۴- معدل ۶ درس یک دانش‌آموز برابر ۱۷ بوده ولی بعداً متوجه شده‌ایم که نمره ۲ درس دیگر او را در نظر نگرفته‌ایم. اگر نمره این دو درس ۱۹ و ۱۵ باشد، معدل واقعی او چه قدر است؟ (برگرفته از متن کتاب درسی)

۳۵- میانگین ۸ داده برابر ۱۴ است. اگر یکی از اعداد کنار گذاشته شود، میانگین ۷ داده باقی‌مانده ۱۵ می‌شود. عدد کنار گذاشته شده کدام است؟ (مدارس تهران)

۳۶- میانگین ۱۰ داده آماری برابر $13/2$ است. اگر داده‌های ۵ و ۷ را از بین آن‌ها بیرون بکشیم، میانگین داده‌های باقی‌مانده را به دست آورید. (برگرفته از متن کتاب درسی)

۳۷- در ۴۵ داده آماری، مقدار میانگین ۱۱۲۴ محاسبه شده است. در بررسی مجدد داده‌ها متوجه می‌شویم که به جای داده ۱۰۲۴ عدد ۱۲۰۴ منظور شده است. با رفع اشتباه، میانگین واقعی چه قدر است؟ (مدارس تهران)

۳۸- اگر میانگین داده‌های $(x_1, x_2, \dots, x_7)$ برابر ۱۰ و میانگین داده‌های $(x_1, x_2, \dots, x_6)$ برابر ۸ باشد، مقدار داده x_7 کدام است؟ (مفهوم علاقه‌مندان)

۳۹- میانگین چهار داده برابر ۸ و میانگین ۸ داده دیگر ۴ است. میانگین کل داده‌ها چه قدر است؟ (مدارس تهران)

۱۶ (۱) ۶ (۲) ۳ (۳) ۴ (۴)

۴۰- اگر میانگین داده‌های $a, a, a, a, a+1$ برابر $\frac{3a}{4}$ باشد، میانگین داده‌های $a, a+1, a+2$ را به دست آورید. (مدارس تهران)

۴۱- در هر جامعه آماری، کدامیک از شاخص‌های زیر، ممکن است منحصر به فرد نباشد؟ (برگرفته از متن کتاب درسی)

۱) میانه ۲) مُد ۳) میانگین ۴) دامنه تغییرات

۴۲- نمرات درس فیزیک دانش‌آموزی در طول سال به صورت مقابل است:

۱۷, ۱۹, ۱۸, ۷, ۱۸, ۱۹

الف) میانه و میانگین نمرات وی را به دست آورید.

ب) کدامیک از شاخص‌های فوق، وضعیت او را در درس فیزیک، بهتر نشان می‌دهد؟

ج) اگر معلم درس فیزیک، برای جبران نمره ۷ دوباره از او امتحان بگیرد، برای آن که میانگین وی در این درس ۱۸ شود، او در این امتحان چه نمره‌ای باید کسب کند؟ (مشابه تمرین کتاب درسی)

۴۳- در آزمون استعداد تحصیلی از ۱۰ دانش‌آموز، به ترتیب یک نفر نمره ۷۵، چهار نفر نمره ۸۰، دو نفر نمره ۸۵ و سه نفر دیگر نمره ۹۰ کسب کرده‌اند. میانگین نمرات آزمون آن‌ها کدام است؟ (برگرفته از متن کتاب درسی)

۸۲/۵ (۱) ۸۳ (۲) ۸۳/۵ (۳) ۸۴ (۴)

۴۴- در جدول داده‌های مقابل، تفاوت مُد از میانگین کدام است؟ (مدارس تهران)

۴۵- طبق جدول زیر، اگر بخواهیم میانگین کل ۷۵ شود، نمره ادبیات چه قدر باید باشد؟ (برگرفته از متن کتاب درسی)

اختصاصی	زبان	معارف	ادبیات	درس
۷۰	۸۱	۹۰	?	نمره از ۱۰۰
۸	۳	۲	۴	ضریب

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

مدیر شرکت	معاون	حسابدار	بازاریاب	فروشنده (۲ نفر)	منشی	مستخدم	سمت
۳۵	۲۰	۳	۲/۵	۱/۳	۱/۲	۰/۷	حقوق (میلیون تومان)

مدیر شرکت در آگهی دعوت به همکاری در روزنامه، اعلام کرده که میانگین پرداخت حقوق در این شرکت تومان است. آیا این عدد می‌تواند دلیلی برای پرداخت حقوق بالا در این شرکت باشد؟ چرا؟ کدامیک از معیارهای گرایش به مرکز برای نشان دادن وضع پرداختی کارکنان این شرکت مناسب‌تر است؟

۴۷- در هفت داده آماری $x, 110, 75, 80, 85, 100$ و ۹۰ اگر میانگین، میانه و مُد همگی با هم برابر باشند، مقدار x را به دست آورید. (مفهوم علاقه‌مندان)

۴۸- در جدول مقابل، اگر فراوانی نسبی دسته دوم ۰/۴ باشد، میانگین داده‌ها را به دست آورید. (مفهوم علاقه‌مندان)

فراوانی آن داده = مجموع فراوانی‌ها

راهنمایی: $\text{فراوانی آن داده} = \frac{\text{مجموع فراوانی‌ها}}{\text{فراوانی نسبی یک داده}}$

x_i	۲	۳	۴	۵
f_i	۱	۳	x	۲

شاخص‌های پراکندگی



هدف از مطالعه این بخش، یادگیری محاسبه دامنه تغییرات، دامنه میان چارکی، انحراف استاندارد (انحراف معیار) و واریانس می‌باشد. این شاخص‌ها میزان دوری یا نزدیکی داده‌ها به یکدیگر را نشان می‌دهند. یعنی هر یک از چهار شاخصی که نام بردیم، هر چه قدر کم‌تر باشند، بیانگر این موضوع هستند که داده‌ها به هم نزدیک‌ترند و هر چه قدر مقدارشان بیشتر باشد، بیانگر دور بودن داده‌ها از هم می‌باشند. (بچه‌های عزیز کتاب در فصل پنجم دامنه تغییرات را بررسی کرده ولی چون جزء شاخص‌های پراکندگی است ما آن را در این فصل آموزش می‌دهیم).

دامنه تغییرات

$$R = \max - \min$$

اختلاف بزرگ‌ترین و کوچک‌ترین داده‌ها می‌باشد و آن را با R نمایش می‌دهیم؛ یعنی می‌توانیم چنین بنویسیم:

به عنوان مثال دامنه تغییرات داده‌های $۱۰، ۱۷، ۳، ۲۰، ۱۴$ و ۸ برابر است با $R = ۴۰ - ۱ = ۳۹$.

نکته دامنه تغییرات، معیاری سریع برای محاسبه پراکندگی است ولی معیار خوبی نیست؛ چون در محاسبه آن فقط از کم‌ترین و بیشترین داده استفاده می‌کنیم و بقیه داده‌ها عملاً بی‌مصرف خواهند ماند.

مثال و پاسخ

مثال در یک سری از داده‌های آماری، ۱۰ واحد به بزرگ‌ترین داده اضافه کرده و ۱۰ واحد از کوچک‌ترین داده کم می‌کنیم. دامنه تغییرات، چه تغییری خواهد کرد؟

(۱) تغییر نمی‌کند. (۲) ۱۰ واحد زیاد می‌شود. (۳) ۱۰ واحد کم می‌شود. (۴) ۲۰ واحد زیاد می‌شود.

$$R_{\text{اولیه}} = \max - \min$$

پاسخ

$$R_{\text{جدید}} = (\max + ۱۰) - (\min - ۱۰) = \max + ۱۰ - \min + ۱۰ = \underbrace{\max - \min}_{R_{\text{اولیه}}} + ۲۰ = R_{\text{اولیه}} + ۲۰ \Rightarrow \text{گزینه (۴) صحیح است.}$$

تذکره اگر تمام داده‌ها با هم برابر باشند، دامنه تغییرات آن‌ها برابر صفر است و برعکس؛ یعنی اگر دامنه تغییرات یک سری از داده‌ها صفر باشد، تمام داده‌ها با هم مساوی‌اند. علت این است که وقتی تمام داده‌ها برابر باشند، $\max$ و $\min$ هم برابرند و در نتیجه حاصل $\max - \min$ برابر صفر است.

مثال و پاسخ

مثال اگر دامنه تغییرات داده‌های $۸، x_1، x_2، x_3$ برابر صفر باشد، میانگین داده‌های $۷x_3، ۶x_2، ۵x_1$ را به دست آورید.

$$\begin{aligned} \text{پاسخ} \quad & \text{چون دامنه تغییرات } ۸، x_1، x_2، x_3 \text{ برابر } x_1 = x_2 = x_3 = ۸ \Rightarrow \begin{cases} 5x_1 = 5(8) = 40 \\ 6x_2 = 6(8) = 48 \\ 7x_3 = 7(8) = 56 \end{cases} \Rightarrow \bar{x} = \frac{40 + 48 + 56}{3} = 48 \\ & \text{صفر است، لذا همه آن‌ها با هم مساوی‌اند. از طرفی یکی از آن‌ها برابر ۸ است، پس بقیه هم باید ۸ باشند، لذا:} \end{aligned}$$

نکته اگر تمام داده‌ها با عددی ثابت جمع یا تفریق شوند، R هیچ تغییری نخواهد کرد؛ ولی اگر تمام داده‌ها در عددی ثابت ضرب یا بر عددی ثابت تقسیم شوند، برای R هم این اتفاق رخ خواهد داد؛ یعنی R فقط نسبت به ضرب و تقسیم تأثیرپذیر است.

مثلاً اگر دامنه تغییرات داده‌های (a, b, c, d) برابر ۱۰ باشد، آن‌گاه برای محاسبه دامنه تغییرات داده‌های $(2a+1, 2b+1, 2c+1, 2d+1)$ اولاً $(+۱)$ ها را در نظر نمی‌گیریم (جمع روی R بی‌تاثیر)، ثانیاً چون تمام داده‌های اولیه در عدد ۲ ضرب شده‌اند؛ لذا R اولیه هم در ۲ ضرب می‌شود:

$$R_{\text{جدید}} = 2 \times R_{\text{اولیه}} = 2 \times 10 = 20$$

واریانس و انحراف معیار (انحراف استاندارد)

دوتا از بهترین شاخص‌های پراکندگی هستند که میزان پراکندگی داده‌ها را نسبت به میانگین، با دقت خوبی نشان می‌دهند. البته کتاب درسی، بیشتر روی انحراف معیار مانور داده است ولی ما هر دوی این شاخص‌ها را بررسی می‌کنیم. اگر داده‌های $(x_1, x_2, \dots, x_n)$ با میانگین $\bar{x}$ را در اختیار داشته باشیم، برای محاسبه واریانس، ابتدا تک‌تک داده‌ها را منهای میانگین می‌کنیم، سپس جواب‌ها را به توان ۲ رسانده با هم جمع کرده و بر تعداد کل داده‌ها تقسیم می‌کنیم. به طور خلاصه می‌توان این طور نوشت:

$$\delta^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N} \quad (\text{واریانس})$$

در فرمول بالا به ۳ موضوع توجه کنید: موضوع اول این که به پرانتزهای $(x_1 - \bar{x})$ ، $(x_2 - \bar{x})$ ، ... و $(x_n - \bar{x})$ انحراف داده‌ها از میانگین می‌گوییم. (انحراف مشاهدات از میانگین هم می‌گوییم) و موضوع دوم این که به کل عبارت $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$ مجموع مجزورات انحراف داده‌ها از میانگین می‌گوییم و موضوع سوم این که اگر از عدد واریانس، جذر بگیریم، انحراف معیار (انحراف استاندارد) به دست می‌آید. مثلاً اگر واریانس ۹ باشد، انحراف معیار ۳ است یا برعکس، اگر انحراف معیار ۵ باشد، واریانس ۲۵ است.

نکته طبق فرمولی که برای واریانس نوشتیم می‌توان گفت واریانس برابر است با میانگین مربعات اختلاف داده‌ها از میانگین. (ممکنه تعریف واریانس در مدرسه از تون فواسته بشه.)

مثال و پاسخ

داده‌ها: ۲، ۳، ۵، ۷، ۸

مثال واریانس و انحراف معیار داده‌های مقابل را محاسبه کنید:

$$2, 3, 5, 7, 8 \Rightarrow \bar{x} = \frac{2+3+5+7+8}{5} = \frac{25}{5} = 5$$

پاسخ

$$\sigma^2 = \frac{\text{میانگین داده پنجم}^2 + \text{میانگین داده چهارم}^2 + \text{میانگین داده سوم}^2 + \text{میانگین داده دوم}^2 + \text{میانگین داده اول}^2}{\text{تعداد داده‌ها}} = \frac{(8-5)^2 + (7-5)^2 + (5-5)^2 + (3-5)^2 + (2-5)^2}{5}$$

$$= \frac{(-3)^2 + (-2)^2 + 0^2 + 2^2 + 3^2}{5} = \frac{9+4+0+4+9}{5} = \frac{26}{5} \xrightarrow{\text{جذر می‌گیریم}} \sigma = \sqrt{\frac{26}{5}} \text{ انحراف معیار}$$

مثال و پاسخ

مثال واریانس داده‌های جدول مقابل را به دست آورید:

داده‌ها	۱	۳	۵	۷
فراوانی	۲	۴	۱	۳

پاسخ

داده‌ها	۱	۳	۵	۷
فراوانی	۲	۴	۱	۳

$$\bar{x} = \frac{(1 \times 2) + (3 \times 4) + (5 \times 1) + (7 \times 3)}{2 + 4 + 1 + 3} = \frac{40}{10} = 4$$

داده‌ها	۱	۳	۵	۷
فراوانی	۲	۴	۱	۳

$$\sigma^2 = \frac{2(1-4)^2 + 4(3-4)^2 + 1(5-4)^2 + 3(7-4)^2}{10} = \frac{50}{10} = 5$$

می‌بینید که برای محاسبه واریانس از روی جدول فراوانی، دقیقاً مانند قبل عمل می‌کنیم. فقط پشت هر پرانتز، فراوانی داده مربوطه را ضرب می‌کنیم. شاکر، معنرت می‌فوام؛ نهی شه داده‌ها رو از حالت موهلی هارج کنیم، بعرض میاگین و واریانس اون‌ها رو پیدا کنیم؟ یعنی من اومدم داده‌ها رو به این صورت نوشتم: ۱، ۱، ۳، ۳، ۳، ۳، ۵، ۷، ۷، ۷. دیر، اونم می‌شه ولی آله فراوانی‌ها بزرگ باشن، راه هلت طولانی می‌شه.

نکته اگر دو یا چند گروه از داده‌ها دارای میانگین مساوی باشند، پراکندگی در گروهی کمتر است که واریانسش (انحراف معیارش) کمتر باشد. (تو کتاب شما همیشه میانگین‌ها مساوی اند.) به عبارت دیگر، وقتی میانگین دو گروه با هم برابر است، در گروهی که واریانس داده‌ها کمتر است، داده‌ها به میانگین نزدیک‌تر هستند.

مثال و پاسخ

مثال نمرات درس زبان کلاس A به صورت {۶۵، ۷۵، ۷۳، ۵۰، ۶۰، ۶۴، ۶۹، ۶۲، ۶۷، ۸۵} و نمرات همین درس در کلاس B به صورت {۸۵، ۷۹، ۵۷، ۳۹، ۴۵، ۷۱، ۶۷، ۸۷، ۹۱، ۴۹} می‌باشد. پراکندگی نمرات در کدام کلاس کمتر است؟ (به عبارت دیگر در کدام کلاس سطح دانش‌آموزان در درس زبان به میانگین نزدیک‌تر است؟)

پاسخ میانگین نمرات کلاس‌های A و B هر دو برابر ۶۷ است. در این گونه سؤالات هم می‌توانید واریانس دو کلاس را با هم مقایسه کنید و هم انحراف معیار را (هر کدوم که دوست دارید) چه واریانس و چه انحراف معیار، هر چه قدر کمتر باشند، داده‌ها به میانگین نزدیک‌تر هستند. (فودتون به عنوان تمرین واریانس و انحراف معیار هر دو کلاس رو به دست بیارین. البته برای محاسبه انحراف معیار، نیاز به ماشین حساب دارید، پس بهتره به همون واریانس اکتفا کنید ولی ما هر دو را حساب کرده‌ایم.)

(مشابه فعالیت کتاب درسی)

$$\sigma^2 = 80/4 \xrightarrow{\text{جذر}} \sigma = 8/97 \text{ کلاس A}$$

$$\sigma^2 = 315 \xrightarrow{\text{جذر}} \sigma = 17/75 \text{ کلاس B}$$

پس پراکندگی نمرات در کلاس A کمتر است چون واریانس و انحراف معیار آن نسبت به واریانس و انحراف معیار کلاس B کمتر است.

نکته: اگر همه داده‌ها با هم مساوی باشند، واریانس و انحراف معیار آن‌ها صفر است؛ زیرا اگر داده‌های $(x_1, x_2, \dots, x_n)$ با هم برابر باشند، آن‌گاه میانگین آن‌ها یعنی $\bar{x}$ برابر است با تک‌تک داده‌ها یعنی $\bar{x} = x_1 = x_2 = \dots = x_n$ لذا در فرمول $\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N}$ حاصل تمام پرانتزها برابر صفر می‌شود، پس σ^2 هم برابر صفر می‌شود. وقتی σ^2 صفر است، جذر آن یعنی σ هم صفر می‌شود. هم‌چنین می‌توان گفت اگر واریانس یا انحراف معیار یک سری از داده‌ها صفر باشد، تمام داده‌ها با هم مساوی‌اند.

مثال و پاسخ

مثال: اگر واریانس داده‌های $(1, 2c, d, \frac{b}{3}, a, 4)$ برابر صفر باشد، میانگین داده‌های (a, b, c, d) را به دست آورید.

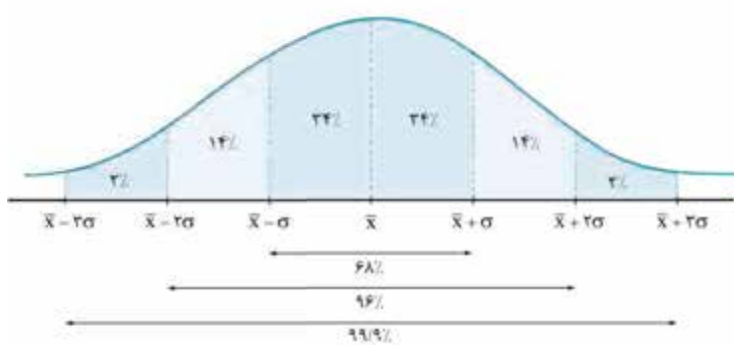
پاسخ: واریانس صفر است؛ پس تمام داده‌ها باید برابر باشند. یکی از داده‌ها ۴ است، پس بقیه داده‌ها نیز باید ۴ باشند:

$$a = 4, \frac{b}{3} = 4 \Rightarrow b = 12, 2c = 4 \Rightarrow c = 2, d - 1 = 4 \Rightarrow d = 5$$

$$\text{داده‌ها: } a, b, c, d \xrightarrow{\text{جای‌گذاری مقادیر}} 4, 12, 2, 5 \Rightarrow \bar{x} = \frac{4 + 12 + 2 + 5}{4} = \frac{23}{4}$$

پراکندگی نرمال (نحوه پخش داده‌ها در اطراف میانگین)

اگر تعداد داده‌ها بسیار زیاد باشد، نمودار خط شکسته آن‌ها به یک منحنی (خم) تبدیل می‌شود حال اگر این منحنی تقریباً به شکل یک زنگ



باشد به آن **منحنی نرمال** (خم بهنجار)

می‌گوییم. نحوه پخش (پراکندگی) داده‌ها در اطراف میانگین

برای منحنی نرمال به صورت مقابل است:

(تمام درصدهای این نمودار رو حفظ کنید.)

نتایج حاصل از نمودار صفحه بالا:

۱) تقریباً ۶۸٪ داده‌ها بین $\bar{x} - \sigma$ و $\bar{x} + \sigma$ قرار دارند ($34 + 34 = 68$). یا می‌توان گفت ۶۸٪ داده‌ها که در فاصله ۱ برابر انحراف معیار از میانگین قرار دارند.

۲) تقریباً ۹۶٪ داده‌ها بین $\bar{x} - 2\sigma$ و $\bar{x} + 2\sigma$ قرار دارند ($14 + 34 + 34 + 14 = 96$). یا می‌توان گفت ۹۶٪ داده‌ها در فاصله دو برابر انحراف معیار از میانگین قرار دارند.

۳) تقریباً ۹۹٫۸٪ درصد داده‌ها (یعنی تقریباً تمام داده‌ها) بین $\bar{x} - 3\sigma$ و $\bar{x} + 3\sigma$ قرار دارند ($2 + 14 + 34 + 34 + 14 + 2 = 100$). یا می‌توان گفت ۹۹٫۸٪ داده‌ها در فاصله ۳ برابر انحراف معیار از میانگین قرار دارند.

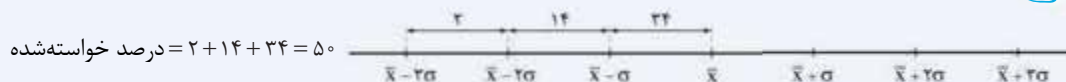
مثال و پاسخ

مثال: با توجه به منحنی نرمال:

الف) چند درصد داده‌ها بین $\bar{x} - 3\sigma$ و $\bar{x}$ قرار دارند؟

ب) چند درصد داده‌ها بین $\bar{x} - \sigma$ و $\bar{x} + 2\sigma$ قرار دارند؟

پاسخ: الف) با توجه به نمودار بالا کافی است درصدهای بین نقطه‌های $\bar{x} - 3\sigma$ و $\bar{x}$ روی محور x را با هم جمع کنیم لذا خواهیم داشت:



$$\text{درصد خواسته شده} = 2 + 14 + 22 = 38$$



$$\text{درصد خواسته شده} = 34 + 34 + 14 = 82$$

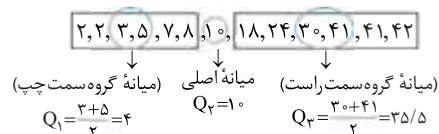
نحوه خواندن اعداد مربوط به وزن که روی اجناس و مواد غذایی درج می‌شود

همیشه روی قوطی یا بسته مواد غذایی، وزن به صورت $\bar{x} \pm 2\sigma$ نوشته می‌شود که البته در این جا منظور از $\bar{x}$ وزن خالص ماده غذایی می‌باشد. (اون $\pm 2\sigma$ به خاطر خطایی که ماشین آلات یا انسان در اندازه‌گیری دارند) پس می‌توان گفت تقریباً در ۹۶ درصد قوطی‌ها وزن ماده داخل آن بین $\bar{x} - 2\sigma$ و $\bar{x} + 2\sigma$ می‌باشد. مثلاً اگر روی یک قوطی کنسرو تن ماهی، عبارت (180 ± 10) گرم نوشته شده باشد، به این معناست که در ۹۶ درصد قوطی‌ها وزن تن ماهی داخل آن‌ها، عددی بین $\frac{180-10}{170}$ و $\frac{180+10}{190}$ یا $(170, 190)$ گرم می‌باشد.

نکته (مفهوم علاقه‌مندان): اگر تمام داده‌ها در عددی مثل k ضرب شوند، آن‌گاه واریانس آن‌ها در k^2 ضرب می‌شود یعنی: $\sigma^2_{\text{جدید}} = k^2 \times \sigma^2$ هم چنین اگر تمام داده‌ها در عددی مثل k ضرب شوند، انحراف معیار آن‌ها در $|k|$ ضرب می‌شود، یعنی: $\sigma_{\text{جدید}} = |k| \times \sigma$ ضمناً جمع و تفریق تمام داده‌ها با یک عدد ثابت تأثیری روی σ^2 و σ ندارد؛ مثلاً اگر واریانس داده‌های (a, b, c, d) برابر ۳ باشد، آن‌گاه برای محاسبه واریانس داده‌های $(5a-2, 5b-2, 5c-2, 5d-2)$ می‌گوییم: اولاً (-2) ‌ها بی تأثیرند، ثانیاً تمام داده‌های اولیه در عدد ۵ ضرب شده‌اند، لذا: $\sigma^2_{\text{جدید}} = k^2 \times \sigma^2_{\text{اولیه}} = 5^2 \times 3 = 75$

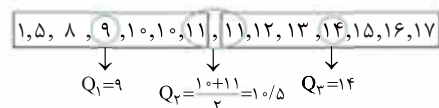
چارک‌های اول، دوم، سوم - دامنه میان چارکی

با مفهوم میانه کاملاً آشنا شده‌اید. نام دیگر میانه، چارک دوم (Q_2) می‌باشد. حال اگر میانه داده‌های کوچک‌تر از Q_2 را حساب کنیم، چارک اول (Q_1) و اگر میانه داده‌های بزرگ‌تر از Q_2 را حساب کنیم، چارک سوم (Q_3) به دست می‌آید. ضمناً به $Q_3 - Q_1$ دامنه میان چارکی می‌گوییم و آن را با IQR نمایش می‌دهیم. به عنوان مثال می‌خواهیم تمام چارک‌ها و دامنه میان چارکی داده‌های زیر را به دست آوریم:

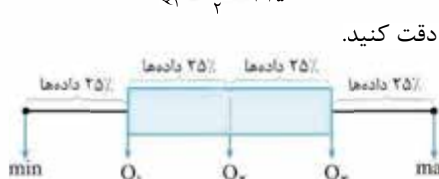


$$IQR = Q_3 - Q_1 = 35/5 - 4 = 31/5$$

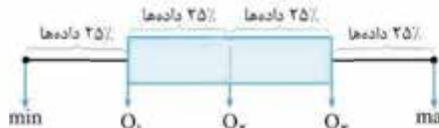
شاگرد، پیش‌شماره ۱۰ رو که میانه بود، نه در گروه سمت چپ آوردین، نه در گروه سمت راست. حالا آگه تعداد داده‌ها زوج بود و دو تا عدد وسط داشتیم چی؟ دیر، به به سوال باحالی، آگه تعداد داده‌ها زوج بود، از دو عدد وسط، یکی شونو می‌بریم به گروه سمت چپ و یکی دیگرم می‌بریم به گروه سمت راست. به مثال بعری دقت کنی فکرت می‌فهمی.



مثال تمامی چارک‌های داده‌های مقابل را به دست می‌آوریم:



نکته چارک‌ها، کل داده‌های آماری را به ۴ قسمت مساوی تقسیم می‌کنند. به نمودار زیر دقت کنید.



فرض بر این است که تمام داده‌ها متمایزند و تعداد داده‌ها مضرب ۴ است.

به این نمودار، نمودار جعبه‌ای می‌گوییم که در فصل بعد به طور مفصل آن را بررسی خواهیم کرد. این نمودار، نحوه پخش داده‌ها را به صورت درصدی و مقداری نشان می‌دهد.

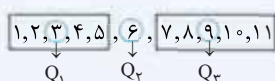
از نمودار جعبه‌ای رسم شده می‌توان به نتایج زیر رسید: (البته فقط با داشتن چارک‌ها نیز می‌شه به این نتایج رسید.)

- الف** ۲۵ درصد داده‌ها ($\frac{1}{4}$ داده‌ها) از چارک اول کوچک‌ترند. (یا می‌توان گفت ۷۵ درصد داده‌ها یعنی $\frac{3}{4}$ داده‌ها از چارک اول بزرگ‌ترند).
- ب** ۲۵ درصد داده‌ها ($\frac{1}{4}$ داده‌ها) بین چارک اول و چارک دوم قرار دارند.
- ج** ۲۵ درصد داده‌ها ($\frac{1}{4}$ داده‌ها) بین چارک دوم و چارک سوم قرار دارند.
- د** ۲۵ درصد داده‌ها ($\frac{1}{4}$ داده‌ها) از چارک سوم بزرگ‌ترند. (یا می‌توان گفت ۷۵ درصد داده‌ها یعنی $\frac{3}{4}$ داده‌ها از چارک سوم کوچک‌ترند).

مثال و پاسخ

۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹, ۱۰, ۱۱

ب) ۷۵ درصد داده‌ها از کدام عدد کوچک‌ترند؟



مثال با توجه به داده‌های مقابل، به سؤالات زیر پاسخ دهید:

الف) ۲۵ درصد داده‌ها از کدام عدد کوچک‌ترند؟

ج) ۵۰ درصد داده‌ها از کدام عدد بزرگ‌ترند؟

پاسخ ابتدا تمام چارک‌ها را به دست می‌آوریم:

الف می‌دانیم ۲۵ درصد از داده‌ها از Q_1 کوچک‌ترند. لذا جواب برابر ۳ است.

ب می‌دانیم ۷۵ درصد داده‌ها از Q_3 کوچک‌ترند پس جواب برابر ۱۰ است.

ج ۵۰ درصد داده‌ها از Q_2 بزرگ‌ترند پس جواب برابر ۶ است.



نکته: اگر در یک سری از داده‌های آماری، یک یا چند داده دور افتاده (پرت) داشته باشیم، برای بررسی میزان پراکندگی داده‌ها (دوری یا نزدیکی داده‌ها به هم) باید از میانه و دامنه میان چارکی استفاده کنیم. (به طور ساده‌تر با داشتن Q_1 ، Q_2 و Q_3 نحوه پراکندگی داده‌ها معلوم می‌شود). البته حتی اگر داده پرت هم نداشته باشیم، همیشه استفاده از میانه و دامنه میان چارکی نسبت به استفاده از میانگین و انحراف معیار برتری داشته و باعث تفسیر سریع‌تر و ساده‌تر اطلاعات خواهد شد.

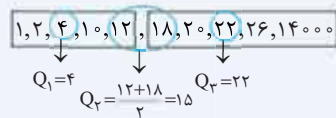
یعنی با نگاهی ساده به چارک‌ها سریعاً می‌فهمیم که داده‌ها چگونه توزیع شده‌اند، ولی اگر $\bar{x}$ و σ را داشته باشیم، باید بازه‌های $(\bar{x}-\sigma, \bar{x}+\sigma)$ ، $(\bar{x}-2\sigma, \bar{x}+2\sigma)$ و $(\bar{x}-3\sigma, \bar{x}+3\sigma)$ را پیدا کنیم که شاید کمی وقت‌گیر باشد. البته هر دو روش معتبر و کاربردی هستند.

مثال و پاسخ

مثال: در داده‌های زیر، برای بررسی میزان پراکندگی داده‌ها بهتر است از چارک‌ها استفاده کنیم یا میانگین و انحراف معیار؟ نحوه پخش (توزیع) داده‌ها را به صورت درصدی نشان دهید.

۱، ۲، ۴، ۱۰، ۱۲، ۱۸، ۲۰، ۲۲، ۲۶، ۱۴۰۰۰

پاسخ: عدد ۱۴۰۰۰ داده دور افتاده است، لذا برای بررسی میزان پراکندگی داده‌ها (نحوه پخش داده‌ها) بهتر است از چارک‌ها استفاده کنیم.



پس ۲۵٪ داده‌ها کوچک‌تر از ۴، ۲۵٪ داده‌ها بین ۴ و ۱۵، ۲۵٪ داده‌ها بین ۱۵ و ۲۲ و در نهایت ۲۵٪ داده‌ها بزرگ‌تر از ۲۲ است.



نکته: اکثر مطالبی که در این فصل بررسی کرده‌ایم از جمله میانگین، میانه، مُد، واریانس، انحراف معیار، چارک‌ها و دامنه میان چارکی تحت عنوان آمار توصیفی شناخته می‌شوند. (فقط گفتیم که تئوری کتاب گفته، شما تلفیق کنید).

سؤال‌های امتحانی

۴۹- جملات درست را با ☒ و جملات نادرست را با ☐ مشخص کنید. (متن کتاب درسی)

(الف) در علم آمار نه تنها معدل (میانگین) یک سری از داده‌ها را می‌خواهیم، بلکه میزان تغییرات و پراکندگی در حوالی و اطراف میانگین نیز مهم است.

(ب) یکی از معیارهای پراکندگی که معمولاً با میانگین بیان می‌شود، دامنه تغییرات است.

(ج) یکی از معیارهای پراکندگی که معمولاً با میانه بیان می‌شود، دامنه میان چارکی است.

(د) میانگین و انحراف معیار، اطلاعات سریعی درباره داده‌ها بدون نیاز به هرگونه محاسبه می‌دهند.

(ه) تقریباً ۷۵ درصد داده‌ها بزرگ‌تر از چارک سوم هستند. (و) تقریباً $\frac{3}{4}$ داده‌ها بزرگ‌تر از چارک اول هستند.

(ز) تقریباً ۲۵ درصد داده‌ها قبل از چارک اول و تقریباً ۲۵ درصد داده‌ها بعد از چارک سوم قرار دارند.

(ح) در منحنی نرمال، تقریباً ۸۰ درصد داده‌ها (مشاهدات) در فاصله ۲ برابر انحراف معیار از میانگین قرار دارند.

۵۰- در یک شرکت، چارک سوم حقوق کارمندان برابر ۸۰۰ هزار تومان است. این به معنای آن است که (مدارس تهران)

(۱) تقریباً $\frac{1}{4}$ کارمندان، کم‌تر از ۸۰۰ هزار تومان دریافت می‌کنند. (۲) تقریباً ۲۵ درصد کارمندان، بیشتر از ۸۰۰ هزار تومان دریافت می‌کنند.

(۳) تقریباً ۵۰ درصد کارمندان، کم‌تر از ۸۰۰ هزار تومان دریافت می‌کنند. (۴) تمام کارمندان، بیشتر از ۸۰۰ هزار تومان دریافت می‌کنند.

۵۱- تمام چارک‌ها و دامنه میان چارکی را برای اعداد صحیح از ۰ تا ۱۲ حساب کنید. (مدارس تهران)

۵۲- اگر دامنه تغییرات داده‌های $x_1, x_2, \dots, x_{19}, 10$ مساوی با صفر باشد، مجموع داده‌ها چه قدر است؟ (مدارس تهران)

۵۳- اگر انحراف معیار داده‌های $x, y, z, t, m, 20$ برابر صفر باشد، میانگین آن‌ها چه قدر است؟ (مدارس تهران)

۵۴- اگر مقداری ثابت را از هر یک از داده‌های آماری کم کنیم، کدام شاخص زیر، ثابت می‌ماند؟ (مدارس تهران)

(الف) میانگین (ب) میانه (ج) دامنه تغییرات

۵۵- در داده‌های ۲۵، ۲۰، ۲۱، ۲۶، ۱۲، ۱۴، ۱۵، ۲۴، ۲۰، ۱۶، ۱۴ و ۱۸ میانگین داده‌های بزرگ‌تر از چارک اول و کوچک‌تر از چارک دوم را محاسبه کنید.

(برگرفته از متن کتاب درسی)

۵۶- اگر تمام داده‌ها در عدد ۱۰ ضرب شوند، آنگاه دامنه تغییرات:
(۱) در عدد ۱۰ ضرب می‌شود. (۲) با ۱۰ جمع می‌شود. (۳) بر ۱۰ تقسیم می‌شود. (۴) هیچ تغییری نمی‌کند.
(مدارس تهران)

۵۷- اگر انحراف معیار داده‌های x, y, z, t و $3x$ و $y - 2, \frac{z}{5}, t, 12$ برابر با صفر باشد، میانه داده‌های (x, y, z, t) چه قدر است؟
(مدارس تهران)

۵۸- اگر داده‌های آماری ۱۱، ۱۵، ۱۷، ۱۶، ۱۴، ۹، ۱۲، ۱۵، ۱۸، ۱۴ و ۱۱ را بانمودار جعبه‌ای نمایش دهیم، انحراف معیار داده‌های داخل جعبه را تعیین کنید.
(مدارس تهران)

۵۹- اگر $\bar{x}$ و σ^2 میانگین و واریانس یک سری از داده‌ها باشند که نمودارشان به صورت منحنی نرمال است، کدام مورد زیر صحیح است؟
(۱) ۶۸ درصد داده‌ها در فاصله $\bar{x} - 2\sigma$ تا $\bar{x} + 2\sigma$ قرار دارند. (۲) ۹۶ درصد داده‌ها در فاصله $\bar{x} - 2\sigma$ تا $\bar{x} + 2\sigma$ قرار دارند.
(۳) ۹۸ درصد داده‌ها در فاصله $\bar{x} - 3\sigma$ تا $\bar{x} + 3\sigma$ قرار دارند. (۴) ۱۰۰ درصد داده‌ها در فاصله $\bar{x} - \sigma$ تا $\bar{x} + \sigma$ قرار دارند.
(متن کتاب درسی)

۶۰- واریانس و انحراف معیار داده‌های ۹، ۱۲، ۱۰، ۹، ۱۱، ۱۰، ۷ و ۱۲ را محاسبه کنید.
(مشابه تمرین کتاب درسی)

۶۱- کدام دو شاخص آماری داده‌های (۱، ۷، ۵، ۹، ۸) با هم برابرند؟
(مدارس تهران)

(۱) میانگین و واریانس (۲) دامنه تغییرات و انحراف معیار (۳) میانگین و انحراف معیار (۴) واریانس و دامنه تغییرات
۶۲- در داده‌های آماری ۱۱، ۶، ۵، ۱۳، ۸، ۷، ۱۰، ۷ و ۱۳ داده‌های کم‌تر از میانه را حذف می‌کنیم. واریانس داده‌های باقی‌مانده چه قدر است؟
(مدارس تهران)

۶۳- اگر میانگین داده‌های $a + 2, a + 1$ و a برابر ۱۰ باشد، انحراف معیار آن‌ها را محاسبه کنید.
(مدارس تهران)

۶۴- در یک سری از داده‌های آماری، انحراف داده‌ها از میانگین عبارت‌اند از: ۵، ۴، ۳، ۲، ۳، ۲، ۱ و -۱. انحراف معیار آن‌ها را به دست آورید.
(مدارس تهران)

۶۵- اختلاف از میانگین (انحراف از میانگین) شش داده آماری به صورت ۳، -۲، -۱، ۰، ۲ و ۴ می‌باشد. واریانس این داده‌ها را به دست آورید.
(کتاب درسی)

۶۶- روی بسته یک نوع ماکارانی عبارت (15 ± 200) گرم نوشته شده است. از هر ۱۰۰ بسته ماکارانی، وزن ۹۶ بسته آن در چه محدوده‌ای قرار دارد؟
(متن کتاب درسی)

۶۷- چه تعداد از داده‌های یک مجموعه ۱۴ عضوی به صورت $(x_1, x_2, \dots, x_{14})$ بین چارک اول و سوم قرار دارند؟
(مفهوم علاقه‌مندان)

(۱) ۹ (۲) ۸ (۳) ۷ (۴) ۶

۶۸- جدول زیر را کامل کرده و به کمک آن، واریانس و انحراف معیار (انحراف استاندارد) داده‌ها را به دست آورید.
(مشابه کار در کلاس کتاب درسی)

مربع انحراف داده‌ها از میانگین	انحراف داده‌ها از میانگین	داده‌ها (مشاهدات)
		۴
		۵
		۶
		۷
		۸

۶۹- اگر واریانس داده‌های ۱، x_1, x_2, x_3, x_4, x_5 برابر صفر باشد، انحراف معیار داده‌های $4 + x_5, 3 + x_4, 2 + x_3, 1 + x_2$ و x_1 را تعیین کنید.
(مفهوم علاقه‌مندان)

۷۰- اگر σ_A^2 واریانس گروه A و σ_B^2 واریانس گروه B باشد، کدام گزینه درست است؟
گروه A: ۱، ۲، ۳، ۴، ۵ گروه B: ۳، ۶، ۹، ۱۲، ۱۵

(۱) $\sigma_A^2 = 9\sigma_B^2$ (۲) $\sigma_B^2 = 2\sigma_A^2 + 1$ (۳) $\sigma_B^2 = \sigma_A^2 + 2$ (۴) $\sigma_B^2 = 9\sigma_A^2$
(مدارس تهران)

۷۱- امتیازات مهارت کاری دو فرد A و B در پنج روز متوالی چنین است:
(برگرفته از متن کتاب درسی)

گروه B: ۲۱، ۲۴، ۲۵، ۲۷، ۲۸ گروه A: ۲۲، ۲۳، ۲۴، ۲۷، ۲۹

دقت عمل کدام فرد بیشتر است؟

(۱) یکسان (۲) غیرقابل بررسی (۳) A (۴) B

۷۲- انحراف معیار داده‌های (۲, ۴, ۵, ۶, ۸) چند برابر دامنه میان چارکی داده‌های ۱۰, ۹, ۸, ۸, ۶, ۶, ۵, ۴, ۲, ۱ می‌باشد؟ (مدارس تهران)

$$\frac{1}{7} \quad (1) \quad \frac{2}{7} \quad (2) \quad \frac{2}{5} \quad (3) \quad \frac{3}{5} \quad (4)$$

۷۳- دو دسته داده زیر، قیمت کالایی را در دو بازار جداگانه برحسب هزار تومان نشان می‌دهد. (مشابه تمرین کتاب درسی)

بازار (الف)	۷	۹	۸	۹	۷
بازار (ب)	۱۰	۸	۶	۷	۹

(الف) در کدام دسته، پراکندگی بیشتر است؟ (یعنی در کدام بازار، قیمت‌ها از میانگین دورترند؟)

(ب) دامنه تغییرات دو گروه را به دست آورید. آیا دامنه تغییرات گروه‌ها با پراکندگی آن‌ها هم‌خوانی دارند؟

(ج) بهتر است از کدام بازار خرید کنیم؟

(د) اگر داده‌ها را در اختیار نداشته باشیم، آیا فقط با داشتن دامنه تغییرات می‌توانیم برای خرید تصمیم‌گیری کنیم؟

۷۴- در جدول زیر، مقدار واریانس و انحراف استاندارد (انحراف معیار) را به دست آورید.

مرکز دسته	۱	۳	۵	۷	۹
فراوانی	۳	۶	۴	۲	۱

۷۵- در جدول زیر، واریانس را به دست آورید.

حدود دسته	۰-۲	۲-۴	۴-۶	۶-۸
فراوانی	۱	۲	۹	۴

پاسخ سؤال‌های امتحانی

- ۱- در مصاحبه و یا تهیه پرسش‌نامه، نباید از پرسش‌های هدایت‌کننده استفاده کنیم. لذا گزینه (۴) صحیح است.
- ۲- سن ازدواج افراد را به کمک اطلاعات از پیش تهیه‌شده (مانند دفتر رسمی ازدواج) می‌توان تهیه کرد. علاقه اشخاص به انجام نوعی ورزش یا رنگ مورد علاقه آن‌ها برای خرید ماشین را می‌توان از طریق پرسش‌نامه بررسی کرد. رابطه رژیم گرفتن با اختلالات عصبی به کمک آزمایش قابل بررسی است. پس جواب، گزینه (۱) است.
- ۳- میزان آلودگی هوا، کمی نسبی است. (هون با عدد بیان می‌شود و افتلاف و نسبت مقادیر اون بامعنی). لذا گزینه (۱) صحیح است.
- ۴- نوع آلودگی هوا (سرب، گوگرد و ...) کیفی اسمی است. دقت کنید که مقدار آلودگی هوا کمی نسبی است. (با عدد بیان می‌شود و افتلاف و نسبت دو مقدار اون بامعنی). لذا گزینه (۳) صحیح است.
- ۵- موضوع اصلی مورد مطالعه، درجه حرارت هوا است. پس گزینه (۳) صحیح است.
- ۶- (الف) در درس‌نامه گفتیم که متغیرهای کیفی به دو دسته اسمی و ترتیبی تقسیم می‌شوند. پس جواب این قسمت می‌شود: کیفی (ب) می‌دانیم متغیرهای کمی به دو دسته فاصله‌ای و نسبتی تقسیم می‌شوند. پس جواب این قسمت می‌شود: کمی
- ۷- پارامتر در یک جامعه آماری همیشه ثابت و بدون تغییر است. ولی مقدار آماره از نمونه‌ای به نمونه دیگر متفاوت است. پس مقدار آماره‌های این ۳ نمونه، در حالت کلی مساوی نیستند.
- ۸- (الف) متغیر، نوع وعده غذایی (صبحانه، ناهار یا شام) می‌باشد. (ب) کیفی است. چون وعده غذایی از جنس عدد نیست. ضمناً چون صبحانه، ناهار و شام دارای یک ترتیب طبیعی است، لذا کیفی با مقیاس ترتیبی می‌باشد. (ج) کل ۱۸۰۰ نفر مشتری رستوران، جامعه آماری محسوب می‌شوند.
- (د) چون تمام افراد جامعه را مورد بررسی قرار داده‌ایم پس سرشماری کرده‌ایم. (آگه قسمتی از این ۱۸۰۰ نفر رو تصادفاً انتخاب می‌کردیم، می‌گفتیم نمونه‌گیری کرده‌ایم).
- ۹- با توجه به متن درس‌نامه، گزینه (۲) صحیح است. مثلاً می‌دانیم وزن متغیر کمی است و می‌توانیم وزن افراد را با هم مقایسه کنیم. ضمناً داده‌های مربوط به متغیر کمی می‌توانند صحیح یا اعشاری باشند (لزوماً عدد صحیح نیستند) هم چنین اختلاف بین وزن دو نفر بامعنی است (مثلاً می‌شود گفت وزن علی ۲۰ کیلوگرم کمتر از وزن مهسره و در نهایت این که متغیرهای کمی از جنس عدد و رقم هستند).
- ۱۰- با توجه به متن درس‌نامه، گزینه (۳) صحیح است. دقت داریم که داده‌های مربوط به متغیرهای کیفی می‌تواند از جنس عدد باشد یا نباشد ولی در حالتی که از جنس عدد باشد، جمع هر دو مقدار آن بی‌معنی است.

۱۱- برای شماره گذاری صندلی‌ها از اعداد طبیعی استفاده می‌شود و چون اعداد طبیعی دارای نوعی ترتیب هستند، لذا نوع متغیر، کیفی و مقیاس آن ترتیبی است. (دقت کنید که جمع دو شماره صندلی، بی‌معنی و پیروی رو نشون نمیده؛ به همین علت، با وجود این که شماره صندلی‌ها از پس عدد هستن ولی متغیر کمی محسوب نمی‌شه.) لذا گزینه (۲) صحیح است.

۱۲- طبق توضیحات درس‌نامه، گزینه (۱) صحیح است.

۱۳- میزان آب هدررفته، مدت زمان رسیدن بسته‌ها، میزان آلودگی هوا و قد دانش‌آموزان همگی متغیر کمی با مقیاس نسبتی هستند. لذا جواب این سؤال، گزینه (۴) خواهد بود. (رتبه کشورها از نظر رفاه اقتصادی، متغیر کیفی با مقیاس ترتیبی.)

- ۱۴- رنگ مو ← کیفی اسمی
وزن ← کمی نسبتی
مدت مکالمه ← کمی نسبتی
سطح سواد ← کیفی ترتیبی
گنجایش (حجم) یک مخزن ← کمی نسبتی
پس در کل ۶ متغیر کمی وجود دارد و لذا گزینه (۳) صحیح است.
- ۱۵- مقاومت یک لامپ ← کمی نسبتی
رتبه کنکور ← کیفی ترتیبی
مراحل زندگی انسان ← کیفی ترتیبی
مزه غذا ← کیفی ترتیبی
پس ۴ متغیر کیفی وجود دارد و لذا گزینه (۲) صحیح است.

۱۶- سابقه کار دبیران: کمی نسبتی - مراحل کشت محصولات کشاورزی: کیفی ترتیبی - میزان سرعت ماشین‌ها: کمی نسبتی - وضع تأهل: کیفی اسمی است. لذا گزینه (۳) صحیح است.

- ۱۷- الف) دادگان (مراجعه به پرونده دانش‌آموزان)
ب) مشاهده
ج) مصاحبه
د) دادگان (مراجعه به آمار وزارت بهداشت)
ه) پرسش‌نامه یا مصاحبه
و) مصاحبه یا پرسش‌نامه
ز) مصاحبه یا پرسش‌نامه
ط) مصاحبه یا پرسش‌نامه (البته ابتدا باید یک نمونه از مردان ایرانی را انتخاب کنیم).

۱۸- الف) کل هندوانه‌ها جامعه آماری و ۱۰۰ عدد هندوانه انتخاب‌شده، نمونه تصادفی است. هم‌چنین وزن، متغیر تصادفی است. (چون موضوع مورد بررسی در این‌جا وزن است.) وزن متوسط (میانگین) تمام هندوانه‌ها پارامتر و وزن متوسط ۱۰۰ هندوانه انتخابی، آماره است.

ب) جامعه آماری: کل هواپیماهای تولیدی در یک روز
نمونه تصادفی: ۲۰ هواپیمای انتخاب‌شده از خط تولید
متغیر تصادفی: بُرد پروازی هواپیماها
پارامتر: نمی‌توان مستقیماً محاسبه کرد، چون از کل هواپیماها، نمی‌دانیم چه تعداد از آن‌ها بُرد زیاد، بُرد متوسط و بُرد کم دارند؛ پس پارامتر را باید از روی آماره‌ها حدس (تخمین) بزنیم.
آماره: تمام کسرهای زیر (نسبت‌های زیر) آماره می‌باشند؛ چون مربوط به نمونه هستند نه کل جامعه:

$$\frac{10}{20} = \frac{1}{2} = \frac{\text{تعداد هواپیماهای با بُرد زیاد در نمونه}}{\text{تعداد کل هواپیماهای نمونه}}$$

$$\frac{7}{20} = \frac{\text{تعداد هواپیماهای با بُرد متوسط در نمونه}}{\text{تعداد کل هواپیماهای نمونه}}$$

$$\frac{3}{20} = \frac{\text{تعداد هواپیماهای با بُرد کم در نمونه}}{\text{تعداد کل هواپیماهای نمونه}}$$

۱۹- الف) جامعه آماری، کل خانوارهای آن شهر هستند. اندازه جامعه همان تعداد خانوارهاست که برابر ۱,۰۰۰,۰۰۰ می‌باشد. متغیر تصادفی، مدت زمان تماشای کانال خاص در هفته است. نمونه تصادفی، خانواده‌هایی هستند که از بین کل خانواده‌ها برای بررسی انتخاب می‌کنیم که تعدادشان برابر ۲۰۰۰ است. (پس اندازه نمونه ۲۰۰۰ است.)

(ب) متغیر تصادفی، مدت زمان تماشای کانال خاص است که متغیری کمتی با مقیاس نسبتی است. مثلاً فرض کنید یک خانواده ۹ ساعت در هفته و یکی دیگر ۳ ساعت آن کانال را می‌بینند، در این صورت اختلاف ۹ و ۳ نشان می‌دهد یک خانواده ۶ ساعت بیشتر آن کانال را دیده و نسبت ۹ و ۳ ($\frac{9}{3} = 3$) نشان می‌دهد یک خانواده ۳ برابر دیگری آن کانال را تماشا کرده. (به همین علت گفتیم کمتی نسبتی)

۲۰- الف) کیفی ترتیبی (ب) کمتی نسبتی (ج) کیفی ترتیبی (چون می‌توان گزینه‌ها را مرتب و مقایسه کرد). د) کیفی اسمی (ه) کمتی فاصله‌ای (مثلاً در یک دانشگاه کلاس ۷:۳۰ شروع میشه و در دانشگاهی دیگر ۸:۰۰ واضح است که اختلاف این دو عدد بامعنیه و نشون میده در یکی از دانشگاه‌ها نیم ساعت زودتر کلاس تشکیل میشه ولی تقسیم (نسبت) این دو عدد بی‌معنیه، پس کمتی فاصله‌ای است). و) کمتی نسبتی (ز) کیفی اسمی (ح) کمتی نسبتی (ط) کیفی ترتیبی (ی) کیفی ترتیبی (جمع سال تولد دو نفر بی‌معنی است؛ پس نوع متغیر کیفی است. ضمناً سال‌های تولد دارای نوعی ترتیب اند و می‌شه اون‌ها رو با هم مقایسه کرد، پس مقیاسش ترتیبی است).

۲۱- متغیر، آن موضوعی است که می‌خواهیم برای جامعه یا نمونه بررسی کنیم مانند وزن، قد و نوع گروه خونی. ولی به نتایج حاصل از بررسی‌های آماری، داده می‌گوییم؛ مثلاً اگر موضوع مورد مطالعه، قد افراد یک خانواده ۳ نفری مشخص باشد، به اعداد ۱۸۰، ۱۶۵ و ۱۴۸ سانتی‌متر داده‌های حاصل از این بررسی می‌گوییم. پارامتر یک مشخصه عددی است که بیان‌کننده جنبه‌ای خاص از جامعه است ولی آماره مشخصه عددی است که بیان‌کننده جنبه‌ای خاص از نمونه است. مثلاً فرض کنید در یک زمین کشاورزی محصول ما هندوانه باشد. فرض کنید متغیر ما مثلاً «مزه هندوانه‌ها» باشد (بد، معمولی، خوب)، حال اگر تمام هندوانه‌ها را بررسی کنیم، نسبت تعداد هندوانه‌هایی که مثلاً مزه بد دارند به تعداد کل هندوانه‌ها، پارامتر نامیده می‌شود ولی چون امتحان کردن مزه تمام هندوانه‌ها غیرممکن است، یک نمونه از آن‌ها را انتخاب و مزه آن‌ها را امتحان می‌کنیم. در این نمونه، نسبت تعداد هندوانه‌هایی که مزه بد دارند به تعداد کل هندوانه‌های این نمونه، آماره نامیده می‌شود.

۲۲- در تمام گزینه‌های (۱)، (۲) و (۳) بررسی تمام اعضای جامعه (سرشماری)، کاری بسیار طولانی و حتی غیرممکن است ولی تعداد دانش‌آموزان یک کلاس، آن قدر زیاد نیست که بخواهیم یک نمونه تصادفی از آن‌ها را انتخاب کنیم؛ یعنی می‌توانیم معدل تمام بچه‌های کلاس را به راحتی محاسبه کنیم. ۲۳- الف) متغیر مزه، کیفی و مقیاس آن ترتیبی است. (پهن مزه خوب، معمولی و بد رو میشه با هم مقایسه کرد).

(ب) چون اعداد ۱۰ و ۵۰ مربوط به نمونه انتخاب‌شده هستند، لذا کسر $\frac{1}{5}$ آماره است نه پارامتر. (پارامتر مربوط به کل جامعه است نه بخشی از آن)
(ج) اگر بخواهیم مزه تک‌تک هندوانه‌ها را امتحان کنیم، دیگر هندوانه‌ای برای فروش باقی نخواهد ماند؛ لذا نمی‌توانیم پارامتر $\frac{\text{تعداد هندوانه‌های بد}}{\text{کل هندوانه‌ها}}$ را مستقیماً محاسبه کنیم ولی می‌توانیم به کمک چند آماره آن را تخمین بزنیم.

۲۴- الف) چون با تمام پلیس‌های پاسگاه مصاحبه شده، سرشماری انجام‌شده نه نمونه‌گیری. پس جامعه آماری در این جا، تمام پلیس‌های پاسگاه مورد نظر هستند و اندازه جامعه، تعداد پلیس‌های آن جاست. ضمناً متغیر تصادفی در این جا، تخلفات می‌باشد.

(ب) در این جا نمونه‌گیری انجام‌شده، چون با همه بچه‌هایی که وارد پارک می‌شوند مصاحبه نشده است و قسمتی از آن‌ها انتخاب شده‌اند. ضمناً کل بچه‌هایی که وارد پارک می‌شوند جامعه آماری محسوب می‌شوند و تعداد آن‌ها، اندازه جامعه است و تمام بچه‌هایی که پنج در میان انتخاب می‌شوند نمونه تصادفی‌اند و تعداد آن‌ها اندازه نمونه است. متغیر تصادفی هم «وسیله بازی مورد علاقه» است.

(ج) چون قسمتی از کل شامپوها بررسی شده لذا نمونه‌گیری انجام شده است. کل شامپوهای تولیدی، جامعه می‌باشد و تعداد آن‌ها اندازه جامعه (حجم جامعه) است. شامپوهایی که ۱۰۰ در میان انتخاب می‌شوند نمونه و تعداد آن‌ها اندازه نمونه است. متغیر هم، مقدار pH شامپوهاست.

۲۵- الف) میانگین (ب) میانه (ج) داده دورافتاده (پرت) - میانه - میانگین

د) میانه - میانه - میانه (ه) میانه - میانگین

۲۶- میانگین یک سری از داده‌ها همیشه بین min و max قرار دارد. در گزینه‌ها تنها عددی که بین ۱۲ و ۲۴ قرار دارد، عدد ۱۸ می‌باشد. پس گزینه (۲) صحیح است.

$$\bar{x} = \frac{\text{مجموع داده‌ها}}{\text{تعداد}} \Rightarrow \bar{x} = \frac{2x - 3 + 7x + 1 + 1 - x}{3} = 16 \Rightarrow \frac{8x - 1}{3} = 16 \Rightarrow 8x - 1 = 48 \Rightarrow 8x = 49 \Rightarrow x = \frac{49}{8} \quad -27$$

$$1, 5, 7, 8, x, 12, 14, 15 \quad -28$$

$$Q_2 = \frac{8+x}{2} = 9 \Rightarrow 8+x = 18 \Rightarrow x = 10$$

$$\left. \begin{aligned} 65 \times 10 = 650 &= \text{تعداد} \times \text{میانگین} = \text{مجموع داده‌های اولیه} \\ 650 + 142 = 792 &= \text{مجموع داده‌های جدید} \end{aligned} \right\} \Rightarrow \bar{x}_{\text{جدید}} = \frac{\text{مجموع جدید}}{\text{تعداد جدید}} = \frac{792}{12} = 66 \text{ کیلوگرم} \quad -29$$

برای تبدیل کیلوگرم به گرم باید ۶۶ را در ۱۰۰۰ ضرب کنیم، پس حاصل برابر با ۶۶۰۰۰ گرم می‌شود و گزینه (۴) صحیح است.