آمار و مدل‌سازی

سال سوم که بودید، یک درس ریاضی بی‌اهمیت داشتید به اسم آمار و مدل‌سازی ریاضی؛ البته بی‌اهمیت از نظر شما! وگرنه آمار مهم‌ترین مطلب ریاضی است که با آن سر و کار دارید. جهت اطلاع، آمار تنها مبحث ریاضی است که در دانشگاه و پایان‌نامه‌تان هم لازم می‌شود. معمولاً دوتا تست ناقابل از آمار در کنکور سراسری می‌آید. گاهی یکی از این سؤالات دشوار و پرمحاسبه هم هست. تا سال ۹۰، آمار در فصل اول ریاضی پیش‌دانشگاهی بود. اما در کتب درسی سال ۹۱-۹۰ حذف شد. برای همین احتمال دارد آمار در برنامه‌ی کلاس‌های درس هم به جای اول کار، به آخر کار تبعید شود. من توصیه می‌کنم آمار را آرام آرام از اول سال بخوانید. از پنج فصل اول کتاب درسی، سؤال‌های متنی هم می‌آید. از کنار این تست‌ها ساده نگذرید.
راستی در این فصل، چند جا مجبور شدم نحوه‌ی محاسبات عددی را توضیح بدهم. لطفاً با دقت و حوصله دنبال کنید. مرسی

مقدمات

۱- مدل‌سازی ریاضی چیست؟

۱) بیان مسئله به زبان ریاضی ۲) تقسیم‌بندی متغیرها ۳) روش محاسبه ۴) نمونه‌ای از سرشماری

۲- کدام نوع مدل‌سازی ریاضی با ارزش‌تر است؟

۱) خطای اندازه‌گیری برابر صفر ۲) نتیجه‌ی حاصل همان پدیده‌ی موردنظر

۳) فقط مفاهیم ریاضی ساده‌تر ۴) مفاهیم ریاضی ساده‌تر ـ نتیجه به پدیده‌های موردنظر نزدیک‌تر

۳- در مورد تفاضل مقدار واقعی و مقدار اندازه‌گیری‌شده، کدام نادرست است؟

۱) کم‌تر از واحد اندازه‌گیری است. ۲) ممکن است منفی باشد. ۳) ممکن است مثبت باشد. ۴) ممکن است صفر باشد.

۴- مدل وزن یک شخص به صورت $W = ۷۲/۵ + E$ کیلوگرم است. خطای این اندازه‌گیری چگونه است؟

۱) نیم کیلوگرم ۲) حداکثر ۱ کیلوگرم ۳) کم‌تر از نیم کیلوگرم ۴) یک کیلوگرم

۵- طول ضلع مربعی را برحسب سانتی‌متر اندازه‌گیری کرده‌ایم، مدل آن به صورت $L = ۶ + E$ است. اگر این طول را برحسب میلی‌متر بیان کنیم، مدل آن چگونه است؟ *(کتاب درسی)*

۱) $L = ۶۰ + \frac{۱}{۱۰}E$ ۲) $L = ۶۰ + ۱۰E$ ۳) $L = ۶۰ + E$ ۴) اندازه‌گیری مجدد انجام شود.

۶- شعاع دایره‌ای به صورت $R = ۳ + E$ مدل‌سازی شده است. مدل مساحت آن چگونه است؟ *(کتاب درسی)*

۱) $S \approx ۹\pi + E$ ۲) $S \approx ۹\pi + ۶E$ ۳) $S \approx ۹\pi + ۶\pi E$ ۴) $S \approx ۹\pi + ۹\pi E$

۷- اگر ضلع مربعی به صورت $۵ + E$ باشد، خطای اندازه‌گیری ضلع در مساحت مربع و محیط آن اثر می‌گذارد. نسبت این تأثیر در مساحت به محیط چه‌قدر است؟

۱) ۲ ۲) ۲/۵ ۳) ۴ ۴) ۵

۸- خطای اندازه‌ی ضلع یک مثلث متساوی‌الاضلاع (a)، در محیط بیشتر اثر گذاشته است تا در مساحت آن. در مورد ضلع این مثلث کدام نتیجه درست است؟

۱) $a < ۲$ ۲) $a < ۲\sqrt{۳}$ ۳) $a < ۳$ ۴) $a < ۴$

۹- در مدل‌سازی ریاضی برای مساحت دایره‌ای به قطر تقریبی ۱۰ واحد، اگر خطای اندازه‌گیری قطر کم‌تر از $\frac{۱}{۶\pi}$ باشد، خطای مساحت تقریباً کم‌تر از چند واحد مربع است؟

۱) $\frac{۵}{۶}$ ۲) $\frac{۵}{۹}$ ۳) $\frac{۵}{۱۲}$ ۴) $\frac{۵}{۱۸}$

۱۰- در مدل‌سازی برای حجم کره‌ای به شعاع ۲cm، می‌خواهیم خطای حجم از واحد اندازه‌گیری ($۱\text{cm}^۳$) بزرگ‌تر نشود، حداکثر خطای شعاع تقریباً چه‌قدر است؟

۱) $\frac{۱}{۱۶\pi}$ ۲) $\frac{۱}{۱۲\pi}$ ۳) $\frac{۱}{۸\pi}$ ۴) $\frac{۱}{۹\pi}$

۱۱- در تابع $y = x^۳ - x$، اگر x را با تقریب $۲ + E$ مدل‌سازی کنیم، مدل y به کدام صورت است؟

۱) $۶ + ۱۲E$
۲) $۸ + ۱۲E$
۳) $۶ + ۱۱E$
۴) $۸ + ۱۱E$

۱۲- اگر مدل طول و عرض یک مستطیل به ترتیب $۵ + E_۱$ و $۳ + E_۲$ باشد، مدل مساحت آن کدام است؟

۱) $۱۵ + E_۱E_۲$
۲) $۱۵ + E_۱ + E_۲$
۳) $۱۵ + ۳E_۲ + ۵E_۱$
۴) $۱۵ + ۳E_۱ + ۵E_۲$

۱۳- مدل شعاع و ارتفاع یک استوانه به ترتیب $R = ۲ + E_۱$ و $h = ۵ + E_۲$ است. در مدل حجم آن $E_۱$ و $E_۲$ به ترتیب چند برابر می‌شوند؟

۱) $۴\pi, ۵\pi$
۲) $۲\pi, ۵\pi$
۳) $۵\pi, ۲۰\pi$
۴) $۴\pi, ۲۰\pi$

۱۴- اضلاع یک مکعب‌مستطیل را به صورت $\begin{array}{l} L_۱ = ۳ + E_۱ \\ L_۲ = ۴ + E_۲ \\ L_۳ = ۶ + E_۳ \end{array}$ **داریم. در مدل حجم آن، کدام جمله را <u>نمی‌بینیم</u>؟** *(کتاب درسی)*

۱) $۱۲E_۳$
۲) $۱۸E_۲$
۳) $۲۴E_۲$
۴) $۲۴E_۱$

۱۵- مجموعه‌ای از افراد یا اشیا که می‌خواهیم درباره‌ی اعضای آن، موضوع خاصی را بررسی کنیم، چه نام دارد؟

۱) نمونه
۲) جامعه‌ی آماری
۳) سرشماری
۴) متغیر

۱۶- در کدام بررسی، اندازه‌ی نمونه برابر اندازه‌ی جامعه است؟ *(تجربی ۸۹)*

۱) نمونه‌ی تصادفی
۲) دسته‌بندی
۳) سرشماری
۴) متغیر کیفی

۱۷- مهم‌ترین بخش آمار را کدام عمل تشکیل می‌دهد؟

۱) دسته‌بندی
۲) نمونه‌گیری
۳) اندازه‌ی جامعه
۴) تعیین شاخص‌ها

۱۸- در یک نمونه‌ی تصادفی ساده چهارتایی از اعداد ۱۰ تا ۲۰ شانس انتخاب کدام عدد از بقیه بیشتر است؟

۱) ۲۰
۲) ۱۲
۳) ۱۳
۴) شانس تمام اعداد با هم برابر است.

۱۹- اگر کامپیوتر عدد تصادفی ۰/۳۷۴ را بدهد، در جامعه‌ای به اندازه‌ی ۲۰۰ کدام عضو باید انتخاب شود؟

۱) ۳۷ام
۲) ۷۴ام
۳) ۷۵ام
۴) ۳۸ام

۲۰- در انتخاب نمونه‌ی تصادفی از ۲۳ تا ۸۷، کامپیوتر عدد تصادفی ۰/۶۱۴ را داده است. کدام عضو انتخاب می‌شود؟

۱) ۳۹
۲) ۴۰
۳) ۶۲
۴) ۵۴

۲۱- کدام طریق برای جمع‌آوری داده‌ها مناسب <u>نیست؟</u> *(خارج ۹۰)*

۱) مصاحبه
۲) الگوی خاص
۳) مشاهده
۴) آزمایش

۲۲- جمع‌آوری داده‌ها به کدام طریق مورد قبول <u>نیست؟</u> *(تجربی ۹۱)*

۱) مصاحبه
۲) مشاهده
۳) انجام آزمایش
۴) پرسش هدایت‌کننده

۲۳- گروه خونی افراد کدام نوع متغیر است؟ *(خارج ۸۷ و تجربی ۹۰)*

۱) کیفی ـ اسمی
۲) کیفی ـ ترتیبی
۳) کمی ـ پیوسته
۴) کمی ـ گسسته

۲۴- خرید و فروش کالا براساس تعداد، تابع کدام نوع متغیر است؟

۱) پیوسته
۲) اسمی
۳) گسسته
۴) ترتیبی

۲۵- کدام داده، کمی گسسته است؟

۱) شدت زلزله
۲) میزان آلودگی هوا
۳) درصد ادبیات یک نفر در کنکور سراسری
۴) هزینه‌ی بیمه‌ی خودرو

۲۶- برای تشکیل یک نمونه‌ی تصادفی ساده، کلید RND یا RAN در ماشین‌حساب را می‌زنیم و عدد تصادفی a به دست می‌آید. سپس شماره‌ی عضو انتخابی برای نمونه (b) را پیدا می‌کنیم. a و b به ترتیب کدام نوع متغیرند؟

۱) کمی پیوسته، کمی پیوسته
۲) کمی پیوسته، کمی گسسته
۳) کمی گسسته، کمی گسسته
۴) کمی گسسته، کمی گسسته

دسته بندی

۲۷- در نوع طبقه‌بندی و تعیین تعداد طبقات داده‌ها، مهم‌ترین ویژگی چیست؟

۱) دامنه‌ی تغییرات ۲) فراوانی ۳) میانه ۴) میانگین

۲۸- کدام عبارت‌ها درست‌اند؟ *(کتاب درسی)*

الف) دامنه‌ی تغییرات جامعه از دامنه‌ی تغییرات نمونه کم‌تر نیست.

ب) از زیادبودن دامنه تغییرات در نمونه، نتیجه می‌شود دامنه تغییرات در جامعه زیاد است.

ج) کوچک‌بودن دامنه‌ی تغییرات در نمونه، نتیجه می‌دهد دامنه‌ی تغییرات جامعه کوچک است.

۱) هر سه نادرست‌اند. ۲) الف و ب ۳) الف و ج ۴) ب و ج

۲۹- دامنه‌ی تغییرات داده‌های ۱۰۱,۱۲۱,۱۱۸,۹۹ کدام است؟

۱) ۱۴ ۲) ۱۸ ۳) ۲۱ ۴) ۲۲

۳۰- اگر به هر داده سه واحد اضافه کنیم، دامنه‌ی تغییرات چه تغییری می‌کند؟

۱) سه واحد اضافه می‌شود. ۲) شش واحد اضافه می‌شود.

۳) سه برابر می‌شود. ۴) تغییری نمی‌کند.

۳۱- اگر در یک دسته‌بندی طول دسته‌ها برابر ۵ و تعداد طبقات ۱۰ باشد، دامنه‌ی تغییرات کدام است؟

۱) ۲ ۲) ۵ ۳) ۱۵ ۴) ۵۰

۳۲- در یک جدول، شش دسته با طول ۵ داریم. اگر بیشترین داده ۵۰ باشد، کم‌ترین داده کدام است؟

۱) ۱۰ ۲) ۲۰ ۳) ۳۰ ۴) ۴۰

۳۳- داده‌های آمار در ۹ طبقه به طول دسته‌ی ۴ ، دسته‌بندی شده‌اند. اگر ۸ داده بین داده‌ی ماکسیمم و مینیمم به آن‌ها اضافه شود و یک واحد از طول دسته کم کنیم، در دسته‌بندی جدید تعداد دسته‌ها کدام است؟

۱) ۱۰ ۲) ۱۱ ۳) ۱۲ ۴) ۱۳

۳۴- اگر داده‌های آماری با ماکسیمم و مینیمم ۷۲ و ۳۲ ، در ۵ دسته، طبقه‌بندی شوند، کران پایین دسته‌ی دوم کدام است؟

۱) ۴۱ ۲) ۴۰ ۳) ۳۷ ۴) ۳۶

۳۵- داده‌های آماری پیوسته، در ۸ دسته تقسیم‌بندی شده‌اند که دسته‌ی آخر ۹۲ – ۸۶ است. کوچک‌ترین این داده‌ها کدام است؟

۱) ۴۰ ۲) ۴۲ ۳) ۴۴ ۴) ۴۸

۳۶- در یک دسته‌بندی با دامنه‌ی تغییرات ۳۰ ، اگر حدود دسته‌ی اول و دوم به صورت [۴۵,۵۱) و [۵۱,۵۷) باشد، تعداد دسته‌ها، کدام است؟

۱) ۴ ۲) ۵ ۳) ۶ ۴) ۷

۳۷- در ۵۶ داده‌ی آماری، بزرگ‌ترین و کوچک‌ترین آن‌ها به ترتیب ۹۶ و ۷۵ هستند. این داده‌ها را در ۷ طبقه دسته‌بندی می‌کنیم. اگر داده‌های هر دسته، یکسان در نظر گرفته شوند، مقدار مشترک آن‌ها در دسته‌ی پنجم چه‌قدر است ؟

۱) ۸۷ ۲) ۸۷/۵ ۳) ۸۸ ۴) ۸۸/۵

۳۸- اگر در یک جدول فراوانی، بزرگ‌ترین داده ۷۰ و طول دسته‌ها ۶ و تعداد دسته‌ها ۵ باشد، ارزش یا اندازه‌ی مشترک داده‌های دسته‌ی اول کدام عدد است؟

۱) ۴۳ ۲) ۴۰ ۳) ۴۲ ۴) ۴۵

۳۹- در جدول فراوانی روبه‌رو، حدود دسته‌ی اول و مرکز دسته‌ی آخر آمده است. کدام درست نیست؟ *(کتاب درسی)*

مرکز	حدود دسته
a	۷ – ۱۱
c	۱۱ – b
f	d – e
k	g – h
۲۵	m – n

۱) e = ۱۹

۲) c = ۱۳

۳) m = ۲۷

۴) k = ۲۱

۴۰- کوچک‌ترین و بزرگ‌ترین داده‌های آماری ۱۷/۲ و ۲۲/۶ هستند. اگر کران پایین دسته‌ی دوم ۱۷/۸ باشد، مرکز دسته‌ی آخر کدام است؟

۱) ۲۱/۷ ۲) ۲۱/۸ ۳) ۲۲/۳ ۴) ۲۲/۴ *(خارج ۸۶)*

۴۱- در یک دسته‌بندی با ۷ دسته، اگر اختلاف بیشترین و کم‌ترین داده ۲۸ و نشان دسته‌ی وسط ۲۶ باشد، حد بالای دسته‌ی آخر کدام است؟

۱) ۳۹ ۲) ۴۰ ۳) ۴۱ ۴) ۴۲

۴۲- داده‌های آماری به ۱۲ طبقه دسته‌بندی شده‌اند. حدود دسته‌ی اول به صورت [۲۳,۲۶) می‌باشد. اگر این داده‌ها به ۹ طبقه دسته‌بندی شوند، مرکز دسته‌ی وسط کدام است؟

۱) ۴۰/۵ ۲) ۴۱ ۳) ۴۱/۵ ۴) ۴۲

۴۳- داده‌های آماری با دامنه‌ی تغییرات ۲۶ در شش دسته با طول صحیح طبقه‌بندی شده‌اند، مجموع طول دسته‌ها چه‌قدر است؟

۱) ۲۶ ۲) ۲۸ ۳) ۳۰ ۴) ۲۴

۴۴- در داده‌های آماری با ماکسیمم و مینیمم ۴۱ و ۱۹، اگر شش دسته‌ی متقارن با طول صحیح داشته باشیم، حدود دسته‌ی دوم کدام است؟

۱) ۲۵ − ۲۲ ۲) ۲۷ − ۲۳ ۳) ۲۶ − ۲۲ ۴) ۲۴ − ۲۱

انواع فراوانی

۴۵- اگر فراوانی طبقات مختلف یک دسته‌بندی ۱,۱,۵,۵,۵,۹,۹,۱۲ باشد، تعداد داده‌ها کدام است؟

۱) ۸ ۲) ۴۷ ۳) ۶ ۴) ۴

۴۶- به منظور از بین بردن تأثیر اندازه‌ی نمونه و امکان مقایسه‌ی نتایج آزمون، محاسبه‌ی کدام مورد لازم است؟

۱) فراوانی نسبی ۲) میانگین وزنی ۳) ضریب تغییرات ۴) فراوانی تجمعی

۴۷- اگر توزیع فراوانی مطلق به صورت زیر باشد، فراوانی نسبی دسته‌ی سوم کدام است؟

x_i	۲	۵	۸	۱۱
f_i	۶	۸	۴	۱۲

۱) $\frac{۲}{۱۵}$ ۲) $\frac{۲}{۵}$

۳) $\frac{۳}{۵}$ ۴) $\frac{۴}{۵}$

۴۸- فراوانی نسبی طبقه‌ای ۰/۰۵ و تعداد داده‌ها ۸۰ است. فراوانی مطلق آن کدام است؟

۱) ۵ ۲) ۴ ۳) ۶ ۴) ۸

۴۹- فراوانی نسبی طبقه‌ای ۰/۱۵ و فراوانی مطلق آن ۱۲ است. در این دسته‌بندی، فراوانی نسبی دسته‌ای با ۱۰ داده چه‌قدر است؟

۱) $\frac{۱}{۶}$ ۲) $\frac{۱}{۷}$ ۳) $\frac{۱}{۸}$ ۴) $\frac{۱}{۹}$

۵۰- هشتاد داده‌ی آماری در ۷ طبقه دسته‌بندی شده‌اند. اگر ۲۰ داده‌ی جدید به این جدول افزوده شود، فراوانی نسبی دسته‌ی وسط تغییر نمی‌کند. نسبت افزایش داده‌های دسته‌ی مذکور به فراوانی مطلق قبلی آن کدام است؟

۱) $\frac{۱}{۸}$ ۲) $\frac{۱}{۵}$ ۳) $\frac{۱}{۴}$ ۴) $\frac{۳}{۸}$

۵۱- در جدول روبه‌رو چه نسبتی از داده‌ها در سه طبقه‌ی اول هستند؟

نماینده‌ی طبقه	۱۷	۲۱	۲۵	۲۹
فراوانی مطلق	۴	۵	۷	۴

۱) ۰/۷ ۲) ۰/۵

۳) ۰/۶ ۴) ۰/۸

۵۲- در داده‌های آماری ۲,۳,۳,۴,۴,۴,۵,۵,۶,۶,۶,۷,۷,۸,۸,۹ درصد فراوانی نسبی مضارب ۳ کدام است؟

۱) ۱۶ ۲) ۲۵ ۳) ۳۷/۵ ۴) ۴۸

۵۳- اندازه‌ی قد ۱۲۰ دانش‌آموز، در جدول زیر دسته‌بندی شده است. فراوانی دسته‌ی چهارم کدام است؟ *(تجربی ۸۳)*

مرکز دسته	۱۵۵	۱۵۸	۱۶۱	۱۶۴	۱۶۷	۱۷۰
درصد فراوانی نسبی	۱۰	۱۵	۱۸	x	۲۰	۱۲

۱) ۲۰
۲) ۲۴
۳) ۲۵
۴) ۳۰

۵۴- در یک نمونه‌گیری از تعداد افراد خانواده‌های یک ساختمان، F تعداد خانواده‌ها با x نفر است. چند درصد از خانواده‌ها ۳ یا ۴ نفری هستند؟

x	۱	۲	۳	۴	۵
F	۹	۱۸	۲۲	۲۶	۵

۱) ۴۵ ۲) ۵۴

۳) ۵۸ ۴) ۶۰

۵۵- دانش‌آموزان یک مدرسه با سال تولد یکسان را وزن‌کشی کرده و عدد صحیح وزن آنان را یادداشت کرده‌ایم. چند درصد آن‌ها وزن کم‌تر از ۵۰ دارند؟ *(خارج ۸۸)*

وزن	۴۶	۴۷	۴۸	۴۹	۵۰	۵۱
تعداد	۸	۹	۱۲	۱۵	۶	۵

۱) ۷۲ ۲) ۷۵

۳) ۷۸ ۴) ۸۰

۵۶- در دسته‌بندی ۱۲۰ داده‌ی آماری در ۹ طبقه، دسته‌ی اول به صورت ۲۲ − ۲۵ می‌باشد. می‌دانیم ۴۵ درصد داده‌ها کم‌تر از ۳۴ و فراوانی نسبی دسته‌ی وسط ۰/۲ است. تعداد داده‌های کم‌تر ۳۷ از کدام است؟

۱) ۶۷ ۲) ۷۶ ۳) ۷۸ ۴) ۸۷

۵۷- ۷۵ داده‌ی آماری در ۷ طبقه دسته‌بندی شده‌اند؛ کوچک‌ترین داده‌ها ۲۷ و بزرگ‌ترین آن‌ها ۴۷/۸ می‌باشد. می‌دانیم ۲۸ درصد داده‌ها کم‌تر از ۳۶ و ۴۰ درصد داده‌ها کم‌تر از ۳۹ می‌باشند. فراوانی مطلق دسته‌ی وسط کدام است؟

۱) ۸ ۲) ۹ ۳) ۱۰ ۴) ۱۲

۵۸- توزیع تعداد فرزندان ۲۰ خانواده به صورت زیر است. فراوانی تجمعی دسته‌ی چهارم کدام است؟

حدود طبقات	۰−۱	۲−۳	۴−۵	۶−۷	۸−۹
فراوانی مطلق	۲	۸	۵	۳	۲

۱) ۵ ۲) ۱۰ ۳) ۱۵ ۴) ۱۸

۵۹- فراوانی نسبی دسته‌ها در یک جدول به ترتیب ۰/۳، ۰/۱۵، ۰/۲ و ... است. اگر تعداد کل داده‌ها ۶۰ باشد، فراوانی تجمعی دسته‌ی دوم کدام است؟

۱) ۱۵ ۲) ۱۸ ۳) ۲۴ ۴) ۲۷

۶۰- داده‌های جدول مقابل، داده‌های آماری پیوسته است. چند درصد از داده‌ها، در فاصله‌ی [۱۸/۵,۲۱/۵) قرار دارند؟ *(تجربی ۸۸)*

مرکز دسته	۱۴	۱۷	۲۰	۲۳	۲۶
فراوانی تجمعی	۵	۱۳	۲۵	۳۴	۴۰

۱) ۲۰ ۲) ۲۵ ۳) ۳۰ ۴) ۴۰

۶۱- در جدول فراوانی تجمعی داده‌های دسته‌بندی‌شده‌ی روبه‌رو، اگر درصد فراوانی نسبی دسته‌ی وسط ۲۴ باشد، فراوانی مطلق دسته‌ی چهارم کدام است؟

مرکز دسته	۱۳	۱۵	۱۷	۱۹	۲۱
فراوانی تجمعی	۵	۱۴	a	۴۱	۵۰

۱) ۱۴ ۲) ۱۵ ۳) ۱۶ ۴) ۱۷

۶۲- در داده‌های روبه‌رو درصد فراوانی تجمعی متناظر با $x_3 = ۴$ کدام است؟

x_i	۲	۳	۴	۵	۶
F_i	۵	۷	۸	۵	۵

۱) ۸ ۲) ۳۰ ۳) ۳۳/۳ ۴) ۶۶/۷

نمودارها

۶۳- در نمودار ساقه و برگ زیر، زمانی که طول می‌کشد تا آمار مسمومیت در فرد ظاهر شود را آورده‌ایم. داده‌ها برحسب دقیقه و با یک رقم اعشار هستند. دامنه‌ی تغییرات کدام است؟

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

۱) ۸/۷ ۲) ۵/۱ ۳) ۳/۶ ۴) ۵/۵

۶۴- در نمودار ساقه و برگ مقابل، اندازه‌ی نمونه برابر کدام است؟

ساقه	برگ
۱	۰ ۳ ۳ ۴
۲	۰ ۲ ۴ ۴ ۸ ۸
۳	۲

۱) ۱۰ ۲) ۱۲ ۳) ۱۱ ۴) ۳۲

۶۵- در نمودار ساقه و برگ روبه‌رو به جای ☐ چند عدد مختلف می‌توان نوشت؟ *(کتاب درسی)*

ساقه	برگ
۱	۲ ☐ ۵ ۶
۲	۷ ۹
۳	۲ ۴ ☐

۱) ۴ ۲) ۳ ۳) ۲ ۴) ۱

۶۶- در نمودار ساقه و برگ ۵۰ داده بین ۱۰ و ۲۰، برخی داده‌ها یک رقم اعشاری دارند. برگ‌های مربوط به ساقه‌ی ۱۴، به صورت ۰۳۵۵۵۵۶۸۸۹ هستند. چند درصد داده‌ها دقیقاً ۱۴/۵ بوده‌اند؟

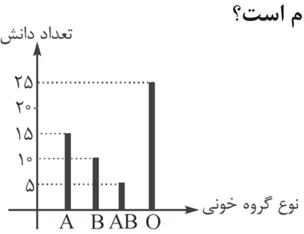
۱) ۴ ۲) ۸ ۳) ۱۰ ۴) ۲۰

۶۷- در نمودار ساقه و برگ مقابل چند درصد داده‌ها ناکم‌تر از ۴۰ و کم‌تر یا مساوی ۴۷ است؟

برگ	ساقه
۲۳۳۳۴۴۵۶۷	۳
۰۱۳۴۴۵۷۷	۴
۱۱۱۲۲۳۳۴	۵

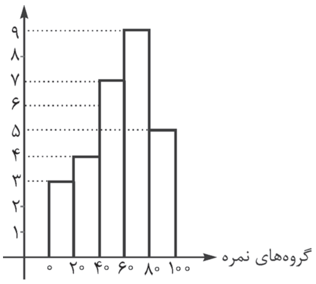
۱) ۲۴
۲) ۲۵
۳) ۲۸
۴) ۳۲

۶۸- در یک کلاس، توزیع گروه خون به شکل نمودار میله‌ای مقابل است. تعداد دانش‌آموزان این کلاس کدام است؟



۱) ۲۵
۲) ۳۰
۳) ۵۰
۴) ۵۵

۶۹- نمرات دانش‌آموزان در یک آزمون تستی، در نمودار زیر آمده است. چند نفر کم‌تر از ۶۰ درصد گرفته‌اند؟

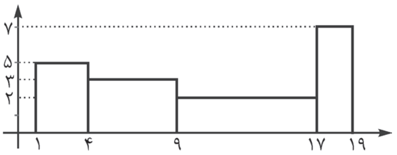


۱) ۷
۲) ۱۲
۳) ۱۳
۴) ۱۴

۷۰- در نمودار مستطیلی با طول دسته‌های مساوی، کدام ویژگی مستطیل‌ها مقایسه می‌شوند؟

۱) قاعده
۲) محیط
۳) مساحت
۴) ارتفاع

۷۱- در نمودار مستطیلی مقابل، فراوانی در کدام دسته بیشتر است؟



۱) [۱,۴)
۲) [۴,۹)
۳) [۹,۱۷)
۴) [۱۷,۱۹]

۷۲- در توزیع فراوانی داده‌های پیوسته، کدام نمودار مناسب است؟ *(تجربی ۸۷)*

۱) مستطیلی
۲) چندبر فراوانی
۳) میله‌ای
۴) دایره‌ای

۷۳- چه نموداری برای توزیع درصدهای انواع هزینه‌های یک شرکت مناسب است؟

۱) چندبر فراوانی
۲) دایره‌ای
۳) چندبر فراوانی نسبی
۴) مستطیلی

۷۴- در یک نمودار چندبر فراوانی نقاط (۱۱/۵,۳)، (۱۴/۵,۶)، (۱۷/۵,۷) و (۲۰/۵,۴) را داریم. کدام درست نیست؟

۱) چهار دسته داریم.
۲) طول دسته‌ها ۳ است.
۳) اندازه‌ی نمونه ۲۰ است.
۴) فراوانی تجمعی دسته‌ی دوم ۱۰ است.

۷۵- در جدول فراوانی داده‌های پیوسته و دسته‌بندی‌شده، نقاط (۲۱,۴۲) و (۲۴,۴۷) متوالیاً از نمودار چندبر هستند. حدود دسته‌ی بعدی چگونه است؟

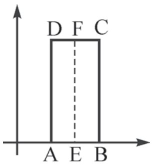
۱) [۲۴,۲۷)
۲) [۴۷,۵۲)
۳) [۲۲/۵,۲۵/۵)
۴) [۲۵/۵,۲۸/۵)

۷۶- برای دسته‌بندی زیر، نقطه‌ی متناظر با دسته‌ی وسط روی نمودار چندبر کدام است؟

۳۲ – ۳۵	۲۹ – ۳۲	۲۶ – ۲۹	۲۳ – ۲۶	۲۰ – ۲۳	حدود دسته
۴۰	۳۲	۲۵	۱۹	۴	فراوانی تجمعی

۱) (۲۷/۵,۷)
۲) (۲۷/۵,۶)
۳) (۲۹,۲۵)
۴) (۲۶,۲۵)

۷۷- مستطیل ABCD قسمتی از نمودار مستطیلی است. کدام خط جزو نمودار میله‌ای است؟



۱) AD
۲) BC
۳) EF
۴) DC

۷۸- در داده‌های آماری دسته‌بندی‌شده، مساحت نمودار مستطیلی آن را S و سطح زیر نمودار چندبر فراوانی را که دو سر آن بر محور افقی باشد، S′ می‌نامیم. نسبت $\frac{S}{S'}$ چگونه است؟ *(خارج ۸۹)*

۱) کوچک‌تر از ۱
۲) بزرگ‌تر از ۱
۳) برابر ۱
۴) اظهارنظر نمی‌توان کرد

۷۹- سطح زیر نمودار چندبر فراوانی و نمودار مستطیلی یک جدول با هم برابرند. در این حالت برای نمودار چندبر فراوانی ۸ نقطه روی محور x ها و ۶ نقطه روی محور y ها داریم. تعداد دسته‌ها کدام است؟

۱) ۵
۲) ۶
۳) ۷
۴) ۸

۸۰- دسته‌ی اول یک جدول فراوانی (۹,۱۵] و فراوانی آن ۷ است. طول نقاط اول و دوم چندبر فراوانی تکمیل‌شده، چه اعدادی هستند؟

۱) صفر و ۶
۲) ۵ و ۶
۳) ۶ و ۱۲
۴) صفر و ۱۲

۸۱- در یک دسته‌بندی، کم‌ترین و بیشترین داده، ۶ و ۱۸ هستند و چهار طبقه داریم. انتهای نمودار چندبر فراوانی تکمیل‌شده در کدام نقطه به محور x ها وصل می‌شود؟

۱) ۲۱
۲) ۲۰/۵
۳) ۲۰
۴) ۱۹/۵

۸۲- رسم نمودار دایره‌ای که ترتیب کنار هم قرارگرفتن نواحی آن اهمیت نداشته باشد، برای کدام متغیر مناسب‌تر است؟

۱) مراحل رشد
۲) تعداد تصادفات
۳) وزن افراد
۴) گروه خون

۸۳- در ۱۸۰ داده‌ی آماری، فراوانی طبقه‌ای ۲۰ است. زاویه‌ی مربوط به این طبقه در نمودار دایره‌ای چه‌قدر است؟

۱) ۲۰°
۲) ۴۰°
۳) ۶۰°
۴) ۸۰°

۸۴- تعداد دانش‌آموزان سال چهارم ریاضی، تجربی و انسانی به ترتیب ۷۰، ۶۰ و ۵۰ نفر است. در نمودار دایره‌ای این سه کلاس، زاویه‌ی مربوط به کلاس انسانی چند درجه است؟

۱) ۵۰°
۲) ۱۵۰°
۳) ۱۲۰°
۴) ۱۰۰°

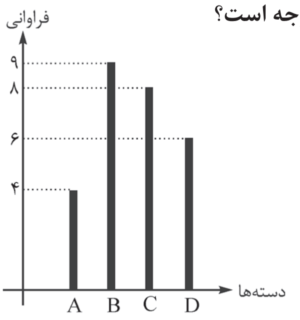
۸۵- در نمودار دایره‌ای وضعیت تحصیلی ۴۸ کارمند یک شرکت، زاویه‌ی مرکزی مربوط به «کارشناسی» ۴۵° بوده است. تعداد کارمندان با مدرک کارشناسی کدام است؟

۱) ۳
۲) ۶
۳) ۱۵
۴) ۱۶

۸۶- داده‌های آماری در ۹ طبقه دسته‌بندی شده‌اند، فراوانی تجمعی نسبی در دسته‌ی چهارم و پنجم به ترتیب ۰/۲۸ و ۰/۴۰ است. در نمودار دایره‌ای، زاویه‌ی مربوط به دسته‌ی پنجم چند درجه است؟

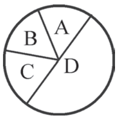
۱) ۴۰/۵
۲) ۴۱/۴
۳) ۴۲/۶
۴) ۴۳/۲

۸۷- شکل مقابل، نمودار میله‌ای داده‌ها در ۴ دسته است. در نمودار دایره‌ای، زاویه‌ی مرکزی دسته‌ی D چند درجه است؟



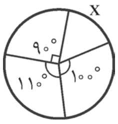
۱) ۶۰
۲) ۷۵
۳) ۸۰
۴) ۹۰

۸۸- در نمودار مقابل تعداد افراد دسته‌های A، B، C و D به نسبت ۶,۳,۲,۱ است. زاویه‌ی دسته‌ی B کدام است؟



۱) ۲۰°
۲) ۳۰°
۳) ۴۰°
۴) ۶۰°

۸۹- در بین ۱۵۰۰۰ نفر، نمودار دایره‌ای مقابل برای ۴ گروه سنی به دست آمده است. چند نفر در گروه سنی x قرار دارند؟



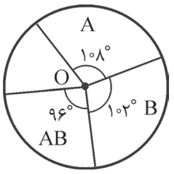
۱) ۶۰۰۰
۲) ۳۰۰۰
۳) ۲۵۰۰
۴) ۱۵۰۰

۹۰- شرکتی ۱۶۰ کارمند دارد که مدارک تحصیلی آنان با ۶ کد متمایز مشخص شده‌اند. در نمودار دایره‌ای، زاویه‌ی مرکزی هر گروه با واحد درجه مطابق جدول روبه‌رو است. تعداد کارکنان با کد ۴ کدام است؟ (خارج ۹۰)

۶	۵	۴	۳	۲	۱	کد
۱۸	۵۴	α	۹۹	۴۵	۲۷	زاویه‌ی مرکزی

۱) ۵۲
۲) ۵۴
۳) ۵۶
۴) ۵۸

۹۱- نمودار دایره‌ای مربوط به اهدای خون افراد مراجعه‌کننده به یک ایستگاه انتقال خون، به شکل مقابل است. چند درصد این افراد در گروه خونی O قرار دارند؟



۱) ۱۵
۲) ۱۶
۳) ۱۸
۴) ۲۰

۹۲- درصد فراوانی نسبی گروه‌های خونی مختلف در یک شهر به صورت زیر است. زاویه‌ی سطح مربوط به گروه خونی O در نمودار دایره‌ای کدام است؟

O	AB	B	A	گروه
a	۳۶	۲۲/۵	۲۴	درصد

۱) ۴۵ ۲) ۵۴

۳) ۶۳ ۴) ۷۲

۹۳- در جدول فراوانی تجمعی داده‌های دسته‌بندی‌شده به شکل زیر، زاویه‌ی مرکزی متناسب با فراوانی مطلق دسته‌ی وسط در نمودار دایره‌ای ۹۰ درجه است. فراوانی مطلق دسته‌ی چهارم کدام است؟

۲۰−۲۲	۱۸−۲۰	۱۶−۱۸	۱۴−۱۶	۱۲−۱۴	حدود دسته
۶۰	۴۸	x	۱۷	۶	فراوانی تجمعی

۱) ۱۴

۲) ۱۵

۳) ۱۶

۴) ۱۸

شاخص‌های مرکزی

۹۴- مد در داده‌های آماری با نمودار ساقه و برگ مقابل کدام است؟

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

کلید: ۱ ۴ = ۱۴

۱) ۳

۲) ۴

۳) ۱۴

۴) ۱۳

۹۵- مد داده‌های ۱,۲,۷,۱,۱,۸,۲,۷,۱,۷ کدام است؟

۱) ۱ ۲) ۲ ۳) ۴ ۴) ۳

۹۶- در جدول داده‌های دسته‌بندی‌شده‌ی زیر، مد کدام است؟

۱۹	۱۷	۱۵	۱۳	۱۱	مرکز دسته
۴۰	۳۲	۲۰	۱۳	۸	فراوانی تجمعی

۱) ۱۲ ۲) ۱۵

۳) ۱۷ ۴) ۱۹

۹۷- در نمودار ساقه و برگ داده‌های آماری مقابل، میانه کدام است؟

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

۱) ۶۱ ۲) ۶۲

۳) ۶۷ ۴) ۷۵

۹۸- وضع نمرات دانش‌آموزان یک کلاس با نمره‌های امتحانی ۱,۴,۷,۱۰,۱۰,۱۰,۱۳,۱۳,۱۴,۱۴,۱۵,۱۵,۱۷ با کدام شاخص بهتر نمایان می‌شود؟

۱) مد ۲) میانه ۳) دامنه‌ی تغییرات ۴) میانگین

۹۹- در اعداد ۷,۱۵,۲۴,۱۱,۵۹,۳۵,۴۱ میانه کدام است؟

۱) ۱۱ ۲) ۲۴ ۳) ۳۵ ۴) ۵۹

۱۰۰- در جدول دسته‌بندی داده‌ها به شکل زیر، میانه کدام است؟

۱۴	۱۲	۱۰	۸	۶	مرکز دسته
۲۵	۱۸	۱۲	۷	۳	فراوانی تجمعی

۱) ۱۰ ۲) ۱۱

۳) ۱۲ ۴) ۱۳

۱۰۱- در کدام گروه از داده‌ها، میانه یکی از داده‌ها نیست ؟

۱) ۱,۲,۳,۴,۷ ۲) ۱,۲,۵,۸,۹ ۳) ۱,۲,۴,۴,۶,۷ ۴) ۱,۲,۴,۶,۷,۸

۱۰۲- در بین ۸ داده، یک مد وجود دارد که فراوانی‌اش ۲ است. اگر میانه و مد داده‌ها با هم برابر باشد «چندمین عضوها» در داده‌های مرتب‌شده، مد هستند؟

۱) سوم و چهارم ۲) چهارم و پنجم ۳) پنجم و ششم ۴) سوم و پنجم

۱۰۳- در داده‌های ۲۵,۳۱,۱۹,۱۲,۱۰,۱۵,۲۳,۱۴,۲۷,۱۶,۳۴,۴۱,۴۳,۳۲,۱۸ چارک اول کدام است؟

۱) ۱۲ ۲) ۱۳ ۳) ۱۴ ۴) ۱۵

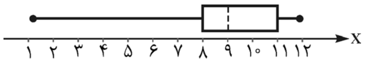
۱۰۴- در داده‌های آماری ۸,۷,۲۱,۱۹,۵,۱۰,۱۲,۹,۱۱,۱۷,۱۴,۱۶,۲۳,۲۰,۱۸ اختلاف چارک اول و سوم کدام است؟

۱) ۱۰ ۲) ۱۱ ۳) ۱۲ ۴) ۱۳

۱۰۵- در نمودار جعبه‌ای برای داده‌های ۱۰,۲۲,۲۳,۹,۱۱,۲۶,۲۰,۱۸,۲۱,۱۹,۱۳,۱۶,۱۴,۷,۱۲ طول دنباله‌ی سمت چپ چه‌قدر است؟

۱) ۷ ۲) ۴ ۳) ۵ ۴) ۱۱

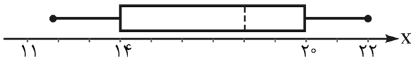
۱۰۶- در نمودار جعبه‌ای شکل مقابل مقدار .... برابر .... است.



۱) چارک اول، ۷ ۲) میانه، ۶/۵

۳) چارک سوم، ۱۱ ۴) دامنه‌ی تغییرات، ۱۲

۱۰۷- در نمودار جعبه‌ای مقابل کدام نادرست است؟



۱) تقریباً ۵۰ درصد داده‌ها بیشتر از ۱۸ هستند.

۲) تقریباً ۵۰ درصد داده‌ها بین ۱۴ تا ۲۰ هستند.

۳) تقریباً ۲۵ درصد داده‌ها از ۱۴ بیشترند.

۴) تقریباً ۲۵ درصد داده‌ها از ۲۰ بیشترند.

۱۰۸- میانگین داده‌های ۱۶,۱۵,۱۴,۱۴,۱۲ کدام است؟

۱) ۱۴ ۲) ۱۴/۱ ۳) ۱۴/۲ ۴) ۱۴/۵

۱۰۹- داده‌های آماری با یک رقم اعشار با نمودارساقه و برگ داده شده‌اند، میانگین آن‌ها کدام است؟ *(تجربی ۸۶)*

ساقه	برگ
۸	۰۰۱۲۲۵۶۷
۹	۰۱۲۳۳۴۵۵
۱۰	۱۱۲۲

۱) ۹/۰۵ ۲) ۹/۰۶

۳) ۹/۰۷ ۴) ۹/۰۸

۱۱۰- نمرات درس فیزیک دانش‌آموزی در طول سال ۱۷,۱۹,۱۸,۷,۱۸/۵,۱۹ شده است. اگر قرار باشد برای جبران نمره‌ی ۷، امکان امتحان مجدد به این دانش‌آموز بدهند، باید چه نمره‌ای بگیرد تا میانگین وی ۱۸/۵ شود؟ *(کتاب درسی)*

۱) ۱۹ ۲) ۱۹/۵ ۳) ۲۰ ۴) امکان‌پذیر نیست

۱۱۱- به ازای کدام مقدار x در اعداد {۸۰,۱۱۰,۷۰,۶۰,x} میانه و مد و میانگین مساوی‌اند؟ *(کتاب درسی)*

۱) ۷۰ ۲) ۸۰ ۳) ۷۵ ۴) ۸۵

۱۱۲- میانگین داده‌های ۱۷۵۰,۱۷۶۰,۱۷۶۶,۱۷۵۵,۱۷۸۰ کدام است؟ *(کتاب درسی)*

۱) ۱۷۶۲/۱ ۲) ۱۷۶۲/۲ ۳) ۱۷۶۲/۳ ۴) ۱۷۶۲/۴

۱۱۳- در داده‌های ۱۸,۱۴,۱۶,۲۰,۲۴,۱۵,۱۴,۱۲,۲۶,۲۱,۲۰,۲۵ میانگین «داده‌های بزرگ‌تر از چارک اول و کوچک‌تر از چارک سوم» تقریباً کدام است؟

۱) ۱۸/۲۵ ۲) ۱۸/۳۳ ۳) ۱۸/۶۶ ۴) ۱۸/۷۵ *(خارج ۸۵)*

۱۱۴- در داده‌های آماری، میانگین را ۱۲ حدس زدیم و تفاضل میانگین حدسی از داده‌ها، اعداد −۴,−۳,−۱,۲,۳ شده است. میانگین واقعی کدام است؟

۱) ۱۱ ۲) ۱۱/۴ ۳) ۱۱/۶ ۴) ۱۲/۴

۱۱۵- در داده‌های آماری با نمودار ساقه و برگ، داده‌های کم‌تر از چارک اول و بیشتر از چارک سوم را حذف می‌کنیم. میانگین داده‌های باقی‌مانده کدام است؟

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

۱) ۴۲/۷ ۲) ۴۲/۹

۳) ۴۳/۲ ۴) ۴۳/۴

۱۱۶- داده‌های آماری، اعداد متمایزند و میانه‌ی آن‌ها در حالت مرتب‌شده، برابر میانگین داده‌های چهارم و پنجم است. اگر مجموع کل داده‌ها ۳۶۰ باشد، میانگین کل داده‌ها کدام است؟

۱) ۳۶ ۲) ۴۰ ۳) ۴۵ ۴) ۳۰

۱۱۷- میانگین داده‌های ۷,۱۳,۱۹,۲۵,...,۶۱ کدام است؟

۱) ۳۰/۵ ۲) ۳۱ ۳) ۳۳ ۴) ۳۴

۱۱۸- میانگین داده‌های a + ۳,a + ۶,a + ۹,...,a + ۱۰۵ کدام است؟ *(خارج ۸۶ و ۸۸)*

۱) ۲a + ۱۰۸ ۲) a + ۵۴ ۳) ۳۵a + ۱۰۸ ۴) ۳۵a + ۵۴

۱۱۹- مجموع ۷ عدد متوالی برابر ۱۴۷ است. اگر میانگین این اعداد از میانه کم شود، چه عددی به دست می‌آید؟ *(کتاب درسی)*

۱) ۲۱ ۲) ۷ ۳) −۷ ۴) صفر

۱۲۰- میانگین داده‌های $a,a+۱,a+۲,a+۳,a+۴$ از میانه‌ی داده‌های $۲a+۳,۲a+۵,۲a+۷,۲a+۹,۲a+۱۱$ چه‌قدر کم‌تر است؟ *(آزاد ۸۵)*

۱) ۵ ۲) $۲a+۱۰$ ۳) $۲a$ ۴) $a+۵$

۱۲۱- نمرات یک دانش‌آموز ۱۸,۱۷,۱۶,۱۵,۱۱,۷,۷ هستند. بیشترین اختلاف بین کدام دو شاخص است؟

۱) میانه و مد ۲) میانگین و مد ۳) مد و دامنه‌ی تغییرات ۴) میانگین و دامنه‌ی تغییرات

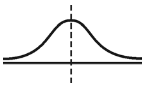
۱۲۲- در داده‌های ۱۲,۹,۷,۳ با تغییر کدام داده به اندازه‌ی یک واحد می‌توان اندازه‌ی میانگین و میانه و دامنه‌ی تغییرات را با هم برابر کرد؟

۱) ۳ ۲) ۷ ۳) ۹ ۴) ۱۲

۱۲۳- در داده‌های ۷,۹,۱۰,۱۰,۱۴ کدام شاخص زیر با تغییر اندازه‌ی ۹ به ۱۰ تغییر می‌کند؟

۱) دامنه‌ی تغییرات ۲) میانه ۳) میانگین ۴) مد

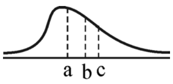
۱۲۴- در شکل روبه‌رو که منحنی نرمال توزیع فروانی است، کدام درست است؟ *(کتاب درسی)*



۱) از تقارن آن نتیجه می‌شود میانگین و مد برابرند. ۲) از تقارن ان نتیجه می‌شود میانگین و میانه برابرند.

۳) چون نقطه‌ی ماکسیمم دارد مد برابر میانه است. ۴) چون نقطه‌ی ماکسیمم دارد میانه برابر میانگین است.

۱۲۵- در شکل روبه‌رو یک توزیع غیرمتقارن را می‌بینید. کدام نتیجه درست است؟



۱) a میانه است. ۲) a میانگین است.

۳) میانگین از میانه کم‌تر است. ۴) میانه از میانگین کم‌تر است.

۱۲۶- در بررسی ۶ داده‌ی آماری، دامنه‌ی تغییرات و میانگین به ترتیب ۵ و ۱۵ هستند. اگر داده‌ها در نیمه‌ی بعد از میانه، اعداد متوالی باشند، مد داده‌ها با فراوانی ۳ کدام عدد می‌تواند باشد؟

۱) ۱۲ ۲) ۱۳ ۳) ۱۴ ۴) ۱۶

۱۲۷- اگر شانزده داده‌ی آماری $x_۱,x_۲,x_۳,...,x_{۱۶}$ داشته باشیم و $\bar{x}$ میانگین آن‌ها باشد، حاصل $(x_۱-\bar{x})+(x_۲-\bar{x})+...+(x_{۱۶}-\bar{x})$ کدام است؟

۱) ۱۶ ۲) -۱۶ ۳) ۸ ۴) صفر

۱۲۸- میانگین داده‌های آماری $x_۱,x_۲,x_۳,...,x_۷,۵$ برابر ۸ است. حاصل $(x_۱-۸)+(x_۲-۸)+...+(x_۷-۸)$ کدام است؟

۱) صفر ۲) ۵ ۳) ۳ ۴) -۵

۱۲۹- از شش داده‌ی آماری، میانگین آن‌ها را کم کردیم و اعداد $-۴,a,-۱,۳,۱,۲$ به دست آمده است. مقدار a کدام است؟

۱) ۱ ۲) -۱ ۳) صفر ۴) ۲

۱۳۰- انحراف شش داده‌ی آماری از میانگین آن‌ها $-۳,-۲,-۱,۰,۳,۳$ است. اگر کم‌ترین داده ۱۷ باشد، میانه‌ی این داده‌ها کدام است؟

۱) ۱۹ ۲) ۱۹/۵ ۳) ۲۰ ۴) ۲۰/۵

۱۳۱- به داده‌های ۳,۶,۷,۸ کدام عدد را اضافه کنیم تا میانگین پنج داده‌ی حاصل یک واحد بیشتر از داده‌های اولیه گردد؟ *(آزاد ۸۳)*

۱) ۸ ۲) ۹ ۳) ۱۰ ۴) ۱۱

۱۳۲- میانگین داده‌های $-۱,-۲,-۳,-۴,-۵$ و داده‌های $۹۵,۹۶,۹۷,۹۸,a$ برابرند. a کدام است؟ *(آزاد ۸۰)*

۱) ۱۰۰ ۲) -۴۰۱ ۳) -۳۹۹ ۴) -۱۰۱

۱۳۳- میانگین داده‌های $۲,۳,۴,۴,a$ برابر ۲ است. میانگین داده‌های $۲,۳,۴,۴,(۶a+۱۰۰)$ چه‌قدر است؟ *(آزاد ۸۰)*

۱) ۱۸ ۲) ۲۰ ۳) ۱۹ ۴) ۲۲

۱۳۴- میانگین داده‌های $۱,۱,۲,۳,۴,...,۹۹,۱۰۰$ چه‌قدر از میانگین داده‌های $۱,۲,۳,...,۹۹,۱۰۰,۱۰۰$ کم‌تر است؟ (در اولی دو بار یک و در دومی دو بار صد تکرار شده.)

۱) $\frac{۹۹}{۱۰۱}$ ۲) $\frac{۹۹}{۱۰۰}$ ۳) ۱ ۴) $\frac{۹۹}{۲۰۲}$

۱۳۵- اگر میانگین داده‌های $x_۱+۱,x_۲+۲,x_۳+۳$ برابر $\bar{x}$ باشد، میانگین داده‌های $۳x_۱+۱,۳x_۲+۲,۳x_۳+۳$ کدام است؟ *(آزاد ۸۱)*

۱) $۳\bar{x}-۶$ ۲) $۳\bar{x}-۴$ ۳) $۳\bar{x}$ ۴) $۳\bar{x}-۲$

۱۳۶- معدل یک دانش‌آموز در هشت درس مختلف ۱۲/۵ بوده. نمرات دو درس دیگر او ۱۴ و ۱۶ هستند. معدل وی در ده درس کدام است؟

۱) ۱۳/۷۵ ۲) ۱۳/۵ ۳) ۱۴ ۴) ۱۳

۱۳۷- اگر میانگین داده‌های $x_۱,x_۲,x_۳,x_۴$ برابر $\bar{x}$ باشد، میانگین داده‌های $۲x_۱+x_۲,۲x_۲+x_۳,۲x_۳+x_۴,۲x_۴+x_۱$ کدام است؟

۱) $\bar{x}$ ۲) $۲\bar{x}$ ۳) $۳\bar{x}$ ۴) $۳\bar{x}+۱$

۱۳۸- اگر میانگین داده‌های $a,b,c,d,۷$ برابر ۱۱ باشد، میانگین داده‌های $a.b,c,d,۳,۹$ کدام است؟

۱) ۱۰	۲) ۱۱	۳) ۱۱/۵	۴) ۱۲

۱۳۹- میانگین شش داده‌ی آماری ۱۸ و میانگین سه داده‌ی دیگر ۱۲ است. میانگین هر ۹ داده کدام است؟ *(آزاد ۹۱)*

۱) ۱۴	۲) ۱۵	۳) ۱۷	۴) ۱۶

۱۴۰- یک گروه ۱۸تایی از داده‌های آماری با میانگین ۲۲ را با گروه دیگری با میانگین ۲۵ ادغام کردیم. اگر میانگین کل ۲۳ شده باشد، تعداد داده‌های گروه با میانگین ۲۵ کدام است؟

۱) ۶	۲) ۹	۳) ۱۸	۴) ۲۷

۱۴۱- میانگین ۱۰ عدد، مساوی ۱۲ شده است. اگر یک عدد را کنار بگذاریم، میانگین ۹ عدد باقی‌مانده مساوی ۱۱ می‌شود. عددی که کنار گذاشته شده است، کدام است؟ *(آزاد ۹۱)*

۱) ۲۱	۲) ۲۰	۳) ۱۲	۴) ۱۱

۱۴۲- اگر میانگین داده‌های $a,a,a,a,a+۱$ برابر $\frac{۳a}{۲}$ باشد، میانگین داده‌های $a,a+۱,a+۲,a+۳,a+۴$ کدام است؟

۱) $\frac{۱۲}{۵}$	۲) $\frac{۹}{۲}$	۳) $\frac{۷}{۲}$	۴) $\frac{۷a}{۲}$

۱۴۳- اگر میانگین اعداد $x_۱,x_۲,...,x_۹$ برابر ۱۰ باشد، میانگین داده‌های $x_۱-۱,x_۲-۳,x_۳-۵,...,x_۸-۱۵,x_۹-۱۷$ کدام است؟

۱) ۱	۲) ۹	۳) ۱۰	۴) ۱۸

۱۴۴- میانگین چند داده برابر ۵۷ است. ابتدا از هر داده ۱۲ واحد کم و سپس داده‌های حاصل را سه برابر کرده‌ایم. میانگین داده‌های نهایی کدام است؟

۱) ۴۵	۲) ۷۰	۳) ۱۳۵	۴) ۱۵۹ *(خارج ۸۴)*

۱۴۵- میانگین داده‌های $۲x_۱-۱,۲x_۲-۱,...,۲x_n-۱$ برابر ۱۷ است. میانگین داده‌های $x_۱+۲,x_۲+۲,...,x_n+۲$ کدام است؟

۱) ۱	۲) ۱۰	۳) ۱۱	۴) ۱۲

۱۴۶- بین x و y رابطه‌ی $y=۳x-۱۷$ برقرار است. اگر میانگین x ها ۸ باشد، میانگین y ها کدام است؟

۱) ۸	۲) ۲۴	۳) ۷	۴) ۱۷

۱۴۷- به هر داده‌ی آماری ۵ واحد اضافه می‌کنیم. کدام مورد تغییر می‌کند؟

۱) دامنه‌ی تغییرات	۲) چارک اول

۳) طول دنباله‌ی سمت راست نمودار جعبه‌ای	۴) اختلاف چارک سوم و اول

۱۴۸- در ده داوطلب، ۱ نفر نمره‌ی ۷۵، ۴ نفر نمره‌ی ۸۰، ۲ نفر نمره‌ی ۸۵ و ۳ نفر دیگر نمره‌ی ۹۰ آورده‌اند. میانگین آن‌ها چه‌قدر است؟

۱) ۸۲/۵	۲) ۸۳/۵	۳) ۸۴/۵	۴) ۸۵/۵

۱۴۹- در داده‌های

x	۲	۳	۴	۵	۶
f	۵	۷	۸	۵	۵

میانگین تقریباً کدام است؟

۱) ۳	۲) ۳/۹	۳) ۴/۱	۴) ۴/۷

۱۵۰- اگر در جدول روبه‌رو میانگین ۱۰ باشد، داده با فراوانی ۹ چه عددی است؟

داده	۷	?
فراوانی	۶	۹

۱) ۱۲	۲) ۱۱	۳) ۸	۴) ۵

۱۵۱- در جدول داده‌های مقابل، تفاوت مد از میانگین کدام است؟

داده	۱۲	۱۴	۱۶	۱۸	۲۰
فراوانی	۲	۳	۲	۲	۱

۱) ۰/۴	۲) ۰/۷	۳) ۱/۲	۴) ۱/۴

۱۵۲- میانگین داده‌های آماری جدول مقابل کدام است؟

حدود دسته	۴−۸	۸−۱۲	۱۲−۱۶	۱۶−۲۰	۲۰−۲۴
فراوانی	۳	۴	۵	۲	۱

۱) ۱۲/۲	۲) ۱۲/۴	۳) ۱۲/۶	۴) ۱۲/۸

۱۵۳- در جدول فراوانی روبه‌رو میانگین کدام است؟

حدود دسته	۱۰−۱۴	۱۴−۱۸	۱۸−۲۲	۲۲−۲۶	۲۶−۳۰
فراوانی تجمعی	۸	۱۸	۲۷	۳۵	۴۰

۱) ۱۹/۱	۲) ۱۹/۲	۳) ۱۹/۳	۴) ۱۹/۴

پاسخ تشریحی آمار و مدل‌سازی

مدل‌سازی

ریاضیات یعنی دقیق‌بودن! برای دقیق‌بودن باید هر چیزی را برحسب اعداد و ارقام بیان کرد. برای همین یوهانس کپلر (ستاره‌شناس آلمانی) گفته: خداوند جهان را به زبان اعداد خلق کرده است. یعنی با دقت و حساب و کتاب!

مثلاً کلمه‌های «بعضی»، «اکثر»، «کم» و «زیاد»، خیلی دقت ندارند؛ ولی با استفاده از اعداد، تکلیف کاملاً روشن می‌شود. می‌گوییم خون طرف غلیظ است! خب این‌که نشد حرف علمی. اما اگر بگوییم هموگلوبین خونش ۱۷گرم بر دسی لیتر است، شنونده درک درست‌تری از غلظت خون آن شخص به دست می‌آورد.

پس قرار است دقیق باشیم و با دقت بررسی کنیم. برای دقت‌داشتن و کار در یک سیستم علمی (!) باید همه‌چیز با اعداد و ارقام و ریاضیات اندازه‌گیری شود. یعنی باید مسئله‌ها را به زبان ریاضی بیان کرد.

به این کار می‌گویند مدل‌سازی!

مدل‌سازی یعنی بیان مسئله به زبان ریاضی

۱- گزینه‌ی ۱

مدل خوب

کتاب درسی گفته بچه‌ها با اتومبیل بازی می‌کنند چون در ذهنشان با اتومبیل اسباب‌بازی، ادای رانندگی درمی‌آورند. یا با عروسک‌بازی، رفتاری شبیه مادر را تقلید می‌کنند. اهمیت زیبایی اسباب‌بازی‌ها این است که تا حد ممکن ساده باشند و بچه بتواند آن‌چه در ذهنش می‌گذرد بهتر نشان بدهد.

پس هر چه مفاهیم ریاضی که به کار می‌رود، ابتدایی و ساده‌تر باشد و نتیجه‌ی کار به پدیده‌ی موردنظر نزدیک‌تر باشد، مدل‌سازی با ارزش‌تر است.

۲- گزینه‌ی ۴

خطای اندازه‌گیری

قرار شد دقیق و ریاضی باشیم و همه‌چیز را به زبان ریاضی مدل کنیم و اندازه بگیریم. حالا با چی اندازه بگیریم؟ مثلاً طول یک خودکار را با خط‌کش اندازه می‌گیریم و جوابی برحسب سانتی‌متر یا میلی‌متر می‌دهیم. اما در مورد چیزهای دقیق‌تر، مثلاً قطعه‌های کوچک توی چرخ خیاطی، از کولیس استفاده می‌کنیم و جواب در حد دهم میلی‌متر است.

توی مسابقات اتومبیل‌رانی، هزارم ثانیه را هم اندازه می‌گیرند. اما هر چه‌قدر این وسیله‌ها دقیق‌تر باشند، هیچ‌وقت نمی‌توانند مقدار دقیق یک کمیت را اندازه بگیرند. پس هر اندازه‌گیری، یک اختلاف با واقعیت دارد. به این اختلاف می‌گوییم خطا .

در واقع خطا یعنی مقدار واقعی منهای مقدار اندازه‌گیری‌شده. خطا را با E نشان می‌دهیم. همیشه قدرمطلق خطا کم‌تر از واحد اندازه‌گیری است.

مثلاً می‌گوییم ۳ ساعت و خرده‌ای، یعنی $۳+E$ و قدرمطلق E از یک ساعت کم‌تر است. یا می‌گوییم وزن این کتاب دو کیلو و نیم است یعنی $۲/۵+E$ و قدرمطلق E از $۰/۵$ کیلو کم‌تر است. خطا می‌تواند مثبت یا منفی بشود ولی هیچ‌وقت صفر نمی‌شود.

۳- گزینه‌ی ۴

۴- گزینه‌ی ۳ در این اندازه‌گیری، واحد اندازه‌گیری $۰/۵$ کیلو بوده و خطا از $۰/۵$ کیلو کم‌تر است.

۵- گزینه‌ی ۴ در سؤال قبل گفتیم خطا از واحد اندازه‌گیری کم‌تر است. پس الان، از $L=۶+E$ (برحسب سانتی‌متر) می‌فهمیم $|E|$ از یک سانتی‌متر کم‌تر است. این یعنی دقت خط‌کش ما، در حد سانتی‌متر بوده.

حالا می‌خواهیم برحسب میلی‌متر بیانش کنیم. خب باید برویم یک خط‌کش جدید با دقت میلی‌متر پیدا کنیم و با آن دوباره اندازه بگیریم. جواب هم معلوم نیست.

فرض کنید طول خط‌کش تا سه رقم اعشار $۶/۲۱۷$ سانتی‌متر باشد. آن‌وقت برحسب سانتی‌متر می‌شود $L=۶+E$ و برحسب میلی‌متر می‌شود $L=۶۲+E'$ که هیچ‌کدام از گزینه‌ها نیست.

محاسبات خطایی در مدل‌سازی

اگر یک پدیده، به چندتا متغیر بستگی داشته باشد، می‌توانیم آن متغیرها را اندازه بگیریم و سپس خودمان جواب پدیده‌ی نهایی را حساب کنیم.

مثلاً ضلع یک مربع را اندازه بگیریم و سپس محیط و مساحتش را حساب کنیم و مدل آن‌ها را بنویسیم.

$$\text{ضلع}\ a=۴+E \Rightarrow \begin{cases} \text{محیط} = ۴a = ۴(۴+E) = ۱۶+۴E & (\text{خطای محیط} = ۴E) \\ \text{مساحت} = a^۲ = (۴+E)^۲ = ۴^۲+۸E+E^۲ = ۱۶+۸E+E^۲ & (\text{خطای مساحت} = ۸E) \end{cases}$$

البته چون قدرمطلق E از یک کم‌تر است، می‌توانیم از جمله‌های شامل $E^۲$ یا $E^۳$ یا ... صرف‌نظر کنیم. پس مدل مساحت می‌شود: $۱۶ + ۸E$

کتاب درسی گفته خطای مساحت (۸E)، هشت برابر بزرگ‌تر از خطای ضلع (E) است. خب درست گفته! یک اصطلاح دیگر هم توی کتاب هست: خطای اندازه‌گیری در مساحت و محیط مربع منتشر می‌شود.

یعنی ما توی ضلع مربع به اندازه‌ی E خطا داشتیم، اما توی محیط به اندازه‌ی (۴E) و در مساحت به اندازه‌ی (۸E) خطا افزایش می‌یابد.

۶- گزینه‌ی ۳ مساحت دایره می‌شود $S = \pi R^۲$. پس داریم:

$$S = \pi R^۲ = \pi(۳ + E)^۲ \xrightarrow{\text{اتحاد اول}} = \pi(۹ + ۶E + E^۲)$$

$$= ۹\pi + ۶\pi E + \pi E^۲$$

راستی $E^۲$ را بی‌خیال می‌شویم. پس جواب می‌شود $S \approx ۹\pi + ۶\pi E$

۷- گزینه‌ی ۲ خب برویم محیط و مساحت مربع را حساب کنیم:

$$\text{محیط} = ۴a = ۴(۵ + E) = ۲۰ + ۴E$$

$$\text{مساحت} = a^۲ = (۵ + E)^۲ = ۲۵ + ۱۰E + E^۲ \approx ۲۵ + ۱۰E$$

$E^۲$ را بی‌خیال می‌شویم.

حالا نگاه کنید ... خطای اندازه‌گیری ضلع مربع، E بوده و در محیط به اندازه‌ی ۴E و در مساحت به اندازه‌ی ۱۰E اثر گذاشت. پس نسبت این تأثیر در مساحت به محیط می‌شود $\frac{۱۰E}{۴E}$ یعنی $\frac{۱۰}{۴}$ یا ۲/۵ .

کتاب درسی پرسیده: خطای ناشی از اندازه‌گیری اضلاع، بیشتر در محیط مربع اثر می‌گذارد یا در مساحت آن؟ خب توی این سؤال که در مساحت بیشتر اثر گذاشت. حالا برویم سؤال بعدی ....

۸- گزینه‌ی ۲ خب فرض کنیم ضلع مثلث $a + E$ باشد. محیط و مساحت را نگاه کنید.

محیط مثلث متساوی‌الاضلاع به ضلع x، می‌شود ۳x. مساحتش هم $\frac{\sqrt{۳}}{۴}x^۲$ است.

$$\text{محیط} = ۳x = ۳(a + E) = ۳a + ۳E$$

$$\text{مساحت} = \frac{\sqrt{۳}}{۴}x^۲ = \frac{\sqrt{۳}}{۴}(a + E)^۲ = \frac{\sqrt{۳}}{۴}(a^۲ + ۲aE + E^۲)$$

$$\approx \frac{\sqrt{۳}}{۴}a^۲ + \frac{\sqrt{۳}}{۲}aE$$

آهان ... خطای ناشی از ضلع (یعنی E)، در محیط به ۳E تبدیل شد. اما در مساحت به $\frac{\sqrt{۳}}{۲}aE$ رسیدیم. سؤال گفته در محیط بیشتر اثر گذاشته است.

پس داریم: $۳E > \frac{\sqrt{۳}}{۲}aE$

بنابراین با حذف E ها می‌توان گفت: $\Rightarrow ۳ > \frac{\sqrt{۳}}{۲}a$

و این یعنی: $\Rightarrow a < \frac{۳}{\frac{\sqrt{۳}}{۲}}$

اما توی گزینه‌ها $\frac{۳}{\frac{\sqrt{۳}}{۲}}$ نداریم. خب نترسید. باید ساده‌تر کنیم:

$$\frac{۳}{\frac{\sqrt{۳}}{۲}} = \frac{۳ \times ۲}{\sqrt{۳}} = \frac{۶}{\sqrt{۳}} = \frac{۶\sqrt{۳}}{۳} = ۲\sqrt{۳}$$

پس در واقع $a < ۲\sqrt{۳}$ جواب است.

نگران علامت E نباشید. منظور از تأثیر خطا در مساحت و محیط، قدرمطلق این اعداد است و مشکلی برای نامساوی‌ها به وجود نمی‌آید.

۹- گزینه‌ی ۱ این چیزی که صورت سؤال گفته یعنی $d = ۱۰ + E$ و $E < \frac{۱}{۶\pi}$. حال مساحت را می‌خواهد. اگر یادتان باشد (توی فیزیک هم دیدید) مساحت دایره $\pi\frac{d^۲}{۴}$ است.

پس داریم: $S = \frac{\pi}{۴}d^۲ = \frac{\pi}{۴}(۱۰ + E)^۲ = \frac{\pi}{۴}(۱۰۰ + ۲۰E + E^۲)$

قرار بود از $E^۲$ هم صرف‌نظر کنیم. پس می‌شود:

$$S = \frac{\pi}{۴}(۱۰۰ + ۲۰E) = ۲۵\pi + \underbrace{۵\pi E}_{E_۱}$$

حالا سؤال گفته بود E از $\frac{۱}{۶\pi}$ کم‌تر است:

$$E < \frac{۱}{۶\pi} \xrightarrow{\times ۵\pi} ۵\pi E < \frac{۵\pi}{۶\pi} \Rightarrow E_۱ < \frac{۵}{۶}$$

پس خطای مساحت تقریبا از $\frac{۵}{۶}$ کم‌تر است.

کلمه‌ی «تقریباً» را برای این گذاشته است که از $E^۲$ صرف‌نظر کرده‌ایم.

۱۰- گزینه‌ی ۱ اگر مدل شعاع به صورت $R = ۲ + E$ باشد، مدل حجم را می‌توانیم پیدا کنیم:

$$V = \frac{۴}{۳}\pi R^۳ \Rightarrow V = \frac{۴}{۳}\pi(۲ + E)^۳ = \frac{۴}{۳}\pi(۸ + ۶E^۲ + ۱۲E + E^۳)$$

از توان‌های E هم چشم می‌پوشیم:

$$V \approx \frac{۴}{۳}\pi(۸ + ۱۲E) = \frac{۴}{۳}\pi(۸) + \underbrace{\frac{۴}{۳}\pi(۱۲E)}_{E'}$$

پس خطای حجم می‌شود: $E' = \frac{۴}{۳}\pi(۱۲E) = ۱۶\pi E$

ما می‌خواهیم $۱۶\pi E < ۱$ باشد، پس $E < \frac{۱}{۱۶\pi}$.

۱۱- گزینه‌ی ۳ خب به جای x، مدلش را بگذاریم:

$$y = x^۳ - x = (۲ + E)^۳ - (۲ + E)$$

$$= ۸ + ۱۲E + ۶E^۲ + E^۳ - ۲ - E$$

$$= ۶ + ۱۱E + ۶E^۲ + E^۳$$

$$\approx ۶ + ۱۱E$$

با عرض شرمندگی فراوان، اتحاد $(a + b)^۳$ را یادآور شدیم:

$$(a + b)^۳ = a^۳ + ۳a^۲b + ۳ab^۲ + b^۳$$

مدل‌سازی‌ها با چند خطا

اگر پدیده‌ای که بررسی می‌کنیم تابعی از چندتا متغیر باشد، ممکن است هر کدام از آن متغیرها خطای اندازه‌گیری خودشان را داشته باشند. یعنی $E_۱$، $E_۲$ و . . . داریم.

مسئله فرق زیادی با قبل ندارد، فقط این‌جا از حاصل‌ضرب E_i ها هم صرف‌نظر می‌شود.

۱۲- گزینه‌ی ۴ خب مساحت می‌شود طول ضرب در عرض:

$$S = ab = (۵ + E_۱)(۳ + E_۲) = ۱۵ + ۳E_۱ + ۵E_۲ + E_۱E_۲$$

قرار شد ضرب E ها را بی‌خیال شویم: $S \approx ۱۵ + ۳E_۱ + ۵E_۲$

۱۳- گزینه‌ی ۴

حجم استوانه می‌شود مساحت قاعده ضرب در ارتفاع، یعنی $\pi R^{۲}h$.

برویم حجم این استوانه را حساب کنیم:

$$V = \pi R^{۲}h = \pi(۲ + E_{۱})^{۲}(۵ + E_{۲})$$
$$= \pi(۴ + ۴E_{۱} + E_{۱}^{۲})(۵ + E_{۲})$$

قرار بود از $E^{۲}$ ها صرف‌نظر کنیم:

$$\approx \pi(۴ + ۴E_{۱})(۵ + E_{۲})$$
$$\approx \pi(۲۰ + ۲۰E_{۱} + ۴E_{۲} + ۴E_{۱}E_{۲})$$

از ضرب E ها هم خوشمان نمی‌آید، پس مدل V می‌شود:

$$V \approx \pi(۲۰ + ۲۰E_{۱} + ۴E_{۲}) = ۲۰\pi + ۲۰\pi E_{۱} + ۴\pi E_{۲}$$

یعنی $E_{۱}$ و $E_{۲}$ به ترتیب ۲۰π و ۴π برابر می‌شوند.

۱۴- گزینه‌ی ۳ برویم حجم این مکعب‌مستطیل را پیدا کنیم:

حجم این مکعب مستطیل، $V = L_{۱}L_{۲}L_{۳}$ یعنی طول × عرض × ارتفاع است.

$$V = L_{۱}L_{۲}L_{۳} = (۳ + E_{۱})(۴ + E_{۲})(۶ + E_{۳})$$
$$= (۱۲ + ۴E_{۱} + ۳E_{۲} + \cancel{E_{۱}E_{۲}})(۶ + E_{۳})$$
$$\approx ۷۲ + ۲۴E_{۱} + ۱۸E_{۲} + ۱۲E_{۳} + \cancel{۴E_{۱}E_{۳}} + \cancel{۳E_{۲}E_{۳}}$$

جملاتی که خط زدیم، شامل ضرب E_{i} ها بود. خب قرار بود از ضرب E ها صرف‌نظر کنیم.

پس در مدل حجم این مکعب، $۲۴E_{۲}$ ندیدیم.

تعاریفی از جامعه و نمونه

جامعه‌ی آماری یعنی تمام افراد یا اشیایی که می‌خواهیم بررسی خاصی در مورد آن‌ها انجام دهیم.

دقت کنید که می‌خواهیم در مورد آن‌ها بررسی کنیم. اگر این بررسی را روی همه‌ی افراد جامعه انجام دهیم می‌گوییم سرشماری کردیم. البته معمولاً زورمان نمی‌رسد همه را بررسی کنیم.

چرا؟ چون گران تمام می‌شود، وقتمان را خیلی می‌گیرد، افراد در دسترس نیستند و جامعه ممکن است از بین برد. مثلاً می‌خواهیم ضربان قلب را در حین سقوط هواپیما بررسی کنیم. سرشماری یعنی همه‌ی مردم جامعه را سوار هواپیما کنیم، بعد سقوط کنیم و به همه‌ی افراد دستگاه ECG (نوار قلب) وصل کنیم!!!

پس سرشماری خیلی عملی نیست. چه‌کار کنیم؟ خب به قول کتاب درسی‌تان، مشت نمونه‌ی خروار است. یعنی به جای همه‌ی جامعه، یک نمونه از آن برمی‌داریم و بررسی می‌کنیم. نمونه در واقع یک زیرمجموعه از جامعه است. راستی به جای تعداد اعضا می‌گوییم اندازه. پس تعداد اعضای جامعه را اندازه‌ی جامعه و تعداد اعضای نمونه را اندازه‌ی نمونه می‌نامیم.

۱۵- گزینه‌ی ۲

۱۶- گزینه‌ی ۳ در سرشماری، همه را بررسی می‌کنیم. اما در نمونه‌گیری، فقط بعضی‌ها را بررسی می‌کنیم.

حالا اگر اندازه‌ی نمونه (یعنی اندازه‌ی زیرمجموعه) برابر اندازه‌ی جامعه بشود، یعنی سرشماری کرده‌ایم.

۱۷- گزینه‌ی ۲ کتاب درسی گفته «عمل نمونه‌گیری مهم‌ترین بخش آمار را تشکیل می‌دهد».

حالا چرا؟ چون نمونه باید خصوصیات جامعه را نمایان کند. برای این‌که گروه کوچکی از اعضا بتوانند خصوصیات عده‌ی بزرگ‌تری را بیان کنند باید به نحو مناسبی انتخاب بشوند. مثلاً نمونه باید به اندازه‌ی کافی بزرگ باشد و انتخاب اعضایش هم از قاعده و قانون خاصی پیروی نکند.

نمونه‌ی تصادفی ساده

یادتان هست که ما می‌خواستیم در مورد جامعه یک چیزی را بررسی کنیم؟!

بعد فهمیدیم زورمان به جامعه نمی‌رسد (پول و وقت و دسترسی‌اش را نداریم)

پس رفتیم سراغ قسمتی از جامعه، یعنی یک نمونه از بین اعضا انتخاب کردیم.

بعد گفتیم این نمونه گرفتن، کار مهمی است و باید انتخاب نمونه، قاعده‌ی خاصی (پارتی‌بازی) نداشته باشد.

کتابتان گفته باید جوری نمونه‌گیری بشود که:

اولاً امکان انتخاب هر عضو جامعه به عنوان عضوی از نمونه باشد.

ثانیاً هر فردی برای شرکت در نمونه، همان‌قدر سهم داشته باشد که بقیه دارند.

به این نمونه‌ی خوب می‌گوییم نمونه‌ی تصادفی ساده؛ یعنی باید شانس تمام افراد جامعه برای انتخاب در نمونه‌ی تصادفی ساده، با هم برابر باشند.

۱۸- گزینه‌ی ۴

اعداد تصادفی

آدم‌ها، به خاطر اعتقادات (مثلاً نحس‌بودن سیزده) یا سلیقه و سابقه‌ی ذهنی، نمی‌توانند نمونه‌ی تصادفی ساده انتخاب کنند. ما هم که دقیق هستیم و می‌خواهیم همه‌ی عضوهای جامعه، شانس مساوی انتخاب‌شدن داشته باشند، پس چه‌کار می‌کنیم؟ به آدمیزاد دو پا نمی‌شود اعتماد کرد، می‌رویم سراغ کامپیوتر. هر چی باشه پارتی‌بازی نمی‌کنه.....

ماشین‌حساب‌های مهندسی و کامپیوترها (زبان‌های برنامه‌نویسی)، یک عملگر به اسم RAN یا RND دارند. (همون random خودمان) که عددی تصادفی بین ۰ و ۱ تولید می‌کند.

این عدد را ضرب در اندازه‌ی جامعه می‌کنیم، بعد اعشارش را حذف کرده و یک واحد اضافه می‌کنیم.

مثلاً کامپیوتر ما ۰/۴۳۱ را می‌دهد. در جامعه‌ای با اندازه‌ی ۵۰۰ نفر داریم: $۰/۴۳۱\times۵۰۰=۲۱۵/۵ \xrightarrow{\text{اعشار حذف}} ۲۱۵ \xrightarrow{+۱} ۲۱۶$

یعنی به عنوان عضو نمونه، باید نفر ۲۱۶ اُم را برداریم.

حالا فرض کنید از ۱ تا ۲۰ ، نمونه‌ای ۳ عضوی می‌خواهیم. سه بار RND را می‌زنیم.

اعداد ۰/۱۰۸، ۰/۵۹۳ و ۰/۸۷۷ را می‌دهد. بنابراین داریم:

$$[۰/۱۰۸\times۲۰]+۱=[۲/۱۶]+۱=۳$$

$$[۰/۵۹۳\times۲۰]+۱=[۱۱/۸۶]+۱=۱۲$$

$$[۰/۸۷۷\times۲۰]+۱=[۱۷/۵۴]+۱=۱۸$$

یعنی باید ۳، ۱۲ و ۱۸ (عضوهای شماره‌ی ۳، ۱۲ و ۱۸) را به عنوان نمونه‌ی ۳ تایی از ۱ تا ۲۰ برداریم.

۱۹- گزینه‌ی ۳ قرار شد عدد رندم را در اندازه ضرب کنیم:

$۲۰۰\times۰/۳۷۴=۷۴/۸$ بعد اعشارش را برداریم: ۷۴ و یکی اضافه کنیم: ۷۵

۲۰- گزینه‌ی ۳

وقتی نمونه‌ی تصادفی از ۱ تا n می‌خواهیم، عدد تصادفی را در n ضرب می‌کنیم، بعد اعشارش را می‌اندازیم و یکی اضافه می‌کنیم. حالا اگر نمونه‌ی تصادفی از a تا b می‌خواهیم، باید عدد تصادفی را ضرب در اندازه‌ی جامعه (یعنی تعداد اعداد از a تا b)کنیم، بعد اعشارش را بیندازیم و a تا اضافه کنیم.

پس اگر فرمول دوست دارید:

عضو انتخابی = [عدد تصادفی × اندازه‌ی جامعه] + شماره‌ی عضو اول

مثلاً در این تست: $=[۰/۶۱۴\times۶۵]+۲۳=[۳۹/۹]+۲۳=۶۲$

از ۲۳ تا ۸۷، تعداد اعداد ۶۵ تا است. $(۸۷-۲۳+۱)$. حتماً می‌دانید...

جمع‌آوری داده‌ها

عرض شود که ما می‌خواستیم در مورد جامعه، یک چیزی را بررسی کنیم. سرشماری (یعنی بررسی همه‌ی افراد) سخت بود.

ما هم رفتیم دنبال نمونه. یک نمونه‌ی تصادفی ساده گرفتیم و می‌خواهیم روی آن بررسی کنیم.

مثلاً می‌خواهیم قد اعضای این نمونه را اندازه بگیریم. اعدادی که به دست می‌آوریم را داده می‌نامند. پس داده یعنی نتایج حاصل از بررسی نمونه‌ی آماری.

حالا این داده‌ها را چه جوری به دست می‌آوریم؟ مثلاً قد را اندازه می‌گیریم؛ رنگ چشم را می‌بینیم؛ معدل را می‌پرسیم؛ در مورد ویژگی‌های اخلاقی، یک فرم پرسش‌نامه می‌دهیم که پر کنند؛ در مورد بعضی چیزها هم از داده‌هایی استفاده می‌کنیم که قبلاً تهیه شده‌اند. مثلاً در مورد زلزله‌هایی که رخ داده‌اند یا چیزهای مربوط به گذشته ...) و پس روش‌های جمع‌آوری داده‌ها عبارت‌اند از:

(۱) پرسش (کتبی با پرسش‌نامه یا شفاهی)

(۲) مشاهده و ثبت وقایع

(۳) آزمایش و اندازه‌گیری

(۴) استفاده از داده‌های موجود (از پیش تهیه‌شده)

۲۱- گزینه‌ی ۲ با توجه به حرف‌هایی که زدم، الگوی خاص مناسب نیست.

طراحی پرسش‌نامه

طرح پرسش‌نامه خیلی مهم است. باید ببینیم چه چیزی را می‌خواهیم بدانیم؟ از چه مطالبی باید بپرسیم؟

دقت کنیم که اطلاعات غیرضروری نگیریم؛ سؤالات ساده و واضح بپرسیم که چند برداشت مختلف نشود؛ جواب‌ها تا حد ممکن یک‌کلمه‌ای باشند؛ در مورد سن (خانم‌ها) و یا درآمد (آقایان) به جای اعداد دقیق، بازه‌های مشخص بدهیم؛ سؤالات یا کلمات هدایت‌کننده ندهیم؛ پاسخ‌های کیفی را دسته‌بندی کنیم و ...

در ضمن باید با مردم مؤدبانه برخورد کرد! یعنی روش پاسخ‌گویی به پرسش‌نامه را هم به آن ضمیمه کنیم و آخرش هم تشکر کنیم.

یک جایی یک پرسش‌نامه به رتبه‌های برتر کنکور داده بودند. این پرسش‌نامه بدترین پرسش‌نامه‌ی آماری در تاریخ ریاضیات بود. سؤال اول پرسیده بودند کدام کتاب‌های ما را استفاده کردید؟ (کتاب‌های خودشان) و ۸ خط جا گذاشته بودند، بعد پرسیده بودند کدام کتاب‌های ناشران دیگر را استفاده کردید؟ و نصف خط جا گذاشته بودند. خداوند همه را به راه راست هدایت کناد.

از عباراتی مثل «بزرگ» و «کوچک» که معیار مشخصی ندارند استفاده نمی‌کنیم.

به پاسخ‌دهنده، نظر خودمان را تحمیل نمی‌کنیم. مثلاً نمی‌پرسیم «نظرتان در مورد برنامه‌ی کامل‌شده چیست؟» چون عبارات «کامل‌شده» نظر مثبت القا می‌کند.

۲۲- گزینه‌ی ۴

متغیر تصادفی

توی آمار، ما یک چیزی را بررسی می‌کنیم. چرا بررسی می‌کنیم؟ چون از یک آدم به یک آدم دیگر با هم فرق دارد. مثلاً هیچ‌وقت در یک بررسی آماری، تعداد چشم‌های افراد سالم را بررسی نمی‌کنند. چون مال همه‌ی افراد یک جور است. بررسی‌های آماری، به ویژگی‌هایی اختصاص دارد که متغیر هستند. به این ویژگی‌ها (یعنی به موضوع مورد مطالعه) می‌گوییم متغیر تصادفی.

دو جور متغیر تصادفی داریم.

الف متغیرهای کمی: مقدار این متغیرها با عدد معلوم می‌شود. یا اندازه‌گیری می‌کنیم یا می‌شماریم.

مثلاً قد، وزن، هزینه، دما، میزان بارندگی، میزان آلودگی هوا، شدت زلزله و ... کمی‌اند.

ب متغیرهای کیفی: این متغیرها با یک کلمه یا نوع معلوم می‌شوند. یعنی فقط اسم دارند. اندازه‌گیری یا شمارش در کار نیست. مثلاً نوع گروه خون، نوع RH، جنسیت، نوع شغل، نوع کشت (آبی باشد یا دیم باشد)، سطح تحصیلات و ... کیفی‌اند.

خود متغیرهای کمی دو جور هستند:

① متغیر پیوسته یعنی متغیری که اگر دو مقدار a و b را بگیرد، هر مقدار بین آن‌ها را هم بتواند بگیرد.

مثلاً قد، وزن، زمان، هزینه، میزان بارندگی، اکثر پدیده‌های فیزیکی مثل سرعت، شتاب، جریان، ولتاژ، مقاومت و اکثر مقادیر هندسی مثل اندازه‌ی زوایا، مساحت، محیط، حجم و ... متغیرهای پیوسته‌اند.

② متغیر گسسته یعنی متغیری که پیوسته نیست. یعنی هر مقدار بین دو مقدارش را نمی‌گیرد.

معمولاً این متغیرها از نوع شمارشی و تعدادی هستند. مثلاً تعداد طبقات ساختمان (دو طبقه داریم، سه طبقه هم داریم اما ۲/۴ طبقه نداریم). تعداد مکالمات، تعداد بیماران، تعداد روزهای بارانی، تعداد تصادفات و ... گسسته‌اند.

کیفی‌ها هم دوجورند.

③ اکثر متغیرهای کیفی، اسمی هستند. یعنی فقط نوع دارند. مثلاً گروه خون، رنگ، وضع تأهل، وضع مسکن، وضع سواد (باسواد یا بی‌سواد)، جنسیت، نوع تلفن (ثابت یا همراه)، نوع بیماری و ...

④ گروه کوچکی از متغیرهای کیفی، یک نوع ترتیب طبیعی دارند. به آن‌ها کیفی ترتیبی می‌گوییم.

مثلاً مراحل تحصیل (دبستان، راهنمایی، دبیرستان، دانشگاه)، مدارک تحصیلی، مراحل رشد، مراحل زندگی، روزهای هفته، فصل‌ها و ماه‌های سال و ... ترتیبی‌اند.

این هم هست:

- داده
  - کمی (عددی)
    - اندازه‌گیری ← پیوسته
    - شمارش ← گسسته
  - کیفی (حالت و نوع)
    - فقط اسم دارد ← اسمی
    - نوعی ترتیب طبیعی دارد ← ترتیبی

۲۳- گزینه‌ی ۱

گروه خون، کیفی اسمی است! این هم سراسری ۹۰.

۲۴- گزینه‌ی ۳

مثلاً گاهی تخم‌مرغ را براساس تعداد می‌خریم. بعضی چیزها را براساس وزن می‌خریم.

خب تعداد، متغیر گسسته است.

۲۵- گزینه‌ی ۳

شدت زلزله، میزان آلودگی هوا و هزینه‌ی بیمه پیوسته‌اند. چون اندازه‌گیری می‌شوند. اما درصد ادبیات کنکور را با شمارش تعداد درست‌ها و غلط‌ها به دست می‌آوریم.

اگر همه‌ی تست‌ها را درست بزنید، درصد ادبیات می‌شود ۱۰۰. حالا اگر یک نزده داشته باشید، درصدتان می‌شود ۹۶. بین ۹۶ تا ۱۰۰، هیچ مقداری وجود ندارد. یعنی هیچ‌کس نمی‌تواند در ادبیات (که ۲۵ سؤال دارد) ۹۷ درصد یا ۹۸ درصد بزند! پس گسسته است.

تصور غلطی وجود دارد که هر داده‌ای که اعشار داشته باشد، پیوسته است. این‌طور نیست!

مثلاً تعداد طبقات ساختمان می‌تواند ۲/۵ هم باشد ولی داده‌ی گسسته است. هم‌چنین درصد کنکور یا نمره‌های امتحانی شما، با این‌که اعشاری هم هست، ولی داده‌ی گسسته است.

۲۶- گزینه‌ی ۲ ماشین حساب، در جواب RAN، عدد تصادفی (a) بین صفر و ۱ می‌دهد که پیوسته است. ما این عدد را در اندازه‌ی جامعه ضرب می‌کردیم و اعشارش را حذف می‌کردیم. پس حاصل b گسسته است.

دامنه‌ی تغییرات

دامنه‌ی تغییرات یعنی اختلاف بیشترین و کم‌ترین داده.

آن‌ها را با R نشان می‌دهند. پس: $R = \max - \min$

وقتی دامنه‌ی تغییرات کوچک است احتیاجی به دسته‌بندی نیست. چون اعضا به هم نزدیک‌اند و اختلاف زیاد ندارند؛ می‌گوییم جامعه یک‌دست است اما هر چه‌قدر دامنه‌ی تغییرات بزرگ‌تر باشد، نمی‌توانیم جامعه را یک‌دست فرض کنیم و مجبوریم دسته‌بندی کنیم.

پس دامنه‌ی تغییرات یعنی طول بازه‌ای که متغیر در آن تغییر می‌کند. اگر دامنه‌ی تغییرات صفر بشود تمام داده‌ها با هم مساوی‌اند.

بعضی وقت‌ها، به همان دلایل که در مورد سرشماری گفتیم، ممکن است نتوانیم دامنه‌ی تغییرات جامعه را حساب کنیم. بلکه به جایش دامنه‌ی تغییرات نمونه را حساب می‌کنیم.

کتاب درسی از شما پرسیده دامنه‌ی تغییرات جامعه بزرگ‌تر است یا نمونه؟ خب دامنه‌ی تغییرات جامعه بیشتر یا مساوی دامنه‌ی تغییرات نمونه است. (اگر گفتی چرا؟)

۲۷- گزینه‌ی ۱

۲۸- گزینه‌ی ۲

الف درسته. چون دامنه‌ی تغییرات جامعه، بیشتر یا مساوی نمونه است. مثلاً اگر از کلاس شما ۴ نفر را انتخاب کنیم، اختلاف قد بلندترین و کوتاه‌ترین آن‌ها ۲۴ سانتی‌متر است. در کل کلاس، اختلاف بلندترین و کوتاه‌ترین نفر، نمی‌تواند کم‌تر از ۲۴ سانتی‌متر باشد. (چون همین ۴ نفر هم توی کلاس‌اند)

ب هم درسته. وقتی بین ۴ نفر از کلاس شما، ۰/۵ متر اختلاف قد هست، خب در کل کلاس هم حداقل نیم متر اختلاف هست.

اما **ج** غلطه، وقتی بین چند نفر انتخابی، دامنه‌ی تغییرات قدها کم است نمی‌شود گفت در بین کل کلاس هم کم است. شاید آدم‌های بلندتر و کوتاه‌تری در افراد کلاس باشند که توی نمونه نیستند.

۲۹- گزینه‌ی ۴ دامنه‌ی تغییرات، اختلاف بیشترین و کم‌ترین بود.

الان بیشترین ۱۲۱ و کم‌ترین ۹۹ است.

پس داریم: $R = ۱۲۱ - ۹۹ = ۲۲$

۳۰- گزینه‌ی ۴ وقتی به هر داده سه واحد اضافه می‌کنیم، همه ۳ واحد زیاد می‌شوند.

پس بیشترین و کم‌ترین داده هم ۳ تا زیاد می‌شوند. مثل این‌که همه بروند روی یک سکوی ۳ سانتی! اختلاف بیشترین قد و کم‌ترین قد که فرقی نمی‌کند. ببینید:

$$R = \max - \min \xrightarrow{\text{به هر کس ۳ تا اضافه کنیم}}$$

$$R' = (\max + \cancel{۳}) - (\min + \cancel{۳}) = \max - \min = R$$

پس اضافه‌کردن عدد ثابتی به همه‌ی داده‌ها (یا کم‌کردن عدد ثابتی از همه‌ی داده‌ها) اثری روی دامنه‌ی تغییرات نمی‌گذارد. یعنی R تغییر نمی‌کند!

دسته‌بندی داده‌ها

بعد از پیداکردن دامنه‌ی تغییرات، باید تعداد دسته‌ها را معلوم کنیم. البته صورت مسئله به ما تعداد دسته‌ها را می‌دهد. اگر دامنه‌ی تغییرات R و تعداد دسته‌ها k باشد، طول هر دسته $C = \frac{R}{k}$ است، پس داریم: $\text{طول دسته} = \frac{\text{دامنه‌ی تغییرات}}{\text{تعداد دسته}}$

مثلا اگر دامنه‌ی تغییرات ۲۴ باشد و ما ۶ دسته بخواهیم، طول دسته‌ها $\frac{۲۴}{۶} = ۴$ می‌شود.

در همین داده‌ها، اگر ۸ دسته می‌خواستیم، طول هر دسته $\frac{۲۴}{۸} = ۳$ می‌شد.

۳۱- گزینه‌ی ۴ گفتیم که $C = \frac{R}{k}$، یعنی:

$$\text{طول دسته} = \frac{\text{دامنه‌ی تغییرات}}{\text{تعداد دسته}}$$

حالا در این مسئله، $C = ۵$ و $k = ۱۰$ است. پس داریم: $۵ = \frac{R}{۱۰}$

بنابراین دامنه‌ی تغییرات برابر است با: $\Rightarrow R = ۵۰$

۳۲- گزینه‌ی ۲ شش دسته با طول ۵ داریم.

پس $C = ۵$ و $k = ۶$، می‌توانیم R را پیدا کنیم:

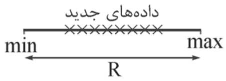
$$C = \frac{R}{k} \Rightarrow ۵ = \frac{R}{۶} \Rightarrow R = ۳۰$$

یعنی دامنه‌ی تغییرات ۳۰ است. سؤال گفته بیشترین داده ۵۰ بوده است؛ پس داریم: $R = \max - \min \xrightarrow[\max=۵۰]{R=۳۰} ۳۰ = ۵۰ - \min \Rightarrow \min = ۲۰$

بنابراین کم‌ترین داده ۲۰ بوده.

۳۳- گزینه‌ی ۳ ۹ طبقه به طول دسته‌ی ۴ یعنی دامنه‌ی تغییرات $R = kC = ۹ \times ۴ = ۳۶$ بوده است.

حالا گفته ۸ داده بین داده‌ی ماکسیمم و مینیمم اضافه می‌کنیم. پس دامنه‌ی تغییرات فرقی نمی‌کند (چون جدیدها بین ماکسیمم و مینیمم هستند، تغییری نداریم):



اما یک واحد از طول دسته کم می‌کنیم. یعنی: $C' = C - ۱ = ۴ - ۱ = ۳$

حالا چندتا دسته داریم؟ خب ۱۲تا، ببین: $k' = \frac{R}{C'} = \frac{۳۶}{۳} = ۱۲$

دسته‌ها

الان که در خدمت شماییم، دامنه‌ی تغییرات را حساب کردیم، تعداد دسته‌ها را می‌دانیم، طول دسته‌ها را هم معلوم کرده‌ایم. حالا وقت دسته‌بندی است. یعنی باید دسته‌ها را پیدا کنیم. مثلاً وقتی ماکسیمم و مینیمم داده‌ها ۲۰ و ۳۰ هستند (دامنه‌ی تغییرات می‌شود ۱۰) برای ۴ تا دسته این‌جوری عمل می‌کنیم؛ طول هر دسته $C = \frac{R}{k} = \frac{۱۰}{۴} = ۲/۵$ است. یعنی باید فاصله‌ی ۲۰ تا ۳۰ را به ۴ قسمت با طول ۲/۵ تقسیم‌بندی کنیم:

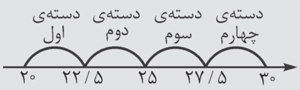
$[۲۰,۲۲/۵),[۲۲/۵,۲۵),[۲۵,۲۷/۵),[۲۷/۵,۳۰]$

دقت کنید که دسته‌ها به صورت $[a,b)$ هستند. یعنی عدد انتهایی هر دسته، جزء دسته‌ی بعدی است. البته در مورد دسته‌ی آخر، چون دسته‌ی بعدی نداریم مجبوریم بازه‌ی بسته بگیریم.

به عدد شروع دسته‌ها می‌گوییم حد پایین یا کران پایین دسته. به عدد آخر هر دسته هم می‌گوییم حد بالا یا کران بالا.

در مثال بالا، دسته‌ی اول $[۲۰,۲۲/۵)$ بود، پس حد بالای دسته‌ی اول می‌شود ۲۲/۵.

من این‌جوری دسته‌ها را نشان می‌دهم:

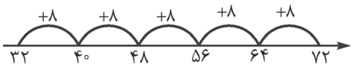


در واقع یک تصاعد حسابی روی محور ساخته می‌شود. از داده‌ی min با قدرنسبت C به طرف داده‌ی max می‌رویم.

۳۴- گزینه‌ی ۲ دامنه‌ی تغییرات می‌شود: $R = ۷۲ - ۳۲ = ۴۰$

حالا ۵ تا دسته داریم، پس طول هر دسته برابر است با: $C = \frac{R}{k} = \frac{۴۰}{۵} = ۸$

برویم سراغ دسته‌بندی: باید از ۳۲ شروع کنیم و ۸ تا ۸ تا جلو برویم:



پس دسته‌ی دوم، $[۴۰,۴۸)$ است و کران پایینش می‌شود ۴۰.

۳۵- گزینه‌ی ۳ دسته‌ی آخر ۸۶-۹۲ است. پس طول دسته‌ها $C = ۶$ بوده (از ۸۶ تا ۹۲، شش‌تا فاصله دارد) سؤال گفته ۸ تا دسته داریم.

پس می‌توانیم دامنه‌ی تغییرات را پیدا کنیم:

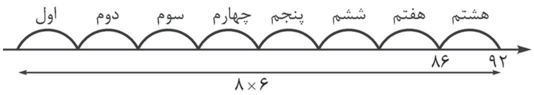
$$C = \frac{R}{k} \xrightarrow[k=۸]{C=۶} ۶ = \frac{R}{۸} \Rightarrow R = ۴۸$$

از طرفی دامنه‌ی تغییرات، اختلاف بیشترین و کم‌ترین بود. الان بیشترین داده (حد بالای دسته‌ی آخر) ۹۲ است.

پس داریم: $R = \max - \min \Rightarrow ۴۸ = ۹۲ - \min \Rightarrow \min = ۴۴$

یعنی داده‌ی مینیمم ۴۴ بوده.

محور را می‌بینیم:



ما باید از دسته‌ی هشتم، به دسته‌ی اول بیاییم. طول هر دسته هم $۹۲ - ۸۶ = ۶$ است. پس باید از ۹۲ به اندازه‌ی ۸×۶ عقب برویم:

$$\min = ۹۲ - \underbrace{۴۸}_{۶\times۸} = ۴۴$$

۳۶- گزینه‌ی ۲ این حدود دسته چی میگن؟ دوتا چیز:

اولاً وقتی دسته‌ی اول $[۴۵,۵۱)$ است، داده‌ی مینیمم ۴۵ بوده.

ثانیاً طول دسته‌ها $C = ۵۱ - ۴۵ = ۵۷ - ۵۱ = ۶$ بوده.

طول دسته‌ها همان اختلاف حد بالا و پایین دسته، اختلاف حدهای بالای دو دسته‌ی متوالی یا اختلاف حدهای پایین آن‌ها است.

دامنه‌ی تغییرات را هم داریم: $R = ۳۰$

پس تعداد دسته‌ها را می‌شود پیدا کرد:

$$C = \frac{R}{k} \xrightarrow[C=۶]{R=۳۰} ۶ = \frac{۳۰}{k} \Rightarrow k = ۵$$

یعنی ۵ تا دسته داریم.

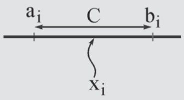
مرکز دسته

هدف از دسته‌بندی داده‌ها صرف‌نظر از اختلاف‌های جزئی است. یعنی داده‌های توی یک دسته، فرق زیادی با هم ندارند و آن‌ها را یکی می‌گیریم. خب مثلاً همه‌ی اعداد دسته‌ی $[۱۲,۱۶)$ را می‌خواهیم یکسان بگیریم.

به نظر شما همه را ۱۲ بگیریم یا همه را ۱۶ بگیریم؟ هیچ‌کدام! بهترین انتخاب این است که مقدار مشترک داده‌ها را عدد وسطی دسته بگیریم. به این عدد می‌گوییم مرکز دسته یا نشان دسته یا نماینده‌ی دسته.

در واقع ارزش مشترک یا اندازه‌ی مشترک داده‌ها در دسته‌ی $[a_i,b_i)$ برابر است با: مرکز دسته $= x_i = \frac{a_i + b_i}{۲}$

روی محور هم ببینیم:



فاصله‌ی مرکز دسته از انتها و ابتدای دسته برابر $\frac{C}{۲}$ است.

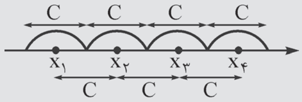
$$x_i = a_i + \frac{C}{۲} = b_i - \frac{C}{۲}$$

این‌جوری هم می‌شود گفت که حدود دسته، از مرکز دسته $\frac{C}{۲}$ بیشتر و کم‌ترند:

$$a_i = x_i - \frac{C}{۲}$$

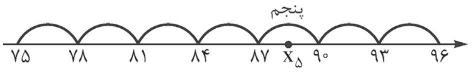
$$b_i = x_i + \frac{C}{۲} \Rightarrow [a_i,b_i) = [x_i - \frac{C}{۲}, x_i + \frac{C}{۲})$$

یک چیز کوچولوی دیگر هم بگوییم و تمام: فاصله‌ی مرکز دسته‌های متوالی از هم برابر C است، ببین:



$$x_۴ - x_۳ = x_۳ - x_۲ = x_۲ - x_۱ = x_n - x_{n-۱} = C$$

۳۷- **گزینه‌ی ۴** دامنه‌ی تغییرات می‌شود $R = ۹۶ - ۷۵ = ۲۱$؛ تعداد دسته‌ها هم ۷ است. (سؤال گفته) پس طول هر دسته می‌شود $C = \frac{R}{k} = \frac{۲۱}{۷} = ۳$، بنابراین دسته‌ها عبارت‌اند از:



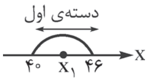
ما مرکز دسته‌ی پنجم را می‌خواستیم: $x_۵ = ۸۷ + ۱/۵ = ۸۸/۵$

۳۸- **گزینه‌ی ۱** طول دسته‌ها $C = ۶$ و تعداد دسته‌ها $k = ۵$ است. پس دامنه‌ی تغییرات می‌شود: $R = kC = ۳۰$

دامنه‌ی تغییرات برابر اختلاف ماکسیمم و مینیمم بود:

$$R = \max - \min \xrightarrow[\max=۷۰]{R=۳۰} ۳۰ = ۷۰ - \min \Rightarrow \min = ۴۰$$

یعنی کم‌ترین داده ۴۰ بوده. پس دسته‌ی اول می شود [۴۰,۴۶) و مرکزش هم $\frac{۴۰+۴۶}{۲}$ یا ۳ + ۴۰ است؛ یعنی $x_۱ = ۴۳$، محور را هم ببینید:



$$x_۱ = ۴۰ + ۳ = ۴۶ - ۳ = \frac{۴۰+۴۶}{۲} = ۴۳$$

۳۹- **گزینه‌ی ۳** این تمرین کتاب در صفحه‌ی ۶۱ جالب بود! طول دسته‌ها $۴ = ۷ - ۱۱$ است. پس جدول کامل این بود:

مرکز	حدود
۹	۱۱ − ۷
c ← ۱۳	۱۵ − ۱۱
۱۷	e ۱۹ − ۱۵
k ← ۲۱	۲۳ − ۱۹
۲۵	m ۲۷ − ۲۳

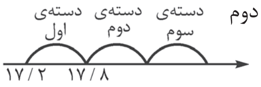
یعنی بین گزینه‌ها، $m = ۲۷$ درست نیست.

۴۰- **گزینه‌ی ۳** وقتی کوچک‌ترین و بزرگ‌ترین داده را می‌دهند، فقط یک کار می‌شود کرد؛ چه‌کاری؟

خب دامنه‌ی تغییرات را حساب می‌کنیم:

$$R = \max - \min = ۲۲/۶ - ۱۷/۲ = ۵/۴$$

حالا خودش گفته کران پایین دسته‌ی دوم ۱۷/۸ است. یعنی این‌جوری:



پس طول دسته هم $C = ۰/۶$ بوده $(۱۷/۸ - ۱۷/۲ = ۰/۶)$.

بیایید تعداد دسته‌ها را حساب کنیم: $k = \frac{R}{C} = \frac{۵/۴}{۰/۶} = ۹$

پس دقیقاً ۹ تا دسته داریم. این‌جوری:

حالا مرکز دسته‌ی آخر را می‌خواهد. یعنی از ۲۲/۶، نصف طول دسته به عقب بیاییم:

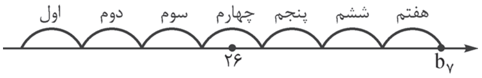
$$\text{مرکز دسته‌ی نهم} = x_۹ = b_۹ - \frac{C}{۲} = ۲۲/۶ - ۰/۳ = ۲۲/۳$$

۴۱- **گزینه‌ی ۲** تعداد دسته‌ها را داده، اختلاف بیشترین و کم‌ترین داده هم که R است. (یعنی همون دامنه‌ی تغییرات).

پس می‌دانیم $k = ۷$ و $R = ۲۸$ و می‌توانیم طول دسته را پیدا کنیم:

$$C = \frac{R}{k} = \frac{۲۸}{۷} = ۴$$

حالا به محور نگاه کنید:



بین ۷ تا دسته، دسته‌ی چهارم وسط است. سؤال گفته نشان دسته‌ی وسط، $x_۴ = ۲۶$ است. ما دنبال حد بالای دسته‌ی آخر هستیم. یعنی $b_۷$ را می‌خواهیم.

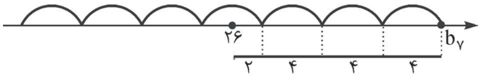
دوباره به محور نگاه کنید ... از ۲۶ تا $b_۷$ چه‌قدر فاصله است؟ چندتا دسته؟ خب سه‌تا دسته‌ی کامل و یک نصفه. یعنی داریم:

$$b_۷ = ۲۶ + ۳C + \frac{C}{۲} = ۲۶ + ۳/۵\,C$$

C هم که ۴ بود. پس می‌توانیم حد بالای آخری را پیدا کنیم:

$$b_۷ = ۲۶ + ۳(۴) + \frac{۴}{۲} = ۲۶ + ۱۴ = ۴۰$$

روی محور هم بد نیست:



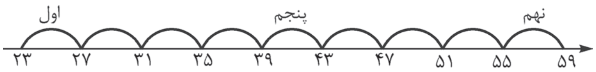
$$b_۷ = ۲۶ + ۱۴ = ۴۰$$

۴۲- **گزینه‌ی ۲** چون حدود دسته‌ی اول [۲۳,۲۶) است، طول دسته $C = ۳$ بوده $(۲۶ - ۲۳ = ۳)$ پس دامنه‌ی تغییرات $R = kC = ۱۲ \times ۳ = ۳۶$ است.

حالا گفته این داده‌ها را در ۹ طبقه دسته‌بندی می‌کنیم. دامنه‌ی تغییرات که فرقی نکرده، یعنی همون ۳۶ است. طول دسته‌های جدید برابر است با:

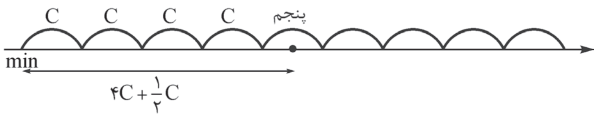
$$C' = \frac{R}{k'} = \frac{۳۶}{۹} = ۴$$

حالا مرکز دسته‌ی وسط را می‌خواهد. از داده‌ی مینیمم یعنی ۲۳ شروع می‌کنیم و دسته‌ها را می‌نویسیم: $\min = ۲۳\,,\, C = ۴$



پس دسته‌ی وسط یعنی دسته‌ی پنجم به صورت [۳۹,۴۳) است و مرکزش می‌شود: $x_۵ = \frac{a_۵ + b_۵}{۲} = \frac{۳۹+۴۳}{۲} = ۴۱$

مرکز دسته‌ی وسط (یعنی مرکز دسته‌ی پنجم)، از داده‌ی مینیمم به اندازه‌ی ۴/۵ C جلوتر است.



پس داریم: $x_۵ = \min + ۴/۵\,C = ۲۳ + ۴/۵\,(۴) = ۲۳ + ۱۶ + ۲ = ۴۱$

وقتی C عدد اعشاری می‌شود!

تا الان هر بار از $C=\frac{R}{k}$ استفاده کردیم، جواب C عدد صحیح شد! ولی بعضی وقت‌ها این‌جور نمی‌شود. مثلاً اگر $R = ۲۵$ و $k = ۷$ باشد، جواب C می‌شود تقریباً ۳/۵۷۱۴.

خب دسته به طول ۳/۵۷۱۴ که به درد نمی‌خورد. چون هدف از دسته‌بندی، از بین بردن اختلاف‌های جزئی داده‌ها بود نه این‌که در حد یک‌هزارم و یک‌صدم دقیق بشویم. پس باید C را عدد مناسب‌تری بگیریم.

مثلاً ۳/۶ یا ۴ خوب است. دقت کنید که نمی‌شود ۳/۵ را برداشت. چون ۷ دسته با طول ۳/۵ فاصله‌ای به طول $۲۴/۵ = ۳/۵ \times ۷$ را می‌پوشاند ولی دامنه‌ی تغییرات ما ۲۵ است.

پس اگر C اعشاری شد، باید عدد مناسب‌تری بیشتر از آن برداریم.

در تست معمولاً به سمت بالا گرد می‌کنیم؛ یعنی عدد صحیح بعدی را می‌گیریم.

۴۳- گزینه‌ی ۳ طول دسته‌ها می‌شود: $C=\frac{R}{k} \xrightarrow[k=۶]{R=۲۶} C=\frac{۲۶}{۶} \approx ۴/۳۳$

اما سؤال گفته دسته به طول صحیح، پس مجبوریم $C = ۵$ بگیریم.

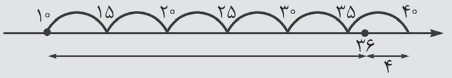
$C = ۴$ کافی نیست. به قول کتاب درسی‌تان: ۶ دسته به طول ۴، فاصله‌ی ۲۶ واحدی را نمی‌پوشاند. حالا مجموع طول دسته‌ها را می‌خواهیم. ۶ تا دسته بود و طولشان را هم ۵ گرفتیم. مجموع طول‌ها می‌شود $۶ \times ۵ = ۳۰$.

دسته‌های متقارن

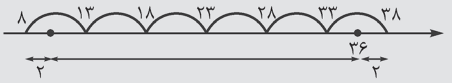
در سؤال قبلی دیدیم که اگر C اعشاری بشود، خودمان درستش می‌کنیم و عدد مناسب‌تری بیشتر از آن می‌گیریم. کتابتان گفته موضوع تقارن از زیباترین مفاهیم ریاضی است. ما هم که دنبال زیبایی هستیم (پناه بر خدا)، پس دسته‌بندی را متقارن می‌کنیم.

قضیه این‌جوری است که وقتی C مناسب‌تری انتخاب می‌کنیم، جمع طول دسته‌ها از دامنه‌ی تغییرات بیشتر است. آن‌وقت حد بالای دسته‌ی آخر، از داده‌ی ماکسیمم بیشتر می‌شود (یعنی دسته‌ها از دامنه‌ی تغییرات بیرون می‌آیند). ما به خاطر تقارن و زیبایی، این مقدار اضافی را دو قسمت می‌کنیم و در سروته می‌گذاریم. یعنی دسته‌ی اول را زودتر شروع می‌کنیم. مثلاً در سؤال قبل، دامنه‌ی تغییرات ۲۶ بود و ما شش دسته به طول ۵ گرفتیم که ۳۰ واحد می‌شود. یعنی ۴ تا بیشتر از ۲۶. این ۴ تا را در دو طرف می‌گذاریم یعنی دسته‌ی اول را دو واحد زودتر شروع می‌کنیم:

دسته‌های عادی:



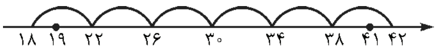
دسته‌های متقارن:



کتاب درسی گفته این دومی خوشگل‌تر است. من بی‌تقصیرم!

۴۴- گزینه‌ی ۳ دامنه‌ی تغییرات این داده‌ها می‌شود $R = ۴۱ - ۱۹ = ۲۲$، برای ۶ تا دسته، باید طول دسته $C=\frac{R}{k}=\frac{۲۲}{۶} \simeq ۳/۶۷$ باشد. چون سؤال طول دسته صحیح خواسته باید $C = ۴$ بگیریم. حالا ۶ تا دسته به طول ۴ داریم که ۲۴ واحد می‌شود. اما دامنه‌ی تغییرات ما ۲۲ است.

پس دو واحد اضافه داریم که قرار شد در سروته بگذاریم. (هر طرف یک واحد) پس دسته‌ها به این صورت است:



بنابراین حدود دسته‌ی دوم [۲۲,۲۶) است.

فراوانی

بعد از دسته‌بندی باید ببینیم چند نفر در هر دسته هستند. مثلاً معلم شما می‌پرسد:

چند نفر بین ۱۰ تا ۱۲ اند؟ چند نفر بین ۱۲ تا ۱۴ هستند و ....

به تعداد دفعات تکرار یک داده یا تعداد داده‌های یک دسته می‌گوییم فراوانی. البته درست‌تر این است که بگوییم فراوانی مطلق (ولی تنبلی است و هزار دردسرا).

فراوانی مطلق داده‌ی x_i را با f_i نشان می‌دهیم. مثلاً اعداد حاصل از ۲۰ بار پرتاب تاس را طبقه‌بندی کردیم:

x_i	۱	۲	۳	۴	۵	۶
f_i	۵	۴	۳	۳	۳	۲

از این جدول می‌فهمیم عدد ۴، ۳ بار ظاهر شده. یا به زبان ریاضی: فراوانی داده‌ی $x_۴ = ۴$ برابر $f_۴ = ۳$ است. بقیه‌ی فراوانی‌ها هم معلوم‌اند:

$f_۱ = ۵ , f_۲ = ۴ , f_۳ = ۳$

$f_۵ = ۳ , f_۶ = ۲$

جمع همه‌ی فراوانی‌ها می‌شود تعداد کل افراد نمونه. یعنی داریم: $f_۱ + f_۲ + ... + f_k = n$

۴۵- گزینه‌ی ۲ چیزی که سؤال گفته یعنی جدول این‌شکلی بوده:

f	۱۲	۹	۹	۵	۵	۵	۱	۱

ما تعداد داده‌ها را می‌خواهیم. پس باید فراوانی‌ها را با هم جمع کنیم:

$n = f_۱ + f_۲ + ... = ۱۲ + ۹ + ۹ + ۵ + ۵ + ۵ + ۱ + ۱ = ۴۷$

فراوانی نسبی

توی یک کلاس ۳ نفر نمره‌ی ۱۸ تا ۲۰ آورده‌اند و توی یک کلاس دیگر، ۷ نفر نمره‌ی ۱۸ تا ۲۰ گرفته‌اند. حالا سهم دانش‌آموزان قوی‌تر در کدام کلاس بیشتر است؟ یک کم فکر کنید... به این سؤال نمی‌شود جواب داد! چون به تعداد کل نفرات هر کلاس بستگی دارد. مثلاً ۳ نفر از بین ۱۰ نفر، خیلی بهتر از ۷ نفر در بین ۱۰۰ نفر است. ($\frac{۳}{۱۰} > \frac{۷}{۱۰۰}$)

یعنی فراوانی مطلق چیز به‌دردبخوری نیست. مهم این است که ببینیم آن دسته (یا داده) به نسبت کل چه‌قدر سهم دارد.

پس فراوانی را تقسیم بر تعداد کل می‌کنیم تا نسبت به دست بیاید. به این نسبت می‌گوییم فراوانی نسبی:

$$\text{فراوانی نسبی} = \frac{f_i}{n} = \frac{\text{فراوانی}}{\text{کل (جمع فراوانی‌ها)}}$$

جواب فراوانی نسبی یک عدد کسری بین ۰ و ۱ می‌شود.

جمع همه‌ی فراوانی‌های نسبی می‌شود ۱. چون: $\frac{f_۱}{n} + \frac{f_۲}{n} + \ldots = \frac{f_۱ + f_۲ + \ldots}{n} = \frac{n}{n} = ۱$

پس برای از بین بردن تأثیر اختلاف نمونه‌ها (کلاس‌های بزرگ و کوچک)، فراوانی را بر کل تقسیم می‌کنیم.

۴۶- گزینه‌ی ۱

۴۷- گزینه‌ی ۱ خب گفتیم فراوانی نسبی یعنی فراوانی تقسیم بر تعداد کل. اول تعداد کل را پیدا کنیم: جمع فراوانی‌ها = تعداد کل

$$n = f_۱ + f_۲ + f_۳ + f_۴ = ۶ + ۸ + ۴ + ۱۲ = ۳۰$$

حالا فراوانی نسبی دسته‌ی سوم: $\frac{f_۳}{n} = \frac{۴}{۳۰} = \frac{۲}{۱۵}$

۴۸- گزینه‌ی ۲ فراوانی نسبی چی بود؟ فراوانی تقسیم بر کل!

سؤال گفته فراوانی نسبی ۰/۰۵ و تعداد $n = ۸۰$ است. پس داریم:

$$۰/۰۵ = \frac{f_i}{۸۰} \xrightarrow{\text{طرفین و وسطین}} f_i = ۰/۰۵ \times ۸۰ = \frac{۵}{۱۰۰} \times ۸۰ = \frac{۴۰۰}{۱۰۰} = ۴$$

یعنی فراوانی مطلق این طبقه، ۴ است.

۴۹- گزینه‌ی ۳ فراوانی نسبی $\frac{f_i}{n}$ بود. پس داریم:

$$۰/۱۵ = \frac{۱۲}{n} \xrightarrow{\text{طرفین و وسطین}} ۰/۱۵n = ۱۲$$

$$\Rightarrow n = \frac{۱۲}{۰/۱۵} = \frac{۱۲}{\frac{۱۵}{۱۰۰}} = \frac{\overset{۴}{\cancel{۱۲}} \times ۱۰۰}{\underset{۵}{\cancel{۱۵}}} = ۴ \times ۲۰ = ۸۰$$

آهان، تعداد کل داده‌ها ۸۰ تا بوده.

حال پرسیده فراوانی نسبی دسته‌ای با ۱۰ داده چه‌قدر است؟ خب کاری نداره: $\text{فراوانی نسبی} = \frac{f}{n} = \frac{۱۰}{۸۰} = \frac{۱}{۸}$

می‌توانیم مستقیماً از تناسب برویم:

فراوانی نسبی	فراوانی
۰/۱۵	۱۲
x	۱۰

مقدار x را هم بلدیم حساب کنیم: $x = \frac{۱۰ \times ۰/۱۵}{۱۲} = \frac{۱/۵}{۱۲}$

اگر به جای ۱/۵ بنویسیم $\frac{۳}{۲}$ کار تمام است: $x = \frac{\frac{۳}{۲}}{۱۲} = \frac{۳}{۲۴} = \frac{۱}{۸}$

۵۰- گزینه‌ی ۳ فراوانی نسبی دسته‌ی وسط می‌شود: $\frac{f_۴}{n} = \frac{f_۴}{۸۰}$

توی ۷ طبقه، دسته‌ی وسط طبقه‌ی چهارم است!

سؤال گفته ۲۰ تا داده‌ی جدید اضافه می‌کنیم، پس n می‌شود: $۸۰ + ۲۰ = ۱۰۰$ ولی نمی‌دانیم چندتا از این ۲۰ داده‌ی جدید، توی دسته‌ی چهارم می‌روند. فرض کنیم x تا به دسته‌ی چهارم اضافه بشوند. فراوانی نسبی دسته‌ی چهارم می‌شود: $\frac{f_۴ + x}{۸۰ + ۲۰} = \frac{f_۴ + x}{۱۰۰}$

اما صورت سؤال می‌گوید فراوانی نسبی دسته‌ی چهارم تغییر نمی‌کند. یعنی این دوتا با هم مساوی‌اند:

$$\frac{f_۴}{۸۰} = \frac{f_۴ + x}{۱۰۰} \xrightarrow{\text{طرفین و وسطین}} ۱۰۰f_۴ = ۸۰(f_۴ + x)$$

$$\xrightarrow{\text{تقسیم بر ۲۰}} ۵f_۴ = ۴(f_۴ + x) = ۴f_۴ + ۴x \Rightarrow f_۴ = ۴x$$

از ما پرسیده «نسبت افزایش داده‌های مذکور به فراوانی مطلق قبلی کدام است؟»

افزایش داده‌ها x بود، فراوانی قبلی هم $f_۴$ بود، پس $\frac{x}{f_۴}$ را می‌خواهیم که می‌شود: $\frac{x}{f_۴} = \frac{x}{۴x} = \frac{۱}{۴}$

۵۱- گزینه‌ی ۴ در سه طبقه‌ی اول روی هم $۴ + ۵ + ۷ = ۱۶$ تا داده وجود دارد.

کل داده‌ها $۴ + ۵ + ۷ + ۴ = ۲۰$ تا هستند.

نسبت هم یعنی فراوانی نسبی. پس نسبت داده‌های سه طبقه‌ی اول می‌شود:

$$\frac{f}{n} = \frac{۱۶}{۲۰} = ۰/۸$$

درصد فراوانی نسبی

$\frac{۳}{۷}$ بیشتر است یا $\frac{۲}{۵}$؟ یک نفر در امتحان عربی $\frac{۳}{۷}$ تست‌ها را درست زده و دیگری $\frac{۲}{۵}$ تست‌ها را درست زده است. کدوم بهترن؟ خب هر کس زیست بیشتر زده باشد!!!

نه پدر جان! عربی کدامشان بهتر است؟ واقعیتش اینه که مردم با کسر راحت نیستند. مقایسه‌ی کسرها یک کم زحمت دارد. اما درصد را همه بهتر می‌فهمند.

پس بیایید فراوانی نسبی را به درصد بیان کنیم. این‌جوری: ٪۱۰۰ × فراوانی نسبی = درصد فراوانی نسبی

پس اگر درصد فراوانی را با P_i (برای داده‌ی i اُم یا دسته‌ی i اُم) نشان بدهیم، داریم: $P_i = \frac{f_i}{n} \times ٪۱۰۰$

جمع درصدهای فراوانی باید بشود ۱۰۰. (با عرض شرمندگی)

جواب درصدهای فراوانی هم یک چیزی بین ۰ و ۱۰۰ است!

۵۲- گزینه‌ی ۳ این عددها ۱۶ تا هستند، پس $n = ۱۶$.

ما دنبال مضرب‌های ۳ هستیم، یعنی ۳، ۶ و ۹ را می‌خواهیم. ۶ تا از داده‌ها مضرب ۳ هستند. پس درصد فراوانی نسبی مضارب ۳ می‌شود:

$$P = \frac{f}{n} \times ۱۰۰ = \xrightarrow[n=۱۶]{f=۶} P = \frac{۶}{۱۶} \times ۱۰۰ = \frac{۳}{\underset{۲}{\cancel{۸}}} \times \overset{۲۵}{\cancel{۱۰۰}} = \frac{۷۵}{۲} = ۳۷/۵$$

۵۳- گزینه‌ی ۴ اول بیایید جدول را کامل کنیم، یعنی x را پیدا کنیم. جمع درصدهای فراوانی باید ۱۰۰ باشد، یعنی جمع همه‌ی اعداد سطر دوم می‌شود ۱۰۰: $۱۰+۱۵+۱۸+x+۲۰+۱۲ = ۷۵ + x = ۱۰۰$

پس: $x = ۲۵$

یعنی درصد فراوانی دسته‌ی چهارم ۲۵ است. فرمول درصد فراوانی را ببینید:

$$P = \frac{f}{n} \times ۱۰۰$$

تعداد کل داده‌ها (یعنی دانش‌آموزان) هم $n = ۱۲۰$ است. پس می‌توانیم f را پیدا کنیم:

$$\xrightarrow[n=۱۲۰]{P=۲۵} \overset{۱}{\cancel{۲۵}} = \frac{f}{۱۲۰} \times \overset{۴}{\cancel{۱۰۰}} \Rightarrow ۴f = ۱۲۰ \Rightarrow f = ۳۰$$

رسیدیم به این‌که درصد فراوانی دسته‌ی چهارم ۲۵ درصد است.

تعداد کل هم که ۱۲۰ بود. پس در این دسته ۲۵ درصد (یعنی $\frac{۱}{۴}$) از ۱۲۰ نفر داریم که می‌شود ۳۰ تا.

۵۴- گزینه‌ی ۴ تعداد کل خانواده‌ها برابر است با جمع کل فراوانی‌ها: $n = F_۱ + F_۲ + \cdots = ۹ + ۱۸ + ۲۲ + ۲۶ + ۵ = ۸۰$

ما خانواده‌های ۳ یا ۴ نفری را می‌خواهیم. پس روی هم $۲۶ + ۲۲$ یعنی ۴۸ تا خانواده موردنظر ما هستند. درصدشان می‌شود:

$$P = \frac{f}{n} \times ۱۰۰ = \frac{\overset{۶}{\cancel{۴۸}}}{\underset{۱۰}{\cancel{۸۰}}} \times ۱۰۰ = ۶۰$$

۵۵- گزینه‌ی ۴ وزن‌های کم‌تر از ۵۰ (مگس‌وزن!) در چهار دسته‌ی اول هستند.

فراوانی این‌ها روی هم می‌شود: $f = ۸ + ۹ + ۱۲ + ۱۵ = ۴۴$

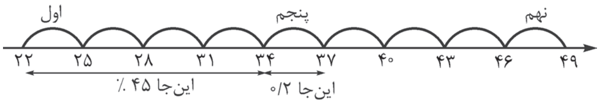
تعداد کل داده هم برابر مجموع فراوانی‌ها است:

$$n = ۸ + ۹ + ۱۲ + ۱۵ + ۶ + ۵ = ۵۵$$

پس درصد این‌ها می‌شود:

$$P = \frac{f}{n} \times ۱۰۰ = \frac{\overset{۴}{\cancel{۴۴}}}{\underset{۵}{\cancel{۵۵}}} \times ۱۰۰ = ۸۰$$

۵۶- گزینه‌ی ۳ اول یک نگاه به محور کنیم:



بین ۹ تا دسته، دسته‌ی پنجم وسط است. سؤال گفته فراوانی نسبی دسته‌ی وسط ۰/۲ است و ۴۵ درصد داده‌ها کم‌تر از ۳۴ هستند.

خب بالاخره تصمیم بگیرید! یا فراوانی نسبی یا درصد فراوانی. نمی‌شود در یک قسمت بگوییم ۴۵ درصد و در قسمت بعدی بگوییم ۰/۲. پس به جای ۰/۲ هم می‌گوییم ۲۰ درصد. روی هم از ۲۲ تا ۳۷ می‌شود $۲۰ + ۴۵$ یعنی ۶۵ درصد. ما تعداد داده‌ها (یعنی فراوانی مطلق) کم‌تر از ۳۷ را می‌خواهیم:

$$\text{درصد} = \frac{\text{فراوانی}}{\text{کل}} \times ۱۰۰$$

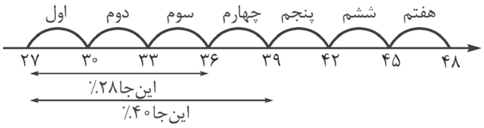
$$۶۵ = \frac{f}{\underset{۶}{\cancel{۱۲۰}}} \times \overset{۵}{\cancel{۱۰۰}} \xrightarrow{\text{طرفین و وسطین}} \cancel{۵}f = ۶ \times \overset{۱۳}{\cancel{۶۵}} = ۷۸$$

۵۷- گزینه‌ی ۲ کوچک‌ترین و بزرگ‌ترین داده را گفته که R را پیدا کنیم: $R = ۴۷/۸ - ۲۷ = ۲۰/۸$

حالا ۷ طبقه داریم. پس می‌توانیم طول دسته‌ها را پیدا کنیم:

$$C = \frac{R}{k} = \frac{۲۰/۸}{۷} \xrightarrow{\text{گرد کنیم}} C = ۳$$

باز برویم سراغ محور جادویی خودمان:



توی هفت‌تا دسته، دسته‌ی وسط چهارمی است.

به محور نگاه کنید، سهم دسته‌ی وسط (چهارم) می‌شود $۱۲ = ۲۸ - ۴۰$ درصد.

تعداد کل داده‌ها $n = ۷۵$ است. پس فراوانی مطلق دسته‌ی وسط را می‌توانیم حساب کنیم: $P = \frac{f}{n} \times ۱۰۰$

$$\xrightarrow[n=۷۵]{P=۱۲} ۱۲ = \frac{f}{\underset{۳}{\cancel{۷۵}}} \times \overset{۴}{\cancel{۱۰۰}} \xrightarrow{\text{طرفین وسطین}} ۴f = ۳۶ \Rightarrow f = ۹$$

فراوانی تجمعی

آقا چند نفر بین ۰ تا ۲۰ درصد زدن؟ چند نفر ۲۰ تا ۴۰ درصد زدن؟ چند نفر ۴۰ تا ۶۰ درصد زدن؟ ...

ول کن! درصد زیر ۶۰ که به درد نمی‌خورد!!! بعضی وقت‌ها به جای فراوانی هر دسته، مجموع فراوانی‌ها از دسته‌ی اول تا آن دسته برای ما مفیدتر است. به این فراوانی می‌گویند تجمعی.

فراوانی تجمعی دسته‌ی اول، همان فراوانی مطلقش است. اما برای دسته‌ی دوم، باید فراوانی دسته‌ی اول و دوم را جمع کنیم. برای دسته‌ی سوم، باید اول و دوم و سوم را جمع کنیم؛ همین‌جور تا آخر. برای دسته‌ی آخر باید همه را با هم جمع کنیم. یعنی فراوانی تجمعی دسته‌ی آخر می‌شود n (تعداد کل داده‌ها).

مثلاً اگر فراوانی‌ها به ترتیب ۶،۴،۵،۳ باشند، فراوانی‌های تجمعی به ترتیب عبارت‌اند از: $\underset{\text{همون مطلق}}{۳}$ و $۸ = ۵ + ۳$ و $۱۲ = ۴ + ۵ + ۳$ و $\underset{\text{تعداد کل (n)}}{۱۸} = ۶ + ۴ + ۵ + ۳$

اگر فراوانی‌های تجمعی را به ما بدهند، می‌توانیم فراوانی‌های مطلق را حساب کنیم.

اولین فراوانی تجمعی که فراوانی مطلق هم هست. برای بعدی‌ها، هر فراوانی تجمعی را منهای قبلی‌اش می‌کنیم تا فراوانی مطلق دربیاید. مثلاً اگر تجمعی‌ها ۲۰،۱۷،۱۱،۴ باشند، مطلق‌ها به ترتیب ۴ و $۷ = ۴ - ۱۱$ و $۶ = ۱۱ - ۱۷$ و $۳ = ۱۷ - ۲۰$ است. راستی تعداد کل داده‌ها هم ۲۰ است. (آخری!)

فراوانی‌های تجمعی همیشه صعودی‌اند؛ یعنی فراوانی تجمعی هر دسته از فراوانی تجمعی قبلی کم‌تر نیست.

۵۸- گزینه‌ی ۴

فراوانی تجمعی دسته‌ی چهارم می‌شود جمع فراوانی‌های دسته‌ی اول تا چهارم: $f_۱ + f_۲ + f_۳ + f_۴ = ۲ + ۸ + ۵ + ۳ = ۱۸$

۵۹- گزینه‌ی ۴ اول بریم سراغ آخر کار!! سؤال از ما فراوانی تجمعی دسته‌ی دوم را خواسته، یعنی جمع فراوانی‌های اول و دوم؛ یا به زبان خودمان $f_۱ + f_۲$ را می‌خواهیم.

حالا چی به ما داده؟ فراوانی نسبی داده.

فراوانی نسبی چی بود؟ فراوانی تقسیم بر کل. آهان! معلوم شد:

$$\text{فراوانی نسبی دسته‌ی اول} = \frac{f_۱}{n} \Rightarrow ۰/۳ = \frac{f_۱}{۶۰} \Rightarrow f_۱ = ۶۰ \times ۰/۳ = ۱۸$$

$$\text{فراوانی نسبی دسته دوم} = \frac{f_۲}{n} \Rightarrow ۰/۱۵ = \frac{f_۲}{۶۰} \Rightarrow f_۲ = ۶۰ \times ۰/۱۵ = ۹$$

بنابراین جواب سؤال یعنی $f_۱ + f_۲$ می‌شود $۱۸ + ۹ = ۲۷$.

۶۰- گزینه‌ی ۳ فاصله‌ی $[۱۸/۵, ۲۱/۵)$ یعنی همان دسته به مرکز ۲۰ (وسط ۲۱/۵ و ۱۸/۵ می‌شود ۲۰). پس دسته‌ی سوم موردنظر سؤال است. از فراوانی‌های تجمعی جدول دوتا چیز می‌فهمیم:

اولاً تعداد کل داده‌ها $n = ۴۰$ است. (آخرین فراوانی تجمعی)

ثانیاً فراوانی دسته‌ی سوم، $۲۵ - ۱۳ = ۱۲$ است.

درصد داده‌های دسته‌ی سوم می‌شود: $P_۳ = \frac{f_۳}{n} \times ۱۰۰ = \frac{\cancel{۱۲}^{۳}}{\cancel{۴۰}_{۱۰}} \times ۱۰۰ = ۳۰$

۶۱- گزینه‌ی ۲ جدول فراوانی می‌گوید تعداد کل داده‌ها $n = ۵۰$ است. (فراوانی تجمعی آخری) و فراوانی مطلق دسته‌ی وسط (یعنی دسته‌ی سوم)، $a - ۱۴$ است. (فراوانی تجمعی را منهای قبلی‌اش می‌کنیم)، حال درصد فراوانی نسبی دسته‌ی وسط می‌شود: $P_۳ = \frac{f_۳}{n} \times ۱۰۰$

$$\Rightarrow ۲۴ = \frac{a - ۱۴}{\cancel{۵۰}} \times \cancel{۱۰۰}^{۲} \Rightarrow ۱۲ = a - ۱۴ \Rightarrow a = ۲۶$$

سؤال فراوانی مطلق دسته‌ی چهارم را می‌خواهد. قرار بود تجمعی را منهای قبلی کنیم: $f_۴ = ۴۱ - a = ۴۱ - ۲۶ = ۱۵$

فراوانی تجمعی نسبی و درصد فراوانی تجمعی

اگر فراوانی‌های مطلق از طبقه‌ی اول تا طبقه‌ی دلخواهی را با هم جمع کنیم، فراوانی تجمعی آن طبقه به دست می‌آید.

حالا اگر فراوانی‌های نسبی از طبقه‌ی اول تا طبقه‌ی دلخواه را با هم جمع کنیم، فراوانی تجمعی نسبی آن طبقه به دست می‌آید و اگر درصدهای فراوانی نسبی از طبقه‌ی اول تا طبقه‌ی دلخواه را جمع کنیم، به درصد فراوانی تجمعی می‌رسیم. بگذارید این کارها را ببینیم:

دسته	اول	دوم	سوم	چهارم	
فراوانی	۳	۱	۴	۲	
فراوانی تجمعی	۳	۴	۸	۱۰	کل
فراوانی نسبی	۰/۳	۰/۱	۰/۴	۰/۲	
فراوانی تجمعی نسبی	۰/۳	۰/۴	۰/۸	۱	کل
درصد فراوانی	۳۰	۱۰	۴۰	۲۰	
درصد فراوانی تجمعی	۳۰	۴۰	۸۰	۱۰۰	

- چهارتا ویژگی برای فراوانی‌های تجمعی (تجمعی عادی، تجمعی نسبی یا درصد تجمعی) داریم:

۱) برای دسته‌ی اول، فراوانی تجمعی و غیرتجمعی برابر است.

۲) برای دسته‌ی آخر فراوانی تجمعی برابر مجموع فراوانی‌ها است. (مطلق: n، نسبی: ۱، درصدی: ۱۰۰)

۳) همیشه فراوانی‌های تجمعی صعودی‌اند (یعنی از یک دسته‌ی به دسته‌ی بعد زیاد می‌شوند و یا ثابت می‌مانند.)

۴) اگر فراوانی تجمعی هر دسته را منهای دسته‌ی قبلی‌اش کنیم، به فراوانی غیرتجمعی می‌رسیم.

- برای درصد فراوانی تجمعی یا فراوانی تجمعی نسبی، می‌توانیم از اول فراوانی‌های نسبی را حساب کنیم و بعد با هم جمع کنیم یا این‌که اول مطلق‌ها را تجمعی کنیم و بعد نسبی‌اش را به دست بیاوریم.

۶۲- گزینه‌ی ۴ $x_۳ = ۴$ یعنی دسته‌ی سوم. فراوانی تجمعی این دسته برابر است با: $۵ + ۷ + ۸ = ۲۰$. ما درصد فراوانی تجمعی را می‌خواهیم. تعداد کل داده‌ها $۵ + ۷ + ۸ + ۵ + ۵ = ۳۰$ است و درصد موردنظر ما می‌شود:

$$P = \frac{۲۰}{۳۰} \times ۱۰۰ = \frac{۲}{۳} \times ۱۰۰ \simeq ۶۶/۷$$

- $\frac{۲}{۳}$ می‌شود ...۰/۶۶۶۶ همین

نمودار ساقه و برگ

از نمودار ساقه و برگ (stem & leaf) وقتی استفاده می‌شود که تعداد ارقام داده‌ها خیلی به هم نزدیک باشد.

مثلاً دمای بدن افراد یک چیزی بین ۳۵ تا ۴۲ است. یعنی حتماً دهگانش ۳ یا ۴ است.

ببینید: دمای بدن ۱۰ نفر : ۳۷,۳۶,۳۶,۳۷,۳۷,۳۷,۳۸,۳۹,۴۰,۴۰

- چهار نفر آخر تب دارند!

حالا می‌توانیم این داده‌ها را با استفاده از ساقه و برگ نشان بدهیم. رقم سمت راست برگ است و سایر ارقام، ساقه هستند. این‌جوری:

ساقه	برگ
۳	۷,۶,۶,۷,۷,۷,۸,۹
۴	۰,۰

البته باید در جلوی هر ساقه، برگ‌هایش را صعودی بنویسیم (یعنی از کم به زیاد):

ساقه	برگ
۳	۶۶۷۷۷۷۸۹
۴	۰۰

به این نمودار نگاه کنید! چه چیزهایی می‌گوید؟ بیشترین داده ۴۰ و کم‌ترین ۳۶ بوده؛ دوتا ۳۶ و چهارتا ۳۷ و یک ۳۸ و یک ۳۹ و دوتا ۴۰ داشتیم.

فایده‌ی نمودار ساقه و برگ این است که از خود داده‌ها استفاده می‌کند و همه‌ی داده‌ها را توی خودش دارد.

اگر قرار باشد از روی نمودار داده‌ها را بخوانیم، یک کلید هم به ما می‌دهند. مثلاً می‌گویند: ۳۷ = ۷ ۳ یا ۳/۴ = ۴ ۳.

-۶۳ گزینه‌ی ۳ نمودار می‌گوید داده‌ها از ۵/۱ تا ۸/۷ دقیقه هستند. پس داریم: $R = ۸/۷ - ۵/۱ = ۳/۶$

-۶۴ گزینه‌ی ۳ کار ساده‌ای است! تعداد داده‌ها یعنی همان تعداد برگ‌ها. در سطر اول ۴ تا، در سطر دوم ۶ تا و در سطر سوم ۱ برگ داریم. پس روی هم ۱۱ داده داریم.

-۶۵ گزینه‌ی ۳ قرارمان این شد که جلوی هر ساقه، برگهایش را بنویسیم. اما از کم به زیاد (یعنی صعودی)؛ پس باید در سطر اول، عدد توی مربع بین ۲ و ۵ باشد. یعنی ۲، ۳، ۴ و ۵ می‌تواند باشد.

اما صبر کنید! توی سطر سوم هم مربع گذاشته، این مربع بعد از ۴ است. پس می‌تواند ۴، ۵، ۶ و ... باشد. حالا برای این‌که هر دوتا شرط تأمین بشود، اشتراک می‌گیریم. پس در مربع‌ها می‌توانیم فقط ۴ یا ۵ بنویسیم. یعنی فقط دوتا عدد.

-۶۶ گزینه‌ی ۲ سؤال می‌گوید جلوی ۱۴، این برگ‌ها نوشته شده: ۰۳۵۵۵۵۶۸۸۹ یعنی چهارتا داده ۱۴/۵ بودند. حالا پرسیده چند درصد از داده‌ها ۱۴/۵ بودند؟ ما کلاً ۵۰ داده داریم، چهارتا از ۵۰ تا می‌شود چند درصد؟ اگر گفتی ...

احتیاجی به فکر نیست! فرمول درصد فراوانی را به یاد بیاورید:

$$P = \frac{۴}{\underset{۱}{\cancel{۵۰}}} \times \overset{۲}{\cancel{۱۰۰}} = ۸$$

-۶۷ گزینه‌ی ٤ ناکم‌تر از ۴۰ یعنی بیشتر یا مساوی ۴۰، پس داده‌های سطر دوم {۴۰,۴۱,۴۳,۴۴,۴۴,۴۵,۴۷,۴۷} را می‌خواهد. (۸ تا هستند.)

تعداد کل برگ‌ها را بشماریم: (۹ تا در سطر اول، ۸ تا سطر دوم و ۸ تا سطر سوم) $n = ۹ + ۸ + ۸ = ۲۵$

پس کلاً ۲۵ تا داده داریم. سؤال پرسیده چند درصدشان توی سطر دوم‌اند؟

فرمول درصد فراوانی را ببینید: $P = \frac{f}{n} \times ۱۰۰ = \frac{۸}{۲۵} \times ۱۰۰ = ۳۲$

نمودارها ـ نمودار میله‌ای

از نمودارها یا شاخص‌های هندسی، برای به تصویر درآوردن و تجسم جامعه استفاده می‌کنیم. نمودار میله‌ای برای متغیرهای کیفی (مثل نوع بیمه، نوع گروه خون و ...) یا متغیرهای کمی گسسته خوب است. در نمودارهای میله‌ای، مقایسه‌ی فراوانی‌ها مهم است.

اگر برای داده‌های کیفی نمودار میله‌ای رسم بشود، ترتیب قرارگرفتن میله‌ها مهم نیست. آن وقت بهتر است طوری نمودار بکشیم که فراوانی‌ها از بزرگ به کوچک یا از کوچک به بزرگ باشند (آن‌وقت مقایسه را سریع‌تر انجام می‌دهیم.)

در نمودار میله‌ای، متغیر تصادفی را روی محور x ها و طول میله‌ها (فراوانی یا درصد فراوانی) را روی محور y ها مشخص می‌کنیم.

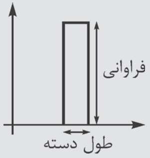
اگر روی محور y ها فراوانی داشته باشیم، اندازه‌ی جامعه برابر است با جمع طول‌های میله‌ها.

-۶۸ گزینه‌ی ٤ ۲۵ نفر O هستند، ۵ نفر AB، ۱۰ نفر B و ۱۵ نفر A، پس روی هم ۵۵ = ۲۵ + ۵ + ۱۰ + ۱۵ دانش‌آموز داریم.

نمودار مستطیلی

این نمودار برای متغیرهای کمی پیوسته خوب است (مثل سن، نمره، هزینه و ...)

در این نمودار برای هر دسته یک مستطیل می‌کشیم. قاعده‌ی این مستطیل روی محور x ها و برای طول دسته‌ها است. ارتفاعش هم متناسب با فراوانی است:



در نمودار مستطیلی، مساحت مستطیل‌ها با هم مقایسه می‌شوند.

هر چه‌قدر مساحت مستطیل بیشتر باشد، تعداد کسانی که در آن دسته قرار می‌گیرند بیشتر است.

اصلاً کتابتان نمودار مستطیلی را این‌جوری تعریف کرده: نمایشی از داده‌های دسته‌بندی‌شده که سطح مستطیل‌ها متناسب با فراوانی دسته‌ها است.

به نمودار مستطیلی، نمودار ستونی یا هیستوگرام یا بافت‌نگار هم می‌گویند!

-۶۹ گزینه‌ی ٤ این نمودار می‌گوید از ۰ تا ۲۰ سه نفر، از ۲۰ تا ۴۰ چهار نفر و از ۴۰ تا ۶۰، هفت نفر داریم.

پس تعداد نمره‌های کم‌تر از ۶۰ (یعنی فراوانی تجمعی دسته‌ی سوم) می‌شود:

۱۴ = ۷ + ۴ + ۳ نفر

تعداد کل نفرات کلاس، مجموع طول همه‌ی ستون‌ها است؛ یعنی:

$n = ۳ + ۴ + ۷ + ۹ + ۵ = ۲۸$

-۷۰ گزینه‌ی ٤ قرار شد در نمودار مستطیلی، مساحت مستطیل‌ها را مقایسه کنیم.

حالا اگر طول مستطیل‌ها با هم برابر بشوند (یعنی طول دسته‌ها مساوی باشند)، ارتفاع مستطیل‌ها (یعنی فراوانی‌ها) را با هم مقایسه می‌کنیم.

۷۱- گزینه‌ی ۳ گفتیم در نمودار مستطیلی، سطح مستطیل‌ها مقایسه می‌شوند. پس باید سطح‌ها را حساب کنیم.

دسته‌ی اول: $S_۱ = (۴-۱)\times ۵ = ۱۵$

دسته‌ی دوم: $S_۲ = (۹-۴)\times ۳ = ۱۵$

دسته‌ی سوم: $S_۳ = (۱۷-۹)\times ۲ = ۱۶$

دسته‌ی چهارم: $S_۴ = (۱۹-۱۷)\times ۷ = ۱۴$

پس دسته‌ی سوم بیشترین مساحت، یعنی بیشترین فراوانی را دارد.

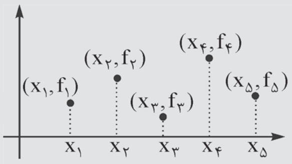
چندبر فراوانی

وقتی می‌خواهیم داده‌های کمی پیوسته را بررسی کنیم، چندبر فراوانی بهترین نمودار است. ماجرا این است که کتابتان از نمودار مستطیلی خیلی راضی نیست! به قول کتاب: در نمودار مستطیلی، فراوانی در دسته‌ها تغییر نمی‌کند. منظورش این است که نمودار مستطیلی یک شکل پلکانی دارد و فراوانی را در طول هر دسته، ثابت می‌گیرد.

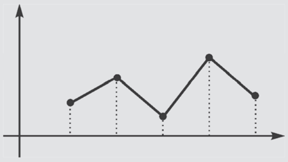
پس برای این‌که تغییرات را بهتر نشان بدهیم، از چندبر استفاده می‌شود.

اول باید مرکز دسته‌ها را مشخص کنیم. بعد نقاط (x_i , f_i) را روی دستگاه محورهای مختصات بگذاریم.

طول این نقاط برابر مرکز دسته و عرض آن‌ها همان فراوانی است. مثلاً این‌جوری:



اگر این نقطه‌ها را با خطی شکسته به هم وصل کنیم، چندبر فراوانی به دست می‌آید:



کتاب درسی تأکید کرده که اگر به جای فراوانی، از فراوانی‌های نسبی استفاده کنیم بهتر است. چون عملاً فراوانی هر دسته را با کل مقایسه می‌کنیم. می‌توانیم چندبر را هم با فراوانی نسبی رسم کنیم. این چندبر فراوانی نسبی هم برای داده‌های پیوسته خوب است.

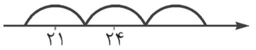
۷۲- گزینه‌ی ۲

۷۳- گزینه‌ی ۳ هزینه، داده‌ی کمی پیوسته است. گفتیم چندبر فراوانی نسبی برای داده‌های پیوسته از همه مناسب‌تر است.

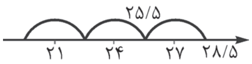
۷۴- گزینه‌ی ۴ در نمودار چندبر، طول نقاط، مرکز دسته است و عرض نقاط فراوانی است.

پس این نقطه‌ها می‌گویند مراکز دسته به صورت ۱۱/۵، ۱۴/۵، ۱۷/۵ و ۲۰/۵ بوده. یعنی ۴ تا دسته با طول ۳ داریم. فراوانی‌ها هم به ترتیب ۳، ۶، ۷ و ۴ بوده، پس: $n = ۴+۷+۶+۳ = ۲۰$ و فراوانی تجمعی دسته‌ی دوم، $۹ = ۶ + ۳$ است.

۷۵- گزینه‌ی ۴ طول نقطه‌ها، مرکز دسته بود. پس الان مرکز دو دسته‌ی متوالی ۲۴،۲۱ است:



این یعنی طول دسته، ۳ بوده و مرکز دسته‌ی بعدی $۲۷ = ۳ + ۲۴$ می‌شود. ما حدود دسته‌ی بعدی را می‌خواهیم. پس از ۲۷، نصف طول دسته را زیاد و کم می‌کنیم:



$[۲۷-۱/۵ , ۲۷+۱/۵) = [۲۵/۵ , ۲۸/۵)$

۷۶- گزینه‌ی ۲ نقطه‌های نمودار چندبر (x_i , f_i) هستند. طولشان مرکز دسته است و عرضشان فراوانی. پس دنبال مرکز دسته و فراوانی دسته‌ی وسط هستیم: (دسته‌ی وسط، دسته‌ی سوم است!)

مرکز: $x_۳ = \frac{۲۶+۲۹}{۲} = ۲۷/۵$

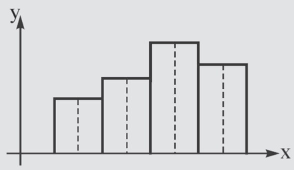
فراوانی: $f_۳ = ۲۵ - ۱۹ = ۶$

$\Rightarrow$ نقطه‌ی نمودار $= (۲۷/۵ , ۶)$

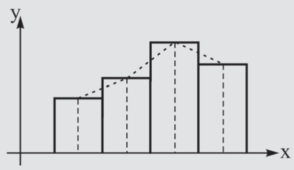
وقتی فراوانی تجمعی داریم، هر فراوانی تجمعی را منهای قبلی‌اش می‌کنیم تا به فراوانی [مطلق] برسیم.

از مستطیلی به میله‌ای و چندبر برسیم

اگر نمودار مستطیلی را داشته باشیم می‌توانیم از روی آن نمودار میله‌ای یا چندبر بکشیم. برای میله‌ای، باید از مرکز هر دسته به وسط عرض بالایی‌اش وصل کنیم:



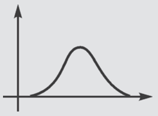
حالا برای رسیدن به چندبر، باید ته میله‌ها را به هم وصل کنیم. این‌جوری:



فقط یک اشکال هست! مساحت زیر این نمودار چندبر، با مساحت نمودار مستطیلی مساوی نیست. این را هم حل می‌کنیم. نگران نباشید.

اگر اندازه‌ی نمونه را بزرگ کنیم، تعداد دسته‌ها بیشتر می‌شود و پاره‌خط‌های بیشتری داریم. چندبر فراوانی کم‌کم به یک منحنی هموار شبیه می‌شود.

این منحنی بیان‌کننده‌ی وضعیت متغیر در جامعه است. در اکثر پدیده‌های طبیعی به شکل یک زنگوله درمی‌آید. این‌جوری:



۷۷- گزینه‌ی ۳

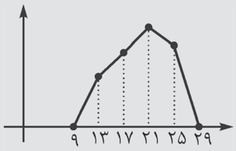
چندبر فراوانی تکمیل شده

یادتان هست که گفتم مساحت زیر نمودار چندبر با مستطیلی مساوی نیست؟ حالا می‌خواهیم این مشکل را حل کنیم، یعنی نمودار چندبر را تکمیل کنیم تا مساحتش (مساحت زیر آن) با نمودار مستطیلی مساوی بشود.

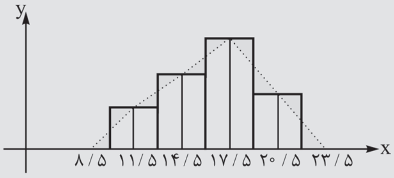
مساحت نمودار مستطیلی برابر است با: $nC =$ طول دسته × اندازه‌ی نمونه $= S$

قضیه از این قرار است: دوتا داده با فراوانی صفر به سروته داده‌ها اضافه می‌کنیم و نمودار را به محور x ها متصل می‌کنیم. این داده‌های جدید $x_۱ - C$ (از مرکز دسته‌ی اول به اندازه‌ی طول دسته کم کنیم) و $x_k + C$ (به مرکز دسته‌ی آخر به اندازه‌ی طول دسته اضافه کنیم) هستند.

پس مثلاً اگر مرکز دسته‌ها ۱۳، ۱۷، ۲۱ و ۲۵ باشد (پس طول دسته‌ها ۴ است)، داده‌ی $x_۱ - C = ۱۳ - ۴ = ۹$ و $x_۴ + C = ۲۵ + ۴ = ۲۹$ را اضافه می‌کنیم و نمودار را به محور x ها متصل می‌کنیم.



با این کار، مساحت زیر مستطیلی و چندبر فراوانی مساوی می‌شود. شکل را ببینید:

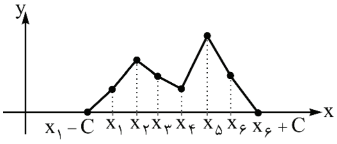


۷۸- گزینه‌ی ۳ سؤال گفته دو سر نمودار چندبر روی محور x ها است. پس خودشان چندبر را تکمیل کرده‌اند.

یعنی مساحت زیر چندبر و مستطیلی مساوی شده است: $S = S'$ و بنابراین $\frac{S}{S'} = ۱$ است.

۷۹- گزینه‌ی ۲

می‌دونید چرا روی محور x ها ۸ تا نقطه و روی محور y ها ۶ تا نقطه است؟ خب معلومه، چون ۲ تا دسته با فراوانی صفر در سروته اضافه کردیم. پس در واقع ۶ تا دسته داریم و ۲ تا x روی محور افقی مال ما نیستند:



۸۰- گزینه‌ی ۳ دسته‌ی اول به ما می‌گوید طول دسته $C = ۱۵ - ۹ = ۶$ است و مرکز دسته‌ی اول هم $x_۱ = \frac{۹ + ۱۵}{۲} = ۱۲$ است. برای تکمیل چندبر، باید یک دسته‌ی جدید با مرکز $x_۱ - C$ اضافه کنیم. پس طول اولین نقطه از چندبر، $x_۱ - C = ۱۲ - ۶ = ۶$ است و دومین نقطه هم خود $x_۱$ است. یعنی طول نقاط اول و دوم ۶ و ۱۲ می‌شود.

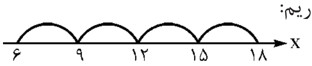
۸۱- گزینه‌ی ۴ از کم‌ترین و بیشترین داده، دامنه‌ی تغییرات را حساب می‌کنیم:

$$R = \max - \min = ۱۸ - ۶ = ۱۲$$

حالا می‌توانیم طول دسته را هم حساب کنیم:

$$C = \frac{R}{k} \xrightarrow[k=۴]{R=۱۲} C = \frac{۱۲}{۴} = ۳$$

حالا حدود دسته‌ها را روی محور می‌آوریم:



پس مرکز دسته‌ی آخر $\frac{۱۵ + ۱۸}{۲} = ۱۶/۵$ است و طول نقطه‌ی آخر نمودار چندبر، $x_k + C = ۱۶/۵ + ۳ = ۱۹/۵$ خواهد بود.

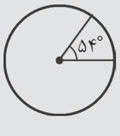
نمودار دایره‌ای

نمودار دایره‌ای (pie diagram) از یک دایره ساخته می‌شود! معمولاً برای داده‌های کیفی رسم می‌شود.

برای هر نوع، یک قاچ از دایره را در نظر می‌گیریم. زاویه‌ی این قاچ $\theta = \frac{f}{n} \times ۳۶۰^\circ$ است.

پس می‌توان گفت: زاویه در نمودار دایره‌ای = فراوانی نسبی × °۳۶۰

مثلاً اگر در ۲۰ داده، دسته‌ای با فراوانی ۳ باشد، زاویه‌ی متناظر آن $\theta = \frac{۳}{۲۰} \times ۳۶۰^\circ = ۵۴^\circ$ خواهد بود. در شکل هم این‌جوری می‌بینیم:



جمع زاویه‌های تمام دسته‌ها روی هم باید °۳۶۰ بشود.

کتاب درسی سؤال کرده: «آیا با تغییر شعاع دایره، تفسیر جدیدی از توزیع داده‌ها به دست می‌آید؟»

خب، نه. بزرگ و کوچک شدن نمودار دایره‌ای توزیع فراوانی را تغییر نمی‌دهد. در واقع توی نمودار دایره‌ای زاویه مهم است نه مساحت. ترتیب نواحی روی دایره هم اهمیتی ندارد.

بیشتر تست‌های کنکور از نمودار دایره‌ای هستند. چون تنها نموداری است که فرمول دارد. پس یک بار دیگر بگوییم: $\theta = \frac{f}{n}\times ۳۶۰^\circ$

۸۲- گزینه‌ی ۴ نمودار دایره‌ای، آن هم بدون ترتیب نواحی، برای داده‌های کیفی اسمی مناسب‌تر است.

۸۳- گزینه‌ی ۲ زاویه‌ی هر طبقه در نمودار دایره‌ای $\theta = \frac{f}{n}\times ۳۶۰^\circ$ بود. پس داریم:

$$\xrightarrow[n=۱۸۰]{f=۲۰} \theta = \frac{۲۰}{۱۸۰}\times ۳۶۰^\circ = ۲۰\times ۲^\circ = ۴۰^\circ$$

۸۴- گزینه‌ی ۴ کلاس انسانی ۵۰ نفرند و کل دانش‌آموزان سال چهارم روی هم $۱۸۰ = ۷۰ + ۶۰ + ۵۰$ نفر.

زاویه‌ی متناظر با گروه انسانی برابر است با:

$$\theta = \frac{f}{n}\times ۳۶۰^\circ \xrightarrow[n=۱۸۰]{f=۵۰} \frac{۵۰}{۱۸۰}\times ۳۶۰^\circ = ۱۰۰^\circ$$

۸۵- گزینه‌ی ۲ به فرمول یک نگاه بیندازید:

$$\theta = \frac{f}{n}\times ۳۶۰^\circ$$

توی این سؤال $\theta = ۴۵^\circ$ و $n = ۴۸$ است. پس داریم:

$$۴۵^\circ = \frac{f}{۴۸}\times ۳۶۰^\circ \xrightarrow{\text{طرفین و وسطین}} \frac{۴۸\times ۴۵^\circ}{۳۶۰^\circ} = f$$

تعداد کارمندان با مدرک کارشناسی برابر است با: $\Rightarrow f = \frac{۴۸}{۸} = ۶$

۸۶- گزینه‌ی ۴

هر جا در زندگی فراوانی تجمعی دیدیم، آن را منهای قبلی‌اش می‌کنیم تا به فراوانی غیرتجمعی برسیم.

الان فراوانی تجمعی نسبی دسته‌ی چهارم و پنجم را داریم. پس فراوانی نسبی دسته‌ی پنجم برابر است با: $\frac{f_۵}{n} = ۰/۴۰ - ۰/۲۸ = ۰/۱۲$

دقت کنید! فراوانی تجمعی دسته‌ی اول همان فراوانی معمولی‌اش است.

حالا فراوانی تجمعی دسته‌ی دوم را منهای تجمعی دسته‌ی اول می‌کنیم و فراوانی دسته‌ی دوم به دست می‌آید.

همین‌طور الی آخر. یعنی فراوانی دسته‌ی چهارم برابر است با فراوانی تجمعی دسته‌ی چهارم منهای فراوانی تجمعی دسته‌ی سوم.

سؤال زاویه‌ی مربوط به دسته‌ی پنجم را می‌خواست:

زاویه = فراوانی نسبی ×۳۶۰°

$$\theta_۵ = \frac{f_۵}{n}\times ۳۶۰^\circ = ۰/۱۲\times ۳۶۰^\circ = \frac{۱۲}{۱۰۰}\times ۳۶۰^\circ = ۴۳/۲^\circ$$

نگران حاصل ضربش نباشید! رقم یکان ۳۶×۱۲ می‌شود ۲. پس رقم آخر سمت راستِ جواب باید ۲ بشود و این فقط در ④ برقرار است.

۸۷- گزینه‌ی ۳ نمودار میله‌ای می‌گوید فراوانی دسته‌ها به ترتیب ۶،۸،۹،۴ است. پس تعداد کل می‌شود: $n = ۶ + ۸ + ۹ + ۴ = ۲۷$. ما زاویه‌ی دسته‌ی D را می‌خواهیم:

$$\theta_D = \frac{f_D}{n}\times ۳۶۰^\circ$$

$$\xrightarrow[n=۲۷]{f_D=۶} \theta_D = \frac{۶}{۲۷}\times ۳۶۰^\circ = \frac{۲}{۹}\times ۳۶۰^\circ = ۲\times ۴۰^\circ = ۸۰^\circ$$

۸۸- گزینه‌ی ۴ زاویه‌ی دسته‌ی B می‌شود: $\theta_B = \frac{f_B}{n}\times ۳۶۰^\circ$ فقط باید تعداد کل داده‌ها (تعداد کل افراد) را پیدا کنیم. خب اگر تعداد افراد A برابر x باشد، تعداد افراد B، C و D به ترتیب ۲x، ۳x و ۶x است. پس داریم: $n = x + ۲x + ۳x + ۶x = ۱۲x$

حالا می‌توانیم زاویه‌ی دسته‌ی B را حساب کنیم:

$$\theta_B = \frac{۲x}{۱۲x}\times ۳۶۰^\circ = \frac{۱}{۶}\times ۳۶۰^\circ = ۶۰^\circ$$

۸۹- گزینه‌ی ۳ جمع همه‌ی زاویه‌ها باید ۳۶۰° باشد. پس زاویه‌ی گروه سنی x را می‌توانیم حساب کنیم:

$$۹۰^\circ + ۱۱۰^\circ + ۱۰۰^\circ + \theta_x = ۳۶۰^\circ \Rightarrow \theta_x = ۶۰^\circ$$

حالا فراوانی گروه x را می‌خواهیم. به فرمول نگاه کنید:

$$\theta = \frac{f}{n}\times ۳۶۰^\circ$$

$$\xrightarrow[n=۱۵۰۰۰]{\theta_x=۶۰^\circ} ۶۰^\circ = \frac{f}{۱۵۰۰۰}\times ۳۶۰^\circ \xrightarrow{\div ۳۶۰^\circ} \frac{f}{۱۵۰۰۰} = \frac{۱}{۶}$$

$$\Rightarrow f = \frac{۱۵۰۰۰}{۶} = ۲۵۰۰$$

یعنی ۲۵۰۰ نفر در گروه سنی x هستند.

۹۰- گزینه‌ی ۱ جمع زاویه‌ها باید ۳۶۰° بشود:

$$۲۷^\circ + ۴۵^\circ + ۹۹^\circ + \alpha + ۵۴^\circ + ۱۸^\circ = ۲۴۳^\circ + \alpha = ۳۶۰^\circ \Rightarrow \alpha = ۱۱۷^\circ$$

۱۸ و ۲۷ روی هم می‌شود ۴۵؛ با اون یکی ۴۵ می‌شود ۹۰؛ با ۵۴ می‌شود ۱۴۴؛ با ۹۹ می‌شود ۲۴۳.

حالا در فرمول زاویه می‌گذاریم:

$$\alpha = \frac{f}{n}\times ۳۶۰^\circ \xrightarrow[n=۱۶۰]{\alpha=۱۱۷^\circ} ۱۱۷^\circ = \frac{f}{۱۶۰}\times ۳۶۰^\circ$$

$$\xrightarrow{\text{طرفین و وسطین}} f = \frac{۱۱۷\times ۱۶۰}{۳۶۰} = \frac{۱۱۷\times ۱۶}{۳۶} = \frac{\overset{۱۳}{\cancel{۱۱۷}}\times ۴}{\underset{۱}{\cancel{۹}}} = ۱۳\times ۴ = ۵۲$$

پس تعداد کارکنان با کد ۴ می‌شود ۵۲ تا.

۹۱- گزینه‌ی ۱ اول زاویه‌ی گروه O را پیدا می‌کنیم:

جمع زوایا = ۳۶۰°

$$۱۰۸^\circ + ۱۰۲^\circ + ۹۶^\circ + \theta_O = ۳۰۶^\circ + \theta_O = ۳۶۰^\circ \Rightarrow \theta_O = ۵۴^\circ$$

برای پیداکردن درصد گروه خونی O باید فراوانی نسبی‌اش را داشته باشیم. یک نگاه به فرمول زاویه کنید:

$$\theta = \frac{f}{n}\times ۳۶۰^\circ \xrightarrow{\theta_O=۵۴^\circ} ۵۴^\circ = \frac{f}{n}\times ۳۶۰^\circ \Rightarrow \frac{f}{n} = \frac{۵۴}{۳۶۰} = \frac{۳}{۲۰}$$

حالا درصد گروه خونی O برابر است با:

$$P = \frac{f}{n}\times ۱۰۰ = \frac{۳}{۲۰}\times ۱۰۰ = ۳\times ۵ = ۱۵$$

نیازی نبود $\frac{f}{n}$ را حساب کنیم. می‌شد درصد را مستقیماً از روی زاویه‌ها پیدا کرد. ببینید:

زاویه	درصد
۵۴°	?
۳۶۰°	۱۰۰

$$? = \frac{۱۰۰\times ۵۴}{۳۶۰} = \frac{۱۰\times ۵۴}{۳۶} = \frac{۱۰\times ۳}{۲} = ۱۵$$

پس ۱۵ درصد افراد در گروه O هستند.

-۹۲ گزینه‌ی ۳ اول درصد گروه O را پیدا می‌کنیم:

$$۲۴+۲۲/۵+۳۶+a=۱۰۰ = \text{جمع درصدها}$$

$$\Rightarrow ۸۲/۵+a=۱۰۰ \Rightarrow a=۱۷/۵$$

من از روش سؤال قبل می‌روم. یعنی تناسب مستقیم بین زاویه‌ها و درصدها:

زاویه	درصد
?	۱۷/۵
۳۶۰°	۱۰۰

$$\Rightarrow ?=\frac{۱۷/۵\times ۳۶۰^\circ}{۱۰۰}=\frac{۱۷۵\times ۳۶^\circ}{۱۰۰}$$

خب ماشین‌حساب که نداریم. ۱۰۰ را می‌نویسیم ۴×۲۵، بعدش ۱۷۵ را با ۲۵ و ۳۶ را با ۴ می‌زنیم:

$$\frac{\cancel{۱۷۵}^{۷}\times \cancel{۳۶}^{۹}}{\cancel{۲۵}_{۱}\times \cancel{۴}_{۱}}=۷\times ۹=۶۳$$

پس زاویه‌ی سطح مربوط به گروه O می‌شود ۶۳°.

-۹۳ گزینه‌ی ۳ برای زاویه‌ی مرکزی دسته‌ی وسط (یعنی دسته‌ی سوم)، باید فراوانی این دسته و فراوانی کل را داشته باشیم:

$$f_۳=x-۱۷ = \text{فراوانی دسته‌ی سوم}$$

$$n=۶۰ = \text{فراوانی کل}$$

فراوانی هر دسته می‌شد فراوانی تجمعی منهای فراوانی تجمعی دسته‌ی قبل. فراوانی تجمعی دسته‌ی آخر هم تعداد کل بود.

حالا زاویه‌ی دسته سوم برابر است با:

$$\theta_۳=\frac{f_۳}{n}\times ۳۶۰^\circ$$

$$\Rightarrow ۹۰^\circ=\frac{x-۱۷}{۶۰}\times ۳۶۰^\circ \Rightarrow \frac{۱}{۴}=\frac{x-۱۷}{۶۰}$$

$$\Rightarrow x-۱۷=۱۵ \Rightarrow x=۳۲$$

و در آخر، فراوانی مطلق دسته‌ی چهارم برابر است با: (اختلاف فراوانی تجمعی چهارم و سوم)

$$f_۴=۴۸-x=۴۸-۳۲=۱۶$$

شاخص‌های مرکزی ـ مد

شاخص‌های مرکزی، اعدادی هستند که محل تمرکز داده‌ها را نشان می‌دهند. مثلاً وضعیت نمرات کلاس شما را با معدل کلاس می شود فهمید. پس معدل، یک شاخص مرکزی است. کلاً شاخص‌های مرکزی اصلی ۳ تا هستند:

میانگین، میانه و مد

اگر از این روش‌های ضایع برای حفظ‌کردن دوست دارید، شاخص‌های مرکزی، میم دارند. مرکزی: میانگین، میانه، مد

اول برویم سراغ مد: الان چی مد شده؟! مد یعنی چیزی که زیاد می‌بینیم؛ یعنی فراوانی‌اش زیاد است.

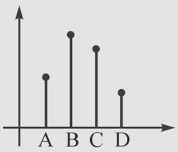
در داده‌های آماری، مد یعنی داده‌ای که بیشترین فراوانی مطلق را داشته باشد. (خود داده، نه فراوانی‌اش!)

مثلاً در

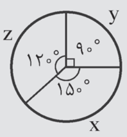
x	۱	۲	۳	۴
f	۳	۷	۴	۲

، داده‌ی ۲ بیشترین فراوانی را دارد. پس مد ۲ است.

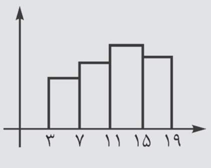
توی جدول فراوانی، مد داده‌ای است که بیشترین f را دارد. توی نمودار میله‌ای، داده‌ای است که بلندترین میله را دارد. در نمودار ستونی با دسته‌های هم‌طول، مرکز دسته‌ای است که بیشترین ارتفاع مستطیل را دارد. در نمودار دایره‌ای، داده‌ای است که بیشترین زاویه را دارد. در نمودار ساقه و برگ، مد داده‌ای است که از همه بیشتر دیده شود (یعنی بیشترین تکرار برگ را در بین سطرها دارد.)



$B = \text{مد}$



$x = \text{مد}$



$۱۳=\frac{۱۱+۱۵}{۲}= \text{مد}$

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

$۴۳ = \text{مد}$

مد به درد فروشنده‌های لباس می‌خورد. در انتخابات و رأی‌گیری‌ها هم مد (نظر اکثریت) مهم است. کتابتان گفته اگر جامعه‌ای دوتا مد (یا بیشتر) داشت دیگر مد اعتباری ندارد. یعنی جامعه یک‌دست نیست.

-۹۴ گزینه‌ی ۳ داده‌های این جدول

۱,۲,۴,۵,۵,۸,۹
۱۳,۱۴,۱۴,۱۴,۱۷,۱۷
۲۰,۲۰,۲۱,۲۴

هستند. حالا مد چی بود؟

داده‌ای که بیشترین فراوانی را دارد. چه عددی بیشتر از همه تکرار شده؟ خب ۱۴ (سه بار تکرار). پس مد می‌شود ۱۴.

-۹۵ گزینه‌ی ۱ در این داده‌ها، ۳ تا عدد ۷، یک عدد ۸، ۲ تا عدد ۲ و ۴ تا عدد ۱ داریم. پس مد (یعنی داده‌ای که بیشترین تکرار را دارد) ۱ است.

-۹۶ گزینه‌ی ۳ اول باید فراوانی‌های مطلق را حساب کنیم.

برای بار صدم، هر فراوانی تجمعی را منهای قبلی‌اش می‌کنیم تا به فراوانی مطلق برسیم.

$$۸,۱۳,۲۰,۳۲,۴۰ \Rightarrow ۸,\underbrace{۱۳-۸}_{۵},\underbrace{۲۰-۱۳}_{۷},\underbrace{۳۲-۲۰}_{۱۲},\underbrace{۴۰-۳۲}_{۸}$$

$\Rightarrow$

x	۱۱	۱۳	۱۵	۱۷	۱۹
f	۸	۵	۷	۱۲	۸

پس بیشترین فراوانی مطلق ۱۲ است و مد (یعنی داده‌ای که بیشترین فراوانی را دارد) می‌شود ۱۷.

شاخص‌های مرکزی ـ میانه

میانه داده‌ای است که در وسط می‌افتد. یعنی وقتی داده‌ها را مرتب می‌کنیم (از زیاد به کم یا از کم به زیاد)، تعداد داده‌های بعد از آن با تعداد داده‌های قبل از آن برابر است.

حالا دوتا وضعیت داریم:

الف اگر تعداد داده‌ها فرد باشد، دقیقاً یک داده وسط می‌افتد. در واقع وقتی n فرد باشد، داده‌ی $\frac{n+۱}{۲}$اُم در وسط است. مثلاً بین ۱۳ تا داده‌ی مرتب‌شده، میانه داده‌ی $(۷=\frac{۱۳+۱}{۲})$ هفتم است.

ب اگر تعداد داده‌ها زوج باشد، هیچ داده‌ای وسط نیست. معدل دوتای وسطی را به عنوان میانه در نظر می‌گیریم.

مثلاً در بین ۲۰ داده، معدل دهمی و یازدهمی را به عنوان میانه در نظر می‌گیریم.

پس در ۱,۲,۴,۷,۸، میانه برابر ۴ است. در بین ۱,۳,۷,۱۰ هم، میانه $۵=\frac{۳+۷}{۲}$ است.

- میانه فراوانی کل را تقریباً نصف می‌کند. نیمی از داده‌ها از میانه بیشترند و نیم دیگر از میانه کم‌ترند.
- در داده‌های ۱,۲,۳ میانه برابر ۲ است. در ۱,۲,۱۰۰ هم میانه ۲ است. کتاب گفته: «میانه نسبت به اندازه‌ی داده‌ها حساسیت نشان نمی‌دهد.»
- برای میانه حساب‌کردن، باید اول داده‌ها را مرتب کرد. اگر نمودار ساقه و برگ بدهند خیلی خوشحال می‌شویم، چون خودبه‌خود مرتب است.

۹۷- گزینه‌ی ۲ این داده‌ها ۲۱ تا هستند (هفت‌تا در سطر اول، شش‌تا در سطر دوم و هشت‌تا در سطر سوم)، پس داده‌ی $۱۱=\frac{۲۱+۱}{۲}$ (یازدهم) میانه است، یعنی عدد ۶۲.

۹۸- گزینه‌ی ۲ کتاب درسی گفته میانه شاخص مرکزی است که می‌تواند در مورد وضعیت جامعه استفاده شود. پس ما دنبال میانه هستیم تا وضع کلاس را متوجه شویم.

در این کلاس ۱۳ تا دانش‌آموز (تعداد داده‌ها) هستند و میانه، داده‌ی $(۷=\frac{۱۳+۱}{۲})$ هفتم است که می‌شود ۱۳.

۹۹- گزینه‌ی ۲ اول باید داده‌ها را مرتب کرد:

۷,۱۱,۱۵,۲۴,۳۵,۴۱,۵۹

- برای این‌که داده‌ای جا نیفتد، تعداد داده‌ها را کنترل کنید؛ در صورت سؤال ۷ تا داده داریم. الان هم ۷ تا داریم. پس چیزی گم‌وگور نشده است.

حالا بین ۷ تا داده، چهارمی میانه است: $\text{میانه} = x_۴ = ۲۴$

۱۰۰- گزینه‌ی ۳ خب این داده‌ها ۲۵ تا هستند، میانه می‌شود داده‌ی $(۱۳=\frac{۲۵+۱}{۲})$ سیزدهم.

حالا سیزدهمین داده چند است؟ بگذارید ببینیم! اول فراوانی‌های مطلق را پیدا می‌کنیم:

x	۶	۸	۱۰	۱۲	۱۴
f	۳	۴	۵	۶	۷

ما دنبال داده‌ی سیزدهم بودیم. بگذارید داده‌ها را دنبال هم بنویسیم:

۶,۶,۶,۸,۸,۸,۸,۱۰,۱۰,۱۰,۱۰,۱۰,۱۲,۱۲,۱۲,۱۲,۱۲,۱۲,۱۴,۱۴,۱۴,۱۴,۱۴,۱۴,۱۴

داده‌ی سیزدهم

پس میانه می‌شود ۱۲.

وقتی فراوانی تجمعی داریم، میانه داده‌ای است که فراوانی تجمعی‌اش $\frac{n+۱}{۲}$ باشد.

در این جدول، ۱۲ تا از داده‌ها ۶,۸,۱۰ بودند و سیزدهمین داده می‌شود اولین عدد ۱۲. پس میانه ۱۲ است.

۱۰۱- گزینه‌ی ۴ آیا میانه یکی از داده‌ها است؟

گفتیم اگر تعداد داده‌ها فرد باشد، میانه داده‌ی $\frac{n+۱}{۲}$اُم است.

مثلاً در ①، پنج‌تا داده داریم و سومی میانه است: $\text{میانه} = ۳$

در ② هم همین‌طور: $\text{میانه} = ۵$

اما وقتی تعداد داده‌ها زوج است، معدل دوتای وسطی را به عنوان میانه می‌گیریم.

در ③، شش‌تا داده داریم که دوتای وسط ۴,۴ هستند. پس:

$۴=\frac{۴+۴}{۲}=$ میانه (باز هم میانه یکی از داده‌ها شد.)

در ④ هم همین‌طور است: $۵=\frac{۴+۶}{۲}=$ میانه (در این‌جا میانه با هیچ‌کدام از داده‌ها مساوی نشد.)

پس میانه در چه صورت از داده‌ها نیست؟ وقتی تعداد داده‌ها زوج باشد و دو عدد وسطی هم مساوی نباشند.

۱۰۲- گزینه‌ی ۲ در بین هشت داده، میانه، معدل داده‌های چهارم و پنجم است.

سؤال گفته میانه با مد برابر است، پس معدل داده‌های چهارم و پنجم، همان داده‌ای است که دو بار تکرار شده، یعنی این‌جوری:

$$\begin{cases} a,b,c,d,d,e,f,g \\ d = \text{مد} = \text{میانه} \end{cases}$$

پس عضوهای چهارم و پنجم در داده‌های مرتب‌شده، مد هستند.

چارک‌ها

یادتان هست که گفتیم میانه درباره‌ی وضعیت جامعه نظر می‌دهد؟ حالا یک کم دقیق‌تر می‌شویم! در واقع می‌خواهیم به پراکندگی داده‌ها در اطراف میانه توجه کنیم. مثلاً ببینیم آیا داده‌ها بیشتر در اطراف میانه‌اند یا بیشتر در اطراف کم‌ترین و بیشترین داده. برای افزایش دقت، از چارک‌ها و نمودار جعبه‌ای استفاده می‌کنیم. چارک‌ها، فراوانی کل را به ۴ قسمت تقسیم می‌کنند.

مراحل کار این‌جوری است: اول داده‌ها را مرتب می‌کنیم و میانه را پیدا می‌کنیم. (وسطی)

حالا میانه را بی‌خیال می‌شویم و در دو طرف میانه، دوباره میانه‌ی هر قسمت را پیدا می‌کنیم.

به میانه‌ی نیمه‌ی اول می‌گوییم چارک اول و به میانه‌ی نیمه‌ی دوم می‌گوییم چارک سوم. خود میانه هم چارک دوم است.

مثلاً در داده‌های $۱۱,۱۲,۱۴,۱۴,۱۷,۱۸,\underline{\underline{۲۰}},۲۱,۲۲,۲۳,۲۵,۲۵,۲۷$ میانه می‌شود داده‌ی هفتم $(\frac{۱۳+۱}{۲})$، یعنی عدد ۲۰. حالا در نیمه‌ی اول داده‌ها $(۱۱,۱۲,۱۴,۱۴,۱۷,۱۸)$ میانه برابر است با $۱۴=\frac{۱۴+۱۴}{۲}$ و در نیمه‌ی دوم داده‌ها $(۲۱,۲۲,۲۳,۲۵,۲۵,۲۷)$ میانه می‌شود: $۲۴=\frac{۲۳+۲۵}{۲}$

پس ۲۰ = میانه = $Q_۲$، ۱۴ = چارک اول = $Q_۱$ و ۲۴ = چارک سوم = $Q_۳$.

چارک به زبان فرنگی می‌شود Quartile، برای همین با Q نشانش می‌دهند.

حالا می‌توانیم یک نمودار بکشیم و میانه و چارک‌ها را ببینیم! چند سؤال جلوتر ...

۱۰۳- گزینه‌ی ۴ خب اول باید مرتب کنیم:

۱۰,۱۲,۱۴,۱۵,۱۶,۱۸,۱۹,۲۳,۲۵,۲۷,۳۱,۳۲,۳۴,۴۱,۴۳

کنترل کردیم که ۱۵ تا داده هستند. همه را نوشتیم؟!

میانه می‌شود داده‌ی $(۸=\frac{۱۵+۱}{۲})$ هشتم، یعنی $Q_۲=۲۳$.

حالا نیمه‌ی اول داده‌ها، ۱۰,۱۲,۱۴,۱۵,۱۶,۱۸,۱۹ است. پس چارک اول، یعنی میانه‌ی نیمه‌ی اول می‌شود چهارمی: $Q_۱=۱۵$

۱۰۴- گزینه‌ی ۱ اول مرتب می‌کنیم:

۵,۷,۸,۹,۱۰,۱۱,۱۲,۱۴,۱۶,۱۷,۱۸,۱۹,۲۰,۲۱,۲۳

۱۵ تا بود، چیزی جا نیفتاده؟!

حالا باید میانه را پیدا کنیم. توی ۱۵ تا داده، میانه هشتمی است:

$Q_۲=x_۸=۱۴$

در نیمه‌ی اول و دوم هم میانه را پیدا می‌کنیم:

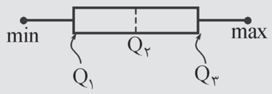
$$\underbrace{۵,۷,۸,\underline{۹},۱۰,۱۱,۱۲}_{Q_۱=۹},\underset{\uparrow\atop Q_۲}{۱۴},\underbrace{۱۶,۱۷,۱۸,\underline{۱۹},۲۰,۲۱,۲۳}_{Q_۳=۱۹}$$

اختلاف چارک اول و سوم هم می‌شود: $Q_۳-Q_۱=۱۹-۹=۱۰$

به $Q_۳-Q_۱$ می‌گویند دامنه‌ی تغییرات بین چارکی!

نمودار جعبه‌ای

نمودار جعبه‌ای یک تصویر است که ۵ تا مقدار را نشان می‌دهد: داده‌ی مینیمم، چارک اول، میانه (چارک دوم)، چارک سوم و داده‌ی ماکسیمم. قیافه‌اش این‌شکلی است:

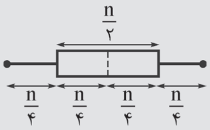


دنباله‌ی سمت چپ یعنی فاصله‌ی داده‌ی مینیمم تا چارک اول. ($\frac{۱}{۴}$ داده‌ها در این قسمت‌اند.)

درون جعبه یعنی از چارک اول تا چارک سوم. ($\frac{۱}{۲}$ داده‌ها توی جعبه‌اند.)

دنباله‌ی سمت راست هم یعنی از چارک سوم تا داده‌ی ماکسیمم. ($\frac{۱}{۴}$ داده‌ها در دنباله‌ی سمت راست هستند.)

توزیع فراوانی توی نمودار جعبه‌ای را ببینید:



کتاب درسی درصدها را پرسیده. در هر یک از دنباله‌ها ۲۵ درصد داده‌ها قرار دارند. در داخل جعبه هم ۵۰ درصد داده‌ها هستند.

۱۰۵- گزینه‌ی ۲ برای نمودار جعبه‌ای به چارک‌ها احتیاج داریم.

پس داده‌ها را مرتب می‌کنیم:

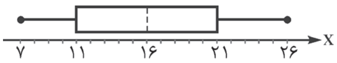
$$\underbrace{۷,۹,۱۰,۱۱,۱۲,۱۳,۱۴}\,,۱۶,\underbrace{۱۸,۱۹,۲۰,۲۱,۲۲,۲۳,۲۶}$$ (۱۵ تا هستند)

چارک دوم یعنی همون میانه، داده‌ی هشتم است: $Q_۲=۱۶$

چارک اول، میانه‌ی هفت‌تای اول است: $Q_۱=۱۱$

چارک سوم هم، میانه‌ی هفت‌تای دوم است: $Q_۳=۲۱$

پس نمودار جعبه‌ای این‌جوری است:



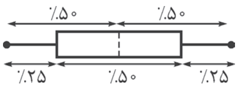
$۴=۷-۱۱$ = طول دنباله‌ی چپ

۱۰۶- گزینه‌ی ۳ بگذارید از روی این نمودار جعبه‌ای بخوانیم.

داده‌ی مینیمم ۱ بوده؛ داده‌ی ماکسیمم هم ۱۲ بوده؛ چارک اول ۸، چارک دوم ۹ و چارک سوم ۱۱ است. پس فقط چارک سوم را درست گفته بود. راستی دامنه‌ی تغییرات می‌شود: $R=۱۲-۱=۱۱$

بلندتربودن دنباله‌ی سمت چپ در این نمودار می‌گوید داده‌های کم‌تر از چارک اول، پراکندگی و تنوع بیشتری دارند.

۱۰۷- گزینه‌ی ۳ یک بار دیگر به درصدهای هر قسمت از نمودار جعبه‌ای نگاه کنید:



پس:

۵۰ درصد داده‌ها بیشتر از میانه‌اند. ۵۰ درصد داده‌ها کم‌تر از میانه‌اند.

۵۰ درصد داده‌ها درون جعبه (بین $Q_۱$ و $Q_۳$) هستند.

۲۵ درصد داده‌ها کم‌تر از $Q_۱$ و ۲۵ درصد داده‌ها بیشتر از $Q_۳$ هستند.

۷۵ درصد داده‌ها بیشتر از $Q_۱$ و ۷۵ درصد داده‌ها کم‌تر از $Q_۳$ هستند.

پس در این سؤال، سهم داده‌های کم‌تر از ۱۴ (یعنی کم‌تر از $Q_۱$)، ۲۵ درصد است و ۷۵ درصد داده‌ها بیشتر از ۱۴ هستند. یعنی ③ غلطه.

میانگین

میانگین یعنی همان معدل داده‌ها، معدل چه‌طوری حساب می‌شود؟ خب جمع داده‌ها تقسیم بر تعداد آن‌ها.

پس می‌توان گفت:

$$\text{میانگین}=\overline{x}=\frac{x_1+x_2+\ldots}{n}=\frac{\text{جمع}}{\text{تعداد}}$$

از نظر آمار، میانگین عددی است که در مرکز داده‌ها قرار دارد.

فعلاً فقط میانگین ساده بلدیم. برای حالت‌هایی که فراوانی هم داشته باشیم بعداً یک کارهایی می‌کنیم.

راستی $\overline{x}$ را می‌خوانند ایکس‌بار.

۱۰۸- **گزینه‌ی ۳** میانگین می‌شود جمع، تقسیم بر تعداد. همین!

$$\text{میانگین}=\frac{\text{جمع}}{\text{تعداد}}\Rightarrow\overline{x}=\frac{۱۶+۱۵+۱۴+۱۴+۱۲}{۵}=\frac{۷۱}{۵}=۱۴/۲$$

۱۰۹- **گزینه‌ی ۲** باز هم باید جمع را بر تعداد تقسیم کنیم. حواستان باشد که داده‌ها $۸/۰,۸/۰,۸/۱,\ldots$ هستند. (یک رقم اعشار دارند.)

جمع داده‌ها برابر است با:

جمع $=۸/۰+\cdots+۸/۷+۹/۰+\cdots+۹/۵+۱۰/۱+\cdots+۱۰/۲$

دقت می‌کنیم که ۸ تا توی سطر اول، ۸ تا توی سطر دوم و ۴ تا توی سطر سوم داریم. یعنی روی هم ۲۰ تا داده.

حساب‌کردن این مجموع خیلی هم ساده نیست! به خصوص که اعشار هم داریم. پس بیایید یک حرکتی بزنیم!!

اول اعشارها را در هر سطر جمع کنیم:

سطر اول: $۰+۰+۰/۱+۰/۲+۰/۲+۰/۵+۰/۶+۰/۷=۲/۳$

سطر دوم: $۰+۰/۱+۰/۲+۰/۳+۰/۳+۰/۴+۰/۵+۰/۵=۲/۳$

سطر سوم: $۰/۱+۰/۱+۰/۲+۰/۲=۰/۶$

پس جمع اعشارها روی هم می‌شود: $۲/۳+۲/۳+۰/۶=۵/۲$

بعد اعداد صحیح را جمع می‌کنیم:

سطر اول: $۸\times۸=۶۴$

سطر دوم: $۸\times۹=۷۲$

سطر سوم: $۴\times۱۰=۴۰$

پس جمع صحیح‌ها روی هم می‌شود: $۶۴+۷۲+۴۰=۱۷۶$

خلاصه، جمع کل می‌شود: $۱۷۶+۵/۲=۱۸۱/۲$ و میانگین برابر است با:

$$\overline{x}=\frac{۱۸۱/۲}{۲۰}=\frac{۱۸۰}{۲۰}+\frac{۱/۲}{۲۰}=۹+۰/۰۶=۹/۰۶$$

در آخرش هم یک کارایی کردیم. ۱۸۱/۲ را نوشتیم ۱۸۰+۱/۲ و بعد هر کدامشان را بر ۲۰ تقسیم کردیم.

روش میانگین حدسی

اگر حساب‌کردن مجموع داده‌ها سخت بود، یک میانگین حدس بزنید. بعد همه‌ی داده‌ها را منهای این حدس کنید و از عددهای جدید معدل بگیرید (آسان‌تر می‌شود) مثلاً برای میانگین اعداد ۱۶,۱۵,۱۴,۱۴,۱۲ حدس می‌زنیم میانگین ۱۴ باشد. همه‌ی اعداد را منهای ۱۴ می‌کنیم:

$$۱۶,۱۵,۱۴,۱۴,۱۲\xrightarrow[\text{کم کنیم}]{\text{از همه ۱۴ تا}}۲,۱,۰,۰,-۲\xrightarrow{\text{معدل اعداد جدید}}\frac{۲+۱+۰+۰-۲}{۵}=\frac{۱}{۵}$$

حالا میانگین اعداد خودمان می‌شود $۱۴+\frac{۱}{۵}$ یعنی ۱۴/۲. پس به حدس اولیه، مقداری که به دست آمده را اضافه می‌کنیم.

حدس اولیه هر عددی باشد اشکالی ندارد. مثلاً اگر توی همین مثال ۱۵ را حدس بزنیم داریم:

$$۱۶,۱۵,۱۴,۱۴,۱۲\xrightarrow{\text{از همه ۱۵ تا کم کنیم}}۱,۰,-۱,-۱,-۳\xrightarrow{\text{معدل اعداد جدید}}\frac{۱+۰-۱-۱-۳}{۵}=\frac{-۴}{۵}$$

پس میانگین اعداد خودمان $۱۵+\frac{(-۴)}{۵}$ است، یعنی همان ۱۴/۲.

به زبان ریاضی: $\overline{x}=y+A$

y میانگین حدسی ما است و A هم میانگین داده‌های جدید (x_i-y) است.

$$A=\frac{(x_1-y)+(x_2-y)+\cdots+(x_n-y)}{n}$$

کتابتان به این روش گفته «روش سریع». در کنکور سراسری هم این اسم را به کار برده‌اند!

۱۱۰- **گزینه‌ی ۲** وقتی امتحان جبرانی را بدهد، نمراتش ۱۹, ۱۸/۵, x, ۱۸, ۱۹, ۱۷ هستند. (x نمره‌ی امتحان جبرانی است.)

حالا معدلش را حساب کنیم: $\overline{x}=\frac{۱۷+۱۹+۱۸+x+۱۸/۵+۱۹}{۶}$

این معدل باید ۱۸/۵ شود: $\frac{۱۷+۱۹+۱۸+x+۱۸/۵+۱۹}{۶}=۱۸/۵$

اما حوصله‌ی جمع‌کردن این عددها را نداریم. پس باز هم می‌رویم به سراغ روش حدسی:

$$\xrightarrow[\text{۱۸/۵ می‌کنیم.}]{\text{همه را منهای}}-۱/۵+۰/۵-۰/۵+(x-۱۸/۵)+۰+۰/۵=۰$$

$$\Rightarrow(x-۱۸/۵)-۱=۰\Rightarrow x=۱۹/۵$$

یعنی باید ۱۹/۵ بشود.

از اولش به نمره‌ی ۱۸/۵ نگاه می‌کنیم.

نمره‌ی ۱۷، ۱/۵ نمره کم دارد؛ نمره‌ی ۱۹، ۰/۵ نمره اضافه دارد؛

نمره‌ی ۱۸، ۰/۵ نمره کم دارد؛ نمره‌ی ۱۸/۵ که درست است؛

نمره‌ی ۱۹ هم ۰/۵ نمره اضافه دارد؛ این کم و زیادها روی هم می‌شود: ۰/۵+۰+۰/۵−۰/۵+۰/۵−۱/۵− یعنی ۱−. پس ۱ نمره کم دارد. یعنی باید ۱ نمره را جبران کند و ۱۹/۵ بشود.

۱۱۱- گزینه‌ی ۲ الان که مد نمی‌بینیم، چون فراوانی هیچ داده‌ای از بقیه بیشتر نیست.

پس x باید یکی از همین اعداد باشد تا فراوانی‌اش بشود ۲ و مد داشته باشیم. یعنی $x=۷۰$ یا $x=۸۰$. خب هر دوتا را امتحان می‌کنیم:

$$x=۷۰ \Rightarrow \{۶۰,۷۰,۷۰,۸۰,۱۱۰\} \Rightarrow \text{مد} = \text{میانه} = ۷۰$$

$$\text{میانگین} = \frac{۶۰+۷۰+۷۰+۸۰+۱۱۰}{۵} = \frac{۳۹۰}{۵} = ۷۸ \quad \text{(این‌که نشد)}$$

$$x=۸۰ \Rightarrow \{۶۰,۷۰,۸۰,۸۰,۱۱۰\} \Rightarrow \text{میانه} = \text{مد} = ۸۰$$

$$\text{میانگین} = \frac{۶۰+۷۰+۸۰+۸۰+۱۱۰}{۵} = \frac{۴۰۰}{۵} = ۸۰ \quad \text{(این خوبه)}$$

پس اگر $x=۸۰$ باشد، اندازه‌ی میانگین و میانه و مد مساوی می‌شوند.

۱۱۲- گزینه‌ی ۲ حدس بزنیم ۱۷۶۰ باشد:

$$۱۷۵۰,۱۷۶۰,۱۷۶۶,۱۷۵۵,۱۷۸۰ \xrightarrow[\text{کم می‌کنیم}]{\text{از همه ۱۷۶۰ تا}} -۱۰,۰,۶,-۵,۲۰$$

$$\xrightarrow{\text{معدل اعداد جدید}} A = \frac{-۱۰+۰+۶-۵+۲۰}{۵} = \frac{۱۱}{۵} = ۲/۲$$

$$\Rightarrow \text{میانگین} = \overline{x} = y + A = ۱۷۶۰ + ۲/۲ = ۱۷۶۲/۲$$

۱۱۳- گزینه‌ی ۲ باید اول چارک‌ها را پیدا کرد. پس برویم دنبال مرتب‌کردن داده‌ها:

$$۱۲,۱۴,\underbrace{۱۴,۱۵},۱۶,۱۸,۲۰,۲۰,\underbrace{۲۱,۲۴},۲۵,۲۶ \quad \text{(۱۲ تا داده داریم)}$$

میانه می‌شود معدل ششمی و هفتمی. یعنی: $Q_۲ = \frac{۱۸+۲۰}{۲} = ۱۹$

$Q_۱$ یعنی میانه‌ی شش‌تای اول. پس داریم: $Q_۱ = \frac{۱۴+۱۵}{۲} = ۱۴/۵$

$Q_۳$ هم میانه‌ی شش‌تای دوم است: $Q_۳ = \frac{۲۱+۲۴}{۲} = ۲۲/۵$

پس داده‌های بیشتر از چارک اول و کم‌تر از چارک سوم (یعنی داده‌های توی جعبه)، عبارت‌اند از: ۱۵,۱۶,۱۸,۲۰,۲۰,۲۱

ما میانگین این‌ها را می‌خواهیم:

$$\overline{x} = \frac{\overbrace{۱۵+۱۶}^{۳۱}+\overbrace{۱۸+۲۰}^{۳۸}+\overbrace{۲۰+۲۱}^{۴۱}}{۶} = \frac{۱۱۰}{۶} = \frac{۵۵}{۳} = ۱۸\frac{۱}{۳} \simeq ۱۸/۳۳$$

اعشارِ عدد $\frac{۱}{۳}$ به صورت ...۰/۳۳۳۳ است.

اعشار $\frac{۲}{۳}$ هم ...۰/۶۶۶۶ است.

گزینه‌ها می‌گویند میانگین هجده و خُرده‌ای است. پس ما $y=۱۸$ را حدس می زنیم:

$$۱۵,۱۶,۱۸,۲۰,۲۰,۲۱ \xrightarrow{\text{از همه ۱۸ تا کم کنیم}} -۳,-۲,۰,۲,۲,۳$$

$$\xrightarrow[\text{داده‌های جدید}]{\text{میانگین}} A = \frac{-۳-۲+۰+۲+۲+۳}{۶} = \frac{۲}{۶} = \frac{۱}{۳} \simeq ۰/۳۳$$

$$\Rightarrow \text{میانگین} = \overline{x} = y + A = ۱۸ + ۰/۳۳ = ۱۸/۳۳$$

۱۱۴- گزینه‌ی ۲ سؤال گفته:

$$\text{داده‌ها} \xrightarrow{\text{همه را منهای ۱۲ کردیم}} -۴,-۳,-۱,۲,۳$$

قرار بود وقتی همه‌ی داده‌ها را منهای حدس کردیم، معدل عددهای جدید را بگیریم:

$$\xrightarrow[\text{داده‌های جدید}]{\text{میانگین}} A = \frac{-۴-۳-۱+۲+۳}{۵} = \frac{-۳}{۵} = -۰/۶$$

پس میانگین واقعی برابر است با: $\overline{x} = y + A = ۱۲ + (-۰/۶) = ۱۱/۴$

۱۱۵- گزینه‌ی ۱ برویم چارک اول و سوم را پیدا کنیم. این داده‌ها ۱۸ تا هستند. (برگ‌ها را شمردیم.)

پس در سمت چپ و راست هر کدام ۹ تا داده داریم. توی ۹ تای اول، پنجمی وسط است، یعنی: $Q_۱ = x_۵ = ۳۸$. در بین ۹ تای دوم هم پنجمی وسط است. پس $Q_۳ = x_{۱۴} = ۵۲$. ببینید:

$$۳۱,۳۴,۳۵,۳۷,\underbrace{۳۸}_{Q_۱},۳۸,۳۹,۴۰,۴۰ \vdots ۴۴,۴۵,۴۵,۴۶,\underbrace{۵۲}_{Q_۳},۵۳,۵۶,۵۶,۵۷$$

حالا سؤال گفته داده‌های بیشتر از $Q_۳$ و کم‌تر از $Q_۱$ را حذف کنیم. پس این‌ها می‌مانند: (ده‌تا داده هستند) ۳۸,۳۸,۳۹,۴۰,۴۰,۴۴,۴۵,۴۵,۴۶,۵۲

گزینه‌ها می‌گویند میانگین حدود ۴۳ است. پس ما هم ۴۳ حدس می‌زنیم!

$$\xrightarrow[\text{۴۳ می‌کنیم}]{\text{همه را منهای}} -۵,-۵,-۴,-۳,-۳,۱,۲,۲,۳,۹$$

$$\xrightarrow{\text{میانگین اعداد جدید}} A = \frac{-۵-۵-۴-۳-۳+۱+۲+۲+۳+۹}{۱۰} = \frac{-۳}{۱۰}$$

$$\Rightarrow \overline{x} = y + A = ۴۳ - \frac{۳}{۱۰} = ۴۲/۷$$

اگر ۴۲ را هم حدس می‌زدیم، همین می‌شد!

۱۱۶- گزینه‌ی ۳

داده‌های مرتب‌شده به صورت: $a,b,c,\underbrace{\overset{\text{چهارم}}{d},\overset{\text{پنجم}}{e}}_{\text{میانه}=\frac{d+e}{۲}},f,g,h$ بوده‌اند.

چون میانه، معدل چهارمی و پنجمی است، پس تعداد داده‌ها ۸ تا بوده؛ یعنی $n=۸$.

پس میانگین برابر است با: $\overline{x} = \frac{\text{جمع}}{\text{تعداد}} = \frac{۳۶۰}{۸} = ۴۵$

میانگین در دنباله‌ی حسابی

دنباله‌ی حسابی، دنباله‌ای است که اختلاف عضوهای متوالی‌اش مقدار ثابتی باشد. مثلاً ۴,۷,۱۰,۱۳,۱۶ یک دنباله‌ی حسابی است (چون ۳ تا ۳ تا زیاد می‌شود) اما ۴,۷,۱۱ دنباله‌ی حسابی نیست.

اگر داده‌های آماری دنباله‌ی حسابی بسازند (یعنی هر داده به اندازه‌ی عدد ثابتی بیشتر از قبلی باشد) آن وقت شاخص‌های مرکزی را می‌توانیم سریع پیدا کنیم.

در دنباله‌ی حسابی میانگین و میانه با هم مساوی‌اند و داریم: $Q_۲ = \overline{x} = \frac{\max + \min}{۲}$

یعنی میانگین و میانه، هر دو، برابر معدل داده‌های اول و آخرند.

پس مثلاً در داده‌های ۴,۷,۱۰,۱۳,۱۶ داریم: $Q_۲ = \overline{x} = \frac{۴+۱۶}{۲} = ۱۰$

۱۱۷- گزینه‌ی ۴ اگر دقت کنید این داده‌ها ۶ تا ۶ تا زیاد شده‌اند. یعنی دنباله‌ی حسابی داریم، بنابراین:

$$\overline{x}=\frac{max+min}{۲}=\frac{۷+۶۱}{۲}=\frac{۶۸}{۲}=۳۴$$

بنابراین میانگین برابر ۳۴ است.

۱۱۸- گزینه‌ی ۲ داده‌ها دنباله‌ی حسابی می‌سازند (۳ تا ۳ تا زیاد می شوند)، پس میانگین برابر است با:

$$\overline{x}=\frac{max+min}{۲}=\frac{\overbrace{(a+۳)}^{\text{اولی}}+\overbrace{(a+۱۰۵)}^{\text{آخری}}}{۲}=\frac{۲a+۱۰۸}{۲}=a+۵۴$$

۱۱۹- گزینه‌ی ۴ هفت عدد متوالی، دنباله‌ی حسابی می‌سازند.

گفتیم در دنباله‌ی حسابی، میانگین با میانه برابر است. پس اگر میانگین از میانه کم شود، صفر می‌شود.

مثلاً ۱,۲,۳,۴,۵,۶,۷ را ببینید: $\overline{x}=Q_۲=۴ \Rightarrow Q_۲-\overline{x}=۰$

۱۲۰- گزینه‌ی ۴ هر دو گروه داده‌ها، دنباله‌ی حسابی می‌سازند. پس میانگین و میانه می‌شود معدل اولی و آخری:

$$a,a+۱,a+۲,a+۳,a+۴ \Rightarrow \overline{x}=\frac{a+a+۴}{۲}=\frac{۲a+۴}{۲}=a+۲$$

$$۲a+۳,۲a+۵,۲a+۷,۲a+۹,۲a+۱۱$$

$$\Rightarrow Q_۲=\frac{۲a+۳+۲a+۱۱}{۲}=\frac{۴a+۱۴}{۲}=۲a+۷$$

حال سؤال پرسیده میانگین اولی‌ها از میانه‌ی دومی‌ها چه‌قدر کم‌تر است؟

یعنی $a+۲$ از $۲a+۷$ چه‌قدر کم‌تر است؟ خب $a+۵$ تا.

۱۲۱- گزینه‌ی ۱

دامنه‌ی تغییرات $R=max-min=۱۸-۷=۱۱$ است. میانه در بین این ۷ داده، چهارمی است: $x_۴=Q_۲=۱۵$

مد هم ۷ است (چون از همه بیشتر تکرار شده است.)

میانگین را هم حساب کنیم:

$$\overline{x}=\frac{\text{جمع}}{\text{تعداد}}=\frac{\overbrace{۱۸+۱۷}^{۳۵}+\overbrace{۱۶+۱۵+۱۱}^{۴۲}+\overbrace{۷+۷}^{۱۴}}{۷}=۵+۶+۲=۱۳$$

پس بیشترین اختلاف بین میانه و مد است. (۱۵ و ۷ بیشترین اختلاف را دارند.)

۱۲۲- گزینه‌ی ۱ الان دامنه‌ی تغییرات برابر است با:

$$R=max-min=۱۲-۳=۹$$

میانه هم معدل دوتای وسطی است: $Q_۲=\frac{۷+۹}{۲}=۸$

و میانگین برابر است با: $\overline{x}=\frac{\text{جمع}}{\text{تعداد}}=\frac{۳+۷+۹+۱۲}{۴}=\frac{۳۱}{۴}$

حالا اگر ۳ را به ۴ تبدیل کنیم، دامنه‌ی تغییرات و میانگین هم ۸ می‌شوند.

شما کنترل کنید که با تغییر ۷ یا ۹، R ثابت می‌ماند و اگر بخواهیم میانه هم ۹ بشود، به میانگین ۹ نمی‌رسیم.

اگر ۱۲ را تغییر بدهیم، میانه ۸ می‌ماند پس باید ۱۲ را به ۱۱ تبدیل کنیم که به میانگین ۸ نمی‌رسیم.

۱۲۳- گزینه‌ی ۳ قرار است ۷,۹,۱۰,۱۰,۱۴ به ۷,۱۰,۱۰,۱۰,۱۴ تبدیل بشوند.

خب دامنه‌ی تغییرات که فرقی نمی‌کند. در هر دو حالت $R=۱۴-۷=۷$ است.

میانه هم در هر دو حالت $Q_۲=۱۰$ است.

مد هم در هر دو حالت ۱۰ است.

اما میانگین در حالت اول $\frac{۷+۹+۱۰+۱۰+۱۴}{۵}$ و در حالت دوم $\frac{۷+۱۰+۱۰+۱۰+۱۴}{۵}$ است. پس در حالت دوم به اندازه‌ی $\frac{۱}{۵}$ بیشتر از حالت اول می‌شود. یعنی فقط میانگین فرق می‌کند.

۱۲۴- گزینه‌ی ۲ منحنی نرمال یک توزیع فراوانی متقارن، شبیه زنگوله است. (متن صفحه‌ی ۱۲۹ کتاب درسی را می‌خوانیم!)

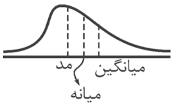
از تقارن آن نتیجه می‌شود میانه و میانگین مساوی‌اند.

چون نقطه‌ی ماکسیمم هم دارد، مد هم برابر میانگین است.

۱۲۵- گزینه‌ی ۳ وقتی میانگین از میانه کم‌تر است، داده‌های کم‌تر از میانه تنوع و پراکندگی بیشتری دارند و بیشتر مساحت نمودار به طرف چپ جابه‌جا می‌شود.

برعکس وقتی میانه از میانگین کم‌تر است، بیشتر مساحت نمودار در سمت راست آن است. در هر حال، نقطه‌ی ماکسیمم نمودار مربوط به مد است.

مثلاً در این شکل داریم:



۱۲۶- گزینه‌ی ۲ صورت سؤال چندتا چیز را گفته:

اولاً ۶ تا داده بود؛ ثانیاً در نیمه‌ی بعد از میانه، یعنی چهارمی و پنجمی و ششمی اعداد متوالی بوده‌اند؛ ثالثاً دامنه‌ی تغییرات و میانگین ۵ و ۱۵ بودند و رابعاً یک مد با فراوانی ۳ داشتیم.

از این حرف‌ها در کنار هم، می‌فهمیم داده‌ها باید $x,x,x,n,n+۱,n+۲$ یا $x,n,n,n,n+۱,n+۲$ باشند. هر کدامشان را بررسی می‌کنیم:

الف در حالت $x,x,x,n,n+۱,n+۲$ میانگین می‌شود:

$$\overline{x}=\frac{\text{جمع}}{\text{تعداد}}=\frac{۳x+۳n+۳}{۶}$$

دامنه‌ی تغییرات هم می‌شود: $R=n+۲-x$

پس باید داشته باشیم:

$$\begin{cases}\frac{۳x+۳n+۳}{۶}=۱۵\\ n+۲-x=۵\end{cases}$$

یک کم ساده‌ترش کنیم:

$$\begin{cases}\frac{۳(x+n+۱)}{۶}=۱۵ \xrightarrow{\text{طرفین و وسطین}} x+n=۲۹\\ n-x=۳\end{cases}$$

از حل این دستگاه داریم: $n=۱۶, x=۱۳$، بنابراین مد می‌شود $x=۱۳$. یعنی ② جواب است.

ب در حالت $x,n,n,n,n+۱,n+۲$ دامنه‌ی تغییرات $n+۲-x$ است و میانگین $\frac{x+۵n+۳}{۶}$ خواهد بود. پس داریم:

$$\begin{cases}\frac{x+۵n+۳}{۶}=۱۵ \Rightarrow x+۵n=۸۷ \xrightarrow{\text{حل}} n=۱۵\ ,\ x=۱۲\\ n+۲-x=۵ \Rightarrow n-x=۳\end{cases}$$

پس مقدار مد می‌شود $n=۱۵$ که در گزینه‌ها نیست.

میانگین در مرکز داده‌ها است

یادتان هست که گفتیم میانگین، محل تمرکز داده‌ها است. یعنی داده‌ها در اطراف میانگین متمرکز هستند. حالا به زبان ریاضی می‌گوییم:

داده‌ها همان مقداری که از میانگین اضافی دارند، همان مقدار را هم از میانگین کم دارند.

یعنی اگر هر داده را منهای میانگین کنیم، عددهایی مثبت و منفی به دست می‌آیند که مجموعشان صفر است. ببینید:

میانگین ۱۴,۱۱,۱۰,۸,۷ برابر ۱۰ است. حالا همه را منهای ۱۰ می‌کنیم: $۷,۸,۱۰,۱۱,۱۴ \xrightarrow{-۱۰} -۳,-۲,۰,۱,۴$

به این اعداد می‌گوییم انحراف از میانگین. مجموع این اعداد همیشه صفر است: $-۳-۲+۰+۱+۴=۰$

پس داریم: $(x_۱-\overline{x})+(x_۲-\overline{x})+\cdots+(x_n-\overline{x})=۰$

۱۲۷- گزینه‌ی ۴ مجموع $(x_i-\overline{x})$ها صفر است.

۱۲۸- گزینه‌ی ۳ در سؤال قبل گفتیم جمع همه‌ی انحراف از میانگین‌ها می‌شود صفر.

پس داریم: $(x_۱-۸)+(x_۲-۸)+\cdots+(x_۷-۸)+(۵-۸)=۰$

داده‌ی آخری که یادتان نرفت! ۵ هم هست.

اما چیزی که سؤال از ما خواسته $(۸-۵)$ را ندارد. پس داریم:

$$(x_۱-۸)+(x_۲-۸)+\cdots+(x_۷-۸)+(-۳)=۰$$

$$\Rightarrow (x_۱-۸)+(x_۲-۸)+\cdots+(x_۷-۸)=۳$$

۱۲۹- گزینه‌ی ۲ باید مجموع این مقادیر صفر باشد. پس داریم:

$$-۴+a-۱+۳+۱+۲=۰ \Rightarrow a-۵+۶=۰ \Rightarrow a=-۱$$

۱۳۰- گزینه‌ی ۲ به نظر شما وقتی کم‌ترین داده را منهای میانگین می‌کنیم، کدام‌یک از این عددها می‌شود؟

خب چون کم‌ترین داده از همه کم‌تر است (عجب کشف مهمی!) پس کم‌ترین جواب را می‌دهد؛ یعنی ۳- ، پس داریم:

$$\min-\overline{x}=۱۷-\overline{x}=-۳ \Rightarrow \overline{x}=۲۰$$

یعنی همه‌ی داده‌ها را منهای ۲۰ کرده بودیم. پس داده‌ها عبارت‌اند از:

$$۲۳,۲۳,۲۰,۱۹,۱۸,۱۷ \xrightarrow{\text{از همه ۲۰ تا کم شده}} ۳,۳,۰,-۱,-۲,-۳$$

میانه می‌شود معدل دوتای وسطی: $Q_۲=\frac{x_۴+x_۳}{۲}=\frac{۲۰+۱۹}{۲}=۱۹/۵$

۱۳۱- گزینه‌ی ۴ این چیزی که سؤال گفته یعنی باید یک x به این‌ها اضافه کنیم و داشته باشیم:

میانگینِ (x,۸,۷,۶,۳) = میانگینِ (۸,۷,۶,۳) +۱

خب میانگین که بلدیم. جمع، تقسیم بر تعداد بود:

$$\frac{x+۸+۷+۶+۳}{۵}=\frac{۸+۷+۶+۳}{۴}+۱$$

این معادله را در سه‌سوت حل می‌کنیم:

$$\Rightarrow \frac{x+۲۴}{۵}=\frac{۲۴}{۴}+۱=۷ \Rightarrow x+۲۴=۳۵ \Rightarrow x=۱۱$$

لازم هست تأکید کنم که اولش ۴ تا بودند، بعد که x آمد شدند ۵ تا؟! مخرج‌ها را درست گذاشتید؟! حتماً ...

۱۳۲- گزینه‌ی ۲ باید میانگین هر دو دسته را حساب بکنیم و مساویِ هم بگذاریم دیگه:

میانگین (۱-,۲-,۳-,۴-,۵-) = میانگین (۹۵,۹۶,۹۷,۹۸,a)

$$\xrightarrow[\text{بر تعداد}]{\text{جمع تقسیم}} \frac{۹۵+۹۶+۹۷+۹۸+a}{۵}=\frac{-۵-۴-۳-۲-۱}{۵}$$

مخرج‌ها را می‌توانیم با هم بزنیم:

$$\underbrace{۹۵+۹۶+۹۷+۹۸}_{۳۸۶}+a=\underbrace{-۵-۴-۳-۲-۱}_{-۱۵}$$

پس داریم: $a=-۱۵-۳۸۶=-۴۰۱$

۱۳۳- گزینه‌ی ۳ اول باید a را پیدا کنیم. گفته میانگین ۲,۳,۴,۴,a برابر ۲ است، پس داریم:

$$\frac{\text{جمع}}{\text{تعداد}}=\frac{۲+۳+۴+۴+a}{۵}=۲ \xrightarrow{\text{طرفین وسطین}} ۱۳+a=۱۰ \Rightarrow a=-۳$$

حالا شد! پس $a=-۳$ بوده. سؤال میانگین داده‌های ۲,۳,۴,۴,(۱۰۰+۶a) را خواسته، پس داریم:

$$\overline{x}=\frac{\text{جمع}}{\text{تعداد}}=\frac{۲+۳+۴+۴+۱۰۰+۶\overbrace{(-۳)}^{a}}{۵}=\frac{۱۳+۱۰۰-۱۸}{۵}=\frac{۹۵}{۵}=۱۹$$

یعنی میانگین گروه جدید می شود ۱۹.

۱۳۴- گزینه‌ی ۱ به میانگین هر دو گروه نگاه کنید. (تعدادشان ۱۰۱ تا است. نه؟!)

$$\overline{x}_۱=\frac{\text{جمع}}{\text{تعداد}}=\frac{۱+(۱+۲+...+۱۰۰)}{۱۰۱}$$

$$\overline{x}_۲=\frac{\text{جمع}}{\text{تعداد}}=\frac{(۱+۲+...+۱۰۰)+۱۰۰}{۱۰۱}$$

توی هر دو کسر، از یک تا صد را توی پرانتز گذاشتم. حالا صورت کسر اولی، عدد ۱ را اضافه دارد و صورت کسر دومی عدد ۱۰۰ را اضافه دارد.

پس اختلافشان در صورت کسر به اندازه‌ی $۱۰۰-۱=۹۹$ است، یعنی اختلاف دو میانگین $\frac{۹۹}{۱۰۱}$ می‌شود.

تغییر ترکیب داده‌ها در میانگین

فرمول میانگین چه‌جوری بود؟ $\text{میانگین}=\overline{x}=\frac{\text{جمع}}{\text{تعداد}}$

اگر این رابطه را طرفین وسطین کنیم این‌شکلی می‌شود: جمع = تعداد × میانگین

پس اگر تعداد داده‌ها و میانگین آن‌ها را داشته باشیم، می‌توانیم جمع داده‌ها را حساب کنیم: (میانگین × تعداد = جمع) $x_۱+x_۲+...+x_n=n\overline{x}$

حالا مجموع داده‌ها به چه درد ما می‌خورد؟ آهان، ماجرا این‌جاست که توی بعضی مسئله‌ها، ترکیب داده‌ها تغییر می‌کند. مثلاً داده‌ای حذف می‌شود؛ داده‌ی جدیدی اضافه می‌شود؛ دو گروه با هم ادغام می‌شوند؛ داده‌ای تغییر می‌کند و...

در این مسئله‌ها باید از مجموع داده‌ها استفاده کنیم تا به میانگین جدید برسیم.

فرض کنید معدل یک کلاس ۱۰نفری، ۱۶ است. پس جمع نمرات کلاس می‌شود ۱۶۰=۱۰×۱۶. حالا می‌خواهیم معدل کلاس را با تغییرات زیر حساب کنیم:

الف یک نفر با نمره‌ی ۱۴به کلاس اضافه شده:

با آمدن این یک نفر، مجموع نمرات می‌شود ۱۷۴=۱۴+۱۶۰، تعداد هم می‌شود ۱۱=۱+۱۰. پس معدل کلاس می‌شود: $\frac{\text{جمع}}{\text{تعداد}}=\frac{۱۷۴}{۱۱}\simeq ۱۵/۸۲$

ب دو نفر با نمره‌های ۱۹ و ۱۸ از کلاس (همان کلاس اولیه) حذف شده‌اند:

وقتی این دو نفر می‌روند، جمع نمرات می‌شود ۱۲۳=۱۹−۱۸−۱۶۰ و تعداد هم می‌شود ۸=۲−۱۰ نفر. پس معدل کلاس جدید می‌شود:

$$\frac{\text{جمع}}{\text{تعداد}}=\frac{۱۲۳}{۸}=۱۵/۳۷۵$$

در **الف** آمدن یک دانش‌آموز ضعیف معدل کلاس را کم کرد. در **ب** هم بیرون‌رفتن شاگردهای خوب معدل کلاس را کم کرد!

ج معلوم شده در کلاس اولیه نمره‌ی یک نفر به جای ۱۴ باید ۱۹ بشود:

خب به نمره‌ی این آقا (یا خانم) ۵ نمره اضافه شده (دلیلش را خدا می‌داند!) پس مجموع نمرات کلاس می‌شود ۱۶۵=۵+۱۶۰، اما تعداد فرقی نکرده، یعنی همان ۱۰ نفر هستند. پس معدل جدید می‌شود: $\frac{\text{جمع}}{\text{تعداد}}=\frac{۱۶۵}{۱۰}=۱۶/۵$

د این کلاس را با یک کلاس ۱۵ نفری با معدل ۱۲ ادغام کردیم:

در کلاس ۱۵ نفری با معدل ۱۲، مجموع نمرات می‌شود: ۱۸۰=۱۲×۱۵= میانگین × تعداد = جمع

مجموع کلاس خودمان هم ۱۶۰ بود. پس وقتی ادغام کنیم جمع نمرات کل دو کلاس می‌شود: ۳۴۰=۱۶۰+۱۸۰. تعداد کل روی هم ۲۵=۱۵+۱۰ نفر است.

$$\text{میانگین}=\frac{\text{جمع کل}}{\text{تعداد کل}}=\frac{۳۴۰}{۲۵}=۱۳/۶$$

پس داریم:

معدل یک کلاس ۱۶و معدل کلاس دیگر ۱۲ بود. وسطِ ۱۲و ۱۶ می‌شود ۱۴، اما چون کلاس ضعیف‌ها (با معدل ۱۲) تعدادش بیشتر بود، میانگین را از ۱۴ کم‌تر کرد. به اصطلاح می‌گویند میانگین را به طرف خودش کشید.

این فرمول‌هایی که در فیزیک و شیمی دیده‌اید مثلاً $\frac{C_۱V_۱+C_۲V_۲}{C_۱+C_۲}$ یا $\frac{N_۱V_۱+N_۲V_۲}{V_۱+V_۲}$ یا $\frac{V_۱t_۱+V_۲t_۲}{t_۱+t_۲}$ همگی از این‌جا درآمده‌اند. صرفاً جهت اطلاع!

۱۳۵- **گزینه‌ی ۲** سؤال گفته میانگین $x_۱+۱, x_۲+۲, x_۳+۳$ می‌شود $\bar{x}$، پس داریم: (تعدادشان ۳ تا است.)

$$\text{جمع}=\text{تعداد}\times\text{میانگین}\Rightarrow x_۱+۱+x_۲+۲+x_۳+۳=۳\bar{x}$$

البته این‌جوری بهتره: $x_۱+x_۲+x_۳=۳\bar{x}-۶$

حال برویم سراغ چیزی که می‌خواهد؛ باید میانگین داده‌های $۳x_۳+۳,$ $۳x_۱+۱, ۳x_۲+۲$ را حساب کنیم:

$$\text{میانگین داده‌های جدید}=\frac{\text{جمع}}{\text{تعداد}}=\frac{۳x_۱+۱+۳x_۲+۲+۳x_۳+۳}{۳}$$

اگر دسته‌بندی کنیم بهتر می‌شود: $=\frac{۳(x_۱+x_۲+x_۳)+۶}{۳}$

بگذارید ساده‌ترش کنیم: $=(x_۱+x_۲+x_۳)+۲$

از داده‌های اولیه فهمیده بودیم که $x_۱+x_۲+x_۳=۳\bar{x}-۶$، این را جای‌گذاری می‌کنیم: $\text{میانگین داده‌های جدید}=(۳\bar{x}-۶)+۲=۳\bar{x}-۴$

۱۳۶- **گزینه‌ی ۴** مجموع نمره‌های این هشت درس را می‌توانیم حساب کنیم:

$$\text{مجموع}=\text{تعداد}\times\text{میانگین}=۸\times ۱۲/۵=۴\times ۲۵=۱۰۰$$

دقت کردید! حالِ ضرب‌کردن ۸×۱۲/۵ را نداشتم. اول ۲×۱۲/۵ را نوشتم ۲۵، بعدش ۴×۲۵ آسان بود.

حالا دوتا نمره‌ی دیگر هم گرفته. پس جمع کل نمراتش می‌شود:

$$۱۰۰+۱۴+۱۶=۱۳۰$$

و میانگین ده درس برابر است با:

$$\bar{x}_{\text{کل}}=\frac{\text{جمع کل}}{\text{تعداد کل}}=\frac{۱۳۰}{۱۰}=۱۳$$

۱۳۷- **گزینه‌ی ۳** چی به ما داده، گفته میانگین $x_۴, x_۳, x_۲, x_۱$ می‌شود $\bar{x}$. پس جمعشان را داریم:

$$\text{جمع}=\text{تعداد}\times\text{میانگین}\Rightarrow x_۱+x_۲+x_۳+x_۴=۴\bar{x}$$

حالا چی از ما می‌خواهد؟ میانگین این داده‌های جدید را ... بگذارید حساب کنیم ببینیم چی می‌شود: $\text{میانگین جدیدها}=\frac{\text{جمع}}{\text{تعداد}}$

$$=\frac{۲x_۱+x_۲+۲x_۲+x_۳+۲x_۳+x_۴+۲x_۴+x_۱}{۴}$$

$$=\frac{۳x_۱+۳x_۲+۳x_۳+۳x_۴}{۴}=\frac{۳\overbrace{(x_۱+x_۲+x_۳+x_۴)}^{۴\bar{x}}}{۴}=\frac{۳(۴\bar{x})}{۴}=۳\bar{x}$$

چه خوب شد! جمع $x_۴, x_۳, x_۲, x_۱$ را هم از قبل داشتیم.

اصلاً می‌توانستیم سراغ $\bar{x}$ نرویم. ببینید:

$$\text{میانگین جدید}=\frac{۲x_۱+x_۲+۲x_۲+x_۳+۲x_۳+x_۴+۲x_۴+x_۱}{۴}$$

$$=\frac{۳(x_۱+x_۲+x_۳+x_۴)}{۴}$$