فصل چهارم: کار با داده های آماری:
درس (۱): گردآوری داده ها

جامعه‌ی آماری: به کلیهٔ افراد، اشیاء یا جانداران که که قرار است آنها را از نظر موضوعی خاص بررسی کنیم جامعه‌ی آماری می گوئیم. و به تعداد آنها اندازه‌ی جامعه می گوئیم.

مثال: اگر بخواهیم معدل دانش آموزان دبیرستان سعدی را (که ۳۷۲ نفر هستند) مورد بررسی قرار دهیم:
الف) جامعه‌ی آماری تمام دانش آموزان دبیرستان سعدی می باشد.
ب) اندازه‌ی جامعه همان تعداد دانش آموزان (۳۷۲ نفر) می باشد.

منظور از سرشماری چیست؟ اگر بخواهیم تمام اعضای جامعه‌ی آماری را مورد بررسی و مطالعه قرار دهیم می گوئیم سرشماری کرده‌ایم. که البته معایب زیر را دارد.

الف) وقت گیر بودن و هزینه‌ی زیاد آن
ب) در دسترس نبودن تمام اعضای جامعه
ج) کم و زیاد شدن اعضا در طول مدت سرشماری
د) داشتن خطا در جمع آوری اطلاعات.

فیروز محمودی @Love_Graph

منظور از نمونه‌گیری چیست؟ به دلیل مشکلاتی که سرشماری دارد، در بسیاری از موارد قسمتی کوچکتر از جامعه‌ی آماری را مورد بررسی و مطالعه قرار می دهند که به این کار نمونه گیری گفته می شود. که باید شرایط زیر را در آن رعایت کرد.

الف) اندازه‌ی نمونه گیری شده باید به مقدار کافی بزرگ باشد، یعنی متناسب با اندازه‌ی جامعه باشد.
ب) اعضای نمونه گیری شده باید کاملاً تصادفی انتخاب شوند یعنی برای انتخاب آنها از هیچ قانون خاصی نباید استفاده کرد.
ج) نمونه‌ی انتخاب شده باید بیانگر خصوصیات اصلی جامعه باشد.

منظور از متغیر تصادفی چیست؟ به موضوع اصلی مورد مطالعه متغیر تصادفی می گوئیم، مثلاً وقتی وزن اعضای یک خانواده را بررسی می کنیم، موضوع اصلی مورد مطالعه وزن است و متغیر تصادفی در اینجا وزن است.

منظور از داده های آماری چیست؟ پس از بررسی متغیرهای تصادفی مربوط به یک جامعه ی آماری به اعدادی می رسیم که به آنها داده های آماری می گوییم.
مثل اعداد مربوط به وزن اعضای یک خانواده.

آمارگیر: شخصی است که با روش های موجود در علم آمار به بررسی و جمع آوری داده ها می پردازد.

روشهای جمع آوری داده ها

الف) مشاهده: یعنی جمع آوری داده ها بدون نیاز به فرد پاسخگو مانند شمارش موتورهایی که در یک ساعت خاص از پل قرمز عبور می کنند

ب) مصاحبه: همان پرسش شفاهی است و معمولاً بین دو نفر اتفاق می افتد (مصاحبه گر و مصاحبه شونده). از مصاحبه وقتی استفاده می شود که آمارگیر از تمام جوابهای ممکن برای سؤالات اطلاعی نداشته باشد، مانند مصاحبه با مردم درباره ی برنامه های درخواستی آنها برای پخش در تلویزیون

ج) پرسشنامه: مجموعه ای از سؤالات است که از قبل تهیه و تنظیم شده است که بهتر است سؤالات ساده و کوتاه و در صورت امکان به صورت چند گزینه ای باشند. (مرکز آمار ایران هر ده سال یکبار به کمک پرسشنامه اطلاعات نفوس و مسکن تمام خانوارهای ایرانی را جمع آوری می کند)

د) دادگان (داده های از قبل تهیه شده): در خیلی از موارد بهتر این است که از اطلاعات قبلاً بدست آمده استفاده کنیم، مثلاً اگر بخواهیم تعداد جوانانی را که در سال ۹۴ ازدواج کرده اند را بدانیم، بهتر است که به اداره ثبت احوال مراجعه کنیم.

لازم به ذکر است که هرکدام از روشهای جمع آوری داده ها معایبی نیز دارند مثلاً:

الف) پرسشنامه: اگر تعداد واحدهای نمونه زیاد باشد، این روش زمان بر است
ب) مشاهده: اگر به دقت زیادی نیاز داشته باشیم، مشاهده روش مناسبی نیست
ج) دادگان: همیشه اطلاعات ثبتی را در اختیار آمارگیر قرار نمی دهند.
د) مصاحبه: وقت گیر است و اطلاعات بدست آمده از این روش را نمی توان به جامعه ی بزرگتری تعمیم داد

فیروز محمودی @Love_Graph

انواع متغیرها

۱- کمّی: متغیرهایی که بر روی آنها عملیات ریاضی مثل جمع و تفریق و معدل گیری و... انجام می شود مثل: نمره، قد، وزن، درآمد.
نکته: متغیرهای کمی به دو دسته تقسیم می شوند.

الف) فاصله‌ای: از جنس عدد هستند و قابل مرتب کردن و مقایسه کردن هستند و ضمنا اختلاف بین مقادیر داده ها با معنی باشد. مثل درجه ی حرارت شهرهای مختلف و نمره ی دانش آموزان

ب) نسبتی: متغیرهایی از جنس عدد هستند که تمام شرایط متغیرهای فاصله‌ای را دارند و علاوه بر آن نسبت (تقسیم) مقادیر داده ها نیز با معنی است. مثل شدت زلزله و وزن افراد. در این جا صفر به معنی نبود و نیز گی در مورد دانش آموز، است مثلاً وقتی می گوییم درآمد شخصی صفر است. به این معنی است که هیچ درآمدی ندارد. پس درآمد متغیر کمی نسبتی است. ولی اگر دمای هوای شهر صفر درجه باشد، نمی توان گفت که شهر دمایی ندارد.

۲- کیفی: متغیرهایی هستند که فقط برای دسته بندی افراد یا اشیاء در گروه ها به کار می رود مانند گروه خونی، جنسیت افراد، رنگ لباسها
نکته: متغیرهای کیفی به دو دسته تقسیم می شوند.

الف) ترتیبی: برای متغیرهایی است که در آنها نوعی ترتیب طبیعی وجود دارد و قابل مرتب کردن و قابل مقایسه کردن باهم هستند. ولی اختلاف آنها غیرممکن است یا بی معنی است. مثل حروف الفبای فارسی، مراحل زندگی انسان (رتبه ی دانش آموزان در کنکور با آن که عدد است ولی کمّی محسوب نمی شود چون مثلاً جمع رتبه ها بی معنی است.) بنابراین رتبه ی دانش آموزان در کنکور کیفی ترتیبی است.

ب) اسمی: متغیرهایی هستند که شامل نام ها، برچسب ها و گروهها می شوند، ضمناً این متغیرها دارای ترتیب خاصی نیستند و نمی توانیم آنها را با هم مقایسه کنیم، مثل گروه خونی، اسامی افراد، ورزش مورد علاقه، رنگ چشم افراد

فیروز محمودی @Love Graph

مثال: نوع متغیرهای زیر را مشخص کنید.

الف) رتبه‌ی داوطلبان آموزان در کنکور ← کیفی ترتیبی، چون جمع و تفریق رتبه‌ها بی معنی است.

ب) درجه حرارت شهرها ← کمی فاصله‌ای، چون اگر صفر باشد، نمی‌توان گفت که شهر دمایی ندارد.

ج) مساحت شکلهای هندسی ← کمی نسبتی.

د) مزه‌ی غذا (خوب، متوسط، بد) ← کیفی ترتیبی

ه) زمان اولین کلاس دانشگاه‌ها در روز شنبه ← کمی فاصله‌ای

ت) شاخص توده‌ی بدن ← کمی نسبتی

ث) مراحل کاشت یک نهال ← کیفی ترتیبی

س) رنگ لباسهای موجود در یک ورزشگاه ← کیفی اسمی

ش) شدت زلزله در یک منطقه ← کمی نسبتی

ع) RH خون ← کیفی اسمی

غ) تغییرات توده‌ی سرطانی در عضو انسان در طی چند ماه ← کیفی ترتیبی

ک) قد دانش آموزان ← کمی نسبتی

گ) ساعت شروع امتحان ← کمی فاصله‌ای

ل) مدت زمان امتحان ← کمی نسبتی

م) ورزش مورد علاقه ← کیفی اسمی

فیروز محمودی @Love_Graph

تعریف علم آمار: به مطالعه‌ی نحوه‌ی جمع آوری، سازماندهی، تحلیل و تفسیر داده‌ها جهت استخراج اطلاعات و تصمیم گیری، علم آمار می‌گوییم.

تعریف پارامتر: یک مشخصه‌ی عددی است که بیان کننده‌ی ویژگی خاصی از جامعه است و در صورتی که داده‌های کل جامعه در اختیار باشند، قابل محاسبه است. مثلاً در یک گروه کوهنوردی بزرگ، نسبت مردان به کل افراد گروه معرف یک پارامتر است لازم به ذکر است که پارامتر همیشه عددی ثابت است.

فیروز محمودی @Love Graph

مثال: یک گروه ۲۰۰۰ نفری از افراد کوهنورد را در نظر بگیرید و فرض کنید که مردان این گروه ۱۲۰۰ نفر باشند، در این صورت نسبت مردان به کل افراد گروه $\frac{1200}{2000}$ می‌باشد و پارامتری است که ویژگی مرد بودن اعضا را مشخص می‌کند (و عددی ثابت است) ولی گاهی اوقات به دلیل زیاد بودن اعضای جامعه معمولاً پارامتر قابل محاسبه نیست، به همین دلیل باید از آماره استفاده کنیم در واقع آماره هم مشخصه‌ای عددی است که بیان کننده‌ی ویژگی خاصی از نمونه است. بنابراین اگر مثلاً یک نمونه‌ی ۲۰۰ نفره از کوهنوردان را (جامعه) انتخاب کنیم و تعداد مردان این نمونه ۱۳۰ نفر باشد. نسبت $\frac{\text{تعداد مردان این نمونه}}{\text{تعداد کل افراد این نمونه}} = \frac{130}{200}$ آماره‌ای است که ویژگی مرد بودن اعضای این نمونه را مشخص می‌کند.

اکنون اگر نمونه‌ای دیگر از این جامعه را در نظر بگیریم حاصل $\frac{\text{تعداد مردان}}{\text{تعداد کل افراد نمونه}}$ تغییر خواهد کرد. بنابراین:

با بررسی چند آماره‌ی مختلف و محاسبه‌ی میانگین آنها می‌توانیم پارامترها را تخمین بزنیم

سوال امتحانی:

فرق بین پارامتر و آماره چیست؟ کدام یک از آنها همیشه مقداری ثابت دارد؟

پارامتر و آماره هر دو مشخصه‌ای عددی هستند ولی پارامتر یک ویژگی از جامعه را نشان می‌دهد و آماره یک ویژگی از نمونه‌ی تصادفی را نشان می‌دهد. ضمناً پارامتر همیشه مقداری ثابت است

سوال امتحانی:

فرق بین داده و متغیر چیست؟

متغیر آن موضوعی است که می‌خواهیم برای جامعه یا نمونه بررسی کنیم مانند وزن یا قد افراد خانواده، ولی به نتایج حاصل از بررسی‌های آماری (اعداد بدست آمده) داده می‌گوییم مانند اعداد مربوط به طول قد افراد یک خانواده که می‌تواند ۱۶۵ یا ۱۷۰ یا .... باشند.

شاخص های مرکزی (معیارهای گرایش به مرکز) عبارتند از:
الف) میانگین (معدل)
ب) میانه
ج) مُد

الف) میانگین: برای محاسبه‌ی میانگین (معدل)، ابتدا باید داده‌های مورد نظر را با هم جمع کنیم و سپس عدد بدست آمده را بر تعداد آنها تقسیم کنیم.
یعنی اگر x_1 ، x_2 ، x_3 ، ... و x_n یک سری از داده های آماری باشند میانگین آنها برابر است با:

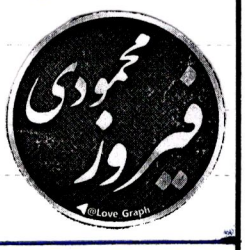
$$\text{میانگین } (\bar{x}) = \frac{\text{مجموع داده‌ها}}{\text{تعداد کل}} = \frac{x_1+x_2+x_3+\cdots+x_n}{N}$$

مثال: میانگین داده‌های ۴ ، ۶ ، ۱۰ ، ۲ ، ۳ را بدست آورید.

$$\text{میانگین } (\bar{x}) = \frac{۳+۲+۱۰+۶+۴}{۵} = \frac{۲۵}{۵} = ۵$$

مثال: اگر میانگین داده‌های ۴ ، $۲x+۱$ ، ۲ ، $۴x-۱$ ، ۴ برابر ۶ باشد، مقدار x چقدر است.

$$\frac{۲x+۱+۲+۴x-۱+۴}{۴} = ۶ \Rightarrow \frac{۶x+۶}{۴} = ۶ \Rightarrow ۶x+۶ = ۲۴$$
$$۶x = ۲۴-۶ = ۱۸$$
$$x = \frac{۱۸}{۶} = ۳$$



ب) میانه: داده‌ای است که نصف داده‌ها از آن کوچکتر و نصف دیگر داده‌ها از آن بزرگترند.

نکته: برای بدست آوردن میانه ابتدا آنها را از کوچک به بزرگ مرتب می‌کنیم و سپس:
الف) اگر تعداد داده‌ها فرد باشد، عدد وسط را به عنوان میانه در نظر می‌گیریم.
ب) اگر تعداد داده‌ها زوج باشد، میانگین دو عدد وسط را به عنوان میانه در نظر می‌گیریم.

مثال: میانه‌ی داده‌های ۱۱ و ۱ و ۲ و ۹ و ۴ را مشخص کنید.
جواب: ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم.

۱ ، ۲ ، [۴] ، ۹ ، ۱۱
میانه = ۴

مثال: میانه‌ی داده‌های ۱۰ و ۶ و ۴ و ۱ و ۱۱ و ۱ را مشخص کنید.
جواب: ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم.

۱ ، ۱ ، [۴ ، ۶] ، ۱۰ ، ۱۱

$$\text{میانه} = \frac{۴+۶}{۲} = \frac{۱۰}{۲} = ۵$$

ج) مُد: داده‌ای است که بیشتر از همه تکرار شده است. به عبارت دیگر مُد، داده‌ای است که بیشترین فراوانی را دارد. البته در یک جامعه‌ی آماری ممکن است دو یا چند مُد داشته باشیم، در این صورت مُد شاخص معتبری نخواهد بود.

الف) صنعت پوشاک: فروشندگان پوشاک، با آمارگیری درمی یابند که چه نوع پوشاکی مورد پسند و نظر مصرف کنندگان است، بنابراین ازهمان پوشاک برای فروش سفارش می دهند.

ب) در رأی گیری ها: در هر رأی گیری، اساس تصمیم گیری مُد است، یعنی موردی که بیشترین فراوانی را داشته باشد انتخاب می شود. مثلاً در انتخاب رئیس جمهور، فردی انتخاب می شود که بیشترین فراوانی (رأی) را داشته باشد.

مثال: در داده‌های ۳، ۵، ۵، ۷، ۴، ۳، ۵، ۶ مُد را مشخص کنید.
جواب: مُد عدد ۵ می باشد، زیرا عدد ۵ بیشتر از داده‌های دیگر تکرار شده است.

مثال: در داده‌های ۶، ۶، ۹، ۲، ۹، ۳، ۹، ۶ مُد را مشخص کنید.
جواب: مُد عبارتست از ۹، ۶ زیرا اعداد ۶ و ۹ به یک اندازه تکرار شده‌اند. (هرکدام سه بار)

میانگین وزن دار: اگر در یک سری از داده‌ها، تکرار وجود داشته باشد، برای محاسبه‌ی میانگین آنها ابتدا هر داده را در فراوانی خودش (تعداد دفعات تکرارش) ضرب می کنیم، سپس جوابهای بدست آمده را با هم جمع کرده و بر فراوانی کل (تعداد کل) تقسیم می کنیم.

مثال: میانگین داده‌های مقابل را بدست آورید.

۳۰، ۳۰، ۳۰، ۳۰، ۳۰، ۳۰، ۲۶، ۳۰، ۳۰، ۳۰، ۲۰، ۱۸، ۱۸، ۱۰، ۱۰، ۱۰

$$\text{میانگین } (\bar{x}) = \frac{3\times10+2\times18+20+26+7\times30}{14} = \frac{322}{14} = 23$$



مثال: جدول مقابل نمرات امیرعلی را نشان می دهد. معدل او چقدر است؟

درس	ریاضی	عربی	فیزیک	زبان
نمره	۱۴	۱۵	۱۸	۱۲
ضریب	۴	۲	۳	۲

$$\text{معدل } (\bar{x}) = \frac{4\times14+2\times15+3\times18+2\times12}{11} = \frac{164}{11} \simeq 14.90$$

مجموع ضرایب $= 4+2+3+2 = 11$

شاخص های پراکندگی (معیارهای پراکندگی) عبارتند از:
الف) دامنه‌ی تغییرات.
ب) واریانس.
ج) انحراف استاندارد (انحراف معیار).
د) دامنه‌ی میان چارکی.

الف) دامنه‌ی تغییرات: با اختلاف بین بزرگترین و کوچکترین داده، دامنه‌ی تغییرات گفته می‌شود. که آن را با حرف R نشان می‌دهند. یعنی:
$R = max - min$

مثال: دامنه‌ی تغییرات را در داده‌های ۳, ۷, ۱۲ و ۳, ۴ را مشخص کنید.
۱۲ = بیشترین داده
۱ = کمترین داده
$R = max - min = ۱۲ - ۱ = ۱۱$

نکته: R معیاری سریع برای محاسبه‌ی پراکندگی داده‌ها است. ولی معیار خوب و قابل اطمینانی نیست زیرا فقط با بیشترین داده (max) و کمترین داده (min) کار دارد. و بقیه‌ی داده‌ها هیچ نقشی در محاسبه‌ی آن ندارند

ب) واریانس: برای محاسبه‌ی واریانس که آن را با σ^2 نشان می‌دهند به این صورت عمل می‌کنیم.
۱- میانگین داده‌ها را بدست می‌آوریم.
۲- تک‌تک داده‌ها را صفهای میانگین می‌کنیم و حاصل آن را به توان ۲ می‌رسانیم.
۳- جوابهای بدست آمده از مرحله‌ی (۲) را با هم جمع می‌کنیم و بر تعدادشان تقسیم می‌کنیم (معدل می‌گیریم)

لـ علامت σ را «سیگما» بخوانید.

مثال: واریانس داده‌های ۱ و ۷ و ۱۱ و ۵ را بدست آورید.

$\bar{x}$ (میانگین) $= \frac{۵+۱۱+۷+۱}{۴} = \frac{۲۴}{۴} = ۶$

$(۶-۵)^۲ = ۱^۲ = ۱$
$(۶-۱۱)^۲ = (-۵)^۲ = (-۵)\times(-۵) = ۲۵$
$(۶-۷)^۲ = (-۱)^۲ = (-۱)\times(-۱) = ۱$
$(۶-۱)^۲ = ۵^۲ = ۲۵$

σ^2 (واریانس) $= \frac{۱+۲۵+۱+۲۵}{۴} = \frac{۵۲}{۴} = ۱۳$

نکته: اگر از واریانس جذر بگیریم، انحراف استاندارد بدست می‌آید. که آن را با حرف σ نشان می‌دهیم.

انحراف استاندارد (σ) $= \sqrt{۱۳} \simeq ۳,۶۰$

تدریس خصوصی ریاضیات دبیرستان (دوره‌ی اول و دوم و ...) فیروز محمودی همراه: ۰۹۱۳۷۰۲۷۲۵۲

نکته‌ی مهم: اگر در یک سری از داده‌ها، دامنه‌ی تغییرات یا واریانس یا انحراف استاندارد صفر باشد. با این معنی است که تمام داده‌ها با هم برابرند. برعکس این موضوع هم درست است یعنی اگر تمام داده‌ها با هم برابر باشند، دامنه‌ی تغییرات و واریانس و انحراف استاندارد صفر می‌باشد.

مثال: در داده‌های ۱۰ و $a+1$ و $b-2$ و $\frac{c}{3}$ واریانس صفر می‌باشد. مقادیر a و b و c را بیابید.

جواب: چون واریانس صفر است، پس تمام داده‌ها با هم برابرند. ولی دانیم که یکی از آن‌ها ۱۰ می‌باشد پس بقیه‌ی داده‌ها نیز باید ۱۰ باشند.

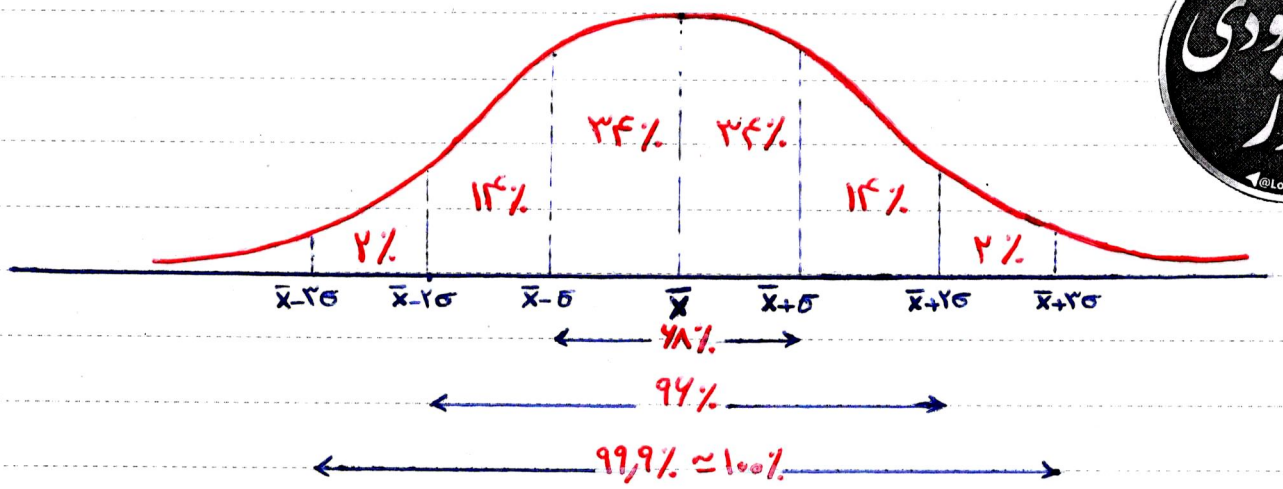
$\frac{c}{3}=10 \Rightarrow c=30$

$b-2=10 \Rightarrow b=10+2=12$

$a+1=10 \Rightarrow a=10-1=9$

نحوه‌ی پخش داده‌ها در اطراف میانگین در منحنی نرمال

منحنی نرمال، یک منحنی به شکل زنگ است که داده‌ها به طور متقارن در آن پخش شده‌اند. در این منحنی نحوه‌ی پراکندگی داده‌ها به صورت زیر است.



فیروز محمودی @Love_Graph

بنابراین: ۶۸ درصد داده‌ها در بازه‌ی $(\bar{x}-\sigma \ , \ \bar{x}+\sigma)$ قرار دارند

۹۶ درصد داده‌ها در بازه‌ی $(\bar{x}-2\sigma \ , \ \bar{x}+2\sigma)$ قرار دارند.

۱۰۰ درصد داده‌ها در بازه‌ی $(\bar{x}-3\sigma \ , \ \bar{x}+3\sigma)$ قرار دارند.

در واقع: $(\bar{x}-\sigma \ , \ \bar{x}+\sigma)$ داده‌هایی هستند که بین یک برابر انحراف معیار از میانگین قرار دارند.

$(\bar{x}-2\sigma \ , \ \bar{x}+2\sigma)$ داده‌هایی هستند که بین دو برابر انحراف معیار از میانگین قرار دارند.

$(\bar{x}-3\sigma \ , \ \bar{x}+3\sigma)$ داده‌هایی هستند که بین دو برابر انحراف معیار از میانگین قرار دارند.

تدریس خصوصی ریاضیات دبیرستان (دوره‌ی اول و دوم و ...) فیروز محمودی همراه: ۰۹۱۳۷۰۲۷۲۵۲

مثال: در یک سری از داده های آماری با منحنی نرمال، $\bar{x}=30$ و $\sigma=4$ می باشد.

الف) ۶۸ درصد داده ها در چه بازه ای هستند؟
ب) ۹۶ درصد داده ها در چه بازه ای هستند؟
ج) ۱۰۰ درصد داده ها در چه بازه ای هستند؟

فیروز محمودی
@Love_Graph

جواب: با توجه به منحنی توزیع نرمال (صفحه ی قبل) واضح است که ۶۸ درصد داده ها در بازه ی $(\bar{x}-\sigma, \bar{x}+\sigma)$ قرار دارند. یعنی؛

جواب (الف) $(\bar{x}-\sigma, \bar{x}+\sigma) = (30-4, 30+4) = (26, 34)$

جواب (ب) $(\bar{x}-2\sigma, \bar{x}+2\sigma) = (30-8, 30+8) = (22, 38)$

جواب (ج) $(\bar{x}-3\sigma, \bar{x}+3\sigma) = (30-12, 30+12) = (18, 42)$

حتماً دیده اید که روی قوطی یا بسته ی مواد غذایی اعدادی مثل 200 ± 2 و امثال آن نوشته می شود. در واقع این اعداد به صورت $\bar{x} \pm 2\sigma$ خوانده می شود که $\bar{x}$ وزن خالص ماده ی غذایی است که باید داخل بسته باشد. ولی چون همیشه کمی خطا در پُر کردن بسته ها وجود دارد $\pm\sigma$ را هم در نظر می گیریم. یعنی ۹۶ درصد بسته ها (قوطی ها) وزنشان در بازه ی $(\bar{x}+2\sigma \text{ و } \bar{x}-2\sigma)$ قرار دارند. مثلاً اگر روی قوطی کنسرو ماهی عبارت 200 ± 10 گرم نوشته شود، به این معنی است که در ۹۶ درصد قوطی کنسروها وزن تن ماهی داخل آن عددی در بازه ی $(200+10 \text{ و } 200-10)$ یعنی $(210, 190)$ گرم می باشد. در واقع اگر ما ۱۰۰ قوطی کنسرو ماهی را وزن کنیم انتظار داریم که حداقل ۹۶ تا از آنها وزنی بین ۱۹۰ تا ۲۱۰ گرم داشته باشند. که درج چنین اعدادی بر روی کالاها از الزامات استاندارد هر کشوری است.

چارک های اول و دوم و سوم - دامنه‌ی میان چارکی

موضوع را با یک مثال توضیح می دهیم. داده‌های مقابل را در نظر بگیرید.

۱, ۴, ۹, ۱۶, ۸, ۱, ۱۳, ۱۱, ۱۷, ۵, ۶

برای بدست آوردن میانه، ابتدا باید آنها را از کوچک به بزرگ مرتب کنیم و عدد وسط را به عنوان میانه در نظر می‌گیریم.

۱, ۱, ۴, ۵, ۶, ⑧, ۹, ۱۱, ۱۳, ۱۶, ۱۷

به عدد ۸ میانه یا چارک دوم گفته می‌شود که آنرا با Q_2 نمایش می دهیم.

حال اگر برای اعداد کمتر از ۸ نیز میانه را مشخص کنیم به Q_1 و اگر میانه را برای اعداد بزرگتر از ۸ نیز مشخص کنیم به Q_3 خواهیم رسید.

۱, ۱, ④, ۵, ۶, ⑧, ۹, ۱۱, ⑬, ۱۶, ۱۷

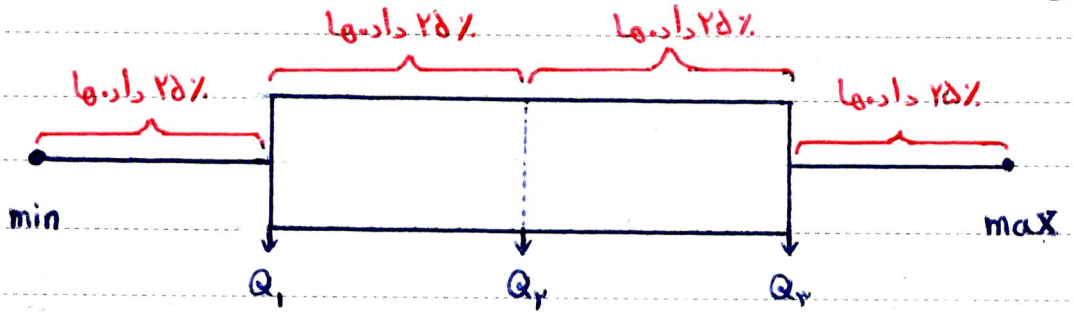
Q_1 چارک اول — Q_2 چارک دوم (میانه) — Q_3 چارک سوم



نکته‌ی مهم: $Q_3 - Q_1$ را دامنه‌ی میان چارکی می گویند و آنرا با IQR نشان می دهند پس

$$IQR = Q_3 - Q_1 = ۱۳ - ۴ = ۹$$

نکته‌ی مهم: چارک ها کل داده‌های آماری را به چهار قسمت مساوی از نظر تعداد تقسیم می کنند که در نمودار زیر این موضوع به خوبی نمایش داده شده است. (با این نمودار، نمودار جعبه‌ای گفته می‌شود)



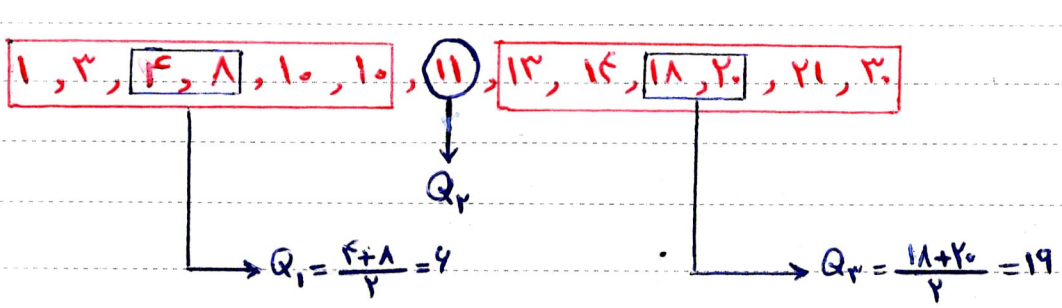
نکته: اگر در یک سری از داده‌های آماری، یک یا چند داده‌ی دور افتاده (پرت) وجود داشته باشد، برای بررسی میزان پراکندگی داده‌ها باید از میانه و دامنه‌ی میان چارکی استفاده کنیم. یعنی در این حالت از $\bar{x}$ و σ استفاده نمی کنیم.

تدریس خصوصی ریاضیات دبیرستان (دوره‌ی اول و دوم و ...) فیروز محمودی همراه: ۰۹۱۳۷۰۲۷۲۵۲

سؤال امتحانی: برای داده های زیر چارک اول، دوم و سوم و دامنه ی میان چارکی را مشخص کنید.

۱۸ و ۴ و ۳ و ۱۱ و ۱۳ و ۲۰ و ۲۱ و ۸ و ۱۰ و ۳۰ و ۱۴ و ۱۰ و ۱

جواب: ابتدا باید داده ها را از کوچک به بزرگ مرتب کنیم.





بنابراین داریم:

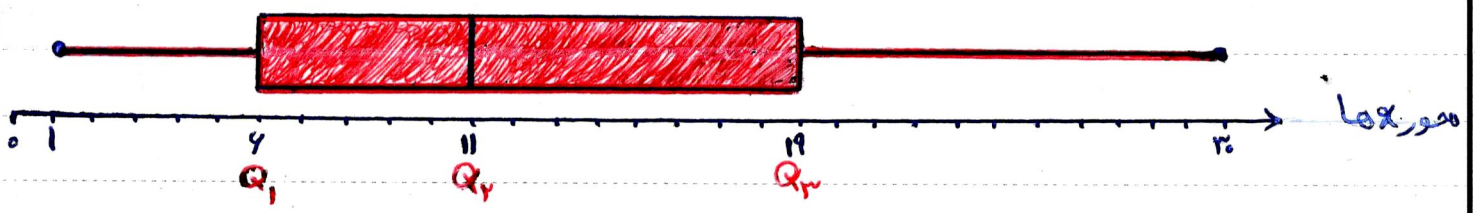
Q_1 = چارک اول = ۶

Q_2 = چارک دوم = ۱۱

Q_3 = چارک سوم = ۱۹

IQR = دامنه ی میان چارکی = $Q_3 - Q_1 = ۱۹ - ۶ = ۱۳$

و نمودار جعبه ای آن به صورت زیر می باشد.



نکته ی مهم: در داخل جعبه لزومی دارد که میانه (Q_2 یا همان) حتماً در وسط جعبه باشد.

نکته ی مهم: همه ی آماره هایی که تا اینجا ذکر شد (میانگین، میانه، انحراف معیار، دامنه ی میان چارکی) را تحت عنوان آمار توصیفی می شناسیم.



داستنی: می دانیم که نمودار جعبه ای به صورت ——[ ]—— می باشد، از آنجا که در خط دو طرف جعبه شبیه به سبیل گربه است. بعضی مواقع به آن نمودار جعبه و سبیل هم گفته می شود.

لازم به ذکر است که سبیل ها کمترین و بیشترین مقادیر داده ها را نشان می دهند.

مثال: امتیازات کسب شده‌ی دو بازیکن بسکتبال به صورت زیراست. نمودار جعبه‌ای مربوط به امتیازات هرکدام را رسم کنید. و مشخص کنید که عملکرد کدام یک از آنها بهتر بوده است.

امتیاز	۳	۶	۷	۱۰	۱۱	۱۳	۳۰
تعداد بازی بازیکن A	۲	۱	۲	۳	۱	۱	۱

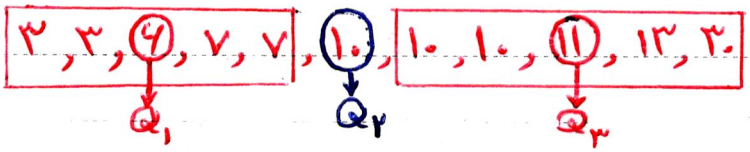
امتیاز	۷	۸	۹	۱۰	۱۱	۱۲	۳۰
تعداد بازی بازیکن B	۱	۱	۲	۲	۲	۲	۱

فیروز محمودی @Love Graph

جواب: ابتدا اعداد مربوط به هر جدول را از حالت جدولی خارج می‌کنیم و آنها را بازنویسی می‌کنیم. (دقت شود که در جدول بالا، عدد ۱۰ سه بار تکرار شده است. پس سه بار عدد ۱۰ را می‌نویسیم)

بازیکن A: ۳, ۳, ۶, ۷, ۷, ۱۰, ۱۰, ۱۰, ۱۱, ۱۳, ۳۰

آنها را به صورت مرتب شده می‌نویسیم.



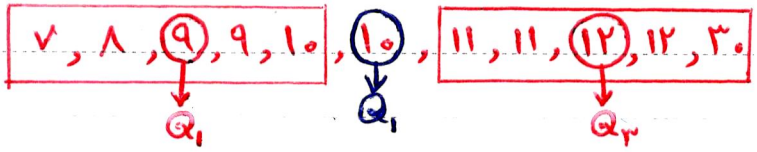
$R = max - min = ۳۰ - ۳ = ۲۷$

$IQR = Q_۳ - Q_۱ = ۱۱ - ۶ = ۵$

به همین ترتیب برای بازیکن B داریم:

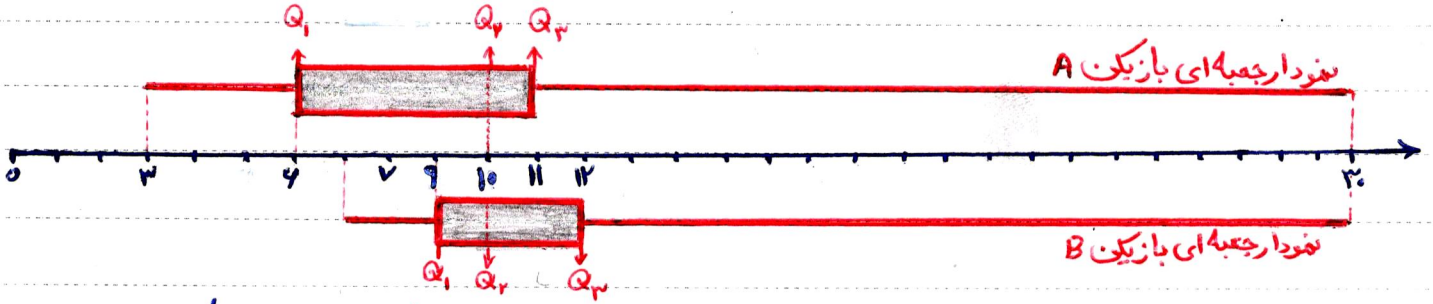
بازیکن B: ۷, ۸, ۹, ۹, ۱۰, ۱۰, ۱۱, ۱۱, ۱۲, ۱۲, ۳۰

آنها را به صورت مرتب شده می‌نویسیم.



$R = max - min = ۳۰ - ۷ = ۲۳$

$IQR = Q_۳ - Q_۱ = ۱۲ - ۹ = ۳$



میانه‌ی امتیازات هردو بازیکن برابرند، (عدد ۱۰) و max ها نیز برابرند (عدد ۳۰) ولی min بازیکن B بیشتر از min بازیکن A است. همچنین $Q_۱$ و $Q_۳$ بازیکن B بیشتر از $Q_۱$ و $Q_۳$ بازیکن A می‌باشد پس در کل عملکرد بازیکن B بهتر از A بوده است. و همچنین R و IQR بازیکن B کمتر از بازیکن A است. به این معنی که پراکندگی امتیازها در بازیکن B کمتر است. یعنی بازیکن B ثبات بیشتری دارد. (به مقدار R و IQR مربوط به هردو بازیکن دقت کنید).

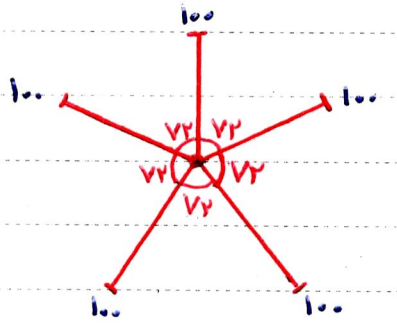
فیروز محمودی @Love_Graph

فصل پنجم: نمایش داده‌ها

- انواع نمودارها
  - الف) نمودارهای یک متغیره
    1. نمودار میله‌ای
    2. نمودار خط شکسته
    3. نمودار مستطیلی (هیستوگرام)
    4. نمودار دایره‌ای (دوبعدی)
    5. نمودار دایره‌ای (سه‌بعدی) یا کلوچه‌ای
    6. نمودار جعبه‌ای
  - ب) نمودارهای چند متغیره
    1. نمودار حبابی
    2. نمودار راداری

در بین انواع مختلف نمودارها تأکید این فصل بر روی نمودارهای حبابی و راداری می باشد که با جهت اهمیت نمودار راداری، فقط این نوع از نمودار را توضیح می دهیم.

نمودار راداری: یک نمودار دوبعدی است که در آن سه متغیر کمی یا بیشتر روی محورهایی که نقطه‌ی شروع مشترک دارند، نمایش داده می شوند. مثلاً اگر بخواهیم ۵ متغیر را نمایش دهیم، زاویه‌ی بین این پنج پاره‌خط را برابر $(72 = \frac{360}{5})$ درجه در نظر می گیریم و همواره طول هر پاره‌خط را برابر ۱۰۰ در نظر می گیریم.



نکته‌ی مهم: محل قرارگیری و زاویه‌ی محورها نسبت به یکدیگر هیچ اطلاعات خاصی را بیان نمی کند.

نکته: هر کدام از پاره‌خط‌های نمودار راداری نشان دهنده‌ی یک متغیر است که با این پاره‌خط‌ها، شعاع‌های نمودار راداری می گویند.

نکته: از این نمودار معمولاً برای مقایسه‌ی عملکرد بازیکنان ورزشی استفاده می شود.

نکته: برای رسم این نمودار در هر متغیر، ابتدا مقدار مورد نظرمان را بر مقدار ماکزیمم داده‌های آن متغیر تقسیم می کنیم و حاصل بدست آمده را در ۱۰۰ ضرب می کنیم، همین کار را برای بقیه‌ی محورها نیز انجام می دهیم، و نقاط بدست آمده را به هم وصل می کنیم.

مثال: برای جدول زیر، نمودار راداری مناسب را رسم کنید.

متغیرها	توپ ربایی	دریبل موفق	پاس گل	گل زده	تعداد بازی
عملکرد بازیکن مورد نظر	۳٫۵	۹۰	۱٫۲۲	۰٫۸	۴۰
مقدار ماکزیمم در بین تمام بازیکنان	۴	۹۰	۳	۱٫۷	۴۵

جواب: طبق جدول، واضح است که پنج متغیر داریم. بنابراین نمودار راداری ما از پنج پاره خط تشکیل می شود که زاویه ی بین هرکدام از آنها برابر است با $۳۶۰ \div ۵ = ۷۲$

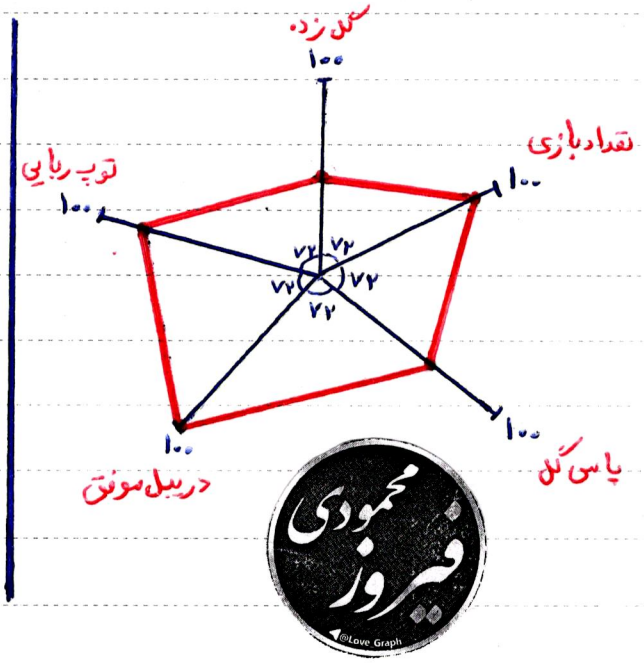
$$\text{درصد توپ ربایی} = \frac{۳٫۵}{۴} \times ۱۰۰ = \frac{۳۵۰}{۴} \simeq ۸۷$$

$$\text{درصد دریبل موفق} = \frac{۹۰}{۹۰} \times ۱۰۰ = \frac{۹۰۰۰}{۹۰} = ۱۰۰$$

$$\text{درصد پاس گل} = \frac{۱٫۶۲}{۳} \times ۱۰۰ = \frac{۱۶۲}{۳} = ۵۴$$

$$\text{درصد تعداد بازی} = \frac{۴۰}{۴۵} \times ۱۰۰ = \frac{۴۰۰۰}{۴۵} = ۸۸$$

$$\text{درصد گل زده} = \frac{٪۸}{۱٫۷} \times ۱۰۰ = \frac{۸۰}{۱٫۷} \simeq ۴۷$$



فیروز محمودی
@Love Graph

مثال: درجای خالی عبارت مناسب بنویسید.

الف) از نمودار راداری برای نمایش متغیرهای کیفی به طور همزمان بکار می رود.

ب) اگر زاویه ی بین دو شعاع مجاور در نمودار راداری ۴۰ درجه باشد، ۹ متغیر در نمودار حضور دارند. زیرا $۴۰ \div ۳۶۰ = ۹$

پ) هریک از خطوط نمودار راداری معرف یک متغیر می باشد.

مثال: کاربرد نمودار راداری در ورزش چیست؟ برای نشان دادن قدرت و ضعف بازیکنان نسبت به همدیگر از نمودار راداری استفاده می شود.

مثال: نمودار راداری بیانگر چه می باشد. الف) کدام مشاهده ها شبیه یکدیگرند. ب) آیا داده ی دور افتاده وجود دارد. ج) مقدار کدام متغیر برای یک مشاهده نسبت به متغیرهای دیگر بیشتر یا کمتر است