

فصل چهارم

کار با داده های آماری

درس اول: گروه آماری داده ها

درس دوم: معیار های گرایش به مرکز

درس سوم: معیار های پراکنندگی

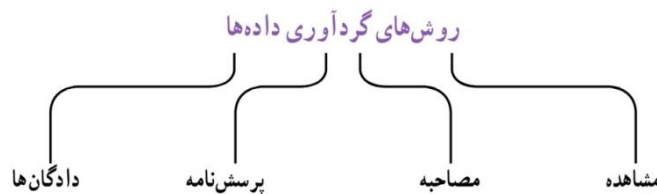
درس اول : گرد آوری داده ها

داده ها : داده ها واقعیت هایی درباره یک چیز هستند که در محاسبه ، استنباط یا برنامه ریزی به کار می روند.

واحد آماری : به هر یک از افراد یا چیزهایی می گویند که داده های مربوط به آنها در یک بررسی آماری گرد آوری می شود .

جامعه آماری : مجموعه کل واحد های آماری را جامعه آماری می گویند .

نمونه : هر زیر مجموعه از جامعه آماری که با روش مشخصی انتخاب شده باشد را نمونه می نامند .



مشاهده : گرد آوری داده ها بدون نیاز به فرد پاسخگو . مانند شمارش تعداد وسایل نقلیه عبوری از یک تقاطع در هر ساعت یا اندازه گیری وزن محصولات یک باغ میوه .

مصاحبه : معمولاً بین دو نفر صورت می گیرد ، یکی مصاحبه گر و دیگری مصاحبه شونده . مانند مصاحبه با صاحب نظران در مورد عوامل افزایش ترافیک شهری .

پرسش نامه : مجموعه سوالات از پیش تعیین شده که توسط تعدادی پاسخ دهنده تکمیل می شود . مانند آمار گیری که هر ۱۰ سال یک بار در ایران انجام می شود .

دادگان : شامل مجموعه ای از اطلاعات ذخیره شده است . مانند استفاده از اطلاعات ثبت شده پلیس راهور در مورد تعداد مجروحین تصادفی در یک سال مشخص .

مثال : کدام روش جمع آوری داده ها برای موارد زیر مناسب تر است ؟

۱. میزان رضایت مشتری های یک بانک از نحوه برخورد کارمندان آن
۲. سن دانش آموزان پایه دهم مدرسه بر حسب ماه
۳. تعداد سرنشینان خودروهای سواری در یکی از محور های خروجی شهر

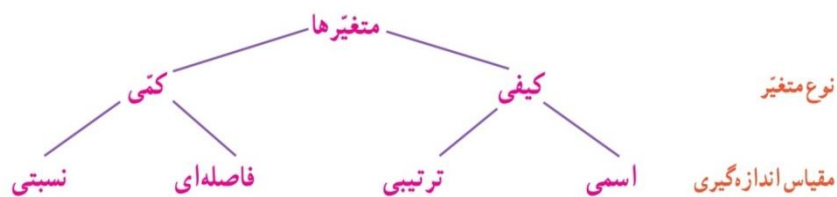
پارامتر : یک مشخصه عددی که توصیف کننده جنبه های خاص از یک جامعه است و فقط در صورتی که داده های کل جامعه در اختیار باشد ، قابل محاسبه است . مثلاً اگر داده های مربوط به تمام کارمندان دولت در دسترس باشد ، نسبت مردان به کل کارمندان یک پارامتر خواهد بود .

آماره: یک مشخصه عددی که توصیف کننده جنبه های خاص از یک نمونه است و در صورتی امکان آمار گیری از کل جامعه وجود نداشته باشد از نمونه استفاده کرده و آماره را برای تخمین پارامتر محاسبه می کنیم. مثلاً اگر قرار باشد در استان خوزستان هندوانه های قابل برداشت از نظر کیفیت مزه (خوب یا بد) بررسی شود این کار در مورد تمام هندوانه ها امری نشدنی است بنابراین با انتخاب یک نمونه و بررسی آنها، نسبت هندوانه های خوب یک آماره خواهد بود.

متغیر: هر ویژگی از اشخاص یا اشیا که قرار است مورد بررسی قرار بگیرد. که به دو دسته کلی کمی و کیفی دسته بندی می شود.

۱. کمی: متغیر هایی که مقدار عددی می گیرند.

۲. کیفی: متغیر هایی که صرفاً برای دسته بندی افراد یا اشیا به کار می روند و لزوماً مقدار عددی نمی گیرند.



نسبی	فاصله ای	ترتیبی	اسمی
این مقیاس برای متغیر هایی است که علاوه بر قابلیت مرتب سازی و اختلاف بیم مقادیر، نسبت نیز با معناست و اغلب متغیر های فیزیکی مانند نمره، وزن، قد و تعداد از این جمله هستند.	این مقیاس برای متغیر هایی است که علاوه بر قابلیت مرتب سازی اختلاف بین داده ها نیز با معناست ولی نسبت معنایی ندارد و صفر آن قراردادی است، یعنی صفر به معنای نبود یک ویژگی در فرد یا شی نیست. مانند درجه حرارت در شهر های مختلف که دمای ۱۰ و ۲۰ درجه دارای اختلاف ۱۰ درجه هستند ولی نمی توان گفت دمای دومی دو برابر اولی است.	این مقیاس با الفاظ ضمن ایجاد افکیک بین افراد و اشیا، ارجحیت نیز قائل می شود. و این متغیر ها قابل مرتب سازی هستند ولی محاسبه اختلاف بین آنها بی معناست. مانند رتبه دانش آموزان یک کلاس که به اختلاف بین رتبه اول و سوم توجه نمی کنیم.	این مقیاس برای متغیر هایی است که شامل نام ها، برجسب ها و گروه ها می شود. در اینجا هیچ معیاری که با آن بتوان داده ها را از کوچک به بزرگ مرتب کرد وجود ندارد. مانند گروه خونی، کد دانش آموزی، جنسیت فرد، پلاک منازل و ماشین ها

مثال: نوع متغیر های زیر را مشخص کنید:

۱. محسن، محمود، محمد و میثم همگی اسامی مذکر هستند.
۲. احمد رتبه بیست و پنجم، رضا رتبه نوزدهم و صادق رتبه دهم را کسب کرده است و رتبه یک بالاترین رتبه است.
۳. دما بدن ماهی قزل آلاهی رودخانه هراز
۴. طول ماهی های قزل آلاهی رودخانه هراز

مثال : موضوعات زیر دارای چه مقیاسی هستند .

۱. مدت زمان پاسخ گویی به سوالات یک امتحان
۲. زمان اولین کلاس
۳. رشته تحصیلی
۴. مقیاس ارزیابی تحصیلی : ضعیف ، معمولی و خوب
۵. نمره آخرین آزمون (از ۱۰۰ امتیاز)
۶. سن دانش آموزان

درس دوم : معیار های گرایش به مرکز

میانگین : یک شاخص عددی است که متوسط داده ها را بیان می کند .
$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

مسئله : مدیر مدرسه ای بر اساس اطلاعات سال های گذشته می گوید که معمولاً خیرین به طور متوسط ۱۰ درصد درآمد سالانه خود را به این امر اختصاص می دهند . فرض کنید در آمد ماهیانه حضار در انجمن خیریه این مدرسه به صورت زیر باشد ، کمک خیرین امسال چقدر خواهد بود ؟

درآمد	نجمه	سبحان	رسول	حسنا	جوانه	احمد	آرمان
(میلیون ریال)	۴۰	۱۲	۲۸	۳۲	۳۰	۲۲	۲۵

حل : میانگین در آمد های ماهانه خیرین برابر است با :
$$\bar{X} = \frac{۲۵ + ۲۲ + ۳۰ + ۳۲ + ۲۸ + ۱۲ + ۴۰}{۷} = \frac{۱۸۹}{۷} = ۲۷$$

یعنی به طور متوسط درآمد ماهانه این افراد ۲۷ میلیون تومان است که در سال برابر است با : $۲۷ \times ۱۲ = ۳۲۴$

و ۱۰ درصد آن برابر است با $۳۲/۴$.

مثال : میانگین داده های ۵۰ و ۴۰ و ۳۰ و ۲۰ و ۱۰ چقدر است ؟

مثال : اگر میانگین داده های ۴۰ و x و ۱۰ برابر ۳۰ شود . مقدار x چقدر است ؟

دور افتاده: داده ای که با سایر داده ها تفاوت اساسی دارد. یعنی بسیار بزرگ تر یا بسیار کوچک تر از بقیه داده هاست.

به عنوان مثال اگر در مسأله قبل یک میلیارد در با درآمد ماهیانه ۱ میلیارد تومان به انجمن خیریه مدرسه بپیوندند میانگین درآمد ها تا حدود ۱۴۸ میلیون تومان در ماه بالا می رود که غیر واقعی به نظر می رسد زیرا این عدد به هیچ وجه بیان گر متوسط درآمد ها نیست.

میانه: شاخصی برای بدست آوردن متوسط داده ها که در واقع عدد وسط در بین داده های مرتب شده است و معمولاً در مواردی استفاده می شود که داده دور افتاده داشته باشیم.

نکته: اگر تعداد داده ها زوج باشد، عدد وسط نداریم و میانه برابر با میانگین دو داده وسط است.

مسأله: در مسأله خیرین، پس از ورود میلیارد، میانه را حساب کنید.

حل: داده ها را به صورت مرتب شده می نویسیم و میانه برابر است با: $\frac{28+30}{2} = 29$

۱۰۰۰	۴۰	۳۲	۳۰	۲۸	۲۵	۲۲	۱۲
------	----	----	----	----	----	----	----

مثال: میانه داده های ۱ و ۹۹ و ۶۸ و ۲ و ۸۶ و ۱۴ و ۱۰ و ۱۰ چقدر است؟

مثال: میانه داده های ۱ و ۹۹ و ۶۸ و ۲ و ۸۶ و ۱۴ و ۱۰ و ۱۱ چقدر است؟

درس سوم: معیارهای پراکندگی

معیارهایی چون میانگین و میانه به تنهایی نمی توانند اطلاعات کاملی در مورد داده ها به ما بدهند مخصوصاً در مورد مقایسه چند گروه که تقریباً دارای شاخص های مرکزی برابر هستند. بنابراین شاخصی را نیاز داریم که میزان پراکندگی داده ها را مشخص کند.

انحراف معیار: این شاخص میزان انحرافات داده ها از میانگین را نشان می دهد و با علامت « زیگما » نمایش داده می شود.

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

مسئله : از سه کلاس دهم در سه دبیرستان آزمونی گرفته شد و از هر کلاس ۱۰ نفر به تصادف انتخاب گردید که نمرات آزمون آنها به ترتیب زیر است .

(الف) {۶۵,۷۵,۷۳,۵۰,۶۰,۶۴,۶۹,۶۲,۶۷,۸۵}

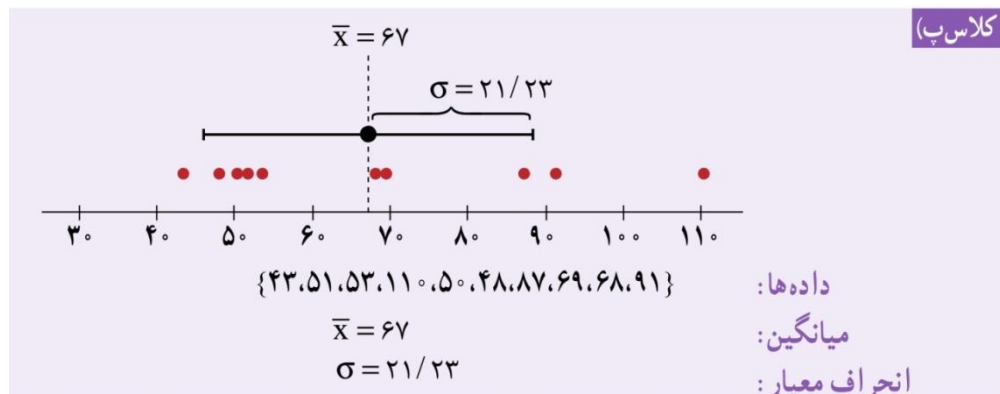
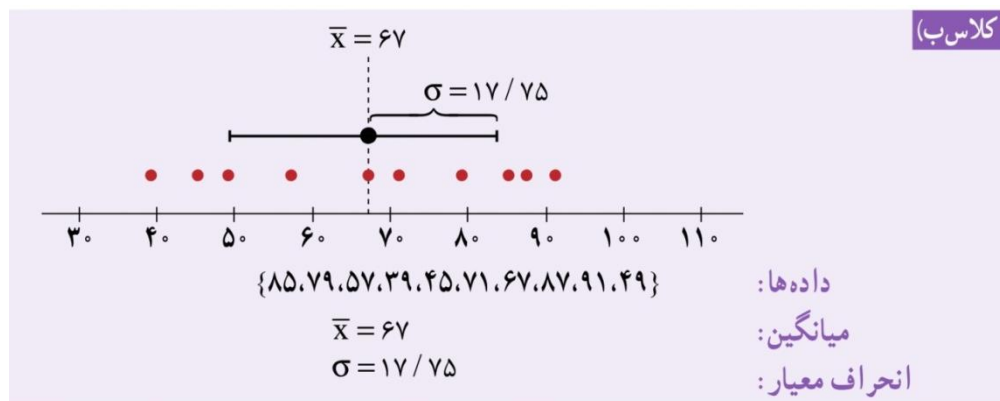
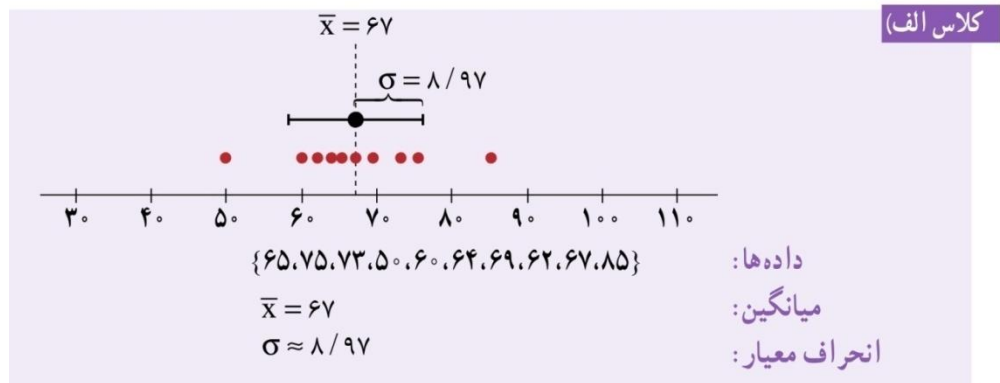
(ب) {۸۵,۷۹,۵۷,۳۹,۴۵,۷۱,۶۷,۸۷,۹۱,۴۹}

(پ) {۴۳,۵۱,۵۳,۱۱۰,۵۰,۴۸,۸۷,۶۹,۶۸,۹۱}

اگر قرار باشد والدینی فرزند بازیگوش خود را در یکی از این مدارس ثبت نام کند کدام را توصیه می کنید ؟

اگر قرار باشد والدینی فرزند خود را به قصد شرکت در المپیاد ، در یکی از این مدارس ثبت نام کند کدام را توصیه می کنید ؟

حل : میانگین هر سه کلاس برابر ۶۷ است . پس معیار میانگین اطلاع دقیقی از تفاوت سه کلاس به ما نمی دهد ولی با محاسبه انحراف معیار سه کلاس داریم .



برای ثبت نام فرزند بازیگوش مدرسه ای بهتر است که پراکندگی نمرات در آن کمتر است زیرا سطح دانش این دانش آموزان نزدیک به هم است و در مقابل برای فرزندی که به قصد شرکت در المپیاد می خواهد مدرسه را انتخاب کند ، مدرسه ای بهتر است که پراکندگی نمرات در آن بیشتر باشد و این فرزند می تواند توانایی و تفاوت خود را به آسانی نشان دهد .

نکته : طبق یک قانون در آمار

۱. تقریباً ۹۹/۹ درصد داده ها بین سه برابر انحراف معیار از میانگین قرار دارند : $(\bar{X} - 3\sigma, \bar{X} + 3\sigma)$

۲. تقریباً ۹۶ درصد داده ها بین دو برابر انحراف معیار از میانگین قرار دارند : $(\bar{X} - 2\sigma, \bar{X} + 2\sigma)$

۳. تقریباً ۶۸ درصد داده ها بین یک برابر انحراف معیار از میانگین قرار دارند : $(\bar{X} - \sigma, \bar{X} + \sigma)$

مثال : در مثال خیرین در درس قبل ، به کمک جدول زیر مقدار انحراف معیار را قبل و بعد از ورود فرد میلیاردر بدست آورید .

مشاهدات		انحراف مشاهدات از میانگین		مربع انحراف مشاهدات از میانگین	
قبل از ورود میلیاردر	بعد از ورود میلیاردر	قبل از ورود میلیاردر	بعد از ورود میلیاردر	قبل از ورود میلیاردر	بعد از ورود میلیاردر
۴۰	۴۰				
۱۲	۱۲				
۲۸	۲۸				
۳۲	۳۲				
۳۰	۳۰				
۲۲	۲۲				
۲۵	۲۵				
	۱۰۰۰				

همان طور که دیدید انحراف معیار قبل و بعد از ورود میلیاردر اختلاف زیادی دارد و این به دلیل داده دور افتاده است . هرگاه داده دور افتاده داشته باشیم از شاخص پراکندگی دیگری به نام **دامنه میان چارکی** استفاده می کنیم .

دامنه میان چارکی : تفاضل چارک اول از چارک سوم را می گویند . (چارک اول در واقع میانه داده های قبل از چارک سوم میانه داده های بعد از میانه است)

مثال : جدول زیر را برای مثال خیرین مدرسه تکمیل کنید .

σ انحراف معیار	$IQR = Q_3 - Q_1$ دامنه میان چارکی	Q_3 چارک سوم	Q_2 چارک دوم	Q_1 چارک اول	
					درآمد ماهیانه اعضای خیریه
					درآمد بعد از ورود میلیاردر

تمرین :

جملات زیر را کامل کنید .

۱. و انواعی از معیار های گرایش به مرکز هستند .
۲. شما معمولاً نه تنها معدل یک مجموعه را می خواهید ، بلکه میزان تغییرات حوالی آن نقطه را هم نیاز دارید که آن معیار است .
۳. معیار پراکندگی که معمولاً با میانگین بیان می شود است .
۴. معیار پراکندگی که معمولاً با میانه بیان می شود است .
۵. آماری که بریا توصیف داده ها ، میانگین ، میانه ، انحراف معیار و دامنه میان چارکی به کار می رود ، نام دارد .
۶. ۵۰ درصد داده ها قبل از و ۵۰ درصد بعد از هستند .
۷. ۷۵ درصد داده ها قبل از و یا بعد از هستند .
۸. ۲۵ درصد داده ها قبل از و یا بعد از هستند .
۹. ۵۰ درصد داده ها بین و قرار دارند .

کار عملی :

نمرات درس فیزیک دانش آموزی در طول سال به صورت ۱۷ و ۱۹ و ۱۸ و ۷ و ۱۸ و ۱۹ است .

الف) میانه و میانگین را حساب کنید . کدام شاخص بیان گر متوسط نمرات این دانش آموز در درس فیزیک است ؟

ب) اگر برای جبران نمره ۷ اجازه دهند امتحان دیگری بدهد . چه نمره ای باید بگیرد تا میانگین نمرات او بیشتر از ۱۸ شود ؟

ج) انحراف معیار و دامنه میان چارکی را بیابید . کدام شاخص بهتری برای بیان میزان پراکندگی نمرات اوست ؟