

اگرچه نیت خوبی است زیستن ...
اما خوشا که دست به تصمیم بهتری بزنیم!

 www.konkursara.com

 ۰۲۱۵۵۷۵۶۵۰۰

دانلود بهترین جزوات در

کنکورسرا

کنکورسرا

مرجع تخصصی قبولی آزمون فرهنگیان و آزمون استخدامی آموزش و پرورش

جزوه آموزش

آمار و احتمال

یازدهم ریاضی

فصل سوم: آمار توصیفی

درس اول : توصیف و نمایش داده ها

درس دوم : معیار های گرایش به مرکز

درس سوم : معیار های پراکندگی

درس اول : توصیف و نمایش داده ها

یاد آوری :

علم آمار : مجموعه ای از روش ها شامل جمع آوری اعداد ، سازماندهی و نمایش ، تحلیل و تفسیر و نتیجه گیری است .

جامعه : مجموعه همه افراد یا اشیاء مورد مطالعه در آمار را جامعه می گویند . به تعداد اعضای جامعه اندازه جامعه گفته می شود .

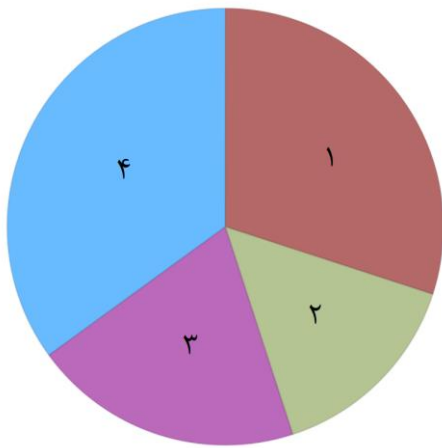
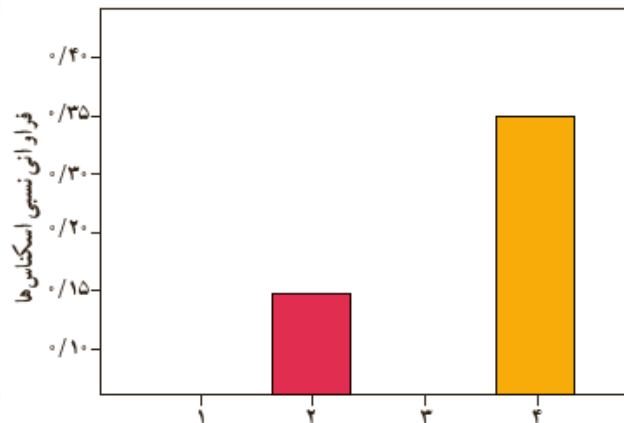
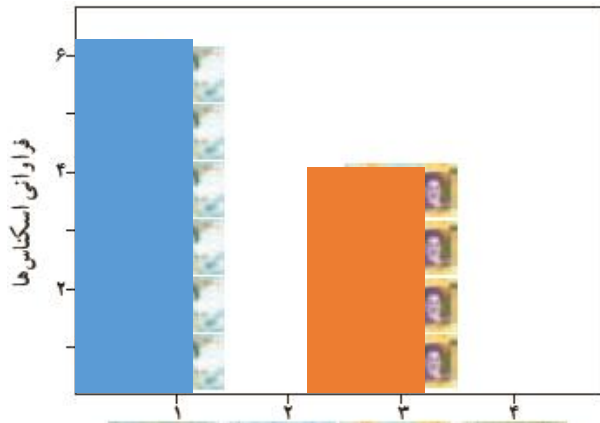
نمونه : بخشی از جامعه که برای مطالعه انتخاب می شود را نمونه می گویند . به تعداد اعضای نمونه اندازه نمونه گفته می شود .

بعد از گرد آوری داده ها به تنظیم و خلاصه کردن آنها در قالب جدول (جدول فراوانی) یا رسم نمودار می پردازیم . قبل از این کار بهتر است با مفاهیم زیر آشنا شوید :

داده ها : واقعیت هایی درباره یک شیء یا فردند که در محاسبه، برنامه ریزی و پیش بینی به کار می روند .
 متغیر : هر ویژگی از اشیا یا اشخاص، که در اعضای جامعه یکسان نیست و معمولاً از یک عضو به عضو دیگر تغییر می کند را **متغیر** می گویند و عددی که به آن ویژگی یک عضو نسبت داده می شود را **مقدار متغیر**، یا **مشاهده** می گویند .
 فراوانی یک داده : تعداد دفعاتی که هر داده مشاهده می شود را **فراوانی** آن داده می گویند .
 فراوانی نسبی یک داده : با تقسیم فراوانی هر داده به تعداد کل داده ها، **فراوانی نسبی** آن داده به دست می آید .
 اگر فراوانی نسبی داده ها در ۱۰۰ ضرب شود، آن گاه درصد داده ها به دست می آید .

تمرین : یک راننده تاکسی در یک روز ۶ عدد اسکناس هزاری ، ۳ عدد اسکناس دو هزاری ، ۴ عدد اسکناس پنج هزاری و ۷ عدد اسکناس ده هزار تومانی از مسافران دریافت کرده است. جدول فراوانی و نمودار میل ای و دایره ای زیر را برای آن کامل کنید .

نوع اسکناس	شماره	فراوانی	فراوانی نسبی
هزاری	۱	۶	۰/۳
دو هزاری	۲		
پنج هزاری	۳		۰/۲
ده هزاری	۴	۷	
تعداد کل اسکناس ها		۲۰	



در نمودار دایره ای مربوط به جدول فراوانی داده شده

هر قسمت چند درجه است ؟

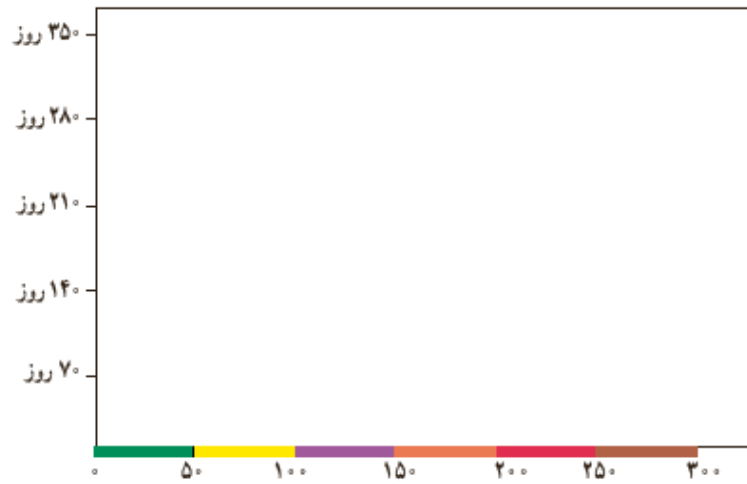
(رانمایی : مثلاً قسمت ۳ دارای فراوانی نسبی ۰/۲ است

یعنی ۰/۲ از کل دایره مربوط به آن است)

تمرین : در سال ۹۳ میزان آلاینده‌گی هوا در طول سال به صورت جدول زیر داده شده است .

وضعیت هوا	شاخص کیفیت هوا	فراوانی	فراوانی نسبی
پاک	$0 \leq AQI \leq 50$	۱۶	
سالم	$50 < AQI \leq 100$	۲۳۳	
ناسالم برای گروه‌های حساس	$100 < AQI \leq 150$	۱۱۲	
ناسالم	$150 < AQI \leq 200$	۴	
بسیار ناسالم	$200 < AQI \leq 250$	۱۰	
خطرناک	$250 < AQI \leq 300$	۰	
تعداد کل روزهای یک سال		۳۶۵	

الف) نمودار فراوانی زیر را برای آن رسم کنید .



ب) نمودار فراوانی نسبی آن را رسم کنید .

ج) چند درصد از روز های سال هوا سالم بوده است ؟

تکلیف : تمرین صفحه ۸۱ و ۸۲ را حل کنید .

درس دوم : معیار های گرایش به مرکز

در آمار توصیفی علاوه بر طبقه بندی اطلاعات و نمایش آنها در قالب نمودار ، سعی می شود به کمک یک شاخص عددی ویژگی های خاص این اطلاعات به صورت خلاصه بیان شود .

معیار های گرایش به مرکز :

شاخص های عددی هستند که میزان گرایش اعداد به یک عدد خاص را نمایش می دهند . در این جزوه ما با سه معیار مهم گرایش به مرکز به نام های میانگین و میانه و مد (نما) آشنا می شویم که بسیار پر کاربرد هستند .

میانگین : میانگین داده ها را با \bar{X} نمایش می دهیم که برابر است با :

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

که x_i ها داده ها و n تعداد داده ها است .

مسئله : تعدادی خیر به طور متوسط ۱۰ درصد در آمد ماهانه خود را به یک خیریه می دهند . فرض کنید در آمد ماهیانه حضار در انجمن خیریه این مدرسه به صورت زیر باشد ، متوسط در آمد خیریه از این خیرین در این سال چقدر خواهد بود ؟

در آمد	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 در آمد ماهانه این افراد ۲۷ میلیون ریال است و ۱۰ درصد آن برابر است با $\frac{۲}{۷}$ میلیون ریال ، معادل ۲۷۰ هزار تومان در ماه و ۳۲۴۰۰۰۰ تومان در سال .

میانگین وزنی (وزنی) : اگر داده ها دارای وزن متفاوت باشند به عبارت دیگر هر x_i دارای تعداد تکرار w_i باشد آنگاه

میانگین وزنی آنها برابر است با :

$$\bar{X}_w = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{n} = \frac{\sum_{i=1}^n w_i x_i}{n}$$

تمرین : در مساله قبل فرض کنید هر نفر فقط در چند ماه خاص قرار است ۱۰ درصد از حقوق خود را صرف خیریه کند . در زیر جدول تعداد ماه هایی که هر شخص در نظر دارد به خیریه کمک کند آمده است . میانگین درآمد خیریه از این خیرین در این سال چقدر خواهد بود ؟

آرمان	احمد	چوانه	حسنا	رسول	سبحان	نجمه	
۲۵	۲۲	۳۰	۳۲	۲۸	۱۲	۴۰	درآمد
۱۲	۴	۵	۳	۸	۵	۲	تعداد ماه

تمرین : دانش آموزی در یک آزمون نتایج زیر را دریافت کرده است و می دانیم هر درس دارای وزنی متفاوت است (ضریب) با توجه به جدول داده شده میانگین درصد مواد امتحانی او را حساب کنید .

مواد امتحانی	ریاضیات	فیزیک	شیمی	زبان انگلیسی	ادبیات و زبان فارسی	دین و زندگی
درصد	۷۱	۶۵	۸۰	۵۲	۹۵	۱۰۰
ضریب درس	۴	۳	۱	۱	۴	۳

نکته : اگر هر یک از داده های آماری با مقدار ثابتی ضرب و جمع شوند ، میانگین آنها نیز با همان مقدار ثابت ضرب و جمع خواهد شد .

تمرین : اگر میانگین تعدادی داده ۵ باشد و داده ها را ۳ برابر کرده و ۵ واحد از آنها کم کنیم میانگین داده های جدید چقدر است ؟

تمرین : میانگین داده های زیر را بیابید .

۱۰۳, ۹۸, ۱۰۵, ۱۱۱, ۹۰, ۹۲, ۱۰۰, ۹۹

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 خیرین ، شخصی با درآمد ماهیانه ۱ میلیارد ریال به انجمن خیریه مدرسه بپیوندد میانگین درآمد ها تا حدود ۱۴۸ میلیون ریال در ماه بالا می رود که غیر واقعی به نظر می رسد زیرا این عدد به هیچ وجه بیان گر متوسط درآمد ها نیست . در صورتی که داده دور افتاده داشته باشیم از شاخص دیگری برای گرایش به مرکز استفاده می شود .

میانگین : 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 . میان را با Q نمایش می دهیم .

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

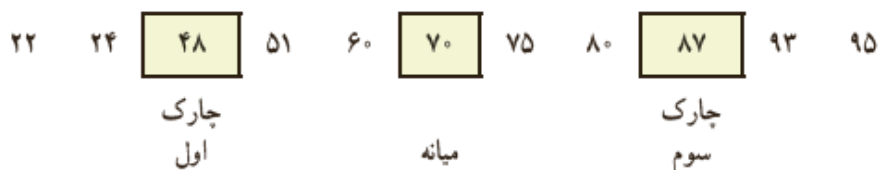
تمرین : میان داده های ۱ و ۹۹ و ۶۸ و ۲ و ۸۶ و ۱۴ و ۱۰ و ۱۱ چقدر است ؟

تمرین : داده های زیر مربوط به تعداد ضریان قلب ۱۲ دانش آموز پایه یازدهم ، قبل از مسابقه دو است . میان و میانگین داده های زیر را مشخص کنید .

۱۰۰ ۹۱ ۸۳ ۷۵ ۱۰۵ ۹۸ ۹۸ ۱۰۱ ۸۹ ۹۳ ۹۷ ۸۶

چارک : مقادیری که داده های مرتب شده را به چهار قسمت مساوی تقسیم می کنند چارک نامیده می شوند . بدیهی است که چارک دوم همان میان است و چارک اول و سوم ، میان داده های سمت چپ و راست میان هستند .

مثال : به داده های زیر توجه کنید . ۷۰ میان است و میان اعداد قبل و بعد از آن به ترتیب ۴۸ و ۸۷ هستند که چارک اول و سوم نامیده می شوند .



تمرین : جاهای خالی را پر کنید .

- ۵۰ درصد داده ها قبل از و ۵۰ درصد بعد از هستند .
- ۷۵ درصد داده ها قبل از و یا بعد از هستند .
- ۲۵ درصد داده ها قبل از و یا بعد از هستند .
- ۵۰ درصد داده ها بین و قرار دارند .

هده یا فها : داده ای که بیشترین تکرار را داشته باشد مد نامیده می شود . اگر همه داده ها دارای تکراریکسان باشند این داده ها دارای مد نیستند و اگر دو داده دارای بیشترین تکرار باشند داده ها دو مدی هستند .

تمرین : در یک مسابقه دات از ۱۰ پرتاب ، امتیازات زیر برای سه نفر حاصل شده است .

۸	۸	۹	۱۰	۹	۵	۷	۱۰	۹	۱۰	نفر اول
۷	۴	۵	۳	۲	۱	۶	۸	۹	۱۰	نفر دوم
۷	۴	۵	۹	۱۰	۱۰	۷	۹	۹	۹	نفر سوم

(الف) مد نفر اول چیست ؟

(ب) مد نفر دوم چیست ؟

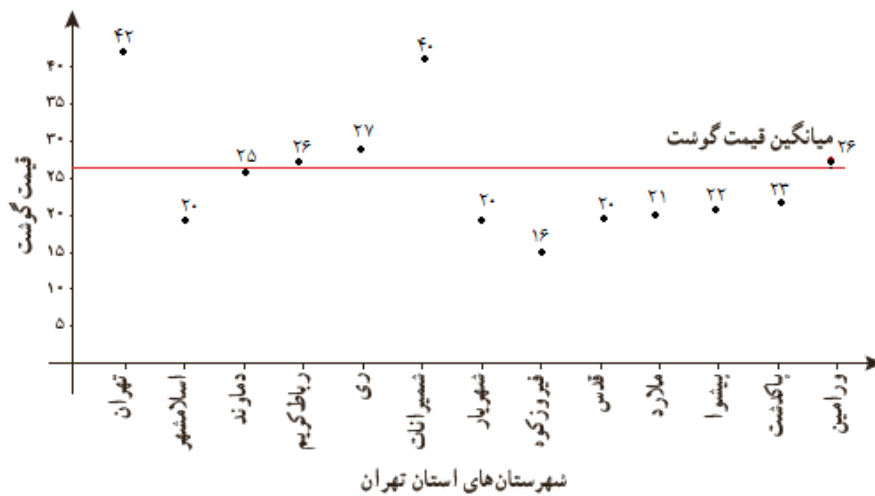
(ج) مد نفر سوم چیست ؟

تکلیف : تمرین صفحه ۹۰ و ۹۱ را حل کنید .

دریس سوم : معیار های پراکندگی

معیار هایی چون میانگین و میان 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 . برای روشن تر شدن موضوع به مثال زیر دقت کنید .

مثال : قیمت گوشت در سال ۹۵ در تهران و شهرستان های آن در نمودار زیر به صورت نقطه آمده و میانگین قیمت گوشت با خط قرمز نمایش داده شده است .



اگر بخواهیم میزان پراکندگی هر کدام از شهر ها را از میانگین اندازه بگیریم $(x_i - \bar{x})$ که به آن انحراف از میانگین گفته می شود مجموع این انحراف ها صفر خواهد بود (همیشه !!!) پس نمی توان به معیاری دست یافت . اگر بخواهیم قدر مطلق آنها را در نظر بگیریم این مشکل رفع می شود و می توان با میانگین گرفتن از اعداد حاصل به یک معیار پراکندگی رسید ولی چون کار با قدر مطلق سخت است مجبور انحراف از میانگین گزینه بهتری خواهد بود یعنی به جای اینکه از $|x_i - \bar{x}|$ ها میانگین بگیریم از $(x_i - \bar{x})^2$ ها میانگین بگیریم .

وار یافتی : میانگین مجذور انحراف داده ها از میانگین را واریانس می نامند و با σ^2 نمایش می دهند .

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

انحراف معیار : بدلیل آنکه مقدار بدست از واریانس بزرگ تر از پراکندگی قابل قبول است و همچنین واحد آن با واحد داده ها برابر نیست از جذر آن که انحراف معیار نامیده می شود بیشتر استفاده می کنند .

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

انحراف معیار

مسئله: از دو کلاس دهم 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}

برای یک معلم تدریس در کدام کلاس بهتر است؟

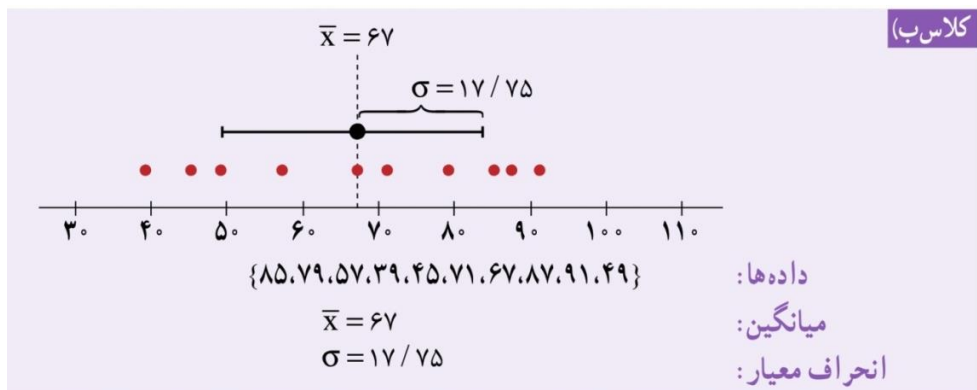
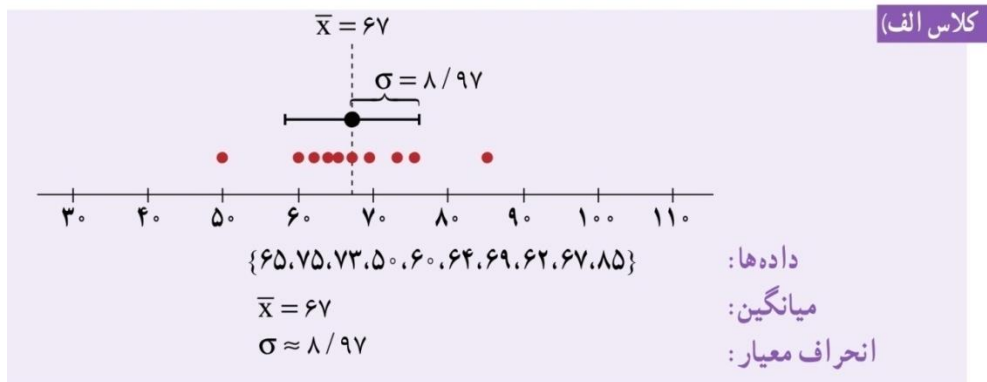
حل: میانگین هر دو کلاس برابر ۶۷ است.

$$\bar{X} = \frac{۶۷۰}{۱۰} = ۶۷$$

پس معیار میانگین اطلاع دقیقی از تفاوت دو کلاس به ما نمی دهد ولی با محاسبه انحراف معیار دو کلاس داریم.

$$\sigma^2 = \frac{۲^2 + ۸^2 + ۶^2 + ۱۷^2 + ۷^2 + ۳^2 + ۲^2 + ۵^2 + ۰^2 + ۱۸^2}{۱۰} = \frac{۸۰۴}{۱۰} = ۸۰/۴ \Rightarrow \sigma = \sqrt{۸۰/۴} \approx ۸/۹۷$$

$$\sigma^2 = \frac{۱۸^2 + ۱۳^2 + ۱۰^2 + ۲۸^2 + ۲۳^2 + ۳^2 + ۰^2 + ۲۰^2 + ۲۴^2 + ۱۸^2}{۱۰} = \frac{۳۱۴۵}{۱۰} = ۳۱۴/۵ \Rightarrow \sigma = \sqrt{۳۱۴/۵} \approx ۱۷/۷۵$$



تمرین : انحراف معیار داده های مربوط به قیمت گوشت در مثال اول را بدست آورید .

ویژگی های واریانس و انحراف معیار :

اگر داده ها با عدد ثابتی جمع شوند واریانس و انحراف معیار آنها تغییر نمی کند .

اگر داده ها در عدد ثابتی مثل a ضرب شوند ، واریانس آنها در a^2 و انحراف معیار در $|a|$ ضرب می شود .

تمرین : واریانس چند داده برابر ۵ است ، اگر داده ها را -2 برابر کرده و 8 واحد به آنها بیافزاییم ، واریانس و انحراف معیار داده های جدید چقدر است ؟

شرط تعمیم : شاخصی بدون واحد است که از تقسیم انحراف معیار بر میانگین داده ها بدست می آید و به همین دلیل می تواند شاخصی خوبی برای مقایسه دو جامعه باشد . هرچه این شاخص کوچکتر باشد ، جامعه بهتر است .

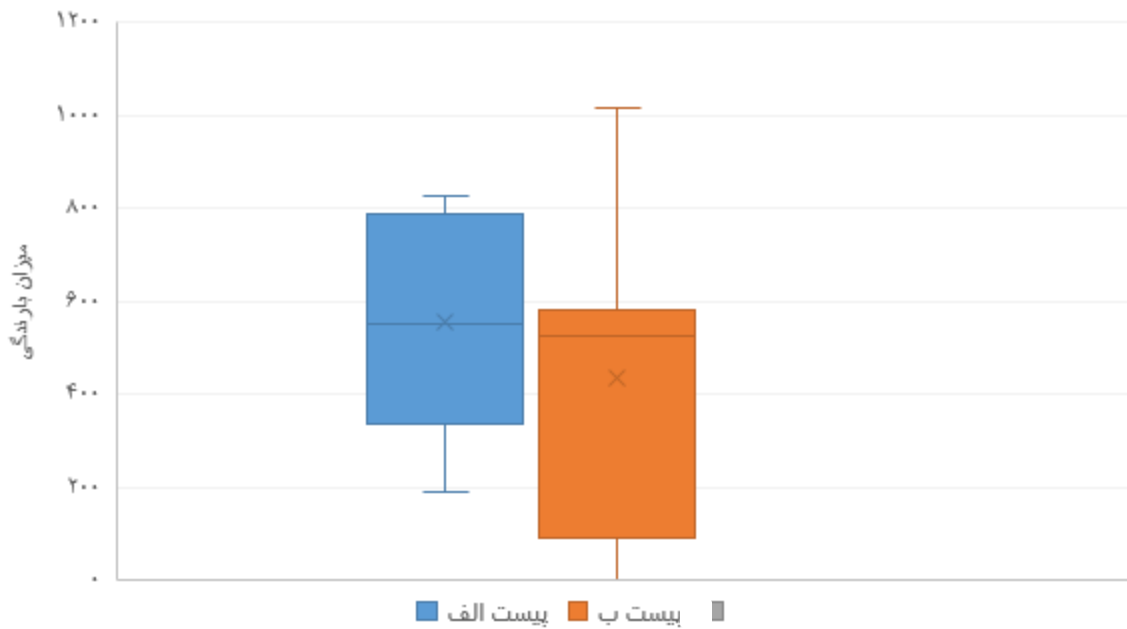
$$CV = \frac{\sigma}{\bar{X}}$$

تمرین : داده های مربوط به مقدار تولید دو شرکت در ماه گذشته به گونه ای است که میانگین اولی 1200 و میانگین دومی 1100 بوده و انحراف معیار آنها نیز به ترتیب 100 و 80 است . کدام شرکت عملکرد بهتری دارد ؟

نمودار چعبه ای : نموداری که نشان می دهد اکثر داده ها در چه محدوده ای قرار گرفته اند (دامنه میان چارکی IQR) و شامل اطلاعاتی مانند کوچکترین و بزرگ ترین داده ، میانه ، چارک اول و سوم است . به این ترتیب که مستطیلی در بازه بین چارک اول و سوم روی محور رسم می شود و مقدار میانه درون مستطیل مشخص می گردد .

مثال : نتایج ۷ سال بارندگی در ۲ پیست اسکی به صورت زیر ثبت شده است .

سال	۱۳۸۸	۱۳۸۹	۱۳۹۰	۱۳۹۱	۱۳۹۲	۱۳۹۳	۱۳۹۴
میزان بارش برف در پیست اسکی الف	۵۵۱	۱۹۰	۳۳۵	۷۸۷	۴۷۲	۷۲۸	۸۲۵
میزان بارش برف در پیست اسکی ب	۲۷۱	۰	۵۲۵	۱۰۱۶	۹۳	۵۸۱	۵۶۶



تکلیف : تمرین های صفحه ۹۹ و ۱۰۰ را حل کنید.