

۱- انحراف معیار و واریانس داده‌ها

فعالیت

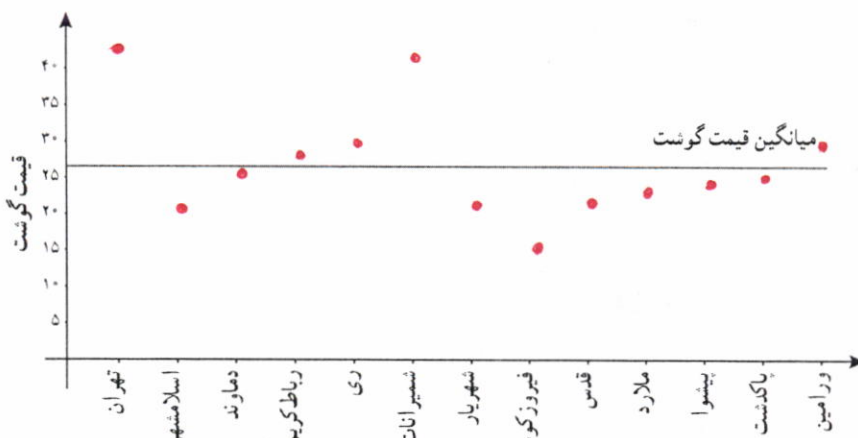


در اقتصاد هر کشوری شاخصی تحت عنوان نرخ تورم، نقش بسیار مهمی را ایفا می‌کند. یکی از اقلام مصرفی مورد نیاز در محاسبه نرخ تورم در یک کشور، قیمت گوشت قرمز است. در جدول روبه‌رو قیمت گوشت قرمز در سال ۱۳۹۵ در شهرستان‌های استان تهران گردآوری شده است.

■ میانگین قیمت گوشت قرمز در شهرستان‌های استان تهران را به دست آوردید؟ $\bar{x} = ۲۵,۲۳$

■ در نمودار زیر، میانگین قیمت گوشت قرمز در شهرستان‌های استان تهران نشان داده شده است. قیمت گوشت قرمز در هر یک از شهرستان‌های استان تهران را با کشیدن نقطه روی نمودار مشخص کنید.

شهرستان‌های استان تهران	قیمت گوشت قرمز (هزار تومان)
تهران	۴۲
اسلام شهر	۲۰
دماوند	۲۵
رباط کریم	۲۶
ری	۲۷
شمیرانات	۴۰
شهریار	۲۰
فیروزکوه	۱۶
قدس	۲۰
ملارد	۲۱
پیشوا	۲۲
پاکدشت	۲۳
ورامین	۲۶



شهرستان‌های استان تهران

میانگین قیمت گوشت $\bar{x} = \frac{۳۲۸}{۱۳} = ۲۵,۲۳$

صیغ کرام

۸ نقطه

۵ نقطه

۱ چند نقطه بالای خط قرمز، چند نقطه پایین خط قرمز و چند نقطه روی خط قرمز قرار دارند؟

۲ منظور از پراکندگی قیمت گوشت قرمز یعنی اینکه قیمت گوشت قرمز در هریک از شهرستان‌های استان تهران چقدر از میانگین قیمت دورتر است. هر چقدر نقاط یا همان قیمت گوشت قرمز در هریک از شهرستان‌های استان تهران حول خط قرمز با همان میانگین قیمت گوشت قرمز نزدیک‌تر باشند، نشان‌دهنده چیست؟ هر چقدر دورتر باشند چطور؟ **پراکندگی قیمت‌ها**

۳ معیاری را برای اندازه‌گیری پراکندگی قیمت گوشت قرمز یا همان نقاط حول خط قرمز می‌توانید معرفی کنید؟

انحراف معیار

دیدیم پراکندگی قیمت گوشت قرمز یعنی اینکه قیمت گوشت قرمز در هریک از شهرستان‌های استان تهران چقدر از میانگین قیمت دورتر است. برای معرفی معیار مناسب یک راه حل ابتدایی این است که تک تک قیمت‌ها را از میانگینشان کم کنیم. این تفاضل‌ها را انحراف از میانگین می‌نامیم. مجموع انحراف از میانگین‌ها برابر با صفر خواهد شد و این به دلیل آن است که برخی از داده‌ها از میانگین بزرگ‌تر و برخی دیگر کوچک‌ترند در نتیجه مقادیر مثبت و منفی حاصل می‌شوند که مجموع آنها همدیگر را خنثی می‌کنند. برای رفع این مشکل، قدر مطلق انحراف از میانگین داده‌ها در نظر گرفته می‌شود. میانگین این مقادیر می‌تواند معیاری برای سنجش پراکندگی داده‌ها باشد، اما کار کردن با قدر مطلق کار آسانی نیست. از این رو، توان دوم انحراف از میانگین داده‌ها در نظر گرفته می‌شود.

در آمار، یک معیار سنجش برای میزان پراکندگی داده‌ها حول میانگینشان، انحراف معیار است.

انحراف معیار به صورت زیر محاسبه می‌شود:

انحراف معیار داده‌ها: اگر n داده از جامعه به صورت x_1, x_2, \dots, x_n داشته باشیم، انحراف آنها را با نماد σ نشان می‌دهیم، که به صورت زیر محاسبه می‌شود:

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

که در آن $x_i - \bar{x}$ را انحراف داده i ام از میانگین داده‌ها می‌گویند.

واریانس داده‌ها: توان دوم انحراف معیار داده‌ها را واریانس داده‌ها گویند و آن را با نماد σ^2 نشان می‌دهیم.

اگر انحراف معیار مجموعه داده‌ها عدد کوچکی باشد، بدین معناست که پراکندگی داده‌ها حول میانگینشان کم و در نتیجه داده‌ها به هم نزدیک‌تر است و اگر انحراف معیار مجموعه داده‌ها عددی بزرگ باشد، بدین معناست که پراکندگی داده‌ها حول میانگینشان زیاد و در نتیجه داده‌ها از هم دورتر است.

مفادیده است که ابتدا واریانس تعریف شود و سپس به کمک آن تعریف انحراف معیار گفته شود.

انحراف معیار و واریانس مربوط به داده‌های قیمت گوشت قرمز در شهرستان‌های تهران را می‌توانید با تکمیل جدول روبه‌رو محاسبه کنید.

$(x_i - \bar{x})^2$	$x_i - \bar{x}$	قیمت گوشت قرمز
۲۸۹	$۴۲ - ۲۵ = ۱۷$	۴۲
۲۵	$۲۰ - ۲۵ = -۵$	۲۰
۰	$۲۵ - ۲۵ = ۰$	۲۵
۱	$۲۶ - ۲۵ = ۱$	۲۶
۴	$۲۷ - ۲۵ = ۲$	۲۷
۲۲۵	$۴۰ - ۲۵ = ۱۵$	۴۰
۲۵	$۲۰ - ۲۵ = -۵$	۲۰
۸۱	$۱۶ - ۲۵ = -۹$	۱۶
۲۵	$۲۰ - ۲۵ = -۵$	۲۰
۱۶	$۲۱ - ۲۵ = -۴$	۲۱
۹	$۲۲ - ۲۵ = -۳$	۲۲
۴	$۲۳ - ۲۵ = -۲$	۲۳
۱	$۲۶ - ۲۵ = ۱$	۲۶
۷,۳۶		σ
۵۴,۲۳		σ^2

$\sum (x_i - \bar{x})^2 = ۷۰۵$

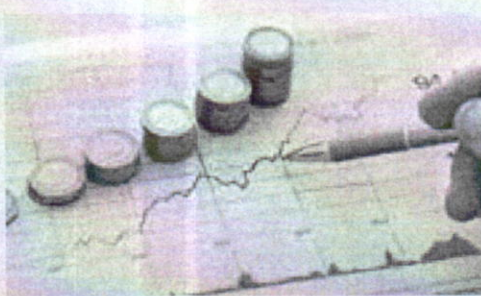
$\bar{x} = ۲۵$ فراموش نکنید

$\sigma^2 = \frac{۷۰۵}{۱۳} = ۵۴,۲۳$

$\sigma = \sqrt{۵۴,۲۳} = ۷,۳۶$

خواندنی

نرخ تورّم



در علوم اقتصادی با استفاده از علم آمار، شاخصی تحت عنوان نرخ تورّم بیان می‌شود. نرخ تورّم، درصد تغییر سطح قیمت مجموعه کالاهای مصرفی مانند خوراک و پوشاک و کالاهای خدماتی مانند مسکن، آب و برق خانوارها در طول زمان را اندازه می‌گیرد. فرض کنید متوسط قیمت مجموعه کالاهای مصرفی

یک خانوار در سال a ، P_1 و متوسط قیمت همان مجموعه کالای مصرفی در سال $a-1$ ، P_0 باشد، در این صورت نرخ تورّم در طی سال a به صورت زیر به دست می‌آید:

نرخ تورّم = $\frac{P_1 - P_0}{P_0} \times ۱۰۰\%$

به عنوان مثال، اگر متوسط قیمت گوشت قرمز به عنوان کالای مصرفی در سال $a-1$ و a به ترتیب، ۲۸ و ۳۲ هزار تومان برای هر کیلو باشد، در این صورت نرخ تورّم برای قیمت گوشت قرمز در سال a برابر:

نرخ تورّم = $\frac{(۳۲ - ۲۸)}{۲۸} \times ۱۰۰ = ۱۴\%$

توجه کننده:

گروه ریاضی دوره ی دوم متوسطه و انجمن مطلقان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

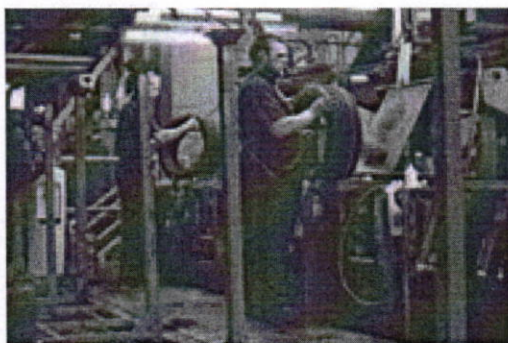
یعنی متوسط قیمت گوشت قرمز در سال a ، ۱۴ درصد نسبت به سال گذشته افزایش یافته است.



لازم به ذکر است هر چقدر نرخ تورم افزایش یابد، قدرت خرید مردم کاهش پیدا می کند. همچنین مرکز آمار ایران برای محاسبه نرخ تورم در یک سال، متوسط قیمت ۱۰۰ قلم کالای گروه خوراکی ها و آشامیدنی ها و ۲۵۹ قلم کالای خدماتی برای سال جاری و سال قبل آن در نظر گرفته و این نرخ را محاسبه می کند.

۲- ضریب تغییرات داده ها

فعالیت



یکی از شاخص های کیفیت در لاستیک های تولید شده اتومبیل توسط یک کارخانه، طول عمر آن لاستیک هاست. هرچقدر متوسط طول عمر لاستیک های تولیدی بیشتر و انحراف معیار طول عمر لاستیک ها کمتر باشد، به این معناست که لاستیک ها کیفیت بالایی از نظر طول عمر دارند.

حال با توجه به مطالب گفته شده، به بررسی کیفیت لاستیک های تولیدی از نظر طول عمر دو کارخانه (الف) و (ب) می پردازیم. براساس داده های به دست آمده میانگین طول عمر لاستیک ها در دو کارخانه و انحراف معیار آنها به شرح جدول روبه رو است:

کارخانه	میانگین	انحراف معیار
کارخانه الف	۵۴۰۰۰ کیلومتر	۵۰ کیلومتر
کارخانه ب	۶۵۰۰۰ کیلومتر	۱۰۰ کیلومتر

■ شما ترجیح می دهید از کدام کارخانه لاستیک بخرید؟
 ■ آیا می توان براساس میانگین و انحراف معیار و نمونه های در نظر گرفته شده قضاوت کرد؟

برای پاسخ به سؤالات فوق نیاز به معرفی معیار جدیدی برای سنجش پراکندگی داده وجود دارد. این معیار را ضریب تغییرات داده ها می نامند.

ضریب تغییرات داده ها: معیاری است که از تقسیم انحراف معیار داده ها (σ) به میانگین داده ها (\bar{x}) به دست می آید و آن را با نماد CV نشان می دهند.

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

هر قدر ضریب تغییرات کمتر باشد، میزان پراکندگی داده ها کمتر خواهد شد که این موضوع برای ما مطلوب است.

کار در کلاس

الف) با کامل کردن جدول زیر، ضریب تغییرات مربوط به طول عمر دو کارخانه را محاسبه کنید.

ضریب تغییرات	انحراف معیار	میانگین	کارخانه
۰/۰۰۰۹	۵۰ کیلومتر	۵۴۰۰۰ کیلومتر	کارخانه الف
۰/۰۰۱۵	۱۰۰ کیلومتر	۶۵۰۰۰ کیلومتر	کارخانه ب

محصولات کدام کارخانه را انتخاب می‌کنید؟ «الف»

ب) حال با تغییر واحد اندازه‌گیری در جدول قبلی میانگین و انحراف معیار طول عمر لاستیک‌ها در دو کارخانه (الف) و (ب) به صورت زیر گزارش داده شده است.

ضریب تغییرات	انحراف معیار	میانگین	کارخانه
۰/۰۰۰۹	۵۰۰۰۰ متر	۵۴۰۰۰۰۰۰ متر	کارخانه الف
۰/۰۰۱۵	۱۰۰۰۰۰ کیلومتر	۶۵۰۰۰ کیلومتر	کارخانه ب

همان‌طور که ملاحظه می‌کنید میانگین و انحراف معیار لاستیک‌ها برای کارخانه (الف) برحسب واحد اندازه‌گیری متر و برای کارخانه (ب) برحسب کیلومتر است. در این حالت نیز ضریب تغییرات را در جدول زیر محاسبه کنید. آیا ضریب تغییرات به واحد اندازه‌گیری وابسته است؟ **خیر، بستگی ندارد.**

نمودار جعبه‌ای

در ابتدای این درس با معیارهای پراکندگی آشنا شدیم، حال می‌خواهیم با استفاده از نمودارهای آماری، معیارهای پراکندگی داده‌ها را به صورت تصویری نشان دهیم.

فعالیت

میزان بارش برف سالانه در دو پیست اسکی «الف» و «ب» برای هفت سال اندازه‌گیری و نتایج، در جدول زیر گردآوری 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

عدد ۰ در جدول به این معناست که میزان بارش کمتر از ۱ سانتی‌متر است.

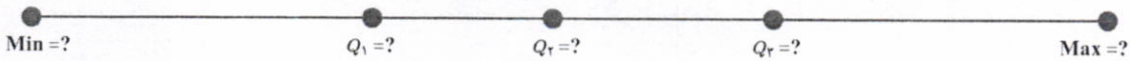


برای رسم نمودار آماری، مراحل زیر را انجام دهید.

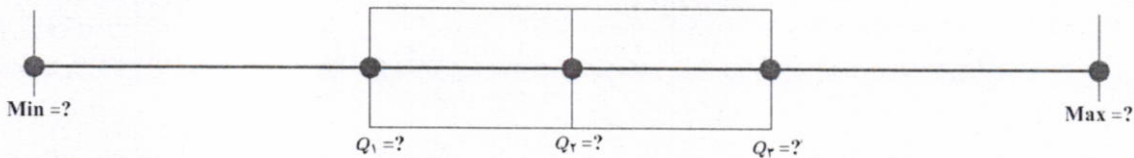
الف) جدول زیر را کامل کنید.

کمترین مقدار	چارک اول	میانه	چارک سوم	بیشترین مقدار	سال
Min	Q_1	Q_2	Q_3	Max	
۱۹۰	۳۳۵	۵۵۱	۷۸۷	۸۲۵	بیست اسکی الف

ب) حال مقادیر جدول را روی یک محور نمایش می دهیم.



پ) برای مشخص کردن حدود دامنه میان چارکی (IQR) یک جعبه به عرض دلخواه رسم می کنیم، سپس با استفاده از یک خط، میانه را در جعبه مشخص می کنیم و در انتها، از دو طرف جعبه به کمترین و بیشترین مقدار داده ها دو خط رسم می کنیم.



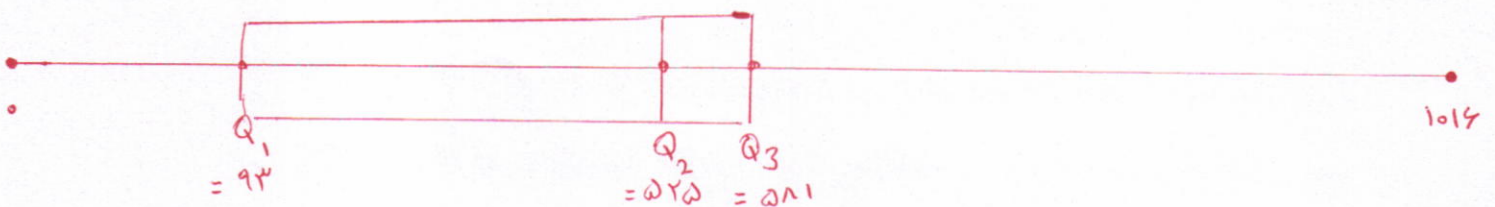
به این نمودار، نمودار جعبه ای می گوئیم. در این نمودار چارک اول، میانه، چارک سوم، بیشترین و کمترین مقدار داده ها به طور هم زمان نشان داده می شود.

کار در کلاس

۱۰۱۶، ۵۸۱، ۵۲۵، ۲۷۱، ۹۳ و ۵
 Q_1 Q_2 Q_3 Max

■ نمودار جعبه ای مربوط به بیست «ب» را رسم کنید. و سپس با نمودار جعبه ای بیست «الف» مقایسه کنید.

■ اگر داده دور افتاده ای در داده ها باشد، نمودار جعبه ای چه تغییری می کند؟ *کشیده تر می شود. (پراکندگی زیاد می شود)*



با مقایسه ای این نمودار با نمودار «الف» معلوم می شود که در نمودار «ب» پراکندگی داده ها زیاد است.

۱ فرض کنید سنّ افرادی که در یک روز سوار اتوبوس شده‌اند، به صورت زیر است :

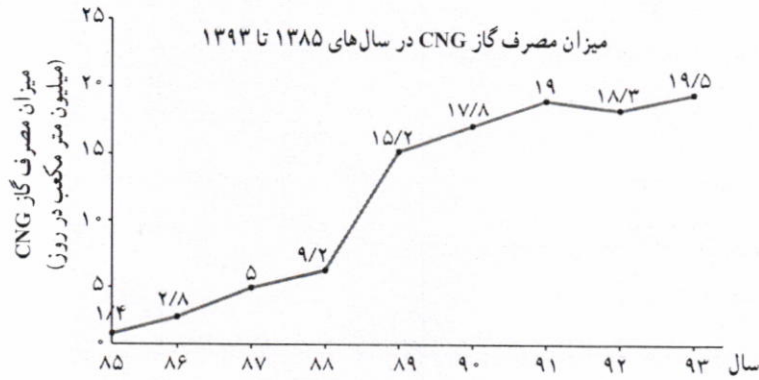
۳۲،۵۹،۲۶،۵۳،۷۴،۱۷،۴۵،۲۳،۶۴،۵۰،۶۱

انحراف معیار، واریانس و ضریب تغییرات سنّ افراد را به دست آورید.



۲ نمودار زیر میزان مصرف گاز CNG را از سال ۱۳۸۵ تا ۱۳۹۳ نشان می‌دهد. با توجه به

این نمودار انحراف معیار، واریانس و ضریب تغییرات میزان مصرف گاز CNG از سال ۱۳۸۵ تا ۱۳۹۳ را به دست آورید.



۳ انحراف معیار، واریانس و ضریب تغییرات را برای هریک از اعداد جدول زیر به دست آورید.

اعداد	انحراف معیار	واریانس	ضریب تغییرات
۱۰۰، ۱۲، ۸، ۱۶، ۱۰، ۴، ۷			
۳، ۲، ۱، ۰، ۰، ۰، -۳، -۲، -۱			
۱۰/۱۱، ۱۱/۳۶، ۱۰/۱۱			
۹/۸۸، ۹/۴۲، ۹/۷۶، ۹/۶۲			
۲، ۳۰۰۰، ۲۵۰۰، ۲۰۰۰			

کلاس اول

۴ اعداد دلخواه را در جدول زیر بنویسید و انحراف معیار، واریانس و ضریب تغییرات را برای هریک از اعداد به دست آورید.

اعداد	انحراف معیار	واریانس	ضریب تغییرات

۲ آنها

۵ اگر ضریب تغییرات ۱۰ داده ۲ باشد و میانگین آن ۴، واریانس داده‌ها را به دست آورید.

۶ اگر داده را c برابر کنیم ضریب تغییرات داده‌ها چند برابر می‌شود؟

۷ فرض کنید ۲۲ بوته گل قرمز را انتخاب و تعداد گل‌های هر بوته را شمرده‌ایم و نتایج زیر به دست آمده است:

۷، ۴، ۳، ۸، ۶، ۴، ۱، ۷، ۴، ۲، ۱، ۱، ۱، ۱، ۳، ۲، ۲، ۲، ۲، ۵، ۵، ۱، ۲

نمودار جعبه‌ای را برای این داده‌ها رسم کنید.

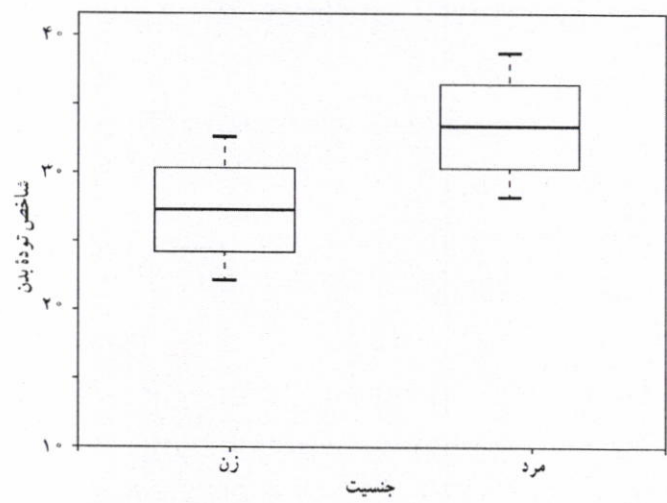
۸ نمودار جعبه‌ای مربوط به شاخص توده بدن (BMI) به تفکیک جنسیت رسم شده است. این نمودار را تفسیر کنید و

به سؤالات زیر پاسخ دهید.

الف) میانگین شاخص توده بدن در خانم‌ها بیشتر است یا آقایان؟

ب) میزان پراکندگی شاخص توده بدن در خانم‌ها بیشتر است یا آقایان؟

شاخص توده بدن



۹ داده‌های زیر مربوط به نرخ بیکاری یک کشور در ده سال گذشته است:

سال	اول	دوم	سوم	چهارم	پنجم	ششم	هفتم	هشتم	نهم	دهم
نرخ بیکاری	۱۱/۵	۱۱/۳	۱۰/۵	۱۰/۴	۱۱/۹	۱۳/۵	۱۲/۳	۱۲/۲	۱۰/۴	۳۰/۱

نمودار جعبه‌ای این داده‌ها را رسم کنید.

حل تمرین های صفحه ۹۹ (آمار و احتمال)

: ۱

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
۳۲	-۱۴	۱۹۶
۵۹	۱۳	۱۶۹
۲۶	-۲۰	۴۰۰
۵۳	۷	۴۹
۷۴	۲۸	۷۸۴
۱۷	-۲۹	۸۴۱
۴۵	-۱	۱
۲۳	-۲۳	۵۲۹
۶۴	۱۸	۳۲۴
۵۰	۴	۱۶
۶۱	۱۵	۲۲۵
جمع = ۵۰۴	---	۳۵۳۴

$$\bar{x} = \frac{\sum x_i}{n} = \frac{۵۰۴}{۱۱} = ۴۵/۱۱ \approx ۴۶$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{۳۵۳۴}{۱۱} = ۳۲۱/۲۷$$

$$\sigma = \sqrt{۳۲۱/۲۷} = ۱۸/۰۹ \quad CV = \frac{\sigma}{\bar{x}} = \frac{۱۸/۰۹}{۴۶} = ۰/۳۹$$

: ۲

سال	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
۸۵	۱/۴	-۱۰/۶	۱۱۲/۳۶
۸۶	۲/۸	-۹/۲	۸۴/۶۴
۸۷	۵	-۷	۴۹
۸۸	۹/۲	-۲/۸	۷/۸۴
۸۹	۱۵/۲	۳/۲	۱۰/۲۴
۹۰	۱۷/۸	۵/۸	۳۳/۶۴
۹۱	۱۹	۷	۴۹
۹۲	۱۸/۳	۶/۳	۳۹/۶۹
۹۳	۱۹/۵	۷/۵	۵۶/۲۵
جمع	۱۰۸/۲	---	۴۴۲/۶۶

تپه کننده:

گروه ریاضی دوره ی دوم متوسطه و انجمن مصلحان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

$$\bar{x} = \frac{\sum x_i}{n} = \frac{108/2}{9} = 12/0.2 \approx 12$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{442/664}{9} = 49/18$$

$$\sigma = \sqrt{49/18} = 7/0.1$$

$$CV = \frac{\sigma}{\bar{x}} = \frac{7/0.1}{12} = 0.58$$

: ۳

ضریب تغییرات	انحراف معیار	واریانس	میانگین	اعداد
۱/۴۲	۳۱/۸۶	۱۰۱۵/۳۹	۲۲/۴۲	۱۰۰ و ۱۲ و ۸ و ۱۶ و ۱۰ و ۴ و ۷
نامعین	۱/۸۷	۳/۵۰	۰	۳ و ۲ و ۱ و ۰ و ۰ و ۰ و -۳ و -۲ و -۱
۰/۰۶	۰/۵۹	۰/۳۵	۱۰/۰۴	۹/۸۸ و ۹/۴۲ و ۹/۷۶ و ۹/۶۲ و ۱۰/۱۱ و ۱۱/۳۶ و ۱۰/۱۱
۰/۶۱	۱۱۳۷/۹۸	۱۲۹۵۰۰۰/۷۵	۱۸۷۵/۵	۲ و ۳۰۰۰ و ۲۵۰۰ و ۲۰۰۰

: ۴

ضریب تغییرات	انحراف معیار	واریانس	میانگین	اعداد
۰	۰	۰	۷	۷ و ۷ و ۷ و ۷ و ۷
۰/۴۲	۱/۹۵	۳/۸۴	۴/۶	۳ و ۷ و ۳ و ۷ و ۳
۰/۶۰	۴/۲۴	۱۸	۷	۱ و ۴ و ۷ و ۱۰ و ۱۳
نامعین	۲/۲۸	۵/۲۰	۰	۳ و ۲ و ۰ و -۲ و -۳

: ۵ (ذکر تعداد داده ها ، لازم نیست.)

$$CV = \frac{\sigma}{\bar{x}} \rightarrow 2 = \frac{\sigma}{4} \rightarrow \sigma = 8 \rightarrow \sigma^2 = 64$$

: ۶

$$y_1 = cx_1 \text{ و } y_2 = cx_2 \text{ و } y_3 = cx_3 \text{ و } \dots \text{ و } y_n = cx_n$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum cx_i}{n} = \frac{\sum cx_i}{n} = \frac{c \sum x_i}{n} = c\bar{x}$$

۱۰۰, ۲

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum (cx_i - c\bar{x})^2}{n} = \frac{\sum c^2 (x_i - \bar{x})^2}{n} = \frac{c^2 \sum (x_i - \bar{x})^2}{n} = c^2 \sigma_x^2$$

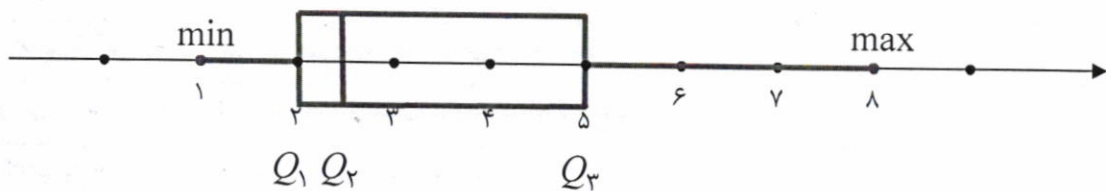
$$\rightarrow \sigma_y = |c| \sigma_x$$

$$CV_y = \frac{\sigma_y}{\bar{y}} = \frac{|c| \sigma_x}{c\bar{x}} = \pm \sigma_x$$

:۷

۱ و ۱ و ۱ و ۱ و ۱ و ۲ و ۲ و ۲ و ۲ و ۲ و ۲ و ۳ و ۳ و ۴ و ۴ و ۴ و ۵ و ۵ و ۶ و ۷ و ۷ و ۸

$$\min = 1, Q_1 = 2, Q_2 = \frac{2+3}{2} = 2.5, Q_3 = 5, \max = 8$$



:۸

الف) با توجه به نمودار، میانگین شاخص توده‌ی بدنی در آقایان بیشتر است.

ب) به نظر می‌رسد، پراکندگی یکسان است.

:۹

۱۰/۲ و ۱۰/۴ و ۱۰/۵ و ۱۱/۳ و ۱۱/۵ و ۱۱/۹ و ۱۲/۲ و ۱۲/۳ و ۱۳/۵ و ۳۰/۱

$$\min = 10/2, Q_1 = 10/5, Q_2 = \frac{11/5 + 11/9}{2} = 11/7, Q_3 = 13/3, \max = 30/1$$



نویسه کننده:

گروه ریاضی دوره ی دوم متوسطه و انجمن مصلحان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

۱۰۰,۳