

۱- انحراف معیار و واریانس داده‌ها

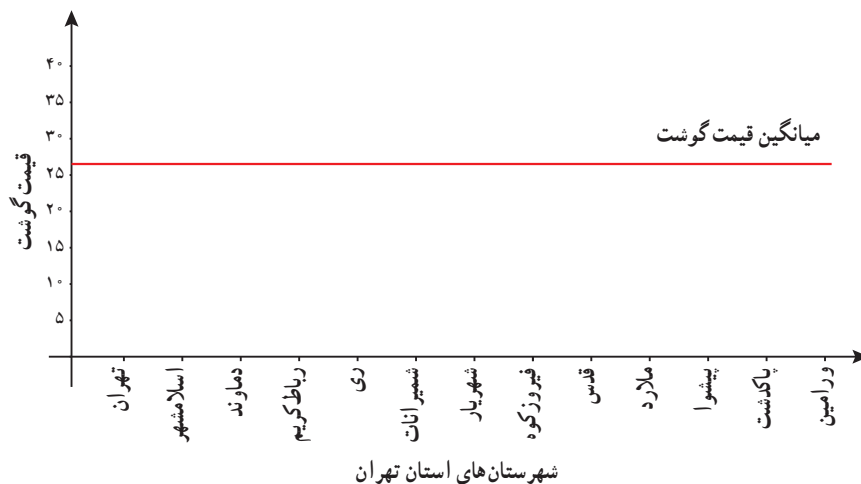
فعالیت



در اقتصاد هر کشوری شاخصی تحت عنوان نرخ تورّم، نقش بسیار مهمی را ایفا می‌کند. یکی از اقلام مصرفی مورد نیاز در محاسبه نرخ تورّم 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 آورید؟

■ در نمودار زیر، میانگین قیمت گوشت قرمز در شهرستان‌های استان تهران نشان داده شده است. قیمت گوشت قرمز در هریک 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

- 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 یا همان میانگین قیمت گوشت قرمز نزدیک‌تر باشند، نشان‌دهنده چیست؟ هر چقدر دورتر باشند چطور؟
- ۳ معیاری را برای اندازه‌گیری پراکندگی قیمت گوشت قرمز یا همان نقاط حول خط قرمز می‌توانید معرفی کنید؟

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.

$(x_i - \bar{x})^2$	$x_i - \bar{x}$	قیمت گوشت قرمز
		۴۲
		۲۰
		۲۵
		۲۶
		۲۷
		۴۰
		۲۰
		۱۶
		۲۰
		۲۱
		۲۲
		۲۳
		۲۶
	σ	
	σ^2	

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

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:

$$\text{نرخ تورّم} = \frac{P_1 - P_0}{P_0} \times 100\%$$

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 هرکیلو باشد، در این صورت نرخ تورّم برای قیمت گوشت قرمز در سال a برابر:

$$\text{نرخ تورّم} = \frac{(32 - 28)}{28} \times 100 = 14\%$$

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.

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

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

محصولات کدام کارخانه را انتخاب می کنید؟

ب) حال با تغییر واحد اندازه گیری در جدول قبلی میانگین و انحراف معیار طول عمر لاستیک ها در دو کارخانه الف) و ب) به صورت زیر گزارش داده شده است.

کارخانه	میانگین	انحراف معیار	ضریب تغییرات
کارخانه الف	۵۴۰۰۰۰۰۰ متر	۵۰۰۰۰ متر	
کارخانه ب	۶۵۰۰۰ کیلومتر	۱۰۰۰۰ کیلومتر	

همان طور که ملاحظه می کنید میانگین و انحراف معیار لاستیک ها برای کارخانه الف) برحسب واحد اندازه گیری 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 اندازه گیری وابسته است؟

نمودار جعبه ای

در ابتدای این درس با معیارهای پراکندگی آشنا شدیم، حال می خواهیم با استفاده از نمودارهای آماری، معیارهای پراکندگی داده ها را به صورت تصویری نشان دهیم.

فعالیت

میزان بارش برف سالانه در دو پیست اسکی «الف» و «ب» برای هفت سال اندازه گیری 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

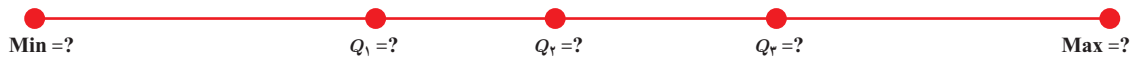
عدد ۰ در جدول به این معناست که میزان بارش کمتر از ۱ سانتی متر است.



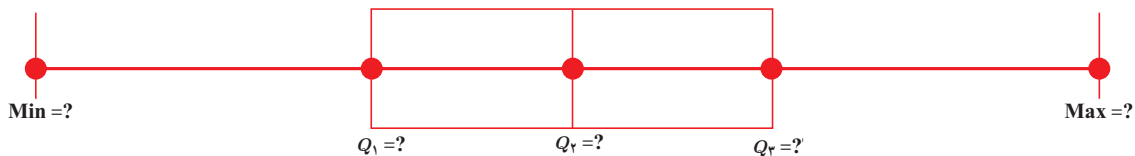
برای رسم نمودار آماری، مراحل زیر را انجام دهید.
الف) جدول زیر را کامل کنید.

کمترین مقدار میزان بارش برف Min	چارک اول میزان بارش برف Q_1	میانه میزان بارش برف Q_2	چارک سوم میزان بارش برف Q_3	بیشترین مقدار میزان بارش برف Max	سال
					پیست 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

- نمودار جعبه ای مربوط به پیست «ب» را رسم کنید. و سپس با نمودار جعبه ای پیست «الف» مقایسه کنید.
- اگر داده دور افتاده ای در داده ها باشد، نمودار جعبه ای چه تغییری می کند؟

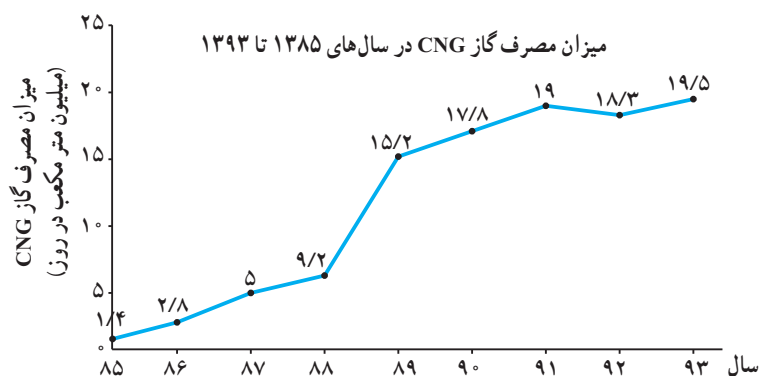
۱ فرض کنید سنّ افرادی که در یک روز سوار اتوبوس شده‌اند، به صورت زیر است :

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 هر یک 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، -۳، -۲، -۱			
۱۰/۱۱، ۱۱/۳۶، ۱۰/۱۱			
۹/۸۸، ۹/۴۲، ۹/۷۶، ۹/۶۲			
۲، ۳۰۰۰، ۲۵۰۰، ۲۰۰۰			

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

اعداد	انحراف معیار	واریانس	ضریب تغییرات

۵ اگر ضریب تغییرات 10 داده 2 باشد و میانگین آن 4 ، واریانس داده‌ها را به دست آورید.

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

۷ فرض کنید 22 بوته گل قرمز را انتخاب و تعداد گل‌های هر بوته را شمرده‌ایم و نتایج زیر به دست آمده است:

$7, 4, 3, 8, 6, 4, 1, 7, 4, 2, 1, 1, 1, 1, 3, 2, 2, 2, 2, 5, 5, 1, 2$

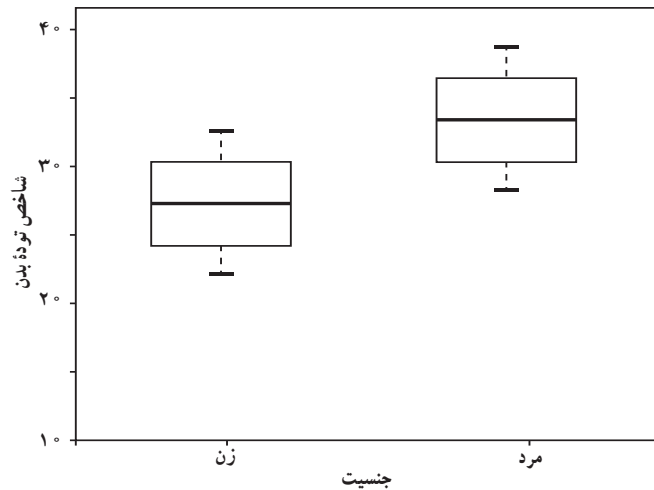
نمودار جعبه‌ای را برای این داده‌ها رسم کنید.

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؟



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.

نرخ بیکاری



امروزه بیکاری یکی از موضوعات مهم در جوامع بشری است که دولتمردان و سیاست‌گذاران تمامی کشورهای جهان به دنبال راهکارهایی برای از بین بردن این مسئله در کشورشان و فراهم کردن زمینه‌ای برای به‌کارگیری استعدادهای مردم کشورشان هستند. در علوم اقتصادی با استفاده از علم آمار، شاخصی تحت عنوان نرخ بیکاری بیان می‌شود. نرخ بیکاری، به صورت زیر تعریف می‌شود.

$$\text{نرخ بیکاری} = \frac{\text{جمعیت بیکار}}{\text{کل جمعیت فعال}}$$

جمعیت بیکار به افراد ۱۰ ساله و یا بالاتر از ۱۰ ساله‌ای گفته می‌شود که سه شرط زیر را توأمأ داشته باشند:

- در هفته مشخص حتی یک ساعت هم کار نکرده باشد.
- آمادگی برای انجام کار داشته باشد.

■ در هفته مشخص و سه هفته قبل از آن جویای کار باشد. (اقدامات مشخصی را به منظور جست‌وجوی اشتغال، مزدبگیری و یا خوداشتغالی به عمل آورده باشد.)

جمعیت شاغل: به افراد ۱۰ ساله، یا بالاتر از ۱۰ ساله‌ای که در طول هفته مشخص (بازه زمانی ۷ روزه‌ای که وضع فعالیت افراد در این بازه زمانی مدنظر باشد) حداقل یک ساعت کار کرده باشد شاغل گویند.

جمعیت فعال: به مجموع جمعیت بیکار و شاغل گفته می‌شود.

