

معیارهای پراکندگی

درس ۳

۱- انحراف معیار و واریانس داده‌ها

فعالیت

ص ۹۳

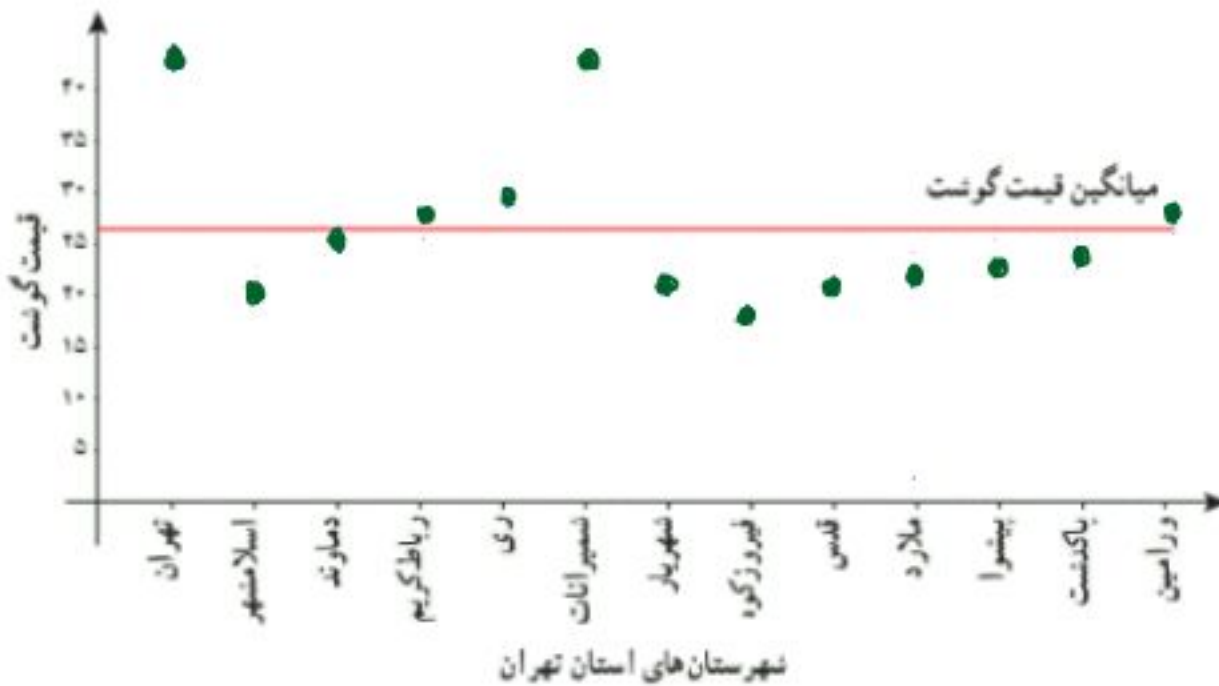


در اقتصاد هر کشوری شاخصی تحت عنوان نرخ توزّم، نقش بسیار مهمی را ایفا می‌کند. یکی از افلام مصرفی مورد نیاز در محاسبه نرخ توزّم در یک کشور، قیمت گوشت قرمز است. در جدول روبه‌رو قیمت گوشت قرمز در سال ۱۳۹۵ در شهرستان‌های استان تهران گردآوری شده است.

■ 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 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
ریاض کریم	۲۶
ری	۲۷
شمیرانات	۴۰
شهریار	۲۰
فیروزکوه	۱۶
قدس	۲۰
ملارد	۲۱
پیشوا	۲۲
پاکدشت	۲۳
ورامین	۲۶



$$\bar{x} = \frac{42 + 20 + 25 + 26 + 27 + 40 + 20 + 16 + 20 + 21 + 22 + 23 + 26}{13} = \frac{328}{13}$$

$$\bar{x} = 25,23$$

ص ۱

۱ نقطه ۲ نقطه ۳ 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 هر یک از شهرستان‌های استان تهران چقدر از میانگین قیمت دورتر است. هر چقدر نقاط با همان قیمت گوشت قرمز در هر یک از شهرستان‌های استان تهران حول خط قرمز با همان میانگین قیمت گوشت قرمز نزدیک‌تر باشند، نشان‌دهنده چیست؟ هر چقدر دورتر باشند چگونه؟

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

اختلاف زیاد قیمت گوشت
قرمز در آن شهرستان با
مقدار میانگین قیمت‌ها

اختلاف کم قیمت گوشت قرمز در آن شهرستان
با مقدار میانگین قیمت‌ها

انحراف از میانگین - انحراف معیار

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

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

توجه: $\bar{x} = ۲۵,۲۳$
 برای راحتی محاسبات بعدی
 میانگین را ۲۵ در نظر می‌گیریم

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$\sigma^2 = \frac{۲۸۹ + ۲۵ + ۰ + ۱ + ۴ + ۲۲۵ + ۲۵ + ۸۱ + ۲۵ + ۱۶ + ۹ + ۴ + ۱}{۱۳}$$

$$\sigma^2 = \frac{۷۰۵}{۱۳} = ۵۴,۲۳$$

$$\sigma = \sqrt{۵۴,۲۳} = ۷,۳۶$$

۲- ضریب تغییرات داده‌ها

فعالیت

ص ۹۶



یکی از شاخص‌های کیفیت در لاستیک‌های تولید شده اتومبیل توسط یک کارخانه، طول عمر آن لاستیک‌هاست. هرچقدر متوسط طول عمر لاستیک‌های تولیدی بیشتر و انحراف معیار طول عمر لاستیک‌ها کمتر باشد، به این معناست که لاستیک‌ها کیفیت بالایی از نظر طول عمر دارند.

حال با توجه به مطالب گفته شده، به بررسی کیفیت لاستیک‌های تولیدی از نظر طول عمر دو کارخانه (الف) و (ب) می‌پردازیم. براساس داده‌های به دست آمده میانگین طول عمر لاستیک‌ها در دو کارخانه و انحراف معیار آنها به شرح جدول روبه‌رو است:

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

- شما ترجیح می‌دهید از کدام کارخانه لاستیک بخرید؟
- آیا می‌توان براساس میانگین و انحراف معیار و نمونه‌های در نظر گرفته شده قضاوت کرد؟ تا حدودی

الف) دلیل کمتر بودن انحراف معیار
ب) بالاتر بودن میانگین طول عمر

کار در کلاس

ص ۹۷

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 می‌کنید؟

$$CV = \frac{6}{x}$$

$$CV \text{ کارخانه الف} = \frac{50}{54000} = 0.0009$$

$$CV \text{ کارخانه ب} = \frac{100}{65000} = 0.0015$$

ص ۹۷

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

کارخانه	میانگین	انحراف معیار	ضریب تغییرات
کارخانه الف	۵۴۰۰۰۰۰ متر	۵۰۰۰۰ متر	
کارخانه ب	۶۵۰۰۰ کیلومتر	۱۰۰ کیلومتر	

همان‌طور که ملاحظه می‌کنید میانگین و انحراف معیار لاستیک‌ها 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؟

$$CV \text{ کارخانه الف} = \frac{50000}{5400000} = 0.009$$

خیر

$$CV \text{ کارخانه ب} = \frac{100}{65000} = 0.0015$$

فعالیت

ص ۹۷

میزان بارش برف سالانه در دو پیست اسکی «الف» و «ب» برای هفت سال اندازه گیری و نتایج، در جدول زیر گردآوری شده است :



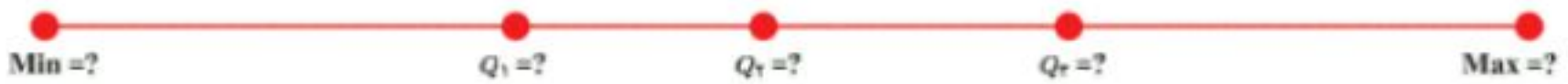
سال	۱۳۸۸	۱۳۸۹	۱۳۹۰	۱۳۹۱	۱۳۹۲	۱۳۹۳	۱۳۹۴
میزان بارش برف در پیست اسکی الف	۵۵۱	۱۹۰	۳۳۵	۷۸۷	۴۷۲	۷۲۸	۸۲۵
میزان بارش برف در پیست اسکی ب	۲۷۱	۰	۵۲۵	۱۰۱۶	۹۳	۵۸۱	۵۶۶

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

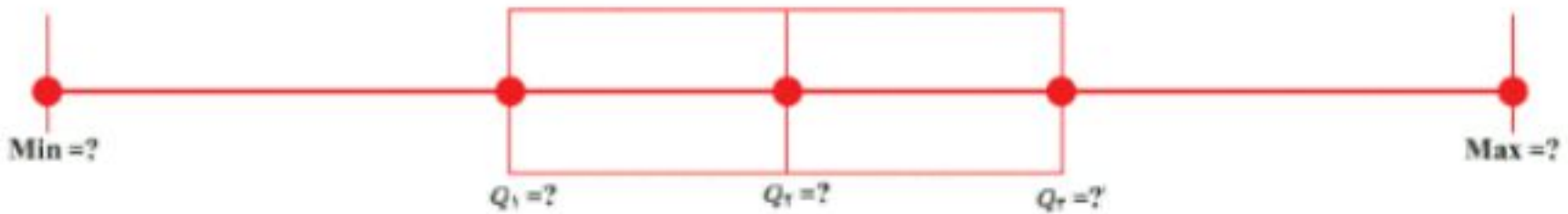
برای رسم نمودار آماری، مراحل زیر را انجام دهید.
الف) جدول زیر را کامل کنید.

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

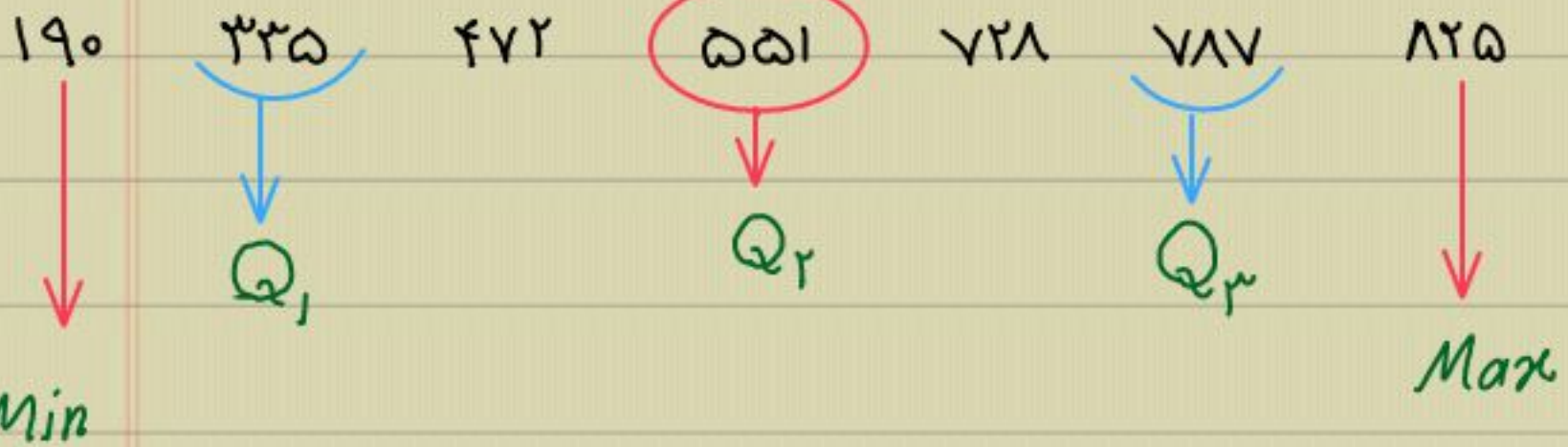
ب) حال مقادیر جدول را روی یک محور نمایش می دهیم.



ب) 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.

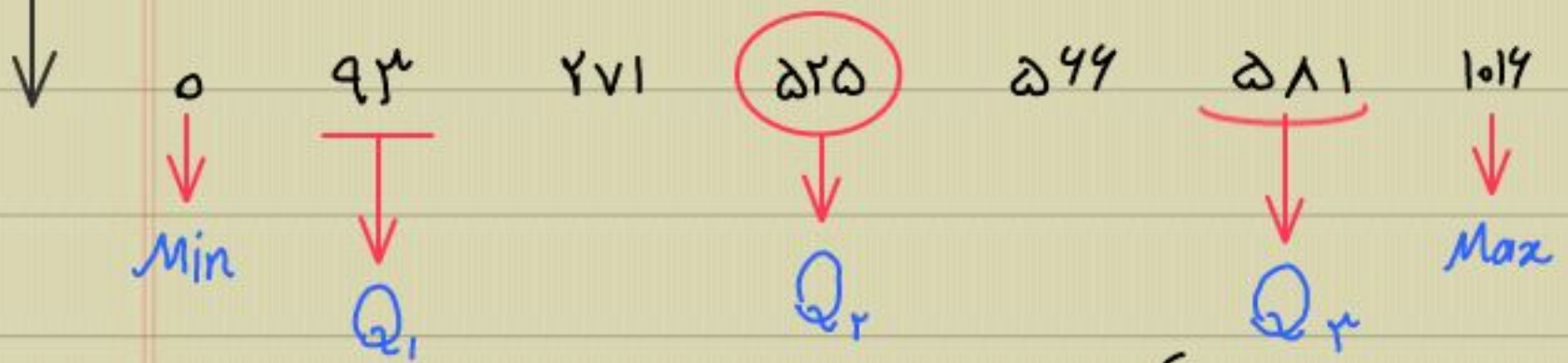


دامنه میان چارکی $IQR = Q_3 - Q_1 = 787 - 335 = 452$

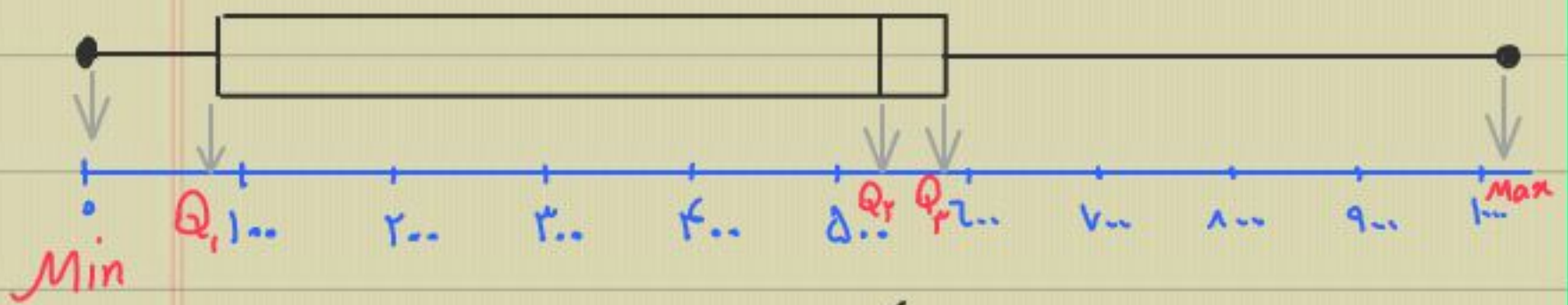
■ نمودار جعبه‌ای مربوط به بیست «ب» را رسم کنید. و سپس با نمودار جعبه‌ای بیست «الف» مقایسه کنید.

■ اگر داده دور افتاده‌ای در داده‌ها باشد، نمودار جعبه‌ای چه تغییری می‌کند؟

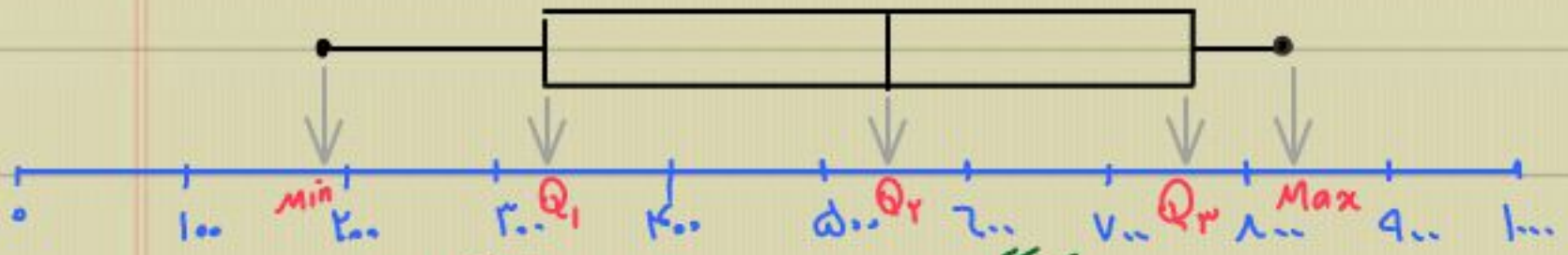
جواب در پایین صفحه



نمودار جعبه‌ای بیست (ب)



نمودار جعبه‌ای بیست (الف)



پراکندگی داده‌ها در بیست (ب) بیشتر است

وجود داده‌ی دور افتاده باعث می‌شود که نمودار جعبه‌ای کشیده‌تر شود

و سبیل نمودار بزرگ‌تر باشد که به معنی پراکندگی بیشتر است

جواب سؤال دوم:

۱ فرض کنید سن 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 سن افراد را به دست آورید.

$$\bar{x} = \frac{32 + 59 + 26 + 53 + 74 + 17 + 45 + 23 + 64 + 50 + 21}{11} = \frac{504}{11} = 45,81$$

میانگین را بصورت تقریبی ۴۶ در نظر می‌گیریم

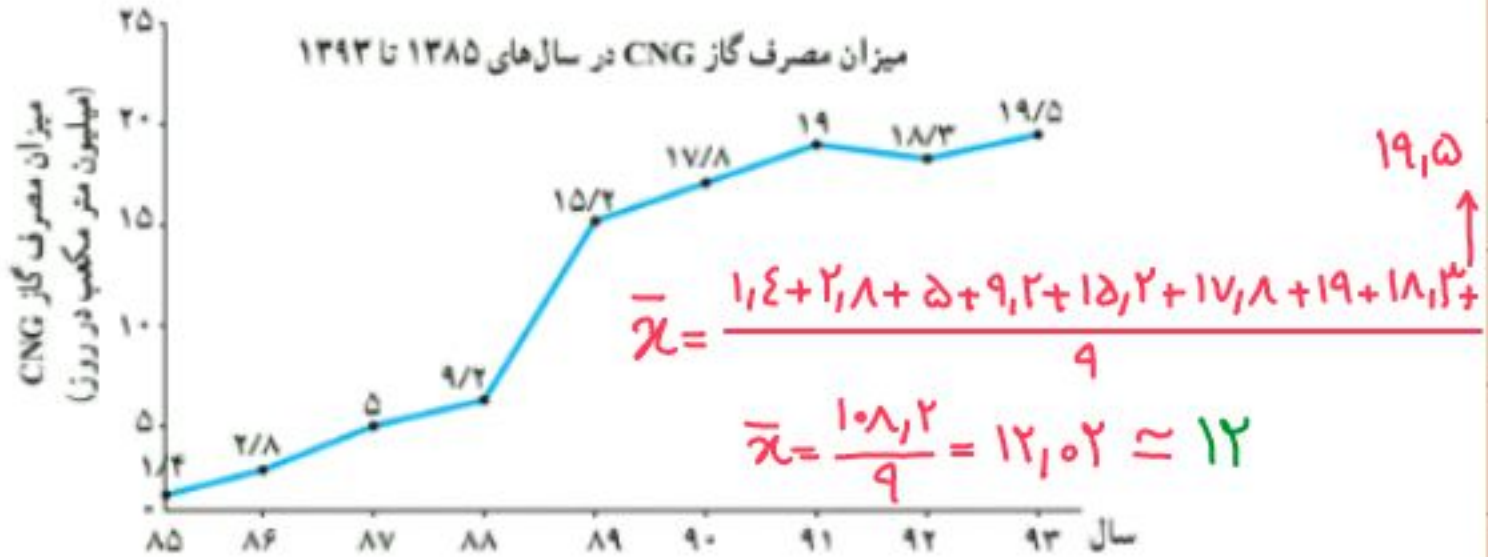
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
۳۲	$32 - 46 = -14$	$(-14)^2 = 196$
۵۹	$59 - 46 = 13$	$(13)^2 = 169$
۲۶	$26 - 46 = -20$	$(-20)^2 = 400$
۵۳	۷	۴۹
۷۴	۲۸	۷۸۴
۱۷	-۲۹	۸۴۱
۴۵	-۱	۱
۲۳	-۲۳	۵۲۹
۶۴	۱۸	۳۲۴
۵۰	۴	۱۶
۲۱	۱۵	۲۲۵
		۳۵۳۴

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{3534}{11} = 321,27 \quad \text{واریانس}$$

$$s = \sqrt{321,27} = 18,109 \rightarrow CV = \frac{s}{\bar{x}} = \frac{18,109}{46} = 0,39 \quad \text{ضریب تغییرات}$$



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.



سال	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
۸۵	۱,۴	$1.4 - 12 = -10.6$	$(-10.6)^2 = 112.36$
۸۶	۲,۸	$2.8 - 12 = -9.2$	$(-9.2)^2 = 84.64$
۸۷	۵	$5 - 12 = -7$	$(-7)^2 = 49$
۸۸	۹,۲	-2.8	7.84
۸۹	۱۵,۲	3.2	10.24
۹۰	۱۷,۸	5.8	33.24
۹۱	۱۹	7	49
۹۲	۱۸,۳	2.3	39.49
۹۳	۱۹,۵	7.5	221.25
			۴۴۲,۴۴

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{442.44}{9} = 49.18 \quad \text{واریانس}$$

$$s = \sqrt{49.18} = 7.01 \quad \text{انحراف معیار}$$

$$CV = \frac{s}{\bar{x}} = \frac{7.01}{12} = 0.58 \quad \text{ضریب تغییرات}$$

۲ انحراف معیار، واریانس و ضریب تغییرات را برای هر یک از اعداد جدول زیر به دست آورید.

ضریب تغییرات	واریانس	انحراف معیار	اعداد	\bar{x}
			۱۰۰، ۱۲، ۸، ۱۶، ۱۰، ۴، ۷	$\frac{۱۵۷}{۷} = ۲۲,۴$
			۳، ۲، ۱، ۰، ۰، ۰، -۳، -۲، -۱	$\frac{۰}{۸} = ۰$
			۱۰/۱۱، ۱۱/۳۶، ۱۰/۱۱ ۹/۸۸، ۹/۴۲، ۹/۷۶، ۹/۶۲	$\frac{۷۰,۲۶}{۷} = ۱۰,۰۳$
			۲، ۳۰۰۰، ۲۵۰۰، ۲۰۰۰	$\frac{۷۵۰۲}{۴} = ۱۸۷۵,۵$

$$s^2 = \frac{(100 - 22,4)^2 + (12 - 22,4)^2 + \dots + (4 - 22,4)^2 + (7 - 22,4)^2}{7}$$

$$s^2 = \frac{7107,75}{7} = 1015,39$$

$$s = \sqrt{1015,39} = 31,87$$

$$CV = \frac{s}{\bar{x}} = \frac{31,87}{22,4} = 1,4$$

برای اعداد هر ردیف همین مراحل را طی می‌کنیم

۲ 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

به عمده رانش آموز

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

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

\downarrow \bar{x} \downarrow CV_p

انحراف معیار $\sigma = \bar{x} \cdot CV_p \Rightarrow \sigma = 4 \times 2 = 8$

واریانس $\sigma^2 = 8^2 = 64$

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

ثابت باقی می‌ماند $x_1, x_2, \dots, x_n \xrightarrow{\times c} cx_1, cx_2, \dots, cx_n$

زیرا: \bar{x} میانگین $\rightarrow c\bar{x}$ میانگین جدید

$k_i = (x_i - \bar{x}) \rightarrow (cx_i - c\bar{x}) = c(x_i - \bar{x}) = ck_i$

ضریب اولیه واریانس $\sigma_1^2 = \frac{\sum k_i^2}{n}$ $\sigma_2^2 = \frac{c \sum k_i^2}{n} = c^2 \sigma_1^2$

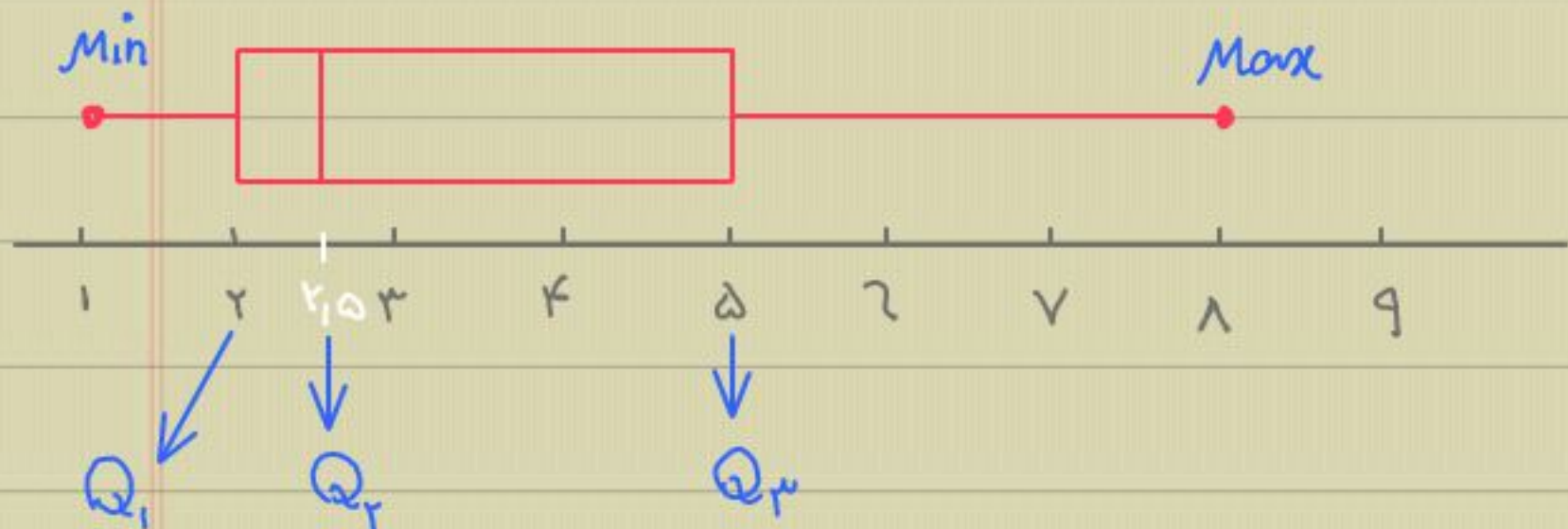
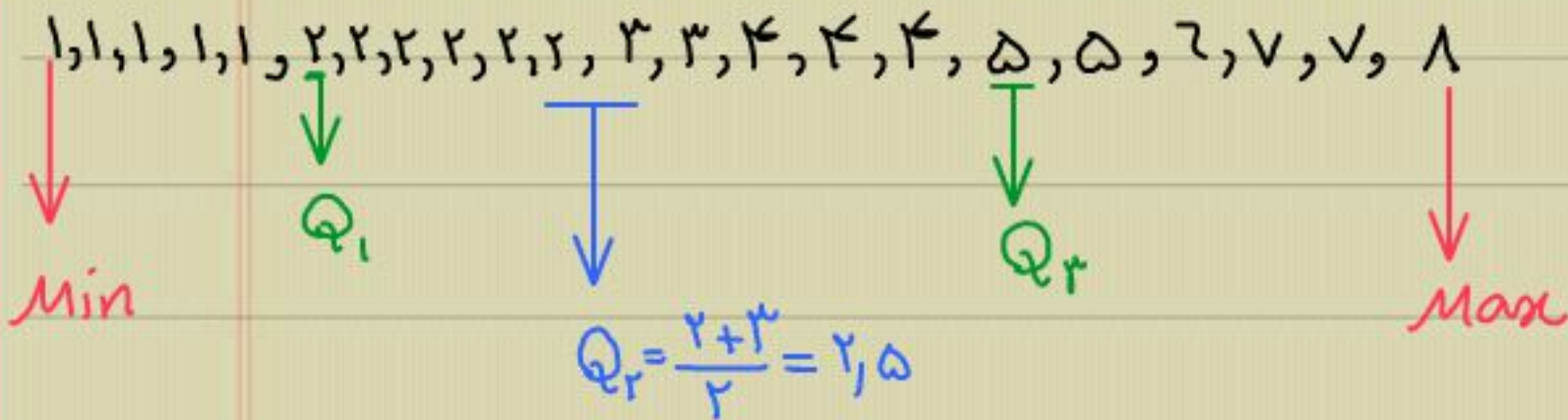
واریانس جدید $\sigma_2 = c \sigma_1$ خیز

ضریب جدید $CV_p = \frac{\sigma_2}{c\bar{x}} = \frac{c\sigma_1}{c\bar{x}} = \frac{\sigma_1}{\bar{x}} = CV_p$

۷ فرض کنید ۲۲ بوته گل قرمز را انتخاب و تعداد گل های هر بوته را شمرده ایم و نتایج زیر به دست آمده است:

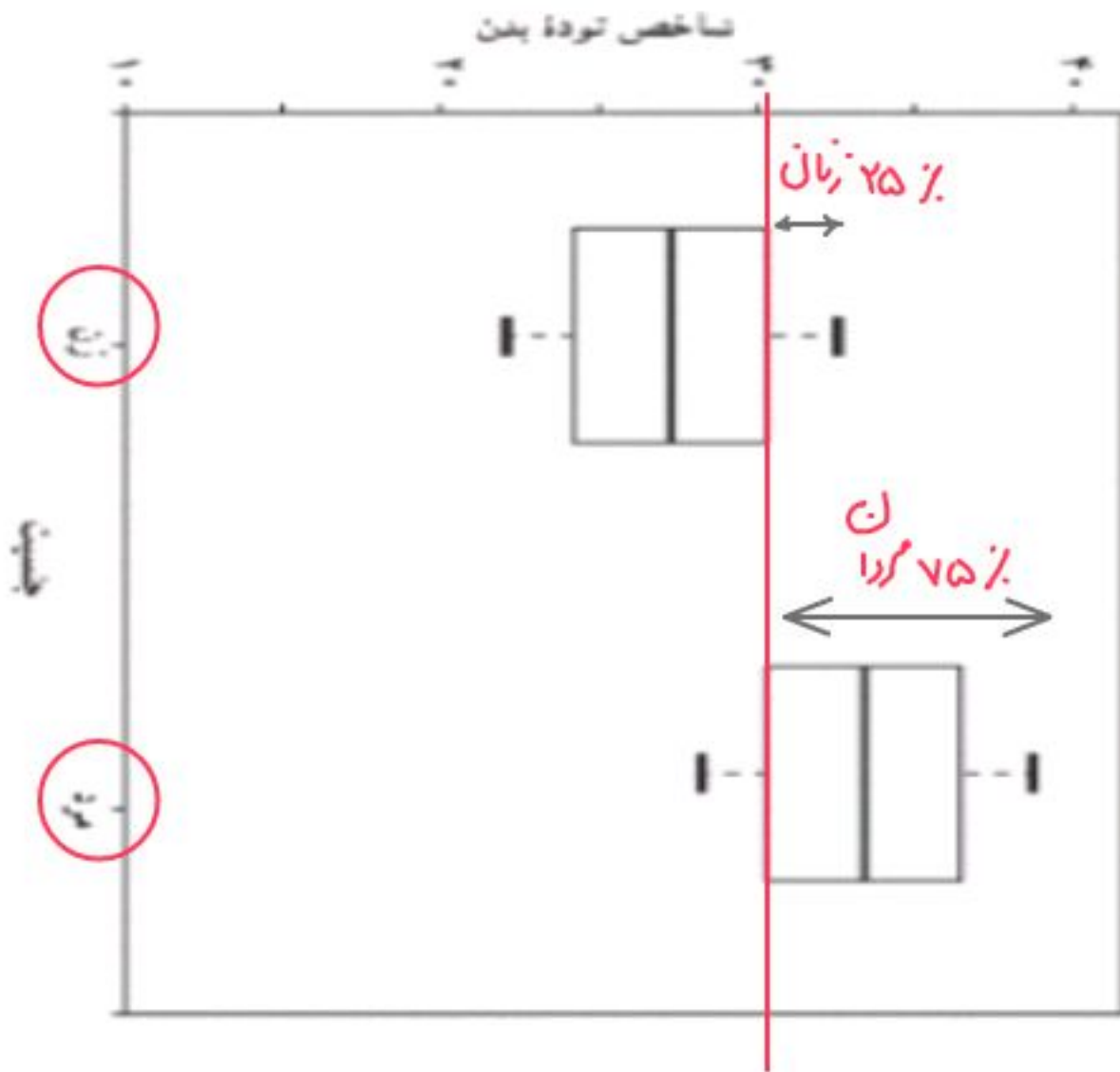
۷، ۴، ۳، ۸، ۶، ۴، ۱، ۷، ۴، ۲، ۱، ۱، ۱، ۳، ۲، ۲، ۲، ۲، ۵، ۵، ۱، ۲

نمودار جعبه ای را برای این داده ها 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؟ **آقایان**
 ب) میزان پراکندگی شاخص توده بدن در خانم‌ها بیشتر است یا آقایان؟ **تقریباً برابر هستند**



الف) ۷۵٪ از مردان شاخص ستر از ۲۵ دارند در صورتی که فقط ۲۵٪ از خانم‌ها شاخص ستر از ۲۵ دارند

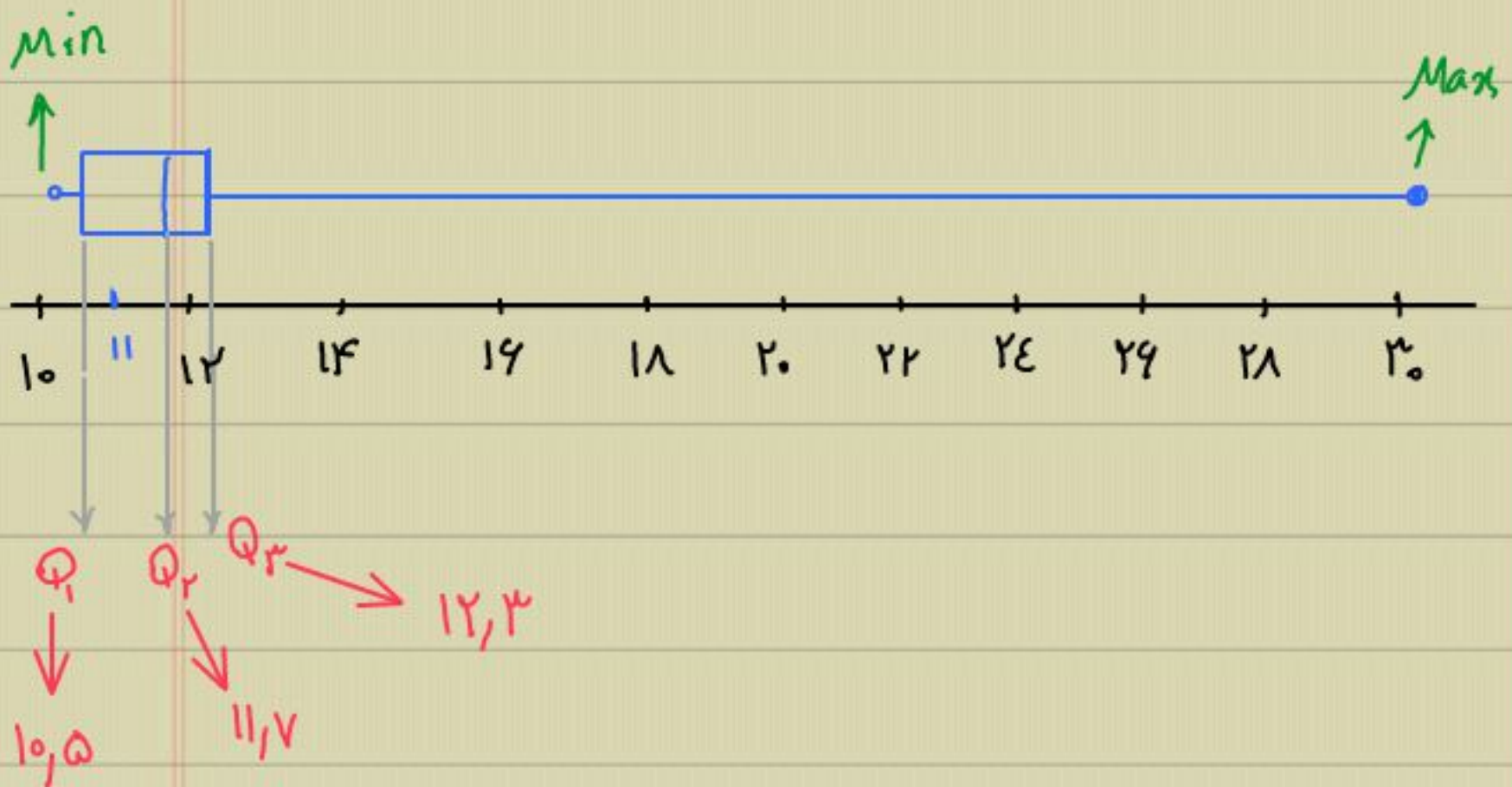
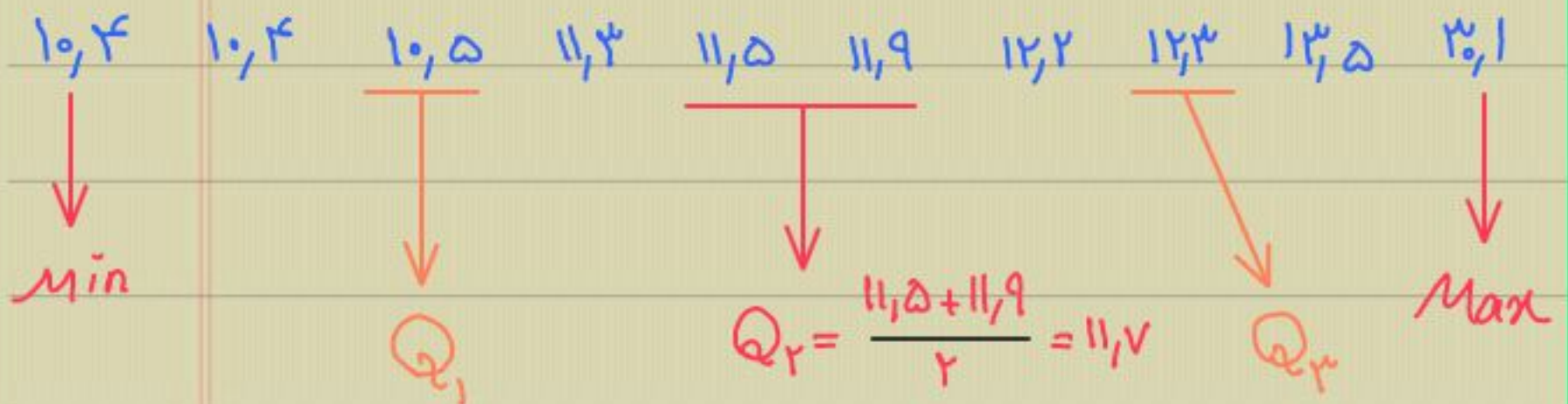
ب) ۱۰٪ پراکندگی شاخص در مردان

۱۰٪ در زنان

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	۱۲/۵	۱۲/۳	۱۲/۲	۱۰/۴	۳۰/۱

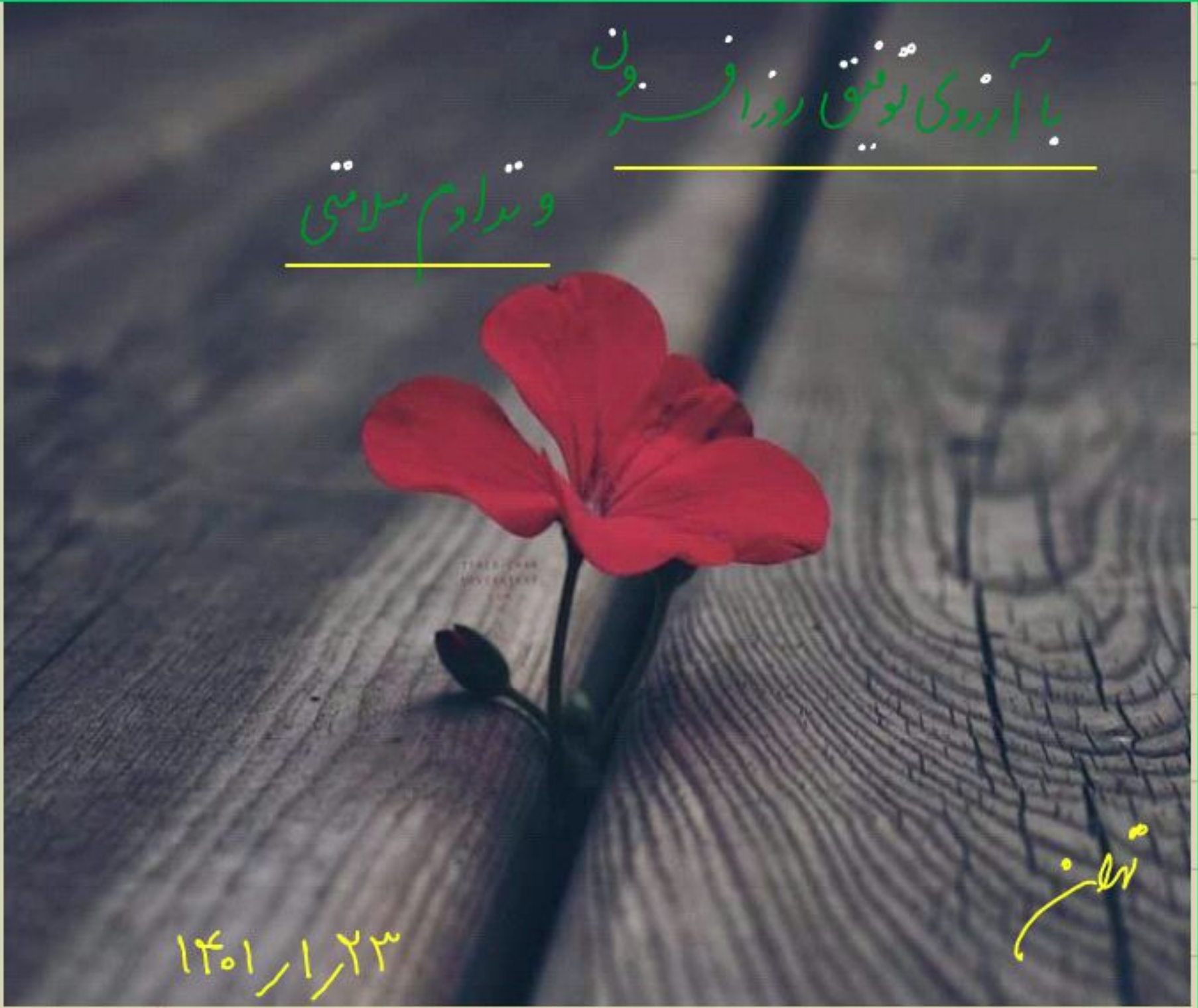
نمودار جعبه‌ای این داده‌ها را رسم کنید.



فصل: سوم درس: سوم : @Mahdi_math_for_all اینستاگرام

با آرزوی توفیق روزی

و تداوم سلامتی



محمد مهدی

ص آفر

09213102271 - 09125102271 - @moharrammahdi - @Riazi101112forall2