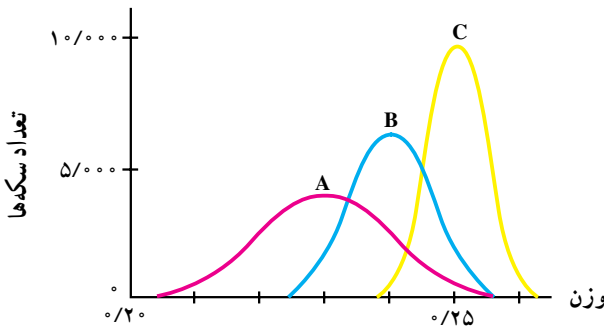


از فرشتگانی که در آسمانهایت جایشان دادی، و از زمین خود برترشان بردی. آنان از دیگر آفریدگانت تو را بهتر شناسند، و از عقاب تو بیشتر می‌هراسند، و به تو نزدیکترند و نه پراکنده‌گردش زمان. آنان با مرتبتی که از آن برخوردارند، و از منزلتی که نزد تو دارند، و یکدله تو را دوست دارند و تو را فراوان طاعت می‌گذارند، و اندک غفلتی در فرمان تو نیارند.



شاخص‌های پراکندگی

ضرابخانه سکه‌های دولتی تصمیم گرفته است که وزن پنج تومانی‌هایی که در جریان بازار قرار می‌دهد، روی نمودار بیاورد. برای این منظور، چندین سکه پنج تومانی که تازه ضرب شده‌اند به طور تصادفی انتخاب و به دقت وزن می‌شوند. همچنین به همان تعداد سکه پنج تومانی که پنج سال پیش به بازار عرضه شده بود و به همان تعداد سکه پنج تومانی که ۱۰ سال پیش به بازار عرضه شده بود به دقت وزن می‌شوند. سه نمودار A و B و C به دست می‌آید.



- فکر می‌کنید کدام نمودار مربوط به سکه‌های نو ارائه شده به بازار و کدام نمودار مربوط به سکه‌های پنج ساله و ده ساله است؟
- با گذشت زمان میانگین وزن سکه‌های پنج تومانی چه تغییری می‌کند؟
- با گذشت زمان مد وزن سکه‌های پنج تومانی چه تغییری می‌کند؟
- چاق و یا لاغر شدن نمودارها را چگونه تعبیر می‌کنید؟



پراکندگی

در فصل قبل با شاخص‌های مرکزی آشنا شدیم. مثلاً شما بر اساس میانگین زمان‌هایی که طول کشیده است تا در ۲۰ روز گذشته از خانه به مدرسه بروید، از منزل خارج شده‌اید. در بین راه اعدادی را که معرف این مدت زمان بوده‌اند در ذهن خود مرور می‌کنید. میانگین اعداد ۳۵ دقیقه شده است و شما درست ۳۵ دقیقه قبل از شروع کلاس‌ها از خانه خارج شده‌اید، ولی اعدادی مانند ۴۰، ۳۸، ۳۰، ۳۲ و یا حتی ۴۵ در بین اعداد دیده می‌شود. شما مایل نیستید که دیر به مدرسه برسید، به ویژه آن‌که روز امتحان هم باشد. یادآوری اعداد بالا آسودگی خیال را به هم می‌ریزد. از خود می‌پرسید نکند امروز یکی از آن روزهایی باشد که بیش از ۳۵ دقیقه طول بکشد. اگر کمی بیش‌تر طول بکشد، (در حد ۵ دقیقه) اشکال عمده‌ای نخواهد بود ولی تأخیر یک ربع و یا بیش‌تر مشکل‌ساز است. هرچه قدر در بین داده‌های شما، اعداد دورتر از میانگین وجود داشته باشد، ناراحتی شما بیش‌تر خواهد بود. اگر قبلاً این اعداد را بهتر و ارسبی می‌کردید و صرفاً میانگین را مدنظر قرار نمی‌دادید الان وضع بهتری داشتید، اگر از ابتدا درمی‌یافتید که اعداد تفاوت دارند می‌توانستید در این روز سرنوشت‌ساز تصمیم بهتری بگیرید. معیاری که می‌تواند این تفاوت‌ها و میزان آن‌ها و به‌خصوص دوری آن‌ها از میانگین را برای ما اندازه‌گیری کند شاخص پراکندگی نامیده می‌شود. ما به چند شاخص پراکندگی در زیر اشاره می‌کنیم:

دامنه تغییرات

دامنه تغییرات که در فصل دوم با آن آشنا شدید یکی از شاخص‌های پراکندگی است. هرچه قدر داده‌ها از هم دورتر باشند، دامنه تغییرات آن‌ها بیش‌تر خواهد بود. دیدیم اگر دامنه تغییرات برابر صفر باشد تمام داده‌ها با هم برابرند. یعنی هیچ تفاوتی بین آن‌ها موجود نیست. حال به دو دسته داده زیر که نمره‌های دانش‌آموزان در درس فیزیک در دو نمونه ۱۰ عضوی را به نمایش می‌گذارند توجه کنید.

الف) ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹

ب) ۱۰، ۱۵، ۱۴، ۱۴، ۱۵، ۱۵، ۱۵، ۱۵، ۱۴، ۱۹

دامنه تغییرات هر دو دسته یک اندازه است:

$$R = 19 - 10 = 9$$

ولی پراکندگی داده‌های دسته دوم کم‌تر است. در داده‌های دسته دوم ما عملاً اعداد ۱۴ و ۱۵ را داریم. به ۱۰ و ۱۹ بیش‌تر باید با شک و تردید نگریست. این دو مقدار ممکن است به خاطر اندازه‌گیری‌های نادرست حاصل شده باشند و یا ممکن است علل دیگری برای وجود آن‌ها در بین داده‌ها باشد. اما در داده‌های دسته اول این دو مقدار کاملاً طبیعی به نظر می‌رسند، زیرا مقادیر نزدیک به آن‌ها نیز داریم. با این توضیحات ما در داده‌های دسته دوم بیش‌تر روی ۱۴ و ۱۵ حساب می‌کنیم و این دو مقدار هم به هم خیلی نزدیک‌اند. این توضیح می‌رساند که دامنه تغییرات در بعضی مواقع به خوبی نمی‌تواند پراکندگی موجود در داده‌ها را نشان دهد. بنابراین به دنبال معیاری هستیم که بهتر از دامنه تغییرات بتواند پراکندگی و تفاوت‌ها را نشان دهد. قبل از آن که به معرفی معیار بعدی برویم ذکر یک نکته ضروری است که وقتی با دید مجرد ریاضی به دو دسته عدد بالا نگاه می‌کنید تفاوت هر دو را در حد ۹ می‌بینید، ولی با دید آماری مسئله کمی فرق می‌کند. در آمار حضور دو مقدار ۱۰ و ۱۹ را در دسته «ب» زیاد جدی نمی‌گیریم چون این‌ها مقادیر تک افتاده‌ای هستند. به نظر می‌رسد این دو مقدار به جامعه تعلق نداشته باشند. ما در این درس سعی می‌کنیم علاوه بر نقطه نظرانی که شما درباره مسائل اطراف خود دارید پایه‌های دید آماری را نیز ایجاد کنیم و آن را توسعه بخشیم.



چارک‌های بالا و پایین

گفتیم که دامنه تغییرات ممکن است در بعضی مواقع تعبیرهای نامناسب از جامعه ارائه کند، مثلاً در جامعه‌ای که عملاً داده‌ها به هم نزدیک‌اند، به علت وجود دو مقدار خیلی کوچک و خیلی بزرگ در جامعه، دامنه تغییرات عدد بزرگی به دست آید، و حال آن که جامعه از دید آماری جامعه متمرکزی باشد. تعداد بازدیدکنندگان از کتابخانه مدرسه در ۱۲ روز کاری در زیر آمده است.

۰, ۱, ۲, ۸, ۷, ۶, ۵, ۹, ۱۰, ۶, ۱۵, ۱۱

دامنه تغییرات این داده‌ها برابر ۱۵ است و به نظر زیاد می‌رسد ولی داده‌ها عموماً از ۵ تا ۱۱ پراکنده شده‌اند که پراکندگی آن‌ها زیاد نیست. از این رو برای حذف تأثیر داده‌های بزرگ و کوچک روی تعدادی از اعداد بالای داده‌ها و تعدادی از اعداد پایین داده‌ها را حذف می‌کنند. این که چه تعدادی حذف شود بستگی به تعداد داده‌ها و نزدیکی داده‌ها به هم دارد. در بعضی مواقع یک دهم داده‌ها از بالا و یک دهم داده‌ها از پایین را حذف می‌کنند و در برخی مواقع یک چهارم داده‌های بالایی و یک چهارم داده‌های پایینی را حذف می‌کنند و بعد از آن دامنه تغییرات داده‌های باقی‌مانده را محاسبه می‌کنند. به عنوان مثال در داده‌های بالا اگر یک چهارم داده‌ها را از بالا و پایین حذف کنیم به داده‌های زیر خواهیم رسید :

۰, ۱, ۲, ۵, ۶, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۵

بنابراین دامنه تغییرات این داده‌ها عبارت خواهد بود از $15 - 0 = 15$ است. اگر توجه کرده باشید در برخی از مسابقات امتیازهایی را که داورها می‌دهند، بیش‌ترین و کم‌ترین امتیاز را حذف می‌کنند و بر اساس امتیازهای باقی‌مانده امتیاز ورزشکار را تعیین می‌کنند.

تمرین

۱- دو کارخانه تولید کننده مواد غذایی A و B شکلات در بسته بندی ۴۸ گرمی می فروشند. پنج بسته شکلات به صورت تصادفی از یک فروشگاه مواد غذایی از دو محصول انتخاب شد و تمام وزن بسته ها به دقت اندازه گیری شد. نتیجه زیر به دست آمد:

A: ۴۸/۰۸, ۴۸/۳۲, ۴۷/۹۶, ۴۷/۸۴, ۴۷/۹۶

B: ۴۹/۱۶, ۴۸/۸۴, ۴۸/۸۸, ۴۹/۰۸, ۴۹

الف - کدام کارخانه شکلات بیش تری در بسته ها می فروشد؟ از چه شاخصی استفاده می کنید؟
ب - کدام کارخانه در توزیع شکلات یکنواخت تر عمل کرده است؟

۲- اگر دامنه تغییرات برابر صفر باشد در باره داده ها چه نتیجه ای می گیرید؟

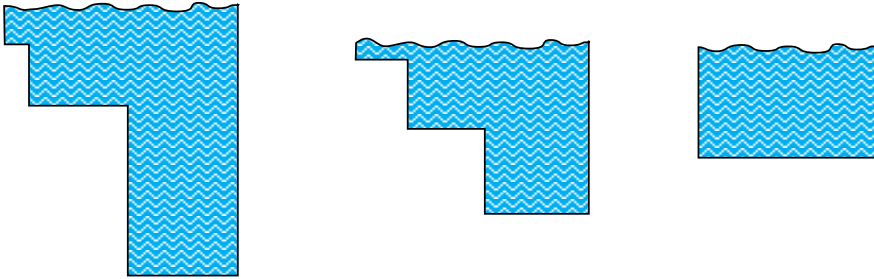
۳- تغییر واحد اندازه گیری چه تأثیری در دامنه تغییرات می گذارد؟ مثلاً اگر شما داده ها را برحسب متر اندازه گیری کرده باشید و سپس آن ها را برحسب سانتی متر به کار برید دامنه تغییرات چگونه تغییر می کند؟

۴- اگر به داده ها یک مقدار ثابت اضافه کنیم دامنه تغییرات چگونه تغییر می کند؟



واریانس

اگر شما خوب شنا نمی‌دانستید و می‌خواستید در استخری شنا کنید که عمق آن متغیر است برای این که اطمینان حاصل کنید که در خطر نخواهید بود چه اطلاعاتی را لازم دارید؟ اگر میانگین عمق استخر را به شما داده باشند و این میانگین برابر ۱/۵ متر باشد، آیا برای شنا در این استخر اطمینان دارید؟ به حالات زیر توجه کنید.



در هر سه استخر بالا میانگین عمق برابر ۱/۵ متر است. می‌خواهیم با یک شاخص عددی پراکندگی متغیر عمق را نسبت به عمق میانگین بررسی کنیم. اگر بدانیم که تغییرات عمق در اطراف عمق میانگین، یعنی ۱/۵ متر زیاد نیست با اطمینان بیش‌تری در آن شنا خواهیم کرد. به شرط آن که این پراکندگی با یک شاخص عددی مناسب معرفی شده باشد.

دامنه تغییرات شاخص مناسبی است. اما در تصمیم‌گیری‌های کلان از ارزش آماری زیادی برخوردار نیست. زیرا ما به شاخص‌هایی نیاز داریم که هم پراکندگی داده‌ها و هم فراوانی آن‌ها را مدنظر قرار دهند.

دیدیم که پراکندگی یعنی این که داده‌ها از مرکز خود چه قدر دور هستند، پس یک راه ابتدایی ممکن است این‌طور به نظر برسد که تک‌تک مقادیر را از میانگین کم کنیم. این تفاضل‌ها را انحراف از میانگین می‌نامیم و سپس مجموع مقادیر حاصل را به‌دست آوریم، یعنی مجموع مقادیر

$$x_1 - \bar{x} \text{ و } x_2 - \bar{x} \text{ و } \dots \text{ و } x_n - \bar{x}$$

را حساب کنیم. قبلاً دیده‌ایم که مجموع مقادیر بالا برابر صفر است. لذا برای هر نوع داده‌ای اعم از آن که داده‌ها به میانگین نزدیک باشند و یا از آن دور باشند، آن چه که از این روش به‌دست می‌آید برابر صفر است. پس معیار مذکور نمی‌تواند معیار مفیدی باشد. البته علت آن که این مجموع برابر صفر شد به این خاطر است که برخی از داده‌ها از میانگین بزرگ‌تر و برخی دیگر کوچک‌ترند. در نتیجه مقادیر مثبت و منفی حاصل می‌شوند که مجموع آن‌ها یک‌دیگر را خنثی می‌کنند. برای جلوگیری از علامت‌های مثبت و منفی یکی از راه‌ها استفاده از قدر مطلق است. یعنی از مقادیر $|x_1 - \bar{x}|$ و \dots و $|x_n - \bar{x}|$

استفاده کنیم. این مقادیر می‌تواند معیاری برای پراکندگی ارائه کند ولی معمولاً کار کردن با قدر مطلق کار آسانی نیست و از این رو از توان دوم مقادیر بالا استفاده می‌کنیم یعنی مجموع مجذورات فاصله از میانگین را در نظر می‌گیریم. پس یک معیار پراکندگی ممکن است به صورت زیر باشد:

$$(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

ملاحظه می‌شود که اگر داده‌ها برابر باشند مقدار بالا برابر صفر است و بالعکس اگر مقدار بالا برابر صفر باشد داده‌ها برابر بوده و در نتیجه هیچ تفاوتی در بین آن‌ها نیست. اما اشکالی که دستور بالا دارد آن است که اگر تعدادی داده دیگر اضافه کنیم به مجموع بالا مقادیر مثبت دیگری اضافه خواهد شد و لذا این مجموع بزرگ‌تر خواهد شد. حال آیا درست است که بگوییم با اضافه شدن داده‌ها پراکندگی بیشتر می‌شود؟ مسلماً پاسخ منفی است، چون پراکندگی چیزی است که به اصل جامعه مربوط است و این نوع محاسبات نباید در مقدار آن‌ها تأثیر بگذارد. پس برای آن که تأثیر تعداد داده‌ها را از بین ببریم مجموع مذکور را بر تعداد آن‌ها تقسیم می‌کنیم و از این‌جا به معیار زیر می‌رسیم:

واریانس

واریانس برابر میانگین مجذور انحرافات از میانگین است و آن را با σ^2 نشان

می‌دهیم. بنابراین:

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

(σ از حروف کوچک یونانی است و سیگما خوانده می‌شود. حرف بزرگ آن Σ است)



تعداد ساعاتی که ۴ دانش‌آموز در طول هفته به ورزش اختصاص داده‌اند در زیر آمده است :
 واریانس این داده‌ها را حساب کنید.

۱، ۵، ۷، ۹

ابتدا میانگین آن‌ها را حساب می‌کنیم.

$$\text{میانگین : } x = \frac{1+5+7+9}{4} = \frac{22}{4} = 5/5$$

سپس انحراف از میانگین را حساب می‌کنیم.

انحرافات از میانگین : $-4/5, -0/5, 1/5, 3/5$

مجدور انحرافات از میانگین : $20/25, 0/25, 2/25, 12/25$

$$\text{میانگین مجدور انحرافات از میانگین : } \frac{20/25 + 0/25 + 2/25 + 12/25}{4} = \frac{35}{4} = 8/75$$

$$\sigma^2 = 8/75 \text{ پس}$$

اگر داده‌ها برابر باشند واریانس آن‌ها صفر است و بالعکس.

اگر داده‌ها برابر باشند، میانگین آن‌ها نیز همان مقدار مشترک داده‌ها خواهد بود. پس انحرافات از میانگین برابر صفر است. لذا مجموع مجدورات انحراف صفر خواهد بود. پس واریانسی که از تقسیم این مجموع بر تعداد داده‌ها به دست می‌آید برابر صفر است. حال فرض کنید واریانس صفر باشد. پس باید مجموع مجدورات انحرافات صفر شده باشد. مجدور انحرافات عددی نامنفی است و چون مجموع آن‌ها صفر است لازم می‌آید هر یک از آن‌ها صفر باشند، چون در غیر این صورت مجموع مجدور انحرافات صفر نخواهد شد. با صفر شدن هر یک از انحرافات نتیجه می‌شود که تمام داده‌ها برابر میانگین هستند. پس مساوی‌اند.



نکاتی دربارهٔ واحد میانگین و واریانس: فرض کنید مطالعه‌ای دربارهٔ قد افراد انجام می‌دهید، و داده‌های شما برحسب سانتی متر اندازه‌گیری و بیان شدند. از آن جایی که میانگین به وسیله مجموع داده‌ها به دست می‌آید آن هم برحسب سانتی متر خواهد بود. یعنی واحد میانگین از نوع همان واحد داده‌هاست. اما در واریانس شرایط فرق می‌کند زیرا در این جا شما مقادیر $(x_i - \bar{x})$ را که برحسب سانتی متر است به توان ۲ می‌رسانید، پس مجذور انحرافات بر حسب سانتی متر مربع است و لذا مجموع آن‌ها و در نتیجه واریانس برحسب سانتی متر مربع خواهد بود. بنابراین واحد واریانس از نوع مجذور واحد متغیر است.

دستور دیگری برای محاسبه واریانس: دیدیم که برای محاسبه واریانس باید میانگین مجذور انحرافات از میانگین را حساب کنیم، یعنی

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

محاسبه واریانس با استفاده از این دستور در بعضی مواقع مشکلاتی دارد.

می‌توان ثابت کرد عبارت بالا را می‌توان به صورت زیر نیز نوشت

$$\sigma^2 = \frac{1}{n} [x_1^2 + \dots + x_n^2] - \bar{x}^2$$

اثبات این دستور چندان مشکل نیست، توجه کنید که

$$\begin{aligned}\sigma^2 &= \frac{1}{n} [(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] \\ &= \frac{1}{n} [(x_1^2 - 2x_1\bar{x} + \bar{x}^2) + \dots + (x_n^2 - 2x_n\bar{x} + \bar{x}^2)] \\ &= \frac{1}{n} [(x_1^2 + \dots + x_n^2) - 2(x_1 + \dots + x_n)\bar{x} + n\bar{x}^2] \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - \frac{2(x_1 + \dots + x_n)}{n}\bar{x} + \bar{x}^2 \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - 2\bar{x} + \bar{x}^2 \\ &= \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2\end{aligned}$$

توجه کنید در $x_1^2 + \dots + x_n^2$ ابتدا مقادیر x_i ها را به توان ۲ می‌رسانیم و سپس آن‌ها را با هم جمع می‌کنیم و در \bar{x}^2 ، \bar{x} (میانگین داده‌ها) را به توان دو می‌رسانیم. حال واریانس داده‌های زیر را حساب می‌کنیم.

۱، ۵، ۶، ۷، ۹

$$\bar{x} = \frac{1+5+6+7+9}{5} = \frac{28}{5} = 5.6 \quad \text{ابتدا میانگین را حساب می‌کنیم:}$$

$$\begin{aligned}x_1^2 + \dots + x_5^2 &= 1 + 25 + 36 + 49 + 81 \\ &= 192\end{aligned}$$

$$\sigma^2 = \frac{192}{5} - (5.6)^2 \quad \text{پس:}$$

$$= 38.4 - 31.36 = 7.04$$

با استفاده مستقیم از دستور واریانس نیز به همین نتیجه می‌رسیدیم

$$\begin{aligned}\sigma^2 &= \frac{(1-5.6)^2 + (5-5.6)^2 + (6-5.6)^2 + (7-5.6)^2 + (9-5.6)^2}{5} \\ &= \frac{21.16 + 0.36 + 0.16 + 1.96 + 11.56}{5} \\ &= 7.04\end{aligned}$$



انحراف معیار

این که واحد واریانس از نوع مجذور واحد متغیر است می تواند مشکلات و سوء تفاهماتی را در پی داشته باشد. فرض کنید شما قدها را برحسب متر اندازه گیری کرده باشید و همکلاسی شما قدها را برحسب سانتی متر، در این صورت واریانسی که همکلاسی شما حساب می کند ۱۰۰۰۰ برابر واریانسی است که شما حساب کرده اید. همکلاسی شما ممکن است این طور نتیجه بگیرد که پراکندگی در قد بسیار زیاد و چشم گیر است و حال آن که نظر شما ممکن است غیر از این باشد. برای آن که این اختلاف نظرها را از بین ببریم سعی می کنیم که تفاوت عمده در واحد واریانس و واحد میانگین را با جذر گرفتن از واریانس از بین ببریم. جذر واریانس را انحراف معیار می گوئیم.

انحراف معیار که با نماد σ نشان داده می شود برابر جذر واریانس است.

واحد انحراف معیار همان واحد متغیر است.

تعداد ساعت هایی در هفته که ۴ نفر به صورت داوطلبانه در یک بیمارستان خدمت می کنند در

زیر آمده است. انحراف معیار داده ها را حساب کنید. ۱, ۳, ۴, ۵, ۶

$$\bar{x} = \frac{1+3+4+5+6}{5} = \frac{19}{5} = 3.8$$

$$\sigma^2 = \frac{(-2/8)^2 + (-0/8)^2 + (0/2)^2 + (1/2)^2 + (2/2)^2}{5}$$

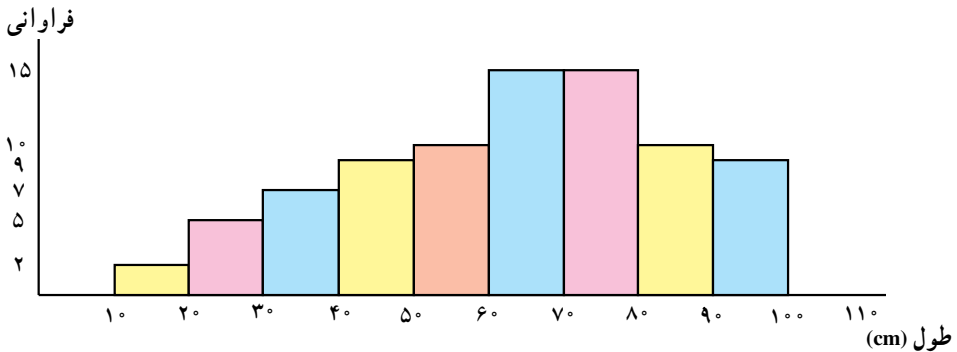
$$= \frac{7/84 + 0/64 + 0/04 + 1/44 + 4/84}{5}$$

$$= \frac{14/80}{5} = 2/96$$

$$\sigma = \sqrt{3/96} \approx 1/7$$

در نتیجه :

- نمودار زیر طول ۸۲ گیاه پرورش داده شده در یک آزمایشگاه را نشان می‌دهد.
 - طول چند گیاه بین ۵۰ تا ۶۰ سانتی‌متر است؟
 - طول چند گیاه بین ۷۰ تا ۹۰ سانتی‌متر است؟
 - میانگین و انحراف معیار طول گیاهان را به دست آورید.
 - توضیح دهید چه ویژگی نمودار بیان‌گر آن است که میانه و میانگین برابر نیستند.



- درجه حرارت بدن ۸ بیمار (به نزدیک‌ترین عدد صحیح گرد شده است) در زیر آمده است:

۳۸ ۳۸ ۳۹ ۳۹ ۴۰ ۴۰ ۴۱ ۴۱

میانگین و انحراف معیار را حساب کنید.

درجه حرارت‌های واقعی به قرار زیر می‌باشند:

۳۸/۰ ۳۸/۴ ۳۹/۰ ۳۹/۴ ۳۹/۸ ۴۰/۲ ۴۰/۹ ۴۱/۲

میانگین و انحراف معیار را حساب کنید.

گرد کردن داده‌ها چه تأثیری در میانگین و انحراف معیار داشت؟

آیا گرد کردن همواره تأثیری بر میانگین و انحراف معیار دارد؟

- در یک نمونه ۹۰۰ نفری از مردان درباره مسافتی که روزانه از منزل به محل کار طی می‌کنند

سؤال شد. اعداد صفحه بعد نتیجه این بررسی را نشان می‌دهد.

درصد	مسافت طی شده (کیلومتر)
۱۷	کمتر از ۱
۳۱	۱ و کمتر از ۳
۳۸	۳ و کمتر از ۸
۷	۸ و کمتر از ۱۰
۵	۱۰ و کمتر از ۱۵
۲	۱۵ و کمتر از ۵۰
۱۰۰	

– نمودار مستطیلی این داده‌ها را رسم کنید.

– میانگین و انحراف معیار را برای این داده‌ها به دست آورید.

– اگر میانگین مسافت طی شده توسط زنان کارمند ۵ کیلومتر باشد، با فرض آن که انحراف

معیار هر دو گروه یکسان است، آیا می‌توانید بگویید که مردان فاصلهٔ بیش‌تری طی می‌کنند یا

زنان؟

۴– دو گروه داده به صورت زیر داریم :

گروه اول : ۲ ۳ ۴ ۵ ۶

گروه دوم : ۴ ۵ ۶ ۷ ۸

الف – واریانس هر گروه را حساب کنید.

ب – این دو گروه داده چه رابطه‌ای با هم دارند؟

ج – رابطه موجود بین دو گروه داده، چه اثری روی واریانس‌ها داشته است؟

د – با توجه به تعریف واریانس، نتیجهٔ حاصل از قسمت ج را توضیح دهید.

۵– دو گروه داده به صورت زیر داریم :

گروه اول : ۱ ۲ ۳ ۴ ۵

گروه دوم : ۳ ۶ ۹ ۱۲ ۱۵

الف – واریانس هر گروه را حساب کنید.

ب – این دو گروه چه رابطه‌ای با هم دارند؟

ج – رابطه موجود بین دو گروه داده، چه تأثیری روی واریانس داشته است؟

د – نتیجه حاصل از قسمت ج را توضیح دهید.

۶- در حالت کلی ثابت کنید که اگر داده‌ها را با عددی جمع کنیم، در واریانس آن‌ها تغییری حاصل نخواهد شد. به زبان نمادها $\sigma_{a+x}^2 = \sigma_x^2$

۷- در حالت کلی ثابت کنید که اگر داده‌ها را در عددی ضرب کنیم واریانس آن‌ها در مجذور این عدد ضرب خواهد شد. به زبان نمادها $\sigma_{ax}^2 = a^2 \sigma_x^2$

۸- با ترکیب دو تمرین بالا نتیجه بگیرید که $\sigma_{ax+b}^2 = a^2 \sigma_x^2$

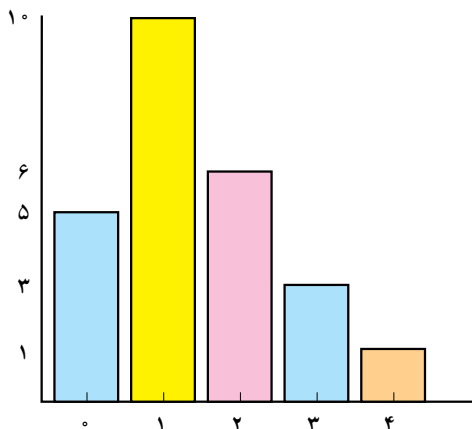
۹- کدامیک از برابری‌های زیر دربارهٔ انحراف معیار صحیح است؟

$$\sigma_{ax+b} = a\sigma_x \quad \sigma_{ax+b} = -a\sigma_x \quad \sigma_{ax+b} = |a|\sigma_x$$

۱۰- در یک بررسی تعداد فرزندان ۲۵ نفر از زنان مورد سؤال بود. نتیجه در نمودار زیر نشان داده شده است.

الف - نشان دهید متوسط تعداد فرزندان برابر $1/4$ است.

ب - نشان دهید انحراف معیار این داده‌ها تقریباً $1/6$ است.



- گروه ۲۵ نفری دیگری از زنان مورد بررسی قرار گرفتند. در این گروه متوسط تعداد فرزندان $2/4$ و انحراف معیار آن‌ها ۲ است. با توجه به اطلاعات به دست آمده در الف و ب، تفاوت بین تعداد فرزندان این دو گروه را بررسی کنید.

ضریب تغییرات

مسلماً قد همکلاسی‌های شما متفاوت است، همچنین وزن آن‌ها نیز چنین است ولی در سن همکلاسی‌های شما اختلاف زیادی نیست. یعنی انتظار می‌رود واریانس سن کم، ولی واریانس قد و وزن زیاد باشد. حال سؤال این است که تغییرات در قد همکلاسی‌ها بیش‌تر است یا در وزن آن‌ها. سن تقریباً ثابت است، زیاد متغیر نیست ولی می‌خواهیم بدانیم از قد و وزن کدام بیش‌تر متغیر است. اگر این دو صفت بر حسب یک واحد می‌بودند مقایسه واریانس آن‌ها می‌توانست جواب‌گوی این سؤال باشد ولی چون واحدها یکی نیستند مقایسه میزان پراکندگی این دو متغیر از طریق واریانس‌ها ممکن نیست. برای آن‌که بتوانیم به سؤال بالا جواب دهیم باید شاخصی معرفی کنیم که بدون واحد باشد تا امکان مقایسه فراهم آید. اگر به خاطر داشته باشید در جدول فراوانی برای آن‌که تأثیر اندازه نمونه را از بین ببریم به محاسبه درصد فراوانی نسبی پرداختیم. در این جا نیز برای از بین بردن واحد اندازه‌گیری از معیار ضریب تغییرات استفاده می‌کنیم.

ضریب تغییرات که با نماد CV نشان می‌دهیم عبارت است از خارج قسمت
انحراف معیار بر میانگین



تذکر: از آنجایی که ضریب تغییرات معیاری برای میزان پراکندگی است باید مثبت باشد، چون پراکندگی منفی معنی ندارد. پس لازم است \bar{x} ، که در مخرج کسر بالا آمده است، مثبت باشد. برای اطمینان از مثبت بودن \bar{x} ، ضریب اطمینان را فقط برای داده‌های مثبت تعریف می‌کنیم. ضریب تغییرات داده‌های زیر را محاسبه می‌کنیم:

۱، ۳، ۴، ۵، ۶

قبلاً دیدیم که $\bar{x} = 3/8$ و $\sigma = 1/7$ پس

$$CV = \frac{1/7}{3/8}$$

تذکر: هرچه قدر داده‌ها بزرگ‌تر باشند زمانی پراکندگی در آن‌ها محسوس است که تفاوت داده‌ها بیش‌تر باشد. ما اختلاف بین ۳ و ۳/۵ را به خوبی درک می‌کنیم ولی اختلاف بین ۳۰۰۰ و ۳۰۰۰/۵ چندان برای ما معنی‌دار نیست. شما در خرید اجناس ارزان قیمت انتظار تخفیف زیاد ندارید ولی در خرید اجناس گران قیمت از بحث و مجادله بر روی مبالغ کم خودداری می‌کنید. دلیل این امر آن است که در مقادیر بزرگ می‌توان از اختلاف‌های کم صرف‌نظر کرد. مثلاً از نظر آماری داده‌های زیر تقریباً با هم برابرند و تفاوت زیادی با هم ندارند.

۱۱۱، ۱۱۱/۵، ۱۱۲، ۱۱۲/۵

ولی داده‌های زیر با هم تفاوت دارند

۱، ۱/۵، ۲، ۲/۵

وقتی از ۱ به ۱/۵ می‌رویم پنجاه درصد به آن اضافه می‌شود ولی وقتی که از ۱۱۱ به ۱۱۱/۵ می‌رویم حدود چهارهزارم ۱۱۱ به آن اضافه می‌شود که ناچیز است. بنابراین پراکندگی به طریقی باید به نسبت بزرگی داده‌ها تعدیل شود. از این رو اگر ما انحراف معیار را بر \bar{x} تقسیم کنیم میزان پراکندگی را برای یک واحد از میانگین حساب کرده‌ایم و به این ترتیب تأثیر بزرگی داده‌ها را از بین برده‌ایم. با ملاحظه بالا می‌توان توصیف زیر را برای ضریب تغییرات به کار برد.

ضریب تغییرات عبارت است از میزان پراکندگی به ازای یک واحد از میانگین

تمرین

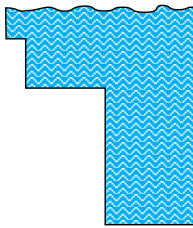
- ۱- اگر میانگین برابر ۴ و انحراف معیار برابر ۶ باشد ضریب تغییرات چه قدر است؟
- ۲- ضریب تغییرات سن دانش‌آموزان کلاس شما ۱۰ سال دیگر چه تغییری می‌کند؟ کم‌تر می‌شود یا بیش‌تر؟
- ۳- اگر ۲۰٪ نمره هر دانش‌آموز به نمره او اضافه شود چه تأثیری روی ضریب تغییرات نمره‌ها حاصل می‌شود؟
- ۴- دو دسته داده زیر قیمت کالایی را در دو بازار جداگانه نشان می‌دهد.

۸, ۱۳, ۹, ۱۲, ۱۰, ۱۱, ۱۲, ۹, ۱۰, ۱۱, ۱۰

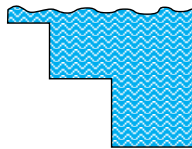
۱۰, ۱۳, ۸, ۱۰, ۹, ۱۱, ۱۰, ۱۰, ۱۱, ۹, ۱۰

- الف - در کدام دسته پراکندگی بیش‌تر مشهود است؟
- ب - دامنه تغییرات را در هر دو دسته حساب کنید.
- ج - آیا مقدار دامنه تغییرات با پراکندگی مشاهده شده در «الف» هم‌خوانی دارد؟
- د - ترجیح می‌دهید از کدام بازار خرید کنید؟ چرا؟
- ه - اگر داده‌های واقعی در اختیار شما نباشند، فقط به صرف داشتن دامنه تغییرات می‌توانستید تصمیم‌گیری کنید؟
- و - کدام یک از شاخص‌های پراکندگی در تصمیم‌گیری شما می‌تواند مؤثرتر باشد؟ این شاخص را معرفی کرده و محاسبه کنید.

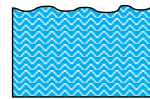
۵- در محله‌ای سه استخر شنا با مشخصات زیر موجود است :



میانگین عمق = ۱/۵ متر
انحراف معیار = ۱/۰۵



میانگین عمق = ۱/۵ متر
ضریب تغییرات = ۰/۵



میانگین عمق = ۱/۵ متر
انحراف معیار = ۰

- الف - برای فردی مبتدی با قد ۱/۴۵ متر شنا کردن در یک از این استخرها امنیت بیش‌تری دارد؟ چرا؟
- ب - در کدام استخر با خطر بیش‌تری روبه‌رو است؟ چرا؟

۶- دو دسته داده زیر ارزش واردات برحسب کشورهای مبدأ را در سال‌های ۶۵ و ۷۰ نشان

می‌دهد. (میلیون ریال)

کشور و منطقه	۱۳۵۵	۱۳۶۰	۱۳۶۵	۱۳۶۶	۱۳۶۷	۱۳۶۸	۱۳۶۹	۱۳۷۰
امارات	۳۲۴۳	۳۲۳۷۱	۴۱۳۴۳	۲۷۴۰۸	۱۹۱۹۹	۶۸۶۷۵	۶۵۴۵۲	۹۵۶۶۱
اندونزی	۴۱	۳۴	۲۱۱	۱۱۰۵	۷۳۴	۵۰۱۵	۴۳۳۵	۶۹۱۱
بنگلادش	۱۰۵۰	۳۰۹۸	۲۲۲۸	۲۶۷۷	۱۱۸۶	۳۱۹۶	۳۳۲۴	۳۰۹۶
پاکستان	۴۴۶۸	۶۷۲۱	۷۰۴۱	۴۳۰۴	۲۷۴۲	۱۰۴۵۵	۳۲۹۶	۱۲۴۰۳
تایلند	۱۳۷۵	۱۹۸۷۵	۴۰۲۵	۹۵۶۹	۴۸۹۳	۱۳۴۳۳	۱۳۲۷۷	۷۲۰۶
تایوان	۹۲۰	۱۰۵۱۰	۷۳۳۰	۴۶۱۹	۴۷۵۴	۵۲۹۱	۲۸۹۸	۱۰۲۹۸
ترکیه	۲۹۳۷	۲۳۳۶۹	۴۵۲۱۶	۳۳۳۳۱	۲۹۵۲۹	۵۰۶۴۵	۴۸۹۳۶	۶۹۶۳۳
چین	۷۰۹۳	۹۶۳۱	۱۹۹۳	۳۷۴۷	۲۴۸۶	۳۴۶۰	۱۱۶۱۳	۱۹۲۶۲
ژاپن	۱۵۵۵۰۰	۱۲۹۶۰۶	۹۷۶۲۷	۷۴۰۶۸	۵۸۰۶۶	۷۰۴۳۹	۱۳۰۲۴۸	۲۱۶۹۲۵
سنگاپور	۲۷۷۳	۸۲۵۴	۵۷۰۰	۳۸۱۶	۵۳۲۸	۵۵۵۴	۷۴۴۶	۹۰۳۰
سوریه	۲۷۳	۱۷۸	۹۱۴	۲۵۸۲	۶۸۸۳	۸۸۷۸	۲۰۰۵	۹۰۷
عربستان	۲۲	۱۶۹۱	۲۰۶۶	۷۳۲	۹۱۰	۶۳۸۲	۴۶۱	۲۰۸۶

– تغییرات ارزش واردات در کدام سال بیش‌تر بوده‌است؟

۷- با مراجعه به جدول صفحه بعد به سؤالات زیر پاسخ دهید.

۱۰ مرکز استان را به تصادف انتخاب کنید و بر اساس میزان بارندگی سالانه در این مراکز، میزان بارندگی سالانه در ایران در این سال را حدوداً به‌دست آورید.

– در این جامعه چند متغیر وجود دارد؟ نام ببرید و نوع آن‌ها را مشخص کنید.

– آیا میزان بارندگی سالانه در مراکز استان‌ها، به میزان بارندگی سالانه در ایران نزدیک است یا خیر؟

– برای اندازه‌گیری تفاوت‌ها از چه معیارهایی استفاده می‌کنید؟

– میزان بارندگی به طور متوسط در هر ماه در رشت چقدر است؟

سطر مربوط به بارندگی در استان گیلان (به مرکز رشت) معرف چه متغیری است؟ آن متغیر را

معرفی کنید و نوع آن را مشخص کنید.

مقدار بارندگی سالانه مراکز استانها

(میلیمتر)

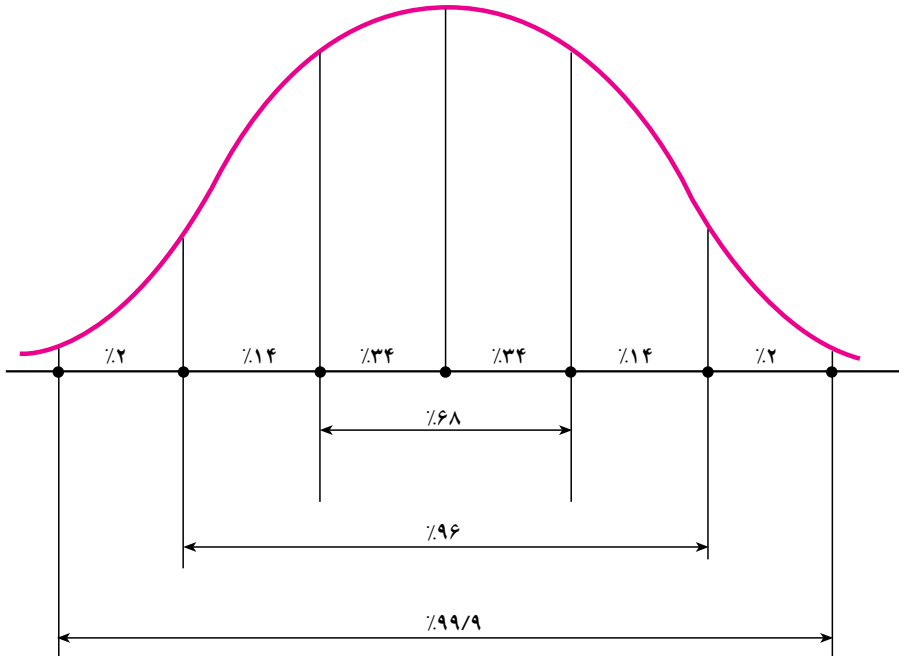
۱۳۷۶	۱۳۷۵	۱۳۷۴	۱۳۷۳	۱۳۷۲	۱۳۷۰	۱۳۶۵	مرکز استان
۲۵۶	۳۰۳	۳۶۷	۳۵۶	۳۲۰	۳۲۰	۳۴۸	اراک
۳۴۲	۳۴۷	۲۳۰	۳۴۴	۳۲۸	۰۰۰	۰۰۰	اردبیل
۲۶۹	۲۶۳	۳۶۷	۴۸۶	۵۳۴	۲۶۹	۳۸۶	ارومیه
۲۱۱	۶۷	۲۰۷	۷۹	۷۷	۱۱۲	۱۳۸	اصفهان
۵۳۹	۱۵۲	۲۳۹	۲۰۶	۱۶۸	۲۹۲	۲۸۶	اهواز
۷۱۷	۵۷۸	۵۴۹	۸۶۱	۷۴۱	۶۷۶	۰۰۰	ایلام
۴۳۷	۱۶۷	۵۰۷	۷۰	۳۶	۲۴۴	۲۳۰	بندرعباس
۸۰۷	۲۱۷	۵۸۸	۲۸۷	۶۴	۱۷۱	۱۱۹	بوشهر
۲۰۲	۲۵۹	۱۵۸	۳۳۶	۳۳۹	۲۳۶	۳۲۳	تبریز
۲۴۴	۱۵۶	۳۱۹	۲۵۵	۲۱۹	۲۶۰	۲۹۲	تهران
۶۰۹	۴۲۸	۵۳۷	۵۸۹	۵۸۷	۵۱۶	۵۹۴	خرم‌آباد
۱۲۹۷	۱۰۱۵	۱۲۱۰	۱۳۱۹	۱۶۸۵	۱۲۹۸	۱۱۸۶	رشت
۱۴۹	۷۳	۱۹۰	۵۶	۵۴	۴۴	۴۲	زاهدان
۳۳۱	۲۴۹	۲۶۶	۳۳۵	۴۲۵	۲۸۶	۳۲۱	زنجان
۱۶۱	۱۲۸	۱۹۲	۱۲۲	۶۷	۱۹۴	۲۲۲	سمنان
۵۷۶	۴۱۳	۵۱۳	۵۶۱	۵۹۹	۳۸۲	۵۴۸	سنندج
۴۲۳	۲۵۸	۳۷۹	۳۶۸	۲۸۲	۴۹۱	۴۹۲	شهرکرد
۴۹۸	۲۰۲	۵۷۸	۴۵۱	۱۵۵	۲۷۲	۴۹۵	شیراز
۳۳۶	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	قزوین
۱۳۸	۷۸	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	قم
۱۷۹	۱۲۳	۲۵۶	۱۷۶	۵۵	۱۱۵	۱۵۹	کرمان
۴۸۸	۳۹۶	۴۳۳	۵۵۹	۶۵۲	۴۵۱	۵۶۹	کرمانشاه
۵۳۶	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	۰۰۰	گرگان
۳۴۸	۱۷۰	۲۲۴	۱۶۷	۳۲۷	۳۲۶	۲۵۰	مشهد
۳۳۶	۲۱۳	۳۱۹	۴۳۸	۲۸۸	۲۹۲	۴۱۷	همدان
۱۱۹۸	۴۷۴	۱۰۲۵	۱۰۶۳	۴۱۷	۱۲۰۲	۱۲۲۸	یاسوج
۴۲	۲۹	۱۰۴	۳۴	۲۱	۶۰	۶۴	یزد

پراکندگی در منحنی نرمال

قبلاً در فصل نمودارها و همچنین در فصل شاخص‌های مرکزی به منحنی نرمال اشاره کردیم. اگر σ انحراف معیار این جامعه باشد پراکندگی افراد در جامعه به صورت زیر است. برای درک بهتر موضوع ابتدا فرض می‌کنیم مثلاً میانگین برابر ۵ و انحراف معیار برابر ۲ باشد در این صورت می‌خواهیم بدانیم چند درصد افراد در فاصله‌ای به مرکز میانگین و به فاصله یک انحراف معیار از آن قرار دارند. در واقع می‌خواهیم بدانیم چند درصد افراد در فاصله $(2-5, 5+2)$ یعنی $(3, 7)$ قرار دارند.

در این منحنی نشان داده می‌شود که این درصد برابر 68% است.

همچنین درصد افرادی که در فاصله‌ای به مرکز میانگین و به طول دو برابر انحراف معیار از میانگین قرار دارند برابر 96% است و سرانجام تقریباً 100% افراد در فاصله‌ای به مرکز میانگین و به طول ۳ برابر انحراف معیار قرار دارند. این اطلاعات را به طور مختصر در شکل زیر ملاحظه می‌کنید (مقادیر روی نمودار تقریبی هستند):



محاسبات شاخص‌های عددی در MINITAB

— داده‌های زیر ساعتی را که ۱۰ دانشجوی (انتخاب شده به‌طور تصادفی) صرف تحقیق در طول هفته می‌کنند نشان می‌دهد:

۷,۱۴,۵,۰,۲,۷,۱۰,۴,۰,۸

با استفاده از MINITAB میانگین، میانه، دامنه تغییرات و انحراف معیار این داده‌ها را به‌دست آورید.

شما می‌توانید از Calc و Stat در لیست انتخاب استفاده نمایید. Stat مقادیر چند شاخص عددی را هم‌زمان ارائه می‌دهد ولی Calc مقدار شاخص‌ها را یکی‌یکی به‌دست می‌آورد. با استفاده از MINITAB WINDOWS برای انتخاب Stat مراحل زیر را دنبال کنید:

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی Stat در لیست انتخاب، کلیک کنید.

قدم سوم: روی Basic Statistics کلیک کنید.

قدم چهارم: روی Descriptive Statistics کلیک کنید.

قدم پنجم: C1 را در کار زیر Variable تایپ کنید.

قدم ششم: روی OK کلیک کنید. نتایج و خروجی Minitab روی صفحه نمایان می‌شود.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، با استفاده از SET داده‌ها را در ستون

C1 وارد کنید. سپس فرمان زیر را تایپ کنید:

```
MTB > DESCRIBE C1
```

با اجرای این فرمان خروجی Minitab ظاهر می‌شود.

فعالیت

فعالیت مثال بالا را اجرا کنید و مقادیر به‌دست آمده را در جدول زیر بنویسید.

Descriptive Statistics

Variable	N	Mean	Median	Tr	Mean	Stder	SE	MEAN
hours	?	?	?	?	?	?	?	?
Variable	Min	Max	Q1	Q3				
hours	?	?	?	?				

همان طور که گفته شد می توان از دستور Calc نیز استفاده کرد، اگر از Minitab Windows استفاده می کنید به ترتیب زیر عمل کنید :

قدم اول: داده ها را در ستون C1 وارد کنید.

قدم دوم: روی Calc در لیست انتخاب، کلیک کنید.

قدم سوم: روی Column Statistics کلیک کنید.

قدم چهارم: در صفحه Column Statistics روی شاخصی که می خواهید محاسبه کنید کلیک کنید. به طور مثال برای پیدا کردن میانگین روی Mean کلیک کنید.

قدم پنجم: در کنار Input Variable در کادر، C1 را تایپ کنید.

قدم ششم: اگر می خواهید پاسخ را به عنوان یک عدد ثابت ذخیره کنید در کادر در کنار Store Result in، K1 را تایپ کنید. اگر از K1 قبلاً استفاده کرده اید K2 را تایپ کنید. اگر نمی خواهید نتیجه را ذخیره کنید، این مرحله را اجرا نکنید.

قدم هفتم: روی OK کلیک کنید. نتیجه روی صفحه نمایان خواهد شد.

اگر از فرمان های زبانی Minitab استفاده می کنید، با استفاده از SET داده ها را در ستون C1 وارد کنید. سپس فرمان زیر را تایپ کنید :

```
MTB > Mean C1
```

در پاسخ، Minitab خروجی زیر را ارائه می کند :

Column Mean

Mean of C1 =

به طور متشابه می توانید میانه، دامنه تغییرات و انحراف معیار را با کلیک کردن روی شاخص مورد نظر محاسبه کنید و فرمان های زبانی Minitab برای شاخص های فوق عبارتند از: MEDIAN C1، RANGE C1، STANDARD DEVIATION، که باید پس از PROMPT تایپ شوند.

نمودار جعبه ای

داده های زیر، درآمد هفتگی نمونه ۱۲ تایی از خانواده ها را نشان می دهد.

۲۳۰۰۰ ۱۷۰۰۰ ۳۲۰۰۰ ۶۰۰۰۰ ۲۲۰۰۰ ۵۲۰۰۰

۲۹۰۰۰ ۳۸۰۰۰ ۴۲۰۰۰ ۹۲۰۰۰ ۲۷۰۰۰ ۴۶۰۰۰

با استفاده از Minitab نمودار جعبه ای این داده ها را رسم کنید. اگر از Minitab Windows

استفاده می‌کنید، مراحل زیر را دنبال کنید :

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی BOXPLOT کلیک کنید.

قدم چهارم: C1 را در کادر زیر Y تایپ کنید.

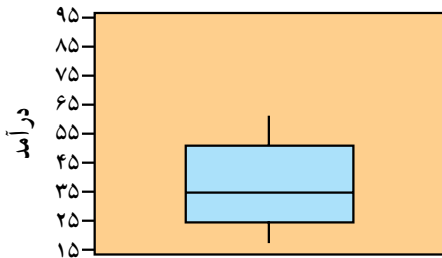
قدم پنجم: روی OK کلیک کنید. نمودار جعبه‌ای مورد نظر نمایان می‌شود. اگر از فرمان‌های

زبانی Minitab استفاده می‌کنید ابتدا با استفاده از فرمان SET داده‌ها را در ستون C1 وارد کنید.

سپس فرمان زیر را تایپ کنید :

MTB > Boxplot C1

با اجرای دستورات فوق نمودار جعبه‌ای زیر رسم می‌شود :



تمرین

۱- نمودار ساقه و برگ زیر موجود است.

- میانگین، انحراف معیار این داده‌ها را با استفاده از Minitab به دست آورید.

- نمودار جعبه‌ای این داده‌ها را با استفاده از Minitab رسم کنید.

ساقه	برگ
۰	۳۶۹
۱	۲۸۵۱۰۵
۲	۵۱۶
۴	۱
۵	
۶	۲