فصل ۷

آمار و احتمال

درس اوّل — احتمال شرطی و پیشامدهای مستقل

درس دوم — نمونه‌گیری

درس سوم — آمار توصیفی

احتمال شرطی و پیشامدهای مستقل

یادآوری : در پایه‌ها قبل با احتمال و کاربرد آن در مسائل آشنا شدیم. برخی مفاهیمی که با آنها آشنایی داریم، به شرح زیر است.

آزمایش: در علم احتمال، به هر عملی که برای جمع آوری داده‌ها صورت می‌پذیرد، آزمایش می‌گوییم.

آزمایش تصادفی: آزمایشی که نتایج آن از قبل قابل پیش بینی نباشد ولی مجموعهٔ نتایج حاصل از آزمایش قابل پیش بینی باشد را آزمایش تصادفی می‌گوییم.

مثال. وقتی یک سکه را پرتاب می‌کنیم از قبل نمی‌توانیم نتیجه بگیریم که روی سکه یا پشت سکه خواهد آمد، ولی مجموعه نتایج حاصل از پرتاب سکه قابل پیش بینی است. این مجموعه عبارت است از {پشت سکه، روی سکه}

فضای نمونه‌ای: مجموعه تمام نتایج ممکن در یک آزمایش تصادفی را فضای نمونه‌ای آن آزمایش می‌گوییم و آن را با S نشان می‌دهیم.

پرسش. فضای نمونه‌ای آزمایش‌های تصادفی زیر را بنویسید.

الف) پرتاب یک تاس

ب) خانوادهٔ دارای ۳ فرزند

پ) یک سکه و یک تاس

پیشامد تصادفی: هر زیر مجموعه از فضای نمونه‌ای را یک پیشامد می‌گوییم. پیشامدها را معمولا با حروف A، B، C و ... نشان می‌دهیم.

مثال. خانواده‌ای دارای سه فرزند است.

الف) فضای نمونه‌ای را بنویسید.

ب) پیشامدی را بنویسید که در آن حداقل ۲ فرزند این خانواده پسر باشد.

پ) پیشامدی را بنوسید که در آن هر سه فرزند دارای یک جنسیت باشند.

نکته. ترکیب‌های k تایی از n شی متمایز، به انتخاب‌های kتایی از آن n شی اطلاق می‌شود که در آنها ترتیب فاقد اهمیت است.

ترکیب‌های kتایی از n شی متمایز را با $\binom{n}{k}$ یا $C(n,\ k)$ (بخوانید انتخاب k از n) نمایش می‌دهیم. تعداد ترکیب‌های kتایی از n شی را از رابطهٔ زیر بدست می‌آوریم : $\binom{n}{k} = \frac{n!}{(n-k)!k!}$

تذکر. اگر در مسئله‌ای بین دو یا چند حالت کلمهٔ «و» بیاید، تعداد هر کدام از حالت‌ها را جداگانه بدست می‌آوریم و در هم ضرب می‌کنیم تا جواب نهایی بدست آید.

همچنین اگر در مسئله‌ای بین دو یا چند حالت کلمهٔ «یا» بیاید، تعداد هر کدام از حالت‌ها را جداگانه بدست می‌آوریم و باهم جمع می‌کنیم تا جواب نهایی بدست آید.

همچنین چنین نیست که همواره در صورت سوال کلمهٔ «و» یا «یا» ذکر شده باشد. بلکه گاها باید با دقت در مفهوم سوال این نکته را متوجه شویم.

مثال. در کیسه‌ای ۳ مهرهٔ سفید، ۴ مهرهٔ سیاه و ۵ مهرهٔ قرمز قرار دارد. از این کیسه ۲ مهره به تصادف و همزمان خارج می‌کنیم. مطلوبست :

الف) فضای نمونه‌ای.

ب) پیشامد آنکه هر دو مهره سیاه باشند.

پ) یکی از مهره‌ها سفید و دیگری قرمز باشد.

ت) هر دو مهره همرنگ باشند.

تعریف احتمال: اگر S فضای نمونه‌ای یک آزمایش و A یک پیشامد از فضای نمونه‌ای باشد، آنگاه احتمال وقوع پیشامد A را با P(A) نشان داده و از رابطهٔ زیر آن را بدست می‌آوریم : $P(A)=\frac{n(A)}{n(S)}$

برای حل مسائل احتمال، ابتدا فضای نمونه‌ای آزمایش مطرح شده و اندازهٔ آن را بدست می‌آوریم. سپس پیشامدی که احتمال وقوع آن در صورت سوال از ما خواسته شده را (پیشامد مطلوب) در صورت امکان نوشته و اندازهٔ آن را بدست می‌آوریم و در آخر از رابطهٔ $\frac{\text{تعداد حالت‌های مطلوب}}{\text{تعداد همهٔ حالت‌های ممکن}}=\frac{n(A)}{n(S)}=P(A)$ احتمال وقوع پیشامد را بدست می‌آوریم.

مثال. مادری دارای ۳ فرزند است. مطلوبست احتمال آنکه :

الف) حداکثر یکی از فرزندان پسر باشد.

ب) دو فرزند آخر پسر باشد.

پ) حداقل دو فرزند دختر باشد.

پرسش. از جعبه‌ای که شامل پنج مهرهٔ سفید و سه مهرهٔ سیاه است، دو مهره به تصادف خارج می‌کنیم. احتمال اینکه هر دو مهره سفید باشد چقدر است؟

پرسش. در آزمایشگاهی ۵ موش سالم و ۳ موش دیابتی نگهداری می‌شوند. اگر دو موش از محفظه گریخته باشند، با کدام احتمال فقط یکی از موش‌های فراری دیابتی است؟ (تجربی ۸۱)

۱) $\frac{۱۵}{۵۶}$　　۲) $\frac{۵}{۱۴}$　　۳) $\frac{۳}{۸}$　　۴) $\frac{۱۵}{۲۸}$

ترکیب پیشامدها: می‌دانیم هر پیشامد زیر مجموعه‌ای از فضای نمونه‌ای است. پس هر پیشامد را می‌توان بعنوان یک مجموعه در نظر گرفت. بنابراین اعمالی مانند متمم یک مجموعه، اجتماع یا اشتراک بین مجموعه‌ها نیز روی پیشامد قابل تعریف است.

الف) متمم یک پیشامد: اگر A پیشامدی از فضای نمونه‌ای $(A \subseteq S)$ باشد، آنگاه متمم آن پیشامدی است که در A رخ ندهد. متمم A را با نمادهای A' یا A^c یا $\overline{A}$ نمایش می‌دهیم و داریم :

$$\begin{cases} P(A') = ۱ - P(A) \\ P(A) = ۱ - P(A') \end{cases} \implies P(A) + P(A') = ۱$$

پرسش. دو تاس را باهم پرتاب می‌کنیم. احتمال اینکه عدد رو شده یکسان نباشد کدام است؟

پرسش. در یک کیسه ۵ مهرهٔ سفید و ۷ مهره سیاه موجود است. ۲ مهره به تصادف خارج می‌کنیم. احتمال اینکه دو مهره همرنگ نباشند کدام است؟ (ریاضی ۸۴)

۱) $\frac{۶}{۱۱}$ ۲) $\frac{۱۹}{۳۳}$ ۳) $\frac{۳۵}{۶۶}$ ۴) $\frac{۳۷}{۶۶}$

ب) اجتماع دو پیشامد: اگر A و B دو پیشامد از فضای نمونه‌ای S باشند، اجتماع آنها را به صورت $A \cup B$ نشان می‌دهیم. پیشامد $A \cup B$ وقتی اتفاق می‌افتد که حداقل یکی از آنها یعنی A یا B اتفاق بیفتد. مانند آنچه در مجموعه‌ها داشتیم، اعضای پیشامد $A \cup B$ یا در A وجود دارند یا در B و یا در هر دو آنها.

تذکر. هرگاه بین دو پیشامد از لفظ «یا» استفاده شود، به معنای اجتماع دو پیشامد است.

همچنین اگر حداقل یکی از دو پیشامد A یا B اتفاق افتاده باشد، بدان معنا است که $A \cup B$ اتفاق افتاده است.

پ) اشتراک دو پیشامد: اگر A و B دو پیشامد از فضای نمونه‌ای S باشند، اشتراک آنها را به صورت $A \cap B$ نشان می‌دهیم. پیشامد $A \cap B$ وقتی اتفاق می‌افتد که A و B هر دو اتفاق بیفتند. اعضای پیشامد $A \cap B$ هم در A وجود دارند و هم در B.

تذکر. هرگاه بین دو پیشامد از لفظ «و» استفاده شود، به معنای دو پیشامد است.

نکته. قانون جمع احتمالات: اگر A و B دو پیشامد از فضای نمونه‌ای S باشند در این صورت :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

ت) پیشامد غیر ممکن: اگر در آزمایشی، پیشامدی را تعریف کنیم که امکان وقوع آن نباشد، آن پیشامد را غیر ممکن می‌گوییم و آن را با نماد Ø نشان می‌هیم. و چون مجموعه تهی عضوی ندارد پس :

$$P(\emptyset) = \frac{n(\emptyset)}{n(S)} = \frac{۰}{n(S)} = ۰$$

ث) پیشامد حتمی: می‌دانیم هر مجموعه، زیر مجموعه خودش است. پس اگر S پیشامدی از فضای نمونه‌ای S باشد در این صورت :

$$P(S) = \frac{n(S)}{n(S)} = ۱$$

مثال. از بین اعداد {۱, ۲, ۳, . . . , ۲۰} عددی به تصادف انتخاب می‌کنیم. مطلوبست محاسبهٔ احتمال اینکه :

الف) عدد انتخابی زوج یا مضرب ۵ باشد.

ب) عدد انتخابی اول و فرد باشد.

پرسش. در یک مدرسه ۲۰٪ دانش آموزان در کلاس فیزیک و ۳۰٪ در کلاس ریاضی و ۱۵٪ در هر دو کلاس ثبت نام کرده‌اند. مشخص کنید چند درصد این دانش آموزان در کلاس فیزیک یا ریاضی ثبت نام کرده‌اند؟

پرسش. فرض کنید در جامعه‌ای درصد گروه‌های خونی به شرح زیر باشد. فردی به تصادف از این جامعه انتخاب می‌شود. احتمال این را بیابید که :

O	AB	B	A	گروه خونی
٪۲۳	٪۲۷	٪۱۲	٪۳۸	درصد

الف) این فرد دارای گروه خونی A یا O باشد.

ب) این فرد دارای گروه خونی AB نباشد.

ج) **پیشامدهای ناسازگار:** اگر دو پیشامد A و B نتوانند باهم رخ دهند، آن دو پیشامد را ناسازگار می‌گوییم. پس A و B دو پیشامد ناسازگارند، اگر داشته باشیم : $A \cap B = \emptyset$

نکته. اگر A_1، A_2، . . . و A_n پیشامدهایی باشند که دو به دو نتوانند باهم رخ دهند، می‌گوییم این پیشامدها دو به دو ناسازگارند. (یعنی هر دوتا از این مجموعه‌ها را در نظر بگیرید، اشتراک آنها تهی است.)

در این صورت داریم : $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

مثال. فرض کنید درون جعبه‌ای ۵ مهرهٔ آبی، ۳ مهرهٔ سبز و ۸ مهرهٔ سیاه باشد. از این جعبه یک مهره به تصادف بر می‌داریم. احتمال اینکه مهرهٔ انتخاب شده، آبی یا سبز یا سیاه باشد را حساب کنید.

چ) **پیشامدهای مستقل :**

سوال. دو تاس را باهم پرتاب می‌کنیم. آیا ۶ آمدن تاس اول تاثیری بر روی ۶ آمدن تاس دوم دارد؟

تعریف. اگر دو پیشامد به گونه‌ای باشند که وقوع یا عدم وقوع یکی در وقوع یا عدم وقوع دیگری تاثیری نداشته باشد، می‌گوییم دو پیشامد مستقل از هم هستند.

اگر A و B دو پیشامد مستقل باشند در این صورت داریم : $P(A \cap B) = P(A) \,.\, P(B)$

در مثال پرتاب دو تاس، نتیجه پرتاب تاس اولی تاثیری بر نتیجهٔ پرتاب تاس دوم ندارد. پس اگر A پیشامد ۶ آمدن تاس اول و B پیشامد ۶ آمدن تاس دوم باشد، احتمال اینکه هر دو تاس ۶ بیاید (یعنی احتمال وقوع پیشامد $A \cap B$)، به صورت زیر محاسبه می‌شود :

$$\begin{cases} P(A) = \frac{۱}{۶} \\ P(B) = \frac{۱}{۶} \end{cases} \implies P(A \cap B) = P(A) \,.\, P(B) = \frac{۱}{۶} \times \frac{۱}{۶} = \frac{۱}{۳۶}$$

تذکر. یک روش دیگر برای حل این مسئله آن است که به کمک نمودار درختی فضای نمونه‌ای را نوشته سپس پیشامد مطلوب را استخراج کرده و از فرمول کلاسیک احتمال، احتمال مورد نظر را محاسبه کنیم.

پرسش. مادری صاحب سه فرزند است.

الف) احتمال آنکه دو فرزند اول پسر باشند چقدر است؟

ب) احتمال آنکه فقط دو فرزند اول پسر باشند چقدر است؟

نکته. برای آنکه فردی دارای RH منفی باشد، لازم است که هر دو ژن منفی داشته باشد. یعنی :

P(RH منفی) = P(هر دو ژن منفی) = P(یک ژن منفی) . P(یک ژن منفی)

مثال. مطالعات ژنتیکی نشان داده است که ۴۰٪ ژن‌های تعیین کنندهٔ عامل RH خونی منفی اند. مطلوبست احتمال آنکه :

الف) فردی دارای RH منفی باشد.

ب) در خانواده‌ای دارای سه فرزند، اولین فرزند با RH منفی، فرزند سوم خانواده باشد.

پرسش. فرض کنید در یک سال احتمال قهرمانی تیم ملی فوتبال ایران در آسیا برابر ۰/۵ و احتمال قهرمانی تیم ملی والیبال ایران در آسیا برابر ۰/۸ باشد. با چه احتمالی حداقل یکی از این تیم‌ها قهرمان خواهد شد؟

پرسش. احتمال اینکه سیما در یک امتحان قبول شود $\frac{۱}{۳}$ و احتمال اینکه مینا در همان امتحان قبول شود $\frac{۱}{۵}$ است. مطلوبست احتمال آنکه سیما یا مینا در این امتحان قبول شوند؟

احتمال شرطی :

عوامل ژنتیک در شکل گیری صفات نقش دارند و از والدین به فرزندان منتقل می‌شوند. آیا تاکنون دقت کرده‌اید که نرمهٔ گوش یک انسان می‌تواند دو حالت داشته باشد، یکی پیوسته و یکی آزاد. با توجه به این موضوع

سوالاتی از این قبیل می‌توانند مطرح باشند :

اگر مردی نرمهٔ آزاد و همسرش نرمهٔ گوش پیوسته داشته باشد، آیا می‌توان پیش بینی کرد که فرزند آنها چه نوع نرمهٔ گوشی خواهد داشت.

در ادامه به کمک علم احتمال به مسئله بالا پرداخته خواهد شد.

سوال. فرض کنید از ظرفی که شامل ۵ مهرهٔ سفید و ۴ مهرهٔ سیاه است، مهره‌ای به تصادف خارج کردیم. احتمال آنکه این مهره سفید باشد چقدر است؟

حل. در این مثال فضای نمونه‌ای ۹ عضو دارد که ۵ عضو آن برای پیشامد مورد نظر مساعد است. پس احتمال آمدن مهرهٔ سفید برابر $\frac{۵}{۹}$ است.

این احتمال بطور مطلق حساب شد و از هیچ اطلاع اضافی استفاده نشد. حال مثال زیر را در نظر بگیرید.

مثال. از جعبهٔ مثال قبل مهره‌ای خارج می‌کنیم ملاحظه می‌شود که رنگ آن سیاه است. این مهره را کنار گذاشته و مهرهٔ دوم را به تصادف خارج می‌کنیم. مطلوبست احتمال آنکه این مهره سفید باشد.

حل. در این مثال پس از کشیدن مهرهٔ اول و با توجه به اطلاعات داده شده (سیاه بودن مهرهٔ خارج شده) شرایط هنگام استخراج مهرهٔ دوم عبارت است از وجود ۵ مهرهٔ سفید و ۳ مهرهٔ سیاه در جعبه، پس احتمال آمدن یک مهرهٔ سفید از این جعبه $\frac{۵}{۸}$ است.

همانطور که ملاحظه می‌شود اگر چه در دو مثال بالا احتمال آمدن مهرهٔ سفید از یک جعبه را حساب کردیم ولی جواب‌ها یکسان نیستند. زیرا در مثال دوم اطلاعاتی داریم که احتمال آمدن مهرهٔ سفید را تغییر می‌دهد.

تعریف : اگر A و B دو پیشامد از فضای نمونه‌ای باشند و $P(B) > ۰$ در این صورت احتمال وقوع پیشامد A به شرط اینکه بدانیم پیشامد B قبلا رخ داده باشد را به صورت $P(A \mid B)$ نوشته و چنین تعریف می-

کنیم : $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

نتیجه. احتمال وقوع پیشامد B به شرط آنکه بدانیم پیشامد A قبلا رخ داده است را به صورت $P(B \mid A)$ نوشته و چنین تعریف می‌کنیم : $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$

مثال. اگر $P(A \cap B) = ۰/۲$ و $P(B) = ۰/۶$ آنگاه $P(A \mid B)$ را حساب کنید.

پرسش. احتمال وقوع نوعی بیماری در یک جامعهٔ مشخص برابر ۰/۰۴ و احتمال اینکه فردی هم دچار این بیماری شود و هم درمان یابد برابر ۰/۰۲ است. اگر فردی به بیماری مذکور دچار شده باشد، احتمال درمان یافتن او چقدر است؟

پرسش. در یک مسابقهٔ اتومبیل رانی احتمال اینکه یک اتومبیل دچار نقص فنی نشود و به خط پایان نیز برسد برابر ۰/۷ است و احتمال اینکه یک اتومبیل دچار نقص فنی نشود برابر ۰/۸ است. اگر بدانیم یک اتومبیل دچار نقص فنی نشده است، با چه احتمالی به خط پایان رسیده است؟

پرسش. فرض کنید احتمال اینکه یک تیم فوتبال اصلی‌ترین رقیبش را ببرد $\frac{۱}{۶}$ باشد. احتمال قهرمانی این تیم در حال حاضر $\frac{۱}{۴}$ و در صورتی که اصلی‌ترین رقیبش را ببرد $\frac{۱}{۳}$ خواهد بود. با چه احتمالی حداقل یکی از دو اتفاق قهرمان شدن یا برد اصلی‌ترین رقیب برای این تیم اتفاق خواهد افتاد؟

پرسش. در علم ژنتیک برای ایجاد برخی صفات در فرزندان دو عامل را موثر می‌دانند که یکی از پدر و یکی از مادر به ارث می‌رسد. فرض کنیم دو عامل مذکور را برای تعیین شکل نرمهٔ گوش فرزند یکی با A و دیگری با a نمایش دهیم که در آن :

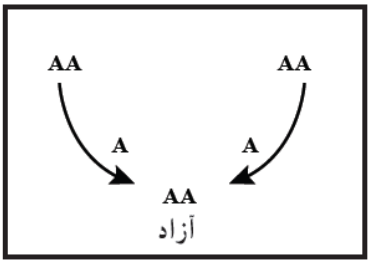
A : عامل به وجود آمدن نرمهٔ گوش آزاد

a : عامل به وجود آمدن نرمهٔ گوش پیوسته

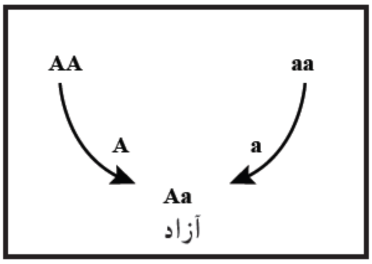
بنابر این هر فرد به یکی از حالت‌های AA یا aa یا $a\mathrm{A}$ می‌تواند باشد که با احتمال $\frac{۱}{۲}$ هر یک از آنها را به فرزند خود می‌تواند انتقال دهد و تاثیر آن بر نرمهٔ گوش فرزند به صورت زیر است :

- اگر عامل‌های فرزند به صورت AA باشد، این فرد دارای نرمهٔ گوش آزاد است.
- اگر عامل‌های فرزند به صورت aa باشد، این فرد دارای نرمهٔ گوش پیوسته است.
- اگر عامل‌های فرزند به صورت $\mathrm{A}a$ باشد، این فرد دارای نرمهٔ گوش آزاد است به همین دلیل عامل A را غالب و عامل a را مغلوب می‌نامند.

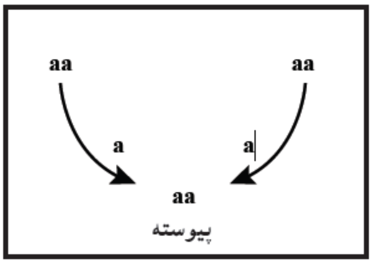
در شکل‌های زیر حالت‌های مختلف انتقال عوامل از پدر و مادر به فرزند نمایش داده شده‌اند.



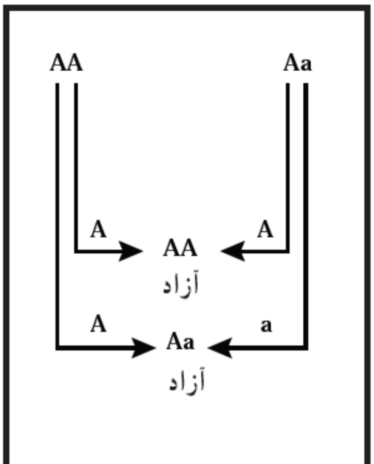
(۱)



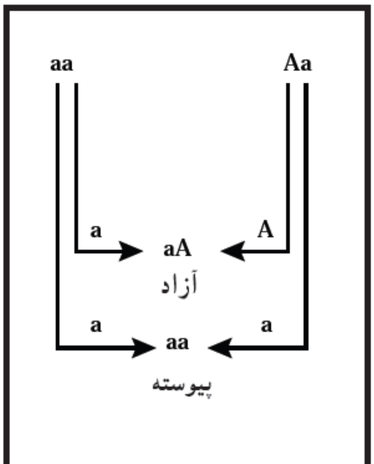
(۲)



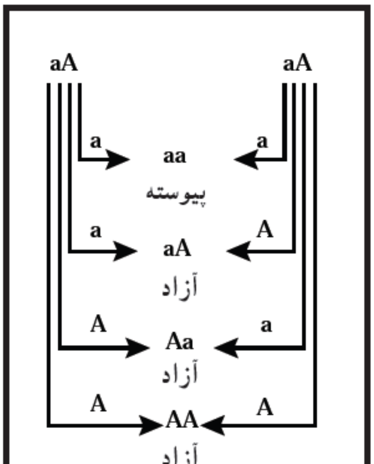
(۳)



(۴)



(۵)



(۶)

فرض کنیم احتمال هر یک از دو عامل هر فرد به فرزندش $\frac{۱}{۲}$ باشد. اگر از میان افراد یک جامعهٔ آماری که نرمهٔ گوش آزاد دارند ۵۰ درصدشان خالص و ۵۰ درصدشان ناخالص باشند، هر یک از احتمال‌های زیر را محاسبه کنید.

اگر علی نرمهٔ گوش آزاد و همسرش نرمهٔ گوش پیوسته داشته باشد، و آنها یک فرزند با نرمهٔ گوش پیوسته داشته باشند با چه احتمالی فرزند دوم آنها نرمهٔ گوش پیوسته خواهد داشت؟

پرسش. ترکیبی از ۴ مادهٔ شیمیایی داریم که دوتا از آنها مواد A و B هستند. احتمال واکنش نشان دادن ماده A، $\frac{۱}{۵}$ و احتمال واکنش دادن مادهٔ B، $\frac{۱}{۷}$ است. اگر مادهٔ A واکنش نشان دهد، احتمال واکنش نشان دادن مادهٔ B، $\frac{۱}{۴}$ خواهد شد. با چه احتمالی حداقل یکی از مواد A یا B واکنش نشان خواهد داد؟

پرسش. در آزمایشگاهی ۳ موش سیاه و ۲ موش سفید داریم، قبلا یک موش سفید را برای آزمایشی انتخاب کرده ایم. حال می‌خواهیم همان آزمایش را روی موش دیگری انجام دهیم. برای این منظور موشی را از بین موش‌هایی که روی آنها آزمایش انجام نشده است انتخاب می‌کنیم. مطلوبست احتمال آنکه این موش نیز سفید باشد.

احتمال شرطی و استقلال پیشامدها :

قبلا گفتیم که اگر دو پیشامد A و B مستقل از هم باشند، وقوع یکی بر وقوع دیگری تاثیری نمی‌گذارد. این مطلب را می‌توانیم با توجه به تعریف احتمال شرطی، به صورت زیر ثابت کنیم :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \implies P(A \mid B) = \frac{P(A).P(B)}{P(B)} \implies P(A \mid B) = P(A)$$

پس احتمال اینکه پیشامد A اتفاق بیفتد به شرط رخ دادن پیشامد B، با احتمال A برابر است. یعنی وقوع پیشامد B تاثیری روی وقوع پیشامد A نداشته است. پس می‌توان نتیجه بگیریم که اگر A و B دو پیشامد مستقل باشند داریم : $P(A \mid B) = P(A) \quad , \quad P(B \mid A) = P(B)$

قاعدهٔ ضرب احتمال : (محاسبهٔ احتمال اشتراک دو پیشامد با استفاده از احتمال شرطی)

از قاعدهٔ احتمال شرطی می‌توانیم $P(A \cap B)$ را بصورت زیر محاسبه کنیم :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Longrightarrow P(A \cap B) = P(A \mid B) \times P(B)$$

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \Longrightarrow P(B \cap A) = P(B \mid A) \times P(A)$$

مثال. دو مهره متوالیا و بدون جایگذاری از جعبه‌ای که شامل ۴ مهرهٔ سفید و ۶ مهرهٔ سیاه است خارج می‌کنیم. مطلوبست احتمال آنکه مهرهٔ اول سفید و مهرهٔ دوم سیاه باشد.

پرسش. دو مهره متوالیا و بدون جایگذاری از جعبه‌ای که شامل ۶ مهرهٔ سفید و ۸ مهرهٔ سیاه است خارج می‌کنیم. مطلوبست احتمال آنکه مهرهٔ اول سفید و مهرهٔ دوم سیاه باشد.

قانون احتمال کل :

اگر فضای نمونه‌ای (S) را به n قسمت $A_1, A_2, \ldots, A_n$ افراز کنیم (یعنی اولا به ازای هر $1 \le i \le n$ داشته باشیم $A_i \neq \emptyset$ ثانیا به ازای هر i,j که $i \neq j$، $A_i \cap A_j = \emptyset$ و ثالثا $A_1 \cup A_2 \cup \ldots \cup A_n = S$)

در این صورت :

$$P(A) = P(A_1)P(A \mid A_1) + P(A_2)P(A \mid A_2) + \ldots + P(A_n)P(A \mid A_n)$$

$$\Longrightarrow P(A) = \sum_{i=1}^{n} P(A_i)P(A \mid A_i)$$

مثال. فرض کنید انتقال نوعی بیماری ارثی از والدین به فرزند پسر ۰/۱۴ و به فرزند دختر ۰/۰۸ باشد. والدینی که حامل این نوع بیماری هستند انتظار فرزندی را دارند. مطلوبست احتمال اینکه فرزند سالم باشد.

پرسش. ۴۵٪ جمعیت کشوری را زنان و ۵۵٪ بقیه را مردان تشکیل می‌دهند. اگر ۸ درصد زنان و ۱۰ درصد مردان تحصیلات دانشگاهی داشته باشند، چند درصد جمعیت این کشور تحصیلات دانشگاهی دارند؟

پرسش. سه ظرف هم شکل و یکسان داریم. در ظرف اول ۲ مهرهٔ سفید و ۳ مهرهٔ سیاه، در ظرف دوم ۳ مهرهٔ سفید و ۴ مهرهٔ سیاه و در ظرف سوم ۴ مهرهٔ سفید و ۲ مهرهٔ سیاه وجود دارد. یکی از ظرفا را به تصادف انتخاب کرده و مهره‌ای از آن خارج می‌کنیم. احتمال اینکه مهره سفید باشد چقدر است؟

پرسش. احتمال موفقیت عمل جراحی برای شخص A برابر ۰/۹ و برای شخص B برابر ۰/۸ است. با کدام احتمال، لااقل عمل جراحی برای یکی از این دو نفر، موفقیت آمیز است؟ (تجربی ۹۵)

۱) ۰/۹۲ ۲) ۰/۹۴ ۳) ۰/۹۶ ۴) ۰/۹۸

پرسش. در جعبه‌ای ۳ مهرهٔ سفید، ۲ مهرهٔ سیاه و ۵ مهرهٔ قرمز موجود است. اگر دو مهره از آن بیرون آوریم، با کدام احتمال این دو مهره همرنگ نیستند؟ (تجربی ۹۴)

۱) $\frac{۲۸}{۴۵}$ ۲) $\frac{۲۹}{۴۵}$ ۳) $\frac{۳۱}{۴۵}$ ۴) $\frac{۳۲}{۴۵}$

پرسش. در کیسه‌ای ۵ مهرهٔ سفید و ۴ مهرهٔ سیاه و ۳ مهرهٔ قرمز موجود است. اگر سه مهره از کیسه خارج کنیم، با کدام احتمال حداکثر ۳ مهره از مهره‌های خارج شده همرنگ هستند؟ (ریاضی ۹۵)

۱) $\frac{۱۷}{۲۲}$ ۲) $\frac{۱۹}{۲۲}$ ۳) $\frac{۳۹}{۴۴}$ ۴) $\frac{۴۱}{۴۴}$

پرسش. از بین ۳ کارت سفید و ۴ کارت سبز یکسان به تصادف یک کارت بدون جایگذاری بیرون می‌آوریم. سپس کارت دوم را خارج می‌کنیم. با کدام احتمال هر دو کارت همرنگ هستند؟ (تجربی ۹۱)

۱) $\frac{۱}{۷}$ ۲) $\frac{۵}{۱۴}$ ۳) $\frac{۳}{۷}$ ۴) $\frac{۴}{۷}$

پرسش. دو تاس سالم را باهم پرتاب می‌کنیم تا برای اولین بار هر دو عدد رو شده زوج باشند. با کدام

احتمال حداکثر در سه پرتاب نتیجه حاصل می‌شود؟ (تجربی ۹۱)

۱) $\frac{۲۷}{۶۴}$ ۲) $\frac{۳۷}{۶۴}$ ۳) $\frac{۱۹}{۳۲}$ ۴) $\frac{۳۹}{۶۴}$

پرسش. در گروه زنان ساکن یک روستا ۶۰ درصد آنان تحصیلات ابتدایی و ۲۵ درصد آنان مهارت قالی بافی دارند. اگر یک فرد از این گروه انتخاب شود، با کدام احتمال این فرد تحصیلات ابتدایی یا مهارت قالی بافی دارد؟ (تجربی ۹۰) ۱) ۰/۷ ۲) ۰/۷۵ ۳) ۰/۸ ۴) ۰/۸۵

پرسش. در یک خانواده ۴ فرزندی با کدام احتمال ۲ فرزند پسر یا ۳ فرزند دختر است؟ (تجربی ۹۰)

۱) $\frac{۳}{۸}$ ۲) $\frac{۹}{۱۶}$ ۳) $\frac{۵}{۸}$ ۴) $\frac{۳}{۴}$

پرسش. دو تاس را باهم می‌اندازیم. با کدام احتمال دو عدد رو شده، متوالی هستند؟ (تجربی ۹۵ خارج)

۱) $\frac{۲}{۹}$ ۲) $\frac{۵}{۱۸}$ ۳) $\frac{۷}{۱۸}$ ۴) $\frac{۴}{۹}$

نمونه گیری

در سال گذشته علم آمار به عنوان مجموعه روش‌هایی شامل جمع آوری اعداد و اطلاعات (بعنوان داده‌ها) ، سازماندهی و نمایش داده‌ها، تحلیل و تفسیر داده‌ها و در نهایت نتیجه گیری، قضاوت و پیش بینی در مورد پدیده‌ها و آزمایش‌های تصادفی تعریف شد. در این درس ضمن مرور مطالب گذشته، با جمع آوری داده‌ها و روش‌های نمونه گیری آشنا خواهیم شد.

جامعه و نمونه :

جامعه آماری : مجموعه‌ای از افراد یا اشیاء است که دربارهٔ اعضای آن می‌خواهیم موضوع یا موضوعاتی را مطالعه کنیم. یا به عبارت دیگر جامعه مجموعه‌ای از اشیا است که در یک یا چند صفت مشترک هستند، و برحسب موضوع مورد مطالعه تعیین می‌شود.

مثال. الف) بانوان ورزشکار استان گلستان   موضوع مورد مطالعه: مدت زمان ورزش بانوان در هفته

ب) محصولات کشاورزی استان گلستان   موضوع مورد مطالعه: انواع محصولات

اندازه جامعه : تعداد اعضای جامعه را اندازه جامعه می‌گویند.

سرشماری : اگر تمام افراد جامعه را مورد مطالعه قرار دهیم، می‌گوییم سرشماری کرده‌ایم.

مشکلات سرشماری : در دسترس نبودن تمام اعضای جامعه، وقت گیر بودن دسترسی به تمام اعضای جامعه، هزینه بر بودن، از بین رفتن جامعه در برخی مطالعات

نمونه آماری : زیرمجموعه‌ای از جامعه آماری است که بتوان بیانگر جامعه باشد.

عمل نمونه گیری : مهمترین بخش آمار را عمل نمونه گیری تشکیل می‌دهد. برای آنکه نمونه بتواند به درستی نمایانگر خصوصیت تمام جامعه باشد، باید به اندازه کافی بزرگ باشد و همچنین اعضای نمونه باید تصادفی انتخاب شوند.

اندازه نمونه : تعداد اعضای نمونه را اندازه نمونه می‌گویند.

مثال. سازمان میراث فرهنگی صنایع دستی و گردشگری، می‌خواهد میزان رضایت گردشگران را از سفر به گلستان در سال ۱۳۹۶ ارزیابی نماید.

جامعه: کلیه گردشگرانی که در سال ۱۳۹۶ به گلستان مسافرت کرده‌اند، جامعه آماری تشکیل می‌دهند.

نمونه: گردشگرانی که در شش ماهه اول سال ۱۳۹۶ از جزیره آشوراده دیدن کرده‌اند، نمونه آماری را تشکیل می‌دهند.

نمونه تصادفی : زیرمجموعه‌ای از جامعه آماری است که اعضای آن به طور تصادفی انتخاب می‌شوند.

روش انتخاب نمونه تصادفی (منظور از اعضای نمونه باید تصادفی انتخاب شوند)

الف) هر یک از اعضا امکان حضور داشته باشند.

ب) قبل از انتخاب نمونه نتوانیم، یا اطمینان بیشتر درباره حضور یا عدم حضور عده‌ای در نمونه قضاوت کنیم.

اعداد تصادفی : نمونه گیری تصادفی ساده به روش‌های مختلفی انجام می‌گیرید، یکی از آن روش‌ها استفاده از ماشین حساب است.

نحوه استفاده از عدد تصادفی : برای این منظور اعداد تولید شده را در اندازه جامعه ضرب می‌کنیم و اگر عدد حاصل قسمت اعشاری داشت، قسمت اعشاری را حذف کرده و یک واحد به آن اضافه می‌کنیم. مثلا اگر جامعه ما ۱۰۰ عضو داشته باشد و اعداد تصادفی تولید شده به وسیله ماشین حساب به صورت زیر باشد، در این صورت خواهیم داشت :

عدد تصادفی	۰/۰۳۴	۰/۰۲۹	۰/۹۵۵	۰/۷۰۲
۱۰۰×عدد تصادفی	۳/۴	۲/۹	۹۵/۵	۷۰/۲
شمار ه انتخاب شده	۴	۳	۹۶	۷۱

داده : نتایج حاصل از اندازه گیری یا بررسی نمونه را داده می‌گویند.

روش‌های جمع‌آوری داده‌ها :

۱) از طریق پرسش (پرسشنامه کتبی، پرسش شفاهی از اشخاص (مصاحبه))

مثلا: «تابستان فصل مورد علاقه بیشتر مردم برای مسافرت است» و «پراید ماشین محبوب ایرانیان برای استفاده است»

بهترین روش جمع آوری داده‌ها استفاده از روش پرسشنامه است.

۲) از طریق مشاهده و ثبت وقایع

مثلا: «افرادی که در ساعت خاص وارد یک ساختمان می‌شوند» و «بررسی استقبال مردم از نمایشگاه»

۳) از طریق آزمایش

مثلا: «اگر در مطالعه‌ای مربوط به میزان هموگلوبین افراد را نیاز داشته باشیم باید آزمایش خون بدهند.» و «اثر نور خورشید بر رشد گیاهان»

۴) استفاده از داده‌های از پیش تهیه شده

مثلا: «اطلاعات بهداشتی پیرامون بیماری‌های واگیر یا غیرواگیر، سلامت مادر و کودک، آلودگی هوا و ... از کشورهای مختلف در درگاه سازمان بهداشت جهانی (WHO[1]) در دسترس است.» و «ارتباط بین نمره درس ریاضی و معدل کل دانش‌آموزا»

بررسی آماری : روش جمع آوری داده‌ها بدون دخالت وضع آنها بررسی آماری نام دارد.

پرسش. برای موضوع‌های زیر جامعه و نمونه مناسب معرفی کنید.

الف) بررسی PH شامپوهای تولیدی یک کارخانه

ب) بررسی علاقه دانش آموزان به ورزش در دبیرستان فراغی بندرترکمن

پ) بررسی علل طلاق در شهر گرگان

ت) مدت زمانی که دانش آموزان کلاس شما در طول یک هفته صرف مطالعه کتاب‌های غیر درسی می‌کنند

ث) هزینه‌های مبارزه با مواد مخدر در کشورهای آسیایی

ج) بررسی وضعیت شغلی (ترس از بیکاری) کارمندان شاغل در شرکت‌های خصوصی

[1] World Health Organization

پرسش. فرض کنید در یک کلاس ۴۰ نفری می‌خواهید با استفاده از عدد تصادفی دو نفر را انتخاب کنید. اگر اعداد تصادفی ۰/۷۶۵ و ۰/۰۲۳ ظاهر شود، نفرهای چندم کلاس انتخاب خواهند شد.

پرسش. بهترین روش جمع آوری داده‌ها را برای بررسی فرضیه‌های زیر بیان کنید.

الف) تاثیر مصرف چیپس در امراض قلبی

ب) سفید، رنگ مورد علاقه بیشتر مردم برای ماشین است.

پ) بیشتر تصادفات اتومبیل‌ها را رانندگان با سن کمتر از ۲۵ سال مرتکب می‌شوند.

ت) بررسی میزان استقبال از انتخابات در روز انتخابات

ج) رژیم گرفتن، موجب کاهش هوش می‌شود.

متغیر تصادفی : همان موضوعی است که روی جامعه یا نمونه مورد مطالعه قرار می گیرد .و این موضوع از فردی به فرد دیگر تغییر می کند.

مثال. فرض کنید می خواهیم بدانیم دانش آموزان یک کلاس چه قدر پول همراه دارند، برای این که بتوانیم مثال را تعقیب کنیم، یک نمونه تصادفی ۸ تایی از این کلاس انتخاب می کنیم، که عبارتند از:

$A =$ {محمدامین، شهاب، موسی، جمشید، تقی، اکبر، رحیم}، اگر پول‌های این افراد را معلوم کنیم، مثلا مجموعه بالا به مجموعه $B = \{۵۰۰, ۱۲۰۰, ۱۳۰۰, ۱۵۰۰, ۲۰۰۰, ۲۵۰۰, ۱۵۰۰\}$ تبدیل می‌شود. آنچه برای ما مهم است مجموعه B می‌باشد.

متغیر تصادفی مجموعه هشت فرد را تبدیل به مجموعه ای از اعداد کرد .یعنی هم چنانچه می بینیم فرد برای ما مهم نیست بلکه آن چه مهم است اطلاعاتی است که در ارتباط با متغیر تصادفی مورد مطالعه این فرد در اختیار ما قرار داده شده است.

متغییر تصادفی :

انواع متغییر تصادفی: به دو دسته تقسیم می‌شوند، متغییر تصادفی کمی و متغییر تصادفی کیفی

الف) متغییر تصادفی کمی : متغییرهایی هستند که قابل اندازه گیری باشند (یعنی می‌توانیم به آنها عدد نسبت دهیم). مانند قد، وزن، سن، مدت زمان استفاده از اینترنت، مقدار شاخص توده بدنی، میزان هموگلوبین و ...

ب) متغییر تصادفی کیفی : متغیرهایی هستند که قابل اندازه گیری نباشند (یعنی نمی‌توانیم به آنها عدد نسبت دهیم). مانند گروه خونی، جنسیت افراد، کیفیت کالای تولید شده و ...

انواع متغییر تصادفی کمی :

۱. متغییر کمی پیوسته : متغیرهایی هستند که اگر دو مقدار a و b را بتواند اختیار کند، هر مقدار بین آنها را نیز بتوانند اختیار کنند. مانند قد دانش آموزان یک کلاس، اگر قد یک دانش آموز ۱۷۰ cm و قد دانش آموز دیگر ۱۷۱ cm باشد، وزن دانش آموزان دیگر می‌تواند بین ۱۷۰ cm و ۱۷۱ cm باشد.

۲. متغییر کمی گسسته : متغییرهای کمی هستند که پیوسته نباشد. مانند تعداد بیماران مراجعه کننده به یک بیمارستان.

نکته. متغییرهای تصادفی پیوسته عموما از راه اندازه گیری (یا ابزار) بدست می‌آیند. مثل قد، وزن، درجه حرارت.

متغییرهای تصادفی گسسته عموما از راه شمارش بدست می‌آیند. مثل تعداد دندان‌های فاسد یک کودک.

انواع متغییر تصادفی کیفی :

۱. مغییر کیفی ترتیبی : مغییرهای کیفی هستند که در آنها نوعی ترتیب طبیعی وجود دارد. مثل مراحل زندگی انسان که شامل نوزادی، کودکی، نوجوانی، جوانی، بزرگسالی و پیری می‌باشد. یا حروف الفبای فارسی که به ترتیب شامل الف، ب، ... می‌باشد.

۲. متغییر کیفی اسمی : متغییرهای کیفی هستند که ترتیبی نباشند. مثل گروه خونی افراد، رنگ ماشن، جنسیت افراد، نوع کشت.

پرسش. نوع متغییرهای تصادفی را در هر یک از عبارات زیر مشخص نمایید.

نوع متغیر	متغیر تصادفی	نوع متغیر	متغیر تصادفی
	تعداد نامه های یک صندوق		متغیری که معمولاً از نوع تعداد می باشد.
	وزن نامه های یک صندوق		غذای مورد علاقه
	زمان انتظار بیمار در مطب		نمره درس آمار
	تعداد بیماران مراجعه کننده		تعداد سوالات امتحانی درس آمار
	جنسیت افراد یک شهر		مدت زمان معاینه یک فرد
	میزان تحصیلات (دیپلم، کارشناسی،...)		مراحل تحصیل
	وضعیت مسکن (مستأجر، مالک)		متغیری که هر نوع عدد اعشاری یا صحیح را می توان به آن ها اختصاص داد.
	میزان مالیات پرداختی سالانه		نوع کشت (برنج، گندم و..)
	میزان آلودگی هوا		تعداد مکالمات تلفنی
	مقاومت یک ترانزیستور		طول مکالمات تلفنی
	تعداد شکایات رسیده به یک پاسگاه		گنجایش آب یک تانکر
	وضعیت تأهل		روزهای هفته

پرسش. می خواهیم مدت زمانی که دانش آموزان کلاس شما صرف مطالعه کتاب های غیر درسی می‌کنند بررسی کنیم.

الف) در این بررسی جامعه را مشخص نمایید.

ب) برای این بررسی یک نمونه مشخص نمایید.

پ) از چه روشی برای جمع آوری اطلاعات استفاده می کنید.

ت) اندازه این جامعه چه قدر است؟

ث) اندازه نمونه مورد بررسی چه قدر است؟

آیا این تعداد برای بررسی مورد نظر مناسب است؟

ج) متغیر مورد مطالعه در این مسأله چیست؟

چ) متغیر مورد مطالعه از چه نوعی است؟

پرسش. در موضوع‌های زیر متغییر تصادفی را مشخص کنید.

الف) بررسی اندازه قد دانش آموزان ب) رنگ لباس موجود در یک فروشگاه

پ) جنسیت افراد یک شهر ت) درجه حرارت هوا در ساعت ۱۰ صبح روزهایی تیرماه سال جاری

پرسش. مراحل تحصیل افراد، کدام نوع مغییر است؟ (انسانی ۸۸)

۱) کمی پیوسته ۲) کمی گسسته ۳) کیفی اسمی ۴) کیفی ترتیبی

پرسش. در کدام بررسی اندازه نمونه برابر اندازه جامعه است؟ (تجربی ۸۹)

۱) نمونه تصادفی ۲) دسته بندی ۳) سرشماری ۴) با مغییر کیفی

آمار توصیفی

آمار توصیفی به خلاصه سازی داده‌ها در قالب نمودار، جدول و یا شاخص‌هایی در قالب معیارهای گرایش به مرکز و معیارهای پراکندگی می‌پردازد. آمار توصیفی اطلاعی از چگونگی داده‌های جمع آوری شده فراهم می‌آورد که بسیار مفید است.

دسته بندی داده‌ها و جدول فراوانی :

فراوانی مطلق : تعداد دفعاتی که یک شی یا عدد تکرار می‌شود را فراوانی مطلق آن شی یا عدد می‌گویند و

با f_i نشان می‌دهیم.

دامنه تغییرات : اختلاف بین بزرگترین داده و کوچکترین داده را دامنه تغییرات می‌گویند و با R نشان می‌هیم. $R =$ بزرگترین داده - کوچکترین داده $= x_{max} - x_{min}$

جدول فراوانی : جدولی که شامل دسته‌ها و یا حالت‌هاست و در مقابل هر یک از آنها فراوانی مطلق آنها نوشته شده است.

نکته. در بررسی جدول توزیع فراوانی دو حالت کلی زیر را در نظر می‌گیریم :

الف) اگر تعداد داده‌های آماری کم باشد، در این حالت جدولی رسم می‌کنیم که دو ستون (یا دو سطر) دارد. در ستون اول (یا سطر اول) مقادیری را می‌نویسیم که متغییر انتخاب می‌کند و عموما با x_i نشان می-دهند و در ستون دوم (سطر دوم) فراوانی مطلق مربوط به هر متغییر را می‌نویسیم.

ب) اگر تعداد داده‌های آماری زیاد باشد، در این حالت باید داده‌ها را دسته بندی کنیم و برای هر دسته حدودی قایل شویم.

باز شبیه حالت اول جدولی رسم می‌کنیم که دو ستون (یا دو سطر) دارد. در ستون اول (یا سطر اول) دسته‌ها را می‌نویسیم و در ستون دوم (یا سطر دوم) فراوانی مطلق مربوط به دادهٔ موجود در هر دسته را می‌نویسیم.

مراحل پیشنهادی برای دسته بندی داده‌ها :

۱) ابتدا تصمیم می‌گیریم که داده‌ها را در چند دسته، دسته بندی کنیم. این تعداد را k بنامید. و برای تعیین k، دستور خاصی پیشنهاد نمی‌کنیم (با اینکه بعضی از روشهای تجربی وجود دار مثلا $k = ۱ + ۳/۳\log n$) ولی توصیه می‌کنیم که دسته‌ها آنقدر کوچک نباشند که هیچ مقداری در آن دسته قرار نگیرد و همچنین آنقدر بزرگ نباشد که تعداد زیادی از داده‌ها را در خود جا دهد. (هرچه نمونه بزرگتر باشد k نیز می‌تواند بزرگتر باشد.)

۲) از رابطهٔ زیر طول دسته را بدست آورید. $\text{طول دسته} = \frac{\text{دامنه تغییرات}}{\text{تعداد دسته}}$ $\quad c = \frac{R}{K}$

۳) حال با داشتن c (طول دسته) و a (کوچکترین داده) می‌توانیم دسته‌ها را تشکیل دهیم.

دسته اول $[a, a + \text{c})$

دسته دوم $[a + \text{c}, a + ۲\text{c})$

.

.

دسته آخر $[\text{b} - \text{c}, \text{b}]$

نکات :

۱- در دسته اول به a کران پایین و به $a + \text{c}$ کران بالا می‌گویند، به همین ترتیب برای سایر دسته‌ها نیز کران بالا و پایین تعریف می‌شود.

۲- در تمامی دسته‌ها طول دسته باهم برابرند یعنی برابر c می‌باشد.

۳- این دسته‌ها باید به گونه‌ای باشد که اشتراک نداشته باشند و تمام داده‌ها را بپوشاند.

مرکز هر دسته : مرکز هر دسته برابر است با نصف مجموع کران پایین و کران بالای همان دسته، یعنی :

$$\text{مرکز هر دسته} = \frac{\text{کران پایین همان دسته} + \text{کران بالای همان دست}}{۲}$$

نکته. تفاضل دو مرکز دسته متوالی برابر طول دسته می‌باشد.

انواع فراوانی :

الف) فراوانی مطلق: تعداد داده‌ها یا افراد موجود در هر دسته را فراوانی مطلق آن دسته می‌گویند و با f_i نشان می‌دهند.

نکته. در یک جدول فراوانی مجموع فراوانی‌های مطلق برابر با تعداد کل داده‌ها (n) می‌باشد.

ب) فراوانی نسبی هر دسته: برابر است با نسبت فراوانی مطلق هر دسته به تعداد کل داده‌ها، یعنی :

$$\text{فراوانی نسبی هر دسته} = \frac{\text{فراوانی مطلق همان دسته}}{\text{تعداد کل داده‌ها}} = \frac{f_i}{n}$$

نکته. در یک جدول فراوانی مجموع فراوانی‌های نسبی برابر یک است.

درصد فراوانی نسبی هر دسته : برابر است با حاصل ضرب فراوانی نسبی هر دسته در ۱۰۰ ، یعنی :

۱۰۰ × فراوانی نسبی همان دسته = درصد فراوانی نسبی هر دسته

نکته. در یک جدول فراوانی مجموع درصد فراوانی‌های نسبی برابر ۱۰۰ است.

فراوانی تجمعی هر دسته : مجموع فراوانی‌های مطلق هر دسته بعلاوه فراوانی‌های مطلق دسته‌های قبل از آن را فراوانی تجمعی آن دسته می‌گویند و با F_{ci} نشان می‌دهند. یعنی مثلا

$$F_{c4} = f_1 + f_2 + f_3 + f_4 = (f_1 + f_2 + f_3) + f_4 = F_{c3} + F_{c4}$$

بطور کلی $F_{cn} = f_1 + f_2 + \dots + f_{n-1} + f_n = F_{cn-1} + f_n \implies f_n = F_{cn} - F_{cn-1}$

فراوانی تجمعی دسته ما قبل آخر − فراوانی تجمعی دسته آخر = فراوانی مطلق دسته آخر

نکته. فراوانی تجمعی دسته اول برابر فراوانی مطلق دسته اول می‌باشد و فراوانی تجمعی دسته آخر برابر تعداد کل داده‌ها است.

پرسش. داده‌های ۱۵۰ ,۱۵۵ ,۱۷۰ ,۱۷۰ ,۱۷۵ ,۱۵۵ ,۱۵۰ ,۱۵۵ مربوط به قد ۸ نفر از دانش آموزان یک کلاس می‌باشد. جدول توزیع فراوانی آن را بنویسید.

پرسش. داده‌های زیر را در یک جدول فراوانی خلاصه کنید.

۴, ۱۵, ۱۴, ۱۰, ۲۱, ۹, ۶, ۵, ۹, ۱۶, ۱۷, ۱۴, ۱۳, ۱۱, ۱۲, ۸, ۱۶, ۱۳, ۱۸, ۱۱

پرسش. نمرات درس راضی۲ چهل دانش آموز به صورت زیر گزارش شده است :

۴, ۶, ۷, ۷, ۷/۵, ۸, ۸/۵, ۸/۵, ۹, ۹, ۹, ۱۰, ۱۰, ۱۱, ۱۱, ۱۱, ۱۱, ۱۲, ۱۲, ۱۲/۵, ۱۳, ۱۳, ۱۳, ۱۳,

۱۳, ۱۳, ۱۳/۵, ۱۴/۵, ۱۴/۵, ۱۵, ۱۵, ۱۵, ۱۶, ۱۶, ۱۷, ۱۷, ۱۸, ۱۸, ۱۸, ۱۸, ۱۹

جدول دسته بندی شده این نمرات را به همراه مرکز دسته‌ها، فراوانی نسبی دسته‌ها، درصد فراوانـی نسـبی دسته‌ها و فراوانی تجمعی دسته‌ها رسم کنید.

پرسش. آمار مراجعه کنندگان به یک سینما در طی یک ماه از سال (سی روز) به صورت زیر می‌باشد،

۲۴۰, ۲۴۲, ۲۴۴, ۲۴۵, ۲۳۰, ۲۲۰, ۱۸۰, ۲۳۱, ۲۴۵, ۲۴۸, ۲۵۰, ۲۵۱, ۲۵۵, ۲۳۰, ۲۲۲, ۱۹۰, ۲۱۱,

۲۲۰, ۲۳۰, ۲۳۲, ۲۴۰, ۲۲۵, ۲۳۰, ۲۶۰, ۲۷۰, ۲۰۴, ۲۲۵, ۲۳۰, ۲۴۰, ۱۹۵

جدول توزیع فراوانی داده‌ها را همراه فراوانی نسبی، درصد فراوانی نسبی و فراوانی تجمعی رسم نمایید.

پرسش. اگر در یک جدول توزیع فراوانی تعداد دسته‌ها ۲۰ و طول دسته‌ها ۵ باشد، تفاوت داده کـوچکتر از داده بزرگتر را بیابید.

پرسش. اگر مراکز دسته‌های جدول فراوانی ۱۸ ,۱۴ ,۱۰ ,۶ ,۲ باشد، دسته‌ها را مشخص کنید.

پرسش. در داده‌های آماری بدست آمده از یک تجربه، کمترین داده ۳۲ و بیشـترین داده ۷۲ بـوده و طـول دسته‌ها ۵ در نظر گرفته شده، حد پایین دسته دوم چقدر است؟

پرسش. اگر فراوانی تجمعی طبقه پنجم ۳۲ و فراوانی مطلق آن ۸ باشد، فراوانی تجمعی طبقه چهارم را بیابید.

پرسش. فراوانی نسبی طبقه‌ای ۰/۰۵ و تعداد داده‌ها برابر ۸۰ است، فراوانی مطلق را بیابید.

پرسش. داده‌های آماری پیوسته در ۸ طبقه دسته بندی شده‌اند بطوری که آخرین دسته به صورت ۹۲-۸۶ نوشته شده است. کوچکترین داده را بیابید.

پرسش. در جدول توزیع ۶۰ داده آماری، فراوانی تجعی دسته پنجم ۲۱ و فراوانی تجمعی دسته ششم ۲۴ است. فراوانی نسبی دسته ششم چقدر است؟

پرسش. فراوانی نسبی ۴۰ نفر از افرادی که در یک روز به یک پایگاه اهداء خون مراجعه کرده‌اند، در جدول زیر آمده است. چه تعداد از این افراد دارای گروه خونی B بوده‌اند؟

گروه خونی	A	B	AB	O
فراوانی نسبی	۰/۲	x	۰/۲۵	۰/۴

نمودارها و تحلیل داده‌ها :

در این بخش سعی می‌کنیم اطلاعات موجود در جدول فراوانی را از طریق هندسی یا نمودارها به نمایش بگذاریم، زیرا از طریق مشاهده نمودار یک توزیع فراوانی، به خصوصیات مجموعه اطلاعات آماری، نسبت به نگرش از طریق جدول توزیع، به آسانی می‌توان پی برد. لذا می‌توان گفت که نمودار ها وسیله سودمند برای به تصویر در آوردن جامعه آماری می‌باشند .که برخی از این نمودارها عبارتند از: نمودار میله ای، نمودار مستطیلی، نمودار چند بر فراوانی (خط شکسته)، نمودار دایره ای، نمودار جعبه‌ای، نمودار ساقه و برگ

نمودار میله‌ای : این نمودار بیشتر برای متغییرهای «کیفی» و «کمی گسسته» مناسب است.

برای رسم نمودار میله‌ای در صفحه محورهای مختصات، روی محور xها، متغییرها (مراکز دسته) را با واحدهای مساوی جدا می‌کنیم و روی محور yها فراوانی مطلق همان دسته را قرار می‌دهیم.

نمودار مستطیلی : این نمودار برای متغییرهای کمی پیوسته مناسب است.

از این نمودار وقتی که جدول دسته بندی توزیع فراوانی وجود داشته باشد و تعداد داده‌ها زیاد باشد استفاده می‌کنیم. برای رسم آن مستطیل‌هایی رسم می‌کنیم که قاعده آنها روی محور xها و برابر طول هر یک از دسته‌ها و ارتفاع آن به موازات محور yها، و متناسب با فراوانی مطلق (نسبی) دسته‌ها است.

نمودار خطی (خط شکسته) : این نمودار برای متغییرهای کمی پیوسته مناسب است.

برای رسم آن نقاط (x_i, f_i)، (x_i مرکز دستهٔ iام و f_i فراوانی مطلق آن دسته) را در صفحه محورهای مختصات رسم می‌کنیم. سپس این نقاط را به وسیله خط مستقیم به هم وصل می‌کنیم.

نکته. نمودار خطی تغییرات متغییر پیوسته را بهتر از نمودار مستطیلی نشان می‌دهد.

تذکر. اگر نمودار خطی را بجای نمودار فراوانی مطلق، از روی فراوانی نسبی رسم کنیم، اطلاعات منسجم تری در اختیار ما قرار می‌گیرد چون می‌توان فراوانی کل را با کل جامعه مقایسه کرد.

نمودار دایره‌ای : این نمودار برای متغییرهای کمی گسسته یا کیفی مناسب است.

برای رسم آن سطح دایره را به قطاع‌هایی تقسیم می‌کنیم که سطح هر قطاع متناسب با مقدار فراوانی داده مورد نظر. اگر فراوانی دسته برابر f_i باشد، زاویهٔ نظیر آن برابر است با $\alpha_i = \frac{f_i}{n} \times ۳۶۰^\circ$ (n مجموع فراوانی‌های مطلق یعنی کل داده‌هاست)

نکته. نمودار دایره‌ای بر مبنای فراوانی نسبی رسم می‌شود.

پرسش. توزیع فراوانی معلمان علوم پایه یک دبیرستان به صورت زیر است. نمودار میله‌ای را رسم کنید.

دبیر	ریاضی	فیزیک	زیست	شیمی
فراوانی	۶	۳	۲	۵

پرسش. آمار تعداد تصادفات اتومبیل در یک جاده معین در سی روز متوالی یک ماه به صورت زیر ارایه شده است. نمودار مستطیلی و نمودار خطی داده‌ها در ۵ دسته رسم کنید.

۶, ۲, ۳, ۱۰, ۷, ۹, ۸, ۱, ۲, ۴, ۹, ۹, ۱۰, ۱۰, ۱۱, ۱۲, ۵,

۶, ۴, ۱۰, ۱۴, ۱۹, ۲۰, ۲۱, ۲۰, ۱۸, ۱۵, ۱۲, ۸, ۴,

پرسش. جدول توزیع نمودار دایره‌ای توزیع دانش آموزان را رسم کنید.

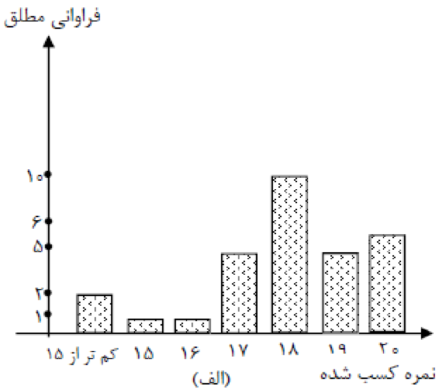
مقطع	اول	دوم	سوم	پیش دانشگاهی
تعداد دانش آموزیک دبیرستان و مرکز پیش دانشگاهی	۷۵	۹۰	۹۰	۴۵

پرسش. در یک نمودار دایره‌ای تعداد افرادی که در دسته‌های B، C و D قرار دارند، به ترتیب ۳، ۲ و ۶ برابر تعداد افرادی است که در دسته A قرار دارند. زاویهٔ مرکزی مربوط به هر دسته را بدست آورید.

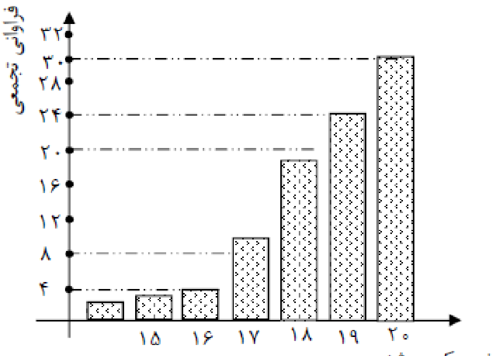
پرسش. در یک امتحان کتبی از درس کمک‌های اولیه، افرادی که نمره کمتر از ۱۵ (از ۲۰ نمره) کسب نکنند، می‌توانند در دوره عملی شرکت کنند. نمودارهای زیر نتیجه آزمون را نشان می‌دهد. (نمرات به شکل اعداد صحیح اند)

الف) با توجه به نمودارها تعداد افرادی را که می‌توانند در دوره شرکت کنند را مشخص کنید.

ب) کدام نمودار شما را سریع به پاسخ هدایت می‌کند؟



(الف)



(ب)

پرسش. جدول زیر، سهم وزنی ترکیبات تشکیل دهنده یک بسته غذایی کنسرو شده را نشان می‌دهد.

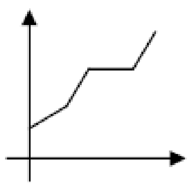
پروتئین	چربی	فیبر	کربوهیدرات	دسته
۶۰°	۳۰°	۷۰°	۲۰۰°	زاویه مرکزی

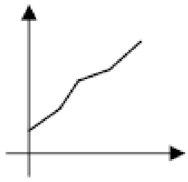
الف) چه کسری از این ترکیبات چربی است؟

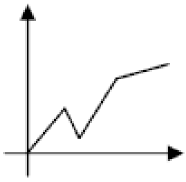
ب) چه کسری از این ترکیبات کربوهیدرات است؟

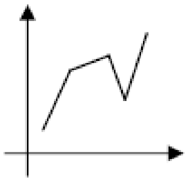
پ) چند گرم پروتئین در یک بسته ۳۶ گرمی از این محصول وجود دارد؟

پرسش. کدام یک از شکل‌های زیر نمودار فراوانی تجمعی است؟

الف) 

ب) 

ج) 

د) 

پرسش. تعداد دانش آموزان سه کلاس مدرسه‌ای ۷۰، ۶۰، و ۵۰ نفر می‌باشد. در نمودار دایره‌ای زاویه مربوط به کلاس ۵۰ نفری چند درجه است؟

پرسش. جدول مقابل درصد فراوانی نسبی گروه خونی افراد یک جامعه است. در نمودار دایره‌ای زاویه سطح مربوط به گروه خونی O چند درجه است؟

O	AB	B	A	گروه خونی
α	۳۶	۲۲/۵	۲۴	درصد فراوانی نسبی

پرسش. در توزیع فراوانی داده‌های پیوسته، کدام نمودار مناسب است؟ (تجربی ۸۷)

۱) مستطیلی ۲) خطی ۳) میله‌ای ۴) دایره‌ای

معیارهای گرایش به مرکز (شاخص‌های مرکزی) :

معمولا سعی می‌شود تا دانسته‌های نهفته در داده‌ها را به صورت یک یا چند عدد شاخص در آورد، تا بتوان هم ایده کلی درباره ویژگی مورد مطالعه بدست آورد و هم نتیجه مطالعات را را به سادگی گزارش کرد. میانگین و میانه بعنوان معیارهای گرایش به مرکز در این کتاب معرفی می‌شوند.

میانگین : اصلی‌ترین و مورد استفاده‌ترین شاخص مرکزی، میانگین است. میانگین یعنی معدل داده‌ها که آن را با μ (یا $\bar{x}$) نشان می‌دهند. یعنی اگر داده‌ها بصورت $x_1, x_2, \ldots, x_n$ باشند، در این صورت میانگین آنها از رابطه زیر بدست می‌آید : $\mu = \frac{x_1 + x_2 + \ldots + x_n}{n}$

نکاتی در مورد میانگین :

۱) عموما برای نشان دادن مجموع داده‌ها از حرف یونانی Σ (سیگما) استفاده میشود. مثلا $x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^{4} x_i$. لذا $\mu = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$

۲) هرگاه داده‌های $x_1, x_2, \ldots, x_n$ تشکیل تصاعد عددی دهند، میانگین آنها از رابطهٔ زیر بدست می‌آید : $\mu = \frac{x_1 + x_n}{2}$

۳) میانگین وزن دار : هرگاه عددهای $x_1, x_2, \ldots, x_n$ دارای فراوانی‌های $f_1, f_2, \ldots, f_n$ باشند، در این صورت برای محاسبه میانگین از رابطه زیر استفاده می‌شود :

$$\mu = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N}$$ (N کل داده‌ها است)

۴) هرگاه جدول توزیع به صورت حدود طبقات باشد، در این صورت در فرمول فوق x_iها همان مراکز دسته‌ها می‌باشند.

۵) در هر جامعه آماری فقط یک میانگین وجود دارد.

۶) کلیه مقادیر متغییر در مقدار میانگین تأثیر دارند.

۷) مجموع جبری تفاضل داده‌ها از میانگین صفر است. $\sum_{i=1}^{n}(x_i - \mu) = 0$

۸) هرگاه داده‌ها را با عدد ثابت b جمع کنیم، میانگین جدید برابر میانگین قبلی بـه اضـافه b خواهـد بود. یعنی : $\mu(x_i + b) = \mu(x_i) + b$

۹) هرگاه داده‌ها را در عدد ثابت a ضرب کنیم، میانگین جدید برابر، a برابر میانگین قبلی خواهد بـود. یعنی : $\mu(a\, x_i) = a\ \mu(x_i)$

۱۰)اگر داده‌ها برابر باشند، میانگین آنها نیز همان مقدار مشترک خواهد بود.

۱۱)روش سریع برای محاسبه میانگین : برای محاسبهٔ میانگین $x_1, x_2, \dots, x_n$ میانگین داده‌هـا را y تخمین میزنیم و اختلاف y را با هر یک از داده‌ها به ترتیـب $m_1, m_2, \dots, m_n$ قـرار مـی‌دهـیم. سپس میانگین مورد نظر از فرمول زیر بدست می‌آید. $\mu = y + \frac{\sum_{i=1}^{n} m_i}{n}$

مثال. میانگین داده‌های ۵ ,۴ ,۳ ,۲ ,۱ را بدست آورید.

پرسش. اگر میانگین ۵ داده برابر ۸ باشد، آنگاه مجموع داده‌ها را محاسبه کنید.

میانه : عبارت است از مقداری که تعداد داده‌های بعد از آن برابر تعداد داده‌های قبل از آن باشد.

روش بدست آوردن میانه : ابتدا داده‌ها را بطور صعودی مرتب می‌کنیم سپس اگر :

الف) تعداد داده‌ها فرد باشد، داده‌ای که وسط قرار می‌گیرد، برابر میانه است.

ب) تعداد داده‌ها زوج باشد، نصف مجموع دو داده‌ای که در وسط قرار گرفته‌اند، برابر میانه است.

روش دوم برای بدست آوردن میانه : ابتدا داده‌ها را بطور صعودی مرتب می‌کنیم، که :

شمارهٔ ردیف میانه $= \frac{n+1}{2}$

نکاتی در مورد میانه :

۱) در هر جامعه آماری فقط یک میانه وجود دارد.

۲) میانه نسبت به اندازه داده‌ها حساسیت نشان نمی‌دهد، به عبارت دیگر میانه به بزرگـی یـا کـوچکی

داده‌ها کاری ندارد.

۳) چنانچه داده‌هایی که در ابتدا یا انتهای توزیع واقع شده‌اند، از لحاظ بزرگی و کوچکی با سایر داده‌ها بطور قابل ملاحظه‌ای تفاوت داشته باشند، از میانه بعنوان شاخص استفاده می‌کنیم.

۴) اگر فراوانی مقادیر به یک نسبت کم یا زیاد شود، مثلا نصف یا دو برابر شود، میانه تغییر نخواهد کرد.

۵) اگر داده‌ها را k برابر کنیم، میانه نیز k برابر می‌شود.

مثال. میانه داده‌های زیر را بدست آورید

الف) ۶, ۷, ۵, ۸, ۲

ب) ۶, ۷, ۵, ۸, ۲, ۱

چارک‌ها :

اگر جامعه آماری به چهار قسمت مساوی تقسیم شود، چارک‌های اول تا سوم مشخص می‌شود.

چارک اول (Q_1): میانه نیمه اول داده‌ها (میانه داده‌های قبل از میانه اصلی) را چارک اول می‌گویند. (به عبارت دیگر، مقداری است که از ۲۵ درصد داده‌ها بزرگتر و از ۷۵ درصد داده‌ها کوچکتر است)

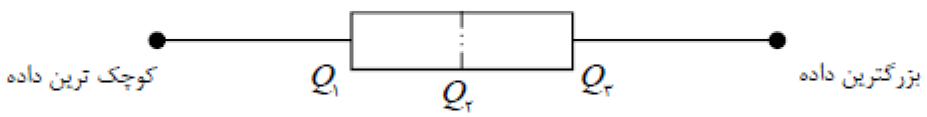
چارک دوم (Q_2): همان میانه اصلی داده‌ها می‌باشد. (به عبارت دیگر، مقداری است که از ۵۰ درصد داده‌ها بزرگتر و از ۵۰ درصد داده‌ها کوچکتر است)

چارک سوم (Q_3): میانه نیمه دوم داده‌ها (میانهٔ داده‌های بعد از میانه اصلی) را چارک سوم می‌گویند. (به عبارت دیگر، مقداری است که از ۷۵ درصد داده‌ها بزرگتر و از ۲۵ درصد داده‌ها کوچکتر است)

نمودار جعبه‌ای : نمودارهایی که تاکنون شناختیم برای مقایسه داده‌ها بسیار مفید بودند اما نمودار جعبه‌ای بهتر از بقیه نمودارها پراکندگی داده‌ها را نشان می‌دهد. و این نمودار، یک نمودار تصویری می‌باشد که داده-

ها را بر اساس پنج مقدار زیر نشان می‌دهد.

۱) کوچکترین داده ۲) چارک اول ۳) چارک دوم (میانه) ۴) چارک سوم ۵) بزرگترین داده



نکته. محل قرار گرفتن Q_2 ، بستگی به داده‌ها دارد. لذا از لحاظ موقعیت قرار گرفتن، لزومی نـدارد درسـت در وسط جعبه باشد. می‌تواند داخل جعبه مستطیلی تغییر نماید.

مثال. نمودار جعبه‌ای هر یک از داده‌های زیر را رسم کنید.

الف) ۶, ۱۲, ۵, ۲۳, ۱۵, ۱۱, ۲۵

ب) ۲۰, ۷, ۸, ۱۳, ۱۱, ۱۸, ۱۷, ۵

پرسش. تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان عبارتند از :

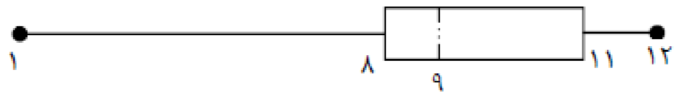
۱۲, ۱۰, ۱۵, ۲۳, ۱۴, ۲۷, ۱۶, ۳۴, ۴۱, ۴۳, ۳۲, ۱۸, ۲۵, ۳۱, ۱۹

الف) نمودار جعبه‌ای داده‌ها را رسم نمایید.

ب) چند درصد از داده‌ها در داخل جعبه قرار دارند؟

پ) چند درصد از داده‌ها در هر یک از دنباله‌ها قرار دارد؟

پرسش. با توجه به نمودار مقابل، به سوالات پاسخ دهید.



الف) میانه را بیابید.

ب) چارک اول چند است و بیان کنید که این عدد نشان دهنده چیست؟

پ) چارک سوم چند است و این عدد نشان دهنده چیست؟

ت) وجود میانه در سمت چپ داخل جعبه نشان دهنده چیست؟

پ) بلندتر بودن سمت چپ دنباله سمت چپ نسبت به دنباله سمت راست نشان دهنده چیست؟

ج) دامنهٔ تغییرات در این نمودار چقدر است؟

چ) حداکثر پراکندگی در سمت چپ جعبه چقدر است؟

پرسش. هرگاه میانگین اعداد ۵ ,a ,۳ ,۱ برابر ۲ باشد، مقدار a را بیابید.

پرسش. میانگین ۵ داده آماری برابر ۸ و میانگین ۱۰ داده آماری دیگر برابر ۵ است. میانگین کل این ۱۵ داده آماری چقدر است؟

پرسش. میانگین ۵ داده آماری ۱۲ است. اگر دو عدد ۸ و ۱۳ را از آن داده‌ها کنار بگذاریم، میانگین ۳ داده حاصل چقدر است؟

پرسش. میانگین داده‌های جدول‌های زیر را بیابید.

x_i	۱	۳	۴	۶
f_i	۵	۱	۳	۵

حدود طبقات	۱۳−۱۱	۱۶−۱۴	۱۹−۱۷
فراوانی	۴	۶	۲

پرسش. اگر میانگین داده‌های $x_۱, x_۲, \ldots, x_n$ برابر ۳ باشد، میانگین داده‌های زیر را بیابید.

$۴x_۱+ ۵, ۴x_۲+ ۵, \ldots, ۴x_n+ ۵$

پرسش. بین x, y رابطهٔ $y_i = ۷x_i - ۳$ برقرار است. اگر میانگین x_iها برابر ۷ باشد، میانگین y_iها چقدر خواهد شد؟

پرسش. میانگین ۴ درس یک دانش‌آموز هر کدام با ضریب ۱ برابر ۱۵/۵ و نمره درس پنجم وی که با ضریب ۲ می‌باشد، چه عددی باشد تا میانگین ۵ درس او ۱۶/۵ گردد؟ (انسانی ۸۴)

۱) ۱۸/۲۵   ۲) ۱۸/۵   ۳) ۱۸/۷۵   ۴) ۱۹

پرسش. در نمودار جعبه‌ای داده‌های آماری ۲۰،۱۸, ۲۳, ۱۶, ۱۴, ۱۷, ۱۱, ۹, ۱۲, ۱۰, ۵, ۱۹, ۲۱, ۷, ۸ دامنه تغییرات داده‌های داخل جعبه کدام است؟ (انسانی ۸۶)

۱) ۱۰ ۲) ۱۱ ۳) ۱۲ ۴) ۱۳

پرسش. در جدول فراوانی مقابل، میانگین به صورت $\mu = ۱۲ + ۲a$ محاسبه شده است. a کدام است؟ (ریاضی ۸۸)

x_i	۸	۱۰	۱۲	۱۴	۱۶
f_i	۲	۵	۵	۹	۳

۱) ۰/۲۵ ۲) ۰/۳۶ ۳) ۰/۴۵ ۴) ۰/۵۴

پرسش. فراوانی تجمعی در داده‌های آماری دسته بندی شده زیر داده شده است. میانگین کدام است؟ (ریاضی ۸۹)

حدود دسته	۱۰−۱۴	۱۴−۱۸	۱۸−۲۲	۲۲−۲۶	۲۶−۳۰
فراوانی تجمعی	۸	۲۰	۲۷	۳۵	۴۰

۱) ۱۹ ۲) ۲۰ ۳) ۱۹/۶ ۴) ۱۹/۵

شاخص‌های پراکندگی :

می‌دانیم که بهترین شاخص مرکزی میانگین است، ولی این شاخص همواره جوابگو نیاز ما نیست. در این صورت می‌توان جوامع آماری را از نظر پراکندگی مورد بررسی و مقایسه قرار داد (منظور از پراکندگی یعنی اینکه داده‌ها از مرکز خود چقدر دور هستند). بعنوان مثال جدول زیر را در نظر بگیرید که در آمد ماهانه (برحسب هزار تومان) سه نفر از جوامع مختلف را نشان می‌دهد.

	جامعه A	جامعه B	جامعه C
درآمد نفر اول	۳۰۰	۲۵۰	۷۰۰
درآمد شخص دوم	۳۰۰	۳۰۰	۱۵۰
درآمد شخص سوم	۳۰۰	۳۵۰	۵۰
میانگین	۳۰۰	۳۰۰	۳۰۰

همانطور که در جدول فوق دیده می‌شود، میانگین در هر سه جامعه برابر سیصد هزار تومان می‌باشد، ولی جوامع از لحاظ در آمد با همدیگر متفاوت‌اند.

در جامعه A همه افراد جامعه در آمدشان یکسان و برابر می‌باشد و اختلاف در آمد افراد از میانگین صفر است.

در جامعه B بین در آمد افراد اختلاف وجود دارد ولی اختلاف آنها نسبت به میانگین زیاد نیست، یعنی جامعه از حیث صفت مورد مطالعه نسبتا یکدست (همگن) می‌باشد.

در جامعه C بین در آمد افراد اختلاف فاحشی وجود دارد و اختلاف در آمد نسبت به میانگین زیاد است.

بنابر این فقط با دانستن میانگین نمی‌توان در مورد جامعه قضاوت کرد. لذا استفاده از شاخص‌های دیگر ضرورت دارد که چگونگی پراکندگی داده‌ها را به صورت کمی ظاهر سازند، این شاخص‌ها به شاخص‌های پراکندگی معروفند.

مهم‌ترین شاخص‌های پراکندگی عبارتند از : دامنه تغییرات، دامنه چارک‌ها، واریانس و انحراف معیار

دامنه تغییرات : اختلاف بین بزرگترین و کوچکترین مقدار داده را دامنه تغییرات می‌گویند و با R نشان می‌دهند، که $R = x_{max} - x_{min}$

نکاتی در مورد دامنه تغییرات

۱) دامنه تغییرات یک معیار سریع برای بدست آوردن پراکندگی بین داده‌هاست ولی معیار خوبی نیست. زیرا فقط بزرگترین داده و کوچکترین داده را در نظر می‌گیرد و اطلاعات داده‌هایی را که بین بزرگترین و کوچکترین داده می‌باشند را در نظر نمی‌گیرد.

۲) اگر همه داده‌ها باهم برابر باشند، دامنه تغییرات برابر صفر می‌شود و اگر دامنه تغییرات صفر باشد در این صورت همه داده‌ها باهم برابر هستند.

۳) اگر همه داده‌ها با عددی جمع یا تفریق شوند، در این صورت در دامنه تغییرات، تغییری ایجاد نمی‌شود.

۴) اگر همه داده‌ها k برابر شوند (در k ضرب یا تقسیم شوند)، در این صورت دامنه تغییرات، k برابر می‌شود. یعنی (دامنه تغییرات قبلی) × k = دامنه تغییرات جدید

۵) واحد اندازه گیری دامنه تغییرات با واحد اندازه گیری داده‌ها یکسان است.

مثال. الف) اگر دامنه تغییرات داده‌های ۲۵ ,m ,۱۳ ,۱۵ ,۱۲ برابر ۱۳ باشد، در این صورت حدود m را بیابید.

ب) اگر دامنه تغییرات داده‌های ۲۵ ,m ,۱۳ ,۱۵ ,۱۲ برابر ۱۵ باشد، در این صورت دامنه تغییرات اعداد زیر رابیابید. ۱۲ ,۳ ,۷ ,m

دامنه چارک‌ها (دامنه میان چارکی) : اختلاف بین چارک‌های اول و سوم را دامنه چارک‌ها می‌گویند و با Q نشان می‌دهند که $Q = Q_۳ - Q_۱$.

مثال. دامنه میان چارکی داده‌های ۲۱ ,۷ ,۱۲ ,۱۴ ,۷ ,۳ ,۱۱ ,۱۵ را بیابید.

نکته. اگر تک تک مقادیر را از میانگین کم کنیم، این تفاضل‌ها را انحراف از میانگین می‌نامیم. یعنی :

انحارف از میانگین $= x_۱ - \mu, x_۲ - \mu, \ldots, x_n - \mu$

واریانس :

شاخص‌های پراکندگی دامنه تغییرات و دامنه چارک‌ها قادر به بیان تمامی تغییرات نیستند پس باید دنبال شاخصی بود که تغییرات کل داده‌ها را اندازه گیری کند. طبیعی است تغییر زمانی مفهوم پیدا می‌کند که هر یک از داده‌ها نسبت به یک مرکز مقایسه شوند و بهترین مرکز برای داده‌ها میانگین است.

واریانس برابر میانگین مجذور انحرافات از میانگین است و آن را با علامت $\sigma^۲$ (بخوانید سیگما دو) نشان می‌دهند و از رابطه زیر بدست می‌آید. (σ علامت سیگمای کوچک و Σ علامت سیگمای بزرگ می‌باشد).

$$\sigma^۲ = \frac{(x_۱-\mu)^۲+(x_۲-\mu)^۲+\ldots+(x_n-\mu)^۲}{n} = \frac{\sum_{i=۱}^{n}(x_i-\mu)^۲}{n}$$

چنانچه جدول فراوانی داشته باشیم در این صورت واریانس از دستور زیر محاسبه خواهد شد :

$$\sigma^۲ = \frac{f_۱(x_۱-\mu)^۲+f_۲(x_۲-\mu)^۲+\ldots+f_n(x_n-\mu)^۲}{f_۱+f_۲+\ldots+f_n} = \frac{\sum_{i=۱}^{n}f_i(x_i-\mu)^۲}{\sum_{i=۱}^{n}f_i}$$

نکاتی در مورد واریانس :

۱) برای محاسبه واریانس دستور بسیار مهم دیگری که از آن می‌توان استفاده کرد، به صورت زیر است. $\sigma^۲ = \frac{\sum_{i=۱}^{n} x_i^۲}{n} - \mu^۲$ لازم به ذکر است که این رابطه از رابطه فوق بدست می‌آید : و چنانچه جدول فراوانی داشته باشیم در این صورت واریانس از دستور زیر محاسبه خواهد شد.

$$\sigma^۲ = \frac{\sum_{i=۱}^{n} f_i x_i^۲}{\sum_{i=۱}^{n} f_i} - \mu^۲$$

۲) اگر تمام داده‌ها باهم برابر باشند، واریانس صفر خواهد شد و بلعکس (اگر واریانس صفر باشد، تمام داده‌ها باهم برابرند.) در نتیجه میانگین و میانه نیز باهم برابر می‌شوند.

۳) اگر به داده‌ها عدد ثابت اضافه کنیم (یا کم کنیم)، واریانس آنها تغییر نمی‌کند $(\sigma^۲_{x+a} = \sigma^۲_x)$ به عبارت دیگر با افزودن مقداری به هر داده، واریانس داده‌های جدید و با واریانس داده‌های قبلی برابر خواهد شد.

۴) اگر داده‌ها را در مقدار ثابتی مانند a ضرب کنیم (یا تقسیم کنیم)، واریانس آنها در مجذور a ضرب می‌شود. $(\sigma^۲_{ax} = a^۲ \times \sigma^۲_x)$ به عبارت دیگر با ضرب کردن مقدار ثابت a در داده‌ها واریانس داده‌های جدید با $a^۲$ برابر واریانس قبلی برابر خواهد شد.

۵) واریانس از نوع مجذور واحد متغییر است. مثلا اگر واحد متغییر سانتی‌متر باشد، واحد واریانس سانتی‌متر مربع است.

مثال. برای داده‌های ۵ ,۴ ,۳ ,۲ ,۱ واریانس را محاسبه کنید.

پرسش. واریانس جدول مقابل را بیابید.

x_i	۲	۴	۶
f_i	۲	۴	۲

انحراف معیار : جذر واریانس را انحراف معیار می‌گویند و آن را با علامت σ نشان می‌دهند و از دستور های

زیر محاسبه می‌شود : $\sigma^2 = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+...+(x_n-\mu)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$ یا

$$\sigma^2 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \mu^2}$$

نکاتی در مورد انحراف معیار :

۱) اگر تمام داده‌ها باهم برابر باشند، انحراف معیار صفر است و بالعکس اگر انحراف معیار صفر باشد، تمام داده‌ها باهم برابرند.

۲) اگر همه داده‌ها با یک عدد ثابت جمع یا تفریق کنیم، انحراف معیار تغییر نمی‌کند.

۳) اگر تمام داده‌ها را در عدد ثابتی مانند a ضرب یا تقسیم کنیم، انحراف معیار آن در $|a|$ ضرب یا تقسیم می‌شود. یعنی : $\sigma_{ax+b} = |a|\sigma_x$

۴) انحراف معیار عدد ثابت صفر است. مثلا انحراف معیار عدد ۲ صفر است.

۵) واحد انحراف معیار از نوع واحد متغییر است.

مثال. انحراف معیار داده‌های ۱۶ ,$x_1, x_2, ..., x_n$ برابر صفر است. میانگین $x_1, x_2, ..., x_n$ را بیابید.

ضریب تغییرات : خارج قسمت انحراف معیار بر میانگین را ضریب تغییرات می‌نامند و با v نشان می‌دهیم.

یعنی : $v = \frac{\sigma}{\mu}$. این ضریب که به واحد اندازه گیری بستگی ندارد (یعنی واحد اندازه گیری به وسیله آن حذف می‌شود) در عمل برای مقایسه به کار می‌رود.

نکاتی در مورد ضریب :

۱) ضریب تغییرات معیاری برای میزان پراکندگی است، لذا باید مثبت باشد. چون پراکندگی منفی معنی ندارد.

۲) هرگاه واریانس دو جامعه یکسان باشد، میزان پراکندگی را بر اساس ضریب تغییرات می‌توان سنجید؟

۳) در دو یا چند جامعه آماری که برحسب یک واحد نباشند، میزان پراکندگی را بر اساس ضریب تغییرات می‌توان سنجید. مثلا برای پاسخ به این سوال که «از دو متغییر قد و وزن در بین همکلاسی هایتان کدام یک بیشتر متغییر است؟» اگر دو صفت بر حسب یک واحد می‌بودند مقایسه واریانس آنها می‌توانست جوابگو این سوال باشد ولی چون واحدها یکی نیستند مقایسه میزان پراکندگی این دو متغییر از طریق واریانس‌ها ممکن نیست. لذا باید دنبال شاخصی باشیم که بدون واحد باشد تا امکان مقایسه فراهم آید، این شاخص به ضریب تغییرات معروف است.

۴) برای ضریب تغییرات توصیف زیر نیز به کار می‌رود: ضریب تغییرات عبارت است از میزان پراکندگی به ازای یک واحد از میانگین.

۵) اگر همه داده‌ها باهم برابر باشند، ضریب تغییرات صفر است.

۶) اگر همه داده‌ها را در یک عدد ثابت (مثبت) ضرب کنیم، ضریب تغییرات تغییر نمی‌کند.

۷) اگر به همه داده‌ها یک عدد مثبت اضافه کنیم، ضریب تغییرات جدید کوچکتر از ضریب تغییرات داده‌های اولیه است. (چون اگر تمام داده‌ها با یک عدد ثابت جمع شود انحراف معیار که در صورت کسر است تغییر نمی‌کند ولی میانگین که در مخرج کسر است با آن عدد جمع و این کار مخرج را بزرگ می‌کند در نتیجه کل کسر کوچک می‌شود).

۸) ضریب تغییرات اغلب به صورت درصر بیان می‌شود.

مثال. ضریب تغییرات داده‌های آماری جدول زیر را بدست آورید.

دسته ها	۰−۲	۲−۴	۴−۶	۶−۸
فراوانی ها	۱	۲	۹	۴

پرسش. اگر واریانس داده‌های x, y, z برابر ۳ باشد، واریانس داده‌های $۲x, ۲y, ۲z$ را بیابید.

پرسش. مجموع ده داده آماری ۹۶/۴ و میانگین آنها ۳ می‌باشد. واریانس داده‌ها را بدست آورید.

پرسش. اگر واریانس قیمت‌ها در سال گذشته ۱۰۰۰ ریال و امسال ۱۰ درصد به قیمت‌ها افزوده شود، واریانس قیمت‌ها در سال جدید را حساب کنید.

پرسش. اگر انحراف از معیار داده‌های $x_1, x_2, \dots, x_n$ برابر σ باشد، در این صورت انحراف از معیار داده‌های $-2x_1-3, -2x_2-3, \dots, -2x_n-3$ را بیابید.

پرسش. انحراف از میانگین ۶ داده آماری عبارتند از : ۴- ,۲- ,۲- ,۱ ,۲ ,۵ ، انحراف از معیار آنها چقدر خواهد بود؟

پرسش. داده‌های آماری با میانگین ۸ و واریانس ۲/۲۵ موجود است. تمام داده‌ها را دو برابر می‌کنیم تا داده‌های جدیدی حاصل شود. انحراف معیار جدید را بیابید.

پرسش. میانگین نمرات ده دانش آموز ۱۶ و انحراف معیار آنها ۰/۴ است. اگر یک نمره ۵ به نمرات اضافه کنیم، واریانس نمرات جدید چقدر است؟

پرسش. میانگین حقوق پرداختی به کارکنان یک شرکت ۲ میلیون تومان و انحراف معیار ۵۰۰ هزار تومان می‌باشد. اگر ۱۰٪ میانگین به حقوق کارکنان شرکت اضافه شود، میانگین و انحراف معیار چه تغییری می‌کند؟

پرسش. اگر $\sigma^2_x = 9$ باشد، در این صورت انحراف معیار $y_i = 4x_i - 3$ چقدر است؟

پرسش. در یک جامعه آماری با میانگین ۱۵۰ و واریانس ۴۹، از کلیه داده‌ها ۱۰۰ واحد کم کنیم، ضریب تغییرات چقدر خواهد شد؟

پرسش. در جامعه‌ای پس از محاسبات لازم، ضریب تغییرات ٪۷/۵ بدست آمده است. اگر متغییرها را در عدد ثابتی مانند ۵ ضرب کنیم، ضریب تغییرات چقدر خواهد شد؟

پرسش. ضریب تغییرات تعدادی داده آماری، اگر از همه داده‌ها یک واحد کم کنیم ۱/۲ برابر وقتی است که به همه داده‌ها یک واحد اضافه می‌کنیم. میانگین این داده‌ها چیست؟

پرسش. در ۱۰۰ داده آماری با میانگین ۱۸ و انحراف معیار ۲، تمام داده‌ها را در ۱/۵ ضرب می‌کنیم. واریانس داده‌های جدید کدام است؟ (انسانی ۸۷)

۱) ۴ ۲) ۶ ۳) ۹ ۴) ۶/۲۵

پرسش. دو نفر در یک آزمایشگاه در ۵ روز متوالی هم زمان شروع به کار کردند. امتیازات دقت کاری آنان مطابق جدول زیر است. دقت کاری کدام یک بیشتر است؟ (تجربی ۸۷)

۷	۹	۸	۹	۷	نفر اول
۹	۷	۶	۸	۱۰	نفر دوم

۱) نفر اول ۲) نفر دوم ۳) یکسان ۴) نیاز به اطلاعات بیشتر

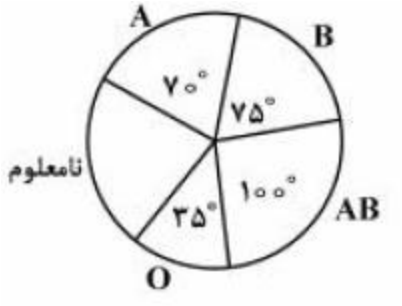
پرسش. اگر داده‌های ۱۱, ۱۵, ۱۷, ۱۶, ۱۴, ۹, ۱۱, ۱۲, ۱۵, ۱۸, ۱۴ را با نمودار جعبه‌ای نشان دهیم، انحراف معیار داده‌های داخل جعبه کدام است؟ (تجربی ۸۸)

۱) ۱/۱ ۲) ۱/۲ ۳) ۱/۲۵ ۴) ۱/۳

پرسش. ۱۵ داده آماری با واریانس ۱۲ و ده داده آماری دیگر با واریانس ۷/۶ را باهم ترکیب می‌کنیم. اگر میانگین هر دو گروه یکسان باشد، انحراف معیار ۲۵ داده حاصل کدام است؟ (ریاضی ۸۹)

۱) ۳/۱ ۲) ۳/۵ ۳) ۳/۲۵ ۴) ۳/۲

پرسش. نمودار دایره‌ای زیر متناسب با تعداد کارکنان سازمانی با گروه خونی متمایز است. گروه خونی ۳۲ نفر از آنان تعیین نشده است. چند نفر از آنها دارای نوع خون B هستند؟ (تجربی ۹۵)



۱) ۲۵ ۲) ۳۰ ۳) ۳۶ ۴) ۴۰

پرسش. داده‌های ۵ ,۴ ,۳ ,۲ ,۱ = x_i مفروض است. ضریب تغییرات داده‌های $u_i = ۱۲x_i + ۶$ کدام است؟ (ریاضی ۹۵)

۱) ۰/۴ ۲) ۰/۴۸ ۳) ۰/۵۲ ۴) ۰/۶

پرسش. در جدول فراوانی تجمعی زیر میانگین داده‌ها کدام است؟ (تجربی ۹۲)

مرکز دسته	۷	۸	۹	۱۰	۱۱
فراوانی تجمعی	۸	۲۴	۴۴	۶۸	۸۰

۱) ۹/۲ ۲) ۹/۳
۳) ۹/۴ ۴) ۹/۵

پرسش. در ۱۵۰ داده آماری با میانگین ۱۲، به دو برابر هر یک از داده‌ها ۳ واحد اضافه می‌کنیم تا داده‌های جدید حاصل شود. ضریب تغییرات داده‌های جدید چند برابر ضریب تغییرات داده‌های قبلی است؟ (تجربی ۹۱)

۱) $\frac{۷}{۹}$ ۲) $\frac{۵}{۶}$ ۳) $\frac{۷}{۸}$ ۴) $\frac{۸}{۹}$

پرسش. در ۳۰ داده آماری، مجموع تمام داده‌ها برابر ۲۴۰ و مجموع مربعات این داده‌ها ۲۱۹۰ می‌باشد. ضریب تغییرات کدام است؟ (تجربی ۹۵ خارج)

۱) ۰/۲۲۵ ۲) ۰/۲۷۵ ۳) ۰/۳۲۵ ۴) ۰/۳۷۵