زَكاةُ العِلمِ بَذلُهُ لِمُستَحِقِّهِ وَإجهادُ النَّفسِ فِی العَمَلِ بِهِ؛

امام علی (ع)

شرح درس

آمار مقدماتی

دکتر مسعود عالمی نیسی

به کوشش: مصطفی امینی

دانشکده علوم اجتماعی و ارتباطات دانشگاه علامه طباطبائی

آمار چیست؟

واژه آمار (Statistics) از کلمه لاتین Status سرچشمه گرفته است که به معنای حالت، وضع یا موقعیت می باشد. از این واژه به عنوان ریشه واژه های Stato (دولت)، Statista (دولت شناسی یا کسی که اطلاعات راجع به دولت دارد)، Statistica (آمار)، که مجموعه معین راجع به دولت می باشد، به وجود آمده است.

علم آمار همانند هر علم دیگر، در نتیجه نیازهای بشر بوجود آمده است و تاریخی غنی دارد بطوریکه از دورانهای گذشته تا کنون رشد و تکامل آن ادامه یافته است. سرشماری های بسیار ابتدایی که به هیچ رو با آمار دموگرافی و سرشماریهای امروزی قیاس شدنی نیست، بنای آمار کنونی را پی ریزی کرده و آغاز نموده است. با ظهور سرمایه داری و گسترش تجارت، آمار در مقابل مسائل مرکب تر و پیچیده تری قرار می گیرد و حجم اطلاعات جمع آوری شده افزایش می یابد و در نتیجه کارهای آماری نیز توسعه می یابد بطوریکه از نظر ماهیت، عمیق تر، از نظر موضوع مورد مطالعه وسیع تر و از نظر وسائلی که به کار گرفته می شود کامل تر میگردد.

در تحقیق های علمی، بیش از همه، این فکر که آمار در قرن هفدهم به خود شکل یک علم می گیرد، طرفدار پیدا کرده است. در اواسط قرن هفدهم در انگلستان یک جریان علمی پدید آمد که نام" حساب سیاسی" به خود گرفت. این جریان علمی را ویلیام پتی و جان گرانت آغاز کردند و بعد از آنها بنام کتب (حسابدانهای سیاسی) نامیده شد. این دانشمندان در بررسی های خود از مشخص کننده های آمار، همچون کمّیت های نسبی و متوسط استفاده می کردند. همزمان با ظهور این مکتب، در آلمان، مکتب (آمار توصیفی) یا (دولت شناسی) توسعه یافت.

ظهور این علم به سال های ۱۶۶۰ مربوط می گردد. دانشمندان این مکتب سعی وافر داشتند که به طور همه جانبه ای با استفاده از اعداد، دولتها و کشورها را تشریح و تفسیر کنند. بین دانشمندان دولت شناس، بیش از همه " آخن وال" استاد دروس حقوق بین الملل و آمار در دانشگاه گوتینگن جلب نظر می کند. بعضی از آمار دانان، آخن وال، را پدر آمار می دانند، البته از بنیانگذاران علم آمار، قبل از دیگران می توان از " کتله" نام برد.

از واژه آمار ۳ معنی می توان برداشت کرد :

الف) اطلاعات عددی:

مجموعه اعدادی که به روش خاصی از جامعه تحت مطالعه، جمع آوری وبه صورت جدول ونمودار باشاخه های عددی ارائه می شود.

ب) تئوری اعداد:

منظور اصول وقواعدریاضی و احتمالی برای ساختن فرمولها ومحاسبه پارامترهاست.

ج) روشهای آماری:

روشهایی که در جمع آوری،تنظیم وتجزیه وتحلیل وتفسیر اطلاعات عددی مورد استفاده قرار می گیرد.

تعریف آمار:

آمار، تعریف واحد و روشنی ندارد زیرا هر شاخه ای از علوم، آنرا وابسته به خود می داند اما اکثر آمارشناسان، عبارت زیر را در تعریف آمار باز گو می کنند : **((آمار علمی است که پیرامون جمع آوری وتنظیم وتحلیل وتفسیر اطلاعات عددی سخن می گوید.))**

آمار، امروزه یک تکنولوژی(فناوری) از روشهای علمی است زیرا ابزار وتکنیک لازم را برای محققین آماده می کند.

آمار، یا توصیفی است یا استنباطی ، که در واقع، آمار توصیفی، تحلیل و توصیف نمونه و نتایج حاصل از آن است وآمار استنباطی، تعمیم نتایج این نمونه به کل جامعه تحت مطالعه.

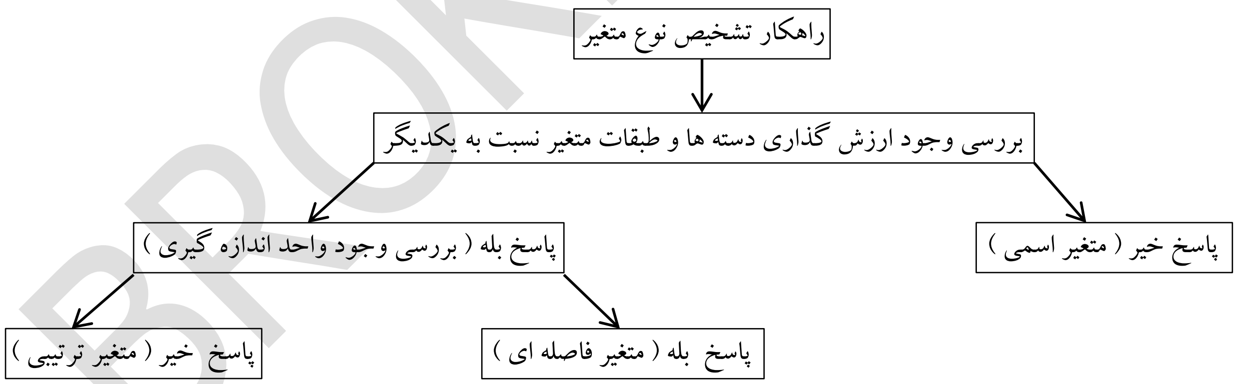
حل مسئله در آمار ، دارای مراحلی است که عبارت است از :

- ۱) تشخیص ابزار آماری (آماره مناسب)
- ۲) محاسبه آماره
- ۳) تفسیر

در آمار ، متغیرهای اندازه گیری را بر اساس دقّت سنجش ، به سه دسته تقسیم بندی می کنند که عبارت است از :

- ۱) متغیرهای اسمی : با قابلیت دسته بندی صرف
  مثال: قومیت (لُر ، ترک ، عرب ، کر)
- ۲) متغیرهای ترتیبی: با قابلیت دسته بندی و ارزش گذاری
  مثال : سنجش اعتماد به نفس ، سنجش معرفت (خیلی زیاد – زیاد – متوسط – کم – خیلی کم)
- ۳) متغیرهای فاصله ای: با قابلیت دسته بندی، ارزش گذاری و دارای واحد اندازه گیری
  مثال: تعداد دانشجویان دانشکده علوم اجتماعی و ارتباطات دانشگاه علامه طباطبائی (رشته ، مقطع ، سطح کیفی آموزشی ، تعداد)

نکته: برای **تشخیص نوع متغیر کاربردی** در مسائل آماری، باید دید که آیا (دسته ها و طبقات) متغیر، نسبت به هم، دارای ارزش گذاری هستند یا خیر. چنانچه پاسخ ما (خیر) بود، بعبارت دیگر، دسته ها و طبقات، نسبت به یکدیگر، ارزش گذاری نداشتند، متغیر مورد نظر، **متغیر اسمی** است و باید از آماره ی متناسب با متغیرهای اسمی استفاده کرد؛ حال چنانچه پاسخ ، بلی بود، یعنی نسبت به یکدیگر دارای ارزش گذاری بودند، باید دید که آیا دسته ها و طبقات یا مقادیر متغیّر، **دارای واحد اندازه گیری هستند یا خیر** ؛ حال، چنانچه مجددا ، پاسخ ، بلی بود، یعنی متغیرها دارای واحد اندازه گیری بودند، متغیر مورد نظر، **متغیر فاصله ای** است و چنانچه، متغیرها، واحد اندازه گیری نداشتند، متغیر مورد نظر، **متغیر ترتیبی** است.



تعریف آماره:

پردازشی است بر روی اعداد، که داده ای جدید را محصول می دهد. آماره ها ، با توجّه به نوع متغیر (اسمی ، ترتیبی ، فاصله ای) دارای اقسام گوناگونی هستند.

آنچه که دانشجو، در درس آمار مقدماتی ، فرا می گیرد، کاربرد همین آماره های مربوطه است در تحقیقات اجتماعی به منظور تحلیل مسائل.

انواع آماره در متغیرهای اسمی:

- **آ) بهرگان**: همانگونه که از نام این آماره مشخص است، تخصیص اجزاء از کل را می سنجد، بعبارت دیگر، افراد جامعه ی آماری، فقط در یکی از دسته ها، به کار می روند و سهم مخصوصی را از کل جامعه، به خود اختصاص می دهند. در بهرگان، جنس و کمیت، برای صورت و مخرج یکسان است، فی الواقع، در بهرگان، بهره ی یک جزء را از کل خودش می سنجند.

$$n۱/N + n۲/N + n۳/N + \cdots = ۱$$

- **ب) درصد(٪) :** از دیگر آماره های کاربردی در متغیرهای اسمی، درصد است که نسبت به بهرگان، آماره ی معمول تری است.

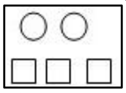
در ریاضیات، یک درصد معادل یک صدم هر چیزی است. و با علامت، «٪» مشخص می‌شود.

برای مثال، ۴۵٪ (بخوانید چهل و پنج درصد) برابر ۱۰۰/۴۵ یا ۰٫۴۵ است.

این واحد در ریاضیات (بدون بُعد) بوده و کاربرد آن در بیان سوددهی یا میزان تغییرات واحد (مخصوصا اقتصادی) می‌باشد.

چنانچه ، تعداد کل داده ها ، کمتر از (۵۰) باشد، از درصد استفاده نمی کنند و یا اینکه ، اعداد داده را در کنار درصد می نویسند، این کار برای جلوگیری از اشتباه و خطا در تفسیر آماره است. درصد نیز ، همانند بهرگان، دارای جنس و کمیت یکسان برای صورت و مخرج می باشد و سهم یک جزء ، از مجموعه ی کل همجنس خودش، برآورد می شود.

$$n۱/N \times ۱۰۰ = \%$$

- **ج) نسبت:** به رابطه بین دو کمیت نسبت می گویند . هرگاه بین دو کمیت رابطه ثابت یا عددی وجود داشته باشد می گوییم بین آنها نسبت برقرار می باشد. به مثال مقابل توجه کنید:  ؛ می بینید که نسبت دایره ها به مربع ها مثل ۲به ۳ است.

یا بطور مثال، می گوییم، برای تهیه نوعی شیرینی نسبت آرد به شکر ، مثل ۶ به ۲ است.

لازم به ذکر است که در نسبت، صورت و مخرج کسر، یکسان نیست و امکان سنجش مقادیر متفاوت، وجود دارد.

تعداد فوت به ازای هر صد هزار نفر ۱۰۰۰۰۰ × تعداد فوت / کل جمعیت

فرمول محاسبه نرخ رشد جرم = ۱۰۰ × ($X۱$ تعداد مجرمان سال اول − $X۲$ تعداد مجرمان سال دوم) / ($X۱$ تعداد مجرمان سال اول)

مثال: بررسی کنید، کدام شهر، مجرمان بیشتری دارد؟

	شهر A	شهر B
مجرمان	۱۰۱	۲۰۵
غیر مجرمان	۴۸۱	۱۰۸۱
مجموع کل	۵۸۲	۱۲۸۶

پاسخ: آماره ی مناسب برای این سوال، بهرگان است ، هر چند که درصد نیز قابل استفاده است.

بهره (سهم) مجرمان شهر A : $n۱/_N = ۱۰۱/_{۵۸۲} = ۰.۱۷۴$

بهره (سهم) مجرمان شهر B : $n۱/_N = ۲۰۵/_{۱۲۸۶} = ۰.۱۵۹$

همانطور که مشاهده می شود، شهر A ، مجرمان بیشتری دارد نسبت به شهر B .

مثال: بررسی کنید، کدام شهر، جرم خیز تر است؟ پاسخ: آماره ی مناسب، برای این سوال نیز، بهرگان است. $n۱/_N$

	شهر A	شهر B
مجرمان سابقه دار	۴۳	۱۳۷
مجرمان تک سابقه	۵۸	۶۸
کل مجرمان	۴۸۱	۱۰۸۱
جمعیت کل شهر	۵۸۲	۱۲۸۶

	بهرگان شهر A	بهرگان شهر B
مجرمان سابقه دار	۰.۰۷	۰.۱
مجرمان تک سابقه	۰.۱	۰.۰۵
بهره ی کل مجرمان	۰.۸۲	۰.۸۴

همانطور که میدانیم، برای جرم خیز بودن یک شهر، باید مجرمان سابقه دار را در نظر گرفت فلذا ، مطابق برآوردها، شهر B ، نسبت به شهر A ، جرم خیزتر است.

نکته: آماره های بهرگان، درصد و نسبت، در مورد متغیرهای ترتیبی نیز، صادق است.

فراوانی: £

در علم آمار، برای تکرار پیشامدهای حاصل از یک آزمایش، فراوانی تعریف می‌کنند. فراوانی ها، اغلب به سه دسته ی مطلق، نسبی و تجمعی، تقسیم می شوند.

فراوانی مطلق یک داده، به تعداد دفعات تکرار آن داده گفته می‌شود. فراوانی مطلق داده ی (xi) را با (**fi**) نمایش می‌دهند.

اگر داده‌ها دسته بندی شده باشند، فراوانی مطلق دسته ی (**i** اُم)، برابر تعداد اعضای این دسته خواهد بود.

فراوانی نسبی: اگر دسته ی(**i** اُم) ، دارای **فراوانی مطلق fi** ، حاصل از n داده باشد، **فراوانی نسبی** این دسته به صورت کسر**fi/n** تعریف می‌شود.

فراوانی تجمعی c£ یک دسته، به تعداد پیشامدهایی گفته می‌شود که مقدارشان از کران بالای آن دسته کمتر باشد.

فراوانی داده‌ها را معمولاً به صورت گرافیکی و در قالب‌هایی مانند هیستوگرام، یا به صورت جدول فراوانی نمایش می‌دهند.

نکته: در آمار، اقتصاد، و علوم تجربی، توزیع فراوانی (Frequency distribution) ، نوعی خلاصه‌سازی پرسودی از داده است که مشاهدات را طبقه‌بندی نموده و بسامد یا فراوانی مشاهدات را برحسب هر طبقه، بصورت درصد یا بصورت عدد، توصیف می کند.

حدود مقرر و واقعی در داده های دسته بندی شده:

حدود مقرر: در داده های دسته بندی شده، فاصله ی بین دسته ها ، بطور مثال ، { ۱۹.۹ — ۱۰ /// ۹.۹ — ۰.۰ } حدود مقرر می گویند ، یعنی محدوده و بازه ای که داده ها امکان قرار گرفتن در دسته ی مورد نظر را داشته باشند.

حدود واقعی: گاهی اوقات، علی الخصوص به منظور محاسبه ی فراوانی تجمعی، احتیاج به حدود واقعی دسته های متغیرها داریم، فلذا برای این منظور، حد مقرر انتهای دسته ی قبل را (کران بالای دسته ی قبل را) از حد مقرر ابتدای دسته ی بعدی (کران پائیین دسته ی بعدی) کم می کنیم و آنرا بر عدد ۲ ، تقسیم می کنیم، سپس نتیجه ی حاصله را به کران بالایی دسته ها ، اضافه و از کران پائینی دسته ها، کم می کنیم، به این ترتیب، حدود واقعی در داده های دسته بندی شده بدست می آید.

بطور مثال: { ۱۹.۹۵ _ ۹.۹۵ /// ۹.۹۵ _ ۰.۰۵- } = $\frac{10-9.9}{2} = \frac{0.1}{2} = 0.05$

نکته: تفاوت فراوانی تجمعی و دیگر انواع فراوانی ها، کاربرد ضروری حدود واقعی در فراوانی تجمعی است.

فراوانی تجمعی بالارونده و پائین رونده: فراوانی تجمعی ، شامل انواع بالارونده و پائین رونده می شود که برای فراوانی تجمعی بالارونده، (مقادیر کمتر از ...) حدود واقعی بالا را در نظر گرفته و با یکدیگر جمع می کنیم ؛ و برای فراوانی تجمعی پائین رونده، (مقادیر بیشتر از ...) حدود واقعی پائین را در نظر می گیریم.

درصد فراوانی تجمعی: هرگاه فراوانی تجمعی حاصله را ، بر تعداد کل داده ها تقسیم نموده و آنرا در عدد ۱۰۰ ضرب کنیم، درصد فراوانی تجمعی بدست می آید.

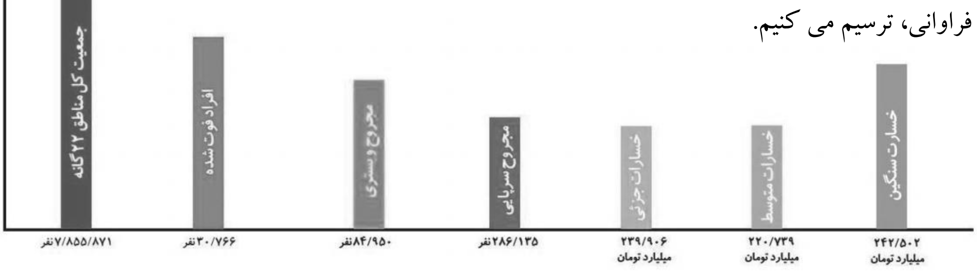
نکته ی امتحانی: هنگام مواجهه با سوالاتی از قبیل : { **چند نفر ، چند عدد ، چه تعداد ، بیشتر از فلان ، کمتر از فلان** } **، به روش فراوانی مراجعه کرده و پاسخ را از این طریق، محاسبه می کنیم.**

نمودار ها :

برای ترسیم تصویری ، اطلاعات آماری از نمودار استفاده می شود. نمودارها، اقسامی دارند که چند مورد آن ، در ذیل ، نمایش داده می شوند.

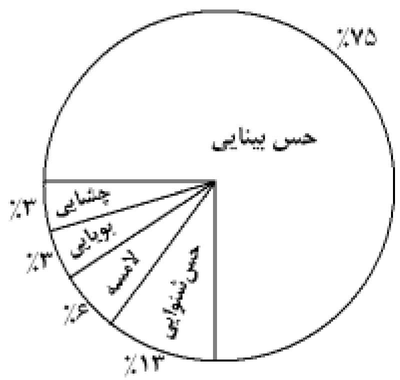
نمودار میله ای (مربوط به داده های با کمیت گسسته ، غیر مرتبط و منقطع) :

این نمودار ، برای نمایش داده های گسسته با مقیاس های اسمی، ترتیبی و طبقه ای استفاده می شودو بصورت میله هایی ترسیم می شود که هر میله، بیانگر یک رده یا یک طبقه از داده است. جهت رسم این نمودار، بر روی محور X ، طبقات متمایز را با فاصله ی مناسب ، معین می کنیم. سپس بر روی محور Y ها، فراوانی مطلق را می نویسیم و در آخر، برای هر یک از ارزش های متغیر ، میله ی متناسب با مقدار فراوانی، ترسیم می کنیم.



نمودار خسارات و تلفات احتمالی تهران در صورت وقوع زلزله ۷ ریشتری

نمودار دایره ای (مربوط به داده های با کمیت گسسته، غیر مرتبط و منقطع) :



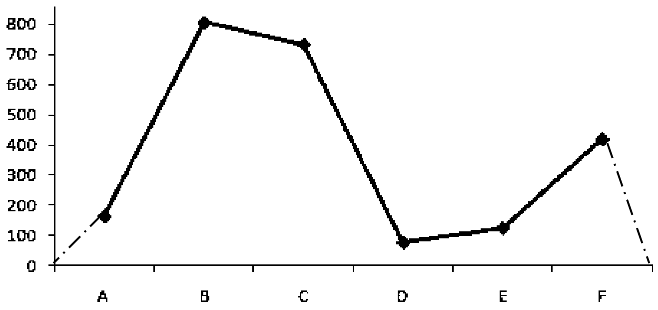
در این نمودار، مجموع فراوانی های مطلق، سطح یک دایره را اشغال می کنند. بطوریکه هر یک از فراوانی های طبقات یا متغیرهای قطاعی از سطح دایره را در بر می گیرند. این نمودار، معمولا برای نشان دادن رابطه ی اجزاء با کل داده ها بکار می رود . برای ترسیم نمودار دایره ای، ابتدا یک دایره می کشیم و سطح آنرا جهت نمایش فراوانی در نظر می گیریم و سپس با استفاده از فرمول مربوطه، سطح هر دسته را ، مشخص می کنیم .

فرمول مقدار سطوح طبقات در نمودار دایره ای :

$$Si = \frac{۳۶۰}{\varepsilon Fi} \times Fi \quad \gg \quad \text{فراوانی بر حسب درجه} = \frac{\text{فراوانی هر طبقه}}{N \text{ مجموع فراوانی ها } \varepsilon Fi} \times ۳۶۰$$

نمودار چندگوش فراوانی « چند ضلعی » (مربوط به داده ها ی با کمیّت های پیوسته ، مرتبط و متصل) :

این نمودار، روش تصویری یا دیداری دیگری است که معمولا بیشترین کاربرد را دارد.

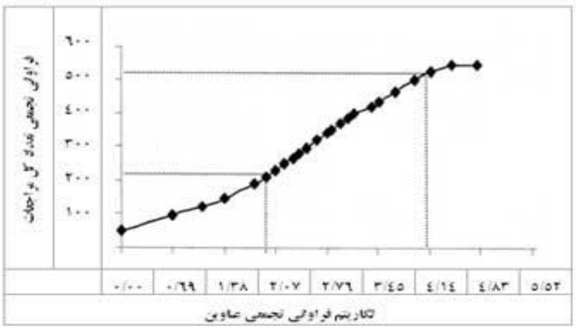


برای ترسیم این نمودار، به سه مولّفه نیاز داریم که عبارت است از :

حدود واقعی طبقات (c.l) ؛ فراوانی مطلق (Fi) ؛ نماینده طبقات (Xi)

که در محور X ها، نماینده طبقات را رسم نموده و در محور Y ها، فراوانی مطلق هر طبقه را . آنگاه، نقاط حاصله را به هم متصل نموده و نمودار مورد نظر را ترسیم می کنیم. (لازم به ذکر است که دو انتهای تمودار را با تعیین واحدهای متناسب مازاد ، به محور X ها ، متصل می کنیم.)

نمودار چندگوش فراوانی تجمعی (مربوط به داده های با کمیت های پیوسته ، مرتبط و متصل) :



نمودار شماره ۳. قانون پراکندگی «برادفورد» در ارتباط با عناوین نشریات ادواری لاتین جاری دانشکده علوم تربیتی و روانشناسی

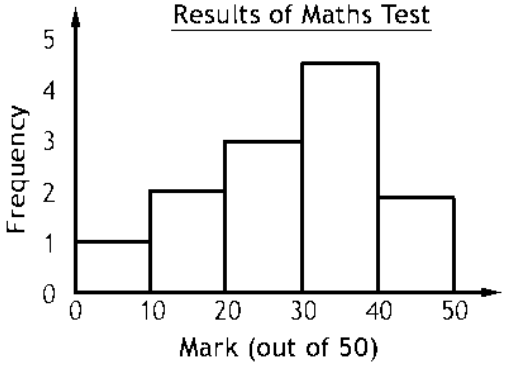
این نمودار ، در شرایطی که محقق ، علاقه مند است بداند ، چند نمره پائین تر از یک فرد یا یک طبقه وجود دارد ، مورد استفاده قرار می گیرد. بنابراین، وضعیت یک نمره یا یک نفر را ، نسبت به افراد پائین تر از خود نشان می دهد. جهت این منحنی ، بالارونده است زیرا فراوانی های تراکمی از طریق جمع متوالی فراوانی ها از پائین به بالا بدست می آید. برای رسم، بر روی محور X ها ، به فاصله ی مساوی ، حدود واقعی طبقات را می نویسیم ، فراوانی تراکمی طبقات را ، بر روی محور Y ها ، می نویسیم. سپس از حد بالای طبقه ی اول، خطی به موازات محور Y ها ، رسم می کنیم و سپس از محور Y ، در نقطه ی فراوانی طبقه ی اول ، خطی به موازات محور X می کشیم ، محل تلاقی ، نقطه ی اول را تشکیل می دهد . این کار را برای همه ی طبقات بعدی به همین صورت انجام می دهیم و در پایان ، نقطه ی اول را با خط چین ، به حد پائین طبقه اول ، متصل می کنیم. برای ترسیم این نمودار، مولفهای زیر مورد نیاز است:

درصد فراوانی تراکمی Cf% ؛ حدود واقعی طبقات R.L ؛ فراوانی تراکمی cFi ؛ فراوانی مطلق Fi ؛ حدود مقرر طبقات C.L.

نمودار هیستوگرام:

در علم آمار هیستوگرام یا بافت‌نگار یک نمودار ستونی و پله‌ای برای نشان دادن داده‌ها است.

برای نمونه بافت‌نگار فراوانی، نمودار مستطیلی با پایه‌ای به پهنای یک واحد بر روی هر مقدار مشاهده شده است که ارتفاع هر ستون آن برابر با فراوانی داده مورد نظر همخوانی دارد



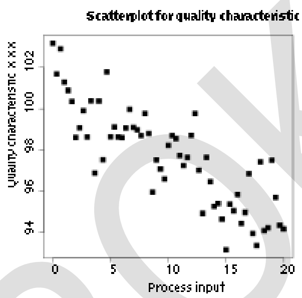
نمودار بافت‌نگار همانند نمودار ستونی است و یگانه اختلافی که بین این دو وجود دارد، نمایش ستون‌هاست. فی الواقع ، نمودار هیستوگرام، از ستون های بهم چسبیده تشکیل می شود و اتصال ستون ها در این نمودار ، باعث شده ، تا این نمودار، وسیله ی مناسبی برای نمایش داده های ناشی از اجرای متغیرهای پیوسته باشد.

در واقع ، این نمودار برای نمایش متغیرهایی که مقیاس آنها فاصله ای یا نسبی هستند بکار می رود . در هیستوگرام ، هر ستون ، نشان دهنده ی یک طبقه از اعداد است . عرض هر ستون بین حد پائین و حد بالای طبقه و ارتفاع هر ستون مساوی فراوانی هر طبقه است. **برای ترسیم** ، این نمودار، ابتدا بر روی محور X ها ، به فاصله ی مساوی ، حدود واقعی طبقات را می نویسیم و فاصله ی بین دو عدد که برابر فاصله ی طبقات است را انتخاب می کنیم. سپس بر روی محور Y ها ، فراوانی مطلق را می نویسیم آنگاه از حد پائین و حد بالای طبقه ی اول ، دو خط به موازات محور Y ، به اندازه ی فراوانی مطلق همان طبقه ترسیم می کنیم و انتهای دو خط را به هم وصل می کنیم و این کار را برای طبقات بعدی نیز ادامه می دهیم. مولفه های مورد نیاز ، برای ترسیم این نمودار ، عبارت است از:

نماینده طبقات Xi ؛ حدود واقعی طبقات R.L. ؛ فراوانی مطلق Fi ؛ حدود طبقات C.L.

نمودار پراکنش:

نموداری است که از نمایش یک نقطه به ازای هر جفت متغیر در دستگاه مختصات دکارتی به دست می‌آید. این نمودار معمولا برای نمایش نحوه پاسخ یک متغیر (متغیر پاسخ) به تغییرات متغیر دیگر (متغیر کنترل) به کار می‌رود.



مقدار یکی از متغیرها (متغیر کنترل) به عنوان مقدار محور افقی و مقدار متغیر دیگر (متغیر پاسخ) به عنوان مقدار محور عمودی در نظر گرفته می‌شود.این نمودار برای تحلیل همبستگی دو متغیر نیز به کار می‌رود.

نکته: اطلاعات بدست آمده از یک تحقیق غالبا توده‌ای از اطلاعات خام، بر معنی و بدون نظم هستند که هر نوع نتیجه گیری و تفسیر آنها غیر ممکن است. بنابراین برای هر نوع تجزیه و تحلیل اطلاعات لازم است داده‌ها (بخصوص داده‌هایی که در سطح مقیاس‌اندازه گیری فاصله‌ای و نسبی به دست آمده‌اند) براساس یک نظم منطقی طبقه بندی (Classification) شوند تا به صورت معنی‌دار و قابل تفسیر در آید.

طبقه بندی داده‌ها مستلزم محاسبه مرحله به مرحله دامنه تغییرات ، تعداد طبقات ، فاصله طبقات ، انواع فراوانی‌ها با استفاده از فرمولهای مشخص است.

در طبقه بندی داده‌ها، تمام اطلاعات در یک جدول به نام جدول توزیع فراوانی (Frequeny Table) گردآوری می‌شود و این جدول باید اساسی برای محاسبه شاخص‌های مرکزی (Centrol Index) ، شاخص‌های پراکندگی (Dispersion Index) و مقایسه گروهی از داده‌ها با گروههای دیگر جهت استنباط آماری است.

مثالی برای محاسبه انواع توزیع فراوانی:

محققی سطح هوشی یک گروه ۱۰۵ نفری از دانش آموزان را با استفاده از آزمون هوشی و کسلر برای کودکان اندازه گیری کرده است. (جدول یک) نمرات به دست آمده دارای پراکندگی بسیاری هستند و بدست آوردن اطلاعات لازم از قبیل "درصد دانش آموزانی که در سطح هوشی معینی قرار دارند" غیر ممکن است. حتی پیدا کردن بالاترین و پایین‌ترین سطح هوشی به سختی امکان‌پذیر است. بنابراین برای معنی دار و قابل تفسیر شدن داده‌ها لازم است نمرات هوشی طبقه بندی شوند. این طبقه بندی در مراحل زیر انجام می‌گیرد:

جدول یک (نمرات بهره هوشی ۱۰۵ نفر از دانش آموزان)

۹۰	۱۰۰	۱۰۰	۸۰	۱۱۲	۱۳۰	۷۰	۷۰	۸۶	۸۹	۹۰	۹۵	۱۳۰	۶۵	۶۵
۷۰	۱۰۰	۹۰	۸۰	۹۰	۸۵	۱۱۲	۱۱۵	۸۵	۷۵	۹۰	۹۵	۱۰۵	۹۵	۱۲۰
۱۰۰	۱۲۰	۱۱۰	۱۱۵	۱۰۵	۸۵	۸۰	۶۰	۸۰	۷۰	۱۰۰	۱۰۰	۱۱۰	۱۲۰	۱۱۱
۹۵	۸۵	۸۰	۱۱۰	۱۱۰	۹۵	۱۰۵	۱۰۰	۸۰	۹۹	۱۱۵	۱۲۵	۹۰	۱۰۵	۱۲۵
۹۵	۸۵	۷۵	۷۰	۷۵	۱۱۰	۱۰۰	۱۰۰	۹۰	۹۵	۱۰۵	۱۱۰	۱۰۵	۱۰۵	۸۵
۶۵	۱۰۰	۱۱۰	۹۰	۸۵	۹۰	۸۰	۱۰۰	۱۱۰	۱۱۵	۱۰۵	۱۳۰	۱۲۵	۱۰۰	۹۰
۷۰	۱۰۰	۱۱۰	۹۵	۱۰۵	۱۰۵	۹۵	۷۵	۶۵	۱۱۵	۹۰	۱۰۰	۱۰۵	۱۰۰	۶۰

محاسبه دامنه تغییرات

اگر متغیر مورد اندازه گیری (مانند هوش) را با حرف "x" نشان دهیم، دامنه تغییرات با استفاده از فرمول (تفاضل بزرگترین عدد از کوچکترین عدد به اضافه یک) بدست می‌آید.

۱ + کوچکترین عدد = xL - بزرگترین عدد = xH : دامنه تغییرات = R

فرمول شماره یک : R=Xh - xL + ۱

دامنه تغییرات سطح هوشی دانش آموزان: R = ۱۳۰ – ۶۰ + ۱ = ۷۱

محاسبه تعداد طبقات

روش تجربی

در این روش تعیین تعداد طبقات در اختیار محقق است ولی معمولا آن را بین ۱۰ تا ۲۰ طبقه انتخاب می‌کنند. چون طبقات کمتر از ۱۰ باعث بزرگتر شدن اندازه طبقات و از دست رفتن اطلاعات می‌شود و طبقات بالاتر از ۲۰ باعث طولانی شدن تهیه و تنظیم جدول می‌شود. در مثال فوق محقق تعداد طبقات به روش تجربی "۱۵۰" طبقه در نظر گرفت.

روش فرمولی

در روش فرمولی تعداد طبقات از طریق فرمولی زیر که به قانون استرژنیر معروف است بدست می‌آید.

لگاریتم بر مبنای ۱۰ = Log ؛ تعداد اعداد = n : تعداد طبقات = K

K = ۱ + (۳.۳ Log n)

محاسبه فاصله طبقات

محاسبه طبقات از تقسیم دامنه تغییرات بر تعداد طبقات از طریق فرمول زیر به دست می‌آید.

i=R/K : فاصله طبقات = i :

i = ۷۱/۱۵ = ۴.۷۳ ≈ **5**

نوشتن طبقات

معمولا نوشتن طبقات را از پایین و با عددی شروع می‌کنند که فاصله طبقات مضربی از آن باشد. در مثال فوق کوچکترین عدد ۴۰ است و فاصله طبقات "۵۰" بنابراین اولیه طبقه با ۴۰ شروع می‌شود (۶۰ مضربی از ۵ است) و به ۶۴ ختم می‌شود (بین "۶۰" تا "۶۴" پنج عدد ۶۰–۶۱–۶۲–۶۳–۶۴ قرار دارد). پس از نوشتن اولین طبقه سایر طبقات را به همان ترتیب می‌نویسند تا به آخرین طبقه برسند (جدول دوم)

جدول دو: (جدول توزیع فراوانی سطح هوشی ۱۰۰ دانش آموز)

طبقات	f	cf	Pf	P	Pcf	Xc	حدود واقعی طبقات
۱۳۰–۱۳۴	۳	۱۰۵	۰.۰۳	۳	۱۰۰	۱۳۲	۱۳۴.۵–۱۲۹.۵
۱۲۵–۱۲۹	۳	۱۰۲	۰.۰۳	۳	۹۷.۱	۱۲۷	۱۲۴.۵–۱۲۹.۵
۱۲۰–۱۲۴	۳	۹۹	۰.۰۳	۳	۹۴.۳	۱۲۲	۱۱۹.۵–۱۲۴.۵
۱۱۵–۱۱۹	۵	۹۶	۰.۰۵	۵	۹۱.۴	۱۱۷	۱۱۴.۵–۱۱۹.۵
۱۱۰–۱۱۴	۱۲	۹۱	۰.۱۱	۱۱	۸۶.۶	۱۱۲	۱۰۹.۵–۱۱۴.۵
۱۰۵–۱۰۹	۱۱	۷۹	۰.۱۰	۱۰	۷۵.۲	۱۰۷	۱۰۴.۵–۱۰۹.۵
۱۰۰–۱۰۴	۱۵	۶۸	۰.۱۴	۱۴	۶۵	۱۰۲	۹۹.۵–۱۰۴.۵
۹۵–۹۹	۱۰	۵۳	۰.۰۹	۹	۵۰.۵	۹۷	۹۴.۵–۹۹.۵
۹۰–۹۴	۱۱	۴۳	۰.۱۰	۱۰	۴۱	۹۲	۸۹.۵–۹۴.۵
۸۵–۸۹	۹	۳۲	۰.۰۸	۸	۳۰.۵	۸۷	۸۴.۵–۸۹.۵
۸۰–۸۴	۷	۲۳	۰.۰۷	۷	۲۲	۸۲	۷۹.۵–۸۴.۵
۷۵–۷۹	۴	۱۶	۰.۰۴	۴	۱۵.۲	۷۷	۷۴.۵–۷۹.۵
۷۰–۷۴	۶	۱۲	۰.۰۶	۶	۹.۵	۷۲	۶۹.۵–۷۴.۵
۶۵–۶۹	۴	۶	۰.۰۴	۴	۵.۷	۶۷	۶۴.۵–۶۹.۵
۶۰–۶۴	۲	۲	۰.۰۲	۲	۲	۶۲	۵۹.۵–۶۴.۵
	N= ۱۰۵		۱	۱۰۰			

محاسبه انواع فراوانی

فراوانی مطلق

فراوانی مطلق که حرف f نشان داده می‌شود از شمار تعداد نمرات یا اعدادی که در یک طبقه قرار می‌گیرند؛ بدست می‌آید.

برای مثال فراوانی مطلق طبقه بندی اول (۶۴–۶۰) "۲" می‌باشد. مجموع فراوانی‌ها تمام طبقات باید برابر با تعداد نمرات یا اعداد (N) باشد.

فراوانی تراکمی

فراوانی تراکمی از جمع کردن فراوانی‌ها به صورت متوالی از پایین‌ترین طبقه تا بالاترین طبقه بدست می‌آید و نشان دهنده آن است که چه تعداد از فراوانی‌ها (نمرات) در پایین نمره یا طبقه خاصی قرار دارند. در این فراوانی که cf نشان داده می‌شود فراوانی تراکمی بالاترین طبقه با مجموع نمرات (N) برابر است.

فراوانی نسبی

فراوانی نسبی که با Pf نشان داده می‌شود نشان دهنده میزان فضایی است که فراوانی یک طبقه نسبت به سایر طبقات به خود اختصاص داده است. این فراوانی از طریق فرمول PF=fi/N محاسبه می‌شود. برای مثال فراوانی نسبی طبقه ششم (۸۹–۸۵) برابر است با: ۰.۰۸ = ۱۰۵ / ۹ = PF

فراوانی نسبی درصدی

فراوانی نسبی درصدی که P نشان داده می‌شود میزان فضای اشغال شده توسط فراوانی های یک طبقه براساس مقیاس صد نشان می‌دهد و از طرف فرمول ۱۰۰ ×p=fi/N به دست می‌آید برای مثال فراوانی نسبی درصدی طبقه مهم (۱۰۴–۱۰۰) برابر است با: ۱۴ = ۱۰۰ × ۱۵/۱۰۵ =p

فراوانی تراکمی درصدی

فراوانی تراکمی درصدی که با Pcf نشان داده می‌شود نشان دهنده درصد اعداد با نمراتی است که در زیر یک طبقه معین قرار دارد این فراوانی از طریق فرمول Pcf=cf/N× ۱۰۰ بدست می‌آید.

برای مثال فراوانی تراکمی درصدی یازدهم (۱۱۴-۱۱۰) برابر است با : ۸۶.۶ = ۱۰۰ × Pcf=۹۱/۱۰۵

محاسبه نماینده طبقه

برای محاسبه نماینده طبقه (نقطه میانی هر طبقه) که با xc نشان داده می‌شود از طریق فرمول { ۲/بالای طبقه-حد پایین طبقه=Xc } بدست می‌آید. برای مثال نماینده دهم (۱۰۹-۱۰۵) برابر است با: ۱۰۷ = ۲ / (۱۰۹ + ۱۰۵) =Xc

محاسبه حدود واقعی طبقه

هر طبقه دارای یک حد واقعی است به صورت "کم کردن نیم نمره از حد پایین طبقه و اضافه نیم نمره به حد بالای طبقه" محاسبه می‌شود.

شاخص های مرکزی (نما یا مُد ، میانه ، میانگین) :

میانگین:

پرکاربردترین، کاراترین و با ثبات ترین شاخص مرکزی است و مقدار متوسط یا همان معدل حسابی داده ها ، محسوب می شود. این شاخص، مشهورترین و معتبرترین اندازه گرایش مرکزی است. علامت آن برای جامعه ی آماری برابر μ (میو) و برای نمونه ی آماری برابر $\bar{X}$ (X بار) است .

میانگین، نوعیت و ویژگی را نمایش می دهد.

لازم به ذکر است که چنانچه همه ی داده ها را از میانگین کسر کنیم و پاسخ های حاصله را با هم جمع کنیم ، پاسخ نهایی برابر صفر می شود.

$$\sum_{i=۱}^{x}(Xi - \bar{X}) = ۰$$

فرمول کاربردی میانگین ، به شرح روبرو است:

$$\bar{X} = \frac{\sum_{i=۱}^{n} Xi}{n} = \frac{\text{مجموع داده ها}}{\text{تعداد داده ها}}$$

نکته: برای بدست آوردن min در داده ها می توان از فرمول زیر استفاده کرد ، لازم به ذکر است که اگر در فرمول مذکور، بجای $\bar{X}$ ، عدد دیگری بگذاریم مثل (۱ + $\bar{X}$) یا (۱ - $\bar{X}$) ، عدد بدست آمده دیگر min نیست و حتما از آن بیشتر است.

مثال: میانگین مجموعه اعداد { ۵۷ و ٦۹ و ۸٦ و ۸۱ و ۷۲ } را حساب نموده و آنگاه، اثبات کنید حاصلجمع کسر داده ها از میانگین، برابر صفر است.

پاسخ:

۱) $$\bar{X} = \frac{\sum_{i=۱}^{n} Xi}{n} = \frac{\mathbf{72+81+86+69+57}}{\mathbf{5}} = ۷۳$$

۲) $$\sum_{i=۱}^{x}(Xi - \bar{X}) = ۰ \;=\; (۷۲ - ۷۳) + (۸۱ - ۷۳) + (۸۶ - ۷۳) + (۶۹ - ۷۳) + (۵۷ - ۷۳) = ۰$$

میانه Md :

عددی است که نیمی از داده ها، در بالای آن و نیمی دیگر در پائین آن قرار دارد و در واقع، شاخصی است که تعداد اندازه ها یا فراوانی های مقادیر را به دو نیم تقسیم می کند . میانه ارزشی است که هنگامیکه داده ها را بصورت صعودی یا نزولی مرتب کنیم ، در وسط آنها قرار می گیرد ، فی الواقع ارزشی است که تعداد داده ها را به دو گروه مساوی تقسیم می کند. میانه، نقطه ی ۵۰ درصدی داده ها را نشان می دهد و آنرا با علامت Md یا Mn نشان می دهند.

از منظر دیگر، میانه ، نوعی شاخص موقعیتی است که پس از مرتب نمودن داده ها بصورت از کمترین تا بیشترین ، نقطه ای است که ۵۰ ٪ داده ها ، کمتر از آن و ۵۰ ٪ داده ها ، بیشتر از آن هستند.

پس از مرتّب کردن ، چنانچه تعداد داده ها فرد باشد، عدد وسط ، میانه است و چنانچه تعداد داده ها زوج باشد ، میانه حاصلجمع دو عدد وسط تقسیم بر عدد ۲ است ، بعبارت دیگر، میانه، میانگین (n/۲) اُمین عدد و (۱ + n/۲) اُمین عدد است.

میانه در تعداد داده های فرد : { ۸۶ و ۸۱ و ۷۲ و ۶۹ و ۵۷ } برابر است با : $\frac{n+1}{2}$ و n برابر است با تعداد داده ها. Md = ۷۲

میانه در داده های زوج : {۸۶ و ۸۱ و۷۲ و۶۹ و۵۷ و۵۵ } برابر است با : $\left\{\left(\frac{n}{۲} + \left(\frac{n}{۲}+۱\right)\right)\right\}$ و n برابر است با تعداد داده ها. Md = ۷۰.۵

محاسبه میانگین $\overline{X}$ ، در داده های طبقه بندی شده به روش نمرات اصلی:

برای محاسبه ی میانگین توزیع فراوانی ، حد میانی هر طبقه، بعنوان نماینده ی آن طبقه به کار برده می شود بعبارت دیگر، توزیع فراوانی ، همواره بر این مفروضه بنا شده است که نمره ی همه ی افرادی که در یک طبقه قرار می گیرند ، به گونه ی یکنواخت در آن طبقه توزیع شده است. در مجموع، چنانچه حد میانی هر طبقه را در فراوانی آن طبقه ضرب کنیم ، مجموع نمرات مربوط به آن طبقه ، بدست خواهد آمد که در چنین شرایطی، میانگین، برابر است با حاصلضربهای بدست آمده بر تعداد کل اعداد.

فرمول محاسبه ی میانگین در داده های طبقه بندی شده عبارت است از : $\left\{ \overline{X} = \frac{\sum_{i=1}^{k} Fi \times mi}{N} \right\}$ ، که در آن ، N برابر است با تعداد کل داده ها و mi یا xi ، برابر است با حد میانی یا نماینده ی طبقات $\left\{ mi = \frac{(\text{ابتدای دسته} + \text{انتهای دسته})}{2} \right\}$ و Fi نیز برابر است با فراوانی هر طبقه.

مثال: میانگین را در داده های طبقه بندی شده ی زیر بدست آورید.

Fi mi	Fi	mi(xi)	حدود مقرر
۴۱۶۵۰	۱۷	۲۴۵۰	۲۰۰۰ - ۲۹۰۰
۸۹۷۰۰	۲۶	۳۴۵۰	۳۰۰۰ - ۳۹۰۰
۱۶۹۱۰۰	۳۸	۴۴۵۰	۴۰۰۰ - ۴۹۰۰
۲۷۷۹۵۰	۵۱	۵۴۵۰	۵۰۰۰ - ۵۹۰۰
۲۳۲۲۰۰	۳۶	۶۴۵۰	۶۰۰۰ - ۶۹۰۰
۱۵۶۴۵۰	۲۱	۷۴۵۰	۷۰۰۰ - ۷۹۰۰
$\sum_{i=1}^{k} Fi \times mi$ ۹۶۷۰۵۰	$\sum_{i=1}^{k} Fi$ n= ۱۸۹		

پاسخ :

$$mi = \frac{(\text{ابتدای دسته} + \text{انتهای دسته})}{2}$$

$$\overline{X} = \frac{\sum_{i=1}^{k} Fi \times mi}{N}$$

$$\overline{X} = \frac{967050}{189} = 5117$$

محاسبه ی میانه (*Md*) در داده های طبقه بندی شده:

در محاسبه ی میانه بر اساس داده های طبقه بندی شده بصورت توزیع فراوانی ، مسئله ی اصلی تعیین ارزشی از متغیر بصورتی است که نیمی از مشاهدات در بالای آن و نیمی دیگر در پائین آن واقع شوند.

فرمول اصلی میانه برابر است با:

$$\mathbf{Md} = \boldsymbol{L} + \frac{\frac{N}{2} - cF_{i-1}}{Fi} \times \boldsymbol{C}$$

که در آن ، میانه = *Md* ؛ حد واقعی پائین طبقه ی میان دار = *L* ؛ نصف افراد جدول *N/۲* ؛ فراوانی تجمعی طبقه ی قبل از طبقه ی میانه دار = cF_{i-1}

فراوانی مطلق طبقه ی میانه دار = Fi ؛ فاصله ی طبقات

همچنین، برای احتساب میانه، می توان از فرمول مقابل نیز بهره گرفت:

$$\mathbf{Md} = \boldsymbol{L} + \left(\frac{\frac{N}{2} - F}{f}\right) \times \boldsymbol{i}$$

که در این فرمول نیز ، میانه برابر *Md* ؛ حد واقعی پائین دسته ای که میانه در آن قرار دارد برابر *L* ؛ فراوانی تجمعی دسته قبل از دسته میانه برابر *F*

فراوانی دسته ای که میانه در آن قرار دارد برابر *f* ؛ تفریق حد واقعی بالا و پائین دسته میانه برابر i

نکته: در این فرمول، فرض ما بر این است، که اعداد با فاصله ی یکسانی در هر دسته قرار دارند. همچنین در این فرمول باید فراوانی بالارونده را محاسبه کرد.

برای احتساب میانه در داده های طبقه بندی شده، به مولفه های زیر ، نیازمندیم:

فراوانی تراکمی بالارونده↓ cFi	فراوانی مطلق Fi	حدود واقعی طبقات *R.L.*	حدود مقرر طبقات *C.L.*

مراحل محاسبه ی میانه در داده های طبقه بندی شده، به شرح زیر است:

۱) حدود واقعی طبقات را تعیین می کنیم.

۲) فراوانی تراکمی را بدست می آوریم.

۳) با مراجعه به ستون فراوانی تراکمی ، طبقه میانه دار را مشخص می کنیم. اگر دقیقا به *N/۲* برسیم ، یعنی حاصل عبارت *N/۲* ، در داده های فراوانی تجمعی باشد، میانه برابر با حد بالای آن طبقه خواهد بود ، در غیر اینصورت ، طبقه ای که فراوانی تراکمی آن ، از *N/۲* بزرگتر است را انتخاب می کنیم. لازم به ذکر است که چنانچه، طبقه ی میانه دار، در طبقه ای قرار بگیرد که فراوانی مطلق آن برابر صفر است ، می توان حد میانی یا نماینده ی آن طبقه را بعنوان میانه پذیرفت. (نزدیک ترین عدد فراوانی تجمعی به *N/۲* برابر است با دسته ای که میانه در آن قرار دارد.)

۴) بر اساس فرمول میانه، میانه ی داده ها را حساب می کنیم.

مثال: میانه را در داده های زیر حساب کنید.

حدود واقعی : *R.L.* = ۳۰۰۰ – ۲۹۰۰ = ۱۰۰ / ۲ = ۵۰

Fi mi	Fi	mi(xi)	*فراوانی تراکمی* cFi	*حدود واقعی*	حدود مقرر
۴۱۶۵۰	۱۷	۲۴۵۰	۱۷	۱۹۵۰ – ۲۹۵۰ ※	۲۰۰۰ – ۲۹۰۰
۸۹۷۰۰	۲۶	۳۴۵۰	۴۳	۲۹۵۰ – ۳۹۵۰ ※	۳۰۰۰ – ۳۹۰۰
۱۶۹۱۰۰	۳۸	۴۴۵۰	۸۱	۳۹۵۰ – ۴۹۵۰ ※	۴۰۰۰ – ۴۹۰۰
۲۷۷۹۵۰	۵۱	۵۴۵۰	۱۳۲	۴۹۵۰ – ۵۹۵۰ ※	۵۰۰۰ – ۵۹۰۰
۲۳۲۲۰۰	۳۶	۶۴۵۰	۱۶۸	۵۹۵۰ – ۶۹۵۰ ※	۶۰۰۰ – ۶۹۰۰
۱۵۶۴۵۰	۲۱	۷۴۵۰	۱۸۹	۶۹۵۰ – ۷۹۵۰ ※	۷۰۰۰ – ۷۹۰۰
$\sum_{i=1}^{k} Fi \times mi$	$\sum_{i=1}^{k} Fi$				
۹۶۷۰۵۰	n= ۱۸۹				

پاسخ:

فراوانی دسته میانه دار *fi* = ۵۱ // *N*= ۱۸۹

اُمین عدد *Md* = ۱۸۹/۲ = ۹۴.۵

حد پائین واقعی دسته میانه دار = *L* = ۴۹۵۰

تفریق حد واقعی بالا و پائین دسته میانه دار :

i = ۵۹۵۰ – ۴۹۵۰ = ۱۰۰۰

$$\mathbf{Md} = \boldsymbol{L} + \left(\frac{\frac{N}{2} - F}{f}\right) \times \boldsymbol{i} \quad = \quad Md = ٤٩٥٠ + \frac{\frac{189}{2} - 81}{59} \times 1000 = 5215$$

نما یا مد (Mo) و محاسبه ی آن داده های طبقه بندی شده و طبقه بندی نشده:

نما یا مد، ساده ترین شاخص گرایش مرکزی است و در یک مجموعه ، به مقداری گفته می شود، که بیشترین فراوانی را داشته باشد. بعبارت دیگر، نما، اندازه یا عددی است که بیش از اعداد یا اندازه های دیگر تکرار شده است . در داده هایی که دارای مقیاس اسمی است، نما تنها شاخص مرکزی است که مورد استفاده قرار می گیرد . نما شکل ظاهری یک توزیع را نشان می دهد . در مجموعه ای از داده ها، ممکن است نما نداشته باشیم و در مجموعه ی دیگری از داده ها ، ممکن است دو یا چند نما داشته باشیم. نما خیلی ناپایاست و صرفا شکل ظاهری یک توزیع را نشان می دهد.

روش های تعیین نما:

۱) در داده های خام یا طبقه بندی نشده و متغیرهای کمّی گسسته، نما عبارت است از مقداری که بیشترین فراوانی را داشته باشد، یعنی دارای تکرار بیشتر باشد.

۲) در داده های طبقه بندی شده، متغیرهای پیوسته و مقیاس های فاصله ای یا نسبتی ، نما در طبقه ای قرار دارد که دارای بیشترین فراوانی است فلذا ابتدا باید طبقه ی نمایی یا طبقه ای که نما در آن قرار دارد را معین کنیم که در این روش ، نما را می توان به دو صورت (بر حسب فراوانی « برآورد نما») و (از طریق فرمول « محاسبه ی نما ») ، محاسبه کرد.

محاسبه ی نما در داده های طبقه بندی شده از طریق برآورد نما (بر حسب فراوانی) :

در این روش، کافی است به جدول توزیع فراوانی مراجعه و طبقه ای که بیشترین فراوانی را دارد ، مشخص کنیم ؛ سپس حد میانی یا نماینده ی طبقه بعنوان نما شناخته می شود ، بعبارت بهتر، نما عبارت است از حد میانی یا نماینده ی طبقه ای که بیشترین فراوانی را داراست که البته به این مقدار، نمای تقریبی گفته می شود زیرا در این روش ، بطور تقریبی نه بطور دقیق، حد میانی یا نماینده ی طبقه ای که بیشترین فراوانی را داراست بعنوان مقدار نما در نظر گفته می شود.

مثال: توزیع فراوانی نمرات درس ریاضی، ۱۶ دانش آموز، در زیر آمده است. نما را بر حسب برآورد فراوانی محاسبه کنید.

نماینده طبقات (Xi)	فراوانی مطلق (Fi)	حدود واقعی طبقات $R.L.$	حدود مقرر طبقات ($C.L$)
۸	۳	۶.۵ – ۹.۵	۷ – ۹
۱۱	۵	۹.۵ – ۱۲.۵	۱۰ – ۱۲
۱۴	۲	۱۲.۵ – ۱۵.۵	۱۳ – ۱۵
۱۷	۴	۱۵.۵ – ۱۸.۵	۱۶ – ۱۸
۲۰	۲	۱۸.۵ – ۲۱.۵	۱۹ – ۲۱
	$N = ۱۶$		$C = ۳$

پاسخ : همانگونه که مشاهده می کنیم ، فراوانی مطلق دسته ی دوم ، بیش از همه است فلذا طبقه ی دارای نماست و در نتیجه عدد ۱۱ که نماینده این طبقه می باشد، بعنوان برآورد نما ، انتخاب می شود. لازم به ذکر است که اگر چنانچه ، دو نما در مقیاس نمرات، خیلی با هم فاصله نداشته باشند (یکی دو طبقه) ، میانگین آن دو را می توان بعنوان برآورد کلی نمای توزیع فراوانی پذیرفت.

محاسبه ی نما در داده های طبقه بندی شده بوسیله ی فرمول (محاسبه ی نما) :

فرمول محاسبه ی نما در داده های طبقه بندی شده، عبارت است از : $Mo = L + \left(\frac{d۱}{d۱+d۲}\right) \times C$ ؛ که در آن ، نما یا مُد برابر است با Mo ؛ و L برابر است با حد پائین واقعی طبقه ای که نما در آن قرار دارد ؛ و $d۱$ برابر است با تفاوت فراوانی مطلق طبقه نمادار از طبقه ما قبل ؛ و $d۲$ برابر است با تفاوت فراوانی مطلق طبقه نمادار از طبقه مابعد ، و C نیز همان فاصله ی طبقات است.

مثال: نما را در مثال قبل، از طریق فرمول ، محاسبه کنید:

$C = ۳$ ؛ $d۲ = ۵ - ۲ = ۳$ ؛ $d۱ = ۵ - ۳ = ۲$ ؛ $L = ۹.۵$

$$Mo = L + \left(\frac{d۱}{d۱ + d۲}\right) \times C = ۹.۵ + \left(\frac{۲}{۲ + ۳}\right) \times ۳ = ۹.۵ + \frac{۶}{۵} = ۹.۵ + \frac{۱}{۲} = ۱۰.۷$$

تفاوت میانه و میانگین:

- در میانگین، مقدار دقیق اعداد استفاده می شود لیکن در میانه، موقعیت نسبی اعداد مورد استفاده قرار می گیرد .

مثال: ۷۲ = میانه و ۷۳ = میانگین { ۸۶ و ۸۱ و ۷۲ و ۶۹ و ۵۷ } ***** ۷۲ = میانه و ۸۱ = میانگین { ۱۲۶ و ۸۱ و ۷۲ و ۶۹ و ۵۷ }

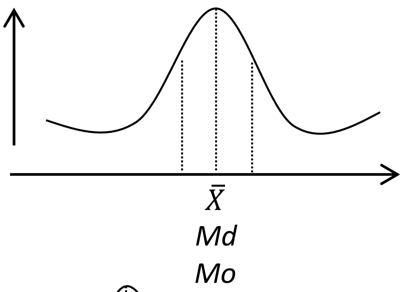
- در میانگین، اطلاعات بیشتری از اعداد استفاده می شود.
- لازم به ذکر است که دقت میانه کمتر از میانگین است
- اغلب اوقات، میانگین ، کاربردی تر است و ترجیح دارد.
- میانگین، صرفا برای سطوح سنجشی فاصله ای کاربرد دارد در حالیکه میانه برای سطوح سنجش ترتیبی نیز قابل استفاده است.

مثال برای میانه در داده های ترتیبی : زیاد ، زیاد ، زیاد ، متوسط ، متوسط ، متوسط ، کم ، کم ، کم ، کم = (n=۱۰)

چولگی یا کجی در شاخص های مرکزی: بین سه شاخص گرایش مرکزی، روابط ساده ای برقرار است که هم برای بررسی نمودار توزیع فراوانی و هم در برخی موارد برای تبدیل یک شاخص به شاخص دیگر ، می توان از آنها استفاده کرد.

اشکال توزیع فراوانی چنین است : ۱) توزیع بهنجار یا نرمال ۲) توزیع دارای کجی مثبت ۳) توزیع دارای کجی منفی

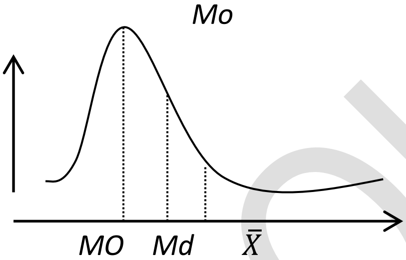
حال، رابطه ی اندازه های گرایش مرکزی را ، در این سه حالت توضیح می دهیم.



۱) توزیع بهنجار یا نرمال:

توزیع بهنجار، یک توزیع نظری نمره هایی است که متقارن است .

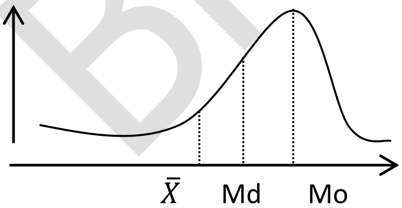
در توزیع فراوانی متقارن یا زنگوله شکل ، مقدار میانگین، میانه و نما ، برابر است و در منحنی بر روی هم منطبق هستند. رابطه ی این توزیع به شکل { $\bar{X} = Md = Mo$ } است.



۲) توزیع دارای کجی مثبت:

اگر دامنه یا دنباله ی منحنی فراوانی ، به طرف اعداد سمت راست (نمرات بالا) رفته باشد توزیع دارای کجی مثبت است . در واقع در کجی مثبت ، بیشتر نمرات ، در سمت چپ یعنی نمرات پائین انباشته شده است . در این کجی ، مقدار میانگین به ترتیب از میانه و نما بزرگتر است . رابطه ی این توزیع به شکل { $\bar{X} > Md > Mo$ } است.

تفسیر این کجی ، این است که افراد مورد مطالعه، نمره ی کمتر از میانگین گرفته اند و اگر موضوع به یک آزمون مربوط باشد ، معنی آن این است که چنین آزمونی برای بیشتر افراد دشوار بوده است .

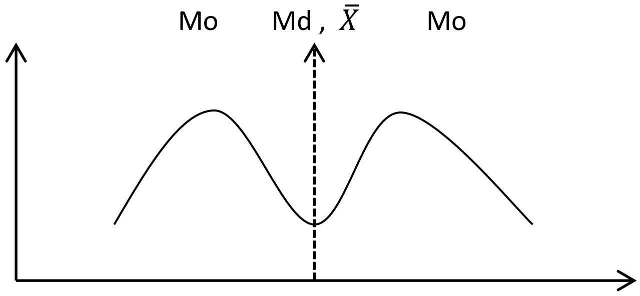


۳) توزیع دارای کجی منفی:

اگر دامنه یا دنباله ی منحنی فراوانی ف به طرف اعداد سمت چپ، یعنی نمرات پائین رفته باشد توزیع دارای کجی منفی است . در واقع در کجی منفی ، بیشتر نمرات در سمت راست ، یعنی نمرات بالا انباشته شده است . در کجی منفی، مقدار نما، به ترتیب از میانه و میانگین بزرگتر است. رابطه ی این توزیع به شکل { $\bar{X} < Md < Mo$ } است.

تفسیر این کجی ، این است که بیشتر افراد مورد مطالعه ، نمره ی بالاتر از میانگین گرفته اند و اگر موضوع به یک آزمون مربوط باشد ، معنی آن این است که چنین آزمونی برای بیشتر افراد، آسان بوده است.

نکته: میانگین به سمت اعداد بزرگ یا کوچک کشیده می شود، میانه کل داده ها را به دو قسمت مساوی تقسیم می کند و نما عددی است که بیشترین فراوانی یا بلندترین قسمت نمودار یا منحنی را نشان می دهد.



٤) توزیع دو نمایی یا متقارن:

در این توزیع، میانگین و میانه برابر هستند
و نماها در طرفین آنها قرار می گیرند.

نکته: هر یک از اندازه های گرایش مرکزی را می توان با استفاده از روابط زیر، از دوتای دیگر ، محاسبه کرد:

آ) محاسبه ی نما از میانه و میانگین : $Mo = 3Md - 2\bar{X}$

ب) محاسبه ی میانه از نما و میانگین: $Md = (Mo + 2\bar{X}) / 3$

ج) محاسبه ی میانگین از میانه و نما : $\bar{X} = (3Md - Mo) / 2$

موارد کاربرد شاخص های مرکزی:

آ) میانگین:

داده ها از نوع مقیاس فاصله ای یا نسبتی است

توزیع داده ها، بهنجار و یا نزدیک به بهنجار است.

نیاز داشتن نقطه ی تعادل یا مرکز ثقل داده ها

معتبرترین اندازه ی گرایش مرکزی را نیاز داشته باشیم

محاسبه ی اندازه های پراکندگی مانند انحراف معیار و یا محاسبه ی ضریب هنبستگی را نیاز داشته باشیم.

هنگامی که مسئله ی تحقیق ایجاب می کند، یک میانگین یا میانگین گروه های دیگر در یک متغیر مشابه ، ترکیب شوند.

برآورد میانگین جامعه آماری از طریق محاسبه ی میانگین نمونه ی آن .

ب) میانه:

داده ها دارای مقیاس رتبه ای باشند

پراکندگی داده ها دارای کجی زیاد باشد

(مناسب ترین شاخص مرکزی جهت تهیه گزارش و تفسیر در داده های دارای کشیدگی)

وقت کافی برای محاسبه ی میانگین نداشته باشیم.

توزیع داده ها ناقص باشد

بخواهیم بدانیم نتایج اندازه گیری

در نیمه ی بالا یا در نیمه ی پایین توزیع قرار می گیرد.

ج) نما:

داده ها دارای مقیاس اسمی باشند

بخواهیم یک برآورد صحیح و فوری از اندازه ی گرایش مرکزی داشته باشیم

بخواهیم بدانیم که کدام نمره ف بیشتر تکرار شده یا دارای بیشترین فراوانی است.

شاخص های پراکندگی:

همانطور که در مباحث قبلی مشاهده کردیم ، شاخص های گرایش مرکزی ، فقط مرکز یا عدد متوسط هر دسته از داده ها را معرفی می کنند و تغییرات یا پراکندگی موجود میان داده ها را پوشش نمی دهند. حال تصور کنید که در دو توزیع ، میانگین های برابری داشته باشیم که پراکندگی داده ها در اطراف میانگین ، دارای اختلاف باشد و یا اینکه بخواهیم در مجموعه ای ، میزان پراکندگی داده ها را بسنجیم، بدیهی است که برای اینکار، نیاز به شاخص های دیگری داریم که به شاخص های پراکندگی معروف اند و میزان پراکندگی یا تغییرات داده های یک توزیع را نشان می دهند.

انواع اندازه های پراکندگی:

۱) **دامنه ی تغییر:** این شاخص، ساده ترین اندازه ی پراکندگی است و عبارت است از اختلاف بین بزرگترین و کوچکترین داده های آماری و بسته به نوع کمیت مورد محاسبه (گسسته یا پیوسته) ، فرمول تخصصی خود را داراست . علامت شاخص دامنه ی تغییر ، برابر با (R) است.

آ) محاسبه دامنه ی تغییر در داده های خام :

A: کمیت های گسسته : $\boxed{R = XH - XL}$ یا $\boxed{R = Max - Min}$ ؛؛ $\boxed{\text{۲۹ = ۵۹ − ۸۶ و (۸۶ و ۵۹)}}$ ؛؛ $\boxed{\text{۸۶و۸۱و۶۹و۸۶و۵۷}}$

B : کمیت های پیوسته (قد ، وزن ، درجه حرارت و ...) : $\boxed{R = (XH - XL) + ۱}$ یا $\boxed{R = (Max - Min) + ۱}$

ب) محاسبه ی دامنه ی تغییرات در داده های دسته بندی شده:

به منظور محاسبه ی دامنه در داده های طبقه بندی شده، ابتدا نقطه ی میانی هر طبقه را مشخص می کنیم (m یا Xi) و آنگاه بیشترین مقدار نقطه ی میانی را از کمترین مقدار آن ، کم می کنیم:

$\boxed{R = XH\ m - XL\ m}$ یا $\boxed{R = Max\ m - Min\ m}$ در داده های طبقه بندی شده

مثال: دامنه ی تغییر را در داده های دسته بندی شده ی زیر ، محاسبه کنید:

mi	Fi	حدود مقرر
۲۴۵۰	۱۷	۲۰۰۰ − ۲۹۰۰
۳۴۵۰	۲۶	۳۰۰۰ − ۳۹۰۰
۴۴۵۰	۳۸	۴۰۰۰ − ۴۹۰۰
۵۴۵۰	۵۱	۵۰۰۰ − ۵۹۰۰
۶۴۵۰	۳۶	۶۰۰۰ − ۶۹۰۰
۷۴۵۰	۲۱	۷۰۰۰ − ۷۹۰۰

پاسخ: $N = ۱۸۰$

$R = Max\ m - Min\ m$ یا $R = XH\ m - XL\ m$ = ۷۴۵۰ − ۲۴۵۰ = ۵۰۰۰

ویژگی های دامنه ی تغییر :

به دلیل اثر پذیری از بزرگترین و کوچکترین عدد ، یک شاخص پایدار نیست و اعتبار چندانی ندارد ؛ قیاس آن ، قیاس با نما در شاخص های مرکزی است.

در کمیت های پیوسته ، مستلزم داشتن مقیاس فاصله ای است ؛ هر چه دامنه ی تغییر به صفر نزدیکتر باشد بیانگر پراکندگی کمتر بین داده هاست.

۲) انحراف چارکی و نقاط درصدی:

دومین اندازه ی پراکندگی، انحراف چارکی است که به آن ، نصف دامنه ی چارک ها یا چارک متوسط هم می گویند . این شاخص ، پراکندگی داده ها را در اطراف میانه توزیع نشان می دهد و از دامنه ی تغییر با ثبات تر است ، قبل از پرداختن به مبحث انحراف چارکی ، توضیح پیرامون نقاط درصدی و چارکها لازم است.

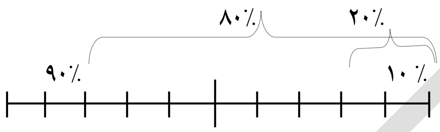
آ) نقاط درصدی (Px) :

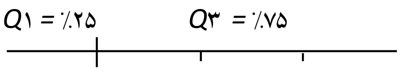
در مباحث قبلی، شیوه ی محاسبه ی میانه بعنوان یک شاخص مرکزی را آموختیم، گفته شد که میانه ، نقطه ی ۵۰ درصدی داده های مرتب شده را نشان می دهد . با استفاده از شیوه ی محاسبه ی میانه ، می توانیم نقاط درصدی از جمله، نقاط درصدی معروف (۲۵٪ یا چارک اول) و (۷۵٪ یا چارک سوم) و یا هر نقطه ی درصدی مورد نیاز را محاسبه کنیم. فی الواقع، نقاط درصدی ، همانند میانه، اندازه های ترتیبی هستند؛ نقطه ی درصدی ، در واقع ، نقطه ای است که نشان می دهد ، چند درصد اندازه ها در زیر آن نمره قرار دارند .

نقاط مهم درصدی ، عبارت اند از :

چارک اول یا $P۲۵$ یا $Q۱$	**چارک دوم یا $P۵۰$ یا $Q۲$**	**چارک سوم یا $P۷۵$ یا $Q۳$**

فرمول کلی برای محاسبه ی نقاط درصدی عبارت است از : $\boxed{Px \text{ یا } n = (P / ۱۰۰) \times N + (۱/۲)}$ که در آن ، تعداد کل اعداد برابر است با : (N) ؛ همچنین برای احتساب دهک ها، با توجّه به دهک مورد نظر ، می توانیم اعداد ۱۰ یا ۲۰ یا ۳۰ یا ٤۰ یا ۵۰ یا ... یا ۱۰۰ را جایگزین (P) کنیم و یا به منظور احتساب چارک های اول و دوم و سوم ، اعداد ۲۵ و ۵۰ و ۷۵ را جایگزین آن کنیم و یا برای احتساب نقاط صدکی ، اعداد ۱ و ۲ و ۳ و ٤ و ۵ و .... و ۱۰۰ را جایگزین مقدار P کنیم.





ب) محاسبه ی چارکها و نقاط درصدی در داده های طبقه بندی نشده:

برای محاسبه ی $Q۱$ و $Q۳$ در داده های طبقه بندی نشده ، ابتدا آنها را بطور صعودی مرتب می کنیم ، پس از مرتب کردن ، میانه (Md یا $Q۲$) را تعیین می کنیم در این صورت، میانه ، اندازه ها را به دو نیمه ی مساوی تقسیم می کند . سپس مجددا، میانه ی اعداد سمت چپ و میانه ی اعداد سمت راست را مشخص می کنیم ، آنگاه، میانه اعداد سمت چپِ میانه اصلی ، می شود (چارک اول یا $P۲۵$ یا $Q۱$) ؛ و میانه اعداد سمت راستِ میانه اصلی ، می شود (چارک سوم یا $Q۳$ یا $P۷۵$)

ج) محاسبه ی چارکها و نقاط درصدی در داده های طبقه بندی شده (دو روش) :

A : روش اول: در این روش، ابتدا، فراوانی تراکمی (تجمعی) توزیع فراوانی را می نویسیم و سپس با استفاده از رابطه ی زیر که به فرمول تعیین نقاط درصدی معروف است ، چارکها یا هر نقطه ی درصدی مورد نیاز را محاسبه می کنیم :

$$Px = L + (Pn - cFi\text{-}۱ / Fi) \times C$$

که در آن ، نقطه ی درصدی مورد نیاز برابر است با (Px) ؛ حد پائین طبقه ای که نقطه ی درصدی مورد نظر در آن قرار می گیرد برابر است با (L) ؛ درصد مورد نیاز از کل داده ها برابر است با (Pn) ؛ فراوانی تراکمی طبقه ی قبل از نقطه ی درصدی برابر است با (cFi-۱) ؛ فراوانی مطلق طبقه ی نقطه ی درصدی برابر است با (Fi) و فاصله ی طبقات برابر است با (C) .

لازم به ذکر است که نحوه ی محاسبه ی Pn به شکل روبروست : {۱۰۰ / (تعداد کل داده ها × نقطه ی درصدی مورد نیاز) } = Pn

مولفه های مورد نیاز برای برآورد نقاط ، از طریق این فرمول عبارت اند از:

فراوانی تراکمی (تجمعی) داده ها cFi	فراوانی مطلق داده ها Fi	حدود واقعی طبقات $R.L.$	حدود مقرر طبقات $C.L.$

B : روش دوم :

در این روش، هنگام مواجهه با داده های دسته بندی شده، برای محاسبه ی چارک ، دهک و صدک ، از فرمول های زیر استفاده می کنیم که دارای کلیت واحدی هستند و صرفا تغییرات مورد نظر را در صورت کسر ، باید اعمال کرد.

فرمول محاسبه ی چارک اول : $Q_1 = L + \left(\frac{\frac{N}{4} - F}{f} \right) \times i$

فرمول محاسبه ی چارک دوم (میانه) : $Q_2 = L + \left(\frac{\frac{2N}{4} - F}{f} \right) \times i$

فرمول محاسبه ی چارک سوم : $Q_3 = L + \left(\frac{\frac{3N}{4} - F}{f} \right) \times i$

فرمول محاسبه ی دهک اول: $\text{دهک اول} = L + \left(\frac{\frac{N}{10} - F}{f} \right) \times i$

فرمول محاسبه ی دهک دوم: $\text{دهک دوم} = L + \left(\frac{\frac{2N}{10} - F}{f} \right) \times i$

فرمول محاسبه ی دهک سوم: $\text{دهک سوم} = L + \left(\frac{\frac{3N}{10} - F}{f} \right) \times i$

فرمول محاسبه ی صدک اول: $\text{صدک اول} = L + \left(\frac{\frac{N}{100} - F}{f} \right) \times i$

فرمول محاسبه ی صدک پنجاه و دوم: $\text{صدک پنجاه و دوم} = L + \left(\frac{\frac{52N}{100} - F}{f} \right) \times i$

فرمول محاسبه ی صدک هشتاد و پنجم: $\text{صدک هشتاد و پنجم} = L + \left(\frac{\frac{85N}{100} - F}{f} \right) \times i$

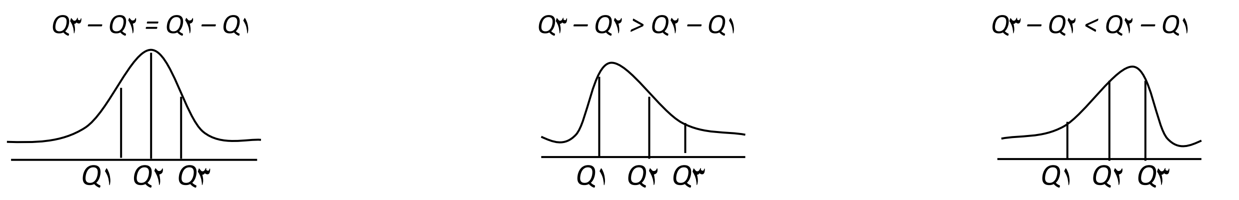
تذکر: برای احتساب نقاط دیگر نیز، به همین منوال عمل می کنیم.

دامنه ی چارک ها : فاصله ی میان چارک های اول و سوم (Q_1 و Q_3) ، را دامنه ی چارک ها می گویند بعبارت دیگر: $R.q = Q_3 - Q_1$

انحراف چارکی: هرگاه چارک سوم را منهای چارک اول نموده و حاصل را بر دو تقسیم کنیم، چارک متوسط بدست می آید: $Q = (Q_3 - Q_1) / 2$

ویژگی انحراف چارکی یا چارک متوسط:

- مقیاس این شاخص، حد اقل رتبه ای است.
- بدلیل تاثیر نپذیرفتن از اعداد بزرگ و کوچک، مناسبترین شاخص پراکندگی است هنگامیکه داده ها به سمت راست یا چپ ، کشیدگی داشته باشند.
- مقدار آن هرچه به صفر نزدیکتر شود، بیانگر پراکندگی کمتر داده ها در اطراف میانه در فاصله ی بین نقطه ی ۲۵٪ و ۷۵٪ است.
- **رابطه ی نوع کشیدگی ها و مقدار چارک ها به شرح زیر است :**



کجی بهنجار | کجی مثبت | کجی منفی

۳) « برای مطالعه ی بیشتر » : انحراف متوسط (انحراف معیار یا انحراف از میانگین) *AD* :

انحراف متوسط یا میانگین قدر مطلق، پراکندگی داده ها را حول میانگین نشان می دهد ، تحت تاثیر تک تک داده ها قرار می گیرد و در مقایسه با انحراف چارکی و دامنه ی تغییرات، از اعتبار بیشتری برخوردار است . در این شاخص، تفاوت هر نمره را نسبت به عددی معین ، همانند میانگین بدست می آورند.

ذکر این نکته خالی از فایده نیست که حاصلجمع انحراف داده ها از میانگین همان داده ها برابر صفر است: $\sum(Xi - \bar{X}) = ۰$

یکی از راههای برطرف کردن این مشکل، آن است که قدر مطلق تفاوت یا انحراف داده ها از میانگین را در نظر بگیریم ، فلذا می توان گفت که انحراف متوسط یا انحراف از میانگین ، عبارت است از مجموعه قدر مطلق انحرافات داده ها از میانگین تقسیم بر تعداد داده های آماری ، فرمول محاسباتی این شاخص ، بسته به نوع داده ها (طبقه بندی شده یا داده های خام) ، متفاوت است.

آ) **فرمول محاسبه انحراف متوسط در داده های طبقه بندی نشده:** $\mathbf{AD = \frac{\sum|Xi - \bar{X}|}{N}}$

که در این فرمول، داده ها برابر است با ($\mathbf{Xi}$) ؛ میانگین داده ها برابر است با ($\mathbf{\bar{X}}$) ؛ و تعداد کل داده ها برابر است با ($\mathbf{N}$)

مولفه های مورد نیاز برای حل مسئله بوسیله ی این فرمول ، عبارت اند از:

قدر مطلق احراف داده ها از میانگین	انحراف داده ها از میانگین	داده ها
$\lvert Xi - \bar{X}\rvert$	$Xi - \bar{X}$	Xi

مثال: انحراف متوسط مجموعه ی { ۳ و ۷ و ۴ و ۲ } را محاسبه کنید:

پاسخ:

قدر مطلق انحراف داده ها از میانگین	انحراف داده ها از میانگین	داده ها
$\lvert Xi - \bar{X}\rvert$	$Xi - \bar{X}$	Xi
۲	۲ – ۴ = – ۲	۲
۰	۴ – ۴ = ۰	۴
۳	۷ – ۴ = ۳	۷
۱	۳ – ۴ = – ۱	۳
$\sum\lvert \mathbf{Xi - \bar{X}}\rvert$ = ۶	$\sum(\mathbf{Xi - \bar{X}})$ = ۰	$\sum(\mathbf{Xi})$

$$\mathbf{\bar{X} = \frac{16}{4} = 4}$$

$$\mathbf{AD = \frac{\sum|Xi - \bar{X}|}{N} = \frac{|2-4| + |4-4| + |7-4| + |3-4|}{4} = \frac{6}{4} = \frac{3}{2} = 1.5}$$

ویژگی های انحراف متوسط یا انحراف از میانگین:

هر چه مقدار انحراف از میانگین به صفر ، نزدیک شود ، نشانه ی پراکندگی کمتر در صفت یا ویژگی مورد مطالعه در بین آزمودنی هاست.

انحراف از میانگین اعداد مساوی برابر صفر است .

انحراف از میانگین ۵ عدد متوالی برابر می شود با ۱.۲ ؛ برای مثال ، { ۱۰۰ و ۱۰۱ و ۱۰۲ و ۱۰۳ و ۱۰۴ }

ب) فرمول محاسبه ی انحراف متوسط در داده های طبقه بندی شده، برابر است با :

$$\mathbf{AD = \frac{\sum Fi\ |Xi - \bar{X}|}{N}}$$

که در آن:

فراوانی داده ها برابر است با (Fi)

نماینده طبقات برابر است با (Xi)

میانگین داده ها برابر است با ($\bar{X}$)

تعداد داده ها برابر است با (N)

مولفه های مورد نیاز جهت حل مسائلی از این قبیل ، عبارت است از:

حدود مقرر طبقات C.L	**Fi**	**Xi**	**Fi × Xi**	$\mathbf{\|Xi - \bar{X}\|}$	$\mathbf{Fi\ \|Xi - \bar{X}\|}$
C	$\sum$ **Fi**		$\sum$ **Xi × Fi**		$\mathbf{\sum Fi\ \|Xi - \bar{X}\|}$

مثال: در جدول توزیع فراوانی زیر، انحراف متوسط یا انحراف از میانگین را محاسبه کنید.

داده ها و پاسخ:

حدود مقرر طبقات C.L	**Fi**	**Xi**	**Fi × Xi**	$\mathbf{\|Xi - \bar{X}\|}$	$\mathbf{Fi\ \|Xi - \bar{X}\|}$
۳ – ۵	۴	۴	۱۶	|۴ − ۱۰.۷۲| = ۶.۷۲	۴ × ۶.۷۲ = ۲۶.۸۸
۶ – ۸	۲	۷	۱۴	|۷ − ۱۰.۷۲| = ۳.۷۲	۲ × ۳.۷۲ = ۷.۴۴
۹ – ۱۱	۹	۱۰	۹۰	|۱۰ − ۱۰.۷۲| = ۰.۷۲	۹ × ۰.۷۲ = ۶.۴۸
۱۲ – ۱۴	۴	۱۳	۵۲	|۱۳ − ۱۰.۷۲| = ۲.۲۸	۴ × ۲.۲۸ = ۹.۱۲
۱۵ – ۱۷	۶	۱۶	۹۶	|۱۶ − ۱۰.۷۲| = ۵.۲۸	۶ × ۵.۲۸ = ۳۱.۶۸
C = ۳	$\sum$ **Fi** N=۲۵		$\sum$ **Xi × Fi** ۲۶۸		$\mathbf{\sum Fi\ \|Xi - \bar{X}\|}$ ۸۱.۶

$$\bar{X} = \frac{\sum FiXi}{N} = \frac{268}{25} = \mathbf{10.72}$$

$$\mathrm{AD} = \frac{\sum \mathrm{Fi}\ |\mathrm{Xi} - \bar{\mathrm{X}}|}{\mathrm{N}} = \frac{۸۱.۶}{۲۵} = ۳.۲۶۴$$

٤) واریانس و انحراف معیار (استاندارد) و تصحیح شپرد { صفحات ۲۲ تا ۲٦ مطالب فرا درسی } :

چهارمین شاخص پراکندگی ، انحراف استاندارد یا انحراف معیار است .

این شاخص مفیدترین و باثبات ترین شاخص پراکندگی است و در مطالعات آماری و تحقیقات کاربرد زیادی دارد و آنرا می توان با میانگین که با ثبات ترین شاخص مرکزی است ، مقایسه کرد .

می دانیم مجموع انحرافات اندازه ها از میانگین همان اندازه ها صفر است ، یعنی ($\sum_{i=1}^{n}(X_i - \bar{X}) = 0$) و به همین دلیل برای محاسبه ی انحراف متوسط و شاخص پراکندگی از قدر مطلق انحرافات استفاده می کنیم.

واریانس: به دلیل طولانی بودن و وقت گیر بودن روش انحراف متوسط، بهتر است انحراف داده ها را به توان ۲ برسانیم ، سپس جمع و بر تعداد داده ها تقسیم کنیم ، ملاکی که به دست می آید برای تعیین پراکندگی بکار می رود و عددی که بدین ترتیب بدست می آید واریانس یا پراش نام دارد که اگر در مورد نمونه باشد با V و SD^2 یا تنها S^2، و اگر در مورد جامعه باشد با δ^2 ، نشان می دهند.

انحراف معیار: لازم به ذکر است که جذر یا ریشه ی دوم واریانس می شود انحراف معیار که آنرا با SD و S یا δ نشان می دهند و برابر است با :

$$(\delta = \sqrt[2]{\delta^2}).$$

آماره و پارامتر : به یاد داشته باشید که شاخص هایی را که از مطالعه ی نمونه بدست می آید ، آماره و شاخص هایی را که از مطالعه ی جامعه بدست می آید ، پارامتر می نامند .

نکته : در فرمول محاسبه ی شاخص های جامعه یا نمونه تفاوتی وجود ندارد به جز فرمول محاسبه ی واریانس که برای جامعه و نمونه متفاوت است .

در زیر فرمولهای محاسباتی واریانس برای دو گروه (جامعه و نمونه) به تفکیک نشان داده شده است .

شماره	نمونه آماری (آماره)	جامعه آماری (پارامتر)
۱	$S^2 = (\frac{\sum(X_i - \bar{X})^2}{n-1})$	$\delta^2 = \frac{(\sum X_i - \mu)^2}{N}$
۲	$S^2 = (\frac{\sum Fi(X_i - \bar{X})^2}{n-1})$	$\delta^2 = \frac{\sum Fi(X_i - \mu)^2}{N}$
۳	$S^2 = \left(\frac{\sum FiXi^2 - ((\sum F_i Xi)^2/n)}{n-1}\right)$	$\delta^2 = \frac{\sum Fi(X_i)^2}{N} - (\frac{\sum FiX_i}{N})^2$
۴	$S^2 = (\frac{n\sum FiXi^2 - (\sum F_i Xi)^2}{n(n-1)})$	$\delta^2 = \frac{\sum FiX_i^2}{N} - \mu^2$

نکته : از روابط شماره های ۱ و ۲ ، معمولا هنگامی استفاده می کنیم که میانگین داده ها عدد صحیح باشد و در غیر اینصورت بهتر است از روابط شماره های (۳)و(٤) استفاده کنیم . در ضمن اگر در فرمول های مربوط به محاسبه ی واریانس ، در نمونه آماری عدد (۱-) از مخرج کسر حذف شود ، می توان برای محاسبه ی واریانس جامعه آماری از آن استفاده کرد .

ویژگی ها و مفروضات واریانس و انحراف معیار :

۱ – مقیاس حد اقل فاصله ای است .

۲ – اگر داده ها کجی یا کشیدگی نداشته باشد ، یعنی توزیع بهنجار یا نزدیک به آن باشد ، مناسبترین شاخصِ پراکندگی است .

۳ – انحراف معیار منفی نداریم و انحراف معیار اعداد منفی هم می شود صفر .

۴ – انحراف معیار اندازه های یک جامعه ی آماری همیشه کوچکتر از انحراف معیار اندازه های نمونه از همان جامعه ی آماری است : ($\delta < S$) .

۵ – انحراف معیار ۵ عدد متوالی از یک جامعه ی آماری می شود $\sqrt{۲}$ و از یک نمونه ی آماری می شود ($\sqrt{۲.۵}$) .

۶ – اگر به هر یک از داده ها عدد ثابت (a) را اضافه کنیم ، واریانس و انحراف معیار داده های جدید تغییر نمی کند ، همچنین اگر از هر یک از داده ها ، عدد ثابت (a) را کم کنیم نیز ، واریانس و انحراف معیار داده های جدید ، تغییر نمی کند ،

بطور مثال : اگر واریانس x و y و z برابر ۴ باشد ، واریانس (x-۲) و (y-۲) و (z-۲) نیز می شود عدد ۴ .

۷ – اگر هر یک از داده ها را در عدد ثابت (a) ضرب کنیم واریانس نیز در $a^۲$ ضرب می شود ، بطور مثال اگر واریانس X برابر ۴ باشد ، واریانس ۲X نیز می شود ($\delta^۲ = ۲^۲ \times ۴ = ۱۶$) .

۸ – اگر هر یک از داده ها را در عدد ثابت (a) تقسیم کنیم واریانس نیز بر $a^۲$ تقسیم می شود .

۹ – بطور کلی در شاخص های پراکندگی ، ضرب و تقسیم داده ها در عدد ثابت (a) در مقدار واریانس و انحراف معیار اثر می گذارد ولی جمع و تفریق داده ها در عدد ثابت (a) در مقدار واریانس و انحراف معیار اثر نمی گذارد و این خاصیت و قاعده در مورد تمامی شاخص های پراکندگی صادق است . درحالیکه در شاخص های گرایش مرکزی ، چهار عمل اصلی ریاضی در مقدار شاخص تاثیر می گذارد .

۱۰ – اگر میانگین چند اندازه صفر باشد ، انحراف معیار از رابطه ی ($\delta = \sqrt{\frac{\sum X_i^۲}{N}}$) بدست می آید .

۱۱ – اگر اندازه های آماری تشکیل تصاعد حسابی بدهند می توان واریانس یا انحراف معیار را از رابطه ی $\delta^۲ = \frac{d^۲(n^۲-۱)}{۱۲}$ ، بدست آورد که در آن d قدر نسبت (فاصله بین داده ها) و n تعداد داده هاست مثل مجموعه (۱۱و۹و۷و۵و۳) ، که در آن d=۲ و n=۵ است .

۱۲ – اگر کلیه ی اندازه های آماری برابر هم باشند ، در این صورت کلیه ی شاخص های گرایش مرکزی (نما،میانه،میانگین،چارکها) برابر و مساوی یکی از اندازه هاست و کلیه ی شاخص های پراکندگی (دامنه ی تغییرات ، انحراف چارکی ، انحراف متوسط ، واریانس و انحراف معیار) برابر صفر است .

بطور مثال در مورد اندازه های a,a,a,a,a داریم
$$\left.\begin{aligned} \mu = Md = Mo = Q۱ = Q۲ = a \\ R = Q = AD = \delta = \delta^۲ = ۰ \end{aligned}\right\}$$

محاسبه ی واریانس و انحراف معیار در داده های طبقه بندی نشده :

الف) روش نمره های انحرافی :

مثال: اندازه های زیر به عنوان نمونه از یک جامعه آماری انتخاب شده است .

واریانس و انحراف معیار داده ها را محاسبه کنید . Xi = ۲,۴,۸,۱۲,۴

Xi	Xi-$\bar{X}$	(Xi-$\bar{X}$)^{۲}
۲	۲-۶=-۴	۱۶
۴	۴-۶=-۲	۴
۸	۸-۶=+۲	۴
۱۲	۱۲-۶=+۶	۳۶
۴	۴-۶=-۲	۴
∑Xi=۳۰	∑Xi-$\bar{X}$=۰	$\sum(X_i-\bar{X})^{۲}=۶۴$

پاسخ :

$$\bar{X}=\frac{۳۰}{۵}=۶ \qquad S^{۲}=\left(\frac{\sum(X_i-\bar{X})^{۲}}{n-۱}\right)=\frac{۶۴}{۵-۱}=\frac{۶۴}{۴}=۱۶\ ,\ S=۴$$

نکته : روش فوق به روش داده های انحرافی معروف است زیرا از طریق داده های انحرافی یا انحراف از میانگین محاسبه می شود .

ب) روش داده های اصلی یا خام :

مثال : اندازه های زیر به عنوان نمونه از یک جامعه ی آماری انتخاب شده است .

واریانس و انحراف معیار آنها را محاسبه کنید : Xi=۱۲,۱۴,۸,۶,۴

پاسخ:

$$S^{۲}=\left(\frac{n\sum Xi^{۲}-(\sum X_i)^{۲}}{n(n-۱)}\right)=\frac{۵\times۴۵۶-۴۴^{۲}}{۵\times(۵-۱)}=\frac{۲۲۸۰-۱۹۳۶}{۲۰}=۱۷.۲\ ,\qquad S=\sqrt{۱۷/۲}=۴.۱۴۷$$

Xi	$X_i^{۲}$
۱۲	۱۴۴
۱۴	۱۹۶
۸	۶۴
۶	۳۶
۴	۱۶
∑Xi=۴۴	∑Xi^{۲}=۴۵۶

خلاصه مراحل محاسبه در روش داده های انحرافی :

۱ – میانگین توزیع را محاسبه می کنیم .

۲ – انحراف هر یک از داده ها را از میانگین بدست می آوریم .

۳ – داده های انحراف محاسبه شده را به توان ۲ می رسانیم .

۴ – مجموع مجذورات انحراف را محاسبه می کنیم .

۵ – حاصل جمع را بر n-1 تقسیم می کنیم .

محاسبه ی واریانس و انحراف معیار در داده های طبقه بندی شده یا جدول توزیع فراوانی :

برای محاسبه این مهم در داده های طبقه بندی شده ، دو روش وجود دارد که عبارت اند از :

۱ – روش مستقیم یا از راه داده های اصلی

۲ – روش غیر مستقیم یا از راه داده های انحرافی .

مثال : در جدول توزیع فراوانی زیر ، واریانس و انحراف معیار داده ها را از دو روش مستقیم (اصلی) و غیر مستقیم (انحرافی) محاسبه کنید و نتایج را با هم مقایسه کنید .

C.L. حدود طبقات	فراوانی مطلق Fi	نماینده طبقات (Xi)	F_iX_i	$(Xi-\mu)^۲$	$Fi(Xi-\mu)^۲$	$F_iX_i^۲$
۳–۷	۴	۵	۲۰	$(۵-۱۶)^۲=۱۲۱$	$۴\times۱۲۱=۴۸۴$	$۴\times۵^۲=۱۰۰$
۸–۱۲	۳	۱۰	۳۰	$(۱۰-۱۶)^۲=۳۶$	$۳\times۳۶=۱۰۸$	$۳\times۱۰^۲=۳۰۰$
۱۳–۱۷	۱۱	۱۵	۱۶۵	$(۱۵-۱۶)^۲=۱$	$۱۱\times۱=۱۱$	$۱۱\times۵^۲=۲۴۷۵$
۱۸–۲۲	۷	۲۰	۱۴۰	$(۲۰-۱۶)^۲=۱۶$	$۷\times۱۶=۱۱۲$	$۷\times۲۰^۲=۲۸۰۰$
۲۳–۲۷	۵	۲۵	۱۲۵	$(۲۵-۱۶)^۲=۸۱$	$۵\times۸۱=۴۰۵$	$۵\times۲۵^۲=۳۱۲۵$
C=۵	N=۳۰		∑FiXi=۴۸۰		∑Fi(Xi-μ)^۲=۱۱۲۰	∑FiX$i^۲$=۸۸۰۰

۱)روش مستقیم (اصلی) :

$$\delta^۲=\frac{\sum Fi(X_i)^۲}{N}-(\frac{\sum FiX_i}{N})^۲=\frac{۸۸۰۰}{۳۰}-(\frac{۴۸۰}{۳۰})^۲=۲۹۳.۳۳-۲۵۶=۳۷.۳۳\quad,\quad \delta=\sqrt{۳۷.۳۳}=۶.۱۱$$

۲)روش غیر مستقیم (انحرافی) : هنگامی استفاده می کنیم که میانگین داده ها عدد صحیح باشد .

$$\mu=\frac{\sum FiXi}{N}=\frac{۴۸۰}{۳۰}=۱۶\ \rightarrow\ \delta^۲=\frac{\sum Fi(X_i-\mu)^۲}{N}=\frac{۱۱۲۰}{۳۰}=۳۷.۳۳\ ,\ \delta=\sqrt{۳۷.۳۳}=۶.۱۱$$

ضریب پراکندگی CV :

عاملی است که پراکندگی نسبی یک صفت را نشان می دهد و معمولا برای مقایسه بین ویژگی های مختلف یک گروه بکار می رود و از فرمول ($CV=\frac{Sx}{\bar{X}}\times۱۰۰$) بدست می آید که در آن ، CVضریب پراکندگی ، Sxانحراف استاندارد x و $\bar{X}$ میانگین مقادیر x است .

مثال : با توجه به جدول زیر ، کدام کلاس از لحاظ صفت مورد اندازه گیری از پراکندگی کمتری برخوردار است ؟

کلاس	$\bar{X}$	S	CV
A	۱۴	۲.۵	۱۷.۸۵
B	۱۶	۵	۳۱.۲۵
C	۱۲	۲	۱۶.۶۷

پاسخ : افراد کلاس C در مقایسه با دو کلاس دیگر نسبت به صفت مورد اندازه گیری همگن تر یا متجانس تر هستند .

تصحیح شپرد (تصحیح خطای انحراف استاندارد) :

بین مقدار میانگین در اعداد طبقه بندی شده و نشده اختلاف وجود دارد که این تفاوت یا اختلاف ناشی از طبقه بندی داده هاست که به آن خطای گروه بندی یا طبقه بندی می گویند .

از خطای طبقه بندی در محاسبه ی میانگین می توان چشم پوشی کرد اما در محاسبه ی انحراف معیار خطای طبقه بندی در نتایج تاثیر می گذارد ، بنابراین باید انحراف معیار محاسبه شده از جدول توزیع فراوانی تعدیل یا اصلاح شود ، لازم به ذکر است خطای طبقه بندی تابع فاصله ی طبقات است و هر چه فاصله ی طبقات بزرگتر شود ، خطا نیز بیشتر می شود و در ضمن تصحیح انحراف استاندارد وقتی مصداق پیدا می کند که تعداد طبقات از ۱۲ طبقه کمتر باشد ، انحراف استاندارد محاسبه شده بوسیله ی فرمول ($S_C = \sqrt{S^2 - \frac{C^2}{12}}$) که به تصحیح شپرد معروف است و در آن (S_C انحراف استاندارد اصلاح شده ، $S^۲$ واریانس اندازه ها ، C فاصله ی طبقات) است ، تصحیح می گردد .

مثال: در یک توزیع فراوانی طبقه بندی شده انحراف استاندارد و فاصله ی طبقات به ترتیب ۴ و ۳ شده است. انحراف استاندارد اصلاح شده چقدر می شود؟

$$S_C = \sqrt{S^۲ - \frac{C^۲}{۱۲}} = \sqrt{۱۶ - \frac{۹}{۱۲}} = ۳.۹$$

اندازه ها	میانگین	واریانس	انحراف معیار
X.Y.Z	$\overline{x}$	$\delta^۲$	$\sqrt{\delta^۲}$
(X+a),(y+a),(z+a)	$\overline{x} + a$	$\delta^۲$	$\sqrt{\delta^۲}$
(X-a),(y-a),(Z-a)	$\overline{x} - a$	$\delta^۲$	$\sqrt{\delta^۲}$
ax,ay,az	$\overline{ax}$	$a^۲\delta^۲$	$a\sqrt{\delta^۲}$
$\frac{x}{a}, \frac{y}{a}, \frac{z}{a}$	$\frac{\bar{X}}{a}$	$\frac{\delta^۲}{a^۲}$	$\frac{\sqrt{\delta^۲}}{a}$

انحراف استاندارد مرکّب :

محقق گاهی اوقات مجبور می شود انحراف استاندارد مرکب یا ادغام شده ی چند متغیر را نیز بداند .

انحراف استاندارد مرکب را با علامت S_T نشان می دهند و بوسیله ی فرمول مک نمار

$$S_T = \sqrt{\frac{\left(n_A(\bar{X}_A^۲+S_A^۲)+n_B(\bar{X}_B^۲+S_B^۲)\right)}{n_A+n_B} - \bar{X}_T^۲}$$ محاسبه می شود

که در آن (S_T انحراف استاندارد مرکب یا گروهی ، n_A , n_B , n_C , ... تعداد داده ها در هر گروه ، $\bar{X}_A$, $\bar{X}_B$, ... میانگین اندازه گروه ها، $\bar{X}_T$ میانگین مرکب) است . مثال : جدول زیر توزیع نمرات دو گروه از دانشجویان (گروه آزمایشی و گروه کنترل) در یک صفت معین را نشان می دهد . انحراف استاندارد یا معیار مرکّب را محاسبه کنید .

X_A	X_B	$X_A^۲$	$X_B^۲$
۲	۱	۴	۱
۴	۵	۱۶	۲۵
۶	۷	۳۶	۴۹
۸	۵	۶۴	۲۵
$\sum X_A = ۲۰$	$\sum X_A = ۱۸$	$\sum X_A^۲ = ۱۲۰$	$\sum X_B^۲ = ۱۰۰$

$$S_A^2 = \frac{4\times120-20^2}{4(4-1)} = \frac{80}{12} = \quad \bar{X}_A = 5 , \bar{X}_B = 4.5 , \bar{X}_T = \frac{5+4.5}{2} = 4.75$$

$$S_B^2 = \frac{4\times100-18^2}{4(4-1)} = \frac{76}{12} = 6.33 \qquad 6.67$$

$$S_T = \sqrt{\frac{\left(4(25+6.67)+4(20.25+6.33)\right)}{4+4} - (4.75)^2} = \sqrt{\frac{(126.68+106.32)}{8} - 22.56} = \sqrt{29.125 - 22.56}$$

$$= \sqrt{6.565} = 2.56$$

۵) انحراف از میانگین (انحراف از معیار) و واریانس (تدریس استاد) :

در علوم اجتماعی ، بیشتر از انحراف معیار استفاده می شود.

فرمول انحراف از میانگین ، عبارت است از: $\frac{\sum_{i=1}^{n}|Xi-\bar{X}|}{n} =$ انحراف از میانگین

فرمول انحراف معیار ، عبارت است از: $\sqrt{\frac{\sum_{i=1}^{n}(Xi-\bar{X})}{n}} = S$ انحراف از معیار

برای محاسبه ی انحراف معیار، مولفه های زیر لازم است:

$$Xi \qquad (Xi-\bar{X}) \qquad (X-\bar{X})^2$$

مثال: انحراف معیار را در داده های زیر ، محاسبه کنید.

Xi	$(Xi-\bar{X})$	$(X-\bar{X})^2$
۷۲	- ۱	۱
۸۱	۸	۶۴
۸۶	۱۳	۱۶۹
۶۹	- ۳	۱۶
۵۷	- ۱۶	۲۵۶

پاسخ: $\bar{X} = 73$

$S = \sqrt{\frac{506}{5}} = 10.06$

مثال: انحراف معیار را در داده های زیر ، محاسبه کنید.

Xi	$(Xi-\bar{X})$	$(X-\bar{X})^2$
۱۰	- ۲۲	۱
۱۷	- ۱۵	۶۴
۲۰	- ۱۲	۱۶۹
۵۱	۱۹	۱۶
۶۲	۳۰	۲۵۶

پاسخ: $\bar{X} = 32$

$S = \sqrt{\frac{2114}{5}} = 20.56$

واریانس:

نکته: لازم به ذکر است که واریانس و انحراف معیار ، در کاربرد، تفاوت چندانی با هم ندارند لیکن ، بصورت قرار دادی ، از انحراف معیار استفاده میکنیم.

نکته: کاربرد واریانس و انحراف معیار، برای درک و فهم میزان تنوع است فلذا در سوالاتی که تنوع داده ها را جست و جو می کرد، باید از این شاخص های پراکندگی استفاده کرد.

فرمول واریانس ، عبارت است از : $$S^2 = \frac{\sum(Xi - \bar{X})^2}{N}$$

نکته : با توجه به فرمول واریانس، می توان ، فرمول دیگری را نیز برای محاسبه ی انحراف معیار در نظر گرفت.

فرمول محاسبه ی انحراف معیار در داده های طبقه بندی نشده (خام) :

$$S = \sqrt{\frac{\sum X^۲}{N} - \left(\frac{\sum Xi}{N}\right)^۲}$$

فرمول محاسبه ی انحراف معیار در داده های طبقه بندی شده :

$$S = \sqrt{\frac{\sum Fi(m_i^۲) - \frac{(\sum Fi\, mi)^۲}{N}}{N}}$$

نکته : مولفه های مورد نیاز برای محاسبه ی انحراف معیار در داده های طبقه بندی شده ، عبارت است از:

حدود مقرر	Fi	Mi	Mi۲	Fi Mi	Fi Mi۲

مثال: انحراف معیار را در داده های طبقه بندی شده ی زیر حساب کنید.

حدود مقرر	Fi	Mi	$Mi^۲$	$Fi\ Mi$	$Fi\ Mi^۲$
۵۰ – ۵۴	۱	۵۲	۲۷۰۴	۵۲	۳۷۰۴
۵۵ – ۵۹	۲	۵۷	۳۱۴۹	۱۱۴	۶۴۹۸
۶۰ – ۶۴	۶	۶۲	۳۸۴۴	۳۷۲	۲۳۰۶۴
۶۵ – ۶۹	۳	۶۷	۴۴۸۹	۲۰۱	۱۳۴۶۷
۷۰ – ۷۴	۱۲	۷۲	۵۱۸۴	۸۶۴	۶۲۲۰۸
۷۵ – ۷۹	۱۴	۷۷	۵۹۲۹	۱۰۷۸	۸۳۰۰۰
۸۰ – ۸۴	۸	۸۲	۶۷۲۴	۶۵۶	۵۳۷۹۳
۸۵ – ۸۹	۴	۸۷	۷۵۶۹	۳۴۸	۳۰۲۷۲
۹۰ – ۹۴	۷	۹۲	۸۴۶۴	۶۴۴	۵۹۲۴۸
۹۵ – ۹۹	۵	۹۷	۹۴۰۹	۴۸۵	۴۷۰۴۵
c=۴	۶۲			۴۸۱۴	۳۸۱۳۰۸

پاسخ:

$$\mathbf{S} = \sqrt{\frac{\sum \mathrm{Fi}(\mathrm{m}_\mathrm{i}^2) - \frac{(\sum \mathrm{Fi\,mi})^2}{\mathrm{N}}}{\mathrm{N}}}$$

$$\mathbf{S} = \sqrt{\frac{381308 - \frac{(4814)^2}{62}}{62}} = \sqrt{۱۲۱.۲۶} = ۱۱.۰۲$$

٦) شاخص پراکندگی برای متغیرهای اسمی و ترتیبی:

شاخص تغییرات کیفی : **IQV (INDEX OF QUALITIVE VALIATION)**

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100}$$

- **این شاخص ، معمولا هنگام مقایسه ی داده های اسمی و ترتیبی ، کاربرد دارد و چگونگی تنوع داده ها را نشان می دهد ، بعبارت بهتر، هر چه *IQV* ، بیشتر باشد ، پراکندگی داده ها کمتر و تنوع بیشتر است. معمولا (*IQV*) بین (۱۰۰ – ۰) است و چنانچه همه ی داده ها در یک دسته باشند ، *OV* صفر می شود و در نتیجه *IQV* نیز صفر می شود.**

مثال: داده های زیر ، مربوط به شرکت کنندگان در دو سمینار مختلف است، بررسی کنید تنوع دانشجویان شرکت کننده در کدام یک از سمینارها بیشتر است؟

رشته ها	سمینار (الف)		سمینار (ب)	
	تعداد شرکت کننده (الف)	مورد انتظار (الف)	تعداد شرکت کننده (ب)	مورد انتظار (ب)
خدمات اجتماعی	۲	۵	۸	۷
جامعه شناسی	۱۷	۵	۶	۷
روان شناسی	۱	۵	۹	۷
مردم شناسی	۰	۵	۵	۷
	N = ۲۰	۲۰ / ۴ = ۵	*N* = ۲۸	۲۸ / ۴ = ۷

نکته: در قسمت مورد انتظارها ، انتظار ما این است که در هر سمینار، به تعداد مساوی ، شرکت کننده از رشته های مختلف داشته باشیم

سمینار (الف)

$$QV = ۲(۱۷+۱+۰)+۱۷\,(\,۱+۰\,)+۱\,(\,۰\,) = ۵۳$$

$$EV = ۵\,(\,۵+۵+۵\,)+۵\,(\,۵+۵\,)+(\,۵\,) = ۱۵۰$$

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100}$$

$$\mathbf{IQV} = \frac{\mathbf{53}}{\mathbf{150}} \times \mathbf{100} = \%\,\mathbf{35.33}$$

سمینار (ب)

$$QV = ۸\,(\,۶+۹+۵\,)+۶\,(\,۹+۵\,)+۹\,(\,۵\,) = ۲۸۹$$

$$EV = ۷\,(\,۷+۷+۷\,)+۷\,(\,۷+۷\,)+۷\,(۷) = ۲۹۴$$

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100}$$

$$\mathbf{IQV} = \frac{\mathbf{289}}{\mathbf{294}} \times \mathbf{100} = \%\,\mathbf{98.3}$$

نتیجه:

همانطور که در نتیجه مشاهده می شود، مقدار IQV در سمینار (ب) بیشتر است فلذا تنوع شرکت کنندگان در سمینار (ب) بیشتر است.

سخن پایانی :

علامات مربوط به شاخص های آماری به شرح زیر می باشد:

علائم مربوط به پارامتر و آماره

ویژگی شاخص	میانگین	واریانس	انحراف استاندارد	نسبت	همبستگی	تعدا مشاهدات (نمونه و جامعه)
آماره	X	S	S	P	r	n
پارامتر و تلفظ	μ(میو)	δ^2 سیگما ۲	δ سیگما	π(پی)	ρ(رو)	N (اِن یا نیو)

برای حل مسائل اجتماعی که نیاز به محاسبات آماری دارد، باید ، به ترتیب ، چهار مرحله را سپری کرد :

آ) تشخیص سطح متغیر (اسمی ، ترتیبی ، فاصله ای)

ب) تشخیص نوع مفهوم مورد سنجش :

چنانچه ، مواردی از قبیل (ویژگی ، خاصیت و ...) را خواست ،

باید از شاخص های مرکزی { میانگین ، میانه و نما } استفاده کرد.

چنانچه ، مواردی از قبیل (تفاوت ، تنوع ، پراکندگی و ...) را خواست ،

از شاخص های پراکندگی { انحراف معیار ، واریانس و *IQV* } استفاده می کنیم.

ج) انجام عملیات محاسباتی

د) تفسیر نتایج بدست آمده و ارائه ی پاسخ سوال.

پایان
شرح درس
آمار مقدماتی

پیوست جزوه ی آمار مقدماتی:

نمونه

سوالات

و

مسائل امتحانی

سوال ۱)

برای هرکدام از متغیرهای اسمی ، ترتیبی و فاصله ای ، به ترتیب پنج مثال بزنید :

آ) متغیر اسمی :

تعریف : از متغیر اسمی در ساده‌ترین سطح اندازه‌گیریها، (یعنی زمانی که داده‌ها براحتی در دسته‌های مختلف قرار گیرند) استفاده می‌شود . متغیرهای اسمی که، پاسخ به آنها یا «بلی» است یا «خیر» را متغیرهای دو حالتی یا dichotomous می‌نامند.

مثال اول: سابقه ابتلا به بیماری های سیستمیک ؛ که پاسخ به صورت (بلی ، خیر) مشخص شده باشد .

مثال دوم: متغیر نوع کم خونی که به سه دسته تقسیم می‌شود: میکروسیتیک، ماکروسیتیک و نرموسیتیک .

مثال سوم: متغیر نوع محل سکونت که به صورت (پدری ، شخصی ، سازمانی ، استیجاری ، غیره) دسته بندی شده باشد .

مثال چهارم : متغیر اشتغال که به صورت (شاغل ، بیکار) دسته بندی شده باشد .

مثال پنجم: متغیر رنگ چشم که به صورت (مشکی ، آبی ، قهوه ای ، میشی) دسته بندی شده باشد .

مثال ششم: متغیر رنگ پوست که به صورت (سفید ، سیاه ، سبزه ، زرد ، سرخ) دسته بندی شده باشد .

مثال هفتم : متغیر نوع معلولیت : ۱- جسم ۲- ذهنی ۳- روانی ۴- حسی

ب) متغیر ترتیبی :

تعریف : اگر بین دسته‌ها ترتیب ذاتی وجود داشته باشد، برای اندازه‌گیری این مشاهدات از مقیاس ترتیبی استفاده می‌شود . یعنی متغیر یا مشاهده از نوع رتبه‌ای است. بنابراین در متغیر ترتیبی، دسته‌بندی با مراعات تقدم و تأخر صورت می گیرد، لازم به ذکر است که یکی از ویژگی های مقیاس های ترتیبی این است که با وجود ترتیب بین دسته‌ها ، دامنه دسته‌های مختلف یکسان نیست .

لازم به تذکر است برای گزارش متغیر ترتیبی مانند مقیاس اسمی اکثراً از درصد و نسبت استفاده می‌شود .

با این مقیاس‌ها می‌توان علاوه بر تشخیص وجود یا عدم وجود صفت، شدت و ضعف آن را نیز سنجید.

مثال اول: متغیر دسته بندی تومورها که بر حسب درجهٔ توسعهٔ آنها مرحله‌بندی می‌شوند.

مثال دوم: دسته‌بندی بین‌المللی برای مشخص کردن مرحلهٔ «سرطان گردن رحم» که شامل یک متغیر ترتیبی پنج گروهی از O تاIV می‌باشد.

مثال سوم: نمرات آپگار که مربوط به ارزیابی تکامل نوزادان در موقع تولد می‌باشد، در این شیوه ارزیابی، تعداد یازده دسته از صفر تا ۱۰ وجود دارد. نمرات پایین مربوط به نوزادانی است که از نظر عملکرد قلبی ریوی و عصبی وضعیت خوبی ندارند. دسته‌های بالاتر نشانه سلامت سیستم قلبی ریوی و عصبی می‌باشد. تفاوت بین دسته ۸ و ۱۰ از نظر مقدار با تفاوت بین دو گروه صفر و ۲ یکسان نمی‌باشد.

مثال چهارم: رتبه بندی ویژگی های شغلی نمونه مورد نظر بر حسب درجه ی اهمیت آن ویژگی ها .

مثال پنجم: سنجش مشارکت دانشجویان در فعالیت‌های آموزشی که قابل تقسیم به سه مقوله (فعالیت بسیار زیاد ، فعالیت متوسط و فعالیت کم) است .

مثال ششم: سنجش میزان پایبندی پیروان یک آئین ، به دین خود که بصورت (خیلی زیاد ، زیاد ، متوسط ، کم ، خیلی کم) دسته بندی شده باشد .

مثال هفتم: متغیر میزان ناتوانی ((۱- جزئی ، ۲- کم ، ۳- متوسط ، ۴- زیاد ، ۵- شدید))

ج) متغیر فاصله ای :

مقیاسی با درجات مساوی است . هنگامی که نتایج یک تحقیق را بر روی چنین مقیاسی وارد می‌کنیم، هم امکان رده‌بندی آنها در دو جهت (از پایین به بالا و از بالا به پائین) وجود دارد و هم مقایسه داده‌ها امکان‌پذیر است چون درجات مساوی هستند .

بنابراین در مقیاس فاصله‌ای علاوه بر دارا بودن یا دارا نبودن یک صفت و شدت و ضعف آن در افراد، می‌توان سنجید که مثلاً یک شخص یک صفت را چقدر بیشتر یا کمتر از شخص دیگر دارا است. فی الواقع مقیاس فاصله‌ای علاوه بر دارا بودن صفات مقیاس‌های اسمی و ترتیبی، دارای این ویژگی است که می‌تواند فواصل بین نمرات را نیز مشخص کند یا بعبارتی آن را کمّی نماید؛ که به این ترتیب می‌توانیم علاوه بر تعیین رتبه هر یک، فاصله زمانی انجام دادن کار بین آنها را نیز مشخص نمائیم.

کمّی شدن ارزش هر یک از صفات در این مقیاس‌ها راه را برای تجزیه و تحلیل آماری داده‌های تحقیق هموار می‌نماید.

مثال اول : مقیاس قراردادی به کار گرفته شده در دماسنجهای سانتی گراد و فارنهایت و یا متغیر درجه ی حرارت

مثال دوم : میزان های قراردادی در اندازه گیری سطوح سنجش بهره ی هوشی دانش آموزان مقطع سوم دبیرستان

مثال سوم : درجه بندی های قراردادی میزان شنوائی(بر حسب دسی بل)

مثال چهارم : درجه بندی های آزمایش های خونی مثل میزان کلسترول

مثال پنجم : سطح سنجی میزان پیشرفت پروژه های عظیم زیرساختی ، بصورت عددی در توالی زمانی

سوال ۲)

با مراجعه به سایت مرکز آمار ایران و استخراج اطلاعات آماری (جمعیت کل ، تعداد باسواد ، تعداد بیسواد) دو شهرستان ، اقدام به احتساب بهرگان پارامترهای مورد نظر نمائید .

شهرستان شماره یک : نجف آباد

جمعیت کل : ۳۰۰۲۸۸ نفر

جمعیت زیر شش سال : ۲۹۹۷۳ نفر

جمعیت کل ۶ ساله به بالا (سنجش سواد) : ۲۷۰۳۱۵ نفر

تعداد با سواد : ۲۳۹۹۷۳ نفر

تعداد بیسوارد : ۲۸۸۴۶ نفر

اظهار نشده : ۱۴۹۶ نفر

شهرستان شماره ی دو : شهر رضا

جمعیت کل : ۱۴۹۵۳۲ نفر

جمعیت زیر شش سال : ۱۳۳۹۷ نفر

جمعیت کل۶ ساله به بالا (سنجش سواد) : ۱۳۶۱۳۵ نفر

تعداد با سواد : ۱۱۷۲۹۶ نفر

تعداد بیسواد : ۱۸۰۹۸ نفر

اظهار نشده : ۷۴۱ نفر

بهرگان به جمعیت کل و ٦ سال به بالا .	شهرستان شهرضا ، اصفهان	شهرستان نجف آباد ، اصفهان	
شهرضا به کل : ۷۸ درصد شهرضا به ٦ سال به بالا : ۸٦ درصد ------------- نجف آباد به کل : ۸۰ درصد نجف آباد به ٦ سال به بالا : ۸۹ درصد	۱۱۷۲۹٦ نفر	۲۳۹۹۷۳ نفر	تعداد افراد باسواد بالای شش سال
شهرضا به کل : ۱۲ درصد شهرضا به ٦ سال به بالا : ۱۳ درصد ------------- نجف آباد به کل : ۱۰ درصد نجف آباد به ٦ سال به بالا : ۱۰.۵ درصد	۱۸۰۹۸ نفر	۲۸۸٤٦ نفر	تعداد افراد بیسواد بالای شش سال
شهرضا به کل : ۰.۵۰ درصد شهرضا به ٦ سال به بالا : ۱ درصد ------------- نجف آباد به کل : ۰.۵۰ درصد نجف آباد به ٦ سال به بالا : ۰.۵ درصد	۷٤۱ نفر	۱٤۹٦ نفر	تعداد افراد اظهار نشده
شهرضا : ۹ درصد ------------- نجف آباد : ۱۰ درصد	۱۳۳۹۷ نفر	۲۹۹۷۳ نفر	جمعیت زیر شش سال
شهرضا : ۹۱ درصد ------------- نجف آباد : ۹۰ درصد	۱۳٦۱۳۵ نفر	۲۷۰۳۱۵ نفر	جمعیت بالای شش سال
شهرضا : ۱ ------------- نجف آباد : ۱	۱٤۹۵۳۲ نفر	۳۰۰۲۸۸ نفر	جمعیت کل

تحلیل آماری :

با توجه به یافته های آماری جدول بالا ، و مقایسه ی دو شهرستان شهرضا و نجف آباد استان اصفهان ، و بررسی تطبیقی میزان سواد ، به این نتیجه می رسیم که :

نسبت سطح سواد ، با فاصله ی اندکی ، در شهرستان نجف آباد ، نسبت به شهرستان شهرضا ، از مطلوبیت بهتری برخوردار است .

فلذا این مهم را می توان در قالب بررسی میزان های مختلف همچون سطح کیفی و کمی آموزشی ، بینش خانواده ها و ... ، مورد واکاوی و تحلیل دقیق تر ، قرار داد .

(منبع داده ها : وبسایت مرکز آمار ایران www.sci.org.ir ؛ نتایج سرشماری عمومی نفوس و مسکن ۱۳۹۰)

سوال ۳)

پس از مراجعه به سایت مرکز آمار ایران و انتخاب داوطلبانه چهار شهر یا استان در چهار جهت جغرافیایی کشور و استخراج اطلاعات آماری (جمعیت در سن ازدواج ، تعداد ازدواج کرده ها) اقدام به احتساب درصد تاهل نموده و نتیجه را تحلیل مقایسه ای نمائید .

فرمول درصد تاهل : ۱۰۰ × (جمعیت در سن ازدواج / ازدواج کرده ها) = درصد تأهل

منطقه ی غرب کشور : ایلام

جمعیت کل : ۲۱۳۵۷۹ نفر

جمعیت در سن ازدواج (۱۰ ساله و بیشتر) : ۱۸۱۲۸۲ نفر — ۵۸ درصد = درصد تأهّل در ایلام

جمعیت ازدواج کرده ها : ۱۰۴۹۰۸

منطقه ی شرق کشور : سربیشه خراسان جنوبی

جمعیت در سن ازدواج (۱۰ ساله و بیشتر) : ۳۱۵۲۶ نفر — ۶۵ درصد = درصد تأهل در سربیشه

جمعیت ازدواج کرده ها : ۲۰۵۶۸ نفر

منطقه ی جنوب کشور : دشتستان بوشهر

جمعیت در سن ازدواج (۱۰ ساله و بیشتر) : ۱۸۹۱۳۰ نفر — ۶۱.۶۴ درصد = درصد تأهل دشتستان

جمعیت ازدواج کرده ها : ۱۱۶۵۹۴ نفر

منطقه ی شمال کشور : ساری مازندران

جمعیت در سن ازدواج (۱۰ ساله و بیشتر) : ۴۱۷۴۲۷ نفر — ۶۹.۶۶ درصد = درصد تأهل در ساری

جمعیت ازدواج کرده ها : ۲۹۰۸۰۸ نفر

تحلیل آماری :

با توجه به نتایج تقریبی که در بالا آمده است ، مشاهده می کنیم که درصد تاهّل در ساری بیش از نقاط دیگر بوده است و سپس سربیشه ، دشتستان و ایلام در جایگاه های بعدی قرار دارند .

می توان با در نظر گرفتن شرایط جغرافیایی ، موقعیت های شغلی و دیگر داده های آماری مربوطه به بررسی علل میزان بالای تاهل در ساری مازندران نسبت به دیگر ناط مورد بررسی پرداخت .

اجمالا ، ذکر این نکته ، خالی از فایده نیست که : تفاوت اندک در میزان تاهل در نقاط مختلف کشور ، بیانگر این نکته است که سطح تاهل در سرزمین ایران ، نسبت مطلوبی را برخوردار است .

(منبع داده ها : وبسایت مرکز آمار ایران www.sci.org.ir ؛ نتایج سرشماری عمومی نفوس و مسکن ۱۳۹۰)

سوال ۴)

سرانه پزشک را را برای استان های مختلف، در دو بازه ی زمانی متفاوت یافته، انحرف معیار و دامنه ی آنرا محاسبه کنید.

121- تعداد و سرانه پزشکان(1و2) و تخت‌هاي مؤسسات درماني فعال: 1384

استان	تعداد پزشکان	سرانه پزشك به ازاي هر 1000 نفر جمعيت
کل کشور........................	*26564*	*39/0*
آذربايجان شرقي........................	1152	33/0
آذربايجان غربي........................	835	28/0
اردبيل........................	350	28/0
اصفهان........................	1933	43/0
ايلام........................	144	26/0
بوشهر........................	379	46/0
تهران........................	5313	44/0
چهارمحال و بختياري........................	424	50/0
خراسان جنوبي........................	(3)000	(3)000
خراسان رضوي........................	(3)2967	(3)46/0
خراسان شمالي........................	(3)000	(3)000
خوزستان........................	1416	33/0
زنجان........................	399	41/0
سمنان........................	416	71/0
سيستان و بلوچستان........................	591	26/0
فارس........................	1470	34/0
قزوين........................	353	30/0
قم........................	306	29/0
کردستان........................	587	37/0
کرمان........................	934	38/0
کرمانشاه........................	587	30/0
کهگيلويه و بويراحمد........................	248	36/0
گلستان........................	556	34/0
گيلان........................	986	41/0
لرستان........................	502	29/0
مازندران........................	1505	53/0
مرکزي........................	507	37/0
هرمزگان........................	492	37/0
همدان........................	633	36/0
يزد........................	579	60/0

(1)منظور پزشكان شاغل در وزارت بهداشت، درمان و آموزش پزشكي است و شامل پزشكان خار

(2) شامل دندانپزشك، دامپزشك و داروساز نيز مي‌باشد.

(3) آمار استان‌هاي خراسان جنوبي و شمالي در استان خراسان رضوي منظور شده است.

مأخذ - وزارت بهداشت، درمان و آموزش پزشكي.

- مرکز آمار ايران.

دامنه سرانه پزشک به ازای هر ۱۰۰۰ نفر در سال ۱۳۸۴ :

D = ۷۱ (SEMNAN) – ۲۶ (ILAM) = ۴۵

انحراف معیار برای سال ۱۳۸۴ :

$$S = \sqrt{\sum_{I=۱}^{X} \frac{(X-\bar{X})^{۲}}{N}}$$

$$\bar{X} = \frac{(X۱ + X۲ + X۳ + \cdots + XN)}{N} = ۳۹$$

$$\sum (Xi - X)^{۲} = \quad ۳۰۲۰$$

$$S = \sqrt{\frac{۳۰۲۰}{۲۷}} = ۱۱۱.۹ \quad = ۱۰.۶$$

همانطور که مشاهده می کنیم

دامنه سرانه : ۴۵

میانگین : ۳۹

انحراف معیار : ۱۰.۶

تفسیر:

همانگونه که مشاهده می کنیم در محاسبات ۲۷ استان پیرامون سرانه ی پزشک دامنه ی تغییرات داده ها، برابر ۴۵ می باشد و با نگاه به کمینه و بیشینه مشخص می گردد که توزیع نه چندان عادلانه ای در سطح کشور از لحاظ سرانه پزشک داریم .

121- تعداد و سرانه‌ي پزشكان[1] و تختهاي مؤسسات درماني فعال: 1383

استان	تعداد پزشكان	سرانه‌ي پزشك به ازاي هر 1000 نفر جمعيت
كل كشور........................	۲۴۶۶۱	۳۷/۰
آذربايجان شرقي........................	۱۷۹۹	۵۲/۰
آذربايجان غربي........................	۷۸۸	۲۷/۰
اردبيل........................	۳۵۴	۲۸/۰
اصفهان........................	۱۳۴۰	۳۰/۰
ايلام........................	۱۲۱	۲۲/۰
بوشهر........................	۳۹۲	۴۸/۰
تهران........................	۴۴۹۸	۳۸/۰
چهارمحال و بختياري........................	۴۱۲	۴۹/۰
خراسان جنوبي[2]........................	...	...
خراسان رضوي[2]........................	۱۸۸۸	۲۹/۰
خراسان شمالي[2]........................	...	...
خوزستان........................	۱۴۱۳	۳۳/۰
زنجان........................	۳۹۱	۴۱/۰
سمنان........................	۳۹۸	۶۹/۰
سيستان و بلوچستان........................	۶۱۶	۲۸/۰
فارس........................	۱۴۳۲	۳۳/۰
قزوين........................	۳۱۰	۲۷/۰
قم........................	۳۱۵	۳۰/۰
كردستان........................	۷۵۱	۴۹/۰
كرمان........................	۹۱۸	۳۹/۰
كرمانشاه........................	۵۵۵	۲۹/۰
كهگيلويه و بويراحمد........................	۲۳۶	۳۵/۰
گلستان........................	۵۳۷	۳۳/۰
گيلان........................	۱۰۰۸	۴۲/۰
لرستان........................	۵۰۸	۲۹/۰
مازندران........................	۱۴۲۳	۵۱/۰
مركزي........................	۵۸۵	۴۳/۰
هرمزگان........................	۴۷۰	۳۷/۰
همدان........................	۶۲۷	۳۶/۰
يزد........................	۵۷۶	۶۱/۰

(1) منظور پزشكان شاغل در وزارت بهداشت، درمان و آموزش پزشكي است و شامل پزشـ
(2) آمار استان‌هاي خراسان جنوبي و خراسان شمالي در استان خراسان رضوي منظور شد
مأخذ ـ وزارت بهداشت، درمان و آموزش پزشكي.
ـ مركز آمار ايران.

دامنه سرانه پزشک به ازای هر ۱۰۰۰ نفر در سال ۱۳۸۳ :

D = ۶۹ (SEMNAN) – ۲۲ (ILAM) = ۴۷

انحراف معیار برای سال ۱۳۸۳ :

$$S = \sqrt{\sum_{I=1}^{X} \frac{(X-\bar{X})^2}{N}}$$

$$\bar{X} = \frac{(X_1 + X_2 + X_3 + \cdots + X_N)}{N} = 37$$

$$\sum (Xi - X)^2 = 3443$$

$$S = \sqrt{\frac{3443}{27}} = 127.5 = 11.3$$

همانطور که مشاهده می کنیم

دامنه سرانه : ۴۷

میانگین : ۷۳

انحراف معیار : ۱۱.۳

تفسیر:

در این سال نیز، با توجه به بیشینه و کمینه
همچنین دامنه ی تغییرات
شاهد عدم توازن و عدالت توزیعی پزشک
نسبت به جمعیت استان ها داریم
و همانگونه که مشاهده می شود
استان های محروم و یا با تراکم جمعیتی بالا
نسبت به سایر استان ها
با مشکلات بیشتری از لحاظ
سرانه پزشک مواجه هستند

سوال۵)

با مراجعه به سایت مرکز آمار ایران و انتخاب یک شهر و استخراج داده های آماری آن ، نسبت پزشکان آن شهر را به کل جمعیت ، دریابید .

جمعیت کل شهرستان اصفهان بر مبنای سرشماری سال نفوس و مسکن سال ۱۳۹۰ : ۲۱۷۴۱۷۲ نفر

جمعیت کل متخصصان حوزه ی بهداشت و رفاه : ۵۵۸۰ نفر

۳۸۹.٦۳ = ۲۱۷٤۱۷۲/۵۵۸۰ = نسبت متخصصان حوزه ی بهداشت و رفاه به کل جمعیت شهرستان اصفهان

تحلیل : همانگونه که نتایج یافته ها ، بیان می کند ، در شهرستان اصفهان ، به ازای هر ۳۸۹.٦۳ نفر ؛ یک متخصص در حوزه ی بهداشت و رفاه وجود دارد . که می توان با مراجعه به نُرم های استاندارد این حوزه ، میزان فاصله ی این یافته را با مقدار استاندارد بدست آورد .

سوال ۶)

با مراجعه به وبسایت مرکز آمار ایران و مراجعه به فیلد سری های زمانی و استخراج داده های پایه ای سال ۱۳٦۵ و ۱۳۹۰ ، و مشاهده ی آمارهای قضایی ، نرخ رشد پرونده های مختومه ی قضایی ؛ نرخ رشد قتل عمد را در فاصله ی سالهای ۱۳٦۳ تا ۱۳۸۷ محاسبه و گزارش دهید .

با مراجعه به سایت مرکز آمار ایران و استخراج اطلاعات آماری (جمعیت کل ، تعداد باسواد ، تعداد بیسواد) دو شهرستان ، اقدام به احتساب بهرگان پارامترهای مورد نظر نمائید .

۳.۸ = ۱۰۰۰۰۰ * ۱۴۲۰۴۹۲ / ۵۵ = شاخص قتل عمد سال ۱۳۶۳ بر پایه ی سال ۱۳۶۵

۱۳.۳ = ۱۰۰۰۰۰ * ۴۸۷۹۳۱۲ / ۶۵۰ = شاخص قتل عمد سال ۱۳۸۷ بر پایه ی سال ۱۳۹۰

٪۲.۵ = ۳.۸ / (۳.۸ - ۱۳.۳) = نرخ رشد شاخص قتل عمد

.

(منبع داده ها : وبسایت مرکز آمار ایران www.sci.org.ir)

سوال ۷)

جدول الف ، مصرف انرژی (بر حسب صد هزار تن) ؛ و جدول ب ، تولید انرژی (بر حسب صد هزار تن) ، در ۱۶۰ کشور را نشان می دهد. آیا می توان گفت ، اکثر کشورها ، انرژی تولید می کنند و کشورهای معدودی ، انرژی تولید شده را مصرف می کنند ؟ آیا می توان گفت تولید انرژی میان تعداد زیادی از کشورها پراکنده است امّا مصرف آن ، صرفا در چند کشور متمرکز شده است؟

جدول (الف) **جدول (ب)**

مصرف انرژی (صد هزار تن معادل نفت) **تولید انرژی (صد هزار تن معادل نفت)**

١٤٣	۰ — ۹	//	۱۴۰	۰ — ۹
۴	۱۰ — ۱۹	//	۸	۱۰ — ۱۹
۵	۲۰ — ۲۹	//	۴	۲۰ — ۲۹
۰	۳۰ — ۳۹	//	۳	۳۰ — ۳۹
۲	۴۰ — ۴۹	//	۱	۴۰ — ۴۹
۳	۵۰ — ۵۹	//	۱	۵۰ — ۵۹
۲	۶۰ — ۶۹	//	۰	۶۰ — ۶۹
۰	۷۰ — ۷۹	//	۲	۷۰ — ۷۹
۰	۸۰ — ۸۹	//	۰	۸۰ — ۸۹
۰	۹۰ — ۹۹	//	۰	۹۰ — ۹۹
۰	۱۰۰ — ۱۰۹	//	۰	۱۰۰ — ۱۰۹
۰	۱۱۰ — ۱۱۹	//	۰	۱۱۰ — ۱۱۹
۱	۱۲۰ — ۱۲۹	//	۱	۱۲۰ — ۱۲۹

پاسخ:

همانگونه که در سوال مستتر است، جهت حل مسئله، ما نیاز به پیمودن مراحل محاسباتی انحراف معیار داریم .

فرمول مورد نیاز عبارت است از :

$$S = \sqrt{\frac{\sum_{I-1}^{N} Fi(m_i^2) - \frac{(\sum Fi\, mi)^2}{N}}{N}}$$

مرحله ی ترسیم جدول توزیع فراوانی برای تولید و مصرف:

{ تولید انرژی }　　　　　　　　　　　　{ مصرف انرژی }

حدود مقرر	Fi	mi	Fi mi	Fi mi^2	*****************	حدود مقرر	Fi	mi	Fi mi	Fi mi^2
۰ – ۹	۱۴۰	٤.٥	٦۳۰	۲۸۳۵	*****************	۰ – ۹	۱٤۳	٤.٥	٦٤۳.٥	۲۸۹٥.۷٥
۱۰ – ۱۹	۸	۱٤.٥	۱۱٦	۱٦۸۲	*****************	۱۰ – ۱۹	۴	۱٤.٥	٥۸	۸٤۱
۲۰ – ۲۹	۴	۲٤.٥	۹۸	۲٤۰۱	*****************	۲۰ – ۲۹	۵	۲٤.٥	۱۲۲.٥	۳۰۰۱.۲٥
۳۰ – ۳۹	۳	۳٤.٥	۱۰۳.٥	۳٥۷۰.۷٥	*****************	۳۰ – ۳۹	۰	۳٤.٥	۰	۰
۴۰ – ۴۹	۱	٤٤.٥	٤٤.٥	۱۹۸۰.۲٥	*****************	۴۰ – ۴۹	۲	٤٤.٥	۸۹	۳۹٦۰.٥
۵۰ – ۵۹	۱	٥٤.٥	٥٤.٥	۲۹۷۰.۲٥	*****************	۵۰ – ۵۹	۳	٥٤.٥	۱٦۳.٥	۸۹۱۰.۷٥
۶۰ – ۶۹	۰	٦٤.٥	۰	۰	*****************	۶۰ – ۶۹	۲	٦٤.٥	۱۲۹	۸۳۲۰.٥
۷۰ – ۷۹	۲	۷٤.٥	۱٤۹	۱۱۱۰۰.٥	*****************	۷۰ – ۷۹	۰	۷٤.٥	۰	۰
۸۰ – ۸۹	۰	۸٤.٥	۰	۰	*****************	۸۰ – ۸۹	۰	۸٤.٥	۰	۰
۹۰ – ۹۹	۰	۹٤.٥	۰	۰	*****************	۹۰ – ۹۹	۰	۹٤.٥	۰	۰
۱۰۰ – ۱۰۹	۰	۱۰٤.٥	۰	۰	*****************	۱۰۰ – ۱۰۹	۰	۱۰٤.٥	۰	۰
۱۱۰ – ۱۱۹	۰	۱۱٥.٥	۰	۰	*****************	۱۱۰ – ۱۱۹	۰	۱۱٥.٥	۰	۰
۱۲۰ – ۱۲۹	۱	۱۲٥.٥	۱۲٥.٥	۱٥۷٥۰.۲٥	*****************	۱۲۰ – ۱۲۹	۱	۱۲٥.٥	۱۲٤.٥	۱٥٥۰۰.۲٥

$$S = \sqrt{\frac{\sum_{I-1}^{N} Fi(m_i^2) - \frac{(\sum Fi\,mi)^2}{N}}{N}}$$

$$S_{\text{تولید نفت}} = \sqrt{\frac{\sum_{I-1}^{N} Fi(m_i^2) - \frac{(\sum Fi\,mi)^2}{N}}{N}} = \sqrt{\frac{42290 - \frac{(1321)^2}{160}}{160}} = 14$$

$$S_{\text{مصرف نفت}} = \sqrt{\frac{\sum_{I-1}^{N} Fi(m_i^2) - \frac{(\sum Fi\,mi)^2}{N}}{N}} = \sqrt{\frac{43430 - \frac{(1330)^2}{160}}{160}} = 14.22$$

تولید نفت $<$ مصرف نفت

تفسیر:

با توجّه به نتایج حاصله، انحراف معیار مصرف ، بیش از انحراف معیار تولید می باشد و در نتیجه ، تمرکز در تولید نفت ، بیش از مصرف آن است ، فی الواقع در مصرف نفت، موازنه ی کمتری نسبت به تولید آن مشاهده می شود فلذا می توان گفت که در مصرف نفت ، تعداد اندکی از کشورها ، مصرف بیشتری دارند حال آنکه ، در تولید نفت، سهم کشورها، دارای موازنه ی بهتری است. همانگونه که میدانیم ، نفت ، خام است و ماده ی اولیه بسیاری از تولیدات صنعتی ، همینطور می دانیم تولید صنعتی در انحصار کشورهای اندکی است و در نتیجه مصرف آنان نیز بیشتر از سایر کشورهاست ، فلذا منطقی است پاسخی که بدست آمده است.

سوال ۸)

یکی از معیارهای زیست محیطی ، هوای پاک است . اطلاعات دو کشور الف و ب در سال ۲۰۱۰ به لحاظ میزان نشر گاز دی اکسید کربن (بعنوان معیار میزان تولید آلودگی هوا) و مساحت جنگلها (بعنوان معیار از بین برنده ی آلودگی هوا) در جدول ذیل داده شده است . کدام کشور برای انسان، محیط پاک تری است؟

	کشور الف	کشور ب
مساحت جنگلها (کیلومتر مربع)	۷۰۰	۲۹۸۸۰
میزان نشر گاز دی اکسید کربن (هزار تن)	۷۲۱٦۱۳	۳٤۸۰
جمعیت (میلیون)	۸۱	۳۳

پاسخ: آماره ی مناسب جهت حل این مسئله ، نسبت است چرا که ما باید نسبت جنگل به آلودگی را برای دو کشور بسنجیم و سپس ضمن مقایسه ی نتایج حاصله ، اقدام به تفسیر نمائیم.

آ) محاسبات کشور الف:

سهم هر یک میلیون نفر از بهره ی جنگل برابر است با ۸.۶۴ کیلومتر مربع : $\rightarrow$ ۸.۶۴ = $\frac{۷۰۰}{۸۱}$ = سهم جنگل برای جمعیت

سهم هر یک میلیون نفر از بهره ی آلودگی برابر است با ۸۹۰۸.۸۰ هزار تن: $\rightarrow$ ۸۹۰۸.۸۰ = $\frac{۷۲۱۶۱۳}{۸۱}$ = سهم CO_2 برای جمعیت

آ) محاسبات کشور ب:

سهم هر یک میلیون نفر از بهره ی جنگل برابر است با ۹۰۵.۴۵ کیلومتر مربع : $\rightarrow$ ۹۰۵.۴۵ = $\frac{۲۹۸۸۰}{۸۱}$ = سهم جنگل برای جمعیت

سهم هر یک میلیون نفر از بهره ی آلودگی برابر است با ۴۲.۹۶ هزار تن: $\rightarrow$ ۴۲.۹۶ = $\frac{۳۴۸۰}{۸۱}$ = سهم CO_2 برای جمعیت

تفسیر: همانگونه که مشاهده می کنیم در کشور(الف)، سهم کمتری از فضای سبز و جنگل و محیط پاک، در اختیار افراد است نسبت به کشور(ب)

سوال ۹)

جمعیت مسلمانان در یک کشور اروپایی در سالهای ۱۹۹۰ و ۲۰۱۰ به قرار زیر است . آیا نگرانی برخی فعالان سیاسی از افزایش مسلمانان در این کشور ، جنبه ی واقعی دارد؟

	۱۹۹۰	۲۰۱۰
جمعیت مسلمانان	۱۱۷۲۰۰۰	۲۸٦۹۰۰۰
جمعیت کشور	۵۷۰۰۰۰۰۰	٦۲۰۰۰۰۰۰

پاسخ: آماره ی مناسب برای این سوال، بهرگان یا درصد (٪) است : { ۱۰۰ × (ni / N) = بهرگان }

درصد جمعیت مسلمانان در سال ۱۹۹۰ = $\frac{۱۱۷۲۰۰۰}{۵۷۰۰۰۰۰۰} \times ۱۰۰ =$ ۲.۰۵ %

درصد جمعیت مسلمانان در سال ۲۰۱۰ = $\frac{۲۸۶۹۰۰۰}{۶۲۰۰۰۰۰۰} \times ۱۰۰ =$ ۴.۶ %

تفسیر: همانگونه که مشاهده می کنیم ، تعداد مسلمانان در سال ۲۰۱۰ ، نسبت به سال ۱۹۹۰ ، حدود (۲.۵ ٪) رشد داشته است فلذا نگرانی برخی از فعالان سیاسی را می توان واقع گرایانه دانست.

سوال ۱۰)

جدول ذیل ، ضریب نفوذ اینترنت یا همان تعداد کاربران اینترنت (در هر ۱۰۰۰۰ نفر) در ۱۷۰ کشور مختلف جهان را برای سال ۱۹۹۷ نشان می دهد .

میانگین ضریب نفوذ اینترنت کشورها چقدر است ؟

ضریب نفوذ چه تعداد از کشورها را می توان با اطمینان گفت که بیشتر از مقدار میانگین است ؟

این تعداد ، چند درصد کشورها را در بر می گیرد ؟ از درصد بدست آمده چه نتیجه ای می گیرید؟

٥	٣١	١٧٣	٥	٨٠	٣٩٢	٥٤	٠	٢٠
١	٧٩	٧	٦	٩١٦	٠	١٠	١٦٣	٣٣٠
٣٠٥	٤٨٣	٤٣٨	٦٧١	٦٠	٤٨	١٠	٢٨٠	٧٣٩
١	١٢٢	٢٩٢	٣	٦	٦٣	٤٠	١٦	٢١٦٢
٣٤٨	٢	١١٣٨	١٨٥	٣	٥٧	١٤	٩٩	٣٣٦
٢٨	١	٨	٧٩٥	٣٦٠	٣	٢٠٧	٩٣	١
١١	١	١٤	١٠	٢١٧	١١	٤٩٥	٠	٣٩
١٦٣٧	١	١١	٠	٢٠٥	٢	١٣٣	١٠٠	٠
٩٥٣	١٥٠٧	٩	٢	١٢٤	١	٣٠٨	٩	٢
٣	٢٤	٢٦	١٣	١	٦	٤٥	٢٣٧٣	١
١٣٦	١	٤	١٧	٠	٢	٤٧	١٥١٠	
١٦٥	٠	١	١٠٥٢	٩٨	١٤٠٧	٠	٣	
٠	١٠٦	٥٧١	١٩٤	٧١٤	٩٩	١٤٢	١	
٧٨	٣	٠	٢٧٤٨	٢٣٦	١٤٦٠	٥	٣٦	
٥	٥٥	٤٦٠	٧	٥٠	٢١	٣	٢١	
٤٩٤	٠	٢٢	١٩	١	٠	١٢٨	٥١	
٢	٠	١٩٤٦	٥	٠	٢	١٣٤٧	١١٧	
٢٤١٩	٠	٤٢٦	٤٠٩	٢٣١	٢٠٤٢	١١٧	٤	
٤٥	١٦٤	٢١	٤٤٠	٣١	٤٤	٧٥٩	٤٧	
٦	٢	٥	٢٢٨	١	٣	٣٩	١	

پاسخ:

همانگونه که در متن سوال مشخص است ، ابتدا باید داده ها را دسته بندی نمود، سپس میانگین داده های دسته بندی شده را محاسبه کرد و آنگاه با توجه به فراوانی تجمعی ترسیم شده، پاسخ مسئله را دریافت:

فراوانی تجمعی پائین رونده	کران پائین حدود واقعی	حدود واقعی	Fi mi	mi	Fi	حدود مقرر
۱۷۰	۰.۵	۰.۵ – ۲۹۹.۵	۱۹۷۳۴	۱۴۹.۵	۱۳۲	۰ – ۲۹۹
۳۸	۲۹۹.۵	۲۹۹.۵ – ۵۹۹.۵	۷۶۴۱.۵	۴۴۹.۵	۱۷	۳۰۰ – ۵۹۹
۲۱	۵۹۹.۵	۵۹۹.۵ – ۸۹۹.۵	۳۷۴۷.۵	۷۴۹.۵	۵	۶۰۰ – ۸۹۹
۱۶	۸۹۹.۵	۸۹۹.۵ – ۱۱۹۹.۵	۴۱۹۸	۱۰۴۹.۵	۴	۹۰۰ – ۱۱۹۹
۱۲	۱۱۹۹.۵	۱۱۹۹.۵ – ۱۴۹۹.۵	۴۰۴۸.۵	۱۳۴۹.۵	۳	۱۲۰۰ – ۱۴۹۹
۹	۱۴۹۹.۵	۱۴۹۹.۵ – ۱۷۹۹.۵	۴۹۴۸.۵	۱۶۴۹.۵	۳	۱۵۰۰ – ۱۷۹۹
۶	۱۷۹۹.۵	۱۷۹۹.۵ – ۲۰۹۹.۵	۳۸۹۹	۱۹۴۹.۵	۲	۱۸۰۰ – ۲۰۹۹
۴	۲۰۹۹.۵	۲۰۹۹.۵ – ۲۳۹۹.۵	۴۴۹۹	۲۲۴۹.۵	۲	۲۱۰۰ – ۲۳۹۹
۲	۲۳۹۹.۵	۲۳۹۹.۵ – ۲۶۹۹.۵	۲۵۴۹.۵	۲۵۴۹.۵	۱	۲۴۰۰ – ۲۶۹۹
۱	۲۶۹۹.۵	۲۶۹۹.۵ – ۲۹۹۹.۵	۲۸۴۹.۵	۲۸۴۹.۵	۱	۲۷۰۰ – ۲۹۹۹
			$\sum Fi\, Mi = 58115$		$\sum Fi = 170$	C = ۲۹۹

$$mi = \frac{(\text{ابتدای دسته} + \text{انتهای دسته})}{2}$$

میانگین ضریب نفوذ اینترنت در کشورها ، برابر است با (۳۴۱.۸۵) :

$$\bar{X} = \frac{\sum_{i=1}^{k} Fi \times mi}{N} = \frac{58115}{170} = 341.85$$

در قسمت بعدی سوال ، به دلیل اینکه ، چنین طرح شده است که مقادیر بیشتر از میانگین را بیابیم ، فلذا نیاز به فراوانی تجمعی پائین رونده داریم و باید کران پائین حدود واقعی را در نظر بگیریم و بسنجیم ، در نتیجه می توان گفت: به دلیل اینکه ، مقدار کران پائین (۵۹۹.۵) حد واقعی دسته ی سوم (۸۹۹ – ۳۰۰) ، اولین مقداریست که ، اندازه ی آن ، از میانگین بدست آمده (۳۴۱.۸۵) بیشتر است ، فلذا ، مقدار فراوانی تجمعی این دسته (۲۱) بعنوان تعداد کشورهایی در نظر گرفته می شود که ضریب نفوذ اینترنت در آنها ، بیشتر از مقدار میانگین است ، فی الواقع می توان گفت: تعداد (۲۱) کشور ، ضریب نفوذ اینترنتشان ، بیشتر از مقدار میانگین است. حال برای اینکه بسنجیم ، این تعداد ، چند درصد کشورها را شامل می شود ، باید از شاخص درصد بروش زیر بهره ببریم:

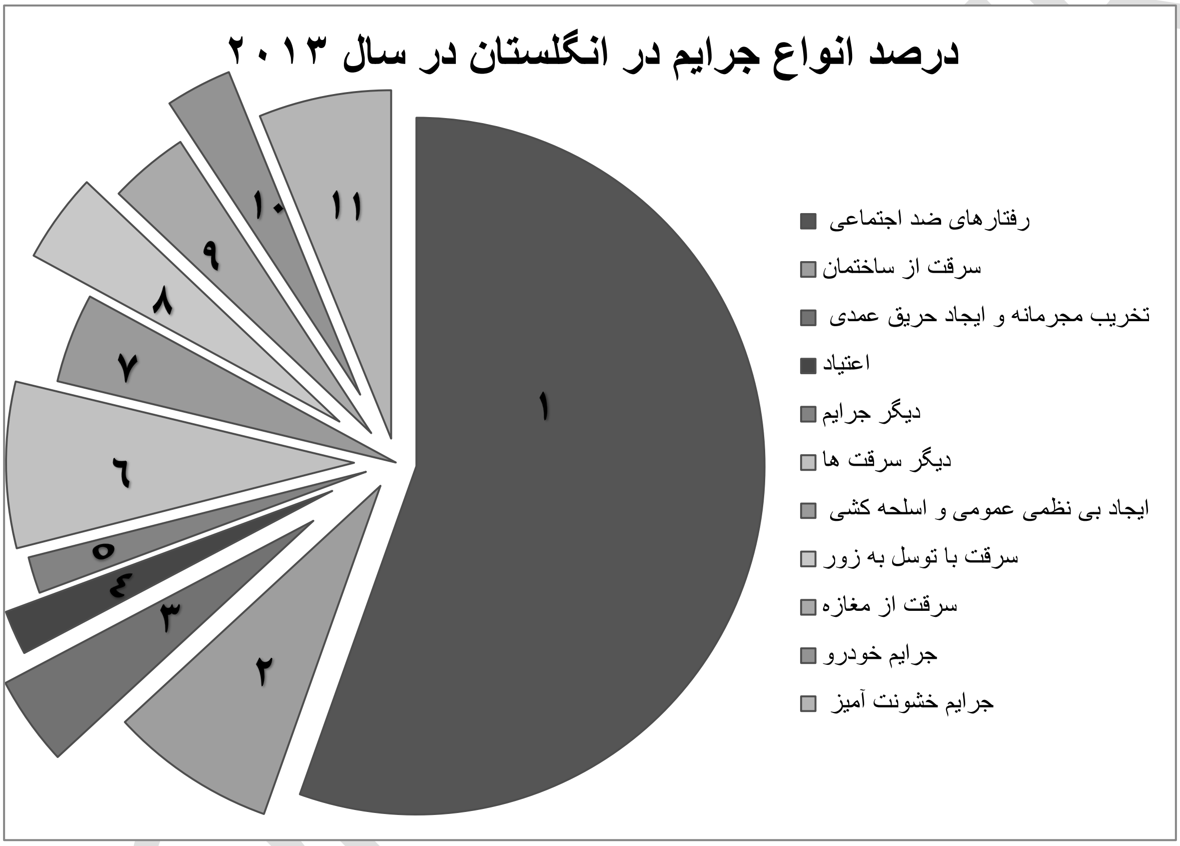
درصد ضریب نفوذ اینترنت کشورهای بیشتر از میانگین $\; n1/_N \times 100 = \% = \frac{21}{170} \times 100 = 12.35\,\%$

همانگونه که نتیجه ی بدست آمده ، گواهی می دهد ، تعداد کمی از کشورها دارای ضریب نفوذ اینترنت می باشند فلذا ضریب نفوذ اینترنت دارای نابرابری بالایی است .

سوال ۱۱)

نمودار ذیل ، درصد انواع جرایم در کشور انگلستان در سال ۲۰۱۳ را نشان می دهد .
آ) بیشترین و کمترین جرایم کدام است کدام هستند ؟
ب) سه جرم اول ، کدام جرایم هستند ؟

(راهنمایی: انواع جرایم در جهت عقربه های ساعت ، داخل نمودار به ترتیب از بالا به پائین در سمت راست چینش شده اند)



پاسخ:

آ) بیشترین جرم متعلق به رفتارهای ضد اجتماعی است و کمترین جرم متعلق به معضل اعتیاد است.

ب) سه جرم اول ، به ترتیب ، عبارت اند از :

رفتارهای ضد اجتماعی

سرقت از ساختمان

تخریب مجرمانه و ایجاد حریق عمدی

سوال ۱۲)

از پنج کشور با توسعه یافتگی متفاوت ، در خصوص میزان اندیشیدن در خصوص هدف زندگی، پیمایشی صورت گرفته است. نتایج بدست آمده در جدول زیر ، آمده است .

آیا این پنج کشور ، به لحاظ آنکه غالبا درباره هدف و معنای زندگی می اندیشند با هم تفاوت قابل ملاحظه ای دارند ؟ چه نتیجه ای می توان گرفت؟

		کشورها				
		ایران	برزیل	غنا	آمریکا	فرانسه
میزان اندیشه درباره هدف و معنای زندگی	غالبا	۱۳	۹	۸	٤	٤
	برخی اوقات	۱۰	٤	٦	٥	۳
	به ندرت	۳	۲	۱	۲	۱
	هرگز	۰	۰	۰	۰	۱
	جمع	۲٦	۱٥	۱٥	۱۱	۹

پاسخ:

همانطوریکه در صورت مسئله مطرح است ، باید تفاوت کشورها را از لحاظ (غالبا) بررسی نمود ، فلذا آماره ی مناسب ، نسبت است:

متغیر غالبا فرانسه $= \frac{4}{9} = \mathbf{0.44}$

متغیر غالبا آمریکا $= \frac{4}{11} = \mathbf{0.36}$

متغیر غالبا غنا $= \frac{\mathbf{8}}{\mathbf{15}} = \mathbf{0.53}$

متغیر غالبا برزیل $= \frac{9}{15} = \mathbf{0.6}$

متغیر غالبا ایران $= \frac{\mathbf{13}}{\mathbf{26}} = \mathbf{0.5}$

همانگونه که در نتایج نیز ، محسوس است ، تفاوت چندانی در توجه غالبی به اندیشه پیرامون هدف و معنای زندگی ، در کشورهای مورد مطالعه ، مشاهده نمی شود ، فلذا ، صرف نظر از احتمال خطای پیمایشی ، می توان به این نتیجه ، دست یافت که متغییر اندیشه پیرامون معنا و هدف زندگی ، برای جوامع مختلف حائز اهمیت است و تقریبا ، این اهمیت ، توازن قابل قبولی دارد.

سوال ۱۳)

با توجه به داده های بالا ، آیا میان مردم ایران ، به لحاظ اندیشیدن درباره ی معنای زندگی تفاوت بیشتر است یا میان مردم فرانسه؟

پاسخ : می دانیم که با یک متغیر کیفی سر و کار داریم و راهکار سنجش پراکنش در متغیرهای کیفی ، شاخص تغییرات کیفی IQV است ، فلذا برای حل سوال باید مراحل محاسباتی این شاخص را پیمود.

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100}$$

آ) محاسبات شاخص تغییر پذیری کیفی در کشور فرانسه:

$$QV = ۴\,(\,۳ + ۱ + ۱\,) + ۳\,(\,۱ + ۱\,) + ۱\,(\,۰\,) = ۲۶$$

$$EV = ۲.۲۵\,(\,۲.۲۵ + ۲.۲۵ + ۲.۲۵\,) + ۲.۲۵\,(\,۲.۲۵ + ۲.۲۵\,) + ۲.۲۵\,(\,۲.۲۵\,) = ۳۰.۴$$

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100} = \frac{\mathbf{26}}{\mathbf{30.4}} \times \mathbf{100} = \mathbf{85}\,\%$$

ب) محاسبات شاخص تغییر پذیری کیفی در کشور ایران:

$$QV = ۱۳\,(\,۱۰ + ۳ + ۰\,) + ۱۰\,(\,۳ + ۰\,) + ۳\,(\,۰\,) = ۱۹۹$$

$$EV = ۶.۵\,(\,۶.۵ + ۶.۵ + ۶.۵\,) + ۶.۵\,(\,۶.۵ + ۶.۵\,) + ۶.۵\,(\,۶.۵\,) = ۲۵۳.۲۵$$

$$\mathbf{IQV} = \frac{\text{تغییرات مشاهده شده}}{\text{تغییرات مورد انتظار}} \times \mathbf{100} = \frac{\boldsymbol{QV}}{\boldsymbol{EV}} \times \mathbf{100} = \frac{\mathbf{199}}{\mathbf{253.25}} \times \mathbf{100} = \mathbf{78}\,\%$$

تفسیر:

همانگونه که می دانیم ، شاخص IQV ، چگونگی تنوع داده ها را نشان می دهد و هر چه مقدار آن بیشتر باشد ، تنوع و تفاوت نیز بیشتر است هر چند که پراکندگی داده ها ، کمتر باشد ، فلذا با توجه به محاسبات انجام شده ، می توان نتیجه گرفت که تنوع پیرامون اندیشیدن درباره ی معنای زندگی ، در کشور ایران بیشتر از کشور فرانسه است .

	ایران		فرانسه	
	مورد انتظار	مشاهده شده	مورد انتظار	مشاهده شده
غالبا	۶.۵	۱۳	۲.۲۵	۴
برخی اوقات	۶.۵	۱۰	۲.۲۵	۳
به ندرت	۶.۵	۳	۲.۲۵	۱
هرگز	۶.۵	۰	۲.۲۵	۱
جمع	۲۶/۴=۶.۵	۲۶	۹/۴=۲.۲۵	۹

سوال ۱۴)

جدول زیر ، تعداد واحدهای مسکن معمولی به ازای هر صد خانوار در استانهای مختلف کشور برای سالهای ۱۳٦۵ و ۱۳۸۵ را نشان می دهد .

آیا می توان گفت وضعیت مسکن خانوار ایرانی در سال ۱۳۸۵ نسبت به سال ۱۳٦۵ بهتر شده است؟

استان / سالهای پژوهش	۱۳۶۵	۱۳۸۵
اصفهان	۸۳	۹۱
ایلام	۷۴	۸۹
آذربایجان شرقی	۸۹	۸۸
آذربایجان غربی	۸۶	۸۸
بوشهر	۹۳	۹۱
تهران	۸۴	۹۶
چهارمحال و بختیاری	۷۹	۸۶
خراسان	۸	۹۰
خوزستان	۸۱	۸۷
زنجان	۸۶	۸۵
سمنان	۸۹	۹۱
سیستان و بلوچستان	۷۶	۷۹
فارس	۸۲	۸۷
کردستان	۸۸	۹۲
کرمان	۸۶	۸۴
کهگیلویه و بویر احمد	۸۵	۹۱
کرمانشاه	۸۰	۸۷
گیلان	۸۶	۹۲
لرستان	۸	۸۴
مازندران	۹۰	۹۲
مرکزی	۸۶	۹۵
هرمزگان	۸۶	۸۷
همدان	۸۰	۸۶
یزد	۹۱	۹۷

پاسخ:

با توجّه به مفهوم کلی سوال، باید ، برآیندی کلی (میانگین هر سال) را بدست آورده و با هم مقایسه کنیم لازم به ذکر است به دلیل اینکه ، اعداد اندک می باشد و سوال محاسبات جزئی را نمی خواهد ، نیازی به طبقه بندی و مواردی از این قبیل نیست . فلذا کافی است از روش محاسباتی میانگین در داده های خام بهره برداری نمود :

مراحل محاسباتی به شرح زیر است:

آ) میانگین داده ها در سال ۱۳۶۵ :

$$1365\ \bar{X} = \frac{\sum_{i=1}^{N} Fi}{N} = \frac{1876}{24}$$

ب) میانگین داده ها در سال ۱۳۸۵ :

$$1385\ \bar{X} = \frac{\sum_{i=1}^{N} Fi}{N} = \frac{2135}{24}$$

تفسیر:

همانگونه که مشاهده می کنیم در سال ۱۳۶۵ ، بطور متوسط در سطح استان های کشور ، به ازای هر صد خانوار ، حدود ۷۸ واحد مسکونی وجود داشته است ، حال آنکه در سال ۱۳۸۵ ، این میزان ، به حدود ۸۹ واحد مسکونی ارتقاء یافته است فلذا می توان گفت ، وضعیت مسکن خانوار ایرانی در سال ۱۳۸۵ ، نسبت به سال ۱۳۶۵ ، بهتر شده است.

سوال ۱۵)

جداول زیر ، نرخ زاد و ولد (متوسط تعداد فرزندان به ازای هر زن) را برای ۱۰۰ کشور جهان در سالهای ۱۹۹۰ و ۲۰۱۰ ؛ نشان داده است .

بررسی کنید که آیا فرهنگ کشورهای مختلف دنیا به لحاظ تعداد فرزندان در این بیست سال به هم نزدیک شده است ؟

با توجه به نتایج حاصله ، آیا می توان گفت که جهان در حال حرکت به سمت فرهنگ واحد در این زمینه است؟

آ) جدول نرخ زاد و ولد برای سال ۲۰۱۰

٤.٦	٢.٤	٤	٢.١	٢.٧	٥.١	١.٦	٣.٤	١.٩	٥.٧
٣	١.٩	١.٩	٤.٢	٢.٩	١.٨	٢.٤	١.٢	١.٩	١.٧
٢	٤.٢	٤.٢	٥.٨	٢.٣	٤.٩	٤.٦	٢.٨	٢.١	٢.٨
١.٢	٢.١	٢.١	١.٨	٥.١	١.٥	٦.٦	١.٨	٢.٣	٦.٢
٢.٣	٣	٣	١.٤	٥	١.٥	١.٤	٢.١	١.٨	٢.١
٢.٧	١.٤	١.٤	٤.١	١.٦	١.٥	١.٩	١.٥	١.٤	٢.٢
٣.١	٢.٣	٢.٣	١.٥	٤.٩	١.٥	١.٧	٥.٩	١.٨	١.٧
٣.٣	١.٤	١.٤	٢.٢	٢.٧	١.٩	٢.٤	٦.٣	٢.٨	١.٧
١.٢	٣.٥	٣.٥	٢.٢	١.٩	٣.٦	٤.٩	٣	٥.١	١.٩
١.٥	٢.٦	٢.٦	٢.٥	٢	٢.٦	٦.٣	٥	٢.٤	١.٤

ب) جدول نرخ زاد و ولد برای سال ۱۹۹۰

٦	٣.١	٥.٦	٣.٤	٣.٨	٥.٣	١.٨	٤.٩	٢.٧	٧.٧
٤.٦	٤.٨	٦.٦	٥.٤	٤.٤	٣.٢	٥.٣	١.٧	٢.٦	٣
٢.٣	٥.٩	٦.٦	٦.١	٤	٦.٤	٥.٨	٤.٧	٣.٧	٤.٨
١.٦	٢.١	٢.٥	٢.٢	٥.٩	١.٦	٧.٣	٢.٨	٤.٦	٧.٢
٣.٩	٢.٨	٥.٤	١.٥	٦.٥	١.٨	١.٥	٣.٥	١.٧	٢.١
٢.٤	١.٣	٥.١	٥.٦	٢	٢.٤	٢.٦	١.٨	١.٩	٣
٣.٧	٢.٩	١.٣	١.٤	٧.٢	١.٩	٢.٥	٧	١.٦	٢.٥
٦.٢	١.٥	١.٨	٢.٤	٣.٤	١.٧	٣.١	٧.٥	٤.٥	٢.٢
٢	٥.٥	٢.٣	٣.٨	١.٨	٦.١	٥.٦	٥.٦	٦.٧	١.٩
٣	٢.٧	٣.٩	٣	١.٨	٣.٥	٧.١	٦.٤	٥.٦	١.٥

توجّه : در صورت دسته بندی ، داده ها را با فاصله ی (۰.۹) دسته بندی کنید.

پاسخ:

همانطوریکه در صورت مسئله، مشخص است ، هدف از سنجش میزان پراکندگی و نزدیکی داده ها به یکدیگر است ، فلذا برای یافتن پاسخ ، باید مراحل محاسباتی انحراف معیار را ، بصورت مجزّا برای هر یک از جداول ، محاسبه و نتایج حاصله را مورد بررسی و تفسیر قرار داد:

آ) محاسبه ی انحراف معیار برای سال ۱۹۹۰ :

C.L.	Fi	mi	$mi^{۲}$	Fi mi	$Fi\,(mi)^{۲}$
۰ — ۰.۹	۰	۰.۴۵	۰.۲	۰	۰
۱ — ۱.۹	۲۲	۱.۴۵	۲.۱	۳۱.۹	۴۶.۲
۲ — ۲.۹	۲۱	۲.۴۵	۶	۵۱.۴۵	۱۲۶
۳ — ۳.۹	۱۷	۳.۴۵	۱۱.۹	۵۸.۶۵	۲۰۲.۳
۴ — ۴.۹	۹	۴.۴۵	۱۹.۸	۴۰.۰۵	۱۷۸.۲
۵ — ۵.۹	۱۴	۵.۴۵	۲۹.۷	۷۶.۳	۴۱۵.۸
۶ — ۶.۹	۱۰	۶.۴۵	۴۱.۶	۶۴.۵	۴۱۶
۷ — ۷.۹	۷	۷.۴۵	۵۵.۵	۵۲.۱۵	۳۸۸.۵
C = ۰.۹	N = ۱۰۰			$\sum Fi\,mi = ۳۷۵$	$\sum Fi\,(mi)^{۲} = ۱۷۷۳$

$$\boldsymbol{S\ 1990 = \sqrt{\frac{1773 - \frac{(140625)}{100}}{100}} = 1.9}$$

ب) محاسبه ی انحراف معیار برای سال ۲۰۱۰ :

C.L.	Fi	mi	$mi^{۲}$	Fi mi	$Fi\,(mi)^{۲}$
۰ — ۰.۹	۰	۰.۴۵	۰.۲	۰	۰
۱ — ۱.۹	۳۶	۱.۴۵	۲.۱	۵۲.۲	۷۵.۶
۲ — ۲.۹	۳۱	۲.۴۵	۶	۷۵.۹۵	۱۸۶
۳ — ۳.۹	۱۰	۳.۴۵	۱۱.۹	۳۴.۵	۱۱۹
۴ — ۴.۹	۹	۴.۴۵	۱۹.۸	۴۰.۰۵	۱۷۸.۲
۵ — ۵.۹	۱۰	۵.۴۵	۲۹.۷	۵۴.۵	۲۹۷
۶ — ۶.۹	۴	۶.۴۵	۴۱.۶	۲۵.۸	۱۶۶.۴
۷ — ۷.۹	۰	۷.۴۵	۵۵.۵	۰	۰
C = ۰.۹	N = ۱۰۰			$\sum Fi\,mi = ۲۸۳$	$\sum Fi\,(mi)^{۲} = ۱۰۲۲.۲$

تفسیر :

همانگونه که مشاهده می کنیم ، انحراف معیار داده ها در سال ۲۰۱۰ ، نسبت به سال ۱۹۹۰ ، افزایش چشمگیری را شاهد می باشد ، فلذا نتیجه می گیریم که فاصله ی فرهنگی در خصوص مسئله ی زاد و ولد در میان کشورها ، بیشتر شده است و نتیجتا ، جهان به سمت یک فرهنگ واحد حرکت نمی کند .

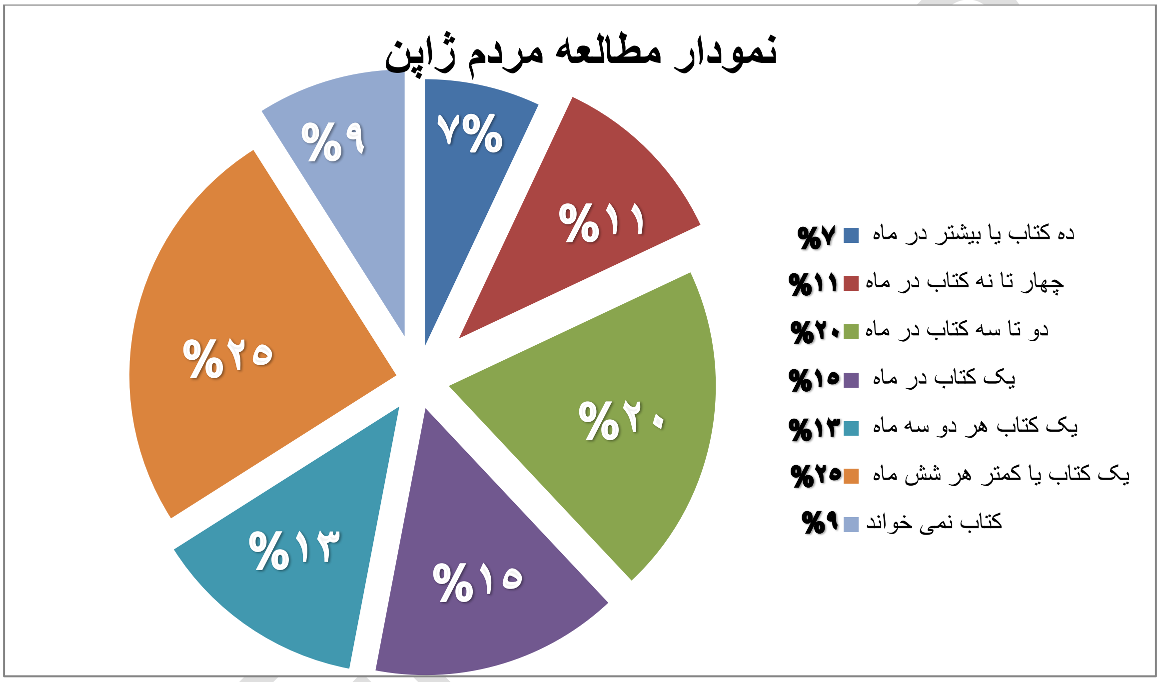
$$\boldsymbol{S\ 2010 = \sqrt{\frac{1022.2 - \frac{(8089)}{100}}{100}} = 3.07}$$

سوال ۱۶)

نمودار ذیل ، مردم ژاپن را به لحاظ میزان مطالعه دسته بندی کرده است .

آمار هر یک از این دسته ها را از روی نمودار بیان کنید.

چه نتیجه یا نتایجی از این نمودار بدست می آورید؟



آ) آمار دسته ها:

۱۰ کتاب یا بیشتر در ماه : ۷ ٪ ؛ **۱ کتاب تا ۹ کتاب در ماه : ۱۱ ٪** ؛ **۲ تا ۳ کتاب در ماه : ۲۰ ٪**

۱ کتاب در ماه : ۱۵ ٪ ؛ **۱ کتاب هر دو سه ماه : ۱۳ ٪** ؛ **۱ کتاب یا کمتر هر شش ماه: ۲۵ ٪**

کتاب نمی خواند : ۹ ٪

ب) نتیجه و تفسیر:

همانگونه که آمارها نشان می دهند ، می توان گفت حدود (۳۴٪) مردمان ژاپن ، رغبتی به مطالعه ی کتاب ندارند چرا که (۲۵٪) آنان در شش ماه ، نهایتا یک کتاب می خوانند و (۹ ٪) آنان ، اصلا کتاب نمی خوانند ،

در ۶۶٪ باقی مانده ی جمعیت مورد مطالعه نیز ، حدود (۱۳٪) درصد آنان ، میل به مطالعه ی کمی دارند ، حال چنانچه میزان مطالعه ی حداقل یک کتاب در ماه را ، مطالعه ی مورد قبول تلقی کنیم ، مشاهده می کنیم که بالای نیمی از مردمان ژاپن ، سطح مطالعه ی قابل قبولی دارند (۵۳٪) و در این ۵۳ درصد ، به نسبت ، شاهد میزان مطالعه ی بسیار خوبی هستیم بطوریکه در ۷٪ این مقدار، میزان مطالعه ی بسیار عالی را می توان دید ؛ لازم به ذکر است که چنین مسائلی را باید با شاخص های جهانی سنجید و نتیجه گیری کرد و از آنجایی که بدین علّت ، مسئله دارای نقص است و با کمبود اطلاعات مواجه ایم ، نمی توان تفسیر دقیقی ارائه نمود.

به پایان آمد این دفتر

حکایت همچنان باقیست

پایان جزوه ی آمار مقدماتی

دکتر عالمی نیسی