آمار در کتابداری و اطلاع‌رسانی

دکتر پرویز نصیری

دکتر ثریا ضیایی

دکتر باقر مقدس‌زاده

فهرست

پیوست‌ها

پیشگفتار

علم آمار امروزه در زمرهٔ علوم و فنون جدیدی است که اخیراً به‌وجود آمده است. شیوه‌ها و روش‌های آن به صورت علمی بسط و تعمیم یافته و علمی است که در تمام رشته‌های علوم اعم از نظری یا عملی مورد استفاده قرار می‌گیرد. رشتهٔ کتابداری و اطلاع‌رسانی یکی از رشته‌هایی است که نیاز مبرم دارد از روش‌های آماری برای معرفی کتاب‌سنجی و علم‌سنجی استفاده کند. لذا با مطالعهٔ سرفصل‌های رشتهٔ کتابداری و اطلاع‌رسانی بر آن شدیم که کتاب پیش رو را به رشتهٔ تحریر درآوریم. بنابراین، سعی شده است با ارائهٔ تئوری علم آمار به صورت مختصر، کاربرد آن با ذکر مثال‌های مرتبط با رشتهٔ کتابداری و اطلاع‌رسانی آورده شود. در تجزیه و تحلیل داده‌ها از نرم‌افزار آماری SPSS استفاده شده است. برای آشنایی با نرم‌افزار آماری SPSS در ضمیمه، یک نسخهٔ آن به همراه محاسبات آورده شده است.

این کتاب در شش فصل تنظیم شده است. که در ادامه، مطالب هر یک از فصول به اختصار توضیح داده می‌شود.

عنوان فصل اول "کلیات و مفاهیم اولیه" است. در این فصل، سعی بر این است که شرح مختصری از واژهٔ کتاب‌سنجی ارائه و در بخش‌های بعدی، مفاهیم اولیهٔ آمار توصیفی به همراه مقیاس‌های اندازه‌گیری آورده شود.

عنوان فصل دوم "طبقه‌بندی داده‌های آماری و شاخص‌های آماری" است. در این فصل، جدول توزیع فراوانی، انواع نمودارهای آمار و نحوهٔ محاسبهٔ شاخص‌های مرکزی و غیرمرکزی ارائه شده است.

عنوان فصل سوم "متغیر تصادفی و توزیع آماری" است. در این فصل، پس از تعریف متغیر تصادفی، توزیع آماری از نوع گسسته و پیوسته آورده می‌شود و در پایان، توزیع‌های نرمال، نرمال استاندارد و توزیع استودنت معرفی می‌شوند.

عنوان فصل چهارم "آزمون فرضیهٔ آماری" است. در این فصل پس از آشنایی با مفاهیم اولیهٔ آزمون فرض، فرض‌های آماری دربارهٔ میانگین جامعهٔ نرمال و نسبت جامعه ارائه و آزمون می‌شود.

عنوان فصل پنجم "همبستگی و انواع آن" است. در این فصل، ضمن تعریف همبستگی دو متغیر تصادفی، انواع ضرایب همبستگی مورد بحث قرار می‌گیرد.

عنوان فصل ششم "رگرسیون و اصول پیش‌بینی" است. در این فصل، پس از آشنایی با نمودار پراکنش و خط رگرسیون، معادلهٔ خط رگرسیون در نمونهٔ برازش و براساس آن متغیر پاسخ پیش‌بینی می‌شود.

امیدواریم این کتاب قابل استفاده خوانندگان باشد و مورد علاقهٔ آنان قرار گیرد، و آرزومندیم خوانندگان به‌ویژه دانشجویان، نظرها و پیشنهادهای خود را برای مؤلفان ارسال دارند تا در تجدیدنظر چاپ‌های بعدی مورد استفاده قرار گیرند.

در پایان از سرکار خانم فاطمه محمدی که حروفچینی و صفحه‌آرایی کتاب را بر عهده داشتند تشکر می‌کنیم.

پرویز نصیری
دانشیار دانشگاه پیام نور
تابستان ۱۳۹۵

فصل اول

کلیات و مفاهیم اولیه

هدف کلی

آشنایی با مفاهیم اولیهٔ کتابداری و آمار توصیفی.

هدف‌های یادگیری

پس از مطالعهٔ این فصل، شما باید بتوانید:

۱. با واژهٔ کتاب‌سنجی آشنا شوید.

۲. با مهم‌ترین مشخصه‌های کتاب‌سنجی آشنا شوید.

۳. با مطالعات آماری بین‌المللی در حوزه‌های سنجشی آشنا شوید.

٤. با مفاهیم اولیهٔ آمار توصیفی آشنا شوید.

۵. با مقیاس‌های اندازه‌گیری آشنا شوید.

مقدمه

کمپبل[1] در سال ۱۸۹٦ با مطالعهٔ پراکندگی موضوعی انتشارات به کمک روش‌های آماری، اقدام به معرفی کتاب‌سنجی[2] کرد. بعد از ایشان تعداد زیادی از نویسندگان ضمن تأیید انواع مطالعات کتاب‌سنجی توسط کمپبل سعی کردند که با محاسبهٔ تعداد مدارک علمی، چشم‌اندازی در تاریخ علم فناوری ایجاد کنند. واژهٔ قدیمی‌تر کتاب‌سنجی، کتابشناسی آماری بود که در سال ۱۹٤۸ رانگاتان[3] واژهٔ کتاب‌سنجی را به کاربرد و آن را ترکیبی از روش‌های کمّی قابل استفاده برای مدیریت و خدمات کتابخانه معرفی کرد [نوروزی چالکی]. اکنون می‌توان ادعا کرد که نزدیک به چهار دهه است که بوفور از واژهٔ کتاب‌سنجی استفاده می‌شود.

هم‌اکنون مطالعات کتاب‌سنجی در سطح وسیع در کتابخانه‌ها و جامعهٔ علمی و پژوهشی به منظور سنجش و ارزیابی انتشارات علمی استفاده و روزبه‌روز نیز بر دامنهٔ فنون و روش‌های آن افزوده می‌شود. البته نباید از تأثیر قوانینی که در قرن بیستم مطرح شد و کتاب‌سنجی را توسعه داد غافل بود. در همین زمینه باید به قانون زیپف[4] اشاره کرد که برای نخستین بار به مطالعهٔ مشخصه‌ها و فرآیندهای مرتبط با مدارک پرداخت که بر توسعهٔ کتاب‌سنجی تأثیر بسزایی گذاشت. او به دلیل معرفی "اصل کم‌ترین کوشش" خود، بسیار معروف شد. در واقع اصل کم‌ترین کوشش به این معناست که یک شخص می‌کوشد مشکلات فوری و به احتمال، دشواری‌های آینده‌اش را از طریقی که به کم‌ترین تلاش نیاز داشته باشد، حل کند. زیپف در سال ۱۹٤۹ از اصل کم‌ترین کوشش برای توصیف نرخ متوسط کم‌ترین کار احتمال استفاده کرد و بر همین مبنا،

---

۱. campbell
۲. Bibliometrice
۳. Rangatan
٤. Zipf

اهمیت کوتاه‌نویسی مقاله‌ها را از طریق به کارگیری واژه‌های کم‌تر بـرای مفهـوم، مورد تأکید قرار داد. اصل زیپف که در ابتدا برای هدایت رفتار اطلاع‌یابی استفاده می‌شد، به طور غیرمستقیم به فرمول زیپف که از مدل قانون توان[1] در زبان‌شناسی کمی برگرفته شده است ربط داده می‌شود؛ به طوری که در زبان‌شناسی کمّی مدل فرضی قانون– توان به صورت زیر ارائه می‌شود:

$$r f = C$$

که در آن r رتبهٔ هر واژه و f فراوانی رخداد هر واژه و C مقدار ثابت است. در واقع حاصل‌ضرب رتبهٔ هر واژه و فراوانی رخداد آن، مقداری است ثابت. قابل ذکر است علاوه بر قانون زیپف می‌توان از قوانین لوتکا، بردفورد[2] و قانون پارتو[3] نیز در مقام زیربناهای فکری کتاب‌سنجی نام برد. قدر مسلم آنکه این قوانین به گونه‌ای با زبان ریاضی بیان شده‌اند که می‌توان آن‌ها را در محیط واقعی به کار گرفت. قابل ذکر است که قانون زیپف فراوانی مشاهده شدهٔ واژگان در یک متن را بررسی می‌کند؛ در حالی که قانون لوتکا[4] بهره‌وری نویسندگان را برحسب انتشارات علمی‌شان مورد مطالعه قرار می‌دهد. در حالی‌که بردفورد، به مطالعهٔ پراکندگی مقاله‌های مجله‌های گوناگون می‌پردازد.

۱–۱ حوزهٔ فعالیت‌های کتاب‌سنجی و کاربرد آمار

در سال ۱۹٦۹ پریچارد[5] نخستین تعریف رسمی را از واژهٔ کتاب‌سنجی ارائه کرد که عبارتست از:

---

۱. Power –law model
۲. bradford
۳. Paretos rule
٤. Lotka
٥. Pritchard

"کاربرد روش‌های ریاضی و آمار برای کتاب‌ها و دیگر رسانه‌های ارتباطی". سن‌گوپتا[1] ریشهٔ کتاب‌سنجی را از معادل فارسی دو واژه کتاب و سنجش می‌داند. ریشهٔ واژهٔ کتاب را از کلمه بیبلیون[2] که خود دارای ترکیبی لاتین – یونانی است و معادل واژهٔ بیبل یا بیبلُس[3] به معنی کتاب یا کاغذ است معرفی می‌کند. همچنین وی واژهٔ سنجش را با مفهوم اندازه‌گیری یکی گرفته و براین باور است که ریشهٔ آن به واژهٔ یونانی "مترکِلوس[4]" یا کلمهٔ لاتین متریکیوس[5] به معنای اندازه‌گیری بازمی‌گردد. در سال ۲۰۰۲ بورگمن و فرنر[6] کتاب‌سنجی را به طور ساده، اندازه‌گیری مشخصه‌ها و فرایندهای مرتبط با مدارک تعریف می‌کنند که از جملهٔ این مشخصه‌ها و فرآیندهای مرتبط با مدارک، می‌توان به موارد زیر اشاره کرد:

الف– تحلیل بسامد یا فراوانی واژه‌ها

ب– تحلیل هم‌واژه‌ها

ج– تحلیل استنادی

د– محاسبهٔ سادهٔ مدارک (مثلاً تعداد آثار منتشر شدهٔ علمی یک نویسنده یا یک گروه پژوهشی یا یک کشور)

قابل ذکر است که کتاب‌سنجی از تحلیل استنادی استفاده می‌کند و به آن توجهی خاص نشان می‌دهد؛ زیرا همواره این باور وجود دارد که یک دانشمند به دنبال شناسایی و مطالعهٔ مقاله‌های مرتبط است. بنابراین شناسایی اینکه کدام مقاله به یکدیگر استناد کرده‌اند، می‌تواند برای او کارساز باشد و به وی در شناسایی مرتبط‌ترین مقاله‌هایی که در یک حوزه وجود دارند، کمک کند. از همین رو است که

---

١. San gupta
٢. Biblion
٣. Biblos
٤. metrokos
٥. metricus
٦. Borgman and furner

نمایه‌های استنادی[1] در مقام ابزاری قدرتمند برای انجام تحلیل‌های استنادی در کتاب‌سنجی اهمیتی ویژه دارد. در واقع، نمایه‌های استنادی علاوه بر بهره‌گیری از تفکر تحلیل استنادی، تا حدّ زیادی با قانون پراکندگی بردفورد نیز سازگارند. افزون بر استفاده از تحلیل‌های استنادی، استفاده از شاخص‌ها و فنون دیگر نیز ضرورت دارد تا بتواند نتایج متنوع‌تر، معنادارتر و مفیدتر را برای سیاست علم فراهم کند. با این وجود، امروزه نیز به طور دائم، فنون مکمل برای کتاب‌سنجی پا به عرصه وجود می‌گذارند. انقلابی که با ورود وب ایجاد شده، بر این امر دامن زد و کتاب‌سنجی را نیز تحت تأثیر خود قرار داد. استفاده از شاخص‌های مبتنی بر استناد، هنوز هم هستهٔ اصلی این نوع مطالعات را تشکیل می‌دهد. گذشته از روش‌های اصلی تحلیل استنادی، بزرگ‌ترین تحول در کتاب‌سنجی از محیط ارتباطات علمی و دسترس‌پذیری به منابع اطلاعاتی جدید و مهمی نظیر پروانه‌های ثبت اختراعات، صفحه‌های وب و آمارهای استفاده از کتابخانه‌ها نشأت گرفته است.

البته باید توجه داشت که کتاب‌سنجی، هرگز تنها به مقاله‌های علمی وابسته نبوده و سایر منابع علمی نیز همواره مورد توجه پژوهشگران این حوزه قرار داشته‌اند.

۱-۲ مهم‌ترین مشخصه‌های کتاب‌سنجی

مطالعات کتاب‌سنجی با هدف توسعهٔ خدمات کتابخانه‌ها و مراکز اطلاع‌رسانی انجام می‌شوند. در حقیقت، کتابداران همیشه فنون کتاب‌سنجی را در امور گوناگون کتابخانه‌ای به کار می‌گرفتند و از دیرباز در امور کتابخانه‌ای رایج بوده است. کتاب‌سنجی سرمنشأ بسیاری از حوزه‌های سنجشی از جمله علم‌سنجی، اطلاع‌سنجی و وب‌سنجی است. بنابراین، کتاب‌سنجی از نمایه‌های استنادی در

---

۱. Citation Indexes

جایگاه یک ابزار استفاده می‌کند و به طور مستقیم روش‌های ریاضی و آمار را برای دستیابی به اهداف خود به کار می‌برد. به این ترتیب، کتاب‌سنجی حوزه‌ای است به طور کامل کمّی که با تکیه بر روش‌های ریاضی و آمار، به سنجش و ارزیابی متون می‌پردازد.

۱-۳ کتاب‌سنجی نوین در ایران

در ایران عده‌ای از صاحب‌نظران کتابداری و اطلاع‌رسانی، توجه به بنیان‌های تاریخی و مفهومی کتاب‌سنجی را در اوایل دهۀ ۱۳٦۰ (مطابق ۱۹۸۰ میلادی) مطرح کردند که می‌توان به مقالات دیانی (۱۳٦۱)، حری (۱۳٦۲)، دیانی و عصاری (۱۳٦۷)، مهراد (۱۳۷۲) و عصاری (۱۳۷٦ الف، ۱۳۷٦ ب و ۱۳۷۷) اشاره کرد. در سال ۱۳۸۹ منصوریان با هدف شناسایی مهم‌ترین تحقیقات کتاب‌سنجی و علم‌سنجی در ایران، بدون جدایی این دو حوزه و با شناسایی چند محور پژوهشی فرعی، کتاب‌سنجی و علم‌سنجی در ایران را در چهار گروه کلی زیر دسته‌بندی کرد:

۱- مطالعۀ ارتباطات علمی و تحلیل استنادی
۲- ارزیابی کمّی و کیفی منابع و انتشارات علمی
۳- سنجش برونداد، بازدهی و تأثیرگذاری علمی
٤- تدوین سیاست‌ها و خط‌مشی‌های علمی و پژوهشی

به لحاظ تاریخی، واژۀ اطلاع‌سنجی[۱] در مقایسه با حوزه‌های کتاب‌سنجی و علم‌سنجی، قدمتی کم‌تر دارد و اولین بار در سال ۱۹۷۹ در کشور آلمان از این واژه استفاده شد.

---

۱. Informetrics

ناکه[1]، از اطلاع‌سنجی، برای پوشش‌دادن آن بخش از علم اطلاعات که با اندازه‌گیری پدیدهٔ اطلاعات و کاربرد روش‌های ریاضی برای حل مسائل رشته‌ها سر و کار دارد، استفاده کرد و آن را در حوزه‌هایی مانند کتاب‌سنجی، در بخش‌هایی از نظریهٔ بازیابی اطلاعات و حتی شاید در سطحی بسیار وسیع‌تر قابل استفاده دانست (نقل در: هود و ویلسون، ۲۰۰۱: ۲۹٤).

ناکه در سال ۱۹۸٤ و در شرایطی که هنوز اطلاع‌سنجی، واژه‌ای برگرفته از کتاب‌سنجی و علم‌سنجی محسوب می‌شد، «مؤسسهٔ مشترک اطلاعات علمی و فنی (وینیتی)[2]»، «کمیتهٔ اطلاع‌سنجی»[3] را در «فدراسیون بین‌المللی دبیزش (فی)»[4]، تأسیس کرد.

بروکس[5] (۱۹۸۸) در «نخستین کنفرانس بین‌المللی کتاب‌سنجی و جنبه‌های نظری بازیابی اطلاعات[6]»، اطلاع‌سنجی را به لحاظ مفهومی به دو حوزهٔ کتاب‌سنجی و علم‌سنجی تقسیم کرد و آن را برای ارزیابی هر دو نوع اطلاعات الکترونیکی و اطلاعات سنتی قابل استفاده دانست؛ با وجود این، در اوایل دههٔ ۱۹۹۰ میلادی، مطالعات اطلاع‌سنجی، گستردگی بیشتری یافت. بروکس (۱۹۹۰) در «دومین کنفرانس بین‌المللی کتاب‌سنجی و جنبه‌های نظری بازیابی اطلاعات[7]»، باز هم، اطلاع‌سنجی را اصطلاحی اعم از علم‌سنجی و کتاب‌سنجی معرفی و در عین‌حال، بر این نکته تأکید کرد که علم‌سنجی بیشتر به سوی مطالعات سیاستی گرایش دارد و حال آنکه کتاب‌سنجی بیشتر در راستای کاربردهای کتابخانه‌ای استفاده می‌شود؛ علاوه بر آن، تگ- ساتکلیف[8] (۱۹۹۲) نیز در همان کنفرانس،

---

۱. Nacke
۲. All-union Institute for Scientific and Technical Information (VINITI)
۳. Committee of Informetrics
٤. Federation Internationale de le Documentation(FID)
٥. Brookes
٦. Fist international conference on bibliometrics and theoretical aspects aspects of information retrieval
۷. Second international conference on bibliometrics and theoretical aspects of information retrieval
۸. Tague- Sutcliffe

حوزهٔ اطلاع‌سنجی را یکی از حوزه‌های مستخرج از کتاب‌سنجی و علم‌سنجی معرفی کرد.

با این حال، در سومین مجموعهٔ مقالات آن کنفرانس که در سال ۱۹۹۲ (با عنوان سومین کنفرانس بین‌المللی اطلاع‌سنجی[1]) برگزار شد، جایگاه اطلاع‌سنجی، استحکام و قوامی بیشتر یافت؛ هر چند، در «چهارمین کنفرانس بین‌المللی کتاب‌سنجی، اطلاع‌سنجی و علم‌سنجی[2]» تعداد کم‌تری مقاله در زمینهٔ اطلاع‌سنجی ارائه شد و بیشتر مقاله‌ها به حوزه‌های علم‌سنجی و کتاب‌سنجی اختصاص داشتند؛ در این میان، سه جلد از چهار جلد مجموعهٔ مقالات چهارمین کنفرانس بین‌المللی کتاب‌سنجی، اطلاع‌سنجی و علم‌سنجی، به طور کامل به مسائل مجله‌های انگلیسی‌زبان اختصاص یافت (گلنزل؛ کرتشمر، ۱۹۹٤ab، ۱۹۹۲)[3]. در سال ۱۹۹۲ و در همان کنفرانس، «انجمن بین‌المللی علم‌سنجی و اطلاع‌سنجی (آی.اس.اس.آی)»[4] تأسیس شد؛ این انجمن از آن پس، هر دو سال یک‌بار «کنفرانس بین‌المللی آی.اس.اس.آی». را برگزار کرد؛ علاوه بر آن، ظهور مجلهٔ «مدیریت و پردازش اطلاعات[5]» در سال ۱۹۹۲ را باید یکی دیگر از مهم‌ترین حرکت‌هایی دانست که در حوزهٔ اطلاع‌سنجی به انجام رسید. در مجموع، باید یادآوری شود که اوایل دههٔ ۱۹۹۰، اصطلاح اطلاع‌سنجی در سطحی گسترده‌تر شناخته شد و پس از آن تداوم یافت (نقل در: هود و ویلسون، ۲۰۰۱: ۲۹٤).

اطلاع‌سنجی اغلب با تولید اطلاعات یا فرآیندهای تولید اطلاعات[6] سر و کار دارد. آن‌ها مصداق‌های اصلی اطلاعات را «منبع»[7] و «فقرهٔ اطلاعاتی»[8] نامیدند

---

۱. Third international conference on informetrics
۲. International conference on bibliometrics, informetrics and scientometrics
۳. GlÄnzel, Kretschmer
٤. Institute for scientific Information
٥. Information processing & management
٦. Information production processes (IPPs)
۷. Source
۸. Item Information

و برای نمونه، مجله را به منزلهٔ «منبع» و مقاله را به منزلهٔ «فقره‌های اطلاعاتی» در نظر گرفتند؛ بنابراین از نظر آن‌ها، اطلاع‌سنجی با اثبات روابط میان منابع و فقره‌های اطلاعاتی و توصیف قواعد موجود در الگوهای تولید اطلاعات سر و کار دارد. در این میان، دایادتو[۱] (۱۹۹٤) مباحث موجود در اطلاع‌سنجی را به سه قسمت تقسیم کرده است:

۱. توزیع‌های کتاب‌سنجی/ اطلاع‌سنجی
۲. تحلیل‌های استنادی
۳. شاخص‌های اطلاع‌سنجی

از سویی دیگر، اِرار (۲۰۰۰)[۲] به کارگیری روش‌های آماری پیشرفته و برخی روش‌های پژوهش عملیاتی و ریاضی را در حوزهٔ علوم کتابداری و اطلاع‌رسانی، به عنوان اطلاع‌سنجی معرفی می‌کند.

۱-٤ مطالعات بین‌المللی در حوزه‌های سنجشی

برای مشخص کردن وضعیت کنونی و گذشتهٔ کتاب‌سنجی و علم‌سنجی از دیدگاه‌های مختلف، جستجو در پایگاه‌های اطلاعاتی سه‌گانهٔ آی.اس.ای تحت SCI، SSCI، HCI و H از طریق وب سایت علوم (Web of Science) امکان‌پذیر است؛ به‌طور نمونه می‌توان به مؤلفان، کشورها، مؤسسه‌ها، زبان، منبع و موارد دیگر اشاره کرد. از آنجایی که رکوردها براساس پارامترهای مختلف از جمله مؤلفان، کشورها، مؤسسه‌ها، زبآن‌ها، منبع و موارد دیگر در پایگاه‌ها تعبیه شده‌اند، می‌توان با استفاده از نرم‌افزار تحلیلی، اطلاعات را مورد تجزیه و تحلیل قرار داد. برای نمونه می‌توان ده نویسندهٔ پرکار، ده کشور برتر، ده دانشگاه یا مؤسسهٔ برتر،

---

۱. Dayadto
۲. Arrar

ده زبان مورد استفاده و یا ده مجلهٔ برتر در تولید مدارک کتاب‌سنجی را به صورت گزارش تهیه کرد.

جدول ۱-۱: جدول توزیع ده نویسندهٔ پرکار در تولید مدارک کتاب‌سنجی

ردیف	نام نویسندگان	تعداد
۱	GALANZEL, W	٤۱
۲	MOED, HF	۳۰
۳	KOSTOFF, RN	٤۰
٤	VAN RAAN, AFJ	۲۱
٥	LEWISON, G	۱۹
٦	BORDONS, M	۱۸
۷	THELWAL, M	۱۸
۸	VAN RAAN, AFJ	۱۸
۹	GOMEZ, I	۱٥
۱۰	NEDEFHOF, AJ	۱٥

جدول ۱-۲: جدول توزیع ده کشور برتر در تولید مدارک کتاب‌سنجی

ردیف	نام کشورها	تعداد مدارک
۱	ایالات متحده آمریکا	٤٥۳
۲	اسپانیا	۲۲۷
۳	انگلستان	۱٥٦
٤	هلند	۱۳٦
٥	آلمان	۱۱۷
٦	ایتالیا	۷۱
۷	کانادا	۷۱
۸	تایوان	٦۷
۹	فرانسه	٦۷
۱۰	بلژیک	٦۷

جدول ۱-۳: جدول توزیع ده دانشگاه یا مؤسسهٔ برتر در تولید مدارک کتاب‌سنجی

ردیف	نام دانشگاه‌ها یا سازمان‌ها	تعداد مدارک
۱	LEIDEN UNIV	۷۸
۲	CSIC	٤۳
۳	INDIANA UNIV	۳٤
٤	UNIV GRANADA	۳۱
٥	HUNGARIAN ACAD SCI	۲۸
٦	OFF NAVAL RES	۲۸
۷	KATHOLIEKE UNIV LEUVEN	۲٥
۸	UNIV SUSSEX	۲٥
۹	CITY UNIV LONDON	۲۳
۱۰	UNIV VALENCIA	۲۲

جدول ۱-٤: جدول توزیع فراوانی زبان‌های مورد استفاده در تولید مدارک کتاب‌سنجی

ردیف	زبان	تعداد مدارک
۱	انگلیسی	۱٥۰۳
۲	اسپانیایی	۱۲۱
۳	آلمانی	٤٥
٤	فرانسوی	۱۹
٥	روسی	۸
٦	ایتالیایی	۳
۷	ژاپنی	۳
۸	کرواتی	۲
۹	قزاقستانی	۲
۱۰	صربی-کرواتی	۲
۱۱	پرتغالی	۲

جدول ۱-۵: جدول توزیع فراوانی مجلات چاپ‌کنندۀ مدارک مرتبط با کتاب‌سنجی

ردیف	عنوان مجلات	تعداد
۱	SCIENTOMETRIC	۱۱۱
۲	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	۲۸
۳	RESEARCH EVALUATION	۱۳
٤	INFORMATION PROCESSING & MANAGEMENT	۱۲
٥	TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE	۱۱
٦	SOCIAL WORK IN HEALTH CARE	۹
۷	JOURNAL OF DOCUMENTATION	۸
۸	JOURNAL OF INFORMATION SCIENCE	۸
۹	TECHNOVATION	۷
۱۰	EUROPEAN JOURNAL OF PUBLIC HEALTH	٦

۱-۵ مفاهیم اولیۀ آمار توصیفی

آمار، علمی است که مشخصات یا خصوصیات جامعه‌های آماری را با توجه به شرایط کیفی مربوط، مورد مطالعه قرار می‌دهد. برای ارتقای توانایی درتوصیف وتفسیر انبوهی از اطلاعات پژوهش‌های علوم مختلف، به آمار نیازمندیم. آمار توصیفی که شاخه‌ای از آمار است، ابزاری برای توصیف انبوهی از داده‌هاست. گردآوری اطلاعات و ارائۀ روش دقیق آن‌ها کمک مهمی در فهم و تفسیر صحیح آن‌ها است. برای ارائۀ مشاهدات کمّی معمولاً از دو روش استفاده می‌شود. یک روش، شامل نشان دادن خلاصه‌شده‌ای از اطلاعات است که معمولاً در قالب جداول آماری ارائه می‌شود. روش دیگر، مشاهدات را در قالب تصویر به صورت نمودارها و یا دیگر شکل‌های مشابه نشان می‌دهد. لذا در این بخش، پس از ارائۀ تعریفی از جامعه، نمونه، روش‌های نمونه‌گیری، انواع متغیرها و مقیاس‌های اندازه‌گیری شرح داده می‌شود.

۱-۵-۱ جامعه[۱]

جامعه عبارت است از گروهی از افراد، اشیاء یا حوادث که حداقل یک صفت یا ویژگی مشترک دارند و به خاطر همان صفت، مورد بررسی قرار می‌گیرند. مفهوم جامعه از نظر آماری خیلی وسیع‌تر از مفهوم واژهٔ آن است. مانند ساکنان کلانشهرها، کلیهٔ کتب در کتابخانه‌های کشور و یا دانشجویان دانشگاه‌ها. برای انجام هر کار آماری روی یک جامعه باید آن جامعه و ویژگی مورد مطالعه، قبلاً بدون هرگونه ابهام مشخص شوند. به عنوان مثال، جامعهٔ دانشجویان دانشگاه پیام‌نور مشهد و میزان استفاده از کتابخانه.

جامعهٔ آماری ممکن است، با پایان (محدود) باشد. یعنی قادر به شمارش اعضای آن جامعه باشیم و یا بی‌پایان (نامحدود) باشد که در این‌صورت قادر به شمارش اعضای آن جامعه نیستیم. مثلاً دانشجویانی که طی دو سال گذشته در دانشگاه پیام نور ثبت نام کرده‌اند با پایان است، ولی جامعهٔ افرادی که از مهرماه سال آینده به بعد در دانشگاه پیام نور ثبت نام خواهند کرد، بی‌پایان است.

نمایش اعضای جامعه برای گرفتن نمونه و یا انجام چهار عمل اصلی بین آن‌ها نقش مهمی در استنباط آماری دارد. در حالتی که جامعهٔ آماری محدود یا متناهی و دارای حجم N باشد، اعضای جامعه را به صورت زیر نمایش می‌دهند:

$$X_۱, X_۲, X_۳, \ldots, X_N$$

که در آن $X_۱$ عضو اول، $X_۲$ عضو دوم و X_i عضو iام جامعه است. برای مثال اگر اعضای انجمن کتابداری در کشور دارای ۱۰ عضو باشد، می‌توان بدون مراجعه به آن‌ها، آن‌ها را به صورت زیر نمایش داده و یا در ذهن شبیه‌سازی کرد.

---

۱. Population

$X_{۱}, X_{۲}, X_{۳}, X_{۴}, X_{۵}, X_{۶}, X_{۷}, X_{۸}, X_{۹}, X_{۱۰}$

البته به این نکته باید اشاره کرد که $X_{۱}$، نفر اول انجمن نیست، بلکه ویژگی‌ای است که می‌تواند برای یکی از ۱۰ نفر انجمن تخصیص داده شود. در حالت کلی X_i می‌تواند، سابقهٔ کار، میزان درآمد، میزان تحصیلات، سن و غیره برای نفر iام باشد.

بنابراین به اطلاعاتی که با روش‌های مختلف به‌دست می‌آید، دادهٔ آماری گفته می‌شود که در واقع، همان اطلاعات خامی است که هنوز هیچ‌گونه بررسی‌ای روی آن‌ها انجام نشده است.

گاهی جامعهٔ مورد مطالعه به قدری انبوه است که عملاً می‌توان آن را نامحدود تلقی کرد؛ مثل جامعهٔ کلیهٔ کتب منتشر شده در بازار نشر ایران طی ده سال گذشته. از آنجایی که اغلب، مطالعهٔ یکایک افراد جامعه به علت هزینه زیاد، کمی وقت، نداشتن امکانات کافی و یا از بین رفتن جامعه بر اثر بررسی، مقدور نیست، قسمتی از جامعه را به جای تمام آن در نظر می‌گیرند و به این بخش در نظر گرفته شده، نمونه می‌گوییم که در بخش بعدی ارائه می‌شود.

۱-۵-۲ نمونه[۱]

قسمتی از جامعه را که طبق ضوابطی مقبول انتخاب می‌شود و مطالعهٔ آن به جای مطالعهٔ تمام جامعه مقدور است، نمونه‌ای از جامعه می‌نامیم. معمولاً به مصداق مشت نمونهٔ خروار است نتیجه حاصل از مطالعهٔ نمونه را به تمام جامعه تعمیم می‌دهند. ولی این کار احتیاط دارد؛ زیرا هر مشتی نمی‌تواند نمونهٔ خروار باشد و قطعاً بی‌غرضی در انتخاب مشت و اندازهٔ مشت در این نمایندگی نقش مهمی دارد. مسئلهٔ انتخاب یک نمونهٔ خوب به قدری مهم است که قسمت زیادی از

---

۱. Sample

نظریه‌های آمار و احتمال به آن اختصاص دارد. اغلب برای کسب بی‌طرفی، نمونه‌هایی در نظر گرفته می‌شود که انتخاب آن‌ها کاملاً شانسی باشد. به این نوع از نمونه‌ها، نمونهٔ تصادفی می‌گوییم. اگر حجم نمونه n باشد، می‌توان اعضای نمونه را که بخشی از جامعه است به صورت زیر نمایش داد:

$$X_1, X_2, X_3, \ldots, X_n$$

که در آن X_1 عضو اول، X_2 عضو دوم و X_i عضو iام نمونه است. باید متذکر شد که X_1 که اولین نمونه است، اولین عضو جامعه نیست و می‌تواند هر کدام از $X_1, X_2, X_3, \ldots, X_N$ باشد $(n \leq N)$. همان‌طور که گفته شد، انتخاب نمونه نقش مهمی در استنباط دارد که در ادامه، انواع روش‌های نمونه‌گیری ارائه می‌شود.

۱-۵-۳ روش‌های نمونه‌گیری

لازم به توضیح است برای داشتن نمونهٔ بهتر، می‌توان از روش‌های نمونه‌گیری استفاده کرد که مهم‌ترین آن‌ها عبارتند از:

۱- نمونه‌گیری تصادفی ساده

۲- نمونه‌گیری طبقه‌ای

۳- نمونه‌گیری منظم

٤- نمونه‌گیری خوشه‌ای

در ادامه، شرح مختصری از روش‌های نمونه‌گیری ارائه می‌شود.

- در نمونه‌گیری تصادفی ساده، در حالی که حجم نمونه متناهی است، انتخاب نمونه به روش نمونه‌گیری تصادفی ساده با جایگذاری و بدون جایگذاری انجام می‌شود که در نمونه‌گیری تصادفی ساده با جایگذاری تکرار اعضای نمونه امکان‌پذیر است؛ در حالی که در نمونه‌گیری بدون جایگذاری امکان‌پذیر نیست.

- در نمونه‌گیری طبقه‌ای، از آنجایی که جامعهٔ مورد بررسی همگن نیست، معمولاً آن را به چند طبقه افراز و از هر طبقه با استفاده از نمونه‌گیری تصادفی ساده اقدام به نمونه‌گیری می‌شود.
- در نمونه‌گیری منظم یا سیستماتیک، از آنجایی که جامعهٔ مورد بررسی دارای لیست یا چارچوب است، پس از انتخاب اولین نمونه به روش نمونه‌گیری تصادفی ساده، بقیهٔ نمونه با فاصلهٔ مکانی انتخاب می‌شوند.
- در نمونه‌گیری خوشه‌ای، از آنجایی که اطلاع کامل از جامعه نیست، جامعه را به چند خوشه افراز و از بین خوشه‌ها، چند خوشه انتخاب و سرشماری روی خوشه‌های انتخاب شده انجام می‌گیرد. (برای اطلاع بیشتر به کتاب نصیری و قربانی مراجعه کنید.)

١-٥-٤ متغیر[١]

متغیر عبارت است از صفت یا صفات مورد بررسی که مقدارش از یک فرد به فرد دیگر تغییر می‌کند. مانند متغیرهای قد و وزن که از فردی به فرد دیگر متفاوت است. در برخی پژوهش‌ها و بررسی‌های آماری، تنها یک متغیر مورد بررسی قرار می‌گیرد که آن را پژوهش تک متغیری می‌نامیم. مثلاً تأثیر استفاده از نشریات جدید در افزایش تعداد مراجعان به کتابخانه. در اینجا نشریات تازه، یک متغیر است که مورد بررسی قرار گرفته است.

در مواردی که پژوهش روی بیش از یک متغیر انجام گیرد، با یک مطالعهٔ چندمتغیره مواجه هستیم. به عنوان مثال، تأثیر تغذیه، ورزش و خواب در رشد کودکان دبستانی.

در اینجا سه نوع طبقه‌بندی برای متغیر در نظر گرفته می‌شود:

---

١. Variable

الف- طبقه‌بندی اول (کمّی و کیفی)

متغیر را می‌توان به دو طبقهٔ کمّی[۱] و کیفی[۲] تقسیم کرد. متغیرهای کمّی به آن‌هایی اطلاق می‌شود که از نظر مقدار یا ارزش متفاوت هستند و مقدار آن‌ها را می‌توان پس از اندازه‌گیری، به عدد نشان داد. مانند سن، نمره، وزن، قد، هزینهٔ زندگی و بودجهٔ خرید کتاب. متغیرهای کیفی آن‌هایی هستند که از نظر چگونگی و کیفی با هم متفاوت هستند. این متغیرها پیچیده‌تر از متغیرهای کمّی هستند و در اندازه‌گیری، این متغیرها فقط با استفاده از نام‌ها و یا اعداد نامگذاری می‌شوند. مانند: جنسیت، رنگ چشم، گروه خونی، شمارهٔ خودرو و ... .

ب- طبقه‌بندی دوم طبق دقت اندازه‌گیری (گسسته و پیوسته)

متغیر از نظر اندازه‌گیری و اختصاص اعداد به آن، به دو دستهٔ گسسته[۳] و پیوسته[٤] تقسیم می‌شود. متغیر گسسته آن است که فقط ارزش آن با اعداد صحیح نشان داده شود. مثلاً تعداد دانش‌آموزان یک کلاس، یک متغیر گسسته است. در متغیر گسسته، ارزش‌های موجود بین دو مقدار یا دو ارزش، دارای معنی و مفهوم نیست. مثلاً تعداد دانش‌آموزان یک کلاس ۲۰ یا ۲۱ است.

متغیر پیوسته، متغیری است که هر ارزش یا عدد اعشاری را می‌توان به مقدار آن نسبت داد. مثلاً وزن یا قد یک متغیر پیوسته است. برای نمونه میزان بارندگی، دمای یک منطقه، مساحت یک کتابخانه، اندازهٔ طول و عرض کتاب و غیره همگی به عنوان متغیر پیوسته اندازه‌گیری می‌شوند و برای اندازه‌گیری مقیاس اندازه‌گیری مورد استفاده قرار می‌گیرند.

---

۱. Quantitative
۲. Qualitative
۳. Discrete
٤. Continues

ج- طبقه‌بندی سوم (مستقل و وابسته)

متغیر بر اساس نقشی که در پژوهش به عهده دارد، به دو دستهٔ مستقل[1] و وابسته[2] تقسیم می‌شود. متغیر مستقل آن است که در پژوهش‌های آزمایشی به وسیله پژوهشگر به عنوان "علت" در نظر گرفته می‌شود و در اختیار پژوهشگر است. متغیر وابسته آن است که ارزش یا مقدار آن به متغیر مستقل بستگی دارد و یا تحت تأثیر متغیر مستقل تغییر می‌کند. متغیر وابسته در اختیار پژوهشگر نیست و او نمی‌تواند در آن دخل و تصرف و دستکاری کند. پژوهشگر، متغیر وابسته را اندازه می‌گیرد و با استفاده از روش‌های آماری میزان اثر متغیر مستقل را بررسی می‌کند. مثلاً وقتی پژوهشگر قصد تعیین میزان تأثیر آموزش مهارت‌های خواندن در میزان یادگیری متون کتاب روانشناسی شخصیت را دارد، آموزش مهارت‌های خواندن متغیر مستقل و یادگیری متون، متغیر وابسته است.

۱-۵-۵ مقیاس‌های اندازه‌گیری

یکی از اساسی‌ترین فعالیت‌ها در هر پژوهش، اندازه‌گیری متغیرهای مورد مطالعه است. اندازه‌گیری، فرایندی است که از طریق آن حاصل مشاهده کمّیت و کیفیت متغیرها به عدد تبدیل می‌شود. به عبارت دیگر، اندازه‌گیری نسبت دادن عددی به یک صفت یا رویداد بر اساس یک قانون معین است.

ماهیت فرایند اندازه‌گیری که اعداد را به‌وجود می‌آورد، روش‌های آماری لازم و چگونگی تفسیر آن‌ها را مشخص می‌کند. چهار مقیاس اندازه‌گیری یا چهار روش تخصیص عدد را می‌توان از یکدیگر متمایز کرد. این چهار مقیاس عبارتند از: مقیاس‌های اسمی، ترتیبی، فاصله‌ای و نسبی.

---

۱. Independent
۲. Dependent

الف- مقیاس اسمی[۱]

این سطح اندازه‌گیری، ابتدایی‌ترین مقیاس است. این مقیاس صرفاً به تعیین طبقاتی می‌پردازد که افراد، اشیاء یا رویدادها را می‌توان در آن جایگزین کرد. در این مقیاس، هر فرد در یکی از طبقه‌ها جای می‌گیرد و قرار گرفتن یک نفر در بیش از یک طبقه امکان‌پذیر نیست.

معمولاً در این مقیاس، طبقات و متغیرهای کیفی با استفاده از اعداد نامگذاری می‌شوند. اعداد در این مقیاس معنای کمّی ندارند و نمی‌توان آن‌ها را جمع یا تفریق کرد و یا بر روی آن‌ها عملیات ضرب و تقسیم انجام داد؛ بلکه هدف استفاده از اعداد، فقط طبقه‌بندی و تمیز طبقات از یکدیگر است. به عنوان مثال، طبقات مختلف مراجعان به مجلات مختلف یک کتابخانه در جدول ۱ به صورت مقیاس اسمی نشان داده شده است.

جدول ۱-٦: طبقات مختلف مراجعان به مجلات مختلف یک کتابخانه طی یک ماه

عنوان مجله / گرایش تحصیلی	کتابداری و اطلاع رسانی	فصلنامهٔ علوم فناوری اطلاعات	اطلاع شناسی
مدیریت اطلاعات	۳	۱۰	٤
کتابخانهٔ عمومی	۲	٤	٤
کتابخانه‌های دانشگاهی	۷	٦	٥
جمع	۱۲	۲۰	۱۳

ب- مقیاس ترتیبی[۲]

مقیاس ترتیبی نیز همانند مقیاس اسمی به طبقه‌بندی و نامگذاری طبقات می‌پردازد، اما علاوه بر آن، به ترتیب طبقات نیز نظر دارد. در این مقیاس، اعداد صرفاً به منظور رتبه‌بندی افراد با توجه به یک متغیر به کار می‌روند. مقیاس

---

۱. Nominal
۲. Ordinal

ترتیبی به منظور مرتب کردن افراد یا اشیاء، از بیشترین میزان مورد اندازه‌گیری به کم‌ترین میزان آن به کار برده می‌شود.

اعدادی که در این مقیاس برای مرتب کردن به کار می‌روند، بیان‌کننده کمّیت و کیفیت مطلق موضوع اندازه‌گیری نیستند و تنها ترتیب فرد را از نظر آن متغیر در گروه مشخص می‌کنند. همچنین فاصلهٔ بین دو عدد را نمی‌توان مساوی فرض کرد. مثلاً دو کودک را که از نظر میزان علاقه‌مندی آن‌ها به فعالیت‌های کلاسی از بالاترین درجه همکاری تا پایین‌ترین درجه، رتبه‌بندی شده‌اند در نظر بگیرید. در این رتبه‌بندی نمی‌توان پنداشت که فاصلهٔ درجهٔ همکاری بین فرد اول و دوم همانند یا برابر فاصلهٔ درجهٔ همکاری بین فرد دوم و سوم است. در این مقیاس نیز نمی‌توان اعداد را با هم جمع و یا آن‌ها را از هم کم کرد. همچنین عملیات ضرب و تقسیم را نیز نمی‌توان به کار برد. در این مقیاس، عدد صفر وجود ندارد.

ج- مقیاس فاصله‌ای[۱]

این مقیاس علاوه بر طبقه‌بندی، نامگذاری و مرتب کردن طبقه‌ها، فاصله‌های موجود بین متغیرها را مشخص می‌کند. در این مقیاس می‌توان اعداد را با هم جمع و یا از هم کم کرد، اما نمی‌توان آن‌ها را در هم ضرب و یا بر هم تقسیم کرد؛ زیرا صفر در این مقیاس، قراردادی است. بنابراین اگر رضا در درس ریاضی نمره ۱۵ و محمود نمره ۵ گرفته باشد، نمی‌توان ۱۵ را بر ۵ تقسیم و اظهار کرد که رضا سه برابر محمود ریاضی می‌داند؛ بلکه فقط می‌توان این دو نمره را از هم کم و بیان کرد که رضا به اندازه ۱۰ نمره از محمود بیشتر ریاضی یاد گرفته است.

د- مقیاس نسبی[۲]

این مقیاس، بالاترین سطح اندازه‌گیری است و حدود فعالیت آن مشتمل بر کلیهٔ عملیاتی است که می‌توان در مقیاس‌های اسمی، ترتیبی و فاصله‌ای انجام داد. در

---

۱. Interval
۲. Ratio

این مقیاس، صفر مطلق وجود دارد. این مقیاس نسبت را حفظ می‌کند. مثلاً فرض کنید وزن کالایی ٦ کیلوگرم است، در صورتی که وزن کالای دیگر ۲ کیلوگرم باشد، حال آنکه با مقیاس گرم وزن اولی ٦۰۰۰ گرم و وزن دومی ۲۰۰۰ گرم است. ملاحظه می‌شود که در هر دو مقیاس، وزن اولی نسبت به دومی سه برابر است.

خلاصهٔ فصل اول

در این فصل پس از آشنایی با حوزهٔ فعالیت‌های کتاب‌سنجی و کتاب‌سنجی نوین در ایران، ضمن آشنایی با جامعه و نمونه، نحوهٔ انتخاب نمونه و انواع روش‌های نمونه‌گیری شامل نمونه‌گیری تصادفی، نمونه‌گیری منظم، نمونه‌گیری طبقه‌ای و نمونه‌گیری خوشه‌ای شرح داده شد و در انتها انواع مقیاس اندازه‌گیری را مورد بحث قرار دادیم.

خودآزمایی تشریحی فصل اول

۱- کدام یک از موارد زیر جامعه و کدام یک نمونه است؟

الف) کتب خطی کتابخانه‌های کشور

ب) کتب خطی منتخب از کتابخانه‌های کشور

ج) اعضای انجمن فارغ‌التحصیلان رشتهٔ کتابداری سال ۱۳۹۲

د) تعداد کتب از رده خارج شدهٔ کتابخانه‌های کشور

هـ) کارکنان کتابخانه‌های کشور که حقوق آن‌ها بین هفتصد و یک میلیون و دویست هزار تومان است.

و- بودجه‌های هزینه شده در کتابخانه‌های کشور تا سال ۱۳۹۲

۲- این ویژگی‌ها را با چه مقیاس‌هایی می‌توان اندازه‌گیری کرد؟ کدام‌یک از متغیرهای مربوط گسسته و کدام‌یک پیوسته هستند؟

زمان، درآمد، شغل، نژاد

۳- در موارد زیر، جامعه، متغیر، نمونه و مقیاس مربوط را مشخص کنید.

الف) دانشجویان دانشگاه پیام نور و تعداد دفعات مراجعه به بخش مرجع

ب) یک پایگاه مقالات و تعداد مقالات مرتبط با رشتهٔ کتابداری در آن پایگاه

ج) دانشجویان کتابداری و میزان یادگیری درس آمار

٤- در موارد زیر مقیاس اندازه‌گیری را مشخص کنید.

الف) مدارک تحصیلی مراجعان به یک کتابخانه

ب) تعداد مراجعان به مجلات در یک کتابخانه

۵- انواع روش‌های نمونه‌گیری را نام ببرید.

٦- کدام‌یک از جوامع زیر محدود و کدام‌یک نامحدود است؟

الف) خانوارهای با درآمد بیشتر از ۳ میلیون تومان

ب) کتب با قدمت کم‌تر از ده سال

ج) آخرین کتاب چاپ شده در نشر دانشگاهی

۷- برای انتخاب یک نمونهٔ ۱۰ تایی از لیست دانشجویان، چه نوع نمونه‌گیری مورد استفاده قرار می‌گیرد؟

الف) نمونه‌گیری تصادفی ساده ب) نمونه‌گیری خوشه‌ای

ج) نمونه‌گیری طبقه‌ای د) نمونه‌گیری منظم

۸- یک جامعه ارائه دهید و نمونه‌گیری از آن جامعه را شرح دهید.

فصل دوم

طبقه‌بندی داده‌ها و شاخص‌های آماری

هدف کلی

آشنایی با نحوهٔ طبقه‌بندی داده‌ها و محاسبهٔ شاخص‌های آماری.

هدف‌های یادگیری

پس از مطالعهٔ این فصل، شما باید بتوانید:

۱. جدول توزیع فراوانی را برای داده‌های آماری تشکیل دهید.

۲. با انواع نمودارهای آماری آشنا شوید.

۳. انواع شاخص‌های مرکزی از جمله انواع میانگین، میانه و نما را محاسبهٔ کنید.

٤. انواع شاخص‌های پراکندگی از جمله دامنه، واریانس و ضریب تغییر را محاسبهٔ کنید.

مقدمه

هنگامی که توده‌ای از اطلاعات کمّی برای تحقیق گردآوری می‌شود، ابتدا سازمان‌بندی و خلاصه کردن آن‌ها به طریقی که به صورت معنی‌داری قابل درک و ارتباط باشند، ضروری است. روش‌های آمار توصیفی به همین منظور به کار برده می‌شوند. غالباً مفیدترین و در عین حال اولین قدم در سازمان داده‌ها مرتب کردن داده‌ها بر اساس یک ملاک منطقی است. اولین گام در نیل به این هدف، تشکیل جدول توزیع فراوانی و محاسبهٔ شاخص‌های مرکزی و پراکندگی است. در یک جمع‌بندی با استفادهٔ مناسب از روش‌های آمار توصیفی می‌توان دقیقاً ویژگی‌های یک دسته از اطلاعات را بیان کرد. آمار توصیفی همیشه برای تعیین و بیان ویژگی‌های اطلاعات پژوهش‌ها به کار برده می‌شود. در این فصل، به چگونگی تشکیل جداول فراوانی و محاسبهٔ تعدادی از شاخص‌های مرکزی و پراکندگی پرداخته می‌شود.

۲-۱ جدول‌های آماری

داده‌های آماری همواره به صورت انبوهی از اعداد و ارقام خودنمایی می‌کنند که اگر به نحوی تنظیم و دسته‌بندی نشوند، استفاده از آن‌ها و حتی تجزیه و تحلیل آن‌ها دشوار و بعضاً غیرممکن می‌شود. نمایش داده‌ها را با نظمی خاص، در چند سطر و ستون، یک جدول آماری می‌گویند. یک جدول آماری باید به نحوی تنظیم شود که بتوان به آسانی پاره‌ای از دانسته‌های نهفته در داده‌ها را از روی آن خواند. شمارهٔ جدول، نام جدول، عنوان سطر و ستون، زیرنویس و بالاخره مأخذ جدول در صورت لزوم باید روشن و گویا باشند. در کارهای آماری برای خلاصه کردن داده‌ها، آن‌ها را در جدولی به نام جدول توزیع فراوانی تنظیم می‌کنند.

مثال ۲-۱: داده‌های زیر، تعداد مراجعات ۲۰۰ دانشجو به بخش مرجع کتابخانه در مدت زمان دو هفته است:

١ ١ ٠ ٣ ٤ ٦ ١ ٠ ٢ ٥ ٢ ٥ ٣ ٤ ٠ ١ ٢ ٤ ٥ ٣

٤ ٤ ٢ ٦ ٥ ٢ ٣ ٢ ١ ٣ ٢ ٠ ٢ ٢ ٣ ٤ ٦ ٥ ١ ٣

٢ ٤ ٣ ٥ ٢ ٠ ١ ٢ ١ ٣ ٤ ٦ ٥ ٢ ٠ ٤ ٢ ٥ ٢ ٢

٤ ٣ ٥ ١ ٤ ٦ ٢ ١ ٠ ٣ ٤ ٦ ٥ ٢ ١ ٠ ٢ ٣ ٠ ١

٢ ٤ ٠ ٢ ٣ ٥ ١ ٤ ٢ ٤ ٠ ١ ٢ ٠ ٣ ٣ ١ ٥ ٢ ٤

١ ٠ ٣ ٣ ٢ ٥ ١ ٠ ٢ ٢ ٣ ٢ ٣ ٥ ٠ ٢ ١ ٦ ٥ ٣

٤ ٢ ٣ ٤ ٥ ٣ ٢ ٤ ٠ ١ ٦ ٢ ٣ ٢ ١ ٢ ٣ ٤ ٢ ١

٢ ٣ ٤ ٢ ٦ ٣ ٤ ٠ ١ ٢ ٤ ٣ ٢ ٣ ٤ ٢ ٥ ٥ ٢ ٣

٢ ١ ٣ ٢ ٠ ١ ٢ ٤ ٣ ٤ ٥ ١ ٣ ٢ ٤ ٥ ٦ ١ ٤ ٢

٣ ١ ٠ ٢ ١ ٠ ٣ ٤ ٤ ١ ٣ ٢ ٣ ٢ ٣ ٣ ١ ٢ ٣ ٣

با مشاهدهٔ داده‌های فوق، درمی‌یابیم که با توده‌ای بی‌نظم از اعداد مواجه هستیم که هیچ نوع اطلاعات منسجمی در مورد تعداد دفعات مراجعهٔ دانشجویان به بخش مرجع در اختیار ما قرار نمی‌دهند. همواره در بررسی‌ها و پژوهش‌های آماری، علاقمند به دانستن اطلاعاتی هستیم که اعداد طبقه‌بندی نشده و خام پاسخگوی آن‌ها نیستند. بنابراین ناچار به تشکیل جدولی آماری هستیم. در جداول آماری علاوه بر داده‌ها، تعداد و درصد تکرار آن‌ها نمایش داده می‌شود. به طوری که با یک نگاه به جدول می‌توان اطلاعات مفیدی در مورد پراکندگی و توزیع داده‌ها به دست آورد. یکی از متداول‌ترین جداول آماری، جدول توزیع فراوانی است. در جدول توزیع فراوانی همواره موارد زیر را خواهیم داشت:

الف- فراوانی: با شمردن تعداد دفعات تکرار هر داده در میان تمامی داده‌ها، فراوانی آن داده به دست می‌آید. فرض کنید که تعداد n داده داشته باشیم. با قرار دادن داده‌های مشابه در یک دسته، در نهایت K رده یا گروه خواهیم داشت $(K \le n)$ که تعداد داده‌ها در هر دسته را فراوانی آن داده می‌نامیم. فراوانی طبقهٔ iام را با f_i نمایش داده، که در آن $1 \le i \le K$ و $1 \le f_i \le n$ و $\sum_{n}^{i=1} f_i = n$، n تعداد کل داده‌هاست.

ب- فراوانی نسبی: از حاصل تقسیم فراوانی هر طبقه بر تعداد کل داده‌ها که آن را با $i = 1,\ldots,k$، $r_i = \frac{f_i}{n}$ نمایش می‌دهیم، به دست می‌آید؛ به طوری که:

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} \frac{f_i}{n} = \frac{1}{n}\sum_{i=1}^{k} f_i = 1$$

همچنین اگر فراوانی نسبی را در عدد ۱۰۰ ضرب کنیم، درصد وقوع هر داده (درصد فراوانی نسبی) به دست می‌آید.

ج- فراوانی تجمعی: حاصل جمع فراوانی هر طبقه با طبقات قبل از آن را فراوانی تجمعی می‌نامیم و با $F_i \, (1 \le i \le k)$ نمایش می‌دهیم.

$$f_1 + f_2 + \ldots + f_i = \sum_{i=1}^{k} f_i = F_i = \text{فراوانی تجمعی طبقه } i \text{ام}$$

به همین ترتیب، حاصل جمع فراوانی نسبی هر طبقه با طبقات قبل آن را فراوانی تجمعی نسبی می‌نامیم و با F_{ci} نمایش می‌دهیم که در آن $(1 \le i \le k)$ است.

۲-۱-۱ جدول فراوانی برای داده‌های گسسته

فرض کنید که داده‌ها گسسته و با تنوع کم باشند. در چنین مواردی هر نوع داده را به عنوان نمایندهٔ دسته یا رده در نظر می‌گیریم.

در مثال زیر چگونگی تشکیل جدول فراوانی برای داده‌های گسسته ارائه می‌شود.

مثال ۲-۲: در یک تحلیل استنادی، تعداد دفعاتی که در پایان‌نامه‌های ۳۰ دانشجوی کارشناسی ارشد به مجلات اقتصادی استناد شده، شمارش شده و به صورت زیر یادداشت شده است:

۳ ۰ ۱ ۲ ٥ ۲ ۳ ۱ ۱ ۰ ۱ ٤ ٤ ۲ ۲ ۲ ۲ ۰ ٥ ۰ ۲ ۳ ۳ ۳ ۱ ۱ ۱ ۲ ٤ ٥

با توجه به مشاهدات ارقام ۰، ۱، ۲، ۳، ٤، ٥ و ٦ رده یا نمایندهٔ کلاس در نظر گرفته می‌شود، نمایندهٔ هر رده در ستون X_i قرار دارد. به این ترتیب جدول توزیع فراوانی زیر را خواهیم داشت:

جدول ۲-۱: تعداد دفعاتی که در پایان‌نامه‌های ۳۰ دانشجوی کارشناسی ارشد به مجلات اقتصادی استناد شده است

رده-کلاس	فراوانی	درصد فراوانی	فراوانی تجمعی	درصد فراوانی تجمعی
۰	٤	۰/۱۳	٤	۰/۱۳
۱	۷	۰/۲٤	۱۱	۰/۳۷
۲	۸	۰/۲۷	۱۹	۰/٦٤
۳	٤	۰/۱۳	۲۳	۰/۷۷
٤	٤	۰/۱۳	۲۷	۰/۹۰
٥	۳	۰/۱۰	۳۰	۱
جمع	۳۰	۱		

با توجه به جدول، می‌توان به نتایج زیر رسید:

۱- با توجه به عدد ۰/۲۷ در ستون فراوانی نسبی می‌توان نتیجه گرفت که در ۲۷ درصد از پایان‌نامه‌ها، ۲ بار به مجلات اقتصادی استناد شده است.

۲- با توجه به عدد ۰/۷۷ در ستون فراوانی تجمعی نسبی می‌توان نتیجه گرفت که در ۷۷ درصد از پایان‌نامه‌ها حداکثر ۳ بار به مجلات اقتصادی مختلف استناد شده است.

۲-۱-۲ جدول توزیع فراوانی برای داده‌های پیوسته

از آنجا که داده‌ها پیوسته هستند، می‌توان در تشکیل جدول توزیع فراوانی از بازه فاصله‌ای استفاده کرد. نحوهٔ تشکیل جدول توزیع فراوانی با ارائهٔ مثال نشان داده می‌شود.

مثال ۲-۳: داده‌های زیر، تعداد ساعاتی است که ۴۰ محقق از بخش مرجع کتابخانه‌ای در طی مدت یک ماه استفاده کرده‌اند (اعداد به نزدیک‌ترین عدد صحیح گرد شده‌اند):

۱۱ ۹ ۱۲ ۱۵ ۲۰ ۱۳ ۱۴ ۱۷ ۲۳ ۲۲ ۸ ۱۶ ۱۷ ۲۱ ۱۱ ۱۸ ۲۱ ۱۲ ۱۱ ۱۰

۱۴ ۱۳ ۱۹ ۱۶ ۱۵ ۱۷ ۲۰ ۸ ۷ ۱۳ ۱۵ ۱۷ ۱۶ ۱۴ ۲۲ ۱۲ ۱۱ ۹ ۱۸ ۱۹

زمانی که با داده‌های پیوسته کار می‌کنیم، داده‌ها را به رده‌هایی (فاصله‌ها) با طول مساوی تقسیم و فراوانی داده‌ها را در هر رده به دست می‌آوریم. در این حالت نیاز به ثبت تمام داده‌ها در جدول نیست، بلکه هر رده را به صورت مرتب از کوچک به بزرگ در جدول ثبت می‌کنیم که در این حالت روند به این صورت است که:

۱- از آنجا که هر عدد به نزدیک‌ترین عدد صحیح گرد شده است، بنابراین عدد ۱۲ در واقع عددی بین (۱۱/۵و۱۲/۵) بوده است. برای اینکه این مقدار نیز در عملیات وارد شود، به جدول فراوانی ستون دیگری به نام حدود واقعی طبقات اضافه می‌کنیم و عمل خط و نشان زدن اعداد را برای تعیین فراوانی‌ها

از ستون بعد آغاز می‌کنیم تا دریابیم هر کدام از نمره‌ها متعلق به کدام طبقه است.

۲- دامنهٔ واقعی داده‌ها بر اساس کوچک‌ترین و بزرگ‌ترین داده به دست می‌آید که به این ترتیب محاسبهٔ می‌کنیم:

$\min =$ کوچک‌ترین داده =۷

$\max =$ بزرگ‌ترین داده =۲۳

$R = \max - \min = 23 - 6 = 11$ دامنهٔ تغییرات داده‌ها

با توجه به مورد اول اولین رده در جدول با ٦/۵ شروع و آخرین رده در جدول باید با ۲۳/۵ تمام شود.

۳- با توجه به دامنهٔ داده‌ها می‌توان طول هر رده را معین کرد. برای دستیابی به این موضوع، طول دامنه را بر تعداد رده‌ها (که به صورت دلخواه قابل انتخاب است) تقسیم کنیم.

$\omega = \dfrac{R}{K} =$ طول رده

که در آن K تعداد رده‌ها یا طبقات است.

معمولاً تعداد رده‌ها طوری انتخاب می‌شوند که هر رده حداقل پنج داده را در بر داشته باشد و معمولاً توصیه می‌شود که تعداد طبقات کم‌تر از ۵ و بیشتر از ۱۵ نباشد.

٤- فرض کنید که در اینجا علاقمند به تشکیل جدول فراوانی با ۷ رده باشیم. در این صورت طول رده را ۳ در نظر می‌گیریم و جدول به این صورت خواهد بود:

جدول ۲-۲: جدول فراوانی مربوط به مثال ۲-۳

درصد فراوانی تجمعی	فراوانی تجمعی	درصد فراوانی	فراوانی	نمایندهٔ (x_i)	خط و نشان	حدود واقعی طبقات	حدود طبقات
۰/۱۲۵	۵	۰/۱۲۵	۵	۸	𝍸	۶/۵–۹/۵	۶–۹
۰/۳۲۵	۱۳	۰/۲	۸	۱۱	𝍸 𝍷𝍷𝍷	۹/۵–۱۲/۵	۱۰–۱۲
۰/۵۵	۲۲	۰/۲۲۵	۹	۱۴	𝍸 𝍷𝍷𝍷𝍷	۱۲/۵–۱۵/۵	۱۳–۱۵
۰/۷۷۵	۳۱	۰/۲۲۵	۹	۱۷	𝍸 𝍷𝍷𝍷𝍷	۱۵/۵–۱۸/۵	۱۶–۱۸
۰/۹۲۵	۳۷	۰/۱۵	۶	۲۰	𝍸 𝍷	۱۸/۵–۲۱/۵	۱۹–۲۱
۱/۰	۴۰	۰/۰۷۵	۳	۲۳	𝍷𝍷𝍷	۲۱/۵–۲۴/۵	۲۲–۲۴
			۴۰				

- x_i معرف نمایندهٔ هر رده است که عضو میانی رده است.
- هرگاه آخرین رده شامل هیچ عضوی نبود، آن را حذف می‌کنیم. مثلاً در این مثال ردهٔ ۲۴/۵–۲۷/۵ شامل هیچ عضوی نیست. بنابراین حذف می‌شود.

برای آشنایی بیشتر با جدول توزیع فراوانی و طریقهٔ تشکیل آن به صورت کلی جدول توزیع فراوانی به صورت زیر معرفی می‌شود:

نمایندهٔ رده یا کلاس	درصد فراوانی تجمعی نسبی	فراوانی تجمعی نسبی	فراوانی تجمعی	درصد فراوانی نسبی	فراوانی نسبی	فراوانی	رده یا کلاس	شمارهٔ رده یا کلاس
$\frac{a_1+b_1}{2}$	$\frac{F_1}{n}\times 100$	$\frac{F_1}{n}$	F_1	$\frac{f_1}{n}\times 100$	$\frac{f_1}{n}$	f_1	a_1-b_1	۱
$\frac{a_2+b_2}{2}$	$\frac{F_2}{n}\times 100$	$\frac{F_2}{n}$	F_2	$\frac{f_2}{n}\times 100$	$\frac{f_2}{n}$	f_2	a_2-b_2	۲
								$\vdots$
$\frac{a_k+b_k}{2}$	$\frac{F_k}{n}\times 100$	$\frac{F_k}{n}$	F_k	$\frac{f_k}{n}\times 100$	$\frac{f_k}{n}$	f_k	a_k-b_k	K
–	–	–	–	۱۰۰	۱	n	–	جمع

که در آن

- $a_k,\ldots,a_2,a_1$ کران پایین رده‌ها یا کلاس‌ها
- $b_k,\ldots,b_2,b_1$ کران بالای رده‌ها یا کلاس‌ها
- $f_k,\ldots,f_2,f_1$ فراوانی یا فراوانی مطلق رده‌ها یا کلاس‌ها
- $\frac{f_k}{n},\ldots,\frac{f_2}{n},\frac{f_1}{n}$ فراوانی نسبی رده‌ها یا کلاس‌ها
- $F_k,\ldots,F_2,F_1$ فراوانی تجمعی رده‌ها یا کلاس‌ها
- $\frac{F_k}{n},\ldots,\frac{F_2}{n},\frac{F_1}{n}$ فراوانی تجمعی نسبی رده‌ها یا کلاس‌ها
- $\frac{a_k+b_k}{2},\ldots,\frac{a_2+b_2}{2},\frac{a_1+b_1}{2}$ نمایندهٔ رده‌ها یا کلاس‌ها

مثال ۲-٤: اگر توزیع نسبی مراجعه‌کنندگان به یک کتابخانهٔ بزرگ در طول یک روز به صورت زیر باشد، جدول توزیع فراوانی را کامل کنید.

شمارهٔ رده	۱	۲	۳	٤	۵
ردهٔ سنی	۱۰-۱۵	۱۵-۲۰	۲۰-۲۵	۲۵-۳۰	۳۰-۳۵
فراوانی	۵۰	۸۰	۲۰۰	۱۰۰	۷۰

حل:

شمارهٔ رده	**ردهٔ سنی**	**فراوانی**	**فراوانی نسبی**	**درصد فراوانی نسبی**	**فراوانی تجمعی**	**فراوانی تجمعی نسبی**	**درصد فراوانی تجمعی نسبی**	**نمایندهٔ رده**
۱	۱۰-۱۵	۵۰	۰/۱۰	۱۰	۵۰	۰/۱۰	۱۰	۱۲/۵
۲	۱۵-۲۰	۸۰	۰/۱٦	۱٦	۱۳۰	۰/۲٦	۲٦	۱۷/۵
۳	۲۰-۲۵	۲۰۰	۰/٤۰	٤۰	۳۳۰	۰/٦٦	٦٦	۲۲/۵
٤	۲۵-۳۰	۱۰۰	۰/۲۰	۲۰	٤۳۰	۰/۸٦	۸٦	۲۷/۵
۵	۳۰-۳۵	۷۰	۰/۱٤	۱٤	۵۰۰	۱	۱۰۰	۳۲/۵
		۵۰۰	۱	۱۰۰				

در ادامه با ارائهٔ یک مثال، طریقهٔ تشکیل جدول توزیع فراوانی شرح داده می‌شود.

مثال ۲-۵: زمان تأخیر عودت ۵۰ عضوی کتابخانه‌ای برحسب روز به صورت زیر داده شده است:

۳	٥	١	٧	٢	١	٣	٤	٦	٨
٤	٢	٦	١	٥	٤	٢	٧	٣	٧
١٠	٤	١	٨	٣	٢	١	٣	٥	٦
٨	٩	٢	٣	٤	٥	٦	٩	٣	٤
٧	٥	٤	١	٣	٩	٨	٧	٤	٥

الف- اطلاعات داده شده را در ۵ رده، رده‌بندی کنید.

ب- پس از تشکیل و تکمیل جدول توزیع فراوانی به سؤالات زیر جواب دهید:

۱. بیشترین فراوانی در کدام رده و با چند درصد؟

۲. چند نفر و چند درصد از اعضا کم‌تر یا مساوی ٦ روز تأخیر در عودت داشتند؟

۳. چند درصد از اعضا تأخیر بین ٤ تا ٨ روز را داشته‌اند؟

حل: برای تشکیل جدول توزیع فراوانی، ابتدا داده‌ها را مرتب و دامنه و طول رده‌ها را محاسبهٔ می‌کنیم.

١	١	١	١	١	١	٢	٢	٢	٢
٢	٣	٣	٣	٣	٣	٣	٣	٣	٤
٤	٤	٤	٤	٤	٤	٤	٥	٥	٥
٥	٥	٥	٦	٦	٦	٦	٧	٧	٧
٧	٧	٨	٨	٨	٨	٩	٩	٩	١٠

دامنهٔ داده‌ها برابر است با: $R=10-1=9$

تعداد کلاس: $k=5$

طول رده: $\omega=\frac{R}{k}=\frac{9}{5}=1/8\Rightarrow \omega\sim 2$

شمارهٔ رده	رده	فراوانی	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	فراوانی تجمعی نسبی	درصد فراوانی تجمعی نسبی	نمایندهٔ رده
۱	۰–۲	۱۱	۰/۲۲	۲۲	۱۱	۰/۲۲	۲۲	۱
۲	۲–٤	١٦	۰/۳۲	۳۲	۲۷	۰/٥٤	٥٤	۳
۳	٤–٦	۱۰	۰/۲۰	۲۰	۳۷	۰/۷٤	۷٤	٥
٤	٦–۸	۹	۰/۱۸	۱۸	٤٦	۰/۹۲	۹۲	۷
٥	۸–۱۰	٤	۸	۸	٥۰	۱	۱۰۰	۹
کل		٥۰	۱	۱۰۰				

با توجه به جدول توزیع فراوانی به دست آمده می‌توان گفت که:

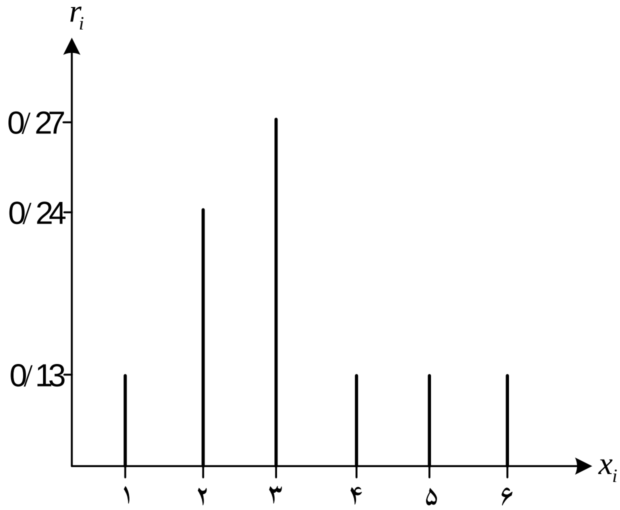
- بیشترین فراوانی متعلق به رده یا گروه ۲ با ۳۲ درصد است.
- با توجه به ستون فراوانی تجمعی می‌توان گفت ۳۷ نفر از اعضا ٦ روز تأخیر در عودت داشته که درصد مربوط به آن ۷٤ درصد است.
- با توجه به ستون‌های سوم و چهارم جدول توزیع فراوانی، ۱۹ نفر با ۳۸ درصد تأخیر بین ٤ تا ۸ روز داشته‌اند.

۲–۲ نمودارهای آماری

یکی دیگر از روش‌های مناسب برای خلاصه کردن داده‌های آماری استفاده از نمودارهای آماری است. نمودارهای آماری بهترین ابزار برای نمایش نحوهٔ توزیع داده‌ها هستند که معروف‌ترین آن‌ها را معرفی می‌کنیم:

۲-۲-۱ نمودارهای آماری برای داده‌های گسسته

الف- نمودار میله‌ای: در این نمودار، محور xها نمایش‌دهندهٔ مقادیر داده‌ها و محور yها نمایش‌دهندهٔ فراوانی نسبی است. چون فراوانی نسبی داده‌های هر جامعهٔ آماری مقادیری بین صفر و یک را می‌گیرند، بنابراین می‌توان چندین نمودار را با هم مقایسه کرد. نمودار ۲-۱ نمودار میله‌ای برای مثال۲-۲ است.



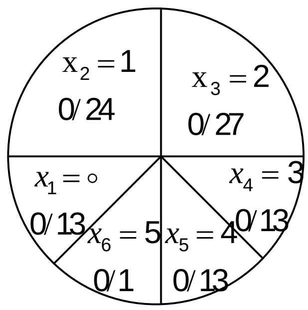
نمودار ۲-۱: نمودار میله‌ای مربوط به مثال۲-۲

ب- نمودار دایره‌ای: این نمودار، درون دایره‌ای رسم می‌شود که به قطاع‌هایی تقسیم شده است و هر قطاع متناسب با مساحت خود، معرف یکی از فراوانی‌های نسبی در جدول فراوانی است. برای تعیین قطاع دایره، هر فراوانی را بر کل فراوانی‌ها تقسیم کرده و حاصل را در مجموع درجات محیط دایره (۳٦۰) ضرب می‌کنیم. در جدول ۲-۳ محاسبات قطاع مربوط به مثال ۲-۲ انجام شده است.

جدول ۲–۳: محاسبات قطاع مربوط به مثال ۲–۲

x_i	f_i	اندازهٔ قطاع دایره
۰	٤	۰/۱۳×۳٦۰=٤٦/۸
۱	۷	۰/۲٤×۳٦۰=۸٦/٤
۲	۸	۰/۲۷×۳٦۰=۹۷/۲
۳	٤	۰/۱۳×۳٦۰=٤٦/۸
٤	٤	۰/۱۳×۳٦۰=٤٦/۸
٥	۳	۰/۱×۳٦۰=۳٦
جمع	۳۰	۳٦۰

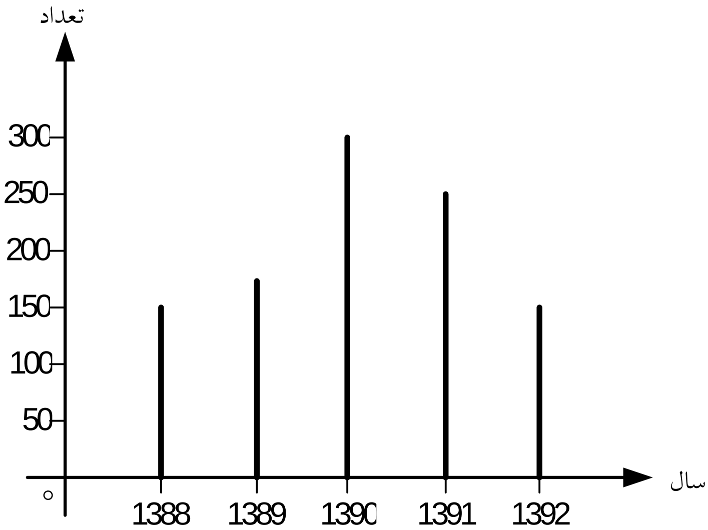
با توجه به جدول فوق می‌توان نمودار دایره‌ای را برحسب زوایای به دست آمده رسم کرد. نمودار ۲–۲، نمایش‌دهندهٔ نمودار دایره‌ای برای جدول ۲–۳ است.



نمودار ۲–۲: نمودار دایره‌ای مربوط به داده‌های مثال ۲–۲

مثال ۲–٦: جدول توزیع فراوانی کتب خریداری شده از نمایشگاه دائمی برای یک مرکز تحقیقات به صورت زیر داده شده است. نمودارهای میله‌ای و دایره‌ای را رسم کنید.

سال	۱۳۸۸	۱۳۸۹	۱۳۹۰	۱۳۹۱	۱۳۹۲
تعداد کتب	۱٤۰	۱٦۰	۳۰۰	۲٥۰	۱٥۰



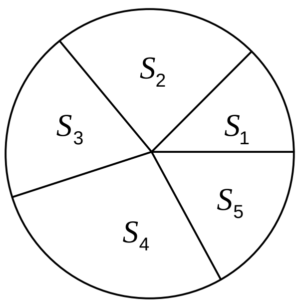
- برای رسم نمودار میله‌ای، محور xها را سال خریداری و تعداد کتب را متناسب با تعداد روی محور yها مشخص می‌کنیم.
- برای رسم نمودار دایره‌ای با برابر کل فراوانی با مساحت کل دایره و استفاده از رابطهٔ زیر قطاع فراوانی مربوط به هر رده به دست می‌آید.

$$S_i = \frac{360}{n} \times f_i = \frac{360}{1000} \times f_i \quad , \quad i = 1, 2, 3, 4, 5$$

$$S_1 = \frac{360}{1000} \times 140 = 50/40 \qquad S_2 = \frac{360}{1000} \times 160 = 57/6$$

$$S_3 = \frac{360}{1000} \times 300 = 108 \qquad S_4 = \frac{360}{1000} \times 250 = 90$$

$$S_5 = \frac{360}{1000} \times 150 = 54$$

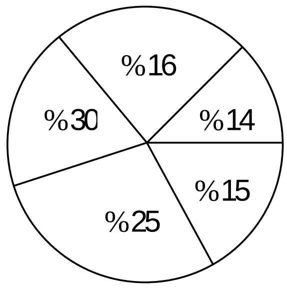


یادآوری: اگر در رابطهٔ $S_i = \frac{360}{n} \times f_i$ به جای ۳٦۰ عدد ۱۰۰ قرار گیرد، مساحت به دست آمده به درصد خواهد بود. بنابراین

$$S_i = \frac{100}{n} \times f_i = \frac{100}{1000} \times f_i = 0/1\, f_i$$

$$S_1 = 0/1 \times 140 = 14 \qquad S_2 = 0/1 \times 160 = 16 \qquad S_3 = 0/1 \times 300 = 30$$

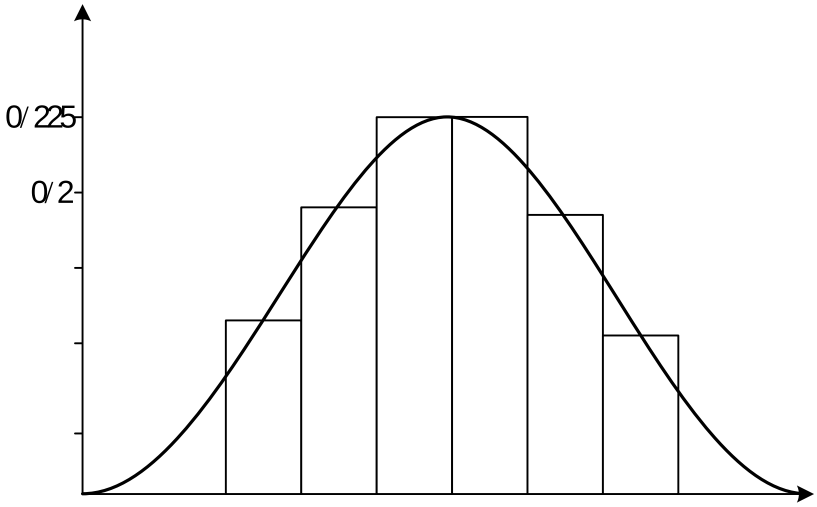
$$S_4 = 0/1 \times 250 = 25 \qquad S_5 = 0/1 \times 150 = 15$$



۲-۲-۲ نمودارهای آماری برای داده‌های پیوسته

همان‌طور که در جدول فراوانی مشاهده کردید، تفاوت عمدهٔ رده‌های پیوسته با گسسته، در به کارگیری رده‌ها است. این امر در نمودارها نیز کاملاً صادق است. در این قسمت سه نمودار را که برای نمایش داده‌های پیوسته به کار می‌روند معرفی می‌شوند:

الف- نمودار هیستوگرام (نمودار ستونی): این نمودار، کاملاً مشابه نمودار میله‌ای است با این تفاوت که محور xها نمایش‌دهندهٔ رده‌های جدول توزیع فراوانی است و به ازای هر رده مستطیلی که عرض آن یک واحد و ارتفاع آن معادل فراوانی نسبی آن رده است، رسم می‌شود. بنابراین مجموع مساحت‌های مستطیل‌های رسم شده برابر واحد است که همان مجموع کل فراوانی‌های نسبی است. نمودار ۲-۳ مربوط به نمودار هیستوگرام برای مثال۲-۳ است.



نمودار ۲–۳: نمودار هیستوگرام مربوط به مثال ۲–۳

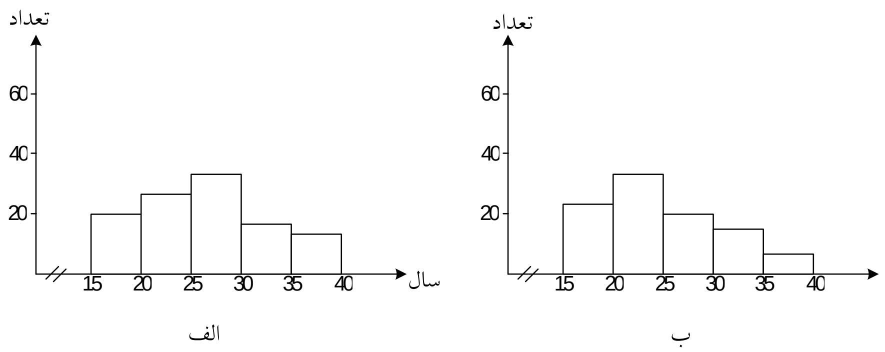
ب– نمودار چندبر فراوانی: اگر در نمودار هیستوگرام وسط قاعده‌های بالای مستطیل‌ها را به یکدیگر توسط خطوطی متوالی متصل کنیم، نمودار چندبر فراوانی به دست می‌آید که در شکل بالا نیز نشان داده شده است. انتهای خطوط به وسط ردهٔ ماقبل و انتهای خطوط به وسط ردهٔ مابعد مستطیل‌های هیستوگرام متصل می‌شوند. به این ترتیب، مساحت زیر نمودار چندبرفراوانی مجدداً برابر واحد خواهد بود.

ج– نمودار منحنی فراوانی تجمعی: نمودار مستطیلی، تصویر واضحی از توزیع را نشان می‌دهد. ولی معمولاً رسم نمودار فراوانی تجمعی ساده‌تر است. برای رسم این منحنی، روی محور xها و در ربع اول، حدود واقعی طبقات و روی محور yها داده‌های موجود در ستون فراوانی تجمعی را منتقل می‌سازیم.

مثال ۲–۷: برای مقایسهٔ قدمت کتب موجود در دو کتابخانه اطلاعات زیر به دست آمده است (تعداد برحسب ۱۰۰۰۰ عدد).

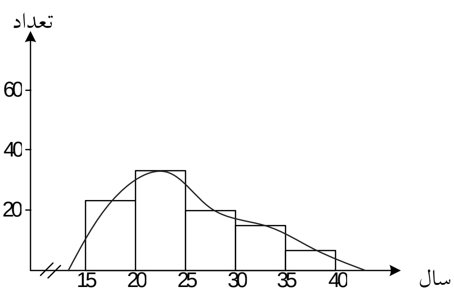
کتابخانه	۳۵–۴۰	۳۰–۳۵	۲۵–۳۰	۲۰–۲۵	۱۵–۲۰	بازه زمانی برحسب سال
الف	۱۰	۱۸	۲۷	۲۵	۲۰	تعداد
ب	۵	۱۵	۲۰	۳۵	۲۵	تعداد

نمودارهای هیستوگرام، چندبر و منحنی فراوانی تجمعی یا اجایو را رسم کنید و آن‌ها را با هم مقایسه کنید.

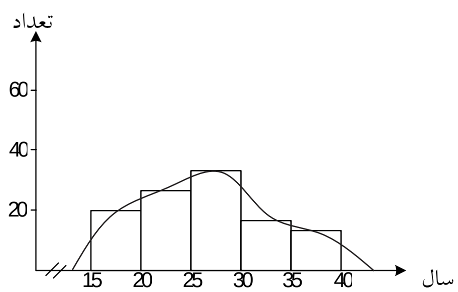


نمودار ۲–۴

برای رسم نمودار چندبر فراوانی، کافی است نقاط میانی نمودارهای مستطیل را به هم وصل و پس از امتداد آن‌ها و وصل به محور xها نمودار حاصل می‌شود.

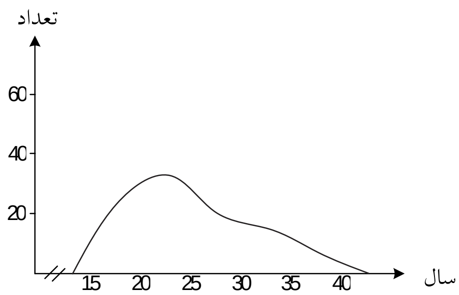


نمودار ۲–۵

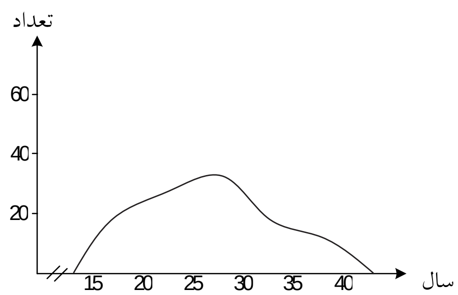


نمودار ۲–۶

اگر در دو نمودار فوق مستطیل‌ها حذف شوند، نمودار چندبر فراوانی به صورت زیر خواهد شد:



نمودار ۲-۷

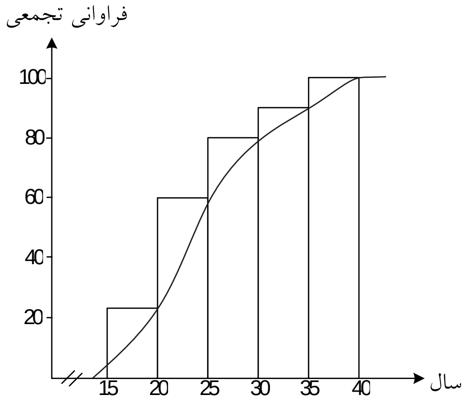


نمودار ۲-۸

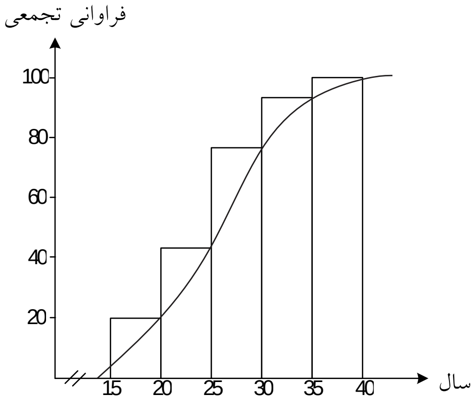
برای رسم نمودار منحنی فراوانی تجمعی، ابتدا ستون‌های فراوانی تجمعی را تشکیل می‌دهیم.

رده (سال)	فراوانی تجمعی کتابخانهٔ الف	فراوانی تجمعی کتابخانهٔ ب
۱۵-۲۰	۲۰	۲۵
۲۰-۲۵	٤٥	٦٠
۲۵-۳۰	۷۲	۸۰
۳۰-۳۵	۹۰	۹۵
۳۵-٤۰	۱۰۰	۱۰۰

برای رسم نمودار منحنی فراوانی تجمعی، نمودار هیستوگرام، با توجه به فراوانی تجمعی رسم می‌شود و از گوشه‌های بالای سمت به هم وصل و امتداد آن تا اتصال به محور xها، نمودار حاصل می‌شود.

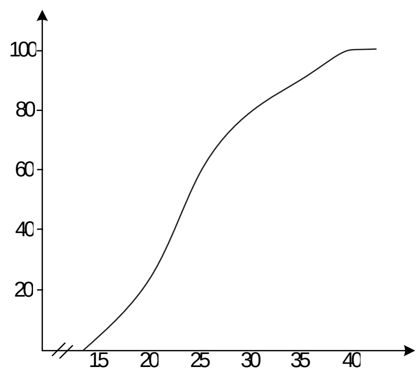


نمودار ۲–۹

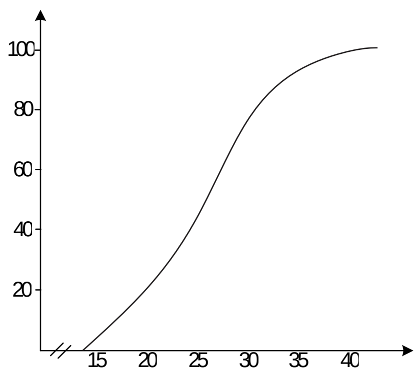


نمودار ۲–۱۰

پس از حذف مستطیل‌ها در دو نمودار، نمودار منحنی فراوانی تجمعی به صورت زیر خواهد بود:

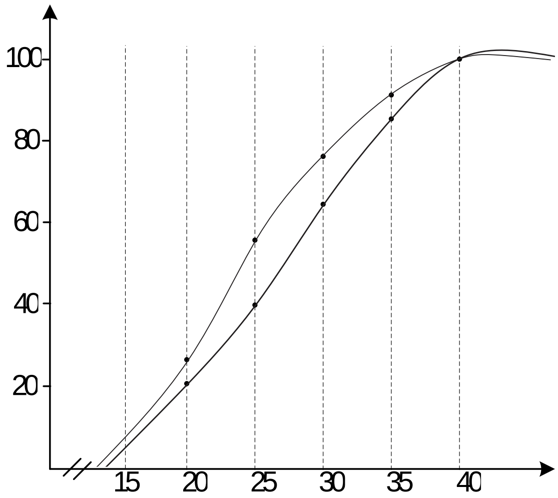


نمودار ۲–۱۱



نمودار ۲–۱۲

برای مقایسهٔ قدمت کتب موجود در دو کتابخانه، نمودار تجمعی را در یک شکل رسم می‌کنیم که خواهیم داشت:



نمودار ۲-۱۳

با توجه به شکل، مشخص است تعداد کتب با قدمت کم‌تر از ۳۵ سال در کتابخانهٔ دوم بیشتر از کتابخانهٔ اول است. در حالی که کتب با قدمت بالای ۳۵ سال تا حدودی در دو کتابخانه برابر است.

۲-۳ شاخص‌های مرکزی و پراکندگی

از آنجا که با مطالعهٔ یک جامعه آماری تعداد زیادی داده به دست می‌آوریم و مطالعه روی تک تک یا قسمتی از این داده‌ها مشکل و حتی غیرممکن است، همواره علاقه داریم براوردی برای پارامترها ارائه دهیم تا بتوان با یک نگاه اجمالی به آن‌ها یک دید کلی نسبت به کل داده‌ها و جامعهٔ آماری به دست آورد. بنابراین در این بخش، شاخص‌های مرکزی را معرفی می‌کنیم.

۲-۳-۱ شاخص‌های مرکزی

تعیین مقدار مرکزی در مطالعهٔ هر جامعهٔ آماری بسیار مهم و ضروری است. هر معیار عددی که معرف مرکز مجموعهٔ داده‌ها باشد، شاخص مرکزی نامیده

می‌شود. عموماً داده‌ها در جامعهٔ آماری یک نوع تجمع و فشردگی حول یک مقدار خاص از صفت مورد مطالعه را به وجود می‌آورند که این مقدار خاص به عنوان یک پارامتر مرکزی معرفی می‌شود. برای اندازه‌گیری شاخص‌های مرکز جامعه‌های آماری، ملاک‌های متعددی وجود دارد که مهم‌ترین آن‌ها عبارتند از: میانگین، میانه، نما (مد).

۱- میانگین حسابی (معدل): اصلی‌ترین و مورد استفاده‌ترین شاخص مرکزی، میانگین است. میانگین، یک اصطلاح آماری است و درست معادل معدل حسابی است که ما در زندگی روزمره با آن سر و کار داریم. میانگین یا معدل حسابی از حاصل جمع داده‌ها تقسیم بر تعداد آن‌ها به دست می‌آید. برای مثال، میانگین ۲۰ و ۴۰ به صورت $\frac{20+40}{2}=30$ محاسبهٔ می‌شود. این محاسبهٔ را می‌توان به شکل عام‌تری تغییر داد که در آن X_1 نشان‌دهندهٔ عدد اول و X_2 نشان‌دهندهٔ عدد دوم باشد، بنابراین میانگین آماری این دو عدد (دادهٔ آماری) به شکل $\frac{X_1+X_2}{2}$ محاسبهٔ می‌شود. حال اگر ۱۰ داده داشته باشیم از X_1 تا X_{10} را با هم جمع و حاصل را بر ۱۰ تقسیم می‌کنیم. در ریاضیات شیوهٔ خلاصه‌تر برای بیان جمع X_1 تا X_{10} وجود دارد که به صورت $\sum_{i=1}^{n} X_i$ است (که در اینجا به جای n عدد ۱۰ را می‌گذاریم)؛ پس فرمول محاسبهٔ میانگین برای تمام داده‌ها به صورت زیر است:

$$\overline{X}=\frac{\sum_{i=1}^{n} X_i}{n}$$

در اینجا $\overline{X}$ معرف میانگین و n نشان‌دهندهٔ تعداد داده‌ها است. این فرمول به ما می‌گوید که تمام داده‌ها از X_1 تا X_n را با هم جمع کنید (Σ علامت در ریاضی سیگما خوانده می‌شود و به معنی مجموع است) و سپس بر تعداد آن‌ها (n) تقسیم کنید.

مثال ۲–۸: میزان درآمد دو کتابفروش در چهار فصل از سال به صورت زیر گزارش شده است:

میزان درآمد کتابفروش اول در یک سال

x_i	x_1	x_2	x_3	x_4
فصل	بهار	تابستان	پاییز	زمستان
قیمت (هزار تومان)	۲۵۰	۳۰۰	٤۵۰	٤۰۰

میزان درآمد کتابفروش دوم در یک سال

x_i	x_1	x_2	x_3	x_4
فصل	بهار	تابستان	پاییز	زمستان
قیمت (هزار تومان)	۲۵۰	٤۰۰	۲۵۰	۳۰۰

میانگین یا متوسط درآمد کتابفروش اول برابر است با:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{250+300+450+400}{4} = 350$$

میانگین یا متوسط درآمد کتابفروش دوم برابر است با:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{250+400+250+300}{4} = 300$$

پس فروشندهٔ اول با میانگین فروش ۳۵۰ هزار تومان در ماه نسبت به فروشندهٔ دوم با میانگین فروش ۳۰۰ هزار تومان در ماه، طی یک سال گذشته درآمد بیشتری داشته است.

ویژگی‌های میانگین: همان‌طور که گفته شد، اگر $X_1, X_2, \dots, X_k$ یک نمونه به حجم n باشد، میانگین نمونه از رابطهٔ زیر به دست می‌آید.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\bar{X}$ دارای ویژگی‌های زیر است:

۱- واحد اندازه‌گیری را حذف می‌کند؛ یعنی اگر داده‌ها برحسب متر باشد، واحد میانگین نیز متر خواهد بود.

۲- اگر به داده‌ها مقدار ثابت a اضافه (یا کم) کنیم، میانگین جدید از اضافه کردن (یا کم کردن) از میانگین قدیم به دست می‌آید. اگر میانگین $X_1, X_2, \ldots, X_k$ برابر با $\bar{X}$ باشد، میانگین $X_1 \pm a, X_2 \pm a, \ldots, X_n \pm a$ برابر است با $\bar{X} \pm a$.

۳- اگر داده‌ها را در مقدار ثابت غیر صفر k ضرب (یا تقسیم) کنیم، میانگین جدید از ضریب یا تقسیم میانگین قدیم k به دست می‌آید.

مثال ۲-۹: میانگین سری داده‌های زیر را به دست آورید.

۹	۵	۴	۲	۰
۱۲	۸	۷	۵	۳
۴۵	۲۵	۲۰	۱۰	۰

حل: میانگین گروه اول برابر است با: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{5}(0+2+4+5+9) = 4$

- میانگین گروه دوم برابر است با: $\bar{X} + 3 = 4 + 3 = 7$

- میانگین گروه سوم برابر است با: $5\bar{X} = 5 \times 4 = 20$

ضمن این که می‌توان مستقیماً آن‌ها را محاسبهٔ کرد.

محاسبهٔ میانگین حسابی در جدول توزیع فراوانی: در صورتی‌که داده‌ها در یک جدول توزیع فراوانی طبقه‌بندی شده باشند و جدول k رده داشته باشد، به‌طوری

که $x_1,\ldots,x_n$ نمایندهٔ طبقات و $f_1,\ldots,f_n$ فراوانی رده‌ها باشد، میانگین در جدول توزیع فراوانی از فرمول زیر به دست می‌آید:

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{k} f_i x_i \qquad , \qquad n=\sum_{i=1}^{k} f_i$$

مثال ۲–۱۰: تعداد کتب امانت گرفته شده از کتابخانهٔ دانشکده‌ای طی مدت ۱۲۰ روز در جدول زیر خلاصه شده است. مایلیم بدانیم که به طور متوسط، روزانه چند کتاب به امانت رفته است؟

برای به دست آوردن میانگین ابتدا باید نمایندهٔ رده‌ها (x_i) و $f_i x_i$ را محاسبهٔ کنیم؛ لذا داریم:

تعداد کتب امانت داده شده	تعداد روزها (f_i)	x_i	$f_i x_i$
۰–۲۰	۵	۱۰	۵۰
۲۱–۳۰	۹	۲۵	۲۲۵
۳۱–۴۰	۸	۳۵	۲۸۰
۴۱–۵۰	۲۱	۴۵	۹۴۵
۵۱–۶۰	۲۲	۵۵	۱۲۱۰
۶۱–۷۰	۱۵	۶۵	۹۷۵
۷۱–۸۰	۱۰	۷۵	۷۵۰
۸۱–۹۰	۱۰	۸۵	۸۵۰
۹۱–۱۰۰	۷	۹۵	۶۶۵
۱۰۱–۱۱۰	۶	۱۰۵	۶۳۰
۱۱۱–۱۲۰	۷	۱۱۵	۸۰۵
مجموع	$120=n$		$\sum_{i=1}^{n} x_i f_i = 7385$

با استفاده از فرمول داریم:

در این مثال $k=۱۱$،

$$\overline{x}=\frac{۱}{n}\sum_{i=۱}^{۱۱} f_i x_i=\frac{۱}{۱۲۰}\sum_{i=۱}^{۱۱} f_i x_i=\frac{۷۳۸۵}{۱۲۰}=۶۱/۵۴\cong ۶۲$$

پس می‌توان این طور نتیجه گرفت که به طور متوسط حدود ٦٢ کتاب در روز، طی یک دوره ۱۲۰ روزه از کتابخانه به امانت گرفته شده است.

۲- میانه (نقطه وسط): به منظور محاسبهٔ میانه، ابتدا داده‌ها را از کم به زیاد یا از زیاد به کم مرتب می‌کنیم. اگر تعداد داده‌ها فرد باشد، میانهٔ مقدار عدد وسط است. برای مثال، میانهٔ مجموعهٔ داده‌های ۳، ۵، ۹،۱۱، ١٤، عدد ۹ است.

اگر تعداد داده‌ها زوج باشد، میانه از میانگین دو عدد وسطی به دست می‌آید. برای مثال، میانه برای داده‌های ۳، ۵، ۹، ۱۱، ١٤ و ۱۸ برابر با $10=\frac{9+11}{2}$ است. از نحوهٔ محاسبهٔ میانه مشخص می‌شود که ۵۰٪ داده‌ها بالای میانه و ۵۰٪ درصد آن‌ها پایین میانه قرار دارند. بنابراین اصطلاح میانه همان مفهوم خط سفید وسط خیابان را دارد که نصف خیابان در سمت چپ آن و نصف دیگر در سمت راست آن قرار دارد.

مثال ۲-۱۱: اگر قیمت ۷ کتاب به صورت زیر گزارش شده باشد:

۱۰۰ ۲۰۰ ۳۰۰ ٤۵۰ ۵۷۰ ۸۲۰ ۹۵۰

میانه را به‌دست آورید.

در اینجا تعداد داده‌ها فرد است؛ لذا بعد از مرتب کردن داده‌ها، دادهٔ وسط را به عنوان میانه برمی‌گزینیم. پس ٤۵۰ میانه مشاهده شده است.

مثال ۲-۱۲: میزان فروش دو کتابفروش معمولی و ممتاز برای چند ماه متوالی به صورت زیر گزارش شده است:

کتابفروش معمولی:	۱۴۰۰	۱۰۵۰	۴۹۵	۶۹۰	۹۵۰	۷۸۰	۴۵۰	۲۰۰
کتابفروش ممتاز:	۱۴۸۰	۲۰۰۰	۱۷۵۰	۴۰۰	۹۰۰	۶۰۰	۳۵۰	

برای پیدا کردن میانه، داده‌ها را مرتب می‌کنیم.

کتابفروش معمولی: ۱۴۰۰ ۱۰۵۰ ۹۵۰ ۷۸۰ ۶۹۰ ۴۹۵ ۴۵۰ ۲۰۰

میانه $= \frac{690+780}{2} = 735$

کتابفروش ممتاز: ۲۰۰۰ ۱۷۵۰ ۱۴۸۰ ۹۰۰ ۶۰۰ ۴۰۰ ۳۵۰

میانه=۹۰۰

میانهٔ فروش کتابفروش معمولی ۷۳۵ است، در حالی که میانهٔ فروش کتابفروش ممتاز ۹۰۰ است. نتیجه می‌گیریم که در کتابفروشی معمولی ۵۰٪ از ماه‌های در نظر گرفته شده فروش کم‌تر از ۷۳۵ هزار تومان بوده است. این در حالی است که این عدد برای کتابفروشی ممتاز ۹۰۰ است.

میانه همچنین به عنوان یک اندازهٔ گرایش مرکزی برای داده‌های رتبه‌ای استفاده می‌شود. محاسبهٔ میانه برای داده‌های رتبه‌ای، درست مانند داده‌های فاصله‌ای یا نسبتی است. فرض کنید از کاربران سؤال می‌شود که کارایی یک نظام جدید بازاریابی را با استفاده از یک مقیاس پنج درجه‌ای عالی، خوب، رضایت‌بخش، متوسط و ضعیف درجه‌بندی کنند که نتایج زیر به دست می‌آید:

۱ ۲ ۲ ۳ ۳ ۳ ۴ ۴ ۴ ۴ ۵ ۵ ۵ ۵ ۵

در اینجا عدد ۱ نشاندهندهٔ ضعیف و عدد ۵ نشاندهندهٔ عالی است. میانه عدد ٤ است که به ما می‌گوید کاربر سیستم را به عنوان یک سیستم خوب ارزیابی کرده است.

در اینجا به دلیل ویژگی‌های مقیاس رتبه‌ای و اینکه میانگین گرفتن از رتبه‌ها بی‌معنی است، نتیجه می‌گیریم که نمی‌توانیم از میانگین استفاده کنیم. برای ارزیابی داده‌های فوق می‌دانیم که ٤، از یک درجه‌بندی بهتر نسبت به ۳ برخوردار است و در مقابل، درجه ۳ بهتر از ۲ است، اما نمی‌توانیم بگوییم که اختلاف بین ۱ و ۲ برابر با اختلاف بین ۳ و ٤ است. در حالی که در محاسبهٔ ریاضی مربوط به میانگین فرض می‌شود که اختلاف‌ها یکسان هستند.

محاسبهٔ میانه در جدول توزیع فراوانی: هرگاه بخواهیم میانهٔ یک سری داده آماری را که درون یک جدول طبقه‌بندی شده‌اند به دست آوریم، باید به روش زیر عمل کنیم. برای راحتی بیشتر، توضیحات با استفاده از جدول توزیع فراوانی مثال ۲–۱۰ ارائه می‌شود. برای دستیابی به مقدار میانه در جدول توزیع فراوانی به مقادیر فراوانی و فراوانی تجمعی نیاز داریم.

ردیف	تعداد کتب امانت داده شده	تعداد روزها (f_i)	F_i (فراوانی تجمعی)
۱	۰–۲۰	۵	۵
۲	۲۱–۳۰	۹	۱٤
۳	۳۱–٤۰	۸	۲۲
٤	٤۱–۵۰	۲۱	٤۳
۵	۵۱–٦۰	۲۲	٦۵
٦	٦۱–۷۰	۱۵	۸۰
۷	۷۱–۸۰	۱۰	۹۰
۸	۸۱–۹۰	۱۰	۱۰۰
۹	۹۱–۱۰۰	۷	۱۰۷
۱۰	۱۰۱–۱۱۰	٦	۱۱۳
۱۱	۱۱۱–۱۲۰	۷	۱۲۰
	مجموع	$120=n$	

بنا بر تعریف، میانه عبارت است از ارزش عددی نقطه‌ای که نیمی از داده‌ها کم‌تر از آن و نیمی دیگر بیشتر از آن هستند؛ لذا $\frac{n}{2}$ را محاسبهٔ مـی‌کنیم. در ایـن مثال داریم $\frac{120}{2}=60$ و در ستون فراوانی تجمعی رده‌هایی را که $\frac{n}{2}=60$ بین آن‌ها می‌افتد پیدا می‌کنیم. در اینجا $65>60>43$ است (یعنی ردیف‌هـای ٤ و۵). عـدد ٤۳ نشاندهندهٔ این است که طی ٤۳ روز نهایتاً ۵۰ کتاب به امانت رفته است؛ پس میانه در طبقه ۵ قرار دارد، اما کجای آن؟

حدّ پایین طبقه شامل میانه (یعنی طبقهٔ ۵) را در نظر می‌گیریم. این مقدار ۵۱ است $(L_5=51)$ فراوانی این طبقه ۲۲ است $(f_5=22)$ و فراوانی تجمعی طبقهٔ قبل از آن $(F_4=43)$ با توجه به اعداد به دست آمده برای محاسبهٔ میانه، از فرمول زیر استفاده می‌کنیم:

$$\text{میانه}=M_{\text{n}}=a_i+\frac{\frac{n}{2}-F_{i-1}\times l}{f_i}$$

که در آن:

a_i: حدّ پایین طبقه‌ای است که میانه در آن قرار دارد.

F_i: فراوانی تجمعی طبقهٔ ماقبل طبقه‌ای که در آن میانه قرار دارد.

l: فاصله طبقات.

f_i : فراوانی طبقه‌ای که میانه در آن قرار دارد.

پس داریم:

$$M_{\text{n}}=51+\frac{\left(\frac{120}{2}-43\right)9}{22}=57/95 \approx 60$$

مثال ۲–۱۳: توزیع سنی کارکنان در بخشی از کتابخانهٔ ملی به صورت زیر داده شده است.

رده	۱۸–۲۲	۲۲–۲٦	۲٦–۳۰	۳۰–۳٤	۳٤–۳۸
فراوانی	۵	۷	۲۰	۱۰	۸

میانهٔ سنی کارکنان را به دست آورید و کاربرد آن را بیان کنید.

حل: همان‌طور که گفته شد میانه از فرمول زیر به دست می‌آید.

میانه $M_n = a_i + \dfrac{\frac{n}{2} - F_{i-1}}{f_i} \times \ell$

برای پیدا کردن کلاسی که میانه در آن قرار دارد، اقدامات زیر را انجام می‌دهیم.

- $\frac{n}{2}$ را محاسبهٔ می‌کنیم.
- $\frac{n}{2}$ را با ستون فراوانی تجمعی مقایسه می‌کنیم. اولین عددی که از $\frac{n}{2}$ بزرگ یا مساوی باشد، فراوانی تجمعی مربوط به کلاس است که میانه در آن قرار دارد. بنابراین

رده	فراوانی	فراوانی تجمعی
۱۸–۲۲	۵	۵
۲۲–۲٦	۷	۱۲
۲٦–۳۰	۲۰	۳۲
۳۰–۳٤	۱۰	٤۲
۳٤–۳۸	۸	۵۰
کل	۵۰	–

بنابراین، به راحتی می‌توان گفت که میانه در کلاس سوم قرار دارد. $(i = 3)$

$$M_n = a_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \times \ell$$

$$= 26 + \frac{25-12}{20} \times 4 = 26 + 2/6 = 28/6$$

با توجه به مقدار میانه $M_n = 28/6$ می‌توان گفت که ۵۰٪ از کارکنان، سن کم‌تر یا مساوی ۲۸/٦ سال دارند.

۳- نما (مد): برای داده‌های اسمی نمی‌توان میانگین و میانه را به کار برد. فرض کنید در یک تحقیق، داده‌هایی را در خصوص جنسیت، جمع‌آوری و مردان را با عدد ۱ و زنان را با عدد ۲ کدگذاری کرده‌ایم. چنانچه میانگین ۱ها و ۲ها را محاسبهٔ کنیم، عددی بین ۱ و ۲ تقریباً ۱/۵ را به دست می‌آوریم. این ۱/۵ چه مفهومی دارد؟ آیا این بدین معنی است که میانگین افراد در تحقیق، فردی است که نصف آن مذکر و نصف آن مؤنث است؟ همین مشکل، زمانی که از میانه به عنوان اندازهٔ گرایش مرکزی استفاده می‌شود، پیش خواهد آمد.

نما تنها اندازه گرایش مرکزی مناسب برای داده‌های اسمی است. نما رخداد فراوان‌ترین عدد در یک توزیع است. رأس یک هیستوگرام نما را نشان می‌دهد. در مجموعهٔ مشاهدات نما یا مد، مشاهده یا مشاهداتی هستند که بیشتر از بقیه تکرار شده باشند. برای مثال در یک کتابخانه ممکن است بیشترین کتاب‌ها به زبان فارسی باشد. در ادامه، مد را با نماد M_o نشان خواهیم داد و یادآوری می‌کنیم، چنانچه کلیهٔ مشاهدات به یک اندازه تکرار شده باشند، مشاهدات مورد بررسی فاقد مد خواهند بود.

مثال ۲-١٤: فرض کنید:

الف- حروف A، B، C و D به ترتیب مشخص‌کنندهٔ زبان کتاب‌ها به ترتیب فارسی، انگلیسی، عربی و فرانسه باشند. در مجموعهٔ مشاهدات زیر:

A B C A B A C D

ب- نمرات ریاضی ۷ دانشجو به صورت زیر گزارش شده است:

۹ ۱۱ ۱۹ ١٦ ۱۳ ۱۷ ۱۲

ج- تعداد کتب ۹ کتابخانه به صورت زیر داده شده است (ارقام به ده هزار)

٦ ۹ ۷ ٥ ۱ ۳ ٤ ۱ ۳

مجموعهٔ مشاهدات الف یک مد، مشاهدات ب فاقد مد است؛ در حالی که مجموعهٔ مشاهدات ج دارای دو مد ۱ و ۳ است.

محاسبهٔ نما در جدول فراوانی: در جدول فراوانی برای محاسبهٔ نما (مد) می‌توان از فرمول زیر استفاده کرد:

$$M_{o} = 3M_n - \bar{X}$$

در مثال ۲-۱۰ داریم:

$$M_n = 60 \qquad \bar{X} = 62$$

$$M_{o} = 3M_n - \bar{X} = (3\times 60) - 62 = 118$$

الف- مقایسهٔ میانگین، میانه و نما

در بعضی موارد، یکی از معیارهای مطرح شده نسبت به بقیه برتری دارد. پاره‌ای از این موارد در زیر آمده است:

۱- یکی از مزیت‌های میانگین نسبت به میانه و نما این است که بر کلیهٔ داده‌ها متکی است، ولی میانه و نما این چنین نیستند. مثلاً داده‌های زیر را در نظر بگیرید:

۲٥۰ ۳۰۰ ٥۰۰ ٥٥۰ ٦۰۰ ٦۰۰ ۷۰۰

در این مثال داریم:

$$\bar{X} = \frac{250+300+500+550+600+600+700}{7} = 500$$

$$M_n = 550 \qquad M_o = 600$$

این در صورتی است که اگر عدد ۷۰۰ را با ۹۰۰ عوض کنیم، میانگین تغییر می‌کند $\left(\bar{X} = 528/6\right)$، اما میانه و نما هیچ تغییری نمی‌کنند.

۲- مزیت دوم میانگین این است که نسبت به میانه و نما پایدارتر است؛ به عبارت دیگر اگر از جامعه‌ای نمونه‌های مختلف گرفته شود و برای هر نمونه مقادیر میانگین، میانه و نما محاسبه شود، مقادیر به دست آمده برای هر کدام در نمونه‌های مختلف متفاوت است؛ به عبارت دیگر می‌گوییم، میانگین میانه و نما دستخوش نوسانات نمونه‌ای است، ولی ثابت می‌شود نوسانات میانگین از میانه و نما کم‌تر است.

۳- یکی از مزیت‌های میانه در صورتی که یک عنصر غیرطبیعی وارد جامعه یا نمونه شود، روی میانه تأثیرگذار نیست، ولی روی میانگین مؤثر است.

مثال ۲-۱۵: مبالغ زیر، مقدار جریمه‌های دیرکرد تحویل کتاب به بخش امانت مربوط به ۱۲ نفر است. این مقادیر به شرح زیر است:

۲۵۰ ۲۵۰ ۵۰۰ ۷۵۰ ۷۵۰ ۷۵۰ ۷۵۰ ۱۰۰۰ ۱۰۰۰ ۱۵۰۰ ۱۷۵۰ ۲۰۰۰۰

میانگین اعداد فوق برابر است با:

$$\bar{X} = \frac{250+250+500+750+750+750+750+1000+1000+1500+1750+20000}{12}$$

$$= \frac{29250}{12} = 2437/5$$

در صورتی که با توجه به تعریف میانگین اعلام شود که هر فرد به طور متوسط مبلغ ۲٤۳۷/۵ تومان جریمه شده است، این عدد برای هیچ کدام از ۱۲ نفر قابل قبول نخواهد بود؛ در حالی که اگر میانه را محاسبهٔ کنیم، خواهیم داشت:

$$M_n = \frac{750+750}{2} = 750$$

پس اگر اعلام شود که متوسط پرداخت، برابر با ۷۵۰ تومان بوده است، رقم مذکور برای اکثریت افراد قابل قبول است؛ به عبارت دیگر در مواردی که عناصر غیرهمگن (غیرمعمول) وارد جامعه شود، بهتر است از میانه به جای میانگین استفاده شود؛ زیرا میانه کم‌تر دستخوش عوامل غیرطبیعی جامعه قرار می‌گیرد.

٤- یکی دیگر از کاربردهای میانه، مواردی است که یکی از ابتدا یا انتهای جامعه در دسترس نباشد.

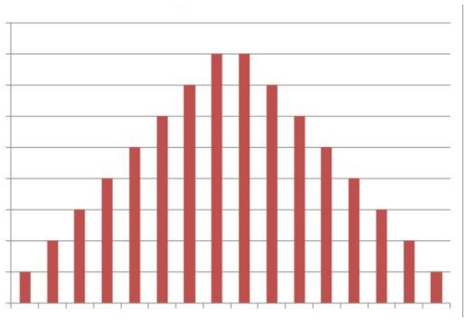
مثال ۲–١٦: تعداد عضویت در یک کتابخانهٔ عمومی بزرگ طی یک ماه گذشته در جدول زیر داده شده است:

تعداد روزها	تعداد عضویت‌ها
۵	۱۰ و کم‌تر
۱۱	۱۱–۲۰
۸	۲۱–۳۰
٤	۳۱–٤۰
۲	٤۱ و بیشتر

در این مثال اگر بخواهیم میانگین را حساب کنیم، غیرممکن خواهد بود؛ زیرا در طبقه‌های اول و آخر قادر به محاسبهٔ نمایندهٔ طبقه نیستیم. اما محاسبهٔ میانه با توجه به تعاریفی که قبلاً ذکر شده، امکان‌پذیر است.

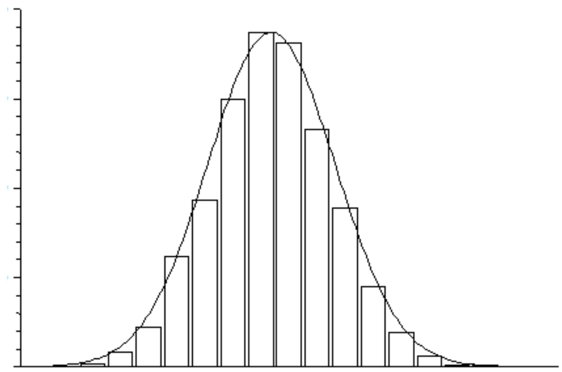
ب- رابطهٔ تجربی بین میانگین، میانه و نما

در صورتی که هیستوگرام فراوانی یک توزیع به صورت زیر باشد، آن را یک توزیع متقارن می‌نامیم.



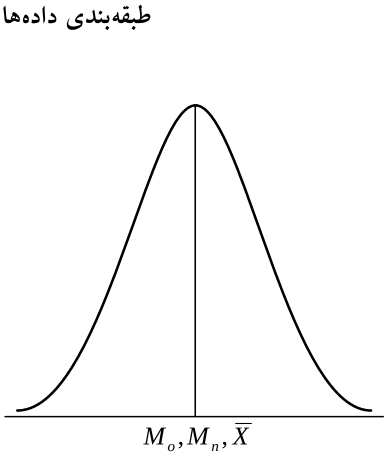
نمودار ۲–۱٤: نمودار هیستوگرام مربوط به یک توزیع متقارن

در صورتی که در هیستوگرام فوق، تعداد طبقات زیاد باشد و نمودار چندبر فراوانی را رسم کنیم، چندبر فراوانی رسم شده بسیار شبیه یک منحنی خواهد بود که مقطعی شبیه یک زنگ دارد.



نمودار ۲–۱۵: منحنی زنگی شکل رسم شده بر نمودار هیستوگرام مربوط به یک توزیع متقارن

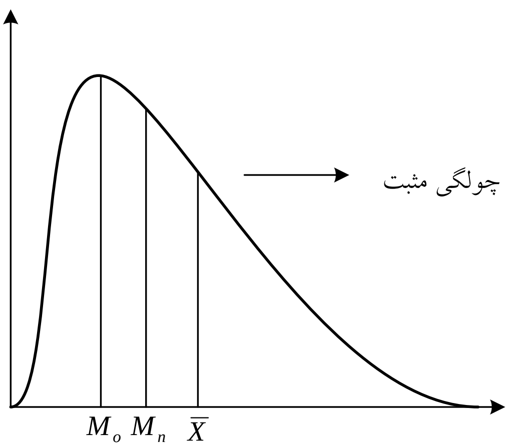
حال اگر از یک جامعهٔ متقارن نمونه‌گیری شود و مقدار میانه، میانگین و نما در آن محاسبهٔ شود، این مقادیر معمولاً با هم برابر نیستند، اما اختلاف چندانی با هم ندارند. البته این تفاوت ناچیز است و به طور حدودی می‌توان گفت که میانه، نما و میانگین بر هم منطبق هستند. در نمودار ۲–۱٦ میانه، میانگین و نما در یک توزیع متقارن نشان داده شده است.



نمودار ۲–۱۶: میانه، میانگین و نما در منحنی نرمال

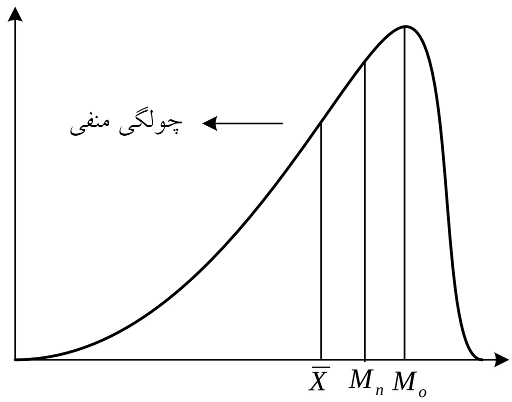
حال اگر بعد از انجام نمونه‌گیری و محاسبۀ میانگین، میانه و نما مشاهده شود که این مقادیر با هم متفاوت بوده و مقدار این تفاوت به ناچیزی قبل نیست، دیگر با توزیع متقارن مواجه نیستیم. چنین توزیع‌هایی را که متقارن نیستند، چوله می‌گویند. در این توزیع‌ها میانه، بین نما و میانگین قرار دارد.

چنانچه میانه از نما بزرگ‌تر باشد، با توزیعی روبه‌رو هستیم که اصطلاحاً به آن چوله به راست (دارای چولگی مثبت) می‌گوییم. در نمودار ۲–۱۷ میانگین، میانه و نما در یک نمونه توزیع چوله به راست مشخص شده است.



نمودار ۲–۱۷: میانه، میانگین و نما در یک توزیع چوله به راست

و اگر میانه از نما کوچک‌تر بود با توزیعی روبرو هستیم که اصطلاحاً به آن چوله به چپ (دارای چولگی منفی) می‌گوییم. در نمودار ۲–۱۸ میانگین، میانه و نما در یک نمونهٔ توزیع چوله به چپ مشخص شده است.



نمودار ۲–۱۸: میانه، میانگین و نما در یک توزیع چوله به چپ

توجه داشته باشید که در یک توزیع که چولگی (کجی) آن زیاد به نظر می‌رسد، از بین شاخص‌های گرایش به مرکز داده‌ها، استفاده از میانه توصیه می‌شود؛ زیرا همان طور که پیش‌تر توضیح داده شد، بر خلاف میانگین حسابی کلیهٔ اعداد بر آن تأثیر نمی‌گذارند.

۲–۳–۲ شاخص‌های پراکندگی

شاخص‌های مرکزی تنها یک منطقه را به عنوان محل تمرکز داده‌ها معرفی می‌کنند. حال آنکه ممکن است دو دسته داده با پراکندگی‌های متفاوت دارای میانگین برابری داشته باشند؛ به عبارتی نیاز به شاخصی برای نمایش میزان پراکندگی داده‌ها خواهیم داشت. در ادامه به معرفی این پارامترها یا شاخص‌ها می‌پردازیم.

اندازه‌های پارامترهای مرکزی مانند میانه، میانگین و نما برای خلاصه کردن اطلاعات به کار می‌روند. آن‌ها به این دلیل مهم هستند که اطلاعاتی دربارهٔ جامعه ارائه می‌دهند، ولی علاوه بر اندازه‌های پارامترهای مرکزی، سایر معرف‌هایی که نمایانگر پراکندگی هستند و ویژگی‌های مشاهدات را توصیف می‌کنند، لازم است بررسی شوند که مهم‌ترین آن‌ها دامنه، واریانس، انحراف معیار و ضریب تغییر است.

۱– دامنه

ساده‌ترین روش برای توصیف پراکندگی، دامنه است. دامنهٔ تغییرات، عبارت است از تفاضل بزرگ‌ترین داده از کوچک‌ترین داده و آن را با R نشان می‌دهند؛ چون از اختلاف بین بزرگ‌ترین و کوچک‌ترین به دست می‌آید، اطلاعات بسیار کمی از هر جامعه ارائه می‌دهد.

$$R = X_{\text{Max}} - X_{\text{min}}$$

که در آن X_{Max} عبارت است از بزرگ‌ترین داده و X_{min} عبارت است از کوچک‌ترین دادهٔ موجود در نمونه.

توجه کنید که این شاخص معیار خوبی برای محاسبهٔ میزان پراکندگی داده‌ها نیست؛ زیرا در محاسبه، تنها کوچک‌ترین و بزرگ‌ترین داده وارد می‌شود.

مثال ۲–۱۷: نمرات ۱۰ دانشجو در دو کلاس مختلف در درس آمار به قرار زیر است:

کلاس A: ۲۰ ۱۶ ۱۶ ۱۴ ۱۲ ۱۲ ۱۲ ۴ ۴ ۰

کلاس B: ۱۴ ۱۳ ۱۳ ۱۲ ۱۲ ۱۲ ۹ ۹ ۸ ۸

با محاسبهٔ مقادیر $\bar{X}$ و M_n و M_o دیده می‌شود که برای هر دو کلاس داریم:

$\bar{X}=11 \quad , \quad M_n=12 \quad , \quad M_o=12$

اما با توجه به مقادیر تک تک نمرات، واضح است که میزان پراکندگی نمرات دو کلاس کاملاً متفاوت است و برای این منظور نیاز به شاخص‌های پراکندگی است تا مطلب را بتوان با مقایسهٔ آن‌ها نشان داد.

برای دو کلاس مقدار دامنهٔ تغییرات را محاسبهٔ می‌کنیم:

کلاس A: $R=20-0=20$

کلاس B: $R=14-8=6$

مثال ۲-۱۸: سن ۵ نفر از کارکنان در یک کتابخانه به صورت زیر گزارش شده است:

٥٤ ۳۷ ۲۰ ۲٥ ۲۱

دامنهٔ سن کارکنان برابر است با: $X_{\min}=20$ و $X_{\text{Max}}=54$ پس:

$R=54-20=34$

۲- انحراف متوسط

یکی از معیارهای اندازه‌گیری پراکندگی میانگین انحرافات از میانگین (انحراف متوسط) است که با $M.D$ نشان داده می‌شود. برای محاسبهٔ آن نیز همان طور که از اسم آن پیداست، باید تمامی داده‌ها را جداگانه از میانگین کم کرده و بین اعداد به دست آمده میانگین بگیریم.

اکنون فرض کنید که مجموعه‌ای از داده‌ها به صورت $x_1, x_2, \ldots, x_3$ در اختیار داریم و می‌خواهیم میزان اختلافات این داده‌ها را با میانگین $\left(\bar{X}\right)$ محاسبه کنیم. برای این منظور باید تک‌تک داده‌ها را از میانگین کم کرده و مقادیر $\left(\bar{X}-x_i\right)$ را برای داده‌ها به دست آوریم. اگر بخواهیم بین اعداد حاصل از

$\left(\bar{X}-x_i\right)$ میانگین بگیریم، این میانگین همواره برابر با صفر خواهد بود و علت آن نیز توزیع متقارن داده‌ها حول میانگین است که باعث می‌شود مجموع اعداد حاصل از $\left(\bar{X}-x_i\right)$ صفر شود.

به عنوان مثال، فرض کنید که ۵ داده به صورت ۱۱، ۸، ٦، ۳، ۲ در اختیار داریم. میانگین این مقادیر عبارت خواهد بود از $\bar{X}=\dfrac{2+3+6+8+11}{5}=6$؛ حال مقادیر اختلاف از میانگین را محاسبه می‌کنیم و خواهیم داشت:

$$\bar{X}=X_i:(2-6),(3-6),(6-6),(8-6),(11-6)=-4,-3,0,2,5$$

حال اگر میانگین اعداد حاصل را به دست آوریم، خواهیم داشت:

$$\bar{X}=\frac{(-4)+(-3)+(0)+(2)+(5)}{5}=0$$

سؤالی که در اینجا پیش می‌آید این است که در صورتی که جواب میانگین فوق، همواره صفر باشد، چگونه می‌توان میانگین انحرافات از میانگین (انحراف متوسط) را به عنوان یکی از معیارهای پراکندگی به حساب آورد؟

برای حل شدن مشکل صفر بودن میانگین اختلافات، از میانگین قدر مطلق استفاده می‌کنیم و مقدار $\bar{X}-X_i$ را به صورت $\left|\bar{X}-X_i\right|$ می‌نویسیم. در این صورت با توجه به تعریف قدر مطلق دیگر برای تفاضلات مقدار منفی حاصل نخواهد شد و انحراف متوسط حاصلی به غیر از صفر خواهد داشت.

در مثال فوق داریم:

$$\left|\bar{X}-X_i\right|=|2-6|,|3-6|,|6-6|,|8-6|,|11-6|=4,3,0,2,5$$

پس فرمول محاسباتی انحرافات از میانگین به صورت زیر خواهد بود:

$$M.D=\frac{\sum_{i=1}^{n}\left|\bar{X}-x_i\right|}{n}$$

که در آن مشاهدات با x_i، میانگین مشاهدات با $\bar{X}$ و تعداد مشاهدات با n نشان داده می‌شود.

لذا در مثال فوق مقدار انحراف متوسط برابر خواهد بود با:

$$M.D=\frac{\sum_{i=1}^{n}\left|\bar{X}-x_i\right|}{n}=\frac{4+3+0+2+5}{5}=2/8$$

یعنی هر داده به طور متوسط با میانگین $\left(\bar{X}\right)$ به اندازهٔ ۲/۸ واحد فاصله دارد.

محاسبهٔ انحراف متوسط در جدول فراوانی: هرگاه بخواهیم انحراف متوسط یک سری دادهٔ آماری را که در جدول فراوانی طبقه‌بندی شده‌اند را به دست آوریم، فرمول محاسباتی مربوط به انحراف متوسط، به صورت زیر تغییر می‌کند:

$$M.D=\frac{\sum_{i=1}^{n}f_i\left|\bar{X}-x_i\right|}{n}$$

که در آن x_i معرف نمایندهٔ هر رده است که عضو میانی رده و f_i فراوانی مربوط به هر طبقه است.

مثال ۲–۱۹: تعداد عضویت‌های در روز مربوط به یک کتابخانه طی مدت ۲۵ روز در جدول زیر آورده شده است؛ داریم:

تعداد عضویت‌ها	تعداد روز $\left(f_i\right)$	x_i	$f_i x_i$	$\left\|\bar{X}-x_i\right\|$	$f_i\left\|\bar{X}-x_i\right\|$
۰–۲	٤	۱	٤	۳/۷٦	۱۵/۰٤
۲–٤	٦	۳	۱۸	۱/۷٦	۱۰/۵٦
٤–٦	۷	۵	۳۵	۰/۲٤	۱/٦۸
٦–۸	۵	۷	۳۵	۲/۲٤	۱۱/۲
۱۰–۸	۳.	۹	۲۷	٤/۲٤	۱۲/۷۲
مجموع	۲۵		۱۱۹		۵۱/۲

از قبل می‌دانیم:

$$\bar{X}=\frac{x_1f_1+x_2f_2+\ldots+x_kf_k}{n}=\frac{1}{n}\sum_{i=1}^{n}x_if_i=\frac{119}{25}=4/76$$

$$M.D=\frac{\sum_{i=1}^{n}f_i\left|\bar{X}-x_i\right|}{n}=\frac{51/2}{25}=2/048\approx 2$$

یعنی هر داده به طور متوسط با میانگین $\left(\bar{X}=4/76\right)$ تقریباً به اندازهٔ ۲ واحد فاصله دارد.

مزیت $M.D$ این است که داده‌ها در محاسبهٔ آن نقش دارند، ولی از آنجایی که کار با قدر مطلق قدری مشکل است، از واریانس و انحراف معیار استفاده می‌کنیم.

۳- واریانس[۱]

یکی از مهم‌ترین شاخص‌های پراکندگی اعداد، واریانس است. واریانس در لغت به معنی تفاوت و تغییر است. در فرمول محاسباتی انحراف متوسط از قدر مطلق استفاده کرده بودیم که به دلیل مشکلات کار با قدر مطلق خود ضعف این معیار به حساب می‌آید. برای رفع این مشکل به جای اینکه از $\left|\bar{X}-x_i\right|$ استفاده کنیم، مقدار $\bar{X}-x_i$ را به توان دو می‌رسانیم و معیاری دیگر را معرفی می‌کنیم که واریانس نام دارد.

اگر با داده‌های خام و طبقه‌بندی نشده سروکار داشته باشیم، آنگاه واریانس از فرمول‌های زیر به دست می‌آید:

$$S^2=\frac{\sum_{i=1}^{n}\left(\bar{X}-x_i\right)^2}{n} \quad \text{یا} \quad S^2=\frac{\sum_{i=1}^{n}x_i^2}{n}-\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2=\frac{\sum_{i=1}^{n}x_i^2}{n}-\left(\bar{X}\right)^2$$

---

۱. Variance

که در آن مشاهدات با x_i، میانگین مشاهدات با $\bar{X}$ و تعداد مشاهدات با n نشان داده می‌شود.

در حقیقت از تک‌تک مشاهدات، میانگین را کم و پس از به توان دو رساندن آن‌ها، جمع و بر تعداد مشاهدات تقسیم می‌کنیم.

مثال ۲-۲۰: سود یک کتابفروشی در ٦ ماه گذشته به صورت زیر گزارش شده است (ارقام به میلیون تومان). واریانس نمونه را محاسبهٔ کنید.

٦ ۲ ۱ ۳ ٤ ۲

راه حل اول:

با استفاده از فرمول: $S^2 = \frac{\sum_{i=1}^{n}\left(\bar{X} - x_i\right)^2}{n}$

برای محاسبهٔ واریانس، ابتدا میانگین را به دست می‌آوریم.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+4+3+1+2+6}{6} = \frac{18}{6} = 3$$

$$S^2 = \frac{\sum_{i=1}^{n}\left(\bar{X} - x_i\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(3 - x_i\right)^2}{6}$$

$$= \frac{[(2-3)^2 + (4-3)^2 + (3-3)^2 + (1-3)^2 + (2-3)^2 + (6-3)^2]}{6}$$

$$= \frac{(1+1+0+4+1+9)}{6} = \frac{16}{6} = 2/67$$

راه حل دوم:

با استفاده از فرمول: $S^2 = \frac{\sum_{i=1}^{n} \bar{X}_i^2}{n} - \left(\frac{\sum_{i=1}^{n} \bar{X}_i}{n}\right)^2$

برای محاسبهٔ واریانس، ابتدا میانگین را به دست می‌آوریم.

$$\bar{X}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{2+4+3+1+2+6}{6}=\frac{18}{6}=3$$

اکنون مقادیر اولیه (داده‌ها) را به توان ۲ می‌رسانیم و داریم:

$$X_i^2=2^2,4^2,3^2,1^2,2^2,6^2=4,16,9,1,4,36$$

اکنون میانگین توان دوم‌ها را محاسبهٔ می‌کنیم، داریم:

$$\frac{\sum_{i=1}^{n}X_i^2}{n}=\frac{4+16+9+1+4+36}{6}=\frac{70}{6}=11/67$$

با استفاده از فرمول داریم:

$$S^2=\frac{\sum_{i=1}^{n}X_i^2}{n}-\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)^2=11/67-\left(3^2\right)=2/67$$

نتایج به دست آمده از روش‌های اول و دوم یکسان است.

ویژگی‌های مهم واریانس عبارتند از:

- همواره بیشتر یا برابر صفر است.
- اگر به همهٔ مشاهدات مقدار ثابتی اضافه یا کم کنیم، مقدار واریانس تغییر نمی‌کند.
- اگر همهٔ مشاهدات را در مقدار ثابت k ضرب یا تقسیم کنیم، مقدار واریانس در k^2 ضرب یا تقسیم می‌شود.

محاسبۀ واریانس در جدول فراوانی: هرگاه بخواهیم واریانس یک سری دادۀ آماری را که در جدول فراوانی طبقه‌بندی شده‌اند، به دست آوریم، از یکی از فرمول‌های زیر استفاده می‌کنیم:

$$S^2 = \frac{\sum_{i=1}^{n} f_i \left(\bar{X} - x_i\right)^2}{n} \quad , \quad S^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{n} - \bar{x}^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} f_i x_i}{n}\right)^2$$

مثال ۲-۲۱: تعداد ۷۵ دانشجوی لیسانس را در طول مدت تحصیل‌شان در دانشکده‌ای را مورد مطالعه قرار داده‌ایم و تعداد کتاب‌هایی را که در طی مدت ۸ ترم از کتابخانه به امانت گرفته‌اند، ثبت کرده‌ایم و نتایج در جدول زیر طبقه‌بندی شده است.

برای محاسبۀ واریانس:

روش اول: از فرمول $S^2 = \frac{\sum_{i=1}^{n} f_i \left(\bar{X} - x_i\right)^2}{n}$ استفاده می‌کنیم و داریم:

تعداد کتاب‌ها	تعداد دانشجویان (f_i)	x_i	$x_i - \bar{X}$	$f_i\left(x_i - \bar{X}\right)^2$
۱-۳۰	۳	۱۵/۵	-۸۲/۸	۲۰۵۶۷/۵۲
۳۱-۶۰	۹	۴۵/۵	-۵۲/۸	۹۰۲۵۰/۵۶
۶۱-۹۰	۲۰	۷۵/۵	-۲۲/۸	۱۰۳۹۶/۸
۹۱-۱۲۰	۲۲	۱۰۵/۵	-۷/۲	۱۱۴۰/۴۸
۱۲۱-۱۵۰	۱۳	۱۳۵/۵	۳۷/۲	۱۷۹۸۹/۹۲
۱۵۱-۱۸۰	۸	۱۶۵/۵	۶۷/۲	۳۶۱۲۶/۷۲
جمع	$75 = n$			$\sum_{i=1}^{n} f_i\left(\bar{X} - x_i\right)^2 = 111312$

لذا مقدار واریانس عبارت خواهد بود از:

$$S^2=\frac{\sum_{i=1}^{n} f_i\left(\bar{X}-x_i\right)^2}{n}=\frac{111312}{75}=38/525\approx 38/5$$

روش دوم: از فرمول $S^2=\frac{\sum_{i=1}^{n} f_i x_i^2}{n}-\bar{x}^2=\frac{\sum_{i=1}^{n} f_i x_i^2}{n}-\left(\frac{\sum_{i=1}^{n} f_i x_i}{n}\right)^2$ استفاده می‌کنیم، داریم:

تعداد کتاب‌ها	تعداد دانشجویان f_i	x_i	$f_i x_i$	$f_i x_i^2$
١-٣٠	٣	١٥/٥	٤٦/٥	٧٢٠/٧٥
٣١-٦٠	٩	٤٥/٥	٤٠٩/٥	١٨٦٣٢/٢٥
٦١-٩٠	٢٠	٧٥/٥	١٥١٠	١١٤٠٠٦
٩١-١٢٠	٢٢	١٠٥/٥	٢٣٢١	٢٤٤٨٦٥/٥
١٢١-١٥٠	١٣	١٣٥/٥	١٧٦١/٥	٢٣٨٦٨٣/٢٥
١٥١-١٨٠	٨	١٦٥/٥	١٣٢٤	٢١٩١٢٢
جمع	$75=n$		$\sum_{i=1}^{n} f_i x_i=7372/5$	$\sum_{i=1}^{n} f_i x_i^2=836028/75$

$$\bar{X}=\frac{\sum_{i=1}^{n} f_i x_i}{n}=\frac{7372/5}{75}=98/3$$

$$S^2=\frac{\sum_{i=1}^{n} f_i x_i^2}{n}-\bar{x}^2=\frac{836028/75}{75}-\left(98/3\right)^2=38/525\approx 38/52$$

٤- انحراف معیار

از آنجا که واحد اندازه‌گیری واریانس توان دوم واحد اصلی متغیر است و بعلاوه تفسیر آن زیاد ساده نیست، از جذر آن، که انحراف معیار نامیده می‌شود،

استفاده می‌شود. اگر S^2 واریانس مشاهدات باشد، S یا انحراف معیـار، دارای فرمول زیر است:

$$\text{انحراف معیار} = \sqrt{\text{واریانس}}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(\bar{X} - x_i\right)^2}{n}}$$

لذا کافی است بعد از محاسبهٔ واریانس که در بخش قبل توضیح داده شد، از مقدار به دست آمده جذر بگیریم، آنگاه حاصل انحراف معیار است. هرگاه جدول فراوانی در اختیار داشتیم و هدفمان محاسبهٔ انحراف معیار بود نیز، ابتدا واریانس را محاسبهٔ می‌کنیم و سپس از آن جذر می‌گیریم.

مثال ۲–۲۲: اگر حقوق دریافتی مدیران ۵ کتابخانه به صورت زیر باشد (ارقام به صدهزار تومان)

٦ ۷ ۵ ۳ ٤

میانگین، واریانس و انحراف معیار حقوق‌های دریافتی را به دست آورید.

$$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i = \frac{۱}{۵}\sum_{i=۱}^{۵} x_i = \frac{۲۵}{۵} = ۵$$

$$S^{۲} = \frac{۱}{n}\sum_{i=۱}^{n}\left(x_i - \bar{x}\right)^{۲} = \frac{۱}{۵}\sum_{i=۱}^{۵}\left(x_i - ۵\right)^{۲} = \frac{۱}{۵}\left(۱ + ۴ + ۰ + ۴ + ۱\right) = ۲$$

$$S = \sqrt{S^{۲}} = \sqrt{۲}$$

مثال ۲–۲۳: انحراف معیار و میانگین را برای تعداد مراجعات ٤۰۰ نفر عضو کتابخانه‌ای به کتابخانه طی مدت ۲ ماه به دست آورید و درصد تعداد مراجعانی را که بین میانگین و یک برابر انحراف معیار هستند، برآورد کنید.

جمع	۲۷–۳۶	۲۳–۲۶	۲۰–۲۲	۱۷–۱۹	۱۴–۱۶	۱۲–۱۳	۱۰–۱۱	۵–۹	تعداد مراجعه
	۳۱/۵	۲۴/۵	۲۱	۱۸	۱۵	۱۲/۵	۱۰/۵	۷	نقاط میانی x
۴۰۰	۵۵	۳۶	۴۲	۵۷	۷۲	۶۲	۵۸	۱۸	تعداد افراد f

حل:

$$\sum xf = (7\times 18)+(10/5\times 58)+\ldots = 7112/5$$

$$\sum x^2 f = (7^2\times 18)+(10/5^2\times 58)+\ldots = 146336/75$$

$$\bar{x} = \frac{7112/5}{400} = 17/781 \approx 17/8$$

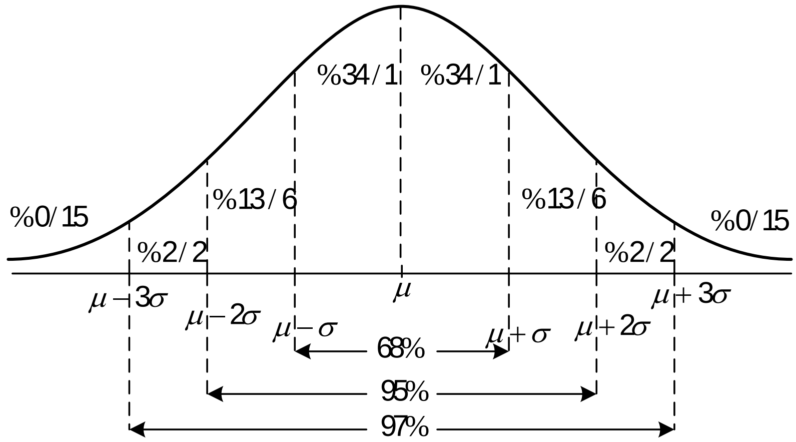
$$S = \sqrt{\frac{146336/75}{400} - \left(\frac{7112/5}{400}\right)^2} = 7/048 \approx 7$$

الف) تعداد مراجعه در طرفین میانگین به فاصلهٔ یک انحراف معیار از آن برابر $17/781 \pm 7/048$ است؛ یعنی ۱۰/۷۳ تا ۲۴/۸۳ فرد. برای یافتن تعداد مراجعان در این فاصله می‌توان از فراوانی تجمعی یا به طور مستقیم از جدول فراوانی به طریق زیر استفاده کرد:

۱۰/۷۳ در طبقهٔ ۹/۵ تا ۱۱/۵ قرار می‌گیرد و تعداد ۵۸ فرد مراجعه‌کننده، در این طبقه وجود دارد. بنابراین بین ۱۰/۷۳ تا ۱۱/۵ نفر به تقریب $\frac{11/5-10/73}{2}\times 58 = 22/3$ فرد وجود دارد. بین ۱۱ تا ۲۲ مراجعه تعداد $62+72+57+42=233$ است، بین ۲۲/۵ و ۲۴/۸۳ تعداد مراجعان برابر $22/3+233+21/0=276/3$ و درصد آن $\frac{276/3}{400}\times 100 = 61/1\%$ است.

در اکثر توزیع‌ها حدود $\frac{2}{3}$ مقادیر در فاصلهٔ $(\bar{X}-S, \bar{X}+S)$ حدود ۹۵٪ مقادیر در فاصلهٔ $(\bar{X}-2S, \bar{X}+2S)$ و تقریباً همه مقادیر در فاصلهٔ $(\bar{X}-3S, \bar{X}+3S)$ قرار می‌گیرند.

نسبت‌های دقیق به شکل توزیع بستگی دارد. روش بالا برای تقریب توزیع‌های متقارن به شکل "زنگ" است.



نمودار ۲–۹: پراکندگی داده‌ها در اطراف میانگین (μ) نسبت به انحراف معیار (σ)

۵- ضریب تغییر

در بسیاری از تحقیقات، ضرورت دارد برای توصیف مشاهدات، پراکندگی به صورت کسری از میانگین بیان شود. یکی از معیارهای پراکندگی نسبی که ضریب تغییر یا تعیین (CV)[۱] نامیده می‌شود، به صورت زیر تعریف می‌شود:

$$C \cdot V = \frac{S}{\bar{x}}$$

ضریب پراکندگی کاربردهایی دارد که واریانس و انحراف معیار فاقد آن‌ها هستند. کاربرد آن وقتی است که بخواهیم دو جامعهٔ آماری را با هم مقایسه کنیم.

مثال ۲–۲٤: برای مقایسه متوسط سود دو کتابفروشی اطلاعات زیر در سال گذشته برحسب میلیون تومان به دست آمده است. پس از محاسبهٔ ضریب تغییر برای هر کتابفروشی آن‌ها را مقایسه کنید.

---

۱. CV: coefficient of variation

کتابفروشی الف: ۱/٤٥ ۱/۲٥ ۲/۷٥ ۲/۰٥ ۱/۹٥ ٦/۷٥ ۲/٥۰ ۱/۰۰ ۲/۲٥ ۱/۷٥ ۲/٥۰ ۳/۲٥
کتابفروشی ب: ٤/۰۰ ۱/۸٥ ٥/۰۰ ۱/۰۰ ۲/٥۰ ۳/٥۰ ۹/٥۰ ٤/۰۰ ۱/۳٥ ۷/۲٥ ٥/۷٥ ۱/٦٥

الف– $\bar{X}_۱ = ۲/۴۵۴$ $\quad s_۱ = ۱/۴۳۵$ $\quad CV_۱ = ۰/۵۸۴۷$

ب– $\bar{X}_۲ = ۳/۹۴۵۸$ $\quad s_۲ = ۲/۴۷۵$ $\quad CV_۲ = ۰/۶۲۷۲$

چون ضریب تغییر کتابفروشی الف کم‌تر از کتابفروشی ب است، لذا متوسط سود کتابفروشی الف از پراکندگی کم‌تری برخوردار است.

خلاصهٔ فصل دوم

در این فصل، جدول توزیع فراوانی برای خلاصه کردن اطلاعات داده‌ها ارائه شده و ضمن تشکیل جدول توزیع فراوانی نحوهٔ محاسبهٔ فراوانی نسبی، فراوانی نسبی تجمعی و نمایندهٔ رده‌ها ارائه شد و به موازات آن در بخش دوم انواع نمودارهای آماری با ارائهٔ مثال شرح داده شده‌اند و در انتها نحوهٔ محاسبهٔ شاخص‌های مرکزی و پراکندگی برای داده‌ها در جدول توزیع فراوانی ارائه شده است.

خودآزمایی تشریحی فصل دوم

۱- مشاهدات زیر از یک بررسی آماری به دست آمده است آن‌ها را به صورت مجموع با استفاده از نماد سیگما بنویسید.

$x_۱, x_۲, x_۳, \ldots, x_n$

$1, 2, 3, 4, 5, 6, \ldots, 125$
$-3, -2, -1, ۰, 1, 2, \ldots, 100$

۲- نمای هر یک از گروه‌های زیر را به دست آورید.

الف): ۳ ۱٤ ٦ ۹ ٤ ۷ ۸ ۵ ۳ ۱۱

ب): ۹ ۷ ۵ ٤ ۳ ۲ ۱ ۰ ۱- ۲- ۳-

۳- کتاب‌های یک کتابخانه با حروف A، B، C، D و E برچسب‌گذاری شده و اطلاعات زیر در یک بررسی آماری از ۲۰ نمونه به دست آمده است. نما یا مد نمونه را به دست آورید.

A B A B C D E E D B A C
D C E A B C D A

٤- میانهٔ داده‌های زیر را به دست آورید.

الف): ٤ ۳ ۲ ۱ ۰ ۱- ۲- ۳- ٤-

ب): ۱٦ ۲۱ ۲۹ ۳۳ ۸٦ ۱۰۵ ۱۵۵ ۸۲ ۷۵ ۹۵

۵- میانهٔ مشاهدات زیر برابر با ۵ است.

$x_۱ \quad x_۲ \quad x_۳ \quad \ldots \quad x_n$

اگر به همهٔ مشاهدات، مقدار ثابت ۷ اضافه کنیم، میانهٔ جدید چقدر می‌شود؟

٦- حقوق ۷ کارمند بر حسب میلیون ریال به صورت زیر داده شده است. متوسط یا میانگین حقوق آن‌ها را محاسبهٔ کنید.

۲/۰ ۳/۰ ۳/۵ ٤/۵ ٦/۵ ٤/۷۵ ۵/۵

۷- در جدول فراوانی زیر، مقدار میانگین حسابی را به دست آورید.

کتاب	۱	۲	۳	٤	۵	٦
قیمت کتاب (هزار ریال)	۲/۷۵	۱/۲۵	۳/۵	۱/۵	۲/۵	٤/۷۵
تعداد (فراوانی)	٤	٦	۲	۵	۳	۵

۸- اگر میانگین حسابی اعداد $x_1, x_2, x_3, \ldots, x_n$ برابر با ۵ باشد، میانگین حسابی اعداد زیر را محاسبهٔ کنید.

الف) $x_1+1, x_2+1, x_3+1, \ldots, x_n+1$

ب) $x_1-7, x_2-7, x_3-7, \ldots, x_n-7$

ج) $2x_1+1, 2x_2+1, 2x_3+1, \ldots, 2x_n+1$

۹- قیمت ۱۲ کتاب جدید در کتابخانه‌ای به صورت زیر داده شده است (ارقام به هزار تومان):

۱/۸۰ ۱/۱۵ ۱/۸۵ ۱/۲۰ ۲/۳۵ ۱/٤۵ ۲/۰۰ ۱/۲۵ ۳/۰۰ ۲/۵۰ ۱/۵۰ ۱/۷۵

الف) دامنهٔ قیمت‌ها را به دست آورید.

ب) متوسط یا میانگین قیمت‌ها را به دست آورید.

ج) واریانس و انحراف معیار قیمت‌ها را به دست آورید.

۱۰- متوسط حقوق ۵۰ نفر از کارکنان ۳۰۰ هزار تومان است.

الف) اگر به حقوق تمام کارمندان ۱۵ هزار تومان اضافه شود، متوسط حقوق جدید چقدر است؟

ب) اگر واریانس و انحراف معیار قبل از افزایش به ترتیب ٦/۲۵ و ۲/۵ باشد، واریانس و انحراف معیار جدید را حساب کنید.

۱۱- اگر s^2 واریانس $x_1, x_2, \ldots, x_n$ باشد، نشان دهید که:

$$s^2 = \frac{1}{n}\sum\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum x_i^2 - x^{-2}$$

۱۲- کتابخانه‌ای در دو سال متوالی کتاب‌هایی با قیمت‌های متفاوت خریداری کرده است. اگر قیمت ٦ کتاب در سال اول و ۹ کتاب در سال دوم به صورت زیر باشد (ارقام به هزار تومان):

سال اول: ۲۷۰ ۲۱۰ ۲٦۵ ۲۳۰ ۲٤۰ ۲٦۰

سال دوم: ۲٤۰ ۲٦۵ ۲٤۰ ۲٦۰ ۲٦۵ ۲٤۰ ۳۰۵ ۲۲۸ ۱۹۰

ضریب تغییر را برای دو سال متوالی محاسبهٔ و آن‌ها را مقایسه کنید.

۱۳- انواع شاخص‌ها یا پارامترهای مرکزی و پراکندگی را نام ببرید.

۱٤- اگر تعداد کتب در طول ٦ روز هفته از دو کتابخانه به صورت زیر به امانت گرفته شده باشد، نما و میانهٔ آن‌ها را به دست آورید (ارقام به هزار).

روز	شنبه	۱ شنبه	۲ شنبه	۳ شنبه	٤ شنبه	۵ شنبه
کتابخانه مرکزی	۳	٤	٦	۵	۱	۷
کتابخانه ارشاد	۲	۳	٤	۸	۹	۵

۱۵- کتابخانهٔ مرکزی در چهار دوره برگزاری نمایشگاه کتاب، آمار کتاب‌های خریداری شدهٔ خود را به صورت زیر گزارش داده است (ارقام به هزار):

دوره	اول	دوم	سوم	چهارم
تعداد کتب خریداری شده	۲/۵	۱/۵	۲	٤

متوسط کتب خریداری شده را به دست آورید.

۱٦- رشد متقاضیان عضویت در کتابخانه‌ای در ۵ سال متوالی به صورت زیر گزارش شده است:

سال	اول	دوم	سوم	چهارم	پنجم
رشد	۳	۲	٤	۱۰	۱۵

متوسط رشد را به دست آورید.

۱۷- اگر میزان جریمهٔ کتب دیرکرد ۵ نفر از اعضا به صورت زیر گزارش شده باشد:

تعداد کتاب	۲	۱	٤	۲	٦
مبلغ دیرکرد (ریال)	٦۰	۱۰۰	۲٤۰	٤۰۰	۱۲۰

رشد متوسط میزان جریمه را به دست آورید.

۱۸- تعداد صفحات گم شدهٔ ۱۸ کتاب به صورت زیر گزارش شده است:

۱۲ ۱۷ ۵ ٤ ۹ ٦ ۲۰ ۱۱ ۱٤ ۳ ۲۳ ۱۱ ۲۱ ۱٦ ۱۳ ۵ ٤ ۷

میانه، نما و میانگین را به دست آورید.

۱۹- اگر ارزش کتاب‌های ۵ کتابخانهٔ کشور بر حسب میلیارد ریال به صورت زیر گزارش شده باشد:

۵ ۳ ۱ ٤ ۲

الف) دامنه تغییرات، چند میلیارد ریال است؟

ب) متوسط ارزش کتب ۵ کتابخانه را به دست آورید.

ج) پس از محاسبهٔ واریانس، انحراف معیار را به دست آورید.

۲۰- اگر واریانس مشاهدات $X_1, X_2, \ldots, X_{15}$ برابر با ۱۷ باشد، واریانس مشاهدات $X_1+5, X_2+5, \ldots, X_{15}+5$ را به دست آورید.

۲۱- اگر میزان یادگیری واژهٔ خارجی دو گروه از دانشجویان به صورت زیر گزارش شده باشد، ضریب تغییر هر گروه را محاسبهٔ و آن‌ها را مقایسه کنید.

گروه اول: ۳۰ ۳۵ ۱۵ ٤۵ ۲۵

گروه دوم: ۲۵ ۳۵ ۵ ۲۵ ۵۵ ۱۵

۲۲- در جدول توزیع فراوانی زیر، میانه، نما و واریانس را محاسبهٔ کنید.

رده	۰-۲	۲-٤	٤-٦	٦-۸	۸-۱۰
فراوانی	۲	۵	۱۰	۷	٦

فصل سوم

متغیرهای تصادفی و توزیع‌های مهم آماری

هدف کلی

آشنایی با متغیرهای تصادفی و توزیع‌های آن.

هدف‌های یادگیری

پس از مطالعه این فصل، شما باید بتوانید:

۱. متغیر تصادفی را تعریف کنید.

۲. تابع چگالی احتمال را برای متغیر تصادفی گسسته و پیوسته بنویسید.

۳. با توزیع‌های مهم گسستهٔ آماری آشنا شوید.

٤. با توزیع‌های مهم پیوستهٔ آماری آشنا شوید.

۵. ویژگی‌های متغیر نرمال استاندارد را بنویسید.

٦. با نحوهٔ استفاده از جدول نرمال استاندارد و توزیع استودنت آشنا شوید.

۷. امید ریاضی و واریانس متغیر تصادفی را محاسبهٔ کنید.

مقدمه

در یک تقسیم‌بندی آمار می‌توان به آمار توصیفی و استنباطی اشاره کرد. مباحثی که در فصل‌های گذشته شامل نمایش گرافیکی داده‌ها، اندازه‌های گرایش‌های مرکزی و پراکندگی مطرح شد، جزو آمار توصیفی محسوب می‌شوند. در آمار توصیفی تنها داده‌ها خلاصه و توصیف می‌شوند، بدون اینکه بر روی داده‌ها استنباطی صورت گیرد. برای مثال، در تحقیقی که از طریق نمونه‌گیری تصادفی بر روی ۲۰۰ نفر از دانشجویان صورت گرفته است، از آن‌ها خواسته می‌شود متوسط تعداد ساعاتی را که هر هفته صرف استفاده از اینترنت می‌کنند، بیان کنند. اگر داده‌های گردآوری شده به صورت هیستوگرام نشان داده و گرایش مرکز و پراکندگی داده‌ها گزارش شوند، از آمار توصیفی استفاده کرده‌ایم. حال احتمال دارد که نخواهیم این فرایند را اینجا متوقف کنیم. احتمالاً هدف تحقیق این است که متوسط ساعات استفاده از اینترنت را در میان کلیه دانشجویان مورد بررسی قرار دهیم، نه فقط ۲۰۰ دانشجویی که از طریق نمونه‌گیری انتخاب شده‌اند. در واقع بر اساس اطلاعات گردآوری شده از ۲۰۰ نفر از دانشجویان، می‌توان متوسط زمان استفاده از اینترنت را در میان کلیهٔ دانشجویان برآورد کرد. برای تعمیم چنین برآوردی از نمونه به کلیهٔ جامعهٔ دانشجویان، باید از آمار استنباطی استفاده کرد، که در واقع نوعی برآورد محاسبهٔ شده یا حدس خردمندانه است. به طور خلاصه، آمار توصیفی، داده‌های گردآوری شده را خلاصه و توصیف می‌کند، در حالی که آمار استنباطی از داده‌های آماری به دست آمده، به عنوان مبنای برآورد یا استنباط استفاده می‌کنند و قابل ذکر است آمار استنباطی در واقع شناخت روی پارامترهای توزیع‌های آماری است. در ادامه پس از تعریف متغیر تصادفی اقدام به معرفی بعضی از این انواع توزیع‌های آماری خواهد شد.

۳-۱ متغیر تصادفی

برای ارائهٔ تعریف سادهٔ متغیر تصادفی، فرض کنید یک تاس متعادل پرتاب می‌شود. می‌دانیم که فضای نمونهٔ آن برابر است با:

$$S = \{۱, ۲, ۳, ۴, ۵, ۶\}$$

اگر متغیر تصادفی X به صورت زیر تعریف شود:

X: وجه ظاهر شده برای تاس

با توجه به فضای نمونه، مقادیر ممکن برای X مجموعهٔ زیر خواهد بود:

$$A_X = \{۱, ۲, ۳, ۴, ۵, ۶\}$$

که A_X دقیقاً همان S است و یا در پرتاب دو تاس اگر متغیر تصادفی X مجموعه وجه‌های ظاهر شده برای دو تاس باشد، خواهیم داشت:

$$A_X = \{۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۲\}$$

تعریف: تابع حقیقی X که دامنهٔ آن S و بردش زیرمجموعه‌ای از اعداد حقیقی است، یک متغیر تصادفی نامیده می‌شود.

با توجه به تعریف فوق می‌توان گفت که تابع X مجموعهٔ S یا فضای نمونه را که ممکن است عددی نباشد به یک مجموعهٔ عددی A_X تبدیل کند. برای توضیح بیشتر در پرتاب دو سکه اگر متغیر تصادفی X تعداد شیرها در دو پرتاب باشد، می‌توان نوشت:

$$S = \{(H,H),(H,T),(T,H),(T,T)\}$$

بنابراین تعداد شیرها $X: 0,1,2$ و به دنبال آن A_X برابر است با:

$$A_X = \{۰, ۱, ۲\}$$

متغیرتصادفی ممکن است گسسته یا پیوسته باشد. از حیث اینکه فضای نمونه گسسته یا پیوسته باشد، متغیر تصادفی، گسسته و پیوسته خواهد بود. با توجه به نوع متغیر تصادفی در ادامه، توزیع‌های احتمالات گسسته و پیوسته ارائه می‌شود.

۳-۲ توزیع احتمالات گسسته

در تعریف متغیر تصادفی روی فضای نمونهٔ گسسته، مقادیر اختیاری برای X با احتمالات معین است. در پرتاب دو سکه، اگر متغیر تصادفی X تعداد شیرها در دو پرتاب باشد، شانس اینکه X مقدار دو را اختیار کند برابر با $\frac{۱}{۴}$ است. چون احتمال اینکه در پرتاب دو سکه، در هر دو سکه شیر ظاهر شود برابر با $\frac{۱}{۴}$ است. برای راحتی می‌توان تمام احتمالات متغیر تصادفی X را با فرمولی نمایش داد. چنین فرمولی تابعی از مقادیر عددی X خواهد بود و آن را معمولاً با $f(x)$ یا $P(X=x)$ نشان می‌دهند.

x	۰	۱	۲
$f(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

مثال ۳-۱: در پرتاب دو تاس، اگر متغیر تصادفی X مجموع وجه‌های ظاهر شدهٔ دو تاس باشد، پس از نوشتن فضای نمونهٔ تابع چگالی یا $f(x)$ را بنویسید.
با توجه به تعریف فضای نمونه

$$S=\{(۱,۱),(۱,۲),(۱,۳),(۱,۴),(۱,۵),(۱,۶),\ldots,(۶,۱),(۶,۲),(۶,۳),(۶,۴),(۶,۵),(۶,۶)\}$$

مقادیری که X اختیار می‌کنند برابرند با:

$X: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$

بنابراین

x	۲	۳	٤	٥	٦	۷	۸	۹	۱۰	۱۱	۱۲
$f(x)$	$\frac{۱}{۳۶}$	$\frac{۲}{۳۶}$	$\frac{۳}{۳۶}$	$\frac{۴}{۳۶}$	$\frac{۵}{۳۶}$	$\frac{۶}{۳۶}$	$\frac{۵}{۳۶}$	$\frac{۴}{۳۶}$	$\frac{۳}{۳۶}$	$\frac{۲}{۳۶}$	$\frac{۱}{۳۶}$

در ادبیات و منابع آماری $f(x)$ یا $P(X=x)$ را تابع چگالی احتمال می‌نامند که در ۲ شرط مهم زیر صدق می‌کند:

$$۱- f(x) \geq ۰ \quad \forall \quad x \in A_x$$
$$۲- \sum_x f(x) = ۱$$

مثال ۳-۲: در پرتاب سه سکه اگر متغیر تصادفی X تعداد خط‌ها باشد، پس از نوشتن A_X، تابع چگالی آن را به دست آورید.

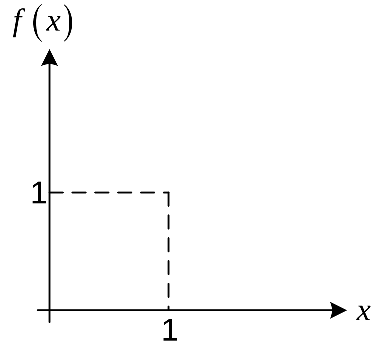
$$S = \{(HHH),(HHT),(HTH),(THH),(HTT),(THT),(TTH),(TTT)\}$$
$$A_X = \{۰,۱,۲,۳\}$$

x	۰	۱	۲	۳
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

۳-۳ توزیع احتمالات پیوسته

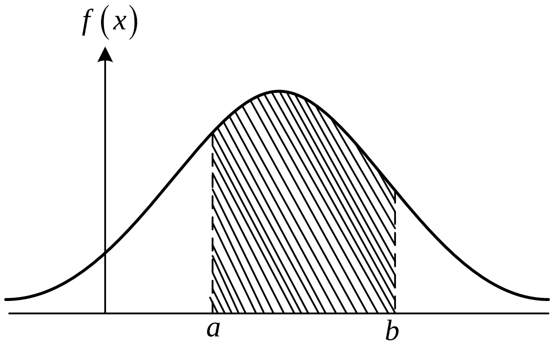
از آنجا که متغیر تصادفی می‌تواند مقادیری را که اختیار می‌کند پیوسته باشد، متغیر تصادفی پیوسته روی فضای نمونه تعریف می‌شود. قابل ذکر است که توزیع احتمالات یک متغیر تصادفی پیوسته را نمی‌توان همانند متغیر تصادفی گسسته به صورت جدول نمایش داد و ضرورت دارد آن را به صورت فرمول ارائه داد. اگر تابع چگالی احتمال متغیر تصادفی تعریف شده روی فضای نمونه

S باشد، آن را تابع چگالی[۱] می‌نامند. تابع چگالی طوری طرح‌ریزی می‌شود که سطح محدود بین منحنی و محور xها برابر یک شود. نمونه‌هایی از تابع چگالی در شکل ۳–۱ نشان داده شده است.



شکل ۳–۱

تعریف تابع $f(x)$، تابع چگالی احتمالات برای متغیر تصادفی پیوستهٔ X نامیده می‌شود اگر مجموع سطوح زیر منحنی آن‌که به وسیلهٔ محور xها احاطه شده، برابر یک است و اگر سطح زیر منحنی بین هر دو طول $x=a$ و $x=b$ احتمالی را بدهد که X بین a و b قرار گیرد؛ یعنی $P(a<X<b)=S$ که در شکل زیر نشان داده شده است:



شکل ۳–۲

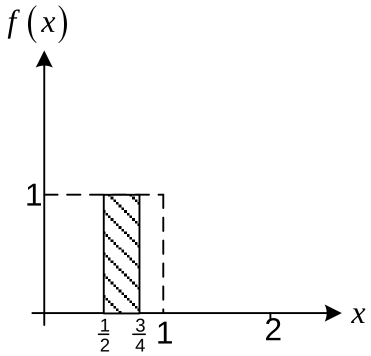
۱. Density function

مثال ۳-۳: فرض کنید متغیر تصادفی X دارای تابع چگالی زیر باشد:

$$f(x)=۱ \quad ۰<x<۱$$

منحنی آن را رسم کنید و مقدار $P\left(\frac{۱}{۲}<X<\frac{۳}{۴}\right)$ را به دست آورید.

مقدار احتمال $P\left(\frac{۱}{۲}<X<\frac{۳}{۴}\right)$ یعنی مساحت بین $\frac{۱}{۲}$ و $\frac{۳}{۴}$ که به صورت زیر محاسبهٔ می‌شود:



مساحت شکل هاشورخورده برابر است با:

$$S=\text{طول}\times\text{عرض}=P\left(\frac{۱}{۲}<X<\frac{۳}{۴}\right)=\left(\frac{۳}{۴}-\frac{۱}{۲}\right)\times۱=\frac{۱}{۴}$$

۳-٤ تابع توزیع

فرض کنید X متغیر تصادفی و دارای تابع چگالی یا تابع چگالی احتمال $f(x)$ باشد. برای یک عدد حقیقی x، احتمالی رخداد پیشامد $(X\leq x)$ را با $F(x)$ نشان می‌دهند و آن را تابع توزیع گویند؛ یعنی

$$F(x)=P(X\leq x)$$

می‌توان گفت $F(x)$ نقش تابع تجمعی در آمار توصیفی را دارد.

مثال ۳-٤: فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد.

x	۰	۱	۲
$f(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

تابع توزیع X یا $F(x)$ را به دست آورید.

با توجه به تعریف $F(x)$، برای $x=۰$،

$$F(x)=P(X\le x)=F(۰)=P(X\le ۰)=P(X=۰)=\frac{۱}{۴}$$

برای $x=۱$،

$$F(x)=P(X\le x)=F(۱)=P(X\le ۱)=P(X=۰\quad X=۱)$$
$$=P(X=۰)+P(X=۱)=\frac{۱}{۴}+\frac{۲}{۴}=\frac{۳}{۴}$$

برای $x=۲$،

$$F(x)=p(X\le x)=F(۲)=p(X\le ۲)$$
$$=p(X=۰)+p(X=۱)+p(X=۲)=\frac{۱}{۴}+\frac{۲}{۴}+\frac{۱}{۴}=۱$$

می‌توان $F(x)$ را به صورت زیر با چند ضابطه نشان داد:

$$F(x)=\begin{cases} ۰ & x<۰ \\ \frac{۱}{۴} & ۰\le x<۱ \\ \frac{۳}{۴} & ۱\le x<۲ \\ ۱ & x\ge ۲ \end{cases}$$

مثال ۳–۵: فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد:

x	-۱	۰	۱	۲
$f(x)$	۰/۲	۰/۳	k	۰/۲

الف- مقدار k را محاسبهٔ کنید.

ب- تابع توزیع را به دست آورید.

حل: چون $\sum_x f(x)=۱$ است. بنابراین

$$۰/۲+۰/۳+k+۰/۲=۱ \quad , \quad k=۰/۳$$

x	-۱	۰	۱	۲
$f(x)$	۰/۲	۰/۳	۰/۳	۰/۲

$$F(x)=\begin{cases} ۰ & x<-۱ \\ ۰/۲ & -۱\leq x<۰ \\ ۰/۵ & ۰\leq x<۱ \\ ۰/۸ & ۱\leq x<۲ \\ ۱ & x\geq ۲ \end{cases}$$

با توجه به تعریف متغیر تصادفی و ارائه قالب احتمال برای آن در ادامه چند توزیع خاص آماری از نوع گسسته و پیوسته آورده می‌شود.

۳-٤-۱ توزیع یکنواخت گسسته

متغیر تصادفی X دارای تابع چگالی احتمال گسسته با پارامتر k است، اگر تابع چگالی آن به صورت زیر باشد:

$$P(X=x)=f(x)=\frac{۱}{k} \quad , \quad x=۱,۲,۳,\ldots,k$$

البته می‌توان با توجه به شانس انتخاب $f(x)$ را به صورت زیر نوشت:

x	۱	۲	۳	...	k
$f(x)$	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{k}$	...	$\frac{1}{k}$

واضح است که:

$$f(۱)+f(۲)+f(۳)+\ldots+f(k)=\sum_{x=۱}^{k} f(x)=۱$$

مثال ۳–٦: برای $k=۶$، تابع چگالی گسسته را بنویسید.

$$P(X=x)=f(x)=\frac{۱}{۶} \quad , \quad x=۱,۲,\ldots,۶$$

در حقیقت این قالب برای پرتاب یک تاس است، اگر متغیر تصادفی X وجه ظاهر شده برای تاس باشد.

۳–٤–۲ توزیع برنولی

برای تعریف تابع چگالی احتمال برنولی، نیاز به تعریف آزمایش برنولی است. آزمایش برنولی یک تجربه است که نتیجهٔ آن می‌تواند مثبت یا منفی، صفر یا یک باشد. اگر شانس پیروزی و شکست به ترتیب p و q و $q=۱-p$ باشد، تابع چگالی احتمال برنولی به صورت زیر خواهد بود:

$$P(X=x)=f(x)=p^x(۱-p)^{۱-x} \quad , \quad x=۰,۱ \quad , \quad ۰<p<۱$$

یا به عبارتی، $f(۰)=۱-p$ و $f(۱)=p$ بنابراین می‌توان $f(x)$ را به صورت زیر نشان داد:

x	۰	۱
$f(x)$	$1-p$	p

مثال ۳–٦: اگر شانس بارانی بودن یک روز بخصوص ۰/۸ باشد و متغیر تصادفی X نشان‌دهندهٔ وضعیت بارندگی در آن روز بخصوص باشد، تابع چگالی احتمال X را مشخص کنید و احتمال اینکه وضعیت بارانی نباشد را محاسبه کنید.

حل:

وضعیت بارانی: $X=1$

وضعیت غیربارانی: $X=0$

$$P(X=x)=0/8 \quad , \quad P(X=0)=0/2$$

x	۰	۱
$f(x)$	۰/۲	۰/۸

۳-٤-۳ توزیع دوجمله‌ای

متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای n و P است، اگر تابع چگالی آن به صورت زیر باشد:

$$P(X=x)=f(x)=\binom{n}{x}p^x(1-p)^{n-x} \quad , \quad x=0,1,2,\dots,n \quad , \quad 0<p<1$$

در واقع تابع چگالی دوجمله‌ای، قالب احتمال برای متغیر تصادفی X است که متغیر تصادفی X داشتن x پیروزی برای n تکرار آزمایش برنولی است.

مثال ۳-۷: شخصی در یک آزمون چهار جوابی که شامل ۷۰ سؤال است شرکت می‌کند، مطلوبست:

الف- احتمال اینکه این شخص به هیچ سؤالی پاسخ صحیح ندهد چقدر است؟

ب- احتمال اینکه این شخص به همه سؤالات پاسخ صحیح دهد چقدر است؟

ج- احتمال اینکه به ۵۰ سؤال پاسخ صحیح دهد چقدر است؟

حل:

$$n=70 \quad , \quad p=\frac{1}{4} \quad , \quad f(x)=\binom{n}{x}p^x(1-p)^{n-x}$$

$$f(x)=\binom{70}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{70-x} \quad , \quad x=0,1,2,\dots,70$$

$$f(0)=p(X=0)=\binom{70}{0}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{70-0}=\left(\frac{3}{4}\right)^{70}$$

$$f(70)=p(X=70)=\binom{70}{70}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{70-0}=\left(\frac{3}{4}\right)^{70}=\left(\frac{1}{4}\right)^{70}$$

$$f(50)=p(X=50)=\binom{70}{50}\left(\frac{1}{4}\right)^{50}\left(\frac{3}{4}\right)^{20}$$

۳-٤-٤ توزیع پواسن

متغیر تصادفی X دارای توزیع پواسن با پارامتر λ است، اگر چگالی احتمال آن به صورت زیر باشد:

$$p(X=x)=f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\quad,\quad x=0,1,2,\ldots\quad,\quad \lambda>0$$

مثال ۳–۸: تجربه نشان داده است که در هر ساعت ۱۰ نفر وارد کتابخانه می‌شود. مطلوبست

- احتمال اینکه در یک ساعت بخصوص شخصی وارد کتابخانه نشود چقدر است؟
- احتمال اینکه در یک ساعت کم‌تر از ۸ نفر وارد کتابخانه شود چقدر است؟
- احتمال اینکه در یک ساعت بیش از ۵ نفر وارد کتابخانه شود چقدر است؟

حل:

$$\lambda=10\quad,\quad p(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}=f(x)\quad,\quad x=0,1,2,\ldots$$

$$f(x)=\frac{e^{-10}(10)^{x}}{x!}$$

$$f(0)=\frac{e^{-10}(10)^{0}}{0!}=e^{-10}$$

$$p(X<8)=f(0)+f(1)+f(2)+\ldots+f(7)=\frac{e^{-10}(10)^{0}}{0!}+\frac{e^{-10}(10)}{1}+\ldots+\frac{e^{-10}(10)^{7}}{7!}$$

$$p\left(X>۵\right)=۱-p\left(X\leq ۵\right)=۱-f\left(۰\right)-f\left(۱\right)-f\left(۲\right)-f\left(۳\right)-f\left(۴\right)-f\left(۵\right)$$

مثال ۳-۹: در کتابخانه‌ای بزرگ در هر دقیقه ۵ کتاب به اعضای امانت داده می‌شود مطلوبست:

الف- احتمال اینکه در یک دقیقه ۳ کتاب امانت داده شود چقدر است؟

ب- احتمال اینکه در ۱۰ دقیقه فقط ۲۷ کتاب امانت داده شود چقدر است؟

ج- احتمال اینکه در یک ساعت کم‌تر از ۳۰ کتاب امانت داده شود چقدر است؟

حل:

$$e=۵\quad,\quad f\left(x\right)=\frac{e^{-\lambda}\lambda^{x}}{x!}+\frac{e^{-۵}\left(۵\right)^{x}}{x!}\quad,\quad x=۰,۱,۲,\ldots$$

$$f\left(۳\right)=\frac{e^{-۵}\left(۵\right)^{۳}}{۳!}=\frac{۱۲۵}{۶}e^{-۵}$$

ب- با توجه به اینکه ضریب یا شدت امانت در هر دقیقه ۳ است، نیاز است ضریب آن را برای ده دقیقه به‌صورت زیر به‌دست می‌آوریم:

۱ ۵

۱۰ $\lambda=۵۰$

$$f\left(۲۷\right)=\frac{e^{-۵۰}\left(۵۰\right)^{۲۷}}{۲۷!}$$

ج-

۱ ۵

$q_{۰}$ $\lambda=۳۰۰$

$$P\left(X<۳۰\right)=\sum_{x=۰}^{۲۹}\frac{e^{-\lambda}\lambda^{x}}{x!}=\sum_{x=۰}^{۲۹}\frac{e^{-۳۰۰}\left(۳۰۰\right)^{x}}{x!}$$

۳-۵ توزیع یکنواخت پیوسته

توزیع یکنواخت در متغیر پیوسته ساده‌ترین توزیع و تابع چگالی آن در فاصلهٔ معین (a, b) به صورت زیر است؛

$$f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

توزیع یکنواخت در آمار ریاضی نقش مهمی دارد. این توزیع در مواردی نظیر مطالعه گرد کردن اعداد وقتی که اندازه‌گیری تا دقت معینی انجام می‌شود و همچنین طول زمان انتظار برای مسافران اتوبوس یا مترو و غیره مورد استفاده قرار می‌گیرد.

مثال ۳-۱۰: درب ورودی کتابخانه طوری تنظیم شده که در هر ۱۵ ثانیه یکبار باز می‌شود؛ احتمال اینکه یک نفر کم‌تر از ۱۰ ثانیه پشت درب بماند چقدر است؟

$$f(x) = \frac{1}{15} \qquad 0 < x < 15$$

$$P(X \le 10) = \int_0^{10} \frac{1}{15}\, dx = \frac{x}{15}\Big|_0^{10} = \frac{10}{15} = \frac{2}{3}$$

۳-٦ توزیع نرمال

یکی از توزیع‌های مهم توزیع احتمال پیوسته، توزیع نرمال است که در بسیاری از فنون استنباط آماری بر این پیش‌فرض متکی است که توزیع فراوانی یک متغیر در جامعهٔ آماری به صورت مناسبی به وسیله توزیع نرمال با یک میانگین و انحراف معیار بخصوص تشریح می‌شود.

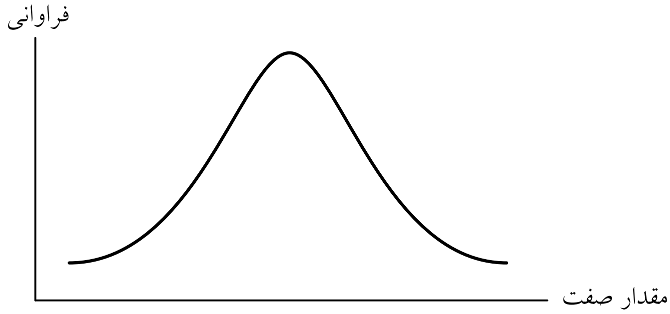
تعریف: متغیرتصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 است، اگر تابع چگالی آن به صورت زیر باشد:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\ ,\ -\infty<x<\infty\ ,\ -\infty<\mu<\infty\ ,\ \sigma>0$$

که در آن $e\approx$ ۲/۷۱۸۲ و $\pi\approx$ ۳/۱۴۱۵ و نمودار $f(x)$ را منحنی نرمال گویند. توزیع نرمال غالباً به نام گوسین به افتخار گوس (۱۸۵۵– ۱۷۷۷) نامیده می‌شود که او معادله آن را از روی مطالعهٔ خطای حاصل از اندازه‌گیری مکرر کیفیت یک کمّیت به دست آورد.

در آمار متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ را با نماد $X \sim N(\mu,\sigma^۲)$ نشان می‌دهند و می‌خوانند: X دارای توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ است. جوامع یا توزیع‌های زیادی است که طبیعت یا توزیع آن‌ها نرمال است؛ از جمله توزیع قد، توزیع وزن، نمرهٔ ضریب هوشی، نمرهٔ دانشجویان، خطا، توزیع درآمد و غیره.

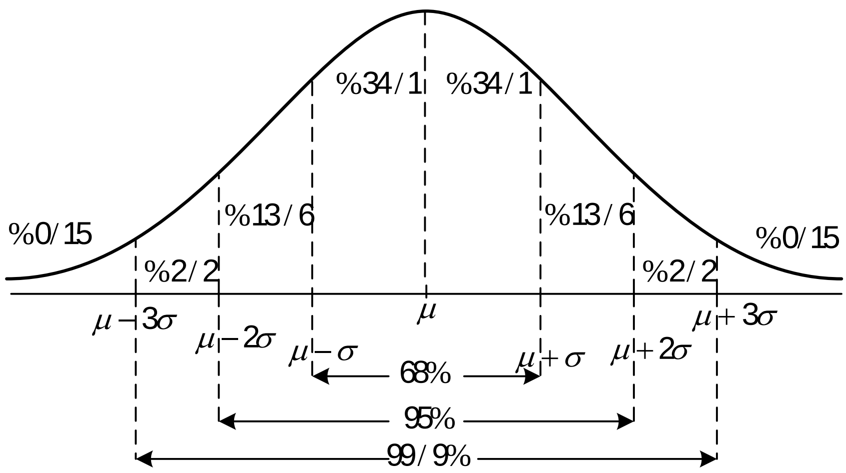
به طور معمول، متغیرها و صفات در طبیعت از توزیع نرمال تبعیت می‌کنند؛ این گفته یعنی آنکه اگر نمودار توزیع فراوانی یک صفت را رسم، روی محور عرضی فراوانی آن صفت و روی محور طولی مقادیر مختلف آن صفت را درج کنیم، شکلی مشابه شکل زیر ایجاد خواهد شد:



نمودار ۳-۳

به عبارت دیگر در چنین وضعیتی ما یک منحنی زنگوله‌ای شکل متقارن خواهیم داشت.

در چنین وضعیتی احتمال آنکه داده‌ها در دو فاصلهٔ مشخص از مقادیر صفت قرار داشته باشند، از روی منحنی زیر قابل محاسبهٔ است:



به عبارت دیگر ٦٨٪ داده‌ها در فاصلهٔ میانگین و یک انحراف معیار قرار دارند.

$(\mu-\sigma, \mu+\sigma)$

۹۵٪ داده‌ها در فاصلهٔ یک میانگین و دو انحراف معیار قرار دارند.

$(\mu-2\sigma, \mu+2\sigma)$

۹۹/۹٪ داده‌ها در فاصلهٔ یک میانگین و سه انحراف معیار قرار دارند.

$(\mu-3\sigma, \mu+3\sigma)$

نکتهٔ مهم در مورد این نمودار، آن است که حول میانگین متقارن است.

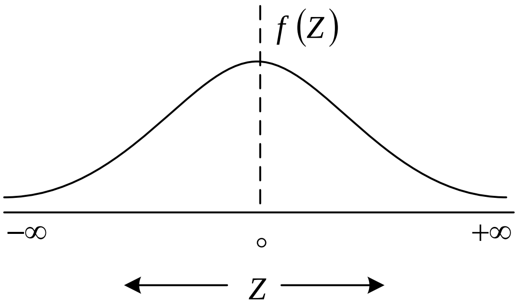
توزیع نرمال استاندارد

هرگاه در شکل توزیع نرمال تغییر متغیر $Z=\dfrac{X-\mu}{\sigma}$ را لحاظ کنیم، تابع چگالی نرمال استاندارد به صورت زیر است:

$$f(z)=\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}z^{2}} \qquad , \qquad -\infty < Z < \infty$$

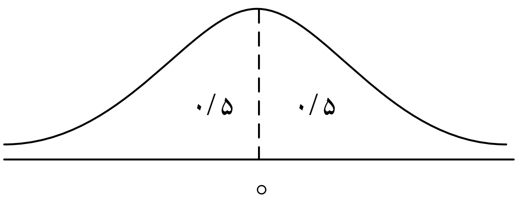
تابع چگالی فوق را تابع چگالی نرمال استاندارد گویند. بنابراین $Z \sim N(0,1)$ است و ویژگی‌های یک متغیر استاندارد عبارتند از:

- میانگین و واریانس آن به ترتیب صفر و یک است.
- منحنی $f(z)$ به صورت زیر است:



نمودار ۳-۴

- مساحت زیر منحنی برابر با یک است.
- منحنی نسبت به محور $Z=0$ یا محور yها متقارن است؛ یعنی:

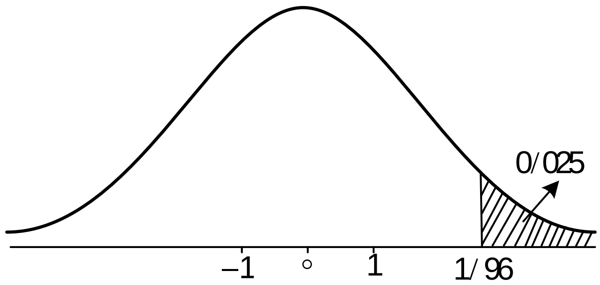


نمودار ۳-۵

بنابراین $P(Z>0)=P(Z<0)=۰/۵$

برای محاسبهٔ احتمال به ازای مقادیر مختلف $Z_۰$ جدولی تحت عنوان جدول نرمال استاندارد تنظیم شده است که به راحتی می‌توان با استفاده از آن جدول سطح زیر منحنی نرمال استاندارد را به دست آورد (جدول۱).

برای مثال، اگر منحنی زیر را در نظر بگیریم:



مساحت بیشتر از ۱/۹۶ یا $P\left(Z > 1/96\right)$ برابر است با ۰/۰۲۵ و مساحت کم‌تر از ۱/۹۶ یا $P\left(Z < 1/96\right)$ برابر با ۰/۹۷۵ است؛ چون ۱=۰/۰۲۵+۰/۹۷۵ است.

اکنون فرض کنید $X \sim N\left(\mu, \sigma^۲\right)$ باشد، برای محاسبهٔ احتمالات ابتدا باید آن را استاندارد و به نرمال $Z \sim N\left(۰,۱\right)$ تبدیل کرد. برای این کار کافی است همان طور که قبل‌تر نیز مطرح شد، به صورت زیر عمل کنیم:

برای استاندارد کردن X کافی است از آن متغیر میانگین را کم و بر انحراف معیار تقسیم کرد. بنابراین

$$Z = \frac{X - \mu}{\sigma} \sim N\left(۰,۱\right)$$

اکنون می‌توانیم احتمالات زیر را با توجه به تبدیلات از جدول نرمال استاندارد محاسبهٔ کنیم:

$$P\left(X < b\right) = P\left(\frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = P\left(Z < \frac{b - \mu}{\sigma}\right)$$

$$P(X>a)=P\left(\frac{X-\mu}{\sigma}>\frac{a-\mu}{\sigma}\right)=P\left(Z>\frac{a-\mu}{\sigma}\right)$$

۳-۷ نحوهٔ استفاده از جدول Z

برای محاسبهٔ احتمالات در توزیع نرمال استاندارد از جدولی با عنوان "جدول Z" استفاده می‌شود. بر روی ستون عمودی سمت چپ این جدول اعداد ۰/۰ تا ۳/۰ و بر روی ردیف افقی بالای این ستون، مقادیر ۰/۰ تا ۰/۰۹ نشان داده شده است.

مثال ۳-۱۱: فرض کنید که $Z\sim N(۰,۱)$، با توجه به جدول، احتمالات زیر محاسبه شده است.

$$P(Z\le ۲/۱۷)=۰/۹۸۵۰$$
$$P(Z\le -۳/۱۲)=۰/۰۰۰۹$$
$$P(Z>۱/۱۷)=۱-P(Z\le ۱/۱۷)=۱-۰/۸۷۹۰=۰/۱۲۱۰$$
$$P(-۱/۰۱<Z<۲/۰۱)=P(Z<۲/۰۱)-P(Z<-۱/۰۱)=۰/۹۷۷۸-۰/۱۵۶۲$$

مثال ۳-۱۲: اگرخطای دستگاه برش کتاب‌ها از توزیع نرمال استاندارد پیروی کند مطلوبست:

الف- احتمال اینکه خطای کم‌تر از ۱/۹۶ میلی‌متر رخ دهد.

ب- احتمال اینکه خطای بین صفر و ۱/۹۶- میلی‌متر رخ دهد.

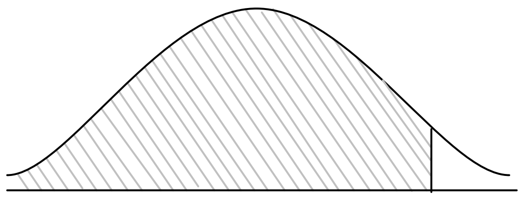
ج- احتمال اینکه خطای بین ۱ و ۱- میلی‌متر رخ دهد.

چون توزیع خطا نرمال استاندارد است، پس:

$$f(Z)=\frac{۱}{\sqrt{۲\pi}}\ e^{-\frac{۱}{۲}Z^{۲}} \qquad , \qquad -\infty<Z<\infty$$

الف) $P(Z<۱/۹۶)$

این احتمال با مساحت هاشورخورده زیر هم‌ارز است.



$$P(Z < ۱/۹۶) = ۰/۹۷۵$$

بنابراین

ب) $P(-۱/۹۶ < Z < ۰) = ۰/۵ - ۰/۲۵ = ۰/۴۷۵$

با توجه به جدول

ج) $P(-۱ < Z < ۱) = ۰/۶۸۲۶$

مثال ۳-۱۳: تجربه نشان داده است که نمرات دانشجویان دارای توزیع نرمال با میانگین ۱۲ و واریانس ٤ است. احتمال اینکه نمرهٔ یک دانشجو که به تصادف انتخاب می‌شود کم‌تر از ۱۵/۹۲ باشد چقدر است؟

- اگر متغیرتصادفی X نمرهٔ دانشجو باشد، با توجه به توزیع آن:

$$X \sim N(12, 4)$$

$$P(X < ۱۵/۹۲) = P\left(\frac{X - ۱۲}{۲} < \frac{۱۵/۹۲ - ۱۲}{۲}\right)$$

$$= P\left(Z < \frac{۳/۹۲}{۲}\right) = P(Z < ۱/۹۶) = ۰/۹۷۵$$

تبصره: اگر X_1 , X_2 , ... , X_n یک نمونه از توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ باشد، میانگین نمونه یا $\overline{X}$ نیز دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^۲}{n}$ است.

یعنی:

$$X \sim N(\mu , \sigma^۲) \qquad , \qquad \overline{X} \sim N\left(\mu , \frac{\sigma^۲}{n}\right)$$

اگر میانگین نمونه استاندارد شود خواهیم داشت:

$$Z_{\overline{X}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

مثال ۳-۱۴: در چاپخانه، اگر طول کتاب‌های چاپ شده دارای توزیع نرمال با میانگین ۲۳ و واریانس ۰/۰۱ باشد، احتمال اینکه متوسط طول ۲۵ کتاب بیشتر از ۲۳/۰۳۹۲ باشد، چقدراست؟

$$X \sim N(۲۳ , ۰/۰۱) \quad , \quad \overline{X} \sim N\left(۲۳ , \frac{۰/۰۱}{۲۵}\right)$$

$$P(\overline{X} > ۲۳/۰۳۹۲) = P\left(\frac{\overline{X} - ۲۳}{\frac{\sigma}{\sqrt{n}}} > \frac{۲۳/۰۳۹۲ - ۲۳}{۰/۱ / \sqrt{۲۵}}\right)$$

$$P\left(Z > \frac{۵ \times (۰/۰۳۹۲)}{۰/۱}\right) = P\left(Z > \frac{۳/۹۲}{۲}\right) = P\left(Z > \frac{۳/۹۲}{۲}\right) = P(Z > ۱/۹۶) = ۰/۰۲۵$$

۳-۸ توزیع میانگین نمونه‌ها و برآورد نقطه‌ای میانگین جامعه

اگر جامعه‌ای متشکل از N نفر باشد و تعدادی نمونهٔ n تایی به صورت تصادفی از جامعه بگیریم و بخواهیم مثلاً میانگین یا واریانس نمونه را به دست آوریم، به طور حتم مقادیر مختلفی به دست می‌آید.

به عنوان مثال، کلاسی را در نظر بگیرید که از ۲۰ نفر تشکیل شده است، این ۲۰ نفر نمرات مختلفی در درس آمار کسب کرده‌اند. فرض کنید می‌خواهیم نمونه‌ای سه تایی از این کلاس بگیریم و میانگین نمرات این سه نفر را محاسبه

کنیم. در اینجا اگر نمونه‌ای که می‌گیریم افرادی باشند که نمرات آن‌ها مثلاً ۲۰، ۱۹ و ۱۸ باشند، میانگین عبارت خواهد بود از:

$$\bar{X} = \frac{۱۸+۱۹+۲۰}{۳} = ۱۹$$

و اگر افرادی که انتخاب می‌کنیم نمرات ۹، ۱۲ و ۱۰ را کسب کرده باشند، میانگین عبارت خواهد بود از:

$$\bar{X} = \frac{۱۰+۱۲+۹}{۳} = ۱۰/۳$$

در نمونه‌گیری، همواره مهم‌ترین اصل این است که نمونه باید معرّف جامعه باشد؛ یعنی مثلاً در مثال فوق بتوانیم این گونه تحلیل کرد که عدد حاصل از نمونه‌گیری برای تمام جامعه مناسب است و یا به عبارتی کل کلاس دارای میانگین نمره به دست آمده هستند.

در مثال فوق می‌توان گفت که کل کلاس به طور میانگین دارای نمره ۱۰/۳ و یا اینکه کل کلاس دارای نمره ۱۹ بوده است؟

نکته‌ای که اینجا پیش می‌آید این است که ما چگونه باید فرضاً میانگین را در یک جامعه که به تمام اعضای آن دسترسی نداریم محاسبهٔ کنیم که مشکلی همانند مشکلی که در مثال فوق پیش آمده بود، پیش نیاید؟ برای حل این مشکل ابتدا دو قضیهٔ زیر را مطرح می‌کنیم.

قضیهٔ اول: اگر از جامعه‌ای که توزیع آن نرمال است، نمونهٔ n تایی استخراج شود، میانگین نمونهٔ $\bar{X}$ نیز دارای توزیع نرمال است.

مثال ۳-۱۵: جامعه‌ای سه عضو دارد: ۲۰، ۲۲و ۲٤ از این جامعه نمونه‌های دوتایی $n=۲$ با جایگذاری استخراج می‌کنیم.

میانگین کل در اینجا عبارت است از:

$$\mu=\frac{\sum X_i}{n}=\frac{۲۰+۲۲+۲۴}{۳}=۲۲$$

پس واریانس جامعه عبارت است از:

$$\sigma^۲=\frac{\sum\left(X_i-\mu\right)^۲}{N}=\frac{\left(۲۰-۲۲\right)^۲+\left(۲۲-۲۲\right)^۲+\left(۲۴-۲۲\right)^۲}{۳}=\frac{۴+۰+۴}{۳}=\frac{۸}{۳}$$

اکنون در این مثال، تمامی نمونه‌های دوتایی ممکن را در نظر می‌گیریم که میانگین آن‌ها در جدول آمده است:

میانگین نمونه‌های دوتایی $\bar{X}$	نمونهٔ دوتایی
۲۰	۲۰,۲۰
۲۱	۲۰,۲۲
۲۲	۲۰,۲٤
۲۲	۲۲,۲۲
۲۳	۲۲,۲٤
۲۱	۲۲,۲۰
۲٤	۲٤,۲٤
۲۲	۲٤,۲۰
۲۳	۲٤,۲۲

اگر بین میانگین‌های دوتایی به دست آمده، دوباره میانگین بگیریم، خواهیم داشت:

$$E\left(\bar{X}\right)=\frac{۲۰+۲۱+۲۲+۲۲+۲۳+۲۴+۲۲+۲۳+۲۱}{۹}=۲۲$$

توجه داریم که مقدار به دست آمده برای $E\left(\bar{X}\right)$ با مقداری که برای μ در ابتدای حل به دست آوردیم، یکسان بود.

قضیهٔ دوم: هرگاه از یک جامعه با تعداد N و میانگین μ، تمام نمونه‌های nتایی استخراج شود و میانگین هر کدام از این نمونه‌ها $\bar{X}_i$ باشد، در این صورت میانگین میانگین‌ها را با $E\left(\bar{X}\right)$ نشان می‌دهیم و تقریباً با μ که میانگین جامعه بود، برابر است؛ یعنی:

$$\mu = E\left(\bar{X}\right)$$

۳-۹ توزیع استودنت (t)

در بخش قبل ملاحظه شد که اگر $X_۱ , X_۲ , \ldots , X_n$ یک نمونهٔ تصادفی از جامعهٔ نرمال با میانگین μ و واریانس $\sigma^۲$ باشد، متغیر تصادفی $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد با میانگین صفر و واریانس ۱ است، مشروط بر این که واریانس $\sigma^۲$معلوم باشد. در حالتی که واریانس جامعه $\left(\sigma^۲\right)$ مجهول باشد، از براورد نقطه‌ای آن یعنی $S^۲ = \frac{۱}{n-۱}\sum\left(x_i-\bar{x}\right)^۲$ استفاده می‌شود. در این حالت متغیرتصادفی $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$دارای توزیع استودنت با $df = n-۱$ [۱] درجهٔ آزادی است و آن را با T نشان می‌دهند.

اگر $f(t)$ تابع چگالی متغیر تصادفی T باشد، $f(t)$ همانند توزیع نرمال استاندارد دارای ویژگی‌های زیراست:

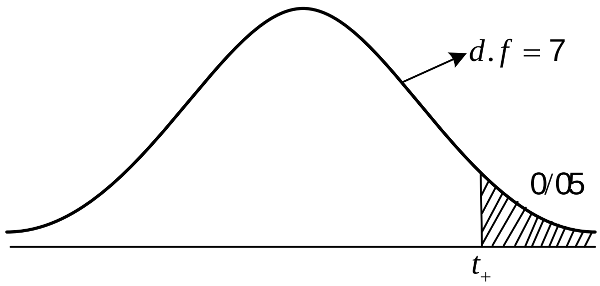
- $\int_{-\infty}^{\infty} f(t)dt = ۱$
- برای $n-۱>۱$ دارای میانگین صفر و واریانس $\frac{n-۱}{n-۳}$
- نسبت به محور yها دارای تقارن است.

۱. df: Degrees of Freedom

- اگر $n-۱$ از حد تصور بزرگ‌تر باشد، توزیع استودنت بر توزیع نرمال استاندارد منطبق می‌شود. (معمولاً بزرگ‌تر از ۳۰)
- برای مقادیر مختلف $n-۱$ می‌توان برای $T=t$ داده شده مساحت زیر منحنی را از جدول توزیع استودنت به دست آورد.

نحوهٔ محاسبهٔ عدد جدول از استودنت برای درجهٔ آزادی $d.f=7$ و $\alpha=0/05$ شرح داده می‌شود.

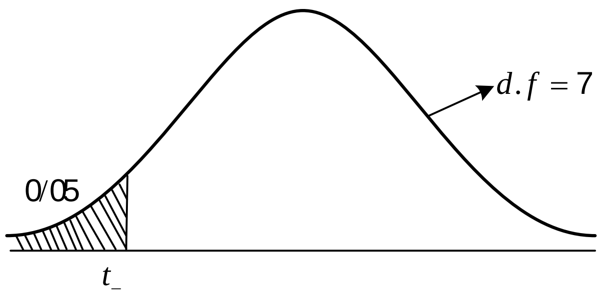
با توجه به منحنی توزیع استودنت می‌توان مساحت هاشورخورده و عدد جدول را به صورت زیر نمایش داد:



که در آن t عدد جدول است. به دلیل همسو بودن شکل با جدول، کافی است. درجه آزادی یا $d.f=r=7$ از ستون اول پیدا کرد و تقاطع آن با ستون $\alpha=0/05$ بنابراین:

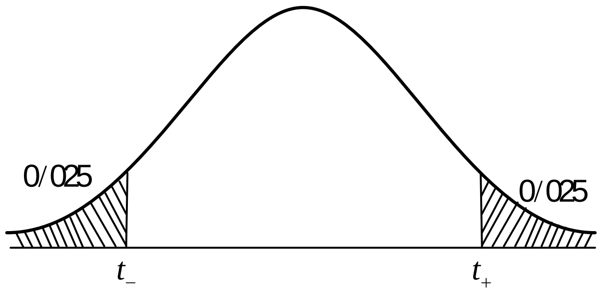
$t=1/895$

حال فرض کنید با توجه به منحنی زیر هدف، پیدا کردن t_- است.

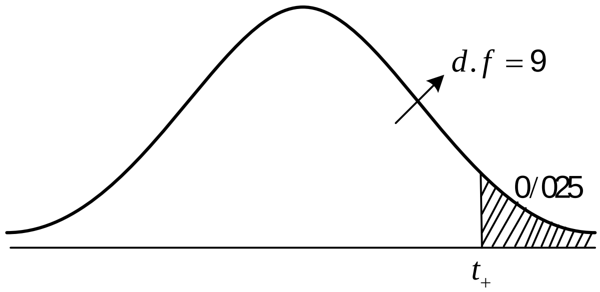


با توجه به درجهٔ آزادی $r = d.f = 7$ و متقارن بودن منحنی نسبت به محور yها می‌توان نوشت: $t_{-} = -1/895$

مثال ۳–۱٦: اگر درجهٔ آزادی $d.f = n-1 = r = 9$ باشد، با توجه به منحنی زیر اعداد t_{+}، t_{-} را پیدا کنید.

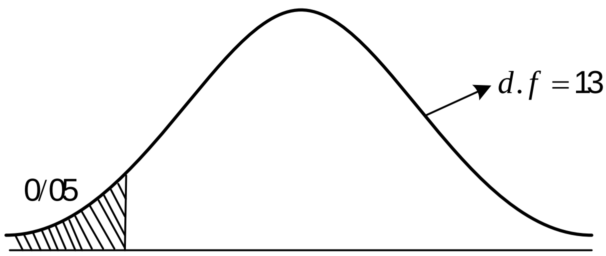


ابتدا می‌توان t_{+} را با توجه به منحنی زیر به دست آورد.

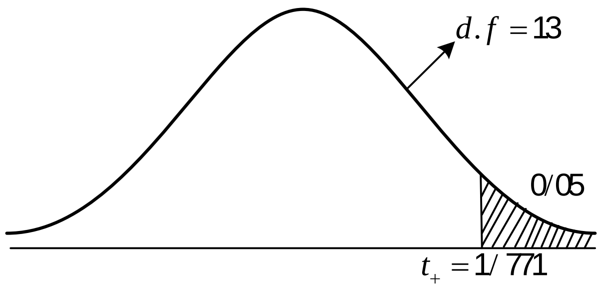


$t_{+} = 2/262$ با توجه به تقارن $t_{-} = -2/262$

مثال ۳–۱۷: با توجه به منحنی زیر و درجهٔ آزادی ۱۳، مقدار t_{-} را به دست آورید.



چون کتب ارائهٔ اعداد جدول برای سمت راست یا اعداد مثبت است، لذا با توجه به تقارن می‌توان منحنی زیر را در نظر گرفت:



با توجه به تقارن $t_- = -1/771$ خواهد بود.

۳–۱۰ امید ریاضی

مفهوم امید ریاضی، در اصل در ارتباط با بازی‌های شانسی به وجود آمده است و در ساده‌ترین صورتش، حاصل‌ضرب مبلغی است که بازیکن احتمال دارد که برنده شود. در واقع، امید ریاضی یک میانگین و یا همان مقدار مورد انتظار یک متغیر تصادفی است.

برای درک مفهوم امید ریاضی مثال زیر ارائه می‌شود:

مثال ۳–۱۸: فرض کنید در یک بازی شرطی شرکت کرده‌ایم، در این بازی تاس سالمی پرتاب می‌شود؛ اگر اعداد ۱ یا ۲ رو شوند، مبلغ ۵ تومان می‌بریم، اگر

اعداد ۳ و ٤ رو شوند، مبلغ ۷ تومان می‌بریم و اگر اعداد ۵ و ٦ ظاهر شد، مبلغ ۱۵ تومان می‌بریم، بنابراین انتظار داریم مبلغ:

$$\left(\frac{۱}{۶}+\frac{۱}{۶}\right)\times ۵+\left(\frac{۱}{۶}+\frac{۱}{۶}\right)\times ۷+\left(\frac{۱}{۶}+\frac{۱}{۶}\right)\times ۱۵=\frac{۵۴}{۶}=۹$$

تومان به طور متوسط برنده شویم.

در فصول گذشته میانگین حسابی داده‌ها $\left(\overline{X}\right)$ معرفی و روشن شد که اگر مقادیر $x_1, x_2, \ldots, x_n$ با فراوانی‌های متناظر $f_1, f_2, \ldots, f_n$ در دست باشد، میانگین آن‌ها از این فرمول محاسبهٔ می‌شود:

$$\overline{x}=\frac{۱}{n}\sum f_i x_i$$

$\frac{f_i}{n}$ فراوانی نسبی داده‌هاست. امید ریاضی یک متغیر تصادفی گسسته X که آن را با $E(X)$ نشان می‌دهند و در واقع همان میانگین توزیع احتمال X است. (μ)، برای تعیین مرکز یک توزیع احتمالی به کار گرفته می‌شود:

$$E(X)=\mu$$

امید ریاضی X را میانگین X نیز می‌نامند و با نماد μ نشان داده می‌شود.

مثال ۳–۱۹: فرض کنید متغیر تصادفی x دارای تابع احتمال زیر است؛ امید ریاضی X را محاسبهٔ کنید.

x_i	۰	۱	۲	۳
f_i	۱	۳	۳	۱
$\frac{f_i}{n}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$\frac{f_i}{n}x_i$	۰	$\frac{3}{8}$	$\frac{6}{8}$	$\frac{3}{8}$

$$E(X)=\sum\frac{f_i}{n}x_i=0+\frac{3}{8}+\frac{6}{8}+\frac{3}{8}=\frac{12}{8}=1/5$$

۳-۱۱ امید ریاضی تابعی از X

فرض کنید که X یک متغیر تصادفی با تابع چگالی $f(x)$ و $g(x)$ تابعی از X باشد؛ امید ریاضی $g(x)$ به صورت زیر خواهد بود:

$$E\left[g(x)\right]=\sum g(x).f(x)$$

مثال ۳-۲۰: فرض کنید متغیر تصادفی X دارای تابع احتمال زیر باشد؛ مطلوبست امید ریاضی $g(x)=4X^2-3X$

x_i	۱	۲	۳	٤	۵
f_i	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$g(x)=4x^2-3x$	۱	۱۰	۲۷	۵۲	۸۵
$g(x).f(x)$	$\frac{1}{8}$	$\frac{20}{8}$	$\frac{81}{8}$	$\frac{52}{8}$	$\frac{85}{8}$

$$\sum g(x).f(x)=\frac{239}{8}=29/875$$

۳-۱۲ ویژگی‌های امید ریاضی

ویژگی ۱: اگر تمامی داده‌ها با یک مقدار ثابت جمع شوند، آن گاه امید ریاضی نیز با آن مقدار ثابت جمع خواهد شد.

به عبارتی اگر میانگین $X_1, X_2, \ldots, X_n$ برابر با $E(X)$ باشد، آن گاه امید ریاضی $X_1+a, X_2+a, \ldots, X_n+a$ برابر خواهد بود با: (a مقدار ثابت است)

$$E(X+a)=E(X)+a$$

مثال۳-۲۱: اگر امید ریاضی $X_1, X_2, \ldots, X_n$ برابر با $E(X)=3$ باشد، آن گاه امید ریاضی را برای مقادیر مشخص شده $X_1+5, X_2+5, \ldots, X_n+5$ عبارت خواهد بود:

$$E(X+5)=E(X)+5=3+5=8$$

ویژگی ۲: اگر تمامی داده‌ها در یک مقدار ثابت ضرب شوند، آنگاه امید ریاضی نیز در آن مقدار ثابت ضرب خواهد شد. به عبارتی اگر میانگین $X_1, X_2, \ldots, X_n$ برابر با $E(X)$ باشد، آن گاه میانگین $bX_1, bX_2, \ldots, bX_n$ برابر خواهد بود با: (b مقدار ثابت است)

$$E(bX)=bE(X)$$

مثال ۳–۲۲: اگر امید ریاضی $X_1, X_2, \ldots, X_n$ برابر با $E(X)=3$ باشد، آن گاه امید ریاضی $5X_1, 5X_2, \ldots, 5X_n$ برابر خواهد بود با:

$$E(5X)=E(X)\times 5=3\times 5=15$$

ویژگی ۳: اگر تمامی داده‌ها با یک مقدار ثابت جمع و در یک مقدار ثابت ضرب شوند، آن گاه امید ریاضی نیز در آن مقدار ثابت، ضرب و با آن مقدار ثابت جمع خواهد شد.

به عبارتی اگر امید ریاضی $X_1, X_2, \ldots, X_n$ برابر با $E(X)$ باشد، آن گاه میانگین $bX_1+a, bX_2+a, \ldots, bX_n+a$ برابر خواهد بود با: (a مقدار ثابت است)

$$E(bX+a)=bE(X)+a$$

مثال ۳–۲۳: اگر امید ریاضی $X_1, X_2, \ldots, X_n$ برابر با $E(X)=3$ باشد، آنگاه امید ریاضی مقدیر $6X_1+5, 6X_2+5, \ldots, 6X_n+5$ برابر خواهد بود با:

$$E(6X+5)=6E(X)+5=(6\times 3)+5=23$$

ویژگی ٤: امید ریاضی یک مقدار ثابت برابر خودش است (a مقدار ثابت است).

$$E(a)=a$$

۳-۱۳ واریانس جامعه

فرض کنید X یک متغیر تصادفی با تابع چگالی $f(x)$ باشد، می‌توان واریانس را از فرمول زیر نیز محاسبهٔ کرد:

$$\sigma^2=E\left(X-E(X)\right)^2 \quad \text{یا} \quad \sigma^2=E\left(X^2\right)-\left[E(X)\right]^2$$

مثال ۳-۲۴: اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد، واریانس را محاسبهٔ کنید.

x_i	۱	۲	۳	٤	۵	٦
$f(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$
$xf(x)$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{15}{36}$	$\frac{28}{36}$	$\frac{45}{36}$	$\frac{66}{36}$
$x^2f(x)$	$\frac{1}{36}$	$\frac{12}{36}$	$\frac{45}{36}$	$\frac{112}{36}$	$\frac{255}{36}$	$\frac{396}{36}$

$$E(X)=\sum xf(x)=\frac{161}{36}=4/47$$

$$E\left(X^2\right)=\sum x^2f(x)=\frac{821}{36}=22/81$$

$$\operatorname{var}(X)=\sigma^2=E\left(X^2\right)-\left[E(X)\right]^2=22/81-(4/47)^2=2/8291$$

۳-۱٤ ویژگی‌های واریانس

ویژگی ۱: اگر تمامی داده‌ها با یک مقدار ثابت جمع شوند، آن گاه واریانس بدون تغییر می‌ماند.

به عبارتی اگر واریانس $X_1, X_2, \dots, X_n$ برابر با $\operatorname{var}(X)$ باشد، آن گاه واریانس $X_1 + \mathrm{a}, X_2 + \mathrm{a}, \dots, X_n + a$ برابر خواهد بود با: (a مقدار ثابت است)

$$\operatorname{var}(X + a) = \operatorname{var}(X)$$

مثال ۳–۲۵: اگر واریانس $X_1, X_2, \dots, X_n$ برابر با $\operatorname{var}(X) = 3$ باشد، آن گاه واریانس $X_1 + 5, X_2 + 5, \dots, X_n + 5$ برابر خواهد بود با:

$$\operatorname{var}(X + 5) = \operatorname{var}(X) = 3$$

ویژگی ۲: اگر تمامی داده‌ها در یک مقدار ثابت ضرب شوند، آن گاه واریانس در توان دوم آن مقدار ثابت ضرب خواهد شد. به عبارتی اگر واریانس $X_1, X_2, \dots, X_n$ برابر با $\operatorname{var}(X)$ باشد، آن گاه واریانس $bX_1, bX_2, \dots, bX_n$ برابر خواهد بود با: (b مقدار ثابت است)

$$\operatorname{var}(bX) = b^2 \operatorname{var}(X)$$

مثال ۳–۲۶: اگر واریانس $X_1, X_2, \dots, X_n$ برابر با $\operatorname{var}(X) = 3$ باشد، آن گاه واریانس $5X_1, 5X_2, \dots, 5X_n$ برابر خواهد بود با:

$$\operatorname{var}(5X) = \operatorname{var}(X) \times 5 = 3 \times 5^2 = 75$$

ویژگی ۳: اگر تمامی داده‌ها با یک مقدار ثابت، جمع و در یک مقدار ثابت ضرب شوند، آن گاه واریانس تنها در توان دوم آن مقدار ثابت ضرب خواهد شد.

به عبارتی اگر واریانس $X_1, X_2, \dots, X_n$ برابر با $\operatorname{var}(X)$ باشد، آن گاه واریانس $bX_1 + a, bX_2 + a, \dots, bX_n + a$ برابر خواهد بود با: (a مقدار ثابت است)

$$\operatorname{var}(bX + a) = b^2 E(X)$$

مثال۳–۲۷: اگر واریانس $X_1, X_2, \ldots, X_n$ برابر با $E(X)=3$ باشد، آن گاه واریانس $6X_1+5, 6X_2+5, \ldots, 6X_n+5$ برابر خواهد بود با:

$$\operatorname{var}(6X+5)=6^2\operatorname{var}(X)=(36\times 3)=108$$

ویژگی ٤: واریانس یک مقدار ثابت برابر صفر است (a مقدار ثابت است).

$$\operatorname{var}(a)=0$$

خلاصهٔ فصل سوم

در این فصل پس از تعریف متغیر تصادفی، قالب‌ها برای متغیرهای تصادفی گسسته و پیوسته ارائه شد. در بخش‌های دوم و سوم توزیع احتمالات گسسته و پیوسته ارائه و منجر به آشنایی چند توزیع مهم آماری شد. در ادامه ضمن معرفی توزیع‌های نرمال، نرمال استاندارد و توزیع استودنت نحوهٔ استفاده از جداول آماری توزیع نرمال استاندارد و استودنت مطرح و در پایان، تعریف میانگین و واریانس جامعه، برای چند توزیع نحوهٔ محاسبهٔ امید ریاضی و واریانس ارائه شد.

خودآزمایی تشریحی فصل سوم

۱- اگر متغیر تصادفی X دارای تابع چگالی زیر باشد:

x	۱	۲	۳	٤	٥	٦
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

مطلوبست: $E(X)$، $E(X^۲)$، $\sigma^۲$

۲- اگر متغیر تصادفی X دارای تابع چگالی زیر باشد:

x	۱-	۰	۱
$f(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$

مطلوبست: $E(X)$، $E(X^۲)$، $\sigma^۲=E\left(X-E(X)\right)^۲$ و $E(X^۲+X+۱)$.

۳- اگر ارزش وجه‌های ظاهر شده برای یک تاس به ترتیب، صفر، ۲، ٤، ٥، ٦ و ۷ باشد، میانگین μ و واریانس جامعه را به دست آورید.

٤- برای تابع چگالی زیر میانگین و واریانس را حساب کنید.

$$f(x)=\frac{1}{3} \qquad x=1,2,3$$

٥- اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد:

x	۱-	۰	۱	۲
$f(x)$	۰/۲	۰/۳	۰/۳	K

الف) مطلوبست: مقدار K

ب) تابع توزیع یا $F(x)$ را به‌دست آورید.

٦- در توزیع برنولی تابع توزیع و $E(X)$ را به‌دست آورید.

۷- ویژگی‌های متغیر تصادفی نرمال استاندارد را نام ببرید.

۸- اگر توزیع کتب نسخه‌های خطی دارای توزیع پواسن با پارامتر ۲ باشد، مطلوبست:

الف) $P(X=5)$ ب) $P(X \ge 10)$ ج) $P(X \le 100)$

۹- با استفاده از جدول توزیع نرمال استاندارد، احتمال‌های زیر را محاسبه کنید:

الف) $P(Z>0)$ ب) $P(Z<1)$ ج) $P(Z<0)$

د) $P(-1<Z<1)$ هـ) $P(|Z|\le 2)$

و) $P(-3/75<Z<3/75)$

۱۰- اگر متغیر تصادفی دارای تابع چگالی زیر باشد:

x	۲-	۱-	۰	۱	۲
$f(x)$	۰/۲	۰/۱	۰/۲	۰/۳	۰/۲

مطلوبست:

الف) $E(X)$ ب) $E(X^2)$ ج) $\text{var}(X)$ د) $\text{var}(5X+2)$

۱۱- با توجه به جدول نرمال استاندارد، احتمالات زیر را به دست آورید:

$P[Z>۱]$, $P[۱<Z<۱/۵]$, $P[Z<۱]$

$P[|Z|>-۱]$, $P[Z>۱/۶۴]$, $P[|Z|<۱/۹۶]$

۱۲- اگر متغیر تصادفی Z دارای توزیع نرمال استاندارد و $P(Z<-K)=a$ باشد، مقدار $P[-k<Z<k]$ را بر حسب a به دست آورید:

۱۳- سن کارکنان کتابخانه‌های کشور دارای توزیع نرمال با میانگین ۳۵ و انحراف معیار ۱۲ یعنی $X \sim N(۳۵,۱۴۴)$ است. اگر خط مشی مدیر کتابخانه‌ها، بازنشسته کردن تمام افرادی باشد که بیش از ۵۵ سال سن دارند، چند درصد از کارکنان بازنشسته می‌شوند.

۱٤- اگر نمرات پذیرفته‌شدگان دانشجویان در دانشگاه از توزیع نرمال با میانگین ۱۵ و واریانس ٤ باشد:

الف) احتمال اینکه نمرهٔ دانشجویی کم‌تر از ۱۷ باشد، چقدر است؟

ب) احتمال اینکه متوسط نمرهٔ ۱٦ دانشجو بیشتر از ۱٤/۵ باشد، چقدر است؟

۱۵- اگر $X_1, X_2, \ldots, X_n$ یک نمونهٔ تصادفی از جامعه نرمال با میانگین μ و واریانس مجهول σ^2 باشد، توزیع $\overline{X}$ را به دست آورید و نشان دهید که توزیع حدی آن دارای توزیع استودنت با $n-1$ درجهٔ آزادی است.

۱٦- اگر درآمد بخش فروش کپی کتاب‌های کتابخانه در یک نمونه به صورت زیر باشد (ارقام به ده هزار سال):

۲۱ ۱۹ ۱٤ ۱٦ ۸ ۱۲

الف) میانگین و واریانس نمونه را به دست آورید.

ب) مقدار $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ را به دست آورید، اگر $\mu = 12$ باشد.

۱۷- اگر $X_1, X_2, \ldots, X_n$ یک نمونه nتایی از توزیع نرمال با میانگین μ و واریانس σ^2 باشد، مطلوبست: $P\left[\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 0\right]$.

۱۸- اگر توزیع ورود اعضا به کتابخانه دارای توزیع یکنواخت زیر باشد، مطلوبست:

$$f(x) = \frac{1}{10} \quad , \quad 0 < x < 10$$

الف) $P(X > 2)$

ب) $P(3 \le X \le 7)$

ج) $P(X \le 8)$

۱۹- اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس ۲۵ باشد، تابع چگالی آن را بنویسید.

۲۰- با توجه به جدول نرمال استاندارد، احتمالات زیر را محاسبهٔ کنید.

$P(Z > ۰)$ $\qquad P(۰ < Z < ۱)$ $\qquad P(Z > ۱/۹۶)$

$P(|Z| \leq ۱/۹۶)$ $\qquad P(Z^۲ > ۲۵)$ $\qquad P(۳ < Z < ۲)$

۲۱- نمرهٔ کارایی کارکنان کتابخانه کشور دارای توزیع نرمال با میانگین ۱۵ و انحراف معیار ۲/۵ است.

الف) احتمال اینکه نمره کارایی یک نفر از کارکنان بین ۱۳ و ۱۷ باشد، چقدر است؟

ب) احتمال اینکه نمره کارایی یک نفر از کارکنان بیشتر از ۱۸ باشد، چقدر است؟

۲۲- وزن کارکنان مدرسه‌ای دارای توزیع نرمال با میانگین ٦٥ و انحراف معیار ٤ کیلوگرم است. از این مؤسسه ١٦ نفر از کارکنان به تصادف انتخاب می‌شوند، مطلوبست:

الف) مجموع وزن آن‌ها بیشتر از ١٠٥٦ کیلوگرم باشد.

ب) متوسط وزن آن‌ها بین ٦٢ و ٦٨ کیلوگرم باشد.

۲۳- فرض کنید میانگین و انحراف معیار نمرات نهایی آزمون ورودی به ترتیب ۷۰ و ١٤ باشد. اگر ۱۰ درصد از داوطلبان نمره الف، ۲۰ درصد نمرهٔ ب و ٤٠ درصد نمرهٔ ج آورده باشند، حداقل نمره برای قرار گرفتن در هر یک از رده‌ها چقدر است؟

٢٤- زمان لازم برای مونتاژ کردن قطعات ماشین، متغیر تصادفی، به طور تقریبی دارای توزیع نرمال با میانگین $\mu = ۱۲/۹$ و انحراف معیار $\sigma = ۲/۰$ است. مطلوبست احتمال آنکه برای مونتاژ کردن قطعات ماشین از این نوع:

الف) حداقل ۱۱/۵ دقیقه صرف شود.

ب) از ۱۱/۰ تا ۱٤/۸ دقیقه صرف شود.

۲۵- فرض کنید قد کارکنان کتابخانه دارای توزیع نرمال با میانگین ۱٦۵ و واریانس $\sigma^۲$ باشد. اگر قد ۹ نفر از کارکنان که به تصادف انتخاب شده‌اند به صورت زیر باشد:

۱۷۷ ۱٦۸ ۱۵۱ ۱٦۵ ۱٦٦ ۱۷۸ ۱۸۰ ۱۷۵

الف) احتمال اینکه یک نفر به تصادف انتخاب شود و دارای قد کم‌تر از ۱۷۲ باشد، چقدر است؟

ب) احتمال اینکه میانگین یک نمونهٔ ۲۵ تایی بین ۱٦۰ و ۱۷۰ باشد، چقدر است؟

۲٦- تجربه نشان داده است که میزان خطا در برش طول جلد کتب دارای توزیع نرمال با میانگین ۲ میلی‌متر است. اگر انحراف معیار یک نمونهٔ ۲۵ تایی برابر با ۰/۰۰۲۸ میلی‌متر باشد، مطلوبست:

الف) میانگین یک نمونهٔ ۱٦ تایی که بیشتر از ۲/۵ میلی‌متر باشد، چقدر است؟

ب) میانگین یک نمونهٔ ۳٦ تایی که بین ۰/۱۸ و ۰/۲۲ باشد، چقدر است؟

فصل چهارم

آزمون فرضیهٔ آماری

هدف کلی

آشنایی با مبانی آزمون فرضیه.

هدف‌های یادگیری

پس از مطالعهٔ این فصل، شما باید بتوانید:

۱. با مفاهیم اولیهٔ آزمون فرض آشنا شوید.

۲. آزمون فرض دربارهٔ میانگین جامعهٔ نرمال را وقتی که واریانس معلوم است، بنویسید و آن را آزمون کنید.

۳. آزمون فرض دربارهٔ میانگین جامعهٔ نرمال را وقتی که واریانس مجهول است، بنویسید و آن را آزمون کنید.

٤. آزمون فرض دربارهٔ نسبت جامعه را بنویسید و آن را آزمون کنید.

مقدمه

در پژوهش علمی، اغلب فرض‌هایی به کار برده می‌شود که ماهیتاً مملو از شک است. یک محقق هرگز نمی‌خواهد چیزی را باور کند، مگر اینکه اثبات شود. در مطالعات رسمی و آزمایش‌ها، دانشمندان نوعاً با این فرض که تفاوتی بین گروه‌های مختلف وجود نخواهد داشت، شروع می‌کنند. در مورد آزمون یک میانگین، نمونه به طور معنی‌دار از پارامتر جامعهٔ آماری تفاوت نمی‌کند. به عبارت دیگر فرض می‌شود که نمونه می‌تواند معرّف جامعهٔ آماری در نظر گرفته شود. همان‌گونه که ما چندین بار مشاهده کرده‌ایم، میانگین‌های نمونه از یکدیگر متفاوت هستند. انتخاب دو نمونه و به دست آوردن دقیق میانگین مشابه واقعاً نادر خواهد بود. بر اساس فرضیهٔ پوچ که معمولاً عدم‌معنی‌داری را بیان می‌کند، فرض می‌شود که این تفاوت‌های مشاهده شده مبتنی بر خطای نمونه‌گیری هستند. به هر حال، شرایطی ممکن است وجود داشته باشد که در آن فرضیهٔ پوچ غیرقابل دفاع باشد. به عنوان مثال ممکن است رویداد جانشین وجود داشته باشد که فرضیهٔ پوچ را نادرست سازد، تحت این شرایط، بایستی فرضیهٔ پوچ را به خاطر یک فرضیه بدیل رد بکنیم. در تحلیل آماری، فرض صفر یا پوچ نوعاً علیه یک فرض مقابل آزمون می‌شود. فرض مقابل معمولاً تأکید دارد که تفاوت‌های مشاهده شده به خطای نمونه‌گیری مربوط نیستند. به طور خلاصه، پژوهشگر تحقیق دو فرضیه برای انتخاب از بین آن‌ها را دارد. فرض مقابل تأکید می‌کند که تفاوت‌های مشاهده شده به شانس مربوط نیستند. انتخاب این دو فرض نیازمند قواعد تصمیم‌گیری اسمی است. برای انجام این تصمیم‌گیری تحلیل‌گر یک طرح ویژه برای عمل قبل از شروع آزمایش بسط می‌دهد. معمولاً فرض صفر را با $H_۰$ و فرض مقابل را با $H_۱$ نشان می‌دهند.

در ادامه به معرفی بعضی از انواع فرض‌ها می‌پردازیم.

۴-۱ مفاهیم پایه در آزمون فرض

فرض آماری: فرضی است که در مورد یک پارامتر، یعنی در مورد یک معیار آماری جامعه در نظر گرفته می‌شود. برای نمونه ممکن است ادعا شود میانگین یک جامعهٔ نرمال، بیشتر از پنج و یا ادعا شود متوسط طول قد در کشور بیشتر از ۱۶۵ است.
در ادامه انواع فرض‌های آماری به همراه انواع خطا ارائه می‌شود.

فرض آماری ساده: فرضی است که توزیع جامعه را کاملاً مشخص کند. برای مثال، اگر توزیع جامعهٔ آماری نرمال با میانگین μ و واریانس ۲/۲۵ باشد و ادعا شود $H_0 : \mu = 2$ است، آن گاه H_0 یک فرض آماری ساده است.

فرض آماری مرکب: فرضی است که توزیع جامعه تحت آن معلوم نیست؛ مثلاً ادعا شود که $H_1 : \mu > 2$ است. H_1 یک فرض مرکب است به لحاظ اینکه بیشتر از عدد ۲، بی‌شمار عدد وجود دارد.

خطای نوع اول: در آزمون فرض آماری، اگر فرض آماری درست باشد و ما آن را رد کنیم، مرتکب خطای نوع اول شده‌ایم.

خطای نوع دوم: در آزمون فرض آماری، اگر فرض آماری نادرست باشد و ما آن را بپذیریم، مرتکب خطای نوع دوم شده‌ایم.

اگر H_0 و H_1 به ترتیب فرض صفر و فرض مقابل صفر باشند، چهار ترکیب ممکن از این حالت‌ها و تصمیم‌ها در جدول زیر نشان داده می‌شود:

حالات ممکن / تصمیم	H_0 درست است	H_1 درست است
H_0 رد می‌شود	خطای نوع اول	تصمیم صحیح
H_1 پذیرفته می‌شود	خطای نوع دوم	تصمیم صحیح

در ادبیات آماری، احتمال ارتکاب خطای نوع اول را با α و احتمال ارتکاب خطای نوع دوم را با β نشان می‌دهند؛ یعنی:

$\alpha = P[H_0$ رد | درست است $H_0]$

$\beta = P[H_0$ قبول | درست است $H_1]$

علامت " | " نماد برای "به شرطی" است.

سطح اطمینان: $1-\alpha$ را سطح اطمینان یا سطح آزمون گویند و برابر است با:

$1-\alpha = 1-P[RH_0|$ درست $H_0] = P(AH_0|$ درست $H_0)$

توان آزمون: $1-\beta$ را توان آزمون گویند و برابر است با:

$1-\beta = 1-P[AH_0|$ درست $H_1] = P(RH_0|$ درست $H_1)$

آزمون‌های دو طرفه در مقابل یک طرفه

فرض مقابل H_0 یعنی H_1 را می‌توان به صورت یک‌طرفه یا دوطرفه بیان کرد. فرض $H_1: \mu \neq 0$ فرض دوطرفه است؛ زیرا فقط بیان می‌کند که μ مساوی صفر نیست، مشخص نمی‌کند که در چه جهتی (بیشتر یا کم‌تر) از صفر انحراف دارد. ولی فرض $H_0: \mu = 0$ در مقابل $H_1: \mu > 0$ فرض مقابل جهت‌دار و یک‌طرفه است.

۴-۲ آزمون فرض دربارهٔ میانگین جامعهٔ نرمال

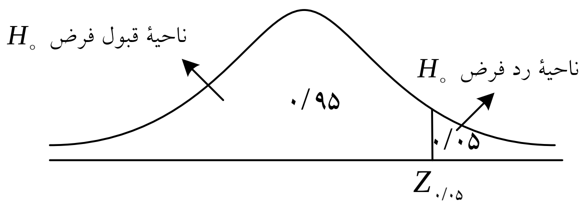
الف- واریانس معلوم

در این بخش، هدف آزمون دربارهٔ میانگین جامعهٔ نرمال وقتی که واریانس یا σ^2 معلوم است. برای مثال یک مدیر کتابخانه ممکن است ادعا کند، کسانی که از سالن مطالعهٔ کتابخانه استفاده می‌کنند متوسط نمرهٔ امتحانی ریاضی آن‌ها بیشتر از

۱۳ است. اگر در مقابل مدیر، ادعا شود متوسط نمره برابر با ۱۳ است، می‌توان ادعاها را به صورت زیر نوشت:

$$H_0 : \mu = 13 \quad , \quad H_1 : \mu > 13$$

فرض H_1 در مقابل H_0 یک فرض یک طرفه است. از آنجایی که توزیع نمرات نرمال است، لذا نواحی ادعا با اطمینان ۹۵٪ به صورت زیر خواهد بود:



برای آزمون فرض H_0 در مقابل H_1 ضرورت دارد از جامعهٔ مورد بررسی، نمونه $X_1, X_2, \dots, X_n$ به حجم n انتخاب شود. به طور شهودی $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$؛ میانگین نمونهٔ یک ملاک برای قبول یا رد H_0 است. با توجه به یک طرفه بودن H_1 مقادیر بزرگ $\bar{X}$ باعث قبول H_1 و رد H_0 می‌شود.

از آنجا که میانگین نمونه دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است، لذا استاندارد $\bar{X}$ ملاک برای تصمیم‌گیری خواهد بود؛ یعنی:

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

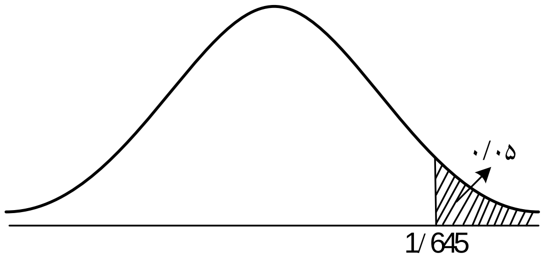
از مقایسهٔ $Z_{\bar{x}}$ و $Z_{0/05}$ می‌توان تصمیم‌گیری کرد. در این حالت اگر $Z_{\bar{x}}$ بیشتر از مقدار $Z_{0/05}$ باشد، فرض H_0 رد می‌شود؛ در غیر این صورت پذیرفته می‌شود.

مثال ٤–۱: اگر نمرهٔ ۹ نفر به صورت زیر گزارش شده باشد، فرض $H_0: \mu = 17$ را در مقابل $H_1: \mu > 17$ با اطمینان ۹۵٪ آزمون کنید. فرض کنید واریانس جامعه ۲/۲۵ باشد.

۱٤ ۱۰ ۱۷ ۱۸ ۱۱ ۱۲ ۱٦ ۱۳ ۱۵

مراحل آزمون

۱. مقدار $Z_{0/05}$ را از جدول نرمال استاندارد محاسبهٔ می‌کنیم که مقدار آن برابر است با ۱/٦٤٥.



۲. پس از محاسبهٔ $\overline{X}$، مقدار $Z_{\bar{x}}$ را به صورت زیر محاسبهٔ می‌کنیم.

$$Z_{\bar{x}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{-3}{\frac{1/5}{\sqrt{9}}} = \frac{-9}{1/5} = -6$$

۳. تصمیم‌گیری: چون مقدار $Z_{\bar{x}} = -6$ از مقدار جدول $Z_{0/05} = 1/65$ کم‌تر است، فرض H_0 پذیرفته می‌شود.

برای آزمون فرض $H_0: \mu = \mu_0$ در مقابل $H_1: \mu < \mu_0$ یا $H_1: \mu \neq \mu_0$، مجدداً مقدار آماره $Z_{\bar{x}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ را محاسبهٔ و آن را با توجه به ناحیهٔ رد تحت H_1 مقایسه و تصمیم‌گیری می‌کنیم. برای فرض $H_1: \mu < \mu_0$ اگر مقدار $Z_{\bar{x}}$ کم‌تر از Z_α باشد، فرض H_0 رد می‌شود و برای فرض $H_1: \mu \neq \mu_0$ اگر مقدار

$Z_{\bar{x}}$ کم‌تر از $-Z_{\frac{\alpha}{2}}$ یا بیشتر از $Z_{\frac{\alpha}{2}}$ باشد، فرض H_0 رد می‌شود، در غیر این صورت پذیرفته می‌شود.

مثال ٤-۲: مدیر خرید مجلات کتابخانه‌های ادعا می‌کند که متوسط قیمت مجلات خریداری شده کم‌تر از ۲ هزار تومان است. اگر انحراف معیار ۱/۵ تومان باشد، براساس نمونهٔ زیر، ادعا را با اطمینان ۰/۹۷۵ آزمون کنید (فرض کنید توزیع مشاهدات نرمال است).

۲۰۰۱ ۲۰۰۹ ۲۰۰۰ ۲۰۰۶ ۲۰۰۴

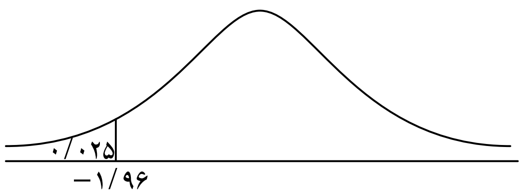
$$H_0 : \mu = 2000$$

$$H_1 : \mu < 2000$$

$$n = 5 \quad , \quad \alpha = 0/025 \quad , \quad z_{0/025} = 1/96 \quad , \quad \sigma = 1/5$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}\left[2001 + 2009 + 2006 + 2004\right] = 2004$$

$$Z_0 = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2004 - 2000}{\frac{1/5}{\sqrt{5}}} = \frac{4\sqrt{5}}{1/5} = \frac{4\sqrt{5}}{1/5} = 5/9628$$



چون $Z_0 > Z_{0/025}$ فرض H_0 پذیرفته می‌شود.

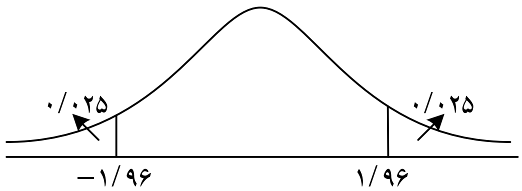
مثال ٤-۳: رئیس کتابخانه‌های ادعا می‌کند که متوسط نمرات آزمون دانشجویان پذیرفته شده در دانشگاه‌ها که عضو کتابخانه بودند ٤٨٠ است؛ اگر واریانس

نمرات ۲/۲۵ باشد و نمرات آزمون ۹ عضو پذیرفته شدهٔ کتابخانه به صورت زیر باشد، ادعا را با اطمینان ۹۵٪ آزمون کنید.

۴۸۱	۴۸۰	۴۷۹	۴۸۲	۴۸۱	۴۸۳	۴۷۵	۴۸۵	۴۸۳

$$n=۹ \quad , \quad \alpha=۰/۰۵ \quad , \quad \sigma^{۲}=۲/۲۵ \quad , \quad \sigma=۱/۵$$

$$\left| \begin{array}{l} H_{\circ}: \mu = ۴۸۰ \\ H_{۱}: \mu \neq ۴۸۰ \end{array} \right.$$



$$\overline{X}=\frac{۱}{n}\sum_{i=۱}^{n} x_i=\frac{۱}{۹}\sum_{i=۱}^{۹} x_i = ۴۸۱$$

$$Z_{\circ}=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{۴۸۱-۴۸۰}{\frac{۱/۵}{\sqrt{۹}}}=\frac{۳}{۱/۵}=۲$$

چون $Z_{\circ}=۲$ بیشتر از ۱/۹٦است، فرض $H_{\circ}$ رد می‌شود.

ب- واریانس مجهول

در بخش قبل، برای آزمون فرض‌های زیر

$$\left| \begin{array}{l} H_{\circ}: \mu = \mu_{\circ} \\ H_{۱}: \mu > \mu_{\circ} \end{array} \right. \qquad \left| \begin{array}{l} H_{\circ}: \mu = \mu_{\circ} \\ H_{۱}: \mu < \mu_{\circ} \end{array} \right. \qquad \left| \begin{array}{l} H_{\circ}: \mu = \mu_{\circ} \\ H_{۱}: \mu \neq \mu_{\circ} \end{array} \right.$$

از آماره آزمون $Z_{\circ}=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ استفاده شد که σ انحراف معیار جامعه از قبل معلوم و یا مقدار آن از تحقیق قبلی موجود بود، استفاده شد. در این بخش هدف آزمون فرض‌های فوق با مجهول بودن انحراف معیار یا واریانس است. معمولاً در جامعهٔ نرمال وقتی که واریانس مجهول است از براوردگر آن یعنی واریانس نمونه استفاده می‌شود؛ یعنی:

$$\hat{\sigma}^{۲} = S^{۲} = \frac{۱}{n-۱}\sum\left(X_i - \bar{X}\right)^{۲}$$

در واقع می‌توان گفت که واریانس نمونهٔ یک تقریب برای واریانس جامعه است و همچنین ثابت می‌شود که $S^{۲}$ یک براوردگر خوبی برای $\sigma^{۲}$ است. در این حالت آماره آزمون فرض‌ها به صورت زیر خواهد بود:

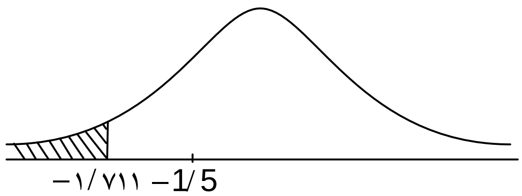
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

که از توزیع استودنت (t) با $d \cdot f = n-۱$ درجهٔ آزادی پیروی می‌کند. بنابراین در آزمون فرض‌ها مقدار T را محاسبهٔ و آن را با مقدار جدول استودنت با $n-1$ درجه آزادی مقایسه و تصمیم‌گیری می‌کنیم.

مثال ٤-٤: مسئول پارکینگ کتابخآن‌های ادعا می‌کند که متوسط زمان برای یافتن مکان برای پارک کم‌تر از ۵ دقیقه است. در یک بررسی از ۲۵ مورد متوسط زمان و انحراف معیار به ترتیب ٤/٧ و ١ دقیقه است. ادعا را با اطمینان ٩٥٪ آزمون کنید (فرض بر اینکه توزیع جامعهٔ نرمال باشد).

$$n = ۲۵ \quad , \quad \bar{X} = ۴/۷ \quad , \quad S = ۱ \quad , \quad S^{۲} = ۱ \quad , \quad \alpha = ۰/۰۵ \quad , \quad d \cdot f = ۲۴$$

$$H_{۰} : \mu = ۵ \qquad\qquad H_{۱} : \mu < ۵$$



$$t_{۰} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{۴/۷ - ۵}{\frac{۱}{\sqrt{۲۵}}} = \frac{-۰/۳ \times ۵}{۱} = -۱/۵$$

$t_{(۲۴,۰/۰۵)} = -۱/۷۱۱$

چون $t_{(۲۴,۰/۰۵)} = -۱/۷۱۱ < t_{\circ} = -۱/۵$ فرض $H_{\circ}$ پذیرفته می‌شود.

مثال ٤-٥: متوسط نمرات دانشجویان دانشگاهی برابر با ١٦ است. نمره ٩ دانشجو به صورت زیر گزارش شده است.

۱۶	۱۹	۱۵	۵	۱۲	۲۰	۱۳	۱۸	۱۷

با توجه به اطلاعات به دست آمده ادعا را با اطمینان ۹۵٪ آزمون کنید.

هدف آزمون فرض $H_{\circ}: \mu = 16$ در مقابل $H_1: \mu \neq 16$

$$n = ۹ \quad , \quad \alpha = ۰/۰۵ \quad , \quad \frac{\alpha}{۲} = ۰/۰۲۵ \quad , \quad d \cdot f = ۸$$

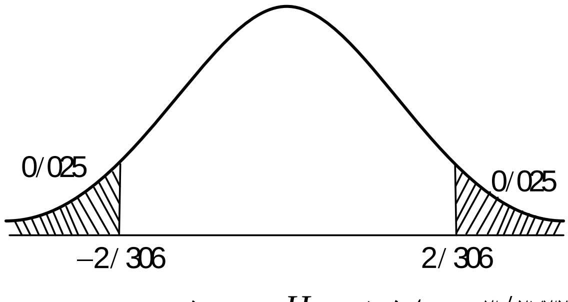
$$\bar{X} = \frac{۱}{n}\sum x_i = \frac{۱}{۹}(۱۳۵) = ۱۵$$

$$S^{۲} = \frac{۱}{n-۱}\sum\left(X_i - \bar{X}\right)^{۲} = \frac{۱}{۸}\sum\left(X_i - ۱۵\right)^{۲}$$

$$S^{۲} = \frac{۱}{n}\left[۴+۹+۴+۲۵+۹+۱۰۰+۰+۱۶+۱\right] = ۲۱$$

$$t_{\circ} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{۱۵-۲۰}{\frac{\sqrt{۲۱}}{\sqrt{۹}}} = \frac{-۱۵}{\sqrt{۲۱}} = \frac{-۱/۵}{۴/۵۸۲۶} = -۳/۲۷۳۲$$

$t_{(۸,۰/۰۲۵)} = ۲/۳۰۶$



چون $t_{\circ} = -۳/۲۷۳۲ < ۲/۳۰۶$ فرض $H_{\circ}$ رد می‌شود.

۴-۳ آزمون فرض دربارهٔ نسبت جامعه

در این بخش، هدف مطالعه نسبت در جامعه است. برای نمونه نسبت افراد کم‌درآمد، نسبت کتب از رده خارج شده، نسبت باسوادان جامعه و غیره. نسبت واحدهای جامعه که ویژگی مورد نظر را دارا هستند با P نشان داده می‌شود. درواقع P برابر با تعداد واحدهایی است که ویژگی مورد نظر را دارند، تقسیم بر تعداد کل واحدهای جامعه. اگر جامعهٔ مورد بررسی متناهی و قابل‌شمارش و دارای حجم N و تعداد واحدهایی که صفت مورد بررسی را دارند، برابر با A باشد، نسبت از رابطه زیر به دست می‌آید:

$$P = \frac{A}{N}$$

از آنجایی که P بین صفر و یک است، معمولاً آن را به درصد بیان می‌کنند؛ به عبارتی درصد واحدهایی که صفت مورد بررسی را دارند، از رابطه زیر به دست می‌آید:

$$\%P = \frac{A}{N} \times ۱۰۰$$

مثال ۴-۶: اگر کتابخانه‌ای دارای ۱۰۰۰۰ جلد کتاب باشد و از این تعداد ۸۰۰۰ جلد به زبان فارسی باشد، نسبت و درصد کتب فارسی برابر است با:

$$N = 10000 \quad , \qquad A = 8000 \qquad P = \frac{A}{N} = \frac{۸۰۰۰}{۱۰۰۰۰} = ۰/۸$$

$$\%P = \frac{A}{N} \times ۱۰۰ = ۰/۸ \times ۱۰۰ = ۸۰$$

در این بخش، هدف آزمون این است که در یک جامعهٔ نسبتاً بزرگ، نسبت جامعه یا P که یک ویژگی بخصوص را داراست برابر با $P_۰$ است یا نه؛ یعنی:

$$H_۰ : P = P_۰ \qquad\qquad H_۱ : P \neq P_۰$$

که تغییر $P_{\circ}$ بین صفر و یک است.

برای آزمون فرض $H_{\circ}$ در مقابل $H_{۱}$ از جامعهٔ مورد بررسی، یک نمونهٔ تصادفی به حجم n انتخاب می‌کنیم. اگر f تعداد واحدهایی باشد که ویژگی مورد نظر را داشته باشند، نسبت نمونه از رابطه زیر به دست می‌آید:

$$p = \frac{f}{n}$$

و p یک براوردکننده برای نسبت جامعه یا P است. برای محاسبهٔ امید ریاضی و واریانس p متغیر دو حالتی یا دو وضعیتی X_i را به صورت زیر تعریف می‌کنیم:

$$X_i = \begin{cases} ۱ & \text{اگر نمونهٔ انتخابی صفت مورد بررسی را داشته باشد.} \\ \circ & \text{اگر نمونهٔ انتخابی صفت مورد بررسی را نداشته باشد.} \end{cases}$$

در این صورت مقدار f از رابطه زیر به دست می‌آید:

$$f = X_{۱} + X_{۲} + \ldots + X_n = \sum_{i=۱}^{n} X_i$$

و میانگین X_i برابر است با:

$$p = \frac{f}{n} = \frac{۱}{n}\sum_{i=۱}^{n} X_i = \overline{X}$$

می‌توان گفت نسبت نمونهٔ p تمام ویژگی‌های میانگین نمونهٔ $\overline{X}$ در براورد میانگین جامعه را دارد. بنابراین می‌توان به راحتی نشان داد که برای n های به اندازه بزرگ روابط زیر برقرار است:

$$E(p) = P \qquad , \qquad \operatorname{var}(p) = \frac{P(۱-P)}{n}$$

برای آزمون فرض $H_{\circ}$ در مقابل $H_{۱}$، آماره آزمون تحت $H_{\circ}$ برابر است با:

$$Z_{\circ} = \frac{p - P_{\circ}}{\sqrt{\frac{P_{\circ}(۱-P_{\circ})}{n}}}$$

لذا برای آزمون فرض H_0، مقدار Z_0 را محاسبهٔ و آن را با مقدار جدول مقایسه و تصمیم‌گیری می‌کنیم.

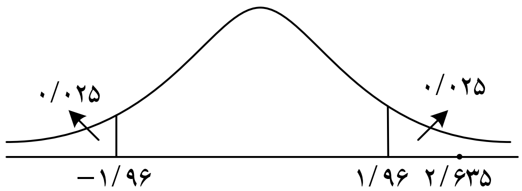
مثال ٤-٧: مدیر کتابخآن‌های ادعا می‌کند که در این کتابخانه ٣٦ درصد از کتب، لاتین است. در یک بررسی آماری از ۱۰۰۰ جلد کتاب مشخص می‌شود که ٤٠٠ جلد از کتب به زبان لاتین هستند. ادعا را با اطمینان ۹۵٪ آزمون کنید.

$$H_0 : P = ۰/۳۶ \quad , \quad H_1 : P \neq ۰/۳۶ \quad , \quad \alpha = ۰/۰۵$$

$$n = ۱۰۰۰ \quad , \quad f = ۴۰۰ \quad , \quad p = \frac{۴۰۰}{۱۰۰۰} = ۰/۴$$

$$Z_0 = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{۰/۴ - ۰/۳۶}{\sqrt{\frac{۰/۳۶(۱-۰/۳۶)}{۱۰۰۰}}} = \frac{۰/۰۴}{\sqrt{\frac{۰/۳۶ \times ۰/۶۴}{۱۰۰۰}}} = \frac{۴۰}{\sqrt{\frac{۳۶ \times ۶۴}{۱۰}}}$$

$$= \frac{۴۰\sqrt{۱۰}}{۶ \times ۸} = ۲/۶۳۵$$



چون مقدار $Z_0 = ۲۰۶۳۵$ بیشتر از مقدار جدول $Z_{۰/۰۲۵} = ۱/۹۶$است، فرض H_0 رد می‌شود.

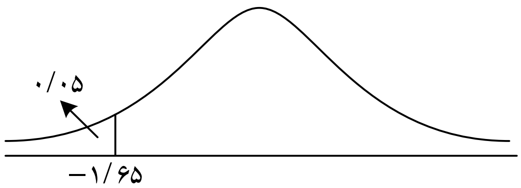
مثال ٤-٨: ادعا می‌شود قیمت کتب نسبت به سال گذشته ۳۰ درصد کاهش یافته است. اگر قیمت پایه ۱۵۰۰ باشد و قیمت ١٦ کتاب که به تصادف انتخاب شده باشد دارای قیمت‌های زیر باشد، فرض زیر را با اطمینان ۹۵٪ آزمون کنید:

$$H_0 : P = ۰/۳ \quad , \quad H_1 : P < ۰/۳$$

۲۴۰۰	۲۵۰۰	۱۹۹۵	۱۶۵۰	۲۰۲۵	۲۰۰۲	۱۸۵۰	۲۱۰۰
۲۶۰۰	۱۹۹۵	۲۲۵۰	۱۴۵۰	۱۸۷۰	۲۰۰۷	۲۱۱۵	۲۷۰۰

$$n = ۱۶ \qquad , \qquad f = ۱۰ \qquad , \qquad p = \frac{۱۰}{۱۶} = ۰/۶۲۵$$

$$Z_\circ = \frac{p - P_\circ}{\sqrt{\frac{P_\circ(۱ - P_\circ)}{n}}} = \frac{۰/۶۲۵ - ۰/۳}{\sqrt{\frac{۰/۳ \times ۰/۷}{۱۶}}} = ۲/۸۳۷$$



چون $Z_\circ = ۲/۸۳۷$ بیشتر از $Z_{۰/۰۵} = -۱/۶۵$ است، فرض $H_\circ$ پذیرفته می‌شود.

خلاصهٔ فصل چهارم

در این فصل پس از ارائهٔ مفاهیم اولیهٔ آزمون فرض شامل خطای نوع اول، خطای نوع دوم و سطح معنی‌داری و توان آزمون نحوهٔ آزمون فرض دربارهٔ میانگین جامعهٔ نرمال در حالتی که واریانس جامعه معلوم و مجهول است ارائه شد و در بخش آزمون فرض دربارهٔ نسبت جامعه ارائه و مورد آزمون قرار گرفت.

خودآزمایی تشریحی فصل چهارم

۱- الف) خطاهای نوع اول و نوع دوم را تعریف کنید.

ب) طریقهٔ محاسبهٔ $\alpha-۱$ و $\beta-۱$ را توضیح دهید.

۲- انواع فرض‌ها دربارهٔ میانگین جامعهٔ نرمال را وقتی که واریانس‌ها معلوم است، بنویسید و نواحی قبول و رد هر فرضی را مشخص کنید.

۳- ادعا می‌شود متوسط نمرات مهارت مدیران سازمانی ۶۰ است. در بررسی آمار، آماری از ۲۵ مدیر، میانگین و انحراف معیار نمونه به ترتیب ۶۲ و ۵ به دست آمده است. ادعا را با اطمینان ۹۵٪ آزمون کنید.

۴- برای آزمون فرض $H_0: \mu = ۱۲$ در مقابل $H_1: \mu \neq ۱۲$ میانگین یک نمونهٔ ۱۶ تایی برابر با ۱۳/۷۵ است. اگر توزیع جامعه، نرمال و دارای واریانس ۱/۶۹ باشد، فرض H_0 را با اطمینان ۹۵٪ آزمون کنید.

۵- ادعا می‌شود که متوسط نمرات کارکنان استخدام شده در کتابخانه بیشتر از ۱۵ است. برای بررسی این ادعا با اطمینان ۹۵٪ اطلاعات نمونه‌ای زیر به دست آمده است:

۱۲ ۱۴ ۱۷ ۱۸ ۱۵ ۱۶ ۱۹ ۲۰ ۱۳

اگر واریانس جامعه برابر با ۶/۲۵ باشد، فرض H_0 را در مقابل H_1 آزمون کنید.

۶- شرط استخدام در کتابخانه‌های کشور برای متقاضیان داشتن نمرهٔ ضریب هوشی، بیشتر از ۹۰ است. اگر نمرهٔ ضریب هوشی ۱۶ نفر به صورت زیر باشد، فرض H_0 را در مقابل H_1 با اطمینان ۹۵٪ آزمون کنید.

۸۵	۸۶	۸۹	۹۰	۹۳	۹۵	۹۶	۱۰۰
۱۰۱	۹۲	۹۱	۸۸	۸۷	۹۲	۹۰	۸۰

۷- برای آزمون فرض $H_0: P = 0/2$ در مقابل $H_1: P \neq 0/2$ اطلاعات زیر آمده است:

$n = 400$ $f = 300$

فرض را با اطمینان ۹۵٪ آزمون کنید.

۸- اگر نمرات ۹ نفر به صورت زیر و ملاک قبولی در آزمون بیشتر یا برابر با ۱۸ باشد، فرض $H_{\circ}: P = 082$ را در مقابل $H_1: P > 0/8$ با اطمینان ۹۵٪ آزمون کنید.

۱۳ ۲۰ ۱۱ ۱۶ ۱۹ ۱۵ ۱۰ ۱۴ ۱۷

۹- در بررسی کتب کتابخانه‌ای بزرگ، اطلاعات زیر بدست آمده است.

۱ ۱ ۱ ۱ ۰ ۱ ۱ ۱ ۱ ۰ ۱ ۱ ۱ ۱ ۱ ۰ ۱ ۱ ۱ ۰

اگر کد ۱ سالم بودن کتاب در نمونه باشد. آیا می‌توان گفت که ۸۰ درصد از کتب سالم هستند.

۱۰- انواع فرض‌های آماری درباره نسبت جامعه را به همراه نواحی قبول و رد بنویسید.

فصل پنجم

همبستگی و انواع آن

هدف کلی

آشنایی با انواع ضریب همبستگی.

هدف‌های یادگیری

پس از مطالعهٔ این فصل، شما باید بتوانید:

۱. همبستگی دو متغیر را تعریف کنید.

۲. نمودار پراکنش بین مشاهدات دو متغیر را رسم کنید.

۳. ضریب همبستگی پیرسون را برای مشاهدات دو متغیر محاسبهٔ کنید.

٤. با ضریب تعیین آشنا و آن را محاسبهٔ کنید.

۵. ضریب همبستگی رتبه‌ای اسپیرمن را برای مشاهدات دو متغیر محاسبهٔ کنید.

٦. آزمون فرض دربارهٔ ضریب همبستگی جامعه (ρ) را مورد آزمون قرار دهید.

مقدمه

همبستگی مفهومی است که در زندگی روزمرهٔ خود با آن بسیار سر و کار داریم. به تعبیر دیگر، ممکن است مفهوم همبستگی را در گفتار روزمره به کار برده باشیم، اما ممکن است تاکنون از نقطه نظر علم آمار به آن توجه نکرده باشیم. به عنوان مثال، می‌دانیم که با توسعهٔ یک کتابخانه تعداد کتب موجود در آن نیز افزوده می‌شود. برای بیان این مثال به صورت آماری می‌توان گفت که ارتباط مستقیمی (مثبتی) میان اندازهٔ کتابخانه و تعداد کتب آن وجود دارد. بر اساس همبستگی موجود میان این دو متغیر، مدیران کتابخانه‌ها و کتابداران تعداد کتب را بر اساس اندازهٔ کتابخانه براورد می‌کنند. از این رو، فعالیت‌های کتابخانه بر این اساس برچسب زده می‌شود. از دیدگاه علم آمار به چنین براوردی، رگرسیون می‌گویند که در فصل ششم به آن خواهیم پرداخت.

البته استثناهایی نیز در ارتباط با این قاعدهٔ کلی (ارتباط مستقیم اندازهٔ کتابخانه و تعداد کتب) وجود دارد. کتابخانه‌هایی وجود دارند که به دلیل عضویت کم و تقاضای محدود برای خریداری منابع جدید، به شکل چشمگیری تعداد منابعشان افزایش نیافته است و نیز تمام کتابخانه‌های بزرگ و یا کوچک دارای تعداد منابع برابر نیستند. از دیدگاه علم آمار، می‌توان گفت همبستگی میان اندازهٔ کتابخانه و تعداد منابع، کامل یا قوی نیست. در حقیقت، همبستگی شدت، نوع و معناداری آماری در ارتباط میان متغیرها را مشخص می‌کند. در این فصل دربارهٔ مفاهیم بنیادی، اندازه گیری و همبستگی بحث می‌کنیم.

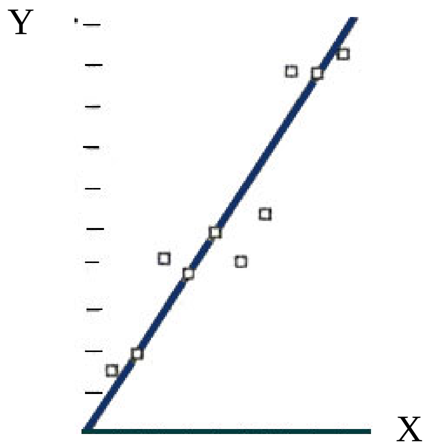
۵-۱ انواع همبستگی

یکی از تعاریف اساسی در علم آمار، تعریف همبستگی و رابطهٔ بین دو متغیر است. به طور کلی شدت وابستگی دو متغیر به یکدیگر را همبستگی تعریف

می‌کنیم. اگر X و Y دو متغیر باشند، ضریب همبستگی بین آن‌ها در جامعه برابر است با:

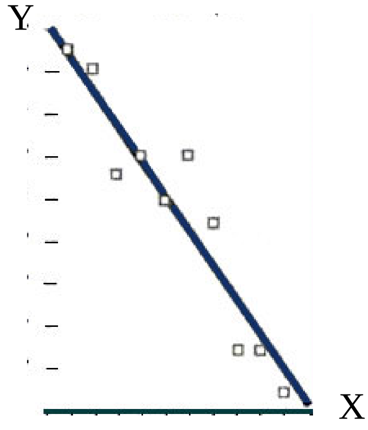
$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X).\operatorname{var}(Y)}}$$

ضریب همبستگی ممکن است مثبت یا منفی باشد. علامت ضریب همبستگی نشان می‌دهد که همبستگی بین دو متغیر مثبت است (یعنی مقادیر هر دو متغیر روندی فزاینده یا هر دو روندی کاهنده دارند). به عنوان مثال با افزایش تعداد عضویت‌های یک کتابخانه (X) تعداد امانت (Y) افزوده می‌شود و یا با کاهش تعداد ساعات کاری یک کتابخانه مراجعه به کتابخانه کم می‌شود.



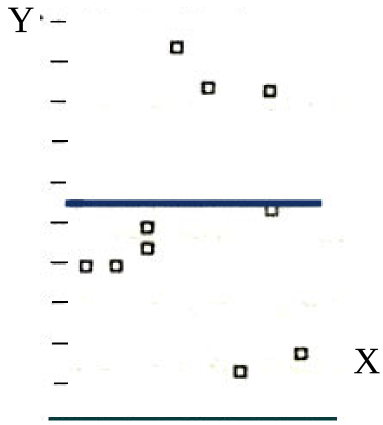
نمودار ۵-۱: همبستگی مثبت

ضریب همبستگی ممکن است منفی باشد؛ (مقادیر یک متغیر روندی فزاینده و مقادیر متغیر دیگر روندی کاهنده دارند). به عنوان مثال با افزایش دسترسی به منابع الکترونیکی (X)، میزان استفاده از منابع چاپی (Y) کاهش یافته است.



نمودار ۵-۲: همبستگی منفی

ضریب همبستگی ممکن است صفرباشد؛ یعنی هیچ رابطهٔ خطی بین متغیرها وجود نداشته باشد.



نمودار ۵-۳: همبستگی وجود ندارد

همبستگی همواره مقداری است بین ۱ و ۱- ؛ اگر این مقدار دقیقاً ۱ و یا ۱- باشد، نشاندهندهٔ یک رابطهٔ کاملاً خطی بین متغیرهای X و Y است و اگر مقدار ضریب همبستگی در نزدیکی صفر باشد، نتیجه می‌شود که بین متغیرها رابطه‌ای وجود ندارد و یا بسیار ضعیف است.

نمودار پراکنش

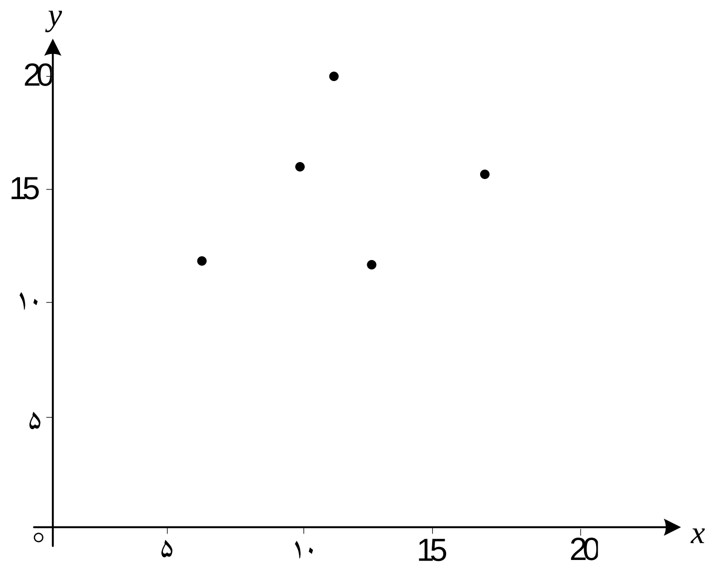
ساده‌ترین روش بـرای تشـخیص رابطـهٔ بـین دو متغیـر X و Y رسـم نمـودار پراکنش است. اگر $x_1, x_2, \ldots, x_n$ و $y_1, y_2, \ldots, y_n$ نمونهٔ مشاهدات از متغیرهای X و Y به صورت زیر باشد:

شماره	۱	۲	۳	٤	...	n
X	$x_۱$	$x_۲$	$x_۳$	$x_۴$	...	x_n
Y	$y_۱$	$y_۲$	$y_۳$	$y_۴$	...	y_n

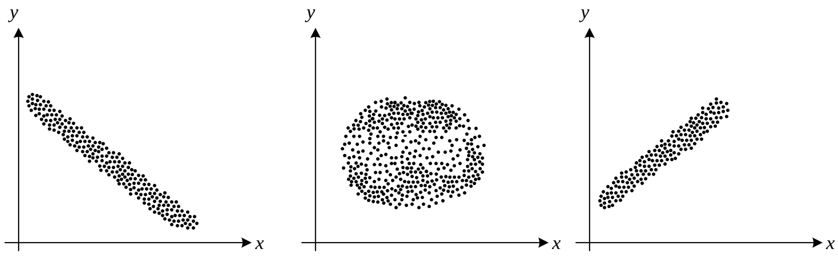
برای رسم نمودار پراکنش، یکی از متغیرها روی محور Xها و دیگری روی محور Yها علامت‌گذاری و نقطه‌یابی می‌شود؛ نقاط حاصل، نمودار پراکنش خواهد شد.

مثال ۵-۱: اگر نمرهٔ امتحانی (Y) متناسب با مدت زمان مطالعه بر حسب ساعت (X) برای ۵ دانشجو به صورت زیر باشد، نمودار پراکنش را رسم کنید.

شماره دانشجو	۱	۲	۳	٤	۵
X	۱۰	۱۲	۱۵	۸	٦
Y	۱۷	۱۹	۱٦	۱۳	۱۳



اگر تعداد مشاهدات زیاد باشد، مثلاً برای بررسی رابطهٔ بین درآمد (X) و هزینهٔ زندگی (Y) بین خانوارها در کشور که شامل حدوداً ۱۷ میلیون خانوار است، ۱۰۰۰ خانوار انتخاب و نمودار پراکنش را رسم کنیم، نمودار پراکنش به یکی از سه صورت زیر خواهد بود:



شکل: ج شکل: ب شکل: الف

در شکل الف، مشخص است که اگر (X) یا درآمد زیاد شود، هزینهٔ زندگی (Y) زیاد می‌شود.

شکل ب مشخص نمی‌کند که اگر (X) زیاد یا کم شود، هزینهٔ زندگی (Y) کم یا زیاد می‌شود.

در شکل ج مشخص است که اگر درآمد (X) زیاد شود، هزینهٔ زندگی (Y) کم می‌شود.

۵-۲ ضریب همبستگی

در بسیاری اوقات نیاز به شاخصی داریم که چگونگی ارتباط بین دو متغیر X و Y را اندازه بگیرد. این شاخص، ضریب همبستگی خطی بین دو متغیر X و Y نامیده می‌شود.

در آمار انواع زیادی از ضرایب همبستگی وجود دارند که هر کدام همبستگی بین دو متغیر را با توجه به نوع داده‌ها و شرایط متغیرها اندازه‌گیری می‌کنند. در ادامه به معرفی بعضی از انواع ضریب همبستگی می‌پردازیم.

۵-۳ ضریب همبستگی پیرسون

با بررسی نمودار پراکندگی می‌توان دریافت که آیا دو متغیر با هم ارتباط دارند و جهت ارتباط آن‌ها چگونه است. اگرچه یک نمودار پراکندگی به تنهایی برای اندازه‌گیری ارتباط بین دو متغیر کافی نیست و رابطه‌ای که با استفاده از نمودار پراکندگی مشخص می‌شود باید به طور عددی توصیف شود. در آمار توصیفی، میزان ارتباط بین متغیرها را ضریب همبستگی می‌نامند. ضریب همبستگی‌ای که برای نمرات در سطوح نسبی و فاصله‌ای به کار می‌رود ضریب همبستگی پیرسون است و آن را با r نشان می‌دهند. ضریب همبستگی پیرسون از فرمول زیر به دست می‌آید.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\cdot\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

این فرمول را می‌توان به صورت زیر نیز نوشت:

$$r = \frac{\sum_{i=1}^{n} X_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\cdot\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

ضریب همبستگی نمونه‌ای یا ضریب همبستگی پیرسون دارای ویژگی‌های زیر است:

۱- مستقل از واحد اندازه‌گیری است و مقدار آن بین $[-1,1]$ است.

۲- یک براورد برای ضریب همبستگی جامعه یا ρ است.

۳- اگر r نزدیک به یک باشد، همبستگی شدید همسو است؛ یعنی یکی از متغیرها زیاد شود، دیگری نیز زیاد می‌شود.

٤- اگر r نزدیک به منهای یک باشد، همبستگی شدید و غیرهمسو است.

مثال ۵-۲: اگر مشاهدات برای نمرهٔ میان‌ترم و پایان‌ترم ۱۰ نفر به صورت زیر باشد، ضریب همبستگی پیرسون را حساب کنید.

i	۱	۲	۳	٤	۵	٦	۷	۸	۹	۱۰
x_i	٤	۲۷	۱۸	۷	۳۰	۱۲	۱۸	۲۳	۱۹	۱۲
y_i	۱٦	۳۷	۳۳	۲۳	۳٤	۳۲	۲٤	۲۹	۲٦	۲٦

برای محاسبهٔ ضریب همبستگی پیرسون بین نمرهٔ میان‌ترم (X) و نمرهٔ پایان‌ترم (Y) از جدول زیر استفاده می‌کنیم:

i	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^{۲}$	y_i	$y_i-\bar{y}$	$(y_i-\bar{y})^{۲}$	$(x_i-\bar{x})(y_i-\bar{y})$
۱	٤	−۱۳	۱٦۹	۱٦	−۱۲	۱٤٤	۱٥٦
۲	۲۷	۱۰	۱۰۰	۳۷	۹	۸۱	۹۰
۳	۱۸	۱	۱	۳۳	٥	۲٥	٥
٤	۷	−۱۰	۱۰۰	۲۳	−٥	۲٥	٥۰
٥	۳۰	۱۳	۱٦۹	۳٤	٦	۳٦	۷۸
٦	۱۲	−٥	۲٥	۳۲	٤	۱٦	−۲۰
۷	۱۸	۱	۱	۲٤	−٤	۱٦	−٤
۸	۲۳	٦	۳٦	۲۹	۱	۱	٦
۹	۱۹	۲	٤	۲٦	−۲	٤	−٤
۱۰	۱۲	−٥	۲٥	۲٦	−۲	٤	۱۰
کل	۱۷۰	۰	٦۳۰	۲۸۰	۰	۳٥۲	۳٦۷

$$n=10 \quad , \quad \bar{X}=\frac{170}{10}=17 \quad , \quad \bar{Y}=\frac{280}{10}=28$$

$$r=\frac{\sum_{i=۱}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^{۲}\cdot\sum(y_i-\bar{y})^{۲}}}=\frac{۳۶۷}{\sqrt{(۶۳۰)(۳۵۲)}}=۰/۷۷۹$$

مقدارعددی r نشاندهندهٔ میزان خطی بودن بین متغیرهای X و Y است، که مثبت بودن آن حاکی از آن است که یک رابطهٔ مثبت و مسقیم بین داده‌ها برقرار است؛ یعنی با افزایش نمرهٔ میان ترم، نمرهٔ پایان ترم نیز افزایش می‌یابد.

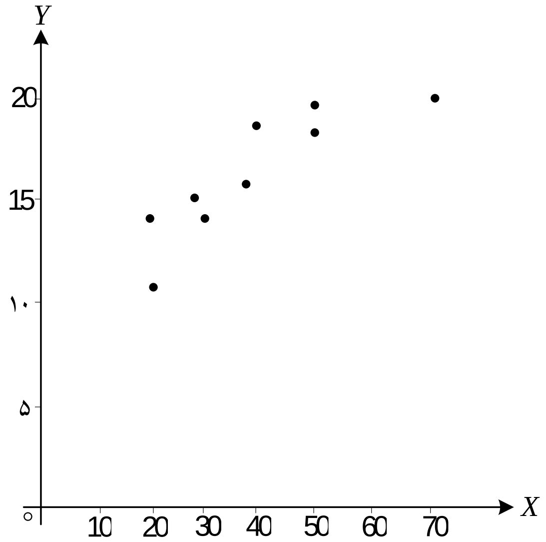
مثال ۵–۲: میزان ساعت بازآموزی و نمرهٔ کارایی ۱۰ نفر از کارکنان کتابخانه‌های کشور به صورت زیر داده شده است:

ساعت بازآموزی (X)	٤٠	٣٠	٥٠	٢٠	٤٠	٥٠	٣٠	٢٠	٧٠	٥٠
نمره کارایی (Y)	١٦	١٥	١٨	١٤	١٧	١٩	١٤	١٢	٢٠	

الف- نمودار پراکنش را رسم کنید.

ب- ضریب همبستگی نمونه‌ای پیرسون را محاسبهٔ کنید.

الف-



ب-

x	y	xy	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
٤٠	١٦	٦٤٠	٠	٠
٣٠	١٥	٤٥٠	١٠٠	١
٥٠	١٨	٩٠٠	١٠٠	٤
٢٠	١٤	٢٨٠	٤٠٠	٤
٤٠	١٧	٦٨٠	٠	١
٥٠	١٩	٩٥٠	١٠٠	٩
٣٠	١٤	٤٢٠	١٠٠	٤
٢٠	١٢	٢٤٠	٤٠٠	١٦
٧٠	٢٠	١٤٠٠	٩٠٠	١٦
٥٠	١٥	٧٥٠	١٠٠	١
٤٠٠	١٦٠	٦٧١٠	٢٢٠٠	٥٦

$$\bar{x} = ۴۰ \quad , \quad \bar{y} = ۱۶ \quad , \quad n = ۱۰$$

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum (x_i - \bar{x})^۲ . \sum (y_i - \bar{y})^۲}} = \frac{۶۷۱۰ - ۱۰(۴۰)(۱۶)}{\sqrt{۲۲۰۰ \times ۵۶}} = \frac{۳۱۰}{۳۵۰/۹۹} = ۰/۸۸۳$$

۵-٤ ضریب تعیین

در مطالعات همبستگی، ضریب تعیین $r^۲$ (مربع r) یک معیار برای اندازه‌گیری میزان ارتباط بین متغیرهای X و Y را فراهم می‌کند. برای مثال اخیر مقدار r برابر با ۰/۸۸۳ است، در نتیجه $r^2 = 0/779$ یعنی ۷۷/۹٪ از این نمرات با هم مشترک هستند؛ به تعبیری دیگر ۷۷/۹٪ از نمرات میان ترم با تغییرات نمرات در پایان ترم مشترک است.

۵-۵ همبستگی و علیت

متغیرهای زیادی وجود دارند که همبستگی بالایی دارند، اما وجود همبستگی به این معنی نیست که متغیر دلیل وجود دیگری است. اگر متغیرهای X و Y همبسته باشند، بنابراین حداقل سه امکان برای رابطه‌ای علّی بین متغیرها وجود دارد که عبارتند از:

- وجود متغیر X، وجود متغیر Y را سبب می‌شود.
- وجود متغیر Y، وجود متغیر X را سبب می‌شود.
- هیچ یک از متغیرهای X و Y سبب وجود یکدیگر نیست بلکه متغیر ثالثی دلیل وجود آن‌ها است.

برای نمونه ضریب همبستگی بین نمرات میان‌ترم و پایان‌ترم در مثال اخیر $r = 0/883$ به دست آمده که نشاندهندهٔ داشتن نمرهٔ بیشتر در میان ترم، داشتن نمرهٔ بیشتر در پایان‌ترم است. هیچ کدام از دو متغیر علت وجود دیگری نیست. بلکه گرفتن نمرهٔ بیشتر و داشتن معدل بالاتر مد نظر دانشجو است.

٥-٦ ضریب همبستگی رتبه‌ای اسپیرمن

استفاده از ضریب همبستگی پیرسن زمانی که متغیرهای همبسته به صورت فاصله‌ای و یا نسبی باشند، مناسب است. گاهی مقادیر مشاهده شده برای متغیرهای X و Y به صورت رتبه‌ای هستند. در این صورت از ضریب رتبه‌ای اسپیرمن برای نشان دادن رابطه‌ای بین مشاهدات استفاده می‌شود. ضریب همبستگی رتبه‌ای اسپیرمن از طریق فرمول زیر محاسبه می‌شود:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

که در آن

d_i: تفاضل جفتی رتبهٔ iام یک آزمودنی

n: تعداد زوج‌ها یا جفت های رتبه‌ای

r_s همانند r ضریب همبستگی پیرسون ممکن است علامت مثبت یا منفی داشته باشد که دامنهٔ تغییرات آن از ۱- تا ۱ است. برای محاسبهٔ ضریب همبستگی رتبه‌ای اسپیرمن یکی از متغیرها را مرتب و رتبه‌گذاری می‌کنیم و رتبه‌های مربوط به متغیر دوم را با رتبه‌های متغیر اول متناظر و d_iها را برای $i = 1,2,3,\dots,n$ محاسبهٔ می‌کنیم.

مثال ٥-۳: سابقهٔ خدمت و درجهٔ کارایی ده نفر از کارکنان کتابخانه‌ای به صورت زیر گزارش شده است:

سابقهٔ خدمت (X)	۲٤	۳۰	۱۲	۲٥	۲۹	۱۹	۱٦	۱۰	۱۱	۷
درجهٔ کارایی (Y)	٦٦	٥٥	۸٤	٦٦	٤٥	۸۱	۷۲	۹۷	۹۲	۷۰

ضریب همبستگی رتبه‌ای اسپیرمن را محاسبهٔ کنید.

سابقه خدمت (X) را به صورت زیر مرتب و به آن‌ها رتبه می‌دهیم:

۳۰	۲۹	۲۵	۲٤	۱۹	۱٦	۱۲	۱۱	۱۰	۷
۱۰	۹	۸	۷	٦	٥	٤	۳	۲	۱

درجهٔ کارایی (Y) را به صورت زیر مرتب و به آن‌ها رتبه می‌دهیم:

۹۷	۹۲	۸٤	۸۱	۷۲	۷۰	٦٦	٦٦	٥۱	٤٥
۱۰	۹	۸	۷	٦	٥	۳/٥	۳/٥	۲	۱

درجهٔ کارایی دو نفر از کارکنان ٦٦ است و رتبه‌های مربوط به آن‌ها ۳ و ٤ است؛ چون درجهٔ کارایی باید یکسان باشد، متوسط رتبه‌ها را برای آن‌ها در نظر می‌گیریم.

بنابراین:

$$\frac{۳+۴}{۲}=۳/۵$$

۹۷	۹۲	۸٤	۸۱	۷۲	۷۰	٦٦	٦٦	٥۱	٤٥
۱۰	۹	۸	۷	٦	٥	۳/٥	۳/٥	۲	۱

برای محاسبهٔ d_iها رتبهٔ درجهٔ کارایی افراد را متناظر با رتبهٔ سابقهٔ آن‌ها در نظر می‌گیریم. لذا

۱۰	۹	۸	۷	٦	٥	٤	۳	۲	۱	**رتبهٔ سابقهٔ خدمت:**
۲	۱	۳/٥	۳/٥	۷	٦	۸	۹	۱۰	٥	**رتبهٔ درجهٔ کارایی:**

d_i:	−٤	−۸	−٦	−٤	−۱	−۱	−۳/٥	−٤/٥	−۸	−۸
$d_i^۲$:	۱٦	٦٤	۳٦	۱٦	۱	۱	۱۲/۲٥	۲۰/۲٥	٦٤	٦٤

$$\sum_{i=1}^{n} d_i = -48 \quad , \quad \sum_{i=1}^{10} d_i^2 = 294/5$$

$$r_s = 1 - \frac{6\sum_{i=1}^{10} d_i^2}{10(100-1)} = 1 - \frac{6 \times 294/5}{990} = 1 - 1/7848 = -0/7848$$

۵–۷ آزمون فرض دربارهٔ ρ

در بررسی همبستگی نمونه‌ای دو متغیر ممکن است ضریب همبستگی نمونه‌ای در همسایگی ۱ یا ۱– باشد که دال بر همبستگی شدید بین دو متغیر باشد، اما دال نظر قطعی نیست و نیاز است از نظر آماری ضریب همبستگی مورد آزمون قرار گیرد که مخالف صفر بودن آن تأیید شود. لذا نیاز است فرض H_0 در مقابل H_1 مورد آزمون قرار گیرد.

$$H_0 : \rho = 0 \quad \text{در مقابل} \quad H_1 : \rho \neq 0$$

از آنجایی که ضریب همبستگی نمونه‌ای r یک براورد برای ρ است، نشان داده می‌شود که آمارهٔ آزمون برابر است با:

$$T_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

که دارای توزیع استودنت با $n-2$ درجهٔ آزادی است. برای تصمیم‌گیری دربارهٔ درستی یا عدم‌درستی فرض H_0 در مقایسهٔ مقدار T_0 با مقدار جدول استودنت اقدام می‌شود.

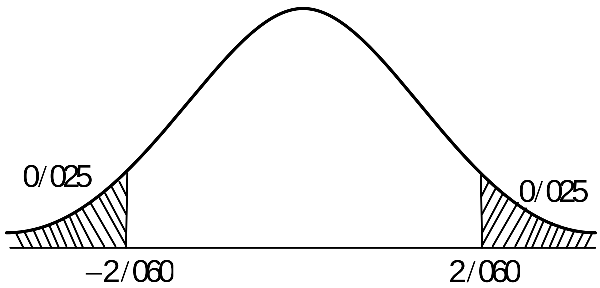
مثال ۵–۳: در بررسی رابطهٔ بین نمرهٔ ضریب هوشی (X) و نمرهٔ کارایی ۲۷ نفر از کارکنان، ضریب همبستگی نمونه‌ای پیرسون برابر با ۰/۸ گزارش شده است. فرض $H_0 : \rho = 0$ را در مقابل $H_1 : \rho \neq 0$ با اطمینان ۹۵٪ آزمون کنید.

حل:

$$n=27 \quad , \quad r=0/8 \quad , \quad \alpha=0/05 \quad , \quad \frac{\alpha}{2}=0/025$$

$$d.f=n-2=27-2=25$$

$$H_\circ : \rho=0 \qquad H_1 : \rho \neq 0$$



عدد جدول $t_{(0/025,25)}=2/060$

مقدار آماره برابر است با:

$$T_\circ=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\frac{0/8\sqrt{27-2}}{\sqrt{1-(0/8)^2}}=\frac{0/8\times 5}{0/6}=\frac{20}{3}=6/67$$

چون $T_\circ=6/67$ بیشتر از $t_{(0/025,25)}=2/060$ است، فرض $H_\circ$ رد می‌شود. نتیجه می‌شود که بین نمرهٔ ضریب همبستگی و نمرهٔ کارایی همبستگی وجود دارد.

مثال ۵-٤: برای بررسی میزان تبلیغات (X) و میزان فروش کتاب در ۷ هفته در یک فروشگاه بزرگ کتابفروشی، اطلاعات زیر به‌دست آمده است (ارقام به میلیون تومان):

۷	٦	۵	٤	۳	۲	۱	هفته
۲/۹	۲/۳	۱/۸	۲/۵	۲	۲/۲	۲	X
۹۳/٤	۸۷	۷۷/۸	۳۹/۹	۷٤/۳	۸۵/٦	۸۰/۲	Y

الف- مقدار ضریب همبستگی نمونه‌ای پیرسون را به‌دست آورید.

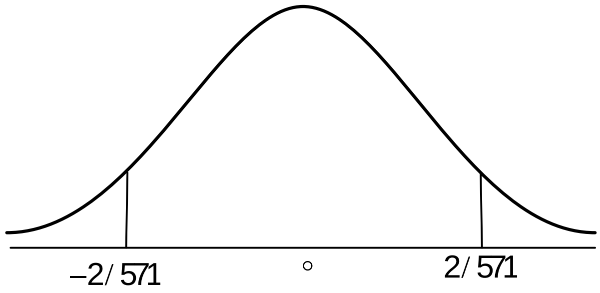
ب- فرض $H_\circ : \rho = 0$ را در مقابل $H_1 : \rho \neq 0$ با اطمینان ۹۵درصد آزمون کنید.

حل:

$$n = 7 \quad , \quad \alpha = 0/05 \quad , \quad \frac{\alpha}{2} = 0/025 \quad , \quad d.f = n - 2 = 5$$

$$t_{(0/025,5)} = 2/571$$

$$\begin{cases} H_\circ : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$



هفته i	X_i	Y_i	$\left(X_i - \bar{X}\right)^2$	$\left(Y_i - \bar{Y}\right)^2$	$\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$
۱	۲	۸۰/۲	۰/۰۵۹	۱۹/۳٦	۱/۰٦۹
۲	۲/۲	۸۵/٦	۰/۰۰۲	۱	-۰/۰٤۳
۳	۲	۷٤/۳	۰/۰۵۹	۱۰٦/۰۹	۲/۵۰۲
٤	۲/۵	۹۳/۹	۰/۰٦٦	۸٦/٤۹	۲/۳۹۱
۵	۱/۸	۷۷/۸	۰/۱۹٦	٤٦/۲٤	۳/۰۱۲
٦	۲/۳	۸۷	۰/۰۰۳	۵/۷٦	۰/۱۳۷
۷	۲/۹	۹۳/٤	۰/٤۳۲	۷۷/٤٤	۵/۷۸۲
کل	۱۵/۷	۵۹۲/۲	۰/۸۱۷	۳٤۲/۳۸	۱٤/۸۵۰

$$r = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum\left(X_i - \bar{X}\right)^2 \sum\left(Y_i - \bar{Y}\right)^2}} - \frac{14/850}{\sqrt{0/817 \times 342/38}} = \frac{14/850}{528/89} = 0/035$$

$$T_{\circ}=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\frac{0/35\sqrt{7-2}}{\sqrt{1-(0/035)^2}}=\frac{0/078}{0/999}=0/078$$

چون مقدار آمارهٔ محاسبه شده، بین دو مقدار جدول قرار دارد فرض $H_{\circ}$ پذیرفته می‌شود.

خلاصهٔ فصل پنجم

در این فصل، پس از بررسی دو متغیر، یکی وابسته و دیگری مستقل، بر اساس یک نمونهٔ n تایی از متغیرهای فوق اقدام به رسم نمودار پراکنش شد و با توجه به شکل حاصله برای ارائهٔ میزان همبستگی به‌صورت عددی، فرمول ضریب همبستگی پیرسون ارائه شد و در ادامه با نحوهٔ محاسبهٔ دو ضریب همبستگی پیرسون و رتبه‌ای اسپیرمن آشنا شدیم.

خودآزمایی تشریحی فصل پنجم

۱- جدول زیر رتبه پنج کتابخانه به همراه ارزش کتاب‌ها را نشان می‌دهد.

شمارهٔ کتابخانه	۱	۲	۳	٤	۵
رتبهٔ کتابخانه	۲	۳	۱	۵	٤
ارزش کتاب‌ها	۳	۲	۷	٦	۵

ضریب همبستگی رتبه‌ای اسپیرمن را محاسبهٔ کنید.

۲- الف) نشان دهید که

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=\sum_{i=1}^{n}\left(X_i-\bar{X}\right)Y_i$$

و

$$\sum\left(X_i-\bar{X}\right)^2=\sum X_i^2-n\bar{X}^2$$

ب) با توجه به بخش الف، نشان دهید فرمول ضریب همبستگی نمونه‌ای پیرسون برابر است با:

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)Y_i}{\sqrt{\left[\sum X_i^2-n\bar{X}^2\right]\left[\sum Y_i^2-n\bar{Y}^2\right]}}$$

۳- اگر $\sum\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=8$ و $\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=32$ و $\sum_{j=1}^{n}\left(Y_i-\bar{Y}\right)^2=42$، مقدار r را محاسبهٔ و آن را تحلیل کنید.

٤- برای بررسی رابطهٔ بین ساعت مطالعه و نمرهٔ دریافتی دانشجویان نمرات ۵ نفر از دانشجویان به صورت زیر گزارش شده است:

دانشجو	۱	۲	۳	٤	۵
ساعت مطالعه	۱۰	۱۳	۷	۱۲	۸
نمرهٔ دریافتی	۱٦	۱۸	۱۲	۱٤	۱۰

الف) مقدار ضریب همبستگی نمونه‌ای پیرسن را به دست آورید.

ب) ضریب تعیین نمونه‌ای را به دست آورده و آن را تحلیل کنید.

۵- برای بررسی میزان همبستگی X و Y اطلاعات زیر به دست آمده است:

X	۷	٤	۳	۱	۰	-۱
Y	۱	٥	٦	۷	۹	۱۰

ضریب همبستگی پیرسون را محاسبهٔ و آن را تحلیل کنید.

٦- رتبهٔ قبولی ۱۱ نفر از دانشجویان به همراه معدل ترم به صورت زیر گزارش شده است:

i	۱	۲	۳	٤	٥	٦	۷	۸	۹	۱۰	۱۱
رتبه X_i	٥	۱	۳	٤	۷	۹	۲	٦	۱۱	۸	۱۲
معدل Y_i	۱٥	۱۸	۱٤	۱٥	۱۷	۱۸	۱۰	۱٦	۱۰	۹	۱۱

پس از محاسبهٔ d_iها، ضریب همبستگی رتبه‌ای اسپیرمن را به دست آورید.

۷- اگر برای بررسی رابطهٔ بین صفت رتبه‌ای، تفاضل جفت‌ها به صورت زیر باشد:

i	۱	۲	۳	٤	٥	٦	۷	۸	۹
d_i	۳	-۱	-۲	۱	۲/٥	-۳/٥	۱/٥	٤	٤/٥

r_s را محاسبهٔ کنید.

۸- برای بررسی رابطهٔ بین درآمد و پس‌انداز ۸ نفر از کارکنان کتابخانه‌های کشور اطلاعات زیر به دست آمده است (ارقام به ده هزار تومان است):

درآمد (X)	۳۸	٤۰	٦۰	۳٦	٥٥	۷۲	٤۰	۳۰
پس‌انداز (Y)	۱۱	۹	۱۰	٤	۱۷	۳۰	۱٥	۱٤

ضریب همبستگی نمونه‌ای پیرسون را محاسبهٔ کنید.

۹- اگر برای بررسی رابطهٔ بین متغیرهای X و Y، رابطهٔ زیر برقرار باشد:

$$Y = ۲X + ۱$$

و مشاهدات زیر برای متغیر X باشد:

$X : -۱ \quad ۰ \quad ۱ \quad ۲ \quad ۳ \quad ۵$

نمودار پراکنش بین X و Y را رسم و مقدار ضریب همبستگی پیرسون را محاسبهٔ کنید.

۱۰- اگر $y = f(x) = x^2 - 1$ باشد، برای مقادیر $x = 1,2,3,4$، مقادیر Y را به‌دست آورید و ضرایب همبستگی نمونه‌ای پیرسون و اسپیرمن را محاسبهٔ و مقایسه کنید.

۱۱- برای آزمون فرض $H_۰ : \rho = ۰$ در مقابل $H_۱ : \rho \neq ۰$ اگر ضریب همبستگی یک نمونه ۲۷تایی ۰/۸ باشد، فرض $H_۰$ را با اطمینان ۹۵ درصد آزمون کنید.

فصل ششم

رگرسیون و اصول پیش‌بینی

هدف کلی

آشنایی با مبانی رگرسیون و اصول پیش‌بینی.

هدف‌های یادگیری

پس از مطالعهٔ این فصل، شما باید بتوانید:

۱. با انواع نمودار پراکنش بین مشاهدات دو متغیر آشنا شوید.

۲. با خط رگرسیون آشنا شوید.

۳. پارامترهای خط رگرسیون برازش شده را براورد کنید.

٤. واریانس خطا را براورد کنید.

۵. با توجه به خط برازش شده، متغیر پاسخ یا وابسته را پیش‌بینی کنید.

مقدمه

در فصل‌های قبلی، بیشتر تجزیه و تحلیل‌های آماری روی یک صفت از جامعه یا متغیر تصادفی متمرکز بود و در این فصل قصد داریم به صورت همزمان دو صفت از جامعه یا دو متغیر تصادفی را مورد بررسی قرار دهیم. این بررسی به پیش‌بینی مقادیر یکی از متغیرها از روی مقادیر متغیر دیگر که به مسئله برگشت یا رگرسیون معروف است، برمی‌گردد.

به عنوان مثال، فرض کنید بخواهیم تأثیر میزان مصرف شیر را در افزایش قد به دست آوریم و یا بخواهیم میزان وزن فرزند را از روی وزن پدرش پیش‌بینی کنیم. ملاحظه می‌کنید که در این گونه مسائل دو متغیر تصادفی مورد مطالعه به نوعی به یکدیگر وابسته هستند. به عبارت دقیق‌تر یک متغیر تصادفی مثل X را مستقل و متغیر تصادفی Y را وابسته به آن در نظر می‌گیریم و یا برعکس Y را مستقل و X را وابسته به آن؛ آشکار است که انتخاب هر یک از دو حالت به نوع مسئله هم بستگی دارد.

در مسائل رگرسیون برای یافتن رابطهٔ بین متغیر تصادفی مستقل X و متغیر وابستهٔ Y، ابتدا یک نمونهٔ nتایی از متغیر تصادفی X جمع‌آوری می‌کنیم که نتایج آن بصورت $x_1, x_2, \ldots, x_n$ هستند. سپس مقادیر متناظر با هر یک از نمونه‌های به دست آمده را که همان مقادیر معادل متغیر تصادفی وابسته Y هستند، به دست می‌آوریم. به این ترتیب برای (x_i)ها مقادیر متناظر $y_1, y_2, \ldots, y_n$ به دست می‌آید. که می‌توان نتیجه را به صورت زوج‌ها به شکل زیر نوشت:

x	x_1	x_2	$\ldots$	x_n
y	y_1	y_2	$\ldots$	y_n

یکی از روش‌های سادهٔ بررسی رابطهٔ بین دو متغیر، استفاده از نمودار پراکنش است که در ادامه آورده می‌شود.

۶–۱ استفاده از نمودار پراکندگی برای نمایش الگوی ارتباط بین داده‌ها

ضریب همبستگی و تحلیل رگرسیونی، تکنیک‌هایی هستند که به طور وسیع در بخش‌های مختلف علوم رفتاری برای یافتن رابطه‌ای بین متغیرها و یافتن متغیری از یک متغیر دیگر مورد استفاده قرار می‌گیرد. برای مثال اگر متغیرهای X و Y به ترتیب نمرات میان‌ترم و پایان‌ترم دانشجو در یک ترم باشند، تمایل محقق برای پیش‌بینی متغیر Y از مقدار متغیر X است. معمولاً متغیر X را به عنوان متغیر پیش‌بینی‌کننده و متغیر Y را به عنوان متغیر پیش‌بینی‌شونده به کار می‌برند. نمرات زوج شده در جدول یک توزیع دو متغیره از نمرات، به صورت زیر ارائه شده است:

جدول ۶–۱

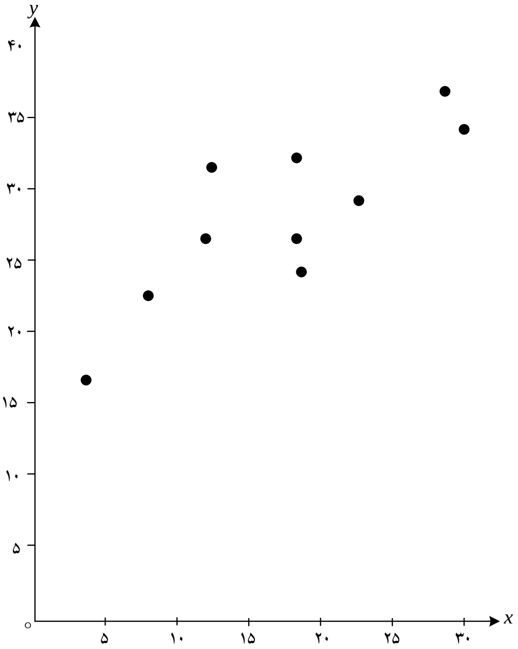
دانشجو	۱	۲	۳	٤	٥	٦	۷	۸	۹	۱۰
نمرهٔ میان‌ترم X	٤	۲۷	۱۸	۷	۳۰	۱۲	۱۸	۲۳	۱۹	۱۲
نمرهٔ پایان‌ترم Y	۱٦	۳۷	۳۳	۲۳	۳٤	۳۲	۲٤	۲۹	۲٦	۲٦

سؤالاتی که مطرح هستند عبارتند از:

- آیا نمرات میان‌ترم و پایان‌ترم به طور مثبت با هم مرتبطند؟
- آیا نمرات میان‌ترم و پایان‌ترم به طور منفی با هم مرتبطند؟
- آیا وابسته نبودن نمرات میان‌ترم و پایان‌ترم امکان‌پذیر است؟

برای اطلاع از ارتباط نمرات میان‌ترم و پایان‌ترم، نمودار پراکندگی رسم می‌شود. در نمودار پراکندگی متغیری که X نامگذاری می‌شود روی محور افقی

و متغیر Y روی محور عمودی نشان داده می‌شود. نمودار پـراکنش جـدول ٦-۱ به‌صورت زیر به‌دست می‌آید:



نمودار ٦-۱: پراکندگی نمرات امتحانات میان ترم و پایان ترم

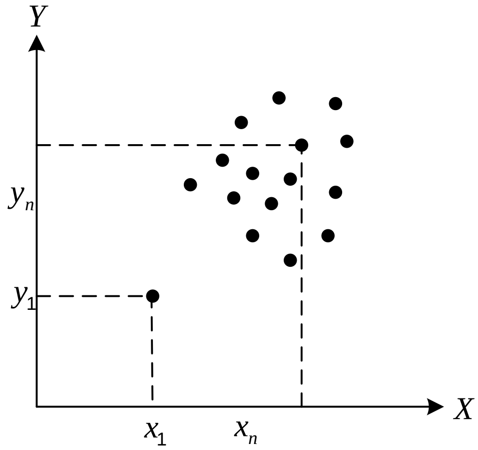
نمودار پراکندگی، ارتباط مثبت بین نمرات میان‌ترم و پایان‌ترم را نشان می‌دهد؛ یعنی اینکه اگر نمرهٔ میان‌ترم دانشجویی بیشتر است، نمرهٔ پایان‌ترم ایشان نیز بیشتر خواهد بود و بالعکس. اما در حالت کلی نمودار پراکنش چه وضعیت‌هایی می‌تواند داشته باشد و در هر حالت چگونه تفسیر می‌شود؟

اکنون فرض کنید که مقادیر x_iها میزان طول قد افراد یک جامعه و مقادیر y_iها میزان مصرف شیر هر یک از نمونه‌ها باشند. به این ترتیب زوج مرتب (۲، ۱۸۰) بیانگر این است که در نمونه‌گیری یکی از افراد جامعه دارای طول قد ۱۸۰ سانتی متر بوده و وی روزانه دو لیوان شیر مصرف کرده است.

پس از به دست آوردن زوج مرتب (x_i, y_i) ملاحظه می‌کنید که هر زوج مرتب می‌تواند معادل یک نقطه در صفحه باشد. با رسم نقاط مورد نظر در صفحه، یک تصویر کلی از رابطهٔ X و Y به دست می‌آوریم.

به شکل‌های زیر که برای سه نمونهٔ جداگانه است، توجه کنید:

۱. در این نمودار ملاحظه می‌کنید که زوج مرتب (x_i, y_i) به صورت کاملاً پراکنده توزیع شده‌اند و به این ترتیب نتیجه می‌گیریم که رابطه‌ای بین متغیرهای تصادفی X و Y وجود ندارد.



نمودار ٦-٢: بین X و Y رابطه‌ای وجود ندارد

۲. اگر داده‌ها به طور کلی حول یک خط توزیع شده باشند، بین مقادیر x_i و y_i یک رابطهٔ خطی برقرار است، حال آنکه دو حالت ممکن است به وجود آید:

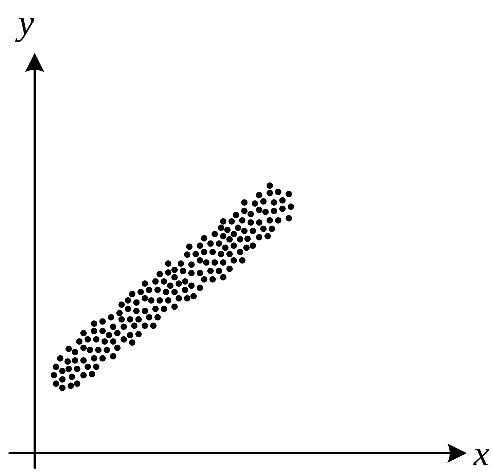
الف) شیب نمودار مثبت باشد. در این حالت با افزایش یک متغیر، متغیر دیگر هم افزایش می‌یابد. به عنوان مثال با افزایش سطح سواد، استفاده از کتابخانه بیشتر می‌شود.

ب) شیب نمودار منفی است. در این حالت با افزایش یک متغیر، مقدار متغیر دیگر کاهش می‌یابد. به عنوان مثال با افزایش استفاده از کتابخانه‌های دیجیتال استفاده از منابع کاغذی کاهش می‌یابد.

در این نمونه یک رابطهٔ خطی بین مقادیر x_i و y_i برقرار است و می‌توان نوشت:

$$y_i = ax_i + b$$

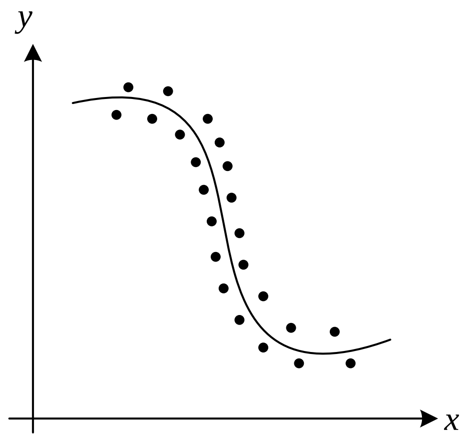
حال آنکه در صورتی که شیب، مثبت باشد، داریم: $\alpha > 0$ و درصورتی که شیب، منفی باشد، همواره $\alpha < 0$ خواهد بود.



نمودار ٦-٣: رابطهٔ خطی بین X و Y وجود دارد.

با توجه به شکل ٦-٣، خط مفروض نشان می‌دهد که رابطهٔ خطی مثبت است.

۱- ممکن است که داده‌ها حول یک منحنی پراکنده شده باشند و بین مقادیر x_i و y_i یک رابطهٔ غیرخطی برقرار باشد. در این نمودار X و Y به یکدیگر وابسته هستند که این وابستگی از نوع خطی نیست.



نمودار ٦-٤: رابطهٔ خطی بین X و Y وجود ندارد.

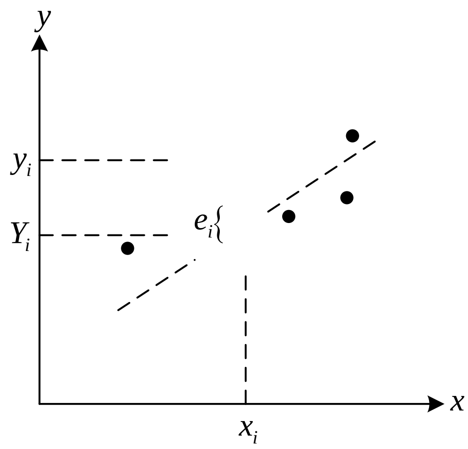
با توجه به دو نمودار ٦-٢ و ٦-٣ این طور به نظر می‌رسد که در حالت کلی دو متغیر تصادفی X و Y اگر از یکدیگر مستقل نباشند یا به صورت خطی و یا به صورت غیرخطی به یکدیگر وابسته خواهند بود.

در مسائل رگرسیون، پیش‌بینی متغیر Y از روی X و یا بالعکس از اهمیت ویژه‌ای برخوردار است. بنابراین نیازمند روشی هستیم که بتوان در صورت نیاز با ثابت در نظر گرفتن یکی از مقادیر X یا Y، مقدار دیگری را به دست بیاوریم.

٦-٢ رگرسیون خطی

اگر بین دو متغیر X و Y یک رابطهٔ خطی وجود داشته باشد، می‌توان یک خط را طوری رسم کرد که نقاط (x_i, y_i) کم‌ترین فاصله را با خط مورد نظر داشته باشد. به این عمل، برازش منحنی می‌گوییم. همان طور که در نمودار ٦-٤ مشاهده می‌کنید منحنی طوری رسم شده که داده در اطراف آن کم‌ترین فاصلهٔ ممکن را از خط دارد. اگر منحنی برازش داده شده به صورت خط باشد، به این عمل، برازش خطی می‌گویند و خط برازش داده شده هم دارای معادله‌ای به

صورت $Y = \alpha X + \beta$ خواهد بود که در آن پارامترهای α و β می‌بایستی طوری محاسبهٔ شوند که مجموع فاصلهٔ نقاط (x_i, y_i) از خط $Y = \alpha X + \beta$ حداقل شود. در این حالت به خط $Y = \alpha X + \beta$ معادلهٔ خط رگرسیونی Y می‌گویند. برای حداقل کردن فاصلهٔ نقاط (x_i, y_i) از خط رگرسیون مقدار خطای e_i را مطابق نمودار زیر به دست می‌آوریم:



نمودار ٦-۵: مقدار خطای e_i

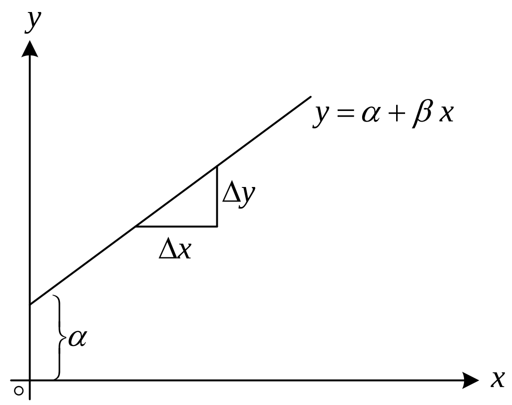
با توجه به نمودار، نماد Y_i نشاندهندهٔ مقدار پیش‌بینی شده توسط خط رگرسیون است. که با مقدار واقعی y_i به اندازهٔ $e_i = |y_i - Y_i|$ فاصله دارد که این فاصله همان خطای پیش‌بینی است. و همچنین باید مقادیر α و β را نیز براورد کرد.

٦-۳ رگرسیون دو متغیره

در رگرسیون دو متغیره، متغیر وابستهٔ Y به وسیله تنها متغیر مستقل X پیش‌بینی می‌شود. براورد خط رگرسیون بین متغیرهای X و Y امکان‌پذیر نخواهد بود،

مگر ابتدا فرض شود بین دو متغیر X و Y یک رابطهٔ خطی به فرم زیر وجود داشته باشد:

$$Y = \alpha X + \beta$$



نمودار ٦-٦: معادله خط رگرسیونی

که در آن

متغیر Y، متغیر وابسته است.

متغیر X، متغیر مستقل است.

پارامتر α عرض از مبدأ است.

پارامتر β شیب خط است که برای دو نقطه می‌توان نوشت:

$$\beta = \frac{\Delta x}{\Delta y} = \frac{x_2 - x_1}{y_2 - y_1}$$

مثال ٦-١: فرض کنید درآمد و هزینهٔ کتابخانه‌ها به وسیلهٔ خط رگرسیون زیر داده شده باشد (ارقام به میلیون تومان):

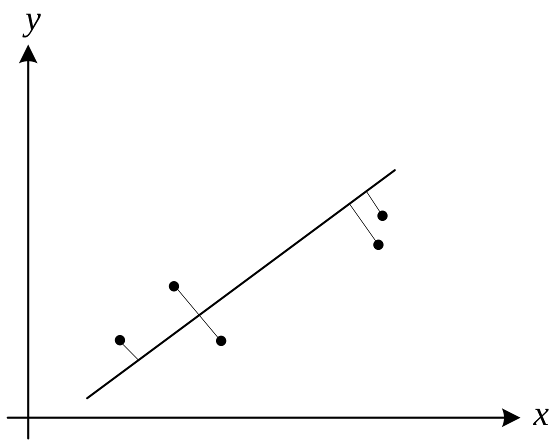
$$Y = 0/3 + 0/8X$$

که متغیر X درآمد و متغیر Y هزینه است.

به این ترتیب، کتابخانه‌ای که ۲ میلیون تومان درآمد دارد مبلغ ۱/۹ میلیون تومان هزینه می‌کند.

$$Y = 0/3 + 0/8X = 0/3 + 0/8(2) = 0/3 + 1/6 = 1/9$$

اما برازش خط رگرسیون با استفاده از یک نمونه همراه با خطا خواهد بود و می‌دانیم که وقتی متغیر وابسته Y به وسیله متغیر مستقل X پیش‌بینی می‌شود که نمودار پراکنش وجود خط رگرسیون را تأیید کنند. اگر برای ۵ زوج $(x_۱, y_۱), (x_۲, y_۲), \ldots, (x_۵, y_۵)$ نمودار پراکنش به همراه خط رگرسیون برازش به صورت زیر باشد:



نمودار ٦-۷

می‌توان گفت خط طوری برازش داده شده است که کم‌ترین فاصله را با نقاط مشاهده شده داشته باشد. در این صورت معادلهٔ خط برازش به صورت زیر خواهد بود:

$$y_i = a + bx_i + e_i$$

که e_i همان جملهٔ خطای پیش‌بینی است و فرض می‌شود که دارای توزیع نرمال با میانگین صفر و واریانس σ_e^2 است.

در آمار وقتی خط برازش شده، بهترین برازش را دارد که مجموع مربعات خطا یا $\sum_{i=1}^{n} e_i^2$ می‌نیمم باشد. با استفاده از آن می‌توان a و b را که براوردی برای α و β هستند از روابط زیر به دست آورد.

$$\hat{\beta} = b = \frac{\sum x_i y_i - n\,\bar{x}\,\bar{y}}{\sum (x_i - \bar{x})^2}$$

$$\hat{\alpha} = a = \bar{y} - b\bar{x}$$

بنابراین خط برازش شده نمونه‌ای است برابر با:

$$\hat{y}_i = a + bx_i$$

اختلاف این معادله $y_i = a + bx_i + e_i$ و $\hat{y}_i = a + bx_i$، e_i است که به صورت زیر است:

$$e_i = y_i - \hat{y}_i$$

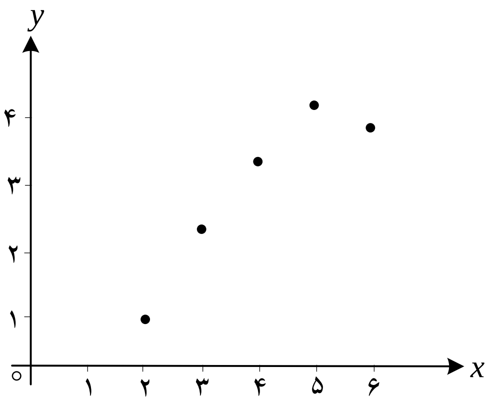
مثال ۶-۲: درآمد و هزینهٔ ۵ تا از کتابخانه‌های کشور به صورت زیر گزارش شده است:

کتابخانه	۱	۲	۳	٤	۵
درآمد (x)	٤	۵	۳	٦	۲
هزینه (y)	۳/۵	٤/۲	۲/۳	٤	۱

الف- نمودار پراکنش را رسم کنید.

ب- اگر $\hat{y}_i = a + bx_i$ خط برازش شده باشد، a و b را به دست آورید.

الف–



ب– برای محاسبهٔ a و b از جدول زیر استفاده می‌شود:

i	x_i	y_i	$x_i y_i$	$(x_i - \bar{x})^2$
۱	٤	۳/٥	۱٤	۰
۲	٥	٤/۲	۲۱	۱
۳	۳	۲/۳	٦/۹	۱
٤	٦	٤	۲٤	٤
٥	۲	۱	۲	٤
کل	۲۰	۱٥	٦۷/۹	۱۰

$$b = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum (x_i - \bar{x})^۲} = \frac{۶۷/۹ - ۵(۴)(۳)}{۱۰} = ۰/۷۹$$

$$a = \bar{y} - b\bar{x} = ۳ - ۰/۷۹(۴) = -۰/۱۶$$

$$\hat{y}_i = a + bx_i = -0/16 + 0/79x_i \quad , \quad i = 1,2,...,n$$

٦–٤ براورد واریانس خطا

در بخش قبل فرض شده که خطاها یعنی e_iها دارای توزیع نرمال با میانگین صفر و واریانس $\sigma_e^۲$ است. بنابراین با توجه به ویژگی توزیع نرمال براورد $\sigma_e^۲$ برابر است با:

$$\sigma_e^2=\frac{1}{n-2}\sum_{i=1}^{n}e_i^2=\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\frac{1}{n-2}\sum_{i=1}^{n}e_i^2$$

مقدار $\sigma_e^۲$ برای مثال اخیر برابر است با:

i	y_i	$\hat{y}_i=-۰/۱۶+۰/۷۹x_i$	e_i	$e_i^۲$
۱	۳/۵	۳	۰/۵	۰/۲۵
۲	٤/۲	۳/۷۹	۰/٤۱	۰/۱٦۸۱
۳	۲/۳	۲/۲۱	۰/۰۹	۰/۰۰۸۱
٤	٤	٤/۵۸	$-$۰/۵۸	۰/۳۳٦٤
۵	۱	۱/٤۲	$-$۰/٤۲	۰/۱۷٦٤
کل	۱۵	۱۵	۰	۰/۹۳۹

$$\sigma_e^۲=\frac{۱}{n-۲}\sum e_i^۲=\frac{۱}{۳}\left(۰/۹۳۹\right)=۰/۳۱۳$$

٦–۵ پیش‌بینی

هنگامی که معادلهٔ رگرسیون شکل گرفت، می‌توان یک متغیر (معمولاً Y) را بر اساس متغیر دیگر (معمولاً X) پیش‌بینی کرد. برای مثال فرض کنید که در نظر داریم کتابخانهٔ جدیدی راه‌اندازی کنیم و می‌خواهیم بدانیم که چه تعداد کامپیوتر برای ۲۵ کارمند کتابخانه نیاز است. این کار می‌تواند به سادگی از طریق قرار دادن مقدار متغیر X در معادله و محاسبهٔ مقدار Y انجام گیرد، با فرض $X=25$ در معادلهٔ رگرسیون $Y=4/25+0/63X$، مقدار $Y=20$ به دست می‌آید؛ یعنی برای راه‌اندازی کتابخانه جدید نیاز به ۲۰ کامپیوتر شخصی است.

مثال ٦–۳: در محاسبهٔ رابطهٔ خطی بین نمرهٔ امتحان داوطلب X و نمرهٔ رضایتمندی مراجعان Y، معادلهٔ رگرسیونی به صورت $Y=23/8+0/79X$ به

دست آمده است. اگر داوطلبی نمرهٔ امتحانی ۸۰ را کسب کرده مقدار رضایتمندی را چقدر پیش‌بینی می‌کنید؟

برای $x = 80$ با توجه به خط برازش خواهیم داشت:

$$Y = 23/8 + 0/79(80) = 87$$

پس در صورتی که نمرهٔ امتحانی ۸۰ توسط داوطلب کسب شود، میزان رضایتمندی وی دارای نمرهٔ ۸۷ خواهد بود.

مثال ٦–٤: برای بررسی رابطهٔ بین قیمت کتاب (X) و تعداد صفحات (Y) اطلاعات زیر به دست آمده است:

قیمت (هزار تومان)	۱۰۰	۷۰	۱۲۰	۱۱۰	۲۰۰
تعداد صفحه	۳۰۰	۲۰۰	٤۰۰	۳٥۰	٥٥۰

الف– معادلهٔ خط برازش را به دست آورید.

ب– تعداد صفحه را برای کتابی که قیمت آن ۱٥۰ تومان است، پیش‌بینی کنید.

حل:

x	y	xy	$(x-\bar{x})^۲$
۱۰۰	۳۰۰	۳۰۰۰۰	٤۰۰
۷۰	۲۰۰	۱٤۰۰۰	۲٥۰۰
۱۲۰	٤۰۰	٤۸۰۰۰	۰
۱۱۰	۳٥۰	۳۸٥۰۰	۱۰۰
۲۰۰	٥٥۰	۱۱۰۰۰۰	٦٤۰۰
٦۰۰	۱۸۰۰	۲٤۰٥۰۰	۹٤۰۰

با توجه به جدول فوق

$$\bar{x} = ۱۲۰ \quad , \quad \bar{y} = ۳۶۰$$

$$b=\frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum(x_i-\bar{x})^۲}=\frac{۲۴۰۵۰۰-۲۱۶۰۰۰}{۹۴۰۰}=\frac{۲۴۵۰۰}{۹۴۰۰}=۲/۶۱$$

$$a=\bar{y}-b\bar{x}=360-(2/61)(120)=46/8$$

$$\hat{y}_i=a+bx_i=۴۶/۸+۲/۶۱\ x_i$$

ب- برای $x_i=150$ داریم:

$$\hat{y}_i=46/8+2/61(150)=46/8+391/5=438/3$$

خلاصهٔ فصل ششم

در این فصل با توجه به میزان همبستگی متغیر وابسته و مستقل معادله خط رگرسیون در جامعه ارائه شد و از آنجایی که خط رگرسیون دارای پارامترهای عرض از مبدأ و شیب خط بوده با استفاده از فرمول‌های ارائه شده، نحوهٔ براورد آن‌ها با ارائهٔ مثال توضیح داده شده است. پس از برازش خط رگرسیون، پیش‌بینی که یکی از اهداف اصلی برازش است ارائه شد.

خودآزمایی تشریحی فصل ششم

۱- جدول زیر تعداد دانشجویان (به هزار) و میزان فروش فصلی غذای دانشگاه را نشان می‌دهد:

تعداد دانشجو (x)	۲	٦	۸	۸	۱۲	۱٦	۲۰	۲۰	۲۲	۲٦
میزان فروش (y)	۵۸	۱۰۵	۸۸	۱۱۸	۱۱۷	۱۳۷	۱۵۷	۱٦۹	۱٤۹	۲۰۲

نمودار پراکنش را رسم کنید.

۲- اگر متغیرهای X و Y دارای مقادیر مشاهده به صورت زیر باشد:

x_i	۱	۲	۳	٤	۵
y_i	۳	۷	۵	۱۱	۱٤

نمودار پراکنش را رسم کنید.

۳- اگر رابطهٔ خطی قیمت (X) و عمر (Y) رایانه‌ها به صورت زیر باشد:

$$Y = ۳ + ۰/۲X$$

الف) رایانه‌ای دارای قیمت $X = ۵$ است، مقدار Y را به‌دست آورید.

ب) برای $X = ۷$ مقدار Y چقدر است؟

۵- جدول زیر، میزان مطالعه و نمرهٔ آزمون ورودی ده نفر از کارکنان کتابخانه را نشان می‌دهد:

میزان مطالعه	٦۲	۵۳	٤٤	۵۰	۵٤	۳۹	٦٦	۵۵	۳٤	۳۹
نمرهٔ آزمون ورودی	۲۸۰۰	۲۸۰۰	۲۷۰۰	۳۵۰۰	۳۳۰۰	۲۰۰۰	٤۰۰۰	۳۰۰۰	۲۵۰۰	۳۰۰۰

الف) اگر $y_i = a + bx_i$ خط برازش شده باشد، a و b را به دست آورید.

ب) $\hat{\sigma}^۲ = \frac{۱}{n-۲}\sum_{i=۱}^{n} e_i^۲$ را محاسبهٔ کنید.

۶- برای بررسی و ارتباط بین نمرهٔ ضریب هوشی و نمرهٔ ریاضی پنج نفر از دانشجویان به صورت زیر گزارش شده است:

نمرهٔ ضریب هوشی	۹۵	۸۵	۱۰۰	۸۰	۹۰
نمرهٔ ریاضی	۱۵	۱٤	۱۸	۱۲	۱٦

الف) پس از محاسبهٔ a و b ،خط برازش شده $\hat{y}_i = a + bx_i$ را بنویسید.
ب) برای شخصی که نمرهٔ ضریب هوشی او ۱۰۱ است، میزان نمرهٔ ریاضی را پیش‌بینی کنید.
۷- نمرهٔ ضریب هوشی، نمرهٔ ورودی به دانشگاه و معدل کل ۵ نفر از دانشجویان که به تصادف انتخاب شده‌اند:

i	۱	۲	۳	٤	۵
نمرهٔ ضریب هوشی	۸۵	۹۰	۹۵	۱۰۵	۷۵
نمرهٔ ورودی	۱۳	۱۵	۱۷	۱۸	۱۲
معدل کل	۱٤/۵	۱۵	۱٦	۱۷	۱۲/۵

الف) معادلهٔ خط برازش بین ضریب هوشی (X) و نمرهٔ ورودی (Y) را به دست آورید.
ب) معادلهٔ خط برازش بین نمرهٔ ورودی و معدل کل را به دست آورید.

پیوست‌ها

پیوست ۱: آشنایی مقدماتی با نرم‌افزار SPSS

- معرفی منوها
- راهنمای گام به گام ورود داده‌ها

مقدمه

امروزه در بسیاری از موارد، خصوصاً در علم اطلاعات و دانش‌شناسی، استفاده از محاسبات آماری، یک بخش اصلی در فرایند انجام بسیاری از فعالیت‌ها و تحقیقات است. از این رو استفاده از نرم‌افزارهای آماری، به دلیل آنکه ساده‌ترین و سریع‌ترین ابزارهای انجام تجزیه و تحلیل‌های مذکور هستند، اجتناب‌ناپذیر است.

نظر به اهمیت انتخاب بهترین نرم‌افزار آماری برای انجام تجزیه و تحلیل‌های مربوط به پژوهش‌ها به ساده‌ترین شکل ممکن، در این قسمت، با توجه به نیاز دانشجویان رشتهٔ علم اطلاعات و دانش‌شناسی، دو قسمت شامل معرفی منوها و چگونگی ورود داده‌ها در نرم‌افزار آماری SPSS به صورت گام‌به‌گام توضیح داده شده است.

Spss نرم‌افزاری تخصصی مربوط به علم آمار است که شهرتی جهانی دارد. کلمهٔ SPSS مخفف Statistical package for social science (نرم‌افزار آماری برای علوم اجتماعی) است. ابتدا این نرم‌افزار برای کاربرد در دانشگاه استانفورد شیکاگو طراحی شد، ولی بعدها در پژوهش‌های بسیار زیادی در زمینه‌های موضوعی متعدد از جمله بازاریابی، علوم بهداشتی، مسائل مالی شرکت‌ها، تعلیم و تربیت و ... به کار گرفته شد و در سال ۱۹۹۰ نسخهٔ حرفه‌ای و در عین حال سادهٔ آن طراحی شد.

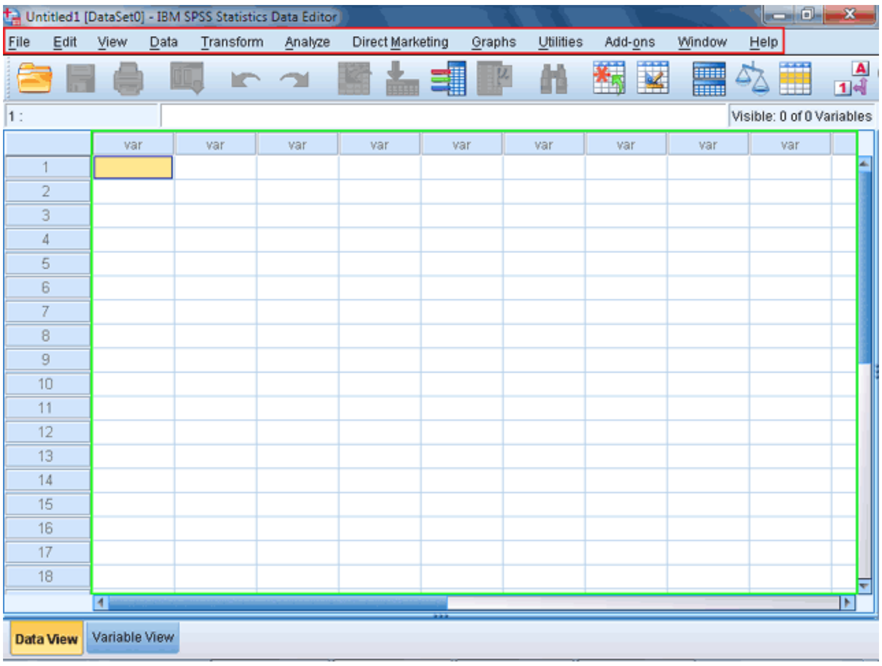
پنجرهٔ کاری در نرم‌افزار spss

پس از نصب و راه‌اندازی برنامه، اولین پنجرهٔ کاری را که در spss مشاهده می‌کنید، یک صفحه، کار برگ (Data Editor) است. در اینجا داده‌هایی را که می‌خواهید با SPSS تحلیل و یا محاسبه شوند، وارد می‌کنید. پس از ایجاد نمودار

یا تحلیل، نتایج کارتان در پنجرۀ دیگری به نام صفحۀ نمایش (Viewer) ظاهر می‌شود.

اکنون به بررسی محیط این نرم‌افزار می‌پردازیم. صفحه‌ای که در شکل (۱) مشاهده می‌کنید، صفحه، کاربرگ است که از آن برای ورود داده‌ها استفاده می‌شود.

قسمتی که با کادر قرمز مشخص شده است منوی اصلی برنامه است. در SPSS تمام فرمان‌ها و عملیات ممکن از طریق منوی اصلی برنامه قابل انجام است.

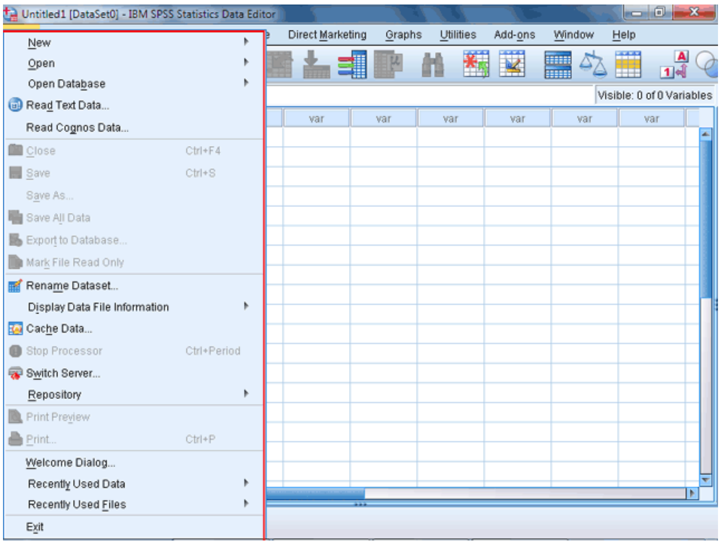


شکل (۱)

معرفی منوها

۱- **منوی File:** هدف از فرمان‌های منوی File کار با پرونده‌ها است. با استفاده از فرمان‌های موجود در این منو، می‌توانید پرونده‌های جدید ایجاد کنید، پرونده‌های

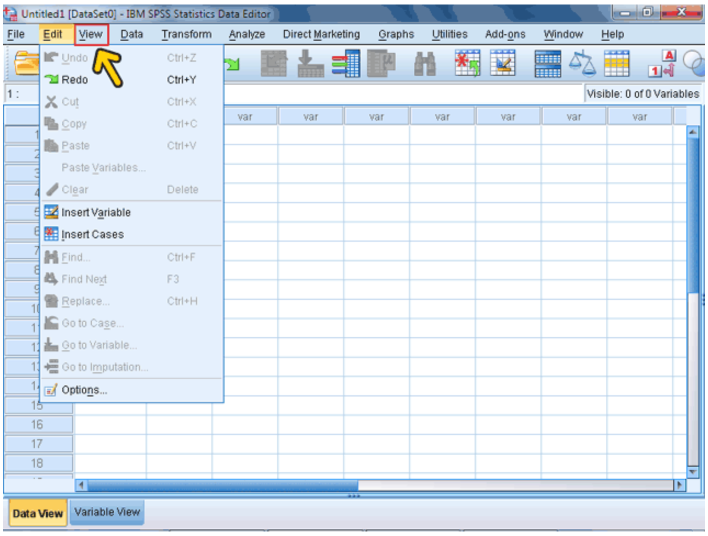
موجود را باز کنید، پرونده‌ها را با فرمت‌های مختلف ذخیره کنید، عملیات چاپ را انجام دهید و از SPSS خارج شوید. در این منو هنگام کار با بانک‌های اطلاعاتی بزرگ روی شبکه، برای کم کردن زمان خواندن اطلاعات، می‌توانید با امکان Data Cache یک کپی موقت از بانک اطلاعاتی فعال ایجاد کنید. خارج از منو کلیک کنید تا بسته شود.



شکل (۲)

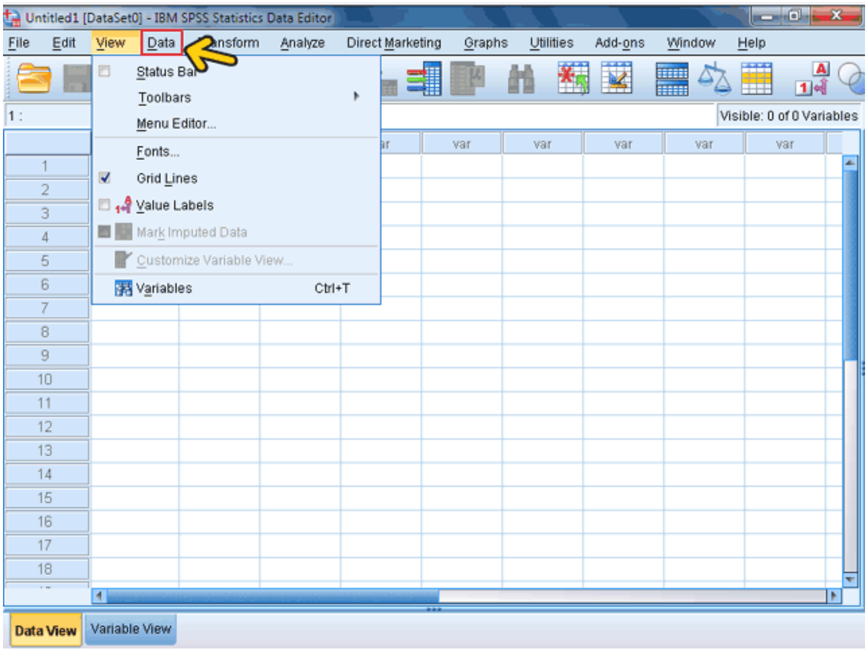
۲- **منوی Edit:** از گزینه‌های منوی Edit برای کپی کردن داده‌ها، جابه‌جایی در همان پرونده یا پروندهٔ دیگر استفاده می‌شود. علاوه بر این از فرمان‌های منوی Edit برای جستجوی داده‌ها یا متن و جایگزینی استفاده می‌شود.

روی منوی View کلیک کنید.



شکل (۳)

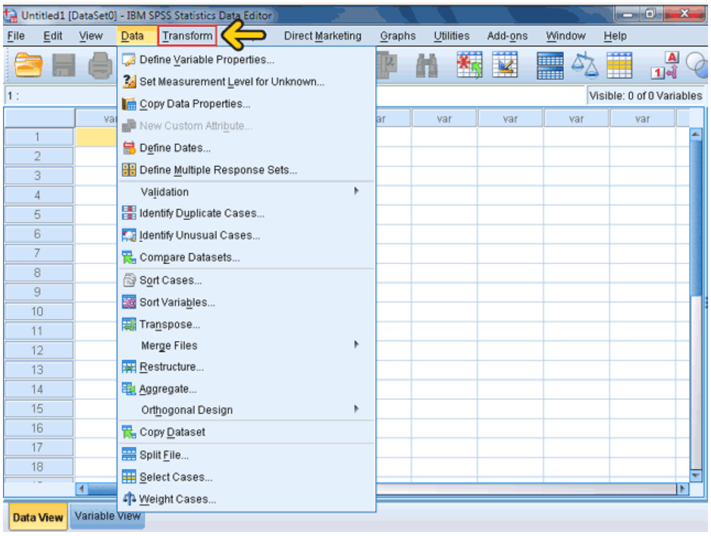
۳- منوی View: با فرمان‌های این منو، آرایش پنجره SPSS را تنظیم می‌کنید. برای نمایش یا پنهان‌سازی میله ابزار (Toolbars)، خط وضعیت (Statues bar)، خطوط زمینه در صفحه کاربرگ (Gridlines)، تغییر قلمها (Fonts) و عنوان مقادیر (Value Labels) از فرمان‌های این منو استفاده می‌کنیم. روی منوی Data کلیک کنید.



شکل (٤)

٤- منوی Data: این منو، دارای فرمان‌هایی برای کار با متغیرها است. این منو دارای فرمان‌هایی برای تعریف و مرتب‌سازی متغیرها، کار با الگوها، رفتن به رکورد خاص، ترکیب و جمع کردن پرونده‌ها و وزن‌گذاری رکوردها است.

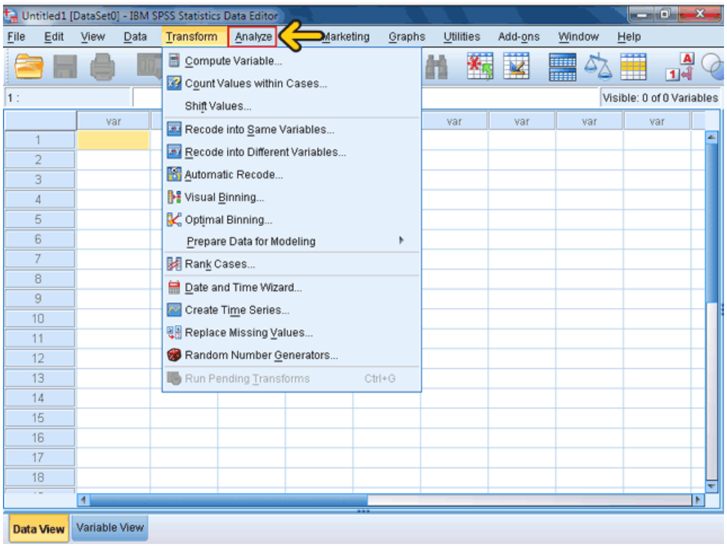
روی منوی Transform کلیک کنید.



شکل (۵)

۵- **منوی Transform:** از فرمان‌های منوی Transform برای محاسبهٔ مقادیر جدید، ایجاد یک سری مقادیر تصادفی، ضبط مقادیر، جایگزینی مقادیر غایب (Missing Value) و غیره استفاده می‌شود.

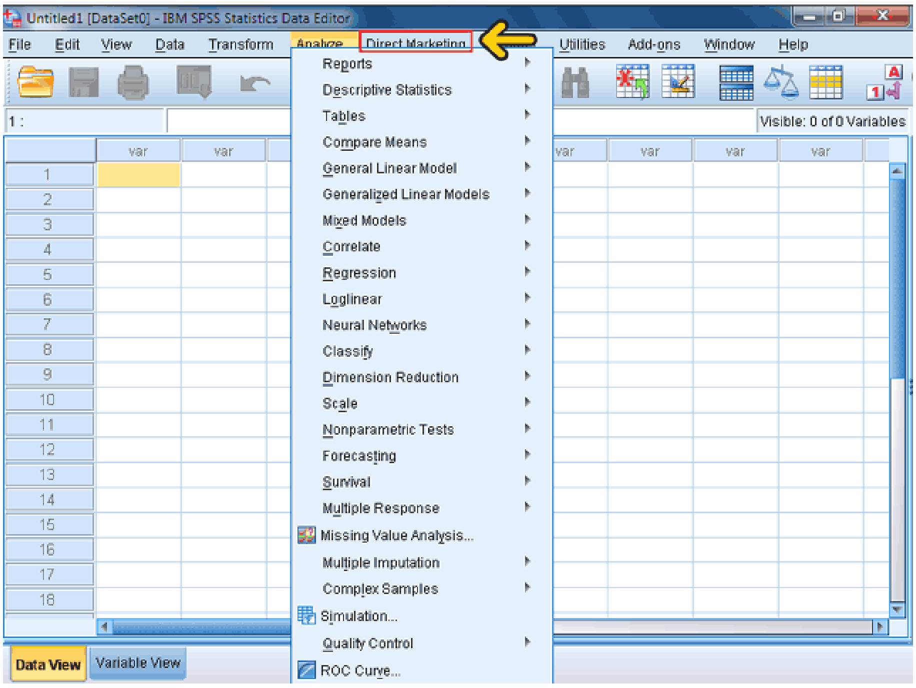
منوی Analyze را باز کنید.



شکل (٦)

٦- منوی Analyz: در این منو چندین امکان مختلف برای انجام بیشتر روش‌های تحلیل داده‌ها وجود دارد. محاسبهٔ یک میانگین و انحراف استاندارد تا تجزیهٔ سری‌های زمانی و رگرسیون چندمتغیره از طریق فرمان‌های این منو قابل اجرا است. برای مثال اگر بخواهید میانگین رتبه‌ها را که استاد شمارهٔ ٦ براساس فرم ارزیابی دریافت کرده با میانگین نمرات متعلق به استاد شمارهٔ ٤ مقایسه کنید. باید از گزینهٔ Compare Means استفاده کنید. در بخش‌های بعدی به بررسی کامل تمامی این تحلیل‌ها می‌پردازیم.

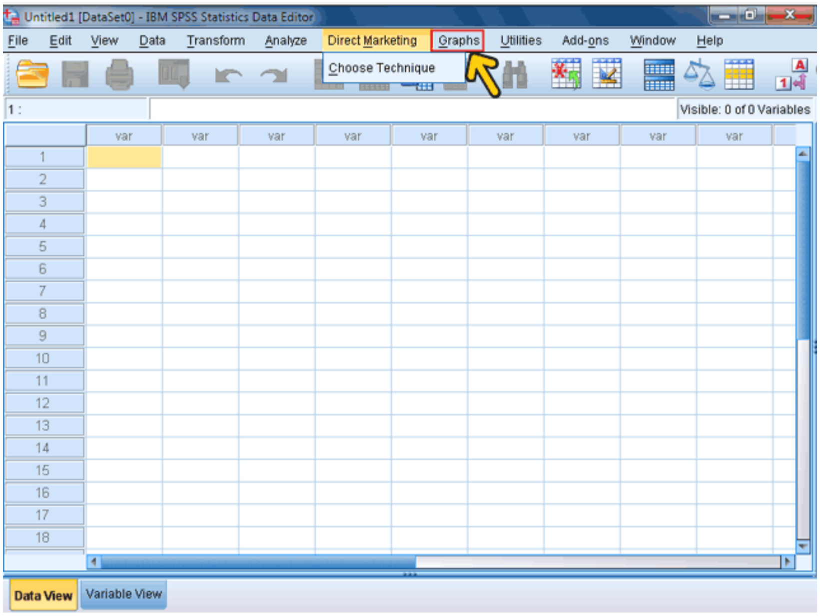
روی منوی Direct Marketing کلیک کنید.



شکل (۷)

۷- **منوی Direct Marketing:** گزینهٔ Direct Marketing یا بازاریابی مستقیم شامل مجموعه ابزارهایی است که برای بهبود نتایج حاصل از بازاریابی استفاده می‌شود. این کار از طریق شناسایی ویژگی‌های جمعیت‌شناختی، خرید کردن و سایر ویژگی‌های گروه‌های مختلف مصرف‌کننده و با هدف قرار دادن گروه‌های خاص به منظور به حداکثر رساندن میزان پاسخ مثبت انجام می‌شود.

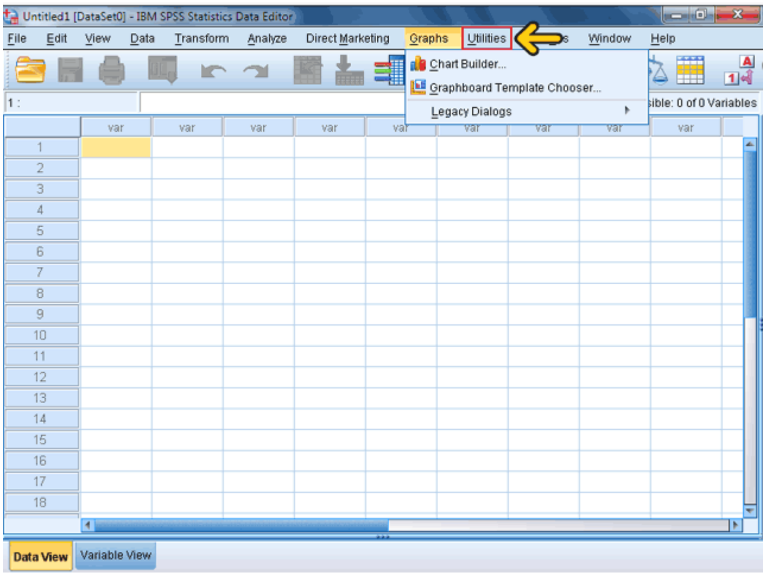
منوی Graphs را باز کنید.



شکل (۸)

۸- **منوی Graphs:** از این منو برای ساخت انواع نمودارها، ویرایش و زیباسازی آن‌ها، به اضافه هر کاری که باید بر روی نمودارها انجام شود استفاده می‌شود. نمودارهای قابل ساخت در SPSS میله‌ای، خطی، دایره‌ای و چندین نوع دیگر است که در بخش‌های بعدی به بررسی آن‌ها می‌پردازیم.

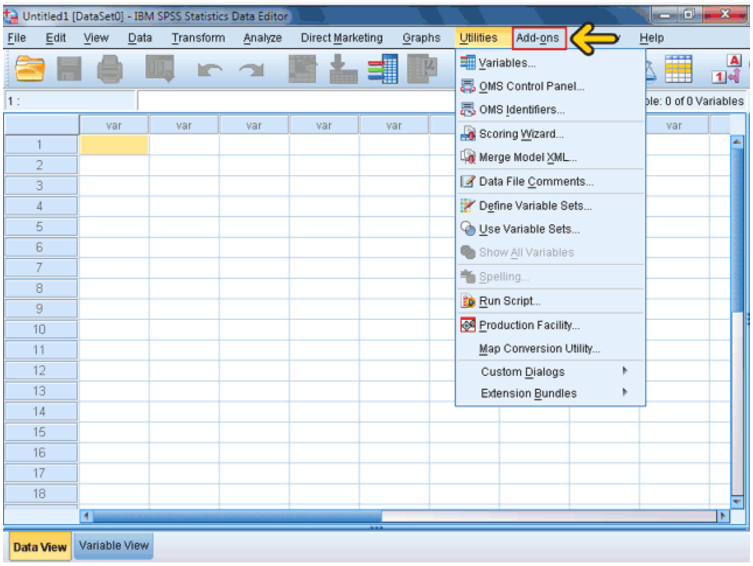
منوی Utilities را باز کنید.



شکل (۹)

۹- منوی Utilities: در این منو امکاناتی برای جستجوی اطلاعات دربارهٔ متغیرها و پرونده‌ها و همچنین تعریف و استفاده از سری متغیرها وجود دارد.

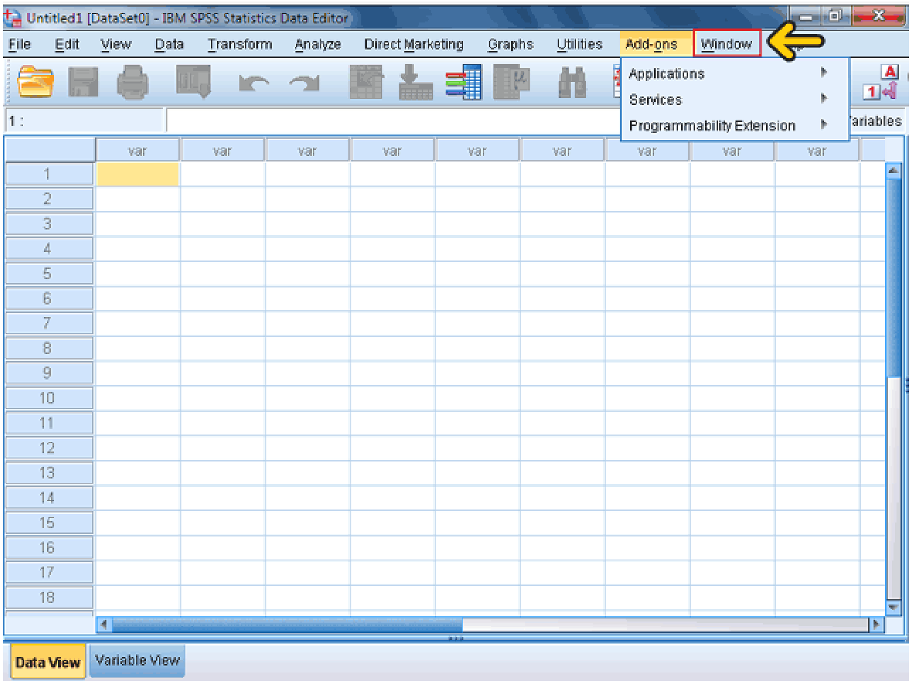
روی منوی Add-ons کلیک کنید.



شکل (۱۰)

۱۰- **منوی Add-ons:** در این منو امکاناتی برای جستجوی اطلاعات دربارهٔ متغیرها و پرونده‌ها و همچنین تعریف و استفاده از سری متغیرها وجود دارد.

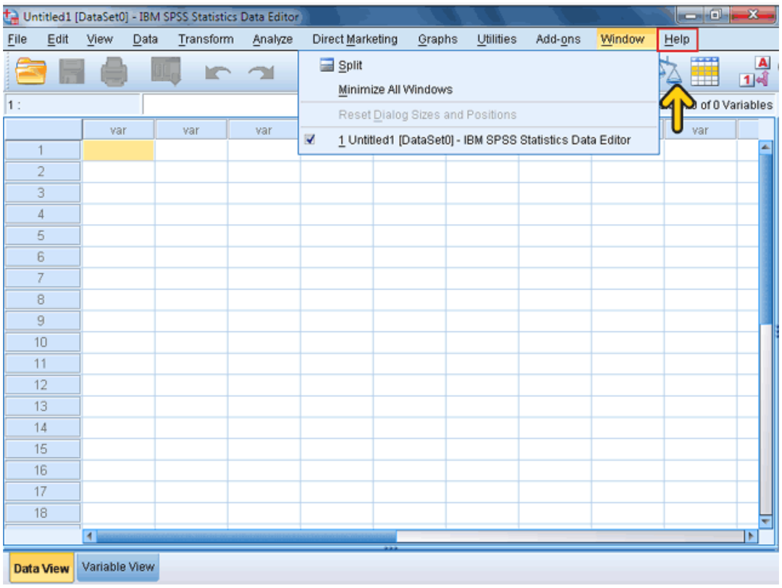
روی منوی Window کلیک کنید.



شکل (۱۱)

۱۱- منوی Window: فرمان‌های موجود در این منو برای انتخاب پنجره فعال و یا کوچک کردن صفحهٔ کاربرگ و نمایش استفاده می‌شود.

منوی Help را باز کنید.

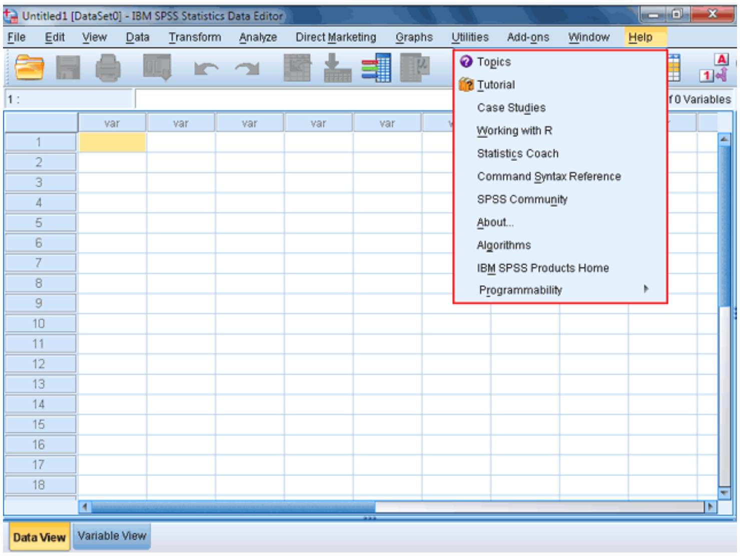


شکل (١٢)

١٢- منوی Help

از طریق این منو انواع راهنمایی‌ها برای کار با SPSS ارائه شده است.

در نقطه‌ای خارج از منو کلیک کنید تا این منو بسته شود.



شکل (١٣)

وارد کردن داده‌ها در SPSS

اولین گام برای کار با نرم‌افزار SPSS وارد کردن داده‌ها است. هنگامی که SPSS را اجرا می‌کنید صفحه کاربرگ بدون نام، باز شده و آمادهٔ وارد کردن متغیرها و مقادیر مربوط به آن‌ها است (شکل ١٤).

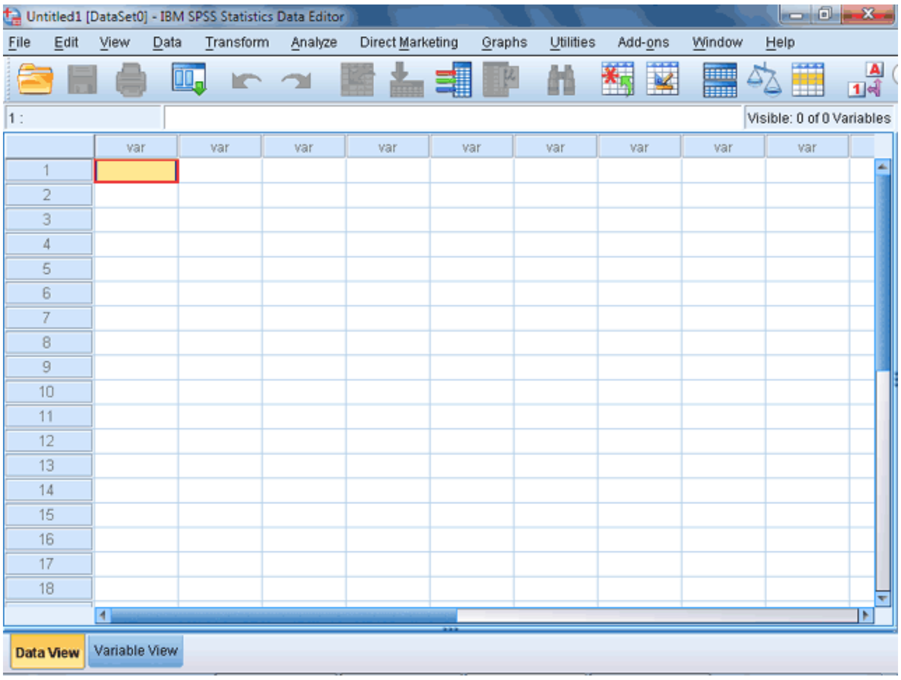
صفحهٔ کاربرگ از چند جزء مهم تشکیل شده است که در ادامه به بررسی تمامی جزئیات آن‌ها می‌پردازیم.

۱- هر سطر یا رکورد شامل یکسری اطلاعات مربوط به موضوع تحقیق یا تحلیل است. در SPSS سطر یا رکورد را Case می‌نامند.

۲- یکی دیگر از اجزای مهم در صفحهٔ کاربرگ، ستون است. هر ستون شامل یکسری اطلاعات یا مقادیر مربوط به یک متغیر است.

۳- سلول یکی دیگر از اجزای مهم صفحهٔ کاربرگ است. مکان ورود داده‌ها را سلول می‌گویند. هر سلول محل تقاطع یک سطر و یک ستون است.

سطرها و ستون‌ها و داده‌های وارد شده در سلول‌ها، یک پرونده را تشکیل می‌دهند. صفحهٔ کاربرگ در SPSS برای تعداد کم داده‌ها بسیار مفید است. در صورت زیاد بودن تعداد داده‌ها و لزوم انجام برخی کارها روی آن‌ها، مانند مرتب‌سازی، بهتر است ابتدا داده‌ها را در یک کاربرگ مانند Excel یا یک بانک اطلاعاتی مانند Access وارد کنید. سپس با آن داده‌ها در SPSS کار کنید.



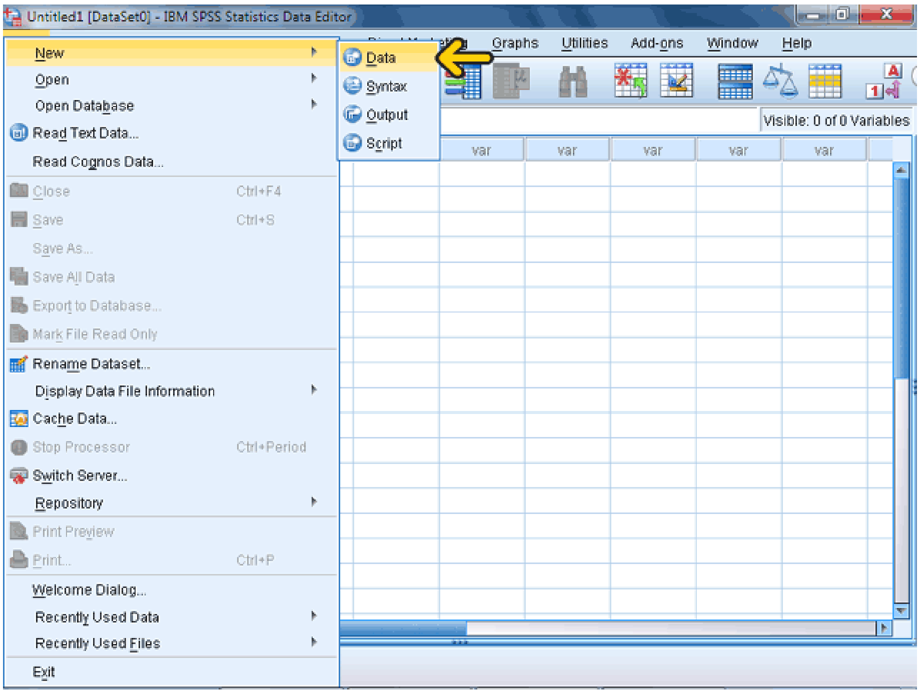
شکل (١٤)

مراحل ایجاد پرونده و ورود داده

۱- هر پروندهٔ داده حداقل باید دارای یک ردیف داده (Case) باشد. اولین مرحله برای ایجاد یک پروندهٔ داده، تعریف متغیرهای آن است. برای ایجاد یک پروندهٔ جدید منوی File را باز کنید.

منوی فرعی New را باز کنید.

روی گزینهٔ Data کلیک کنید.

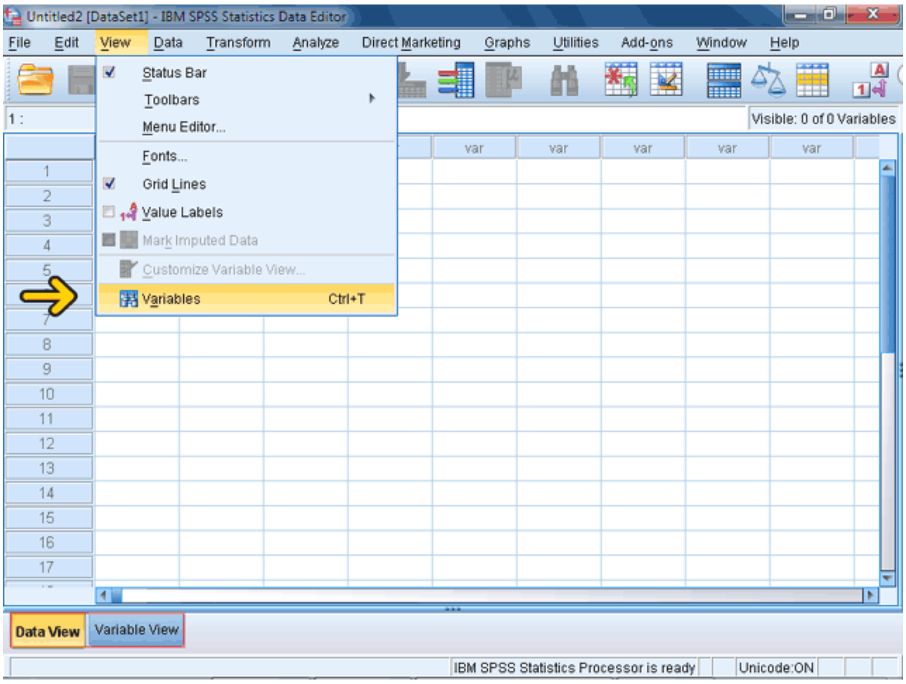


شکل (۱۵)

مشاهده می‌کنید که صفحهٔ کاربرگ جدید برای ورود داده‌ها باز شده است. در این صفحه متغیرها نامگذاری نشده و با نام Var نمایش داده شده‌اند.

۲- تا زمانی که متغیرها تعریف نشده باشند و داده‌ها در آن‌ها قرار نگرفته باشد SPSS هیچ کاری برای شما انجام نمی‌دهد. اکنون برای تعریف متغیرها منوی View را باز کنید.

برای وارد شدن به قسمت متغیرها روی گزینهٔ Variables کلیک کنید.

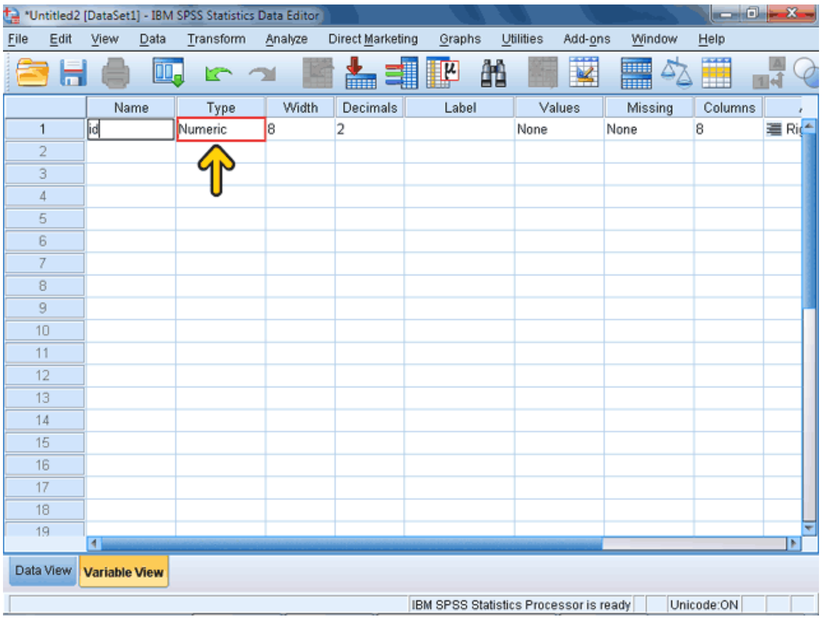


شکل (۱٦)

همان‌طور که مشاهده می‌کنید اکنون در قسمت Variable View قرار داریم. این پنجره نیز مانند صفحه کاربرگ است. در این صفحه در ستون Name نام متغیر را وارد می‌کنیم. اکنون id را به عنوان نام متغیر وارد کنید.

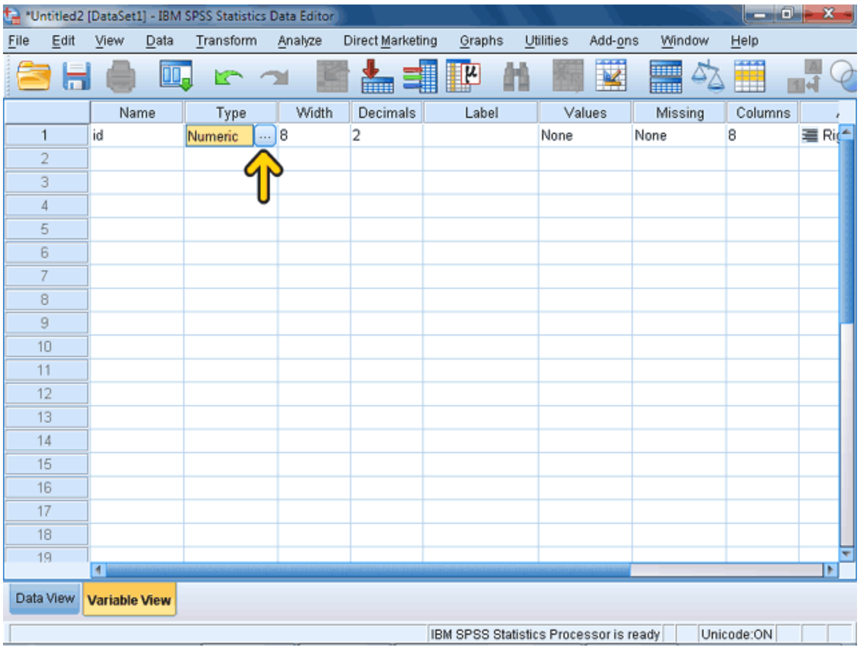
در ستون‌های بعدی این صفحه مشخصات دیگری مانند نوع متغیر (Type)، مقادیر متغیر (Values)، مقادیر غایب (Missing)، عرض ستون (Columns)، تراز (Align) و نوع مقایسه اندازه‌گیری (Measure) قابل تغییر است. در این قسمت نوع متغیر به طور پیش‌فرض از نوع عدد انتخاب شده است.

برای تعریف نوع متغیر روی ستون Type کلیک کنید.



شکل (۱۷)

سپس روی دکمهٔ مشخص شده کلیک کنید (شکل ۱۸).



شکل (۱۸)

اکنون، پنجرهٔ دیگری برای تعیین نوع متغیر باز شده است. در این پنجره انواع مختلف داده‌ها نمایش داده شده است. (شکل۱۹).

اولین گزینه Numeric است که جهت نمایش عدد به کار می‌رود. توسط گزینه‌های Width و Decimal Places می‌توانید تعداد ارقام اعشار و اعداد صحیح را تعیین کنید. این ارقام با علامت . از یکدیگر جدا می‌شوند.

اگر گزینهٔ Comma را انتخاب کنید، اعداد به همان صورت Numeric نمایش داده می‌شوند؛ با این تفاوت که ارقام عدد صحیح سه رقم سه رقم با کاما، جدا می‌شوند.

اگر گزینهٔ Dot را انتخاب کنید، عدد صحیح و اعشار با کاما، از هم جدا می‌شوند و ارقام صحیح سه رقم سه رقم با نقطه از هم جدا می‌شوند.

با انتخاب گزینهٔ Scientific notation ارقام به صورت نماد علمی نمایش داده می‌شوند.

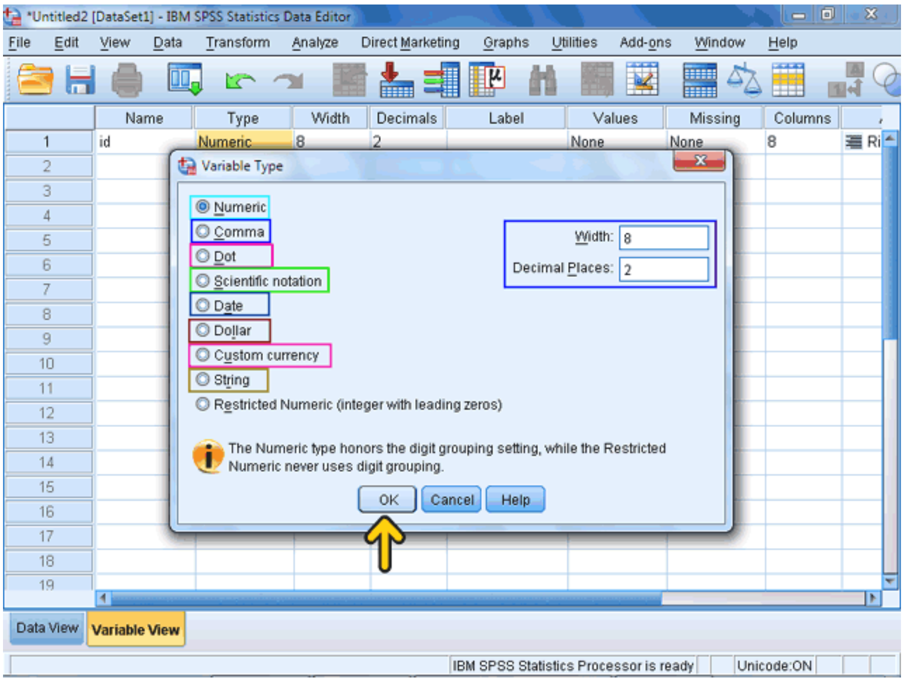
اگر گزینهٔ Date را انتخاب کنید، فرمت‌های مختلف نوشتن تاریخ در جلوی این گزینه ظاهر خواهد شد که با انتخاب فرمت دلخواه می‌توانید تاریخ را وارد کنید.

با انتخاب گزینهٔ Dollar علامت $ در سمت چپ عدد ظاهر خواهد شد و با توجه به فرمت انتخاب شده عدد با نقطه یا کاما جدا خواهد شد.

گزینهٔ Custom Currency نوعی عدد است که مقادیر آن به صورت یکی از فرمت‌های دلخواه تعریف شده در برگهٔ Currency در پنجرهٔ Options است.

با انتخاب گزینهٔ String می‌توانید داده را به صورت رشته وارد کنید. توجه کنید که این نوع داده عدد نیست و نمی‌توانید در محاسبات از آن استفاده کنید. حداکثر تعداد کاراکترهای این نوع داده نیز در قسمت Length تعریف می‌شود.

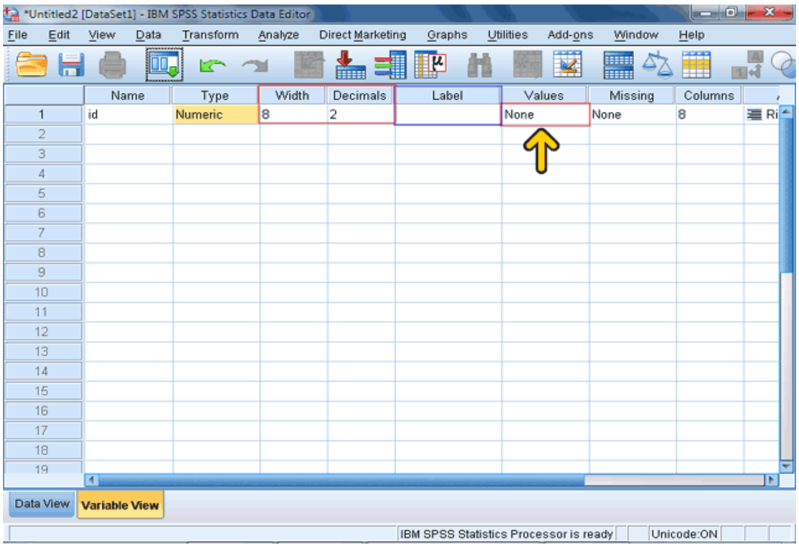
روی دکمهٔ OK کلیک کنید.



شکل (۱۹)

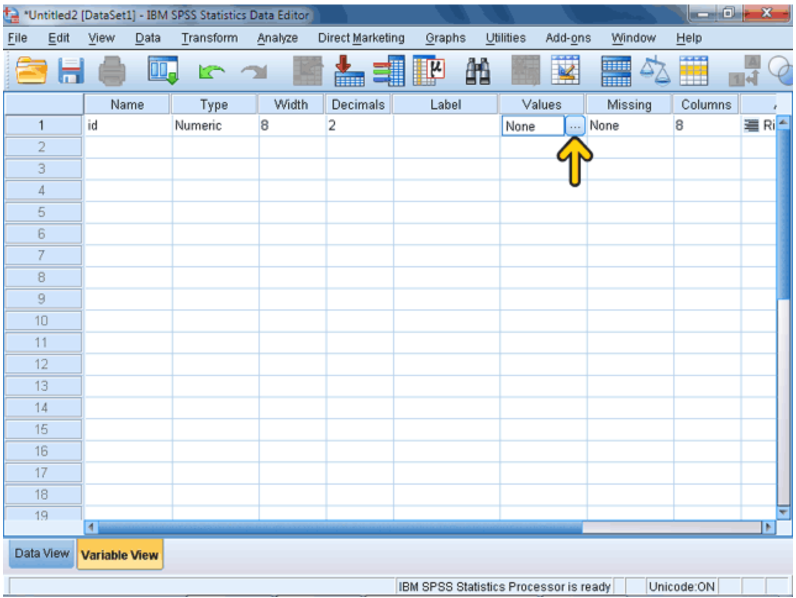
در قسمت مشخص شده توسط کادر قرمز گزینه‌های Width و Decimals را مشاهده می‌کنید (شکل ۲۰) که در این پنجره نیز وجود دارد و می‌توانید مقادیر آن‌ها را تغییر دهید. در قسمت Label می‌توانید عبارتی را برای متغیر وارد کنید که در نمودارها و ... این عبارت به جای نام متغیر ظاهر شود. در قسمت Values برای مقادیر خاصی می‌توانید حروف یا کلمهٔ خاصی را وارد کنید تا در نمودار آن مقدار با این کلمه نمایش داده شوند. برای مثال به کلمهٔ Male عدد ۱ و به کلمهٔ Female عدد ۲ را اختصاص دهید تا در نمودار به جای دو عدد ۱ و ۲، دو عبارت Male و Female بر روی محور مختصات نمایش داده شود. پس برای چنین مواردی لازم نیست داده را از نوع String تعیین کنید و برای ورود

اطلاعات هر بار کلمهٔ Male یا Female را تایپ کنید؛ فقط کافیست عدد ۱ یا ۲ را وارد کنید. اکنون برای تعریف یک مقدار بر روی ستون Values کلیک کنید.



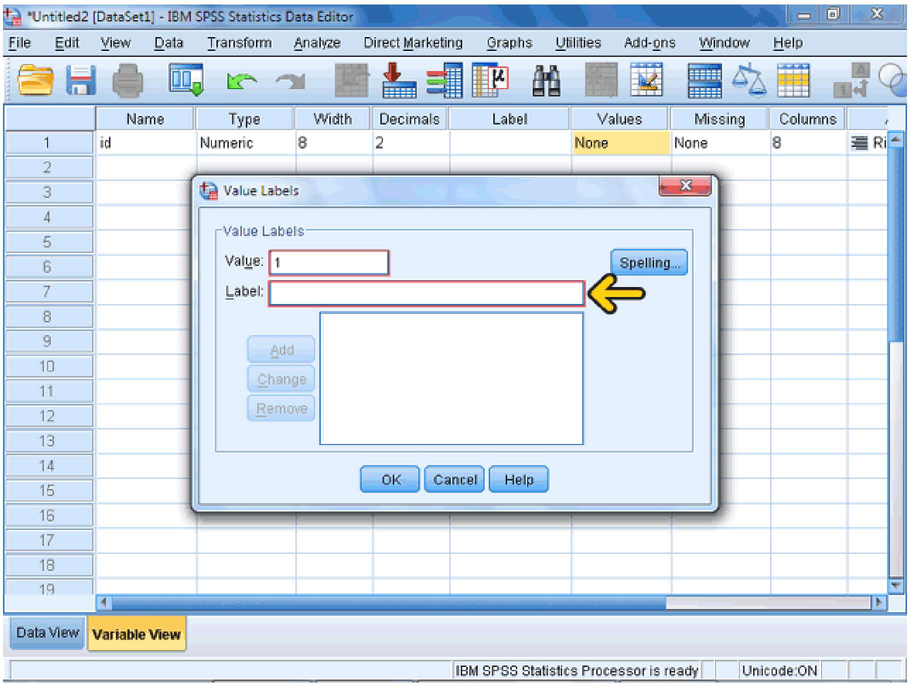
شکل (۲۰)

در مرحلهٔ بعد، روی علامت کنار ستون کلیک کنید (شکل ۲۱).



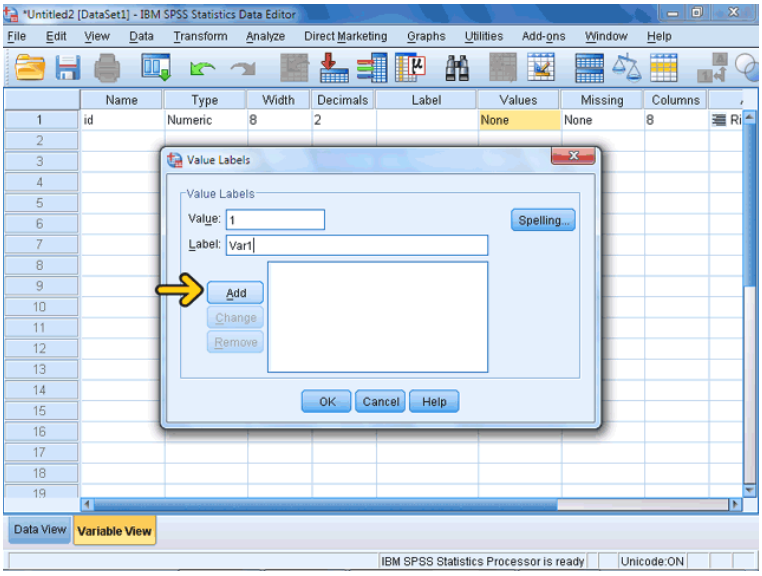
شکل (۲۱)

در صفحهٔ باز شده، روی جعبهٔ متن جلوی گزینهٔ Label کلیک کنید .(شکل ۲۲).
عدد ۱ را وارد کنید.



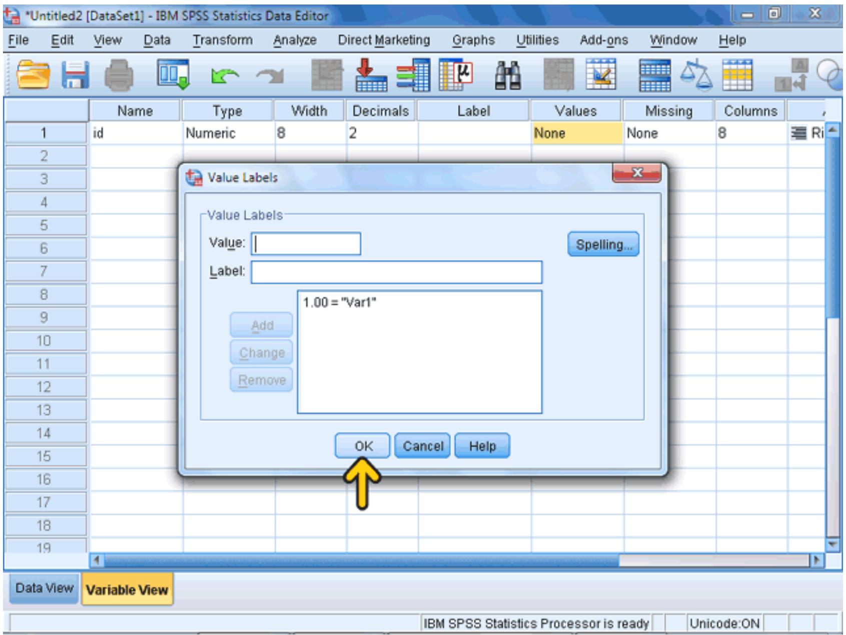
شکل (۲۲)

در این مرحله، در قسمت Label یک نام دلخواه برای مقداری که می‌خواهیم تعریف کنیم وارد می‌کنیم. برای مثال عبارت Var۱ را تایپ کنید.
روی دکمهٔ Add کلیک کنید (شکل ۲۳).



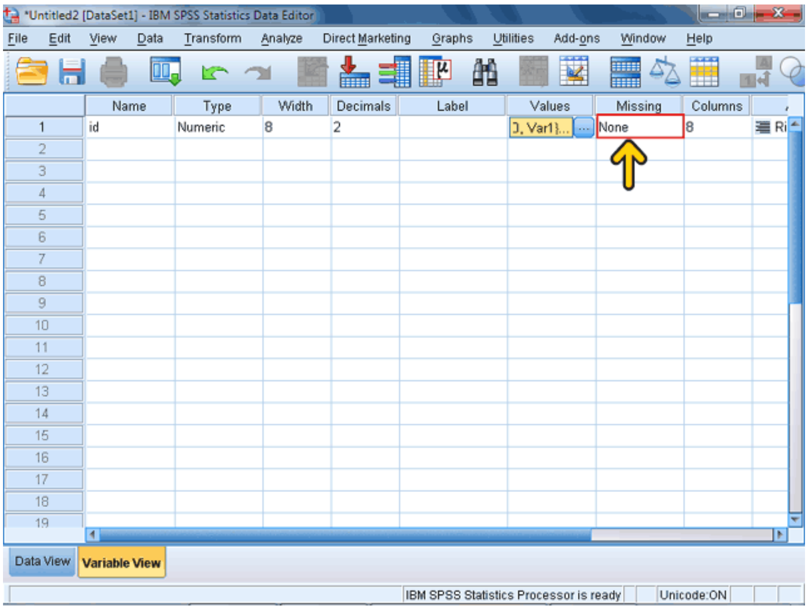
شکل (۲۳)

با این کار یک متن به نام Var۱ تعریف کرده‌ایم که برابر عدد یک است. روی دکمهٔ OK کلیک کنید (شکل ۲٤).



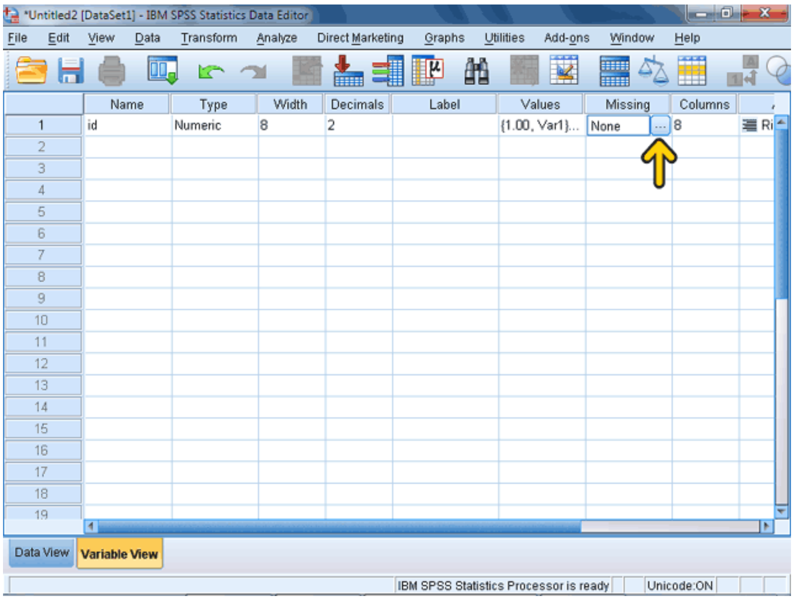
شکل (۲٤)

با توجه به شکل ۲۵، در قسمت دادهٔ Missing کلیک کنید.



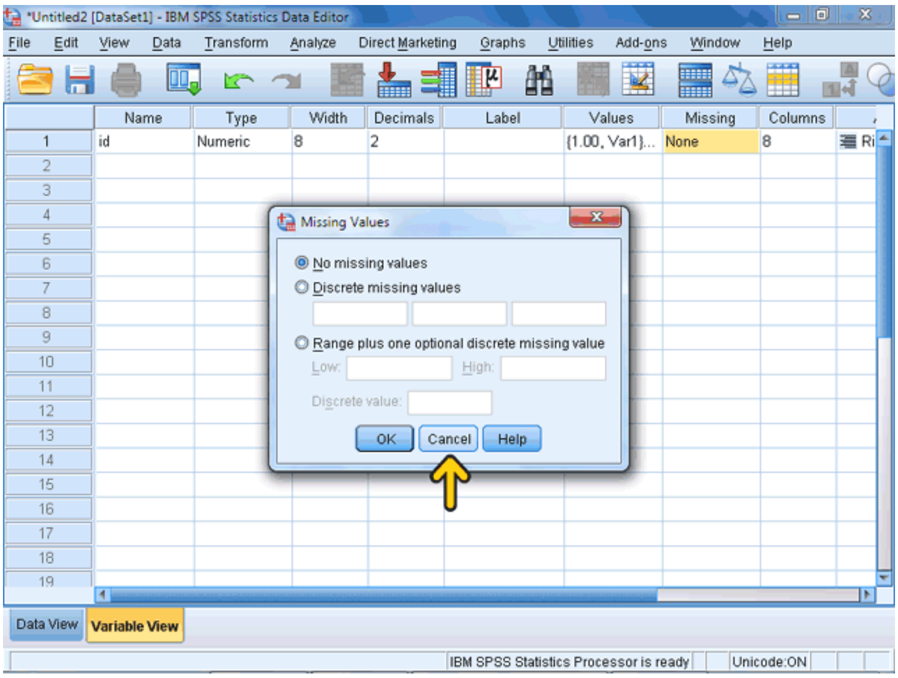
شکل (۲۵)

سپس، روی دکمهٔ ... کلیک کنید (شکل ۲٦).



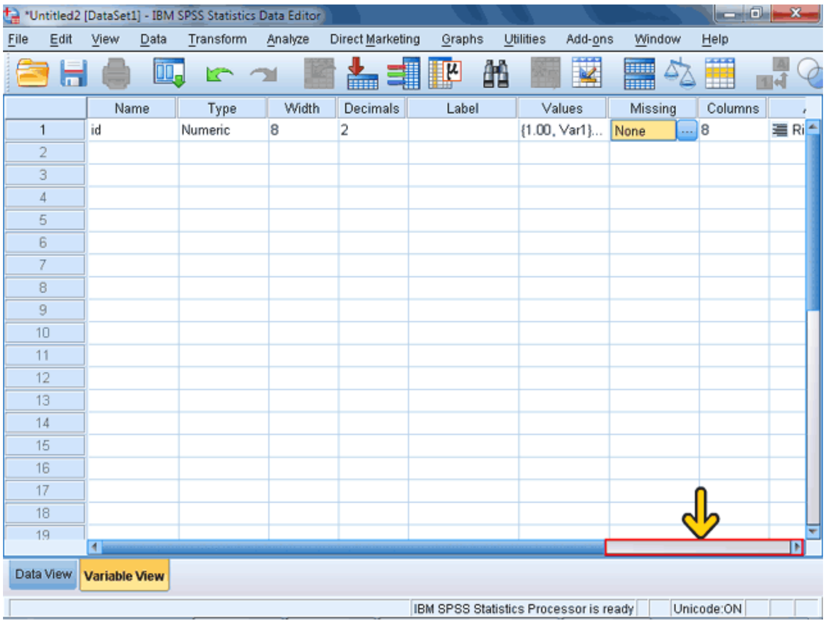
شکل (۲٦)

اکنون می‌توانید در این پنجره روش برخورد SPSS با مقادیر غایب را تعیین کنید. با انتخاب گزینهٔ اول (No Missing Values) یعنی هیچ مقدار غایب در این پرونده وجود ندارد. با انتخاب گزینهٔ Discrete Missing Values می‌توانید حداکثر سه داده به عنوان مقادیر غایب وارد کنید. گزینه Range Plus Discrete MissingValues با انتخاب این گزینه می‌توانید دامنهٔ مقادیر بین کوچک‌ترین و بزرگ‌ترین عدد، به اضافه یک دادهٔ خارج از دامنه را به عنوان مقادیر غایب وارد کنید. روی دکمهٔ Cancel کلیک کنید (شکل ٢٧).



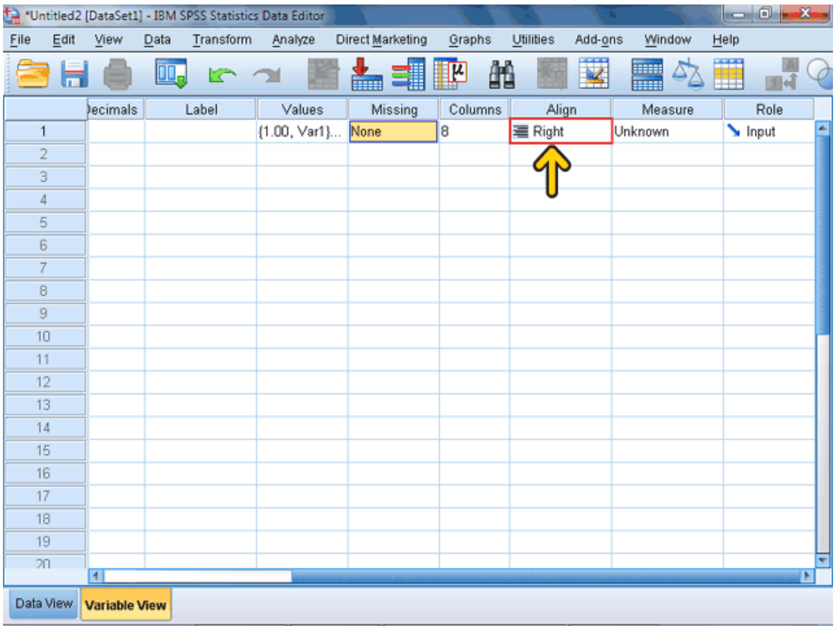
شکل (٢٧)

روی میلهٔ لغزان (کادر قرمز) کلیک کنید (شکل ٢٨).



شکل (۲۸)

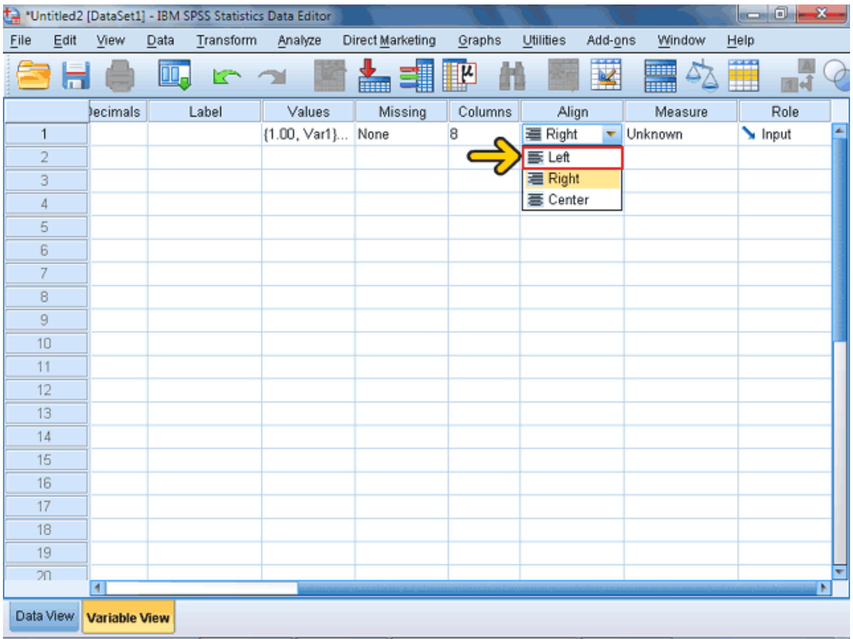
برای تعیین تراز ستون روی ستون Align کلیک کنید.



شکل (۲۹)

روی لیست بازشونده کلیک کنید (شکل ۳۰).

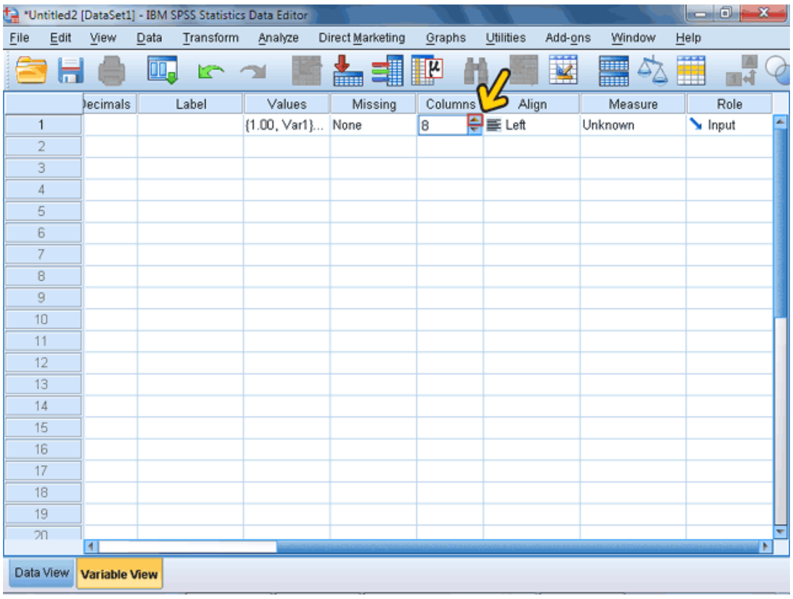
مشاهده می‌کنید این لیست دارای سه گزینه است که هر یک از این گزینه‌ها نحوهٔ قرارگیری داده‌ها در سلول‌ها را مشخص می‌کنند. برای آنکه داده‌های موجود در این ستون در سمت چپ ستون قرار بگیرند روی گزینهٔ Left کلیک کنید.



شکل (۳۰)

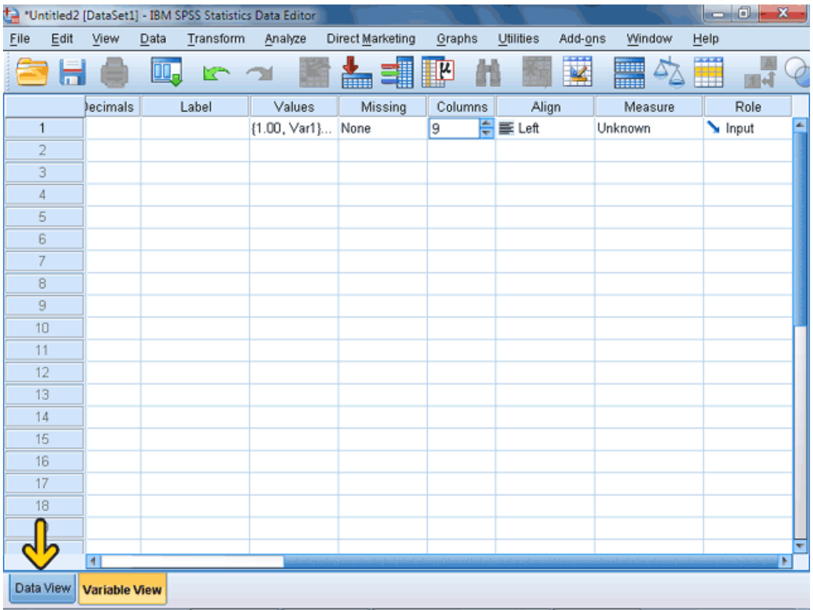
برای تعریف عرض ستون‌ها روی ستون Columns کلیک کنید (شکل ۳۱).

در این قسمت می‌توانید عرض ستون‌ها را به دلخواه تغییر دهید. برای مثال، روی دکمهٔ افزایش مشخص شده برای اضافه شدن عرض ستون‌ها کلیک کنید.



شکل (۳۱)

با توجه به شکل۳۲، مشاهده می‌کنید مقدار ۸ به ۹ تغییر پیدا کرده است. برای برگشت به کاربرگ، روی برگهٔ Data View کلیک کنید.

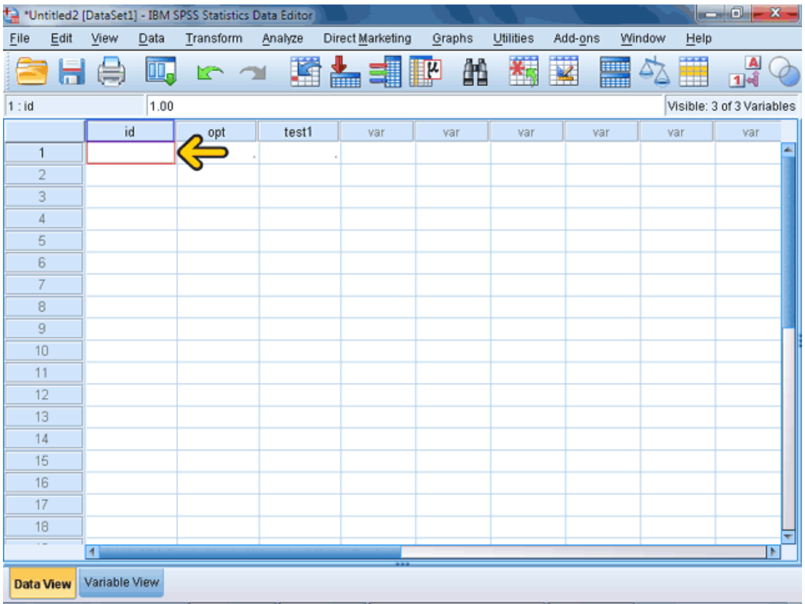


شکل (۳۲)

در نهایت مشاهده می‌کنید این متغیر به کاربرگ اضافه شده است. اکنون می‌توان داده‌های مورد نیاز را در این متغیر ذخیره کرد. در ادامه دو متغیر دیگر نیز به همین روش اضافه می‌کنیم.

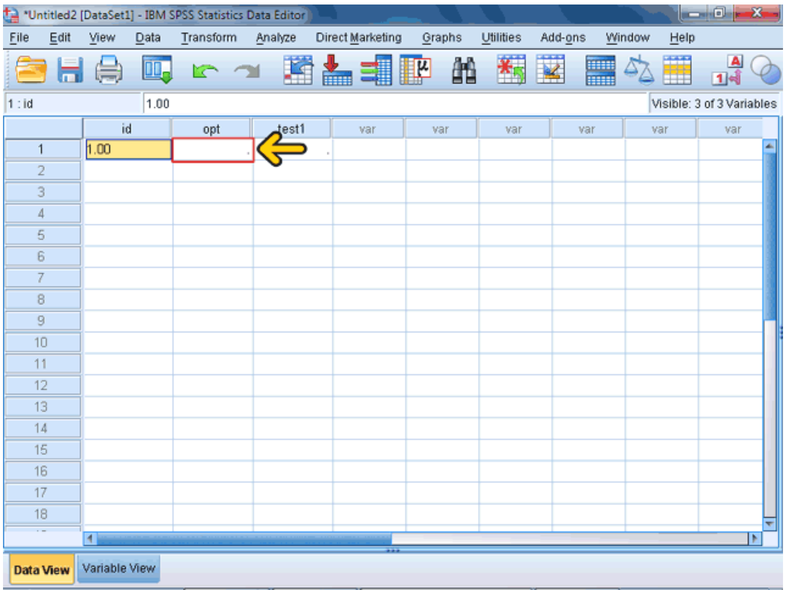
در این قسمت برای راحتی کار، قبل از ورود و ویرایش داده‌ها متغیرهای مورد نیاز برنامه را ایجاد کرده‌ایم. برای ورود داده‌ها روی سلول مشخص شده کلیک کنید.

عدد ۱ را وارد کنید (شکل ۳۳).



شکل (۳۳)

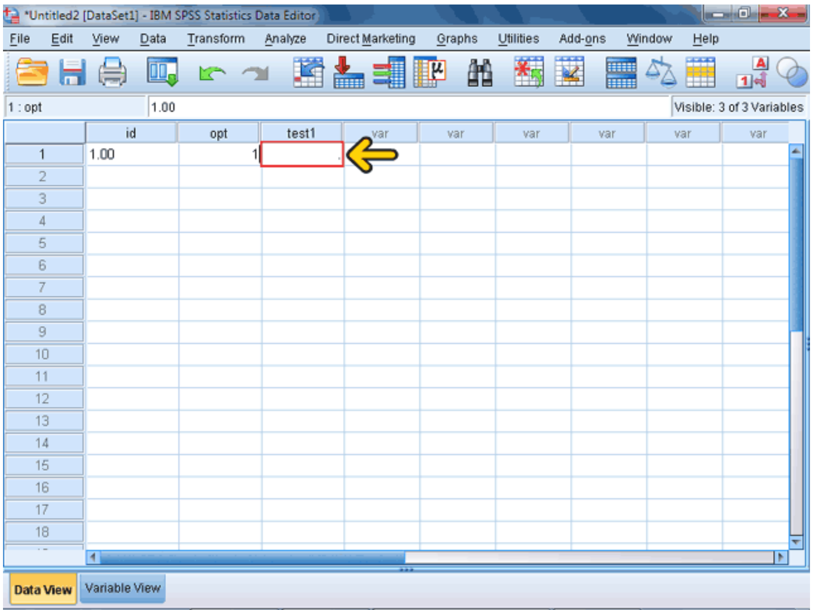
روی سلول opt کلیک کنید (شکل ۳٤).



شکل (۳٤)

در همان جا عدد ۱ را وارد کنید.

روی سلول test۱ کلیک کنید (شکل ۳۵).

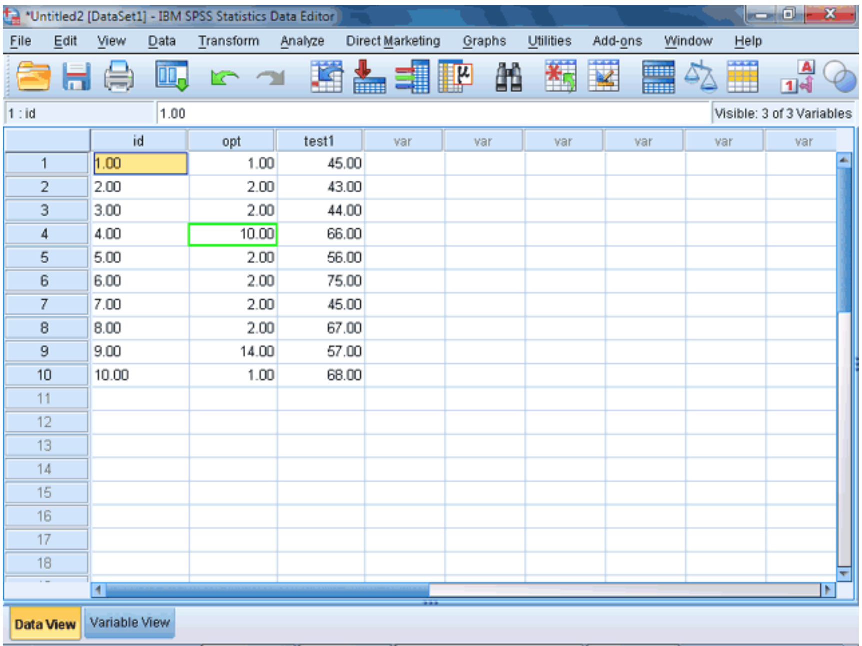


شکل (۳۵)

عدد ٤٥ را تایپ کنید.

به همین ترتیب چندین عدد را برای مثال وارد می‌کنیم.

اکنون می‌خواهیم مقدار موجود در سلول مشخص شده را تغییر دهیم. برای این کار روی سلول مشخص شده کلیک کنید.

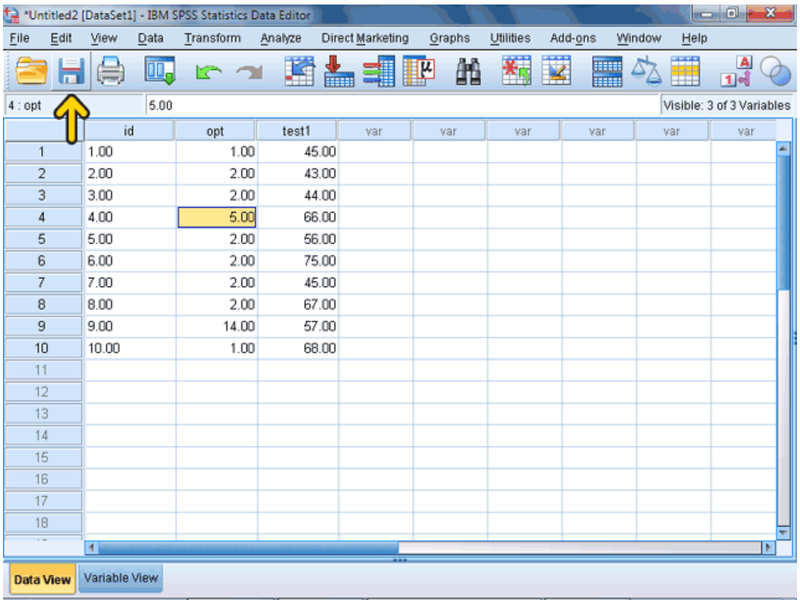


شکل (۳٦)

عدد ٥ را تایپ کنید (شکل ۳۷).

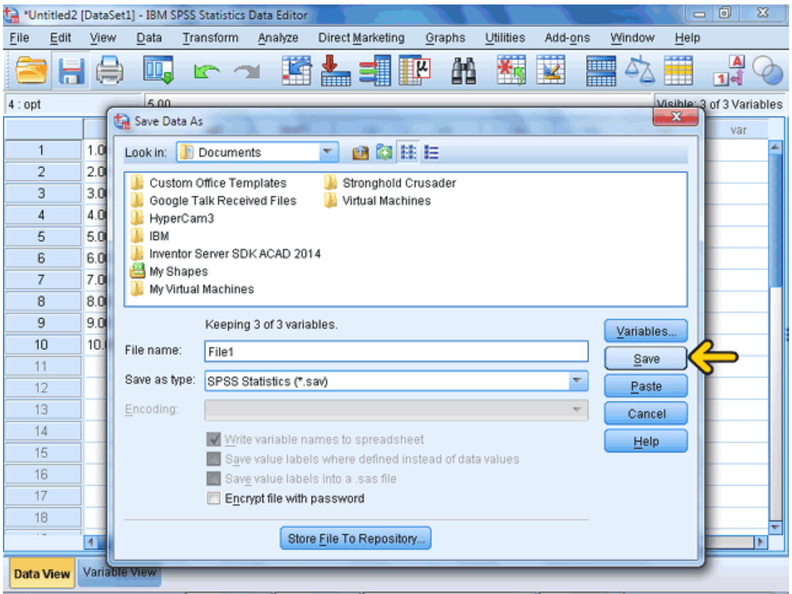
مشاهده می‌کنید مقدار داخل سلول تغییر پیدا کرده است.

بعد از ورود داده‌ها و ویرایش‌های لازم باید روی آن‌ها پرونده را ذخیره کرد. بر روی دکمهٔ Save File کلیک کنید.



شکل (۳۷)

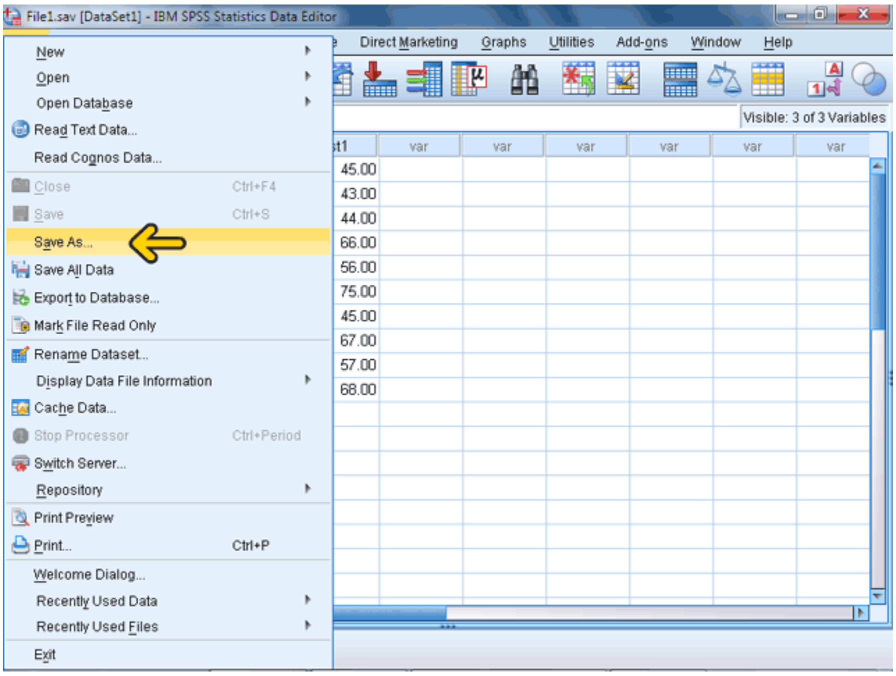
در ادامه در قسمت File Name عبارت file۱ را به عنوان نام فایل وارد می‌کنیم. روی دکمهٔ Save کلیک کنید. (شکل ۳۸).



شکل (۳۸)

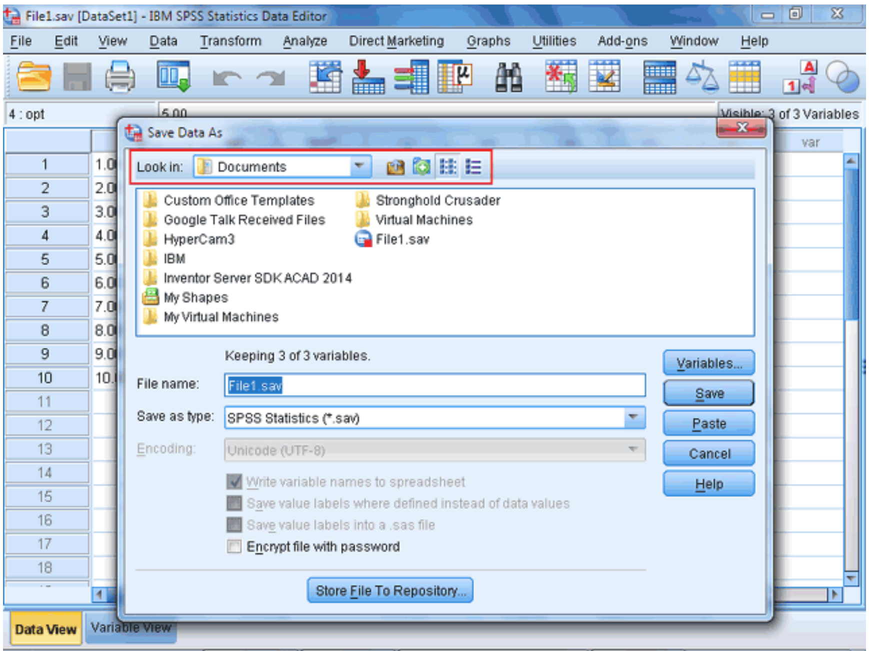
اکنون این پرونده با نام file۱ و پسوند sav ذخیره شده است. برای ذخیرهٔ این پرونده تحت یک نام دیگر یا در مکان دیگری مانند دیسکت، باید از گزینه Save as استفاده کنیم. برای این کار منوی File را باز کنید.

روی گزینهٔ Save As کلیک کنید (شکل ۳۹).



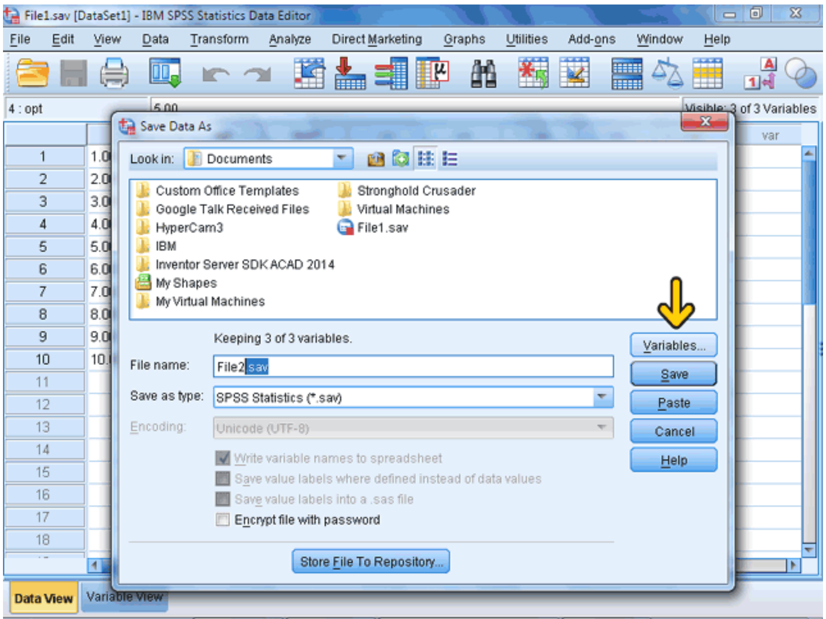
شکل (۳۹)

با استفاده از بخش مشخص شده توسط کادر قرمز می‌توانید شاخه یا درایوی را که می‌خواهید فایل در آنجا ذخیره شود، تعیین کنید. عبارت file۲ را به عنوان نام جدید پرونده تایپ کنید (شکل ٤۰).



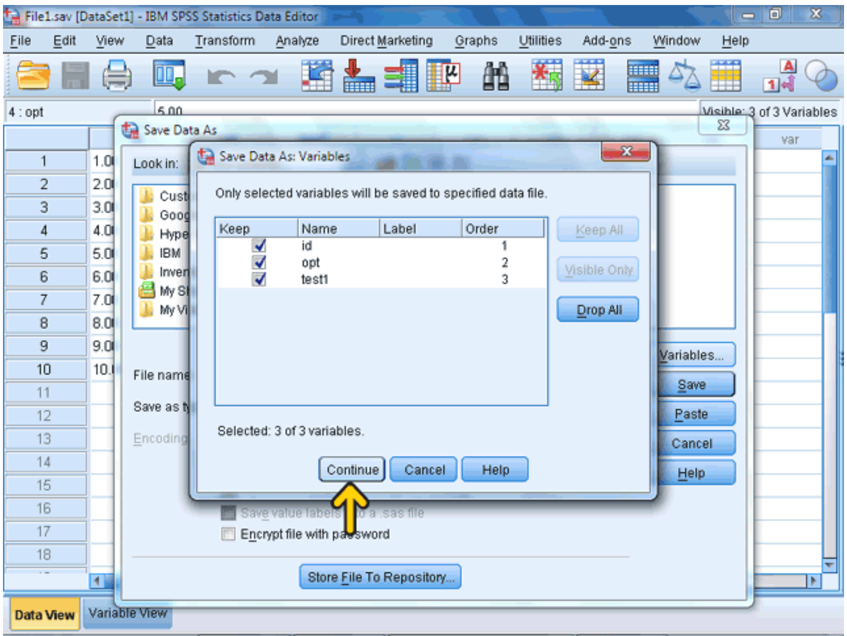
شکل (٤٠)

روی دکمهٔ Variables کلیک کنید (شکل ٤١).



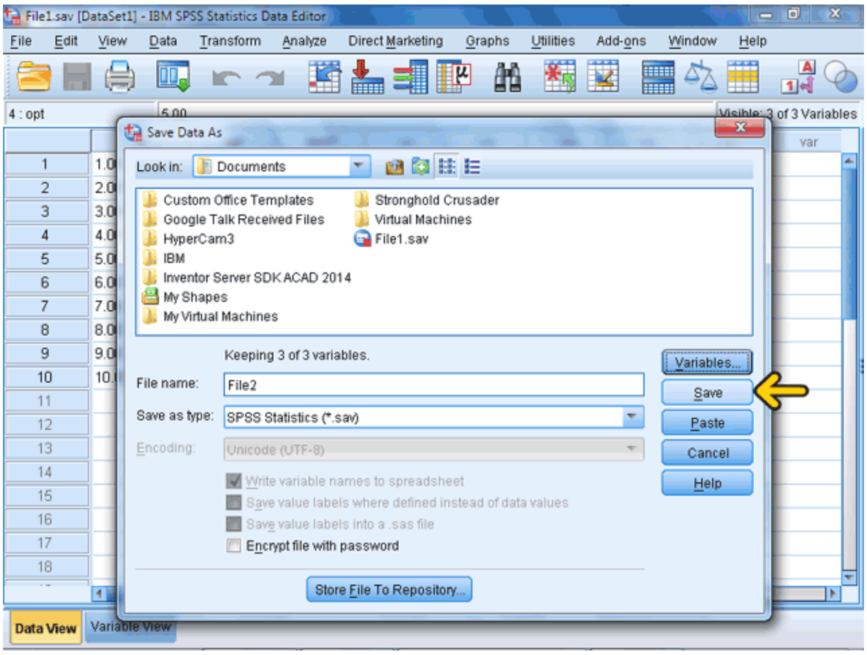
شکل (٤١)

در این پنجره می‌توانید تعیین کنید که کدام متغیرها ذخیره شوند. اگر بخواهید داده‌های متغیر Opt ذخیره نشود، کافیست روی مربع قسمت Keep این متغیر کلیک کنید. با استفاده از دکمهٔ Keep All تمامی متغیرها انتخاب شده و توسط دکمهٔ Drop All تمامی متغیرها حذف می‌شوند. برای تأیید تنظیمات باید روی دکمهٔ Continue کلیک کنید و در صورتی که می‌خواهید انصراف دهید روی دکمهٔ Cancel کلیک کنید. اما در اینجا روی دکمهٔ Continue کلیک کنید (شکل ٤٢).



شکل (٤٢)

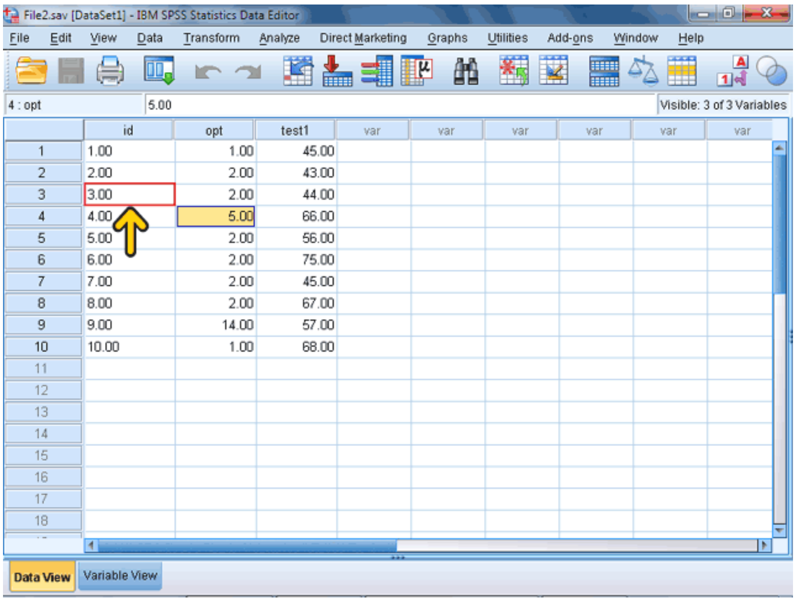
اکنون عبارت Keeping ۳ of ۳ variables بالای نام فایل نمایش داده شده است. این عبارت نشاندهندهٔ این است که سه متغیر از سه متغیر موجود در این فایل ذخیره می‌شوند. روی دکمهٔ Save کلیک کنید (شکل ٤٣).



شکل (٤٣)

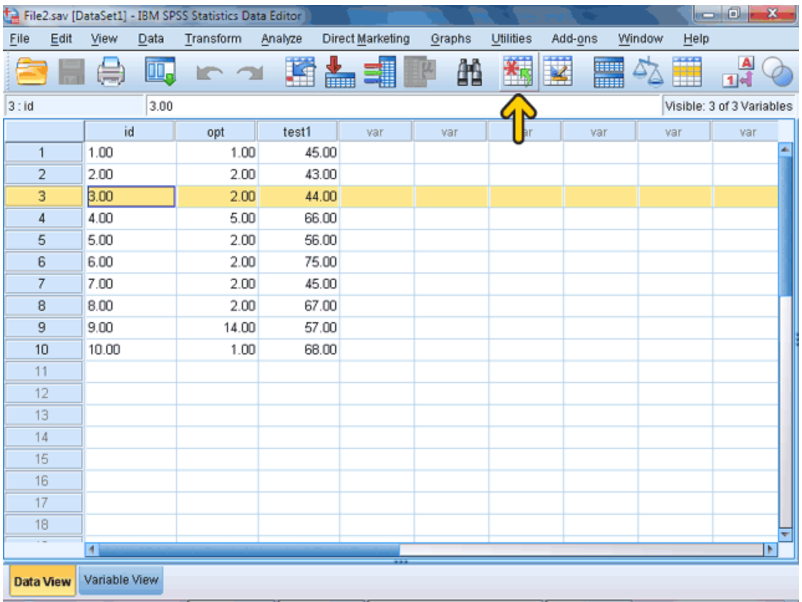
اکنون این پرونده با دو نام file۱ و file۲ ذخیره شده است.

اکنون می‌خواهیم در این پرونده یک سطر جدید ایجاد کنیم. برای این منظور باید روی سلولی از ردیفی که می‌خواهید سطر جدید در بالای آن قرار بگیرد، کلیک کنید. برای مثال روی سلول مشخص شده کلیک کنید (شکل ٤٤).



شکل (٤٤)

روی دکمهٔ Insert Case کلیک کنید. برای اجرای این دستور می‌توانید منوی Edit را باز و گزینهٔ Insert Case را انتخاب کنید (شکل ٤٥).

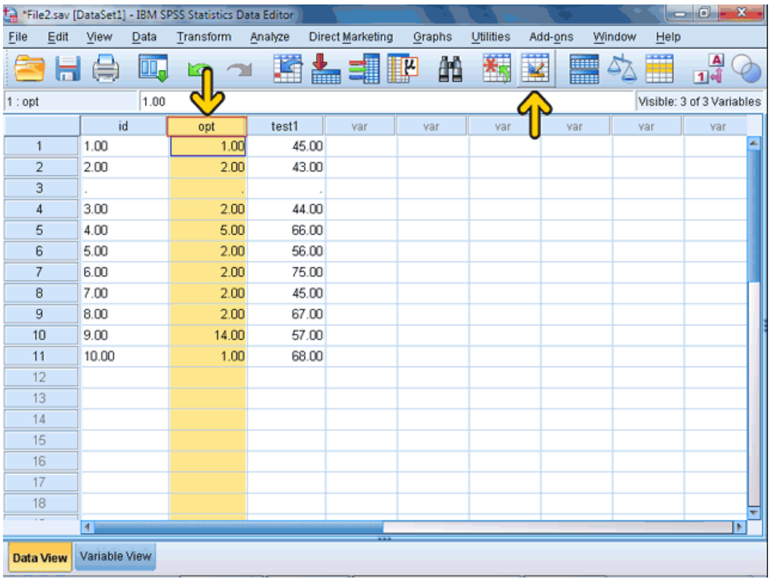


شکل (٤٥)

مشاهده می‌کنید یک سطر به کاربرگ اضافه شده است و می‌توانید در این سطر مقادیر دلخواه خود را وارد کنید.

برای افزودن یک متغیر جدید به صفحهٔ کاربرگ، روی سلولی که می‌خواهید متغیر جدید قبل از آن اضافه شود کلیک کنید. برای مثال روی سلول مشخص شده کلیک کنید.

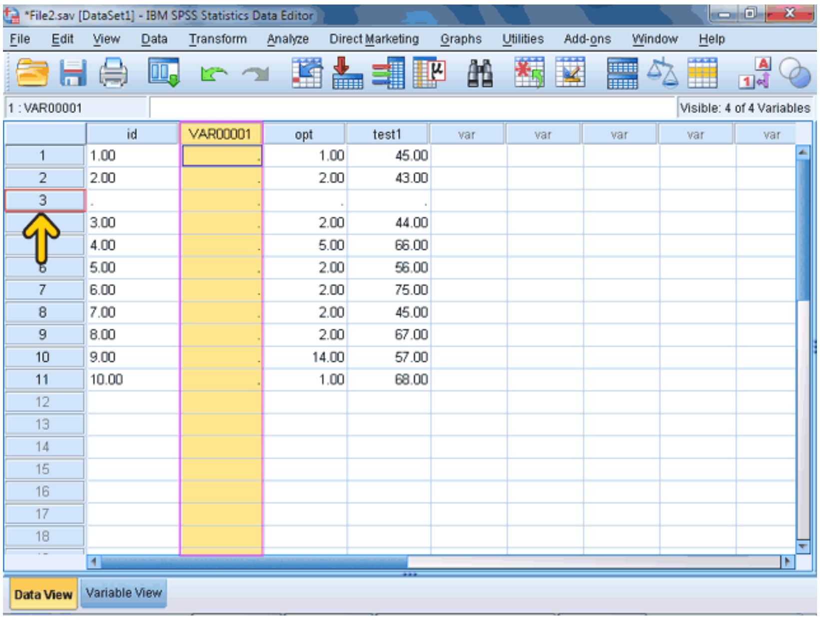
روی دکمهٔ Insert Variable کلیک کنید. برای اجرای این دستور می‌توانید منوی Edit را باز و گزینهٔ Insert Variable را انتخاب کنید.



شکل (٤٦)

مشاهده می‌کنید که یک متغیر جدید به صفحهٔ کاربرگ افزوده شده است.

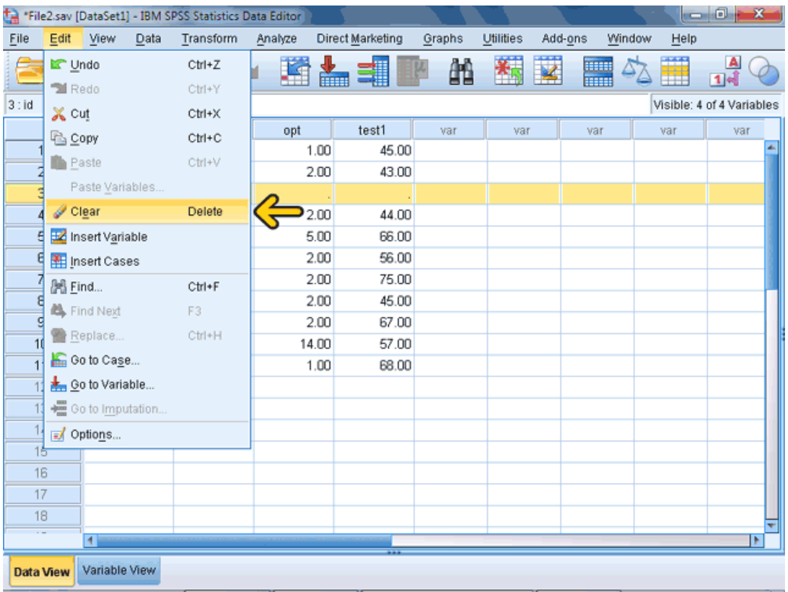
برای حذف سطر یا رکورد ابتدا باید سطر مورد نظر خود را انتخاب کنیم. برای حذف سطر شمارهٔ سه روی شمارهٔ سطر که با کادر قرمز مشخص شده است کلیک کنید (شکل ٤٧).



شکل (٤٧)

منوی Edit را باز کنید.

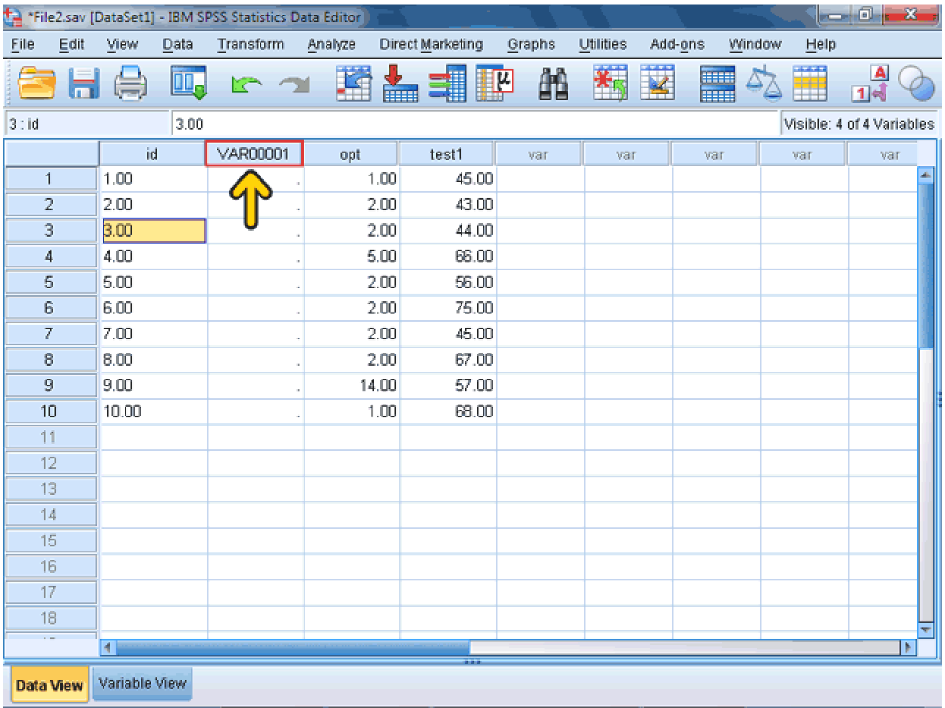
روی گزینهٔ Clear کلیک کنید (شکل ٤٨).



شکل (٤٨)

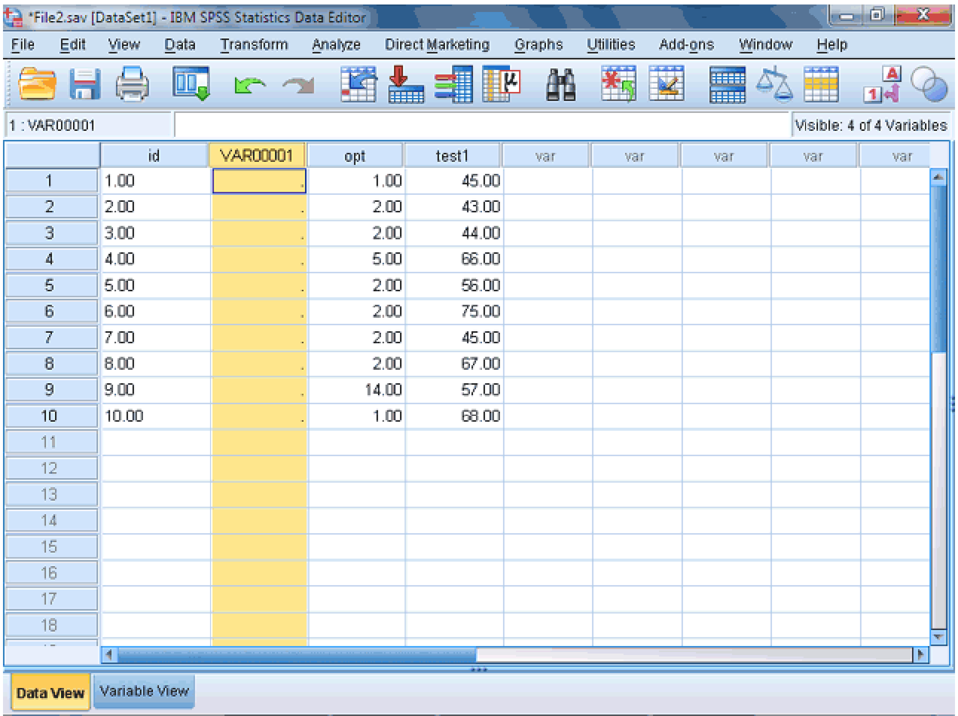
مشاهده می‌کنید که با این کار، سطر انتخاب شده، حذف شده است.

برای حذف یک متغیر یا ستون روی عنوان متغیر در صفحهٔ کاربرگ کلیک کنید (شکل ٤٩).



شکل (٤٩)

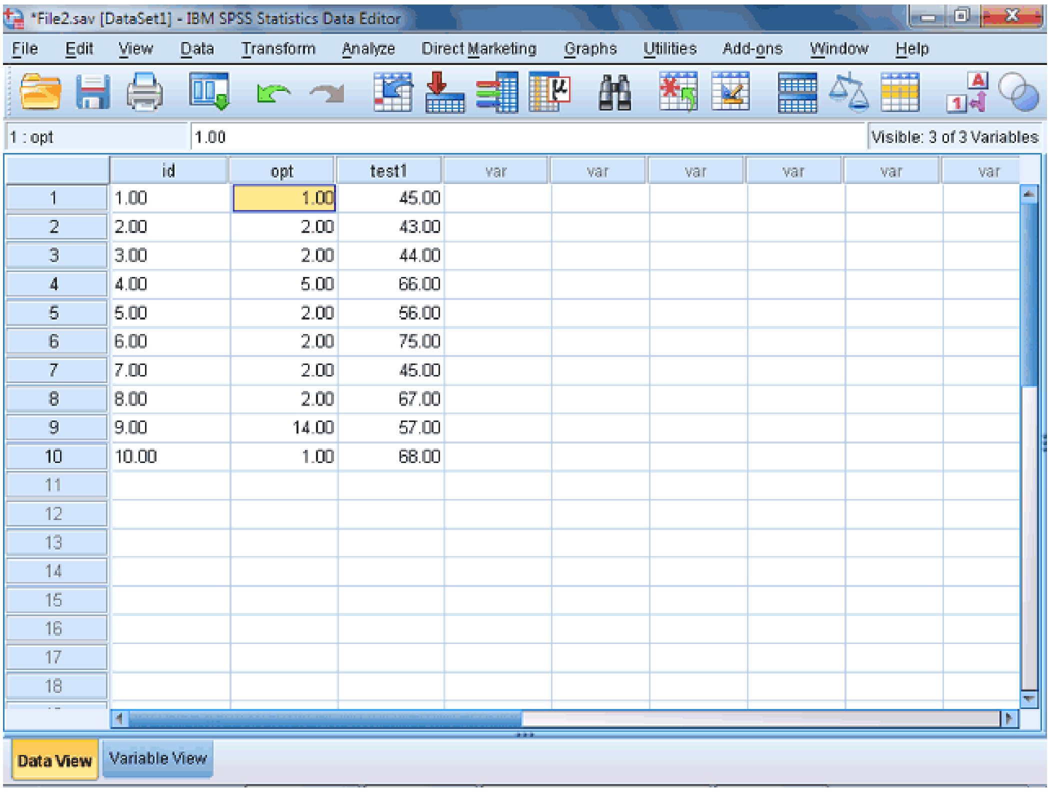
در اینجا نیز می‌توانیم با استفاده از منوی Edit و انتخاب گزینهٔ Clear این متغیر را حذف کنیم. راه دیگر، استفاده از دکمهٔ Delete صفحه کلید است. کلید Delete صفحهٔ کلید را فشار دهید تا این متغیر حذف شود (شکل ٥٠).



شکل (۵۰)

مشاهده می‌کنید متغیر مورد نظر از صفحه حذف شده است. برای حذف بیش از یک سطر یا ستون در زمان انتخاب سطر یا ستون، به جای انتخاب یک سطر یا ستون، در حالی که نشانگر ماوس را روی شمارهٔ اولین سطر یا عنوان اولین ستون قرار داده‌اید، کلید سمت چپ را فشرده نگه دارید و سپس آن را روی بقیهٔ سطرها یا ستون‌ها حرکت دهید تا با هم انتخاب شوند. سپس با استفاده از کلید Delete از صفحهٔ کلید یا فرمان Clear از منوی Edit می‌توانید ردیف‌ها یا ستون‌های مورد نظر خود را حذف کنید.

شما اکنون در پایان این بخش قرار دارید (شکل ۵۱).


*File2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
1 : opt
1.00
Visible: 3 of 3 Variables

	id	opt	test1	var	var	var	var	var	var
1	1.00	1.00	45.00						
2	2.00	2.00	43.00						
3	3.00	2.00	44.00						
4	4.00	5.00	66.00						
5	5.00	2.00	56.00						
6	6.00	2.00	75.00						
7	7.00	2.00	45.00						
8	8.00	2.00	67.00						
9	9.00	14.00	57.00						
10	10.00	1.00	68.00						
11									
12									
13									
14									
15									
16									
17									
18									

Data View
Variable View

شکل (۵۱)

حل مسئله

۱- اگر نمرات ارزشیابی ۵۰ نفر از کارکنان کتابخانه‌های کشور در یک بررسی آماری به صورت زیر داده شده باشد:

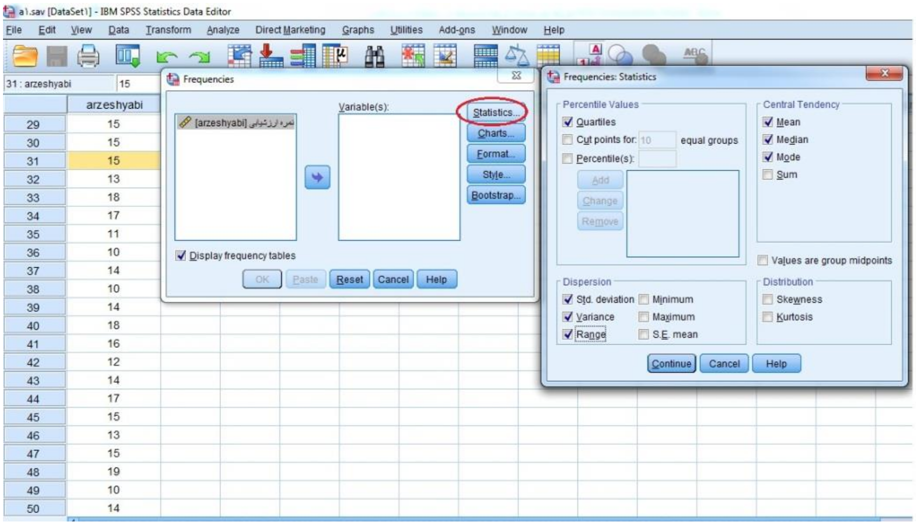
۱۳ ۱٦ ۱۰ ۱۵ ۱۸ ۱٤ ۱۵ ۱۷ ۱۲ ۱۹
۱۵ ۱۲ ۱٤ ۱۳ ۱۱ ۱۷ ۱۳ ۱٦ ۱۳ ۱۷
۱۹ ۱٤ ۱۰ ۱۸ ۱٦ ۱۱ ۲۰ ۱۵ ۱٤ ۱۷
۱۰ ۱۷ ۱٤ ۱۷ ۱۵ ۱٦ ۱٤ ۱۷ ۱۸ ۱۳
۱٤ ۱۵ ۱۸ ۱۱ ۱۵ ۱۸ ۱۲ ۱۹ ۱۰ ۱٤

الف- شاخص‌های مرکزی نمونه‌ای شامل نما، میانه، میانگین و چارک‌ها را محاسبهٔ کنید.

ب- شاخص‌های پراکندگی نمونه‌ای شامل دامنه، واریانس، انحراف معیار و ضریب تغییر را محاسبهٔ کنید.

ج- جدول توزیع فراوانی را تشکیل دهید.

حل:



Statistics

نمرهٔ ارزشیابی

N	Valid	۵۰
	Missing	۰
Mean		۱۴,۸۲
Median		۱۵,۰۰
Mode		۱۴
Std. Deviation		۲,۶۷۸
Variance		۷,۱۷۱
Range		۱۰
Percentiles	۲۵	۱۳,۰۰
	۵۰	۱۵,۰۰
	۷۵	۱۷,۰۰

نمرهٔ ارزشیابی

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	۱۰	۴	۸,۰	۸,۰	۸,۰
	۱۱	۳	۶,۰	۶,۰	۱۴,۰
	۱۲	۳	۶,۰	۶,۰	۲۰,۰
	۱۳	۵	۱۰,۰	۱۰,۰	۳۰,۰
	۱۴	۸	۱۶,۰	۱۶,۰	۴۶,۰
	۱۵	۷	۱۴,۰	۱۴,۰	۶۰,۰
	۱۶	۴	۸,۰	۸,۰	۶۸,۰
	۱۷	۷	۱۴,۰	۱۴,۰	۸۲,۰
	۱۸	۵	۱۰,۰	۱۰,۰	۹۲,۰
	۱۹	۳	۶,۰	۶,۰	۹۸,۰
	۲۰	۱	۲,۰	۲,۰	۱۰۰,۰
	Total	۵۰	۱۰۰,۰	۱۰۰,۰	

۲ – تعداد کتب موجود در بیست و یک کتابخانه به‌صورت زیر گزارش شده است(ارقام به هزار عدد).

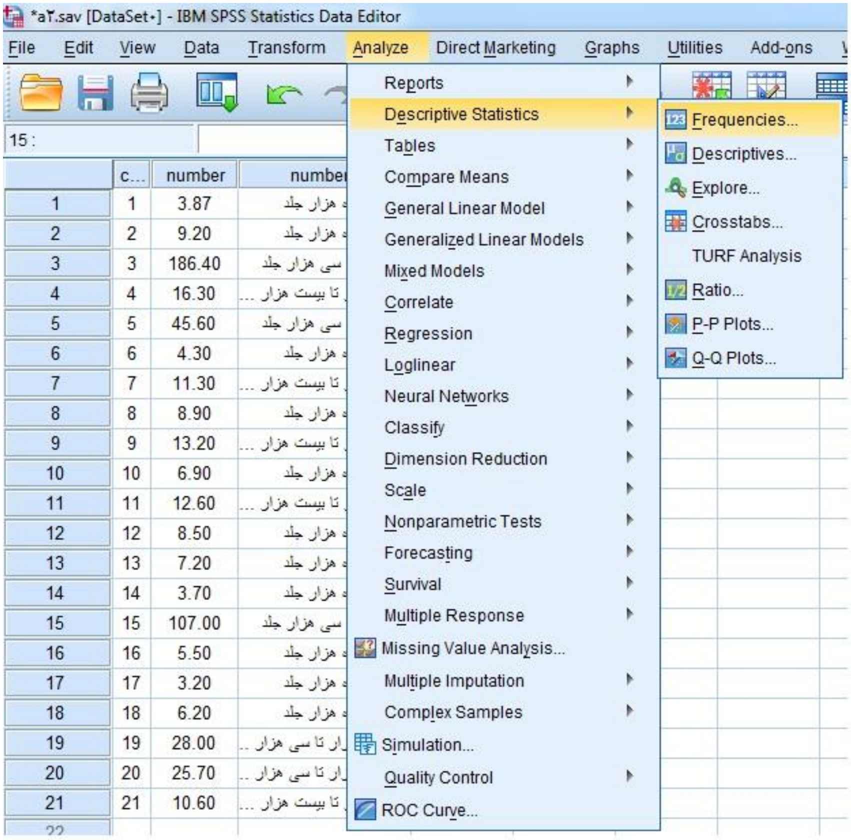
تعداد کتب	کتابخانه	تعداد کتب	کتابخانه	تعداد کتب	کتابخانه
١٠٧	١٥	٨/٩	٨	٣/٨٧	١
٥/٥	١٦	١٣/٢	٩	٩/٢	٢
٣/٢	١٧	٦/٩	١٠	١٨٦/٤	٣
٦/٢	١٨	١٢/٦	١١	١٦/٣	٤
٢٨	١٩	٨/٥	١٢	٤٥/٦	٥
٥٧/٧	٢٠	٧/٢	١٣	٤/٣	٦
١٠/٦	٢١	٣/٧	١٤	١١/٣	٧

الف- جدول توزیع فراوانی را تشکیل دهید.

ب- نمودار دایره‌ای را رسم کنید.

ج- نمودار مستطیلی و چندضلعی را رسم کنید.

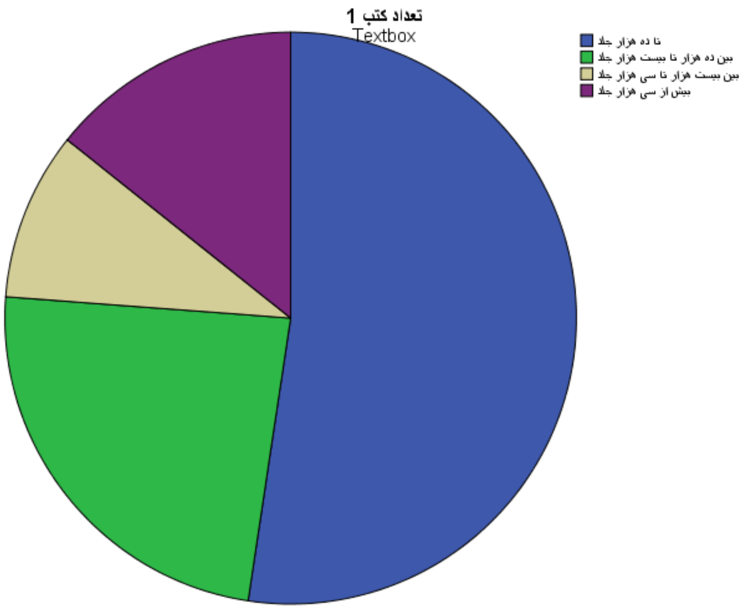
حل:

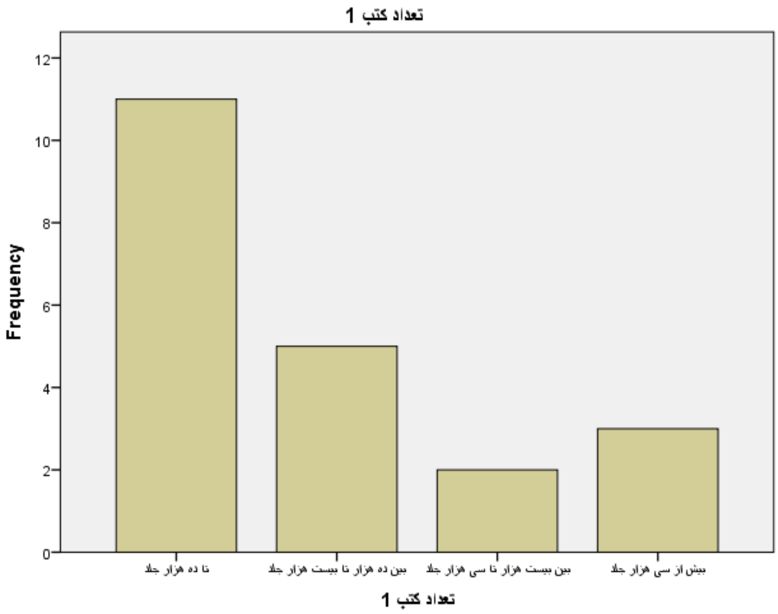

*a۲.sav [DataSet۰] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...

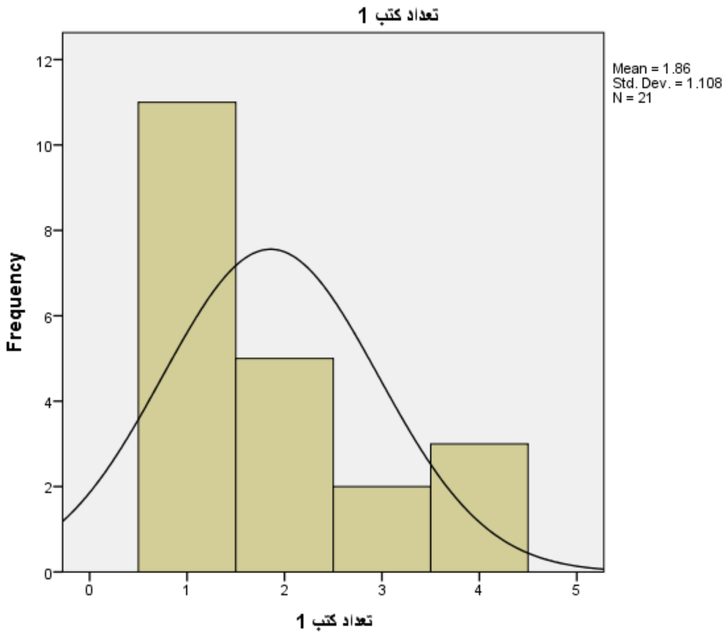

Frequencies
[case] شماره کتابخانه
Variable(s):
[number1] 1 تعداد کتب
تعداد کتب موجود در ...
Statistics...
Charts...
Format...
Style...
Bootstrap...
Display frequency tables
OK Paste Reset Cancel Help
Frequencies: Charts
Chart Type
None
Bar charts
Pie charts
Histograms:
Show normal curve on histogram
Chart Values
Frequencies Percentages
Continue Cancel Help

تعداد کتب ۱

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	تا ده هزار جلد	۱۱	۵۲,٤	۵۲,٤	۵۲,٤
	بین ده هزار تا بیست هزار جلد	۵	۲۳,۸	۲۳,۸	۷٦,۲
	بین بیست هزار تا سی هزار جلد	۲	۹,۵	۹,۵	۸۵,۷
	بیش از سی هزار جلد	۳	۱٤,۳	۱٤,۳	۱۰۰,۰
	Total	۲۱	۱۰۰,۰	۱۰۰,۰	




تعداد کتب 1
Frequency
12
10
8
6
4
2
0
تا ده هزار جلد
بین ده هزار تا بیست هزار جلد
بین بیست هزار تا سی هزار جلد
بیش از سی هزار جلد
تعداد کتب 1

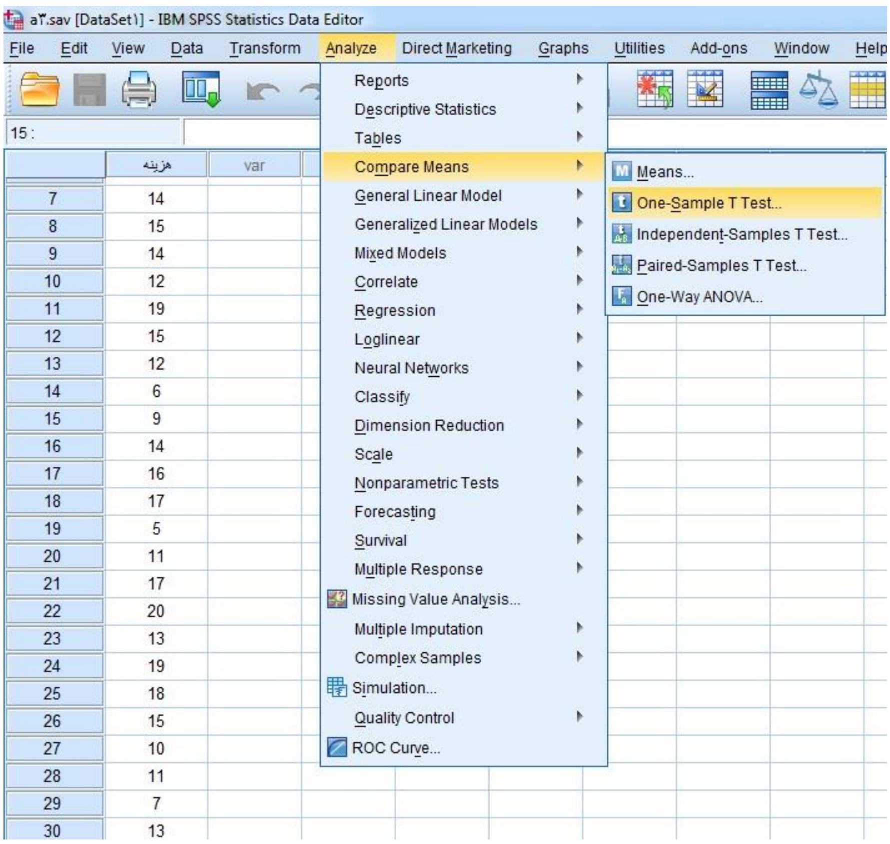

تعداد کتب 1
Mean = 1.86
Std. Dev. = 1.108
N = 21
Frequency
12
10
8
6
4
2
0
0
1
2
3
4
5
تعداد کتب 1

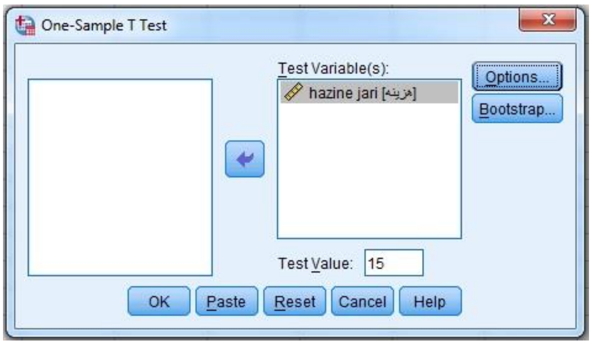
۳- اطلاعات زیر، هزینه‌های جاری ۳۰ کتابخانه را نشان می‌دهد (ارقام به میلیون تومان):

۱۰	۱۵	۱۳	۱۲	۸	۱۴	۱۴	۱۵	۱۴	۱۲
۱۹	۱۵	۱۲	۶	۹	۱۴	۱۶	۱۷	۵	۱۱
۱۷	۲۰	۱۳	۱۹	۱۸	۱۵	۱۰	۱۱	۷	۱۳

فرض $H_0: \mu = 15$ را در مقابل $H_1: \mu \neq 15$ با اطمینان ۹۵ درصد آزمون کنید.

حل:





براساس مسیر ارائه شده در تصاویر فوق، خروجی زیر به دست آمده است که چون مقدار معنی‌داری آزمون $sig = 0/012 < 0/05$ فرض صفر رد می‌شود. یعنی $H_1: \mu \neq 15$ تأیید می‌شود.

لازم به ذکر است مقدار آماره t-student، ۲٫۶۶۹– با درجهٔ آزادی ۲۹ است.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
hazine jari	۳۰	۱۳٫۱۳	۳٫۸۳۰	.۶۹۹

One-Sample Test

	Test Value = ۱۵					
	t	df	Sig. (۲-tailed)	Mean Difference	۹۵٪ Confidence Interval of the Difference	
					Lower	Upper
hazine jari	-۲٫۶۶۹	۲۹	.۰۱۲	-۱٫۸۶۷	-۳٫۳۰	-.۴۴

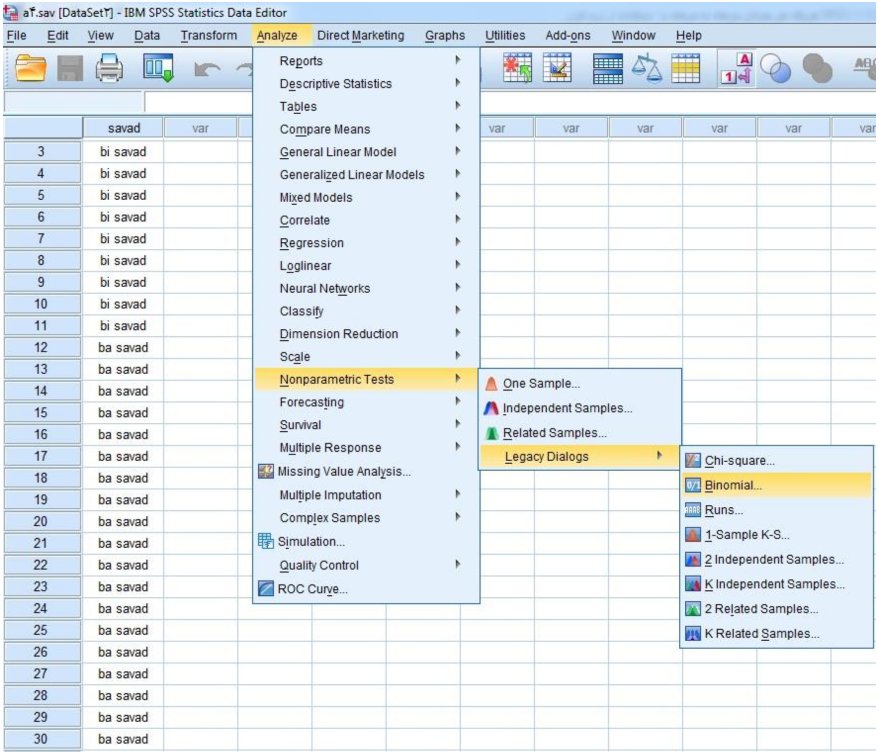
٤ – ادعا می‌شود که بیش از ۸۰ درصد از جمعیت شهری با سواد است. برای بررسی این ادعا اطلاعات زیر به دست آمده است. (کد ۱ دال بر باسواد بودن):

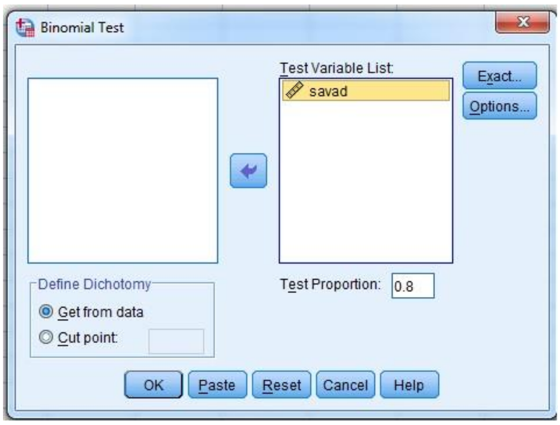
۱ ۰ ۱ ۱ ۰ ۰ ۰ ۱ ۱ ۱ ۱ ۰ ۱ ۰ ۱

۰ ۱ ۰ ۰ ۱ ۱ ۱ ۰ ۱ ۰ ۱ ۱ ۱ ۱ ۱

فرض $H_0: P = 0/8$ را در مقابل $H_1: P > 0/8$ با اطمینان ۹۵ درصد آزمون کنید.

حل:





براساس مسیر فوق مقدار $P=\frac{19}{30}=0/6$ به دست آمده است. چون $sig=0<0/05$ فرض مخالف رد می‌شود؛ یعنی $H_1: P<0/8$ تأیید می‌شود.

Binomial Test

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (۱-tailed)
Group ۱	bi savad	۱۱	.٤	.۸	.۰۰۰[a]
savad Group ۲	ba savad	۱۹	.٦		
Total		۳۰	۱,۰		

a. Alternative hypothesis states that the proportion of cases in the first group < .۸.

۵- برای بررسی رابطه بین میزان تأخیر و نمرهٔ کارایی کارکنان در بخش امانت کتابخانه‌های کشور اطلاعات زیر به دست آمده است:

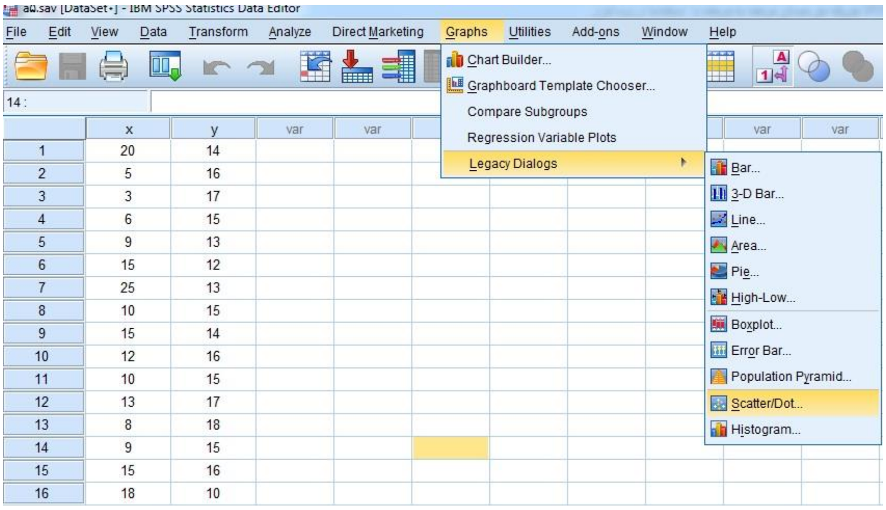
میزان تأخیر به دقیقه (X)	۱۸	۱۵	۹	۸	۱۳	۱۰	۱۲	۱۵	۱۰	۲۵	۱۵	۹	٦	۳	۵	۲۰
نمرهٔ کارایی (Y)	۱۰	۱٦	۱۵	۱۸	۱۷	۱۵	۱٦	۱٤	۱۵	۱۳	۱۲	۱۳	۱۵	۱۷	۱٦	۱٤

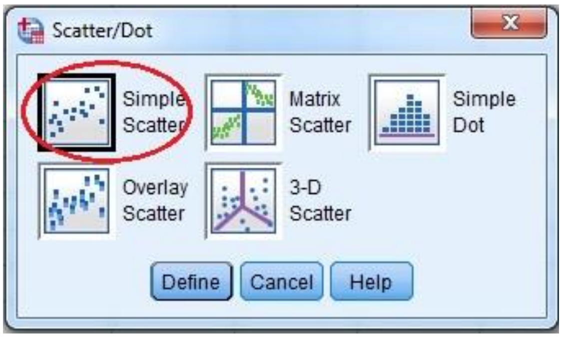
الف- نمودار پراکنش را رسم کنید.

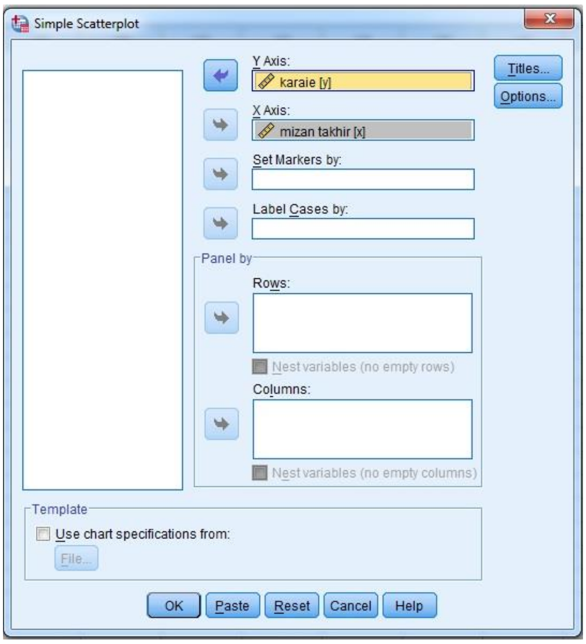
ب- ضریب همبستگی پیرسون را محاسبهٔ کنید.

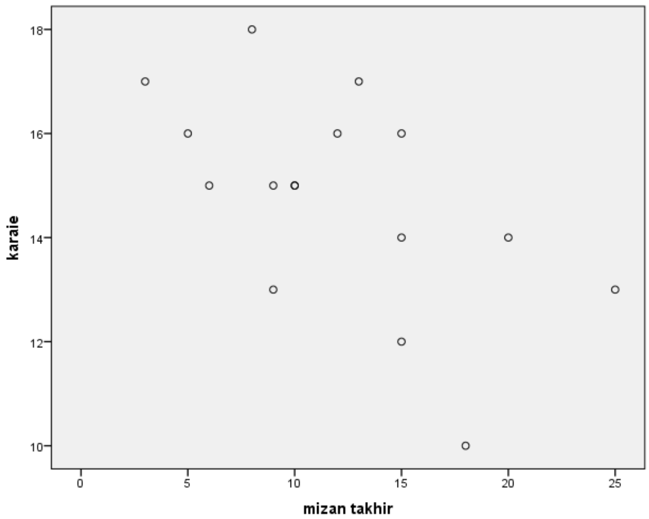
ج- ضریب همبستگی اسپیرمن را محاسبهٔ کنید.

حل:

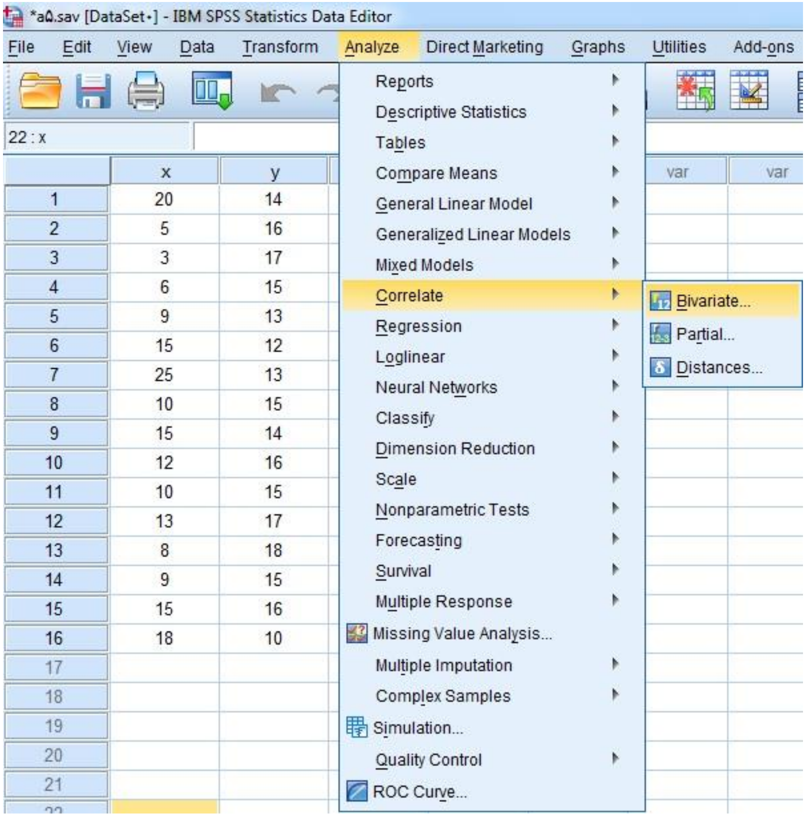

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Chart Builder...
Graphboard Template Chooser...
Compare Subgroups
Regression Variable Plots
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
x y
1 20 14
2 5 16
3 3 17
4 6 15
5 9 13
6 15 12
7 25 13
8 10 15
9 15 14
10 12 16
11 10 15
12 13 17
13 8 18
14 9 15
15 15 16
16 18 10

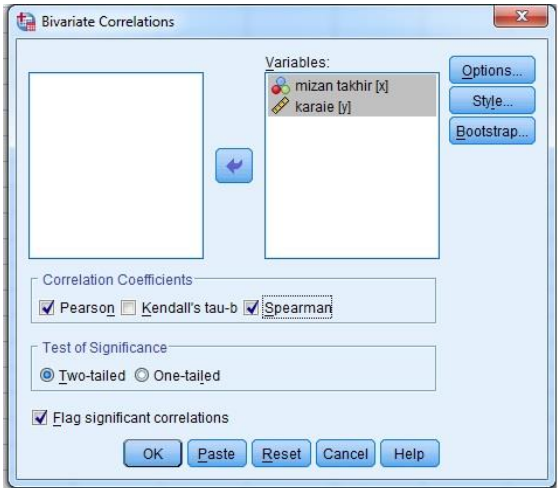

Scatter/Dot
Simple Scatter
Matrix Scatter
Simple Dot
Overlay Scatter
3-D Scatter
Define
Cancel
Help





براساس اطلاعات مسئله و نمودار رسم شدهٔ فوق، همبستگی بین دو متغیر میزان تأخیر و کارایی منفی است. در ادامه مقدار ضریب همبستگی پیرسون و اسپیرمن را محاسبهٔ می‌کنیم.


*a۵.sav [DataSet۰] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons
22 : x
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Bivariate...
Partial...
Distances...


Bivariate Correlations
Variables:
mizan takhir [x]
karaie [y]
Options...
Style...
Bootstrap...
Correlation Coefficients
Pearson Kendall's tau-b Spearman
Test of Significance
Two-tailed One-tailed
Flag significant correlations
OK Paste Reset Cancel Help

مقدار ضریب همبستگی پیرسون ۰/۵۷- است و چون مقدار معنی‌داری برای آن $sig = 0/021 < 0/05$ است، معنی‌دار بودن همبستگی منفی بین دو متغیر میزان کارایی و تأخیر تأیید می‌شود. در مورد ضریب همبستگی اسپیرمن مقدار ضریب همبستگی ۰/۵۸۳- به دست آمده است و چون مقدار معنی‌داری برای آن $sig = 0/015 < 0/05$ مانند ضریب همبستگی پیرسون، وجود همبستگی منفی بین دو متغیر تأیید می‌شود.

Correlations

		mizan takhir	karaie
mizan takhir	Pearson Correlation	۱	-.۵۷۰*
	Sig. (۲-tailed)		.۰۲۱
	N	۱۶	۱۶
karaie	Pearson Correlation	-.۵۷۰*	۱
	Sig. (۲-tailed)	.۰۲۱	
	N	۱۶	۱۶

*. Correlation is significant at the ۰,۰۵ level (۲-tailed).

Nonparametric Correlations

Correlations

			mizan takhir	karaie
Spearman's rho	mizan takhir	Correlation Coefficient	۱,۰۰۰	-.۵۸۳*
		Sig. (۲-tailed)	.	.۰۱۸
		N	۱۶	۱۶
	karaie	Correlation Coefficient	-.۵۸۳*	۱,۰۰۰
		Sig. (۲-tailed)	.۰۱۸	.
		N	۱۶	۱۶

*. Correlation is significant at the ۰,۰۵ level (۲-tailed).

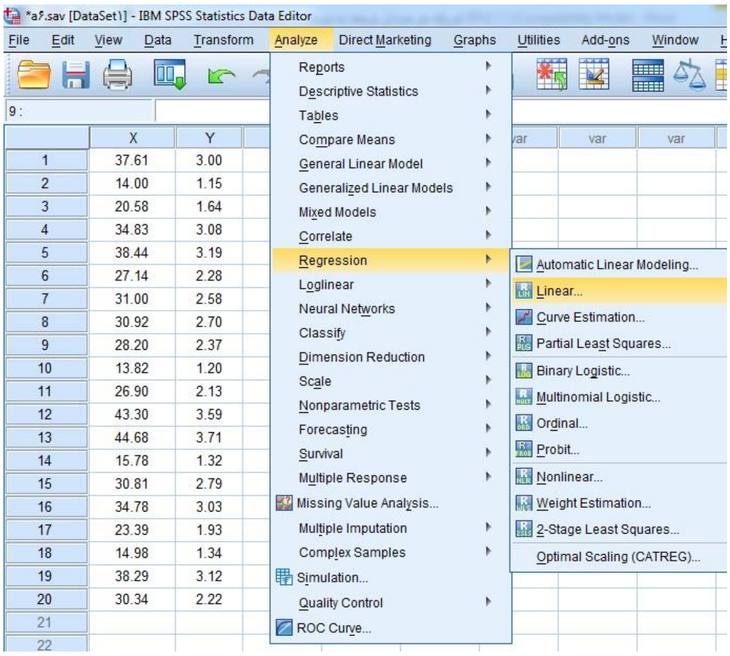
۶- درآمد (X) و هزینهٔ اجاره‌ای (Y) بیست مرکز آموزشی به صورت زیر داده شده است:

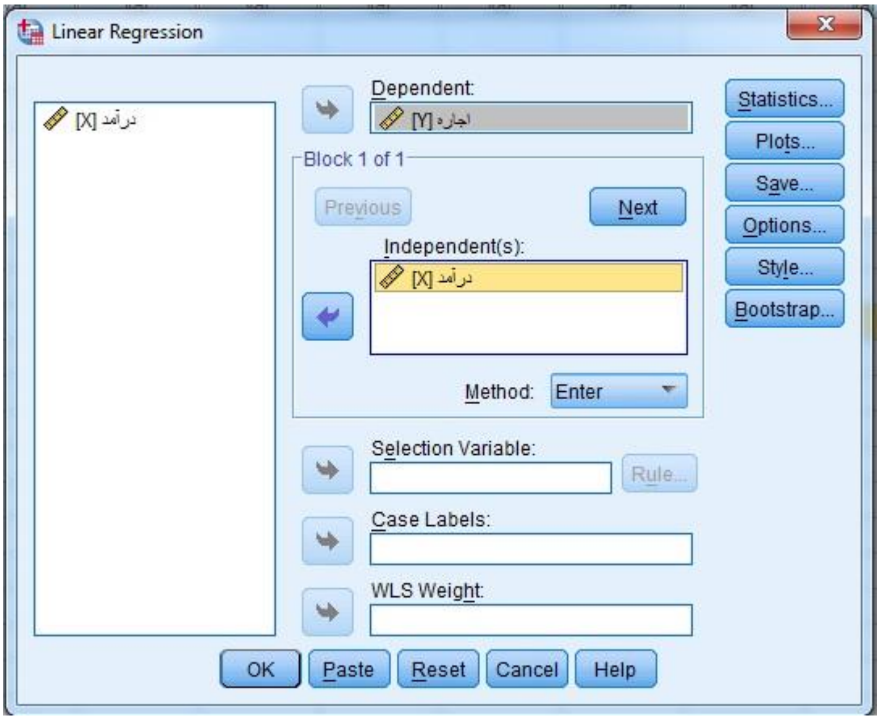
X	Y	X	Y
۲۶/۹	۲/۱۳	۳۷/۶۱	۳
۴۳/۳	۳/۵۹	۱۴	۱/۱۵
۴۴/۶۸	۳/۷۱	۲۰/۵۸	۱/۶۴
۱۵/۷۸	۱/۳۲	۳۴/۸۳	۳/۰۸
۳۰/۸۱	۲/۷۹	۳۸/۴۴	۳/۱۹
۳۴/۷۸	۳/۰۳	۲۷/۱۴	۲/۲۸
۲۳/۳۹	۱/۹۳	۳۱	۲/۵۸
۱۴/۹۸	۱/۳۴	۳۰/۹۲	۲/۷
۳۸/۲۹	۳/۱۲	۲۸/۲	۲/۳۷
۳۰/۳۴	۲/۲۲	۱۳/۸۲	۱/۲

الف– معادلهٔ خط برازش را به دست آورید.

ب– برای درآمد مرکز آموزشی ۲۵، مقدار اجاره را پیش‌بینی کنید.

حل:





براساس مسیر فوق، نتیجهٔ زیر به دست آمد که معادلهٔ خط زیر را نشان می‌دهد:

$$y = 0/025 + 0/083 \times x$$

البته مقدار ضریب تعیین مدل $R^2 = 0/978$ است و مقدار معنی‌داری مدل براساس جدول ANOVA، $sig = 0 < 0/05$ که نشاندهندهٔ معنی‌داری مدل است، برای درآمد ۲۵ مقدار اجاره ۲/۱ به دست می‌آید.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
۱	درآمد[b]	.	Enter

a. Dependent Variable: اجاره

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
۱	.۹۸۹[a]	.۹۷۸	.۹۷۷	.۱۲۰۵۲

a. Predictors: (Constant), درآمد

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
۱	Regression	۱۱,۷۱۲	۱	۱۱,۷۱۲	۸۰۶,۳۲۰	.۰۰۰[b]
	Residual	.۲۶۱	۱۸	.۰۱۵		
	Total	۱۱,۹۷۳	۱۹			

a. Dependent Variable: اجاره

b. Predictors: (Constant), درآمد

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
۱	(Constant)	.۰۲۵	.۰۸۹		.۲۸۰	.۷۸۳
	درآمد	.۰۸۳	.۰۰۳	.۹۸۹	۲۸,۳۹۶	.۰۰۰

a. Dependent Variable: اجاره

پیوست ۲: پاسخ خودآزمایی‌ها

فصل اول

۱. الف) جامعه ب) نمونه ج) نمونه د) جامعه هـ)نمونه و) جامعه

۲. زمان: متغیر پیوسته و مقیاس اندازه‌گیری آن مقیاس سنی است.
درآمد: متغیر پیوسته و مقیاس اندازه‌گیری آن مقیاس نسبی است.
شغل: برای بیان آن از مقیاس اسمی استفاده می‌شود.
نژاد: برای بیان آن از مقیاس طبقه‌ای استفاده می‌شود.

۳. به عهدهٔ دانشجو

٤. به عهدهٔ دانشجو

۵. نمونه‌گیری تصادفی ساده، نمونه‌گیری منظم، نمونه‌گیری طبقه‌ای، نمونه‌گیری خوشه‌ای

٦. الف) جامعهٔ محدود ب) جامعهٔ محدود ج) جامعهٔ نامحدود

۷. نمونه‌گیری منظم

۸. تعداد کتب چاپ شده در سال ۱۳۹۵ در کشور که یک جامعه آماری است، انتخاب تعداد کتب با فروش بیشتر از ۱۰/۰۰۰ با استفاده از نمونه‌گیری طبقه‌ای امکان‌پذیر است.

فصل دوم

۱. $\sum_{i=۱}^{n} x_i \quad , \quad \sum_{i=۱}^{۱۲۵} i \quad , \quad \sum_{i=-۳}^{۱۰۰} i$

۲. الف) ۳ ب) فاقد نما

۳.

حروف	A	B	C	D	E
تعداد تکرار	۵	٤	٤	٤	۳

چون حرف A بیشتر از بقیه تکرار شده، نما حرف A است.

٤. الف) صفر ب) مرتب

۱٦ ۲۱ ۲۹ ۳۳ ۷۵ ۸۲ ۸٦ ۹۵ ۱۰۵ ۱۵۵

$$\text{میانه} = \frac{۷۵+۸۲}{۲} = ۷۸/۵$$

۵. اگر m میانهٔ مشاهدات $x_۱, x_۲, \ldots, x_n$ باشد، آنگاه

$$\text{میانهٔ جدید} = ۵m+۷$$

٦. $$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i = \frac{۱}{۷}\sum_{i=۱}^{۷} x_i = \frac{۱}{۷}(۲۹/۷۵) = ٤/۲۵$$

۷. $$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{k} f_i x_i = \frac{۱}{۲۵}\sum_{i=۱}^{٦} f_i x_i = \frac{۱}{۲۵}(f_۱ x_۱ + f_۲ x_۲ + \ldots + f_٦ x_٦)$$

$$= \frac{۱}{۲۵}\left(٤(۲/۷۵)+٦(۱/۲۵)+۲(۳/۵)+۵(۱/۵)+۳(۲/۵)+۵(٤/۷۵)\right) = \frac{۱}{۲۵}(٦٤/۲۵) = ۲/۵۷$$

۸. الف) ٦ ب) ۲- ج) ۱۱

۹. الف) دامنه: $۳-۱/۱۵ = ۱/۸۵$

ب) $\bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i = \frac{۱}{۱۲}\sum_{i=۱}^{۱۲} x_i = ۱/۸۱۷$

ج) $S^۲ = \frac{۱}{n}\sum_{i=۱}^{n} (x_i - \bar{x})^۲ = ۰/۲۹۸$ $\qquad S = ۰/۵٤۵$

۱۰. الف) $300+15=315$

ب) چون اضافه و کم کردن در انحراف معیار و واریانس تأثیری ندارد، لذا ۶/۲۵ و ۲/۵

۱۱.

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2+x^{-2}-2x^{-}x_i\right)$$

$$=\frac{1}{n}\sum_{i=1}^{n}x_i^2+x^{-2}-\frac{2x^{-}}{n}\sum x_i=\frac{1}{n}\sum_{i=1}^{n}x_i^2+x^{-2}-2x^{-2}=\frac{1}{n}\sum x^2-x^{-2}$$

۱۲. میانگین، واریانس و ضریب تغییر سال اول به ترتیب برابر است با:

$$\bar{x}_1=245/83 \quad , \quad S_1^2=543/47 \quad , \quad S_1=21/395 \quad , \quad cv_1=\frac{21/295}{245/83}=0/087$$

میانگین، واریانس و ضریب تغییر سال دوم برابر است با:

$$\bar{x}_2=246/625 \quad , \quad S_2^2=970/984 \quad , \quad S_2=31/161 \quad , \quad cv_1=\frac{31/161}{246/625}=0/126$$

چون ضریب تغییر cv_1 کم‌تر از cv_2 است، تغییرات قیمت در سال اول کم‌تر از سال دوم است.

۱۳. نما، میانه، انواع میانگین و دامنه، واریانس، انحراف معیار و ضریب تغییر

۱۴. کتابخانهٔ مرکزی: نما ندارد میانه: ۴/۵

کتابخانهٔ ارشاد: نما ندارد میانه ۴/۵

۱۵.

$$\bar{x}=\frac{1}{4}\sum_{i=1}^{n}x_i=\frac{1}{4}\left(2/5+1/5+2+4\right)=2/5$$

$2/5\times1000=2500$

۱۶.

سال	اول	دوم	سوم	چهارم	پنجم
رشته	۳	۲	۴	۱۰	۱۵

رشتهٔ متوسط $=\sqrt[5]{3\times2\times4\times10\times15}=\sqrt[5]{3600}=5/144$

۱۷. همانند سؤال ۱۶ حل شود.

۱۸. به عهدهٔ دانشجو

۱۹. الف) $۴ = ۱ - ۵ =$ دامنه

ب) $\bar{x} = \frac{۱}{۵}(۲+۴+۱+۳+۵) = ۳$

ج) $= \frac{۱}{۵}(۱+۱+۴+۰+۴) = ۲ \quad , \quad S = \sqrt{۲}$

۲۰. ثابت می‌ماند.

۲۱. همانند سؤال ۱۲ حل شود.

۲۲. به عهدهٔ دانشجو

فصل سوم

.۱

$$E(X) = \sum f(x) = \frac{۱}{۶}(۱+۲+۳+۴+۵+۶) = \frac{۲۱}{۶} = ۳/۵$$

$$E(X^۲) = \sum x^۲ f(x) = \frac{۱}{۶}(۱+۴+۹+۱۶+۲۵+۳۶) = \frac{۹۱}{۶}$$

$$\sigma^۲ = E(X^۲) - (E(X))^۲ = \frac{۹۱}{۶} - (۳/۵)^۲ = ۲/۹۱۶۷$$

.۲

$$E(X) = \sum xf(x) = \frac{۱}{۴}$$

$$E(X^۲) = \sum x^۲ f(x) = \frac{۳}{۴}$$

$$\sigma^۲ = E(X^۲) - (E(X))^۲ = \frac{۳}{۴} - \left(\frac{۱}{۴}\right) = \frac{۹}{۱۶} - \frac{۱}{۱۶} = \frac{۸}{۱۶} = ۰/۵$$

$$E(X^۲ + X + ۱) = E(X^۲) + E(X) + ۱ = \frac{۳}{۴} + \frac{۱}{۴} + ۱ = \frac{۸}{۴} = ۲$$

.۳

x	۰	۲	۴	۵	۶	۷
$f(x)$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$

$$E(X)=\sum x\,f(x)=\frac{۱}{۶}(۰+۲+۴+۵+۶+۷)=۴$$

$$\mu=E(X^۲)=\sum x^۲ f(x)=\frac{۱}{۶}(۰+۴+۱۶+۲۵+۳۶+۴۹)=۲۱/۶۷$$

$$\sigma^۲=E(X^۲)-(E(X))^۲=۲۱/۶۷-(۴)^۲=۲۱/۶۷-۱۶=۵/۶۷$$

.۴ $$E(X)=\sum x\,f(x)=(۱)\left(\frac{۱}{۳}\right)+(۲)\left(\frac{۱}{۳}\right)+(۳)\left(\frac{۱}{۳}\right)=۲$$

$$E(X^۲)=\sum x^۲ f(x)=\frac{۱}{۳}(۱+۴+۹)=\frac{۱۴}{۳}$$

$$\sigma^۲=E(X^۲)-(E(X))^۲=\frac{۱۴}{۳}-۴=\frac{۱۴-۱۲}{۳}=\frac{۲}{۳}$$

.۵ الف) $k=۰/۲$ (چون $\sum f_i=۱$)

$$F(x)=P(X\le x)=\begin{cases}۰ & x<-۱\\ ۰/۲ & -۱\le x<۰\\ ۰/۵ & ۰\le x<۱\\ ۰/۸ & ۱\le x<۲\\ ۱ & x\ge ۲\end{cases}$$

.۶

x	۰	۱
$f(x)$	$۱-P$	P

$$E(X)=\sum x\,f(x)=P$$

$$E(X^۲)=\sum x^۲ f(x)=P$$

$$\sigma^۲=E(X^۲)-(E(X))^۲=P-P^۲=P(۱-P)$$

.۷ به عهدهٔ دانشجو

.۸ $\lambda=۲$

$$P(X=x)=f(x)=\frac{e^{-\lambda}\lambda^x}{x!}=\frac{e^{-۲}(۲)^x}{x!}\qquad,\quad x=۰,۱,۲,\ldots$$

$$P(X=۰)=f(۰)=\frac{e^{-۲}(۲)^۰}{۰!}=۰/۰۳۶$$

$$P\left(X \geq ۱۰\right) = \sum_{x=۱۰}^{\infty} \frac{e^{-۲}\left(۲\right)^{x}}{x!} = e^{-۲}\left[\frac{\left(۲\right)^{۱۰}}{۱۰!} + \frac{\left(۲\right)^{۱۱}}{۱۱!} + \ldots\right]$$

$$P\left(X \leq ۱۰۰\right) = \sum_{x=۰}^{۱۰۰} \frac{e^{-۲}\left(۲\right)^{x}}{x!} = e^{-۲}\left[۱ + \frac{۲}{۱} + \frac{٤}{۲۱} + \ldots + \frac{\left(۲\right)^{۱۰۰}}{۱۰۰!}\right]$$

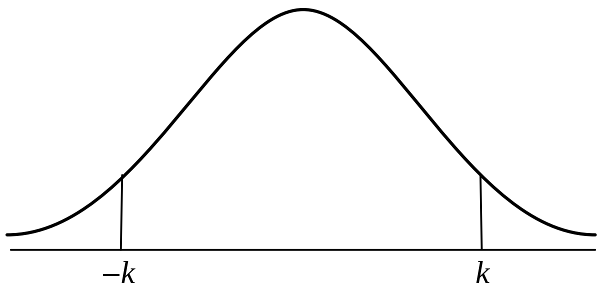
۹. به عهدهٔ دانشجو

۱۰. به عهدهٔ دانشجو

۱۱. به عهدهٔ دانشجو

۱۲.

$$\begin{aligned} P\left(-k < z < k\right) &= P\left(z < k\right) - P\left(z < -k\right) \quad , \quad P\left(z < -k\right) = a \\ &= P\left(z < k\right) - a \\ &= ۱ - P\left(z > k\right) - a \\ &= ۱ - a - a = ۱ - ۲a \end{aligned}$$



۱۳.

$$X \sim N\left(۳۵, ۱۴۴\right)$$

$$P\left(X > ۵۵\right) = P\left(\frac{X - \mu}{\sigma} > \frac{۵۵ - ۳۵}{۱۲}\right) = P\left(z > \frac{۲۰}{۱۲}\right) = P\left(z > ۱/۶۷\right) \approx ۰/۰۵$$

یا حدوداً ۵ درصد

فصل چهارم

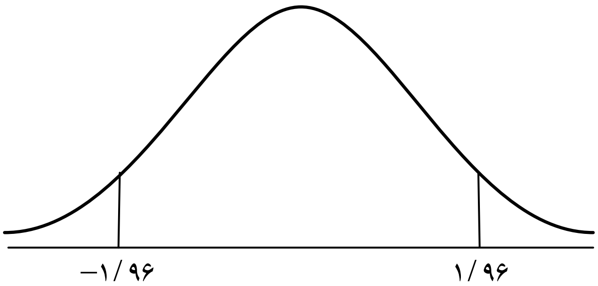
۱. به عهدهٔ دانشجو

۲. به عهدهٔ دانشجو

۳.
$$\begin{vmatrix} H_0 : \mu = ۶۰ \\ H_1 : \mu \neq ۶۰ \end{vmatrix}$$

$$n = ۲۵ \quad , \quad \bar{x} = ۶۲ \quad , \quad \sigma = ۵ \quad , \quad ۱ - \alpha = ۰/۹۵ \quad , \quad \alpha = ۰/۰۵ \quad , \quad \frac{\alpha}{۲} = ۰/۰۲۵$$

$$z_{۰/۰۲۵} = ۱/۹۶ \qquad z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{۶۲ - ۶۰}{۵ / \sqrt{۲۵}} = ۲$$



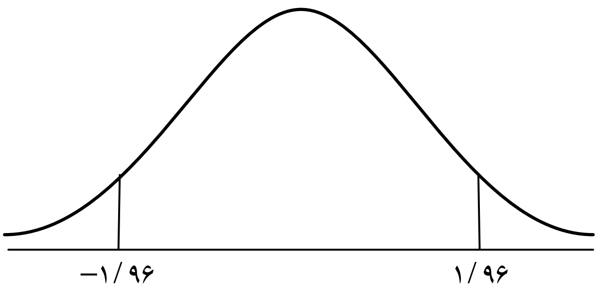
چون ۲>۱/۹۶، فرض H_0 رد می‌شود.

۴.
$$\begin{vmatrix} H_0 : \mu = ۱۲ \\ H_1 : \mu \neq ۱۲ \end{vmatrix}$$

$$n = ۱۶ \quad , \quad \bar{x} = ۱۳/۷۵ \quad , \quad \sigma^2 = ۱/۶۹ \quad , \quad ۱ - \alpha = ۰/۹۵$$

$$\alpha = ۰/۰۵ \quad , \quad Z_{۰/۰۵} = ۱/۹۶$$

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{۱۳/۷۵ - ۱۲}{\frac{۱/۳}{۴}} = ۵/۳۸۵$$

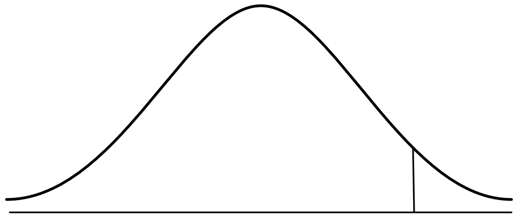


چون ۵/۳۸۵>۱/۹۶، فرض H_0 رد می‌شود.

.۵
$$\begin{vmatrix} H_{\circ}: \mu = ۱۵ \\ H_{۱}: \mu > ۱۵ \end{vmatrix}$$

$$n = ۹ \quad , \quad \bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i = ۱۶ \quad , \quad \sigma^{۲} = ۶/۲۵ \quad , \quad \alpha = ۰/۰۵$$

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{۱۶ - ۱۵}{\frac{۲/۵}{۳}} = \frac{۳}{۲/۵} = ۱/۲$$



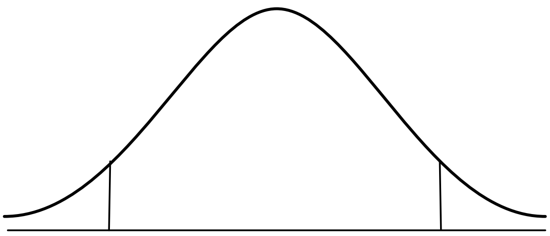
چون Z=۱/۲ <۱/٦٤٥، فرض $H_{\circ}$ رد می‌شود.

٦. به عهدهٔ دانشجو

.۷
$$\begin{vmatrix} H_{\circ}: p = ۰/۲ \\ H_{۱}: p \neq ۰/۲ \end{vmatrix}$$

$$\alpha = ۰/۰۵ \quad , \quad n = ۴۰۰ \quad , \quad f = ۳۰۰ \quad , \quad p = \frac{f}{n} = \frac{۳۰۰}{۴۰۰} = ۰/۷۵$$

$$z = \frac{p - ۰/۲}{\sqrt{\frac{p(۱-p)}{n}}} = \frac{۰/۷۵ - ۰/۲}{\sqrt{\frac{۰/۷۵ \times ۰/۲۵}{۴۰۰}}} = ۲۵/۴$$



فرض $H_{\circ}$ رد می‌شود.

۸. به عهدهٔ دانشجو

۹. به عهدهٔ دانشجو

۱۰. به عهدهٔ دانشجو

فصل پنجم

۱. به عهدهٔ دانشجو

۲. الف)
$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=\sum_{i=1}^{n}\left[X_i\left(Y_i-\bar{Y}\right)-\bar{X}\left(Y_i-\bar{Y}\right)\right]$$
$$\sum_{i=1}^{n}X_i\left(Y_i-\bar{Y}\right)-\bar{X}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)$$
$$\sum_{i=1}^{n}X_i\left(Y_i-\bar{Y}\right)-\bar{X}\left(Y_i-\bar{Y}\right)=\sum X_i\left(Y_i-\bar{Y}\right)$$
$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\sum_{i=1}^{n}\left(X_i^2+\bar{X}^2-2\bar{X}X_i\right)=\sum X_i^2+n\bar{X}^2-2n\bar{X}^2$$
$$=\sum X_i^2+n\bar{X}^2$$

ب) با توجه به قسمت الف به راحتی نتیجه می‌شود.

۳.
$$r=\frac{\sum\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum\left(X_i-\bar{X}\right)^2\sum\left(Y_i-\bar{Y}\right)^2}}=\frac{8}{\sqrt{32\times 42}}=0/218$$

همبستگی مثبت است.

۴.

X_i	Y_i	X_iY_i	$\left(X_i-\bar{X}\right)^2$	$\left(Y_i-\bar{Y}\right)^2$
۱۰	۱۶	۱۶۰	۰	٤
۱۳	۱۸	۲۳٤	۹	۱٦
۷	۱۲	۸٤	۹	٤
۱۲	۱٤	۱٦۸	٤	۰
۸	۱۰	۸۰	٤	۱٦
۵۰	۷۰	۷۲٦	۲٦	٤۰

$$n=۵ \quad , \quad \bar{X}=۱۰ \quad , \quad \bar{Y}=۱۴$$

$$r=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^۲\sum(Y_i-\bar{Y})^۲}}=\frac{\sum X_iY_i-n\bar{X}\bar{Y}}{\sqrt{\sum(X_i-\bar{X})^۲\sum(Y_i-\bar{Y})^۲}}$$

$$=\frac{۷۲۶-۵(۱۰)(۱۴)}{\sqrt{۲۶\times ۴۰}}=\frac{۲۶}{\sqrt{۲۶\times ۴۰}}=۰/۸۱$$

ضریب تعیین $r^۲=۰/۶۵۶۱$

۵. به عهدهٔ دانشجو

٦. به عهدهٔ دانشجو

۷. $$r_s=۱-\frac{۶\sum d_i^۲}{n(n^۲-۱)}=۱-\frac{۶(۷۲)}{۹(۸۱-۱)}=۱-\frac{۴۳۲}{۷۲۰}=۱-۰/۶=۰/۴$$

۸. به عهدهٔ دانشجو

۹.

X	۱-	۰	۲	۳	۵
Y	۱-	۱	۵	۷	۱۱

با توجه به معادلهٔ $Y=۲X+۱$، Yها محاسبهٔ می‌شوند.

رسم نمودار به عهدهٔ دانشجو است.

X	Y	XY
۱-	۱-	۱
۰	۱	۰
۲	۵	۱۰
۳	۷	۲۱
۵	۱۱	۵۵
۹	۲۳	۸۷

$$n=۵ \quad , \quad \bar{X}=۱/۸ \quad , \quad \bar{Y}=۴/۶ \quad , \quad r=۱$$

.۱۰

x	y	xy	$(x_i-\bar{x})^۲$	$(y_i-\bar{y})^۲$
۱	۰	۰	۲/۲٥	٤۲/۲٥
۲	۳	٦	۰/۲٥	۱۲/۲٥
۳	۸	۲٤	۰/۲٥	۲/۲٥
٤	۱٥	٦۰	۲/۲٥	۷۲/۲٥
۱۰	۲٦	۹۰	٥	۱۲۹

$n=۴ \quad , \quad \bar{x}=۲/۵ \quad , \quad \bar{y}=۶/۵$

$$r=\frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum(x_i-\bar{x})^۲\sum(y_i-\bar{y})^۲}}=\frac{۹۰-۴(۲/۵)(۶/۵)}{\sqrt{۵\times۱۲۹}}=۰/۹۸۴$$

برای به دست آوردن ضریب همبستگی اسپیرمن می‌توان نوشت:

رتبه X	۱	۲	۳	٤
Y:	۰	۳	۸	۱٥
رتبه Y	۱	۲	۳	٤
ریشه X_iها	۱	۲	۳	٤
رتبه Y_iها	۱	۲	۳	٤
d_i :	۰	۰	۰	۰

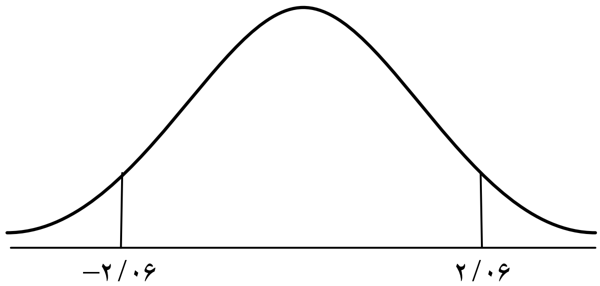
$$r_s=۱-\frac{۶\sum d_i^۲}{n(n^۲-۱)}=۱-\frac{۶(۰)}{۴(۱۶-۱)}=۱$$

چون رابطه بین X و Y برقرار است همبستگی قوی است.

.۱۱
$$\begin{cases} H_۰:\rho=۰ \\ H_۱:\rho\neq۰ \end{cases}$$

$n=۲۷ \quad , \quad df=۲۵ \quad , \quad r=۰/۸ \quad , \quad \mathrm{t}(۰/۰۲۵,۲۵)=۲/۰۶$

$$T=\frac{r\sqrt{n-۲}}{\sqrt{۱-r^۲}}=\frac{۰/۸\sqrt{۲۵}}{\sqrt{۱-۰/۶۴}}=\frac{۰/۸\times۵}{۰/۶}=۶/۶۷$$

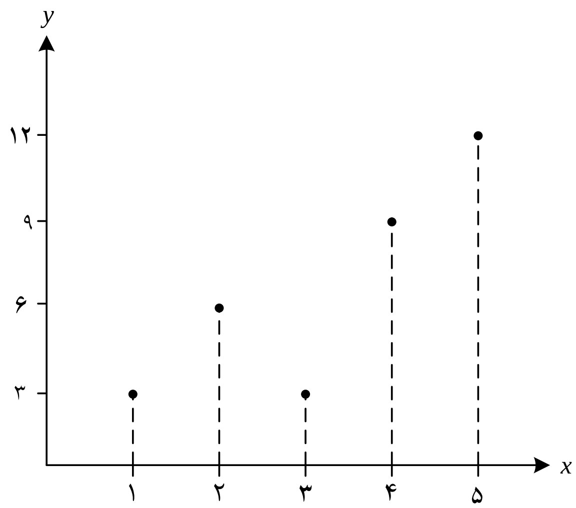


فرض $H_۰$ رد می‌شود.

فصل ششم

۱. به عهدهٔ دانشجو

۲.



۳. با توجه به مدل خط برازش $y = ۳ + ۰/۲x$ داریم

برای $x = ۵$، $y = ۳ + ۰/۲(۵) = ۴$

برای $x = ۷$، $y = ۳ + ۰/۲(۷) = ۴/۴$

٤. به عهدهٔ دانشجو

۵. به عهدهٔ دانشجو

.٦

x	y	xy	$(x_i-\bar{x})^۲$
٩٥	١٥	١٤٢٥	٢٥
٨٥	١٤	١١٩٠	٢٥
١٠٠	١٨	١٨٠٠	١٠٠
٨٠	١٢	٩٦٠	١٠٠
٩٠	١٦	١٤٤٠	
٤٥٠	٧٥	٦٨١٥	٢٥٠

$$n=۵ \quad , \quad \bar{x}=۹۰ \quad , \quad \bar{y}=۱۵$$

$$b=\frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum (x_i-\bar{x})^۲}=\frac{۶۸۱۵-۵(۹۰)(۱۵)}{۲۵۰}=۰/۲۶$$

$$a=\bar{y}-b\bar{x}=۱۵-۰/۲۶(۹۰)=-۸/۴$$

$$\hat{y}_i=a+bx_i \quad , \quad \hat{y}_i=-۸/۴+۰/۲۶x_i$$

.٧ الف) $\hat{y}_i=-۴/۸+۰/۲۲x_i$

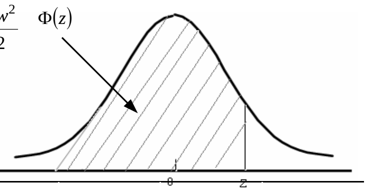
ب) $\hat{y}_i=۵/۴۸+۰/۶۳۴x_i$

پیوست ۳: جداول آماری

جدول۱. توزیع نرمال استاندارد

$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-w^2}{2}}$$

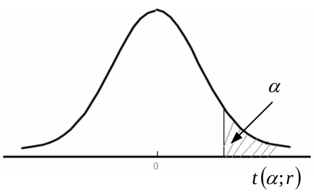
$$\Phi(-z) = 1 - \Phi(z)$$



z	۰,۰۰	۰,۰۱	۰,۰۲	۰,۰۳	۰,۰٤	۰,۰٥	۰,۰٦	۰,۰۷	۰,۰۸	۰,۰۹

٠,٠	٠,٥٠٠٠	٠,٥٠٤٠	٠,٥٠٨٠	٠,٥١٢٠	٠,٥١٦٠	٠,٥١٩٩	٠,٥٢٣٩	٠,٥٢٧٩	٠,٥٣١٩	٠,٥٣٥٩
٠,١	٠,٥٣٩٨	٠,٥٤٣٨	٠,٥٤٧٨	٠,٥٥١٧	٠,٥٥٥٧	٠,٥٥٩٦	٠,٥٦٣٦	٠,٥٦٧٥	٠,٥٧١٤	٠,٥٧٥٣
٠,٢	٠,٥٧٩٣	٠,٥٨٣٢	٠,٥٨٧١	٠,٥٩١٠	٠,٥٩٤٨	٠,٥٩٨٧	٠,٦٠٢٦	٠,٦٠٦٤	٠,٦١٠٣	٠,٦١٤١
٠,٣	٠,٦١٧٩	٠,٦٢١٧	٠,٦٢٥٥	٠,٦٢٩٣	٠,٦٣٣١	٠,٦٣٦٨	٠,٦٤٠٦	٠,٦٤٤٣	٠,٦٤٨٠	٠,٦٥١٧
٠,٤	٠,٦٥٥٤	٠,٦٥٩١	٠,٦٦٢٨	٠,٦٦٦٤	٠,٦٧٠٠	٠,٦٧٣٦	٠,٦٧٧٢	٠,٦٨٠٨	٠,٦٨٤٤	٠,٦٨٧٩
٠,٥	٠,٦٩١٥	٠,٦٩٥٠	٠,٦٩٨٥	٠,٧٠١٩	٠,٧٠٥٤	٠,٧٠٨٨	٠,٧١٢٣	٠,٧١٥٧	٠,٧١٩٠	٠,٧٢٢٤
٠,٦	٠,٧٢٥٧	٠,٧٢٩١	٠,٧٣٢٤	٠,٧٣٥٧	٠,٧٣٨٩	٠,٧٤٢٢	٠,٧٤٥٤	٠,٧٤٨٦	٠,٧٥١٧	٠,٧٥٤٩
٠,٧	٠,٧٥٨٠	٠,٧٦١١	٠,٧٦٤٢	٠,٧٦٧٣	٠,٧٧٠٣	٠,٧٧٣٤	٠,٧٧٦٤	٠,٧٧٩٤	٠,٧٨٢٣	٠,٧٨٥٢
٠,٨	٠,٧٨٨١	٠,٧٩١٠	٠,٧٩٣٩	٠,٧٩٦٧	٠,٧٩٩٥	٠,٨٠٢٣	٠,٨٠٥١	٠,٨٠٧٨	٠,٨١٠٦	٠,٨١٣٣
٠,٩	٠,٨١٥٩	٠,٨١٨٦	٠,٨٢١٢	٠,٨٢٣٨	٠,٨٢٦٤	٠,٨٢٨٩	٠,٨٣١٥	٠,٨٣٤٠	٠,٨٣٦٥	٠,٨٣٨٩
١,٠	٠,٨٤١٣	٠,٨٤٣٨	٠,٨٤٦١	٠,٨٤٨٥	٠,٨٥٠٨	٠,٨٥٣١	٠,٨٥٥٤	٠,٨٥٧٧	٠,٨٥٩٩	٠,٨٦٢١
١,١	٠,٨٦٤٣	٠,٨٦٦٥	٠,٨٦٨٦	٠,٨٧٠٨	٠,٨٧٢٩	٠,٨٧٤٩	٠,٨٧٧٠	٠,٨٧٩٠	٠,٨٨١٠	٠,٨٨٣٠
١,٢	٠,٨٨٤٩	٠,٨٨٦٩	٠,٨٨٨٨	٠,٨٩٠٧	٠,٨٩٢٥	٠,٨٩٤٤	٠,٨٩٦٢	٠,٨٩٨٠	٠,٨٩٩٧	٠,٩٠١٥
١,٣	٠,٩٠٣٢	٠,٩٠٤٩	٠,٩٠٦٦	٠,٩٠٨٢	٠,٩٠٩٩	٠,٩١١٥	٠,٩١٣١	٠,٩١٤٧	٠,٩١٦٢	٠,٩١٧٧
١,٤	٠,٩١٩٢	٠,٩٢٠٧	٠,٩٢٢٢	٠,٩٢٣٦	٠,٩٢٥١	٠,٩٢٦٥	٠,٩٢٧٩	٠,٩٢٩٢	٠,٩٣٠٦	٠,٩٣١٩
١,٥	٠,٩٣٣٢	٠,٩٣٤٥	٠,٩٣٥٧	٠,٩٣٧٠	٠,٩٣٨٢	٠,٩٣٩٤	٠,٩٤٠٦	٠,٩٤١٨	٠,٩٤٢٩	٠,٩٤٤١
١,٦	٠,٩٤٥٢	٠,٩٤٦٣	٠,٩٤٧٤	٠,٩٤٨٤	٠,٩٤٩٥	٠,٩٥٠٥	٠,٩٥١٥	٠,٩٥٢٥	٠,٩٥٣٥	٠,٩٥٤٥
١,٧	٠,٩٥٥٤	٠,٩٥٦٤	٠,٩٥٧٣	٠,٩٥٨٢	٠,٩٥٩١	٠,٩٥٩٩	٠,٩٦٠٨	٠,٩٦١٦	٠,٩٦٢٥	٠,٩٦٣٣
١,٨	٠,٩٦٤١	٠,٩٦٤٩	٠,٩٦٥٦	٠,٩٦٦٤	٠,٩٦٧١	٠,٩٦٧٨	٠,٩٦٨٦	٠,٩٦٩٣	٠,٩٦٩٩	٠,٩٧٠٦
١,٩	٠,٩٧١٣	٠,٩٧١٩	٠,٩٧٢٦	٠,٩٧٣٢	٠,٩٧٣٨	٠,٩٧٤٤	٠,٩٧٥٠	٠,٩٧٥٦	٠,٩٧٦١	٠,٩٧٦٧
٢,٠	٠,٩٧٧٢	٠,٩٧٧٨	٠,٩٧٨٣	٠,٩٧٨٨	٠,٩٧٩٣	٠,٩٧٩٨	٠,٩٨٠٣	٠,٩٨٠٨	٠,٩٨١٢	٠,٩٨١٧
٢,١	٠,٩٨٢١	٠,٩٨٢٦	٠,٩٨٣٠	٠,٩٨٣٤	٠,٩٨٣٨	٠,٩٨٤٢	٠,٩٨٤٦	٠,٩٨٥٠	٠,٩٨٥٤	٠,٩٨٥٧
٢,٢	٠,٩٨٦١	٠,٩٨٦٤	٠,٩٨٦٨	٠,٩٨٧١	٠,٩٨٧٥	٠,٩٨٧٨	٠,٩٨٨١	٠,٩٨٨٤	٠,٩٨٨٧	٠,٩٨٩٠
٢,٣	٠,٩٨٩٣	٠,٩٨٩٦	٠,٩٨٩٨	٠,٩٩٠١	٠,٩٩٠٤	٠,٩٩٠٦	٠,٩٩٠٩	٠,٩٩١١	٠,٩٩١٣	٠,٩٩١٦
٢,٤	٠,٩٩١٨	٠,٩٩٢٠	٠,٩٩٢٢	٠,٩٩٢٥	٠,٩٩٢٧	٠,٩٩٢٩	٠,٩٩٣١	٠,٩٩٣٢	٠,٩٩٣٤	٠,٩٩٣٦
٢,٥	٠,٩٩٣٨	٠,٩٩٤٠	٠,٩٩٤١	٠,٩٩٤٣	٠,٩٩٤٥	٠,٩٩٤٦	٠,٩٩٤٨	٠,٩٩٤٩	٠,٩٩٥١	٠,٩٩٥٢
٢,٦	٠,٩٩٥٣	٠,٩٩٥٥	٠,٩٩٥٦	٠,٩٩٥٧	٠,٩٩٥٩	٠,٩٩٦٠	٠,٩٩٦١	٠,٩٩٦٢	٠,٩٩٦٣	٠,٩٩٦٤
٢,٧	٠,٩٩٦٥	٠,٩٩٦٦	٠,٩٩٦٧	٠,٩٩٦٨	٠,٩٩٦٩	٠,٩٩٧٠	٠,٩٩٧١	٠,٩٩٧٢	٠,٩٩٧٣	٠,٩٩٧٤
٢,٨	٠,٩٩٧٤	٠,٩٩٧٥	٠,٩٩٧٦	٠,٩٩٧٧	٠,٩٩٧٧	٠,٩٩٧٨	٠,٩٩٧٩	٠,٩٩٧٩	٠,٩٩٨٠	٠,٩٩٨١
٢,٩	٠,٩٩٨١	٠,٩٩٨٢	٠,٩٩٨٢	٠,٩٩٨٣	٠,٩٩٨٤	٠,٩٩٨٤	٠,٩٩٨٥	٠,٩٩٨٥	٠,٩٩٨٦	٠,٩٩٨٦
٣,٠	٠,٩٩٨٧	٠,٩٩٨٧	٠,٩٩٨٧	٠,٩٩٨٨	٠,٩٩٨٨	٠,٩٩٨٩	٠,٩٩٨٩	٠,٩٩٨٩	٠,٩٩٩٠	٠,٩٩٩٠

جدول۲. توزیع استودنت



r	$\alpha = ۰{,}۱۰$	$\alpha = ۰{,}۰۵$	$\alpha = ۰{,}۰۲۵$	$\alpha = ۰{,}۰۱$	$\alpha = ۰{,}۰۰۵$

١	٣,٠٧٨	٦,٣١٤	١٢,٧٠٦	٣١,٨٢١	٦٣,٦٥٧
٢	١,٨٨٦	٢,٩٢٠	٤,٣٠٣	٦,٩٦٥	٩,٩٢٥
٣	١,٦٣٥	٢,٣٥٣	٣,١٨٢	٤,٥٤١	٥,٨٤١
٤	١,٥٣٣	٢,١٣٢	٢,٩٩٦	٣,٧٤٧	٤,٦٠٤
٥	١,٤٧٦	٢,٠١٥	٢,٥٧١	٣,٣٦٥	٤,٠٣٢
٦	١,٤٤٠	١,٩٤٣	٢,٤٤٧	٣,١٤٣	٣,٧٠٧
٧	١,٤١٥	١,٨٩٥	٢,٣٦٥	٢,٩٩٨	٣,٤٩٩
٨	١,٣٩٧	١,٨٦٠	٢,٣٠٦	٢,٨٩٦	٣,٣٥٥
٩	١,٣٨٣	١,٨٣٣	٢,٢٦٢	٢,٨٢١	٣,٢٥٠
١٠	١,٣٧٢	١,٨١٢	٢,٢٢٨	٢,٧٦٤	٣,١٦٩
١١	١,٣٦٣	١,٧٩٦	٢,٢٠١	٢,٧١٨	٣,١٠٦
١٢	١,٣٥٦	١,٧٨٢	٢,١٧٩	٢,٦٨١	٣,٠٥٥
١٣	١,٣٥٠	١,٧٧١	٢,١٦٠	٢,٦٥٠	٣,٠١٢
١٤	١,٣٤٥	١,٧٦١	٢,١٤٥	٢,٦٢٤	٢,٩٧٧
١٥	١,٣٤١	١,٧٥٣	٢,١٣١	٢,٦٠٢	٢,٩٤٧
١٦	١,٣٣٧	١,٧٤٦	٢,١٢٠	٢,٥٨٣	٢,٩٢١
١٧	١,٣٣٣	١,٧٤٠	٢,١١٠	٢,٥٦٧	٢,٨٩٨
١٨	١,٣٣٠	١,٧٣٤	٢,١٠١	٢,٥٥٢	٢,٨٧٨
١٩	١. ٣٢٨	١,٧٢٩	٢,٠٩٣	٢,٥٣٩	٢,٨٦١
٢٠	١,٣٢٥	١,٧٢٥	٢,٠٨٦	٢,٥٢٨	٢,٨٤٥
٢١	١,٣٢٣	١,٧٢١	٢,٠٨٠	٢,٥١٨	٢,٨٣١
٢٢	١,٣٢١	١,٧١٧	٢,٠٧٤	٢,٥٠٨	٢,٨١٩
٢٣	١,٣١٩	١,٧١٤	٢,٠٦٩	٢,٥٠٠	٢,٨٠٧
٢٤	١,٣١٨	١,٧١١	٢,٠٦٤	٢,٤٩٢	٢,٧٩٧
٢٥	١,٣١٦	١,٧٠٨	٢,٠٦٠	٢,٤٨٥	٢,٧٨٧
٢٦	١,٣١٥	١,٧٠٦	٢,٠٥٦	٢,٤٧٩	٢,٧٧٩
٢٧	١,٣١٤	١,٧٠٣	٢,٠٥٢	٢,٤٧٣	٢,٧٧١
٢٨	١,٣١٣	١,٧٠١	٢,٠٤٨	٢,٤٦٧	٢,٧٦٣
٢٩	١,٣١١	١,٦٩٩	٢,٠٤٥	٢,٤٦٢	٢,٧٥٦
٣٠	١,٣١٠	١,٦٩٧	٢,٠٤٢	٢,٤٥٧	٢,٧٥٠
٤٠	١,٣٠٣	١,٦٨٤	٢,٠٢١	٢,٤٢٣	٢,٧٠٤
٦٠	١,٢٩٦	١,٦٧١	٢,٠٠٠	٢,٣٩٠	٢,٦٦٠
١٢٠	١,٢٨٩	١,٦٥٨	١,٩٨٠	٢,٣٥٨	٢,٦١٧
∞	١,٢٨٢	١,٦٤٥	١,٩٦٠	٢,٣٢٦	٢,٥٧٦

منابع

- بهبودیان، جواد (۱۳۸۵) آماراحتمال مقدماتی (ویرایش سوم)، مشهد، دانشگاه امام‌رضا (ع).
- حری، عباس. ۱۳٦۲، تحلیل استنادی و شباهت‌های آن با علم الحدیث. نشر دانش ٤:(۲) ۱۱–۱۷
- دیانی، محمد حسین. ۱۳٦۱، کتاب‌سنجی، نشردانش ۳ (۲): ٤۰–٤۷.
- دیانی، محمدحسین، فریده عصاره. ۱۳٦۷، انتشارات دانشگاه‌های ایران در سال‌های ۱۳٤۰–۱۳٦۵. مجله علوم تربیتی و روانشناسی دانشگاه شهید چمران اهواز ۱ (۲). ۵۱–٦۹
- عصاره، فریده. ۱۳۷٦، بررسی مختصر کتاب‌سنجی، فصلنامه کتاب ۸ (٤): ۹۰–۹۷.
- عصاره، فریده. ۱۳۷٦، کتاب‌سنجی، مجله علوم تربیتی و روانشناسی، دانشگاه شهید چمران اهواز ۳ (٤): ٦۳–۷٤.
- عصاره، فریده. ۱۳۷۷، تحلیل استنادی، فصلنامه کتاب ۳۵ و ۳٦: ۳٤–٤۸.
- گادفری، ام جی، روباک، ای ام، شرلوک، ای جی (۱۳۸۷)، نخستین درس آمار، ترجمه: انیس ایرانمنش، ابولقاسم بزرگنیا، مشهد: دانشگاه آزاد اسلامی (مشهد).
- مدنی، علی (۱۳٦٦)، مفاهیم اساسی آمار، جلد اول، مشهد: انتشارات فروردین.
- منصوریان، یزدان. ۱۳۸۹، پنجاه محور پژوهشی در مطالعات علم‌سنجی، کتاب ماه کلیات ۱۳ (۱۰): ٦٤–۷۱.
- نصیری پرویز، ابراهیمی محمدعلی (۱۳۹۱)، روش‌های پیشرفته آماری در علوم زیستی و کشاورزی، یادواره کتاب.
- نصیری پرویز، احمدی سیدعلی‌اکبر، صالحی علی (۱۳۹۲)، تحلیل آماری، چاپ دوم، دانشگاه پیام نور.
- نصیری پرویز، رضایی‌پور الماسی غلامرضا، خسروی راد محمدهادی (۱۳۹۰)، مفاهیم و روش‌های آماری، چاپ پنجم، دانشگاه پیام نور.

- نصیری پرویز، شقاقی فرهاد، (۱۳۹۰) استنباط آماری در روان‌شناسی و علوم تربیتی، چاپ چهارم، دانشگاه پیام نور.
- نصیری، پرویز (۱۳۸٤)، آمار و احتمالات مهندسی، چاپ دوازدهم، دانشگاه پیام نور.
- نصیری پرویز، قربانی علی، (۱۳۹٤) آمار و احتمالات مهندسی (رشته مهندسی مدیریت پروژه)، چاپ سوم، دانشگاه پیام نور.
- واگان، لیون (۱۳۸٤)، روش‌های آماری برای متخصصان کتابداری و اطلاع رسانی، ترجمه: محمدرضا قانع، کیوان کوشا، تهران: چاپار.
- هویدا، علیرضا (۱۳۷۸)، آمار و روشهای کمی در کتابداری و اطلاع رسانی، تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت).

- Brookes, B. C L. Egghe, R. Rousseau (Eds) (۱۹۸۸), *Informetrics ۸۷/۸۸. Select Proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*. Comments on the scope of bibliometrics, In:, Amsterdam, Elsevier Science, pp. ۲۹–٤۱.
- Brookes, B. C. (۱۹۹۰), Biblio-, Sciento-, Infor-metrics??? What are we talking about? In: L. Egghe, R. Rousseau (Eds), *Informetrics ۸۹/۹۰. Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics,* Amsterdam, Netherlands, Elsevier, pp. ۳۱–٤۳.
- Campbell,F. (۱۸۹٦), Theory of the National and International Bibliography. London.
- GlÄnzel, W., H. Kretschmer (۱۹۹۲), (Eds) Selected Papers Presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; ۱۹۹۳ September ۱۱–۱٥; Berlin, Germany, *Research Evaluation*, ۲ (۳): ۱۲۱–۱۸۸.
- GlÄnzel, W., H. Kretschmer (۱۹۹٤a), (Eds) Selected Papers Presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; ۱۹۹۳ September ۱۱–۱٥; Berlin, Germany.*Scientometrics*, ۳۰ (۱).
- GlÄnzel, W., H. Kretschmer (۱۹۹٤b), (Eds) Selected Papers Presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; ۱۹۹۳ September ۱۱–۱٥; Berlin, Germany, *Science and Science of Science*, ۳ (٥).
- Hood, W. W. & Wilson, C. S. (۲۰۰۱), The literature of bibliometrics, scientometrics,and informetrics. Scientometrics, ٥۲ (۲) , p ۲۹۱–۳۱٤
- Tague-Sutcliffe, J. M. (۱۹۹۲), An introduction to informetrics, *Information Processing & Management*, ۲۸: ۱–۳.