به نام خدا

وزارت علوم تحقیقات و فناوری

دانشگاه فنی و حرفه ای

دانشگاه فنی و حرفه ای خراسان شمالی

آموزشکده کشاورزی سمنگان

جزوه درس:

آمار

مقطع: کاردانی

مدرس: میرحسن زاده

فایده آمار: در صورتی که بنحو درستی صورت گیرد به محققان امکان خواهد داد در جمع آوری داده ها و اطلاعات تصمیمات درست تری بگیرند.
مسائل مختلف امروزه شناخت آماری مسائل برای محققان یکی از ضروریات می باشد، زیرا این امر به آنان کمک می نماید که مشکلات را شناخته و حتی المقدور راه مقابله با آنها را پیش بینی کنند.

مراحل مختلف بررسی های آماری:
در آمار به طور کلی اندازه گیری داده های تحقیق و تبدیل داده ها و مشاهدات تجزی به علائم و نشانه های قراردادی مورد بحث قرار می گیرد. پس از اندازه گیری، تنظیم داده ها و اطلاعات می رسد.
قدم بعدی در بررسی های آماری، محاسبه پارامترها و شاخص ها و معیارهایی است که خصوصیات توزیع را بیان کند.
قدم اساسی بعدی تفسیر، تعبیر و استنباط از روی داده هاست. بوسیله توصیف اطلاعات می توان جهات مختلف پدیده های تحقیق را به صورت دقیق بیان داشت.

روشهای آماری:
روشهای آماری به مجموعه اصول، روابط و فرمولهایی گفته می شود که به عنوان ابزاری برای تجزیه و تحلیل و شناخت بهتر جامعه ها به کار می رود.

تعریف جامعه:
عبارت است از مجموعه ای از افراد یا اشیاء یا عناصری که حداقل در یک صفت مشترک باشند. بنابراین مجموعه ممکن است بصورتهای مختلف مانند جامعه ایرانی، جامعه دانشجویان، جامعه زمینهای کشاورزی منطقه و ... باشد.

داده های آماری:
برای اینکه بتوان یک مسئله را بررسی کرد، در قدم اول احتیاج به مشاهدات و گرفتن اطلاعات و آزمایشهایی در زمینه مورد بررسی داریم. نتیجه جمع آوری این مشاهدات و اطلاعات و آزمایشها را میتوان بصورت کمی درآورد. مجموعه این کمیتها که بصورت عددی بیان می گردد داده های آماری نامیده می شود.

متغیر پیوسته و متغیر گسسته یا جدا:
چنانچه یک متغیر بتواند همه اندازه های ممکن در یک فاصله را اختیار کند، متغیر پیوسته نامیده می شود مانند میزان عملکرد گندم یا طول قد افراد یک جامعه.
متغیری که تنها اندازه های خاص را در یک دامنه بپذیرد، متغیر گسسته یا ناپیوسته نام دارد. مثل تعداد اتاقهای منازل یک شهر، تعداد فرزندان یک خانواده.

- داده های آماری در صورتیکه بصورت پیوسته باشند با اندازه گیری بدست می آیند ولی در صورتی که بصورت جدا باشد بوسیله شمارش محاسبه می گردد.

تعداد دسته ها:
برای خلاصه کردن اطلاعات، لازم است که داده ها را به دسته های مختلفی تقسیم کنیم. تقسیم داده ها به دسته های مختلف بدین منظور است که اولاً بتوان تا آنجا که ممکن است محاسبات را به سهولت انجام داد و ثانیاً بوسیله جدولی که تنظیم می گردد تصویری منطقی از توزیع افراد جامعه بدست آورد و ثالثاً با استفاده از آن، توزیع افراد جامعه را بوسیله نمودار نمایش داد. جدولی که برای این منظور تنظیم می شود، جدول توزیع فراوانی نام دارد.
معمولاً هیچ قاعده خاصی برای تعداد دسته ها وجود ندارد و معمولاً این دسته بندی ها بر اساس نیاز تحقیق و نظرات خاص محقق یا آمارشناس صورت می گیرد.

فاصله دسته ها:
با در اختیار داشتن دامنه تغییرات یعنی R و تعداد دسته ها یعنی K می توان فاصله بین دسته ها را بصورت زیر بدست آورد:

$$a = \frac{R}{K} = \frac{\text{دامنه تغییرات}}{\text{تعداد دسته ها}}$$

که در آن a فاصله بین دسته هاست.

فاصله دسته ها را در توزیع فراوانی بطور مختلف می توان نشان داد. مثلاً بصورت:

$$a < x < b$$

و معنی آن این است که متغییر x کلیه مقادیر بین a و b را می تواند اختیار کند، اما خود a و b را اختیار نمی کند. چنین فاصله ای را فاصله باز گویند.
اگر x بتواند کلیه مقادیر بین فاصله ای که a و b را اختیار کند و خود a و b را هم اختیار کند چنین فاصله ای را فاصله بسته گویند و آن را به صورت زیر نمایش می دهند

$$a \leq x \leq b$$

- در جداول فراوانی فواصل باز کمتر مورد استفاده قرار می گیرد و فواصل بسته اکثراً در مورد متغییرهای جدا و فواصل نیم بسته یا نیم باز بیشتر در مورد متغییرهای پیوسته بکار می رود.
تذکر:
الف: حدود دسته ها را کناره دسته ها یا مرز دسته ها یا طبقات نیز می گویند.
ب: حد پایین هر دسته را کناره پایین و حد بالا را کناره بالای آن دسته می گویند.
ج: برای سهولت در محاسبات معمولاً تعداد دسته ها و فواصل دسته ها طوری در نظر گرفته می شود که حدود دسته ها به اعداد صحیح ختم گردد.

جدول فراوانی:

به طوری که در مثال قبل دیدیم دامنه تغییرات، تعداد دسته ها و حدود دسته ها تعیین گردید. از انتساب افراد به این دسته ها، جدولی بدست می آید که جدول فراوانی نامیده می شود.

در این جداول معمولاً در ستون اول حدود دسته ها و در ستون دوم تعداد افراد هر دسته درج می گردد. مقادیر این ستون به F نمایش داده می شود. F_i را فراوانی مطلق هر دسته گویند و از نظر سهولت تنها واژه فراوانی برای معرفی آن بکار گرفته می شود.

به جدول فراوانی میتوان ستون دیگری اضافه نمود. در این ستون اگر فراوانی هر دسته به تعداد کل افراد مورد مطالعه تقسیم شود، ملاکی بدست می آید که آنرا فراوانی نسبی می نامند و با $f_i(x)$ نمایش داده می شود.

$$f_i(x_i) = \frac{F_i}{N}$$

→ تعداد افراد نظیر دسته iام

→ تعداد کل افراد

فراوانی تجمعی:

اگر به جدول فراوانی ستون دیگری اضافه شود و آن را با F_c نمایش دهیم، F_c همان فراوانی تجمعی و یا تراکمی می باشد.

طریقه محاسبه فراوانی تجمعی بدین صورت است که فراوانی مطلق هر دسته را با مجموع فراوانیهای مطلق دسته های قبل جمع می نمائیم، در نتیجه:

$F_{c_1} = F_1$ فراوانی تجمعی دسته اول

$F_{c_2} = F_{c_1} + F_2$ فراوانی تجمعی دسته دوم

$F_{c_K} = F_{c_{K-1}} + F_K = N$ فراوانی تجمعی دسته آخر

فراوانی تجمعی نسبی:

اگر فراوانی تجمعی هر دسته را به کل فراوانی تقسیم کنیم، فراوانی تجمعی نسبی بدست می آید که با $f_c(x)$ نمایش می دهیم.

$$f_c(x) = \frac{F_c}{N}$$

مثال: بمنظور مطالعه تاثیر یک نوع قارچ کش ۵۰ مزرعه گندم از هر کدام از مزارع به مساحت ۱ متر مربع نمونه برداری به عمل آمده و تعداد بوته های آلوده یادداشت گردیده است که شرح زیر می باشد.

— چون متغیر جدا است لذا از فواصل بسته استفاده می کنیم.

۵۱	۶۸	۷۰	۷۷	۶۱	۵۵	۴۸	۴۶	۶۱	۶۷
۸۹	۶۰	۵۹	۵۲	۶۵	۶۲	۶۶	۴۵	۷۸	۳۰
۶۵	۶۸	۵۹	۵۵	۶۳	۵۷	۶۴	۷۰	۶۷	۷۴
۶۲	۵۸	۵۰	۵۷	۶۷	۶۹	۸۸	۵۷	۸۲	۶۱
۴۴	۶۰	۷۶	۶۴	۶۴	۶۷	۴۸	۸۳	۶۲	۵۴

جدول فراوانی را تشکیل داده و فراوانیهای مطلق ، نسبی و تجمعی را بدست آورید ، درصورتیکه تعداد دسته ها $K=6$ باشد.

حل:

$R = ۸۹ - ۳۰ = ۵۹$

$a = \frac{R}{K} = \frac{۵۹}{۶} = ۹٫۸ \simeq ۱۰$

حدود دسته ها	F_i	$F_r(x)$	F_c
۳۰-۳۹	۱	۰/۰۲	۱
۴۰-۴۹	۵	۰/۱۰	۶
۵۰-۵۹	۱۲	۰/۲۴	۱۸
۶۰-۶۹	۲۲	۰/۴۴	۴۰
۷۰-۷۹	۶	۰/۱۲	۴۶
۸۰-۸۹	۴	۰/۰۸	۵۰
	۵۰	۱	

نمودارهای فراوانی:

نمودار ~~های فراوانی~~ بافتی یا هیستوگرام

اگر کناره بالا و پائین دسته ها بطور مناسب روی محور متغیرها منتقل شود و بر روی هر فاصله مستطیلی ساخته شود که قاعده اش فاصله آن دسته و ارتفاع آن فراوانی خطی آن دسته باشد ، مجموعه مستطیلهایی که بدین ترتیب بدست می آید نمودار بافتی یا هیستوگرام خوانده می شود.

نکته۱: هیستوگرام بیشتر در مواردی که متغیرها بصورت پیوسته باشد بکار می رود.

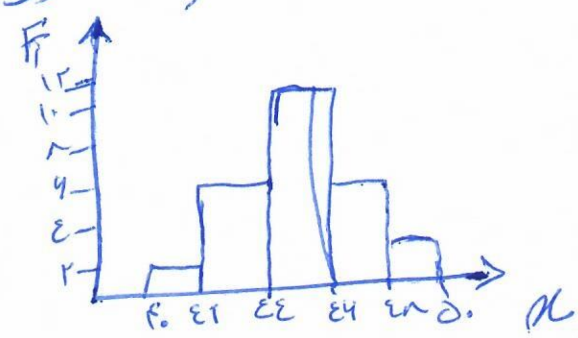
نکته ۲: اگر متغیر جدا باشد و فواصل بصورت دسته انتخاب شده باشد چنانچه با این فواصل نمودار رسم شود ، فاصله کوچکی بین هر دسته پیدا می شود. برای رفع این اشکال باید

فواصل بصورت نیم باز تبدیل شود.
برای اینکار ابتدا مراکز دسته ها یا فواصل میانی هر دسته بصورت زیر محاسبه می گردد.

$\frac{a_1+b_1}{2}$ مرکز دسته اول

$\frac{a_2+b_2}{2}$ مرکز دسته دوم

حال فاصله هر دسته (مثلا اگر $a=10$ باشد) را بر عدد ۲ تقسیم کرده و عدد بدست آمده را اگر از مرکز دسته ها کم کنیم کناره پایینی طبقه و اگر به مرکز دسته اضافه شود کناره بالای طبقه بدست خواهد آمد. این عمل برای سایر دسته ها نیز انجام می شود. حدود جدیدی که بدین طریق بدست می آید، بصورت فاصله نیم باز می باشد که بوسیله آن میتوان هستوگرام را رسم کرد.



نمودار چند بر فراوانی یا پلی گان:
در یک صفحه کاغذ دو محور عمود بر هم اختیار و بر روی محور افقی مقادیر متغیر را با انتخاب واحد مناسب منتقل می نمائیم. توجه کنید که برای نقل متغیرها از مراکز دسته ها استفاده می گردد و بر روی محور y ها یا محور F ها فراوانی نظیر مقادیر هر دسته درج می شود.

مقایسه دو نمودار هیستوگرام و پلی گان:
تفاوت معناداری با یکدیگر ندارند و هر دو نمودار یک وضعیت مشخص می نمایند جز اینکه هیستوگرام به دلیل ساده بودن آن بیشتر در روزنامه ها و مجلات برای نشان دادن وضعیت اقتصادی معمولا استفاده قرار می گیرد. اما پلی گان به دلیل علمی تر بودن و قابلیت انطباق آن با الگوهای آماری (مانند منحنی نرمال) بیشتر برای مقایسه ها بکار برده می شود.
- وقتی بخواهیم چند توزیع فراوانی را در یک محور مختصات نمایش دهیم از پلی گان استفاده می کنیم زیرا دو منحنی را به راحتی میتوان در یک محور رسم کرد. در صورتیکه اگر از هستوگرام استفاده کنیم به دلیل اینکه ممکن است مستطیلها روی یکدیگر قرار گیرند، وضعیت دو توزیع قابل تشخیص نخواهد بود.

نمودار میله‌ای :

برای رسم این نمودار بر روی F ها تعداد افراد نظیر هر دسته و بر روی محور X ها مقادیر متغیر نظیر آن دسته درج می‌گردد.

به ازاء هر متغیر و F نظیر آن نقطه‌ای در صفحه مختصات بدست می‌آید که اگر از نقاط مذبور به موازات محور F عمودی بر X ها رسم شود، شکل حاصل نمودار میله‌ای را مشخص می‌کند.

مثال: آمار سطح زیر کشت گندم و جو در سال‌های زراعی ۸۹-۹۰-...-۹۳

نمودار دایره‌ای :

این نمودار به شکل دایره‌ای است و وسیله بهتری برای تقسیم بشمار می‌رود. برای رسم آن لازم است ابتدا ~~فراوانیها بصورت درصد~~ مقادیر به صورت فراوانی نسبی محاسبه شود و سپس دایره‌ای رسم می‌کنیم و درصدهای مربوط به صفت مورد نظر را از دایره جدا می‌کنیم. (طبیعی است که کل دایره برابر ۱۰۰٪ است)

تقسیم بندی درصدها از قسمت تمامی دایره شروع شده و در جهت عقربه‌های ساعت ادامه می‌یابد و چون دایره به ۳۶۰ درجه تقسیم شده است، لذا مقادیر ذکر می‌ فراوانیهای نسبی می‌باشند با ضرب فراوانی‌های نسبی در ۳۶۰ تبدیل به درجه می‌شود. سپس بوسیله نقاله به اندازه درجه‌ای که بدینوسیله بدست می‌آید جدا می‌کنیم.

مثال: سطح زیر کشت یک نوع گیاه در سال ۲۰۱۴ طبق جدول زیر بوده است. نمودار دایره‌ای آن را رسم کنید.

قاره	سطح زیر کشت	F(%)	360 F(%)
آسیا	۱۸۶۲	~~۰/۵۷۸~~	۲۰۸
اروپا	۶۰۸	۰/۱۸۹	۶۸
امریکا	۴۳۸	۰/۱۳۶	۴۹
افریقا	۲۹۶	۰/۰۹۲	۳۳
استرالیا	۱۷/۵	۰/۰۰۵	۲
جمع	۳۲۲۱/۵	۱۰۰	۳۶۰

نمودار دایره‌ای

شاخصهای تمایل مرکزی

مقدمه:

به منظور دسترسی به نتایج تحقیق و بررسی آن در بسیاری از تحقیقات، ضرورت دارد کمیت یا شاخصی به عنوان معرف یا نماینده کلی داده ها و مشاهدات ارائه و محاسبه گردد.

برای محاسبه مقدار متوسط میتوان از روشهای آماری زیادی استفاده نمود. نما یا مد، میانه، چندک ها و میانگینهای حسابی و هندسی از مهمترین شاخصهای مرکزیت یا نقطه ثقل داده ها می باشند. به هریک از این شاخصها، اندازه های اصلی تمایل مرکزی نیز گفته میشود که از آنها به منظور تعیین جایگاه و موقعیت کلیه داده ها استفاده میشود.

نما:

نما ساده ترین شاخص گرایش مرکزی است که اغلب در دامنه میانی قرار دارد. در بین مشاهدات حاصل از یک بررسی آماری، نما مشاهده ای است که دارای بیشترین فراوانی باشد.

در یک توزیع فراوانی طبقه بندی شده، نما برابر نمره متوسط طبقه ای است که بیشترین فراوانی را داراست.

مثال: مد یا نما را برای هر دو مجموعه مشاهدات زیر، تعیین کنید.

الف = {۱۵و۱۷و۸و۱۵و۱۷و۱۵و۱۱و۱۴و۱۲و۱۱}

ب = {۱۵و۳و۲و۱۵و۱و۸و۱۵و۷و۳و۲}

برای هر دو مجموعه برابر با ۱۵ می باشد.

مثال: در جدول توزیع فراوانی زیر، نما برابر با ۸ می باشد:

طبقات	۴	۵	۶	۷	۸	۹	۱۰
فراوانی	۱	۳	۱	۲	۴	۱	۱

نکته ۱: اگر در توزیع تنها یک عدد وجود داشته باشد که دارای بیشترین فراوانی باشد، توزیع را یک نمایی می نامند. توزیعی که دارای دو نما باشد را توزیع دو نمایی گویند.

نکته ۲: در مواردی که توزیع فراوانی دو نمایی باشد، چنانچه دو نما با هم تفاوت زیادی نداشته باشند، میتوان از میانگین آنها به عنوان برآورد نمای توزیع فراوانی استفاده نمود.

مد یا نما شاخص آماری معتبری نیست، زیرا تنها با توجه به فراوانی مشاهدات محاسبه میشود و کمیت داده ها در مقدار آن دخالت ندارند.

میانه: در داده های مرتب شده میانه عددی است که رشته اعداد را از نظر تعداد به دو قسمت مساوی تقسیم می‌کند به طوری که نیمی از اعداد کمتر و نیم دیگر بیشتر از آن باشند.

میانه نمونه‌ای مرکب از n اندازه x_1 ، ... و x_n عبارت است از اندازه وسطی ، در صورتی که اندازه ها به ترتیب از کوچکترین به بزرگترین مقدار مرتب شده باشد. اگر n عدد فردی باشد ، یک مقدار وسطی وجود دارد که میانه است. در مواردی که تعداد داده ها زوج می باشد ، دو مقدار وسطی وجود دارند که در این صورت میانه متوسط دو داده وسطی می باشد.

مثال ۱: میانه اعداد زیر را به دست آورید.

۱۲ و ۸ ۲۰ ۴۸ ۱۸ ۴۵ ۲ ۱۸ ۳۰ ۱ ۳

ابتدا باید داده ها به ترتیب صعودی مرتب گردند:

۱ ۲ ۳ ۸ ۱۲ ۱۸ ۱۸ ۲۰ ۳۰ ۴۵ ۴۸

چون تعداد اعداد فرد هستند ، عدد ششم (۱۸) میانه است.

مثال ۲: میانه داده های زیر چند است

x_i = ۳ ۴۰ ۴ ۶ ۲۹ ۲۱ ۹ ۷ ۱۶ ۵

پس از مرتب کردن داده ها معلوم می گردد که عدد ۸ یعنی میانگین دو عدد ۷ و ۹ میانه است.

۳ ۴ ۵ ۶ $\boxed{۷ \quad ۹}$ ۱۶ ۲۱ ۲۹ ۴۰

مثال ۳: داده های مربوط به طبقه بندی طول ۳۰ بوته برحسب سانتی متر در جدول زیر آمده است ، میانه را محاسبه کنید.

شماره طبقه	۱	۲	۳	۴	۵	۶	۷	۸
X_i	۱۰-۱۴٫۵	۱۵-۱۹٫۵	۲۰-۲۴٫۵	۲۵-۲۹٫۵	۳۰-۳۴٫۵	۳۵-۳۹٫۵	۴۰-۴۴٫۵	۴۵-۴۹٫۵
F_i	۱	۳	۲	۱۰	۳	۵	۳	۳
F_c	۱	۴	۶	۱۶	۱۹	۲۴	۲۷	۳۰

میانه این داده نقطه ای است که ۱۵ عدد کوچک را از ۱۵ عدد بزرگ جدا می سازد. پانزدهمین داده میانه است ($۳۰ \times \frac{1}{۲} = ۱۵$)، بنابراین میانه در طبقه ای قرار دارد که فراوانی تجمعی آن ۱۷ است، یعنی طبقه ای که در آن اعداد ۲۹٫۵-۲۵ سانتی متر قرار دارند. در بین مشاهدات تا قبل از عدد ۲۵ سانتی متر، شش مشاهده و تا ۲۹٫۵ سانتی متر، شانزده مشاهده وجود دارد. پس میانه در این طبقه قرار دارد. بنابراین با توجه به فراوانی تجمعی صعودی داده ها، میانه در طبقه چهارم قرار دارد. طبقه ای که ۱۵ عدد کوچک را از ۱۵ عدد بزرگ مجزا می کند. روشن است که این عدد بین ۲۵ و ۲۹٫۵ سانتی متر واقع شده است.

روش کلی برای بدست آوردن میانه:

برای تعیین میانه در جدول توزیع فراوانی، ابتدا باید طبقه ای که عدد میانه در آن قرار دارد تعیین شود. چون نقطه میانه، نقطه ای است که پس از مرتب کردن داده ها، آنها را به دو بخش مساوی تقسیم می کند، بنابراین باید از رابطه ($n \times \frac{1}{۲}$) شماره عددی میانه را پیدا کرد (تعداد کل داده ها $= n$). بنابراین با توجه به این عدد می توان طبقه ای را که در آن میانه واقع است، تشخیص داد و عدد میانه را تعیین کرد. برای این منظور باید فاصله طبقه ای که میانه در آن واقع شده را به فراوانی آن طبقه تقسیم کرده و قسمت لازم را به حد پائین میانه اضافه نمود تا به نقطه میانی مشاهدات رسید. این عمل را میانگیری می نامند. این عمل به کمک رابطه زیر به راحتی امکان پذیر است و از آن می توان برای بدست آوردن عدد دقیق میانه استفاده نمود.

$$Md = L + \left(\frac{\frac{n}{2} - F_{c-1}}{F_i} \right) i$$

$n =$ تعداد داده ها $\quad$ $i =$ عرض طبقه

$L =$ حد پائینی واقعی طبقه ای که میانه در آن است

$F_{c-1} =$ فراوانی تجمعی طبقه قبلی

$F_i =$ فراوانی مطلق طبقه ای که میانه در آن است.

$$Md = L + \left(\frac{\text{محل قرار گرفتن میانه} - \text{فراوانی تجمعی طبقه قبلی}}{\text{فراوانی مطلق طبقه ای که میانه در آن قرار دارد}} \right) \text{عرض طبقه}$$

مثال: جدول توزیع فراوانی زیر را در نظر بگیرید و میانه را حساب کنید

x_i	۴	۵	۶	۷	۸	۹	۱۰
F_i	۱	۱	۱	۳	۱	۲	۱

ابتدا باید جدول را با توجه به حدود طبقات تنظیم نمود، پس فراوانی های تجمعی را محاسبه کرده و از فرمول فوق، میانه را بدست آورد.

داده ها	حدود طبقات	فراوانی ساده (مطلق)	فراوانی تجمعی
۴	۳٫۵ - ۴٫۴	۱	۱
۵	۴٫۵ - ۵٫۴	۱	۲
۶	۵٫۵ - ۶٫۴	۱	۳
۷	۶٫۵ - ۷٫۴	۳	۶
۸	۷٫۵ - ۸٫۴	۱	۷
۹	۸٫۵ - ۹٫۴	۲	۹
۱۰	۹٫۵ - ۱۰٫۴	۱	۱۰

$$N\tfrac{1}{2} = 10 \times \tfrac{1}{2} = 5$$

$$Md = 6.5 + \left(\frac{5-3}{3}\right) \times 1 =$$

$$Md = 6.5 + 0.67 = 7.17$$

چندکها:

با تعیین نقطه میانه، رشته اعداد به دو قسمت تقسیم می‌شود. رشته مرتب شده داده‌ها را می‌توان به هر تعدادی تقسیم کرد. به طور کلی با تعیین n شاخص در یک سلسله مرتب شده از داده‌های آن سلسله به $n+1$ قسمت تقسیم می‌شود. هر یک از n شاخص که به وسیله آنها رشته اعداد تقسیم می‌شود یک چندک است. بدین ترتیب با انتخاب ۳، ۹ و ۹۹ شاخص، رشته اعداد به ۴، ۱۰ و ۱۰۰ چندک تقسیم می‌گردد که هر یک از این تقسیمات به ترتیب چارک (یک چهارم)، دهک (یک دهم) و صدک (یک صدم) نامیده می‌شوند.

بدیهی است در مشاهدات طبقه بندی نشده می‌توان بعد از مرتب کردن آنها، چندک مورد نظر را تعیین نمود. در جداول توزیع فراوانی از روش محاسبه میانه برای محاسبه چندکها استفاده می‌شود.

$$P_p = L + \left(\frac{P_n - F_{c-1}}{F_i}\right) i$$

P_p = چندک مورد نظر

L = حد پایینی واقعی دسته‌ای که چندک در آن قرار دارد

P_n = تعداد مشاهداتی که در چندک مورد نظر قرار دارند.

F_{c-1} = فراوانی تجمعی طبقه ما قبل طبقه‌ای که چندک مورد نظر در آن است.

F_i = فراوانی مطلق طبقه‌ای که شامل چندک مورد نظر است.

چندک	چارک	دهک	صدک
Partile	Quartile	Decile	Percentile

مثال: در مورد مثال طبقه بندی طول ۳۰ بوته، چارک اول، دهک دوم و صدک هشتاد و پنجم را محاسبه کنید

$$P_n = ۳۰ \times \frac{1}{4} = ۷/۵ \Rightarrow P_{1/4} = Q_1 = ۲۵ + \left(\frac{۷/۵ - ۶}{۱۰}\right) \times ۵ = ۲۵/۷۵$$

$$P_n = ۳۰ \times \frac{۲}{۱۰} = ۶ \Rightarrow P_{۲/۱۰} = D_2 = ۲۰ + \left(\frac{۶ - ۴}{۲}\right) \times ۵ = ۲۵$$

$$P_n = ۳۰ \times \frac{۸۵}{۱۰۰} = ۲۵/۵ \Rightarrow P_{۸۵/۱۰۰} = P_{85} = ۴۰ + \left(\frac{۲۵/۵ - ۲۴}{۳}\right) \times ۵ = ۴۲/۵$$

مثال: جدول توزیع فراوانی مربوط به قد ۱۰۰ نفر را در نظر بگیرید و میانه، چارک اول، صدک هفتاد و پنجم و دهک سوم را محاسبه کنید.

حدود طبقات	F_i	F_c
۱۴۹/۵ - ۱۵۶/۴	۱۵	۱۵
۱۵۶/۵ - ۱۶۳/۴	۲۰	۳۵
۱۶۳/۵ - ۱۷۰/۴	۳۰	۶۵
۱۷۰/۵ - ۱۷۷/۴	۲۵	۹۰
۱۷۷/۵ - ۱۸۴/۴	۱۰	۱۰۰

$$Md = ۱۶۳/۵ + \left(\frac{۵۰ - ۳۵}{۳۰}\right) \times ۷ = ۱۶۷$$

$$Q_1 = ۱۵۶/۵ + \left(\frac{۲۵ - ۱۵}{۲۰}\right) \times ۷ = ۱۶۰$$

$$P_{75} = ۱۷۰/۵ + \left(\frac{۷۵ - ۶۵}{۲۵}\right) \times ۷ = ۱۷۳/۳$$

$$D_3 = ۱۵۶/۵ + \left(\frac{۳۰ - ۱۵}{۲۰}\right) \times ۷ = ۱۶۱/۷۵$$

میانگین حسابی:

که از این پس آن را میانگین خواهیم نامید، عبارت از حاصل جمع داده ها تقسیم بر تعداد آنها می باشد. اگر میانگین با علامت μ و داده ها با X_i و تعداد کل داده ها یا مشاهدات با N نشان داده شود، میانگین طبق فرمول زیر محاسبه می شود:

$$\mu = \frac{\sum X_i}{N}$$

وقتی داده ها در جدول فراوانی بدون طبقه بندی تنظیم شده باشد از فرمول

$$\mu = \frac{\sum F_i X_i}{\sum F_i} = \frac{\sum F_i X_i}{N}$$

برای محاسبه میانگین استفاده می شود.

مثال: میانگین X را برای یک سری عدد با فراوانی های جدول زیر محاسبه کنید؟

X	۰	۱	۲	۳	۴	۵	۶	۷	۸
F_i	۱	۹	۲۶	۵۹	۷۲	۵۲	۲۹	۷	۱
F_iX_i	۰	۹	۵۲	۱۷۷	۲۸۸	۲۶۰	۱۷۴	۴۹	۸

$$M = \frac{\sum F_i X_i}{\sum F_i}$$

$$M = \frac{۱۰۱۷}{۲۵۶} = ۳/۹۷$$

میانگین هندسی:
اگر مشاهدات حاصل از یک بررسی به صورت درصد، نسبت و امثال اینها بیان شده باشد، معمولاً از این میانگین استفاده می شود.
میانگین هندسی n داده آماری برابر است با ریشه n ام حاصلضرب این داده ها و آن را با M_g یا $\bar{X}_g$ نشان می دهند.

$$M_g = \sqrt[n]{x_1 x_2 x_3 \cdots x_n}$$

مثال: در چهار مزرعه نمونه برداری از مزرعه ای میزان خسارت ناشی از یک آفت ۲۵، ۳۰، ۲۴ و ۴۵٪ ثبت شده است. مقدار متوسط خسارت چقدر است؟

$$M_g = \sqrt[4]{25 \times 30 \times 24 \times 45} = 30$$

شاخصهای پراکندگی:
مهمترین شاخصهای پراکندگی عبارتند از: دامنه کلی تغییرات، چارک متوسط، انحراف متوسط و انحراف معیار که شاخص اخیر از نظر اصول و مبانی آمار و کاربرد مهمترین آنها می باشد.

انحراف معیار:
بر مبنای انحراف مشاهدات از میانگین محاسبه می شود و متوسط پراکندگی مشاهدات را در اطراف میانگین نشان می دهد. برخی از انحرافهای داده ها از میانگین، مثبت و برخی منفی می باشند و جمع جبری آنها مساوی با صفر است. میانگین توان های دوم انحرافها را (برای حذف علامت جبری مثبت و منفی) واریانس گویند. با این روش، مقیاس اندازه گیری داده ها به توان ۲ می رسد. با این حال می توان با جذر واریانس به مقیاس اصلی دست یافت که به آن

... S با δ (سیگما) نشان داده می‌شود عبارت
انحراف معیار گفته می‌شود بنابراین انحراف معیار که
است از جذر واریانس که خود میانگین ~~مجموع~~ مربع انحراف‌ها یا به اختصار میانگین مربعات
می‌باشد.
برای محاسبه واریانس دو روش کلی وجود دارد:

۱- فرمول تعریفی:
از این روش در مواردی که میانگین مشاهدات عددی بدون رقم اعشاری است استفاده می‌گردد

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

برای محاسبه واریانس برای داده‌های دارای فراوانی از فرمول زیر استفاده می‌گردد:

$$S^2 = \frac{\sum F_i (x_i - \bar{x})^2}{\sum F_i \text{ یا } N}$$

۲- فرمول کاربردی:
وقتی میانگین مشاهدات دارای رقم اعشار باشد از فرمول کاربردی محاسبه واریانس استفاده می‌گردد

$$S^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N}$$

محاسبه واریانس برای داده‌های دارای فراوانی از فرمول زیر استفاده می‌گردد:

$$S^2 = \frac{\sum F_i x_i^2 - \frac{(\sum F_i x_i)^2}{\sum F_i}}{\sum F_i}$$

مثال: داده های زیر را در نظر بگیرید: مجموع مربع انحراف داده ها از میانگین، واریانس و انحراف معیار آنها چقدر است؟

x_i	۳	۲	۵	۶	۹

$$\bar{x} = \frac{\sum x_i}{N} = \frac{۲۵}{۵} = ۵$$

x_i	۳	۲	۵	۶	۹	$\sum x_i = ۲۵$
$(x_i - \bar{x})$	−۲	−۳	۰	۱	۴	$\sum (x_i - \bar{x}) = ۰$
$(x_i - \bar{x})^2$	۴	۹	۰	۱	۱۶	$\sum (x_i - \bar{x})^2 = ۳۰$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{۳۰}{۵} = ۶ \qquad s = \sqrt{s^2} = \sqrt{۶} = ۲٫۴۵$$

واریانس داده های فوق با استفاده از فرمول کاربردی به صورت زیر است:

x_i	۳	۲	۵	۶	۹
x_i^2	۹	۴	۲۵	۳۶	۸۱

$\sum x_i^2 = ۱۵۵$

$\sum x_i = ۲۵$

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N} = \frac{۱۵۵ - \frac{(۲۵)^2}{۵}}{۵} = \frac{۱۵۵ - \frac{۶۲۵}{۵}}{۵} = \frac{۱۵۵ - ۱۲۵}{۵} = ۶$$

مثال: اعداد زیر نمرات دروس مختلف یک دانشجو هستند. واریانس و انحراف معیار نمرات را به کمک فرمول کاربردی و تئوری بدست آورید

x_i = ۶و۱۰و۸و۷و۱۲و۸و۱۵و۱۲و۱۷و۱۵

با استفاده از فرمول کاربردی

x_i	۶	۱۰	۸	۷	۱۲	۸	۱۵	۱۲	۱۷	۱۵
x_i^2	۳۶	۱۰۰	۶۴	۴۹	۱۴۴	۶۴	۲۲۵	۱۴۴	۲۸۹	۲۲۵

$\sum x_i = ۱۱۰$

$\sum x_i^2 = ۱۳۴۰$

$$s^2 = \frac{۱۳۴۰ - \frac{(۱۱۰)^2}{۱۰}}{۱۰} = \frac{۱۳۰}{۱۰} = ۱۳ \Rightarrow s = \sqrt{s^2} = \sqrt{۱۳} = ۳٫۶۱$$

با استفاده از فرمول تئوری

x_i	۶	۱۰	۸	۷	۱۲	۸	۱۵	۱۲	۱۷	۱۵
$(x_i - \bar{x})$	−۵	−۱	−۳	−۴	۱	−۳	۴	۱	۶	۴
$(x_i - \bar{x})^2$	۲۵	۱	۹	۱۶	۱	۹	۱۶	۱	۳۶	۱۶

$$\bar{x} = \frac{\sum x_i}{N} = \frac{۱۱۰}{۱۰} = ۱۱$$

$$s^2 = \frac{۱۳۰}{۱۰} = ۱۳$$

$$s = \sqrt{s^2} = \sqrt{۱۳} = ۳٫۶۱$$

مثال: برای مشاهدات زیر مقدار انحراف معیار را براساس فرمول تئوری تعیین کنید

فاصله دسته x	F_i
۸-۱۰	۱۲
۱۱-۱۳	۱۰
۱۴-۱۶	۲
۱۷-۱۹	۲
۲۰-۲۲	۱۰

$$\bar{x} = \frac{\sum F_i x_i}{\sum F_i} = \frac{۵۰۴}{۳۶} = ۱۴$$

$$S^2 = \frac{\sum F_i (x_i - \bar{x})^2}{N}$$

فاصله دسته	مرکز دسته x_i	F_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$F_i (x_i - \bar{x})^2$
۸-۱۰	۹	۱۲	(۹-۱۴)	۲۵	۳۰۰
۱۱-۱۳	۱۲	۱۰	(۱۲-۱۴)	۴	۴۰
۱۴-۱۶	۱۵	۲	(۱۵-۱۴)	۱	۲
۱۷-۱۹	۱۸	۲	(۱۸-۱۴)	۱۶	۳۲
۲۰-۲۲	۲۱	۱۰	(۲۱-۱۴)	۴۹	۴۹۰

$$S^2 = \frac{۸۶۴}{۳۶} = ۲۴ \Rightarrow S = \sqrt{S^2} = \sqrt{۲۴} = ۴,۹$$

تفسیر انحراف معیار:

انحراف معیار، معتبرترین و مرسوم‌ترین شاخص برای نشان دادن پراکندگی داده‌ها یا مشاهدات است. انحراف معیار نشان دهندهٔ متوسط پراکندگی مشاهدات در اطراف میانگین می‌باشد و معین می‌کند که چه نسبتی از مشاهدات در فواصل مختلف نسبت به میانگین قرار گرفته‌اند.