

آمار توصیفی

مقدمه

آمار دانشی است که به مطالعه گردآوری، تحلیل، تفسیر، ارائه و سازماندهی **داده‌ها** می‌پردازد. آمار را باید علم و عمل استخراج، بسط، و توسعه **دانشهای** تجربی **انسانی** با استفاده از روش‌های گردآوری، تنظیم، پرورش، و تحلیل **داده‌های** تجربی (حاصل از **اندازه‌گیری** و آزمایش) دانست.

علم آمار، مبتنی است بر دو شاخه **آمار توصیفی** و **آمار استنباطی**. در آمار توصیفی با داشتن تمام اعضا جامعه به بررسی خصوصیت‌های آماری آن پرداخته می‌شود در حالی که در آمار استنباطی با بدست آوردن نمونه‌ای از جامعه که خصوصیات اصلی جامعه را بیان می‌کند در مورد جامعه استنباط آماری انجام می‌شود.

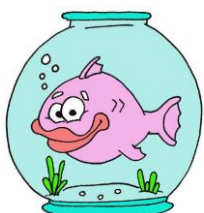
تعاریف اولیه

حل هر مسئله آماری مستلزم یادگیری چند مفهوم اولیه است که در ادامه به آن‌ها می‌پردازیم.

داده‌ها: واقعیت‌هایی درباره یک شیء یا فردند که در محاسبه، برنامه ریزی و پیش‌بینی به کار می‌روند.

متغیر: هر ویژگی از اشیا یا اشخاص، که در اعضای جامعه یکسان نیست و معمولاً از یک عضو به عضو دیگر تغییر می‌کند را **متغیر** می‌گویند و عددی که به آن ویژگی یک عضو نسبت داده می‌شود را **مقدار متغیر**، یا **مشاهده** می‌گویند.

فراوانی یک داده: تعداد دفعاتی که هر داده مشاهده می‌شود را **فراوانی** آن داده می‌گویند.



نکته: فراوانی داده را با f_i نمایش می دهند.

فراوانی نسبی یک داده: با تقسیم فراوانی هر داده به تعداد کل داده ها، فراوانی نسبی آن داده به دست می آید.

اگر فراوانی نسبی داده ها در ۱۰۰ ضرب شود، آنگاه درصد داده ها به دست می آید.

$$\text{فراوانی نسبی} = \frac{\text{فراوانی}}{\text{تعداد کل داده ها}}$$

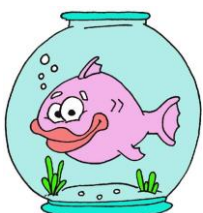
$$\text{درصد داده} = 100 \times \text{فراوانی نسبی}$$

برای مطالعه و فهم بیشتر

↓
طبقه بندی متغیرها بر حسب ماهیت متغیر

۱- متغیر کمی و کیفی

متغیر یک مفهوم است که می تواند مشاهده یا اندازه گیری شود. این اندازه گیری ممکن است به صورت کیفی یا کمی انجام شود. **متغیر کمی** به متغیری اطلاق می شود که از نظر کمیّت تغییر می کند و اختلاف مقادیر آن را می توان با استفاده از **عدد** ثبت کرد و آنها را می توان با هم جمع کرد. متغیرهای کمی، متغیرهایی هستند که انسان توانسته است برای آنها واحد و مبدأ اندازه گیری معین کند، مانند: قد، وزن، سن.



متغیر کیفی، به متغیری اطلاق می‌شود که اختلاف و تغییرات بین میزان‌های مختلف آن کیفی است و برای ثبت آن ممکن است از روش‌های دیگری غیر از به کار بردن عدد استفاده شود. به عبارت دیگر **پژوهشگر**، **توانایی اندازه‌گیری متغیر کیفی را ندارد** و ویژگی‌های آن را نمی‌تواند به وسیله ارقام ریاضی نمایش دهد. برای ثبت مشاهدات یا اندازه‌گیری‌هایی که از این متغیر به عمل می‌آید، از حروف الفبا یا کد استفاده می‌شود. این گونه متغیرها را نمی‌توان جمع و تفریق کرد و برای آنها مبدأ اندازه‌گیری نیز وجود ندارد. رنگ مو، رنگ چشم، و جنس متغیرهای کیفی هستند.

۲- متغیر گسسته و پیوسته

همان‌طور که قبلاً ذکر شد، متغیرها به وسیله عدد یا ارزش، مشخص می‌شوند. ماهیت اعداد و ارزش‌ها بستگی به این دارد که متغیر موردنظر پیوسته است یا گسسته.

متغیر گسسته می‌تواند اعداد یا ارزش‌هایی را که مشخص کننده یک وجه مشخص و معین از یک مقیاس هستند، به خود اختصاص دهد. به عنوان مثال، جنس یک متغیر گسسته است: یک شخص یا زن است یا مرد. اختصاص هر نوع ارزش دیگری بین این دو نوع ارزش امکان‌پذیر نیست. تعداد بازیکنان یک تیم فوتبال نیز یک متغیر گسسته است، زیرا فقط امکان داشتن ۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰، ۱۱ بازیکن وجود دارد و نه ۷/۵ نفر بازیکن.

متغیر پیوسته، متغیری است که بین دو واحد آن هر نقطه یا ارزشی را می‌توان انتخاب کرد. در این متغیر درجات مختلف اندازه‌گیری وجود دارد و دقت وسیله اندازه‌گیری، تعداد این درجات را تعیین می‌کند. به عنوان مثال، وزن یک متغیر پیوسته است و می‌تواند بین صفر تا بی‌نهایت باشد. قد، زمان، طول یا ارتفاع پرش، درصد چاقی بدن، و سطح هموگلوبین خون متغیرهای پیوسته هستند. ناگفته نماند که در عمل تشخیص بین متغیر پیوسته و گسسته به صورت نظری امکان‌پذیر نیست. دلیل این امر فقدان وسایل اندازه‌گیری دقیق و مناسب است. در خیلی از متغیرهای پیوسته



ما ناگزیریم اعداد را به صورت کلی برای اندازه‌گیری به کار ببریم. بهره‌ هوشی از نظر تئوری یک متغیر پیوسته است. اما در عمل، آزمودنی که برای اندازه‌گیری هوش به کار برده می‌شود، به گونه‌ای است که نمره‌ها را به صورت کلی یا نمره‌های گسسته نشان می‌دهد.

(این قسمت از سایت ایران پژوهان (www.iranresearches.ir) اتخاذ شده است.)

جدول فراوانی

اطلاعات بدست آمده از یک تحقیق غالباً **توده‌ای از اطلاعات خام، بی معنی و بدون نظم** هستند که هر نوع نتیجه‌گیری و تفسیر آنها غیر ممکن است. بنابراین برای هر نوع تجزیه و تحلیل اطلاعات لازم است داده‌ها براساس **یک نظم منطقی** طبقه‌بندی شوند تا به صورت معنی‌دار و قابل تفسیر در آید. طبقه‌بندی داده‌ها تمام اطلاعات در یک جدول به نام جدول فراوانی گردآوری می‌شود.

مراحل رسم جدول فراوانی

۱) به هر یک از متغیرها یک **عدد نسبت** می‌دهیم و آن را در ستون **شماره** می‌نویسیم.

۲) **فراوانی** هر متغیر (تعداد تکرار هر متغیر) را بدست می‌آوریم و آن را در ستون فراوانی می‌نویسیم.

۳) **فراوانی نسبی** هر متغیر را بدست می‌آوریم و آن را در ستون فراوانی نسبی می‌نویسیم.

۴) در صورت لزوم، **درصد هر متغیر** را محاسبه کرده و آن را در ستون درصد متغیر وارد می‌کنیم.



@just_jozve

دانلود از اپلیکیشن پادرس



جدول فراوانی طبقه بندی شده

اگر داده های ما **پیوسته** و یا **گسسته** با دامنه تغییرات بزرگ باشد، برای رسم جدول فراوانی داده ها ابتدا لازم است داده ها را **گروه بندی** کنیم. در گروه بندی داده ها باید موارد زیر را در نظر داشته باشید:

- (۱) گروه بندی باید به گونه ای باشد که از **کوچک ترین** داده آغاز و به **بزرگ ترین** داده ختم شود.
- (۲) **تعداد گروه ها** متناسب با **تعداد داده ها** باشد، بهتر است تعداد گروه ها از ۷ یا ۸ گروه بیشتر نشود.

انواع نمودارها

(۱) نمودارهای میله ای (ستونی)

برای رسم این نمودار ابتدا باید داده ها و یا شماره آن ها را روی **محور افقی** مشخص کنیم بعد روی نام هر داده (یا شماره آن) میله ای رسم می کنیم که طول آن متناسب با فراوانی آن داده باشد. طول میله ها می تواند بیانگر **فراوانی داده** یا **درصد** آن باشد. عرض میله ها باید با هم برابر باشد.

(۲) نمودار دایره ای

در این نوع نمودار هر **قطاعی** از دایره بیانگر **درصد** مربوط به همان داده می باشد. برای محاسبه زاویه قطاع از رابطه زیر استفاده کنید.

$$۳۶۰ \times \text{فراوانی نسبی هر داده} = \text{زاویه هر قطاع}$$



۳) نمودار بافت نگاشت

برای رسم نمودار بافت نگاشت برای **داده های پیوسته** ابتدا داده ها را گروه بندی کرده سپس محدوده هر گروه از داده ها را روی محور افقی نمودار مشخص می کنیم و روی آن محدوده مستطیلی رسم می کنیم به طوری که طول مستطیل متناسب با فراوانی یا فراوانی نسبی باشد.

نکته: از نمودار بافت نگاشت فقط برای **متغیر های پیوسته** استفاده می شود.



معیار های گرایش به مرکز

مجموع داده ها

اگر n داده x_1, x_2, \dots, x_n داشته باشیم، مجموع آن داده ها را با نماد **سیگما** (Σ) نمایش می دهیم و داریم:

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$$

و عبارت $\sum_{i=1}^n x_i$ ، سیگمای از 1 تا n ، x_i می خوانیم.

میانگین یا متوسط داده ها

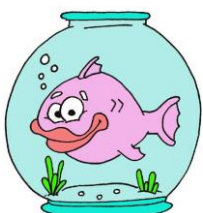
میانگین یا متوسط داده ها را با نماد \bar{x} نشان می دهیم و آن را به صورت زیر تعریف می کنیم:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

میانگین موزون داده ها

اگر n داده x_1, x_2, \dots, x_n داشته باشیم به طوری که هر یک از داده ها دارای تعداد تکرار w_1, w_2, \dots, w_n باشند که به هر یک از آنها وزن داده متناظر با آن می گوئیم. **میانگین موزون** داده ها را با نماد \bar{x}_w نشان می دهیم و آن را به صورت زیر تعریف می کنیم:

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n}$$



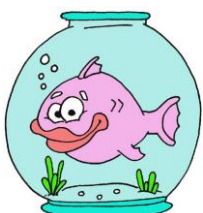
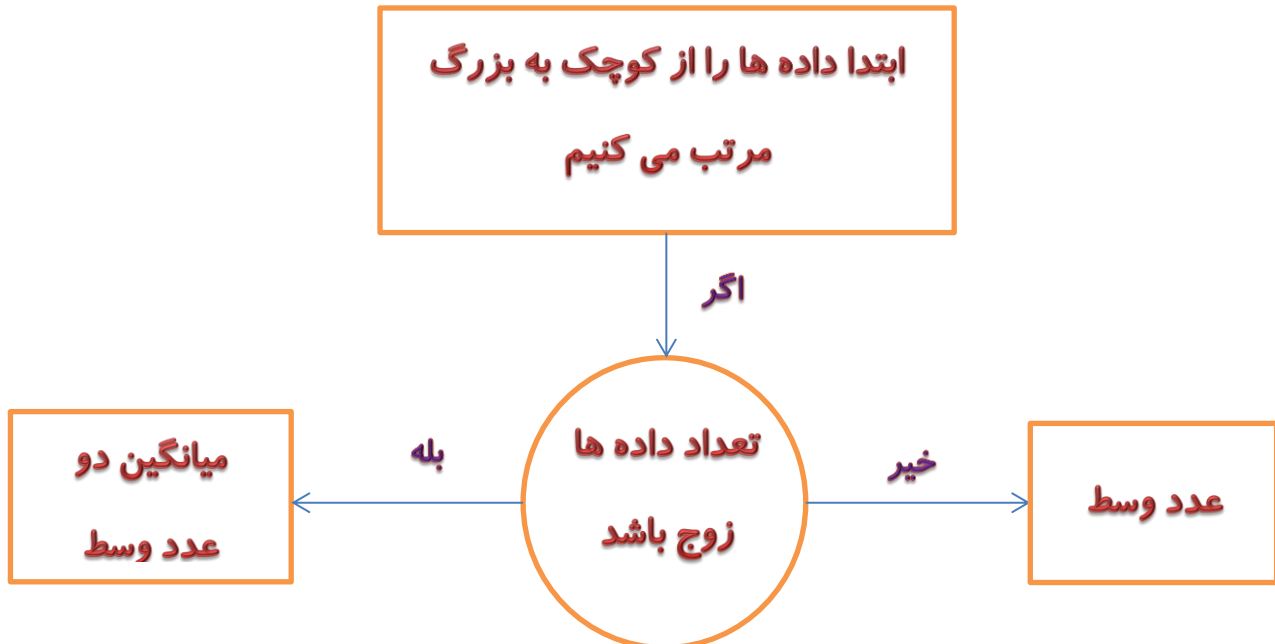
نکته: برای محاسبه میانگین داده ها در یک جدول فراوانی ، باید از رابطه بالا استفاده کنیم که w ها بیانگر تعداد تکرار داده مربوط به خودش است.

نکته: اگر جدول فراوانی ما متشکل از داده های پیوسته گروه بندی شده باشد، در رابطه $\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$ به جای x_i ها ، **مرکز دسته** را قرار می دهیم.

میان، چارک اول و چارک سوم:

عدد وسط مجموعه ای از داده ها را که از کوچک به بزرگ مرتب شده باشند **میان** داده ها می گوئیم و آن را با نماد Q_2 نشان می دهیم. میان **یک چهارم** اول داده های مرتب شده را چارک اول داده ها گوئیم و آن را با Q_1 نشان می دهیم. همچنین میان **سه چهارم** داده های مرتب شده را **چارک سوم** گوئیم و آن را با Q_3 نشان می دهیم.

نحوه بدست آوردن میان داده ها



راهنمایی

برای به دست آوردن میانه، چارک اول و چارک سوم، ابتدا داده ها را از کوچک به بزرگ مرتب می کنیم بعد میانه آن ها را بدست می آوریم (چارک دوم) بعد میانه داده های قبل از میانه را بدست می آوریم و آن را **چارک اول** می نامیم و در نهایت میانه داده های بعد از میانه را بدست می آوریم و آن را **چارک سوم** می نامیم!!

مد

داده ای که **بیش ترین فراوانی** را داشته باشد، مد یا نما داده ها نام دارد.

نکته: اگر در داده هایی، همه داده ها یک فراوانی داشته باشند، آن گاه این داده ها **مد ندارند**. و اگر در داده هایی، دو داده بیشترین فراوانی را داشته باشند، آن گاه این داده ها **دو مد** دارند.

داده دور افتاده:

داده ای که مقدار متفاوت با سایر داده ها را دارد، **داده دور افتاده** نامیده می شود. منظور از مقدار متفاوت، تفاوت بسیار آن داده با سایر داده ها می باشد.

نکته: داده دور افتاده **میانگین** داده ها را تحت تاثیر قرار می دهد در حالیکه تاثیری بر میانه و مد ندارد.

نکته: در تفسیر و تحلیل مسائل آماری، در نظر گرفتن تنها یک شاخص گرایش به مرکز کافی نیست. می بایست هر سه معیار میانگین، میانه و مد محاسبه شود و بر اساس هدف مورد بررسی، معیار مناسب انتخاب و از آن برای انجام تفسیر، قضاوت و پیش بینی مورد استفاده قرار گیرد.



نکته: اگر تمامی داده ها با عدد **a** جمع شوند، همه ی معیار های گرایش به مرکز (میانگین، میانه و مد) نیز با عدد **a** جمع می شود.

نکته: اگر تمامی داده ها در عدد **a** ضرب شوند، همه ی معیار های گرایش به مرکز (میانگین، میانه و مد) نیز در عدد **a** ضرب می شوند.

@just_jozve

