بسمه تعالی

موضوع : مبانی آمار و احتمال برای ورود به علم داده و تحلیل داده

بخش چهارم:

- روش‌های جمع‌آوری داده
- سرشماری و نمونه‌گیری

در این پارت از اموزش به بررسی **روش های جمع اوری دیتا و سازماندهی آنها** می پردازیم .

در جلسات گذشته با **جامعه آماری، نمونه آماری و انواع متغیرها** آشنا شدیم.

اما یک سؤال مهم مطرح می‌شود.

فرض کنید قصد دارید یک پژوهش انجام دهید یا عملکرد یک سازمان را تحلیل کنید.

قبل از هر نوع تحلیل آماری، ابتدا باید داده‌ها را از کجا تهیه کنیم؟

اگر داده‌های ما صحیح نباشند، حتی بهترین روش‌های آماری نیز نمی‌توانند نتیجه درستی ارائه دهند.

به همین دلیل گفته می‌شود:

"کیفیت تحلیل آماری، به کیفیت داده‌های جمع‌آوری‌شده بستگی دارد."

برای همین قبل از ورود به بخش امار توصیفی و استنتاجی به بررسی مرحله جمع آوری دیتا و پیش پردازش آن می پردازیم:

در گام اول بهتر است فرق بین داده و اطلاعات را بررسی کنیم:

داده یا Data به اطلاعات خامی گفته می‌شود که هنوز تحلیل نشده‌اند. زمانی که این داده‌ها را خلاصه و تحلیل کنیم و به نتیجه برسیم، آن‌ها به **اطلاعات (Information)** تبدیل می‌شوند.

داده ها را می توانیم به طرق مختلف جمع آوری کرد:



حال که داده جمع آوری شد ممکن است این داده ها ایراد و خطا داشته باشه . بنابراین به روشی نیاز است که ارور داده ها را برطرف کند و به اصطلاح داده ها را تمیز کند . روند این بدین صورت است که :

مرحله اول: مرتب‌سازی داده‌ها

ابتدا داده‌ها را بر اساس مقدار مرتب می‌کنیم.

مثلاً از کوچک به بزرگ.

این کار باعث می‌شود الگوی داده‌ها راحت‌تر دیده شود.

مرحله دوم: کدگذاری داده‌ها

گاهی لازم است داده‌های کیفی را به صورت عددی ثبت کنیم.

مرحله سوم: تشکیل جدول داده‌ها

معمولاً داده‌ها به صورت سطری و ستونی ثبت می‌شوند.

هر سطر نشان‌دهنده یک مشاهده است.

هر ستون یک متغیر را نمایش می‌دهد.

مرحله چهارم: کنترل کیفیت داده‌ها

قبل از شروع تحلیل باید داده‌ها بررسی شوند.

برای مثال:

- آیا داده‌ای ثبت نشده است؟
- آیا عدد اشتباه وارد شده است؟
- آیا داده پرت وجود دارد؟
- آیا کدگذاری صحیح انجام شده است؟

اگر این مرحله انجام نشود،

نتایج تحلیل آماری می‌تواند کاملاً اشتباه باشد.

هر کدام از مراحل گفته شده خودشون دارای روش های مختلف هستند که به دلیل گستردگی موضوع ازش عبور می کنیم.