

## چند متغیره پیوسته و گسسته

معمولا پژوهشگران علوم زیست شناسی، فیزیک، علوم رفتاری و اجتماعی، پزشکان و سایر افراد درمورد موضوع مورد مطالعه خود، داده هایی را مربوط به چندین متغیر جمع آوری می کنند. به عبارت دیگر به جای داشتن یک متغیر، چند متغیر دیگر را مورد مطالعه قرار می دهند. داده های مربوط به این متغیرها را، داده های چند متغیره می نامند و از روشهای آماری چند متغیره، برای تجزیه و تحلیل آنها استفاده می کنند. این متغیرها یا پیوسته هستند یا گسسته و یا برخی پیوسته و برخی دیگر گسسته. به عنوان مثال یک پزشک همزمان قد، وزن، جنسیت، گروه خونی و فشار خون یک بیمار را جهت بررسی یک نوع بیماری مورد مطالعه قرار می دهد. در ادامه با یکی از روشهای آماری چند متغیره آشنا می شویم.

### تحلیل مولفه های اصلی (PCA) Principle Component Analysis

تحلیل مولفه های اصلی یکی از انواع روشهای تحلیل داده های چند متغیره است که هدف اصلی آن تقلیل بعد مساله مورد مطالعه است. یکی از کاربردهای مهم تحلیل مولفه های اصلی، در رگرسیون است. با استفاده از تحلیل مولفه های اصلی می توان تعداد زیادی متغیر توضیحی (متغیر مستقل) همبسته را با تعداد محدودی متغیر توضیحی جدید که مولفه های اصلی نامیده می شوند و ناهمبسته اند، جایگزین نمود. به این ترتیب نه تنها بعد مساله تقلیل می یابد بلکه مساله چند همخطی پیش نمی آید.

#### مولفه های اصلی

فرض کنید  $\underline{X} = (X_1, X_2, \dots, X_p)^T$  یک بردار تصادفی با ماتریس کواریانس معین نامنفی  $\Sigma$  و  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  مقادیر ویژه  $\Sigma$  باشد. همچنین  $a_1, a_2, \dots, a_p$  بردارهای ویژه یکامتعامد به ترتیب متناظر با  $\lambda_1, \lambda_2, \dots, \lambda_p$  باشند. متغیرهای  $Y_1, Y_2, \dots, Y_p$  تعریف شده در زیر را مولفه های اصلی گویند.

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$$

$$Y_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p$$

$$Y_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p$$

$Y_i$  را مولفه اصلی  $i$ ام گویند. بردار  $\underline{Y} = (Y_1, Y_2, \dots, Y_p)^T$  بردار مولفه های اصلی نام دارد.