



دانشگاه سindh  
پي

# آمار و داده پردازی

(کارشناسی ارشد مهندسی کشاورزی)

دکتر امیر تیمور پاینده نجف آبادی      دکتر مریم امیدی نجف آبادی

امروزه کتاب‌خوانی و علم‌آموزی، نه تنها یک وظیفه‌ی ملی، که یک واجب دینی است.

### مقام معظم رهبری

در عصر حاضر یکی از شاخصه‌های ارزیابی رشد، توسعه و پیشرفت فرهنگی هر کشوری میزان تولید کتاب، مطالعه و کتاب‌خوانی مردم آن مرز و بوم است. ایران اسلامی نیز از دیرباز تاکنون با داشتن تمدنی چندهزارساله و مراکز متعدد علمی، فرهنگی، کتابخانه‌های معتبر، علما و دانشمندان بزرگ با آثار ارزشمند تاریخی، سرآمد دولت‌ها و ملت‌های دیگر بوده و در عرصه‌ی فرهنگ و تمدن جهانی به‌سان خورشیدی تابناک همچنان می‌درخشد و با فرزندان نیک‌نهاد خویش هنرنمایی می‌کند. چه کسی است که در دنیا با دانشمندان فرزانه و نام‌آور ایرانی همچون ابوعلی سینا، ابوریحان بیرونی، فارابی، خوارزمی و ... همچنین شاعران برجسته‌ای نظیر فردوسی، سعدی، مولوی، حافظ و ... آشنا نباشد و در مقابل عظمت آنها سر تعظیم فرود نیاورد. تمامی این افتخارات ارزشمند، برگرفته از میزان عشق و علاقه فراوان ملت ما به فراگیری علم و دانش از طریق خواندن و مطالعه منابع و کتاب‌های گوناگون است. به شکرانه‌ی الهی، تاریخ و گذشته ما، همیشه درخشان و پر بار است. ولی اکنون در این زمینه در چه جایگاهی قرار داریم؟ آمار و ارقام ارائه‌شده از سوی مجامع و سازمان‌های فرهنگی در مورد سرانه‌ی مطالعه‌ی هر ایرانی، برایمان چندان امیدوارکننده نمی‌باشد و رهبر معظم انقلاب اسلامی نیز از این وضعیت بارها اظهار گله و ناخشنودی نموده‌اند.

کتاب، دروازه‌ای به سوی گستره‌ی دانش و معرفت است و کتاب خوب، یکی از بهترین ابزارهای کمال بشری است. همه‌ی دستاوردهای بشر در سراسر عمر جهان، تا آنجا که قابل کتابت بوده است، در میان دست‌نوشته‌هایی است که انسان‌ها پدید آورده و می‌آورند. در این مجموعه‌ی بی‌نظیر، تعالیم الهی، درس‌های پیامبران به بشر، و همچنین علوم مختلفی است که سعادت بشر بدون آگاهی از آنها امکان‌پذیر نیست. کسی که با دنیای زیبا و زندگی‌بخش کتاب ارتباط ندارد بی‌شک از مهم‌ترین دستاورد انسانی و نیز از بیشترین معارف الهی و بشری محروم است. با این دیدگاه، به‌روشنی می‌توان ارزش و مفهوم رمزی عمیق در این حقیقت تاریخی را دریافت که اولین خطاب خداوند متعال به پیامبر گرامی اسلام (ص) این است که «بخوان!» و در اولین

سوره‌ای که بر آن فرستاده‌ی عظیم‌الشان خداوند، فرود آمده، نام «قلم» به تجلیل یاد شده‌است: «اقْرَأْ وَ رَبُّكَ الْأَكْرَمُ. الَّذِي عَلَّمَ بِالْقَلَمِ» در اهمیت عنصر کتاب برای تکامل جامعه‌ی انسانی، همین بس که تمامی ادیان آسمانی و رجال بزرگ تاریخ بشری، از طریق کتاب جاودانه مانده‌اند.

دانشگاه پیام‌نور با گستره‌ی جغرافیایی ایران‌شمول خود با هدف آموزش برای همه، همه‌جا و همه‌وقت، به‌عنوان دانشگاهی کتاب‌محور در نظام آموزش عالی کشورمان، افتخار دارد جایگاه اندیشه‌سازی و خردورزی بخش عظیمی از جوانان جویای علم این مرز و بوم باشد. تلاش فراوانی در ایام طولانی فعالیت این دانشگاه انجام پذیرفته تا با بهره‌گیری از تجربه‌های گرانقدر استادان و صاحب‌نظران برجسته کشورمان، کتاب‌ها و منابع آموزشی درسی شاخص و خودآموز تولید شود. در آینده هم، این مهم با هدف ارتقای سطح علمی، روزآمدی و توجه بیشتر به نیازهای مخاطبان دانشگاه پیام‌نور با جدیت ادامه خواهد داشت. به‌طور قطع استفاده از نظرات استادان، صاحب‌نظران و دانشجویان محترم، ما را در انجام این وظیفه‌ی مهم و خطیر یاری‌رسان خواهد بود. پیشاپیش از تمامی عزیزانی که با نقد، تصحیح و پیشنهادهای خود ما را در انجام این وظیفه‌ی خطیر یاری می‌رسانند، سپاسگزاری می‌نماییم. لازم است از تمامی اندیشمندانی که تاکنون دانشگاه پیام‌نور را منزلگاه اندیشه‌سازی خود دانسته و ما را در تولید کتاب و محتوای آموزشی درسی یاری نموده‌اند، صمیمانه قدردانی گردد. موفقیت و بهروزی تمامی دانشجویان و دانش‌پژوهان عزیز آرزوی همیشگی ما است.

دانشگاه پیام‌نور

## فهرست مطالب

پیشگفتار ..... یازده

|                                                             |   |
|-------------------------------------------------------------|---|
| فصل اول. آمار توصیفی: کلیات، شاخص‌ها و نمودارهای آماری..... | ۱ |
| هدف کلی.....                                                | ۱ |
| هدف‌های یادگیری.....                                        | ۱ |
| مقدمه.....                                                  | ۱ |
| ۱-۱-۱ مراحل مختلف تحقیق.....                                | ۲ |
| ۱-۱-۱-۱ مشخص کردن هدف.....                                  | ۲ |
| ۱-۱-۲ جمع‌آوری اطلاعات.....                                 | ۲ |
| ۱-۱-۳ تحلیل اطلاعات.....                                    | ۳ |
| ۱-۱-۴ مرحله نتیجه‌گیری.....                                 | ۳ |
| ۲-۱ مفاهیم اولیه علم آمار.....                              | ۳ |
| ۱-۲-۱ انواع روش‌های آماری.....                              | ۳ |
| ۲-۲-۱ جامعه آماری.....                                      | ۴ |
| ۳-۲-۱ نمونه.....                                            | ۴ |
| ۳-۱ اندازه‌گیری و مقیاس‌سازی.....                           | ۴ |
| ۱-۳-۱ مقیاس اسمی.....                                       | ۴ |
| ۲-۳-۱ مقیاس ترتیبی.....                                     | ۵ |
| ۳-۳-۱ مقیاس فاصله‌ای.....                                   | ۵ |
| ۴-۳-۱ مقیاس نسبی.....                                       | ۵ |
| ۴-۱ انواع فراوانی.....                                      | ۶ |
| ۵-۱ مشاهده یا داده سانسور شده.....                          | ۷ |
| ۶-۱ مشاهده یا داده گم‌شده.....                              | ۷ |

|                                          |    |
|------------------------------------------|----|
| ۷-۱ داده دورافتاده.....                  | ۷  |
| ۸-۱ متغیر تصادفی.....                    | ۸  |
| ۸-۱-۱ متغیر تصادفی و انواع آن.....       | ۸  |
| ۹-۱ خلاصه سازی و مقایسه داده ها.....     | ۹  |
| ۱۰-۱ انواع شاخص های تمرکز.....           | ۱۰ |
| ۱۰-۱-۱ نما.....                          | ۱۰ |
| ۱۰-۱-۲ میانه.....                        | ۱۰ |
| ۱۰-۱-۳ چندک $p$ .....                    | ۱۰ |
| ۱۰-۱-۴ میانگین.....                      | ۱۰ |
| ۱۱-۱ انواع شاخص های پراکندگی.....        | ۱۱ |
| ۱۱-۱-۱ دامنه تغییرات.....                | ۱۱ |
| ۱۱-۱-۲ واریانس.....                      | ۱۱ |
| ۱۲-۱ شاخص های مقایسه.....                | ۱۲ |
| ۱۲-۱-۱ ضریب تغییرات.....                 | ۱۲ |
| ۱۳-۱ نمودارهای آماری.....                | ۱۵ |
| ۱۳-۱-۱ نمودار دایره ای.....              | ۱۵ |
| ۱۳-۱-۲ نمودار میله ای.....               | ۱۶ |
| ۱۳-۱-۳ نمودار بافت نگار (هیستوگرام)..... | ۱۷ |
| ۱۳-۱-۴ نمودار جعبه ای.....               | ۱۸ |
| ۱۳-۱-۵ نمودار پراکنش.....                | ۲۰ |
| ۱۳-۱-۶ نمودارهای چندمتغیره.....          | ۲۱ |
| ۱۳-۱-۷ نمودار سری زمانی.....             | ۲۲ |
| خودآزمایی تشریحی فصل اول.....            | ۳۲ |

|                                                      |    |
|------------------------------------------------------|----|
| فصل دوم. معرفی اجمالی برخی از توزیع های احتمالی..... | ۳۵ |
| هدف کلی.....                                         | ۳۵ |
| هدف های یادگیری.....                                 | ۳۵ |
| مقدمه.....                                           | ۳۵ |
| ۱-۲ توزیع های مهم گسسته.....                         | ۳۶ |
| ۱-۲-۱ توزیع دو جمله ای.....                          | ۳۶ |
| ۱-۲-۲ توزیع هندسی.....                               | ۳۷ |
| ۱-۲-۳ توزیع دو جمله ای منفی.....                     | ۳۷ |
| ۱-۲-۴ توزیع پواسون.....                              | ۳۸ |
| ۲-۲ توزیع های مهم پیوسته.....                        | ۳۸ |
| ۲-۲-۱ توزیع نمایی.....                               | ۳۸ |
| ۲-۲-۲ توزیع گاما.....                                | ۳۹ |
| ۲-۲-۳ توزیع نرمال.....                               | ۴۰ |
| ۲-۲-۴ توزیع $t$ -استیودنت.....                       | ۴۱ |
| ۲-۲-۵ توزیع $F$ (یا توزیع فیشر).....                 | ۴۲ |

|       |                                            |    |
|-------|--------------------------------------------|----|
| ۳-۲   | محاسبه احتمال با استفاده از نرم افزار SPSS | ۴۲ |
| ۴-۲   | روش پیدا کردن توزیع یک متغیر تصادفی        | ۴۷ |
| ۵-۲   | برآورد پارامترهای یک توزیع احتمالی         | ۴۹ |
| ۶-۲   | برآورد پارامتر مجهول یک توزیع آماری        | ۵۰ |
| ۱-۶-۲ | برآورد نقطه‌ای                             | ۵۰ |
| ۲-۶-۲ | برآورد بازه‌ای                             | ۵۰ |
| ۵۳    | خودآزمایی تشریحی فصل دوم                   | ۵۳ |

|                                                                |    |
|----------------------------------------------------------------|----|
| فصل سوم. آزمون فرض در مورد میانگین یک یا دو جامعه آماری نرمال  | ۵۵ |
| هدف کلی                                                        | ۵۵ |
| هدف‌های یادگیری                                                | ۵۵ |
| مقدمه                                                          | ۵۵ |
| ۱-۳ آزمون فرض در مورد میانگین یک جامعه آماری نرمال             | ۵۸ |
| ۱-۳-۱ چگونگی انجام آزمون شماره یک با استفاده از نرم افزار SPSS | ۵۸ |
| ۲-۳ مقایسه دو جامعه آماری نرمال براساس آزمون مقایسه میانگین‌ها | ۶۲ |
| ۲-۳-۱ روش اول (انتخاب تصادفی از دو جامعه)                      | ۶۳ |
| ۲-۳-۲ روش دوم (روش زوج‌های جور شده)                            | ۶۴ |
| خودآزمایی تشریحی فصل سوم                                       | ۶۹ |

|                                                         |     |
|---------------------------------------------------------|-----|
| فصل چهارم. مدل سازی رگرسیونی و تحلیل مسیر               | ۷۱  |
| هدف کلی                                                 | ۷۱  |
| هدف‌های یادگیری                                         | ۷۱  |
| مقدمه                                                   | ۷۱  |
| ۱-۴ رگرسیون خطی                                         | ۷۲  |
| ۲-۴ همبستگی خطی                                         | ۷۲  |
| ۳-۴ تحلیل رگرسیونی                                      | ۷۸  |
| ۴-۴ مدل رگرسیون خطی ساده                                | ۷۸  |
| ۵-۴ مدل رگرسیون خطی چندگانه                             | ۸۹  |
| ۶-۴ چگونگی معرفی مدل رگرسیونی چندگانه به نرم افزار SPSS | ۹۰  |
| ۷-۴ انتخاب بهترین مدل رگرسیون خطی چندگانه               | ۹۱  |
| ۸-۴ تحلیل مسیر                                          | ۱۰۰ |
| خودآزمایی تشریحی فصل چهارم                              | ۱۰۶ |

|                               |     |
|-------------------------------|-----|
| فصل پنجم. تحلیل عاملی اکتشافی | ۱۰۹ |
| هدف کلی                       | ۱۰۹ |
| هدف‌های یادگیری               | ۱۰۹ |
| مقدمه                         | ۱۰۹ |
| ۱-۵ مفاهیم و تعریف مقدماتی    | ۱۱۰ |

|     |                                                                |
|-----|----------------------------------------------------------------|
| ۱۱۱ | ۲-۵ تحلیل عاملی اکتشافی کلاسیک                                 |
| ۱۱۱ | ۱-۲-۵ فرضیات اساسی تحلیل عاملی اکتشافی                         |
| ۱۱۲ | ۲-۲-۵ چگونگی انجام تحلیل عاملی اکتشافی با نرم افزار SPSS       |
| ۱۲۰ | خودآزمایی تشریحی فصل پنجم                                      |
| ۱۲۷ | فصل ششم. کاربردهای آزمون خی دو و روش های ناپارامتری            |
| ۱۲۷ | هدف کلی                                                        |
| ۱۲۷ | هدف های یادگیری                                                |
| ۱۲۷ | ۱-۶ آزمون خی دو                                                |
| ۱۲۷ | ۱-۱-۶ کاربرد ۱. در آزمون نیکویی برازش                          |
| ۱۳۰ | ۱-۲-۶ کاربرد ۲. در آزمون استقلال جدول های پیشایندی             |
| ۱۳۴ | ۲-۶ روش های ناپارامتری                                         |
| ۱۳۵ | ۱-۲-۶ آزمون کلموگروف - اسمیرنوف                                |
| ۱۳۶ | ۲-۲-۶ آزمون های ناپارامتری برای مقایسه دو یا چند جامعه آماری   |
| ۱۴۷ | خودآزمایی تشریحی فصل ششم                                       |
| ۱۴۹ | فصل هفتم. رگرسیون های لوژستیک                                  |
| ۱۴۹ | هدف کلی                                                        |
| ۱۴۹ | هدف های یادگیری                                                |
| ۱۴۹ | مقدمه                                                          |
| ۱۵۰ | ۱-۷ رگرسیون لوژستیک (ساده یا دوتایی)                           |
| ۱۵۲ | ۱-۱-۷ برآورد پارامترهای مدل رگرسیون لوژستیک                    |
| ۱۵۳ | ۲-۱-۷ فرضیات مدل رگرسیون لوژستیک و آماره های نیکویی برازش      |
| ۱۵۵ | ۳-۱-۷ استنباط در مورد ضرایب مدل های رگرسیون لوژستیک            |
| ۱۵۵ | ۴-۱-۷ استفاده از نرم افزار SPSS برای برازش مدل رگرسیون لوژستیک |
| ۱۶۳ | ۵-۱-۷ تحلیل خروجی نرم افزار                                    |
| ۱۶۷ | ۲-۷ رگرسیون لوژستیک اسمی                                       |
| ۱۷۸ | ۳-۷ رگرسیون لوژستیک ترتیبی                                     |
| ۱۹۳ | خودآزمایی تشریحی فصل هفتم                                      |
| ۱۹۵ | فصل هشتم. سری های زمانی                                        |
| ۱۹۵ | هدف کلی                                                        |
| ۱۹۵ | هدف های یادگیری                                                |
| ۱۹۵ | مقدمه                                                          |
| ۱۹۶ | ۱-۸ مفاهیم و تعریف مقدماتی                                     |
| ۱۹۶ | ۱-۱-۸ تعریف فرایند تصادفی                                      |
| ۱۹۹ | ۲-۱-۸ تعریف خودهمبستگی                                         |
| ۱۹۹ | ۳-۱-۸ تعریف خودهمبستگی جزئی                                    |

|     |                                                                                                 |
|-----|-------------------------------------------------------------------------------------------------|
| ۲۰۱ | ۸-۱-۴ تعریف شوک تصادفی                                                                          |
| ۲۰۲ | ۸-۲ مدل‌های ARIMA غیرفصلی                                                                       |
| ۲۰۲ | ۸-۲-۱ مدل‌های اتورگرسیو                                                                         |
| ۲۰۲ | ۸-۲-۲ مدل‌های میانگین متحرک                                                                     |
| ۲۰۳ | ۸-۲-۳ مدل‌های اتورگرسیو - میانگین متحرک ARMA                                                    |
| ۲۰۳ | ۸-۲-۴ چگونگی برازش یک مدل $ARIMA(p,d,q)$ به یک دسته داده سری‌زمانی با استفاده از نرم‌افزار SPSS |
| ۲۰۵ | ۸-۲-۵ چگونگی تعیین رتبه‌های یک مدل ARIMA                                                        |
| ۲۰۹ | ۸-۲-۶ چگونگی ارزیابی مدل برازش‌شده                                                              |
| ۲۱۱ | ۸-۲-۷ پیش‌بینی رفتار سری‌زمانی                                                                  |
| ۲۱۱ | ۸-۳ مدل‌های ARIMA فصلی                                                                          |
| ۲۱۸ | ۸-۴ مدل‌های رگرسیونی-ARIMA                                                                      |
| ۲۲۲ | خودآزمایی تشریحی فصل هشتم                                                                       |
| ۲۲۵ | فصل نهم. رگرسیون چندمتغیره چندگانه                                                              |
| ۲۲۵ | هدف کلی                                                                                         |
| ۲۲۵ | هدف‌های یادگیری                                                                                 |
| ۲۲۵ | مقدمه                                                                                           |
| ۲۲۷ | ۹-۱ رگرسیون چند متغیره چندگانه                                                                  |
| ۲۲۷ | ۹-۱-۱ فرضیات یک مدل رگرسیونی چندمتغیره چندگانه                                                  |
| ۲۲۸ | ۹-۲ بررسی نرمال چندمتغیره بودن خطاها                                                            |
| ۲۳۴ | ۹-۳ چگونگی برازش یک مدل رگرسیون چندمتغیره چندگانه                                               |
| ۲۳۹ | ۹-۳-۱ بررسی صحت فرضیات مدل رگرسیونی چندمتغیره چندگانه                                           |
| ۲۴۳ | خودآزمایی تشریحی فصل نهم                                                                        |
| ۲۴۵ | فصل دهم. تحلیل واریانس (کوواریانس) یک یا چندمتغیره                                              |
| ۲۴۵ | هدف کلی                                                                                         |
| ۲۴۵ | هدف‌های یادگیری                                                                                 |
| ۲۴۵ | مقدمه                                                                                           |
| ۲۴۵ | ۱۰-۱ مقایسه بیش از دو جامعه نرمال براساس یک صفت                                                 |
| ۲۴۶ | ۱۰-۱-۱ تحلیل واریانس یک‌متغیره ANOVA                                                            |
| ۲۵۹ | ۱۰-۱-۲ تحلیل کواریانس یک متغیره ANCOVA                                                          |
| ۲۶۹ | ۱۰-۲ مقایسه بیش از دو جامعه نرمال براساس چند صفت                                                |
| ۲۶۹ | ۱۰-۲-۱ تحلیل واریانس چندمتغیره MANOVA                                                           |
| ۲۸۳ | ۱۰-۲-۲ تحلیل کواریانس چندمتغیره MANCOVA                                                         |
| ۲۹۴ | خودآزمایی تشریحی فصل دهم                                                                        |
| ۲۹۷ | پیوست الف. چگونگی برقرارکردن فرضیه‌های مدل رگرسیونی                                             |

خودآزمایی تشریحی پیوست الف ..... ۳۱۰

پیوست ب. مقدمه‌ای بر نرم‌افزار SPSS ..... ۳۱۱

مقدمه ..... ۳۱۱

ب. ۱. معرفی محیط‌های نرم‌افزار ..... ۳۱۱

ب. ۱-۱ محیط داده ..... ۳۱۲

ب. ۱-۲ محیط خروجی ..... ۳۱۴

ب. ۱-۳ محیط ویرایش ..... ۳۱۵

منابع ..... ۳۲۳

## پیشگفتار

آمار به عنوان ابزاری توانمند و کارا، محققان را قادر به انجام تحقیقات خود می‌سازد. مانند هر علم دیگر، کاربران آمار برای استفاده از آن، مستلزم آشنایی با مفاهیم اولیه آمار هستند. متأسفانه در بسیاری از تحقیقات علوم اجتماعی بدون توجه به زیرساخت‌ها و پیش‌نیازهای روش‌های آماری مبادرت به استفاده از آن روش‌ها می‌کنند و بارها مشاهده شده‌است که کاربران آمار بدون توجه به ماهیت متغیرها مبادرت به استفاده از تحلیل رگرسیونی می‌نمایند درحالی‌که متغیرهای تحقیق، ترتیبی است و نمی‌توان از مدل‌سازی رگرسیونی معمولی استفاده کرد. یا مثلاً برای خلاصه‌سازی داده‌های ترتیبی از شاخص‌های میانگین یا واریانس (که شاخص‌های مناسب برای داده‌های نسبتی یا فاصله‌ای است) استفاده شده‌است.

نویسندگان این کتاب براساس تجربه تدریس و تحقیق در علوم اجتماعی، خلأ وجود کتابی ساده در عین حال دقیق برای داده‌پردازی را احساس کرده و مبادرت به نگارش این کتاب نمودند. در این کتاب سعی شده‌است که روش‌های آماری با حداقل فرمول‌های ریاضی و چگونگی استفاده از نرم‌افزار SPSS ارائه شود.

فصل اول کتاب حاضر به بررسی آمار توصیفی و روش‌های خلاصه‌سازی داده‌ها می‌پردازد. مقدمه‌ای ساده بر نظریه احتمال، رسم نمودارهای احتمالی و برآورد پارامترهای یک توزیع احتمالی در فصل دوم ارائه شده‌است. آزمون فرض در مورد میانگین یک یا دو جامعه آماری در فصل سوم مورد بررسی قرار گرفته‌است. مدل‌سازی رگرسیونی و تحلیل مسیر به همراه روش‌های بررسی صحت فرضیات یک مدل رگرسیونی در فصل چهارم به زبان ساده و در عین حال دقیق ارائه شده‌است. فصل‌های

پنجم و ششم به ترتیب تحلیل عاملی اکتشافی و آزمون‌های نیکویی برازش را به همراه روش‌های ناپارامتری مورد بررسی قرار داده‌است. انواع رگرسیون‌های لوژستیک در فصل هفتم مورد بررسی قرار گرفته‌اند. سری‌های زمانی ARIMA به دو صورت فصلی و غیرفصلی، در فصل هشتم معرفی شده‌اند. فصل نهم به آموزش رگرسیون‌های چندمتغیره می‌پردازد. سرانجام آنالیز واریانس و کوواریانس یک و چندمتغیره در فصل دهم مورد بررسی قرار گرفته‌اند. ضمناً در دو پیوست این کتاب، چگونگی برقرارکردن فرضیه‌های مدل رگرسیونی به همراه مقدمه‌ای بر نرم‌افزار SPSS گنجانده شده‌است. همانند هر اثر دیگری، این اثر نیز علی‌رغم صرف زمان زیاد جهت نگارش آن، عاری از نقص نیست. قبلاً از خوانندگان این کتاب که با پیشنهادات و نظرات اصلاحی خود باعث ارتقا کیفیت این اثر خواهند شد قدردانی می‌کنیم. امیدواریم که این کتاب مورد استفاده دانشجویان و محققان عزیز قرار گیرد.

دکتر امیر تیمور پاینده نجف‌آبادی

دانشیار دانشگاه شهید بهشتی

دکتر مریم امیدی نجف‌آبادی

دانشیار دانشگاه آزاد واحد علوم و

تحقیقات تهران

# فصل اول

## آمار توصیفی: کلیات، شاخص‌ها و نمودارهای آماری

### هدف کلی

بررسی مفاهیم اساسی آمار توصیفی

### هدف‌های یادگیری

انتظار می‌رود دانشجو پس از مطالعه این فصل بتواند:

۱. انواع مقیاس‌های اندازه‌گیری را بیان کند.
۲. مفهوم آمارهای تمرکز و پراکندگی را بیان کند.
۳. موارد استفاده از هر نمودار آماری را بیان کند.

### مقدمه

هر فرد متناسب با کار و علائق خود به‌گونه‌ای از علم آمار استفاده می‌کند. مثلاً یک مدیر در هنگام توجیه کارایی شیوه مدیریتی خود، ناگزیر به استفاده از علم آمار است. مروج کشاورزی برای قانع کردن کشاورزان در مورد مؤثر بودن شیوه جدید کاشت گندم، با استفاده از علم آمار، شیوه جدید را با شیوه‌های قبلی مقایسه می‌کند. بنابراین به‌سادگی نمی‌توان تعریف جامع و کاملی، که بتواند تمام جنبه‌های علم آمار را در برگیرد، برای علم آمار ارائه کرد. ولی تعریف زیر تعریفی تا اندازه‌ای دقیق برای علم آمار ارائه می‌کند.

«آمار علم جمع‌آوری، دسته‌بندی، خلاصه‌سازی و بالاخره نتیجه‌گیری از داده‌ها است.»

آمار به صورت توصیفی، اولین بار توسط بابلیان و مصریان مورد استفاده قرار گرفت. سپس هخامنشیان در دربار خود طبقه‌ای به نام دبیران ایجاد کردند که کار این طبقه، جمع‌آوری اطلاعات و تهیه گزارش در مورد تولد و مرگ، خراج سالیانه و امور قشون و غیره بود.

منشأ پیدایش علم کلاسیک ریاضی‌دانان بزرگی نظیر فرما و پاسکال بودند. با استفاده از مسائل حل شده توسط این ریاضی‌دانان، محققان دیگر نظیر مندل و گالتون برخی مفاهیم علم اساسی آمار نظیر همبستگی، نظریه رگرسیون و قانون وراثت را پایه‌گذاری کردند. فیشر از بزرگ‌ترین آماردانانی است که بسیاری از مفاهیم کلاسیک علم آمار را به طور دقیق تعریف و روش‌های محاسباتی برای آن‌ها ارائه کرده است.

در سال‌های اخیر، علم آمار در شاخه‌های بسیار متنوع نظیر نمونه‌گیری، نظریه تصمیم، آمار بیز، آمار اجتماعی و غیره به سرعت در حال رشد و توسعه است. علم آمار در تمام تحقیقاتی که بر پایه مشاهدات انجام می‌پذیرد، نقشی بسیار پررنگ و غیرقابل انکار ایفا می‌کند. در ادامه مراحل انجام یک تحقیق مورد بررسی قرار می‌گیرد و نقش علم آمار در هر مرحله یادآوری می‌شود.

## ۱-۱ مراحل مختلف تحقیق

به زبان ساده مراحل اصلی یک تحقیق علمی مبتنی بر مشاهدات را می‌توان به صورت زیر مرحله‌بندی کرد:

### ۱-۱-۱ مشخص کردن هدف

در این مرحله محقق به همراه تیمی تحقیقاتی پس از جلسات متعدد و تبادل اطلاعات (اتاق فکر) هدف اصلی تحقیق و فرضیات تحقیق را به طور دقیق به همراه جامعه آماری مشخص می‌کند.

### ۱-۱-۲ جمع‌آوری اطلاعات

در این مرحله، پس از مشورت با یک متخصص آمار (آماردان) چگونگی جمع‌آوری اطلاعات را از جامعه آماری بر پایه یک روش نمونه‌گیری مشخص کرده و برای جمع‌آوری اطلاعات پرسش‌نامه‌ای طراحی می‌کند.

### ۳-۱-۱ تحلیل اطلاعات

پس از جمع‌آوری اطلاعات، محقق با استفاده از تکنیک‌های آماری اطلاعات به‌دست‌آمده را تحلیل می‌کند. محقق در این مرحله نیز همانند مرحله جمع‌آوری اطلاعات برای انجام تحلیل داده‌ها، نیازمند مشاوره با یک آماردان است، زیرا داده‌های آماری را می‌توان به‌روش‌های مختلفی تحلیل کرد. محقق باید مناسب‌ترین تحلیل را انتخاب کند. مناسب‌ترین تحلیل، تحلیلی است که منجر به نتایجی شود که محقق براساس آن نتایج بتواند در مورد صحت یا سقم فرضیات تحقیق تصمیم‌گیری کرده، همچنین بتواند یک نتیجه‌گیری نهایی برای موضوع تحقیق ارائه کند.

### ۴-۱-۱ مرحله نتیجه‌گیری

در این مرحله، محقق نتایج به‌دست‌آمده از مرحله تحلیل را به‌دقت مورد بررسی قرار داده و با کمک محققان دیگر به تفسیر نتایج به زبان ساده می‌پردازد. پیش‌نیاز هر علم دانستن زبان آن علم است. در ادامه برخی از مفاهیم اولیه و کلیدی علم آمار بررسی می‌شود.

### ۲-۱ مفاهیم اولیه علم آمار

#### ۱-۲-۱ انواع روش‌های آماری

روش‌های موجود در علم آمار را می‌توان به دو دسته آمار توصیفی و آمار استنباطی دسته‌بندی کرد. مشخص‌کردن مرز دقیقی بین این دو روش معمولاً کاری مشکل و غیرممکن است. ولی به‌طور غیردقیق می‌توان آمار توصیفی را قسمتی از علم آمار دانست که بیشتر بر روی توصیف داده‌ها و اطلاعات تمرکز دارد و حداقل استفاده را از روابط و فرمول‌های پیچیده ریاضی می‌کند. حال آنکه در آمار استنباطی سعی بر آن است که با استفاده از روش‌های ریاضی به ارائه نتیجه‌گیری از اطلاعات پرداخته شود.

بسیاری از محققان مبتدی فکر می‌کنند یک تحقیق خوب تحقیقی است که در آن، از آمار توصیفی حداقل استفاده شده‌باشد و بیشتر نتیجه‌گیری‌ها براساس روش‌های آمار استنباطی ارائه‌شده‌باشد. درحالی‌که این تصور کاملاً اشتباه است، زیرا در بسیاری از مواقع نتایج بسیار جالبی را می‌توان از آمار توصیفی گرفت که به‌سادگی نمی‌توان آن

نتایج را از آمار استنباطی به دست آورد. به نظر می‌رسد تحقیق‌های جامع و کامل، آن‌هایی هستند که از هر دو روش آماری استفاده مناسب و بهینه را داشته باشند.

### ۲-۲-۱ جامعه آماری

جامعه آماری به مجموعه کاملی از افراد، اشیاء یا اجزا اطلاق می‌شود که حداقل در یک صفت مشترک باشند.

### ۳-۲-۱ نمونه

در بسیاری از موارد، جمعیت مورد مطالعه به قدری بزرگ است که به علت هزینه زیاد و کمی وقت، مطالعه کلی جمعیت مقدور نیست، لذا قسمتی از جامعه را (تحت ضوابط خاصی) انتخاب کرده و بر روی آن مطالعات انجام می‌گیرد. به این بخش از جامعه، نمونه می‌گویند.

مثلاً جامعه آماری تحقیق محقق که علاقمند به بررسی شناسایی نیازهای آموزشی گندم‌کاران است، تمام گندم‌کاران است. مسلماً او نمی‌تواند تمامی جامعه آماری (N) را مورد بررسی قرار دهد. به همین دلیل او تعدادی از افراد جامعه آماری را، با استفاده از روش نمونه‌گیری مناسب، انتخاب می‌کند. به افراد انتخاب‌شده، نمونه (n) گفته می‌شود.

### ۳-۱ اندازه‌گیری و مقیاس‌سازی

مقیاس‌ها نقش بسیار تعیین‌کننده‌ای در انتخاب روش‌های آماری برای انجام تحلیل داده‌های آماری دارند. انواع مقیاس‌ها شامل موارد زیر است:

### ۱-۳-۱ مقیاس اسمی

داده‌هایی که متغیر آن‌ها کیفی است و در عین حال نمی‌توان هیچ‌گونه ترتیبی بین سطوح آن متغیر در نظر گرفت، دارای مقیاس اسمی هستند.

مثلاً متغیر جنسیت دارای مقیاس اسمی است زیرا اولاً: جنسیت صفتی کیفی است و ثانیاً: هیچ‌گونه ارزشی نمی‌توان بین طبقات این متغیر (زن و مرد) قائل شد.

معمولاً برای نمایش سطوح متغیر اسمی از اعداد به عنوان کد استفاده می‌کنیم. لازم به ذکر است که مقدار این کدها نشان‌گر هیچ‌گونه ارزش یا ترتیبی بین سطوح نیست.

### ۱-۳-۲ مقیاس ترتیبی

متغیری که مربوط به یک صفت کیفی است ولی برخلاف سطح اسمی، بین سطوح آن می‌توان نوعی ترتیب قائل شد، دارای مقیاس ترتیبی است.

مثلاً علاقه به تحصیل را می‌توان یک متغیر با مقیاس ترتیبی نامید. همانند مقیاس اسمی برای سادگی، از اعداد جهت نمایش سطوح یک متغیر با مقیاس ترتیبی استفاده می‌کنیم. اعداد اختصاص‌یافته به طبقات باز هم فاقد ارزش عددی بوده و صرفاً نشان‌گر ترتیب اهمیت طبقات هستند و نه میزان ارزش هر طبقه. مثلاً اگر طبقات به صورت ۱، ۲ و ۳ کدگذاری شوند، نمی‌توان نتیجه‌گیری کرد که طبقه اول سه‌برابر طبقه سوم ارزشمند است!

### ۱-۳-۳ مقیاس فاصله‌ای

متغیری که مربوط به یک صفت کمی است که مقادیر آن دارای ارزش عددی بوده و صفر آن غیرمطلق<sup>۱</sup> باشد، دارای مقیاس فاصله‌ای است. مثلاً درجه حرارت دارای مقیاس فاصله‌ای است.

### ۱-۳-۴ مقیاس نسبتی

متغیری که مربوط به یک صفت کمی است که مقادیر آن دارای ارزش عددی بوده و صفر آن مطلق باشد، دارای مقیاس نسبتی است.

مثلاً درجه نمره دانشجویان دارای مقیاس نسبتی است.

لازم به ذکر است:

- تمایز قائل‌شدن بین مقیاس‌های اسمی و ترتیبی بسیار مشکل است و معمولاً به دیدگاه محقق در مورد وجود یا عدم وجود ترتیب بین سطوح متغیر باز می‌گردد.
- اکثر روش‌های آماری که تاکنون ارائه‌شده، مربوط به متغیرهایی با مقیاس فاصله‌ای

---

۱. صفر مطلق صفری است که با تغییر واحد اندازه‌گیری مقدار آن تغییر پیدا نکند. مثلاً صفر در سیستم اندازه‌گیری طول یا وزن صفر مطلق است زیرا اگر مثلاً واحد وزن از گرم به کیلوگرم تغییر پیدا کند، مقدار صفر بدون تغییر باقی می‌ماند. حال آنکه در سیستم اندازه‌گیری دما صفر، مطلق نیست. زیرا اگر واحد دما از سلسیوس به فارنهایت تغییر پیدا کند، صفر درجه سلسیوس به ۳۲- درجه حرارت فارنهایت تبدیل می‌شود.

و نسبتی است. روش‌های بسیار محدودی برای تحلیل داده‌های ترتیبی و اسمی وجود دارد که بیشتر آن‌ها ناپارامتریک<sup>۱</sup> هستند.

- اگر بتوان داده‌ها را به‌صورت مقیاس فاصله‌ای یا مقیاس ترتیبی جمع‌آوری کرد، اولویت مقیاس فاصله‌ای است چرا که در مقیاس ترتیبی اصطلاحاً اطلاعات داده‌ها گم می‌شود.
- برای مقیاس اسمی تنها می‌توان از انواع فراوانی‌ها و برای مقیاس ترتیبی می‌توان از انواع فراوانی‌ها و میانه استفاده کرد. حال آنکه برای مقیاس فاصله‌ای و نسبتی تمامی فراوانی‌ها و شاخص‌های تمرکز و پراکندگی قابل استفاده است.

#### ۴-۱ انواع فراوانی

فراوانی درواقع میزان تکرار را در جامعه آماری نشان می‌دهد. با توجه به کاربرد فراوانی در آمار چهار نوع فراوانی زیر مورد استفاده قرار می‌گیرند:

- **فراوانی (فراوانی مطلق):** تعداد مشاهدات هر طبقه را فراوانی مطلق آن طبقه نامیده و آن را معمولاً با نماد  $f$  نمایش می‌دهند.
- **فراوانی نسبی:** نسبت فراوانی مطلق  $f$  به کل داده‌ها  $(n)$  را فراوانی نسبی گفته و آن را با نماد  $r$  نمایش می‌دهند.  $(r = f/n)$ .
- **فراوانی تجمعی:** به‌مجموع فراوانی‌های ماقبل هر طبقه، به‌همراه فراوانی خود آن طبقه، فراوانی تجمعی گفته می‌شود و آن را با نماد  $F$  نمایش می‌دهند.
- **فراوانی تجمعی نسبی:** به نسبت فراوانی تجمعی به کل داده‌ها فراوانی تجمعی نسبی گفته می‌شود و آن را با نماد  $R$  نمایش می‌دهند  $(R = F/n)$ .

**مثال ۱-۱.** برای دفع نوعی آفت سه نوع کود  $A$ ،  $B$  و  $C$  وجود دارد. برای تعیین متدوال،ترین کود یک نمونه تصادفی ۲۰ تایی از جامعه آماری انتخاب می‌شود. نوع کود مصرفی به‌صورت زیر است:

$A, A, B, A, C, C, A, B, C, B, C, A, A, B, C, A, B, B, A, C$

۱. آمار ناپارامتریک قسمتی از علم آمار است که آماردان بدون اطلاع از ماهیت و توزیع متغیرهای تصادفی به تحلیل و استنباط آماری می‌پردازد.

جدول ۱-۱ انواع فراوانی را برای داده‌های بالا ارائه می‌کند.

جدول ۱-۱. جدول مربوط به انواع فراوانی‌ها

| نوع کود | فراوانی | فراوانی نسبی   | فراوانی تجمعی | فراوانی تجمعی نسبی  |
|---------|---------|----------------|---------------|---------------------|
| A       | ۸       | $\frac{۸}{۲۰}$ | ۸             | $\frac{۸}{۲۰}$      |
| B       | ۶       | $\frac{۶}{۲۰}$ | ۱۴            | $\frac{۱۴}{۲۰}$     |
| C       | ۶       | $\frac{۶}{۲۰}$ | ۲۰            | $\frac{۲۰}{۲۰} = ۱$ |

### ۵-۱ مشاهده یا داده سانسور شده

به داده‌ای که مقدار آن به گونه‌ای از مجموعه مشاهدات حذف شده است، داده سانسور شده گویند.

مثلاً مدیری که علاقمند است نشان دهد که درآمد کارکنانش بالای ۳۰۰ هزار تومان در هر ماه است، ممکن است هر داده‌ای را که خلاف این مقدار را نشان دهد از نمونه تصادفی حذف کند، تا نظر او با استفاده از آن نمونه تصادفی (که البته پس از سانسور کردن داده‌ها دیگر نمونه تصادفی نخواهد بود) مورد تأیید قرار گیرد.

### ۶-۱ مشاهده یا داده گم شده

به داده‌ای که مقدار آن غیر قابل مشاهده باشد و محقق عملاً نتواند مقدار واقعی آن داده را ثبت یا اندازه‌گیری کند، داده گم شده گویند.

داده‌های گم شده را می‌توان به وفور در تحقیقاتی که مبتنی بر پرسش‌نامه است، مشاهده کرد. روش‌های متنوعی برای مواجهه با داده‌های گم شده وجود دارد که برخی از متداول‌ترین این روش‌ها در کتاب نمونه‌گیری عمیدی (۱۳۷۸) مورد بررسی قرار گرفته است.

### ۷-۱ داده دور افتاده

به داده‌ای است که مقدار آن به طور معناداری با بقیه داده‌ها متفاوت است. مثلاً نمرات ۰ و ۲۰ داده‌های دور افتاده در یک کلاس درسی محسوب می‌شوند. داده‌های دور افتاده

معمولاً براساس اشتباه اندازه‌گیری به‌وجود می‌آیند و یا ممکن است واقعیتی را در جامعه آماری بیان کنند که محقق تاکنون از آن غافل بوده‌است.

داده دورافتاده را می‌توان با استفاده از نمودار جعبه‌ای و یا براساس نظر محقق شناسایی کرد. روش‌های متفاوتی برای مواجهه با داده‌های دورافتاده وجود دارد که در فصل‌های آینده مورد بررسی قرار می‌گیرند.

## ۸-۱ متغیر تصادفی

### ۱-۸-۱ متغیر تصادفی و انواع آن

متغیر تصادفی متغیری است که تغییرات یک صفت را در جامعه آماری نشان می‌دهد. واضح است که واژه تصادفی بودن متغیر تصادفی به‌دلیل آن است که مقدار یک صفت از فردی به فرد دیگر در جامعه آماری تغییر می‌کند. متغیرهای تصادفی را معمولاً با حروف بزرگ  $X$ ،  $Y$  و  $Z$ ، ... نمایش می‌دهند.

متغیر تصادفی را می‌توان به‌صورت‌های زیر دسته‌بندی کرد:

- دسته‌بندی براساس مقادیری که متغیر تصادفی اختیار می‌کند:
  - متغیر تصادفی گسسته: متغیری که مقادیر آن قابل شمارش است. جنسیت و تعداد استان‌هایی که در سال ۸۸ خشکسالی داشته‌اند (۰، ۱، ۲، ...، ۳۱ استان) مثال‌هایی از متغیر گسسته هستند.
  - متغیر تصادفی پیوسته: متغیری است که مقادیر آن غیرقابل شمارش باشد و یا به‌عبارت ساده‌تر، بین هر دو مقدار متغیر بتوان مقدار دیگری در نظر گرفت. مثلاً: قد یک متغیر پیوسته است زیرا می‌توان بین دو مقدار ۱۵۲ سانتی‌متر و ۱۵۲/۵ سانتی‌متر مقدار ۱۵۲/۲۵ را در نظر گرفت.
- دسته‌بندی براساس نوع صفتی که متغیر تصادفی بر آن اساس تعریف می‌شود:
  - متغیر تصادفی کیفی: متغیری است که براساس یک صفت کیفی تعریف می‌شود. مثلاً جنسیت و رنگ ماشین مثال‌هایی از متغیرهای کیفی هستند.
  - متغیر تصادفی کمی: متغیری است که براساس یک صفت کمی تعریف می‌شود. مثلاً تعداد دانشجویان آمار در سال ۱۳۸۸ و متوسط درجه‌حرارت شهر تهران در سال ۱۳۸۸ مثال‌هایی از متغیرهای کمی هستند.

- با توجه به دسته‌بندی بالا و همچنین مقیاس‌های اندازه‌گیری می‌توان نتیجه گرفت:
- متغیرهای کیفی دارای مقیاس اسمی یا ترتیبی هستند که گاهی اوقات براساس نوع مقیاس یک متغیر کیفی، آن را متغیر کیفی اسمی یا ترتیبی می‌نامیم.
  - گسسته یا پیوسته بودن یک متغیر تصادفی تنها برای متغیرهای کمی قابل تعریف است.
  - مقیاس فاصله‌ای یا نسبتی تنها برای متغیرهای کمی قابل تعریف است.

### ۹-۱ خلاصه‌سازی و مقایسه داده‌ها

در علم آمار، همواره علاقمند به ارائه نتیجه‌گیری براساس داده‌ها و اطلاعات موجود در مورد یک یا چند جامعه آماری هستیم. واضح است که این‌گونه نتیجه‌گیری و مقایسه بین جوامع آماری براساس کلیه اطلاعات موجود، امری ناممکن است. برای انجام این‌گونه نتیجه‌گیری و یا مقایسه‌ها، نیازمند معیارهایی هستیم که براساس آن‌ها بتوان داده‌ها را خلاصه کرد.

خلاصه‌سازی داده‌ها را می‌توان براساس دو عامل تمرکز و پراکندگی انجام داد. به شاخص‌هایی که براساس عامل تمرکز عمل می‌کنند «شاخص تمرکز» می‌گویند. به زبان ساده: شاخص تمرکز عددی است که نشان می‌دهد داده‌ها در کجای جامعه آماری بیشتر متمرکز شده‌اند. شاخص دیگری که براساس عامل پراکندگی عمل می‌کند، و از آن با عنوان شاخص پراکندگی یاد خواهیم کرد، عددی است که نشان می‌دهد به چه میزان داده‌های موجود حول شاخص تمرکز مناسب، پراکنده شده‌اند.

با توجه به تعاریف ارائه‌شده، واضح است که هر قدر میزان یک شاخص تمرکز در یک جامعه آماری بیشتر باشد، آن جامعه آماری از نظر ما مناسب‌تر است. حال آنکه براساس شاخص پراکندگی جامعه‌ای مناسب‌تر است که شاخص پراکندگی آن‌ها جامعه کوچک باشد. در بسیاری مواقع با جوامع آماری روبه‌رو هستیم که شاخص تمرکز آن‌ها بزرگ و پراکندگی آن‌ها نیز بزرگ است. در این‌گونه مواقع نیازمند شاخص دیگری با عنوان ضریب تغییرات هستیم که همزمان هر دو شاخص پراکندگی و تمرکز را مورد توجه قرار می‌دهد. در این بخش به معرفی تمامی این شاخص‌ها پرداخته و همچنین در مورد مناسب بودن این شاخص‌ها با توجه به نوع و مقیاس داده‌ها بحث خواهد شد.