فهرست

کلام آغازین ۱۷

بخش اول: کلیات علم آمار

فصل ۱- مفاهیم و اصطلاحات آماری ۲۱

مقدمه ۲۱
تعریف علم آمار ۲۱
آمار توصیفی و آمار استنباطی ۲۱
خلاصه‌ای از آمار توصیفی ۲۲
چندک‌ها: چارک‌ها، دهک‌ها و صدک‌ها ۲۴
محاسبه شاخص‌های آمار توصیفی در SPSS ۲۹
جامعه آماری و نمونه آماری ۳۰
آماره و پارامتر ۳۱
ثابت و متغیر ۳۱
انواع متغیر از نظر ماهیت مقادیر آنها ۳۱
انواع متغیرها از نظر نقش آنها در تحقیق ۳۲
فرضیه و انواع آن ۳۲
سطح معناداری یا سطح خطا ۳۴
انواع خطا در استنباط آماری ۳۵
آزمون‌های پارامتریک و آزمون‌های غیرپارامتریک ۳۶
ملاک‌ها (یا مفروضه‌های) آزمون‌های پارامتریک ۳۷
روش کار در نرم‌افزار **SPSS** **۴۱**
اینترنت و آمار ۴۳
خودآزمایی و تمرین‌های فصل اول ۴۴

فصل ۲- منحنی توزیع نرمال ۴۸

مقدمه ۴۸
ویژگی‌های منحنی طبیعی ۴۹

منحنی نرمال استاندارد ۴۹
تقسیم بندی سطوح زیر منحنی طبیعی ۵۰
استفاده از سطوح زیر منحنی طبیعی ۵۱
خودآزمایی و تمرین‌های فصل دوم ۵۳

بخش دوم: آزمون‌های پارامتریک
(قسمت الف: آزمون‌های مربوط به بررسی تفاوت‌ها)

آزمون‌های خانواده **t** استیودنت ۵۶

فصل۳- آزمون t تک‌نمونه‌ای ۵۷
توصیف ۵۷
مفروضه‌ها ۵۸
مراحل محاسبه ۵۸
محاسبه اندازه اثر ۵۹
نحوه انجام آزمون **t** تک‌نمونه‌ای در SPSS **۶۱**
نحوه گزارش نتایج ۶۱
خودآزمایی و تمرین‌های فصل سوم ۶۲

فصل ۴- آزمون t مستقل ۶۴
توصیف ۶۴
مفروضه‌ها ۶۴
مراحل محاسبه آزمون ۶۵
محاسبه اندازه اثر ۶۷
نحوه انجام آزمون t مستقل با نرم‌افزار SPSS **۶۸**
نحوه گزارش نتایج آزمون ۷۰
خودآزمایی و تمرین‌های فصل چهارم ۷۰

فصل ۵- آزمون t وابسته ۷۳
توصیف ۷۳
مفروضه‌ها ۷۳
مثال برای موقعیت الف ۷۳
مراحل محاسبه آزمون ۷۴

محاسبه اندازه اثر ۷۵
نحوه انجام آزمون t وابسته با نرم‌افزار SPSS ۷۶
نحوه گزارش نتایج آزمون ۷۷
مثال برای موقعیت ب ۷۷
مراحل محاسبه ۷۷
محاسبه اندازه اثر ۷۹
نحوه گزارش نتایج آزمون ۷۹
خودآزمایی و تمرین‌های فصل پنجم ۷۹

فصل ۶- تحلیل واریانس یک‌سویه (بین‌گروهی) ۸۳
توصیف ۸۳
مفروضه‌ها ۸۳
مراحل محاسبه آزمون ۸۴
آزمون‌های تعقیبی ۸۸
آزمون تعقیبی .L.S.D ۸۸
آزمون تعقیبی توکی ۹۰
محاسبه اندازه اثر از طریق مجذور اِتا (اِتادو) ۹۱
تفسیر مقادیر مجذور اتا ۹۲
انجام تحلیل واریانس یک‌سویه با SPSS ۹۲
روش دیگر برای انجام آزمون آنوا در SPSS ۹۴
نحوه گزارش نتایج آزمون ۹۴
خودآزمایی و تمرین‌های فصل ششم ۹۵

فصل ۷- تحلیل واریانس بین گروهی دوسویه یا عاملی ۹۹
توصیف ۹۹
مفروضه‌ها ۹۹
مراحل محاسبه آزمون ۱۰۱
آزمون‌های تعقیبی ۱۰۵
آزمون تعقیبی شِفه ۱۰۵
نحوه انجام تحلیل واریانس بین گروهی دوسویه در نرم‌افزار SPSS ۱۰۷
نحوه گزارش نتایج ۱۱۰
خودآزمایی و تمرین‌های فصل هفتم ۱۱۱

فصل ۸- تحلیل واریانس نمره‌های همبسته (اندازه‌گیری‌های مکرر) ۱۱۴
توصیف ۱۱۴
مفروضه‌ها ۱۱۴
مراحل محاسبه آزمون ۱۱۵
نحوه انجام تحلیل واریانس یکسویه با اندازه‌گیری‌های مکرر در نرم‌افزار SPSS ۱۱۸
نحوه گزارش نتایج ۱۲۱
خودآزمایی و تمرین‌های فصل هشتم ۱۲۲

فصل ۹- تحلیل واریانس دوراهه آمیخته یا مرکب ۱۲۴
توصیف ۱۲۴
مفروضه‌ها ۱۲۵
مراحل محاسبه آزمون ۱۲۶
نحوه انجام تحلیل واریانس آمیخته در SPSS ۱۳۰
نحوه گزارش نتایج آزمون ۱۳۵
خودآزمایی و تمرین‌های فصل نهم ۱۳۵

فصل ۱۰- تحلیل کوواریانس (آنکوا) ۱۳۷
توصیف ۱۳۷
مفروضه‌ها ۱۳۸
بررسی وجود مفروضه‌های استفاده از تحلیل کوواریانس در داده‌ها ۱۳۹
مراحل محاسبه آزمون ۱۴۵
روش انجام تحلیل کوواریانس در SPSS ۱۵۱
نحوه گزارش کوواریانس یک‌طرفه ۱۵۳
خودآزمایی و تمرین‌های فصل دهم ۱۵۴

فصل ۱۱- تحلیل واریانس چندمتغیری (مانوا) ۱۵۶
توصیف ۱۵۶
دلایل استفاده از مانوا ۱۵۶
چه موقع نباید از مانوا استفاده کرد ۱۵۷
مفروضه‌ها و محدودیت‌های آماری در تحلیل واریانس چندمتغیری (مانوا) ۱۵۷
انواع تحلیل واریانس چندمتغیره (مانوا) ۱۵۸
مثال شماره ۱: مانوا با متغیر مستقل دارای دو سطح ۱۵۹

الف: بررسی نرمال بودن تک‌متغیری از طریق آزمون کولموگروف- اسمیرنوف ۱۶۰
ب: بررسی نرمال بودن چندمتغیری از طریق شاخص ماهلانوبیس ۱۶۱
نحوه گزارش نتایج مانوا ۱۶۸
مثال شماره ۲ : مانوا با متغیر مستقل دارای بیش از دو سطح ۱۶۹
روش انجام مانوا برای مثال شماره ۲ در SPSS **۱۶۹**
خودآزمایی و تمرین‌های فصل یازدهم ۱۷۵

فصل ۱۲- تحلیل آشیانه‌ای (سلسله‌مراتبی) ۱۷۷
توصیف ۱۷۷
مفروضه‌ها ۱۷۸
مراحل محاسبه آزمون ۱۷۸
نحوه انجام تحلیل آشیانه‌ای در SPSS **۱۸۲**
نحوه گزارش نتایج ۱۸۲
خودآزمایی و تمرین‌های فصل دوازدهم ۱۸۲

فصل ۱۳- تحلیل واریانس بلوک‌های تصادفی ۱۸۴
توصیف ۱۸۴
مفروضه‌ها ۱۸۴
مراحل محاسبه آزمون ۱۸۶
نحوه انجام طرح بلوکی در SPSS **۱۸۹**
نحوه گزارش نتایج طرح بلوکی ۱۹۲
خودآزمایی و تمرین های فصل سیزدهم ۱۹۳

فصل ۱۴- طرح مربع لاتین ۱۹۵
توصیف ۱۹۵
مفروضه‌ها ۱۹۵
مراحل محاسبه آزمون ۱۹۶
نحوه انجام مربع لاتین در SPSS **۲۰۰**
نحوه گزارش نتایج ۲۰۲
خودآزمایی و تمرین‌های فصل چهاردهم ۲۱۰

بخش دوم: آزمون‌های پارامتریک
(قسمت ب: آزمون‌های مربوط به بررسی رابطه‌ها)

فصل ۱۵- روابط همبستگی بین متغیرها ۲۱۳

مقدمه ۲۱۳

دسته ۱: ضرایب مبتنی بر مقیاس اسمی ۲۱۷

دسته۲: ضرایب همبستگی مبتنی بر مقیاس‌های رتبه‌ای ۲۲۷

دسته ۳: ضرایب مبتنی بر مقیاس فاصله‌ای/ نسبی ۲۳۴

مفروضه‌ها ۲۳۸

مراحل محاسبه آزمون ۲۳۸

آزمون معنادار بودن ضریب همبستگی ۲۳۹

بررسی معنادار بودن تفاوت دو ضریب همبستگی ۲۴۰

نحوه گزارش نتایج ۲۴۲

خودآزمایی و تمرین‌های فصل پانزدهم ۲۴۵

فصل ۱۶- رگرسیون خطی ۲۵۰

الف: رگرسیون خطی ساده ۲۵۰

مفروضه‌های رگرسیون ساده ۲۵۱

مراحل محاسبه رگرسیون ۲۵۱

خط رگرسیون ۲۵۵

محاسبه ضریب رگرسیون تفکیکی استانداردشده ۲۵۶

آزمون معناداری مدل رگرسیون ۲۵۷

روش محاسبه رگرسیون ساده دو متغیری در SPSS ۲۵۹

نحوه گزارش نتایج رگرسیون ساده ۲۶۱

خودآزمایی و تمرین‌های فصل شانزدهم: بخش الف (رگرسیون خطی ساده) ۲۶۲

ب: تحلیل رگرسیون خطی چندگانه (چندمتغیره) ۲۶۴

توصیف ۲۶۵

مفروضه‌ها ۲۶۵

نحوه گزارش نتایج ۲۸۱

خودآزمایی و تمرین‌های فصل شانزدهم: بخش ب (رگرسیون خطی چندگانه) ۲۸۲

فصل ۱۷- تحلیل عاملی ۲۸۴

توصیف ۲۸۴

انواع تحلیل عاملی ۲۸۵
مدل‌های عاملی ۲۸۵
مفاهیم کلیدی در تحلیل عاملی ۲۸۶
مفروضه‌ها ۲۸۷
مراحل اجرای تحلیل عاملی در SPSS ۲۹۱
نحوه گزارش نتایج تحلیل عاملی ۳۰۰
خودآزمایی و تمرین‌های فصل هفدهم ۳۰۰

فصل ۱۸- تحلیل مسیر ۳۰۳
توصیف ۳۰۳
مفروضه‌های تحلیل مسیر ۳۰۳
متغیرهای درون‌زا و برون‌زا ۳۰۴
نحوه انجام تحلیل مسیر در SPSS ۳۰۷
نحوه انجام تحلیل مسیر در نرم‌افزار AMOS ۳۱۰
خودآزمایی و تمرین‌های فصل هجدهم ۳۱۴

فصل ۱۹- مدل‌یابی معادلات ساختاری ۳۱۶
توصیف ۳۱۶
اصطلاحات پرکاربرد حوزه مدل‌یابی معادلات ساختاری ۳۱۷
متغیرهای وابسته و مستقل ۳۱۸
انواع مدل در روش مدل‌یابی معادلات ساختاری ۳۲۰
تدوین مدل ۳۲۱
حجم نمونه در مدل‌یابی معادلات ساختاری ۳۲۱
مراحل مدل‌یابی معادلات ساختاری ۳۲۲
آزمون‌های برازندگی مدل کلی ۳۲۳
شاخص‌های مطلق مدل ۳۲۳
شاخص‌های تطبیقی مدل ۳۲۴
شاخص‌های اقتصادی مدل ۳۲۴
سایر شاخص‌ها ۳۲۴
الف: برخی از شاخص‌های برازش مطلق ۳۲۶

ب: برخی از شاخص‌های تطبیقی ۳۲۷
ج: برخی از شاخص‌های اقتصادی ۳۲۸
سایر شاخص‌ها ۳۲۹
نحوه انجام مدل‌یابی معادلات ساختاری در AMOS..... **۳۳۱**
انجام تحلیل ۳۳۶
مشاهده خروجی ۳۳۷
الف: خروجی به‌صورت نمودار ۳۳۷
ب: خروجی به‌صورت جداول..... ۳۳۷
نحوه گزارش نتایج مدل‌یابی معادلات ساختاری ۳۴۱
خودآزمایی و تمرین‌های فصل نوزدهم ۳۴۲

بخش سوم

فصل ۲۰- آزمون‌های غیرپارامتریک..... ۳۴۵
مقدمه ۳۴۵
آزمون‌های غیرپارامتری فرضیه‌های تفاوتی تک‌متغیری (تک‌نمونه‌ای) ۳۴۶
آزمون مجذور کای (χ^2) تک‌متغیری ۳۴۶
آزمون کولموگروف- اسمیرنوف یک‌متغیری ۳۵۱
آزمون‌های غیرپارامتری فرضیه‌های تفاوتی دومتغیری..... ۳۵۶
آزمون کولموگروف-اسمیرنوف دومتغیری..... ۳۵۶
آزمون U من- ویتنی ۳۶۳
آزمون مجذور کای دومتغیری (آزمون استقلال) ۳۶۹
مجذور کای با جدول توافقی دو در دو **(۲×۲)** ۳۷۴
نحوه انجام مجذور کای دو در دو در SPSS..... **۳۷۶**
آزمون ویل‌کاکسون..... ۳۷۹
آزمون دوجمله‌ای ۳۸۳
آزمون علامت یا نشانه (آزمون غیرپارامتری مقایسه دو گروه وابسته)..... ۳۸۸
آزمون مک‌نمار ۳۹۳
آزمون تفاوت دو نسبت همبسته..... ۳۹۸
آزمون‌های غیرپارامتری فرضیه‌های تفاوتی چندمتغیری (چندنمونه‌ای) ۴۰۰
آزمون کروسکال - والیس..... ۴۰۰

آزمون کوکران یا آزمون کیو ۴۰۵
آزمون فریدمن ۴۰۸
روش لگاریتم خطی ۴۱۰
خودآزمایی و تمرین‌های فصل بیستم ۴۲۴

پیوست‌ها ۴۲۸
پاسخ‌نامه سؤالات چندگزینه‌ای و تشریحی فصل‌ها ۴۲۹
علایم و نمادهای آماری ۴۵۷
مروری بر عملیات محاسباتی پایه ۴۵۸
مباحث آماری در کتب آماری ۴۷۰
راهنمای انتخاب آزمون آماری ۴۷۴
فهرست جداول آماری ۴۸۵

منابع ۵۱۹

واژه‌نامه: فارسی به انگلیسی ۵۲۳

واژه‌نامه: انگلیسی به فارسی ۵۲۵

کلام آغازین

خوشبختانه در مقایسه با سایر زمینه‌ها، منابع بسیار خوبی در خصوص روش‌های آماری در علوم رفتاری به‌ویژه در خصوص نرم‌افزارهای آماری همچون SPSS تألیف و ترجمه شده است. از سوی دیگر، با عنایت به گسترش مقاطع تحصیلی در علوم رفتاری در اغلب دانشگاه های کشور و استقبال خوب از این رشته‌ها، همچنان نیاز به وجود منابع گوناگون و تخصصی در زمینه روش‌های آماری احساس می‌شود. با این‌وجود، هر کتاب جدیدی که در این زمینه چاپ می‌شود باید از ویژگی‌های خاصی نسبت به کتاب‌های قبلی برخوردار باشد تا بتواند مفید واقع شود.

شاکله این کتاب که خدمت خوانندگان محترم تقدیم می‌شود، مطالبی است که در دوره دکترا از استاد بزرگوارم جناب آقای دکتر حسن‌پاشا شریفی در کلاس درس روش‌های آماری آموختم. بنده با توجه به علاقه‌ای که به مباحث آماری و روش تدریس استاد داشتم، به این فکر افتادم که با مبنا قرار دادن روش و مطالب استادم، کتابی را تألیف کنم تا مورد استفاده همه علاقه‌مندان قرار گیرد. بنابراین ویژگی خاص کتاب حاضر این است که مطالب آن به شیوه تدریس استادی تشکیل شده است که سبک تدریس ایشان به نظر بنده و همه شاگردانش، آموزش پیچیده‌ترین مباحث آماری به بهترین شکل ممکن است.

از ویژگی‌های دیگر این کتاب، آموزش اغلب روش‌های آماری به دو شکل محاسبه معمولی و محاسبه با کمک نرم‌افزار SPSS و AMOS است. استفاده از مثال‌های بسیار ساده، انجام آزمون‌های آماری به شکل گام‌به‌گام، وجود تمرین‌های محاسباتی و چندگزینه‌ای دوره کنکور کارشناسی ارشد و دکتری برای هر روش آماری نیز می‌تواند از امتیازات این کتاب محسوب شود. در نهایت، بخش جدیدی با عنوان مباحث آماری در کتب آماری شامل فهرست مطالب اغلب کتاب‌های معتبر آماری است که به افرادی که در جستجوی یک بحث خاص آماری هستند کمک می‌کند تا آن مبحث را در کتب مختلف آماری پیدا کنند.

افزون بر استاد بزرگوارم، در تهیه این کتاب از محضر استادان و سروران زیادی بهره گرفتم: جناب آقای دکتر صیاد گلشن، جناب آقای دکتر یداله ولی‌نژاد، جناب آقای دکتر یداله عباس‌زاده، جناب آقای دکتر جواد امانی در ویراستاری علمی کتاب مساعدت فرمودند و همچنین برادر بزرگوارم مهندس مهدی زارعی در تایپ و تنظیم مطالب کمک کردند که برحسب وظیفه سپاسگزار این سروران هستم.

در چاپ این کتاب همه مسئولان محترم انتشارات ارجمند به‌ویژه جناب آقای دکتر ارجمند و سرکار خانم مرضیه شکی نهایت لطف و همکاری را داشتند بدین وسیله از همه آنها صمیمانه تشکر و قدردانی می‌کنم.

در تألیف کتاب حاضر، بیش از خودم، همسر وفادارم و فرزندان عزیزم مرا یاری کردند. مخصوصاً از این بابت‌که، وقت‌هایی را که باید در اختیار آنها می‌بودم، صرف این کتاب کردم. لذا بابت همه‌چیز از آنها متشکرم.

حیدر علی زارعی
(alizarei@iaukhoy.ac.ir)

بخش اول

کلیات علم آمار

فصل ۱: مفاهیم و اصطلاحات آماری

فصل ۲: منحنی توزیع نرمال

فصل ۱

مفاهیم و اصطلاحات آماری

مقدمه

هنگامی‌که حقایق به‌صورت اعداد و ارقام ارائه شوند از اصطلاح آمار استفاده می‌شود. واژه آمار در لغت به معنی شمردن و حساب کردن است. ساده‌ترین عمل در آمار نیز شمارش و تعیین اجزاء یا آحاد تشکیل‌دهنده یک مجموعه است. واژه انگلیسی آمار statistics است که از ریشه لاتینی status به معنی ارقام و اطلاعات عددی مربوط به وضع اجتماعی و اقتصادی کشور که برای اداره حکومت لازم و مفیدند، گرفته شده است. باید توجه داشت که واژه دیگری از همین ریشه با عنوان statistic وجود دارد که برای نامیدن اندازه‌های حاصل از محاسبات آماری مانند میانگین و میانه به کار می‌رود که معادل فارسی آن شاخص آماری یا آماره است (شریفی و نجفی زند، ۱۳۷۶).

تعریف علم آمار

علم آمار فعالیتی گسترده‌تر از شمارش و ارائه حقایق به‌صورت اعداد و ارقام است. علم آمار به مجموعه‌ای از فنون و روش‌های علمی- ریاضی گفته می‌شود که برای جمع‌آوری، تنظیم، ارائه، تحلیل و تفسیر اطلاعات کمی و کیفی و نتیجه‌گیری از آنها در جهت هدفی معین به کار می‌رود (شریفی و نجفی زند، ۱۳۷۶).

آمار توصیفی و آمار استنباطی

طبق یک طبقه‌بندی قدیمی، فعالیت آماری دو بخش اصلی دارد:

الف: جمع‌آوری، تلخیص، تنظیم و ارائه اطلاعات به‌صورت روشن و قابل‌درک و در صورت لزوم تعیین روابط موجود بین اطلاعات جمع‌آوری‌شده. این بخش از آمار را که بیشتر به مشخص کردن داده‌ها، تنظیم و ارائه آنها به‌صورت جدول‌بندی یا ترسیمی، محاسبه آماره‌ها و تعیین ارتباط بین اطلاعات می‌پردازد، آمار

توصیفی[1] می‌نامند. به‌طورکلی در این بخش از آمار، آمارشناس به توصیف خصایص نمونه(ها) می‌پردازد.
ب: در بیشتر فعالیت‌های آماری جمع‌آوری، تنظیم و ارائه یافته‌ها و یا تعیین آماره‌ها و روابط بین داده‌ها کفایت نمی‌کند، بلکه لازم است بر اساس این اطلاعات جمع‌آوری و تنظیم‌شده، تحلیل‌ها و استنباط‌هایی برای تبیین و تصمیم‌گیری صورت گیرد. این بخش از آمار که به تحلیل، تفسیر و تعمیم نتایج حاصل از تنظیم و محاسبه مقدماتی آماری تکیه دارد، آمار استنباطی[2] خوانده می‌شود. با استفاده از روش‌های آمار استنباطی می‌توان مشخصات جامعه آماری را از روی ویژگی‌های نمونه(ها) استنباط کرد.

خلاصه‌ای از آمار توصیفی

با توجه به این‌که آمار توصیفی پیش‌نیاز آمار استنباطی محسوب می‌شود لذا در سطور پایین مختصری از آمار توصیفی جهت یادآوری ارائه می‌شود.

به‌طورکلی آمار توصیفی شامل جدول توزیع داده‌ها (طبقات، فراوانی، درصد، درصد تراکمی یا تجمعی)، شاخص‌های گرایش مرکزی (میانگین، میانه، مد)، شاخص‌های پراکندگی (واریانس، انحراف معیار، دامنه تغییرات)، چارک‌ها، دهک‌ها و صدک‌ها، شاخص‌های کجی و کشیدگی، حداقل و حداکثر توزیع داده‌ها، تعداد و مجموع نمرات به همراه نمایش تصویری داده‌ها یعنی نمودارهای توزیع داده‌ها است(شسکین[3]، ۲۰۰۳). در ادامه هرکدام از مفاهیم و شاخص‌های آماری مذکور به‌طور خلاصه شرح داده می‌شود:

جدول توزیع داده‌ها: چون اغلب در علم آمار، با داده‌های زیادی مواجه هستیم لذا برای خلاصه کردن و قابل‌درک کردن آنها نیاز است که داده‌ها را در قالب یک جدول تنظیم کنیم. در حقیقت، جدول توزیع داده‌ها، انبوهی از داده‌های منفرد را در یک جدول سازمان‌دهی می‌کند. جدول توزیع داده‌ها معمولا به شکل زیر بوده و از سمت چپ به راست تنظیم می‌شود:

جدول ۱-۱: نمونه‌ای فرضی از جدول توزیع فراوانی در آمار توصیفی

طبقات	فراوانی ساده	درصد	فراوانی تراکمی	درصد تراکمی
۱ – ۴	۳	۱۰	۳	٪۱۰
۵ – ۸	۵	۱۷	۸	٪۲۷
۹ – ۱۲	۷	۲۳	۱۵	٪۵۰
۱۳ – ۱۶	۱۰	۳۳	۲۵	٪۸۳
۱۷ – ۲۰	۵	۱۷	۳۰	٪۱۰۰

طبقات می‌تواند به فاصله یک یا بیشتر از یک طبقه‌بندی شود. تعداد مجاز طبقات معمولاً بین هفت

1- descriptive statistic
2- inferential statistics
3 -Sheskin

تا بیست‌ویک طبقه را شامل می‌شود و اغلب به‌صورت یک عدد فرد در نظر گرفته می‌شود مثلاً هفت یا نه طبقه. نرم‌افزارهای آماری مثل SPSS فاصله طبقات را برابر یک در نظر می‌گیرند. فراوانی تراکمی و درصد تراکمی از مجموع فراوانی و درصد همان طبقه به‌اضافه‌ی فراوانی طبقات و درصدهای طبقات قبلی به دست می‌آید. فراوانی و درصد تراکمی آخرین طبقه باید به ترتیب برابر با تعداد کل نمرات توزیع و ۱۰۰٪ باشد.

میانگین: انواع مختلفی از میانگین مثل هندسی[1]، وزنی[2]، هارمونیک[3] و حسابی[4] در علم آمار وجود دارد. پرکاربردترین نوع آن میانگین حسابی است؛ به‌طوری‌که هر زمان از اصطلاح میانگین استفاده می‌شود منظور همان میانگین حسابی است. میانگین کمیتی است که مقدار متوسط و یا به‌بیان‌دیگر مرکز ثقل داده‌های یک توزیع معین را نشان می‌دهد. محاسبه میانگین حسابی بسیار آسان است و از تقسیم مجموع نمرات بر تعداد کل به دست می‌آید.

میانه: میانه عددی است که توزیع داده‌ها را به دو نصف برابر تقسیم می‌کند. برای محاسبه میانه ابتدا باید نمرات را از کوچک به بزرگ مرتب کرد و سپس در توزیع‌هایی که تعداد نمرات آن یک عدد فرد است، همان عددی که در وسط توزیع قرار می‌گیرد به‌عنوان میانه محسوب می‌شود؛ و در توزیع‌هایی که تعداد آن یک عدد زوج است، میانه از مجموع دو عدد وسطی تقسیم بر دو به دست می‌آید:

توزیع فرد	۱،۲،۳	Mn = ۲
توزیع زوج	۱،۲،۳،۴	$Mn = \frac{۲+۳}{۲} = ۲/۵$

اگر داده‌ها با فاصله طبقاتی بزرگ‌تر از یک طبقه‌بندی‌شده باشند آن موقع باید برای محاسبه میانه، فراوانی تراکمی طبقات را هم محاسبه کرد و از یک فرمول دیگر که در کتاب‌های آمار توصیفی و همچنین در جدول ۱-۲ این کتاب وجود دارد، استفاده کرد.

نکته: زمانی که توزیع داده‌ها نرمال نیست و دارای کجی است میانه به عنوان یک شاخص گرایش مرکزی مورد توجه قرار می‌گیرد و همان نقشی را بازی می‌کند که میانگین در توزیع‌های نرمال بر عهده دارد.

مد یا نما: در یک توزیع داده‌ها، مد عدد یا نماینده طبقه‌ای است که بیشترین تکرار یا فراوانی را داشته باشد. ساده‌ترین و تنها شاخص گرایش مرکزی است که در مورد داده‌هایی با مقیاس اسمی به کار می‌رود. اگر در یک توزیع، همه داده‌ها با فراوانی یک باشند، درآن صورت مدی وجود نخواهد داشت؛ یعنی توزیع داده‌ها فاقد مد خواهد بود و نرم‌افزار SPSS در چنین صورتی، کوچک‌ترین عدد توزیع را به‌عنوان مد در نظر می‌گیرد. اگر در یک توزیع دو عدد دارای فراوانی برابر باشند در آن صورت توزیع داده‌ها دارای دو مد خواهد بود که به آن توزیع دو مدی گفته می‌شود (آذر و مؤمنی، ۱۳۹۳).

واریانس: متوسط یا میانگین مجذور انحراف‌ها از میانگین را واریانس (پراش) می‌نامند (شریفی و نجفی زند، ۱۳۷۶). واریانس میزان پراکندگی نمرات را در پیرامون میانگین نشان می‌دهد. هر چه قدر مقدار

1- geometric mean
2- weighted mean
3- harmonic mean
4- arithmetic mean

واریانس در یک توزیع بیشتر باشد، نشان‌گر آن است که مقدار پراکندگی در آن داده‌ها زیاد است. اصولاً در آزمون‌های آماری هر چه قدر مقدار واریانس بیشتر باشد کاربرد آزمون بهتر جواب خواهد داد. برای مثال واریانس نمرات هوش، پیشرفت تحصیلی در دوره دبیرستان بیشتر از واریانس همان متغیرها در دوره کارشناسی است. چون افرادی که وارد دوره کارشناسی می‌شوند همگن‌تر از افراد دوره دبیرستان هستند و بنابراین پراکندگی نمرات افراد در دانشگاه کمتر از دبیرستان است.

فرمول‌های متعددی برای محاسبه واریانس وجود دارد یکی از این فرمول‌ها که با داده‌های خام کار می‌کند به شکل زیر است:

$$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$$

در فرمول فوق، n به تعداد نمره‌ها یا فراوانی‌ها، $\sum x^2$ مجموع مجذور تک‌تک نمره‌ها و $(\sum x)^2$ به مجذور مجموع نمره‌ها اشاره دارد.

انحراف معیار: ریشه دوم واریانس را انحراف استاندارد یا انحراف معیار می‌نامند. به‌عبارت‌دیگر اگر از واریانس جذر بگیریم، انحراف معیار به دست می‌آید و برعکس اگر انحراف معیار را به توان دو برسانیم، واریانس حاصل می‌شود؛ بنابراین فرمول واریانس و انحراف معیار دقیقاً یکسان است و فقط در علامت رادیکال با هم فرق می‌کنند. پس فرمول انحراف معیار به‌صورت زیر خواهد بود:

$$s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

اگر مقدار ثابتی را به تک‌تک اعداد توزیع، اضافه یا منها کنیم، انحراف معیار و واریانس هیچ تغییری نخواهد کرد؛ اما اگر همه اعداد یک توزیع را در عدد ثابتی مثل c ضرب کنیم، انحراف استاندارد آن توزیع نیز در آن عدد ضرب خواهد شد؛ و اگر همه اعداد را در عدد ثابتی تقسیم کنیم، انحراف معیار جدید نیز، حاصل تقسیم انحراف معیار قبلی بر همان عدد جدید خواهد بود.

توجه: برای محاسبه میانگین، واریانس و انحراف معیار مقیاس داده‌ها باید به‌صورت فاصله‌ای یا نسبی باشد و محاسبه این شاخص‌ها در مقیاس‌های اسمی و ترتیبی بی‌معنا بوده و درست نیست.

دامنه تغییرات: درواقع این شاخص پراکندگی تفاوت میان بیشترین و کمترین مقدار حاصل در یک سری از داده‌ها است. فرمول محاسبه آن به شرح زیر است:

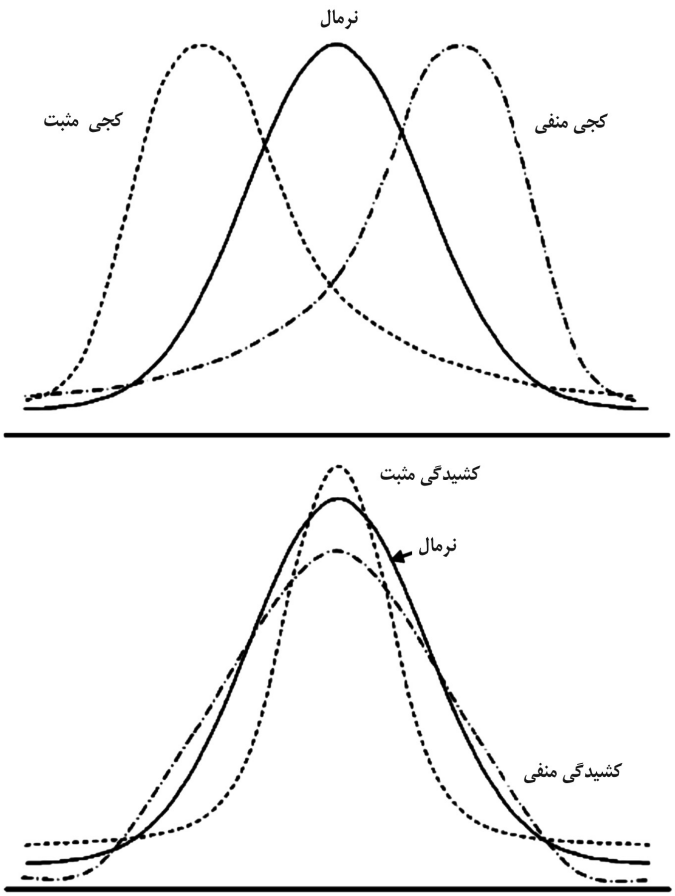
دامنه تغییرات (R) = بزرگ‌ترین داده (Maximum) − کوچک‌ترین داده (Minimum)

چندک‌ها: چارک‌ها، دهک‌ها و صدک‌ها

چارک‌ها، دهک‌ها (نقاط دهکی) و صدک‌ها (نقاط صدکی) شاخص‌هایی هستند که به ترتیب توزیع داده‌ها را به چهار، ده و صد قسمت مساوی تقسیم می‌کنند. برای نمونه چارک‌ها یک توزیع داده را به سه قسمت چارک اول، دوم و سوم تقسیم می‌کنند و بدین ترتیب توزیع داده‌ها به چهار قسمت تقسیم می‌شود. چارک اول ۲۵ درصد پایین‌ترین نمره‌ها، چارک دوم نقطه پنجاه درصدی توزیع نمره را نشان می‌دهد. به همین جهت چارک دوم همیشه

مساوی با میانه است. چارک سوم نقطه‌ای را در توزیع نمرات نشان می‌دهد که ۷۵ درصد آنها پایین‌تر از آن و ۲۵ درصد بالای آن نقطه قرار می‌گیرند. این مسئله در مورد دهک‌ها و صدک‌ها هم صدق می‌کند. به‌عنوان مثال دیگر می‌توان افراد جامعه را بر اساس درآمد به ده قسمت مثل دهک اول، دوم و ... تقسیم‌بندی کرده و سپس در خصوص اختصاص یارانه و سایر امکانات رفاهی در مورد آنها تصمیم‌گیری کرد.

کجی و کشیدگی: شاخص کجی[1] و کشیدگی[2] میزان انحراف توزیع نمرات را از توزیع منحنی نرمال نشان می‌دهد. توزیع منحنی نرمال به تفصیل در فصل دوم توضیح داده خواهد شد. مشخص کردن این دو شاخص برای متغیرها بسیار لازم و ضروری است. چون یکی از مفروضه‌های آزمون‌های پارامتریک نرمال بودن توزیع متغیرها است. معمولاً اگر مقدار این دو شاخص بین +۱ تا −۱ باشد نشان‌گر نرمال بودن توزیع داده‌ها است. البته ما انتظار نداریم توزیع همه نمرات در همه متغیره‌ها یا در همه جوامع آماری نرمال باشد. برای مثال متغیر قد در بین بازیکنان بسکتبال و والیبال و متغیر هوش و پیشرفت تحصیلی در مدارس تیزهوشان حتماً نرمال نخواهد بود و توزیع اغلب داده‌ها در سمت راست منحنی تراکم پیدا کرده و دنباله منحنی به سمت چپ کشیده شده و ما شاهد کجی منفی خواهیم بود. در چنین مواردی می‌توان از تبدیل‌های آماری استفاده کرده و مقدار کجی و کشیدگی را تعدیل کرد؛ اما انتظار می‌رود توزیع این متغیرها در بین افراد عادی



شکل ۱-۱: مفاهیم توزیع نرمال، کجی و کشیدگی (اقتباس از کلاین، ۲۰۱۱، ص. ۶۱)

1 -skewness
2 -kurtosis

جامعه و دانش‌آموزان مدارس معمولی نرمال باشد.

در شکل ۱-۱، مفاهیم توزیع نرمال، کجی (مثبت و منفی) و کشیدگی (مثبت و منفی) به شکل تصویری نشان داده شده است:

حداکثر توزیع داده‌ها، تعداد و مجموع نمرات: اغلب برای توصیف بهتر توزیع داده‌ها بهتر است، کوچک‌ترین و بزرگ‌ترین داده، تعداد کلی و مجموع نمرات گزارش شود. این کار باعث می‌شود متخصصان آمار شناخت کلی از توزیع داده‌ها داشته باشند و در صورت وجود اشکالات محاسباتی در داده‌ها، آنها را شناسایی کنند.

نمودارها: در علم آمار از دو ابزار جهت نمایش داده‌های انبوه استفاده می‌شود که یکی جداول توزیع فراوانی و دیگری نمودارها هستند. انواع مختلفی از نمودارها شامل میله‌ای، دایره‌ای برای داده‌هایی در سطح مقیاس اسمی و رتبه‌ای و نمودار هیستوگرام، چندضلعی، ساقه و برگ برای داده‌هایی در سطح مقیاس فاصله‌ای و نسبی وجود دارد. نمودارها وضعیت داده‌های توزیع را به شکل تصویری نشان می‌دهند لذا درک آنها برای افرادی که سواد آماری کمی دارند، آسان‌تر است.

در جدول زیر، انواع پرکاربرد شاخص‌های گرایش مرکزی آمار توصیفی همراه با روش محاسبه یا فرمول‌های آنها، ارائه شده است:

جدول شماره ۲-۱: مشخصات شاخص‌های گرایش مرکزی پرکاربرد در آمار توصیفی

نام شاخص	نام انگلیسی	نشان آماری	فرمول	توضیح فرمول
میانگین	Mean	$\bar{x}$ یا M	$\tilde{X}=\frac{\sum x}{n}$	مجموع نمرات تقسیم بر تعداد.
میانه	Median	MDیاMn	$Mn=L+\frac{\frac{N}{۲}-cf}{f}(i)$	L= حد پایین طبقه‌ای که میانه در آن قرار دارد. میانه در طبقه‌ای قرار دارد که برابر $\frac{N}{۲}$ است. N= تعداد نمونه Cf= فراوانی تراکمی تا طبقه‌ای که میانه در آن قرار دارد. F= فراوانی ساده طبقه‌ای که میانه در آن قرار دارد. I= فاصله طبقاتی توزیع
مد (نما)	Mode	Mo		عددی که بیشترین فراوانی را در توزیع نمرات دارد.
تعداد	number	N		تعداد نمرات
مجموع	sum	Σ		جمع همه نمرات

در جدول زیر، انواع پرکاربرد شاخص‌های پراکندگی در آمار توصیفی همراه با روش محاسبه یا فرمول‌های آنها، ارائه شده است:

جدول ۳-۱: مشخصات شاخص‌های پراکندگی پرکاربرد در آمار توصیفی

نام شاخص	نام انگلیسی	نشان آماری	فرمول	توضیح فرمول
واریانس	Variance	S^2	$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$	n = تعداد نمونه. $\sum x^2$ = مجموع تک تک نمرات توزیع که به توان دو رسیده‌اند. $(\sum x)^2$ = مجموع نمرات به توان دو
انحراف معیار (انحراف استاندارد)	Standard deviation	SDیاS یا σ	$s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$	توضیح عناصر فرمول دقیقه همان توضیح عناصر فرمول واریانس (خانه بالا) است.
دامنه تغییرات	range	R	R= max−min	حداکثر نمره منهای حداقل نمره
چارک اول	Quartile۱	Q_1	$Q_1 = L + \frac{\frac{N}{4} - cf}{f}(i)$	L= حد پایین طبقه‌ای که چارک مدنظر (اول، دوم یا سوم) در آن قرار دارد. N= تعداد نمونه Cf= فراوانی تراکمی تا طبقه‌ای که چارک مدنظر در آن قرار دارد. F= فراوانی ساده طبقه‌ای که چارک مدنظر در آن قرار دارد. I= فاصله طبقاتی توزیع
چارک سوم	Quartile۳	Q_3	$Q_3 = L + \frac{\frac{3N}{4} - cf}{f}(i)$	عناصر فرمول همانند چارک اول است.
چارک دوم	Quartile۲	Q_2	$Q_2 = \frac{Q_3 + Q_1}{2}$	چارک دوم برابر با مقدار میانه است.
نقاط دهکی: دهک اول	Decile point	D.P	$D_1 = L + \frac{\frac{1N}{10} - cf}{f}(i)$	شاخصی که توزیع نمرات را به ده قسمت مساوی تقسیم می‌کند. فرمول نقاط دهکی دیگر از طریق تغییر در صورت فرمول صورت می‌گیرد برای مثال نقطه دهکی ۶ از طریق تغییر به $\frac{6n}{10}$ صورت می‌گیرد.
نقاط صدکی	Percentile point	Px	$px = L + \left(\frac{pn - cf}{fi}\right)(i)$	شاخصی که توزیع نمرات را به صد قسمت مساوی تقسیم می‌کند. P صدک مدنظر مثلا صدک ۶۵ام.

جدول ۳-۱: مشخصات شاخص‌های پراکندگی پرکاربرد در آمار توصیفی *(ادامه)*

نام شاخص	نام انگلیسی	نشان آماری	فرمول	توضیح فرمول
کجی	skewness	s.k	$S.K=\frac{\left(\sum d^{۳}\right)\times N}{(SD)^{۳}(n-۱)(n-۲)}$ $S.K=\frac{m۳}{m۲\sqrt{m۲}}=\frac{\frac{\sum x^{۳}}{n}}{(\sqrt{\frac{\sum x^{۲}}{n}})^{۳}}$	میزان کجی توزیع نمرات را از توزیع نرمال نشان می‌دهد
کشیدگی	kurtosis	k.u	$k\mu=\frac{m_{۴}}{(m_{۲})۲}-۳=\frac{\frac{\sum x^{۴}}{n}}{\left(\frac{\sum x^{۲}}{n}\right)^{۲}}-۳$	میزان کشیدگی یا پهن بودن توزیع نمرات را از توزیع نرمال نشان می‌دهد x انحراف نمره از میانگین
حداقل	minumum	Min	فرمول ندارد	کوچک‌ترین یا کمترین عدد موجود در توزیع نمرات
حداکثر	maximum	Max	فرمول ندارد	بزرگ‌ترین یا بیشترین عدد موجود در توزیع نمرات

برای مثال، آمار توصیفی توزیع نمرات: "۱-۲-۳-۴-۵" عبارت خواهد بود از:

جدول ۴-۱: محاسبه شاخص‌های توصیفی توزیع نمرات "۱-۲-۳-۴-۵"

میانگین	$\frac{۱۵}{۵}=۳$
میانه	۳
مد (نما)	توزیع فاقد مد است، چون فراوانی همه اعداد برابر با یک است.
واریانس	$s^{۲}=\frac{۵\times ۵۵-(۱۵)^{۲}}{۵(۵-۱)}=۲/۵$
انحراف معیار (انحراف استاندارد)	$s=\sqrt{\frac{۵\times ۵۵-(۱۵)^{۲}}{۵(۵-۱)}}=۱/۵۸$
دامنه تغییرات	۵-۱=۴
چارک در داده‌های طبقه‌بندی نشده مثل مثال حاضر بهتر است ابتدا میانه را محاسبه کرده و سپس میانه داده‌های سمت چپ و سمت راست توزیع را که در حقیقت چارک اول و چارک سوم هستند، محاسبه کنیم. قبلاً هم گفته شده است که چارک دوم همان میانه است.	$Q_{۱}=\frac{۱+۲}{۲}=۱/۵$ $Q_{۳}=\frac{۴+۵}{۲}=۴/۵$ $Q_{۲}=\frac{۴/۵+۱/۵}{۲}=۳$

جدول ۴-۱: محاسبه شاخص‌های توصیفی توزیع نمرات "۱-۲-۳-۴-۵" *(ادامه)*

نقاط دهکی	به دلیل کمی تعداد نمونه (۵ مورد) از محاسبه نقاط دهکی صرف‌نظر کردیم.
نقاط صدکی	به‌دلیل کمی تعداد نمونه (۵ مورد) از محاسبه نقاط صدکی صرف‌نظر کردیم.
کجی	چون M_3 صفر می‌شود لذا مقدار کجی برابر صفر خواهد بود (نحوه محاسبه شاخص کجی در ادامه همین فصل و در بخش مفروضه‌های آزمون‌های پارامتریک مطرح شده است).
کشیدگی	$k\mu=\frac{m_4}{(m_2)2}-3=\frac{\frac{34}{5}}{\left(\frac{10}{5}\right)^2}-3=-1/30$
حداقل	۱
حداکثر	۵
تعداد	۵
مجموع	۱۵

محاسبه شاخص‌های آمار توصیفی در SPSS

جهت محاسبه شاخص‌های آماری توصیفی با استفاده از نرم‌افزار SPSS لازم است پس از نصب نرم‌افزار، داده‌ها را در یک ستون در صفحه data view وارد کرده و مسیر زیر را انتخاب کنید:

Analyze ⇒ Descriptive statistics ⇒ frequencies

در کادر باز شده، متغیر مورد نظر را از کادر سمت چپ به راست منتقل می‌کنیم و در همان پنجره گزینه statistics را کلیک می‌کنیم و از بین گزینه‌ای موجود در آن، گزینه‌های .std ،sum ،Mode ،Median ،Mean Quartiles، Kurtosis، skewness، maximum، minimum، Range، Variance، deviation را با گذاشتن علامت تیک در کنار این شاخص‌ها انتخاب کرده و گزینه continue را کلیک کرده و به پنجره اصلی برمی‌گردیم. جهت انتخاب نمودار گزینه chart را کلیک کرده و برحسب مقیاس داده‌ها، نمودار مناسب را انتخاب می‌کنیم. در نهایت ok می‌کنیم. خروجی SPSS به شکل زیر خواهد بود:

جدول ۵-۱: خروجی دستور فراوانی (frequencies) در نرم افزار SPSS

VAR00001

طبقه		Frequency (فراوانی)	Percent (درصد)	Valid Percent (درصد معتبر)	Cumulative Percent (درصد تراکمی)
Valid	1.00	1	20.0	20.0	20.0
	2.00	1	20.0	20.0	40.0
	3.00	1	20.0	20.0	60.0
	4.00	1	20.0	20.0	80.0
	5.00	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

در جدول فراوانی فوق، طبقات، فراوانی، درصد، درصد معتبر و درصد تراکمی توزیع داده‌ها نشان داده شده است. تفاوت بین درصد ساده با درصد معتبر در خصوص مقادیر داده‌های گمشده است. در شاخص درصد معتبر داده‌های گمشده محسوب نمی‌شوند.

در جدول زیر، شاخص‌های توصیفی شامل شاخص‌های مرکزی و پراکندگی توزیع داده‌ها نشان داده شده است:

جدول ۶-۱: خروجی SPSS مربوط به شاخص‌های توصیفی و پراکندگی

Statistics

VAR00001

N (تعداد)	Valid (تعداد معتبر)	5
	Missing (داده گمشده)	0
Mean (میانگین)		3.0000
Median (میانه)		3.0000
Mode (مد)		1.00[a]
Std. Deviation (انحراف معیار)		1.58114
Variance (واریانس)		2.500
Skewness (کجی)		.000
Std. Error of Skewness (خطای معیار کجی)		.913
Kurtosis (کشیدگی)		-1.200
Std. Error of Kurtosis (خطای معیار کشیدگی)		2.000
Range (دامنه تغییرات)		4.00
Minimum (حداقل)		1.00
Maximum (حداکثر)		5.00
Sum (جمع)		15.00
Percentiles	25	1.50
	50	3.00
	75	4.50

a. Multiple modes exist. The smallest value is shown

جامعه آماری و نمونه آماری

جامعه آماری[1] مجموعه‌ای از اشیا، افراد، اعداد و یا چیزهایی است که دارای یک یا چند ویژگی مشترک هستند. مثل همه دانش‌آموزان پایه سوم ابتدایی مدارس دولتی شهر یزد در سال تحصیلی ۱۳۹۵-۱۳۹۴ و یا کل مقالات منتشر شده اساتید دانشگاه آزاد در مجلات علمی و پژوهشی در سال ۲۰۱۶. نمونه آماری[2]،

1- statistical universe
2- statistical population